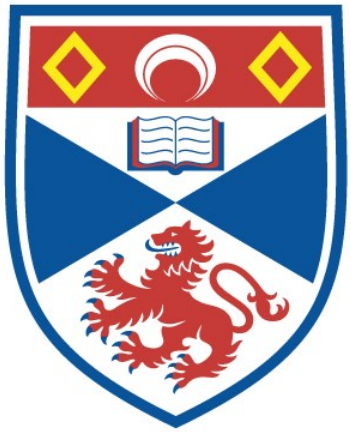
SHORT-TERM ADAPTATION IN THE LAB: AN ANALYSIS OF EVOLUTIONARY TRAJECTORIES USING TIME-SERIES DATA

Carolina de Castro Barbosa Rodrigues Barata

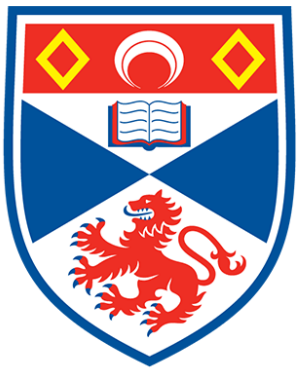
A Thesis Submitted for the Degree of PhD
at the
University of St Andrews



2022

Short-term adaptation in the lab: an analysis of evolutionary trajectories using time-series data

Carolina de Castro Barbosa Rodrigues Barata



University of
St Andrews

This thesis is submitted in partial fulfilment for the degree of

Doctor of Philosophy (PhD)

at the University of St Andrews

May, 2022

Abstract

Experimental evolution has led to remarkable discoveries on the evolution of adaptive phenotypes. However, an understanding of the genetic basis of adaptation is only just starting to emerge, especially in eukaryotes. My thesis aims to take full advantage of genomic time-series data by examining Evolve & Resequence (E&R) experiments. This approach typically monitors the genomes of lab populations over the course of several generations.

Firstly, I introduce Bait-ER: a fully Bayesian approach to investigate allele frequency trajectories while testing for selection. It implements a Moran model to estimate selection paramaters at each biallelic site. Bait-ER models varying coverage explicitly to account for sampling error. My method is robust even when studying small populations, and it proved accurate in both real and simulated genomic datasets.

Secondly, I investigate the response to sexual selection in lab populations of *Drosophila pseudoobscura* using an E&R approach. These flies were subject to an altered mating regime where the intensity of sexual selection was either relaxed in monogamous populations - M - or intensified in polyandrous populations - E. I resequenced the genomes of each population at five separate time points. I estimated diversity and the effective population size (N_e) across the genome, and found that overall estimates of N_e match neutral expectations. However, N_e was lower at the start of the experiment, especially on the X chromosome in E populations. This indicates that sexual selection is strongest at the start in E lines, and that the X is responsible for early adaptation. Lastly, I performed a genome scan using Bait-ER to detect any potential targets of selection on the full genomic time-series. E lines show a stronger signal of selection where the X dominates, but the 3$^{\text{rd}}$ chromosome also seems to be a hotspot.

In summary, I found clear signals of adaptation in the genome of lab populations at several levels. This is in spite of strong genetic drift acting to slow down the adaptive process.

Declarations

Candidate's declaration

I, Carolina de Castro Barbosa Rodrigues Barata, do hereby certify that this thesis, submitted for the degree of PhD, which is approximately 29,000 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for any degree. I confirm that any appendices included in my thesis contain only material permitted by the 'Assessment of Postgraduate Research Students' policy. I was admitted as a research student at the University of St Andrews in September 2017. I received funding from an organisation or institution and have acknowledged the funder(s) in the full text of my thesis.

12.May.2022

Date

Signature of candidate

Supervisor's declaration

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree. I confirm that any appendices included in the thesis contain only material permitted by the 'Assessment of Postgraduate Research Students' policy.

12.May.2022

Date Signature of supervisor

Permission for publication

In submitting this thesis to the University of St Andrews we understand that we are giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. We also understand, unless exempt by an award of an embargo as requested below, that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that this thesis will be electronically accessible for personal or research use and that the library has the right to migrate this thesis into new electronic forms as required to ensure continued access to the thesis. I, Carolina de Castro Barbosa Rodrigues Barata, confirm that my thesis does not contain any third-party material that requires copyright clearance.

The following is an agreed request by candidate and supervisor regarding the publication of this thesis: We request an embargo on all of print and electronic copy for a period of 1 year on the grounds that the publishing of this thesis would preclude future publication of any research chapters as separate journal articles. We agree to the title and abstract of this thesis being published.

12.May.2022

Date Signature of candidate

12.May.2022

Date Signature of supervisor

Underpinning Research Data or Digital Outputs

Candidate's declaration

I, Carolina De Castro Barbosa Rodrigues Barata, hereby certify that no requirements to deposit original research data or digital outputs apply to this thesis and that, where appropriate, secondary data used have been referenced in the full text of my thesis.

12.May.2022

Date Signature of candidate

Authorship statement

Chapter 1 This chapter was written by me with some general advice from my supervisor.

Chapter 2 This chapter is a research article co-authored by myself, Dr. Rui Borges and Dr. Carolin Kosiol. My contribution was to test the software on simulated data and apply it to real data, as well as to lead the writing and revision of the manuscript. Dr. Borges was responsible for developing the initial statistical framework and software. All authors contributed with suggestions to the final piece of writing.

Chapter 3 This chapter is a collaborative project with Professor Michael Ritchie, Professor Rhonda Snook and Dr. Carolin Kosiol who have all contributed to conceptualising the data analysis and manuscript editing. I have extracted the DNA samples in the laboratory from frozen fly stocks to create a genomic time-series. I have parsed and performed all further analyses on the resequencing data. Finally, I led the manuscript writing and submission processes. This chapter is currently under revision in *Genome Biology and Evolution*.

Chapter 4 This final chapter is my own work with helpful suggestions from my supervisor.

Acknowledgements

I would like to start by thanking Dr. Carolin Kosiol for supervising me throughout these four years. I also want to thank Professor Mike Ritchie and Professor Rhonda Snook for welcoming me into the *Drosophila pseudoobscura* project.

I am thankful for having collaborated with Dr. Rui Borges and for his endless enthusiasm about anything science-related. Also, thank you to Dr. Axel Wiberg and Dr. Paris Veltsos for helpful discussions on the genomic data.

Thanks also go to Tanya Sneddon and Lianne Baker whose support with DNA extractions and admin-related struggles made my time in the molecular evolution lab all the more enjoyable.

Thank you to the friendly CBD community most of whom became an instant family. A special thank-you is due to everyone in the Theory Lab - especially to Anna, Darren, Gonçalo, João, Kerstin, Lynette, Tom and Vivienne. Room 22 was home to me, not because of the occasional nap in the one comfy chair, but because of my good friends who turned it into a home. I am thankful to my friends Claudia, Edith, Isaac, Jack, Karina, Leeban, Mariana, Petri, Raquel and Vanessa who shared countless adventures with me. These are now memories that I shall always cherish.

I would also like to thank those who were there from the start - Pedro, Joana and Daniela, and the Poogens. This journey would not have been the same without your support.

I am thankful to my partner, Fiachna. Thank you for bringing me so much happiness amongst all the chaos. All days are better when you are there.

E por fim, o meu obrigada à avó Conceição, ao pai e à mãe, que me ensinaram a ver o mundo com um olhar sempre curioso. O saber é melhor compartilhado tal como o amor deve ser. Urgentemente.

Funding

This work was supported by the Vienna Science and Technology Fund (WWTF) through project MA16-064 and the School of Biology at the University of St Andrews that funded my studentship. DNA extraction and sequencing (as described in chapter 3) was supported by a Carnegie Trust Research Incentive Grant (RIG007474).

Contents

Chapter 1

Introduction

1.1 The experimental evolution paradigm

Geneticists have been concerned with describing evolutionary processes for over a century. Ever since the DNA molecule was described as being responsible for inheritance, Darwinian selection was put into context over tens if not thousands of generations. Strikingly, we have since been able to quantify how much mutation, selection, migration and drift allow lineages to diverge and adapt. This is due to the extensive work of population geneticists, who have created mathematical models to describe evolutionary phenomena across generations; and experimentalists, who have thoroughly conducted numerous studies on natural and laboratory populations.

It was the bout of high-throughput next generation sequencing during this last decade that has allowed genes and genomes to be thoroughly examined. However, it was only recently that genome sequencing costs decreased sharply, which has removed certain prohibitive costs from laboratory budgets. This has led to an increasing availability of whole-genome datasets from population samples. Thus, we can investigate patterns of diversity throughout the genome, and, from those, pick out characteristic signatures of selection and demography. This data availability offers an invaluable opportunity to couple population genetics theory with novel statistical approaches to shed light on the genomics of adaptation.

Over the years, experimental populations have been set up in the lab and studied for many generations. This has been an alternative to field studies where it might

be difficult to control any, or all, environmental variables. Evolution studies have imposed countless selective pressures on experimental populations, ranging from varying concentrations of micronutrients in the medium, such as iron (e.g. Figueiredo et al., 2021), cadmium, nickel or zinc (e.g. Gorter et al., 2017), to selection for antibiotic resistance (e.g. Toprak et al., 2012). Other studies in insects focused on selecting for increased body size (e.g. Reeve et al., 2000), starvation resistance (e.g. Chippindale et al., 1996), novel temperature regimes (see Maynard Smith, 1956 for an early instance and Barghi et al., 2019 for a more recent example) or parasitoid resistance (e.g. Jalvingh et al., 2014). This allows one to investigate several categories of phenotypes, such as resource allocation, immunity responses and some social traits. It might even be possible to assess population fitness using proxies like offspring quality and quantity (e.g. Crudgington et al., 2009; Firman and Simmons, 2010). These evolution experiments have investigated not only phenotypic changes but also changes at the genotype level. There are several examples of studies that have characterised such changes both at functional loci and other molecular markers (Teotónio et al., 2009; Turner et al., 2011; Burke et al., 2010; Orozco-terWengel et al., 2012).

The power of experimental evolution stems from the ability to compare the divergence between experimental treatments to the variation among replicate populations (Kawecki et al., 2012). Combining high-throughput sequencing technologies with sequencing pools of individuals at once - pool-seq - has reduced lab costs even further. This has contributed to many successful Evolve & Resequence – E&R – experiments (e.g. Teotónio et al., 2009; Burke et al., 2010; Turner et al., 2011; Orozco-terWengel et al., 2012). Pool-seq is a cost-effective way of obtaining accurate population allele frequencies (Anand et al., 2016; Schlötterer et al., 2014, 2015) such that it has become quite commonplace. The field of experimental evolution has transformed substantially with these technical advancements, and their importance has been discussed for over a decade (Futschik and Schlötterer, 2010; Phillips and Burke, 2021). This has led us to the brink of understanding the complexities of adaptation at a genomic level.

The genetic basis of adaptation in the lab varies substantially depending on what sort of organism one is concerned with. Organisms vary greatly in complexity namely in

terms of how they reproduce and propagate which can result in very distinct modes of adaptation (reviewed in Long et al., 2015). At a genome-wide scale, adaptation in asexual systems typically occurs from new mutations via clonal interference (see Barrick et al., 2009 for an example in *Escherichia coli* and (Lang et al., 2013) for an example in *Saccharomyces cerevisiae*). When there is no recombination between genetic backgrounds, multiple beneficial alleles arise in a population leading to competition among clones (Muller, 1932). This will most likely lead to the fixation of a single such clone. Clonal interference thus causes diversity to be eliminated until other mutations arise at different loci. Alternatively, in sexual systems, adaptation from new mutations in large populations can result in the typical classical selective sweep signature. This genomic pattern is characterised by an advantageous mutation that arises in the population and sweeps to fixation along with linked neutral variants. After a few generations of stochastic frequency shifts, the allele is then picked up by selection and starts behaving deterministically until it reaches fixation (Maynard Smith and Haigh, 1974). The hitchhiking phenomenon leaves a marked signature of reduced genetic diversity around the selected site (Stephan et al., 1992). The characteristic trough in diversity is affected by the strength of selection. This trough gradually flattens as mutation generates new diversity and recombination brings mutations on different backgrounds together. The selective sweep signature is therefore a hallmark of the adaptive process in large populations. By contrast, the mode of adaptation in small populations of recombining organisms differs substantially the sweep paradigm. Studies in outcrossing yeast (Burke et al., 2014), insects, such as *Drosophila* (Hoedjes et al., 2019), and plants (Scarcelli and Kover, 2009) have shown that these rely on standing genetic variation to respond to selection. The genomic signatures left behind are still largely uncharacterised and expected to differ from the classical selective sweep.

As discussed above, contrary to what happens in large population experiments, studies in small populations of outbred eukaryotes have shown that rapid adaptation is mediated via standing genetic variation. There is evidence from laboratory experiments that selection acts on extant polymorphism to produce a swift response to the experimental setup (Franssen et al., 2015, 2017b). Extant additive genetic variance for fitness permits for fitter individuals to be quickly picked up by selection

in the new environment. In experiments where individuals are collected from the wild, samples are likely to exhibit a substantial amount of genetic diversity. The reason standing genetic variation allows for adaptation to occur swiftly within a short time frame is twofold. First, variance in fitness increases at each generation due to recombination. It generates added fitness variance which accelerates adaptation (Barton, 2010; Sharp and Otto, 2016). Recombination breaks up linkage disequilibria (LD) between advantageous and deleterious mutations rescuing beneficial variants from low fitness backgrounds. As a result, advantageous mutations can recombine into genetic backgrounds with other positively selected variants facilitating adaptation. Alternatively, when populations are small, recombination might not act fast enough as to generate fitter combinations of alleles. In that case, if recombination is low compared to selection, it may pick up on the variation in fitness generated by extended haplotypes. Instead of a marked reduction in neutral diversity around a limited set of selected sites spread throughout the genome, one will observe an increase in frequency of a particular haplotype. In E&R studies with several replicate populations, multiple haplotypes can be selected for in different replicates (Franssen et al., 2017a). This causes a heterogeneous response amongst replicates and is a direct cause of the readily available standing genetic variation at the bout of selection.

We now have a better grasp of the complexity of short term adaptation in laboratory experiments (Schlötterer et al., 2015). However, for a long time, the main goal of E&R studies was to pinpoint specific SNPs or structural variants as putative causative sites. After detecting potential candidate loci, one would then perform a functional analysis by introducing the mutant allele in an invariant genetic background (e.g. Martins et al., 2014; Turner et al., 2013). This would allow one to confirm the key role of said variant in the response to selection. In recent years, it has become quite clear that the genetic background has a much more relevant role than previously thought. This has shifted the focus of E&R experiments which has moved on to investigating large-scale patterns of genomic adaptation. If linked to other selected sites, the ability of a positively selected allele to spread through the population can be hindered (Kim and Stephan, 2003). Without recombination to break up associations between closely linked selected sites, haplotypes might com-

pete with one another in smalls populations. However, beneficial alleles might be linked to slightly deleterious variants. This can cause the fitness of the whole haplotype to be reduced compared to the fitness of the advantageous mutation. As a result, linked selection found in recombining sexual systems, either in the form of background selection or hitchhiking, can have a substantial impact in both the outcome of selection and the genomic signature it may leave (Charlesworth and Jensen, 2021). From early population genetics theory, we also know that each selected allele interacts with nearby loci through epistasis (Hill and Robertson, 1966). Directional epistasis can increase evolvability or hinder adaptation, and it decays when selection is relaxed (Hansen, 2013). An integrated view of how all these evolutionary forces interact with one another has helped us put forward numerous hypotheses that one can both test in the lab and develop models for (Hermisson and Pennings, 2017; Höllinger et al., 2019; Gompert, 2021).

Not only do we have a better understanding of the underlying population genetics but there is also increasing E&R data availability. Thus, new and improved statistical methods are needed. The purpose of these is to detect and infer selection whilst making the most of the replicated experimental design. New methods should account for any confounding factors to do with the experimental set up that could cause spurious results. This includes the effects of small population sizes and insufficient sampling or sequencing depth. The E&R design is a valuable resource when natural population studies are underpowered from lack of replication. I have shown here the relevance of evolution studies for investigating rapid adaptation to selection pressures imposed in the lab. The next sections will introduce the evidence that has been collected so far to elucidate the adaptive genomic response, as well as examples of selection inference methods that have been put forward.

1.2 The genomics of adaptation to the lab

Researchers have studied a plethora of organisms in the lab with the goal of understanding genomic adaptation. That has led us to the realisation that depending on which organism and population size one chooses to work with the adaptive response might differ. It can range from almost deterministic in large populations to mostly affected by genetic drift when recombination is low and populations are small. Studies were thus needed to understand what aspects of the experimental design might influence statistical power to find selection in the genome. With more guidance on how to set up more optimal evolution experiments (Baldwin-Brown et al., 2014; Kofler and Schlötterer, 2014), E&R experiments are now suited to gather evidence for short-term adaptation from genomic signatures.

After many attempts at investigating allele frequency changes, there was consensus that the process of adaptation in microbial E&R experiments occurs in two distinct phases (reviewed in Orr, 2005, and Tenaillon, 2014). The initial phase of large-effect advantageous mutations that sweep through the population is typically followed by various small effect variants increasing in frequency. These fine-tuning mutations arise in multiple genetic backgrounds as the population approaches the phenotypic optimum. In other words, multi-step mutational pathways contribute to further adaptation as the long-term fate of an individual mutation may not depend solely on its immediate effect (Barrick and Lenski, 2013). In sexually reproducing microbes, selective sweeps at the start of an experiment are accompanied by strong neutral hitchhiking signal. This process is substantially different from that in small eukaryote populations. Whereas in microbes, adaptation occurs due to new mutations arising in the population (e.g. Kosheleva and Desai, 2018), adaptation in higher order eukaryotes relies mostly on standing genetic variation (reviewed in Long et al., 2015). This suggests that the pace at which either can adapt is largely determined by the mutation and recombination rates, respectively. In other words, adaptation in microbes depends on how fast new variance in fitness is generated via new mutations (reviewed in Barrick and Lenski, 2013). In contrast, small populations of eukaryotes adapt at a pace that is limited by the initial genetic variation of the founders and the rate at which advantageous variants recombine into fitter

genetic backgrounds (Kosheleva and Desai, 2018; Otte et al., 2021).

The concept of hitchhiking has been around since Maynard Smith and Haigh (1974) described it in a deterministic model. They were inspired by Lewontin (1974) who observed that allozyme variability levels were not sufficiently associated with population size. This would contradict the principles of the neutral theory suggesting there had to be another evolutionary phenomenon causing diversity to drop. Fast forward over a decade, to 1991 when the term *selective sweep* was coined by Berry et al. (1991). Berry et al. used it to encompass the whole process, from the rise in frequency until fixation of the causative locus to the decrease in neutral diversity surrounding it that is the hitchhiking phenomenon. The characteristic trough in neutral diversity around the selected site has been explored extensively by researchers attempting to model the hard sweep process. These models were often incorporated into genome scans for detecting selection in population-level datasets (reviewed in Stephan, 2019).

It is clear from studies in both natural and laboratory populations (Jha et al., 2015; Elyashiv et al., 2016; Barghi et al., 2019; Kelly and Hughes, 2019; Garud et al., 2021) that the classical sweep scenario can only partially capture the complexity of adaptation. That is especially true for rapid adaptation that relies heavily on standing genetic variation. The lack of marked troughs in neutral polymorphism has left scientists baffled (Barghi et al., 2020; Jain and Stephan, 2017a). One of two things could be causing this pattern. It could be that one is observing an incomplete sweep. Since one is merely taking a snapshot of a population at a given point in time, it is possible that not enough time has passed between the mutational event that produced the new beneficial allele and the bout of selection for a marked reduction in neutral variation. If enough generations had passed, one would have detected a reduction in neutral diversity that would match the hitchhiking model's predictions. Alternatively, one could be observing the tracks of a soft sweep. A soft sweep can be the result of adaptation through standing genetic variation, either due to polymorphism at one locus or multiple variants at loci with potentially similar fitness effects (Hermisson and Pennings, 2017). In an E&R experiment, effectively neutral mutations that were segregating in the founder population become advantageous in the new environment. Extant neutral variants

might suddenly provide a fitness advantage to those individuals that carry them in the presence of a new environmental stressor (Hermisson and Pennings, 2005). Such variants may simply evolve in a nearly-neutral manner or be present in mutation-selection balance at an appreciable frequency (Stephan, 2019) thus contributing to the evolvability of the experimental population.

Soft sweeps can be caused by directional selection on complex traits where the population fitness trajectory is first characterised by a steep increase until the population reaches the new phenotypic optimum. After this optimum is reached, the fitness trajectory plateaus. These two phases of the trajectory can be accompanied by allele frequency changes. The first step is caused by large-effect alleles that were segregating in the population. As these alleles move towards fixation, other neutral polymorphic sites present in the genetic background linked to the causative loci will also increase in frequency. However, because these large-effect mutations might find themselves in different genetic backgrounds, neutral allele diversity will not be reduced to cause a hard selective sweep. If multiple haplotypes containing one or more beneficial alleles are segregating before the environmental change, adaptation will produce a soft sweep signature (Hermisson and Pennings, 2005; Messer and Petrov, 2013). This involves the haplotypes containing said advantageous variants to increase in frequency with other less fit haplotypes being eliminated from the population. Since different combinations of advantageous alleles are already present in various genetic backgrounds, the end of this phase will result in only a few haplotypes present. The second stage after the trait mean has shifted towards the new optimum involves smaller-effect alleles. These will change in frequency as stabilising selection acts to maintain the new trait optimum (Franssen et al., 2017b).

With adaptation from standing genetic variation, one will most likely not be looking for the typical signature of a hard selective sweep. There might be multiple segregating haplotypes as well as interference between selected sites. Whilst the former facilitates rapid adaptation, the latter hinders the process. In the case of interference, the effect of hitchhiking in reducing neutral variation also decreases, maintaining some levels of polymorphism (Kim and Stephan, 2003). As we have seen, the process is complex and it most likely involves simultaneous soft sweeps that can become selected haplotypes across long stretches of the genome. A few

studies have reported that there is a heterogeneous response to selection within the replicated E&R design (Orozco-terWengel et al., 2012; Barghi et al., 2019). Experimental replication should help tease apart soft sweep signatures from other possible phenomena such as polygenic adaptation involving potentially infinite loci. This observed heterogeneous response can also be caused by sampling of multiple haplotypes that were segregating before the foundation of the experimental replicates. In which case, each replicate would have received a different haplotype that is possibly equally as advantageous.

There seems to be no clear signal of narrow regions of the genome that respond to selection in E&R experiments from genome scans. This is likely caused by the pervasive effects of linkage. With adaptation from standing genetic variation, recombination might have not had enough time to break up associations between linked selected sites that are present in one of the extant haplotypes. However, experimental populations seem to reach their new phenotypic optimum quite swiftly. Alternatively, when there is a variant sweeping through the population, genome scans might not have enough power to detect which are the true targets of selection. When selection is very strong, linkage between neutral and selected sites will cause neutrally evolving allele frequency trajectories to be shaped very much like if they were the real targets. In other words, linkage between neutral and selected sites has been hypothesised and proven to cause skewed neutral site frequency spectra. This is a well known caveat of genome scans. By analysing individual site trajectories, one can only assume that the highest scoring SNPs will be in linkage with the true causative locus or loci. One option to overcome this hurdle is to account for the effect of linkage disequilibrium. However, that is far from trivial. One can implement a multi-locus version of classical models of allele frequency evolution (e.g. Wright-Fisher or Moran) to a mechanistic model including selection, or explore correlations and covariances between allele frequency trajectories. Especially in eukaryotes with sexual reproduction, it is key to investigate the time-dependency of these correlated changes in allele frequencies whilst taking the recombinational landscape into account.

In addition to the effects of linked selection, polygenic adaptation has recently taken centre stage in E&R studies. It has helped explain why classical selective sweeps do

not fully describe the process of adaptation to the laboratory. More often than not, evolution experiments involve selection on complex traits, which typically have a polygenic basis (Schlötterer et al., 2015). Evidence has been mounting from genome wide association studies – GWAS – that most traits are highly polygenic, suggesting that considering the hard sweep signature alone will underestimate selection (reviewed in Sella and Barton, 2019). I have described the key forces that have a role in allowing populations to adapt readily to environmental shifts. Is it now timely to introduce the reader to relevant statistical methods that have been developed for detecting signatures of selection in population-level genomic datasets.

1.3 Methods for time series data

E&R studies with multiple sampling events have created a need for more sophisticated time series analyses. Making use of full allele frequency trajectories should increase one's power to detect selection in polymorphism datasets (Burke and Long, 2012). If combined with population genetics models of evolution, one might also be able to estimate the effective population size from allele frequency changes. Current methods range from more empirical statistical approaches to explicit models of evolutionary forces (reviewed in Vlachos et al., 2019). These sizeable efforts have pushed a relatively new field forward into making the most of the data that is currently available from large scale laboratory experiments.

Single-locus evolution has often been described mathematically using the Wright-Fisher (WF) model. It characterises the deterministic trajectory of a new allele in a WF population. Changes in allele frequency are often caused by other forces that violate WF principles - i.e., infinite and constant population size, no mutation, selection, migration nor drift. These violations in real populations have led scientists to propose inference methods for estimating parameters. Such parameters describe how a population deviates from the neutral WF properties. Many researchers have developed their own approximations to the process. Bollback et al. (2008) have put forward a method that uses the diffusion approximation to the process so they can estimate N_e and the selection coefficient for a given allele frequency trajectory. Others have approximated the diffusion process with a one-step method where there

is a finite number of allele frequency states (Malaspinas et al., 2012). Whilst others have created a hidden-Markov-model to produce a maximum likelihood estimate of the selection coefficent (Mathieson and McVean, 2013). Mathieson and McVean employ an expectation-maximisation algorithm to maximise the likelihood function. In addition, Steinrücken et al. (2014) developed a spectral representation of the transition density function, while Lacerda and Seoighe (2014) devised a delta method to approximate the mean and variance of the WF process. Others have considered the importance of other loci in the genome on the success of a advantageous mutation. Whereas Illingworth and Mustonen (2012) and Nené et al. (2018) have inferred the fitness of multiple haplotypes that arise in an experimental population via mutation, Terhorst et al. (2015) developed an approximation to the multi-locus WF model using Gaussian processes (GP). The latter addresses pooled sequencing since it accounts for ascertainment bias due to sampling effects. This feature is one of the key aspects of finding suitable methods for analysing allele frequency trajectories from E&R studies.

Whilst making use of these mechanistic approaches to describe adaptation, many inferential methods were developed in this context. These methods use statistical frameworks to estimate relevant parameters and can be used to analyse polymorphism datasets. As an example, Ferrer-Admetlla et al. (2016) developed a very comprehensive approach that allows them to estimate a number of evolutionary parameters, including N_e and the selection coefficient, as well as sequencing error and mutation rates. The authors introduce a discrete approximation for the diffusion process. The method also employs a Bayesian inference approach that performs the joint inference of the model parameters. Their results suggest the model is highly accurate at estimating selection coefficients in medium to large population sizes. In a study aimed at ancient DNA time series data, Schraiber et al. (2016) present a Bayesian method to infer both selection and allele age. The authors use a path augmentation approach coupled with a Markov Chain Monte Carlo (MCMC) method for integrating over the latent allele frequency state space. The key advantage to this is that it imputes allele frequencies between sampled time points if sampling is sparse.

Other authors have developed statistical approaches to test for selection explicitly.

The work published by Feder et al. (2014) presented an empirical likelihood ratio test (LTR) for rejecting neutral evolution and a frequency increment test (FIT) that looks at distribution of allele frequency increments. The authors showed that both methods were good at detecting moderate to strong selection. In 2014, Foll et al. presented a method for inferring demography and selection from time series data. The authors introduced an Approximate-Bayesian Computation (ABC) method for inferring selection using Jorde and Ryman's 2007 frequency-based N_e estimator - WFABC. Not unlike other ABC approaches, WFABC involves sampling parameter values from prior probability distributions, then simulating evolution based on those parameter values, and finally calculating summary statistics for the simulated and observed data. Shim et al. (2016) extended WFABC to include a change point in the signal of selection assuming that the selected allele is segregating at the time of the change point. This approach requires resequencing data before and after the onset of selection. Zinger et al.'s (2019) paper proposed an approach similar to Foll et al. (2014), where they designed a method suited for short-term E&R experiments or rapidly recombining virus populations. Their tool - FITS (Flexible Inference from Time-Series) - was implemented such that one has to choose one of three parameters to estimate. These are the fitness of a given allele, the mutation rate or the population size. FITS builds on Foll et al. (2014) as it is also an ABC approach that implements a two-step WF model with selection and recurrent mutation. The first step applies selection, whereas the second step incorporates drift through binomial sampling. FITS uses a default prior distribution that is based on empirical distribution of fitness effects obtained from studies in viruses.

There was a lack of any methods that were aimed at analysing E&R datasets that combine pooled sequencing with a replicated experimental setup. Iranmehr et al. (2017) proposed a method that addresses uneven coverage along the genome which causes added sampling noise. The author's approach estimates N_e as well as the dominance and the selection coefficients. It is a composition of likelihoods for E&R experiments (CLEAR). It uses a hidden Markov model (HMM) with discrete frequency states and outputs a log-odds ratio of the likelihoods for a neutral evolution and a directional selection model. Hypothesis testing relies on obtaining a distribution of those statistics and performing genome-wide drift simulations to compute an

empirical null-distribution. This allows CLEAR to output a p-value for each variant. Another use of Gaussian processes was reported by Topa et al. (2015) where the authors apply a beta-binomial GP model to rank allele frequency trajectories that have a significant non-random variation in abundance over time. They apply it to E&R experiments where there are allele frequency changes across sampled time points. The GP model is coupled with a beta-binomial prior for the allele counts that represents the uncertainty generated by finite sequencing depth. The program outputs a Bayes factor (BF) that is a ratio of two likelihoods: one for a time-dependent model where allele frequencies depend on previous states, and another for a time-independent model where allele frequencies changes are governed by drift alone. In other words, a model where an allele increases in frequency consistently through time due to directional selection is compared to another where it changes in abundance randomly in either direction. Taus et al. (2017) have taken a different approach to modelling allele frequency trajectories where they employ linear least squares (LLS) regression to fit allele frequencies to a purely deterministic selection model. The method infers the selection coefficient and starting allele frequency by fitting a LLS model to logit-transformed allele frequency data. The method also simulates neutral trajectories so it can output a p-value for each trajectory. More recently, Kojima et al. (2020) have coupled WF parameter estimation with formal statistical testing. The authors applied an expectation-maximisation algorithm to the Kolmogorov forward equation that is associated with the diffusion approximation of the WF model. They then describe a LRT to detect selected variants.

The process of adaptation is inherently complicated given that each variant in the genome interacts with other variants it is linked to. Its genomic context might be just as important as its direct effect on fitness. That is why an increasing amount of studies have tried to report the impact of linked selection in both neutral polymorphism and other selected sites. Some have developed methods multi-locus approximations of the WF model (Terhorst et al., 2015; Sohail et al., 2021). Terhorst et al. (2015) use their multi-locus GP approximation of the WF process and incorporate it into a LRT that uses an empirical null distribution that calculated by running additional simulations. In a later study, Sohail et al. (2021) presented a multi-locus model using the diffusion approximation to infer selection from time series data. They call

their method Marginal Path Likelihood (MPL) as it computes the probability of an evolutionary path, that is described by the set of mutant allele frequencies present at each sampled time point. MPL uses Bayesian inference to calculate the posterior distribution of possible selection coefficients given an observed path of allele frequencies. Because it is a multi-locus model it helps disentangle the effects of linkage, but it does require phased genotype data.

Others tried to quantify the impact of linked selection in causing allele frequency change by measuring temporal autocovariances (Buffalo and Coop, 2019). The method estimates the additive genetic variance for fitness using the property of temporal autocovariances that are caused by LD between a neutral allele and the fitness of its genetic background which persists over generations. Their approach also employs a least squares method to estimate drift-N_e. Buffalo and Coop (2020) apply their statistical method to several evolution experiments, particularly to Barghi et al. (2019). They found that over 20% of allele frequent changes were caused by selection, and that covariances between adjacent time points were positive initially but decay as one looks at more distant time intervals. This is consistent with positive selection affecting linked neutral trajectories until enough time has passed that LD is broken up. Instead of including neutral variation in their analyses, Franssen et al. (2017a) take on the task of reconstructing selected haplotypes from time series data. Their approach is targeted for replicated experimental setups that resequence using pool-seq. The method identifies selected haplotypes from finding correlated allele frequency trajectories. Correlated trajectories are maintained by physical linkage between selected sites.

As I have described, the past decade has been incredibly fruitful in terms of investigating short-term adaptation using time series data. Coupled with this need to analyse complex data, a number of versatile statistical approaches were developed. Most rely on population genetics models that describe the evolution of idealised populations. However, lab experimental populations commonly violate several assumptions made by any such models. For example, it is often the case that population sizes are small such that genetic drift determines the fate of low frequency alleles even when these are beneficial. This might hinder adaptation as genetic diversity is quickly erased reducing the adaptive potential of the population. Additionally,

when resequencing, one rarely samples the whole population. This results in increased sampling noise that can be further aggravated by uneven genome coverage. Both of these features of the experimental design can severely underpower the statistical analysis. That would cause genome scans to have poor resolution at pin pointing which are the true targets of selection. Since experiments are underpowered to detect selection on individual SNPs, the focus has started to shift towards describing larger scale patterns namely the role of structural variants in adaptation. The adaptive process is further complicated by linkage and interference where recombination plays a decisive role. The future of E&R data analysis thus lies on new mechanistic models that can incorporate the complexities of the genetic background.

1.4 The genomics of sexual selection in *D. pseudoobscura*

As early as the 19th century, Darwin (1871) wrote about the special case of natural selection that is traits evolving differently between the sexes. Sexual selection thus results from competition among same sex individuals for mates of the opposite sex. Individuals exhibiting some trait variant outcompete other members of the same sex gaining access to mates. With better understanding of heritable variation through DNA, Parker (1979) reasoned that there must be conflict between the sexes since they share one genome but might have divergent evolutionary interests. There is sexual conflict when there is sexually antagonistic (SA) selection acting on shared traits. These traits can have similar – intralocus conflict – or different – interlocus conflict – genetic basis in each sex (Perry and Rowe, 2015; Rowe et al., 2018). Many have investigated how this intragenomic conflict might arise whilst others focused on how it may be resolved.

Sexually antagonistic selection can thus lead to a co-evolutionary arms race between the sexes (Holland and Rice, 1998). Evidence for an arms race has been found in multiple behavioural traits, particularly in species of *Drosophila*, which have been repeatedly chosen as a model organism to study in the lab for well over a century. Holland and Rice (1999) set out to investigate the consequences of imposing monogamy on the naturally promiscuous *Drosophila melanogaster*. After 47 genera-

tions of selection, Holland and Rice found that males had evolved to be less harmful to females. There was also evidence that females had evolved to be less resistant to male harm as a result of relaxed sexual selection under monogamy. Other studies of *Drosophila* matings also showed that seminal fluid proteins have numerous fitness-related effects on female flies (Chapman et al., 1995; Chapman and Davies, 2004). These proteins can cause physiological changes to females, including altered immune gene expression and ovulation, as well as altered aggression and feeding behaviours. All these factors will affect female survival and overall reproductive rate.

One of the main advantages of modern experimental evolution is that it allows investigating the genetic basis of traits that respond to sexual selection. Despite it being especially difficult to detect genetic variation that might be associated with behavioural traits, there are few examples that relate mating behaviour with polymorphism (reviewed in Boake et al., 2002, and Wilkinson et al., 2015). Ritchie (2000) found that female preference for male song was determined by sex-linked genes in the bushcricket *Ephippiger ephippiger*. Later, Noor et al. (2001) were able to map inversions on the X and second chromosomes of *Drosophila pseudoobscura* and *Drosophila persimilis* which they associated successfully with female preference and male courtship wing vibration. Turner et al. (2013) performed an association study between over 13k SNPs and interpulse interval (IPI) of male courtship song. The authors even validated a few candidate genes. Chenoweth et al. (2015) set out to assess the impact of sexual selection in comparison to that of natural selection in *Drosophila serrata* populations. They found that, after genotyping over 1,400 SNPs, sexual selection had affected many of the same genomic regions as natural selection. Even more recently, Ruzicka et al. (2019) used sex-specific fitness data from fully sequenced *D. melanogaster* lines. The authors identified 226 clusters of potential candidate SNPs throughout the genome that responded to sexually antagonistic selection. However, the authors found no evidence that the X chromosome was a hot spot for sexually antagonistic variation as some theory had previously suggested (Rice, 1984).

It had been originally proposed by Rice (1984) that sexually antagonistic variation should be more common on the X chromosome in species where males are the heterogametic sex. This would be in contrast with the autosomes where sexually

antagonistic polymorphism would be rare. Rice (1984) proposed a theoretical model where a sexually antagonistic allele will readily invade the population if dominant and beneficial to females or partially recessive and favouring males. The paper then predicts that X-linked genes will increase in frequency within a wider parameter space when compared to autosomal genes assuming parallel dominance between the sexes. Despite this expectation, there is mixed evidence that the X is a hotspot for sexually antagonistic variation (e.g. see Gibson et al., 2002, Innocenti and Morrow, 2010, and Lucotte et al., 2016, for evidence for X-linked SA polymorphism, and Delcourt et al., 2009, Ruzicka et al., 2019, and Ruzicka and Connallon, 2022, for genome-wide SA variation). The answer to the question of where SA variation would be predominantly located may lie in understanding the evolution of sex-specific dominance (Fry, 2009; Mullon et al., 2012; Spencer and Priest, 2016; Grieshop and Arnqvist, 2018). On the question of genetic variation, sexually antagonistic selection was also proposed to cause allele frequency differences between males and females. Lucotte et al. (2016) found that allele frequencies differed between the sexes in humans. More variants that differed significantly between the sexes were found on the X chromosome. The authors suggest that this is evidence of SA selection on viability. Another study of sexually antagonist selection in humans focused on autosomal variants and found that loci with intermediate degrees of sex-bias in gene expression also showed the greatest allele differentiation between the sexes (Cheng and Kirkpatrick, 2016). In other words, ongoing sex-specific selection is strongest for loci with intermediate sex-biased expression - a pattern the authors refer to as 'Twin Peaks'. Thus far, studies have shown that the genomic outcome of sexual conflict varies drastically depending on several key conditions, namely the genetic basis of the conflict and sex-specific dominance coefficients. Ultimately, this conflict may cause polymorphism to be maintained, or for mutations to invade and lead to adaptive substitutions.

In a large population where drift is minimised, adaptive substitutions have been proposed to accumulate more readily on the X chromosome. This is caused by new recessive mutations being directly exposed to selection in the hemizygous sex, i.e. allelic effects will no longer be masked in the males. Theoretical models of rates of evolution predict an accelerated rate of adaptive change in X-linked loci

relatively to autosomal alleles - faster-X effect (Charlesworth et al., 1987; Vicoso and Charlesworth, 2006). The genomic response to sexual conflict can thus be tied in with the faster-X effect. If variants affected by sexual selection are mostly found on the X chromosome, faster-X evolution may facilitate adaptation to altered mating systems. This effect was found to promote the fixation of advantageous mutations depending on the ratio of the effective population size of the X chromosome in comparison to that of autosomes (Vicoso and Charlesworth, 2009). The same model found that an N_{eX}/N_{eA} greater than $^3/_4$ promoted a faster-X effect under a much wider range of dominance levels. Some empirical studies have found that N_{eX}/N_{eA} is typically close to 1 in African populations of *Drosophila melanogaster* (reviewed in Charlesworth, 2012). If selection, mutation and migration increase N_{eX}/N_{eA} ratios, these evolutionary forces might in fact contribute to a faster-X effect. Predicting how N_{eX}/N_{eA} ratios might change under different sex-determination and mating systems is not a trivial task. However, understanding how differences between autosomes and sex chromosomes can facilitate adaptation may guide research on intralocus sexual conflict.

Understanding the mechanisms via which sexual conflict may be resolved has thus been the focus of recent research. This is especially true now that we have the ability to resequence not only whole genomes but also transcriptomes. The genetic mechanisms that can be used to resolve sexual conflict may differ depending on whether its genetic basis is intra- or interlocus. Intralocus sexual conflict results from selection on a shared trait with a genetic basis that is common between the sexes. When trait optima vary between males and females, sexual conflict arises. As discussed above, this conflict is thought to involve allele frequency differences between the sexes. It is possible that, under some restrictive conditions where linkage between loci is tight and alleles are strongly selected, genomic islands of differentiation on sex chromosomes might arise (Otto, 2019). However, these are likely to be rare and located in the recombining region of the sex chromosomes. A similar result was found when modelling genomic differentiation in autosomal loci (Kasimatis et al., 2019). In such cases, intralocus sexual conflict might be resolved through the decoupling of male and female gene expression patterns (Wright et al., 2018, 2019). Hollis et al. (2014) used *D. melanogaster* to show that enforced

monogamy does change gene expression profiles. They found that after more than 100 generations of monogamy gene expression became feminised. In contrast, if there is interlocus sexual conflict, it may as well be fully resolved by allele changes between the sexes (Cheng and Kirkpatrick, 2016; Lucotte et al., 2016). To help understand this process, E&R experimental setups should prove very useful in disentangling the genetic basis of sexual conflict resolution.

The primary data set I will focus on in this thesis is of a particular social environment experiment designed by Snook et al. (2005) using small populations of *Drosophila pseudoobscura*. There were two treatments that influenced the strength of sexual selection: monogamy (M) or elevated polyandry (E). It is known that *D. pseudoobscura* females mate with 2 to 3 males within their lifetime (Dobzhansky and Pavlovsky, 1967). Therefore, a regime where a female is housed with 6 males instead is expected to elevate sexual conflict. In contrast, monogamy should lead to relaxed selection since there is no male-male competition nor any need for further female preferences to evolve. Overall, E males were faster to sing (Snook et al., 2005; Debelle et al., 2017), courted females more often (Crudgington et al., 2010) and sired the most amount of progeny (Crudgington et al., 2009).

In addition to examining behavioural and fitness-related traits, there were several studies that tried to find patterns of adaptation to this new social environment in the genome. Both gene expression levels (Immonen et al., 2014; Veltsos et al., 2017) and allele frequency changes (Wiberg et al., 2021) were examined. Immonen et al. showed that 14% of the transcriptome was differentially expressed between treatments and that 70% of these genes were sex biased. When Veltsos et al. further investigated the transcriptome, they found that male and female abdomens where sex-specific reproductive tissues are located as well as female heads were masculinised in monogamy lines. This is somewhat unexpected since relaxed sexual selection should cause a feminisation of the transcriptome (Hollis et al., 2014).

In a very recent study, Wiberg et al. (2021) investigated allele frequency differences between treatments during the last quarter of the experiment. They found that there were "islands" of divergence between selection lines that were more often present on the X chromosome. The authors also found that there was lower diversity on the

X, as well as some evidence for selective sweeps. Overall, diversity was lower in E populations, which might suggest stronger selection acting on allele trajectories in comparison to M line flies. Genes located within 10kbp of any top scoring polymorphism were investigated to find whether these were differentially expressed between lines. This set of genes showed a significant overlap with differentially expressed genes in ovaries and testes between M and E populations.

This experiment sets an interesting example of how adaptation to a new social environment shapes traits and genes. There was evidence for a strong response to selection in both life-history traits and courtship behaviour when comparing the two treatments. Transcriptome analysis also showed differences between the treatments and indicated that males and females responded differently to the selection regime. Evidence also mounted that, in addition to phenotypic changes at the trait and molecular level, population allele frequencies responded to the selection treatments. There was signal of genomic differentiation, which might indicate that selection did produce a detectable signature. Nevertheless, most studies on this system focused on comparing either the two treatments at a single time point, or the response between two time points. Given the availability of new methods for analysing time series data, it would be very relevant to perform a resequencing study of these populations throughout the whole experiment. This should give us an idea of the impact of both selection and drift on allele frequency trajectory shapes. It would be interesting to quantify the real impact of genetic drift in producing a noticeable signal of differentiation between not only selection lines, but also amongst experimental replicates. One would expect this impact to be non-negligible since effective population size estimates from neutral markers are lower than 200 individuals. Another interesting aspect of this experimental design is that it allows us to explicitly quantify how strong selection actually is. This can be achieved by computing selection coefficients for each variant throughout the genome.

With more user-friendly inference software for analysing E&R data, one can, and should, revisit old questions that have puzzled evolutionary biologists. As we gather more complete and thorough population-level datasets, analysing allele frequency changes in lab experiments will certainly provide insight into the complex phenomenon of adaptation.

1.5 Thesis outline

In this thesis, I will describe the process of developing a new statistical method that fills the need to analyse time series data, as well as report the results of a time series study in *D. pseudoobscura*.

This thesis will consist of two research chapters. The first presents a new method for estimating selection coefficients aimed at E&R polymorphism datasets. We released this method – Bait-ER – as an open-source software that is freely available for download. It quantifies the strength of selection as it tests each variant formally to reject neutral evolution assuming a Moran model with overlapping generations. It focuses, specifically, on experimental designs with small population sizes. For the second chapter, I will discuss the findings of an extensive genome-wide time series study in *D. pseudoobscura*. The dataset produced consists of allele frequencies across five time points that span from approximately generation 20 to generation 200. I will focus on differences in estimates of the effective population size, both at a genome- and chromosome-wide level. I will aim to disentangle the effects of drift and sexual selection in a replicated experimental setup where each small replicate population is exposed to changes in their mating system. I will also aim to find specific targets of selection throughout the genome that might indicate adaptation to a new social environment. Them, I will compare and discuss the genetic basis of adaptation to monogamy and polyandry in *D. pseudoobscura*.

Finally, I will discuss my findings in the context of the latest available literature. I will discuss caveats related to the power of current experimental setups and the limitations associated with our understanding of the evolutionary phenomena underlying short-term adaptation. In particular, I will focus on how adaptation of complex traits has yet to be fully described. Moreover, I will argue that collecting data more thoroughly in an E&R setting might help bridge the gap between our current modelling approaches to polygenic adaptation and the real complexities of the process.

Chapter 2

Bait-ER: a Bayesian method to detect targets of selection in Evolve-and-Resequence experiments

Abstract

For over a decade, experimental evolution has been combined with high-throughput sequencing techniques in so-called Evolve-and-Resequence (E&R) experiments. This allows testing for selection in populations kept in the laboratory under given experimental conditions. However, identifying signatures of adaptation in E&R datasets is far from trivial, and it is still necessary to develop more efficient and statistically sound methods for detecting selection in genome-wide data. Here, we present Bait-ER – a fully Bayesian approach based on the Moran model of allele evolution to estimate selection coefficients from E&R experiments. The model has overlapping generations, a feature that describes several experimental designs found in the literature. We tested our method under several different demographic and experimental conditions to assess its accuracy and precision, and it performs well in most scenarios. Nevertheless, some care must be taken when analysing trajectories where drift largely dominates and starting frequencies are low. We compare our method with other available software and report that ours has generally high accuracy even for trajectories whose complexity goes beyond a classical sweep model. Furthermore, our approach avoids the computational burden of simulating an empirical

null distribution, outperforming available software in terms of computational time and facilitating its use on genome-wide data.

