The Genetic Basis of Complex Traits Studied via Analysis of Evolve and Resequence Experiments

by

Stefanie Belohlavy

March, 2022

A thesis submitted to the
Graduate School
of the
Institute of Science and Technology Austria
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

Committee in charge:

Gašper Tkačik, Chair
Nicholas H. Barton
Beatriz Vicoso
Reinhard Bürger

The thesis of Stefanie Belohlavy, titled *The Genetic Basis of Complex Traits Studied via Analysis of Evolve and Resequence Experiments*, is approved by:

Supervisor: Nicholas H. Barton, IST Austria, Klosterneuburg, Austria

Signature: _____

Committee Member: Beatriz Vicoso, IST Austria, Klosterneuburg, Austria

Signature: _____

Committee Member: Reinhard Bürger, University of Vienna, Vienna, Austria

Signature: _____

Defense Chair: Gašper Tkačik, IST Austria, Klosterneuburg, Austria

Signature: _____

Signed page is on file

IST Austria Thesis, ISSN: 2663-337X

ISBN: 978-3-99078-018-3

I hereby declare that this thesis is my own work and that it does not contain other people's work without this being so stated; this thesis does not contain my previous work without this being stated, and the bibliography contains all the literature that I used in writing the dissertation.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee, and that this thesis has not been submitted for a higher degree to any other university or institution.

I certify that any republication of materials presented in this thesis has been approved by the relevant publishers and co-authors.

Signature: _____

Stefanie Belohlavy
March, 2022

Signed page is on file

Abstract

In evolve and resequence experiments, a population is sequenced, subjected to selection and then sequenced again, so that genetic changes before and after selection can be observed at the genetic level. Here, I use these studies to better understand the genetic basis of complex traits - traits which depend on more than a few genes.

In the first chapter, I discuss the first evolve and resequence experiment, in which a population of mice, the so-called "Longshanks" mice, were selected for tibia length while their body mass was kept constant. The full pedigree is known. We observed a selection response on all chromosomes and used the infinitesimal model with linkage, a model which assumes an infinite number of genes with infinitesimally small effect sizes, as a null model. Results implied a very polygenic basis with a few loci of major effect standing out and changing in parallel. There was large variability between the different chromosomes in this study, probably due to LD.

In chapter two, I go on to discuss the impact of LD, on the variability in an allele-frequency based summary statistic, giving an equation based on the initial allele frequencies, average pairwise LD, and the first four moments of the haplotype block copy number distribution. I describe this distribution by referring back to the founder generation. I then demonstrate how to infer selection via a maximum likelihood scheme on the example of a single locus and discuss how to extend this to more realistic scenarios.

In chapter three, I discuss the second evolve and resequence experiment, in which a small population of *Drosophila melanogaster* was selected for increased pupal case size over 6 generations. The experiment was highly replicated with 27 lines selected within family and a known pedigree. We observed a phenotypic selection response of over one standard deviation. I describe the patterns in allele frequency data, including allele frequency changes and patterns of heterozygosity, and give ideas for future work.

Acknowledgements

I want to thank my advisor Nick Barton for all his support, and the inspiring and motivating conversations we had about evolution and surrounding topics. I thank Beatriz Vicoso and Reinhard Bürger for their valuable feedback on my work and general support as part of my committee. Also a huge thank you to my collaborators Frank Chan and his group, as well as Guy Reeves and Diethart Tautz for the productive team work and new perspectives.

IST has a very special community and I want to thank everyone who makes that possible, most of all the grad school and everyone in administrative support, to whom I am very grateful for being flexible and solution oriented when things did not go as planned, as is sometimes unavoidable in this line of work.

I thank Andrea from the Vicoso group for her part in the Drosophila project and the nice and interesting conversations we had as a result. I am of course very grateful to everyone in the Barton group, all amazing researchers and people. Here I need to emphasize Michal Hledik's contribution, who gave great feedback and helped me advance my projects, most of all the SNP on Haplotypes one, as well as Oluwafunmilola who also was a big help. Also, Gemma Puixeu Sala, who is my awesome office mate and helped me in all kinds of difficult situations, as well as Dasha for being a very encouraging voice when it was needed most. I also thank Julia for being a really cool office mate and discussing all kinds of systemic issues that irk us over lunch, because who doesn't need to vent sometimes.

Of course I want to thank all my amazing friends, who kept me going when things got tough. Also great thanks to my family - my parents and my sister Babsi, for always supporting me and believing in me. But most of all, thanks to my partner Mel, who helped me with basically everything and continues to be my biggest support. After the defense I will go back to also doing our laundry sometimes, I promise.

About the Author

Stefanie Belohlavy completed a BSc in Physics at the University of Vienna. She continued her studies at the Technical University Dresden as a Master student in Nanobiophysics and wrote her Master thesis at the MPI for Physics of Complex Systems, on "A Model for the Maternal-to-Zygotic Transition" supervised by Prof. Vasily Zaburdaev. In 2015 Stefanie joined IST Austria as a PhD Candidate in the group of Prof. Nick Barton, where she works on understanding the genetic basis of complex traits through mathematical models and simulations. In her main project, she used data from artificial selection experiments to investigate the connection between genotype and phenotype for traits which depend on many genes, leading to a publication in eLife (Castro et al, 2019). She presented her work at various international conferences throughout her PhD and was involved in several scientific outreach projects at IST Austria.

List of Collaborators and Publications

Collaborators:
Yingguang Frank Chan, Friedrich Miescher Laboratory, Tübingen, Germany
Layla Hiramatsu, previously in Frank Chan's group
Guy Reeves, Max Planck Institute for Evolutionary Biology, Plön, Germany
Diethard Tautz, Max Planck Institute for Evolutionary Biology, Plön, Germany

Publications:
J. P. Castro, M. N. Yancoskie, M. Marchini, S. Belohlavy, L. Hiramatsu, M. Kučka, W. H. Beluch, R. Naumann, I. Skuplik, J. Cobb, N. H. Barton, C. Rolian, and Y. F. Chan. An integrative genomic analysis of the Longshanks selection experiment for longer limbs in mice. *eLife*, 8:e42014, June 2019. ISSN 2050-084X. doi: 10.7554/eLife.42014. URL https://doi.org/10.7554/eLife.42014

Table of Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

Sequence data has now become abundant, and is used to answer the question: How much selection is acting in a population and where is it acting along the genome? In my thesis, I investigated this question with a focus on how reliably we can find regions under selection in principle and in practice. Reliably finding these regions currently depends on our ability to correctly identify outliers in summary statistics. This is complicated by correlations in sequence data, which we also refer to as linkage disequilibrium (LD). I derived an equation to estimate the effect of LD in sequence data on the variance in summary statistics. The smaller this variance, the higher the statistical power to find regions under selection. I also analyzed real data from two artificial selection experiments in which populations were sequenced, selected for a specific trait and then sequenced again (evolve and resequence) in order to:

1. quantify the sources of error that occur when looking for selected regions,
2. better understand genetic basis of quantitative traits under selection,
3. quantifying the advantage of haplotype data over individual genotype data, and over allele frequency data when looking for signatures of selection, and
4. discuss the added value, if any, that time series data of the evolve and resequence type has compared to GWAS.

The main reason information gained from sequence data is more limited than the sheer volume of data might suggest is due to the nature of genome evolution: During Mendelian inheritance in eukaryotes, blocks of genome recombine and are passed on to offspring genomes. As a consequence, evolutionary forces act on these blocks, not on individual base pairs or genes. On such blocks, mutation creates single nucleotide polymorphisms (SNP), which are correlated, or in other words, in LD.

1.1 Selection Influences Variance in Sequence Data

In this thesis, I study models with two alleles A and a at each locus exclusively. In the simple haploid case, selection occurs when $w_A \neq w_a$ and the difference in fitness between alleles is called the selection coefficient, $s_A = w_A - w_a$.

In the diploid model, the appropriate one for the data sets I deal with here, there are three possible genotypes: AA, aa and Aa. Selection occurs whenever their fitnesses w_{AA}, w_{Aa} and w_{aa} are not equal. When either $w_{AA} > w_{Aa} > w_{aa}$ or $w_{AA} < w_{Aa} < w_{aa}$, "directional selection" is acting (Nielsen, 2005). We expect this to be the case for the data sets analyzed here (chapters 2 and 4), since in both cases in each generation individuals with the highest trait values (within each family) were selected. The mode by which the researcher picks the top individuals in this type of experiment is called truncation selection (a method) - which then leads to directional selection (an evolutionary force).

This means that one of the two alleles, say A, increases the trait value and is beneficial. This would lead to an increase of diplotypes AA and Aa over time. Directional selection tends to eliminate overall variation within populations (Nielsen, 2005), which is why we expect to see a drop in heterozygosity in our data as well.

The selection experiments analyzed in this thesis were both too short to produce an appreciable amount of new mutation, and we therefore expect adaptation to come almost exclusively from standing variation, i.e. preexisting genetic variation. Adaptation from standing variation is likely to lead to faster evolution and the fixation of more alleles of smaller effects, as well as the spread of more recessive alleles (Barrett and Schluter, 2008).

In general, one of the main effects of selection is to modify the levels of variability in species (Nielsen, 2005). When discussing patterns in genomic diversity that are caused by recent adaptation, we speak of a "selective sweeps" (Hermisson and Pennings, 2017). In the case of a strongly selected new advantageous mutation, we speak of a "hard selective sweep". Selective sweeps outside the mutation-limited scenario, as is the case for our data, are termed "soft sweeps". They start from standing genetic variation, and while they leave less distinctive signatures of selection in the genome, they still produce patterns of altered variation which can in principle be detected with sufficient statistical power (i.e sample size, limited LD etc.).

Hermisson and Pennings (2017) discuss the different types of selective sweeps in terms of the underlying genealogy, or in other words the coalescent history. In terms of general patterns we can say that soft sweeps, like hard sweeps, lead to a decrease in variance at and around the selected site, and an increase in the proportion of low frequency variants in the site frequency spectrum (a histogram counting the number of variants in the population in each frequency class), however the signal is overall weaker, with a narrower core region and nearby flanking regions not strongly dominated by low-frequency variants (Hermisson and Pennings, 2017).

So how should we go about detecting genomic signatures of selection? When an observed locus is not affected by selection, it is said to be neutral; many tests for signatures of selection test against neutrality. However in our case, we know that selection was acting and will therefore employ a different null model, aiming to reject the claim that selection was completely homogeneously distributed, as is expressed by the infinitesimal model, see section 1.5. In other words, we aim to detect patterns of variation which look more heterogeneous than they would under the infinitesimal model, so that we can find evidence for discrete loci of appreciable effect playing a role in the response to selection.

It is also informative to look at the distribution of frequencies after selection, conditional on starting frequencies - an excess of alleles which swept from low to high frequency compared to the infinitesimal expectation can be counted as evidence for discrete loci of major effect. To determine what counts as an excess, we can calculate the expected distribution under the infinitesimal model from simulations and compare to analytical predictions. Including the effects of LD however, which are important in this case, is only possible in simulations.

1.2 Linkage Disequilibrium and Haplotype Structure

Linkage disequilibrium (LD) is defined as the nonrandom association of alleles at two or more loci. However, this does not necessarily mean that there is also linkage, and neither does linkage imply LD (Slatkin, 2008). LD is usually defined by pairwise measures derived from the linkage disequilibrium coefficient,

$$D_{AB} = p_{AB} - p_A \cdot p_B \tag{1.1}$$

where p_{AB} is the frequency of gametes carrying the allele pair AB at two loci, and p_A and p_B are the individual frequencies at these loci. The value of D itself depends on the allele frequencies, which is why scaled measures are often used, such as the correlation coefficient between pairs of loci squared, r^2. This is defined as

$$r^2 = \frac{D^2}{p_A\,(1-p_A)\,p_a\,(1-p_a)} \tag{1.2}$$

While pairwise measures of LD have been shown to fall off quickly along the genome, as in Kruglyak (1999) and Dunning et al. (2000), long range correlations are common (Abecasis et al., 2001). LD patterns vary a lot and the many pairs of loci in a typical 10 kB window of genome might accumulate to an overall large effect.

Correlations in sequence data are mostly due to haplotype structure. A haplotype is defined as a group of alleles which is inherited together from a single parent. This can mean slightly different things depending on context. Here, I think about haplotype structure in terms of ancestry and will therefore in my thesis define a "haplotype block" as a stretch of DNA derived from a specific set of haploid DNA in a particular reference generation - the "founder generation". In an artificial selection experiment, we might consider the point at which selection started as the "founder" or "reference generation", as I do in what follows. For natural populations the definition can be somewhat more tricky and arbitrary; in any case, we ignore any events preceding the reference generation.

A haplotype block as defined here, is bounded by two recombination break points - or one recombination break point and the end of the chromosome. Such blocks of DNA, which are inherited together, carry sites which due to physical linkage are highly correlated. The terms "Haplotype structure" and "LD" are closely related, however "LD" is often used to refer to pairwise measures of correlation, while "haplotype structure" includes higher-order correlations.

The connection between LD and ancestry is deep, as illustrated in McVean (2002); the author quantifies the relationship between LD and the genealogy of DNA sequences of interest. They showed that there is a direct correspondence between the covariance in coalescence times at different parts of the genome and the degree of LD, as measured by r^2 (McVean, 2002).

Haplotype information is highly desirable for a number of applications, e.g. detecting signatures of selection, or informing us about demography, since it gives a fuller picture of the evolutionary history of a population. However, haplotypes are often hard to obtain directly, as sequencing and phasing many individuals is difficult and costly.

It is therefore of interest to estimate haplotypes from data which is easier to obtain, e.g. allele frequency data, for example from pool-sequencing (pool-seq data). When multiple time points and founder haplotypes are available, estimating their relative frequencies is possible, e.g. via maximum likelihood estimation as in Excoffier and Slatkin (1995). Pelizzola et al. (2021) developed a method called haploSep, which uses multiple time points to reconstruct major

haplotypes, and is broadly applicable, even when no candidate haplotypes from other sources are available. It only requires allele frequency data from multiple samples as input and has a run time linear in the number of SNP. Their method was developed for evolve and resequence experiments. How well reconstruction works depends on some design parameters; for example it will be more accurate for bigger haplotype frequency changes, as is expected under strong selection, a moderate number of haplotypes at sufficiently high frequencies, and a sufficiently high number of samples. Since the method requires haplotypes to be present in the same form for a sufficient number of samples, recombination can pose a problem, and therefore DNA segments need to be short enough to safely neglect it. In general, methods for long regions of genome are still lacking, but methods for haplotype block partitioning have been proposed (Zhang et al., 2002; Pelizzola et al., 2021).

To approach the challenges haplotype structure poses, in my thesis, I am concerned with the concept of a recent reference generation. Recent selection is of special interest, because we often want to understand how current population's genomic features came to be and were influenced in the past tens to hundreds of generations. For simplicity, we would like to ignore the complex ancestry that has shaped a sample, while still taking correlations into account. This is tricky, because samples will show some inscrutable LD structure, which has built up over very long time scales. Such LD needs to be taken into account and corrected for in analyses, and might come up for example in the form of population structure (Marchini et al., 2004; Tiwari et al., 2008).

However, in artificial selection, LD can be estimated from the founder individuals (the first breeding population) of the experiment directly. Also it is possible to have replicates - this allows us to observe parallel evolution, which often gives the most convincing evidence of selection and eases some of the challenges posed by correlation structure. In addition, the hard questions about how population history confounds signatures of selection can be avoided at least partially. From the starting point of the experiment onward, we can know the full pedigree and other specifics about the experimental population. For this reason, data from this type of experiment is appropriate to answer questions about haplotype structure and LD.

GWAS versus Evolve and Resequence

In genome-wide association studies (GWAS), a large population sample is genotyped at many marker SNP, which are then tested for statistical associations with the trait of interest. The goal is to advance our understanding of genotype-phenotype links. GWAS made it first possible to study the genetic basis of complex traits, shifting the focus from the macroscopic to the microscopic level, as earlier analyses had mostly focused on phenotypic observations (Risch and Merikangas, 1996). While associations alone do not provide a direct link to function or mechanism, analyses can sometimes be complemented by other techniques to establish such a link and provide functional information (Visscher et al., 2017).

One major challenge for the study of complex traits in general is that most of the variance these traits exhibit tends to come from small contributions at a large number of loci (Park et al., 2010). In many cases, an impossible sample size would be needed to provide enough statistical power to detect all of these small effects (Rockman, 2012).

In general, the distribution of effect sizes of genes involved in common traits is not well understood though often assumed to be exponential (rather than normal). There is some limited empirical support for this from GWAS (Park et al., 2010). The assumption is also

justified by Orr's model of adaptive walks (Orr, 1998). Turelli (1984) and subsequent work indicate that the distribution has a kurtosis much higher than gaussian.

Statistical power depends on sample size, variant frequency, effect sizes, and LD between the potentially unobserved causal variant and the observed marker SNP. While in this sense GWAS rely on LD, LD also *limits* the power to detect variants, as discussed in the previous section 1.2. Due to these issues with statistical power, a weakness of GWAS is that it mostly says something about alleles of intermediate frequency, which is typically a minority (Visscher et al., 2017). Despite these limitations, GWAS have overall been very successful at identifying loci reproducably associated with complex traits of interest, as long as these loci have appreciable effect sizes and frequencies (Visscher et al., 2017).

In evolve and resequence experiments, a population is sequenced, exposed to selection and sequenced again. The core of the method is to use allele frequency change to infer selection. However, the term itself is not always without confusion, as there exists a continuum from experiments in which a population is simply exposed to a new environment and the resultant changes are observed, to experiments in which the full pedigree is recorded, and selection is for a specific trait. Even when selection is for a specific trait, there is usually some additional natural selection going on; natural in the sense that the experimenter has no control over it. For example, in the Longshanks experiment (Chapter 2), selection was for increased tibia length, but some litters failed independently of this specific trait. In my thesis, I focus on the type of E&R experiment that involves selection for a specific trait. My question here is, what are the potential advantages of the evolve and resequence approach over GWAS, or how can it complement GWAS in a meaningful way?

Some of the limitations of GWAS, e.g. how hard it is to distinguish variants affected by direct selection from those affected by linked selection, and therefore to pinpoint which variants are truly causal, translate directly to the E&R design. However, there are several benefits to following allele frequencies through successive generations, rather than looking at only one snapshot in time, as is done in GWAS. For example consider e.g. how Pelizzola et al. (2021) use multiple time points to reconstruct haplotypes, as discussed in section 1.2. Also, the approach can potentially improve on GWAS when it comes to detecting alleles at high or low frequencies, since starting frequencies can be manipulated experimentally. Also, sustained change in the trait over many generations can be more sensitive than measuring the trait directly at just one point in time. The main drawback are additional experimental effort and cost for both keeping an experiment going for the required number of generations, and additional sequencing effort (Vlachos and Kofler, 2019). In addition to factors relevant for GWAS, statistical power in E&R experiments also depends on other experimental design parameters, like the duration of the experiment or the strength of selection. If the full pedigree is available in addition to time-series data, there are even more possibilities to explore the evolutionary processes acting on the population. Impressive examples include the use of pedigrees in the prediction of the probability of gene loss in captive breeding populations (MacCluer et al., 1986) or understanding genetic models of human disease (Byard, 1986). Especially relevant for this project, pedigree analyses are increasingly used to understand the genetic basis of complex traits (Pemberton, 2008; Chen et al., 2019).

It is possible to study the temporal development of allelic copy numbers via pedigrees as well, as long as genomic data of sufficient quality is available. This has been done by Chen et al. (2019). In their paper, the authors could link individual reproductive success to long-term genetic contribution and allele frequency change, being able to quantify the relative roles of drift, gene flow and selection. However, while the full pedigree allows us to find the exact

effective population size (N_e) and at least in principle makes phasing possible, is is not obvious to what extent it can increase the statistical power for finding selected variants.

Overall, while there are some clear advantages to using time-series data as is available in E&R experiments, it remains somewhat unclear to what extent the additional information can make up for the much larger amount of resources required. If experimental design can be optimized, the advantages might outweigh the drawbacks (Vlachos and Kofler, 2019).

In this thesis, I discuss two E&R experiments with selection for a specific trait and the full pedigree, as well as whole-genome data for (almost) all individuals in the founder and last generations available. In both experiments, selection was within-family, meaning that in each selected generation, individuals with the highest trait values *from each family* were selected and used to establish the next breeding generation, with no sib-matings ever allowed. This specific experimental design choice means that selection does not distort the pedigree, and it is possible to do simulations *conditional* on the pedigree. Intuitively, this is possible because blocks of genome may take different "paths" through the pedigree, and because the randomness produced by Mendelian segregation and variation in the location of recombination break points lead to varying results, even when we start with the same founder haplotypes.

1.3 Windows-based Statistics and Inferring Selection

To feasibly investigate whole genomes, data is often divided into windows. This way, information can easily be simplified by calculating summary statistics, such as the inbreeding coefficient, pairwise nucleotide diversity, or the average change in allele frequency for each window. This is a natural approach, since data tend to cluster due to underlying haplotype structure. Summary statistics for data in each window are usually averaged over statistics calculated on individual SNP and so do not involve LD. This is the most common approach, but there are also summary statistics based on other types of data, including haplotype data. Variability in these summary statistics influences the baseline from which the significance of candidate windows must be judged. The window's size also has to be chosen with care - large enough to limit correlation between windows, but small enough to give detailed information.

Overall there is an abundance of methods to infer selection in genome data analysis. While traditionally mostly simple summary statistics were used, there is now a new generation of methods using inferred gene trees and ancestral recombination graphs (Hejase et al., 2020). These take into account the recent ancestry of a sample as well as LD and contain much richer features than summary statistics derived from SNP alone.

1.4 Complex Traits in Population Genetics and Quantitative Genetics

The term "complex trait" refers to the kind of trait which is influenced by more than a few genes. We are interested in them because many important features fall into this category, such as disease liability scores or many traits important for adaptation to environmental changes in small natural populations of endangered species. We want to understand their genetic basis basis better by learning how selection has acted on them, which is often through small shifts of allele frequencies at many loci. Observing genetic changes after selection is revealing about the distribution of genetic effect sizes which contribute to variance in such traits.

Once we have identified interesting windows of genome or other signatures of selection, we aim to connect them to phenotypic changes. There are many different approaches to do this. Two sub-fields of evolutionary biology and genetics which tackle these questions are population genetics and quantitative genetics. My research is at the interface of these two, since I ask questions about changes in allele and haplotype frequencies which cause phenotypic changes in response to selection.

Quantitative genetics (QG) is the study of traits which are affected by the action of more than a few major genes, called quantitative or complex traits, as described above. Notably, QG and statistics have been intricately connected from the beginning of both disciplines, as early on they were both developed in large part by the same person, namely Ronald A. Fisher.

QG usually focuses on variation in phenotype and its changes under selection. It has been very successful at predicting the response to selection, by which we mean the difference of mean phenotypic value between the offspring of the selected parents and the whole parental generation before selection. These predictions were possible even before the notion of genes was widely used and accepted. QG uses strong assumptions, some of which are biologically unrealistic; yet it has produced methods which work extremely well, as evidenced by their application and success in breeding programs. These programs aim to maximize the response to selection in order to obtain certain breeding goals, for example maximization of milk yield (Hill, 2010; Brotherstone and Goddard, 2005).

The related discipline of population genetics (PG) deals with genetic differences within and between populations. It focuses on frequencies of alleles or combinations of such. Population genetic models can be used for statistical inference from DNA sequence data and to quantify the forces of evolution in general. Analytical results often exist only for a small number of loci, but there are useful approximations for many loci.

Here, my focus is on artificial selection experiments which involve quantitative traits and their genetic basis.

1.5 The Infinitesimal Model

To investigate how selection and random genetic drift change complex traits and distinguish the respective impacts of these forces, it is necessary to define a null model for the genetic basis. How to do this is not obvious; conceptually, for a typical common trait we expect something in-between a few loci with large effect size and the limit of perfectly homogeneously distributed contributions to trait variance along the genome. This is because we know that there is evidence for the involvement of many genes of small effect sizes, but some loci of larger effect sizes have also been found. We also know that there is linkage. In artificial selection experiments such as the ones discussed in this thesis, we additionally know that selection has been acting and was effective, which excludes a purely neutral model as appropriate null model.

In the data sets discussed in this thesis, we also know that linkage is present and that most, if not all chromosomes contributed to the selection response, as would be the case for a trait with a genetic basis consisting of many loci of small effects. Minor shifts in allele frequencies at many loci can lead to signatures of selection that look very different from the type seen in selective sweeps of one major effect allele (Pritchard et al., 2010; Elyashiv et al., 2016). To appropriately model this type of selection response, we therefore choose an infinitesimal model with linkage (Robertson, 1977) as the appropriate null hypothesis.

In the infinitesimal model, we assume an effectively infinite number of loci, each with an infinitesimal effect on the trait under selection (Bulmer, 1980). The classic infinitesimal model does not include linkage, but it can be expanded to do so (Bulmer, 1974; Santiago, 1998). One immediate consequence of the infinitesimal assumption is that the effect of selection on individual variants is negligible. Therefore, the genic variance can be assumed to be constant over short time scales, and in approximations, we expect a small and smooth decline. The infinitesimal model gives the limit of any distribution of effect sizes as we go to smaller and smaller values of these effects distributed over a larger and larger number of loci, keeping the overall trait value constant (Sachdeva and Barton, 2018; Barton et al., 2017).

Following Robertson (1977) and Sachdeva and Barton (2018) consider a trait determined by an effectively infinite number of loci, uniformly distributed on a genome block of a certain length. All loci are weakly selected, though there may be appreciable selection on the block as a whole. This model is parameterized by only one parameter V_0, namely the genic variance per unit map length. Under infinitesimal conditions, long term selection is possible, as standing genetic variance takes a long time to be depleted in this case.

If it is possible to reject the infinitesimal model with linkage, it is worth thinking about a hierarchy of models, bridging the gap between this model and one of, in the extreme case, only one major effect locus. One might compare the likelihood ratios of these different scenarios to find a better approximation to the actual genetic basis. In such in-between cases, one could, starting from the infinitesimal model with linkage, make the variance distribution along the genome more and more heterogeneous. This would show up as a higher variance between replicates which could be tested in simulations. At some point, there would be spikes of variance (from individual major effect size loci) with long stretches of genome contributing zero variance in-between, as one approaches the extreme of a single quantitative trait locus (QTL) scenario (this is unless one of the "smoother" models can not be rejected).

It should be noted that there is also family of methods which use linear mixed models (LMM) to attribute genetic variance to regions of genome and test hypotheses as to the heterogeneity of this distribution (Kang et al., 2010; Listgarten et al., 2012; Zhou and Stephens, 2012).

In chapter 2 of this thesis, I discuss an E&R selection experiment in mice and what we could learn about the genetic basis of the trait of tibia length by observing genomic changes. This was done in collaboration with Frank Chan and his group at the Friedrich-Miescher Institute in Tübingen, especially Layla Hiramatsu during her tenure as a Postdoc there. Chapter 3 discusses the problem of correlations between SNP associated with specific haplotypes and the increased variance in SNP-based summary statistics which follows. This chapter was inspired by the high variance between chromosomes observed in the Longshanks experiment of Chapter 2. Finally, Chapter 4 discusses an evolve and resequence experiment in Drosophila in which we investigated the genetic basis of pupal length and which had much higher replication than the mouse experiment. This was done in collaboration with Guy Reeves and Diethart Tautz at the MPI for Evolutionary Biology in Plön, Germany.

CHAPTER 2

Analysis of an Evolve and Resequence Experiment in Mice

The work presented in this chapter was done in collaboration with Nick Barton, with whom I worked on the statistical analysis and simulations, Frank Chan and his group, who did the genomic analysis, and Campbell Rolian and his group, who did the experiment, and is based on our paper in eLife (Castro et al., 2019).

2.1 Introduction

In this chapter, I discuss the Longshanks selection experiment. Its goal was to understand the genetic basis of tibia length, as an example of a complex trait in mice. The so-called "Longshanks" mice were selected for increased tibia length while aiming to keep body mass constant (Marchini et al., 2014). Originally, Marchini et al. (2014) investigated how selection could break the correlation between tibia length and body mass, such that an increase in tibia length over the first 14 generations of the experiment could be observed, while body mass was essentially held constant. This previous study focused on phenotypic change and inferred genetic correlations indirectly using the pedigree (Marchini et al., 2014). In contrast, our paper and this chapter, while using the same data set, have a different focus, namely to understand the underlying genetic changes. The Longshanks experiment is still ongoing, but the current study focuses on the first 17 generations.

The main questions discussed in this chapter are:

- To what extent are allele frequency changes influenced by selection, to what extent by random genetic drift?
- What can we learn about the genetic basis of complex traits on the example of tibia length - what portion of the variance contributed to the trait is due to large effect loci, small effect loci, and what portion cannot be distinguished from an infinitesimal null model?
- What type of selection signature is associated with this process?

- When parallel trait changes occur in replicates, to what extent do we also observe *genetic* parallelism, e.g. do these parallel phenotypic changes correspond to changes in the same genes?

2.1.1 Background

The understanding of how populations adapt genetically to changing environments is relevant in a number of fields, from animal breeding to conservation biology. While insight into adaptation from any organism is important, mammals are of particular interest, both because they are close to humans genetically and because they are particularly vulnerable to rapid changes in the environment, as might occur because of climate change, due to often small population sizes, and/or fragmented habitats. There is also a lot of general interest in genetic parallelism, especially with regard to the question of whether parallel phenotypes are the result of parallel genetic changes or different genomic solutions leading to the same result (Elmer and Meyer, 2011; Stern, 2013; Schluter et al., 2004; Ravinet et al., 2016; Deagle et al., 2012).

Beneficial alleles in mammals typically arise from standing variation, namely alleles which were present in the population before selection started. This means that their reservoir would be quickly depleted under strong selection if the genetic basis of the trait or traits under selection consisted of only few loci of large effects. However, it is known that for many traits in mammals and other organisms adapting from standing variation, the response to selection can be significant and consistently sustained over many generations (Keightley et al., 1996; Laurie et al., 2004; Hill and Kirkpatrick, 2010).

The question of how polygenic traits will respond to selection has typically belonged to the realm of quantitative genetics, which uses methods conceptually based on the infinitesimal model (Barton et al., 2017) and see 1.5 for a short explanation, to predict that response. As stated in the introduction, the infinitesimal model assumes a virtually infinite number of loci with infinitesimally small effects which contribute to a trait. It is a continuous model, though discrete approximation is possible, which we used in the simulations described in this chapter. The selection response under the infinitesimal model is smooth and predictable. Under this model, the breeder's equation (Lush, 1943) predicts the response. One can also estimate the total expected gain in the trait value, which is equal to $2N_e$ times the response in the initial generation (Robertson, 1960). There is impressive empirical evidence for this, see Fig.4 in Weber and Diggins (1990).

In general, the infinitesimal model has been very successful in predicting selection response across a wide range of conditions in selection experiments, and is the basis for commercial breeding (Walsh and Lynch, 2018). However, going beyond the phenotypic response to selection, many open questions remain as to the actual underlying genetic basis. After all, we know that as long as there is sufficient statistical power, it has to be possible to reject the infinitesimal model in favor of a more realistic model, since in reality genomes do not consists of infinitely many genes and neither can effect sizes be infinitesimally small.

In understanding the fate of adaptive variants under different conditions, the existing body of theory (Walsh and Lynch (2018, Ch. 5)) is far ahead of our understanding of the empirical data, especially when it comes to natural populations where many theoretical expectations are hard to test. In artificial selection experiments, there is a much better chance to get at these questions; here we can avoid missing data and control conditions completely, from population size to selection regime.

Allele Frequency-Based Statistic Δz^2

In what follows, we use a windows-based statistic to summarize information about the allele frequency changes from the first to the last generation under selection (F0 to F17). We use the Fisher-transformed allele frequency $z = 2\arcsin(\sqrt{p})$. Δz^2 is the square of the arcsine transformed allele frequency difference between F0 and F17. This has an expected variance of $1/2N_e$ per generation independently of starting frequency, and ranges from 0 to π^2. Δz^2 was averaged over 10 kbp windows.

2.2 The Experiment

2.2.1 Description of the Longshanks Selection Experiment

At the start of experiment, three base populations of 14 breeding pairs each were established (Marchini et al., 2014). This was done by sampling from a commercial mouse stock known as CD1; derived from mixed breeding of classic laboratory mice (Yalcin et al., 2010). Two replicate "Longshanks" lines (LS1 and LS2) were set up, as well as one control line (Ctrl) without selection. In the two Longshanks lines, the 14 breeding pairs each were established initially, and then selected in each subsequent generation, never allowing sibling crosses (later population size was kept to 16 pairs - it was planned to establish 16 pairs, but some mice got lost unfortunately). The next breeding population was chosen by picking the son and daughter with the longest tibia length relative to the cube root of body mass from each family. The number of chosen individuals corresponds to 15-20% of all offspring. In Ctrl the identical breeding scheme was applied, except that breeders were chosen at random (Marchini et al., 2014).

