Dense seasonal sampling of an orchard population uncovers population turnover, adaptive tracking, and structure in multiple Drosophila species

Alyssa Sanae Bangerter
Beaverton, Oregon

Bachelor of Science, University of Utah, 2015

A Dissertation presented to the Graduate Faculty
of the University of Virginia in Candidacy for the Degree of
Doctor of Philosophy

Department of Biology

University of Virginia
August, 2021

ABSTRACT

Organisms require genetic variation in order to evolve, and balancing selection is one way by which genetic variation is maintained in genomes. Adaptation to temporally varying selection is a mechanism of balancing selection that is not well characterized in natural populations. In my dissertation, I utilize the fruit fly *Drosophila melanogaster*, which adapts to seasonally varying selection, to better understand balancing selection through adaptation to temporally varying selection. In Chapter 1, I utilize frequent sampling across three years to identify the kinds of loci involved in adaptation to seasonal selection as well as the environmental drivers of selection. I find that coding loci and previously identified seasonally adaptive loci are involved in adaptive differentiation through time. I also identify extreme hot and extreme cold temperatures to have the strongest selective force on allele frequency changes through time. In Chapter 2, I investigate the concept that spatial population structure can bolster the ability of temporally varying selection to maintain large amounts of genetic variation across long periods of time by asking if natural populations of fruit flies have population structure. I identify signals of population structure in natural fruit fly populations, which contradicts long standing assumptions about fruit fly migration behavior. In Chapter 3, I ask if seasonal adaptation is a unique phenomenon, or if other species of fruit flies also exhibit signals of adaptation to seasonally varying selection. I find that other species of fruit flies look like established, overwintering populations, indicating that they are likely also adapting to seasonal change. Overall, my dissertation deepens our understanding of this model system of balancing selection.

ACKNOWLEDGEMENTS

None of this work would have been possible without the help of an army of undergraduates who valiantly helped with fly collections. My army of undergrads across three years of collection trips include: Cynthia Ong, Helen Stone, Matt Campbell, Danny Song, Warren Wheaton, Liam Miller, and Yasmin Khodaei. Thank you for the help and the good memories.

AnhThu Nguyen helped with or completed the preparation of every single sequencing library generated over the course this dissertation. Without her help, none of my sequencing data would exist. She went above and beyond to help when things went wrong and was always patient in helping me troubleshoot issues.

Joaquin Nunez performed the modeling discussed in Chapter 1, so I thank him for contributing to the interpretation of my data.

I also would like to thank the Chiles family as well as all the workers at Carter Mountain Orchard. They were incredibly accommodating and helpful in answering my many questions. I am grateful for their permission to allow me to regularly collect flies on their property.

Finally, I would like to acknowledge Research Computing at The University of Virginia. The bulk of my analysis was done on the university computing cluster, Rivanna. The staff, with a special shout-out to Karsten, were amazing at helping address computing questions and issues I ran into.

TABLE OF CONTENTS

INTRODUCTION

All organisms require genetic variation in order for evolution to take place. Balancing selection is any selection that maintains genetic variation, and can occur via several different mechanisms. One of the mechanisms of balancing selection is adaptation in response to temporal heterogeneity in selection pressures. Specifically, this adaptation must include changes in allele frequencies through time to track the shifting phenotypic optima, also known as adaptive tracking. In some theoretical work, adaptive tracking of temporally heterogeneous selection is predicted to not maintain genetic variation at a large scale in the genome or for long periods of time (Cvijović, Good, Jerison, & Desai, 2015; Ewing, 1979; Hedrick, Ginevan, & Ewing, 1976). However, newer theoretical frameworks (Bertram & Masel, 2019; Botero, Weissing, Wright, & Rubenstein, 2015; Wittmann, Bergland, Feldman, Schmidt, & Petrov, 2017) and observations of natural populations (Bergland, Behrman, O'Brien, Schmidt, & Petrov, 2014; Chakraborty & Fry, 2016; Ghosh, Andersen, Shapiro, Gerke, & Kruglyak, 2012; Mojica, Lee, Willis, & Kelly, 2012; Rodríguez-Trelles, Alvarez, & Zapata, 1996; Rodríguez-Trelles, Tarrío, & Santos, 2013) show that adaptive tracking of temporal heterogeneity can lead to the maintenance of genetic variation. Despite this newer evidence, we lack detailed understanding of how balancing selection through adaptive tracking to temporal heterogeneity occurs in natural populations, what selective agents lead to the maintenance of functionally important genetic variation, and what other demographic processes also play a role.

Examples of balancing selection in response to temporal heterogeneity are few, partially because it can be difficult to identify the type of balancing selection (i.e.

adaptation to temporal or spatial heterogeneity, heterozygote advantage, negative frequency dependent selection, etc.; Spurgin & Richardson, 2010). There are organisms where adaptation to temporal heterogeneity is the theorized reason for the maintenance of genetic variation (Hawley & Fleischer, 2012; Pespeni, Chan, Menge, & Palumbi, 2013; Pespeni & Palumbi, 2013) but few where the connection between genetic variation and temporally varying selection is clear (Bergland et al., 2014; Chakraborty & Fry, 2016; Ghosh et al., 2012; Mojica et al., 2012; Rodríguez-Trelles et al., 1996; Rodriguez-Trelles, Tarrio, & Santos, 2013). Most organisms that exhibit balancing selection do not represent predictable or repeatable systems of adaptation to temporal heterogeneity where changes in the direction of selection within an ecological timescale are guaranteed, making them non-ideal systems to study the details of balancing selection due to temporally varying selection. An ideal system to better understand the details of balancing selection through adaptation to temporal heterogeneity would be an organism with repeatable responses to changing selection, results generalizable beyond the population studied, and detectable changes on an ecological timescale.

One study system used to better understand adaptive tracking in response to temporally varying selection is the fruit fly *Drosophila melanogaster*. Temperate populations of *D. melanogaster* experience temporal heterogeneity in the form of seasonally varying selection. Each year, *D. melanogaster* is estimated to undergo approximately 10 generations in summer conditions and 1-2 generations in winter conditions. Observations of phenotypic variation (Ives, 1970) and genotypic variation (Machado et al., 2016) indicate that *D. melanogaster* is a species that overwinters. These overwintering individuals then re-initiate natural populations when resources become

more abundant in the spring, though there is little known about how fruit flies overwinter and survive the harsh winter months. Because *D. melanogaster* undergoes multiple generations across similar seasonal conditions where selection can result in changed allele frequencies, and they experience repeated seasonal cycles, fruit flies are an ideal model system in which to observe adaptive tracking to temporally varying selection.

Many phenotypes change in response to temporally variable selection in fruit flies. In experimental evolution regimes, flies evolved a reduced tolerance to extreme hot and cold temperatures after adaptation to a fluctuating temperature environment relative to populations evolved under constant extreme temperatures (Condon et al., 2015). Wild-caught flies, or close descendants, show a range of phenotypes that change depending on the season flies are collected. Flies collected closer to winter environmental conditions show increased desiccation tolerance in males (Rajpurohit et al., 2018), increased cold tolerance (Stone, Erickson, & Bergland, 2020), greater propensity to enter diapause (Schmidt & Conde, 2006), increased innate immunity (Behrman et al., 2018), altered cuticular hydrocarbon composition (Rajpurohit et al., 2017), and higher stress tolerance (Behrman, Watson, O'Brien, Heschel, & Schmidt, 2015) than flies collected after summer environmental conditions.

Many of the phenotypes that change in *D. melanogaster* in response to temporally varying selection are quantitative traits with many underlying loci contributing to changing trait values. We observe hundreds of SNPs that oscillate in allele frequency in seasonal end-point sampling across multiple years in one population (Bergland et al., 2014) and across globally collected seasonal samples (Machado et al., 2021). *D.*

melanogaster presents an ideal system in which to track somewhat repeatable and predictable adaptive tracking to temporal heterogeneity on a manageable time scale.

While *D. melanogaster* is an ideal study system in which to study adaptation to temporal heterogeneity, there is still much we do not understand about this system and how balancing selection is acting to maintain genetic variation. We do not know what environmental factors are driving temporal shifts in allele frequencies in natural populations. Additionally, while we have learned much using allele frequencies, there is other information to be gained if genotypes of individuals can be obtained, such as changes in genotype frequencies and tests of Hardy-Weinberg Equilibrium. Finally, seasonal adaptation in *D. melanogaster* has been shown to be a global phenomenon, but it is less certain how generalizable this work is to other species, including other species of *Drosophila*.

With my dissertation, I aim to increase our fine-grain understanding of adaptive tracking to seasonally varying selection in *Drosophila melanogaster*. I chose a local peach and apple orchard to be the focal orchard for my dissertation. Across three years, 2016-2018, I collected samples of Drosophilids from Carter Mountain Orchard in central Virginia, USA (37.99N, 78.47W) on at least a bi-weekly basis starting in mid-late June and ending in late November to early December. For each collection I brought wild Drosophilids into the lab, separated out *D. melanogaster*, and placed single wild-caught females into isolated vials in order to start isofemale lines. The offspring from these lines were then sequenced with two different strategies and those results form the first two chapters of my dissertation. In 2017 and 2018 I identified and collected the many different Drosophilids found in my samples, then characterized the variability in species

composition and sequenced wild-caught individuals to look for signals of established populations in Drosophilids that are not *D. melanogaster*. Overall, these different sequencing projects address three sets of questions found in the three chapters of my dissertation.

In my first chapter, I leverage the three years of dense sampling in a pooled sequencing project to gather fine-scale observations of allele frequency changes. I find genetic differentiation through time within my samples that point to year-to-year population turnover and provide insight into overwintering dynamics. Additionally, I identify outliers for adaptive differentiation through time that are enriched for coding loci and previously identified seasonally oscillating loci. This points to functional relevance and validation of the outliers found within the data. Finally, I find that the proportion of extreme hot days and the proportion of extreme cold days prior to collection are the most important selective agents associated with temporal changes in allele frequencies out of the abiotic environmental factors tested. This chapter expands our understanding of the model system of balancing selection as a consequence of adaptively tracking seasonally varying selection in *D. melanogaster*.

In my second chapter, I perform individual sequencing on male offspring of isofemale lines from each collection time point in 2016 to gather genotypic information rather than allele frequency estimates. Using genotypes, I identify an excess of homozygosity in the data. This observation could be the result of one, or all, of the following processes: artifacts of low read depth, selection, and cryptic population structure. I perform simulations to assess the potential artifactual increase in homozygosity and find that the data show an excess of homozygosity beyond the null

expectation generated by simulations. I then perform a sliding window analysis on elevated homozygosity across the genome to delineate between selection and cryptic population structure as drivers of elevated homozygosity. I find that some time points show signals of cryptic population structure while others do not. This suggests that there may be stochastic cryptic population structure within a year in a single orchard. Observing cryptic population structure in a highly mobile insect typically assumed to be locally panmictic could have implications for the potential of population structure in other insect species. Additionally, the interplay between population structure and temporal heterogeneity has been previously theorized to bolster the maintenance of genetic variation in systems of adaptation to temporal heterogeneity. Thus, the presence of cryptic population structure in *D. melanogaster*, which contradicts assumptions made about local panmixia, adds insight to the fruit fly model system of balancing selection under seasonal variability.

In my third chapter, I do pooled sequencing on a subset of Drosophilid species collected in 2017 and 2018 to compare observations in other Drosophilids to known seasonal patterns in *D. melanogaster*. I find that the species composition and abundance of different Drosophilids dramatically change within and between years. As species composition changes in the community, there is potential for alterations in interspecific competition to be a temporally varying driver of selection, thus pointing to the importance of measuring interspecific competitors when addressing adaptation to temporal heterogeneity. With the sequencing performed, I identify one species, *D. simulans*, that shows strong signals of population turnover while the others do not. *D. simulans* is known to die out in the winter and is recolonized by mass migration in the

following summers. The species without strong signals of population turnover are on a spectrum from known established populations to predicted to be extirpated and recolonized each year. My observations point to new understanding of overwintering behavior, which has downstream implications in the potential for seasonal adaptation, in other Drosophilid species.

Overall, my dissertation contributes to our understanding of adaptive tracking to temporally varying selection in *Drosophila melanogaster*, which in turn contributes to our understanding about the mechanisms and targets of balancing selection in a natural population. In identifying particular environmental factors driving temporal shifts in allele frequency, we can then follow how specific loci respond to those environmental drivers and their evolutionary trajectories. By also starting to dig in to the functional relevance to specific environmental drivers we have the starting point to identify which loci are the biologically relevant loci versus what loci might be oscillating in frequency as a result of genetic draft. The potential presence of cryptic population structure within a single orchard not only contradicts long-standing assumptions about fruit fly population dynamics, but it also points to a potential interaction between adaptation to temporal heterogeneity and spatial heterogeneity. Theoretical work predicts that alleles that may be lost in a subset of subpopulations through successive rounds of adaptation to temporal heterogeneity, like in the summer months for fruit flies, may be rescued through migrants from other subpopulations where those alleles have not been lost (Ewing, 1979). This interplay between temporal heterogeneity and population structure is thought to increase the likelihood that adaptation to temporal heterogeneity can maintain genetic variation in the longer term. Finally, by showing that there are other species that are established

populations like *D. melanogaster*, we can start expanding observations in fruit flies out to other important insect species. In this dissertation, I further develop the model system of balancing selection in response to temporal heterogeneity in *Drosophila melanogaster*.

Literature Cited

Behrman, E L, Watson, S. S., O'Brien, K. R., Heschel, M. S., & Schmidt, P. S. (2015). Seasonal variation in life history traits in two Drosophila species. *Journal of Evolutionary Biology*, *28*(9), 1691–1704.

Behrman, Emily L., Howick, V. M., Kapun, M., Staubach, F., Bergland, A. O., Petrov, D. A., Lazzaro, B. P., et al. (2018). Rapid seasonal evolution in innate immunity of wild *Drosophila melanogaster*. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1870), 20172599.

Bergland, A. O., Behrman, E. L., O'Brien, K. R., Schmidt, P. S., & Petrov, D. A. (2014). Genomic Evidence of Rapid and Stable Adaptive Oscillations over Seasonal Time Scales in *Drosophila*. (D. Bolnick, Ed.)*PLoS Genetics*, *10*(11), e1004775-19.

Bertram, J., & Masel, J. (2019). Different mechanisms drive the maintenance of polymorphism at loci subject to strong versus weak fluctuating selection. *Evolution*, *101*, 3490–14.

Botero, C. A., Weissing, F. J., Wright, J., & Rubenstein, D. R. (2015). Evolutionary tipping points in the capacity to adapt to environmental change. *Proceedings of the National Academy of Sciences*, *112*(1), 184–189.

Chakraborty, M., & Fry, J. D. (2016). Evidence that Environmental Heterogeneity Maintains a Detoxifying Enzyme Polymorphism in *Drosophila melanogaster*. *Current Biology*, *26*(2), 219–223.

Condon, C., Acharya, A., Adrian, G. J., Hurliman, A. M., Malekooti, D., Nguyen, P., Zelic, M. H., et al. (2015). Indirect selection of thermal tolerance during experimental evolution of *Drosophila melanogaster*. *Ecology and Evolution*, *5*(9), 1873–1880.

Cvijović, I., Good, B. H., Jerison, E. R., & Desai, M. M. (2015). Fate of a mutation in a fluctuating environment. *Proceedings of the National Academy of Sciences*, *112*(36), E5021–E5028.

Ewing, E. P. (1979). Genetic Variation in a Heterogeneous Environment VII. Temporal and Spatial Heterogeneity in Infinite Populations. *The American Naturalist*, *114*(2), 197–212.

Ghosh, R., Andersen, E. C., Shapiro, J. A., Gerke, J. P., & Kruglyak, L. (2012). Natural Variation in a Chloride Channel Subunit Confers Avermectin Resistance in *C. elegans*. *Science*, *335*(6068), 574–578.

Hawley, D. M., & Fleischer, R. C. (2012). Contrasting Epidemic Histories Reveal Pathogen-Mediated Balancing Selection on Class II MHC Diversity in a Wild Songbird. *PLoS ONE*, *7*(1), e30222.

Hedrick, P. W., Ginevan, M. E., & Ewing, E. P. (1976). Genetic Polymorphism in Heterogeneous Environments. *Annual Review of Ecology and Systematics*, *7*, 1–32.

Ives, P. T. (1970). Further Genetic Studies of the South Amherst Population of *Drosophila melanogaster*. *Evolution*, *24*(3), 507–518.

Machado, H., Bergland, A. O., Taylor, R., Tilk, S., Behrman, E., Dyer, K., Fabian, D., et al. (2021). Broad geographic sampling reveals the shared basis and environmental correlates of seasonal adaptation in *Drosophila. eLife*, *10*(e67577), 1–21.

Machado, H. E., Bergland, A. O., O'Brien, K. R., Behrman, E. L., Schmidt, P. S., & Petrov, D. A. (2016). Comparative population genomics of latitudinal variation in *Drosophila simulans* and *Drosophila melanogaster*. *Molecular Ecology*, *25*(3), 723–740.

Mojica, J. P., Lee, Y. W., Willis, J. H., & Kelly, J. K. (2012). Spatially and temporally varying selection on intrapopulation quantitative trait loci for a life history trade-off in *Mimulus guttatus*. *Molecular Ecology*, *21*(15), 3718–3728.

Pespeni, M. H., Chan, F., Menge, B. A., & Palumbi, S. R. (2013). Signs of Adaptation to Local pH Conditions across an Environmental Mosaic in the California Current Ecosystem. *Integrative and Comparative Biology*, *53*(5), 857–870.

Pespeni, Melissa H., & Palumbi, S. R. (2013). Signals of selection in outlier loci in a widely dispersing species across an environmental mosaic. *Molecular Ecology*, *22*(13), 3580–3597.

Rajpurohit, S., Hanus, R., Vrkoslav, V., Behrman, E. L., Bergland, A. O., Petrov, D., Cvačka, J., et al. (2017). Adaptive dynamics of cuticular hydrocarbons in *Drosophila*. *Journal of Evolutionary Biology*, *30*(1), 66–80.

Rajpurohit, Subhash, Gefen, E., Bergland, A. O., Petrov, D. A., Gibbs, A. G., & Schmidt, P. S. (2018). Spatiotemporal dynamics and genome-wide association analysis of desiccation tolerance in *Drosophila melanogaster*. *Molecular Ecology*, *27*(17), 3525–3540.

Rodríguez-Trelles, F., Alvarez, G., & Zapata, C. (1996). Time-Series Analysis of Seasonal Changes of the O Inversion Polymorphism of *Drosophila subobscura*. *Genetics*, *142*, 179–187.

Rodríguez-Trelles, F., Tarrío, R., & Santos, M. (2013). Genome-wide evolutionary response to a heat wave in *Drosophila*. *Biology Letters*, *9*(4), 20130228.

Rodriguez-Trelles, F., Tarrio, R., & Santos, M. (2013). Genome-wide evolutionary response to a heat wave in *Drosophila*. *Biology Letters*, *9*(4), 20130228–20130228.

Schmidt, P. S., & Conde, D. R. (2006). Environmental Heterogeneity and the Maintenance of Genetic Variation for Reproductive Diapause in *Drosophila melanogaster*. *Evolution*, *60*(8), 1602–1611.

Spurgin, L. G., & Richardson, D. S. (2010). How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1684), 979–988.

Stone, H. M., Erickson, P. A., & Bergland, A. O. (2020). Phenotypic plasticity, but not adaptive tracking, underlies seasonal variation in post-cold hardening freeze tolerance of *Drosophila melanogaster*. *Ecology and Evolution*, *10*(1), 217–231.

Wittmann, M. J., Bergland, A. O., Feldman, M. W., Schmidt, P. S., & Petrov, D. A. (2017). Seasonally fluctuating selection can maintain polymorphism at many loci via segregation lift. *Proceedings of the National Academy of Sciences*, *54*, 201702994–10.

CHAPTER ONE

Dense temporal sampling identifies yearly population turnover and confirms that temperature is an important driver of seasonal adaptive tracking in *D. melanogaster*

Abstract

The adaptive strategy taken to survive temporal environmental heterogeneity depends on the life history of the organism in relation to the duration and predictability of environmental change. Adaptive tracking, where changes in allele frequency change to better match the shifting optimal phenotype, is a strategy that when employed in response to temporal heterogeneity can lead to balancing selection. The fruit fly *Drosophila melanogaster* is a model system in which to study adaptive tracking and balancing selection in response to seasonally oscillating selective pressures. Using novel bi-weekly sampling across three years and pooled sequencing of a single *D. melanogaster* orchard population, I ask a series of questions to better understand fine-scale patterns of population differentiation, what type of loci are involved in seasonal adaptation, and what are some of the specific environmental selective agents. I find population turnover from year to year that points to severe overwintering bottlenecks. Adaptively differentiated outlier SNPs are enriched for coding loci and previously identified seasonally varying loci. I also see that the proportion of extremely hot and extremely cold days prior to collection are dominant selective agents. With my findings we better understand some of the dominant forces at play in the system of balancing selection in response to seasonally varying selection in *D. melanogaster*.

Introduction

Temporal environmental heterogeneity is a common condition that most organisms experience, whether it takes the form of wet and dry seasons or temperate climate seasonality. There are several strategies employed by organisms experiencing temporal environmental heterogeneity in order to survive changing conditions: plasticity, bet-hedging, and adaptive tracking. Adaptive tracking, or changes in allele frequency from one generation to the next to track the changing phenotypic optimum, will occur when enough generations pass under similar environmental conditions (Botero, Weissing, Wright, & Rubenstein, 2015). Investigations into rapid adaptation that tracks temporal environmental heterogeneity have become more frequent in the past two decades, but we know little about the loci that underlie such adaptation and the abiotic factors that drive rapid adaptive changes in allele frequency.

Adaptive tracking has been observed in several organisms. Long-term observations of Darwin's Finches show repeated phenotypic shifts in response to changing environmental conditions that alter the available food supply in the highly heritable traits of beak and body morphology (Gibbs & Grant, 1987; Grant & Grant, 1993; Grant & Grant, 2002; Price, Grant, Gibbs, & Boag, 1984). In stick insects, *Timema cristinae*, the melanistic morphs increase in frequency in response to several selective agents, including warm spring temperatures (Nosil et al., 2018). In wild populations of *Drosophila pseudoobscura*, large chromosomal inversions changed in frequency across seasons (Dobzhansky, 1948). Rapid adaptive tracking to fluctuating selection pressures is one way by which balancing selection can act in a population to maintain genetic variation (Gillespie, 1973). Some theoretical work on balancing selection through

adaptation to temporally fluctuating selection concluded that such forces would not be able to maintain large amounts of genetic variation under certain biological circumstances (Cvijović, Good, Jerison, & Desai, 2015; Ewing, 1979; Hedrick, Ginevan, & Ewing, 1976). However, more recent theoretical work (Bertram & Masel, 2019; Botero et al., 2015; Wittmann, Bergland, Feldman, Schmidt, & Petrov, 2017) and observations of natural populations (Bergland, Behrman, O'Brien, Schmidt, & Petrov, 2014; Chakraborty & Fry, 2016; Ghosh, Andersen, Shapiro, Gerke, & Kruglyak, 2012; Mojica, Lee, Willis, & Kelly, 2012; Rodríguez-Trelles, Alvarez, & Zapata, 1996; Rodriguez-Trelles, Tarrio, & Santos, 2013) show that early theoretical work did not capture what occurs in nature in response to temporal heterogeneity in selection. Adaptive tracking to temporal heterogeneity has the potential to result in balancing selection at many loci throughout a genome, thus impacting overall evolutionary patterns.

Drosophila melanogaster is an excellent model in which to study these processes as it undergoes several generations within a single season and across seasonal transitions, indicating it could adaptively track environmental changes. Fruit flies exhibit many heritable phenotypic changes that appear to repeatedly track seasonal change. When flies collected at different seasonal time points are tested in a lab common garden environment they show several phenotypes that track seasonal changes. For example, wild-caught flies or their descendants collected closer to the winter exhibit altered cuticular hydrocarbon (Rajpurohit et al., 2017), a slight increase in desiccation tolerance in males (Rajpurohit et al., 2018), greater propensity to enter diapause (Schmidt & Conde, 2006), higher stress tolerance (Behrman, Watson, O'Brien, Heschel, & Schmidt, 2015), and increased innate immunity (Behrman et al., 2018) when compared to flies collected shortly after or during

the summer. In addition to seasonal changes in phenotypes, *D. melanogaster* shows repeated oscillations in the allele frequencies of hundreds of SNPs in the genome when assessing spring and fall samples across multiple years in wild populations (Bergland et al., 2014). Globally collected spring and fall seasonal samples of *D. melanogaster* show that there is some predictability in which loci appear to oscillate with seasonal change (Machado et al., 2021). Other work has identified specific functional loci that oscillate seasonally (Cogni et al., 2014; Cogni et al., 2013; Paaby, Bergland, Behrman, & Schmidt, 2014). While we observe somewhat predictable seasonal oscillations in allele frequency globally in *D. melanogaster*, we know little about the functional relevance of these seasonally oscillating loci or the speed at which these loci track unknown specific selective pressures.

All the previous genetic work in *D. melanogaster* seasonal evolution focuses on seasonal end points (Bergland et al., 2014; Machado et al., 2021). These seasonal end points aim to include a single sample as early as can be collected and then a single sample before temperatures start to drop in the fall (but see frost samples in Bergland et al., 2014). While these samples provide insight into broad seasonal evolutionary dynamics, they can be somewhat misleading. It has been observed that early samples can appear more summer-adapted if they are collected following a very warm spring and early onset of summer while late samples can appear more winter-adapted if collected following sharp downturns in temperature (frost samples in Bergland et al., 2014; flipped model in Machado et al., 2021). This indicates that while seasonal end points serve an important function and purpose in looking at broad signals of seasonal evolution, they can show somewhat contrary signals that are potentially an artifact of the environmental

conditions prior to collection. To overcome the limitations of seasonal end-points, a set of more dense temporal collections could be informative of the important environmental cues that drive patterns of seasonal evolution.

Tying driving environmental factors to loci with seasonally varying allele frequencies is an important goal in understanding how rapid adaptation to fluctuating selection works to maintain genetic variation and impact broad patterns of evolution. Genotype-environment association (GEA) models that tie changes in allele frequencies to driving environmental factors (Coop, Witonsky, Rienzo, & Pritchard, 2010; Frichot, Schoville, Bouchard, & François, 2013; Gautier, 2015; Gompert, 2015; Günther & Coop, 2013) have been used to link abiotic factors like temperature, wind, and rainfall to differences in allele frequency across the globe in *D. melanogaster* populations (Bogaerts-Márquez, Guirao-Rico, Gautier, & González, 2020). Using GEA models for dense temporal sampling of a single population could allow us to associate seasonal changes in allele frequency to abiotic factors that drive adaptive tracking in seasonal evolution.

Here, I harness bi-weekly sampling of a single, wild *D. melanogaster* population across three years to further investigate questions of seasonal adaptation: (1) What are the whole-genome patterns of population differentiation through time? (2) What are the characteristics of functional loci that are differentiated through time? (3) Can I connect environmental selective agents with temporal shifts in allele frequency? I find that there are several clusters of loci within the genome that show signals of adaptive differentiation through time, and that this adaptive differentiation shows patterns in year-to-year turnover that may provide further insight into poorly understood overwintering dynamics.

I show that some of these adaptively differentiated loci are driven by temperature extremes experienced by flies prior to collection. I also present a dataset that can be utilized to address further questions in the context of global seasonal sampling efforts of *D. melanogaster*.