2.1 Introduction

Natural selection is a complex process that can dramatically alter phenotypes and genotypes over remarkably short timescales. Researchers have successfully tested theoretical predictions and collected evidence for how strong laboratory selection acting on phenotypes can be. However, it is not as straightforward to measure selection acting on the genome. Many confounding factors can lead to spurious results. This is particularly relevant if we are interested in studying how experimental populations adapt to laboratory conditions within tens of generations, in which case we need to take both experiment- and population-related parameters into account.

A powerful approach to gathering data on the genomics of adaptation is to combine experimental evolution, where populations are exposed to a controlled laboratory environment for some number of generations (Kawecki et al., 2012), with genome resequencing throughout the experiment. This approach is referred to as Evolve-and-Resequence (E&R, **fig. 2.1**). E&R studies are becoming increasingly more common and have already made remarkable discoveries on the genomic architecture of short-term adaptation. Examples of experimental evolution studies include those on yeast (Burke et al., 2014), red flour beetles (Godwin et al., 2017) and fruit flies (Turner et al., 2011; Debelle et al., 2017). The E&R set-up allows for describing the divergence between experimental treatments while accounting for variation among replicate populations (Schlötterer et al., 2015). This is true both at the phenotype and genotype level. Consequently, the optimal approach to finding signatures of selection, is to not only monitor allele frequency changes but to also search for consistent changes across replicates. Moreover, experimental populations are often sampled and pooled for genome sequencing. The motivation for sequencing pooled samples of individuals (pool-seq) is that it is cost-effective and it produces largely accurate estimates of population allele frequencies (Futschik and Schlötterer, 2010). Thus, statistical methods tailored for E&R studies are especially valuable, notably so when investigating allele frequency trajectories originating from pooled samples

of small populations.

Several statistical approaches have been proposed to analyse these data and detect signatures of selection across the genome. A few such methods consider allele frequency changes between two time points. These simply identify those loci where there is a consistent difference in frequency between time points. One such approach is the widely-used Cochran-Mantel-Haenszel (CMH) test (Cochran, 1954). Such tests are often preferred since they are very fast, which makes them suitable for genome-wide datasets. Other approaches allow for more than two time points: for example, Wiberg et al. (2017) used generalised linear models, and introduced a quasi-binomial distribution for the residual error to quantify allele frequency differences between treatments; and Topa et al. (2015) employed Gaussian Process models in a Bayesian framework to test for selection by identifying time dependency from earlier sampling events while accounting for sampling and sequencing noise. While the latter methods use more sophisticated statistical approaches, they remain descriptive with respect to the underlying evolutionary processes. In contrast, mechanistic approaches explicitly model evolutionary forces, such as genetic drift and selection. Such models have the advantage that they can properly account for drift, which may generate allele frequency changes that can easily be mistaken for selection. Indeed, this is usually the case for E&R experimental populations with low effective population sizes (N_e), where genetic drift is the main evolutionary force determining the fate of most alleles.

The Wright-Fisher (WF) model is the most used mechanistic model for allele frequencies from time series data. There have been numerous studies that rely on approximations of the WF process, e.g., its diffusion limit (Bollback et al., 2008), a one-step process where there is a finite number of allele frequency states (Malaspinas et al., 2012), a spectral representation of the transition density function (Steinrücken et al., 2014), or a delta method to approximate the mean and variance of the process (Lacerda and Seoighe, 2014). Others have additionally considered the importance of haplotypes arising in a population via mutation (Illingworth and Mustonen, 2012; Nené et al., 2018), or implemented an approximation to the multi-locus WF process over tens of generations (Terhorst et al., 2015). Amongst these methods, most infer selection parameters in the form of selection coefficients, whilst some can also

estimate the population size, allele age, mutation rate and even the dominance coefficient. Such parameters are key for understanding the process of genetic adaptation. Nonetheless, there are only a few approaches that couple parameter estimation with explicitly testing for selection (Feder et al., 2014; Terhorst et al., 2015; Iranmehr et al., 2017; Taus et al., 2017; Kojima et al., 2020). While these approaches are useful for detecting selected variants whilst estimating the strength of selection, not all of them are implemented in software packages that can be used genome-wide for E&R experiments.

Most approaches assume linkage equilibrium, and consequently each trajectory is analysed independently from the effects of neighbouring sites. In reality, allele frequencies at linked loci co-vary which causes selection to be overestimated around selected sites. Some have tried to measure the impact of linked selection through analysing autocovariances between adjacent sites (Buffalo and Coop, 2019), and others have investigated the correlation between nearby loci to identify selected haplotypes (Franssen et al., 2017a). Whilst these efforts are a step in the right direction, neither approaches directly estimate selection coefficients nor do they test for selection. These two approaches do not rely on modelling evolutionary processes explicitly.

To provide a review of methods that are available for analysing E&R experiments, Vlachos et al. (2019) have produced a comprehensive benchmarking analysis of such methods. Their study compares the programs in terms of overall performance including parameter estimation using simulated data. It features a number of approaches, but not all estimate selection coefficients whilst performing statistical testing for each locus individually. Based on Vlachos et al.'s work, three mechanistic methods are thus particularly relevant in an E&R context: Wright-Fisher Approximate Bayesian Computation (WFABC, Foll et al. (2015)), Composition of Likelihoods for E&R experiments (CLEAR, Iranmehr et al. (2017)) and LLS (Linear Least Squares, Taus et al. (2017)). These methods differ in how they model drift and selection, the inferential approach to estimate selection coefficients, the hypothesis testing strategy, and the extent to which they consider specific experimental conditions (**table 2.1**). WFABC employs an ABC approach that uses summary statistics to compare simulated and real data. It jointly infers the posterior of both N_e and

	WFABC	CLEAR	LLS	Bait-ER
Inference approach	Approximate Bayesian computation	Maximum likelihood	Linear least squares*	Bayesian
Hypothesis testing	Bayes factors Depends heavily on summary statistics	Likelihood-ratio tests Empirical p-values based on genome-wide drift simulations	Empirical simulated p-values based on simulations of allele trajectories	Bayes factors Based on the posterior distribution
Assumptions	WF model	WF model	WF and Moran model Allele frequencies vary linearly with the selection coefficients Weak selection	Time-continuous Moran model
Accounts for replicates	No	Yes	Yes	Yes
Accounts for sequencing noise	Yes	Yes	No	Yes
Reference	Foll et al. (2015)	Iranmehr et al. (2017)	Taus et al. (2017)	This study

Table 2.1: **Currently available software for estimating selection coefficients in E&R experiments.** The table describes several features of each method namely: i) the approach used for inferring selection coefficients, ii) whether it performs hypothesis testing or not, iii) what sort of assumptions are made about the underlying population genetics model, iv) its overall computational and inference performance, v) whether it accounts for multiple replicate populations, and vi) whether it accounts for sampling variance due to sequencing noise. WF: Wright-Fisher. *LLS under the assumption of linearity is equivalent to a maximum likelihood approach.

the selection coefficient at some locus in the genome using allele frequency trajectory simulations. Real and simulated summary statistics must agree to a certain predefined scale. This makes WFABC computationally intensive. CLEAR computes maximum-likelihood estimates of selection parameters using a hidden Markov model tailored for small population sizes. LLS assumes that allele frequencies vary linearly with selection coefficients such that the slope provides the coefficient estimate. Although all three methods have been shown to accurately estimate selection coefficients, they rely heavily on empirical parameter distributions to perform hypothesis testing: (i) WFABC is highly dependent on how accurately the chosen set of summary statistics describes the underlying evolutionary forces determining the observed trajectories; (ii) CLEAR relies on genome-wide simulations to calculate an empirical likelihood-ratio statistic to assess significance; and (iii) LLS computes an empirical distribution of p-values simulated under neutrality. One other common thread amongst these tools is that they do not account for linked selection. Be it background selection or hitchhiking, these software estimate selection without looking into how linked loci might affect other sites' trajectories. Additionally, the three software vary substantially in computational effort. Therefore, currently available methods are still limited in their use for genome-wide hypothesis testing.

Here, we propose a new Bayesian inference tool – Bait-ER – to estimate selection coefficients in E&R time series data. It is suitable for large genome-wide polymorphism datasets and particularly useful for small experimental populations. As our new approach was implemented in a Bayesian framework, it gives posterior distributions of any selection parameters while considering sources of experimental uncertainty. Bait-ER jointly tests for selection and estimates selection contrary to other state-of-the-art methods. It does not rely on empirical or simulation-based approaches that might be computationally intensive, and it properly accounts for specific shortcomings of E&R experimental design. As it currently stands, Bait-ER is not concerned with the impact of linked selection as it models individual allele frequency trajectories. To test Bait-ER and other software, we explore individually simulated trajectories, whole chromosome arm simulations with linkage and an analysis of real data. We show that Bait-ER is faster than other available software, when accounting for hypothesis testing, and still performing accurately in some

particularly difficult scenarios.

2.2 Material and Methods

2.2.1 Method outline

E&R experiments produce a remarkable amount of data, namely allele frequencies for thousands to millions of loci. We created a Bayesian framework to infer and test for selection at an individual locus that is based on the Moran model. It estimates the selection coefficient, σ, for each allele frequency trajectory, which relies on the assumption that the variant in question is a potential causative locus. The Moran model is especially useful for studies that have overlapping generations, such as insect cage experimental designs (**fig. 2.1**). Such cage experiments are easier to maintain in the lab and allow for larger experimental population sizes avoiding potential inbreeding depression and crashing populations (Kawecki et al., 2012). Furthermore, Bait-ER combines modelling the evolution of an allele that can be under selection while accounting for sampling noise to do with pooled sequencing and finite sequencing depth. Our method takes allele count data in the widely-used sync format (Kofler et al., 2011b) as input. Each locus is described by allele counts per time point and replicate population. The algorithm implemented includes the following key steps:

1. Bait-ER calculates the virtual allele frequency trajectories accounting for N_e that is provided by the user. This step includes a binomial, or beta-binomial, sampling process that corrects for pool-seq-associated sampling noise.

2. The log posterior density of σ is calculated for a given grid of σ-values. This step requires repeatedly assessing the likelihood function (equation eq. (2.3) in section section 2.2.2).

3. The log posterior values obtained in the previous step are fitted to a gamma surface (details on surface fitting can be found in **supplementary fig. A.1**).

4. Bait-ER returns a set of statistics that describe the posterior distribution of σ per locus. In particular, the average σ and the log Bayes Factor (BF) are

the most important quantities. In this case, BFs test the hypothesis that σ is different from 0. Bait-ER also returns the posterior shape and rate parameter values, α and β, respectively. These can be used to compute other relevant statistics (e.g., credible intervals, variance).

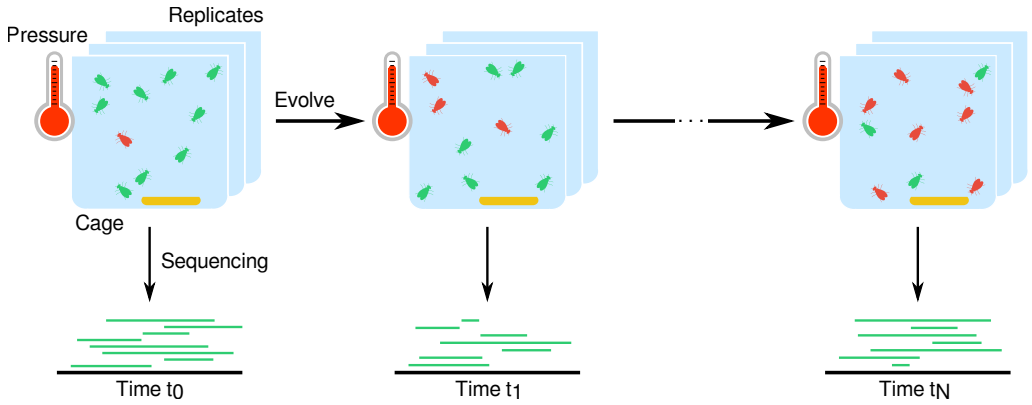


Figure 2.1: **Example of an E&R experimental setup.** E&R experiments expose several replicated populations (e.g., of flies, yeast, viruses) to a selective pressure (e.g., temperature, food regimes) for a specific number of generations t_N. The replicated populations are surveyed at several time points by whole-genome sequencing, which allows one to quantify changes in allele frequencies over time.

2.2.2 Model description

Let us assume that there is a biallelic locus with two alleles, A and a. The evolution of allele A in time is fully characterised by a frequency trajectory in the state space $\{nA, (N-n)a\}$, where n is the total number of individuals that carry allele A (in a population of size N). Supposing the allele evolves according to the Moran model where a randomly chosen individual reproduces as another is randomly drawn from the population for death, the transition rates for the process are the following

$$\begin{aligned} n \to n-1 &: \ \tfrac{n(N-n)}{N} \\ n \to n+1 &: \ \tfrac{n(N-n)}{N}(1+\sigma) \end{aligned} \quad , \tag{2.1}$$

where $1+\sigma$ is the fitness of any A-type offspring and σ the selection coefficient for allele A. If $\sigma = 0$, i.e. A is evolving neutrally, then none of the alleles is preferred at reproduction. Let X_t be the number of copies of A at time t in a population of N individuals and x_t the observed counts of A at that time; the probability of a given

allele trajectory $\boldsymbol{X}$ can be defined using the Markov property as

$$p(\boldsymbol{X} \mid \sigma) = p(X_0 = x_0) \prod_{t=1}^{T} p(X_t = x_t \mid X_{t-1} = x_{t-1}, \sigma) \quad , \tag{2.2}$$

where T is the total number of time points measured in generations at which the trajectory was assayed. The conditional probability on the left-hand side of the equation has one calculating $X_t = e^{Qd_t} X_{t-1}$, where Q is the rate matrix defined in (eq. (2.1)) and d_t the difference in number of generations between time point t and $t-1$. The probability of a single allele frequency trajectory can be generalised for R replicates by assuming their independence

$$p(\boldsymbol{X} \mid \sigma) = \prod_{r=1}^{R} p(X_0^r = x_0^r) \prod_{t=1}^{T} p(X_t^r = x_t^r \mid X_{t-1}^r = x_{t-1}^r, \sigma) \quad . \tag{2.3}$$

The main caveat for pool-seq data is the fact that it provides estimates for allele frequencies, not true frequencies. For that reason, we assume that the allele counts are generated by a binomial or beta-binomial sampling process which depends on the frequency of allele A and the total sequencing depth C obtained by pool-seq. We then recalculate the probability of the Moran states given an observed allele count c, which becomes the following with binomial sampling

$$p(\{nA, (N-n)a\} \mid \{c, C\}) \propto \binom{C}{c} \left(\frac{n}{N}\right)^c \left(1 - \frac{n}{N}\right)^{C-c}, \quad n = 0, \ldots, N \quad . \tag{2.4}$$

This step is key, for it corrects for sampling noise generated during data acquisition. This is particularly relevant for low frequency alleles and poorly covered loci.

2.2.3 Inferential framework

We used a Bayesian framework to estimate σ. It requires allele counts and coverage for each time point and replicate population $\{\boldsymbol{c}, \boldsymbol{C}\}$ at each position as input. The posterior distribution can be obtained by

$$p(\sigma|\{\boldsymbol{c}, \boldsymbol{C}\}) \propto p(\sigma) p(\{\boldsymbol{c}, \boldsymbol{C}\}|\sigma) \quad . \tag{2.5}$$

Our algorithm is defined using a gamma prior on σ. The posterior cannot be formally obtained, hence we define a grid of σ values for which we calculate the posterior density. Estimating the posterior distribution $p(\sigma|\{\boldsymbol{c}, \boldsymbol{C}\})$ is a time consuming part of our algorithm because the likelihood is computationally costly to compute. To avoid this burden, we fit the posterior to a gamma density

$$\log p(\sigma|\{\boldsymbol{c}, \boldsymbol{C}\}) = c + (\alpha - 1)\log\sigma - \beta\sigma \quad , \tag{2.6}$$

where α and β are the shape and rate parameters, respectively, and c the normalization constant. The gamma fitting represents a good trade-off between complexity and flexibility, since it only requires two parameters, but its density may take many shapes. As one requires the values of α and β that best fit the gamma density for further analyses, we find the least squares estimates of α and β (and c), such that the error is minimal. The estimation is as follows

$$\begin{aligned} \hat{\alpha} =& \frac{-(s_2 s_4 + s_4^2 - s_6 - s_7)(s_1^2 - s_8) - (s_3 + s_1 s_2 + s_1 s_4 + s_5)(s_1 s_4 - s_5)}{s_7 s_1^2 - 2 s_4 s_5 s_1 + s_5^2 + s_4^2 s_8 - s_7 s_8} \quad \wedge \\ \hat{\beta} =& \frac{-s_3 s_4^2 + s_2 s_5 s_4 + s_1 s_6 s_4 - s_5 s_6 - s_1 s_2 s_7 + s_3 s_7}{s_7 s_1^2 - 2 s_4 s_5 s_1 + s_5^2 + s_4^2 s_8 - s_7 s_8} \quad , \end{aligned} \tag{2.7}$$

where $s_1 = \sum_i x_i/N$, $s_2 = \sum_i y_i/N$, $s_3 = \sum_i x_i y_i/N$, $s_4 = \sum_i \log x_i/N$, $s_5 = \sum_i x_i \log x_i/N$, $s_6 = \sum_i y_i \log x_i/N$, $s_7 = \sum_i \log^2 x_i/N$ and $s_8 = \sum_i x_i^2/N$. We evaluated the fitting of the gamma density for neutral and selected loci, and observed that a gamma surface with five points describes the log posterior of selected and neutral loci well (**supplementary fig. A.1**).

Bait-ER was implemented with an allele frequency variance filter that is applied before performing the inferential step of our algorithm. This filtering process excludes any trajectories that do not vary or vary very little throughout the experiment from further analyses. To do that, we assess the trajectories' frequency increments and exclude loci with frequency variance lower than 0.01. These correspond to cases where trajectories are too flat to perform any statistical inference on. Trajectories such as these typically have both inflated $\hat{\sigma}$ and BFs. This filtering step allows us to improve computational efficiency as we remove trajectories that are statistically uninformative since allele frequencies are essentially constant. Such trajectories are still

included in the output file, despite Bait-ER not performing the selection inference step on them. This results in Bait-ER being suitable for large genome-wide datasets without losing any relevant information on trajectories that might be initially flat but can eventually escape drift very quickly.

Bait-ER is implemented in C++ and freely available for download at: `https://github.com/mrborges23/Bait-ER` (accessed on April 9th 2021). There, we provide a tutorial on how to compile and run Bait-ER, including a toy example with 100 loci taken from Barghi et al. (2019).

2.2.4 Simulated data

We tested our algorithm's performance under several biologically relevant scenarios using (1) a Moran model allele frequency trajectory simulator, and (2) the individual-based forward simulation software MimicrEE2 (Vlachos and Kofler, 2018).

The Moran model simulator was used, firstly, for benchmarking Bait-ER's performance across a range of experimental conditions, and, secondly, to compare our estimates of σ to those of CLEAR (Iranmehr et al., 2017), LLS (Taus et al., 2017) and WFABC (Foll et al., 2015). We started out by testing Bait-ER under different combinations of experimental and population parameters. A full description of these parameters can be found in **supplementary table A.2**. Scenarios that explored several experimental designs included those with varying coverage (20x, 60x and 100x), number of replicate populations (2, 5 and 10) and number of sampled time points (2, 5 and 11). In addition to simulating even sampling throughout the experiment, we tested our method on trajectories where we varied sampling towards the start or towards the end of said experiment. Total study length might also affect Bait-ER's estimation, therefore we tracked allele frequency trajectories for $0.2N_e$ and $0.4N_e$ generations.

We set out to compare Bait-ER to other selection estimation software using experimental parameters that resemble realistic E&R designs. Our base experiment replicate populations consist of 300 individuals that were sequenced to 60x coverage. There are five such replicates that were evenly sampled five times throughout the experiment. We then simulated 100 allele frequency trajectories for all starting

frequencies and selection coefficients mentioned above. We simulated trajectories for $0.25N_e$ as well as $0.5N_e$ generations.

The performance of both CLEAR and LLS was assessed by running the software with a fixed population size of 300 individuals (flag `--N=300` and `estimateSH(..., Ne = 300)`, respectively). Additionally, to estimate the selection coefficient under the LLS model, we used the `estimateSH(...)` function assuming allele codominance (argument `h = 0.5`). WFABC was tested with a fixed population size of N_e individuals (flag `-n 300`), lower and upper limit on the selection coefficient of -1 and 1, respectively (flags `-min_s -1` and `-max_s 1`), maximum number of simulations of 10000 (flag `-max_sims 10000`) and four parallel processes (flag `-n_threads 4`). The program was run for 1200 seconds, after which the process timed out to prevent it from running indefinitely in case it fails to converge. This caused trajectories with starting allele frequencies of 5% and 1% not to be analysed at all. We have thus only been able to include results for alleles starting at 10% and 50% frequencies.

Finally, we used data simulated with MimicrEE2 (Vlachos and Kofler, 2018) by Vlachos et al. (2019) to benchmark Bait-ER and compare it extensively with other relevant statistical methods. MimicrEE2 allows for whole chromosomes to be simulated under a wide range of parameters mimicking the effects of an E&R set-up on allele frequencies (see **supplementary figs. A.16 to A.21, A.25 and A.26** for a comparison of population parameters, including nucleotide diversity, with real experimental data). This dataset consisted of 10 replicate experimental populations, and each experimental population consisted of 1,000 diploid organisms evolving for 60 generations. The haplotypes used to find the simulated populations were based on 2L chromosome polymorphism patterns from *Drosophila melanogaster* fly populations (Bastide et al., 2013). Recombination rate variation was based on the *D. melanogaster* recombination landscape (Comeron et al., 2012). 30 segregating loci were randomly picked to be targets of selection. Sites were initially segregating at a frequency between 0.05 and 0.95. Benchmarking Bait-ER using these data also allowed us to look into our method's robustness when the data generating model is not Moran: the first scenario includes allele frequency trajectories simulated under a Wright-Fisher model of a selective sweep; and the second consists of trajectories simulated under a quantitative trait model with truncating selection. In the former,

each of the targets of selection were simulated with a selection coefficient of 0.05. For the latter, 80% of the individuals with the largest trait values were chosen to reproduce.

2.2.5 Application

We applied our algorithm to the recently published dataset from an E&R experiment in 10 replicates of a *Drosophila simulans* population to a hot temperature regime for 60 generations (Barghi et al., 2019). All populations were kept at a census size of 1000 individuals. The experimental regime consisted of light and temperature varying every 12 hours. The temperature was set at either 18°C or 28°C to mimic night and day, respectively. The authors extracted genomic DNA from each replicate population every 10 generations using pool-seq. The polymorphism datasets are available at `https://doi.org/10.5061/dryad.rr137kn` in sync format. The full dataset consists of more than 5 million SNPs. We subsampled the data such that Bait-ER was tested on 20% of the SNPs. Subsampling was performed randomly across the whole genome.

2.3 Results

2.3.1 Prior fitting with Bait-ER

Bait-ER employs a Bayesian approach outlined in section 2.2.1 – Method outline – and described in further detail in the section 2.2.2 – Model description. Bayesian model fitting depends on the prior distribution implemented and requires further testing. Bait-ER uses a gamma prior for which the shape α and rate β parameters have to be defined beforehand. We tested the impact of uninformative ($\alpha = \beta = 0.001$) and informative ($\alpha = \beta = 10^5$) gamma priors on the posterior distribution of σ under standard (60x coverage, 5 time points and 5 replicates) and sparse (20x coverage, 2 time points and 2 replicates) E&R experiments. Our results show that the prior parameters have virtually no impact on the posterior estimates of unscaled σ when $\alpha = \beta < 100$ (**fig. 2.2** and **supplementary fig. A.1**), and thus, by default, Bait-ER sets both prior parameters to 0.001.

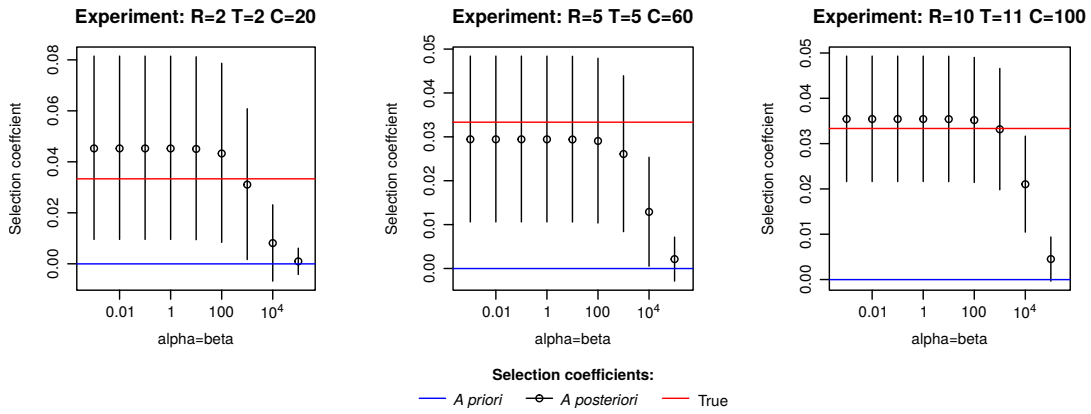


Figure 2.2: **Impact of the prior on the posterior estimates of the selection coefficients.** The posterior distribution of unscaled σ was calculated using gamma priors $G(\alpha, \beta)$, where α and β are the shape and rate parameters. We set $\alpha = \beta$ and allowed β to vary from 0.001 to 10^5 (i.e. ranging from a very uninformative to a very informative prior, respectively). The different priors were tested under three E&R experiment scenarios: the first was a sparse experimental design (coverage (C) = 20x, number of time points (T) = 2 and number of replicates (R) = 2), while the second mimicked a standard set up (C = 60x, T = 5 and R = 5). Finally, the third scenario had the most thorough experimental conditions (C = 100x, T = 11 and R = 10). Red lines indicate the true value of σ. Blue lines point to the mean of the prior imposed on σ. Black lines and points correspond to the posterior mean of unscaled σ and credibility intervals at 0.95.

Calculating the posterior distribution of σ is a computationally intensive step because it requires solving the exponential Moran matrix for several σ-values. To

reduce the number of times Bait-ER assesses the log-posterior, we fit the posterior density to a gamma distribution. We found that a gamma surface fits the posterior well, and further that five points are enough to provide a good estimate of its surface. This remains valid even for neutral scenarios, where the log-likelihood functions are generally flatter (**supplementary fig. A.1**).

2.3.2 Impact of E&R experimental design on detecting targets of selection

Bait-ER not only models the evolution of allele frequency trajectories but it also considers aspects of the experimental design specific to E&R studies. Bait-ER can thus be used to gauge the impact of particular experimental conditions in pinpointing targets of selection. We simulated allele frequency trajectories by considering a range of experimental parameters, including the number and span of sampled time points, the number of replicated populations, and coverage. Each of these settings was tested in different population scenarios that we defined by varying population size, starting allele frequency, and selection coefficient. We assessed the error of the estimated selection coefficients by calculating the absolute bias in relation to the true value. In total, we investigated 576 scenarios (supplementary **table A.2**). Heatmaps in **fig. 2.3 (A-C)** show the error for each scenario.

Heatmaps A, B, and C in **fig. 2.3** show that the initial frequency is a determining factor in the accuracy of $\hat{\sigma}$ in E&R experiments. We observed that trajectories starting at very low frequencies (around 0.01) may provide unreliable estimates of σ. However, $\hat{\sigma}$'s accuracy on those trajectories can be improved by either increasing the sequencing depth (**supplementary fig. A.4**) or the number of replicates (**supplementary fig. A.3**). Similar results have been obtained using other methods such as in Kofler and Schlötterer (2014) and Taus et al. (2017). Designs with high coverage and several replicates may be appropriate when potential selective loci appear at low frequencies (e.g., in dilution experiments). Surprisingly, alternative sampling schemes do not seem to substantially impact the accuracy of σ (see **supplementary text in appendix A**). These results have practical importance because sampling additional time points is time-consuming and significantly

increases the cost of E&R experiments.

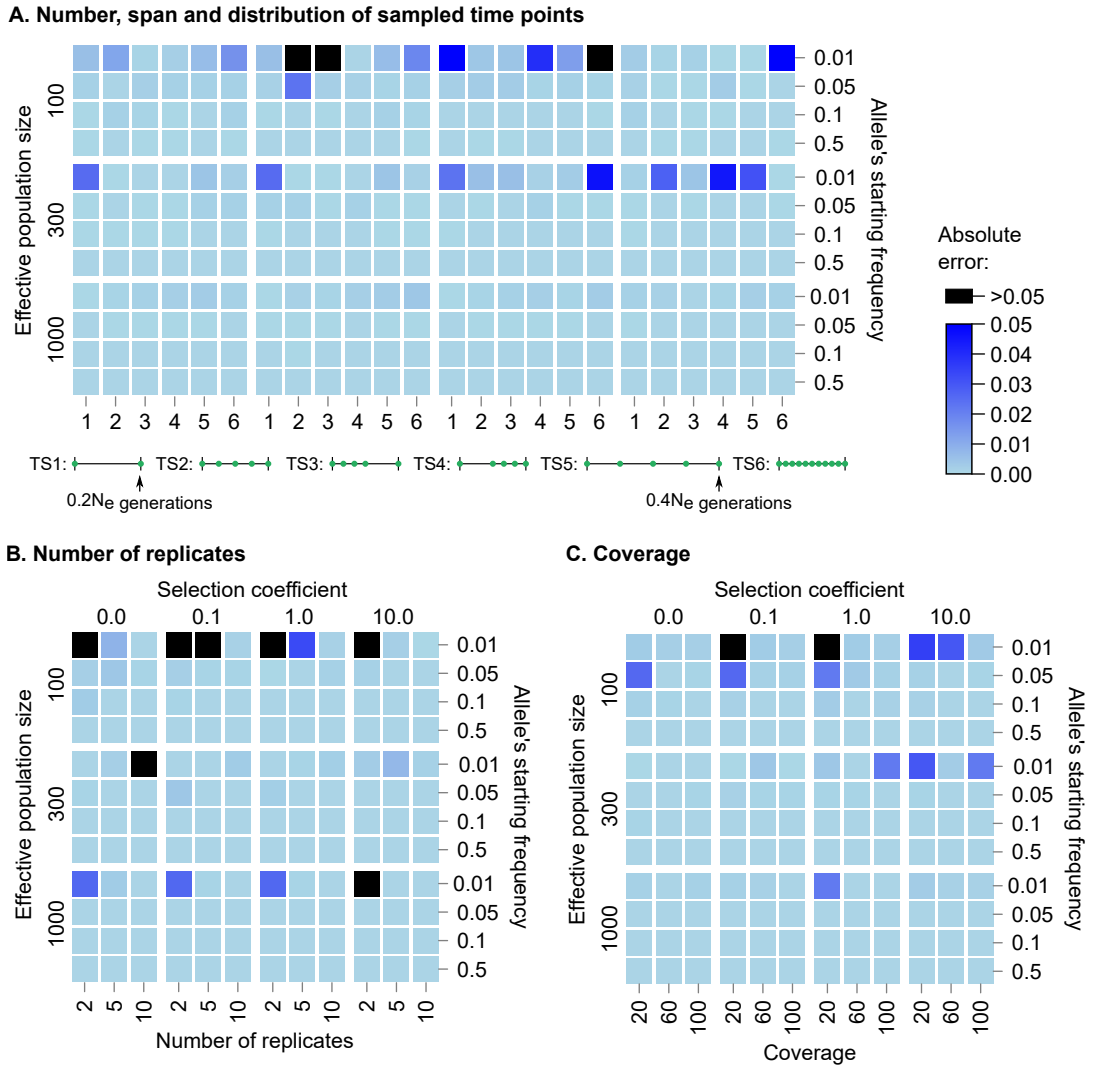


Figure 2.3: **Impact of E&R experimental design on N_e-scaled estimated selection coefficients.** Each square of the heatmap represents the error of the scaled estimated selection coefficients, i.e., the absolute difference between the estimated and the true simulated $N_e\sigma$: $N_e \times |\hat{\sigma} - \sigma|$, for a range of population dynamics and E&R experimental conditions. **(A)** Number, span and distribution of sampled time points. The six time schemes differ according to the following criteria: most time schemes have five sampling events, except for TS1 and TS6, which have two and eleven time points, respectively; all time schemes have a total span of $N_e/5$ generations, except for TS5, which has double the span ($2N_e/5$); uniform sampling was used in most scenarios but for TS3, which is more heavily sampled during the first half of the experiment, and TS4, during the second half. The two maximum experiment lengths considered ($0.2N_e$ and $0.4N_e$) were chosen based on typical E&R experimental designs. **(B)** number of replicates. **(C)** coverage. To test the experimental conditions, we defined a base experiment with five replicates, five uniformly distributed time points (total span of $0.20Ne$ generations) and a coverage of 60x. The complete set of results is shown in **supplementary fig. A.2-fig. A.5**.

A note on population size

When using Bait-ER to estimate selection coefficients, one needs to specify the effective population size, N_e. However, as effective population size and strength of selection are intertwined, misspecifying N_e will directly affect estimates of selection. The effective population size is often not known at the start of the experiment, but plenty of methods can estimate it from genomic data (e.g., Jónás et al., 2016). To assess the impact of misspecifying N_e on σ posterior, we simulated allele frequency trajectories using a fixed population size of 300 individuals. We then ran Bait-ER setting the effective population size to 100 or 1000. By doing so, we are increasing and decreasing, respectively, the strength of genetic drift relative to the true simulated population.

Bait-ER produces highly accurate estimates of σ regardless of varying N_e (**fig. 2.4** and **supplementary fig. A.5**). Misspecifying it merely rescales time in terms of Moran events rather than changing the relationship between N_e and the number of Moran events in the process. Further, we observed that the BFs are generally higher when the specified N_e is greater than the true value, suggesting an increased false positive rate. The opposite pattern is observed when the population size one specifies is lower than the real parameter. Additionally, we investigated the relationship between BFs computed with the true N_e and those produced under a misspecified N_e. We found that these BFs are highly correlated (Spearman's correlation coefficients were always higher than 0.99; **fig. 2.4** and **fig. A.5**). Taken together, our results indicate one should use a more stringent BF acceptance threshold if estimates of the real N_e have wide confidence intervals.

Furthermore, we assessed Bait-ER's computational performance by comparing the relative CPU time while varying several user-defined experimental parameters. We found that increasing N_e affects our software's computational performance most substantially (31-fold increase in CPU time when increasing the simulated population size from 300 to 1000 individuals; see **supplementary table A.1**).

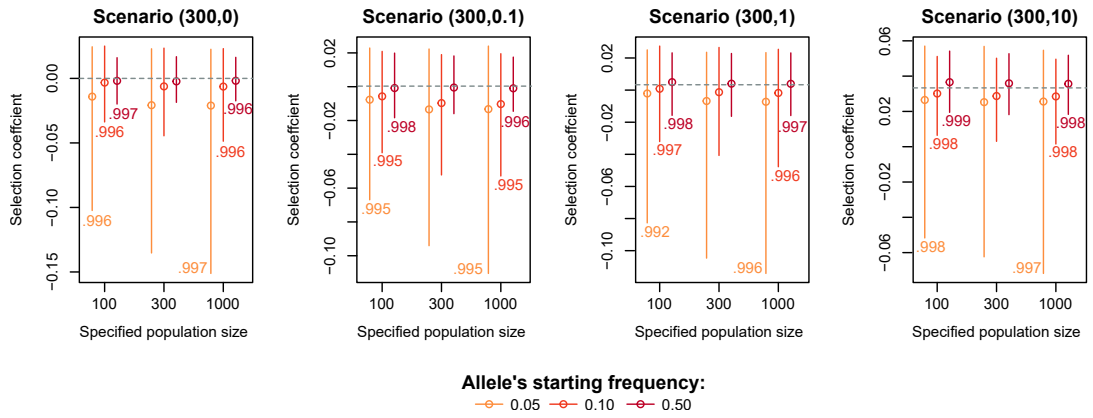


Figure 2.4: **Impact of the user-specified population size on the estimation of selection coefficients.** The plots show the distribution of unscaled estimated selection coefficients where the population size is misspecified. Vertical lines and points indicate the interquartile range and median selection coefficient. Each plot represents a specific scenario that was simulated by varying the population size, the true simulated selection coefficient (indicated within brackets $(N_e, N_e\sigma)$) and starting allele frequency (indicated by the yellow-to-red colour gradient). The numbers next to each bar correspond to the Spearman's correlation coefficient, which correlates the BFs of the 100 replicated trajectories between the cases where we have either under- and overspecified the population size ($N_e = 100$ or 1000, respectively) and the case where we use the true population size ($N_e = 300$). Regarding simulated experimental design, we defined a base experiment with five replicates, five uniformly distributed time points (total span of $0.20Ne$ generations) and a coverage of 60x.

2.3.3 Benchmarking Bait-ER with LLS, CLEAR and WFABC

Simulated Moran trajectories

To compare the performance of Bait-ER to that of other relevant software, we set out to simulate Moran frequency trajectories under the base experiment conditions described above. We tested Bait-ER as well as CLEAR (Iranmehr et al., 2017), LLS (Taus et al., 2017) and WFABC (Foll et al., 2015) on 100 trajectories for four starting frequencies (from 1% to 50%) and four selection coefficients ($0 \leqslant N_e\sigma \leqslant 10$). All population parameters were tested for both 75 and 150 generations of experimental evolution. **Figure 2.5** shows the σ estimates for Bait-ER, LLS and CLEAR under two starting frequency scenarios – 10% and 50% – and two $N_e\sigma$. CLEAR and LLS largely agree with Bait-ER's estimates of σ, even though the level of statistical significance is often not the same. It is evident that LLS produces estimates that are not as accurate as CLEAR's. This might have to do with the former not explicitly considering sampling bias in pool-seq data as a direct source of error. On the other hand, WFABC systematically disagrees with Bait-

ER's estimates because its distribution is very skewed towards high $N_e\sigma$ (greater than 180; see **supplementary fig. A.6**). This is perhaps unsurprising given that WFABC does not consider replicate populations nor finite sequencing depth unlike the other three methods. We have included WFABC in our study to compare Bait-ER with another Bayesian method. However, WFABC was not designed for E&R experiments with multiple replicates, hence its poor performance.

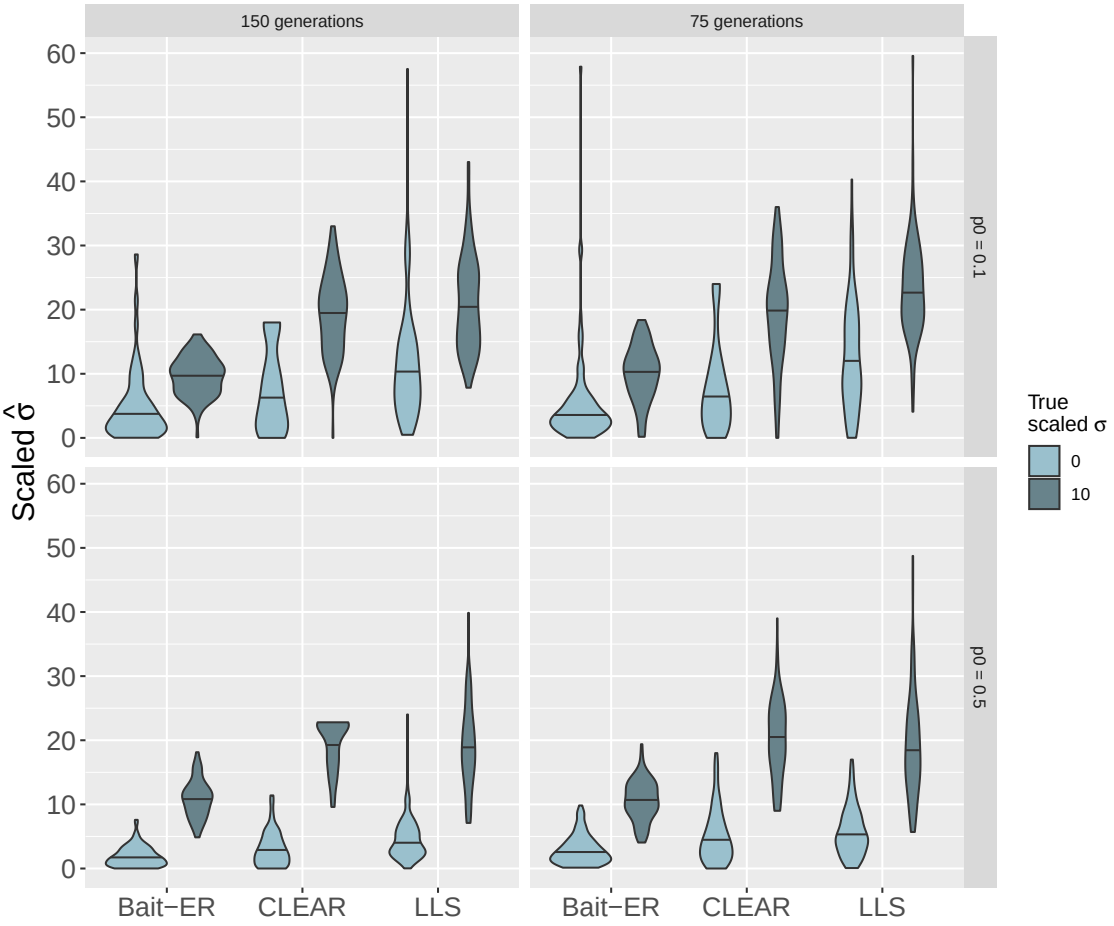


Figure 2.5: **Comparison of estimates of σ produced by Bait-ER versus CLEAR and LLS**. These plots include estimates for those Moran trajectories simulated with starting frequencies of 10% and 50% (top and bottom row, respectively). Only neutrally evolving ($N_e\sigma = 0$) and strongly selected alleles were considered here ($N_e\sigma = 10$). The left and right hand side panels correspond to two different experiment lengths: 150 and 75 generations, respectively. LLS returned NA's for 3 out of 800 trajectories which were excluded from these graphs.

Regarding computational performance, Bait-ER seems to be the fastest of the four methods, even though it is comparable to WFABC (see **fig. 2.6**). However, we tested WFABC on the first replicate population data rather than the five experimental replicates used for the remaining methods. Additionally, WFABC does not provide any statistical testing output such as a Bayes Factor. In contrast, CLEAR and LLS are slower than the other two approaches. While CLEAR takes less than 40 seconds on average to analyse 100 sites, LLS is the slowest of the four, averaging around 4 minutes. Overall, these results suggest Bait-ER is just as accurate and potentially faster than other currently available approaches, which makes it a good resource for testing and inferring selection from genome-wide polymorphism datasets.

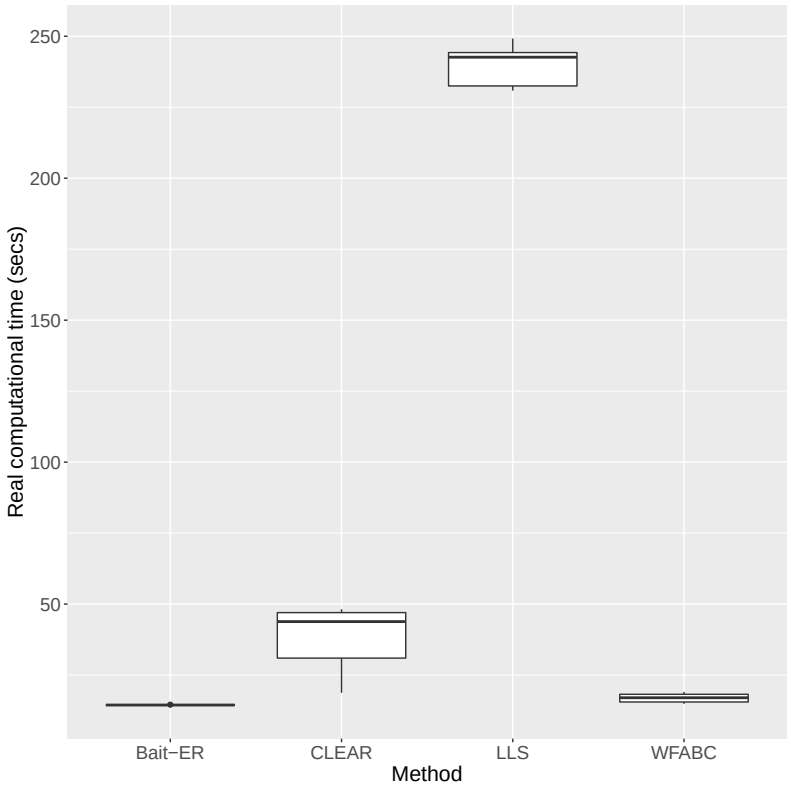


Figure 2.6: **Real computational time for Bait-ER and the other three approaches tested.** From left to right, computational time in seconds including both inference and hypothesis testing for Bait-ER, CLEAR, LLS and WFABC is shown here. Similarly to figs. fig. 2.5 and fig. A.6, these boxplots include estimates for those trajectories simulated with starting frequencies of 10% and 50%, as well as both study lengths investigated, i.e. 150 and 75 generations. Four NA's produced by LLS were again removed from these plots.

Complex simulation scenarios with recombination

For a more comprehensive study of Bait-ER's performance, we have analysed a complex simulated dataset produced by Vlachos et al. (2019). The authors simulated an E&R experiment inspired by the experimental set-up of Barghi et al. (2019) and used polymorphism data from a *Drosophila melanogaster* population. Vlachos et al. (2019) have produced this dataset to standardise software benchmarking by simulating a series of experimental scenarios that are relevant in an E&R context. We have used it to assess Bait-ER's performance at inferring selection under linkage and varying recombination rates. In particular, we chose to focus on the classic sweep scenario as well as a quantitative trait model with truncating selection, which are two of three complex scenarios simulated in Vlachos et al. (2019). Each experiment had 30 targets of selection randomly distributed along the chromosome arm.

ROC (Receiver Operating Characteristic) curves are compared for five methods, Bait-ER, CLEAR, the CMH test (Agresti, 2003), LLS and WFABC, similarly to 2A in Vlachos et al. (2019). Bait-ER performs well with an average true positive rate of 80% at a 0.2% false positive rate (**fig. 2.7 (a)**). Its performance is as good as the CMH test's, but it does underperform slightly in comparison to CLEAR. Bait-ER, CLEAR and the CMH test greatly outperform LLS and WFABC. FIT1 and FIT2 (Feder et al., 2014) are also included for comparison. These methods both use a t-test for allele frequencies and are inaccurate in a classical sweep dataset. A similar picture to that of the sweep simulation emerges for the truncating selection scenario (**fig. 2.7 (b)**). Bait-ER is amongst the top three methods despite the generating quantitative trait model being completely misspecified during inference. It is only slightly outperformed by CLEAR.

To assess why Bait-ER seems to be outperformed by CLEAR, we further investigated CLEAR's selection coefficient estimates. We note that Bait-ER assumes a continuous-time Moran model, whilst CLEAR uses a WF model for inference, much like the simulated data analysed here. Comparison of selection coefficients estimated by Bait-ER and CLEAR showed that Bait-ER is slightly more accurate on average at estimating true targets' σ (**fig. A.7**). In addition, those trajectories that scored highest with CLEAR also produced the highest Bait-ER $\hat{\sigma}$ (**fig. A.8**). True targets

of selection mostly score in the top half of Bait-ER's $N_e\sigma$ scale (**fig. A.23**). Overall, Bait-ER and CLEAR perform to a similar high standard. However, the frequency variance filter implemented in Bait-ER seems to explain our method's slight underperformance shown in **fig. 2.7**. Despite having excluded fewer than 70 (out of 300) targets of selection, Bait-ER's filtering step has also classified approximately the same amount of neutral trajectories for being too flat for inferring selection. Whilst the two method's false positive rates seem to be comparable, Bait-ER excluded a few selected sites from further analyses as they had changed very little in frequency throughout the experiment.

