THE UNIVERSITY OF CHICAGO

TWO PROBLEMS IN POPULATION GENETICS

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES
AND THE PRITZKER SCHOOL OF MEDICINE
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

COMMITTEE ON GENETICS, GENOMICS, AND SYSTEMS BIOLOGY

BY
MARYN OLIVIA CARLSON

CHICAGO, ILLINOIS
AUGUST 2023

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

I am grateful to my PhD advisor, Matthias Steinrücken, for accepting me into his lab after a difficult first year, for his openness to impromptu meetings and generosity of time, and for enabling me to pursue several of my own research ideas. My thesis committee, Matthew Stephens, John Novembre, and Jeremy Berg, have provided invaluable feedback on the research that constitutes this thesis, as well as, several other research projects. Daniel Rice for sharing his careful and exacting approach to science and his contributions to the first chapter of this thesis. Howard Judelson for his insightful feedback on the work constituting the second chapter of this thesis. My friends and colleagues who through myriad conversations have contributed substantially to this work. And finally, my mentors at Cornell University, William Fry, Elodie Gazave, and Michael Gore, without whom I would not have began or completed this PhD.

ABSTRACT

Genome-wide association (GWA) studies conducted in large human cohorts have revealed that many traits are highly polygenic, with numerous loci throughout the genome contributing small additive effects. These findings imply a simple prediction model, often referred to as a polygenic score, in which an individual's predicted phenotype is the inner product of their genotype and a set of weights corresponding to the effects of every site in the genome.

It is now widely appreciated that the accuracy of a focal individual's polygenic score depends on their genetic relationship to the population in which the prediction model was generated. Genetic predictions are usually less accurate for more distantly related, out-of-sample individuals. When environmental conditions are constant, this reduction in accuracy can largely be attributed to differences in allele frequencies and patterns of linkage disequilibria between the GWA study population and the population from which the focal individual was sampled. The primary aim of the first project is to determine what proportion of reduced accuracy can be attributed to the former—referred to as allelic turnover—in isolation. In Chapter 2, I develop a theoretical framework to investigate the effects of allelic turnover on the accuracy of out-of-sample polygenic scores.

The second thrust of my thesis describes the distribution of mitotic loss of heterozygosity (LOH) throughout the genome of the plant pathogen, *Phytophthora capsici*. In Chapter 3, I develop a procedure to infer LOH events simultaneously in multiple members of a clonal lineage. As isolates within a clonal lineage differ only by mutations accrued during mitotis, LOH events identified in lineages can be more readily attributed to mitotic processes, distinct from meiotic processes, such as inbreeding, which also produce runs of homozygosity. I apply the method to two large genotyping-by-sequencing data sets of *P. capsici*.

CHAPTER 1

INTRODUCTION

This thesis unites two disparate topics in population genetics: (1) Polygenic score accuracy in out-of-sample predictions, and (2) Mitotic loss of heterozygosity (LOH) in an oomycete plant pathogen.

◇ ◇ ◇

My work on the first topic was a response to the increasingly prevalent refrain that out-of-sample genetic predictions of individuals' phenotypes should be less accurate than their in-sample counterparts (Martin et al., 2017). In the human genetics community, this sentiment emerged largely from reports of substantial reductions in prediction accuracy when prediction models generated in white European samples were applied to individuals of distinct ancestry. These observations were bolstered by simulations showing that cross-population prediction accuracy was significantly reduced under a realistic human demography (Martin et al., 2017).[1] With environmental factors held constant, all accuracy reductions in these simulations could be attributed to differences in allele frequencies and patterns of linkage disequilibria (LD) between the GWA study population and the population of the focal individual(s). Reductions in cross-population accuracy have been replicated in additional empirical (Bitarello and Mathieson, 2020), simulation (Durvasula and Lohmueller, 2021; Ragsdale et al., 2020), and combined (Wang et al., 2020) studies, and see Fig. 1.1.

1. It was later shown that a programming error in the specification of the demographic model inflated the simulated accuracy reductions observed in (Martin et al., 2017). Though attenuated, reductions in cross-population accuracy persisted when the demographic model was correctly specified (Ragsdale et al., 2020). In addition, the empirical differences in polygenic score accuracy reported in (Martin et al., 2017) were also likely inflated by bias in the effect estimates used to compute the polygenic scores (Sohail et al., 2019; Berg et al., 2019).

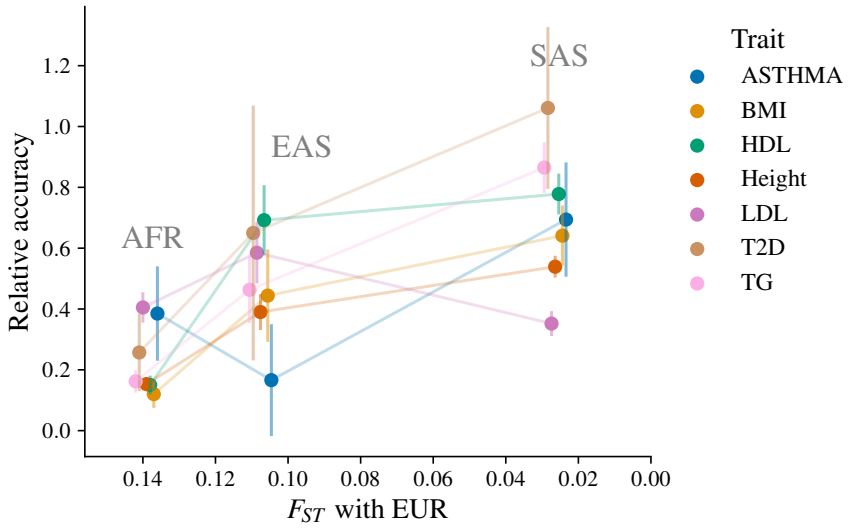


Figure 1.1: **Relative accuracy for different ancestry groups in the UK Biobank.** For each ancestry group South Asian (SAS), East Asian (EAS), and African (AFR), relative accuracy is computed with respect to the white European (EUR) samples in the UK Biobank, where relative accuracy is defined as the correlation between the polygenic score and the true phenotype in the focal population, e.g., EAS, divided by the same quantity computed in EUR for seven traits. The relative accuracies are plotted as a function of F_{ST} between EUR and the ancestry group. Each color represents a different phenotype, and the error bars show the standard error. All data are from Wang et al. (2020).

This phenomenon, however, had been described much earlier in the plant and animal breeding literature in the context of genomic selection (Meuwissen et al., 2001; Habier et al., 2007). In genomic selection, breeders use genetic predictions of the phenotype as the basis for selection, whereas, in traditional breeding, individuals are selected on the basis of their phenotypic values. If the genetic prediction models are accurate, genomic selection promises more efficient phenotype gains—in many plants, several rounds of selection could occur within a growing season and less frequent phenotyping would be required (Meuwissen et al., 2001; Heffner et al., 2010). Early simulation studies, however, showed that prediction accuracy rapidly decays over several generations of breeding. This reduction in accuracy was explained by decreases in linkage disequilibrium (LD) between the genotyped and causal sites (Meuwissen et al., 2001). In other words, recombination events between the tagging, geno-

typed markers and the causal sites rendered the genotyped sites less informative about the causal sites. In addition, it was appreciated that if only a subset of the training population, analogous to the GWA study population, was used to generate the prediction model, prediction errors would likely be higher for more distantly related individuals (Meuwissen et al., 2001). Motivated in part by these considerations, subsequent empirical and computational efforts sought to optimize genomic selection schemes in terms of the phenotyping frequency, e.g., how often the prediction model should be reestimated, and the genetic diversity of the training population, among other factors (De Roos et al., 2009; Isidro et al., 2015; Heffner et al., 2010).[2] While human geneticists and plant breeders differ in their aims, both must address similar challenges to achieve accurate predictions.

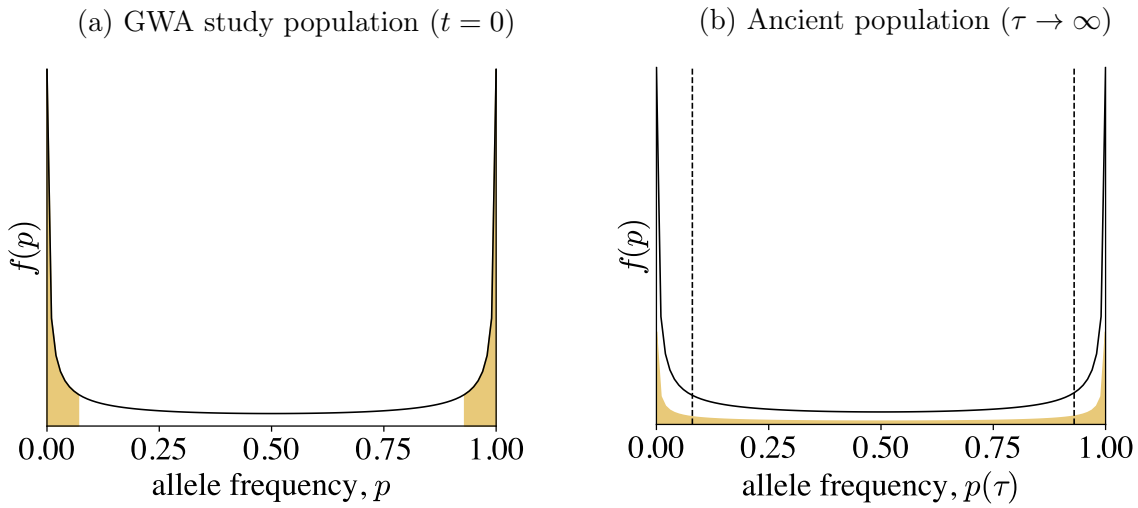


Figure 1.2: **Effects of allele frequency changes on the** *bias* **and** *mse*. In (a), we show the distribution of allele frequencies p in the GWA study population ($t = 0$). The area in gold represents the sites which are inaccessible to the GWA study, i.e., $\hat{\beta}_\ell = 0$ for all sites in the gold areas. For exposition, we consider that the ancient individual was sampled a long time in the past, such that the allele frequencies in (b) have relaxed to stationarity. Many of the gold sites are in interior, within the dashed lines, of the allele frequency spectrum, contributing substantially to trait variation.

Indeed, the concerns enumerated above are equally pertinent to the application of poly-

2. Notably, researchers were concerned with prediction bias due to population structure early in the development of genomic selection, e.g., (De Roos et al., 2009; Isidro et al., 2015).

genic scores to ancient individuals, who may have lived more than 10,000 years in the past. In this context, polygenic scores promise to illuminate the extent to which phenotypic changes over time have been predicated on genetic changes (Cox et al., 2019, 2021). And, where phenotypic information is not available, polygenic scores provide otherwise inaccessible insights into the phenotypes of ancient individuals (Colbran et al., 2019). As in human genetics, application preceded a thorough investigation of the statistical properties of polygenic scores. This uncertainty was expressed in statements such as, "... even in the absence of bias, variance explained by the [polygenic score] is likely to decrease as we move back in time and ancient populations become less closely related to present-day populations" (Cox et al., 2019).

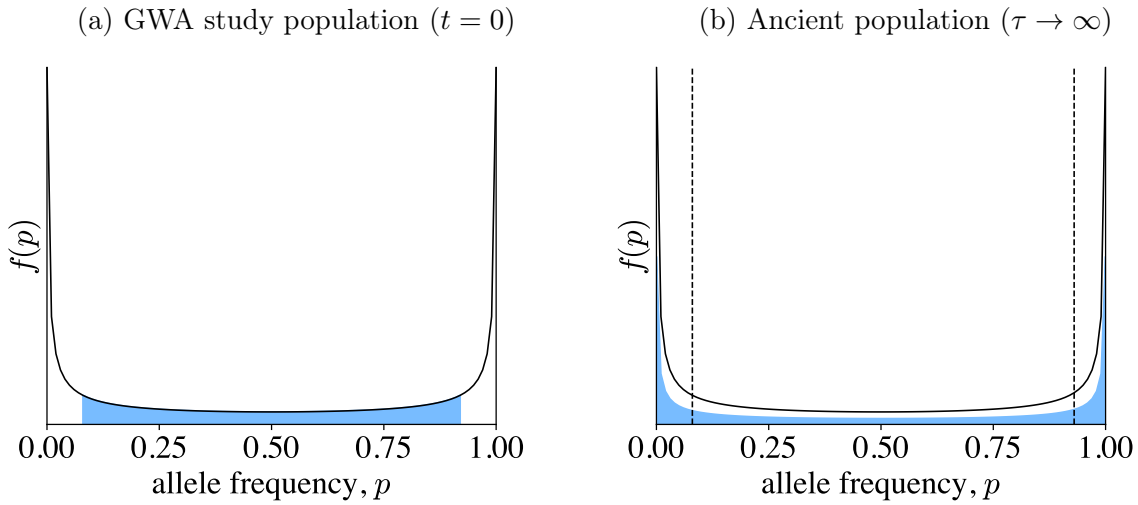


Figure 1.3: **Effects of allele frequency changes on the estimated additive genetic variance.** In (a), we show the distribution of allele frequencies p in the GWA study population ($t = 0$). The area in blue represents the sites which are accessible to the GWA study, i.e., $\hat{\beta}_\ell = \beta_\ell$ for all sites in the blue area. As in Fig. 1.2, we consider that the ancient individual was sampled a long time in the past, such that the allele frequencies in (b) have relaxed to stationarity. Many of the blue sites are in the boundaries of the allele frequency spectrum, outside of the dashed lines. This implies that many of the blue sites were contributing less to trait variation in the past.

My first aim, summarized in Chapter 2, was to use population genetic theory coupled with a simple model of a GWA study to provide a precise answer to the following related question: How much does prediction accuracy decay as a function of the ancient sampling

time? More specifically, I aimed to quantify reductions in prediction accuracy due solely to changes in allele frequency over time, deemed allelic turnover. To do so, I focused on a simple and often used definition of the polygenic score, where the polygenic score of individual i is given by,

$$\hat{Y}_i = \sum_{\ell=1}^{L} X_{i\ell}\hat{\beta}_\ell, \tag{1.1}$$

where $X_{i\ell}$ is the individual's genotype at site ℓ, $\hat{\beta}_\ell$ is an estimate of the additive effect of site ℓ on the trait from a genome-wide association (GWA) study, and the sum is over a subset of L sites in the genome. An important step in my work was to recognize that the sum in Eq. (1.1) could be defined with respect to *all* loci which contribute to trait variation, even those for which $\hat{\beta}_\ell = 0$. This allowed for comparison of $\hat{Y}_i$ (Eq. (1.1)) with the true phenotype Y_i, defined with respect to all loci with non-zero effects on the trait. In addition, I assumed a simple model for the GWA study, where the effects are either estimated perfectly, i.e., $\hat{\beta}_\ell = \beta_\ell$, when the allele frequency p_ℓ is sufficiently high, or the estimated effect is zero.

In Chapter 2, I use a mathematical tool, the spectral decomposition of the transition density function (see Section 2.8.6), to solve for several accuracy metrics. Here, I provide a conceptual overview of my findings, without recourse to the calculations of Chapter 2. For ease of exposition, we consider a scenario where the ancient individual is sampled very far in the past. In Fig. 1.2a, we show the allele frequency distribution in the GWA study population ($t = 0$) under the assumptions of neutrality and recurrent mutations and symmetric mutation rates. The areas in gold near the boundaries represent the sites which can not be detected in the GWA study, i.e., $\hat{\beta}_\ell = 0$ for all of the gold sites. To find the *bias* of the polygenic score, defined as,

$$bias(\tau) = \mathbb{E}[\hat{Y}_i - Y_i], \tag{1.2}$$

we need only consider the frequencies of the gold sites in the ancient population, as the effects of all white sites in the interior were estimated perfectly. For large τ, the gold sites

will have relaxed to stationarity backwards-in-time, spreading out along the allele frequency spectrum. This implies that some of the gold sites will contribute substantially to trait variation in the ancient population—yet, their estimated effects are still zero. However, as the allele frequency changes preserve symmetry, the *bias* of the polygenic score is zero. This is not the case for the mean-squared error, defined as,

$$mse(\tau) = \mathbb{E}\left[(\hat{Y}_i - Y_i)^2\right], \tag{1.3}$$

where each gold site will contribute approximately β_ℓ^2 times the variance of the change in allele frequency to the total mse.

At the same time, the polygenic score will explain less phenotypic variance in the past. To illustrate this, we plot the allele frequency spectrum in the present day in Fig. 1.3a, and color the sites which are contributing to the estimated additive genetic variance,

$$\hat{V}_A = 2\sum_{\ell=1}^{L} \hat{\beta}_\ell^2 p_\ell(1-p_\ell), \tag{1.4}$$

in the GWA study population in blue. In the ancient population ($\tau \to \infty$), we observe that many of the blue sites have migrated to the boundaries of the allele frequency distribution. This implies that the expected contribution of a site to the estimated additive genetic variance is necessarily smaller in the past relative to the present, i.e.,

$$\mathbb{E}[\hat{\beta}_\ell^2 p_\ell(\tau)(1-p_\ell(\tau))] < \mathbb{E}[\hat{\beta}_\ell^2 p_\ell(1-p_\ell)], \tag{1.5}$$

where p_ℓ and $p_\ell(\tau)$ are the allele frequencies in the GWA study and ancient populations, respectively.

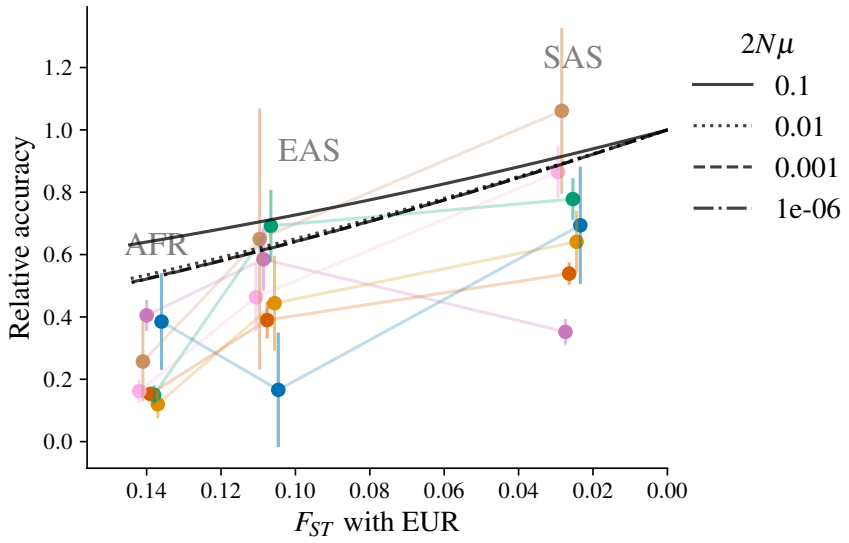


Figure 1.4: **Relative accuracy for different ancestry groups in the UK Biobank with theoretical predictions.** Fig. 1.1 is re-plotted with theoretical results from Chapter 2. Each black line corresponds to predicted relative accuracy as a function of F_{ST} between the ancient and GWA study populations for different mutation rates, $2N\mu \in \{10^{-1}, 10^{-2}, 10^{-3}, 10^{-6}\}$, a GWA study sample size of 10^5, and a detection threshold of 500, i.e., $\hat{\beta}_\ell = \beta_\ell$ when the allele count in the GWA study is more than 500 and less than $2n - 500$.