Methods

Sample Collection

Approximately every other week I collected a variety of Drosophilids from a local peach and apple orchard near Charlottesville, Virginia, USA (Carter Mountain, 37.99N, 78.47W). I began collections when peaches came into season in mid-late June where fruit flies have emerged from overwintering and are colonizing early ripening fruit, then ended collections when the orchard closed in late November to mid-December. Orchard closure corresponded with freezing temperatures and approximately a month without new apple varieties coming into season. These collections started in June 2016 and finished in November 2018 totaling 37 collections. See Table 1 for a summary of collections and Figure 3 for a temporal spread of sampling.

Samples collected in the orchard were brought to the lab in vials containing cornmeal-molasses food. After sitting in the lab for roughly 24 hours, wild-caught Drosophilids were sorted by species. I set isofemale lines using all wild-caught females identified as either *Drosophila melanogaster* or its visibly indistinguishable sister-species *Drosophila simulans*. After the offspring of isofemale lines reached adulthood, I used the genitalia of male offspring to distinguish species identity between *D. melanogaster* and

D. simulans. The first-generation offspring of the *D. melanogaster* isofemale lines were split by sex and then frozen in ethanol and stored at -20°C.

Sample Preparation & Sequencing

Pooled samples per collection time point were made with one randomly selected male offspring from every *D. melanogaster* isofemale line (see Table 1 for counts of flies per pool). I performed DNA extractions on pools using the extraction protocol outlined in Bergland et al., 2014. Extracted DNA was diluted in a 1:1 DNA:water mixture, then sonically sheared to create fragments 500 bp in length using a Covaris sonicator. Libraries were prepared with a NEBNext Ultra II kit using dual indices and following the manufacturer's protocol, using 8 cycles of PCR in the final PCR enrichment step. Following library preparation, we quantified the concentration of each library, and all libraries were pooled in equal concentrations. After pooling all libraries, we size-selected the pooled library on a Pippen for DNA in the 600-750 bp range. The sequencing library was sequenced on a NovaSeq using 2x150 paired-end reads.

Dataset Filtering & Preparation

Sequencing data was mapped and filtered for contamination following steps found in Machado et al., 2021. In brief: I filtered raw sequencing data for adaptor sequences using cutAdapt (Martin, 2011), followed by mapping with BWA-mem (Li & Durbin, 2010; Li, 2013) to the *D. melanogaster* 6.12 release genome assembly (Hoskins et al., 2015). Because of the high phenotypic similarity between *D. melanogaster* and *D. simulans* males, I removed any reads that appeared to be sequence contamination by *D.*

simulans individuals. To do that, high quality reads were reformatted into a raw fastq file and then were competitively mapped, again using BWA-mem (Li & Durbin, 2010; Li, 2013), to a *D. melanogaster*-*D. simulans* (Hu, Eisen, Thornton, & Andolfatto, 2013) combined genome to remove any reads that map preferentially to *D. simulans*. After mapping and decontamination, samples were run through samtools *mpileup*. I identified variable sites using the mpileup file as input for VarScan v2.3.9 (Koboldt et al., 2012) which generated a VCF file.

I filtered the VCF file following steps in Machado et al., 2021 to obtain a set of high confidence SNPs and allele frequencies. In short, I removed: sites +/- 10 bp of VarScan-identified indels, known repetitive element regions listed in a RepeatMasker bed file (A.F.A. Smit, R. Hubley & P. Green; RepeatMasker at http://repeatmasker.org), sites in regions of zero recombination (Comeron, Ratnappan, & Bailin, 2012), and non-biallelic sites. Following these filtering steps, I removed any SNPs in the upper 5% quantile of read depths at the site (summed across all 37 samples). Next, within each sample, any SNPs with a read depth in the upper 5% quantile were turned into missing data. Any sites with an average minor allele frequency (MAF) below 0.05 were removed from the dataset. Finally, large portions of heterochromatic and repetitive regions were added to chromosomal ends in the updated dm6 genome assembly (Hoskins et al., 2015); I removed variants in those regions from the dataset.

After the SNP-level filtering was completed, I looked at pool-level summaries of sequencing quality to determine if any of the pools were clearly of poor quality relative to the other pools. A plot of effective coverage (EC = (number of chromosomes * read depth)/(number of chromosomes + read depth)) (Bergland et al., 2014; Kolaczkowski,

Kern, Holloway, & Begun, 2011; Machado et al., 2021; Machado et al., 2016) versus missing data rate (Figure 1) showed three visible outliers with high rates of missing data and low effective coverage. I removed those three pools (2016_0624, 2016_1003, and 2018_0628) from the analysis. This final VCF was annotated with SnpEff (Cingolani et al., 2012).

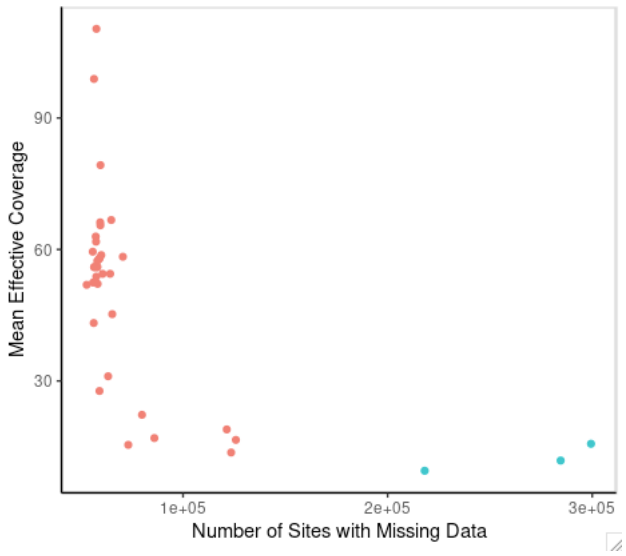


Figure 1: ***The missing data rate and effective coverage relationship indicates three clear outliers of poor quality.*** Missing data rate was calculated as the proportion of sites that do not have data after all rounds of filtering. The pooled samples are color coded by if they were included (red) or excluded from the analysis (blue).

Principal Components Analysis

I performed a principal components analysis on the allele frequency estimates of the samples to look for broad patterns of clustering due to genetic similarity. Principal components analyses were conducted on the SNPs in each autosomal arm individually and across all four autosomal arms combined (referred to as all autosomes) using the R package LEA (Frichot & François, 2015). Initial PCA indicated that samples with low

effective coverage were stark outliers and drowned out any discernible patterns. I therefore conducted the PCA and subsequent analysis on the PCs using only samples with higher effective coverage (mean + 1 standard deviation > 50; sample is above and/or crosses the dashed line in Figure 3B).

Following the PCA, I performed ANOVAs or linear models on PC loading values to test which principal components are associated with biological explanatory factors or artifactual drivers of clustering. I looked at the following possible biological explanatory values: year of collection, month of collection, and average inversion status. Average inversion status was calculated by using inversion-specific SNPs previously identified (Corbett-Detig & Hartl, 2012) and calculating the mean allele frequency of inversion-specific alleles (coordinates from Corbett-Detig & Hartl 2012 are in dm3, and were lifted over to dm6 using the UCSC browser tool). I also looked at the following artifactual explanatory factors: DNA extraction batch (there were two separate “batches" of DNA extractions done about a month apart), and row and column in the plate when building libraries.

Genetic Differentiation Through Time: Pairwise F_{ST} Calculations

I wanted to investigate patterns of genetic differentiation through time using F_{ST}, both within a single year and between years, including older collections to get a longer-term view of genetic differentiation. I took the filtered samples described above and added spring and fall samples of pooled sequencing data from the same focal population collected in 2012, 2014, and 2015 (Machado et al., 2021). I applied a simpler filtering approach to this dataset, retaining all sites that were retained in the 2016-2018-only

dataset; the more complex steps of converting high coverage sites into missing data were not applied to this larger dataset. This combined VCF was run through the R package poolfstat (Hivert, Leblois, Petit, Gautier, & Vitalis, 2018) using the *computePairwiseFSTmatrix()* command to calculate F_{ST} between every pairwise combination of samples. In displaying the data, spring and fall collection designations were kept for the 2012, 2014, and 2015 data, while the 2016-2018 data were classified as bi-weekly based on their collection strategy.

Genotype-Environment Associations

Environmental Data Gathering

In order to test putative associations between allele frequencies and environmental factors, I obtained climate data collected at or near the focal orchard. Climate data was gathered from Weather Underground weather stations. For the majority of the field season, data was available from weather stations located at or near the orchard the collections came from (Carters Mountain (CM) - KVACHARL73; Daniel Morris Lane (DML) - KVACHARL33). Whenever possible, I used data from the station at the orchard (CM), but when data was missing from the CM station, I estimated data based on DML station. DML weather data was tightly correlated with CM weather data (adjusted R^2 from a linear model of DML predicting CM weather data ranging from 0.7297 to 0.9810), so it was reasonable to fill in the gaps using a predictive linear regression.

I used measures of temperature and humidity as the environmental covariates, estimated across 14, 21, and 28 days prior to collection date. The 14, 21, and 28-day estimations were all calculated to capture different temperature-dependent lengths of *D.*

melanogaster development (Powsner, 1935). I investigated average temperatures (averages of low, average, and high temperatures), a proportion of the days preceding collections below or above approximate thermal limits for *D. melanogaster* (Chown, Jumbam, Sørensen, & Terblanche, 2009; Garcia, Littler, Sriram, & Teets, 2020; Lecheta, 2020; Overgaard, Hoffmann, & Kristensen, 2011; Petavy, David, Gibert, & Moreteau, 2001; Ransberry, MacMillan, & Sinclair, 2011), the average range of temperatures, and average relative humidity. I initially ran the model (described below) using conservative thermal limits (lower limit, 5°C; upper limit, 32°C) when calculating the proportion of extreme hot and extreme cold days, but this produced an inflation in the model output values. Therefore, I adjusted the thermal limits (lower limit, 8°C; upper limit, 30°C) in order to reduce the number of zeroes in the model environmental covariate inputs. I used my own records about what fruit variety the flies were collected from (Table 1). Fruit type was encoded as 1 for peaches, 2 for apples, and 1.5 for collections done on both peaches and apples. I used year and Julian date as covariates.

BayPass

BayPass v2.2 takes pooled sequencing data and uses a scaled genetic covariance matrix and Bayesian modeling to assess adaptive differentiation and association with environmental covariates (Gautier, 2015). Prior to running BayPass, the VCF datafile was converted to the BayPass format using R package poolfstat (Hivert et al., 2018). The core model, which generates the genetic covariance matrix, was run on autosomal SNPs (2L, 2R, 3L, 3R) using default settings (Figure 2). The output genetic covariance (relatedness matrix) file from the core model was used in the AUX covariate model, which identifies

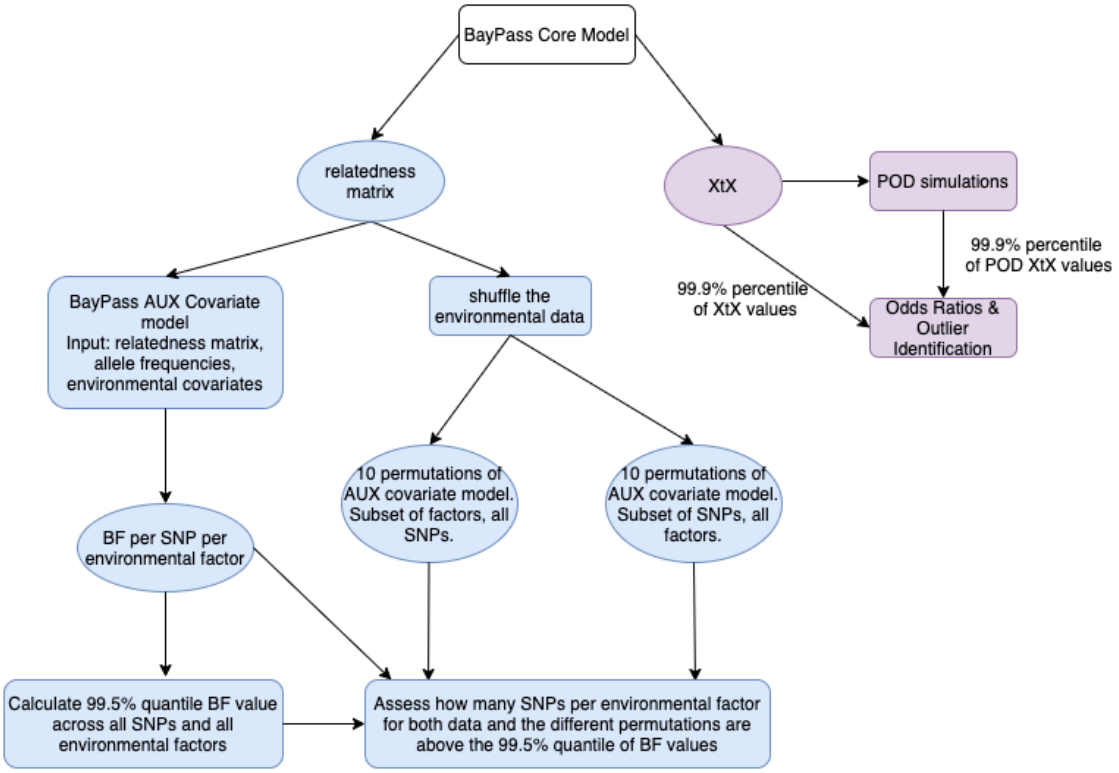


Figure 2: ***Workflow for BayPass analysis.*** The blue represents the analysis on the environmental covariate model output and the purple represents the analysis on XtX statistics. Circles represent outputs and rectangles represent modeling or analysis steps. BF=Bayes Factor.

connections between allele frequency changes in SNPs to each environmental covariates independently while taking into account the genetic covariance between samples. The AUX covariate model was run three times (set seed manually to 1,2,3 for the corresponding run) using the default settings.

I also ran the AUX covariate model with permutations of the environmental covariates in order to obtain a null distribution. Ten permutations were run for the covariate model by shuffling the environmental data and re-running the AUX covariate model with the permuted environmental data. The permutations of the environmental data were done by taking all the environmental data points from a collection and shuffling

them together as a block to maintain the correlation structure within a single time point for the various factors. Environmental data was shuffled within each year (e.g. 2016 environmental data was shuffled and assigned to new 2016 time points). However, year of collection was shuffled completely between all samples. Following the environmental data shuffling, and to reduce computational burden, I randomly selected one subset of SNPs per autosome arm (1/25th of the SNPs on a given chromosomal arm). Each permutation was run with three random seeds, comparable to how the data was modeled. An additional ten permutations were run on all SNPs using the same model structure but with a subset of six covariates that stood out in the data: the proportion of extreme hot days and proportion of extreme cold days across 14, 21, and 28 days.

Analysis of BayPass XtX Output

XtX statistics are designed to measure the degree of differentiation between populations that are due to non-demographic processes. Large XtX values, which correspond to a greater degree of adaptive differentiation (Gautier, 2015; Günther & Coop, 2013) were used to identify outlier SNPs that represent SNPs undergoing adaptive differentiation across time. I used two different outlier thresholds (Figure 2). The more conservative threshold was calculated as the top 0.1% of SNPs (99.9^{th} percentile; XtX>104.7178). The second threshold was based on pseudo-observed data sets (PODs) simulated using the *simulate.baypass()* command in R. PODs are generated and analyzed with the *simulate.baypass()* command by randomly sampling SNPs from the dataset that will maintain overall data parameters calculated by the BayPass core model (Gautier, 2015). Simulations were run on 10,000 SNPs using actual read depth data from the

pooled sequencing data. After analyzing the PODs with BayPass, the top 0.1% of POD SNPs (99.9th percentile) equaled a threshold at XtX>48.17961.

To characterize XtX outliers, I assessed levels of enrichment (i.e. odds ratios) across different loci categories: coding vs. non-coding, nonsynonymous vs. synonymous, globally seasonal vs. all other SNPs, and tandem repeat rich regions vs. the rest of the genome. Using the two different outlier thresholds, I looked for enrichments among the XtX outliers vs. 100 random samples of matched controls. Matched controls were identified by choosing random SNPs matched for chromosome, average minor allele frequency, and total SNP read depth. Global seasonality was categorized by using previously identified globally seasonal SNPs (Machado et al., 2021) at different significance thresholds. Because I observed an enrichment of tandem repeats in the XtX outliers, I re-ran the enrichment tests after excluding tandem repeat regions based on the simple repeat information from the UCSC genome browser (https://genome.ucsc.edu/cgi-bin/hgTables; *D. melanogaster* dm6, Variation & Repeats; Simple Repeats).

Analysis of BayPass Covariate Model Output

I assessed the output of the covariate model to identify genome-wide patterns of association with environmental covariates and to identify outliers and what environmental covariates they are associated with. After the covariate models on the data and permutations were completed, Bayes Factors for each environmental factor per SNP were averaged across the three BayPass runs. To assess which environmental factors appear to be the most important to temporal shifts in allele frequencies, I calculated the 99.5% quantile of Bayes Factors across all SNPs and environmental factors. Using the

99.5% quantile threshold, I calculated the proportion of SNPs for both the data and permutations that was above this threshold per environmental factor. This was done for the downsampled permutations initially, and then again for the full permutations of the 14-, 21-, and 28- day proportion of extreme hot and extreme cold days environmental factors.

To assess the distribution of outliers in the Bayes Factors with the 6 environmental factors I identified as being most tied with shifts in allele frequencies, I checked for overdispersion. Outliers were defined as a Bayes Factor above the 99.5 percentile for that dataset, whether it was the data or one of the ten full permutations. The overdispersion model looked at 1000-SNP windows with a 500-SNP step and counted up how many SNPs were above the outlier threshold of Bayes Factors either for the data, or for that permutation depending on what was being assessed. Following the counting of outliers, each window was compared to an expected even distribution of outliers using a one-tailed binomial test in R. P-values from this assessment of windows were then adjusted for multiple testing using R function *p.adjust()* with the default settings, the Holm correction (Holm, 1979).

Results

Basic dataset quality and description

From 2016 to 2018, I conducted 37 different sampling efforts at a local peach and apple orchard in central Virginia, US (37.99°N, 78.47°W), yielding approximately 2,700 *D. melanogaster* isofemale lines. From these flies, I selected male offspring for pooled

sequencing. The number of flies entering each pool was variable given the fluctuations in *D. melanogaster* density over the field season and from year to year (Figure 3A) and ranged from 10 to 150 individuals in the different time-point sample pools. The effective coverage, which accounts for the size of the pool, of most samples was either above or within one standard deviation the target effective coverage of 50. A few samples did not reach the target effective coverage due to a low number of flies included in the pool (Figure 3B). After removing pools with insufficient effective coverage (see methods), I retained 34 pooled samples containing 1,333,338 autosomal SNPs.

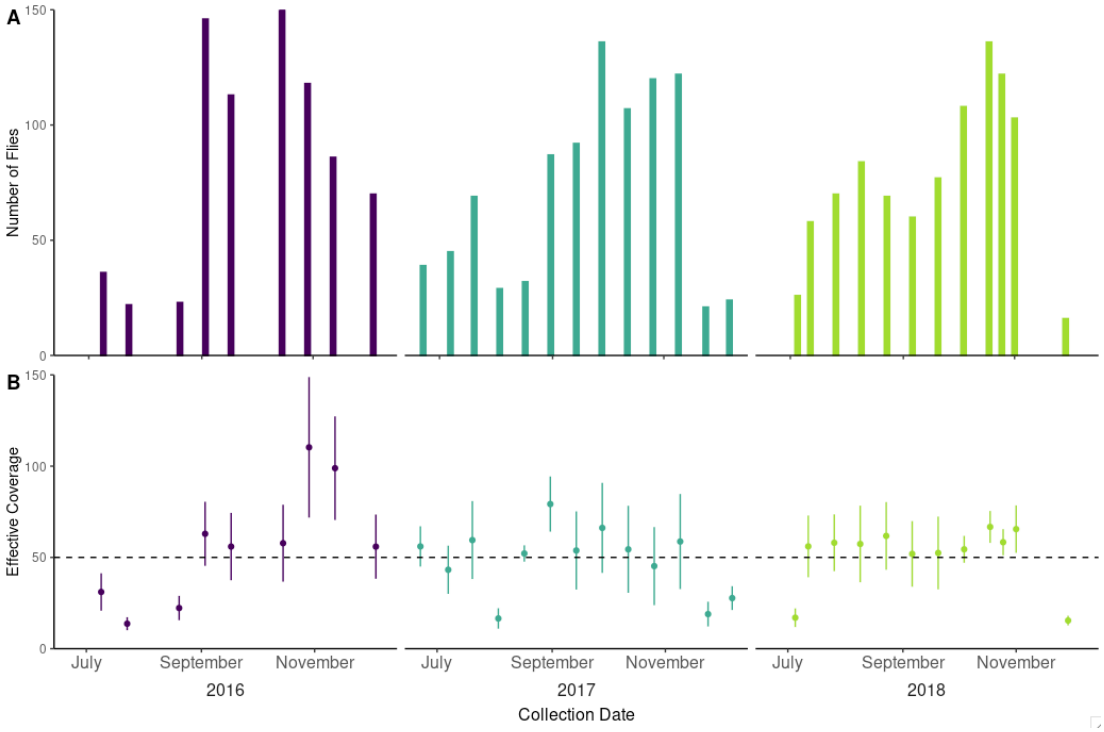


Figure 3: ***A basic description of the dataset used in the analysis.*** A) The number of male flies pooled into each sequenced time point sample. B) The effective coverage of each pooled sample. The point is the mean effective coverage across all autosomal SNPs included in the analysis and the line represents one standard deviation away from the mean. The dashed line represents the target effective coverage.

Whole genome patterns of differentiation

Principal components analysis (PCA) on autosomal allele frequencies revealed that both the year of the collection and the estimated Inv(2L)t inversion frequency separate the collections along principal components 1-3. Chromosomes 2R, 3L, 3R show samples clustering by year while 2L and the combination of all autosomes show samples clustering by Inv(2L)t inversion frequency (Figure 4). An ANOVA conducted on the PC loadings indicate that other potential confounding factors (position in the library making plate, DNA extraction batch, etc.) were not significantly associated with any other PCs (Bonferroni $\alpha<0.05$).

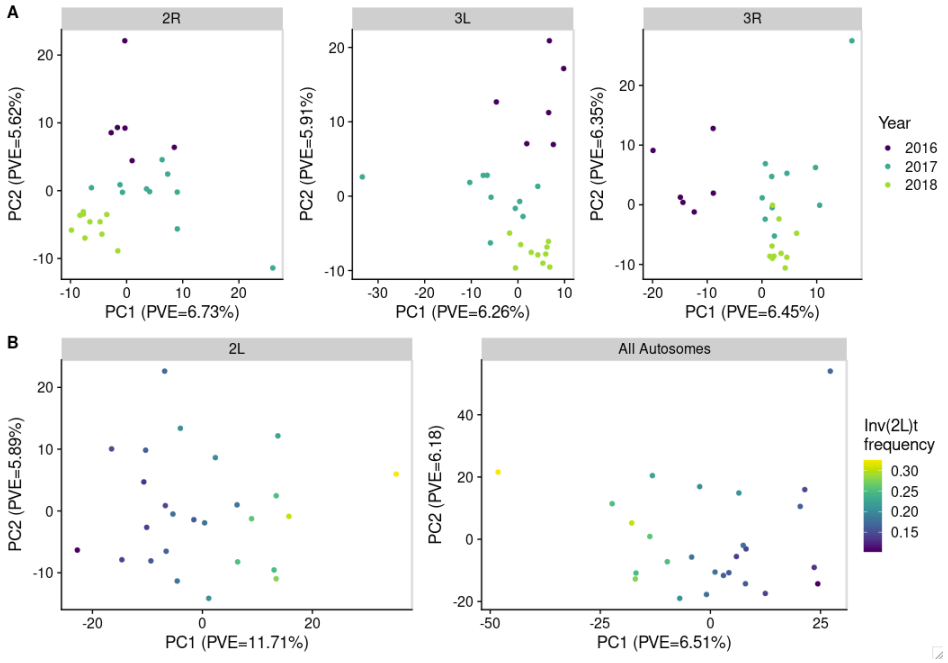


Figure 4: ***The principle components analysis show that year of collection and frequency of Inv(2L)t unique SNPs are the dominant explanatory factors for PC loading values.*** A) PC1 vs. PC2 for three main autosomal arms separately, colored by year. B) PC1 vs. PC2 for 2L and All Autosomes, colored by average estimated inversion status of Inv(2L)t.

Pairwise F_{ST} through time appears to plateau when comparing samples collected during different years, no matter what year the oldest sample in the pairwise comparison comes from (Figure 5). However, there is also a general increase through time (Mantel test, P-value=0.001).

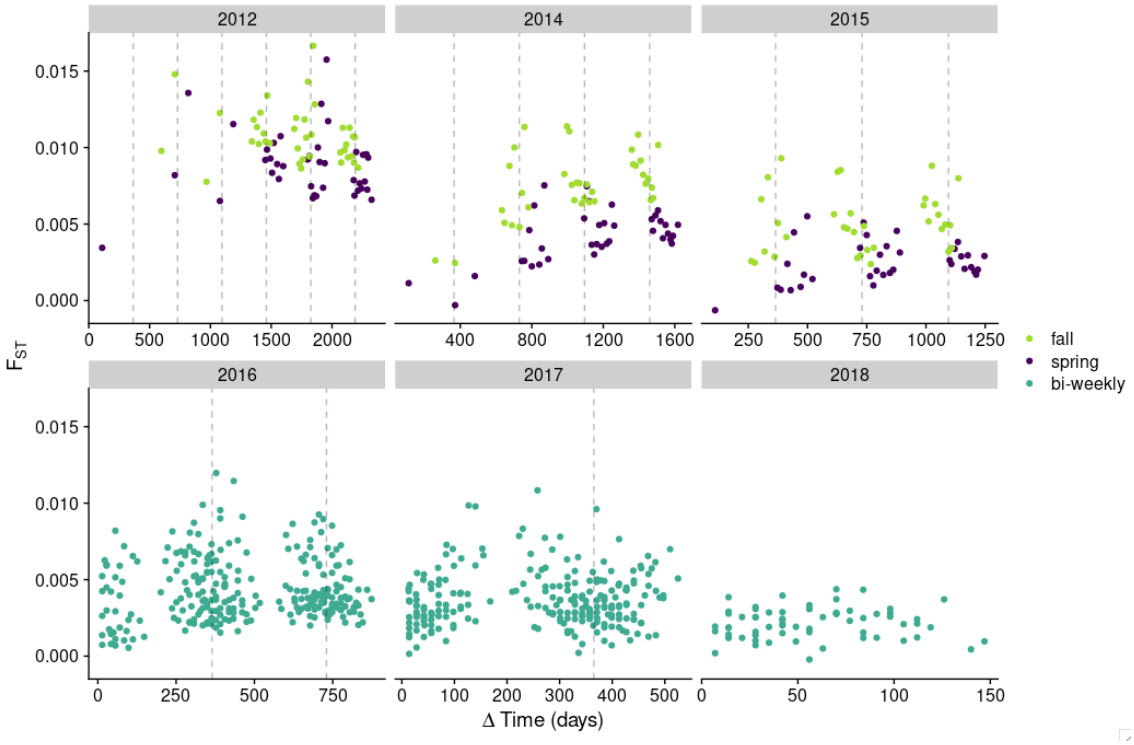


Figure 5: ***F_{ST} through time appears to plateau after one year.*** Pairwise F_{ST} between 2016-2018 samples and other data from 2012, 2014, and 2015. Data points are sorted into facets by the year of the oldest sample in the pairwise comparison. Similarly, the fall, spring, and bi-weekly collection categorization is based on the oldest sample in the pairwise comparison. Δ Time is measured in the days between the oldest and youngest sample in the pairwise comparison. The dashed vertical lines align with every 365 days.