Our results also indicate that there is interference between linked selected sites. This phenomenon hinders adaptation as it reduces the fixation probability for each locus - Hill-Robertson Interference, HRI (Hill and Robertson, 1966). It can result in both incomplete and soft sweeps, which are often hard to detect because neither causes the characteristic trough in diversity around causative sites. Bait-ER estimated scaled selection coefficients ranged from 5.85 to 43.2, which suggests each target was under strong selection. Such values should be enough for selection to overcome genetic drift unless there is some degree of interference between selected sites within a 16Mb region. Nevertheless, even with realistic amounts of linked selection, Bait-ER identifies most targets along the chromosome arm and results in narrow peaks of significant BFs (**supplementary fig. A.15**). For the undetected targets of selection, HRI and inconsistent responses between replicate populations might cause Bait-ER not to perform optimally.

2.3.4 Analysing E&R data from hot adapted *Drosophila simulans* populations

We have applied Bait-ER to a real E&R dataset that was published by Barghi et al. (2019). The authors exposed 10 experimental replicates of a *Drosophila simulans* population to a new temperature regime for 60 generations. Each replicate was surveyed using pool-seq every 10 generations. This dataset is particularly suited to demonstrate the relevance of our method, as Barghi et al. (2019) observed a strikingly heterogeneous response across the 10 replicates. The highly polygenic

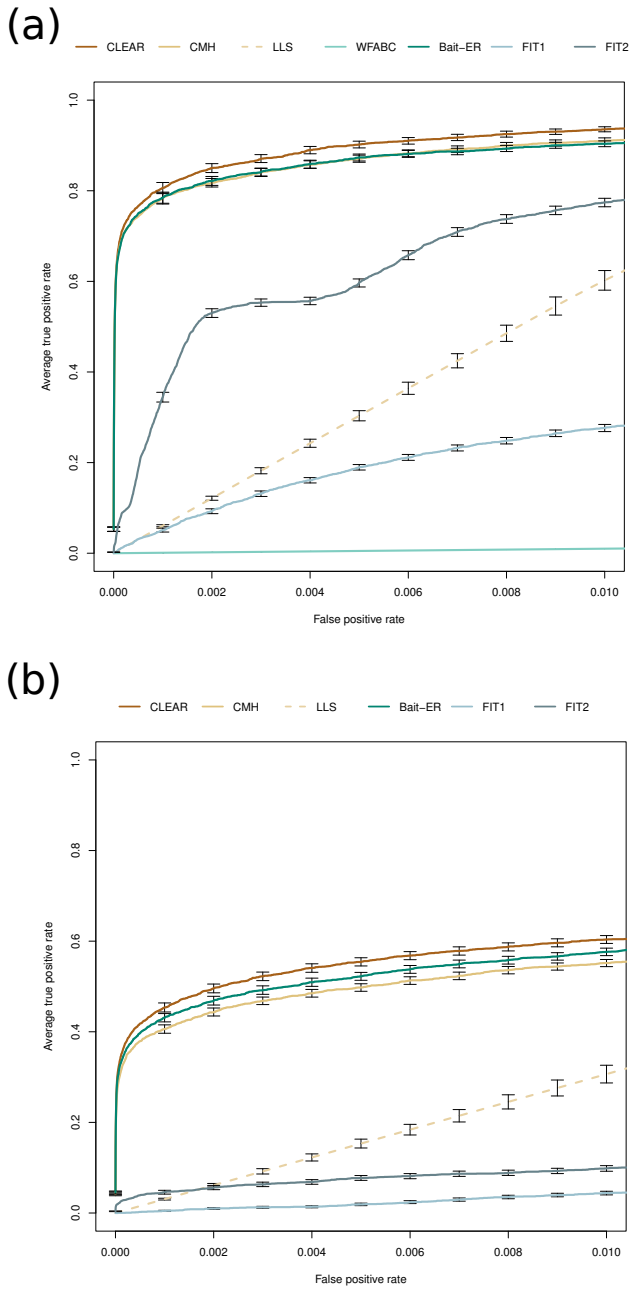


Figure 2.7: **Performance of Bait-ER and other software at testing for selection in data simulated by Vlachos et al. (2019).** ROC (Receiver Operating Characteristic) curves for Bait-ER, CLEAR, CMH, LLS, WFABC, FIT1 and FIT2 under (a) the classic sweep scenario and (b) a scenario with truncating selection. Note that LLS and WFABC were run on a subset of SNPs in (a), and that WFABC was not included in (b) for it was prohibitively slow and only finished runs for 29 replicate experiments.

basis of adaptation has proved challenging to measure and summarise thus far.

The *D. simulans* genome dataset is composed of six genomic elements: chromosomes

2-4 and chromosome X. For each element, we have estimated selection parameters using Bait-ER (mean $\hat{\sigma}$ distributions can be found in **fig. A.9**). **Figure 2.8** shows a Manhattan plot of BFs for the right arm of chromosome 3. We can observe that there are two distinct peaks across the chromosome arm that seem highly significant (BF greater than 9). These two peaks – one at the start and another just before the centre of the chromosome – should correspond to loci that responded strongly to selection in the new lab environment. Such regions display a consistent increase in frequency across replicate populations (see **supplementary fig. A.22** for the relationship between allele frequency changes and σ). Overall, there are only a few other peaks that exhibit very strong evidence for selection across the genome (**supplementary fig. A.10**). Those are located on chromosomes 2L, 2R and 3L. When compared to the CMH test results as per Barghi et al., Bait-ER's most prominent peaks seem to largely agree with those produced by the CMH (see **supplementary fig. A.11**). The same is true for high BF regions on chromosomes 2L and 2R where there are similarly located p-value chimneys at the start of these genomic elements (**supplementary fig. A.12**). Both Bait-ER and the CMH test did not produce clear signals of selection on chromosomes 3L, 4 and on the X.

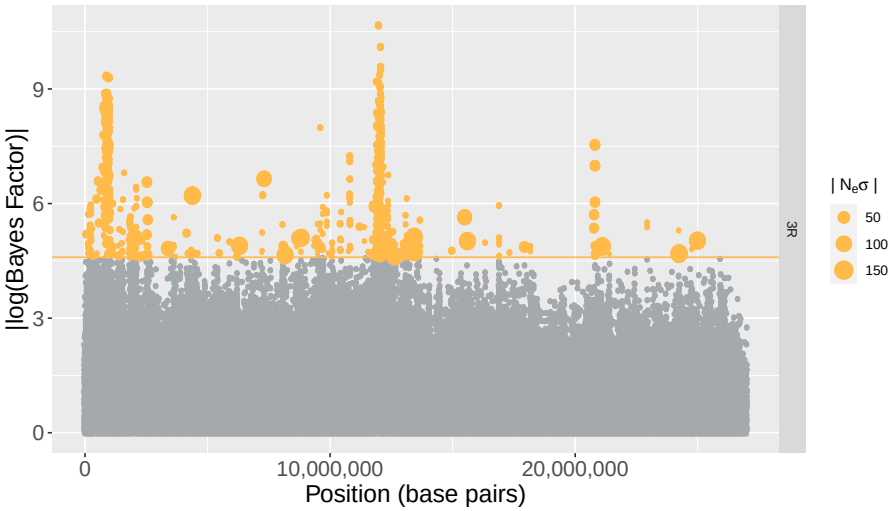


Figure 2.8: **Bayes Factors on chromosome 3R.** This Manhattan plot shows log-transformed Bayes Factors computed by Bait-ER for loci along the right arm of the 3rd chromosome in the Barghi et al. (2019) time series dataset. The orange line indicates a conservative threshold of approximately 4.6, which corresponds to log(0.99/0.01), meaning all points in orange have very strong evidence for these to be under selection. The SNPs that are significant at this level are sorted by size according to how strong Bait-ER's selection coefficients are. In other words, points are sized according to how strong the large selection coefficient is estimated to be.

One of the advantages of Bait-ER is that we have implemented a Bayesian approach

for estimating selection parameters, which means we can calculate both the mean and variance of the posterior distributions. To examine both of these statistics, we looked into how the posterior variance varies as a function of mean σ. **Fig. 2.9** shows the relationship between variance and mean selection coefficient for the X chromosome. We observe that the highest mean values also correspond to those with the highest variance. This suggests that the strongest response to selection, i.e. the highest estimated σ values, are also those showing a fairly heterogeneous response across replicates. The remaining genomic elements seem to show similar patterns, apart from chromosome 4 (see **fig. A.13**). This is consistent with other reports that inferring selection on this chromosome is rather difficult due to its size and low levels of polymorphism (Jensen et al., 2002).

Finally, we compared the p-values obtained by Barghi et al. (2019) and the BFs computed by Bait-ER. Barghi and colleagues performed genome-wide testing for targets of selection between first and last time points using the CMH test. The tests seem to largely agree for the most significant BFs correspond to the most significant p-values. However, Bait-ER appears to be more conservative than the CMH test. This follows from the finding that there is quite a substantial proportion of loci (less than 10% of all loci) that are deemed significant by a p-value threshold of 0.01, which are not accepted as such by Bait-ER. This is true even for a BF threshold of 2 such as that shown in **fig. 2.10** for chromosome 2L. Similar patterns are found in other genomic elements (**supplementary fig. A.14**).

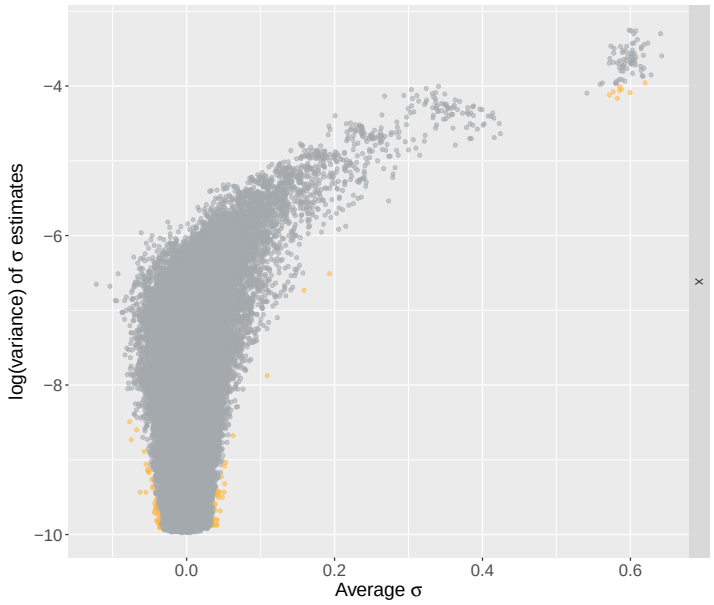


Figure 2.9: **Variance versus mean sigma on the X chromosome.** This graph compares log transformed variances in σ estimates to average σs. The variance is calculated using the inferred rate and shape parameters for the beta distribution, and the average σ is the mean value of the posterior distribution estimated by Bait-ER. Orange coloured points are significant at a conservative BF threshold of log(0.99/0.01), approx. 4.6.

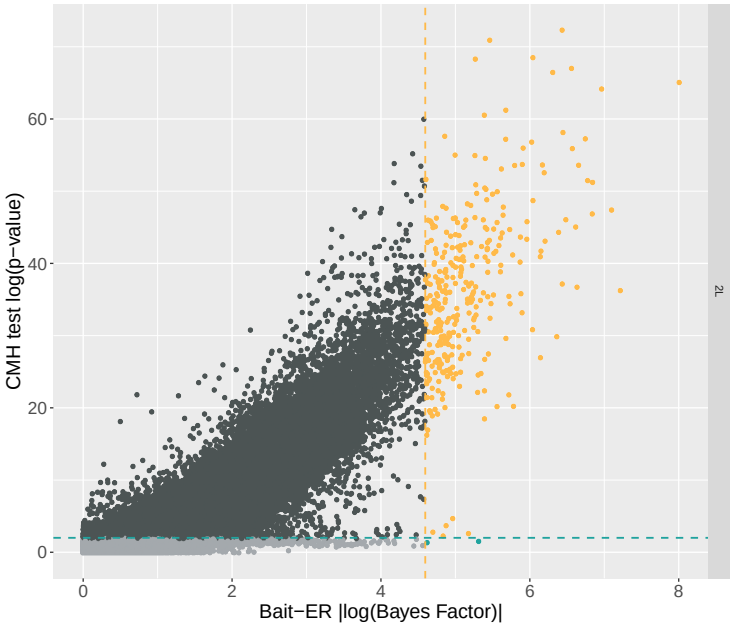


Figure 2.10: **Bait-ER's Bayes Factors versus CMH test's p-values on chromosome 2L.** Orange coloured points correspond to BFs which are greater than log(0.99/0.01) (approx. 4.6) and p-values less than or equal to 0.01, i.e. those that are considered significant by both tests. Blue coloured points indicated that the computed BF is greater than our threshold, but not significant according to the CMH test. Additionally, dark grey points are significant according to the CMH test, but not to Bait-ER, and light grey points are inferred not significant by both tests.

2.4 Discussion

One of the main aims of E&R studies is to find targets of selection in genome-wide datasets. For that, we developed an approach that uses time series allele frequency data to test for selection whilst estimating selection parameters for individual loci. As Bait-ER does not rely on simulations for statistical testing, it sets itself apart from other currently available methods. Bait-ER's implementation of the time-continuous Moran model makes it especially suitable for experimental set-ups with overlapping generations. In addition, we designed Bait-ER to be well suited for small population experiments where genetic drift has a substantial impact on the fate of polymorphisms. This is because random frequency fluctuations can force alleles to be more readily lost and, thus, overlooked by selection. When considering such polymorphisms, our stochastic modelling approach to describe their frequency trajectory is most fitting. We assume that the effect of drift is pervasive and that there is added noise from sampling a pool of individuals from the original population. We show that Bait-ER is faster and just as accurate as other relevant software. Overall, these features make it a desirable approach that can be used in many E&R designs.

Firstly, we addressed Bait-ER's performance at inferring selection. For that, we carried out a comprehensive analysis of simulated trajectories where we explored the parameter space for coverage, number of experimental replicates, user-defined population size, starting allele frequency and sampling scheme (**figs. 2.3, 2.4 and A.2 to A.5**). Our results suggest that Bait-ER's inference is mostly affected by low starting allele frequencies. This can be overcome should the sequencing depth or the number of experimental replicates be increased. Our simulations show that Bait-ER estimates selection coefficients accurately even if an allele's starting frequency is low but provided coverage is high and there are at least 5 replicates (**fig. 2.3**). Although increasing the number of replicates increases the cost of setting up an E&R experiment substantially, improving sequencing depth is certainly within reach. This interesting result might help guide future research. Encouragingly, Bait-ER performed well with small manageable population sizes, suggesting replication is key, but large populations are not necessarily required for achieving good results.

We also assessed Bait-ER's performance on a complex chromosome arm dataset simulated by Vlachos et al. (2019). We then compared it to other selection inference programs of which most are suited for time series allele frequency data. Despite numerous similarities, they vary substantially in terms of model assumptions and what sort of experimental set-up they are a good fit for. For example, WFABC seems to underperform in comparison with the other methods for E&R experiments. This is likely to be the case because it was modelled for relatively large populations. As Foll et al. (2015) show in their original study, WFABC is accurate for population sizes of 1000 individuals and for both weak and strong selection coefficients. Despite this being low in comparison to experiments in bacteria or yeast, which easily range from 10^5 to 10^8, that is not the standard population size we consider in our work. Bait-ER has been shown to perform well for such large populations (see bottom rows of each graph in **fig. 2.3**), as well as small census sizes. In reality, N_e is predicted to be a lot smaller than the census sizes reported in typical E&R studies. In contrast to WFABC, CLEAR and LLS set a better standard to which one should compare new software to. Whilst CLEAR accounts for uneven coverage, LLS only considers consistency between experimental replicates. In terms of overall performance, Bait-ER and CLEAR are similar in accuracy but Bait-ER runs substantially faster. This indicates that inferring selection from WF trajectories simulated with MimicrEE2 produces similar results regardless of whether a WF or a Moran model is used to describe the evolution of such trajectories.

To investigate Bait-ER's ability to detect selected sites in a real time series dataset, we analysed the *D. simulans* E&R experiment by Barghi et al. (2019). Bait-ER performs well on this dataset as it is rather conservative and produces only a few very significant peaks across the genome, which suggests it has a low false positive rate. It was designed to account for strong genetic drift, hence the use of a discrete-population state space. Most of the genome produced BFs greater than 2, indicating that there is not enough resolution to narrow down candidate regions to specific genes despite those very significant peaks. Barghi et al. (2019) argue that there is strong genetic redundancy caused by a highly polygenic response to selection in their experiment. Despite Bait-ER modelling sweep-like scenarios rather than the evolution of a quantitative trait using an infinitesimal model, the somewhat elevated

BF signal across the genome might indicate that the genetic basis of adaptation to this new temperature regime is indeed polygenic. Our results also suggest that the impact of linked selection might be non-negligible and worth investigating further.

We used ROC curves to compare Bait-ER's performance to six other methods'. They serve the purpose of showing the performance of a binary classification model at all significance thresholds, regardless of the statistical measurement used, may it be a p-value or a BF. ROC curves address whether the method places the true targets of selection amongst its highest scoring hits. While this is informative, it fails to account for the importance of finding an adequate significance threshold when analysing experimental data. For example, **fig. 2.7** suggests that Bait-ER and the CHM test perform very similarly. However, the CMH test returns more potential targets than Bait-ER when comparable thresholds are used for both methods (e.g. **fig. 2.10** that shows the comparison between Bait-ER logBFs and CMH test p-values for a real *D. simulans* dataset). This is consistent with other reports of the CMH test producing overinflated false positive rates on account of it confounding heterogeneity across replicates with a main effect (Wiberg et al., 2017). Additionally, whilst the CMH might be more prone to identifying high coverage sites, Bait-ER is not affected by sequencing depth (**fig. A.24**). Altogether, this indicates that Bait-ER is more conservative and that the CMH test is more prone to producing false positives.

Linkage between selected and neutral variants has long been shown to cause skewed neutral site frequency spectra (Fay and Wu, 2000). Our analysis of the Barghi et al. (2019) experiment indicates that linked selection might be the cause of a similar skew in this dataset. Of the six genomic elements in the *D. simulans* genome, five show significant SNPs all throughout the chromosome. In fact, Buffalo and Coop (2020) have analysed temporal covariances in Barghi et al.'s dataset to quantify the impact of linked selection in a model of polygenic adaptation. They found that the covariances between adjacent time points are positive but do decay towards zero as one examines more distant time intervals. This would be predicted under a scenario where directional selection is determining the fate of linked neutral loci. Over 20% of genome-wide allele frequency changes were estimated to be caused by selection, particularly linked selection. Linkage disequilibrium (LD) between neutral

and selected sites is likely to have a substantial impact on genome scans such as Bait-ER that assume independence between sites. Bait-ER producing an elevated signal throughout the genome is consistent with this prediction. Linked selection is especially evident in highly significant peaks of BFs (**figs. 2.8 and A.10**). The trajectories within such peaks will have similar sweep-like shapes and will likely consist of causative loci as well as closely linked neutral sites. These results are in contrast to what we obtained from analysing Vlachos et al. (2019) where linked selection does not generally affect Bait-ER's ability to detect the true targets of selection. Bait-ER's BFs are able to pinpoint the majority of causative loci without producing peaks due to LD (**fig. A.15**). This indicates that the data simulated by Vlachos et al. might not fully reproduce the complexity of real genomic data.

Barghi et al. (2019) claim that their experiment showed a very distinctive pattern of heterogeneity amongst replicate populations. In other words, they found that different replicates had different combinations of alleles changing in frequency together throughout the whole experiment. This heterogeneous response can be the result of sufficient standing genetic variation followed by haplotype segregation amongst replicates. If there was enough time for the multiple beneficial mutations to spread to different genetic backgrounds before the onset of laboratory selection, the haplotypes that were present at the foundation of the population replicates could have been segregating from the start. In an independent study, Buffalo and Coop (2020) found that there is a substantial proportion of the initial allele frequency change in response to selection that is shared between replicates in the Barghi et al. dataset, but this pattern is overturned rapidly. This can be caused by the population swiftly reaching the new phenotypic optimum, thereby hitchhiker alleles spread through the population along with adaptive sites, which reach high frequency very quickly. These linked loci eventually recombine on to other genetic backgrounds causing linkage to dissipate. After the adaptive phase, recombination is coupled with stabilising selection which will then act to maintain those beneficial alleles at high frequencies whilst allowing for neutral variation to accumulate.

The consequences of replicate heterogeneity on genome scans are twofold. On the one hand, different segregating haplotypes could be selected for in different replicates. This will cause genome scans not to find any convergent genotype frequencies, sug-

gesting the response to selection is varied across replicates. The process is difficult to characterise unless there is sufficient data on the founder haplotypes. However, numerous studies have time series data that does not include full sequences of those starting haplotypes, e.g. Barghi et al. (2019) and Burke et al. (2014). On the other hand, if there was enough diversity at the start of the experiment, it is possible that multiple interacting beneficial mutations are already present in the standing genetic variation. Interference between linked selected sites through epistasis can reduce the effectiveness of selection (Hill and Robertson, 1966). This will be more prevalent if there are large effect loci in the vicinity. Our results indicate that that might be the case in the sweep simulated by Vlachos et al. (2019), where the authors simulate a little over 10% of the *D. melanogaster* total genome length. Each simulated segment had 30 selected targets. For moderate to strong selection, that might be enough for linkage to hinder rapid adaptation and produce signatures that are not readily captured in genome scans.

Bait-ER estimates and tests for selection. However, σ estimates are not to be taken literally as linked selection might be inflating individual selection coefficients. Such is the case that nearby sites are not independent from one another that extended haplotypes might be rising to fixation at once. In a short timescale such as that of an evolution study, recombination is unlikely to have had the chance to have broken up haplotypes present in the standing genetic variation. In addition, one expects drift to exacerbate the effect of linked selection in experiments where populations are small. Selection inference methods will likely be affected when the combined effect of linkage and drift is pervasive. Maximum likelihood estimates of selection coefficients were shown to be unaffected by demography in populations as small as 500 individuals (Jewett et al., 2016). However, it is common that N_e in laboratory experiments is lower than the census population size. For example, Barghi et al. (2019) have reared flies in populations of roughly 1000 individuals, but they have estimated N_e to be around 300. Collectively, our results suggest that drift should not be neglected as it might inflate selection coefficient estimates since it exacerbates the extent of linked selection. Its impact can be substantial especially for populations with low polymorphism levels.

Regardless of demographic factors, adaptation of complex traits in and of itself is a

challenging process to characterise. This is because trait variation is influenced by numerous genes and gene networks. There is now some evidence in the literature suggesting that polygenic adaptation is key in a handful of laboratory evolution studies. The genomic signature left by such a complex process is still hard to describe in its entirety even in a replicated experimental design. It depends on numerous factors, including the total number of causative loci, as well as on the levels of standing genetic variation within the initial population. These are not independent of each other, as the more polygenic a trait is the more likely linkage between selected sites is to generate selected haplotypes. Nevertheless, directional selection will cause some proportion of selected sites to behave as sweep-like trajectories. It is those that Bait-ER is aiming to characterise. In short-term evolution experiments, theoretical studies have shown that a shift in the phenotypic optimum can result in sweep signatures provided the effect size is large (Jain and Stephan, 2017b).

One aspect of time series polymorphism datasets that is worth of attention is that of missing data. It is sometimes the case that there is no frequency data at consecutive time points for a given trajectory. In the future, we will extend Bait-ER to allow for missing time points. Such a feature will enable one not to discard alleles for which not all time points have been sequenced. By using a probabilistic approach to estimate missing allele frequencies, Bait-ER has inherently the potential to cope with missing data when estimating selection parameters.

Results from genome-scans in E&R studies of small populations generally tend to underperform. Since drift is pervasive and LD is extensive, genome scans might suffer from low power and high false positive rates. For that reason, we plan to extend Bait-ER to explicitly account for linkage, which decays with distance from any given locus under selection. Accounting for linkage should help disentangle the effects of local directional selection on specific variants versus polygenic selection on complex traits. Modelling the evolution of linked sites by including information on the recombination landscape will further clarify the contribution of each type of selection.

Chapter 3

Selection on the fly: short term adaptation to an altered sexual selection regime in *Drosophila pseudoobscura*

Abstract

Experimental evolution studies are powerful approaches to unveil the evolutionary history of lab populations. Such studies have shed light on how selection changes phenotypes and genotypes. Most of these studies have not examined the time course of adaptation under sexual selection manipulation, by resequencing the populations' genomes at multiple time points. Here, we analyse allele frequency trajectories in *Drosophila pseudoobscura* where we altered their sexual selection regime for 200 generations and sequenced pooled populations at 5 time points. The intensity of sexual selection was either relaxed in monogamous populations (M) or elevated in polyandrous lines (E). We present a comprehensive study of how selection alters population genetics parameters at the chromosome and gene level. We investigate differences in the effective population size – N_e – between the treatments, and perform a genome-wide scan to identify signatures of selection from the time-series data.

We found genomic signatures of adaptation to both regimes in *D. pseudoobscura*. There are more significant variants in E lines as expected from stronger sexual

selection. However, we found that the response on the X chromosome was substantial in both treatments, only more marked in E and restricted to chromosome arm XR in M. N_e is lower on the X at the start of the experiment, which might indicate a swift adaptive response at the onset of selection. Additionally, we show that the third chromosome was also affected by elevated polyandry. Its distal end harbours a region showing a strong signal of adaptive divergence in E lines.

3.1 Introduction

Evolutionary biologists have put considerable effort into uncovering how social environments shape evolution, especially those that change sexual selection pressures. Studies over the years have found differences in courtship phenotypes as well as other fitness-related traits caused by altered mating systems (Chapman et al., 1995; Wigby and Chapman, 2004; Chenoweth and Blows, 2005; Hollis et al., 2017). Due to the effects of mate competition, male harm has also been found to evolve under specific environmental conditions (Holland and Rice, 1999; Yun et al., 2019).

A few key studies have tried to identify the genetic basis of adaptation to a new sexual selection regime. Sexually antagonistic loci were initially hypothesised to be more prevalent on the X chromosome (Rice, 1984). In a model with equal dominance in both sexes, Rice proposed that a sexually antagonistic variant that is either dominant and female-beneficial or recessive and advantageous to males should increase in frequency. This would then result in X-linked sexually antagonistic variation to invade more readily when compared to autosomal loci. Rice's prediction was confirmed in several evolution experiments (e.g. Chippindale et al., 2001; Innocenti and Morrow, 2010). However, subsequent theoretical predictions suggested that autosomes are just as likely to harbour sexually antagonistic polymorphism as the X under certain conditions namely relaxing the assumption of parallel dominance between the sexes (Fry, 2009; Ruzicka and Connallon, 2020). Others have subsequently shown that sexual selection seems to affect many of the same genomic regions as those affected by natural selection (Chenoweth et al., 2015). More surprisingly, the X chromosome was found not to be a hotspot for sexually antagonistic variation in inbred *Drosophila melanogaster* (Ruzicka et al., 2019). We also know that the res-

olution of sexual conflict over gene expression optima is involved in the response to selection (Innocenti and Morrow, 2010). Evidence indicates that sexual antagonism can lead to sex-biased gene expression within a relatively short timescale (Wright et al., 2018). The importance of sexual selection in shaping the genomic landscape of a population is still largely undiscovered. Our study will characterise the adaptive response of polymorphic sites throughout the genome. It will address some of general patterns in short-term adaptation that have started to emerge. In particular, we will focus on whether the response to an altered mating system is restricted to the X chromosome or more evenly distributed along the genome.

Here, we investigate patterns of genetic adaptation of *Drosophila pseudoobscura* flies in a socially manipulated environment across 200 generations of evolution. The experiment consisted of rearing replicated populations under either monogamy – M – or elevated polyandry – E. These two treatments should relax or increase sexual selection, respectively. It has been shown that behavioural and physiological traits have diverged between these lines throughout the experiment. These include courtship song and male mating and courtship rates. In summary, E males produced more attractive song, showing decreased singing latency and faster songs over longer periods of time (Snook et al., 2005; Debelle et al., 2017). These males also had higher courtship rates (Crudgington et al., 2010). In contrast, M males had smaller accessory glands and sired fewer progeny (Crudgington et al., 2009). Interestingly, female preference also seems to have coevolved with male signal in opposite directions between the two selection regimes (Debelle et al., 2014).

Earlier studies have demonstrated that sexual selection affected multiple traits substantially as populations adapted. It soon became clear that a better understanding of the genetic mechanisms responsible for differences in phenotype was needed. Analyses of gene expression patterns in virgin M and E females showed that 14% of the transcriptome was differentially expressed (Immonen et al., 2014). In addition, 70% of these differences were sex biased. This suggests loci under sexually antagonistic selection might be contributing to divergence between the treatments. Nevertheless, more evidence is required to make a compelling case for sexual conflict as a major driver of adaptation in E line flies. The prediction that sexual selection should increase the number of male-biased genes had thus to be tested (Veltsos et al., 2017).

Both female and male transcriptomes were sequenced and revealed that the majority of differentially expressed genes was found in males' heads, which is consistent with the importance of behavioural traits. Conversely, M treatment flies were predicted to exhibit a feminisation of the transcriptome. In M populations, there was indeed a feminisation of male heads but, contrary to expectations (Haig, 2006; Hollis et al., 2014), male abdomens and both female heads and abdomens were masculinised. This is relevant since the abdomens house the sex-specific reproductive tissues.

Transcriptome evolution therefore seems to be a large part of the adaptive response to sexual selection (Connallon and Knowles, 2005; Hollis et al., 2014). However, it is still unclear how sexual selection can cause allele frequencies to change in the short-term. Genomic time-series data can provide a missing link between phenotypic changes and proof of selection acting on the genome. For this reason, investigating allele frequency trajectories alongside experimental phenotypes in an Evolve & Resequence (E&R) design can prove very useful. They can help determine the rate and strength of selection driving genomic responses. E&R studies focus on adaptation from standing genetic variation and can help reveal signatures of selection by investigating allele frequency changes. Experimental populations are typically sampled and resequenced repeatedly within a certain number of generations. Samples at two time points can be used to test for selection by finding allele frequency changes that differ significantly between treatments (e.g. Pearson's chi-square, χ^2, test as in Griffin et al., 2017; Fisher's Exact test as in Burke et al., 2010; or the Cochran-Mantel-Haenszel, CMH, test as in Barghi et al., 2017). However, such approaches lack the ability to take advantage of the allele's frequency trajectory. In contrast, more probabilistic modelling frameworks use time-series data to fully describe frequency trajectories. In particular, time-series approaches gain a lot from accounting for sampling noise typical of E&R experimental designs.

There are theoretical predictions on the genetic basis of adaptation to an altered mating system that we must consider. First, diversity on the autosomes (A) is expected to differ from X-linked diversity due to differences in the effective population size. Under monogamy, X/A diversity ratios are predicted to be roughly $^3/_4$. Because males only carry one copy of the X chromosome, $N_{eX} = \frac{3}{4}N_{eA}$ under the assumption of equal variance in reproductive success between the sexes. This affects the efficacy

of selection and, consequently, diversity ratios. Under polyandry, however, these ratios are expected to shift towards even lower values especially if populations are founded following a bottleneck (Pool and Nielsen, 2008). This effect should perhaps be counter-acted by the experimental design in our study. The family size for E and M populations was set as to ensure that N_e on the autosomes was roughly the same in both lines (addressed in Snook et al., 2009). In addition, if most beneficial mutations on the X chromosome are partially recessive, diversity on the X is predicted to be lower compared to that on the autosomes (Betancourt et al., 2004; Vicoso and Charlesworth, 2009). These effects combined with sexual selection pressures should result in a marked reduction in diversity on the X chromosome. The X chromosome is known to be enriched in genes that are important for mating success, namely accessory gland proteins (Acps) which are involved in sperm production along with other seminal fluid proteins. One would thus expect sexual selection signatures to be especially prominent on the X where genes that are key for successful mating are located.

With the appropriate statistical framework, we can now characterise allele frequency changes caused by sexual selection in these *D. pseudoobscura* populations. Here, we looked for evidence of adaptation in M and E line females both at the chromosome and gene level. Our study builds on the work by Wiberg et al. (2021) who examined genomic variation between the two treatments after $\approx$ 160 generations of selection. Wiberg et al. (2021) found "islands" of differentiation between the lines located on the X and 3rd chromosomes. This work offers a more comprehensive analysis of full allele frequency trajectories. Whilst resequencing early generations was not possible, we produced and analysed a pool-seq time-series consisting of five time points throughout those 15 years of evolution in the lab. Starting at generation 21, this time-series allows us to better understand both short-term and long-term effects. Our study assumes that adaptation has proceeded from standing genetic variation in these populations so that the effect of new mutations is negligible. We estimated the effective population size – N_e – for the four main *D. pseudoobscura* chromosomes: 2, 3, 4 and X. We used a Bayesian modelling approach – Bait-ER (Barata et al., 2020) – to infer selection on individual SNPs that allows for finding potential targets of selection. We combined individual SNP tests for selection with

window-based estimates of the effective population size, N_e, which gave us a clearer view of the process of adaptation. We found that there was a substantial response to selection with the two treatments differing in their rate of adaptation.

3.2 Material and Methods

3.2.1 Experimental setup

The experiment was established in 2002 and reached approximately 200 generations (Snook et al., 2005). The ancestral population was established from 50 wild-caught females collected in Arizona, USA. The selection experiment was set up after four generations of "common-garden" laboratory evolution. Each of the two treatment lines (M and E) was replicated four times. For each selection regime, recently eclosed offspring were collected and combined given the appropriate sex ratio at every generation. All experimental populations were kept with standard food media and added live yeast at 22°C on a 12-L:12-D light cycle. For a more detailed description, see Snook et al. (2005) and Crudgington et al. (2005).

Both selection regimes were established based on the observation that *D. pseudoobscura* females are assumed to carry sperm from two males at any given time in the wild (Cobbs, 1977). Therefore, for each E treatment group, two M line groups were established. For each replicate population, 80 and 40 groups of flies remained after culling in M and E treatment lines, respectively. Consequently, we expect no differences in the potential for adaptation between treatments due to reduced effective population size.

3.2.2 Sequencing

The time-series dataset consists of 5 time points and it includes all four replicates for each selection regime. These time points are fairly evenly distributed throughout the study. The experimental setup was such that the replicates were established in a staggered fashion. Therefore, fly sampling did not occur at the same for all replicate populations: time point 1 was sampled at generations 21 and 22, time point 2 between 59-63, time point 3 between 112-116, time point 4 ranged between

160-164, and time point 5 at generation 200. For more details on the generation at which each replicate was sampled and, thus, sequenced, see **table B.1**. Samples at time point 4 were sequenced as part of Wiberg et al. (2021).

All fly samples were stored at -80°C immediately after collection in the Department of Animal and Plant Sciences at the University of Sheffield. The samples were then collected and kept at -80°C storage in the Centre for Biological Diversity at the University of St Andrews up until DNA extraction.

For each DNA sample, 40 female flies were pooled from the frozen stocks. Females were sexed and collected approximately at time of emergence, thus, we assume these to be virgin. DNA extraction was performed using a DNeasy Blood & Tissue Kit (QIAGEN) for 20 individuals. Firstly, flies were homogenised at room temperature using a Bullet Blender homogeniser with zirconium beads. After adding Proteinase K, all samples were left to incubate overnight at 56°C. The step which involves adding buffer AW1 was repeated, and the elution with buffer AE (150 µL) was also repeated to maximise DNA yield. At the end of the extraction protocol, the two 20-female samples were combined to make up a pool of 40 females and stored at -20°C.

DNA sequencing was carried out at Novogene (Hong Kong) using an Illumina HiSeq X Ten platform. The library preparation protocol resulted in a 350bp insert DNA library. For each sample, there is a set of raw paired-end reads all 150bp long.

3.2.3 Read mapping

Raw reads were filtered and trimmed using Trimmomatic (version 0.38, Bolger et al., 2014). After trimming, time points 1, 2, 3, and 5 had an average read length of 148 (min = 36, max = 150), whilst time point 4 had shorter reads with an average length of 97.3 (min = 36, max = 100). Trimmed reads were then mapped to both the complete *Drosophila pseudoobscura* genome assembly Dpse_4.0 (FlyBase, GenBank accession GCA_000001765.3, June 2018) and the X chromosome sequence of the UCBerk_Dpse_1.0 assembly (UC Berkeley, GenBank accession GCA_004329205.1, March 2019). The former reference was assembled with Illumina (150x) and PacBio (70x) reads, and the latter consists of Oxford Nanopore MinION (40x) long reads.

All paired-end reads were mapped separately using two mappers: bwa mem (version 0.7.17, default parameters, Li, 2013) and novoalign (version 4.00.Pre-20190624, Novocraft Technologies, http://novocraft.com/). Regardless of which mapper was used, over 98% of reads were mapped successfully to the reference genome (see **table B.2** for more details). The SAM files produced by the two mappers were re-aligned to around indels using GATK (Genome Analysis Tool Kit, version 3.8.1, Van der Auwera and O'Connor, 2020) as recommended in Schlötterer et al. (2014).

3.2.4 Variant calling and filtering

Variants were then called with both bcftools (mpileup and call functions, version 1.9, Li, 2011; Danecek et al., 2021) and freebayes (version 1.3.3, Garrison and Marth, 2012) (see **table B.3** for details on variants called by both callers). Variants called by bcftools were filtered according to the following criteria suggested by Kofler et al. (2016a):

1. Minimum mapping quality of 40;
2. Minimum base quality of 30;
3. Minimum allele count of 1/F at the first time point, where F is the total number of founder haplotypes or the sample size, i.e. MAF = 1/40 = 0.025;
4. Remove sites not called by FreeBayes;

In addition, we have filtered out those variants not present in both the bwa mem and novoalign mapped datasets. Such a two mapper approach to producing pool-seq data is rather conservative and preferred to ensure good quality datasets (Kofler et al., 2016a). Genome-wide, novoalign alignments led to fewer called mean SNPs genome-wide: 2,309,226 with bwa mem versus 2,194,721 with novoalign for M lines, and 2,281,034 with bwa mem versus 2,167,341 with novoalign for E populations. A similar trend is observed in the X chromosome alignment (bwa mem 862,881 and 841,935 vs novoalign 807,689 and 787,561 for M and E, respectively). Fewer variants were called as time progresses. On average, a little over a third of variants that were called on autosomes were intergenic (chromosome 2: 34.1%; chromosome 3: 34.1%; chromosome 4: 38.6%). A higher percentage was observed on the X chromosome

where 47.5% of variants were intergenic. We also considered strand bias (**fig. B.4**) and overall coverage for each analysed locus as a measure of quality (**fig. B.5**). After filtering for mapping and base calling quality, as well as retaining any variants called by bcftools and Freebayes in the two alignments produced with bwa mem and novoalign, median sequencing depth is 45x. For this reason, we decided not to filter variants based on coverage any further. Finally, we have only considered biallelic sites and of those only the ones that were found to be polymorphic at the first sampled time point were used for further analyses. This allowed us to describe allele frequency changes throughout the experiment. The total number of SNPs remaining after quality filtering as well as those called in both bwa mem and novoalign alignments for the whole-genome and the X chromosome assemblies can be found in **table B.3**. After filtering, SNPs present across all replicates between the five time points (or the first three) were considered for further selection inference. We analysed 38,065 and 51,339 full five time point trajectories in M and E lines, respectively. For N_e estimation, those variants present between any two time points were considered. More details on how these were distributed across time points and chromosomes can be found in **table B.4** and **fig. B.6**.

To investigate how quickly alleles were fixing throughout the experiment, we calculated experimental fixation rates that correspond to the number of SNPs that become fixed from one time point to the next. Experimental fixation rates were calculated as follows:

$$\text{Fixation rate} = \frac{f_{T_n} - f_{T_{n-1}}}{t_n}$$

where f_{T_n} and $f_{T_{n-1}}$ are the number of fixed sites at time point n and time point $n-1$, and t_n is the total number of trajectories being analysed.

3.2.5 Genetic diversity

We used Grenedalf to calculate estimates of nucleotide diversity and Tajima's D (Czech and Exposito-Alonso, 2021). Grenedalf computes measures of diversity corrected for pool-seq. This includes θ_π, hereinafter referred to as π. The program

follows the approach implemented in PoPoolation (Kofler et al., 2011a) and PoPoolation2 (Kofler et al., 2011b) that accounts for any bias introduced by sampling and sequencing error. Absolute π is the sum of estimates for all SNPs within a given window, and relative π the average per window. We used sync format files, which include all replicates and time points, as input. We considered sample sizes of 40 individuals and computed π as well as Tajima's D for each replicate at each time point in windows of 250kbp with a 25kbp overlap.

3.2.6 Selection inference and N_e estimation

For estimating the effective population size, N_e, we used a moment-based estimator (Jónás et al., 2016) implemented in the R package poolSeq (Taus et al., 2017). The estimator uses temporal data to investigate any allele frequency changes. It accounts for the effect of pooled sequencing by introducing variance from two sampling events: first, when individuals are sampled from an experimental population for sequencing, and second, due to uneven coverage throughout the genome. The approach models drift variance to obtain a temporal estimator for N_e. Estimates were computed for 2k SNP windows with a 10% overlap assuming that populations were sampled according to Jónás et al.'s plan II, where sampling takes place before reproduction and sampled individuals' genomes do not contribute to the next generation. Replicate medians are computed first. These are, in turn, used to calculate chromosome-level median estimates. Similarly, each genome-wide N_e estimate is a median of all replicate medians.

Finally, we investigated potential targets of selection using a Bayesian genome scan on the time-series - Bait-ER (Barata et al., 2020). It was designed for E&R experiments as it accounts for added binomial, or beta-binomial, sampling noise from pool-seq. Bait-ER models the evolution of an allele using a Moran model with overlapping generations. It estimates parameters of selection, namely selection coefficients (σ), whilst also testing each allele frequency trajectory for selection. The program outputs a Bayes Factor (logBF) per site, which is a ratio of the likelihoods of two alternative models: one where genetic drift is the main driver of allele frequency changes, and another where there is positive selection favouring a particular allele. Similar to Grenedalf, both Bait-ER and poolSeq take sync files as input.

3.2.7 Gene feature analysis

In order to obtain a complete annotation of gene features for our time-series dataset, we used NCBI's Remap tool (www.ncbi.nlm.nih.gov/genome/tools/remap). This tools allowed us to perform coordinate remapping between the latest annotated reference genome uploaded to NCBI's repository (University of California, Irvine, Dpse_MV25, accession number GCA_009870125.2) and the two assemblies described in section 3.2.2. The software outputs gff3 format annotation files which we converted to bed format using BEDOPS' (Neph et al., 2012) gff2bed tool.

3.3 Results

3.3.1 Diversity and allele frequency changes

We first investigated the behaviour of allele trajectories by looking at allele frequency spectra throughout the experiment. The time-series consists of five time points from generation 21 to generation 200 (T1: 21-22; T2:59-63; T3:112-116; T4:160-164; T5:200; see **table B.1** for more details). Frequency spectra at the start are flat distributions with maxima roughly at 0.5-0.6 (**fig. 3.1**). Alleles fixed at high rates, with the most fixations between time point 3 and time point 4 for M lines and time points 2 and 3 for E lines. Up to 29.5% and 16.9% more fixed sites than in the previous time point were observed for E and M, respectively. This indicates that diversity was more swiftly erased in E populations, as expected if sexual selection is strongest in this regime. Allele frequency changes between first and last time point show distributions that are highly skewed towards low values (**fig. B.7**). This is especially true in the case of the X chromosome.

Nucleotide diversity was measured in 250kbp windows for each chromosome separately and at each time point. Diversity distributions show a marked reduction in diversity as time passes, particularly from time point 1 to time point 2 (**fig. 3.2**). All chromosomes' densities peak at 0.4-0.5 per site at the first time point. At the end, densities for chromosomes 3 and X flatten out, especially for E flies. Interestingly, in M lines, π on the 3rd chromosome becomes skewed towards very low values in later generations. In contrast, chromosomes 2 and 4 maintain more diversity. Taken together, these results suggest that selection was pervasive on the 3rd and X chromosomes resulting in more windows of very low π across treatments. Window estimates along each chromosome exhibit some diversity peaks, particularly on the 3rd chromosome, that become flat towards the end of the experiment (**fig. B.8**).

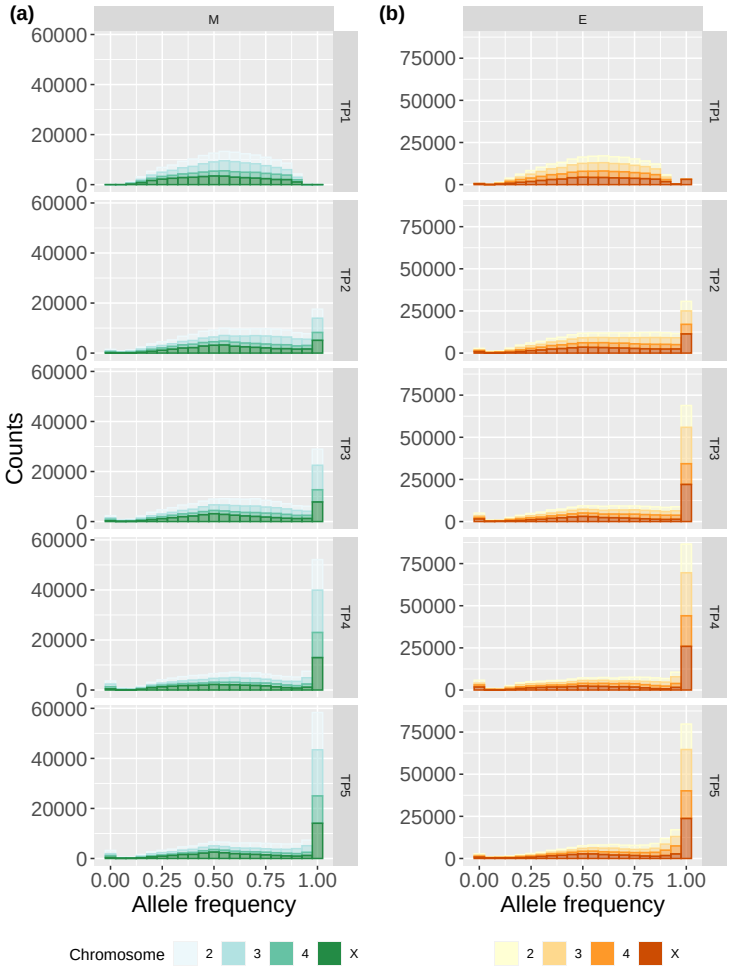


Figure 3.1: **Allele frequency spectra in M and E populations.** Allele frequency spectra in (a) for M and (b) for E populations per time point (rows) for each replicate population (columns). Each chromosome is coloured in a different shade of green (M) or orange (E) as seen on the bottom legend.