2.2.2 Description of the Data

The full pedigree was recorded during the experiment, phenotypes were measured for all breeding individuals and tissue samples were taken. Tibia length and body mass were measured as described in Marchini et al. (2014), namely by weighing the mice and radiographing them in a cabinet X-ray system, acquiring images with a digital X-ray scanner. A total of 1332 Ctrl, 3054 LS1 and 3101 LS2 mice were recorded. Five outlier individuals with abnormal skeletal development were removed from LS2 and excluded from further analysis. For simulations, missing data in LS2 were filled in with random individuals which best matched the pedigree. The initial (F0) and final (F17) breeding populations were sequenced at low coverage (2.91-fold coverage (range: 0.73–20.6×; n = 169 with <10% missing F0 individuals; Supplementary file 1 in Marchini et al. (2014), revealing on average 6.7 million (M) segregating SNP per line (this corresponds to approximately 0.025%, or 1 SNP per four kbp). The reference genome used was *Mus musculus* reference mm10, which is derived from GRCm38.

2.3 Results

In the two selected lines LS1 and LS2, a strong and significant response to selection in tibia length was observed, corresponding to 0.29 and 0.26 standard deviations (s.d.) per generation respectively (fig.2.1. Over 17 generations increases of 5.27 s.d. (LS1) and 4.81 s.d. (LS2) were observed. That corresponds to 12.7% in LS1 and 13.1% in LS2. There was a modest decrease in body mass, -1.5% in LS1 and -3.7% in LS2. In contrast, Ctrl showed no significant

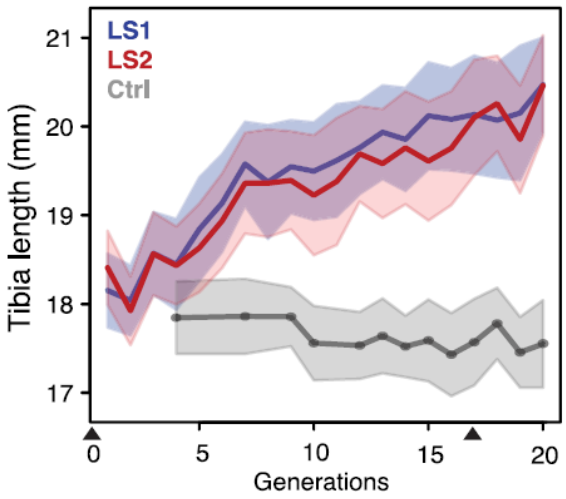

Figure 2.1: Tibia length increased by 5.27 s.d. overall in LS1, by 4.81 s.d. in LS2, and remaind roughly constant in Ctrl.

directional change in tibia length or body mass at significance level $p < 0.05$ (Student's t-test). This approximate 5 s.d. change in 17 generations is fast compared to typical rates observed in nature, see e.g. Hendry and Kinnison (1999) and Reznick et al. (1997), but in line with typical responses seen in selection experiments, see examples in Garland and Rose (2009).

2.3.1 Simulations and Statistical Analysis

In the LS experiment, we see a rapid and smooth increase in tibia length. As explained in section 2.1.1, this points to a highly polygenic architecture, with no individual locus explaining the bulk of the selection response. It also shows that selection was acting and effective. Therefore, the appropriate null hypothesis for the genetic response to selection in this case should capture "polygenic adaptation", rather than neutrality.

In this case, the rejection criterion is more subtle than "reject if selection is detected". Rather, the null hypothesis will be rejected if loci of larger effect sizes are detected. "Larger effect size" here means that they are distinguishable from the infinitesimal background i.e. their effect size exceeds a threshold set by what we expect to see under the infinitesimal model with linkage. Simulations were based on these considerations.

Simulation Setup

The simulations we developed incorporate the trait measurements, the correct selection regime and the full pedigree. We therefore expect it to produce an accurate expectation for the genetic selection response. The simulations follow one chromosome at a time - this is possible because of a special feature of the Longshanks breeding scheme: the highest-ranking male and the highest-ranking female from each family were chosen to breed with the highest ranking mice from other families within a line. Therefore, if we disregard non-Mendelian segregation and the fraction of failed litters ($\sim 15\%$), selection acted solely *within families* and on the measured traits. Such selection does not distort the pedigree and therefore allows us to follow the evolution of each chromosome separately (Falconer and Mackay, 1981, Ch. 13).

Each chromosome was represented by a set of "junctions", which arise as recombination breakpoints and create boundaries between pieces of genome inherited from different founders

genomes. These junctions are inherited like Mendelian loci and therefore easy to follow (Fisher, 1954). In principle, we should simulate selection under the infinitesimal model by following the contributions of all blocks of chromosome across the whole genome. However, this is computationally challenging and would make the simulation prohibitively slow, since the contributions of all blocks in every individual defined by every recombination event would have to be tracked.

Our solution is to follow a large number of biallelic loci and checking that their number is sufficiently large to approach the infinitesimal limit. Compare to Fig. 2.4 for confirmation that the bulk of the response coincides with the diffusion limit expectation. We made a further slight approximation by only explicitly modeling discrete loci on one chromosome at a time (possible because of within-family selection). After this process, SNP genotypes were recovered by seeding each genome block with the appropriate ancestral haplotype. This is much more computationally efficient than following all the neutral SNP markers individually. Crossovers were distributed according to the standard genetic map in Cox et al. (2009).

Trait values were modeled as a sum of three variance components: a component due to the infinitesimal background, called V_g and determined by a large number of unlinked loci on all *but* the focal chromosome, V_s, a component determined by the sum of effects of 10^4 evenly spaced loci *along* the focal chromosome, and a Gaussian non-genetic component V_e, which represents variation due to the environment. The first two, namely the genetic components, sum to the breeding value if each individual. The values of V_g are normally distributed among offspring around the mean of the parents breeding values, and its variance is:

$$V_M = (V_A/2)(1-\beta)(1-F_{ii}-F_{jj}) \tag{2.1}$$

(Falconer and Mackay, 1981, Ch. 5),where V_A is the initial (total) genetic variance, F_{ii}, F_{jj} are the probabilities of identity between distinct genes within each parent which are calculated from the pedigree, and β is the fraction of genome on the focal chromosome.

Component V_s was determined by 10000 loci, which contribute a fraction β of the initial additive variance. We initially choose to let these loci have equal effect sizes with random signs, $\pm\alpha$, such that the initial allele frequencies all are $p_0 = q_0 = \frac{1}{2}$, and $\beta \cdot V_{A,0} = 2\sum_{i=1}^{n} \alpha^2 p_{i,0} q_{i,0}$. The choice of equal effects approaches the infinitesimal most closely. Matching the experiment, the initial population consists of 28 diploid individuals, and alternative alleles at loci have initial frequencies of $\frac{1}{56}$, $\frac{4}{56}$, $\frac{12}{56}$ and $\frac{28}{56}$ in equal proportions. Inheritance is assumed to be autosomal with no sex-linkage.

The variance due to the two genetic components was proportional to the corresponding map lengths. Heritability was estimated from observed trait values. For the selection scheme in each generation, the actual (i.e. observed) number of male and female offspring were generated from each breeding pair (though with varying simulated genotypes) and the male and female with the largest trait value were chosen to become part of the breeding population in the next generation.

SNP genotypes were assigned to founder genomes with their observed frequencies. However, in order to reproduce the correct variances, founder *haplotypes* had to be assigned. This essentially means to to assign SNP to also be consistent with the individual diplotypes and then make an additional choice at all heterozygous sites to produce two haplotypes per diplotype. To do this in the absence of phased data, three procedures to produce such haplotypes were compared, see Fig. 2.2: assigning haplotypes in linkage equilibrium (consistent with given frequencies, but ignoring individual genotypes), using the genotypes, and assigning the two

alleles at heterozygous sites in each individual at random to its two haplotypes (this minimizes LD in a way consistent with observed diploid genotypes) and finally using the diploid genotypes and then assigning alleles at heterozygous sites in each individual to the "reference" and "alternative" haplotypes consistently within an individual, which maximizes LD consistently with the given genotypes. We call these three procedures "no LD", "min LD" and "max LD" respectively. Note that even in the "no LD" case there will technically always be some small amount of LD, because there are always some correlations between SNP due to chance.

Simulation Results

The decrease in genetic variance due to random genetic drift was measured by the inbreeding coefficient F, defined as the probability of identity by descent relative to the initial population We distinguish the identity between two distinct genes *within* a diploid individual, F_w (or F_{ii} in index notation as above), from the probability of identity between two genes in different individuals, F_b (F_{ij} above). The overall mean identity between two genes chosen independently and at random from all $2N$ genes is then:

$$\bar{F} = \frac{2\left(N-1\right)F_b + F_w + 1}{2N}. \tag{2.2}$$

The proportion of heterozygotes decreases by a factor of $1 - F_w$, and the variance in allele frequency increases with $\bar{F}$. Then, the expected genetic diversity, $\mathbb{E}\left[2pq\right]$ decreases as $1 - \bar{F}$.

Figure 2.7 shows that in the absence of selection, the identity F_b (light shaded lines) increases slower than expected under the Wright-Fisher model (black lines). These differences are a consequence of the circular mating scheme, which was designed to slow the loss of heterozygosity. The dotted lines show the average F, estimated from the loss of heterozygosity in 50 replicate neutral simulations, each with 10^4 loci on a chromosome of map length of $R = 1$ Morgan. This is close to the prediction from the pedigree (light shaded lines), which validates the simulations. The thick-colored lines show F estimated in the same way, but with $V_s/V_e = 0.584$, which was calculated from the observed selection response from parent-offspring regression. This works because we know that the top male and female were selected from each family, and the intensity of selection that this causes is determined by the genetic variance segregating, relative to the environmental variance. Thus, given V_s/V_e, we know the strength of selection. In simulations, the actual numbers of males and females in each family were used. V_s/V_e was estimated to fit the response to selection observed in the experiment.

The rate of drift, as measured by F over time, is significantly faster under selection: by 6.7% in LS1 and 9.8% in LS2 (Student's t-test $P \leq 0.008$ in LS1 and $P \leq 0.0005$ in LS2). However, this is not apparent from individual replicates, since the standard deviation of the rate of drift, relative to the mean rate, is $\sim 13\%$ between replicates. Interestingly, while the observed loss of heterozygosity in the actual data fits that expected from the pedigree closely (large dot with error bars in 2.7), there is extremely wide variation among chromosomes for these data (filled dots), substantially higher than seen in any simulation seeded with SNP at linkage equilibrium (open dots, slightly to the right), another hint at the importance of LD to estimate variation.

Since the breeding scheme was not set up only to increase tibia length, but also to keep body mass constant, the simulation needed to capture both of these factors. This was done by mapping fitness onto tibia length $\boldsymbol{T}$ and body mass $\boldsymbol{B}$ as a single composite trait $\boldsymbol{ln(TB^{\phi})}$. The scaling ϕ was estimated from the actual data as -0.57, chosen to closely match the actual ranking used to select breeders in the experiment. In the experiment, the ranking was established by combining a ranking based on absolute tibia length with a ranking based on

relative tibia length, namely tibia length divided by the third root of the body mass (Marchini et al., 2014). For simulations ϕ was chosen such that the actual breeders would have the lowest ranks overall.

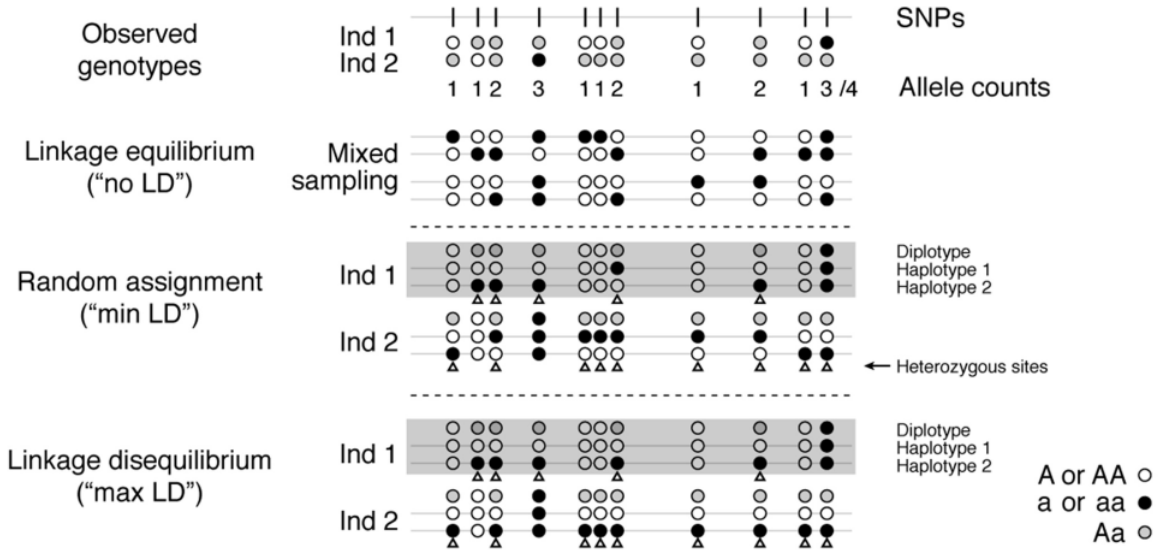

Figure 2.2: Assigning SNP to founder haplotypes: In the absence of phased data, simulations were seeded with three different types of founder haplotypes, all consistent with the allele frequency data. Black and white dots indicate either of the two homozygous states and grey dots indicate the heterozygous state. Founder haplotypes were sampled directly from founder individuals in each of the experimental LS lines in the following three ways: 1. The linkage equilibrium scheme: Here, we sampled from the list of allele counts at all SNP. This produces founder haplotypes that carry essentially no LD (however, some correlation is always created by chance, even under this scheme). 2. The "min LD" scheme: Here, founder haplotypes are kept consistent with individual genotype data, and only in the case of heterozygous loci, the two alleles are assigned at random to the two haplotypes. This produces founder haplotypes with minimal LD ("min LD") *consistent* with the observed genotypes. 3. The maximum LD ("max LD") scheme: Here, founder haplotypes are also sampled consistently with the individual genotype data, but more than that, haplotypes 1 and 2 are also consistently assigned with the reference and alternative alleles respectively (so once the alternative allele is assigned to haplotype 1 at the first SNP, we will keep assigning it to haplotype 1 for all other heterozygous sites in this individual as well). This maximizes LD in the founder haplotypes, again consistent with individual genotypes.

We tested a range of models with varying selection intensity and initial LD, running 100 simulated replicates for each one, to determine the significance of allele frequency changes. Since there was much more variation in chromosomes than expected from simulations seeded with the "no LD" scheme, clearly initial LD plays an important role. LD between ancestral SNP greatly increases random variation, (see Chapter 3). Simulations were seeded with SNP drawn according to the "max LD" scheme, which fit variability between windows observed in the actual data best. Significance threshold tests were also based on this scheme (Fig.2.3). This is also the more conservative choice, since more LD leads to more variability between replicates and higher significance thresholds.

Figure 2.3 shows the distributions for $z = 2\arcsin(\sqrt{p})$ for 100 replicates each for scenarios with and without selection and the different ways of setting up initial LD. The two columns

correspond to replicates LS1 and LS2. The three rows correspond to the three ways of assigning SNP to the starting set of haplotypes described in figure 2.2: "no LD", "minimal LD" or "maximal LD". The "no LD" scheme is only consistent with the given allele frequencies (first row), while the other two processes are consistent with individual genotypes as well. The blue and red bars correspond to results from simulations done under neutrality or selection respectively (with selection strength consistent with the response observed in the experiment). Significance thresholds in black mark the values above which changes in the scaled allele frequency statistic Δz^2 are unlikely to have been produced under neutrality at $p = 0.05$ significance level. While increased selection pressure leads to greater shifts in allele frequencies, the figure shows that the impacts on Δz^2 due to *initial LD* are much larger.

A main conclusion from the simulations is that overall allele frequencies were hardly perturbed by varying selection intensity, apart from a few significant loci. Even under infinitesimal selection however, a very weak but detectable signal can be generated. It can show up as a (small) excess of alleles which swept from low to high frequency compared to strict neutrality (see Fig. 2.4, specifically SNP classes $\frac{1}{56}$ and $\frac{4}{56}$). However, it would take many replicates for such an excess to become statistically significant. These results echo other E&R experiments based on diverse base populations which also showed only weak evidence of selective sweeps at individual loci, e.g. Burke et al. (2010); Orozco-terWengel et al. (2012).

Simulations assume infinitesimal effects of loci, so allele frequency shifts exceeding the stringent "max LD" threshold (see Figure 2.3) would suggest the presence of discrete loci contributing significantly to the selection response. An excess of such loci, particularly if present in both replicates, would thus imply a mixed genetic basis of few distinguishable large-effect loci on an infinitesimal background.

2.3.2 Genomic Signatures of Selection

This section summarizes the main genomic results of the analysis. I am brief in some places, especially when summarizing the functional genomics part, as this topic is not the focus of this thesis.

Sequencing and Genome-Wide Diversity

To detect genomic changes in the LS experiment, all individuals of the initial (F0) and last generations (F17) were sequenced to an average 2.91-fold coverage (range: 0.73–20.6×; n = 169 with <10% missing F0 individuals). For each line and generation, all individuals were barcoded and pooled for sequencing. Since the CD-1 mice were founded by an original import of 7 inbred female mice and two inbred males, a maximum of 18 segregating haplotypes at any given locus was expected (see Figure 1 in Yalcin et al. (2010)) Across the three lines in our setup, similar levels of diversity were found, with an average of 6.7 million SNP segregating. This corresponds to about 1 SNP per four kbp. Globally, diversity decreased by 13%, while F17 populations still retained ~ 5.8 million segregating SNP. This drop is sufficiently explained by drift alone. The simulations confirmed this and showed that selection contributed negligibly to the drop in diversity 2.7.

There was negligible population differentiation between the three founder populations with an across-line F_{ST} on the order of $1 \cdot 10^{-4}$.This increased to 0.18 in generation F17, which is consistent with random sampling from an outbred breeding stock. Overall, despite strong selection in the experimental lines, there was little perturbation of genome-wide diversity; apparently changes in global diversity have little power to distinguish selection from drift. This

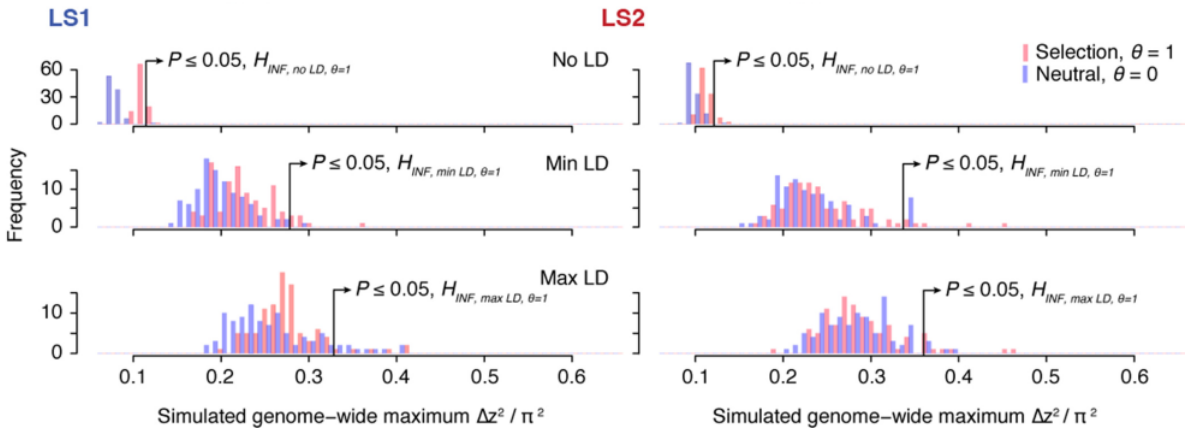

Figure 2.3: The significance threshold for Δz^2 depends on the initial LD. We use the square since selection might lead to allele frequency shifts in a positive *or* negative direction, and we do not know how individual genes functionally affect the trait. The figure shows significance threshold values under varying LD in the starting population from 100 simulated replicates. blue: no selection, red: selection according to the observed selection response in the actual experiment. The three rows on either side correspond to the three allele-assignment methods discussed in figure 2.2. The two columns correspond to replicates LS1 and LS2. Increasing selection pressure produces greater shifts in Δz^2 even on the *same* pedigree, due to a relatively greater proportion of additive genetic variance V_s on the overall variance in the trait. One can imagine that different blocks are passed down the same pedigree, through the same individuals - blocks which might carry varying numbers of alleles affecting the trait. However, a *far* greater shift in Δz^2 can be observed due to differences in initial LD. This is because even weak associations between large numbers of SNP in a window can add up greatly to inflate the variance of Δz^2 (compare to chapter 3). Subsequently, of the three initial LD levels used in these simulations, "max LD" should produce overly conservative thresholds, and "min LD" should produces overly permissive thresholds, which can lead to more false positives. For no initial LD, thresholds might substantially underestimate the correct significance threshold. In our analysis, the "maximal LD" thresholds were used.

is consistent with the simulation results and notable as there was a strong *phenotypic* selection response.

Significance Thresholds

In order to obtain the significance thresholds used in the analysis, we simulated replicates of the LS1 and LS2 lines, and calculated Δz^2 shifts averaged over 10 kB windows. The maximum genome-wide Δz^2 value was then found for each replicate. This was done for a range of selection intensities and the 3 different LD models ("no LD", "min LD" and "max LD", explained in Fig. 2.2). Under each selection and LD model, we obtained the critical Δz^2 value from the distribution of genome-wide maximum Δz^2, corresponding to $p = 0.05$, the 95th quantile. This procedure controls for hitchhiking due to linkage, line specific pedigree, and selection strength.

Sequencing

Genomic DNA was extracted from ear clips. Each sample was individually bar-coded and pooled for high-throughput sequencing. Sequenced data were pre-processed using a pipeline

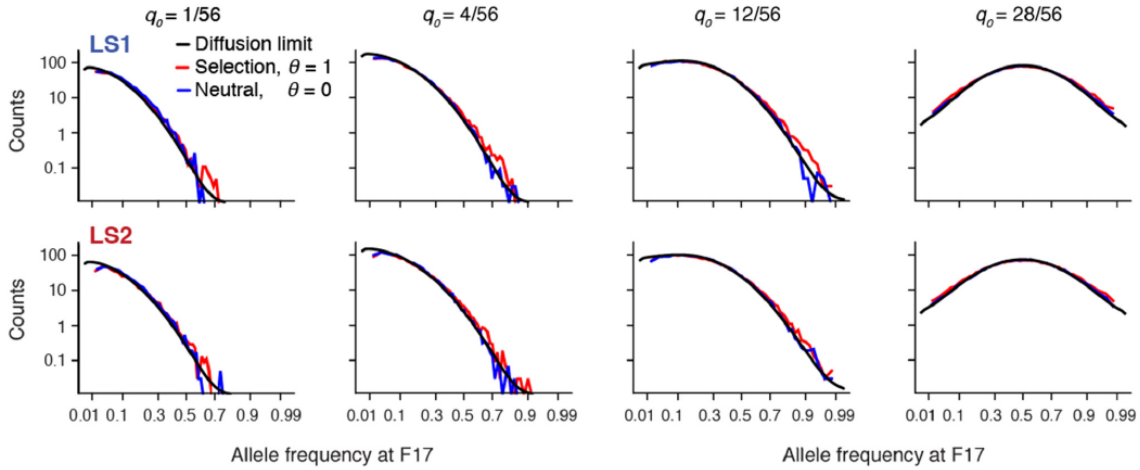

Figure 2.4: The simulated distribution of minor allele frequencies q_0 at generation 17 with no selection (blue line) is compared to the simulated distribution with selection (red line) and the theoretical expectation under the diffusion limit. These distributions are conditional on an initial minor allele frequency of 1, 4, 12 or 28 copies out of 56 (columns left to right). The black line shows the diffusion limit, calculated for scaled time $\frac{17}{2N_e}$, since we look at the change over 17 generations. N_e was estimated from the rate of increase in the inbreeding coefficient F, calculated from the pedigree, giving values of $N_e = 51.7$ for LS1 and 48 for LS2. Under a within-family selection scheme we would expect N_e to be higher than the census size, up to twice as high as the value obtained under random mating. We observe a very slight excess of alleles going from low to high copy numbers under selection, with substantial overlap between the distributions.

consisting of data clean-up, mapping, base-calling and analysis.

There was a combined $\sim 100\times$ coverage which we consider sufficient to recover any of the (maximum) 18 CD-1 founding haplotypes still segregating at a given locus. The raw genotypes were phased with Beagle v4.1, but phasing remained unreliable (Browning and Browning, 2016). For more details on the methods used, see our paper (Castro et al., 2019).

Windows with Significant Allele Frequency Changes

We asked whether specific loci reveal more definitive differences between the experimental lines and the Ctrl, such that they are clearly distinguishable from the infinitesimal background. To do this, we identified windows with significantly elevated Δz^2 values.

169 windows were found which had significant shifts in allele frequency in LS1 and/or LS2 at $p < 0.05$ under the infinitesimal model simulation with initial SNP assigned under the "max LD" scheme (see section 2.3.1 and figure 2.2 for details). They belonged to 8 clusters in LS1 and/or LS2. This corresponds to $\Delta z^2 \geq 0.33\pi^2$ compared to a genome-wide background of $\Delta z^2 = 0.02 \pm 0.03\pi^2$. In Ctrl, 8 significant windows, belonging to three clusters were identified, compared to a genome-wide background of $\Delta z^2 = 0.01 \pm 0.02\pi^2$.

The eight clusters in the selected lines overlapped with 2-179 genes each. Together they contained 11 candidate genes with known roles in bone, cartilage and/or limb development. Some genes with "short-tibia" knock-out phenotypes were found. Gene-regulation likely played an important role in the selection response.

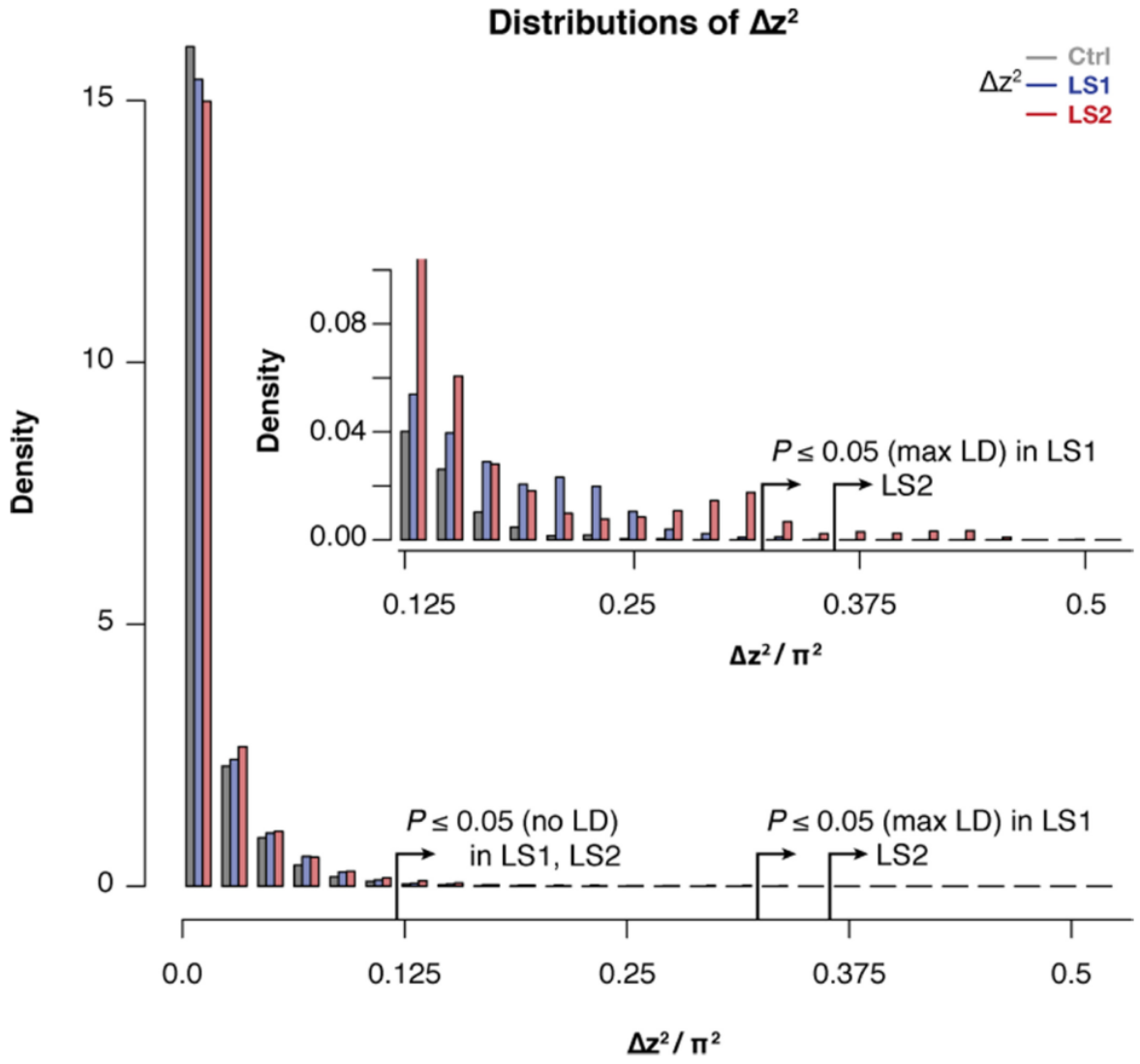

Figure 2.5: The histograms show the distributions of Δz^2 in 10 kbp windows across the genome for the Ctrl (gray), LS1 (blue) and LS2 (red). While the distributions are quite similar, the two selected lines showed more extreme values than Ctrl, with larger values of Δz^2 overrepresented. The inset zooms in on the region above $\Delta z^2/\pi^2 = 0.125$. Two different significance thresholds are marked, both for $p \leq 0.05$. These thresholds were obtained from simulations which reproduced the experiment faithfully as much as possible, and used three different ways to represent initial LD. These are explained in Fig. 2.2. More details on the simulations follow in section 2.3.1. Under the "max LD" scheme, maximal LD between SNP in the founder generation is assumed, whereas under the "no LD" scheme, there is (almost) not initial LD. The "no LD" scheme produces a much more lenient significance threshold. Above the more lenient of these thresholds, the pattern of higher Δz^2 values from the selected lines as opposed to Ctrl is clear; above the more stringent threshold, specifically shifts in Δz^2 from LS2 are overrepresented. Greater distortion in the Δz^2 spectra is expected if discrete loci contribute to the selection response. Accordingly, we read this as evidence for discrete loci contributing in LS2.

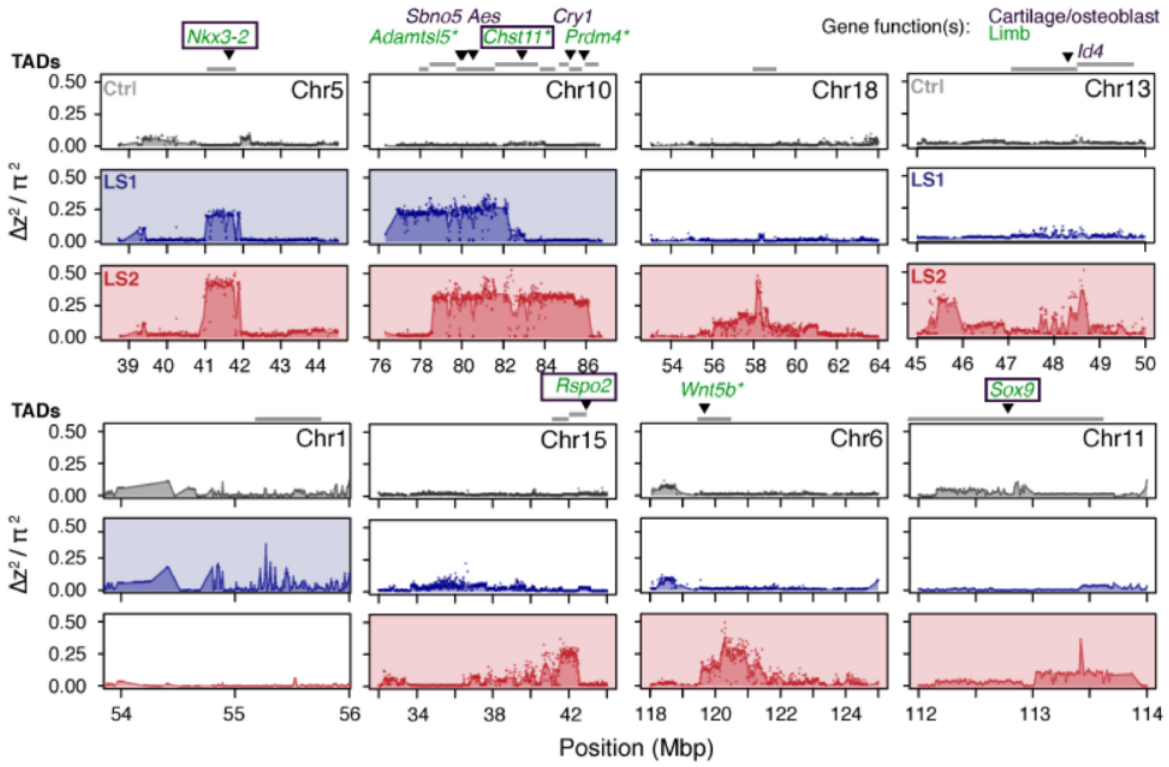

Figure 2.6: For each cluster, Δz^2 profiles for Ctrl (gray, LS1 (blue) and LS2 (red) are shown. Plots are shaded if the cluster is significant in the given line. None of the three significant clusters in Ctrl overlapped with these 8 clusters.