Demographic modeling of population turnover

To test hypotheses of the cause of population turnover being due to either severe bottlenecks or complete extirpation and recolonization, we conducted simple

demographic model simulations (Bangerter, Nunez, & Bergland, 2021). In brief, we performed forward simulations using SLiM 3.0 (Haller & Messer, 2019) of three demographic situations: constant population size, cyclical bottlenecks, and cyclical extirpation followed by recolonization from a source population. Only a cyclical bottleneck to 2% of the maximum population size somewhat recapitulates the PCA results of samples clustering by year of collection (Bangerter, Nunez, & Bergland, *in prep.*).

Adaptive outlier characterization

The XtX statistic identifies outlier loci that exhibit allele frequency changes different from expectations based on demography alone. With the XtX statistic, I found several outlier loci that are candidates of adaptive differentiation across the temporal samples (Figure 6B). I investigated XtX outliers using two different thresholds of significance: the top 0.1% of XtX values in the data (XtX>104.7178; 613 SNPs), and the top 0.1% of SNPs from simulated data (XtX>48.17961; 98,144 SNPs; PODS, see methods; Figure 2). XtX outliers at both thresholds are enriched for previously identified seasonal loci (Machado et al., 2021), when compared to randomly chosen matched control SNPs (Figure 6A). The top 0.1% of XtX values in the data have a median 4.15-fold enrichment for coding loci ($p=7.91x10^{-29}$), a median 2.08-fold enrichment for nonsynonymous loci (p=0.0072), and a median 2.89-fold nonsignificant enrichment for seasonal loci (p=0.235; seasonal threshold of p=0.001). I also see a median 94.1-fold enrichment for tandem repeats ($p=3.57x10^{-65}$). When looking at the less conservative set of XtX outliers, top 0.1% of XtX values in the simulated data, I see a median 1.1-fold

enrichment for coding loci ($p=1.82x10^{-13}$), a median 1.2-fold enrichment for seasonal loci ($p=3.19x10^{-7}$; seasonal threshold of p=0.005), and a 2.1-fold enrichment for tandem repeats ($p=9.36x10^{-34}$). No enrichment for nonsynonymous loci was observed (p=0.225) at the less conservative outlier threshold.

The patterns of enrichment persist for some categories after removing loci that are tandem repeats, though all were reduced. After the removal of tandem repeats, in the XtX outliers in the top 0.1% of XtX values in the data showed no significant enrichments for coding loci (p=0.10), nonsynonymous loci, or seasonal loci (p=0.44; seasonal threshold of p=0.001). In the less conservative set of XtX outliers, top 0.1% of XtX values in the simulated data, I see a slight median 1.08-fold enrichment for coding loci ($p=2.03x10^{-8}$) and a slight median-1.29 fold enrichment for seasonal loci (p=0.0019; seasonal threshold of p=0.001). There was no enrichment of nonsynonymous loci (p=0.68). Given the change in enrichment patterns after the removal of tandem repeats, the enrichment in coding and nonsynonymous loci appear to be driven by variants in tandem repeat-rich genes for the more conservative outlier threshold, though they are not the sole driver of enrichment for the less conservative outlier threshold.

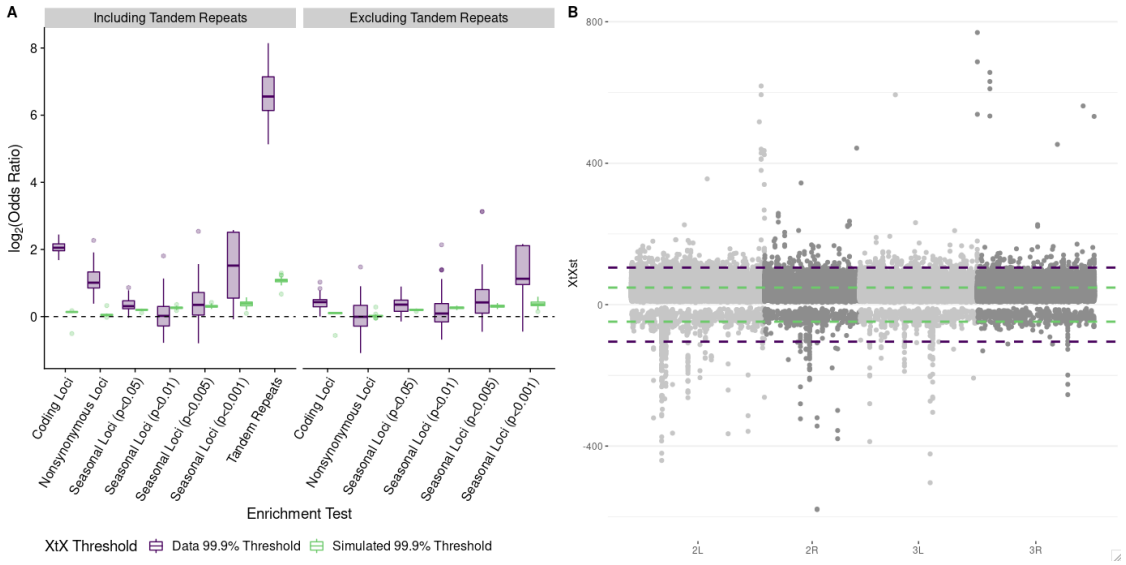


Figure 6: ***XtX outliers within the genome show varying degrees of enrichment of coding, non-synonymous, tandem repeat, and seasonal SNPs.*** A) The $\log_2$ transformed odds ratio of the enrichment of XtX outliers vs. matched control SNPs for a variety of types of loci for two different thresholds. B) Manhattan plot of XtX values, where the points with "negative" XtX values are SNPs within tandem repeat regions. The green dashed line indicates the threshold for POD-based simulated 99.9% threshold of significance. The purple dashed line indicates the top 99.9% of XtX values of the data.

Connecting allele frequency changes to environmental factors

I used BayPass to test associations between shifts in allele frequencies and a range of environmental factors: low temperatures, average temperatures, high temperatures, temperature range, humidity, year of collection, day of the year, fruit variety that the flies were collected off of, as well as the number of extreme hot days and extreme cold days. The tests were conducted using values estimated across 14, 21, and 28 days prior to collection. Initial analysis of the BayPass AUX model and down-sampled permutations showed that most environmental factors show very few associated outlier SNPs (Figure 7A). The average temperature measures, relative humidity, temperature ranges, year, and

date all showed a small proportion of SNPs that are outliers (range of proportions: $1.21x10^{-6}$ to 0.0009). Fruit type and the proportion of extreme hot and extreme cold days showed an order of magnitude greater proportion of outlier SNPs relative to the other factors (range of proportions: 0.004-0.04).

I chose to focus on the measures of the proportion of extreme hot and extreme cold days as fruit type may have been associated with population turnover through time rather than being a response to specific selection pressures by the fruit variety. The 14-, 21-, and 28-day proportion of extreme hot and extreme cold days prior to collection did have a large number of associated outlier SNPs, however permutations also showed an elevated number of outliers (range of proportions: 0.002-0.03). To investigate outlier signals associated with the 14-, 21-, and 28-day proportion of extreme hot and extreme cold days, I performed permutations of the full genome rather than a down-sampled set of SNPs. When comparing the data with the full permutations, I observed that the number of outliers associated with the 14- and 28-day proportion of extreme hot days and 28-day proportion extreme cold days was still beyond the null expectation produced by the permutations (Figure 7B). The 28-day proportion of extreme cold days prior to collection and the 14- and 28-day proportion of extreme hot days prior to collection showed 1.78, 2.21, and 1.46 times the proportion of outlier SNPs beyond the mean of the permutations, respectively. The 14- and 21-day proportion of extreme cold days prior to collection and the 21-day proportion of extreme hot days prior to collection showed that the proportion of outlier SNPs in the data was squarely in the distribution of proportion of outlier SNPs in the permutations.

Figure 7: ***The proportion of hot and cold extreme days are the dominant environmental factors associated with shifts in allele frequencies.*** A) The proportion of SNPs assessed using down-sampled permutations (green) compared to the data (purple) that have a greater Bayes Factor value than the 99.5% percentile of all Bayes Factor values from the data. B) The proportion of SNPs assessed using the full permutations of the proportion of hot and cold extreme days (green) compared to the data (purple) that have a greater Bayes Factor value than the 99.5% percentile of Bayes Factor values from the extreme hot and extreme cold data.

I assessed the dispersion of BayPass outliers throughout the genome using a sliding window analysis. The purpose of the sliding window analysis was to understand how outliers are distributed across the genome and to compare to the same kind of outlier

distribution in the full permutations. I observed clustering throughout the genome in the data when compared to the expectation of an even dispersal of outliers (Figure 8A). I also observed a much larger number of significantly clustered windows in the data than in the permutations (Figure 8B).

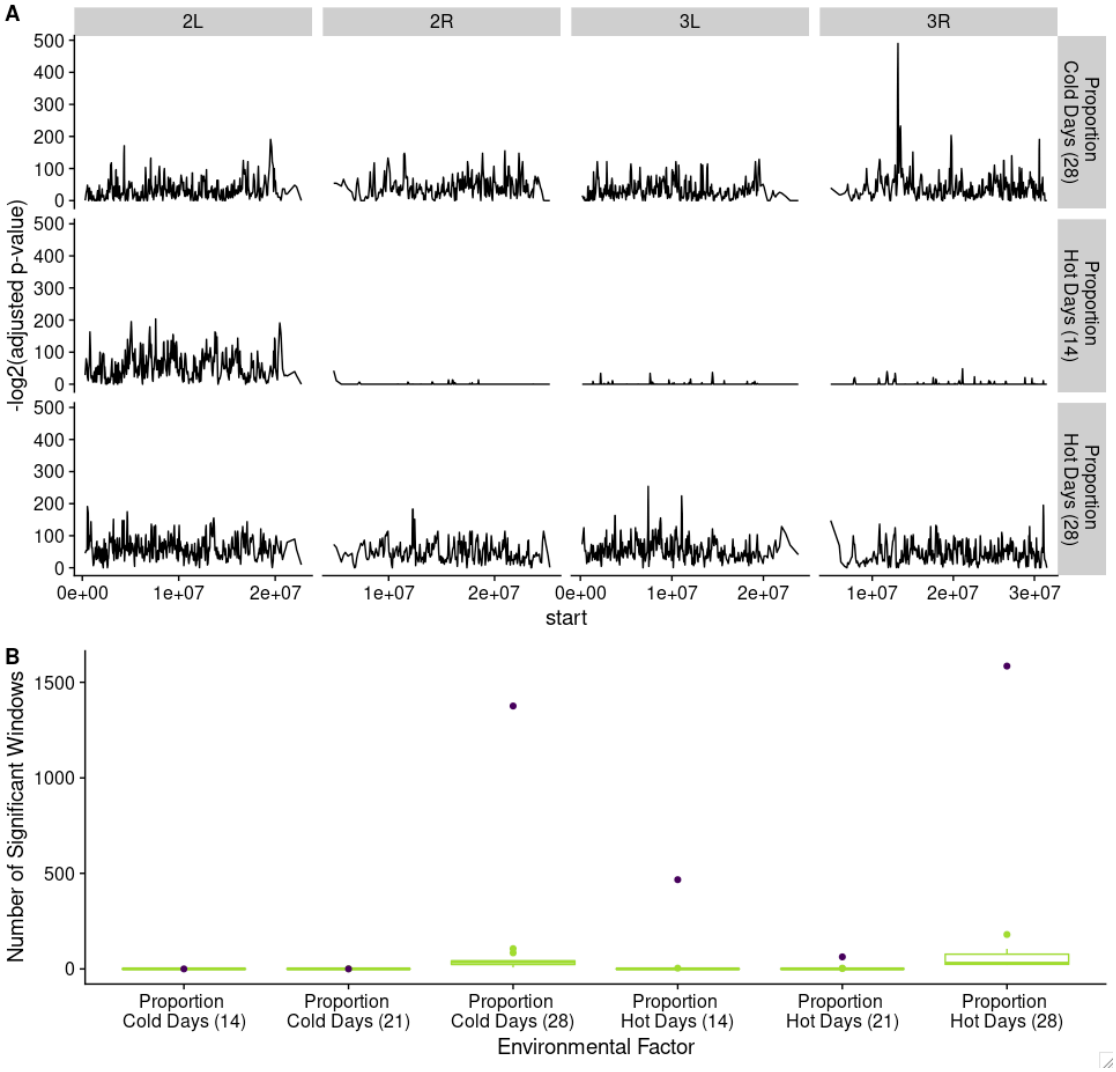


Figure 8: ***The proportion of hot and cold days have many more clustered windows of the genome significantly associated with allele frequency shifts than permutations.*** A) The -$\log_2$(adjusted p-value) for 1000 SNP windows and a 500 SNP step across the autosomes for three environmental factors tested. B) For a p-value threshold of $p<0.001$, the count of the number of windows that deviate from expectation for both the data (purple) and permutations (green).

Discussion

I generated then analyzed a novel, finely-sampled dataset of a focal wild *Drosophila melanogaster* population and outline three findings. Genome-wide patterns of differentiation indicate signals of potential year-to-year population turnover that appear to be a consequence of severe winter bottlenecks. I identify outliers of adaptive differentiation through time that exhibit enrichments for different coding regions, previously identified seasonally oscillating SNPs, and other categories of loci. Finally, I find that the proportion of extreme hot and extreme cold days prior to collection are associated with shifts in allele frequencies through time.

Patterns of differentiation point to population turnover and winter bottlenecks

The degree of differentiation between samples through time can provide insight into the rate of population turnover. Samples cluster by year of collection in the PCA and whole genome pairwise F_{ST} estimates asymptote with time. These results suggest that the focal *D. melanogaster* orchard population experiences population turnover each year. This could be driven by either winter bottlenecks or complete extirpation and recolonization by a genetically different source of local migrants each year. The focal orchard does not have clear places of refugia (e.g. large compost piles that retain warmth through decomposition) where adults could overwinter safely, though little is understood about overwintering behavior in *D. melanogaster*; the focal orchard also is a part of a network of orchards within central Virginia that experience regular transport of fruit and other produce between locations. Observations of the focal orchard population and the

lack of understanding of overwintering behavior do not make it clear if the focal population undergoes winter bottlenecks or is extirpated and recolonized every year.

The simple demographic models we performed imply that the focal population has some mechanism of overwintering and recolonization despite the lack of obvious overwintering sites. I conclude based on the demographic modeling that the signal of population turnover is the result of bottlenecks over the winter rather than complete extirpation and recolonization. More modeling work would be needed to reach conclusions about estimated overwintering population sizes.

Adaptive outlier enriched for coding and previously identified seasonal loci

To assess broad patterns of adaptive differentiation in the samples, I used the XtX statistic to identify outliers that represent loci that are adaptively differentiating through time. The enrichment of coding loci indicates that outliers for adaptive differentiation may be functional, something not certain in previous work assessing seasonal samples of *D. melanogaster* (Bergland et al., 2014; Machado et al., 2021). The enrichment of previously identified seasonally oscillating loci (Machado et al., 2021) suggests that while the XtX statistic is not specifically designed to assess temporal fluctuations in adaptive differentiation, the results are verified and therefore valid. Adaptively differentiating outliers appear to be both functional and globally seasonal, which is a first step in providing evidence that seasonally oscillating SNPs are functional.

Extreme temperatures, not other environmental variables, are associated with temporal shifts in allele frequencies

The observation that the proportion of extreme hot and extreme cold days prior to collection has the strongest association with temporally changing allele frequencies is consistent with the known biology of *D. melanogaster*. For example, thermal extremes predict species distributions in Australian Drosophilids (Overgaard, Kearney, & Hoffmann, 2014) and models show that thermal extremes are more influential on the evolution of thermal performance curves than thermal means in temperate Drosophila populations (Buckley & Huey, 2016). Additionally, average air temperatures are not always a good proxy for either body or microhabitat temperature (Bakken, 1992; Helmuth et al., 2010; Sears, Raskin, & Angilletta, 2011) which may explain why they do not appear to be drivers of temporally changing allele frequencies. Temperature extremes being more associated with changing allele frequencies than temperature averages confirm expectations based on previous work.

The different time periods across which I estimated the environmental factors provide potential insight into the developmental biology of wild *D. melanogaster*. We see that the 14- and 28-day proportion of hot extremes prior to collection as well as the 28-day proportion of extreme cold days prior to collection drive the most outlier SNPs, relative to null expectations. At colder temperatures, *D. melanogaster* develops more slowly (Powsner, 1935). Therefore, it is reasonable that a 28-day estimation of the proportion of extreme cold days is a dominant environmental selective pressure during the colder months. Conversely, at warmer, but non-lethal, temperatures *D. melanogaster* develops more quickly (Powsner, 1935), therefore the 14-day proportion of extreme hot

days being dominant is reasonable as the typical egg-to-adult life cycle for *D. melanogaster* is 10 days. The less intuitive dominance of the 28-day proportion of extreme hot days prior to collection may be explained by an additive effect of two generations in a row of directional selection as 28 days can capture at least two generations of adaptation if there is a long period of extreme heat. While we know that temperature impacts developmental timing (Powsner, 1935) and the factors most associated with temporal changes in allele frequencies seem to reflect that, more experimental methods may be needed to make this link more explicit.

While fruit type did appear to be an environmental factor with a strong association with temporally changing allele frequencies, fruit type might be a proxy for evolution across the year. Sample collections always began on peaches and then transitioned to apples mid-season. However, a closer investigation of the fruit variety from which flies are collected may be of interest as peaches and apples clearly have a different makeup, and different varieties of apples have different sugar content (Hecke et al., 2006) and sugar consumption can impact several fitness related traits in *D. melanogaster* (Chng, Hietakangas, & Lemaitre, 2017). During sample collection, I also observed that fruit flies do not colonize certain apple varieties. Fruit type may be an important selective factor, but the way the data was collected cannot accurately determine the relative contribution of fruit type to seasonal adaptation.

Caveats and future lines of inquiry

There are some caveats to the interpretation and analysis of the data with BayPass. Based on the observations of the raw data, I speculate that because the

proportion of days with temperature extremes had many zeroes, I am limited in the interpretation of the specific identities of outliers as the outlier signals appeared somewhat artificially inflated. Additionally, while there is a history of substituting space for time (Blois, Williams, Fitzpatrick, Jackson, & Ferrier, 2013; Srivathsa, Puri, Kumar, Jathanna, & Karanth, 2018), BayPass is not explicitly built to assess time-series data. I admit this is a limitation, but I do not think that it takes away from the broad conclusions as I looked at general patterns rather than specific outlier loci. As a final caveat to the data analysis with BayPass, I did not assess any biotic factors. Particularly, we know that the Drosophilid species distributions change within years and between years (Gleason, Roy, Everman, Gleason, & Morgan, 2019) and that variation in the microbiomes of *D. melanogaster* may be an agent of selection (Rudman et al., 2019). Given the broad interpretation of the BayPass analysis and the use of permutations, I posit that the conclusions are still robust.

This dataset is a unique tool to investigate the details of seasonal evolution. Such fine-scale sampling and sequencing of a single population of Drosophila has not been previously done. Global efforts to capture spring and fall samples of wild *D. melanogaster* populations (Kapun et al., *in prep*.) have shown some limitations derived from how fluctuations in the environment prior to collections may have a greater impact on the seasonal patterns observed (Machado et al., 2021). Fine temporal sampling provides a closer perspective of environmental selective pressures that can be tied to and investigated in global seasonal collections (Kapun et al., *in prep*.). The sampling effort of wild-caught Drosophilids described exceeds the sequencing effort described. Further uses of the sampling can be tracking genotypic changes rather than pooled allele frequency

estimates, sequencing individuals at the end and start of seasons to more closely investigate overwintering dynamics, and investigate the dynamics of other Drosophilid species or Drosophila commensals. Also, the focal orchard is part of a network of nearby orchards and vineyards and perhaps spatial sampling can provide insight into migration dynamics on a local scale tied with changes in allele frequencies in the focal population. With the dataset generated here, the sampling conducted, and the focal orchard connected to a network of orchards, I have found an ideal population in which to study many questions related to balancing selection in response to seasonally oscillating selection.

Conclusion

Overall, I have laid out evidence to support three findings. First, the patterns of genome-wide differentiation through time indicate year-to-year population turnover. Through straightforward demographic models of winter bottlenecks, we can generate patterns of clustering by year that mimic what we observe in the natural population suggesting that winter bottlenecks are sufficient to generate year-to-year population turnover. Second, the outliers of adaptive differentiation are enriched for coding regions and previously identified seasonally oscillating SNPs, which adds more confirmation that seasonally varying selection is repeatable globally and acts through biologically relevant loci. Third, the proportion of extreme hot and cold days prior to collections are the dominant environmental factors that are associated with temporally changing allele frequencies. In particular, the days across which the proportions were taken (14- and 28-day windows) align with expectations according to developmental timing. This shows that thermal extremes play an important role in seasonal adaptation in addition to limiting

broad patterns of species distributions. With these three claims, I have extended our understanding of seasonal evolution in *D. melanogaster*, thus allowing us to better understand this model system of balancing selection through adaptively tracking seasonal change.

Data Availability

All scripts can be found at https://github.com/abangerter/3yr_Dmel_poolseq

All raw sequencing data can be found under NCBI SRA BioProject ID PRJNA728438. Data is embargoed until peer-reviewed publication.

Tables

Table 1: ***Metadata for the samples.***

Sample ID	Collection Date (MM/DD/YY)	SRA #	Fruit Collected From	Number of Sequenced Flies
TYS_2016_0624	6/24/16	SRR14475045	peaches	37
TYS_2016_0708	7/8/16	SRR14475044	peaches	36
TYS_2016_0722	7/22/16	SRR14475033	peaches	22
TYS_2016_0819	8/19/16	SRR14475022	peaches	23
TYS_2016_0902	9/2/16	SRR14475014	peaches & apples	146
TYS_2016_0916	9/16/16	SRR14475013	apples	113
TYS_2016_1003	10/3/16	SRR14475012	apples	40
TYS_2016_1014	10/14/16	SRR14475011	apples	150
TYS_2016_1028	10/28/16	SRR14475010	apples	118
TYS_2016_1111	11/11/16	SRR14475009	apples	86

TYS_2016_1203	12/3/16	SRR14475043	apples	70
TYS_2017_0622	6/22/17	SRR14475042	peaches	39
TYS_2017_0707	7/7/17	SRR14475041	peaches	45
TYS_2017_0720	7/20/17	SRR14475040	peaches	69
TYS_2017_0803	8/3/17	SRR14475039	peaches & apples	29
TYS_2017_0817	8/17/17	SRR14475038	peaches & apples	32
TYS_2017_0831	8/31/17	SRR14475037	peaches & apples	87
TYS_2017_0914	9/14/17	SRR14475036	apples	92
TYS_2017_0928	9/28/17	SRR14475035	apples	136
TYS_2017_1012	10/12/17	SRR14475034	apples	107
TYS_2017_1026	10/26/17	SRR14475032	apples	120
TYS_2017_1109	11/9/17	SRR14475031	apples	122
TYS_2017_1124	11/24/17	SRR14475030	apples	21
TYS_2017_1207	12/7/17	SRR14475029	apples	24
TYS_2018_0628	6/28/18	SRR14475028	peaches	10
TYS_2018_0705	7/5/18	SRR14475027	peaches	26
TYS_2018_0712	7/12/18	SRR14475026	peaches	58
TYS_2018_0726	7/26/18	SRR14475025	peaches	70
TYS_2018_0809	8/9/18	SRR14475024	peaches	84
TYS_2018_0823	8/23/18	SRR14475023	peaches & apples	69
TYS_2018_0906	9/6/18	SRR14475021	peaches & apples	60
TYS_2018_0920	9/20/18	SRR14475020	apples	77
TYS_2018_1004	10/4/18	SRR14475019	apples	108
TYS_2018_1018	10/18/18	SRR14475018	apples	136
TYS_2018_1025	10/25/18	SRR14475017	apples	122
TYS_2018_1101	11/1/18	SRR14475016	apples	103
TYS_2018_1129	11/29/18	SRR14475015	apples	16

Literature Cited

Bakken, G. S. (1992). Measurement and Application of Operative and Standard Operative Temperatures in Ecology. *American Zoologist*, *32*(2), 194–216.

Bangerter, A., Nunez, J., Bergland, A. O. (2021). Dense temporal sampling identifies yearly population turnover and confirms that temperature is an important driver of seasonal adaptive tracking in *D. melanogaster*. *Manuscript in preparation*.

Behrman, E L, Watson, S. S., O'Brien, K. R., Heschel, M. S., & Schmidt, P. S. (2015). Seasonal variation in life history traits in two *Drosophila* species. *Journal of Evolutionary Biology*, *28*(9), 1691–1704.

Behrman, Emily L., Howick, V. M., Kapun, M., Staubach, F., Bergland, A. O., Petrov, D. A., Lazzaro, B. P., et al. (2018). Rapid seasonal evolution in innate immunity of wild *Drosophila melanogaster*. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1870), 20172599.

Bergland, A. O., Behrman, E. L., O'Brien, K. R., Schmidt, P. S., & Petrov, D. A. (2014). Genomic Evidence of Rapid and Stable Adaptive Oscillations over Seasonal Time Scales in *Drosophila*. (D. Bolnick, Ed.)*PLoS Genetics*, *10*(11), e1004775-19.

Bertram, J., & Masel, J. (2019). Different mechanisms drive the maintenance of polymorphism at loci subject to strong versus weak fluctuating selection. *Evolution*, *101*, 3490–14.

Blois, J. L., Williams, J. W., Fitzpatrick, M. C., Jackson, S. T., & Ferrier, S. (2013). Space can substitute for time in predicting climate-change effects on biodiversity. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(23), 9374–9379.

Bogaerts-Márquez, M., Guirao-Rico, S., Gautier, M., & González, J. (2020). Temperature, rainfall and wind variables underlie environmental adaptation in natural populations of *Drosophila melanogaster*. *Molecular Ecology*.

Botero, C. A., Weissing, F. J., Wright, J., & Rubenstein, D. R. (2015). Evolutionary tipping points in the capacity to adapt to environmental change. *Proceedings of the National Academy of Sciences*, *112*(1), 184–189.

Buckley, L. B., & Huey, R. B. (2016). How Extreme Temperatures Impact Organisms and the Evolution of their Thermal Tolerance. *Integrative and Comparative Biology*, *56*(1), 98–109.

Chakraborty, M., & Fry, J. D. (2016). Evidence that Environmental Heterogeneity Maintains a Detoxifying Enzyme Polymorphism in *Drosophila melanogaster*. *Current Biology*, *26*(2), 219–223.

Chng, W. A., Hietakangas, V., & Lemaitre, B. (2017). Physiological Adaptations to Sugar Intake: New Paradigms from *Drosophila melanogaster*. *Trends in Endocrinology & Metabolism*, *28*(2), 131–142.

Chown, S. L., Jumbam, K. R., Sørensen, J. G., & Terblanche, J. S. (2009). Phenotypic variance, plasticity and heritability estimates of critical thermal limits depend on methodological context. *Functional Ecology*, *23*(1), 133–140.

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., et al. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly*, *6*(2), 80–92.

Cogni, R, Kuczynski, K., Lavington, E., Koury, S., Behrman, E. L., O'Brien, K. R., Schmidt, P. S., et al. (2014). Variation in *Drosophila melanogaster* central metabolic genes appears driven by natural selection both within and between populations. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1800), 20142688–20142688.

