UNIVERSITY OF CALIFORNIA,
IRVINE

The Genomics of Experimental Evolution

DISSERTATION

submitted in satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Biological Sciences

by

Kathreen Eve Bitner

Dissertation Committee:
Professor Laurence D. Mueller, Chair
Professor Michael R. Rose
Associate Professor Jose M. Ranz

2020

Dedication

To my parents, Mariola and Chris Bitner and
my grandma, Aniela Boiska,
for your unconditional love and support.

To

My advisor Dr. Laurence Mueller

And to

The undergraduate researchers of the Rose and Mueller lab

for all your hard work

TABLES OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Laurence Mueller, for his endless support, guidance, and inspiration during my masters and dissertation. I am grateful to my committee members, Dr. Michael R. Rose and Dr. Jose Ranz, for always giving great advice on my research and providing me with many resources to complete my experiments.

Thank you to Dr. Parvin Shahrestani of the College of Natural Sciences and Mathematics at CSU Fullerton. It was a privilege to work together on Chapter 1, as her knowledge and background was very helpful in the design of the experiment. Thank you to Dr. Larry Cabral, my own graduate advisor while I was completing my undergraduate studies at UCI, and whose guidance was fundamental to my completion of my graduate degree.

In addition, I would like to acknowledge the graduate students of the Rose and Mueller Laboratories for their help. Specifically, Dr. Mark A. Philips, Dr. Grant Rutledge, Dr. Thomas T. Barter, Dr. James N. Kezos, Marjan Koosha, and Kevin H. Phung.

A big thank you to the administrative staff in the Ecology and Evolutionary Department at UCI for always being there to help, explain and make sure the logistical aspects of my dissertation were completed.

To my Dad, Chris Bitner, for his love and support during my life in all my endeavors. To my Mom, Mariola Bitner, for not only being my backbone and strength in times of need, but for always willing to help in any way that I needed and accompanying me to complete experiments at 2 am. To my grandma, Aniela Boiska, whose enthusiasm for watching larvae eat and learning about the fruit flies of my experiments is unparalleled.

A huge thanks to the undergraduate researchers of the Rose and Mueller Labs. This research would not have been possible without your hard work and dedication.

CURRICULUM VITAE

Kathreen Eve Bitner

EDUCATION

2016 – 2020	Ph.D., Biological Sciences, University of California, Irvine
2013 – 2015	Masters, Biological Sciences, University of California, Irvine
2008 – 2012	B.S., Ecology and Evolutionary Biology, University of California, Irvine
2008 – 2012	B.A., Music Performance, University of California, Irvine

RESEARCH

2010-2012, 2013-2015, 2016-2020	Dr. Mueller & Dr. Rose lab University of California, Irvine Life History Evolution and Ecology and Evolutionary Biology.

I used experimental evolution to study the trade-offs that occur in the *Drosophila melanogaster* populations over time and how they adapt to new, stressful environments that contain elevated ammonia and urea levels in larval food and crowding conditions for my Masters. For my doctorate, I studied the Smurf assay for identifying a physical characteristic of the death spiral. I also compared the urea-adapted populations of *D. melanogaster* to study their phenotypic and genomic changes.

TEACHING EXPERIENCE

2013-2015, 2016-2020	Graduate Teaching Assistant. Department of Ecology and Evolutionary Biology, University of California, Irvine **Bio 97 Genetics (6 quarters)** **Bio E179L Limnology and Freshwater Lab (6 quarters)** **Bio 94 Organismal Biology (4 quarters)** **Bio E179 Limnology and Freshwater Ecology (2 quarters)** **Bio E112L Physiology lab (2 quarters)** **Bio E106 Ecology and Evolution (1 quarter)**

APPOINTMENT HISTORY:

2018	Laboratory Manager Laboratory of Dr. Michael Rose and Dr. Laurence Mueller University of California, Irvine

2016 – 2020 Graduate Teaching Assistant, Department of Ecology and Evolutionary Biology, University of California, Irvine

2013 – 2015 Graduate Teaching Assistant, Department of Ecology and Evolutionary Biology, University of California, Irvine

MENTORING EXPERIENCE:

2013 – 2020 Mentoring undergraduate researchers.
Laboratory of Dr. Laurence Mueller,
University of California, Irvine

PUBLICATIONS

Bitner, K., Mueller L.D. 2015. The Evolution of Ovoviviparity in a Temporally varying environment. The American Naturalist 186, Vol. 6

Bitner K, Shahrestani P, Pardue E, Mueller LD. 2020. Predicting death by the loss of intestinal function. PLoS ONE 15(4): e0230970. https://doi. org/10.1371/journal.pone.0230970

ABSTRACT OF THE DISSERTATION

The Genomics of Experimental Evolution

By

Kathreen Eve Bitner

Doctor of Philosophy in Biological Sciences

University of California, Irvine, 2020

Professor Laurence Mueller, Chair

This dissertation focused on two central theories in experimental evolution; studying the death spiral and looking at trade-offs in adapted populations to a toxin in their environment. These results are summarized in Chapter 1. Chapter 2 focused on quantifying the efficacy of ingested dyes (using the 'Smurf' assay) to predict when an individual would die. This can be extremely useful for many research problems in aging. The death spiral is a short period prior to death that is marked by a dramatic decline in physiological health. The results show three key conclusions: that all blue dyes used had significant negative effects on mean longevity, only a small fraction of the flies showed the Smurf phenotype prior to death, and among the small fraction that did become Smurfs most (40-60%) become blue during their last 24 hours of life. The research in Chapters 3-5 focused on the phenotypic and genomic evolution of urea adapted populations. Insects are often exposed to many toxic substances in their environments. Urea and ammonia differ from most studies of resistance to toxic pesticides because of their wide-ranging effects on organismal and cellular physiology. The UX and UTB selection groups (adapted to high levels of urea in the larval stage of development), displayed slower larval feeding rates, slower larval growth rates, and lower starvation resistance compared to the other populations reared that had

not adapted to urea. The UX and UTB selection regimes also showed higher viability, faster development time, and lower starvation resistance when reared on food supplemented with urea compared to controls. Chapter 4 of this dissertation studied genomic differentiation among populations subjected to different types of selection for tolerance to urea in their larval food. The RUX, reverse selected of the urea adapted populations still maintained measurable levels of adaptation to urea suggesting any reverse selection is slow. We saw parallel evolution occurring in the replicates of the same environmental populations, as well as genomic differentiation in the urea adapted lines, UX, UTB and RUX versus the control population of AUC. The starting conditions for the important gene regions might have been important and very different for the UX and UTB populations. These differences may account for why these populations share only 50% of the same significantly differentiated SNPS (single nucleotide polymorphisms). Chapter 5 sought to connect potential genes as the causative agents for specific phenotypes studied in Chapter 3. FLAM, the fused lasso additive model, is a statistical learning tool for determining which genes may affect differentiated phenotypes. The FLAM analysis provided us with 53 SNPS that had large effects on the four phenotypes studied – larval feeding rate, larval growth rate, development time and viability. Almost 500 genes were located to be potentially responsible for the 53 SNPS identified by FLAM.

CHAPTER 1

Introduction

My dissertation centers on two different research topics. Chapter 2 was focused on the potential SMURF assay technique for identifying when a *Drosophila melanogaster* fruit fly would be in the death spiral using a blue dye. After the completion of the experiment, my research was the focused on the urea-adapted populations of flies in the Rose and Mueller lab, and the phenotypic (Chapter 3), and genomic (Chapter 4) changes that occurred as a result of their adaptation to a toxin in the environment. Finally, Chapter 5 focused on using the Fused Lasso Additive Model, FLAM to distinguish significant SNP's that may directly affect specific urea adapted phenotypes (Mueller 2018).

Chapter 2 focused on quantifying the efficacy of ingested dyes to predict when an individual would die. This can be extremely useful for many research problems in aging. Under protected conditions, late-life is characterized by a plateau in age-specific mortality (Carey *et al.* 1992, Curtsinger *et al.* 1992, Rose *et al.* 2006, Vaupel *et al.* 1998) female fecundity (Mueller *et al.* 2007, Rauster *et al.* 2005), male virility (Shahrestani *et al.* 2012), and age related motor performance decline and specific late-life motor disabilities (Gaitanidis *et al.* 2019). It has been suggested there is a fourth stage of adult life called the death spiral (Mueller *et al.* 2007, Rauser *et al.* 2005, Mueller *et al.* 2016). The death spiral is a short period prior to death that is marked by a dramatic decline in physiological health. There is evidence of this decline in fecundity (Rauser *et al.* 2005, Muller *et al.* 2001, Rogina *et al.* 2007), supine behavior (Papadopoulos *et al.* 2002), activity and desiccation resistance (Shahrestani *et al.* 2012a), and male virility (Shahrestani *et al.* 2012b). We have previously shown that the decline in fecundity can be used to predict death (Mueller *et al.* 2016). We were looking for a better and faster way to predict death as current phenotypic methods are cumbersome.

A process described by Rera *et al.* 2012 as the 'Smurf Assay' would have presented an easier and faster way to visually observe a fly in the death spiral. According to (Rera *et al.* 2012) and (Martins *et al.* 2018) individuals fed food with a blue dye (FD&C blue dye #1) would maintain their ability to prevent the dye from permeating the intestinal barrier until a few days before death. At that time the entire body of the fly would become blue, leading Rera *et al.* 2012 to identify the individuals so colored as "Smurfs" (Dambrose *et al.* 2016, Darby *et al.* 2019, Kissoyan *et al.* 2019, Klichko *et al.* 2019, Liang *et al.* 2019, Rodriguez-Fernandez *et al.* 2019). The original paper Rera *et al.* 2012, presented no quantitiatve demographic data for this assay, and we set out to run a large, multi-population, multi-dye experiment to see if this method would have useful applications we were hoping for in gene expression studies.

The study in Chapter 2 established a number of important conclusions. (1) All dyes used had significant negative effects on mean longevity, with decreases ranging from 5 to almost 10 days. (2) Only a small fraction of the flies showed the Smurf phenotype prior to death. Over all populations and dyes, 22% of males and 34% of females became Smurfs. (3) Among the small fraction that did become Smurfs most (40-60%) become blue during their last 24 hours of life. Thus, even with daily checks most of the Smurf flies would be dead when initially identified as Smurf making their utility for gene expression studies nil. There was substantial replication, with each population undergoing 6 different dyes and a control. These results would tend to limit the utility of this method depending on the application the method was intended for.

After the completion of the Smurf experiment in Chapter 2, my research changed to focusing on the urea-adapted populations (Chapter 3 – 5). Insects are often exposed to many toxic substances in their environments, which include insecticides, heavy metals, chemicals or feeding deterrents synthesized by plants. Urea and ammonia differ from most studies of

resistance to toxic pesticides because of their wide-ranging effects on organismal and cellular physiology (Somero and Yancey, 1997). Fitness components in flies are frequently correlated with traits that confer the ability to resist and survive various stresses (Prasad and Joshi 2003, Flatt 2020). Larval competitive ability in *D. melanogaster* depends on many factors, such as feeding rate, initial weight, relative time spent on molting, minimum food requirement for pupation, and resistance to crowding. Past research has shown that feeding rates in *Drosophila* larvae are affected by environmental factors

An important model looked at the evolution of feeding rates in *D. melanogaster* larvae in stressful environments found they would decrease in response to adaptation to high levels of ammonia, urea, and the presence of parasitoids (Mueller *et al* 2015). If there are toxic compounds in the larval food, energy is required for detoxification, and larvae can maximize food intake by slowing their feeding rate (Mueller *et al* 2015). While it may be argued that a slower feeding rate does not necessarily prove that the larva will grow slower, previous studies (Joshi and Mueller 1996; Mueller *et al*. 1991) showed that fast feeding rate larvae required more food to reach the same critical minimum size. Mueller *et al* 1990 showed a negative correlation between feeding efficiency and feeding rate.

Our goal in Chapter 3 was not only to study larval feeding rates of our urea-adapted lines, but to also look at various other phenotypes, such as viability, development time, starvation and desiccation resistance, larval growth rate and adult size. The UX and UTB selection groups displayed slower larval feeding rates, slower larval growth rates, and lower starvation resistance compared to the other populations reared that had not adapted to urea. The UX and UTB populations also showed higher viability, faster development time, and lower starvation resistance when reared on food supplemented with urea compared to controls. With the

significant slower feeding rates of the urea adapted lines – UTB, UX and RUX, it is quite clear that larval feeding rates evolve in response to environmental factors, and are correlated with a slower growth rates.

After establishing various phenotypic adaptations in the urea adapted selection groups, next we looked at the genomic aspect of these populations in Chapter 4. Genome-wide sequencing of experimentally evolved populations has emerged as a powerful method for understanding the genetic basis of adaptation (Burke *et al.* 2010; Turner *et al.* 2011; Tenaillon *et al.* 2012; Schlotterer *et al.* 2015, Graves *et al*. 2017). Previous work has shown that *Drosophila* populations can evolve resistance to high levels of environmental urea (Joshi *et al.* 1996, Shiotsugu *et al.*, 1997, David *et al.* 1999, Etienne *et al.* 2001), a compound that stresses the *Drosophila* larvae.

Chapter 4 of this dissertation studied genomic differentiation among populations subjected to different types of selection for tolerance to urea in their larval food. The results, we showed some depressions in heterozygosity in all four populations, but very few regions had variations that had been completely expunged. The RUX populations still maintained measurable levels of adaptation to urea suggesting any reverse selection is slow. We saw parallel evolution occurring in the replicates of the same environmental populations, as well as genomic differentiation in the Urea adapted lines, UX, UTB and RUX versus the control population of AUC. Of the 70,980 differentiated SNPs in the UX populations 50.8% were also differentiated in the UTB populations. The starting conditions for the important gene regions might have been very different and this may have led to different trajectories for this variation once the environment was changed.

Chapter 5 combined the phenotypic results of Chapter 3 with understanding the potential genes that might be playing a role in the adaptation to the urea toxin. Understanding the relationship between genes and phenotypes is the base of much genetic and experimental evolutionary work. Lewontin's (1974) goal was understanding this relationship, which now may be possible with genome-wide sequencing in the field of evolutionary biology. Other experiments have suggested that the genomic response to selection can involve many selected SNPs that show unexpectedly complex evolutionary trajectories, possibly due to nonadditive effects. (Wengel et al 2012). The efficacy of FLAM, the Fused Lasso Additive Model, is improved with increased number of independent populations, reduced environmental phenotypic variation, and increased within-treatment among-replicate variation (Mueller *et al.* 2018). With the simulations in FLAM, we are picking up the genes that have the most direct impact on a phenotype.

FLAM was applied to SNP variation measured in 40 populations of *D. melanogaster* – 20 of which are the urea adapted selection regimes and controls, UX, UTB, RUX and AUC, and 20 that make up the demographic lines subjected to selection for age-at reproduction or starvation/ desiccation resistance – CO, nCO, TSO, and TDO. The FLAM analysis provided us with 53 SNPS that had large effects on the four phenotypes studied – larval feeding rate, larval growth rate, development time and viability. Almost 500 genes were located to be potentially responsible for the significant SNPS differentiated.

Three of the four phenotypes compared had high correlation coefficients between the FLAM predictions and the observed phenotypes. These high correlations were based on separating observations into testing sets, used to estimate FLAM parameters, and prediction sets. Larval feeding rate had the highest correlation coefficient. The lowest correlation coefficient for

the larval growth rate was 0.64. When we looked directly at the growth rate parameter among the 40 populations (Chapter 3), there were no significant differences. Genes with smaller influences that are not being picked up by the FLAM analysis, may be playing role in the correlation of a faster feeding rate with a faster growth rate (Chapter 3).

Three SNP's were found to be pleiotropic, affecting the viability and development time phenotype. These two phenotypes reflect the physiological impact of urea on either survival or development time also show overlapping genes. The other two phenotypes, larval feeding rate and larval growth rate are not necessarily a response to the same stress. While we don't see extensive pleiotropy of genes, we are also dealing with a limitation of FLAM, to identify more significant SNPs than the number of pupations examined. Thus, our analysis might be missing many pleiotropic genes limiting our ability to make strong inferences about their relative importance.

The analysis of genetic variation at the *for* locus showed that this variation was helpful in clustering the UX and RUX vs. the AUC populations, but did a poor job differentiating the UTB and all the demographic populations. The UX populations (and its close relative, the RUX populations) may have taken a different evolutionary trajectory than the UTB populations.

As the FLAM technique gets modified and more advanced, or replaced by a different analytical technique, more genes may be discovered with attributions to the four physiological characteristics we studied – larval feeding rate, development time, viability and larval growth rate.

REFERENCES

Burke MK, Dunham JP, Shahrestani P, Thornton KR, Rose MR, Long AD. 2010. Genome-wide analysis of a long-term evolution experiment with Drosophila. Nature 467:587–590.

Carey JR, Liedo P, Orozco D, Vaupel JW (1992) Slowing of mortality rates at older ages in large medfly cohorts. Science. 1992 Oct 16;258(5081):457-61. doi: 10.1126/science.1411540

Curtsinger JW, Fukui HH, Townsend DR, Vaupel JW. Demography of genotypes: failure of the limited life-span paradigm in Drosophila melanogaster. Science. 1992 Oct 16;258(5081):461-3. doi: 10.1126/science.1411541

Dambroise E, Monnier L, Ruisheng L, Aguilaniu H, Joly J.-S, Tricoire H, et al. Two phases of aging separated by the Smurf transition as a public path to death. Sci Rep6, 23523 (2016) doi:10.1038/srep23523

Darby TM, Owens JA, Saeedi BJ, Robinson BS, Naudin CR, Jones RM. Lactococcus Lactis Subsp. cremoris Is an Efficacious Beneficial Bacterium that Limits Tissue Injury in the Intestine. iScience 2019; 12: 356– 367. doi: 10.1016/j.isci.2019.01.030.

David, C.L., Pierce, V.A., Aswad, D.A., Gibbs, A.G., 1999. The effect of urea exposure on isoaspartyl content and PIMT activity in Drosophila melanogaster. Comparative Biochemistry and Physiology B. 124(4):423-7

Etienne, R., Fortunat, K., V. Pierce. 2001. Mechanisms of urea tolerance in urea-adapted populations of *Drosophila melanogaster*. The Journal of Experimental Biology 204, 2699–2707

Gaitanidis A., Dimitriadou A., Dowse H., Sanyal S. Duch C., Consoulas C. (2019). Longitudinal assessment of health-span and pre-death morbidity in wild type Drosophila. Aging, 11:1850-1873).

Graves, J.L, K.L. Hertweck, M.A. Phillips, M.V. Han, L.G. Cabral, T.T. Barter, L.F. Greer, M.K. Burke, L.D. Mueller, and M.R. Rose. 2017. Genomics of experimental evolution. Mol. Biol. Evol. 34(4): 831-842

Joshi, A, J. Shiotsugu, L.D. Mueller. 1996. Phenotypic enhancement of Longevity by environmental Urea in *Drosophila melanogaster*. Experimental Gerontology. 31:4, 533-544

Kissoyan KAB, Drechsler M, Stange EL, Zimmerman J, Kaleta C, Bode HB, Dierking K. Natural *C. elegans* Microbiota Protects against Infection via Production of a Cyclic Lipopeptide of the Viscosin Group. Current Biology 2019; 29: 1030–1037

Klichko VI, Safonov VL, Safonov MY, Radyuk SN. Supplementation with hydrogen-producing

composition confers beneficial effects on physiology and life span in Drosophila. Heliyon 5 (2019) E01679.

Lewontin RC. 1974. The Genetic Basis of Evolutionary Change. New York: Columbia University Press.

Liang ST, Audira G, Juniardi S, Chen JR, Lai YH, Du ZC, et al. Zebrafish Carrying pycr1 Gene Deficiency Display Aging and Multiple Behavioral Abnormalities. Cells 8: 453 (2019) doi: 10.3390/cells8050453

Martins RR, McCracken AW, Simons MJP, Henriques CM, Rera M. How to catch a Smurf? in vivo assessment of intestinal permeability in multiple model organisms. Bio Protoc. 2018; 8:e2722. doi: 10.21769/BioProtoc.2722

Mueller LD (1990) Density-dependent natural selection does not increase efficiency. Evol Ecol 4:290–297

Mueller LD, Gonza´lez-Candelas F, Sweet VF (1991) Components of density-dependent population dynamics: models and tests with Drosophila. Am Nat 137:457–475

Mueller LD, Rauser CL, Rose MR. An evolutionary heterogeneity model of late-life fecundity in Drosophila. Biogerontology. 2007 Apr;8(2):147-61.

Mueller, L.D. and T.T. Barter. 2015. A model of the evolution of larval feeding rate in Drosophila driven by conflicting energy demands. Genetica 143:93-100

Mueller LD, Shahrestani P, Rauser CL, Rose MR. The death spiral: predicting death in Drosophila cohorts. Biogerontology. 2016 Nov;17(5-6):805-816. Epub 2016 Feb 25

Mueller LD, Phillips MA, Barter TT, Greenspan ZS, Rose MR. 2018. Genome-wide mapping of gene phenotype relationships in experimentally evolved populations. Mol Biol Evol 35(8):2085–2095

Muller HG, Carey JR, Wu D, Liedo P, Vaupel JW. Reproductive potential predicts longevity of female Mediterranean fruit flies. Proc Biol Sci. 2001 Mar 7; 268(1466): 445–450, doi: 10.1098/rspb.2000.1370

Papadopoulos NT, Carey JR, Katsoyannos BI, Kouloussis NA, Muller HG, Liu X. Supine behaviour predicts the time to death in male Mediterranean fruitflies (Ceratitis capitata). Proc Biol Sci. 2002 Aug 22;269(1501):1633-7

Prasad, N. G., and A. Joshi, 2003 What have two decades of laboratory life-history evolution studies on Drosophila melanogaster taught us? J. Genet. 82: 45–76. https://doi.org/10.1007/BF02715881

Rauser CL, Abdel-Aal, Y, Shieh JA, Suen CW, Mueller LD, Rose MR. Lifelong heterogeneity in

fecundity is insufficient to explain late-life fecundity plateaus in Drosophila melanogaster. Exp Gerontol. 2005 Aug-Sep;40(8-9):660-70. doi:10.1016/j.exger.2005.06.006

Rera M, Clarck RI, Walker DW. Intestinal barrier dysfunction links metabolic and inflammatory markers of aging to death in Drosophila. Proc Natl Acad Sci U S A. 2012 Dec 26;109(52):21528-33. doi: 10.1073/pnas.1215849110.

Rodriguez-Fernandez IA, Qi Y, Jasper H. Loss of a proteostatic checkpoint in intestinal stem cells contributes to age-related epithelial dysfunction. Nat Commun 10, 1050 (2019) doi:10.1038/s41467-019-08982-9

Rogina B, Wolverton T, Bross T, Chen K, Muller H-G, Carey JR. Distinct biological epochs in the reproductive life of female Drosophila melanogaster. Mech Ageing Dev 128, 477–485, doi:10.1016/j.mad.2007.06.004 (2007).

Rose MR, Drapeau MD, Yazdi PG, Shah KH, Moise DB, Thakar RR, et al. Evolution of late-life mortality in Drosophila melanogaster. J Evol Biol. 2006 Jan;19(1):289-301. doi: 10.1111/j.1420-9101.2005.00966.x

Schlotterer C, Kofler R, Versace E, Tobler R, Franseen SU. 2015. Combining experimental evolution with next-generation sequencing: a powerful tool to study adaptation from standing genetic variation. Heredity 114:331–440.

Shahrestani P, Tran X, Mueller LD,. Patterns of male fitness conform to predictions of evolutionary models of late life. J Evol Biol. 2012a Jun;25(6):1060-5. doi: 10.1111/j.1420-9101.2012.02492.x

Shahrestani P, Tran X, Mueller LD,. Patterns of male fitness conform to predictions of evolutionary models of late life. J Evol Biol. 2012b Jun;25(6):1060-5. doi: 10.1111/j.1420-9101.2012.02492.x

Shiotsugu, J., Leroi, A.M., Yashiro, H., Rose, M.R., Mueller, L.D., 1997. The symmetry of correlated selection responses in adaptive evolution: an experimental study using *Drosophila*. Evolution 51, 163–172.

Somero, G.N., Yancey, P.H., 1997. Osmolytes and cell-volume regulation: physiological and evolutionary principles. In: Hoffman, F.F., Jamieson, J.D. (Eds.), Handbook of Physiology. Oxford University Press.

Tenaillon O, Rodriguez-Verdugo A, Gaut RL, McDonald P, Bennett AF, Long AD, Gaut BS. 2012. The molecular diversity of adaptive convergence. Science 335:457–461.

Turner TL, Steward AD, Fields AT, Rice WR, Tarone AM. 2011. Population-based resequencing of experimentally evolved populations reveals the genetic basis of body size variation in Drosophila melanogaster. PLoS Genet. 7:e10001336.

Vaupel JW, Carey JR, Christenson K, Johnson TE, Yashin AI, Holm NV, et al. Biodemographic trajectories of longevity. Science. 1998 May 8;280(5365):855-60. doi: 10.1126/science.280.5365.855

Orozcoter Wengel P, Kapun M, Nolte V, Kofler R, Flatt T, Schlotterer C. 2012. Adaptation of Drosophila to a novel laboratory environment reveals temporally heterogeneous trajectories of selected traits. Mol Ecol. 21:4931–4941.

CHAPTER 2

Predicting death by the loss of intestinal function

Bitner K, Shahrestani P, Pardue E, Mueller LD (2020) Predicting death by the loss of intestinal function. PLoS ONE 15(4): e0230970. https://doi. org/10.1371/journal.pone.0230970

Abstract

The ability to predict when an individual will die can be extremely useful for many research problems in aging. A technique for predicting death in the model organism, *Drosophila melanogaster*, has been proposed which relies on an increase in the permeability of the fly intestinal system, allowing dyes from the diet to permeate the body of the fly shortly before death. In this study, we sought to verify this claim in a large cohort study using different populations of *D. melanogaster* and different dyes. We found that only about 50% of the individuals showed a visible distribution of dye before death. This number did not vary substantially with the dye used. Most flies that did turn a blue color before death did so within 24 hours of death. There was also a measurable effect of the dye on the fly mean longevity. These results would tend to limit the utility of this method depending on the application the method was intended for.

Introduction

Evolutionary biologists recognize three phases of adult life in organisms that reproduce multiple times. The first phase occurs prior to reproduction and can be called development. During this phase, we expect natural selection to oppose any genetically based reductions in survival since death at these ages means zero fitness. In the second phase, called aging, the strength of natural selection declines with age as first outlined by Hamilton [1]. Under protected conditions, we typically see an age-dependent increase in mortality and a decline in fertility [2]. Finally, at advanced ages organisms enter late-life [3 - 6]. Again, under protected conditions late-life is characterized by a plateau in age-specific mortality [3-5, 7] female fecundity [6, 8], male virility [10], and age related motor performance decline and specific late-life motor disabilities [11].

Recently we have suggested there is a fourth stage of adult life called the death spiral [6**,** 8, 12]. The death spiral is a short period prior to death that is marked by a dramatic decline in physiological health. There is evidence of this decline in fecundity [8, 13 -14], supine behavior [15], activity and desiccation resistance [9], and male virility [10]. We have previously shown that the decline in fecundity can be used to predict death [12]. We are looking for a better and faster way to predict death as current phenotypic methods are cumbersome.

Additional study of the death spiral and a more detailed understanding of the physiological systems that are under decline could be done if there was a reliable and easy way to identify individuals that were about to die. This would permit one to do destructive assays on individuals in the death spiral such as gene expression studies and compare them to similarly aged individuals who are not about to die. Rera *et al.* [16] describes such a process for *D. melanogaster*. According to [16] and [17] individuals fed food with a blue dye (FD&C blue dye

#1) will maintain their ability to prevent the dye from permeating the intestinal barrier until a few days before death. At that time the entire body of the fly will become blue, leading [16] to identify the individuals so colored as “Smurfs” [18 – 23]

This technique, in principle, offers exactly the assay needed for more detailed analysis of the death spiral. Unfortunately, prior work with the technique has not laid out any detailed analysis of the demographic features of the Smurf phenotype. For instance, what is the average and distribution of the time interval between becoming a Smurf and death? Do these properties change with chronological age? In large samples what fraction of flies become Smurfs prior to death? How does the appearance of the Smurf phenotype vary with other dyes and different populations of *D. melanogaster*? The goal of this study is to answer these questions.

Methods

Populations

Five large independent populations of *D. melanogaster* were used in this experiment. Two of these populations, ACO and CO, are large, outbred populations that have been maintained on different age-at-reproduction schedules for hundreds of generations. The ACO population was maintained on 9-day discrete generation cycles. The CO population was kept on 28-day discrete generation cycles. The remaining populations, S93, A4 3852 and Canton S (CAS), were inbred lines raised on three week cycles in the Long lab at the University of California, Irvine. All populations were raised in identical conditions of temperature, food, cultures and density for three generations prior to these experiments.

Mortality Assay

Adult, 14 day old (from egg) flies were knocked out with CO_2 gas and placed into individual plastic straws about 4 inches in length and capped with plastic pipette tips on both ends (Fig. 2.1). During anesthetization, a steady supply of CO_2 was flowing through a semi-porous plate. The flies were placed on the plate and separated by gender and each fly was gently swept into the plastic straw using a fine painters brush. An equal number of females and males were used per population. Food was provided to each fly at one end of the straw. Each fly was transferred to a new straw with new food and new pipette tips every 3 days to maintain a clean environment. The straw length and girth permitted individuals to fly from one end to the other.

The process of transferring the flies, as well as daily checking of the flies, required a light tapping of the fly into the pipette tip. Cohorts of about 56 adult flies, equal numbers of males and females from each of the five populations were exposed to either control food or food with one

of six dyes (table 2.1) added to their food. Substantial replication was used. Thus, the original dye, SPS Alfachem, was replicated in 5 different populations, and each population was replicated in 6 different dye environments. The use of different FDA FD&C Blue dye #1's permitted us to determine if the development of the SMURF phenotype was sensitive to the particular dye used. By using a combination of different populations of *D. melanogaster*, which varied in levels of inbreeding, we could determine if the development of the SMURF phenotype was limited to inbred populations.

The flies were exposed to the blue dyes from day 14 (from egg) continuously to their death. Each fly was individually checked underneath a microscope and light to see if it had become a 'Smurf'. Smurf status required that the entire body changed to any variation of a blue color. This was an important distinction as all the *Drosophila* flies fed food with a blue dye would have visible blue coloring in only the gut portion when they weren't a Smurf. Some of the dyes resulted in a slight variation in blue color in the Smurfs. Every day under a microscope with a light we looked for any change of color in the fly thorax, head and abdomen. If the fly was any shade of blue in all three sections, it was marked as a Smurf and was then checked daily to see when it died. We did not limit our observations to individual sections of the fly, such as only the thorax, for our evaluation of when a fly became a Smurf.

Tapping

We did the tapping experiment to see if the physical disruption, the process of tapping the fly into the pipette tip, affected the mean longevity and lifespan of the fly. A total of 164 ACO flies were chosen for this assay – 83 males and 81 females. The 164 flies were placed into regular food straws with no dye. A total of 84 flies (42 male and 42 female) were tapped 5 times daily, mimicking the checking that occurred in the original experiment, and the other 81 (41

males and 39 females) flies were not tapped. The flies were transferred to new straws, with fresh food and new pipette tips every 3 days. Each fly was checked daily for movement and if no movement was detected, the fly was classified as deceased on that day. Only ACO flies were used as the purpose of the Tapping experiment was to see if our methods for checking for Smurf flies would affect the mean longevity of the fly.

Food & Dyes

Flies were provided with banana-molasses food with one of the dyes added. The control flies received only banana molasses food in their respective straws. The recipe for the banana molasses food used in the lab, as well as the experiment, can be found in the Supplemental Portion. Food with dye was prepared by mixing 2.5 grams of each dye to create a 100-ml solution of the banana molasses food mixed with the dye (2.5% wt/vol). Food was always prepared the day before it was needed and stored in a refrigerator until it was used. The dyes were kept separate and carefully handled so no cross-contamination occurred during the preparation and food blending process.

Statistical Analysis

To analyze the effects of dye, sex and population on longevity we let y_{ijkl} be the age at death of the lth individual of sex-i (i=1 (female), 2 (male)), treatment-j (j=1,,..,7 (see table 1, 7=control)), and population-k (k=1,..,5 (see table 1)). Then a linear model for longevity is,

$$y_{ijkl} = \alpha + \delta_i\beta + \delta_j\gamma_j + \delta_k\pi_k + \delta_j\delta_k\theta_{jk} + \varepsilon_{ijkl}, \quad (1)$$

where δ_s=0 if s=1, and 1 otherwise, ε_{ijkl} is an error term assumed to have normal distribution with mean 0 and variance σ^2. An initial test showed no significant differences between sexes so the

final model tested did not include the β parameter. These tests were run with R (version 3.4.3, R Core team, 2017) and the *lm* function. Pairwise tests with Bonferroni corrections for simultaneous tests were conducted with the R *emmeans* function.

At the time of death each fly was classified according to their sex, population, treatment, and Smurf status (blue: yes or no). Using hierarchical log-linear models (*loglm* function in the R MASS package) we tested in succession whether sex, treatment, and population would have an effect on Smurf status at the time of death.

A *t*-test was performed on the Tapping Experiment results, comparing the mean longevity of the tapped flies versus the non-tapped flies to see if the mechanical disruption would affect their mean longevity.

RESULTS

We tested the difference in mean longevity for each population in the control environment vs. each of six dyes yielding a total of 30 hypothesis tests (table S2.1, Fig. S2.1). We found the control populations lived longer in all cases and 7 out of 30 of these tests were significantly different (using a Bonferroni correction for multiple testing). The log-rank tests detected 9 significant differences (table S2.2). Median longevity and maximum longevity were also calculated for each population-dye combination (table S2.3 and S2.4). Averaged over the six dye- treatments there were significant differences in mean longevity between all populations and their controls except CO and s93 (p=0.16, Bonferroni corrected for 5 different tests). Averaged over the five different populations, the control treatments lived significantly longer than every other dye treatment. The controls lived from 4.9 to 9.8 days longer, depending on which dye they were compared to, or about 8% to 17% of the control fly mean longevity (Fig. 2.2).

All FD&C blue dyes did show a Smurf phenotype, though the fraction of flies becoming Smurfs varied considerably from dye to dye, as well as among populations (Fig. 2.3 and Table S2.5). These results are consistent with a preliminary study we conducted on 172 ACO flies (Fig. S2.1).

The null loglinear model with no interactions was compared to a model with an interaction between sex and Smurf status and showed that sex has a significant effect on whether a fly becomes a Smurf ($\chi_1^2 = 30$, p<10^{-5}). Averaged over all blue dye treatment populations, 22% of the males became Smurfs and 34% of the females became Smurfs. If we add an interaction between dye treatment and Smurf status to the previous model with the sex interaction there is a significant effect of dye treatment ($\chi_5^2 = 13.0$, p< 0.022). Finally, adding an interaction between population and Smurf status to the previous model with interactions between

sex, treatment and Smurf status there is a significant effect of population ($\chi_4^2 = 28$, p=0.00001). Thus, achieving the Smurf phenotype before death is significantly affected by sex, dye, and population. However, the majority of flies never showed the Smurf phenotype prior to death.

The flies became a distinct Blue color in their abdomen, thorax and head when they became a 'Smurf' (Fig 2.1). This could be seen in some flies as much as 3 or 4 days before their death. However, of the flies that became a Smurf, the majority did so on the day of their death or one day before death (Fig. 2.4).

A *t*-test was run on the tapping experiment, comparing the mean longevity of the tapped flies versus the non-tapped flies to see if the mechanical disruption would affect their mean longevity. The males were not affected by tapping, with a mean longevity of 53.2 days for those tapped and 53.9 days for those not tapped (p=0.83, Table 2.2). Likewise females were not affected due to the tapping mechanism either, with a mean longevity of 54.5 days for the tapped females and 54.4 days for the untapped females (p=0.99, Table 2.2).

Discussion

This study has established a number of important conclusions. (1) All dyes used have significant negative effects on mean longevity, with decreases ranging from 5 to almost 10 days. (2) Only a small fraction of the flies show the Smurf phenotype prior to death. Over all populations and dyes 22% of males and 34% of females became Smurfs. (3) Among the small fraction that do become Smurfs most (40-60%) become blue during their last 24 hours of life. Thus, even with daily checks most of the Smurf flies will be dead when initially identified as Smurf making their utility for gene expression studies useless. As can be seen from Table 2.1, there was substantial replication, with each population undergoing 6 different dyes and a control. The original dye used in Rera *et al.* [16], SPS Alfachem, was replicated in 5 different populations, and each population was replicated in 6 different dye environments, allowing for substantial replication across the whole experiment. The three results cited above were consistently seen across all the replicates suggesting that these findings are robust.

These results certainly contradict prior claims [16]. Rera et al [16] suggested that essentially all flies become Smurfs prior to death and that the dyes do not affect survival. Certainly, one can claim there were differences in handling or techniques used in these studies [16]. This is challenging to evaluate. We note that the food used in our study has 2.5 grams of dye per 100 mL of food (2.5 % wt/vol), which is the same dye concentration that Rera et al [16] put in their food [17]. Thus, our observations of increased mortality due to dye cannot be attributed to overdosing. We also tested whether the tapping employed in our experimental technique could explain the mean longevity differences. That experiment showed no detectable effects of tapping on either male or female longevity. We only tested one population, ACO, for an effect of tapping on mean longevity. Thus, for the ACO population it is clear that dyes are

responsible for their reduced longevity not tapping. While it is theoretically possible that the other populations are not affected by the dyes but are affected by tapping, we believe this is an unlikely possibility.

There might be variation in how much food a fly consumes, but since each fly was in their own environment with only the dyed food, the flies had no other option but to consume the food or die from starvation. The purpose of this assay technique is to identify flies about to die under normal husbandry protocols. Future experiments can focus on whether the dead 'non-Smurfed' flies consumed food before death or not. But if the majority of the flies are dying without the distinguishing blue body color, then the technique is of little practical use. Lastly, most flies which did turn blue did so during their last 24 hours of life. This also renders the technique less useful for collecting live flies shortly before their death.

Many experiments have used the Smurf Assay technique [18 – 23]. However, the widespread use of the Smurf assay to differentiate between aging flies and young flies is not justified. At older ages, less flies Smurfed than flies that were younger. Prior research has demonstrated that the technique will differentiate between individuals that loose intestinal integrity and become Smurfs and those that don't, but they fail to provide exact details on how many individuals become Smurfs prior to death.

