DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD
DASH.HARVARD.EDU

HARVARD LIBRARY
Office for Scholarly Communication

The Role of Environment in Drosophila Behavioral Variability and the Genetics of Life History Variation

Citation

Akhund-Zade, Jamilla. 2020. The Role of Environment in Drosophila Behavioral Variability and the Genetics of Life History Variation. Doctoral dissertation, Harvard University, Graduate School of Arts & Sciences.

Permanent link

https://nrs.harvard.edu/URN-3:HUL.INSTREPOS:37365765

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. Submit a story .

Accessibility

The role of environment in *Drosophila* behavioral variability and the genetics of life history variation

A dissertation presented

by

Jamilla Akhund-Zade

to

The Department of Molecular and Cellular Biology

In partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in the subject of

Biology

Harvard University

Cambridge, Massachusetts

April 2020

© 2020 Jamilla Akhund-Zade

All rights reserved

Dissertation Advisor: Dr. Benjamin L. de Bivort Jamilla Akhund-Zade

The role of environment in *Drosophila* behavioral variability and the genetics of life history variation

Abstract

Phenotypic differences among individuals are ubiquitous, but our understanding of their origins and evolutionary consequences is still not complete. Previous work has identified within-genotype behavioral differences that persist even in constant laboratory environments (behavioral variability). Causes of variability have been investigated, though its response to the environment and its adaptive value are largely uncharacterized. In Chapter I, I tested the hypothesis that *Drosophila melanogaster* behavioral variability could be increased under an enriched environment with more micro-environmental variation. I found that behavioral variability slightly increased under environmental enrichment, but that genotype-by-enrichment effects dominated the behavioral variability response. In Chapter II, I investigated, using a combination of simulations and measurements of wild fly behavior, whether variability in *D. melanogaster* thermal preference reflects bet-hedging, a risk-spreading evolutionary strategy. I found that the seasonal dynamics of mean preference and genetic determination of variability in wild populations supported a pervasive bet-hedging strategy as predicted by modeling. Differential thermal preference heritability across six sites distributed across the continental U.S. supported modeling results that there are regional differences in the adaptive value of bet-hedging. Unlike behavioral variability, life history trait variation is largely genetic in origin - in Chapter III, I found that the genetic basis of offspring number and weight is highly polygenic, implicating genes previously unknown to have roles in life history. Overall, the results of my studies point to

complex relationships among genes and genes and environment in determining phenotypes, as well as an evolutionary role for non-genetic phenotype determination.

Acknowledgements

To begin, I'd like to thank my dissertation advisor, Benjamin de Bivort, for six years of mentorship and support while I completed my doctoral studies. He believed in my abilities as a scientist, even when I had doubts about whether or not I was cut out for graduate school. He taught me how to think critically about scientific problems, keep sight of the bigger picture, and communicate clearly and effectively. His advice and guidance helped me grow as a scientist and I am truly grateful to have been part of his lab.

I would like to thank all the members of the de Bivort lab, past and present, who helped me with my work and made my experience in lab so wonderful. You are the most talented, generous, and fun-loving people I know, and I find that it is a real shame that I have to leave your company.

I'd also like to thank my committee, Dr. Aravinthan Samuel, Dr. Dan Hartl, and Dr. Paul Garrity, for all your helpful advice and feedback over the years.

To my family, thank you for supporting me while I pursued my graduate degree instead of coming home after college. Even though you were far away, I never felt lonely, knowing you were cheering me on.

To Nico – who would have thought that when I met you six years ago, my life would change so much? Thank you for being my rock during the hard times, for cheering me on during the good times, and for taking me on so many adventures, big and small. I love you and I can't wait to see what the future holds for us.

Table of Contents

List of Figures

List of Tables

Introduction

Variation can be observed at all levels of the biological world, from stochastic fluctuations at the cellular level to divergences between populations. Research on variation among organisms spurred on seminal findings in genetics and evolutionary biology, such as the concept of trait inheritance and theories of natural selection and speciation. Even now, trait variation remains an active field of study - we have expanded our tools in model organisms to probe the molecular basis of the relationship between genes and phenotypes and leveraged large population samples and novel statistical approaches to understand the influence of genes and the environment on variation. Throughout this document, I will use variation to refer to genetic and environmental differences (and the resulting trait differences) between individuals, variability to refer to trait differences observed within a genotype and under a controlled environment, and variance to refer to statistical measures of phenotype, genetic, and environmental differences within a population.

Trait variation can arise from several different mechanisms. Genetic differences between individuals have been extensively investigated as a cause of trait variation. In some cases, strong effects on traits can come from alleles of a single gene, as in human genetic disorders[1], rodent coat colors[2,3], and fruit fly pigmentation and morphology[4–6], leading to distinct phenotype classes. More often, observed trait differences are a result of a complex interplay between many genes and the environment. Genetic mapping studies, such as QTL mapping[7–10] and genome-wide association studies (GWAS)[11–13], have been instrumental in determining the genetic basis of complex traits. With the advent of easier and cheaper sequencing technology, genome-wide association studies have been employed to map complex traits across a variety of organisms, including humans[14–16], dogs[17,18], mice[19], fish[20,21], birds[22,23], insects[24,25], plants[26,27], and microbes[28]. In model organisms, GWAS coupled with functional validation of candidate genes

using amorphic alleles[29,30], overexpression, and/or gene knockdowns[31] in genetically tractable organisms has expanded our overall knowledge of the genetic basis of complex traits, often with direct implications on our understanding of human biology and disease.

Drosophila melanogaster, commonly known as the fruit fly or the vinegar fly, has an incredible array of tools[32] to tackle the questions of what genes underlie complex traits. One of the more important resources for genetic mapping in *D. melanogaster* has been the *Drosophila* Genetic Reference Panel (DGRP), a collection of fully-sequenced inbred fly lines established using wild-caught individuals from Raleigh, North Carolina, USA[33]. The DGRP has been a cornerstone for many quantitative genetic analyses - as of 2017, there were 61 different GWA studies using the DGRP collection[25]. The inbred line structure of the DGRP increases additive genetic variance (loci are almost all homozygous within an inbred line, so dominance effects are practically non-existent) and allows for detection of common allelic variants with modest to large effects on traits. Overall, the DGRP studies reveal that complex traits usually have a very polygenic basis, with most of the identified associated variants present in intronic or intergenic regions rather than in coding sequences. Another valuable resource for genetic mapping has been the *Drosophila* Synthetic Population Resource (DSPR)[34]. The DSPR is a large collection of recombinant inbred lines (RILs) that were created through 50 generations of recombination of eight inbred founder lines, followed by inbreeding. Mapping using the DSPR allows for identification of associated quantitative trait loci (QTLs), and putatively causative variants could be identified using the assigned founder genotypes from the RILs at each QTL. Interestingly, when the same complex traits are mapped using the DGRP and the DSPR, the resulting associated variants and QTLs do not clearly overlap between the two panels[35,36]. Differences in mapping approaches can influence what loci are able to be detected - the DSPR can map QTLs consisting of multiple variants, but if the variants under the QTL are of small effect, then a GWAS using the DGRP is unlikely to pick them up due to the stringent genome-wide false

positive correction[37]. In addition to this, the DSPR captures global genetic variation in *D. melanogaster* with eight founder lines, while the DGRP is a single sample of genetic variation within one population. Therefore, it is likely that associated variants found in one panel are simply not segregating in the other panel or are present at a low allele frequency and do not pass the statistical threshold to be detected[38]. The non-overlapping results from the DSPR and DGRP mapping, as well as the lack of overlap between top hits of multiple DGRP studies on the same trait[35], support the idea that complex traits are determined by a large number of variants of small effect.

The genetics of life history traits are of particular interest, as they are the most directly connected to organismal fitness[39]. Examples of life history traits include fecundity, longevity, and body size. Given the genetic toolkit of *D. melanogaster*, as well as its short development time and large numbers of offspring, it has been a popular model organism for studying life history trait variation. Life history traits in *D. melanogaster* show a large contribution of genetic variance. For example, in longevity, mutant screens[40], QTL analysis[40], and GWAS using the DGRP[41] reveal a highly polygenic basis. Genes involved in stress response, insulin signaling, growth, and metabolism all contribute to variation in longevity. Mutations that decrease gene expression in the insulin signaling pathway produce a longer lifespan, and differences in the allele frequency of a deletion at an insulin receptor coincide with differences in average lifespan in natural populations. DGRP lines show heritable variation in embryonic development time[42] with a direct impact on fitness - decreased development time increased egg-to-adult viability. Once again, a variety of genes determine development time, most with known roles in the cell cycle, cell signaling, and development. In addition, variants associated with development time are found in regulatory regions, showing that for traits closely related to fitness, changing gene expression is likely a more common approach to regulating trait variation than changing the coding sequence. Fecundity[43], body size[44], nutritional indices[45], and immunity[46] also have a

polygenic basis - with traits that are closely related to fitness, it would not be surprising that almost every gene in the genome would contribute in some small way (an omnigenic basis)[47]. In Chapter III, I use the DGRP to examine the genetic basis of a trade-off between offspring number and body weight in a resource-limited scenario in order to see what complement of genes plays a role post-egg laying to determine offspring viability and robustness.

In addition to genetic variation, differential environmental exposure can play a large role in trait variation among individuals. Trait variation or phenotypic variance (V_P) is a combination of genetic variance (V_G), environmental variance (V_E), and genotype-by-environment variance (V_{GxE}). If a substantial proportion of phenotypic variance comes from V_E, the trait is regarded as phenotypically plastic[48]. A way to evaluate the plasticity of a trait is to look at its reaction norm, the phenotypic response of a genotype under different environmental treatments. A plastic trait will have significant responses to the environmental treatments, whereas a non-plastic trait will hold the same value over all treatments. Variable reaction norms across genotypes show that there is a V_{GxE} component to the trait. Plastic traits are varied in scope - traits can be considered active or passive, reversible or irreversible, and, related to the former, can be caused by environmental cues acting at different points in development e.g. embryonic vs. adult stages[48,49]. Active plastic response is thought to occur in response to an environmental cue in order to anticipate a changing environment, such as increase in predators, a change in resource abundance, or seasonal changes. A passive plastic response comes from a direct influence of the environment on the physiology of the organism, without any anticipation of changing circumstances, e.g. poor nutrition leading to small size. The environment can influence an organism at multiple stages of its life; plastic responses during development are likely to be irreversible commitments to a particular trait, whereas responses during the adult stages are likely to be flexible and reversible e.g. metabolic switches, behavioral changes, and

acclimation[50]. The environment can have a large impact on phenotype - there is evidence to support that for certain traits, environmental changes can phenocopy genetic mutations[51].

D. melanogaster shows developmental plasticity in morphological traits[52]. When flies were raised at 19ºC, they had a longer thorax and wings (due to a larger body size) and also a higher ratio of wing to thorax length. Bristle number was higher in flies reared at 25ºC. The plastic response was heritable in all three traits and correlated across the thorax and wing length measurements. In a sister species, *Drosophila simulans*, morphological plasticity was evaluated in a population sampled from the wild (exposed to natural seasonal temperature fluctuations) and a wild-derived lab population exposed to comparable temperature regimes in the laboratory. The reaction norms were qualitatively similar, though less pronounced in the populations sampled from the wild (perhaps due to the conflicting influences of food availability and seasonal temperatures in the wild on the trait plasticity)[53]. In neuroanatomical traits, *D. melanogaster* shows plasticity in brain size in response to certain environmental pressures[54]. Mushroom body size decreases when larva are heat stressed, but is unaffected by larval crowding, yet antenna lobe, olfactory lobe, and central complex size decreases with increased larval crowding. An enriched environment for adults did not affect olfactory associative learning or brain structure size. Plastic responses in the brain are not always consistent among studies. Previous work found an increase in the mushroom body calyx of females with a longer developmental time when subjected to higher larval densities, and an increase in brain size under enriched living conditions[55]. It is likely that there are genotype effects on brain size plasticity that contribute to the inconsistency in results - previous work using the DGRP has found that there is a genetic basis for variation in mushroom body size[56], and that some of the functionally validated genes play a role in mushroom body plasticity in the adult. In agreement with potential genotype-by-environment interactions in brain plasticity, genotype-by-environment interactions have been shown to be present in behaviors, such as male aggression[57]. Male flies

from the DGRP lines were reared post-eclosion either in mixed-sex housing or in isolation and then assayed for aggressive behaviors towards other male flies. The results show that social isolation had variable effects on aggressive behaviors - some lines showed a decrease in aggression after isolation, some showed an increase, and some showed no change. Annotated genes associated with variation in genotype-by-environment interactions in male aggression are involved in nervous system growth and morphogenesis, as well as sleep and memory. Overall results from *D. melanogaster* show both developmental and adult plasticity across a variety of traits, though the effects of the interaction of environment, genotype, and specific trait can be complex.

The studies on plasticity mentioned above focused on controlled large-scale environmental changes, such as different diets, social exposure, or stressors, to measure the change in the mean trait response. Yet, there is also an uncontrolled component of environmental variance that exists within the treatment regime. The two components can be thought of as macro-environmental variance and micro-environmental variance, respectively[58]. In laboratory studies, macro-environmental variance is imposed by the researcher as experimental treatments, and micro-environmental variance is the component responsible for within-treatment phenotypic variability (phenotypic variance observed within a genotype and under controlled environmental conditions). Under natural conditions, it would be more appropriate to think of environmental variance as existing on a continuum from macro- to micro-scale. Micro-environmental variance is thought to arise from stochastic differences among individuals, ranging from the internal, such as epigenetic[59] or neuron wiring differences[60], to the external, such as micro-habitat differences in developmental temperature or food availability. Since micro-environmental variance by definition comes from uncontrolled and stochastic influences, it is difficult to rank the possible factors in order of importance for trait variability - it is also very likely that the external and internal factors are interconnected e.g. small differences in developmental nutrition or

temperature could manifest in epigenetic differences, which in turn produces phenotype differences between individuals. The degree of trait variability which a particular genotype shows under a controlled environment has been termed its micro-environmental plasticity or intra-genotypic variability. Fascinatingly, there are genotype differences in micro-environmental plasticity[61–64]. A GWAS using the DGRP mapped the genetic basis of micro-environmental plasticity in starvation resistance, chill coma recovery time, and startle response[63]. Micro-environmental plasticity in the traits was highly heritable, comparable to the heritability of the trait means, and the genetic basis was trait-specific, meaning that there was no genetic correlation in micro-environmental plasticity among the three traits measured. Examination of variants associated with micro-environmental plasticity in the three traits showed a polygenic basis, with no presence of a single variant of large effect (though the power of the study was diminished due to the noise inherent in measures of micro-environmental plasticity). Studies using the DGRP combined with studies on plasticity show that *D. melanogaster* is a powerful model organism for teasing apart the complex interactions between genotype, macro-, and micro-environmental plasticity. In Chapter I, I consider the effects of enriched environments during the developmental and adult stages of *D. melanogoster* on genotype- and behavior-specific micro-environmental plasticity.

While the proximal causes of trait variation can be broadly agreed upon, the adaptive value of trait variation is not always clear cut. Under what conditions is increased phenotypic variance maintained in the population? Should the variation come from plastic responses, genetic determination, or a mix of both? What are the costs and limits to the different strategies for establishing trait variation?

An initial proposal for the maintenance of genetic variance is that it arises from a balance between variation created by mutations and elimination of variation by purifying selection[65].

While some studies have shown that observed levels of genetic variance can be maintained by the mutation-selection balance (though with some restrictive assumptions)[65,66], there are other scenarios where high levels of genetic variance can be maintained. Balancing selection at loci and fluctuating spatial and/or temporal environments[67] can also maintain genetic variance. Balancing selection can maintain genetic variance through overdominance, or selection for high-fitness heterozygous genotypes, or through negative frequency-dependent selection, where the rarer allele has a fitness advantage. To maintain genetic variance under fluctuating environmental pressures, genotype-by-environment effects on fitness must exist i.e. an allele that has a positive effect on fitness in one environment has a negative (or no) effect on fitness in another. As discussed above, there are many examples of genotype-by-environment effects in traits that lend support for this scenario. Just as fluctuating environments can cause maintenance of genetic variance, the level of genetic variance may also be important in the persistence of populations under fluctuating environments[68]. In an environment with long-period cyclical fluctuations, increased genetic variance allows populations to track the phenotypic optimum more closely. The advantage of genetic variance diminishes as the cyclical fluctuations get more rapid and the strength of selection increases. Modeling even suggests that under rapid periodic fluctuations in the phenotypic optimum, it is advantageous to dampen the genetic and phenotypic response to selection through a modifier locus[69] so that previous responses to selection do not negatively impact future responses.

Since phenotypic variance does not only depend on genetic variance, we can also consider the adaptive value of environmental variance. Bet-hedging is a non-genetic strategy for producing phenotypic variance as a way to cope with the risk of environmental fluctuations[70–73]. The driving force behind the advantage of a bet-hedging strategy is that, over the long-term, a reduction in mean fitness is overcome by a reduction in the fitness variance[72,73]. In essence, a bet-hedging genotype would not have optimal fitness under every environment, but it would perform

consistently well over time, thereby increasing its geometric mean fitness at the expense of arithmetic mean fitness. There are two types of bet-hedging strategy - conservative and diversifying[72]. Conservative bet-hedging is a way to deal with environmental uncertainty that avoids risk altogether, for example, always migrating early in the year to avoid a potentially early winter. A diversifying bet-hedging strategy mitigates risk by investing in multiple phenotypes for one genotype, a classic example being different seed germination timing to prevent complete failure to reproduce in a bad year[70]. Diversifying bet-hedging (just referred to as bet-hedging from now on) is of particular interest here, because of its direct parallels to observed micro-environmental plasticity[74]. The fluctuating selection pressures that favor bet-hedging differ from those that favor maintenance of genetic variance. A bet-hedging strategy is advantageous when the variance in the phenotypic optimum across time outcompetes the purifying selection pressure against deviations from the optimum at a particular time[75]. Overall, the common thread of many modeling studies is that bet-hedging requires a higher and more rapid fluctuations of the environment to be a favorable strategy[75–79], whereas genetic variance is favored under slower fluctuations[79]. In addition to the extensive theoretical discussion on bet-hedging, there have been empirical studies on putative bet-hedging traits across many organisms[73], though evidence for bet-hedging across the studies is variable, and with few studies looking at behavior as a possible bet-hedging trait. In Chapter II, I examine empirical evidence for bet-hedging in thermal preference of *Drosophila melanogaster*, as well as whether bet-hedging in thermal preference is dependent on climate.

Future directions of study on trait variation still remain quite open. Understanding trait variation is less tied to a purely genetic approach, but rather a more holistic approach that incorporates the effect of environment at both large and small scales and its interaction with genotype. Previously unexplained components of trait variation have now become fascinating avenues of new research. The research presented in the following chapters will examine the interactions

between genetic and environmental influences on variation in behavior and life history traits in *Drosophila melanogaster*.

References

1. Chong, J. X. *et al.* The Genetic Basis of Mendelian Phenotypes: Discoveries, Challenges, and Opportunities. *Am. J. Hum. Genet.* **97**, 199–215 (2015).

2. Silvers, W. K. *The coat colors of mice: a model for mammalian gene action and interaction*. (Springer-Verlag, 1979).

3. Miltenberger, R. J., Mynatt, R. L., Wilkinson, J. E. & Woychik, R. P. The Role of the agouti Gene in the Yellow Obese Syndrome. *J. Nutr.* **127**, 1902S-1907S (1997).

4. Wittkopp, P. J., True, J. R. & Carroll, S. B. Reciprocal functions of the Drosophila yellow and ebony proteins in the development and evolution of pigment patterns. *Dev. Camb. Engl.* **129**, 1849–1858 (2002).

5. Hurd, T. R., Liang, F.-X. & Lehmann, R. Curly Encodes Dual Oxidase, Which Acts with Heme Peroxidase Curly Su to Shape the Adult Drosophila Wing. *PLoS Genet.* **11**, (2015).

6. Hammonds, A. S. & Fristrom, J. W. Mutational Analysis of Stubble-stubbloid Gene Structure and Function in Drosophila Leg and Bristle Morphogenesis. *Genetics* **172**, 1577–1593 (2006).

7. Paterson, A. H. *et al.* Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. *Nature* **335**, 721–726 (1988).

8. Zeng, Z. B. Precision mapping of quantitative trait loci. *Genetics* **136**, 1457–1468 (1994).

9. Lander, E. S. & Botstein, D. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**, 185–199 (1989).

10. Li, H., Hearne, S., Bänziger, M., Li, Z. & Wang, J. Statistical properties of QTL linkage mapping in biparental genetic populations. *Heredity* **105**, 257–267 (2010).

11. Bush, W. S. & Moore, J. H. Chapter 11: Genome-Wide Association Studies. *PLOS Comput. Biol.* **8**, e1002822 (2012).

12. Goddard, M. E., Kemper, K. E., MacLeod, I. M., Chamberlain, A. J. & Hayes, B. J. Genetics of complex traits: prediction of phenotype, identification of causal polymorphisms and genetic architecture. *Proc. R. Soc. B Biol. Sci.* **283**, 20160569 (2016).

13. Visscher, P. M. *et al.* 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am. J. Hum. Genet.* **101**, 5–22 (2017).

14. Xue, A. *et al.* Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. *Nat. Commun.* **9**, 1–14 (2018).

15. Okbay, A. *et al.* Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* **533**, 539–542 (2016).

16. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights

from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).

17. MacLean, E. L., Snyder-Mackler, N., vonHoldt, B. M. & Serpell, J. A. Highly heritable and functionally relevant breed differences in dog behaviour. *Proc. R. Soc. B Biol. Sci.* **286**, 20190716 (2019).

18. Hayward, J. J. *et al.* Complex disease and phenotype mapping in the domestic dog. *Nat. Commun.* **7**, 1–11 (2016).

19. Flint, J. & Eskin, E. Genome-wide association studies in mice. *Nat. Rev. Genet.* **13**, 807–817 (2012).

20. Gutierrez, A. P., Yáñez, J. M., Fukui, S., Swift, B. & Davidson, W. S. Genome-Wide Association Study (GWAS) for Growth Rate and Age at Sexual Maturation in Atlantic Salmon (Salmo salar). *PLOS ONE* **10**, e0119730 (2015).

21. Li, N. *et al.* Identification of novel genes significantly affecting growth in catfish through GWAS analysis. *Mol. Genet. Genomics* **293**, 587–599 (2018).

22. Silva, C. N. S. *et al.* Insights into the genetic architecture of morphological traits in two passerine bird species. *Heredity* **119**, 197–205 (2017).

23. Husby, A. *et al.* Genome-wide association mapping in a wild avian population identifies a link between genetic and phenotypic variation in a life history trait. *Proc. R. Soc. B Biol. Sci.* **282**, 20150156 (2015).

24. Weetman, D. *et al.* Candidate-gene based GWAS identifies reproducible DNA markers for metabolic pyrethroid resistance from standing genetic variation in East African Anopheles gambiae. *Sci. Rep.* **8**, 1–12 (2018).

25. Mackay, T. F. C. & Huang, W. Charting the genotype–phenotype map: lessons from the Drosophila melanogaster Genetic Reference Panel. *WIREs Dev. Biol.* **7**, e289 (2018).

26. Ando, K. *et al.* Genome-wide associations for multiple pest resistances in a Northwestern United States elite spring wheat panel. *PLOS ONE* **13**, e0191305 (2018).

27. Brachi, B., Morris, G. P. & Borevitz, J. O. Genome-wide association studies in plants: the missing heritability is in the field. *Genome Biol.* **12**, 232 (2011).

28. Power, R. A., Parkhill, J. & de Oliveira, T. Microbial genome-wide association studies: lessons from human GWAS. *Nat. Rev. Genet.* **18**, 41–50 (2017).

29. Thibault, S. T. *et al.* A complementary transposon tool kit for Drosophila melanogaster using P and piggyBac. *Nat. Genet.* **36**, 283–287 (2004).

30. Bellen, H. J. *et al.* The Drosophila Gene Disruption Project: Progress Using Transposons With Distinctive Site Specificities. *Genetics* **188**, 731–743 (2011).

31. Zirin, J. *et al.* Large-Scale Transgenic Drosophila Resource Collections for Loss- and Gain-of-Function Studies. *Genetics* (2020) doi:10.1534/genetics.119.302964.

32. Hales, K. G., Korey, C. A., Larracuente, A. M. & Roberts, D. M. Genetics on the Fly: A

Primer on the Drosophila Model System. *Genetics* **201**, 815–842 (2015).

33. Mackay, T. F. C. *et al.* The *Drosophila melanogaster* Genetic Reference Panel. *Nature* **482**, 173 (2012).

34. King, E. G. *et al.* Genetic dissection of a model complex trait using the Drosophila Synthetic Population Resource. *Genome Res.* **22**, 1558–1566 (2012).

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

36. Najarro, M. A., Hackett, J. L. & Macdonald, S. J. Loci Contributing to Boric Acid Toxicity in Two Reference Populations of Drosophila melanogaster. *G3 Bethesda* **7**, 1631–1641 (2017).

37. Long, A. D., Macdonald, S. J. & King, E. G. Dissecting complex traits using the Drosophila Synthetic Population Resource. *Trends Genet.* **30**, 488–495 (2014).

38. King, E. G. & Long, A. D. The Beavis Effect in Next-Generation Mapping Panels in Drosophila melanogaster. *G3 Bethesda* **7**, 1643–1652 (2017).

39. Flatt, T. Life history Evolution and the Genetics of Fitness Components in Drosophila melanogaster. *Genetics* **214**, 3–48 (2020).

40. Paaby, A. B. & Schmidt, P. S. Dissecting the genetics of longevity in Drosophila melanogaster. *Fly (Austin)* **3**, 29–38 (2009).

41. Ivanov, D. K. *et al.* Longevity GWAS Using the Drosophila Genetic Reference Panel. *J. Gerontol. Ser. A* **70**, 1470–1478 (2015).

42. Horváth, B., Betancourt, A. J. & Kalinka, A. T. A novel method for quantifying the rate of embryogenesis uncovers considerable genetic variation for the duration of embryonic development in Drosophila melanogaster. *BMC Evol. Biol.* **16**, 200 (2016).

43. Durham, M. F., Magwire, M. M., Stone, E. A. & Leips, J. Genome-wide analysis in *Drosophila* reveals age-specific effects of SNPs on fitness traits. *Nat. Commun.* **5**, 1–8 (2014).

44. Vonesch, S. C., Lamparter, D., Mackay, T. F. C., Bergmann, S. & Hafen, E. Genome-Wide Analysis Reveals Novel Regulators of Growth in Drosophila melanogaster. *PLOS Genet.* **12**, e1005616 (2016).

45. Unckless, R. L., Rottschaefer, S. M. & Lazzaro, B. P. A Genome-Wide Association Study for Nutritional Indices in Drosophila. *G3 Genes Genomes Genet.* **5**, 417–425 (2015).

46. Unckless, R. L., Rottschaefer, S. M. & Lazzaro, B. P. The Complex Contributions of Genetics and Nutrition to Immunity in Drosophila melanogaster. *PLOS Genet.* **11**, e1005030 (2015).

47. Boyle, E. A., Li, Y. I. & Pritchard, J. K. An expanded view of complex traits: from polygenic to

omnigenic. *Cell* **169**, 1177–1186 (2017).

48. Fusco, G. & Minelli, A. Phenotypic plasticity in development and evolution: facts and concepts. *Philos. Trans. R. Soc. B Biol. Sci.* **365**, 547–556 (2010).

49. Forsman, A. Rethinking phenotypic plasticity and its consequences for individuals, populations and species. *Heredity* **115**, 276–284 (2015).

50. Piersma, T. & Drent, J. Phenotypic flexibility and the evolution of organismal design. *Trends Ecol. Evol.* **18**, 228–233 (2003).

51. West-Eberhard, M. J. *Developmental Plasticity and Evolution*. (OUP U.S.A. , 2003).

52. Scheiner, S. M., Caplan, R. L. & Lyman, R. F. The genetics of phenotypic plasticity. III. Genetic correlations and fluctuating asymmetries. *J. Evol. Biol.* **4**, 51–68 (2002).

53. Negoua, H., Chakir, M., David, J. R. & Capy, P. Climatic adaptation in Drosophila: phenotypic plasticity of morphological traits along a seasonal cycle. *Ann. Société Entomol. Fr. NS* **55**, 48–60 (2019).

54. Wang, X., Amei, A., Belle, J. S. de & Roberts, S. P. Environmental effects on Drosophila brain development and learning. *J. Exp. Biol.* **221**, jeb169375 (2018).

55. Heisenberg, M., Heusipp, M. & Wanke, C. Structural plasticity in the *Drosophila* brain. *J. Neurosci.* **15**, 1951–1960 (1995).

56. Zwarts, L. *et al.* The genetic basis of natural variation in mushroom body size in *Drosophila melanogaster*. *Nat. Commun.* **6**, 10115 (2015).

57. Rohde, P. D., Gaertner, B., Ward, K., Sørensen, P. & Mackay, T. F. C. Genomic Analysis of Genotype-by-Social Environment Interaction for Drosophila melanogaster Aggressive Behavior. *Genetics* **206**, 1969–1984 (2017).

58. Debat, V. & David, P. Mapping phenotypes: canalization, plasticity and developmental stability. *Trends Ecol. Evol.* **16**, 555–561 (2001).

59. Wong, A. H. C., Gottesman, I. I. & Petronis, A. Phenotypic differences in genetically identical organisms: the epigenetic perspective. *Hum. Mol. Genet.* **14**, R11–R18 (2005).

60. Honegger, K. & de Bivort, B. Stochasticity, individuality and behavior. *Curr. Biol.* **28**, R8–R12 (2018).

61. Dworkin, I. A Study of canalization and developmental stability in the sternopleural bristle system of *Drosophila melanogaster*. *Evolution* **59**, 1500–1509 (2005).

62. Hall, M. C., Dworkin, I., Ungerer, M. C. & Purugganan, M. Genetics of microenvironmental canalization in Arabidopsis thaliana. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 13717–13722 (2007).

63. Morgante, F., Sørensen, P., Sorensen, D. A., Maltecca, C. & Mackay, T. F. C. Genetic Architecture of Micro-Environmental Plasticity in *Drosophila melanogaster*. *Sci. Rep.* **5**, 9785 (2015).

64. Stamps, J. A., Saltz, J. B. & Krishnan, V. V. Genotypic differences in behavioural entropy: unpredictable genotypes are composed of unpredictable individuals. *Anim. Behav.* **86**, 641–649 (2013).

65. Burger, R., Wagner, G. P. & Stettinger, F. How Much Heritable Variation Can be Maintained in Finite Populations by Mutation-Selection Balance? *Evolution* **43**, 1748–1766 (1989).

66. Lynch, M., Latta, L., Hicks, J. & Giorgianni, M. Mutation, Selection, and the Maintenance of life history Variation in a Natural Population. *Evolution* **52**, 727–733 (1998).

67. Turelli, M. & Barton, N. H. Polygenic variation maintained by balancing selection: pleiotropy, sex-dependent allelic effects and G x E interactions. *Genetics* **166**, 1053–1079 (2004).

68. Lande, R. & Shannon, S. The Role of Genetic Variation in Adaptation and Population Persistence in a Changing Environment. *Evolution* **50**, 434–437 (1996).

69. Kawecki, T. J. The Evolution of Genetic Canalization Under Fluctuating Selection. *Evolution* **54**, 1–12 (2000).

70. Cohen, D. Optimizing reproduction in a randomly varying environment. *J. Theor. Biol.* **12**, 119–129 (1966).

71. Slatkin, M. Hedging one's evolutionary bets. *Nature* **250**, 704–705 (1974).

72. Hopper, K. R. Risk-spreading and bet-hedging in insect population biology. *Annu. Rev. Entomol.* **44**, 535–560 (1999).

73. Simons, A. M. Modes of response to environmental change and the elusive empirical evidence for bet hedging. *Proc. R. Soc. B Biol. Sci.* **278**, 1601–1609 (2011).

74. Simons, A. M. & Johnston, M. O. Developmental Instability as a Bet-Hedging Strategy. *Oikos* **80**, 401–406 (1997).

75. Bull, J. J. Evolution of Phenotypic Variance. *Evolution* **41**, 303–315 (1987).

76. Haccou, P. & Iwasa, Y. Optimal mixed strategies in stochastic environments. *Optim. Mix. Strateg. Stoch. Environ.* (1995).

77. Svardal, H., Rueffler, C. & Hermisson, J. Comparing environmental and genetic variance as adaptive response to fluctuating selection. *Evol. Int. J. Org. Evol.* **65**, 2492–2513 (2011).

78. Scheiner, S. M. Bet-hedging as a complex interaction among developmental instability, environmental heterogeneity, dispersal, and life history strategy. *Ecol. Evol.* **4**, 505–515 (2014).

79. Tufto, J. Genetic evolution, plasticity, and bet-hedging as adaptive responses to temporally autocorrelated fluctuating selection: A quantitative genetic model. *Evolution* **69**, 2034–2049 (2015).

Chapter I - The effect of environmental enrichment on behavioral variability depends on genotype, behavior, and type of enrichment

I would like to acknowledge the work of all of those who contributed to this project. Sandra Ho and Chelsea O'Leary collected the data. Benjamin L. de Bivort conceptualized the experimental design. I carried out the data analysis and interpretation. I am grateful to Kyle Honegger for discussions of Bayesian inference. I also thank Carolyn Elya for moral support when the Hamiltonian Monte Carlo sampler misbehaved. All of the measures of individual behavior and all analysis scripts are publicly available at https://zenodo.org/record/2573158. They are also hosted on our lab website at http://lab.debivort.org/enrichment.

Abstract

Non-genetic individuality in behavior, also termed intragenotypic variability, has been observed across many different organisms. A potential cause of intragenotypic variability is sensitivity to minute environmental differences during development, even as major environmental parameters are kept constant. Animal enrichment paradigms often include the addition of environmental diversity, whether in the form of social interaction, novel objects, or exploratory opportunities. Enrichment could plausibly affect intragenotypic variability in opposing ways: it could cause an increase in variability due to the increase in microenvironmental variation, or a decrease in variability due to elimination of aberrant behavior as animals are taken out of impoverished laboratory conditions. In order to test our hypothesis, we assayed five isogenic *Drosophila melanogaster* lines raised in control and mild enrichment conditions, and one isogenic line under both mild and intense enrichment conditions. We compared the mean and variability of six behavioral metrics between our enriched fly populations and the laboratory housing control. We found that enrichment often caused a small increase in variability across most of our behaviors, but that the ultimate effect of enrichment on both behavioral means and variabilities was highly dependent on genotype and its interaction with the particular enrichment treatment. Our results support previous work on enrichment that presents a highly variable picture of its effects on both behavior and physiology.

Introduction

Stable behavioral differences among conspecifics are seen in a wide array of species. These differences, caused (definitionally) by some combination of genetic and environmental factors, are commonly referred to as individuality[1–3]. Yet, even after experimentally homogenizing genotype and environment, individuality still persists [4], often undiminished or even increased[5,6]. Multiple studies across different organisms demonstrate non-genetic individuality, which we refer to as intragenotypic variability (fruit flies:[5,7]; pea aphid:[8]; nematodes:[9]; fish:[10]; crayfish:[11]; mice:[12,13]).

This intragenotypic variability may originate in sensitivity to stochastic microenvironmental effects that persist even when large-scale differences in environment across individuals are removed[14–16]. Along these lines, environmental causes of phenotypic differences can be decomposed into deterministic (macro) and stochastic (internal or micro) aspects[16,17]. Examples of macroenvironmental effects are different levels of fertilizer or different temperatures across treatments. Examples of microenvironmental effects include whether an individual animal ate more food in the morning or evening, or (in the case of flies) whether they pupated on the plastic vial or food media surface. Generally, microenvironmental effects exist within a treatment regime[14] and are hard to measure[15]. For individuals of the same genotype raised in a homogenous experimental environment, trait differences would be primarily due to microenvironmental effects, and the propensity to this variation is known as microenvironmental plasticity[18]. For a given trait, it would seem that intragenotypic variability would be maladaptive since some individuals are far from the trait optimum. Yet, in unpredictable and/or fluctuating environments, having intragenotypic variability can be advantageous[19]. In such so-called diversifying bet-hedging strategies, variability can protect against sudden environmental changes, by increasing the likelihood that at any time a subset of the population has high fitness[20,21].

Many studies have focused on characterizing the intragenotypic variability in morphological, physiological, and behavioral traits[5,7,10,12,18,22–29]. In *Drosophila melanogaster*, intragenotypic variability was found in chill coma recovery time, starvation resistance, sternopleural bristles, wing traits, and neuronal morphology in the larval ventral nerve cord[18,24,25,27,28]. Morphological variations present in the ventral nerve cord and optic lobes are of particular note as they respectively correlate with the timing of flight initiation and visually-guided locomotor biases, providing a link between morphological and behavioral intragenotypic variability[27,30]. Our research has identified intragenotypic variability in isogenic lines of *D. melanogaster* for turning bias, phototaxis, and thermotaxis[5,7,26]. Outside of flies, intragenotypic variability in behavior has been studied in inbred mice and clonal fish (Amazon molly), with mice showing variation in exploratory behavior and fish showing variation in activity[10,12]. If these examples of intragenotypic variability have their basis in microenvironmental differences, it may be hard to attribute the behavioral outcomes of specific individuals to their micro-causal underpinnings. It is, however, possible to test whether changing in the degree of microenvironmental variation predicts changes in the amount of intragenotypic variability.

As most lab organisms are already raised in heavily standardized environments where microenvironmental variation is minimized, it is feasible to increase microenvironmental variation and examine the effects on behavior. “Enrichment” treatments include a variety of different modifications to regular laboratory housing, such as opportunities for exercise, novel object interaction, and socialization[31]. Enrichment may add microenvironmental variation to a particular treatment, potentially affecting both the mean and variance of phenotypic traits[32]. Typically, enrichment treatments are hypothesized to more closely match an animal's natural habitat, increasing mean well-being and cognition (while perhaps increasing intragenotypic variability). For mice and rats, enrichment has been shown to enhance mean gliogenesis,

neurogenesis, and synapse formation in the cortex, hippocampus, and cerebellum, leading to improved memory and cognition[31,33–36]. While there are some consistent enrichment effects in rats and mice, physiological and behavioral strain differences in responses to enrichment have been observed[37–39]. There is conflicting evidence for the effect of enrichment on brain size in fish, with enrichment having no effect on three-spined sticklebacks, but decreasing brain size in eastern mosquitofish[40,41]. Early studies in *D. melanogaster* have found that changing the social milieu affects the size of brain structures, with social isolation leading to decreased sizes and numbers of Kenyon cell fibers[42–44]. In addition, social isolation in *D. melanogaster* leads to faster cancer progression, suggesting that a stimulating social environment buffers against stresses[45]. In crickets, mushroom body neurogenesis is higher in enriched environments with complex visual, olfactory, and auditory stimuli as compared to impoverished environments[46]. On the other hand, enriched olfactory environments did not change mushroom body calyx size or affect odor learning in *D. melanogaster*[47].

The effect of enrichment on trait variability has been primarily studied in mice and rats and focused on understanding whether enriched rearing and housing conditions would decrease the statistical power to detect treatment effects by increasing within-sample variance[38]. The evidence presented from behavioral and physiological studies has been conflicting - studies have shown that enrichment can increase, decrease, or have no effect on variability depending on the trait in question[38,39,48–51]. A recent study by Körholz et al. chose to focus more directly on whether intragenotypic variability in behavior and brain plasticity is influenced by the diversity of experiences that results from an enriched environment. They found that enrichment increases variation in specific domains - mice from enriched environments showed higher variation in exploratory behavior (object interaction times, habituation, but not locomotion), adult neurogenesis, and motor cortex thickness[32]. They attributed this increase in variation directly to

the diversity of experiences or diversity of microenvironments that individuals could explore in an enriched environment.

Given the conflicting evidence for the effects of enrichment on variability, we propose two hypotheses for how enrichment may influence variation in traits. The first hypothesis is that enrichment introduces microenvironmental differences which in turn increase trait differences through microenvironmental plasticity. Our second hypothesis is that by more closely matching natural conditions to which organisms are adapted, enrichment increases the robustness of development and somatic maintenance, with a corresponding reduction in variation (due to the removal of aberrant phenotypes that may appear in impoverished laboratory conditions). Even though they predict opposite outcomes, both of these hypotheses are intuitive, and have some support in the literature. We chose to test them by measuring intragenotypic variability in *D. melanogaster* under control and enriched treatments. This species is a good model system for this work because of the ease of rearing large experimental groups from isogenic lines, and its suitability for automated behavioral phenotyping. We were also interested in testing whether the observed effects of enrichment were dependent on genotype. We measured a variety of behavioral metrics associated with spontaneous locomotion and phototaxis[5–7] in one isogenic line across two enrichment treatments and five isogenic lines in a single enrichment treatment to examine the effects of enrichment and the interaction of genotype and enrichment on behavioral variability. We found that while enrichment often caused a small increase in intragenotypic variability, the predominant determinants of behavioral means and variabilities were genotype and its interaction with the particular enrichment treatment.

Results

Intragenotypic variability is evident in locomotor and phototactic behaviors

In order to measure intragenotypic variability, we employed two automated assays (which measure spontaneous locomotion and locomotor responses to light) to rapidly collect many behavioral observations from many individual flies. We first confirmed that intragenotypic variability was present in a standard lab wild-type strain, Canton-S, in left-right turn bias and light-choice probability (Figure 1.1). Indeed, the observed distributions of these measures were significantly broader ($p < 1E-3$ by bootstrapping, χ^2, and Kolmogorov-Smirnov tests) than expected under null models in which all individuals behaved identically, i.e., sequences of behavior drawn from identical distributions (see Methods).

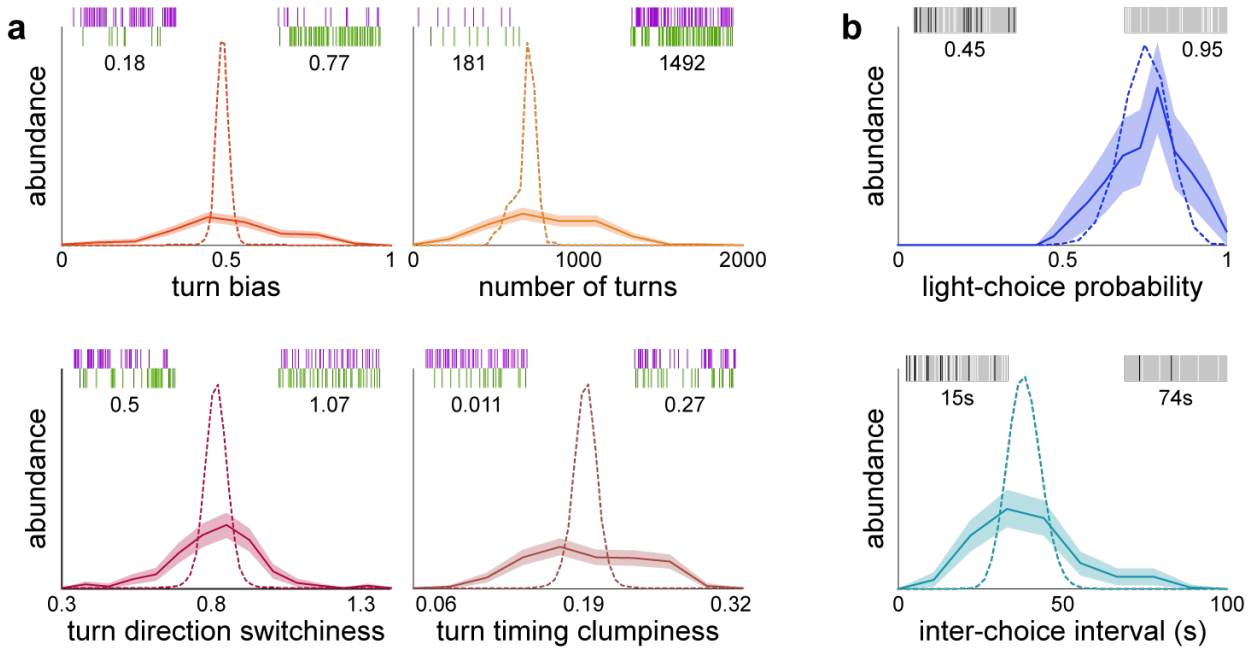


Figure 1.1 - Observed and null hypothesis distributions of Y-maze and FlyVac behavioral measures for Canton-S (wild type) flies.
Dotted lines represent the distributions expected under null hypotheses in which all individuals exhibit behaviors drawn from identical distributions. The solid line represents the observed distribution, with the shaded region representing +/- 1 standard error of the distribution, as estimated by bootstrap resampling. Insets show 10 minutes of representative data of the original behavior traces of extreme individuals and the corresponding value of that metric. **a)** Metrics from the Y-maze assay: turn bias is the fraction of turns made to the right, number of turns is the number of left-right choices made in the 2 hour test, turn direction switchiness is a turn bias-normalized measure of the mutual information between successive turns (for higher values, left turns are more predictive of subsequent left turns and vice versa), and turn timing clumpiness is a normalized measure of the irregularity of turns (the mean absolute deviation of the inter-turn intervals divided by the mean inter-turn interval). Purple ticks represent left turns and green represent right turns. **b)** Metrics from the FlyVac phototaxis assay: light-choice probability is the fraction of choices toward light, inter-choice interval is the mean time between choices. White ticks indicate a light choice, dark ticks show a dark choice, and shaded areas represent regions of time where no choice is made. 151 flies were analyzed for Y-maze behaviors, and 175 flies for FlyVac behaviors.

We next asked if there is evidence of intragenotypic variability in other measurements taken while measuring turn bias and phototactic preference in these assays. As with turn bias, the observed distributions of number of turns, turn switchiness, and turn clumpiness were significantly broader than expected under null models in which all flies behaved identically (Figure 1.1a). Using FlyVac data, we observed that the distribution across flies of the average interval between phototactic choices also was broader than expected if all flies were behaving identically (Figure 1.1b). Thus, intragenotypic variability was evident in all six behavioral traits examined.

Enrichment affects behavioral means in a genotype-, measure-, and enrichment-dependent manner

We developed enrichment protocols that were either “mild” or “intense” (Figure 1.2). Our mild enrichment treatment, the fly jungle gym, was designed to provide a variety of textures, colors, and light conditions, whereas the intense cage enrichment was designed for flies to experience natural weather conditions, in addition to several different foods and a greater variety of biotic and abiotic substrates.