While I primarily framed my work in terms of ancient polygenic scores, it is equally applicable to a scenario in which the focal individual is sampled from a population which diverged at some time in the past from the population in which the GWA study was conducted—the scenario described in the opening of this introduction. Even in this "best-case" scenario, I found that polygenic score accuracy decayed substantially over time scales relevant to human applications. Yet, allelic turnover cannot explain all of the observed accuracy reductions, as shown in Fig. 1.4. Indeed, additional factors may contribute to accuracy decay, such as differences in patterns of linkage disequilibria, environmental conditions, or trait architectures.

◇ ◇ ◇

While studying the oomycete plant pathogen *Phytophthora capsici*, I unwittingly encountered the genetic phenomenon of mitotic loss of heterozygosity (LOH). As a Master's

student, I was investigating temporal genetic changes in an experimental population of *P. capsici* founded by two parents of opposite mating types, *A*1 and *A*2, with no subsequent introductions of the pathogen.

Like its more infamous relative, *P. infestans*, *P. capsici* is heterothallic. This means that despite the fact that *P. capsici* is hermaphroditic, it requires the presence of two distinct mating types to stimulate sexual reproduction. Current evidence suggests that *P. capsici* preferentially outcrosses, but isolates presumed to result from self-fertilization have been observed in the laboratory and field (Dunn et al., 2014; Carlson et al., 2017). The outcome of sexual reproduction is hardy-long lived oospores, which can withstand cold temperatures and survive for many years in the soil in the absence of a host (Bowers et al., 1990; Lamour and Hausbeck, 2003; Granke et al., 2012). Thus, oospores produced in one growing season can seed epidemics in subsequent years. While sexual reproduction ensures the survival of *P. capsici* across years, rapid asexual reproduction within a season inflicts crop damage. In favorable conditions, *P. capsici* exhibits extensive filamentous growth and readily produces large quantities of asexual spores, referred to as sporangia. Each sporangium contains approximately 20-40 swimming zoospores, which when released from the sporangium, can each initiate a new infection (Hausbeck and Lamour, 2004). In contrast to the oospores, asexual propagules of *P. capsici* cannot survive cold temperatures. As a result, perennial populations should be sustained solely by the recombinant oospores.

Our understanding of *P. capsici*'s population biology was borne out in the biparental field experiment: Samples from the years following inoculation consisted of F_1 and isolates presumed to be the result of intermating among the F_1 and subsequent generations (Dunn et al., 2014; Carlson et al., 2017). The parental isolates were not recovered from the field, as expected.

When I analyzed segregation in the F_1 isolates, certain regions exhibited elevated Mendelian errors with respect to the parental genotypes, i.e., errors in segregation under the assumption

of Mendelian inheritance. Closer inspection revealed that haplotypes segregating in the F_1, were not present in the genotypes of the parental replicates. In the affected regions, one or both parents were instead homozygous for one of the segregating haplotypes. In one case, earlier sequences of the *A*2 parent retained the missing haplotype. Thus, what appeared as aberrant segregation in the F_1 was actually due to the loss of genetic information in the parental isolates. Altogether, these observations suggested that mitotic LOH events must have occurred in the parental isolates during culturing in the laboratory post-inoculation of the field experiment in 2009, and prior to the initiation of sequencing in 2014. At the time, mitotic LOH represented a nuisance as much as a curiosity. My results added to existing evidence for mitotic LOH in *P. capsici* (Hulvey et al., 2010; Lamour et al., 2012), and *Phytophthora* species more broadly (Chamnanpunt et al., 2001), but failed to provide much additional insight into the phenomenon. Subsequent studies have provided additional evidence of mitotic LOH incidence in *P. capsici* (Vogel et al., 2021) and other *Phytophthora* species (Dale et al., 2019). In addition, Vogel et al. (2021) reported mitotic LOH in regions containing SNPs associated with mating type (referred to as the MTR) coincident with mating type discordance within a clonal lineage. In this case, the two *A*1 isolates within the lineage exhibited LOH events in the MTR, whereas the *A*2 isolate remained heterozygous. This observation is consistent with a model of mating type proposed by Carlson et al. (2017) in which the *A*2 mating type is heterozygous for the mating type determining gene(s) and the *A*1 type is homozygous. This model was based on the segregation of alleles in the MTR in the Biparental population and implies that conversion from *A*2 to *A*1 can occur due to LOH, but not vice versa. In summary, investigations of mitotic LOH to date suggests that mitotic LOH occurs relatively frequently in *P. capsici* and may induce mating type switching. However, a large-scale quantitative investigation of LOH in *P. capsici* has heretofore been lacking.

In Chapter 3, I systematically investigate the incidence of mitotic LOH in *P. capsici*. To

do so, I develop an inference procedure to identify mitotic LOH events which have occurred among members of a single clonal lineage. The method takes advantage of the fact that, barring sequencing errors, genetic variation within a clonal lineage can be attributed exclusively to somatic mutations. Importantly, this method can, in theory, distinguish mitotic LOH unequivocally from runs of homozygosity (ROH) due to inbreeding.

I analyzed 35 lineages sampled from the Biparental experimental population described above, and 48 lineages sampled from across New York state. My results provide additional evidence that LOH is a relatively common phenomenon in *P. capsici*, affecting on average 2-3% of an isolate's genome. In some cases, inferred LOH events spanned more than 10% of an isolate's genome, comparable with the expected reductions in heterozygosity due to mating between F_1 isolates (sib-mating).[3] While LOH was widespread across the genome, incidence was fairly similar across data sets. Several regions exhibited elevated LOH incidence in both populations, suggesting that shared genomic features may predispose certain regions to LOH. In addition, we identified four more lineages in which LOH in the MTR was coincident with mating type discordance within the lineage, adding to the one example of Vogel et al. (2021).

Altogether, these observations provoke more questions than they answer: For example, what is the predominant mechanism underlying LOH in *P. capsici*, and how is it regulated? What are the phenotypic consequences of LOH? Does it incur a fitness cost? In the introduction of Chapter 3, I introduce the mechanisms underlying mitotic LOH and dwell on its population genetic consequences. By uniting mechanistic and population genetic considerations with detailed data analysis, I attempt to begin to answer these questions.

3. We note that these estimates of LOH incidence cannot be translated to LOH rates as the number of cell divisions separating members of a clonal lineage is unknown and likely highly variable among lineages.

CHAPTER 2

POLYGENIC SCORE ACCURACY IN ANCIENT SAMPLES: QUANTIFYING THE EFFECTS OF ALLELIC TURNOVER

This chapter has been published as Carlson et al. (2022), and is reproduced here in accordance with the copyright.

2.1 Abstract

Polygenic scores link the genotypes of ancient individuals to their phenotypes, which are often unobservable, offering a tantalizing opportunity to reconstruct complex trait evolution. In practice, however, interpretation of ancient polygenic scores is subject to numerous assumptions. For one, the genome-wide association (GWA) studies from which polygenic scores are derived, can only estimate effect sizes for loci segregating in contemporary populations. Therefore, a GWA study may not correctly identify all loci relevant to trait variation in the ancient population. In addition, the frequencies of trait-associated loci may have changed in the intervening years. Here, we devise a theoretical framework to quantify the effect of this allelic turnover on the statistical properties of polygenic scores as functions of population genetic dynamics, trait architecture, power to detect significant loci, and the age of the ancient sample. We model the allele frequencies of loci underlying trait variation using the Wright-Fisher diffusion, and employ the spectral representation of its transition density to find analytical expressions for several error metrics, including the expected sample correlation between the polygenic scores of ancient individuals and their true phenotypes, referred to as polygenic score accuracy. Our theory also applies to a two-population scenario and demonstrates that allelic turnover alone *may* explain a substantial percentage of the reduced accuracy observed in cross-population predictions, akin to those performed in human genetics. Finally, we use simulations to explore the effects of recent directional

selection, a bias-inducing process, on the statistics of interest. We find that even in the presence of bias, weak selection induces minimal deviations from our neutral expectations for the decay of polygenic score accuracy. By quantifying the limitations of polygenic scores in an explicit evolutionary context, our work lays the foundation for the development of more sophisticated statistical procedures to analyze both temporally and geographically resolved polygenic scores.

2.2 Introduction

Decay in linkage disequilibrium (LD) between tagging and causal sites, population stratification, variation in allele frequencies within and across populations, and environmental heterogeneity, among other factors, are all thought to negatively impact the prediction accuracy of polygenic scores (see e.g., Habier et al. (2007); De Roos et al. (2008); Hamblin et al. (2011); Erbe et al. (2012); Carlson et al. (2013); Wray et al. (2013); Guo et al. (2014), and more recently in humans, e.g., Galinsky et al. (2019); Berg et al. (2019); Sohail et al. (2019); Mostafavi et al. (2020); Bitarello and Mathieson (2020); Durvasula and Lohmueller (2021)). Many of these issues likely influence both within- *and* out-of-sample predictions, where out-of-sample may refer to an individual sampled from a distinct time or location relative to that of the GWA study. While empirical (Martin et al., 2017; Bitarello and Mathieson, 2020) and simulation (Habier et al., 2007; Ragsdale et al., 2020; Durvasula and Lohmueller, 2021) or combined (Wang et al., 2020) studies have explored particular population genetic scenarios or experimental contexts, we still do not know the extent to which each of these factors compromises prediction accuracy in general.

In this work, we address an issue pertinent to out-of-sample prediction: that causal loci may have different allele frequencies in the GWA study and focal populations. Variants common in the GWA study may be rare in the focal population, and vice versa. We refer to this phenomenon as *allelic turnover*. Allelic turnover implies that effect estimates ported across

space and time, or both, may not reflect all of the genetic variation relevant to phenotypic variation in an ancient or geographically distinct population. Allelic turnover further suggests that the statistical properties of ancient polygenic scores depend on when an ancient individual was sampled—a feature not currently accounted for in ancient DNA analyses. Similarly, statistical properties of geographically disparate polygenic scores depend on the divergence time between the GWA study and focal populations. An understanding of allelic turnover in these contexts may ultimately improve statistical analyses of temporally (e.g., Swarts et al. (2017); Cox et al. (2019); Colbran et al. (2019); Cox et al. (2021)) and geographically resolved polygenic scores (e.g., Sohail et al. (2019); Berg et al. (2019)), analyses which are increasingly commonplace.

We aim to quantify the effect of allelic turnover on the polygenic scores of such out-of-sample individuals when they are computed using effect estimates from a contemporary population. We expect that increases in ancient sampling time or divergence time will be associated with declines in polygenic score accuracy due exclusively to allelic turnover. The question is, by how much does accuracy decline? And, can allelic turnover alone explain the reduced accuracy of out-of-sample predictions observed in numerous human (e.g., Wang et al. (2020); Ragsdale et al. (2020)), animal (e.g., Habier et al. (2007); De Roos et al. (2008); Erbe et al. (2012)) and plant (e.g., Windhausen et al. (2012); Lorenz et al. (2012)) experiments and simulation studies. The answer is likely to depend on the particular population genetic, trait, and GWA study features of the system under study (Hamblin et al., 2011). We attempt to capture some important aspects of this diversity in our modeling framework.

Here, we consider a standard implementation of the polygenic score $\hat{Y}$ which attributes non-zero effects to a particular set of loci, $\mathcal{S}$. An individual's polygenic score is a weighted sum of its genotype, where the weights are the estimated allelic effects. The loci in $\mathcal{S}$ and their estimated effects are usually identified in large-scale GWA studies, often performed in regional biobanks with sample sizes in the tens to hundreds of thousands of individuals (e.g.,

the UK Biobank (Bycroft et al., 2018), BioBank Japan (Kanai et al., 2018)). Frequently, the set $\mathcal{S}$ includes loci which are approximately independent and surpass some allele frequency and p-value thresholds. Though there are numerous ways to define a polygenic score (e.g., Meuwissen et al. (2001); de los Campos et al. (2013a) and see Section 2.8.4), the "prune and threshold" method is commonly used and proves analytically tractable in our framework.

Previous quantitative genetic approaches, such as (Daetwyler et al., 2008) and (Wang et al., 2020), largely ignore the underlying population genetic dynamics. For example, Wang et al. (Wang et al., 2020) estimate the reduction in polygenic score accuracy in a focal population relative to the GWA study population as a function of the fixed population-specific trait heritabilities, allele frequencies, and LD patterns, and the estimated per-locus effects. In contrast, we embed the ancient polygenic score in an explicit population genetic framework, allowing us to take into account changes in allele frequency as well as the statistical constraint imposed by a finite GWA study sample size. And, distinct from previous approaches to the evolutionary modeling of polygenic scores (Berg and Coop, 2014), we track the frequencies of *all loci* that potentially contribute to a trait—not just the loci included in the polygenic score (i.e., loci in $\mathcal{S}$).

Henceforth, we frame our study in terms of ancient polygenic scores. However, we formally demonstrate that our theoretical results apply to out-of-space polygenic scores, where the population divergence time multiplied by two is analogous to the ancient sampling time (see Fig. 2.1 and Section 2.8.1). The latter scenario can represent an ancient individual sampled from a population not directly ancestral to that of the GWA study as the two populations must have diverged at some point in the past. This scenario, to a first approximation, describes the population displacement events thought to be ubiquitous in the history of humans (e.g., Liu et al. (2021)). However, human history is additionally characterized by numerous admixture events and population size changes (e.g., Liu et al. (2021)) which are not yet captured within our modeling framework.

We use several statistics to characterize ancient polygenic score error in distinct population genetic and GWA study scenarios. Each statistic is indexed by the ancient sampling time τ: the bias, $bias(\tau)$, mean-squared error, $mse(\tau)$, estimated additive genetic variance, $\hat{V}_A(\tau)$, and polygenic score accuracy, $\rho^2(\tau)$, which approximates the expectation of the squared sample correlation coefficient between the polygenic scores and phenotypes of an ancient sample. In addition, we can readily express these statistics as functions of the genetic divergence between the ancient and GWA study populations, as measured by the fixation index, $\mathrm{F_{ST}}$ (Section 2.8.13). We first derive general forms for these statistics that are agnostic to almost all of our modeling assumptions and which provide conceptual insights into the effects of allelic turnover. Next, we derive explicit, parameter-dependent expressions for each statistic when the trait is neutrally evolving in a population of constant size subject to recurrent mutation—which for small mutation rates approximates the infinite sites model. We take advantage of the spectral representation of the transition density function of the Wright-Fisher diffusion (tdf) to execute these computations (Ewens, 2004; Durrett, 2008; Griffiths and Spano, 2010; Song and Steinrücken, 2012). We then find interpretable linear approximations for the initial rate of increase (or decrease) of the metrics under study. These approximations apply for the small ancient sampling times typical of ancient humans remains (e.g., see Cox et al. (2019)).

Consistent with our expectations, $mse(\tau)$ increases and the estimated additive genetic variance $\hat{V}_A(\tau)$ decreases with increasing sampling age τ. Despite the fact that $mse(\tau)$ and $\hat{V}_A(\tau)$ are measuring distinct quantities—and indeed have different functional forms—our linear approximations reveal that, under our assumptions, both statistics initially change at approximately the same rate. This rate is proportional to the product of the mutation rate and the power to detect trait-associated loci in the GWA study, which in turn, is influenced by both study size, the magnitude of the true per-locus effect, and the underlying distribution of the allele frequencies of causal loci.

Moreover, we show that polygenic score accuracy $\rho^2(\tau)$ is proportional to $\hat{V}_A(\tau)$, which, as stated, is sensitive to the GWA study and evolutionary parameters. Unlike $\hat{V}_A(\tau)$, $\rho^2(\tau)$ depends on the trait heritability h^2, with larger values of h^2 increasing its rate of decay. In contrast, for small mutation rates, relative accuracy, defined as the ratio of $\rho^2(\tau)$ to accuracy measured in a present-day sample $\rho^2(0)$, is insensitive to h^2, the true per-locus effect size, and the GWA study parameters, as long as the GWA study size n exceeds some minimum threshold. We show that this result likely holds for an arbitrary distribution of effects. Importantly, accuracy and relative accuracy decay considerably over the short time spans characteristic of ancient human samples and geographically distinct human populations.

With equal probability of detecting positive versus negative effect alleles, and under neutrality, the bias of the polygenic score is zero for all ancient sampling times. In practice, both of these conditions are likely violated. For example, detection imbalances have been observed in case-control GWA studies (Chan et al., 2014), and many polygenic traits are likely under some form of selection (Pritchard et al., 2010; Boyle et al., 2017b). While unequal thresholds do not precisely capture the phenomena described in (Chan et al., 2014), they do yield a non-zero $bias(\tau)$ within our framework. The magnitude of this bias is small, implying that other perturbations would be necessary to explain an observed, appreciable bias. To relax the neutrality assumption, we simulate recent directional selection. We find that when the selection coefficient is large enough ($4Ns \geq 1$), selection indeed yields biased polygenic scores. Though this selection-induced bias is several orders of magnitude larger than that induced by asymmetry in the detection thresholds, it is still small relative to the variance explained by segregating genetic variants. Additionally, weak selection only induces small deviations from neutral theoretical expectations for the other statistics, suggesting that our neutral theory may still accurately capture accuracy declines in the presence of weak directional selection. Altogether, our theoretical results suggest that allelic turnover may make large contributions to out-of-sample reductions in accuracy, even under neutrality.

2.3 Model and metrics

Our modeling framework readily encompasses two demographic scenarios. In the first, the focal individual is sampled from the same population in which the GWA study was performed, but at a previous point in time τ (Fig. 2.1a). We specify τ in coalescent time units: An ancient sampling time of τ corresponds to $2N \cdot \tau$ generations in the past, with $2N$ as the diploid population size. When $\tau = 0$, the focal individual is an independent sample from the GWA study population. In the second scenario (Fig. 2.1b), the focal individual is sampled at τ' from a population that diverged from the GWA study population at τ_{split} (in coalescent time units) in the past. However, we show in Section 2.8.1 that scenario (A) is equivalent to scenario (B) if the ancient sampling time τ is equal to $2\tau_{\text{split}} - \tau'$. Therefore, we proceed according to the first scenario, while emphasizing that our conclusions readily translate to the second.