Our results also demonstrate significant effects of fly population of origin and dye on both mean longevity and frequency of Smurfs. However, these effects are essentially background noise to the major observations that only about 28% of flies ever become Smurfs and those that do only do so on their day of death or one day before death.

ACKNOWLEDGMENTS

A big thank you to the undergraduates in the lab who helped with the experiment and the UCI School of Biological Science and the University of California HBCU Fellowship program for financial support.

REFERENCES

1. Hamilton WD. The moulding of senescence by natural selection. J Theor Biol. 1966 Sep;12(1):12-45. doi: 10.1016/0022-5193(66)90184-6

2. Comfort, A. (1979) The Biology of Senescence. London (UK), Livingstone

3. Carey JR, Liedo P, Orozco D, Vaupel JW (1992) Slowing of mortality rates at older ages in large medfly cohorts. Science. 1992 Oct 16;258(5081):457-61. doi: 10.1126/science.1411540

4. Curtsinger JW, Fukui HH, Townsend DR, Vaupel JW. Demography of genotypes: failure of the limited life-span paradigm in Drosophila melanogaster. Science. 1992 Oct 16;258(5081):461-3. doi: 10.1126/science.1411541

5. Rose MR, Drapeau MD, Yazdi PG, Shah KH, Moise DB, Thakar RR, et al. Evolution of late-life mortality in Drosophila melanogaster. J Evol Biol. 2006 Jan;19(1):289-301. doi: 10.1111/j.1420-9101.2005.00966.x

6. Mueller LD, Rauser CL, Rose MR. An evolutionary heterogeneity model of late-life fecundity in Drosophila. Biogerontology. 2007 Apr;8(2):147-61.

7. Vaupel JW, Carey JR, Christenson K, Johnson TE, Yashin AI, Holm NV, et al. Biodemographic trajectories of longevity. Science. 1998 May 8;280(5365):855-60. doi: 10.1126/science.280.5365.855

8. Rauser CL, Abdel-Aal, Y, Shieh JA, Suen CW, Mueller LD, Rose MR. Lifelong heterogeneity in fecundity is insufficient to explain late-life fecundity plateaus in Drosophila melanogaster. Exp Gerontol. 2005 Aug-Sep;40(8-9):660-70. doi:10.1016/j.exger.2005.06.006

9. Shahrestani P, Tran X, Mueller LD. Physiological decline prior to death in Drosophila melanogaster. Biogerontology. 2012 Oct;13(5):537-45. doi: 10.1007/s10522-012-9398-z

10. Shahrestani P, Tran X, Mueller LD,. Patterns of male fitness conform to predictions of evolutionary models of late life. J Evol Biol. 2012 Jun;25(6):1060-5. doi: 10.1111/j.1420-9101.2012.02492.x

11. Gaitanidis A., Dimitriadou A., Dowse H., Sanyal S. Duch C., Consoulas C. (2019). Longitudinal assessment of health-span and pre-death morbidity in wild type Drosophila. Aging, 11:1850-1873).

12. Mueller LD, Shahrestani P, Rauser CL, Rose MR. The death spiral: predicting death in Drosophila cohorts. Biogerontology. 2016 Nov;17(5-6):805-816. Epub 2016 Feb 25

13. Muller HG, Carey JR, Wu D, Liedo P, Vaupel JW. Reproductive potential predicts longevity of female Mediterranean fruit flies. Proc Biol Sci. 2001 Mar 7; 268(1466): 445–450, doi: 10.1098/rspb.2000.1370

14. Rogina B, Wolverton T, Bross T, Chen K, Muller H-G, Carey JR. Distinct biological epochs in the reproductive life of female Drosophila melanogaster. Mech Ageing Dev 128, 477–485, doi:10.1016/j.mad.2007.06.004 (2007).

15. Papadopoulos NT, Carey JR, Katsoyannos BI, Kouloussis NA, Muller HG, Liu X. Supine behaviour predicts the time to death in male Mediterranean fruitflies (Ceratitis capitata). Proc Biol Sci. 2002 Aug 22;269(1501):1633-7

16. Rera M, Clarck RI, Walker DW. Intestinal barrier dysfunction links metabolic and inflammatory markers of aging to death in Drosophila. Proc Natl Acad Sci U S A. 2012 Dec 26;109(52):21528-33. doi: 10.1073/pnas.1215849110.

17. Martins RR, McCracken AW, Simons MJP, Henriques CM, Rera M. How to catch a Smurf? in vivo assessment of intestinal permeability in multiple model organisms. Bio Protoc. 2018; 8:e2722. doi: 10.21769/BioProtoc.2722

18. Darby TM, Owens JA, Saeedi BJ, Robinson BS, Naudin CR, Jones RM. Lactococcus Lactis Subsp. cremoris Is an Efficacious Beneficial Bacterium that Limits Tissue Injury in the Intestine. iScience 2019; 12: 356– 367. doi: 10.1016/j.isci.2019.01.030.

19. Kissoyan KAB, Drechsler M, Stange EL, Zimmerman J, Kaleta C, Bode HB, Dierking K. Natural *C. elegans* Microbiota Protects against Infection via Production of a Cyclic Lipopeptide of the Viscosin Group. Current Biology 2019; 29: 1030–1037

20. Klichko VI, Safonov VL, Safonov MY, Radyuk SN. Supplementation with hydrogen-producing composition confers beneficial effects on physiology and life span in Drosophila. Heliyon 5 (2019) E01679.

21. Liang ST, Audira G, Juniardi S, Chen JR, Lai YH, Du ZC, et al. Zebrafish Carrying pycr1 Gene Deficiency Display Aging and Multiple Behavioral Abnormalities. Cells 8: 453 (2019) doi: 10.3390/cells8050453

22. Rodriguez-Fernandez IA, Qi Y, Jasper H. Loss of a proteostatic checkpoint in intestinal stem cells contributes to age-related epithelial dysfunction. Nat Commun 10, 1050 (2019) doi:10.1038/s41467-019-08982-9

23. Dambroise E, Monnier L, Ruisheng L, Aguilaniu H, Joly J.-S, Tricoire H, et al. Two phases of aging separated by the Smurf transition as a public path to death. Sci Rep6, 23523 (2016) doi:10.1038/srep23523

FIGURES

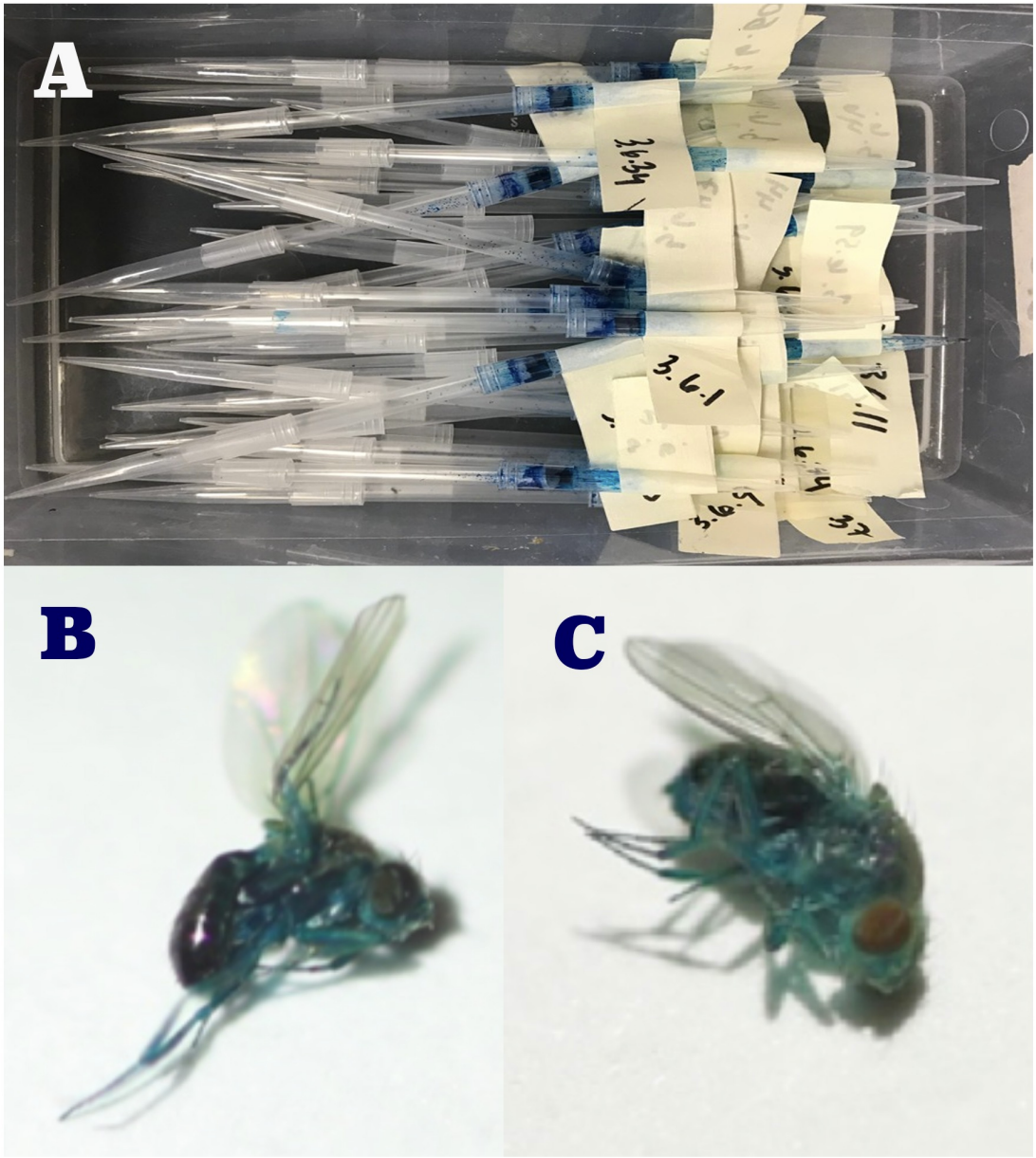

Figure 2.1 Setup of Smurf experiment and 'Smurfed' flies. A**.** Adult, 14 day old (from egg) flies were placed into individual plastic straws about 4 inches in length and capped with plastic pipette tips on both ends. Each straw was labeled with a number that allowed us to keep track of each fly, **B.** Blue male Smurf at the time of death, C. Blue female Smurf at time of death. The head, thorax and abdomen have all visibly become blue for both the male and female *D. melanogaster* flies.

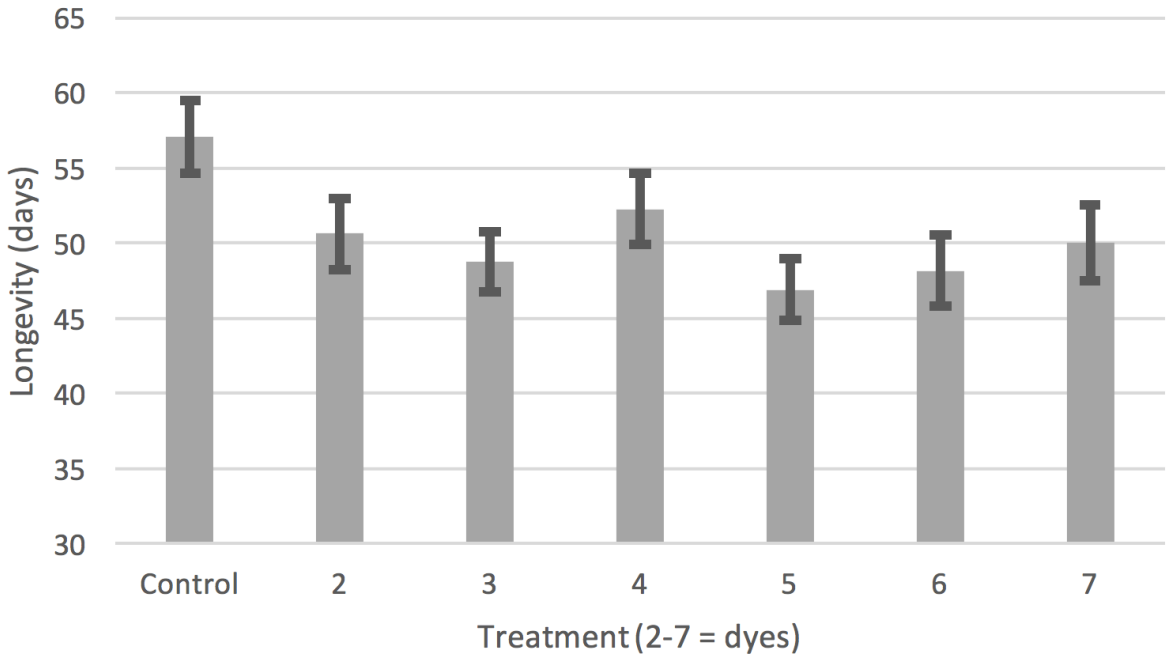

Figure 2.2 The mean longevity of five populations of *D. melanogaster* in the control and six different dye treatments. Bars are 95% confidence intervals calculated from a pooled variance estimate by the *emmeans* R function. Each dye treatment resulted in a significant reduction in mean longevity compared to the control treatment. Treatment 2: SPS Alfachem Blue, 3: Sigma Aldrich, 4: Spectrum Blue, 5: Flavors and Color Blue, 6: Chemistry Connection Blue, 7: Electric Blue

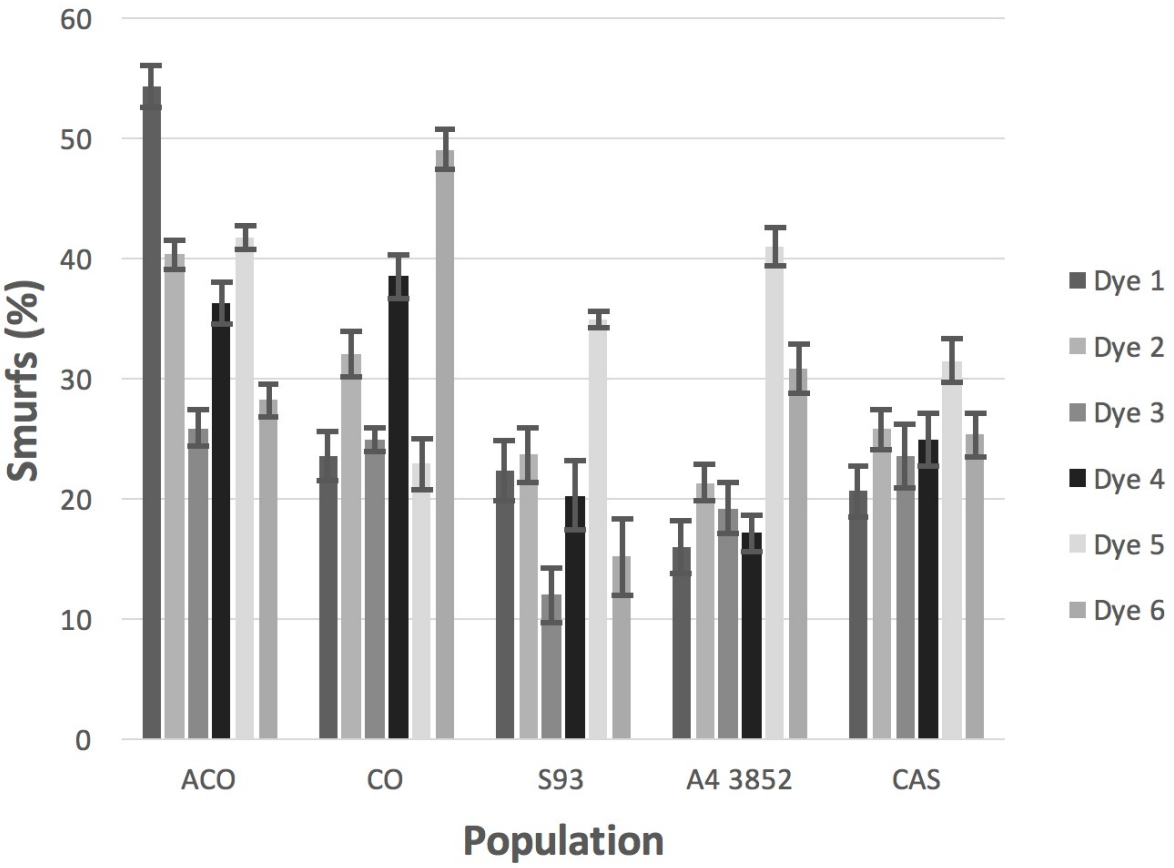

Figure 2.3 Percentage of flies that became Smurfs for each population and dye. Bars are standard errors.

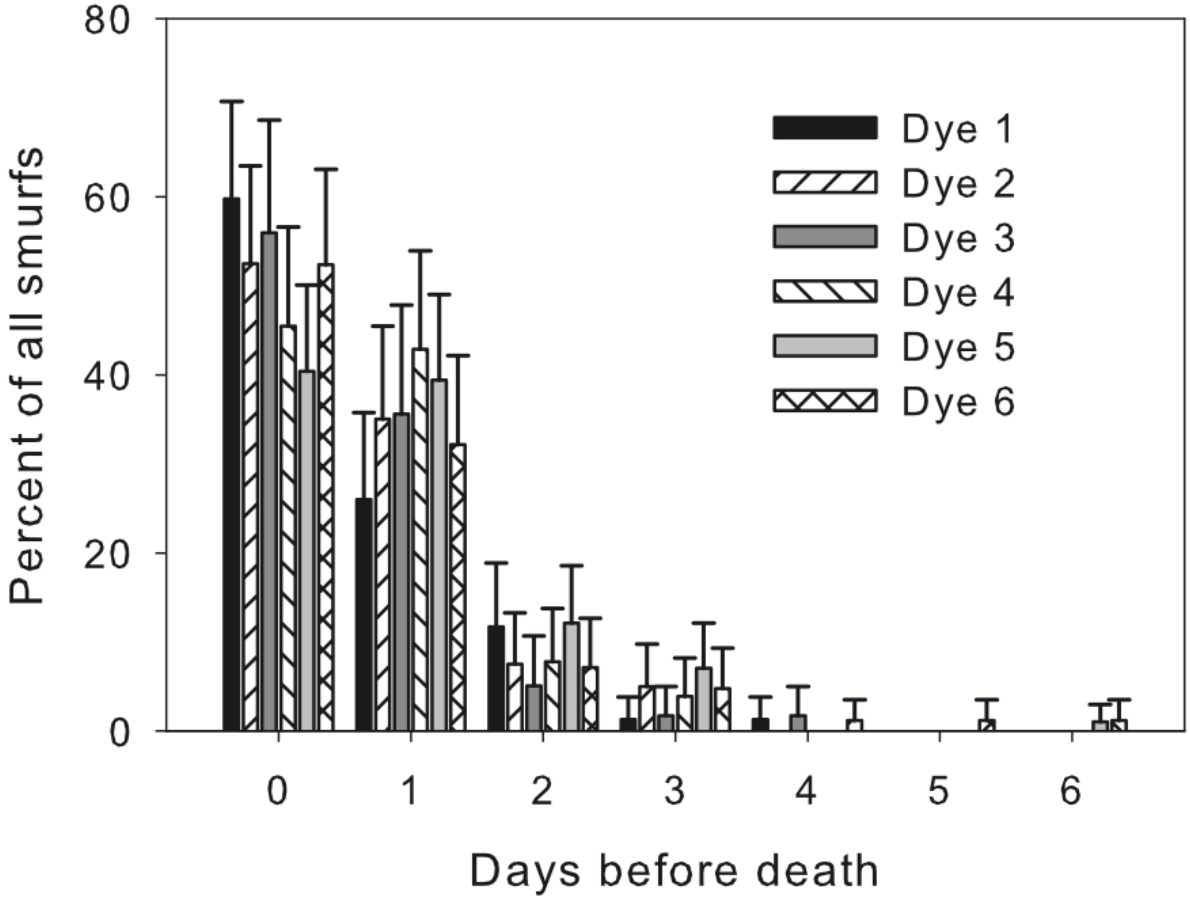

Fig. 2.4 Timing of SMURF appearance. We used only flies that satisfied our criteria for being a Smurf. Most flies appeared to be Smurfs on the day they are found dead or 1 day before death. Of the *D. melanogaster* that became a Smurf, the majority of them did so on the day of their death or one day before. Bars are simultaneous 95% confidence intervals. Dye 1: SPS Alfachem Blue, Dye 2: Sigma Aldrich, Dye 3: Spectrum Blue, Dye 4: Flavors and Color Blue, Dye 5: Chemistry Connection Blue, and Dye 6: Electric Blue

TABLES

Table 2.1 The number of flies used in the experiment per population of *D. melanogaster* as well as the total number and sex (M = Male, F = Female) of flies exposed to each dye. The left column has the five populations used: ACO, CO, S93, A4 3852 and Canton S, and the top row has the medium that the flies were fed: either the control banana molasses food or the banana molasses food with the indicated dye. This table excludes one CO individual in dye 1 whose sex was unknown.

Population	**Control**		**Dye 1**		**Dye 2**		**Dye 3**		**Dye 4**		**Dye 5**		**Dye 6**		**Total Flies per Population**
	Regular Banana Molasses Food		**Food & SPS Alfachem Blue**		**Food & Sigma Aldrich**		**Food & Spectrum Blue**		**Food & Flavors & Color Blue**		**Food & Chemistry Connection**		**Food & Electric Blue**		
1: ACO	Total Flies		Total Flies		Total Flies		Total Flies		Total Flies		Total Flies		Total Flies		382
	57		56		52		54		55		55		53		
	M	F	M	F	M	F	M	F	M	F	M	F	M	F	
	30	27	28	28	27	25	27	27	29	26	27	28	28	25	
2: CO	Total Flies		Total Flies		Total Flies		Total Flies		Total Flies		Total Flies		Total Flies		395
	54		54		56		56		57		61		57		
	M	F	M	F	M	F	M	F	M	F	M	F	M	F	
	27	27	25	29	29	27	27	29	29	28	29	32	28	29	
3: S93	Total Flies		Total Flies		Total Flies		Total Flies		Total Flies		Total Flies		Total Flies		410
	57		58		59		58		59		60		59		
	M	F	M	F	M	F	M	F	M	F	M	F	M	F	
	27	30	29	29	30	29	27	31	30	29	30	30	31	28	
4: A4 3852	Total Flies		Total Flies		Total Flies		Total Flies		Total Flies		Total Flies		Total Flies		396
	57		57		56		57		58		56		55		
	M	F	M	F	M	F	M	F	M	F	M	F	M	F	
	28	29	30	27	27	29	29	28	29	29	27	29	27	28	
5: Canton S (CAS)	Total Flies		Total Flies		Total Flies		Total Flies		Total Flies		Total Flies		Total Flies		398
	55		58		58		55		56		57		59		
	M	F	M	F	M	F	M	F	M	F	M	F	M	F	
	27	28	29	29	29	29	24	31	28	28	28	29	29	30	
Total Flies per dye	**280**		**283**		**281**		**280**		**285**		**289**		**283**		**1982**

Table 2.2**.** Mean longevity from the tapping experiment. The *p*-value is for a *t*-test for different mean longevities of same sex treatments. There is no discernible difference in the mean longevity of the flies that were tapped versus those that were not tapped.

Sex	**Tapping**	**Mean Longevity (Days)**	**95% Confidence Interval**	***p* value**
Male	Yes	53.2	(47.8, 58.5)	0.83
Male	No	53.9	(49.4, 58.5)	
Female	Yes	54.5	(49.0, 60)	0.88
Female	No	54.4	(49.0, 59.8)	

SUPPLEMENTAL FIGURES

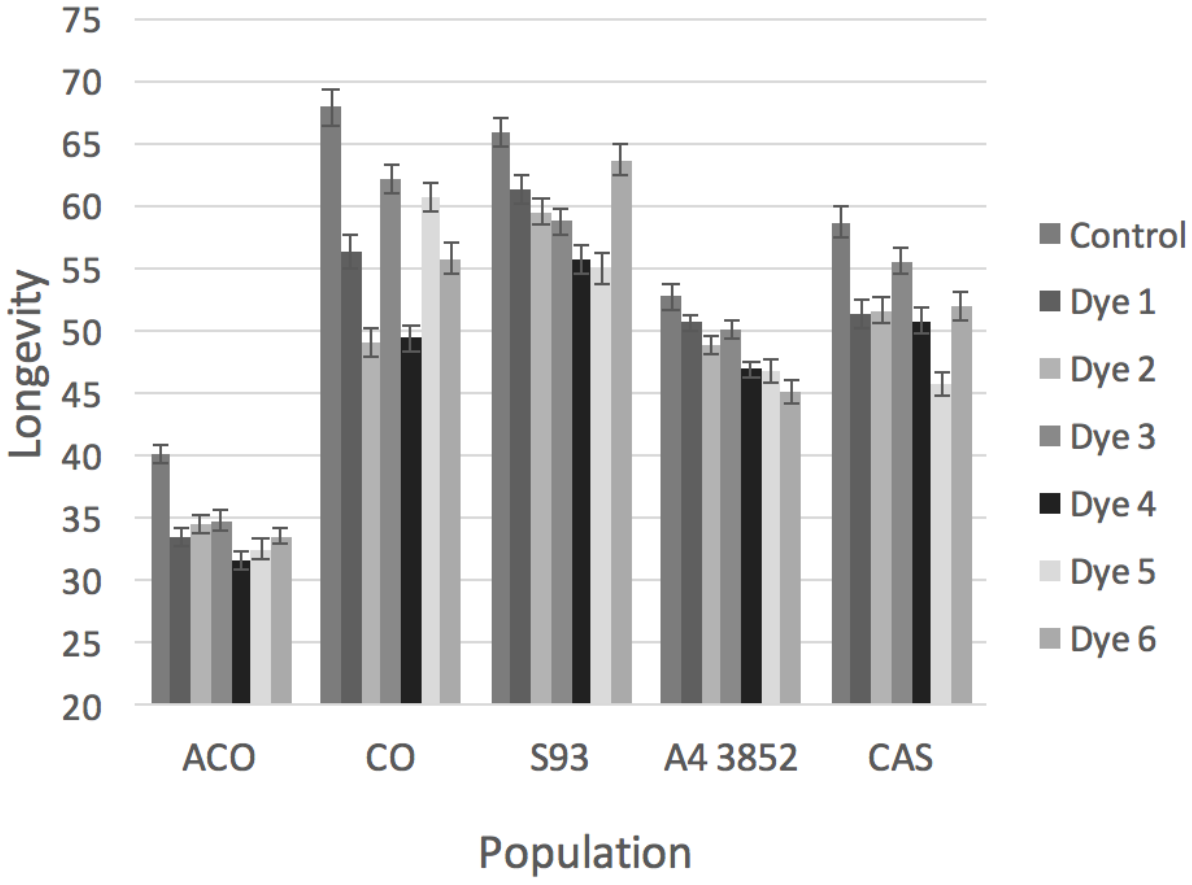

Figure S2.1. Mean longevity in days of flies in the control environment and dyes across all 5 populations. Standard error bars. When pooling the dyes against the control, the control flies lived significantly longer than the flies in an environment with dye in the food. The dyes used: Dye 1: SPS Alfachem Blue, Dye 2: Sigma Aldrich, Dye 3: Spectrum Blue, Dye 4: Flavors and Color Blue, Dye 5: Chemistry Connection Blue, and Dye 6: Electric Blue

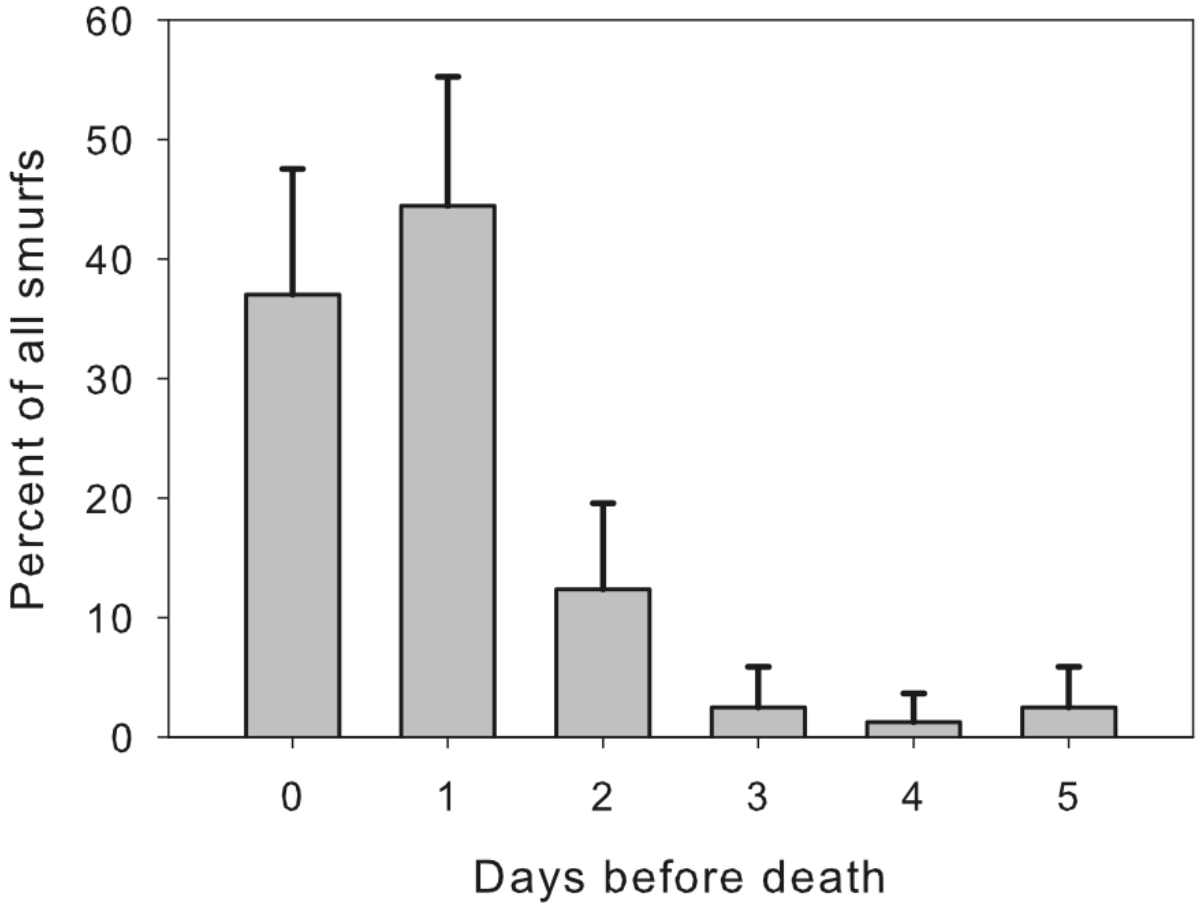

Figure S2.2. Results of the preliminary study of 172 ACO1 adults. The percent of first appearance of all 81 Smurfs as a function of the days before death when raised on food with dye 1. The bars are simultaneous 95% confidence intervals. The majority became Smurfs on the day they were found dead (day 0) or 1 day before death. A total of 47% (95% confidence interval, (39%, 54%)) eventually became Smurfs. The mean longevity (from egg) of all flies in this experiment was 32.9 days (95% confidence interval ±1.4 days). These results are consistent with those in the full experiment. Specifically, less than 50% of all flies became Smurfs prior to death and those that did become Smurfs most frequently did so on the day or day before they died.

SUPPLEMENTAL TABLES

Table S2.1. Average longevity for every population and every dye used in the experiment. Also provided are the average day of Smurfing.

Population		Dye 1	Dye 2	Dye 3	Dye 4	Dye 5	Dye 6	Control
ACO	Average Day of Death	33.518	34.558	34.815	31.655	32.545	33.566	40.053
	Average day of Smurf	30.129	31.143	26.571	26.8	28.913	31.2	N/A
CO	Average Day of Death	56.255	49.107	62.143	49.439	60.689	55.719	67.87
	Average day of Smurf	43.538	41.944	61.857	48.545	63.143	53.929	N/A
S93	Average Day of Death	61.293	59.508	58.724	55.712	54.95	63.678	65.86
	Average day of Smurf	59.154	56.857	61.571	46.083	48.524	46.889	N/A
A4 3852	Average Day of Death	50.632	48.821	50.07	46.948	46.821	45.091	52.719
	Average day of Smurf	47.444	49.5	50	46.5	44.739	40	N/A
CAS	Average Day of Death	51.362	51.655	53.582	50.786	45.754	52.051	58.691
	Average day of Smurf	50.333	54.933	45.308	46.5	41.444	41.067	N/A

Table S2. Log-rank *p*-values comparing the control treatment to each of the other populations at each of the dye treatments. These results are from the R *survdiff* function in the *survival* package. The significant results are shown in bold (using the Bonferroni correction for 30 tests).

Population	Dye 1	Dye 2	Dye 3	Dye 4	Dye 5	Dye 6
ACO	9.1×10^{-3}	9.4×10^{-3}	4.9×10^{-2}	$\mathbf{2.5\times10^{-4}}$	2.0×10^{-3}	$\mathbf{3.7\times10^{-4}}$
CO	8.8×10^{-3}	$\mathbf{5.9\times10^{-8}}$	4.6×10^{-2}	$\mathbf{4.9\times10^{-8}}$	1.1×10^{-2}	$\mathbf{3.89\times10^{-4}}$
S93	0.17	3.2×10^{-2}	2.7×10^{-2}	1.6×10^{-2}	1.5×10^{-2}	0.58
A4 3852	1.2×10^{-2}	9.6×10^{-3}	0.051	$\mathbf{2.9\times10^{-4}}$	4.1×10^{-3}	3.4×10^{-3}
CAS	3.02×10^{-2}	$\mathbf{9.9\times10^{-4}}$	1.01×10^{-2}	$\mathbf{4.5\times10^{-4}}$	$\mathbf{3.4\times10^{-6}}$	3.2×10^{-2}

Table S2.3. The median longevity for each population at every dye treatment. The median longevity for each population was greater in the control treatment in than every other dye except for A4 3852 and dye 1.

Population	Dye 1	Dye 2	Dye 3	Dye 4	Dye 5	Dye 6	Control
A4 3852	54	49	51	49	50	47	53
ACO	33	33.5	36	31	33	33	37
CAS	51	54	57	54.5	45	51	63
CO	56	54	65	49	59	55	68
S93	63.5	63	62.5	56	54	68	68

Table S2.4. The maximum longevity for each population, dye combination.

Population	Dye 1	Dye 2	Dye 3	Dye 4	Dye 5	Dye 6	Control
ACO	62	61	64	63	71	55	74
CO	104	86	105	88	123	95	100
S93	100	100	89	85	100	96	105
A4 3852	70	71	72	65	67	73	86
CAS	89	84	89	82	79	86	89

Table S2.5. The number of flies that Smurfed per population and per dye, as well as the number of total flies per population and dye.

Population		**Dye 1**	**Dye 2**	**Dye 3**	**Dye 4**	**Dye 5**	**Dye 6**
ACO	**TOTAL Smurf**	30	21	14	20	23	15
	Total Flies	56	52	54	54	55	53
CO	**TOTAL Smurf**	13	18	14	22	14	28
	Total Flies	55	56	56	57	61	57
S93	**TOTAL Smurf**	13	14	7	12	21	9
	Total Flies	58	59	58	59	60	59
A4 3852	**TOTAL Smurf**	9	12	11	10	23	17
	Total Flies	56	56	57	58	56	55
CAS	**TOTAL Smurf**	12	15	13	14	18	15
	Total Flies	58	58	55	56	57	59

STANDARD ROSE AND MUELLER LAB BANANA FOOD RECIPE

The laboratory is a controlled environment with respect to temperature and so on, and we must take care that the food given the flies is equally uniform in quality. Most of the adult phenotype is determined by the larval stages, and the larvae feed, bathe, and live in the banana food. It is therefore absolutely critical that the food be consistent, particularly when the flies are to be assayed in an experiment. Limited variation in banana ripeness & quality, slight burning (&c) can be tolerated for routine stock maintenance, but should be avoided as much as possible for experimental-generation flies. The ideal banana is ripe but not rotting, with full yellow color and the first hints of browning spots.

INGREDIENTS

Cook Size	**4.4 Liters**	**6.6 Liters**
STEP 1		
Distilled water	4.4 Liters	6.6 Liters
Agar	66.7 grams	100 grams
STEP 2		
Bananas unpeeled	600 grams	900 grams
Distilled water	267 ml	400 ml
Light Karo syrup	1 & 1/3 scoop *	2 scoops *
Dark Karo syrup	1 & 1/3 scoop *	2 scoops *
Barley Malt	2 scoops *	2 scoops *
STEP 3		
Distilled water	307 ml	460 mL
95% Ethanol	107 ml	160 mL
Yeast	160 grams	240 grams
Step 4		
95% Ethanol	104 ml	156 mL
p-hydrobenzoic acid	10.4 grams	15.6 grams

Table S2.6 Ingredients used in the Rose and Mueller laboratory food recipe. * A "scoop" refers to the ice-cream scooper used in the Rose Lab, which is 55 mL, volumetrically

DIRECTIONS

STEP 1

Start by measuring the water & agar and combining them in the cook pot with a stir-bar (a large one) and put the hot-plate on "high". Don't turn the stirring speed too high or the stir-bar will come off-center. Allow the pot to boil; this may take better than an hour, so you should start the cook while still attending to other things. The Bunsen burner may be used at this stage to speed up the boiling, but even at this stage over-boiling and burning are possible. Do not leave the lab with the flame on.

STEP 2

When the agar-water has boiled vigorously for at least 5 minutes you may proceed with step two; do not start this step too early, or else the mixture will congeal into nasty lumps and the bananas will oxidize. For the small (4.4L) cook one blender is sufficient, but two are necessary for the 6.6L cook. Add the water to the blender(s), which should be well cleaned, and then the weighed bananas one by one while blending on low. Make sure the bananas are smoothly blended in by switching to a high setting. Then add the other ingredients (karo & barley malt). Mix everything well and then add to the boiling agar-water. Do not wash the blenders. Using a wooden spoon to scrub the areas of the pot outside of the stir-bar's reach (don't knock stir-bar off center) and rotate the whole pot slightly help prevent burned spots which slow down re-boiling.