Figure 2.5 shows the genome-wide distributions of Δz^2 in the three lines. The distributions are quite similar, but an excess of Δz^2 in the higher bins above significance thresholds set by the infinitesimal model with linkage, especially for LS2, suggests that there were some bigger effect loci playing an important role in the experimental lines. Overall, these results are consistent with a very polygenic genetic basis of the trait, with some loci of larger effects contributing.

Parallel Selection Response

To quantify genetic parallelism, we compared genetic responses in the two replicates LS1 and LS2. In mice, as in other organisms with relatively long generation times, most adaptation is from standing genetic variation. Therefore, if the starting populations shared beneficial alleles, these may sweep in both replicates, as long as selection is strong enough to overcome drift and favorable alleles are initially frequent enough not to be lost. Therefore, observing parallel changes is strong evidence for selection.

Figure 2.8 shows Δz^2 profiles across the genome. Looking at Ctrl (in gray) it is interesting that some peaks appeared here as well. This confirms that drift, inbreeding and linkage together can generate large allele frequency changes, even with no selection involved. This is in accordance with similar observations in simulations and shows how important it is to consider LD. However, LS1 (blue) and LS2 (red) (Fig. 2.8) showed more and stronger parallel shifts than Ctrl. Δz^2 profiles of the two selected lines were generally more similar than between each line and Ctrl, with Pearson's correlation in Δz^2 for 10 kbp windows: LS1-LS2: 0.21, LS1-Ctrl: 0.06 and LS2-Ctrl: 0.05. More specifically, the handful of peaks coinciding

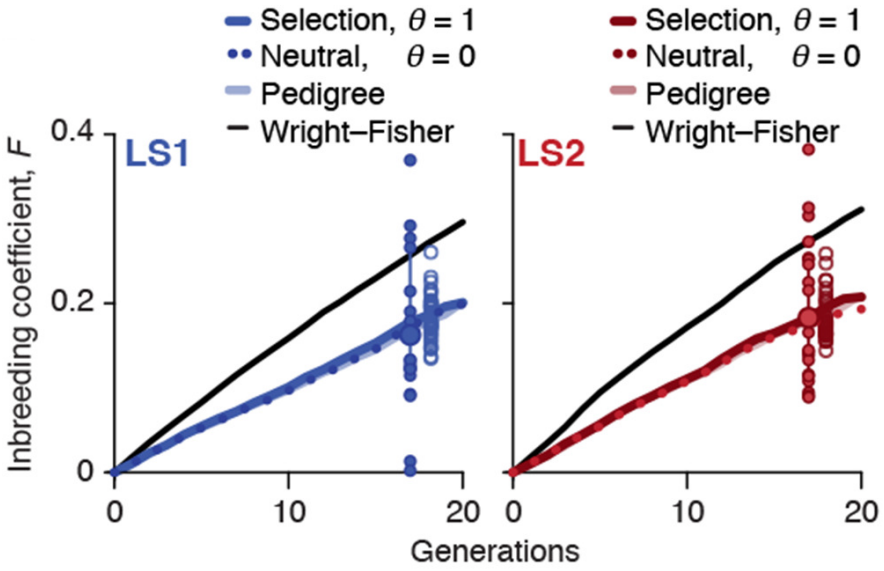

Figure 2.7: This plot shows the increase in inbreeding over the course of the Longshanks experiment. The lines follow the change in F_b ("F between"), the probability of identity between two alleles in two different diploid individuals, over 20 generations. The light lines follow F_b as calculated from the pedigree, the dotted lines follow F_b calculated from the average value over 50 neutral simulations, and the thick dark line follows F_b as calculated from the average of 50 simulations with selection response consistent with selection intensity at $V_s/V_e = 0.584$, where V_s is the segregating variance within families and V_e is the environmental variance. The F_b trajectories based on the pedigree and on neutral simulations are basically indistinguishable, which is a good check for simulations. Under selection, inbreeding increases slightly faster. The black line shows the increase in F_b, as expected under a Wright-Fisher model with actual population sizes. Under this model, F_b is close to F_w, which is the probability of identity between two allelic copies at the same locus within diploid individuals. Under this model, they are also both close to $1 - \left(1 - \frac{1}{2N_e}\right)^t$ over time, where N_e is calculated as the harmonic mean, 24.8. The large dots (with error bars) show the actual F_b calculated from decline in average $2p(1-p)$ in the data over 17 generations, with the error bars showing the interquartile range among chromosomes. The small dots show the estimates from each of the 20 chromosomes. The open dots show F_b from 40 replicate simulations, conditional on the same pedigree and the same selection response and taking the actual map length of each mouse chromosome into consideration (Cox et al., 2009). The simulations agree well with the observed genome-wide average, while the actual variance between chromosomes is much larger even than the one found from chromosome-wise simulation. LD is likely the source of this excess variance.

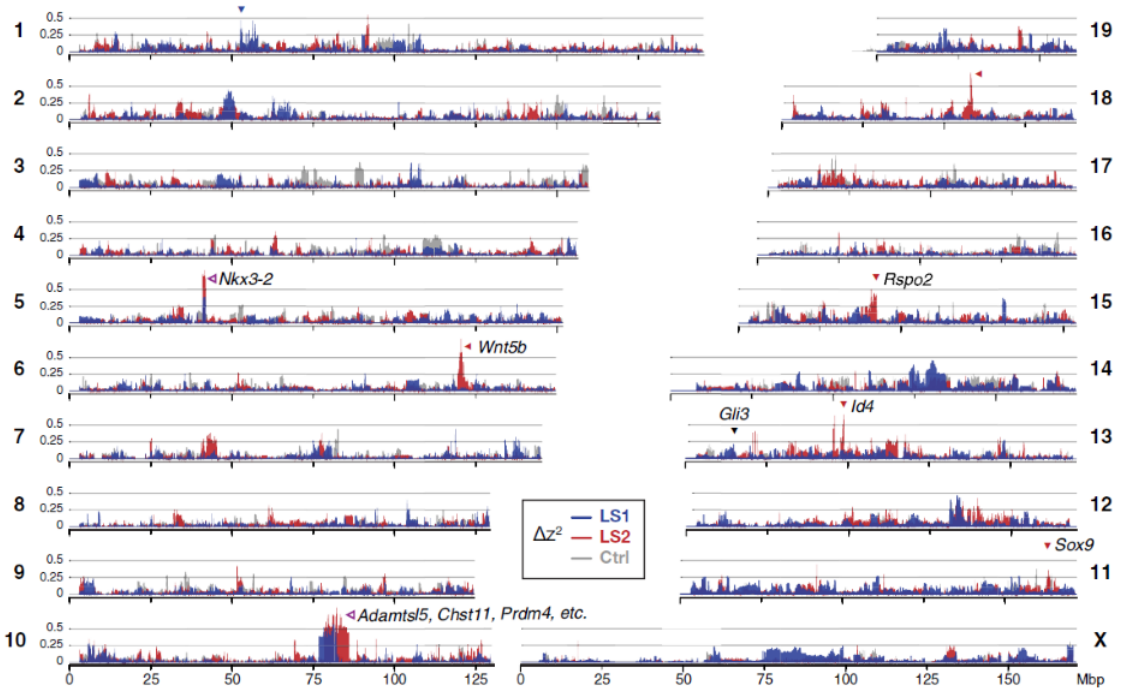

Figure 2.8: Allele frequency shifts between initial (F0) and final (F17) generation in LS1 (blue), LS2 (red) and Ctrl (gray) are shown as Δz^2 profiles along the genome. Units on the y-axis are fraction of the full range of Δz^2, from 0 to π^2. Some large shifts in the Ctrl are due to chance and most likely caused by LD in addition to the relatively small population size and limited diversity. Still, LS1 and LS2 show a handful of parallel selective sweeps on top of the infinitesimal background. Candidate genes are highlighted in the plot.

in LS1 and LS2 appear to correspond to parallel selective sweeps, which stand out from the infinitesimal background.

In contrast to previous selection experiments with replicates, which focused mainly on detecting parallel *loci* (Burke et al., 2010; Jones et al., 2012; Chan et al., 2012; Kelly and Hughes, 2018), the LS experiment allowed us to quantify parallelism more broadly, in addition to determining the selection coefficient of the loci with the strongest allele frequency shifts. The significance threshold used in most analyses was calculated under the infinitesimal model with "max LD".

Six of the eight significant loci were line-specific, despite the fact that all eight selected alleles were present in both lines in the founder generation F0. The two remaining loci were parallel in LS1 and LS2 and ranked highest with an estimated selection coefficient of $s > 0.2$. The estimated selection coefficient should be taken with a grain of salt, as it may be a substantial overestimate, because significant values are expected to overestimate s.

We should ask, why were only two out of the eight discrete loci with significant shifts found to have parallel Δz^2 signals? This seems low, especially given that LS1 and LS2 are genetically very similar in the founding generation and the fact that identical, and strong, selection was applied to them. Since the infinitesimal model with linkage recreates the bulk of the selection response accurately (see Figures 2.9, 2.7), we attempt an infinitesimal explanation: there are simply very many ways to increase tibia length under the infinitesimal model, or a genetic model close to it, with subtle shifts in many gene frequencies producing a strong phenotypic

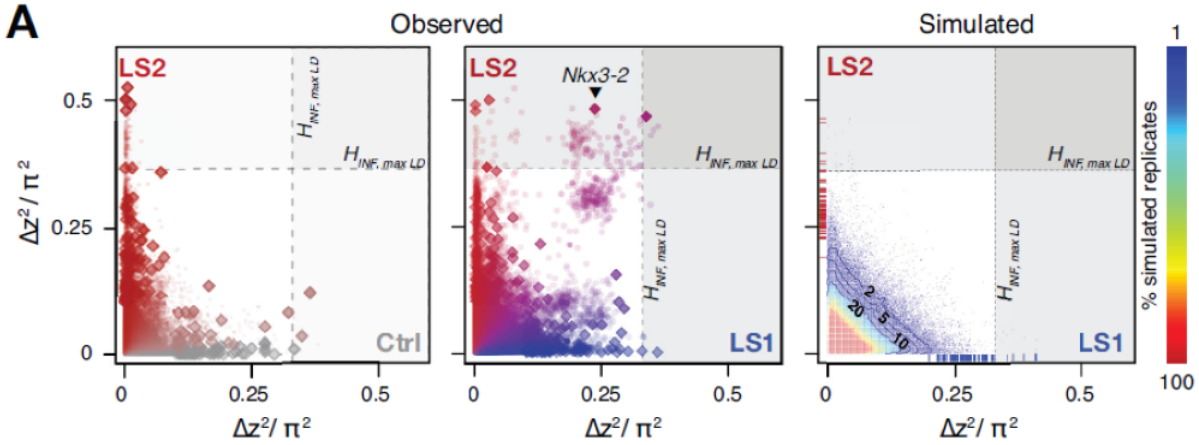

Figure 2.9: Here we show the genomic response compared between selected lines and Ctrl, as well as simulations. The overall genomic response represented in Δz^2 profiles matches the simulated "max LD" results closely, as can be seen comparing the right panel (from simulations) to the middle and left panels (observed data). Right panel: the heatmap summarizes percentages as seen in 100 simulated replicates. Most windows show little to no shift, in concordance with the data. The colored tick marks along the axes show the genome-wide maximum Δz^2 shifts in each of the 100 replicate simulations in LS1 (x-axis, blue) and LS2 (y-axis, red). Line-specific significance thresholds at $p \leq 0.05$ are indicated by dashed gray lines. While the simulations matched the bulk of observed data well, *no* simulation recovered the strong parallel shifts observed in LS1 and LS2 data (compare middle panel (data) to right panel (simulation)). Despite a largely line-specific genomic response in the experiment, the *largest* signals did occur in parallel in LS1 and LS2. The left panel shows, that shifts in LS2 (red) were greater than in Ctrl (gray) and changes in these two lines were not correlated (Pearson's correlation coefficient of 0.05). In contrast, there where many parallel changes as seen in the joint distribution of LS1-LS2 (mid-panel). Note that adjacent windows cluster due to hitchhiking. The strongest parallel shift could also be associated with functionality. It occurred in a gene called Nkx3-2, marked with an arrow in the mid-panel (more on this in the main text and 2.12). Significance thresholds as calculated under the "max LD" scheme are indicated.

response. In addition, the effect of drift is expected to be strong in a small population like the current one, obscuring signatures of selection. In larger populations, selection is be expected to be more effective and able to pick out increasingly small effect size loci, which are also much more likely to experience parallel frequency shifts.

To summarize, we observed only subtle differences between the line's changes in global diversity between F0 and F17, as shown in Figures 2.8 and 2.5. Using a standard χ^2 test, we found that parallelism is much more prevalent in the comparison between selected replicates than in the comparison of each selected line and the control: χ^2 test, LS1-LS2: $p \leq 1 \cdot 10^{-10}$, LS1-Ctrl: $p > 0.01$ and LS2-Ctrl: $p > 0.2$. The comparisons between experimental lines and Ctrl were both found to be non-significant after correcting for multiple testing, as were comparisons between simulated replicates, see Fig. 2.11. Since the loci selected in parallel in LS1 and LS2 also have the highest estimated selection coefficients, and parallelism is not generally expected from infinitesimal simulations, these loci provide the strongest evidence yet for the role of discrete major loci in this study, and also to reject the infinitesimal model. Also, in line with theoretical expectations, even though we did not observe widespread parallelism, these results

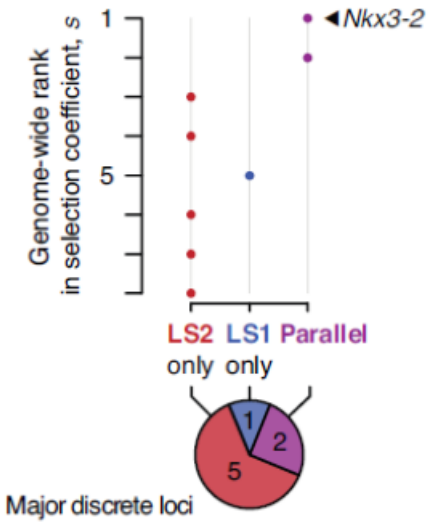

Figure 2.10: This shows the genome-wide ranking of candidate loci at $p \leq 0.05$ significance threshold. The ranking is derived from infinitesimal "max LD" simulations. Ranking is based on estimated selection coefficients *s* (see example in 2.3.2). Six out of the eight loci (clusters of windows containing candidate genes) showed significant shifts in only one line, LS1 or LS2, but the two loci with the highest estimated selection coefficients shifted in parallel in both LS1 and LS2, providing strong evidence for parallel selection in both lines. See also the middle panel in figure 2.9.

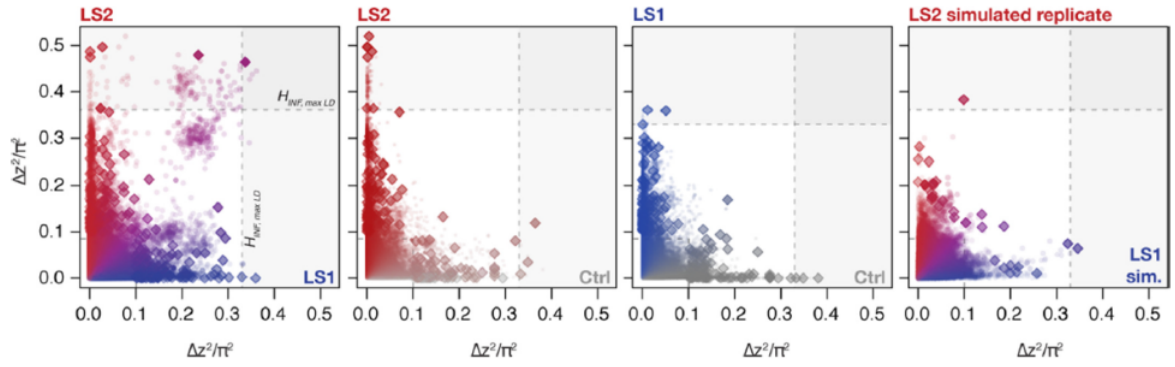

Figure 2.11: The joint Δz^2 distributions show that comparing Δz^2 profiles overall, they were very similar, with a slight skew toward higher values in replicate LS2. The joint LS1-LS2 distribution however, shows that many windows shifted in parallel in both lines (left panel, parallel-shifted windows in purple). This is in contrast to the joint distributions of experimental and control lines: LS1-Ctrl (second panel) and LS2-Ctrl (third panel), which show very few parallel shifts. The right panel shows the joint distribution of two individual replicate simulations using the LS1 and LS2 pedigrees respectively. These replicates were chosen for having among the greatest extent of parallel Δz^2 windows among simulated replicates, showing how even so, there is much less parallelism than in the data. Infinitesimal simulations were done with selection pressure corresponding to $V_s/V_e = 0.58$, see 2.3.1 under "max LD". Significance thresholds under the "max LD" scheme and "min LD" scheme are indicated as dashed lines, and the excess in parallel loci seen in the data is significant under both thresholds.

support the idea that the probability of parallelism can be high among loci with the greatest selective advantage (Orr, 2005).

Nkx3-2, The top-ranking locus

This section deals with the top-ranked locus in terms of estimated selection coefficient, Nkx3-2, which was selected in parallel in LS1 and LS2 and molecularly dissected by our collaborators.

The cluster of significant windows on chromosome 5 (see Fig. 2.6) contains three genes, including Nkx3-2, which is a known regulator of bone maturation (Provot et al., 2006). Nkx3-2 is a broadly expressed pleiotropic transcription factor which is lethal when knocked out. At this locus, the pattern of variation resembled a selective sweep spanning 1 Mbp, see Fig. 2.6. Through *in situ* hybridization, we showed robust expression of Nkx3-2 in the developing fore-and hind limb buds of Ctrl, LS1 and LS2. We hypothesized that at the Nkx3-2 locus the F17 allele causes de-repression of bone and/or cartilage formation by reducing enhancer activity and Nkx3-2 expression. The dissection yielded identification of up to six candidate quantitative trait nucleotides (QTNs), which provides a rare example of genetic dissection of a trait to the base-pair level in mice.

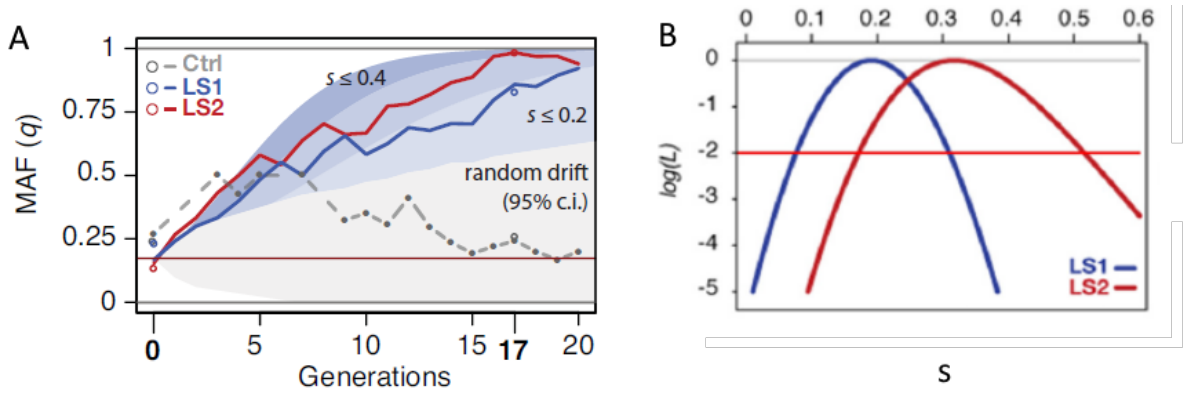

Figure 2.12: Selection at Nkx3-2: (A) Allele frequency of a selected allele (minor allele at F0, named q) within Nkx3-2 over 20 generations in LS1 (blue), LS2 (red) and Ctrl (gray, broken line). Observed frequencies from genotyped generations in Ctrl are marked with filled circles. Dashed lines indicate missing Ctrl generations. Open circles in generations F0 and F17 correspond to allele frequencies from whole genome sequencing. The allele fluctuated in Ctrl but followed a linear upwards trend in the selected lines, from around 0.17 starting frequency in both selected lines to 0.85 in LS1 and 0.98 in LS2 by generation 17. Shaded contours mark the expected allelic trajectories under varying selection coefficients starting from 0.17 (red horizontal line). The light gray shaded region marks the 95% confidence interval under random drift. (B) Log-likelihood curves of the selection coefficient, s, for LS1 (blue curve) and LS2 (red curve) respectively, based on the transition probabilities from a Wright-Fisher process with appropriate N_e (see main text). The horizontal red line marks a loss of 2 units in log-likelihood, setting conventional 2-unit support limits. Note that genome-wide analysis was done on the first 17 generations, but the trajectory included genotypes from this particular locus from the first 20 generations.

Selection Coefficient of Nkx3-2

To estimate a selection coefficient for Nkx3-2, the allele frequency trajectory of an allele at Nkx3-2 was produced by genotyping 1569 mice at the locus in question in all 20 currently

available generations. The trajectory shows the selected allele's steady increase across 20 generations in LS1 and LS2. The starting frequency was 0.17 in both experimental lines and rose to 0.85 in LS1 and 0.98 in LS2. In Ctrl it fluctuated around 0.25. Assuming two alternative alleles at any locus, the selection coefficient is implied by the observed parallel changes in allele frequency. We also need to set bounds on the estimate by accounting for random genetic drift.

Under constant selection, $\log \frac{q}{p}$ changes linearly with time, at a rate equal to the selection coefficient s. Therefore, a naive estimate of the selection coefficient is given by $\hat{s} = \frac{1}{T} \log \left[\frac{q_{17}}{p_{17}} \frac{p_0}{q_0} \right]$ (Haldane, 1932). This is not the MLE, but a simpler estimate only valid under constant selection. This gives, $\hat{s} = 0.19$ in LS1 and $\hat{s} = 0.32$ in LS2, giving an average of $\hat{s} = 0.26$ for the focal allele in Nkx3-2.

Constant selection is a reasonable assumption in our case: The strength of selection on an additive allele in the Longshanks experiment will depend on the fraction selected and within-family variance. The former is kept constant as much as possible and there is little loss of variance due to drift (F at generation F17 is about 0.17). Therefore, we can assume s to be constant, as long as there is no strong dominance.

We now have to account for random genetic drift to set bounds on this estimate: We obtain N_e, which is needed to find the dimensions of a transition matrix describing the process, by calculating the predicted loss of diversity over 17 generations from the pedigree. This gives $N_e = 44.9$ and $N_e = 44.4$ for LS1 and LS2 respectively. We round this to the next integer and calculate the transition matrix for the Wright-Fisher process with $2N$ rounded to 90 and 89 copies in LS1 and LS2 respectively. This transition matrix calculation gives the likelihood to see the observed allele frequency changes from $16/90$ to $75/90$ in LS1, and from $14/89$ to $87/89$ in LS2, given N_e and varying the selection coefficient. Overall, this gives $\hat{s} = 0.24$ in both lines with no significant loss of likelihood. We also investigated the effects of linked selection, but it did not appreciably change the distribution of allele frequencies at the focal locus.

Via this analysis the locus was found to be responsible for $\sim 9.4\%$ of the total selection response (2-unit support limits $3.6 - 15.5\%$). There might be some inflation due to multiple testing, which is hard to estimate, but this will not introduce estimation bias if the effect was large enough that it would certainly be detected in this study. The estimate of selection made here should be regarded as an effective value which may reflect a more complex reality, as we have not excluded more complex models, for example including more than two alleles at a locus.

We also estimated the contribution of Nkx3-2 using an animal model, namely:

$$V_p = \text{fixed effects} + V_A + V_R, \tag{2.3}$$

where the fixed effects were sex, generation, litter size, genotype at the locus and the replicate line, V_A is the additive genetic variance and V_R is the residual variance. This gave a small but significant effect of the genotype on the composite trait, with mean effect $= 0.36\%$ per additional copy of the selected allele in generation F17, and 95% CI: $0.069 - 0.64\%$ at $p = 0.017$. This corresponds to $\sim 1\%$ of variance in tibia length at generation F01. The observed increase in allele frequency from ~ 0.18 to 0.91, averaged over the two lines, implies that it accounts for about $\sim 4\%$ of the total selection response (12.9% increase in tibia length). This estimate is lower, but within the bounds estimated from simulation (2-unit support limits gave an interval of 3.6-15.5%). This estimate, unlike the previous one, does

control for ascertainment bias. However, the exact effect of the allele is difficult to pinpoint in any given generation or population due to the nature of the composite trait and its change in variance over time. Either way, both estimates support that Nkx3-2 contributes substantially to the selection response.

2.4 Discussion

Here, we looked at a replicated artificial selection experiment with small population size over 17 generations and characterized the observed genomic changes as selection acted on a combined trait of tibia length in relation to body size.

It is notable that the selection response was steady and robust even though we expect drift to play a major role in such a small population. Even with just 14-16 breeding pairs per generation, tibia length increased readily and in both experimental replicates in response to selection, while body mass stayed on a comparable level, and there was no such increase observed in the Ctrl. There was apparently enough standing variation present in F0 to fuel this increase in the trait. Also, under the infinitesimal model we expect to reach the limit for selection response at $\sim 2N_e$ generations (Robertson, 1960), which would correspond to around 90 generations here, which is of course far from the 17 generation duration of this experiment. Other artificial selection experiments on mice using a similar base population encountered selection limits already after 20-25 generations - potentially due to countervailing selection rather than loss of genetic variance (Careau et al., 2013, a study on high voluntary wheel running behavior). Here all evidence suggests that the Longshanks mice would continue to show increases in tibia length for many more generations.

N_e in the Longshanks experiment was estimated to be 46, which is larger than the census population size of about 30, due to the rotational breeding scheme, which minimized inbreeding. This population size is small, but comparable to those in some natural populations like the Soay sheep (McRae et al., 2005), Darwin's finches (Grant and Grant, 1992) or Tasmanian devils (Epstein et al., 2016). These populations all spent different amounts of time at small population sizes, from a few decades in the recent case of Tasmanian devils to likely many millions of years in the case of Darwin's finches.

These different time scales lead to different expectations with respect to the selection response. Little to no new mutations have a chance to influence adaptation in the type of short-term selection seen in the Longshanks experiment, which might be compared to the case of the Tasmanian devils in that regard. The response will depend almost completely on standing variation, which will change in the longer term (» 20 generations), when *de novo* mutations will contribute more and more (Hill, 1982; Weber and Diggins, 1990).

Using both the complete pedigree as well as the founder individual's sequences, the data in this experiment was detailed enough to allow for precise modeling of the trait response - predicted shifts in the allele frequency distribution closely matched the results, with some deviations potentially due to selection on a few major loci, see Fig. 2.4.

Some loci of major effect were detected and their ranking according to an estimated selection coefficient determined, see Fig. 2.10. The Nkx3-2 locus was identified as the locus associated with the largest allele frequency shift parallel in both selected lines, connecting the trait change directly to allele frequency shifts at a particular locus which contributes a modest amount of variance to the selection response, but nevertheless is clearly differentiated from the infinitesimal background. Some studies that attempt to link explicit trait changes to changes

in allele frequencies are Keightley et al. (1996); Chen et al. (2019) and Nuzhdin et al. (1999). However none of these systematically tested against an infinitesimal background.

Our results imply a mixed genetic architecture with a few discrete loci amid an infinitesimal background. It remains to be seen whether other evolve and resequence studies with different parameters may reveal similar results. It is not clear how one might parameterize a generally applicable model for mixed genetic architectures. While we consider a small number of major effect loci together with a polygenic background the most likely genetic architecture in this study, we cannot reject other alternative models that could also account for the observed response. For example, an effectively infinitesimal model with linkage where larger allele frequency changes can happen by chance (Sachdeva and Barton, 2018), or on the other end of the spectrum, a model with few major trait loci, though we consider this unlikely - one would have to determine the minimum number of QTLs necessary for the response. Of course, there is a whole range in-between these extremes, which it would be interesting to explore in further study.

How do our results relate to other studies on complex traits and which types of genetic architectures have been observed in those cases? The classic example of a complex trait is human height. The strength of selection acting on this trait has been debated, see e.g. Turchin et al. (2012); Berg and Coop (2014), and Barton et al. (2019). It shows high heritability and a very polygenic genetic architecture, with the top contributing locus identified accounting for only 0.8% of the variation explained in European populations (Weedon et al., 2007; Wood et al., 2014).

The situation is completely different in populations of horses or dogs. Both these species have been under strong and sustained selection for hundreds of years and breed-specific populations tend to be small. Only 4-6 loci account for 83% and 50% of variation in height for horses and dogs respectively (Makvandi-Nejad et al., 2012; Rimbault et al., 2013). Interestingly, the major allele at the *IGF1* locus in dogs, which is a major determinant for small size, stems from standing genetic variation, as did the major effect alleles in the Longshanks experiment. In *Drosophila*, results from many selection experiments suggests a highly polygenic basis for a number of traits e.g. pupal size (Reeves and Tautz, 2017), egg size (Jha et al., 2015) or body size (Turner et al., 2011).

In the Longshanks experiment, the combined effects of low diversity in the founders and small founding populations are probably what leads to the extreme tail of the Δz^2 distribution (see e.g. Fig.2.3 and Fig.2.4) contributing a substantial part to the selection response, i.e. a considerable part of the response is due to individual loci of moderate effect sizes. Parallel selection provides especially convincing evidence for detecting selection (Chan et al., 2012; Schluter et al., 2004; Chan et al., 2010; Martin and Orgogozo, 2013), and also did so in the Longshanks experiment for the largest effect loci.

Overall however, there was little parallelism observed between each selected line and the Ctrl line, and also between simulated selected replicates. This is somewhat surprising, considering how simulated haplotypes were sampled directly from the actual founders which came from the same population. It goes to show that parallelism depends not only on shared selection pressure, but also the availability of large-effect alleles at appreciable frequency, which confer a substantial advantage - the latter of course are absent in the simulated replicates (right panel in Fig.2.9 and Fig.2.11). In the observed data, the two largest-effect loci did show parallelism, (figure 2.10). Next, it would be especially interesting to determine the extent to which linkage places a fundamental limit on our inference of signatures of selection, in addition to experimental design parameters like the population size and number of replicates.

Using the Longshanks selection experiment and contributions from theory, empirical data and molecular genetics, we showed that it is possible to identify *some* significant individual SNP which contributed to the selection response, while confirming that most of the variation contributed to the trait is spread over the whole genome and most likely due to many loci of small effect sizes. We showed that loci with the highest estimated effect sizes and selection coefficients increased in parallel in frequency in the two selected lines, giving convincing evidence for selection at specific loci. We hinted at the importance of LD in population genetic inference, and how it places fundamental limits on inference, an idea which will be developed further in the next chapter.

CHAPTER 3

Association between SNPs and Haplotypes

3.1 Introduction

This chapter was inspired by the analysis we did in Castro et al. (2019) which I describe in the previous chapter, where we looked for signatures of selection by identifying outlier windows via statistics based on allele frequency change under simple directional selection. The idea is that without selection, large allele frequency changes will be unlikely and thus Δp^2 (or its scaled version Δz^2 which was used in the LS analysis, see table 3.1) can be used as a test statistic for selection.

It was hard to distinguish signals of selection at individual loci, even though we knew that selection was acting and effective. We took this, and the fact that most of the genome showed some response to selection, as an indication of the polygenic basis of the selected trait, pointing to many small allele frequency shifts at many loci. However, even a more easily detectable signal of selection, such as hard selective sweep, can still be hard to detect. In this chapter, I examine the sources of noise in sequence data analysis which lead to these complications.