We summarize the full model in Fig. 2.1c and detail its constituent parts in the proceeding subsections. Briefly, the genotype of the ancient individual is sampled conditional on the population allele frequencies at τ. The ancient individual's phenotype is then sampled conditional on its genotype. Population allele frequencies for all loci that potentially affect the trait evolve until present day, at which point the GWA study is conducted. In particular, the effect sizes included in the polygenic score model are estimated from the genotypes and phenotypes of n contemporary individuals. Finally, the ancient polygenic score is computed from the ancient individual's genotype and the polygenic score model derived from the results of a contemporary GWA study.

2.3.1 Sampling the genotype of a time-indexed individual

We assume that each site is at most bi-allelic, with possible alleles A_1 and A_2. We denote the genotype of an individual sampled at some time t (in coalescent units) as $X_{i\ell}(t)$, where i indexes the individual, and ℓ the locus. For the ancient individual(s), $t = \tau$; for the

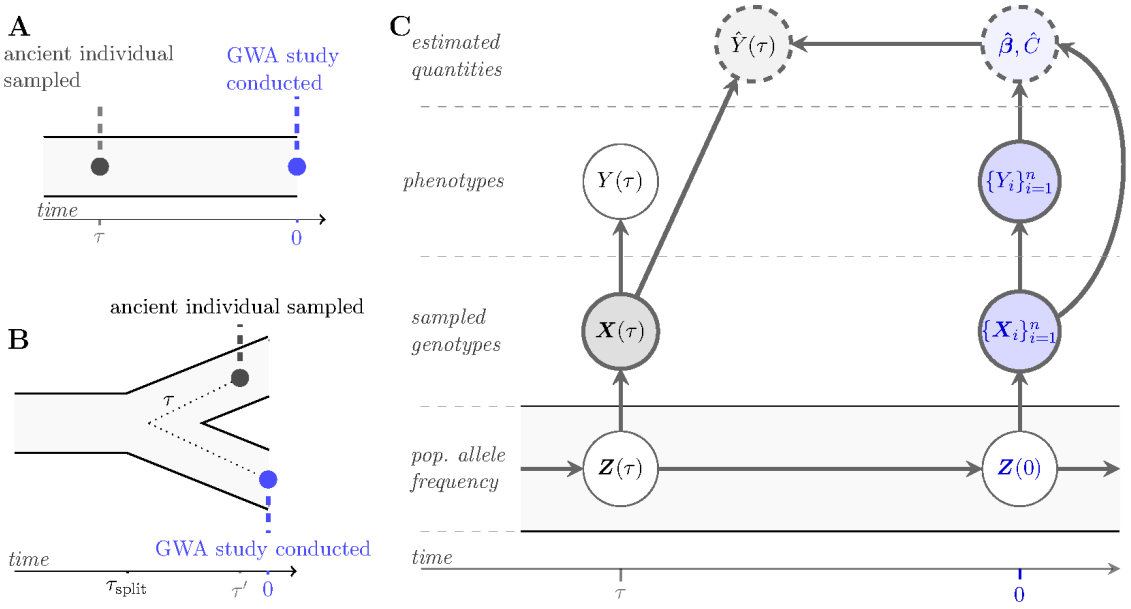


Figure 2.1: **A population genetic model for an ancient polygenic score.** Figures (A) and (B) portray the two demographic scenarios encompassed by our modeling framework. In (A), the ancient individual is sampled at an earlier time τ from the same population in which the GWA study is conducted. In (B), the ancient individual is sampled at an arbitrary time τ'. The ancient individual's population split from the GWA study population at τ_{split} in the past. The two demographic scenarios are equivalent under our assumptions when τ of (A) is equal to $2\tau_{\text{split}} - \tau'$, a quantity denoted by the dotted line in (B). In (C), a graphical model relates the random variables explicit and implicit in the polygenic score $\hat{Y}(\tau)$ and phenotype $Y(\tau)$ of an ancient individual sampled τ generations in the past, as in (A). Darkly shaded and thickly bordered nodes are observed quantities. Unshaded and thinly bordered nodes are unobserved. Lightly shaded nodes bordered by dashed lines denote estimated quantities. Edges denote direct dependencies between connected nodes. For example, conditional on the ancient genotype $\boldsymbol{X}(\tau)$, the polygenic score $\hat{Y}(\tau)$ is independent of the population allele frequencies $\boldsymbol{Z}(\tau)$. Quantities in blue are associated with the present day only, and include the population allele frequencies $\boldsymbol{Z}(0)$; the genotypes of the n individuals in the GWA study, $\{\boldsymbol{X}_i\}_{i=1}^n$ and their phenotypes, $\{Y_i\}_{i=1}^n$; and, the effects and intercept term estimated in the GWA study, $\hat{\boldsymbol{\beta}}$ and $\hat{C}$, respectively.

participants in the GWA study, $t = 0$. For mathematical convenience, we use a symmetric genotype encoding, that is $X_{i\ell}(t) \in \{-1, 0, 1\}$, corresponding to genotypes A_1A_1, A_1A_2, and A_2A_2, respectively. Conditional on the population allele frequency of allele A_2 at t, $Z_\ell(t)$, the distribution of $X_{i\ell}(t)$ is given by the Hardy-Weinberg sampling probabilities: $\mathbb{P}\{X_{i\ell}(t) = -1|Z_\ell(t) = z\} = (1-z)^2$, $\mathbb{P}\{X_{i\ell}(t) = 0|Z_\ell(t) = z\} = 2z(1-z)$, and $\mathbb{P}\{X_{i\ell}(t) = 1|Z_\ell(t) = z\} = z^2$.

2.3.2 Modeling the true phenotype

The genetic basis of a polygenic trait, Y, is determined by a set $\mathcal{L}$, consisting of L distinct genetic loci ($|\mathcal{L}| = L$), each with a true per-locus additive effect $\beta_\ell \in \mathbb{R}$ (for $\ell = 1, 2, \ldots, L$). We further assume that the L loci contribute linearly to the trait, such that the true phenotype of the i-th individual sampled at t is specified by the commonly used additive genetic model (Lynch and Walsh, 1998),

$$Y_i(t) = C + \sum_{\ell=1}^{L} X_{i\ell}(t)\beta_\ell + \epsilon_i(t), \tag{2.1}$$

where C is a constant; β_ℓ is the true additive effect of locus ℓ; and $\epsilon_i(t) \sim \mathcal{N}(0, \sigma_e^2)$ is a normally distributed random variable that incorporates variance in the phenotype due to the environment. The summation in Eq. (2.1) is often referred to as an individual's *genetic value* (Meuwissen et al., 2001). A locus ℓ contributes $\pm\beta_\ell$ to the genetic value (and phenotype) of an individual who is homozygous at ℓ, and zero to that of a heterozygous individual. C is thus the phenotype of an hypothetical all heterozygous individual. Without loss of generality, we set $C = 0$. In addition, we assume, without loss of generality, that all $\beta_\ell \geq 0$ such that locus ℓ contributes $-\beta_\ell$ to the genetic values of A_1A_1 individuals and $+\beta_\ell$ to the genetic values of A_2A_2 individuals.

A fixed locus, $Z_\ell(t) \in [0, 1]$, will affect the mean phenotype of the population at t by

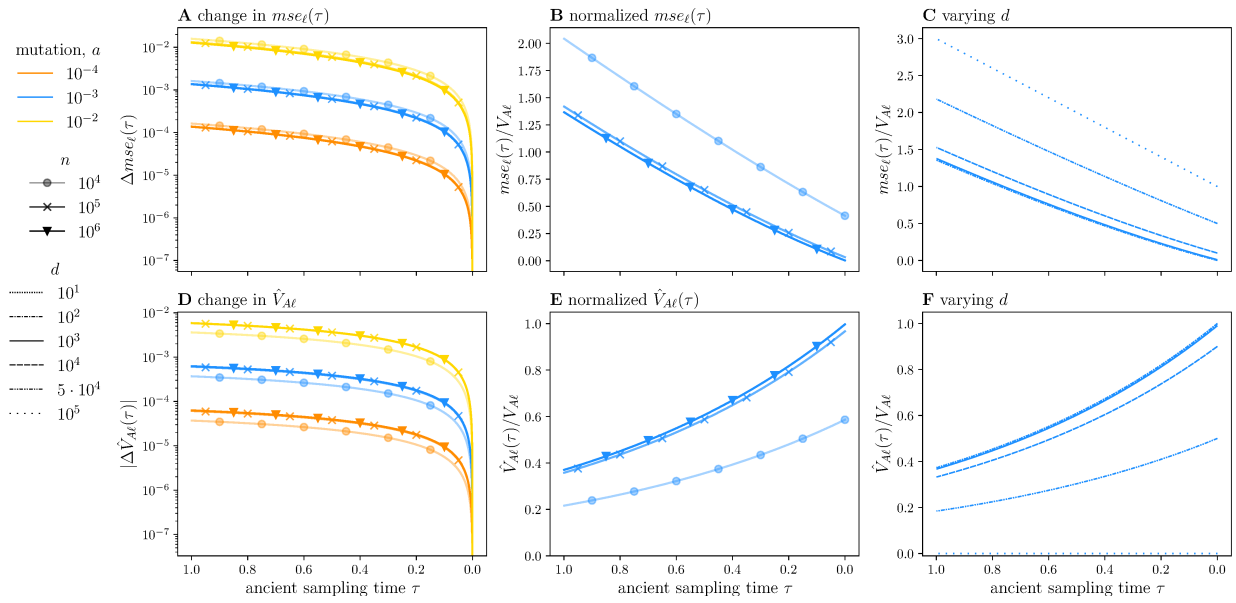


Figure 2.2: **Per locus contributions to the mean-squared error and estimated additive genetic variance across sample sizes, mutation rates, and detection thresholds.** In (A), we plot the per-locus increase in *mse*, $\Delta mse_\ell(\tau)$, normalized by β^2, for three mutation rates $a = 10^{-4}, 10^{-3}, 10^{-2}$ by color, and for the three sample sizes, $n = 10^4, 10^5, 10^6$ by shape, respectively. For a squared effect size of $\beta^2 = 0.01$, each sample size, in part, specifies a value of d_ℓ, with $d = 4142, 3340, 3290$, or sample allele frequencies of approximately 0.2, 0.02, and 0.002, in order of increasing sample size. In (B-C), we restrict ourselves to $a = 10^{-3}$ as the lines for different mutation rates would otherwise largely coincide. In (B), we plot $mse_\ell(\tau)$ normalized by the expected additive genetic variance at stationarity, $\mathbb{E}[V_{A\ell}] = \beta^2(a/2a+1)$. In (C), we fix $n = 10^4$ and vary the detection threshold over several orders of magnitude, $d \in \{10, \ldots, 10^5\}$, plotting $mse_\ell(\tau)$ normalized by $\mathbb{E}[V_{A\ell}]$. In (D-F), we repeat (A-C), but for the statistic $\hat{V}_{A\ell}(\tau)$, with the following exception: Because $\hat{V}_{A\ell}(\tau)$ decreases with τ, we plot the absolute value of its difference from $\hat{V}_{A\ell}(0)$ in (A). For all plots the ancient sampling time $\tau \in [1, 0]$, which corresponds to a time span of $2N$ generations.

$\pm\beta_\ell$ but will not contribute to phenotypic variation. We illustrate this fact by conditioning on the allele frequencies of all loci in $\mathcal{L}$ at t, $\boldsymbol{Z}(t) \in [0,1]^L$. Assuming linkage equilibrium between loci as well as independence between the environmental and genetic effects, we have,

$$\mathbb{V}\left[Y_i(t)|\boldsymbol{Z}(t)\right] = 2\sum_{\ell=1}^{L} \beta_\ell^2 Z_\ell(t)(1 - Z_\ell(t)) + \sigma_e^2. \tag{2.2}$$

The summation in Eq. (2.2) is the additive genetic variance at t, $V_A(t)$. For a segregating site, the summand is proportional to $Z_\ell(t)(1 - Z_\ell(t))$, with $0 < Z_\ell(t)(1 - Z_\ell(t)) < 1$. For a fixed site, the summand is zero and the site does not contribute to the additive genetic variance $V_A(t)$. An important feature of our model is that some of the L loci may not exhibit genetic variation in the population at a given time. More concretely, the set of loci with non-zero estimated effects on the polygenic score, $\mathcal{S}$, may only be a small subset of $\mathcal{L}$. Thus, we assume that $\mathcal{L}$ is a superset of $\mathcal{S}$.

2.3.3 Constructing a model for the polygenic score

As our aim is to isolate the effects of allelic turnover on the statistical properties of polygenic scores, we make the additional assumption that the genotyped sites are the causal sites. (We have already assumed that all loci are in linkage equilibrium.) Akin to (Simons et al., 2018), we employ a simple threshold model for the effect estimates. For a GWA study consisting of n individuals (and $2n$ chromosomes),

$$\hat{\beta}_\ell := \begin{cases} \beta_\ell & \text{if } D_\ell \in (d_{\ell 1}, 2n - d_{\ell 2}), \\ 0 & \text{else,} \end{cases} \tag{2.3}$$

where D_ℓ is the allele count of the trait-increasing allele A_2 at the ℓ-th site in the GWA study sample; and $d_{\ell 1}$ and $d_{\ell 2}$ are the site-specific detection thresholds. In this simplified model, the true effect is estimated perfectly for all sites with allele counts within the intervals

$(d_{\ell 1}, 2n - d_{\ell 2})$ for $\ell \in \mathcal{L}$. In Section 2.8.4, we relate Eq. (2.3) to two alternative estimation procedures: maximum likelihood estimation (MLE) and the best linear unbiased predictor (BLUP).

We allow the two thresholds to differ in order to encompass scenarios in which power is an asymmetric function of the sample allele frequencies, e.g., there is more power to detect low frequency ($D_\ell < n$) versus high frequency ($D_\ell > n$) trait-increasing alleles. Such situations may arise with polygenic disease inheritance and imbalanced case and control sample sizes (Chan et al., 2014). In most cases, however, we will consider symmetric detection thresholds, with $d_{\ell 1} = d_{\ell 2} = d_\ell$. The threshold d_ℓ depends on on the phenotypic variance, genome-wide significance threshold, true per-locus effect β_ℓ, and GWA study size n. In Section 2.8.2, we give an explicit form for this dependency for a continuous focal trait and equal detection thresholds.Varying d_ℓ while keeping the GWA sample size fixed is equivalent to varying the true per-locus effect β_ℓ. Varying the GWA study size n while keeping β_ℓ and the other parameters fixed is akin to varying the GWA study's power to detect loci of a particular effect size. In **Analytical Results**, we do both.

The threshold model arises in the large GWA study size n limit for the model of $\hat{\beta}_\ell$ provided in Eq. (2.25). Namely, as long as D_ℓ is not too small, the variance of $\hat{\beta}_\ell$ goes to zero as n grows. Thus, the threshold model in Eq. (2.3) will necessarily underestimate the true variance of $\hat{\beta}$ (Section 2.8.4). Still, this model captures the dependency of $\hat{\beta}_\ell$ on the GWA study sample size n and the true per-locus effect β_ℓ, while facilitating our analytical treatment.

In order to compare the polygenic score with an individual's true phenotype, we need to account for all sites in the mutational target $\mathcal{L}$, not just those in $\mathcal{S}$, the set of sites with non-zero effect estimates in the polygenic score. As $\hat{\beta}_\ell = 0$ for any site in $\mathcal{L}$ but not $\mathcal{S}$, we express the polygenic score as a function of all loci in $\mathcal{L}$. The ancient polygenic score of

individual i sampled τ generations in the past is then given by,

$$\hat{Y}_i(\tau) \coloneqq \hat{C} + \sum_{\ell=1}^{L} X_{i\ell}(\tau)\hat{\beta}_\ell, \tag{2.4}$$

where $\hat{C}$ is the average phenotype of the GWA sample after subtracting the estimated genetic effects at all loci,

$$\hat{C} \coloneqq \bar{Y} - \sum_{\ell=1}^{L} \hat{\beta}_\ell \bar{X}_\ell, \tag{2.5}$$

with $\bar{Y} = \frac{1}{n}\sum_{j=1}^{n} Y_j$ and $\bar{X}_\ell = \frac{1}{n}\sum_{j=1}^{n} X_{j\ell}$ as the mean phenotype and genotype at locus ℓ in the GWA study sample, respectively. Here, and in the remainder of our study, we omit time-indexing for random variables associated with the GWA study at $t = 0$. By design, the estimated intercept $\hat{C}$ absorbs the effects of all loci which were not detected as significant in the GWA study, i.e., those sites for which $\hat{\beta}_\ell = 0$. Its presence in the polygenic score of Eq. (2.4) is necessitated by the fact that, to facilitate our analytical treatment, we did not center nor scale the genotypes and phenotypes in the GWA study. Importantly, all of our results are independent of this choice (Section 2.8.5). Henceforth, unless otherwise noted, we refer to Eq. (2.4) as the *polygenic score* and to the summation in Eq. (2.4) as the *genetic prediction*.

2.3.4 Modeling population genetic dynamics

Population genetic processes govern the correlations between allele frequencies at distinct points in time. We model this correlation using the Wright-Fisher diffusion with recurrent mutation. As we assumed all loci were in linkage equilibrium, their allele frequencies evolve forward in time independently, subject to genetic drift and mutation. At each site, alleles mutate from $A_1 \to A_2$ with rate μ, and from $A_2 \to A_1$ with rate ν. While our results readily generalize to arbitrary μ and ν, we restrict ourselves to equal mutation rates, $\mu = \nu$.

We further assume that the population is at equilibrium. In this setting, the marginal allele frequencies are beta-distributed, with shape and scale parameters specified by the population-scaled mutation rate; we denote the latter quantity by a, with $a = 4N\mu = 4N\nu$.

The relative magnitudes of mutation and genetic drift determine which force dominates an allele frequency trajectory. For example, as a approaches 0, the effects of mutation on the frequencies of segregating mutations become negligible and genetic drift dominates. In this low mutation regime ($a \ll 1$, or equivalently $\mu \ll \frac{1}{2N}$), the recurrent mutation model approximates the infinite sites model, while still retaining the features that make it attractive for our analytical treatment. In particular, the stationary allele frequency distribution is a well-defined probability distribution under the recurrent mutation model, but not under the infinite sites model. We concern ourselves almost exclusively with the low mutation regime.

2.3.5 Quantifying out-of-sample prediction errors

To quantify how well the polygenic score approximates the true phenotype of an individual sampled uniformly at random from the population at time τ before the present, we use several statistics:

Bias. We define the bias as the expectation of the difference between the polygenic score and true phenotype,

$$bias(\tau) \coloneqq \mathbb{E}\big[\hat{Y}(\tau) - Y(\tau)\big], \tag{2.6}$$

where, here and elsewhere, we omit the subscript when there is only one sample. The expectation in Eq. (2.6) is with respect to the entire random process, encompassing the underlying population genetic dynamics, estimation of the per-locus effects in the GWA study, and computation of the ancient polygenic score (illustrated in Fig. 2.1c).

