Natural Selection in *Drosophila melanogaster:* A New Detection Method, Impact on Demographic Inference, and Short-Term Evolution

By
Jeremy D. Lange

A dissertation submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy
(Genetics)

at the

UNIVERSITY OF WISCONSIN – MADISON

2021

Date of final oral examination: 01/13/2021
The dissertation is approved by the following members of the Final Oral Committee:
John E. Pool, Associate Professor, Genetics
Bret A. Payseur, Professor, Genetics
Sean D. Schoville, Associate Professor, Entomology
Karl W. Broman, Professor, Biostatistics and Medical Informatics
Michael A. Newton, Professor, Biostatistics and Medical Informatics

Acknowledgements

The Ph.D. is not a solo journey: it takes a village. I have been truly fortunate to have a village of brilliant, supportive, and encouraging people guiding me along the way. I would first like to thank Dr. John Pool for his incredible mentorship. Dr. Pool is the one who first suggested that I was capable of pursuing a Ph.D. in genetics. What started as an undergraduate job tending the fly stocks in Dr. Pool's lab developed into a passion for population genetics, and ultimately (what I imagine will be) a fulfilling career as a bioinformatic scientist. I am thankful for the countless hours Dr. Pool has spent patiently teaching, mentoring, and guiding me over the years. I owe much to Dr. Pool for helping me develop into the scientist and, more importantly, the person I am today. I am profoundly grateful to be a recipient of his time and talents throughout my undergraduate and graduate careers.

To my doctoral committee, thank you for your ongoing mentorship, guidance, and support. It has been an honor working with so many brilliant minds. I have appreciated the time you have spent reviewing my work and offering feedback. I would also like to acknowledge my various collaborators over the years who have graciously shared their life's work with me so that I could learn and benefit from their knowledge. I am indebted to my peers in the Pool Lab who have all been exceedingly helpful and supportive of my work. I would particularly like to thank Justin Lack, Amir Yassin, Heloise Bastide, Yuheng Huang, Quentin Sprengelmeyer, Chris McAllester, Tiago Ribeiro, and Matt Lollar.

To my life partner, Mara Stewart, my parents, my family, and the Stewart and Bradley families (including Barley, my favorite goldendoodle), thank you for your endless love and encouragement as I pursued this dream. Stressful years of working on my Ph.D. were offset by

enjoyable times with all of you. I am looking forward to having more time for visits to Minnesota, card games, and Boundary Waters camping trips now.

And finally, I would like to acknowledge all the educators, coaches, music teachers, leaders, and mentors throughout my life who have worked tirelessly and often thanklessly so that I could succeed in earning a Ph.D. It takes a truly selfless person to dedicate their life to the success of others, and I am fortunate to have crossed paths with many people who so readily extended this selflessness to me.

Abstract

Understanding how natural selection works in nature has been a goal of population geneticists for many decades. This thesis offers an exploration of natural selection in the fruit fly *Drosophila melanogaster*. In Chapter 1, we present a novel haplotype statistic that assesses whether pairwise haplotype sharing at a locus in one population is unusually large compared with another population relative to genome-wide trends. Using simulation of *Drosophila*-like parameters, we show that this statistic has power to detect both hard and soft selective sweeps. We demonstrate that its broad utility and computational simplicity makes this a valuable tool to discover instances of recent adaptation.

In Chapter 2, we examine the effects of recurrent hitchhiking on demographic inference. We show that neutralist assumptions made by a common demographic inference method is indeed biased by high rates of natural selection, but such biases are weaker for parameters relating recently diverged populations, resolving the utility of estimated demographies.

In Chapters 3 and 4, we utilize temporal genetic sampling to study the population genomics of two different populations of *D. melanogaster*. Studying temporal changes in allele frequencies can better illuminate the role of natural selection on very short time scales. In the first of these studies, the subject of Chapter 3, we use whole genome sequencing of isofemale *D. melanogaster* lines originally collected 35 years ago and compare genetic variation to modern samples collected from the same location. We reveal recent targets of adaptation to insecticide resistance alleles and uncover a shift toward Northern-associated alleles at well-studied clinal SNPs, possibly due to continued local adaptation favoring alleles of European ancestry in this relatively cool environment.

In a second study, the subject of Chapter 4, we analyze genomic data collected from eight museum specimens collected in the 1840s. Comparing these samples with modern populations, we reveal potential targets of recent adaptation, and again find evidence of adaptation of resistance to insecticides. We also show limited evidence that inversions may have been at a lower frequency than modern populations, giving additional evidence to the hypothesis that inversions are a more recent arrival into modern European populations.

Table of Contents

Introduction

Population genetics offers the theoretical framework to study gene frequencies in populations and make important inferences about the neutral and non-neutral forces shaping genetic variation. As a field, population genetics grew in the first half of the 20th century, long before the sequencing revolution. With limited empirical data, the earliest population geneticists developed and refined mathematical models that laid the foundation for future scientists. Now, armed with whole genome datasets, population geneticists have the ability to test and further refine the models originally designed by early titans of the field.

The field of population genetics has proven to be critically important in the advancement of modern society. In medicine, for instance, population genetics can be used to help map diseases and can offer important insights into allele frequency distributions for researchers in the field of personalized medicine. In agriculture, population genetics can be used to identify natural genetic variation that can be brought into a breeding program, or to help identify new avenues for pest control. Possibly most relevant in 2020 is the use of population genetics to identify risk factors for severe reactions to Covid-19 (Zeberg and Pääbo 2020) or to infer selection underlying the evolution of SARS-CoV-2 (Cagliani *et al.* 2020).

The work described in this thesis uses the fruit fly, *Drosophila melanogaster,* as a model organism to help answer fundamental questions about how evolution works in nature. *Drosophila melanogaster* is an invaluable tool in the advancement of the field of population genetics. Its compact genome (~120 Megabases) means that sequencing an entire genome is cost effective. The short generation time of *D. melanogaster*, about 300-fold shorter than humans, allows for substantially more evolution on a significantly shorter time scale. Further, the species' recent expansion out of its ancestral sub-Saharan African range and into more diverse

environments provides excellent case studies for local adaptation. Finally, the well-annotated genome and many genomic tools developed for *D. melanogaster* provide ample opportunity for downstream functional analyses after identifying recent adaptation.

It is essential to acknowledge that this thesis is part of a much broader scientific context. The work described here builds on both past and contemporary research and adds to our understanding of population genetics in a new and meaningful way. While there is still much to learn, the work described in this thesis begins to answer some important scientific questions, including the ones discussed below:

How effectively can we separate natural selection from neutral forces?

There are well-studied models on which natural selection acts (Figure 1). The first is selection on a new mutation, resulting in a single haplotype rising in frequency. This phenomenon, known as a "hard sweep," has a strong effect on linked variation and leaves a much more discernible pattern on the genome. Natural selection can also act on recurrent mutations or on standing genetic variation. Also known as a "soft sweep," multiple haplotypes rise in frequency, reducing genetic variation relative to neutrality but less so compared to a hard selective sweep. The relative frequencies and importance of these two major models is intensely debated in the field today (e.g., Jensen 2014, Harris *et al*. 2018, Feder *at al*. 2020) and is not a focus of this dissertation.

Many methods exist that attempt to identify targets of recent adaptation, including site frequency-based statistics (Kim and Stephan 2002) and diversity-based statistics (Schlötterer & Dieringer 2005). Perhaps most popular in the field today are haplotype-based statistics (Voight *et al*. 2006, Sabeti *et al.* 2007, Garud *et al*. 2015, Schrider and Kern 2016), which rely on long range linkage disequilibrium between sites to identify recent targets of adaptation.

In Chapter 1 of this dissertation, we present a novel haplotype statistic that assesses whether pairwise haplotype sharing at a locus in one population is unusually large compared with another population relative to genome-wide trends. This statistic relies on a signal of reduced genetic diversity that is caused by a selective sweep. Using simulation, we show that this statistic has power to detect both hard and soft selective sweeps, and outperforms other haplotype and allele frequency-based statistics in a variety of challenging scenarios. Its simplicity and broad utility make it a compelling and useful tool compared to other published statistics.

Selectively neutral events, such as population bottlenecks, also decrease genetic diversity, potentially obscuring signals of adaptation. In typical inference pipelines, researchers leverage neutral genetic variation to estimate demographic change and determine whether genetic variation at a particular region of the genome can be explained using neutral demographic simulation. The impact of natural selection on genomic diversity may be particularly significant for species with very large population sizes, such as *Drosophila melanogaster*. In abundant taxa, the population adaptive mutation rate is elevated, and the weak influence of genetic drift may allow natural selection to favor alleles with modest selection coefficients. Therefore, in populous species, the demography estimated from supposedly neutral data may be biased.

We extensively explore this bias in Chapter 2 of this thesis. We simulate different models of recurrent hitchhiking (RHH) - in which selective sweeps occur at a given rate randomly across the genome - and demonstrate that these RHH scenarios can bias demographic parameter estimation. However, we also find that such biases are weaker for parameters relating recently diverged populations. The weak bias observed give credibility to the utility of demographic models estimated using neutralist assumptions.

What types of genes are targeted by selection in a natural population? And how much evidence for selection does a population contain?

Typical genome-wide analyses of genetic variation make inferences across a wide interval of evolutionary time by examining a single snapshot of genetic variation. Adaptation, however, can occur over much shorter ecological timescales in natural populations. In Chapters 3 and 4, we utilize temporal genetic sampling to study the population genomics of two different populations of *D. melanogaster*. Such temporal sampling can help clarify the relative roles of neutral and non-neutral forces on very short time scales. Using whole-genome sequencing, both studies give us an unusually direct way of quantifying change in genomic variation across decades, allowing us to ask important questions about how evolution works in nature in a populous insect species. We measure difference in genetic variation using the Population Branch Statistic (Yi *et al.* 2010), which uses three population samples to quantify genetic differentiation on one population's branch (see Figure 2).

In Chapter 3, we use whole genome NGS data collected from isofemale *D. melanogaster* strains collected in the 1970s and 1980s and compare genetic variation to modern populations. In Chapter 4, we take a similar approach but instead analyze NGS data from museum specimens. Though these studies are similar in nature, they offer distinctly different contributions to the scientific literature. In the first study we examine a population of North American *D. melanogaster,* while in the second study we examine a population from Northern Europe. In both cases, we observe significant changes in genetic differentiation at two well-known insecticide resistance genes, *Cyp6g1* and *Ace*. We also find evidence of distinct genetic change at other insecticide resistance genes, implying that adaptation is not always predictable. We argue that an

enrichment of statistically significant genomic windows may begin to tell us about the frequency of selective sweeps in a population.

The research described in Chapters 3 and 4 are the first of their kind to quantify change in genomic variation across decades in a large population, providing a fundamental understanding of how evolution works in nature. These results will help future researchers more accurately model the genetic basis of adaptation of human-associated insect species, including crop pests and invasive species. These studies also reveal targets of the genome that have undergone selection in recent decades, which could inform more precise investigations into genes associated with insecticide resistance. While we have started to answer some of these important questions, future generations of population geneticists must continue to build on and contextualize the work described in this thesis.

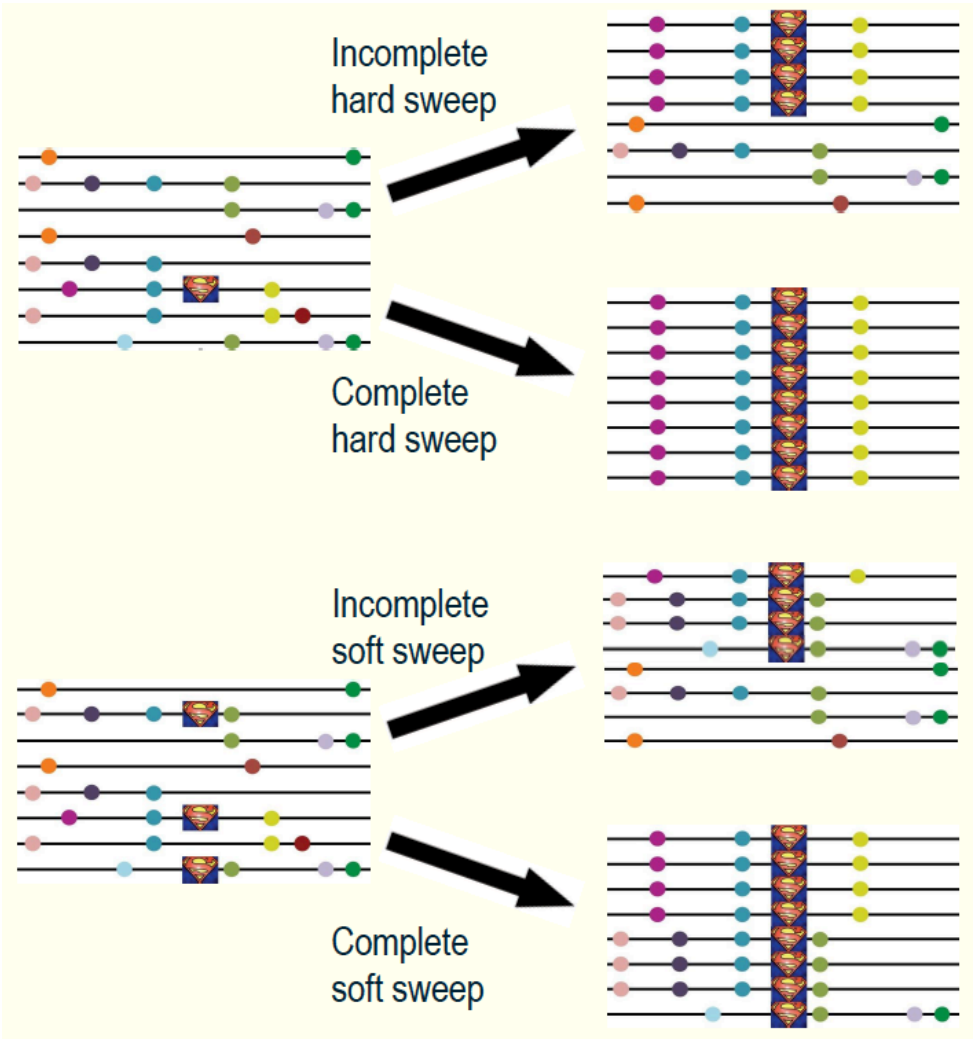


Figure 1: Visual representation of different types of selective sweeps. In a hard sweep, a single haplotype is selected for and rise in frequency. In a soft sweep, a mutation on more than one genetic background is selected for, resulting in multiple haplotypes rising in frequency. Genetic variation is more greatly reduced following a hard sweep relative to a soft sweep.

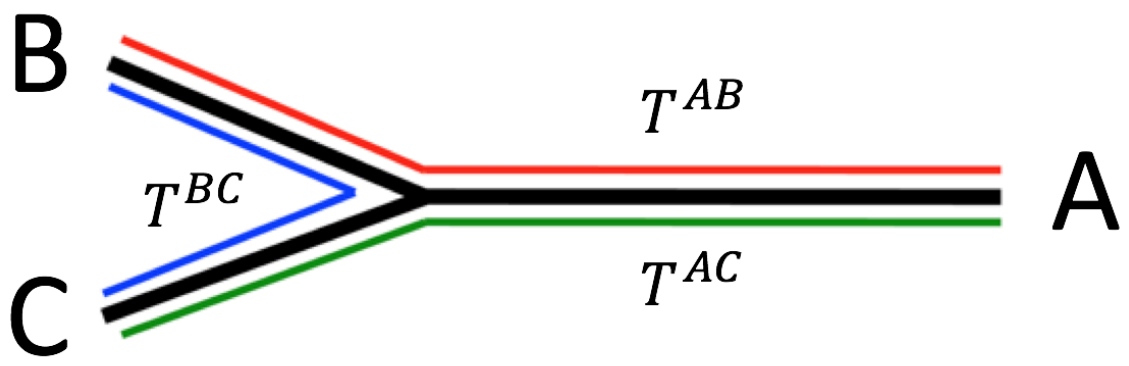


Figure 2: Illustration of the Population Branch Statistic approach to detecting population-specific allele frequency change. Starting with three population samples, we calculate the three pairwise values of F_{ST} between these samples. F_{ST} is not a linear function of genetic differentiation, but it can be transformed into a more additive quantity that scales with population divergence time under a simple isolation model, as $T = -1 - \log(F_{ST})$. Below, our interest is in the length of the branch connecting the sample A to the node representing the ancestor of sample B and C. This quantity can be estimated by adding the length of the red and green branches, subtracting the length of the blue branch, and dividing by two.

References

Feder AF, Pennings PS, Petrov DA. (2020) The clarifying role of time series data in the population genetics of HIV. PLoS Genetics, (in press).

Jensen JD. (2014) On the unfounded enthusiasm for soft selective sweeps. Nature Communications, 5(1), 1-10.

Harris RB, Sackman A, Jensen JD. (2018) On the unfounded enthusiasm for soft selective sweeps II: Examining recent evidence from humans, flies, and viruses. PLoS genetics, 14(12), e1007859.

Zeberg H, Pääbo S. (2020) The major genetic risk factor for severe COVID-19 is inherited from Neanderthals. Nature, 587, 610–612.

Cagliani R, Forni D, Clerici M, Sironi M. (2020) Computational Inference of Selection Underlying the Evolution of the Novel Coronavirus, Severe Acute Respiratory Syndrome Coronavirus 2. Journal of Virology, 94(12).

Kim Y, Stephan W. (2002) Detecting a local signature of genetic hitchhiking along a recombining chromosome. Genetics, 160(2), 765-777.

Schlötterer C, Dieringer D. (2005) A novel test statistic for the identification of local selective sweeps based on microsatellite gene diversity. In Selective Sweep (pp. 55-64). Springer, Boston, MA.

Voight BF, Kudaravalli S, Wen X, Pritchard JK. (2006) A Map of Recent Positive Selection in the Human Genome. PLoS Biol 4(3): e72.

Garud NR, Messer PW, Buzbas EO, Petrov DA. (2015) Recent selective sweeps in North American *Drosophila melanogaster* show signatures of soft sweeps. PLoS Genet. 11: e1005004.

Sabeti PC, Varilly P, Fry B et al. (2007) Genome-wide detection and characterization of positive selection in human populations. Nature, 449, 913–918.

Schrider DR, Kern AD. (2016) S/HIC: Robust identification of soft and hard sweeps using machine learning. PLoS Genetics. 12: e1005928.

Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZXP, Pool JE, *et al*. (2010) Sequencing of fifty human exomes reveals adaptation to high altitude. Science 329:75-78.

Chapter 1: A haplotype method detects diverse scenarios of local adaptation from genomic sequence variation

A version of this chapter has been published in *Molecular Ecology.*

Citation: Lange JD, Pool JE (2016) A haplotype method detects diverse scenarios of local adaptation from genomic sequence variation. Molecular Ecology, 25:3081-3100.

Abstract

Identifying genomic targets of population-specific positive selection is a major goal in several areas of basic and applied biology. However, it is unclear how often such selection should act on new mutations versus standing genetic variation or recurrent mutation, and furthermore, favoured alleles may either become fixed or remain variable in the population. Very few population genetic statistics are sensitive to all of these modes of selection. Here, we introduce and evaluate the Comparative Haplotype Identity statistic (χ_{MD}), which assesses whether pairwise haplotype sharing at a locus in one population is unusually large compared with another population, relative to genomewide trends. Using simulations that emulate human and *Drosophila* genetic variation, we find that χ_{MD} is sensitive to a wide range of selection scenarios, and for some very challenging cases (e.g. partial soft sweeps), it outperforms other two-population statistics. We also find that, as with F_{ST}, our haplotype approach has the ability to detect surprisingly ancient selective sweeps. Particularly for the scenarios resembling human variation, we find that χ_{MD} outperforms other frequency- and haplotype-based statistics for soft and/or partial selective sweeps. Applying χ_{MD} and other between-population statistics to published population genomic data from *D. melanogaster*, we find both shared and unique genes

and functional categories identified by each statistic. The broad utility and computational simplicity of χ_{MD} will make it an especially valuable tool in the search for genes targeted by local adaptation.

Introduction

Detecting instances of population- specific natural selection from patterns of genetic variation is a critically important task in evolutionary biology. Research of this nature has identified genes that contributed to human adaptation to local environments (*e.g*. Yi *et al*. 2010; Fumagalli *et al*. 2011; Hancock *et al*. 2011). In model organisms, adaptive differences between closely related populations offer a promising avenue for uncovering the genetics of adaptation (*e.g*. Rebeiz *et al*. 2009; Will *et al*. 2010). And in species of conservation interest, the identification of adaptive population differences may inform conservation strategies that account for the maintenance of functional genetic diversity (*e.g*. Bonin *et al*. 2007).

Although conventionally referred to as 'local adaptation', causes of population- specific selective sweeps may include ecological adaptation, sexual selection or selfish genetic elements. Comparisons of genetic variation between closely related populations offer a highly promising approach for detecting positive selection. Whereas the power of population genetic tests in a single population is limited by the substantial evolutionary variance expected from one locus to the next under neutrality, comparisons between closely related populations help control for the shared history of the ancestral population. However, stochastic variance may still be a factor even for comparisons of recently diverged populations if a population bottleneck has occurred since their split. In addition to neutral explanations for apparent signals of population- specific selection, such signals may also be produced in the flanking regions of complete sweeps shared between populations (Santiago & Caballero 2005; Roesti *et al*. 2014).

Signatures of positive selection present in one population but not another can be detected through comparisons of diversity levels (*e.g*. Schlötterer & Dieringer 2005), allele frequency differentiation (*e.g*. using F_{ST} and related approaches) and by comparing linkage disequilibrium or haplotype patterns (*e.g*. Sabeti *et al*. 2007; Storz & Kelly 2008). Haplotype statistics have strong potential to detect positive selection, because under a wide range of adaptive scenarios, natural selection causes random pairs of alleles in a population to have recent common ancestry more often than expected under neutrality. This recent common ancestry leaves less time for recombination and mutation events to differentiate the alleles, and hence, they display longer shared haplotypes.

Immediately following a complete hard sweep, all individuals in the population should have haplotype identity for some interval containing the selected site. In the case of a partial/incomplete sweep from a new mutation, a subset of individuals will show the haplotype identity pattern. Hence, haplotype statistics such as *iHS* (integrated haplotype score) and the related *EHH* (extended haplotype homozygosity), which quantify haplotype identity around a focal SNP allele, have been used to detect partial sweeps from human SNP data (Sabeti *et al*. 2002; Voight *et al*. 2006).

Haplotype statistics may also have utility for the detection of soft sweeps, which refer to selective sweeps in which the beneficial allele rises in frequency on more than one haplotype, either because it arose multiple times by mutation, or because it had time to recombine in the population before it became adaptive. Recently, Ferrer- Admetlla *et al*. (2014) found that haplotype statistics including *nSL*, which is analogous to a diversity- scaled *iHS*, can detect soft sweeps in addition to complete and incomplete hard sweeps. While the above statistics analyze a single population, additional power might be obtained from comparing closely related

populations in cases of local adaptation. Indeed, Pennings & Hermisson (2006) suggested that linkage statistics that compare populations might have the best prospects to detect soft sweeps.

Population comparisons of haplotype identity have therefore been utilized in the search for adaptive population differentiation (*e.g*. Fariello *et al*. 2013; Roesti *et al*. 2014). For example, the cross- population *EHH* analysis (*XP- EHH*; Sabeti *et al*. 2007) compares the lengths of identical haplotypes radiating from a focal SNP between two or more populations. *XP-EHH* was presented as a method of detecting population- specific classic sweeps and was found to be reasonably robust to nonequilibrium demographic history. One limitation of this and related approaches is that the power of statistics requiring complete haplotype identity may decay very quickly after a sweep, as new mutation and recombination events begin to occur. A second challenge, especially for genomic resequencing studies, is that SNP- oriented tests become computationally more demanding and produce a larger number of tests when the total number of SNPs is very large (with implications for statistical power if correcting for multiple testing).

We attempt to overcome these challenges by introducing a straightforward, window-based metric called Comparative Haplotype Identity, or χ. The χ statistic sums the lengths of pairwise identical haplotypes that exceed a specified threshold and compares this quantity between two populations (as in Pool & Aquadro 2007). Windows with unusually high haplotype identity in one population compared with the other are candidates for local positive selection. The window approach improves computational efficiency and reduces multiple testing concerns. By excluding rare variation from the analysis, the temporal horizon of the method is substantially extended. In addition to the ability to detect relatively older sweeps, simulations indicate that χ is sensitive to a wide variety of adaptive scenarios, including classic sweeps, sweeps in bottlenecked populations, partial sweeps and soft sweeps.

Materials and Methods

Statistics

In its simplest form, the χ statistic compares the summed length of identical haplotype blocks among individuals in one population versus another, within a particular genomic window. Here, the goal is to identify genomic regions that may have been subject to recent directional selection in population 1, but not in population 2. As natural selection raises the frequency of a beneficial allele more quickly than under genetic drift, chromosomes carrying this allele will have unusually recent common ancestry, implying longer stretches of identical haplotypes where mutation and recombination have not had time to generate haplotype diversity. Hence, a window showing far more haplotype identity in one population compared with another (relative to genome-wide observations for these samples) is a candidate for recent population-specific selection.

First, each pairwise combination of chromosomes in a population sample is evaluated, and the lengths of sequence intervals within the window that are identical between these chromosomes (*i.e*. shared haplotype blocks) are noted. Shared blocks that are longer than a specified threshold length are added to compile the population's summed haplotype sharing. The threshold length is chosen such that it will exceed the average scale of haplotype identity expected under neutrality, although some neutral haplotype sharing beyond this length is acceptable. Stated more formally, for S_k, the sum of haplotype identity for population k, in a sample of n_k chromosomes, indexed by i and j,

$$S_k = \sum_{i=1}^{n_k-1} \sum_{j=i+1}^{n_k} \sum_{1}^{b} H_{L \geq a},$$

where $H_L \geq a$ indicates the length of each of the b identical haplotype blocks between a pair of chromosomes that are greater than or equal to the threshold length a. In this study, we will refer to a in terms of the threshold proportion of total window length that must be identical.

In cases of unequal sample size, the summed haplotype sharing of each population can be made comparable by dividing each sum by the number of pairwise individual comparisons in that population. If missing data is present heterogeneously, the number of pairwise site comparisons in each population can instead be used as the divisor for each population. Here, the proportion of a population's pairwise site comparisons that are part of an identical haplotype block can be written as:

$$P_k = \frac{S_k}{\sum_{i=1}^{n_k-1}\sum_{j=i+1}^{n_k} C},$$

where C is the number of site comparisons (with data present) between individuals i and j. However, these rescalings will not affect a case with uniform sample sizes across windows and no missing data, as investigated under our simulations below. Ultimately, the haplotype sharing of the focal population 1 (for which local selection is being tested) is divided by that of the 'reference' population 2, yielding $\chi = {P_1}/{P_2}$. Ideally, the reference population is closely related to the focal population, but does not share a selective pressure of interest.

Aside from the haplotype length threshold, χ also utilizes an allele frequency threshold to enable the exclusion of variants that are rare across both populations. Because new mutations may quickly disrupt the long identical haplotypes produced by positive selection, their exclusion may significantly extend the temporal signal of haplotype- based neutrality tests. In most of the simulations described below, we specifically exclude singletons (polymorphisms that occur on just one allele across both populations) from the calculation of χ. For a subset of the simulated scenarios, we increased the allele frequency threshold to explore its effects on the power of χ.

Based on preliminary analyses, we noticed that when summed haplotype identity in the reference population had elevated stochastic variation (*e.g.* due to small sample size), outliers for χ could be driven by low values in the denominator (unusually low haplotype identity in the reference population), instead of a high numerator from the focal population. Conceivably, elevated stochastic variance in S_2 might also result from nonequilibrium demography in the reference population. Hence, we also calculated a modified version of χ, applicable for genomic or large multilocus analyses. In this alternative, the focal population's haplotype sharing is divided by the larger of: (i) the reference population's haplotype sharing in this window, or (ii) the median value of the reference population's haplotype sharing across all windows (or in this case, all simulated replicates). We refer to this 'median denominator' version of the statistic as χ_{MD}. Thus,

$$\chi_{MD} = \frac{P_1}{\max\left(P_2, median(P_2)\right)}.$$

Although results for χ are reported, χ_{MD} is the primary focus of the present analysis. In addition to avoiding denominator- driven χ outliers, the median denominator approach also avoids the possibility of an undefined statistic when $P_2 = 0$ (an outcome that could also be circumvented by defining P_2 as having a minimum value equal to the threshold length, but should be uncommon with appropriate choice of threshold and window lengths; see Results and Discussion). Scripts calculating this statistic are available at: https://github.com/jeremy-lange/CHI-Statistic.

We compare the performance of χ and χ_{MD} against two well- known statistics for the detection of local selection. As an indicator of allele frequency differentiation between populations, we evaluate the F_{ST} formulation of Hudson, Slatkin and Maddison

(Hudson *et al*. 1992). As an alternative approach to population haplotype comparisons, we also assess *XP- EHH* (Sabeti *et al*. 2007), as implemented by Pickrell *et al*. (2009).

Simulation Strategy

A simulation program, *msms* version 3.2rc (Ewing & Hermisson 2010), was used to test the power and robustness of χ. *msms* utilizes the functionality of *ms* (Hudson 2002), a coalescent simulator used to generate structured populations under neutrality. *msms* builds on *ms* by allowing selection at a single diploid locus to be simulated. A multitude of population scenarios and parameters were simulated in this study. In all cases, simulations involved two populations that split from a common ancestral population at a specific time (0.05 coalescent units ago, unless otherwise stated). Except where specified below, no subsequent migration occurred. At a specific time after the split, one population begins to experience positive selection at a target site in the middle of the simulated locus (using the '- SFC' option to condition against loss of the adaptive allele), while the other continues to evolve neutrally until sampling.

As sample cases for outcrossing species with lower and higher effective population size (N_e), we simulated scenarios with parameters inspired by human and *Drosophila* genetic diversity. For the high N_e case, 5- kb windows were generated with a per- site population mutation rate (θ) of 0.01 and a per- site population recombination rate (ρ) of 0.05. This ratio of ρ to θ is compatible with ratios of recombination and mutation rates estimated from recent studies of *D. melanogaster* (Comeron *et al*. 2012; Schrider *et al.* 2013). For low N_e scenarios, 100- kb windows were simulated with ρ and θ, both equal to 0.001. The difference in window size between these cases reflects the importance of both recombination and mutation rate differences for the scale and detection of selective sweeps.

For a subset of cases, lengths of simulated loci were increased ten- fold, and χ_{MD} and F_{ST} were calculated in sliding windows along the simulated locus. ρ and θ were scaled accordingly (increased ten- fold) and location of selection remained at the center of the locus. The sliding windows overlapped half of the previous window, and the windows were the same lengths as the full analyses. In total, 19 windows were analyzed in each simulation of longer loci. As *XP- EHH* utilizes SNPs surrounding a focal SNP, edge effects can alter *XP-EHH* calculations for windows at either end of the simulated locus. To correct for this issue, simulated locus lengths were further increased threefold to 150 kb for the high N_e population and 3 Mb for the low N_e population, with the beneficial mutation occurring in the center of the simulated region. *XP- EHH* was calculated on 19 windows sliding along the middle third of the simulated locus. Thus, SNPs in the added flanking regions could be utilized in the *XP-EHH* calculations to minimize edge effects.

In each population scenario, strong selection (s = 0.01) and weak selection (s = 0.001) were simulated for high N_e data, while only strong selection (s = 0.01) was simulated for low N_e data (too few simulated replicates reached fixation within the desired time interval in weaker selection simulations). Analyzed sample sizes were typically 50 chromosomes per population, but for a subset of cases other sample sizes (n = 12, 25, 100, 200) were also assessed. For this same subset that sample size was varied, we ran separate simulations varying locus length and haplotype length threshold proportion (a). Simulated window lengths were increased to 2X and 4X the original length as well as decreased to 0.5X and 0.25X the original lengths. Threshold proportions (the proportion of a window that need be identical) were also varied on these subsets (a = 0.025, 0.05, 0.1, 0.15, 0.2). In all other simulations, a threshold proportion of 0.1 was used. Command lines for all simulated cases are given in Table S1 (supplemental files).

For each scenario, a completely neutral set of simulations was also conducted, in which neither population experienced selection. A total of 10,000 replicates were simulated for each case with and without selection. Due to the heavy computational demands of calculating *XP-EHH* for each nonsingleton SNP across a window, only 1000 replicates were evaluated for this statistic. Power for each statistic was defined as the proportion of replicates giving a more extreme value (in the direction predicted by local adaptation in the first population) than 95% of the neutral replicates (implying a 5% false- positive rate). For *XP- EHH*, which is applied to each SNP in a window, we tested whether the maximum SNP *XP- EHH* obtained from a particular selection replicate was higher than 95% of neutral max(*XP- EHH*) values.

Simulation of Selective Sweeps from New Mutations

For each scenario in which a complete sweep was simulated, a large sample of 102 simulated chromosomes was split into selected and neutral populations of 52 and 50 chromosomes, respectively. To simulate a complete sweep, only replicates in which the beneficial allele appeared in 50 or more chromosomes were used in the analysis. In these cases, two chromosomes were thrown out so that a sample of 50 chromosomes (all with the beneficial allele) could be analyzed. This method of simulating extra chromosomes was used because of the difficulty in simulating recent complete sweeps due to the long stochastic phase at the end of a sweep.

Complete hard sweeps where populations split at 0.2 coalescent time units in the past as well as more ancient splits of 0.5 coalescent time units in the past were simulated. Selection initiation times were varied between 0.025 and 0.2 for the more recent split scenarios, while initiation times were varied between 0.2 and 0.5 for the more ancient split. Allele frequency thresholds were also studied for these more ancient splits. Instead of excluding only singletons,

allele counts of 2, 5, 10, 20, 25, 30, 35 and 40 (across both populations) were iteratively excluded in simulations where selection began between 0.2 and 0.5 coalescent time units in the past.

Complete hard sweeps with differing strengths of population bottlenecks were also simulated. In these cases, the populations split 0.05 coalescent time units in the past. The focal population immediately experienced a bottleneck and returned to its original effective population size at 0.04 coalescent time units in the past before undergoing selection at 0.025 coalescent time units in the past. The ratio of the bottlenecked population size to the original size was varied at 0.005, 0.01, 0.025, 0.05 and 0.1. We treat this ratio as a proxy for relative bottleneck strength.

Ongoing hard sweeps where the beneficial allele had not approached fixation (*i.e.* incomplete or partial sweeps) were also simulated. Here, simulated replicates were retained if the final frequency of the beneficial allele fell within a desired range around a target frequency (*e.g.* within 5% of 30%), and selection initiation times were chosen to generate such cases frequently (Table S2, supplemental files).

Simulations of Selective Sweeps from Standing Genetic Variation

For complete soft sweeps from standing genetic variation, we simulated different starting beneficial allele frequencies. These starting frequencies differed by species and selection strength (Table S3, supplemental files), in order to vary the number of unique adaptive alleles contributing to a sweep and to observe a range of power for the statistics examined. Population bottlenecks in combination with soft sweeps were simulated as described above, for a subset of the previously examined initial beneficial allele frequencies (Table S3, supplemental files).

Partial soft sweeps with varying starting and ending beneficial allele frequencies were also simulated. As with partial hard sweeps, selection was chosen to begin such that the

beneficial allele would often reach a target frequency range by the time of sampling, and only replicates within this range were accepted. Starting and ending beneficial allele frequencies, selection initiation times and the number of unique adaptive alleles they entailed are listed in Table S4 (supplemental files).

Simulations with Migration

For a subset of hard and soft sweep scenarios, symmetric migration between diverged populations was simulated. The population migration rate, $4N_em$, was varied at $4N_em$ = (1000, 2000, 3000, 4000, 5000) for the high N_e population and $4N_em$ = (100, 200, 300, 400, 500) for the low N_e population. The hard sweep scenario for the high N_e population involved a population split and an onset of selection 0.5 and 0.2 coalescent time units ago, respectively, while for the low N_e population these events occurred 0.2 and 0.1 coalescent units ago. For both soft sweep scenarios, the population split and onset of selection occurred 0.05 and 0.025 units in the past, respectively. The initial beneficial allele frequency was 0.001 for the high N_e population and 0.005 for the low N_e population. These scenarios were chosen to represent intermediate statistical power, such that performance could be compared between statistics. Selection of equal magnitude against the mutation was simulated in population 2 (the reference population). Full command lines for these simulations can be found in Table S1 (supplemental files).

Comparison with Single-Population Statistics

χ_{MD}, *XP- EHH,* and F_{ST} compare genetic variation between populations. A subset of simulations was analyzed with single- population statistics to examine how power is affected by the utilization of only a single population. Four single- population statistics were used: the numerator of the χ_{MD} statistic (P_1, the haplotype sharing of population 1), nucleotide diversity

(π), Tajima's *D* (Tajima 1989) and Fay and Wu's *H* (Fay & Wu 2000). Here, the 95th percentile values under neutrality for high P_1 and low π, *D*, and *H* were used as the detection threshold.

There were four scenarios for both high and low N_e populations that we tested single-population statistics on: a complete hard sweep, a partial hard sweep, a complete soft sweep and a partial soft sweep. Simulation parameters were chosen based on having relatively high power for between- population statistics. For complete hard sweep simulations, population divergence time and selection onset occurred at times 0.5 and 0.3 coalescent units before the present for the high N_e scenario, and at times 0.2 and 0.1 for the low N_e case (the more ancient selection in the high N_e case being necessary to focus on statistical powers below one). We simulated partial hard sweeps with a final beneficial allele frequency of 0.4 for both population sizes. Complete soft sweeps were simulated from initial beneficial allele frequencies of 0.001 and 0.02 for the high N_e and low N_e populations, respectively. For partial soft sweeps, the beneficial alleles rose in frequency from 0.0001 to 0.5 for the high N_e population and from 0.001 to 0.5 for the low N_e population. All other parameters corresponded to the default values elaborated in the above sections.

Application to an Empirical Dataset

χ_{MD}, *XP- EHH* and F_{ST} were applied to a *Drosophila melanogaster* genomewide data set, specifically the two largest African population samples from the *Drosophila* Genome Nexus (Lack *et al*. 2015). In this case, population 1 (the population of interest) is a collection of flies from Rwanda, while population 2 (the reference population) is a collection of fly lines from Zambia, which is thought to represent an ancestral range population (Pool *et al*. 2012). Window size was chosen so that 100 nonsingleton SNPs were contained in each window. In line with our high N_e simulations, these windows averaged approximately 5 kb in length, and we used 500

base pairs as the haplotype length threshold for χ_{MD}. For each statistic, an empirical '*P* value' (quantile) for a particular window was calculated as the proportion of windows on the same chromosome arm with more extreme statistic values than the focal window.

Using the results of the genomewide data set, we performed a Gene Ontology (GO) enrichment using the approach described by Pool *et al*. (2012). Outlier regions were defined as a set of windows in the 5% tail for a given statistic, separated by at most four nonoutlier windows. For each GO category, the number of outlier regions containing one or more genes associated with this category was noted. Based on 100,000 random permutation of outlier region locations, a *P* value was then calculated, representing the probability of randomly observing as many (or more) outliers from that category. The overlap of detected GO categories between statistics was visualized using euler*APE* (Micallef & Rodgers 2014).