In the first part of this chapter, section 3.2, I focus on the effect of correlations between SNPs in the founder generation on the variance between replicates of summary statistics. In the second part, section 3.3, I then discuss how inference based on sequence data can be affected by the different sources of extra variability in the data.

3.2 Associations of SNPs with Haplotypes Contribute to Variance in an Allele Frequency-based Summary Statistic

As discussed in the general introduction (Chapter 1) evolution acts directly on blocks of genome, rather than individual loci. However, we can only observe the SNP on these blocks, which are a result of the random mutational process, followed by random shuffling due to Mendelian inheritance in sexual organisms. SNP share correlations due both to chance and the ancestral relationships between blocks with which they are associated. While it is convenient

Variable	Definition
p^{α}	Allele frequency at locus α in the ancestral population
L	Total number of SNP in a window
$\Omega = \frac{1}{L}\sum_{\alpha=1}^{L}\left(\Delta p^{\alpha}\right)^2$	Average allele frequency change squared over L loci
n_0	Initial number of haplotypes
n_T	Final number of haplotypes at generation T
k_i	Copy number of haplotype i; $\sum_{i=1}^{n_0} k_i = n_T$
$g_i^{\alpha} \in \{0,1\}$	Genotype at locus α, haplotype i
$D^{\alpha,\beta}$	LD coefficient between loci α and β
$\chi^{\alpha,\beta} = p^{\alpha}p^{\beta} + D^{\alpha,\beta}$	Haplotype frequency in the ancestral population
$\mathbb{E}\left[g_i^{\alpha}\right] = p^{\alpha}$	Expectation over randomly drawn SNP
$\mathbb{E}\left[g_i^{\alpha}g_i^{\beta}\right] = \chi^{\alpha,\beta}$	Expectation over randomly drawn pairs of SNPs
$p_0^{\alpha} = \frac{1}{n_0}\sum_{i=1}^{n_0} g_i^{\alpha}$	initial allele frequency at locus α in the sample
$p_T^{\alpha} = \frac{1}{n_T}\sum_{i=1}^{n_0} k_i g_i^{\alpha}$	final allele frequency at locus α in the sample
$\delta_i = \frac{k_i}{n_T} - \frac{1}{n_0}$	Change in frequency of haplotype i
$\Delta p^{\alpha} = \sum_{i=1}^{n_0} \delta_i \cdot g_i^{\alpha}$	Change in allele frequency in the final population
$S_k = \sum_{i=1}^{n_0}\left(\delta_i\right)^k$	Moments of the changes in haplotype frequency; $S_1 = 0$
r	Recombination rate
s	The selection coefficient
π	The average pairwise number of nucleotide differences per site
$X\left(t\right)$	The average allelic copy number at generation t
T_{neut}	Number of generations under neutrality
T_{sel}	Number of generations under selection
j_0	Allelic copy number in generation 0
j_T	Allelic copy number in generation T
$\ell\left(s\right)$	Likelihood of value s, given the data
H_0	The null hypothesis
$\Phi = \max_i\left(\frac{k_i}{n_T} - \frac{1}{n_0}\right)^2 = \max_i\left(\delta_i\right)^2$	The maximum haplotype frequency change per window squared.

Table 3.1: Table of definitions used in this chapter

to analyze data in windows and calculate summary statistics, it is also important to be mindful of the information lost in this process due to correlations between SNP, or in other words, LD. Correlations built up in the past continue to influence the present results of sequence data analysis (refer also to sections 1.2 and 2.1.1).

3.2.1 A window of genome

Let us now consider a window of genome small enough to neglect recombination break points, as illustrated in Fig.3.1. We can justify this approach by theory showing that over tens of

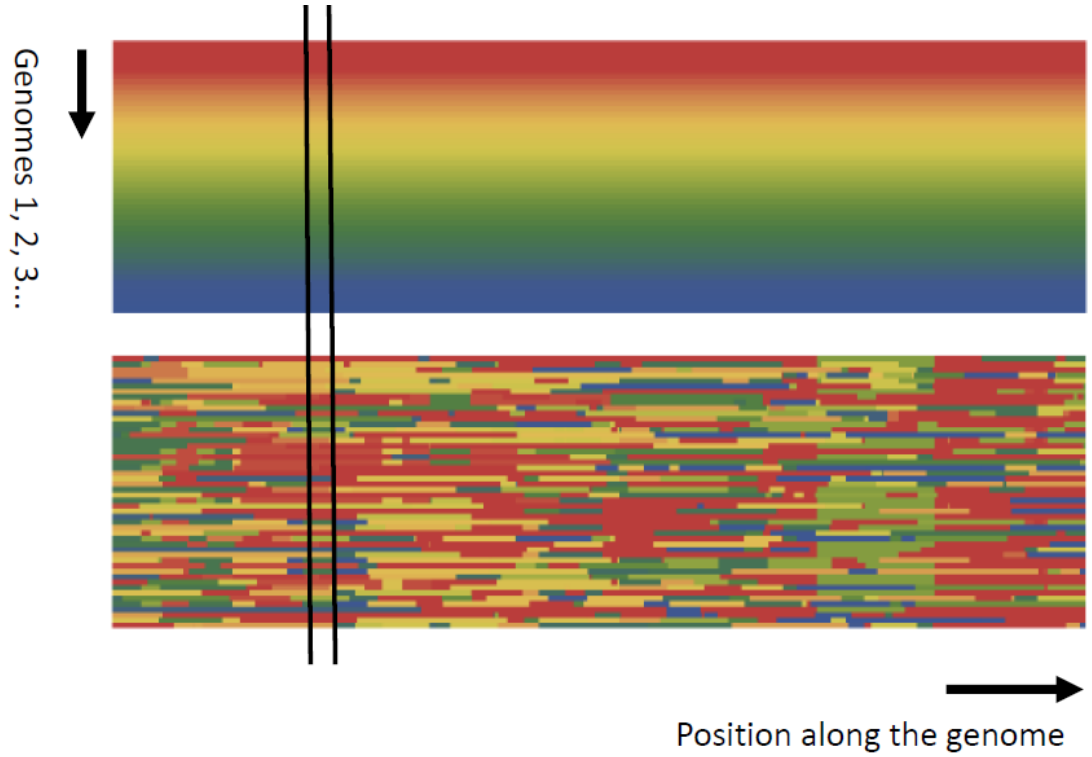

Figure 3.1: Choosing a window of genome small enough to be able to neglect recombination: The illustration shows individual genomes, each represented by a different color, lined up horizontally. In the upper part of the figure, we see the genomes before recombination. After a number of generations, genomes get broken up into blocks and rearranged into new configurations in each following generation. In the lower part of the figure, we see the result: blocks of a certain color, all descended from the same founder, start to increase in frequency in certain regions of the genome and will eventually fix, if there is no mutation introducing new variation. This will transform the image from all horizontal to all vertical "stripes". Since these haplotype blocks retain a certain minimum size up until fixation, it is possible to choose a window size for analysis which is small enough to reduce the chance of capturing a recombination break point within its bounds to the point where it can be neglected. This choice can be made from data after exploring the effects of window size on relevant statistics.

generations, large blocks of genome are passed on without recombination. Their size decreases as $1/r$ each generation, were r is the recombination rate. These blocks therefore retain a size where they still carry many SNP (see e.g. (Martin and Hospital, 2011; Chapman and Thompson, 2003; Tiret and Hospital, 2017; Sachdeva and Barton, 2018) for a more detailed analysis of haplotype block length distributions).

In the toy model I describe, I continue to refer to a "founding generation". Let us assume that haplotypes are known for this generation. I will continue to refer back to the founding generation and describe how the state of the population changed with respect to this first generation. We ignore the state of the population at earlier times.

Think of our window of genome as a collection of haplotype blocks, which at the start of our "experiment", are all distinct. I describe the "haplotype block copy number distribution" (HBCND) the following way: each haplotype has an index, (1, ...n_0), where n_0 is the number of haplotypes in the founding generation. From each generation to the next, the copy number of haplotypes may change, some will increase in frequency, while others may decrease and get lost. We therefore name the copy number of haplotype i k_i, where i is its index in the founding generation. On these haplotypes there are associated SNP, which I assume are biallelic. Their frequency changes in concert with the haplotypes they are associated with. Figure 3.2 illustrates how I record haplotype and SNP frequency changes over several generations.

When we have E&R data, meaning genotype data at at least the first and last observed generation, a possible (yet arbitrary) choice of summary statistic is Ω, the mean squared allele frequency change between first and last generation:

$$\Omega = \frac{1}{L}\sum_{\alpha=1}^{L}(\Delta p^{\alpha})^2 \tag{3.1}$$

In the previous chapter, we used $(\Delta z)^2$, the arcsin-transformed allele frequency squared, which has certain advantages (see section 2.1.1), but here we base our analysis on p, because it connects more straightforwardly to theory. I use the square because we usually do not know if selection will increase or decrease a certain allele (for this, we would have to have functional information), so direction should not matter in a first attempt to detect outliers.

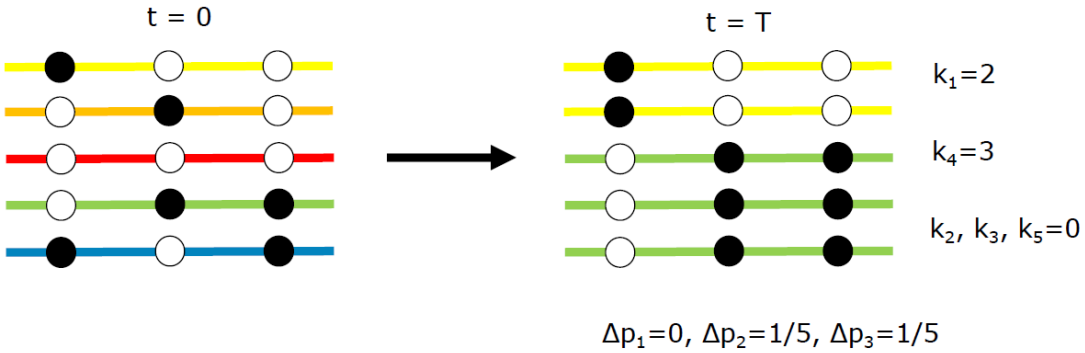

Figure 3.2: Illustration of how I describe allele and haplotype frequency changes: Allele frequency changes from generation $t = 0$ to $t = T$ and haplotype copy number distribution at generation $t = T$ with respect to generation t=0 are shown - assuming all haplotypes were distinct initially, haplotype $k_m = n$ means the m^{th} haplotype in the founder generation is present in n copies at generation T. In this example, the haplotype block copy number distribution is (1,1,1,1,1) in the first generation and (2, 0, 0, 3, 0) in the last one.

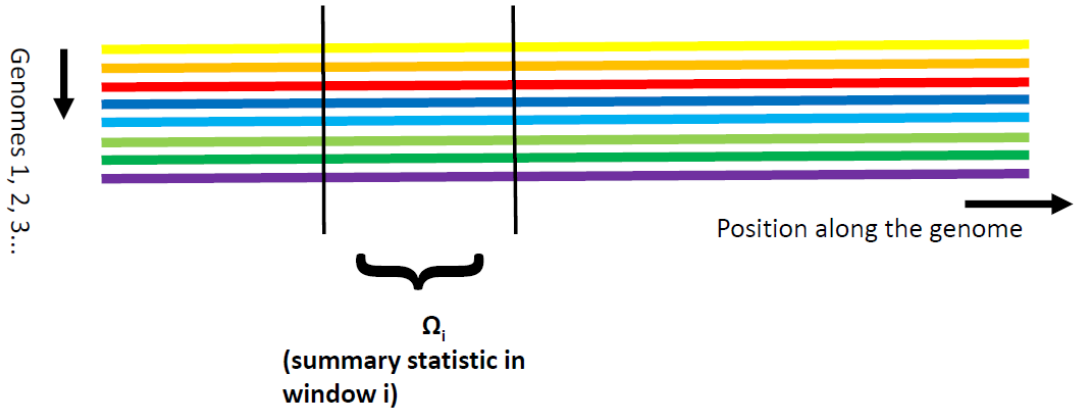

Figure 3.3: Window of genome with summary statistic $\Omega = \frac{1}{L}\sum_{\alpha=1}^{L}(\Delta p^{\alpha})^2$ where L is the number of SNPs in the window and Δp^{α} is the allele frequency change at locus α.

3.2.2 Sources of Noise in Allele Frequencies

We now examine how evolutionary processes' action on haplotype blocks leads to extra variability in allele frequencies and their various measures; "extra" in addition to variation due to the actual signal. Much of this extra variability is due to LD. LD is created in two main ways:

First, due to LD between SNPs in the founding generation. To explain this, I will state here the key result of this chapter, which gives the variance in summary statistic Ω in terms of LD between SNPs in the initial generation and initial allele frequencies; the equation is derived in full below, in section 3.2.3.

$$Var\,(\Omega) = \frac{1}{L^2}\Big(\sum_{\alpha=1}^{L} p^{\alpha}q^{\alpha}(2p^{\alpha}q^{\alpha}S_2^2 + (1-6p^{\alpha}q^{\alpha})S_4) + \sum_{\alpha,\beta=1,\alpha\neq\beta}^{L} D^{\alpha,\beta}(p^{\alpha}-q^{\alpha})(p^{\beta}-q^{\beta})S_4 + 2(D^{\alpha,\beta})^2(S_2^2-S_4)\Big) \tag{3.2}$$

The first term in the equation gives the contribution to the overall variance due to allele frequencies at individual loci. The second term can be ignored if LD is not directional. The third term, containing $(D^{\alpha,\beta})^2$ gives the contribution to the variance due to LD between SNP in the founding generation. This result shows that LD in the founder population inflates the variance in the statistic, and it will dominate as the number of loci increases - even if these correlations are very weak, there can be very many pairs of loci in a single window, and thus the effect adds up. Also, note the there will always be some correlation between SNP due to chance, even if they are assigned in linkage equilibrium at the start. These correlations, present already in the founding generation, inflate variance over time.

Second, due to random genetic drift, some haplotypes increase in frequency over time and others are lost, as illustrated in Fig. 3.2. This changes the haplotype block copy number distribution over time, and the total number of haplotypes goes down (as long as we can neglect mutation). Through this process, LD increases over time, because now SNP on all

haplotypes which descend from the same founding haplotype are correlated. This type of LD does not appear in equation 3.2.

In addition, because the precise location of recombination breakpoints is random, the way SNP are grouped together on haplotype blocks bounded by these breakpoints is random too, which also adds to the overall variance. The inflation of variance will show up as increased variance between replicates in simulations or real data. One important conclusion we can draw from this immediately is that working with haplotypes directly, we can always decrease the amount of uncertainty in our results, because this way the first source of LD and thus uncertainty disappears: correlations between SNP in the founding generation become irrelevant. In what follows I will express the variance in the summary statistic in terms of the first four moments of the HBCND, the initial allele frequencies and the average pairwise LD in the founder generation. Simulation of populations with varying levels of initial LD were used to check this.

3.2.3 The Variance in the Summary Statistic due to LD

My first aim is to obtain the expected variance in an allele frequency-based summary statistic due to LD produced by the evolutionary forces acting on haplotype blocks.

Analytical Derivation

Consider a window of genome as illustrated in figures 3.1 through 3.2. The collection of haplotype blocks making up this window changes composition over time. This corresponds to a standard Wright-Fisher process.

After T generations, we describe the resulting population in terms of the distribution of haplotype blocks (the HBCND), referring back to the founding generation. This distribution of haplotype block copy numbers can be written as a list: $K = \{k_1, k_2, ...k_i, ...k_{n_0}\}$, and $\sum_{i=1}^{n_0} = n_T$, where element k_i is the number of copies of haplotype i from the first generation left at generation T. n_0 is the initial total number of haplotype blocks and n_T is the total number of haplotype blocks at generation T. To make the connection to allele frequencies, we aim to express Ω (see 3.1) in terms of K. The variance in Ω can be decomposed into a variance and a covariance term:

$$Var(\Omega) = \frac{1}{L^2}\left(\sum_{\alpha=1}^{L} Var\left(\Delta p^{\alpha}\right)^2 + \sum_{\alpha,\beta=1,\alpha\neq\beta}^{L} Cov((\Delta p^{\alpha})^2, \left(\Delta p^{\beta}\right)^2\right) \tag{3.3}$$

Define $g_i^{\alpha} \in \{0, 1\}$ as the genotype at locus α on haplotype i. Note that we assume all loci to be biallelic. The genotype is defined in relation to a reference genome: 0 if the individual carries the reference allele, 1 if the individual carries the alternative allele. The frequency of haplotype $\chi \in (00, 01, 10, 11)$ at loci α and β is

$$\chi^{\alpha,\beta} = p^{\alpha}p^{\beta} + D^{\alpha,\beta} \tag{3.4}$$

where p^{α}/p^{β} are the allele frequencies at loci α/β in the initial generation and $D^{\alpha,\beta}$ is the sample average pairwise LD between loci α and β for many randomly "assigned" SNP configurations. Taking the expectation gives:

$$\mathbb{E}\left[g_i^{\alpha}\right] = p^{\alpha} \tag{3.5}$$

$$\mathbb{E}\left[g_i^{\alpha} \cdot g_i^{\beta}\right] = \chi^{\alpha,\beta} \tag{3.6}$$

The change in allele frequency at locus α in the sampled population is:

$$\Delta p^{\alpha} = \sum_{i=1}^{n_0} \delta_i \cdot g_i^{\alpha} \tag{3.7}$$

where

$$\delta_i = \frac{k_i}{n_T} - \frac{1}{n_0} \tag{3.8}$$

Then, the k^{th} moment of the distribution of haplotype frequency changes is defined as

$$S_k = \sum_{i=1}^{n_0} (\delta_i)^k \tag{3.9}$$

The goal is to obtain an expression for $Var(\Omega)$ in terms of moments of the haplotype block copy number changes Δk_i, initial LD between SNPs $D^{\alpha,\beta}$, and initial allele frequencies p^{α} and p^{β}. First, I express the sums over products of distinct elements in the list of haplotype frequency changes in terms of the moments of these changes, see table 3.2.

Indices	Numbers of Permutations
$\{i,i,i,i\}$	n_0
$\{i,i,i,j\}$	$4n_0(n_0-1)$
$\{i,i,j,j\}$	$3n_0(n_0-1)$
$\{i,i,j,k\}$	$6n_0(n_0-1)(n_0-2)$
$\{i,j,k,l\}$	$n_0(n_0-1)(n_0-2)(n_0-3)$

Table 3.2: Number of ways of choosing distinct sets of elements for four indices that sum to a total number of n_0^4 elements.

Computing the individual terms:

$$\begin{aligned}&\mathbb{E}\left[(\Delta p^{\alpha})^2\right] = \sum_{i,j=1}^{n_0} \delta_i\delta_j \mathbb{E}\left[g_i^{\alpha} g_j^{\alpha}\right] = \sum_{i}^{n_0} \delta_i^2 \mathbb{E}\left[(g_i^{\alpha})^2\right] + \sum_{i\neq j} \delta_i\delta_j \mathbb{E}\left[g_i^{\alpha} g_j^{\alpha}\right] = \\ &S_2 p^{\alpha} + (p^{\alpha})^2 \sum_{i\neq j} \delta_i\delta_j = S_2 p^{\alpha} q^{\alpha}\end{aligned} \tag{3.10}$$

$$\begin{aligned}&\mathbb{E}\left[(\Delta p^{\alpha})^4\right] = \sum_{i,j,k,l=1}^{n_0} \delta_i\delta_j\delta_k\delta_l \mathbb{E}\left[g_i^{\alpha} g_j^{\alpha} g_k^{\alpha} g_l^{\alpha}\right] = \\ &\sum_{i}^{n_0} \delta_i^4 \mathbb{E}\left[(g_i^{\alpha})^4\right] + 4\sum_{i\neq j} \delta_i^3\delta_j \mathbb{E}\left[(g_i^{\alpha})^3 g_j^{\alpha}\right] + 3\sum_{i\neq j} \delta_i^2\delta_j^2 \mathbb{E}\left[(g_i^{\alpha})^2 \left(g_j^{\alpha}\right)^2\right] + \\ &6\sum_{i\neq j\neq k} \delta_i^2\delta_j\delta_k \mathbb{E}\left[(g_i^{\alpha})^2 g_j^{\alpha} g_k^{\alpha}\right] + \sum_{i\neq j\neq k\neq l} \delta_i\delta_j\delta_k\delta_l \mathbb{E}\left[g_i^{\alpha} g_j^{\alpha} g_k^{\alpha} g_l^{\alpha}\right] = \\ &S_4 p^{\alpha} - 4S_4 (p^{\alpha})^2 + 3\left(S_2^2 - S_4\right)(p^{\alpha})^2 + 6\left(2S_4 - S_2^2\right)(p^{\alpha})^3 + 3\left(S_2^2 - 2S_4\right)(p^{\alpha})^4 = \\ &p^{\alpha} q^{\alpha}\left(3p^{\alpha} q^{\alpha} S_2^2 + \left(1 - 6p^{\alpha} q^{\alpha}\right) S_4\right)\end{aligned} \tag{3.11}$$

Sums over Distinct Elements	
Sum	In terms of moments of the distribution of haplotype frequency changes
$\sum_{i\neq j}\delta_i\delta_j$	$S_1^2 - S_2$
$\sum_{i\neq j}\delta_i^3\delta_j$	$S_1S_3 - S_4$
$\sum_{i\neq j}\delta_i^2\delta_j^2$	$S_2^2 - S_4$
$\sum_{i\neq j\neq k}\delta_i^2\delta_j\delta_k$	$S_2S_1^2 + 2S_4 - S_2^2 - 2S_3S_1$
$\sum_{i\neq j\neq k\neq l}\delta_i\delta_j\delta_k\delta_l$	$S_1^4 - 4(S_1S_3 - S_4) - 3(S_2^2 - S_4) - 6(S_2S_1^2 + 2S_4 - S_2^2 - 2S_3S_1) - S_4 = S_1^4 - 6S_1^2S_2 + 3S_2^2 + 8S_1S_3 - 6S_4$

Table 3.3: Sums over distinct elements of products of haplotype frequency changes in terms of the haplotype block copy number distribution. For example, this is how to derive the first expression:
$\left(\sum_{i,j}\delta_i\right)^2 = S_1^2$
$\sum_{i=j}\delta_i^2 + \sum_{i\neq j}\delta_i\delta_j = S_1^2$
$S_2 + \sum_{i\neq j}\delta_i\delta_j = S_1^2$
$\sum_{i\neq j}\delta_i\delta_j = S_1^2 - S_2$
Similarly for the other expressions.

And we obtain the variance:

$$Var\left[(\Delta p^\alpha)^2\right] = \mathbb{E}\left[(\Delta p^\alpha)^4\right] - \mathbb{E}\left[(\Delta p^\alpha)^2\right]^2 = p^\alpha q^\alpha\left(2p^\alpha q^\alpha S_2^2 + (1-6p^\alpha q^\alpha)\,S_4\right) \quad (3.12)$$

$$\begin{aligned}
&\mathbb{E}\left[(\Delta p^\alpha)^2\left(\Delta p^\beta\right)^2\right] = \sum_{i,j,k,l=1}^{n_0}\delta_i\delta_j\delta_k\delta_l\mathbb{E}\left[g_i^\alpha g_j^\alpha g_k^\beta g_l^\beta\right] = \\
&\sum_i^{n_0}\delta_i^4\mathbb{E}\left[(g_i^\alpha)^2\left(g_i^\beta\right)^2\right] + 2\sum_{i\neq j}\delta_i^3\delta_j\left(\mathbb{E}\left[(g_i^\alpha)^2 g_i^\beta g_j^\beta\right] + \mathbb{E}\left[\left(g_i^\beta\right)^2 g_i^\alpha g_j^\beta\right]\right) + \\
&\sum_{i\neq j}\delta_i^2\delta_j^2\left(\mathbb{E}\left[(g_i^\alpha)^2\left(g_j^\beta\right)^2\right] + 2\mathbb{E}\left[\left(g_i^\alpha g_i^\beta\right)\left(g_j^\alpha g_j^\beta\right)\right]\right) + \\
&\sum_{i\neq j\neq k}\delta_i^2\delta_j\delta_k\left(\mathbb{E}\left[(g_i^\alpha)^2 g_j^\beta g_k^\beta\right] + \mathbb{E}\left[\left(g_i^\beta\right)^2 g_j^\alpha g_k^\alpha\right] + 4\mathbb{E}\left[\left(g_i^\alpha g_i^\beta\right) g_j^\alpha g_k^\beta\right]\right) + \\
&\sum_{i\neq j\neq k\neq l}\delta_i\delta_j\delta_k\delta_l\mathbb{E}\left[g_i^\alpha g_j^\alpha g_k^\alpha g_l^\alpha\right] = \\
&S_4\left(p^\alpha p^\alpha + D^{\alpha,\beta}\right) - 2S_4\left(p^\alpha p^\alpha + D^{\alpha,\beta}\right)\left(p^\alpha + p^\beta\right) + \left(S_2^2 - S_4\right)\left(p^\alpha p^\beta + \right. \\
&\left.2\left(p^\alpha p^\beta + D^{\alpha,\beta}\right)^2\right) + \left(2S_4 - S_2^2\right)\left(p^\alpha p^\beta\left(p^\alpha + p^\beta\right) + \right. \\
&\left.4p^\alpha p^\beta\left(p^\alpha p^\beta + D^{\alpha,\beta}\right)\right) + 3\left(S_2^2 - 2S_4\right)\left(p^\alpha p^\beta\right)^2
\end{aligned} \quad (3.13)$$

This gives

$$Var\left((\Delta p^\alpha)^2\right) = p^\alpha q^\alpha\left(2p^\alpha q^\alpha S_2^2 + (1-6p^\alpha q^\alpha)S_4\right) \quad (3.14)$$

and

$$Cov\left((\Delta p^{\alpha})^2, \left(\Delta p^{\beta}\right)^2\right) = D^{\alpha,\beta}(p^{\alpha}-q^{\alpha})(p^{\beta}-q^{\beta})S_4 + 2(D^{\alpha,\beta})^2(S_2^2 - S_4) \tag{3.15}$$

For the variance of Ω we then have:

$$\begin{aligned} Var\left(\Omega\right) = \frac{1}{L^2}\Big(&\sum_{\alpha=1}^{L} p^{\alpha}q^{\alpha}(2p^{\alpha}q^{\alpha}S_2^2 + (1-6p^{\alpha}q^{\alpha})S_4) + \\ &\sum_{\alpha,\beta=1,\alpha\neq\beta}^{L} D^{\alpha,\beta}(p^{\alpha}-q^{\alpha})(p^{\beta}-q^{\beta})S_4 + 2(D^{\alpha,\beta})^2(S_2^2-S_4)\Big) \end{aligned} \tag{3.16}$$

As L, the total number of SNP, gets larger and larger, the third term $2(D^{\alpha,\beta})^2(S_2^2 - S_4)$ will start to dominate the equation. This is because for more and more pairs of SNP in a given window, the LD in the founder population inflates the variance more and more.

Numerical Calculation of the Variance of Ω

For a specified allele frequency at a given locus, there are many different ways alleles can be associated with haplotype blocks. To calculate the variance of Ω numerically, I repeatedly assign SNP to the genome window at random, but consistent with defined allele frequencies at each locus. I analyze this for different values of LD, which are produced by mixing two populations with a different constant starting frequency each. In this case the difference between the minor allele frequencies matters - the larger this difference, the higher LD.

The variance in Ω is higher for higher levels of LD. Even as the number of SNP per window gets very large, the variance will never go to zero, because adding SNP does not add new information if they are fully correlated with SNP that have already been scored.

3.3 Inference via Allele Frequency Data is Noisier than Inference via Haplotype Frequency Data

In this section, I illustrate the issues of inferences for recent selection when only indirect observation via SNP is possible, and how much better we can do using haplotypes if they are available.

To properly estimate the variability in summary statistics based on allele frequencies, I use the example of an E&R experiment with selection for a specific trait like the ones examined in Chapters 1 and 3. It should be noted, however, that the issue of correlations between SNP inflating variance is more general and influences all types of analysis using allele frequency based summary statistics in windows, not only E&R type data.

Say we aim to do inference from statistics based on changes in allele frequencies. Let us consider first the case of no LD - here, one can simply aim to identify the causal SNP by looking for loci which show exceptionally large changes in Δp (or any sensible statistic based on it). Usually, we rely on neutral marker SNP to identify causal loci - this means that while long range LD can obscure the signal, we need short-range LD to be able to detect causal SNP at all, because if neutral markers are not linked to a causal SNP, we will not detect selection.

If LD is only present on a very short scale, as is often assumed and has been found in some studies (Kruglyak, 1999; Dunning et al., 2000), we can look at exceptional windows, and

count on neutral marker SNP to be linked to the causal SNP and therefore show exceptional changes. However, reality is often more complicated and, due to the reasons discussed in the first part of this chapter, even relatively clear signals of selection such as hard sweeps can be hard to detect.

Due to those same reasons, values of the statistics we use will be correlated and highly variable between replicates. Commonly used statistics such as π or F_{ST}, are "local", i.e. they do not use information from LD, so we expect increased noise due to correlations between allele frequencies.

Here, we assume two fitness classes of haplotypes - they either carry or do not carry the beneficial allele and have fitnesses $1 : 1+s$. I describe the single locus case to present the main ideas. One of the two alleles at the locus, coded as 1, is selected with selection coefficient s. I simulate data and check how well I can infer the selection coefficient just from allele frequency change at the locus with a maximum likelihood scheme. Of course, one would then like to progress to inference in a window with many SNP and LD, using the mean Δp^2 for inference, or a statistic based on the haplotype frequency changes. However, since this problem still remains in the almost-solved stage, where it has been lingering for several months now, only an outline of this more interesting analysis has made it into the Discussion and Outlook section of this chapter, section 3.4.

In terms of statistics, we can do inference either by using likelihood directly, comparing the likelihood to see the data in different scenarios, or by using a frequentist significance test. These should lead to similar quantitative outcomes, but one might argue that using the full likelihood curve rather than an arbitrary cut-off preserves more information carried by the data. In the single SNP case with replicates, the signal comes from the distribution of replicate Δp values at the locus under a certain strength of selection. The idea is to simulate the distribution of values under different s, and then use a single Δp value from the "actual data" for inference, by calculating its likelihood under each selection coefficient, interpolate over these values, find the maximum likelihood estimate (MLE) and determine support limits (which we can choose to correspond to 95% significance levels).

3.3.1 Producing data with selection

As stated above, to include selection in the Wright-Fisher model I assume a simple two class model: there is one locus with a selected allele and haplotype blocks fall into two categories - they either carry this allele or not. Say the alternative allele, coded as 1, is selected. Haplotypes carrying the reference allele have fitness 1, while haplotypes carrying the selected alternative allele have fitness $1+s$. The fitness advantage of some haplotypes is implemented by adjusting the probabilities so that they still sum to 1 when drawing from the multinomial distribution which is appropriate for the Wright-Fisher process. If we want to simulate a hard sweep, only one haplotype should initially carry the beneficial allele.

In the simulations, the window of genome first goes through a certain number of generations of neutral drift. This is to simulate the ancestral relationships SNP will have with each other in any natural sample for the simplest scenario of the standard coalescent. This neutral process builds up LD. With very low initial LD, signals are hard to detect, so some appreciable amount of LD is needed. After this neutral period, selection starts. In each generation of multinomial sampling, haplotypes carrying the selected allele have a slightly higher probability to be selected. The number of different *types* of haplotypes goes down quickly, but there can still be a lot of diversity as measured e.g. by π, the average pairwise diversity - see figure 3.4.

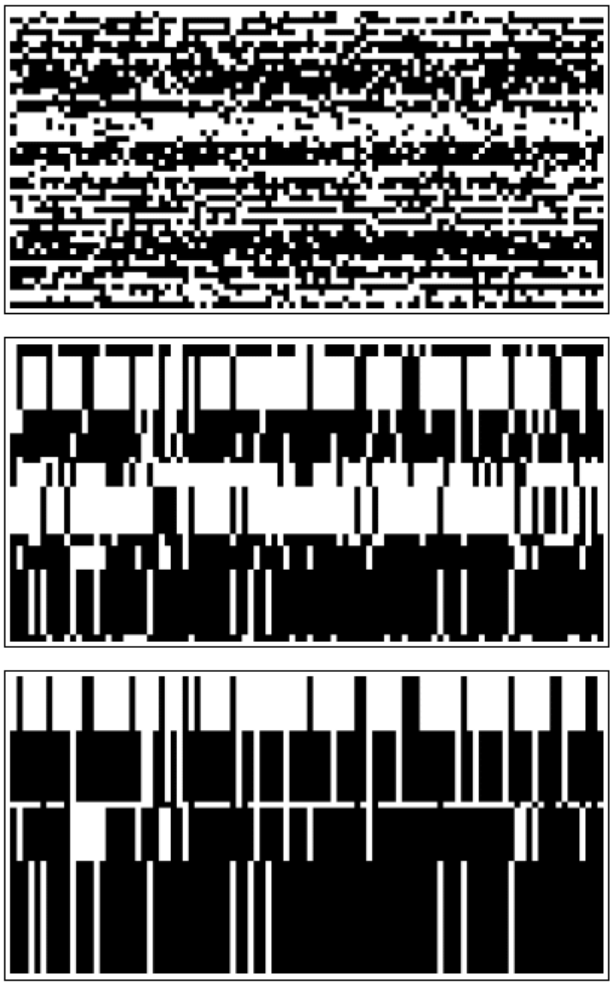

Figure 3.4: To illustrate how a window of genome might change under drift and selection, here I show such a window with 50 haplotypes and 100 SNP. The initial population was produced by mixing two source populations with constant initial copy numbers at all SNPs: j_0 in source population 1 and $j_0 + \Delta j_0$ in source population 2. Here, $j_0 = 10$ and $\Delta j_0 = 30$ were used. The larger the difference between initial copy numbers in the source populations, the higher LD. The reference alleles is white, the alternative allele is black. First row: The initial window, second row: the same window after 6 generations of drift without mutation, final row: the same window after 6 generations of selection without mutation.