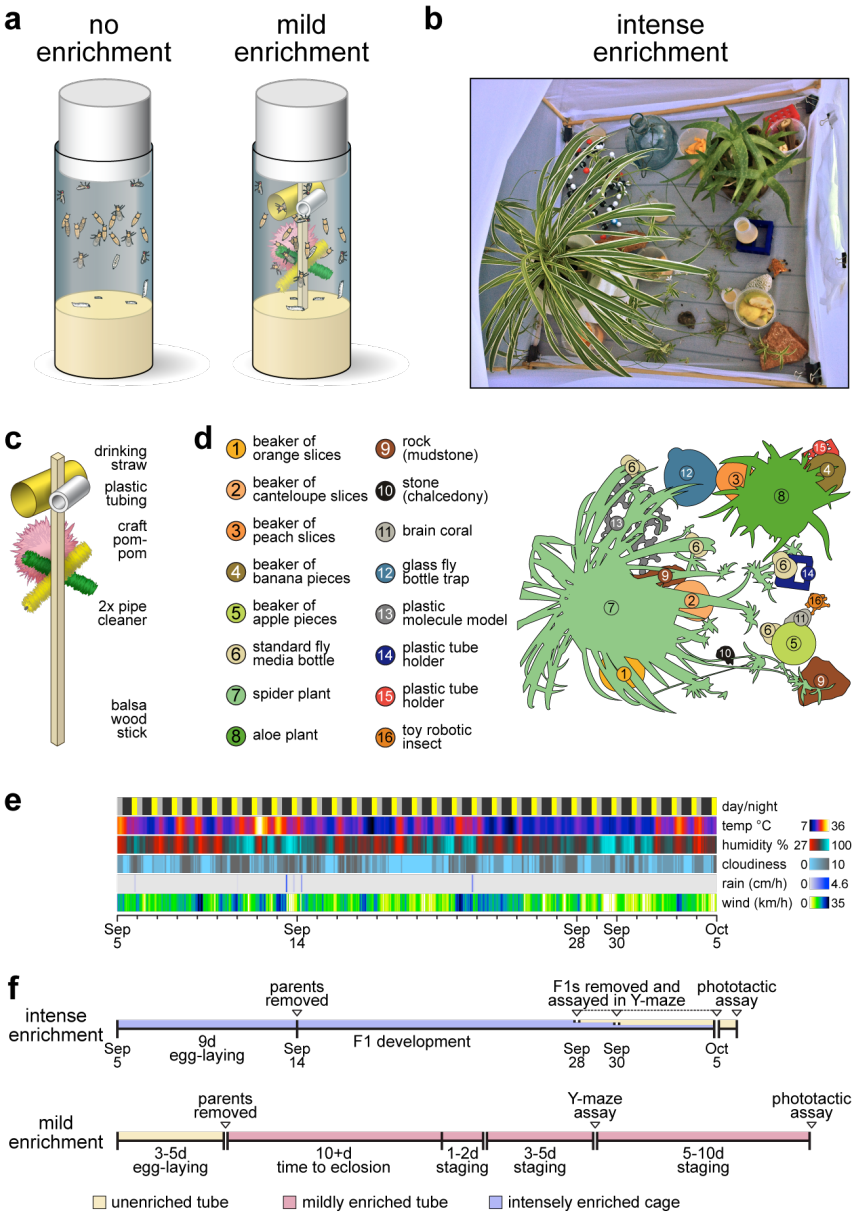


Figure 1.2 - Illustration of enrichment paradigms used.
a) Control vials vs. mild enrichment vials. **b)** Photo of the inside of the intense enrichment cage. **c)** Diagram of the mild enrichment jungle gym components. **d)** Diagram of the intense enrichment cage components. **e)** Weather conditions that the intense enrichment cage was subject to for the experimental period. For the daylight timeline, yellow indicates potential direct sunlight on the cage, grey periods where the cage was shaded by our building, and black shows nighttime. The cloudiness timeline reflects the NOAA 10-point scale where 0 is clear skies and 10 is full cloud cover. **f)** Timelines of experiments showing the development, staging, and behavioral testing of the experimental animals for both mild and intense enrichment treatments. Each contiguous horizontal line indicates the time spent in a fresh container. The no enrichment control was the same as the mild enrichment, except all vials used were unenriched.

To confirm that the enrichment treatments had an effect on our flies, we examined the mean values of our six behavioral phenotypes under each treatment. We used a Bayesian framework with a weakly informative prior to estimate the posterior distributions of the means and variances of each behavioral metric under the mild and intense enrichment treatments. We used the 99% highest density interval, also termed credible interval (see Methods), to assess whether the posterior distributions of the means of each behavioral metric were different from each other. Intense enrichment caused strong decreases in the mean of number of turns and inter-choice interval, and a strong increase in the turn switchiness when compared to mild enrichment and the control (Figure 1.3a).

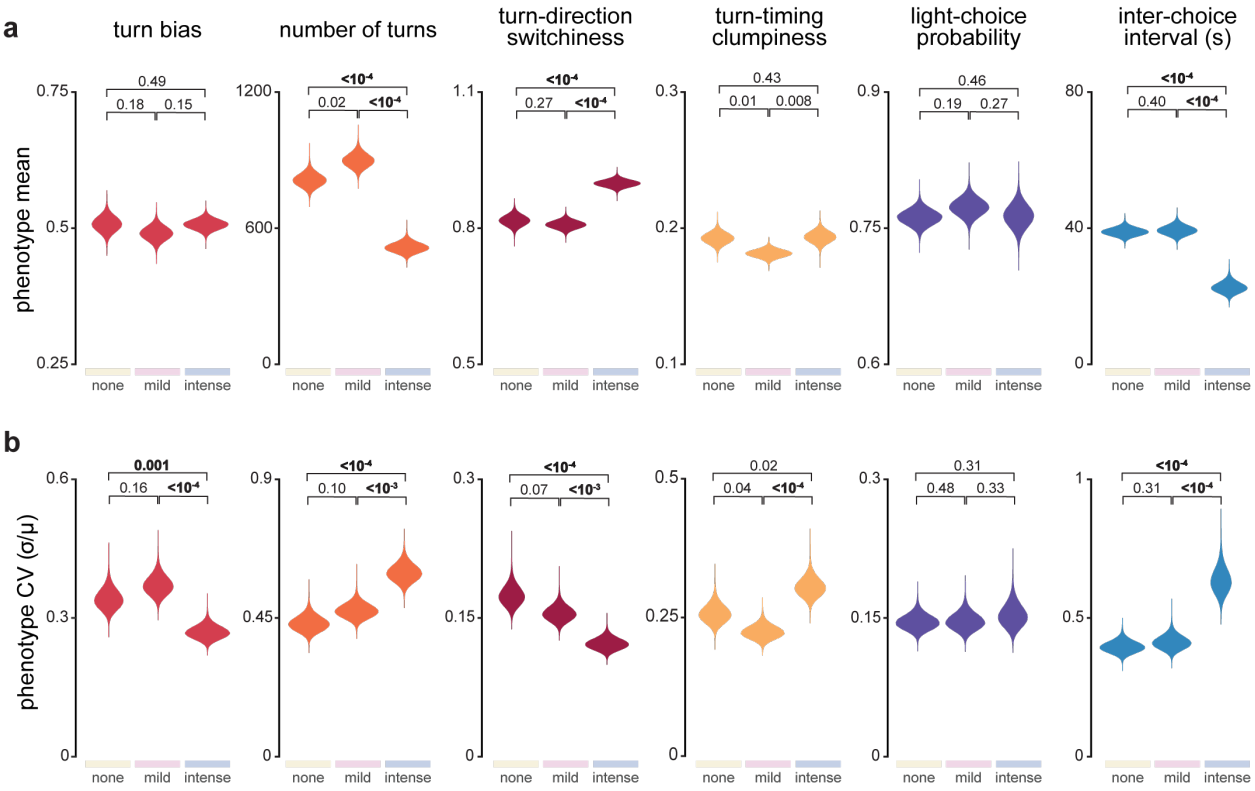


Figure 1.3 - Posterior distributions of a) mean and b) intragenotypic variability (coefficient of variation; CV) for Canton-S flies under both enrichment treatments.
Values shown at the top of each plot are the fraction of the posterior distribution of differences between two treatments (e.g., control and mild enrichment) that lies either below or above zero, depending on the direction of change. Given our finite posterior sampling, we cannot estimate fractions of distributions accurately below approximately 1E-4. Bold values indicate treatments for which the 99% credible interval of the treatment effect does not include 0. Sample sizes of each experiment are provided in the Methods.

For these behaviors, intense enrichment had a larger effect on the mean than mild enrichment. Mild enrichment had a less pronounced effect on the mean number of turns and turn clumpiness. There was no apparent effect of enrichment on turn bias and light-choice probability, though the FlyVac assay has lower power than the Y-maze assay. We viewed these observed mean changes as a positive control that the flies were sensitive to our enrichment treatments. Our results were supported by a non-parametric test of mean differences (Figure 1.4).

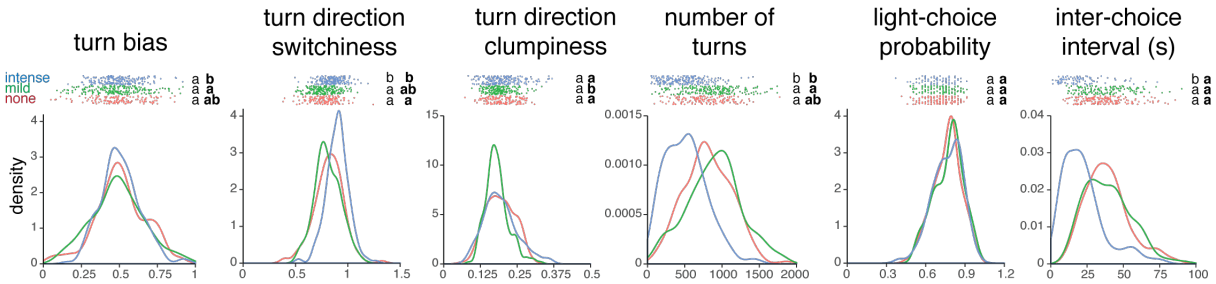


Figure 1.4 - Non-parametric tests of mean and variance differences across the unenriched, mildly enriched, and intensely enriched treatment groups in Canton-S flies for the six behaviors examined.
Kernel density plots and raw data were plotted for each behavior and each treatment group. Significant differences in treatment group means (as shown by the letter code) were determined using a Kruskal-Wallis test followed by a Dunn's test of multiple comparisons with a Bonferroni correction. Significant differences among group variances (bolded letter code) were determined by a Fligner-Killeen test of homogeneity of variances with a Bonferroni correction to control for multiple comparisons.

We estimated the effects of genotype, mild enrichment and genotype-by-mild enrichment on behavioral means using four Drosophila Genetic Reference Panel (DGRP)[52] lines (45, 105, 535, 796) and Canton-S (Figure 1.5). Genotype had an effect (i.e., zero was not in the 99% credible interval of the posterior distribution) on all behaviors except turn bias. Mild enrichment caused a genotype-independent increase in number of turns and switchiness, but no other behaviors. There were genotype-by-mild enrichment effects on number of turns, switchiness, and clumpiness, and the direction of those effects were variable.

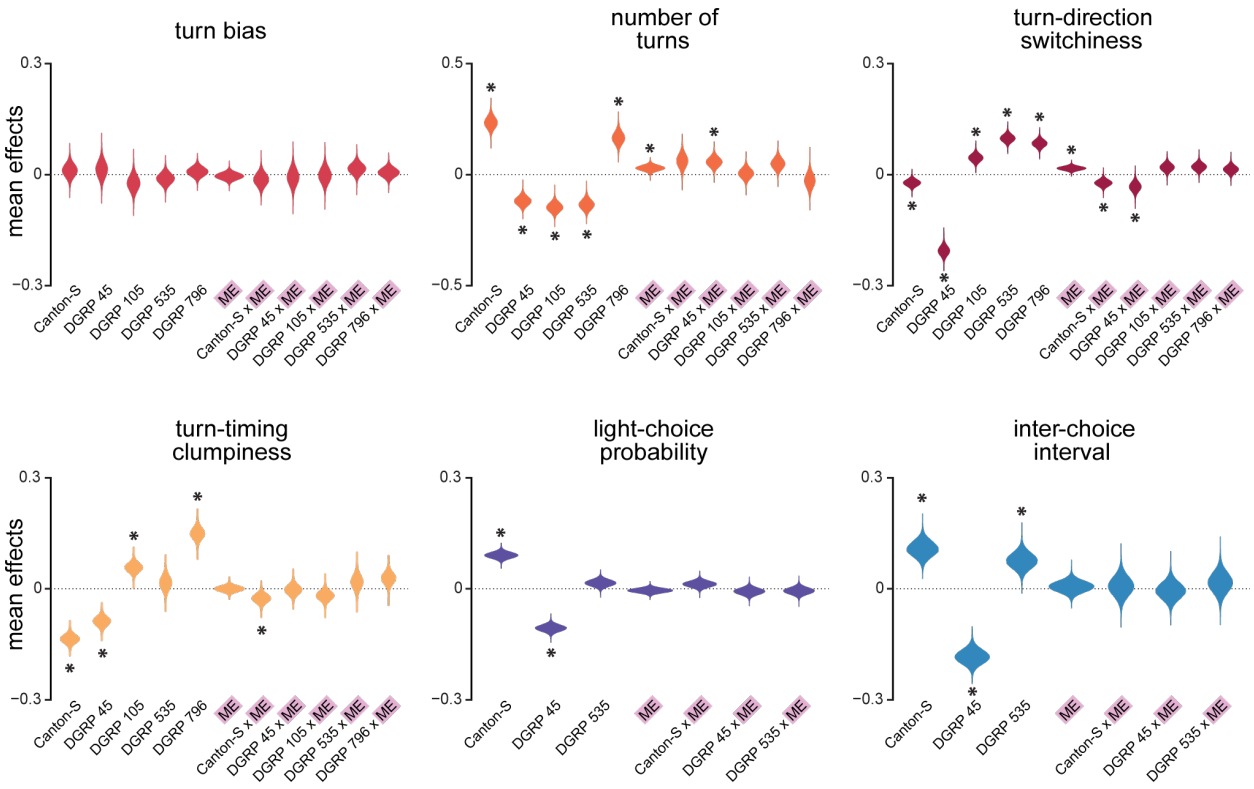


Figure 1.5 - Genotype, mild enrichment, and genotype-by-mild enrichment effects on behavioral measure means.
Asterisks mark those effects whose 99% credible interval does not include zero. All effects were normalized by the grand mean of all treatments and genotypes, so these values can be interpreted as effect sizes of each condition on the mean. Sample sizes of each experiment are provided in the Methods. ME, mild enrichment.

Genotype, behavioral measure, enrichment, and their interactions determine intragenotypic variability

We examined the effect of mild and intense enrichment on intragenotypic variability in our behavioral measures (Figure 1.3b). We chose to look at the coefficient of variation as our measure of intragenotypic variability in order to standardize it across multiple types of measures and control for mean effects (estimates of the posterior distributions of variance effects, not normalized by the treatment means, are included in Figure 1.6 and Figure 1.7).

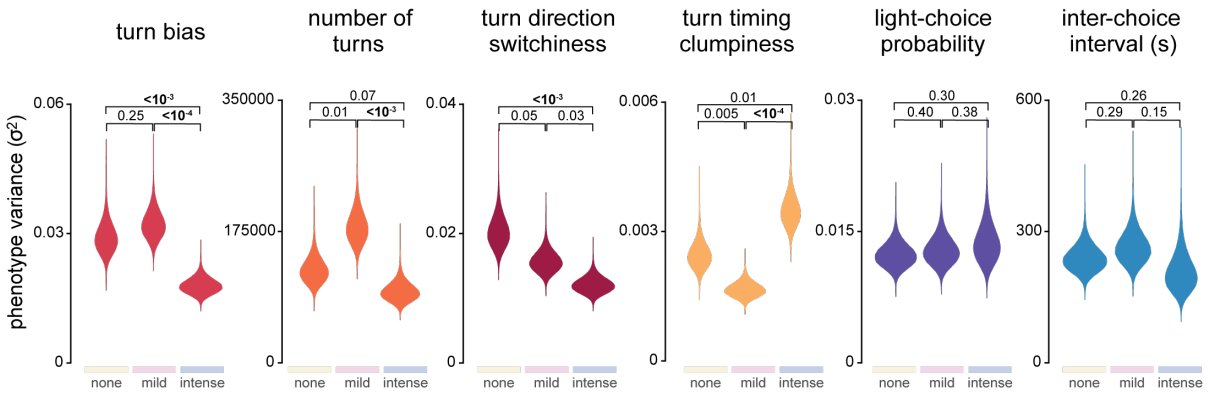


Figure 1.6 - Posterior distributions of behavioral measure variance for Canton-S flies under two enrichment treatments.
Numbers at the tops of the panels are the fraction of the posterior distribution of differences between two treatments (e.g., control and mild enrichment) that lies either below or above zero, depending on the direction of change. Given our finite posterior sampling, we cannot accurately estimate fractions of distributions below approximately 1E-4. Bold values indicate treatments for which the 99% credible interval of the treatment effect does not include 0. See Methods. Sample sizes of each experiment are provided in the Methods.

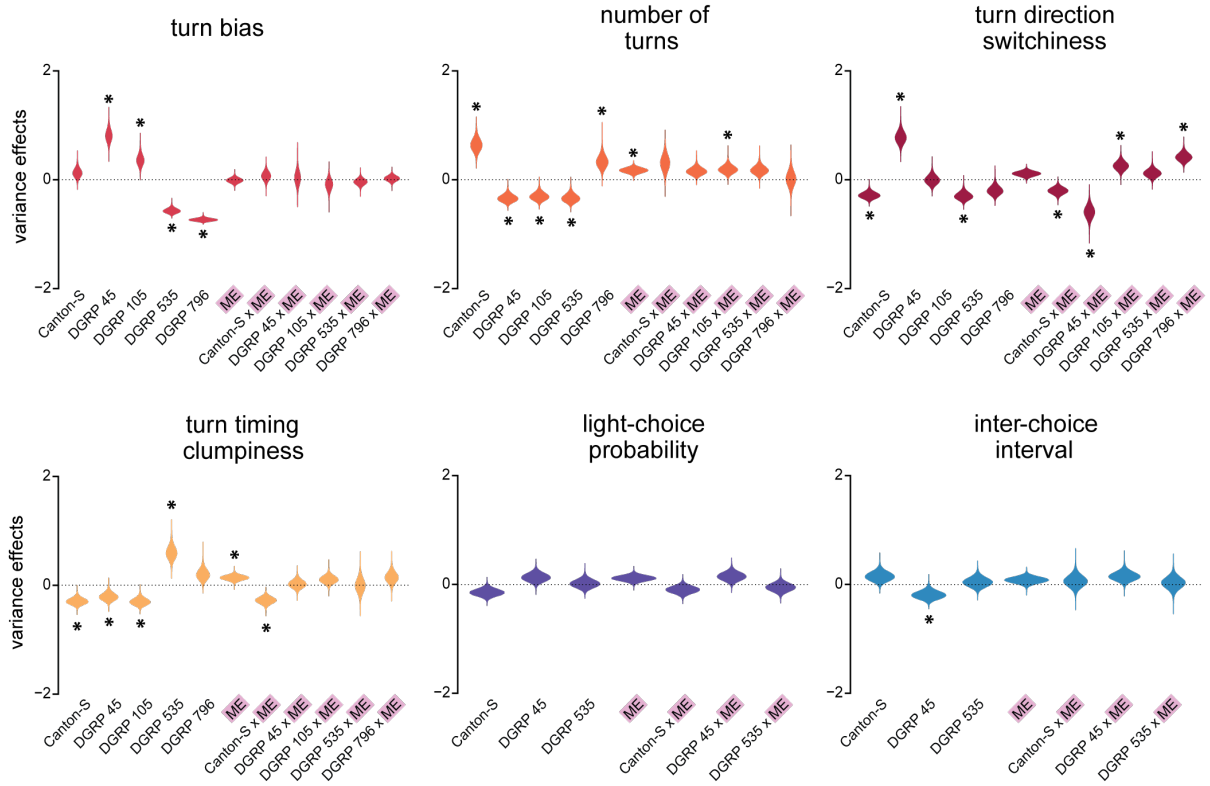


Figure 1.7 - Genotype, mild enrichment, and genotype-by-mild enrichment effects on behavioral metric variance.
Asterisks mark those effects whose 99% credible interval does not include zero. All effects were normalized by the grand variance of all treatments and genotypes. Sample sizes of each experiment are provided in the Methods. ME, mild enrichment.

For nearly all behaviors, intense enrichment had a larger effect on variability than mild enrichment, but these effects were not all in the same direction (results of a non-parametric test of differences in variance are shown in Figure 1.4). Intense enrichment decreased the variability of turn bias and turn direction switchiness but increased the variability of number of turns and inter-choice interval, when compared to the control and mild enrichment treatments. Intense enrichment increased variability in clumpiness, though the effect was more pronounced upon comparison to mild enrichment as opposed to the control.

Mild enrichment caused small or no differences (zero effect was within the 99% credible interval) when compared to the control treatment for all the behavioral measures from both assays, with turn direction switchiness and turn timing clumpiness the most likely behavioral measures to be affected by mild enrichment. Variability in turn bias and number of turns probably increased slightly under mild enrichment, while clumpiness and switchiness probably decreased slightly. In two of these cases the direction of the effect matched the direction of the intense enrichment effect; in the other two cases, it did not. To summarize, intense enrichment had stronger effects on variability than mild enrichment, and the direction of these effects was behavior-dependent.

In our analysis of five isogenic lines (four DGRP and Canton-S) under unenriched and mildly enriched conditions, we found that genotype, mild enrichment and genotype-by-mild enrichment all had effects on intragenotypic variability (Figure 1.8). We found that the variability of all behavioral measurements, except the number of turns, were affected by genotype. The variability of number of turns, clumpiness, and light choice increased in a genotype-independent manner under mild enrichment. Variability of switchiness and inter-choice interval were probably also increased in a genotype-independent manner by mild enrichment (a large majority of their respective posterior distributions was above zero). We observed genotype-by-mild enrichment

effects for number of turns, switchiness, clumpiness, and light-choice probability. Of all the behavioral measures, switchiness showed the most variable and the strongest genotype-by-enrichment effects. To summarize, mild enrichment often increased variability in a genotype-independent fashion, but there were also frequently genotype-by-mild enrichment effects.

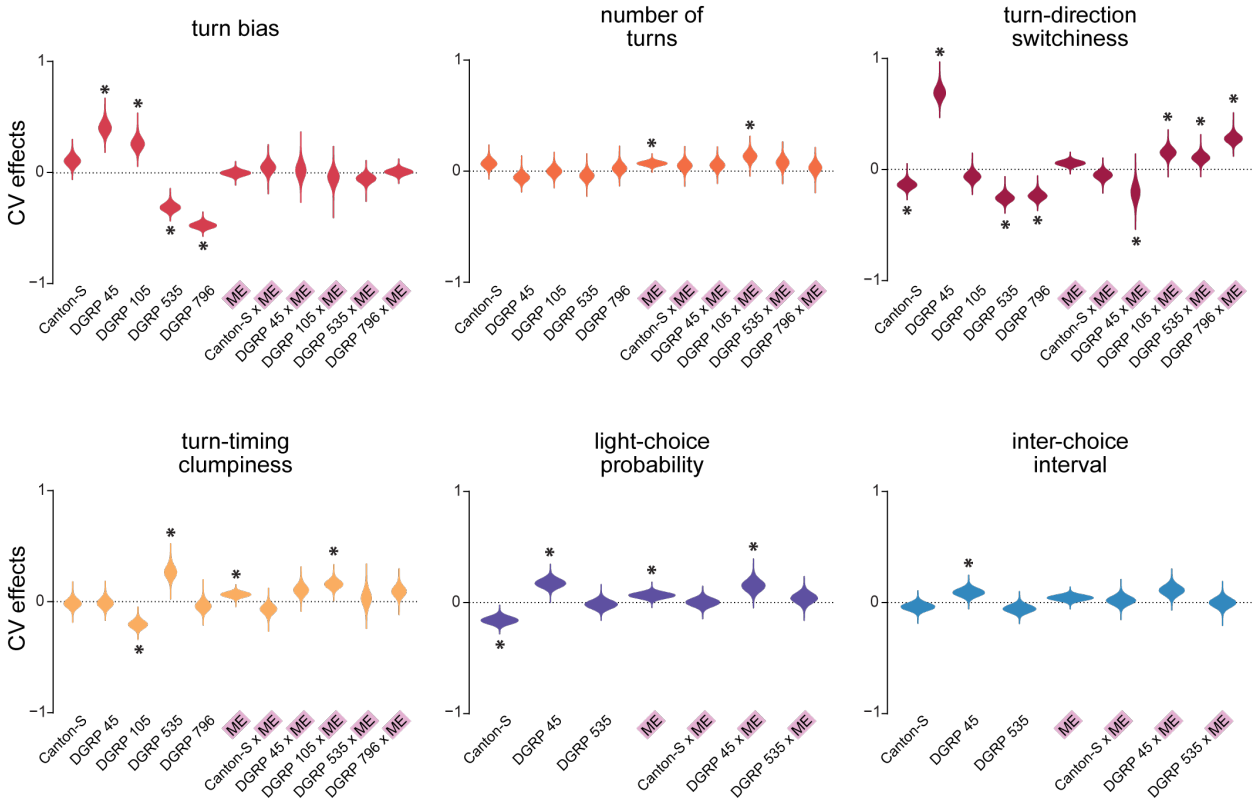


Figure 1.8 - Genotype, mild enrichment, and genotype-by-mild enrichment effects on measures of behavioral intragenotypic variability.
Asterisks mark those effects whose 99% credible interval does not include zero. All effects were normalized by the grand variability of all treatments and genotypes, so these values can be interpreted as effect sizes of each condition on intragenotypic variability. Sample sizes of each experiment are provided in the Methods. ME, mild enrichment.

Interestingly, we found that the average magnitudes of mean effects were smaller than the average magnitudes of variability effects (Figure 1.9a,b). For both mean and variability, genotype effects tended to be larger than the mild enrichment or genotype-by-mild enrichment effects. This pattern was especially prominent for mean effects. We also found that the sizes of

the effects on behavioral means were uncorrelated with sizes of effects on variability (Figure 1.9c).

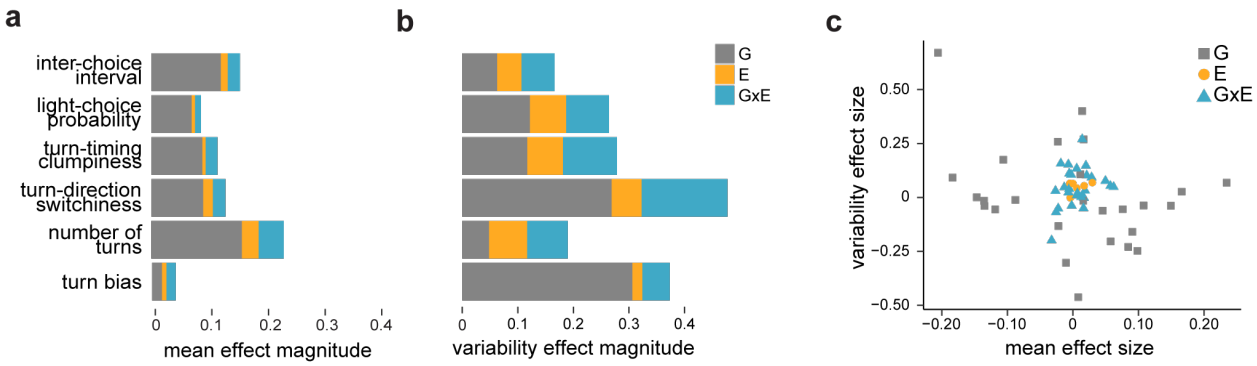


Figure 1.9 - Summary of the genotype, mild enrichment, and genotype-by-mild enrichment effects on measures of behavioral mean and intragenotypic variability.
a-b) Length of the bar represents the average magnitude of the effect; G is genotype effect, E is mild enrichment effect, and GxE is genotype-by-mild enrichment effect. **c)** Correlation of variability effect sizes and mean effect sizes across all behaviors, separated into genotype, mild enrichment, and genotype-by-mild enrichment effects. All effects were normalized by either the grand mean or variability of all treatments and genotypes, so these values can be interpreted as effect sizes of each condition.

Discussion

The goal of our study was to test two opposing hypotheses about the effects of enrichment on intragenotypic behavioral variability. We hypothesized that enrichment could increase variability due to the increase in microenvironmental diversity or decrease variability due to enriched environments more closely mimicking natural conditions (resulting in more robust development of behaviors and elimination of extremes that can occur in impoverished conditions)[32]. To this end, we examined six behaviors in *Drosophila melanogaster* across several genotypes and employed two levels of enrichment. We found that for five of the six behaviors, when examined across several genotypes, mild enrichment via a fly jungle gym likely led to an increase in intragenotypic variability, supporting the hypothesis that enrichment causes an increase in behavioral variability due to an increase in microenvironmental diversity. However, these genotype-independent effects were generally smaller than the effects of genotype or genotype-by-enrichment interactions. Therefore, the effect of enrichment on intragenotypic variability appears to depend on the particular genotype and behavior being assayed. When we examine the effects of enrichment on variability within a genotype, we see support for both hypotheses depending on the behavioral measure and enrichment treatment examined (mild or intense). Therefore, while it is broadly true that mild enrichment causes a small increase in the intragenotypic variability, a more granular look at the effects of enrichment reveals both strong increases and decreases in variability. We also found that the effects of enrichment on behavioral means and variabilities were largely independent of each other, with variability effects having larger magnitudes than mean effects (Figure 1.9). This finding confirms that our enrichment paradigm was able to affect both mean and variability, and that these effects are potentially independent. From our experiments, it remains uncertain which aspects of enrichment influence mean and which influence variability, or indeed, if these aspects are one and the same. Behavioral variability was also more strongly impacted by enrichment, which may underscore a biological flexibility that is not present in determining mean behavior.

With respect to both mean and variability, we found that genotype usually had a larger effect than the mild enrichment. This was especially obvious when looking at the genotype effects on behavioral means, where all behaviors except turn bias showed large genotype effects (Figure 1.5, Figure 1.9). The lack of effect of genotype on turn bias is consistent with previous work that found no differences in the mean turn biases of 159 DGRP lines[5]. Genotype also had strong effects on intragenotypic variability (Figure 1.8), as expected[5–7].

We found evidence of interactions between genotype and enrichment for practically all the behavioral measures examined, though the magnitude of these interactions was behavior-dependent (Figure 1.9). For example, turn bias and inter-choice interval showed very little genotype-by-enrichment effect for variability, but large effects were seen for turn switchiness. Dependence of variability on the particular parameter measured was previously noted in mouse enrichment studies[32,50]. Behavioral parameters may fall into different categories with respect to their response to enrichment. For example, switchiness is a measure of intraindividual variability (Figure 1.1), hinting at a link between the biological mechanism controlling variability from trial-to-trial and individual-to-individual[53]. Our results also make it clear that in assessing the effects of enrichment on a particular measure of behavior, genotype cannot be ignored. These interactions are consistent with previous findings in rats and mice[37–39], where the effects of enrichment differed between strains. Our results also imply that there is genotype-dependent plasticity in variability. In essence, phenotype variability is not a static feature of a genotype, but depending on the trait measured, the environment can have a large effect. Evolution of plasticity has usually been examined in trait means. Our findings suggest such inquiries should extend to variability. Going forward, thinking about variability as a flexible, evolvable trait may be necessary to understand how phenotype distributions arise.

Our fly jungle gym enrichment (mild enrichment) featured an array of perching sites and perching materials. The goal was that flies in the jungle gyms would experience a diversity of perching sites and textures, as well as be forced to navigate a more complex environment. Flies would therefore be subject to a diversity of experiences closer to what they might have in the wild while still under the constraints of laboratory conditions. We expected that the mild enrichment would mostly affect locomotion and activity behaviors, such as turn bias and inter-choice interval. We see that mild enrichment caused an increase in the mean of the number of turns and switchiness, as well as an increase in the variability of all behaviors except turn bias (Figure 1.5, Figure 1.8). Surprisingly, we see that variability in light-choice probability also increased under the mild enrichment treatment, which leads us to believe that our jungle gym construction may have also created differential light conditions in the vial or stimulated phototaxis variability via more indirect means.

The six behavior measures examined were chosen largely because they could be measured at scale across many individuals, a requirement for measuring effects on variability. Still, we can speculate on the ecological relevance of variability in several of the phenotypes measured. We suspect that variability in turn bias could be potentially advantageous for exploration, dispersal, and/or foraging via a bet-hedging mechanism. Individuals with stronger turning biases move through the environment with lower effective diffusion constants. If the spatial scale of resources in the environment fluctuates unpredictably, variation in turn bias could reflect a matched strategy of diversifying diffusion constants. Light-choice bias in our assay may be reflective of an escape response since the fly is startled prior to the light choice[7]. Variability in moving toward light upon being startled could reflect a bet-hedging mechanism as well, if the threats faced by flies are variable i.e., if it is alternatively advantageous to seek light or dark after a startle. Predator escape behavior has been considered to be a possible bet-hedging trait - for example, clonal pea aphids show variability in predator escape behavior among individuals[8]. We have

also found that variability in light preference under non-startled conditions could influence the thermal experience of a fruit fly in nature, and therefore also be part of a bet-hedging strategy[26].

We examined whether the effects on variability would change with a different type of enrichment. We raised one cohort of flies in a naturalistic setting, subject to the environmental fluctuations of the outdoors and with access to numerous organic and inorganic substrates (Figure 1.2). Compared to the jungle gyms of mild enrichment, this intense enrichment treatment had more structural complexity and diversity of biotic (fruits, plants, spider predators) and abiotic (sunlight, temperature) factors. By increasing the microenvironmental diversity along several different axes, we expected to observe stronger effects on variability on flies reared in this treatment. In general, this intense enrichment did have stronger effects on our behavioral measures than the mild enrichment. Even though the effects of intense enrichment were more pronounced, the direction of these effects was behavior-dependent (Figure 1.3). For example, we saw a decrease in intragenotypic variability for turn bias and turn switchiness under intense enrichment, but an increase in the variability for number of turns and turn clumpiness. The directions of these effects varied by behavior, even relative to the direction of the mild enrichment effect.

The intense enrichment treatment was created to have a higher level of microenvironmental diversity than the mild enrichment treatment, yet we observe that the behavioral variability does not always change in the same direction between these treatments. For example, we saw a small increase in variability under mild enrichment but a large decrease under intense enrichment such as for turn bias (Figure 1.3). This leads us to believe that the relationship between the mild and intense enrichment is not just a simple increase in “enrichment intensity.” Still, we recorded some direct evidence that flies in the intense enrichment indeed experienced at least one dimension of increased microenvironmental variation compared to the mild

enrichment: flies recovered from the intense enrichment cage exhibited variation in the color of their gut contents, consistent with their having recent fed on different food sources (Figure 1.10). This variation was absent in flies subject to mild enrichment. One of our predictions was that a more naturalistic enrichment treatment could lead to a decrease in variability because of an increase in robustness, but it could also be that naturalistic enrichments cause fly populations to exhibit more natural behaviors in general, whether or not that corresponds to a decrease in behavioral variability. Future studies could address what constitutes natural fly behavior in more detail, whether by making field-deployable assays or bringing wild flies directly to the lab for testing, though any comparisons with our current enrichment paradigm would need to carefully consider population genotypic variance.

Figure 1.10 - Differences in food consumption in intense enrichment.
Flies recovered from the intense enrichment cage displayed differences in the type of food consumed, as shown by the differences in the gut color (black arrowhead).

Overall, our results support the hypothesis that enrichment increases intragenotypic variability, though this effect is highly dependent on the particular genotype, enrichment, and behavior in question. We also conclude that genotype is likely to remain the main determinant of intragenotypic variability. Our findings make it apparent that the genotype used, and behavior

measured will affect the inferred relationship between environmental variability and behavioral variability. Moreover, the type of enrichment (e.g., mild vs intense enrichment) can qualitatively and quantitatively alter this relationship. This, and the effects of genotype and behavioral measure, could be why effects observed in one enrichment study may not be replicated in another[38]. While the multifactorial nature of enrichment provides challenges, its specific effects continue to be of great interest in behavioral research, and high throughput, data-driven approaches have the potential to illuminate the complex relationships between environmental variability and behavioral variability.

Methods

Behavior and enrichment protocols

Fly stocks were cultured in vials on Caltech formula media at 25°C in temperature-controlled incubators on a 12h-12h light-dark cycle. For our isogenic populations, we used Canton-S and four lines from the Drosophila Genetic Reference Panel[52]. These lines (numbers 45, 105, 535 and 796) were derived from different wild-caught gravid females and then inbred for 20 generations. Thus, there is significant genetic variation between the lines, but not within them. We chose to work with these particular lines because we have previously observed that they vary in intragenotypic variability in Y-maze turn bias[5]. Flies were subjected to two enrichment treatments: mild enrichment (the addition of a small "jungle gym" to each culture vial) and intense enrichment (growth of the flies in a $1m^3$ cage filled with many rotting fruit substrates, plants, rocks etc.) (Figure 1.2).

For mild enrichment, in each vial of media where experimental animals were to develop, 3 female and 2 male parental flies were housed for 3-5 days. The parents were removed and the jungle gym enrichment was inserted. The enrichment object consisted of plastic tubing, pipe cleaners and fuzzy pom poms that were hot-glued to a balsa wood applicator stick that was inserted into the media (Figure 1.2c). These were identically constructed for ~30 vials, with the exception of the pom-pom color, which in some vials at random was white and in the others pink. F_1 experimental progeny developed in this enriched environment for around 10 days in incubators. Once they began eclosing, they were allowed to accumulate for 1-2 days, after which they were removed and mixed under cold anesthetization with other flies from the same genotype. They were then sorted into cohorts of 40 males and 40 females and placed in mildly enriched vials for 3-5 days. At this point, their behavior was measured in the Y-mazes for 2 hours according to the methods in Buchanan *et al*., 2015[6] (though here we loaded anesthetized experimental animals on ice to transfer them into the Y-mazes, rather than CO_2). After the Y-

maze assay, they were anesthetized on ice and returned to their mildly enriched vials for 5-10 days at which point their phototactic preferences were measured using FlyVac according to the methods in Kain *et al*. (2012)[7]. This mild enrichment procedure was used for flies of all genotypes. See Figure 1.1 for representations of the six behavioral metrics we acquired in these two assays, and their distributions across Canton-S flies.

For intense enrichment, we prepared a population cage using 1m wooden dowels to make a cubic frame, with sheer white polyester drapery material as walls (Figure 1.2b). A tube of this material, normally held closed by binder clips, provided access to the inside of the cage. The items shown in Figure 1.2d were introduced to the cage at the time of its construction: six kinds of fly food (a variety of decomposing fruits as well as bottles of standard cornmeal media), houseplants, stones, varied plastic objects, etc The cage was placed outside on a deck where it experienced natural fluctuations in luminance, temperature, rainfall, wind, humidity, etc. during the course of our experiment (Figure 1.2e). For the experiment, a parental Canton-S population of 200 males and 200 females was placed in the cage on September 5th 2013, and removed 9 days later. F_1s were collected on September 28th, 30th, and October 5th 2013 and assayed in the Y-mazes on those days respectively. Flies were recovered from the Y-mazes using cold anesthetization and stored in unenriched standard media tubes in groups of ~30 individuals until testing with FlyVac on October 6th 2013 (Figure 1.2f). For all assays, males and females were tested in equal proportions.

For Y-maze enrichment experiments with Canton-S, we assayed 151 control flies, 203 mildly enriched flies, and 206 intensely enriched flies. For Canton-S FlyVac experiments, we assayed 175 control flies, 140 mildly enriched flies, and 86 intensely enriched flies.

For Y-maze enrichment experiments with the DGRP lines, we assayed:

DGRP 45: 166 control, 133 mildly enriched

DGRP 105: 130 control, 148 mildly enriched

DGRP 535: 113 control, 111 mildly enriched

DGRP 796: 132 control, 128 mildly enriched

For FlyVac enrichment experiments with DGRP lines, we assayed:

DGRP 45: 157 control, 144 mildly enriched

DGRP 535: 122 control, 140 mildly enriched

Behavior measures and null model distributions

Behavior measures from the Y-maze assay (turn bias, number of turns, turn direction switchiness and turn timing clumpiness) were calculated from the vectors of turn directions and times that each fly produced in the experiment. Behavior measures from FlyVac (light-choice probability and inter-choice interval) were calculated from the FlyVac data output file[7]. These measures were computed and/or collected into a common data structure in MATLAB 2013a (The Mathworks, Inc., Natick, MA). With Y-maze arrays[6], we measured the left-vs-right free locomotion turning bias of individual flies. With FlyVac, we measured the locomotory response to light cues of agitated flies ("fast phototaxis"[54]). We have previously used both of these assays to detect genetic and neural circuit regulators of intragenotypic variability[5–7], and between them, we examined both spontaneous and stimulus-evoked behaviors. Beyond turn bias in the Y-maze assay, we assessed 1) the number of turns completed by individual flies within the two hour trials, 2) flies' tendencies to alternate between left and right turns successively ("switchiness"), and 3) the extent to which their turning events were clustered in time ("clumpiness"). For the FlyVac dataset, we measured the average interval between phototactic choices in addition to the light-choice probability. While the Y-maze arrays and FlyVac were primarily designed to measure locomotor turning and phototaxis, respectively, they also produce

precise estimates of these other individual behavioral measures, so for the purpose of this study, we do not emphasize any of these measures over the others.

Null hypothesis distributions were generated in MATLAB 2013a by resampling (with replacement) a million values for each distribution as follows: 1) For turn bias and light-choice probability, a) all observed choice values (i.e. left vs. right and light vs. dark) were pooled across individuals, b) an individual was chosen at random from all tested, and a vector of length equal to the number of behavioral choices performed by that individual during the experiment was populated randomly with values from the pool, and c) the turn bias or light-choice probability for that vector was recorded. 2) For number of turns, a) the observed inter-turn intervals (ITIs) were pooled across individuals, b) ITIs were chosen randomly one at a time until their cumulative sum exceeded 7200000ms, the length of an experiment, and c) the number of turns in this sequence was recorded. The moderate discrepancy in mean between the null hypothesis distribution and experimental distribution in this analysis arises from the disproportionate number of short intervals contributed to the total pool by more active animals. 3) For switchiness, which arises from slight dependence between consecutive turns in the LR turn sequence, a) we implemented a Markov chain in which the L-L (= R-R) transition probabilities yielded LR sequences with mutual information between successive turns equal to the observed mutual information (0.018 bits, P(L-L) = 0.592). b) An individual was chosen at random from all tested, and a choice sequence of length equal to the number of choices performed by that individual during the experiment was generated using the Markov chain, and c) the switchiness of this sequence (= (#LR+#RL)/(2*turn bias*(1-turn bias)*num turns) was recorded. 4) For clumpiness, a) the observed ITIs were pooled across individuals, b) an individual was chosen at random from all tested, and a vector of length equal to the number of choices performed by that individual during the experiment was populated randomly with values from the ITI pool, and c) clumpiness (= MAD(ITIs)/(sum(ITIs)/num turns), where MAD is the median absolute deviation from the

median) was recorded. Thus, this approach reflects sampling variation in number of turns and mean ITI as well as clumpiness. 5) For inter-choice interval (ICI), a) the observed ICIs were pooled across individuals, b) an individual was chosen at random from all tested, and a vector of length equal to the number of choices performed by that individual during the experiment was populated randomly with values from the ICI pool, c) the mean ICI across this vector was recorded.

Bayesian inference of mean and variance effects

To get the estimates of the posterior distributions of behavioral mean and variance, and the effects on the observed distributions of enrichment treatment, genotype, and their interactions, we employed linear and generalized linear models in R's Stan interface v.2.18.2. The Stan platform allows the user to specify desired models and performs full Bayesian inference using Hamiltonian Monte Carlo with the No U-Turn sampler[55]. To get the posterior distributions of the mean and variance for turn bias, light-choice probability, switchiness, and clumpiness under different enrichment-genotype conditions, we specified the following model:

$$
\begin{aligned}
y_n &\sim Normal(\mu_n, \sigma_n) \\
\mu &= a + \boldsymbol{X} \cdot \boldsymbol{b} \\
\sigma^2 &= v_0 + X \cdot v
\end{aligned}
$$

where y_n is the behavioral outcome of an individual n that comes from a normal distribution with parameters μ_n (mean) and σ_n (standard deviation). μ and σ^2 are vectors specified via linear models, where a and v_0 are intercepts, X is a logical predictor matrix specifying the genotype and/or enrichment treatment for each individual, and b and v are vectors of coefficients of the linear model.

Since the distribution of the number of turns was right-skewed, bounded to real positive integers, and overdispersed compared to a Poisson distribution, we chose to model this measure with a negative binomial as follows:

$$
\begin{aligned}
y_n &\sim NegBinomial(\mu_n, \varphi_n) \\
\mu &= exp(X \cdot \beta) \\
\phi &= exp(X \cdot \gamma) \\
\sigma^2 &= \frac{\mu + \mu^2}{\phi}
\end{aligned}
$$

Here, y_n is the number of turns made by individual n modeled by a negative binomial distribution with parameters μ_n(mean) and φ_n (dispersion). Both parameter vectors are related to the coefficients of a generalized linear model (vectors β and γ) via a log-link function. $X$is the experimental design matrix, as above. σ^2 is the variance vector calculated from the mean and dispersion parameter vectors.

To model inter-choice intervals, we chose a gamma distribution since the data is right-skewed and positive continuous:

$$
\begin{aligned}
y_n &\sim \Gamma(a_n, b_n) \\
\mu &= exp(X \cdot \beta) \\
\sigma^2 &= exp(X \cdot \gamma) \\
a &= \frac{\mu^2}{\sigma^2} \\
b &= \frac{\mu}{\sigma^2}
\end{aligned}
$$

Where y_n is the inter-choice interval of individual n and a_n and b_n are the shape and rate parameters of the gamma distribution, respectively, and the rest of the parameters are as above.

For the estimation of all posterior distributions, we set our priors on the coefficients to broad Cauchy distributions centered at 0 to allow them to be weakly informative[56]. Our qualitative findings were robust to the choice of prior, as an uninformative uniform prior told the same story (Figure 1.11).

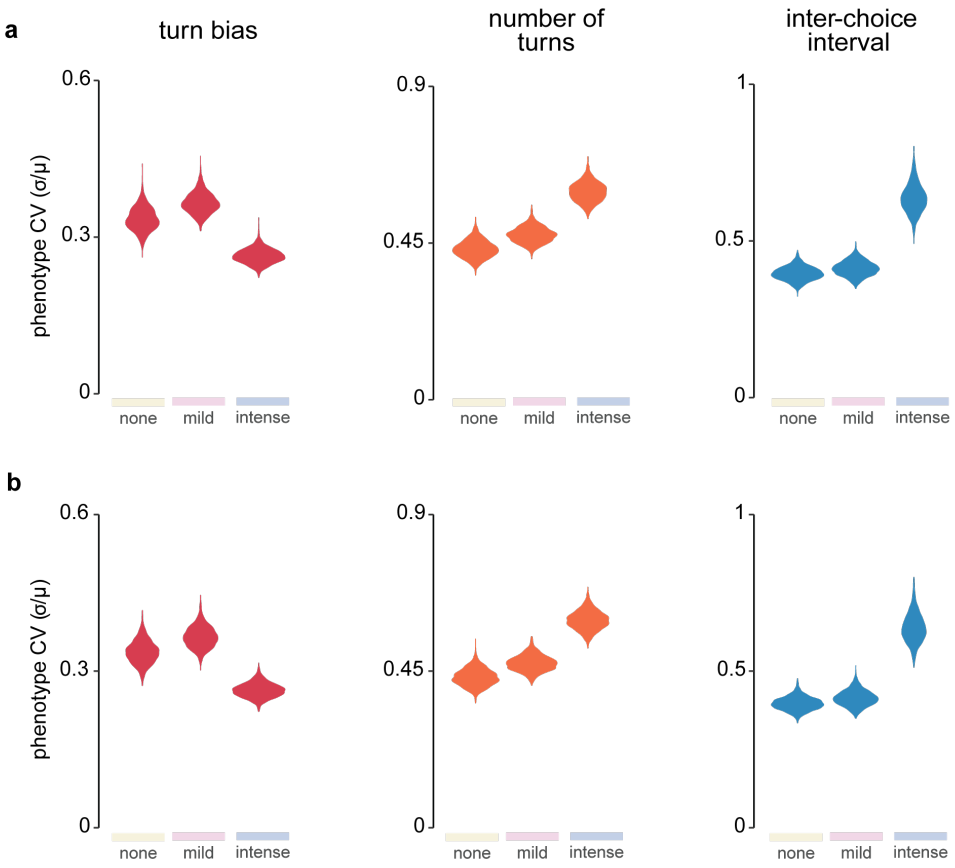


Figure 1.11 - Weakly informative Cauchy and uniform priors result in similar posterior distributions.
a) Posterior distributions generated under a Cauchy prior. **b)** Posterior distributions generated under a uniform prior. Thus, our choice of a weakly informative prior, rather than non-informative prior, is not driving our conclusions about the treatment effects. Turn bias, number of turns, and inter-choice interval were chosen to represent the three types of models used: normal, negative binomial, and gamma, respectively. Sample sizes of each experiment are provided in the Methods.