Results

As detailed above, we conducted coalescent simulations under a wide range of scenarios with and without positive selection, using parameters motivated by human and *Drosophila* genetic variation as examples of species with lower or higher N_e. These simulations allowed us to gauge the empirical power of the χ_{MD} statistic relative to *XP-EHH* (Sabeti *et al*. 2007) and window F_{ST} (Wright 1931; Hudson *et al*. 1992) and to compare the power of these population comparison statistics against single- population statistics. As *XP-EHH* is a per- SNP analysis, we compared it to the window statistics by comparing the maximum *XP- EHH* in each window from selection versus neutral simulations. Results illustrating intermediate power are highlighted in the figures and text below, while full results (including those for the raw χ statistic with no denominator adjustment) are given in Table S5 (supplemental files). For the subset of scenarios where the sliding window (as well as the locus

length and threshold) analyses were performed, distributions under both neutrality and selection are provided (Figures S1 and S2, Appendix 1). Default simulation parameters are given in Table 1; these values were used except when explicitly varied in the sections described below.

Older Hard Sweeps

Previous simulation analysis of single- population summary statistics for detecting selective sweeps has pointed to a fairly brief window for their detection. For example, by 0.15 coalescent units (*i.e*. $0.6N_e$ generations) after a selective sweep, Przeworski (2002) found that the power of Tajima's (1989) *D* had been reduced to around 30%, while the rejection rate of Fay & Wu's (2000) *H* was close to the false- positive rate. As expected, our analysis of population-specific classic sweeps also showed that power for each statistic decreased as selection initiation was pushed further back (Fig. 1; Table S5, supplemental files). However, the temporal signal of selection was notably extended for these population comparison statistics. F_{ST} showed the strongest performance, with an exceptionally long- lasting signal of selection. Thus, even in the absence of ongoing selection against migrant alleles, sweeps that differentiate fairly anciently isolated populations may still be detectable. χ_{MD}, which excludes singleton polymorphisms to avoid loss of power due to new mutations, outperformed *XP- EHH* for older hard sweeps. χ_{MD} still retained approximately 50% power at 0.2 coalescent units after a sweep in the low N_e, $s = 0.01$ case and maintained this performance until 0.5 coalescent units after the high N_e, $s = 0.001$ sweep scenario.

Partial Hard Sweeps

Statistical power from partial hard sweep simulations (Fig. 2; Table S5, supplemental files) showed an intuitive increase from a final adaptive allele frequency of 10% (for which power was minimal) to 50% (for which all statistics had strong power). χ_{MD} displayed superior

power for the low N_e case. The statistics had generally similar performance for high N_e partial sweeps, with χ_{MD} and F_{ST} ahead of *XP- EHH* in some instances.

Complete and Partial Soft Sweeps

Soft sweeps act on standing variation (as simulated here) or else on recurrent mutations in very large populations. Hence, the adaptive allele may persist on multiple genetic backgrounds within a population after selection, reducing haplotype sharing relative to hard sweeps, and making it more difficult to detect local adaptation. Relatively speaking, softer sweeps are those where the adaptive alleles present at the time of sampling trace back to a larger number of unique chromosomes at the onset of selection. Concordant with previous findings (Pennings & Hermisson 2006), we observed that sweeps become more difficult to detect with increasing softness (Fig. 3). In cases where statistical power was neither uniformly high nor uniformly low, χ_{MD} generally outperformed *XP- EHH*. For the low N_e case, χ_{MD} also outperformed F_{ST}, while in the high N_e case F_{ST} had an advantage for softer sweeps. Notably, χ_{MD} was able still to detect a signal of selection in more than 20% of the low N_e replicates when the starting beneficial allele frequency was as high as 10%.

Naturally, incomplete soft sweeps were found to be even more challenging to detect than complete soft sweeps, especially with high initial frequencies and/or low final frequencies of the favored allele (Fig. 4). The χ_{MD} statistic performed particularly impressively in the low N_e simulations, often outperforming *XP- EHH* and F_{ST} by significant margins. For high N_e data, performance of the three statistics was more similar, with χ_{MD} and F_{ST} often slightly exceeding the power of *XP- EHH*.

Hard and Soft Sweeps in Bottlenecked Populations

Until now, we have considered cases of positive selection in populations of constant size. However, we also evaluated a series of population bottleneck scenarios affecting the same population subject to a complete hard or soft sweep (models of particular interest with regard to domestication and the colonization of new environments). In general, bottlenecks are known to reduce genetic variation and to increase the stochastic variance among loci. This increased homozygosity (and, therefore, increased haplotype sharing) in the neutral simulations created higher threshold levels, lowering the power of the tested statistics.

Although bottlenecks presented a challenge for all statistics, *XP- EHH* often showed the highest power, especially for the low N_e simulations (Fig. 5; Table S5, supplemental files). Here, the focus of *XP- EHH* on a specific haplotype configuration (as opposed to all haplotype identity) may have helped preserve more discriminatory power. F_{ST} typically performed worse than either haplotype statistic in the presence of bottlenecks, in agreement with the notion that linkage information may be generally helpful in differentiating nonequilibrium demography from positive selection (Jensen *et al*. 2007).

Detecting Local Adaptation in the Presence of Migration

We also investigated scenarios in which migration occurred between diverged populations (Table S5, supplemental files), under a standard isolation–migration model. Results from very high migrations rates are presented here, because the power of each statistic was mostly unaffected until migration rates were increased enough to keep F_{ST} close to 0 under neutrality. Figure 6 illustrates statistical performance in cases of very high migration rates that typically prevent the beneficial allele from becoming a fixed difference. Particularly for F_{ST}, selection scenarios with lower migration rates were often easier to detect than those with no migration, suggesting that ongoing selection against migrant alleles may have increased power

(note that the onset of selection in some scenarios was fairly ancient; Materials and Methods). For the low N_e cases, all three statistics showed similar performance in the presence of migration. For the high N_e scenarios, F_{ST} gave the highest power, potentially due to ongoing differentiation at the target site and very closely linked variants (leading to modest window F_{ST} values that still exceeded the even smaller values under neutrality). *XP- EHH* also shows an advantage over χ_{MD}, particularly for the ancient hard sweep case examined. Here, the association between long haplotypes and a specific allele at the target site may preserve a signal for *XP- EHH*, even if overall levels of haplotype sharing become relatively similar between the two populations.

Effects of Allele Frequency Threshold and Sample Size

We found that the power to detect old sweeps, already notable for these statistics relative to single- population approaches, could be substantially improved for χ_{MD} by increasing our allele frequency threshold to exclude more than just singletons (Fig. 7). This result is intuitive because as time passes after a sweep, new mutations start drifting to higher frequencies, and nonsingleton SNPs disrupt otherwise identical haplotypes that had been homogenized by the sweep. Frequency thresholds as high as 20 or 25% (out of the combined two- population sample size of 100) were favored for sweeps as ancient as 0.4 or 0.5 coalescent units. These results suggest that a localized absence of intermediate frequency alleles may carry a previously unappreciated signal of ancient positive selection.

As would be predicted, power for each statistic increased with increasing sample size (Fig. 8; Table S5, supplemental files). In general, the sample size of 50 chromosomes per population used in the preceding analyses appears to represent a good compromise between sequencing effort and power. Additional power was observed with larger samples, but with some diminishing returns.

Impact of Window Length and Threshold Proportion

Simulated window lengths and threshold proportions were investigated for the χ_{MD} statistic (Fig. 9; Table S5, supplemental files). Here, threshold proportion refers to the fraction of the window that must be identical between a pair of haplotypes to count towards the total. Diagonal 'ridges' of high power are sometimes observed in Fig. 8, suggesting an optimum threshold *length* (*i.e*. window length × threshold proportion) for a given selection scenario. However, this optimum depends not only on the species, but also on the nature of selection (*e.g*. hard vs. soft sweeps), suggesting that no single configuration is universally advantageous. It should be noted that the scenarios simulated in this study involved relatively strong selection (s = 0.001 and s = 0.01), so that sweeps would finish within a proscribed time frame. If selection is typically weaker in the species of interest, the shorter shared haplotypes that result could favor a smaller threshold length than indicated by Fig. 9 (see 4).

Sliding Window Analyses

All three statistics were evaluated in sliding windows along a locus so that the effects of physical distance from the selected site could be observed. Intuitively, powers for all three statistics decreased with distance from the site of selection (Fig. 10; Table S5, supplemental files). Minor differences were observed in the spatial extent of the three statistics' signals. The two haplotype signals often displayed wider signals than F_{ST}, and χ_{MD} sometimes showed a slightly broader signal than *XP- EHH*.

Comparison with Single Populations Statistics

In general, single- population statistics were outperformed by cross- population statistics (Fig. 11; Table S5, supplemental files), underscoring the advantage of controlling for shared history in the ancestral population. An exception was power for the haplotype statistic P_1, which

was essentially unaffected by the use of only one population. Thus, under the conditions simulated, the P_1 statistic (quantifying the haplotype sharing of population 1) is quite sensitive a wide range of selective sweep scenarios. However, adding a second population may add important robustness to empirical studies. In these simulations, a specific known recombination rate was used. Using a second population helps control for the historical recombination rate, which would not necessarily be known in a real data set, making it difficult to predict how a single- population haplotype statistic should behave under neutrality. Further, the use of a second population can also control for demographic and selective events in the ancestral population, which were not simulated in this study.

Nucleotide diversity (π), Tajima's *D*, and Fay and Wu's *H* had varying power in each sweep scenario. Fay and Wu's *H*, for instance, showed moderately high power in partial and/or soft sweep scenarios, but low power in the complete hard sweep scenarios (particularly in the large N_e case, where the longer time since selection erases the signal of high- frequency-derived alleles; Przeworski 2002). In contrast, the between- population statistics showed relatively high power in each sweep scenario, a critical advantage because we do not know which kind of selection to expect in a real data set.

Empirical Analysis of Drosophila Genomes

To examine the performance of cross- population statistics on empirical data, we analysed fully sequenced *D. melanogaster* genomes from the Drosophila Genome Nexus (Lack *et al*. 2015). Specifically, we compared variation between the Rwanda Gikongoro population sample (27 genomes) and the Zambia Siavonga population sample (197 genomes). Being sequenced to averaged depths of >27X (Lack *et al*. 2015) from haploid female gametes (Langley *et al*. 2011), these genomes have the advantage of clearly defined haplotypes.

Zambia appears to represent an ancestral range population, while Rwanda and other equatorial African populations may reflect range expansion (Pool *et al*. 2012). The range of selective pressures that may differ between these populations is unknown, but geographic and climate differences do exist. The Rwanda location features a higher altitude (1930 vs. 530 m above sea level) and greater rainfall, while Zambia has more seasonal variation in temperature and a longer dry season.

Applying χ_{MD}, *XP- EHH* (again bounded as a window statistic), and F_{ST} to this genomic data set, we were able to study statistic correlations as well as perform a GO enrichment analysis. Each genomic window has a value for χ_{MD}, *XP-EHH* and F_{ST} (Table S6, supplemental files) and thus has an associated quantile or empirical *P* value for each statistic as well. Moderately strong correlations were observed between all three statistics (Table 2; Figure S3, appendix 1), with the highest correlation between *XP- EHH* and F_{ST}.

Figure 12 depicts the most extreme outlier regions for each statistic as well as their flanking regions. The vMD outlier, located within the Insulin-like receptor gene, was also detected by FST but not by XP-EHH. XP-EHH and FST identified the same maximal outlier region, among a group of cuticle protein genes, which was also flagged by vMD.

We performed Gene Ontology enrichment analysis on the results for each statistic (Materials and Methods). Our primary goal for this exploratory analysis was to investigate the degree to which different statistics find evidence for selection in the same functional categories of genes. We found fairly strong overlap between the biological processes implicated by χ_{MD}, *XP- EHH*, and F_{ST} (Figure S4, appendix 1). Complete results are given in Table S7 (supplemental files), while a set of the most enriched terms for each statistic is given in Table 3. While each statistic implicated a unique combination of GO categories, all lists included

functions related to sensory perception and apoptosis. Differences in the genes and categories detected by each statistic may reflect both false positives and differences in the type and timing of selection impacting different genes and functional categories.

Discussion

Detecting cases of local selection is critical for the study of agricultural domestication, conservation and human biology, as well as our basic understanding of adaptation and its genetic basis. However, positive selection can have different forms at the population genetic level (hard vs. soft sweeps, complete vs. partial sweeps) and may or may not have occurred very recently in population genetic time; especially when data from only one population is available, it can be very difficult to find statistical methods able to detect such a wide variety of adaptive scenarios. Here, we show that detecting diverse modes of positive selection is often possible when comparing genetic variation from two populations with adaptive differences.

We have introduced a statistic, χ_{MD}, that compares the total pairwise haplotype identity within each of two populations and compared its performance against another haplotype statistic (*XP- EHH*) and an index of allele frequency differentiation (F_{ST}). F_{ST} often had fairly similar power to detect local selection as the haplotype statistics. Although joint approaches are not a focus of the present study, it may be advantageous in many scenarios to use F_{ST} and a haplotype metric as complementary statistics. Relative to the haplotype approaches, F_{ST} often had stronger performance for older (hard) sweeps and weaker power for population bottleneck scenarios with hard or soft sweeps.

Focusing on the differences between the χ_{MD} and *XP- EHH* haplotype statistics, the primary performance advantage observed for *XP- EHH* was for selection in bottlenecked populations. *XP- EHH* had important advantages for certain bottleneck and migration scenarios.

Hence, the specific haplotype configuration sought by the *EHH* approach appears to confer some robustness against demographic sources of haplotype identity.

Notably, however, χ_{MD} showed superior power to *XP- EHH* in most other scenarios. For hard sweeps, the statistical signal of χ_{MD} is more enduring than for *XP- EHH*. The longer-lasting signal of χ_{MD} may stem partly from the masking of rare variation, which prevents postselection mutations from interrupting haplotype identity. χ_{MD} may also be more tolerant of recombination during or after selection, as identical haplotype blocks do not need to maintain their original linkage configuration in order to contribute to summed haplotype identity.

In addition, χ_{MD} displayed greater power than *XP- EHH* for many cases of partial and/or soft sweeps. For the low N_e partial and soft sweep cases, χ_{MD} showed performance advantages over both *XP- EHH* and F_{ST}. These results underscore the versatility of χ_{MD} for detecting population- specific selection. This flexibility reflects a very basic signal of directional selection that χ_{MD} responds to haplotype sharing between alleles with unusually recent common ancestry. This signal is produced even if multiple haplotypes carry the beneficial mutation, or if this mutation has not reached high frequency.

Being a window- based approach, χ_{MD} is particularly well suited to analyzing fully sequenced genomes. Although implemented in kilobase- defined windows in this simulation study, in real genomes it may be preferable to apply χ_{MD} in windows scaled by genetic distance or by numbers of variable sites (as implemented in the *Drosophila* case studied here). The window orientation of χ_{MD} also makes it dramatically more computationally efficient than *XP-EHH*, which must be evaluated separately for every variable site that passes filtering criteria. This difference also implies that many fewer tests need to be performed for a genomewide

analysis of χ_{MD} in comparison with *XP- EHH*, although we have shown that *XP- EHH* still maintains significant power when applied in a window- maximum format.

When applying χ_{MD} to empirical data, two general issues should be carefully considered. One is the parameterization of χ_{MD} in terms of window and threshold length, and allele frequency threshold. Although we offer preliminary guidance through the simulation analyses shown here, we recommend that potential users conduct similar simulations reflecting the genetic properties and demographic histories of their own study populations, along with selective sweep models of potential interest (in terms of strength, hardness and timing), in order to fine-tune χ_{MD} settings.

A second major issue, relevant to any population genomic analysis, is the determination of statistical significance. If demographic parameter estimates are available that are reliable, or at least conservative with respect to intrapopulation shared haplotype lengths, then neutral simulations can be performed to obtain the probability of observing a given χ_{MD} value without selection. If researchers need to establish whether a given value is unexpected genomewide, then clearly a multiple testing correction is also needed (*e.g*. Storey & Tibshirani 2003). If no credible demographic model is available, then the user is most likely restricted to an outlier framework to identify preliminary candidates for local adaptation.

Throughout this analysis, we have assumed that the phase of each haplotype is known with certainty. In some organisms, including *Drosophila*, it is possible to sequence completely or mostly homozygous genomes (*e.g*. Langley *et al*. 2011; Mackay *et al*. 2012). But for many diploid nonlaboratory organisms, including humans, it is not yet practical to empirically obtain genomewide phasing data unless family groups (*e.g*. parent–child trios) are sequenced. Although haplotype phasing can be estimated computationally (*e.g*. Scheet & Stephens 2006), the bias

entailed by such methods for haplotype statistics like χ_{MD} is unclear. Alternatively, an unphased counterpart to χ_{MD} could be envisioned in which homozygosity runs shared between individuals are totaled.

The simple χ_{MD} statistic appears to be quite useful in its current form, but future advances over the present approach are certainly conceivable. The probability of a specific shared haplotype length under the null hypothesis could be evaluated via theory (Harris & Nielsen 2013) or simulation, potentially eliminating the need for a threshold length. Information could also be combined across windows to delineate the boundaries of non- neutral regions, or window size could be adjusted based on observed genetic variation (Pavlidis *et al*. 2010). Lastly, the signal of haplotype identity could be combined with information from the two- population allele frequency spectrum and other aspects of genetic variation. Still, the present work represents a 'proof of concept' that haplotype identity tracts efficiently capture the signal of diverse modes of positive selection, often performing as well or better than published statistics at distinguishing neutral from non- neutral histories.

Acknowledgements

We thank the UW-Madison Center for High Throughput Computing (CHTC) for access to the computing cluster that facilitated our simulations. Funding was provided by an NIH grant (R01 GM111797) and a USDA Hatch award (WIS01900) to JEP.

References

Bonin A, Nicole F, Pompanon F, Miaud C, Taberlet P (2007) Population Adaptive Index: a new method to help measure intraspecific genetic diversity and prioritize populations for conservation. Conservation Biology, 21, 697–708.

Comeron JM, Ratnappan R, Bailin A (2012) The many landscapes of recombination in Drosophila melanogaster. PLoS Genetics, 8, e1002905.

Ewing G, Hermisson J (2010) MSMS: a coalescent simulation program including recombination, demographic structure and selection at a single locus. Bioinformatics, 26, 2064–2065.

Fariello MI, Boitard S, Naya H, SanCristobal M, Servin B (2013) Detecting signatures of selection through haplotype differentiation among hierarchically structured populations. Genetics, 193, 929–941.

Fay JC, Wu CI (2000) Hitchhiking under positive Darwinian selection. Genetics, 155, 1405–1413.

Ferrer-Admetlla A, Liang M, Korneliussen T, Nielsen R (2014) On detecting incomplete soft or hard selective sweeps using haplotype structure. Molecular Biology and Evolution, 31, 1275–1291.

Fumagalli M, Sironi M, Pozzoli U, Ferrer-Admettla A, Pattini L, Nielsen R (2011) Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution. PLoS Genetics, 7, e1002355.

Hancock AM, Witonsky DB, Alkorta-Aranburu G et al. (2011) Adaptations to climate-mediated selective pressures in humans. PLoS Genetics, 7, e1001375.

Harris K, Nielsen R (2013) Inferring demographic history from a spectrum of shared haplotype lengths. PLoS Genetics, 9, e1003521.

Hudson RR (2002) Generating samples under a Wright-Fisher neutral model of genetic variation. Bioinformatics, 18, 337–338.

Hudson RR, Slatkin M, Maddison WP (1992) Estimation of levels of gene flow from DNA sequence data. Genetics, 132, 583–589.

Jensen JD, Thornton KR, Bustamante CD, Aquadro CF (2007) On the utility of linkage disequilibrium as a statistic for identifying targets of positive selection in nonequilibrium populations. Genetics, 176, 2371–2379.

Lack JB, Cardeno CM, Crepeau MW et al. (2015) The Drosophila Genome Nexus: a population genomic resource of 623 Drosophila melanogaster genomes, including 197 from a single ancestral range population. Genetics, 199, 1229–1241.

Langley CH, Crepeau M, Cardeno C, Corbett-Detig R, Stevens K (2011) Circumventing heterozygosity: sequencing the amplified genome of a single haploid Drosophila melanogaster embryo. Genetics, 188, 239–246.

Mackay TFC, Richards S, Stone EA et al. (2012) The Drosophila melanogaster genetic reference panel. Nature, 482, 173–178. Micallef L, Rodgers P (2014) eulerAPE: Drawing area-proportional 3-Venn diagrams using ellipses. PLoS ONE, 9, e101717

Pavlidis P, Jensen JD, Stephan W (2010) Searching for footprints of positive selection in whole-genome SNP data from nonequilibrium populations. Genetics, 185, 907–922.

Pennings PS, Hermisson J (2006) Soft sweeps III: the signature of positive selection from recurrent mutation. PLoS Genetics, 2, e186.

Pickrell JK, Coop G, Novembre J et al. (2009) Signals of recent positive selection in a worldwide sample of human populations. Genome Research, 19, 826–837.

Pool JE, Aquadro CF (2007) The genetic basis of adaptive pigmentation variation in Drosophila melanogaster. Molecular Ecology, 16, 2844–2851.

Pool JE, Corbett-Detig RB, Sugino RP et al. (2012) Population genomics of sub-Saharan Drosophila melanogaster: African diversity and non-African admixture. PLoS Genetics, 8, e1003080.

Przeworski M (2002) The signature of positive selection at randomly chosen loci. Genetics, 160, 1179–1189.

Rebeiz M, Pool JE, Kassner VA, Aquadro CF, Carroll SB (2009) Stepwise modification of a modular enhancer underlies adaptation in a Drosophila population. Science, 326, 1663–1667.

Roesti M, Gavrilets S, Hendry AP, Salzburger W, Berner D (2014) The genomic signature of parallel adaptation from shared genetic variation. Molecular Ecology, 23, 3944–3956.

Sabeti PC, Reich DE, Higgins JM et al. (2002) Detecting recent positive selection in the human genome from haplotype structure. Nature, 419, 832–837.

Sabeti PC, Varilly P, Fry B et al. (2007) Genome-wide detection and characterization of positive selection in human populations. Nature, 449, 913–918.

Santiago E, Caballero A (2005) Variation after a selective sweep in a subdivided population. Genetics, 169, 475–483.

Scheet P, Stephens M (2006) A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. American Journal of Human Genetics, 78, 629–644.

Schlotterer C, Dieringer D (2005) A novel test statistic for the identification of local selective sweeps based on microsatellite gene diversity. In: Selective Sweep (ed Nurminsky D), pp. 55–64. Kluwer Academic/Plenum Publishers, New York.

Schrider DR, Houle D, Lynch M, Hahn MW (2013) Rates and genomic consequences of spontaneous mutational events in Drosophila melanogaster. Genetics, 194, 937–954.

Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. Proceedings of the National Academy of Sciences, 100, 9440–9445.

Storz JF, Kelly JK (2008) Effects of spatially varying selection on nucleotide diversity and linkage disequilibrium: insights from deer mouse globin genes. Genetics, 180, 367–379.

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics, 123, 585– 595.

Voight BF, Kudaravalli S, Wen X, Pritchard JK (2006) A map of recent positive selection in the human genome. PLoS Biology, 4, e72.

Will JL, Kim HS, Clarke J, Painter JC, Fay JC , Gasch AP et al. (2010) Incipient balancing selection through adaptive loss of aquaporins in natural Saccharomyces cerevisiae populations. PLoS Genetics, 6, e1000893.

Wright S (1931) Evolution in Mendelian populations. Genetics, 16, 97–159.

Yi X, Liang Y, Huerta-Sanchez E et al. (2010) Sequencing of 50 human exomes reveals adaptation to high altitude. Science, 329, 75–78.

Table 1

Default simulation parameters, used except where otherwise noted.

Parameter	Low N_e	High N_e
locus length (kilobases)	100	5
a (threshold proportion)	0.1	0.1
haploid sample size (per population)	50	50
N_e (effective population size)	10,000	2,500,500
θ (population mutation rate)	0.001	0.01
ρ (population recombination rate)	0.001	0.05
$4N_em$ (population migration rate)	0	0
population split time (coalescent units)	0.05	0.05
onset of selection (coalescent units)	0.025	0.025
s selection coefficient	0.01	0.001

Table 2

Window quantile correlations are shown among the three between-population statistics evaluated for the Drosophila genomic data set. Conditional probability refers to the probability that a window is within the 5% tail of one statistic, given that it is within the 5% tail of another statistic. Because the number of outliers is the same for each statistic, these probabilities are symmetric.

Statistics	Correlation Coefficient	Conditional Probability
χ_{MD}, *XP-EHH*	0.5395	0.3529
χ_{MD}, F_{ST}	0.4827	0.4029
XP-EHH, F_{ST}	0.5713	0.4837

Table 3

Selected biological processes enriched for outliers of each statistic are given. For each statistic, biological process GO categories represented in at least five outlier regions were identified. Of those with raw permutation P value below 0.01, the categories with the highest proportion of outliers are listed here. Highly similar GO categories were omitted to minimize redundancy.

GO ID	Description	Windows w/Genes	χ_{MD} Outliers	χ_{MD} *P*	*XP-EHH* Outliers	*XP-EHH* *P*	F_{ST} Outliers	F_{ST} *P*
χ_{MD} Enrichment								
43 524	Negative regulation of neuron apoptotis	13	7	0.001	4	0.103	4	0.054
32 006	Regulation of TOR signalling cascade	11	5	0.008	5	0.006	3	0.095
48 190	Wing disc dorsal/ventral pattern formation	46	16	0.004	13	0.031	9	0.159
9582	Detection of abiotic stimulus	55	18	0.002	17	0.002	13	0.016
45 448	Mitotic cell cycle, embryonic	28	9	0.004	1	0.977	6	0.050
7602	Phototransduction	43	13	0.007	13	0.004	11	0.007
31 124	mRNA 3′-end processing	30	9	0.002	5	0.201	3	0.551
6289	Nucleotide excision repair	24	7	0.003	6	0.013	4	0.102
6401	RNA catabolic process	32	8	0.009	7	0.026	3	0.507
22 613	Ribonucleoprotein complex biogenesis	35	8	0.006	7	0.021	7	0.011
42 451	Purine nucleoside biosynthetic process	44	10	0.007	6	0.209	10	0.002
6260	DNA replication	66	14	0.004	7	0.466	13	0.002
70 647	Protein modified by small protein conjugation/removal	91	19	0.002	17	0.009	16	0.004
8340	Determination of adult lifespan	145	30	0.001	19	0.261	26	0.001
6310	DNA recombination	56	11	0.006	9	0.039	8	0.049
XP-EHH enrichment								
32 006	Regulation of TOR signalling cascade	11	5	0.008	5	0.006	3	0.095
6917	Induction of apoptosis	20	5	0.251	8	0.007	6	0.029
71 453	Cellular response to oxygen levels	28	7	0.130	10	0.003	7	0.044
6816	Calcium ion transport	31	9	0.050	11	0.003	10	0.002
7369	Gastrulation	31	8	0.133	11	0.004	10	0.003
46 662	Regulation of oviposition	18	1	0.909	6	0.009	3	0.238
9581	Detection of external stimulus	60	19	0.002	19	0.001	15	0.005
10 942	Positive regulation of cell death	53	9	0.619	16	0.006	13	0.013
8344	Adult locomotory behaviour	60	16	0.018	16	0.010	11	0.098
7291	Sperm individualization	45	9	0.047	11	0.005	10	0.004
52 548	Regulation of endopeptidase activity	47	9	0.051	11	0.005	10	0.004
50 906	Detection of stimulus involved in sensory perception	84	14	0.167	19	0.003	19	<0.001
9416	Response to light stimulus	119	25	0.014	26	0.003	25	<0.001
7349	Cellularization	94	16	0.064	20	0.002	18	0.002
43 900	Regulation of multi-organism process	81	9	0.689	17	0.009	17	0.001
F_{ST} enrichment								
35 072	Ecdysone-mediated induction of salivary gland cell autophagic cell death	11	3	0.622	5	0.083	6	0.005
7157	Heterophilic cell–cell adhesion	25	10	0.098	11	0.019	13	<0.001
61 057	Peptidoglycan recognition protein signalling pathway	10	2	0.228	3	0.052	5	<0.001
35 073	Pupariation	11	2	0.386	1	0.740	5	0.002
51 260	Protein homo-oligomerization	17	6	0.034	5	0.090	7	0.002
35 303	Regulation of dephosphorylation	13	4	0.033	4	0.029	5	0.002
6963	Positive regulation of antibacterial peptide biosynthesis	21	3	0.612	3	0.565	8	0.001
43 279	Response to alkaloid	22	2	0.859	5	0.165	8	0.002
12 502	Induction of programmed cell death	31	7	0.318	11	0.006	11	0.001
43 523	Regulation of neuron apoptotic process	17	8	0.001	5	0.064	6	0.007
10 950	Positive regulation of endopeptidase activity	17	5	0.045	5	0.039	6	0.004
45 088	Regulation of innate immune response	20	3	0.468	2	0.712	7	0.002
71 897	DNA biosynthetic process	27	6	0.036	5	0.099	9	<0.001
50 911	Detection of chemical stimulus involved in sensory perception of smell	40	5	0.552	10	0.011	13	<0.001
16 337	Cell–cell adhesion	80	25	0.083	25	0.028	26	<0.001

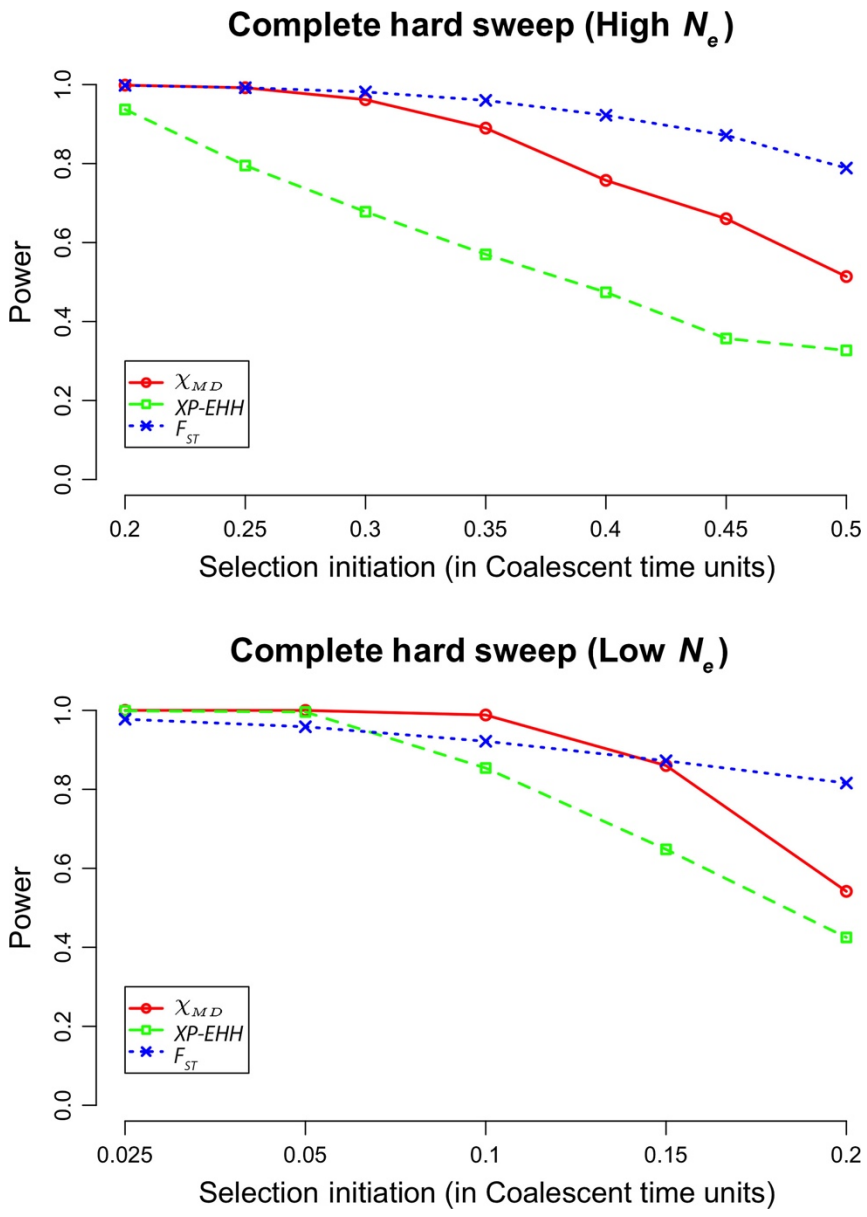


Figure 1. Power of each statistic for complete hard sweeps for high N_e (top) and low N_e cases (bottom). Note the difference in selection initiation times of the x-axes.

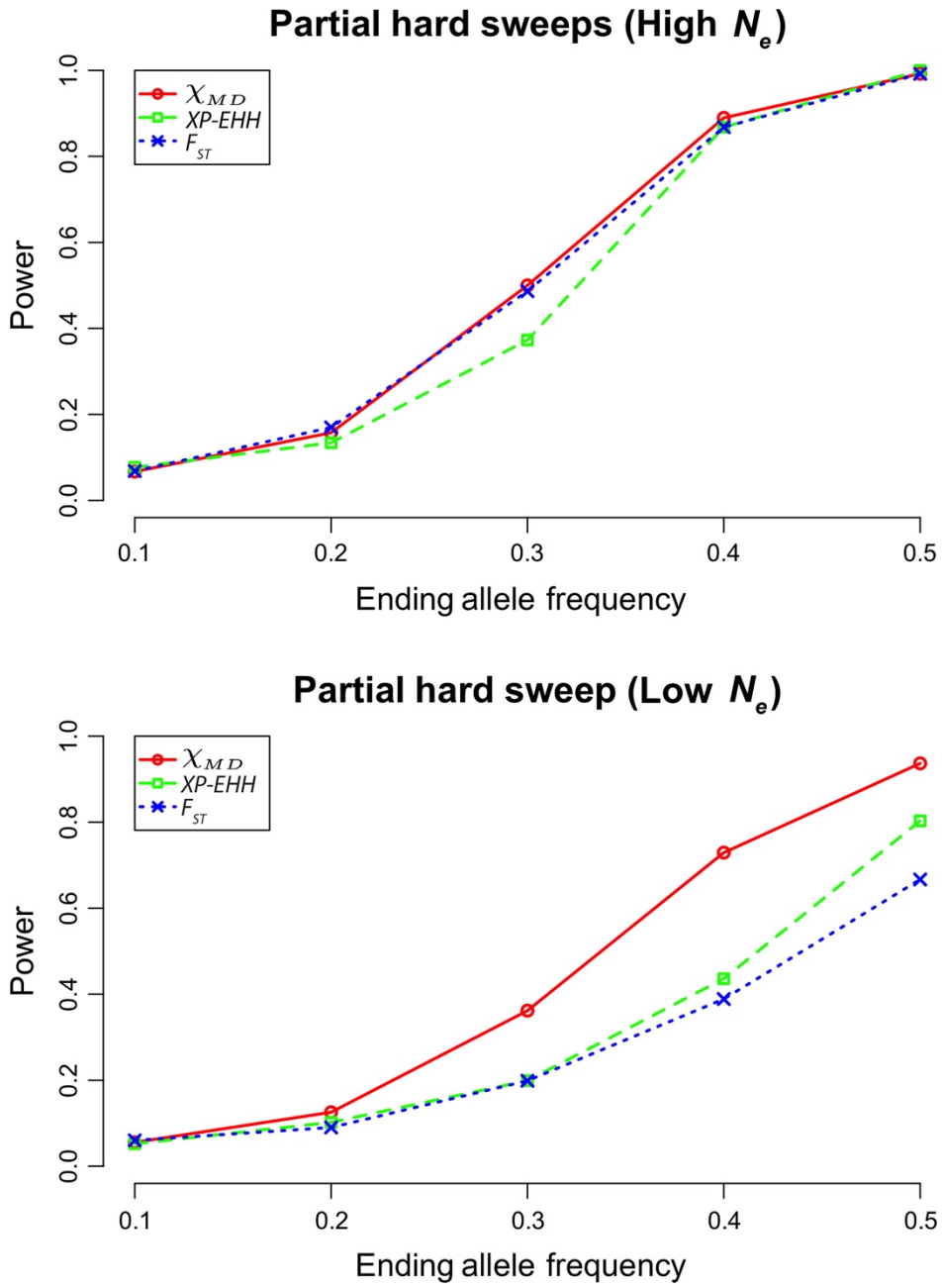


Figure 2. Power of each statistic tested for partial hard sweeps, for high N_e (top) and low N_e cases (bottom).

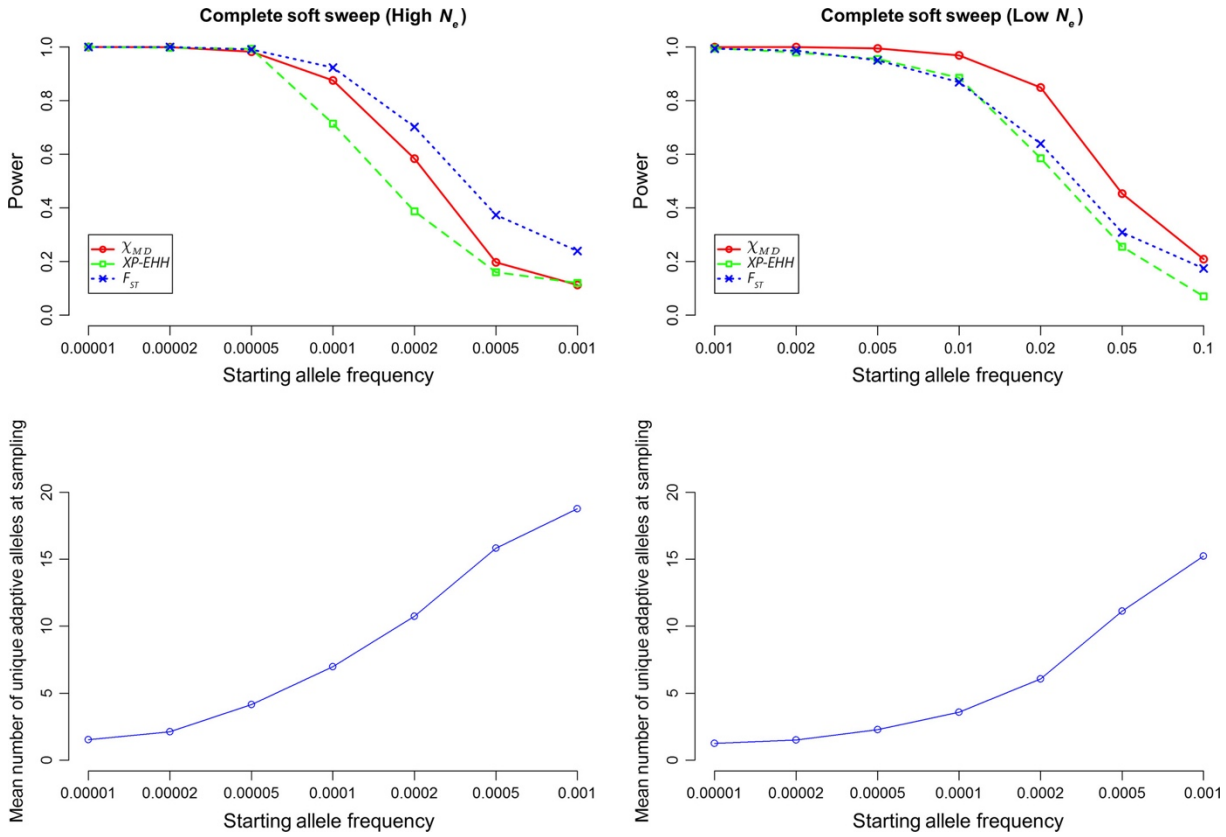


Figure 3. For complete soft sweeps, the top two panels depict power of each tested statistic. The bottom two panels depict the number of unique adaptations of derived allele at the time of sampling to help distinguish the softness of the sweep (where a value close to 1 indicates mostly hard sweeps). Note the change in scale of x-axes between the two N_e cases simulated (left and right).