3.3.2 The Transition Matrix and Likelihood

The transition matrix in the context of this thesis is the right stochastic matrix associated with the Wright-Fisher process - a discrete time Markov process. Say the allelic copy number at generation t is $X(t) = i$ and the transition probability P_{ij} to go from $X_t = i$ to $X_{t+1} = j$ from generation t to $t + 1$ for the Markov chain X_n is drawn from the binomial distribution with population size n:

$$P_{ij} = \binom{2n}{j} \left(\frac{i}{2n}\right)^j \left(1 - \frac{i}{2n}\right)^{2n-j} \quad \text{for} \quad i, j \in (1, 2, ...2n) \tag{3.17}$$

For a population of size n with a selection coefficient s (using fitnesses $1 : 1 + s$) the binomial probabilities are given by the standard population genetic equation for allele frequency change under selection. Here, p_t is the allele frequency at a given locus in generation t, and p_{t+1} the allele frequency at the same locus in generation $t + 1$:

$$p_{t+1} = p_t + \frac{sp_tq_t}{1 + sp_t} \tag{3.18}$$

To calculate the likelihood of a given pair of allele frequencies at generations $t = 0$ and $t = T$ after drift and selection, take the matrix exponential of the transition matrix to the power of T.

3.3.3 Inference in the Single SNP Case

I demonstrate the process of inferring the selection coefficient via the case of a single locus, going through the process as described in section 3.3 and subjecting the population to T_{neut} generations of random genetic drift and T_{sel} generations of selection. Every replicate is a pair of numbers (j_0, j_T), i.e. the allelic copy number in the first and last generations. In the one locus case I can use the transition matrix to calculate the likelihood.

As described in 3.3.2, the likelihood of a given pair of copy numbers can be calculated at any value of s. I calculate the log-likelihood at $s = (0.0, 0.1, 0.2, 0.3, 0.4, 0.5)$ and interpolate to obtain the full log-likelihood curve. The curve's maximum is the maximum likelihood estimate (MLE) for s. Support limits are given by the two s-values corresponding to the points at which the likelihood curve would intersect a horizontal line two units below the MLE. These are the 2-unit support limits, which correspond to 95% confidence intervals for large samples.

Using replicates, the estimates can be improved by simply adding up the log-likelihood values at each s-value and finding the maximum of the resulting curve. Note that by definition the likelihood of the true model will always be lower than at the maximum likelihood estimate, see Fig.3.5.

Likelihood Ratio Test

As a check and to calculate the power to detect selection, I use a likelihood ratio test, in which the likelihood of the data under the null model is compared to a certain critical value. In a likelihood ratio test, there is a defined null model. In my analysis this is s=0, meaning the data was produced under neutrality. The difference between the likelihood at the MLE and at s=0 can be used to judge whether the data is consistent with the null hypothesis or not. The statistic based on which this decision is made is:

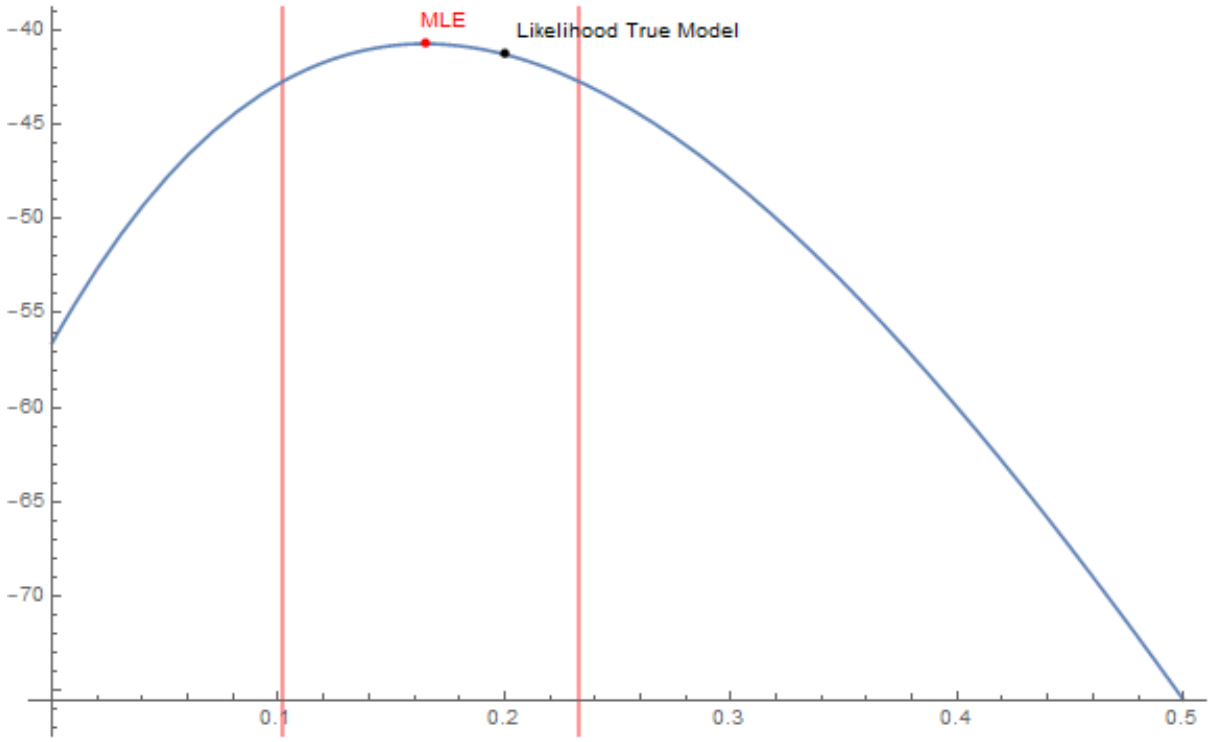

Figure 3.5: The Log-likelihood curve at a single locus for going from 10 to 24 copies after 6 generations of selection with selection coefficient s=0.2 for diploid population size Ne = 50 (example). The 2-unit support limits in red are at s=(0.102, 0.233).

$$\lambda = 2 \cdot (\ell(s_{MLE}) - \ell(s_{H_0})) \tag{3.19}$$

Due to the asymptotic normality of the maximum likelihood estimate for large sample sizes, we can form confidence intervals using the "asympotic normal approximation". For large samples, the log-likelihood of θ, some parameter pertaining to a discrete probability distribution (for example s producing discrete copy number changes in haplotypes or alleles), approaches a parabola centered at $\hat{\theta}$, the MLE.

According to Wilk's theorem, the statistic λ is approximately distributed like a χ^2_k distribution, where k is the number of degrees of freedom equal to the difference in dimensionality between θ and θ_{H_0} (1 in the current analysis, since s ranges in one dimension and its estimate at any single point is a single value and therefore has 0 dimensions) (Wilks, 1938). To reject H_0 at the α significance level, one has to check if the statistic exceeds the $100 \cdot (1-\alpha)^t h$ percentile of the χ^2 distribution. This is the value $\chi^2_\alpha(k)$, such that the area under the curve to the right of $\chi^2_\alpha(k)$ is α. To calculate critical value x_{crit} compute the x-value at which the cumulative distribution function (CDF) of $\chi^2_\alpha(k)$ reaches value 1-α. At significance level $\alpha = 0.05$ for one degree of freedom, the critical value x_{crit} is 3.841.

Likelihood Ratio-based Confidence Intervals and Support Limits

To find likelihood-ratio based confidence intervals, I find all values s^* for which the log-likelihood $\ell(s)$ is within a given tolerance of the MLE. An approximate $100 \cdot (1-\alpha)\,\%$ confidence interval of s consists of all the possible s^* for which the null hypothesis $H_0 : s = s_0$ can not be rejected at the α level. For the 5% significance level, the interval consists of all s^* for which $2 \cdot \ell(s^*) - \ell(s_0) <= 3.84$ or $\ell(s^*) >= \ell(s^*) - 1.92$. In other words, the 95% confidence

interval includes all values of s^* for which the log-likelihood function drops off by no more than 1.92 units. This corresponds closely to the 2-unit support limit, which includes all values for which the log-likelihood drops by no more than 2.

The Power to Detect Selection

The power to detect selection is the probability that the null hypothesis is rejected given that it is false. Fig. 3.6 shows the increase in power for increasing numbers of replicates for the single SNP case.

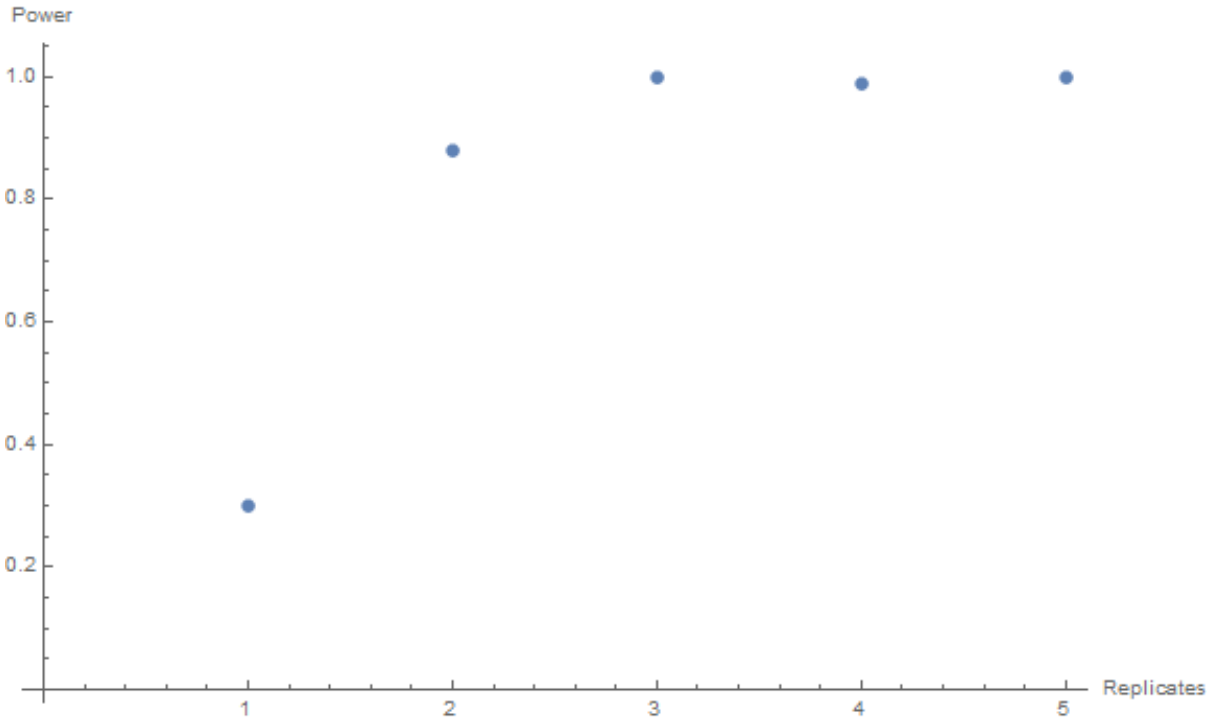

Figure 3.6: This shows the power to detect selection at a single locus for 100 haplotypes and initial copy number j0 = 10 at the locus. After 5 generations of neutral evolution without mutation, 5 generations of selection without mutation at selection coefficient s=0.2 were applied. A likelihood ratio test is done based on a likelihood curve built from 1, 6, 11, 16 or 21 replicates. Statistic over 100 repetitions each.

Independent Inference for SNP in a window

Here I simulate a genome window without linkage to make the simple point that even in this case the causal SNP is not necessarily the one with the highest allele frequency change, and therefore would not be associated with the highest inferred selection coefficient, figure 3.7.

3.4 Discussion and Outlook

In this chapter, I introduced a simple model in which we observe a window of genome without recombination and follow frequency changes of both haplotype blocks and alleles in the window. I discussed the different sources of noise in statistics based on allele frequencies and derived an equation to quantify the variance in one such statistic in terms of the first four moments of the HBCND, the initial allele frequencies, and LD between alleles in the founding generation,

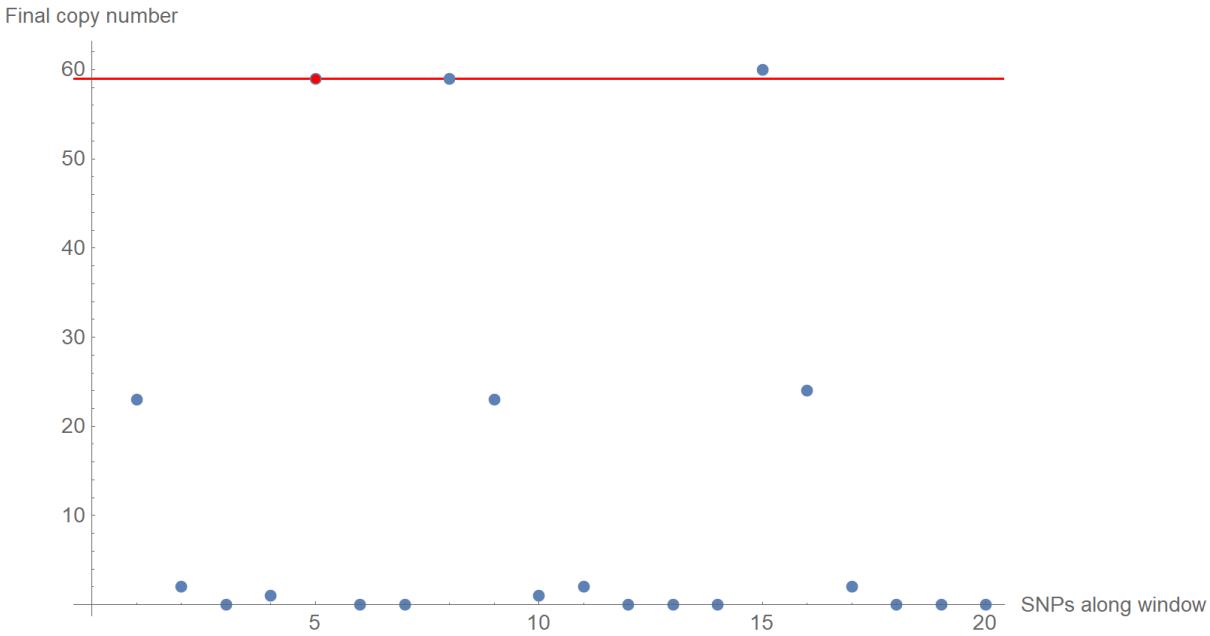

Figure 3.7: Simulations show that the causal SNP does not necessarily show the highest allele frequency changes. Here, one particular replicate is shown. The simulated window of genome has 100 haplotypes and 20 SNP. The alternative allele starts at copy number 10, with copies assigned in linkage equilibrium. 10 generations of neutral evolution without mutation are followed by 10 generations under selection at selection coefficient s=0.2. The red point represents the causal SNP at the 5th position along the genome. Here 2 SNP experienced a higher positive frequency change than the selected SNP, despite being neither selected nor linked to the selected SNP. For 100 replicates, the mean number of SNPs experiencing a higher positive frequency change than the causal allele is ~ 4, with a very large variance of ~ 23.

the latter of which starts to dominate the as the number of SNP in a window grows. I then discussed various issues relating to inference from such data. I illustrated inference to detect selection in the single SNP case to prepare the analysis of windows of genome with many correlated SNP.

Connecting to this analysis, let me discuss the next steps in this project. To investigate inference in windows with many correlated SNP in a window, I am planning to compare statistics Ω, based on allele frequency changes and Φ (see table 3.1), based on haplotype frequency changes. To produce data in this scenario, I would again assume two haplotype fitness classes by assigning a single beneficial allele to (in the simplest case) only one haplotype. I would then draw haplotypes according to a Wright-Fisher model, and then assign SNP to them consistently. In other words, data is then produced in two steps - First the haplotype trajectory both of the fitter haplotype class and the other classes, and second by overlaying SNP consistently.

One could then either condition on non-loss or also include the information about how often the SNP is lost to use it for the inference. It will need many replicates in a range of parameters where there is enough power, to detect a signal at all. The general idea would then be to find the distribution of these statistics as a function of Ns by simulations, where N is the population size and s is the selection coefficient. One can then interpolate over the likelihood values for different s values, which gives a continuous likelihood surface.

Then, using a single Ω or Φ value, the log likelihood curve under each value of s could be

calculated. Finding its maximum again gives the MLE, and finding the points at which the log likelihood has dropped by 2 units gives the 2-unit support limits. Looking at the (average) curvature of the log likelihood function would allow me to find the power of each statistic for estimating selection. For regular problems (i.e. the likelihood around the MLE is closely approximated by a concave quadratic function) this gives an indication as to how sharply peaked and therefore informative the log-likelihood curve is. Other statistics could be tested as well to find the best one. It is important to choose parameter ranges that have enough power to distinguish the signal, which can be tricky. Fundamentally, the distributions of the statistic under different s have to be *different enough* to be distinguished.

The issues discussed in this chapter are closely related to the study of "selective sweeps" i.e. the increase of beneficial alleles due to positive selection. The signature of a classic sweep (due to one new mutation originating on a single haplotype) is in principle detectable as a reduction neutral variation surrounding the sweep site (Maynard Smith and Haigh, 1974). Other more complex types of sweeps are common (Hermisson and Pennings, 2017). There are a number of issues which may obscure the signal of selection left by a selective sweeps (see e.g. Bamshad and Wooding (2003)) which are closely related to the topics discussed in this chapter.

It can be hard to detect selection even when haplotypes are available, which might seem somewhat surprising. As discussed in section 3.2.2, using haplotypes directly should be at least a little bit more informative, since we completely avoid one of the sources of noise in the data: the correlations between SNP in the founding generation. However, there are still several sources of excess variability. For example, there may be more than two haplotype fitness classes, which makes it harder to detect a particular beneficial allele, and LD between any one haplotype and the fittest class might be incomplete. This could happen when the beneficial allele is spread over multiple haplotypes, or there is a more complicated distribution of fitnesses across haplotypes rather than the simplest case of two haplotype fitness classes. Also, sweeps may not be complete, leading to a weaker signal. Therefore, even using the full haplotype structure, we will not necessarily have power to detect selection.

In general, there is a big difference between looking at statistics which estimate the general degree of change in haplotype frequencies, without identifying the causal SNP and looking at the "true" Δp at the causal locus/haplotype, which of course remains an unobtainable gold standard when working with real data. In simulations, one should be able to quantify the different sources of noise to estimate how much information we lose, and which kinds of statistic minimize this noise. The topic is extremely relevant for our quest to try to understand adaptation via sequence data and will remain so even as more methods for obtaining haplotypes and more appropriate methods for sequence data analysis are being developed.

CHAPTER 4

Analysis of an Evolve and Resequence Experiment in Drosophila

4.1 Introduction

In this chapter I describe the analysis of an evolve and resequence experiment in *Drosophila melanogaster*. The experiment had a small founding population size (n=4) and selection over only a few generations, but many replicates.

The goal of our analysis in this project is to learn about the genetic basis of a complex trait: pupal size in flies. This trait was chosen because it is highly heritable in *Drosophila* and phenotyping can be automated. Signatures of selection can be informative about the number of genes impacting a trait and their effect on the pupal size distribution. Here, we investigate the genome for such signatures by observing the genetic changes - specifically allele frequency or haplotype frequency changes, summarized by windows-based statistics as described in the introduction, section 1.3.

Here I will present the background, first results and end with a discussion of the limitations of the current approach and a plan for further research, since this project is still in progress.

4.1.1 Main Questions

The main questions in this project are:

- Where along the genome can we detect signatures of selection?
- What can we say about the distribution of effect sizes of genes contributing to variance in the trait?
- What limits to the statistical inference of selection or other population genetic parameters do we encounter? Are these limits of a technical or fundamental nature?

For this data set, we know that selection was acting and effective, since pupal case length changed on average by 0.306mm $\pm$ SD 0.188 (standard deviation of the per family change, taken over all families) between G6, the first selected generation, to G11, the last generation. Trait variance seems to have decreased very little This corresponds to 1.62 std.dev., so it should

be possible to reject a neutral model in favor of a model including selection. Going one step further, similarly to the analysis in Chapter 2, one could aim to reject the infinitesimal model with linkage in favor of a different model which more accurately captures the distribution of variance contributed to the selected trait along the genome.

The theoretical questions discussed in chapter 3 are confronted with actual experimental data in this chapter - unlike in the Longshanks experiment, discussed in chapter 2, there are many, namely 27, replicates in this case. Here I am focused on allele frequency data analysis: I describe the data and the distribution of allele frequencies before and after selection as well as allele frequency changes conditional on initial frequencies. I also look at the distribution of these changes along the genome and finally describe the drop in heterozygosity on average and along each of four chromosomes. I relate these results to the analyses done in the previous chapters and suggest next steps.

The Drosophila Genome

The *Drosophila melanogaster* genome sequence was first published in 2000 by Adams et al. (2000). It consists of autosomes 2, 3 and 4 and sex chromosomes X and Y. The 4th chromosome is very small and therefore not of much interest. The total genome size is about 180 Mb, with roughly two thirds heterochromatic DNA and one third euchromatic DNA. Almost all euchromatin is on the two large autosomes 2 and 3 and on the X chromosome. The genome carries about 14000 protein-coding genes. Most of the Y-chromosome consists of heterochromatin, and there are no cross-overs in males (Adams et al., 2000).

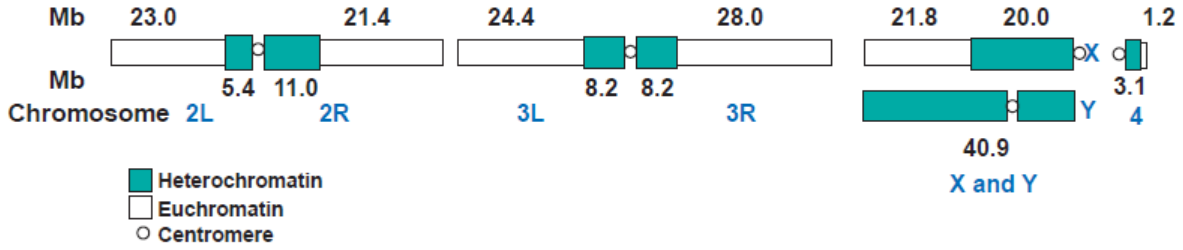

Figure 4.1: Chromosomes of *D. melanogaster*, showing euchromatic regions, heterochromatic regions and centromeres. The autosome arms are called 2L, 2R, 3L, 3R, and 4. The Y chromosome consists almost entirely of heterochromatin (Adams et al., 2000), Fig.1

.

4.2 The Experiment

4.2.1 Description of the Experiment

Selection was, similarly to the Longshanks analysis, within families, meaning that from each breeding pair, the largest male and female individual were used to contribute to the next breeding generation. Again this is important, because selection within families does not distort the pedigree, so it is possible to analyze chromosomes separately and condition on the pedigree when doing simulations. We can assume that selection was only on the trait, apart from a small amount of pairs which did not produce offspring.

There were two phases of the experiment - first a neutral phase, lasting for 5 generations, and then a selective phase, for an additional 6 generations. *Drosophila melanogaster* were mated

Generation	Crosses
G1	[*329* female× *340* male] **family a** and [*330* female× *335* male] **family b**
G2	15 single pair crosses between G1 **family a** and G1 **family b**
G3	81 single pair crosses between randomly selected individuals from different G2 vials
G4	78 single pair crosses between randomly selected individuals from different G3 vials
G5*	88 single pair crosses between randomly selected individuals from different G4 vials. 8 individuals at a time from this generation were chosen to establish 31 lines*
G6	Overall 154 single pair crosses from different G5 vials
G7	139 crosses, selecting the largest available males and females from the previous generation within the groups of 4 vials of a given line (within-family)
G8	141 crosses, selecting the largest available males and females from G7 within the groups of 4 vials of a given line
G9	130 crosses, selecting the largest available males and females from G8 within the groups of 4 vials of a given line
G10	128 crosses, selecting the largest available males and females from G9 within the groups of 4 vials of a given line
G11	149 crosses, selecting the largest available males and female from G10 within the groups of 4 vials of a given line
Later	3 pairs of siblings (one vial per pair) for each line were established using G11 individuals, and these stocks have been maintained until today.

Table 4.1: Breeding scheme in the *Drosophila* selection experiment. First, "family a" and "family b" were established using one female Japanese individual and one male African individual each. These four individuals were all sequenced at high coverage.
*31 times 8 Pupae from generation G5 were chosen based on their size and sex to generate groups of G5 parents drawn from across the phenotype distribution (4 from the middle of the distribution, then 2 short males and 2 long females each). 31 lines were established of which 27 were ultimately successful (Guy Reeves, personal communication).

Stock name	Stock Center	Original Stock Center Stock Number	Collection location	Available vial average pupal length
S-329	EHIME (Watada)	TS48	Fukuoka, Kyusyu, .Japan	3.74mm±0.1 SD(n=162)
S-330	EHIME (Watada)	TS95	Fukuoka, Kyusyu, .Japan	3.56mm±0.12 SD(n=47)
S-335	EHIME	ZI178-Davis	Siavonga, Zambia	2.98mm±0.08 SD(n=98)
S-340	EHIME	UM6-Davis	Masindi, Uganda	3.01mm±0.08 SD(n=43)

Figure 4.2: Founder stocks from which the 4 individuals in G1 were taken: The two individuals from the African stocks showed substantial heterozygosity and were shorter by about half a millimeter than the two individuals from Japanese stocks, which showed lower heterozygosity. Both Y chromosomes are African, both mtDNA Japanese, and 4 of the 6 original X chromosomes are Japanese (Guy Reeves, personal communication).

in single pairs and sib-mating was never allowed, see table 4.1 and figures 4.2 and 4.3. After the neutral phase, in generation G6, 27 replicate lines were established, with a population size of 8 individuals each. The full pedigree was recorded, see also figure 4.3.

Pupal length was chosen as an appropriate trait, since it is highly heritable and its phenotyping can be automated. In fact, in the experiment itself, narrow sense heritability (h^2) ranged from 0.44 to 0.5, and broad sense heritability (H^2) ranged from 0.58 to 0.61, which puts pupal case size among the most heritable traits in Drosophila (Reeves and Tautz, 2017).

In the first 5 generations (G1-G5), parents were selected at random with respect to pupal size. This first, neutral phase of the experiment increased the overall number of individuals, as well as the total number of recombination events. From this larger population size, the replicate lines were established. The founders (G1) consisted of two individuals from African stocks with substantial heterozygosity and two Japanese individuals with much lower heterozygosity, as well as slightly larger pupae 4.2. The four G1 founders were sequenced at high coverage (16-30 fold). No large inversions were found to be segregating in these founder individuals.

Selection started at generation G6. 31 lines, of which 4 lines ended up failing, where established from G5 parents. Lines were established by choosing 8 individuals per line from G5 based on their phenotype, see table 4.1. For each line, 4 pupae from the middle of the size distribution, 2 shorter males and 2 longer females were picked. From this point on, the longest individuals of each family - effectively the longest 4% of individuals overall - were selected in each generation (G6-G11). No sibling crosses were ever allowed. This selection scheme was used in order to reduce inbreeding (compared to random mating). After selection, 3 pairs of siblings, one vial per pair for each line, were established, using G11 individuals. These stocks have been maintained until today.

Selection was effective, with the offspring's mean pupal size increasing by over 0.3 mm in 6

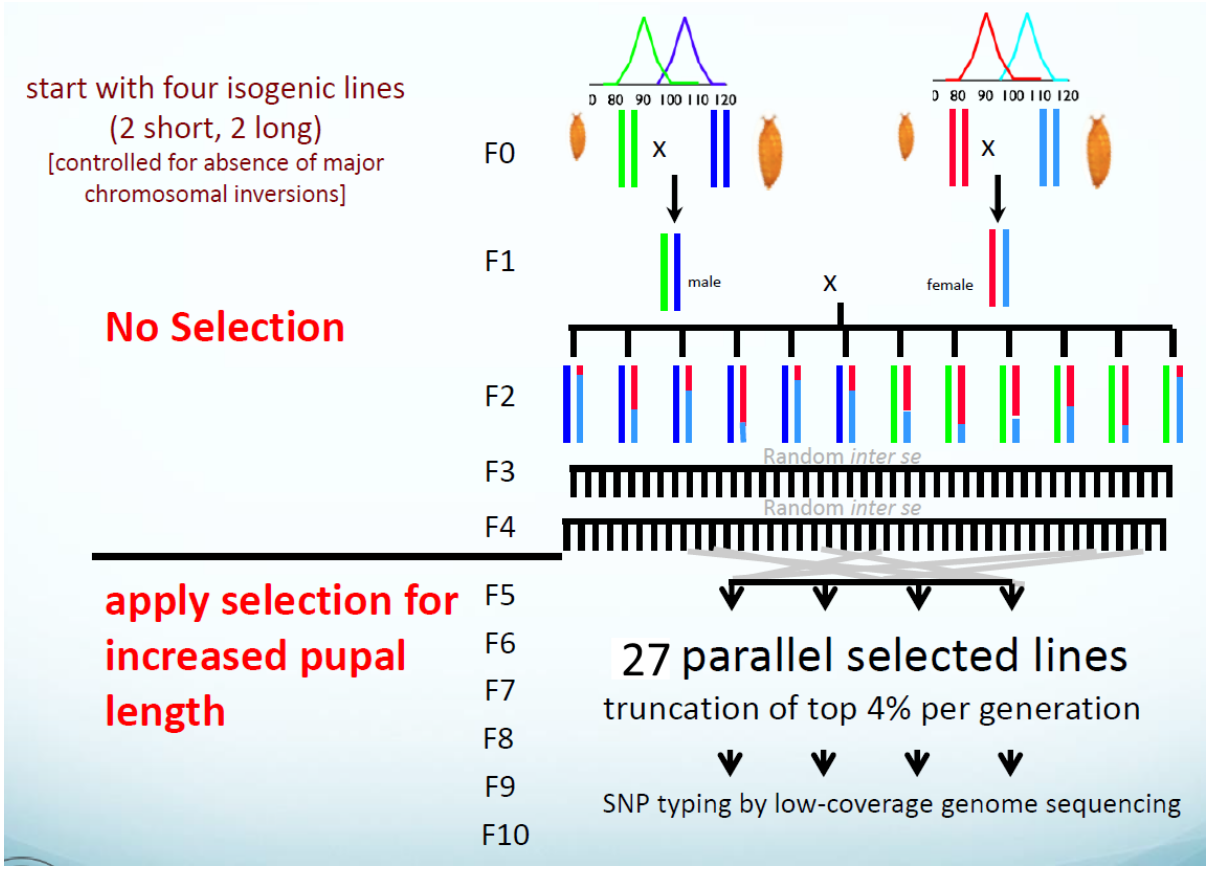

Figure 4.3: Experimental design for *D. melanogaster* selection experiment. 2 African and 2 Japanese individuals from opposing ends of the pupal size spectrum, which were also sequenced at high coverage (16-30 fold), were used to establish populations which went through 5 generations of neutral reproduction and 6 generations of truncation selection, where about the top 4% of individuals were chosen (Guy Reeves, personal communication).

generations (from G5 to G11), which corresponds to 1.62 standard deviations. The typical size range of pupal cases within *D. melanogaster* as a species is 1.1 mm and was 0.8-0.9 mm in this experiment, which is large compared to the smallest size differences which can still be discerned through measurements, namely 0.04 mm (Reeves and Tautz, 2017). Observed broad-sense heritability estimates compared to those expected if a purely additive model is assumed seem to indicate that non-additive genetic effects are modest (Reeves and Tautz, 2017; Mackay, 2014).