STEP 3

While it was critical in step one to allow the mixture to boil, the third step may be started before the food re-boils. Add most of the water to one of the blenders (at least 3/4), saving a bit in the

graduated cylinder, and rinse it out before pouring it into the other blender (it only takes one) to get most of the residual banana mix out. Add ethanol, and, gradually while blending, the yeast. Blend until all particles of yeast are dissolved (5-10 minutes); too short and the yeast will precipitate out, too long and it will start growing in the blender. Use the bit of water you saved to rinse out the blender and add it to the cook. Allow the final mixture to boil for at least five minutes. As the food boils at the final stage it will foam up and overflow if the lid is on; you must watch the food as it nears a boil. Generally, it is best to take the lid off when the food comes to a boil and let it boil for at least five minutes. It is absolutely critical that the food boil after the yeast has been added.

STEP 4 anti-fungal solution

Remove the food and allow it to cool -- a cold water bath speeds this up. Mix up the ethanol and hydrobenzoic acid and cover it to prevent evaporation. The food should be stirred occasionally while cooling. When the temperature hits 48oC the anti-bacterial solution should be added, and mixed in thoroughly before pouring.

STEP 5 - pouring

Vials: For most purposes, we require vials for egg-collection/larval rearing. The standard depth in an 8-dram vial is about 2cm (or 6mL), which is well in excess of the amount eaten by the larvae; within reason, any level of food above about 1cm is adequate. A 6.6L cook will make about 41/2 racks of each. food if carefully poured; a 4.4L cook will make about 3 racks. Vials which are only for temporary feeding of adults (transfer vials) can be poured low.

Plates: Plates are usually poured up to the upper lip of a 15x100mm petri dish. A cook will produce about 12 - 14 plates / liter. Plates are almost always needed in the lab, so if you have

extra food pour plates.

STEP 6 - cooling & storage

Putting the food in front of the fan will speed its cooling and help prevent stray flies from getting into it. Egg-collection vials should be covered with the screening and plates should not be left out for flies to lay on. When vials are room temperature (1/2 - 1 hour), they should be inverted with paper towel on top into a new & clearly labelled & dated rack. Plates should be closed, bagged, labelled & dated. All food to be kept for more than two days should be refrigerated. Don't expect a fan or a screen to keep out loose flies: check.

CHAPTER 3

Correlated response to Urea Adaptation

INTRODUCTION

It is crucial in evolutionary biology to understand the role the natural environment has in molding adaptations and affecting allele frequency change, especially in insects (Partridge and Harvey 1988; Roff 1992; Stearns 1992). Insects are often exposed to many toxic substances in their environments, which include insecticides, heavy metals, chemicals or feeding deterrents synthesized by plants. Studies of resistance to toxic pesticides usually focus on compounds that target a specific enzyme or biological process. Urea and ammonia differ from these because of their wide-ranging effects on organismal and cellular physiology (Somero and Yancey, 1997). Many life-history characters are interrelated and rooted in the common network of physiological processes constituting metabolism and determining reproductive success or fitness (Lande 1982, Lande and Arnold 1983, Service, 1987, 1989; Rose, 1991; Graves *et al.,* 1992; Charlesworth 1993, Chippindale *et al*., 1993, Borash *et al*., 2000, Flatt 2020).

Fitness components in flies are frequently correlated with traits that confer the ability to resist and survive various stresses (Prasad and Joshi 2003, Flatt 2020). Larval competitive ability in *D. melanogaster* depends on many factors, such as feeding rate, initial weight, relative time spent on molting, minimum food requirement for pupation, and resistance to crowding. Experimental studies indicate that, in laboratory populations of *D. melanogaster,* the outcome of larval competition for food depends primarily on the feeding rate (Bakker, 1961). They found that an important component of the feeding rate is the number of bites the larva takes from the food medium per unit time, which can be measured as the number of sclerite retractions per minute. It was also demonstrated that there is a fair amount of additive genetic variation for retraction rates in populations of *D. melanogaster (*Sewell *et al*. 1975). The implication of their

finding meant that sclerite-retraction rate could be a character that responds to natural selection. Joshi *et a*l. (1988) found significant genetic variation for feeding rates in *D. melanogaster*.

Past research has shown that feeding rates in *Drosophila* larvae are affected by environmental factors. Experiments with crowded conditions (hundreds of larvae vs. 50-80 larvae per vial) have shown to lead to a major food shortage, intense competition, and increasing levels of toxic metabolic waste (ammonia) produced by the larvae (Shiotsugu *et al*. 1997; Borash *et al*. 1998). Under such crowded cultures, early emerging larvae have high larval feeding rates and lower absolute viability in food laced with ammonia while the late emerging types have reduced feeding rates but higher absolute survival (Joshi *et al* 1988, Borash *et al*. 1998).

Previous experiments in our research laboratory have examined adaptations in populations of *D. melanogaster* to high levels of environmental urea and ammonia. Although Botella *et al*. (1985) suggested urea might be the primary nitrogen waste product of *Drosophila* larvae, more recent work by Borash *et al*. (1998) showed that only ammonia accumulated in the food of crowded *Drosophila* cultures. Urea appears to have general cytotoxic effects due to interference with translation, action as protein denaturant (Somero and Yancey, 1997), and its reduction of enzyme activity and thermostability (Bowlus and Somero, 1979; Yancey and Somero, 1979; Yancey, 1985; Yancey, 1992; Somero and Yancey, 1997). Other experiments have found that increased levels of isoaspartyl residues, a type of protein, are found in larvae reared on urea-containing media (David *et al*. 1999). However, urea is not a common nitrogen waste produce in insects. Borash *et al* 1998 found that urea levels were very low in regular environments, and even in crowded environments the levels were not high enough to be toxic. Borash *et al*, 2000 has shown that feeding rates decline as populations adapt to urea and ammonia, while feeding rates decline also in the presence of parasites (Fellowes *et al*. 1999).

Previous work has shown that *Drosophila* populations can evolve resistance to high levels of environmental urea (Joshi *et al*. 1996, Shiotsugu *et al.*, 1997, David *et al.* 1999, Etienne *et al.* 2001).

An important model that looked at the evolution of feeding rates in *D. melanogaster* larvae in stressful environments found they were altered as they adapted to high levels of ammonia, urea, and the presence of parasitoids (Mueller *et al* 2015). If there are toxic compounds in the larval food, energy is required for detoxification, and larvae can maximize food intake by slowing their feeding rate (Mueller *et al* 2015). While it may be argued that a slower feeding rate does not necessarily prove that the larva will grow slower, previous studies (Joshi and Mueller 1996; Mueller *et al*. 1991) showed that fast feeding rate larvae required more food to reach the same critical minimum size. Mueller *et al.* 1990 showed a negative correlation between feeding efficiency and feeding rate.

Our goal was not only to study feeding rates of our urea-adapted lines, but to also look at various other phenotypes, such as viability, development time, starvation and desiccation resistance, larval growth rate and adult size. Numerous studies have measured phenotypic variation in fitness components by assaying these traits under standard laboratory conditions. These various studies (referenced in Flatt 2020) revealed that there generally exist large amounts of phenotypic variation for many components of fitness, including traits such as development time, larval competitive ability, viability, size at eclosion, age-specific fecundity, lifetime reproductive success, age-specific mortality, life span, and various stress resistance traits.
We developed 3 replicated lines of *D. melanogaster* that are currently, or have at some point in the past, been exposed to many generations of urea exposure in the larval food. Previous research on urea adapted lines have shown changes in larval feeding rates, larval foraging paths, and

viability in urea laced environments (Shiotsugu *et al.*, 1997; Borash *et al.*, 2000; Mueller *et al.*, 2005). Some of the same phenotypes are examined in this study. In addition, we have undertaken the measurements of larval growth rates. This will allow us to test whether larval feeding rates affect growth rates as suggested by Mueller and Barter (2015). The relationship between larval feeding rates and food consumption rates has been questioned (Kaun *et al.*, 2007). The experiments described next will present decisive evidence on this issue.

MATERIALS AND METHODS

1.1 Populations

All the populations in the lab are derived from an established Rose IV population (Rose, 1984), Figure 3.1. They are maintained on a banana-molasses food (Rose, 1984) at 25C (24 h light), uncontrolled humidity, and having a generation time of approximately 3 - 4 weeks (Table 3.1). All selection regimes are five-fold replicated, uncrowded as larva (60–80 eggs per 8-dram vial), with emergent adults kept at a low density of approximately 50–60 flies per 8-dram vial, and transferred to a cage environment with fresh food given about every other day for approximately 1 week. Effective populations sizes for each line were in the range of 700-1000 (Mueller *et al.*, 2013) every generation and maintained at large population sizes (1000), with discrete generations. The CO and nCO are control groups of populations that are on a 28-day life cycle. The TSO selected populations are a starvation selected population of five replicates, while the TDO's are a starvation and desiccation selected treatment population group of five replicates, both on 28-day life cycles. Both the TSO and TDO have now been in a control environment.

The UX and the UTB populations were subjected to selection for increased larval tolerance to the presence of toxic levels of urea in the food. The levels of urea were increased every few generations, when it was observed that a great proportion of larvae were surviving to adulthood. The derivation of the urea-tolerant (UX) and unselected controls (AUC) selection regimes was created in the Fall of 1996. Both populations were derived from a five-fold replicated set of populations called UU, which had a 3-week generation time, and were reared at low larval and adult densities. The UU group of populations were derived in 1990, from the Rose B populations (Rose, 1984; Chippindale *et al.*, 1994, 1996). The RUX are a reverse selected line

of the UX. They have been in a control environment since January 2013 The UTB line were created in October 2013. The AUC, RUX, UX, and UTB lines are all 21-day cycle flies.

For every assay, eggs were collected from the 8 groups of population types and passed through two generations of common, standard conditions – low larval density, 1000 adult density, discrete generation times and regular banana-molasses food (Bitner *et al.* 2020).

1.2 Feeding Rates

Feeding rates were collected on a total of eight groups of populations (Table 3.2). Eggs were collected from adults who had undergone a two-generation standardization procedure – low density, regular banana-molasses food. To measure the feeding rate, individual larvae around 48 hours old were gently moved onto a 3% agar coated with a 10% live yeast suspension. The larvae were given 60 seconds to adjust to the new surroundings, and their sclerite retractions were recorded for 60 seconds and counted for twenty larvae per population. The procedure for measuring sclerite rates is similar to Sewell *et al.* (1975) and described in Joshi *et al.* (1988).

The feeding rates were done in two blocks. The TSO populations were used in each block and therefore the feeding rates were presented as a difference from the TSO mean feeding rate. To control for block differences all feeding rates for individuals were converted to feeding rate differences by subtracting the mean feeding rate of all TSO populations which were tested in both block 1 and block 2. Then we can let the feeding rate difference for selection regime-i (i= 1,..,7), population-j (k= 1,..,35) and individual-k (k= 1,..,25 or 50) be y_{ijk}. The linear mixed effects model for y_{ijk} is,

$$y_{ijk} = \mu + \alpha_i \delta_i + B_j + \varepsilon_{ijk}$$

where $\delta_i = 0$, if i=1, and 1 otherwise, B_j and ε_{ijk} are independent normally distributed random variables with zero mean and variances σ_B^2 and σ_ε^2. The model was analyzed with the linear

mixed effects *R* function *lme* (R Core Team, 2017). Pairwise differences were evaluated after adjusting for multiple comparisons using Tukey's method implemented by the *R* function *lsmeans*.

The UX, RUX and UTB groups of populations all have a current or past history of being raised as larvae in food with urea. Since past research (Borash *et al*, 2000) has shown that feeding rates decline as populations adapt to urea, we were interested in testing the feeding rates of these three selection regimes as a group. Thus, we pooled all 15 UX, RUX, and UTB selection treatments into one urea population and repeated the analysis as described above comparing the urea adapted groups of populations to the AUC control.

1.3 Starvation and Desiccation

Individual female flies were collected from the eight selected groups, a total of 30 flies per replicate per assay (150 flies per population for one assay). Each fly was placed into a straw with pipette tips on both ends. The straws were wide enough for the flies to be able to move from one end to the other. The flies selected for the starvation assay were placed into straws with 3% agar while the flies selected for the desiccation assay were placed into straws with desiccant separated by cheese cloth. Desiccant can clog their airways and the cheese cloth functioned to dry out the environment, but not kill the flies directly. Flies subjected to desiccation were checked hourly, and flies undergoing starvation were checked every 4 hours.

Time to death by starvation and desiccation was collected on a total of eight populations groups. We can let the desiccation (or starvation) time for selection regime-i (i= 1,..,8), population-j (k= 1,..,40) and individual-k (k= 1,..,30) be y_{ijk}. The linear mixed effects model for y_{ijk} is,

$$y_{ijk} = \mu + \alpha_i \delta_i + B_j + \varepsilon_{ijk}$$

where $\delta_i = 0$, if i=1, and 1 otherwise, B_j and ε_{ijk} are independent normally distributed random variables with zero mean and variances σ_B^2 and σ_ε^2. The model was analyzed with the linear mixed effects *R* function *lme* (R Core Team, 2017). Pairwise differences were evaluated after adjusting for multiple comparisons using Tukey's method implemented by the *R* function *lsmeans*.

1.4 Viability vs. Food type

The viability experiment started with 50 first instar larvae. However, due to the size of this experiment and technical difficulty of counting out exactly 50 larvae the actual number of larvae put in each vial is more properly thought of as a random variable which could be both higher or lower than 50. Thus, in this analysis we have analyzed the total number of larvae that survived to become adults under the assumption that the mean number of input larvae was the same in all treatments.

Larvae were raised under two experimental treatments, a control environment with regular food and food with added urea. Ten vials were used per population per environment, for a total of 20 vials per population. Each vial held the approximate 50 larvae. Ultimately, we are interested in the testing the effects of urea on survival for each population as well as differences between the different selection regimes. These experiments were also run in three blocks each separated by about one years' time: (i) AUC, RUX and UX, (ii) TSO, CO, NCO, and UTB, (iii) UTB, UX, CO, and nCO.

The analysis of the differences among the urea selected lines (RUX, UX, and UTB) and their control (AUC) was done with blocks (i) and (iii). For this analysis, viability was scaled to the mean viability of the UX group of populations in the urea food environment. The analysis of demographically selected lines (TSO, CO, and nCO) was done with blocks (ii) and (iii). Viability

for the demographically selected populations was scaled to the mean viability UTB populations in urea.

Let y_{ijkm} be the number of survivors in food type-*i* (control (1), urea (2)), selection regime-*j* (AUC (1), RUX (2), and UTB (3) in the urea analysis and TSO (1), CO (2), and nCO (3) in the demographic analysis), population-*k* (*k*= 1,..,15) and replicate *m* (*m*=1,…,10). Let $\bar{y}_{urea}$ be the mean viability in the urea control populations (UX for blocks (i) and (iii) and UTB for blocks (ii) and (iii)). We analyzed the differences, $\Delta_{ijkm} = y_{ijkm} - \bar{y}_{urea}$. The effects of selection regime and food type were studied with the linear mixed effects model,

$$\Delta_{ijkm} = \alpha + \delta_i\beta_i + \delta_j\gamma_j + \delta_i\delta_j\pi_{ij} + b_k + c_{ijkm}$$

where δ_i=0 if *i*=1 and 1 otherwise, b_k and c_{ijkm} are the population and residual error terms assumed to have a mean of zero and different variances. Parameter estimates for these terms were made by the *lme* R function (R core team, 2017). All pairs of selection regimes were compared for significant differences in viability using the R *emmeans* program.

We also developed a statistical analysis of the relative level of adaptation to urea for each selection regime. Let y_{ijkm} be the number of survivors in food type-*i* (control, urea), selection regime-*j* (TSO, AUC, CO, NCO, RUX, UX, UTB), population-*k* (*k*= 1,..,35) and replicate *m* (*m*=1,…,10). Then we analyzed the differences, $\Delta_{jkm} = \bar{y}_{controljk} - y_{ureajkm}$ where $\bar{y}_{controljk}$ is the mean in selection regime-*j* and population-*k*.

1.5 Developmental Time vs Food type

This portion of the assay was focused on looking at differences in the development time of the larvae. The measurements were made from when first instar larvae were collected to when the fly eclosed in the control banana molasses environment compared to the urea environment. For each population, 10 vials were set up with 50 freshly hatched larvae each in a banana

molasses environment. Another 10 vials were set up with urea food and 50 freshly hatched larvae. Eclosing adults were collected every 6 hours separated by male and female and recorded.

The methods for the analysis of the development time experiment were the same as the viability analysis except there was one additional fixed effect, sex. The first analysis was on the difference between the urea development time and the control development time. Here the mean control development time for each selection/sex/population/rep sample was calculated and then subtracted from the corresponding development time in urea.

In this first section, we analyzed blocks (i) and (iii) using UX-urea as the standard. Thus, in block (i) we computed the mean UX-urea, and this mean was subtracted from the other control and urea observations. The same was done for block (iii) using the UX-urea mean from block (iii). In the second section, we analyzed blocks (ii) and (iii) using UTB as the standard. Thus, in block (ii) we computed the mean UTB-urea, and this mean was subtracted from the other control and urea observations. The same was done for block (iii) using the UTB-urea mean from block (iii).

1.6 Larval Growth Rate

Pairs of populations, matched by the replicate number, were assayed together in a randomized block design and the assay followed the same procedures as mentioned in Santos *et al.* (1997). With a two-generation lead in for each population, 45 newly hatched first instar larvae were collected with a fine paint brush and placed onto non-nutritive agar petri dishes with 3 ml of yeast paste (188 grams of yeast in 500 ml of DI water) and placed randomly into a 25 C incubator with 24-hour lighting. There were 13 different "hour numbers" larvae were sampled at: 24, 30, 36, 42, 48, 54, 60, 66, 72, 78, 84, 90, and 105 hours after the larvae were added to the petri dish. At the designated hour, the larvae were washed with DI water and then allowed to air

dry. The wet weight was taken of the larvae and they were then placed into an 80-degree C drying oven, after which their dry weights were recorded.

Empirically, larval growth follows a logistic trajectory. We used a three-parameter logistic model to model basic growth dynamics. Under this model the size of individual larvae after *t*-hours of growth is given by,

$$f(\boldsymbol{\varphi}, t) = \frac{\varphi_1}{1+exp\left[{}^{(\varphi_2 - t)}/_{\varphi_3}\right]}. \quad (2)$$

In model (2) the asymptotic maximum size is equal to φ_1, and the time to reaching half the maximum size is φ_2.With this model we let y_{ijkt} be the average size of a larva from selection regime-*i* (*i*=1 (AUC), 2 (CO), 3 (nCO), 4 (RUX), 5 (UTB), 6 (UX), 7 (C) and 8(D)), population-*j* (*j*=1,…,40), and block-*k* at time-*t*. Random variation arises due to both population effects (random genetic drift) and individual variation. Consequently, the size of larvae from selection regime-*i* and population-*j* at time-*t* is $y_{ijt} = f(\boldsymbol{\varphi}_{ij}, t) + \varepsilon_{ijkt}$, and,

$$\begin{aligned} \varphi_{i1} &= \alpha_1 + \delta_i \gamma_{1i} \\ \varphi_{ij2} &= \alpha_2 + \delta_i \gamma_{2i} + b_j + c_k \\ \varphi_{i3} &= \alpha_3 + \delta_i \gamma_{3i}, \end{aligned} \quad (3)$$

where δ_i =0, if *i*=1 and 1 otherwise. The within population variation, ε, is assumed to be normally distributed with a zero mean. This variation increases as the larvae get larger so we assumed that $\mathrm{Var}(\varepsilon) = \sigma^2 |t|^{2\Delta}$ where Δ is estimated from the data. Population variation, *b*, and block variation, *c*, is assumed to only affect parameter φ_2. We tested models with population variation in the other parameters and the model with variation in φ_2 was chosen due to having the lowest Akaike and Bayesian information criterion (Pinheiro and Bates, 2000, chapter 8). The population variation is assumed independent of the within population variation and has a normal distribution

with zero mean and variance, σ_b^2. Parameters of equation (3) were estimated by the restricted maximum likelihood techniques implemented by the *nlme* function in R (R Core Team, 2017).

When displaying the predicted larval size from equations (2-3) we also calculated 95% confidence intervals. With eight different selection regimes we have 24 maximum likelihood parameters estimates and their covariance matrix estimates, $\widehat{\boldsymbol{\mu}} = \left(\hat{\beta}_1, \hat{\beta}_2, .., \hat{\beta}_{24}\right)$ and $\widehat{\boldsymbol{\Sigma}}$. These were assumed to have a *t*-distribution. From these distributions we drew samples of the parameter vectors, $\tilde{\mu}_k$, (*k*= 1,…,*m*). For each sampled parameter vector, we made size predictions for each selection regime for ages, 42, 48, 54, and 60 hours. At a specific age let the *kth* (out of *m*) prediction for selection regime-*i* be $\tilde{y}_{ki}$. From these *m* predictions, we generated order statistics, $\Delta^s(\tilde{y}_{ki})$, where $\Delta^1(\tilde{y}_{ki})$ is the smallest predicted value at *t* and $\Delta^m(\tilde{y}_{ki})$ is the largest. From the order statistics we then used $\Delta^l(\tilde{y}_{ki})$ as the lower confidence limit and $\Delta^u(\tilde{y}_{ki})$ as the upper confidence limit. In our simulations we set m=5,000. Therefore, a 95% confidence interval corresponds to $\Delta^l(\tilde{y}_{ki}) = \Delta^{125}(\tilde{y}_{ki})$ and $\Delta^u(\tilde{y}_{ki}) = \Delta^{4876}(\tilde{y}_{ki})$.

One hypothesis of interest was whether there was a relationship between the larval feeding rates and the growth of larvae. To test this, we fitted a line to the larval size vs. feeding rate observations at 4 larval ages around 48 hours – the age our feeding rates were measured. A significant positive slope for these lines were taken as evidence consistent with our hypothesis.

1.7 Adult size

The adult size was collected from the same larvae collected for the larval growth rate assay (section 1.6). Following Santos *et al.* (1997), at hour 105, 30 pupa were collected and placed into non-nutritive agar vials to allow for their development. When they had eclosed, flynap was used to anesthetize the flies to record their wet weights. Afterwards, they were placed into a drying oven at 80 degrees Celsius for 48 hours. The dry weight of the adult flies was then

recorded after the 48 hours had passed. The statistics for analyzing the adult dry size of the *D. melanogaster* flies are the same as the development time differences.

RESULTS

Feeding rates

Toxic environments can lead to various trade-offs in *D. melanogaster.* Larval feeding rates in *D. melanogaster* are highly correlated with competitive ability (Sewell *et al.*, 1975; Burnet *et al.*, 1977; Joshi *et al.*, 1988). A significant slower feeding rate was observed in the urea-adapted lines – UX, UTB and RUX. Figure 3.2 shows that the RUX, UX and UTB groups of populations are feeding at similar rates.

The UX, RUX and UTB populations all have a current or past history of being raised as larvae in food with urea. Since past research (Borash *et al*, 2000) has shown that feeding rates decline as populations adapt to urea, we were interested in testing the feeding rates of these three selection regimes as a group. Thus, we pooled all 15 UX, RUX, and UTB populations into one urea population and repeated the analysis. In the Urea vs AUC contrast the *p*-value for just that one test was 0.028. So, the urea lines feed significantly more slowly than the AUC control (Fig. 3.3).

Starvation and Desiccation Resistance

In looking at starvation and desiccation resistance, two of the selection treatments, TSO and TDO, had undergone selection in their evolutionary history for resistance to starvation (TSO) and starvation and desiccation (TDO) but were now reverse selected culturing in the control environment. Regarding desiccation resistance, there was no significant difference between any of the 8 groups of populations (Figure 3.4). Starvation resistance showed no difference between the selection treatments under the horizontal bars, but significant differences were seen in some population comparisons (Figure 3.5). UTB had a significant difference between TSO (p=0.0002), TDO (p=0.0026), nCO (p=0.0029) and CO (p=0.039). UX had a

significant difference between TSO (p=0.0003), TDO (p=0.0054) and NCO (p=). RUX (p values=) and AUC had a significant difference with TSO. (p =0.0061)

Viability

As a measure of adaptation to urea we compared the viability of each population in the control environment to the urea environment by computing the difference in the two (Fig 3.6). If this difference is positive and significantly different from zero it indicates sensitivity to the toxic effects of urea. The TSO, AUC, CO, and nCO selected groups show significant sensitivity to urea (Figure 3.6). The viability difference is not significantly different from zero in the RUX, UX and UTB groups of populations. This difference points to an adaptation of the urea environment in the RUX, UX and UTB populations, and little to no adaptation in the AUC and demographic populations. It is of some interest that despite more than 100 generations in the control environment the RUX populations still retain an ability to survive well in urea laced food.

The viability of the AUC in the control environment was not significantly greater than UX and RUX group of populations, but was significantly greater than UTB (Figure 3.7). In the urea environment, AUC viability was significantly less then UX, RUX and UTB (Figure 3.7). In both environments, the TSO selection regime has lower viability than the CO and nCO regimes. There were no significant differences between the CO and nCO regimes in either environment (Fig 3.8).

Developmental Time

The developmental time of the *D. melanogaster* is calculated as the time it took for the first instar larvae to pupate and eclose. These experiments were run in three blocks: (i) AUC, RUX and UX, in Spring 2017 (ii) TSO, CO, NCO, and UTB in Spring 2018, (iii) UTB, UX, CO,

and NCO in Spring of 2019. The block structure will become important when we try to compare populations for significant differences. As an example, to compare all the urea adapted populations and the control, AUC populations require observations from blocks (i) and (iii). These two blocks have in common the UX populations and thus they can be used as a standard to assess differences among the group as a whole.

The first analysis is on the difference between the urea development time and the control development time (Figure 3.9). Here the mean development time in the control environment for each selection/sex/population/replicate sample was calculated and then subtracted from the corresponding development time in urea. Thus, a positive value for this difference indicates that the larva takes longer to develop in urea. Prior evidence suggests urea slows development. The populations that have been selected for urea resistance, RUX, UX and UTB, show the smallest development time difference consistent with their adaptation to urea. However, all populations show a development time difference that is positive and significantly different than zero (see confidence intervals in Figure 3.9. Thus, even populations adapted to urea show delayed development in urea laced food.

We assessed differences in development time among demographically selected populations separately from urea selected populations. To evaluate the demographic populations, we used blocks (ii) and (iii) making UTB the standard. Thus, in block (ii) we computed the mean UTB-urea development time, and subtracted this from the TSO, CO, and nCO, control and urea observations. The same was done for block (iii) using the UTB-urea mean from block (iii) and subtracting that mean from the CO and nCO observations. The actual mean development times for the relevant UTB populations were: females: UTB-urea(ii) 301.8, UTB-urea(iii) 243.0; Males: UTB-urea (ii) 302.8, UTB-urea (iii) 245.8. For the control environment: Females: UTB-

control (ii) 262.5, UTB-control (iii) 227.9; Males: UTB-control (ii) 266.0, UTB-control (iii) 231.4. In the control environment, there are no significant differences between males and females from the CO, nCO and TSO selected groups (Figure 3.10). In the urea environment, the TSO females developed significantly slower than the CO and nCO flies. Likewise, the TSO males developed more slowly than the CO and nCO flies. There were no significant differences between the nCO and CO flies.

Then we analyzed the blocks (i) and (iii) using UX-urea as the standard to assess development time differences among the urea populations. Thus, in block (i) we computed the mean UX-urea, and this mean was subtracted from the other control and urea observations. The same was done for block (iii) using the UX-urea mean from block (iii). The actual mean development times in urea for the relevant UX selection treatment were: UX-female (i) 270.7, UX-male (i) 274.7, UX-female (iii) 255.5, UX-male (iii) 257.4. For the control environment: ux-female (i) 247.8, UX-male (i) 251.7, UX-female (iii) 229.6, UX-male (iii) 232.9. Thus, in the urea environment the UX phenotype is 0 in the control environment it is: females (-24.4), males (-23.75). We found that in the urea environment, AUC females and males had significantly greater development times than their respective sexes in UTB, but no difference with UX or RUX (Figure 3.10). In the control environment, the development time of the AUC females were not significantly greater than UTB, RUX or UX, while AUC males were significantly less than RUX, but not different than UX or UTB (Figure 3.11).

Larval Growth Rate

The larval growth rate assay provided dry weight of larvae for all 8 groups of populations – TSO, TDO, AUC, UX, RUX, UTB, CO and nCO at hours 24, 30, 36, 42, 48, 54, 60, 66, 72, 78, 84, 90 and 105. The pupa that were collected at the 105-hour mark were weighed and dried.

A separate collection of pupa was collected for the adult weights. The mean dry weights of each selection regime and the fitted growth equation [eq. (2)] are shown in Figure 3.12.

We connected the predicted individual larval dry weights, with 95% confidence intervals, as a function of feeding rate (Figure 3.13). The *p*-values are for tests of the hypothesis that the slope of the line is zero. The order of the eight selection regimes from slowest feeding to fastest was RUX, UX, UTB, AUC, nCO, TSO, CO, and TDO. The *x*-axis coordinates are offset to improve readability. A significant correlation was seen in all 8 groups of populations at the hours 48 (p=0.004), 54 (p=0.002) and 60 (p=0.002), showing that a slower feeding rate results in a slower growth rate.

Dry Weight

From the pairwise tests comparing the adult female dry weights from the flies of the larval growth rate experiment, we see no significant differences between any selection regimes (Figure 3.14). The males divided into three groups which had one or more members that showed significant differences from the other groups (Figure 3.15). The TDO was significantly different with TSO, RUX, UX and UTB. CO was significantly different with TSO, nCO, and UTB. Finally, nCO was significantly different with RUX, UTB and UX.

DISCUSSION

Previous research on urea adapted lines have shown changes in larval feeding rates, larval foraging paths, and viability in larval urea environments (Shiotsugu *et al.*, 1997, Borash *et al.* 2000, Mueller *et al.*, 2005). However, the relationship between larval feeding rates and food consumption rates has been questioned (Kaun *et al.*, 2007). Our goal was to test whether larval feeding rates affect growth rates as suggested by Mueller and Barter (2015).

We found significant decreases in feeding rate correlated with a slower growth rate. Feeding efficiency, the fraction of ingest food that is digested by larvae, is a decreasing function of feeding rate. Empirical support for this claim comes from a comparison of two different sets of crowding-adapted (fast feeding) and control (slow feeding) *Drosophila* populations (Joshi and Mueller 1996; Mueller *et al.* 1991). These studies showed that fast feeding larvae required more food to reach the same critical minimum size as control larvae. Other experiments performed on feeding rate (Joshi *et al.* 1988, Bakker 1961 1961) support the view that larval feeding rate is an important component of competitive ability of *Drosophila* in high-density conditions. The results of Joshi *et al.* 1988 and Burnet *et al.* 1977 indicate that *Drosophila* larva retraction rates are good indicators of the ability to compete for limiting resources. With the significant slower feeding rates of the urea adapted lines – UTB, UX and RUX, it is quite clear that feeding rates are affected by environmental factors, as well as a correlation with a slower growth rate.

While allele fixation could have occurred in the urea-adapted populations, other traits such as viability, and development time, do not show any evidence of inbreeding depression. Also, both the UX, and the UTB groups of populations, created at different times in the phylogeny of the populations in our research laboratory, display similar traits (feeding rate, developmental time, viability, larval growth rate) in their larval urea environments. Some of

these traits, like development time, have reverted to their ancestral state in the reverse selected RUX population group. The RUX populations have been in a control environments for 108 generations.

Etienne *et al.* 2001 found that larvae of urea adapted lines appear to have adapted to urea exposure by decreasing the ability of urea to enter the body larvae. This correlation with a slower feeding rate might be a reasonable explanation for the lower levels of urea in the *D. melanogaster* larvae that Etienne found. This would be something to test in the future in our UTB, UX and RUX populations of *D. melanogaster* larvae.

All populations were removed from selection for two generations prior to the experiments. This eliminates non-genetic influences from confounding the results, such as maternal and environmental effects. Therefore, any phenotypic differences between populations reflect genetic differences that have arisen due to the selection regime. The UX and UTB populations displayed slower feeding rates, slower growth rates, higher viability, faster developmental time, and lower starvation resistance when reared on food supplemented with urea, compared to the other populations reared, as larvae, on urea.

Though the TSO and TDO lines reverted to their ancestral state regarding desiccation resistance, starvation resistance was still significantly higher in both of these selection treatments. *D. melanogaster* can evolve increased resistance to desiccation by decreasing water loss rates and by increasing bulk water content but not by increasing metabolic water content or dehydration tolerance (Archer *et al* 2007). While glycogen is involved in water storage, its primary role is in water binding, not the production of metabolic water (Archer *et al* 2007). Future studies can focus on glycogen storage in the urea-adapted lines to see if they have lower glycogen levels and high water loss rates due to their significantly lower resistance to starvation

in the environment.

The study of trade-offs has been a fundamental component of modern theory of life-history evolution (van Noordwijk *et al.* 1986, Zera *et al.* 2001, 2011, Roff *et al.* 2007). Trade-offs occur in the *D. melanogaster* larvae and adult flies. The urea has toxic effects on the larvae, and there is some observed toxic effect on the adult flies, such as lower starvation resistance. Joshi *et al* 1996 found that the urea was not toxic to the adult flies, since they only suffered a decline in fecundity and not longevity. A model proposed by Mueller *et al.* 2015 stated that lower larval feeding rates could be favored by natural selection if the resulting larvae had a higher viability in a toxic environment. Not only do we see a significant difference in the feeding rates of the larvae, but this is correlated with a slower growth rate, supporting the theory proposed in Mueller *et al.* (2015).

Future studies will focus on the genomic analysis of the urea adapted lines compared to various control population groups. Potential projects would be to look at urea consumption in the larvae vs concentration of urea in the larvae and looking at glycogen storage.

REFERENCES

Archer, M.A., Bradley, T.J., Mueller, L.D, and M.R. Rose. 2007. Using Experimental Evolution to study the physiological mechanisms of desiccation resistance in Drosophila melanogaster. Physiological and Biochemical Zoology 80(4):386–398.

Bakker, K. 1961. An analysis of factors which determine success in competition for food among larvae of *Drosophila* melanogaster. Archs. Neerl. Zool. 14: 200-281

Borash, D.J., Gibbs, A.G., Joshi, A., Mueller, L.D., 1998. A genetic polymorphism maintained by natural selection in a temporally varying environment. American Naturalist 151, 148–156.

Borash, D. J., Teotónio, H., M. R. Rose, and L. D. Mueller. 2000. Density-dependent natural selection in Drosophila: correlations between feeding rate, development time, and viability. Journal of Evolutionary Biology 13:181-187.

Botella, L.M., Moya, A., Gonzalez, M.C., Mensua, J.L., 1985. Larval stop, delayed development and survival in overcrowded cultures of *Drosophila melanogaster*. Effect of urea and uric acid. Journal of Insect Physiology 31, 179–185.

Bowlus, R. D. and Somero, G. N. (1979). Solute compatibility with enzyme function and structure: rationales for the selection of osmotic agents and end-products of anaerobic metabolism in marine invertebrates. J. Exp. Zool. 208, 137–152.

Burnet, B., D. Sewell, and M. Bos. 1977. Genetic analysis of larval feeding rate behavior in *Drosophila melanogaster.* II. Gen. Res. 30: 149-161

Charlesworth, B., 1993 Natural selection on multivariate traits in age-structured populations. Proc. Biol. Sci. 251: 47–52. https://doi.org/10.1098/rspb.1993.0007

Chippindale, A.K., Leroi, A.M., Kim, S.B., and Rose, M.R. 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., Hoang, D.T., Service, P.M., Rose, M.R., 1994. The evolution of development in *Drosophila melanogaster* selected for postponed senescence. Evolution 48, 1880–1899.

Chippindale, A.K., Chu, J.F., Rose, M.R., 1996. Complex trade-offs and the evolution of starvation resistance in *Drosophila*. Evolution 50, 753–766.

Cooper, A.J.L., Plum, F., 1987. Biochemistry and physiology of brain ammonia. Physiology Review 67, 440–519.

David, C.L., Pierce, V.A., Aswad, D.A., Gibbs, A.G., 1999. The effect of urea exposure on

isoaspartyl content and PIMT activity in Drosophila melanogaster. Comparative Biochemistry and Physiology B. 124(4):423-7

Etienne, R., Fortunat, K., V. Pierce. 2001. Mechanisms of urea tolerance in urea-adapted populations of *Drosophila melanogaster*. The Journal of Experimental Biology 204, 2699–2707

Flatt, Thomas. 2020. Life-History Evolution and the Genetics of Fitness Components in *Drosophila melanogaster*. Genetics, Vol. 214, 3–48, https://doi.org/10.1534/genetics.119.300160

Fellowes MDE, Kraaijeveld AR, Godfray HCJ (1999) Association between feeding rate and parasitoid resistance in Drosophila melanogaster. Evolution 53:1302–1305

Graves, J.L., E.C. Toolson, C. Jeong, L.N. Vu, and Rose. M.R. 1992. Dessicaation, flight, glycogen, ad postponed senescence in *Drosophila melanogaster*. *Physiol. Zool.* 65, 268-286.

Joshi, A, L.D. Mueller. 1988. Evolution of Higher Feeding rate in Drosophila due to density-dependent natural selection. Evolution, 42(5): 1090-1093

Joshi, A, J. Shiotsugu, L.D. Mueller. 1996. Phenotypic enhancement of Longevity by environmental Urea in *Drosophila melanogaster*. Experimental Gerontology. 31:4, 533-544

Joshi, A, Do, M.H., and L.D. Mueller. 1999. Poisson distribution of male mating success in laboratory populations of Drosophila melanogaster. Genet. Res., Camb. (1999), 73, pp. 239–249.

Kaun, K. R., C.A.L. Riedl, M. Chakaborty-Chatterjee, A.T. Belay, S.J. Douglas, A.G. Gibbs, and M.B. Sokolowski. 2007. Natural variation in food acquisition mediated via a *Drosophila* cGMP-dependent protein kinase. J. Exp. Biol. 210: 3547-3558

Lande, R., 1982 A quantitative genetic theory of life history evolution. Ecology 63: 607–615. https://doi.org/10.2307/1936778

Lande, R., and S. J. Arnold, 1983 The measurement of selection on correlated characters. Evolution 37: 1210–1226. https:// doi.org/10.1111/j.1558-5646.1983.tb00236.x

Mueller LD (1990) Density-dependent natural selection does not increase efficiency. Evol Ecol 4:290–297

Mueller LD, Gonza´lez-Candelas F, Sweet VF (1991) Components of density-dependent population dynamics: models and tests with Drosophila. Am Nat 137:457–475

Mueller, L.D., A. Joshi, M. Santos, and M.R. Rose. 2013. Effective population size and

evolutionary dynamics in outbred laboratory populations of Drosophila. Journal of Genetics 92:349-361.