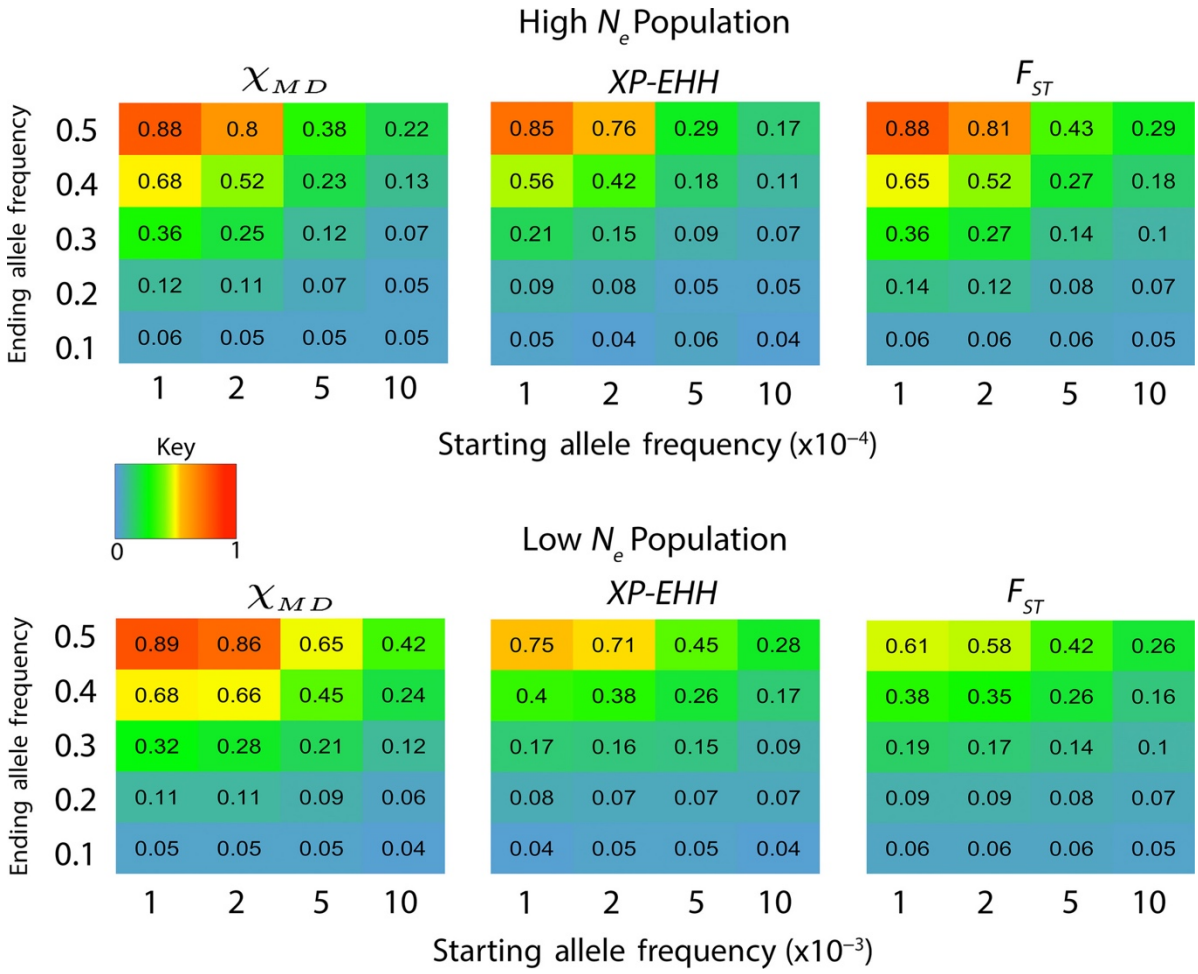


Figure 4. Heat map depicting power for each statistic for partial soft sweeps. The key refers to powers ranging from 0 to 1. The x-axis represents the number of copies of the beneficial allele in the population when the populations split. Note the change in x-axes between the two N_e cases (starting frequency per 10 000 or per 1000). The y-axis represents the ending allele frequency at sampling.

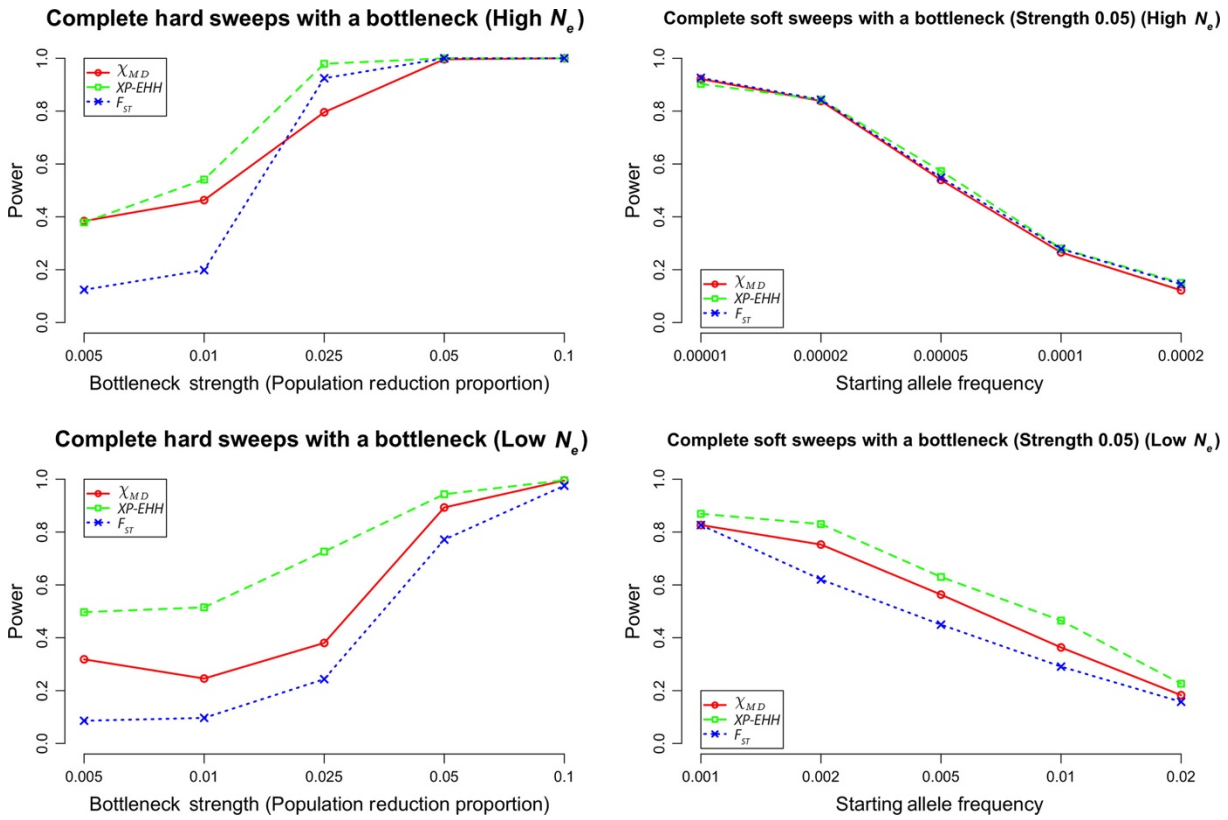


Figure 5. Depicted here are power for scenarios with bottlenecks simulated. The left panels depict hard sweeps, with varying strengths of bottlenecks indicated on the x-axis. The right panels depict a single bottleneck strength (0.05) with varying starting allele frequencies. Additional cases are summarized in Table S5 (supplemental files).

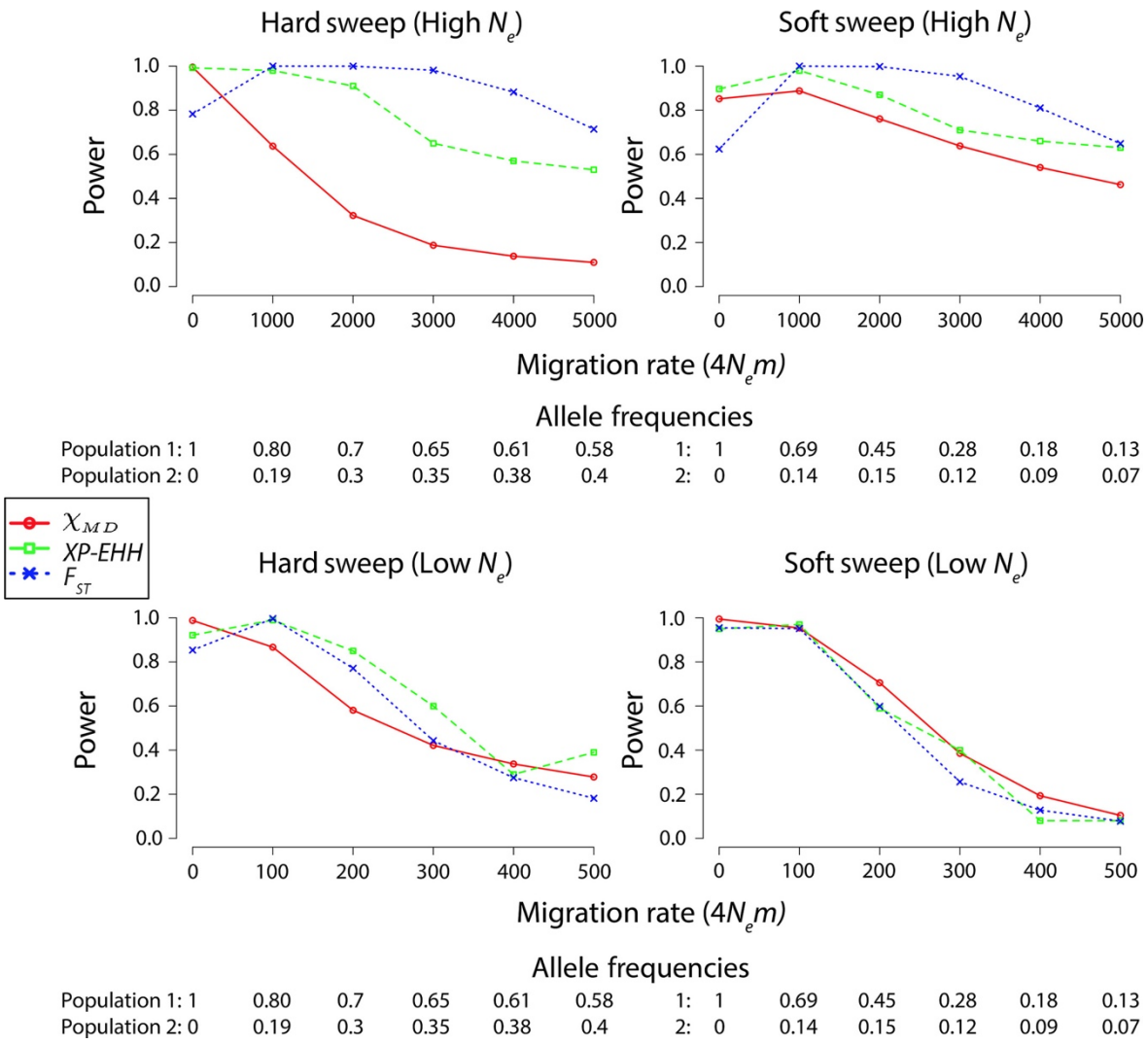


Figure 6. Migration was simulated for a subset of scenarios. The high levels of migration that affected statistical performance were sufficient to prevent fixed differences at the target site. Allele frequencies at sampling for both populations are shown below each migration rate.

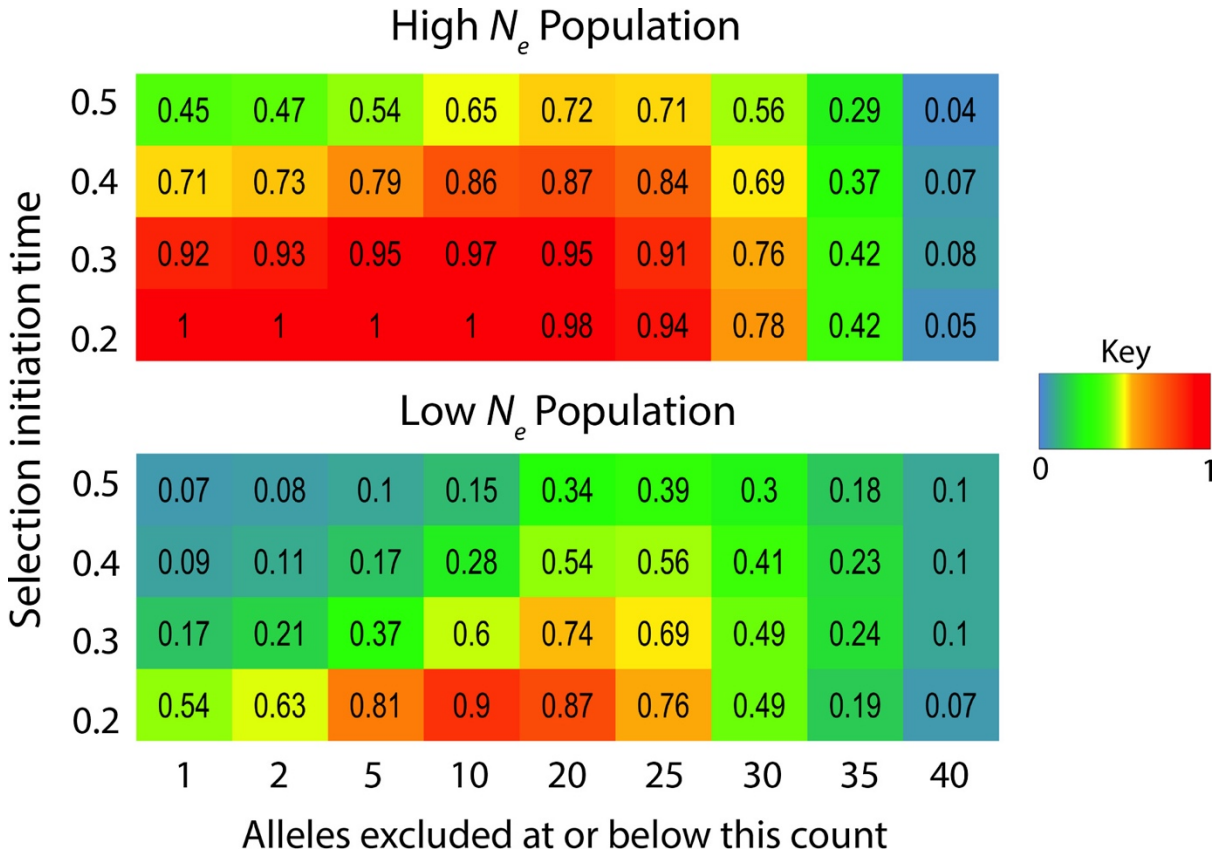


Figure 7. This heat map depicts power of the χ_{MD} statistic as a function of allele frequency threshold (minimum frequency of allele to be included in analysis) and the time (in coalescent units) since the initiation of a complete hard selective sweep. The exclusion of all but intermediate frequency alleles yields surprising power to detect very ancient sweeps.

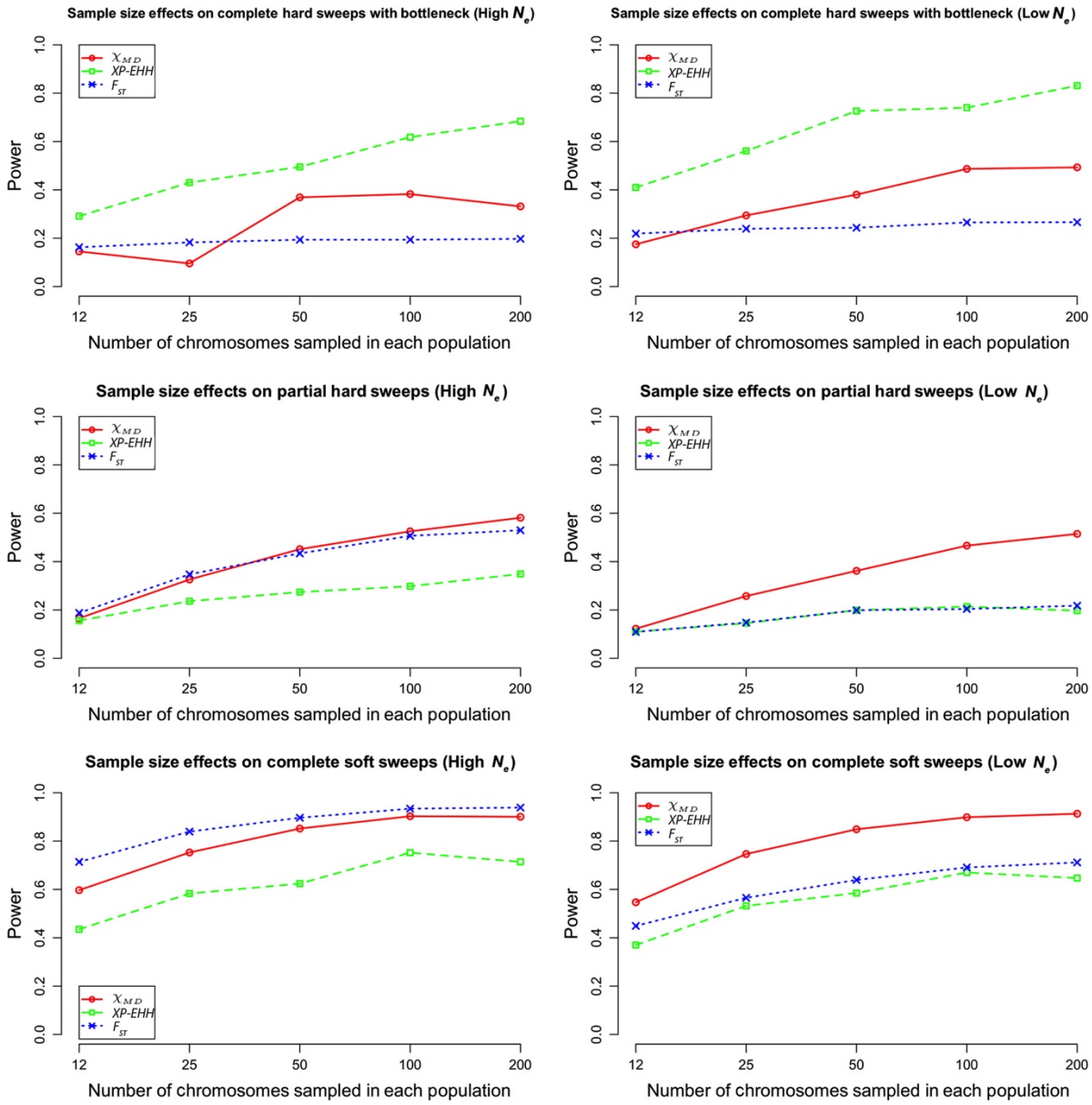


Figure 8. Sample size effects on each statistic. Bottleneck strength in the high N_e case is 0.01 while in the low N_e case it is 0.025. Ending frequency of partial hard sweeps is 0.3. Starting allele frequency is 0.001 for the high N_e complete soft sweep is and 0.02 for the low N_e case.

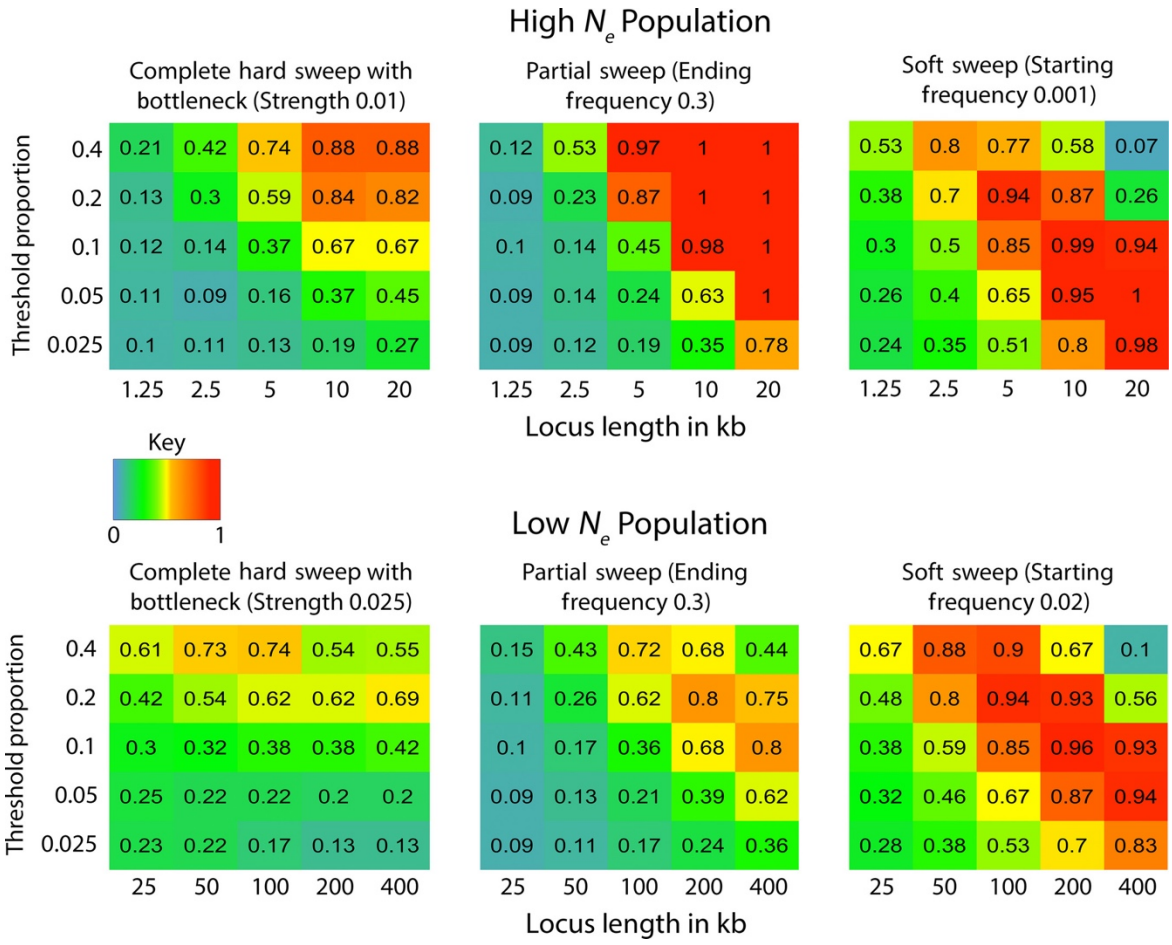


Figure 9. For selected sweep scenarios, this heat map shows χ_{MD} power for differing window lengths and threshold proportions (the fraction of a window that must be identical between two haplotypes).

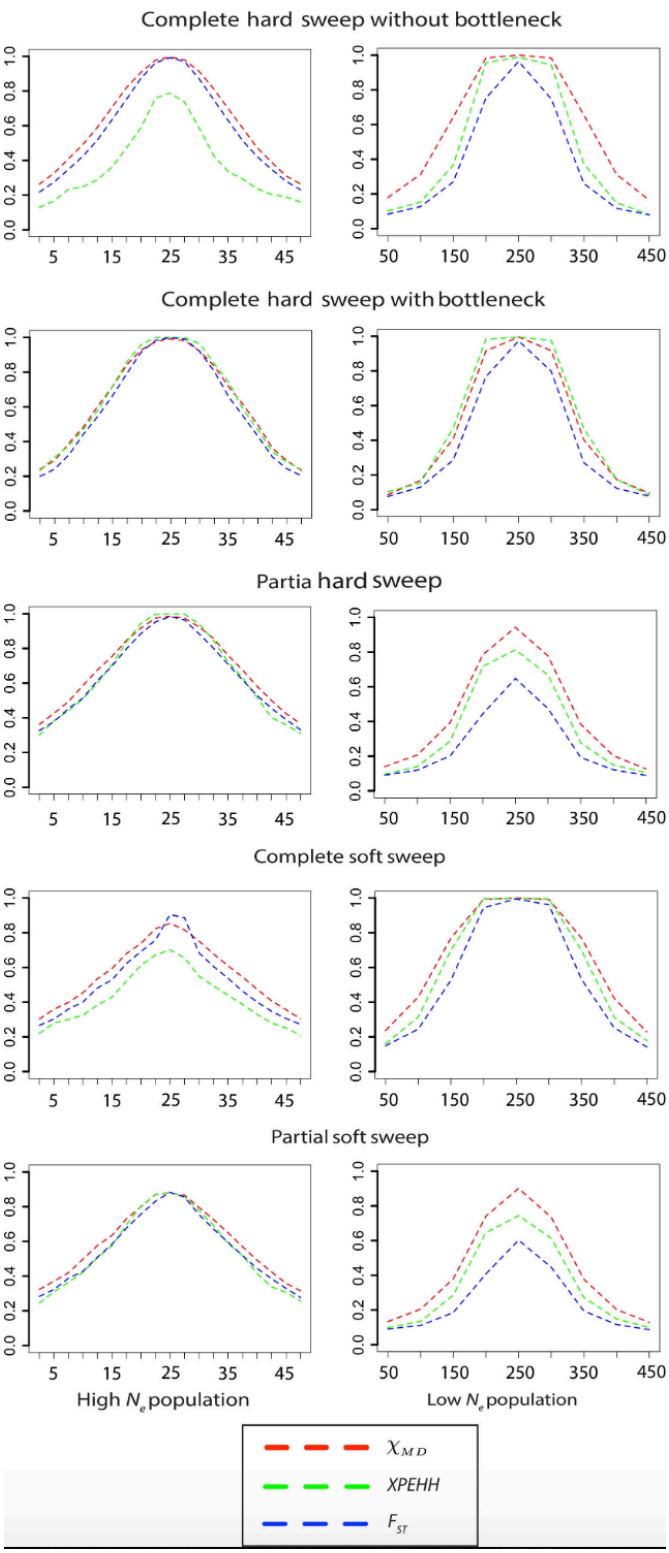


Figure 10. For a subset of sweep scenarios, this figure illustrates the decay of all three statistics' power by distance (kilobases on the x-axis). In the nonbottleneck complete hard sweep, the high N_e populations split at 0.5 time units in the past and selection (s = 0.001) began at 0.2 time units in the past. In the low N_e population, the populations split at 0.2 time units in the past and selection (s = 0.01) began immediately. The bottleneck strength in the high N_e case is 0.05 and in the low N_e case is 0.1. In both cases of the partial hard sweep, the ending allele frequencies were 0.5. In the complete soft sweep cases for both populations, starting frequency was 0.001. In the partial soft sweep cases for the high N_e case, starting allele frequency was 0.0001 and ending allele frequency was 0.5. For the high N_e case, beneficial starting allele frequency was 0.001 and ended at 0.5.

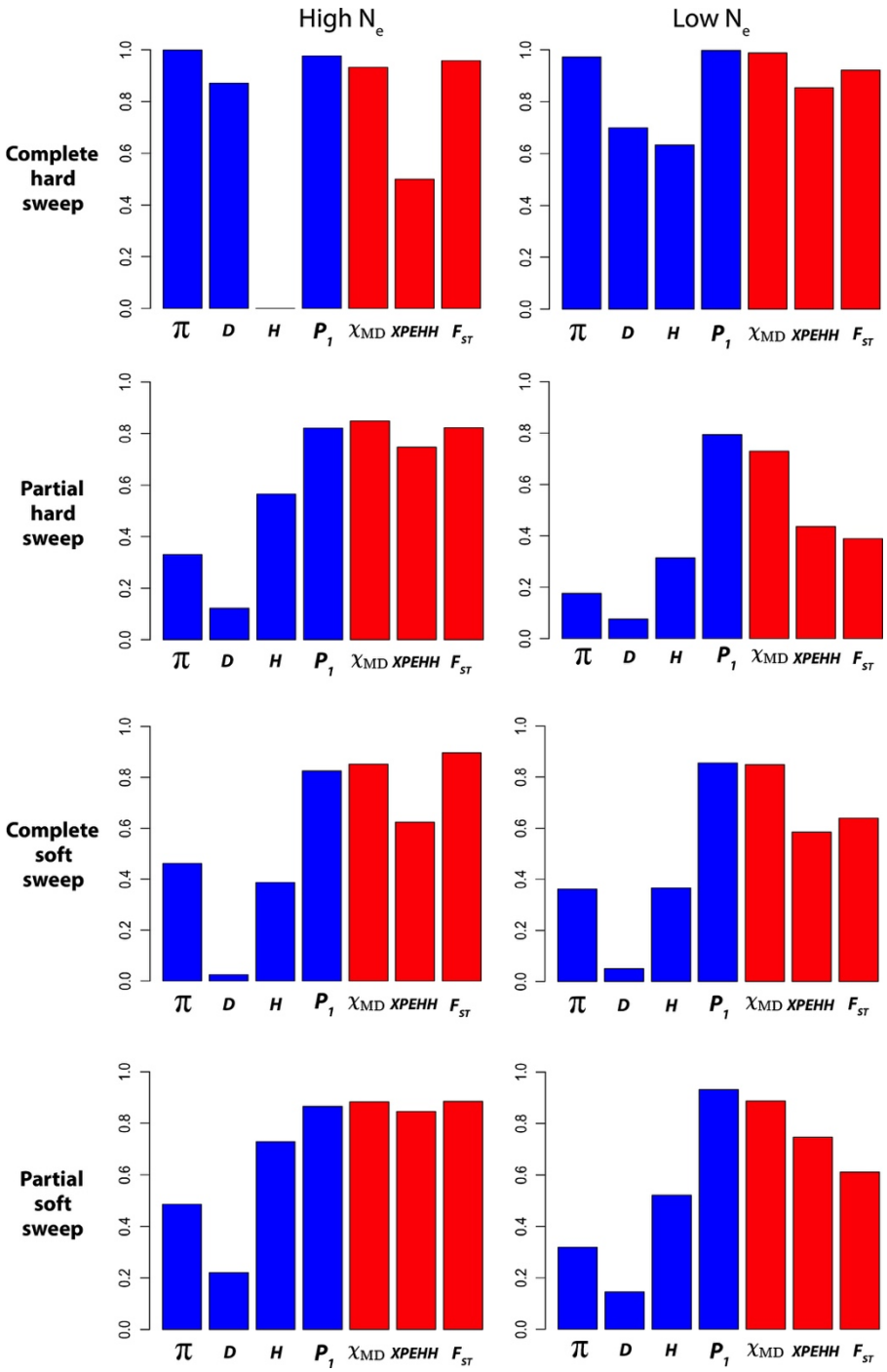


Figure 11. The power of four single-statistics calculated for an older complete hard sweep, a partial hard sweep, a complete soft sweep and a partial soft sweep. Note that simulation parameters differ between the high N_e and low N_e cases (Materials and Methods).

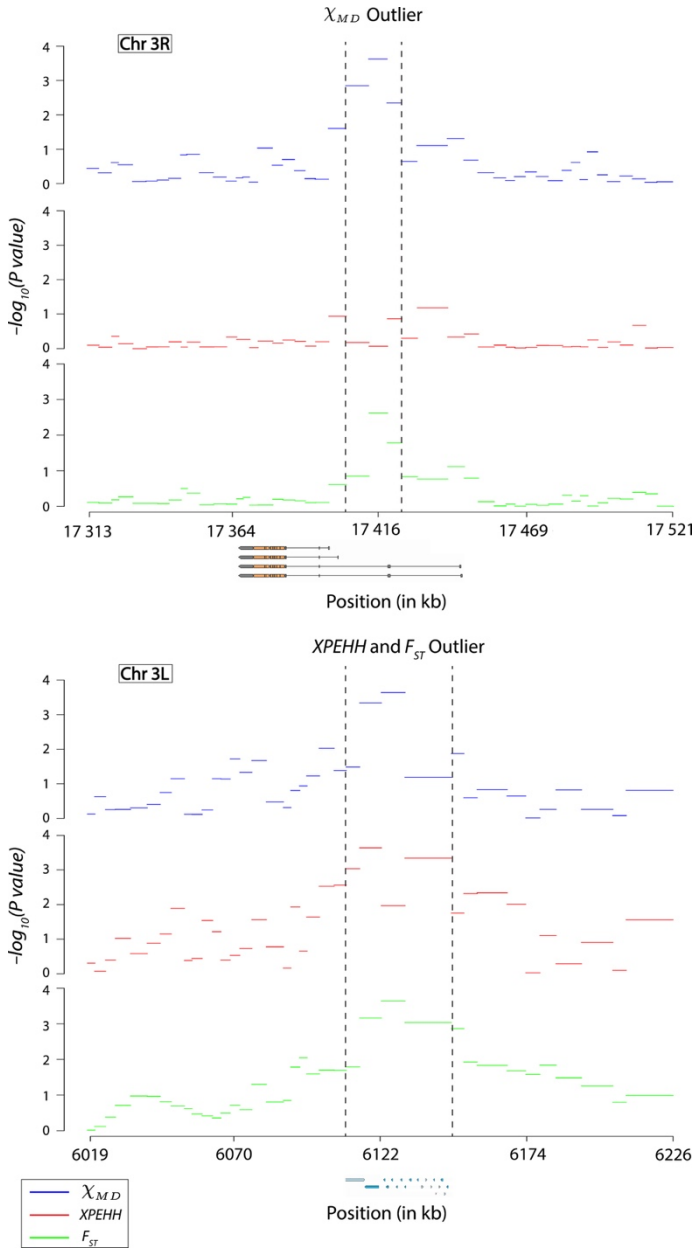


Figure 12. The top outlier regions and flanking windows for the empirical analysis of χ_{MD}, *XP-EHH*, and F_{ST} are shown. Above, the χ_{MD} outlier resides within a transcript region of the insulin receptor gene (*InR* alternative transcripts are shown). Below, *XP-EHH* and F_{ST} reached their maxima in the same outlier region (at adjacent windows), within a cluster of cuticle-related genes.

Chapter 2: Impacts of Recurrent Hitchhiking on Divergence and Demographic Inference in *Drosophila*

A version of this chapter has been published in *Genome Biology and Evolution.*

Citation: Lange JD, Pool JE (2018) Impacts of Recurrent Hitchhiking on Divergence and Demographic Inference in *Drosophila.* Genome Biology and Evolution, 10(8):1882-1891.

Abstract

In species with large population sizes such as *Drosophila*, natural selection may have substantial effects on genetic diversity and divergence. However, the implications of this widespread nonneutrality for standard population genetic assumptions and practices remain poorly resolved. Here, we assess the consequences of recurrent hitchhiking (RHH), in which selective sweeps occur at a given rate randomly across the genome. We use forward simulations to examine two published RHH models for *D. melanogaster*, reflecting relatively common/weak and rare/strong selection. We find that unlike the rare/strong RHH model, the common/weak model entails a slight degree of Hill–Robertson interference in high recombination regions. We also find that the common/weak RHH model is more consistent with our genome-wide estimate of the proportion of substitutions fixed by natural selection between *D. melanogaster* and *D. simulans* (19%). Finally, we examine how these models of RHH might bias demographic inference. We find that these RHH scenarios can bias demographic parameter estimation, but such biases are weaker for parameters relating recently diverged populations, and for the common/weak RHH model in general. Thus, even for species with important genome-wide

impacts of selective sweeps, neutralist demographic inference can have some utility in understanding the histories of recently diverged populations.

Introduction

The advancement of DNA sequencing technology, along with computational capacity and methodology, continues to revolutionize the field of population genetics. Harnessing the power of whole genome data sets, researchers have begun to explore a wider variety of evolutionary models. One such model that has received considerable attention recently is a model of recurrent hitchhiking, where genetic diversity at neutral regions is reduced due to repeated selective sweeps at nearby loci. This reduction in diversity has been explored theoretically (Kaplan et al. 1989; Stephan et al. 1992; Wiehe and Stephan 1993) showing that the expected reduction in diversity can be approximated as a function of RHH model parameters: $\gamma = 2N_e s$ and λ, where N_e is the effective population size, s is the selection coefficient, and λ is the rate of positively selected substitutions. Subsequent studies have examined such RHH models using forward simulation, focusing attention on how Hill–Robertson interference (HRI; Hill and Robertson 1966) between linked beneficial mutations on different haplotypes reduces the probability of fixation (Gerrish and Lenski 1998; Chevin and Hospital 2008).

The impact of natural selection on genomic diversity may be particularly significant for species with very large population sizes, such as *Drosophila melanogaster* (e.g., Sella et al. 2009; Langley et al. 2012). In abundant taxa, the population adaptive mutation rate is elevated and the weak influence of genetic drift may allow natural selection to favor alleles with modest selection coefficients. By estimating RHH parameters, Jensen et al. (2008) suggested that selective sweeps may reduce genomic diversity in *D. melanogaster* to half of neutral levels.

While this study implicated a model of relatively strong and infrequent sweeps, the study of Andolfatto (2007) instead favored a model of substantially weaker but more frequent adaptive substitutions. Though both of these studies utilized the same genomic data (synonymous polymorphism data at 137 X-linked loci in *D. melanogaster*) to infer selection strength and adaptive mutation rate, the methods of the two studies led to distinct conclusions. The Jensen study utilized an Approximate Bayesian Computation method to jointly infer adaptive mutation rate and selection strength. The Andolfatto study, however, used a maximum likelihood approach to estimate the product of the adaptive mutation rate and selection strength, followed by a McDonald–Kreitman approach to separate the two (McDonald and Kreitman 1991). While these models should imply strongly different proportions of substitutions driven by positive selection, their alignment with estimates of this quantity from genome-wide *Drosophila* data is unclear. And likewise, the predictions of each model for the role of HRI and for the fixation of neutral variants have not been investigated. We therefore sought to clarify the relationship between published *Drosophila* RHH models and adaptive divergence.

If linkage to natural selection substantially impacts *Drosophila* genetic diversity at neutral sites, the accuracy of demographic inference methods that assume neutrality is not assured. Most sites in the fly genome experience direct functional constraint (Halligan and Keightley 2006), which may lead to an excess of rare alleles from deleterious polymorphisms. Many sites that are not under as much direct selection pressure, such as synonymous sites and middles of short introns, are by definition very close to nonsynonymous sites and other functional sites that may experience natural selection. Selective sweeps could skew the genome-wide allele frequency spectrum, in particular by generating a skew toward rare alleles (Braverman et al. 1995) that may resemble the predictions of recent population growth. In line

with these concerns, Schrider et al. (2016) found that the presence of positive selection can bias demographic parameter estimates for a single population's history, and can lead to misidentification of demographic models. However, much interest centers on the inference of demographic parameters between recently diverged populations, and it remains unclear whether *Drosophila*-like RHH on shorter time-scales is sufficient to bias parameters concerning population divergence times, population-specific size changes, and migration rates. We therefore use RHH simulations to investigate the impact of RHH on estimation of these parameters.