For phenotype measurements, pupae were attached to a transparent film and photographed. The pupal case measurements were taken via automatic image analysis. In addition, results in Reeves and Tautz (2017) indicate that pupal case size is a polygenic trait - or at least not controlled by just a few large effect loci, as no SNP of significance at level α=0.05 were identified in a whole genome scan; this was done using recombinant inbred lines (RIL) from a second data set studied in Reeves and Tautz (2017), which was established using the same founders. The scan was for genomic regions associated with mean RIL pupal length (unweighted mean of replicate vial means) with a density of one SNP per 10kb region, excluding mtDNA and the Y chromosome (Reeves and Tautz, 2017).

4.2.2 Description of the Data

Sequencing

Only the 4 G1 founder individuals were sequenced at high coverage (16-30 fold). Most of the breeding populations in generation G6 (198 sequences, a few individuals failed to sequence), as well as the pupae at generation G11 (488), have been sequenced at low coverage of approximately $\times 1$ depth. Imputation - the reconstructing from low coverage sequencing data, sites with missing information and the conversion of sites estimated to be heterozygotes - was attempted at 231640 SNP using the program BEAGLE 4.1. It is possible to estimate the error rate of imputation by looking at the Mendelian error rate in trios (mother, father and offspring). An example of a Mendelian error would be if mother was 01 the father 11 but the child was 00 (rather than the expected 11 or 01). The Mendelian error rate for over 120 trios gave an average value of 0.3%. In this data set, chromosomes 2L, 2R, 3L and 3R are included, the sex chromosomes had yet to be sequenced at the time of analysis and writing.

Pedigree and data files

The full pedigree was recorded over the course of the whole experiment separately for each line (see figure 4.4), as well as sex, family size, genome coverage, and phenotype (pupal length). The sequencing data was collected in a variant call format (vcf) file, which also contains information about coverage. This format allows for storing variations between genomes along with a reference genome. I produced genotype and allele frequency files, as well as lists of individuals in each line and SNP positions using vcftools (Danecek et al., 2011). I then imported the files into Python and analyzed them there.

4.3 Results

In the experiments, alleles were labeled according to a reference genome (Hoskins et al. (2015), accession number GCA_0000012154). Any position carrying the same allele as the reference genome is labeled 0 at the given SNP, and 1 if it carries the alternative allele. We assume that the data are biallelic throughout. Since the Japanese individual's genomes were closer to the reference genome, they carry more 0 alleles. In a preliminary check, our collaborators found that these alleles tended to increase, which is as expected, since the Japanese individuals were chosen for their especially long pupae, and selection was on pupa length. This is a clear indication that selection was acting and we should be able to identify some of its signatures.

In what follows I analyzed data from only 14 out of the 27 replicate lines, because the other lines had missing sequences in the first generation of selection. The full dataset will become available for analysis as more samples are continually being sequenced.

4.3.1 Distribution of Allele Frequencies

In figure 4.5, a histogram of allele frequencies summarizing all lines is shown for generation 5 (blue) and generation 11 (red), the two generations at the beginning and end of selection. The asymmetry in generation five is due to the labeling method as described in the previous section 4.3. Here, I plot the frequency distributions of the "alternative alleles" (with respect to the reference genome), labeled as 1. After six generations of selection, many of these alleles have fixed and even more have been lost, as would be expected if 1 alleles tend to come from

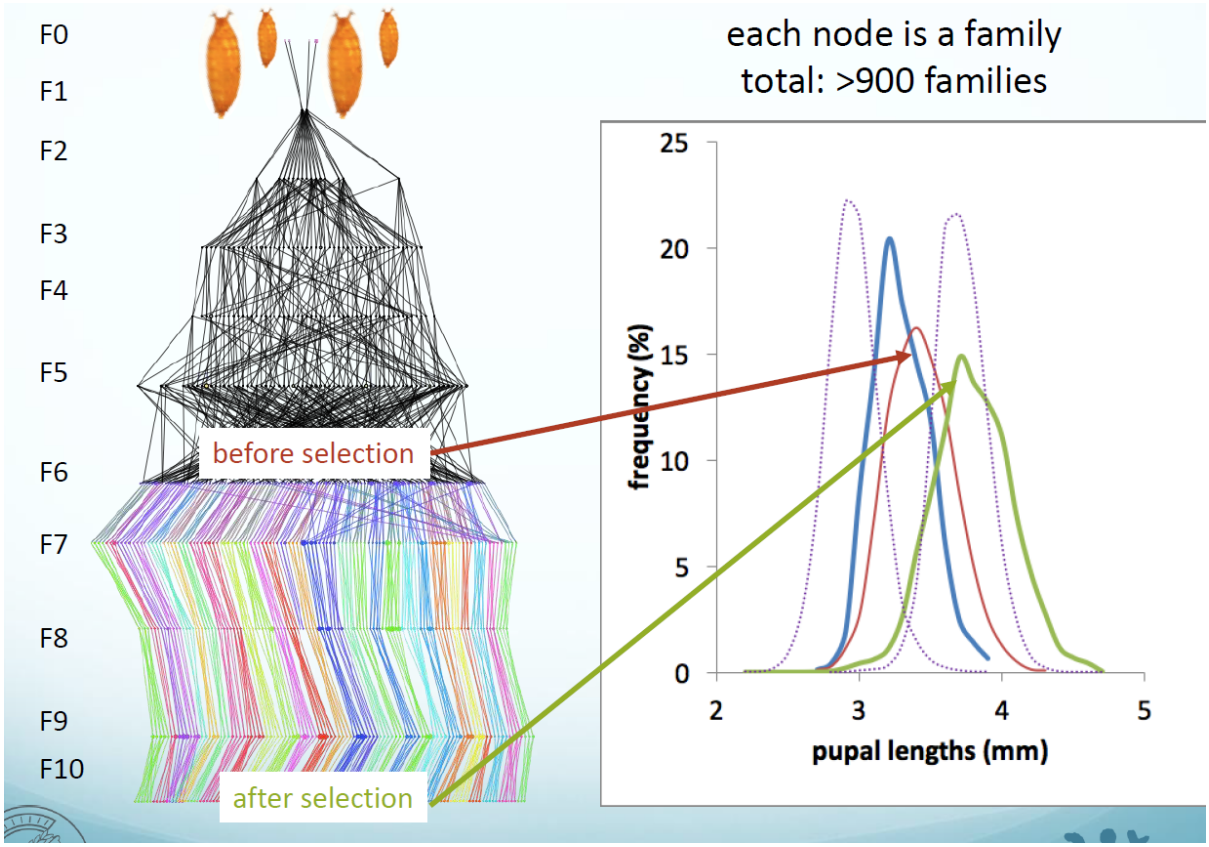

Figure 4.4: Visual representation of the pedigree and increase in the trait value. At F0, 4 founder individuals from 4 different stocks, see figure 4.2 were used to establish a larger population of Drosophila, excluding sib-mating. At generation F5, individuals were chosen for 31 lines on which selection was then applied for 6 generations. The pupal length distributions are shown in the inset. The different colors represent the individual lines, and each node represents a family. The two purple dotted lines represent size distributions from the African and Japanese stocks respectively, the blue line gives the size distribution of the F1, the red curve gives the size distribution in generation of the F6, before selection, and the green curve gives the size distribution after selection, of the F11. The irregular shape is due to the small number of individuals in F1 probably due to uneven sampling for each parallel line and family in F11, so we might not have a Gaussian expectation. Refer also to table 4.1. (Guy Reeves, personal communication).

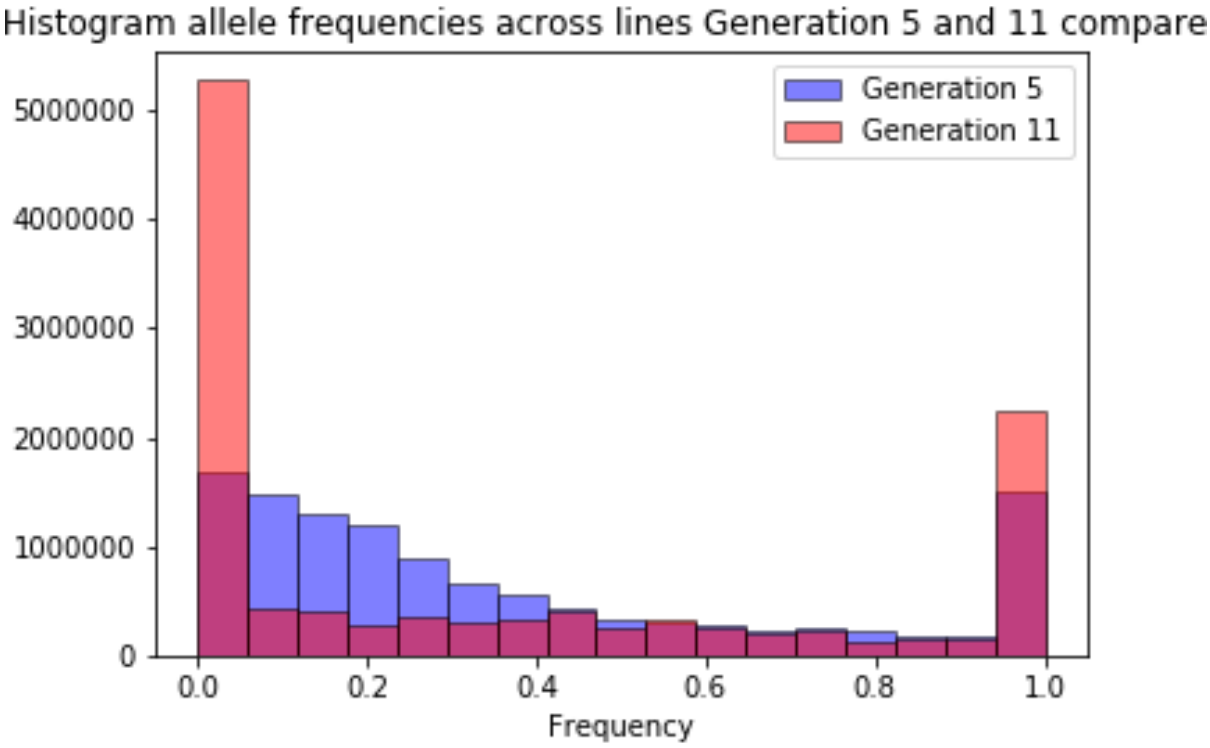

Figure 4.5: Histogram of Allele frequency distributions in generations 5 and 11. Intermediate frequency alleles have become rarer in generation 11, which is as expected. The distributions are asymmetric due to labeling, which was done with respect to a reference genome.

the genomes associated with the initially shorter phenotype, which fits with our knowledge of the founders.

4.3.2 Allele Frequency Changes

Figure 4.6 shows a histogram of the allele frequency changes at all SNPs which were polymorphic in at least one line from generation 5 to generation 11, including all 14 analyzed lines. The peak in the middle is due to loci fixed from the beginning and could be taken out. Some invariant loci are part of the data set because they are polymorphic in at least one line and therefore included, but may be invariant in other lines.

The asymmetric pattern in this plot may have a similar explanation to the asymmetry in figure 4.5; since alternative alleles tend to be associated on average with the shorter pupal case phenotype, they might decline in frequency at a higher rate, leading to the bias towards negative changes here. However, this still needs to be tested to quantify the effect.

As can be seen in figure 4.7, frequency changes varied a lot between lines. We do expect drift to play a big role, since population size was so small. To evaluate these data quantitatively, a more sophisticated analysis will be needed, as described in the discussion section 4.4. Allele frequency changes are also shown for one exemplary line across chromosomes two and three to show the variance between chromosomes (figure 4.8). Chromosomes 2 and 3 are the largest chromosomes - We did not have the data for the X chromosome yet, and chromosome 4 in Drosophila is much smaller than the other ones and has therefore been neglected for now. Note how there are only few distinct horizonal "levels". The likely explanation is the presence

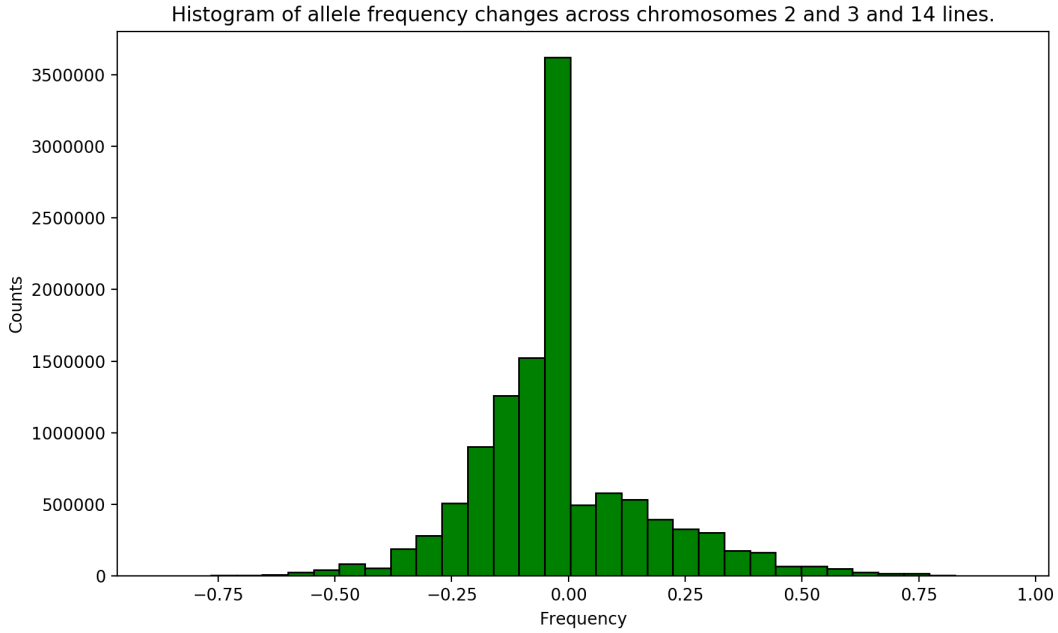

Figure 4.6: Histogram of the allele frequency change from generation 5 to 11 at all SNPs, across all 14 analyzed lines combined. The high peak in the middle is due to loci which are invariant in some lines. Bias towards negative change might be due to the alternative 1 alleles being on average more often associated with the short pupal case phenotype, which is selected against.

of only few haplotypes. When a haplotype has many SNP associated with it, they will all show the same change in frequency when their haplotype block changes frequency. These linked SNP are highly correlated, which reduces the overall amount of information available. In fact, in many regions, there will be fewer than the maximum number of haplotypes (which is 2 times the population size), since some of them will have already been lost in generations G1 to G5. Accordingly, there will be even fewer values by which allele frequency can change in certain regions.

In a few regions, the plot even appears to have only two distinct lines around which points cluster. To illustrate this, imagine that in a certain region we start out with only two haplotypes, one of which starts at frequency p_0 and the other at $1 - p_0$. If there was enough time for it to either fix or be lost by chance, then the only values Δp could take would be p_0 if this haplotype fixes, or q_0 if it is lost.

Rather than looking at the overall changes in allele frequencies it is more informative to condition on initial allele frequency and see where alleles in a certain frequency class end up after the 6 generations of selection. Due to selection, we expect an excess of alleles which started at low frequency and ended up at high frequency compared to the neutral distribution. From this, selection can be inferred via the transition matrix and/or simulations. The transition matrix, which contains the transition probabilities to go from copy number $i \in \{1, ...2n\}$ to copy number $j \in \{1, ...2n\}$ at a single locus, can, as long as transition probabilities remain constant, be raised to power t to obtain the overall expected distribution of allele frequencies after t generations. This can be used to calculate the likelihoods of different selection coefficient values, given the data in terms of allele frequency changes at

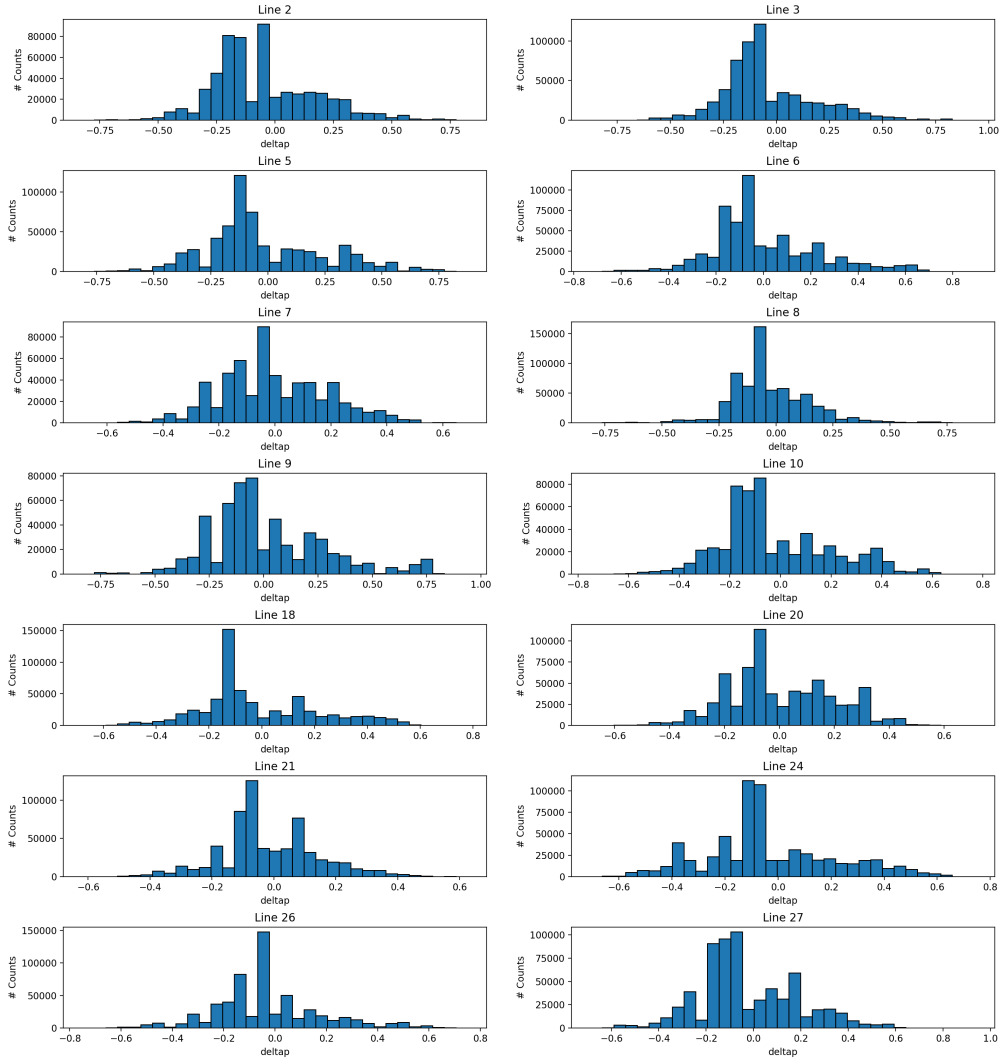

Figure 4.7: Histograms of allele frequency changes at all SNPs for each of the 14 analyzed lines. There is a lot of variation between lines, in line with the expectation of drift playing a large role. Bins correspond to frequency differences of one copy. Neighboring histogram bars sometimes have very uneven sizes. This puzzling, as we expect the probability distribution (which the histogram should approximate) to be smooth. However, connecting to the topics discussed in the previous chapter 3, the reason might be haplotypes structure: when many SNP associated with the same haplotype move in the same way, this could explain the clustering. The result is more noise between replicates.

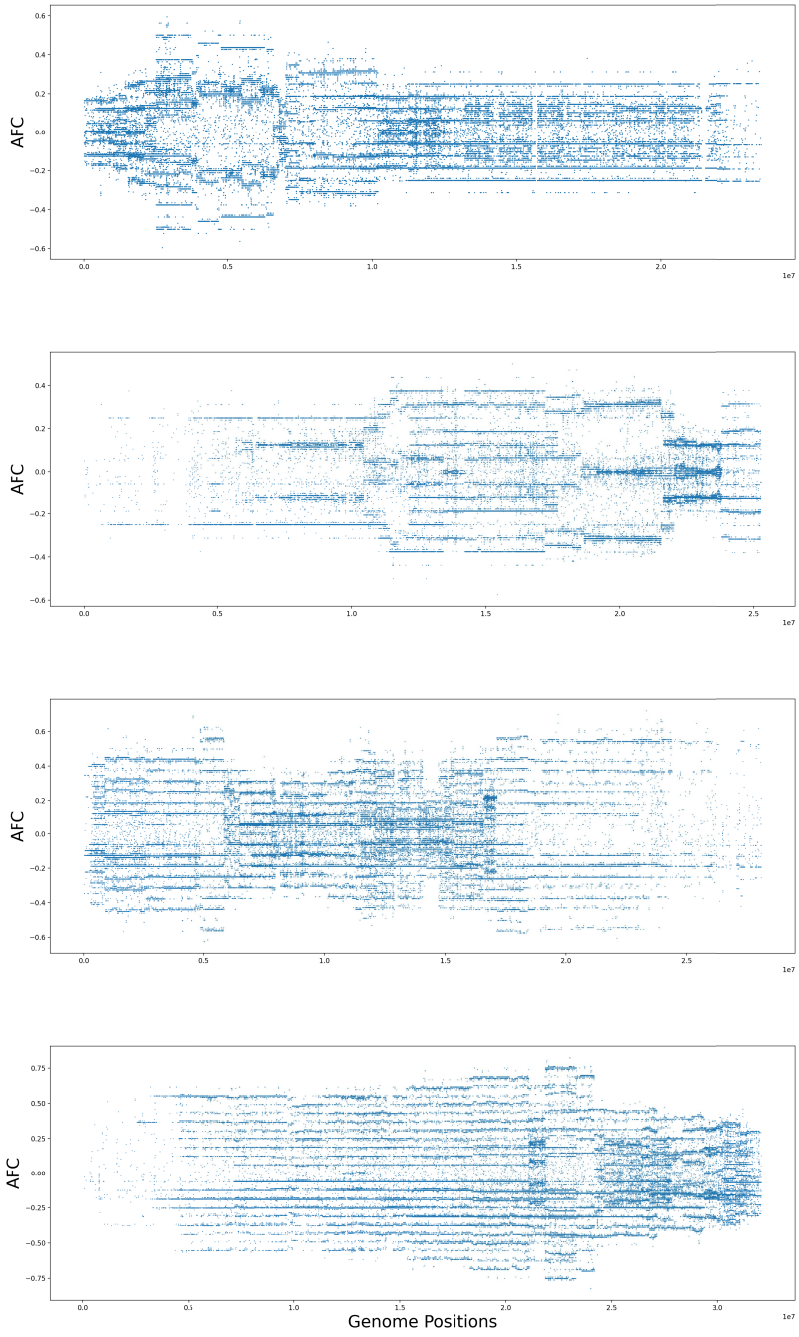

Figure 4.8: Allele frequency changes in line 1 for chromosome arms 2L, 2R, 3L and 3R. Few starting haplotypes lead to a low number of distinct lines in this plot around which data points cluster; if a a specific haplotype has many SNP associated with it, it makes them all move in the same direction when it changes frequency, leading to fewer distinct horizontal "levels". As expected, there is a lot of variance between chromosomes, probably also due to haplotype structure - correlations reduce the overall amount of information available.

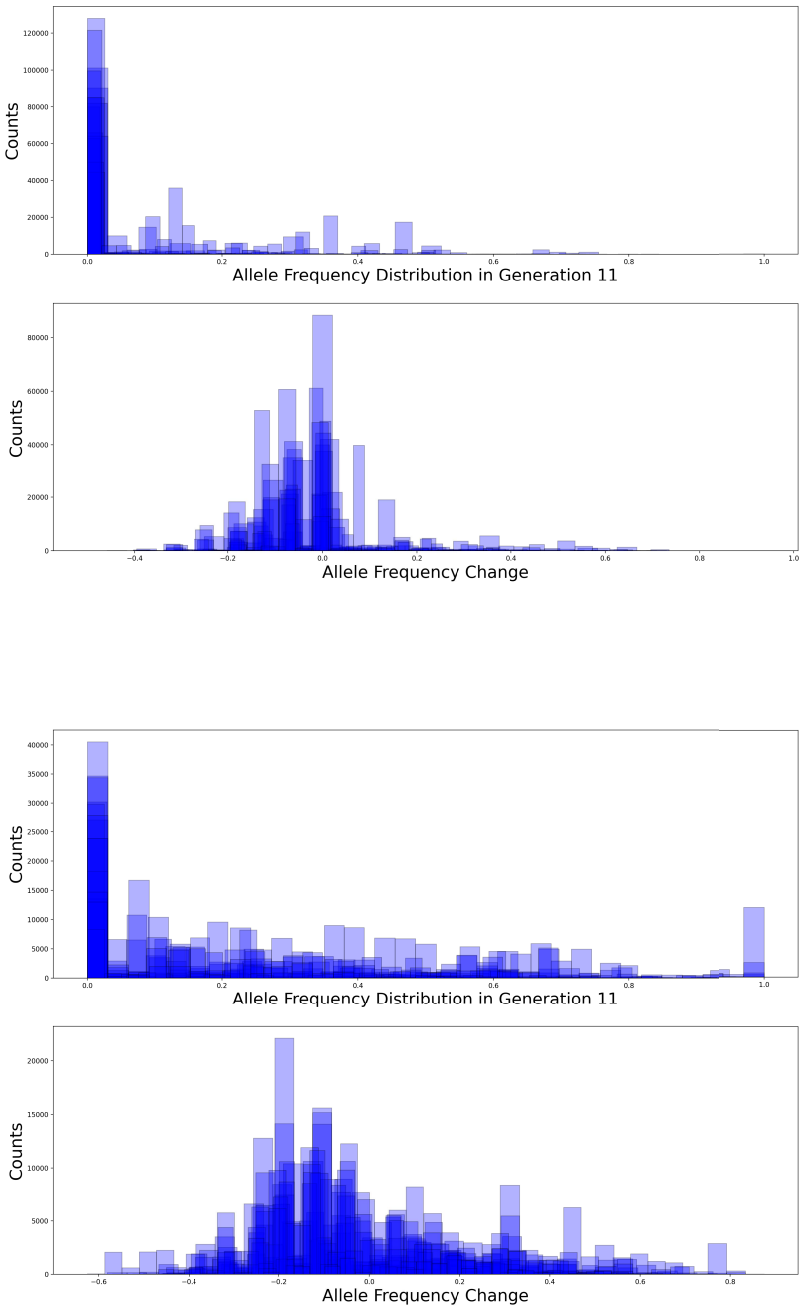

Figure 4.9: This plot shows the conditional histograms of all 14 lines: they show the distribution of allele frequencies at generation 11 for only those alleles that started with frequency between 0 and 0.1 (first plot). In our data this corresponds exactly to starting with one copy of the given allele. The second plot show allele frequency changes from generation 5 to 11 for the same class of alleles that started at frequency < 0.1. The third and fourth row plots show the same conditional distribution, but for the frequency class corresponding to 4 starting copies. Bin size was chosen such that for the maximum number of haplotypes possible (16) each bin corresponds to one specific copy number, even though the scale is in terms of frequencies.

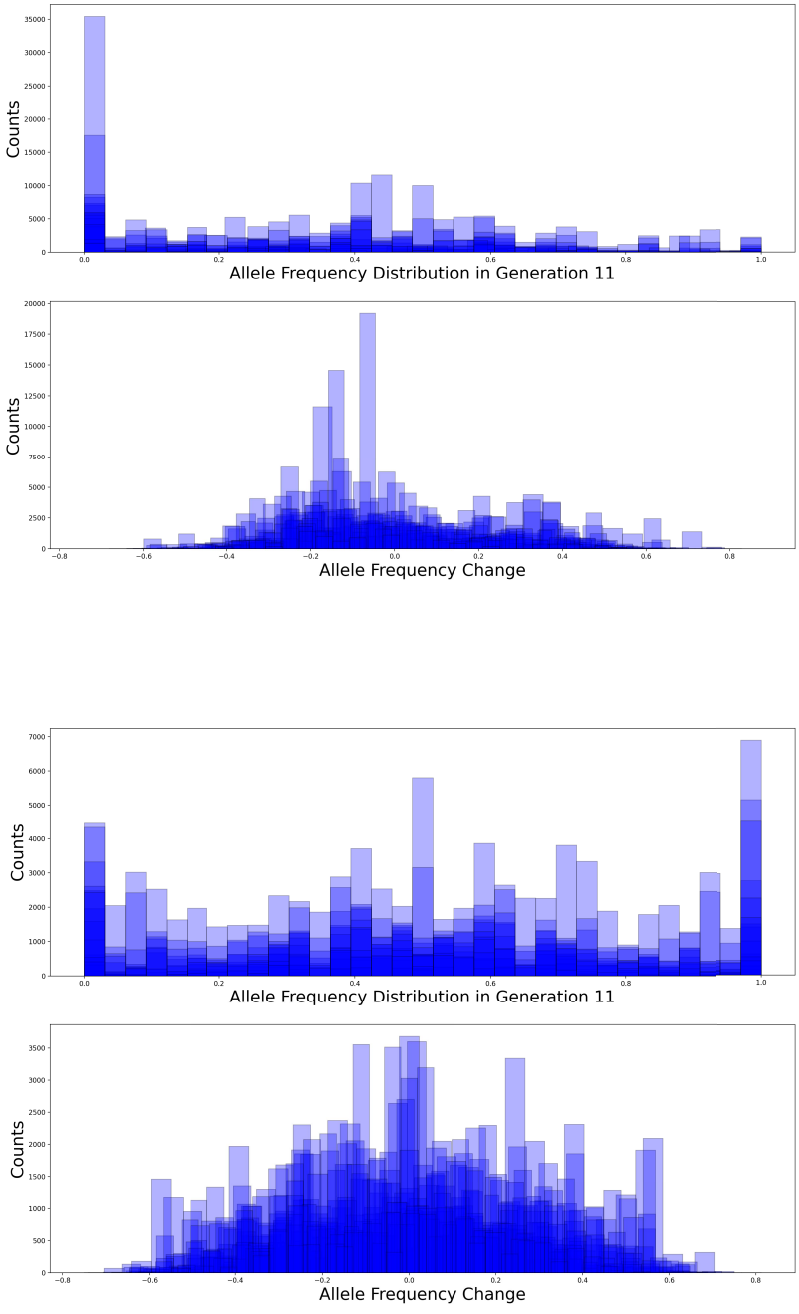

Figure 4.10: This plot shows the conditional histograms of all 14 lines: they show the distribution of allele frequencies at generation 11 for only those alleles that started with 6 initial copies (first row) or 8 initial copies (third row). The second and fourth rows show allele frequency changes from generation 5 to 11 for the same class of allele frequency classes (6 and 8 copies respectively). Bin size was chosen such that for the maximum number of haplotypes possible (16) each bin corresponds to one specific copy number, even though the scale is in terms of frequencies.

individual SNPs. The expected distribution can be calculated via the transition matrix both under neutrality or selection.

Figure 4.9 shows the distribution of allele frequencies at generation 11 for alleles that started out at 1 copy which here corresponds exactly to the frequency <0.1 class. The second row shows the allele frequency changes for these frequency classes. The third and fourth plot in this figure repeat the same, but for alleles which started at 4 initial copies. Figure 4.10 shows the same thing, but for 6 and 8 initial copies. The bulk of alleles that start at one copy get lost, as expected, but we can also see that at some loci they do reach high frequency. In further analyses, I will compare this to the expectation: a transition matrix analysis can tell us the expected neutral distribution, which is independent of LD. This way one can show whether more alleles increased in frequency than expected under neutrality and also determine an estimate for the strength of selection. Similarly it is possible to define a statistic like the mass in the tail of the distribution: the total number of counts in the histogram above some critical value. This could measure the excess in the low-to-high frequency class by comparing the statistic to the expectation found via the transition matrix or simulations conditional on the pedigree. LD, while not changing the expectation, will increase variance between replicates, especially since *Drosophila*, with its few chromosomes has strong linkage.

The lower subplots in figures 4.9 and 4.10 each show the allele frequency change between generations five and eleven for the same class of alleles - 1, 4, 6 or 8 respectively: Again, in the first two cases, more alleles seem to have undergone a negative rather than positive frequency change, which might be due to more of the one alleles being associated with the "shorter pupae" phenotype and getting lost at higher rates.