To sample posteriors, we used four chains and 50,000 - 100,000 iterations per chain, with the target average proposal acceptance probability of 0.8-0.9 and maximum tree depth of 10-15, to

generate a posterior distribution of 100,000 - 200,000 samples (50% of chain iterations were used for tuning the Hamiltonian Monte Carlo sampler parameters and were discarded as the burn-in period). The ratio of chain effective sample size to sample size was in the range of 0.7 - 1.2, indicating that posterior estimate error due to autocorrelation was minimal. To get posterior distributions for the coefficients of variation, we took the square root of the variance and divided it by the mean at each step in the chain. To check our model fits, we carried out graphical posterior predictive checks (Figure 1.12) and found that our models fit the data well.

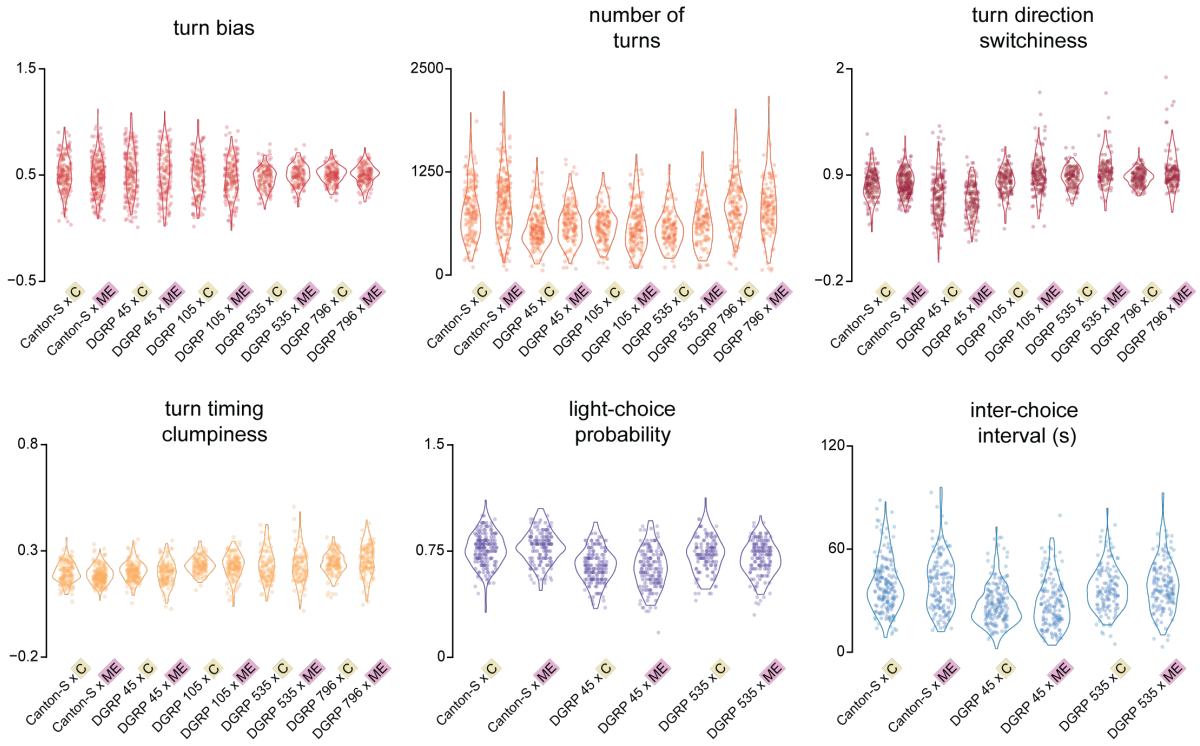


Figure 1.12 - Concordance between values sampled from our models of treatment effect parameters and the experimental data (posterior predictive checks).
Violin plots are kernel density estimates from a sample of values drawn from distributions whose parameters are drawn from our effect posteriors. Points are the original data. The general agreement between the modeled and empirical distributions indicates that we have appropriately modeled the effects and that our choice of priors was not wholly inappropriate. Sample sizes of each experiment are provided in the Methods. C, control, ME, mild enrichment.

We adapted the methodology used by Kruschke's BEST method[57] to our own posterior distributions in order to determine which effects were inferred to differ from zero. To estimate

the posterior distribution of a treatment effect, we subtracted the parameter values of one treatment condition from the control (or the other treatment condition) at each step in the chain and took the distribution of that difference. We calculated the 99% highest density interval (the credible interval) of the posterior of treatment effects to evaluate whether the treatments had an effect - if the 99% highest density interval excluded 0, we inferred an effect between the treatments. Since this approach is subject to multiple comparisons concerns, we chose the 99% credible interval (rather than e.g., 95%) as a more stringent indicator of effects. While we think that the 99% credible interval is a useful guide to pulling out the strongest effects we observe, we believe that the strength of Bayesian inference lies in being able to examine the posterior distributions as a whole and observing their trends (rather than applying a threshold to identify effects).

To determine the contribution of genotype, mild enrichment, and genotype-by-mild enrichment effects to the variability (coefficient of variance) and mean of each behavior, we used the following formulas:

$$CV_{ij} = CV_0 + G_i + E_j + G_i \times E_j$$
$$\mu_{ij} = \mu_0 + G_i + E_j + G_i \times E_j$$

CV_{ij} and $\mu_{ij}$are the variability and mean, respectively, of genotype *i* in treatment group *j*. CV_0 and μ_0 are the grand variability and mean, averaged over all genotypes and treatments. $G_i$is the deviation of the variability or mean of genotype *i* from the grand parameter in question, calculated over all treatment groups. E_j is the deviation of treatment group *j* from the grand parameter, calculated over all genotypes. The treatment groups in this experiment were mild enrichment or control vials. $G_i \times E_j$is the specific deviation of genotype *i* in treatment group *j* after accounting for the main effects of genotype *i* and treatment group *j*. All deviations were

standardized by dividing them by the grand parameter value in order to interpret them as effect sizes.

References

1. Dall, S. R. X., Houston, A. I. & McNamara, J. M. The behavioural ecology of personality: consistent individual differences from an adaptive perspective. *Ecol. Lett.* **7**, 734–739 (2004).

2. Sih, A., Bell, A. M., Johnson, J. C. & Ziemba, R. E. Behavioral Syndromes: An Integrative Overview. *Q. Rev. Biol.* **79**, 241–277 (2004).

3. Wolf, M. & Weissing, F. J. An explanatory framework for adaptive personality differences. *Philos. Trans. R. Soc. B Biol. Sci.* **365**, 3959–3968 (2010).

4. Gärtner, K. A third component causing random variability beside environment and genotype: a reason for the limited success of a 30 year long effort to standardize laboratory animals? *Int. J. Epidemiol.* **41**, 335–341 (2012).

5. Ayroles, J. F. *et al.* Behavioral idiosyncrasy reveals genetic control of phenotypic variability. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 6706–6711 (2015).

6. Buchanan, S. M., Kain, J. S. & de Bivort, B. L. Neuronal control of locomotor handedness in *Drosophila*. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 6700–6705 (2015).

7. Kain, J. S., Stokes, C. & de Bivort, B. L. Phototactic personality in fruit flies and its suppression by serotonin and white. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 19834–19839 (2012).

8. Schuett, W. *et al.* Personality variation in a clonal insect: the pea aphid, *Acyrthosiphon pisum*. *Dev. Psychobiol.* **53**, 631–640 (2011).

9. Stern, S., Kirst, C. & Bargmann, C. I. Neuromodulatory control of long-term behavioral patterns and individuality across development. *Cell* **171**, 1649-1662.e10 (2017).

10. Bierbach, D., Laskowski, K. L. & Wolf, M. Behavioural individuality in clonal fish arises despite near-identical rearing conditions. *Nat. Commun.* **8**, 15361 (2017).

11. Vogt, G. *et al.* Production of different phenotypes from the same genotype in the same environment by developmental variation. *J. Exp. Biol.* **211**, 510–523 (2008).

12. Freund, J. *et al.* Emergence of individuality in genetically identical mice. *Science* **340**, 756–759 (2013).

13. Kurikawa, T., Haga, T., Handa, T., Harukuni, R. & Fukai, T. Neuronal stability in medial frontal cortex sets individual variability in decision-making. *Nat. Neurosci.* 1 (2018) doi:10.1038/s41593-018-0263-5.

14. Debat, V. & David, P. Mapping phenotypes: canalization, plasticity and developmental stability. *Trends Ecol. Evol.* **16**, 555–561 (2001).

15. Honegger, K. & de Bivort, B. Stochasticity, individuality and behavior. *Curr. Biol.* **28**, R8–R12 (2018).

16. Willmore, K. E., Young, N. M. & Richtsmeier, J. T. Phenotypic variability: its components, measurement and underlying developmental processes. *Evol. Biol.* **34**, 99–120 (2007).

17. Clarke, G. M. The genetic basis of developmental stability. V. Inter- and intra-individual character variation. *Heredity* **80**, 562–567 (1998).

18. Morgante, F., Sørensen, P., Sorensen, D. A., Maltecca, C. & Mackay, T. F. C. Genetic Architecture of Micro-Environmental Plasticity in *Drosophila melanogaster*. *Sci. Rep.* **5**, 9785 (2015).

19. Simons, A. M. Modes of response to environmental change and the elusive empirical evidence for bet hedging. *Proc. R. Soc. B Biol. Sci.* **278**, 1601–1609 (2011).

20. Hopper, K. R. Risk-spreading and bet-hedging in insect population biology. *Annu. Rev. Entomol.* **44**, 535–560 (1999).

21. Simons, A. M. Fluctuating natural selection accounts for the evolution of diversification bet hedging. *Proc. R. Soc. B Biol. Sci.* **276**, 1987–1992 (2009).

22. Abley, K., Locke, J. C. W. & Leyser, H. M. O. Developmental mechanisms underlying variable, invariant and plastic phenotypes. *Ann. Bot.* **117**, 733–748 (2016).

23. Blasco, A., Martínez-Álvaro, M., García, M.-L., Ibáñez-Escriche, N. & Argente, M.-J. Selection for environmental variance of litter size in rabbits. *Genet. Sel. Evol. GSE* **49**, 48 (2017).

24. Dworkin, I. A Study of canalization and developmental stability in the sternopleural bristle system of *Drosophila melanogaster*. *Evolution* **59**, 1500–1509 (2005).

25. Sztepanacz, J. L., McGuigan, K. & Blows, M. W. Heritable micro-environmental variance covaries with fitness in an outbred population of *Drosophila serrata*. *Genetics* **206**, 2185–2198 (2017).

26. Kain, J. S. *et al.* Variability in thermal and phototactic preferences in Drosophila may reflect an adaptive bet-hedging strategy. *Evolution* **69**, 3171–3185 (2015).

27. Mellert, D. J., Williamson, W., Shirangi, T. R., Card, G. M. & Truman, J. W. Genetic and environmental control of neurodevelopmental robustness in *Drosophila*. *PLoS ONE* **11**, (2016).

28. Sørensen, P., de los Campos, G., Morgante, F., Mackay, T. F. C. & Sorensen, D. Genetic control of environmental variation of two quantitative traits of *Drosophila melanogaster* revealed by whole-genome sequencing. *Genetics* **201**, 487–497 (2015).

29. Tonsor, S. J., Elnaccash, T. W. & Scheiner, S. M. Developmental instability Is genetically correlated with phenotypic plasticity, constraining heritability and fitness. *Evolution* **67**, 2923–2935 (2013).

30. Linneweber, G. *et al.* A neurodevelopmental origin of behavioral individuality. *bioRxiv* 540880 (2019) doi:10.1101/540880.

31. van Praag, H., Kempermann, G. & Gage, F. H. Neural consequences of environmental enrichment. *Nat. Rev. Neurosci.* **1**, 191–198 (2000).

32. Körholz, J. C. *et al.* Selective increases in inter-individual variability in response to environmental enrichment in female mice. *eLife* **7**, e35690 (2018).

33. Bruel-Jungerman, E., Laroche, S. & Rampon, C. New neurons in the dentate gyrus are involved in the expression of enhanced long-term memory following environmental enrichment. *Eur. J. Neurosci.* **21**, 513–521 (2005).

34. Garthe, A., Roeder, I. & Kempermann, G. Mice in an enriched environment learn more flexibly because of adult hippocampal neurogenesis. *Hippocampus* **26**, 261–271 (2016).

35. Leger, M. *et al.* Environmental enrichment duration differentially affects behavior and neuroplasticity in adult mice. *Cereb. Cortex N. Y. N 1991* **25**, 4048–4061 (2015).

36. Mohammed, A. H. *et al.* Environmental enrichment and the brain. *Prog. Brain Res.* **138**, 109–133 (2002).

37. Konkle, A. T., Kentner, A. C., Baker, S. L., Stewart, A. & Bielajew, C. Environmental-enrichment–related variations in behavioral, biochemical, and physiologic responses of Sprague–Dawley and Long Evans rats. *J. Am. Assoc. Lab. Anim. Sci.* **49**, 427–436 (2010).

38. Toth, L. A., Kregel, K., Leon, L. & Musch, T. I. Environmental enrichment of laboratory rodents: the answer depends on the question. *Comp. Med.* **61**, 314–321 (2011).

39. van de Weerd, H. A., Baumans, V., Koolhaas, J. M. & van Zutphen, L. F. Strain specific behavioural response to environmental enrichment in the mouse. *J. Exp. Anim. Sci.* **36**, 117–127 (1994).

40. Toli, E. A., Noreikiene, K., DeFaveri, J. & Merilä, J. Environmental enrichment, sexual dimorphism, and brain size in sticklebacks. *Ecol. Evol.* **7**, 1691–1698 (2017).

41. Turschwell, M. P. & White, C. R. The effects of laboratory housing and spatial enrichment on brain size and metabolic rate in the eastern mosquitofish, *Gambusia holbrooki*. *Biol. Open* **5**, 205–210 (2016).

42. Barth, M. & Heisenberg, M. Vision affects mushroom bodies and central complex in *Drosophila melanogaster*. *Learn. Mem.* **4**, 219–229 (1997).

43. Heisenberg, M., Heusipp, M. & Wanke, C. Structural plasticity in the *Drosophila* brain. *J. Neurosci.* **15**, 1951–1960 (1995).

44. Technau, G. M. Fiber number in the mushroom bodies of adult *Drosophila melanogaster* depends on age, sex and experience. *J. Neurogenet.* **1**, 113–126 (1984).

45. Dawson, E. H. *et al.* Social environment mediates cancer progression in *Drosophila*. *Nat. Commun.* **9**, 3574 (2018).

46. Scotto Lomassese, S. *et al.* Influence of environmental stimulation on neurogenesis in the adult insect brain. *J. Neurobiol.* **45**, 162–171 (2000).

47. Wang, X., Amei, A., Belle, J. S. de & Roberts, S. P. Environmental effects on Drosophila brain development and learning. *J. Exp. Biol.* **221**, jeb169375 (2018).

48. Freund, J. *et al.* Association between exploratory activity and social individuality in genetically identical mice living in the same enriched environment. *Neuroscience* **309**, 140–152 (2015).

49. van de Weerd, H. A., van Loo, P. L., van Zutphen, L. F., Koolhaas, J. M. & Baumans, V. Nesting material as environmental enrichment has no adverse effects on behavior and physiology of laboratory mice. *Physiol. Behav.* **62**, 1019–1028 (1997).

50. van de Weerd, H. A. *et al.* Effects of environmental enrichment for mice: variation in experimental results. *J. Appl. Anim. Welf. Sci.* **5**, 87–109 (2002).

51. Wolfer, D. P. *et al.* Laboratory animal welfare: cage enrichment and mouse behaviour. *Nature* **432**, 821–822 (2004).

52. Mackay, T. F. C. *et al.* The *Drosophila melanogaster* Genetic Reference Panel. *Nature* **482**, 173 (2012).

53. Stamps, J. A., Saltz, J. B. & Krishnan, V. V. Genotypic differences in behavioural entropy: unpredictable genotypes are composed of unpredictable individuals. *Anim. Behav.* **86**, 641–649 (2013).

54. Scott, J. P. Effects of single genes on the behavior of *Drosophila*. *Am. Nat.* **77**, 184–190 (1943).

55. Carpenter, B. *et al.* Stan: A Probabilistic Programming Language. *J. Stat. Softw.* **76**, 1–32 (2017).

56. Gelman, A., Jakulin, A., Pittau, M. G. & Su, Y.-S. A weakly informative default prior distribution for logistic and other regression models. *Ann. Appl. Stat.* **2**, 1360–1383 (2008).

57. Kruschke, J. K. Bayesian estimation supersedes the t test. *J. Exp. Psychol. Gen.* **142**, 573–603 (2013).

Chapter II - Wild flies hedge their thermal preference bets in response to seasonal fluctuations

I would like to thank all of those that made this project possible. While I designed and carried out the following experimental work, including the data analysis and interpretation, individuals in the de Bivort lab and across the country assisted me carrying out experiments, wild fly collections, and data analysis. Denise Yoon and Anna Soloshenko assisted me in carrying out the Cambridge, Massachusetts, U.S.A. seasonal collections and variability assays. Denise Yoon assisted me in carrying out the heritability experiment. Nikolaos Polizos and Thomas Zhang carried out the seasonal collections and weekly assays in Coral Gables, Florida, U.S.A. Alyssa Bangerter and Matthew Campbell carried out the seasonal collections and weekly assays in Charlottesville, Virginia, U.S.A. Eric Wice (Houston, Texas, U.S.A), Ashley R. Albright (Oakland, California, U.S.A), Aditi Narayanan (Pasadena, California, U.S.A), Paul Schmidt (Philadelphia, Pennsylvania, U.S.A) all contributed wild-caught gravid females to establish isofemale lines. I also thank Nick Keiser (Gainesville, Florida, U.S.A), Alex Keene (Boca Raton, Florida, U.S.A), Sophie Caron (Salt Lake City, Utah, U.S.A), John Tuthill (Seattle, Washington, U.S.A), and Rob Unckless (Lawrence, Kansas, U.S.A) for kindly supplying wild-derived isofemale lines that were not included in the following experiments. Brian J. Arnold and Luisa Pallares provided invaluable assistance with sequencing data analysis and Elena Filippova provided expert help with genomic library preparation. I thank Edward Soucy and Brett Graham of the Center for Brain Science Neurotechnology Core for help with instrument manufacturing and design.

Abstract

Fluctuating environmental pressures challenge organisms by shifting the optimum phenotype. Two contrasting strategies to cope with these pressures are either via evolution of the mean phenotype to follow the optimum (adaptive tracking) or via diversifying phenotypes of individuals to hedge against the fluctuations (bet-hedging). Stable differences in the behavior of individuals are a ubiquitous phenomenon and are present even when genotype and environment are held constant. Instead of being simply “noise”, behavioral individuality may reflect a bet-hedging strategy of phenotype diversification. Using geographically diverse wild-derived fly strains and high-throughput assays of individual preference, we tested whether thermal preference in *Drosophila melanogaster* reflected a bet-hedging strategy and whether populations from different regions preferentially adopt bet-hedging or adaptive tracking strategies. Our modeling predicted regional differences in the advantage of bet-hedging, and in support of that, we found that heritability in thermal preference across six locations is negatively correlated with predicted bet-hedging advantage. We also found that dynamics of mean preference and variability in preference support the existence of bet-hedging in thermal preference. Our empirical results point to bet-hedging in thermal preference as an important evolutionary strategy in wild *Drosophila* populations.

Introduction

Individuals differ in their behavior – these differences can be caused by local environment, age, sex, and genetic variation. There exists a wide range of behavioral differences between individuals that are consistent across environments and time in every species examined[1–5]. The ubiquity of this variability is evidence for its potential ecological and evolutionary importance, which raises the question as to how it arises and is maintained.

Temporal fluctuations in the environment are a way that interindividual differences in a polygenic population can be maintained by selection – at different points in time, environmental pressures will select for a different optimum[6–9]. If interindividual differences are determined by genetic polymorphisms segregating in the population, polymorphism frequencies will change due to adaptation to the new selection pressures, and as a consequence, the behaviors of individuals may also change[10,11]. This process is referred to as adaptive tracking[12]. When these temporal fluctuations are relatively rapid, the population lags behind the selective pressures, which can lead to lower relative fitness[13–15]. In this case, it has been proposed that it is advantageous to reduce the phenotype expression of the segregating variants, thereby minimizing the effect of previous selection in the current environment[15].

Diversifying bet-hedging (from here on out referred to as bet-hedging) is an alternative strategy that overcomes the limitations faced by adaptive tracking in rapidly fluctuating environments, because it ensures that there will always be fit individuals should a drastic or unpredictable environmental change occur[7,16–18]. Under this strategy, a single genotype produces multiple phenotypes as a way to mitigate risk i.e. “don’t put all your eggs in one basket”. This bet-hedging strategy reduces fitness variance between generations, increasing geometric mean fitness at the expense of arithmetic mean fitness[16,19–21]. Intuitively, this means that although in a single generation a bet-hedger may not be optimally fit given the environment, the stability of

fitness over generations (a bet-hedger potentially always having decent fitness no matter the circumstances) results in success for the strategy over the long term. As environmental variance increases, bet-hedging becomes an evolutionarily optimal strategy that can explain the maintenance of interindividual differences[22–26]. There is a variety of evidence for bet-hedging traits across organisms, though there are few examples of bet-hedging in behavioral traits[12].

We expect that individuals from a bet-hedging genotype would exhibit stable idiosyncratic behavioral biases, even reared in the same environment. Work in *Drosophila melanogaster* has shown that in a constant environment and genetic background, individuals exhibit consistent, but starkly different, behaviors - examples include turning bias[27,28], phototaxis[29], and thermal preference[30]. These differences are termed intragenotypic variability, and they can reflect a bet-hedging strategy. Kain *et al*. used a model that translated observed phototaxis and thermal preference intragenotypic variability into simulated variability in life history, under either a bet-hedging or adaptive tracking strategy in natural populations of *Drosophila melanogaster*[30]*.* They found that bet-hedging is more advantageous than adaptive tracking when in environments with a high variance in seasonal temperatures and a short breeding season. Despite this effort, empirical, rather than computational evidence that animals use bet-hedging strategies, particularly with respect to behavior, is scarce.

We present findings from empirical tests of the predictions made by the Kain *et al.* model to test the hypothesis that thermal preference in *Drosophila melanogaster* follows a bet-hedging strategy and that bet-hedging and adaptive tracking strategies are geographically determined. Measuring the thermal preferences of many individual flies collected wild from multiple geographic sites, we show that 1) patterns of mean thermal preference over the season support the existence of a bet-hedging strategy, 2) variability across locations doesn't correlate with their

predicted bet-hedging advantage, and 3) heritability of thermal preference is inversely correlated with the predicted bet-hedging advantage of a location.

Results

Individual thermal preference is idiosyncratic and persistent

If individual variation in thermal preference behavior reflects an evolutionary bet-hedging strategy, individual preferences would represent stable phenotypes. As such, they would be quantifiable and stable over time. We created a two-choice preference assay to measure individual thermal preferences of many flies in parallel (Figure 2.1a). We measured an individual fly's thermal preference as the average of time spent on the hot and cold sides of the assay (Figure 2.1b). We compared the observed distribution of thermal preferences within a wild-type line (isogenic animals raised in a temperature-controlled incubator) to a null distribution in which all flies have identical preferences and sampling error alone generates dispersion in the measurement of that distribution (null distribution was generated by bootstrap resampling of walking bouts). The observed distribution was significantly broader than the null distribution (Kolmogorov-Smirnov test, p = 9.1E-9 ; Figure 2.1c), indicating that thermal preferences are overdispersed, i.e., the flies exhibited individuality in thermal preference. We also observed significant positive correlation (Figure 2.1d; r = 0.50, p = 6.2E-5) in thermal preference measured on consecutive days, consistent with individual measured thermal preferences representing stable phenotypes.

Life history modeling predicts that the adaptive value of bet-hedging varies geographically

With this new, high-throughput method for measuring thermal preference, we sought to update a temperature-dependent life history model of fly development and mortality from Kain *et al*.[30] (Figure 2.1e). This model estimates the dynamics of populations implementing pure bet-hedging or adaptive tracking strategies, under specific climate conditions. We updated the model in two ways: 1) we updated the temperature-dependent life history equations with fits to new data collected from isofemale lines established using wild-caught females (Figure 2.1e) and 2) we implemented more realistic rules to convert individual thermal preference, in combination with

the range of temperatures available on a given day, to individual thermal experience (See Methods). With these improvements, we recalculated the predicted bet-hedging advantage (Figure 2.1f) across the continental U.S.A. and Puerto Rico using the 1981-2010 climate normals from 7112 weather stations (U.S. National Oceanic and Atmospheric Administration [NOAA][31]). To make predictions at sites between weather stations, and take into account local dispersal of flies, we employed a Gaussian convolution on the station data with a standard deviation of 0.04 decimal degrees (equivalent to 3-4km depending on the latitude). We chose this value based on empirical data from a release and recapture study[32] that found appreciable frequencies (~15%) of marked flies 3-6km from the release site.

Bet-hedging advantage was calculated as the natural log of the final size of the bet-hedging population divided by the final size of the adaptive tracking population. Overall, the predictions of this updated model are in qualitative agreement with the original model: we predicted that bet-hedging advantage is higher in the north. In the eastern half of the U.S.A., we predicted a decline in bet-hedging advantage towards the south, while in the western half, bet-hedging advantage persisted in the southern latitudes. The most adaptive-tracking favored (least bet-hedging favored) regions were in the south and south-east, along the Gulf Coast, Texas, Florida, and Puerto Rico. It is notable that within these regions, we still predicted heterogeneity in bet-hedging advantage, likely due to local microclimates.

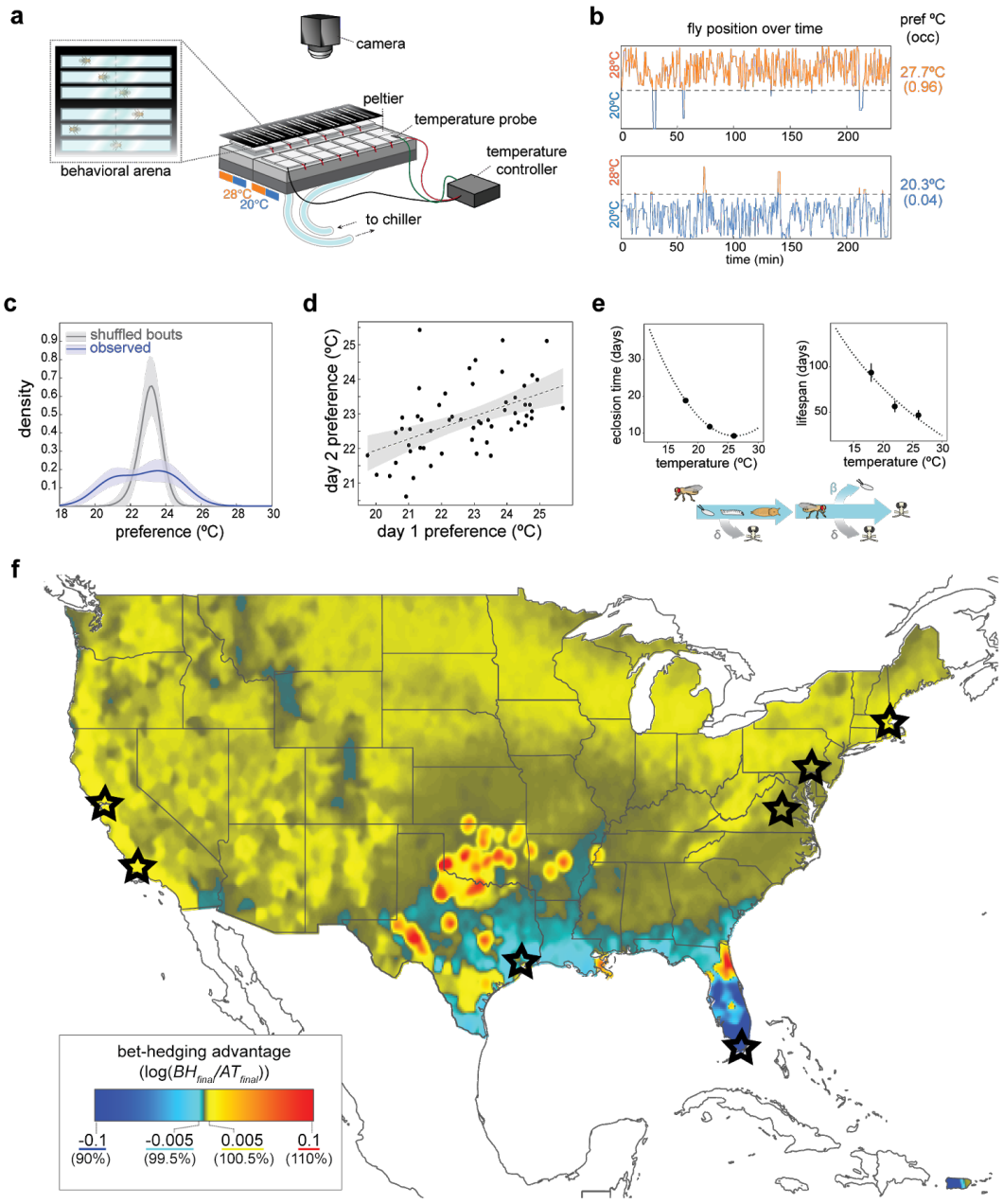


Figure 2.1 - Individual thermal preference is persistent over time in isogenic and wild *D. melanogaster* populations.

a) Diagram of the automated two-choice assay used to measure individual thermal preference. **b)** Examples of position over time in the assay for extreme cold- and warm-preferring flies. Thermal preference is calculated as the proportion of trial time spent on the hot side of the tunnel. ºC metric is the average temperature experienced. **c)** Observed (blue) vs null distribution (gray) of thermal preference for an isogenic line. Null distribution is created by simulating single flies by randomly assembling bouts from the hot and cold sides to build a 4-hour experiment. Solid lines show the observed data kernel-density estimate or the mean kernel density estimate of 100 simulated null distributions. Shaded areas show +/- 2 s.d. of kernel-density estimates of 100 bootstrap resamples of observed data (n=57) or 100 simulated null distributions. **d)** Persistence of thermal preference over 24hrs in an isogenic line (n = 57). Shaded area shows the 95% CI of a linear fit to the data.

Figure 2.1 (continued) - e) (top) Life history vs. temperature relationships used in the bet-hedging vs. adaptive tracking model. Error bars show the 95% confidence interval of the mean. **(bottom)** Temperature dependent life history model used to simulate fly population evolution: β, birth rate; δ, death rate; *M*, metamorphosis time; *A*, adult life span; *T*, thermal preference index. "Fly skull and crossbones" icons indicate death (adapted with permission from Kain *et al.*[30]). **f)** Map of bet-hedging advantage across the continental U.S.A. and Puerto Rico calculated using a Gaussian convolution on the predicted bet-hedging advantage at 7112 weather stations ($\mu = 0.44$, $\sigma^2 = 0.015$ for all stations). Sampling locations are overlaid in black stars.

Seasonal patterns in mean thermal preference are consistent with a bet-hedging strategy

We hypothesized that seasonal patterns in mean thermal preference would reflect the bet-hedging advantage of the particular locale, with locations where bet-hedging is predicted to be advantageous exhibiting a more constant mean preference (phenotypic variation in strictly bet-hedging traits is not heritable). To test our hypothesis, we assayed wild-caught flies weekly over from late June to late October/early November in Cambridge, Massachusetts, U.S.A. (MA), Charlottesville, Virginia, U.S.A. (VA), and Coral Gables, Florida, U.S.A. (FL). Flies were captured in residential areas of Cambridge and Coral Gables and in an orchard on the outskirts of Charlottesville; assays were performed in a laboratory environment (Figure 2.2, Figure 2.3a).

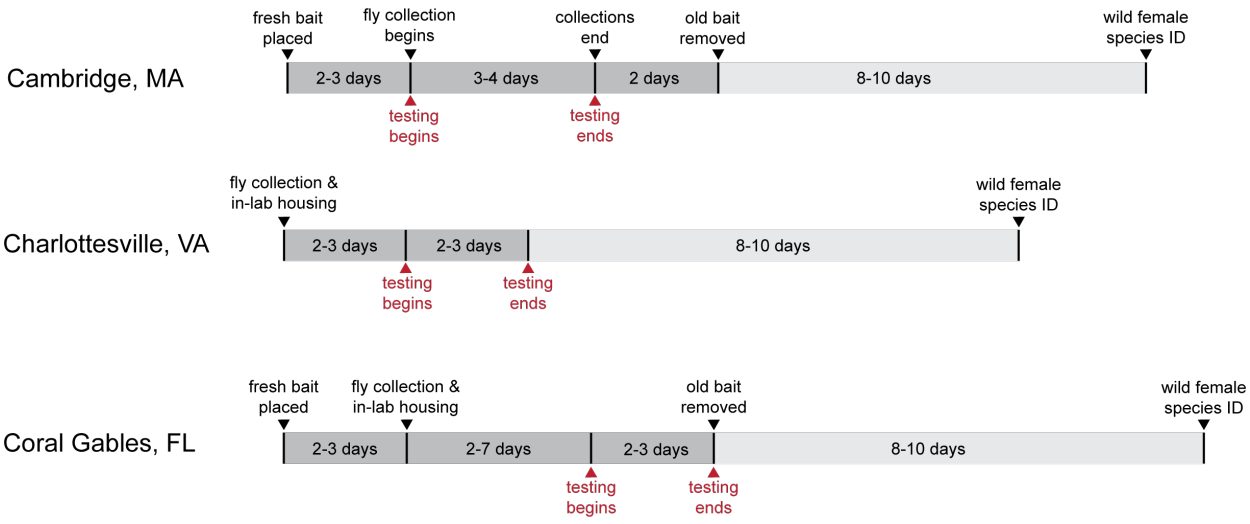


Figure 2.2 - Experimental timelines for seasonal collections across the three locations. Each timeline shows the process of collection and testing for flies collected from a particular week.

We chose these locations due to their differences in predicted bet-hedging advantage - Cambridge is predicted to be the most bet-hedging advantageous (BH advantage = 0.0042, corresponding to an annual growth advantage of 0.42% for a bet-hedging strategy), followed by Charlottesville (BH advantage = 0.0020 or annual growth advantage of 0.20% for a bet-hedging strategy), while Coral Gables is strongly favored for adaptive tracking (BH advantage = -0.54 or annual growth disadvantage of 58% for a bet-hedging strategy) (Figure 2.1f). To test whether the seasonal patterns in mean preference followed bet-hedging or adaptive tracking predictions, we plugged daily temperature data from 2018 into our model to generate site-and-year-specific predicted patterns in mean preference for a purely bet-hedging or a purely adaptive tracking population. We calculated a log-likelihood ratio to gauge whether the observed dynamics are more likely under a bet-hedging or an adaptive tracking strategy.

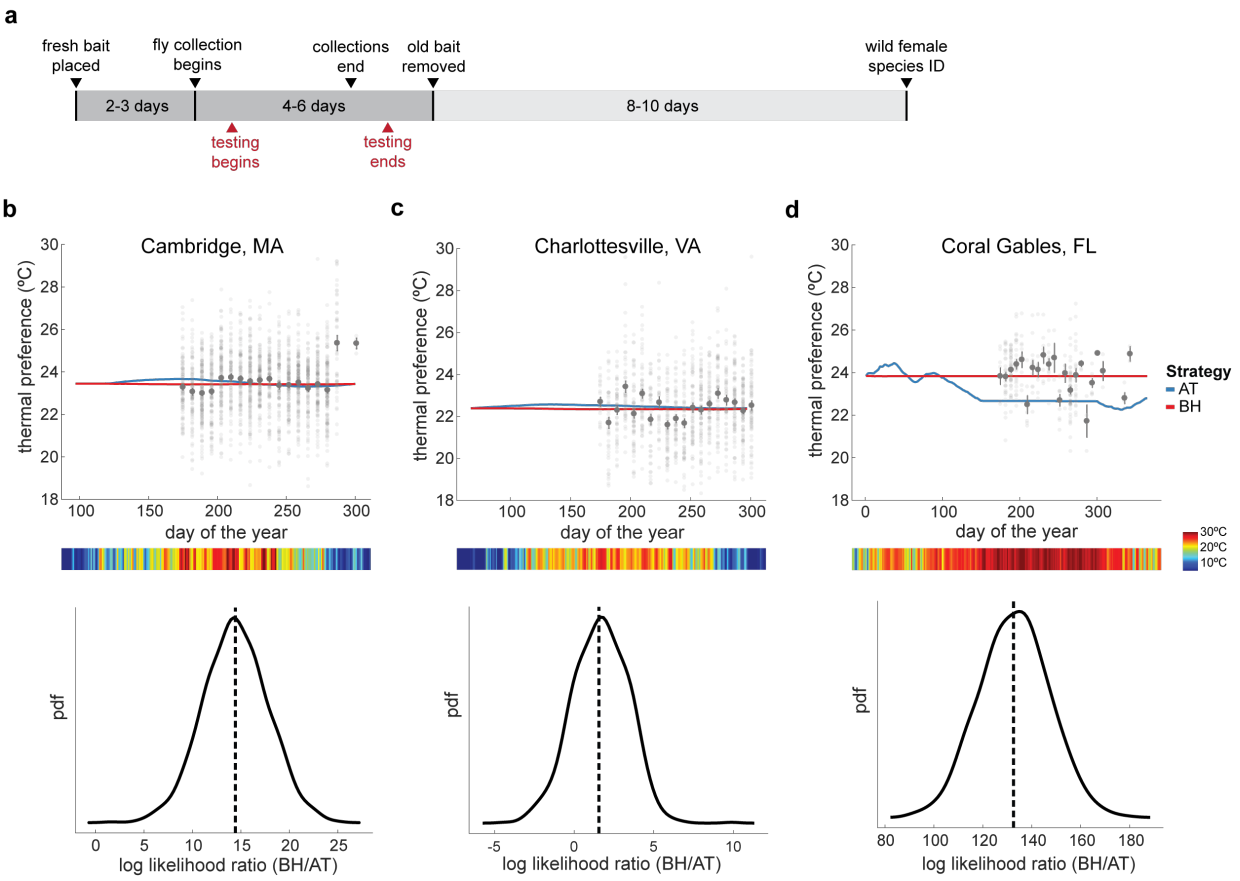


Figure 2.3 - Dynamics of mean thermal preference over the course of the fly breeding season reflect a bet-hedging strategy in both bet-hedging and adaptive tracking favored regions.
a) General experimental timeline for the seasonal sampling - for further detail see Figure 2.2. **b) (top)** Flies were collected in residential neighborhoods near Harvard University, Cambridge, MA. Solid lines show the predicted mean thermal preference under bet-hedging (BH, red) and adaptive tracking (AT, blue) modeling predictions. To create these predictions, we used 2018 daily average temperatures[31] from April to November and empirically determined $\mu = 0.45$, $\sigma^2 = 0.016$. The light gray points represent the thermal preference of individual flies caught, with the dark gray point and error representing the mean +/- 1 S.E of that week's collection. The heatmap at the bottom of the plot shows the temperature of each day in the breeding season. **(bottom)** Dashed line shows the log likelihood ratio (BH/AT) of the observed data, while the kernel-density estimate shows the log likelihood ratios calculated from 1000 bootstrap resamples of the observed data. **c)** Flies collected in Carter Mountain Orchard, Charlottesville, VA. Top and bottom panels are the same as in (b) with the fly breeding season established to be from March until November and $\mu = 0.37$, $\sigma^2 = 0.019$. **d)** Flies collected in residential neighborhoods next to University of Miami, Coral Gables, FL. Top and bottom panels are the same as (b) with $\mu = 0.49$, $\sigma^2 = 0.010$.

We found that in Cambridge, MA, the log-likelihood ratio supported bet-hedging over adaptive tracking (log-likelihood ratio [LLR] = 14.43; Figure 2.3b). These findings were in line with our modeling predictions for Cambridge, MA being a bet-hedging favored locale. In Charlottesville, VA, the observed data were still more likely under a bet-hedging strategy, though the likelihood of a bet-hedging strategy was less strong than in Cambridge, MA (LLR = 1.56; Figure 2.3c). In agreement with the observed dynamics, Charlottesville was predicted by our model to be bet-hedging favored, though less so than Cambridge. In Coral Gables, FL, the observed dynamics of mean thermal preference are much more likely under the bet-hedging than the adaptive tracking model (LLR = 132.90; Figure 2.3d). Our model predicts that flies in Coral Gables will exhibit adaptive tracking. However, we observed mean temperature preference dynamics that are more likely under a bet-hedging strategy. The predicted mean preference dynamics for Coral Gables show a clear selection for colder thermal preference, but this was not observed empirically. Interestingly, the model predicted a steep decline in total population during the hottest part of the summer - with the exception of a couple of weeks, we observed a sharp drop in the number of collected *D. melanogaster* during this time (Figure 2.4).

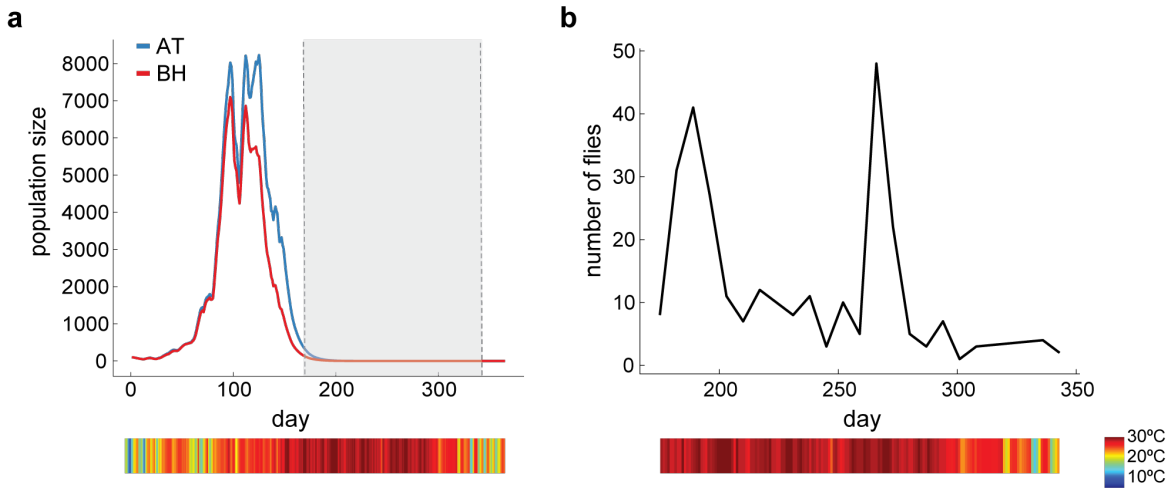


Figure 2.4 - Predicted population crash in Coral Gables, FL coincides with low numbers of caught *D. melanogaster*.
a) Predicted population sizes under adaptive tracking (AT) and bet-hedging (BH) strategies for the entire 2018 Coral Gables, FL breeding season. Gray area denotes the collection period. **b)** The number of collected *D. melanogaster* over the 20-week period from June to November 2018. The heatmaps at the bottom of the plots show the daily temperatures.

We assessed the potential role of developmental temperature on the dynamics of mean preference we observed. We reared offspring from a single isofemale line at 18ºC, 22ºC, and 26ºC from egg to adulthood. We found that rearing temperature had a minimal and non-linear effect on the preference mean or variance (Figure 2.5), leading us to believe that plasticity in thermal preference is not a driving force for the patterns observed.

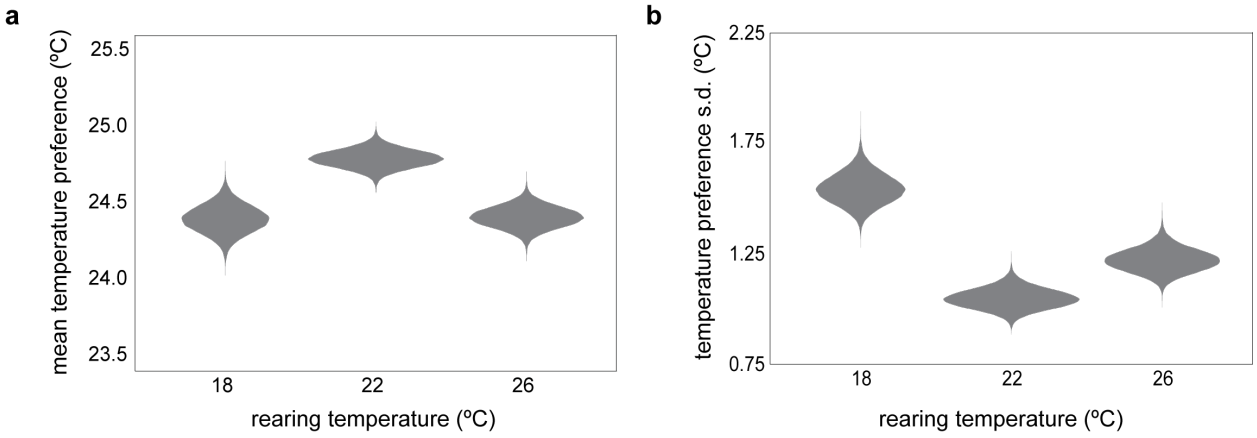


Figure 2.5 - Plasticity in mean and standard deviation of thermal preference.
a) Posterior estimates of mean temperature preference under three different rearing temperatures (18ºC: n=134; 22ºC: n=166; 26ºC: n=145). **b)** Posterior estimates of variability (s.d.) under three different rearing temperatures.

Variability in thermal preference differs among isofemale lines, but does not correlate with predicted bet-hedging advantage

When comparing phenotypic variability within a single genotype, one might expect that a bet-hedging genotype would show higher variability than an adaptive tracking genotype, as the adaptive tracking genotype may have been subject to purifying selection. Under adaptive tracking, deviations from the mean phenotype are maladaptive if a genotype is adapted to the current environment. Phenotypic diversity for a bet-hedging genotype is an essential part of the diversification strategy to avoid risk, and therefore would be maintained. Therefore, we hypothesized that variability in thermal preference would be higher in locales where the bet-hedging strategy is predicted to be advantageous. We established isofemale lines from gravid females sampled from seven locations across the U.S.A. (three of which were used for the weekly seasonal sampling experiment). We measured variability of each isofemale line as the standard deviation in thermal preference (Figure 2.6a). Using Bayesian inference on a hierarchical model which treated isofemale lines as nested within their respective location (population), we estimated a posterior distribution for the variability in thermal preference for individual lines as well as for the overall sampling location (Figure 2.6b). Variability of a particular location shows no correlation (r = -0.04, p = 0.92) to the predicted bet-hedging advantage (Figure 2.6c). We did observe strong line-by-line differences in variability in the isofemale lines measured. We also found that the largest variability estimate was from Pennsylvania, U.S.A. (PA). Since we were not using isogenic lines for this analysis, we examined if there was a correlation between variability and genetic diversity, estimated as Watterson's θ_s. We found no significant correlation between θ_s measured at the level of isofemale lines (r = 0.14, p = 0.55) and at the level of location (r = 0.027, p = 0.95) and the corresponding variability estimate (Figure 2.7a-c).