Materials and Methods

McDonald–Kreitman Analysis

To estimate the proportion of substitutions in the *D. melanogaster* genome fixed by natural selection, we applied a genome-wide asymptotic McDonald–Kreitman analysis (McDonald and Kreitman 1991; Messer and Petrov 2013) using the web tool from Haller and Messer (2017). Here, we surveyed 197 genomes from a Zambian population of *D. melanogaster* (Lack et al. 2015), which is believed to be within the ancestral range of the species (Pool et al. 2012). These genomes are masked for identity by descent, apparent heterozygosity, and recent cosmopolitan admixture. In our analysis, we require any given site to be called in at least 50% of the genomes. We also applied a more conventional McDonald–Kreitman analysis, in which we required the minor allele to be segregating above 10% frequency in our sample. Applying this filter to both putatively neutral and selected site classes should reduce bias from deleterious polymorphisms. To estimate the number of substitutions, we used a *Drosophila simulans* genome aligned to the *D. melanogaster* genome (Stanley and Kulathinal 2016).

In this analysis, we estimated the proportion of substitutions driven to fixation by natural selection as $\alpha = \frac{D_s P_n}{D_n P_s}$ (Smith and Eyre-Walker 2002). Here, D_s is the number of synonymous

substitutions, P_n is the number of nonsynonymous polymorphic sites, D_n is the number of nonsynonymous substitutions, and P_s is the number of synonymous polymorphic sites. α was calculated for nine site classes (nonsynonymous, 2-fold synonymous, 3-fold synonymous, 5′ untranslated regions, 3′ untranslated regions, intron, intergenic, and RNA-coding) and individually for each major chromosome arm (2 L, 2 R, 3 L, 3 R, and X). Four-fold synonymous sites were evaluated as proxies for neutral evolution. Site classes were taken from flybase.org for release 5.56 of the *D. melanogaster* genome.

Simulations

In this study, we are interested in the effects of recurrent hitchhiking on demographic inference. To examine this, we ran forward simulations using SLIM version 2.5 (Haller and Messer 2016) to model recurrent hitchhiking. Because full-forward simulations are memory intensive and slow when simulating large populations, it is necessary to rescale simulation simulation parameters. We started by running test simulations to get an idea of the largest population size that we could simulate in a reasonable amount of time. We concluded that diploid populations of 50,000 individuals were a sensible target. This results in 50× rescaling assuming an effective population size of roughly 2,500,000 (e.g., Duchen et al. 2013). Following the results of Uricchio and Hernandez (2014), we determined that under the RHH models of interest, a size reduction to 50,000 individuals should closely maintain the genetic variation of a nonrescaled population. Further, both algorithms provided in the cited paper yielded near identical scaled parameters when reducing the population size from 2,500,000 to 50,000 individuals. Because of this, we used the simpler Algorithm 1 of Uricchio and Hernandez (2014). The main idea behind this rescaling method is that patterns of genetic diversity are maintained when population-scaled parameters $\theta = 4N_e\mu$, $\rho = 4N_e r$, and $\gamma = 2N_e s$ are fixed while N_e is varied.

As such, if we decrease N_e=2,500,000 by 50× to N_e=50,000, then θ, ρ, and α must be increased by 50×. The algorithm is laid out in step form below.

Let $s_0 = \alpha_0/2N_0$; $r_0 = \rho_0/4N_0$; $a = s_0/L_0r_0$
$\gamma_1 = \gamma_0$
$s_1 = \gamma_1/2N_1$
$r_1 = s_0/aL_1$
$\lambda_1 = r_1\lambda_0/r_0$

Here, the subscripts refer to before rescaling (subscript 0) and after rescaling (subscript 1). s_1 is the selection strength, N_1 is the population size, r_1 is the per base pair per generation per chromosome recombination rate, and L_1 is the simulated locus length.

We ran simulations under two different models of RHH, both of which were estimated from *D. melanogaster* data. Since the rate of adaptive substitutions and the average selective advantage are highly confounded in terms of their impact on diversity levels, we wanted to examine complementary models. The first model we chose to study is from Jensen et al. (2008). Here, the rate of incoming adaptive substitutions (λ) is low (λ = 4.2E-11; $2N\lambda$ = 2.1E-4) and the average strength of selection (s = 0.002; 2Ns=10,000) is high. The second model, from Andolfatto (2007), consisted of a high rate of adaptive substitutions (λ = 6.9E-10; $2N\lambda$ = 3.45E-3) with a very low average selection strength (s = 1.2E-5; $2Ns$ = 60). For both models, we used a mutation rate μ = 3.27E-9 ($4N\mu$ = 0.0327) (Schrider et al. 2013) and a recombination rate r = 2.5E-8 ($4Nr$ = 0.25). We also modeled gene conversion at a rate of 6.25E-8/bp/generation. The tract length of the gene conversion was drawn from a geometric distribution with a mean length of 518 base pairs (Comeron 2012). In forward simulations, it is not possible to directly specify the rate of adaptive substitutions. Instead, one must input the rate of beneficial

mutations v. In the absence of interference among selected mutations, this can be derived using λ and the probability of fixation (Kimura 1962):

$$\lambda = vP_{fix}$$
$$v = \frac{\lambda}{P_{fix}}$$
$$v = \lambda \Big/ \frac{1 - e^{-2s}}{1 - e^{-2\gamma}}$$

In our simulations, each beneficial mutation had its selection coefficient drawn randomly from an exponential distribution with a mean equal to about half of the rescaled selection strength. The means of these distributions were chosen such that the average selection strength of a fixed mutation is the rescaled selection strength, s_1 (since the more strongly beneficial mutations drawn from this distribution are more likely to fix). The variation in selection coefficients helps to avoid the artificial scenario of interference between mutations with precisely identical fitness.

We wanted to analyze ten kilobases (kb) from each simulation for demographic inference. Because a sweep can affect regions far from the target of a sweep, we simulated extra flanking regions for each side of the 10 kb that was used for the demographic inference, while analyzing only the middle region. For the common/weak sweep model, we simulated 480 base pairs on each side of the 10 kb for a total of 10,960 base pairs simulated. For the rare/strong model, we simulated 20 kb flanking loci for a total of 50 kb simulated. Using the formula $(2N_e s)^{-2r/s}$ (Smith and Haigh 1974; Barton 2010), we expect a 1% reduction in neutral diversity from a sweep 20 kb away under the rare/strong model, while under the common/weak model a sweep should reduce diversity 480 bp away by only 0.03%. Hence, these simulations should incorporate a large majority of sweep effects on neutral diversity predicted by the associated RHH models,

while maintaining computational tractability. RHH simulation parameters are provided in table 1.

Demographies Simulated

Forward simulations require a burn-in period to generate appropriate genetic variation. Thus, both recurrent hitchhiking models were run for 500,000 ($10N_e$) generations. These "trunk" simulations were then used for the demographic simulations. There are two relevant two-population demographic models that we were interested in. These demographies included a bottleneck model and an isolation with migration (IM) model. In our bottleneck model, the populations split and one population experiences a bottleneck. The parameters of the bottleneck model were taken from Thornton and Andolfatto (2006). In this model, a bottleneck occurs 0.0516 coalescent time units in the past (5,160 generations) and lasts for 0.042 coalescent time units (4,200 generations). During the bottleneck, the population decreases to 4.7% of its original population size. The IM model consisted of a population split $0.5N_e$ coalescent units in the past (25,000 generations) and subsequent migration of $2N_em = 0.25$. We simulated two different kinds of IM models, a "shared sweep" and a "private sweep" model. The "shared sweep" model allowed selective mutations to have an equal selective advantage if they migrated into the other population. The "private sweep" model multiplied s by -1 if it migrated into the other population, making the allele deleterious.

Demographic Inference

To examine how recurrent hitchhiking affects demographic inference, we used δαδι version 1.6.3 (Gutenkunst et al. 2009) to estimate demographic model parameters. We first attempted to fit a two-epoch and three-epoch size change model to the trunk simulations (where there were no size changes simulated) to examine whether recurrent hitchhiking can misidentify

demographic models. For the bottleneck simulations, we fit two and three parameter bottleneck models. The three parameter bottleneck model consisted of a population size reduction, a length of time as the reduced population size, and an instantaneous size change back to the original size that occurs some time in the past. In the two population bottleneck model, the length of the bottleneck is fixed and not optimized. We examined each bottleneck model both with and without fitting the ancestral size change models as well. In this way, we could better parse ancient parameters from more recent demographic parameters post population split. Finally, we tested the IM models with both shared and private selective sweeps. In these cases, the timing of the population split and the migration rate are estimated. As in the bottleneck cases, we tested both IM models with and without ancestral size changes. In total, 14 demographic models were investigated for both hitchhiking models and 11 demographic models were tested for the neutral simulations (since there is no shared/private sweep distinction in the neutral case).

We ran 1,000 simulations of both the common/weak RHH and the rare/strong RHH model. For any given demographic model tested, we randomly chose 50 simulations to generate a site frequency spectrum to run δαδι on. δαδι was then run ten times on each SFS. The parameters of the δαδι run with the highest likelihood across the ten runs were chosen as the inferred parameters. We repeated this process 200 times for each demography tested.

Data Deposition

This study produced no empirical data. All scripts necessary to recapitulate the analyses presented can be found at http://github.com/jeremy-lange/RHH_project.

Results

Effects of Hill–Robertson Interference

In order to investigate the effects of recurrent hitchhiking on divergence and diversity in *Drosophila*, we performed forward simulations reflecting two published models of RHH representing relatively common/weak selection (Andolfatto 2007) and rare/strong selection (Jensen et al. 2008), respectively. Before proceeding with further analysis, we checked to see if our initial adaptive mutation rates based on these models were producing the prescribed rates of adaptive substitution, or if instead an important impact of interference must be accounted for. While the above studies assumed no interference between positively selected mutations, Andolfatto (2007) suggested that an interesting next step would be to examine how the presence of interference influences the observed versus expected adaptive substitution rate (λ) and selection coefficient of fixed beneficial substitutions (s) in the simulations.

Under a model of no interference, we can expect approximately λLgP_{fix} adaptive substitutions as a product of the adaptive substitution rate at L sites across g generations and their probability of fixation. However, if multiple beneficial mutations are sweeping simultaneously, competition among sweeping haplotypes will lead fewer mutations to fix. Further, mutations that do fix will tend to have a higher selection strength than the input distribution of selective effects. The dynamics of the common/weak and rare/strong models that we tested are very different. In the rare/strong model, on average, we expect 554 beneficial mutations to occur per simulation (over the full 50 kb locus), with 52.2 fixing on average during the 500,000 generations. We expect one beneficial fixation approximately every 9,523 generations. It is unlikely, therefore, that any given beneficial mutation would experience interference from another. In the common/weak model, however, we expect 315,197 beneficial mutations per simulation. Under a model of no interference, we would expect on average 189 of these beneficial mutations to fix. This makes it very likely that more than one beneficial mutation would be sweeping at any given

time. Thus, we expect interference to be more significant in the common/weak model relative to the rare/strong model.

To test this expectation, we ran both the common/weak and the rare/strong models as described in the methods section, using the published λ and s to generate input parameters. We tracked every mutation that fixed across all simulations and recorded the average selection coefficient. To accurately reflect the RHH models that we were simulating, our goal for our simulations was to approximately match the expected number of fixed beneficial mutations as described above. Across the 1,000 simulations of the rare/strong model, an average of 51.8 beneficial mutations fixed per simulation compared to an expectation of 52.2. For the common/weak RHH model, we found that, on average, 182 beneficial mutations fixed, corresponding to a modest 3.7% reduction in the expected number of adaptive fixations. We attribute this reduction to interference between positively selected mutations. In order to emulate the properties of the common/weak model, we increased the adaptive mutation rate by 3.7%. This recovered the desired rate of adaptive fixations, averaging 188 adaptive fixations per simulation.

We also conducted simulations to test whether our inclusion of gene conversion was crucial, and we found that it indeed had an important effect on the degree of Hill–Roberton interference. In simulations without gene conversion, the retuned common/weak model averaged 182 adaptive fixations. Thus, the absence of gene conversion reduced the rate of adaptive substitution by 3.3%.

We used the above retuned ν parameter and ran 1,000 simulations as described in the methods section to more accurately reflect the common weak sweep RHH model. Output of these simulations was used for the simulation analyses detailed below. Note that for a

nonrescaled *Drosophila* population, our adjusted RHH mutational parameters for the Andolfatto (2007) model would correspond to a beneficial mutation rate of $v = 5.5E\text{-}12$.

Impacts of RHH on Adaptive Divergence

In these simulations, we can calculate the proportion of substitutions fixed by selection. In the rare/strong RHH model, 1.23% of all fixations are adaptive while in the retuned common/weak model, 17.5% of fixations are adaptive (table 2). These proportions can be compared with the same quantity (α) estimated from extended McDonald–Kreitman analyses of empirical data (Smith and Eyre-Walker 2002). Our genome-wide analysis estimated that approximately 18.8% of substitutions in the *D. melanogaster* genome were driven to fixation by natural selection (fig. 1; supplementary table S1, supplemental files). Our estimates of α are largely concordant with other studies (Andolfatto 2005, Begun et al. 2007, but see Mackay et al. 2012). Perhaps surprisingly, α estimates for 2 and 3-fold synonymous site classes were very low and hence indicated evolutionary patterns similar to 4-fold synonymous sites. It is unclear why these sites may not have been frequent targets of adaptive protein evolution and, further, it is unclear why α estimates did not converge to a particular range when the minimum allele frequency threshold was increased in the asymptotic McDonald–Kreitman method (Messer and Petrov 2013). For full results, see supplementary table S1 (supplemental files), for nonasymptotic α estimates with a simple 10% frequency threshold. Overall, this empirical analysis suggests that the common/weak RHH model seems more compatible with adaptive divergence estimates in *D. melanogaster*.

Hitchhiking Effects on Demographic Inference

The final goal of this study was to examine how models of recurrent hitchhiking affect inferences on demography. We tested whether selective sweep models involving substantial

hitchhiking effects would violate assumptions of neutrality made by demographic inference tools and bias parameter estimation. We ran our simulations for 500,000 generations as a single population before we added a population split and distinct demographies. As previously demonstrated (Jensen et al. 2008), the rare/strong model entails a large impact on nucleotide diversity (π reduced by over 50% vs. neutrality; supplementary table S2, supplemental files). This model also generates a notable excess of rare alleles (supplementary fig. S1, appendix 2), in line with theoretical expectations (Braverman et al. 1995). By comparison, the common/weak model reduces π by just 15% and does not produce an excess of rare alleles.

For the "trunk" simulations, we first asked if δαδι would fit a model with population size changes over the true model of constant size. Here, we tested both two and three epoch size change models. The two epoch model consisted of two parameters to maximize: a single size change (reduction or expansion) that occurs some time in the past. The three epoch model has three parameters: a size change (reduction or expansion), a length of time at this new size, and some time in the past when the population recovers. The parameters of these models were most affected by the RHH models. Under neutrality, δαδι prefers a model of constant population size (the true model) for both the two and the three epoch models (fig. 2). In the two epoch size change model, δαδι infers a population expansion for the rare/strong sweep model, in line with the observed excess of rare alleles. For the three epoch model, a bottleneck ending $1.2N_e$ generations ago is instead favored. Qualititatively similar single-population results have recently been reported (Schrider et al. 2016). However, for common/weak sweep model, estimated population size changes are subtle (<2-fold) and quite ancient (>$6N_e$ generations ago), in agreement with the lesser impact of RHH on genetic variation under this model.

Our primary interest was to assess whether the demographic parameters that relate recently diverged populations are similarly biased by RHH. We therefore attempted to infer two distinct bottleneck models occurring after the 500,000 generation burn-in: a two and a three parameter model. Both models consisted of a population size contraction, a bottleneck length, and a length in time since recovery back to original population size. All three parameters are inferred in the three parameter model while bottleneck length is fixed and not inferred in the two parameter model. Results showed that in the two parameter bottleneck model, bottleneck strength was accurately inferred under neutrality and both selection models (supplementary table S3, supplemental files). In the three parameter bottleneck case, bottleneck strength and bottleneck duration were accurately inferred under neutrality and in the common/weak RHH model, with little upward bias in the rare/strong RHH case (fig. 3; supplementary table S3, supplemental files). Time since recovery was most affected by the presence of selection. Under neutrality, this parameter was accurately recapitulated for both the two and three parameter bottleneck models. Under selection, however, this parameter was overestimated by at least twofold. This result is in line with the fact that both post-bottleneck growth and selective sweeps leave behind an excess of rare alleles. Thus, if both are occurring in our simulation but δαδι assumes that only neutral events have occurred, it makes sense that this inference method is biased towards overestimating bottleneck recovery times in the presence of positive selection. Models were also run in which ancestral size change parameters and bottleneck parameters were estimated simultaneously, with qualitatively similar results (supplementary table S4, supplemental files).

We simulated two distinct IM models: one with shared selective sweeps between populations, and one simulating local adaptation. For both cases, we estimated the time since the

population split and the migration rate. It should be noted that the timescale simulated here, $0.5N_e$, is much longer than the estimated divergence of any *D. melanogaster* populations (Duchen et al. 2013; Kern and Hey 2017). This scenario, therefore, should be viewed as an extreme scenario in terms of divergence time. Under neutrality, δαδι correctly estimated both the divergence time and the migration rate. Under both RHH models, however, these parameter estimates were biased in both the shared sweep and local adaptation cases. In the shared sweep model, δαδι overestimated migration rates, with less bias for divergence time (fig. 4; supplementary table S3, supplemental files). In contrast, for the local adaptation case, the divergence time was greatly overestimated while the migration rate was less biased. As with the bottleneck models, we also made demographic inferences by combining the ancestral size change models with the recent divergence IM models, again with comparable results (supplementary table S4, supplemental files).

Discussion

Simulation provides us with flexible tools for studying the impact of selective sweeps on genetic diversity and divergence. In this study, we have examined two published RHH models and their consequences for divergence and demographic inference. We found that a model of common and weak positive selection appears to fit the genomic estimates of adaptive divergence better than a rare/strong sweep model, and that this favored model appears to entail a slight degree of Hill–Robertson interference. After retuning the common/weak model to yield adaptive substitution rates similar to the published parameters, we investigated the impact of both RHH models on demographic inference. Here, we confirmed that RHH can bias demographic parameters, but we found that the magnitude of such bias was greater for long term population parameters and lesser for more recent parameters.

Although we find agreement between the common/weak model and McDonald–Kreitman estimates of adaptive divergence, we do not argue for a specific quantitative RHH model in light of some important caveats. First, the model that gives α estimates in line with our genome-wide estimates is based on an analysis of nonsynonynous sites specifically (Andolfatto 2007). While this model may share traits in common with the true genomic RHH model for this species, it is best viewed as a qualitative example of the type of model compatible with this aspect of data. Second, we are assessing RHH models based on their general agreement with McDonald–Kreitman-based estimates of α, but such estimates can be biased depending on the history of population size change (Eyre-Walker 2002) and recombination rate change (Comeron et al. 2014). Finally, estimates of α derived from RHH models depend not only on adaptive parameters, but also upon the *neutral* mutation rate that we use in simulations. The raw mutation rate may be somewhat higher than our simulated μ (Schrider et al. 2013; Huang et al. 2016). However, more than half of these mutations should be prevented from fixing by selective constraint (Halligan and Keightley 2006). If theμ that we use in our simulations is too high, the predictions for α may be too low, and vice versa. In spite of these quantitative uncertainties, we argue that our analyses provide general insight into the RHH models that are most plausible for this species.

Although this is not an inference study, our results suggest the importance of HRI in a common/weak RHH model like that of Andolfatto (2007). We found that this model led to a loss of approximately 3.7% of beneficial substitutions before retuning the adaptive mutation rate. These results raise the possibility that the effects of interference in *D. melanogaster* may be stronger than previously suggested (Castellano et al. 2016). In this 2016 study, the authors conclude that approximately 27% of beneficial substitutions are lost due to HRI. This estimate is

based on the assumption that there is no interference in regions of recombination that exceed 2 cM/Mb. However, our results raise the prospect of slight interference in regions of high recombination (2.5 cM/Mb). Therefore, the genomewide effects of interference on the adaptive substitution rate may be somewhat stronger than previously estimated.

Our results also shed light on the impact of RHH on linked neutral variation and its consequences for demographic estimation. It is clear that strong effects of natural selection across the genome can violate the neutralist assumptions of typical demographic inference methods, whether due to background selection (Ewing and Jensen 2016) or selective sweeps (Schrider et al. 2016). Our results support the biasing effect of RHH on ancient parameter estimation and demographic model choice. This is in line with a previous study using single population simulations (Schrider et al. 2016). However, we show that there can be relatively less impact on recent parameters in two population models. Thus, such methods may retain utility even for populous species like *D. melanogaster*, but the biases that do exist should be borne in mind for this species and investigated for other taxa.

Our simulations emulate a population of *D. melanogaster*, a populous species with a compact genome. It has been shown that selection may be prevalent in large fly populations (Sella et al. 2009; Karasov et al. 2010; Langley et al. 2012). Further, due to the compactness of the genome, a selective sweep can affect relatively large regions. It is less clear how much of an impact natural selection would have on demographic inference in smaller N_e species, such as humans. It is presumed that much of the genetic variation in the human genome is most affected by genetic drift instead of natural selection. Recent studies, however, have argued that an appreciable rate of adaptive substitution has shaped genetic variation in humans (Enard et al. 2014) and that soft sweeps play the dominant role in adaption in human evolution (Schrider and

Kern 2017). Thus, even though theory suggests that demographic inference should be more accurate in a less populous species where the effects of natural selection should be lessened, further study is needed to delimit the parameter space in which RHH biases demographic inference.

The simulations in this study entail specific, important caveats. First, we simulated data reflecting the highly recombining portion of the *Drosophila* genome. The inclusion of low recombination regions would presumably exacerbate the effects of selective sweeps, HRI, and background selection on genetic variation and hence demographic estimates. Our simulations also did not model selective constraint. The biases we observed, therefore, would presumably be worse if analyzed sites had an excess of rare alleles due to deleterious variants. The RHH models that we simulated invoke an arbitrary distribution around a point estimate of s. Importantly, we do not know the true distribution of fitness effects in nature. Theoretical work has suggested that advantageous mutations may be exponentially distributed (Gillespie 1984, Orr 2003), as we have done in this study, although processes such as migration-selection balance may lead to departures from this prediction (Yeaman and Whitlock 2011). There are important caveats to such a distribution in regards to our inference. For instance, the size of the flanking region was determined by the point estimate of the selection strength. Any selective sweep with a selective strength substantially greater than the mean of the distribution will not fully be captured by our analyzed region.

The ABC method used in the Jensen study attempts to capture the locus-to-locus variance in genetic variation to infer adaptive mutation rate and selection strength. Such a method is highly dependent on the size of the genomic regions used in the study. Longer loci entail higher discriminatory power to distinguish between a rare/strong RHH model and a common/weak

RHH model because common/weak selection reduces variance from one locus to the next even in small genomic windows. The loci used in the Jensen and Andolfatto studies were on average 680 base pairs, potentially biasing the ABC method of the Jensen study towards high values of *s*. The Andolfatto (2007) study utilized a McDonald–Kreitman approach to estimate *s*, which is not sensitive to the size of locus length. Our estimates of adaptive divergence are in accordance with the results of both the Andolfatto (2007) study and also a more recent study (Keightley et al. 2016). Using *D. melanogaster* polymorphism data, Keightley suggested a scaled selection strength of $N_e s = 12$. This is very close to our common/weak model's scaled selection strength of $N_e s = 15$ Further, the inferred probability that a new mutation is beneficial from the Keightley study (0.5%) is very close to the retuned estimate from this study (0.4%). Of course, while we draw these conclusions about weak/common RHH models, it is still possible that there could be a number of strong sweeps in the genome.

Our simulations only modeled complete sweeps from new mutations. It has been argued that this is not necessarily the dominant adaptive model in nature (Pritchard et al. 2010; Hernandez et al. 2011; Schrider and Kern 2017), so it is worth considering how other models of natural selection may alter conclusions drawn in this study. If there are two simultaneous soft sweeps, for instance, it is more likely that there exists a haplotype with both favorable variants prior to selection starting, reducing the impact of HRI. Likewise, because soft sweeps have more limited impacts on genetic variation (Pennings and Hermisson 2006), their impact on demographic estimation should be less severe as well.

It is also important to note that complete sweeps may only be a single component of Darwinian selection in nature. Other models of selection may impact genetic variation in fly populations with less effect on divergence, including fluctuating selection (Mustonen and Lässig

2010; Bergland et al. 2014) and diminishing selection (Vy et al. 2017). Thus, depending on the modes of positive selection that are prevalent in nature, the total impact of hitchhiking on genetic variation and demographic inference may be greater or lesser than simulated here—underscoring the need for further investigation of this topic.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgements

The UW-Madison Center for High Throughput Computing provided computational assistance and resources for this work. This work was funded by USDA Hatch grant WIS01900 and NIH NIGMS training grant T32 GM007133.

References

Andolfatto P. 2005. Adaptive evolution of non-coding DNA in *Drosophila*. Nature 437:1149–1152.

Andolfatto P. 2007. Hitchhiking effects of recurrent beneficial amino acid substitutions in the *Drosophila melanogaster* genome. Genome Res. 17:1755–1762.

Begun DJ, et al. 2007. Population genomics: whole-genome analysis of Polytene Chromosome Maps in *Drosophila* 1651 polymorphism and divergence in *Drosophila simulans*. PLoS Biol. 5:2534–2559.

Bergland AO, Behrman EL, O'Brien KR, Schmidt PS, Petrov DA.. 2014. Genomic evidence of rapid and stable adaptive oscillations over seasonal time scales in *Drosophila*. PLoS Genet. 10:e1004775.

Braverman JM, Hudson RR, Kaplan NL, Langley CH, Stephan W.. 1995. The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. Genetics 140:783–796.

Castellano D, Coronado-Zamora M, Campos JL, Barbadilla A, Eyre-Walker A.. 2016. Adaptive evolution is substantially impeded by Hill–Robertson interference in *Drosophila*. Mol Biol Evol. 33:442–455.

Chevin LM, Hospital F.. 2008. Selective sweep at a quantitative trait locus in the presence of background genetic variation. Genetics 180:1645–1660.

Comeron JM, Ratnappan R, Bailin S.. 2012. The many landscapes of recombination in *Drosophila melanogaster*. PLoS Genet. 8:e1002905,

Comeron JP. 2014. Background selection as baseline for nucleotide variation across the *Drosophila* genome. PLoS Genet. 10:e1004434..

Duchen P, Zivkovic D, Hutter S, Stephan W, Laurent S.. 2013. Demographic inference reveals African and European admixture in the North American *Drosophila melanogaster* population. Genetics 193:291–301.

Enard D, Messer PW, Petrov DA.. 2014. Genome-wide signals of positive selection in human evolution. Genome Res. 24:885–895.

Ewing GB, Jensen JD.. 2016. The consequences of not accounting for background selection in demographic inference. Mol Ecol. 25:135–141.

Eyre-Walker A. 2002. Changing effective population size and the McDonald–Kreitman test. Genetics 162:2017–2024.

Gerrish P, Lenski R.. 1998. The fate of competing beneficial mutations in an asexual population. Genetica 102–103:127–144.

Gillespie JH. 1984. Molecular evolution over the mutational landscape. Evolution 38:1116–1129.

Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD.. 2009. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. PLoS Genet. 5:e1000695.

Haller BC, Messer PW.. 2016. Slim 2: flexible, interactive forward genetic simulations. Mol Biol Evol. 34:230–240.

Haller BC, Messer PW.. 2017. asymptoticMK: a web-based tool for the asymptotic McDonald–Kreitman test. G3 75:1569–1575.

Halligan DL, Keightley PD.. 2006. Ubiquitous selective constraints in the *Drosophila* genome revealed by a genome-wide interspecies comparison. Genome Res. 16:875–884.

Hernandez RD, et al. 2011. Classic selective sweeps were rare in recent human evolution. Science 331:920–924.

Hill WG, Robertson A.. 1966. The effect of linkage on limits to artificial selection. Genet Res. 83:269.

Huang W, et al. 2016. Spontaneous mutations and the origin and maintenance of quantitative genetic variation. eLife 5:e14625.

Jensen JD, Thornton KR, Andolfatto P.. 2008. An approximate Bayesian estimator suggests strong, recurrent selective sweeps in *Drosophila*. PLoS Genet. 49:e1000198.

Kaplan NL, Hudson R, Langley C.. 1989. The “hitchhiking effect” revisited. Genetics 1234:887–899.

Karasov T, Messer SW, Petrov DA.. 2010. Evidence that adaptation in *Drosophila* is not limited by mutation at single sites. PLoS Genet. 66:e1000924.

Kern AD, Hey J.. 2017. Exact calculation of the joint allele frequency spectrum for isolation with migration models. Genetics 2071:241–253.

Kimura M. 1962. On the probability of fixation of mutant genes in a population. Genetics 476:713–719.

Lack JB, et al. 2015. The *Drosophila* genome nexus: a population genomic resource of 623 *Drosophila melanogaster* Genomes, including 197 from a single ancestral range population. Genetics 1994:1229–1241.

Langley CH, et al. 2012. Genomic variation in natural populations of *Drosophila melanogaster*. Genetics 1922:533–598.

Mackay TF, et al. 2012. The *Drosophila melanogaster* genetic reference panel. Nature 4827384:173–178.

Smith JM, Haigh J.. 1974. The hitch-hiking effect of a favourable gene. Genet Res. 2301:23–35.

McDonald JH, Kreitman M.. 1991. Adaptive protein evolution at the Adh locus in *Drosophila*. Nature 3516328:652–654.

Messer PW, Petrov DA.. 2013. Frequent adaptation and the McDonaldKreitman test. Proc Natl Acad Sci U S A. 11021:8615–8620.

Mustonen V, Lässig M.. 2010. Fitness flux and ubiquity of adaptive evolution. Proc Natl Acad Sci U S A. 1079:4248–4253.

Orr HA. 2003. The distribution of fitness effects among beneficial mutations. Genetics 1634:1519–1526.

Pennings PS, Hermisson J.. 2006. Soft sweeps III: the signature of positive selection from recurrent mutation. PLoS Genet. 212:e186.

Pool JE, et al. 2012. Population genomics of Sub-Saharan *Drosophila melanogaster*: african diversity and non-African admixture. PLoS Genet. 812:e1003080.

Pritchard JK, Pickrell JK, Coop G.. 2010. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. Curr Biol. 204:R208–R215.

Schrider DR, Houle D, Lynch M, Hahn MW.. 2013. Rates and genomic consequences of spontaneous mutational events in *Drosophila melanogaster*. Genetics 1944:937–954.

Schrider DR, Shanku AG, Kern AD.. 2016. Effects of linked selective sweeps on demographic inference and model selection. Genetics 2043:1207–1223.

Schrider DR, Kern AD.. 2017. Soft Sweeps are the dominant mode of adaptation in the human genome. Mol Biol Evol. 348:1863–1877.

Sella G, Petrov DA, Przeworski M, Andolfatto P.. 2009. Pervasive natural selection in the *Drosophila* genome? PLoS Genet. 56:e1000495.

Smith NGC, Eyre-Walker A.. 2002. Adaptive protein evolution in *Drosophila*. Nature 4156875:1022–1024.

Stanley CE, Kulathinal RJ.. 2016. Genomic signatures of domestication on neurogenetic genes in *Drosophila melanogaster*. BMC Evol Biol. 16:6.

Stephan W, Wiehe TH, Lenz MW.. 1992. The effect of strongly selected substitutions on neutral polymorphism: analytical results based on diffusion theory. Theor Popul Biol. 412:237–254.

Thornton K, Andolfatto P.. 2006. Approximate Bayesian inference reveals evidence for a recent, severe bottleneck in a Netherlands population of *Drosophila melanogaster*. Genetics 1723:1607–1619.

Uricchio LH, Hernandez RD.. 2014. Robust forward simulations of recurrent hitchhiking. Genetics 1971:221–236.

Vy HMT, Won YJ, Kim Y.. 2017. Multiple modes of positive selection shaping the patterns of incomplete selective sweeps over African populations of *Drosophila melanogaster*. Mol Biol Evol. 3411:2792–2807.

Wiehe T, Stephan W.. 1993. Analysis of a genetic hitchhiking model, and its application to DNA polymorphism data from *Drosophila melanogaster*. Mol Biol Evol. 104:842–854.

Yeaman S, Whitlock MC.. 2011. The genetic architecture of adaptation under migration–selection balance. Evolution 657:1897–1911.

Table 1

Summary of simulation parameters.

Table 1

Summary of Simulation Parameters

	Prerescale	Postrescale	Prerescale	Postrescale	Prerescale	Post-rescale
	Rare Strong Sweep		Common Weak Sweep		Retuned Common Weak Sweep	
Population size	2,500,000	50,000	2,500,000	50,000	2,500,000	50,000
Neutral mutation rate	3.27E-09	1.64E-07	3.27E-09	1.64E-07	3.27E-09	1.64E-07
Beneficial mutation rate	4.21E-15	2.23E-13	4.89E-12	5.75E-10	5.50E-12	5.97E-10
Recombination rate	2.50E-08	1.25E-06	2.50E-08	1.25E-06	2.50E-08	1.25E-06
Gene conversion rate	6.25E-08	3.125E-06	6.25E-08	3.125E-06	6.25E-08	3.125E-06
Selection strength (Target)	0.002	0.1	1.20E-05	6.00E-04	1.20E-05	6.00E-04
Selection strength (Input)	NA	0.053	NA	3.00E-04	NA	3.00E-04
Predicted adaptive substitution rate	4.2E-11	2.1E-09	6.90E-10	3.45E-08	7.76E-10	3.58E-08

NOTE.—The prerescale retuned common weak sweep parameters show the parameters for a full-size *Drosophila* population.

Table 2
The observed and expected proportion of substitutions driven by positive selection are shown.

Model	Observed	Expected
Rare/strong	0.0123	0.0127
Common/weak	0.1742	0.1742
Retuned common/weak	0.1745	0.01796
Empirical	0.01884	

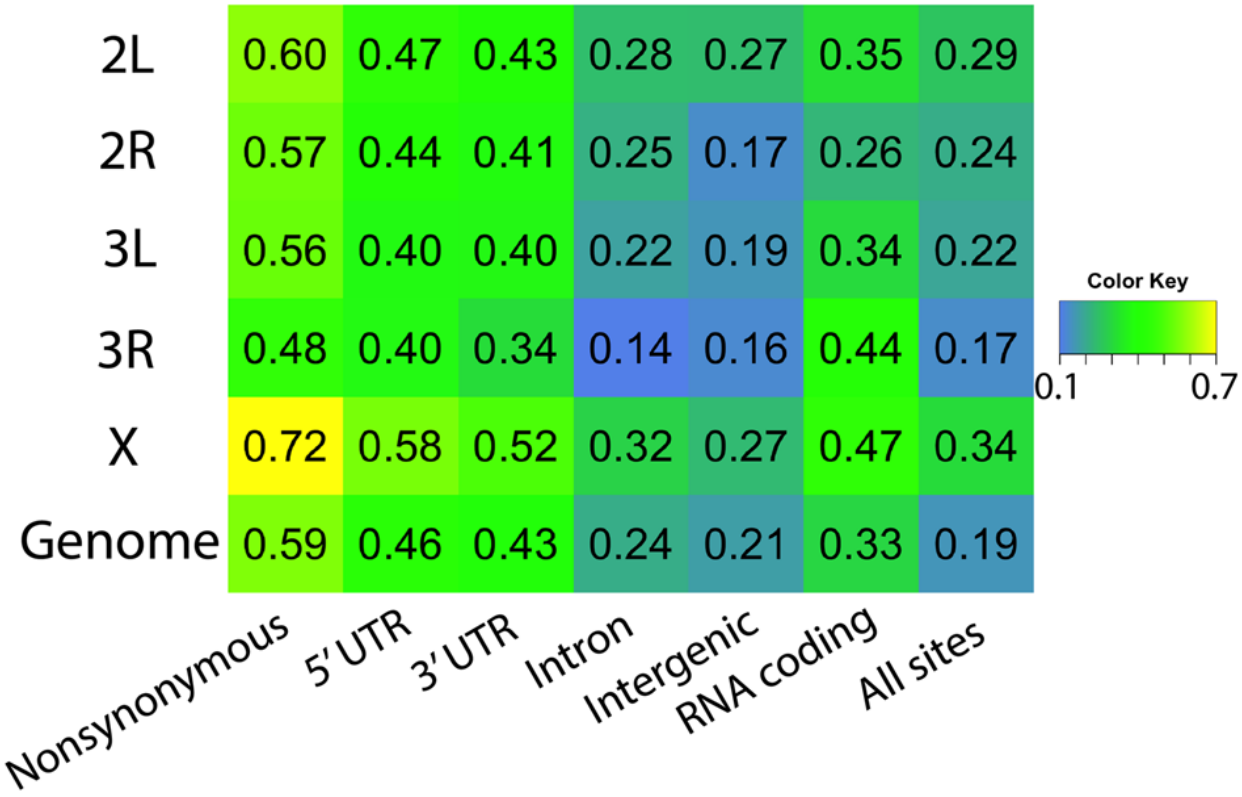


Figure 1. The estimated proportion of substitutions driven by positive selection (α), as estimated from *Drosophila* genomic data, is shown for each chromosome arm and site functional class, as well as the genome-wide average across all arms.

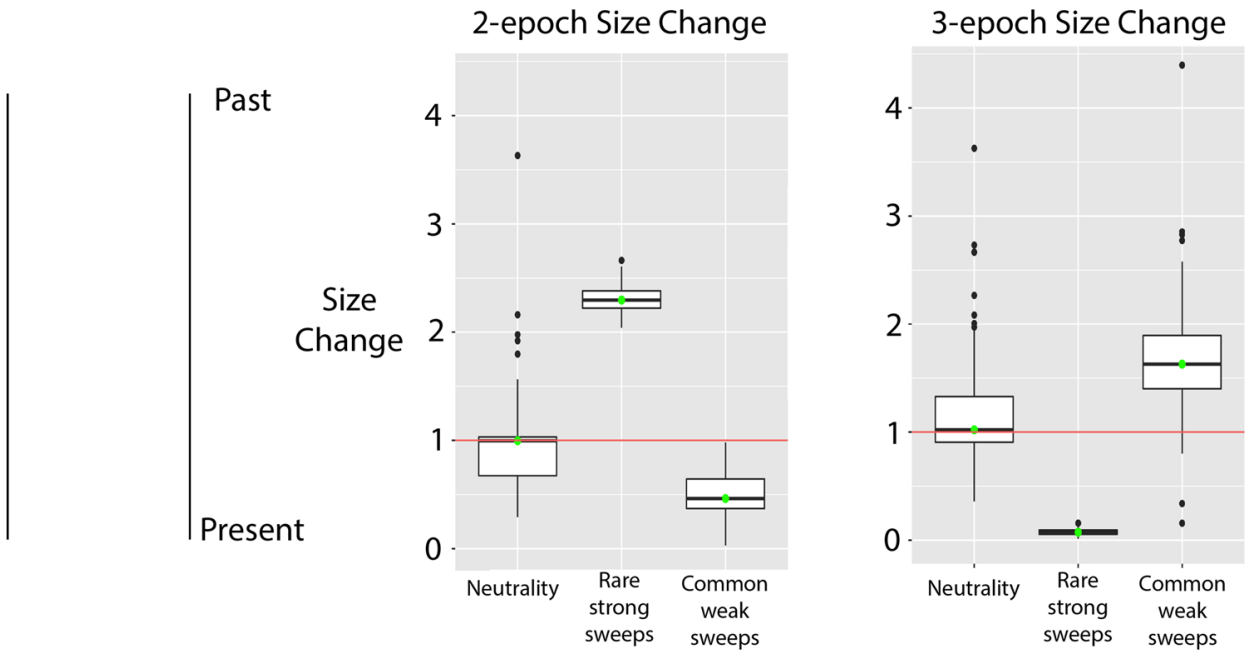


Figure 2. Falsely inferred population size changes based on RHH simulations are illustrated. Population size change estimates are shown for the trunk simulations (no true size changes). In the two-epoch scenario, we allow δαδι to infer a single population size change. In the three-epoch scenario, we allow δαδι to infer a size change and require it to return to its original effective population size.