Cogni, Rodrigo, Kuczynski, C., Koury, S., Lavington, E., Behrman, E. L., O'Brien, K. R., Schmidt, P. S., et al. (2013). The Intensity of Selection Acting on the Couch Potatogene-Spatial-Temporal Variation in a Diapause Cline. *Evolution*, *68*(2), 538–548.

Comeron, J. M., Ratnappan, R., & Bailin, S. (2012). The Many Landscapes of Recombination in *Drosophila melanogaster*. (D. A. Petrov, Ed.)*PLoS Genetics*, *8*(10), e1002905-21.

Coop, G., Witonsky, D., Rienzo, A. D., & Pritchard, J. K. (2010). Using Environmental Correlations to Identify Loci Underlying Local Adaptation. *Genetics*, *185*(4), 1411–1423.

Corbett-Detig, R. B., & Hartl, D. L. (2012). Population Genomics of Inversion Polymorphisms in *Drosophila melanogaster*. (H. S. Malik, Ed.)*PLoS Genetics*, *8*(12), e1003056-15.

Cvijović, I., Good, B. H., Jerison, E. R., & Desai, M. M. (2015). Fate of a mutation in a fluctuating environment. *Proceedings of the National Academy of Sciences*, *112*(36), E5021–E5028.

Dobzhansky, T. (1948). Genetics of Natural Populations. XVI. Altitudinal and Seasonal Changes Produced by Natural Selection in Certain Populations of *Drosophila pseudoobscura* and *Drosophila persimilis*. *Genetics*, *33*, 158–176.

Ewing, E. P. (1979). Genetic Variation in a Heterogeneous Environment VII. Temporal and Spatial Heterogeneity in Infinite Populations. *The American Naturalist*, *114*(2), 197–212.

Frichot, E., & François, O. (2015). LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution*, *6*(8), 925–929.

Frichot, É., Schoville, S. D., Bouchard, G., & François, O. (2013). Testing for Associations between Loci and Environmental Gradients Using Latent Factor Mixed Models. *Molecular Biology and Evolution*, *30*(7), 1687–1699.

Garcia, M. J., Littler, A. S., Sriram, A., & Teets, N. M. (2020). Distinct cold tolerance traits independently vary across genotypes in *Drosophila melanogaster*. *Evolution*, *331*, 192–14.

Gautier, M. (2015). Genome-Wide Scan for Adaptive Divergence and Association with Population-Specific Covariates. *Genetics*, *201*, 1555–1579.

Ghosh, R., Andersen, E. C., Shapiro, J. A., Gerke, J. P., & Kruglyak, L. (2012). Natural Variation in a Chloride Channel Subunit Confers Avermectin Resistance in *C. elegans*. *Science*, *335*(6068), 574–578.

Gibbs, H. L., & Grant, P. R. (1987). Oscillating selection on Darwin's finches. *Nature*, *327*(6122), 511–513.

Gillespie, J. (1973). Polymorphism in random environments. *Theoretical Population Biology*, *4*(2), 193–195.

Gleason, J. M., Roy, P. R., Everman, E. R., Gleason, T. C., & Morgan, T. J. (2019). Phenology of *Drosophila* species across a temperate growing season and implications for behavior. (N. Desneux, Ed.)*PLOS ONE*, *14*(5), e0216601-21.

Gompert, Z. (2015). Bayesian inference of selection in a heterogeneous environment from genetic time-series data. *Molecular Ecology*, *25*(1), 121–134.

Grant, B. R., & Grant, P. R. (1993). Evolution of Darwin's finches caused by a rare climatic event. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *251*(1331), 111–117.

Grant, P. R., & Grant, B. R. (2002). Unpredictable Evolution in a 30-Year Study of Darwin's Finches. *Science*, *296*(5568), 707–711.

Günther, T., & Coop, G. (2013). Robust Identification of Local Adaptation from Allele Frequencies. *Genetics*, *195*(1), 205–220.

Haller, B. C., & Messer, P. W. (2019). SLiM 3: Forward Genetic Simulations Beyond the Wright–Fisher Model. *Molecular Biology and Evolution*, *36*(3), 632–637.

Hecke, K., Herbinger, K., Veberič, R., Trobec, M., Toplak, H., Štampar, F., Keppel, H., et al. (2006). Sugar-, acid- and phenol contents in apple cultivars from organic and integrated fruit cultivation. *European Journal of Clinical Nutrition*, *60*(9), 1136–1140.

Hedrick, P. W., Ginevan, M. E., & Ewing, E. P. (1976). Genetic Polymorphism in Heterogeneous Environments. *Annual Review of Ecology and Systematics*, *7*, 1–32.

Helmuth, B., Broitman, B. R., Yamane, L., Gilman, S. E., Mach, K., Mislan, K. A. S., & Denny, M. W. (2010). Organismal climatology: analyzing environmental variability at scales relevant to physiological stress. *Journal of Experimental Biology*, *213*(6), 995–1003.

Hivert, V., Leblois, R., Petit, E. J., Gautier, M., & Vitalis, R. (2018). Measuring Genetic Differentiation from Pool-seq Data. *Genetics*, genetics.300900.2018-51.

Holm, S. (1979). A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics*, *6*(2), 65–70.

Hoskins, R. A., Carlson, J. W., Wan, K. H., Park, S., Mendez, I., Galle, S. E., Booth, B. W., et al. (2015). The Release 6 reference sequence of the *Drosophila melanogaster* genome. *Genome Research*, *25*(3), 445–458.

Hu, T. T., Eisen, M. B., Thornton, K. R., & Andolfatto, P. (2013). A second-generation assembly of the *Drosophila simulans* genome provides new insights into patterns of lineage-specific divergence. *Genome Research*, *23*(1), 89–98.

Kapun, M., Nunez, J. C., Bogaerts-Márquez, M., Murga-Moreno, J., Paris, M., Outen, J., Coronado-Zamora, M., et al. *Drosophila* Evolution in Space and Time (DEST) - A New Population Genomics Resource. *Manuscript in preparation*.

Koboldt, D. C., Zhang, Q., Larson, D. E., Shen, D., McLellan, M. D., Lin, L., Miller, C. A., et al. (2012). VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Research*, *22*(3), 568–576.

Kolaczkowski, B., Kern, A. D., Holloway, A. K., & Begun, D. J. (2011). Genomic differentiation between temperate and tropical Australian populations of *Drosophila melanogaster*. *Genetics*, *187*(1), 245–260.

Lecheta, M. C. (2020). Integrating GWAS and Transcriptomics to Identify the Molecular Underpinnings of Thermal Stress Responses in *Drosophila melanogaster*. *Frontiers in Genetics*, *11*, 1–17.

Li, H, & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, *26*(5), 589–595.

Li, Heng. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM, 1–3.

Machado, H., Bergland, A. O., Taylor, R., Tilk, S., Behrman, E., Dyer, K., Fabian, D., et al. (2021). Broad geographic sampling reveals the shared basis and environmental correlates of seasonal adaptation in *Drosophila. eLife*, *10*(e67577), 1–21.

Machado, H. E., Bergland, A. O., O'Brien, K. R., Behrman, E. L., Schmidt, P. S., & Petrov, D. A. (2016). Comparative population genomics of latitudinal variation in *Drosophila simulans* and *Drosophila melanogaster*. *Molecular Ecology*, *25*(3), 723–740.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, *17*(1), 10–12.

Mojica, J. P., Lee, Y. W., Willis, J. H., & Kelly, J. K. (2012). Spatially and temporally varying selection on intrapopulation quantitative trait loci for a life history trade-off in *Mimulus guttatus*. *Molecular Ecology*, *21*(15), 3718–3728.

Nosil, P., Villoutreix, R., Carvalho, C. F. de, Farkas, T. E., Soria-Carrasco, V., Feder, J. L., Crespi, B. J., et al. (2018). Natural selection and the predictability of evolution in *Timema* stick insects. *Science*, *359*(6377), 765–770.

Overgaard, J., Hoffmann, A. A., & Kristensen, T. N. (2011). Assessing population and environmental effects on thermal resistance in *Drosophila melanogaster* using ecologically relevant assays. *Journal of Thermal Biology*, *36*(7), 409–416.

Overgaard, J., Kearney, M. R., & Hoffmann, A. A. (2014). Sensitivity to thermal extremes in Australian *Drosophila* implies similar impacts of climate change on the distribution of widespread and tropical species. *Global Change Biology*, *20*(6), 1738–1750.

Paaby, A. B., Bergland, A. O., Behrman, E. L., & Schmidt, P. S. (2014). A highly pleiotropic amino acid polymorphism in the *Drosophila* insulin receptor contributes to life-history adaptation. *Evolution*, *68*(12), 3395–3409.

Petavy, G., David, J. R., Gibert, P., & Moreteau, B. (2001). Viability and rate of development at different temperatures in *Drosophila*: a comparison of constant and alternating thermal regimes. *Journal of Thermal Biology*, *26*, 29–39.

Powsner, L. (1935). The Effects of Temperature on the Durations of the Developmental Stages of *Drosophila melanogaster*. *Physiological Zoology*, *8*(4), 474–520.

Price, T. D., Grant, P. R., Gibbs, H. L., & Boag, P. T. (1984). Recurrent patterns of natural selection in a population of Darwin's finches. *Nature*, *309*(5971), 787–789.

Rajpurohit, S., Hanus, R., Vrkoslav, V., Behrman, E. L., Bergland, A. O., Petrov, D., Cvačka, J., et al. (2017). Adaptive dynamics of cuticular hydrocarbons in *Drosophila*. *Journal of Evolutionary Biology*, *30*(1), 66–80.

Rajpurohit, Subhash, Gefen, E., Bergland, A. O., Petrov, D. A., Gibbs, A. G., & Schmidt, P. S. (2018). Spatiotemporal dynamics and genome-wide association analysis of desiccation tolerance in *Drosophila melanogaster*. *Molecular Ecology*, *27*(17), 3525–3540.

Ransberry, V. E., MacMillan, H. A., & Sinclair, B. J. (2011). The Relationship between Chill-Coma Onset and Recovery at the Extremes of the Thermal Window of *Drosophila melanogaster*. *Physiological and Biochemical Zoology*, *84*(6), 553–559.

Rodríguez-Trelles, F., Alvarez, G., & Zapata, C. (1996). Time-Series Analysis of Seasonal Changes of the O Inversion Polymorphism of *Drosophila subobscura*. *Genetics*, *142*, 179–187.

Rodriguez-Trelles, F., Tarrio, R., & Santos, M. (2013). Genome-wide evolutionary response to a heat wave in *Drosophila*. *Biology Letters*, *9*(4), 20130228–20130228.

Rudman, S. M., Greenblum, S., Hughes, R. C., Rajpurohit, S., Kiratli, O., Lowder, D. B., Lemmon, S. G., et al. (2019). Microbiome composition shapes rapid genomic adaptation of *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences*, *116*(40), 20025–20032.

Schmidt, P. S., & Conde, D. R. (2006). Environmental Heterogeneity and the Maintenance of Genetic Variation for Reproductive Diapause in *Drosophila melanogaster*. *Evolution*, *60*(8), 1602–1611.

Sears, M. W., Raskin, E., & Angilletta, M. J. (2011). The World Is not Flat: Defining Relevant Thermal Landscapes in the Context of Climate Change. *Integrative and Comparative Biology*, *51*(5), 666–675.

Srivathsa, A., Puri, M., Kumar, N. S., Jathanna, D., & Karanth, K. U. (2018). Substituting space for time: Empirical evaluation of spatial replication as a surrogate for temporal replication in occupancy modelling. *Journal of Applied Ecology*, *55*(2), 754–765.

Wittmann, M. J., Bergland, A. O., Feldman, M. W., Schmidt, P. S., & Petrov, D. A. (2017). Seasonally fluctuating selection can maintain polymorphism at many loci via segregation lift. *Proceedings of the National Academy of Sciences*, *54*, 201702994–10.

CHAPTER TWO

Individual whole genome sequencing of a single orchard population of *D. melanogaster* shows stochastic population structure throughout a single year

Abstract

The degree of population structure, in the form of genetic differentiation between subpopulations or isolation by distance, can impact the potential for local adaptation to occur. In many species, the capability of long-distance dispersal leads to assumptions that only large-scale patterns like latitudinal clines will emerge and that local population structure will not be observed. Local panmixia has been a long-standing assumption in *Drosophila melanogaster.* However, there is evidence that dispersal rates in *D. melanogaster* may be resource-dependent, leading to questions about whether population structure can arise in a well-resourced, natural orchard population. I ask: does a temperate orchard population of *D. melanogaster* exhibit signs of population structure, and does any observed structure vary through time? Using bi-weekly sampling from a local fruit orchard, I performed whole-genome sequencing of individual male offspring of wild-caught isofemale lines. I observed an excess of homozygosity, as measured by F_{IS}, relative to simulated expectations. Positive F_{IS} beyond the simulated expectation was observed genome-wide in a subset of time points. The temporally inconsistent elevated genome-wide positive F_{IS} is indicative of stochastic shifts between cryptic population structure and panmixia in the focal orchard population. The presence of stochastic population structure in wild *D. melanogaster* populations has implications for the understanding of evolution in short-lived, mobile organisms and provides a rich model system in which to study rapidly changing metapopulation dynamics.

Introduction

In nature, population substructure, population size, and migration rates of many organisms are dynamic through time. The short-term demographics of a population can impact patterns of genetic diversity and therefore differentiation (Ellegren & Galtier, 2016). In addition, many populations experience spatial structuring into smaller clusters of individuals called subpopulations. Structuring of a population is considered a metapopulation if it includes dynamics of migration between subpopulations as well as extinction and recolonization of subpopulations (Hanski, 1994, 1999). Metapopulation structure can also impact measures of genetic diversity and differentiation while providing important insight into the evolution of natural populations (Pannell & Charlesworth, 2000).

For metapopulations where migration rates are high, it is expected that genetic differentiation between subpopulations (hereafter referred to as population structure) will decline due to high gene flow. In metapopulations, the degree of observed population structure depends on the degree of migration. For example, the degree of genetic differentiation between subpopulations in the butterfly *Parnassius mnemosyne* is dependent upon the connectivity of patches, determined by the spatial distribution of habitat types. Higher connectivity between the subpopulation patches leads to more migration that results in measurable gene flow and a lowered measure of genetic differentiation between patches (Keyghobadi, Roland, & Strobeck, 2005). In the moth *Cydia pomonella*, genetic differentiation between subpopulations found in different orchards within the same valley is minimal, but at a regional scale genetic differentiation between subpopulations is observed, consistent with understanding of flight capability of

the moths (Chen & Dorn, 2010). Dispersal events in bed bugs are mediated through human movement resulting in strong founder effects in newly recolonized infestations. These strong founder effects result in different infestations, or subpopulations, exhibiting high levels of differentiation (Fountain, Duvaux, Horsburgh, Reinhardt, & Butlin, 2014). While some species do exhibit clear, observable spatial delineation between subpopulations with measurable population structure between them, there are other instances where the existence of subpopulations is not clear.

Natural populations where there are no apparent spatial divisions may still have genetically distinct subpopulations, also known as cryptic population structure. One way to identify cryptic population structure is by looking for genome-wide positive F_{IS}. F_{IS}, the inbreeding coefficient, is often calculated using the number of observed heterozygotes versus the expected number of heterozygotes based on allele frequencies. More positive values of F_{IS} point to a more inbred population. Due to both selection and neutral processes, allele frequencies will be different between subpopulations when gene flow is not high. When two or more subpopulations are collected and treated as a single population (i.e. cryptic population structure), genome-wide positive F_{IS} would be produced from an excess of homozygotes. This excess of homozygotes would strongly deviate from Hardy-Weinberg as there would be fewer heterozygotes than expected based on combined allele frequencies. Positive F_{IS} because of cryptic population structure is also called a Wahlund effect (Wahlund, 1928). Observing a Wahlund effect in samples from natural populations can signal the presence of cryptic population structure.

There is a broad assumption that the fruit fly *Drosophila melanogaster* is a panmictic species on a moderately local level (e.g. an orchard) with no fine-scale

population structure, but this assumption might be flawed. The assumptions of general panmixia in *D. melanogaster* are based on observations of dispersal up to 25 km (Coyne & Milstead, 1987), but migration distances may be fewer than 5 m in the presence of readily available food (Wallace, 1970). Other Drosophilids support resource-dependent migration rates: where resources are abundant, flies remain on those local food sources over their lifetime, but where food resources are lacking, flies tend to migrate at high rates (Coyne & Milstead, 1987; Dubinin & Tiniakov, 1946; Johnston & Heed, 1975; Markow & Castrezana, 2000). These observations on differing migration rates could imply that after reaching adulthood, *D. melanogaster* individuals rarely leave that microhabitat if resources are plenty. Fruit fly offspring eclosing from a single fruit likely descend from two-three females ovipositing successfully on the fruit (Hoffmann & Nielson, 1985; Jaenike & Selander, 1979). There is a possibility that in a well-resourced environment *D. melanogaster* may not disperse far before reproduction, and given the localized nature of oviposition behavior, the flies in one region that then mate with each other may be related to one another. Therefore, despite long-held assumptions, population structure may exist within a single fruit orchard as a result of local inbreeding of close relatives on small patches of fruit.

Resource abundance changes throughout a growing season in a fruit orchard, potentially altering dispersal and population structure over time within a single orchard. A metapopulation model was developed (Shpak, Wakeley, Garrigan, & Lewontin, 2010) based upon seasonal observations of a single *D. melanogaster* population made by Ives and colleagues (Band & Ives, 1961, 1963; Ives, 1945, 1954, 1970; Ives & Band, 1986). This model posits that metapopulation structure and dynamics would change with

seasonal change (Shpak, Wakeley, Garrigan, & Lewontin, 2010). The model begins with subpopulations, also called demes, recolonized by few individuals in the spring, that reproduce and grow in size and migration rates between demes increase until they overlap and the metapopulation collapses into a single panmictic population. However, aside from the initial studies done by Ives and colleagues, we have few natural observations of how population structure may change with seasonal change in *D. melanogaster*. With limited and contradictory evidence for migration distances and untested models of how population structure in *D. melanogaster* may vary seasonally, this presents a set of questions ripe for the harvest.

Looking for cryptic, seasonally variable population structure in wild *D. melanogaster* populations would be impossible to do with available data despite having years of seasonal samples (Machado et al., 2021) and broad global sampling (Huang et al., 2014; Lack, Lange, Tang, Corbett-Detig, & Pool, 2016; Mackay et al., 2012). The available seasonal sampling of wild *D. melanogaster* is in the form of pooled sequencing data that offers allele frequency estimates of seasonal end-points, however, genotypes are needed to ask questions of fine-scale population structure. Here, I conduct biweekly sampling across a single fruit orchard and then perform whole genome sequencing on the offspring of sampled individuals. With these data I ask: (1) Do we observe cryptic population structure within a single temperate orchard? (2) If so, does that population structure begin to collapse over time? I observe positive F_{IS} greater than the null expectations. In some cases, the signal of positive F_{IS} is elevated genome wide, indicating stochastic population structure rather than a collapse in structure over time.

Methods

Sample Collection

In 2016, I collected samples of assorted Drosophilids on an approximately bi-weekly basis from Carter Mountain, a peach and apple orchard near Charlottesville, VA, USA (37.99°N, 78.47°W). Collections began June 24th, the start of peach season, and continued until December 16th, far past the end of apple season. See Table 1 for a summary of collections. Flies were caught across a variety of fruit types (early to late season peaches and an assortment of apple varieties) as different sections of the orchards opened for harvest in a scattered manner across the orchard.

Upon collection, wild-caught samples were deposited into vials containing cornmeal-molasses food in the field. Vials containing the wild-caught flies were brought back to the lab and held for about 24 hours to increase the probability that all females had mated. Wild-caught females that appeared to be either *Drosophila melanogaster* or *Drosophila simulans* were used to set isofemale lines. Male offspring of isofemale lines were used to distinguish between *D. melanogaster* and *D. simulans*. Male and female offspring of the *D. melanogaster* isofemale lines were placed in 70% ethanol and stored at -20°C until sequencing.

Table 1: ***Summary of all the collections done in 2016 and the number of D. melanogaster isofemale lines set.*** The number of individuals in the analysis dataset is the count of sequenced individuals kept after filtering of poor-quality individuals was complete. I sequenced 266 individuals and retained 119 individuals after filtering.

Collection Date (2016)	Number of Isofemale Lines Set	Number Individuals sequenced	Number Individuals in analysis dataset
June 24th	37	25	17
July 8th	39	25	8

July 22nd	23	23	8
August 9th	6	0	0
August 19th	23	22	13
September 2nd	147	24	9
September 16th	115	24	14
October 3rd	41	25	12
October 14th	157	24	9
October 28th	120	24	8
November 11th	91	23	11
December 3rd	70	25	10
December 16th	2	2	0

Sample Preparation & Sequencing

DNA was extracted from individual male offspring from a random subset of isofemale lines for every collection time point using an Agencourt DNAdvance kit (Beckman-Coulter A48705). Sequencing libraries were constructed using a scaled-down Nextera kit (Illumina FC-131-1024; Baym et al., 2015) through the help of a liquid handling robot. Unique index sequences were generated (Meyer & Kircher, 2010) so that each library was tagged with a single unique index. Libraries of the individual flies were pooled into a single sequencing library which was sequenced on a full flowcell of Illumina HiSeqX with a target of 10X coverage for each sample library.

Dataset Filtering & Preparation

Prior to alignment, index sequences were trimmed using trimmomatic v0.36 (Bolger, Lohse, & Usadel, 2014) followed by merging of overlapping reads with PEAR version 0.9.11 (Zhang et al., 2014). Sequencing reads were aligned to a dm6 *Drosophila melanogaster* reference genome using BWA-mem (Li & Durbin, 2010; Li, 2013). Following alignment, PCR duplicates were marked using Picardtools version 2.20.4

(http://broadinstitute.github.io/picard). GATK v3.8-0 was used to call variants (McKenna et al., 2010). *HaplotypeCaller* generated GVCF files for each individual. *CombineGVCF* followed by *GenotypeGVCF* generated a VCF file of all individuals. A random subset of DGRP SNPs (Mackay et al., 2012) were used to calibrate the SNP calls with *VariantRecalibrator* and *ApplyRecalibration*.

Following mapping, I used a rigorous and conservative set of filtering steps to generate a set of SNPs with high confidence. Initial filtering of sites in the VCF file included removing SNPs that did not pass quality filters, removal of indels, and removing any SNPs +/-50 base pairs of any identified indels. Repetitive DNA regions were removed using regions identified by RepeatMasker (Smit, Hubley, & Green, 2013; RepeatMasker at http://repeatmasker.org). Low read depth individuals were eliminated from the dataset if the median read depth of every autosome for that individual was less than 4. Sites were eliminated from the dataset if they were in the bottom or top 1.5% quantile of total site read depth across all samples. Then, each individual sample was assessed to identify SNPs in the upper and lower 1% quantile of read depth for that individual. SNPs were eliminated if more than 20 samples at that site failed the individual-based 1% quantile threshold. Any other SNPs above or below the individual 1% read depth cut-off were turned into missing data for the individual that failed the threshold cut-off. Structural variants were detected in samples using Manta v1.4.0 (Chen et al., 2016). All deletions, insertions, and duplications that passed internal Manta scoring were accumulated into a list of structural variants used for filtering. Structural variants that were at or below 10000 bp in length or were identified in at least 15 individual samples were used to filter SNPs from the dataset. Any SNPs within +/- 50bp of the start

and stop positions of these structural variants were removed from the dataset using R package valR (Riemondy et al., 2017) and vcftools v0.1.15 (Danecek et al., 2011).

Finally, individuals were removed from the dataset if they exhibited unusual patterns relative to the rest of the data. Two individuals were identified as females due to having roughly equal coverage on the X and autosomes, and therefore were removed from the dataset (Figure 1A). Another individual had roughly 1.5X the level of heterozygosity than all the other samples relative to read depth and I speculate that the sample was contaminated with an unintended second fly during DNA extraction and therefore it was removed from the dataset (Figure 1B). Lastly, I noticed that many samples had non-zero heterozygosity on the X chromosome, which is unexpected given the individuals sequenced are male. I inferred there might be some low level of contamination in those individuals with high X heterozygosity, therefore I removed any individuals that had more than 1000 heterozygous sites on the X chromosome (Figure 1C). This resulted in a set of 119 individuals in the final analysis dataset. Most filtering thresholds were determined using R v3.5.1 with packages SeqArray (Zheng et al., 2012, 2017) and valr (Riemondy et al., 2017). For some analyses, an LD-pruned set of SNPs were selected using R package SNPRelate (Zheng et al., 2012; parameters of slide.max.bp=1000, maf=0.05, ld.threshold=0.3). A summary of how the filtering and LD-pruning altered the size of the dataset can be found in Table 2.

Table 2: ***The number of sequenced individuals and SNPs remaining after each filtering step was completed.*** NAs are listed for the number of SNPs remaining in the first two rows as up until that point, the total number of variable sites included indels.

Filtering step	N individuals remaining	N SNPs remaining
Raw Sequencing	263	NA
Raw VCF	263	NA
Indel removal	263	4031781
Remove non-passing SNPS (post VQSR)	263	3138621
Indels +/- 50 bp removal	263	1546027
Repetitive regions removal	263	1464685
Read-based individual filtering	156	1464685
Total read depth 1.5% and 98.5% quantile	156	1421388
Individual RD 1% & 99% quantile pass/fail	156	1176009
Only biallelic sites	156	1155082
CNV-based filtering	156	1067960
Last check for outlier individuals	119	1067960
LD pruned SNP set for analysis	119	165518

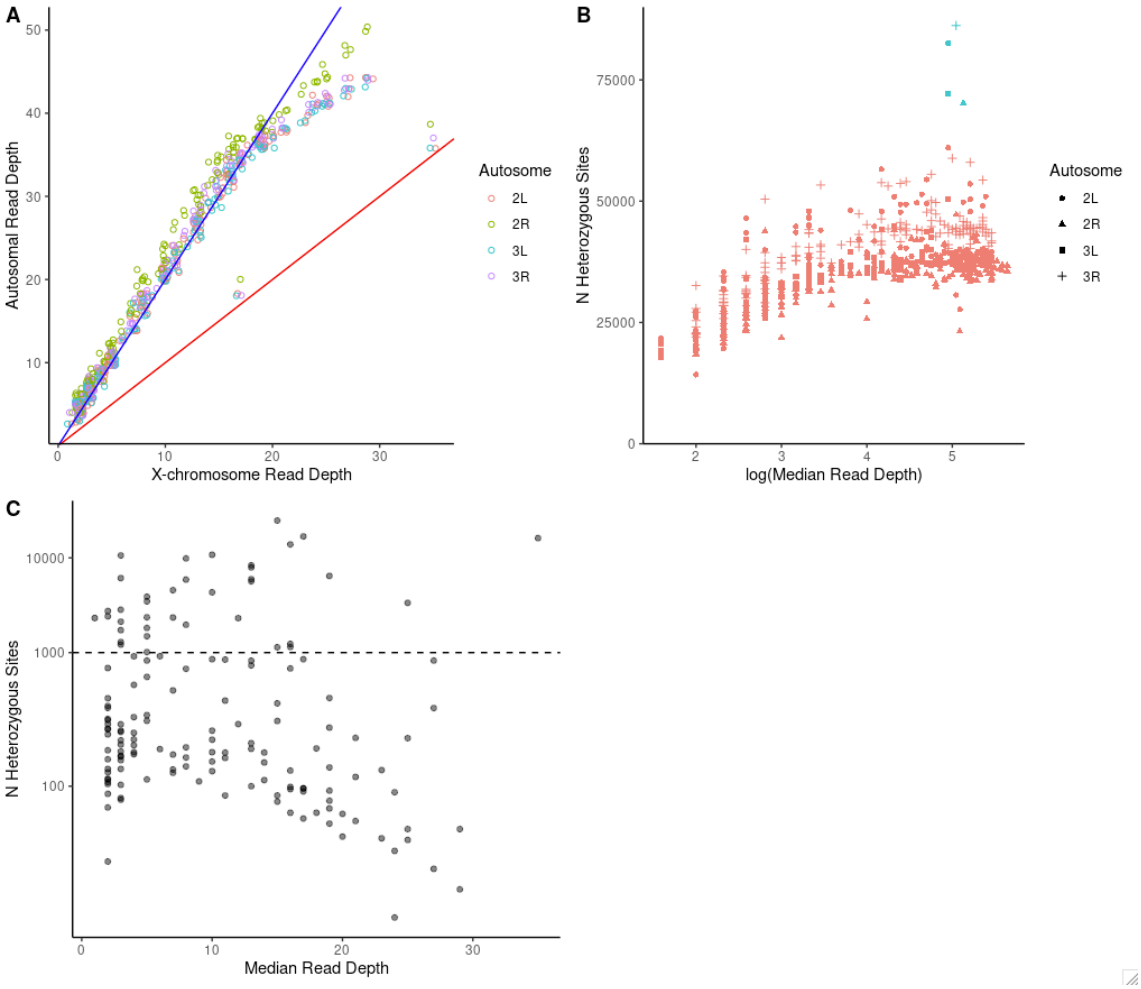


Figure 1: ***Filtering steps.*** A) Read depth on the X chromosome vs. each autosomal arm. Two putatively female samples fall along the 1:1 (red) line instead of the 2:1 (blue) line and were removed from the analysis. B) The log of the median read depth vs. the number of heterozygous sites of the different chromosomes. One outlier individual was removed from the analysis due to much higher levels of heterozygosity (blue) compared to the rest of the individuals (red) across all chromosomes. C) The median read depth of X chromosome per individual vs. the number of heterozygous sites on a log10 scale. Individuals with more than 1000 heterozygous sites on the X chromosome (dashed line) were excluded from the analysis.