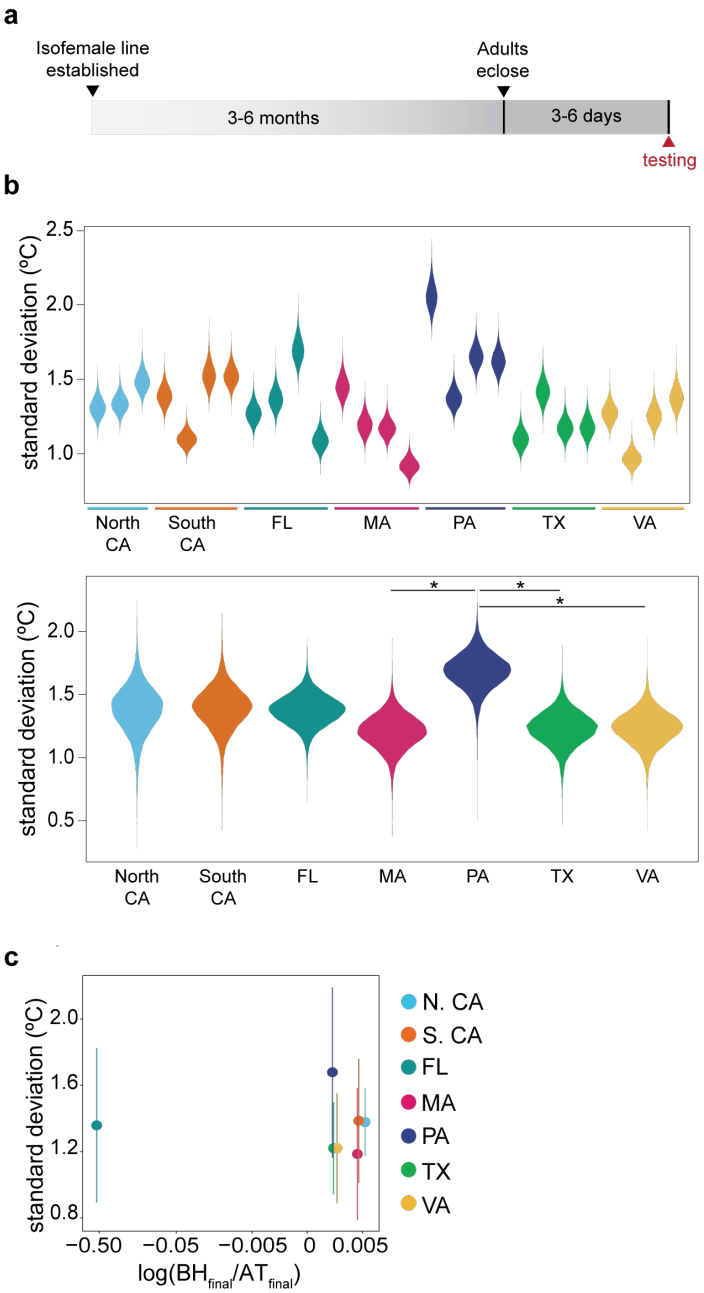


Figure 2.6 - Variability of thermal preference in geographically diverse isofemale lines does not show a relationship with the bet-hedging advantage of the geographic origin but does show line-to-line differences.

a) Experimental timeline for assessing variability. **b) (top)** Posterior distributions of thermal preference standard deviation for isofemale lines chosen from each sampling location (origin). Posterior distributions were generated using a Bayesian modeling approach that takes into account the nested nature of the dataset (lines come from a single origin) and "sampling error" (inflated variance measures that come from inactivity in the assay). **(bottom)** Posterior distributions of standard deviation for each sampling location (origin). Asterisks indicate where the 95% credible interval of the difference between locations does not include 0. **c)** Variability within a location vs. the bet-hedging advantage estimate for that location. Vertical error bars show +/- 2 standard deviations of the variability posterior distribution.

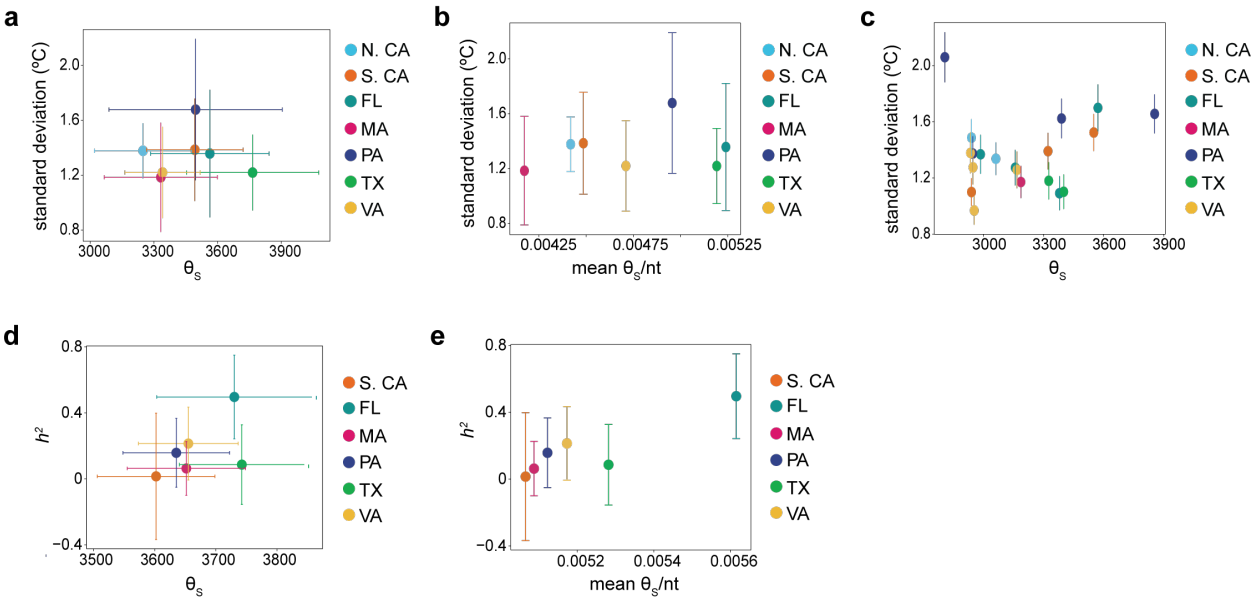


Figure 2.7 – Relationship of population genetic diversity compared with variability and heritability. **a)** Resampling estimate of θ_s within a location vs. the variability estimate for that location (n = 7; r = 0.027, p = 0.95). Vertical error bars show +/- 2 s.d. of the variability posterior distribution and horizontal error bars show +/- 2 s.d. of the bootstrap distribution of θ_s. **b)** Average PoPoolation estimate of θ_s/nt within a location vs. the variability estimate (r = 0.23, p = 0.62). **c)** θ_s within an isofemale line vs. the variability estimate of that line (n = 20; r = 0.14, p = 0.55). Vertical error bars show +/- 2 s.d. of the variability posterior distribution. **d)** Relationship between heritability (h^2) and resampling estimate of θ_s (n=6; r = 0.54, p = 0.27) Vertical error bars show the 95% confidence interval on the h^2 estimate (regression slope) and horizontal error bars show +/- 2 s.d. of the bootstrap distribution of the θ_s estimate. **e)** Average PoPoolation estimate of θ_s/nt within a location vs. the h^2 estimate (r = 0.90, p = 0.015). Removing the Coral Gables, FL data point diminished the positive correlation (r = 0.33, p = 0.59).

Geographic variation in heritability of thermal preference is consistent with predicted bet-hedging advantage

An integral aspect of bet-hedging is that phenotypic variation does not arise due to genetic variation. Therefore, we hypothesized that heritability of thermal preference would be higher in those locations that are predicted to follow the adaptive tracking strategy. We used isofemale lines from six locations to perform midparent-offspring regression in order to measure the narrow-sense heritability (h^2) (Figure 2.8a). We found that the highest heritability was in Coral Gables, FL (h^2 = 0.50, 95%CI = [0.24, 0.75]; Figure 2.8b). Across all sites, h^2 is inversely correlated with the predicted bet-hedging advantage of the geographic origin. Sites predicted to favor adaptive tracking have higher thermal preference heritability than sites predicted to favor

bet-hedging (r = -0.90, p = 0.011; Figure 2.8c). Removing the Coral Gables, FL data point still produced a negative correlation, though the magnitude was smaller and the correlation was no longer significant (r = -0.75, p = 0.14). There was no significant positive correlation of h^2 with the θ_s estimate (r = 0.54, p = 0.27) calculated using our bootstrap resampling method (see Methods), but there was a significant positive correlation using the PoPoolation θ_s estimate (r = 0.90, p = 0.015) (Figure 2.7d,e). The significance of the correlation between h^2 and θ_s was highly influenced by the presence of the Coral Gables, FL data point (r = 0.33, p = 0.59 with FL data point removed).

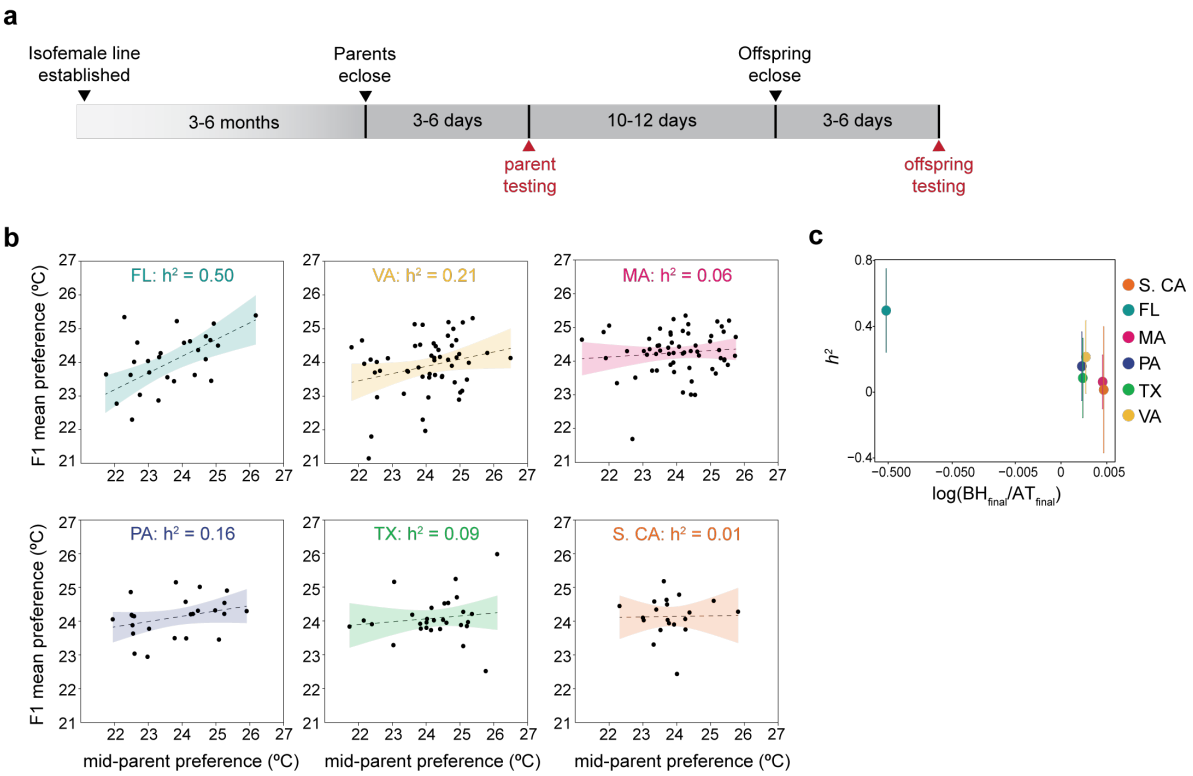


Figure 2.8 - Heritability (h^2) of thermal preference varies with geographic origin and is correlated with the origin's bet-hedging advantage.
a) Experimental timeline for assessing heritability. **b)** h^2, as measured by the slope of the midparent-offspring regression, varies with geographic origin of the isofemale lines. Coral Gables, FL had the only slope that was significantly different from zero (p = 4.1E-4), though Charlottesville, VA was close to significant (p = 0.057). **c)** Correlation between h^2 and predicted bet-hedging advantage of the location. Vertical error bars show the 95%CI on the heritability estimate (95%CI of the parent-offspring regression slope).

Discussion

Bet-hedging, in which a single genotype produces a distribution of phenotypes to maintain fitness in the face of fluctuating selection, is a potential explanation of observed inter-individual phenotypic variability, including behavioral variability[16,30]. While this framework has strong theoretical foundations, empirical evidence, particularly in animals and with respect to behavioral phenotypes is lacking. The goal of our study was to help close this gap by examining whether thermal preference variability in *Drosophila melanogaster* reflects either a bet-hedging or adaptive-tracking strategy depending on regional climate. We found that dynamics of mean preference were consistent with bet-hedging, regardless of geographic prediction, variability varied by line, but not with geographic prediction, and that heritability of thermal preference varied geographically, in agreement with the predictions of the life history model. Overall, we find this to be strong empirical evidence that bet-hedging underlies thermal preference variability in *Drosophila*.

More specifically, we observed persistent individual thermal preferences that do not depend on genetic or macro-environmental differences, showing that thermal preference is a potential bet-hedging trait. Modeling the breeding seasons of purely bet-hedging and adaptive-tracking fly populations using local climate data across the U.S.A., we predicted the existence of regional differences in preferred strategy. We predicted that for most of the contiguous U.S., bet-hedging strategy has an advantage over adaptive tracking. Adaptive tracking was a favored strategy only in the deep south of the Gulf Coast (south-eastern U.S.A.), where mild temperatures allow for year-round breeding seasons, as well as slow and mild seasonal fluctuations that give time for populations to track the phenotypic optimum. We did not observe similar predictions in the south-west, a region also marked by warm weather, though with cooler average temperatures and/or higher temperature variance than in the Gulf Coast. These patterns supported our

previous findings that bet-hedging advantage depended on both average seasonal temperature and standard deviation[30], with bet-hedging at a disadvantage in very warm and stable seasons.

Given our predictions of regional differences in strategy, we collected wild flies both across different sites and seasons to test for the presence of a bet-hedging strategy using 1) dynamics of mean thermal preference, 2) variability in thermal preference, and 3) heritability of thermal preference. We found that heritability of thermal preference depended on the geographic population sampled. Low thermal preference heritability was observed in the populations collected from sites predicted to be favored for bet-hedging by the life history model, while high heritability was observed in the FL population that was collected from a site predicted to be highly favored for adaptive-tracking (Figure 2.8b,c). We found that temporal patterns in mean preference over a 20-week period for MA, VA, and FL sampling sites supported the bet-hedging strategy, which aligned with our life history model bet-hedging predictions for MA and VA, but not FL. In contrast, variability in thermal preference was not associated with predicted bet-hedging advantage.

In Coral Gables, FL, the finding that dynamics of mean preference supported the bet-hedging strategy was at odds with our estimate of high heritability in thermal preference, which supports an adaptive-tracking strategy (variation between individuals in a strictly bet-hedging trait is non-genetic) (Figure 2.3d, Figure 2.8b). The predicted mean preference pattern under a purely adaptive-tracking strategy for Coral Gables was different than in Cambridge, MA and Charlottesville, VA, showing more extreme fluctuations and a shift of the mean preference to colder temperatures (Figure 2.3d). The year-round high temperatures in Coral Gables are the likely culprit, resulting in strong selection for cold thermal preference in our model under the adaptive tracking strategy. We had lower sampling success in FL, as compared to MA and VA, which remained low over the course of the collection period as the high daily temperatures

persisted (Figure 2.4). The observations of *D. melanogaster* abundance and the modeling predictions together suggest that the high temperatures could have led to a decrease in the available population through strong selection against warm-preferring flies. Since our traps were set outside, it is possible that we failed to capture a representative sample if the majority of the fly population (potentially consisting of cool-preferring flies) had migrated into human residences. This biased sampling could potentially lead to the appearance of a "bet-hedging" pattern in the Coral Gables fly population. As an alternative hypothesis, we note that *Drosophila simulans* repopulates the northern limits of its range via seasonal migration[33]. The existence of a corresponding seasonal migratory repopulation in *D. melanogaster* could also explain these sampling observations.

Even though we found that there was little difference between sites in thermal preference variability, we observed a high level of heterogeneity in thermal preference variability across the isofemale lines. Differences in variability across lines is consistent with there being a genetic component to variability, previously noted in fruit fly turn bias behavior[28]. A genetic basis for variability may be indicative of a bet-hedging trait, as optimal levels of variability could be selected for[19,21–24]. Interestingly, we also observed that rearing temperature caused plasticity in the variability of an isofemale line (Figure 2.5). We previously found plasticity in variability of several locomotion and phototaxis behavior metrics when comparing flies that were raised in standard and enriched food vials[34]. Environmental plasticity in variability, in addition to a genetic contribution to variability, indicates that both genetics and environment could play a role in determining the degree of variability of behavioral traits.

The lack of relationship between the estimated variability of a site and the predicted bet-hedging advantage suggests that the difference in optimal variability under bet-hedging and adaptive tracking may be low. Measuring variability on its own may not be diagnostic as to which

evolutionary strategy a population is employing. At a population-level, we observed consistency in the level of variability, with all but one population having a similar level of variability (s.d. 1.2-1.4ºC). PA's variability estimate was biased strongly upward by one isofemale line (Figure 2.6b), which was influential in a limited sample size. Since we identified differences in thermal preference heritability among our populations, we propose that the underlying differences between individuals may be primarily due to stochastic microenvironmental forces for populations with lower heritability and allelic variation for thermal preference in populations with higher heritability.

When examining thermal preference heritability and variability, we were cognizant that genetic diversity could affect our estimates. We evaluated genome-wide levels of variation (Watterson's θ_s) using individuals from isofemale lines with a custom bootstrapping approach in addition to PoPoolation software. For both methods of estimating θ_s, we did not see a strong relationship with variability, though we observed a stronger positive relationship with heritability (Figure 2.7). A positive relationship with heritability is likely due to both heritability and θ_s following the same geographical patterns, with higher heritability and θ_s in our sampled south-eastern populations (TX, FL) as opposed to the northern populations. Higher θ_s in southern latitudes has been found in previous studies[35–37]. Coral Gables, FL was an influential point in the relationship between θ_s and heritability, given that it was the only sampling site where we observed a strong heritability of thermal preference. We do not claim that there is a causal relationship between θ_s and heritability - we would guess that the relationship between them is likely mediated through geography.

Overall, our investigations provide evidence for bet-hedging in thermal preference, showing 1) high levels of non-genetic individual differences within lines and a genetic basis for the degree of variability across lines, 2) seasonal mean preference patterns consistent with bet-hedging

predictions, and 3) low trait heritability. Strikingly, our heritability results also align with our modeling predictions of regional differences in adaptive-tracking vs. bet-hedging strategy advantage, with the adaptive-tracking favored region showing high heritability in thermal preference, but the bet-hedging favored regions showing low heritability. Our findings put behavioral individuality into an ecological and evolutionary context, showing that putative “noise” may reflect an adaptive strategy for dealing with risky environments.

Methods

Fly husbandry

Unless otherwise stated, all stocks and Isofemale lines were maintained at 22-23ºC and 45% relative humidity (RH) in temperature controlled incubators in 12L:12D conditions on a yeast, cornmeal, and dextrose media (23g yeast/L, 30g cornmeal/L, 110g dextrose/L, 6.4g agar/L, and 0.12% Tegosept).

Thermal preference assay

A two-choice assay was created to assay thermal preference, where occupancy (time spent) was the metric of an individual's thermal preference. The behavioral assay consisted of 20 tunnels, with each half of the set of tunnels under independent temperature control. Peltier elements (Custom Thermoelectric 12711-5L31-09CQ) were wired in series and put under PID control by either a commercial (AccuThermo FTC100D) or a custom Arduino-based temperature controller to create one-half of an arena. Two Peltier sets under independent control were put across from each other to create a full arena. This design gave the ability to precisely control temperature, as well as to switch the orientations of the cold and hot sides. Unless otherwise stated, the setpoint for the cold side was kept at 20ºC and the set point for the hot side was kept at 28ºC. The set points were chosen to be within the fly's innocuous temperature range, so as to avoid activating any noxious stimuli receptors[38], as well as being amenable to an hours-long experiment where an excessively high temperature can lead to desiccation or an excessively low temperature can lead to cessation of movement.

In order to measure an individual's occupancy over time, single flies were placed into each tunnel and allowed to explore for 4 hours. Their positions were monitored and recorded under far infrared lighting (940nm) in an enclosed box using the beta version of the Massively

Automated Real-time GUI for Object-tracking software[39] in MATLAB 2018a (Mathworks, Inc). Two behavioral arenas were simultaneously tracked by one camera, resulting in 40 flies assayed in a single camera box. Three boxes were set up to work in parallel to facilitate higher throughput.

An initial thermal preference metric ranging from 0 to 1 was calculated for each individual at the end of the trial by measuring the proportion of time spent on the hot side over the total trial time. In order to correct for any bias induced by long periods of inactivity, pauses longer than 5 minutes were filtered out from the tracks and the thermal preference was recalculated. Flies which had less than 1 hour total activity throughout the trial post-filtering were removed from further analysis. Since minor temperature differences were observed among the Peltier elements at the given setpoints, a tunnel temperature correction was applied to the 0-1 metric. The tunnel correction gives a thermal preference metric in ºC, which translates to the average of time spent at the cold and hot temperatures.

Bet-hedging and adaptive tracking model

Predictions for seasonal patterns in mean thermal preference and calculations of bet-hedging advantage were done using a modified modeling approach based on Kain *et al.*[30] In brief, the approach in Kain *et al.* used a system of difference equations coupled to empirically determined relationships between temperature and development time/lifespan to model fly populations over a breeding season. Thermal preference within each generation of the modeled fly populations could be determined by one of two pure strategies: bet-hedging (no heritability in thermal preference, thermal preference of the new generation determined by sampling from a beta distribution) or adaptive tracking (heritability of 1, thermal preference of offspring is determined entirely by parents). The model also incorporated a birth rate (β) for the new generation and a death rate (δ) from random events unrelated to thermal experience. These were calibrated

under the adaptive tracking strategy using a hill-climbing algorithm to ensure that 1) the initial population size matched the final population size at the end of the breeding season and 2) the mean preference at the start of the season matched the mean preference at the end of the season. These constraints were chosen to ensure that the population is at equilibrium and not evolving across breeding seasons. Using climate normals, a breeding season was established to start when the temperatures exceed 6.5ºC and end when they drop below 10ºC. To apply these thresholds to daily temperature measurements, a two-month moving average window was used to smooth the daily fluctuations before applying the threshold.

Two aspects of the original model were modified here: 1) determination of thermal experience given a thermal preference and the available environmental temperature range and 2) empirically determined relationships between temperature and development time/lifespan. Under the Kain *et al.* approach, thermal experience (τ) of fly *i* on day *j* was calculated as:

$$\tau(i,j) = pref_i * 7ºC * cloudCover_j + temp_j$$

Where $pref_i$ is thermal preference of fly *i* (0-1 scale), 7ºC is a typical empirical difference between sun and shade temperatures, $cloudCover_j$ is the fraction of cloud cover on day *j,* and $temp_j$ is the average in-shade temperature on day *j*[30]. This coding of thermal experience produces a 7ºC difference in thermal experience between flies at the thermal preference extremes, without consideration of flies avoiding noxiously hot and cold temperatures[38] (Figure 2.9d). Under our approach, thermal experience was coded as a piecewise function to avoid instances where flies with extreme thermal preferences would be experiencing noxious temperatures.

$$\tau(i,j) = (1-a) * temp_j + a * \left(temp_j + 3.5ºC * cloudCover_j\right):$$
$$temp_j \geq pref_i$$
$$\tau(i,j) = (1-a) * \left(temp_j + 7ºC * cloudCover_j\right) + a * \left(temp_j + 3.5ºC * cloudCover_j\right):$$
$$temp_j + 7ºC \leq pref_i$$
$$\tau(i,j) = (1-a) * pref_i + a * \left(temp_j + 3.5ºC * cloudCover_j\right):$$
$$temp_j < pref_i < temp_j + 7ºC$$

Where $temp_j$ and $cloudCover_j$ are as above, $pref_i$ is thermal preference of fly *i* in ºC, and *a* is a value between 0 and 1. $Pref_i$ ranges between 18ºC and 30ºC, the limits of the thermal preference assay in Kain *et al.* The new formula specifies that when the preference of fly *i* in ºC is between the in-shade and in-sun temperatures of day *j,* the fly's thermal experience is a combination of the proportion of time spent at the thermal preference temperature, $pref_i$ (thermoregulation), and proportion of time spent at the average daily temperature, $temp_j$ + 3.5ºC x $cloudCover_j$ (non-thermoregulated activities, such as predator avoidance or foraging). When the in-sun temperature for day *j*, $temp_j$ + 7ºC x $cloudCover_j$, is less than or equal to the fly's thermal preference, the fly will spend the thermoregulation portion of the day at the in-sun temperature (maximum temperature it can achieve). When the in-shade temperature, $temp_j$, is greater than or equal to the fly's thermal preference, the fly will spend the thermoregulation portion of the day at the in-shade temperature (minimum temperature it can achieve). For a sun vs. shade temperature difference of 7ºC, an *a* of 0.4 was chosen, signifying that flies spend 60% of their day doing thermoregulated activities. An *a* of 0.4 was chosen so that the maximum standard deviation in thermal experience over the entire breeding season would be ~1.5ºC, the average measured standard deviation in the laboratory thermal assay (Figure 2.9, Figure 2.6b). The relative bet-hedging advantages calculated from the model are robust to different *a* values (Table 2.1).

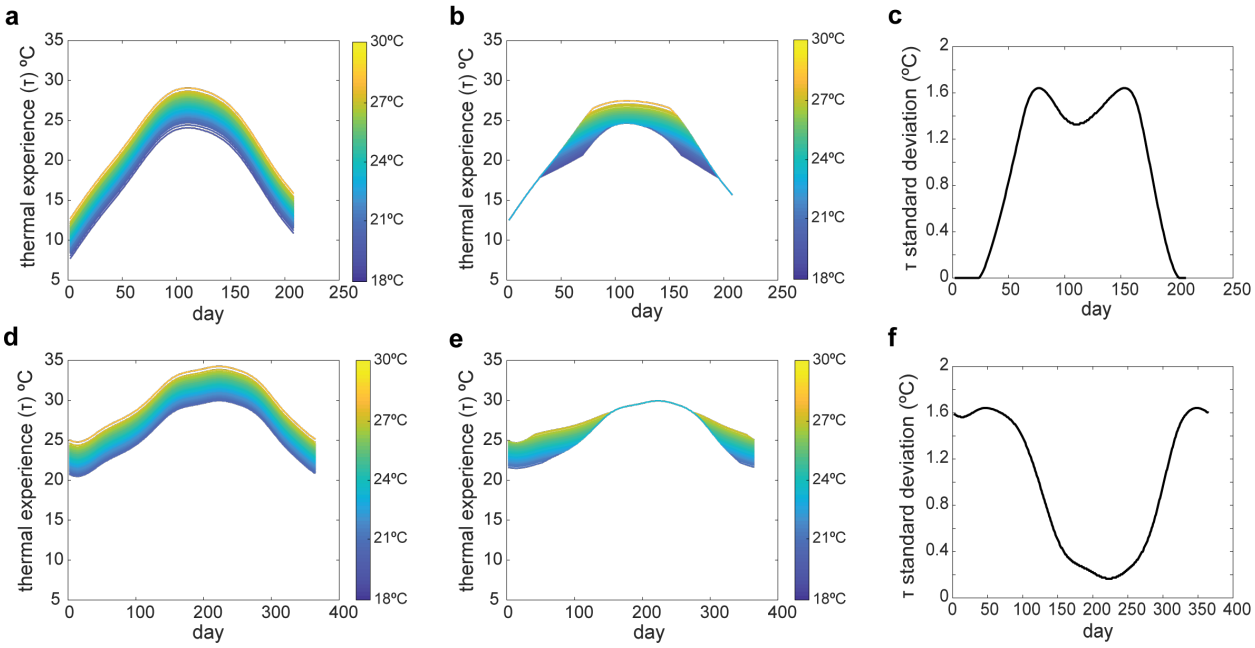


Figure 2.9 - Relationships between thermal preference and thermal experience (τ) in the original Kain *et al.* model and the updated model.
Colored lines in **a-b)** and **d-e)** reflect individual flies with particular thermal preferences bounded by 18ºC and 30ºC (color scale). **a)** τ under an average Boston, MA breeding season in the Kain *et al.* model. **b)** τ under an average Boston, MA breeding season in the updated model. **c)** Standard deviation in τ over the Boston, MA breeding season for the updated model. **d)** τ under an average Miami, FL breeding season in the Kain *et al.* model. **e)** τ under an average Miami, FL breeding season in the updated model. **f)** Standard deviation in τ over the Miami, FL breeding season for the updated model.

Table 2.1 - Bet-hedging advantage under different *a* values.
Bet-hedging advantage was calculated as $\log(BH_{\text{final pop size}}/AT_{\text{final pop size}})$. Station-specific climate normals were used to calculate the bet-hedging advantage ($\mu = 0.44$, $\sigma^2 = 0.015$). A dash (-) signifies that for that combination of location and *a* value, β and δ parameters were not able to be calibrated and no estimate of bet-hedging advantage was calculated.

Location (Station ID)	***a*=0.7**	***a*=0.5**	***a*=0.4**	***a*=0.3**	***a*=0.2**
Berkeley, CA (USC00040693)	8.6E-4	2.6E-3	3.7E-3	5.3E-3	7.0E-3
Boston, MA (USW00014739)	1.3E-3	2.9E-3	3.8E-3	4.7E-3	5.7E-3
Philadelphia, PA (USW00013739)	3.4E-4	7.2E-4	9.1E-4	1.1E-3	1.3E-3
Charlottesville, VA (USW00003759)	8.1E-4	1.6E-3	1.9E-3	2.4E-3	3.0E-3
Pasadena, CA (USC00046719)	3.3E-4	6.5E-4	9.9E-4	1.1E-3	1.3E-3
Houston, TX (USW00012918)	4.7E-3	5.4E-3	7.8E-3	-	-
Miami, FL (USC00085667)	-0.17	-0.35	-0.45	-0.57	-0.68

The relationships of development time and lifespan to temperature experience were updated based on data collected from three isofemale lines from Coral Gables, FL, Cambridge, MA, and Charlottesville, VA (Figure 2.10). For the relationship between development time and temperature, a quadratic fit on combined data from the three isofemale lines was used. For lifespan vs. temperature, a natural logarithm fit to the combined data was used. For fly *i* and on day *j,* development time and lifespan are determined by the following equations:

$$devTime(i,j) = 0.1445 * \tau(i,j)^2 - 7.5636 * \tau(i,j) + 108.1585$$
$$lifespan(i,j) = 459.9 - 128.0 * log(\tau(i,j))$$

Where $\tau(i, j)$ is the thermal experience of fly *i* on day *j.*

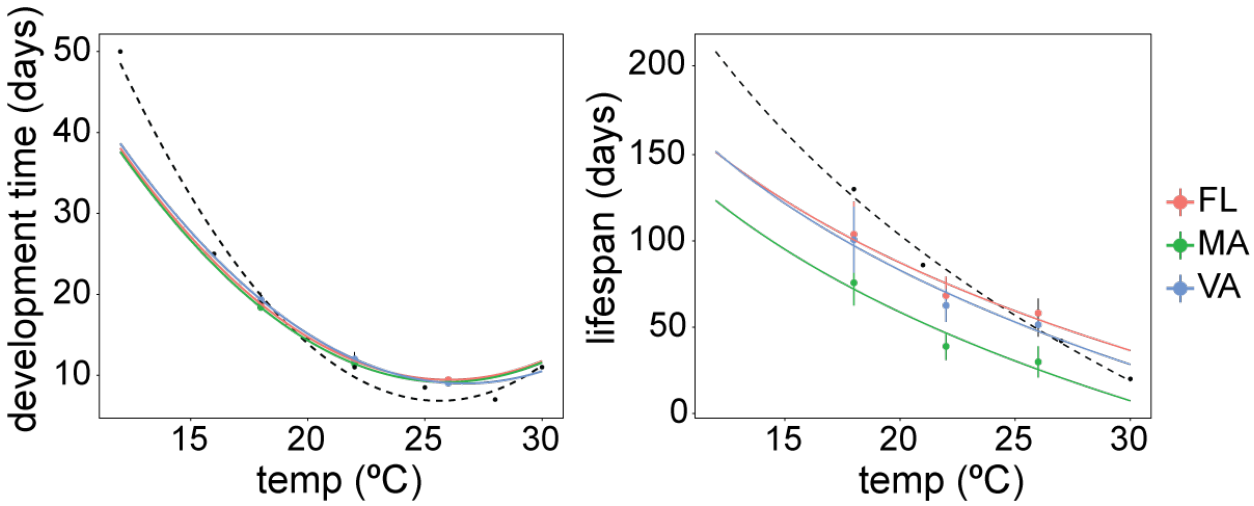


Figure 2.10 - Life history relationships used in the model.
Life history data used in the Kain *et al.* model (black dashed lines) and updated model life history data using isofemale lines from three locations (colored lines). Error bars show the 95% confidence interval of the mean.

Seasonal patterns of mean thermal preference

Weekly collections were carried out from June 24, 2018 until November 1, 2018 in Cambridge, Massachusetts, U.S.A. (MA), Charlottesville, Virginia, U.S.A (VA), and Coral Gables, Florida, U.S.A. (FL). In VA, *Drosophila* were collected via aspiration from rotting fruit at Carter Mountain Orchard (37.99° N, 78.47° W). In MA and FL, traps were set out to capture *Drosophila* around the residential areas in the vicinity of Harvard University (42.38° N, 71.12° W) and University of Miami (25.72° N, 80.28° W).

The MA and FL traps were created by cutting an approximately 2 inch flap into an empty one-gallon plastic ethanol jug (Koptec, Decon Labs) and baiting them with a fruit and wine mixture. Flies were collected by placing an empty fly food bottle over the neck of the jug and allowing flies in the trap to move upwards into the bottle. Bait for a single trap was made by mixing two sliced bananas with ½ sliced orange and soaking them in 50mL of 8.5% alc/vol red wine overnight. Bait was added to the trap, sprinkled with yeast, and the trap was hung on a fence or railing 2-3 days prior to the start of the week's collections. At the end of the week's collections,

the trap was removed and thoroughly washed to get rid of any bait/larvae/pupae that remained before fresh bait was put back in.

Collected flies were taken back to the local lab and sorted by sex and species (see Figure 2.2 for detailed experimental timeline for each location). For the thermal preference assay, *D. melanogaster* males and *D. melanogaster/ D. simulans* females were chosen (visual species identification of female *melanogaster* and *simulans* was too difficult to perform at a large-scale). The female flies were housed individually after the completion of the assay to allow for species identification to be performed on the male offspring. Females that did not produce an F1 generation were species identified through sequencing of *CoII* gene (forward primer: 5' -ATGGCAGATTAGTxGCAATGG; reverse primer: 5' -GTTTAAGAGACCAGTACTTG).

The chosen flies were assayed for thermal preference immediately after capture in MA during the entire sampling period. Due to mortality in the assay preventing species identification, VA and FL females were switched to testing 1-7 days post-collection to allow time for egg laying to occur. Caught flies were stored in vials with mixed-sex groups at 22-23ºC in ambient laboratory conditions. A mean thermal preference was determined from all the flies sampled on a particular week.

Observed mean preferences over the 20 week sampling period were compared to predicted patterns given a purely bet-hedging population or a purely adaptive tracking population as specified in Kain *et al.*[30] The predicted patterns were calculated using location-specific 2018 daily average temperatures[31] into a difference equations model simulating an infinite population reproducing over a breeding season with thermal preference determined by 1) a random draw from the observed distribution (bet-hedging) or 2) direct inheritance (adaptive tracking). The mean and variance parameters for the location-specific model run were empirically determined

using the assayed flies from each location. Birth (β) and death (δ) rates were calibrated for each population independently using daily average 2018 temperatures for Boston, MA and Charlottesville, VA and NOAA climate normals for Coral Gables, FL (due to failure of calibration on 2018 daily temperatures). To determine whether the bet-hedging model or adaptive tracking model fits the observed patterns, a log likelihood ratio was calculated on the observed data. 1000 bootstrap resamples were done to determine the estimation error. The findings are robust to different *a* values, as well as different relationships between temperature and life history (Figure 2.11).

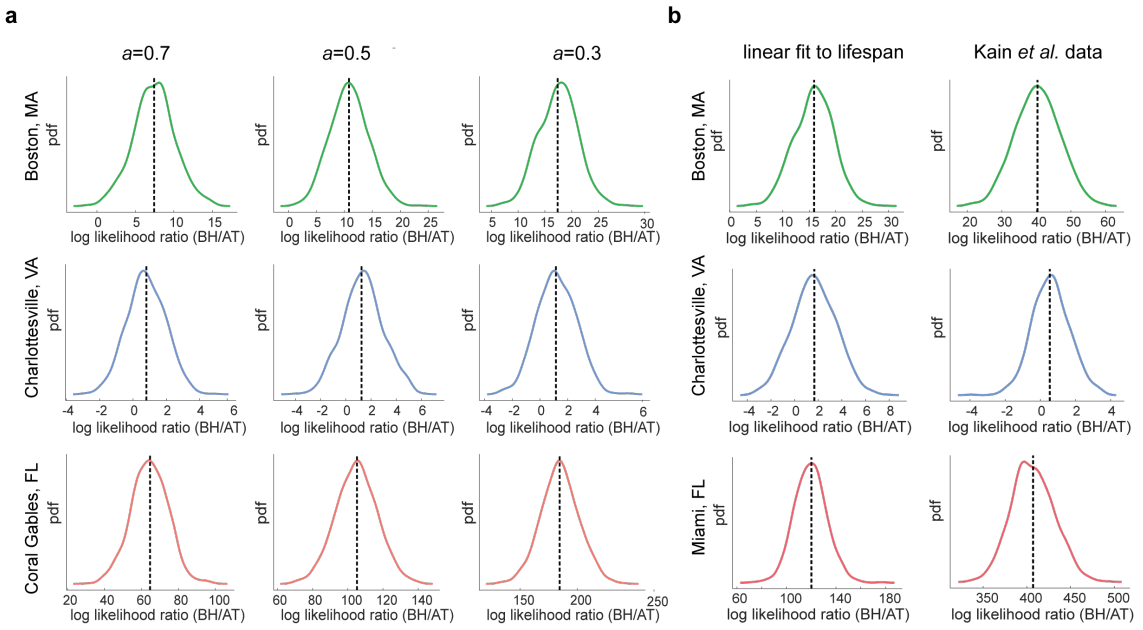


Figure 2.11 - Robustness of log-likelihood ratios to *a* value and life history parameters. β and δ parameters for Boston, MA and Charlottesville, VA were calibrated using the 2018 historical temperature data. β and δ parameters for Coral Gables, FL were calibrated using climate normals, as it was not feasible to calibrate using the 2018 data. Dashed lines show the observed **a)** Log-likelihood ratio (bet-hedging vs. adaptive tracking) of observed seasonal patterns of thermal preference under different *a* values. **b)** Log likelihood ratios (bet-hedging vs. adaptive tracking) using a linear fit to the isofemale lifespan data and using the life history data from thel Kain *et al.* model (*a* = 0.4).

Thermal preference variability of geographically diverse lines

Wild gravid *D. melanogaster* females caught in Houston, Texas, U.S.A (TX) (29.76° N, 95.36° W) and Oakland, California, U.S.A (CA) (37.80° N, 122.27° W) in September 2018 and

Pasadena, California, U.S.A (CA) (34.15° N, 118.14° W) and Media, Pennsylvania, U.S.A (PA) (39.89° N, 75.41° W) in October 2018 were used to establish isofemale lines in order to extend our study to more locales that may vary in bet-hedging advantage. In addition, as part of the seasonal collection experiment, two gravid females per week from June to October 2018 in FL, MA, and VA were retained to establish isofemale lines. Isofemale lines were maintained with overlapping generations. Four isofemale lines from each location (with the exception of Berkeley, where there were three) were chosen for evaluation of thermal preference variability. Lines chosen from FL, MA, and VA were established from gravid females collected from the start of July to the end of August.

200-250 mated female flies (aged 3-6 days) from FL, MA, TX, and VA isofemale lines were tested in two batches (two lines from each location per batch) in November and December 2018. Northern CA (Oakland), Southern CA (Pasadena), and PA flies were tested in two batches in late January 2019 and April 2019. Two MA isofemale lines previously tested earlier were retested alongside these lines to control for batch effects on variability. Batch effects were calculated as the average difference in variability for the internal control lines between the November/December 2018 trials and the January 2019 or April 2019 trials. Batch effects were added to the variability estimates of the isofemale lines to remove the effect of the time of year on variability.

Bayesian inference was used on a hierarchical model to estimate the mean and variance of thermal preference (in ºC) in each isofemale line and for each location. In the hierarchical model, isofemale lines were nested within sampling location, such that the prior on the line mean and variance was dependent on the hyper-prior for the location:

$$\mu_{location} \sim N(0,1)$$
$$\sigma^2{}_{location} \sim \Gamma(2, 0.1)$$

$$\mu_{line} \sim N(X_{location} \cdot \mu_{location}, \sigma_{\mu_{line}})$$
$$\sigma^2{}_{line} \sim N(X_{location} \cdot \sigma^2{}_{location}, \sigma_{\sigma^2 line})$$

The likelihood was specified as follows:

$$L(y \mid \mu_{line}, \sigma) \sim N(X_{line} \cdot \mu_{line}, \sigma)$$
$$\sigma = \sqrt{(X_{line} \cdot \sigma^2{}_{line}) + \phi D^{\psi}}$$

Where y is the vector of observed thermal preferences (ºC), X is dummy-coded predictor matrix for either line or location categories, and D is a vector of distance traveled (px) over the course of the experiment. σ depends on both the line variance ($\sigma^2{}_{line}$) and a sampling error component (ϕD^{ψ}) that depends on distance traveled during the experiment (rarely moving flies register fewer effective data points about their preference)[40]. ϕ and ψ constants were calculated by fitting a power function to the relationship between variance and distance traveled for flies experiencing no temperature stimulus. This allows us to differentiate variance that is inherent to the line from variance that comes from sampling noise due to low activity in the assay.

Bayesian inference was done using R's Stan interface v.2.18.2. The Stan platform does full Bayesian inference using a Hamiltonian Monte Carlo approach with the flexibility of custom model specification[41]. Posterior distributions for mean and variance for both lines and locations were generated by sampling using four chains - 25,000 iterations per chain, with target average proposal acceptance probability of 0.9, and maximum tree depth of 10. Every other sample from the chain was saved to the posterior distribution to reduce autocorrelation between the samples, thereby increasing the ratio of effective sample size to actual sample size and reducing

uncertainty in our estimate. To estimate variability, we used standard deviation, which was calculated by taking the square root of the variance at each step in the chain. Model fits were qualitatively evaluated using graphical posterior predictive checks, where mock data that was generated using values from the posterior distributions were compared to our observed data.

To establish whether variability estimates between two locations were different from each other, we generated the posterior distribution of differences by subtracting variability estimates for one location from the other at each step in the chain. If the 95% credible interval for the distribution of differences did not include 0, the two locations were considered to be different from each other in terms of variability[34,42].

Heritability of thermal preference

Narrow-sense heritability was calculated using parent-offspring regression. Males and females from 5 isofemale lines from the Southern CA, PA, and TX sampling locations were chosen for the parental generation. Males and females from 10 isofemale lines (original females collected between mid-August and late October) from FL, MA, and VA sampling locations were also chosen for the parental generation. Parents were collected, separated by sex to maintain virginity of females, and aged 3-6 days before testing for thermal preference. 10 crosses were made from the Southern CA, PA, and TX parents, and 20 crosses were made from the FL, MA, and VA parents (Figure 2.12). Each cross between two lines was replicated 2-3 times with independent sets of parents. Parent flies had to pass the activity thresholds for the thermal preference assay to be included in the crossing scheme.

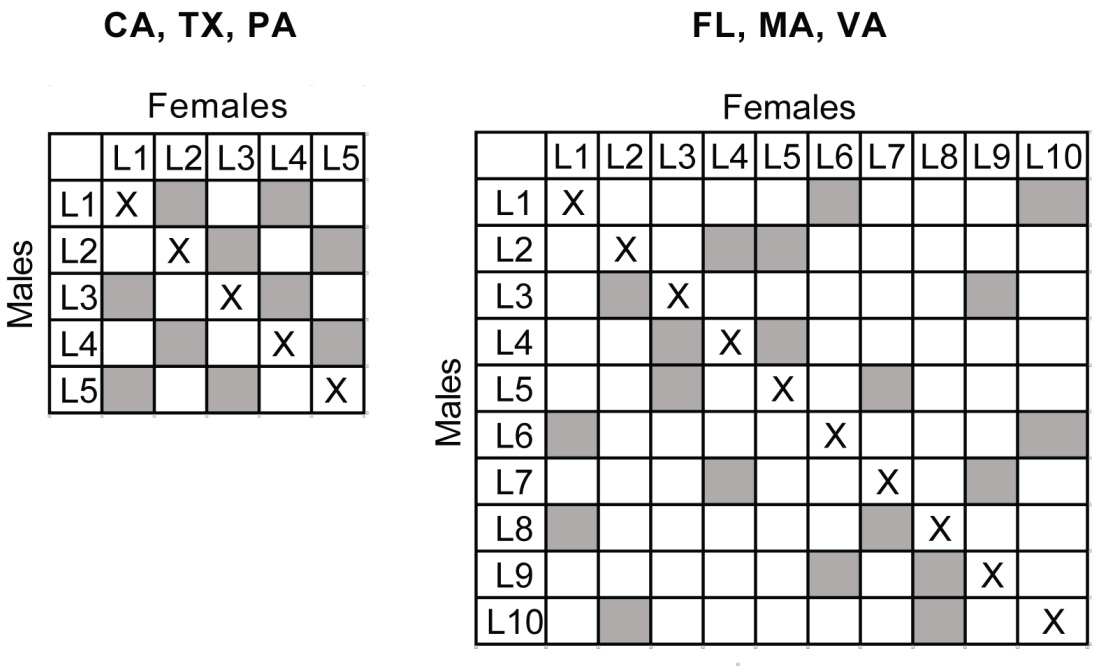


Figure 2.12 - Cross scheme for heritability analysis.
Gray squares indicate a cross between two lines and an "X" indicates the diagonal (no crosses were made along the diagonal).

Male F1s from each cross were collected and aged 3-7 days prior to testing. For each cross, a minimum of four male F1s were tested, with 10 male F1s tested for 96% (221/230) of crosses. All crosses with fewer than three male F1s passing the thermal preference filtering threshold were excluded from further analysis. Narrow-sense heritability was estimated from the slope of the regression of F1 mean thermal preference on the mid-parent thermal preference.