Mean-squared error (*mse*). We define the *mse* as the expectation of the squared prediction

error,

$$mse(\tau) \coloneqq \mathbb{E}\left[\left(\hat{Y}(\tau) - Y(\tau)\right)^2\right]. \tag{2.7}$$

As in Eq. (2.6), the expectation in Eq. (2.7) is with respect to all sources of randomness in the model. The variance of the prediction error equals the difference of the *mse* and the square of the *bias*, and thus it is fully characterized by these two metrics.

Expected estimated additive genetic variance ($\hat{V}_A$). The estimated additive genetic variance is an estimate of the amount of phenotypic variance in the ancient population explained by additive genetic effects alone. We use $\hat{V}_A(\tau)$ to represent the expectation of this quantity,

$$\hat{V}_A(\tau) \coloneqq \sum_{\ell=1}^{L} \hat{V}_{A\ell}(\tau) = 2\sum_{\ell=1}^{L} \mathbb{E}\left[\hat{\beta}_\ell^2 \hat{Z}_\ell(\tau)(1 - \hat{Z}_\ell(\tau))\right], \tag{2.8}$$

where $\hat{Z}_\ell(\tau)$ is an estimate of the ancient population allele frequency computed from a sample of n_a individuals sampled at τ. The expected true additive genetic variance, $\mathbb{E}[V_A]$, can be found by taking the expectation of the summation in Eq. (2.2).

Polygenic score accuracy (ρ^2). Practitioners often compute the sample correlation coefficient r^2 to measure the accuracy of a predictor in a sample. Here, our sample is n_a ancient individuals sampled at time τ, thus,

$$r^2(\tau) \coloneqq \frac{Cov[\hat{\boldsymbol{Y}}(\tau), \boldsymbol{Y}(\tau)]^2}{Var[\hat{\boldsymbol{Y}}(\tau)]Var[\boldsymbol{Y}(\tau)]}, \tag{2.9}$$

where $Cov[\cdot,\cdot]$ and $Var[\cdot]$ are the sample covariance and variance operators, respectively, and $\hat{\boldsymbol{Y}}(\tau), \boldsymbol{Y}(\tau) \in \mathbb{R}^{n_a}$ are the n_a-dimensional vectors of polygenic scores and phenotypes of the ancient individuals, respectively. Ideally, we would compute the expectation of this quantity—but, this is challenging due to the common difficulty of computing an expectation of a ratio of random variables. Thus, we approximate the expectation of $r^2(\tau)$ as the ratio

of expectations,

$$\mathbb{E}\left[r^2(\tau)\right] \approx \frac{\mathbb{E}[Cov[\hat{\boldsymbol{Y}}(\tau), \boldsymbol{Y}(\tau)]]^2}{\mathbb{E}[Var[\hat{\boldsymbol{Y}}(\tau)]]\mathbb{E}[Var[\boldsymbol{Y}(\tau)]]} =: \rho^2(\tau), \tag{2.10}$$

where, as above, the covariance and variances are taken with respect to the sample of n_a ancient individuals, while the expectation is over all sources of randomness in Fig. 2.1c (see Section 2.8.9 for more details). We present simulations in the section *Polygenic score accuracy* of **Analytical Results** showing that $\rho^2(\tau)$ is a good approximation for the expectation of $r^2(\tau)$ in the parameter regimes of interest.

2.4 Analytical Results

By how much does the prediction accuracy of a polygenic score decrease as the time between sampling the ancient individual and conducting the GWA study increases? To answer this question, we consider a trait potentially influenced by L genetic loci, each with true effect $\beta_\ell \geq 0,\ \ell = 1, \ldots, L$. The forward evolution of sites underlying this trait is modulated by a per site, per generation mutation rate, μ, and a population scaled rate of $a = 4N\mu$. The diploid population of size $2N$ chromosomes is assumed to be at equilibrium. The parameters dictating the GWA study are the sample size n and the detection thresholds specified by $\boldsymbol{d}_1, \boldsymbol{d}_2 \in \{1, \ldots, n\}^L$. The metrics are indexed by the ancient sampling time τ in coalescent time-units. An ancient sampling time of τ corresponds to $2N \cdot \tau$ generations in the past. We omit the time index for variables associated with the GWA study, which occurs at present day ($t = 0$). (We show in Section 2.8.13, that the metrics can also be expressed as a function of divergence or F_{ST} between the ancient and contemporary populations.)

Each subsection is structured as follows: We first derive a general expression for the statistic that does not depend on how we model the population genetic dynamics nor the GWA study. Second, we derive an analytical expression for the statistic under the population

genetic assumptions and the GWA study threshold model described in **Model and metrics**.

2.4.1 Bias

We can rewrite the sampling time-dependent bias defined in Eq. (2.6) as,

$$bias(\tau) = \sum_{\ell=1}^{L} bias_\ell(\tau) = \sum_{\ell=1}^{L} \mathbb{E}\left[\left(\bar{X}_\ell - X_\ell(\tau)\right)\left(\beta_\ell - \hat{\beta}_\ell\right)\right], \tag{2.11}$$

where $bias_\ell(\tau)$ is the contribution of locus ℓ to $bias(\tau)$. From Eq. (2.11), we see that $bias_\ell(\tau) \approx 0$ when either or both of $\hat{\beta}_\ell \approx \beta_\ell$ and $\bar{X}_\ell \approx X_\ell(\tau)$ are true. Thus, $bias_\ell(\tau)$ is minimal when (i) effect estimates are accurate, and (ii) the allele frequencies have not changed substantially in the interval $[\tau, 0]$.

Under the assumption of equal mutation rates and detection thresholds ($d_{\ell 1} = d_{\ell 2}$), $bias_\ell(\tau) = 0$ for $\tau \geq 0$ for a reason distinct from those stated above. Trait-increasing alleles at high frequencies ($D_\ell > n$) and low frequencies ($D_\ell < n$) are detected as significant ($\hat{\beta}_\ell \neq 0$) with equal probability. An equivalent assumption is that power is not affected by whether the most prevalent allele is trait-increasing or decreasing. Subsequent evolution of the allele frequencies preserves this symmetry and $bias(\tau)$ remains equal to zero for all τ. It follows that in the absence of additional perturbing forces, an estimate of the mean polygenic score from a sample of n_a ancient individuals will also be unbiased, and therefore will on average accurately reflect the lack of change in the mean phenotype.

However, if we introduce asymmetry in the detection thresholds ($d_{\ell 1} \neq d_{\ell 2}$), $bias(\tau)$ is non-zero for all τ (Section 2.8.7). Using the spectral representation of the transition density of the Wright-Fisher diffusion (*tdf*), we derive the per-locus contribution to the bias, $bias_\ell(\tau)$ (Section 2.8.7). For a small population-scaled mutation rate a and a large GWA study size

n, we approximate this expression (given in Eq. (2.65)) as,

$$bias_\ell(\tau) \approx \left(e^{-a\tau} - 1\right)\left(P^{(d_{\ell 1})} - P^{(d_{\ell 2})}\right), \tag{2.12}$$

where,

$$P^{(d_{\ell i})} = \sum_{i=0}^{d_{\ell i}-1} \binom{2n}{i} \frac{B(a+i, a+2n-i)}{B(a,a)} \tag{2.13}$$

is the probability that the allele count of site ℓ is less than $d_{\ell i}$, i.e., $D_\ell < d_{\ell i}$ for $i = 1, 2$; and, $B(\cdot, \cdot)$ is the beta function. Thus, the magnitude of $bias_\ell(\tau)$ is approximately proportional to the difference in the probability of detecting high ($D_\ell > n$) versus low ($D_\ell < n$) frequency alleles, and increases exponentially with τ. With a large GWA study size n and a small mutation rate a, this difference is small relative to the square root of the additive genetic variance—the ratio of these two quantities is smaller than $\mathcal{O}(a)$ (Fig. 2.5a). This is due to the fact that when the mutation rate is small, most alleles are close to fixation or fixed. The stationary population allele frequency density $\kappa(z) \propto z^{a-1}(1-z)^{a-1}$ behaves like $z^{-1}(1-z)^{-1}$ for small a. Varying $d_{\ell i}$ then has relatively little impact on $P^{(d_{\ell i})}$, constraining the difference between the one-sided detection probabilities (Fig. 2.5b).

2.4.2 Mean-squared error

The sampling time-dependent mean-squared error $mse(\tau)$ can be expressed as,

$$\begin{aligned} mse(\tau) &= \sum_{\ell=1}^{L} mse_\ell(\tau) + \left(\frac{n-1}{n}\right)\sigma_e^2 \\ &= \sum_{\ell=1}^{L} \mathbb{E}\left[\left(X_\ell(\tau) - \bar{X}_\ell\right)^2 \left(\hat{\beta}_\ell - \beta_\ell\right)^2\right] + \left(\frac{n-1}{n}\right)\sigma_e^2, \end{aligned} \tag{2.14}$$

where σ_e^2 is the variance in the phenotype due to the environment (Section 2.8.7). Note the similarity of the left term in Eq. (2.14) to the form of $bias(\tau)$ given in Eq. (2.11)—

similar heuristics apply. Under the threshold model specified in Eq. (2.3), sites at moderate frequencies in the GWA study sample, $D_\ell \in [d_\ell, 2n - d_\ell]$, will not contribute to $mse(\tau)$ since $\hat{\beta}_\ell = \beta_\ell$. Only sites with frequencies outside this interval (including sites invariant in the GWA study sample) will contribute, and their contributions will be proportional to the squared difference between $X_\ell(\tau)$ and $\bar{X}_\ell$. In practice, moderate frequency loci will also contribute to $mse(\tau)$ due to errors in the estimation of the effect estimates and any difference between the ancient genotypes and the average genotypes in the GWA study sample at these sites (Section 2.8.4).

We use the spectral representation of the *tdf* (Section 2.8.6) to derive an analytical expression for $mse_\ell(\tau)$, the per-locus contribution to the *mse* (Section 2.8.7). From this expression, Eq. (2.70), we derive a linear approximation for the initial per-locus increase in this statistic, $\Delta mse_\ell(\tau)$. With a symmetric detection threshold ($d_{\ell 1} = d_{\ell 2} = d_\ell$) we have,

$$\Delta mse_\ell(\tau) \coloneqq mse_\ell(\tau) - mse_\ell(0) \approx 2\beta_\ell^2 a P^{(d_\ell)} \tau, \tag{2.15}$$

where $mse_\ell(0)$ is the contribution of site ℓ to $mse(\tau)$ for $\tau = 0$ (Eq. (2.96)); and $2P^{(d_\ell)}$, defined in Eq. (2.13), is the probability that the allele count of site ℓ is outside the detection interval such that $\hat{\beta}_\ell = 0$. Both $mse_\ell(0)$ and $P^{(d_\ell)}$ depend on the mutation rate a, the GWA study size n, and the detection threshold d_ℓ.

$\Delta mse_\ell(\tau)$ reflects the time-dependent contributions of sites *not* detected in the GWA study. To see this, we condition on the effect estimate $\hat{\beta}_\ell$, $mse_\ell(\tau) = \beta_\ell^2 \mathbb{E}\big[(X_\ell(\tau) - \bar{X}_\ell)^2 | \hat{\beta}_\ell = 0\big] \cdot 2P^{(d_\ell)} + 0 \cdot (1 - 2P^{(d_\ell)})$. Thus, Eq. (2.15) implies that $\frac{d\mathbb{E}\left[(X_\ell(\tau) - \bar{X}_\ell)^2 | \hat{\beta}_\ell = 0\right]}{d\tau} \approx a$ for small τ, and consequently, the combined effects of drift and mutation on $mse_\ell(\tau)$ are captured in the product of the mutation rate and sampling time $a\tau$.

In addition, Eq. (2.15) suggests that the rate at which $mse_\ell(\tau)$ increases will be shared across parameter regimes when $aP^{(d_\ell)}$ is similar (Fig. 2.4a). To illustrate this, we use our analytic formula (given in Eq. (2.70)) to compute $mse_\ell(\tau)$ for several low mutation rates,

$a \in \{10^{-4}, 10^{-3}, 10^{-2}\}$, and three GWA study sizes, $n \in \{10^4, 10^5, 10^6\}$ (Fig. 2.2a). These mutation rates and sample sizes span the range of parameter values appropriate for human data. We depict our results in two ways: (i) we plot the change in $mse_\ell(\tau)$, and (ii) we plot $mse_\ell(\tau)$ normalized by the expected additive genetic variance contributed by a single site. At stationarity the expected additive genetic variance is constant and equal to,

$$\mathbb{E}\left[V_{A\ell}\right] = \mathbb{E}\left[2\beta_\ell^2 Z_\ell(1 - Z_\ell)\right] = \beta_\ell^2(a/(2a+1))$$

for a scaled-mutation rate a. The plot of the former, Fig. 2.2a, exhibits the functional relationship revealed by Eq. (2.15), while the latter, Fig. 2.2b, approximates the noise-to-signal ratio. In Section 2.8.11, we demonstrate that Eq. (2.15) is a good approximation to $mse(\tau)$ for $\tau \leq 0.2$, particularly when the GWA study size n is large (in particular, see Fig. 2.9).

To find the GWA study size specific detection thresholds used in Figs. 2.2a and 2.2b, we solve Eq. (2.31) for a given effect size β, phenotypic variance V_p, and significance threshold α, while varying the GWA study sample size. For $\beta^2 = 0.01$, $V_p = 1$, and $\alpha = 10^{-8}$, the detection thresholds are $d = 4142, 3340, 3290$ in order of increasing sample size, which corresponds to sample allele frequencies of approximately 0.2, 0.02, amd 0.002, respectively. Thus, for a given effect size, larger sample sizes will lead to the detection of alleles at more extreme allele frequencies, while smaller samples will restrict detection to alleles at more intermediate frequencies. Due to non-identifiability, the parameter choices are fairly arbitrary.

We find that for small mutation rates, the cumulative change in the *mse*, $\Delta mse_\ell(\tau)$, is mostly insensitive to differences in the GWA study sample size (Figs. 2.2a and 2.2b). The approximation in Eq. (2.15) helps to explain this result. The rate of increase is approximately proportional to $2aP^{(d_\ell)}\tau$. For small mutation rates ($a \ll 1$) and an arbitrary detection threshold d_ℓ, the probability of *not* detecting a locus as significantly associated with the

trait is roughly $2P^{(d_\ell)} \approx 1$ for all sufficiently large n (Fig. 2.5b). In this regime, increasing the GWA study sample size only yields small increases in the probability of detecting a locus as significant. Thus, for small mutation rates, the product of this quantity with the mutation rate is $2aP^{(d_\ell)} \approx a$, and indeed, we observe a cumulative increase in $mse_\ell(\tau)$ that is $\mathcal{O}(a)$ for $\tau = 1$ (Fig. 2.2a). We note that increasing the GWA study sample size does enable detection of loci with smaller effects.

The result in Fig. 2.2a, however, hides the fact that a small absolute increase in $mse(\tau)$ may correspond to a substantial increase in the noise-to-signal ratio. Indeed, for $a = 10^{-3}$ (blue lines throughout), $mse_\ell(\tau)$ ultimately exceeds the expected additive genetic variance $\mathbb{E}[V_{A\ell}]$ for all GWA study sample sizes (Fig. 2.2b). By $\tau = 0.2$, a sampling time characteristic of ancient humans, $mse_\ell(\tau)$ due to allelic turnover is approximately 20% of the additive genetic variance $\mathbb{E}[V_{A\ell}]$. For sufficiently large τ, $mse_\ell(\tau)$ is at least the same order of magnitude as the expected additive genetic variance. In addition, while $mse_\ell(\tau)$ increases at approximately the same rate irrespective of study size, its initial value $mse_\ell(0)$ is sample size dependent (Fig. 2.2b and see Figs. 2.4b and 2.4e for a larger parameter space). Yet, for a given value of d_ℓ, reductions in $mse_\ell(0)$ mediated by sample size diminish once n is large enough (Figs. 2.4b and 2.4e).

Further, Fig. 2.2a obscures the fact that different mutation rates may yield similar noise-to-signal ratios. As discussed, for small a, $mse_\ell(\tau)$ increases with τ at a rate that is $\mathcal{O}(a)$. For small a, the additive genetic variance is likewise $\mathcal{O}(a)$, yielding a relative increase that is mostly insensitive to the mutation rate. Normalized $mse_\ell(0)$ is also similar across small mutation rates (Figs. 2.4b and 2.4e), rendering relative $mse_\ell(\tau)$ mostly insensitive to a. We thus omitted the other two mutation rates from Fig. 2.2b.

Lastly, we fix the GWA study sample size at $n = 10^5$ and vary the detection threshold d (Fig. 2.2c). Varying d while keeping n fixed is analogous to varying the true per-locus effect size β, or keeping β fixed while varying the significance threshold α. The minimum

threshold is $d = 10$, whereas $d = n = 10^5$ maximizes $mse_\ell(\tau)$ since $\hat{\beta}_\ell$ would equal zero for all ℓ. Consistent with our analysis above, for small a, (i) $mse_\ell(0)$ depends critically on d, while (ii) $mse_\ell(\tau)$'s approximately linear growth rate is largely insensitive to d. Furthermore, by our previous arguments, relative $mse_\ell(\tau)$ is similar across small mutation rates, and they are also omitted in Fig. 2.2c. For independent and identically distributed (*iid*) loci and $\sigma_e^2 = 0$, the per-locus $mse_\ell(\tau)$ values presented in Figs. 2.2b and 2.2c are equal to the corresponding trait-wide statistics $mse(\tau)$.

2.4.3 Additive genetic variance

The per-locus contribution to the expected estimated additive genetic variance $\hat{V}_A(\tau)$ is,

$$\hat{V}_{A\ell}(\tau) = 2\mathbb{E}\left[\hat{\beta}_\ell^2 \hat{Z}_\ell(\tau)(1 - \hat{Z}_\ell(\tau))\right] = 2\left(\frac{2n_a - 1}{2n_a}\right)\mathbb{E}\left[\hat{\beta}_\ell^2 Z_\ell(\tau)(1 - Z_\ell(\tau))\right], \tag{2.16}$$

where $\hat{Z}(\tau) = \frac{1}{2n_a}\sum_{i=1}^{n_a}(X_i(\tau) + 1)$ is the estimated allele frequency at τ, computed in a sample of n_a ancient individuals. When $\hat{\beta}_\ell = 0$ or $Z_\ell(\tau) \in \{0, 1\}$, site ℓ will not contribute to $\hat{V}_A(\tau)$. Thus, a site ℓ has a non-zero contribution to the estimated additive genetic variance only when it is segregating at both the present day and τ. This condition is necessary for both $\hat{Z}_\ell(\tau)(1 - \hat{Z}_\ell(\tau)) > 0$ and $\hat{\beta}_\ell \neq 0$ to be true.