Mueller, L.D. and T.T. Barter. 2015. A model of the evolution of larval feeding rate in Drosophila driven by conflicting energy demands. Genetica 143:93-100

Partridge, L. and P. H. Harvey. 1988. The ecological context of life history evolution. Science (Washington D.C.) 241:1449-1445

Pinheiro, J.C. and D. M. Bates. 2000. Mixed-effects models in S and S-PLUS. New York:Springer.

Prasad, N. G., and A. Joshi, 2003 What have two decades of laboratory life-history evolution studies on Drosophila melanogaster taught us? J. Genet. 82: 45–76. https://doi.org/10.1007/BF02715881

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Roff, D. A. 1992 The evolution of life histories. Chapman & Hall, New York.

Roff DA, Fairbairn DJ (2007a) The evolution of trade-offs: Where are we? J Evol Biol 20:433–447. doi:10.1111/j.1420-9101.2006. 01255.x

Rose, M.R., 1984. Laboratory evolution of postponed senescence in *Drosophila melanogaster*. Evolution 38, 1004–1010.

Rose, M.R., *Evolutionary Biology of Aging.* Oxford University Press. New York, 1991

Santos, M., Borash, D. J., Joshi, A., Bounlutay, N. & Mueller, L. D. (1997). Density-dependent natural selection in Drosophila: evolution of growth rate and body size. Evolution 51, 420–432.

Service, P.M. 1987 Physiological mechanisms of increased stress resistance in Drosophila melanogaster selected for postponed senescence. Physiol. Zool. 60: 321-326

Service, P.M. 1989The effect of mating status on lifespan, egg laying, and starvation resistance in *Drosophila melanogaster* sin relation to selection on longevity. J. Insect. Physiol. 35: 447-452

Sewell, D., B. Burnet and K. Conolly. 1975. Genetic Analysis of larval feeding behavior in Drosophila melanogaster. Gen. Res. 24: 163-173

Shiotsugu, J., Leroi, A.M., Yashiro, H., Rose, M.R., Mueller, L.D., 1997. The symmetry of correlated selection responses in adaptive evolution: an experimental study using *Drosophila*. Evolution 51, 163–172.

Somero, G.N., Yancey, P.H., 1997. Osmolytes and cell-volume regulation: physiological and evolutionary principles. In: Hoffman, F.F., Jamieson, J.D. (Eds.), Handbook of Physiology. Oxford University Press.

Stearns, S.C. 1992. The evolution of life histories. Oxford University Press, New York.

van Noordwijk AJ, de Jong G (1986) Acquisition and allocation of resources: their influence on variation in life history tactics. Am Nat 128:137–142

Yancey, P . H. (1985). Organic osmotic effectors in cartilaginous fishes. In Transport Processes, Iono- and Osmoregulation (ed. R. Gilles and M. Gilles-Ballien), pp. 424–436. Berlin: Springer-Verlag.

Yancey, P. H. (1992). Compatible and counteracting aspects of organic osmolytes in mammalian kidney cells in vivo and in vitro. In Water and Life: A Comparative Analysis of Water Relationships at the Organismic, Cellular, and Molecular Levels (ed. G. N. Somero, C. B. Osmond and C. L. Bolis), pp. 19–32. Berlin: Springer-V erlag.

Yancey, P. H. and Somero, G. N. (1979). Counteraction of urea destabilization of protein structure by methylamine osmoregulatory compounds of elasmobranch fishes. Biochem. J. 182, 317-323.

Zera AJ, Harshman LG (2001) The physiology of life-history trade- offs in animals. Annu Rev Ecol Syst 32:95–126

Zera AJ, Harshman LG (2011) Intermediary metabolism and the biochemical-molecular basis of life history variation and trade- offs in two insect models. In: Flatt T, Heyland A (eds) Mechanisms of life history evolution: the genetics and physiology of life history traits and trade-offs. Oxford University Press, Oxford

FIGURES

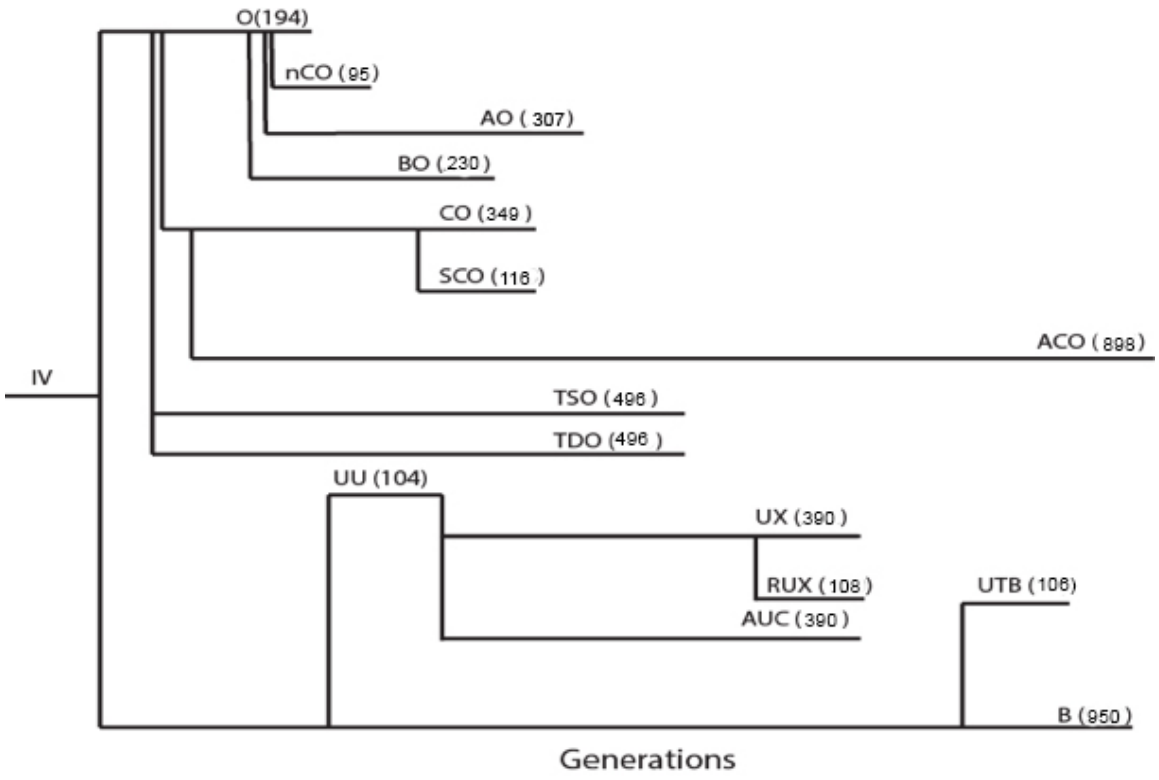

Figure 3.1. The phylogeny of all current populations in the Rose/Mueller lab. The number by each population group is the total number of generations since their derivation.

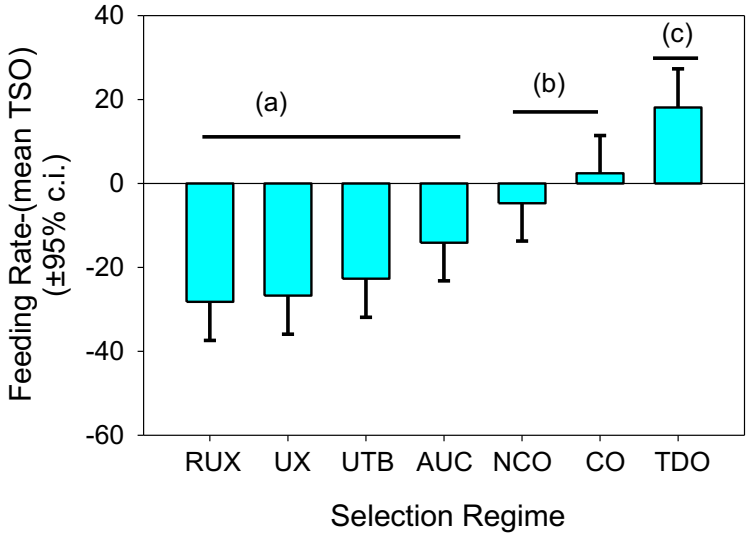

Figure 3.2 Feeding rate comparison between all 8 groups of populations compared to the TSO group. The horizontal bars mark groups that are not significantly differentiated form each other. Between the separate groups are significant differences, with TDO feeding the fastest and RUX feeding the slowest.

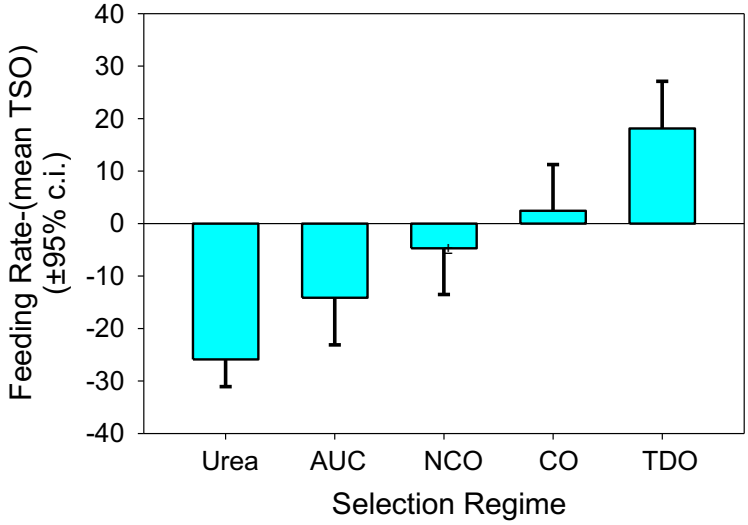

Figure 3.3 Feeding rate of the urea selection treatments grouped together compared to the demographic lines. A one-sided test comparing the Urea pooled populations (UX, UTB and RUX) against the AUC control, nCO, CO, TSO and TDO. The mean feeding rate difference for the fifteen urea populations was -25.88. This is significantly less than the AUC feeding rate difference (-14.13) with p=0.014. The urea also fed significantly slower than the nCO, CO, TSO and TDO.

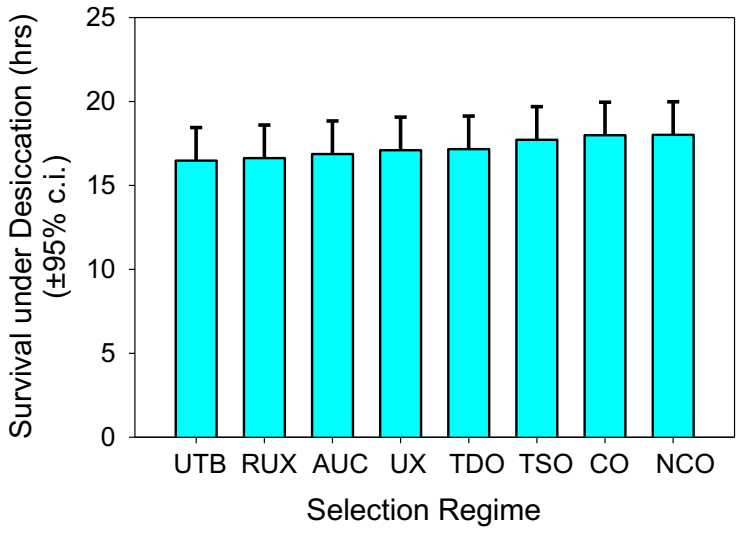

Figure 3.4 Desiccation resistance results. No significant difference was observed between any of the populations in their resistance to desiccation.

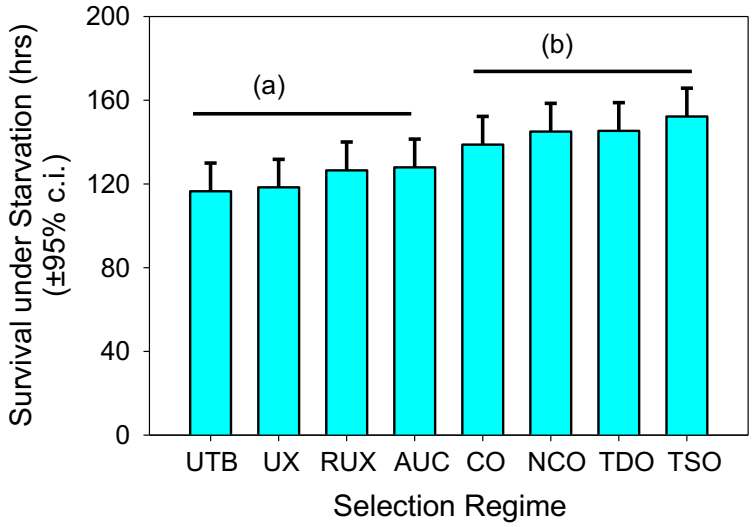

Figure 3.5 S**tarvation resistance results.** No difference between the populations under the horizontal bars, but significant difference was seen in some population comparisons. UTB had a significant difference between TSO, TDO, NCO and CO. UX had a significant difference between TSO, TDO and NCO. RUX and AUC had a significant difference with TSO.

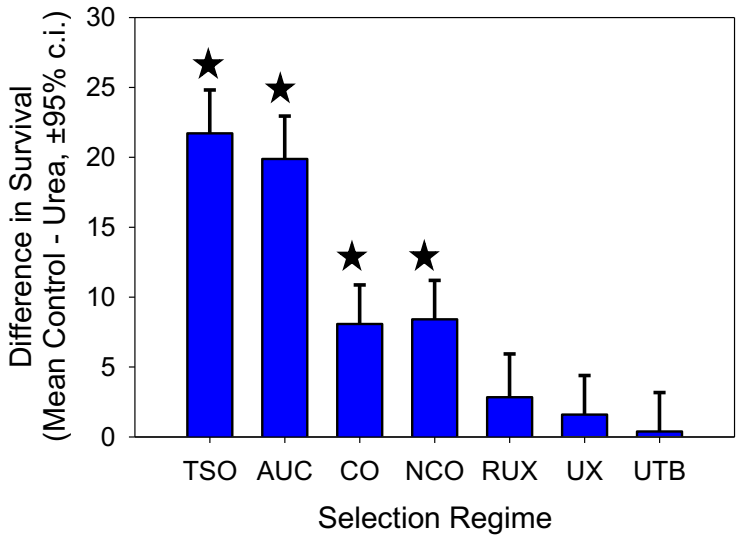

Figure 3.6 Difference in survival subtracting the urea viability of the population from the control viability of the population. The * indicates differences that are significantly different from 0. The bars are 95% confidence intervals. The difference of AUC is significantly greater than RUX, UX, and UTB in each case with a $p<0.0001$.

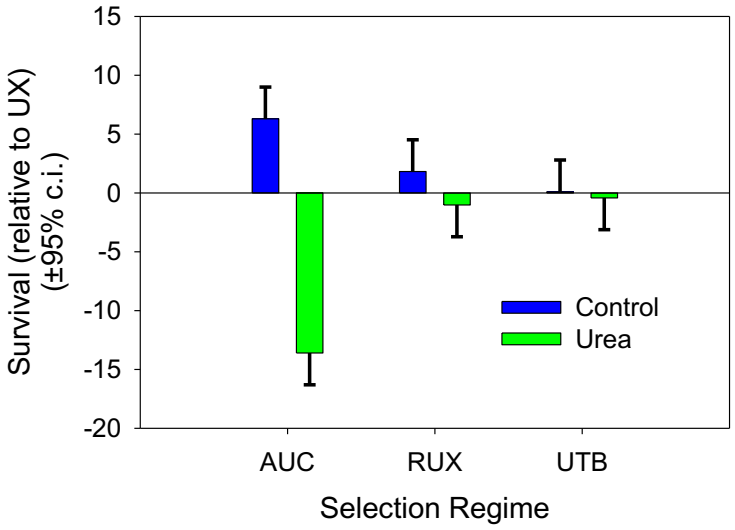

Figure 3.7 Survival of the urea adapted lines compared to AUC. The urea lines (RUX, UTB, UX) and control (AUC). Survival is shown relative to the UX- urea mean survival in each block. This makes the mean UX survival in urea 0 and in the control environment, 1.5. In the control environment, AUC did not have a significant survival than UX, RUX, but was significantly greater than UTB. In the urea environment, AUC is significantly less then UX, RUX and UTB.

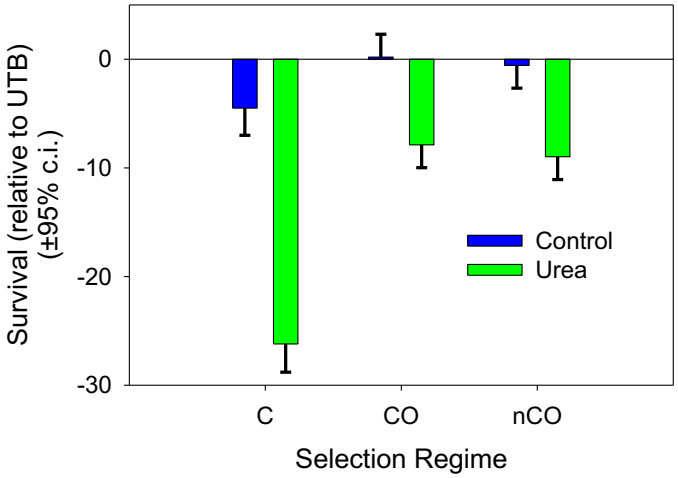

Figure 3.8 Survival of the demographic lines (TSO, CO, nCO) related to UTB. Survival is shown relative to the UTB- urea mean survival in each block. This makes the mean UTB survival in urea 0 and in the control environment, 0.39. In both environments the TSO selection regime has lower viability than the CO and nCO regimes. There are no significant differences between the CO and nCO regimes in either environment.

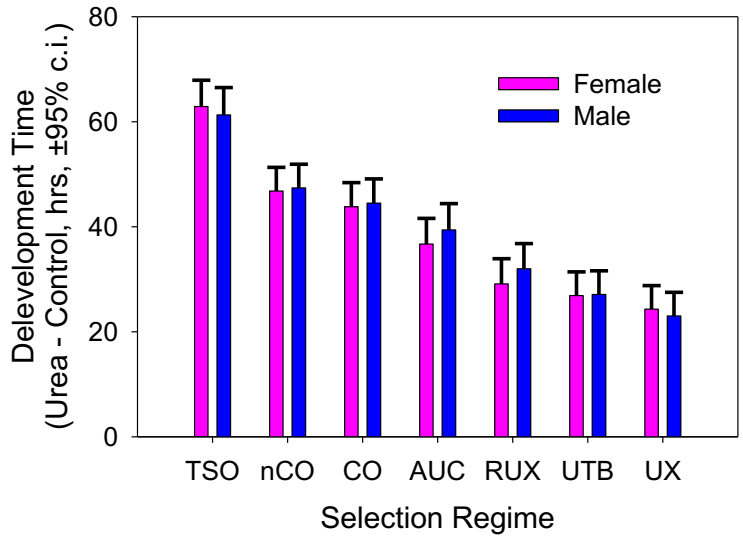

Figure 3.9 Development time of all 7 groups of populations. Development times of each group of population in urea food relative to the development time in control (banana) food for 7 groups of populations – TSO, nCO, CO, AUC, RUX, UTB, and UX.

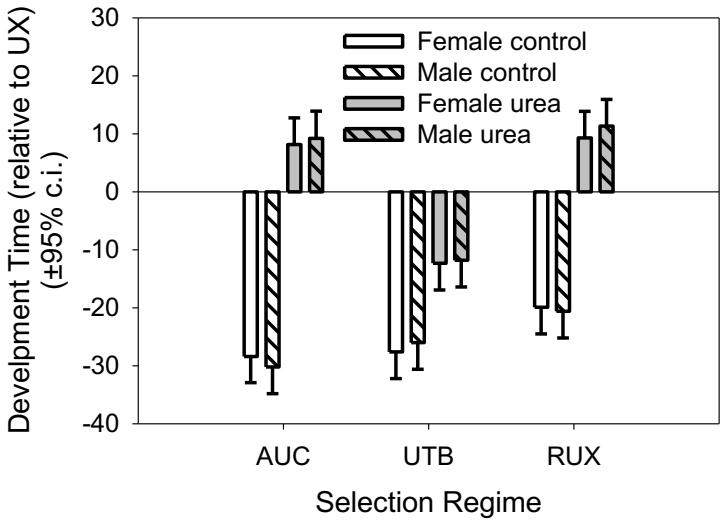

Figure 3.10 Development time of the urea adapted lines (RUX, UTB, UX) and control (AUC). Development time, in hours, is shown relative to the UX- urea mean development time in each block. Thus, fast developing selected groups have large negative times and slow developing populations have large positive values.

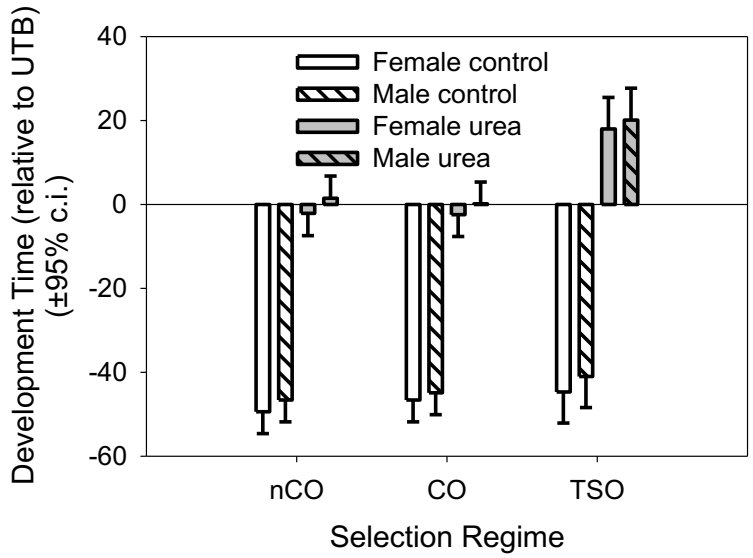

Figure 3.11 Development time of demographic lines relative to UTB. Development time, in hours, is shown relative to the UTB- urea mean development time in each block for the demographic lines – nCO, CO and TSO. Thus, fast developing populations have large negative times and slow developing populations have large positive values.

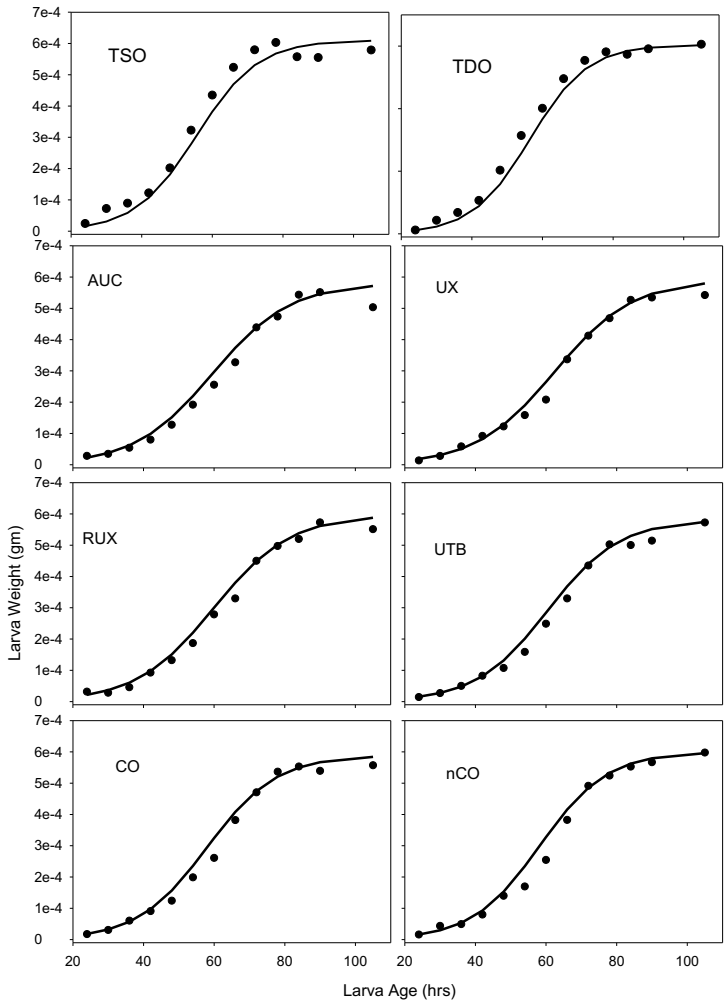

Figure 3.12 Dry weight of larvae from larval growth rate experiment. The dry weight of larvae at various ages for all 8 groups of populations – TSO, TDO, AUC, UX, RUX, UTB, CO and nCO. The points are the average weights over the five replicate populations. The lines are the predicted weights from model larva.nlme6.

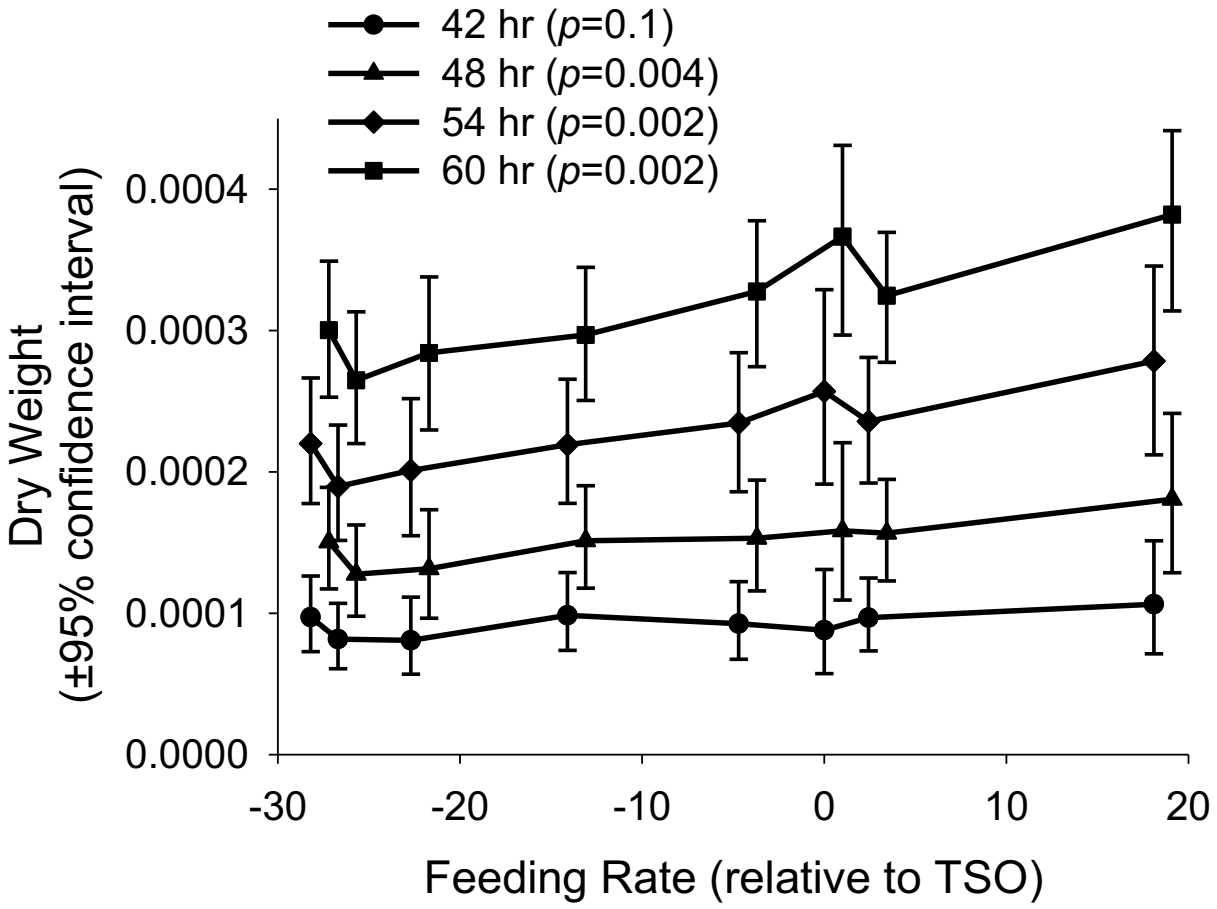

Figure 3.13 Predicated larval dry weights as a function of feeding rate. The predicted individual larval dry weight, with 95% confidence intervals, as a function of feeding rate. The *p*-values are for tests of the hypothesis that the slope of the line is zero. The order of the eight selection regimes from slowest feeding to fastest is RUX, UX, UTB, AUC, nCO, TSO, CO, and TDO. The *x*-axis coordinates are offset to improve readability.

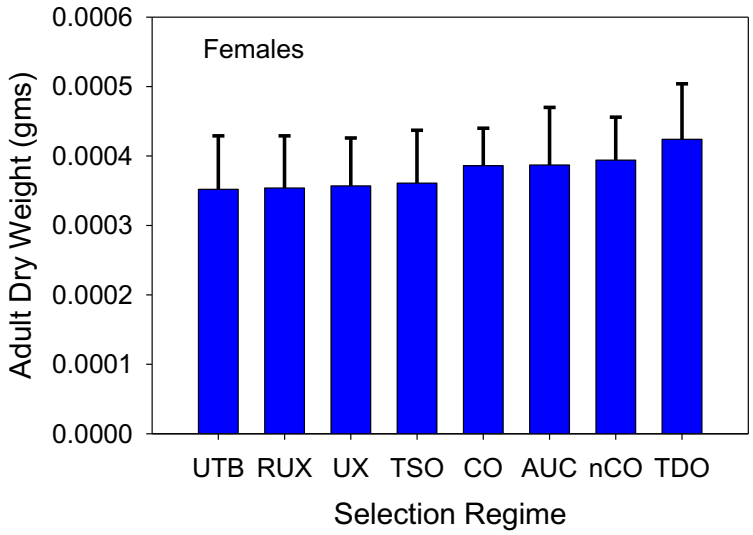

Figure 3.14 The adult female dry weights from the flies of the larval growth rate experiment. From the pairwise tests, we see no significant differences between any selection regimes.

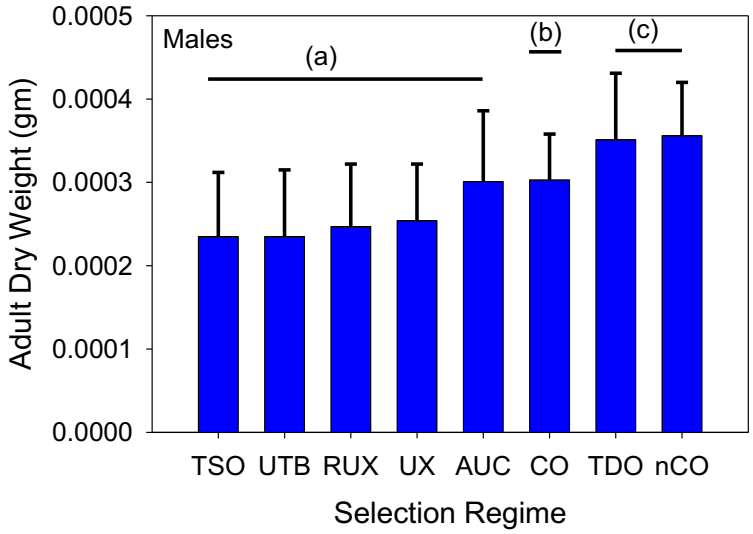

Figure 3.15 The adult male dry weights from the flies of the larval growth rate experiment. Males divide into three groups which have one or more members that show significant differences from the other groups.

TABLES

Table 3.1. All 8 groups populations of *D. melanogaster* used in the phenotypic assays. The number of days spent in vials, number of days spent in cages, their generation time and if they lived in a larval environment with urea is specified per population group. See section 2.1 of the Materials and Methods for a description of each population group.

Population*	Days in Vials	Days in cages before egg collect	Generation Time	Urea in larval environment
AUC	14	7	21	No
RUX	14	7	21	Previously yes
UX	14	7	21	Yes
UTB	14	7	21	Yes
CO	14	14	28	No
nCO	14	14	28	No
TSO	14	14	28	No
TDO	14	14	28	No

Table 3.2. Phenotypic assays performed on the 8 populations

Population	Feeding Rate	Developmental Time	Viability	Larval growth rate	Starvation Resistance	Desiccation Resistance
AUC	Yes	Yes	Yes	Yes	Yes	Yes
CO	Yes	Yes	Yes	Yes	Yes	Yes
nCO	Yes	Yes	Yes	Yes	Yes	Yes
RUX	Yes	Yes	Yes	Yes	Yes	Yes
TSO	Yes	Yes	Yes	Yes	Yes	Yes
TDO	Yes	No	No	Yes	Yes	Yes
UTB	Yes	Yes	Yes	Yes	Yes	Yes
UX	Yes	Yes	Yes	Yes	Yes	Yes

CHAPTER 4

Genomic Differentiation Among Urea Selected Populations

INTRODUCTION

Genome-wide sequencing of experimentally evolved populations has emerged as a powerful method for understanding the genetic basis of adaptation (Burke *et al.* 2010; Turner *et al.* 2011; Tenaillon *et al.* 2012; Schlotterer *et al.* 2015, Graves e*t al.* 2017). Various environmental stresses affect both vertebrate and invertebrate organisms throughout their lifetime, and both will experience and adapt to such stresses, usually affecting their physiological machinery. This stress can be manipulated in a laboratory setting, allowing us to create differentiated populations with which to study adaptation and measure physical robustness and evolution.

Previous work has shown that *Drosophila* populations can evolve resistance to high levels of environmental urea (Joshi *et al.* 1996, Shiotsugu *et al.*, 1997, David *et al.* 1999, Etienne *et al.* 2001), a compound that stresses the *Drosophila* larvae. Previous experiments in our research laboratory have examined adaptations in populations of *D. melanogaster* to high levels of environmental urea and ammonia. Although Botella *et al*. (1985) suggested urea might be the primary nitrogen waste product of *Drosophila* larvae more recent work by Borash *et al*. (1998) showed that only ammonia accumulated in the food of crowded *Drosophila* cultures. Urea appears to have general cytotoxic effects due to interference with translation, action as protein denaturant (Somero and Yancey, 1997), and its reduction of enzyme activity and thermostability (Bowlus and Somero, 1979; Yancey and Somero, 1979; Yancey, 1985; Yancey, 1992; Somero and Yancey, 1997).

Previous research on urea adapted lines have shown changes in larval feeding rates, larval foraging paths, and viability in larval urea environments (Shiotsugu *et al.*, 1997, Borash *et al.* 2000, Mueller *et al.*, 2005). However, the relationship between larval feeding rates and food

consumption rates has been questioned (Kaun *et al.*, 2007). A model proposed by Mueller *et al*. 2015 predicted that lower larval feeding rates could be favored by natural selection if the resulting larvae had a higher viability in a toxic environment. In chapter 3, experiments were presented to support the clear differentiation of the urea adapted lines compared to the control populations for their viability in urea laced food and other phenotypes. Not only do we see a significant difference in the feeding rates of the larvae, but this is correlated with a slower growth rate, supporting the theory proposed in Mueller *et al.* 2015.

We developed three replicated selection regimes of *D. melanogaster* that are currently, or have at some point in the past, been exposed to many generations of urea exposure in their larval food. Previous research on urea adapted lines have shown changes in larval feeding rates, larval foraging paths, and viability in urea laced environments (Shiotsugu *et al.*, 1997; Borash *et al.*, 2000; Mueller *et al.*, 2005). The UX, and UTB populations, both currently exposed to urea in the larval environment, were created at different times in the phylogeny of populations studied in our laboratory. The UX and UTB group of populations display phenotypic differentiation from the controls, the AUC populations, for feeding rate, developmental time, viability, and larval growth rate (see Chapter 3). The RUX selection treatment populations, which are the reverse selected urea adapted lines, were found to have reverted back to an ancestral state for development time, but still were differentiated in feeding rate, viability and larval growth rate (see Chapter 3).

We studied genomic differentiation among populations subjected to different types of selection for tolerance to urea in their larval food. The five AUC populations served as controls and were fed normal food and kept on a three-week generation cycle like the other populations. The UX and UTB populations were raised on larval food containing urea up to the time that genomic samples were collected. The RUX populations were derived from the UX populations

after 282 generation of selection for urea tolerance and were placed back on standard food for 108 generations prior to the genomic samples.

METHODS

1.1 Populations

All the populations in the lab are derived from an established Rose IV population (Rose, 1984, Figure 4.1). They were maintained on a banana-molasses food (Rose, 1984) at 25^{o}C (24 h light) and uncontrolled humidity, and have a generation time of approximately 3 - 4 weeks (Table 4.2). All selection regimes were five-fold replicated, uncrowded as larva (60–80 eggs per 8-dram vial), with emergent adults kept at a low density of approximately 50–60 flies per 8-dram vial, and transferred to a cage environment with fresh food given about every other day for approximately 1 week. Effective populations sizes for each line were in the range of 700-1000 (Mueller *et al.*, 2013) every generation.

The UX and the UTB populations were subjected to selection for increased larval tolerance to the presence of toxic levels of urea in the food. The levels of urea were increased every few generations, when it was observed that a great proportion of larvae were surviving to adulthood. The derivation of the urea-tolerant (UX) and unselected controls (AUC) selection regimes was in the Fall of 1996. Both populations were derived from a five-fold replicated set of populations called UU, which had a 3-week generation time, and were reared at low larval and adult densities. The UU populations were derived in 1990, from the Rose B populations (Rose, 1984; Chippindale *et al.*, 1994, 1996). The RUX are a reverse selected line of the UX. They have been in a control environment for 108 generations. The UTB line were created in October 2013. The AUC, RUX, UX, and UTB lines are all 21-day cycle flies.

1.2 DNA Extraction and Sequencing

Genomic DNA was extracted in May 2019 from a large sample of 200 female flies using the Qiagen/Gentra Puregene kit. The manufacturer's protocol for bulk DNA purification were

followed. 30 gDNA pools were prepared as standard 200–300bp fragment libraries for Illumina sequencing. A DNA pooled sample of about 30 female flies was sequenced in June 2019 from the large sample of 200 female flies. A second sequencing was conducted on the other 180 approximate female flies in December 2019. The DNA was stored in a -20 C freezer during the time between June and December. Libraries were run across PE100 lanes of an Illumina HiSEQ 2000 at the UCI Genomics High throughput Sequencing Facility and constructed such that each five replicate populations of a treatment (e.g., UX1–5) were given unique barcodes, normalized, and pooled together (Graves *et al.* 2017). Each 5-plex library was run on individual PE100 lanes of an Illumina HiSEQ 2000 at the UNC High Throughput Sequencing Facility and the resulting data were 200 bp paired-end reads. Each population was sequenced two times; data from both runs were combined for analysis. Combining reads from two independent sequencing runs likely alleviates the effects of possible bias introduced from running all replicates for each population in the same lane.