4.3.3 Heterozygosity and N_e

I calculated the average heterozygosities H_t at generations five and eleven from allele frequencies. Figure 4.11 shows the drop in average heterozygosity $(H_{11} - H_5)$. For a first rough estimate of N_e from these data, I used

$$N_e = \frac{1}{2\left(1 - \sqrt[t]{\frac{H_t}{H_0}}\right)}. \tag{4.1}$$

This gave $N_e = 7$, close to the census population size. However, this estimate is unreliable as it does not consider the specific features of this design, including within-family selection, which was specifically chosen to reduce inbreeding. It is therefore reasonable to assume that N_e will be significantly larger and it can be calculated from the pedigree. To get a better estimate of N_e and its changes over time, it can also be estimated directly from the pedigree, by iteratively working out identity-by-descent and again using a version of 4.1, with $H_t = 1 - F_t$, where F_t is the inbreeding coefficient at generation t, or estimate N_e with a gene dropping simulation. Beyond the expected overall loss in heterozygosity, there will also be substantial variation in heterozygosity along the genome and between replicates. This is true in *D. melanogaster* even more than in mice, since, as mentioned earlier, flies have fewer chromosomes and therefore linkage plays a bigger role. Figure 4.12 shows the average heterozygosity in windows of 100 SNP each along the genome in line 2, again compared between the main autosomes in generations five (top) and eleven (bottom). This shows that overall heterozygosity has gown down in this line, but there is interesting variation along the genome. For example, in the first windows of chromosome 3L, heterozygosity has been completely lost over a whole stretch of DNA. This could be explained by a single block in that region having reached fixation over the 5 generations.

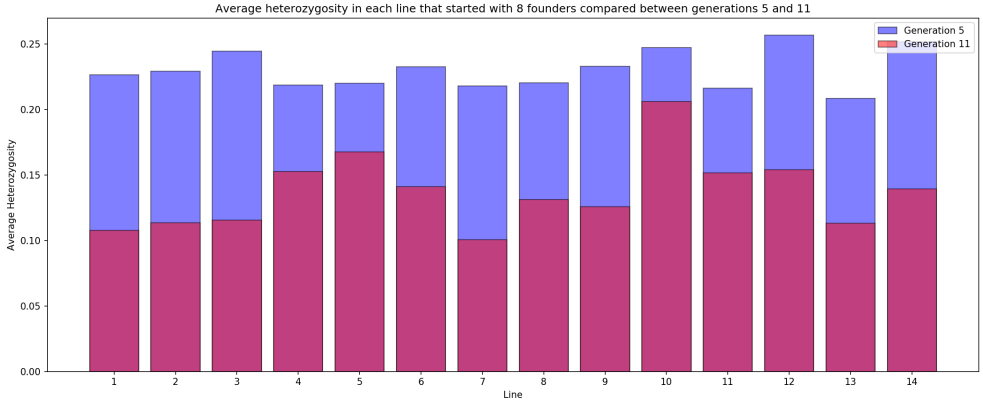

Figure 4.11: Drop in average heterozygosity between generations five and eleven.

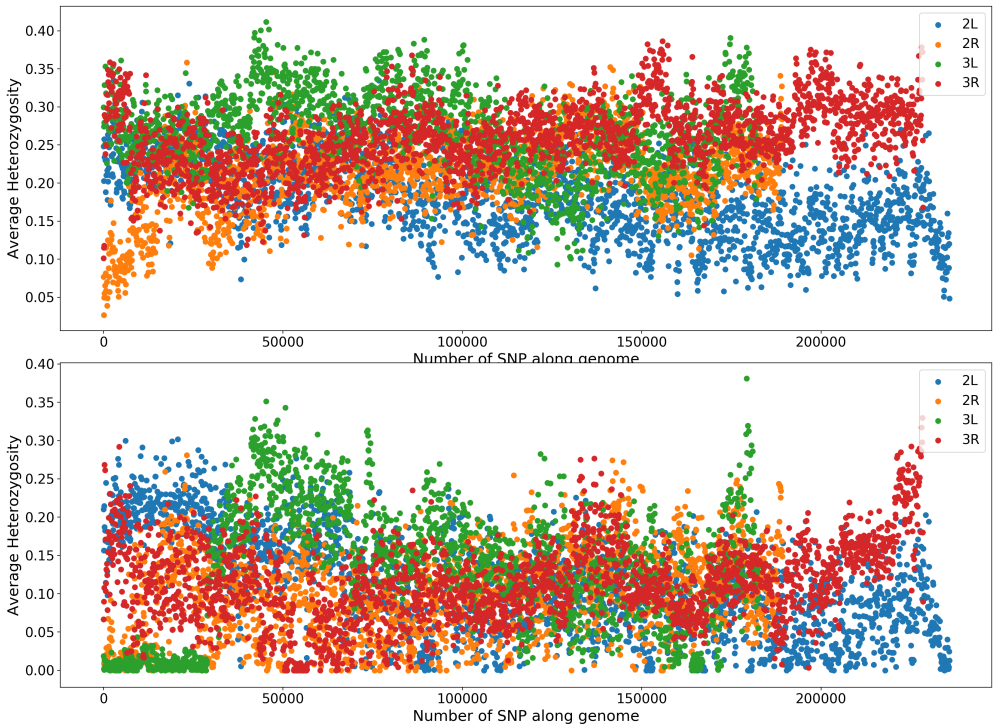

Figure 4.12: Heterozygosity along the genome in windows of 100 SNP each, for Line 2, in generations 5 (top) and 11 (bottom), for chromosomes 2L, 2R, 3L and 3R. Overall heterozygosity has gone down, with interesting variation along the genome: For example, the first windows in chromosome 3L seem to have lost all heterozygosity in this line, whereas on the same chromosome at about 180000 SNP, one window seems to have increased in heterozygosity, which can easily happen by chance, since the system is very stochastic.

Figure 4.13: Heterozygosity in windows of 500 SNP each on chromosome 2L for all analyzed lines in generation 5.

Figure 4.14: Heterozygosity in windows of 500 SNP each on chromosome 2L for all analyzed lines in generation 11.

Figures 4.13 and 4.14 show the heterozygosity in windows of 500 SNP each on chromosome 2L for all the lines I analyzed, comparing generation 5 and 11. As expected, heterozygosity along the genome varies much more in generation 11 than in generation 5, and there is substantial variation between lines. What seems clear from inspection is that there is a very polygenic response - many regions, distributed over most if not all chromosomes, seem to change in response to selection. This is similar to what we saw in the Longshanks data, but needs to be checked systematically.

4.4 Discussion and Outlook

The current data set is special in that it has many replicates. Due to this, I anticipate increased power for inference in this experiment compared to e.g. the Longshanks analysis. However, selection was applied over only a few generations, which is limiting in several regards. For one, when interested in the genetic response along the genome, the smallest observable unit can not be smaller than the average block length, which in turn depends on the recombination rate and its variance. However, during a short selection period, only a very limited number of recombination events can take place. Also, for the very small population size and this short time period, the influence of drift is strong, and it might be hard to pick out alleles which established due to selection, as opposed to chance.

While it is true that selection applied by my collaborators in the current experiment is on pupa length only, one should note that there is always some amount of "natural" selection going on as well. For example, some of the crosses that were set up were infertile. Also, populations might take a number of generations to adapt to the laboratory conditions they were placed in. Therefore, there is a small amount of selection response that can not be accounted for by selection on the trait. Note that in our case, estimating overall selection and learning about the genetic basis of the trait of pupal length are directly related questions, since most selection was on the trait itself (with only a little selection from infertile crosses). The effect on the trait is related to selection strength via the selection gradient.

In order to give a complete and quantitative description of the genetic changes in the population, the next step is to produce the expected neutral distribution. In principle it is possible to calculate the TM from the pedigree directly, by finding the probability that i copies in one generation go to j in the next in every generation. This is very general, as the pedigree can capture any population structure. However, the calculation is very difficult and might not add much insight. I use the Wright-Fisher approximation instead, which assumes random mating. N_e can be calculated from the pedigree and used to determine the dimensions of the transition matrix for a Wright-Fisher process. The transition matrix gives the expected distribution at a single locus under neutrality or selection, conditional on starting frequency. One can then distinguish between changes due to drift or selection or, using simulations, reveal to what extent such a distinction might not be possible.

Possible statistical tests, as mentioned in section 4.3 could for example be for an excess of alleles which went from low to high frequency, using summary statistics to detect outlier windows, similarly to the Longshanks analysis (section 2.3.1). Once phased data is available, one can look at the expected versus actual number of haplotypes in any given region of the genome after selection.

To obtain the proper variance between replicates, linkage needs to be included, which can be done in simulations. The pedigree and correct linkage map for *Drosophila melanogaster* can

both be included, as well as the full information about diploid genotypes, or, once available, haplotypes. Starting with a simple model one can then reject increasingly complex null models step-by-step as alluded to in previous chapters, to find a good trade-off between simplicity and realistic representation of the data. The "hierarchy of models" of increasing complexity could be:

1. Start with a neutral model,
2. continue with an infinitesimal model, assuming a completely homogeneous distribution of variance accounted for by regions of genome contributing to the trait under selection, and using the known strength of selection (top 4% of individuals with the longest pupal cases were selected).
3. and finally deviate more and more from said homogeneous distribution by adding linkage, include varying levels of "spikiness" which could be measured e.g. by variance between windows along the genome.

Going through 4.4, ideally one would test each pair of successive models and ask whether one is significantly more likely than the other. In general, a common criterion for rejecting one model in favor of another is their likelihood ratio - if this ratio is bigger than e^2 to e^3 (depending on ones prior beliefs) the model with smaller likelihood can be rejected in favor of the alternative model (Edwards, 1984). Likelihood needs to be traded against degrees of freedom in order to avoid overfitting; how to do this in each case can be tricky to determine and depends on prior beliefs. A first simple simulation in SLiM (Haller and Messer, 2019) was already implemented by Andrea Mrnjavac during a lab rotation with our group.

In addition, simulations can give the distributions of any chosen statistic under different models. When calculating such statistics for our data, there are 27 replicates for each window, which makes it possible to measure both mean and variance of each statistic to get a signal plus an estimate of its reliability. Here the window size plays an important role - it needs to be small enough to catch the relevant details, but large enough to have many SNP per window in order to have decent statistics. How to choose this parameter is not obvious.

As it remains hard to detect selection even with a lot of data, it will be interesting to compute population genetic observables like F_{IS}, the inbreeding coefficient, or π, the average pairwise nucleotide diversity in windows, using either allele frequency, individual diploid genotype data or haplotype data and see to what extent inference of selection is possible in each case.

Sex chromosomes should of course also be included in the analysis. We will also check whether there was consistent change across lines window by window (using the high replication) and whether lines were independent, or to what extent we can observe parallelism. While allele frequency data alone can yield a lot of insight already, we could go much further using haplotype data, since it is more informative than allele frequency or even diploid genotype data, because it removes noise due to correlations between SNP and gives a fuller picture of the genetic state of a population. Our collaborator Guy Reeves (at the MPI for Evolutionary Biology, Plön, Germany) is working on phasing the data, hopefully allowing us to ask even more advanced questions and build on previous experience. Data from all parents in the pedigree has also been extracted and sequences will be available in the near future. Since phenotype data are available, one can use the comparison between simulations and data to attempt to estimate the so-called "missing heritability", i.e. the amount of heritability explained by loci with an effect too small to detect individually, or due to other effects which are hard to identify (Lee et al.,

2011; Manolio et al., 2009). Another way to view this is as a comparison between estimates of heritability from quantitative genetics versus estimates from individual SNP effects, which is tricky in the context of the infinitesimal model with linkage. The analysis relies on most of the effects being additive, since there is probably not enough power to estimate nonlinear, e.g. epistatic effects, and we would not be able to distinguish them from effects due to unidentified small-effect alleles. It should be insightful to compare results to those found in the Longshanks data analysis and see if we can improve on the baseline set by that analysis.

In this chapter, I presented data from an evolve and resequence experiment in Drosophila, in which 27 small populations of 8 individuals each were selected for longer pupal case size. I described the data on allele frequency distributions and changes over the period of selection, both overall and conditional on starting allele frequency, which give first indications that selection was acting. I analyzed the changes in heterozygosity on chromosomes 2 and 3 and compared them between the 14 analyzed lines. I discussed what can be done next. The high replication in this experiment gives hope that we can explain the genetic basis of pupal case size in detail in the future.

CHAPTER 5

Conclusion

In this thesis, I presented the analysis of sequence data from two evolve and resequence studies, with the goal of learning more about the genetic basis of complex traits. I emphasized the importance of understanding correlations in the data which stem from underlying haplotype structure.

In the first chapter I discussed our analysis of a selection experiment in mice for tibia length. Selection was acting and effective as we could observe an increase in tibia length of 5.27 s.d. and 4.81 s.d. in the two replicates respectively. Adaptation depended completely on standing genetic variation, since the experiment's duration was only 17 generations. The phenotypic response was relatively smooth and sustained. We thus expected a polygenic genetic response of small shifts at many loci, and used the infinitesimal model with linkage as a null model. The selection response was mostly in accordance with the expectations produced by this model, with a few deviations most likely due to larger effect size loci. One locus in particular, Nkx3-2, was functionally validated by our collaborators and we estimated its selection coefficient. In contrast to the phenotypic parallelism we observed in the two experimental lines, not much parallelism was observed genetically, but the largest shifts at individual loci occurred in parallel. Observing the increase in inbreeding, it became clear that there was large variation between chromosomes, probably due to linkage disequilibrium, though the average value was according to the theoretical expectation. While phasing was attempted by our collaborators, results remained unreliable, but phasing might be attempted again with more advanced methods in the future.

From high variability between chromosomes observed in the Longshanks analysis we followed that LD and linkage must have played a large role, an idea that inspired the analysis in chapter 3. Analyzing a window without recombination, I followed frequency changes in haplotypes and alleles and quantified the excess variability added to a summary statistic based on allele frequencies due to correlations between SNP in the founding generation. I also discussed the excess variability added by drift. I showed that working with haplotypes directly should always be the superior way to do inference, since it completely avoids the noise due to LD between founder SNP, though it is still not clear that even then there will be enough statistical power to distinguish selection at individual regions or loci. I illustrated the inference of selection via a maximum likelihood-based scheme on the case of a single locus and laid out some of the next steps.

I then showed first results from the analysis of a second evolve and resequence experiment,

selecting for pupal case length in *Drosophila*. While population size was even smaller, this experiment was highly replicated with 27 independent lines from the same founder population undergoing the same selective process. Again, selection was acting and effective, with an increase in mean pupal case size of about 1.62 s.d. over only 6 generations under selection. I described allele frequency distributions before and after selection and allele frequency changes conditional on initial starting frequency, which should be very informative once they are tested against the expected distribution under neutrality, and then the infinitesimal model with linkage. LD will play an even larger role here than in mice, as we are effectively dealing with only three chromosomes carrying the bulk of euchromatin. Allele frequency changes along the chromosomes showed interesting patterns possibly suggesting the presence of only few segregating haplotypes. Here, more research is needed to confirm or reject this hypothesis. I also showed that the loss in heterozygosity seemed to be somewhat correlated along the genome, further evidence for strong LD. For this experiment, phased haplotype data should soon be available and it will be interesting to see how much further one can go inferring selection once haplotypes are available. This will also help us to quantify the sources of noise due to drift or LD more precisely.

Both experiments shared the special feature of within-family selection, in which the son and daughter with the highest trait values from each breeding family are chosen to contribute to the next breeding generation. This has the perk that this type of selection does not distort the pedigree and so we can look at one chromosome at a time, and simulate conditional on the pedigree, to get more accurate expectations for the selection response.

Since parallelism gives more convincing evidence of selection than large changes in summary statistics alone, one conclusion for the experimental design is that replication is very important. To what extent it can make up for small population sizes and a short selective period, features also shared by both studies though to different extent, remains to be seen. Either way, we saw that even with an abundance of data given, it remains hard to distinguish selected variants individually.

Future research should focus on haplotype data, which gives more direct insight into the LD structure of populations, and a much fuller picture of their genetic state. This is very topical, as haplotype data has become important for imputing variants, inferring selection more accurately and allowing insight into the genetic basis of complex traits, among others (Browning and Browning, 2011). While obtaining these data is expensive and challenging both experimentally and computationally, costs are coming down as methods like haplotagging (Meier et al., 2021) are being developed, and it will be increasingly feasible to phase many samples quickly at acceptable cost (Browning and Browning, 2011). In addition, working with simulations can help us to get at the inherent limits to inference, and determine how much power there is, even in principle, to distinguish distributions of small effect loci.

Another point worthy of discussion is the choice of the infinitesimal model with linkage as the appropriate null model. In principle any educated guess which does not contradict the data and incorporates the known facts could be a valid starting point. For example, one might choose a model including background selection or demographic history of the population. If these forces were known to present in the given data set, not including them might actually distort results. Both examples do not apply in our case though, as background selection is likely to be unimportant over so few generations and no demographic model is needed since the full pedigree is available. However, we do know that linkage is present and that all chromosomes show a selection response, pointing to a very polygenic basis of the trait, so the infinitesimal model with linkage seems to be an appropriate starting point, see also

section 2.3.1. Starting with this model, we can then work to incorporate other elements in an increasingly complex series of models to be tested. We can then use the likelihood criterion to test them. Starting with a completely homogeneous distribution of variance constitutes one extreme on a spectrum of possible genetic architectures, the opposite of which would be the case of one Mendelian locus accounting for all the variance in the trait. Starting from the infinitesimal side, we can carefully approach the other extreme by investigating different, and more heterogeneous, distributions of variance.

There are many avenues for future research, as understanding sequence data is challenging and important, especially in the context of recent and rapid adaptation. We rely on it in the context of medical research, plant and animal breeding, biodiversity and conservation, adaptation to climate change and many more - therefore these topics will continue to be relevant.

Bibliography

G. R. Abecasis, E. Noguchi, A. Heinzmann, J. A. Traherne, S. Bhattacharyya, N. I. Leaves, G. G. Anderson, Y. Zhang, N. J. Lench, A. Carey, L. R. Cardon, M. F. Moffatt, and W. O. C. Cookson. Extent and Distribution of Linkage Disequilibrium in Three Genomic Regions. *The American Journal of Human Genetics*, 68(1):191–197, Jan. 2001. ISSN 0002-9297. doi: 10.1086/316944. URL https://www.sciencedirect.com/science/article/pii/S0002929707624835.

M. D. Adams, S. E. Celniker, R. A. Holt, C. A. Evans, J. D. Gocayne, P. G. Amanatides, S. E. Scherer, P. W. Li, R. A. Hoskins, R. F. Galle, R. A. George, S. E. Lewis, S. Richards, M. Ashburner, S. N. Henderson, G. G. Sutton, J. R. Wortman, M. D. Yandell, Q. Zhang, L. X. Chen, R. C. Brandon, Y. H. Rogers, R. G. Blazej, M. Champe, B. D. Pfeiffer, K. H. Wan, C. Doyle, E. G. Baxter, G. Helt, C. R. Nelson, G. L. Gabor, J. F. Abril, A. Agbayani, H. J. An, C. Andrews-Pfannkoch, D. Baldwin, R. M. Ballew, A. Basu, J. Baxendale, L. Bayraktaroglu, E. M. Beasley, K. Y. Beeson, P. V. Benos, B. P. Berman, D. Bhandari, S. Bolshakov, D. Borkova, M. R. Botchan, J. Bouck, P. Brokstein, P. Brottier, K. C. Burtis, D. A. Busam, H. Butler, E. Cadieu, A. Center, I. Chandra, J. M. Cherry, S. Cawley, C. Dahlke, L. B. Davenport, P. Davies, B. de Pablos, A. Delcher, Z. Deng, A. D. Mays, I. Dew, S. M. Dietz, K. Dodson, L. E. Doup, M. Downes, S. Dugan-Rocha, B. C. Dunkov, P. Dunn, K. J. Durbin, C. C. Evangelista, C. Ferraz, S. Ferriera, W. Fleischmann, C. Fosler, A. E. Gabrielian, N. S. Garg, W. M. Gelbart, K. Glasser, A. Glodek, F. Gong, J. H. Gorrell, Z. Gu, P. Guan, M. Harris, N. L. Harris, D. Harvey, T. J. Heiman, J. R. Hernandez, J. Houck, D. Hostin, K. A. Houston, T. J. Howland, M. H. Wei, C. Ibegwam, M. Jalali, F. Kalush, G. H. Karpen, Z. Ke, J. A. Kennison, K. A. Ketchum, B. E. Kimmel, C. D. Kodira, C. Kraft, S. Kravitz, D. Kulp, Z. Lai, P. Lasko, Y. Lei, A. A. Levitsky, J. Li, Z. Li, Y. Liang, X. Lin, X. Liu, B. Mattei, T. C. McIntosh, M. P. McLeod, D. McPherson, G. Merkulov, N. V. Milshina, C. Mobarry, J. Morris, A. Moshrefi, S. M. Mount, M. Moy, B. Murphy, L. Murphy, D. M. Muzny, D. L. Nelson, D. R. Nelson, K. A. Nelson, K. Nixon, D. R. Nusskern, J. M. Pacleb, M. Palazzolo, G. S. Pittman, S. Pan, J. Pollard, V. Puri, M. G. Reese, K. Reinert, K. Remington, R. D. Saunders, F. Scheeler, H. Shen, B. C. Shue, I. Sidén-Kiamos, M. Simpson, M. P. Skupski, T. Smith, E. Spier, A. C. Spradling, M. Stapleton, R. Strong, E. Sun, R. Svirskas, C. Tector, R. Turner, E. Venter, A. H. Wang, X. Wang, Z. Y. Wang, D. A. Wassarman, G. M. Weinstock, J. Weissenbach, S. M. Williams, n. WoodageT, K. C. Worley, D. Wu, S. Yang, Q. A. Yao, J. Ye, R. F. Yeh, J. S. Zaveri, M. Zhan, G. Zhang, Q. Zhao, L. Zheng, X. H. Zheng, F. N. Zhong, W. Zhong, X. Zhou, S. Zhu, X. Zhu, H. O. Smith, R. A. Gibbs, E. W. Myers, G. M. Rubin, and J. C. Venter. The genome sequence of Drosophila melanogaster. *Science (New York, N.Y.)*, 287(5461): 2185–2195, Mar. 2000. ISSN 0036-8075. doi: 10.1126/science.287.5461.2185.

M. Bamshad and S. P. Wooding. Signatures of natural selection in the human genome. *Nature Reviews Genetics*, 4(2):99–110, Feb. 2003. ISSN 1471-0064. doi: 10.1038/nrg999. URL

https://www.nature.com/articles/nrg999. Number: 2 Publisher: Nature Publishing Group.

R. D. H. Barrett and D. Schluter. Adaptation from standing genetic variation. *Trends in Ecology & Evolution*, 23(1):38–44, Jan. 2008. ISSN 0169-5347. doi: 10.1016/j.tree.2007.09.008.

N. Barton, J. Hermisson, and M. Nordborg. Why structure matters. *eLife*, 8:e45380, Mar. 2019. ISSN 2050-084X. doi: 10.7554/eLife.45380. URL https://doi.org/10.7554/eLife.45380. Publisher: eLife Sciences Publications, Ltd.

N. H. Barton, A. M. Etheridge, and A. Véber. The infinitesimal model: Definition, derivation, and implications. *Theoretical Population Biology*, 118:50–73, Dec. 2017. ISSN 0040-5809. doi: 10.1016/j.tpb.2017.06.001. URL https://www.sciencedirect.com/science/article/pii/S0040580917300886.

J. J. Berg and G. Coop. A Population Genetic Signal of Polygenic Adaptation. *PLOS Genetics*, 10(8):e1004412, Aug. 2014. ISSN 1553-7404. doi: 10.1371/journal.pgen.1004412. URL https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1004412. Publisher: Public Library of Science.

S. Brotherstone and M. Goddard. Artificial selection and maintenance of genetic variance in the global dairy cow population. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1459):1479–1488, July 2005. doi: 10.1098/rstb.2005.1668. URL https://royalsocietypublishing.org/doi/full/10.1098/rstb.2005.1668. Publisher: Royal Society.

B. L. Browning and S. R. Browning. Genotype Imputation with Millions of Reference Samples. *The American Journal of Human Genetics*, 98(1):116–126, Jan. 2016. ISSN 0002-9297, 1537-6605. doi: 10.1016/j.ajhg.2015.11.020. URL https://www.cell.com/ajhg/abstract/S0002-9297(15)00491-7. Publisher: Elsevier.

S. R. Browning and B. L. Browning. Haplotype phasing: existing methods and new developments. *Nature Reviews Genetics*, 12(10):703–714, Oct. 2011. ISSN 1471-0064. doi: 10.1038/nrg3054. URL https://www.nature.com/articles/nrg3054.

M. G. Bulmer. Linkage disequilibrium and genetic variability. *Genetics Research*, 23(3):281–289, June 1974. ISSN 1469-5073, 0016-6723. doi: 10.1017/S0016672300014920. URL https://www.cambridge.org/core/journals/genetics-research/article/linkage-disequilibrium-and-genetic-variability/79AC7FDECBA9A75E78795473553B2D52. Publisher: Cambridge University Press.

M. G. Bulmer. Mathematical theory of quantitative genetics. *Oxford University Press, Oxford*, 1980.

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, Sept. 2010. ISSN 1476-4687. doi: 10.1038/nature09352. URL https://www.nature.com/articles/nature09352.

P. J. Byard. Pedigree analysis in human genetics. By E. A. Thompson. Baltimore: The Johns Hopkins University Press. 1986. xix + 213 pp., figures, tables, appendices, references, indices. $35.00 (cloth). *American Journal of Physical Anthropology*, 71(3):381–381, 1986. ISSN

1096-8644. doi: 10.1002/ajpa.1330710314. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/ajpa.1330710314.

V. Careau, M. E. Wolak, P. A. Carter, and T. Garland Jr. Limits to Behavioral Evolution: The Quantitative Genetics of a Complex Trait Under Directional Selection. *Evolution*, 67(11):3102–3119, 2013. ISSN 1558-5646. doi: 10.1111/evo.12200. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/evo.12200. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/evo.12200.

J. P. Castro, M. N. Yancoskie, M. Marchini, S. Belohlavy, L. Hiramatsu, M. Kučka, W. H. Beluch, R. Naumann, I. Skuplik, J. Cobb, N. H. Barton, C. Rolian, and Y. F. Chan. An integrative genomic analysis of the Longshanks selection experiment for longer limbs in mice. *eLife*, 8:e42014, June 2019. ISSN 2050-084X. doi: 10.7554/eLife.42014. URL https://doi.org/10.7554/eLife.42014.

Y. F. Chan, M. E. Marks, F. C. Jones, J. Guadalupe Villarreal, M. D. Shapiro, S. D. Brady, A. M. Southwick, D. M. Absher, J. Grimwood, J. Schmutz, R. M. Myers, D. Petrov, B. Jónsson, D. Schluter, M. A. Bell, and D. M. Kingsley. Adaptive Evolution of Pelvic Reduction in Sticklebacks by Recurrent Deletion of a Pitx1 Enhancer. *Science*, Jan. 2010. doi: 10.1126/science.1182213. URL https://www.science.org/doi/abs/10.1126/science.1182213. Publisher: American Association for the Advancement of Science.

Y. F. Chan, F. C. Jones, E. McConnell, J. Bryk, L. Bünger, and D. Tautz. Parallel selection mapping using artificially selected mice reveals body weight control loci. *Current biology: CB*, 22(9):794–800, May 2012. ISSN 1879-0445. doi: 10.1016/j.cub.2012.03.011.

N. H. Chapman and E. A. Thompson. A model for the length of tracts of identity by descent in finite random mating populations. *Theoretical Population Biology*, 64(2):141–150, Sept. 2003. ISSN 0040-5809. doi: 10.1016/S0040-5809(03)00071-6. URL https://www.sciencedirect.com/science/article/pii/S0040580903000716.

N. Chen, I. Juric, E. J. Cosgrove, R. Bowman, J. W. Fitzpatrick, S. J. Schoech, A. G. Clark, and G. Coop. Allele frequency dynamics in a pedigreed natural population. *Proceedings of the National Academy of Sciences*, 116(6):2158–2164, Feb. 2019. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1813852116. URL https://www.pnas.org/content/116/6/2158. Publisher: National Academy of Sciences Section: PNAS Plus.

A. Cox, C. L. Ackert-Bicknell, B. L. Dumont, Y. Ding, J. T. Bell, G. A. Brockmann, J. E. Wergedal, C. Bult, B. Paigen, J. Flint, S.-W. Tsaih, G. A. Churchill, and K. W. Broman. A New Standard Genetic Map for the Laboratory Mouse. *Genetics*, 182(4):1335–1344, Aug. 2009. ISSN 0016-6731. doi: 10.1534/genetics.109.105486. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2728870/.

P. Danecek, A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, R. Durbin, and 1000 Genomes Project Analysis Group. The variant call format and VCFtools. *Bioinformatics*, 27(15):2156–2158, Aug. 2011. ISSN 1367-4803. doi: 10.1093/bioinformatics/btr330. URL https://doi.org/10.1093/bioinformatics/btr330.

B. E. Deagle, F. C. Jones, Y. F. Chan, D. M. Absher, D. M. Kingsley, and T. E. Reimchen. Population genomics of parallel phenotypic evolution in stickleback across stream–lake ecological transitions. *Proceedings of the Royal Society B: Biological Sciences*, 279(1732):1277–1286, Apr. 2012. doi: 10.1098/rspb.2011.1552. URL https://royalsocietypublishing.org/doi/full/10.1098/rspb.2011.1552. Publisher: Royal Society.

A. M. Dunning, F. Durocher, C. S. Healey, M. D. Teare, S. E. McBride, F. Carlomagno, C.-F. Xu, E. Dawson, S. Rhodes, S. Ueda, E. Lai, R. N. Luben, E. J. Van Rensburg, A. Mannermaa, V. Kataja, G. Rennart, I. Dunham, I. Purvis, D. Easton, and B. A. J. Ponder. The Extent of Linkage Disequilibrium in Four Populations with Distinct Demographic Histories. *The American Journal of Human Genetics*, 67(6):1544–1554, Dec. 2000. ISSN 0002-9297. doi: 10.1086/316906. URL https://www.sciencedirect.com/science/article/pii/S0002929707632224.

A. W. F. Edwards. *Likelihood*. CUP Archive, 1984.

K. R. Elmer and A. Meyer. Adaptation in the age of ecological genomics: insights from parallelism and convergence. *Trends in Ecology & Evolution*, 26(6):298–306, June 2011. ISSN 0169-5347. doi: 10.1016/j.tree.2011.02.008. URL https://www.sciencedirect.com/science/article/pii/S0169534711000528.

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, Aug. 2016. ISSN 1553-7404. doi: 10.1371/journal.pgen.1006130. URL https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1006130. Publisher: Public Library of Science.

B. Epstein, M. Jones, R. Hamede, S. Hendricks, H. McCallum, E. P. Murchison, B. Schönfeld, C. Wiench, P. Hohenlohe, and A. Storfer. Rapid evolutionary response to a transmissible cancer in Tasmanian devils. *Nature Communications*, 7(1):12684, Aug. 2016. ISSN 2041-1723. doi: 10.1038/ncomms12684. URL https://www.nature.com/articles/ncomms12684. Number: 1 Publisher: Nature Publishing Group.

L. Excoffier and M. Slatkin. Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Molecular Biology and Evolution*, 12(5):921–927, July 1995. ISSN 0737-4038. doi: 10.1093/oxfordjournals.molbev.a040269. URL https://doi.org/10.1093/oxfordjournals.molbev.a040269.

D. S. Falconer and T. F. C. Mackay. Introduction to quantitative genetics. Longman, New York. *Introduction to quantitative genetics. 2nd ed. Longman, New York.*, 1981.

R. A. Fisher. A fuller theory of "Junctions" in inbreeding. *Heredity*, 8(2):187–197, Aug. 1954. ISSN 1365-2540. doi: 10.1038/hdy.1954.17. URL https://www.nature.com/articles/hdy195417. Number: 2 Publisher: Nature Publishing Group.

T. Garland and M. R. Rose, editors. *Experimental Evolution: Concepts, Methods, and Applications of Selection Experiments*. University of California Press, Dec. 2009. ISBN 978-0-520-26180-8.

P. R. Grant and B. R. Grant. Demography and the Genetically Effective sizes of Two Populations of Darwin's Finches. *Ecology*, 73(3):766–784, 1992. ISSN 1939-9170.

doi: 10.2307/1940156. URL https://onlinelibrary.wiley.com/doi/abs/10.2307/1940156. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.2307/1940156.

J. B. Haldane. *The Causes of Evolution*. Princeton University Press, 1932. ISBN 978-0-691-02442-4. Google-Books-ID: JupUCPvgO6AC.

B. C. Haller and P. W. Messer. SLiM 3: Forward genetic simulations beyond the Wright–Fisher model. *Molecular Biology and Evolution*, 36(3):632–637, 2019.

H. A. Hejase, N. Dukler, and A. Siepel. From Summary Statistics to Gene Trees: Methods for Inferring Positive Selection. *Trends in Genetics*, 0(0), Jan. 2020. ISSN 0168-9525. doi: 10.1016/j.tig.2019.12.008. URL https://www.cell.com/trends/genetics/abstract/S0168-9525(19)30269-0.