As with the two previous statistics, we use the spectral representation of the *tdf* to derive an analytical expression for $\hat{V}_A(\tau)$ under our population genetic assumptions (Section 2.8.8). The resulting expression, Eq. (2.74), indicates that the expected additive genetic variance decays exponentially. We then, to first order in the ancient sampling time τ, approximate the initial decrease in the per-locus estimated additive genetic variance $\Delta\hat{V}_{A\ell}(\tau)$,

$$\Delta\hat{V}_{A\ell}(\tau) \coloneqq \hat{V}_{A\ell}(\tau) - \hat{V}_{A\ell}(0) = -2\left(\frac{2n_a - 1}{2n_a}\right)\beta_\ell^2 a P^{(d_\ell)}\tau, \tag{2.17}$$

where $\hat{V}_{A\ell}(0)$ is $\hat{V}_{A\ell}(\tau)$ evaluated at $\tau = 0$ (Eq. (2.97)); and $2P^{(d_\ell)}$, defined in Eq. (2.6), is

the probability that $\hat{\beta}_\ell = 0$. The factor due to finite sampling, $2n_a/(2n_a - 1)$, is ≈ 1 when the ancient sample size n_a is large. Thus, apart from sign, $\Delta\hat{V}_{A\ell}(\tau)$ is equal to $\Delta mse_\ell(\tau)$ of Eq. (2.15). Therefore, for small τ, $\hat{V}_A(\tau)$ decreases at approximately the same rate as $mse(\tau)$ increases. This result further suggests that for $a \ll 1$ and a large GWA study size n, $\hat{V}_{A\ell}(\tau)/\mathbb{E}[V_{A\ell}] \approx 1 - mse_\ell(\tau)/\mathbb{E}[V_{A\ell}]$ for small τ (Figs. 2.2c and 2.2f). Although, this relationship trivially breaks down for large τ as $mse_\ell(\tau)$ is not bounded by one.

To compare $\hat{V}_{A\ell}(\tau)$ across mutation rates, we mirror our treatment of $mse_\ell(\tau)$ in the previous section. We plot (i) its increase $\Delta\hat{V}_{A\ell}(\tau)$ (Fig. 2.2d); (ii) $\hat{V}_{A\ell}(\tau)$ normalized by the expectation of the true additive genetic variance at stationarity (Fig. 2.2e); and (iii) normalized $\hat{V}_{A\ell}(\tau)$, varying the detection threshold for a fixed GWA study sample size (Fig. 2.2f). Akin to $mse_\ell(\tau)$, normalized $\hat{V}_{A\ell}(\tau)$ is very similar across small mutation rates. And, while the GWA study size n and the detection threshold d influence the initial estimated additive genetic variance $\hat{V}_{A\ell}(0)$, its rate of change is mostly insensitive to the two GWA study parameters.

As $\hat{V}_A(\tau)$ largely recapitulates our results for $mse(\tau)$ with opposing sign, we focus on their differences. Indeed, they have different functional forms and behave differently for modest or large τ (see Eqs. (2.70) and (2.74), respectively). Conceptually, this discrepancy is not unexpected: In the previous section, we showed that a site only contributes to $mse(\tau)$ if its allele count falls outside the detection interval and $\hat{\beta}_\ell = 0$. Thus, $mse(\tau)$ increases with τ due to alleles shifting from intermediate frequencies in the ancient population to frequencies *outside* of the detection region in the contemporary population. For the expected estimated additive genetic variance $\hat{V}_A(\tau)$, the converse is true: The slope represents the decline in $\hat{V}_A(\tau)$ due to alleles changing from frequencies near or at fixation in the ancient population to frequencies *within* the detection interval in the contemporary population. While our results reveal similar functional behavior for these two quantities (with opposing signs) that applies for small τ, we caution that statements about $\hat{V}_A(\tau)$ do not immediately translate

to statements about $mse(\tau)$, particularly for $\tau \gtrapprox 0.2$.

2.4.4 Polygenic score accuracy

While our framework, in principle, encompasses a trait with varying effect sizes, we will first assume that all sites are *iid* with true effect size β. Our approximation to the expectation of the sample correlation coefficient simplifies to,

$$\rho^2(\tau) = \frac{L\beta\mathbb{E}\left[\hat{\beta}(X(\tau) - \bar{X}(\tau))^2\right]}{L\beta^2\mathbb{E}\left[(X(\tau) - \bar{X}(\tau))^2\right] + \sigma_e^2} = \frac{\mathbb{E}\left[\hat{\beta}Z_\ell(\tau)(1 - Z_\ell(\tau))\right]/\beta}{\mathbb{E}\left[Z_\ell(\tau)(1 - Z_\ell(\tau))\right] + \sigma_{e'}^2}, \tag{2.18}$$

where the compound parameter $\sigma_{e'}^2 = \sigma_e^2/L\beta^2$ is the environmental variance normalized by the product of the number of loci in the mutational target L and the squared per-locus effect size β (Section 2.8.9). By comparing Eq. (2.18) with Eq. (2.16), we can see that $\rho^2(\tau)$ is closely related to the estimated additive genetic variance. Thus, like $\hat{V}_A(\tau)$, $\rho^2(\tau)$ will decrease with τ due to loci having changed from frequencies close to zero or one in the ancient population to intermediate frequencies in the contemporary population. However, unlike $\hat{V}_A(\tau)$, $\rho^2(\tau)$ does not depend on the ancient sample size. Therefore, to relate the two statistics, we multiply by the inverse of the ancient sample size dependent factor implicit in $\hat{V}_A(\tau)$,

$$\rho^2(\tau) = \left(\frac{2n_a}{2n_a - 1}\right)\frac{\hat{V}_{A\ell}(\tau)/\beta^2}{\mathbb{E}\left[V_{A\ell}(\tau)\right]/\beta^2 + \sigma_{e'}^2}. \tag{2.19}$$

For $\sigma_e^2 = 0$, barring the sample size factor, Eq. (2.19) is equal to $\hat{V}_A(\tau)$ normalized by the expected additive genetic variance. By extension, this quantity approximates the expected sample correlation coefficient $r^2(\tau)$. By invoking our additional population genetic and GWA study assumptions, we arrive at an approximation for the decrease in polygenic score accuracy,

$$\Delta\rho^2(\tau) \coloneqq \rho^2(\tau) - \rho^2(0) \approx -\frac{2aP^{(d_\ell)}\tau}{\frac{a}{2a+1} + \sigma_{e'}^2}. \tag{2.20}$$

Now, to relate our theory to empirical and simulation studies, we compute $\rho^2(\tau)$ for a given narrow-sense heritability h^2 and mutation rate a pair. We define h^2 for a trait with a mutational target of L loci of equal effects β,

$$h^2 \coloneqq \frac{\mathbb{E}\left[V_A\right]}{\mathbb{E}\left[V_A\right] + \sigma_e^2} = \frac{a/(2a+1)}{a/(2a+1) + \sigma_{e'}^2},$$

where the equality follows from our population genetic assumptions. Together with a, h^2 fully specifies the compound parameter $\sigma_{e'}^2$ with,

$$\sigma_{e'}^2 = \left(\frac{a}{2a+1}\right)\left(\frac{1-h^2}{h^2}\right).$$

We plot our analytical expressions for both accuracy (Fig. 2.3a) and relative accuracy (Fig. 2.3b), defined as the ratio of $\rho^2(\tau)$ to $\rho^2(0)$ for $\tau \in [1, 0]$ spanning $2N$ generations. For humans, this time span corresponds to approximately 500,000 years in the past, encompassing the "Out-of-Africa" migration event estimated to have occurred 50,000-100,000 years ago (Jouganous et al., 2017). As with the preceding statistics, when $\tau = 0$, $\rho^2(\tau)$ approximates the accuracy of the polygenic score *within* the GWA study population. Relative accuracy then directly measures reductions in accuracy relative to the GWA study population. We set $h^2 = 0.5$ and $a = 10^{-3}$, and fix the GWA study sample size at $n = 10^5$. We then compute $\rho^2(\tau)$, varying the detection threshold over several orders of magnitude (Fig. 2.3a). (See Fig. 2.10 for accuracy as a function of the fixation index, or F_{ST}.) Our results for $\rho^2(\tau)$ necessarily recapitulate those of $\hat{V}_A(\tau)$: While increasing the detection threshold d reduces accuracy substantially, it does not have a large impact on relative accuracy for $n = 10^5$ (Fig. 2.3a). Indeed, for small mutation rates, relative accuracy is insensitive to the mutation rate and threshold, and is well approximated by $e^{-\tau}$ (Eq. (2.88)). Thus, its derivative is also exponential. Absolute accuracy $\rho^2(\tau)$ likewise decays exponentially, but its derivative

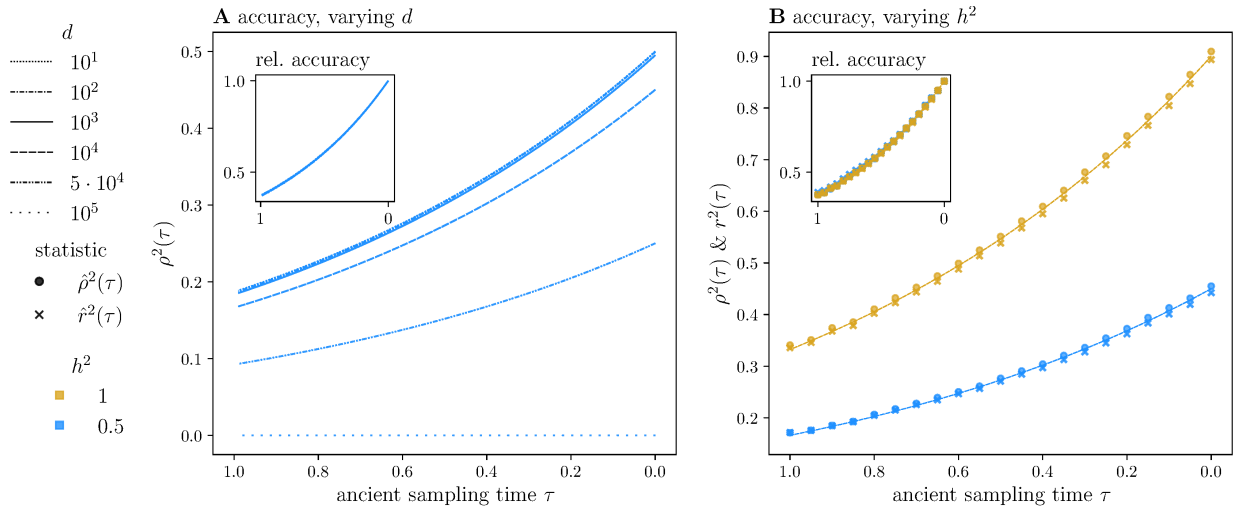


Figure 2.3: **Polygenic score accuracy.** We plot our theoretical results for both absolute (A, main) and relative accuracy $\rho^2(\tau)$ (A, inset) for ancient sampling times $\tau \in [1, 0]$ (or a time span of $2N$ generations) with a mutation rate of $a = 10^{-3}$. The GWA study size is shared in all plots, with $n = 10^5$. In (A), we vary the detection threshold over the range of possible values, $d_\ell \in \{10, \dots 10^5\}$. In (B), we compare our theoretical expectations with simulated estimates of the approximate sample correlation coefficient $\rho^2(\tau)$ (circles) and the statistic itself $r^2(\tau)$ (crosses) for a threshold of $d = 10^4$ (a minimum sample allele frequency of 0.05), and two values of heritability, $h^2 = 0.5, 1$ (in blue and gold, respectively). The ancient sample size is $n_a = 100$. In the inset of (B), we normalize the estimates by their initial (estimated) values. Theoretical expressions for $\rho^2(\tau)$ are also plotted in (B). Each simulated point is the average of $K = 5000$ simulations of $L = 5000$ *iid* loci.

is scaled by a quantity that reflects features of the GWA study and the phenotypic variance. For a small mutation rate $a \ll 1$, its derivative is approximately $2P^{(d)}(a/(a+\sigma^2_{e'}))e^{-\tau}$, which, in turn, is approximately $2P^{(d)}h^2e^{-\tau}$ (Eq. (2.87)). The latter expression suggests that the probability of not detecting a significant association $P^{(d)}$ and trait heritability h^2 are the key determinants of prediction accuracy. Importantly, $\rho^2(\tau)$ declines considerably over the interval $\tau \in [1, 0]$ irrespective of the detection threshold d.

In addition, we glean from Eq. (2.18) that while heritability affects the magnitude of $\rho^2(\tau)$ through the compound parameter $\sigma^2_{e'}$, it does not influence the relative accuracy, consistent with previous results (Wang et al., 2020). Our simulations suggest that this is also true of the sample correlation coefficient, as simulated estimates of $r^2(\tau)$ agree extremely well with our theory for $\rho^2(\tau)$ (Fig. 2.3b). We note that this result is contingent on the fact that the environmental variance σ^2_e only enters our simple threshold model in the specification of the threshold d (Eq. (2.31)), and does not contribute directly to the variance of the polygenic score (Section 2.8.9). Therefore, we expect this result to hold only for large GWA study sample sizes for which the threshold model is a good approximation to the distribution of $\hat{\beta}$. While the finding that relative accuracy is insensitive to the GWA study parameters relies on the assumption that all loci are *iid* and share a causal effect β, we provide preliminary theoretical evidence that our results will hold when β varies across loci (see Eq. (2.89) and ensuing comments).

2.5 Simulation results for recent directional selection

We use simulations to explore if and how the statistics under study deviate from their neutral expectations in the presence of recent directional selection. Each copy of the A_2 allele at the ℓ-th site confers a fitness advantage of $+s_\ell$, and so the fitness ratio of the three possible genotypes A_1A_1:A_1A_2:A_2A_2 is 1:$(1+s_\ell)$:$(1+2s_\ell)$. In our simulations, the population evolves neutrally until the onset of selection at N generations (or $\tau_s = 0.5$ in coalescent time

units) before present. Thereafter, the population evolves according to discrete Wright-Fisher dynamics with selection.

In the presence of selection, the allele frequency distribution is no longer symmetric; rather, it is skewed toward the beneficial allele. The severity of the skew depends on the selection coefficient and mutation rate, as well as the amount of time that selection has been acting. As we restrict s_ℓ to positive values, designating the A_2 or + allele as beneficial, the allele frequency distribution will be skewed toward one. If we instead designated the A_1 allele as the beneficial allele, the allele frequency distribution would be skewed toward zero. The former models "positive" selection whereas the latter models "negative" selection. Because $bias(\tau)$ is proportional to β, its sign will be sensitive to this choice, but its magnitude will be unaltered. The other statistics will not be affected as long as the detection thresholds are symmetric. Therefore, our results are general up to the sign of $bias(\tau)$.

We conduct simulations over a range of selection coefficients, $\sigma = 4Ns \in \{0, 0.1, 1, 10\}$, for a mutation rate of $a = 10^{-3}$. Under directional selection, σ is proportional to the locus effect size β; mutations with larger effect sizes will be more likely to establish and achieve appreciable frequencies (Chevin and Hospital, 2008). In addition, we plot results for two different detection thresholds, $d \in \{10^3, 10^4\}$, in a GWA study sample of size $n = 10^4$. More details on the simulation procedures are provided in Section 2.8.3.

When $\sigma \geq 1$, the polygenic score is biased towards positive values for $\tau > 0$ for both detection thresholds (Fig. 2.4a). In other words, with directional selection acting to increase the trait value, $\hat{Y}(\tau)$ tends to overestimate $Y(\tau)$. The magnitude of $bias_\ell(\tau)$ depends critically on the strength of selection relative to mutation: We observe a larger bias for $\sigma = 10$ relative to $\sigma = 1$, and likewise the bias is larger for $\sigma = 1$ relative to $\sigma = 0.1$. In fact, the smaller selection coefficient $\sigma = 0.1$ is not distinguishable from neutral expectations. For $0 \leq \tau < \tau_s$, $bias_\ell(\tau)$ increases at an accelerating rate; for $\tau \geq \tau_s$, $bias(\tau)$ appears constant in this parameter regime.

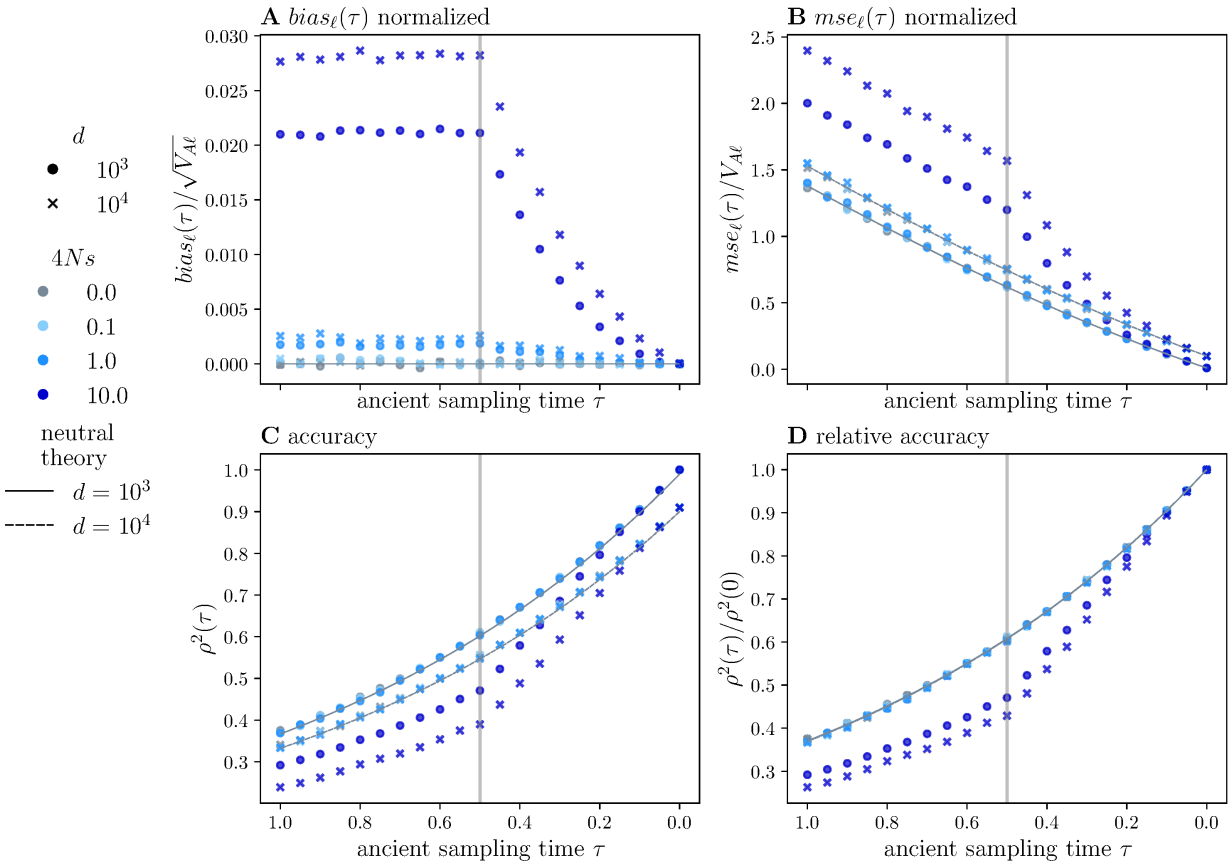


Figure 2.4: **Ancient polygenic scores in the presence of genic selection.** We conduct $K = 5000$ simulations, each with a mutational target of $L = 5000$ loci, in a population of size $2N = 2 \cdot 10^3$, with a population-scaled mutation rate, $a = 10^{-3}$. We consider four selection coefficients, $\sigma = 4Ns \in \{0, 0.1, 1, 10\}$ (indicated by color). The GWA study sample size is $2n = 2 \cdot 10^5$, with d equal to either 10^3 or 10^4. In (A-D), we plot the various simulated statistics along with their neutral expectations (solid or dashed black lines). The vertical gray lines indicate the onset of selection at $\tau_s = 0.5$ which corresponds to $N = 1000$ generations. The ancient sample times are $\tau \in [1, 0]$, corresponding to a time span of $2N = 2000$ generations. We computed, but did not plot, 95% confidence intervals for $bias(\tau)$, $mse(\tau)$, and $r^2(\tau)$, as they largely overlapped with the symbols. We note that the oscillations observed in (A) and (B) are not statistically significant.