Plasticity in thermal preference

Males and females (0-2 days old) were collected from a MA isofemale established in mid-August and allowed to lay eggs for 48 hours at 26ºC (45% RH) and 22ºC (45% RH) and 96 hours at 18ºC (50-55% RH). Female F1s were collected daily and aged 3-6 days with males (to ensure mated status) at the treatment temperature prior to testing for thermal preference. Mean and variability in thermal preference was estimated using Bayesian inference on R's Stan interface, as mentioned above. The priors and likelihood function were as follows:

$$\mu \sim N(0,1)$$
$$\sigma^2 \sim \Gamma(2, 0.1)$$

$$L(y \mid \mu, \sigma) \sim N(X \cdot \mu, \sigma)$$
$$\sigma = \sqrt{(X \cdot \sigma^2) + \phi D^{\psi}}$$

Where y is the vector of observed thermal preferences, X is dummy-coded predictor matrix for the temperature treatments, and D is a vector of distance traveled (px) over the course of the experiment. As for the variability experiments, σ is partitioned into the line variance (σ^2) and a component (ϕD^{ψ}) that denotes the variance due to sampling noise.

Estimating genetic diversity in the sampled populations

Genomic DNA from 4 female flies from each isofemale line tested in the heritability and variability assays was extracted using bead-beating and the ZYMO *Quick*-DNA kit (cat. no. D3012). DNA was made into libraries using a liquid handling robot (Analytic Gena CyBi-Felix Model 30-5015-100-24). Library preparation was done using a tagmentation protocol with Tn5 transposase[43,44]. Genomic DNA from 273 individual flies was made into per-individual libraries. 150bp paired-end reads were sequenced on an Illumina NovaSeq platform with mean 0.02x - 8x coverage per individual. Alignment of reads was done using the BWA-MEM algorithm (v0.7.15; default parameters)[45] to the *Drosophila melanogaster* reference genome 6.28 release. PCR and optical duplicates were flagged using Picard's MarkDuplicates (v2.20.6). HaplotypeCaller in GATK version 4.1.3.0 was used to call variants[46]. Given the low sequencing coverage, minimum pruning support and minimum dangling branch length were set to 1; all other parameters were kept at default values. Mean coverage depth and fraction of missing genotypes per individual was quantified using VCFtools[47]. Individuals with mean coverage depth less than 2x were excluded from further analysis, leaving a total of 246 individuals.

Variants were filtered for biallelic SNPs with a minor allele frequency > 5%. 15,080 SNPs (distributed across the genome) with called genotypes for all individuals were chosen. This set of SNPs was used to calculate Watterson's θ_s in the downstream analysis. A genotype matrix of variants by individuals was created - the cells of the matrix contain whether a particular individual is homozygous for the reference or alternate allele or heterozygous. Since the number of individuals and number of lines in each population will influence the θ_s estimate, a subsampling approach was taken.

The individuals were divided into subsets: those used in the heritability analysis and those used in the variability analysis. For each subset, the geographic population that had the fewest lines and individuals was chosen as the subsampling benchmark. To estimate θ_s in a population and its uncertainty, a bootstrapping approach was employed in addition to the subsampling. For each bootstrap iteration, a subsample was taken from the target population that had the same number of lines and the same number of individuals per line as the benchmark population (the same individual could be chosen more than once). For each variant, a homozygous individual would contribute either a reference or alternate allele, and a heterozygous individual would contribute a reference allele with a 50% probability. A variant was counted as segregating if individuals contributed both reference and alternate alleles. The number of segregating sites and Watterson's theta (θ_s) were then calculated from the chosen individuals:

$$\theta_s = \frac{S}{\sum_{i=1}^{n-1} \frac{1}{i}}$$

where S is the number of segregating sites and n is the number of chromosomes (i.e. individuals). The final result was a metric of genetic diversity that could be compared across populations of different individual and line compositions (Figure 2.13).

To calculate within-line genetic diversity, the same general approach as above for calculating θ_s was employed. Since there was a maximum of four individuals per line, a bootstrapping approach was not used. Instead, θ_s was calculated only for lines that had the full complement of four individuals (Figure 2.13).

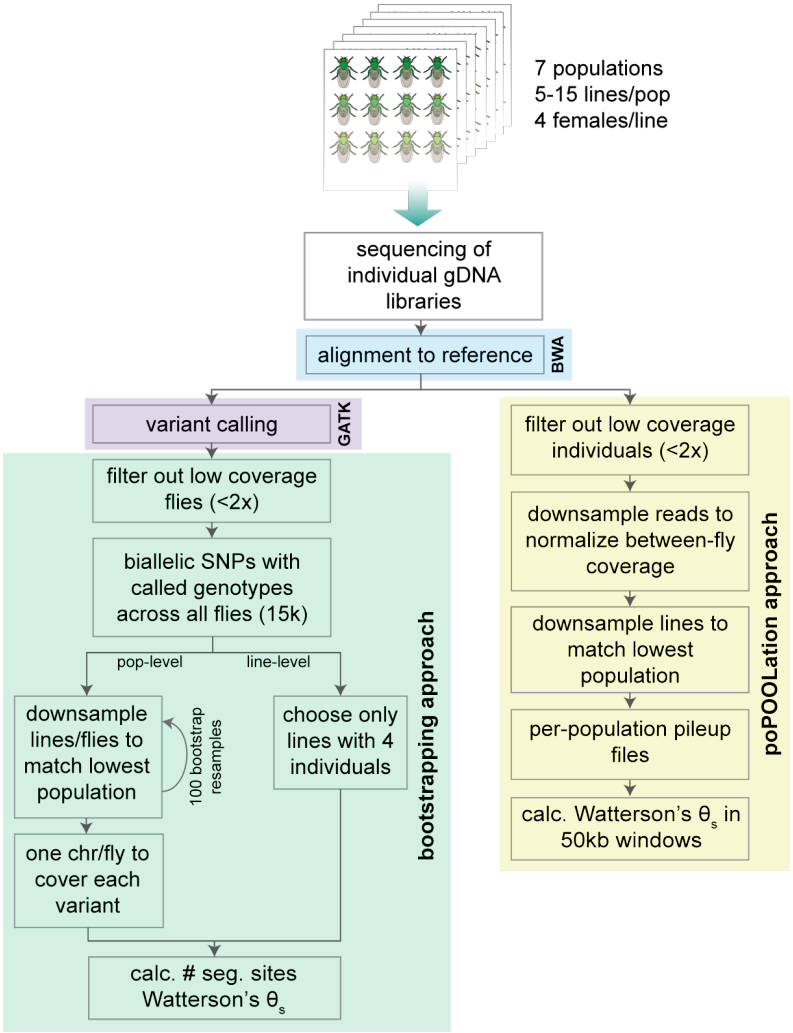


Figure 2.13 - Flowchart of the two approaches used to estimate Watterson's θ_s.

As a complement to the bootstrapping approach, θ_s was also calculated using PoPoolation[48] for both the heritability and variability flies. Using only the 246 individuals with mean coverage > 2x, the number of reads per individual was downsampled using SAMtools[49] to match the individual with the lowest coverage in order to standardize coverage across individuals. As in the bootstrapping approach, populations with more isofemale lines were subsampled to match the population with the fewest isofemale lines prior to making the population pileup file. The pileup file was then filtered using the *identify-genomic-indel-regions.pl* and *filter-pileup-by-gtf.pl* functions to remove indels and the variants within 5bp of them. θ_s was calculated from the pileup file in 50kb non-overlapping windows using only SNPs with a minimum minor allele count of 2, minimum site coverage of 4, maximum site coverage of 400, and reads with a minimum quality score of 20. 60% of the 50kb window had to have coverage between 4 and 400 for θ_s to be calculated. To get a single population θ_s, the mean of θ_s across all 50kb windows for chromosomes 2L, 2R, 3L, 3R, X, and 4 was taken (Figure 2.13). The θ_s from the PoPoolation analysis is reported as θ_s/nt to distinguish it from our bootstrapping approach.

References

1. Dall, S. R. X., Houston, A. I. & McNamara, J. M. The behavioural ecology of personality: consistent individual differences from an adaptive perspective. *Ecol. Lett.* **7**, 734–739 (2004).

2. Sih, A., Bell, A. M., Johnson, J. C. & Ziemba, R. E. Behavioral Syndromes: An Integrative Overview. *Q. Rev. Biol.* **79**, 241–277 (2004).

3. Sih, A. *et al.* Animal personality and state–behaviour feedbacks: a review and guide for empiricists. *Trends Ecol. Evol.* **30**, 50–60 (2015).

4. Wolf, M. & Weissing, F. J. An explanatory framework for adaptive personality differences. *Philos. Trans. R. Soc. B Biol. Sci.* **365**, 3959–3968 (2010).

5. Dochtermann, N. A., Schwab, T. & Sih, A. The contribution of additive genetic variation to personality variation: heritability of personality. *Proc. R. Soc. B Biol. Sci.* **282**, 20142201 (2015).

6. Ellner, S. & Sasaki, A. Patterns of Genetic Polymorphism Maintained by Fluctuating Selection with Overlapping Generations. *Theor. Popul. Biol.* **50**, 31–65 (1996).

7. Meyers, L. A. & Bull, J. J. Fighting change with change: adaptive variation in an uncertain world. *Trends Ecol. Evol.* **17**, 551–557 (2002).

8. Turelli, M. & Barton, N. H. Polygenic variation maintained by balancing selection: pleiotropy, sex-dependent allelic effects and G x E interactions. *Genetics* **166**, 1053–1079 (2004).

9. Wittmann, M. J., Bergland, A. O., Feldman, M. W., Schmidt, P. S. & Petrov, D. A. Seasonally fluctuating selection can maintain polymorphism at many loci via segregation lift. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E9932–E9941 (2017).

10. Dingemanse, N. J., Both, C., Drent, P. J. & Tinbergen, J. M. Fitness consequences of avian personalities in a fluctuating environment. *Proc. R. Soc. Lond. B Biol. Sci.* **271**, 847–852 (2004).

11. Penke, L., Denissen, J. J. A. & Miller, G. F. The evolutionary genetics of personality. *Eur. J. Personal.* **21**, 549–587 (2007).

12. Simons, A. M. Modes of response to environmental change and the elusive empirical evidence for bet hedging. *Proc. R. Soc. B Biol. Sci.* **278**, 1601–1609 (2011).

13. Bell, G. & Collins, S. Adaptation, extinction and global change. *Evol. Appl.* **1**, 3–16 (2008).

14. Lande, R. & Shannon, S. The Role of Genetic Variation in Adaptation and Population Persistence in a Changing Environment. *Evolution* **50**, 434–437 (1996).

15. Kawecki, T. J. The Evolution of Genetic Canalization Under Fluctuating Selection. *Evolution* **54**, 1–12 (2000).

16. Hopper, K. R. Risk-spreading and bet-hedging in insect population biology. *Annu. Rev. Entomol.* **44**, 535–560 (1999).

17. Donaldson-Matasci, M. C., Lachmann, * Michael & Bergstrom, C. T. Phenotypic diversity as an adaptation to environmental uncertainty. *Evol. Ecol. Res.* **10**, 493–515 (2008).

18. Starrfelt, J. & Kokko, H. Bet-hedging—a triple trade-off between means, variances and correlations. *Biol. Rev.* **87**, 742–755 (2012).

19. Sasaki, A. & Ellner, S. The Evolutionarily Stable Phenotype Distribution in a Random Environment. *Evolution* **49**, 337–350 (1995).

20. Starrfelt, J. & Kokko, H. Bet-hedging—a triple trade-off between means, variances and correlations. *Biol. Rev.* **87**, 742–755 (2012).

21. Simons, A. M. Fluctuating natural selection accounts for the evolution of diversification bet hedging. *Proc. R. Soc. B Biol. Sci.* **276**, 1987–1992 (2009).

22. Bull, J. J. Evolution of Phenotypic Variance. *Evolution* **41**, 303–315 (1987).

23. Haccou, P. & Iwasa, Y. Optimal Mixed Strategies in Stochastic Environments. *Theor. Popul. Biol.* **47**, 212–243 (1995).

24. Svardal, H., Rueffler, C. & Hermisson, J. Comparing environmental and genetic variance as adaptive response to fluctuating selection. *Evol. Int. J. Org. Evol.* **65**, 2492–2513 (2011).

25. Tufto, J. Genetic evolution, plasticity, and bet-hedging as adaptive responses to temporally autocorrelated fluctuating selection: A quantitative genetic model. *Evolution* **69**, 2034–2049 (2015).

26. Scheiner, S. M. Bet-hedging as a complex interaction among developmental instability, environmental heterogeneity, dispersal, and life-history strategy. *Ecol. Evol.* **4**, 505–515 (2014).

27. Buchanan, S. M., Kain, J. S. & de Bivort, B. L. Neuronal control of locomotor handedness in *Drosophila*. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 6700–6705 (2015).

28. Ayroles, J. F. *et al.* Behavioral idiosyncrasy reveals genetic control of phenotypic variability. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 6706–6711 (2015).

29. Kain, J. S., Stokes, C. & de Bivort, B. L. Phototactic personality in fruit flies and its suppression by serotonin and white. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 19834–19839 (2012).

30. Kain, J. S. *et al.* Variability in thermal and phototactic preferences in Drosophila may reflect an adaptive bet-hedging strategy. *Evolution* **69**, 3171–3185 (2015).

31. Arguez, A. *et al. U.S. Daily Climate Normals (1981-2010)*. doi:10.7289/V5PN93JP (2010).

32. Coyne, J. A. & Milstead, B. Long-Distance Migration of Drosophila. 3. Dispersal of D. melanogaster Alleles from a Maryland Orchard. *Am. Nat.* **130**, 70–82 (1987).

33. Machado, H. E. *et al.* Comparative population genomics of latitudinal variation in Drosophila simulans and Drosophila melanogaster. *Mol. Ecol.* **25**, 723–740 (2016).

34. Akhund-Zade, J., Ho, S., O'Leary, C. & Bivort, B. de. The effect of environmental enrichment on behavioral variability depends on genotype, behavior, and type of enrichment. *J. Exp. Biol.* **222**, (2019).

35. Adrion, J. R., Hahn, M. W. & Cooper, B. S. Revisiting classic clines in Drosophila melanogaster in the age of genomics. *Trends Genet. TIG* **31**, 434–444 (2015).

36. Fabian, D. K. *et al.* Genome-wide patterns of latitudinal differentiation among populations of Drosophila melanogaster from North America. *Mol. Ecol.* **21**, 4748–4769 (2012).

37. Reinhardt, J. A., Kolaczkowski, B., Jones, C. D., Begun, D. J. & Kern, A. D. Parallel geographic variation in Drosophila melanogaster. *Genetics* **197**, 361–373 (2014).

38. Barbagallo, B. & Garrity, P. A. Temperature sensation in Drosophila. *Curr. Opin. Neurobiol.* **34**, 8–13 (2015).

39. Werkhoven, Z., Rohrsen, C., Qin, C., Brembs, B. & Bivort, B. de. MARGO (Massively Automated Real-time GUI for Object-tracking), a platform for high-throughput ethology. *PLOS ONE* **14**, e0224243 (2019).

40. Honegger, K. S., Smith, M. A.-Y., Churgin, M. A., Turner, G. C. & Bivort, B. L. de. Idiosyncratic neural coding and neuromodulation of olfactory individuality in Drosophila. *Proc. Natl. Acad. Sci.* (2019) doi:10.1073/pnas.1901623116.

41. Carpenter, B. *et al.* Stan: A Probabilistic Programming Language. *J. Stat. Softw.* **76**, 1–32 (2017).

42. Kruschke, J. K. Bayesian estimation supersedes the t test. *J. Exp. Psychol. Gen.* **142**, 573–603 (2013).

43. Pallares, L. F., Picard, S. & Ayroles, J. F. TM3'seq: A Tagmentation-Mediated 3' Sequencing Approach for Improving Scalability of RNAseq Experiments. *G3 Genes Genomes Genet.* **10**, 143–150 (2020).

44. Picelli, S. *et al.* Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res.* **24**, 2033–2040 (2014).

45. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv13033997 Q-Bio* (2013).

46. McKenna, A. *et al.* The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).

47. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).

48. Kofler, R. *et al.* PoPoolation: A Toolbox for Population Genetic Analysis of Next Generation Sequencing Data from Pooled Individuals. *PLOS ONE* **6**, e15925 (2011).

49. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinforma. Oxf. Engl.* **25**, 2078–2079 (2009).

Chapter III - Genetic basis of offspring number and body weight in *Drosophila melanogaster*

I would like to thank those individuals that helped me carry out the experimental work in this project. Erika A. Gajda worked with me on the screen of the Drosophila Genetic Reference Panel (DGRP) lines. Denise Yoon assisted me in with the parental density experiments. Shraddha Lall carried out the functional validation tests for the candidate genes. In addition to the experimental work, I designed the experiments and analyzed the data. I also would like to thank Rob Unckless and Alan O. Bergland for sharing raw data on DGRP phenotypes.

Abstract

Drosophila melanogaster egg production, a proxy for fecundity, has been an extensively studied life history trait with a strong genetic basis. As eggs develop into larva and adult offspring, space and resource constraints can put pressure on the developing offspring, leading to a decrease in viability, body size, and lifespan. We were interested in testing for and mapping the genetic basis of offspring number and weight using the restrictions of a standard laboratory vial.

We screened 143 lines from the *Drosophila* Genetic Reference Panel for offspring numbers and weights to create an 'offspring index' that captured the number vs. weight trade-off. We found 30 associated variants in 18 genes - validation of *hid, Sox21b, CG8312,* and *mub* candidate genes using gene disruption mutants showed them to be important in adult stage viability, while mutations in *Ih* and *Rbp* increased offspring number and increased weight, respectively. The polygenic basis of offspring number and weight with many variants of small effect, as well as the involvement of genes with varied functional roles, support an omnigenic model for this life history trait.

Introduction

Life history traits, such as fecundity, lifespan, and body size, are major contributors to fitness. In *Drosophila melanogaster*, the genetics, plasticity, and evolution of life history traits have been extensively studied[1]. *Drosophila* fecundity, measured through egg-laying behavior, was previously shown to have a strong genetic component that is different between young and old flies[2–5], but to also be influenced by temperature and nutrition[6,7]. In support of a resource allocation model of life history, fecundity has also been shown to trade-off with longevity[5,8], showing that investment into the next generation comes at a cost to somatic maintenance via maintained energy reserves. A genome-wide association study revealed that age-specific fecundity is associated with variants present across a large set of candidate genes, enriched for genes involved in development, morphogenesis, neural function, and cell signaling[4]. Connecting fecundity with neural function, QTL and deficiency mapping revealed that expression of a *Drip* aquaporin in corazonin neurons was positively correlated with fecundity by modulating the neurohormone balance between corazonin and dopamine[9].

While the vast majority of *Drosophila* studies have focused on using egg production as a measure of fecundity, the number of eggs laid may not translate perfectly to viable offspring due to potential mortality at the larval stages. Under both natural and laboratory conditions, larvae must contend with a finite space and resource limitations given the constraints of either the rotting food substrate[10,11] or the laboratory food and competition between larvae. Increased larval density decreases egg-to-adult viability[12–14], body size[13,15–18] and longevity[19], while increasing development time[13,14,16,18] and lowering starvation resistance[15]. Highly fecund flies that lay a large number of eggs may end up negatively affecting the larval and adult stages of their offspring due to the increased density. On the other end of the spectrum, flies producing fewer eggs may have large offspring capable of weathering stress[20,21], but at a decrease to their

competitive ability with more fecund individuals. Therefore, imposed space and resource limitation may impose a trade-off between the number of and investment into offspring.

Given the fitness ideal of having a large number of high-quality offspring, we were interested in whether, downstream of egg production, there was a genetic basis for adult offspring number and their quality under resource limitation. We used the standard laboratory food vial to impose both a space and food limitation on the developing offspring. As a measure of quality, we measured the wet weight of the eclosing offspring - increased body weight is correlated with increased starvation resistance[21], increased nutrient stores, and increased immunity[22], which indicate an investment in somatic maintenance. We scored lines from the *Drosophila* Genetic Reference Panel (DGRP)[23] for numbers of adult offspring and their weight. We found candidate genes with variants significantly associated with a combined metric of offspring number and weight - disruption of these genes in most cases caused lethality or impaired survival at the adult stage, but in other cases shifted the balance between offspring weight and number.

Results

Genome-wide association mapping for offspring index

We collected four fecundity and body weight phenotypes from 143 DGRP lines: total number of female progeny, total number of male progeny, and their respective mean weights (in mg). We found inter-line differences for all four phenotypes measured (Figure 3.1a), as well as strong correlations between the phenotypes (Figure 3.1b). The estimated broad-sense heritability of mean female weight (0.639, 95%HPDI: 0.553 - 0.722) was lower than for mean male weight (0.725, 95%HPDI: 0.650 - 0.796). Both weight phenotype heritabilities were higher than heritabilities previously estimated for body weight[21,24]. The heritabilities of the total number of female progeny (0.465, 95% HPDI: 0.369 - 0.567) and male progeny (0.484, 95%HPDI: 0.390 - 0.584) were higher than the heritabilities previously estimated on number of eggs laid[4] (Table 3.1).

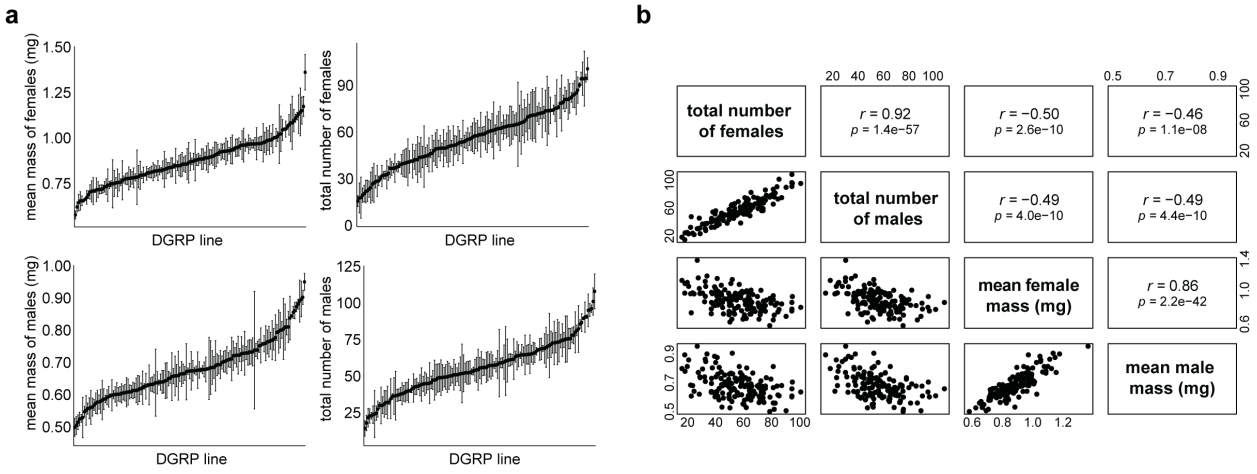


Figure 3.1 - DGRP lines show variation in offspring number and weight.
a) Plot of phenotypes measured (+/- 1 s.e.) with the DGRP lines sorted ascending order for each phenotype (3 replicates/line). **b)** Correlation matrix of the phenotypes measured.

Table 3.1 - Heritability estimates for the means of the four phenotypes.
HPD = highest posterior density.

Phenotype	Heritability Estimate	95% HPD Interval
Mean ♀ Weight	0.639	0.553 - 0.722
Mean ♂ Weight	0.725	0.650 - 0.796
Number of ♀	0.465	0.369 - 0.567
Number of ♂	0.484	0.390 - 0.584

As we predicted based on resource limitation, we found that the number of offspring was negatively correlated with the offspring weight (Figure 3.1b). The strong correlations between the offspring number and weight phenotypes measured allowed us to use the first principal component to combine the four measurements into a single metric, which we termed the offspring index. The first principal component explained 71% of the variance and is negatively loaded for offspring number and positively loaded for offspring weight (Table 3.2).

Table 3.2 - Variance proportion explained by each principal component and their loadings.

	PC1	PC2	PC3	PC4
St. dev	1.686	0.965	0.385	0.282
Variance prop.	0.711	0.233	0.037	0.020
Number of ♀	-0.505	0.501	-0.185	0.678
Number of ♂	-0.510	0.483	0.211	-0.680
Mean ♀ Weight	0.497	0.494	0.688	0.192
Mean ♂ Weight	0.488	0.521	-0.670	-0.205

A negative index value indicated many low-weight offspring, a positive index value indicated few high-weight offspring, and an index value close to zero indicated a balance between offspring number and weight. We used this phenotype for the genome-wide association study. Thirty variants were associated with the offspring index using a threshold of $p < 1E\text{-}5$ (Table 3.3). 12 of

the significant variants were located on the third chromosome, while 16 were located on the second chromosome and only 2 on the X chromosome. 23 of 30 of the variants were within 1,000 bp of a gene, of which about half (14/23) were located in introns. Five of the associated variants are present in genes that were previously associated with fecundity[4]. Among the candidate genes, we did not find significant enrichment for particular biological processes or molecular functions using PANTHER's Overrepresentation Test with the GO-Slim annotation sets[25]. The quantile-quantile plot shows no systematic bias and a slight enrichment for $p < 1E\text{-}5$ (Figure 3.2a). The linkage disequilibrium plot reveals no long-distance linkage between variants (Figure 3.2b).

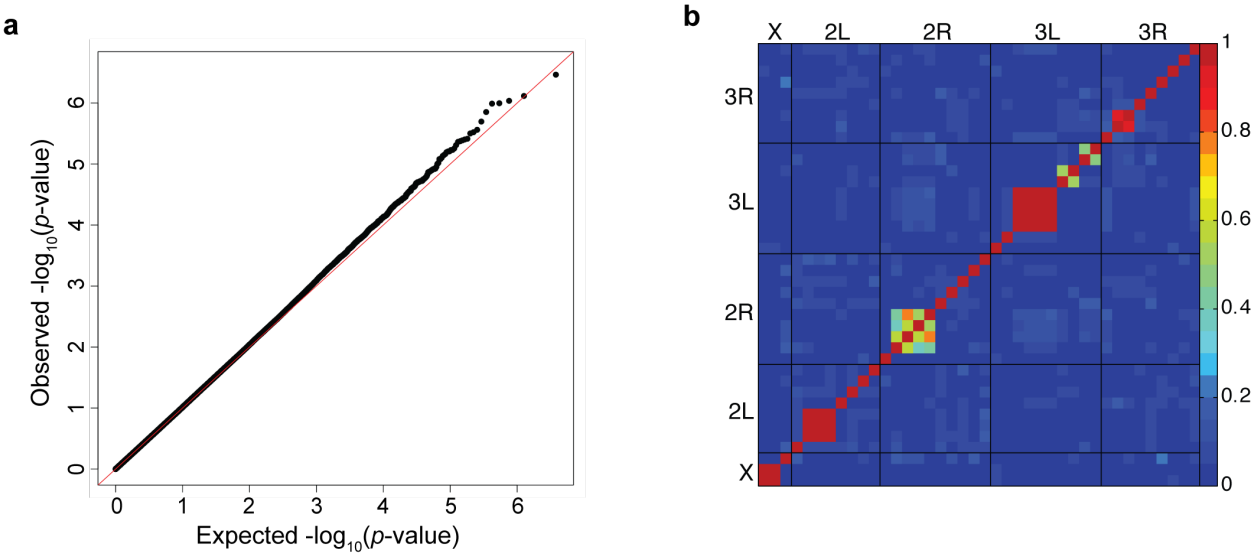


Figure 3.2 – QQ plot and linkage disequilibrium of associated variants.
a) Quantile-quantile plot comparing observed and expected *p*-values, with the red line showing a 1:1 relationship. **b)** Linkage disequilibrium heat map for all the variants identified ($p < 1E\text{-}4$).

Table 3.3 - Variants significantly associated with the offspring index (p < 1E-5).
Chromosome coordinates represented in dm5 assembly coordinates. MAF is minor allele frequency, Del is deletion, and numbers in parentheses represent the number of basepairs to the closest gene. Genes in bold were previously identified to contain variants associated with age-specific fecundity[4].

Chr	Pos	MAF	Effect	*p*-value	Gene	Class
3L	18174169	0.492	0.697	3.43E-07	*hid*	Intron
2L	10356660	0.101	-1.112	7.65E-07	***CG5367***	Upstream (63bp)
3L	14106797	0.456	-0.597	9.20E-07	***Sox21b***	Del (12bp - Intron)
3R	5440058	0.102	-1.075	1.01E-06	*CG8312*	Intron
3R	5437737	0.090	-1.099	1.02E-06	*CG8312*	Intron
3R	11212396	0.066	-1.260	1.40E-06	*Rbp*	Intron
3R	10285026	0.052	-1.473	2.01E-06	***cv-c***	Intron
3R	2150285	0.123	-0.980	2.74E-06	*Osi17*	Intron
2R	19814320	0.130	-0.907	3.01E-06	*CG2812*	3' UTR
2L	22137883	0.347	-0.700	3.15E-06	*CG42748*	Intron
2R	8813359	0.291	0.703	3.85E-06	*sug*	Intron
2R	10191983	0.058	-1.326	3.96E-06	—	—
2R	16280567	0.346	-0.597	4.09E-06	—	—
2R	10186017	0.130	-0.935	4.23E-06	*lh*	3' UTR
X	16619471	0.338	0.652	4.33E-06	*CG32572*	Intron
2R	18428402	0.050	-1.392	4.90E-06	***px***	Synonymous
2R	16643265	0.157	-0.823	5.56E-06	—	—
3L	21865887	0.121	-0.932	5.86E-06	*mub*	Intron
2R	10185377	0.058	-1.333	5.90E-06	*lh*	Intron
2L	14413190	0.050	-1.315	6.21E-06	—	—
2L	14413193	0.057	-1.251	6.23E-06	—	—
3L	16206105	0.159	-0.843	6.38E-06	*CG13073*	Downstream(46bp)
2R	10354544	0.050	-1.438	6.87E-06	—	—
3L	16206075	0.154	-0.851	7.11E-06	*CG13073*	Downstream(76bp)
X	16619495	0.333	0.642	7.36E-06	*CG32572*	Intron
3R	6652348	0.210	-0.682	7.91E-06	*Cad86C*	Upstream(453bp)
2L	19610094	0.072	-1.142	8.28E-06	***Lar***	Intron
2R	10185828	0.079	-1.101	8.30E-06	*lh*	Intron
2L	14413263	0.051	-1.358	9.60E-06	—	—
3R	14116444	0.346	0.650	9.93E-06	*l(3)05822*	Synonymous

Fecundity differences are stable under different parental densities

We assessed whether differences among DGRP lines in the offspring index would persist under different parental densities. We chose six lines from our screen that were representative of negative offspring index (many low-weight offspring), zero offspring index, and positive offspring index (few high-weight offspring) to assay for offspring phenotypes at different densities of parents (Figure 3.3). We find that parental density is a significant predictor of offspring number (females: p = 1.43E-7, males: p = 3.14E-6) and offspring weight (females: p = 2.42E-11, males: p = 2.35E-11). As expected, increasing parental density increased offspring number and decreased offspring weight, though the effects diminished with increasing density for most lines. Accounting for the effect of density, we see that the DGRP line still has a significant impact on offspring number (females: p = 2.58E-5, males: p = 1.84E-5) and offspring weight (females: p = 8.57E-8, males: p = 2.48E-7). DGRP lines with positive index (RAL 812: +3.58; RAL 894: +3.79) maintained a low offspring number and high offspring weight under different densities. RAL 237, a DGRP line with a negative index (-2.87), had consistently high offspring numbers and low offspring weights. Surprisingly, RAL 176, a DGRP line with a negative index (-3.18) in our screen had fewer high-weight offspring compared to RAL 49, a line with an index close to zero (-0.04). We also could not detect significant line-by-density interactions on the phenotypes measured (female number: p = 0.663, female weight: p = 0.857, male number: p = 0.572, male weight: p = 0.540).

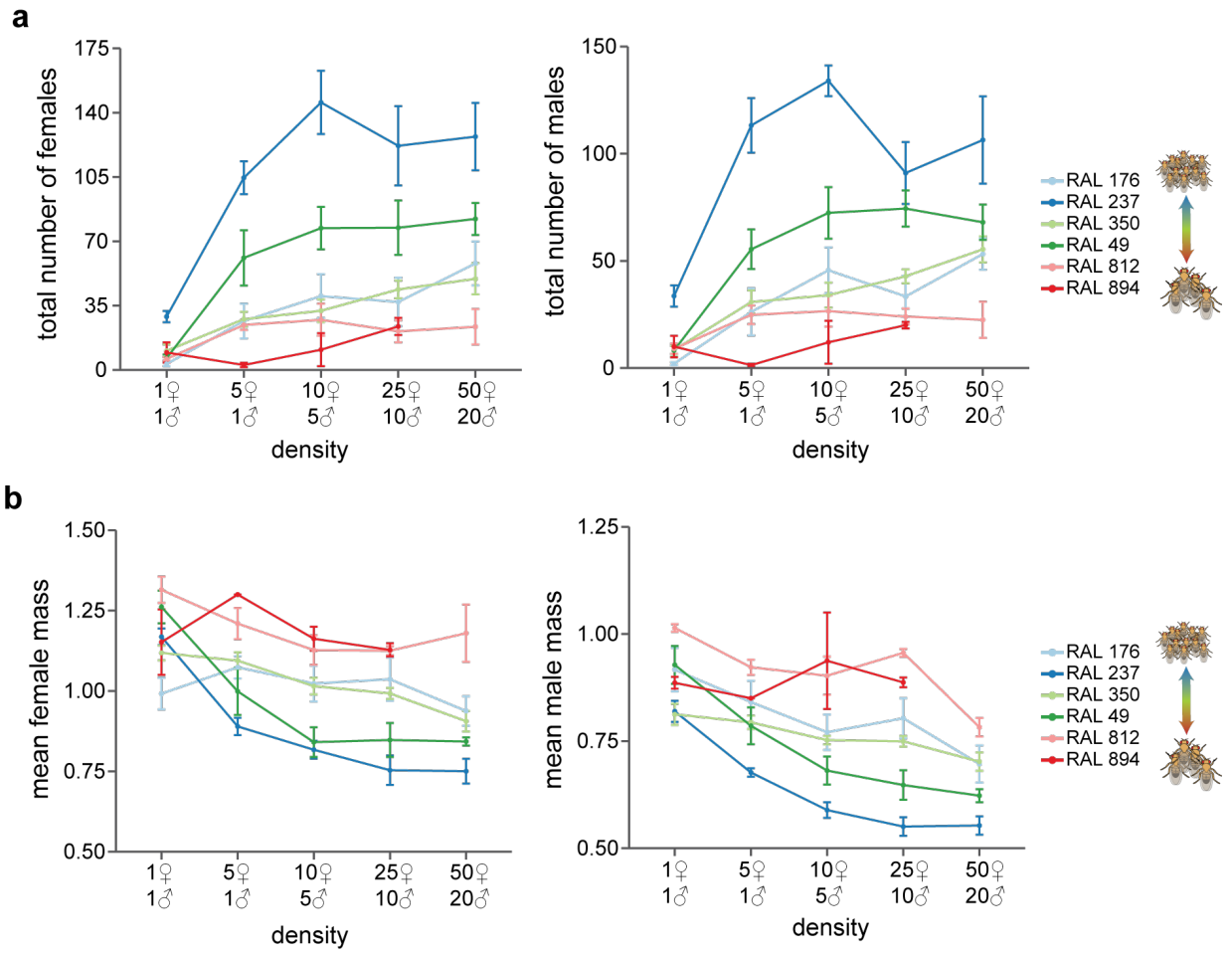


Figure 3.3 - Independent effects of parental density and line on offspring number and weight. **a)** Relationship between number of offspring and the density of parents. Each point represents the mean phenotype for a line at a particular density and errors bars show +/- 1 s.e. (n=2-7). **b)** Relationship between mean offspring weight and density of parents.

Functional validation of associated variants shows

We chose six candidate genes to validate for involvement with our fecundity phenotype - *hid, Sox21b, Rbp, CG8312, mub,* and *Ih*. Genes for validation were chosen based on having a variant with a high association with our offspring index, presence of multiple associated variants, and availability of mutant lines. We used mutant lines from the Exelixis gene disruption panel with a *piggyBac* construct in the gene of interest to validate our candidate genes. For four of the six candidate genes (*hid, Sox21b, CG8312,* and *mub*), we found that the available insertions severely impacted pupal and adult viability, to the point where we were unable to generate a

stable homozygous line to use in our validation experiments (Table 3.4). With the remaining genes, *Rbp* and *Ih*, we found opposite effects on the offspring index, with the *Ih* insertion strongly decreasing the offspring index, while the *Rbp* insertion slightly increased it (Figure 3.4a). Breaking it down to individual phenotypes (Figure 3.4b), we see that the *Ih* insertion significantly increases the number of offspring (reflected also in the strong mean weight decrease). Disrupting *Rbp* did not have a statistically significant impact on the number of offspring, but there was a small, but statistically significant increase, in offspring weight. Both *Ih* and *Rbp* play a role in nervous system function. *Ih* encodes a voltage-gated potassium channel and *Ih* mutants show defects in locomotion, proboscis extension, circadian rhythm, and lifespan[26,27]. *Rbp* encodes a protein involved in the organization of the presynaptic active zone and is instrumental in proper vesicle release[28] - mutations in *Rbp* can result in neurological and locomotor defects, and in some cases, lethality.

For genes where we had multiple mutant lines, we could qualitatively compare the effect of insertion site on the fecundity phenotypes. We find that for *Ih, Ih^{e01599}* was a viable line with increased fecundity, while Ih^{f01485} has impaired adult viability (Table 3.4). While both insertions are in introns, *Ih* transcription has been shown to be disrupted in Ih^{f01485}, but not in Ih^{e01599} [29]. The effect of the insertion on the final protein function of Ih^{e01599} is unknown, but RT-PCR shows that intronic insertion in Ih^{f01485} results in a null mutation[29].

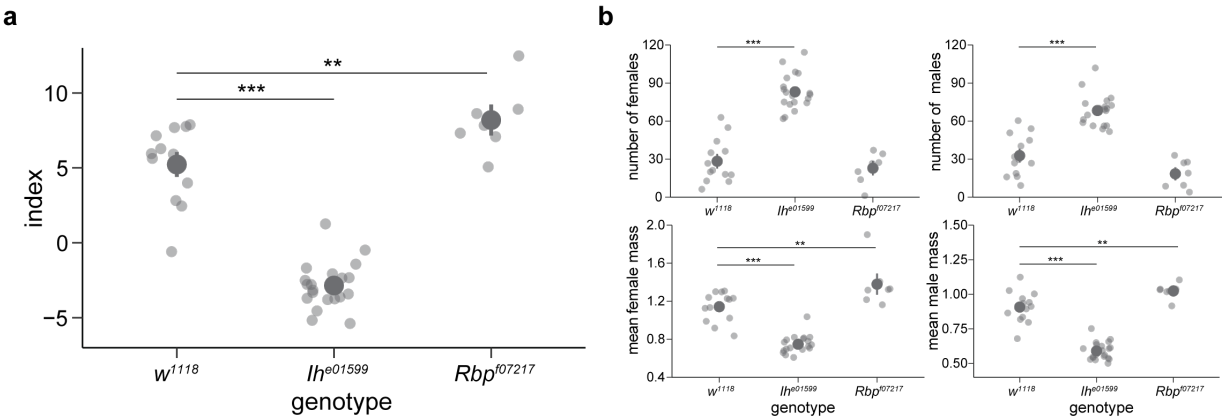


Figure 3.4 - Candidate gene validation using PBac{RB}*Ih*^{e01599} (n=10) and PBac{WH}*Rbp*^{f07217} (n=7), compared to their genetic background control, *w*^{1118}.
a) Impact of gene disruption on offspring index. **b)** Impact of gene disruption on individual phenotypes. The dark gray points and bars show the mean +/- 1 s.e., while the lighter gray points are replicates. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. *P*-values were calculated using Dunnett's test with a family-wise confidence level of 95%.

Table 3.4 - Mutant genotypes used in candidate gene validation and their fecundity phenotypes and gene functions.
Gene function information was retrieved from FlyBase[30]. Number of + or - is a qualitative representation of how much the measured phenotype increased or decreased as compared to the control genetic background line (only statistically significant differences are shown).

Genotype	Pupa?	Adult survival?	Fecundity phenotype	Gene function
PBac{PB}*hid*^{c01591}	yes	**no**	—	BIR domain binding protein; apoptosis regulator
PBac{WH}*Sox21b*^{f06429}	yes	**low**	—	transcription factor; transcription regulation and development
PBac{WH}*Rbp*^{f07217}	yes	yes	+ offspring weight	RIM-binding protein; presynaptic active zone organization
PBac{RB}*CG8312*^{e01204}	yes	**low**	—	transcription regulation
PBac{WH}*CG8312*^{f02825}	yes	**no**	—	transcription regulation
PBac{WH}*mub*^{f02647}	**no**	—	—	regulation of RNA splicing
PBac{WH}*Ih*^{f01485}	yes	**low**	—	voltage-gated K^+ channel;
PBac{RB}*Ih*^{e01599}	yes	yes	+++ offspring number - - - offspring weight	voltage-gated K^+ channel

Correlations among candidate gene expression and traits

We examined whether the expression levels of candidate genes used in the validation were correlated. Five of the six candidate genes were used - *CG8312* did not have available expression data. Using female expression data, we identified that *Ih* expression was positively correlated with *hid* expression and *Rbp* expression, following a multiple test correction (*Ih-hid: r* = 0.247, *p* = 2.5E-3; *Ih-Rbp: r* = 0.238, *p* = 3.1E-3) (Figure 3.5a). Using the male expression data and a multiple test correction, we find that *mub* is positively correlated with *hid, Rbp,* and *Ih* and negatively correlated with *Sox21b* (Figure 3.5b). We wanted to see whether the number of significant correlations we observed was higher than what you would get with random chance. To this end, we sampled five random genes from the expressed data and calculated how many significant correlations we would observe post-correction. We find that for both male and female expression, we have not enriched for a highly-correlated cluster of genes within our validation set (Figure 3.5c,d). Using the same approach on 17 of the candidate genes, we find a similar pattern (Figure 3.6).

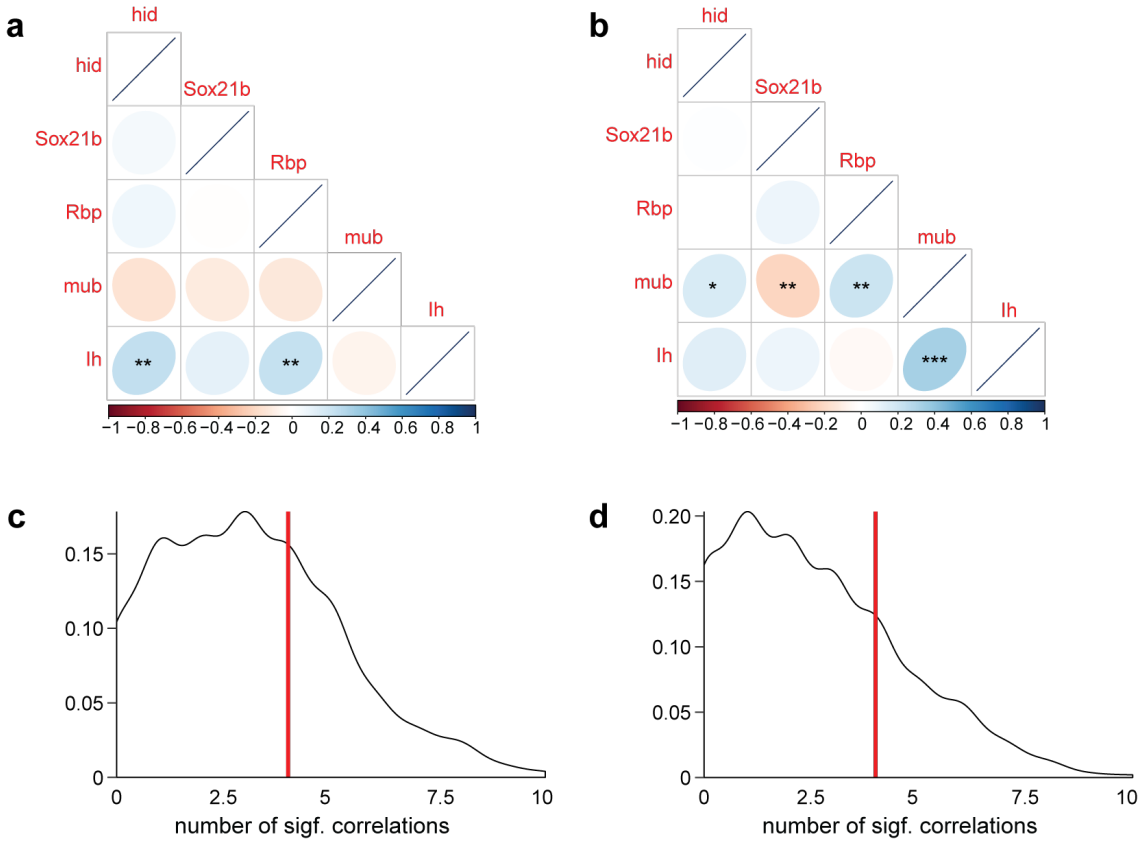


Figure 3.5 - Correlation of expression among candidate genes used in the validation experiment. **a)** DGRP female expression data. **b)** DGRP male expression data. Kernel density plot of number of significant correlations of expression of 5 randomly chosen genes (1000 samples) from **c)** female expression data and **d)** male expression data. Red line shows our observed number of significant correlations. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. *P*-values shown are corrected for multiple tests using the Benjamini-Hochberg method.

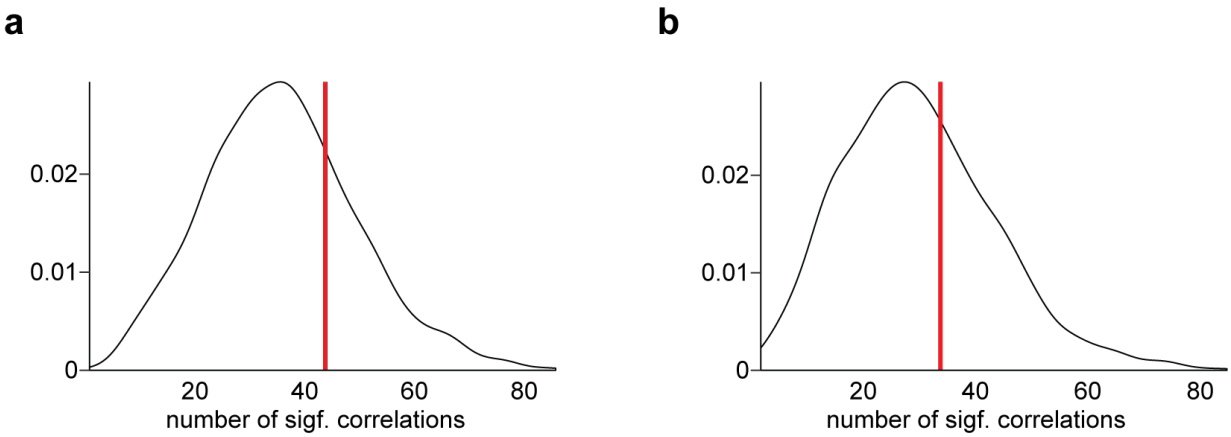


Figure 3.6 - Kernel density plots of the number of significant correlations of expression of 17 randomly chosen genes (1000 samples) from a) female expression data and b) male expression data.
Red line shows our observed number of significant correlations among 17 candidate genes examined. *P*-values were corrected for multiple tests using the Benjamini-Hochberg method prior to determining significance.

We also examined whether expression of the candidate genes was significantly correlated to the phenotypes measured in this study. We did not find evidence of strong correlations between the gene expression of our candidate genes and the phenotypes measured (Figure 3.7).