Identifying Inversions

D. melanogaster has many known large cosmopolitan inversions. Using inversion specific SNPs (Corbett-Detig & Hartl, 2012), I determined whether inversion alleles were present in the samples. For each sample, I calculated the probability that the sample was in a standard, heterozygous, or homozygous-inverted state for each main inversion. If there was missing data at these inversion-specific SNPs, this reduced the ability to confidently call inversion status. After generating a probabilistic call of inversion status for each sample, I validated inversion status with a principal components analysis. Initial probabilistic calls reliably identified the presence of Inv(2L)t, but did not detect the presence of any other known inversions (Corbett-Detig & Hartl, 2012). Using R package LEA (Frichot & François, 2015), I ran a PCA on the LD pruned set of SNPs from chromosome 2L. After running the PCA (Figure 2) I labelled three samples (CM.025.0722, CM.017.0819, CM.029,0819) as inversion homozygotes despite one being called an inversion heterozygote by the probabilistic method.

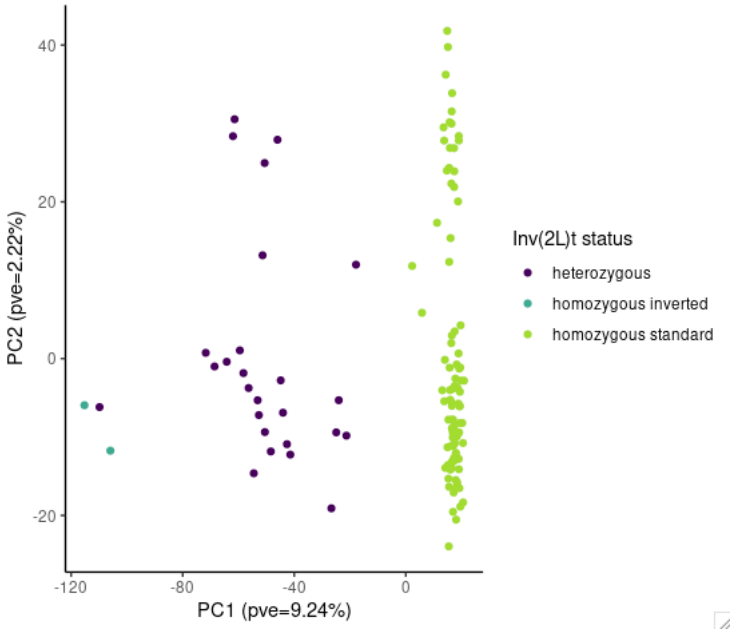


Figure 2: ***Identifying inversion status on chromosome 2L.*** A principal components analysis of LD-pruned SNPs on chromosome 2L, colored by the inversion status assigned probabilistically. One individual probabilistically called as a Inv(2L)t heterozygote clearly clusters with probabilistically called Inv(2L)t homozygotes.

Simulations

The data had several samples with low to moderate read depth, which can lead to heterozygous sites being called as homozygous when binomial sampling of chromosomes during sequencing and library preparation results in missing one chromosome entirely. Additionally, GATK uses likelihoods to call genotypes, therefore it occasionally calls homozygous genotypes despite reads that potentially support a heterozygous call. Because both could impact the genotype calls, I wanted to ensure that any patterns in summary statistics and analyses were not being driven by low read depth and genotype caller biases. To do that, I ran simulations to get a null expectation of how low read depth and genotype caller biases would impact baseline estimates of summary statistics.

I simulated genotypes using information about read depth as well as the distribution of genotype-calling probabilities to obtain null expectations. Simulations were run using the LD pruned set of SNPs to reduce computation time and to reduce signal inflation from linkage blocks. Prior to running the simulations, I assembled summary information from the data to use as input. Using the R package SeqArray (Zheng et al., 2017), I generated a table of read depths for every individual at every SNP and a table of reference allele frequencies at every SNP calculated for both all the individuals together and for each collection date. To account for the biases of the genotype caller in these simulations, I generated a probability table of reference and alternate read depths and the observed probability of which genotypes were called based on that set of read depths. To make this table, I grouped the data according to different combinations of reference and alternate allele read depths (i.e. ref=1 & alt=1; ref=2 & alt=1, ref=1 & alt=2, etc.) and then calculated the observed frequency of the three different genotypes for each observed combination of reference and alternate allele read depths. For each pair of reference and alternate read depths observed, there is an associated set of probabilities to observe each of the homozygotes (reference or alternate) and a heterozygote (Figure 3).

Simulated data included the generation of a simulated true genotype and a subsequently simulated observed genotype. The simulated true genotypes were generated using the allele frequencies from the data. Simulated observed genotypes were generated based on the simulated true genotype, using read depths from the data and the genotype probability table (Figure 4). If the simulated true genotype was homozygous, the depth of the allele it was homozygous for was set to the total read depth. If the simulated true

genotype was heterozygous, allele read depths were generated as a random draw of the two different alleles with the number of draws being equivalent to the read depth. The allele read depths were then used to generate the simulated observed genotype by using the genotype probability table previously described.

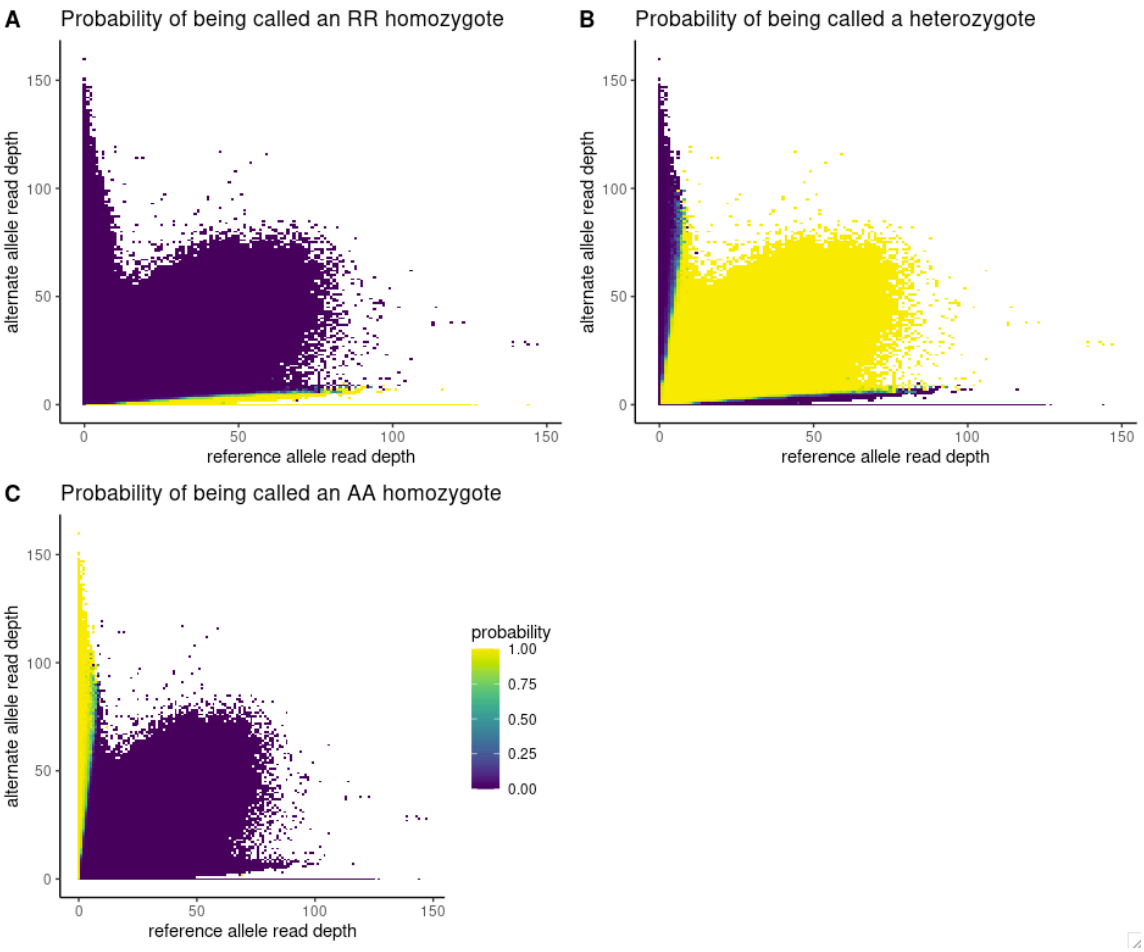
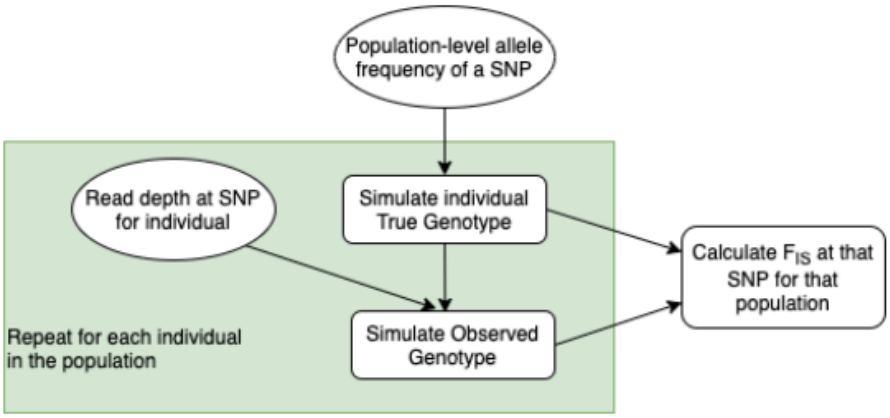


Figure 3: Visualization of the probability table used in the simulations. The panels show the reference allele read depth on the X and the alternate allele read depth on the Y, with the color scale inside each combination filled according to A) the probability of being called a reference-allele homozygote (RR), B) the probability of being called a heterozygote, and C) the probability of being called an alternate-allele homozygote (AA).

At the end of a simulation, I had a simulated true genotype and a simulated observed genotype for each SNP for each individual. Simulations were split into 2 different categories: all individuals (treating all 119 individuals as if they were one population) and by-population (treating each collection time-point as a single population). Each simulation was run 1000 times for every iteration to generate a null expectation based on read depth and genotype calling errors.

A.



B. Simulation Steps

Individual	1	2	3	4	5	6
Simulated True genotype	RR	RA	RA	RR	RA	RR
True read depth	10	8	13	17	16	9
Simulated Reference allele RD	10	7	8	17	3	9
Simulated Alternate allele RD	0	1	5	0	13	0
Simulated Observed genotype (generated with probability table)	RR	RR	RA	RR	RA	RR

F_{IS} calculated on Simulated Observed genotypes

Genotype	RR	RA	AA
Observed	4	2	0

Figure 4: ***Walk through of the simulations***. A) Flowchart of the steps of the simulations. Steps within the green box are repeated for every individual in a population per SNP. B) A toy example of a single iteration of a simulation where for all individuals in that population, a simulated true genotype is generated from the population allele frequency in the

data. Then reference (ref) and alternate (alt) allele read depths (RD=read depth) are generated as described in the text. The simulated reference and alternate allele depths are then used to pull probabilities of different genotypes from the probability table. Once simulated observed genotypes are generated, they are then used in calculations of FIS.

F_{IS} Analysis

Nei's F_{IS} (Nei & Chesser, 1983) was calculated for every SNP in both the data and in the simulations to assess levels of homozygosity. Equation 1 was applied to each SNP (N_{het} is the number of heterozygotes in the population, N_{total} is the total number of individuals in the population), treating different groups as the population: each time point as a population and then all the samples as a whole population (referred to as "all individuals"). Following the per-SNP Nei's F_{IS} calculation, I obtained an average F_{IS} by calculating the mean F_{IS} across the LD pruned set of SNPs for each population. Mean F_{IS} was calculated for both the data and the simulations. For the simulations, mean Nei's F_{IS} was calculated for each population per each simulation both for the simulated observed genotypes and the simulated true genotypes. To get a relative difference estimate of the difference between the data and the simulations, I used Equation 2 and calculated the difference between the empirical dataset and the simulated data relative to the simulations for all 1000 simulations.

$$(Eq.1) \qquad Nei's\ F_{IS} = 1 - \frac{{}^{N_{het}}/_{N_{Total}}}{({}^{N_{Total}}/_{N_{Total}} - 1) * (2pq - \frac{{}^{N_{het}}/_{N_{Total}}}{2 * N_{Total}})}$$

$$(Eq.2) \quad relative\ difference = \frac{empirical\ mean\ F_{IS} - simulated\ observed\ mean\ F_{IS}}{simulated\ observed\ mean\ F_{IS}}$$

A jackknife analysis was performed for each collection time point; Nei's F_{IS} was calculated per SNP then averaged across the whole genome, dropping each individual once. The jackknife Nei's F_{IS} calculation was also done on a random set of 10 simulations to reduce the computational burden of doing the jackknife Nei's F_{IS} calculation on all 1000 simulations. The relative difference calculation was performed similarly as described above, but again only using the random set of 10 simulations.

I also investigated whether individuals that were homozygous for Inv(2L)t were the sole drivers of the patterns of F_{IS}. Inv(2L)t was the only large cosmopolitan inversion detected in the samples, which could generate homozygosity across large stretches of chromosomes. I performed the same relative difference of Nei's F_{IS} described above between the empirical and simulated F_{IS} after removing the 3 individuals that were identified as homozygous for Inv(2L)t using inversion specific SNPs (Corbett-Detig et al., 2012) and the PCA.

Relatedness Analysis

To assess the relatedness of the individuals in the sample, I used the *snpgdsIBDKING()* function in the SNPRelate R package (Zheng et al., 2012). This function calculates both the KING estimate of kinship (Manichaikul et al., 2010) as well as proportion of sites that are identical-by-state 0 (IBS0). A site is IBS0 when the two individuals being compared share zero alleles that are identical by state (i.e. the two individuals are homozygous for opposite alleles). Plotting the proportion of sites that are IBS0 versus the KING estimate of kinship is one way to identify clusters of individuals that belong to specific classes of relatedness (Manichaikul et al., 2010).

After obtaining estimates of KING kinship, I also wanted to compare values of kinship for comparisons of individuals from the same time point (within) vs. in different time points (between). I tested whether individuals within a time point may be more related than individuals between time points using a t-test to compare the means of the two different distributions using the R function *t.test()*.

Sliding Window Analysis

I used a sliding window analysis to assess fine-scale patterns of mean F_{IS} across smaller windows of the genome. I used a window size of 1000 SNPs and a step size of 500. Using the LD pruned set of SNPs, I calculated the mean F_{IS} per window for each population and each chromosome for both the empirical data and the 1000 simulated observed genotypes.

I then compared the empirical estimates of window mean F_{IS} to the window mean F_{IS} estimates of the 1000 simulations to understand how the data compared to the null expectations. For each window, I determined the quantile rank of the observed F_{IS} relative to the simulated F_{IS} values. I then determined the mean and standard deviation of this ranking for each chromosome and population, allowing us to determine whether signals of positive F_{IS} beyond null expectation are genome-wide, or in clustered peaks.

Results

Dataset description and quality

I imposed strict and conservative filtering on the dataset to ensure that I obtained a set of high-quality SNPs. I began with 263 individuals, and after filtering I retained 119 individuals and 1,067,960 SNPs (Table 1; Table 2). In the LD-pruned dataset, I retained 165,518 SNPs across the autosomes (Table 2). Among the samples, I only detected heterozygotes and homozygotes for the known cosmopolitan inversion Inv(2L)t. In initial data exploration, I observed an excess of homozygosity among the genotypes in the samples.

Measuring F_{IS} to characterize excess homozygosity relative to a simulated null expectation

To quantify patterns of excess homozygosity, I obtained estimates of mean F_{IS} by averaging SNP-level estimates of F_{IS} across the whole genome. When treating all individuals as a single population, we see positive F_{IS} relative to the null expectation, which is expected as the population is evolving across the season and allele frequencies are changing due to both selection and neutral processes. We observe that for each individual time point, mean F_{IS} of the empirical data is greater than zero (Figure 5) which points to an excess of homozygosity in the samples. A mean F_{IS} greater than zero could be indicative of cryptic population structure (i.e. a Wahlund effect), strong selection within the focal orchard, or other biases.

I used a measure of F_{IS} that is designed to correct for the small sample sizes (Nei & Chesser, 1983) on the empirical dataset, simulated true genotypes, and simulated

observed genotypes (see methods for description of simulations). The simulated true genotypes should have a mean F_{IS} of zero given that genotypes were randomly assigned based on average allele frequencies within each time point. We see that estimates of mean F_{IS} for the simulated true genotypes is near zero for all time points indicating that the estimator of F_{IS} is unbiased (Figure 5A).

To eliminate some of the potential artifactual drivers of positive F_{IS}, I performed simulations to generate null expectations. I simulated observed genotypes where inflation of F_{IS} values may be driven by low read depth or improper genotype calls made by the genotype calling algorithm. To determine if the signal of positive F_{IS} in the data was greater than what we would expect due to artifactual homozygosity from read depth and genotype caller issues, I compared estimates of mean F_{IS} in the data to mean F_{IS} estimates from 1000 simulations. We observe that while F_{IS} is positive in the simulated observed genotypes (Figure 5A), the actual data mean F_{IS} is always more positive (Figure 5B). Specific time points show greater differences compared to the null expectation generated by the simulations. For instance, collections from August 19th and October 14th show the greatest deviations from null expectation, while the collection from June 24th shows the smallest deviation from null expectation (Figure 5B).

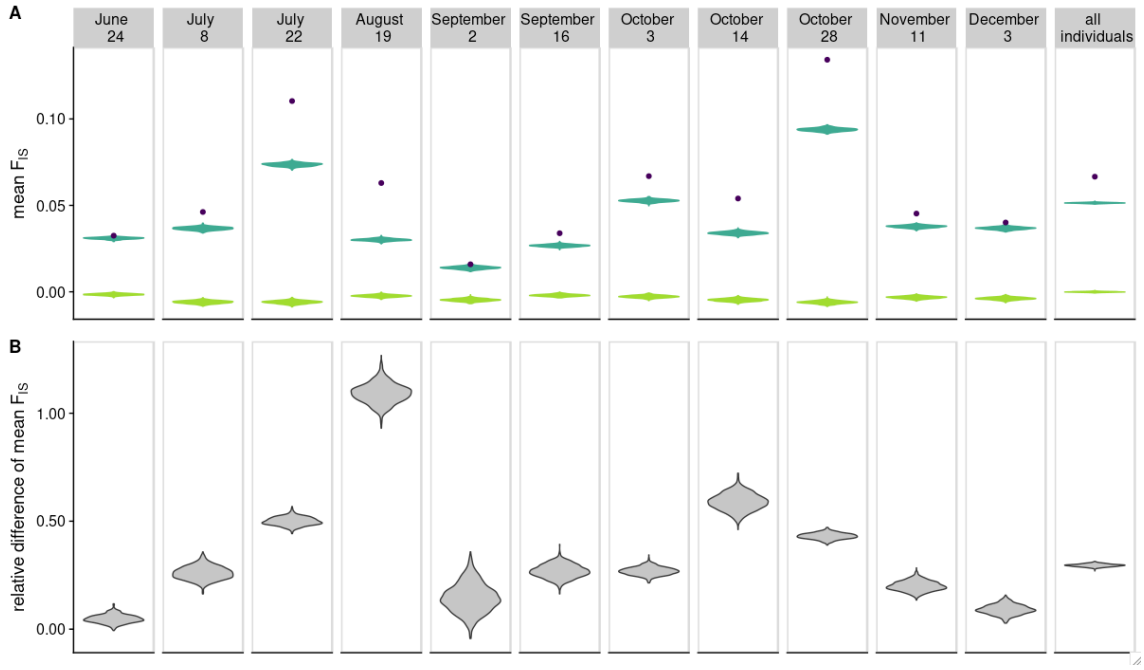


Figure 5: ***Mean F_{IS} in the data is more positive than the null expectation.*** A) Mean F_{IS} (calculated as the mean of F_{IS} across all SNPs using Nei's F_{IS} estimator) calculated for the data (purple), simulated observed genotypes (teal), and the simulated true genotypes (green). Each violin in the plot represents the distribution of 1000 simulations. B) The relative difference between the data mean F_{IS} and the individual simulation mean F_{IS} calculated by equation 2. Each violin in the plot represents the distribution of 1000 simulations.

I then assessed whether certain individuals in a time point drove patterns of positive F_{IS} since the focal orchard is part of a network of orchards and capturing migrants could result in positive F_{IS}. I performed a jackknife analysis calculating F_{IS} for each time point after dropping each individual one at a time. This analysis showed that for each time point, no single individual alone drives the patterns of positive F_{IS} (Figure 6).

Additionally, I tested whether individuals homozygous for the known large cosmopolitan inversion Inv(2L)t drove patterns of positive F_{IS}. Inversion homozygotes

were only found in two time points, July 22nd and August 19th. I performed the same assessment of F_{IS} in the data vs. simulations after removing Inv(2L)t homozygotes and found that Inv(2L)t homozygotes do not alter any of the general positive F_{IS} patterns where the data is greater than the simulated mean F_{IS}. Though there is a reduction in the relative difference of mean F_{IS} (Figure 7).

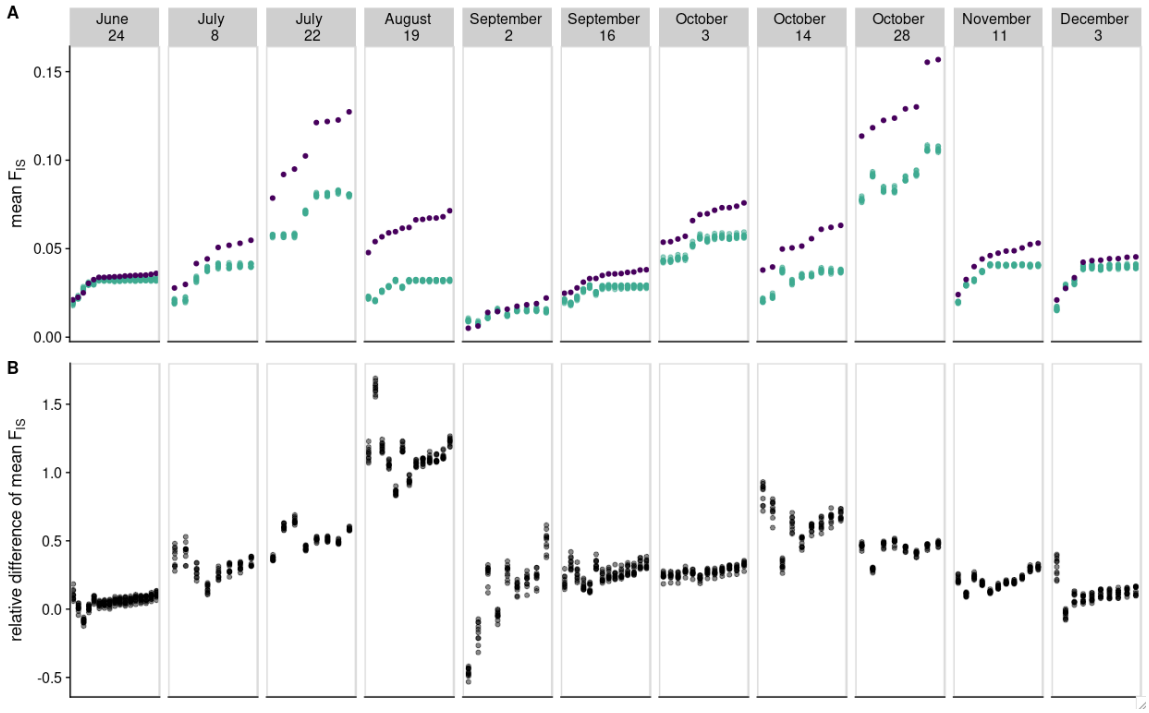


Figure 6: ***Jackknife analysis of mean F_{IS} shows that no single individual within a time point drives mean F_{IS} patterns.*** A) mean F_{IS} for the data (purple) and a subset of 10 simulations (teal) where each cluster of points along the x-axis corresponds to dropping the same sample. Samples are ordered by the empirical mean F_{IS}. B) The relative difference between the data and the 10 random simulations calculated by equation 2, where again each cluster of points corresponds to the same sample being dropped from the analysis, again ordered by the empirical mean F_{IS}.