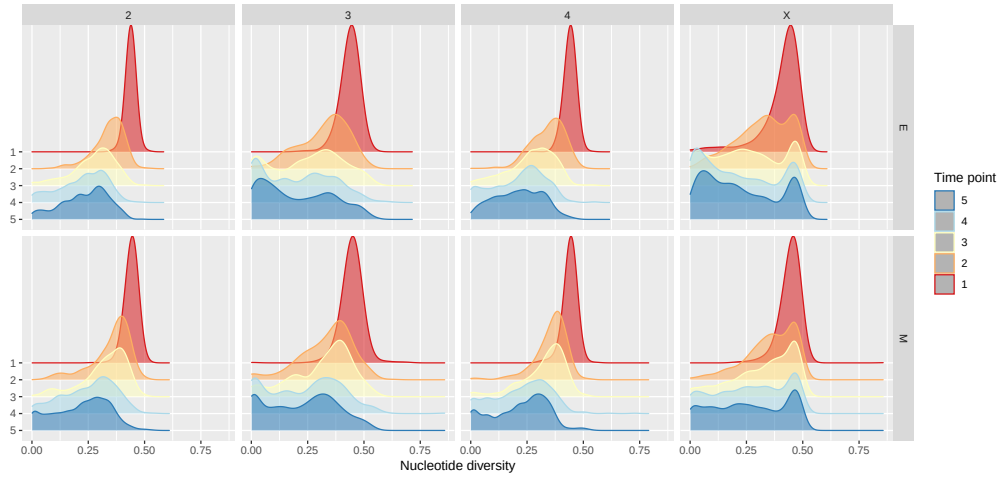


Figure 3.2: **Relative π densities in M and E lines.** Estimates were computed with Grenedalf (Czech and Exposito-Alonso, 2021). Rows correspond to the two different treatments and columns to chromosomes. Each individual plot has five densities coloured in as per side legend that correspond to one of the five time points.

Tajima's D estimates throughout the genome are typically greater than 0 but show substantial variation amongst windows (**figs. 3.3 and B.9**). These were also computed for 250k SNP windows as for π. This result suggests that there is a lack of rare alleles in our dataset which is perhaps unsurprising in a pool-seq experiment using such stringent filtering criteria. For example, we filtered variants a minimum allele frequency of 0.025 at the first time point which removes very low frequency variants. This would partly explain an elevated Tajima's D especially at the start of the experiment. One region worthy of note is the centre of the X chromosome where Tajima's D is consistently elevated in comparison to surrounding stretches (**fig. 3.3**). The pattern is present in both E and M populations throughout the whole experiment.

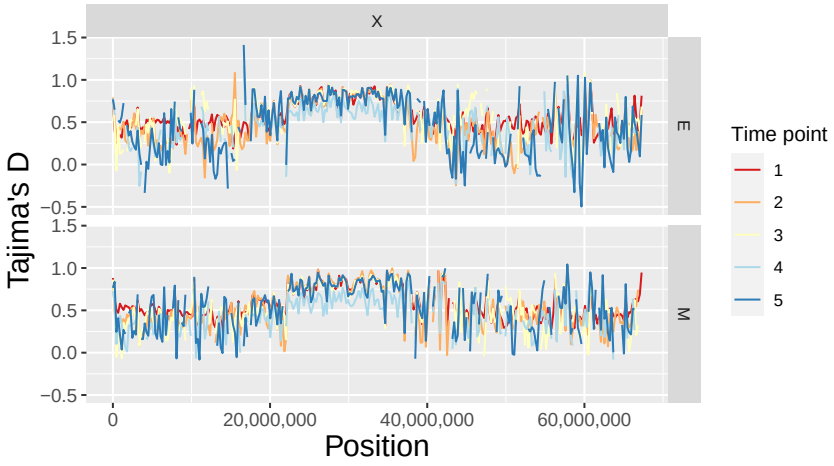


Figure 3.3: **Tajima's D estimates along the X chromosome for E and M lines.** Rows correspond to the two different treatments. Estimates were calculated in 250k SNP windows with Grenedalf (Czech and Exposito-Alonso, 2021). Lines are coloured per time point according to the side legend.

3.3.2 Estimating the effective population size

Time interval	Median - M	Median - E
Overall	144.6 (n = 134)	149.6 (n = 258)
T1T2	100.6 (n = 292)	90.7 (n = 360)
T2T3	67.1 (n = 307)	110 (n = 288)
T3T4	66.5 (n = 167)	120.3 (n = 139)
T4T5	134.2 (n = 140)	140.1 (n = 140)

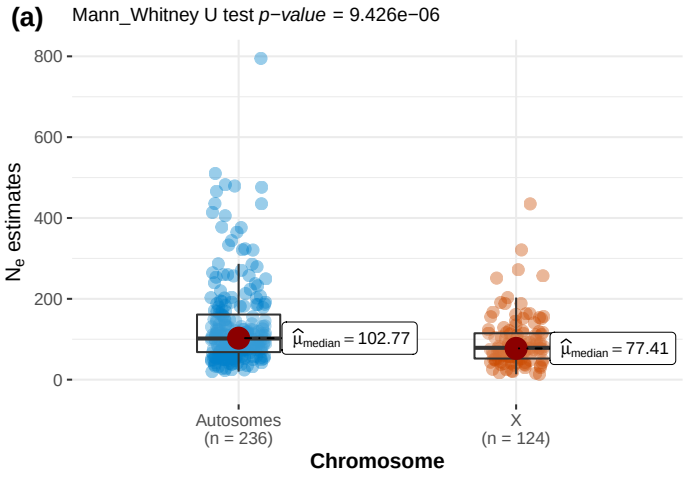
Table 3.1: **Median genome-wide N_e estimates for M and E lines at different time point intervals.** Medians were calculated using 2k SNP window estimates from all of the four experimental replicates. 'Overall' corresponds to N_e estimates based on allele frequency changes between the first and last time point. The total number of windows considered in each replicate is found in brackets.

Estimating the effective population size in windows across the genome should shed light on how fast selection and drift act together to cause allele frequencies to change. Using an estimator that relies on frequencies changing between any two time points (Jónás et al., 2016), we looked for differences in N_e between chromosomes and treatments. This approach uses data on any polymorphic sites and computes an estimate of variance-N_e which does not use any information on fixed loci. Previous results using molecular marker-based estimators, suggest that N_e is similar between lines, ranging from 141.2 (s.d. 27.4) to 110.5 (s.d. 19.2) for M and E, respectively (Snook et al., 2009). Our results confirm this: genome-wide median estimates considering allele frequency changes between time point 1 and 5 are 145 for M populations and 150 for E lines (**table 3.1**).

In E populations, N_e drops most during the first 20 to 60 generations implying that selection is strongest then. N_e starts to recover from time point 3 onwards reaching ≈ 140 at the end of the experiment. Such a result is only possible because variance-N_e is estimated from observed temporal shifts in allele frequencies. This is not caused by new mutations as we assume that adaptation occurs from standing genetic variation throughout this study. This same pattern is not found in M lines, where N_e estimates suffer a continuous and drastic reduction until time point 4 (from 101 at generation $\sim$21 to 66 at generation $\sim$161), after which neutral levels are nearly recovered (**table 3.1**). Similar patterns are recovered when using only

intergenic variants to estimate N_e (**table B.5**). Since finding that N_e recovers towards the end of the experiment, we compared estimates between the first and last time point. These are significantly different (Mann-Whitney U test p-value $= 4.4 \times 10^{-7}$), suggesting that (i) selection acting at the start might be causing a significant reduction in N_e, and (ii) selection during the last intervals of the experiment is less effective in altering allele frequencies, allowing N_e to recover.

Reduced N_e during the first half of the experiment - time points 1 to 2 - might indicate a strong selective response as E populations reach the new phenotypic optimum very swiftly. In contrast, selection under monogamy seems to act slower, causing low N_e until the third quarter of the experiment. Interestingly, the first interval is marked by low N_e on the X chromosome especially in E populations (**fig. 3.4**, panel (b)). We investigated whether the estimates on the X differed from those on the autosomes and this difference is statistically significant (**fig. 3.4**, panel (a), Mann-Whitney U test p-value $= 9.4 \times 10^{-6}$). Thus, the X chromosome shows a fast adaptive response in E lines from the onset of selection. This result is not as clearly replicated in M populations where N_e estimates on the X are more similar to those on chromosomes 3 and 4 from T1 to T4 (Mann-Whitney U test p-value $= 0.051$).



(b)

Chromosome	Treatment	Median
2	E	113.7
3	E	82.7
4	E	94.5
X	E	73.7
2	M	129.9
3	M	99.4
4	M	93.6
X	M	100.2

Figure 3.4: N_e **estimate comparison between autosomes and the X chromosome during the first 20 to 60 generations (T1 to T2).** Panel (a) shows boxplots of N_e window estimates for the two categories - autosomes and X - in E populations (Mann-Whitney U test very significant). Twelve outliers were removed from (a) for visualisation purposes. Panel (b) shows median estimates per chromosome within this first time point interval.

3.3.3 Estimating selection

We performed a genome scan of the time-series across all time points using Bait-ER (Barata et al., 2020). The signal of selection is substantially higher in E versus M lines (**fig. 3.5**), which suggests that selection is indeed stronger under elevated polyandry. In total, 350 (0.9% of all sites in the time-series) and 770 (1.5%) SNPs were statistically significant (at a threshold of $log(99)$) for M and E lines, respectively. If considering the first 3 time points alone, M lines had 570 (0.6%) significant SNPs whilst E had 1591 (1.5%). Regardless of whether you consider the complete time-series or a shorter dataset with the first 3 time points only, E populations have a similar percentage of sites - 1.5% - that are considered to be under selection. They consistently show approximately double the number of loci with evidence of selection than the M lines.

When comparing the different chromosomes, it is clear that there are far more significant peaks along the X chromosome in comparison to autosomes in both treatments (M: 322; E:563). Of those 322 top X candidates in M lines, 309 (96%) were located in intergenic regions (versus 50.8% in E). This is a surprising result given that half of the variants called on the X can be found within intergenic regions. Significant trajectories on chromosomes 2 and 4 were never more than 60 (2: 9 and 43; 4: 3 and 54; for M and E, respectively). In addition, signal on the 3rd chromosome is also markedly elevated in E populations where there are 110 significant SNPs versus only 16 in M (**fig. 3.5**). Taken together, these results suggest that whilst selection is stronger under elevated polyandry, the X chromosome is also responsible for adaptation to a strict monogamy regime.

In E lines, 417 (54.2%) of those statistically significant variants were found within genes, whereas only 35 SNPs (10%) were mapped on to genes in M lines. This difference is striking given that approximately the same number of significant variants were located in intergenic regions (E: 353; M: 315). A large proportion of the E line top variants that were found in genes locate to the X chromosome (426, 72.3%). Fourteen genes in E populations have 5 or more significant SNPs (up to 23; **table B.7**). Selection coefficients for individual trajectories as estimated with Bait-ER ranged from 0.03 to 0.11. Genes with the highest number of top SNPs were

found on the X and third chromosomes. Of these, the second gene with the most top variants (11) codes for hemicentin-1 (NCBI: 4813557; FlyBase: FBgn0076932) with an ortholog in *Drosophila melanogaster* - neuromusculin which is a protein that is expressed in the muscle system and the peripheral nervous system. Allele frequency trajectories for these top SNPs typically start at relatively high frequency and fix within two or three time points (**fig. 3.6**).

A previous study (Wiberg et al., 2021) identified 480 variants as having a significant allele frequency differences between M and E replicates at time point 3. We determined which genes these top SNPs were located in as well as any genes in the vicinity of intergenic top SNPs. We then compared these genes with those found significant in our genome scan. There were 21 genes in common between the two studies (**table B.6**). These include genes involved in neural and muscle development, as well as other biological regulation processes.

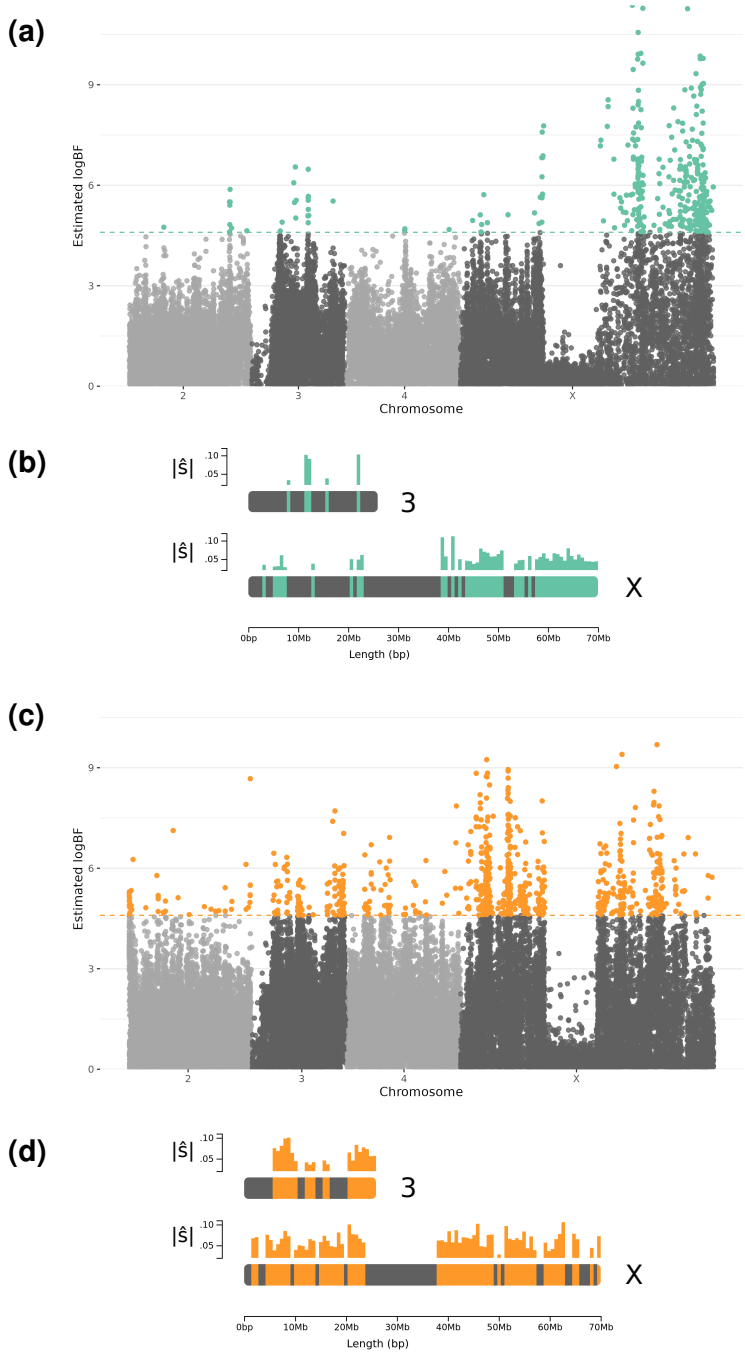


Figure 3.5: **Genome scan for signatures of adaptation throughout the genome for M (top) and E (bottom) lines.** (a) and (c) are manhattan plots of Bait-ER (Barata et al., 2020) logBF for each allele frequency trajectory. Statistically significant SNPs are coloured in green (M, top) or orange (E, bottom). Dashed lines correspond to a threshold of $log(99) \approx 4.6$. (b) and (d) are diagrams of chromosomes 3 (top) and X (bottom) that illustrate which regions of each chromosome harboured the most number of significant hits. Average estimated selection coefficients ($|\hat{s}|$) for each interval can be found above each diagram as a bar plot. Data excludes chromosome 5.

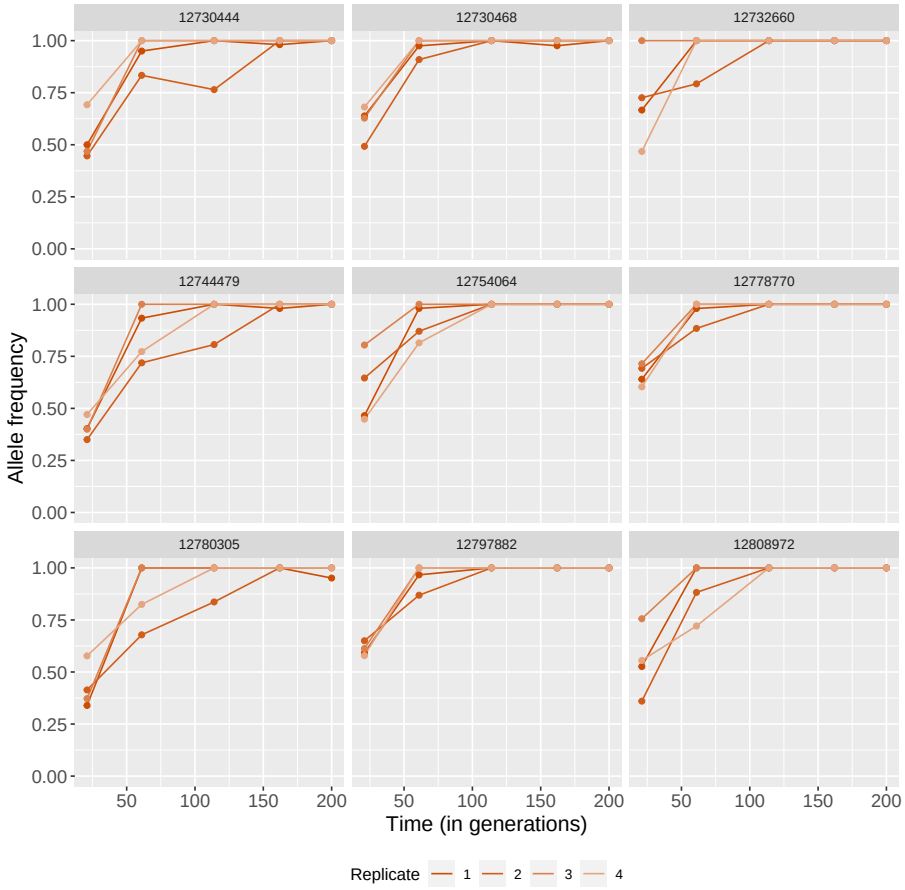


Figure 3.6: **Nine (out of 11) allele frequency trajectories of significant SNPs located on a single gene coding for neuromusculin.** Each replicate is coloured differently as per bottom legend. Individual variant coordinates can be found at the top of each graph.

3.4 Discussion

Sexual selection can cause substantial divergence between populations and even is hypothesised to be involved in speciation (Seehausen et al., 2014). It has been repeatedly implicated in altered ratios of genetic diversity between sex chromosomes and autosomes (e.g Corl and Ellegren, 2012). Here, we used an E&R experimental design in *D. pseudoobscura* to help elucidate the process of adaptation when the strength of sexual selection is altered in the short-term. We altered the intensity of sexual selection by reducing it in monogamous populations (M) or elevating it in a polyandrous regime (E). Signals of selection were strongest in E populations where sexual selection is elevated. This response is accompanied by a reduction in nucleotide diversity and alleles becoming fixed as time progresses. In addition, N_e estimates suffer a reduction as populations adapt, but recover towards neutral

levels.

While M lines should exhibit relaxed selection since competition for mates is eliminated, E lines are likely subject to elevated selection. Increasing the number of males a single female is housed with should cause sexual selection to be stronger. The elevated polyandry regime thus results from the observation that *D. pseudoobscura* are naturally polyandrous and each female has been found to mate with 2 to 3 males within its lifetime (Dobzhansky and Pavlovsky, 1967). Therefore, housing a female with 6 males instead should increase intrasexual competition. One would predict that increased promiscuity would facilitate the evolution of traits involved in mating or fertilisation success and perhaps pre-zygotic isolation mechanisms between the two treatments. Our *D. pseudoobscura* populations show a lack of assortative mating either between treatments or within lines (Debelle et al., 2016). A possible explanation for this might be that male-male competition overcomes the coevolution of female preference in this experiment. This observation could help generate expectations regarding some of the genomic response to selection. Under monogamy, competition for mates is eliminated. In a system where female preference is overshadowed by competition amongst males, sexually antagonistic selection is reduced to a minimum in M lines. If sexual conflict is promptly resolved, the adaptive response to enforced monogamy should be dominated by sexually antagonistic variation being drastically reduced.

Adaptation to an altered mating system will shape patterns of genetic variation in somewhat unpredictable ways. Understanding how allele frequency changes occur within given haplotype structures is instrumental to finding putative targets of selection. Is the signal of adaptation to sexual selection even throughout the genome and consistent across time? Our approach to understanding the adaptive process relied on taking snapshots of the replicates at several time points throughout the 200 generation experiment. These snapshots were allele frequencies estimated from a pool-seq dataset of each of the four replicate populations. Our *two-mapper two-variant caller* approach ensures that only high quality SNPs are present in both the full time-series and the two time point datasets. In particular, selection scan results are based on a time-series that is comprised of SNPs that were polymorphic at the first time sampling point. This causes our results to be focused on polymorphisms

with the most adaptive potential, since most would have possibly overcome the counteracting effects of drift over the first 20 generations of selection. In small populations such as these, drift will cause alleles to shift such that most low frequency polymorphisms will get lost within a few generations.

Taken together, our results support the hypothesis that strong sexual selection in E lines causes a substantial adaptive response. Not only did alleles become fixed more promptly in E populations (**fig. 3.1**), but also nucleotide diversity was depleted faster (**fig. 3.2**). In addition, our genome scan showed more than double the number of target candidates in E versus M. These top SNPs were found mostly across the X and the 3rd chromosomes (**fig. 3.5**). These results are consistent with the findings by Wiberg et al. (2021) whose SNPs showing significantly consistent allele frequency differences between E and M clustered along chromosomes 3 and X. The selection signatures we find are more pervasive in comparison to Wiberg et al.'s "islands" of differentiation. This is perhaps suggesting that LD has a substantial impact in our genome scan. In addition, effective population size estimates are consistent with a swift response to selection from the start of the experiment, especially in E. N_e estimated from allele frequency changes between time points 1 and 2 indicate that E lines suffer a more drastic reduction from the onset of selection.

At the end of the first half of the experiment - between time point 2 and 3 - E populations showed far more fixations than M lines (**fig. 3.1**). We hypothesised that this could result in a more marked response to selection in E which could, in turn, manifest as a reduction in N_e within that time interval. Interestingly, the pattern is reversed when comparing N_e between the two treatments: M lines have a much lower overall N_e on average than E. This result is statistically significant (M versus E at T2 to T3 Mann-Whitney U test p-value = 8×10^{-13}). This could mean that, despite more overall fixed loci, variance-N_e at the end of this first half indicates a more substantial reduction amongst M populations. A pattern such as this might be caused by one of two things. First, drift could be stronger in M overall resulting in drift variance that is picked up by the N_e estimator. Other monogamous regimes have been shown to result in lower overall N_e which augments the extent of drift (Wigby and Chapman, 2004). N_e in monogamous populations of *Drosophila melanogaster* lines was found to be only 16.2% smaller than that in polyandrous lines

(Rice and Holland, 2005). This prediction was tested previously in our lines and N_e was not significantly different between treatments (Snook et al., 2009). Secondly, our estimates could be biased if selection affects most allele frequency changes. This effect should dissipate if one would estimate N_e with sites that are evolving neutrally. We tried to overcome this by computing N_e using intergenic SNPs alone. General trends remained unchanged with similar median N_e for M and E between T2 and T3 as that using the complete dataset (M: 68.2 and E: 111.9; **table B.5**).

As costs of promiscuity arise from different mating frequency optima between the sexes, monogamy lines should show signs of resolved sexual conflict. This could, in turn, result in a more even distribution of nucleotide diversity along the genome, as well as less variance in N_e. Monogamous populations do exhibit less response to selection with fewer significant genome scan hits across the genome. This finding can be evidence for relaxed sexual selection due to mate competition being eliminated. Those regions showing selection signatures may have harboured a substantial portion of the sexually antagonistic variation. With most phenotypic optima that differed between the sexes having converged by the end of the experiment, such sexually antagonistic loci would have likely been targets of selection. Interestingly, median N_e in M throughout the experiment is severely reduced until time point 4 to values lower than those found in E. An accelerated rate of genetic drift due to monogamy cannot be ruled out here. However, this could, alternatively, be evidence of a delayed response where new phenotypic optima are reached towards the last quarter of the experiment.

Most of the adaptive signal is found on the X chromosome, which is unsurprising given that most genes involved in the response to both intra- and intersexual conflict are usually expected to be X-linked (Gibson et al., 2002; Connallon and Jakubowski, 2009; Mank and Ellegren, 2009). This signal is widespread across the X in E lines. Evidence for a faster-X effect (Charlesworth et al., 1987) is also supported by a low X/A N_e ratio during the first time point interval. Here, N_e estimates on the X are significantly lower than those on the autosomes (**fig. 3.4**). Indeed, the substitution of favourable mutations on the X appears swifter in comparison to autosomes. Diversity ratios are also predicted to be reduced as a result of hitchhiking favouring recessive beneficial alleles on the X (Betancourt et al., 2004), an effect that should

be elevated under polyandry (Pool and Nielsen, 2008). However, reduced recombination on chromosome X favours the increase in frequency of extended haplotypes. This could be causing the widespread signal that our genome scan is producing. Determining the extent of linkage disequilibrium in this experiment proves difficult as there is no data on the haplotypes present in the founder populations. Nevertheless, signatures of adaptation emerging from examining the X chromosome are still compelling.

Our results suggest that the predicted ¾ reduction in N_e on the X chromosome is not prevalent throughout the experiment. N_{eX} is just as high as on the autosomes in M lines regardless of the time point interval in question. Take the example of the T1 and T2 interval, median N_e is very similar for chromosomes 3, 4 and X (**fig. 3.4**). A similar trend is observed in E lines where N_{eX} is only found to be the lowest in comparison to any autosome between T1 and T2. If considering any other interval, N_e on the X is as high at that calculated for the autosomes. This is a striking result as theoretical studies predict a lower N_{eX} (Pool and Nielsen, 2008; Betancourt et al., 2004), especially under polyandry. Higher N_{eX} indicates that genetic diversity on the X is similar to that found in the autosomes. Accordingly, the elevated Tajima's D we observe throughout the genome is also consistent with an excess of heterozygosity. One possible explanation might be that balancing selection could be maintaining higher levels of polymorphism which could increase estimates of variance-N_{eX}. In such cases, pervasive balancing selection might arise especially if a large proportion of this variation is antagonistic between the sexes, which would favour heterozygotes. Alternatively, any excess heterozygosity may be caused by associative overdominance where there is an apparent heterozygote advantage (or pseudo-overdominance) at effectively neutral loci due to linked selection. Associative overdominance results from linkage desequilibrium between a neutral polymorphic locus and other loci under balancing selection (Ohta and Kimura, 1970). Purifying selection against recessive deleterious variants can also cause associative overdominance (Ohta, 1971).

As far as the X chromosome is concerned, the response found on the two chromosome arms differs between the treatments. M lines had a more pronounced response towards the distal end of the chromosome, whereas E lines' signal was more evenly

distributed along the chromosome. This suggests that the genetic basis of adaptation to either elevated or relaxed sexual selection on the X is distinct. This pattern would not be detected in most studies which simply compare divergence in allele frequencies between any two lines. The distal end of the X chromosome (chromosome arm XR in previous *D. pseudoobscura* assemblies) is equivalent to Muller element D in the *D. melanogaster* genome. This chromosome arm is known to have fused with the ancestral X chromosome to form the "neo-X". This could indicate that most of the sexually antagonistic variation is found on chromosome arm XR. The signatures found could be a signal of resolved sexual conflict. The ancestral X arm (XL) may have had enough evolutionary time to resolve any intragenomic conflict through mutation and recombination prior to the fusion. Moreover, both treatments exhibited a marked valley of signals of selection (**fig. 3.5**) in the centre of the X. This was coupled with a positive and elevated Tajima's D within the same region (**fig. 3.3**). Such a pattern suggests that, in spite of evidence for positive selection on both chromosome arms, the centromere region could be under balancing selection. This could perhaps be the result of sex-specific allele differentiation on the X between the two sexes. The centre of the X also contains some of the highest coverage regions across the genome (**fig. B.10**). This indicates that it might be a highly repetitive portion of the chromosome. Phased data from long read sequencing technology would be necessary to resolve this issue.

One other interesting finding is that populations recovered to neutral levels of N_e towards the middle or end of the experiment for E and M lines, respectively. Here, drift variance caused by any allele frequency changes that match neutral expectations might indicate the end of an initial strong selection phase. It is possible that phenotypic optima are reached at that point and allele frequencies might even plateau. In other words, directional selection becomes less effective towards the end of the experiment as populations reach the new phenotypic optima. Selection coefficients are reduced as polymorphism is eliminated and other modes of selection may arise. Stabilising selection will prevent trait means from moving away from the new optima. Again, balancing selection may now be present and act to maintain some genetic diversity. Any remaining genetic variation at the end behaved in such a way that allele frequency variance matched drift expectations. This is independent

from any previous allele fixations that might have occurred as a result of directional selection. While more in-depth studies of the genetic basis of sexual conflict would be required to elucidate this matter, our findings support the potential that it has largely been resolved.

Wiberg et al. (2021) found a cluster of top SNPs on chromosome 3 that showed significant differentiation between M and E lines. High levels of nucleotide diversity observed at the start of the experiment (**fig. B.8**) make chromosome 3 a good candidate for harbouring selection targets. This would facilitate adaptation due to increased fitness variance amongst individuals in the population. The region at the end of the chromosome was identified by Wiberg et al. as also showing a steep rate of decay in LD. This suggests that this peak region exhibits high recombination which is unexpected given that telomeres are typically low recombination regions. Our study confirms these results. There is evidence for positive directional selection within this region. Interestingly, the signal seems to be caused solely by directional selection in E lines. Increased recombination at the end of chromosome 3 relative to neighbouring areas could have contributed to the slightly elevated nucleotide diversity (**fig. B.8**). Additionally, this region also seems not to overlap with any known inversions (Wallace et al., 2011). However, it is still possible that extant structural variants have had an impact in our results which we have not accounted for here.

In summary, we showed that the response to an altered mating system in populations of *D. pseudoobscura* is found mostly to the X chromosome, but also on chromosome 3. This is consistent with previous work that focused on comparing E and M lines at time point 4 in our analysis (Wiberg et al., 2021). Selection signal is strongest when mate competition is strongest due to elevated polyandry. Such a pattern indicates that most allele frequency changes observed were in fact caused by elevated sexual selection and not solely adaptation to lab conditions. Our study showed the power of investigating allele frequency trajectories and their usefulness when estimating selection parameters and the effective population size.

Chapter 4

Discussion

4.1 Experimental evolution in the genomics era

Experimental evolution has been key in an era dominated by the search for the genetic architecture of the response to selection. E&R experiments have combined strong selection pressures with genome resequencing. These have also taken full advantage of replicated populations in order to understand allele frequency changes due to some selection regime. The question of how selection in the lab shapes allele frequencies has been the subject of heated debate for some time now. Adaptation to lab environments within tens of generations seems to cause phenotypic optima to be reached quickly. As phenotypes evolve, allele frequencies change to match them. Perhaps even whole haplotypes can increase in frequency as a result of the adaptive process (Franssen et al., 2015). How are these allele frequency trajectories shaped? How can we detect potential selection targets?

Several properties of adaptation to strong lab selection in the short-term have been uncovered as a consequence of decades of experimental evolution. The adaptive process depends mainly on population ancestry, on the effective population size and on the genetic basis of the traits in question. The field is rapidly moving towards unifying population genetics principles with quantitative trait modelling. Additionally, recognising sources of noise when analysing real time-series data is paramount. Understanding and modelling these sources can help improve the power of any statistical methods for detecting selection targets. Whilst some studies have

also drawn attention to the complexity of quantitative trait adaptation, we are only just starting to unveil the intricacies of the process.

In this thesis, I have investigated the genomic architecture of short-term adaptation in lab populations. I have studied properties of allele frequency trajectories over the course of tens of generations across experimental replicates. In particular, I worked to develop a statistical method for inferring selection parameters, thus confirming the usefulness of characterising full frequency trajectories. In addition, I analysed signatures of sexual selection in *Drosophila pseudoobscura* populations in an E&R setup. I compared the signal found on the X chromosome to that on the autosomes in populations where sexual selection was either relaxed or elevated.

4.2 Statistical power in an E&R framework

As more and more evolution experiments require appropriate statistical frameworks to analyse them, researchers have examined several aspects of the E&R experimental setup (Schlötterer et al., 2015; Kofler and Schlötterer, 2014; Rode et al., 2018). What could limit the power of an E&R experiment to detect signals of adaptation in the genome? Plenty of factors have been put forward to explain the low power of most tests to detect targets of selection. One specific aspect of pool-seq data is that there are multiple sources of sampling error associated with it. For example, sources of variation can arise from unequal contributions of individuals to the DNA pool. These can occur when sampling individuals from the population or during the DNA extraction procedure. If sampled individuals vary in body size or developmental stage, their respective DNA contribution might vary too. As far as noise is concerned, variance in sequencing depth can also affect allele frequency estimates. Rode et al. (2018) showed that even when samples are small, allele frequency estimates can be just as precise provided loci have enough coverage. Typically, DNA library preparation methods do not affect estimates more than 1% which is far less than the effect of coverage (Kofler et al., 2016b). This added variation can be especially troublesome when investigating low frequency variants as these might go unnoticed.

While some sources of pool-seq noise cannot be resolved until individual-level se-

quencing costs decrease substantially, statistical frameworks can be modified to better account for added variance. Our approach to circumventing sampling error was to incorporate a binomial (or, alternatively, beta-binomial) sampling process to observed Moran allele frequency states. Bait-ER produced accurate estimates of the selection coefficient, σ, for average and high simulated coverage (60x and 100x, respectively; **fig. 2.3C** and **fig. A.4**). This result suggests that Bait-ER is robust to varying sequencing depth, provided starting allele frequencies are not too low and populations are not too small. We also show that low statistical power in estimating σ for alleles that start at low frequencies can be overcome if there are enough experimental replicates (**fig. A.3**) or by increasing sequencing coverage. Moreover, Bait-ER's power to detect selected trajectories is not affected by coverage variation in real pool-seq time-series as Barghi et al.'s (2019) (**fig. A.24**). In general, we prove that our method avoids estimation bias caused by sampling error during pooled sequencing. As with other approaches, Bait-ER still produces inaccurate σ estimates for low starting frequency alleles especially in populations with 100 individuals or less.

Moreover, small lab populations can cause trouble of their own. First, if the founder population lacks diversity, experimental populations might not be able to reach the new phenotypic optimum set by the imposed selection pressure. A population lacking genetic diversity is a population with reduced genetic variance for fitness. Ultimately, this will hinder adaptation to the new environment. Secondly, linkage disequilibrium throughout the genome will most likely be extensive. If a population is small and the experiment is short, extended haplotypes will probably be passed down. This is because recombination has not had the ability to breakup these haplotypes. Recombination is low compared to selection in experimental evolution. Consequently, the power of genome scan-type methods will potentially be reduced as most do not take LD into account. As peaks of statistically significant candidate SNPs are spread throughout the genome, it is difficult to pinpoint the actual causative locus (or loci). Finally, as genetic drift is stronger in small populations, beneficial alleles might be lost causing the response to selection to be slower. This can result in soft or partial sweeps being the most common signature. As most methods search for sweep-like trajectories, drift can become a limitation.

As mentioned above, for a method such as Bait-ER, strong drift can be inconvenient as it might result in biased σ estimates. The issue lies in low starting frequency variants. In large populations, these variants rarely overcome the drift barrier. There is then a small proportion of variants that reaches a high enough frequency for selection to cause them to increase in frequency and, ultimately, to fix. In small populations, however, drift may cause considerable allele frequency shifts from one generation to the next. Consequently, alleles may fix or get lost within a few generations. This affects σ estimates because it will change the shape of allele frequency trajectories. Estimates will likely be downwardly biased since a larger proportion of ancestral polymorphism will appear to be swiftly fixed by selection. We have shown that drift does affect estimates for low frequency variants in populations with up to 300 individuals. However, Bait-ER's accuracy is unaffected by more prevalent alleles (from 0.5%; **fig. 2.3**). This is true regardless of the size of the population one considers. While Bait-ER is robust to drift variance in small populations, its performance is still affected by LD. In the laboratory, populations are more prone to high levels of LD due to reduced census sizes and limited generations of evolution. Taken together, our results on Barghi et al.'s (2019) *D. simulans* experiment and our study on *D. pseudoobscura* (described in chapter 3) suggest that there is substantial LD given the overall elevated logBF signal throughout the genome (**figs. 2.8, 3.5 and A.10**). Applying a suitably conservative threshold to Bait-ER's logBF results guarantees most of the hitchhiking noise is removed. We are then left with a few very significant peaks. This indicates that, despite extensive LD, Bait-ER is still identifying the significant regions where there is marked response to selection. However, our genome scan is less conclusive at detecting individual loci that might be potential targets of selection. Strategies to improve this might involve investigating genes within significant peaks. This could help understand what gene networks could be involved in the response to the selection regime. Gene ontology analyses can be further complemented with expression studies and gene knockout assays.

Some statistical methods for detecting genomic selection in E&R experiments have attempted to look for consistent behaviour across experimental replicates. Alleles that start at a low frequency are problematic when quantifying shared responses across replicates. Such alleles are more likely to be lost by drift. If low starting

frequency variants are the most likely selection targets, the lack of parallel responses across replicate populations will certainly reduce the power of any statistical test that accounts for it. Strong genetic drift will affect the rate at which diversity is eroded, causing these low frequency alleles to be even more readily lost. Ultimately, larger populations show a higher degree of repeated evolution compared to smaller populations. Bait-ER searches for consistency across replicate populations insofar as it calculates the likelihood of a given selection coefficient as a product of all replicates. In other words, strongly selected trajectories are those that exhibit a similarly shaped trajectory across replicates. This is evident in our genome-scan of the *D. pseudoobscura* time-series where top significant SNP trajectories were quite similar resulting in alelle fixations on all four replicates (**fig. 3.6**).

One way to avoid making assumptions about the shape of a selected allele's trajectory is to develop methods that model phenotypic selection. Gompert (2016) started off by developing a hidden Markov model to estimate variance-based N_e and quantify selection in heterogeneous environments. Gompert's method relies on the association between allele frequency changes and some environmental variable at each generation. N_e is estimated using a Bayesian bootstrap approach coupled with a linear model to estimate selection coefficients. Gompert (2021) has expanded this idea of allele frequency change and environmental covariance to model phenotypic selection as an explicit function of the state of the environment in an ABC approach. This method models and quantifies fluctuating selection on complex traits. Here, phenotypic selection depends on an understanding of the trait's genetic architecture, which requires a genotype-phenotype association study. The method assumes that mutation and migration are negligible and that selection is directional. Additionally, each locus only explains a small proportion of the trait variance and any causal variants are unlinked. This is a different approach to estimating the effects of selection as it models selection acting directly on the phenotype rather than the genetic process of adaptation via allele frequency changes. It does not remove the genotype from its organismal context as it tries to explain phenotypic change through time. However, current evolution experiments still lack the appropriate experimental setups to collect the required data. It would be necessary to perform some thorough GWA study using founder populations in order to find any variants that might be

associated with the complex trait of interest. Such an experimental design would become feasible as individual-level sequencing costs decrease with the release of new high-throughput sequencing technologies and new library preparation methods like haplotagging (Meier et al., 2021). While this might still be a limitation, it is certainly one to inform on how phenotypic data collection should be set up along with genome resequencing in the future.

4.3 Architecture of complex traits

One would argue that selective sweeps should dominate in E&R datasets owing to strong selection pressures, but how common are selective sweeps if most quantitative traits have a complex genetic basis? In a system where selection on a trait is strong, it is still possible that selection on individual genes is weak (Walsh and Lynch, 2018), especially if we are studying complex traits. Jain and Stephan (2017a,b) model a quantitative trait subject to polygenic adaptation after a sudden shift in the phenotypic optimum. The strength of directional selection depends on the distance to the new optimum. They find that dramatic changes in an allele's frequency - similar to a sweep-like trajectory - can occur if the size of the phenotypic effect is suitably large. When most effects were large, considerable frequency shifts over a long timescale were also periodically found. These shifts resembled sweeps but did not occur within the short-term as sweeps are typically expected to. This work thus suggests that subtle allele frequency shifts are not necessarily the rule in the adaptive process of quantitative traits.

The results I presented throughout this thesis are consistent with this prediction that sweep-like trajectories are still frequent in E&R time-series. Bait-ER's genome scan for selection in both Barghi et al.'s (2019) and our *D. pseudoobscura* experiment is powerful enough to detect some very significant peaks (**figs. 2.8, 3.5 and A.10**). In the case of Barghi et al. (2019), Bait-ER produces peaks that are largely in agreement with Barghi et al.'s original CMH test analysis (**figs. A.11 and A.12**). The most striking difference is that, for an equivalent significance threshold, the CMH test is substantially more lenient resulting in more false positives. This could be because Bait-ER's most significant trajectories are those that show consistent sweep-like

behaviour across replicates. The CMH test has been previously shown to have a high false positive rate since a real effect is often confounded with heterogeneity amongst replicates (Wiberg et al., 2017). It is also affected by overdispersion causing p-values to be systematically low (Spitzer et al., 2020). In addition, our results for the *D. pseudoobscura* genomic time-series paint an even clearer picture of adaptation (**fig. 3.5**). The response to selection was evidently more marked in the regime where sexual selection is elevated (E lines), and it is mostly found on chromosomes 3 and X. The data was parsed as to include only sites that are polymorphic at time point 1. Since 20 generations of selection had passed since the onset of selection, the dataset is reduced in comparison to Barghi et al.'s. This could cause the signal to be clear since low frequency variants might have already been lost and thus not be part of the time-series. Nevertheless, there are distinctive peaks marking regions containing putative selected targets.

The adaptive process can also be shaped by epistatic interactions between loci. When causative alleles at separate loci are inherited non-independently due to fitness constraints, overall selection strength on a specific locus can be attenuated. Hill-Robertson Interference (HRI) describes the phenomenon whereby multiple loci under selection interfere with each other (Hill and Robertson, 1966). This causes a reduction in the efficacy of selection on a specific locus - a property that arises because the effective population size of that locus decreases. Theoretical predictions suggest that epistatic effects of a certain strength will affect linkage disequilibrium the most when selection is weak (Kouyos et al., 2006). As recombination can only aid selection if it increases genetic variance in fitness, it can play a decisive role in such cases. It does so by breaking down negative associations between advantageous variants and thus generating fitter combinations of alleles (Barton, 2010). Whilst very important in the long-term, adaptation in the short-term might be hindered by epistasis and low recombination rates.

The extent to which pleiotropy can cause trade-offs between correlated traits is largely undetermined for most complex traits. A model where traits can be correlated via gene network pleiotropy has been put forward as an explanation for this phenomenon (Boyle et al., 2017). The omnigenic model for a quantitative trait assumes that the trait in question is affected by a modest number of genes. However,

these are still constrained by cell regulatory networks that regulate these genes' expression. These networks are themselves interconnected. Whilst the original model was thought of in the context of disease, it is likely to be relevant for other complex traits. In an E&R context, one could predict that regulatory elements, such as promoter and enhancer sequences, would consistently show signatures of adaptation as these contribute to the fine-tuning of cell regulatory element expression. Moreover, those genes with the largest fitness effects might show up in genome scan analyses as potential candidates as they are the most likely elements in this model to show a sweep-like trajectory.

In an E&R setup, there may be several haplotypes segregating in the founder population. If that is the case, the adaptive response amongst replicates might show genetic redundancy. Redundancy is when more than one combination of segregating polymorphisms is able to produce the same phenotype (also referred to as genotypic redundancy; see Láruson et al. 2020). If we are concerned with polygenic traits, there might be enough redundancy amongst small effect loci that each haplotype is completely different from the other. This is in line with the quantitative genetics paradigm where small effect loci will dominate trait adaptation if the starting population is not very far from the fitness optimum. Genetic redundancy is an effect that requires sequencing founder populations to fully characterise the extant haplotypes. Fitness assays of founder haplotypes further contribute to understanding adaptation in the presence of redundancy.

Let us now focus on the length of an evolution experiment. As with most complex traits, it is difficult to pinpoint when a population has reached a new phenotypic optimum, partly because we do not know what that optimum is. We are typically able to quantify changes in fitness to some degree by investigating longevity or offspring number. However, if we are concerned with behavioural traits, for example, it is challenging to predict what the phenotypic outcome should be. Unless we perform a real-time analysis of allele frequency changes, it remains problematic to map these to any phenotypic differences through time. We are thus left with making an educated guess regarding how long our study should be. In experiments where populations of *Drosophila* flies were subject to strong lab selection, studies found that adaptation was complete by 60 generations (Barghi et al., 2017; Hsu et al.,

2020). Others chose to characterise changes for over 600 generations (Burke et al., 2010). Perhaps longer experiments experience a third mode of selection that is balancing selection. This phase would follow that where directional and stabilising selection have moved the population to the new phenotypic optimum. Such a phase would be characterised by an overall increase in genetic diversity and even N_e. As directional selection becomes less effective in altering allele frequencies, additive genetic variance begins to accumulate as replicate populations diverge freely.

A simulation study has helped illustrate this step process of genomic adaptation in the context of the E&R framework. Franssen et al. (2017b) simulated the evolution of a quantitative trait under a typical E&R experimental setup. They found that, after some environmental change, there were three phases of adaptation to a new trait optimum. The first, where there are directional frequency changes until the population reaches the new optimum, is followed by a second phase where allele frequencies plateau and only small shifts are observed. The third and final step is when replicate populations begin to diverge possibly even at causative loci. These trends are consistent with selection changing its mode as time progresses. Initially, there is directional selection causing drastic allele frequency changes. Towards the end, populations should be fully adapted as additive genetic variance for fitness is reduced by beneficial alleles reach fixation. This step is followed by stabilising selection keeping alleles from changing in frequency any further. Our drift-N_e analysis across the *D. pseudoobscura* time-series is consistent with a phased adaptive process that ends with stabilising selection acting to maintain genotypes at the new fitness optimum. After 100-160 generations of selection, N_e recovered to neutral levels similar to the census population size. This could only be the case if the remaining polymorphism experienced a relaxation of selection where alleles changed in frequency similar to drifting variants.

4.4 Sexual selection in *Drosophila pseudoobscura*

Sexual selection is thought to affect autosomes and sex chromosomes differently. Sex differences can be further promoted by sex chromosome inheritance and generate sexual conflict. These differences can include sexually antagonistic selection,

sex-specific mutation (Li et al., 2002; Charlesworth et al., 2018) and recombination rates (Sardell and Kirkpatrick, 2020), as well as sex-specific dominance coefficients (Spencer and Priest, 2016) and sex-specific variance in reproductive success Caballero (1995). Population size changes (Pool and Nielsen, 2007) and sex-biased migration (Laporte and Charlesworth, 2002) may also promote differences between autosomes and sex chromosomes. Any of these factors can contribute to the maintenance or reduction of genetic diversity and, thus, change the outcome of selection. In the context of our *D. pseudoobscura* experiment (chapter 3), we have gathered predictions about X:Autosome (or X:A) diversity ratios. This is especially relevant since the outcome of the experiment was affected not only by positive selection but also by other evolutionary and demographic factors. In general, N_e is predicted to be reduced to $^3/_4$ on the X chromosome relatively to the autosomes. This applies to large populations with an equal sex ratio. Our results contradict this expectation where N_{eX} is almost always just as high on autosomes. Overall, more signatures of adaptation were found in E lines where sexual selection is elevated. The response to selection is especially marked on the X in both treatments. Below is a discussion of our findings in a wider context.