A. P. Hendry and M. T. Kinnison. Perspective: The Pace of Modern Life: Measuring Rates of Contemporary Microevolution. *Evolution*, 53(6):1637–1653, 1999. ISSN 1558-5646. doi: 10.1111/j.1558-5646.1999.tb04550.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1558-5646.1999.tb04550.x. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1558-5646.1999.tb04550.x.

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. ISSN 2041-210X. doi: 10.1111/2041-210X.12808. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.12808. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/2041-210X.12808.

W. G. Hill. Predictions of response to artificial selection from new mutations. *Genetics Research*, 40(3):255–278, Dec. 1982. ISSN 1469-5073, 0016-6723. doi: 10.1017/S0016672300019145. URL https://www.cambridge.org/core/journals/genetics-research/article/predictions-of-response-to-artificial-selection-from-new-mutations/6DF4FC2EFE63E1AC49768C953F3E40DE. Publisher: Cambridge University Press.

W. G. Hill. Understanding and Using Quantitative Genetics. *Philosophical Transactions of the Royal Society B*, 365:73–85, 2010.

W. G. Hill and M. Kirkpatrick. What Animal Breeding Has Taught Us about Evolution. *Annual Review of Ecology, Evolution, and Systematics*, 41(1):1–19, 2010. doi: 10.1146/annurev-ecolsys-102209-144728. URL https://doi.org/10.1146/annurev-ecolsys-102209-144728. _eprint: https://doi.org/10.1146/annurev-ecolsys-102209-144728.

R. A. Hoskins, J. W. Carlson, K. H. Wan, S. Park, I. Mendez, S. E. Galle, B. W. Booth, B. D. Pfeiffer, R. A. George, R. Svirskas, M. Krzywinski, J. Schein, M. C. Accardo, E. Damia, G. Messina, M. Méndez-Lago, B. de Pablos, O. V. Demakova, E. N. Andreyeva, L. V. Boldyreva, M. Marra, A. B. Carvalho, P. Dimitri, A. Villasante, I. F. Zhimulev, G. M. Rubin, G. H. Karpen, and S. E. Celniker. The Release 6 reference sequence of the Drosophila melanogaster genome. *Genome Research*, 25(3):445–458, Mar. 2015. ISSN 1549-5469. doi: 10.1101/gr.185579.114.

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, Oct. 2015. ISSN 0737-4038. doi: 10.1093/molbev/msv136. URL https://doi.org/10.1093/molbev/msv136.

F. C. Jones, M. G. Grabherr, Y. F. Chan, P. Russell, E. Mauceli, J. Johnson, R. Swofford, M. Pirun, M. C. Zody, S. White, E. Birney, S. Searle, J. Schmutz, J. Grimwood, M. C. Dickson, R. M. Myers, C. T. Miller, B. R. Summers, A. K. Knecht, S. D. Brady, H. Zhang, A. A. Pollen, T. Howes, C. Amemiya, E. S. Lander, F. Di Palma, K. Lindblad-Toh, and D. M. Kingsley. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, 484(7392):55–61, Apr. 2012. ISSN 1476-4687. doi: 10.1038/nature10944. URL https://www.nature.com/articles/nature10944. Number: 7392 Publisher: Nature Publishing Group.

H. M. Kang, J. H. Sul, S. K. Service, N. A. Zaitlen, S.-y. Kong, N. B. Freimer, C. Sabatti, and E. Eskin. Variance component model to account for sample structure in genome-wide association studies. *Nature Genetics*, 42(4):348–354, Apr. 2010. ISSN 1546-1718. doi: 10.1038/ng.548. URL https://www.nature.com/articles/ng.548. Number: 4 Publisher: Nature Publishing Group.

P. D. Keightley, T. Hardge, L. May, and G. Bulfield. A Genetic Map of Quantitative Trait Loci for Body Weight in the Mouse. *Genetics*, 142(1):227–235, Jan. 1996. ISSN 1943-2631. doi: 10.1093/genetics/142.1.227. URL https://doi.org/10.1093/genetics/142.1.227.

J. K. Kelly and K. A. Hughes. An examination of the evolve-and-resequence method using Drosophila simulans. Technical report, bioRxiv, Oct. 2018. URL https://www.biorxiv.org/content/10.1101/337188v2. Section: New Results Type: article.

L. Kruglyak. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nature Genetics*, 22(2):139–144, June 1999. ISSN 1546-1718. doi: 10.1038/9642. URL https://www.nature.com/articles/ng0699_139. Number: 2 Publisher: Nature Publishing Group.

C. C. Laurie, S. D. Chasalow, J. R. LeDeaux, R. McCarroll, D. Bush, B. Hauge, C. Lai, D. Clark, T. R. Rocheford, and J. W. Dudley. The Genetic Architecture of Response to Long-Term Artificial Selection for Oil Concentration in the Maize Kernel. *Genetics*, 168 (4):2141–2155, Dec. 2004. ISSN 0016-6731. doi: 10.1534/genetics.104.029686. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1448749/.

S. H. Lee, N. R. Wray, M. E. Goddard, and P. M. Visscher. Estimating Missing Heritability for Disease from Genome-wide Association Studies. *The American Journal of Human Genetics*, 88(3):294–305, Mar. 2011. ISSN 0002-9297. doi: 10.1016/j.ajhg.2011.02.002. URL https://www.sciencedirect.com/science/article/pii/S0002929711000206.

J. Listgarten, C. Lippert, C. M. Kadie, R. I. Davidson, E. Eskin, and D. Heckerman. Improved linear mixed models for genome-wide association studies. *Nature Methods*, 9(6):525–526, June 2012. ISSN 1548-7105. doi: 10.1038/nmeth.2037. URL https://www.nature.com/articles/nmeth.2037. Number: 6 Publisher: Nature Publishing Group.

J. L. Lush. Animal breeding plans. *Animal breeding plans.*, 1943. URL https://www.cabdirect.org/cabdirect/abstract/19440100562. Publisher: Iowa State College Press.

J. W. MacCluer, J. L. VandeBerg, B. Read, and O. A. Ryder. Pedigree analysis by computer simulation. *Zoo Biology*, 5(2):147–160, 1986. ISSN 1098-2361. doi: 10.1002/zoo.1430050209. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/zoo.1430050209.

T. F. C. Mackay. Epistasis and quantitative traits: using model organisms to study gene–gene interactions. *Nature Reviews Genetics*, 15(1):22–33, Jan. 2014. ISSN 1471-0064. doi: 10.1038/nrg3627. URL https://www.nature.com/articles/nrg3627.

S. Makvandi-Nejad, G. E. Hoffman, J. J. Allen, E. Chu, E. Gu, A. M. Chandler, A. I. Loredo, R. R. Bellone, J. G. Mezey, S. A. Brooks, and N. B. Sutter. Four Loci Explain 83% of Size Variation in the Horse. *PLOS ONE*, 7(7):e39929, July 2012. ISSN 1932-6203. doi: 10.1371/journal.pone.0039929. URL https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0039929. Publisher: Public Library of Science.

T. A. Manolio, F. S. Collins, N. J. Cox, D. B. Goldstein, L. A. Hindorff, D. J. Hunter, M. I. McCarthy, E. M. Ramos, L. R. Cardon, A. Chakravarti, J. H. Cho, A. E. Guttmacher, A. Kong, L. Kruglyak, E. Mardis, C. N. Rotimi, M. Slatkin, D. Valle, A. S. Whittemore, M. Boehnke, A. G. Clark, E. E. Eichler, G. Gibson, J. L. Haines, T. F. C. Mackay, S. A. McCarroll, and P. M. Visscher. Finding the missing heritability of complex diseases. *Nature*, 461(7265):747–753, Oct. 2009. ISSN 1476-4687. doi: 10.1038/nature08494. URL https://www.nature.com/articles/nature08494. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 7265 Primary_atype: Reviews Publisher: Nature Publishing Group.

J. Marchini, L. R. Cardon, M. S. Phillips, and P. Donnelly. The effects of human population structure on large genetic association studies. *Nature Genetics*, 36(5):512–517, May 2004. ISSN 1546-1718. doi: 10.1038/ng1337. URL https://www.nature.com/articles/ng1337. Number: 5 Publisher: Nature Publishing Group.

M. Marchini, L. M. Sparrow, M. N. Cosman, A. Dowhanik, C. B. Krueger, B. Hallgrimsson, and C. Rolian. Impacts of genetic correlation on the independent evolution of body mass and skeletal size in mammals. *BMC Evolutionary Biology*, 14(1):258, Dec. 2014. ISSN 1471-2148. doi: 10.1186/s12862-014-0258-0. URL https://doi.org/10.1186/s12862-014-0258-0.

A. Martin and V. Orgogozo. The Loci of Repeated Evolution: A Catalog of Genetic Hotspots of Phenotypic Variation. *Evolution*, 67(5):1235–1250, 2013. ISSN 1558-5646. doi: 10.1111/evo.12081. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/evo.12081. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/evo.12081.

O. C. Martin and F. Hospital. Distribution of Parental Genome Blocks in Recombinant Inbred Lines. *Genetics*, 189(2):645–654, Oct. 2011. ISSN 0016-6731. doi: 10.1534/genetics.111.129700. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3189807/.

J. Maynard Smith and J. Haigh. The hitch-hiking effect of a favourable gene. *Genetics Research*, 23(1):23–35, Feb. 1974. ISSN 1469-5073, 0016-6723. doi: 10.1017/S0016672300014634. URL https://www.cambridge.org/core/journals/genetics-research/

article/hitchhiking-effect-of-a-favourable-gene/918291A3B62BD50E1AE5C1F22165EF1B.

A. F. McRae, J. M. Pemberton, and P. M. Visscher. Modeling Linkage Disequilibrium in Natural Populations: The Example of the Soay Sheep Population of St. Kilda, Scotland. *Genetics*, 171(1):251–258, Sept. 2005. ISSN 1943-2631. doi: 10.1534/genetics.105.040972. URL https://doi.org/10.1534/genetics.105.040972.

G. A. T. McVean. A Genealogical Interpretation of Linkage Disequilibrium. *Genetics*, 162(2): 987–991, Oct. 2002. ISSN 1943-2631. doi: 10.1093/genetics/162.2.987. URL https://doi.org/10.1093/genetics/162.2.987.

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*, 118(25), June 2021. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.2015005118. URL https://www.pnas.org/content/118/25/e2015005118. Publisher: National Academy of Sciences Section: Biological Sciences.

R. Nielsen. Molecular Signatures of Natural Selection. *Annual Review of Genetics*, 39(1):197–218, 2005. doi: 10.1146/annurev.genet.39.073003.112420. URL https://doi.org/10.1146/annurev.genet.39.073003.112420. _eprint: https://doi.org/10.1146/annurev.genet.39.073003.112420.

S. V. Nuzhdin, C. L. Dilda, and T. F. Mackay. The genetic architecture of selection response. Inferences from fine-scale mapping of bristle number quantitative trait loci in Drosophila melanogaster. *Genetics*, 153(3):1317–1331, Nov. 1999. ISSN 0016-6731. doi: 10.1093/genetics/153.3.1317.

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. ISSN 1365-294X. doi: 10.1111/j.1365-294X.2012.05673.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-294X.2012.05673.x. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-294X.2012.05673.x.

H. A. Orr. The Population Genetics of Adaptation: The Distribution of Factors Fixed during Adaptive Evolution. *Evolution*, 52(4):935–949, 1998. ISSN 0014-3820. doi: 10.2307/2411226. URL https://www.jstor.org/stable/2411226. Publisher: [Society for the Study of Evolution, Wiley].

H. A. Orr. The Probability of Parallel Evolution. *Evolution*, 59(1):216–220, 2005. ISSN 1558-5646. doi: 10.1111/j.0014-3820.2005.tb00907.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.0014-3820.2005.tb00907.x. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.0014-3820.2005.tb00907.x.

J.-H. Park, S. Wacholder, M. H. Gail, U. Peters, K. B. Jacobs, S. J. Chanock, and N. Chatterjee. Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. *Nature Genetics*, 42(7):570–575, July 2010. ISSN 1546-1718. doi: 10.1038/ng.610. URL https://www.nature.com/articles/ng.610.

M. Pelizzola, M. Behr, H. Li, A. Munk, and A. Futschik. Multiple haplotype reconstruction from allele frequency data. *Nature Computational Science*, 1(4):262–271, Apr. 2021. ISSN 2662-8457. doi: 10.1038/s43588-021-00056-5. URL https://www.nature.com/articles/s43588-021-00056-5. Number: 4 Publisher: Nature Publishing Group.

J. Pemberton. Wild pedigrees: the way forward. *Proceedings of the Royal Society B: Biological Sciences*, 275(1635):613–621, Mar. 2008. doi: 10.1098/rspb.2007.1531. URL https://royalsocietypublishing.org/doi/full/10.1098/rspb.2007.1531.

J. K. Pritchard, J. K. Pickrell, and G. Coop. The Genetics of Human Adaptation: Hard Sweeps, Soft Sweeps, and Polygenic Adaptation. *Current Biology*, 20(4):R208–R215, Feb. 2010. ISSN 0960-9822. doi: 10.1016/j.cub.2009.11.055. URL https://www.sciencedirect.com/science/article/pii/S0960982209020703.

S. Provot, H. Kempf, L. C. Murtaugh, U.-i. Chung, D.-W. Kim, J. Chyung, H. M. Kronenberg, and A. B. Lassar. Nkx3.2/Bapx1 acts as a negative regulator of chondrocyte maturation. *Development (Cambridge, England)*, 133(4):651–662, Feb. 2006. ISSN 0950-1991. doi: 10.1242/dev.02258.

M. Ravinet, A. Westram, K. Johannesson, R. Butlin, C. André, and M. Panova. Shared and nonshared genomic divergence in parallel ecotypes of Littorina saxatilis at a local scale. *Molecular Ecology*, 25(1):287–305, 2016. ISSN 1365-294X. doi: 10.1111/mec.13332. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/mec.13332. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/mec.13332.

R. G. Reeves and D. Tautz. Automated Phenotyping Indicates Pupal Size in Drosophila Is a Highly Heritable Trait with an Apparent Polygenic Basis. *G3: Genes\textbarGenomes\textbarGenetics*, 7(4):1277–1286, Mar. 2017. ISSN 2160-1836. doi: 10.1534/g3.117.039883. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5386876/.

n. Reznick, n. Shaw, n. Rodd, and n. Shaw. Evaluation of the Rate of Evolution in Natural Populations of Guppies (Poecilia reticulata). *Science (New York, N.Y.)*, 275(5308):1934–1937, Mar. 1997. ISSN 1095-9203. doi: 10.1126/science.275.5308.1934.

M. Rimbault, H. C. Beale, J. J. Schoenebeck, B. C. Hoopes, J. J. Allen, P. Kilroy-Glynn, R. K. Wayne, N. B. Sutter, and E. A. Ostrander. Derived variants at six genes explain nearly half of size reduction in dog breeds. *Genome Research*, 23(12):1985–1995, Dec. 2013. ISSN 1088-9051, 1549-5469. doi: 10.1101/gr.157339.113. URL https://genome.cshlp.org/content/23/12/1985. Company: Cold Spring Harbor Laboratory Press Distributor: Cold Spring Harbor Laboratory Press Institution: Cold Spring Harbor Laboratory Press Label: Cold Spring Harbor Laboratory Press Publisher: Cold Spring Harbor Lab.

N. Risch and K. Merikangas. The future of genetic studies of complex human diseases. *Science (New York, N.Y.)*, 273(5281):1516–1517, Sept. 1996. ISSN 0036-8075. doi: 10.1126/science.273.5281.1516.

A. Robertson. A theory of limits in artificial selection. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 153(951):234–249, Nov. 1960. doi: 10.1098/rspb.1960.0099. URL https://royalsocietypublishing.org/doi/10.1098/rspb.1960.0099. Publisher: Royal Society.

A. Robertson. Artificial selection with a large number of linked loci. In *Proceedings of the 1st International Conference on Quantitative Genetics*, pages 307–322. Iowa State University Press Ames, IA, 1977.

M. V. Rockman. The Qtn Program and the Alleles That Matter for Evolution: All That's Gold Does Not Glitter. *Evolution*, 66(1):1–17, 2012. ISSN 1558-5646. doi: 10.1111/j.1558-5646.2011.01486.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1558-5646.2011.01486.x. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1558-5646.2011.01486.x.

H. Sachdeva and N. H. Barton. Introgression of a Block of Genome Under Infinitesimal Selection. *Genetics*, 209(4):1279–1303, Aug. 2018. ISSN 1943-2631. doi: 10.1534/genetics.118.301018. URL https://doi.org/10.1534/genetics.118.301018.

E. Santiago. Linkage and the maintenance of variation for quantitative traits by mutation–selection balance: an infinitesimal model. *Genetics Research*, 71(2):161–170, Apr. 1998. ISSN 1469-5073, 0016-6723. doi: 10.1017/S0016672398003231. URL https://www.cambridge.org/core/journals/genetics-research/article/linkage-and-the-maintenance-of-variation-for-quantitative-traits-by-mut 2085C53021580D05FCA785172FEE0D5D. Publisher: Cambridge University Press.

D. Schluter, E. Clifford, M. Nemethy, and J. McKinnon. Parallel Evolution and Inheritance of Quantitative Traits. *The American Naturalist*, 163(6):809–822, June 2004. ISSN 0003-0147. doi: 10.1086/383621. URL https://www.journals.uchicago.edu/doi/10.1086/383621. Publisher: The University of Chicago Press.

M. Slatkin. Linkage disequilibrium — understanding the evolutionary past and mapping the medical future. *Nature reviews. Genetics*, 9(6):477–485, June 2008. ISSN 1471-0056. doi: 10.1038/nrg2361. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5124487/.

D. L. Stern. The genetic causes of convergent evolution. *Nature Reviews Genetics*, 14(11):751–764, Nov. 2013. ISSN 1471-0064. doi: 10.1038/nrg3483. URL https://www.nature.com/articles/nrg3483. Number: 11 Publisher: Nature Publishing Group.

M. Tiret and F. Hospital. Blocks of chromosomes identical by descent in a population: Models and predictions. *PLOS ONE*, 12(11):e0187416, Nov. 2017. ISSN 1932-6203. doi: 10.1371/journal.pone.0187416. URL https://dx.plos.org/10.1371/journal.pone.0187416.

H. K. Tiwari, J. Barnholtz-Sloan, N. Wineinger, M. A. Padilla, L. K. Vaughan, and D. B. Allison. Review and Evaluation of Methods Correcting for Population Stratification with a Focus on Underlying Statistical Principles. *Human Heredity*, 66(2):67–86, 2008. ISSN 0001-5652, 1423-0062. doi: 10.1159/000119107. URL https://www.karger.com/Article/FullText/119107. Publisher: Karger Publishers.

M. C. Turchin, C. W. Chiang, C. D. Palmer, S. Sankararaman, D. Reich, and J. N. Hirschhorn. Evidence of widespread selection on standing variation in Europe at height-associated SNPs. *Nature Genetics*, 44(9):1015–1019, Sept. 2012. ISSN 1546-1718. doi: 10.1038/ng.2368. URL https://www.nature.com/articles/ng.2368. Number: 9 Publisher: Nature Publishing Group.

M. Turelli. Heritable genetic variation via mutation-selection balance: Lerch's zeta meets the abdominal bristle. *Theoretical Population Biology*, 25(2):138–193, Apr. 1984. ISSN 0040-5809. doi: 10.1016/0040-5809(84)90017-0. URL https://www.sciencedirect.com/science/article/pii/0040580984900170.

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, Mar. 2011. ISSN 1553-7404. doi: 10.1371/journal.pgen.1001336. URL https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1001336.

P. M. Visscher, N. R. Wray, Q. Zhang, P. Sklar, M. I. McCarthy, M. A. Brown, and J. Yang. 10 Years of GWAS Discovery: Biology, Function, and Translation. *The American Journal of Human Genetics*, 101(1):5–22, July 2017. ISSN 0002-9297, 1537-6605. doi: 10.1016/j.ajhg.2017.06.005. URL https://www.cell.com/ajhg/abstract/S0002-9297(17)30240-9.

C. Vlachos and R. Kofler. Optimizing the Power to Identify the Genetic Basis of Complex Traits with Evolve and Resequence Studies. *Molecular Biology and Evolution*, 36(12):2890–2905, Dec. 2019. ISSN 0737-4038. doi: 10.1093/molbev/msz183. URL https://doi.org/10.1093/molbev/msz183.

B. Walsh and M. Lynch. *Evolution and Selection of Quantitative Traits*. Oxford University Press, June 2018. ISBN 978-0-19-256664-5. Google-Books-ID: L2liDwAAQBAJ.

K. E. Weber and L. T. Diggins. Increased selection response in larger populations. II. Selection for ethanol vapor resistance in Drosophila melanogaster at two population sizes. *Genetics*, 125(3):585–597, July 1990. ISSN 0016-6731. doi: 10.1093/genetics/125.3.585.

M. N. Weedon, G. Lettre, R. M. Freathy, C. M. Lindgren, B. F. Voight, J. R. B. Perry, K. S. Elliott, R. Hackett, C. Guiducci, B. Shields, E. Zeggini, H. Lango, V. Lyssenko, N. J. Timpson, N. P. Burtt, N. W. Rayner, R. Saxena, K. Ardlie, J. H. Tobias, A. R. Ness, S. M. Ring, C. N. A. Palmer, A. D. Morris, L. Peltonen, V. Salomaa, G. D. Smith, L. C. Groop, A. T. Hattersley, M. I. McCarthy, J. N. Hirschhorn, and T. M. Frayling. A common variant of HMGA2 is associated with adult and childhood height in the general population. *Nature Genetics*, 39(10):1245–1250, Oct. 2007. ISSN 1546-1718. doi: 10.1038/ng2121. URL https://www.nature.com/articles/ng2121. Number: 10 Publisher: Nature Publishing Group.

S. S. Wilks. The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses. *The Annals of Mathematical Statistics*, 9(1):60–62, Mar. 1938. ISSN 0003-4851, 2168-8990. doi: 10.1214/aoms/1177732360. URL https://projecteuclid.org/journals/annals-of-mathematical-statistics/volume-9/issue-1/The-Large-Sample-Distribution-of-the-Likelihood-Ratio-for-Testing/10.1214/aoms/1177732360.full.

A. R. Wood, T. Esko, J. Yang, S. Vedantam, T. H. Pers, S. Gustafsson, A. Y. Chu, K. Estrada, J. Luan, Z. Kutalik, N. Amin, M. L. Buchkovich, D. C. Croteau-Chonka, F. R. Day, Y. Duan, T. Fall, R. Fehrmann, T. Ferreira, A. U. Jackson, J. Karjalainen, K. S. Lo, A. E. Locke, R. Mägi, E. Mihailov, E. Porcu, J. C. Randall, A. Scherag, A. A. E. Vinkhuyzen, H.-J. Westra, T. W. Winkler, T. Workalemahu, J. H. Zhao, D. Absher, E. Albrecht, D. Anderson, J. Baron,

M. Beekman, A. Demirkan, G. B. Ehret, B. Feenstra, M. F. Feitosa, K. Fischer, R. M. Fraser, A. Goel, J. Gong, A. E. Justice, S. Kanoni, M. E. Kleber, K. Kristiansson, U. Lim, V. Lotay, J. C. Lui, M. Mangino, I. M. Leach, C. Medina-Gomez, M. A. Nalls, D. R. Nyholt, C. D. Palmer, D. Pasko, S. Pechlivanis, I. Prokopenko, J. S. Ried, S. Ripke, D. Shungin, A. Stancáková, R. J. Strawbridge, Y. J. Sung, T. Tanaka, A. Teumer, S. Trompet, S. W. van der Laan, J. van Setten, J. V. Van Vliet-Ostaptchouk, Z. Wang, L. Yengo, W. Zhang, U. Afzal, J. Ärnlöv, G. M. Arscott, S. Bandinelli, A. Barrett, C. Bellis, A. J. Bennett, C. Berne, M. Blüher, J. L. Bolton, Y. Böttcher, H. A. Boyd, M. Bruinenberg, B. M. Buckley, S. Buyske, I. H. Caspersen, P. S. Chines, R. Clarke, S. Claudi-Boehm, M. Cooper, E. W. Daw, P. A. De Jong, J. Deelen, G. Delgado, J. C. Denny, R. Dhonukshe-Rutten, M. Dimitriou, A. S. F. Doney, M. Dörr, N. Eklund, E. Eury, L. Folkersen, M. E. Garcia, F. Geller, V. Giedraitis, A. S. Go, H. Grallert, T. B. Grammer, J. Gräßler, H. Grönberg, L. C. P. G. M. de Groot, C. J. Groves, J. Haessler, P. Hall, T. Haller, G. Hallmans, A. Hannemann, C. A. Hartman, M. Hassinen, C. Hayward, N. L. Heard-Costa, Q. Helmer, G. Hemani, A. K. Henders, H. L. Hillege, M. A. Hlatky, W. Hoffmann, P. Hoffmann, O. Holmen, J. J. Houwing-Duistermaat, T. Illig, A. Isaacs, A. L. James, J. Jeff, B. Johansen, A. Johansson, J. Jolley, T. Juliusdottir, J. Junttila, A. N. Kho, L. Kinnunen, N. Klopp, T. Kocher, W. Kratzer, P. Lichtner, L. Lind, J. Lindström, S. Lobbens, M. Lorentzon, Y. Lu, V. Lyssenko, P. K. E. Magnusson, A. Mahajan, M. Maillard, W. L. McArdle, C. A. McKenzie, S. McLachlan, P. J. McLaren, C. Menni, S. Merger, L. Milani, A. Moayyeri, K. L. Monda, M. A. Morken, G. Müller, M. Müller-Nurasyid, A. W. Musk, N. Narisu, M. Nauck, I. M. Nolte, M. M. Nöthen, L. Oozageer, S. Pilz, N. W. Rayner, F. Renstrom, N. R. Robertson, L. M. Rose, R. Roussel, S. Sanna, H. Scharnagl, S. Scholtens, F. R. Schumacher, H. Schunkert, R. A. Scott, J. Sehmi, T. Seufferlein, J. Shi, K. Silventoinen, J. H. Smit, A. V. Smith, J. Smolonska, A. V. Stanton, K. Stirrups, D. J. Stott, H. M. Stringham, J. Sundström, M. A. Swertz, A.-C. Syvänen, B. O. Tayo, G. Thorleifsson, J. P. Tyrer, S. van Dijk, N. M. van Schoor, N. van der Velde, D. van Heemst, F. V. A. van Oort, S. H. Vermeulen, N. Verweij, J. M. Vonk, L. L. Waite, M. Waldenberger, R. Wennauer, L. R. Wilkens, C. Willenborg, T. Wilsgaard, M. K. Wojczynski, A. Wong, A. F. Wright, Q. Zhang, D. Arveiler, S. J. L. Bakker, J. Beilby, R. N. Bergman, S. Bergmann, R. Biffar, J. Blangero, D. I. Boomsma, S. R. Bornstein, P. Bovet, P. Brambilla, M. J. Brown, H. Campbell, M. J. Caulfield, A. Chakravarti, R. Collins, F. S. Collins, D. C. Crawford, L. A. Cupples, J. Danesh, U. de Faire, H. M. den Ruijter, R. Erbel, J. Erdmann, J. G. Eriksson, M. Farrall, E. Ferrannini, J. Ferrières, I. Ford, N. G. Forouhi, T. Forrester, R. T. Gansevoort, P. V. Gejman, C. Gieger, A. Golay, O. Gottesman, V. Gudnason, U. Gyllensten, D. W. Haas, A. S. Hall, T. B. Harris, A. T. Hattersley, A. C. Heath, C. Hengstenberg, A. A. Hicks, L. A. Hindorff, A. D. Hingorani, A. Hofman, G. K. Hovingh, S. E. Humphries, S. C. Hunt, E. Hypponen, K. B. Jacobs, M.-R. Jarvelin, P. Jousilahti, A. M. Jula, J. Kaprio, J. J. P. Kastelein, M. Kayser, F. Kee, S. M. Keinanen-Kiukaanniemi, L. A. Kiemeney, J. S. Kooner, C. Kooperberg, S. Koskinen, P. Kovacs, A. T. Kraja, M. Kumari, J. Kuusisto, T. A. Lakka, C. Langenberg, L. Le Marchand, T. Lehtimäki, S. Lupoli, P. A. F. Madden, S. Männistö, P. Manunta, A. Marette, T. C. Matise, B. McKnight, T. Meitinger, F. L. Moll, G. W. Montgomery, A. D. Morris, A. P. Morris, J. C. Murray, M. Nelis, C. Ohlsson, A. J. Oldehinkel, K. K. Ong, W. H. Ouwehand, G. Pasterkamp, A. Peters, P. P. Pramstaller, J. F. Price, L. Qi, O. T. Raitakari, T. Rankinen, D. C. Rao, T. K. Rice, M. Ritchie, I. Rudan, V. Salomaa, N. J. Samani, J. Saramies, M. A. Sarzynski, P. E. H. Schwarz, S. Sebert, P. Sever, A. R. Shuldiner, J. Sinisalo, V. Steinthorsdottir, R. P. Stolk, J.-C. Tardif, A. Tönjes, A. Tremblay, E. Tremoli, J. Virtamo, M.-C. Vohl, The Electronic Medical Records and Genomics (eMERGE) Consortium, The MIGen Consortium, The PAGE Consortium, The LifeLines Cohort Study, P. Amouyel, F. W. Asselbergs, T. L.

Assimes, M. Bochud, B. O. Boehm, E. Boerwinkle, E. P. Bottinger, C. Bouchard, S. Cauchi, J. C. Chambers, S. J. Chanock, R. S. Cooper, P. I. W. de Bakker, G. Dedoussis, L. Ferrucci, P. W. Franks, P. Froguel, L. C. Groop, C. A. Haiman, A. Hamsten, M. G. Hayes, J. Hui, D. J. Hunter, K. Hveem, J. W. Jukema, R. C. Kaplan, M. Kivimaki, D. Kuh, M. Laakso, Y. Liu, N. G. Martin, W. März, M. Melbye, S. Moebus, P. B. Munroe, I. Njølstad, B. A. Oostra, C. N. A. Palmer, N. L. Pedersen, M. Perola, L. Pérusse, U. Peters, J. E. Powell, C. Power, T. Quertermous, R. Rauramaa, E. Reinmaa, P. M. Ridker, F. Rivadeneira, J. I. Rotter, T. E. Saaristo, D. Saleheen, D. Schlessinger, P. E. Slagboom, H. Snieder, T. D. Spector, K. Strauch, M. Stumvoll, J. Tuomilehto, M. Uusitupa, P. van der Harst, H. Völzke, M. Walker, N. J. Wareham, H. Watkins, H.-E. Wichmann, J. F. Wilson, P. Zanen, P. Deloukas, I. M. Heid, C. M. Lindgren, K. L. Mohlke, E. K. Speliotes, U. Thorsteinsdottir, I. Barroso, C. S. Fox, K. E. North, D. P. Strachan, J. S. Beckmann, S. I. Berndt, M. Boehnke, I. B. Borecki, M. I. McCarthy, A. Metspalu, K. Stefansson, A. G. Uitterlinden, C. M. van Duijn, L. Franke, C. J. Willer, A. L. Price, G. Lettre, R. J. F. Loos, M. N. Weedon, E. Ingelsson, J. R. O'Connell, G. R. Abecasis, D. I. Chasman, M. E. Goddard, P. M. Visscher, J. N. Hirschhorn, and T. M. Frayling. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature Genetics*, 46(11):1173–1186, Nov. 2014. ISSN 1546-1718. doi: 10.1038/ng.3097. URL https://www.nature.com/articles/ng.3097.

B. Yalcin, J. Nicod, A. Bhomra, S. Davidson, J. Cleak, L. Farinelli, M. Østerås, A. Whitley, W. Yuan, X. Gan, M. Goodson, P. Klenerman, A. Satpathy, D. Mathis, C. Benoist, D. J. Adams, R. Mott, and J. Flint. Commercially Available Outbred Mice for Genome-Wide Association Studies. *PLOS Genetics*, 6(9):e1001085, Sept. 2010. ISSN 1553-7404. doi: 10.1371/journal.pgen.1001085. URL https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1001085. Publisher: Public Library of Science.

K. Zhang, M. Deng, T. Chen, M. S. Waterman, and F. Sun. A dynamic programming algorithm for haplotype block partitioning. *Proceedings of the National Academy of Sciences*, 99(11): 7335–7339, May 2002. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.102186799. URL https://www.pnas.org/content/99/11/7335. Publisher: National Academy of Sciences Section: Physical Sciences.

X. Zhou and M. Stephens. Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics*, 44(7):821–824, July 2012. ISSN 1546-1718. doi: 10.1038/ng.2310. URL https://www.nature.com/articles/ng.2310. Number: 7 Publisher: Nature Publishing Group.