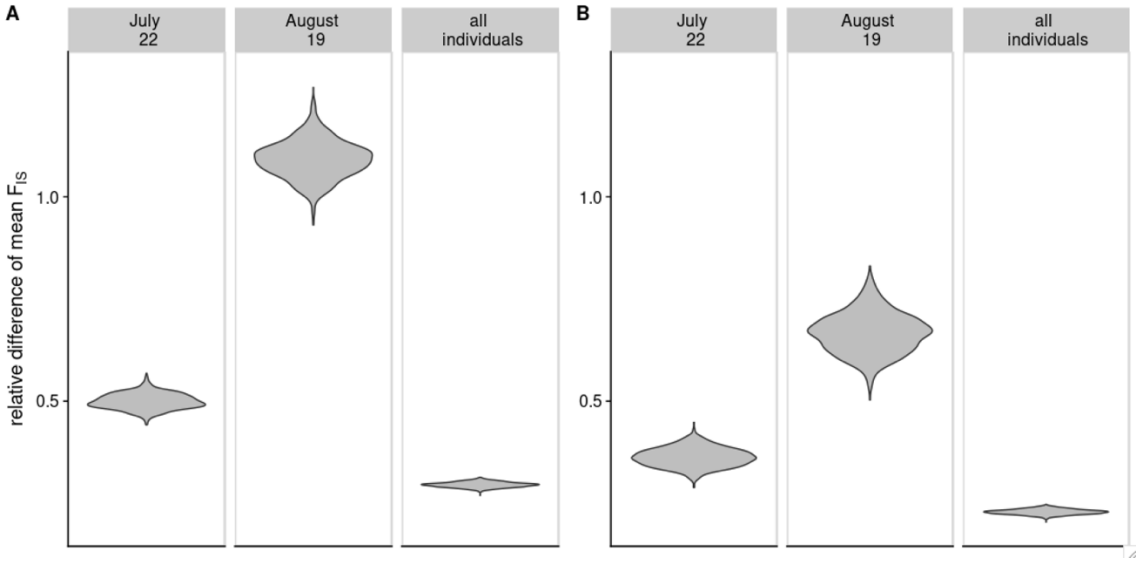


Figure 7: ***Inversion homozygotes do not drive mean F_{IS} patterns, but do accentuate them.*** Mean F_{IS} values were re-calculated for the time points that contained individuals homozygous for Inv(2L)t after removing the homozygotes. The relative difference between the data and the 10 random simulations calculated by equation 2, (A) before and (B) after removing Inv(2L)t homozygotes from the dataset. Each violin represents the distribution of 1000 simulations.

Assessing kinship to determine whether highly related individuals influence F_{IS} estimates

To assess whether the pattern of positive F_{IS} was driven by the presence of highly related individuals, I looked at patterns of relatedness with KING estimate of kinship (Manichaikul et al., 2010) and IBS0. I find that based on both values of KING kinship and IBS0, there are no closely related individuals, per criteria laid out by Manichaukul et al. (Figure 8A). We observe many strong negative values of kinship, which can be indicative of population structure (Manichaikul et al., 2010). The distribution of the estimates of kinship are not significantly different for estimates calculated within and between timepoints (Figure 8B; p=0.15). This is contrary to the intuitive assumption that

individuals within a single time point would be more closely related than individuals between time points. When split into individual time points, most distributions of kinship within versus between time points were not significantly different from each other. A few time points did show significant differences: July 22nd ($p=0.03$), September 2nd ($p=1.22x10^{-6}$), September 16th ($p=2.23x10^{-5}$), and October 28th ($p<0.008$).

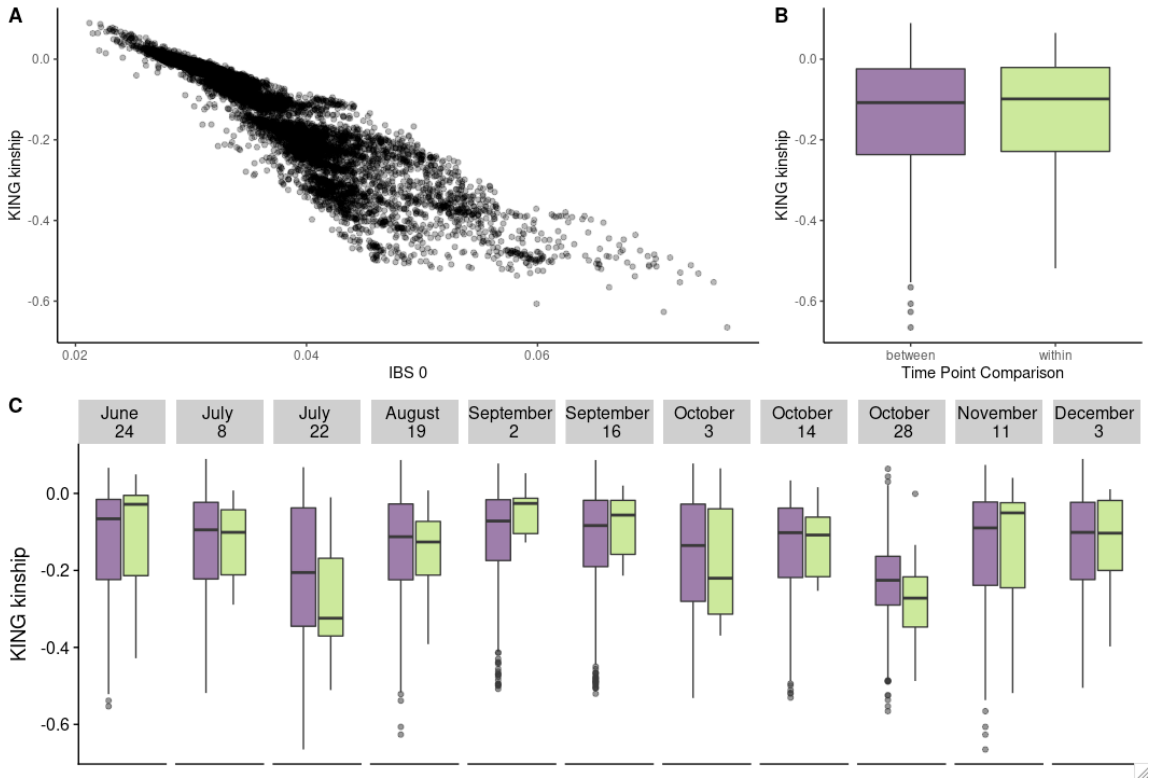


Figure 8: ***Kinship analysis shows that there are no closely related individuals in the sample, either within or between time points.*** A) Identity by state 0 vs. the KING estimates of kinship for all pairwise individual comparisons where each point is the estimate for a comparison between two individuals. B) Boxplots of the KING estimate of kinship for within- (green) or between-time (purple) point comparisons. C) Boxplots of KING estimates of kinship for within- (green) and between-time (purple) point comparisons, split into each time point.

Genome-wide distribution of F_{IS} to assess population structure vs. selection as drivers of F_{IS}

After eliminating potential artifactual drivers of positive F_{IS}, I wanted to know whether the pattern of positive F_{IS} beyond the null expectation is the product of cryptic population structure (i.e. a Wahlund effect) or selection. Genome-wide elevation of F_{IS} would point to a Wahlund effect, while sharper peaks of positive F_{IS} would point to selection. To test these two options, I assessed mean F_{IS} across the different chromosome arms for each time point using a sliding window (Figure 9). After getting sliding-window-based mean F_{IS} estimates, I assessed the quantile rank of the empirical F_{IS} relative to the simulations. For every chromosome and population, I calculated the mean and standard deviation of that rank across all windows for each chromosome separately. We observe that when treating all individuals as one population and some chromosomes from different time-point populations, the empirical window mean F_{IS} is consistently greater than 50% of the simulations, indicative of cryptic population structure driving positive F_{IS} patterns. Other populations show short windows of high empirical F_{IS} relative to simulations, indicative of selection driving positive F_{IS} patterns. This could imply that different processes of population structure or selection may be dominant forces at different sampling time points.

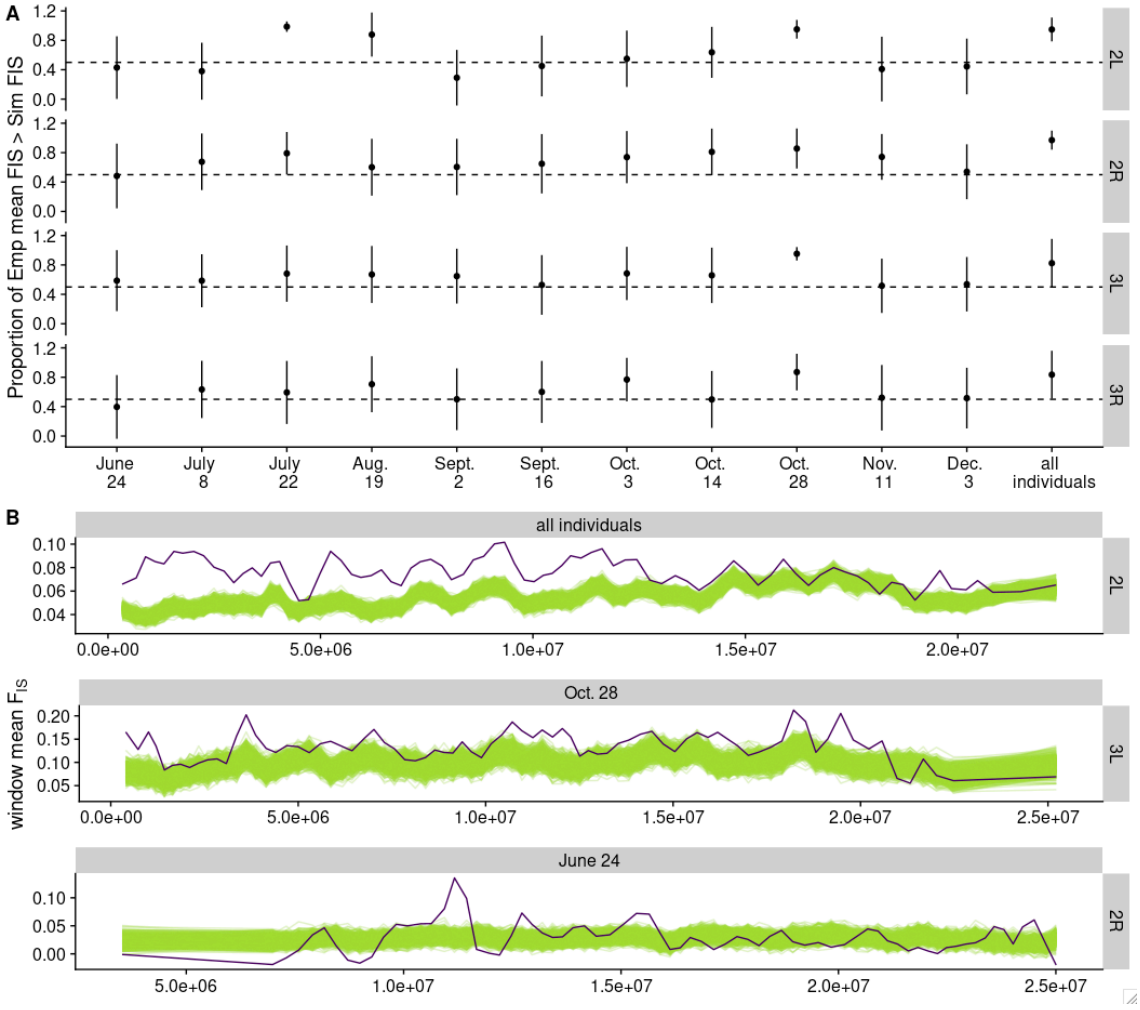


Figure 9: ***Sliding window analysis shows variability between populations that imply differences in demography and selection.*** A) For each population, the mean and +/- 1 standard deviation of the proportion of simulations where the empirical window mean F_{IS} (purple in panel B) is greater than the simulation window mean F_{IS} (green in panel B). The dashed line is set at 0.5, where the empirical data would equal the median simulated estimate of window mean F_{IS}. B) Sliding window means of F_{IS} for three representative samples. The three sliding windows show the empirical estimates (purple) and the 1000 simulation estimates (green). The representative samples of the sliding window include all individuals being treated as one population on chromosome 2L, a population (October 28th, chromosome 3L) showing whole-chromosome positive F_{IS} indicative of population structure, and a population (June 24th, chromosome 2R) lacking chromosome-wide positive F_{IS} indicating selection as the driver of positive F_{IS}.

Discussion

Here I asked whether a natural orchard population of *Drosophila melanogaster* exhibited signs of population structure and if the presence of population structure varied through time. To do this, I conducted whole-genome individual sequencing of the male offspring of isofemale lines caught on an approximately bi-weekly basis. When compared to a null expectation generated by simulations, I observe positive F_{IS} greater than the null expectation, which is indicative of the presence of cryptic population structure (i.e. a Wahlund effect). Genome-wide elevation of F_{IS} is observed when I treat all temporally spaced samples as if they were collected at one time, demonstrative of a temporal Wahlund effect as the population evolves over time. However, genome-wide signals of positive F_{IS} are not observed at all time points. The results suggest that there are stochastic shifts in the degree of population structure and panmixia through time within a single, seasonally-evolving orchard population of *D. melanogaster*.

The observation of stochastic shifts in the degree of structure do not match predictions based on prior work and contradicts long-standing assumptions that single populations of *D. melanogaster* will be panmictic at the local level because of long-distance dispersal capabilities (Coyne & Milstead, 1987). The results support older observations of population structure in naturally occurring wild populations of *D. melanogaster* (Band & Ives, 1961, 1963; Ives, 1945, 1954, 1970; Ives & Band, 1986). Shpak et al. (2010) quantified the observations of Ives and colleagues into a model where the early emergence of *D. melanogaster* starts with the establishment of few demes by few founders in the early spring. Based on this model, one prediction could be that population structure would be observed in the early season, and then would collapse into

panmixia as resources become more abundant and population sizes increase. However, this is not what I observe in the orchard population. I instead observe stochastic shifts between population structure and panmixia throughout the entire growing season. However, the sampling I performed was not on a fine spatial scale and therefore cannot outright reject predictions based on the Shpak and Ives model of population structure in wild *D. melanogaster* populations.

This raises the question: is population structure constantly shifting along a spectrum of genetically structured to local panmixia or is this stochasticity due to the sampling method? Stochasticity in population structure is not a new phenomenon, and has been observed in different organisms. In black bears, observations across several years identified a Wahlund effect in a year with limited food resources while other years did not show a Wahlund effect (McCall et al., 2013). In the cotton bollworm, temporal shifts in genetic differentiation between subpopulations and changes in the genetic structuring within populations occur and are predicted to track changes in crop host availability (Behere, Tay, Russell, Kranthi, & Batterham, 2013). In both these organisms, the shifts in population structure through time is correlated with shifts in important resource availability, something that also true in determining the rates of dispersal and migration, therefore impacting potential population structure, in *D. melanogaster* (Dubinin & Tiniakov, 1946; Johnston & Heed, 1975; Markow & Castrezana, 2000; Wallace, 1970).

The sampling scheme I performed did not cover the whole orchard and cannot directly test for the details of metapopulation structure and dynamics through time. I sampled from the varieties of fruit that were made open to the public a week or so before

the collections as those areas of the orchard had enough flies. As we entered the late fall, only one easily accessible sector of the orchard had appreciable numbers of flies to provide a large enough sample. Sampling of different small sectors of the orchard every collection could have resulted in stochastic sampling of any potential group structure. Sampling across a whole orchard for the entire season would be the best sampling scheme to address this question of stochasticity of population structure vs. sampling strategy. Until this question is answered though, those doing sampling of wild Drosophilids ought to take caution in how they collect their samples in order to capture a truly representative sample as panmixia may not always be the state of an orchard *D. melanogaster* population. One suggestion to obtain a representative sample of an orchard population would be to sample systematically across the entirety of an orchard and making certain all sampling does not come from a small subsample of closely related individuals (Hoffmann & Nielson, 1985; Jaenike & Selander, 1979). Though, interpretation of population level-statistics should be cautious until the degree of structure in temperate *D. melanogaster* populations is further investigated.

One question that arises is: at what level is a local deme? A deme could be a single tree, a cluster of trees or gradation across a single orchard, or it could be at the whole orchard level. Different degrees of spatial sampling have been performed to identify the level of a deme in many species, including in coral (Costantini, Fauvelot, & Abbiati, 2007), domesticated chickens (Wilkinson, Wiener, Teverson, Haley, & Hocking, 2012), and trees (Morand, Brachet, Rossignol, Dufour, & Frascaria-Lacoste, 2002). The orchard I sampled from is part of a network of interconnected orchards with frequent transport of produce between them. While I did not see any obvious signals of migrants

from other orchards as I did not observe any outlier individuals in the jackknife F_{IS} calculations, I cannot exclude the possibility that some of these signals are due to large amounts of migration between orchards. However, in terms of the extinction and recolonization of demes, individual trees or rows of similar fruit varieties that come into season at the same time may be the level at which distinct demes can be measured. Detailed spatial sampling through time both within an orchard and at multiple connected orchards would be required to explore the stochasticity of population structure and level of demes in *D. melanogaster*.

Population structure on a small, local level in wild populations of *D. melanogaster* could have broader implications for the evolutionary history of *D. melanogaster*. Many assumptions about the potential for evolution in natural populations of *D. melanogaster* rely on very large estimates of N_e in a panmictic population. However, population subdivision with even modest levels of migration between demes can impact both estimates of N_e, genetic diversity, and evolutionary potential (Pannell & Charlesworth, 2000).

If the presence of population structure can be confirmed and the scale of demes quantified in wild populations of *D. melanogaster*, we could use fruit flies as a model system to study rapid changes in metapopulation structure and how shifts in metapopulation dynamics impact population genetic diversity and differentiation. Additionally, the seasonal dynamics in temperate populations of fruit flies are repeated from year to year (Bergland, Behrman, O'Brien, Schmidt, & Petrov, 2014), leading to the potential to study how predictable the consequences and outcomes may be of metapopulation dynamics or how different ecological selective forces, such as species

competition or temperature, might alter those metapopulation dynamics of a single species. We observe that the community composition of other Drosophila species changes within a year and between years dramatically (Gleason, Roy, Everman, Gleason, & Morgan, 2019; chapter 3), thus changing degrees of interspecific competition may impact metapopulation dynamics. With confirmation and quantification of the presence and degree of population structure, temperate populations of *D. melanogaster* could be a new system of studying metapopulation dynamics on ecological time scales.

I have shown that *D. melanogaster* populations may not always be as panmictic as previously assumed. The signal of population structure was stochastic over time, contrary to the predictions based on the Ives & Shpak model of fly population temporal dynamics (Shpak et al., 2010). The signal of stochastic population structure could be a consequence of the sampling strategy, and thus those performing sampling of wild populations of *D. melanogaster* should take caution in how they perform samples if they want a true representative sample of an orchard population. Further detailed sampling within orchards and at connected orchards may elucidate metapopulation dynamics and at what level demes arise in natural *D. melanogaster* populations. Additionally, the observation of stochastic population structure can have broader implications for how we interpret the evolutionary history of *D. melanogaster* as well as could provide a new system to study rapid changes in metapopulations across ecological time scales.

Data Availability

All scripts and some information tables can be found at

https://github.com/abangerter/individualWGS_wildDmel

All raw sequencing data can be found under NCBI SRA BioProject ID PRJNA727484.

Data is embargoed until peer-reviewed publication.

Literature Cited

Band, H. T., & Ives, P. T. (1961). Correlated Changes in Environment and Lethal Frequency in a Natural Population of *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences*, *47*(2), 180–185.

Band, H. T., & Ives, P. T. (1963). Genetic Structure of Populations I. On the Nature of the Genetic Load in the South Amherst Populations of *Drosophila melanogaster*. *Evolution*, *17*(2), 198–215.

Baym, M., Kryazhimskiy, S., Lieberman, T. D., Chung, H., Desai, M. M., & Kishony, R. (2015). Inexpensive Multiplexed Library Preparation for Megabase-Sized Genomes. (S. J. Green, Ed.)*PLOS ONE*, *10*(5), e0128036-15.

Behere, G. T., Tay, W. T., Russell, D. A., Kranthi, K. R., & Batterham, P. (2013). Population Genetic Structure of the Cotton Bollworm *Helicoverpa armigera* (Hübner) (Lepidoptera: Noctuidae) in India as Inferred from EPIC-PCR DNA Markers. *PLoS ONE*, *8*(1), e53448.

Bergland, A. O., Behrman, E. L., O'Brien, K. R., Schmidt, P. S., & Petrov, D. A. (2014). Genomic Evidence of Rapid and Stable Adaptive Oscillations over Seasonal Time Scales in *Drosophila*. (D. Bolnick, Ed.)*PLoS Genetics*, *10*(11), e1004775-19.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120.

Chen, M. H., & Dorn, S. (2010). Microsatellites reveal genetic differentiation among populations in an insect species with high genetic variability in dispersal, the codling moth, *Cydia pomonella* (L.) (Lepidoptera: Tortricidae). *Bulletin of Entomological Research*.

Chen, X., Schulz-Trieglaff, O., Shaw, R., Barnes, B., Schlesinger, F., Källberg, M., Cox, A. J., et al. (2016). Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics*, 1220–1222.

Corbett-Detig, R. B., & Hartl, D. L. (2012). Population Genomics of Inversion Polymorphisms in *Drosophila melanogaster*. (H. S. Malik, Ed.)*PLoS Genetics*, *8*(12), e1003056-15.

Costantini, F., Fauvelot, C., & Abbiati, M. (2007). Fine-scale genetic structuring in *Corallium rubrum*: evidence of inbreeding and limited effective larval dispersal. *Marine Ecology Progress Series*, *340*, 109–119.

Coyne, J. A., & Milstead, B. (1987). Long-Distance Migration of Drosophila. 3. Dispersal of *D. melanogaster* Alleles from a Maryland Orchard. *The American Naturalist*, *130*(1), 70–82.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., et al. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158.

Dubinin, N. P., & Tiniakov, G. G. (1946). Inversion gradients and natural selection in ecological races of *Drosophila funebris*. *Genetics*, *31*, 537–545.

Ellegren, H., & Galtier, N. (2016). Determinants of genetic diversity. *Nature Reviews Genetics*, *17*(7), 422–433.

Fountain, T., Duvaux, L., Horsburgh, G., Reinhardt, K., & Butlin, R. K. (2014). Human-facilitated metapopulation dynamics in an emerging pest species, *Cimex lectularius*. *Molecular Ecology*, *23*(5), 1071–1084.

Frichot, E., & François, O. (2015). LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution*, *6*(8), 925–929.

Gleason, J. M., Roy, P. R., Everman, E. R., Gleason, T. C., & Morgan, T. J. (2019). Phenology of *Drosophila* species across a temperate growing season and implications for behavior. (N. Desneux, Ed.)*PLOS ONE*, *14*(5), e0216601-21.

Hanski, I. (1994). A Practical Model of Metapopulation Dynamics. *Journal of Animal Ecology*, *63*(1), 151–162.

Hanski, I. (1999). Habitat Connectivity, Habitat Continuity, and Metapopulations in Dynamic Landscapes. *Oikos*, *87*(2), 209–219.

Hoffmann, A. A., & Nielson, K. M. (1985). The Effect of Resource Subdivision on Genetic Variation in *Drosophila*. *The American Naturalist*, *125*(3), 421–430.

Huang, W., Massouras, A., Inoue, Y., Peiffer, J., Ramia, M., Tarone, A. M., Turlapati, L., et al. (2014). Natural variation in genome architecture among 205 *Drosophila melanogaster* Genetic Reference Panel lines. *Genome Research*, *24*(7), 1193–1208.

Ives, P. T. (1945). The Genetic Structure of American Populations of *Drosophila melanogaster*. *Genetics*, *30*(2), 167–196

Ives, P. T. (1954). Genetic Changes in American Populations of *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences*, *40*(2), 87–92.

Ives, P. T. (1970). Further Genetic Studies of the South Amherst Population of *Drosophila melanogaster*. *Evolution*, *24*(3), 507–518.

Ives, P. T., & Band, H. T. (1986). Continuing Studies on the South Amherst *Drosophila melanogaster* Natural Population During the 1970's and 1980's. *Evolution*, *40*(6), 1289–1302.

Jaenike, J., & Selander, R. K. (1979). Ecological Generalism in *Drosophila falleni*: Genetic Evidence. *Evolution*, *33*(2), 741–748.

Johnston, J. S., & Heed, W. B. (1975). Dispersal of *Drosophila*: The Effect of Baiting on the Behavior and Distribution of Natural Populations. *The American Naturalist*, *109*(966), 207–216.

Keyghobadi, N., Roland, J., & Strobeck, C. (2005). Genetic differentiation and gene flow among populations of the alpine butterfly, *Parnassius smintheus*, vary with landscape connectivity. *Molecular Ecology*, *14*(7), 1897–1909.

Lack, J. B., Lange, J. D., Tang, A. B., Corbett-Detig, R. B., & Pool, J. E. (2016). A Thousand Fly Genomes: An Expanded *Drosophila* Genome Nexus. *bioRxiv*, 063537.

Li, H, & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, *26*(5), 589–595.

Li, Heng. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM, 1–3.

Machado, H., Bergland, A. O., Taylor, R., Tilk, S., Behrman, E., Dyer, K., Fabian, D., et al. (2021). Broad geographic sampling reveals the shared basis and environmental correlates of seasonal adaptation in *Drosophila. eLife*, *10*(e67577), 1–21.

Mackay, T. F. C., Richards, S., Stone, E. A., Barbadilla, A., Ayroles, J. F., Zhu, D., Casillas, S., et al. (2012). The *Drosophila melanogaster* Genetic Reference Panel. *Nature*, *482*(7384), 173–178.

Manichaikul, A., Mychaleckyj, J. C., Rich, S. S., Daly, K., Sale, M., & Chen, W.-M. (2010). Robust relationship inference in genome-wide association studies. *Bioinformatics*, *26*(22), 2867–2873.

Markow, T. A., & Castrezana, S. (2000). Dispersal in cactophilic *Drosophila*. *Oikos*, *89*(2), 378–386.

McCall, B. S., Mitchell, M. S., Schwartz, M. K., Hayden, J., Cushman, S. A., Zager, P., & Kasworm, W. F. (2013). Combined use of mark-recapture and genetic analyses reveals response of a black bear population to changes in food productivity. *The Journal of Wildlife Management*, *77*(8), 1572–1582.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303.

Meyer, M., & Kircher, M. (2010). Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing. *Cold Spring Harbor Protocols*, *2010*(6), 5448–5448.

Morand, M.-E., Brachet, S., Rossignol, P., Dufour, J., & Frascaria-Lacoste, N. (2002). A generalized heterozygote deficiency assessed with microsatellites in French common ash populations. *Molecular Ecology*, *11*, 377–385.

Nei, M., & Chesser, R. K. (1983). Estimation of fixation indices and gene diversities. *Annals of Human Genetics*, (47), 253–259.

Pannell, J. R., & Charlesworth, B. (2000). Effects of metapopulation processes on measures of genetic diversity. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *355*(1404), 1851–1864.

Riemondy, K. A., Sheridan, R. M., Gillen, A., Yu, Y., Bennett, C. G., & Hesselberth, J. R. (2017). valr: Reproducible genome interval analysis in R. *F1000Research*, *6*, 1025.

Shpak, M., Wakeley, J., Garrigan, D., & Lewontin, R. C. (2010). A Structured Coalescent Process for Seasonally Fluctuating Populations. *Evolution*, *64*(5), 1–15.

Smit, A., Hubley, R., & Green, P. (2013). RepeatMasker Open-4.0.

Wahlund, S. (1928). Zusammensetzung von populationen und korrelationserscheinungen vom standpunkt der vererbungslehre aus betrachtet. *Hereditas*, *11*, 65–106.

Wallace, B. (1970). Essays in Evolution and Genetics in Honor of Theodosius Dobzhansky: Observations on the Microdispersion of *Drosophila melanogaster*. In M. K. Hecht & W. C. Steere (Eds.), (pp. 381–399). Meredith Corporation.

Wilkinson, S., Wiener, P., Teverson, D., Haley, C. S., & Hocking, P. M. (2012). Characterization of the genetic diversity, structure and admixture of British chicken breeds. *Animal Genetics*, *43*(5), 552–563.

Zheng, X., Gogarten, S. M., Lawrence, M., Stilp, A., Conomos, M. P., Weir, B. S., Laurie, C., et al. (2017). SeqArray—a storage-efficient high-performance data format for WGS variant calls. (I. Birol, Ed.)*Bioinformatics*, *33*(15), 2251–2257.