A higher detection threshold decreases the detection probability. Thus, we expect that the magnitude of $bias_\ell(\tau)$ will increase with the detection threshold. Indeed, $bias_\ell(\tau)$ is larger and increases more quickly for the larger detection threshold $d = 10^4$ compared to $d = 10^3$ (Fig. 2.4a). Further, our simulations suggest that the detection threshold coupled with the time of the onset of selection govern the magnitude of the bias for $\tau > \tau_s$. For some large τ, $bias_\ell(\tau)$ will reach an equilibrium value that depends approximately on the asymmetry of the detection thresholds at the present day, which in turn, depends on both the timing and strength of selection (Section 2.8.12).

The underlying allele frequency dynamics provide some insight into these patterns. Before the onset of selection, the allele frequency distribution is stationary and symmetric around 0.5. After the onset of selection, trait-increasing alleles tend to increase in frequency, skewing the distribution toward one. Thus, alleles *not* detected in the GWA study will tend be at higher versus lower frequencies at $t = 0$, yielding $\mathbb{E}[\bar{X}_\ell|\hat{\beta}_\ell = 0] > 0$ for $\sigma > 0$. For large τ, the allele frequencies of sites not detected in the GWA study, i.e., with $\hat{\beta}_\ell = 0$, may have been substantially different in the ancient population. Each one of these sites will make a contribution to $bias(\tau)$ that is proportional to $\beta_\ell \mathbb{E}\big[(\bar{X}_\ell - X_\ell(\tau))|\hat{\beta}_\ell = 0\big]$ (Eq. (2.11)). Looking backward in time, the shift in the allele frequency distribution ensures that the conditional expectation of $X_\ell(\tau)$ is smaller than that of $\bar{X}_\ell$, yielding a positive $bias_\ell(\tau)$ for $\tau > 0$. Notably, the magnitude of $bias_\ell(\tau)$ induced by selection is several orders of magnitude larger than that induced by asymmetry in the detection threshold alone (Fig. 2.5a).

The effects of selection on $mse_\ell(\tau)$ are qualitatively consistent with those on $bias_\ell(\tau)$ (Fig. 2.4b). Although, here, the only selection coefficient which induces significant deviations from neutral expectations is $\sigma = 10$. And, $mse(\tau)$ is larger for $d = 10^4$ compared to $d = 10^3$. As with $bias(\tau)$, for $0 \leq \tau < \tau_s$, $mse_\ell(\tau)$ increases at an accelerating rate; before τ_s ($\tau \geq \tau_s$), $mse_\ell(\tau)$ appears to increase linearly. Values of $\sigma < 10$ do not induce noticeable deviations from neutrality for the correlation coefficient $\rho^2(\tau)$ either (Fig. 2.4c).

However, strong selection ($\sigma = 10$) does lead to substantially larger reductions in accuracy relative to our neutral expectations. In addition, for $\sigma = 10$, relative accuracy is sensitive to the detection threshold, with accuracy decreasing faster for the larger detection threshold (Fig. 2.4d).

2.6 Discussion

In this work, we devised a theoretical framework to quantify the effect of allelic turnover on the error and accuracy of out-of-sample polygenic scores. Unlike previous theoretical approaches (Wang et al., 2020; Daetwyler et al., 2008), we averaged over the evolutionary process governing trait evolution, the GWA study from which a polygenic score model is constructed, and the ancient individual's genotype and phenotype. In doing so, we found explicit expressions for several commonly used metrics that depend on the focal individual's sampling time, as well as the parameters governing the population genetic dynamics and power to detect trait-associated loci in the GWA study. Mathematical properties of the recurrent mutation model at stationarity enabled us to compute analytical expressions for the metrics of interest under neutrality, and approximations thereof.

Our analytical expressions suggest that allelic turnover alone may be responsible for large reductions in accuracy: For small mutation rates, $\rho^2(\tau)$ (and $r^2(\tau)$) decreases substantially within short time-spans, by about 20 percent in $0.2N$ generations (corresponding to approximately 120,000 years in humans). In addition, increasing the detection threshold yielded lower polygenic score accuracy, as a locus was less likely to have a non-zero effect. These results are broadly consistent with a concurrent study by Yair and Coop (Yair and Coop, 2021), in which the authors used simulations to assess cross-population prediction accuracy, defined as the ratio of the variance of and individual's polygenic score to that of their genetic value, under neutrality and in the presence of stabilizing selection. When Yair and Coop restricted the polygenic score to the top one percent of SNPs, roughly analogous to

altering the detection threshold, they similarly found that the accuracy declined in the focal population.

Yet, while the detection threshold influenced the magnitude of the polygenic score accuracy, relative accuracy was insensitive to this parameter. In other words, under neutrality, relative accuracy is insensitive to the magnitude of the per-locus effect and only depends on the underlying allele frequency distribution. In addition, relative accuracy was independent of the size of the mutational target when the constituent loci were *iid*. Our theory suggests that these results will hold for arbitrary distributions of the true effect β. Consideration of several effect size distributions in a parameter regime consistent with the UK Biobank further supports this conjecture (Section 2.8.10). Although more work is required to fully substantiate this claim.

Selection, however, induces a dependency between an allele's effect and its frequency, and may thereby render relative accuracy sensitive to the detection threshold. Our simulations provide preliminary evidence in support of this claim. For a small mutation rate of $a = 4N\mu = 10^{-3}$ and a large per-locus selection coefficient $\sigma = 4Ns = 10$, relative accuracy was lower for the larger detection threshold of $d = 10^4$ compared to $d = 10^3$. Yet, the difference between detection thresholds was small relative to that induced by selection, and was negligible for smaller selection coefficients. Indeed, smaller selection coefficients ($\sigma \leq 1$) did not yield appreciable deviations from our neutral expectations for the *mse*, accuracy, nor relative accuracy. Therefore, excluding strong selection ($\sigma > 1$), our neutral expectations for these statistics appear to be good approximations to their true values. Our theoretical results under neutrality thus may prove an accurate description of temporally-resolved polygenic scores when polygenic adaptation is achieved by concurrent small frequency changes at numerous small effect loci—a plausible scenario (Pritchard et al., 2010; Berg and Coop, 2014). In addition, the simple patterns revealed by our simulations suggest that it may be possible to derive (approximate) analytic expressions for the given metrics in the presence

of strong selection, when loci exhibit selective sweep-like behavior.

It is unclear whether our neutral expectations will hold in the context of more sophisticated polygenic trait modeling. In our simulation study, as in our theoretical work, we focus on dynamics at a single locus. Thus, our results are most relevant to scenarios in which single locus dynamics can be decoupled from the evolution of the mean phenotype and the genetic background (Chevin and Hospital, 2008). Namely, the effect of an individual locus must be small relative to the mean phenotype (Chevin and Hospital, 2008; Simons et al., 2018). Future work will assess polygenic score accuracy under more sophisticated models of polygenic adaptation (e.g., Simons et al. (2018); Hayward and Sella (2022)).

Of the two bias-inducing processes explored, detection threshold asymmetry and directional selection, the latter induced much larger deviations from our neutral expectation for the bias, i.e., under neutrality $bias(\tau) = 0$ for all ancient sampling times τ. In the presence of detection asymmetry, $bias(\tau)$ is approximately proportional to the difference between the one-sided detection probabilities, which in turn is constrained by the shape of the allele frequency distribution. Under neutrality, and for small mutation rates, most alleles are at very low frequencies or fixed, such that changing the detection threshold minimally influences the one-sided detection probabilities. Selection, however, perturbs the underlying allele frequency density. At equilibrium, this density is proportional to $e^{\sigma z} z^{-1}(1-z)^{-1}$ for small a, where $\sigma = 4Ns$. Depending on σ, the one-sided detection probabilities may differ markedly, yielding larger values of $bias(\tau)$. We thus suspect that detection asymmetry has the potential to further exacerbate any bias induced by selection. These results are interesting in light of those of Chan et al. 2014 (Chan et al., 2014), who demonstrated that polygenic disease inheritance under the liability threshold model induced differences in the power to detect protective versus susceptible alleles. In Chan et al., this effect was further increased by imbalances in the case and control sample sizes in the GWA study. Additional work is needed to incorporate these features of case-control studies into our modeling framework.

The effects of selection on the bias have implications for assessments of mean differences between ancient polygenic scores from distinct time points. In particular, our results suggest that sufficiently strong positive directional selection will lead to overestimation of the difference between the polygenic scores of ancient individuals sampled before and after the onset of selection. Likewise, in the presence of negative selection, the polygenic score will underestimate this difference. At the same time, as discussed above, estimation error increases (as measured by $mse(\tau)$) and accuracy (as measured by $\rho^2(\tau)$) decreases as the ancient sampling time increases.

Our results clarify relationships between various commonly used metrics of prediction error and accuracy. For example, we demonstrated an approximate functional relationship between the mean-squared error $mse(\tau)$ and the expected additive genetic variance $\hat{V}_A(\tau)$ that applies for small ancient sampling times and mutation rates. This shared initial rate emerged despite fundamental differences between these statistics: $mse(\tau)$ measures error due to variants near or at fixation in the contemporary sample, which were segregating at intermediate frequencies in the ancient sample. In contrast, $\hat{V}_A(\tau)$ measures error due to variants segregating in the contemporary sample, which were near or at fixation in the ancient sample. This conceptual result does not rely on any of our population genetic or GWA modeling assumptions, and perhaps could be exploited to learn about the genetic architecture of quantitative traits from multi-population data. In addition, we showed formally that polygenic score accuracy $\rho^2(\tau)$, an approximation to the expectation of the sample correlation coefficient $r^2(\tau)$, is proportional to the ratio of $\hat{V}_A(\tau)$ to the total phenotypic variance. We believe that these relations, and their evolutionary and GWA study dependent forms, may facilitate the development of novel, more principled statistical procedures for the analysis of out-of-sample polygenic scores.

At the same time, the simplifying assumptions underlying our results indicate that significant challenges remain. For one, our model does not incorporate the complex demographic

processes, such as admixture and population size changes, inherent in human history. This implies that an ancient sampling time of t years in the past likely does not correspond to a sampling time of $\tau = t/2N$ in our model, where $2N$ is the contemporary population size. Indeed, allelic turnover cannot explain all of the reductions in accuracy observed in out-of-sample predictions in humans. For example, our neutral theory predicts an approximately fifty percent reduction in accuracy when F_{ST} between the focal and GWA study populations is comparable to African-European divergence ($F_{ST} \approx 0.1$). This more severely overestimates the prediction accuracy of height in a sample of individuals with African ancestry compared to the Wang et al. predictions, which take into account both LD and allele frequency changes (Section 2.8.14). Thus, to achieve the same accuracy reductions observed in both simulated, e.g., (Ragsdale et al., 2020; Wang et al., 2020) and empirical, e.g., (Martin et al., 2017; Duncan et al., 2019; Wang et al., 2020), studies of cross-population polygenic scores for contemporary humans, allelic turnover under neutrality would require population divergence times that far exceed their estimated values (Fig. 2.11).

Differences in LD between contemporary human populations may largely explain this discrepancy as most trait-associated loci are likely to be tagging rather than causal sites (Wang et al., 2020; Bitarello and Mathieson, 2020). As with geographically distinct populations, if LD between the genotyped and causal sites differed in the ancient population, then polygenic score accuracy would suffer (Habier et al., 2007). We did not model this effect and assumed that the genotyped site was the causal site. This assumption may be justified when ancient sampling or population divergence times are recent, as high marker density in the GWA study may mitigate accuracy losses due to LD decay, but more theoretical work is required to substantiate this claim. While our framework can readily incorporate LD, it is difficult to obtain analytical results when the genotyped marker is *not* the causal site. In lieu of theoretical results, large-scale simulations in simple population genetic scenarios may provide insight into the relative contributions of LD—which depends on the allele frequencies

of the tagging and causal sites—and allelic turnover to declines in polygenic score accuracy.

Furthermore, our assumption of linkage equilibrium between loci roughly equates to assuming that each LD block contains only a single causal site. Thus, our results will be most applicable to traits with relatively sparse genetic architectures for which the distance between any two causal loci is large compared to the scale of LD. In contrast, when the trait architecture is dense, a large number of variants have non-zero effect on the trait. Causal sites in close proximity are necessarily linked, and our assumption of linkage equilibrium would be violated. In addition, under a dense trait architecture, the "prune and threshold" polygenic score described herein may achieve lower accuracy than a best linear unbiased predictor (BLUP) that allows all segregating loci to have non-zero effects. In Section 2.8.4, we speculate on the accuracy of BLUP in the context of our modeling framework when the trait has a dense architecture.

In addition, we assumed that per-locus causal effects were shared by the ancient and contemporary samples. Differences in causal effects across contemporary populations, perhaps due to changes in the environment, epistasis, or gene-by-environment interactions, likely contribute to accuracy reductions (Galinsky et al., 2019; Bitarello and Mathieson, 2020). Indeed, Cox et al. (Cox et al., 2019) found that trends in the polygenic scores of temporally disparate ancient samples did not always recapitulate those of the true phenotype. We conjecture that fluctuations in the per-locus effects would increase $mse(\tau)$ and decrease accuracy, but not profoundly alter our conclusions. Perhaps, if the fluctuations were asymmetric, e.g., effect sizes tended to increase in time, then $bias(\tau)$ may be non-zero under neutrality. Population stratification in the GWA study population may also lead to biased ancient polygenic scores, as has been observed in cross-population predictions in humans (Berg et al., 2019; Sohail et al., 2019). Lastly, technical challenges inherent to the extraction and sequencing of ancient DNA often result in noisy estimates of the ancient genotypes. This additional source of randomness is likely to reduce accuracy and increase $mse(\tau)$, but otherwise should not

substantially alter our conclusions.

2.7 Author contributions

We provide the author contributions in the same format as they appear in the published manuscript (Carlson et al., 2022).

Maryn O. Carlson

Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing - original draft, Writing - review & editing

Daniel P. Rice

Conceptualization, Formal analysis, Methodology, Writing - review & editing

Jeremy J. Berg

Methodology, Writing - review & editing

Matthias Steinrücken

Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Supervision, Writing - review & editing

2.8 Extended model, methods, and results

2.8.1 Joint allele frequency density in the population split scenario

In the main text, we claim that under our assumptions—namely, neutrality, constant population size, and stationarity—the modeling framework readily encompasses a simple population split scenario in which two populations diverged some τ_{split} generations ago and the ancient individual was sampled at τ (Fig. 2.1B). Specifically, the split scenario is analogous to the single population scenario (Fig. 2.1A) in which the ancient individual is sampled at $2\tau_{\text{split}} - \tau$. We derive the form of the joint allele frequency density for the demographic split scenario

(Fig. 2.1B) as proof.

For the scenario in Fig. 2.1A, the joint allele frequency density at τ and the present is $f(z_\tau, z_0) = p(z_\tau, z_0; \tau)\kappa(z_\tau)$, where $\kappa(\cdot)$ is the stationary density,

$$\kappa(z) \coloneqq \frac{1}{B(a,a)} z^{a-1}(1-z)^{b-1} \coloneqq \frac{\pi(z)}{B(a,a)}. \tag{2.21}$$

For the split scenario, we must condition on the allele frequency at the time of the split. The joint density of a single site is then,

$$\begin{aligned} f(z_\tau, z_0) &= \int_{z_s=0}^{1} f(z_\tau, z_0 | Z(\tau_s) = z_s)\kappa(z_s)dz_s \\ &= \int_{z_s=0}^{1} p(z_s, z_\tau; \tau - \tau_{\text{split}})p(z_s, z_0; \tau_{\text{split}})\kappa(z_s)dz_s, \end{aligned} \tag{2.22}$$

where z_s is the integration variable for the allele frequency at the time of the split. We can further simplify Eq. (2.22) by twice substituting the spectral representation of the transition density (Section 2.8.6),

$$\begin{aligned} &f(z_\tau, z_0) \\ &= \int_{z_s=0}^{1} \left(\sum_{k=0}^{\infty} \frac{e^{-\lambda_k(\tau_{\text{split}} - \tau)}}{\langle B_k, B_k \rangle_\pi} B_k(z_s)B_k(z_\tau)\pi(z_\tau) \right) \\ &\qquad \left(\sum_{m=0}^{\infty} \frac{e^{-\lambda_m \tau_{\text{split}}}}{\langle B_m, B_m \rangle_\pi} B_m(z_s)B_m(z_0)\pi(z_0) \right) \frac{\pi(z_s)}{B(a,b)} dz_s \\ &= \sum_{k=0}^{\infty} \frac{e^{-\lambda_k(2\tau_{\text{split}} - \tau)}}{\langle B_k, B_k \rangle_\pi} B_k(z_\tau)B_m(z_0)\frac{\pi(z_\tau)}{B(a,b)}\pi(z_0), \end{aligned} \tag{2.23}$$

where we exchanged integration and summation. We recognize that this equation is equal to,

$$f(z_\tau, z_0) = p(z_\tau, z_0; 2\tau_{\text{split}} - \tau)\frac{\pi(z_\tau)}{B(a,b)} = p(z_\tau, z_0; 2\tau_{\text{split}} - \tau)\kappa(z_s). \tag{2.24}$$

Thus, the joint density in the instance of a population split is of the same form as in the

single population scenario, but with a modified time argument: $2\tau_{\text{split}} - \tau$ instead of τ.