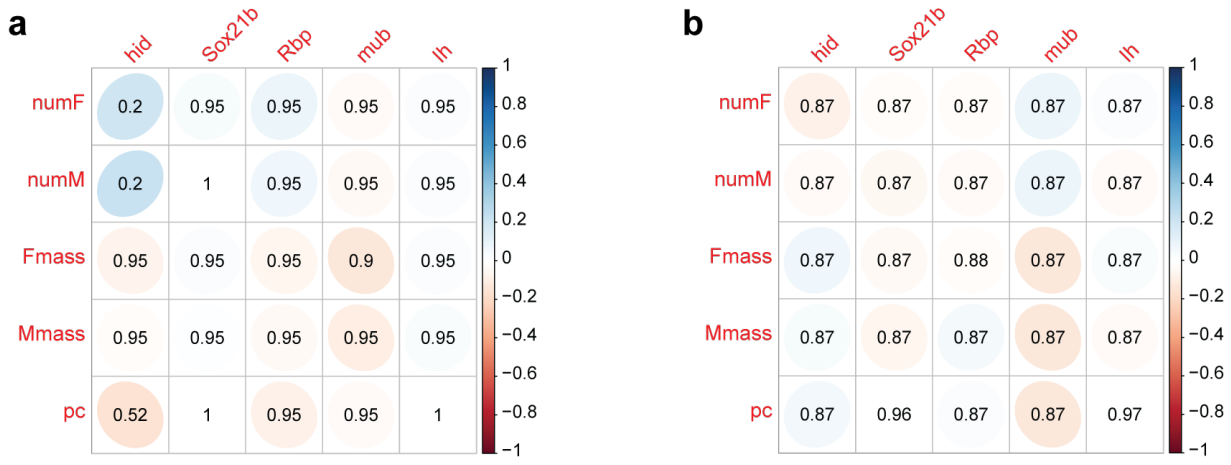


Figure 3.7 - Correlation of expression of candidate genes used in the validation experiment to phenotypes measured (pc = offspring index).
a) DGRP female expression data. **b)** DGRP male expression data. *P*-values shown are corrected using the Benjamini-Hochberg method.

We also measured correlations between traits measured in this study and traits measured in other DGRP studies. We looked at the following phenotypes: starvation resistance[23], chill coma recovery time[23], food intake[31], fecundity and body size[4], nutritional indices and weight[22], and developmental time and egg-to-adult viability under different densities[14]. Using a less stringent within-study multiple testing correction, we find that offspring weight measured in our study was significantly correlated with body size and mean weight measurements made in previous DGRP studies (Table 3.5). There was no correlation found between our measurements of total progeny number and fecundity measurements ($p > 0.1$ for all comparisons). In addition, there was no correlation found between progeny weight and food intake ($p > 0.1$ for all comparisons). We found significant positive correlations between female starvation resistance and female ($r = 0.184$, $p = 0.028$) and male ($r = 0.168$, $p = 0.046$) weight measured in our study, but the correlations did not remain significant after a multiple testing correction. In addition, we found that male and female development time under a high larval density treatment (measured in 31 DGRP lines) was negatively correlated with the number of offspring, but positively correlated with offspring weight. Egg-to-adult viability under high larval density treatment was positively correlated with offspring number and negatively correlated with offspring weight. Though we observed a trend in the relationships, only a few correlations were significant, and none remained significant after multiple testing correction ($p > 0.1$ for all correlations).

Table 3.5 - Correlations of phenotypes measured in our study (phenotype) with traits measured in previous DGRP studies (phenotype (ref)).
Only correlations that remained significant post-multiple testing correction are presented. The symbols are as follows: "-y" denotes a low yeast diet, "+y" denotes a high yeast diet, "+" denotes a high glucose diet, "-" denotes a low glucose diet, and "cb" = overall effect when glucose diets are combined. *P*-values are corrected for multiple tests (within each study) using the Benjamini-Hochberg method.

Phenotype	Phenotype (ref)	*r*	*p*-value	Reference
mean weight(♂)	body size(-y)	0.264	1.03E-2	Durham et al., 2014[4]
mean weight(♂)	body size(+y)	0.277	1.03E-2	Durham et al., 2014[4]
mean weight(♀)	mean weight(♂-)	0.409	1.38E-4	Unckless et al., 2015[22]
mean weight(♂)	mean weight(♂-)	0.387	2.74E-4	Unckless et al., 2015[22]
mean weight(♀)	mean weight(♂+)	0.442	5.49E-5	Unckless et al., 2015[22]
mean weight(♂)	mean weight(♂+)	0.412	1.38E-4	Unckless et al., 2015[22]
mean weight(♀)	mean weight(♂^{cb})	0.459	2.54E-5	Unckless et al., 2015[22]
mean weight(♂)	mean weight(♂^{cb})	0.430	5.49E-5	Unckless et al., 2015[22]

Discussion

We investigated whether there was a genetic basis for adult offspring number and weight under space and resource limitation. We found that DGRP lines varied in the number of offspring produced and their mean weight, with the number negatively correlated with weight. Using a single "offspring index" derived from the first principal component of our offspring phenotypes, we found variants associated with variation in the offspring index among the lines i.e. variation from many low-weight offspring to few high-weight offspring. We examined the effects of gene disruption on six candidate genes and found that for all tested alleles of *hid, Sox21b, CG8312,* and *mub*, as well as one insertion allele of *Ih*, gene disruption caused phenotypes ranging from pupal lethality to low adult survival. Disruption of *Rbp* caused a small increase in offspring weight and an insertion allele of *Ih* caused a large increase in offspring number coupled with a decrease in offspring weight. While we did find significant correlations in gene expression between the candidate genes, the number of significant correlations did not exceed what could be found by random chance. When comparing our measured phenotypes to life history phenotypes measured in other DGRP studies, we found consistency in our body weight measurements and other measurements of weight and body size, but surprisingly, we did not find a relationship between offspring number and fecundity.

We found that, similar to fecundity[4], there is a polygenic basis to offspring number and weight. While a few of the candidate genes we found were previously associated with fecundity[4], most were novel. Along with the polygenic nature of the traits, we were not able to find significant enrichment of genes involved in particular biological processes or molecular functions, though given our limited sample size of genes, only a very strong enrichment would have been able to pass the significance threshold. Even though only two of the six of the candidate genes tested in our validation experiments had roles in developmental processes, we found that gene disruption was severe enough to the survival of the pupal or adult stages in most cases to prevent us from

carrying out the validation experiments. These results indicate that offspring viability may be strongly dependent on the proper functioning of a varied suite of genes, such that even a disruption to a single gene is enough to completely impair successful development of the next generation.

We hypothesize that an omnigenic model[32,33] could explain the varied suite of genes associated with offspring number and weight phenotypes. One of the predictions of the omnigenic model is that all genes expressed at the appropriate developmental point and/or in a particular tissue contribute to the trait, leading to many loci that are weakly associated with the trait of interest. The omnigenic model also predicts that there should be a network of genes whose action is essential. Interestingly, even though we observe that disruptions to our candidate genes lead to strong effects, we do not find an enrichment for correlated expression among them. We were also unable to find enrichment for particular biological pathways or molecular processes among our candidate genes. Since we were unable to identify network relationships between our candidate genes, our results do not seem to explicitly support an omnigenic model, but perhaps reflect what was proposed in Fisher's "infinitesimal model"[34], where at its limits, a quantitative trait is made up of infinitely small contributions of infinitely many genes.

We find that mutations in *Rbp* and *Ih* caused opposite phenotypes. Mutations in *Ih* increased the offspring number, and given the constraints of the vial, decreased offspring weight. *Ih* has not been previously implicated in fecundity phenotypes - while that does not preclude that *Ih* mutants may have laid more eggs, our results could indicate that there was a higher egg-to-adult viability that led to the increase in offspring number and increased larval competition. Disruption of *Rbp* only increased weight without significantly affecting offspring number. A small effect on offspring number could have been obscured by limited sample size, but our results would still indicate that the weight increase was more prominent than the decrease in offspring

number. A decoupling of weight from offspring number shows that there is an independent axis where offspring weight can increase even though the level of larval competition and other density effects remain the same. This notion is supported by the second principal component of our dataset which is positively loaded for both weight and offspring number, indicating that a trade-off between the two is not mandatory (Table 3.2).

We note that even within the same gene, different disruptions can lead to different phenotypes. Both the lh^{e01599} and lh^{f01485} alleles are intronic insertions that affect most transcripts, but lh^{f01485} shows impaired viability, while lh^{e01599} shows an increase in offspring number. The lh^{f01485} allele was shown to abolish transcription of all *lh* transcripts[29], but the lh^{e01599} allele appeared to still have wild-type levels of expression[29]. Based on the insertion position, the lh^{e01599} allele would affect eight of the eleven *lh* transcripts - the phenotype difference between the alleles could depend on specific transcripts, though this is speculative. These results highlight a general caveat about using gene disruption lines to make conclusions about the exact role of the gene in determining the phenotype, rather than a general conclusion about whether or not the gene plays a role at all.

When comparing our measured phenotypes with phenotypes measured in other DGRP studies, we found that there was correspondence between our weight metrics and body weight/size metrics from other studies, affirming that differences between DGRP lines remain consistent across different study environments. We did not, however, see a relationship between the number of offspring and fecundity measures - this is unexpected, as one would expect that the number of eggs laid should correspond to the number of offspring. A possible explanation for the lack of correlation could be due fecundity being assayed in individual females[4], whereas we housed females in groups of 10 for our assay. Number of eggs laid per female was shown to decrease in more crowded conditions[12,35], and with the presence of differential genotype

effects[35], could lead to the lack of correlation observed. We also observe that, within 31 DGRP lines, more offspring is correlated with a lower developmental time and higher egg-to-adult viability under a high-density treatment. Though these correlations were not significant under the limited sample size of lines, it does indicate that lines producing more offspring may have adapted to do well under high-density conditions[14,18].

Overall, our results point to a polygenic basis for offspring number and weight, with validation of the candidate genes implicating an array of biological processes in controlling adult stage viability. Combining our results with results from studies on other *Drosophila* life-history traits, we find support for the idea that traits closely related to fitness may be directly and/or indirectly affected by a large set of genes, perhaps ultimately encompassing the vast majority of functional genes.

Methods

Drosophila stocks and husbandry

We analyzed 143 lines from the *Drosophila* Genetic Reference Panel[23]. All stocks were maintained at 23ºC, 12L:12D cycle and reared on a yeast, cornmeal, and dextrose media (23g yeast/L, 30g cornmeal/L, 110g dextrose/L, 6.4g agar/L, and 0.12% Tegosept). Experiments were carried out in polystyrene narrow culture vials (25 x 95 mm, #32-109, Genesee Scientific). Mutant lines for validation were obtained from the Exelixis collection at Harvard Medical School (Boston, MA)[36] The w^{1118} (#6326) genetic background was obtained from the Bloomington Drosophila Stock Center (BDSC) as a control for the candidate gene validation mutant lines.

Phenotypic measurements of DGRP lines

Bottles were seeded with 15 females and 15 males at 12L:12D to generate the parent flies that would go on to lay eggs for the experiment. 10 females and 5 males (2-5 days old) from the parent flies were placed in each of the three vials (along with ~30 grains of dry yeast) and left to lay for 2 days at 23ºC, 12L:12D. The parent flies were removed, and the vials were kept at 23ºC, 12L:12D until progeny began to eclose. From the start of eclosion, all of the vials were examined every day over the course of 10 days. The number of females and males for each vial was recorded, as well as the total wet weight (to 0.1 mg accuracy) of the females and the total combined wet weight of the females and males. 10 days was chosen as a conservative time course to prevent the next generation from being included in the analysis. 143 lines were tested in two batches. 134 lines were done in the first batch - the second batch included the 35 lines that did not have three complete vial replicates in the first batch, and 9 lines that were not tested at all in the first batch.

Genome-wide association mapping for offspring index

The four phenotypes measured in this screen were the total number of females (males) eclosed and the average weight of a female (male) fly. Since lines were assayed in two batches, the 35 lines that were tested in both batches were used to check for a batch effect. The batch effect for each of the four phenotypes was corrected by applying an offset (difference of mean phenotype between the first and second batch) calculated from the overlapping set of lines.

For offspring total counts, a random intercept linear model was used to calculate the random effect of each DGRP line.

$$Y_a = \mu + line_a + \varepsilon$$

where $Y_a$is the phenotype measure for that particular line combination, $line_a$ is the random effect of DGRP line, and ε is the error term. For mean offspring weights, a random intercept generalized linear model was used (model formula as above), assuming a gamma distribution of mean weights and a logarithm link function. The LME4 package (v1.1-21) in R (v3.5.3) was used to make the models. To estimate heritability for each phenotype, we used the R package brms (v2.8.0) and our models to estimate variance explained by line and divided it by the sum of all sources of variance.

Since we were interested in a single metric to summarize the number of offspring and their average weight, we used *prcomp* with scaling in R's factoextra package (v1.0.5) to generate the principal components of the dataset. The first principal component explained 71% of the variance of the dataset, so we chose to use the value of the rotated data (line phenotype values

multiplied by the rotation values of the first principal component) as our summary phenotype, which we called the offspring index (Table 3.2).

We used the DGRP2 webtool (Huang *et al*. 2014) to perform a mapping of variants associated with the offspring index. The webtools controls for inversion and *Wolbachia* infection status prior to mapping. We chose a significance threshold of $p < 1E\text{-}5$ to identify variants for further consideration.

Parental density analysis

Six DGRP lines were chosen from the overall screen based on their offspring index values - two lines with an extreme negative index (RAL 176: -3.18, RAL 327: -2.87), two lines with an index around zero (RAL 49: -0.04, RAL 350: -0.22), and two lines with a highly positive index (RAL 812: +3.58, RAL 894: +3.79). Five density treatments were chosen: 1♀,1♂, 5♀,1♂, 10♀,5♂, 25♀,10♂, 50♀,20♂. Lines were reared in bottles at 23ºC, 12L:12D and a minimum of 5 (maximum of 10) replicates per line-density combination were set up, with the exception of RAL 894 at the highest density where no replicates were set up due to an insufficient number of flies. Parent flies were 3-5 days old at the time of experiment set-up and the egg-laying conditions were matched to the DGRP phenotype screen. Given the slowed eclosion of offspring in the high-density treatments, an extended window of 25 days was used to evaluate offspring phenotypes. Daily records after the 12th day were monitored for unusual spikes in offspring number that could be indicative of a large number offspring from the next generation, but as flies were removed daily from the vials and not allowed to lay for extended periods of time, a strong influence of next-generation offspring on the counts was deemed unlikely.

To evaluate the impact of DGRP line and density on offspring counts, a generalized linear model was used, assuming a negative binomial distribution of the response along with a

logarithm link function. A negative binomial model was used because of the large spread and right skew of the offspring count distribution that made a regular linear model a poor fit. For offspring weight, a generalized linear model with a gamma response distribution and logarithm link function was used. The model formula was as follows:

$$Y_{ab} = \mu + line_a + density_b + line_a * density_b + \varepsilon$$

where Y_{ab} is the phenotype measure for a particular line-density combination, $line_a$ is the fixed effect of a DGRP line, $density_b$ is the fixed effect of a density treatment, $line_a * density_b$ is the interaction term, and ε is the error term. DGRP line and density were treated as ordered factors in the model. As above, the LME4 package was used, in addition to the MASS package (v7.3-51.1) for the negative binomial model. To evaluate the significance of a predictor, a likelihood-ratio test from the *anova* function in R's STATS (v3.5.3) package was used to compare models with and without the predictor.

Validation of candidate genes

Mutant lines for validation of candidate genes were obtained from the Exelixis collection[36] for six genes containing variants associated with the offspring index at a $p < 1E\text{-}5$ threshold. The mutant lines are as follows: PBac{PB}*hid*^{c01591}, PBac{WH}*Sox21b*^{f06429}, PBac{WH}*Rbp*^{f07217}, PBac{RB}*CG8312*^{e01204}, PBac{WH}*CG8312*^{f02825}, PBac{WH}*mub*^{f02647}, PBac{WH}*lh*^{f01485}, and PBac{RB}*lh*^{e01599} (#17970 BDSC). All lines were made homozygous for the insertion prior to testing. The genetic background for this gene disruption panel was *w*^{1118}. To generate the parental flies for each mutant line and the control, ~30 females and 10 males were placed in bottles at 22ºC, 12L:12D to lay for 7-10 days to generate the experimental flies. 10 females and 5 males from the experimental flies were put into a single vial with ~30 yeast pellets (ten

replicates per mutant line) and allowed to lay for 2 days at 22ºC, 12L:12D. The parental flies were removed, and the progeny was phenotyped the same way as for the DGRP phenotype screen. The validations were done in two batches staggered by one week, but there was no significant batch effect, therefore replicates were combined across batches. Mutant lines were compared to the w^{1118} genetic background control using Dunnett's test with a family-wise confidence level of 95%. Offspring index for the mutant lines was calculated using the principal component loadings calculated from the DGRP data.

Gene expression correlations and correlations with other traits

Gene expression data was obtained from Huang *et al.* (2015)[37]. We calculated the Pearson correlation between the expression of candidate genes tested in the validation experiment, as well as the correlation between the expression of those genes and the phenotypes measured in our screen. We also correlated the traits measured in this study against similar or putatively related traits measured in other DGRP studies[4,22,23,31]. All analysis was done in R (v3.5.3).

References

1. Flatt, T. Life-History Evolution and the Genetics of Fitness Components in Drosophila melanogaster. *Genetics* **214**, 3–48 (2020).

2. Rose, M. R. & Charlesworth, B. Genetics of Life History in Drosophila Melanogaster. Ii. Exploratory Selection Experiments. *Genetics* **97**, 187–196 (1981).

3. Leips, J., Gilligan, P. & Mackay, T. F. C. Quantitative Trait Loci With Age-Specific Effects on Fecundity in Drosophila melanogaster. *Genetics* **172**, 1595–1605 (2006).

4. Durham, M. F., Magwire, M. M., Stone, E. A. & Leips, J. Genome-wide analysis in *Drosophila* reveals age-specific effects of SNPs on fitness traits. *Nat. Commun.* **5**, 1–8 (2014).

5. Tatar, M., Promislow, D. E. L., Khazaeli, A. A. & Curtsinger, J. W. Age-Specific Patterns of Genetic Variance in Drosophila melanogaster. II. Fecundity and Its Genetic Covariance With Age-Specific Mortality. *Genetics* **143**, 849–858 (1996).

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

7. Nunney, L. & Cheung, W. The Effect of Temperature on Body Size and Fecundity in Female Drosophila melanogaster: Evidence for Adaptive Plasticity. *Evolution* **51**, 1529–1535 (1997).

8. Djawdan, M., Sugiyama, T. T., Schlaeger, L. K., Bradley, T. J. & Rose, M. R. Metabolic Aspects of the Trade-Off between Fecundity and Longevity in Drosophila melanogaster. *Physiol. Zool.* **69**, 1176–1195 (1996).

9. Bergland, A. O., Chae, H., Kim, Y.-J. & Tatar, M. Fine-Scale Mapping of Natural Variation in Fly Fecundity Identifies Neuronal Domain of Expression and Function of an Aquaporin. *PLOS Genet.* **8**, e1002631 (2012).

10. Grimaldi, D. & Jaenike, J. Competition in Natural Populations of Mycophagous Drosophila. *Ecology* **65**, 1113–1120 (1984).

11. Prout, T. & Barker, J. S. Ecological aspects of the heritability of body size in Drosophila buzzatii. *Genetics* **123**, 803–813 (1989).

12. Barker, J. S. F. Adult population density, fecundity and productivity in Drosophila melanogaster and Drosophila simulans. *Oecologia* **11**, 83–92 (1973).

13. Scheiring, J. F., Davis, D. G., Ranasinghe, A. & Teare, C. A. Effects of Larval Crowding on Life History Parameters in *Drosophila melanogaster* Meigen (Diptera: Drosophilidae). *Ann. Entomol. Soc. Am.* **77**, 329–332 (1984).

14. Horváth, B. & Kalinka, A. T. Effects of larval crowding on quantitative variation for development time and viability in Drosophila melanogaster. *Ecol. Evol.* **6**, 8460–8473 (2016).

15. Baldal, E. A., van der Linde, K., van Alphen, J. J. M., Brakefield, P. M. & Zwaan, B. J. The effects of larval density on adult life-history traits in three species of Drosophila. *Mech. Ageing Dev.* **126**, 407–416 (2005).

16. Economos, A. C. & Lints, F. A. Growth rate and life span in Drosophila. I. Methods and mechanisms of variation of growth rate. *Mech. Ageing Dev.* **27**, 1–13 (1984).

17. Miller, R. S. & Thomas, J. L. The Effects of Larval Crowding and Body Size on the Longevity of Adult Drosophila Melanogaster. *Ecology* **39**, 118–125 (1958).

18. Santos, M., Fowler, K. & Partridge, L. Gene–environment interaction for body size and larval density in *Drosophila melanogaster*: an investigation of effects on development time, thorax length and adult sex ratio. *Heredity* **72**, 515 (1994).

19. Moghadam, N. N., Holmstrup, M., Manenti, T. & Loeschcke, V. Phospholipid fatty acid composition linking larval-density to lifespan of adult *Drosophila melanogaster*. *Exp. Gerontol.* **72**, 177–183 (2015).

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

21. Jumbo-Lucioni, P. *et al.* Systems genetics analysis of body weight and energy metabolism traits in Drosophila melanogaster. *BMC Genomics* **11**, 297 (2010).

22. Unckless, R. L., Rottschaefer, S. M. & Lazzaro, B. P. The Complex Contributions of Genetics and Nutrition to Immunity in Drosophila melanogaster. *PLOS Genet.* **11**, e1005030 (2015).

23. Mackay, T. F. C. *et al.* The *Drosophila melanogaster* Genetic Reference Panel. *Nature* **482**, 173 (2012).

24. Nelson, C. S. *et al.* Cross-phenotype association tests uncover genes mediating nutrient response in Drosophila. *BMC Genomics* **17**, 867 (2016).

25. Mi, H., Muruganujan, A., Ebert, D., Huang, X. & Thomas, P. D. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res.* **47**, D419–D426 (2019).

26. Chen, Z. & Wang, Z. Functional study of hyperpolarization activated channel (Ih) in Drosophila behavior. *Sci. China Life Sci.* **55**, 2–7 (2012).

27. Gonzalo-Gomez, A. *et al.* Ih Current Is Necessary to Maintain Normal Dopamine

Fluctuations and Sleep Consolidation in Drosophila. *PLOS ONE* **7**, e36477 (2012).

28. Liu, K. S. Y. *et al.* RIM-Binding Protein, a Central Part of the Active Zone, Is Essential for Neurotransmitter Release. *Science* **334**, 1565–1569 (2011).

29. Hu, W., Wang, T., Wang, X. & Han, J. Ih Channels Control Feedback Regulation from Amacrine Cells to Photoreceptors. *PLOS Biol.* **13**, e1002115 (2015).

30. Thurmond, J. *et al.* FlyBase 2.0: the next generation. *Nucleic Acids Res.* **47**, D759–D765 (2019).

31. Garlapow, M. E., Huang, W., Yarboro, M. T., Peterson, K. R. & Mackay, T. F. C. Quantitative Genetics of Food Intake in Drosophila melanogaster. *PLOS ONE* **10**, e0138129 (2015).

32. Boyle, E. A., Li, Y. I. & Pritchard, J. K. An expanded view of complex traits: from polygenic to omnigenic. *Cell* **169**, 1177–1186 (2017).

33. Zhang, W., Reeves, R. G. & Tautz, D. Testing the omnigenic model for a behavioral trait in Drosophila melanogaster. *bioRxiv* 639955 (2019) doi:10.1101/639955.

34. Turelli, M. Commentary: Fisher's infinitesimal model: A story for the ages. *Theor. Popul. Biol.* **118**, 46–49 (2017).

35. Ohnisni, S. Effects of Population Density and Temperature Condition on Fitness in *Drosophila melanogaster*. *Jpn. J. Genet.* **51**, 305–314 (1976).

36. Thibault, S. T. *et al.* A complementary transposon tool kit for Drosophila melanogaster using P and piggyBac. *Nat. Genet.* **36**, 283–287 (2004).

37. Huang, W. *et al.* Genetic basis of transcriptome diversity in Drosophila melanogaster. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E6010-6019 (2015).

Discussion

The work in my dissertation focused on the role of environment in behavioral trait variability, the adaptive value of behavioral variability, and the impact of genetics on life-history trait variation. While we found that enriching the environment of *D. melanogaster* produced a small overall increase in behavioral variability, the impact of environment alone was overwhelmed by the effect of genetic background and a three-way interaction between genotype, environment, and behavioral metric. Next, we focused on an ecologically relevant behavioral trait, thermal preference, in order to examine the adaptive value of environmentally produced behavioral variability as a bet-hedging strategy. Our modeling supported a climate-dependent advantage for bet-hedging vs. adaptive tracking (a strategy where thermal preference variation is primarily due to a genetic component). Bet-hedging was the predicted preferred strategy in most of the contiguous U.S., while the predicted adaptive tracking advantage was limited to southern areas with warm and mild climates. Empirical tests of modeling predictions found that heritability in thermal preference in a particular region matched with the predicted bet-hedging advantage - regions with low bet-hedging advantage (favored for adaptive tracking) had the highest heritability. Seasonal patterns in mean preference and variability in thermal preference, while yielding more mixed results across the sampled regions, still supported the notion that thermal preference in wild *D. melanogaster* populations can be determined by a bet-hedging strategy. While behavioral trait variation that impacts fitness may not need to have a large genetic component, we found that life-history trait variation has a large genetic component. The many genes associated with life-history trait variation are not confined to a particular cellular process - we observed strong impacts on offspring metrics for mutations in genes that have no known roles in fecundity or growth. Our findings, in addition to previous studies on life-history traits, leads us to speculate that for traits closely related to fitness, practically every gene in the genome will contribute through either indirect or direct action on the cellular pathways involved

in the trait. I believe that the work discussed here has provided valuable insights on the origins and adaptive value of micro-environmental contributions to behavioral variability and additional support for the polygenic basis of life-history traits.

One of the common themes that emerged from our studies of behavioral variability is that there is a substantial amount of interaction between genotype, macro-, and micro-environmental influences in determining the observed degree of variability. Genotype-by-(macro-)environment interactions have been commonly observed under laboratory-imposed treatment regimes, and genotype has also been shown to play a role in the degree of micro-environmental plasticity. Our studies demonstrated another connection, one that is perhaps not surprising given the inter-relationships already described - genotype and macro-environment interact to influence micro-environmental plasticity. We tested this idea explicitly by examining the impact of enrichment on the variability of six behavioral metrics (four for turning bias and two for phototactic preference). To our surprise, the impact of the enrichment treatment, after controlling for behavioral metric and genotype, was quite small, though always positive. The biggest contributing factors to the degree of variability were genotype and genotype-by-enrichment interactions, though the specific magnitudes of genotype, enrichment, and genotype-by-enrichment contributions depended on the behavior metric. Even within a particular suite of behavioral metrics, for example those related to turning bias locomotion, there was no predictable pattern of how much each component contributed to variability and/or the sign of the contribution. While this variety of effects on variability could potentially be attributed to the particular nature of rearing flies in an enriched environment, we also found that variability in thermal preference was affected by rearing temperature. In this case, besides temperature, the only other unavoidable difference was the incubator used. Evidence seems to point strongly to the existence of not only the canonical genotype-by-environment effects on trait means, but also genotype-by-environment effects on trait variabilities.

This finding opens up a discussion on what are the direct and indirect causes of micro-environmental plasticity (also referred to as variability in this text). Organisms do not develop in a vacuum, and establishment of the final phenotype is a result of innumerable internal and external interacting factors. While there is an undeniable determinism in certain aspects of development - for example, flies that emerge from the same vial all have the same body plan and do not each resemble a unique Picasso interpretation - the research in this dissertation and previous work[1] demonstrated that not all traits are under strict control. Of the potential factors that can shape a phenotype, factors that can be reliably measured and controlled compose only a subset - the rest, which shape observed variability, are stochastic and unobservable[2]. Potential factors at all levels of organization could be implicated - transcriptional fluctuations, epigenetic modifications, and cell differentiation (in the case of behavior, neuronal wiring[2]), for example. One can imagine potential cascading effects where small perturbations early on in gene expression affect epigenetic modifications, which then create variability in cell differentiation, more permanently altering gene expression patterns in later stages of development. For thermal preference, stochastic differences in expression of peripheral warming[3,4] and cooling[3] sensors could be the proximal cause that underlies the individual differences, or in deeper brain areas, such as differential cyclic adenosine mono-phosphate[5] or dopamine signaling[6] in the mushroom body. In support of this hypothesis, I found preliminary evidence of a negative correlation between expression of *trpA1*, a warmth sensor, and thermal preference within an isogenic line (Figure 4.1). What is currently not well-described is how the degree of stochasticity in these proximal mechanisms (e.g. gene expression) is established through genetic and/or environmental means.

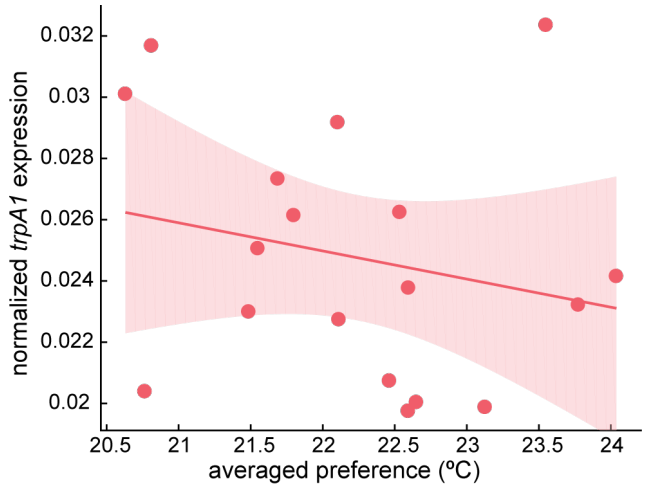


Figure 4.1 - *trpA1* expression is negatively correlated with thermal preference.
Pearson correlation (r = -0.23, p = 0.37) of *trpA1* expression with thermal preference of females from an isogenic line (n=18). *trpA1* expression was measured using digital droplet PCR on RNA from individual brains and normalized using expression of a pan-neuronal gene CG16779.

A possible process by which genotype and external environment can influence the degree of variability is by genotype affecting how robust the developmental program is under the pressure of environmental perturbations. A developmental program that is robust is said to be canalized[7], or able to produce consistent phenotypes despite mutational and/or environmental forces. Canalizing mechanisms that buffer phenotypes should be insensitive to the type of force they are buffering against (genetic or environmental)[8], since genotype and environment act together to determine phenotype - buffering against just one would be less effective. In addition, mutations and environmental fluctuations are unpredictable. Therefore, it is likely that the buffering mechanism is non-specific and participates in many cellular processes. A promising candidate is *Hsp90*, a heat-shock protein that is involved in many signaling processes as a stabilizer of conformationally plastic signaling proteins[9]. Downregulation of *Hsp90* in *Drosophila*[9] and *Arabidopsis*[10,11] increases morphological variance, likely due to *Hsp90*-regulated processes destabilizing under stochastic environmental fluctuations (either internal or external to the organism) and, in certain instances, previously cryptic genetic variation. The chaotic dynamics of cellular processes can amplify potentially small deviations in signaling cascades into large-

scale changes. Interestingly, *Hsp90* is not the only identified candidate to regulate variability. In *Arabidopsis,* QTL mapping for variability in flowering traits revealed tight links to the *ERECTA (ER)* locus[12], which encodes a protein kinase that functions in cell-cell signaling and developmental processes. In contrast to *Hsp90*, which acts to canalize traits, the wild-type *ER* allele decanalizes the leaf number trait. Therefore, *ER* can be considered as an amplifier of environmental perturbations[12]. In *Drosophila,* a GWAS on variability in three traits[1] revealed 36 associated genes in addition to *Hsp90* that are associated with variability in startle response - these genes function in nervous system development and behavior. GWAS for variability in chill coma recovery time and starvation resistance did not show enrichment for significantly associated variants, even though broad-sense heritability was high. The overall message from the described studies is that the genes involved in determining the degree of variability act during development, even for behaviors[1,13].

This result for behaviors is interesting as behaviors are usually considered to be more plastic than morphological traits in the adult organism e.g. experience-dependent learning, but evidence points to a component of behavioral variability established prior to the adult stage. A GWAS on *D. melanogaster* turning bias variability also found associated variants in genes enriched for functional roles in the nervous system[13]. One such candidate gene is *Tenascin accessory* (*Ten-a*), which encodes for a signaling protein that functions in proper formation of neuromuscular junctions and patterning of a locomotion center in the central brain. Knockdown of *Ten-a* increases variability in turning bias, but only when the knockdown was done during pupal development, when *Ten-a* expression peaks and brain structures responsible for locomotion form[13]. Encoding of variability in the developmental stages can explain why we observe persistence in behavioral traits (a “personality”[14]) over an organism’s lifetime. On the other hand, we also observe some decay in behavioral persistence in *D. melanogaster* over

time[15], which indicates that there is plasticity in behavioral traits. The decay depends on the behavior and may reflect on how much input adult experience has on shaping the behavior.

Given that there is a genetic basis to variability, it is a trait that can respond to selection pressures. As a bet-hedging strategy, trait variability should ideally evolve to match the environmental variance[16–18]. On the other hand, in the extreme case of an unchanging environment with one optimal phenotype, a genotype that produces a variable phenotype will be at a fitness disadvantage compared to one that does not. We found that maintenance under general laboratory conditions does not select against different degrees of variability, likely because of relaxed selection pressures (*ad libitum* food, minimal pathogen load, etc.). Yet, due to the environmental component in the degree of variability, the expression of variability that we see under standard laboratory conditions does not necessarily equate to what we would observe under natural conditions. With this in mind, our laboratory-based approach in assessing the degree of variability in thermal preference across different regions is limited in how well the results translate to wild populations. We found heterogeneity in thermal preference variability when comparing isofemale lines, but barely any differences in variability when comparing sites, especially as they correlate to predicted bet-hedging advantage. It is possible that working with laboratory-reared flies and deconstructing variability by isofemale line altered the genotype-by-environment interactions that determine thermal preference variability, possibly obscuring an actual effect present in wild populations.

The environmental component can also complicate selection for an optimal level of variability, given the broad functions of genes that may be responsible for establishing the degree of variability. For example, there may exist an optimal variability in thermal preference given fluctuations in seasonal weather patterns, yet observed variability in the wild is likely a function of the interaction between genotype and environmental fluctuations across many axes, possibly

orthogonal or on different time scales to the seasonal weather patterns e.g. nutrition, predation. There is no clear evidence that the genes responsible for setting the degree of variability in a particular trait are specific to the trait in question and that the variability is only responsive to environmental fluctuations that are directly associated with the trait e.g. predation risk and startle response or temperature changes and thermal preference. *Hsp90* has been identified as a locus for controlling variability at numerous traits - if evolution toward an increased variability in startle response takes place via changing allele frequencies at *Hsp90* (perhaps because of unpredictable predation risk[19]), it may be counteracted by suboptimal increased variability in morphology. Counter to the *Hsp90* example, no cross-trait genetic correlation in variability was found when a GWAS was performed on startle response, starvation resistance, and chill coma recovery variability[1]. In this case, it may be possible to increase variability in one trait without affecting others. The *ER* locus in *Arabidopsis* is currently implicated only in the variability of a single flowering trait, but it is unclear if there is specificity in what environmental perturbations cause the variability. In *Drosophila, Ten-a* is proposed to control turning bias through its functions in nervous system development, but it is also unclear if there is specificity in what external/internal environmental forces turn bias variability is sensitive to. Given variability in a myriad of traits is still observed under standardized laboratory conditions, far removed from most natural experiences, we can surmise that there is no strict specificity. Therefore, it becomes difficult to imagine a scenario where variability in a trait can be specifically tuned to a single external influence. It seems likely that environmental pressures can favor increased or decreased variability, but that the actual degree of variability is set by complex interactions between multiple environmental and genetic factors.

So far, the discussion has centered around the complexity of interactions between the environment and genotype in determining trait variation. Yet, trait variation can also depend on interactions among genes. For life-history traits, including offspring number and weight studied

here, trait variation comes from the small contributions of many variants across the genome, as opposed to a few variants with large effects. Interestingly, the set of variants affecting a particular life-history trait seems to expand with every new study. In our study on offspring number and weight, we found associated genes and variants that had not been previously identified to affect fecundity or growth. The omnigenic model of complex traits[20], discussed mostly in the context of disease risk, stipulates that complex traits are affected by a limited number of core genes that play a direct role in the trait, but that due to the interconnectedness of cellular pathways, most of the heritability for a trait comes from genes outside the core pathway that have indirect impacts on the trait. In this case, there is practically no gene that does not somehow contribute to trait variation. In addition to complex diseases, life-history traits could also be considered to have an omnigenic basis. Our findings showed a highly polygenic basis to the relationship between offspring number and body weight, but we were unable to find an enrichment in our candidate genes for a particular biological process. Therefore, we lacked evidence for a core gene network as suggested under an omnigenic model. The omnigenic model takes inspiration from Fisher's "infinitesimal model"[21], where at the extreme limit, a complex trait is made up of infinitesimally small contributions from infinitely many genes. An interesting question is whether all complex traits reflect the biological reality of the extreme of the "infinitesimal model" or if there is a continuum along which traits can be ranked. If it is the latter, then why do life-history traits seem to trend towards the extreme? A possible reason could be that these traits, which are closely tied to fitness, are the products of many biological pathways - for example, there is likely a myriad cellular processes that can be perturbed to affect lifespan or viability, but fewer that affect wing shape or bristle count. Research beyond genetic variants and into the networks of molecular interactions in cells may help in our understanding of why the genetic basis of some traits is so broad.

Future Directions

In his review on the empirical evidence for bet-hedging[22], Simons introduced six evidence categories used to evaluate a potential bet-hedging trait. The last two evidence categories, for which the fewest studies exist, are direct tests to determine the adaptive value of bet-hedging. Our current work has established the existence of a bet-hedging trait and the climate patterns that favor bet-hedging, as well as provided observational evidence that bet-hedging may be advantageous only in certain climates. As a next step in our investigation of thermal preference variability as a bet-hedging trait, we can consider a direct test of fitness under fluctuating or constant environments. A simple starting hypothesis is that a fluctuating environment, modeled after the weather patterns of a predicted bet-hedging favored location, would favor increased thermal preference variability. We can then compete a high variability genotype against a low variability genotype under laboratory-established fluctuating conditions (and, in parallel, under constant conditions) for several generations to see whether the high variability genotype will be overrepresented in fluctuating conditions and the low variability genotype in constant conditions. Higher fitness of one genotype over another would be evaluated using sequencing of offspring across the experiment to see whether there are reproducible increases in haplotypes originating from a particular parental line (high or low variability). Sequencing and haplotype assignment would also allow us to determine genetic regions that control thermal preference variability, as, under our hypothesis, those would be the regions with the strongest allele frequency changes over time. While this direct test of bet-hedging would make a great contribution toward our understanding of the evolution of bet-hedging traits, there are strong considerations that need to be examined. The timescale of the selection experiment should be established using simulations to determine feasibility - perhaps, in order to see appreciable changes in allele frequency, either the time to carry out the experiment or the selection coefficient will have to be prohibitively large. Replicates and controls will have to be maintained in order to control for genetic drift and generally adaptive alleles. In addition, because of the known genotype-by-

environment component in variability, it would have to be established that the high and low variability genotypes still express the same relative phenotypes in the fluctuating and constant environmental conditions. Heterozygosity and recombination in the offspring may influence the expression of variability in unpredictable ways during the experiment due to dominance or epistatic interactions. These caveats underscore the difficulty in carrying out direct tests of bet-hedging and likely explain why there are so few studies that attempt them. Yet, questions about the adaptive value of variability remain - with a rigorous approach and a consideration of limitations, future work on variability and bet-hedging may provide us with the answers.

Another question that deserves further attention is whether there exists a direct link between a specific environmental factor and variability in a related trait. Even though environmental contributions to variability seem multifaceted, could one specific factor e.g. light, temperature, odor, have an outsize influence on variability and/or in a positively correlated fashion? The difficulties in these studies would be choosing the appropriate environmental factor-trait pair and extensively controlling all the other factors. Evidence from bees shows that enriched odor experience heightened inter-individual variation in odor responses[23]. Odor enrichment in fruit flies did not show mean effects on brain anatomy or odor learning[24], but trait variability was not examined. Odor exposure and odor preference would be solid candidates for an experimental study on this question - there is a strong *a priori* link between the two and enriching the environment for odors is more straightforward and may introduce fewer confounds with other factors than, for example, trying to affect locomotion. Further studies could dig down to understand how this specific correspondence occurs on a genetic and molecular level. Understanding exactly how the environment influences variability is necessary for a more complete picture of how traits are shaped by genetics and the environment.

The overall conclusions of my research into behavioral variability and variation in life-history underline the interconnectedness of genetics and the environment, the adaptive value of environmentally-caused variability, and a holistic perspective on the genetics of life-history. Future studies will unravel some of the complex interrelationships proposed here, to the benefit of our understanding of observed biological diversity.

References

1. Morgante, F., Sørensen, P., Sorensen, D. A., Maltecca, C. & Mackay, T. F. C. Genetic Architecture of Micro-Environmental Plasticity in *Drosophila melanogaster*. *Sci. Rep.* **5**, 9785 (2015).

2. Honegger, K. & de Bivort, B. Stochasticity, individuality and behavior. *Curr. Biol.* **28**, R8–R12 (2018).

3. Budelli, G. *et al.* Ionotropic Receptors Specify the Morphogenesis of Phasic Sensors Controlling Rapid Thermal Preference in Drosophila. *Neuron* (2019) doi:10.1016/j.neuron.2018.12.022.

4. Hamada, F. N. *et al.* An internal thermal sensor controlling temperature preference in Drosophila. *Nature* **454**, 217–220 (2008).

5. Hong, S.-T. *et al.* cAMP signalling in mushroom bodies modulates temperature preference behaviour in Drosophila. *Nature* **454**, 771–775 (2008).

6. Bang, S. *et al.* Dopamine signalling in mushroom bodies regulates temperature-preference behaviour in Drosophila. *PLoS Genet.* **7**, e1001346 (2011).

7. Waddington, C. H. Canalization of Development and the Inheritance of Acquired Characters. *Nature* **150**, 563–565 (1942).

8. Meiklejohn, C. D. & Hartl, D. L. A single mode of canalization. *Trends Ecol. Evol.* **17**, 468–473 (2002).

9. Rutherford, S. L. & Lindquist, S. Hsp90 as a capacitor for morphological evolution. *Nature* **396**, 336–342 (1998).

10. Queitsch, C., Sangster, T. A. & Lindquist, S. Hsp90 as a capacitor of phenotypic variation. *Nature* **417**, 618–624 (2002).

11. Sangster, T. A. *et al.* HSP90 affects the expression of genetic variation and developmental stability in quantitative traits. *Proc. Natl. Acad. Sci.* **105**, 2963–2968 (2008).

12. Hall, M. C., Dworkin, I., Ungerer, M. C. & Purugganan, M. Genetics of microenvironmental canalization in Arabidopsis thaliana. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 13717–13722 (2007).

13. Ayroles, J. F. *et al.* Behavioral idiosyncrasy reveals genetic control of phenotypic variability. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 6706–6711 (2015).

14. Dall, S. R. X., Houston, A. I. & McNamara, J. M. The behavioural ecology of personality: consistent individual differences from an adaptive perspective. *Ecol. Lett.* **7**, 734–739 (2004).

15. Werkhoven, Z. *et al.* The structure of behavioral variation within a genotype. *bioRxiv* 779363 (2019) doi:10.1101/779363.

16. Bull, J. J. Evolution of Phenotypic Variance. *Evolution* **41**, 303–315 (1987).

17. Tufto, J. Genetic evolution, plasticity, and bet-hedging as adaptive responses to temporally autocorrelated fluctuating selection: A quantitative genetic model. *Evolution* **69**, 2034–2049 (2015).

18. Svardal, H., Rueffler, C. & Hermisson, J. Comparing environmental and genetic variance as adaptive response to fluctuating selection. *Evol. Int. J. Org. Evol.* **65**, 2492–2513 (2011).

19. Schuett, W. *et al.* Life-history trade-offs mediate 'personality' variation in two colour morphs of the pea aphid, Acyrthosiphon pisum. *J. Anim. Ecol.* **84**, 90–101 (2015).

20. Boyle, E. A., Li, Y. I. & Pritchard, J. K. An expanded view of complex traits: from polygenic to omnigenic. *Cell* **169**, 1177–1186 (2017).

21. Turelli, M. Commentary: Fisher's infinitesimal model: A story for the ages. *Theor. Popul. Biol.* **118**, 46–49 (2017).

22. Simons, A. M. Modes of response to environmental change and the elusive empirical evidence for bet hedging. *Proc. R. Soc. B Biol. Sci.* **278**, 1601–1609 (2011).

23. Jernigan, C. M. *et al.* Experience-dependent tuning of early olfactory processing in the adult brain. *bioRxiv* 558734 (2019) doi:10.1101/558734.

24. Wang, X., Amei, A., Belle, J. S. de & Roberts, S. P. Environmental effects on Drosophila brain development and learning. *J. Exp. Biol.* **221**, jeb169375 (2018).

Appendix: Variability in thermal and phototactic preferences in Drosophila may reflect an adaptive bet-hedging strategy

Jamey S. Kain[1], Sarah Zhang[1], Jamilla Akhund-Zade[2,3], Aravinthan D. T. Samuel[2,4], Mason Klein[2,4,5], Benjamin L. de Bivort[1,2,3]

[1]Rowland Institute at Harvard, Cambridge, Massachusetts 02142, U.S.A.
[2]Center for Brain Science, Harvard University, Cambridge, Massachusetts 02138, U.S.A.
[3]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts 02138, U.S.A.
[4]Department of Physics, Harvard University, Cambridge, Massachusetts 02138, U.S.A.
[5]Department of Physics, University of Miami, Coral Gables, Florida 33124, U.S.A.

We thank Chris Stokes for help with fabricating the slow photobox and we thank Julien Ayroles, Sarah Kocher, Greg Lang and Sean Buchanan for helping analyze the data and model. We thank Associate Editor Tadeusz Kawecki, Björn Brembs, Axel Gorostiza and an anonymous reviewer for comments that substantially improved the manuscript.

All raw data used in this study, as well as all code used for data acquisition, statistical analysis and modeling are available at http://lab.debivort.org/bet-hedging-seasons-and-the-evolution-of-behavioral-diversity-in-Drosophila.

Abstract

Organisms use various strategies to cope with fluctuating environmental conditions. In diversified bet-hedging, a single genotype exhibits phenotypic heterogeneity with the expectation that some individuals will survive transient selective pressures. To date, empirical evidence for bet-hedging is scarce. Here, we observe that individual *Drosophila melanogaster* flies exhibit striking variation in light- and temperature-preference behaviors. With a modeling approach that combines real world weather and climate data to simulate temperature preference-dependent survival and reproduction, we find that a bet-hedging strategy may underlie the observed inter-individual behavioral diversity. Specifically, bet-hedging outcompetes strategies in which individual thermal preferences are heritable. Animals employing bet-hedging refrain from adapting to the coolness of spring with increased warm-seeking that inevitably becomes counterproductive in the hot summer. This strategy is particularly valuable when mean seasonal temperatures are typical, or when there is considerable fluctuation in temperature within the season. The model predicts, and we experimentally verify, that the behaviors of individual flies are not heritable. Finally, we model the effects of historical weather data, climate change, and geographic seasonal variation on the optimal strategies underlying behavioral variation between individuals, characterizing the regimes in which bet-hedging is advantageous.