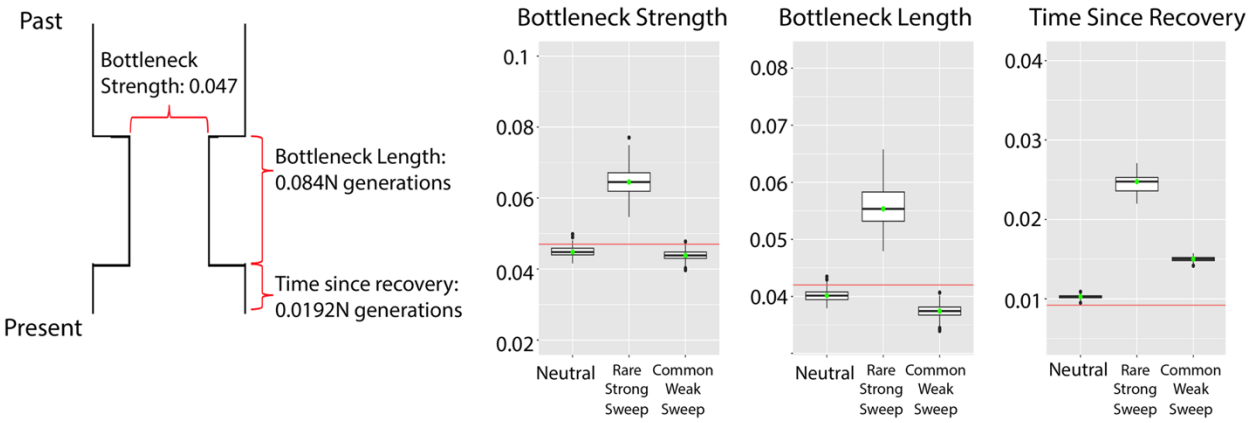


Figure 3. Demographic parameter estimates are shown for bottleneck simulations based on the true model shown on the left. Simulations with RHH showed relatively little bias for the duration and time since the bottleneck, but moderate bias for the time since recovery.

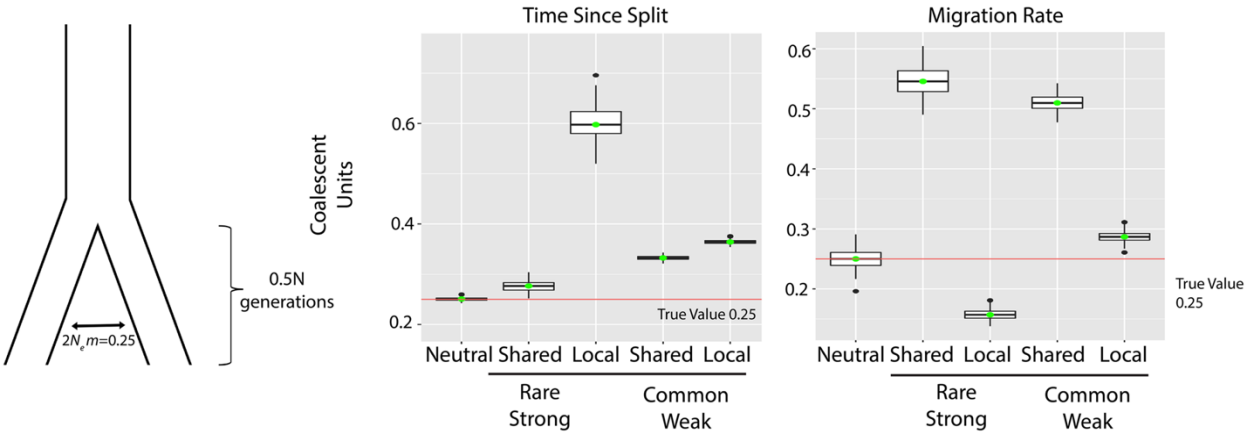


Figure 4. Demographic parameter estimates for simulations of the depicted isolation-migration model are shown. Some bias was observed for this relatively ancient population split time even under neutrality. However, differences from neutral estimates for RHH cases with shared sweeps or local sweeps were consistent with the effects of decreased or increased genetic differentiation, respectively.

Chapter 3: Population Genomics of Short-Term Evolution in a Population of *Drosophila melanogaster*

Abstract

A central goal in population genomics is to understand the relative contributions of neutral and non-neutral forces in shaping genetic variation. Typical genome-wide analyses of genetic variation investigate such forces across a wide interval of evolutionary time by examining a single snapshot of genetic variation. Adaptation, however, can occur over much shorter ecological timescales in natural populations. Utilizing a second sampling time to study temporal changes in allele frequencies can help clarify the relative roles of neutral and non-neutral forces on very short time scales. The research described here compares genomic variation of a recently collected sample of *Drosophila melanogaster* to samples of a collection made approximately 35 years ago. Using whole-genome sequencing, this study gives us an unusually direct way of quantifying change in genomic variation across decades, allowing us to ask critically important questions about how evolution works in nature in a populous insect species. The allele frequency changes between the time points suggests a small local effective population size on the order of <10,000, significantly smaller than the global effective population size of the species. Using the population branch statistic (PBS), we reveal targets of the genome that appear to have undergone very recent directional selection. The top genome-wide PBS outlier includes a locus that has previously been associated with resistance to pyrethroid insecticides. Upon closer examination, we find that structural variation is likely driving this change. We also find evidence for a shift in the latitudinal allele frequency cline along the North American east coast. Here, southern-associated alleles have decreased by an average of nearly 2.5% compared to a genome-

wide average of 0%. This project underscores the importance of utilizing multiple sampling time points to examine how genetic variation changes over ecologically relevant timescales.

Introduction

A central goal in population genetics is to understand the relative contributions of neutral and non-neutral forces on genetic variation. Typical genome-wide analyses of genetic variation examine these forces across a wide interval of evolutionary time by examining a single snapshot of genetic variation, reflecting events from roughly the last $4N_e$ generations (where N_e is the effective population). Adaptation, however, can occur over much shorter ecological timescales in natural populations (Daborn *et al.* 2001, Colosimo *et al.* 2005, Hoekstra *et al.* 2006, Campbell-Staton *et al.* 2017, Pélissié *et al.* 2018). While researchers have long sought to understand short term adaptation, decreasing sequencing costs in recent years have sparked a renewed interest in short-term adaptation studies.

Utilizing multiple sampling times to study temporal changes in allele frequencies can clarify the relative roles of neutral and nonneutral forces on very short time scales. This technique helps minimize neutral genetic differences when comparing genomic variation between time points. Evolve and resequence (E&R) studies use a multiple sampling time technique by evolving laboratory populations in controlled environments and observing changes in genetic variation over perhaps dozens of generations. E&R studies have been conducted on multiple species including *E. coli* (Barrick *et al.* 2009), influenza (Foll *et al.* 2014), *S. cerevisiae* (Parts *et al.* 2011), and *D. melanogaster* (Turner *et al.* 2011) to uncover important functions of natural selection. One disadvantage of E&R studies is that they may not always reveal how selection works in nature because they are typically derived from laboratory populations. Laboratory populations generally contain only a subset of natural genetic diversity, and are kept

in environments free of natural enemies, with at most a single environmental stress (potentially minimizing the pleiotropic consequences of laboratory adaptation).

There exists a multitude of temporal population genomic studies in humans (Burger *et al*. 2007, Mathieson *et al*. 2015, Hofmanová *et al.* 2016) and other mammals (Noonan *et al*. 2005, Lindqvist *et al*. 2010, Castañeda-Rico *et al*. 2020). Organisms with shorter generation times, such as *Drosophila melanogaster*, allow the study of substantially more evolution over decadal time scales. To that end, several studies have applied temporal sampling to natural populations of organisms with shorter generation times. A study by Bergland *et al.* (2014) found evidence for dozens of genomic loci showing seasonal allele frequency changes in *D. melanogaster* that could not be explained by drift alone. In 2015, researchers sequenced museum specimens of the North American honey bee to study short-term genomic changes in response to the introduction of a parasitic mite (Mikheyev *et al*. 2015). More recently, Feder *et al.* (2016) used multiple sampling time points in humans to examine the roles of hard and soft selective sweeps in HIV over short time scales. Additionally, Chen *et al.* (2019) examined a natural pedigreed population of 3984 Florida scrub jays over a period of 24 years. The authors discovered several SNPs under directional selection during this brief interval, suggesting the importance of rapid adaptation in natural populations.

To date, no studies have examined change in genetic variation across multiple decades in *Drosophila melanogaster*. The research described here compares genomic variation of a recently collected sample of *Drosophila melanogaster* to samples from the same locality collected approximately 35 years ago. This study gives us an unusually direct way of quantifying change in genomic variation across decades, allowing us to ask critically important questions about how evolution works in nature in a populous insect species. Because we can focus on a much

narrower time scale than typical population genomic analyses, we can begin to examine how much selection has occurred over the last ~35 years. We also investigate regions of the genome that may have undergone very recent selection, which could inform more precise investigations into genes associated with, for example, insecticide resistance or climate adaptation.

Results and Discussion

Whole genomic sequences were extracted from 64 wild-derived isofemale strains originally collected from Providence, Rhode Island between 1975 and 1983. Mean sequencing depth per individual genome ranged from 5.26X to 71.79X. We returned to Providence, Rhode Island in Fall 2014 and Spring 2015 to collect seasonal samples. 247 flies divided into 6 pools were sequenced in the fall sample and 408 flies divided into 18 pools were sampled in the spring sample. Depth in these pools ranged from 13X to 134X.

Ancestry is Consistent Across Time Points

African and European admixture is a well-studied phenomenon in North American *Drosophila melanogaster* populations. This admixture may have resulted from secondary contact between African migrants introduced to the Caribbean or neighboring regions and European migrants potentially introduced to the northeast US (Keller 2007), generating an ancestry cline along the east coast of North America (Bergland *et al.* 2014; Kao *et al.* 2015). Such admixture can have evolutionary consequences by, for example, introducing novel adaptive variants (Racimo *et al.* 2015) or generating epistatic interactions between alleles of different ancestry (Pool 2015). It is therefore important to estimate the extent of admixture in this population at the two time points. A significant change in ancestry over the 35-year time period could signal a shift in the east coast ancestry cline. A change in admixture could also be an indicative of

migration into the Providence population in the last 35 years. Both would affect downstream interpretation of results, so careful examination was warranted.

Overall, we estimated a genome-wide average of 13.85% African ancestry in the old genomes and a weighted average of 12.91% in the 18 seasonal pools (Table 1, Figure 1). We found little variability from one genome to the next in the old samples, with a standard deviation of 2.3% on chromosome 2, 3.98% on chromosome 3, and 4.04% on chromosome X. This variability from genome to genome is slightly lower than previous estimates from inversion-free DGRP chromosomes (Pool 2015), though when accounting for slightly higher African ancestry in DGRP, the individual variability is quite similar.

We can complement this genome-wide examination of the ancestry cline by looking for SNP frequency changes at highly clinal outlier SNPs that may reflect the action of spatially-varying selection. Using the dataset provided from Machado *et al.* (2016), we examined the temporal dynamics of all SNPs with a clinal P value below p=0.001. By cross-referencing these SNPs with the dataset provided in Bergland (2014), we were able to identify "southern-associated" and "northern-associated" alleles and ask whether there was an appreciable change between the two time points. To help reduce the role of linkage between sites, we further pared down SNPs so that no two SNPs were within 10kb of one another. This left 1671 examined SNPs. Notably, the southern-associated alleles decreased in frequency by an average of 2.44%. Of the 1671 clinal SNPs examined, 62% of (1036) southern-associated alleles decreased in frequency (binomial $P < 0.00001$ assuming a 50% null expectation). This is in contrast to a random sample of non-clinal SNPs showing an average shift of -0.11%, in line with an expectation of no shift between time points (Figure 2, Table 2). Since we did not find any genome-wide evidence of a strong shift toward European ancestry, demographic explanations for

this shift appear unlikely. The drastic, nearly two-to-one ratio of frequency decreases of southern-associated alleles suggest a non-neutral explanation at these clinal SNPs. We hypothesize that the ancestry cline has solidified in the last 35 years due to continued local adaptation favoring alleles of European ancestry in this relatively cool environment.

Inversion Analysis

Because chromosomal inversions suppress recombination in a heterozygous state (Sturtevant and Beadle 1936), they have long been believed to play an important role in the evolutionary process. Indeed, research has established that inversions can play a role in natural selection (Kirkpatrick and Barton 2006; Kapun *et al.* 2016) and speciation (Noor *et al.* 2001). We therefore wanted to measure common inversion frequencies in both the old and new samples.

Inversion frequencies at all time points are shown in Figure 3A. Of note, we only observed inversions in their heterozygous form in the old samples, and none of the four inversions on chromosome 3R were observed in the same strain. This statistically unlikely observation may point to tradeoffs for an individual organism bearing an inversion and, therefore, frequencies at the population level may be held in balance. *In(2L)t* increased in frequency between the time points while the other tested inversions all appeared to decrease. Our inversion frequency estimates for the seven inversions tested in the old samples are highly correlated with what was reported by Mettler *et al.* (1977) in a Portland, Maine population in 1977 (Spearman's =0.691), though our frequency estimates were considerably higher (average frequency 0.069 vs 0.023). Inversion frequencies may have shifted within the isofemale lines since the original collection, potentially because they help maintain heterozygosity and reduce inbreeding efforts via associative overdominance.

We also asked whether inversions varied seasonally between our fall and spring samples. Here, we simulated sampling based on empirical SNP frequencies and empirical depth of coverage at inversion-associated SNPs and asked how often we observed inversion frequency differences between seasons as large as what we observed in the real data (see Materials and Methods). We found that sampling variation could not explain seasonal inversion frequency differences for three inversions: *Inv(2R)Ns, Inv(3R)Mo, and Inv(3R)Payne*. The observed frequencies of *Inv(2R)Ns* in fall and spring seasonal samples was 0.068 and 0.029, a shift that our resampling simulation indicated was extremely unlikely to be due to sampling variance ($p=0.0033$). For *Inv(3R)Mo*, we measured seasonal inversion frequencies of 0.055 (fall) and 0.0895 (spring), for an associated P value of 0.0301. Finally, we measured the frequency of *Inv(3R)Payne* at 0.0394 (fall) and 0.0887 (spring), with an associated P value of 0.0016. The frequency differences we observed in other common inversions *(Inv(2L)t, Inv(3L)P, Inv(3R)C,* and *Inv(3R)K)* could all be explained by sampling variance. Our results are mostly in line with Kapun *et al.* (2016), who described significant seasonal differences in *Inv(2R)Ns* and *Inv(3R)Payne,* but not *Inv(3R)Mo.*

Although we appear to have observed real frequency changes at these three inversions, it is unclear why. One hypothesis is that there was drift caused by seasonal population boom and busts. Under such a model, one could expect inversion frequency differences between sampling locations in the spring sample, since census population size should be lower in the spring than in the fall. We thus asked whether inversions differed between sampling locations within season (as described in fly collection methods). Here, we employed a similar simulation sampling strategy as the seasonal inversion analysis. We asked how often we observed simulated inversion frequency differences between sampling locations as large as what we observed empirically (see

Materials and Methods). We did not find evidence of inversion frequency differences between sampling locations in the fall sample; all differences could be explained by sampling variance. This was not the case for the spring data (Figure 3B). All three inversions where we found significant seasonal differences (*Inv(2R)Ns, Inv(3R)Mo,* and *Inv(3R)Payne),* also showed significant frequency differences between sampling locations. Additionally, *Inv(3R)C* also showed significant differences between sampling locations, although its seasonal difference could be explained by sampling variation (p=0.703). Therefore, although some inversions could have seasonally-variable fitness consequences, we cannot rule out genetic drift (due to low population sizes coming out of a cold winter season) as a primary driver of the seasonal frequency differences we observe in some common inversions.

Estimated Local Population Size is Relatively Small

The long-term effective population size of *D. melanogaster* has been shown to be on the order of 1,500,000 to 2,500,000 (Duchen 2013; Sprengelmeyer *et al.* 2020), but one study suggested a much larger recent effective population size on the order of 10^8 (Karasov *et al.* 2010). These population-scale sizes are likely much larger than local deme sizes, especially in a temperate environment with seasonal population size fluctuations. Smaller local census sizes have been estimated on the order of 1,000-10,000 (McInnis *et al.* 1982). It is important to have an accurate estimate of local population size to identify regions of recent directional selection that deviate from the predictions of genetic drift. To estimate local population size, we simulated allele frequency trajectories of SNPs based on a simple Wright-Fisher model and fit the distribution of observed genome-wide frequency changes to distributions of simulated frequency changes corresponding to differing population sizes. Here we assumed that the bulk of the empirical distribution reflected neutral evolution over this short time period. We found that a

population size of 9,500 individuals best recapitulated the empirical distribution of SNP frequency differences between the time points. Here, the empirical standard deviation was 0.0781 and the simulated standard deviation, with a population size of 9,500, was 0.0782.

Although the amount of genetic variation observed in *D. melanogaster* reveals that the long-term effective population size of the species is very large, our results suggest that not necessarily all of this genetic variation is observed within a single local population. If natural selection is sufficiently widespread in the genome on short time-scales, it could bias our local N_e estimate downward. Such a bias would lead us to overestimate the strength of genetic drift, which is conservative with regard to identifying potential targets of recent natural selection.

Potential Adaptive Differences Between Sampling Points

It has been shown that dozens of loci in the *D. melanogaster* genome exhibit evidence of seasonally-varying selection, with one allele becoming more common by spring and the alternate allele rising in frequency by fall (Bergland *et al.* 2014). We therefore wanted to separate seasonal allele frequency change from directional evolution across the ~35 years. The latter signal can be isolated by the Population Branch Statistic (PBS) (Yi *et al.* 2010), which uses three population samples to quantify genetic differentiation on one population's lineage. We applied PBS to ~5 kilobase windows and individual polymorphic sites along the genome (Figure 4A). We ran a Gene Ontology (GO) enrichment analysis on PBS outliers (the top 1% of windows) to identify functional categories that may hold adaptive differences between the sampling times (Figure 4B). One of the top categories in this analysis was response to insecticide, which we discuss in more detail below. Three other categories were related to the nervous system. Nervous system GO categories have been enriched in other genomic scans for positive selection in *D.*

melanogaster (Langley *et al.* 2012, Pool *et al.* 2012, Pool 2015), with one synaptic gene showing evidence for selection to cold tolerance adaptation (Pool *et al.* 2017).

We used simulation to assign a P value to every empirical window. Empirical windows were divided into five bins, and 2.5 million simulations were run for each bin. Briefly, we simulated the species demography from Sprengelmeyer *et al.* (2020). This demography consists of 9 populations sampled throughout Africa and Europe. To include the North American Rhode Island population, we added a 10th population consisting of an admixed population of Cameroon and France migrants. The simulated admixture proportions reflected estimated admixture from our historic samples.

As a null hypothesis, we expected our neutral simulations to recapitulate our empirical distributions of window PBS and, therefore, we anticipated a uniform distribution of P values. Instead, we saw a slight enrichment of low P values for the window PBS statistic (Figure 5), suggesting the role of natural selection during this time period. Below, we briefly discuss some outlier window regions of interest.

Insecticide Resistance as a Likely Target of Selection

In Table 3, we display our top outlier regions. Our strongest genome-wide outlier region had a window-PBS value of 0.278 and spanned 78 windows. The P value associated with the top window in this region was below a Bonferroni-corrected genome-wide significance threshold (0.05 divided by the total number of windows). Notably, this window contained a pair of cytochrome p450 genes, *Cyp6a17* and *Cyp6a23*. Previous research has identified chimeric alleles of these genes (in which a derived deletion affecting part of each of these paralogs reduces them to a single gene) segregating at high frequency in the DGRP population (Good *et al.* 2014). Later research showed that disruption of *Cyp6a17* confers resistance to the deltamethrin class of

insecticides in *D. melanogaster* (Battlay *et al.* 2018). Given the clear signal of genetic differentiation we observed at this locus, we estimated the frequency at which this gene was intact in both our old and new samples. In the old sample, we observed an intact *Cyp6a17* at 23.53% frequency.

We also examined the frequency of the intact *Cyp6a17* in each of the sampling years, which revealed a striking result. We found that the frequency of the intact *Cyp6a17* increased over the 8-year period in which the older population sample was collected, rising from 0% to >50% (Figure 6A), possibly revealing a selective sweep in action. By examining coverage at *Cyp6a17* in our seasonal samples, we estimated that *Cyp6a17* is intact at around 50% frequency in modern Providence, Rhode Island populations. Thus, after its rapid rise, it appears that the frequency of an intact *Cyp6a17* did not meaningfully change between 1983 and 2014. These patterns might reflect changes in insecticide usage with time, or potentially represent the impact of balancing selection due to fitness costs associated with insecticide resistance, a well-studied phenomenon in insects (Kliot and Ghanim 2012).

We performed approximate Bayesian computation to investigate the strength of natural selection required to observe the empirical frequency shift over 8 years. We simulated a simple Wright-Fisher model over an 8-year period, assuming 15 generations per year (Turelli and Hoffmann 1995; Pool 2015), and sampled to match our empirical counts. We varied the starting allele frequency and the selection strength and rejected any simulation that did not exactly match empirical counts of the sampled allele. This analysis revealed that strong selection (around 5%), as well as an appreciable starting frequency (also around 5%), is required to explain the observed results (Figure 6B). A relatively high initial frequency of this beneficial allele is perhaps not

surprising since it appears to reflect the ancestral arrangement (Good *et al.* 2014) as opposed to a derived variant entering the population at very low frequency.

Our third highest outlier region contained the well-known insecticide resistance gene *Cyp6g1*, which has been shown to confer resistance to Dichlorodiphenyltrichloroethane (DDT). This region spanned 42 windows, and the top window had a PBS value of 0.102. The P value assigned to this window (p=3.64E-5) was just above the Bonferroni-corrected significance threshold. This window was not directly over *Cyp6g1* – it was about 27 kb downstream – but the closely related gene *Cyp6g2* was less than 10 kb from the top window. *Cyp6g2* has also been shown to confer insecticide resistance (Daborn *et al.* 2007). *Cyp6g1* confers resistance to DDT via the insertion of a transposon upstream of the transcription start site (Daborn *et al.* 2002; Chung 2007). It has also been shown that ongoing selection at this locus is caused by a duplication and additional transposable element insertions at this locus (Schmidt *et al.* 2010).

A third well-known locus that confers resistance to insecticides was also amongst our top 20 regions genome-wide. This region spanned 10 windows, and the top window included *Cyp12d1-p* and *Cyp12d1-d*. The PBS value at this window was 0.0677 and its P value was 0.00096. It has been shown that overexpression of *Cyp12d1* increases resistance to DDT (Daborn *et al.* 2007). Structural variation at this locus has been shown to confer resistance to xenobiotics including caffeine (Najarro *et al.* 2015), but this copy number variation is not associated with resistance to DDT (Schmidt *et al.* 2017). When mapped reads were viewed with IGV, few to no reads mapped uniquely to the *Cyp12d1-p* and *Cyp12d1-d* locus at either time point, potentially due to high sequence identity between them. Instead, we observed high PBS SNPs flanking the locus (Figure 7A). These high PBS SNPs may be in linkage disequilibrium (LD) with copy number variation or unmapped SNP variation that differs in frequency between time points.

Another well-known insecticide resistance gene, *Acetylcholine esterase (Ace),* offers resistance to organophosphates and carbamate insecticides (Mutero *et al.* 1994). Windows associated with this gene had a PBS value in the top 1% genome-wide. Known insecticide resistance mutations in this gene have been previously reported (Mutero *et al.* 1994, Menozzi *et al.* 2004, Karasov *et al.* 2010). A set of three mutations within twenty base pairs of each other make up a resistant haplotype. All three of these resistance alleles segregated at low frequency in the old samples (F330Y at 0%, G265A at 2.08%, and I161 at 6.11%) and segregated between 23% and 42% in the new, seasonal samples. Over this relatively short time period, these three mutations that confer insecticide resistance increased by an average of 28.33%.

Our data suggests that insecticides have been a major driver of evolution in the *D. melanogaster* genome over the last 35 years. Interestingly, all four insecticide loci described here have recently been identified as candidate regions of recent adaptive introgression into African *D. melanogaster* populations (Svedberg *et al.* 2020), underscoring the likely importance of insecticide evolution in admixed populations. Additionally, the temporal nature of our data reveals important information about the trajectory of an allele and its underlying selection coefficient, allowing researchers to more accurately model the genetic basis of adaptation in human-associated insect species including crop pests and invasive species.

Male Fertility as a Target of Natural Selection

It appears that male fertility may have been an important evolutionary target between our time points. At least three genes in our top 20 PBS outliers may play important roles in fertilization. Our second highest PBS outlier region was on chromosome X and covered 26 windows. The top window in this region had a window PBS of 0.144 and a P value of 1.2E-6, which was below the Bonferroni-corrected significance threshold. This window contained two

pseudogenes as well as *Hexosaminidase 2 (Hexo2)*. Zooming in to the SNP level of this region, we observed a collection of high PBS SNPs as well as a modest PBS peak in the intergenic regions between *hexo2* and the pseudogenes (Figure 7B). The gene product of *hexo2* is found in the plasma membrane of sperm in *D. melanogaster* (Cattaneo *et al.* 2006) and has a possible role in fertilization and sperm-egg interactions (Intra *et al.* 2017).

Two other genes in our top 20 regions have possible effects on reproduction as well. The first, *Darkener of apricot* (Doa), was within the top window of our fifth highest outlier region. The PBS at this window was 0.101 and had a P value of 4.68E-5. *Doa* spanned four windows, though the PBS signal appeared to localize toward the end of the gene (Figure 7C). Researchers have used artificial selection experiments to show that this gene affects aggression in flies (Edwards *et al.* 2006, Zwarts *et al.* 2011). Later studies have demonstrated that mutations at *Doa* disrupt sex-specific splicing of *doublesex* pre-mRNA, resulting in the feminization of male cuticular hydrocarbon profiles and the masculinization of female cuticular hydrocarbon profiles, along with disruption of associated courtship behavior (Fumey and Thomas 2017).

A third gene that affects fertility was also within our top outlier regions. Growth arrest specific protein 8 (gas8) spanned the top two windows in our fifth highest outlier region. This gene has been associated with sperm fertility in mice (Yeh *et al.* 2002). In *D. melanogaster*, knockdown of this gene causes infertility in males (Zur Lage *et al.* 2019).

Other Targets of Interest

A region on chromosome X was our 7th highest PBS outlier (PBS=0.0849, P value 0.000156). Zooming in on the window revealed a SNP pattern that localized over two genes, *CG4991* and *CG16700* (Figure 7D). This region has previously been identified as a target of

positive selection between African and non-African populations (Svetec *et al.* 2011), likely due to cold tolerance adaptations (Ayroles *et al.* 2009, Wilches *et al.* 2014).

A Complementary Scan to Confirm Window Outliers and Identify SNP-level Outliers

Because we were comparing two distinct types of sequencing data (individual sequences and pool sequences), we wanted to complement our original scan with a scan where all datasets went through the same pipeline. We down-sampled raw reads from our 64 old lines so that they all had the same number of reads. We then combined the reads to emulate a pooled set of sequences and performed a PBS scan with this "pseudo-pool" set. The PBS outliers at the window level were largely consistent across the two scans. All outliers in Table 3 were also outliers in the "pseudo-pool" scan.

Especially in the case of soft sweeps, it is possible that some positive selection signals spanned less than a full window and were missed by the window scan described above. However, when we examined top outliers for maximum SNP PBS (the highest SNP PBS value within a given window) from our primary scan that were not also window PBS outliers, we found that these SNP-specific outliers were not well-replicated in our pseudo-pool analysis (results not shown). Because these outliers could reflect artifacts driven by differences in data processing between pool and individual genomes, we did not examine them individually. Instead, we focused our SNP-oriented analysis solely on the psuedo-pool scan outliers for maximum SNP PBS.

In Table 4, we present 14 windows where the maximum SNP PBS value was in the top 1% genome-wide in our initial scan, and also in the top 2.5% genome-wide in the complementary pseudo-pool scan, while the window PBS value was a non-outlier (in the *bottom* 95% genome-wide) in the original scan. The top SNP PBS signal in this group was from the gene

heavyweight (hwt). Variation at this gene is associated with body mass among DGRP strains (Nelson *et al.* 2016). Both *Neurospecific receptor kinase (nrk*; associated with the second-highest SNP PBS signal) and Dystrophin (*Dys*) also contained a high SNP PBS but lacked a window signal. These two genes interact genetically to control neuron behavior in the eye (Marrone *et al.* 2011).

Conclusions

In this study, we compared genomic variation of a recently collected sample of *Drosophila melanogaster* to samples of a collection made approximately 35 years ago. This unique dataset has given us insight into local effective population sizes, and has allowed us to identify recent targets of adaptation. Namely, we have shown evidence of possible adaptation at four separate insecticide resistance loci. We have also observed a shift toward northern alleles at well-studied clinal SNPs, possibly due to ongoing local adaptation favoring alleles of European ancestry in this relatively cool environment. These results may merit further study in other time-sampled *D. melanogaster* natural populations. Mapping resistance to insecticide in other populations around the world has important real-world applications in agriculture, and overlapping targets of adaptation could inform researchers about the genetic predictability of adaptation.

Materials and Methods

Fly Collection

Flies were collected from 5 traps in Providence, Rhode Island in Fall 2014 and Spring 2015. In the fall samples, 41 flies were collected from each of the 5 traps. A sixth set of 42 flies (7 daughter progeny from 6 flies) was also sequenced. This totaled 6 pools from the fall sample. The spring samples consisted of 12 total pools, each with 34 flies.

Genomic Sequence Data Collection

A separate library was prepared for each of the 64 individual strains and the 18 seasonal pools. Mean sequencing depth per individual genome ranged from 5.26X to 71.79X. Reads were aligned as described in Lack *et al.* (2015), except with a single round of mapping to the *D. melanogaster* (v5.57) reference genome instead of a second round of mapping to a genome-specific reference genome. We chose a single round of mapping to more closely align with the single round of mapping that the pooled sequences necessarily will go through. Briefly, reads were aligned using BWA aln v0.5.9 (Li and Durbin 2010), and unaligned reads were then mapped with Stampy v1.0.20 (Lunter and Goodson 2011). All reads with a mapping quality score below 20 were discarded. We then used Picard version v1.79 (http://picard.sourceforge.net/) to sort the alignment by coordinates and remove optical duplicates. Assemblies were improved around InDels using the GATK v3.2 InDel Realigner (Depristo *et al.* 2010, McKenna *et al.* 2010;).

To generate the "pseudo-pool" dataset, we down-sampled each of the 64 bam files from the individual sequences dataset so that each strain had an equal number of aligned reads. We did the downsampling with the command line "samtools view -s fraction -b data.bam > downsampled_data.bam" where "fraction" was the proportion of aligned reads kept. We then merged all 64 downsampled bam files using samtools and generated a pileup and subsequent sync file as we did with our other pool sequences.

Mean sequencing depth for the pools are provided in Table 5. Reads were aligned using the same pipeline as the individual genomes up through InDel realignment. Following InDel realignment, pileup files were then generated using Samtools v1.3.1, and sync files were

generated using PoPoolation2 v1.201 (Kofler *et al.* 2011), requiring a minimum quality score of 20.

Quality Assurance Checks (IBD, PCA)

Relatedness between sampled individuals violates assumptions of many population genomic models. Such relatedness could result from initial sampling of related individuals or from mishandling or mislabeling of lab samples in the last 35 years. To identify such instances of identity by descent (IBD), all pairwise comparisons between old samples were made. Chromosomes were compared in 500 kb windows sliding in 100 kb increments. Any window with fewer than 0.0005 pairwise differences per site was considered putatively IBD. Some genomic intervals (e.g., centromeric regions) exhibit large scale IBD across populations, suggesting explanations other than relatedness. These regions were excluded from this analysis, unless they extended outside these regions. We identified instances of "relatedness IBD" between two genomes when genome-wide IBD tracts exceeded 5 Mb. Such instances were masked to 'N' for one of the two genomes. One pair of genomes had IBD tracts covering the entire genome, suggesting that these lines were duplicated at some point in the last 35 years. One of these lines was discarded from our analysis, leaving 64 samples from our older time point.

We applied principal components analysis (PCA) to each major chromosome arm of the old samples to identify any strain with aberrant divergence. We also included several African populations (10 Cameroon strains, 10 Gabon, 6 Nigeria, 5 Guinea), 98 France strains, and two North American populations (19 from Ithaca, New York, 131 from Raleigh, North Carolina). Putatively heterozygous regions in all strains were masked for this PCA analysis. See methods of Lack *et al.* (2015) for masking heterozygosity in the non-Rhode Island genomes and see methods in this paper for masking of the Rhode Island genomes. We used SNPRelate release 3.9 (Zheng

et al. 2012) for the PCA analysis. Overall, we did not find evidence for aberrant divergence in our old samples (Figure 8).

Effective Pool Size

One of the major drawbacks of utilizing pooled sequencing is that we cannot assume that every read is a random draw from a population. Though allele frequency estimates are accurate in a large enough sample (Gautier *et al.* 2013), sample size estimates may not be due to library prep error. Because sample size is an important parameter when estimating Fst, we must take measures to obtain an unbiased sample size estimate. One approach is to explicitly estimate the number of *j* unique lineages at a site given n_r reads and n_c equally contributing chromosomes in a pool, such that $P(j|n_r, n_c) = \frac{n_c! S(n_r, j)}{(n_c - j)! n_c^{n_r}}$ (Ferretti *et al.* 2013) where $S(n_r, j)$ are the Stirling numbers of the second kind, defined as the number of ways to partition n_r reads into *j* nonempty sets. We used the R package GMP v0.5-13.6 to calculate these probabilities and extracted the expected number of unique chromosomal draws based off of these probabilities. Experimental error, especially the overrepresentation of one or more chromosomes, can bias $P(j|n_r, n_c)$ because n_c refers to the number of equally contributing chromosomes in a pool. We applied the method introduced in Gautier *et al.* (2013) to estimate the effective pool size, defined as the number of equally contributing diploid individuals in a pool. Our fall data contains 5 pools of 41 individuals from 5 distinct sampling locations and a 6th pool containing 42 flies that are the progeny of 6 flies from the 5 sampling locations. The effective pool size method estimated no error in the first 5 pools ($n_e = 41$ for all pools) and an $n_e = 10$ for the 6th pool. The 6th pool offered a good control of the method, as we expected this pool to have a lower n_e due to the relatedness of individuals in the pool. The method suggested a lower effective pool size number in the spring data. Here, each pool consisted of 34 flies. The effective pool size ranged from

$n_e = 23$ to $n_e = 34$ with a median of ne=31. Utilizing twice this effective pool size (to account for diploid individuals) as the number of equally contributing chromosomes in a pool, we could then estimate sample size at a site as the number of j unique lineages at a site given n_r reads and n_c equally contributing chromosomes in a pool.

Ancestry

We implemented a hidden Markov model (Corbett-Detig and Nielsen 2017) to estimate the proportion of African versus European ancestry. This method is general and can be used on individual genomes or on high ploidy data (i.e., pooled data). It utilizes short read pileup data to model ancestry across the genome as a function of sample allele frequencies within an admixed population. The maximum likelihood probability of each ancestry state at each panel SNP is output. For a diploid genome, for example, maximum likelihood states for homozygous European, heterozygous European and African, and homozygous African are output. This method also allows for variable ploidy across the genome to account for partially inbred chromosomes, allowing us to model inbred segments as a single haploid chromosome and outbred segments as a diploid chromosome individually in each old genome. We generated ploidy maps as described in the methods section. For the high ploidy datasets (pooled data), maximum likelihood probabilities are output for each ploidy + 1 states. Inversions can strongly affect ancestry due to their suppression of recombination in large chromosomal regions (Corbett-Detig and Hartl 2012). Because of this, results shown in Figure 1 of the old genomes are estimated from inversion-free chromosomes arms.

Heterozygosity

Residual heterozygosity often persists in fly stocks even after many generations of full-sibling mating. It is important to identify such regions of heterozygosity, since a putatively

heterozygous region constitutes two random allele draws from a population. We thus sought to identify such regions of heterozygosity in order to identify correct sample sizes at any given site in the genome. We applied a hidden Markov model (https://github.com/russcd/Heterozygosity_HMM) to annotate inbred and outbred regions in the partially inbred old samples.

Identification of Common Inversions

We implemented the method introduced in Kapun *et al.* (2014) to estimate inversion frequencies in each of the 18 pooled samples. For each studied inversion, this method required a set of fixed differences between inverted and non-inverted karyotypes. We used the inversion-specific markers given in table S4 in Kapun *et al.* (2014). To estimate the inversion frequency in an individual pool, we calculated the average frequency across all inversion-associated SNPs whose coverage exceeded a minimum of 10 reads. Seasonal inversion estimates were made by weighting pools by the estimated effective pool size, such that $freq_{seasonal} = \sum_{i=1}^{n_{pools}} f_{pool_i} * \frac{ploidy_{pool_i}}{ploidy_{total}}$. Here, n_{pools} is the total number of pools in a given season, f_{pool_i} is the average frequency across inversion-associated SNPs in the i^{th} pool, $ploidy_{pool_i}$ is the effective pool size in the i^{th} pool, and $ploidy_{total}$ is the sum of effective pool sizes across all pools.

To determine inversion status of the individual genomes in the old sample, we examined the diploid calls at each inversion-associated SNP described above. These diploid calls fell into two classes: most of the SNPs were homozygous for the non-inversion SNPs or most of the SNPs were heterozygous for the inversion SNPs. The former we classified as free of inversion and the latter we classified as heterozygous for the inversion. The inversion frequencies in the old genomes were calculated as the number of inversion heterozygotes divided by twice the total number of genomes in the dataset.

Frequency Differences in Common Inversions

There is some evidence to suggest that inversion frequencies differ seasonally in *Drosophila melanogaster*. Our dataset provided a unique opportunity to evaluate whether seasonal differences in inversion frequencies that we observed in our data could be explained by sampling variance. We simulated a sampling scheme that emulated the empirical data to determine how often we observed seasonal differences in some inversion *I* as large as we observed empirically. To accomplish this, we resampled inversion-associated SNP frequencies based on the empirical coverage in each pool. For each pool, we assumed the true inversion frequency, f_{inv}, is the midpoint of the seasonal point estimates. Using this assumed inversion frequency, we sampled the number of inversion-bearing chromosomes in a simulated pool as $chr_{inv} \sim binom(N_{eff}, f_{inv})$ where N_{eff} is the effective pool size (as described above) of the simulated pool. For each inversion-associated SNP *j* in the empirical dataset, we simulated inversion-bearing reads based on the empirical coverage in the pool. Thus, $reads_{inv} \sim binom(cov_j, \frac{chr_{inv}}{N_{eff}})$ where cov_j is the coverage observed empirically at the j^{th} SNP. Just as we had in the empirical data, we then had a vector of simulated inversion-associated SNP frequencies. The mean of this vector was the simulated estimated inversion frequency of the resampled pool for inversion *i*, and the resampled seasonal estimate was calculated by weighting each resampled pool by effective ploidy. The P values shown in Table 2 are defined as the proportion of resampled seasonal differences that were greater than what we observed empirically. Essentially, we asked whether seasonal differences in inversion frequencies could be explained by sampling variance. We also used this sampling scheme to determine whether sampling variance could explain differences in inversion estimates by sampling location. Here,

we noted the largest estimated inversion frequency difference and asked how often this difference was larger than what was observed empirically.