1.3 SNP Analysis Read Mapping and Preprocessing.

We first trimmed the reads to remove low-quality bases using a script provided in the PoPoolation software package (Kofler, Orozcoter Wengel, *et al.* 2011) separately of the June and then the December data sets. We then mapped reads with the Burrows-Wheeler Aligner (BWA) software package (Li and Durbin 2009) against the *D. melanogaster* reference genome (release 6.31) using bwa mem with default settings with the following mapping parameters: -n 0.01 (error rate), -o 2 (gap opening), -d 12 and -e 12 (gap length), and -l 150 to effectively disable the seed option. The SAM files were filtered for reads mapped in proper pairs with minimum quality of 20 and converted to the BAM format using SAMtools (Li *et al.* 2009). The combined December and August databases yielded an average coverage of 69.1, assuming a genome size of 137 Mb.

The rmdup command in SAMtools was then used to remove potential PCR duplicates. The two BAM files from each population's two sequencing runs were merged using BAMtools to maximize coverage (Barnett *et al.* 2011). These merged BAM files were then all combined in the mpileup format once again using SAMtools. Using PoPoolation2 (Kofler, Pandy *et al.* 2011), the resulting mpileup was converted to "synchronized" files, which is a format that allele counts for all bases in the reference genome and for all populations being analyzed.

1.4 Heterozygosity

We calculated and plotted heterozygosity across the five major chromosome arms to see if we could find any evidence of selective sweeps and to determine if there was convergence in overall patterns of variation. To do this, SNPs were first called across all 20 populations used in this study from our synchronized file. SNPs where discarded if coverage in any of the populations was less than 4X or greater than 500X. We also required a minimum minor allele frequency of 2% across all eight populations. A SNP table with major and minor allele counts for each SNP in each population was then generated. Using these counts, heterozygosities were calculated and plotted over 150 kb non-overlapping windows. We also performed *t*-tests comparing mean genome-wide heterozygosities between different groups of populations.

1.5 F_{ST} estimates

F_{ST} estimates for replicate populations were obtained using the formula: $F_{ST} = \frac{(H_T - H_S)}{H_T}$ where H_T is heterozygosity based on total population allele frequencies, and H_S is the average subpopulation heterozygosity in each of the replicate populations (Hedrick, 2009). F_{ST} estimates were made at every polymorphic site in the data set for a given set of replicate populations. This was done to quantify the level of similarity between replicates of our four sets of selection regimes.

1.6 SNP Sampling Simulation

Two separate samples of very different sizes have been analyzed by pooled sequencing techniques. Each of the pooled samples has about the same coverage. Thus, can these two samples be combined to analyze SNP frequencies and if so, how should they be combined?

We assume that the major nucleotide frequency at a genomic site is P in a relatively large population. Two samples of chromosomes are taken from the population, N_1 and N_2. From these samples a pooled DNA sample has DNA fragments amplified at random giving rise to n_1 and n_2 DNA copies in the two samples. We assume that the count of the major nucleotide in sample 1 and 2 is x_1 (out of the n_1 samples) and x_2 (out of the n_2 samples) respectively. There are two obvious ways to estimate P, a simple average from both samples or a weighted average,

$$\hat{P}_1 = \frac{x_1+x_2}{n_1+n_2} \quad (1)$$

$$\hat{P}_2 = \frac{x_1 N_1}{n_1(N_1+N_2)} + \frac{x_2 N_2}{n_2(N_1+N_2)} \quad (2)$$

To determine which of these two estimation procedures is best we generated 10,000 dual samples, a sample of N_1 and N_2 chromosomes and a sample of DNA fragments of n_1 and n_2. We estimated the mean squared error (MSE) for each method at nine different values of P, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, and 0.9 (figure 4.2). These results show that the simple average (Eq. 1) has the smallest MSE at all allele frequencies.

The next question is, how does the use of Eq. 1 using two samples, compare to a single sample of size N_1 with different coverage values (n_1)? With this comparison, we can estimate an effective coverage value by merging the two samples as done by Eq. 1. The effective coverage will be the value of n_1 that gives the same MSE as Eq. 1. We see that two samples with samples of 180 and 30 females has about the same MSE of a single sample of 180 females with a

coverage of 59 (Figure 4.3). Thus, we can consider the effective coverage in our study to be about 59.

1.7 Statistical Tests of Differentiation

We performed two tests to identify significantly differentiated single nucleotide polymorphisms (SNP). The Generalized Cochran- Mantel-Haenszel (CMH, Landis *et al.*, 1978) test examines whether the SNP counts from two or more replicated selection regimes come from the same multinomial statistical distribution. The second test is called the quasibinomial test. It has been suggested that due to random genetic drift in replicated populations and pseudoreplication of SNPs in genomic analysis the CMH test may not be appropriate (Wiberg *et al.*, 2017). The quasibinomial test is a general linear model with selection regimes as treatments and a quasibinomial error distribution. The test statistic is the difference between the null deviance and the model deviance.

The CMH test was implemented with the *mantelhaen.test* R function (R Core Team, 2019). The quasibinomial test was implemented with the *glm* R function using the quasibinomial family and logit link.

These tests are applied to each of several hundred thousand SNP's and consequently the control of false positives becomes an important issue. One simple approach is the Bonferroni inequality which divides the type-I error rate by the number of tests done to determine the *p*-value required for significance. However, this *p*-value may be too severe and result in reduced power. An approach that may be more useful is to control the number of false positives among all positives. These positive results may become the focus of further investigations and thus it would be important not to pursue false leads.

We used a method called the plug-in method for what? (Hastie *et al.*, 2009, chpt. 18).

Suppose we did M total hypothesis tests. If we let V be the number of positive results when the null hypothesis is true and S be the number of positive test results when the null hypothesis is false, then the false discovery rate (FDR) is defined as $V/(V+S)$. A critical test statistic, C, was chosen and the plug-in method computes the FDR for that critical point. Step 1 in the estimation procedure is the estimation of $V+S$. This was simply the total number of test statistics from the M hypothesis tests that exceed C. Let this number be $\widehat{\mathrm{R}}$. To estimate V, we permuted the labels of the replicated populations assigning them genetic data at random. The number of significant test statistics greater than C was then saved. This permutation process was repeated 100 times.

RESULTS

Tests were done on the four different selection regimes contrasting two selection regimes at a time. A summary of these results is shown in Table 4.1. The greatest number of significantly differentiated SNPs are seen between the control populations (AUC) and the urea selected populations (UX and UTB, Table 4.1). The selection regimes with the smallest differences are UX vs. RUX (Table 4.1). Although the RUX populations originated from the UX populations the 108 generations of reverse selection would presumably be the cause of most of the loss of differentiation. However, the RUX populations still maintain measurable levels of adaptation to urea suggesting any reverse selection is slow. Hence, the RUX populations show roughly similar levels of differentiation from the AUC populations, relative to the UX populations. By contrast the UX populations show very large numbers of differentiated SNPs when compared to AUC.

F_{ST} estimates were made from each polymorphic site among the five replicate populations within a selection regime (Table 4.3). All F_{ST} fall in the same range as similar populations in the Rose and Mueller laboratory (Table 4.3). Based on frequencies at each SNP and averaged together, they indicate there is a high degree of similarity between replicate populations with parallel evolutionary histories, as indicated by mean genome wide F_{ST} estimates that are all less than 0.10 from SNP data. There are some notable depressions consistent across replicates that may be indicative of soft sweeps.

In general, we see that the quasibinomial test typically has lower *FDR* but detects a larger number of differentiated SNPs suggesting greater statistical power. The CMH results can be seen in Figures 4.5 through 4.7, and the quasibinomial results in Figures 4.8 through 4.10 comparing AUC with UX, UTB and RUX. In the next chapter the differentiated SNPs will be identified in

50kb regions for use with the FLAM statistical learning tool. We will use the apparent superior performance of the quasibinomial tests to do this screening.

DISCUSSION

We see some depressions in heterozygosity in all four populations, but very few regions have variations that has been completely expunged (Figure 4.4). With consistent sweeps in depressions in heterozygosity across replicate populations, parallel selection may be acting across the replicates. Mean heterozygosity ranged from 0.228 in the UX populations to 0.258 in the RUX population. F_{ST} values were consistently low. below 0.10, indicating a high degree of similarity between the replicates of the populations. The UX populations had the highest F_{ST} values indicating a reduced, but still significant, similarity between the replicates. The RUX populations still maintain measurable levels of adaptation to urea suggesting any reverse selection is slow. This will allow us to combine all 3 populations, the UX, UTB and RUX, into a grouping that we will call the 'Urea populations' for further testing in the following chapter.

Our extraction process resulted in a large amount of DNA collected. The DNA of 30 flies from the large pool were sent off for testing in June 2019, and the DNA of the rest of the approximate 180 flies from the same large pool were sent off for testing in December 2019. Combining the two data sets resulted in the identification of over 1.4 million SNP's that are polymorphic for the Urea populations (UX, UTB and RUX), with a fraction of 1.4/120, or 0.007. The two samples were merged using equation 1, giving us an effective average coverage of 59.

The RUX populations have maintained some of the adaptive phenotypic differentiation of the other urea selected populations like reduced larval feeding rates and increased viability in urea laced food. There are also 10,463 SNPs that are significantly differentiated between the AUC and RUX populations. Of these 10,463 differentiated SNPs in the RUX populations, 6,406 are also differentiated in the UX and UTB populations. These would presumably be SNPs at or close to genes that are important to the urea adaptations that are still differentiated in the RUX

populations. There are also 1,143 differentiated genes in the RUX populations that are not differentiated in either the UX or UTB populations. These may have become differentiated due to drift and some may be false positive. Indeed, about 220 of these differentiated SNPs would be expected to be false positives. The remaining 2,914 differentiated SNPs in the RUX populations are differentiated in just one of the selection regimes UX (1,940) or UTB (974). These may also represent gene regions that are important to urea adaptation, but they have not achieved the requisite level of significance in all three selection regimes or due to different starting conditions became important in adaptation in only one set of populations.

Of the 70,980 differentiated SNPs in the UX populations 50.8% are also differentiated in the UTB populations. This result does not seem to be sensitive to the critical test statistic level. If we reduce the critical quasi-binomial test statistic to 75, although there are many more SNPs that achieve significance only 50.6% of the UX differentiated SNPs are also differentiated in the UTB populations. The UX and UTB populations can trace their ancestry back to the B populations but the samples but were derived at very different times. Genetic variation in the gene regions important for urea adaptation may not have been under strong stabilizing selection. Thus, it is not unreasonable to assume the starting conditions for the important gene regions were very different and this may have led to different trajectories for this variation once the environment was changed.

The idea that natural selection may follow different trajectories depending on their initial conditions was first outlined in Wright's shifting balance theory (Wright, 1982). In this theory Wright postulates a fitness landscape in which populations with different gene frequencies may be located near different fitness peaks. Populations then come under the forces of the local fitness peak and are taken to different local fitness maxima. These ideas have been made

concrete with detailed multilocus population genetic models of viability selection (Franklin and Lewontin, 1970; Feldman *et al.*, 1974). These studies revealed that selection at multiple loci can have multiple stable equilibria. Each of these stable points have their own domain of attraction and hence the end points of natural selection will depend on the initial conditions. We envisage a similar situation with the UX and UTB populations. Nearly 582 generation of B selection passed between the founding of the UX and UTB populations, enough time to generate the different starting conditions that could affect these trajectories.

Further analysis will occur in Chapter 5, using the Fussed Lasso Additive Model to look at polymorphic SNP differentiation. In conclusion, we can see parallel evolution occurring in the replicates of the same environmental populations, as well as genomic differentiation in the Urea adapted lines, UX, UTB and RUX versus the control population of AUC.

REFERENCES

Barnett DW, Garrison EK, Quinlan AR, Stromberg MP, Marth GT. 2011. BamTools: a C þþ API and toolkit for analyzing and managing BAM files. Bioinformatics 27:1691–1692.

Berriz G.F., J.E. Beaver, C. Cenik, M. Tasan, and F.P. Roth. 2009. Next generation software for functional trend analysis. Bio- informatics 25:3043–3044.

Botella, L.M., Moya, A., Gonzalez, M.C., Mensua, J.L., 1985. Larval stop, delayed development and survival in overcrowded cultures of Drosophila melanogaster. Effect of urea and uric acid. Journal of Insect Physiology 31, 179–185.

Bowlus, R. D. and Somero, G. N. (1979). Solute compatibility with enzyme function and structure: rationales for the selection of osmotic agents and end-products of anaerobic metabolism in marine invertebrates. J. Exp. Zool. 208, 137–152.

Burke MK, Dunham JP, Shahrestani P, Thornton KR, Rose MR, Long AD. 2010. Genome-wide analysis of a long-term evolution experiment with Drosophila. Nature 467:587–590.

Burke MK, Barter TT, Cabral LG, Kezos JN, Phillips MA, Rutledge GA, Phung KH, Chen RH, Nguyen HD, Mueller LD, Rose MR. 2016. Rapid convergence and divergence of life-history in experimentally evolved Drosophila Melanogaster. Evolution 70:2085–2098.

David, C.L., Pierce, V.A., Aswad, D.A., Gibbs, A.G., 1999. The effect of urea exposure on isoaspartyl content and PIMT activity in Drosophila melanogaster. Comparative Biochemistry and Physiology B. 124(4):423-7

Feldman, M.W., I. Franklin, and G.J. Thomson. 1974. Selection in complex genetic systems I. the symmetric equilibria of the three-locus symmetric viability model. Genetics 76: 135-162.

Franklin, I. and R.C. Lewontin. 1970. Is the gene the unit of selection? Genetics 65: 707-734.

Graves, J.L, K.L. Hertweck, M.A. Phillips, M.V. Han, L.G. Cabral, T.T. Barter, L.F. Greer, M.K. Burke, L.D. Mueller, and M.R. Rose. 2017. Genomics of experimental evolution. Mol. Biol. Evol. 34(4): 831-842

Hastie T., R. Tibshirani, J. Friedman. 2009. The elements of statistical learning, 2nd ed. Springer, New York

Hedrick PW. Genetics of populations. Massachusetts: Jones & Bartlett Learning Press; 2009.

Joshi, A, J. Shiotsugu, L.D. Mueller. 1996. Phenotypic enhancement of Longevity by environmental Urea in Drosophila melanogaster. Experimental Gerontology. 31:4, 533-544

Kofler R, Orozco-terWengel P, De Maio N, Pandey RV, Nolte V, Futschik A, Kosiol C, Schlotterer C. 2011. PoPoolation: a toolbox for population genetic analysis of next generation sequencing data from pooled individuals. PLoS ONE 6:e15925.

Kofler R, Pandey RV, Schlotterer C. 2011. PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). Bioinformatics 27:3435–3436.

Kofler R, Betancourt AJ, Schlotterer C. 2012a. Sequencing of pooled DNA samples (Pool-Seq) uncovers complex dynamics of transposable element insertions in Drosophila melanogaster. PLoS Genet 8(1):e1002487.

Landis J.R., E.R. Heyman, G.G. Koch. 1978. Average partial association in three-way contingency tables: a review and discussion of alternative tests. Int Stat Rev. 46:237–254.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760.

Orozcoter Wengel P, Kapun M, Nolte V, Kofler R, Flatt T, Schlotterer C. 2012. Adaptation of Drosophila to a novel laboratory environment reveals temporally heterogeneous trajectories of selected traits. Mol Ecol. 21:4931–4941.

Phillips, M.A., G. A. Rutledge, J. N. Kezos, Z. S. Greenspan, A. Talbott, S. Matty, H. Arain,

L.D Mueller, M. R. Rose and P. Shahrestani. 2018. Effects of evolutionary history on genome wide and phenotypic convergence in Drosophila populations. BMC Genomics 19: 743

R Core Team. 2019. R: A language and environment for statistical computing. R Foundation for Statistical

Rose MR, Passananti HB, Matos M. 2004. Methuselah flies: a case study in the evolution of aging. Singapore: World Scientific Publishing.

Schlotterer C, Kofler R, Versace E, Tobler R, Franseen SU. 2015. Combining experimental evolution with next-generation sequencing: a powerful tool to study adaptation from standing genetic variation. Heredity 114:331–440.

Shiotsugu, J., Leroi, A.M., Yashiro, H., Rose, M.R., Mueller, L.D., 1997. The symmetry of correlated selection responses in adaptive evolution: an experimental study using Drosophila. Evolution 51, 163–172.

Somero, G.N., Yancey, P.H., 1997. Osmolytes and cell-volume regu- lation: physiological and evolutionary principles. In: Hoffman, F.F., Jamieson, J.D. (Eds.), Handbook of Physiology. Oxford University Press.

Stern DL. 2013. The genetic causes of convergent evolution. Nat Rev: Genet. 14:751–764.

Computing, Vienna, Austria. URL https://www.R-project.org/. Storey, J.D. 2002. A direct approach to false discovery rates. J. Roy. Stat. Soc. Series B. 64: 479-498.

Tenaillon O, Rodriguez-Verdugo A, Gaut RL, McDonald P, Bennett AF, Long AD, Gaut BS. 2012. The molecular diversity of adaptive convergence. Science 335:457–461.

Turner TL, Steward AD, Fields AT, Rice WR, Tarone AM. 2011. Population-based resequencing of experimentally evolved populations reveals the genetic basis of body size variation in Drosophila melanogaster. PLoS Genet. 7:e10001336.

Wiberg, R.A.W., O. E. Gaggiotti, M.B. Morrissey, and M.G. Ritchie. 2017. Identifying consistent allele frequency differences in studies of stratified populations. Methods Ecol. Evol. 8: 1899-1909

Wright, S. 1982. The shifting balance theory of macroevolution. Annu. Rev. Genet. 16: 1-19.

Yang X., J. Li, Y. Lee, and Y.A. Lussier. 2011. GO-Module: functional synthesis and improved interpretation of gene ontology patterns. Bioinformatics 27:1444–1446.

Yancey, P . H. (1985). Organic osmotic effectors in cartilaginous fishes. In Transport Processes, Iono- and Osmoregulation (ed. R. Gilles and M. Gilles-Ballien), pp. 424–436. Berlin: Springer-Verlag.

Yancey, P. H. (1992). Compatible and counteracting aspects of organic osmolytes in mammalian kidney cells in vivo and in vitro. In Water and Life: A Comparative Analysis of Water Relationships at the Organismic, Cellular, and Molecular Levels (ed. G. N. Somero, C. B. Osmond and C. L. Bolis), pp. 19–32. Berlin: Springer-Verlag.

Yancey, P . H. and Somero, G. N. (1979). Counteraction of urea destabilization of protein structure by methylamine osmoregulatory compounds of elasmobranch fishes. Biochem. J. 182, 317-323.

FIGURES

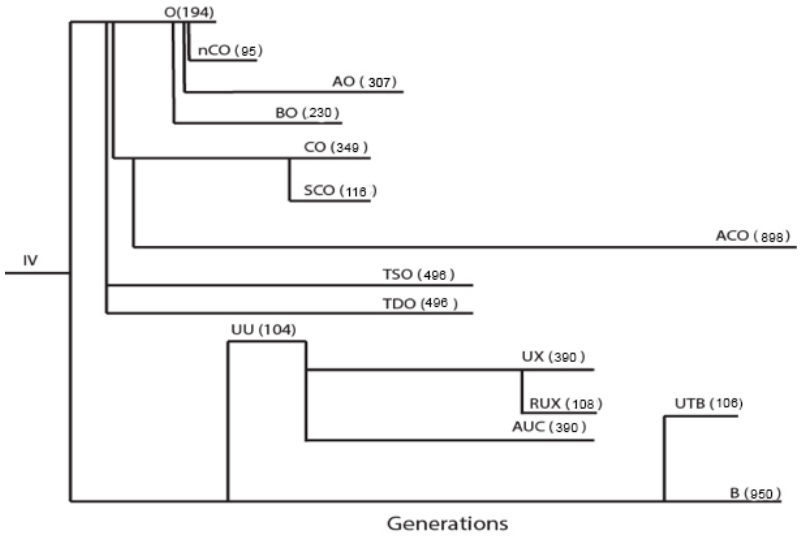

Figure 4.1 The phylogeny of all current populations in the Rose/Mueller lab. The number by each population is the total number of generations since their derivation and the time when the DNA was extracted for populations AUC, RUX, UX and UTB.

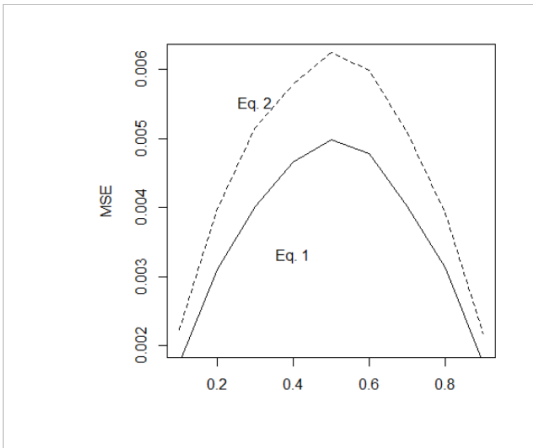

Figure 4.2 The MSE from Eq. 1 is lower at all SNP frequencies due mostly to its lower variance. Parameter values used were, N_1=360 (180 females), N_2=60 (30 females), n_1=n_2=34.

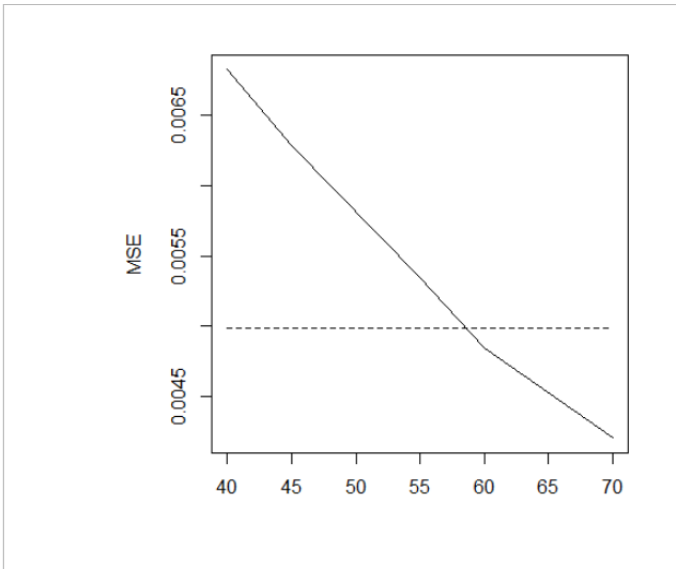

Figure 4.3 MSE of SNP frequency estimates as coverage decreases. Using an initial sample of 360 chromosomes (180 females), and a population SNP frequency of 0.5, the MSE of a SNP frequency estimate decreases as the coverage increases (solid line). The dashed line shows the MSE for the combined sample of 180 and 30 females each with a coverage of 34. The dashed line and the solid line cross at a coverage of about 59.

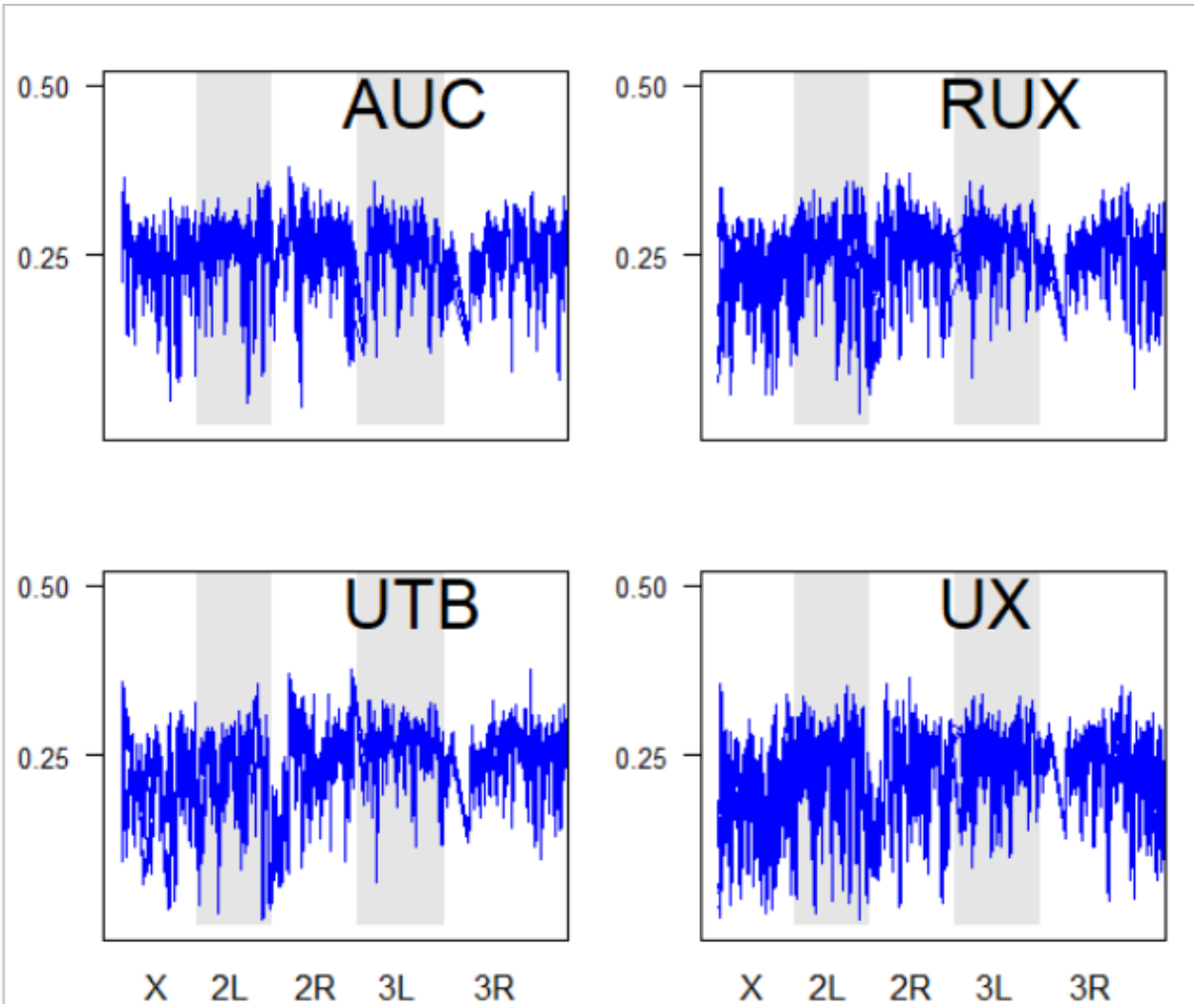

Figure 4.4 The average heterozygosity across all 5 replicates and two database averages.

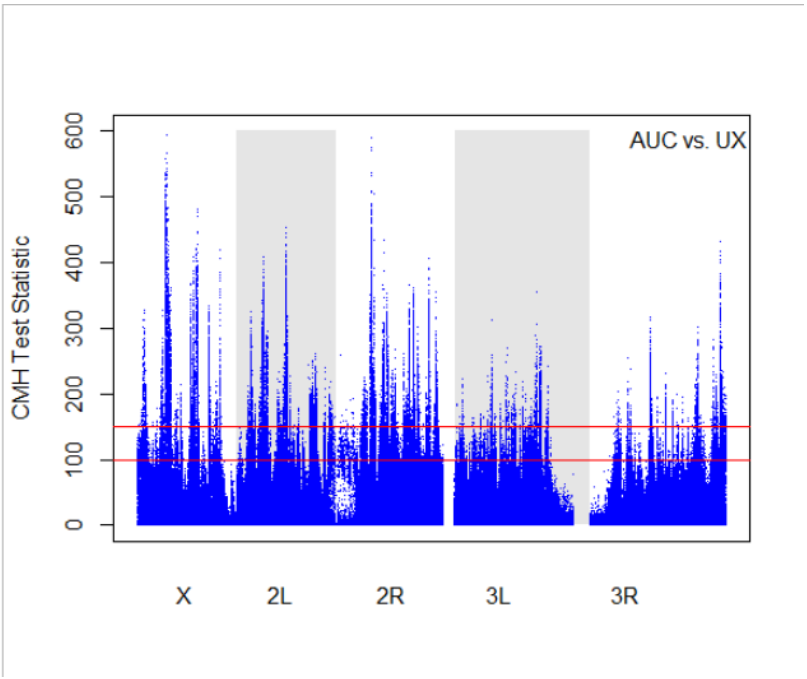

Figure 4.5 CMH test comparison of AUC and UX populations across both data sets from June and December 2019. The CMH test statistic is shown on the *y*-axis. Two thresholds for significance are shown as red lines at 50 and 100. The FDR for these two threshold are summarized in Table 4.1.

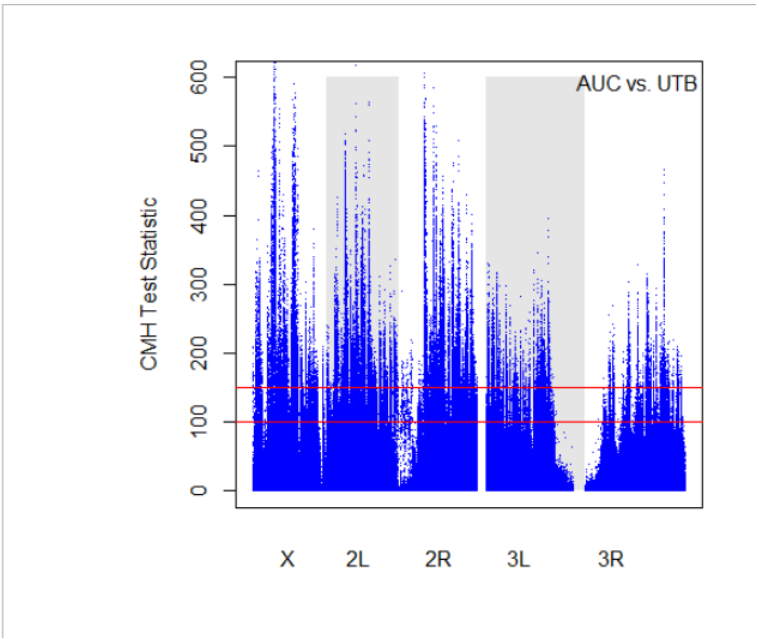

Figure 4.6 CMH test comparison of AUC and UTB populations across both data sets from June and December 2019. The CMH test statistic is shown on the *y*-axis. Two thresholds for significance are shown as red lines at 50 and 100. The FDR for these two threshold are summarized in Table 4.1.

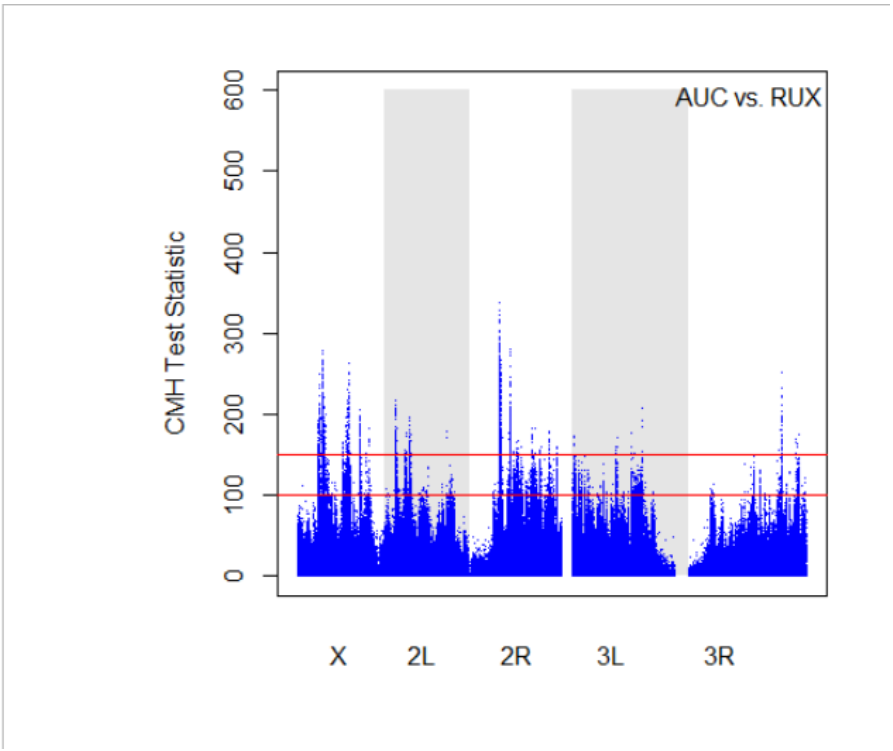

Figure 4.7 CMH test comparison of AUC and RUX populations across both data sets from June and December 2019. The CMH test statistic is shown on the *y*-axis. Two thresholds for significance are shown as red lines at 50 and 100. The FDR for these two threshold are summarized in Table 4.1.

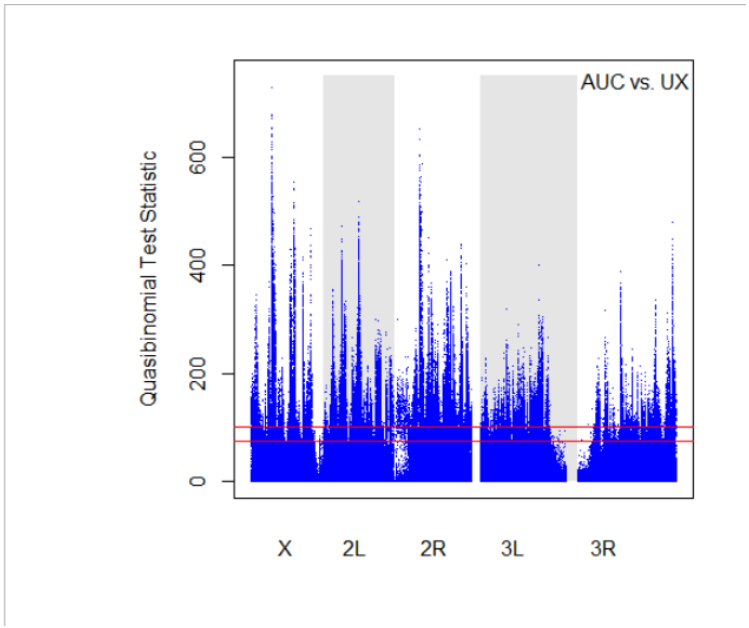

Figure 4.8 Quasibinomial test of AUC and UX populations across both data sets from June and December 2019. The Quasibinomial test statistic is shown on the *y*-axis. Two thresholds for significance are shown as red lines at 50 and 75. The FDR for these two threshold are summarized in Table 4.1

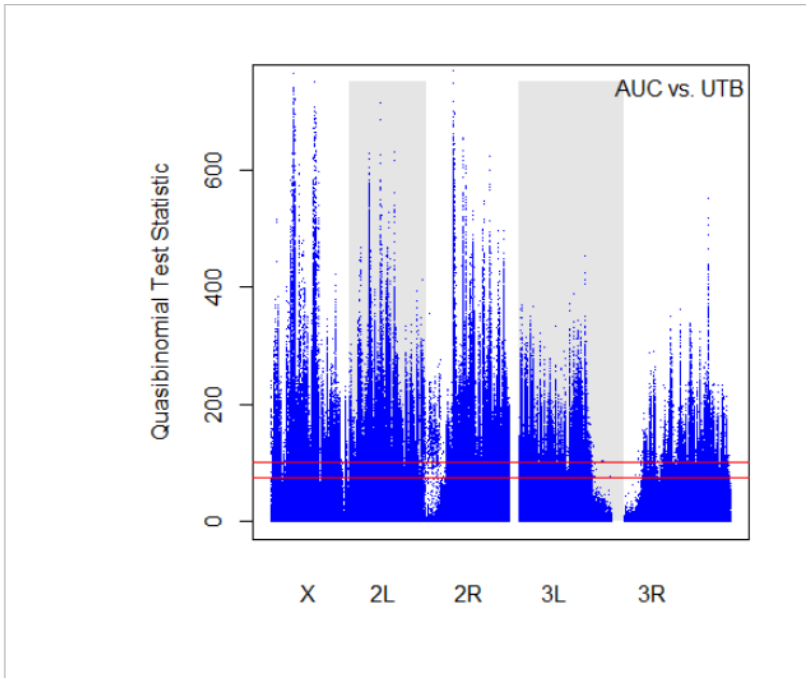

Figure 4.9 Quasibinomial test of AUC and UTB populations across both data sets from June and December 2019. The Quasibinomial test statistic is shown on the *y*-axis. Two thresholds for significance are shown as red lines at 50 and 75. The FDR for these two threshold are summarized in Table 4.1

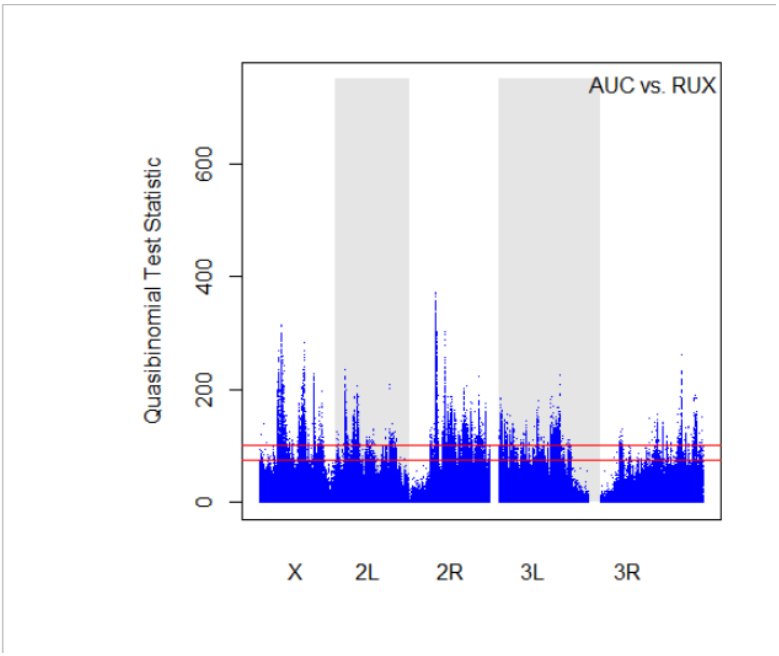

Figure 4.10 Quasibinomial test of AUC and RUX populations across both data sets from June and December 2019. The Quasibinomial test statistic is shown on the *y*-axis. Two thresholds for significance are shown as red lines at 50 and 75. The FDR for these two threshold are summarized in Table 4.1

TABLES

Table 4.1. Counts of significantly differentiated SNPs from the CMH and Quasibinomial tests from both databases in June and December 2019. Associated with each significance level is a false discovery rate (FDR) estimated from the plug-in method. The total number of polymorphic SNP's examined 1,389,302

	CMH Test			Quasibinomial Test		
Comparison	50	100	150	50	75	100
AUC v UX						
# significant tests	197,610	60,382	21296	218,415	121,455	70980
FDR	0.169	0.0788	0.052	0.0347	0.0131	0.0063
AUC v UTB						
# significant tests	275937	110340	110340	306643	194437	128913
FDR	0.11	0.047	0.030	0.021	0.0057	0.0020
AUC v RUX						
# significant tests	58599	8054	1654	70157	25989	10463
FDR	0.093	.030	0.016	0.086	0.039	0.021
UX v UTB						
# significant tests	167586	39925	9938	191140	93850	48383
FDR	0.21	0.10	0.065	0.030	0.0098	0.0041
UX v RUX						
# significant tests	31779	4594	1085	34532	12190	5095
FDR	0.60	0.38	0.23	0.22	0.12	0.074
UTB v RUX						
# significant tests	157656	38736	10311	178862	91657	48297
FDR	0.14	0.059	0.034	0.031	0.0093	0.0037

Table 4.2. All four populations of *D. melanogaster* used in the genomic analysis. The number of days spent in vials, number of days spent in cages, their generation time and if they lived in a larval environment with urea is specified per population.