2.8.2 Power to detect a significant association in a GWA study

We follow (Simons et al., 2018) (who follow (Sham and Purcell, 2014)) in modeling the power of a GWA study to detect a significant association. We assume that conditional on the true effect size β_ℓ, and the population allele frequency Z_ℓ (implicitly assuming $\hat{Z}_\ell \approx Z_\ell$) the estimated marginal effect $\hat{\beta}_\ell$ is normally distributed,

$$\hat{\beta}_\ell | \beta_\ell, Z_\ell \sim \mathcal{N}\left(\beta_\ell, \frac{V_p}{2nZ_\ell(1-Z_\ell)}\right), \tag{2.25}$$

where V_p is the total phenotypic variance which includes both genetic and environmental effects. Under the null hypothesis, i.e., $\beta_\ell = 0$, $\hat{\beta}_\ell$ is normally distributed with mean zero and the same variance as Eq. (2.25). The estimated contribution of locus ℓ to the phenotypic variance, $\hat{v}_\ell$, is $\hat{v}_\ell \coloneqq 2\hat{\beta}_\ell^2 Z_\ell(1-Z_\ell)$. When normalized by $\frac{V_p}{n}$, $\hat{v}_\ell$ is chi-squared distributed with one degree of freedom,

$$\frac{\hat{v}_\ell}{V_p/n} = \frac{2\hat{\beta}_\ell^2 Z_\ell(1-Z_\ell)}{V_p/n} \sim \chi_1^2. \tag{2.26}$$

It follows that there is some threshold contribution to variance, v_*, such that the test statistic given in Eq. (2.26) is statistically significant. Specifically, fixing the significance threshold α,

$$v_* = F^{-1}(1-\alpha) = 2\left(\text{erf}^{-1}(1-\alpha)\right)^2, \tag{2.27}$$

where $F^{-1}(\cdot)$ is the inverse cumulative distribution function (*cdf*) of a chi-squared distributed random variable with one degree of freedom, and erf is the error function (eq. *A82* in (Simons

et al., 2018)). This implies that a locus which satisfies,

$$\frac{2\hat{\beta}_\ell^2 Z_\ell(1-Z_\ell)}{V_p/n} > v_*, \tag{2.28}$$

will yield a statistically significant association. Eq. (2.28) further implies that if, for a fixed Z_ℓ,

$$|\hat{\beta}_\ell| > \sqrt{\frac{v_*(V_p/n)}{2Z_\ell(1-Z_\ell)}}, \qquad \text{or, for a fixed } \hat{\beta}, \qquad Z_\ell(1-Z_\ell) > \frac{v_*(V_p/n)}{2\hat{\beta}_\ell^2}, \tag{2.29}$$

site ℓ will yield a significant association. If we substitute the true effect β_ℓ for $\hat{\beta}_\ell$ in Eq. (2.29), we can define these thresholds with respect to the true effect.

And, for a fixed β our condition is,

$$\frac{1}{2} - \frac{1}{2}\sqrt{1 - \frac{2v_*(V_p/n)}{\beta_\ell^2}} < Z_\ell < \frac{1}{2} + \frac{1}{2}\sqrt{1 - \frac{2v_*(V_p/n)}{\beta_\ell^2}}, \tag{2.30}$$

We define,

$$\gamma_\ell = 1 - \sqrt{1 - \frac{2v_*(V_p/n)}{\beta_\ell^2}}, \qquad \text{and,} \qquad \beta_* = \sqrt{\frac{v_*(V_p/n)}{2Z_\ell(1-Z_\ell)}}. \tag{2.31}$$

Eq. (2.31) specifies the threshold model used in our model, given in Eq. (2.3), $D_\ell \in [d_\ell, 2n - d_\ell]$, where D_ℓ is the allele count in the GWA study and $d_\ell = \lceil n\gamma_\ell \rceil$. The distributional assumptions in Eq. (2.25) imply that the threshold model will be a good approximation when n is large relative to V_p and the detection threshold is not too small. For example, if the allele frequency in the GWA study was at its minimum frequency of $\frac{1}{2n}$, then the variance of $\hat{\beta}$ would be proportional to V_p—which may be large.

2.8.3 Simulation procedures

In this section, we describe how we simulated the ancient polygenic scores (i) under neutrality and (ii) in the presence of genic selection.

Neutrality. To assess the accuracy of our theoretical results for the various statistics presented in Section 2.4, we simulated realizations of the polygenic score for ancient individuals according to our model (Section 2.3).

Initialization. To initialize each realization, we sampled L population allele frequencies, $\boldsymbol{Z}(\tau) \in [0,1]^L$, from a Beta-distribution with parameters $a = b = 4N\mu$ and $b = 4N\nu$. As the population size N is finite, the beta-distribution is a continuous approximation to the discrete probability mass function governing the allele frequencies. Thus, we conduct one round of binomial sampling to obtain frequencies in the set $\{0, \frac{1}{2N}, \ldots, 1 - \frac{1}{2N}, 1\}$.

Allele frequency evolution. Allele frequencies then evolve forward-in-time until the present ($t = 0$) when the GWA study is conducted. For forward and backward mutation rates μ and ν, the transition probability of the discrete Wright-Fisher process is,

$$\psi_\mu(z) = (1-z)\mu + z(1-\nu) = \mu(1-2z) + z, \tag{2.32}$$

for an allele frequency $z \in [0,1]$, and where the second equality follows for $\mu = \nu$. Conditional on the allele frequency at t (generations in the past), the allele frequency in the subsequent generation is given by $(t-1)$,

$$Z_\ell(t-1)|Z_\ell(t) \sim Bin\left(2N, \psi_\mu\left(Z_\ell(t)\right)\right), \tag{2.33}$$

until $t - 1 = 0$.

Genome-wide association study. To conduct the GWA study, we sample n diploid genotypes, $\boldsymbol{X}_i(0) \in \{-1,0,1\}^L$, for $i \in \{1,\ldots,n\}$ conditional on the allele frequencies, $\boldsymbol{Z}(0)$. Condi-

tional on $\boldsymbol{Z}(0)$, each genotype is *iid*, $\boldsymbol{X}_i(0)|\boldsymbol{Z}(0) \sim \prod_{\ell=1}^{L} Bin(2, Z_\ell(0))$. We then sample their phenotypes, $\boldsymbol{Y}(0) \in \mathbb{R}^n$ conditional on the their genotypes, according to Eq. (2.1). This set of n genotypes and phenotype comprise the GWA study sample.

To estimate the effects we first compute the allele count at each site D_ℓ in the study sample. If D_ℓ is within the specified interval $[d_\ell, 2n - d_\ell]$, we set the effect estimate to β_ℓ, as in Eq. (2.3). If D_ℓ falls outside of the interval, then the effect estimate is set to 0. We then estimate $\hat{C}$ using Eq. (2.5).

Sampling the ancient individual(s). We sample the genotype $\boldsymbol{X}(\tau)$ and phenotype $Y(\tau)$ of a single ancient individual conditional on the population allele frequencies $\boldsymbol{Z}(\tau)$.

Computing estimates of the statistics. We compute method of moments estimators for each of the statistics defined in Section 2.3.5. For each ancient sampling time, $\tau = \{\tau_1, \tau_2, \ldots, \tau_T\}$, we conduct K simulations. For $bias(\tau)$, $mse(\tau)$, and $\hat{V}_A(\tau)$ we are interested in per-locus statistics, we average over all $L \times K$ independent locus trajectories for each time point. For example, the estimator of the *bias* is given by,

$$\overline{bias_\ell}(\tau) \coloneqq \frac{1}{KL} \sum_{k=1}^{K} \sum_{\ell=1}^{L} \left(\bar{X}_{k\ell} - X_{k\ell}(\tau)\right) \left(\beta - \hat{\beta}_{k\ell}\right),$$

where k indexes the simulation and ℓ the locus. The estimator's $(1-\alpha)\%$ confidence interval is given by,

$$bias_\ell(\tau) \in [\overline{bias}_\ell(\tau) \pm z_{\alpha/2} s_{bias_\ell}]$$

where s_{bias_ℓ} is the estimated standard deviations of $bias_\ell(\tau)$ and $z_{\alpha/2}$ is the inverse cumulative distribution function of a standard normally distributed random variable evaluated at $\alpha/2$.

Our estimators for the sample correlation coefficient $r^2(\tau)$ and its approximation $\rho^2(\tau)$

are computed for each replicate of L loci. For example, for the k-th replicate,

$$r_k^2 \coloneqq \frac{Cov[\hat{\boldsymbol{Y}}_k(\tau), \boldsymbol{Y}_k(\tau)]}{Var[\hat{\boldsymbol{Y}}_k(\tau)]Var[\boldsymbol{Y}_k(\tau)]} = \frac{Cov[\sum_{\ell=1}^{L} \boldsymbol{X}_{k\ell}(\tau)\hat{\beta}_{k\ell}, \sum_{\ell=1}^{L} \boldsymbol{X}_{k\ell}(\tau)\beta_{k\ell} + \boldsymbol{\epsilon}]}{Var[\sum_{\ell=1}^{L} \boldsymbol{X}_{k\ell}(\tau)\hat{\beta}_{k\ell}]Var[\sum_{\ell=1}^{L} \boldsymbol{X}_{k\ell}(\tau)\beta_{k\ell} + \boldsymbol{\epsilon}]}, \quad (2.34)$$

where $\hat{\boldsymbol{Y}}_k(\tau)$ and $\boldsymbol{Y}_k(\tau)$ are the n_a-length vectors of ancient polygenic scores and phenotypes, respectively; $\boldsymbol{X}_{k\ell}(\tau)$ is the vector of ancient genotypes at the ℓ-th site; and $\boldsymbol{\epsilon}$ is the vector of environmental contributions to each individual's phenotype. Our estimator $\hat{r}^2$ is an average of the K realizations of r_k^2. To estimate $\rho^2(\tau)$, we first find estimators for the covariance and variance terms in Eq. (2.34) by averaging over the K simulations. We then compute the ratio of these quantities to compute $\hat{\rho}^2(\tau)$.

Genic selection. To investigate how positive selection influences the statistical properties of polygenic scores, we simulated a recent directional selection scenario. The population evolves neutrally until the onset of selection τ_s years in the past. The A_2 allele confers a fitness advantage of s, such that the relative fitnesses of the genotypes A_1A_1:A_1A_2:A_2A_2 are given by $1{:}1+s{:}1+2s$.

Initialization. To initialize each realization, we sample L population allele frequencies, $\boldsymbol{Z}(\tau) \in [0,1]^L$, from the stationary distribution of the neutral Wright-Fisher diffusion with recurrent mutation. If the ancient sampling time τ is greater than τ_s then we simulate 50 generations of neutral evolution before the onset of selection.

Allele frequency evolution. Allele frequencies evolve neutrally until the onset of selection at τ_s. At this juncture, the allele frequencies begin to evolve according to,

$$\psi_{\mu s}(z) = \frac{\left[(1-z)^2 + z(1-z)(1+s)\right]\mu + \left[z(1-z)(1+s) + z^2(1+2s)\right](1-\mu)}{\bar{w}(z)}, \quad (2.35)$$

where μ is the forward and backward per-locus, per-generation mutation rate, and the denominator is the mean fitness in a population with A_2 allele frequency z, up to the present

day. The simulations with selection are otherwise identical to those under neutrality.

2.8.4 Alternative prediction models

In the main text, we introduced a simple threshold model for the effect estimates in Eq. (2.3). Here, in Section 2.8.4, we consider a more realistic model in which the effect estimate $\hat{\beta}_\ell$ is the maximum-likelihood estimate (MLE) of β_ℓ. We give expressions for the first two moments of $\hat{\beta}_\ell$ conditional on the contemporary allele frequencies Z_ℓ under each model. In doing so, we illustrate the additional challenges posed by the MLE model, and why we ultimately opted to pursue the simpler threshold model presented in the main text in Eq. (2.3). In addition, in Section 2.8.4, we relate the threshold model to the Best Linear Unbiased Predictor, or BLUP (e.g. Meuwissen et al. (2001); de los Campos et al. (2013b)). In doing so, we provide a brief exposition of BLUP following (de los Campos et al., 2013b).

Moments of the maximum-likelihood and threshold models

Maximum-likelihood threshold model. We define the MLE threshold model,

$$\hat{\beta}_\ell \coloneqq \begin{cases} \frac{Cov[\boldsymbol{X}_\ell, \boldsymbol{Y}]}{Var[\boldsymbol{X}]} = \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})(X_{i\ell} - \bar{X}_\ell)}{\sum_{i=1}^{n}(X_{i\ell} - \bar{X}_\ell)^2} & \text{if } D_\ell \in [d_\ell, 2n - d_\ell] \\ 0 & \text{else,} \end{cases} \tag{2.36}$$

where each genotype, phenotype pair $(\boldsymbol{X}_i, Y_i)$ for $i \in \{1, \ldots, n\}$ is associated with an individual in the GWA study; $Cov[\cdot, \cdot]$ and $Var[\cdot]$ are the sample covariance and variance, respectively; and, as before, $\bar{X}_\ell$ and $\bar{Y}$ are the average genotype at the ℓ-th locus and phenotype in the GWA sample.

First moment of $\hat{\beta}$. For both models, it can be shown that,

$$\mathbb{E}\left[\hat{\beta}_\ell | Z_\ell, \beta_\ell\right] = \beta_\ell \wp(Z_\ell, 2n, d_\ell), \tag{2.37}$$

where Z_ℓ is the contemporary population allele frequency; and

$$\wp(z_\ell, 2n, d_\ell) = \sum_{i=d_\ell}^{2n-d_\ell} \binom{2n}{i} z_\ell^i (1-z_\ell)^{2n-i} \tag{2.38}$$

is the probability that the allele count in the GWA study falls at or above the threshold d_ℓ. Thus, when the site is segregating at a sufficiently high frequency in the GWA study sample, the estimator is unbiased. Unconditionally, for an allele count threshold d_ℓ, $\mathbb{E}[\hat{\beta}_\ell|\beta_\ell] = \beta_\ell(1-2P^{(d_\ell)})$, where $P^{(d_\ell)}$ is the *cdf* of a beta-binomial random variable and is defined in Eq. (2.62).

Second moment of $\hat{\beta}$. The two models yield different second moments.

Simple threshold model. It can be shown that under the simpler model,

$$\mathbb{E}\left[\hat{\beta}_\ell^2|\boldsymbol{Z}, \beta_\ell\right] = \beta_\ell^2 \wp(Z_\ell, 2n, d_\ell). \tag{2.39}$$

As Eq. (2.39) only involves the allele frequency of site ℓ, it is not influenced by variation at other sites in $\mathcal{L}$. And, unconditionally, $\mathbb{E}[\hat{\beta}_\ell^2|\beta_\ell] = \beta_\ell^2\left(1-2P^{(d_\ell)}\right)$.

MLE model. It can be shown that under the MLE model,

$$\mathbb{E}\left[\hat{\beta}_\ell^2|\boldsymbol{Z}(0)\right] \approx \left(\beta_\ell^2 + \sum_{\ell' \neq \ell} \beta_{\ell'}^2 \frac{Z_{\ell'}(1-Z_{\ell'})}{nZ_\ell(1-Z_\ell)} + \frac{\sigma_e^2}{2nZ_\ell(1-Z_\ell)}\right) \wp(Z_\ell, 2n, d_\ell), \tag{2.40}$$

where the sum is over all loci $\ell' \in \mathcal{L}$ such that $\ell' \neq \ell$. The approximation is due to approximating the expectation of a ratio with the ratio of expectations, and as such, comes with all of the corresponding dangers of such an approximation. In addition, we can see from Eq. (2.40) that the second moment of $\hat{\beta}_\ell$ depends on the allele frequencies at all other loci in $\mathcal{L}$. While were able to compute the metrics under this approximate MLE model, we concluded that its reliance on strict assumptions about the genetic architecture (via the

second moment) obscured the effects of allelic turnover. In addition, a threshold model arises naturally as the large n limit of the MLE model which is, up to a sample size factor, equivalent to Eq. (2.25) when the allele frequency Z_ℓ is not too small.

A comparison to the Best Linear Unbiased Predictor

A brief overview of BLUP. We follow the Supporting Information of de los Campos et al. (de los Campos et al., 2013b) in introducing the Best Linear Unbiased Predictor (BLUP). However, we use notation consistent with the notation of our investigation.

The model underlying BLUP is of the form,

$$\tilde{\boldsymbol{Y}} = \tilde{\boldsymbol{X}}\boldsymbol{\beta} + \epsilon, \tag{2.41}$$

where for a GWA study consisting of n individuals, $\tilde{\boldsymbol{Y}} \in \mathbb{R}^n$ is a vector of centered and scaled phenotypes, $\boldsymbol{\beta} \in \mathbb{R}^{L'}$ is a vector of marker effects, and $\tilde{\boldsymbol{X}} \in \mathbb{R}^{n \times L'}$ is a centered and scaled matrix of the genotypes at all L' *segregating* sites in the genome (potentially above some minimum allele frequency), with the $i\ell$-th element $\tilde{X}_{i\ell} = \frac{X_{i\ell} - \bar{X}_\ell}{\sqrt{2\hat{Z}(1-\hat{Z})}}$, where $X_{i\ell} \in \{-1, 0, 1\}$ is the individual's genotype, and $\bar{X}_\ell$ and $\hat{Z}_\ell$ are the average genotype and estimated allele frequency in the GWA study sample. (Note that centering and scaling of the genotypes and phenotypes is not necessary to the formulation of BLUP (de los Campos et al., 2013b).) The marker effects are assumed *iid*, $\beta_\ell \sim \mathcal{N}(0, \sigma_\beta^2)$, where σ_β^2 is the prior variance of the marker effects. Similarly, the residuals are assumed *iid* with $\epsilon_i \sim \mathcal{N}(0, \sigma_e^2)$, where σ_e^2 is the prior residual variance. Under this model, the effect estimates and phenotypes follow a multivariate normal distribution,

$$\begin{bmatrix} \boldsymbol{\beta}|\tilde{\boldsymbol{X}} \\ \tilde{\boldsymbol{Y}}|\tilde{\boldsymbol{X}} \end{bmatrix} \sim MVN\left(0, \begin{bmatrix} \boldsymbol{I}_{L'}\sigma_\beta^2 & \tilde{\boldsymbol{X}}^T\sigma_\beta^2 \\ \tilde{\boldsymbol{X}}\sigma_\beta^2 & \tilde{\boldsymbol{X}}\tilde{\boldsymbol{X}}^{\boldsymbol{T}}\sigma_\beta^2 + \boldsymbol{I}_n\sigma_e^2 \end{bmatrix}\right), \tag{2.42}$$

such that,

$$\mathbb{E}[\hat{\boldsymbol{\beta}}|\tilde{\boldsymbol{Y}}, \tilde{\boldsymbol{X}}] = (1/L')\tilde{\boldsymbol{X}}^T[\boldsymbol{G} + \boldsymbol{I}_n(\sigma_\epsilon^2/(L'\sigma_\beta^2)]\tilde{\boldsymbol{Y}} \tag{2.43}$$

with $\boldsymbol{G} = (1/L')\tilde{\boldsymbol{X}}\tilde{\boldsymbol{X}}^T$. And thus,

$$\mathbb{E}[\hat{\beta}_\ell|\tilde{\boldsymbol{Y}}, \tilde{\boldsymbol{X}}] = (1/L')\tilde{\boldsymbol{X}}_{\boldsymbol{\ell}}^T[\boldsymbol{G} + \boldsymbol{I}_n(\sigma_\epsilon^2/(L'\sigma_\beta^2)]\tilde{\boldsymbol{Y}}. \tag{2.44}$$

Relating the MLE estimate. When we compare Eq. (2.44) to the MLE estimate (for centered and scaled genotypes and phenotypes),

$$\hat{\beta}_\ell = \tilde{X}_\ell^T\tilde{\boldsymbol{Y}}, \tag{2.45}$$

we see that they are analogous up to the bracketed expression in Eq. (2.44). This term, $\boldsymbol{G}+\boldsymbol{I}_n(\sigma_\epsilon^2/L'\sigma_\beta^2)$, models the relationship between individuals in the GWA study and affects the effect estimates of all loci. If the relationship matrix $\boldsymbol{G}$ were the identity, than the bracketed term would simply scale the effect estimates by a factor $\sigma_\epsilon^2/\sigma_\beta^2$. However, in the likely instance that $\boldsymbol{G}$ deviates from the identity, the primary contribution to the per-locus effect estimate is still given by the covariance between the genotype of locus ℓ and the phenotype, i.e., by Eq. (2.45). Thus, the MLE of the effect size is closely related to the effect estimate derived from BLUP, at least when the variance components are known.