Effective Population Size Estimate

In order to draw conclusions about frequency changes over the ~35 year time period, we first needed to determine how much frequency change could be expected due to genetic drift. This expectation depends on the local size of the Rhode Island population, where a smaller population size would lead to a higher variance in the temporal frequency shift. To estimate this local population size, we simulated allele frequency trajectories of SNPs based on a simple Wright-Fisher model and fit the distribution of observed frequency changes to distributions of differing population sizes. The Wright-Fisher simulation emulated the empirical observations.

The site frequency spectrum of the old samples matched the simulated site frequency spectrum at the first time point. We then simulated 465 generations between the temporal samplings, corresponding to 15 generations per year (Pool 2015) for 31 years. The SNP frequency at each generation, $f_{i+1} = \frac{\nu_{i+1}}{2N_e}$, was drawn from a binomial distribution where $\nu_{i+1} \sim binom(2N_e, f_i)$ i+1 is the number of individuals in the next generation bearing the allele, N_e is the effective population size in the simulation, and f_i is the allele frequency in generation i.

The sampling at the latter two time points emulated the sampling observed in the real data. For each SNP, the number of chromosomes in pool j that bear the allele (ψ) is drawn from a binomial distribution. Here, $\psi \sim binom(2N_{eff_j}, f_i)$ where N_{eff_j} is the effective pool size of pool j. To further emulate the real data, coverage for all simulated pools is drawn from the empirical distribution of coverage. We then sampled reads bearing the allele, ϕ, based the

coverage and N_{eff_j} such that $\phi \sim binom(cov_j, \frac{\psi}{2N_{eff_j}})$. Simulated SNP frequency in pool j is then calculated as $\frac{\phi}{cov_j}$ and the overall seasonal frequency is weighted by ploidy.

Identification of Candidate Regions for Recent Directional Selection

The PBS statistic was used to quantify genetic differentiation specific to the newly collected seasonal samples when compared against the old population samples. Because differences in inversion frequencies between the sampling points can bias estimates of allele frequency differences, we weighted the old samples to match inversion frequencies of the pooled samples and corrected for sample size using Kish's effective sample size (Kish 1965). PBS was applied in diversity-scaled genomic windows containing 200 nonsingleton SNPs in the Zambia sample of the *Drosophila* Genome Nexus. Simulations were run for autosomal and X chromosome data based on the model from Sprengelmeyer *et al.* (2020) and adding a North American arm with admixture (both proportion and timing) based on empirical data. This North American arm was admixed from the France and Cameroon arms of the model. We ran 5 million ms simulations (Hudson 2002) for 9 recombination bins corresponding to 0-4 cM/Mb, increasing by 0.5 cM/Mb. We also simulated a 10^{th} bin for >4 cM/Mb. For each simulation, we randomly chose a locus length from the distribution of empirical windows. Recombination rate for each window was based on estimates from Comeron *et al.* (2012). The mutation rate for the autosomal model was 5.21e-9 and the X chromosome model was 5.07e-9 (Huang *et al.* 2016). Number of sampled chromosomes was also based on empirical data. To emulate pooled sequencing, we resampled simulated chromosomes based on coverage in the empirical data. With these simulations, we were able to assign a genome-wide P value for each window PBS by asking how many times simulated PBS values were greater than the observed PBS value. We assigned

significance to any P value below the Bonferroni corrected critical value of 0.05/(number of genome-wide windows).

Gene Ontology Enrichment

The top 1% of PBS quantiles were considered outliers for GO enrichment analysis under the hypothesis that these outliers will be enriched for genuine targets of adaptation. GO enrichment was assessed as previously described in Pool *et al.* (2012). Two or more outlier windows were merged into the same outlier window region if they were separated by no more than four nonoutlier windows (to conservatively avoid counting the same selective sweep more than once). Locations of outlier regions were then randomly permuted, while maintaining their lengths, to properly account for the arrangement and lengths of genes in each functional category. Each outlier region was only allowed to vote for a given GO category one time (from both the empirical and permuted outlier regions), to avoid spurious results from clusters of functionally linked paralogs. For each GO term, a raw P value was defined by the proportion of 1,000,000 randomized data sets in which a greater or equal number of outliers from that category was obtained. Then, by comparing across these randomized data sets, the lowest raw P value for each of them was obtained, and a threshold for analysis-wide significance was defined based on a minimum raw P value observed in 5% or fewer randomized data sets.

ABC Analysis

We used approximate Bayesian computation to estimate the strength of selection and the initial frequency of the intact *Cyp6a17* gene during the 8 year sampling period. We employed a very simple Wright-Fisher simulator and simulated 120 generations (15 generations over 8 years). For each simulation, we randomly selected a selection strength from a uniform distribution between 0 and 0.3 and an initial frequency from a uniform distribution between 0

and 0.2. We only accepted a simulation that exactly matched our empirical counts of an intact *Cyp6a17* gene. We stopped our simulation after 500,000 successes, which was around 100 billion simulations.

Acknowledgments

We are especially grateful to Rayla Temin for the gift of the original population sample of Providence lines. This work benefitted substantially from computational resources and support from the UW-Madison Center for High Throughput Computing (CHTC). It was funded by USDA Hatch grant WIS01900 and by NIH grant R35 GM136306.

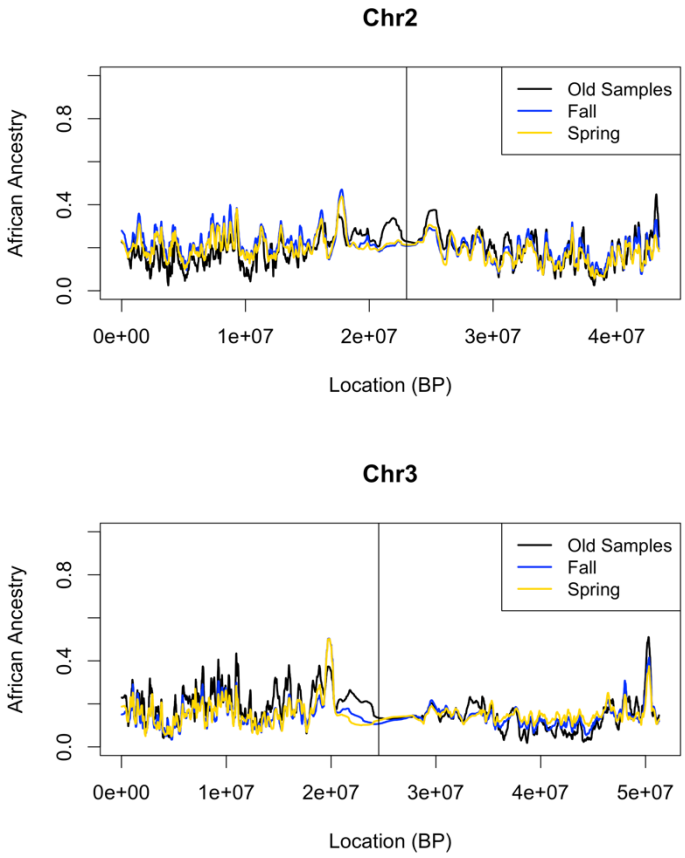


Figure 1. Ancestry appears to have not shifted between time points. African ancestry was averaged across inversion-free strains in the original samples (black line), averaged across 6 pools in the fall sample (blue), and 12 pools in the spring sample (yellow). Note the size difference of the X chromosome plot. For each of the autosomal plots, we are displaying both arms of the chromosome, which are about twice the length of the single arm of the X chromosome.

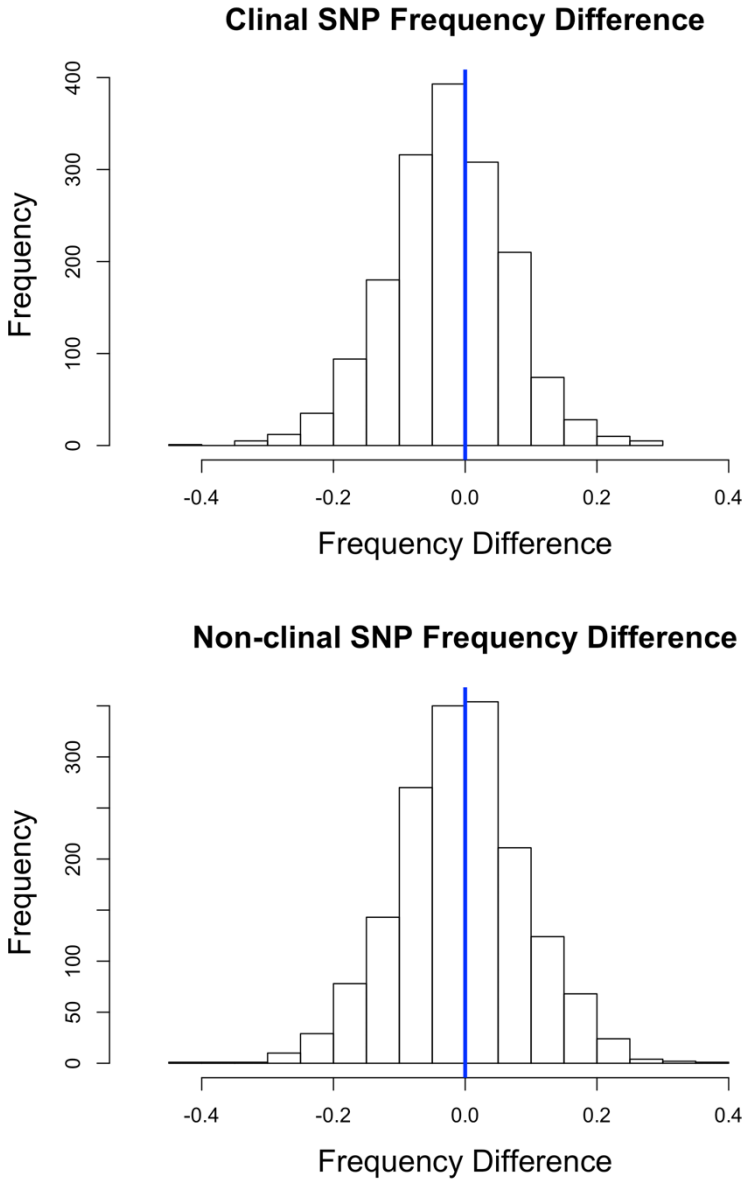


Figure 2. Southern-clinal associated alleles appeared to decrease in frequency over time. The top histogram depicts 1,671 differences of southern-associated clinal SNPs. alleles at bottom The depicts the histogram difference at frequency 1671 SNPs chosen at random across the genome.

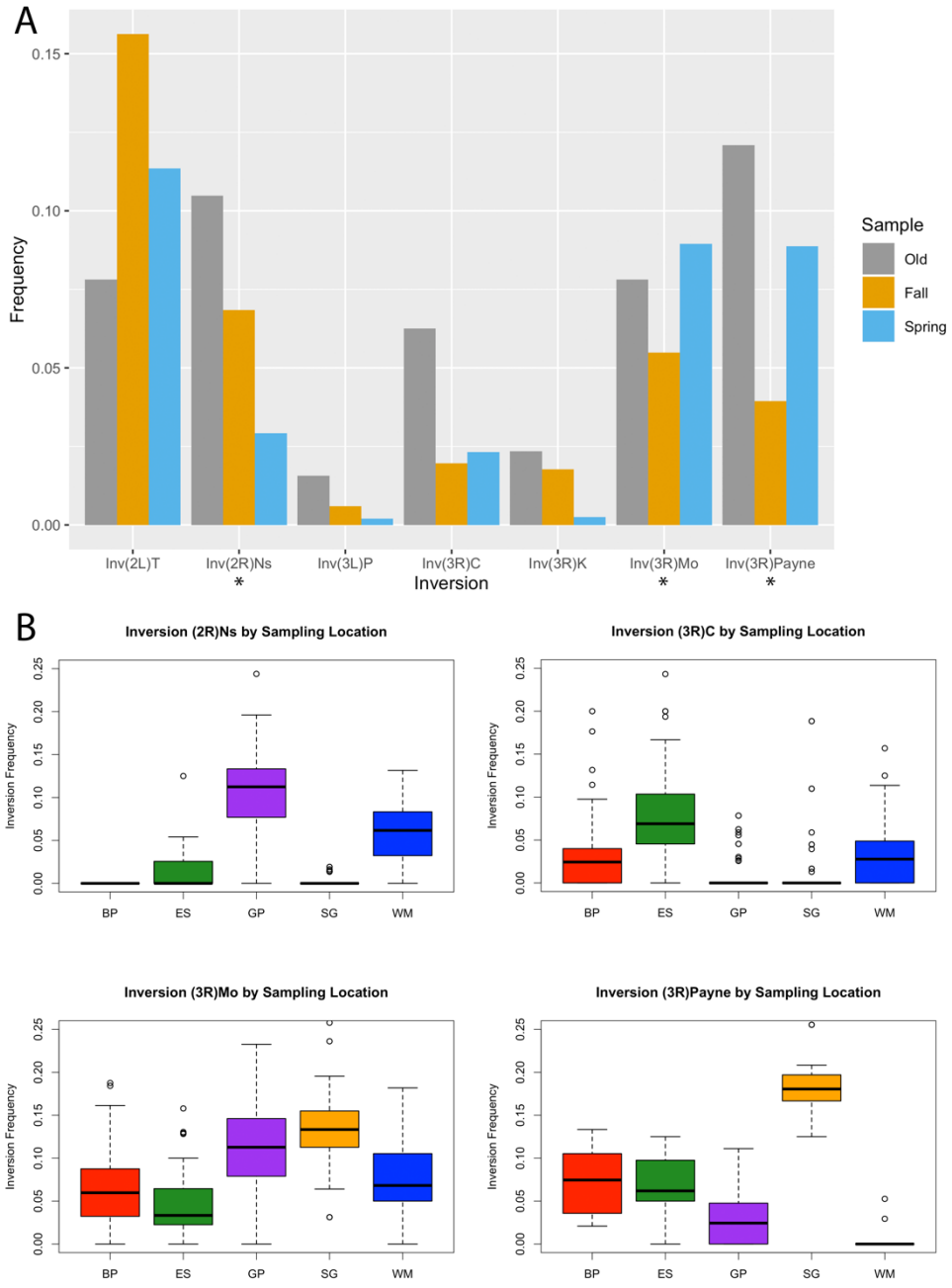
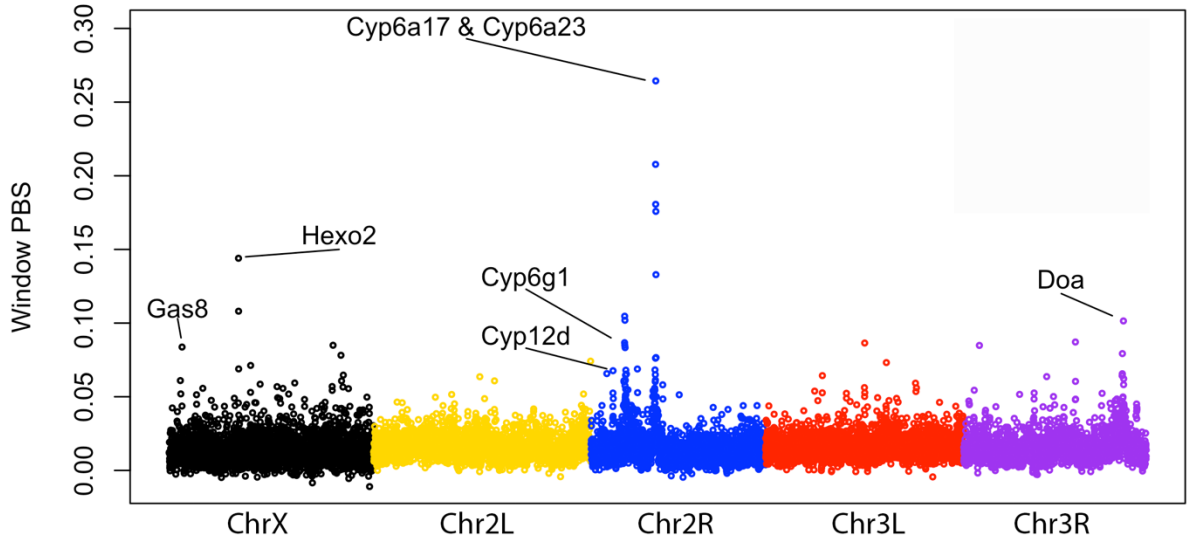


Figure 3: Inversion frequencies across time and space. In panel A, we show inversion frequencies at 8 well-studied chromosomal inversions. The asterisks underneath *Inv(2R)Ns*, *Inv(3R)Mo*, and *Inv(3R)Payne* indicate statistically significant frequencies between seasons. In panel B, we show the 4 inversions with with significant frequency differences between sampling locations in the spring. No inversions displayed significant frequency differences between locations in the fall samples.

A



B

GO Category	Number of Outlier Regions	Raw P Value
rhythmic synaptic transmission	2	0.000326817
integral to synaptic vesicle membrane	2	0.003433584
iron ion binding	7	0.004231579
response to insecticide	3	0.004615539
nucleotidase activity	2	0.007111779
RNA polymerase II core promoter proximal region sequence-specific DNA binding	2	0.007193985
peptidyl-amino acid modification	5	0.007576942
cytokine binding	2	0.007844612
oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	7	0.007851629
synaptic transmission	7	0.007890727

Figure 4: Population Branch Statistic and Gene Ontology Enrichment. In panel A, we show PBS at each window genome-wide. Gene names are discussed in the text. In Panel B, we list the top 10 categories in our GO enrichment analysis.

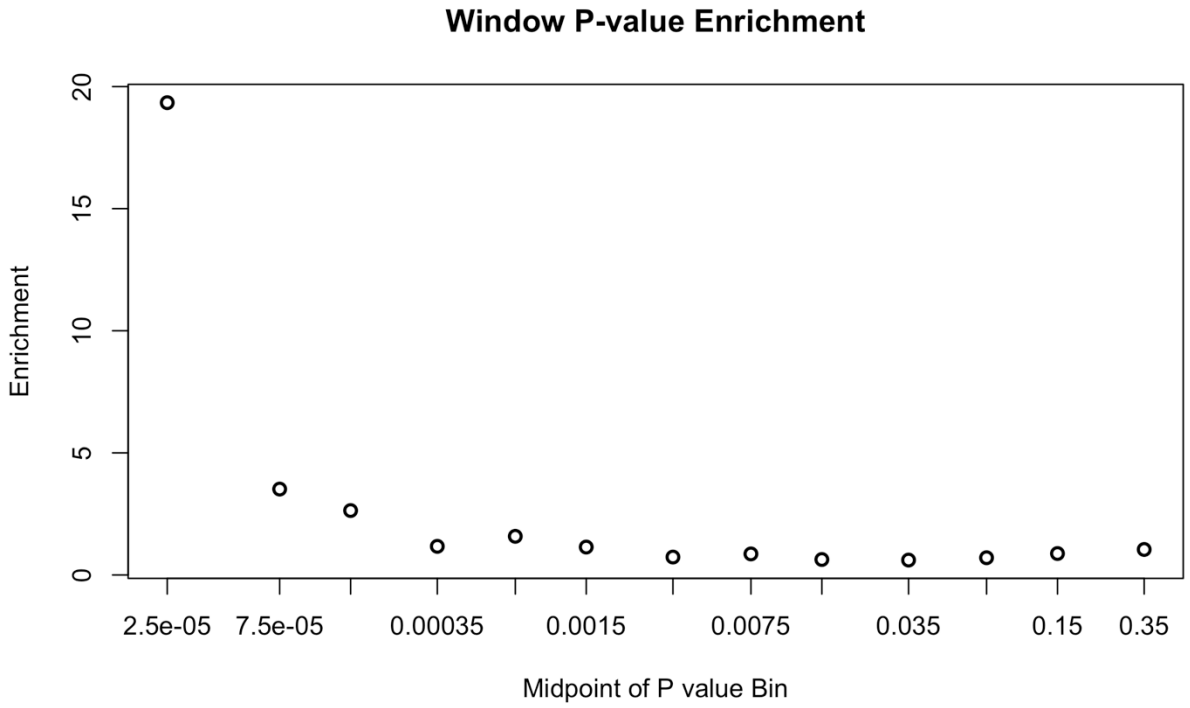


Figure 5: P value enrichment of low window-PBS P values. Enrichment is the multiple of the expected number of windows based on the expectation of a uniform distribution of P values. We would expect, on average, fewer than 1 window with a P value between 0 and 5E-5. Instead, we find nearly 20 times more windows in this P value bin than we expected.

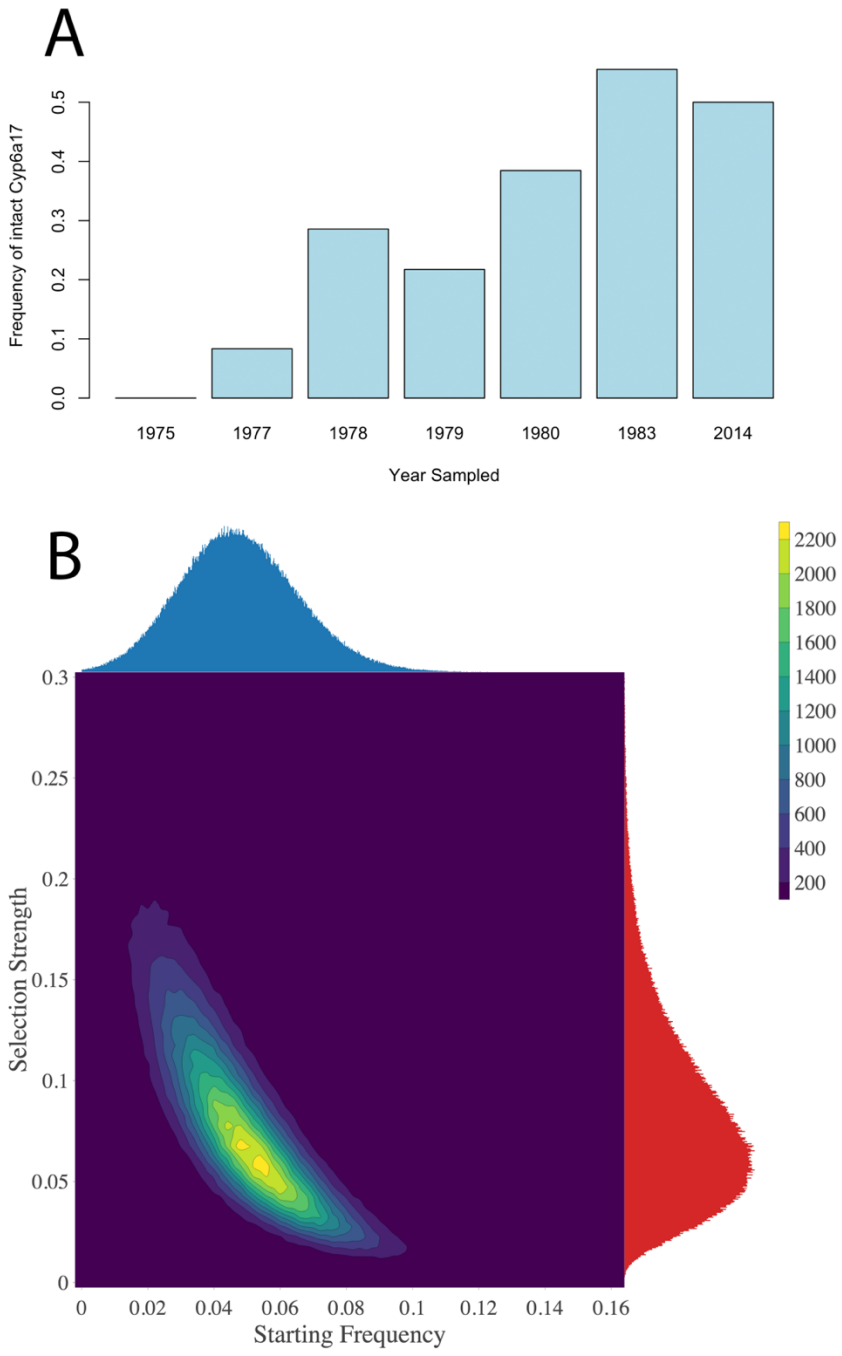


Figure 6: Frequency shift of the *Cyp6a17* allele. In panel A, we display the frequency of the intact *Cyp6a17* in each of the 6 sampling years and an estimate of the frequency in modern populations. In panel B, we show results of an ABC analysis to infer the selection strength and starting frequency of the intact allele that best recapitulate our empirical sampling results.

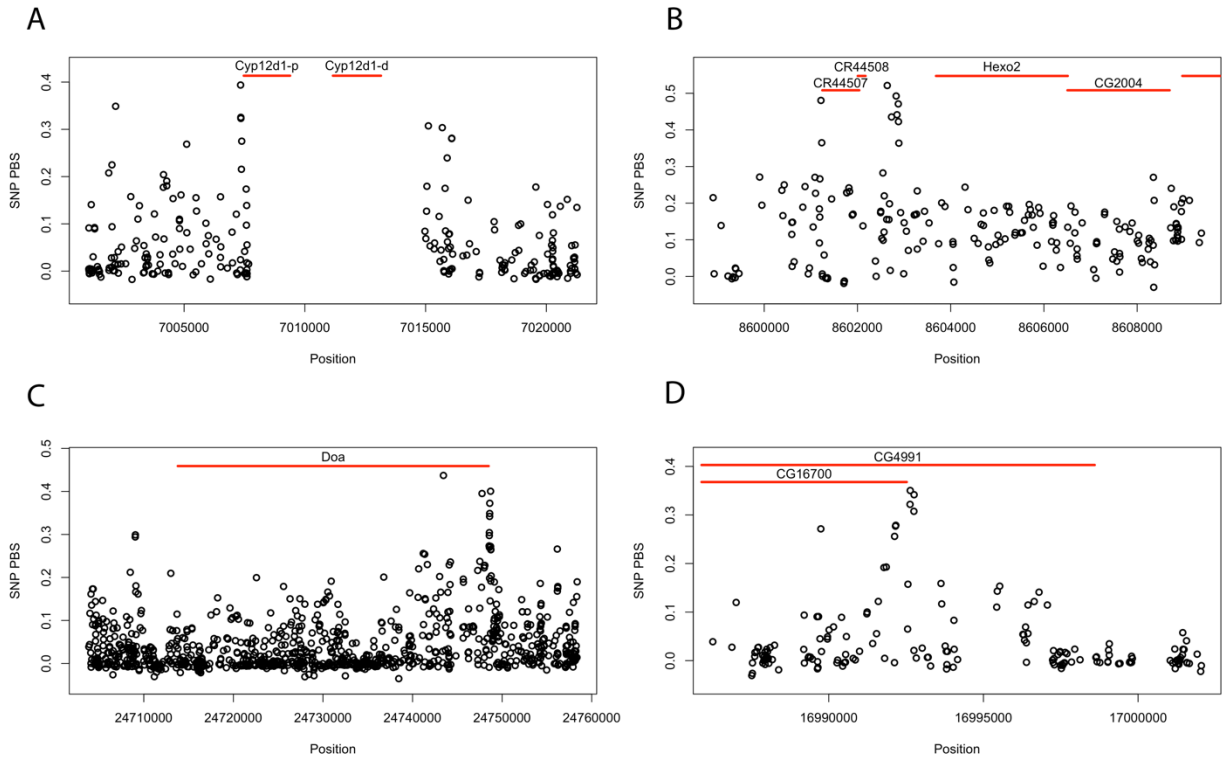
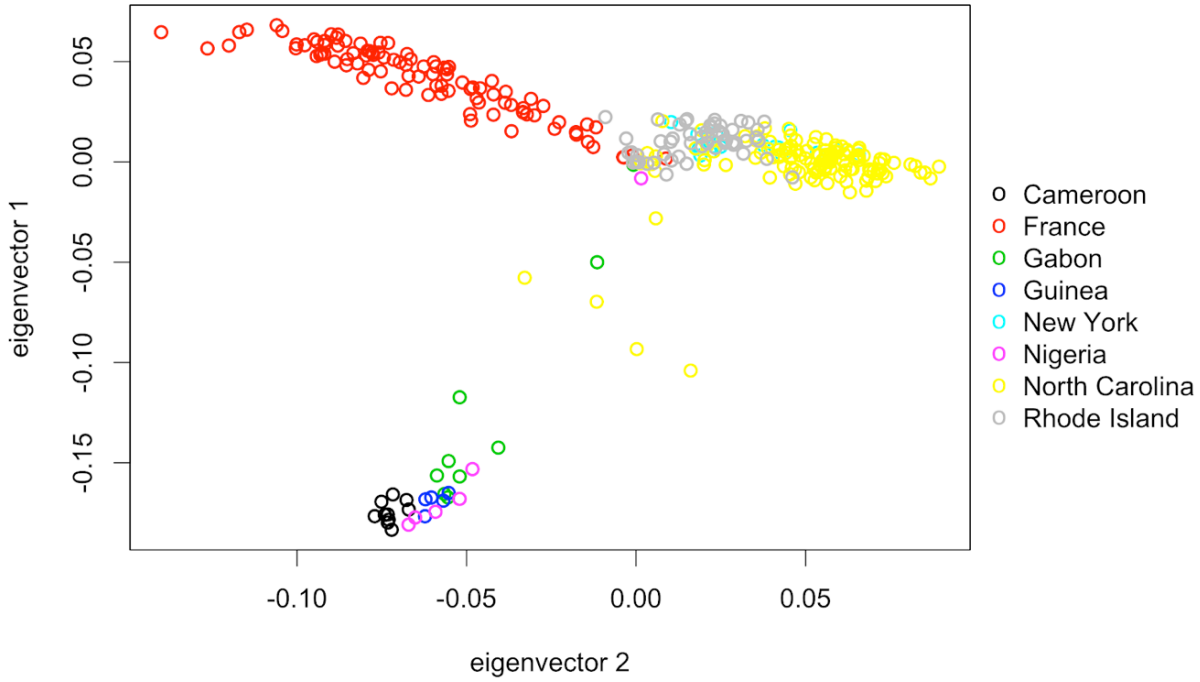


Figure 7: SNP PBS plots. Panels A, B, C, and D show PBS plots at the SNP level for 4 window outliers.

Figure 8: A PCA plot reveals no abberrant divergence among the old strains. Inversion-free



strains from 7 additional Northa American, European, and African populations were used in this analysis.

Table 1:
African ancestry proportions

Sample	Chr2	Chr3	ChrX
Old Sample	0.171188	0.1711506	0.071403
Fall Sample	0.197540209	0.1415769	0.04537338
Spring Sample	0.182260946	0.150928574	0.055800583

Table 2:
Inversion frequencies and P values for frequency differences

Sample	Inv(2L)T	Inv(2R)Ns	Inv(3L)P	Inv(3R)C	Inv(3R)K	Inv(3R)Mo	Inv(3R)Payne
Old	0.078125	0.10483871	0.015625	0.0625	0.0234375	0.078125	0.120967742
Fall	0.1562824	0.06842004	0.00591042	0.01957136	0.01775695	0.05489982	0.03944041
Spring	0.1134851	0.02919873	0.001942423	0.0231948	0.002546296	0.08952676	0.0887487
Between Season P value	0.0513	0.0033	0.2934	0.7033	0.0578	0.0301	0.0016
Within Season P value (fall)	0.8047	0.8814	0.3119	0.8428	0.4765	0.8936	0.8537
Within Season P value (spring)	0.1486	0.0023	0.9261	0.0036	0.7653	0.039	0

Table 3:
Top window PBS outliers

Chromosome Arm	Start	Stop	Number of windows	Top window PBS	Top window P value	Genes within 10 KB of peak
2R	10450310	11002593	78	0.207742416	4.00E-07	Inr-a, Cyp6a22, Cyp6a17, Cyp6a23, Cyp6a19, Cyp6a9, Cyp6a20, Cyp6a21, Cyp6a8
X	8549576	8715139	26	0.144007414	1.20E-06	CR44507, CR44508, Hexo2, CG2004, CG1785, l(1)G0020
2R	7935690	8252783	42	0.101931344	3.64E-05	Cyp6g2, Cyp6t3, CG8858, RpS11, Sr-CII, CG13171
3R	24459659	25160619	88	0.10145107	4.68E-05	Doa, CG11828, DIP-gamma
X	3159368	3404025	35	0.083742934	8.88E-05	CG10802, CG14270, CG10803, Gas8, DIP-alpha, CG13021
3R	19412431	19432121	4	0.087161167	0.0001364	CG10182, CG33337, CG16723, CG10183, CG10184, CG31145
X	16919721	17013928	14	0.084945674	0.000156	CG4991, CG16700, Arpc3B, CG5004
3R	8856383	8882683	3	0.084843975	0.000224805	ry, CG11668, snk, CG11670, Hsc70-2, CG31157, CG7966, pic, sim
X	18020584	18110533	14	0.078167946	0.000228	CG32553, mir-369, mir-210, CG34133, ari-1, CG43229

X	9526894	9557457	4	0.071249591	0.0003616	Ptpmeg2, CG3106, nej
3L	10456488	10564064	20	0.073195577	0.0004108	A2bp1
2L	22212036	22940522	16	0.074164529	0.0006188	IR40a, CR12628
2R	6988766	7043416	10	0.067670948	0.00096	Cyp12d1-p, Cyp12d1-d, BBS4
X	18302386	18449416	21	0.064695885	0.0009684	CG32548, CG6290, CG32551, CG34841, CG32547
2R	9169960	9219511	7	0.068846351	0.0010652	Dh31-R, CG4734, CG17047
X	10212149	10403273	22	0.058532159	0.0012208	CG32681, CG17841, Psf3, flw
X	12025461	12179466	23	0.056873235	0.0014892	Ten-a, CG1924
2R	6448155	6451899	1	0.065713104	0.001506	psq
X	12439232	12514399	10	0.055667386	0.001704	Sec16, CG1463, Fpgs, CG11085, CR44568
3L	4892819	4920139	4	0.06434811	0.0017332	Dnah3, CG13705, CG13704, Rh50

Table 4:
Top SNP outliers that are NOT window outliers

Chromosome Arm	Window Start	Window Stop	Window PBS	Max SNP PBS	Max SNP position	Frequency in Old Samples	Frequency in Fall	Frequency in Spring	Location of max SNP
X	12427570	12433228	0.0240281	0.4909074	12428485	0.354545455	0.006297415	0.058388449	Intron of hwt
2R	9039861	9049013	0.0321952 6	0.4554215	9047945	0.207090582	0	0	Exon of Nrk
2L	11494111	11498366	0.0115067 6	0.4333918	11497953	0.338325215	0.026731642	0.038016501	Intron of Cog8
2R	9022224	9039860	0.0268164 8	0.4100635	9038358	0.185749627	0	0	Intron of Ack-like
2R	11997669	12004050	0.0105064 8	0.4003973 21	12000155	0.194912781	0.003228307	0.003990455	Exon of Asph
3R	15348248	15355832	0.0171768 8	0.3972696 03	15353076	0.490249851	0.176797191	0.042977868	Intron of Dys
2L	6009243	6014163	0.0184062 7	0.3870668	6011436	0.351536901	0.066171751	0.031144078	Intron of CG9098
X	17218724	17225200	0.0200047 7	0.3791857	17221539	0.460317466	0.112060171	0.105419639	Intergenic Region, close gene B-H2
2L	12722715	12729407	0.0289372 9	0.349127	12725414	0.621738626	0.252107242	0.183309081	Intron of MRP1
2R	11931244	11941771	0.0179711 6	0.3457392	11935869	0.155825299	0	0	Exon of CG8405
2R	6608167	6617712	0.0175865	0.3418838	6616156	0.333755348	0.072867198	0.037034912	5' UTR ofr Rab3
X	19627455	19638361	0.0178796 6	0.3370488	19629237	0.774509804	0.19753754	0.486251693	Intergenic Region, flanked by AP-1-2beta and CG14234

X	1041 6048	1043 0411	0.023 1622 7	0.334 2488	104182 89	0.35	0.0634 45092	0.03436 1189	Intron of spri
2L	1680 1630	1680 8036	0.030 0385 7	0.327 5541	168019 68	0.3699137 36	0.0960 55911	0.03279 7253	Exon of CG13280

Table 5:
Mean coverage of sequencing depth in each of the 18 pools.

Season	Pool	Chr2L	Chr2R	Chr3L	Chr3R	ChrX	Number of Flies in Pool
Fall	Pool 1	104.8601	123.8057	107.1681	117.9607	111.8739	41
Fall	Pool 2	37.36802	44.97712	38.38659	42.39372	40.9902	41
Fall	Pool 3	36.82978	42.60032	37.52397	40.97629	38.54016	41
Fall	Pool 4	43.52798	48.73852	44.19712	47.87891	44.08533	41
Fall	Pool 5	37.84135	44.23504	38.70059	42.49779	39.97496	41
Fall	Pool 6	96.86587	114.3169	98.67793	109.5035	102.2905	42
Spring	Pool 1	19.05707	19.23281	19.21127	19.49868	18.71161	34
Spring	Pool 2	15.12735	15.27883	15.24809	15.4329	14.86802	34
Spring	Pool 3	16.42694	16.59842	16.57783	16.80964	16.11497	34
Spring	Pool 4	15.36229	15.60561	15.48307	15.77644	15.11505	34
Spring	Pool 5	20.52848	21.13244	20.67186	21.34098	20.04423	34
Spring	Pool 6	13.43736	13.60501	13.52322	13.79348	13.16437	34
Spring	Pool 7	13.28058	13.78171	13.46408	13.93349	12.98276	34
Spring	Pool 8	15.36458	15.80972	15.60859	15.99578	15.09753	34
Spring	Pool 9	14.58336	14.8238	14.74561	15.00834	14.38315	34
Spring	Pool 10	16.19566	16.47395	16.37867	16.64322	15.84015	34
Spring	Pool 11	16.39418	16.58594	16.52368	16.75612	16.05663	34
Spring	Pool 12	16.55574	16.82262	16.69146	17.08097	16.18909	34

References

Ayroles JF, Carbone MA, Stone EA, Jordan KW, Lyman RF, Magwire MM, Rollmann SM, Duncan LH, Lawrence F, Anholt RR, Mackay TF. (2009) Systems genetics of complex traits in Drosophila melanogaster. Nature Genetics, 41(3):299-307.

Barrick JE, Yu DS, Yoon SH, Jeong H, Oh TK, Schneider D, Lenski RE, Kim JF. (2009) Genome evolution and adaptation in a long-term experiment with Escherichia coli. Nature, 461(7268):1243-7.

Battlay P, Leblanc PB, Green L, Garud NR, Schmidt JM, Fournier-Level A, Robin C. (2018) Structural Variants and Selective Sweep Foci Contribute to Insecticide Resistance in the *Drosophila* Genetic Reference Panel. G3 (Bethesda), 8(11):3489-3497.

Bergland AO, Behrman EL, O'Brien KR, Schmidt PS, Petrov DA. (2014) Genomic evidence of rapid and stable adaptive oscillations over seasonal time scales in Drosophila. PLoS Genetics, 10(11):e1004775.