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, *28*(24), 3326–3328.

CHAPTER THREE

A natural community of multiple Drosophilid species shows changes in species abundance and different degrees of genetic turnover

Abstract

Species composition and abundances in a community can rapidly fluctuate within a year and vary between years. Fluctuations in composition and abundance can provide insight into the biology and ecology of the species in a community, particularly when paired with genetic data. I observed and recorded species composition and abundance for multiple Drosophilid species in a single orchard community across two years on at least a bi-weekly basis. Additionally, I perform pooled sequencing of a subset of collections and species from the orchard community. I focused on three cosmopolitan Drosophilid species (*Drosophila melanogaster*, *D. simulans*, and *D. hydei*) and two recent invasive species (*D. suzukii* and *Zaprionus indianus*). Drosophilid species composition and abundance dramatically fluctuate within and between years. With the pooled sequencing data, I find differences in population turnover that may point to differences in overwintering and migratory behavior. Relatively few SNPs oscillated in frequency depending on the season of collection, likely due to reduced power from small sample sizes. I find that most species of fruit fly tested, with the exception of *D. simulans*, show signals of overwintering and therefore have the potential to be adapting to seasonally varying selection pressures.

Introduction

In a community, species composition may fluctuate across time. These fluctuations can be a result of both abiotic factors and biotic factors. One concept of community assembly posits that a single community is composed of a subset of species from the surrounding region that can survive the abiotic conditions in that community environment with further "filtering" of species based on biotic interactions and niche filling (Weiher et al., 2011). There is a lack of predictability in species composition because of the many different factors and assumptions in models of community assembly (Ozinga et al., 2005). Changes in species composition and abundance over time can reflect their degree of adaptation to their current environment, priority effects, biotic interactions with other species, rates of extirpation and recolonization versus established populations, and many other factors (Weiher et al., 2011).

Shifts in species composition can influence the evolutionary trajectory of established populations (Dieckmann & Doebeli, 1999) and can drive rapid evolution (Grant & Grant, 2006; Hart, Turcotte, & Levine, 2019) in response to interspecific competition. For example, experimental evolution of *Colpoda*, a ciliated protozoan, found that interspecific competition resulted in smaller cell sizes and higher population growth rates (TerHorst, 2011). Across several generations of competition between *Drosophila melanogaster* and *D. simulans* on food with different concentrations of ethanol, *D. simulans* increased in competitive ability through adaptation to the ethanol-rich environment (Joshi & Thompson, 1995). Experimental outdoor cages of *D. melanogaster* showed that competition with *Zaprionus indianus* during summer months influenced the later evolution of a variety of traits during the onset of winter (Grainger,

Rudman, Schmidt, & Levine, 2021). Community composition can be a source of selection that is temporally variable, similar to abiotic factors like temperature.

One model system for understanding the response to temporal variability is seasonal adaptation in the fruit fly, *Drosophila melanogaster* (Bergland, Behrman, O'Brien, Schmidt, & Petrov, 2014; Machado et al., 2021). In fruit flies, aside from findings that innate immunity changes seasonally (Behrman et al., 2018), the biotic agents of selection are not well characterized (Rudman et al., 2019). Different natural environments put *D. melanogaster* in contact with a variety of Drosophilid species and community composition that can change from month to month, likely in response to thermal limits of different Drosophilids (Gleason, Roy, Everman, Gleason, & Morgan, 2019). Previous studies show that species composition can change depending on the year of observation (Yoshimoto, 1952). However, many of these observations that span across decades do not follow a single population in consecutive years, do not include recent invasive species, and do not indicate how recent invaders may impact abundance.

Some of the species that coexist and compete with *D. melanogaster* have limited investigations into their ecology and whether they are likely similar to *D. melanogaster*, which adapts to seasonally varying selection and overwinters. Among these species are *Drosophila simulans*, *D. hydei*, *D. suzukii*, and *Zaprionus indianus*. We know that *D. melanogaster* shows seasonal genetic changes that are replicated globally (Bergland et al., 2014; Machado et al., 2021; Chapter 1) and has established populations that do not experience total extirpation and recolonization every year (Ives, 1970; Machado et al., 2016). Conversely, *D. simulans* shows a lack of a stable genetic latitudinal cline from year to year and a weak isolation by distance pattern along the North American east coast

that is indicative of the lack of established populations, lower winter survival or complete extirpation, and a recurrent northward migration to recolonize northern populations (Machado et al., 2016).

D. hydei is considered a cosmopolitan Drosophilid like *D. melanogaster* and *D. simulans*; it is not native to North America, but have spread across the globe and are now well established in most continents (Atkinson & Shorrocks, 1977). *D. hydei* is known for its large sperm and unique mating patterns (Markow, 1985; Pitnick, Markow, & Spicer, 1995), but also exhibits high temperature tolerance (Mitchell & Hoffmann, 2010; Overgaard, Kristensen, Mitchell, & Hoffmann, 2011), cold temperature tolerance (Overgaard, Kristensen, et al., 2011), and fast chill coma recovery times (Gibert, Moreteau, Ptavy, Karan, & David, 2001) relative to many other Drosophila species.

Z. indianus is a recent invasive species that was first observed in North America in 2005 (Linde et al., 2006), and is considered to be an agricultural pest, particularly in fig crops. Despite being invasive and experiencing the expected reduction in genetic diversity in invasive populations, *Z. indianus* still has high levels of genetic diversity (Comeault et al., 2020). When comparing invasive populations of *Z. indianus* to native populations, there was no significant reduction in performance across a breadth of thermal environments (Comeault et al., 2020). It is believed that *Z. indianus* does not establish populations and overwinter because samples from temperate populations in North America do not reproduce at lower temperatures (Comeault et al., 2020). There is also evidence that *Z. indianus* might die during winter conditions as it disappears from abundance estimates in colder fall and winter months (Gleason et al., 2019; Joshi, Biddinger, Demchak, & Deppen, 2014).

D. suzukii is also a recent invasive species first observed in North America in 2008 (Bolda, Goodhue, & Zalom, 2009; Goodhue, Bolda, Farnsworth, Williams, & Zalom, 2011), and first observed on the east coast of North America in Florida in 2009 (Iglesias, Price, Roubos, Renkema, & Liburd, 2016). Across the globe, *D. suzukii* has become a major agricultural pest to soft fruits and thin-skinned berries because females have a serrated ovipositor that facilitates the laying of eggs into soft fruits. The thermal limits of *D. suzukii* are comparable to the thermal limits of *D. melanogaster,* though tests of mortality at extreme temperatures are contradictory (Asplen et al., 2015). They appear to have a female reproductive diapause in winter conditions (Wallingford & Loeb, 2016; Zhai et al., 2016), have a winter morph that is phenotypically distinct from summer populations (Shearer et al., 2016), and may overwinter in leaf litter in woodland habitats (Pelton et al., 2016), thus potentially facilitating winter survival and established populations.

Here, I track species abundance and composition shifts across two years in a single orchard community. For the species described above, I perform pooled sequencing across time in two years. I find that the distribution of species present dramatically fluctuates within a single year, but also between years. Using temporal pooled sequencing data, I also find that different species exhibit different degrees of population turnover. The differing degree of population turnover can provide insight into overwintering and migratory patterns of these species.

Materials and Methods

Sample Collection

Roughly bi-weekly sampling in 2017 and weekly sampling in 2018 was conducted at a peach and apple orchard near Charlottesville, VA (Carter Mountain; 37.99°N, 78.47°W). In 2017, the number of *Zaprionus indianus* collected exceeded the time and ability to record, and therefore not all individuals were counted and collected. Thus, the 2017 counts of *Z. indianus* are an underestimate of their abundance, and therefore all other species abundances are an overestimate. The weekly sampling in 2018 varied in effort; every other week I did a large collection and in the off weeks I did a smaller collection, which is reflected in the number of individuals collected (Supplemental Table 1). I conclusively identified 8 common species in the collections and also identified up to 5 other species that are grouped together as "unknown" (Supplemental Table1). Species identification was done by consulting *Drosophilids of the Midwest and Northeast* (Werner & Jaenike, 2017). After sorting, flies were preserved in ethanol at -20°C in species-specific vials.

Species Selection

From the different species of Drosophilids that I found in the collection, I chose to sequence temporal samples from 5 species. Species were selected to be sequenced depending on the availability of a reference genome so that the dataset preparation pipeline could be uniform. I selected *Drosophila melanogaster*, *Drosophila simulans*, *Drosophila hydei*, *Drosophila suzukii*, and *Zaprionus indianus*.

Sample Preparation & Sequencing

Spring and fall pooled sequencing samples for 2017 and 2018 were selected so that samples from the different species were collected around the same time. Spring samples were selected from approximately the first month of collections in June or July. Fall samples were selected from mid-late fall months, prior to the first frost and the first night below freezing temperatures. DNA extractions were done on all flies of that species that were collected at the selected time point (Table 1).

We performed DNA extractions using the extraction protocol outlined in Bergland et al., 2014 and followed the library preparation protocol described in Chapter 1. We diluted the extracted DNA with water in a 1:1 ratio prior to shearing with a Covaris sonicator to create fragments of 500 bp in length. Using the NEBNext Ultra II kit and dual indices following manufacturer protocols, we made libraries for each sample with 8 cycles of PCR during the PCR enrichment step. We quantified the concentration of each library and then pooled the libraries in equal concentrations. After pooling all libraries, we used a Pippen to size-select the DNA in the 600-750 bp range. We then sequenced the pooled library on a NovaSeq with 2x150 paired-end reads.

Table 1: ***Metadata for sequenced samples.***

Species	Season	Year	Date Collected (MM/DD)	N Flies in Sequencing Pool	SRA Accession
D. melanogaster	spring	2017	07/20	120	SRR14476487
D. suzukii	spring	2017	07/20	36	SRR14476475
Z. indianus	spring	2017	08/03	189	SRR14476482
D. suzukii	fall	2017	09/28	37	SRR14476474
D. melanogaster	fall	2017	10/26	299	SRR14476486
Z. indianus	fall	2017	10/26	126	SRR14476481

D. simulans	fall	2017	10/26	65	SRR14476478
D. hydei	fall	2017	11/09	21	SRR14476485
D. hydei	spring	2018	07/05	64	SRR14476484
D. simulans	spring	2018	07/05	82	SRR14476477
D. suzukii	spring	2018	07/19	32	SRR14476473
D. suzukii	fall	2018	09/06	46	SRR14476472
D. hydei	fall	2018	09/20	50	SRR14476483
D. melanogaster	fall	2018	10/11	81	SRR14476479
D. simulans	fall	2018	10/11	303	SRR14476476
Z. indianus	fall	2018	10/11	372	SRR14476480

Dataset Preparation & Filtering

Raw fastq files went through a mapping pipeline comparable to Machado et al., 2018 and Chapter 1, but modified to work for non-*D. melanogaster* species. I removed adaptor sequences from the raw sequencing data using cutAdapt (Martin, 2011) and then mapped to the appropriate reference genome (Table 2) with BWA-mem (Li & Durbin, 2010; Li, 2013). I removed contaminating reads by competitively mapping each species with the speculated most likely contaminants (Table 2). Any reads that mapped to the potential contaminating species were removed from the sample. After mapping and removal of potential contaminants, I then ran bam files through samtools mpileup for each cluster of species samples. The mpileup files were then used to call variants into the VCF file with VarScan v2.3.9 (Koboldt et al., 2012).

The VCF files for each species were then filtered in similar fashion, though independently for each species. VarScan v2.3.9 was used to call de novo indel variants using default parameters (Koboldt et al., 2012). I then filtered out any SNP variants that were within +/- 10 bp of de novo indels. I removed repetitive regions identified by repeat masker (Smit, Hubley & Green; RepeatMasker at http://repeatmasker.org) and then

removed non-biallelic sites. After this, in the final filtering step I removed any SNPs that had a total read depth across all samples above the 99th percentile in order to remove potential copy number variants or remnant repetitive regions. Effective coverage of samples was calculated as (number of chromosomes * read depth)/(number of chromosomes + read depth) (Bergland et al., 2014; Kolaczkowski, Kern, Holloway, & Begun, 2011; Machado et al., 2021; Machado et al., 2016). Effective coverages of sex-linked SNPs were adjusted on the X chromosome for *D. melanogaster*, *D. simulans*, and *D. suzukii* where the identities of X chromosome scaffolds are known.

Table 2: ***Reference genomes used and species used in competitive mapping.*** Species selected for competitive mapping were selected based on either phenotypic similarity or abundance in the sample. *D. melanogaster* and *D. simulans* are phenotypically similar sister species. For the other species that are more phenotypically distinct I selected *D. melanogaster* and *D. simulans* as they were among the most abundant species present and therefore more likely to contaminate other samples.

Species	Competitively mapped against	Reference genome	Citation or NCBI accession for genome
D. melanogaster	*D. simulans*	dm6.12	Hoskins et al., 2015
D. simulans	*D. melanogaster*	Prin_Dsim_3.0	PRJNA377886
D. suzukii	*D. melanogaster & D. simulans*	Dsuz-WT3_v1.0	Paris et al., 2020
D. hydei	*D. melanogaster & D. simulans*	DhydRS2	PRJNA475270
Z. indianus	*D. melanogaster & D. simulans*	z_indianus_16GNV01_v02	Comeault et al., 2020

Analysis

To estimate genetic differentiation through time, I used R package poolfstat (Hivert, Leblois, Petit, Gautier, & Vitalis, 2018) to obtain a whole genome estimate of pairwise F_{ST} for each sample within species. To identify seasonally oscillating SNPs

(Bergland et al., 2014) I ran a linear model on effective-coverage corrected allele frequencies. I only ran the linear model on SNPs with a minimum of 3 samples with allele frequency estimates and removed SNPs with a corrected allele frequency below 0.1 in any time point. Using R function *glm()*, I ran a generalized linear model of allele frequency on season weighted by effective coverage for each SNP. Following the linear model, I performed Bonferroni and false discovery rate P-value corrections.

Results

Species distribution through time

In the bi-weekly collections in 2017 and weekly collections in 2018, I observed dramatic changes in species distributions (Figure 1). In 2017, I observed *Drosophila melanogaster* and *Drosophila tripunctata* (Supplemental Table 1) starting the season as dominant species. By mid-summer 2017, *Zaprionus indianus* was the dominant species, despite the undercount (see Materials & Methods). By the end of fall 2017, *D. melanogaster* and *D. simulans* were the dominant species. In contrast, *D. simulans* was dominant at all time points in 2018 with the exception of the first collection of the year, while *D. melanogaster* did not reach appreciable frequencies until late fall. I also saw greater proportions of *D. hydei* in 2018 than I did in 2017. *D. suzukii* appears to drop off in collections at the onset of fall in both 2017 and 2018.

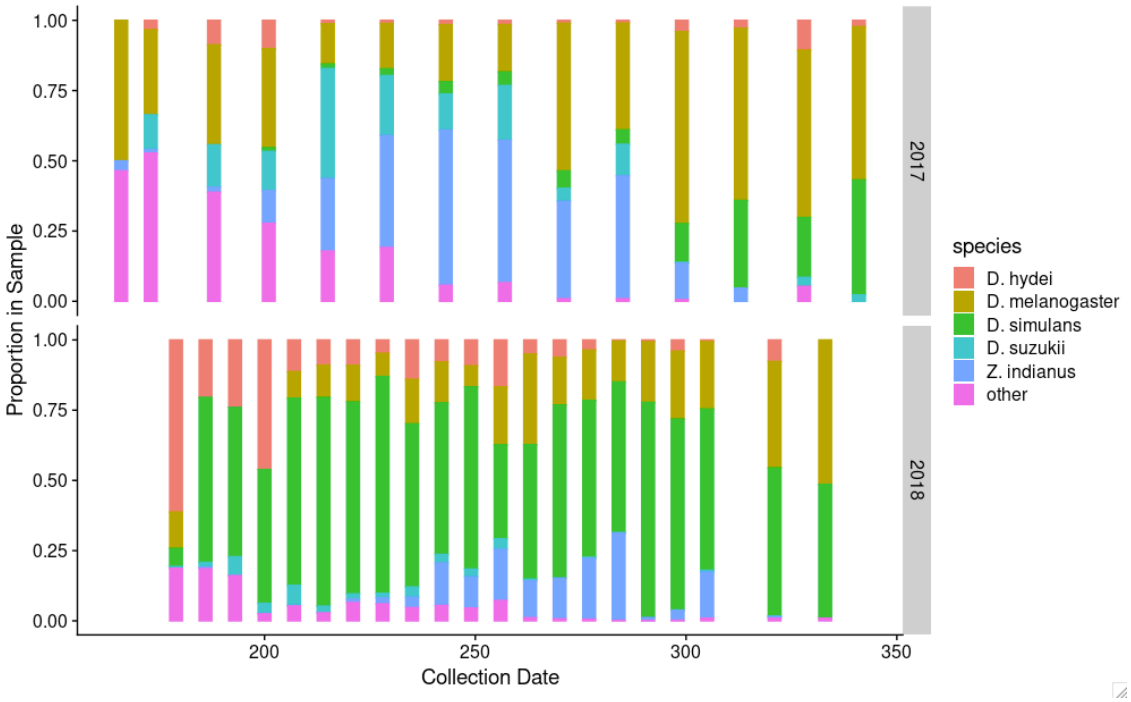


Figure 1: ***Species distributions change within and between years***. The collections of multiple species for 2017 and 2018 show that the proportion of species change through time, collection date referring to the day of year. Note that the proportions of *Zaprionus indianus* in 2017 are an undercount.

Sequencing data quality

After filtering, most of the samples reach or exceed the target effective coverage of 50 (Figure 2A). Most samples that did not reach the target effective coverage had low sample size (Table 1). After filtering I retained 1,722,778 SNPs for *D. melanogaster*, 3,548,696 SNPs for *D. simulans*, 8,195,100 SNPs for *D. suzukii*, 2,294,055 SNPs for *D. hydei*, and 4,376,346 SNPs for *Z. indianous* (Figure 2B).

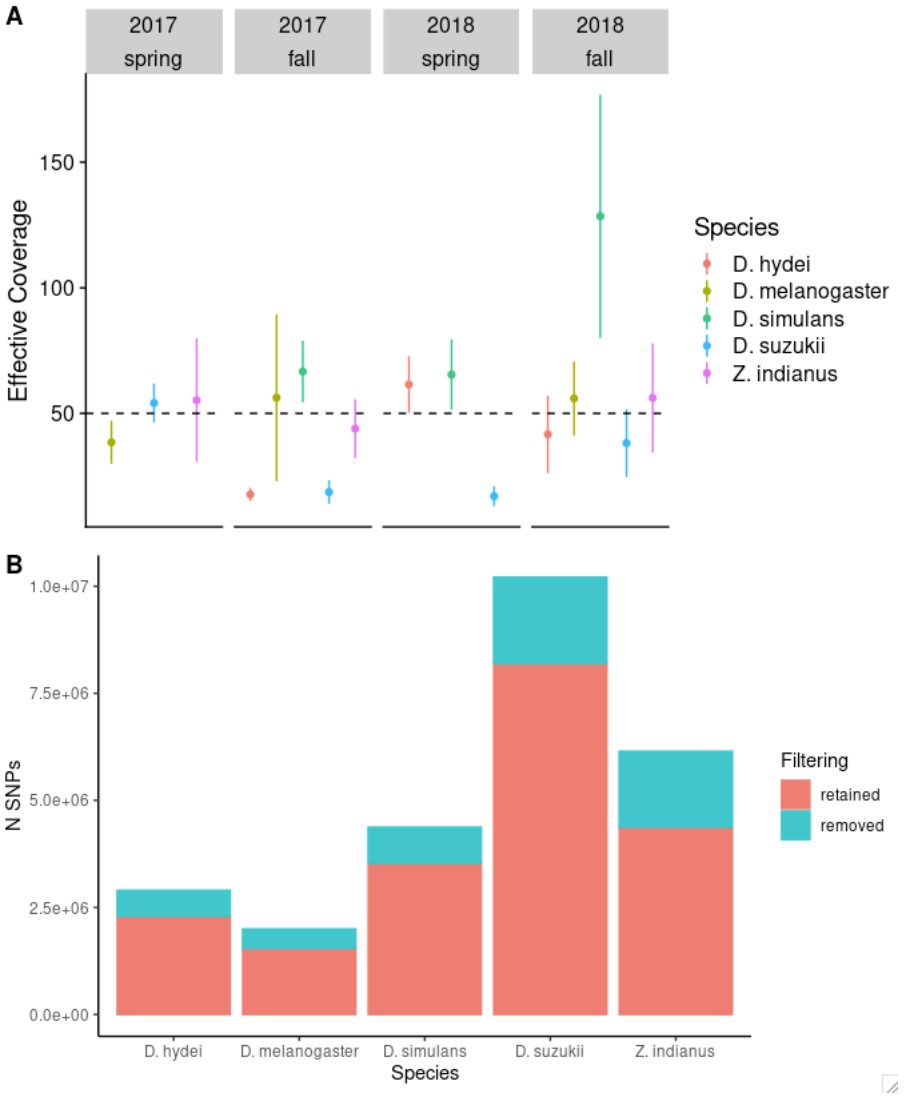


Figure 2*: Samples are scattered across the four time points with good coverage and number of SNPs retained post-filtering.* A) Each point represents the mean and +/- 1 standard deviation of effective coverage across all SNPs per sample. Colors demarcate the species. The target effective coverage was 50 (dashed line). B) The number of SNPs removed (blue) and retained (red) for the final datasets during filtering.

F_{ST} through time

Whole-genome estimates of pairwise F_{ST} over time show a sharp divide between species (Figure 3). *D. simulans* shows a relatively large estimate of F_{ST} between samples collected in different years, while within-year F_{ST} is low. The other four species show a

very different pattern with no large increase in F_{ST} going from within-year to between-year comparisons.

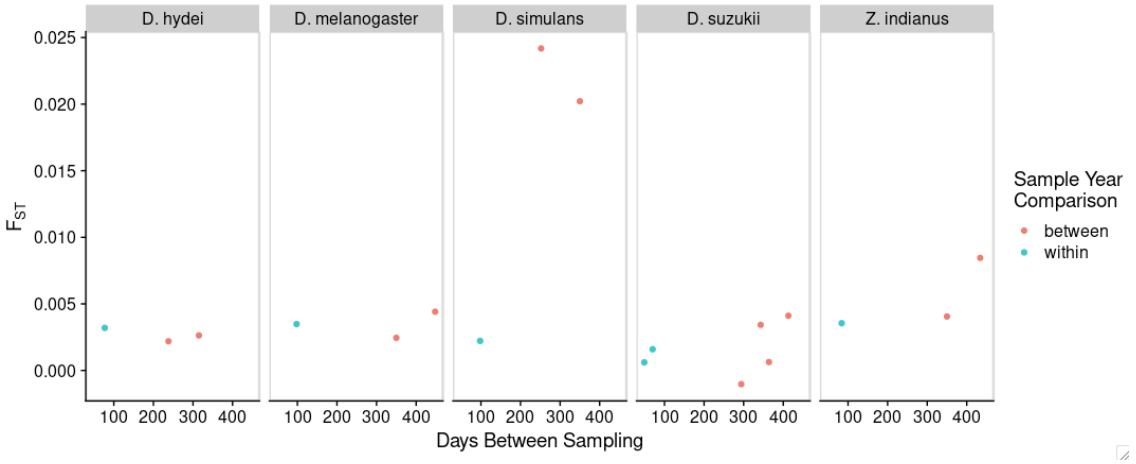


Figure 3: ***Pairwise F_{ST} may reflect the degree of population turnover from year to year.*** Whole genome estimates of pairwise F_{ST} between samples are plotted against the number of days between sampling time points. Samples are color-coded to indicate if the pair of samples being collected were collected in the same year (blue) or in different years (red).

Identifying seasonally oscillating SNPs

I conducted a general linear model of effective coverage corrected allele frequencies per SNP on season of collection (spring vs. fall) for all species where season was encoded as a 0 for spring and a 1 for fall. I identified a subset of SNPs for each species where changes in allele frequencies are significantly associated with season (Table 3; Figure 4). However, following two different multiple-test corrections on the p-values (FDR and Bonferroni) most SNPs were no longer statistically significant (Table 3).

Table 3: ***Counts of SNPs where allele frequency changes are significantly associated with season at different significance thresholds.***

Species	Total Number Tested SNPs	N SNPs p<0.05	N SNPs FDR<0.05	N SNPs FDR<0.1	N SNPs Bonferroni α<0.05
D. melanogaster	1169311	72040	0	0	0
D. simulans	2345168	247340	415	2837	6
D. suzukii	4528392	224524	2	2	1
D. hydei	1510790	88323	8	10	5
Z. indianus	3588541	287021	29	111	6

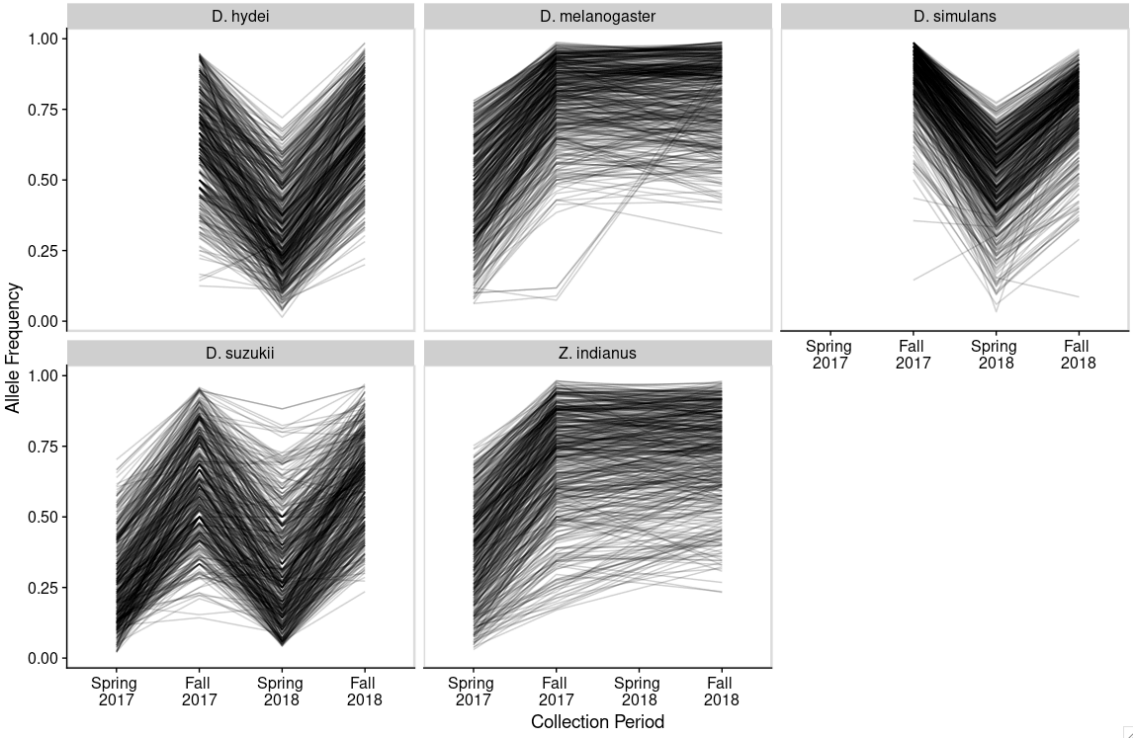


Figure 4: ***Changes in allele frequency in the most significantly seasonally oscillating loci.*** Each panel shows the 500 most significant seasonally oscillating SNPs for that species, prior to any p-value corrections. A single line is a single SNP.