The X chromosome is conventionally thought to value female fitness twice as much as male fitness in comparison to autosomes that have been viewed as placing equal value on each of the sexes. This is a direct consequence of the X spending $^2/_3$ of its evolutionary time in the bodies of females and only $^1/_3$ in males (Rice, 1984; Charlesworth et al., 1987). Intuitively, in XY systems, the X should favour mean phenotypes that are closer to the female optimum, e.g. feminisation of the transcriptome under monogamy, whereas the autosomes should favour phenotypes closer to the male optimum for a trait expressed in both sexes (Haig, 2006). However, an X-linked recessive allele that is advantageous in males will increase in frequency even if deleterious in females (Rice, 1984). Male-beneficial alleles have been shown to invade a population more readily (Patten, 2019). The X chromosome can accumulate more beneficial mutations because the effect of recessive (or partially recessive) mutations is not masked by the ancestral alleles in the heterogametic sex (Charlesworth et al., 1987). This faster-X effect can thus contribute to the fixation of sexually antagonistic mutations.

Under monogamy, variance in reproductive success between the sexes is identical. This allows one to make predictions about the expected levels of genetic diversity, N_e and sexually antagonistic variation (reviewed in Ellegren, 2009). In a scenario where there are multiple sweeps of positively selected mutations, adaptive rates on the X are predicted to be higher than on the autosomes (Betancourt et al., 2004). Evidence for a faster-X effect has been found across several taxa (reviewed in Meisel and Connallon (2013)). *Drosophila* show the most similar adaptive rate between X and autosomal loci, with mammals and birds having a more marked faster-X effect (Mank et al., 2010). It is pervasive amongst the *melanogaster* clade (Ávila et al., 2014), and it has been shown to be prevalent in female-biased genes (Avila et al., 2015; Campos et al., 2018) contrary to theoretical expectations (Vicoso and Charlesworth, 2009).

The predictions described above hold only in large populations where drift is negligible. In finite populations where drift dominates, a larger fraction of mutations is effectively neutral. This effect coupled with a lower N_e contributes to a decrease in the effectiveness of selection on the X. Moreover, low overall recombination due to the short-term nature of the experiment as well as a small population size can contribute to this effect (Betancourt et al., 2009). If the X chromosome is to be responsible for most of the adaptive response in our *D. pseudoobscura* lines, low N_e might hinder adaptation substantially.

Long-term diversity ratios of X/autosome diversity for *D. pseudoobscura* were found to be roughly 0.81 and relative N_e ratios to be ¾ (Haddrill et al., 2011), which matches neutral expectations under an even sex ratio mating system (Charlesworth, 2012). Our experimental evolution results for the monogamy regime show drift N_{eX}/N_{eA} ratios of ≈ 1 regardless of the time interval considered. In other words, N_e on the X is similar to N_e on the autosomes. This suggests that there is no faster-X effect in M lines which would be consistent with a relaxation of sexual selection under monogamy. In addition, if beneficial mutations are dominant (or partially dominant), the X will be more variable than the autosomes, which could also cause similar N_{eX} and N_{eA}.

An overall moderate adaptive response in M lines might have alternative explana-

tions. On the one hand, it might be a direct consequence of background selection which is expected to reduce autosomal more than X-linked diversity (Charlesworth, 1994). This would be even more marked if slightly deleterious mutations are able to drift to appreciable frequencies due to associative overdominance caused by linkage to sites under balancing (Ohta and Kimura, 1970) or purifying selection (Ohta, 1971). On the other hand, moderate signals of adaptation could be caused by selection varying in time as the populations navigate through the adaptive landscape.

Predictions for polyandrous populations are less straightforward. In a scenario where females are polyandrous and populations experience consecutive bottlenecks, X-linked diversity is expected to be further reduced relative to autosomal levels (Pool and Nielsen, 2008). This scenario does not match our experimental setup where *D. pseudoobscura* populations experience a single bottleneck at the time of sampling and are kept at a constant census size throughout the experiment. Having said that, one would perhaps still predict a reduced N_{eX}/N_{eA} ratio in E lines (≈ 0.29) relatively to the $^3/_4$ ratio expected under monogamy as there are far fewer copies of the X in comparison to autosomal copies in a male-biased population. This might result in adaptation being less effective on the X under polyandry. However, this seems to be the opposite to what we found. E lines show a stronger selection signal relative to M populations (**fig. 3.5**). This indicates that, in spite of a biased sex ratio that could weaken the effectiveness of selection, selection was pervasive in the polyandry regime especially on the X chromosome.

The two selection regimes in this *D. pseudoobscura* experiment were predicted to differ mainly in the intensity of sexual conflict. The adaptive response in monogamy lines is expected to consistently work towards resolving most existing sexual conflict as male-male competition was instantly eliminated. Under monogamy, interests shared between the sexes should converge since variance in reproductive success is identical in males and females. M males were found to lead to greater oviposition and hatching rates when mated to ancestral females (Crudgington et al., 2005), which is consistent with monogamous male mating behaviour converging towards female optima. By contrast, flies reared under elevated polyandry experience stronger conflict as males have to compete for access to the female. As E males evolved a higher courtship frequency (Snook et al., 2005) and a greater mating capacity (Crudging-

ton et al., 2009), conflict over the optimal mating frequency escalated with females mating at suboptimal rates. Such elevated conflict under polyandry could lead to more adaptive signal on male-biased genes. Sexual antagonism for traits expressed in both sexes could, in turn, cause further sex-biased gene expression (Connallon and Knowles, 2005). Resolution of sexual conflict in E lines would require enough generations for male and female expression levels to be decoupled. However, pleiotropy for these traits is likely to be extensive which would hinder this process.

Taken together, our results seem to indicate that response to an altered mating system is mostly located on the X chromosome. This is consistent with previous reports of the X being a hot spot for sexually antagonistic selection (Gibson et al., 2002). Despite demographic factors and a male-biased sex ratio hindering adaptation on the X, E lines still experience a faster-X effect. This result contradicts earlier findings that adaptive evolution should be slower on the X relative to the autosomes should populations adapt from standing genetic variation (Orr and Betancourt, 2001). If adaptation starts at mutation-selection balance, 'Haldane's sieve' against recessive mutations should not hold. Adaptation from SGV would, thus, reduce the added effectiveness of selection on the X due to hemizygosity. The faster-X effect we observe could be the result of a founder population not at equilibrium. As founder populations suffered a bottleneck when collected from the field and established in the lab, four generations of common-garden might have not been enough for these to be fully adapted to the new laboratory environment. On the question of dominance, it is possible that sex-specific dominance reversal could have maintained enough fitness variance in the founder populations that facilitated adaptation (Grieshop and Arnqvist, 2018).

4.5 Concluding remarks

Allele frequency dynamics over the short-term is a very active field of research. Whilst long-term trajectories have been thoroughly characterised, the behaviour of selection targets in lab experiments remains elusive. Statistical approaches that tackle essential issues to do with evolutionary constraints in real experimental populations are increasing in complexity. This is coupled with an increasing knowledge

of the deterministic and stochastic factors that come into play.

During these times of increasing data availability, it is crucial for researchers to produce user-friendly software packages. This facilitates their usage by other scientists who might not have extensive training in bioinformatics. Making toy data sets available along with easy to follow tutorials is key. One other aspect that would improve software usage would be to agree on a specific input data format. This would be paramount to make sure software comparison is transparent and encouraged. Reproducible research should be one of our aims as a scientific community.

With sequencing technology costs sharply decreasing, now is a very exciting time to explore time series data. Studying time-series in the lab has illustrated the complexity of the genomic architecture of adaptation. Investigating patterns of LD and how these are associated with the signal of selection identified by genome scans is the next logical step for the field. Data from founder populations until the end of the adaptive response coupled with phenotypic data should provide us with a clear picture of short-term adaptation. This will feed back into our statistical methodology and software, allowing us to extend its application to populations both in the lab and in the wild in the future.

Chapter 5

Bibliography

A. Agresti. *Categorical data analysis*. John Wiley & Sons, Hoboken, New Jersey, USA, 2003. ISBN 978-0-470-46363-5.

S. Anand, E. Mangano, N. Barizzone, R. Bordoni, M. Sorosina, F. Clarelli, L. Corrado, F. M. Boneschi, S. D'Alfonso, and G. De Bellis. Next generation sequencing of pooled samples: Guideline for variants' filtering. *Scientific Reports*, 6(August): 1–9, 2016.

V. Ávila, S. Marion de Procé, J. L. Campos, H. Borthwick, B. Charlesworth, and A. J. Betancourt. Faster-X Effects in Two Drosophila Lineages. *Genome Biology and Evolution*, 6(10):2968–2982, 2014.

V. Avila, J. L. Campos, and B. Charlesworth. The effects of sex-biased gene expression and X-linkage on rates of adaptive protein sequence evolution in Drosophila. *Biology Letters*, 11(4):20150117–20150117, 2015.

J. G. Baldwin-Brown, A. D. Long, and K. R. Thornton. The power to detect quantitative trait loci using resequenced, experimentally evolved populations of diploid, sexual organisms. *Molecular Biology and Evolution*, 31(4):1040–1055, 2014.

C. Barata, R. Borges, and C. Kosiol. Bait-ER: A Bayesian method to detect targets of selection in evolve-and-resequence experiments. *bioRxiv*, 2020. URL https://doi.org/10.1101/2020.12.15.422880.

N. Barghi, R. Tobler, V. Nolte, and C. Schlötterer. Drosophila simulans: A Species with Improved Resolution in Evolve and Resequence Studies. *G3: Genes—Genomes—Genetics*, 7(7):2337–2343, 2017.

N. Barghi, R. Tobler, V. Nolte, A. M. Jakšić, F. Mallard, K. A. Otte, M. Dolezal, T. Taus, R. Kofler, and C. Schlötterer. Genetic redundancy fuels polygenic adaptation in Drosophila. *PLOS Biology*, 17(2):e3000128, 2019.

N. Barghi, J. Hermisson, and C. Schlötterer. Polygenic adaptation: a unifying framework to understand positive selection. *Nature Reviews Genetics*, 21(12): 769–781, 2020.

J. E. Barrick and R. E. Lenski. Genome dynamics during experimental evolution. *Nature Reviews Genetics*, 14(12):827–839, 2013.

J. E. Barrick, D. S. Yu, S. H. Yoon, H. Jeong, T. K. Oh, D. Schneider, R. E. Lenski, and J. F. Kim. Genome evolution and adaptation in a long-term experiment with Escherichia coli. *Nature*, 461(7268):1243–1247, 2009.

N. H. Barton. Genetic linkage and natural selection. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 365(1552):2559–2569, 2010.

H. Bastide, A. Betancourt, V. Nolte, R. Tobler, P. Stöbe, A. Futschik, and C. Schlötterer. A Genome-Wide, Fine-Scale Map of Natural Pigmentation Variation in Drosophila melanogaster. *PLOS Genetics*, 9(6):e1003534, 2013.

A. J. Berry, J. W. Ajioka, and M. Kreitman. Lack of polymorphism on the drosophila fourth chromosome resulting from selection. *Genetics*, 129(4):1111–1117, 1991.

A. J. Betancourt, Y. Kim, and H. A. Orr. A pseudohitchhiking model of X vs. autosomal diversity. *Genetics*, 168(4):2261–2269, 2004.

A. J. Betancourt, J. J. Welch, and B. Charlesworth. Reduced Effectiveness of Selection Caused by a Lack of Recombination. *Current Biology*, 19(8):655–660, 2009.

C. R. B. Boake, S. J. Arnold, F. Breden, L. M. Meffert, M. G. Ritchie, B. J. Taylor, J. B. Wolf, and A. J. Moore. Genetic Tools for Studying Adaptation and the Evolution of Behavior. *The American Naturalist*, 160(S6):S143–S159, 2002.

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

J. P. Bollback, T. L. York, and R. Nielsen. Estimation of 2Nes from temporal allele frequency data. *Genetics*, 179(1):497–502, 2008.

E. A. Boyle, Y. I. Li, and J. K. Pritchard. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell*, 169(7):1177–1186, 2017.

V. Buffalo and G. Coop. The Linked Selection Signature of Rapid Adaptation in Temporal Genomic Data. *Genetics*, 213(3):1007–1045, 2019.

V. Buffalo and G. Coop. Estimating the genome-wide contribution of selection to temporal allele frequency change. *Proceedings of the National Academy of Sciences of the United States of America*, 117(34):20672–20680, 2020.

M. K. Burke and A. D. Long. What paths do advantageous alleles take during short-term evolutionary change? *Molecular Ecology*, 21(20):4913–4916, 2012.

M. K. Burke, J. P. Dunham, P. Shahrestani, K. R. Thornton, M. R. Rose, and A. D. Long. Genome-wide analysis of a long-term evolution experiment with Drosophila. *Nature*, 467(7315):587–590, 2010.

M. K. Burke, G. Liti, and A. D. Long. Standing genetic variation drives repeatable experimental evolution in outcrossing populations of saccharomyces cerevisiae. *Molecular Biology and Evolution*, 31(12):3228–3239, 2014.

A. Caballero. On the effective size of populations with separate sexes, with particular reference to sex-linked genes. *Genetics*, 139(2):1007–1011, 1995.

J. L. Campos, K. J. A. Johnston, B. Charlesworth, and K. Kunte. The Effects of Sex-Biased Gene Expression and X-Linkage on Rates of Sequence Evolution in Drosophila. *Molecular Biology and Evolution*, 35(3):655–665, 2018.

T. Chapman and S. J. Davies. Functions and analysis of the seminal fluid proteins of male Drosophila melanogaster fruit flies. *Peptides*, 25(9):1477–1490, 2004.

T. Chapman, L. F. Liddle, J. M. Kalb, M. F. Wolfner, and L. Partridge. Cost of mating in Drosophila melanogaster females is mediated by male accessory gland products. *Nature*, 373(6511):241–244, 1995.

B. Charlesworth. The effect of background selection against deleterious mutations on weakly selected, linked variants. *Genetical Research*, 63(3):213–227, 1994.

B. Charlesworth. The role of background selection in shaping patterns of molecular evolution and variation: Evidence from variability on the Drosophila X chromosome. *Genetics*, 191(1):233–246, 2012.

B. Charlesworth and J. D. Jensen. Effects of Selection at Linked Sites on Patterns of Genetic Variability. *Annual Review of Ecology, Evolution, and Systematics*, 52: 177–197, 2021.

B. Charlesworth, J. A. Coyne, and N. H. Barton. The Relative Rates of Evolution of Sex Chromosomes and Autosomes. *The American Naturalist*, 130(1):113–146, 1987.

B. Charlesworth, J. L. Campos, and B. C. Jackson. Faster-X evolution: Theory and evidence from Drosophila. *Molecular Ecology*, 27(19):3753–3771, 2018.

C. Cheng and M. Kirkpatrick. Sex-Specific Selection and Sex-Biased Gene Expression in Humans and Flies. *PLOS Genetics*, 12(9):e1006170, 2016.

S. F. Chenoweth and M. W. Blows. Contrasting Mutual Sexual Selection on Homologous Signal Traits in Drosophila serrata. *The American Naturalist*, 165(2): 281–289, 2005.

S. F. Chenoweth, N. C. Appleton, S. L. Allen, and H. D. Rundle. Genomic Evidence that Sexual Selection Impedes Adaptation to a Novel Environment. *Current Biology*, 25(14):1860–1866, 2015.

A. K. Chippindale, T. J. Chu, and M. R. Rose. Complex trade-offs and the evolution of starvation resistance in Drosophila melanogaster. *Evolution*, 50(2):753–766, 1996.

A. K. Chippindale, J. R. Gibson, and W. R. Rice. Negative genetic correlation for adult fitness between sexes reveals ontogenetic conflict in Drosophila. *Proceedings of the National Academy of Sciences of the United States of America*, 98(4):1671–1675, 2001.

G. Cobbs. Multiple Insemination and Male Sexual Selection in Natural Populations of Drosophila pseudoobscura. *The American Naturalist*, 111(980):641–656, 1977.

W. G. Cochran. Some Methods for Strengthening the Common χ 2 Tests. *Biometrics*, 10(4):417–451, 1954.

J. M. Comeron, R. Ratnappan, and S. Bailin. The Many Landscapes of Recombination in Drosophila melanogaster. *PLOS Genetics*, 8(10):e1002905, 2012.

T. Connallon and E. Jakubowski. Association between sex ratio distortion and sexually antagonistic fitness consequences of female choice. *Evolution*, 63(8):2179–2183, 2009.

T. Connallon and L. L. Knowles. Intergenomic conflict revealed by patterns of sex-biased gene expression. *Trends in Genetics*, 21(9):495–499, 2005.

A. Corl and H. Ellegren. The genomic signature of sexual selection in the genetic diversity of the sex chromosomes and autosomes. *Evolution*, 66(7):2138–2149, 2012.

H. S. Crudgington, A. P. Beckerman, L. Brüstle, K. Green, and R. R. Snook. Experimental Removal and Elevation of Sexual Selection: Does Sexual Selection Generate Manipulative Males and Resistant Females? *The American Naturalist*, 165(S5):S72–S87, 2005.

H. S. Crudgington, S. Fellows, N. S. Badcock, and R. R. Snook. Experimental manipulation of sexual selection promotes greater male mating capacity but does not alter sperm investment. *Evolution*, 63(4):926–938, 2009.

H. S. Crudgington, S. Fellows, and R. R. Snook. Increased opportunity for sexual conflict promotes harmful males with elevated courtship frequencies. *Journal of Evolutionary Biology*, 23(2):440–446, 2010.

L. Czech and M. Exposito-Alonso. Grenedalf: Genome Analyses of Differential Allele Frequencies. *Manuscript in preparation*, 2021.

P. Danecek, J. K. Bonfield, J. Liddle, J. Marshall, V. Ohan, M. O. Pollard, A. Whitwham, T. Keane, S. A. McCarthy, R. M. Davies, and H. Li. Twelve years of SAMtools and BCFtools. *GigaScience*, 10(2):1–4, 2021.

C. Darwin. *The descent of man, and selection in relation to sex*. Princeton Univ. Press, Princeton, NJ, 1871.

A. Debelle, M. G. Ritchie, and R. R. Snook. Evolution of divergent female mating preference in response to experimental sexual selection. *Evolution*, 68(9):2524–2533, 2014.

A. Debelle, M. G. Ritchie, and R. R. Snook. Sexual selection and assortative mating: An experimental test. *Journal of Evolutionary Biology*, 29(7):1307–1316, 2016.

A. Debelle, A. Courtiol, M. G. Ritchie, and R. R. Snook. Mate choice intensifies motor signalling in Drosophila. *Animal Behaviour*, 133:169–187, 2017.

M. Delcourt, M. W. Blows, and H. D. Rundle. Sexually antagonistic genetic variance for fitness in an ancestral and a novel environment. *Proceedings of the Royal Society B: Biological Sciences*, 276(1664):2009–2014, 2009.

T. Dobzhansky and O. Pavlovsky. Repeated Mating and Sperm Mixing in Drosophila pseudoobscura. *The American Naturalist*, 101(922):527–533, 1967.

H. Ellegren. The different levels of genetic diversity in sex chromosomes and autosomes. *Trends in Genetics*, 25(6):278–284, 2009.

E. Elyashiv, S. Sattath, T. T. Hu, A. Strutsovsky, G. McVicker, P. Andolfatto, G. Coop, and G. Sella. A Genomic Map of the Effects of Linked Selection in Drosophila. *PLOS Genetics*, 12(8):e1006130, 2016.

J. C. Fay and C. I. Wu. Hitchhiking under positive Darwinian selection. *Genetics*, 155(3):1405–13, 2000.

A. F. Feder, S. Kryazhimskiy, and J. B. Plotkin. Identifying signatures of selection in genetic time series. *Genetics*, 196(2):509–522, 2014.

A. Ferrer-Admetlla, C. Leuenberger, J. D. Jensen, and D. Wegmann. An approximate markov model for the wright–fisher diffusion and its application to time series data. *Genetics*, 203(2):831–846, 2016.

A. R. T. Figueiredo, A. Wagner, and R. Kümmerli. Ecology drives the evolution of diverse social strategies in Pseudomonas aeruginosa. *Molecular Ecology*, 30(20): 5214–5228, 2021.

R. C. Firman and L. W. Simmons. Experimental evolution of sperm quality via postcopulatory sexual selection in house mice. *Evolution*, 64(5):1245–1256, 2010.

M. Foll, Y.-P. Poh, N. Renzette, A. Ferrer-Admetlla, C. Bank, H. Shim, A.-S. Malaspinas, G. Ewing, P. Liu, D. Wegmann, D. R. Caffrey, K. B. Zeldovich, D. N. Bolon, J. P. Wang, T. F. Kowalik, C. A. Schiffer, R. W. Finberg, and J. D. Jensen. Influenza virus drug resistance: a time-sampled population genetics perspective. *PLOS genetics*, 10(2):e1004185, 2014.

M. Foll, H. Shim, and J. D. Jensen. WFABC: A Wright-Fisher ABC-based approach for inferring effective population sizes and selection coefficients from time-sampled data. *Molecular Ecology Resources*, 15(1):87–98, 2015.

S. U. Franssen, V. Nolte, R. Tobler, and C. Schlotterer. Patterns of linkage disequilibrium and long range hitchhiking in evolving experimental drosophila melanogaster populations. *Molecular Biology and Evolution*, 32(2):495–509, 2015.

S. U. Franssen, N. H. Barton, and C. Schlotterer. Reconstruction of haplotype-blocks selected during experimental evolution. *Molecular Biology and Evolution*, 34(1):174–184, 2017a.

S. U. Franssen, R. Kofler, and C. Schlötterer. Uncovering the genetic signature of quantitative trait evolution with replicated time series data. *Heredity*, 118(1): 42–51, 2017b.

J. D. Fry. The Genomic Location of Sexually Antagonistic Variation: Some Cautionary Comments. *Evolution*, 64(5):1510–1516, 2009.

A. Futschik and C. Schlötterer. The next generation of molecular markers from massively parallel sequencing of pooled DNA samples. *Genetics*, 186(1):207–218, 2010.

E. Garrison and G. Marth. Haplotype-based variant detection from short-read sequencing. *arXiv*, 2012. URL http://arxiv.org/abs/1207.3907.

N. R. Garud, P. W. Messer, and D. A. Petrov. Detection of hard and soft selective sweeps from Drosophila melanogaster population genomic data. *PLOS Genetics*, 17(2):e1009373, 2021.

J. R. Gibson, A. K. Chippindale, and W. R. Rice. The X chromosome is a hot spot for sexually antagonistic fitness variation. *Proceedings of the Royal Society B: Biological Sciences*, 269(1490):499–505, 2002.

J. L. Godwin, R. Vasudeva, Ł. Michalczyk, O. Y. Martin, A. J. Lumley, T. Chapman, and M. J. G. Gage. Experimental evolution reveals that sperm competition intensity selects for longer, more costly sperm. *Evolution Letters*, 1(2):102–113, 2017.

Z. Gompert. Bayesian inference of selection in a heterogeneous environment from genetic time-series data. *Molecular Ecology*, 25(1):121–134, 2016.

Z. Gompert. A population-genomic approach for estimating selection on polygenic traits in heterogeneous environments. *Molecular Ecology Resources*, 2021.

F. A. Gorter, M. F. Derks, J. Van Den Heuvel, M. G. Aarts, B. J. Zwaan, D. De Ridder, and J. A. G. De Visser. Genomics of Adaptation Depends on the Rate of Environmental Change in Experimental Yeast Populations. *Molecular Biology and Evolution*, 34(10):2613–2626, 2017.

K. Grieshop and G. Arnqvist. Sex-specific dominance reversal of genetic variation for fitness. *PLOS Biology*, 16(12):e2006810, 2018.

P. C. Griffin, S. B. Hangartner, A. Fournier-Level, and A. A. Hoffmann. Genomic trajectories to desiccation resistance: Convergence and divergence among replicate selected Drosophila lines. *Genetics*, 205(2):871–890, 2017.

P. R. Haddrill, K. Zeng, and B. Charlesworth. Determinants of Synonymous and Nonsynonymous Variability in Three Species of Drosophila. *Molecular Biology and Evolution*, 28(5):1731–1743, 2011.

D. Haig. Intragenomic politics. *Cytogenetic and Genome Research*, 113(1-4):68–74, 2006.

T. F. Hansen. Why epistasis is important for selection and adaptation. *Evolution*, 67(12):3501–3511, 2013.

J. Hermisson and P. S. Pennings. Soft sweeps: Molecular population genetics of adaptation from standing genetic variation. *Genetics*, 169(4):2335–2352, 2005.

J. Hermisson and P. S. Pennings. Soft sweeps and beyond: understanding the patterns and probabilities of selection footprints under rapid adaptation. *Methods in Ecology and Evolution*, 8(6):700–716, 2017.

W. G. Hill and A. Robertson. The effect of linkage on limits to artificial selection. *Genetical Research*, 8(03):269, 1966.

K. M. Hoedjes, J. van den Heuvel, M. Kapun, L. Keller, T. Flatt, and B. J. Zwaan. Distinct genomic signals of lifespan and life history evolution in response to postponed reproduction and larval diet in Drosophila. *Evolution Letters*, 3(6):598–609, 2019.

B. Holland and W. R. Rice. Perspective: Chase-away Sexual Selection: Antagonistic Seduction versus Resistance. *Evolution*, 52(1):1–7, 1998.

B. Holland and W. R. Rice. Experimental removal of sexual selection reverses intersexual antagonistic coevolution and removes a reproductive load. *Proceedings of the National Academy of Sciences*, 96(9):5083–5088, 1999.

I. Höllinger, P. S. Pennings, and J. Hermisson. Polygenic adaptation: From sweeps to subtle frequency shifts. *PLOS Genetics*, 15(3):e1008035, 2019.

B. Hollis, D. Houle, Z. Yan, T. J. Kawecki, and L. Keller. Evolution under monogamy feminizes gene expression in Drosophila melanogaster. *Nature Communications*, 5:1–5, 2014.

B. Hollis, L. Keller, and T. J. Kawecki. Sexual selection shapes development and maturation rates in Drosophila. *Evolution*, 71(2):304–314, 2017.

S. K. Hsu, A. M. Jakšić, V. Nolte, M. Lirakis, R. Kofler, N. Barghi, E. Versace, and C. Schlötterer. Rapid sex-specific adaptation to high temperature in drosophila. *eLife*, 9:1–16, 2020.

C. J. R. Illingworth and V. Mustonen. A method to infer positive selection from marker dynamics in an asexual population. *Bioinformatics*, 28(6):831–837, 2012.

E. Immonen, R. R. Snook, and M. G. Ritchie. Mating system variation drives rapid evolution of the female transcriptome in Drosophila pseudoobscura. *Ecology and Evolution*, 4(11):2186–2201, 2014.

P. Innocenti and E. H. Morrow. The Sexually Antagonistic Genes of Drosophila melanogaster. *PLOS Biology*, 8(3):e1000335, 2010.

A. Iranmehr, A. Akbari, C. Schlötterer, and V. Bafna. CLEAR: Composition of Likelihoods for Evolve and Resequence Experiments. *Genetics*, 206(2):1011–1023, 2017.

K. Jain and W. Stephan. Modes of rapid polygenic adaptation. *Molecular Biology and Evolution*, 34(12):3169–3175, 2017a.

K. Jain and W. Stephan. Rapid adaptation of a polygenic trait after a sudden environmental shift. *Genetics*, 206(1):389–406, 2017b.

K. M. Jalvingh, P. L. Chang, S. V. Nuzhdin, and B. Wertheim. Genomic changes under rapid evolution: Selection for parasitoid resistance. *Proceedings of the Royal Society B: Biological Sciences*, 281(1779), 2014.

M. A. Jensen, B. Charlesworth, and M. Kreitman. Patterns of genetic variation at a chromosome 4 locus of Drosophila melanogaster and D. simulans. *Genetics*, 160 (2):493–507, 2002.

E. M. Jewett, M. Steinrücken, and Y. S. Song. The Effects of Population Size Histories on Estimates of Selection Coefficients from Time-Series Genetic Data. *Molecular Biology and Evolution*, 33(11):3002–3027, 2016.

A. R. Jha, C. M. Miles, N. R. Lippert, C. D. Brown, K. P. White, and M. Kreitman. Whole-genome resequencing of experimental populations reveals polygenic basis of egg-size variation in drosophila melanogaster. *Molecular Biology and Evolution*, 32(10):2616–2632, 2015.

Á. Jónás, T. Taus, C. Kosiol, C. Schlötterer, and A. Futschik. Estimating the effective population size from temporal allele frequency changes in experimental evolution. *Genetics*, 204(2):723–735, 2016.

P. E. Jorde and N. Ryman. Unbiased estimator for genetic drift and effective population size. *Genetics*, 177(2):927–935, 2007.

K. R. Kasimatis, P. L. Ralph, and P. C. Phillips. Limits to Genomic Divergence Under Sexually Antagonistic Selection. *G3: Genes—Genomes—Genetics*, 58: 2160–1836, 2019.

T. J. Kawecki, R. E. Lenski, D. Ebert, B. Hollis, I. Olivieri, and M. C. Whitlock. Experimental evolution. *Trends in Ecology & Evolution*, 27(10):547–60, 2012.

J. K. Kelly and K. A. Hughes. Pervasive Linked Selection and Intermediate-Frequency Alleles Are Implicated in an Evolve-and-Resequencing Experiment of Drosophila simulans. *Genetics*, 211(3):943–961, 2019.

Y. Kim and W. Stephan. Selective Sweeps in the Presence of Interference Among Partially Linked Loci. *Genetics*, 164(1):389–398, 2003.

R. Kofler and C. Schlötterer. A guide for the design of evolve and resequencing studies. *Molecular Biology and Evolution*, 31(2):474–483, 2014.

R. Kofler, P. Orozco-terWengel, N. De Maio, R. V. Pandey, V. Nolte, A. Futschik, C. Kosiol, and C. Schlötterer. PoPoolation: A Toolbox for Population Genetic Analysis of Next Generation Sequencing Data from Pooled Individuals. *PLOS ONE*, 6(1):e15925, 2011a.

R. Kofler, R. V. Pandey, and C. Schlötterer. PoPoolation2: Identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics*, 27(24):3435–3436, 2011b.

R. Kofler, A. M. Langmüller, P. Nouhaud, K. A. Otte, and C. Schlötterer. Suitability of Different Mapping Algorithms for Genome-Wide Polymorphism Scans with Pool-Seq Data. *G3: Genes—Genomes—Genetics*, 6(November):3507–3515, 2016a.

R. Kofler, V. Nolte, and C. Schlötterer. The impact of library preparation protocols on the consistency of allele frequency estimates in Pool-Seq data. *Molecular Ecology Resources*, 16(1):118–122, 2016b.

Y. Kojima, H. Matsumoto, and H. Kiryu. Estimation of population genetic parameters using an EM algorithm and sequence data from experimental evolution populations. *Bioinformatics*, 36(1):221–231, 2020.

K. Kosheleva and M. M. Desai. Recombination Alters the Dynamics of Adaptation on Standing Variation in Laboratory Yeast Populations. *Molecular Biology and Evolution*, 35(1):180–201, 2018.

R. D. Kouyos, S. P. Otto, and S. Bonhoeffer. Effect of varying epistasis on the evolution of recombination. *Genetics*, 173(2):589–597, 2006.

M. Lacerda and C. Seoighe. Population genetics inference for longitudinally-sampled mutants under strong selection. *Genetics*, 198(3):1237–1250, 2014.

G. I. Lang, D. P. Rice, M. J. Hickman, E. Sodergren, G. M. Weinstock, D. Botstein, and M. M. Desai. Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. *Nature*, 500(7464):571–574, 2013.

V. Laporte and B. Charlesworth. Effective population size and population subdivision in demographically structured populations. *Genetics*, 162(1):501–519, 2002.

Á. J. Láruson, S. Yeaman, and K. E. Lotterhos. The Importance of Genetic Redundancy in Evolution. *Trends in Ecology & Evolution*, 35(9):809–822, 2020.

R. C. Lewontin. The genetic basis of evolutionary change. *Columbia University Press, New York*, 560:1–10, 1974.

H. Li. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, 27(21):2987–2993, 2011.

H. Li. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*, 2013. URL http://arxiv.org/abs/1303.3997.

W. H. Li, S. Yi, and K. Makova. Male-driven evolution. *Current Opinion in Genetics and Development*, 12(6):650–656, 2002.

A. Long, G. Liti, A. Luptak, and O. Tenaillon. Elucidating the molecular architecture of adaptation via evolve and resequence experiments. *Nature Reviews Genetics*, 16(10):567–582, 2015.

E. A. Lucotte, R. Laurent, E. Heyer, L. Ségurel, and B. Toupance. Detection of allelic frequency differences between the sexes in humans: A signature of sexually antagonistic selection. *Genome Biology and Evolution*, 8(5):1489–1500, 2016.

A. S. Malaspinas, O. Malaspinas, S. N. Evans, and M. Slatkin. Estimating allele age and selection coefficient from time-serial data. *Genetics*, 192(2):599–607, 2012.

J. E. Mank and H. Ellegren. Sex-linkage of sexually antagonistic genes is predicted by female, but not male, effects in birds. *Evolution*, 63(6):1464–1472, 2009.

J. E. Mank, B. Vicoso, S. Berlin, and B. Charlesworth. Effective Population Size and the Faster-X Effect: Empirical Results and their Interpretation. *Evolution*, 64(3):663–674, 2010.

N. E. Martins, V. G. Faria, V. Nolte, C. Schlötterer, L. Teixeira, É. Sucena, and S. Magalhães. Host adaptation to viruses relies on few genes with different cross-resistance properties. *Proceedings of the National Academy of Sciences of the United States of America*, 111(16):5938–43, 2014.

I. Mathieson and G. McVean. Estimating selection coefficients in spatially structured populations from time series data of allele frequencies. *Genetics*, 193(3):973–984, 2013.

J. Maynard Smith. Acclimation to high temperatures in inbred and outbred Drosophila Subobscura. *Journal of Genetics*, 84(1):37–45, 1956.

J. Maynard Smith and J. Haigh. The hitch-hiking effect of a favourable gene. *Genetical Research*, 23(1):23–35, 1974.

J. I. Meier, P. A. Salazar, M. Kučka, R. W. Davies, A. Dréau, I. Aldás, O. B. Power, N. J. Nadeau, J. R. Bridle, C. Rolian, N. H. Barton, W. O. McMillan, C. D. Jiggins, and Y. F. Chan. Haplotype tagging reveals parallel formation of hybrid races in two butterfly species. *Proceedings of the National Academy of Sciences of the United States of America*, 118(25), 2021.

R. P. Meisel and T. Connallon. The faster-X effect: Integrating theory and data. *Trends in Genetics*, 29(9):537–544, 2013.

P. W. Messer and D. A. Petrov. Population genomics of rapid adaptation by soft selective sweeps. *Trends in Ecology & Evolution*, 28(11):659–669, 2013.

H. J. Muller. Some genetic aspects of sex. *The American Naturalist*, 66(703):118–138, 1932.

C. Mullon, A. Pomiankowski, and M. Reuter. The effects of selection and genetic drift on the genomic distribution of sexually antagonistic alleles. *Evolution*, 66 (12):3743–3753, 2012.

N. R. Nené, A. S. Dunham, and C. J. Illingworth. Inferring fitness effects from time-resolved sequence data with a delay-deterministic model. *Genetics*, 209(1): 255–264, 2018.

S. Neph, M. S. Kuehn, A. P. Reynolds, E. Haugen, R. E. Thurman, A. K. Johnson, E. Rynes, M. T. Maurano, J. Vierstra, S. Thomas, R. Sandstrom, R. Humbert, and J. A. Stamatoyannopoulos. BEDOPS: High-performance genomic feature operations. *Bioinformatics*, 28(14):1919–1920, 2012.

M. A. Noor, K. L. Gratos, L. A. Bertucci, and J. Reiland. Chromosomal inversions and the reproductive isolation of species. *Proceedings of the National Academy of Sciences of the United States of America*, 98(21):12084–12088, 2001.

T. Ohta. Associative overdominance caused by linked detrimental mutations. *Genetical Research*, 18(3):277–286, 1971.

T. Ohta and M. Kimura. Development of associative overdominance through linkage disequilibrium in finite populations. *Genetical Research*, 16(2):165–177, 1970.

P. Orozco-terWengel, M. Kapun, V. Nolte, R. Kofler, T. Flatt, and C. Schlötterer. Adaptation of Drosophila to a novel laboratory environment reveals temporally heterogeneous trajectories of selected alleles. *Molecular Ecology*, 21(20):4931–4941, 2012.

H. A. Orr. The genetic theory of adaptation: A brief history. *Nature Reviews Genetics*, 6(2):119–127, 2005.

H. A. Orr and A. J. Betancourt. Haldane's sieve and adaptation from the standing genetic variation. *Genetics*, 157(2):875–884, 2001.

K. A. Otte, V. Nolte, F. Mallard, and C. Schlötterer. The genetic architecture of temperature adaptation is shaped by population ancestry and not by selection regime. *Genome Biology*, 22(1):1–25, 2021.

S. P. Otto. Evolutionary potential for genomic islands of sexual divergence on recombining sex chromosomes. *New Phytologist*, 224(3):1241–1251, 2019.

A. Papkou, T. Guzella, W. Yang, S. Koepper, B. Pees, R. Schalkowski, M.-C. Barg, P. C. Rosenstiel, H. Teotónio, and H. Schulenburg. The genomic basis of Red Queen dynamics during rapid reciprocal host–pathogen coevolution. *Proceedings of the National Academy of Sciences*, 116(3):923–928, 2019.

G. Parker. Sexual Selection and Sexual Conflict. In M. S. Blum and N. A. Blum, editors, *Sexual Selection and Reproductive Competition in Insects*, pages 123–166. Elsevier, London, UK, 1979. ISBN 978-0-12-108750-0.

M. M. Patten. The X chromosome favors males under sexually antagonistic selection. *Evolution*, 73(1):84–91, 2019.

J. C. Perry and L. Rowe. The evolution of sexually antagonistic phenotypes. *Cold Spring Harbor Perspectives in Biology*, 7(6):1–18, 2015.

M. A. Phillips and M. K. Burke. Can laboratory evolution experiments teach us about natural populations? *Molecular Ecology*, 30(4):877–879, 2021.

J. E. Pool and R. Nielsen. Population size changes reshape genomic patterns of diversity. *Evolution*, 61(12):3001–3006, 2007.

J. E. Pool and R. Nielsen. The impact of founder events on chromosomal variability in multiply mating species. *Molecular Biology and Evolution*, 25(8):1728–1736, 2008.

M. W. Reeve, K. Fowler, and L. Partridge. Increased body size confers greater fitness at lower experimental temperature in male Drosophila melanogaster. *Journal of Evolutionary Biology*, 13(5):836–844, 2000.

W. R. Rice. Sex chromosomes and the evolution of sexual dimorphism. *Evolution*, 38(4):735–742, 1984.

W. R. Rice and B. Holland. Experimentally enforced monogamy: inadvertent selection, inbreeding, or evidence for sexually antagonistic coevolution? *Evolution*, 59 (3):682–685, 2005.

M. G. Ritchie. The inheritance of female preference functions in a mate recognition system. *Proceedings of the Royal Society B: Biological Sciences*, 267(1441):327–332, 2000.

N. O. Rode, Y. Holtz, K. Loridon, S. Santoni, J. Ronfort, and L. Gay. How to optimize the precision of allele and haplotype frequency estimates using pooled-sequencing data. *Molecular Ecology Resources*, 18(2):194–203, 2018.

L. Rowe, S. F. Chenoweth, and A. F. Agrawal. The genomics of sexual conflict. *American Naturalist*, 192(2):274–286, 2018.

F. Ruzicka and T. Connallon. Is the X chromosome a hot spot for sexually antagonistic polymorphisms? Biases in current empirical tests of classical theory: Sexual antagonism on the X chromosome? *Proceedings of the Royal Society B: Biological Sciences*, 287(1937), 2020.

F. Ruzicka and T. Connallon. An unbiased test reveals no enrichment of sexually antagonistic polymorphisms on the human X chromosome. *Proceedings of the Royal Society B: Biological Sciences*, 289(1967):in press, 2022.

F. Ruzicka, M. S. Hill, T. M. Pennell, I. Flis, F. C. Ingleby, R. Mott, K. Fowler, E. H. Morrow, and M. Reuter. Genome-wide sexually antagonistic variants reveal

long-standing constraints on sexual dimorphism in fruit flies. *PLOS Biology*, 17 (4):e3000244, 2019.

J. M. Sardell and M. Kirkpatrick. Sex differences in the recombination landscape. *American Naturalist*, 195(2):361–379, 2020.

N. Scarcelli and P. X. Kover. Standing genetic variation in FRIGIDA mediates experimental evolution of flowering time in Arabidopsis. *Molecular Ecology*, 18 (9):2039–2049, 2009.

C. Schlötterer, R. Tobler, R. Kofler, and V. Nolte. Sequencing pools of individuals-mining genome-wide polymorphism data without big funding. *Nature Reviews Genetics*, 15(11):749–763, 2014.

C. Schlötterer, R. Kofler, E. Versace, R. Tobler, and S. U. Franssen. Combining experimental evolution with next-generation sequencing: a powerful tool to study adaptation from standing genetic variation. *Heredity*, 114(5):431–440, 2015.

J. G. Schraiber, S. N. Evans, and M. Slatkin. Bayesian inference of natural selection from allele frequency time series. *Genetics*, 203(1):493–511, 2016.

O. Seehausen, R. K. Butlin, I. Keller, C. E. Wagner, J. W. Boughman, P. A. Hohenlohe, C. L. Peichel, G. P. Saetre, C. Bank, Å. Brännström, A. Brelsford, C. S. Clarkson, F. Eroukhmanoff, J. L. Feder, M. C. Fischer, A. D. Foote, A. D. Foote, P. Franchini, C. D. Jiggins, F. C. Jones, A. K. Lindholm, K. Lucek, M. E. Maan, D. A. Marques, S. H. Martin, B. Matthews, J. I. Meier, M. Möst, M. W. Nachman, E. Nonaka, D. J. Rennison, J. Schwarzer, E. T. Watson, A. M. Westram, and A. Widmer. Genomics and the origin of species. *Nature Reviews Genetics*, 15(3):176–192, 2014.

G. Sella and N. H. Barton. Thinking About the Evolution of Complex Traits in the Era of Genome-Wide Association Studies. *Annual Review of Genomics and Human Genetics*, 20(1):461–493, 2019.

N. P. Sharp and S. P. Otto. Evolution of sex: Using experimental genomics to select among competing theories. *BioEssays*, 38(8):751–757, 2016.

H. Shim, S. Laurent, S. Matuszewski, M. Foll, and J. D. Jensen. Detecting and Quantifying Changing Selection Intensities from Time-Sampled Polymorphism Data. *G3 Genes—Genomes—Genetics*, 6(4):893–904, 2016.

R. R. Snook, A. Robertson, H. S. Crudgington, and M. G. Ritchie. Experimental manipulation of sexual selection and the evolution of courtship song in Drosophila pseudoobscura. *Behavior Genetics*, 35(3):245–255, 2005.

R. R. Snook, L. Brüstle, and J. Slate. A test and review of the role of effective population size on experimental sexual selection patterns. *Evolution*, 63(7):1923–1933, 2009.

M. S. Sohail, R. H. Y. Louie, M. R. McKay, and J. P. Barton. MPL resolves genetic linkage in fitness inference from complex evolutionary histories. *Nature Biotechnology*, 39(4):472–479, 2021.

H. G. Spencer and N. K. Priest. The evolution of sex-specific dominance in response to sexually antagonistic selection. *American Naturalist*, 187(5):658–666, 2016.

K. Spitzer, M. Pelizzola, and A. Futschik. Modifying the Chi-square and the CMH test for population genetic inference: Adapting to overdispersion. *The Annals of Applied Statistics*, 14(1):202–220, 2020.

M. Steinrücken, A. Bhaskar, and Y. S. Song. A novel spectral method for inferring general diploid selection from time series genetic data. *Annals of Applied Statistics*, 8(4):2203–2222, 2014.

W. Stephan. Selective Sweeps. *Genetics*, 211(1):5–13, 2019.

W. Stephan, T. H. Wiehe, and M. W. Lenz. The effect of strongly selected substitutions on neutral polymorphism: Analytical results based on diffusion theory. *Theoretical Population Biology*, 41(2):237–254, 1992.

T. Taus, A. Futschik, and C. Schlötterer. Quantifying Selection with Pool-Seq Time Series Data. *Molecular Biology and Evolution*, 34(11):3023–3034, 2017.

O. Tenaillon. The Utility of Fisher's Geometric Model in Evolutionary Genetics. *Annual Review of Ecology, Evolution, and Systematics*, 45(1):179–201, 2014.

H. Teotónio, I. M. Chelo, M. Bradić, M. R. Rose, and A. D. Long. Experimental evolution reveals natural selection on standing genetic variation. *Nature Genetics*, 41(2):251–257, 2009.

J. Terhorst, C. Schlötterer, and Y. S. Song. Multi-locus Analysis of Genomic Time Series Data from Experimental Evolution. *PLOS Genetics*, 11(4):e1005069, 2015.

H. Topa, Á. Jónás, R. Kofler, C. Kosiol, and A. Honkela. Gaussian process test for high-throughput sequencing time series: Application to experimental evolution. *Bioinformatics*, 31(11):1762–1770, 2015.

E. Toprak, A. Veres, J. B. Michel, R. Chait, D. L. Hartl, and R. Kishony. Evolutionary paths to antibiotic resistance under dynamically sustained drug selection. *Nature Genetics*, 44(1):101–105, 2012.

T. L. Turner, A. D. Stewart, A. T. Fields, W. R. Rice, and A. M. Tarone. Population-Based Resequencing of Experimentally Evolved Populations Reveals the Genetic Basis of Body Size Variation in Drosophila melanogaster. *PLOS Genetics*, 7(3): e1001336, 2011.

T. L. Turner, P. M. Miller, and V. A. Cochrane. Combining Genome-Wide Methods to Investigate the Genetic Complexity of Courtship Song Variation in Drosophila melanogaster. *Molecular Biology and Evolution*, 30(9):2113–2120, 2013.

G. A. Van der Auwera and B. D. O'Connor. *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra*. O'Reilly Media, Sebastopol, CA, USA, 2020. ISBN 9781491975169.

P. Veltsos, Y. Fang, A. R. Cossins, R. R. Snook, and M. G. Ritchie. Mating system manipulation and the evolution of sex-biased gene expression in Drosophila. *Nature Communications*, 8(1):1–11, 2017.

B. Vicoso and B. Charlesworth. Evolution on the X chromosome: unusual patterns and processes. *Nature Reviews Genetics*, 7(8):645–653, 2006.

B. Vicoso and B. Charlesworth. Effective population size and the faster-X effect: An extended model. *Evolution*, 63(9):2413–2426, 2009.

C. Vlachos and R. Kofler. MimicrEE2: Genome-wide forward simulations of Evolve and Resequencing studies. *PLOS Computational Biology*, 14(8):e1006413, 2018.

C. Vlachos, C. Burny, M. Pelizzola, R. Borges, A. Futschik, R. Kofler, and C. Schlötterer. Benchmarking software tools for detecting and quantifying selection in evolve and resequencing studies. *Genome Biology*, 20(1):1–11, 2019.

A. G. Wallace, D. Detweiler, and S. W. Schaeffer. Evolutionary history of the third chromosome gene arrangements of drosophila pseudoobscura inferred from inversion breakpoints. *Molecular Biology and Evolution*, 28(8):2219–2229, 2011.

B. Walsh and M. Lynch. *Evolution and Selection of Quantitative Traits*, volume 1. Oxford University Press, 2018. ISBN 9780198830870.