Population	Days in Vials	Days in cages before egg collect	Generation Time	Urea in larval environment
AUC	14	7	21	No
RUX	14	7	21	Previously yes
UX	14	7	21	Yes
UTB	14	7	21	Yes

Table 4.3 F_{st} values from our analysis of the two databases compared to other Rose lab populations analyzed (* Graves *et al.* 2017, # Phillips *et al.* 2018)

Population	F_{ST} *(and potential confidence interval)*
AUC 1-5	0.04268 ± 0.000056
RUX 1-5	0.06120, +- 0.000085
UX 1-5	0.09951, +- 0.00015
UTB 1-5	0.06059, +- 0.00012
ACO 1-5 *	0.062
AO 1-5 *	0.087
B 1-5 *	0.058
BO 1-5 *	0.041
CO 1-5 *	0.028
TSO #	0.04
TDO #	0.07

Table 4.4 The average heterozygosity of the four population across both databases – AUC, RUX, UTB and UX.

AUC	0.2565553
RUX	0.2576476
UTB	0.2410314
UX	0.2275179

Table 4.5. The number of significantly differentiated SNPs shared between selection regimes. Differentiation was based on a quasi-binomial test statistic greater than 100 when the AUC populations were compared to a specific urea selection regime.

Contrasts	Shared SNPs	Maximum Possible Shared
RUX and UTB	7,380	10,463
RUX and UX	8,346	10,463
UX and UTB	36,080	70,980
UX, UTB, and RUX	6,406	7,380

CHAPTER 5

Using FLAM Analysis to distinguish between
differentiated genes of Urea Adapted populations

Introduction

Understanding the relationship between genes and phenotypes is the base of much genetic and experimental evolutionary work. Lewontin's (1974) goal was understanding this relationship, which now may be possible with genome-wide sequencing in the field of evolutionary biology. Understanding adaptation can be studied through three different genomic approaches: 1) QTL mapping; 2) population genetics; and 3) experimental evolution. Experimental evolution can produce extreme phenotypic differentiation among traits in replicated populations (Garland and Rose 2009). This method allows for controlling various aspects of an organism's environment, as well as its population history.

Model organisms that have undergone sustained experimental evolution can show high levels of phenotypic differentiation and can be genomically characterized by DNA sequencing technology. With genome-wide sequencing and Pool-Seq, we can trace trajectories. Experimental evolution has been successfully used in microorganisms, such as Escherichia coli (Elena & Lenski 2003) and yeast (Zeyl 2006; Parts *et al.* 2011), and also in multicellular organisms, including *Drosophila* (Hoffmann *et al.* 2003; Turner *et al.* 2011; Zhou *et al.* 2011).

Combining experimental evolution in *D. melanogaster* with genome-wide next generation sequencing of DNA pools (Pool-Seq) can help to identify SNPs or alleles that play a role in the adaptive process. Sequencing of DNA from pooled individuals (Pool-Seq) provides an excellent tool to determine allele frequencies on a genomic scale (Futschik & Schlotterer 2010; Kolaczkowski *et al.* 2011, Wengel *et al.* 2012)

We studied phenotypic differentiation (chapter 3) and genomic differentiation (chapter 4) among populations subjected to different types of selection for tolerance to urea in their larval food. The five AUC populations served as controls and were fed normal food and kept on a

three-week generation cycle like the other populations. The UX and UTB populations were raised on larval food containing urea up to the time that genomic samples were collected. The RUX populations were derived from the UX populations after 282 generation of selection for urea tolerance and were placed back on standard food for 108 generations prior to the genomic samples. We saw parallel evolution occurring in the replicates of the same environmental populations, as well as genomic differentiation in the Urea adapted lines, UX, UTB and RUX versus the control population of AUC. Furthermore, in chapter 4, we compared the Urea lines (UX, UTB, RUX and control AUC) to the demographic lines; CO, nCO, TSO, and TDO. Our next step is to look at polymorphic SNP differentiation using FLAM.

The 'fused lasso additive model' (FLAM) technique is a discovery tool for determining which genes may affect differentiated phenotypes. Evidence from computer simulations has shown that this statistical model can effectively sort out those loci that are differentiated and have a causal effect on a phenotype versus those that are differentiated but do not have a causal effect (Mueller *et al.* 2018). It minimizes cross-validation error to find a subset of genomics SNPs that can be used to predict phenotypes. Other experiments have suggested that the genomic response to selection can involve many selected SNPs that show unexpectedly complex evolutionary trajectories, possibly due to nonadditive effects. (Wengel *et al*. 2012). The efficacy of FLAM is improved with increased number of independent populations, reduced environmental phenotypic variation, and increased within-treatment among-replicate variation (Mueller *et al.* 2018).

Our focus with FLAM was to distinguish significant SNP's that influence the phenotype, a process that FLAM can do well with even small samples (Mueller 2018). FLAM was applied to SNP variation measured in 40 populations of *D. melanogaster* – 20 of which are the urea

adapted selection regimes, UX, UTB, RUX and AUC, and 20 that make up the demographic lines subjected to selection for age-at reproduction or starvation/ desiccation resistance – CO, nCO, TSO, and TDO.

MATERIALS AND METHODS

1.7 Populations

All the populations in the lab are derived from an established Rose IV population (Rose, 1984), Figure 3.1. They are maintained on a banana-molasses food (Rose, 1984) at 25C (24 h light), uncontrolled humidity, and having a generation time of approximately 3 - 4 weeks (Table 5.1). All selection regimes are five-fold replicated, uncrowded as larva (60–80 eggs per 8-dram vial), with emergent adults kept at a low density of approximately 50–60 flies per 8-dram vial, and transferred to a cage environment with fresh food given about every other day for approximately 1 week. Effective populations sizes for each line were in the range of 700-1000 (Mueller *et al.*, 2013) every generation and maintained at large population sizes (1000), with discrete generations. The CO and nCO are control populations that are on a 28-day life cycle. The TSO population is a starvation selected population, while the TDO is a starvation and desiccation selected population, both on 28-day life cycles. Both the TSO and TDO have now been in a control environment for hundreds of generations.

The UX and the UTB populations were subjected to selection for increased larval tolerance to the presence of toxic levels of urea in the food. The levels of urea were increased every few generations, when it was observed that a great proportion of larvae were surviving to adulthood. The derivation of the urea-tolerant (UX) and unselected controls (AUC) selection regimes was done(?) in the Fall of 1996. Both populations were derived from a five-fold replicated set of populations called UU, which had a 3-week generation time, and were reared at low larval and adult densities. The UU populations were derived in 1990, from the Rose B populations (Rose, 1984; Chippindale *et al.*, 1994, 1996). The RUX are a reverse selected line of

the UX. They have been in a control environment for 108 generations. The UTB line were created in October 2013. The AUC, RUX, UX, and UTB lines are all 21-day cycle flies.

For every assay, eggs were collected from the 8 selection regime types and passed through two generations of common, standard conditions – low larval density, 1000 adult density, discrete generation times and regular banana-molasses food (Bitner *et al.* 2020).

1.2 Phenotypes

Four different phenotypes were compared between all 8 sets of selection regimes: viability, growth rate, feeding rate and developmental time. Feeding rate is the rate at how often the sclerite retractions occur in *D. melanogaster* larvae as they are observed over a standard period of time, 60 seconds. To measure the feeding rate, individual larvae around 48 hours old were gently moved onto a 3% agar coated with a 10% live yeast suspension. The larvae were given 60 seconds to adjust to the new surroundings, and their sclerite retractions were recorded for 60 seconds and counted for twenty larvae per population. The procedure for measuring sclerite rates is similar to Sewell's *et al.* (1975) procedure and described in Joshi *et al.* (1988). Feeding rates were collected on all 8 selection regimes– AUC, RUX, UX, UTB, CO, NCO, TSO and TDO.

Viability looked at how many larvae survived to adulthood in a certain environment. Larvae were raised under two experimental treatments, a control environment with regular food and food with added urea. Ten vials were used per population per environment, for a total of 20 vials per population. Each vial held the approximate 50 larvae. Ultimately, we are interested in the testing the effects of urea on survival for each population as well as differences between the seven selection regimes, AUC, RUX, UX, UTB, CO, NCO and TSO. Developmental time focuses looking at differences in the development time of the larvae. The measurements were

made from when first instar larvae were collected to when the fly eclosed in the control banana molasses environment compared to the urea environment. For each population, 10 vials were set up with 50 freshly hatched larvae each in a banana molasses environment. Another 10 vials were set up with urea food and 50 freshly hatched larvae. Eclosing adults were collected every 6 hours separated by male and female and recorded. Development time was collected on the AUC, RUX, UX, UTB, CO, NCO and TSO populations.

The larval growth rate assay focused on measuring the growth of larvae until pupation. 45 newly hatched first instar larvae were collected with a fine paint brush and placed onto non-nutritive agar petri dishes with 3 ml of yeast paste (188 grams of yeast in 500 ml of DI water) and placed randomly into a 25 C incubator with 24-hour lighting. There were 13 different "hour numbers" larvae were sampled at: 24, 30, 36, 42, 48, 54, 60, 66, 72, 78, 84, 90, and 105 hours after the larvae were added to the petri dish. At the designated hour, the larvae were washed with DI water and then allowed to air dry. The wet weight was taken of the larvae and they were then placed into an 80 degree C drying oven, after which their dry weights were recorded. A logistic growth model was estimated from these data. The parameter which describes the age at which a larva reaches 50% of its maximal size was used in the FLAM analysis.

1.3 FLAM

We assume that SNP frequencies across *m* loci has been measured in *n* independent populations along with phenotypes for each population, $\mathbf{P}=(P_1, P_2,..,P_n)$. At locus-*j*, for instance, we assume that the allele frequencies can be ordered as, $\hat{p}_{1j} < \hat{p}_{2j} < \ldots < \hat{p}_{nj}$. The regression relationship, $E[P_i|\hat{p}_{ij}] = \theta_i$, of the fused lasso additive model (FLAM, Peterson et al., 2016) will yield estimates of the parameter vector, $\boldsymbol{\theta}_j=(\theta_{1j}, \ldots, \theta_{nj})$ subject to,

$$\underset{\boldsymbol{\theta}_j \in \mathbb{R}^n}{minimize} \frac{1}{2}\left\|\mathbf{P} - \boldsymbol{\theta}_j\right\|_2^2 + \lambda\left\|\boldsymbol{D\theta}_j\right\|_1 \quad (1)$$

where

$$\mathbf{D}=\begin{pmatrix} 1 & -1 & 0 & \ldots..0 \\ 0 & 1 & -1 & \ldots..0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0\ldots. & 1 & -1 \end{pmatrix}$$ (Peterson *et al.* 2016).

Large values of the tuning parameter λ will tend to make $|\theta_{i\text{-}1,j} - \theta_{i,j}|$ equal to zero. Hence, the final function will be a series of steps with jumps or knots that are adaptively chosen.

Over all loci, we add to the optimization problem in equation (1) a group lasso penalty function that will encourage whole $\boldsymbol{\theta}_j$ vectors to be zero and thus serve to eliminate uninformative loci yielding,

$$\underset{\theta_0 \in \mathbb{R}, \boldsymbol{\theta}_j \in \mathbb{R}^n, 1 \leqslant j \leqslant m}{minimize} \frac{1}{2}\left\|\mathbf{P} - \sum_{j=1}^m \boldsymbol{\theta}_j - \theta_0 1\right\|_2^2 + \alpha\lambda \sum_{j=1}^m \left\|\boldsymbol{DM}_j\boldsymbol{\theta}_j\right\|_1 + (1 - \alpha)\lambda \sum_{j=1}^m \left\|\boldsymbol{\theta}_j\right\|_2, \quad (2)$$

where $\mathbf{M}_j$ is a matrix that orders the values of $\widehat{\boldsymbol{p}}_j$ from smallest to largest. Equation (2) adds a second tuning parameter α, which ranges from 0 to 1 (Peterson et al. 2016). The R-function, *flamCV* (in the *flam* package), will search for the best λ based on the cross-validation error rates for a given value of α. In our analyses, we used a grid of 19 α-values to find the best model (α= 0.05, 0.1, …,0.95).

The solutions to (2) are characterized by a sparse set of loci where $\left\|\boldsymbol{\theta}_j\right\|_2 \neq 0$. Although there is a global minimum for the objective function (2), there is not a unique solution. One method for finding this minimum is called block coordinate descent ("BCD," Friedman *et al.* 2007). While it is fast and robust compared to other methods, the sparse set of loci that are identified by BCD depend on the arrangements of loci in the matrix of independent variables.

We identified causative loci with the following algorithm. Find a solution to (2) using BCD and save the sparse set. Then permute the columns of the matrix of independent variables and solve (2) again. Repeat the permutation and solution steps 100 times. Enumerate the frequency of occurrence of each SNP among these 100 sparse sets. Let the frequency of the most common SNP among the 100 set be C_{max}. Identify as the causative loci only those that occur greater than $C_{max}/2$. The utility of FLAM and the modifications described here for finding SNP's located in causative genes has been explored by computer simulations (Mueller *et al.*, 2018).

1.4 Identification of Differentiated SNPs due to Urea Adaptation

To identify SNPs that will be used in FLAM for phenotype-genome associations we used the only SNPs that exceed the critical value of 100 on the quasibinomial test. With this criteria the false discovery rate (FDR) was less than, often much less than, 0.07 in all pairwise comparisons. We required that a differentiated SNP show significant differentiation in both a comparison of AUC vs. UX and AUC vs. UTB. UX and UTB comprise the populations that were still under selection for larval urea resistance at the time of the genomic assays and thus the most important for identifying the loci involved in urea adaptation. However, the phenotypes that were tested show variation in many other populations so when using FLAM we included all 35 or 40 populations that have been assayed for the important phenotypes related to urea adaptation, feeding rates, viability in urea, development time in urea, and larval growth rates.

The ability of FLAM to identifying causal loci will be hindered if too many SNPs from a small region of the genome in linkage disequilibrium are used. Thus, following a procedure used in Mueller *et al.* (2018) we utilized only the most differentiated SNP in a region of 50kb. To do this we searched each chromosome arm until a SNP that satisfied the criteria for differentiation was found. Then we continued to search over the next 50kb identifying the most differentiated

SNP in that region and saving its location. This procedure was applied to the SNPs identified as differentiated in the AUC vs UX comparison and the AUC vs UTB comparison. Using the shorter AUC vs UX list we identified SNPs in the AUC vs UTB list that were within 25kb of the AUC vs UX SNPs. From this reduced list of pairs of close SNPs in the AUC vs UX and AUC vs UTB lists we chose the single SNP from each pair which had the highest test statistic.

We independently produced a SNP table for the TSO, TDO, CO and nCO populations. Genomic data was extracted at various time points of these populations. Libraries were run across PE100 lanes of an Illumina HiSEQ 2000 at the UCI Genomics High throughput Sequencing Facility and constructed such that each five replicate populations of a treatment (e.g., UX1–5) were given unique barcodes, normalized, and pooled together (Graves *et al.* 2017). The reads were trimmed to remove low-quality bases using a script provided in the PoPoolation software package (Kofler, Orozcoter Wengel, *et al.* 2011). The reads were mapped with BWA (Li and Durbin 2009) against the *D. melanogaster* reference genome (6.31). The SAM files were converted to BAM and potential PCR duplicates were removed and the BAM files combined. Using PoPoolation2 (Kofler, Pandy *et al.* 2011), the resulting mpileup was converted to "synchronized" files, which is a format that allele counts for all bases in the reference genome and for all populations being analyzed. The SNP's were compiled and a SNP table with major and minor allele counts for each SNP in each population was then generated. From this table and the previous differentiated list of SNPs a total of 673 SNPs were used in the FLAM analysis.

To summarize the precision of the FLAM predictions we estimated the correlation between predictions and observations as follows. Depending on the phenotype we had either 7 or 8 different selection regimes, each replicated five times. We deleted one population from each selection regime forming two data sets: a test data set of 7 or 8 observations and a training data

set of 28 or 32 observations. FLAM parameters were then estimated using the training data set and the sparse list of SNPs. From this fitted FLAM model, we then predicted the 7 or 8 observations that had not been used to fit the FLAM parameters. This process was repeated five time so that all observations were eventually predicted from a training set of observations. The Pearson correlation coefficient was then estimated from the observed and predicted phenotypes.

SNP to Gene

We downloaded the FlyBase genome assembly of *D. melanogaster* release 6.31. Comparing the locations of the SNP locations to the gene start and stop locations, we found all genes that were within the 50 kb window of the SNP location. We reproduced the annotations for the genes based off the information we found on FlyBase.

Variation at the foraging locus

One locus which we have an a-priori interest in is the foraging locus (*for*). This locus affects foraging behavior of *Drosophila* larvae (Sokolowski *et al.*, 1983, 1997). Previous work with the AUC and UX populations has shown a decline in the foraging path length – the primary phenotype *for* is believed to affect (Mueller *et al.*, 2005). Foraging path length and feeding rates tend to change in a parallel fashion and thus the *for* locus may account for some of the phenotypic differentiation in feeding rates documented in chapter 3.

The *for* locus is located on chromosome 2L at positions 3,622,074 to 3,656,953. We identified 13 SNPs that were significantly differentiated between the AUC and UX populations and 11 SNPs that were differentiated between the AUC and UTB populations. The differentiation was assessed using SNPs with a quasi-binomial test statistic greater than 75. The false discovery rate using this criterion is less than 0.013. Since these were non-overlapping sets all 24 SNP's were used in a principal component analysis. Allele frequencies at the 24 SNPs

were transformed by taking the arcsin, square root of the allele frequencies and then centering and scaling these values. Principal components were estimated using the R *prcomp* function (R Core team, 2019) on all 20 urea populations and all 20 demographic populations. The first three principal components, which explain 76% of the total variance, were used to generate a hierarchical cluster of all 40 populations using the R *hclust* function.

RESULTS

A sparse list of the SNP's was generated with the cytogenic location for the four phenotypes (Table 5.1). We enumerated for each SNP all the genes that occur +/- 25kb of the SNP picked out by FLAM (Table 5.2) with the FLAM associated attributions, according to FlyBase. Pooling samples from two different DNA sequencing events resulted in the urea lines having over 1,000,000 SNPs, and the demographic lines also had over 1,000,000.

The FLAM step function predicted feeding rates as a function of SNP frequency at 22 genomic locations identified by FLAM (Figure 5.1). Each figure (Figures 5.1, 5.3, 5.5 and 5.7) is a compilation between 7-22 smaller figures. Each one of these smaller figures contains an integer which equals the number of times that SNP was included in the 100 permutations of the genetic matrix. The frequency of a SNP appearing is a suggestion of its relative importance. Our cut off rate was at 50%. Using lower values of maybe 30 or even 25% might include SNP's that are more likely to be false positives. In some of the figures the curves change dramatically, furthering the assumption they are the more important SNP's. In figure 3, viability rates were predicated as a function of SNP frequency at 13 genomic locations identified by FLAM. In figure 5 the FLAM step function predicted the larval growth rates as a function of SNP frequency at 7 genomic locations identified by FLAM. Finally, in figure 7, the FLAM step function predicted the development time as a function of SNP frequency at 14 genomic locations identified by FLAM.

The only way FLAM can distinguish causal loci from noncausal differentiated loci is if there is between replicate genetic variation within well-differentiated groups of populations that leads to small differences in average phenotypes among the replicate populations of such groups (Mueller *et al.* 2018).

In additional to deciding which SNPs are informative, it arranges the allele frequencies from smallest to large and tries to fit a step function to the phenotype as a function of SNP frequency. Using FLAM, we can look at correlations between phenotypes and the predictions from FLAM (Figures 5.2, 5.4, 6, and 5.8). In each of these four graphs, we have a correlation coefficient with 95% confidence intervals. The correlation between the predicted and observed phenotypes was 0.96 (Figure 5.4) for viability and 0.95 for development time (Figure 5.8). Feeding rates resulted in a lower correlation coefficient 0.89 (Figure 5.2). Larval growth rate parameters had the lowest correlation at 0.64 (Figure 5.6).

For the two highest correlation coefficients, viability and developmental time, they shared 3 of the same SNPS - 3L_11894035 (CEN 68E), 3L_15323813 (CEN 71C), and 3L_16272757 (CEN 72D), found in Table 5.2. SNP 3L_11894035 could be associated with the following genes: *CG44837* (protein-coding gene), *CG5906* (protein-coding gene), *Sprn* (which encodes a testis-specific mitochondrial lumen protein that first appears during the spermatocyte stage of spermatogenesis, but is absent from mature sperm), *CCDC151* (protein-coding gene for the coiled-coil domain), *CG5897* (protein-coding gene for Bestrophin 4) and *CR43624* (a long non-coding RNA:CR43624). SNP 3L_15323813 could be associated with *Best4* (Bestrophin 3 which encodes a paralog of the produce of Best1, a CA-activated *CL* channel), *CG7255* (protein-coding gene), *CR46213* (long non-coding RNA:CR46213), *CR43247* (long non-coding RNA:CR43247), *CR43992* (RNA:CR43992), *Toll-6* (Toll-6 encodes a member of the Toll-like receptor family that has neurotrophin receptor activity and contributes to dendrite guidance) and *CG33259* (protein-coding gene). SNP 3L_16272757 might have functional attributes of *CG13055, CG13054, CG13053, CG34247, CG13071, CG13070, CG13051, CG13069, CG4950, CG13068, CG34248, CG13067, CG13066, CG13065, CG13050, CG13064, CG13049,*

CG13048, CG13047, CG13046, CG13045, CG4962, all of which are protein-coding genes (Table 5.2).

The cluster analysis successfully grouped the UX and RUX populations (Figure 5.9) as well as the AUC populations. However, the UTB populations are not grouped together or consistently with the UX and RUX populations (Figure 5.9).

DISCUSSION

FLAM functions in finding a subset of genomic SNP's that are important and using them to predict phenotypes. FLAM can only distinguish between causal and noncausal differentiated loci when there is genetic variation between the replicates of the selection regimes in differentiated populations. Even when all conditions are favorable, an increased number of independent populations, reduced environmental phenotypic variation, and increased within-treatment among-replicate variation (Mueller *et al.* 2018), FLAM will not identify all causative loci. Mueller *et al.* 2018 showed that FLAM does an exceptionally good job of predicting phenotypes. Ultimately, when we predict the phenotype based of the genetic data, like we do in the correlation figures, we are using the allele frequencies from each population in combination with all the important SNPs, and getting a predicted phenotype from it. With the use of a minimal cross-validation error, that subset of genomic significant SNPs can be used to predict phenotypes.

Two of the four phenotypes compared had high correlation coefficients, signifying a significant correlation between the SNP's discovered, and their role in the phenotypic observances, which can be seen in Chapter 3. Development time and viability both had correlation coefficients above 0.9. Feeding rate had a high coefficient of 0.89, with the lowest correlation coefficient of 0.64 for the larval growth rate. When we looked at the growth rate parameter among the 40 populations (Chapter 3), there were no significant differences. There is not a strong differentiation of the growth rate parameter among the 8 selection regimes according to FLAM (Figure 5.6). That there is a weak correlation since differences among the 40 populations are not tremendously large is not surprising. Chapter 3 presented us with a correlated relationship between feeding rate and growth rate; a slower feeding rate resulted in a

slower growth rate. How fast a larva eats, and how quickly it grows is most likely influenced by a larger subset of genes, which might not be discovered by FLAM.

We created a list showing gene regions that are within a distance where there might be reasonable levels of linkage disequilibrium to each of these SNP. The results did not show extensive pleiotropy. The same genes were not showing up for the same traits, except for the urea-based traits – viability and development time. Three SNP's were found to be pleiotropic, affecting the viability and development time phenotype. These two phenotypes reflect the physiological impact of urea on either survival or development time also show overlapping genes. This is not unreasonable that these two phenotypes show pleiotropy as they are a different response to the same stress – urea toxin in the food and may be relying on similar physiological mechanisms. The other two phenotypes, larval feeding rate and larval growth rate are not necessarily a response to the same stress.

With the simulations in FLAM, we are picking up the genes that have the most direct impact on a phenotype. While we don't see extensive pleiotropy of genes, we are also dealing with a limitation of FLAM, and the number of populations we have, to only identify a handful of possible genes that have pretty large effects. We cannot detect all the loci with the FLAM analysis. It is possible that there might be pleiotropy that has not been uncovered with smaller effects. There might be a lot of genes that have small effects are going to be missed by this technique. As this technique gets modified and more advanced, or replaced by a different analytical technique, more genes may be discovered with attributions to the four physiological characteristics we studied – larval feeding rate, development time, viability and larval growth rate.

The analysis of genetic variation at the *for* locus showed that this variation was helpful in clustering the UX and RUX vs. the AUC populations. However, it did a poor job differentiating the UTB and all the demographic populations. This may suggest that the evolutionary trajectory that lead to the genetic differentiation of the UX populations (and hence their close relatives RUX) from the AUC populations was consistent among the UX populations but different from the route taken by the UTB populations.

Future studies would include further analysis and focusing on certain genes and their immediate effects. A transcriptomic analysis of the urea adapted lines could bring more information in further identifying genomic regions active in the adaptation of the urea adapted *D. melanogaster.*

REFERENCES

Bitner K, Shahrestani P, Pardue E, Mueller LD (2020) Predicting death by the loss of intestinal function. PLoS ONE 15(4): e0230970. https://doi. org/10.1371/journal.pone.0230970

Chippindale, A.K., Hoang, D.T., Service, P.M., Rose, M.R., 1994. The evolution of development in *Drosophila melanogaster* selected for postponed senescence. Evolution 48, 1880–1899.

Chippindale, A.K., Chu, J.F., Rose, M.R., 1996. Complex trade-offs and the evolution of starvation resistance in *Drosophila*. Evolution 50, 753–766.

Elena SF, Lenski RE. 2003. Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. Nature Reviews Genetics, 4, 457–469.

Futschik A, Schlotterer C. 2010. The next generation of molecular markers from massively parallel sequencing of pooled DNA samples. Genetics, 186, 207–218.

Garland T, Rose MR. 2009. Experimental evolution. University of California Press, Berkeley, CA

Graves, J.L, K.L. Hertweck, M.A. Phillips, M.V. Han, L.G. Cabral, T.T. Barter, L.F. Greer, M.K. Burke, L.D. Mueller, and M.R. Rose. 2017. Genomics of experimental evolution. Mol. Biol. Evol. 34(4): 831-842

Hoffmann AA, Sorensen JG, Loeschcke V. 2003. Adaptation of Drosophila to temperature extremes: bringing together quantitative and molecular approaches. Journal of Thermal Biology, 28, 175–216.

Joshi, A, L.D. Mueller. 1988. Evolution of Higher Feeding rate in Drosophila due to density-dependent natural selection. Evolution, 42(5): 1090-1093

Kofler R, Orozcoter Wengel P, De Maio N, Pandey RV, Nolte V, Futschik A, Kosiol C, Schlotterer C. 2011. PoPoolation: a toolbox for population genetic analysis of next generation sequencing data from pooled individuals. PLoS ONE 6:e15925.

Kofler R, Pandey RV, Schlotterer C. 2011. PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). Bioinformatics 27:3435–3436.

Kolaczkowski B, Kern AD, Holloway AK, Begun DJ. 2011. Genomic differentiation between temperate and tropical Australian populations of Drosophila melanogaster. Genetics, 187, 245–260.

Lewontin RC. 1974. The Genetic Basis of Evolutionary Change. New York: Columbia University Press.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760.

Mueller, L.D., D.G. Folk, N. Nguyen, P. Nguyen, P. Lam, M.R. Rose, and T. Bradley. 2005. Evolution of larval foraging behavior in Drosophila and its effects on growth and metabolic rate. Physiological Entomology 30: 262-269. (pdf file 124 kb)

Mueller, L.D., A. Joshi, M. Santos, and M.R. Rose. 2013. Effective population size and evolutionary dynamics in outbred laboratory populations of Drosophila. Journal of Genetics 92:349-361.

Mueller LD, Phillips MA, Barter TT, Greenspan ZS, Rose MR. 2018. Genome-wide mapping of gene phenotype relationships in experimentally evolved populations. Mol Biol Evol 35(8):2085–2095

Orozco-terWengel, P., M. Kapun, V. Nolte, R. Kofler, T. Flatt et al., 2012 Adaptation of Drosophila to a novel laboratory environment reveals temporally heterogeneous trajectories of selected alleles. Mol. Ecol. 21: 4931–4941. https://doi.org/10.1111/ j.1365-294X.2012.05673.x

Parts L, Cubillos FA, Warringer J, Jain K, Salinas F, Bumpstead SJ, Molin M, Zia A, Simpson JT, Quail MA, Moses A, Louis EJ, Durbin R, Liti G. 2011. Revealing the genetic structure of a trait by sequencing a population under selection. Genome Research, 21, 1131–1318.

Petersen A, Witten D, Simon N. 2016. Fused lasso additive model. J Comput Graph Stat 25(4):1005–1025

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rose, M.R., 1984. Laboratory evolution of postponed senescence in Drosophila melanogaster. Evolution 38, 1004–1010

Sewell, D., B. Burnet and K. Conolly. 1975. Genetic Analysis of larval feeding behavior in Drosophila melanogaster. Gen. Res. 24: 163-173

Sokolowski, M.B. and Hansell, R.I.C. 1983. Larval foraging behavior. The sibling species Drosophila melanogaster and D. simulans. Behavior Genetics 13: 159-168.

Sokolowski, M.B., Pereira, H.S., and Hughes, K. 1997. Evolution of foraging behavior in Drosophila by density dependent selection. PNAS 94: 7373-7377.

Turner TL, Stewart AD, Fields AT, Rice WR, Tarone AM. 2011. Population-based resequencing of experimentally evolved populations reveals the genetic basis of body size variation in

Drosophila melanogaster. PLoS Genetics, 7, e1001336.
Orozco-ter Wengel P, Kapun M, Nolte V, Kofler R, Flatt T, Schlotterer C. 2012. Adaptation of Drosophila to a novel laboratory environment reveals temporally heterogeneous trajectories of selected alleles. Molecular Ecology 21, 4931–4941

Zeyl C (2006) Experimental evolution with yeast. FEMS Yeast Research, 6, 685–691.
Zhou D, Udpa N, Gersten M et al. 2011. Experimental selection of hypoxia-tolerant Drosophila melanogaster. Proceedings of the National Academy of Sciences, 108, 2349–2354.

Zhou D, Udpa N, Gersten M et al. (2011) Experimental selection of hypoxia-tolerant Drosophila melanogaster. Proceedings of the National Academy of Sciences, 108, 2349–2354.

FIGURES

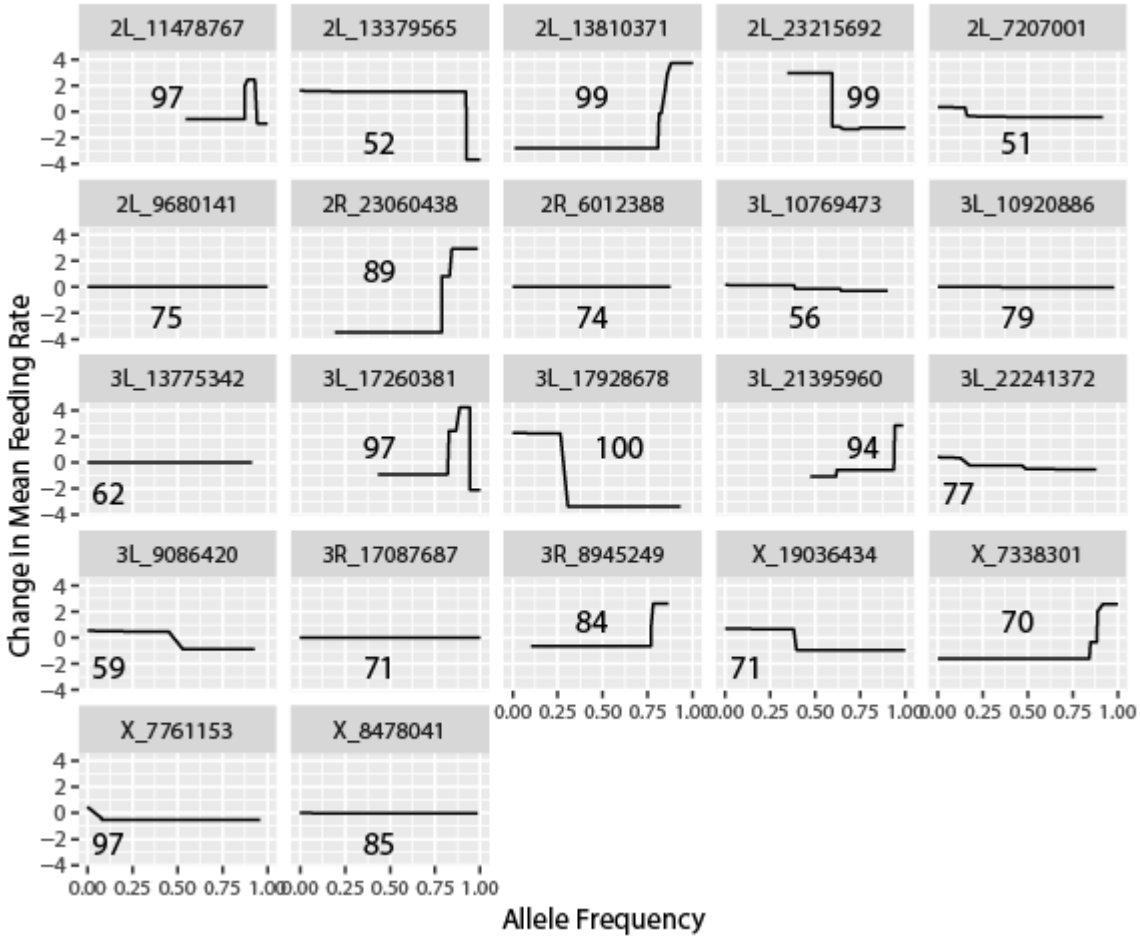

Figure 5.1 The FLAM step function, $E[P_i|\hat{p}_{ij}] = \omega_i$, predicting feeding rates as a function of SNP frequency at 22 genomic locations identified by FLAM. Each figure contains an integer which equals the number of times that SNP was included in the 100 permutations of the genetic matrix. The sparse set using the 50% criteria.