Introduction

How do organisms thrive in the face of fluctuating environmental conditions? Understanding their strategies is a major challenge in evolutionary ecology. One versatile adaptive "solution" is phenotypic plasticity – in which an individual adjusts its phenotype in direct response to the current environmental condition, such as modulation of leaf size in response to lighting conditions[1]. In principle, plasticity can embody perfect solutions to any environmental challenge, as animals can employ a "lookup table," producing the perfect response to any condition. However, there are limitations to plasticity[2,3], such as the metabolic cost of encoding a lookup table, and the speed with which an organism can change its phenotype. The latter constraint, phenotypic inflexibility, applies particularly to animals, like insects, that attain a final adult life stage. That said, behavioral phenotypes specifically have the potential to be quite flexible.

Populations can also survive changing conditions by having diversified phenotypes as a result of genetic variation; this also allows organisms to readily evolve/adapt to new conditions. This is termed "adaptive tracking." However, if the environmental changes are transient, as one would observe with seasonal variation, it would be detrimental to rapidly adapt to their local/temporal environment (summer adapted animals would not fare well during the winter). Instead, an adaptive response to fluctuating selection can be to suppress the phenotypic expression of genetic variation, reducing heritability[4]. While genetic variation can be maintained under some circumstances, recent evidence suggests that temporal environmental fluctuations may reduce polymorphism through most of the genome more severely than even constant environments[5].

A third possible solution to the problem of uncertainty is to utilize a bet-hedging strategy (also called risk-spreading), in which developmental stochasticity produces a distribution of adult phenotypes. In diversified bet-hedging, a single genotype can (stochastically) generate a distribution of phenotypes, guaranteeing that at least some individuals are well suited to any

environmental condition[6–8]. More formally, bet-hedging can be defined as evolutionary strategies that reduce the variance in fitness (maximizing the geometric mean of fitness, at the expense of the arithmetic mean of fitness) across time and environmental conditions.

Some individuals in bet-hedging populations will have reduced fitness for any given environmental condition. The adaptive value of bet-hedging increases with increased environmental variation[9], provided that the fluctuations are not brief compared to animal lifespans[10]. An elegant example is the timing of seed germination[11]. If all the seeds from a desert plant germinated after the first rain of the season, they would be vulnerable to extinction if there is an extensive drought before the second rain. Conversely, if the seeds all germinate later in the season, they will be at a disadvantage relative to other seeds that had germinated at the first opportunity (in typical seasons without an early drought). Thus, an optimal strategy may be for the plant to hedge its bets and have a fraction of seeds delay germination while the others respond to the first rain. Of course, this is biology, and real organisms surely employ a combination of plasticity, adaptive tracking and bet-hedging[12]. Yet, bet-hedging in animal systems remains poorly studied, in part because of the difficulties of studying intra-genotypic variability within a common environment, let alone in more complex and biologically realistic scenarios.

The evolutionary optimality of bet-hedging can explain why a single genotype gives rise to a distribution of phenotypes[13]. This question has also been addressed within behavioral ecology from the perspective of animal personality. Genetic variants are often assumed to underlie the behavioral differences described as personality variants, and indeed animal personality syndromes may be largely heritable (up to 52% of variance)[14]. However, to explain the remaining variance in individual behavior, stochastic mechanisms generating intra-genotypic variability are almost certainly at play, including bet-hedging. Thus, in explaining variation in the

personality of individual animals, it is essential to assess the degree to which bet-hedging is itself under genetic control.

While animal personality is typically evaluated along axes that correspond to dimensions of variation in human personality, such as shyness vs. boldness, behavioral variation is richly multidimensional[15]. We assert that if there is 1) variation in a behavior among closely related individuals, and 2) these idiosyncratic differences persist within the lifetime of those individuals, this is an example of a facet of animal personality, broadly construed. As an example, fruit flies exhibit life-long locomotor biases (preferring to turn left or right on an individual-by-individual basis[15,16]). This variation has no clear relationship with the bold-shy axis but represents one orthogonal axis of "personality" among many.

Fruit flies are one of the most studied organisms for many aspects of biology, including the basis of behavioral diversity. We chose to study bet-hedging using the fly's positional response to thermal gradients (thermotaxis) and asymmetric illumination (phototaxis). The thermotactic and phototactic responses of *Drosophila* depend on a wide range of environmental and stimulus parameters[17], such as humidity[18], directionality of the light source[19], and agitation state of the flies[19–21]. The type of phototactic response is particularly sensitive to the state of agitation. In most *Drosophila* species, agitated animals exhibit "fast phototaxis" toward the light source, while unagitated animals exhibit "slow phototaxis" as a preference to stay in shaded areas. The former response is thought to reflect a predator evasion instinct to move skyward[22], while the latter reflects a thermoregulatory and anti-desiccation instinct during rest[23].

Thermal experience has dramatic effects on the life history of *Drosophila*[24–26]. Individuals can control this experience through a variety of behaviors[27,28] including shade-seeking phototaxis and direct positional response to thermal gradients. Thus, the net resting behavior of flies will

greatly affect the amount of heat they experience across their lifetime, and consequently their vulnerability to unusual weather, season and climate fluctuations. The light versus shade and thermal gradient resting preferences of animals can be readily quantified in laboratory experiments.

Recent results from several groups hint that fluctuating temperature specifically could favor bet-hedging. The optimal preferred temperature of ectotherms may not be the single temperature that yields the fastest growth, if the fitness function on temperature is skewed[29]. Selection for heat-resistance indirectly increased cold resistance[30], suggesting that evolutionary solutions to extreme temperatures may act on the absolute deviation from mean temperatures as much as the direction of deviation. Moreover, populations evolved specifically in fluctuating environments acquired thermal resistance to temperatures outside the selected temperature range, even when the fluctuating temperatures were moderate, and centered on the animals' preferred temperature[30,31].

We found considerable variation in the slow phototactic and thermotactic responses of very recently domesticated *Drosophila* from Cambridge, Massachusetts, U.S.A. . Some individual flies strongly preferred to rest in the shaded portion of the phototactic arena (or the cool portion of the thermotactic arena), others strongly preferred the lit portion (or the warm portion). We wondered whether this behavioral diversity represented a bet-hedging strategy to maximize fitness in the face of fluctuating seasonal or weather conditions. In order to compare the performance of bet-hedging versus a strategy in which the individual behavioral preferences are heritable (i.e. adaptive tracking *sensu* Simons[7]) we developed a model incorporating our behavioral data with local weather and climate data from historical records. Phenotypic plasticity in response to environmental fluctuations is unlikely to explain the behavioral differences we observed between individuals reared in essentially identical lab environments; under phenotypic

plasticity, we would expect animals to adopt similar behaviors as their response to a similar environment, but this is not what we observe. Our scope here is to specifically consider a head-to-head comparison of bet-hedging and adaptive tracking strategies, both of which remain plausible explanations of the observed behavioral variation. Thus, we test the hypothesis that the observed individual behavioral differences reflect a bet-hedging strategy, rather than genetic variation underlying an adaptive-tracking strategy.

We find that the bet-hedging strategy generally outcompetes adaptive tracking. Since the generation time of *Drosophila* is short relative to the seasons, seasonal temperature fluctuations can induce genetic adaptations in the spring[32] which could then decrease fitness in the summer. This reversal of selective pressures throughout the year renders adaptive tracking counterproductive. The alternative bet-hedging strategy is particularly valuable when there is high fluctuation in temperature throughout the season. Adaptive tracking is advantageous, however, during seasons that are consistently warm or cold throughout, because it engenders long-term changes to average behaviors by altering genotypic frequencies. Interestingly, since global climate change will bring about an increase in mean temperatures, we predict that the optimal strategy will change in approximately 100 years, and adaptive tracking will become more advantageous than bet-hedging.

Results

Drosophila exhibit more behavioral variability than expected by chance alone

We sought to directly measure the slow phototactic and thermotactic response of recently domesticated *D. melanogaster* flies and assess to what extent there was individual-to-individual variability in this behavior. An isofemale line ("CamA") was established from a single fertilized female caught in Cambridge, Massachusetts, U.S.A. . To assess phototaxis, age- and sex-matched CamA adults, cultured on standard fly media, were assayed individually in our "slow photobox" (Figure 5.1A), where their light versus shade preference was measured by automated image analysis 24 times per fly (Figure 5.1B), once every 10 minutes. We tested 219 individuals in total, and found that their average light-choice probability was 0.32 with a standard error of 0.032, indicating a preference for resting in the shade. The observed distribution of light-choice probabilities was considerably overdispersed compared to the null hypothesis that all animals were choosing the light with identical probabilities of 0.32 (p = 4E-6, 1E-11 and < 0.001 by Kolmogorov-Smirnov (KS) test, χ^2 test of variance and bootstrap resampling respectively; Figure 5.1C), indicating considerable individual-to-individual behavioral variability. We estimated 44.2% of the experimental variance was due to individual differences, corresponding to a preference index standard deviation across individuals of 0.085 (95% CI = [0.74, 0.94], estimated by bootstrap resampling). These results are similar to our previous findings on agitated phototaxis where we observed significant individual-to-individual variability that was not explainable by differences in age, sex, reproductive status, birth order, social interactions, or previous exposure to light[33].

To assess thermotaxis, similarly cultured animals were tested individually on a linear thermal gradient ranging from 30°C to 18°C (Figure 5.1D), which spans most of the range of flies' natural environment. The position of each of 41 flies within this gradient was measured 20 times per animal, once every 10 minutes, with their position indicating their per-trial thermotactic

preference (Figure 5.1E). The mean average preference was 23.1°C with a standard error of 0.22°C. We observed considerable inter-individual variation in mean thermotactic preferences ($F = 3.07$, d.f. = 40, $p < 1E\text{-}6$ by 1-way ANOVA on fly identities; Figure 5.1F). We estimated 14.7% of the experimental variance was due to individual differences, corresponding to a standard deviation across individuals of 1.4°C (95% CI = [1.15, 1.77], estimated by bootstrap resampling).

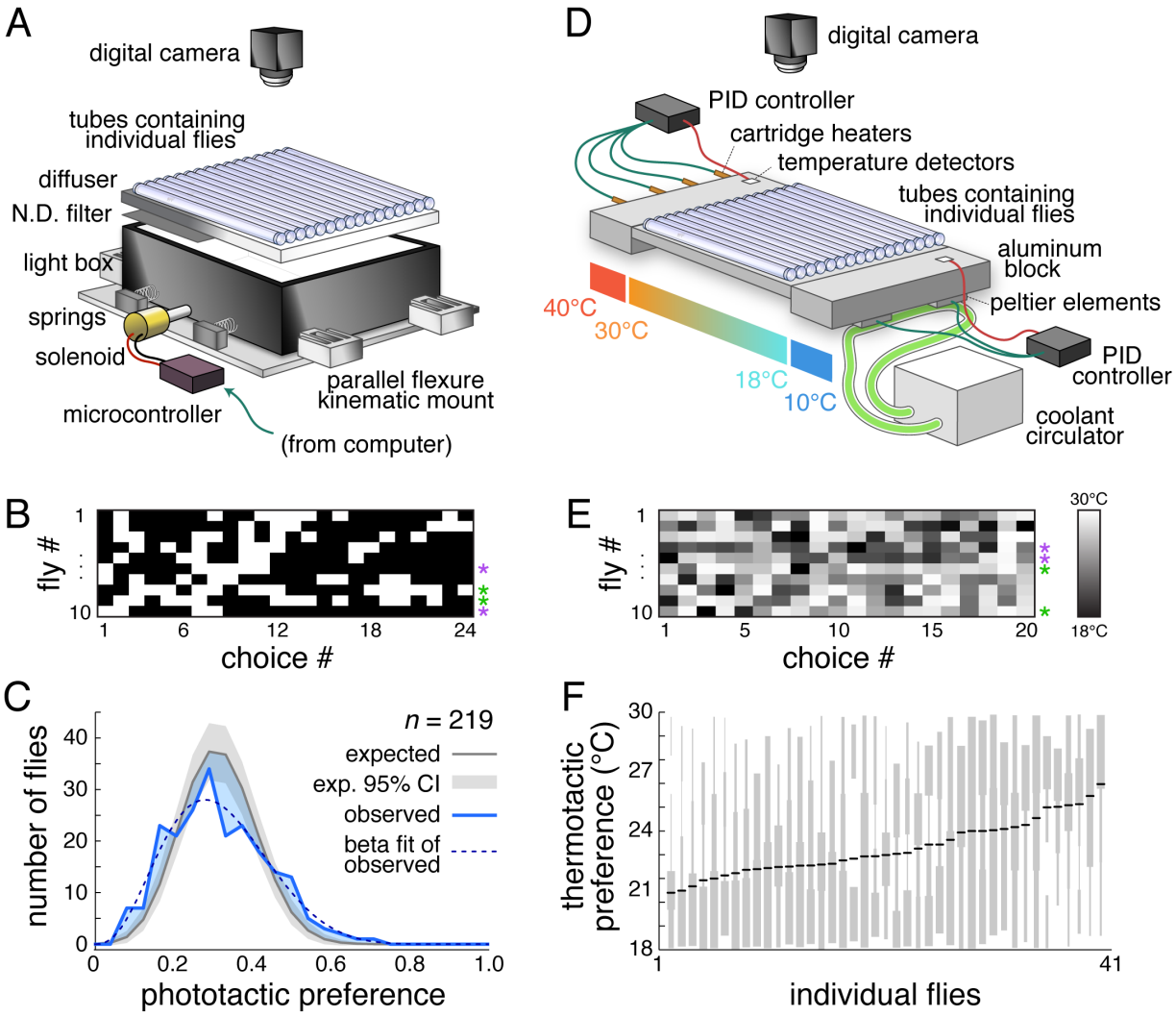


Figure 5.1 - Measurement of phototactic and thermotactic variation and a model of their effect on fitness.

(A) Schematic of the “slow photobox” – a device for the high-throughput characterization of slow phototaxis. Animals were placed individually into clear tubes with a lit and shady side. Their position in the tube was recorded by a camera. (B) Example of data from the slow photobox. Each row represents an individual fly’s phototactic preferences at 24 instances, spaced at 10 minute intervals. White boxes indicate lit choice and black boxes indicate shaded choice. Purple and green asterisks indicate examples of shade- and light-preferring individuals, respectively. (C) Observed histogram of the phototactic preference across individual flies (blue line). Dashed gray line indicates a best-fit beta-binomial distribution for the observed data. Gray line indicates expected distribution for the same flies if they were each to choose light with identical probabilities. Gray shaded region indicates 95% confidence interval of the expected distribution given sampling error. Shaded blue areas indicate discrepancies between the observed and expected histograms consistent with behavioral heterogeneity. (D) Schematic of the “slow thermobox.” (E) Example data from the slow thermobox, as in B. Grayscale indicates thermotactic preference over time. Purple and green asterisks indicate examples of cool- and warm-preferring individuals, respectively. (F) Histograms of thermotactic preference values across all trials (vertical, grey) for individual flies, sorted by mean preference (black bars). Black bars indicate individual mean preferences.

We performed day-to-day persistence experiments to see if the individual differences in thermotaxis and phototaxis were stable across time, rather than arising from transient state differences such as satiety. Individual scores for both phototactic and thermotactic preference were significantly correlated across 24 hours inter-test intervals (Figure 5.2; $r = 0.71$, $p < 0.0001$, d.f. = 70 and $r = 0.48$, $p = 0.002$, d.f. = 36 respectively).

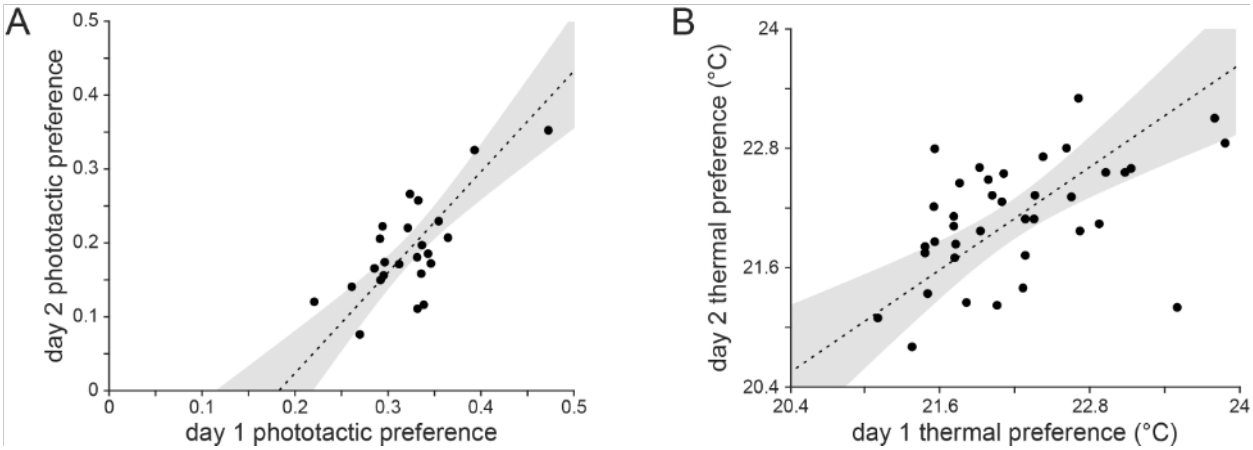


Figure 5.2 - Persistence of individual behavioral phenotypes.
(A) Scatter plot of phototactic preference on day 2 versus phototactic preference on day 1. Flies were assayed for 40 hours continuously. Their position with respect to a phototactic gradient was recorded every 10 minutes. Daily preference values were calculated as the average of all observations in the first 24 hours (day 1) and all the points in the last 16 hours of the recording (day 2). Points are individual flies; dotted line is the best geometric mean regression fit; shaded region is the 95% confidence interval on the fit as determined by bootstrap resampling. Pearson $r = 0.71$, $n = 24$, $p < 0.0001$. (B) As in (A) for thermal preference. Rather than continuous recording, in the thermal persistence experiment, thermal preference was measured as in all other thermal experiments, over four hours, and fly identity was maintained across successive days by individual housing. Regression fits as in A. Pearson $r = 0.48$, $n = 37$, $p = 0.002$. Genotypes of the inbred flies tested here are w^{1118} and DGRP line #796 respectively.

A model to compare adaptive tracking and bet-hedging strategies

Could the observed behavioral individuality represent a bet-hedging strategy to increase the probability that at least some individuals will be well adapted to the current weather conditions? To test this, we proposed a model of fly development and reproduction (Figure 5.3A) in which an individual animal's behavior could be treated either as perfectly inherited from the mother (i.e. adaptive tracking – AT), or as non-heritable/stochastic variation indicative of a bet-hedging

strategy (BH). Holding the magnitude of variation constant, we can evaluate which is more advantageous, adaptive tracking or bet-hedging, and under what conditions.

In considering how thermal experience might affect fitness, we recognized that the metamorphosis time from egg to adulthood depends on the temperature experienced during that period, in a relationship determined by previous experimental work[24,25], with flies developing fastest at 25°C (Figure 5.3B). The expected total lifespan of flies also depends on temperature[26], with flies living considerably longer at cooler temperatures (Figure 5.3C). We assume that the effective temperature experienced throughout adulthood depends on the integrated results of many behavioral choices for each individual fly. By contrast, we assume the temperature experienced during growth from egg through pupa depends on the thermal preference index of each fly's mother (the alternative, that developmental temperature depends on progeny preference, yields qualitatively identical results). These are clearly simplifying assumptions – the total amount of thermal energy integrated across a lifespan and the choice of oviposition site depend on more behaviors than just phototaxis and thermotaxis. But, constraining the model with empirical data on these behaviors allows us to investigate their roles in fitness. We lastly assume that throughout metamorphosis and adulthood flies face a constant risk of death (by e.g. predation, disease, fly swatter, etc.), and after reaching adulthood, flies produce new offspring at a constant rate. Thus, temperature choices represent a tradeoff for the fly: warm-preferring animals will have the benefit of faster development at the cost of shorter lifespan, the kinetics of which are temperature-dependent.

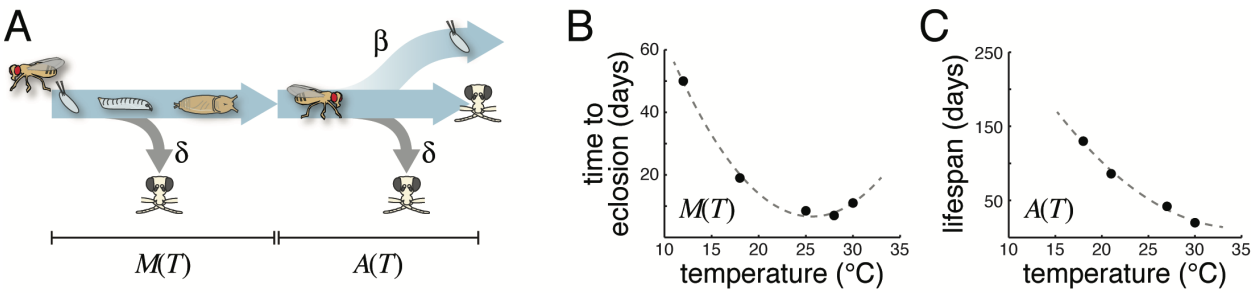


Figure 5.3 - A fly temperature-dependent life history model.
(A) Diagram of the fly life history model; see description in text. β: birth rate, δ: death rate, M: metamorphosis time, A: adult lifespan, T: thermal preference index. "fly skull and crossbones" icons indicate death. (B) Time to eclosion plotted as a function of temperature, as used by the model. Data points from (Ashburner, 1978). (C) Lifespan plotted as a function of temperature, as used by the model. Data points from (Miquel et al., 1976).

In order to formulate a single variable representing the diversity of temperature experience due to all dimensions of behavioral variability, we compared our phototactic and thermotactic observations. The effect of phototactic preference on temperature experience depends on the temperature difference between shade and sunlight. This in turn depends on numerous factors, including weather conditions, latitude, season, wind, substrate composition and duration of exposure to the sun. We measured this directly and determined that a 7°C difference between sun and shade was attained quickly after exposure to sunlight on both natural and artificial substrates in Cambridge, Massachusetts, U.S.A. This estimate is well within the range of previous estimates of the temperature difference between insects in sunlight vs shade[27]. We observed that the mean light-choice probability of flies in the slow phototaxis assay was 0.32, with a standard deviation of 0.13 (Figure 5.1C). Assuming that a fly which spends *x*% of the time in the light would spend *x*% of the time in the sun and 7°C warmer than the remaining 100-*x*% of the time, this phototactic variance implies a standard deviation of temperature experience of 0.89°C. The mean observed thermotactic preference was 23.1°C with a standard deviation of 1.4°C. These two observations are in agreement that individual flies have substantially different temperature experiences. For the model, we let the "thermal preference index" of individual flies, integrated across all behaviors determining temperature experience, vary from 0 to 1,

corresponding to a 7°C temperature range. This index has the same range as the phototactic data. Allowing it to follow the same distribution across flies as the phototactic data (beta-distributed, with mean 0.32 and variance 0.22), it reflects a conservative estimate of thermal experience variability, compared to the direct thermotactic measurements.

The model contains two unknown parameters, the lifelong risk of death from causes other than thermal experience-dependent mortality (δ), and the birth rate at which new eggs are laid by sexually mature flies in the wild (β). We have no empirical data from which to assert these values, but the behavior of the model constrains them under two reasonable assumptions – 1) that the population size of flies is the same at the end of each season as the beginning, and 2) that the mean thermal preference index of the population is the same at the end of the season as the beginning, i.e. they are adapted to average conditions. These assumptions constrain the random death probability of flies in the wild to 0.013-0.044/day, and the birth probability to 0.037-0.11/mother/day (depending on which weather model is used; Table 5.1. See Methods for details); both of these ranges seem plausible.

Table 5.1 - Weather variables simulated in each implementation of the model, and associated values of the fit birth and death rate parameters.

Figure Panel(s)	Model Type	Variables Simulated	β	δ
2A-E	stochastic	daily temperature normals	0.1062	0.0435
2D	stochastic	constant seasonal temperature	0.0366	0.0203
4A-B	difference equation	daily temperature normals	0.04480243	0.012755
5A-C	difference equation	daily temperature normals historical daily temperature deviations historical daily cloud cover fractions	0.05388375	0.013635
5D-E	difference equation	daily temperature normals simulated daily temperature deviations simulated daily cloud cover fractions	0.05388375	0.013635
6A	difference equation	daily temperature normals simulated daily temperature deviations simulated daily cloud cover fractions 100 sequential seasons	0.04661	0.01168
6B	difference equation	daily temperature normals	0.04480243	0.012755
6C	difference equation	daily temperature normals from 1469 different locations	varies	varies

Bet-hedging outperforms adaptive tracking

We simulated a stochastic (agent-based) implementation of this model, tracking 100 individual flies experiencing the average seasonal temperature fluctuations[34] of a typical fly breeding season in Boston Massachusetts, U.S.A., lasting approximately from April 1 to October 31 (Figure 5.4). We implemented two versions of the model. 1) For the adaptive tracking strategy (AT; Figure 5.4A), the thermal preference index of new flies equaled that of their mother. 2) For the bet-hedging strategy (BH; Figure 5.4B), the preference of each new fly was drawn at random from a beta distribution fitting the observed behaviors (mean thermal preference index =

0.32 and standard deviation 0.13; Figure 5.1C). The initial population of all simulations also followed this distribution, irrespective of strategy.

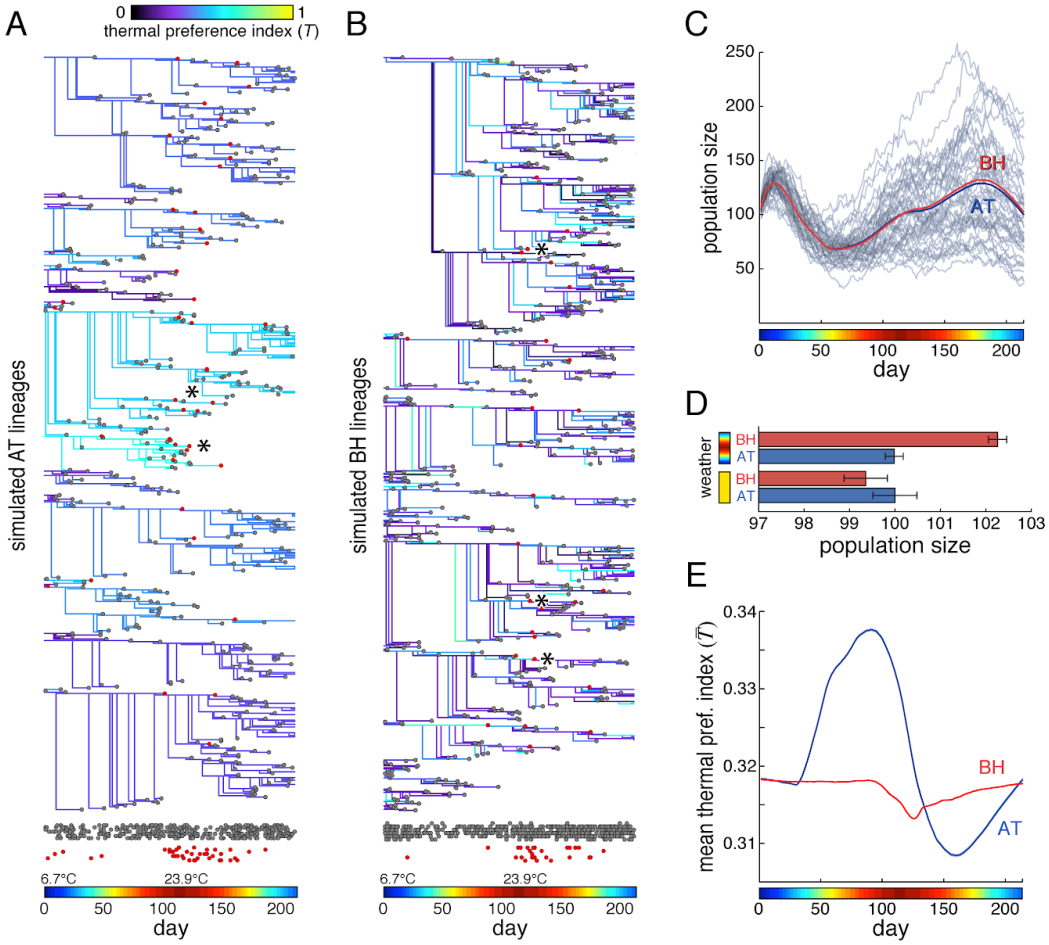


Figure 5.4 - Performance of the bet-hedging and adaptive tracking versions of the stochastic model.
(A) Subset of simulated lineages from one run of the model, under the AT strategy. Branch points indicate the birth of new flies; colors indicate thermal preference index; gray dots indicate death events for reasons other than thermal experience-dependent mortality (due to parameter δ); red dots indicate death events due to thermal experience-dependent mortality. Rows of dots at bottom are projected from above for comparison, with random y-scatter for visibility. Asterisks indicate thermal experience-dependent death events associated with high summer temperatures in light-preferring lineages. Temperature at each day is indicated by the colored bar here and in all other panels. (B) As in (A), but for the BH strategy. (C) The mean performance of a bet-hedging (BH) (red line) and adaptive tracking (AT) (blue) version of the model over time. Gray lines represent a sampling of 100 individual simulated seasons. (D) Mean final population size produced by each version of the model for either constant average weather (yellow) or seasonal weather (colored bar). Error bars are +/-1 one standard error of the mean; n = 40,000 simulations per group. (E) Mean thermal preference index of the population over time for each version of the model. Shaded regions (barely wider than plot lines) are +/-1 one standard error of the mean.

We measured fitness by calculating the population size at the end of the breeding season compared to the beginning. On average, the BH strategy outperformed the AT strategy by just over 2% (Figure 5.4C, D, $p < 0.0001$ by *t*-test), an effect that is completely absent (and non-significantly reversed, $p = 0.64$ by *t*-test) in simulations of constant seasonal temperatures. The reason for the greater population growth of flies using BH is evident in an inspection of the average thermal preference index of the fly population across the breeding season (Figure 5.4E). (The average preference changes even under BH due to temperature-dependent shortening of the lifespan of warm-seeking individuals). In the AT strategy, the cool spring selects for warm-preferring flies because their progeny will develop to maturity more quickly. However, at the onset of summer, the selection is reversed in favor of cool-preferring flies, which have a longer overall lifespan. Once the direction of selection switches, the BH strategy begins to outperform the AT strategy, because AT responds to even transient selective pressures by shifting the population mean.

Individual phototactic preference is not heritable

The model establishes that bet-hedging is a plausible explanation for the behavioral diversity seen experimentally in thermotactic and phototactic preference. However, if the observed individuality we see truly represents bet-hedging, then the differences in preference between individual flies are probably not due to genetic polymorphisms or trans-generational epigenetic effects, which would be heritable. This hypothesis generates two predictions: 1) reducing genetic diversity by inbreeding a polygenic stock should have no effect on the breadth of its behavioral distribution, and 2) the progeny of light- (or shade-) preferring parents should exhibit the same distribution of behaviors as the entire parental generation, not their specific parents. (These predictions were tested in the phototactic paradigm because of its higher throughput and our use of its parameter values in the model). We compared the behavioral distribution of our polygenic isofemale CamA line with that of the line “inbred-CamA”, which was inbred by sibling

matings for 10 generations. Inbreeding had no significant effect on the mean or variance of the behavioral distribution (Figure 5.5A). Using inbred-CamA we set up multiple crosses comprising a male and a virgin female that both either prefer the shade or the light (Figure 5.5B, C). If their individual photopreferences are due to genetic polymorphisms between flies, then their progeny should have a correspondingly shifted mean photopreference relative to the original population. However, we found there was no difference in the mean photopreferences of broods derived from shade-preferring parents versus light-preferring (Figure 5.5B-E). Using Fisher's selection estimator of heritability (h^2 = R/S), we estimated h^2 = -0.026, with a standard error of 0.048. Thus, heritable polymorphisms determine at most a small component of each individual's behavior, consistent with a bet-hedging strategy. Moreover, the distributions of brood photopreferences were indistinguishable from the parental distribution, in variance as well as mean (Figure 5.5D,E).

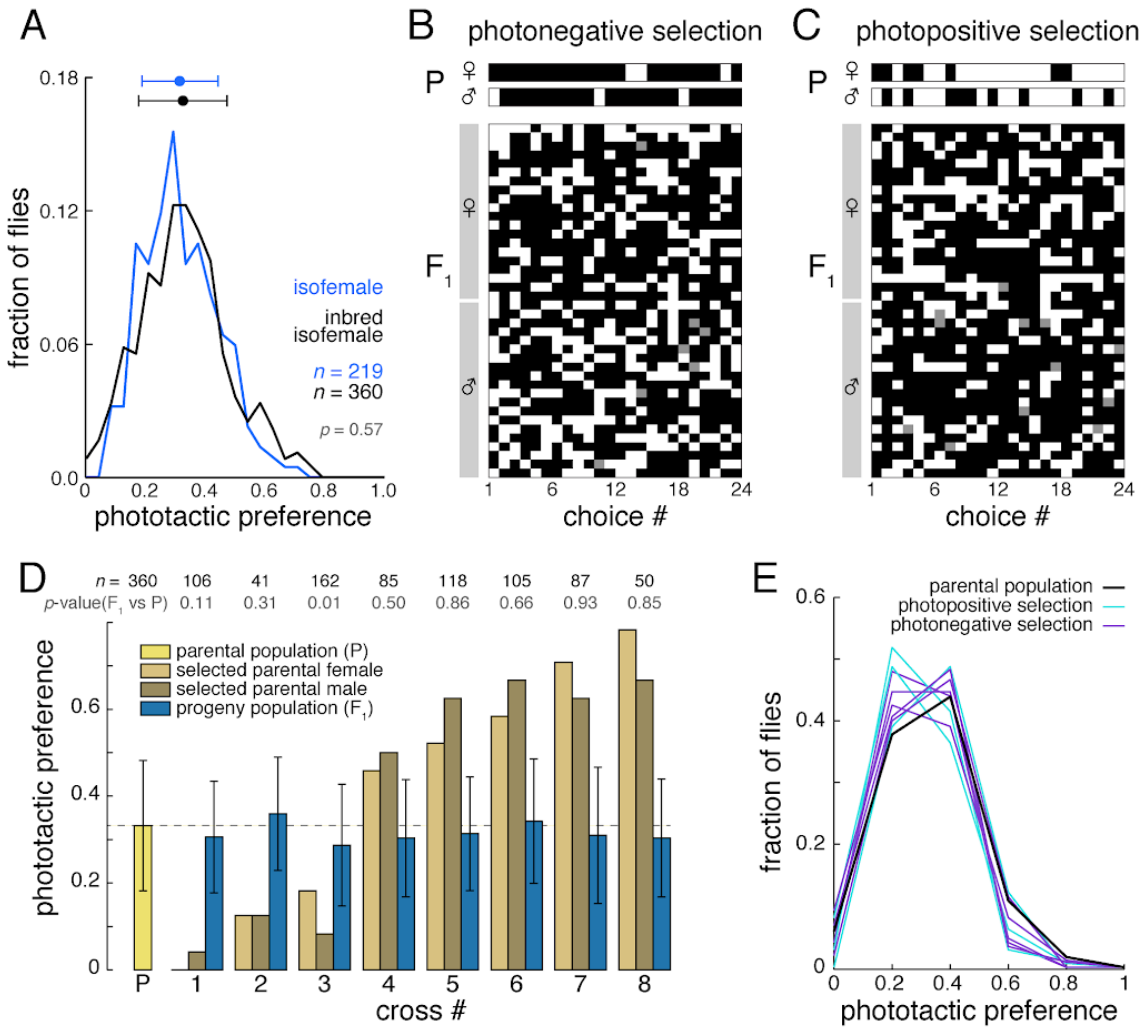


Figure 5.5 - Individual phototactic preference is not heritable.
(A) Observed histogram of the phototactic preference across individual CamA flies (blue) and inbred-CamA flies (black). Points and bars represent the distribution mean and +/- 1 standard deviation. *p*-values from two-sample Kolmogorov-Smirnov test comparing each progeny distribution to the parental distribution. (B) Representative samples of the phototactic scores of a shade-preferring male and female (top) and the phototactic scores of their resulting progeny (bottom). Each row represents an individual fly's phototactic preference over time. White boxes indicate lit choice, black boxes indicate shaded choice, and gray boxes a missing value. (C) As in (B), but for light-preferring parents and their progeny. (D) Phototactic indices for strongly biased shade- or light-preferring individuals (tan and brown bars) and their resultant progeny (dark blue bars). The dashed line and yellow bar indicate the original pool of animals from which strongly biased individual parents were selected. Numbers above bars indicate sample size, with *p*-values from KS test uncorrected for multiple comparisons. Error bars are +/-1 one standard deviation. (E) Histograms of phototactic preferences within the respective progeny.

Deterministic model shows that the bet-hedging advantage is population size invariant

The heritability intrinsic to the AT strategy means that in a finite population simulation (such as in our model population of 100 virtual flies; Figure 5.4) the mean thermal preference index of the population can vary significantly from replicate to replicate due to the stochastic nature of the model (Figure 5.4C). AT may lock in maladaptive thermal preference indices due to drift, and the rate at which this happens depends critically on the simulated population size[35]. Since it was arbitrary to simulate 100 animals, and effective population sizes in the wild are unknown (and perhaps far too large to simulate efficiently[36]), we developed a difference equation version of the model, in which the population size was effectively infinite and immune to stochastic effects. In this implementation, sub-populations of flies with specific thermal preference indices were determined by a set of difference equations (see Methods). The difference equation versions of the BH and AT strategies performed similarly to the simulations of individuals (Figure 5.6A, B), with BH outperforming AT by 1.1% by the end of the summer, and the AT model undergoing two selective sweeps of opposite direction.

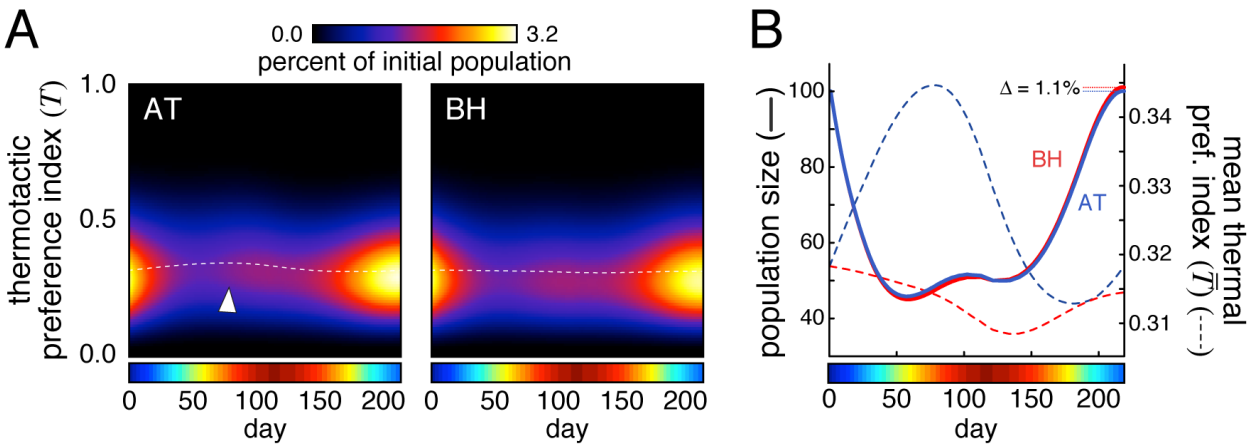


Figure 5.6 - Performance of BH and AT using a difference equation implementation of the model. (A) Abundance of flies as a function of thermal preference index and time for AT and BH strategies under the difference equation model. Arrowhead indicates adaptive thermal positivity during the spring. Dashed white line indicates the mean thermotactic preference. (B) Population size (solid lines) and mean thermal preference index (dashed lines) over time of BH (red) and AT (blue) versions of the difference equation model.

Using this variant of the model, we confirmed that populations utilizing either a BH or AT strategy performed best with intermediate levels of variability (Figure 5.7; for these analyses, we relaxed the constraint of matching simulated variability to experimentally observed variability). Performance diminished when variability was too low or too high, supporting the hypothesis that the observed thermotactic and phototactic preference variability is adaptive. The qualitative results of this model are robust to most assumptions, but sensitive, as expected, to seasonal weather conditions and the range of temperatures accessible by behavioral choices (Table 5.2). The model is qualitatively sensitive to the mean thermal preference index value, which is not surprising, since altering this value means flies are mismatched to their life history tradeoff optimum. An AT strategy allows them to adaptively counter this mismatch.

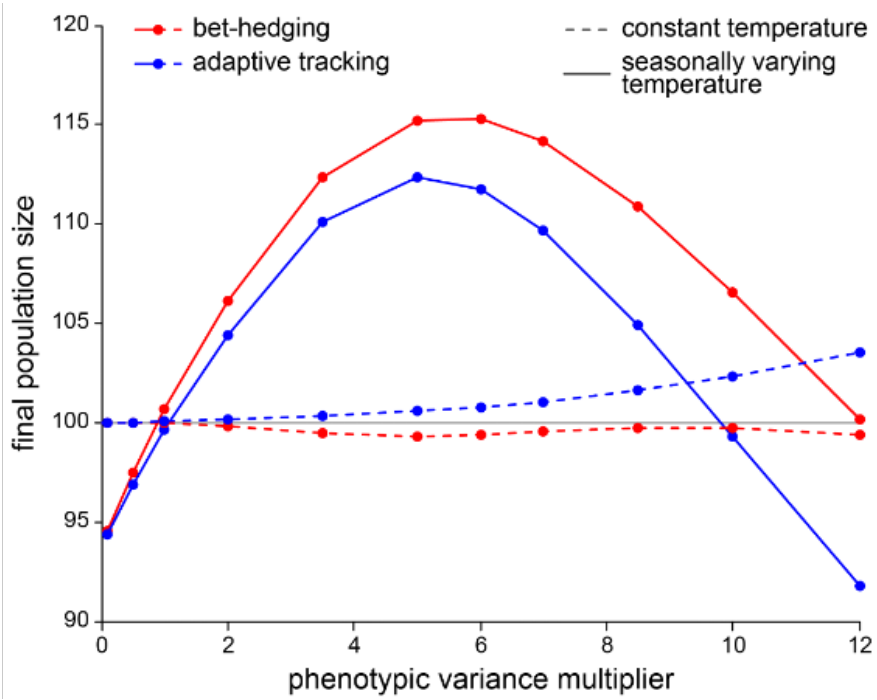


Figure 5.7 - End-of-season population size, as a function of phenotypic variance.
Solid lines reflect performance under seasonally varying temperatures (following the mean daily deviations for Boston, MA). Dotted lines reflect performance in constant temperature seasons. Red indicates bet-hedging populations, blue adaptive tracking. In all conditions the initial phenotypic distribution followed a beta distribution with mean equal to the experimental mean thermal preference, and variance equal to the experimental variance multiplied by the multiplier indicated on the x-axis.

Table 5.2 - Assessment of model qualitative robustness to various parameters and assumptions. * indicates robustness in the relative performance of bet-hedging and adaptive tracking, but unrealistic output otherwise, such as rapidly expanding populations when $\boldsymbol{\beta}$ is increased on its own. Bold figures indicate conditions with BH out performing AT at the end of the season. Numbers in parentheses indicate default values. All parameters or assumptions were changed singly, while holding all others at their default values.

Parameter or Assumption	Robust or Sensitive	Evidence
thermal preference mean (0.3183)	sensitive	TPM = 0.16 … BH vs AT = -47% TPM = 0.26 … BH vs AT = -11% TPM = 0.38 … BH vs AT = -1.7%
thermal preference variance (0.0162)	robust	TPV = 0.012 … BH vs AT = **0.82%** TPV = 0.032 … BH vs AT = **1.7%**
birth rate parameter ($\boldsymbol{\beta}$) (0.04480243)	robust*	$\boldsymbol{\beta}$ = 0.040 … BH vs AT = **1.2%** $\boldsymbol{\beta}$ = 0.060 … BH vs AT = **0.05%**
death rate parameter ($\boldsymbol{\delta}$) (0.012755)	robust*	$\boldsymbol{\delta}$ = 0.008 … BH vs AT = **0.62%** $\boldsymbol{\delta}$ = 0.016 … BH vs AT = **0.047%**
shade temperature difference (7°C)	robust	STD = 4°C … BH vs AT = **0.27%** STD = 10°C … BH vs AT = **1.5%**
time to eclosion vs. temperature = $M(T)$	robust	$M(T) = 0.23T^2 – 11.8T + 168$ (+10d offset) … BH vs AT = **1.5%** $M(T) = 90 – 3T$ (linear, decreasing) … BH vs AT = **0.24%** $M(T) = 100000T^{-3}$ (asymptotically decreasing) … BH vs AT = **1.5%**
adult lifespan vs. temperature = $A(T)$	robust	$A(T) = 0.41T^2 – 28.4T + 606$ (+100d offset) … BH vs AT = **0.035%** $A(T) = 130 – 3T$ (linear, decreasing) … BH vs AT = **0.17%** $A(T) = 120 – 0.12T^2$ (concave-down, decreasing) … BH vs AT = **0.22%**
immature fly thermal experience determined by parental preference	robust	Stochastic simulation run with immature fly thermal experience determined by their own preference. Results statistically indistinguishable. Not applicable to difference equation model.
flies overwinter as adults	robust	Stochastic simulation run with initial fly ages either uniformly distributed > maturity age, or 0. Results statistically indistinguishable. Not applicable to difference equation model.
seasonal weather conditions	varies	Figure 5.4D, Figure 5.8, Figure 5.9
length of breeding season	robust	Figure 5.8F
inclusion of plastic behavioral responses	robust	Figure 5.7

Incorporating historical weather data into the model

To test the effects of daily temperature fluctuations and cloud cover, we ran the difference equation model against historic weather data collected in Boston, Massachusetts, U.S.A. (U.S. National Oceanic and Atmospheric Administration [NOAA][34]; Figure 5.8A). The temperature in each day of the simulation was taken from actual historical data from that day, on a year-by-year basis. Cloud cover was implemented by assuming that the temperature difference available for flies to respond to (i.e. between sun and shade) each day was proportional to the average cloud cover of that day. Not surprisingly, reducing the temperature difference available to flies (due to cloud cover) reduced the magnitude of the advantage of the BH strategy (to around 0.2% for years 2007-2010) (Figure 5.8B). We initially thought that short-term heat waves (or cold spells) might be enough to confer an advantage to bet-hedging, but no clear conclusions about the impact of short-term fluctuations could be drawn from this historical data. However, it was clear that some years were more conducive to bet-hedging than others. For example, in 2010 the BH advantage was comparatively low (Figure 5.8B). The weather that year was consistently warmer than in the others, particularly in the spring and fall (Figure 5.8A, C), exerting a comparatively uniform selective pressure for cool-seeking, thereby reducing the advantage of bet-hedging. Consequently, the AT population exhibited a more consistent trend of decreasing mean thermal preference index across the entire year (Figure 5.8C), although overall BH still outperformed AT.