R. A. W. Wiberg, O. E. Gaggiotti, M. B. Morrissey, and M. G. Ritchie. Identifying consistent allele frequency differences in studies of stratified populations. *Methods in Ecology and Evolution*, 8(12):1899–1909, 2017.

R. A. W. Wiberg, P. Veltsos, R. R. Snook, and M. G. Ritchie. Experimental evolution supports signatures of sexual selection in genomic divergence. *Evolution Letters*, 5(3):214–229, 2021.

S. Wigby and T. Chapman. Female resistance to male harm evolves in response to manipulation of sexual conflict. *Evolution*, 58(5):1028–1037, 2004.

G. S. Wilkinson, F. Breden, J. E. Mank, M. G. Ritchie, A. D. Higginson, J. Radwan, J. Jaquiery, W. Salzburger, E. Arriero, S. M. Barribeau, P. C. Phillips, S. C. Renn, and L. Rowe. The locus of sexual selection: Moving sexual selection studies into the post-genomics era. *Journal of Evolutionary Biology*, 28(4):739–755, 2015.

A. E. Wright, M. Fumagalli, C. R. Cooney, N. I. Bloch, F. G. Vieira, S. D. Buechel, N. Kolm, and J. E. Mank. Male-biased gene expression resolves sexual conflict through the evolution of sex-specific genetic architecture. *Evolution Letters*, 2(2): 52–61, 2018.

A. E. Wright, T. F. Rogers, M. Fumagalli, C. R. Cooney, and J. E. Mank. Phenotypic sexual dimorphism is associated with genomic signatures of resolved sexual conflict. *Molecular Ecology*, 28(11):2860–2871, 2019.

L. Yun, M. Bayoumi, S. Yang, P. J. Chen, H. D. Rundle, and A. F. Agrawal. Testing for local adaptation in adult male and female fitness among populations evolved under different mate competition regimes. *Evolution*, 73(8):1604–1616, 2019.

T. Zinger, M. Gelbart, D. Miller, P. S. Pennings, and A. Stern. Inferring population genetics parameters of evolving viruses using time-series data. *Virus Evolution*, 5 (1):1–11, 2019.

Appendix A

Supplement to chapter 2

Impact of E&R experimental design on detecting targets of selection - sampling schemes, replication and coverage

We tested six different time schemes that allow for varying number, span, and distribution of sampled time points (schematically represented in **fig. 2.3A**). Surprisingly, we observed that there is no substantial increase in accuracy when more time points are sampled. Experiments with two time points estimate quite accurately σ for most of the simulated scenarios (**fig. 2.3A** and **fig. A.2**). Worth of note are the two time span regimes we have investigated: a span of $0.2N_e$, i.e., a common E&R study length (as seen in Barghi et al., 2019; Burke et al., 2014; Papkou et al., 2019), and $0.4N_e$ generations. Similarly to the number of sampled time points, we do not observe a substantial increase in the accuracy of σ by increasing the length of the experiment.

Most E&R studies have uniform sampling schemes (Burke et al., 2014; Barghi et al., 2019). Nevertheless, we set out to compare a range of uniform and biased time schemes. The results show that there are no substantial improvements when using biased sampling schemes for neutral trajectories. However, in the presence of selection, the scheme that samples more often at the start of the experiment provides better estimates of σ (time scheme number 3 for $N_e\sigma = 0.1$ to 10.0; **fig. 2.3A**).

We then set out to test the impact of the number of experimental replicate populations. We observed that the accuracy of the estimated selection coefficients increases

with the number of replicated populations. The most inaccurate estimates were observed in scenarios with two replicates, both with neutral and weak selection (two replicates, $N_e\sigma = 0.0$ and 0.1; **fig. 2.3B**). This suggests that a higher number of replicates may help decrease the false positive rate of E&R experiments. Our results are consistent with previous studies showing that replicated populations improve the detection of selective targets (Kofler and Schlötterer, 2014) and the accuracy of estimated selection coefficients (Taus et al., 2017).

Additionally, we considered three levels of coverage to represent experiments with low to high sequencing depth. We observed that low coverage can decrease the accuracy of σ in the case of small populations and low starting frequency trajectories (first row of the heatmap in **fig. 2.3C**). In contrast, the accuracy does not differ substantially in the case of average and high sequencing depths, which suggests that a coverage around 60x provides enough resolution for detecting targets of selection in E&R experiments. This is consistent with results obtained by Kofler and Schlötterer (2014), who showed that a coverage of approximately 50x is sufficient to identify strongly selected loci.

Real versus simulated data - comparing Barghi et al. (2019) and Vlachos et al. (2019)

To assess whether the simulation study by Vlachos et al. (2019) mimicked the experimental data gathered by Barghi et al. (2019), we examined allele frequency changes, nucleotide diversity and variance in allele frequencies amongst replicates independently for the two datasets. Overall, Barghi et al. (2019) show higher nucleotide diversity (**fig. A.16** and **fig. A.17**) as well as greater allele frequency changes (**fig. A.20** and **fig. A.21**). Smaller allele frequency changes are consistent with a less polymorphic population. In addition, we investigated patterns of nucleotide diversity along the chromosomes in Barghi et al. (**fig. A.18**) and the genomic element simulated by Vlachos et al. (**fig. A.19**). We found that both show similar regions of very low diversity at the start and/or end of the chromosome corresponding to the telomeres. In the case of Vlachos et al., the location of actual selected sites does not match dips in diversity around them, suggesting the simulation does not

capture any low diversity regions that would be the result of selective sweeps. This might indicate that there is too much interference between selected sites to allow for selective sweeps or that the low polymorphism is hindering adaptation.

For both studies, we investigated the relationship between Bait-ER's scaled selection coefficient, $N_e\sigma$, and any allele frequency changes. The two show a different pattern with Barghi et al. chromosomes producing a tail of much higher $N_e\sigma$ values for trajectories that have relatively low allele frequency changes (**fig. A.22**). In Vlachos et al., selected loci in yellow have allele frequency changes that range from very low to some of the highest amongst replicate populations (**fig. A.23**) and we do no observe a tail of high scoring trajectories. Here, there is no apparent linear relationship regardless of replicate experiment (mean r^2 = -0.05, min = -0.11, max = 0.24). Moreover, we analysed how differences in coverage amongst loci would affect Bait-ER's accuracy at identifying targets of selection. This would only be an issue in Barghi et al. as all sites in the Vlachos et al. dataset have identical sequencing depth (1000x). For that reason, we visually inspected the relationship between logBFs and coverage (**fig. A.24**). We found no semblance of any linear or quadratic relationship in any of the five main chromosomes, suggesting that Bait-ER is not biased towards higher coverage sites. Finally, we compared the shape of allele frequency spectra at the start (generation 0) and at the end of the experiment (generation 60) for both studies (**fig. A.25** and **fig. A.26**). Barghi et al. spectra are markedly U-shaped with a non-negligible amount of intermediate frequency alleles. In contrast, Vlachos et al. simulations show distribution that is much more biased towards low frequency alleles which is compatible with the previous finding of low nucleotide diversity in this dataset.

Supplementary tables and figures

Conditions	Relative CPU time
Base experiment	1.00
2 replicates	0.97
10 replicates	1.15
100 individuals	0.05
1000 individuals	31.97
20X coverage	1.00
100X coverage	1.00
2 time points	0.92
11 time points	1.16

Table A.1: **Relative CPU time for Bait-ER under several experimental/population conditions.**[a]

[a]The relative CPU time was calculated considering a base E&R experiment with 5 replicates, 5 uniformly distributed time points, 300 individuals and an average coverage of 60X. Unless otherwise specified, the remaining simulation parameters were the same as those under base experiment conditions.

Parameter	Simulated values	Notes
Population parameters		
N_e effective population size	100, **300** and 1000	representing a small, a typical and a large in E&R study population
p_0 allele's initial frequency	**0.01, 0,05, 0.1 and 0.5**	representing rare, low frequency and common alleles
σ selection coefficient	$\mathbf{0, 1/10N_e, 1/N_e}$ **and** $\mathbf{10/N_e}$	representing regimes of neutrally evolving, drift-dominated, and selection-dominated allele trajectories
Experimental parameters		
C Coverage	20x, **60x** and 100x	low, medium and high coverage for pool-seq data
R Number of replicates	2, **5** and 10	
T Number of time points	2, **5** and 11 time points, assessed at generations	represents different combinations of total number of time points, experiment lengths and distribution of sampling events (uniform/non-uniform)
	(0.0, 0.2),	
	(0.00, 0.05, 0.10, 0.15, 0.20), (0.00, 0.04 0.08 0.12 0.20), (0.00, 0.08 0.12 0.16 0.20), (0.0, 0.1 0.2 0.3 0.4) and (0.00, 0.02 0.04 0.06 0.08 0.10 0.12 0.14 0.16 0.18 0.20) relative to N_e.	

Table A.2: **Simulated scenarios.**[b]

[b]The simulated parameters can be divided into two categories: those which are related with the population dynamics (effective population size, selection coefficient, and allele's starting frequency) and those related to the experimental design (coverage, number of time points and number of replicates). To test the experimental conditions, we defined a base experiment with 5 replicates, 5 uniformly distributed time points (total span of $0.20Ne$ generations) and a coverage of 60x. This base experiment is highlighted in bold. The two maximum experiment lengths considered ($0.2N_e$ and $0.4N_e$) were chosen based on typical E&R experimental designs.

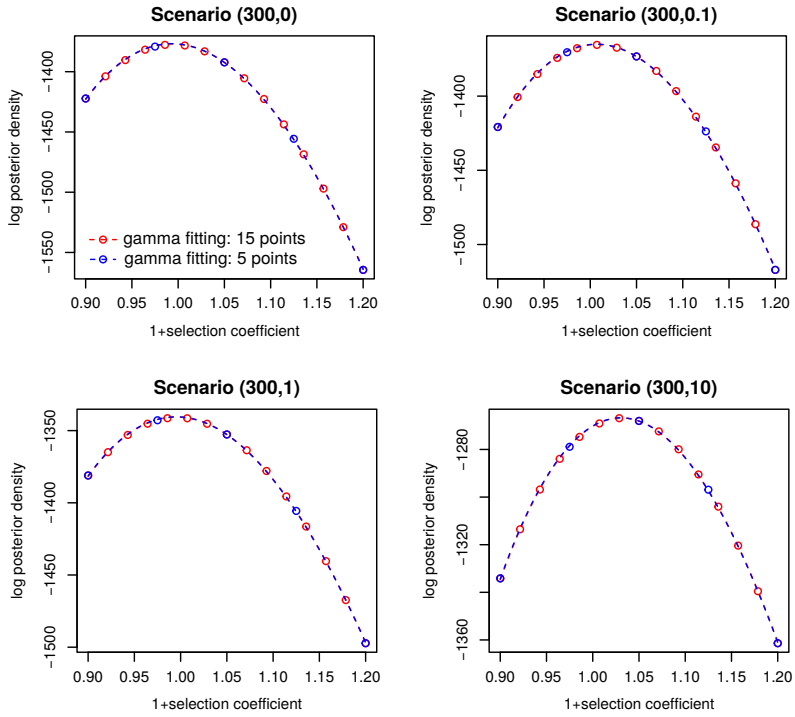


Figure A.1: **Gamma fit to the log posterior distribution of σ.** The fitting of the gamma density to the posterior distribution of $1 + \sigma$ was evaluated for both neutral and selected trajectories. Each plot represents the different scenarios that were simulated by varying the population size, selection coefficient (indicated within brackets $(N_e, N_e\sigma)$). The points represent either 15 or 5 assessments of the posterior distribution (red and blue points, respectively) obtained whilst computing the log-likelihood. The dashed lines represent the gamma fitting to the log posterior obtained by using either 15 or 5 points. Both fitting methods match quite well, suggesting Bait-ER's current approach using 5 points is a good choice.

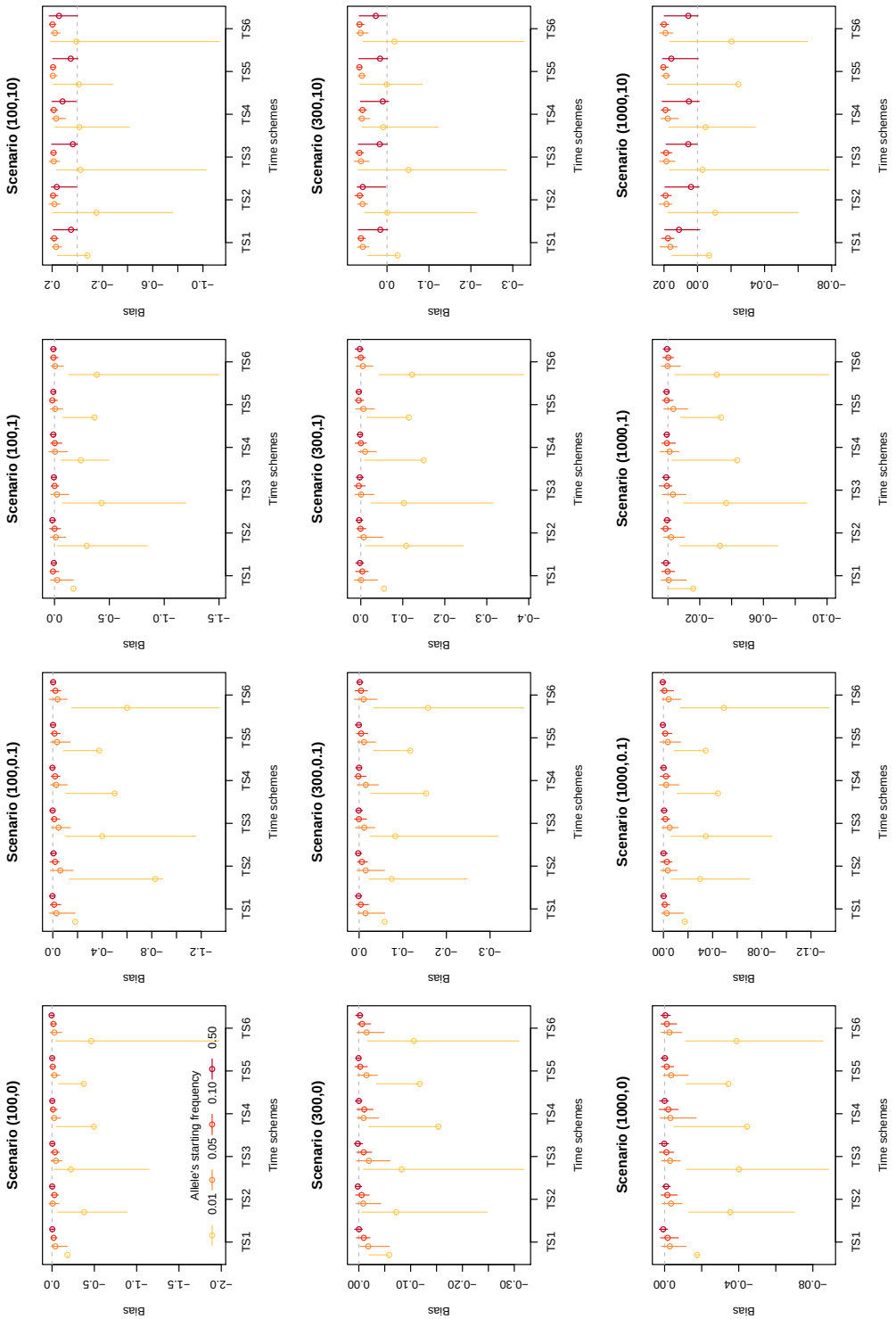


Figure A.2: **Impact of the number and distribution of sampled time points, as well as experimental span on the estimated selection coefficients.** The plots illustrate the bias of the estimated selection coefficients for different time sampling schemes of an E&R experiment. This bias was calculated by computing the difference between the Bait-ER estimated and the true σ: $\hat{\sigma} - \sigma$. The vertical lines and points indicate the interquartile range and median bias. Each plot represents a scenario that was simulated by varying the population size, the true selection coefficient (indicated within brackets (N_e, $N_e\sigma$)) and starting allele frequency (indicated by the yellow-to-red colour gradient). Here is a brief explanation of how the different time schemes might vary from one another: most time schemes have five sampling times, apart from TS1 and TS6, which had two and eleven sampling points, respectively; the majority of time schemes have a total span of $N_e/5$ generations, except for TS5 which had double the span ($2N_e/5$); lastly, uniform sampling was used across most scenarios, but for TS3, which is more heavily sampled during the first half of the experiment, and TS4, during the second half.

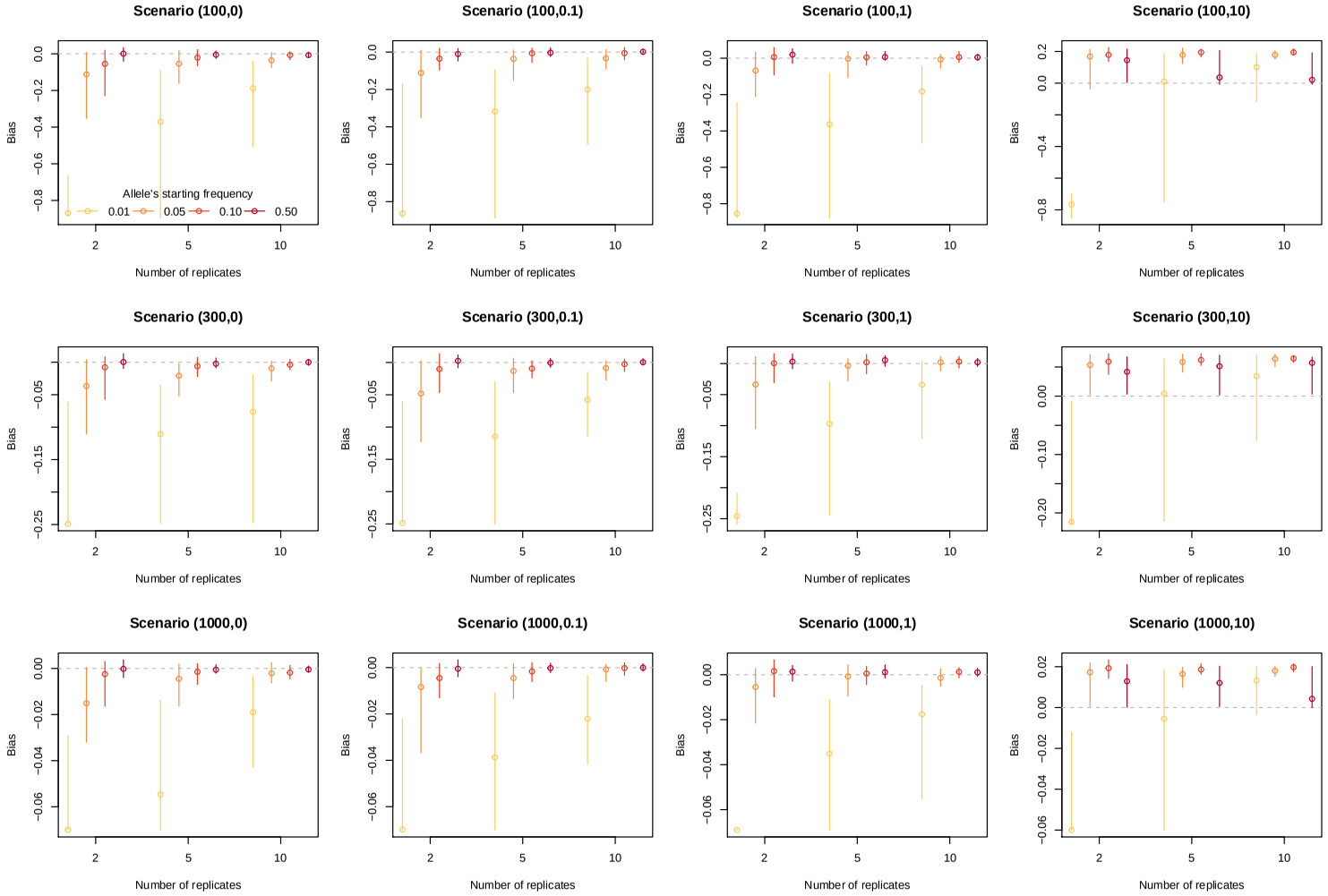


Figure A.3: **Impact of the number of replicates on the estimation of selection coefficients.** The plots show the bias of the estimated selection coefficients for several scenarios with different number of replicate experimental populations. This bias was calculated by computing the difference between the Bait-ER estimated and the true σ: $\hat{\sigma} - \sigma$. Vertical lines and points indicate the interquartile range and median bias. Each plot represents the different scenarios that were simulated by varying the population size, the true selection coefficient (indicated within brackets $(N_e, N_e\sigma)$) and starting allele frequency (indicated by the yellow-red gradient of colors).

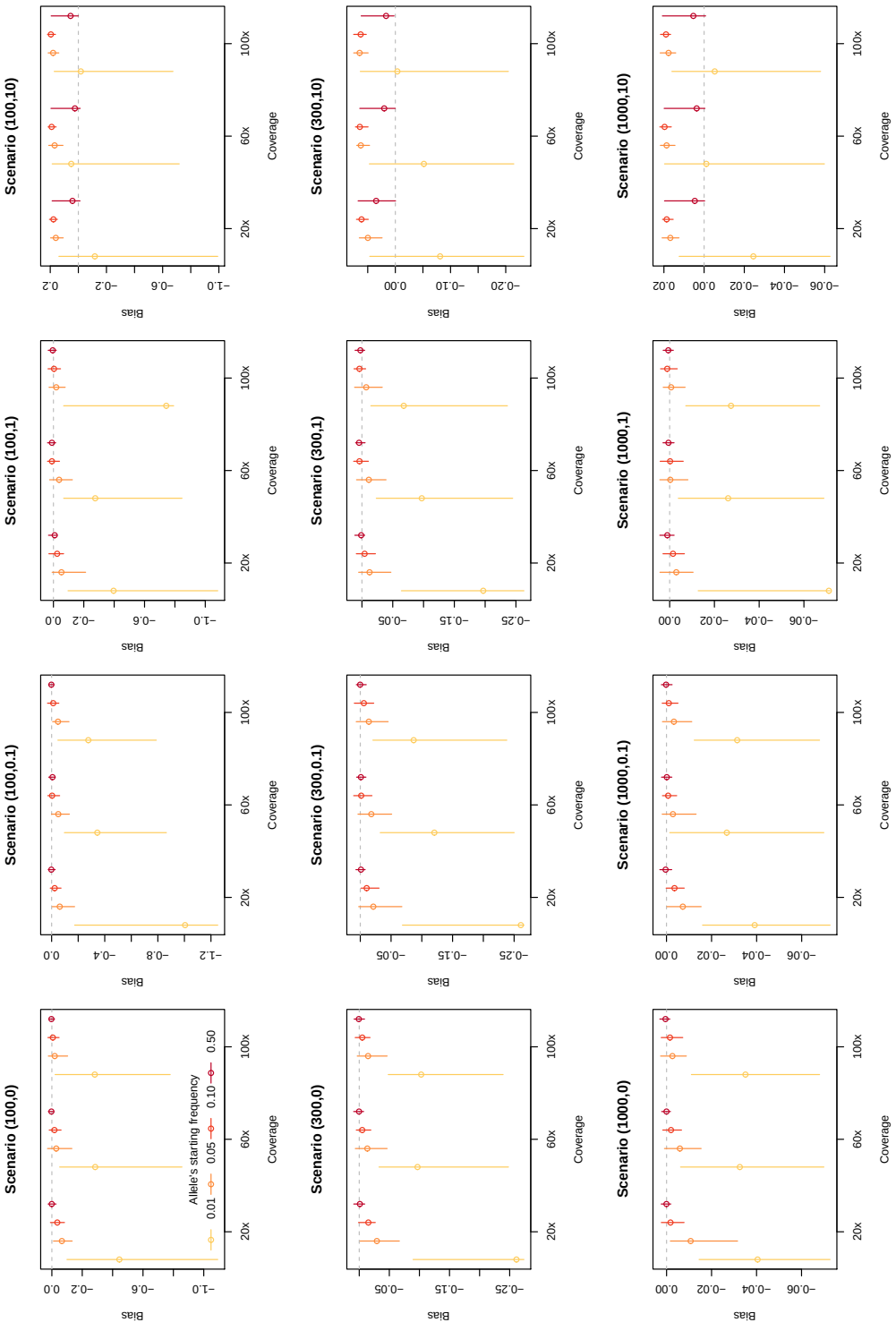


Figure A.4: **Impact of coverage on the estimation of selection coefficients** The plots show the bias of the estimated selection coefficients obtained for different sequencing depths of an E&R experiment. This bias was calculated by taking the difference between the Bait-ER estimated and the true σ: $\hat{\sigma} - \sigma$. The vertical lines and points indicate the interquartile range and median bias. Each plot represents the different scenarios that were simulated by varying the population size, selection coefficient (indicated within the brackets $(N_e, N_e\sigma)$) and starting allele frequency (indicated by the yellow-to-red colour gradient).

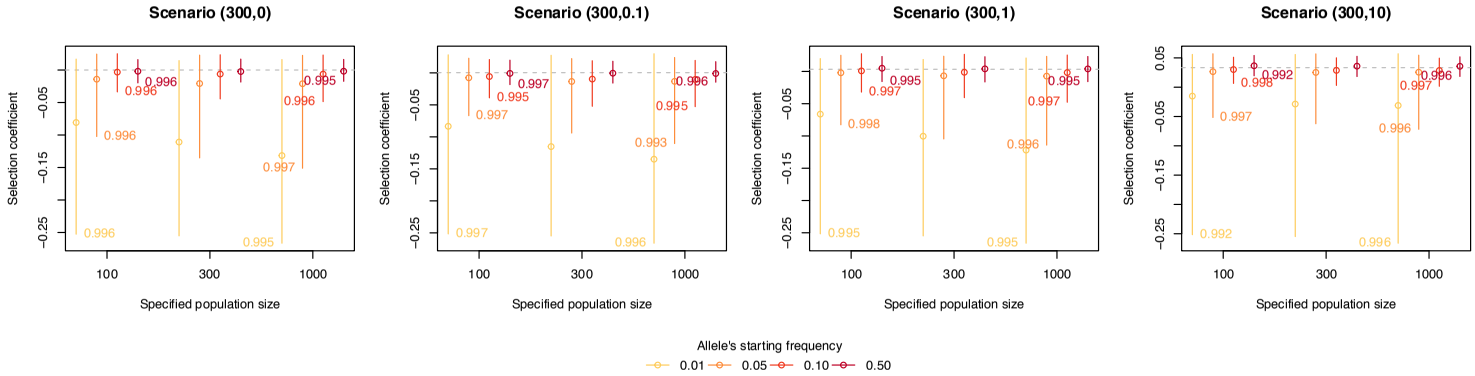


Figure A.5: **Impact of the user-specified population size on the estimation of selection coefficients.** The plots show the distribution of the estimated selection coefficients where the population size is misspecified. Vertical lines and points indicate the interquartile range and median selection coefficient. Each plot represents a specific scenario that was simulated by varying the population size, the true selection coefficient (indicated within brackets $(N_e, N_e\sigma)$) and starting allele frequency (indicated by the yellow-to-red colour gradient). The numbers next to each bar correspond to the Spearman's correlation coefficient, which correlate the BFs of the 100 replicated trajectories between the cases where we have either under- and overspecified the population size ($N_e = 100$ or 1000, respectively) and the case where we use the true population size ($N_e = 300$).

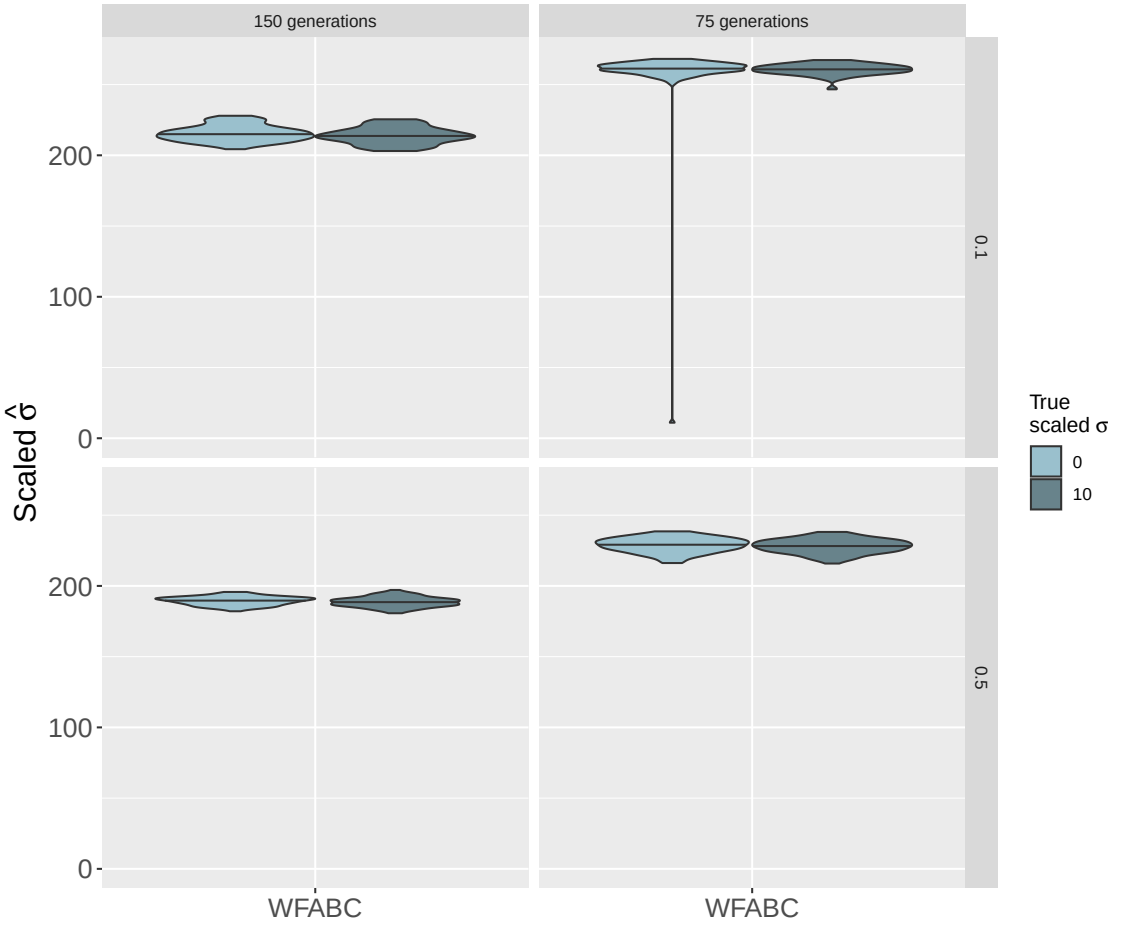


Figure A.6: **Comparison of estimates of σ produced by WFABC**. These plots include estimates for those Moran trajectories simulated with starting frequencies of 10% and 50% (top and bottom row, respectively). Only neutrally evolving ($N_e\sigma = 0$) and strongly selected alleles were considered here ($N_e\sigma = 10$). The left and right hand side panels correspond to two different experiment lengths: 150 and 75 generations, respectively.

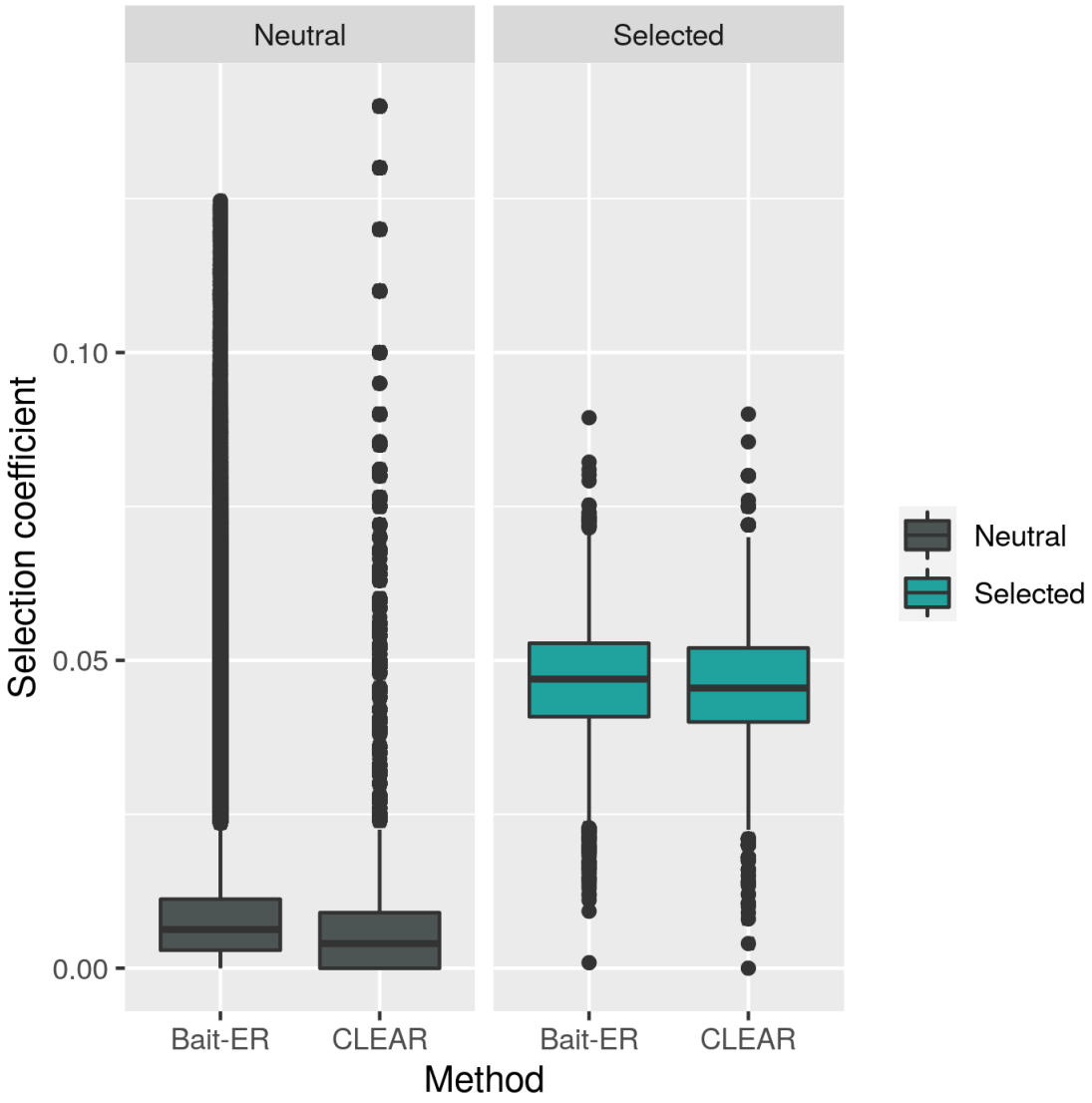


Figure A.7: **Distribution of $\hat{\sigma}$ for Bait-ER and CLEAR.** These boxplots show the comparison between estimates of selection coefficients for both selected (right panel) and neutral (left panel) sites from the Vlachos et al. (2019) selective sweep scenario. Data filtered out by Bait-ER was excluded from the analysis.

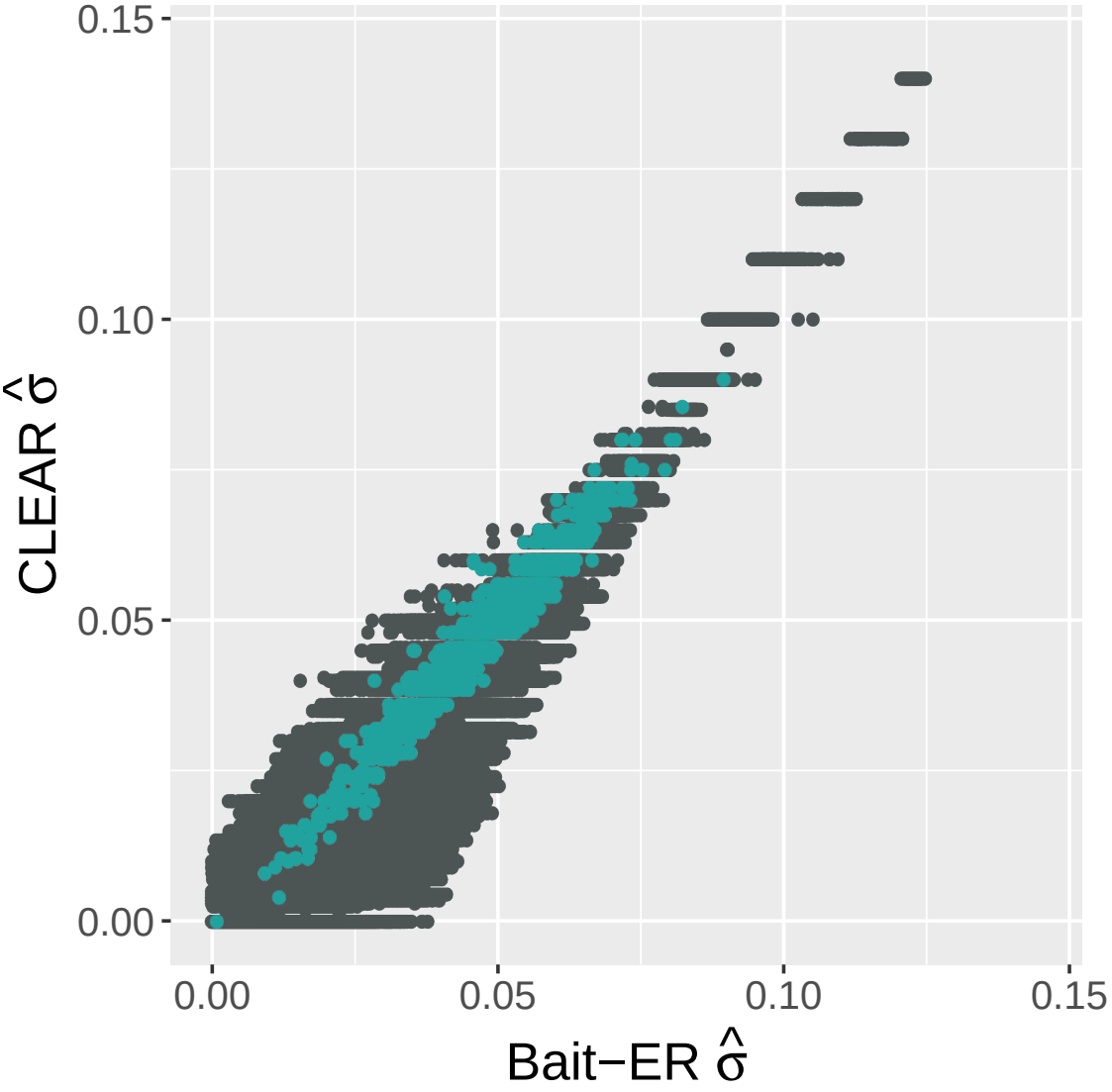


Figure A.8: **Comparison of Bait-ER vs CLEAR $\hat{\sigma}$ in the Vlachos et al. sweep scenario.** This scatterplot is the comparison of selection coefficient estimates for both methods, where blue and grey points are selected and neutral sites, respectively. Data filtered out by Bait-ER was excluded from the analysis.

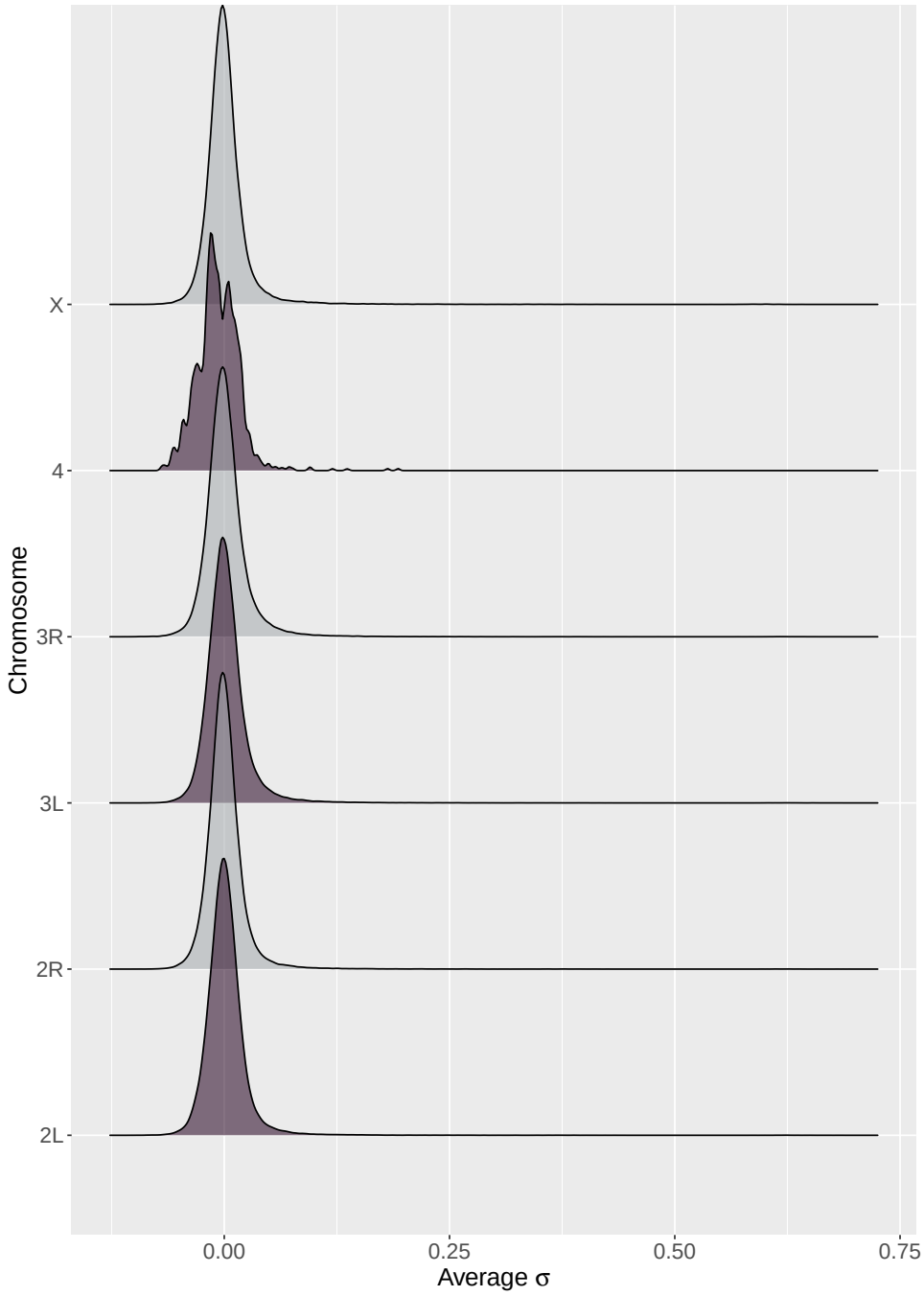


Figure A.9: **Distribution of the selection coefficients estimated by Bait-ER for each of the chromosomes in the Barghi et al. (2019) dataset.** From bottom to top row, the figure shows the distribution of $\hat{\sigma}$ on chromosomes 2L, 2R, 3L, 3R, 4 and X. All distributions seem to be centered around 0, which is unsurprising given that we expect selection not to be widespread across the genome but rather restricted to some genomic windows rather.

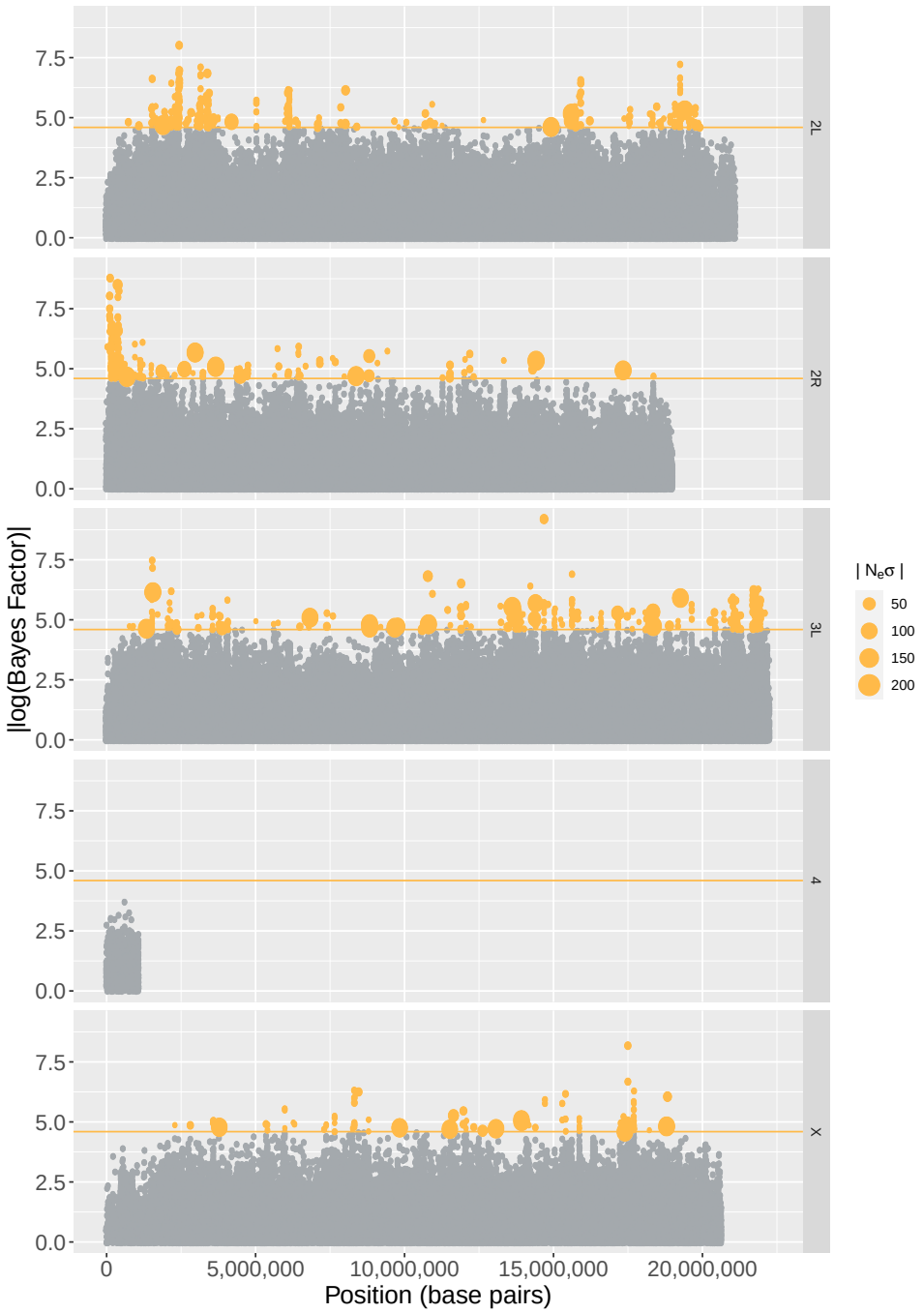


Figure A.10: **Bayes Factors on chromosomes 3L, 2, 4, 5 and X.** Similarly to figure fig. 2.8, these Manhattan plots show log-transformed Bayes Factors computed by Bait-ER for loci along the left arm of the 3rd chromosome, as well as chromosomes 2, 4, 5 and X in the Barghi et al. (2019) time series dataset. The orange line indicates a conservative threshold of approximately 4.6, which corresponds to log(0.99/0.01), meaning all points in orange have very strong evidence for these to be under selection. The SNPs that are significant at this level are sorted by size according to how strong Bait-ER's selection coefficients are. In other words, points are sized according to how strong the large selection coefficient is estimated to be.