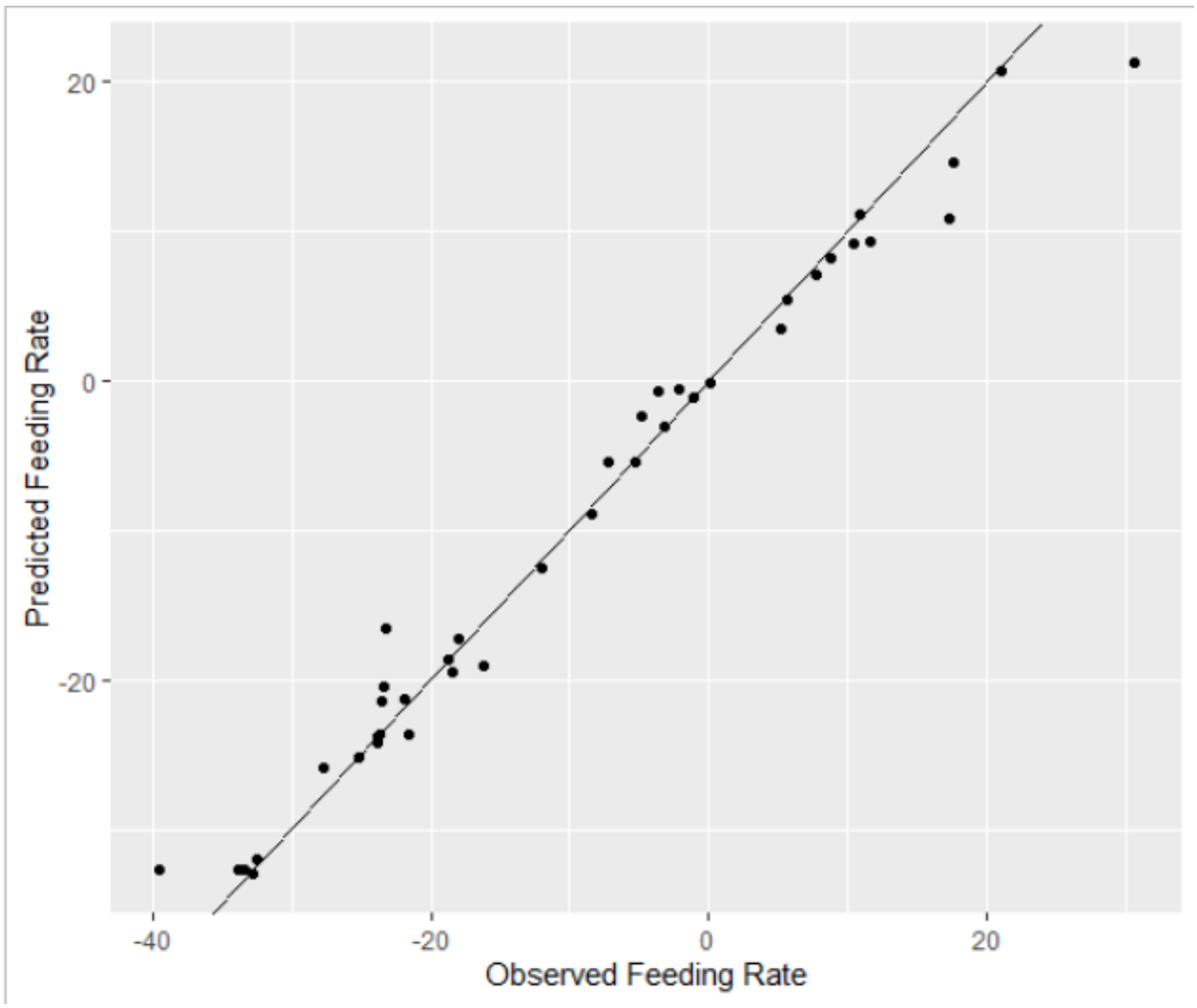

Figure 5.2 Predicted feeding rates in the original transformed units for each of the 40 populations. These predictions were based on FLAM fitted to the 22 SNPs shown in figure 5.1. The correlation between feeding rate predictions on test data and the observed feeding rates was 0.89 with 95% confidence intervals of (0.80 0.94)

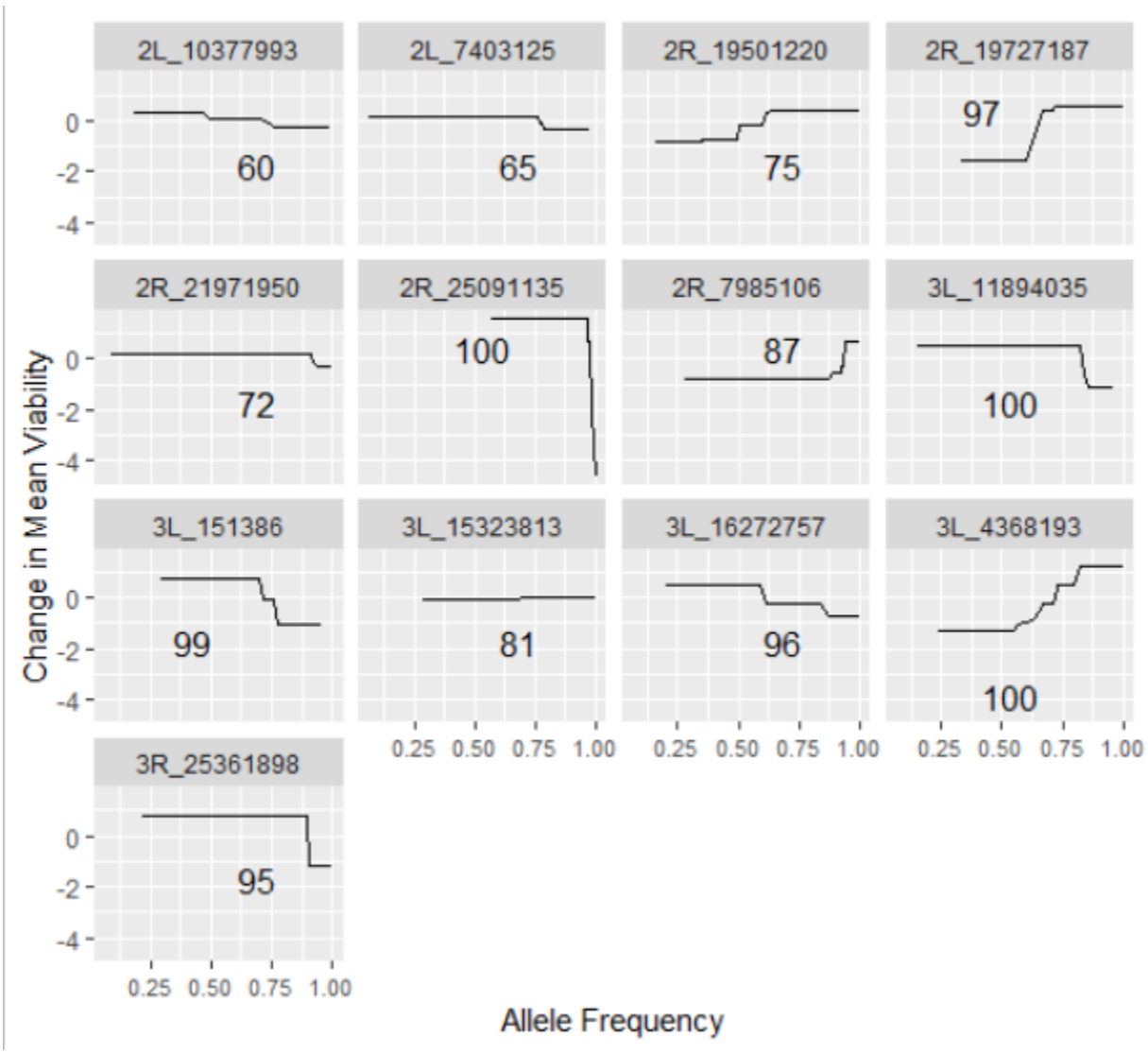

Figure 5.3 The FLAM step function, $E[P_i|\hat{p}_{ij}] = \Omega_i$, predicting viability rates as a function of SNP frequency at 13 genomic locations identified by FLAM. Each figure contains an integer which equals the number of times that SNP was included in the 100 permutations of the genetic matrix. The sparse set using the 50% criteria.

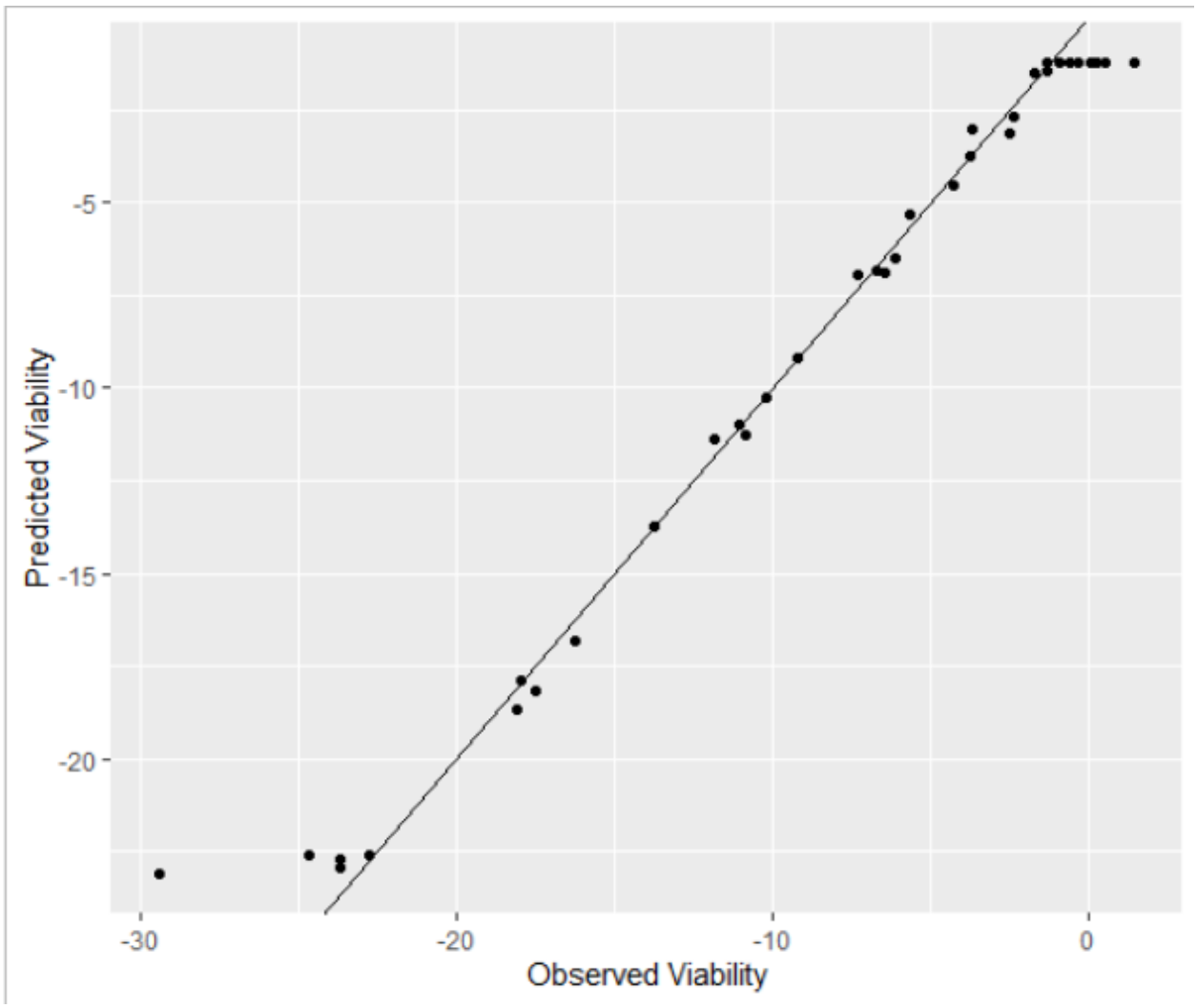

Figure 5.4 Predicted viability rates in the original transformed units for each of the 40 populations. These predictions were based on FLAM fitted to the 13 SNPs shown in figure 5.3. The correlation between viability predictions on test data and the observed viability data was 0.96 with a 95% confidence interval of (0.92, 0.98).

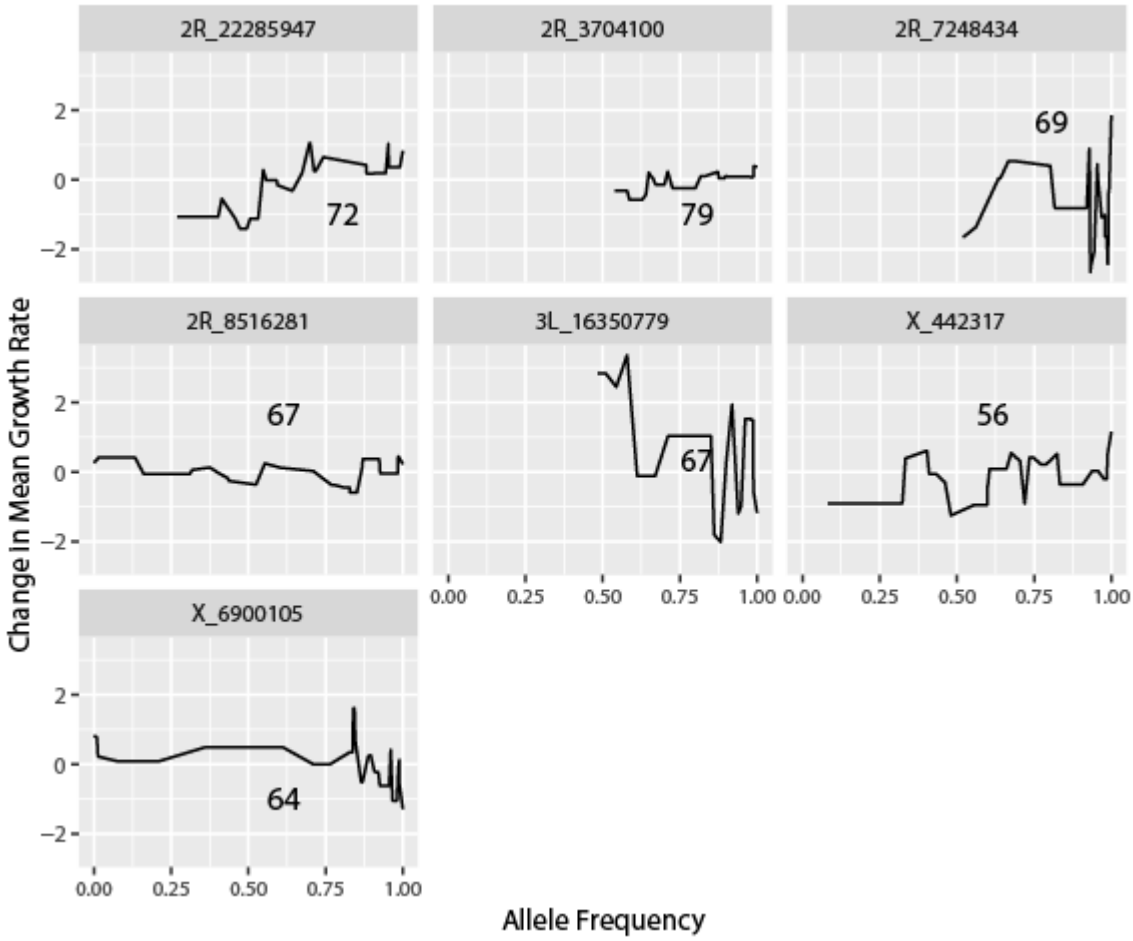

Figure 5.5 The FLAM step function, $E[P_i|\hat{p}_{ij}] = \omega_i$, predicting larval growth rates as a function of SNP frequency at 7 genomic locations identified by FLAM. Each figure contains an integer which equals the number of times that SNP was included in the 100 permutations of the genetic matrix. The sparse set using the 50% criteria.

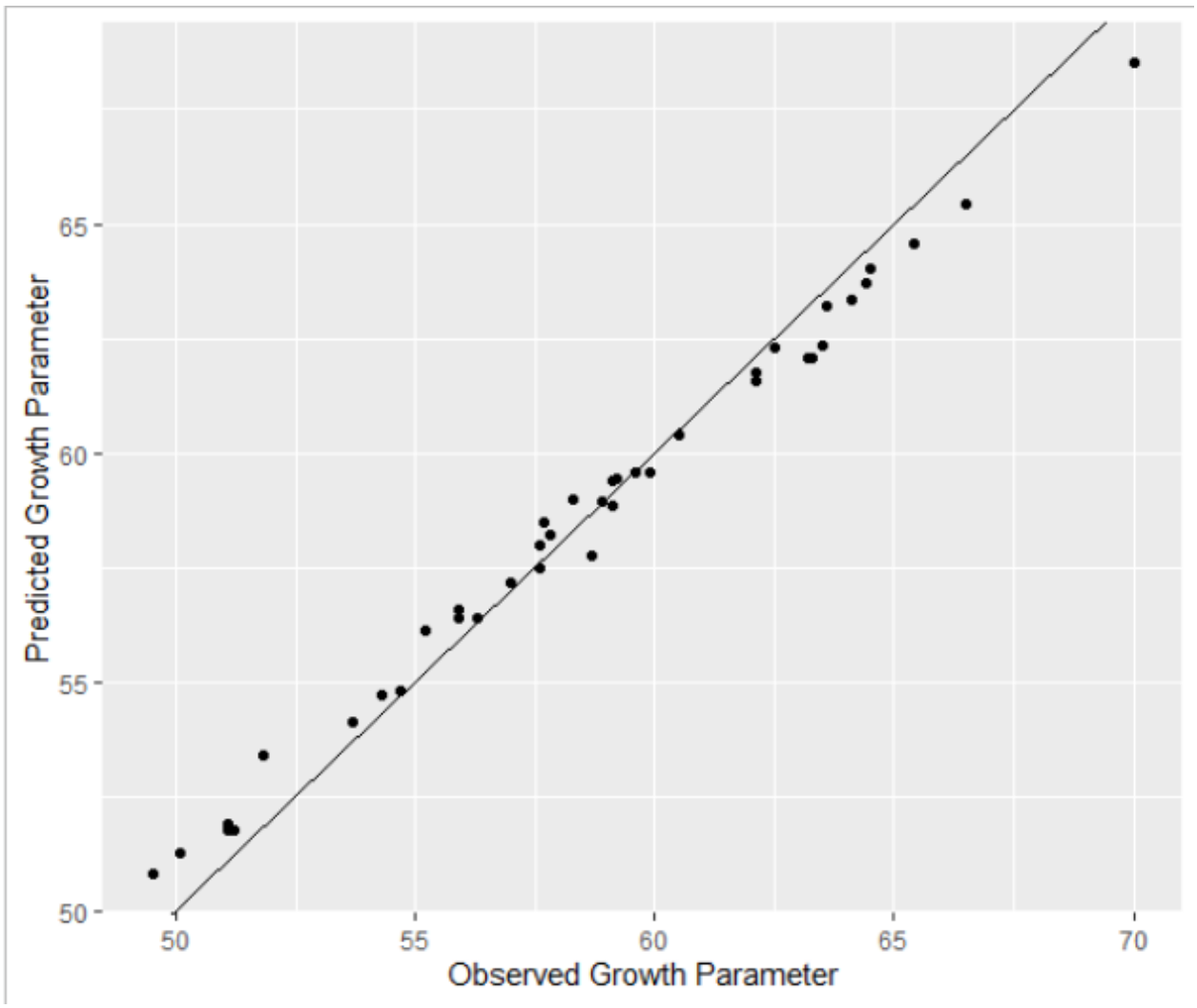

Figure 5.6 Predicted larval growth rates in the original transformed units for each of the 40 populations. These predictions were based on FLAM fitted to the 7 SNPs shown in figure 5.5. T he correlation between larval growth rate predictions on test data and the observed larval growth rates was 0.64 with a 95% confidence interval (0.41, 0.79)

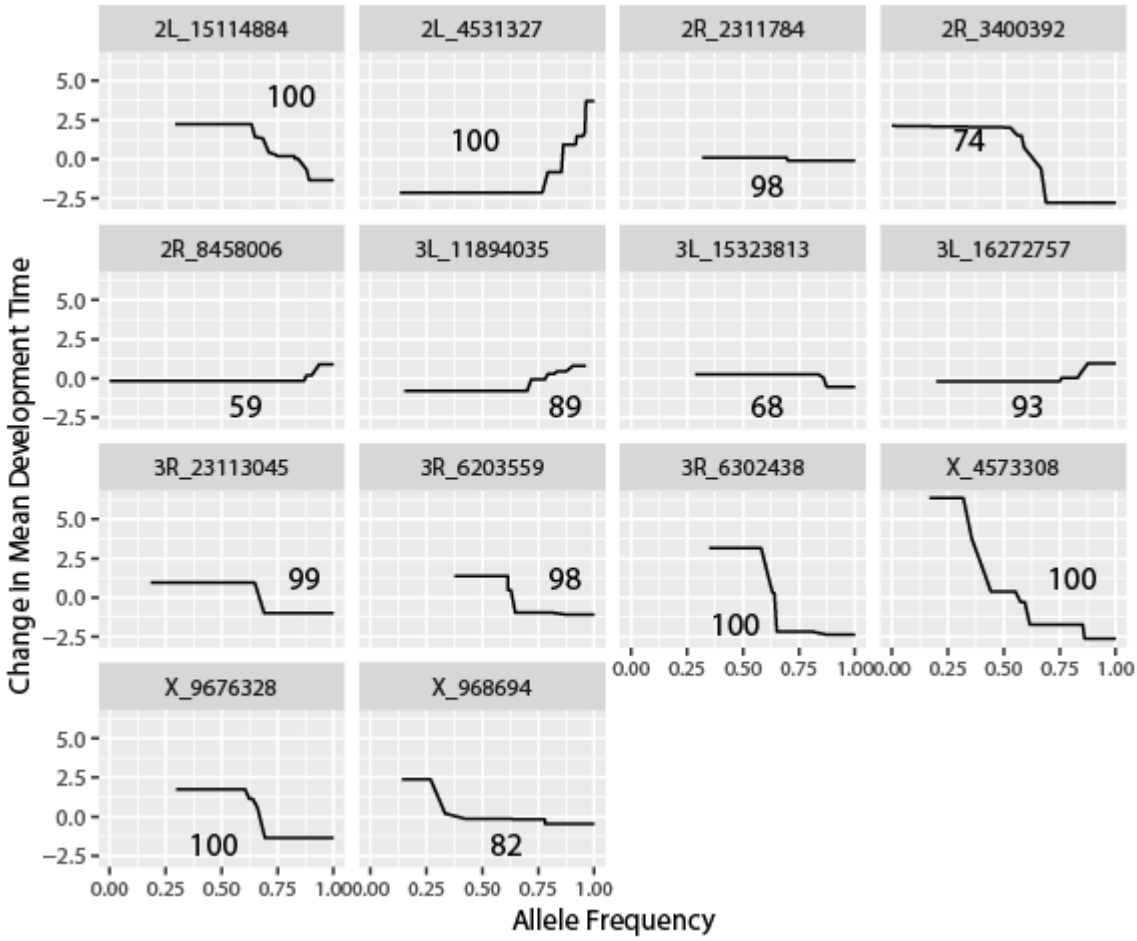

Figure 5.7. The FLAM step function, $E[P_i|\hat{p}_{ij}] = \omega_i$, predicting development time as a function of SNP frequency at 14 genomic locations identified by FLAM. Each figure contains an integer which equals the number of times that SNP was included in the 100 permutations of the genetic matrix. The sparse set using the 50% criteria.

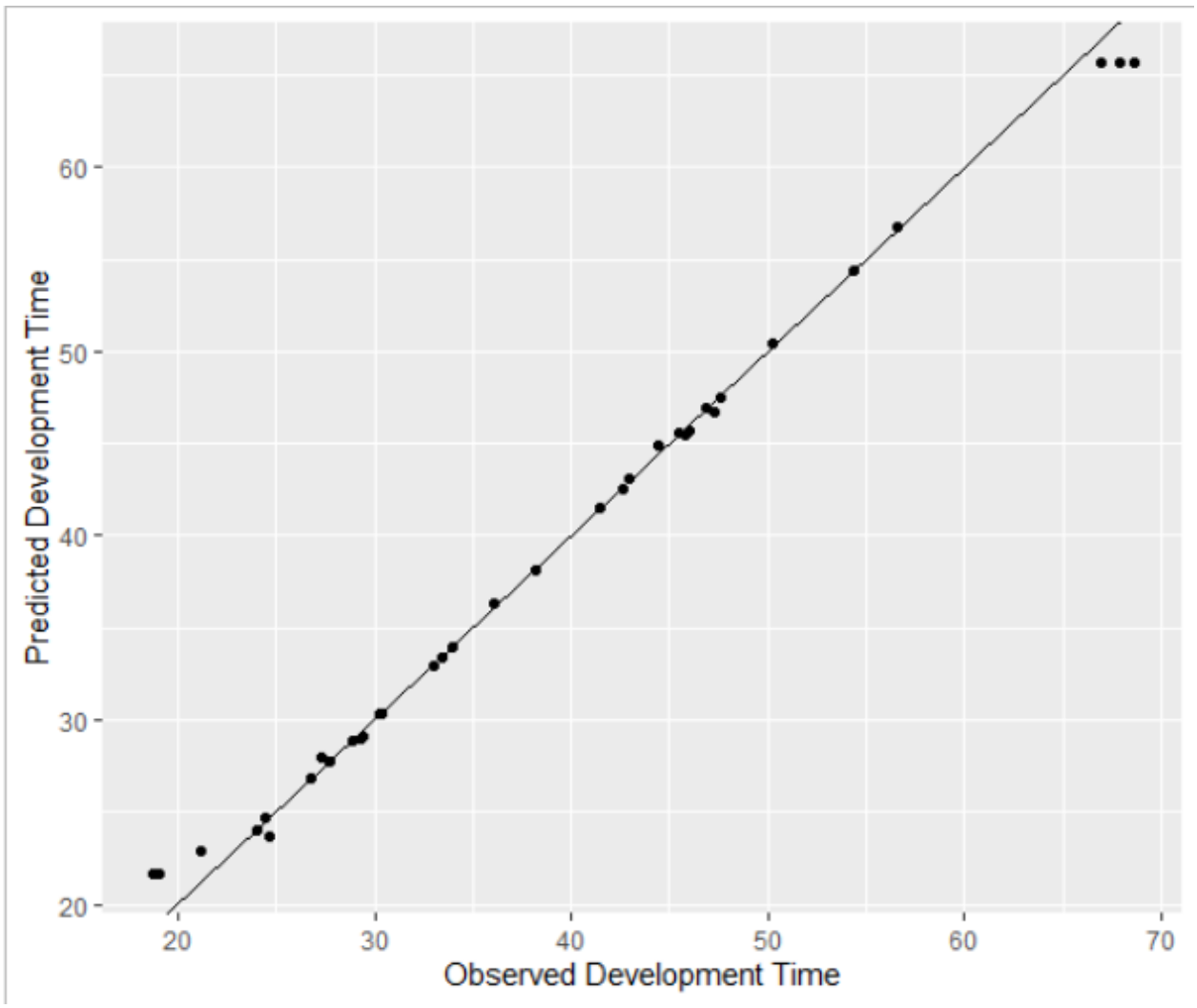

Figure 5.8 Predicted development time in the original transformed units for each of the 40 populations. These predictions were based on FLAM fitted to the 14 SNPs shown in figure 5.7. The correlation between development time rate on test data and the observed development time was 0.95 with a 95% confidence interval (0.90, 0.97).

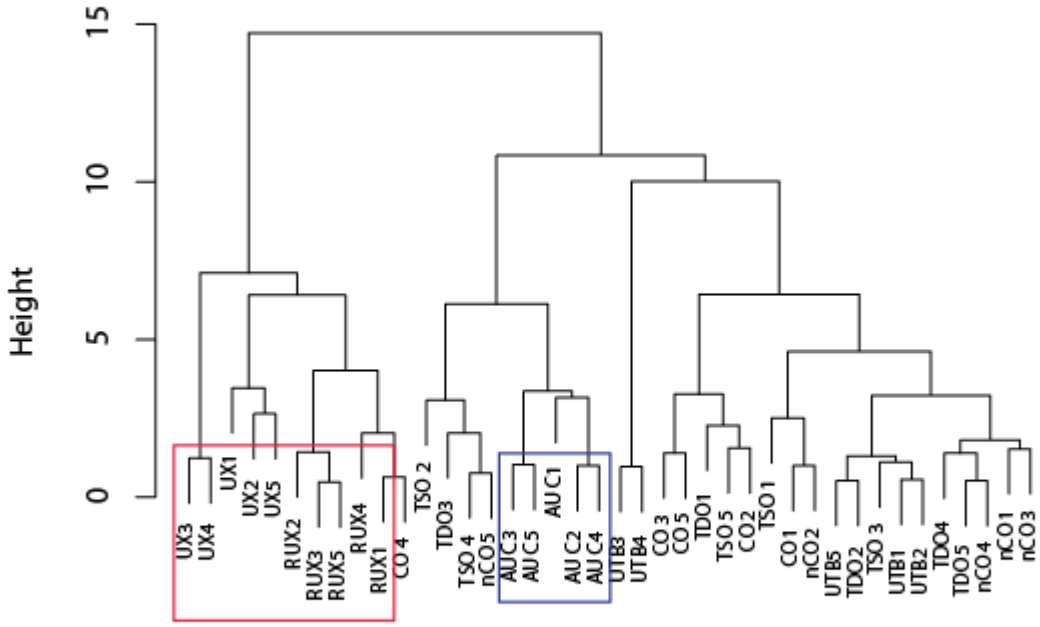

Figure 5.9 A cluster analysis of 40 populations based on genetic variation at the *for* locus. Distances between populations and clusters were based on the Euclidian distance. Group joining was determined by the largest pairwise distance between pairs of populations within each group (the compete method). The red box outlines the grouping of the 10 UX and RUX populations while the blue box outlines the grouping of the five AUC populations.

TABLES

Table 5.1 The sparse list of SNPs for four phenotypes with their SNP and cytogenic location (Cen): feeding rate, viability, development time and growth rate.

SNP	Cen	Feeding Rate	Viability	Development Time	Growth Rate
2L_4531327	24F			X	
2L_7207001	27E	X			
2L_7403125	27F		X		
2L_9680141	30C	X			
2L_10377993	31D		X		
2L_11478767	32F	X			
2L_13379565	34B	X			
2L_13810371	34E	X			
2L_15114884	35C			X	
2L_23215692	40F	X			
2R_2311784	41A			X	
2R_3400392	41A			X	
2R_3704100	41A				X
2R_6012388	42A	X			
2R_7248434	43A				X
2R_7985106	44A		X		
2R_8458006	44D			X	
2R_8516281	44D				X
2R_19501220	56D		X		
2R_19727187	56F		X		
2R_21971950	58A		X		
2R_22285947	58D				X
2R_23060438	59C	X			
2R_25091135	60F		X		
3L_151386	61B		X		
3L_4368193	64B		X		
3L_9086420	66F	X			
3L_10769473	67E	X			
3L_10920886	67F	X			
3L_11894035	68E		X	X	
3L_13775342	70C	X			
3L_15323813	71C		X	X	
3L_16272757	72D		X	X	
3L_16350779	72E				X
3L_17260381	74A	X			
3L_17928678	75A	X			
3L_21395960	78D	X			
3L_22241372	79D	X			
3R_6203559	83E			X	
3R_6302438	83E			X	
3R_23113045	94D			X	

3R_8945249	66D	X					
3R_17087687	89E	X					
3R_25361898	96E		X				
X_442317	1B					X	
X_4573308	4C			X			
X_6900105	6E					X	
X_7338301	7B	X					
X_7761153	7B	X					
X_8478041	7F	X					
X_9676328	8F			X			
X_19036434	18A	X					

Table 5.2. SNP location, FlyBase ID of potential associated gene, symbol, and FlyBase functional attribution from FLAM. Non-coding genes marked with *. Genes with no known FlyBase attribution are labeled N/A in the fourth column.

SNP	Gene (FB ID)	Symbol	FlyBase Functional Attribution from FLAM
2L_4531327	FBgn0053196	dpy	dumpy (dpy) encodes an extracellular protein involved in epidermal-cuticle attachment, aposition of wing surfaces and trachea development
2L_7207001	FBgn0031893	CG4495	N/A
	FBgn0031894	CG4496	N/A
	FBgn0031895	CG4497	N/A
	FBgn0031896	CG4502	E2 Ubiquitin conjugating enzymes
	FBgn0031897	CG13784	cellular organization/biogenesis, transport/localization and endomembrane system
	FBgn0263133	ico	mitochondrial translation elongation factor G 1 (mEFG1) encodes a mitochondrial elongation factor that catalyzes the translocation of the mRNAs and tRNAs along the ribosomes via GTP hydrolysis.
	FBgn0266324	CR44989	N/A*
	FBgn0259111	Ndae1	Na[+]-driven anion exchanger 1 (Ndae1) encodes an intrinsic membrane protein that reversibly mediates the exchange of 1 Na[+] and 2 HCO[[3]][-] for 1 Cl[-] at basolateral membranes of gut and renal epithelia as well as CNS neurons. In the CNS, the product of Ndae1 can lower resting intracellular Cl[-] controlling GABA(A) Cl-channel activity
	FBgn0264343	CG43799	N/A
2L_7403125	FBgn0002938	ninaC	Myosin function
	FBgn0031904	CG5149	N/A
	FBgn0031905	gudu	N/A
	FBgn0031906	CG5160	N/A
	FBgn0264888	CR44079	pseudogene attribute
	FBgn0031907	CG5171	trehalose-phosphatase activity
	FBgn0267924	CR46205	N/A*
	FBgn0031908	CG5177	haloacid dehalogenases
	FBgn0031909	CG5181	N/A
	FBgn0031910	CG15818	c-type lectin-like
	FBgn0028387	chm	chameau (chm) encodes a histone acetyl transferase involved in epigenetic mechanisms of transcriptional control and regulation of replication origin activity. Its functions include modulation of JNK activity and gene regulation by Polycomb group genes.
	FBgn0264439	CR43857	N/A*
	FBgn0262453	mir-275	miRNA gene
	FBgn0262458	mir-305	miRNA gene
	FBgn0031912	CG5261	N/A

2L_9680141	FBgn0032136	Apoltp	Apolipoprotein lipid transfer particle (Apoltp) encodes a apolipoprotein of the ApoB family. It is assembled with lipids to form high density, low abundance lipoproteins named lipid transfer particles (LTP). LTP catalyzes the transfer of lipids from the gut to circulating lipophorin and from lipophorin to peripheral tissues, such as imaginal discs and ovaries
	FBgn0053299	CG33299	N/A
	FBgn0267927	CR46208	N/A*
	FBgn0264999	CR44150	N/A*
	FBgn0000273	Pka-C1	Protein kinase, cAMP-dependent, catalytic subunit 1 (Pka-C1) encodes a serine/threonine kinase that contributes to axis specification, rhythmic behavior and synaptic transmission.
	FBgn0015393	hoip	RNA-binding protein hoip accelerates polyQ-induced neurodegeneration in Drosophila
	FBgn0011207	pelo	N/A
	FBgn0051710	CG31710	N/A
2L_10377993	FBgn0085396	CG34367	N/A
	FBgn0032228	CG5367	N/A
	FBgn0032229	CG5045	N/A
	FBgn0027568	Cand1	CAND family
	FBgn0003087	pim	pimples (pim) encodes the Drosophila Securin homolog. Securins bind to the endoprotease Separase. In some species including Drosophila, this binding appears to promote proper folding of Separase. In addition, securin binding inhibits separase endoprotease activity in all species via an inhibitory pseudosubstrate region.
	FBgn0032230	lft	lowfat (lft) encodes a protein that influences Dachsous-Fat signaling by elevating levels of the products of ds and ft at apical membranes. Mutation of lft causes mild planar cell polarity and growth phenotypes.
	FBgn0032231	CG5056	N/A
	FBgn0041723	rho-5	N/A
	FBgn0053303	CG33303	N/A
	FBgn0032233	dpr19	N/A
	FBgn0032234	gny	garnysstan (gny) encodes a glucosyltransferase involved in cuticle barrier formation and thought to have a role in N-glycan biosynthesis and epidermal differentiation.
	FBgn0032235	CG5096	N/A

	FBgn0004106	Cdk1	Cyclin-dependent kinase 1 (Cdk1) encodes a catalytic protein kinase subunit that can only become active after association with either CycA, CycB or CycB3 products. The protein kinase activities of these complexes (CycA-Cdk1, CycB-Cdk1, CycB3-Cdk1) control important aspects of progression through the cell cycle. Functionally, the different Cdk1 complexes are partially redundant. They phosphorylate hundreds of target proteins and are most important for progression into and through mitotic and meiotic M phases.
	FBgn0032236	mRpS7	N/A
	FBgn0267821	da	daughterless (da) encodes a class I bHLH protein important for sex determination and dosage compensation by controlling the feminizing switch gene Sxl. It participates in transcriptional regulation of a wide variety processes, including oogenesis, neurogenesis, myogenesis and cell proliferation
	FBgn0267439	CR45789	antisense RNA*
	FBgn0262782	Mdh1	Malate dehydrogenase 1 (Mdh1) encodes a L-malate dehydrogenase involved in the interconversion of malate and oxaloacetate
	FBgn0051716	Cnot4	CCR4-NOT transcription complex subunit 4 (Cnot4) encodes a positive regulator of the Jak/Stat pathway which acts by stabilizing the DNA binding of the product of Stat92E.
2L_11478767	FBgn0003313	sala	N/A
	FBgn0263087	CG43355	N/A
	FBgn0032361	CG6488	Component of oligomeric golgi complex 8
	FBgn0032362	spz4	Spätzle related
2L_13379565	FBgn0032506	CG9395	N/A
	FBgn0032507	CG9377	N/A
	FBgn0022069	Nnp-1	RRP1
	FBgn0032509	CG6523	Family of disulfide oxidoreductases
	FBgn0051852	Tap42	Two A-associated protein of 42kDa (Tap42) encodes a phosphatase inhibitor involved in wing disc morphogenesis and the organization of the mitotic spindle
	FBgn0051855	CG31855	N/A
	FBgn0032511	ND-B22	NADH dehydrogenase (ubiquinone) B22 subunit
	FBgn0032512	CG9305	B double prime 1 (Bdp1) encodes a transcription factor involved in snRNA gene transcription
	FBgn0032513	CG6565	N/A
	FBgn0032514	CG9302	N/A
	FBgn0025724	beta'COP	structural molecule activity

	FBgn0032515	loqs	loquacious (loqs) encodes a double-stranded RNA binding protein with multiple splicing isoforms. loqs-PB and loqs-PA associate with the product of Dcr-1 to facilitate processing of precursor miRNA into mature miRNA, whereas loqs-PD interacts with the product of Dcr-2 to enhance processing of dsRNA into siRNA.
	FBgn0265810	CR44599	antisense RNA*
	FBgn0032516	CG9293	Inhibitor of growth family member 5 (Ing5) encodes a histone acetyltransferase that forms part of the Enok complex, which interacts with the product of elg1 to promote the G1/S transition via proliferating cell nuclear antigen (PCNA) unloading.
	FBgn0032517	CG7099	N/A
	FBgn0032518	RpL24	Ribosomal protein L24
	FBgn0032519	CG16957	N/A
	FBgn0032520	CG10859	N/A
	FBgn0265811	CR44600	antisense RNA:CR44600*
	FBgn0032521	CG7110	N/A
2L_13810371	FBgn0028539	CG31731	Adenosinetriphosphatase
	FBgn0266141	CR44847	antisense RNA:CR44847*
	FBgn0028931	CG16863	N/A
	FBgn0028538	Sec71	Secretory 71
	FBgn0027844	CAH1	Carbonic anhydrase 1 (CAH1) encodes an alpha-class carbonate dehydratase that catalyzes the reversible hydration/dehydration of carbon dioxide to bicarbonate and protons.
	FBgn0028658	Adat1	adenosine deaminases
	FBgn0028919	CG16865	N/A
	FBgn0032533	CG16888	N/A
	FBgn0001965	Sos	Son of sevenless
	FBgn0000153	b	black (b) encodes an aspartate 1-decarboxylase. Mutations of b result in viable flies with a cuticular melanization phenotype.
	FBgn0004406	tam	tamas (tam) encodes the catalytic subunit of mitocondrial DNA polymerase γ, which is responsible for the replication and repair of the mitochondrial genome.
	FBgn0001961	Arpc1	Actin-related protein 2/3 complex, subunit 1 (Arpc1) encodes one of seven components of the highly conserved Arp2/3 complex. The Arp2/3 complex is required for many cellular and developmental processes requiring F-actin polymerization, including germline ring canal growth, embryonic cellularization, myoblast fusion, and endocytosis.
	FBgn0015271	Orc5	Origin recognition complex subunit 5
	FBgn0004407	DNApol-gamma35	DNA polymerase gamma 35kD subunit (DNApol-γ35) encodes the accessory subunit of mitocondrial DNA polymerase γ, which is responsible for the replication and repair of the mitochondrial genome.