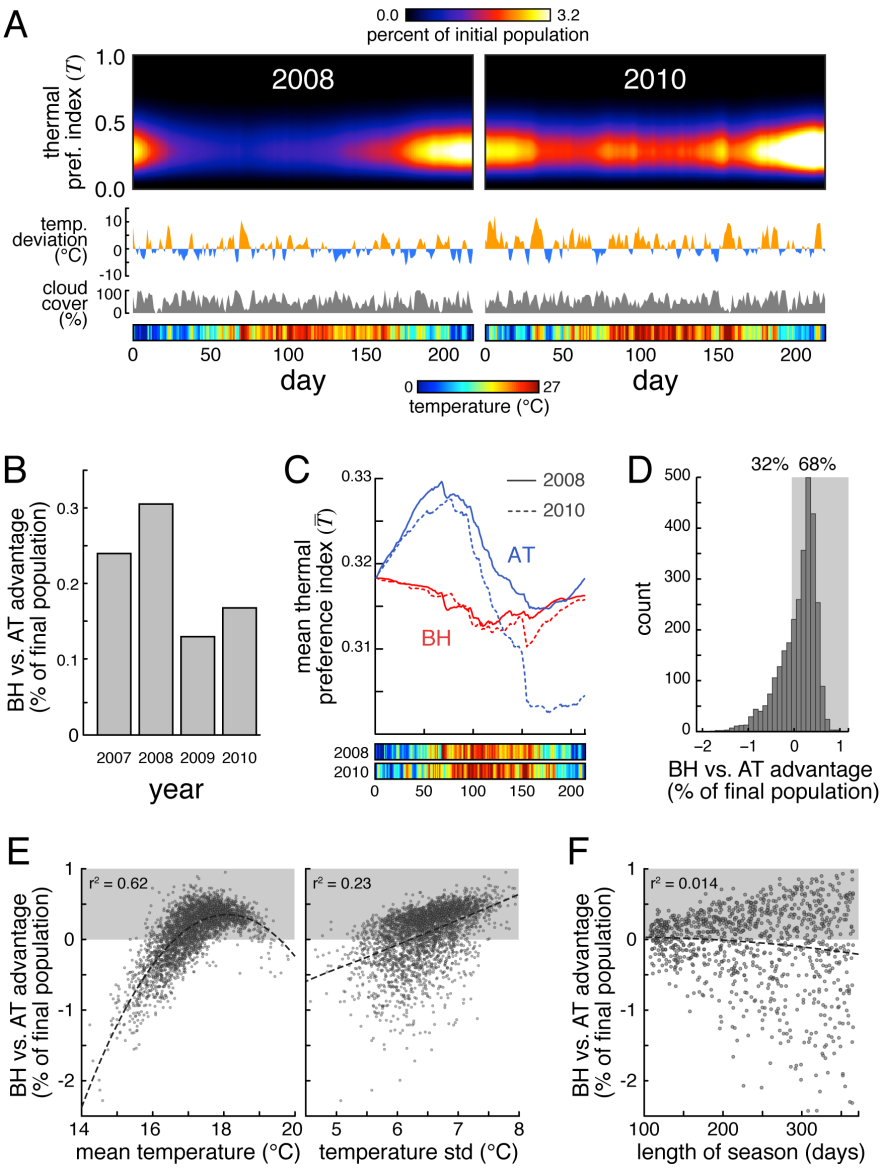


Figure 5.8 - BH vs AT in historical and modeled breeding seasons.
(A) Abundance of flies as a function of thermal preference index and time in the BH version of the difference equation model, applied to historical weather data (temperatures and cloud cover) from 2008 and 2010. Orange and blue traces indicate temperature deviation from daily normals. Gray traces indicate daily cloud cover percentage. Colored bars indicate the daily mean temperature. (B) BH versus AT advantage as a percent of the final population vs. year ((pop_{BH}-pop_{AT})/pop_{AT}*100). (C) Mean thermal preference indices for AT (blue) and BH (red) versions of the real weather difference equation model for 2008 (solid lines) and 2010 (dashed lines). Colored bar as in (A). (D) Histogram of BH versus AT advantage as a percent of final population using the difference equation model across 3000 simulated seasons. Shaded region indicates the simulations in which BH outperformed AT. (E) Scatterplot of BH versus AT advantage versus mean temperature (left panel) or the standard deviation of the temperature (right panel), across 3,000 simulated seasons. Shaded region indicates the simulations in which BH outperformed AT. r^2 values reflect quadratic fits (dashed lines). (F) Scatterplot of BH versus AT advantage versus breeding season length, across 1000 simulated weather seasons. Shaded region and r^2 value as in E.

Mean temperature and temperature range are most predictive of the BH vs AT advantage

We developed statistical models of the daily temperature fluctuations and cloud cover that allowed us to simulate realistic random breeding seasons, and systemically tested the factors favoring the BH and AT strategies. Across 3000 random seasons, BH outperformed AT 68% of the time (Figure 5.8D). We examined numerous metrics describing the simulated seasons (Figure 5.9) and found two in particular that were predictive of the magnitude of the BH vs AT advantage (Figure 5.8E): the temperature mean and standard deviation. BH outperformed AT when the season has a typical temperature while exceptionally hot or cold seasons favored the AT strategy. Additionally, AT performs poorly during “high variance” seasons (those with cold springs and falls, and hot summers) because it engenders large, lagged fluctuations in genotype frequency.

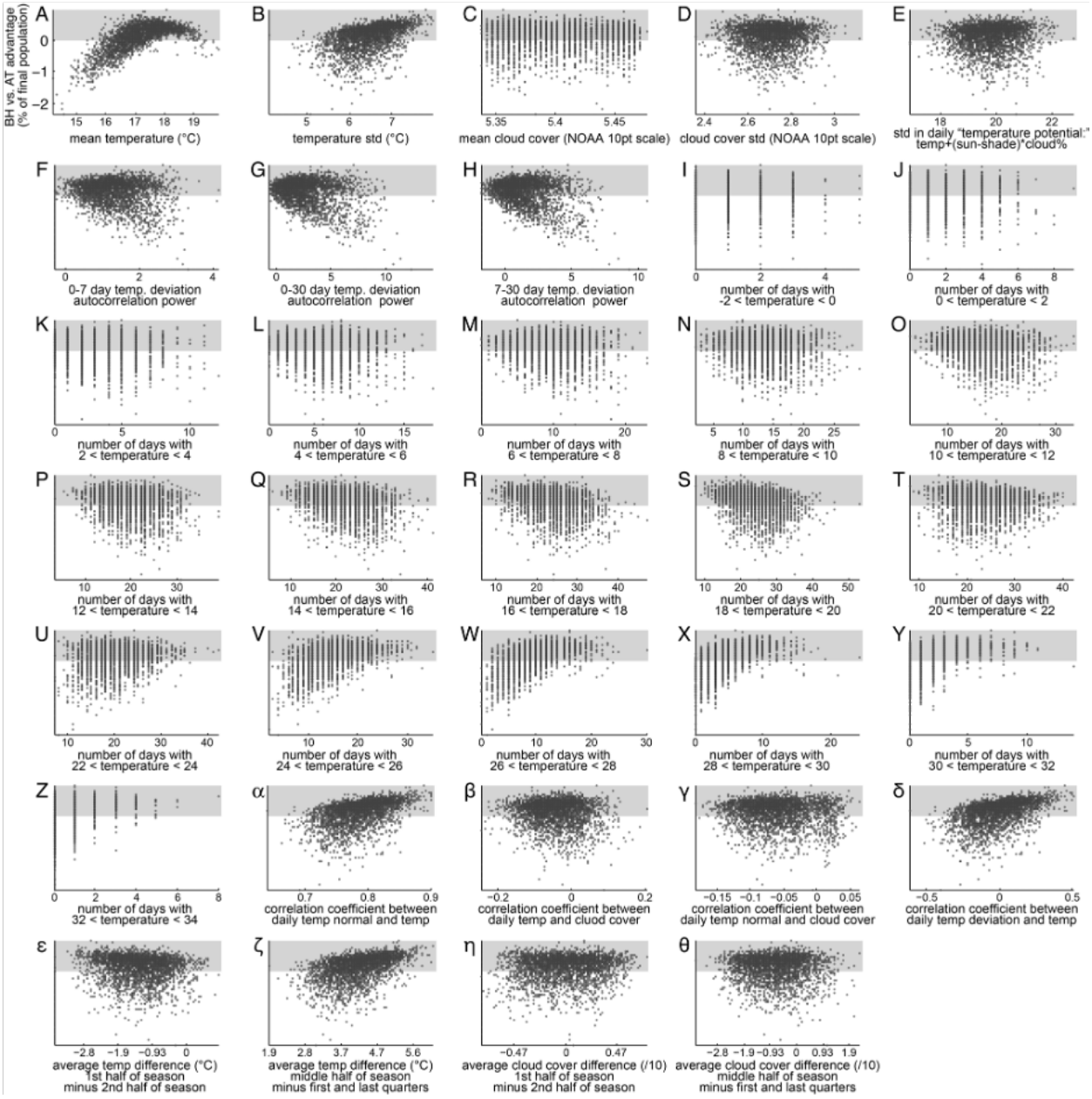


Figure 5.9 - BH versus AT advantage versus various seasonal measures.
All Y-axes as in the first panel. F-H were calculated by summing the normalized autocorrelation vector of the daily temperature deviations for the specified range of offsets. Most measures showing clear relationships with BH-AT advantage reduce to either mean temperature (A) or temperature standard deviation (B). As examples: Seasons with many days of moderate temperature (R, S) correspond to seasons of low temperature standard deviation. Conversely, seasons with more very high temperature days (X-Z) correspond to hot years (A). Seasons with greater correlation between daily temperature and temperature normal (α, γ) have more extreme temperature ranges, corresponding to (B).

We also analyzed the effects of shorter or longer breeding seasons by compressing or stretching random temperature and cloud cover histories into seasons ranging from 107 to 365 days (Figure 5.8F). The mean relative advantage of BH versus AT did not depend on season length, however the variance of BH advantage increased with season length. Only long seasons exhibited strong advantages for either BH or AT, presumably because increasing the number of generations per season increases the potential for adaptation, whether it be productive or counterproductive. The shortest seasons exhibited little difference between BH and AT.

Global climate change is predicted to shift evolutionary strategy from BH to AT

Across the 3000 random seasons, the BH vs. AT advantage never exceeds ~1% per season but could drop as low as ~-2% in some seasons (Figure 5.8D). Despite the longer negative tail in this distribution, the small advantage of BH over AT in most summers quickly accumulated across simulations of multiple sequential seasons (Figure 5.10A), indicating this strategy was highly favored on longer timescales. However, we found that an increase of only 2°C to the mean seasonal temperature was sufficient to change the evolutionary dynamic in favor of adaptive tracking (Figure 5.10B). Conservative models of global climate change predict increases in this range in the Boston area by the end of the century[37]. Thus, while seasonal weather fluctuations generally favor bet-hedging in thermal preference behavior, climate change will likely cause a phase-shift in the evolutionarily optimal strategy toward adaptive tracking.

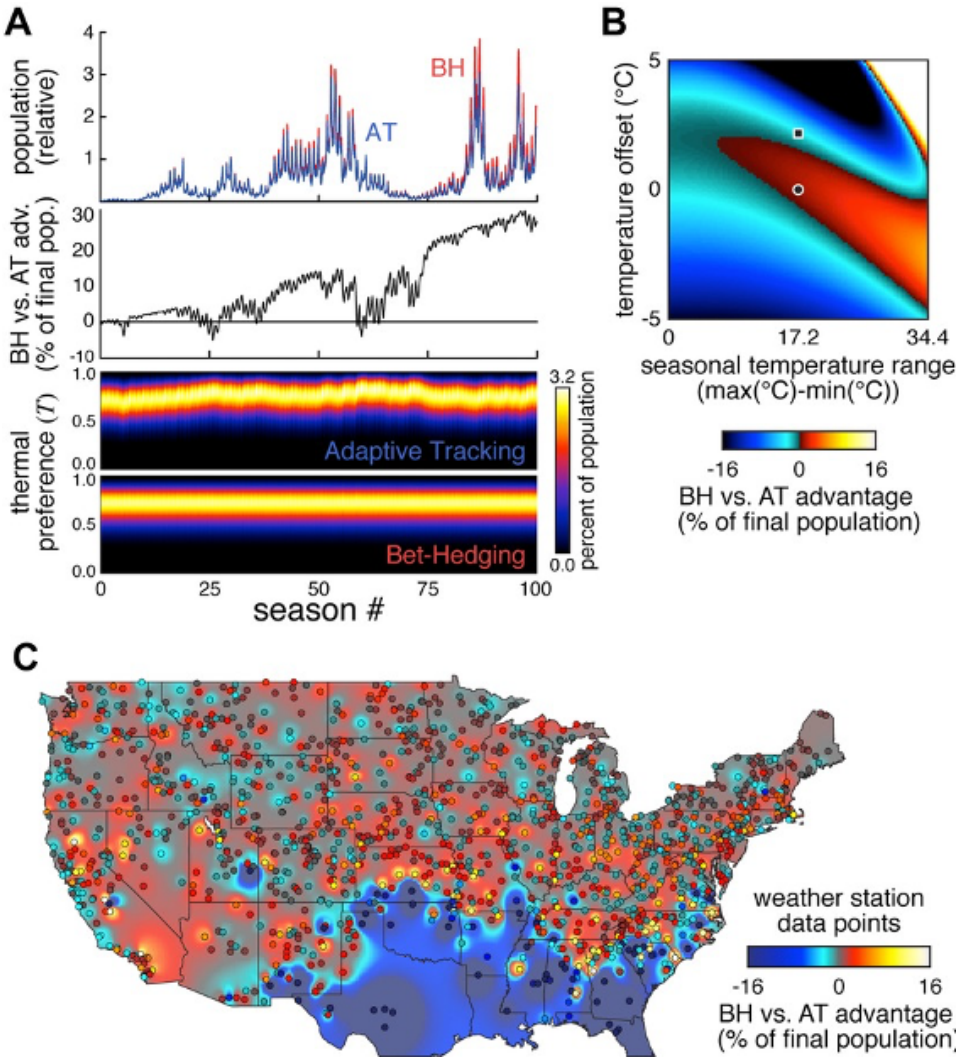


Figure 5.10 - Climatic and geographic variation in BH vs AT advantage.
(A) Relative population sizes for the BH (red) and AT (blue) versions of the model (top) and cumulative BH vs. AT advantage (middle) over the same 100 random simulated seasons. Bottom panel shows the corresponding abundance of flies as a function of thermal preference index and time, across 100 seasons, for each strategy. (B) Phase space of BH vs. AT advantage as a function of the two most predictive metrics. Color indicates magnitude of the advantage. Circle indicates current state while the square indicates the state if the average temperature were to increase 2°C. (C) Geographic map of BH vs AT advantage. Data points correspond to specific NOAA weather stations; background coloration is interpolated.

Geographical variation in BH vs AT advantage

Lastly, we considered to what extent the BH vs AT advantage we saw with Boston weather data was location specific. We ran the model using mean daily temperature data from more than 1400 weather stations across the continental United States[34] and compared the performance of

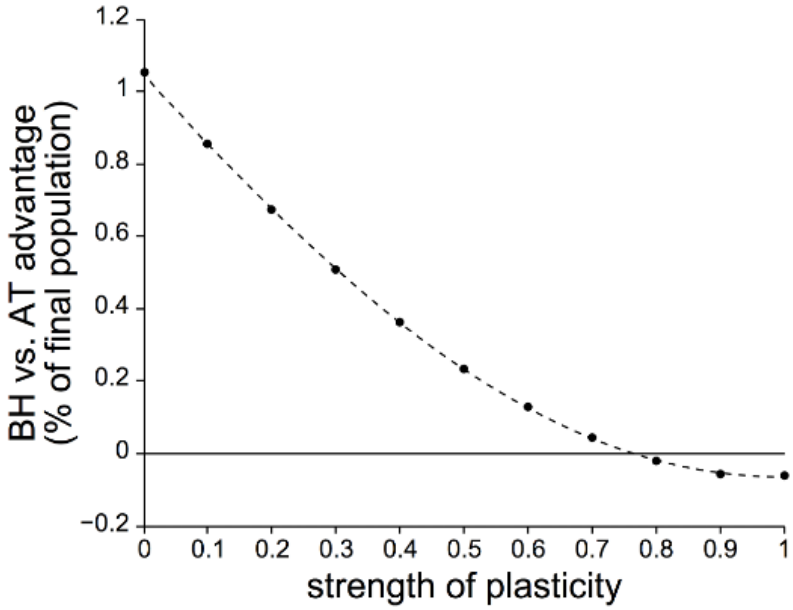
the BH and AT strategies (Figure 5.10C). Our model predicts substantial regional variation in the optimal strategy. In most locations, BH maintains a small advantage. In the deep south, where the breeding season is year-long, allowing more time for adaptation, BH performance is much worse than AT. However, the temperature extremes and shortened breeding season of regions just north (or at higher elevation in the southern Appalachians) renders BH strongly advantageous. This is consistent with the observation that a long breeding season can strongly favor either AT or BH (Figure 5.8F). Consistently, the short breeding seasons of higher latitudes and the Rocky Mountains favor neither AT nor BH strongly. AT appears to be favored along the Pacific coast, which is characterized by low temperature fluctuations.

Discussion

Here we explored whether a bet-hedging strategy could explain the large observed variation in temperature preference in *Drosophila*, as measured in phototactic and thermotactic paradigms. We find that in the face of fluctuating seasonal temperature selective pressures, adaptive tracking (in which progeny inherit the thermal preference index of their parent) always lags; by the time the population has adapted to the cool springtime with increased warm-preference, summer arrives. By contrast, the population grows faster if the behavioral preference of individual flies is non-heritable so that there are always spring-adapted and summer-adapted animals being born. The bet-hedging advantage is strongest under two conditions. 1) Highly variable temperatures (cool springs coupled to hot summers) magnify the selective pressure on the adaptive tracking population and thus produce larger counterproductive changes in genotype frequency as the temperature fluctuates throughout the season. This is consistent with the observations of seasonally fluctuating allele frequencies in flies[32]. 2) When mean temperatures are typical, the ability of the AT strategy to adaptively evolve is superfluous. In one example, the year 2010 was warmer on average, and its spring was particularly warm, reducing the seasonal temperature variability. Both of these factors gave the AT strategy a relative boost for being able to evolve, and thus reduced the overall BH advantage (Figure 5.8B,C).

Beyond adaptive tracking and bet-hedging, another major strategy for dealing with environmental heterogeneity is plasticity, in which organisms adaptively tune their phenotype in direct response to environmental fluctuations. The set of plasticity strategies can even include hybrid strategies such as the moment-to-moment regulation of the extent of bet-hedging in response to environmental conditions. In the absence of constraints, such as metabolic cost or limits on achievable phenotypes, a plasticity strategy is tautologically optimal[38], though such constraints surely exist. Generating an empirical estimate of the costs imposed on *Drosophila* in response to environmental fluctuations is beyond our capabilities.

Instead, we offer three lines of evidence suggesting plasticity cannot explain away the apparent adaptive advantages of bet-hedging. First, we simulated flies that were able to use behavioral choices to achieve a preferred thermal experience, bounded by the environmental temperature range available between shade and sun. Varying strengths of this strategy were combined with bet-hedging, and we found that over a wide range of strengths of plasticity, bet-hedging continued to offer a relative advantage (Figure 5.11). Second, we observed striking behavioral variation in populations of animals grown in essentially identical conditions (laboratory culture); to first approximation, there were no environmental fluctuations (e.g. variations in ambient temperature or luminance) to which a plasticity strategy could respond. Third, under conditions of convex fitness functions (i.e. with a single predominant mode of fit phenotypes), plasticity strategies can be at a disadvantage compared to bet-hedging strategies even if they come with low costs[38]. The unimodal relationships between temperature and eclosion time and lifespan (Figure 5.3B, C) yield a convex fitness function in our case, suggesting that plasticity may be outcompeted by bet-hedging (or even adaptive tracking), even if it comes at a relatively low cost.



$$ATD = \begin{cases} -shadeDiff & \text{if } S - T < -shadeDiff \\ S - (T + x * shadeDiff * cloudCover) & -shadeDiff < S - T < shadeDiff \\ shadeDiff & S - T > shadeDiff \end{cases}$$

Figure 5.11 - Relative performance of bet-hedging vs. adaptive tracking, as a function of the strength of a concurrent plasticity strategy.
Top) Simulated flies experienced temperatures corresponding to the weighted average of their baseline strategy (either BH or AT) and a plasticity strategy in which they could choose to rest either in the shade, or the sun, in whatever portion brought their thermal experience closest to the species-wide preferred temperature (25°C). X-axis corresponds to the relative weighting of the plasticity strategy, with 0 indicating entirely BH or AT, and 1 indicating entirely the plastic strategy. Bottom) We simulated this strategy by assuming that the thermal experience of flies on a given day was *T* + *x* * *shadeDiff* * *cloudCover* + *p* * *ATD*, where *T* is the daily in-shade temperature, *x* is the animal's thermal preference, *shadeDiff* is the shade vs. sunlight temperature difference (7°C), cloud cover is the fraction of cloud cover (0-10 in NOAA coding, divided by 10 in our calculations), *p* is the strength of plasticity, and *ATD* is the "achievable temperature delta" (the temperature offset attainable by behavioral choices, equation given above).

Our analysis focused on *Drosophila melanogaster*, a species with a relatively short reproductive cycle capable of producing several generations within the breeding season. It is likely that species generating fewer generations per season (i.e. K-selected species) would be less subject to the pitfalls of an adaptive tracking strategy since they would respond less to any temperature fluctuation. While our model did not permit us to realistically change the life history of our simulated *Drosophila* in the context of real weather data, we were able to simulate

changes in the length of the breeding season (Figure 5.8F). Shorter seasons are comparable to a K-selected life history because they yield fewer generations per season. We found that, as hypothesized, shorter seasons reduce the difference between adaptive-tracking and bet-hedging strategies, while long seasons can favor either strategy depending on other factors (i.e. Figure 5.8E).

This modeling highlights the importance of population-level properties, namely the amount of variation and the heritability of that variation. Population-level traits touch on the topic of group selection[39], and indeed aspects of bet-hedging were sometimes conflated with group selection in the literature[6]. However, our models do not directly address this controversial issue because they have no reliance on specific population structures, (the concept of which largely evaporates when considering non-heritable traits). Importantly, selection still operates, in all implementations of our model, at the level of the individual.

Two avenues for future investigation emerge from our results. First, flies captured and assayed at different time points throughout the season should show differences in their mean thermotactic preference (Figure 5.4E), that reflect their mode of inheritance. Specifically, flies using a AT strategy and caught in the early summer would be comparatively warm-seeking, while flies using a BH strategy would be comparatively cool-preferring at the height of the summer, when the high temperature selectively shortens the lifespan of warm-seeking individuals. However, analysis of behavior across the breeding season must consider seasonal changes in allelic frequencies[32].

Second, flies from locales with large seasonal weather changes (e.g., Boston, Massachusetts, U.S.A.) may have greater behavioral variation than those from milder, less variant climates (e.g., coastal central California, U.S.A.; Figure 5.10C). This prediction plays out on a variety of

spatial scales, the largest being a latitudinal cline in the east and midwest, where southern climates favor AT and northern climates BH. This prediction is consistent with recent experiments showing that northern strains of *Drosophila subobscura* are more resistant to high intensity fluctuating thermal stress, but more sensitive to prolonged (but milder) constant offset conditions[40]. Further experiments are needed to test these hypotheses, as other groups have found no latitudinal signal across several measures of thermal tolerance and plasticity in Australian *Drosophila simulans*[41]*.* Moreover, both of these studies examined isofemale lines; examination of isogenic lines would more directly permit the detection of a relationship between latitude and bet-hedging derived behavioral variability.

There is also a third prediction from these models concerning the effect of climate change on these strategies. Due to incrementally increasing mean temperatures over time, AT becomes the more evolutionarily advantageous option as the organisms continually adapt to the new normal. An increase of 2°C will be sufficient to favor adaptive tracking over bet-hedging, a change predicted to take approximately one hundred years. As both phototactic[42] and thermotactic[17] preference are heritable in outbred populations, we expect that flies will be able to adapt to climate change, but not by employing bet-hedging. Heritability of individual behaviors is a prerequisite for the evolution of AT, and it is plausible that a switch in selective pressure on strategies could increase adaptive tracking by favoring individuals with deeper developmental canalization. This would reduce the phenotypic variance associated with any single genome and allow the distribution of genetic variation for behavioral traits to more directly determine the phenotypic distribution.

The underlying basis of individual differences in thermal preference also remains to be discovered. Many mechanisms are possible, such as variation in thermoreceptor expression or propensity to stop and rest, but our model is indifferent to underpinnings of individual variability.

The conclusions drawn from the models here are not meant to say that bet-hedging is the sole explanation for behavioral variation. However, we have found that under the constraint of experimental data on the magnitude of behavioral variability between individuals, and with a minimal set of assumptions, bet-hedging appears to be a more adaptive explanation of behavioral variation than deterministic genetic heterogeneity. Indeed, we believe that real *Drosophila* probably utilize at least three strategies – bet-hedging, adaptive tracking, and phenotypic plasticity – to optimize its survival in an uncertain world.

Methods

Behavior

The *Drosophila melanogaster* line CamA was established from a single mated female caught from the wild in Cambridge, MA U.S.A. and propagated in the lab for approximately two generations at typical *Drosophila* culture densities prior to behavioral testing. The line inbred-CamA was derived by 10 generations of sibling matings. All flies were cultured on standard growth medium (Scientiis) in 25°C incubators at 30-40% relative humidity on a 12-12h light-dark cycle. Phototactic experiments were conducted at 23°C. Both behavioral assays were conducted at 30-40% humidity in environmental rooms. We found no difference in the behavioral responses of males versus females and merged their data. For both assays, only those flies registering 10 or more choices were analyzed (flies with only a small number of choices yield noisy estimates of individual preference).

Age- and sex-controlled flies were placed singly into 30 tubes in the “slow photobox,” which is illuminated from below by diffused white LEDs (5500K, LuminousFilm) (Figure 5.1A). A 50% neutral density filter was used to generate a lit half and shaded half for each tube. The lit portion of the arenas were slightly (0.1-0.5°C) warmer than the shaded portion. The arenas, illuminators, and diffusers are mounted on kinematic flexure mounts allowing ~1cm translation parallel to the testing tubes, under the control of a solenoid/microcontroller system driving vibration at 20 Hz. Agitation of the animals induced them to run and thereby reset their position between successive measurements of their light/shade preference. Each trial consisted of agitation (three 2s pulses, each separated by a 1s pause), an interval of 577s, acquisition of the photo used to score animal position, and a 15s interval completing the 10m trial. Animal position was determined by subtracting the background image of the rig and calculating the centroid of all pixels that had changed relative to the background (on a tube-by-tube basis), subject to a

noise-eliminating threshold. Animals not registering more than 10 choices were excluded from analysis.

The slow thermobox (Figure 5.1D) was fabricated by placing the acrylic tray of choice tubes used in the slow photobox down on a slab of aluminum with thermal grease. The aluminum slab was in contact with two larger aluminum blocks, one warmed to 40°C with resistive heating elements, and one cooled to 10°C with thermoelectric coolers (Peltier elements). The temperature of both larger blocks was held constant by PID controllers reading insulated resistance temperature detectors (3-wire, 100ohm). The 30-18°C gradient achieved within the choice tubes was measured using an infrared thermometer gun and was highly linear. For each of 20 trials, animals were first agitated by flowing air into the choice tubes, dislodging the animals toward the warm end. After 9.5 minutes the tubes were photographed, and the position of each animal measured digitally.

Day-to-day persistence of phototactic preferences was measured in a modified apparatus in which the floors of the imaging tubes were open at either end unto a surface of standard fly food poured in a ~0.5cm thick layer. This way, the flies could feed during an extended 40h trial. Day-to-day persistence of thermotactic preference was measured by the standard assay, individual housing of flies overnight, and retesting under the standard protocol.

Temperature measurement

Temperature differences between sun and shade were measured using an infrared thermometer gun on partly cloudy days in the summer and autumn. In one set of comparisons we measured the temperature of substrates in the shade of clouds, and then waited until ~5 minutes after the cloud had passed and measured their temperature in sunlight. In another set of comparisons, we compared adjacent sunlit and shaded (e.g. by a building or road sign)

substrates of the same orientation. Measured substrates included grass, brick, pine branches, tree bark, gravel etc.

Statistics

Data from individual flies that did not move upon agitation for 3 or more successive trials were discarded since these measurements were clearly non-independent from trial to trial. Sequential slow phototactic choices were found to have an average of 0.054 bits of mutual information across individuals, indicating effective independence (0 bits indicates complete independence in every animal, 1 complete dependence). This justifies treating behavioral choices as independent events and shows that the agitation protocol succeeded in rousing the animals between trials. We therefore modeled the expected distribution of light-choices with a binomial distribution with parameter p equal to the average light-choice probability of all animals tested, and parameter n equal to the number of trials, 24.

Modeling

See Results and Figure 5.12 for descriptions of the model. In the bet-hedging implementations of the model, each fly was randomly assigned a thermal preference index drawn from the experimentally observed preference distribution (fit by a beta distribution; Figure 5.1C). In adaptive-tracking implementations, the seed population was initialized in that way, but all subsequent animals were assigned a preference identical to their mother's preference (thus the model is asexual). Stochastic simulations of finite populations were seeded with 100 flies with ages uniformly distributed on $[M(T), A(T)]$ – respectively the temperature-dependent mean ages of eclosion and death – since flies may overwinter as adults[43]. We also implemented a version of the model in which the seed population was synchronized to the egg stage. This model was qualitatively indistinguishable. Flies in this initial population were assigned to have developed at random in the sun versus the shade with a probability equal to the population mean thermal

preference index. Individual flies were simulated, removed from the virtual population at random according to the parameter δ, and born stochastically at a rate β from mature flies already in the population. The temperature experience of fly i on day j was determined as $p_i * shadeDiff * cloudCover_j + T_j$, where p_i is the thermal preference index of fly i, $shadeDiff$ is the temperature difference between light and shade, $cloudCover$ is the average fraction of cloud cover on day j, and T_j is the in-shade temperature on day j. The birth and death rate parameters were identified (by grid search or hill-climbing algorithm) as the unique pair of values that satisfy two assumptions: 1) the fly population neither grows nor diminishes across the breeding season, i.e it is at numerical equilibrium, and 2) the mean thermal preference index does not evolve across the breeding season, i.e. flies are adapted to typical conditions. For every distinctive weather model, parameter fitting was independently performed using the adaptive tracking implementation. See Table 5.1 for parameter values.

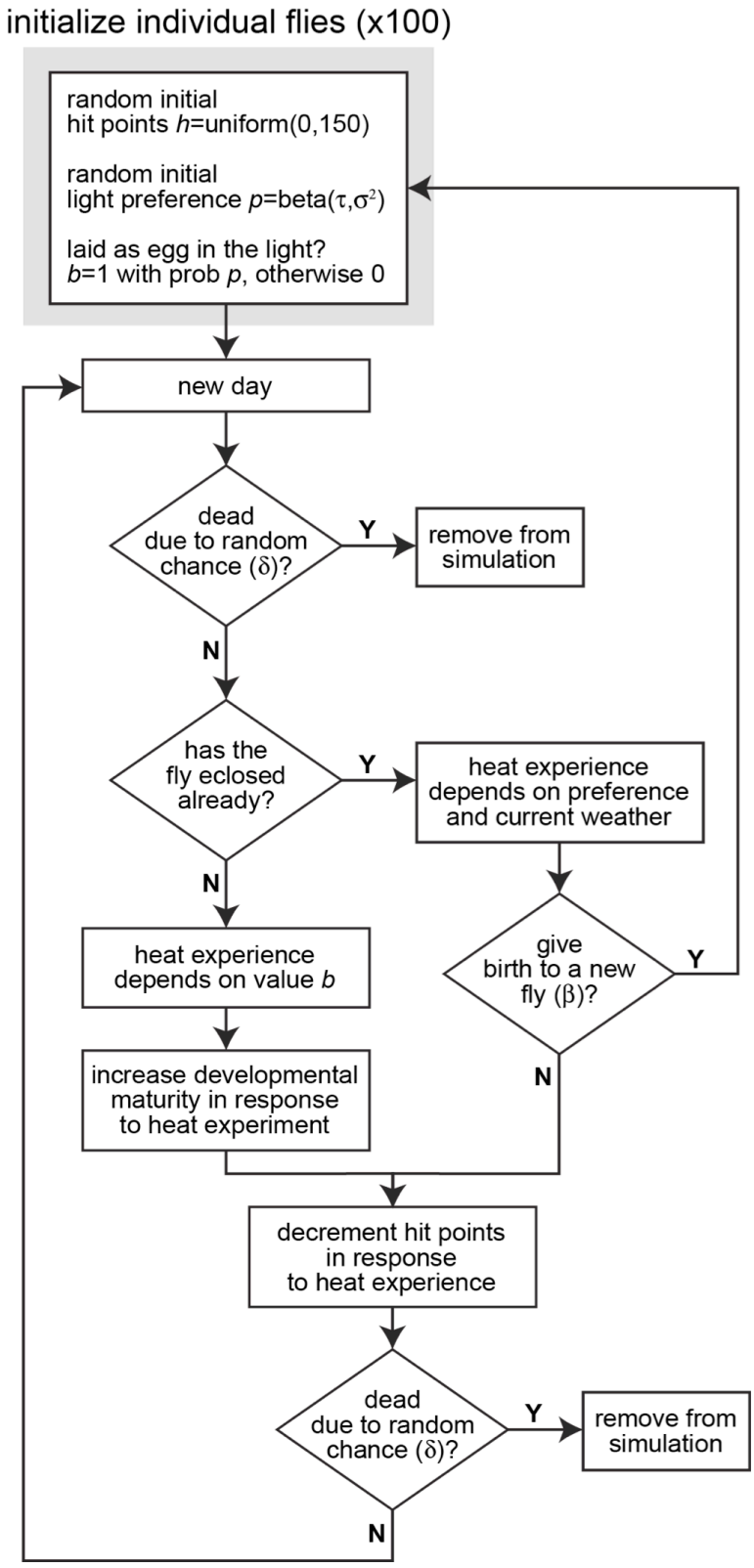


Figure 5.12 - Flowchart of the stochastic agent-based implementation of the fly life history model. See Results for additional explanation.

A related version of the model, which simulates infinite population sizes, was implemented analogously using a system of difference equations, but could be used to efficiently evaluate historical and simulated daily temperature deviations and cloud-cover values. In this implementation of the model, clouds reduced the maximum ambient temperature difference attainable by individual flies in proportion to the mean daily cloud cover fraction. Historical daily temperature deviations were normally distributed and modeled using a 30 parameter autoregression filter of normally distributed white noise. Random cloud cover was generated by drawing a season-long sequence of values from the observed (non-Gaussian) distribution of cloud cover fractions. These values were then shuffled until the new cloud cover sequence was no longer correlated with the original sequence ($r < 0.1$), under the constraint that the autocorrelation of the simulated sequence was correlated to that of historical cloud data with $r>0.998$, thus preserving temporal statistical structure of the sequence. Historical cloud and temperature deviation data were uncorrelated ($r = 0.02$), so simulated sequences of these variables were derived independently.

Difference equation model

The difference equation model was:

$$N_j(t) = N_j(t-1) * (1 - \delta - \Delta(j)) + \sum_i (N_i(t-1) * F(i) * P_j(i) * \beta)$$

$$\Delta(j) = 1/(0.04074\tau(j)^2 - 28.356\tau(j) + 506.2)$$

$$\tau(j) = \mathrm{Temp}(t) + j * 5°C * \mathrm{Sky}(t)$$

$$F(i) = 1 - D(i)/L(i)$$

$$L(i) = 1/(\Delta(i) + \delta)$$

$$D(i) = 0.2306\tau_D(i)^2 - 11.828\tau_D(i) + 158.34$$

$$\tau_D(i) = \frac{1}{t_{D_2} - t_{D_1}} \sum_{t'=t_{D_1}}^{t_{D_2}} (\mathrm{Temp}(t') + i * 5°C * \mathrm{Sky}(t'))$$

$$t_{D_1} = \lfloor 0.5 + \max(1, t - L(i)/2) \rfloor$$

$$t_{D_2} = \lfloor 0.5 + \min(t, t_{D_1} + D(\tau'_D)) \rfloor$$

$$\tau'_D = \frac{1}{t'_{D_2} - t_{D_1}} \sum_{t''=t_{D_1}}^{t'_{D_2}} \mathrm{Temp}(t'')$$

$$t'_{D_2} = \min(t, t_{D_1} + 21)$$

$$P_j(i) = \begin{cases} \mathrm{BetaPDF}(i, a, b) : h^2 = 0 \\ \begin{cases} 1 : i = j \\ 0 : i \neq j \end{cases} : h^2 = 1 \end{cases}$$

Here, $N_j(t)$ is the number of flies alive at time t with thermal preference index j. $\Delta(j)$ is the rate at which flies die due to thermal experience-dependent mortality as a function of j. $\tau(j)$ is the effective temperature experienced by flies with thermal preference index j, with Temp(t) indicating temperature and Sky(t) indicating respectively the temperature and cloud cover fraction at time t. The summation term in $N_j(t)$ indicates the number of flies born at time t with thermal preference index j, born from parents with thermal preference index i, which depends on the population sizes of flies with thermal preference index i at time t-1($N_i(t\text{-}1)$), the fraction of each of those parental subpopulations which are fertile ($F(i)$) and the probability densities of

parental thermal preference index ($P_j(i)$) conditioned on the thermal preference index of the progeny (j), and given the alternative BH vs AT strategies. ($P_j(i)$ is coded as a matrix with probability entries in row j, column i. For strategy AT, it is the identity matrix; for strategy BH, every row of $P_j(i)$ equals the beta-fit distribution from Figure 1C.) $F(i)$ depends on the ratio of development time $D(i)$ to total lifespan $L(i)$ of flies with thermal preference index i. $D(i)$ depends on the effective temperature experienced by parents (as this determines egg laying site) during development $\tau_D(i)$ which we approximate as the mean effective temperature across a range starting at time t minus half the typical lifespan, and ending $D(\tau'_D)$ days later (bounded by the time endpoints of the simulation). Development time is dependent on integrated temperature, which in turn depends on the length of development, given temperature's temporal fluctuation. The calculation of $D(\tau'_D)$ reflects one level of recursion in the calculation of this feedback. τ'_D is calculated as the average temperature from t'_{D1} through 21 days later, an interval approximating half a typical lifespan. The results of the difference equation model are very robust to the choice of the intervals in this recursion approximation, as well as the number of recursive levels implemented.

In comparisons of populations with differing initial thermal preference index distributions, we could not use the same values of δ and β for both conditions. Thus, an approach of fitting those parameters to satisfy the constraint of constant population size from beginning to end of the breeding season would not work - both populations would tautologically have identical populations at the end of the season. In these cases, we calibrated δ and β using a different assumption: identical mean population sizes across the breeding season, thus allowing the final population size to vary, and allowing us to assess relative growth rates. The other fitting assumption, that the mean thermal preference index did not evolve, was invoked in all cases.

In simulations of sequential seasons, the mean thermal preference index of the initial population of each season was set to the mean of population at the end of the previous season, but the variance was reset to match the empirical data. In geographic simulations, breeding seasons were defined as all days between the first day of the year in which temperatures reach 6.5°C and the first day when mean temperatures fall below 10°C, the same thresholds used in the Boston season. The non-parity in these values reflect our understanding that the first thaw suffices to end diapause while the first frost is sufficient to trigger it. The specific predictions associated with some stations are sensitive to these bounds, but the overall geographic patterns are not. The *β* and *δ* parameters were fit independently for each station automatically using a hill-climbing algorithm. Included stations were chosen at random from the 7500 stations in the NOAA data set, however, the algorithm was unable to fit the model parameters for some stations in very hot regions, i.e. some of the deep south and the Mojave Desert, so station geographic sampling is not unbiased. Background interpolation in geographic maps was done pixel by pixel using the function $b_{x,y} = \sum_i b_i * w_i / \sum_i w_i$, i.e. the average of all stations indexed by *i* and weighted by w_i, where $w_i = ((x - x_i)^2 + (y - y_i)^2)^{-2/3}$, i.e. inverse Euclidean distance from the pixel (*x*,*y*) to station *i* raised to the third power. This exponent was chosen to ensure a sharp drop-off with distance from the station but is otherwise arbitrary.

References

1. Sultan, S. E. Phenotypic plasticity for plant development, function and life history. *Trends in Plant Science* **5**, 537–542 (2000).

2. DeWitt, T. J., Sih, A. & Wilson, D. S. Costs and limits of phenotypic plasticity. *Trends Ecol. Evol.* **13**, 77–81 (1998).

3. Murren, C. J. *et al.* Constraints on the evolution of phenotypic plasticity: limits and costs of phenotype and plasticity. *Heredity* **115**, 293–301 (2015).

4. Kawecki, T. J. The Evolution of Genetic Canalization Under Fluctuating Selection. *Evolution* **54**, 1–12 (2000).

5. Huang, Y., Wright, S. I. & Agrawal, A. F. Genome-Wide Patterns of Genetic Variation within and among Alternative Selective Regimes. *PLOS Genet.* **10**, e1004527 (2014).

6. Hopper, K. R. Risk-spreading and bet-hedging in insect population biology. *Annu. Rev. Entomol.* **44**, 535–560 (1999).

7. Simons, A. M. Modes of response to environmental change and the elusive empirical evidence for bet hedging. *Proc. R. Soc. B Biol. Sci.* **278**, 1601–1609 (2011).

8. Levy, S. F., Ziv, N. & Siegal, M. L. Bet Hedging in Yeast by Heterogeneous, Age-Correlated Expression of a Stress Protectant. *PLOS Biol.* **10**, e1001325 (2012).

9. Haccou, P. & Iwasa, Y. Optimal Mixed Strategies in Stochastic Environments. *Theor. Popul. Biol.* **47**, 212–243 (1995).

10. Müller, J., Hense, B. A., Fuchs, T. M., Utz, M. & Pötzsche, Ch. Bet-hedging in stochastically switching environments. *Journal of Theoretical Biology* **336**, 144–157 (2013).

11. Cohen, D. Optimizing reproduction in a randomly varying environment. *J. Theor. Biol.* **12**, 119–129 (1966).

12. Svardal, H., Rueffler, C. & Hermisson, J. Comparing environmental and genetic variance as adaptive response to fluctuating selection. *Evol. Int. J. Org. Evol.* **65**, 2492–2513 (2011).

13. Sasaki, A. & Ellner, S. The Evolutionarily Stable Phenotype Distribution in a Random Environment. *Evolution* **49**, 337–350 (1995).

14. Dochtermann, N. A., Schwab, T. & Sih, A. The contribution of additive genetic variation to personality variation: heritability of personality. *Proc. R. Soc. B Biol. Sci.* **282**, 20142201 (2015).

15. Ayroles, J. F. *et al.* Behavioral idiosyncrasy reveals genetic control of phenotypic variability. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 6706–6711 (2015).

16. Buchanan, S. M., Kain, J. S. & de Bivort, B. L. Neuronal control of locomotor handedness in *Drosophila*. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 6700–6705 (2015).

17. Dillon, M. E., Wang, G., Garrity, P. A. & Huey, R. B. Thermal preference in Drosophila. *J. Therm. Biol.* **34**, 109–119 (2009).

18. Waddington, C. H., Woolf, B. & Perry, M. M. Environment Selection by Drosophila Mutants. *Evolution* **8**, 89–96 (1954).

19. Rockwell, R. F. & Seiger, M. B. Phototaxis in Drosophila: A Critical Evaluation. *Am. Sci.* **61**, 339–345 (1973).

20. Lewontin, R. C. On the Anomalous Response of Drosophila pseudoobscura to Light. *Am. Nat.* **93**, 321–328 (1959).

21. Seiger, M. B., Seiger, L. A. & Kertesz, J. A. Photoresponse in Relation to Experimental Design in Sibling Sympatric Species of Drosophila. *Am. Midl. Nat.* **109**, 163–168 (1983).

22. Scott, J. P. Effects of single genes on the behavior of *Drosophila*. *Am. Nat.* **77**, 184–190 (1943).

23. Pittendrigh, C. S. Adaptation, natural selection and behavior. in *Behavior and Evolution* 390–416 (Yale University Press, 1958).

24. Ashburner, M. *Drosophila: a laboratory handbook*. (Cold Spring Harbor Laboratory Press, 2005).

25. Ashburner, M. & Thompson, J. N. Jr. Laboratory culture of Drosophila. *Genet. Biol. Drosoph.* **2a**, 1–109 (1978).

26. Miquel, J., Lundgren, P. R., Bensch, K. G. & Atlan, H. Effects of temperature on the life span, vitality and fine structure of Drosophila melanogaster. *Mech. Ageing Dev.* **5**, 347–370 (1976).

27. Parry, D. A. Factors Determining the Temperature of Terrestrial Arthropods in Sunlight. *J. Exp. Biol.* **28**, 445–462 (1951).

28. Digby, P. S. B. Factors Affecting the Temperature Excess of Insects in Sunshine. *J. Exp. Biol.* **32**, 279–298 (1955).

29. Martin, T. L. & Huey, R. B. Why 'suboptimal' is optimal: Jensen's inequality and ectotherm thermal preferences. *Am. Nat.* **171**, E102-118 (2008).

30. Condon, C. *et al.* Indirect selection of thermal tolerance during experimental evolution of Drosophila melanogaster. *Ecol. Evol.* **5**, 1873–1880 (2015).

31. Tobler, R., Hermisson, J. & Schlötterer, C. Parallel trait adaptation across opposing thermal

environments in experimental Drosophila melanogaster populations. *Evol. Int. J. Org. Evol.* **69**, 1745–1759 (2015).

32. Bergland, A. O., Behrman, E. L., O'Brien, K. R., Schmidt, P. S. & Petrov, D. A. Genomic Evidence of Rapid and Stable Adaptive Oscillations over Seasonal Time Scales in Drosophila. *PLOS Genet.* **10**, e1004775 (2014).

33. Kain, J. S., Stokes, C. & de Bivort, B. L. Phototactic personality in fruit flies and its suppression by serotonin and white. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 19834–19839 (2012).

34. Arguez, A. *et al. U.S. Daily Climate Normals (1981-2010)*. doi:10.7289/V5PN93JP (2010).

35. Wright, S. Evolution in Mendelian Populations. *Genetics* **16**, 97–159 (1931).

36. Karasov, T., Messer, P. W. & Petrov, D. A. Evidence that adaptation in Drosophila is not limited by mutation at single sites. *PLoS Genet.* **6**, e1000924 (2010).

37. Meehl, G. A. *et al.* Global climate projections. in *IPCC, 2007: Climate Change 2007: the physical science basis. contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* 747–846 (Cambridge University Press, 2007).

38. DeWitt, T. J. & Langerhans, R. B. Integrated solutions to environmental heterogeneity: theory of multimoment reaction norms. in *Phenotypic Plasticity: Functional and Conceptual Approaches* 98–111 (Oxford University Press, 2004).

39. Wilson, D. S. & Wilson, E. O. Evolution 'for the Good of the Group'. *Am. Sci.* **96**, 380–389 (2008).

40. Castañeda, L. E., Rezende, E. L. & Santos, M. Heat tolerance in Drosophila subobscura along a latitudinal gradient: Contrasting patterns between plastic and genetic responses. *Evolution* **69**, 2721–2734 (2015).

41. van Heerwaarden, B., Lee, R. F. H., Overgaard, J. & Sgrò, C. M. No patterns in thermal plasticity along a latitudinal gradient in Drosophila simulans from eastern Australia. *J. Evol. Biol.* **27**, 2541–2553 (2014).

42. Dobzhansky, T. & Spassky, B. Artificial and natural selection for two behavioral traits in Drosophila pseudoobscura. *Proc. Natl. Acad. Sci. U. S. A.* **62**, 75–80 (1969).

43. Izquierdo, J. I. How does Drosophila melanogaster overwinter? *Entomol. Exp. Appl.* **59**, 51–58 (1991).