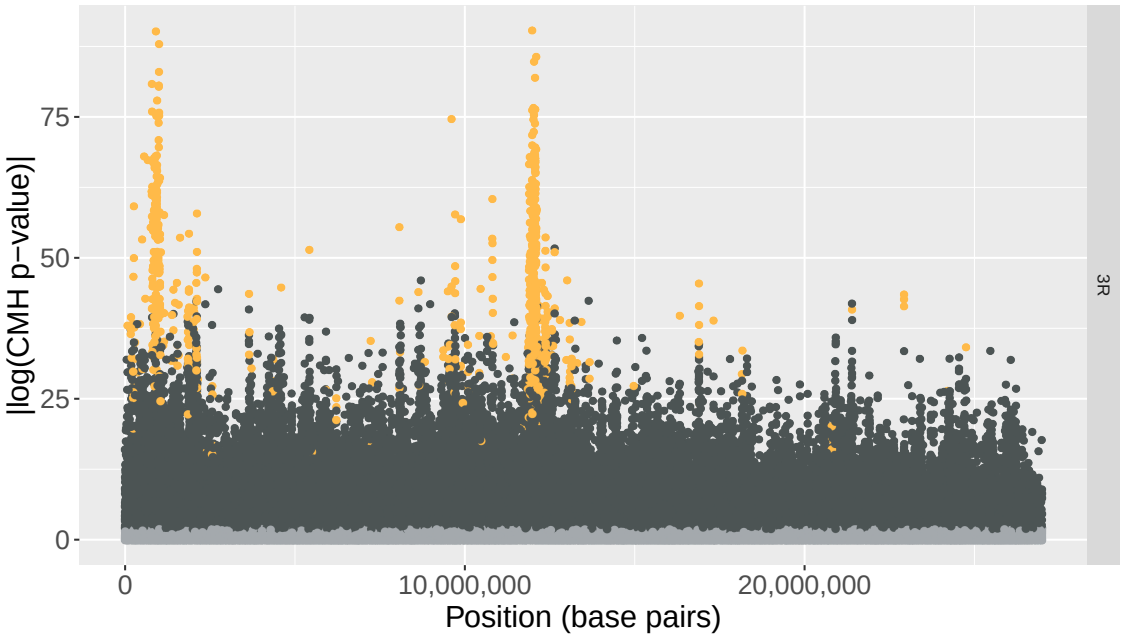


Figure A.11: **CMH test p-values along chromosome 3R.** Data presented are the log-transformed p-values produced by the CMH test as per Barghi et al. for comparison with fig. 2.8. The most pronounced peaks seem to be common to both methods, although the CMH test produces many more significant values when compared to Bait-ER. Orange coloured points correspond to BFs which are greater than log(0.99/0.01) (approx. 4.6) and p-values less than or equal to 0.01, i.e., those that are considered significant by both tests. Blue coloured points indicated that the computed BF is greater than our threshold but not significant according to the CMH test. Additionally, dark grey points are significant according to the CMH test, but not to Bait-ER, and light grey points are inferred not significant by both tests.

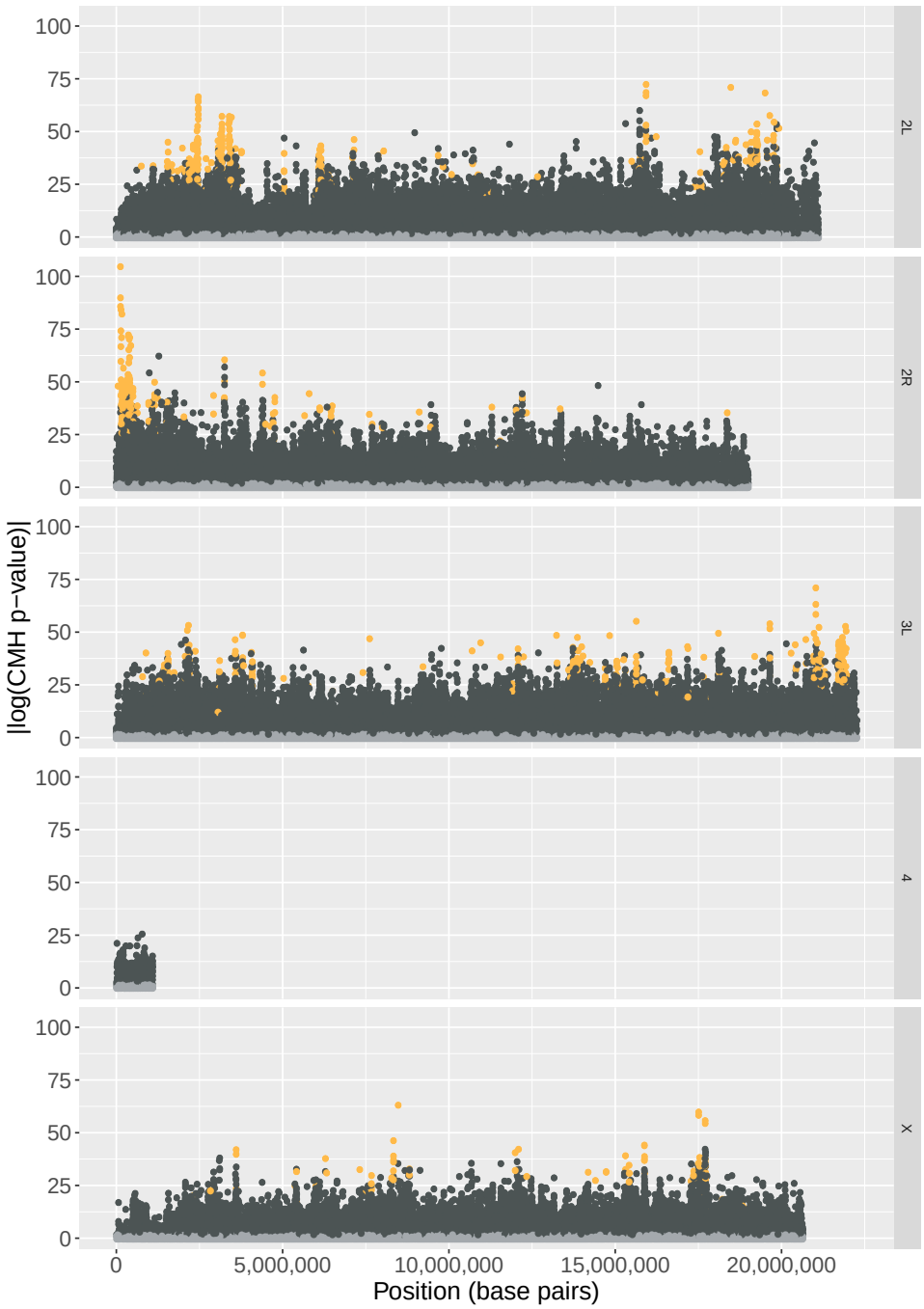


Figure A.12: **CMH test p-values along chromosomes 3L, 2, 4, 5 and X.** Data presented are the log-transformed p-values produced by the CMH test as per Barghi et al. for comparison with fig. A.10. Orange coloured points correspond to BFs which are greater than log(0.99/0.01) (approx. 4.6) and p-values less than or equal to 0.01, i.e., those that are considered significant by both tests. Blue coloured points indicated that the computed BF is greater than our threshold but not significant according to the CMH test. Additionally, dark grey points are significant according to the CMH test, but not to Bait-ER, and light grey points are inferred not significant by both tests.

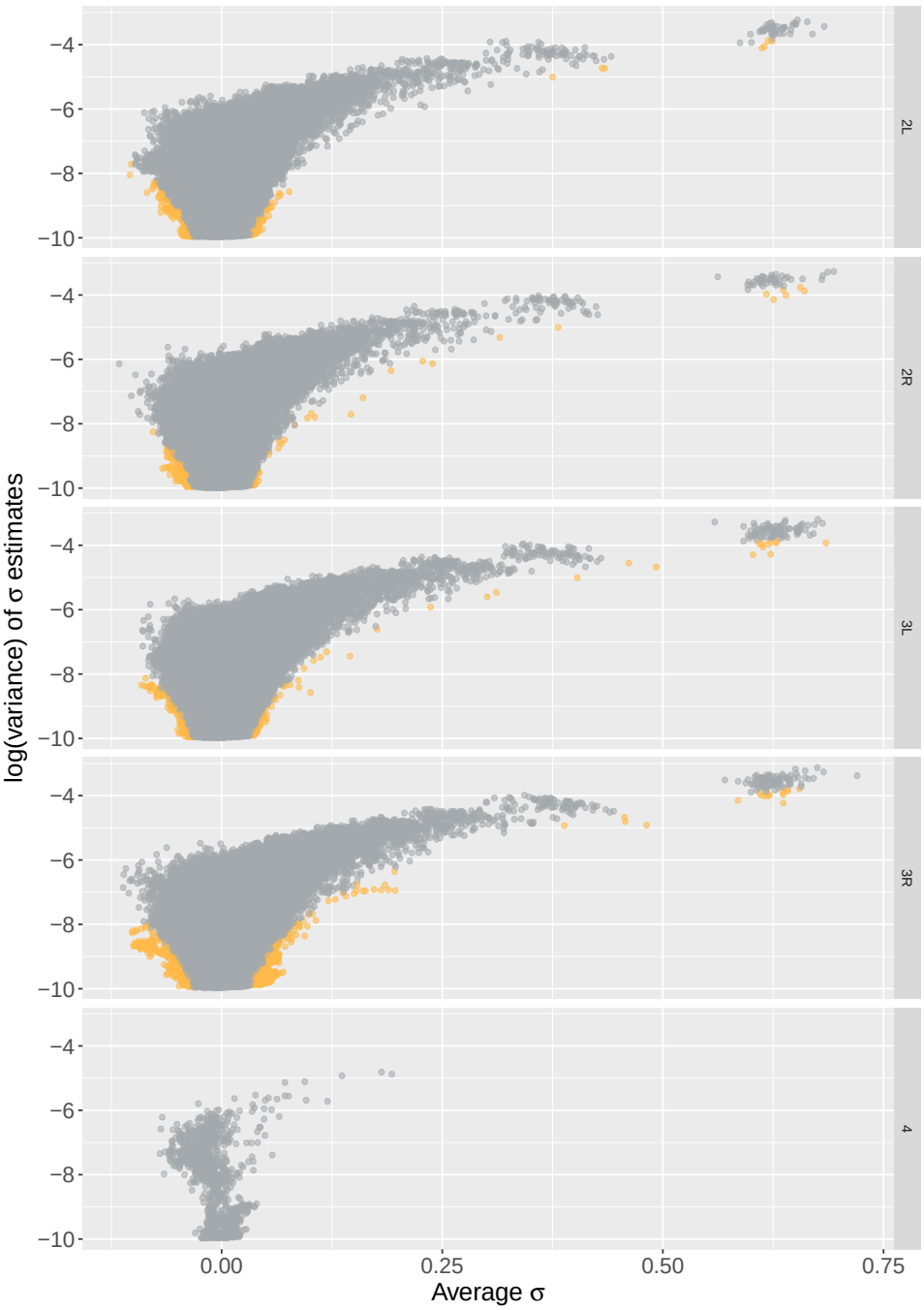


Figure A.13: **Variance versus mean sigma on chromosomes 3R, 3L, 4 and 5.** Similarly to fig. 2.9, these graphs compare log transformed variances in σ estimates to average σs. The variance is calculated using the inferred rate and shape parameters for the beta distribution, and the average σ is the mean value of the posterior distribution estimated by Bait-ER. Orange coloured points are significant at a conservative BF threshold of log(0.99/0.01), approx. 4.6. Data in Barghi et al. (2019).

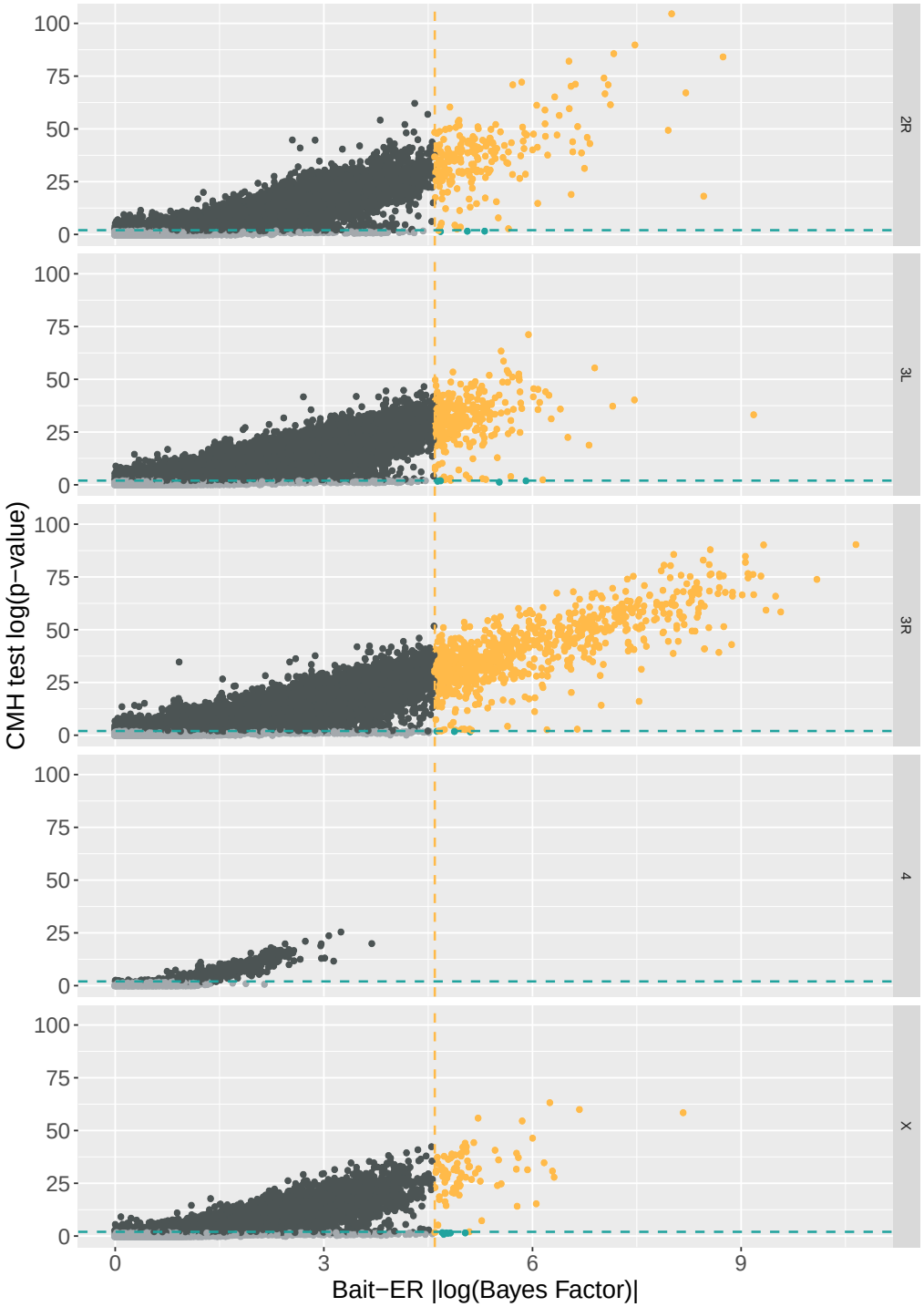


Figure A.14: **Bait-ER's Bayes Factors versus CMH test's p-values on chromosomes 2L, 3R, 3L, 4 and X.** Orange coloured points correspond to BFs which are greater than log(0.99/0.01) (approx. 4.6) and p-values less than or equal to 0.01, i.e., those that are considered significant by both tests. Blue coloured points indicated that the computed BF is greater than our threshold but not significant according to the CMH test. Additionally, dark grey points are significant according to the CMH test, but not to Bait-ER, and light grey points are inferred not significant by both tests.

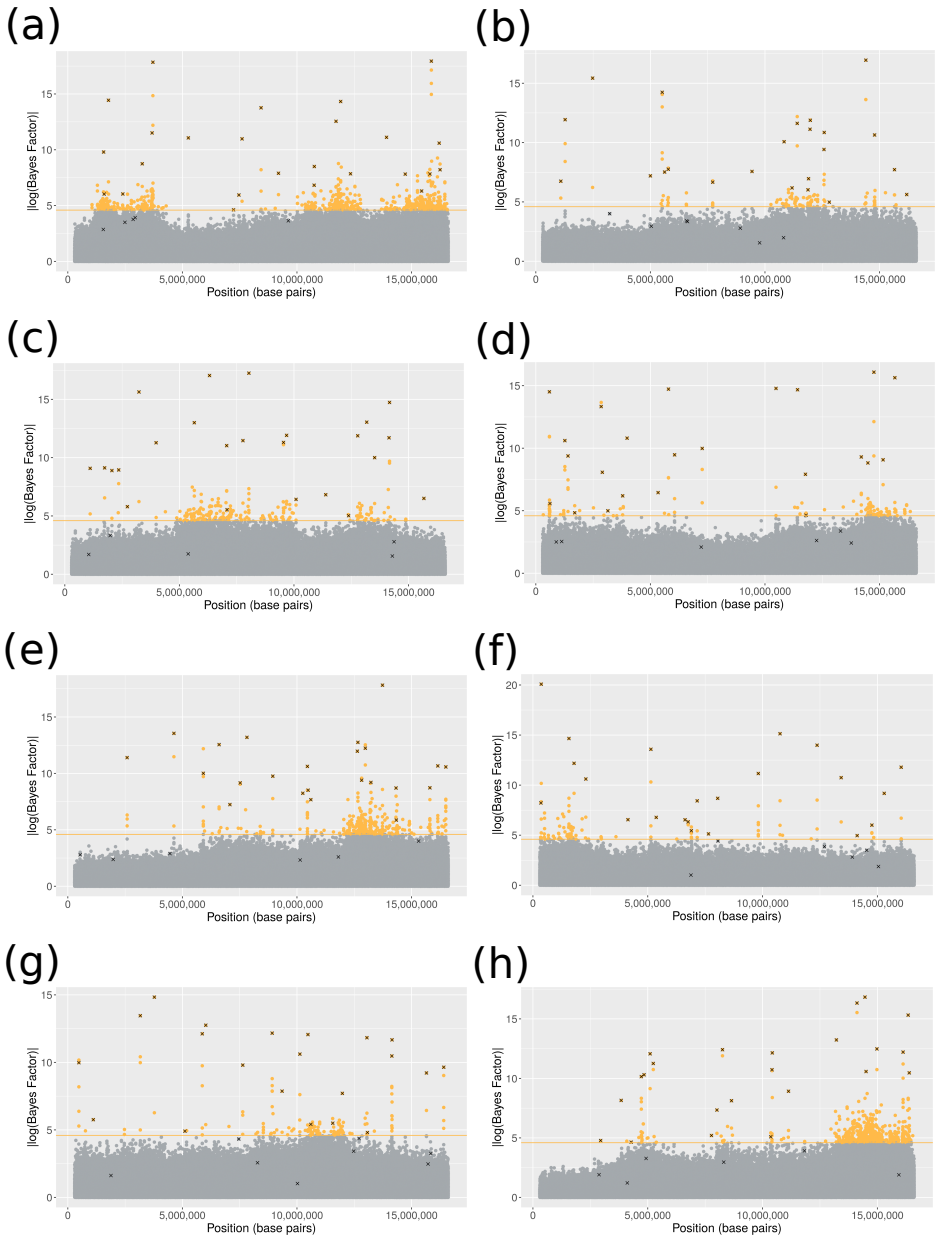


Figure A.15: **Bayes Factors on 8 replicates experiments on the chromosome arm simulated in Vlachos et al. (2019).** This Manhattan plot shows log-transformed Bayes Factors computed by Bait-ER. The orange line indicates a conservative threshold of approximately 4.6, which corresponds to log(0.99/0.01). Data points which are marked by an x are those that were the true targets of selection. These were randomly chosen and simulated with a constant selection coefficient of 0.5. Plots a-h correspond to replicate experiments #4, #9, #39, #50, #51, #77, #84 and #93, respectively.

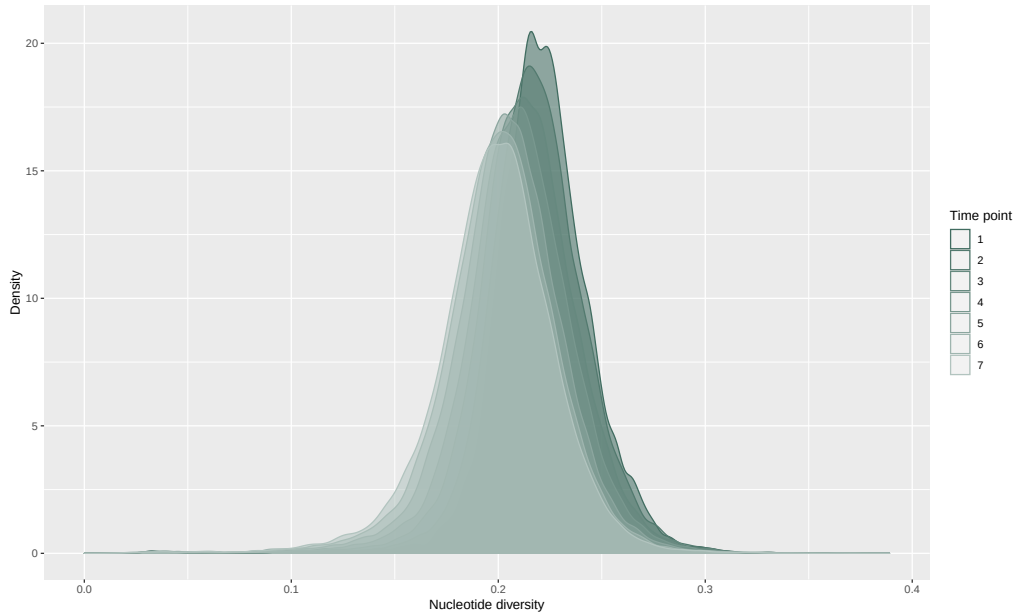


Figure A.16: **Relative nucleotide diversity density per time point for all chromosomes (excluding 4) in the Barghi et al. (2019) dataset.** Diversity values presented are per site.

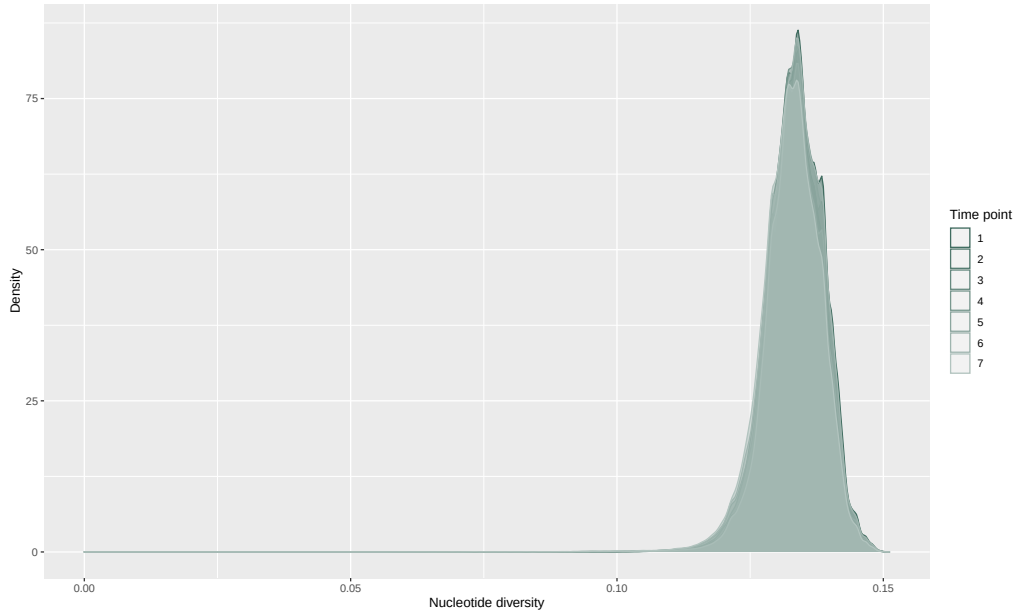


Figure A.17: **Relative nucleotide diversity density per time point for all replicate experiments and populations in the Vlachos et al. (2019) dataset.** Diversity values presented are per site.

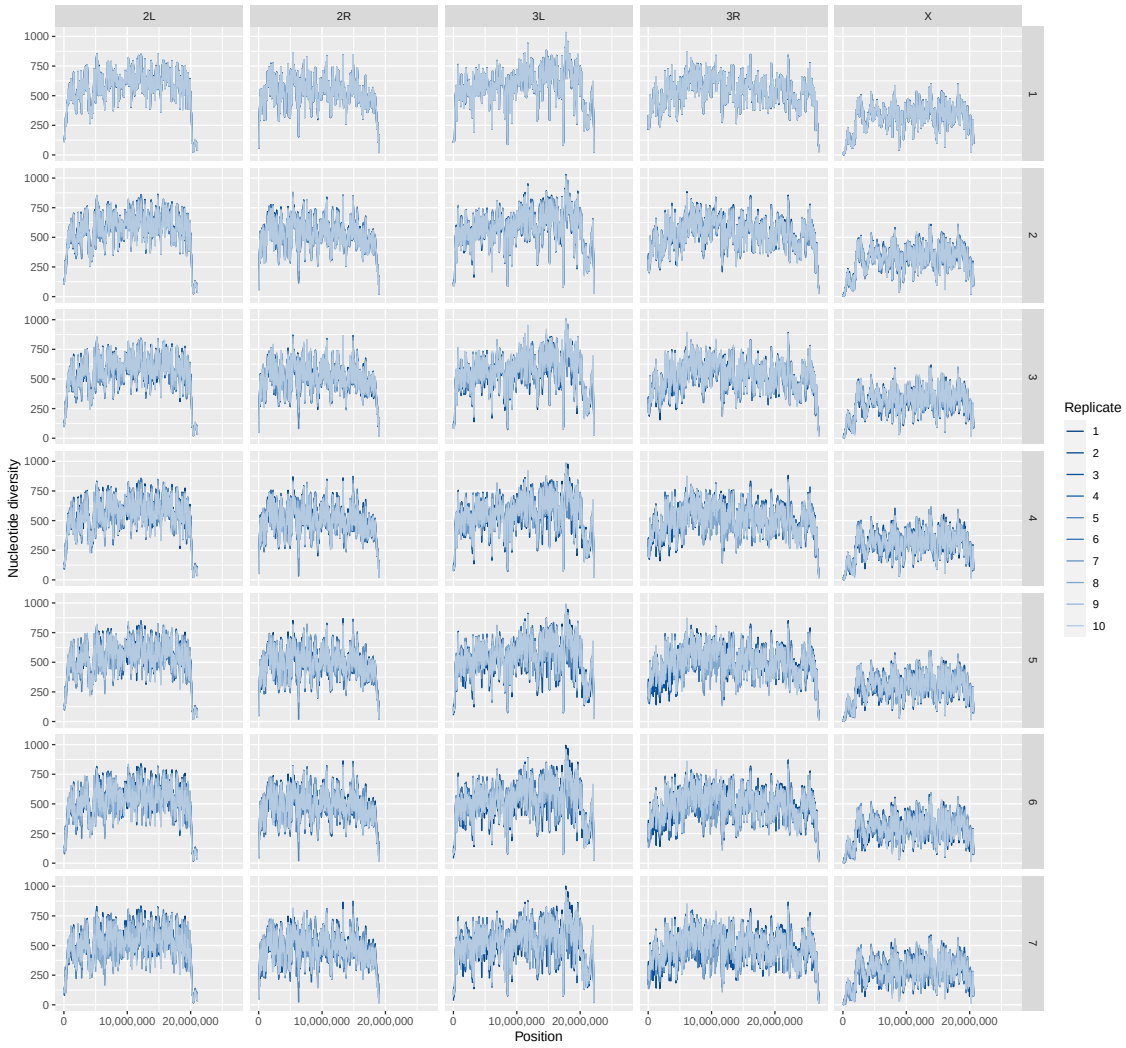


Figure A.18: **Absolute nucleotide diversity along each chromosome (excluding 4) in the Barghi et al. (2019) dataset.** Diversity values presented are per window. Rows correspond to consecutive time points.

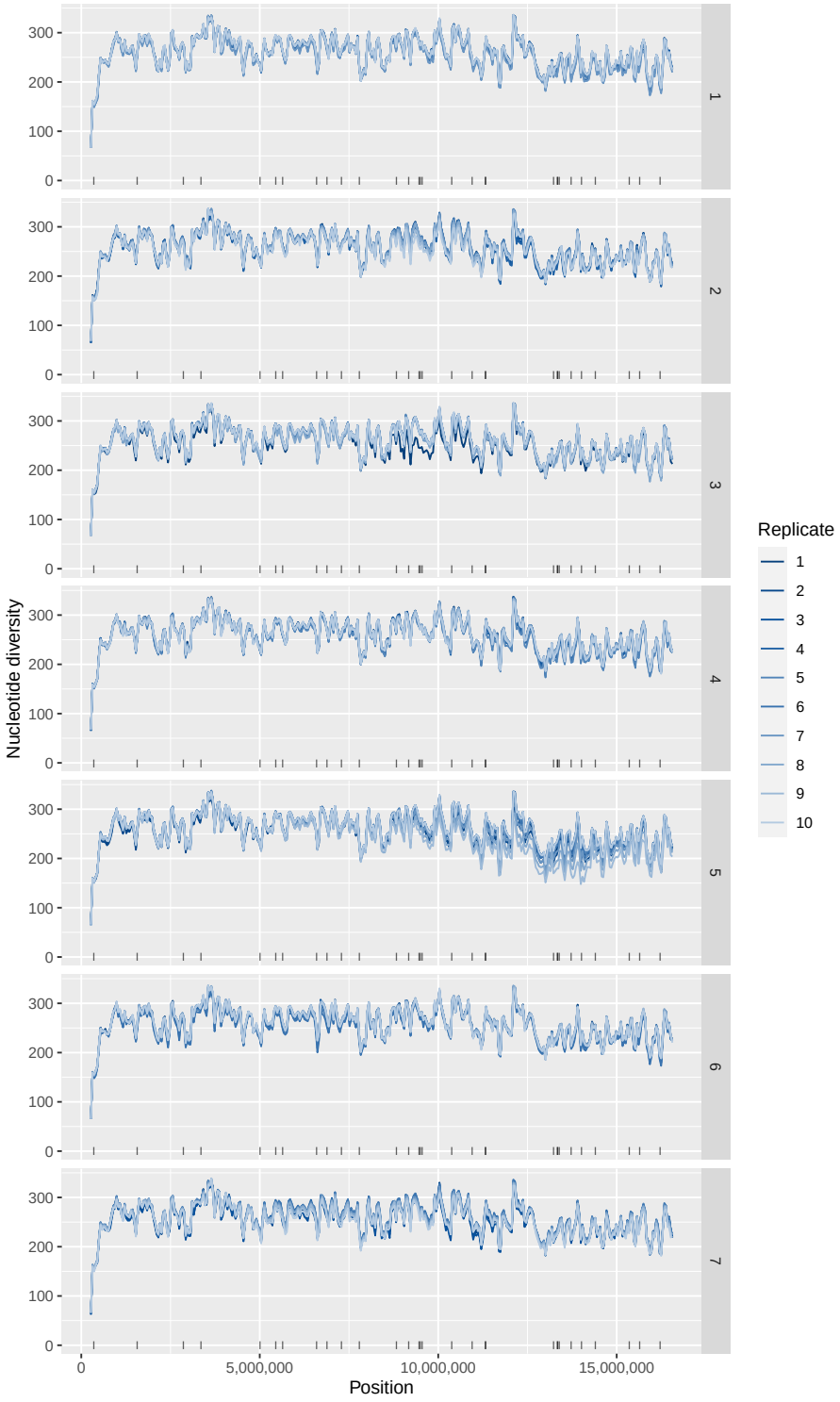


Figure A.19: **Absolute nucleotide diversity along the genomic element simulated by Vlachos et al. (2019) – replicate experiment #18.** Diversity values presented are per window. Rows correspond to consecutive time points. Black solid bars at the bottom of each individual plot indicate the location of a selected site.

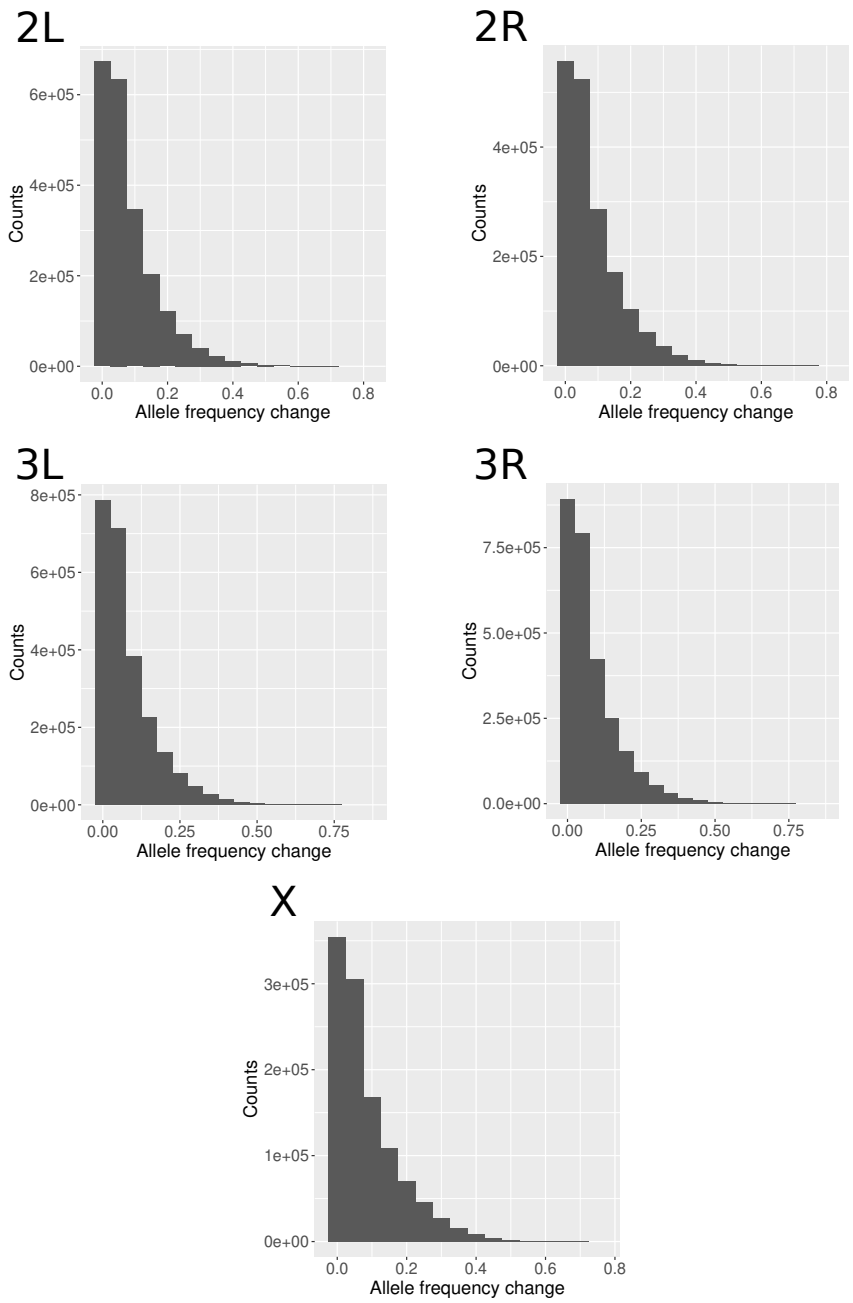


Figure A.20: **Histogram of total allele frequency changes in Barghi et al. (2019) chromosomes.** Excludes data on the fourth chromosome.

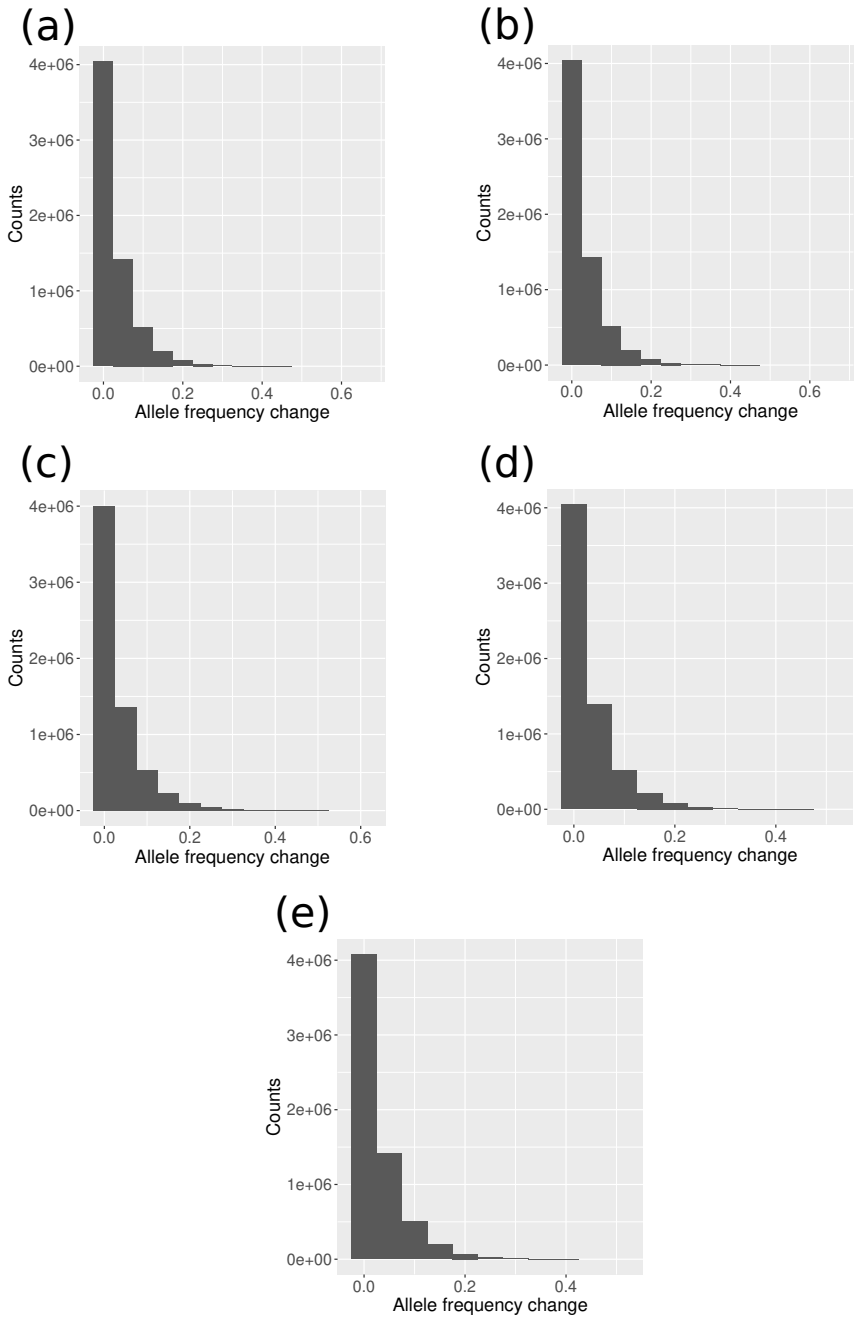


Figure A.21: **Histogram of total allele frequency changes in five Vlachos et al. (2019) sweep scenario experiments.** Includes data on replicate experiments (a) #2, (b) #7, (c) #49, (d) #76 and (e) #100.

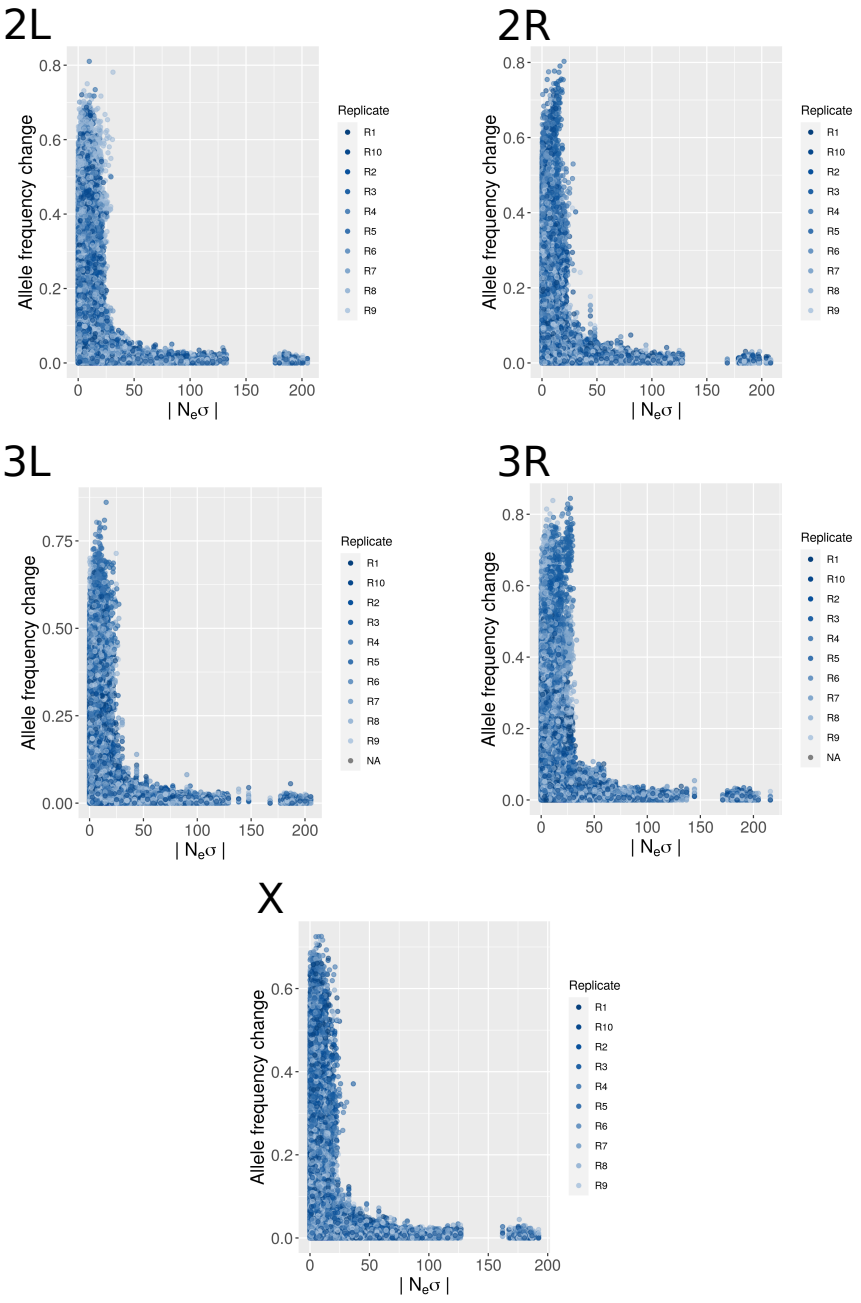


Figure A.22: **Total allele frequency changes for each locus in the Barghi et al. (2019) dataset versus $N_e\sigma$.** Excludes data on the fourth chromosome.

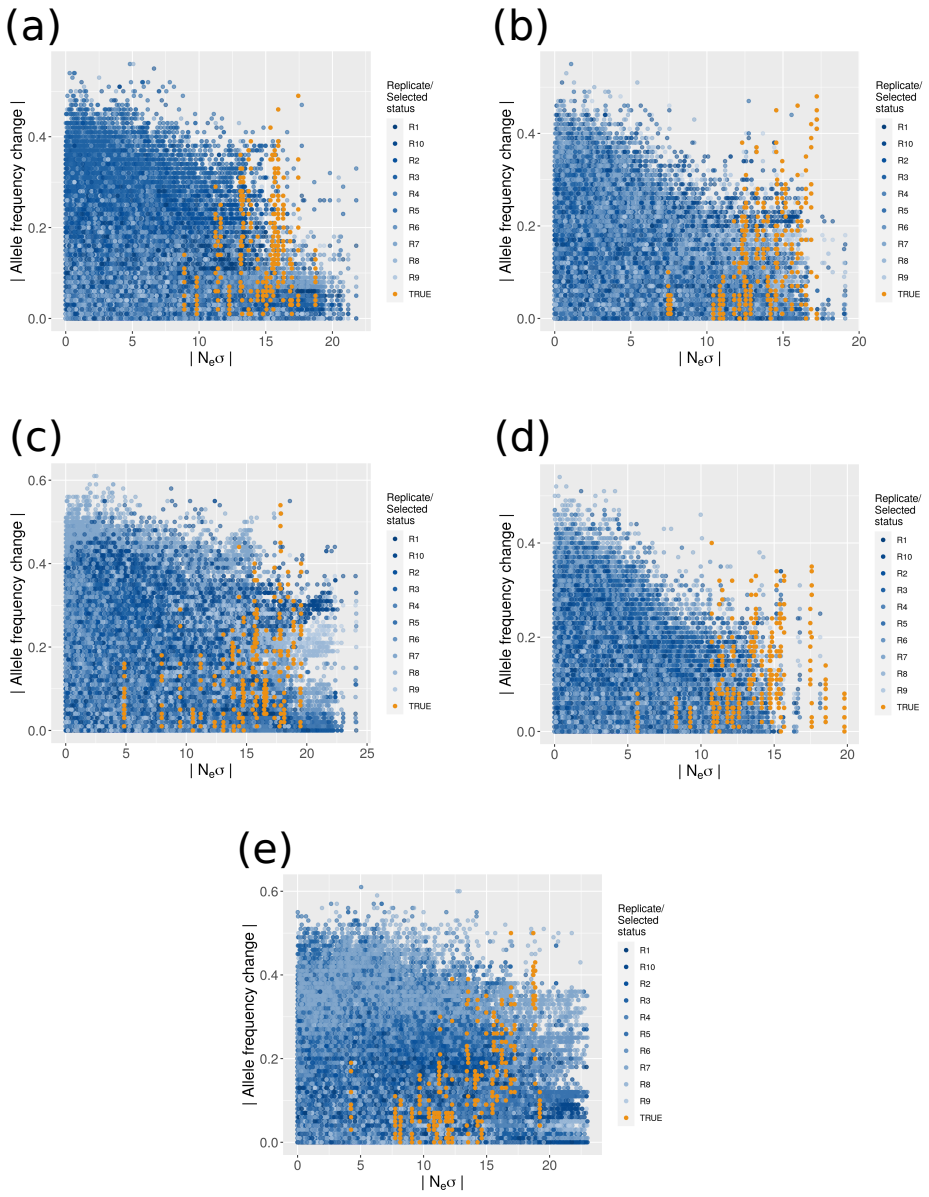


Figure A.23: **Total allele frequency changes for each locus in the Vlachos et al. (2019) dataset versus $N_e\sigma$.** Includes data on replicate experiments (a) #31, (b) #31, (c) #47, (d) #54 and (e) #76. Orange points indicate true selected sites.