Burger J, Kirchner M, Bramanti B, Haak W, Thomas, MG. (2007) Absence of the lactase-persistence-associated allele in early Neolithic Europeans. Proceedings of the National Academy of Sciences of the United States of America, 104:3736–3741

Campbell-Staton SC, Cheviron ZA, Rochette N, Catchen J, Losos JB, Edwards SV. (2017) Winter storms drive rapid phenotypic, regulatory, and genomic shifts in the green anole lizard. Science, 357(6350):495-498.

Castañeda-Rico S, León-Paniagua L, Edwards CW, Maldonado JE. (2020) Ancient DNA From Museum Specimens and Next Generation Sequencing Help Resolve the Controversial Evolutionary History of the Critically Endangered Puebla Deer Mouse. Frontiers in Ecology and Evolution, 8:94.

Cattaneo F, Pasini ME, Intra J, Matsumoto M, Briani F, Hoshi M, Perotti ME. (2006) Identification and expression analysis of *Drosophila melanogaster* genes encoding beta-hexosaminidases of the sperm plasma membrane. Glycobiology, 16(9):786-800.

Chen N, Juric I, Cosgrove EJ, Bowman R, Fitzpatrick JW, Schoech SJ, Clark AG, Coop G. (2019) Allele frequency dynamics in a pedigreed natural population. Proceedings of the National Academy of Sciences of the United States of America, 116(6):2158-2164.

Chung H, Bogwitz MR, McCart C, Andrianopoulos A, Ffrench-Constant RH, Batterham P, Daborn PJ. (2007) Cis-regulatory elements in the Accord retrotransposon result in tissue-specific expression of the *Drosophila melanogaster* insecticide resistance gene Cyp6g1. Genetics, 175(3):1071-7.

Colosimo PF, Hosemann KE, Balabhadra S, Villarreal G Jr, Dickson M, Grimwood J, Schmutz J, Myers RM, Schluter D, Kingsley DM. (2005) Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. Science, 307(5717):1928-33.

Comeron JM, Ratnappan R, Bailin S. (2012) The many landscapes of recombination in Drosophila melanogaster. PLoS Genetics, 8(10):e1002905.

Corbett-Detig RB, Hartl DL. (2012) Population genomics of inversion polymorphisms in Drosophila melanogaster. PLoS Genetics, 8(12):e1003056.

Corbett-Detig R, Nielsen R. (2017) A hidden Markov model approach for simultaneously estimating local ancestry and admixture time using next generation sequence data in samples of arbitrary ploidy. PLoS Genetics, 13(1):e1006529.

Daborn P, Boundy S, Yen J, Pittendrigh B, ffrench-Constant R. (2001) DDT resistance in *Drosophila* correlates with Cyp6g1 over-expression and confers cross-resistance to the neonicotinoid imidacloprid. Molecular Genetics & Genomics, 266(4):556-63.

Daborn PJ, Lumb C, Boey A, Wong W, ffrench-Constant RH, Batterham P. (2007) Evaluating the insecticide resistance potential of eight *Drosophila melanogaster* cytochrome P450 genes by transgenic over-expression. Insect Biochemistry and Molecular Biology, 37(5):512-519.

Daborn PJ, Yen JL, Bogwitz MR, Le Goff G, Feil E, Jeffers S, Tijet N, Perry T, Heckel D, Batterham P, Feyereisen R, Wilson TG, ffrench-Constant RH. (2002) A single p450 allele associated with insecticide resistance in Drosophila. Science, 297(5590):2253-6.

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature Genetics, 43(5):491-8.

Duchen P, Zivkovic D, Hutter S, Stephan W, Laurent S. (2013) Demographic inference reveals African and European admixture in the North American *Drosophila melanogaster* population. Genetics, 193(1):291-301.

Edwards AC, Rollmann SM, Morgan TJ, Mackay TF. (2006) Quantitative genomics of aggressive behavior in *Drosophila melanogaster*. PLoS Genetics, 2(9):e154.

Feder AF, Rhee S-Y, Holmes SP, Shafer RW, Petrov DA, Pennings PS. (2016) More effective drugs lead to harder selective sweeps in the evolution of drug resistance in HIV-1. eLife, 5:e10670-e.

Ferretti L, Ramos-Onsins SE, Pérez-Enciso M. (2013) Population genomics from pool sequencing. Molecular Ecology, 22(22):5561-76.

Foll M, Poh YP, Renzette N, Ferrer-Admetlla A, Bank C, Shim H, Malaspinas AS, Ewing G, Liu P, Wegmann D, Caffrey DR, Zeldovich KB, Bolon DN, Wang JP, Kowalik TF, Schiffer CA, Finberg RW, Jensen JD. (2014) Influenza virus drug resistance: a time-sampled population genetics perspective. PLoS Genetics, 10(2):e1004185.

Fumey J, Wicker-Thomas C. (2017) Mutations at the Darkener of Apricot locus modulate pheromone production and sex behavior in Drosophila melanogaster. Journal of Insect Physiology, 98:182-187.

Gautier M, Gharbi K, Cezard T, Foucaud J, Kerdelhué C, Pudlo P, Cornuet JM, Estoup A. (2013) The effect of RAD allele dropout on the estimation of genetic variation within and between populations. Molecular Ecology, 22(11):3165-78.

Good RT, Gramzow L, Battlay P, Sztal T, Batterham P, Robin C. (2014) The molecular evolution of cytochrome P450 genes within and between drosophila species. Genome Biology and Evolution, 6(5):1118-34.

Hoekstra HE, Hirschmann RJ, Bundey RA, Insel PA, Crossland JP. (2006) A single amino acid mutation contributes to adaptive beach mouse color pattern. Science, 313(5783):101-4.

Hofmanová Z, Kreutzer S, Hellenthal G, Sell C, Diekmann Y, Diez-del-Molino D, *et al*. (2016) Early farmers from across Europe directly descended from Neolithic Aegeans. Proceedings of the National Academy of Sciences. 133:6886–91.

Huang W, Lyman RF, Lyman RA, Carbone MA, Harbison ST, Magwire MM, Mackay TF. (2016) Spontaneous mutations and the origin and maintenance of quantitative genetic variation. eLife 5:e14625.

Hudson RR. (2002) Generating samples under a Wright-Fisher neutral model of genetic variation. Bioinformatics, 18(2):337-8.

Intra J, Veltri C, De Caro D, Perotti ME, Pasini ME. (2017) In vitro evidence for the participation of *Drosophila melanogaster* sperm β-*N*-acetylglucosaminidases in the interactions with glycans carrying terminal *N*-acetylglucosamine residues on the egg's envelopes Archives of Insect Biochemsitry and Physiology, 96 (2017), p. e21403.

Kao JY, Zubair A, Salomon MP, Nuzhdin SV, Campo D. (2015) Population genomic analysis uncovers African and European admixture in Drosophila melanogaster populations from the south-eastern United States and Caribbean Islands. Molecular Ecology, 24(7):1499-509.

Kapun M, van Schalkwyk H, McAllister B, Flatt T, Schlotterer C. (2014) Inference of chromosomal inversion dynamics from Pool-Seq data in natural and laboratory populations of *Drosophila melanogaster*. Molecular Ecology 23: 1813–1827.

Kapun M, Fabian DK, Goudet J, Flatt T. (2016) Genomic evidence for adaptive inversion clines in *Drosophila melanogaster*. Molecular Biology and Evolution, 33(5):1317-36.

Karasov T, Messer PW, Petrov DA. (2010) Evidence that adaptation in Drosophila is not limited by mutation at single sites. PLoS Genetics, 6(6):e1000924.

Keller A. (2007) Drosophila melanogaster's history as a human commensal. Current Biology, 17: R77–81.

Kirkpatrick M, Barton N. (2006) Chromosome inversions, local adaptation and speciation. Genetics, 173(1):419-34.

Kish L. (1965). *Survey sampling.* New York: Wiley.

Kliot A, Ghanim M. (2012) Fitness costs associated with insecticide resistance. Pest Management Science, 68(11):1431-7.

Kofler R, Orozco-terWengel P, De Maio N, Pandey RV, Nolte V, Futschik A, Kosiol C, Schlötterer C. (2011) PoPoolation: a toolbox for population genetic analysis of next generation sequencing data from pooled individuals. PLoS One, 6(1):e15925.

Lack JB, Cardeno CM, Crepeau MW, Taylor W, Corbett-Detig RB, Stevens KA, Langley CH, Pool JE. (2015) The Drosophila genome nexus: a population genomic resource of 623 *Drosophila melanogaster* genomes, including 197 from a single ancestral range population. Genetics, 199(4):1229-41.

Langley CH, Stevens K, Cardeno C, Lee YCG, Schrider DR, Pool JE, Langley SA et al. (2012) Genomic variation in natural populations of *Drosophila melanogaster*. Genetics, 192:533–598.

Li H, Durbin R. (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics, 26(5):589-95.

Lindqvist C, Schuster SC, Sun Y, Talbot SL, Qi J, *et al*. (2010) Complete mitochondrial genome of a Pleistocene jawbone unveils the origin of polar bear. Proceedings of the National Academy of Sciences of the United States of America, 107:5053–5057.

Lunter G, Goodson M. (2011) Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. Genome Research, 21(6):936-9.

Machado HE, Bergland AO, O'Brien KR, Behrman EL, Schmidt PS, Petrov DA. (2016) Comparative population genomics of latitudinal variation in *Drosophila simulans* and *Drosophila melanogaster*. Molecular Ecology, 25(3):723-40.

Marrone AK, Kucherenko MM, Rishko VM, Shcherbata HR. (2011) New dystrophin/dystroglycan interactors control neuron behavior in Drosophila eye. BMC Neuroscience, 12:93.

Mathieson I, Lazaridis I, Rohland N, Mallick S, Patterson N, Roodenberg S, *et al*. (2015) Genome-wide patterns of selection in 230 ancient Eurasians. Nature 528:499-503.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research, 20(9):1297-303.

McInnis DO, Schaffer HE, Mettler LE. (1982). Field dispersal and population sizes of native *Drosophila* from North Carolina. The American Naturalist, 119(3):319-330.

Menozzi P, Shi MA, Lougarre A, Tang ZH, Fournier D. (2004) Mutations of acetylcholinesterase which confer insecticide resistance in *Drosophila melanogaster* populations. BMC Evolutionary Biology, 4:4.

Mettler LE, Voelker RA, Mukai T. (1977) Inversion clines in populations of *Drosophila melanogaster*. Genetics, 87(1):169-76.

Mikheyev AS, Tin MMY, Arora J, Seeley TD. (2015) Museum samples reveal rapid evolution by wild honey bees exposed to a novel parasite. *Nature Communications*, **6**: 1–8.

Mutero A, Pralavorio M, Bride JM, Fournier D. (1994) Resistance-associated point mutations in insecticide-insensitive acetylcholinesterase. Proceedings of the National Academy of Sciences of the United States of America, 91(13):5922-6.

Najarro MA, Hackett JL, Smith BR, Highfill CA, King EG, Long AD, Macdonald SJ. (2015) Identifying loci contributing to natural variation in xenobiotic resistance in Drosophila. PLoS Genetics, 11(11):e1005663.

Nelson CS, Beck JN, Wilson KA, Pilcher ER, Kapahi P, Brem RB. (2016) Cross-phenotype association tests uncover genes mediating nutrient response in Drosophila. BMC Genomics 17(1): 867.

Noonan JP, Hofreiter M, Smith D, Priest JR, Rohland N, *et al*. (2005) Genomic sequencing of Pleistocene cave bears. Science 309: 597–599.

Noor MA, Grams KL, Bertucci LA, Reiland J. (2001) Chromosomal inversions and the reproductive isolation of species. Proceedings of the National Academy of Sciences of the United States of America, 98(21):12084-8.

Parts L, Cubillos FA, Warringer J, Jain K, Salinas F, Bumpstead SJ, Molin M, Zia A, Simpson JT, Quail MA, Moses A, Louis EJ, Durbin R, Liti G. (2011) Revealing the genetic structure of a trait by sequencing a population under selection. Genome Research, 21(7):1131-8.

Pelissie B, Crossley MS, Cohen ZP, Schoville SD. (2018) Rapid evolution in insect pests: the importance of space and time in population genomics studies. Current Opinion in Insect Science, 26, 8–16.

Pool JE, Corbett-Detig RB, Sugino RP, Stevens KA, Cardeno CM, Crepeau MW, Duchen P, Emerson JJ, Saelao P, Begun, DJ, Langley CH. (2012) Population genomics of sub-

Saharan *Drosophila melanogaster*: African diversity and non-African admixture. *PLoS Genetics* 8:e1003080.

Pool JE. (2015) The mosaic ancestry of the Drosophila genetic reference panel and the D. melanogaster reference genome reveals a network of epistatic fitness interactions. Molecular Biology and Evolution, 32(12):3236-51.

Pool JE, Braun DT, and Lack JB (2017). Parallel evolution of cold tolerance within *Drosophila melanogaster*. Molecular Biology and Evolution, 34, 349–360.

Racimo F, Sankararaman S, Nielsen R, Huerta-Sánchez E. (2015) Evidence for archaic adaptive introgression in humans. Nature Reviews: Genetics, 16(6):359-71.

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. (2011) Integrative genomics viewer. Nature Biotechnology, 29(1):24-6.

Schmidt JM, Battlay P, Gledhill-Smith RS, Good RT, Lumb C, Fournier-Level A, Robin C. (2017) Insights into DDT resistance from the *Drosophila melanogaster* genetic reference panel. Genetics, 207(3):1181-1193.

Sprengelmeyer QD, Mansourian S, Lange JD, Matute DR, Cooper BS, Jirle EV, Stensmyr MC, Pool JE (2020). Recurrent collection of *Drosophila melanogaster* from wild African environments and genomic insights into species history. Molecular Biology and Evolution. 37(3):627–638.

Sturtevant AH, Beadle GW. (1936) The relations of inversions in the X chromosome of *Drosophila melanogaster* to crossing over and disjunction. Genetics, 21(5):554-604.

Svedberg J*,* Shchur V, Reinman S, Nielsen R, Corbett-Detig R. (2020) Inferring adaptive introgression using Hidden Markov Models. [Ahead of print]

Svetec N, Werzner A, Wilches R, Pavlidis P, Alvarez-Castro JM, Broman KW, Metzler D, Stephan W. (2011) Identification of X-linked quantitative trait loci affecting cold tolerance in *Drosophila melanogaster* and fine mapping by selective sweep analysis. Molecular Ecology, 20(3):530-44.

Turelli M, Hoffmann AA. (1995). Cytoplasmic incompatibility in *Drosophila simulans*: dynamics and parameter estimates from natural populations. Genetics, 140(4):1319-1338.

Turner TL, Stewart AD, Fields AT, Rice WR, Tarone AM. (2011) Population-based resequencing of experimentally evolved populations reveals the genetic basis of body size variation in *Drosophila melanogaster.* PLoS Genetics, 7(3):e1001336.

Wilches R, Voigt S, Duchen P, Laurent S, Stephan W. (2014) Fine-mapping and selective sweep analysis of QTL for cold tolerance in *Drosophila melanogaster*. G3 (Bethesda), 4(9):1635-45.

Yeh SD, Chen YJ, Chang AC, Ray R, She BR, Lee WS, Chiang HS, Cohen SN, Lin-Chao S. (2002) Isolation and properties of Gas8, a growth arrest-specific gene regulated during male gametogenesis to produce a protein associated with the sperm motility apparatus. Journal of Biological Chemistry, 277(8):6311-7.

Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZX, Pool JE, *et al.* (2010) Sequencing of 50 human exomes reveals adaptation to high altitude. Science, 329(5987):75-8.

Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. (2012) A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics, 28(24):3326-8.

Zur Lage P, Newton FG, Jarman AP. (2019) Survey of the ciliary motility machinery of *Drosophila* sperm and ciliated mechanosensory neurons reveals unexpected cell-type specific variations: a model for motile ciliopathies. Frontiers in Genetics, 10:24.

Zwarts L, Magwire MM, Carbone MA, Versteven M, Herteleer L, Anholt RR, Callaerts P, Mackay TF. (2011) Complex genetic architecture of Drosophila aggressive behavior. Proceedings of the National Academy of Sciences of the United States of America, 108(41):17070-5.

Chapter 4: Curating the past: Next generation sequencing of museum specimens reveal recent adaptive targets in *Drosophila melanogaster*

Abstract

Natural history collections (NHC) offer important insights into *Drosophila melanogaster's* geographic and genomic past. Researchers can leverage these collections to study specimens from hundreds of years ago, helping to determine important evolutionary changes and migration patterns. In this study, we analyze genomic data collected from 8 *D. melanogaster* museum specimens housed at the Lund University Natural History Museum in Lund, Sweden. The historic collection, dating back to the 1840's, offers some of the earliest glimpses into what genetic variation looked like for possibly the first *D. melanogaster* colonizers of Europe. Using the allele frequency-based Population Branch Excess (PBE) statistic, we begin to reveal recent targets of adaptive evolution. We observe significant change in genetic variation at a well-known insecticide resistance gene, *Cyp6g1*. We also uncover significant change at *Choline acetyltransferase* (*ChAT*), which breaks down the neurotransmitter acetylcholine. Overall, we reveal that insecticide resistance genes may be one of the most common targets of recent adaptation in modern Lund populations. This study also provides limited evidence that inversions may have been at a lower frequency than modern populations, giving additional evidence to the hypothesis that inversions are a more recent arrival into modern European populations.

Introduction

Natural history collections (NHC) are an important and rich data source for evolutionary biologists. In addition to preserving up to several hundred years of biological history, NHC

provide a wealth of information about geographic and temporal factors that impact a species' evolution. Many historical collections also include important meta-data, such as field notes and contemporary observations about sampling locations. Scientists can use this information to inform a variety of research topics, including questions about biodiversity, climate, historical events, predation, and much more.

The earliest genetic studies utilizing NHC examined the spread of the *medionigra* gene in the scarlet tiger moth (Fisher and Ford 1947), providing clear evidence of natural selection in action. In more recent decades, genetic studies have utilized sequencing technology to examine temporal and geographic changes in genetic variation across numerous species. For instance, researchers amplified a portion of the mitochondrial genome of the extinct New Zealand bird, the moa, to show that it is not related to the extant kiwi (Cooper *et al.* 1992), thus providing evidence of parallel adaptation of flightless birds.

Genetic studies using NHC have also proven to be important in conservation efforts, as museum specimens are often representative of genetic diversity just prior to human-associated environmental changes. To that end, researchers have used genetic data from NHC to infer population size changes (Miller and Waits 2003; Nyström *et al.* 2006), introgression (Saltonstall 2002), and migration (Flagstad *et al.* 2003). These insights can help clarify the impact of environment and human activity on a population and inform conservation efforts for endangered species.

The studies cited above, and in fact most studies to date, are limited to small DNA fragments from relatively few loci. This is due to technical limitations of PCR amplification due to degraded and heavily fragmented DNA, a typical characteristic of DNA obtained from NHC. Such degradation is less of an issue when utilizing next generation sequencing (NGS)

technologies such as Illumina sequencing, which is designed for short, fragmented DNA. Advancement in NGS technologies has allowed researchers to extract much more data from NHC and begin to identify specific genes where allele frequencies have shifted between the collection date and modern samples. For example, Parejo *et al.* (2020) recently used NGS to sequence whole genomes from an NHC collection of 22 Swiss honeybees. The authors revealed evidence of recent bottlenecks and targets of adaptation despite only having heavily fragmented DNA available for sequencing.

The research presented in this study examines change in genetic variation between modern European samples of *Drosophila melanogaster* and 8 museum specimens collected in 1845 (1 specimen), 1859 (1 specimen), and 1933 (6 specimens). The earlier of the samples may represent some of the earliest colonizers of *D. melanogaster* in Europe (Keller 2007). With this unique data set, we begin to reveal targets of recent adaptation in modern Lund populations. We are also able to measure frequencies of well-known chromosomal inversions, helping to elucidate the timing of the arrival of inversions into modern European D. melanogaster populations.

Results and Discussion

Collection and Quality Analysis of Museum-Derived Genome Sequences

Whole genomic sequences were extracted from 8 specimens from the Lund University Natural History Museum in Lund, Sweden. Sequencing depth was highly variable across the 8 samples, ranging from a median of 2 to 62 (Table 1). In the sample with the lowest depth, 46% of the genome had enough coverage to call an allele. A principal components analysis revealed no aberrant divergence among the 8 strains sequenced (Figure 1).

DNA from museum specimens degrades over time for a multitude of reasons. The DNA samples from museum specimens in this study were heavily fragmented and exhibited adapter read through resulting from insert sizes being smaller than the read length. After read trimming, the median read length was 47 base pairs (Figure 2A).

A second concern when sequencing museum specimens is cytosine deamination, resulting in modifications that cause cytosine to be misread by DNA polymerases. Deamination of cytosine creates uracil, which directs the incorporation of adenine during amplification. This results in apparent C to T substitutions and G to A substitutions, both of which have been found to be greatly increased in other studies of ancient DNA (Stiller *et al.* 2006; Gilbert *et al.* 2007). Despite this, we did not observe an excess of C to T or G to A substitutions in our data (Figures 2B and 2C). With these quality assurance considerations, we believe that the sequenced museum specimens offer an accurate snapshot of genetic variation from decades prior.

No Evidence of Inversions in Museum Specimens

The museum specimens used in this study offer a rare look at inversion frequencies in some of the earliest migrants into Europe. Inversion status in each museum specimen was called based on alleles at inversion-associated SNPs (Kapun *et al.* 2016). We examined the diploid calls at each inversion-associated SNP to determine the presence of the inversion in the individual strain. Perhaps surprisingly, we did not find evidence for any inversions in any of the 6 strains from the Lund samples nor in the earlier non-Swedish samples. In the modern Lund sample used in this study, *Inv(2L)t* was at 15.2% frequency, while other autosomal inversions varied between 0.69% and 4.1% (Kapun *et al.* 2020). The probability of a sampling deviation in our 6 Lund museum sample as far as we observe here, assuming that inversion frequencies have not shifted between time points, is 7.216% (2-tailed binomial test). One of our older museum samples is

from Denmark, very close geographically to Lund, Sweden, and also did not contain inversions. If we add this sample to our binomial test, our P value drops to 0.0415. Thus, it is possible that inversion frequencies have increased since 1933.

Inversion frequencies could be partly modulated by local environment. We know that inversions are at lower frequency in northern *D. melanogaster* populations (Kapun *et al.* 2016). It is possible that the presence of an inversion offers tradeoffs for an individual organism and frequencies at the population level are held in balance. Northern Europe has a significantly colder climate than the ancestral sub-Saharan range of *D. melanogaster*, and likely was even harsher for flies 100 years ago due to having less developed infrastructure and fewer heated indoor spaces. With increasing infrastructure and climate control, flies with inversion-bearing chromosomes may be more successful. This hypothesis is also supported by population genomic studies of inversions in *D. melanogaster*. Recent studies have shown that inverted haplotypes in France have higher genetic diversity than their non-inverted counterparts (Pool *et al.* 2012) and that they are more genetically related to African populations (Corbett-Detig and Hartl 2012). This has led to the suggestion that inverted chromosomes are a more recent migrant into Europe than the initial out of Africa bottleneck. However, the phenotypic consequences of these inversions remain largely unknown.

Potential Adaptive Differences Between Sampling Points

To identify genes that have been selected over the last several decades, we used the Population Branch Excess (PBE) statistic, which uses F_{ST} values among three populations to quantify genetic differentiation specific to one of them. Our modern, non-focal populations were both from Sweden. The first was a set of 26 individually sequenced isofemale lines collected in 2011 from Stockholm, Sweden. The second was a set of 40 flies collected in 2014 from Lund,

Sweden. These flies were pool sequenced and collected from the same city as 6 of the 8 museum specimens. Due to careful considerations in our bioinformatic pipeline, pool-seq-driven data artifacts are not expected to yield false PBE outliers. Genome-wide PBE results are shown in Figure 3A, and further discussed below.

We ran a gene ontology (GO) enrichment analysis on PBE outliers (the top 1% of windows) to identify functional categories that may hold adaptive differences between the time points (Figure 3B). Among the most enriched categories was response to organophosphorus, a compound commonly used in insecticides. A second enriched region, acetylcholine metabolism, may also be related to insecticide resistance. The gene that encodes the product that catalyzes the breakdown of acetylcholine, *Ace*, is a well-studied insecticide resistance gene in multiple insects (Mutero *et al.* 1994, Zhu *et al*. 1996, Nabeshima *et al*. 2004).

To estimate local population size of the Sweden sample since the collection of the museum specimens, we simulated allele frequency trajectories of SNPs based on a simple Wright-Fisher model and fit the distribution of observed genome-wide frequency changes to distributions of simulated frequency changes corresponding to differing population sizes. We found that a population size of 5,750 individuals best recapitulated the empirical distribution of SNP frequency differences between the time points.

We used the coalescent simulator *ms* (Hudson 2002) to assign a P value to every empirical window. Empirical windows were divided into 5 bins, and 2.5 million simulations were run for each bin. Briefly, we simulated the species demography from Sprengelmeyer *et. al* (2020). This demography consists of 9 populations sampled throughout Africa and Europe. Because we do not have an estimated demography between the modern Lund and Stockholm samples, we considered the France population in the simulation to be a stand-in for the modern

sequences and assigned a P value to each empirical window F_{ST} value between the museum specimens and modern Lund. We used the estimate of N_e (described above) as the population size of the Sweden population.

Insecticide Resistance as a Likely Target of Selection

Our highest window PBE contained the well-studied insecticide resistance gene *Cyp6g1*. The F_{ST} between the museum specimens and modern Lund samples had a P value of 9.61E-0.07, making this statistically significant even after a genome-wide Bonferroni correction based on the total number of windows. This gene confers resistance to Dichlorodiphenyltrichloroethane (DDT) via the insertion of a transposon upstream of the transcription start site (Daborn *et al*., 2001; Chung 2007). In Chapter 3 of this thesis, we found evidence of genetic differentiation at this region in a North American *D. melanogaster* population. Given that there has been evidence of selection at this locus in *D. melanogaster* populations around the world, it is becoming increasingly clear that this gene is important for offering resistance to insecticides in modern *Drosophila* populations.

A second window in our PBE scan may also potentially be an adaptive target of insecticide resistance. This window had our third highest window PBE value and contained *Choline acetyltransferase* (*ChAT*). The window F_{ST} between museum and modern Lund had a P value of 1.38E-5, which was not quite statistically significant after a Bonferroni correction. Nevertheless, the function of this gene makes it an interesting outlier, and there is a visible SNP signal at this gene (Figure 4A). The product of *ChAT* catalyzes the biosynthesis of the neurotransmitter acetylcholine (*ACh*), and its activity is strongly correlated with *ACh* levels in *Drosophila* (Salvaterra and McCaman 1985). These results are important because the enzyme that catalyzes the breakdown of *ACh* is the product of a well studied insecticide resistance gene,

Ace. Indeed, the window containing *Ace* had a relatively high PBE in our data as well, as it was in the top 1.5% genome-wide. To date, *ChAT* has not been shown in the literature to confer resistance to insecticides in natural populations, but given its relationship to *Ace* and the signal of differentiation we observed between our time points, further exploration may be warranted.

Other Possible Targets of Directional Selection

The second highest window PBE was on chromosome X and contained the gene *corkscrew* (*csw*). SNPs spanning the entire window had elevated PBE (Figure 4B). The product of *csw* contributes to growth regulation (Perkins *et al.* 1996, Johnson-Hamlet and Perkins 2001), life span (Ruzzi *et al.* 2020), and other functions. Mutations in the ortholog in humans, *PTPN11*, cause Noonan Syndrome (Tartaglia *et al.* 2002). We also point out that this window is very close to a region previously identified in a genomic scan for selection in *D. melanogaster* (Beisswanger *et al*. 2006), and the likely target is a pair of tandem paralogs *ph-p* and *ph-d* (Beisswanger and Stephan 2008). The window containing these genes also had elevated PBE.

At least two genes important for nervous system function were outliers in our PBE scan. The first, *beethoven*, has been implicated in male courtship behavior (Eberl *et al.* 1997) and sound perception (Eberl *et al.* 2000). The second, *Ctr9* is essential for development in *D. melanogaster* (Chaturvedi *et al.* 2016) and is a particularly crucial component in developing the nervous system (Bahrampour and Thor 2016). The window containing this gene was a PBE outlier, with a SNP peak in the gene region (Figure 4C). In light of nearby missing data, the peak here may be associated with copy number variation or unmapped SNP variation that differs in frequency between time points.

Enrichment of Low F_{ST} P Values is Observed

As a null hypothesis, we expected our simulations to emulate our empirical F_{ST} between the museum specimens and modern Lund samples. Our window P values, therefore, should have been uniformly distributed between 0 and 1. Instead, we observed a rather stark enrichment of low P values (Figure 6). Based on this observation, we asked how many regions could be removed before the enrichment disappeared. We defined a region starting at a low P value window and extending in each direction until hitting a string of 10 consecutive windows with a P value greater than 0.1. Windows were removed until the bin containing P values between 0 and 0.05 held fewer windows than the bin containing P values between 0.05 and 0.1. In order to ensure accuracy and minimize unaccounted for bias, we utilized two different approaches to removing regions, and they yielded remarkably similar results. In a deterministic approach, we iteratively removed the lowest P value region until no P value enrichment remained. In a random approach, we randomly chose a window whose P value was less than 0.05 and defined a region to remove around that window. In the deterministic approach, 55 regions were removed, accounting for 6.18% of the genome. We ran 1,000 iterations of the random approach. Across these 1,000 runs, an average of 61 regions were removed, accounting for 6.42% of the genome. Thus, the allele frequency shift between time points that we observed in just over 6% of the genome cannot be explained solely by our neutral simulations.

Conclusions

In this study, we analyze genomic data collected from 8 museum specimens whose collection dates to the 1840s. Comparing these samples with modern populations, we have revealed potential targets of recent adaptation, including adaptation of resistance to insecticides. We also show limited evidence that inversions may have been at a lower frequency than modern

populations, giving additional evidence to the hypothesis that inversions are a more recent arrival into modern European populations

Methods

Genomic Sequence Data Collection

Next generation sequencing reads from the modern Stockholm, Sweden sample were downloaded from the NCBI Sequence Read Archive. The individual strains from the museum specimens as well as strains from the downloaded Stockholm, Sweden strains were first trimmed using Trimmomatic (v0.39, Bolger *et al.* 2014). They were then aligned to the *D. melanogaster* reference genome (v5.57) using the same sequencing pipeline described in Lack *et al.* (2016) up to consensus sequence generation. Upon consensus sequence generation, we added an extra filter to require at least 25% of reads at a site to support an alternative allele in order to call the site heterozygous. This step may help protect against false positive heterozygous calls driven by copy number variation and may make comparison between pooled sequencing and individual sequences more accurate.

Because the museum specimens are wild-derived flies and the modern Stockholm samples are from isofemale lines not subject to intentional inbreeding, most of their genomes should be outbred. After consensus sequence generation, we identified regions of heterozygosity using the hidden Markov model published by Corbett-Detig and Nielsen (2017). Such outbred regions yield 2 random allele draws from a population, giving us a maximum sample size of 16 in the museum specimen sample and 52 in the Stockholm sample.

Sequencing reads from the modern Lund, Sweden pooled sample were downloaded from the NCBI sequence read archive. Reads were trimmed using fastp (v0.21.0, Chen *et al.* 2018)

and bases with a quality score <20 were removed with a custom perl script. Reads were mapped to the *D. melanogaster* reference genome (v5.57) using bwa aln v0.5.9 (Li and Durbin 2010), unaligned reads were mapped with Stampy v1.0.20 (Lunter and Goodson 2010).

We then used Picard version v1.79 (http://picard.sourceforge.net/) to sort the alignment by coordinates and remove optical duplicates. Assemblies were improved around InDels using the GATK v3.2 Indel Realigner (McKenna *et al.* 2010; Depristo *et al.* 2011). We used samtools to generate an mpileup file. We called SNPs using the PoolSNP (Kapun *et al.* 2020), requiring a minimum of 20 reads to call a SNP on the X chromosome (because males were sequenced and coverage was lower on the X chromosome) and 40 reads for the 4 autosomal arms. We discarded sites with coverage in the 99th percentile for each chromosome arm to protect against duplications affecting allele frequency estimates. Using scripts from the DrosEu bioinformatic pipeline (Kapun *et al.* 2020), we detected and masked around InDels by using DetectIndels.py and by requiring 20 reads supporting an InDel to call it. We filtered 5 bases on each side of an InDel. After InDel calling, we used FilterPosFromVCF.py to generate a filtered VCF. We used VCF2sync.py to generate sync files used for downstream population genomic analyses.

Principal Components Analysis

We applied principal components analysis (PCA) to each major chromosome arm of the old samples to better determine their relationship to modern populations and to identify any strain with aberrant divergence. We included 4 modern European and North African population samples from Stockholm, Sweden (Mateo *et al.* 2018), Castellana Grotte, Italy (Mateo *et al.* 2018), Leon, France (Lack *et al.* 2016), and Egypt (Lack *et al.* 2016). We used SNPRelate release 3.9 (Zheng *et al.* 2012) for the PCA analysis.

Population Branch Excess

We used the population branch excess (PBE) statistic (Yassin *et al.* 2016), a modified version of the population branch statistic (PBS), to quantify genetic differentiation specific to the museum samples when compared to modern samples. PBE quantifies the degree to which PBS exceeds its predicted value, based on differentiation between the other two populations at a locus and the typical patterns observed at other loci genome-wide.

Gene Ontology Enrichment

The top 1% of PBE quantiles were considered outliers for GO enrichment analysis under the hypothesis that these outliers will be enriched for genuine targets of adaptation. GO enrichment was assessed as previously described in Pool *et al.* (2012). Two or more outlier windows were merged into the same outlier window region if they were separated by no more than 4 nonoutlier windows (to conservatively avoid counting the same selective sweep more than once). Locations of outlier regions were then randomly permuted, while maintaining their lengths, to properly account for the arrangement and lengths of genes in each functional category. Each outlier region was only allowed to vote for a given GO category one time (from both the empirical and permuted outlier regions) to avoid spurious results from clusters of functionally linked paralogs. For each GO term, a raw P value was defined by the proportion of 1,000,000 randomized data sets in which a greater or equal number of outliers from that category was obtained. Then, by comparing across these randomized data sets, the lowest raw P value for each of the data sets was obtained, and a threshold for analysis-wide significance was defined based on a minimum raw P value observed in 5% or fewer randomized data sets.

Table 1

Sequencing results from museum specimens

Name	Location Collected	Year Collected	Number of Reads (Millions)	Mean Depth of Coverage	Median Read Length Post-Trimming	Percent of Genome with Called Sites
Lun1	Lund, Sweden	1933	40	16.7	49	89%
Lun2	Lund, Sweden	1933	48	17.9	44	91%
Lun3	Lund, Sweden	1933	83	37.3	53	93.1%
Lun4	Lund, Sweden	1933	55	21.9	47	91.6%
Lun5	Lund, Sweden	1933	7	2.4	41	46%
Lun6	Lund, Sweden	1933	11.5	4.3	44	71%
Den1	Zealand, Denmark	1859	11	4	43	64%
Ger1	Passau, Germany	1840's (exact date unknown)	56	22.3	47	91.7%

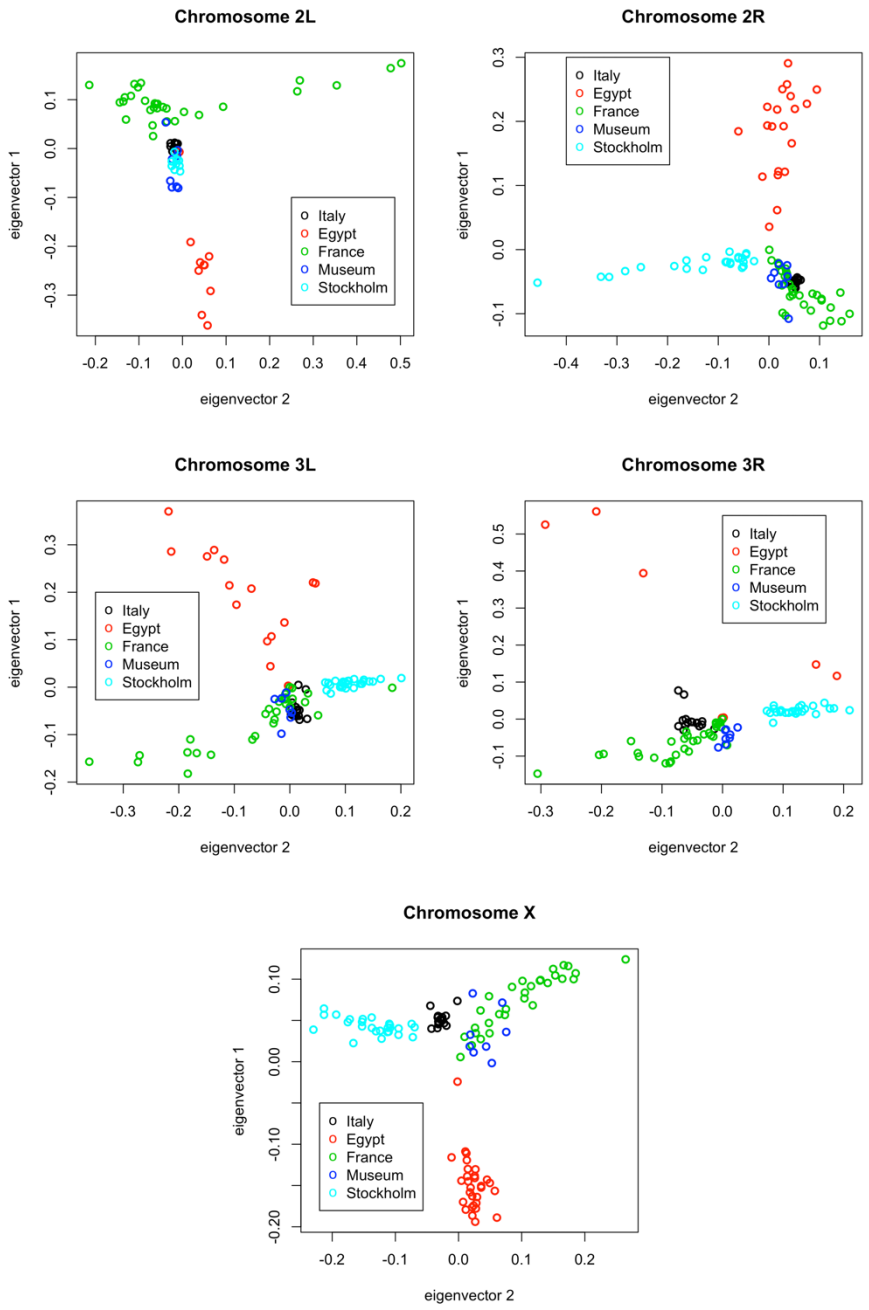


Figure 1: PCA plots of each of the 5 main chromosome arms. Inversion-free strains from 4 additional European and North African populations were used in this analysis.