	FBgn0064115	CG33649	Glutaminyl-tRNA synthase (glutamine-hydrolyzing)
	FBgn0026373	RpII33	RNA polymerase II 33kD subunit
	FBgn0260407	mRpS23	mitochondrial ribosomal protein S23
2L_15114884	FBgn0028879	CG15270	N/A
	FBgn0028878	CG15269	N/A
	FBgn0001978	stc	shuttle craft (stc) encodes a zinc finger transcription factor of the NFX1 family. Maternally contributed protein is essential for segmentation. During the late stages of embryonic neurogenesis, zygotically expressed stc product is involved in axon guidance, which is vital for proper muscle innervation. stc expression in the embryo is important for the adult life span and aging.
	FBgn0264435	CR43853	long non-coding RNA:CR43853
	FBgn0265917	CR44706	pseudogene attribute
	FBgn0264436	CR43854	long non-coding RNA:CR43854
	FBgn0028889	tRNA:L:35C	tRNA gene
2L_23215692	FBgn0267986	CR46253	N/A
2R_2311784	No known association	No known association	No known association
2R_3400392	No known association	No known association	No known association
2R_3704100	No known association	No known association	No known association
2R_6012388	FBgn0264959	Src42A	Src oncogene at 42A (Src42A) encodes the none-receptor tyrosine kinase regulating cell proliferation, cell adhesion and morphogenetic processes including dorsal closure, tracheal tube size control and germ band elongation. It is regulated by both inhibitory and activating tyrosine phosphorylation
	FBgn0000043	Act42A	Actin 42A
	FBgn0033050	Pngl	PNGase-like
	FBgn0033051	Strica	Ser/Thr-rich caspase (Strica) encodes a member of initiator caspases. It contributes, redundantly or cooperatively with the product of Dronc, to the apoptosis of selective larval neurons during metamorphosis and nurse cells during oogenesis, and competitive apoptosis of heterozygous mutant cells.
	FBgn0033052	SCAP	SREBP cleavage activating protein
	FBgn0033054	CG14591	N/A
2R_7248434	FBgn0026389	Or43a	Odorant receptor 43a (Or43a) encodes a multi-transmembrane chemoreceptor that mediates response to volatile chemicals. It is sensitive to a number of volatile small molecule odorants including cyclohexanol, cyclohexanone, benzaldehyde, and benzyl alcohol
	FBgn0026602	Ady43A	Adenosine kinase
	FBgn0033153	Gadd45	growth arrest and DNA damage-inducible gene 45

	FBgn0266818	CR45280	long non-coding RNA:CR45280
	FBgn0033154	CG1850	N/A
	FBgn0033155	Br140	Bromodomain-containing protein, 140kD (Br140) encodes a histone acetyltransferase that is a component of the Enok complex, which interacts with the product of elg1 via Br140 to promote the G1/S transition via proliferating cell nuclear antigen (PCNA) unloading.
2R_7985106	FBgn0026361	Sep5	Septin 5 (Sep5) encodes a member of the septin family of GTP-binding proteins. Septins form hetero-oligomeric filaments and rings that have roles in cytokinesis, cell polarity and membrane rigidity. Sep5 is not essential for development, but mutants display a synthetic lethal pupal phenotype when combined with mutations in its paralog Sep2.
	FBgn0027548	nito	spenito
	FBgn0033243	CG14763	N/A
	FBgn0033244	CG8726	N/A
	FBgn0027054	CSN4	COP9 signalosome subunit 4 (CSN4) encodes the subunit 4 of the COP9 signalosome (CSN) and is essential for the stability of the complex. The CSN plays a central role in the regulation of E3-cullin RING ubiquitin ligases.
	FBgn0033246	ACC	Acetyl-CoA carboxylase (ACC) encodes a ubiquitous metabolic enzyme. It catalyzes the carboxylation of acetyl-CoA to malonyl-CoA, the rate-limiting substrate for fatty acid synthesis. It is essential in the embryo and in the oenocytes (specialized abdominal cells) for the watertightness of the respiratory system. It is required in the fat body for triglyceride storage and in the muscles for locomotor activity.
	FBgn0033247	Nup44A	Nucleoporin at 44A (Nup44A) encodes a nuclear pore protein involved in TORC1 signaling, autophagy and oogenesis.
	FBgn0265299	CR44272	long non-coding RNA:CR44272
	FBgn0033248	Dic3	Dicarboxylate carrier 3 (Dic3) encodes a protein of the inner mitochondrial membrane that belongs to a subfamily of mitochondrial dicarboxylate carriers. It transports only phosphate, sulphate, and thiosulphate but not dicarboxylates (like the canonical dicarboxylate carrier encoded by Dic1). The product of Dic3 is exclusively present in the pupal stage
	FBgn0027788	Hey	Hairy/E(spl)-related with YRPW motif (Hey) encodes a transcription factor involved in neuron fate determination. Heyexpression is regulated by Notch signalling in the embryonic and larval central nervous system
	FBgn0033249	CG11191	N/A
	FBgn0013307	Odc1	Ornithine decarboxylase
	FBgn0013308	Odc2	Ornithine decarboxylase 2
	FBgn0033250	CG14762	N/A

	FBgn0011812	Lcp1Psi	Larval cuticle protein 1 pseudogene
	FBgn0002533	Lcp2	Larval cuticle protein 2
	FBgn0002534	Lcp3	Larval cuticle protein 3
	FBgn0002535	Lcp4	Larval cuticle protein 4
	FBgn0033292	Cyp4ad1	Cyp4ad1
	FBgn0014469	Cyp4e2	Cytochrome P450-4e2
	FBgn0262316	mir-986	mir-986 stem loop
	FBgn0015034	Cyp4e1	Cytochrome P450-4e1
2R_8458006	FBgn0002570	Mal-A1	Maltase A1, Alpha-glucosidase
	FBgn0002569	Mal-A2	Maltase A2, Alpha-glucosidase
	FBgn0002571	Mal-A3	Maltase A3, Alpha-glucosidase
	FBgn0033294	Mal-A4	Maltase A4, Alpha-glucosidase
	FBgn0050359	Mal-A5	Maltase A5, Alpha-glucosidase
	FBgn0050360	Mal-A6	Maltase A6, Alpha-glucosidase
	FBgn0033296	Mal-A7	Maltase A7, Alpha-glucosidase
	FBgn0033297	Mal-A8	Maltase A8, Alpha-glucosidase
	FBgn0050361	mtt	metabotropic glutamate receptors
2R_8516281	FBgn0050361	mtt	metabotropic glutamate receptors
	FBgn0260934	par-1	par-1 (par-1) encodes a protein kinase involved in multiple processes, including microtubule cytoskeleton organization, axis specification and cell polarity. It regulates hippo signaling and osk mRNA localization
	FBgn0034451	TBCB	tubulin-binding cofactor B (TBCB) encodes one of the cofactors required for the assembly of functional α/β-Tubulin dimers needed for microtubule assembly. It is essential for microtubule-associated transport and cell polarity, but not for cell division.
	FBgn0034452	Oseg6	N/A
	FBgn0026378	Rep	Rab escort protein
2R_19501220	FBgn0261456	hpo	hippo (hpo) encodes a kinase in the Salvador-Warts-Hippo pathway. It controls tissue growth by controlling cell growth, proliferation and apoptosis. It has several roles in post-mitotic cells including fate specification of photoreceptors and tiling of dendritic neurons.
	FBgn0034454	CG15120	N/A
	FBgn0040732	CG16926	N/A
	FBgn0034455	CG11007	N/A
	FBgn0265665	CR44472	long non-coding RNA:CR44472
	FBgn0050223	tRNA:H:56E	transfer RNA:Histidine-GTG 1-5
	FBgn0034456	Ir56b	Ionotropic receptor 56b
	FBgn0034457	Ir56c	Ionotropic receptor 56c
	FBgn0034458	Ir56d	Ionotropic receptor 56d
	FBgn0003435	sm	smooth (sm) encodes an RNA binding, heterogeneous nuclear ribonucleoprotein involved in axon guidance, mRNA processing, chemosensation, determination of lifespan and feeding behavior.
2R_19727187	FBgn0034470	Obp56d	Odorant-binding protein 56d

	FBgn0265702	CR44509	antisense RNA:CR44509
	FBgn0034471	Obp56e	Odorant-binding protein 56e
	FBgn0043533	Obp56f	Odorant-binding protein 56f
	FBgn0034472	CG8517	N/A
	FBgn0050215	tRNA:CR30215	tRNA gene, transfer RNA:Isoleucine-AAT 1-9
	FBgn0050452	tRNA:CR30452	tRNA gene, transfer RNA:initiator Methionine-CAT 1-1
	FBgn0050218	tRNA:CR30218	tRNA gene, transfer RNA:initiator Methionine-CAT 1-2
	FBgn0011850	tRNA:E4:56Fc	tRNA gene, transfer RNA:Glutamic acid-TTC 1-4
	FBgn0065076	snoRNA:185	snoRNA_gene
	FBgn0011849	tRNA:E4:56Fb	tRNA gene, transfer RNA:Glutamic acid-TTC 1-5
	FBgn0011848	tRNA:E4:56Fa	tRNA gene, transfer RNA:Glutamic acid-TTC 1-6
	FBgn0050454	tRNA:CR30454	tRNA gene, transfer RNA:Glutamic acid-CTC 2-1
	FBgn0050220	tRNA:CR30220	tRNA gene, transfer RNA:Glutamic acid-CTC 2-2
	FBgn0050449	tRNA:CR30449	tRNA gene, transfer RNA:Glutamic acid-CTC 5-1 pseudogene
	FBgn0053452	5SrRNA:CR33452	rRNA gene, 5SrRNA:CR33452
	FBgn0053451	5SrRNA:CR33451	rRNA gene, 5SrRNA:CR33451
	FBgn0053450	5SrRNA:CR33450	rRNA gene, 5SrRNA:CR33450
	FBgn0053449	5SrRNA:CR33449	rRNA gene, 5SrRNA:CR33449
	FBgn0053448	5SrRNA:CR33448	rRNA gene, 5SrRNA:CR33448
	FBgn0053447	5SrRNA:CR33447	rRNA gene, 5SrRNA:CR33447
	FBgn0053446	5SrRNA:CR33446	rRNA gene, 5SrRNA:CR33446
	FBgn0053445	5SrRNA:CR33445	rRNA gene, 5SrRNA:CR33445
	FBgn0053444	5SrRNA:CR33444	rRNA gene, 5SrRNA:CR33444
	FBgn0053443	5SrRNA:CR33443	rRNA gene, 5SrRNA:CR33443
	FBgn0053442	5SrRNA:CR33442	rRNA gene, 5SrRNA:CR33442
	FBgn0053441	5SrRNA:CR33441	rRNA gene, 5SrRNA:CR33441
	FBgn0053440	5SrRNA:CR33440	rRNA gene, 5SrRNA:CR33440
	FBgn0053439	5SrRNA:CR33439	rRNA gene, 5SrRNA:CR33439
	FBgn0053438	5SrRNA:CR33438	rRNA gene, 5SrRNA:CR33438
	FBgn0053437	5SrRNA:CR33437	rRNA gene, 5SrRNA:CR33437
	FBgn0053436	5SrRNA:CR33436	rRNA gene, 5SrRNA:CR33436
	FBgn0053435	5SrRNA:CR33435	rRNA gene, 5SrRNA:CR33435
	FBgn0053434	5SrRNA:CR33434	rRNA gene, 5SrRNA:CR33434
	FBgn0053433	5SrRNA:CR33433	rRNA gene, 5SrRNA:CR33433
	FBgn0053432	5SrRNA:CR33432	rRNA gene, 5SrRNA:CR33432
	FBgn0053431	5SrRNA:CR33431	rRNA gene, 5SrRNA:CR33431
	FBgn0053430	5SrRNA:CR33430	rRNA gene, 5SrRNA:CR33430
	FBgn0053429	5SrRNA:CR33429	rRNA gene, 5SrRNA:CR33429
	FBgn0053428	5SrRNA:CR33428	rRNA gene, 5SrRNA:CR33428
	FBgn0053427	5SrRNA:CR33427	rRNA gene, 5SrRNA:CR33427
	FBgn0053426	5SrRNA:CR33426	rRNA gene, 5SrRNA:CR33426

	FBgn0053425	5SrRNA:CR33425	rRNA gene, 5SrRNA:CR33425
	FBgn0053424	5SrRNA:CR33424	rRNA gene, 5SrRNA:CR33424
	FBgn0053423	5SrRNA:CR33423	rRNA gene, 5SrRNA:CR33423
	FBgn0053422	5SrRNA:CR33422	rRNA gene, 5SrRNA:CR33422
	FBgn0053421	5SrRNA:CR33421	rRNA gene, 5SrRNA:CR33421
	FBgn0053420	5SrRNA:CR33420	rRNA gene, 5SrRNA:CR33420
	FBgn0053419	5SrRNA:CR33419	rRNA gene, 5SrRNA:CR33419
	FBgn0053418	5SrRNA:CR33418	rRNA gene, 5SrRNA:CR33418
	FBgn0053417	5SrRNA:CR33417	rRNA gene, 5SrRNA:CR33417
	FBgn0053416	5SrRNA-Psi:CR33416	pseudogene attribute, 5SrRNA-Ψ:CR33416
	FBgn0053415	5SrRNA:CR33415	rRNA gene, 5SrRNA:CR33415
	FBgn0053414	5SrRNA:CR33414	rRNA gene, 5SrRNA:CR33414
	FBgn0053413	5SrRNA:CR33413	rRNA gene, 5SrRNA:CR33413
	FBgn0053412	5SrRNA:CR33412	rRNA gene, 5SrRNA:CR33412
	FBgn0053411	5SrRNA:CR33411	rRNA gene, 5SrRNA:CR33411
	FBgn0053410	5SrRNA:CR33410	rRNA gene, 5SrRNA:CR33410
	FBgn0053409	5SrRNA:CR33409	rRNA gene, 5SrRNA:CR33409
	FBgn0053408	5SrRNA:CR33408	rRNA gene, 5SrRNA:CR33408
	FBgn0053407	5SrRNA:CR33407	rRNA gene, 5SrRNA:CR33407
	FBgn0053406	5SrRNA:CR33406	rRNA gene, 5SrRNA:CR33406
	FBgn0053405	5SrRNA:CR33405	rRNA gene, 5SrRNA:CR33405
	FBgn0053404	5SrRNA:CR33404	rRNA gene, 5SrRNA:CR33404
	FBgn0053403	5SrRNA:CR33403	rRNA gene, 5SrRNA:CR33403
	FBgn0053402	5SrRNA:CR33402	rRNA gene, 5SrRNA:CR33402
	FBgn0053401	5SrRNA:CR33401	rRNA gene, 5SrRNA:CR33401
	FBgn0053400	5SrRNA:CR33400	rRNA gene, 5SrRNA:CR33400
	FBgn0053399	5SrRNA:CR33399	rRNA gene, 5SrRNA:CR33399
	FBgn0053398	5SrRNA:CR33398	rRNA gene, 5SrRNA:CR33398
	FBgn0053397	5SrRNA:CR33397	rRNA gene, 5SrRNA:CR33397
	FBgn0053396	5SrRNA:CR33396	rRNA gene, 5SrRNA:CR33396
	FBgn0053395	5SrRNA:CR33395	rRNA gene, 5SrRNA:CR33395
	FBgn0053394	5SrRNA:CR33394	rRNA gene, 5SrRNA:CR33394
	FBgn0053393	5SrRNA:CR33393	rRNA gene, 5SrRNA:CR33393
	FBgn0053392	5SrRNA:CR33392	rRNA gene, 5SrRNA:CR33392
	FBgn0053391	5SrRNA:CR33391	rRNA gene, 5SrRNA:CR33391
2R_21971950	FBgn0034670	CG13488	N/A
	FBgn0263324	CR43405	N/A
	FBgn0034671	CG13494	N/A
	FBgn0085398	ppk9	pickpocket 9
	FBgn0054029	CG34029	BP1066
	FBgn0034674	CG9304	N/A
	FBgn0041237	Gr58c	Gustatory receptor 58c
	FBgn0041238	Gr58b	Gustatory receptor 58b
	FBgn0041239	Gr58a	Gustatory receptor 58a

	FBgn0050401	CG30401	distal antenna-young (dany) encodes a chromatin-binding protein required in spermatocytes for a normal gene expression profile, meiosis and sperm production.
	FBgn0085399	CG34370	N/A
2R_22285947	FBgn0020307	dve	defective proventriculus (dve) encodes a transcriptional repressor that binds to the K50 site. It is involved in developmental patterning, cell-type specification, and functional differentiation
	FBgn0263341	CR43422	long non-coding RNA:CR43422
	FBgn0034717	CG5819	N/A
	FBgn0034718	wdp	windpipe (wdp) encodes a regulator of the JAK/STAT pathway. It is involved in intestinal homeostasis, trachea development and synaptic target recognition
2R_23060438	FBgn0261705	CG42741	N/A
	FBgn0265187	CG44252	Long-chain-fatty-acid--CoA ligase, ABC-type fatty-acyl-CoA transporter
	FBgn0265188	CG44253	N/A
	FBgn0010622	DCTN3-p24	Dynactin 3, p24 subunit
	FBgn0034814	CG9890	N/A
	FBgn0261363	PPO3	Prophenoloxidase 3 (PPO3) is expressed in lamellocytes (a type of hemocyte cell involved in encapsulation) and its product involved in the melanization reaction during wasp encapsulation.
	FBgn0021979	l(2)k09913	lethal (2) k09913
	FBgn0050201	tRNA:CR30201	tRNA gene, transfer RNA:Serine-TGA 1-1
	FBgn0050202	tRNA:CR30202	tRNA gene, transfer RNA:Serine-TGA 2-1
	FBgn0003062	Fib	Fibrillarin (Fib) encodes an rRNA 2'-O-methyltransferase required for pre-rRNA processing
	FBgn0065073	snoRNA:229	snoRNA_gene
	FBgn0034816	CG3085	N/A
	FBgn0034817	Art7	Type II protein arginine methyltransferase, Type III protein arginine methyltransferase
	FBgn0263707	CG43659	N/A
	FBgn0045483	Gr59a	Gustatory receptor 59a
	FBgn0045482	Gr59b	Gustatory receptor 59b
2R_25091135	FBgn0001148	gsb	gooseberry, paired homeobox transcription factors
	FBgn0004919	gol	goliath (gol) encodes an E3 ubiquitin ligase, whose cellular substrates are unknown. During embyrogenesis gol is expressed in fusion-competence myoblasts of the somatic and visceral mesoderm
	FBgn0050198	tRNA:CR30198	tRNA gene, transfer RNA:Alanine-CGC 1-1
	FBgn0050199	tRNA:CR30199	tRNA gene, transfer RNA:Alanine-CGC 1-2
	FBgn0050200	tRNA:CR30200	tRNA gene, transfer RNA:Alanine-CGC 1-3
	FBgn0266129	lov	jim lovell
3L_151386	FBgn0266949	CR45400	long non-coding RNA:CR45400

	FBgn0020386	Pdk1	Phosphoinositide-dependent kinase 1
	FBgn0267455	CR45805	long non-coding RNA:CR45805
	FBgn0035099	CG6845	N/A
	FBgn0263988	Dic61B	Dynein intermediate chain at 61B (Dic61B) encodes an axonemal dynein intermediate chain protein expressed specifically in male germ cells. It is required for the development and precise assembly of sperm axonemes, thus it is essential for male fertility.
	FBgn0035101	p130CAS	p130CAS (p130CAS) encodes an SH3 domain-containing protein phosphorylated by Src kinases. It contributes to integrin-mediated adhesion and in vertebrate cells is phosphorylated in response to mechanical stretch
	FBgn0262681	CR43151	long non-coding RNA:CR43151
	FBgn0035102	CG7049	Formylglycine-generating enzyme
	FBgn0267456	CR45806	antisense RNA:CR45806*
	FBgn0035103	Vdup1	Vitamin D3 up-regulated protein 1
3L_4368193	FBgn0035533	Cip4	Cdc42-interacting protein 4 (Cip4) encodes an F-BAR protein that functions as an adaptor protein regulating membrane curvature and dynamics.
	FBgn0266649	CR45156	long non-coding RNA:CR45156
	FBgn0035534	mRpS6	mitochondrial ribosomal protein S6
	FBgn0053514	CG33514	N/A
	FBgn0035537	CG11342	N/A
	FBgn0035538	DopEcR	Dopamine/Ecdysteroid receptor (DopEcR) encodes a GPCR that shows ligand-biased activation. It can be activated by dopamine to increase cyclic AMP levels and by the insect steroid ecdysone to activate the MAPKinase pathway. It is widely expressed in the nervous system and can modulate a wide variety of complex behaviors including male courtship, locomotion, the response to stressful social interactions and the regulation of appetite.
	FBgn0265467	CR44360	Dopamine/Ecdysteroid receptor
	FBgn0052240	CG32240	N/A
	FBgn0035539	slow	slowdown (slow) encodes a protein secreted from tendon cells and is required for integrin-mediated muscle-tendon adhesion.
3L_9086420	FBgn0000116	Argk	Arginine kinase
	FBgn0035959	CG4911	F box only proteins
	FBgn0035960	CG4942	N/A
	FBgn0023479	Tequila	tequila
	FBgn0043806	CG32032	N/A
	FBgn0040827	CG13315	N/A
	FBgn0011206	bol	boule (bol) encodes a translational regulator required in spermatogenesis for entry into meiosis and spermatid differentiation. It has an additional role in mushroom body γ neurons as a negative regulator of axon pruning
3L_10769473	FBgn0262890	CG43245	N/A

	FBgn0265737	CR44544	long non-coding RNA:CR44544
	FBgn0265738	CR44545	long non-coding RNA:CR44545
	FBgn0265739	CR44546	long non-coding RNA:CR44546
	FBgn0267950	CR46230	long non-coding RNA:CR46230
3L_10920886	FBgn0036111	Aps	Aps (Aps) encodes a diphosphoinositol-polyphosphate diphosphatase that hydrolyses a range of (di)nucleoside polyphosphates. It is involved in glucose and lipid homeostasis.
3L_11894035	FBgn0266100	CG44837	Beta-lactamase
	FBgn0036217	CG5906	N/A
	FBgn0036218	Sprn	Spermitin (Sprn) encodes a testis-specific mitochondrial lumen protein that first appears during the spermatocyte stage of spermatogenesis, and persists until late stages but is absent from mature sperm
	FBgn0036219	CCDC151	Coiled-coil domain containing protein 151
	FBgn0036220	CG5897	Bestrophin 4
	FBgn0263615	CR43624	long non-coding RNA:CR43624
3L_13775342	No known association	No known association	N/A
3L_15323813	FBgn0036491	Best4	Bestrophin 4
	FBgn0036492	Best3	Bestrophin 3 (Best3) encodes a paralog of the product of Best1, a Ca-activated Cl channel. A C-terminal domain inhibits the Cl channel function so that the product of Best3 does not produce current at physiological transmembrane voltages.
	FBgn0036493	CG7255	N/A
	FBgn0267933	CR46213	long non-coding RNA:CR46213
	FBgn0262892	CR43247	long non-coding RNA:CR43247
	FBgn0264724	CR43992	long non-coding RNA:CR43992
	FBgn0036494	Toll-6	Toll-6 (Toll-6) encodes a member of the Toll-like receptor family. It has neurotrophin receptor activity and contributes to dendrite guidance. Genetic interaction with 18w and Tollo suggests a Toll-6 role in convergent extension during early embryogenesis
	FBgn0036495	CG33259	N/A
3L_16272757	FBgn0036583	CG13055	N/A
	FBgn0036584	CG13054	N/A
	FBgn0040801	CG13053	N/A
	FBgn0085276	CG34247	N/A
	FBgn0036585	CG13071	N/A
	FBgn0036586	CG13070	N/A
	FBgn0040799	CG13051	N/A
	FBgn0040798	CG13069	N/A
	FBgn0036587	CG4950	N/A
	FBgn0036588	CG13068	N/A
	FBgn0085277	CG34248	N/A
	FBgn0036589	CG13067	N/A
	FBgn0040797	CG13066	N/A

	FBgn0036590	CG13065	N/A
	FBgn0036591	CG13050	N/A
	FBgn0040796	CG13064	N/A
	FBgn0036592	CG13049	N/A
	FBgn0036593	CG13048	N/A
	FBgn0036594	CG13047	N/A
	FBgn0036595	CG13046	N/A
	FBgn0036596	CG13045	N/A
	FBgn0036597	CG4962	N/A
3L_16350779	FBgn0036608	CG13040	N/A
	FBgn0036609	CG13039	N/A
	FBgn0040795	CG13038	N/A
	FBgn0036610	CG13058	N/A
	FBgn0053061	CG33061	N/A
	FBgn0053060	CG33060	N/A
	FBgn0040074	retinin	retinin
	FBgn0040794	CG13056	CPLCA cuticle protein family
	FBgn0036612	CG4998	S1A serine protease homologs
	FBgn0267553	CR45893	long non-coding RNA:CR45893
	FBgn0053257	CG33257	N/A
	FBgn0260460	mRpS34	mitochondrial ribosomal protein S34
	FBgn0036614	CG4925	Golgin 104
	FBgn0052357	tRNA:CR32357	transfer RNA:Methionine-CAT 1-4
	FBgn0036615	Tcs3	N/A
	FBgn0011693	Pdh	Photoreceptor dehydrogenase (Pdh) encodes a retinal pigment cell dehydrogenase involved in retinol metabolism
	FBgn0267554	CR45894	antisense RNA:CR45894
	FBgn0266417	CG45057	ringmaker (ringer) encodes a protein involved in microtubule bundle formation and polymerization.
	FBgn0036617	Cpr72Ea	Cuticular protein 72Ea
	FBgn0036618	Cpr72Eb	Cuticular protein 72Eb
	FBgn0036619	Cpr72Ec	Cuticular protein 72Ec
	FBgn0036620	CG4842	Alcohol dehydrogenase
	FBgn0042137	CG18814	Alcohol dehydrogenase
	FBgn0036621	roq	roquin
3L_17260381	FBgn0036702	CG6512	N/A
	FBgn0036703	CG7707	F box only proteins
	FBgn0264462	CR43870	long non-coding RNA:CR43870
	FBgn0036704	CG6497	N/A
	FBgn0264466	CR43874	long non-coding RNA:CR43874
	FBgn0036705	CG13723	N/A
	FBgn0266985	CR45436	long non-coding RNA:CR45436
3L_17928678	FBgn0036771	CG14353	WD repeat domain 92
	FBgn0036772	CG5290	N/A
	FBgn0267794	CR43174	long non-coding RNA:CR43174
	FBgn0266942	CR45393	long non-coding RNA:CR45393
3L_21395960	FBgn0261258	rgn	regeneration

	FBgn0267615	CR45953	long non-coding RNA:CR45953
	FBgn0002842	sa	spermatocyte arrest
	FBgn0267616	CR45954	long non-coding RNA:CR45954
	FBgn0037081	barc	barricade
3L_22241372	FBgn0266395	CR45035	long non-coding RNA:CR45035
	FBgn0028500	Rich	RIC1 homolog
	FBgn0015075	Ddx1	Dead-box-1 (Ddx1) encodes a member of the DEAD box family of RNA helicases that bind and unwind double-stranded RNA. Ddx1 product depletion is associated with small size and aberrant gametogenesis, possibly through alternative splicing of Sirup RNA transcript.
	FBgn0037156	CG11523	N/A
3R_6203559	FBgn0051562	CR31562	pseudogene attribute
	FBgn0037408	NPFR	Neuropeptide F receptor
	FBgn0037409	Osi24	Osiris 24
	FBgn0037410	Osi2	Osiris 2
	FBgn0037411	Osi3	Osiris 3
	FBgn0037412	Osi4	Osiris 4
	FBgn0037413	Osi5	Osiris 5
3R_6302438	FBgn0037419	Osi12	Osiris 12
	FBgn0037420	CG15597	N/A
	FBgn0037421	CG15594	N/A
	FBgn0037422	Osi13	Osiris 13
	FBgn0040279	Osi14	Osiris 14
	FBgn0037424	Osi15	Osiris 15
	FBgn0051561	Osi16	Osiris 16
	FBgn0267690	CG46026	N/A
	FBgn0051560	CG31560	N/A
	FBgn0037427	Osi17	Osiris 17
3R_23113045	FBgn0039054	Cow	Carrier of Wingless (Cow) encodes a secreted heparan sulfate proteoglycan that binds the product of wg and increases its extracellular mobility. The binding with the ligand encoded by wg is dependent on heparan sulfate modification of the product of Cow.
	FBgn0039049	CG6726	N-acyl-aliphatic-L-amino acid amidohydrolase
	FBgn0039050	CG17110	N-acyl-aliphatic-L-amino acid amidohydrolase
	FBgn0039051	CG17109	N-acyl-aliphatic-L-amino acid amidohydrolase
	FBgn0039052	CG6733	N-acyl-aliphatic-L-amino acid amidohydrolase
	FBgn0039053	CG6738	N-acyl-aliphatic-L-amino acid amidohydrolase
	FBgn0039055	Rassf	Ras association family member (Rassf) encodes a RASSF (ras-association domain family) protein that binds to the kinase encoded by hpo and promotes its dephosphorylation by the STRIPAK PP2A complex
	FBgn0039056	CenB1A	Centaurin beta 1A
	FBgn0051365	CG31365	N/A
	FBgn0051457	CG31457	N/A

	FBgn0004644	hh	hedgehog (hh) encodes the Hh signaling pathway ligand. It acts as a morphogen contributing to segment polarity determination, stem cells maintenance and cell migration. Post-translational modifications of the product of hh are essential for its restrictive spreading and signaling activity.
3R_8945249	FBgn0262604	CR43130	long non-coding RNA:CR43130
	FBgn0040532	CG8369	N/A
	FBgn0037637	IscU	Iron-sulfur cluster assembly enzyme (IscU) encodes a ferrous iron binding protein involved in iron-sulfur cluster assembly.
	FBgn0037638	CG8379	N/A
	FBgn0262614	pyd	polychaetoid (pyd) encodes a broadly acting protein that is associated with multiple proteins at the surface and within the cytoskeleton, connecting events between the two.
	FBgn0011966	tRNA:R:85Ca	transfer RNA:Arginine-TCG 2-3
	FBgn0053551	tRNA:R:85Cb	transfer RNA:Arginine-TCG 2-4
	FBgn0264790	CR44020	antisense RNA:CR44020
	FBgn0003015	osk	oskar
	FBgn0037643	skap	Succinyl-coenzyme A synthetase β subunit, ADP-forming (ScsβA) encodes a subunit of the ligase that converts succinate to succinyl-CoA in the TCA/Krebs cycle. It contributes to apoptosis and sperm individualization
	FBgn0037644	CG11964	spliceosome complex B, C and P
	FBgn0037645	CG11966	Transcriptional activator
3R_17087687	FBgn0020493	Dad	Daughters against dpp (Dad) encodes the inhibitory SMAD in the BMP/Dpp pathway. It is involved in growth regulation and developmental patterning.
	FBgn0038473	Ns1	Nucleostemin 1 (Ns1) encodes a GTPase that is enriched in the peripheral granular components of nucleoli of most larval and adult cells. It is required for nucleolar release of the large ribosomal subunit. Depletion of the Ns1 product reduces viability of midgut imaginal island cells and ribosome abundance in polyploid cells
	FBgn0038474	mRpS11	mitochondrial ribosomal protein S11
	FBgn0038475	Keap1	Keap1 (Keap1) encodes a protein that interacts with the product of cnc to regulate the activation of genes by oxidative stress
	FBgn0038476	kuk	kugelkern (kuk) encodes a nuclear envelope protein required for nuclear elongation during cellularization. It shares structural and functional similarities to lamins.
	FBgn0266465	GckIII	Germinal centre kinase III (GckIII) encodes a serine/threonine kinase of the STE20 superfamily. It is the sole Drosophila germinal center kinase type 3 family member. Loss of GckIIII results in tube dilation defects that are specific to the terminal cells of the tracheal system

	FBgn0038478	cal1	chromosome alignment defect 1 (cal1) encodes a protein that is required for maintenance of the epigenetic marks specifying centromeres. It binds to newly synthesized centromere-specific histone H3 variant encoded by cid and recruits it to the centromere by binding to the centromere protein encoded by Cenp-C
	FBgn0014141	cher	cheerio (cher) encodes a dimeric F-actin crosslinking protein of the filamin protein family. It functions to organize the F-actin cytoskeleton in multiple contexts including ovarian germline ring canals, migrating somatic cells, and neuronal growth cones.
3R_25361898	FBgn0039329	CG10669	C2H2 Zinc finger transcription factors
	FBgn0028647	CG11902	C2H2 Zinc finger transcription factors
	FBgn0027376	rha	rha (rha) encodes a protein that is postmeiotically present in spermatid nuclei and might be involved in the chromatin substitution process removing histones from the chromatin during sperm development.
	FBgn0261575	tobi	target of brain insulin
	FBgn0004509	Fur1	Furin 1
	FBgn0039331	ND-49L	NADH dehydrogenase, mitochondrial complex I = NADH: Ubiquinone Oxidoreductase complex subunits
	FBgn0039332	alrm	astrocytic leucine-rich repeat molecule
	FBgn0051437	CG31437	N/A
X_442317	FBgn0011822	pcl	Pepsinogen C-like (Pgcl) encodes a protein predicted to belong to a family of aspartic proteases and controls planar cell polarity. Pgcl mutants result in loss of wing margins.
	FBgn0000137	ase	asense (ase) encodes a transcription factor in the achaete-scute complex. It acts together with other proneural genes in nervous system development, which involves N-mediated lateral inhibition. ase is expressed in the CNS type-I neuroblasts and the PNS sensory organ precursors (SOPs) but not in the proneural clusters that give rise to the SOP via lateral inhibition.
	FBgn0010019	Cyp4g1	Cytochrome P450-4g1 (Cyp4g1) encodes the terminal oxidative decarbonylase in cuticular hydrocarbon biosynthesis within oenocytes.
X_4573308	FBgn0029711	Usf	Usf (Usf) encodes a bHLH family transcription factor that recognizes E-boxes in the Ste promoter.
	FBgn0029712	CG15912	N/A
	FBgn0029713	CG11436	N/A
	FBgn0029714	CG3527	N/A
	FBgn0029715	CG11444	N/A
	FBgn0029716	CG3546	N/A
	FBgn0029717	CG12684	N/A
X_6900105	FBgn0263993	CG43736	N/A

	FBgn0029922	CG14431	N/A
X_7338301	FBgn0029943	Atg5	Autophagy-related 5
	FBgn0029944	Dok	Downstream of kinase (Dok) encodes a membrane-associated protein that functions upstream of the product of Shark to activate Jun kinase signaling during embryonic dorsal closure.
	FBgn0029945	CG18155	Long-chain-fatty-acid--CoA ligase, Malonate--CoA ligase
	FBgn0261848	CG42780	N/A
	FBgn0026144	CBP	sarcoplasmic calcium-binding protein
	FBgn0029946	CG15034	N/A
	FBgn0029947	CG1999	N/A
X_7761153	FBgn0029961	Ir7a	Ionotropic receptor 7a
	FBgn0016041	Tom40	Translocase of outer membrane 40 (Tom40) encodes a protein predicted to be a channel component of the TOM complex, which is involved in protein import into mitochondria. It is predicted to be involved in the cellular response to hypoxia.
	FBgn0085367	CG34338	N/A
	FBgn0052718	CG32718	N/A
X_8478041	FBgn0021767	org-1	optomotor-blind-related-gene-1 (org-1) encodes a T-box transcription factor involved in the combinatorial activation of somatic muscle lineage-specific targets.
	FBgn0052713	CG32713	N/A
	FBgn0030040	CG15347	N/A
	FBgn0023506	Es2	ES2
	FBgn0014032	Sptr	Sepiapterin reductase (Sptr) encodes a sepiapterin reductase involved in the biosynthesis of tetrahydrobiopterin
	FBgn0053223	CG33223	N/A
	FBgn0030041	CG12116	Sepiapterin reductase
	FBgn0014464	Cp7Fa	Chorion protein a at 7F
	FBgn0014465	Cp7Fb	Chorion protein b at 7F (Cp7Fb) belongs to a family of nine Chorion protein genes clustered in two chromosomes (X and 3rd). It is expressed in the follicle cells in a dynamic pattern.
	FBgn0014466	Cp7Fc	Chorion protein c at 7F (Cp7Fc) belongs to a family of nine Chorion protein genes clustered in two chromosomes (X and 3rd). It is expressed in the follicle cells in a dynamic pattern.
	FBgn0000359	Cp36	Chorion protein 36 (Cp36) encodes an abundant structural protein of the eggshell whose expression is enhanced by gene amplification in ovarian follicle cells. It is essential for fertility and for endochorion integrity.
	FBgn0000360	Cp38	Chorion protein 38
	FBgn0264382	CR43834	long non-coding RNA:CR43834
	FBgn0003023	otu	ovarian tumor proteases
	FBgn0053181	CG33181	N/A

	FBgn0028341	Ptpmeg2	Protein tyrosine phosphatase Meg2 (Ptpmeg2) encodes a phosphatase involved in border follicle cell migration.
	FBgn0030148	CG3106	N/A
X_9676328	FBgn0261617	nej	nejire (nej) encodes the transcriptional co-activator CBP. It acetylates several nuclear proteins, including the histone encoded by His3 on K18, K27, and H4 on K8. By regulating gene expression, the product of nej has roles in cell proliferation, cell signaling and differentiation, and in developmental patterning
	FBgn0261522	CR42657	long non-coding RNA:CR42657
	FBgn0264786	CR44016	long non-coding RNA:CR44016
	FBgn0000233	btd	buttonhead (btd) encodes a triple C(2)H(2) zinc finger protein structurally and functionally related to the human transcription factors Sp5, Sp8, and Sp9. The product of btd contributes to embryonic head segmentation, leg development, embryonic hematopoiesis, PNS formation, and maintenance of the functional heterogeneity of brain neural stem cells.
X_19036434	FBgn0264090	CG43759	N/A
	FBgn0085358	Diedel3	Diedel 3
	FBgn0030979	CG14190	N/A