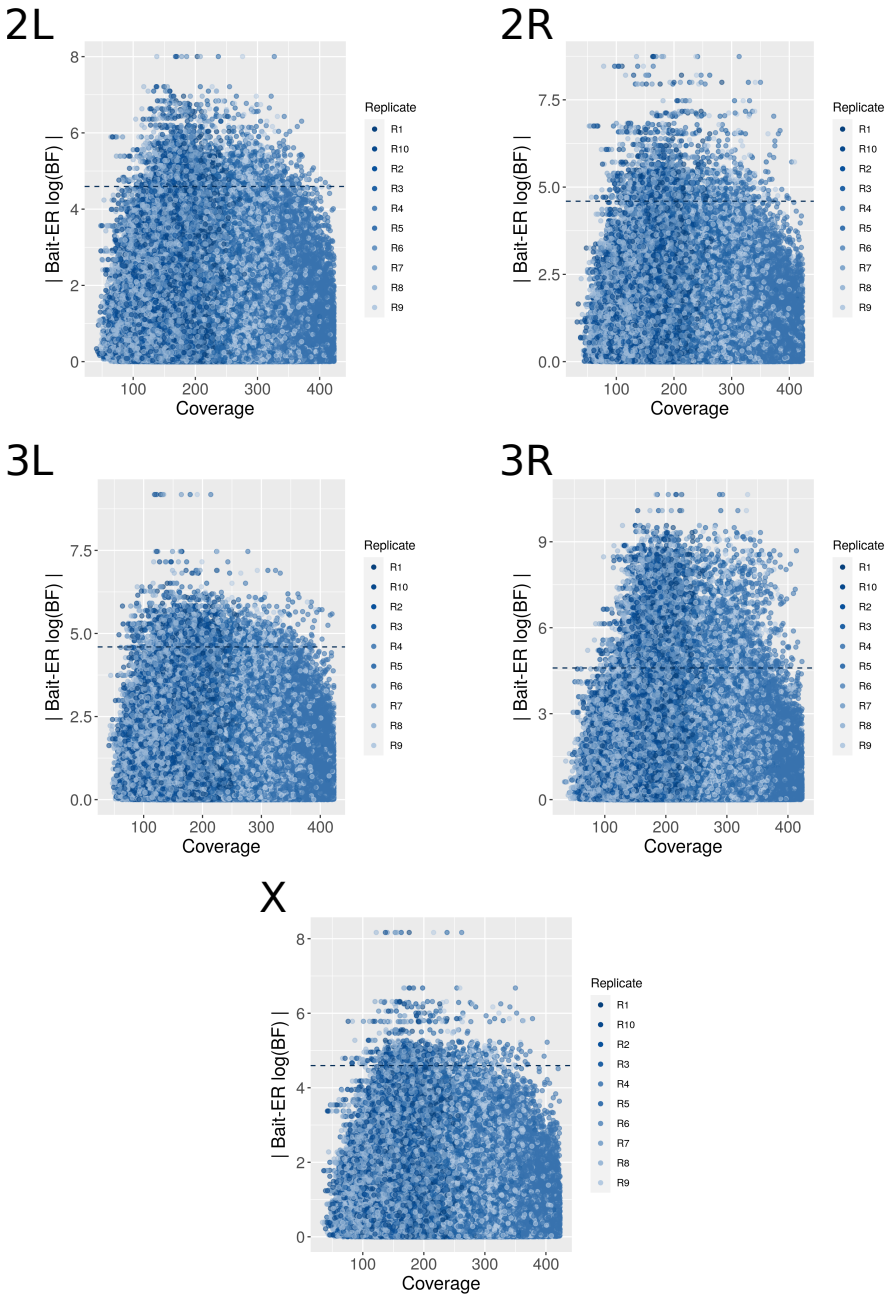


Figure A.24: **Scatterplot of the relationship between coverage at generation 0, i.e. time point 1, and Bait-ER logBFs for chromosomes in the Barghi et al. (2019) dataset.** Excludes data on the fourth chromosome.

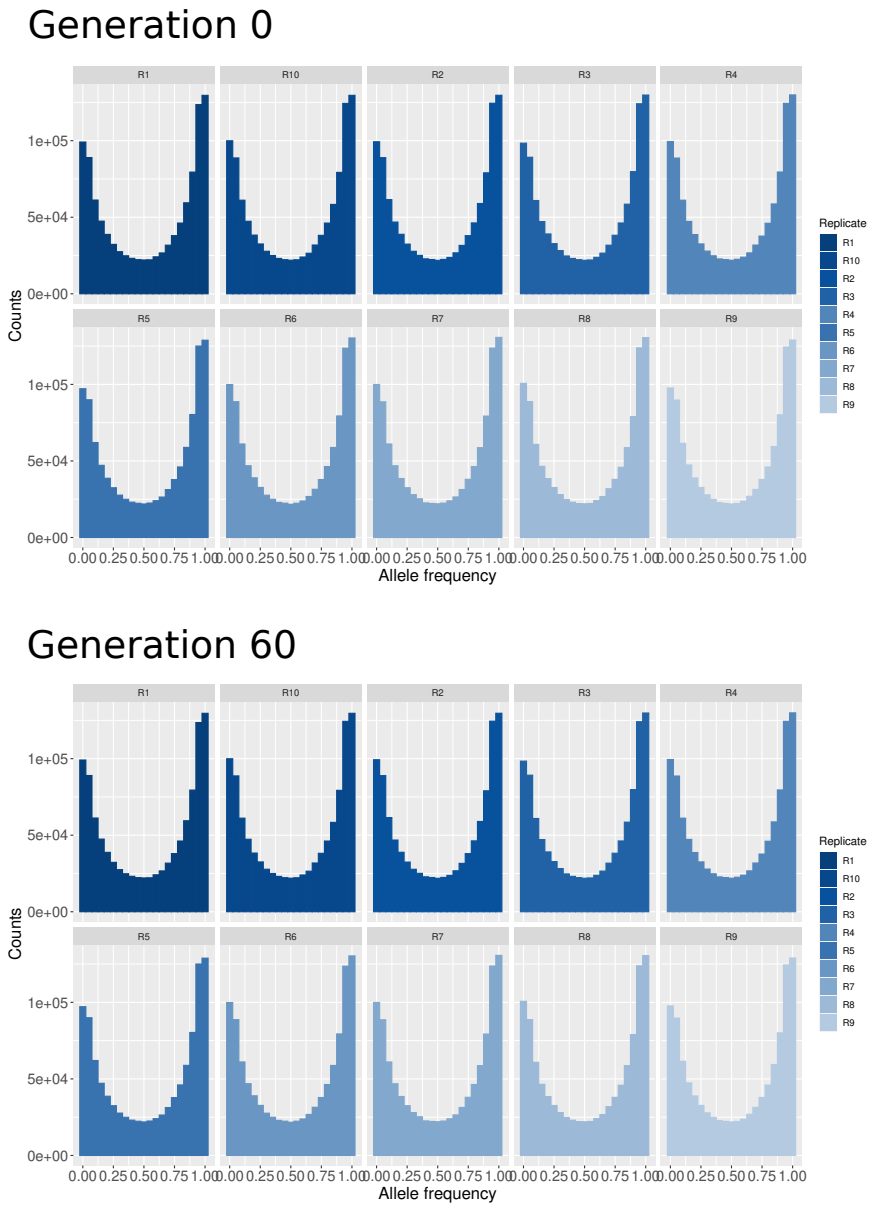


Figure A.25: **Allele frequency spectra of chromosome 2L in Barghi et al. (2019) at generations 0 (top) and 60 (bottom), i.e., time points 1 and 7.** Each replicate population is represented in its own individual histogram plot.

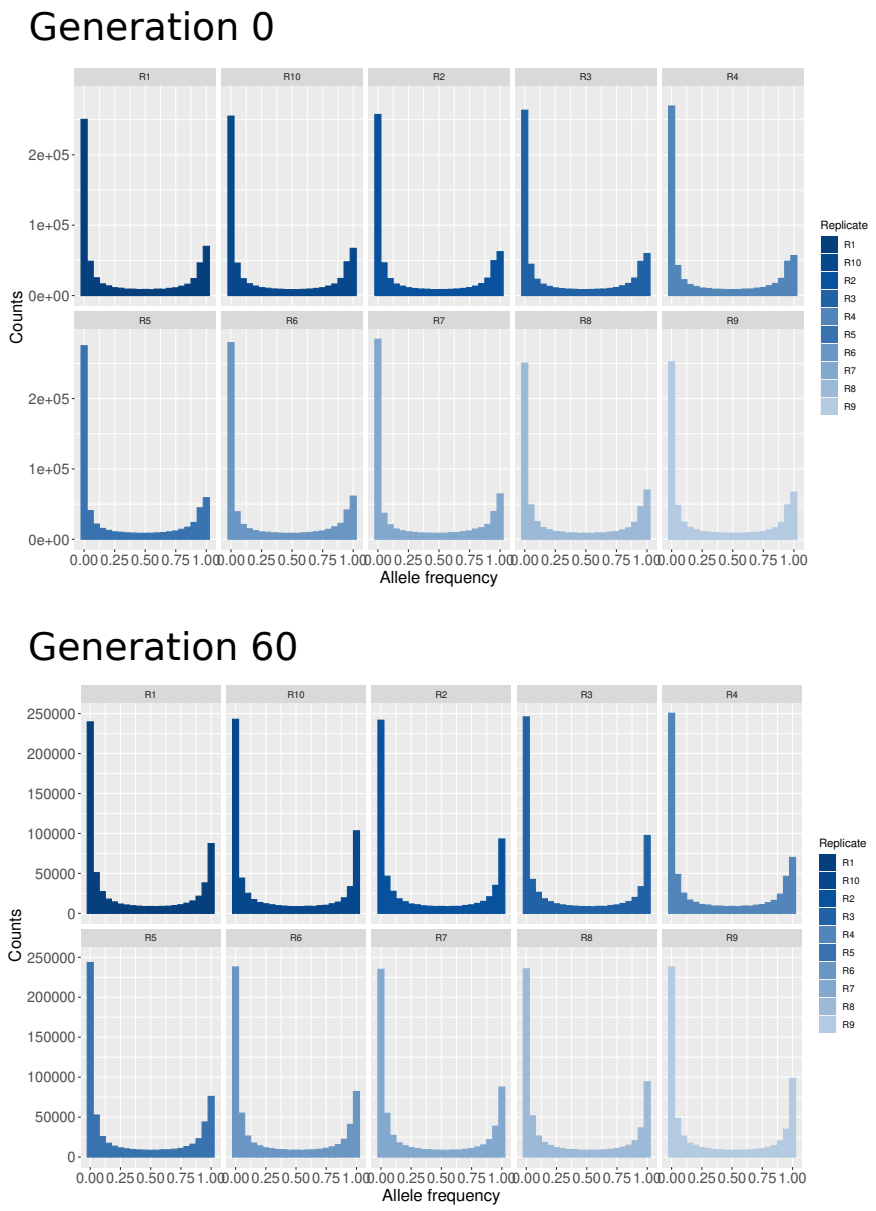


Figure A.26: **Allele frequency spectra in replicate experiment #98 of the Vlachos et al. (2019) study at generations 0 and (top) and 60 (bottom), i.e., time points 1 and 7.** Each replicate population is represented in its own individual histogram plot.

Appendix B

Supplement to chapter 3

	Generation				
Replicate	**TP 1**	**TP 2**	**TP 3**	**TP 4**	**TP 5**
M/E 1	21	63	116	164	200
M/E 2	21	62	115	163	200
M/E 3	21	61	114	160	200
M/E 4	22	59	112	160	200

Table B.1: **Description of which generations were sampled at each time point.** Both M and E lines were sampled at the same generation for each corresponding replicate population. TP: time point.

Treatment	Mapper	Average no.	Average %	Min no. (%)	Max no. (%)	Assembly
M	bwa	48,315,592	99.0	37,798,247 (98.6)	62,704,196 (98.8)	Genome
M	novoalign	48,629,415	99.2	38,092,806 (98.9)	61,869,582 (99.2)	Genome
E	bwa	52,169,450	99.0	43,105,534 (98.8)	65,961,589 (99.2)	Genome
E	novoalign	52,568,742	99.2	43,479,141 (99.1)	66,521,236 (99.3)	Genome
M	bwa	22,269,730	98.4	17,630,965 (98.1)	28,450,846 (98.3)	X chr
M	novoalign	22,712,923	98.8	17,883,361 (98.6)	29,118,289 (98.8)	X chr
E	bwa	24,014,102	98.4	19,722,067 (98.3)	29,666,606 (98.8)	X chr
E	novoalign	24,494,392	98.9	20,185,805 (98.8)	30,279,204 (99.0)	X chr

Table B.2: **Mapping statistics for both mappers.** This includes average number of mapped reads as well as the average percentage across samples for M and E lines. Data on genome and X chromosome level assemblies can be found on this table.

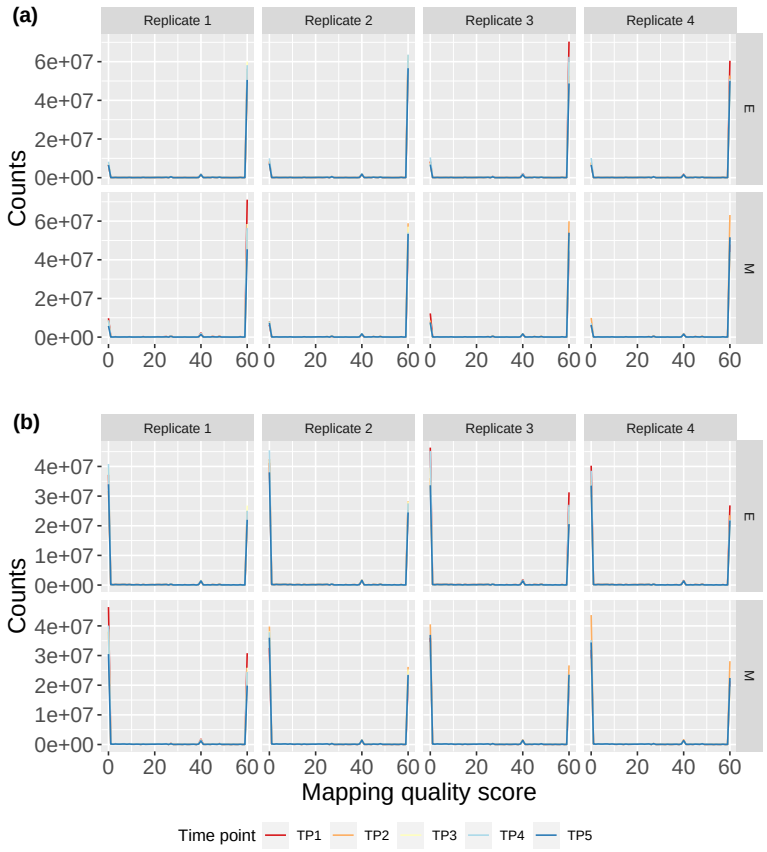


Figure B.1: **Mapping quality score distribution at the genome (a) and X chromosome (b) level assemblies.** Rows correspond to the two treatments, E (top) and M (bottom), and columns to the four experimental replicates. Each time point is coloured differently as per legend at the bottom.

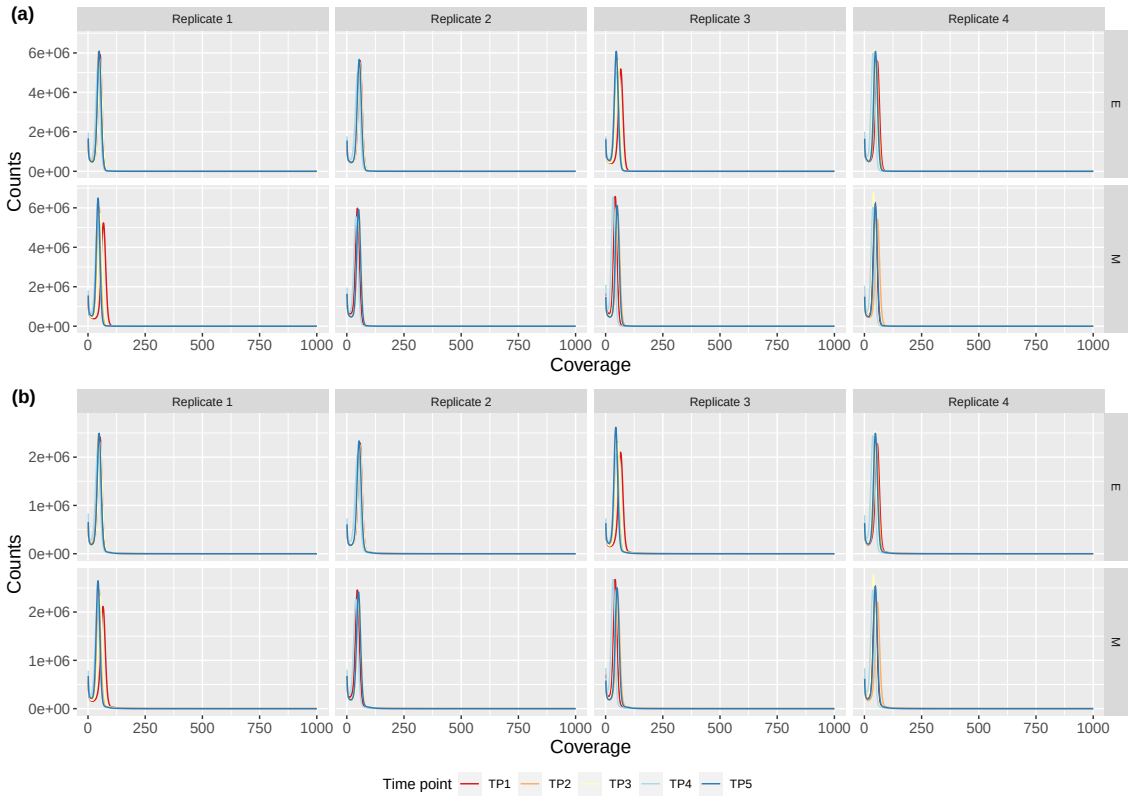


Figure B.2: **Read coverage distribution at the genome and X chromosome level assemblies.** Each column corresponds to an different replicate population and each row to either E (top) or M (bottom) lines. The top panel - (a) - shows genome level data and the bottom panel - (b) - X chromosome data. Different time points are coloured according to legend at the bottom of the figure.

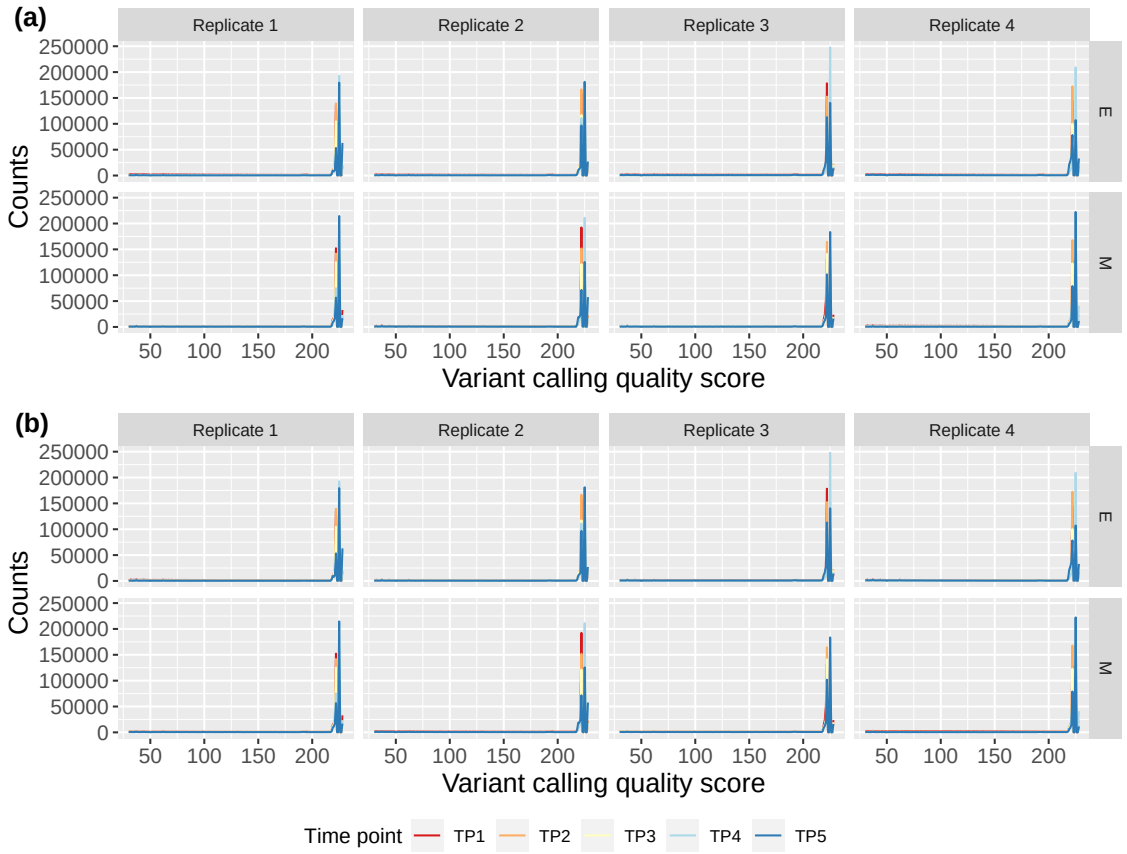


Figure B.3: **Variant calling phred quality score per time point for each replicate.** (a) shows quality score distributions at the genome level and (b) at the X chromosome level assemblies. Each row corresponds to a different treatment - E on top row and M on bottom - and each column to an individual replicate. Time points are visible by differently coloured lines as per legend at the bottom of the graph.

Treatment	Time point	Unfiltered	Filtered round #1 (vs unfiltered)		Both callers (vs filtered round #1)		Filtered round #2 (vs both callers)		Freebayes (vs unfiltered)		Assembly level
M	TP1	2,787,319	2,477,272	(-310,046.25)	1,466,325	(-1,010,947.75)	1,166,114.75	(-300,209.75)	3,156,979	(+369,660.75)	genome
M	TP2	2,491,572	2,303,865	(-187,707)	1,435,954	(-867,911.25)	1,387,945.25	(-48,008.5)	2,305,271	(-186,300.75)	genome
M	TP3	2,351,323	2,182,542	(-168,781.5)	1,384,611	(-797,930.75)	1,358,823.50	(-25,787.5)	2,117,929	(-233,394.75)	genome
M	TP4	1,967,133	1,847,275	(-119,857.75)	1,240,078	(-607,197.75)	1,204,845.00	(-35,232.5)	1,632,634	(-334,499)	genome
M	TP5	1,948,784	1,842,966	(-105,817.75)	1,209,587	(-633,379.25)	1,189,898.50	(-19,688.5)	1,754,641	(-194,142.75)	genome
E	TP1	2,843,322	2,547,546	(-295,776)	1,498,669	(-1,048,876.25)	1,214,163.25	(-284,506)	2,910,488	(+67,166.5)	genome
E	TP2	2,453,723	2,279,650	(-174,072.75)	1,430,774	(-848,876)	1,386,282.50	(-44,491.5)	2,279,557	(-174,165.75)	genome
E	TP3	2,135,627	2,013,440	(-122,186.5)	1,301,792	(-711,648)	1,283,460.00	(-18,332.25)	1,861,553	(-274,074.25)	genome
E	TP4	1,910,089	1,802,535	(-107,553.75)	1,210,829	(-591,705.75)	1,176,927.25	(-33,901.75)	1,573,896	(-336,193)	genome
E	TP5	2,062,411	1,929,839	(-132,571.5)	1,242,848	(-686,991.5)	1,214,033.25	(-28,814.5)	1,979,512	(-82,898.75)	genome
M	TP1	1,047,522	942,455	(-105,067)	565,788	(-376,666.5)	–	–	1,182,418	(-134896)	X
M	TP2	925,122	852,754	(-72,368.5)	528,325	(-324,429)	–	–	900,689	(+24433.25)	X
M	TP3	871,490	804,422	(-67,068)	506,213	(-298209)	–	–	832,029	(+39461)	X
M	TP4	731,140	676,547	(-545,92.75)	450,819	(-225728.5)	–	–	654,219	(+76920.5)	X
M	TP5	739,132	687,412	(-517,19.75)	445,051	(-242361.75)	–	–	726,722	(+12410.25)	X
E	TP1	1,055,999	949,923	(-106,076)	560,418	(-389505)	–	–	1,100,105	(-44105.5)	X
E	TP2	915,083	846,588	(-68,494.5)	530,034	(-316553.75)	–	–	897,880	(+17202.25)	X
E	TP3	777,659	724,817	(-52,841.5)	465,100	(-259717.25)	–	–	733,010	(+44648.25)	X
E	TP4	690,257	642,877	(-47,379.25)	430,364	(-212513.75)	–	–	611,128	(+79128.25)	X
E	TP5	770,680	711,135	(-59,544.75)	452,685	(-258450)	–	–	810,071	(-39391.25)	X

Table B.3: **Number of SNPs at different stages of parsing for M and E lines.** 'Unfiltered' are the average total number of SNPs at each time point called by bcftools before any parsing. Subsequent filtering for variants called with a quality score of at least 30 resulted in the 'Filtered round #1' column, where numbers in brackets are the number of SNPs lost in this parsing step. Only SNPs that were called by both bcftools and Freebayes were retained after the first filtering step - 'Both callers'. Variants lost here are in brackets. 'Filtered round #2' included keeping biallelic sites only, as well as retaining solely those variants that were called both in the bwa mem and the novoalign alignment. In first time point samples, only polymorphic sites with a MAF ≥ 0.025 were kept. SNPs called by Freebayes are included for comparison ('Freebayes'). Figures in brackets here are the difference to 'Unfiltered' polymorphisms.

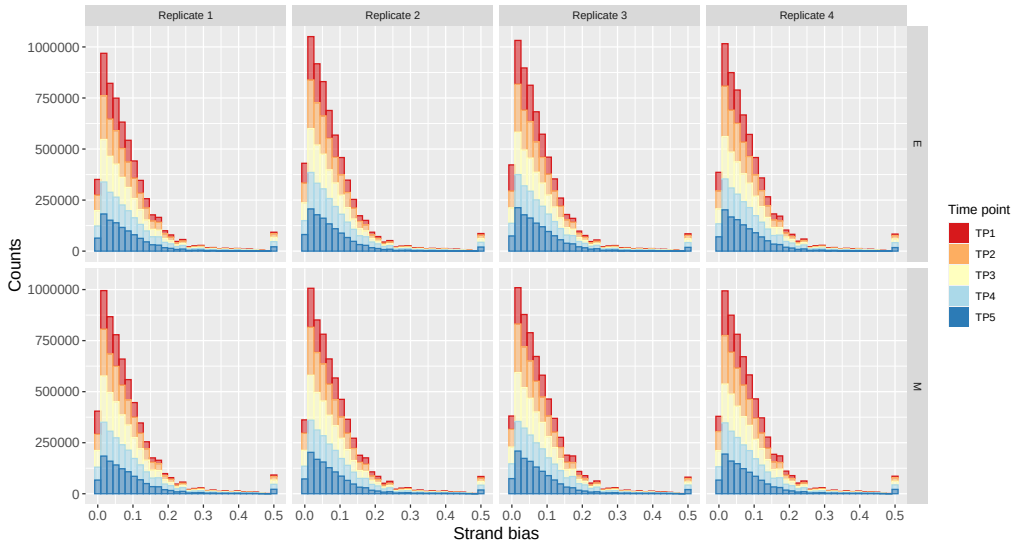


Figure B.4: **Strand bias after filtering per time point for each of the four experimental replicates.** E populations are on the top tow, and M on the bottom. Each column corresponds to an individual replicate. Time point distributions are coloured differently as per side legend.

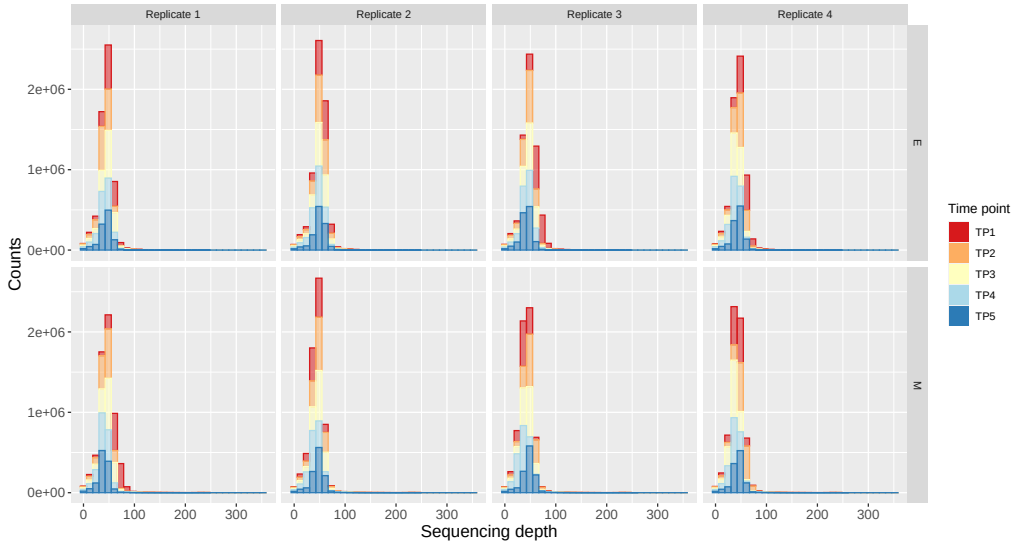


Figure B.5: **Sequencing depth distribution per time point for all variants called and retained after filtering.** Replicates can be found in columns, and treatments in rows (E: top, M: bottom). Distributions for each time point were coloured differently as per side legend.

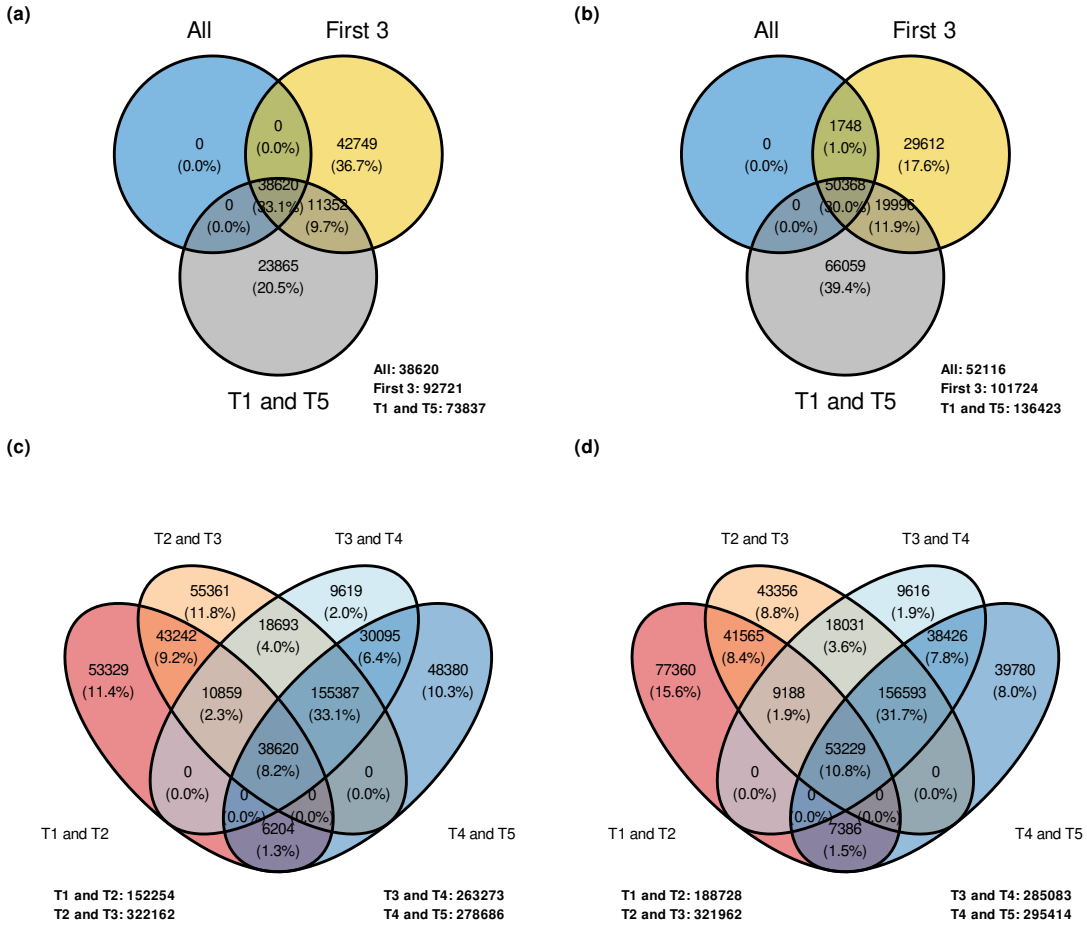


Figure B.6: **Venn diagrams that compare the final number of SNPs analysed between different time point interval datasets.** Panels (a) and (b) show number of SNPs found in a five time point time series ('All'), a three time point time series ('First 3') or the first and last time points (T1 and T5) for M and E, respectively. Bottom panels (c) – M lines – and (d) – E lines – compare two time point intervals: T1 and T2, T2 and T3, T3 and T4, and T4 and T5.

Interval	Treatment	Chr 2	Chr 3	Chr 4	Chr X	Total
All time points	M	11128	10537	6593	9807	38065
First three	M	25300	22547	18568	25191	91606
T1T5	M	21084	17144	13186	21396	72810
T1T2	M	43477	33217	30954	42911	150559
T2T3	M	78318	58094	63784	119452	319648
T3T4	M	68139	45920	51847	95260	261166
T4T5	M	71413	45159	57190	102471	276233
All time points	E	12189	14021	11422	13707	51339
First three	E	25358	23347	23278	28438	100421
T1T5	E	32883	26935	27344	47587	134749
T1T2	E	48018	35596	38963	64042	186619
T2T3	E	78890	54384	75011	111092	319377
T3T4	E	70816	46865	68182	96712	282575
T4T5	E	70666	47206	75381	99531	292784

Table B.4: **Final number of SNPs per treatment for several time point intervals used for subsequent analyses.**

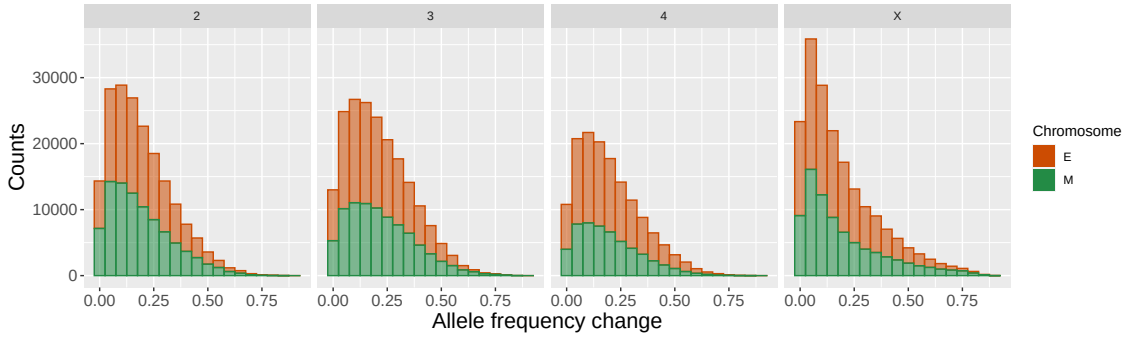


Figure B.7: **Allele frequency change histograms in M (in green) and E (in orange) populations for each chromosome (columns).** These are calculated as the difference in allele frequency between first and last time point for each individual SNP.

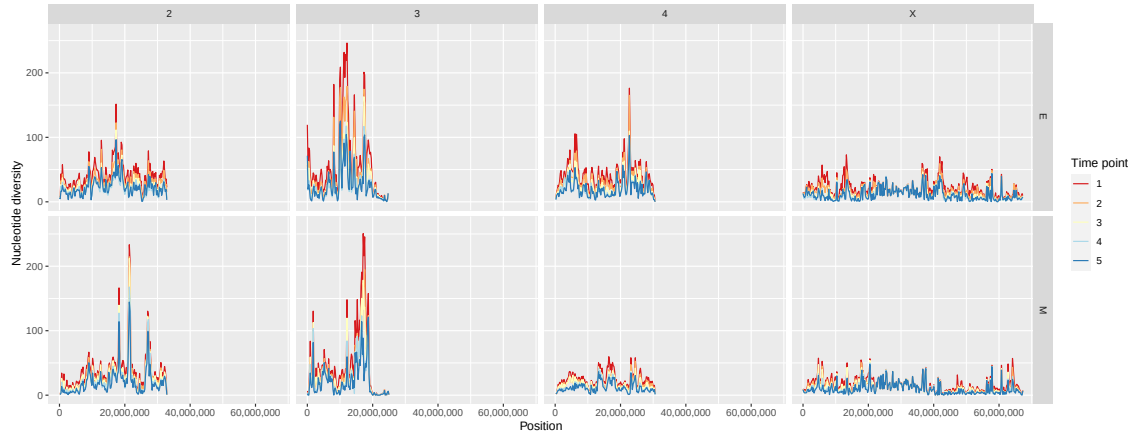


Figure B.8: **Average nucleotide diversity, π, along the genome.** Columns correspond to chromosomes and rows to the two different treatments (top: E; bottom: M). Lines are coloured as to show variation across time. Averages were calculated across replicates for each 250k SNP window separately.

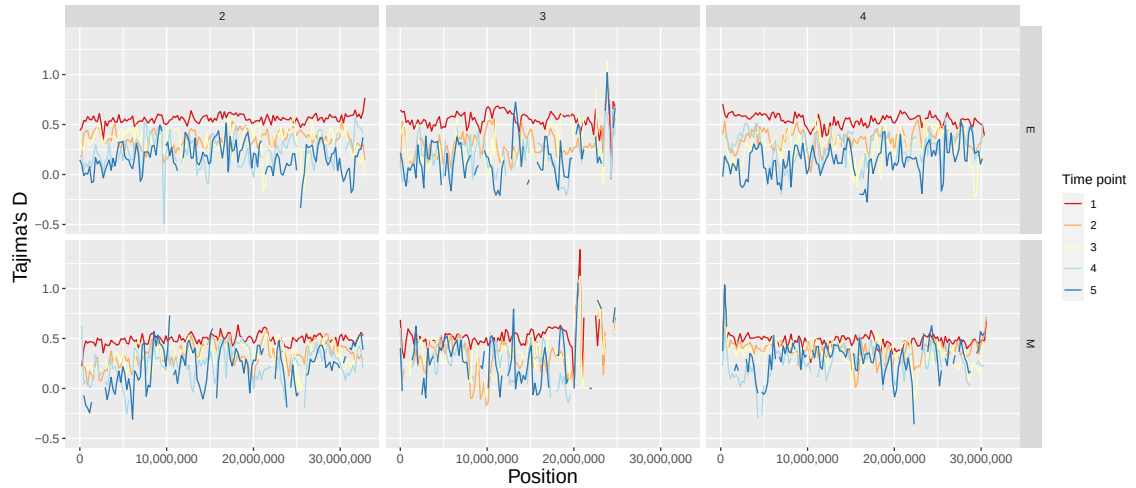


Figure B.9: **Tajima's D estimates along chromosomes 2, 3 and 4 for E and M lines.** Rows correspond to the two different treatments and columns to chromosomes. Estimates were calculated in 250k SNP windows with Grenedalf (Czech and Exposito-Alonso, 2021). Lines are coloured per time point according to the side legend.

Time interval	Median - M	Median - E
Overall	151.0 (n = 100)	159.2 (n = 223)
T1T2	90.0 (n = 212)	85.7 (n = 308)
T2T3	68.2 (n = 246)	111.9 (n = 239)
T3T4	73.8 (n = 131)	102.9 (n = 112)
T4T5	134.8 (n = 107)	145.8 (n = 116)

Table B.5: **Median genome-wide N_e estimates for M and E lines at different time point intervals using intergenic SNPs only.** Medians were calculated using 1k intergenic SNP window estimates from all of the four experimental replicates. 'Overall' corresponds to N_e estimates based on allele frequency changes between the first and last time point. The total number of windows considered in each replicate is found in brackets.

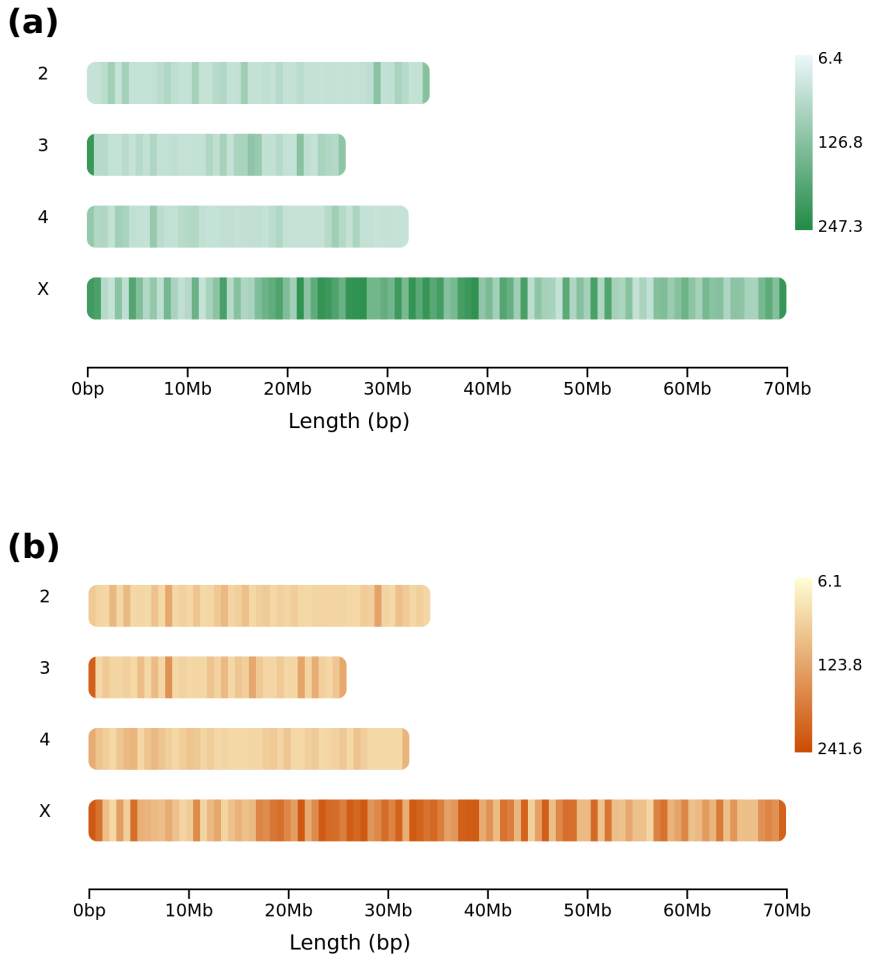


Figure B.10: **Maximum coverage chromosome plots for (a) M and (b) E lines.** Each diagram represents the maximum coverage of any given interval in each of the four chromosomes analysed. M lines are coloured in green and E in orange. Highest coverage regions can indicate repetitive elements and are common in telomeres and centromeres. Data exclude chromosome 5.

Treatment	Chromosome	Location	NCBI ID	FB ID	*D. mel* ortholog	Gene name	#top SNPs	Distance
E	2	Gene	6896726	FBgn0262041	hdc	headcase protein	1	NA
E	3	Gene	4805669	FBgn0073138	SmydA-5	SET domain-containing protein SmydA-8	1	NA
E	3	Gene	4805666	FBgn0080959	Cyp4aa1	cytochrome P450 4aa1	1	NA
E	3	Gene	6899100	FBgn0246519	Pgant9	polypeptide N-acetylgalactosaminyltransferase 9	1	NA
E	3	Gene	6899112	FBgn0246524	Verprolin 1	WAS/WASL-interacting protein family member 3	1	NA
E	3	Gene	6899088	FBgn0263807	Strn-Mlck	myosin light chain kinase	2	NA
E	X	Gene	4815279	FBgn0075585	fzr	fizzy-related protein homolog	3	NA
E	X	Gene	4814581	FBgn0078871	forked	espin	3	NA
E	X	Gene	4815373	FBgn0080302	CG7378	dual specificity protein phosphatase 3	1	NA
E	X	Gene	4814961	FBgn0081931	CG9657	sodium-coupled monocarboxylate transporter 1	1	NA
E	X	Gene	4814480	FBgn0243547	Atg5	autophagy protein 5	4	NA
E	X	Gene	6901434	FBgn0247979	dpr8	zwei Ig domain protein zig-8	5	NA
M	X	Gene	6902114	FBgn0244416	CG32532	homeobox protein Hmx	1	NA
E	3	Intergenic region	4805672	FBgn0073140	CCHa2-R	neuropeptide CCHamide-2 receptor	1	-421
E	3	Intergenic region	6899092	FBgn0245480	resilin	pro-resilin	2	-579
E	4	Intergenic region	4816335	FBgn0077513	CG3528	cilia- and flagella-associated protein 299	1	-9236
E	X	Intergenic region	4815048	FBgn0074111	unc-119	protein unc-119 homolog	1	2760
E	X	Intergenic region	4814454	FBgn0080287	Flacc	fl(2)d-associated complex component	1	1833
E	X	Intergenic region	4814479	FBgn0081927	brinker	J domain-containing protein DDB_G0295729	1	-3996
E	X	Intergenic region	6901756	FBgn0249747	CG15034	uncharacterised protein	1	-516
E	X	Intergenic region	6901755	FBgn0249791	-	antigen 5 like allergen Cul n 1	1	1613

Table B.6: **Common genes amongst top scoring variants in this study and Wiberg et al. (2021).** This includes data on genes where significant variants are located in or genes near top SNPs. Each gene is described by an NCBI ID, a FlyBase ID, a *D. melanogaster* ortholog, the gene's name, how many top SNPs mapped on to the gene or the distance between any intergenic SNPs and the nearest gene.

Chromosome	Treatment	# top SNPs	Gene NCBI ID	Gene FB ID	*D. mel* ortholog	Gene name
X	E	7	4813031	FBgn0076699	Lmx1a	LIM homeobox transcription factor 1-beta
X	E	9	4813494	FBgn0076929	SpoCk	calcium-transporting ATPase type 2C member 1
X	E	11	4813557	FBgn0076932	neuromusculin	hemicentin-1
X	E	6	4814416	FBgn0077610	Fas2	fasciclin-2
X	E	5	4814942	FBgn0079214	Nep1	neprilysin-1
3	E	5	6899052	FBgn0263814	jing	zinc finger protein jing
3	E	5	6899426	FBgn0250039	luna	Krueppel-like factor luna
X	E	23	6901139	FBgn0243670	CG43867	uncharacterised protein
X	E	5	6901434	FBgn0247979	dpr8	zwei Ig domain protein zig-8
X	E	7	6901448	FBgn0247929	Sh	potassium voltage-gated channel protein Shaker
X	E	7	6901459	FBgn0247897	CG5921	uncharacterised protein
X	E	5	6901716	FBgn0245264	-	voltage-dependent T-type calcium channel subunit alpha-1G
X	E	9	6901746	FBgn0245096	SK	small conductance calcium-activated potassium channel protein

Table B.7: **Genes with the most significant variants.** Each gene is described by the number of top SNPs located in said gene, an NCBI ID, a FlyBase ID, a *D. melanogaster* ortholog and the gene's name.