Discussion

Through dense temporal sampling across two years, I find that the composition and relative abundance of Drosophilid species vary dramatically both within and between years. Additionally, I see through pooled sequencing that different species display different degrees of population turnover pointing to some species more likely to be established populations rather than extirpated and recolonized by migrants each year. Finally, while I lacked power to detect seasonal oscillations in allele frequencies with confidence, there was still a signal that seasonal oscillations in allele frequencies may be present in species other than *Drosophila melanogaster*.

Approximately bi-weekly sampling in 2017 and weekly sampling in 2018 showed dramatic shifts in which Drosophilids were abundant within a single year and between years. Interestingly, *Zaprionus indianus* was dominant for a large portion of 2017 and *Drosophila simulans* was dominant for almost all of 2018. These dramatic shifts in species composition are consistent with expectations from community ecology where the species composition of a community is not predictable (Ozinga et al., 2005). Additionally, the sampling was not evenly distributed across the whole orchard, so a lack of capture does not mean zero abundance in the orchard community of Drosophilid species (MacKenzie, 2005). Some of the species I focused on are considered broad generalists, especially *Drosophila suzukii* which is a harmful agricultural pest that can survive on many different substrates and therefore can escape pesticide usage by colonizing other nearby resources when conditions within an orchard become less ideal (Asplen et al., 2015). While the observations of within and between year shifts in species composition and abundance are dramatic, it is not necessarily unexpected.

The observed changes in species composition and abundance could potentially alter the degree of interspecific competition from month to month and from year to year. Interspecific competition in *D. melanogaster* has been previously shown to impact the evolutionary trajectory of several traits, including those important to survival in winter conditions (Grainger et al., 2021). The changing biotic environment due to changing interspecific competitive dynamics could influence the shape and trajectory of many functional loci within a year and between years, contributing to patterns of seasonal adaptation and balancing selection. In future studies trying to identify drivers of seasonality in *D. melanogaster*, it may be important to ensure broad sampling of all present Drosophilid species to be able to understand the competitive environment and tie present species to loci under selection.

In the pooled sequencing performed, I showed that different species exhibit differing degrees of population turnover using pairwise F_{ST} through time. For *D. melanogaster* and *D. simulans*, the observations align with what is known about the different species. *D. melanogaster* is known to have established populations along the entire North American east coast, and therefore large degrees of population turnover are not expected between years. In contrast, *D. simulans* is known to be extirpated in the winter and then recolonized through mass northward migration each year, which is reflected in the sharp jump in F_{ST} when looking at between-year sample comparisons (Machado et al., 2016). *D. hydei* is known to have good thermal performance at both low and high temperatures (Gibert, Moreteau, & David, 2000; Mitchell & Hoffmann, 2010; Overgaard, Hoffmann, & Kristensen, 2011; Overgaard, Kristensen, et al., 2011), which could be indicative of the ability to persist as an established population. *D. suzukii*, is also

expected to overwinter and have established populations as they exhibit comparable thermal limits to *D. melanogaster* (Asplen et al., 2015), have a female reproductive diapause in winter conditions (Wallingford & Loeb, 2016; Zhai et al., 2016), and are believed to have the ability to overwinter in woodland leaf litter (Pelton et al., 2016). In contrast, *Z. indianus* is expected to be extirpated and recolonized every year similar to *D. simulans*, but here it lacks a strong signal of population turnover. The evidence for the predicted extirpation and recolonization in *Z. indianus* is limited, but rooted in limited adaptation to cooler temperate environments and species abundance observations that show it disappearing from Drosophilid collections at the onset of fall and winter (Comeault et al., 2020; Gleason et al., 2019; Joshi et al., 2014). However, none of the existing evidence can rule out any overwintering strategy or recent adaptation. *Z. indianus* is a recent invasive species in North America with an abundance genetic variation (Comeault et al., 2020), so I cannot rule out that it has started to establish more permanent refugia or populations further north. This evidence of the lack of population turnover in some Drosophilid species adds evidence for the overwintering and establishment in temperature environments, which is important knowledge to have for agricultural pests and how these invasive species are adapting to a new temperate environment.

Due to low sample sizes and therefore low power, I was unable to confidently identify seasonally oscillating SNPs. I did see some moderate signal with some species showing in the range of single-digit to thousands of significantly seasonally oscillating SNPs. Given the limited sampling, more sampling and sequencing would need to be done to obtain enough power to detect high-confidence seasonally oscillating SNPs

comparable to Bergland et al., 2014. Given that many of these species appear to be established populations like *D. melanogaster*, these other species may also respond to changing abiotic factors. Adaptation to changing abiotic conditions can occur through plasticity rather than adaptive tracking. Lab-based studies into how seasonal traits like chill coma recovery or body size could determine how many survival-related traits change plastically versus through fixed differences in seasonally spaced samples, if they change at all, in these other species.

In *D. melanogaster*, hundreds of loci change in allele frequency corresponding with the seasonal change (Bergland et al., 2014). *D. simulans*, while not established as permanent populations in northeast North America, still shows clinal variation in SNPs that also vary clinally in *D. melanogaster* (Machado et al., 2016). Similar patterns might exist in seasonal samples of *D. simulans* where SNPs found to oscillate seasonally in *D. melanogaster* may oscillate seasonally despite not experiencing the same balancing selection that *D. melanogaster* experiences through a continually evolving population that overwinters. Cold temperatures and a winter photoperiod can plastically induce a winter morph in *D. suzukii* in lab experiments. The *D. suzukii* winter morph confers greater cold temperature tolerance (Shearer et al., 2016), signaling that perhaps plasticity might drive adaptation to temperate environments for this invasive species. With the other species, only lab tests of thermal limits have been done, and not enough is known about if any potential adaptation to temperate environments could be driven by plasticity or facilitated by adaptive tracking (but see Comeault et al., 2020).

Overall, I have found that the species composition of the focal orchard community dramatically changes through time, thus providing insight into a potential

biotic driver of selection that also varies temporally for *D. melanogaster*. The observation that most species appear to lack the dramatic population turnover found in a species that is extirpated and recolonized each year indicates that Drosophilid species other than *D. melanogaster* may overwinter and continually adapt to seasonally varying selection. However, I cannot know for certain if adaptation to seasonality in other Drosophilid species acts through plasticity or adaptive tracking without further sampling and experimentation.

Data Availability

All scripts and some information tables can be found at https://github.com/abangerter/multiSpecies_seasonality

All raw sequencing data can be found under NCBI SRA BioProject ID PRJNA728438. Data is embargoed until published through peer-review.

Supplemental Tables

Supplemental Table 1: ***Counts of how many individuals of different species were collected on each sampling date.*** Counts of *D. melanogaster* and *D. simulans* only include male individuals while the counts for other species include both male and female individuals.

Species	Year	Collection Date (MM/DD)	Total Count
Zaprionus indianus	2017	06/15	1
Chymomyza amoena	2017	06/15	5
Drosophila melanogaster	2017	06/15	7
unknown	2017	06/15	7
Drosophila tripunctata	2017	06/15	1
Drosophila suzukii	2017	06/22	65
Drosophila immigrans	2017	06/22	124
unknown	2017	06/22	50
Drosophila hydei	2017	06/22	18
Zaprionus indianus	2017	06/22	7
Drosophila tripunctata	2017	06/22	68
Chymomyza amoena	2017	06/22	37
Drosophila melanogaster	2017	06/22	80
Chymomyza amoena	2017	07/07	3
Drosophila tripunctata	2017	07/07	60
Drosophila hydei	2017	07/07	71
unknown	2017	07/07	30
Zaprionus indianus	2017	07/07	15
Drosophila immigrans	2017	07/07	223
Drosophila suzukii	2017	07/07	123
Drosophila melanogaster	2017	07/07	145
Drosophila suzukii	2017	07/20	105
Zaprionus indianus	2017	07/20	88
Drosophila immigrans	2017	07/20	191
Drosophila hydei	2017	07/20	76
Chymomyza amoena	2017	07/20	3
Drosophila tripunctata	2017	07/20	7
unknown	2017	07/20	8
Drosophila simulans	2017	07/20	5
Drosophila melanogaster	2017	07/20	133
Zaprionus indianus	2017	08/03	191

Drosophila suzukii	2017	08/03	289
Drosophila immigrans	2017	08/03	79
Drosophila hydei	2017	08/03	9
Drosophila tripunctata	2017	08/03	1
Chymomyza amoena	2017	08/03	24
unknown	2017	08/03	28
Drosophila simulans	2017	08/03	6
Drosophila melanogaster	2017	08/03	53
Zaprionus indianus	2017	08/17	425
Drosophila immigrans	2017	08/17	95
Drosophila suzukii	2017	08/17	227
unknown	2017	08/17	57
Drosophila tripunctata	2017	08/17	8
Chymomyza amoena	2017	08/17	44
Drosophila hydei	2017	08/17	12
Drosophila simulans	2017	08/17	13
Drosophila melanogaster	2017	08/17	86
unknown	2017	08/31	38
Drosophila suzukii	2017	08/31	158
Drosophila immigrans	2017	08/31	12
Drosophila tripunctata	2017	08/31	7
Drosophila hydei	2017	08/31	19
Zaprionus indianus	2017	08/31	680
Drosophila simulans	2017	08/31	27
Drosophila melanogaster	2017	08/31	125
Chymomyza amoena	2017	08/31	13
Zaprionus indianus	2017	09/14	565
Drosophila suzukii	2017	09/14	217
Chymomyza amoena	2017	09/14	18
Drosophila immigrans	2017	09/14	2
Drosophila hydei	2017	09/14	17
Drosophila tripunctata	2017	09/14	4
unknown	2017	09/14	50
Drosophila simulans	2017	09/14	27
Drosophila melanogaster	2017	09/14	94
Zaprionus indianus	2017	09/28	269
Drosophila immigrans	2017	09/28	2
Drosophila suzukii	2017	09/28	36

Drosophila hydei	2017	09/28	8
unknown	2017	09/28	4
Chymomyza amoena	2017	09/28	1
Drosophila simulans	2017	09/28	24
Drosophila melanogaster	2017	09/28	204
Zaprionus indianus	2017	10/12	444
Drosophila suzukii	2017	10/12	114
unknown	2017	10/12	7
Drosophila hydei	2017	10/12	10
Drosophila simulans	2017	10/12	26
Drosophila melanogaster	2017	10/12	193
Drosophila immigrans	2017	10/12	2
Zaprionus indianus	2017	10/26	126
Drosophila hydei	2017	10/26	39
Drosophila suzukii	2017	10/26	3
Drosophila simulans	2017	10/26	67
Drosophila melanogaster	2017	10/26	333
unknown	2017	10/26	6
Zaprionus indianus	2017	11/09	35
Drosophila hydei	2017	11/09	21
Drosophila melanogaster	2017	11/09	230
Drosophila simulans	2017	11/09	117
Drosophila hydei	2017	11/24	10
Drosophila immigrans	2017	11/24	3
Drosophila suzukii	2017	11/24	3
Chymomyza amoena	2017	11/24	1
Drosophila melanogaster	2017	11/24	28
Drosophila simulans	2017	11/24	10
Drosophila tripunctata	2017	11/24	1
Drosophila suzukii	2017	12/07	1
Drosophila hydei	2017	12/07	1
Drosophila melanogaster	2017	12/07	12
Drosophila simulans	2017	12/07	9
Chymomyza amoena	2018	06/28	12
Drosophila immigrans	2018	06/28	16
Drosophila hydei	2018	06/28	95
Drosophila suzukii	2018	06/28	1
Drosophila tripunctata	2018	06/28	1

Drosophila simulans	2018	06/28	5
Drosophila melanogaster	2018	06/28	10
Drosophila hydei	2018	07/05	64
Drosophila immigrans	2018	07/05	27
Drosophila tripunctata	2018	07/05	24
Drosophila suzukii	2018	07/05	5
Chymomyza amoena	2018	07/05	7
Drosophila simulans	2018	07/05	92
Zaprionus indianus	2018	07/05	2
Drosophila hydei	2018	07/12	199
Drosophila immigrans	2018	07/12	107
Drosophila tripunctata	2018	07/12	21
Drosophila suzukii	2018	07/12	55
Chymomyza amoena	2018	07/12	5
Drosophila simulans	2018	07/12	220
Zaprionus indianus	2018	07/12	1
Drosophila suzukii	2018	07/19	32
Drosophila immigrans	2018	07/19	19
Drosophila hydei	2018	07/19	407
Drosophila simulans	2018	07/19	210
Chymomyza amoena	2018	07/19	3
Zaprionus indianus	2018	07/19	1
Drosophila suzukii	2018	07/26	126
Drosophila immigrans	2018	07/26	72
Drosophila hydei	2018	07/26	198
Chymomyza amoena	2018	07/26	11
Drosophila tripunctata	2018	07/26	10
Zaprionus indianus	2018	07/26	3
Drosophila simulans	2018	07/26	585
Drosophila melanogaster	2018	07/26	83
Drosophila simulans	2018	08/02	356
Drosophila hydei	2018	08/02	86
Zaprionus indianus	2018	08/02	1
Drosophila suzukii	2018	08/02	22
Drosophila immigrans	2018	08/02	27
Drosophila melanogaster	2018	08/02	55
Drosophila simulans	2018	08/09	1298
Drosophila tripunctata	2018	08/09	1

Drosophila suzukii	2018	08/09	67
Drosophila immigrans	2018	08/09	247
Drosophila hydei	2018	08/09	342
Zaprionus indianus	2018	08/09	48
Drosophila melanogaster	2018	08/09	248
Drosophila simulans	2018	08/16	728
Drosophila immigrans	2018	08/16	111
Drosophila hydei	2018	08/16	90
Drosophila suzukii	2018	08/16	29
Drosophila tripunctata	2018	08/16	3
Zaprionus indianus	2018	08/16	42
Drosophila melanogaster	2018	08/16	78
Drosophila immigrans	2018	08/23	53
Drosophila suzukii	2018	08/23	56
Drosophila hydei	2018	08/23	218
Drosophila simulans	2018	08/23	451
Zaprionus indianus	2018	08/23	58
Drosophila melanogaster	2018	08/23	123
Chymomyza amoena	2018	08/23	20
Drosophila hydei	2018	08/30	52
Zaprionus indianus	2018	08/30	100
Drosophila immigrans	2018	08/30	36
Drosophila simulans	2018	08/30	178
Drosophila melanogaster	2018	08/30	48
Drosophila suzukii	2018	08/30	20
Drosophila hydei	2018	09/06	158
Drosophila immigrans	2018	09/06	76
Chymomyza amoena	2018	09/06	2
Drosophila suzukii	2018	09/06	47
Zaprionus indianus	2018	09/06	190
Drosophila simulans	2018	09/06	555
Drosophila melanogaster	2018	09/06	64
Zaprionus indianus	2018	09/13	83
Drosophila hydei	2018	09/13	76
Drosophila suzukii	2018	09/13	17
Drosophila immigrans	2018	09/13	13
Chymomyza amoena	2018	09/13	14
Drosophila tripunctata	2018	09/13	6

Drosophila simulans	2018	09/13	76
Drosophila melanogaster	2018	09/13	47
Drosophila hydei	2018	09/20	50
Zaprionus indianus	2018	09/20	131
Drosophila tripunctata	2018	09/20	2
Chymomyza amoena	2018	09/20	7
Drosophila immigrans	2018	09/20	2
Drosophila suzukii	2018	09/20	6
Drosophila simulans	2018	0920	239
Drosophila melanogaster	2018	09/20	161
Drosophila hydei	2018	09/27	36
Zaprionus indianus	2018	09/27	84
Drosophila immigrans	2018	09/27	2
Chymomyza amoena	2018	09/27	2
Drosophila simulans	2018	09/27	178
Drosophila melanogaster	2018	09/27	49
Drosophila hydei	2018	10/04	57
Chymomyza amoena	2018	10/04	2
Drosophila immigrans	2018	10/04	5
Drosophila suzukii	2018	10/04	6
Zaprionus indianus	2018	10/04	342
Drosophila simulans	2018	10/04	440
Drosophila melanogaster	2018	10/04	141
Drosophila tripunctata	2018	10/04	2
Zaprionus indianus	2018	10/11	379
Drosophila hydei	2018	10/11	5
Drosophila tripunctata	2018	10/11	1
Drosophila suzukii	2018	10/11	3
Chymomyza amoena	2018	10/11	1
Drosophila simulans	2018	10/11	327
Drosophila melanogaster	2018	10/11	89
Zaprionus indianus	2018	10/18	20
Chymomyza amoena	2018	10/18	1
Drosophila hydei	2018	10/18	13
Drosophila immigrans	2018	10/18	1
Drosophila simulans	2018	10/18	721
Drosophila melanogaster	2018	10/18	203
Drosophila hydei	2018	10/25	60

Zaprionus indianus	2018	10/25	53
Drosophila suzukii	2018	10/25	1
Drosophila immigrans	2018	10/25	3
Drosophila simulans	2018	10/25	507
Drosophila melanogaster	2018	10/25	180
Zaprionus indianus	2018	11/01	263
Drosophila immigrans	2018	11/01	15
Drosophila hydei	2018	11/01	10
Drosophila suzukii	2018	11/01	10
Drosophila simulans	2018	11/01	460
Drosophila melanogaster	2018	11/01	192
unknown	2018	11/17	2
Drosophila hydei	2018	11/17	69
Drosophila immigrans	2018	11/17	5
Chymomyza amoena	2018	11/17	1
Zaprionus indianus	2018	11/17	6
Drosophila suzukii	2018	11/17	1
Drosophila melanogaster	2018	11/17	170
Drosophila simulans	2018	11/17	237
Chymomyza amoena	2018	11/29	1
Drosophila melanogaster	2018	11/29	27
Drosophila simulans	2018	11/29	25

Literature Cited

Asplen, M. K., Anfora, G., Biondi, A., Choi, D.-S., Chu, D., Daane, K. M., Gibert, P., et al. (2015). Invasion biology of spotted wing Drosophila (*Drosophila suzukii*): a global perspective and future priorities. *Journal of Pest Science*, *88*(3), 469–494.

Behrman, E. L., Howick, V. M., Kapun, M., Staubach, F., Bergland, A. O., Petrov, D. A., Lazzaro, B. P., et al. (2018). Rapid seasonal evolution in innate immunity of wild *Drosophila melanogaster*. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1870), 20172599.

Bergland, A. O., Behrman, E. L., O'Brien, K. R., Schmidt, P. S., & Petrov, D. A. (2014). Genomic Evidence of Rapid and Stable Adaptive Oscillations over Seasonal Time Scales in *Drosophila*. (D. Bolnick, Ed.)*PLoS Genetics*, *10*(11), e1004775-19.

Bolda, M. P., Goodhue, R. E., & Zalom, F. G. (2009). *Spotted Wing Drosophila: Potential Economic Impact of a Newly Established Pest* (pp. 5–8). Giannini Foundation of Agricultural Economics; University of California.

Comeault, A. A., Wang, J., Tittes, S., Isbell, K., Ingley, S., Hurlbert, A. H., & Matute, D. R. (2020). Genetic Diversity and Thermal Performance in Invasive and Native Populations of African Fig Flies. *Molecular Biology and Evolution*, *37*(7), 1893–1906.

Dieckmann, U., & Doebeli, M. (1999). On the origin of species by sympatric speciation. *Nature*, *400*, 354–357.

Gibert, P., Moreteau, B., & David, J. R. (2000). Developmental constraints on an adaptive plasticity: reaction norms of pigmentation in adult segments of *Drosophila melanogaster*. *Evolution and Development*, *2*(5), 249–260.

Gibert, P., Moreteau, B., Ptavy, G., Karan, D., & David, J. R. (2001). Chill-Coma Tolerance, A Major Climatic Adaptation Among *Drosophila* Species. *Evolution*, *55*(5), 1063–1068.

Gleason, J. M., Roy, P. R., Everman, E. R., Gleason, T. C., & Morgan, T. J. (2019). Phenology of *Drosophila* species across a temperate growing season and implications for behavior. (N. Desneux, Ed.)*PLOS ONE*, *14*(5), e0216601-21.

Goodhue, R. E., Bolda, M., Farnsworth, D., Williams, J. C., & Zalom, F. G. (2011). Spotted wing drosophila infestation of California strawberries and raspberries: economic analysis of potential revenue losses and control costs. *Pest Management Science*, *67*(11), 1396–1402.

Grainger, T. N., Rudman, S. M., Schmidt, P., & Levine, J. M. (2021). Competitive history shapes rapid evolution in a seasonal climate. *Proceedings of the National Academy of the Sciences of the United States of America*, *118*(6), 1–9.

Grant, P. R., & Grant, B. R. (2006). Evolution of Character Displacement in Darwin's Finches. *Science*, *313*(5784), 224–226.

Hart, S. P., Turcotte, M. M., & Levine, J. M. (2019). Effects of rapid evolution on species coexistence. *Proceedings of the National Academy of Sciences*, *116*(6), 2112–2117.

Hivert, V., Leblois, R., Petit, E. J., Gautier, M., & Vitalis, R. (2018). Measuring Genetic Differentiation from Pool-seq Data. *Genetics*, genetics.300900.2018-51.

Hoskins, R. A., Carlson, J. W., Wan, K. H., Park, S., Mendez, I., Galle, S. E., Booth, B. W., et al. (2015). The Release 6 reference sequence of the *Drosophila melanogaster* genome. *Genome Research*, *25*(3), 445–458.

Iglesias, L. E., Price, J. F., Roubos, C. R., Renkema, J. M., & Liburd, O. E. (2016). *Spotted Wing Drosophila in Florida Berry Culture*. University of Florida IFAS Extension.

Ives, P. T. (1970). Further Genetic Studies of the South Amherst Population of *Drosophila melanogaster*. *Evolution*, *24*(3), 507–518.

Joshi, A., & Thompson, J. N. (1995). Alternative Routes to the Evolution of Competitive Ability in Two Competing Species of *Drosophila*. *Evolution*, *49*(4), 616–625.

Joshi, N. K., Biddinger, D. J., Demchak, K., & Deppen, A. (2014). First Report of *Zaprionus indianus* (Diptera: Drosophilidae) in Commercial Fruits and Vegetables in Pennsylvania. *Journal of Insect Science*, *14*(1), 259.

Koboldt, D. C., Zhang, Q., Larson, D. E., Shen, D., McLellan, M. D., Lin, L., Miller, C. A., et al. (2012). VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Research*, *22*(3), 568–576.

Kolaczkowski, B., Kern, A. D., Holloway, A. K., & Begun, D. J. (2011). Genomic differentiation between temperate and tropical Australian populations of *Drosophila melanogaster*. *Genetics*, *187*(1), 245–260.

Li, H, & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, *26*(5), 589–595.

Li, Heng. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM, 1–3.

Linde, K. V. der, Steck, G. J., Hibbard, K., Birdsley, J. S., Alonso, L. M., & Houle, D. (2006). First records of *Zaprionus indianus* (Diptera: Drosophilidae), a pest species on commercial fruits from Panama and the United States of America. *Florida Entomologist*, *89*(3), 402–404.

Machado, H., Bergland, A. O., Taylor, R., Tilk, S., Behrman, E., Dyer, K., Fabian, D., et al. (2021). Broad geographic sampling reveals the shared basis and environmental correlates of seasonal adaptation in *Drosophila. eLife*, *10*(e67577), 1–21.

Machado, H. E., Bergland, A. O., O'Brien, K. R., Behrman, E. L., Schmidt, P. S., & Petrov, D. A. (2016). Comparative population genomics of latitudinal variation in *Drosophila simulans* and *Drosophila melanogaster*. *Molecular Ecology*, *25*(3), 723–740.

MacKenzie, D. I. (2005). What are the issues with presence-absence data for wildlife managers? *Journal of Wildlife Management*, *69*(3), 849–860.

Markow, T. A. (1985). A comparative investigation of the mating system of *Drosophila hydei*. *Animal Behaviour*, *33*(3), 775–781.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, *17*(1), 10–12.

Mitchell, K. A., & Hoffmann, A. A. (2010). Thermal ramping rate influences evolutionary potential and species differences for upper thermal limits in *Drosophila*. *Functional Ecology*, *24*(3), 694–700.

Overgaard, J., Hoffmann, A. A., & Kristensen, T. N. (2011). Assessing population and environmental effects on thermal resistance in *Drosophila melanogaster* using ecologically relevant assays. *Journal of Thermal Biology*, *36*(7), 409–416.

Overgaard, J., Kristensen, T. N., Mitchell, K. A., & Hoffmann, A. A. (2011). Thermal Tolerance in Widespread and Tropical Drosophila Species: Does Phenotypic Plasticity Increase with Latitude? *The American Naturalist*, *178*(S1), S80–S96.

Ozinga, W. A., Schaminée, J. H. J., Bekker, R. M., Bonn, S., Poschlod, P., Tackenberg, O., Bakker, J., et al. (2005). Predictability of plant species composition from environmental conditions is constrained by dispersal limitation. *OIKOS*, *108*(3), 555–561.

Paris, M., Boyer, R., Jaenichen, R., Wolf, J., Karageorgi, M., Green, J., Cagnon, M., et al. (2020). Near-chromosome level genome assembly of the fruit pest *Drosophila suzukii* using long-read sequencing. *Scientific Reports*, *10*(1), 11227.

Pelton, E., Gratton, C., Isaacs, R., Timmeren, S. V., Blanton, A., & Guédot, C. (2016). Earlier activity of *Drosophila suzukii* in high woodland landscapes but relative abundance is unaffected. *Journal of Pest Science*, *89*(3), 725–733.

Pitnick, S., Markow, T. A., & Spicer, G. S. (1995). Delayed male maturity is a cost of producing large sperm in *Drosophila*. *Proceedings of the National Academy of Sciences of the United States of America*, *92*, 10614–10618.

Rudman, S. M., Greenblum, S., Hughes, R. C., Rajpurohit, S., Kiratli, O., Lowder, D. B., Lemmon, S. G., et al. (2019). Microbiome composition shapes rapid genomic adaptation of *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences*, *116*(40), 20025–20032.

Shearer, P. W., West, J. D., Walton, V. M., Brown, P. H., Svetec, N., & Chiu, J. C. (2016). Seasonal cues induce phenotypic plasticity of *Drosophila suzukii* to enhance winter survival. *BMC Ecology*, *16*(1), 11.

TerHorst, C. P. (2011). Experimental evolution of protozoan traits in response to interspecific competition. *Journal of Evolutionary Biology*, *24*(1), 36–46.

Wallingford, A. K., & Loeb, G. M. (2016). Developmental Acclimation of *Drosophila suzukii* (Diptera: Drosophilidae) and Its Effect on Diapause and Winter Stress Tolerance. *Environmental Entomology*, *45*(4), 1081–1089.

Weiher, E., Freund, D., Bunton, T., Stefanski, A., Lee, T., & Bentivenga, S. (2011). Advances, challenges and a developing synthesis of ecological community assembly

theory. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1576), 2403–2413.

Werner, T., & Jaenike, J. (2017). *Drosophilids of the Midwest and Northeast* (1st ed.). Open Access Books.

Yoshimoto, C. M. (1952). *Studies on the Species of Vinegar Gnats (Diptera: Drosophilidae) in Kansas [MS Thesis]*.

Zhai, Y., Lin, Q., Zhang, J., Zhang, F., Zheng, L., & Yu, Y. (2016). Adult reproductive diapause in *Drosophila suzukii* females. *Journal of Pest Science*, *89*(3), 679–688.