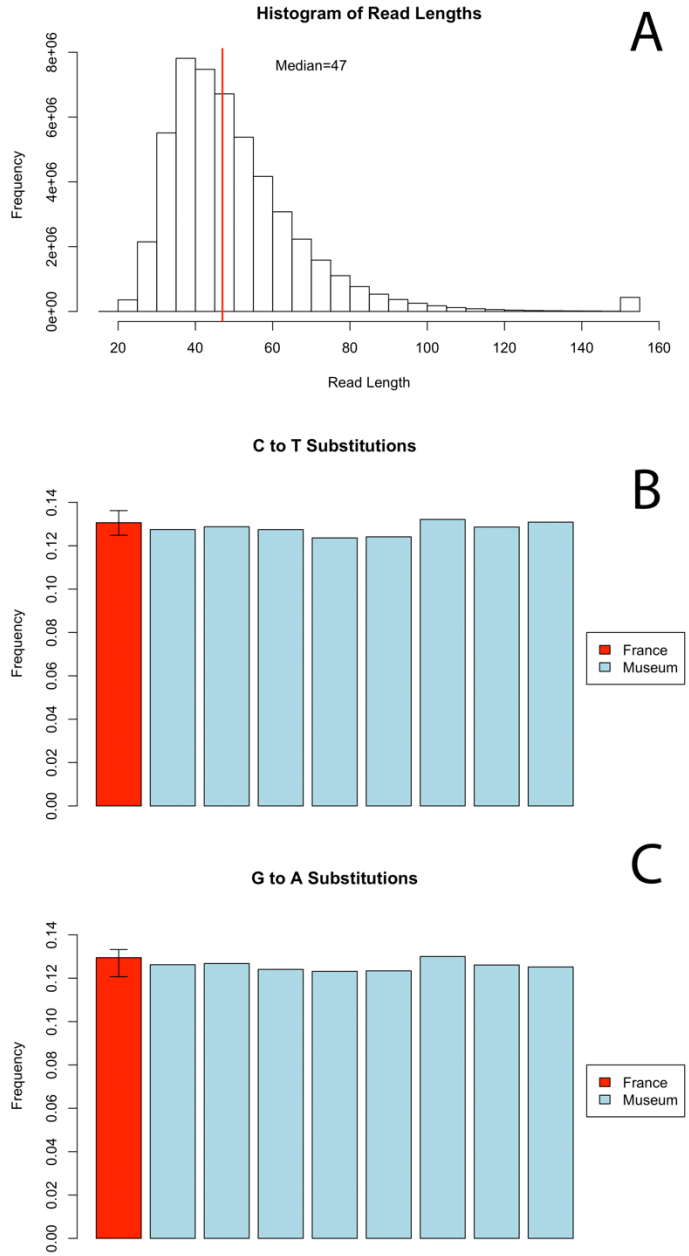
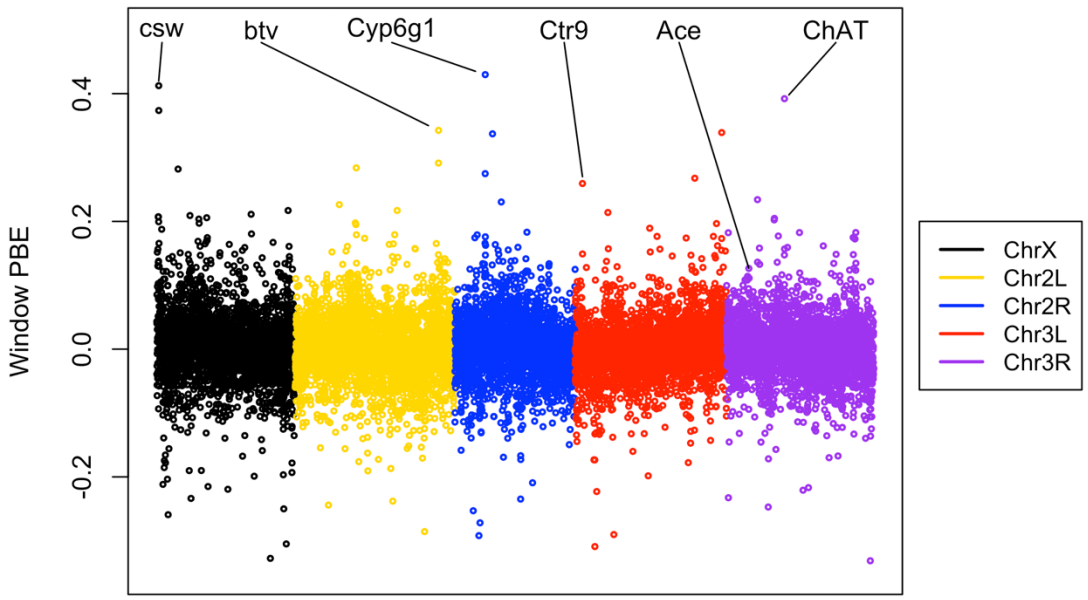


Figure 2: Quality assurance of museum specimen sequencing. In panel A, we show the distribution of read lengths of one sequencing library after trimming. The red vertical line represents the median read length. In panel B, we show C to T substitutions (as a proportion of all substitutions) in each of the museum specimens (blue) and across 81 France strains (red) that went through the same bioinformatics pipeline. The red bar itself represents the average C to T substitution rate across France samples while the error bar represents the minimum and maximum rates. Note that all museum specimens are within the range of the France error bars. Panel C represents G to A substitutions.

A



B

GO Category	Number of Outlier Regions	Raw P Value
endopeptidase regulator activity	7	0.000493506
microtubule binding	7	0.000587413
female meiosis I	2	0.000751249
tubulin binding	7	0.000795205
response to organophosphorus	2	0.000923077
peptidase regulator activity	7	0.000969031
acetylcholine metabolic process	2	0.00236963
acetate ester metabolic process	2	0.00236963
enzyme inhibitor activity	7	0.002407592
regulation of microtubule cytoskeleton organization	4	0.002791209

Figure 3: Population Branch Excess and Gene Ontology Enrichment. In panel A, we show PBE at each window genome-wide. Gene names are discussed in the text. In Panel B, we list the top 10 categories in our GO enrichment analysis.

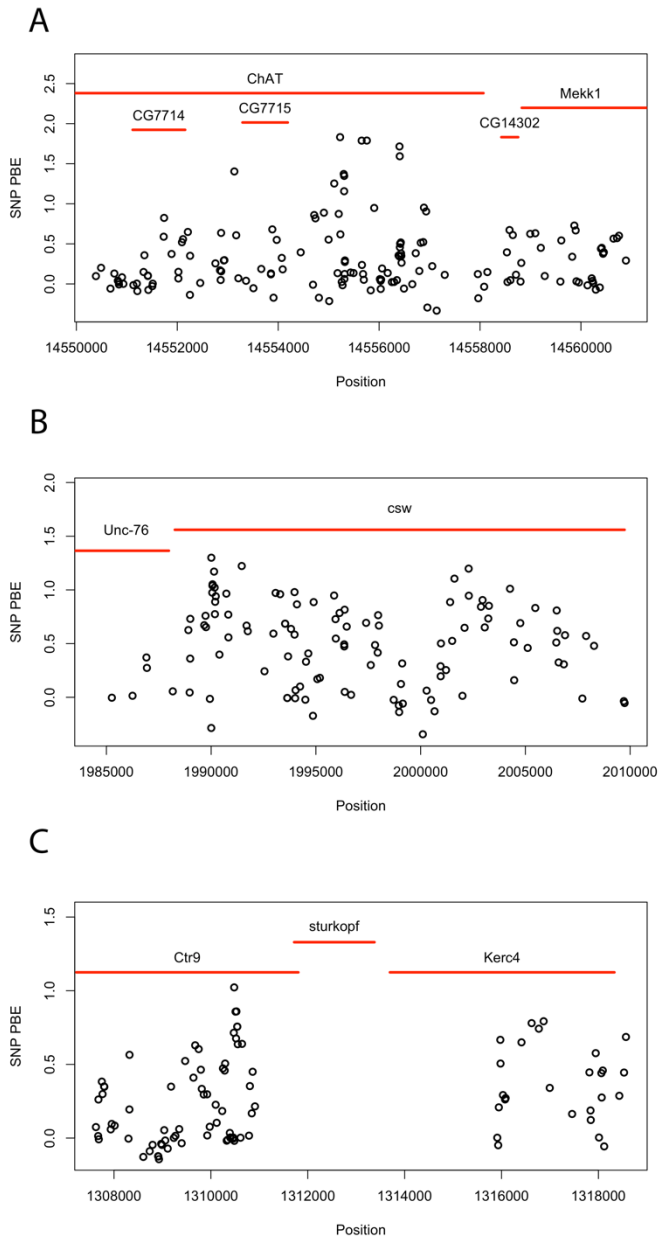


Figure 4: SNP PBE plots. Panels A, B, and C show PBE plots at the SNP level for 3 window outliers.

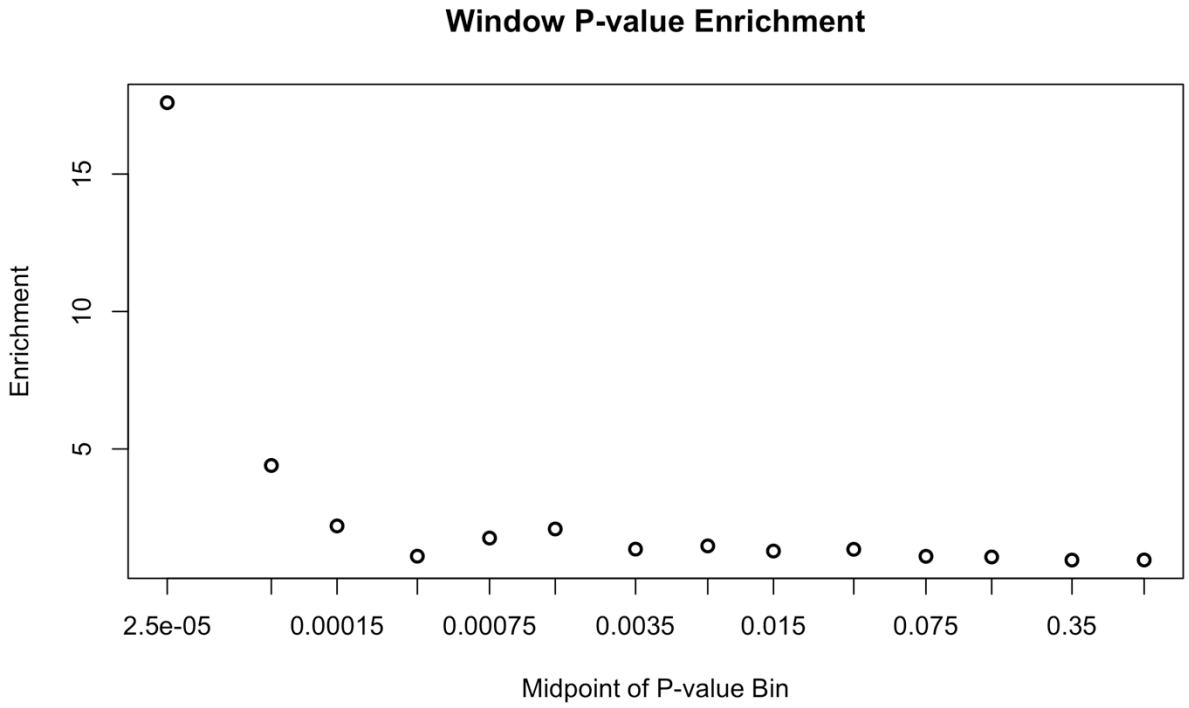


Figure 5: P value enrichment of low window-Fst P values. Here, the Fst is calculated between the modern Lund sample and the museum specimens.

References

Bahrampour S, Thor S. (2016) Ctr9, a key component of the Paf1 Complex, affects proliferation and terminal differentiation in the developing *Drosophila* nervous system. G3 (Bethesda), 6(10):3229-3239.

Beisswanger S, Stephan W, De Lorenzo D. (2006) Evidence for a selective sweep in the wapl region of Drosophila melanogaster. Genetics, *172*(1), 265-274.

Beisswanger S, Stephan W. (2008) Evidence that strong positive selection drives neofunctionalization in the tandemly duplicated polyhomeotic genes in Drosophila. PNAS 105(14): 5447–52

Bolger AM, Lohse M, Usadel B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics, 30(15):2114-20.

Chaturvedi D, Inaba M, Scoggin S, Buszczak M. (2016) *Drosophila* CG2469 encodes a homolog of human CTR9 and is essential for development. G3 (Bethesda), 6(12):3849-3857.

Chen S, Zhou Y, Chen Y, Gu J. (2018) fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics, 34(17):i884-i890.

Chung H, Bogwitz MR, McCart C, Andrianopoulos A, Ffrench-Constant RH, Batterham P, Daborn PJ. (2007) Cis-regulatory elements in the Accord retrotransposon result in tissue-specific expression of the *Drosophila melanogaster* insecticide resistance gene Cyp6g1. Genetics, 175(3):1071-7.

Cooper A, Mourer-Chauviré C, Chambers GK, von Haeseler A, Wilson AC, Pääbo S. (1992) Independent origins of New Zealand moas and kiwis. Proceedings of the National Academy of Sciences of the United States of America, 89(18):8741-4.

Corbett-Detig RB, Hartl DL. (2012) Population genomics of inversion polymorphisms in *Drosophila melanogaster.* PLoS Genetics, 8(12):e1003056.

Corbett-Detig R, Nielsen R. (2017) A hidden Markov model approach for simultaneously estimating local ancestry and admixture time using next generation sequence data in samples of arbitrary ploidy. PLoS Genetics, 13(1):e1006529.

Daborn P, Boundy S, Yen J, Pittendrigh B, ffrench-Constant R. (2001) DDT resistance in *Drosophila* correlates with Cyp6g1 over-expression and confers cross-resistance to the neonicotinoid imidacloprid. Molecular Genetics & Genomics, 266(4):556-63.

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature Genetics, 43(5):491-8.

Eberl DF, Duyk GM, Perrimon N. (1997) A genetic screen for mutations that disrupt an auditory response in *Drosophila melanogaster*. Proceedings of the National Academy of Sciences of the United States of America, 94(26):14837-42.

Eberl DF, Hardy RW, Kernan MJ. (2020) Genetically similar transduction mechanisms for touch and hearing in *Drosophila*. The Journal of Neuroscience, 20(16):5981-8.

Fisher RA and Ford EB. (1947) The spread of a gene in natural conditions in a colony of the moth *Panaxia dominula L.* Heredity, 1:143–174.

Flagstad Ø, Walker CW, Vilà C, Sundqvist AK, Fernholm B, Hufthammer AK, Wiig Ø, Koyola I, Ellegren H. (2003) Two centuries of the Scandinavian wolf population: patterns of genetic variability and migration during an era of dramatic decline. Molecular Ecology, 12(4):869-80.

Gilbert MT, Binladen J, Miller W, Wiuf C, Willerslev E, Poinar H, Carlson JE, Leebens-Mack JH, Schuster SC. (2007) Recharacterization of ancient DNA miscoding lesions: insights in the era of sequencing-by-synthesis. Nucleic Acids Research, 35(1):1-10.

Hudson RR. (2002) Generating samples under a Wright-Fisher neutral model of genetic variation. Bioinformatics, 18(2):337-8.

Johnson Hamlet MR, Perkins LA. (2001) Analysis of corkscrew signaling in the *Drosophila* epidermal growth factor receptor pathway during myogenesis. Genetics, 159(3):1073-87.

Kapun M, Fabian DK, Goudet J, Flatt T. (2016) Genomic evidence for adaptive inversion clines in *Drosophila melanogaster*. Molecular Biology and Evolution, 33(5):1317-36.

Kapun M, Barrón MG, Staubach F, Obbard DJ, Wiberg RAW, Vieira J, et al. (2020) Genomic analysis of European *Drosophila melanogaster* populations reveals longitudinal structure, continent-wide selection, and previously unknown DNA viruses. Molecular Biology and Evolution, 37(9):2661-2678.

Keller A. (2007) Drosophila melanogaster's history as a human commensal. Current Biology, 17: R77–81.

Lack JB, Lange JD, Tang AD, Corbett-Detig RB, Pool JE. (2016) A thousand fly genomes: an expanded *Drosophila* genome nexus. Molecular Biology and Evolution, 33(12):3308-3313.

Lunter G, Goodson M. (2011) Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. Genome Research, 21(6):936-9.

Mateo L, Rech GE, González J. (2018) Genome-wide patterns of local adaptation in Western European *Drosophila melanogaster* natural populations. Scientific Reports, 8(1):16143.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research, 20(9):1297-303.

Miller CR, Waits LP. (2003) The history of effective population size and genetic diversity in the Yellowstone grizzly (Ursus arctos): implications for conservation. Proceedings of the National Academy of Sciences of the United States of America, 100(7):4334-9.

Mutero A, Pralavorio M, Bride JM, Fournier D. (1994) Resistance-associated point mutations in insecticide-insensitive acetylcholinesterase. Proceedings of the National Academy of Sciences of the United States of America, 91(13):5922-6.

Nabeshima T, Mori A, Kozaki T, Iwata Y, Hidoh O, et al. (2004) An amino acid substitution attributable to insecticide-insensitivity of acetylcholinesterase in a Japanese encephalitis vector mosquito, Culex tritaeniorhynchus. Biochem Biophys Res Commun 313: 794–801.

Nyström V, Angerbjörn A, Dalén L. (2006) Genetic consequences of a demographic bottleneck in the Scandinavian arctic fox. Oikos, 114(1):84-94.

Parejo M, Wragg D, Henriques D, Charrière JD, Estonba A. (2020) Digging into the genomic past of Swiss honey bees by whole-genome sequencing museum specimens. Genome Biology and Evolution, 12(12):2535-2551.

Perkins LA, Johnson MR, Melnick MB, Perrimon N. (1996) The nonreceptor protein tyrosine phosphatase corkscrew functions in multiple receptor tyrosine kinase pathways in *Drosophila*. Developmental Biology, 180(1):63-81.

Pool JE, Corbett-Detig RB, Sugino RP, Stevens KA, Cardeno CM, Crepeau MW, et al. (2012) Population genomics of sub-saharan *Drosophila melanogaster*: African diversity and non-African admixture. PLoS Genetics, 8(12):e1003080.

Ruzzi LR, Schilman PE, San Martin A, Lew SE, Gelb BD, Pagani MR. (2020) The phosphatase CSW controls life span by insulin signaling and metabolism throughout adult life in *Drosophila*. Frontiers in Genetics, 11:364.

Saltonstall K. (2002) Cryptic invasion by a non-native genotype of the common reed, *Phragmites australis*, into North America. Proceedings of the National Academy of Sciences of the United States of America, 99(4):2445-9.

Salvaterra PM, McCaman RE. (1985) Choline acetyltransferase and acetylcholine levels in *Drosophila melanogaster*: a study using two temperature-sensitive mutants. Journal of Neuroscience, 5(4):903-10.

Sprengelmeyer QD, Mansourian S, Lange JD, Matute DR, Cooper BS, Jirle EV, Stensmyr MC, Pool JE (2020). Recurrent collection of *Drosophila melanogaster* from wild African

environments and genomic insights into species history. Molecular Biology and Evolution. 37(3):627–638.

Stiller M, Green RE, Ronan M, Simons JF, Du L, He W, et al. (2006) Patterns of nucleotide misincorporations during enzymatic amplification and direct large-scale sequencing of ancient DNA. Proceedings of the National Academy of Sciences of the United States of America, 103(37):13578-84.

Tartaglia M, Kalidas K, Shaw A, Song X, Musat DL, van der Burgt I, et al. (2002) PTPN11 mutations in Noonan syndrome: molecular spectrum, genotype-phenotype correlation, and phenotypic heterogeneity. American Journal of Human Genetics, 70(6):1555-63.

Yassin A, Debat V, Bastide H, Gidaszewski N, David JR, Pool JE. (2016) Recurrent specialization on a toxic fruit in an island *Drosophila* population. Proceedings of the National Academy of Sciences of the United States of America, 113(17):4771-6.

Younis HM, Abo-El-Saad MM, Abdel-Razik RK, Abo-Seda SA. (2002) Resolving the DDT target protein in insects as a subunit of the ATP synthase. Biotechnology and Applied Biochemistry, 35(1):9-17.

Younis HM, Serrano R, Abdel-Razik RK, Rydström J. (2011) The insecticide DDT targets the OSCP and subunit D of the *Apis mellifera* ATP synthase. Journal of Bioenergetics and Biomembranes, 43(5):457-63.

Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. (2012) A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics, 28(24):3326-8.

Zhu KY, Lee SH, Clark JM (1996) A Point Mutation of Acetylcholinesterase Associated with Azinphosmethyl Resistance and Reduced Fitness in Colorado Potato Beetle. Pestic Biochem Physiol 55: 100–108.

Chapter 5: Discussion of thesis work

Understanding how natural selection works at the molecular level remains a fundamental area of research in biology. In this thesis, we have begun to answer fundamental questions about how natural selection works in nature. We have presented a novel way to measure genetic variation and, using simulation, have shown that this statistic has strong power to detect instances of local adaptation. This statistic relies on a signal of reduced genetic variation that is common after a selective sweep. Because neutral demographic events can also affect genetic variation, we tested whether common demographic inference methods are biased in the presence of recurrent natural selection. We have shown that though these methods are indeed biased by selection, we have argued for their utility even in populous species such as Drosophila melanogaster. Using results from these chapters, we have studied natural selection in two empirical populations of Drosophila melanogaster, and utilized time-series genomic data to help identify instances of adaptation.

During the introduction of this thesis, we posed unanswered questions in the field of population genetics. Here, we briefly expand on these questions and attempt to contextualize our results in a broader picture and offer future directions for the next generation of scientists.

How effectively can we separate selection from non-neutral forces?

Many methods exist to identify targets of recent adaptation. The most promising of these statistics may be methods that rely on patterns of linkage disequilibrium, such as haplotype methods. The novel statistic introduced in chapter 1 of this thesis utilizes shared haplotype identity to separate neutral from non-neutral forces. We have shown that our method can detect a wide variety of sweeps, and has particularly high power to detect soft sweeps in data that resembles human genomic data. Other published haplotype methods exist that have been shown

to be powerful to detect soft sweeps as well, including the *H12* statistic (Garud *et al*. 2015). This method, which measures the frequencies of the two most common haplotypes in a sample, has been used in *D. melanogaster* to identify instances of soft sweeps in the DGRP population (Garud *et al*. 2015). Because of its reliance on the two most common haplotypes, however, its utility in scenarios of softer sweeps may deteriorate quicker than χ. A close inspection comparing χ to the more recent advances in haplotype methods may be warranted future study.

Comparing the genetic variation of closely related populations helps to control for demography and can better elucidate regions of natural selection. Closely related populations are often still separated by thousands of years of evolution, however. This thesis has utilized temporal sampling on very short time scales to identify targets of recent adaptation. Both studies of short term evolution in this thesis are the first to study change in genetic variation over decades in *D. melanogaster*, offering fundamental insights into the evolution of this important model organism. Typical genomic scans for selection utilizing a single time point often identify outliers, but it is rare that these outliers are statistically significant. By utilizing such short time scales, we can better control for neutral events. We have shown that we are able to identify targets of selection that are statistically significant even after a conservative genome-wide multiple testing correction.

What types of genes are targeted by selection? And how much evidence for selection does a population contain?

These questions were explored thoroughly in chapters 2-4 of this thesis. In chapter 2, our simulations of recurrent hitchhiking entailed two distinct models: one of rare and strong sweeps and one of weak and common sweeps. Perhaps the most surprising result from these simulations of recurrent hitchhiking is that, even in the common/weak model of selection, we observed a

mild amount of Hill-Robertson interference (HRI). Previous estimates of HRI in *D. melanogaster* have relied on the assumption that there is *no* interference in regions of high recombination (Castellano *et al*. 2016). Our simulations entailed high recombination, and the observation of HRI suggests that more sweeps may be lost than previously estimated. How often sweeps go to fixation remains an unresolved question in the field, and our HRI results may help to answer this.

Chapters 3 and 4 of this thesis explore what types of genes have been targeted by selection in two empirical populations of *D. melanogaster*. We have shown that, across both datasets, there is evidence of natural selection at well-studied insecticide resistance loci. A particularly striking result was observed in our population of Rhode Island flies, where selection of a structural variant may have been driven by particularly strong selection over an 8 year period. Though *D. melanogaster* is not considered a crop pest, our Approximate Bayesian Computation analysis of this region may help researchers model selection of other populous insect species including crop pests.

To date, we only have 8 genomic strains from our museum specimens. We anticipate 16 more genomic samples in the near future. There is ample opportunity for future direction of this study. An examination of geographic differentiation could yield potential interesting results. The PCA results suggest that the museum specimens may be more genetically related to modern southern European populations than modern northern European populations. If we understand where favored alleles come from geographically, we can better illuminate models of selection and theorize why selection occurred.

Museum specimens may provide an untapped source of genetic information for evolutionary biologists. Advancement in Next Generation Sequencing technologies is now

allowing researchers to extract ample genetic data from museum specimens, helping to inform important evolutionary changes and species migration patterns. These museum collections allow researchers to ask questions that cannot be answered utilizing a single time point of a modern sample. For instance, demographic models inferred from solely modern samples can be refined, and hypotheses about a species' response to climate change, or other environmental shifts, can be directly tested. Moreover, museum samples that represent a species' first arrival to a new environment allow researchers to theorize about adaptations that have occurred over time.

Works Cited

Castellano D, Coronado-Zamora M, Campos JL, Barbadilla A, Eyre-Walker A.. 2016. Adaptive evolution is substantially impeded by Hill–Robertson interference in *Drosophila*. Mol Biol Evol. 33:442–455.

Garud NR, Messer PW, Buzbas EO, Petrov DA. (2015) Recent selective sweeps in North American *Drosophila melanogaster* show signatures of soft sweeps. PLoS Genet. 11: e1005004.

Karasov T, Messer PW, Petrov DA. (2010) Evidence that adaptation in Drosophila is not limited by mutation at single sites. PLoS Genetics, 6(6):e1000924.

Appendix 1: Supplemental Figures for Chapter 1

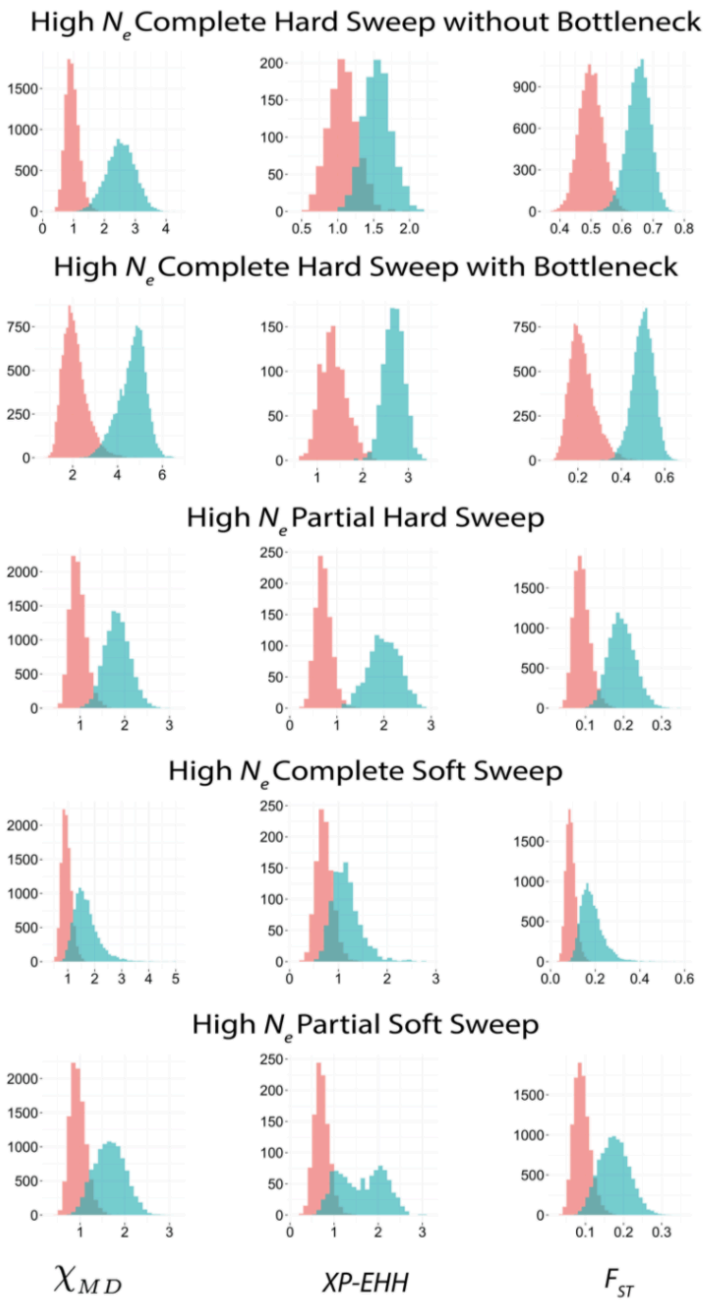


Figure S1. Depicted here are distributions of all three statistics for various scenarios of high Ne scenarios. In the non-bottleneck hard sweep case, the populations split at 0.5 coalescent time units in the past and selection began at 0.2 time unites in the past. In the bottleneck scenario, the bottleneck strength was 0.05. The ending frequency of the partial hard sweep is 0.5. The starting frequency of the complete soft sweep was 0.001. The starting allele frequency was 0.0001 and ended at 0.5 for the partial soft sweep case.

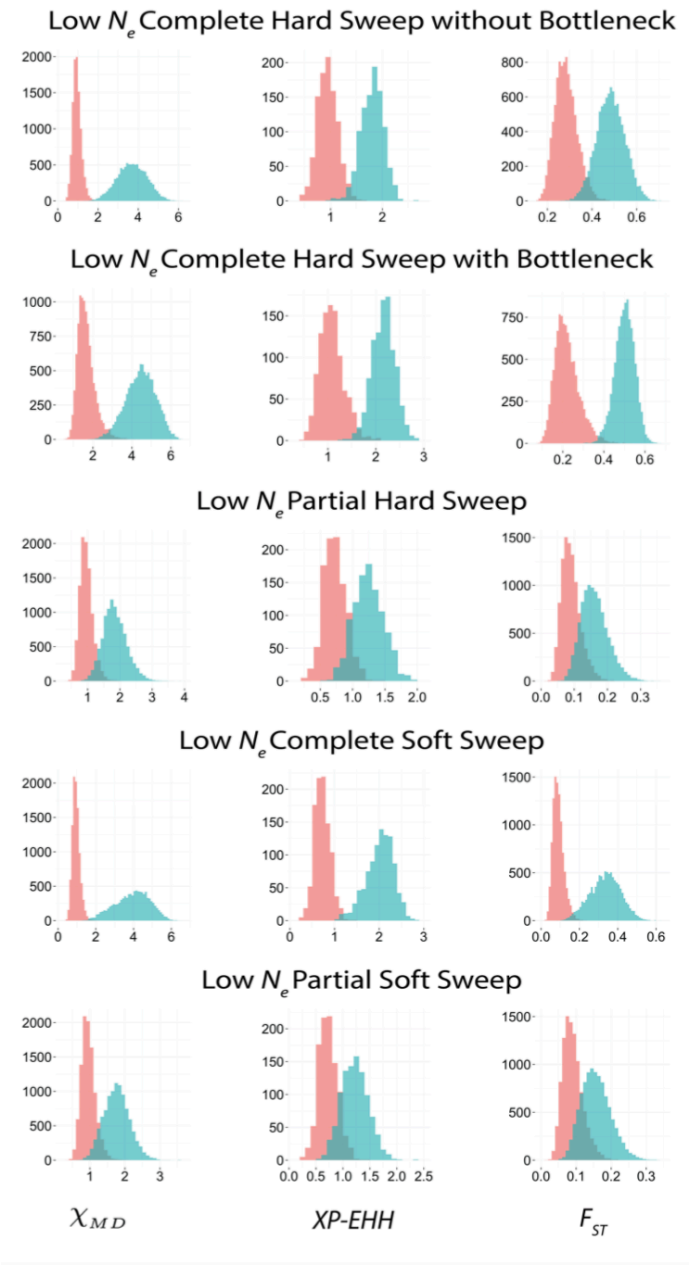


Figure S2. Depicted here are distributions of all three statistics for various scenarios of low *Ne* scenarios. In the non-bottleneck hard sweep case, the populations split at 0.2 coalescent time units in the past and selection began immediately. In the bottleneck scenario, the bottleneck strength was 0.1. The ending frequency of the partial hard sweep is 0.5. The starting frequency of the complete soft sweep was 0.001. The starting allele frequency was 0.001 and ended at 0.5 for the partial soft sweep case.

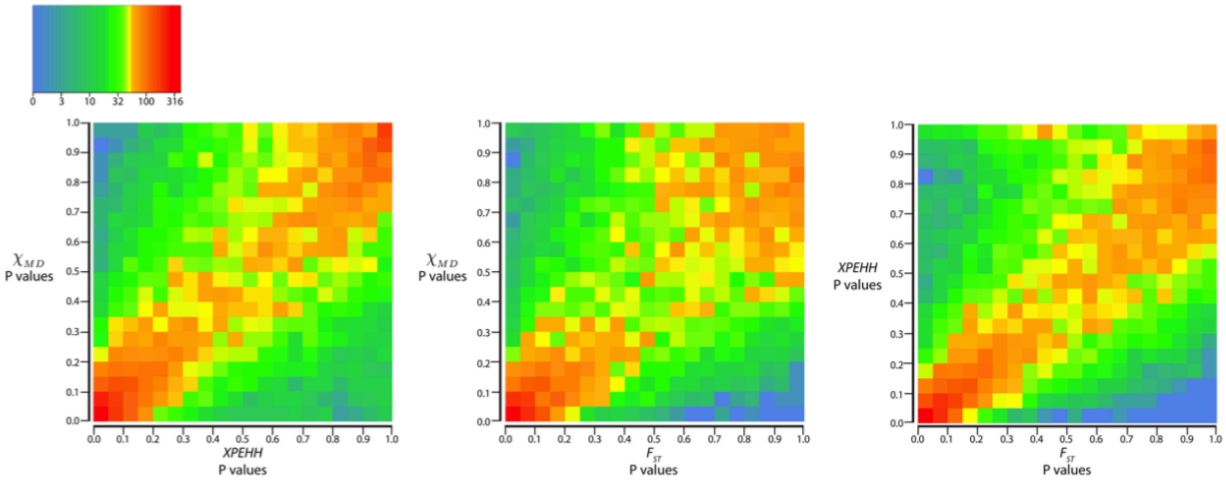


Figure S3. Based on evaluating the three between-population statistics in published *Drosophila* genomes, this figure shows the correlation between their empirical *P* values across all autosomal windows. Each cell represents the number of autosomal genomic windows whose statistic *P* values are within the corresponding range of the cell.

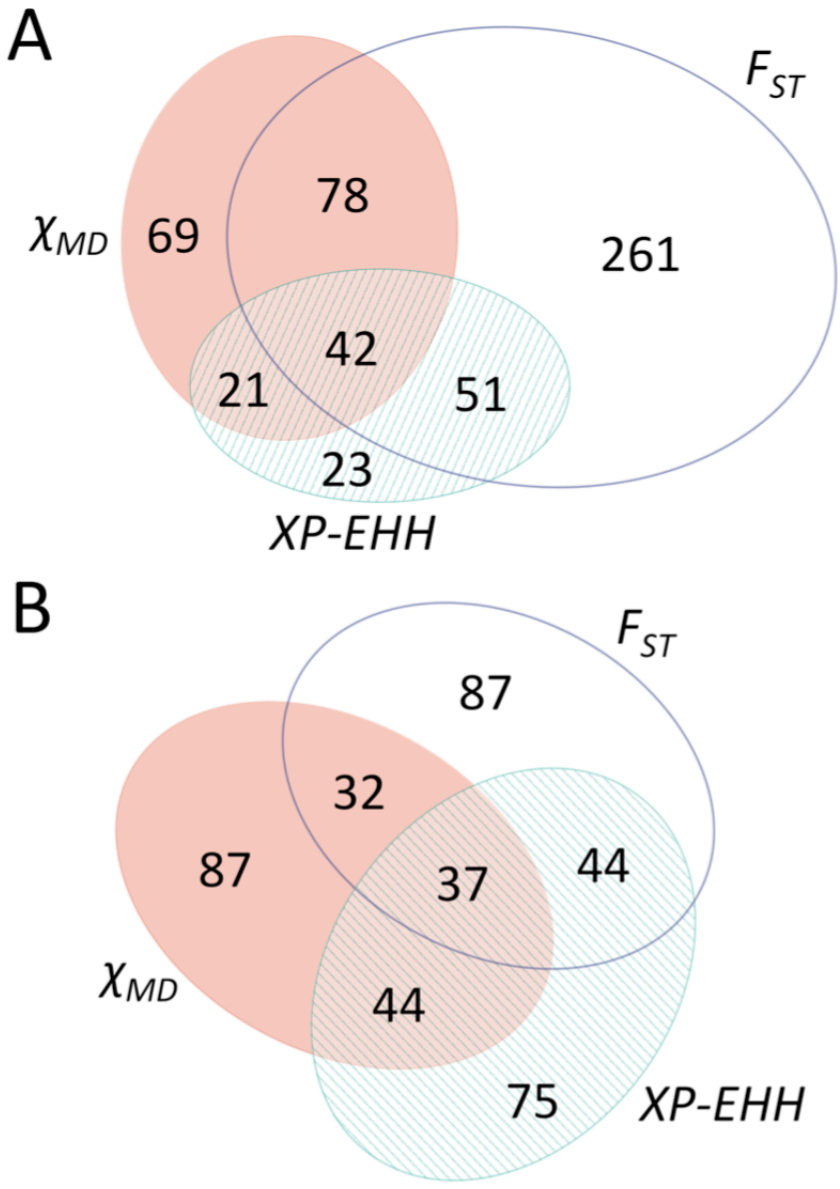


Figure S4. The gene ontology categories identified by each statistic show substantial overlap. This analysis focuses on the 1,988 biological process GO categories represented in at least ten different genomic windows. Part A depicts the number of such categories with a raw permutation *P* value below 0.05 for each statistic. Part B includes the lowest 200 *P* values for each statistic.

Appendix 2: Supplemental Figures for Chapter 2

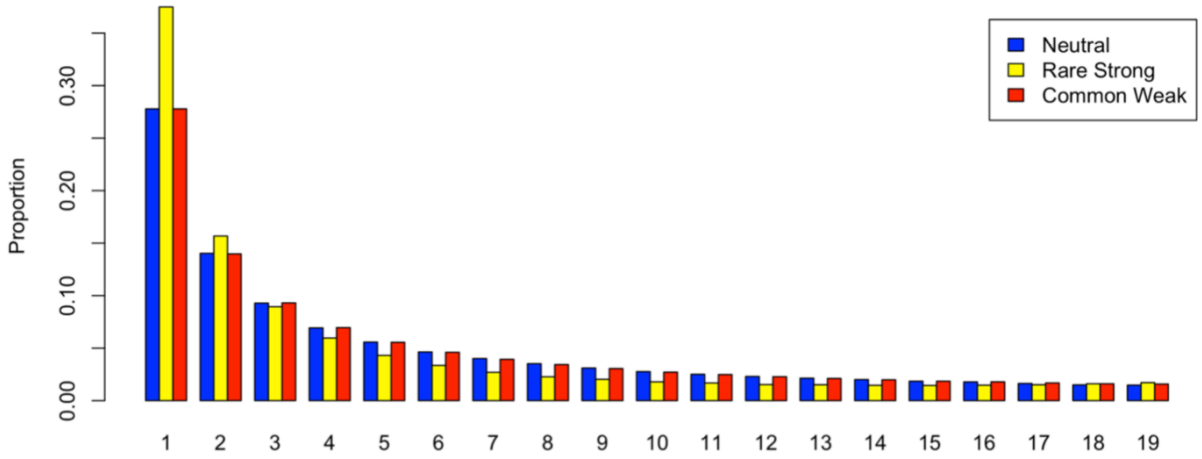


Figure S1: Site frequency spectra for simulations of neutral and RHH models are shown.