A dense trait architecture. When the trait architecture is dense, a large number of segregating sites each impart a small effect on the trait (see (Barton et al., 2017) for the limiting behavior, referred to as the infinitesimal model). In this setting, BLUP—which allows all loci to have non-zero effects—will likely yield predictions with higher accuracy than the "prune and threshold" model—which assumes only one "causal" SNP within a given genomic window. As noted in the *Discussion*, our assumption of linkage equilibrium between loci necessarily breaks down under a dense architecture, and the allele frequency trajectories

cannot be modeled independently. (As before, we assume that all loci are evolving neutrally in a constant size population.) In addition, the marginal effect estimate for a given locus will absorb the effects of its neighbors in proportion to the LD between these loci (e.g. see Equation A87 of Simons et al. (2018)). Genome-wide association study and prediction methods that model LD between loci, e.g. (Vilhjálmsson et al., 2015), aim to reduce to the effects of the latter.

Let d be the allele frequency threshold at which point a locus is considered "segregating" in the GWA study, e.g. $d = 1$. (Though, standard quality control pipelines may impose a threshold $d > 1$ to remove very low frequency sites that are more susceptible to false positives.) The above two LD-induced complications aside, we can loosely approximate the BLUP model by using d as the allele frequency threshold *for all loci* irrespective of their effect sizes. Accuracy under the BLUP can then be roughly approximated as,

$$\begin{aligned}\rho^2(\tau) &\approx \frac{\frac{1}{2a+1}\sum_{\ell'=1}^{L'}\beta_{\ell}^2[a(1-2P^{(d)})+2e^{-(2a+1)\tau}P_3^{(d)}]}{\frac{a}{2a+1}\sum_{\ell'=1}^{L'}\beta_{\ell}^2+\sigma_e^2} \\ &= \frac{\frac{1}{2a+1}[a(1-2P^{(d)})+2e^{-(2a+1)\tau}P_3^{(d)}]}{\frac{a}{2a+1}+\sigma_{e''}^2},\end{aligned} \tag{2.46}$$

where $\sigma_{e''}^2 = \sigma_e^2/(\sum_{\ell'}\beta_{\ell'}^2)$, and the sum is over all segregating sites $\ell' \in \{1,\ldots,L'\}$. As the threshold d does not depend on the per-locus effect, we can factor the sum of the squared effect sizes to arrive at the second line of Eq. (2.46). Relative accuracy readily follows from Eq. (2.46),

$$\rho^2(\tau)/\rho^2(0) \approx \frac{a(1-2P^{(d)})+2e^{-(2a+1)\tau}P_3^{(d)}}{a(1-2P^{(d)})+2P_3^{(d)}}. \tag{2.47}$$

We can compare these expressions with those derived in Section 2.8.9. (1) As before, relative accuracy does not depend on the effect sizes, and will cohere with our previous results for $d_{\ell'} = d$ for all ℓ'. When the mutation rate is small, most sites are fixed for either the A_1 or A_2 allele. Thus, many sites will evade detection even when $d \leq 10$, and both accuracy

and relative accuracy will decay substantially over time. In addition, relative accuracy appears insensitive to the threshold d (see insets in Fig. 2.3), thus relative accuracy, given in Eq. (2.47), will likely behave similarly under BLUP. And (2), while Eq. (2.46) removes the relationship between an effect size $\beta_{\ell'}$ and its threshold $d_{\ell'}$, we still expect it to behave qualitatively similar to our previous results. Though, as $d \leq d_{\ell'}$ for all ℓ' under our previous parameterization (see Section 2.8.2), BLUP will achieve higher accuracy than the threshold model.

Simply setting $d_{\ell'} = d = 1$ for all $\ell' \in \{1, \ldots, L'\}$ to model BLUP does not, however, capture an important way in which BLUP deviates from the "prune and threshold" model. While not captured explicitly in our threshold model, the standard error of the effect estimate $\hat{\beta}_{\ell'}$ for locus ℓ' depends on both the magnitude of its true effect and the allele frequency. In particular, all else being equal, a variant at low frequency will have a larger standard error (and thus larger p-value) compared to a variant at moderate frequency with the same effect. We justify ignoring noise in the effect estimates in our modeling of the "prune and threshold" approach as variants exceeding the p-value threshold must either be at intermediate frequency and/or of large effect. Thus, this effect should be mitigated by the fact that standard errors of loci with non-zero effects in the polygenic score are necessarily small relative to the estimated effect size.

In contrast, BLUP allows *all* variants to have non-zero effects, implying that low frequency variants will have systematically larger standard errors relative to moderate frequency variants. As low frequency variants shift towards higher frequencies in the ancient population, these variants will disproportionately—relative to their frequencies in the ancient population—contribute to the noisiness of BLUP. Similarly, as moderate frequency variants shift towards lower frequencies in the ancient population, they will contribute disproportionately less to the prediction noise. Future work may precisely quantify the confluence of allelic turnover and effect estimate uncertainty induced by the BLUP model (and perhaps

under the "prune and threshold" model as well). In addition, a more rigorous analysis would necessarily take into account LD between loci.

2.8.5 Polygenic scores from centered and scaled GWAS data

In the main text, we chose not to center and scale the genotypes and phenotypes of sampled individuals when conducting the GWA study. In this section, we show that our conclusions are robust to this choice. Our calculations also demonstrate that procedures convenient for statistical analysis, namely scaling, prove inconvenient when evolutionary processes are taken into account.

Centering and scaling. We center and scale to unit variance the phenotypes and genotypes in the GWA study,

$$\tilde{Y}_i = \frac{Y_i - \bar{Y}}{s_Y} \qquad \text{and} \qquad \tilde{X}_{i\ell} = \frac{X_{i\ell} - \bar{X}_\ell}{s_\ell}, \tag{2.48}$$

where s_Y and s_ℓ are the sample standard deviations of the phenotype and genotype at locus ℓ, respectively. In this case, the marginal effect estimate of locus ℓ will be,

$$\tilde{\beta}_\ell = \frac{Cov[\tilde{\boldsymbol{X}}_\ell, \tilde{\boldsymbol{Y}}]}{Var[\tilde{\boldsymbol{X}}_\ell]} = \frac{1}{n}\sum_{i=1}^{n}(\tilde{X}_{i\ell} - 0)(\tilde{Y}_i - 0) = \frac{1}{n s_\ell s_Y}\sum_{i=1}^{n}(X_{i\ell} - \bar{X}_\ell)(Y_i - \bar{Y}) = \frac{s_\ell \hat{\beta}_\ell}{s_Y}. \tag{2.49}$$

Thus, with normalized genotypes and phenotypes, the effect estimate is scaled by a factor $\frac{s_\ell}{s_Y}$, but is otherwise unaltered.

The polygenic score in the transformed case $\hat{Y}^*$, ignoring any intercept term (which would be 0), is,

$$\hat{Y}_i^* = \sum_{\ell=1}^{L} \tilde{\beta}_\ell \tilde{X}_{i\ell} = \frac{1}{s_Y}\sum_{\ell=1}^{L} \hat{\beta}_\ell (X_{i\ell} - \bar{X}_\ell). \tag{2.50}$$

With centering and scaling, the polygenic score is a genetic prediction less the average genetic prediction in the GWA study sample, both scaled by a factor of s_Y.

Bias. We can compute the *bias* of the rescaled polygenic score as,

$$\begin{aligned}\mathbb{E}\left[s_Y\hat{Y}_i^* - (Y_i - \bar{Y})\right] &= \sum_{\ell=1}^{L}\mathbb{E}\left[\hat{\beta}_\ell X_{i\ell}\right] - \sum_{\ell=1}^{L}\mathbb{E}\left[\hat{\beta}_\ell \bar{X}_\ell\right] - \mu - \sum_{\ell=1}^{L}\left[\beta_\ell X_{i\ell}\right] + \mathbb{E}\left[\bar{Y}\right] \\ &= \sum_{\ell=1}^{L}\mathbb{E}\left[\left(\hat{\beta}_\ell - \beta_\ell\right)X_{i\ell}\right] + \sum_{\ell=1}^{L}\mathbb{E}\left[\hat{\beta}_\ell \bar{X}_\ell\right] - \mu + \mu + \sum_{\ell=1}^{L}\beta_\ell \bar{X}_\ell \quad (2.51) \\ &= \sum_{\ell=1}^{L}\mathbb{E}\left[\left(\hat{\beta}_\ell - \beta_\ell\right)\left(X_{i\ell} - \bar{X}_\ell\right)\right]\end{aligned}$$

Eq. (2.51) shows that when we center and scale the data, we arrive at the same result. (One must rescale either $\tilde{Y}_i$ or $Y_i - \bar{Y}$ by s_Y or its inverse, respectively, to put the polygenic score and the phenotype on the same scale. The former is more mathematically convenient.)

Mean-squared error. The proof for the *mse* is almost identical to that for the *bias*. We thereby omit it.

Additive genetic variance. If the effect estimates $\tilde{\boldsymbol{\beta}}$ are used instead of $\hat{\boldsymbol{\beta}}$, one must rescale the estimate of heterozygosity from the ancient sample by that estimated in the GWA study (see Equation 1 of Supplementary Note 1 of Wang et al. (Wang et al., 2020) for a related procedure),

$$\tilde{V}_A(\tau) = 2\sum_{\ell=1}^{L}\mathbb{E}\left[\tilde{\beta}_\ell^2\left(\frac{1}{s_\ell^2}\right)\hat{Z}_\ell(\tau)\left(1 - \hat{Z}_\ell(\tau)\right)\right] = 2\sum_{\ell=1}^{L}\mathbb{E}\left[\left(\frac{1}{s_Y}\right)^2\hat{\beta}_\ell^2\hat{Z}_\ell(\tau)\left(1 - \hat{Z}_\ell(\tau)\right)\right]. \quad (2.52)$$

This formulation is less convenient because it has random quantities in both the numerator and the denominator. Given that the units in which Y is measured are arbitrary, we forewent coping with the additional complexity imposed by this scaling.

Correlation coefficient. By similar arguments, one can show that the sample correlation

coefficient is not influenced by centering and scaling.

2.8.6 Spectral representation of the transition density

Because the *spectral representation* of the *transition density* of an allele frequency is so central to our work, we provide a concise exposition here. For a lengthier treatment, we refer the reader to (Griffiths and Spano, 2010) and (Song and Steinrücken, 2012).

We represent the Wright-Fisher diffusion by its backward generator, $\mathscr{L}$. Introducing the quantities $a = 4N\mu$ and $b = 4N\nu$ for the population scaled mutation rates, $\mathscr{L}$ is given by,

$$\mathscr{L}f(z) = \frac{1}{2}z(1-z)\frac{\partial^2}{\partial z^2}\{f(z)\} + \frac{1}{2}[a(1-z) - bz]\frac{\partial}{\partial z}\{f(z)\}, \tag{2.53}$$

where z is frequency of the A_2 allele, and f is a twice continuously differentiable bounded function on $[0, 1]$ (Griffiths and Spano, 2010).

The transition density of the Wright-Fisher diffusion, $p(z, z'; t)$, specifies the likelihood of transitioning from allele frequency z to z' in a time interval $[t, 0]$. The spectral representation expresses the transition density as an infinite sum,

$$p(z, z'; t) = \sum_{j=0}^{\infty} c_j(z')e^{-\lambda_j t}R_j(z), \tag{2.54}$$

where, for $j = 0, 1, 2, \ldots$, $c_j(\cdot)$ is a constant factor that depends on the initial condition, defined below in Eq. (2.56); λ_j is the eigenvalue that corresponds to eigenfunction $R_j(\cdot)$; and $R_j(\cdot)$ is the j-th eigenfunction. The function, $\pi(\cdot)$, is the stationary measure and $\langle \cdot, \cdot \rangle_\pi$ is the inner product with respect to this measure, defined in Eq. (2.57). In the neutral, recurrent mutation model, the stationary measure $\pi(\cdot)$ is given by,

$$\pi(z) = z^{a-1}(1-z)^{b-1}, \tag{2.55}$$

where a and b are the population-scaled mutation rates defined in Eq. (2.53). Notice that Eq. (2.55) is equivalent to the unnormalized density of a beta-distributed random variable with shape parameters a and b; when normalized to integrate to one, $\pi(\cdot)$ is the stationary density $\kappa(\cdot)$, first defined in Eq. (2.21). When the initial condition is a point mass at the initial allele frequency, z, the factor $c_j(z')$ is,

$$c_j(z') = \frac{R_j(z')\pi(z')}{\langle R_j, R_j\rangle_\pi}, \tag{2.56}$$

where $\langle R_j, R_j\rangle_\pi$ is the inner product of $R_j(\cdot)$ with itself. We also refer to an inner product of the form $\langle R_j, R_j\rangle$ as the squared norm of the j-th eigenfunction. More generally, we define the inner product of two arbitrary functions, $f(\cdot)$ and $g(\cdot)$ as,

$$\langle f, g\rangle_\pi \coloneqq \int_{y=0}^{1} f(y)g(y)\pi(y)dy. \tag{2.57}$$

The inner product of two eigenfunctions, R_j and R_k, is then a special case of Eq. (2.57), with

$$\langle R_j, R_k\rangle_\pi = \begin{cases} \Delta_j(a,b) & \text{for } k = j, \\ 0 & \text{else,} \end{cases} \tag{2.58}$$

where,

$$\Delta_j(a,b) = \frac{\Gamma(j+a)\Gamma(j+b)}{(2j+a+b-1)\Gamma(j+a+b-1)\Gamma(j+1)}, \tag{2.59}$$

and $\Gamma(z) = \int_0^\infty x^{z-1}e^{-x}dx$ for $z \in \mathbb{R}$. Our work involves many inner products of the form, $\langle R_j, P_k\rangle_\pi$, where $P_k(\cdot)$ is a polynomial of degree k.

In the neutral recurrent mutation model, the eigenfunctions of the Wright-Fisher diffusion are Jacobi polynomials (Griffiths and Spano, 2010). The Jacobi polynomials are polynomials of increasing order coincident with their indices, $j = 0, 1, 2, \ldots$, and obey a three-term

recurrence relation,

$$
\begin{aligned}
zR_j(z) = &\frac{(j+a-1)(j+b-1)}{(2j+a+b-1)(2j+a+b-2)}R_{j-1}(z) \\
&+ \left[\frac{1}{2} - \frac{b^2-a^2-2(b-a)}{2(2j+a+b)(2j+a+b-2)}\right] R_j(z) \\
&+ \frac{(j+1)(j+a+b-1)}{(2j+a+b)(2j+a+b-1)}R_{j+1}(z).
\end{aligned} \tag{2.60}
$$

For $j = 0$,

$$
zR_0(z) = \frac{a}{a+b}R_0(z) + \frac{1}{a+b}R_1(z), \tag{2.61}
$$

with $R_0(z) \equiv 1$.

In our work we will exploit two properties of the Jacobi polynomials: (i) the orthogonality of the eigenfunctions, i.e., Eq. (2.58), and (ii) the fact that a Jacobi polynomial of degree j is orthogonal to all lower order polynomials (of degree k, $k < j$).

2.8.7 Detailed derivations of the metrics

We provide detailed derivations of the metrics used to characterize ancient polygenic scores. For all but $bias(\tau)$, we restrict ourselves to equal detection thresholds, although our framework readily accommodates asymmetric detection thresholds. In order to represent the

metrics succinctly, we introduce several variables:

$$
\begin{aligned}
P^{(d)} &:= \sum_{i=0}^{d-1} \binom{2n}{i} \frac{B(a+i, a+2n-i)}{B(a,a)} \\
P_1^{(d)} &:= \sum_{i=0}^{d-1} \left(\frac{i-n}{n}\right)^2 \binom{2n}{i} \frac{B(a+i, a+2n-i)}{B(a,a)} \\
P_2^{(d)} &:= \sum_{i=0}^{d-1} \left(\frac{(i-n)^2}{n(a+n)}\right) \binom{2n}{i} \frac{B(a+i, a+2n-i)}{B(a,a)} = \left(\frac{n}{a+n}\right) P_1^d \\
P_3^{(d)} &:= \sum_{i=0}^{d-1} \binom{2n}{i} \frac{B(a+i, a+2n-i)}{B(a,a)} \left(\frac{(2a+1)i(i-2n) + an(2n-1)}{(2a+2n+1)(a+n)}\right) \\
P_4^{(d)} &:= \sum_{i=0}^{d-1} \binom{2n}{i} \frac{B(a+i, a+2n-i)}{B(a,a)} i.
\end{aligned}
\tag{2.62}
$$

The sums of the variables defined in Eq. (2.62) for $d = 2n + 1$ are,

$$
S = a + 1, \qquad S_1 = \frac{a+n}{(1+2a)n}, \qquad S_2 = \frac{1}{2a+1}, \qquad S_3 = 0, \qquad S_4 = n. \tag{2.63}
$$

respectively, and with S_3 provided for completeness. The pervasive beta functions in Eq. (2.62) are a consequence of the sampling polynomial implicit in the threshold model, see Eqs. (2.3) and (2.37). For example, $P^{(d)}$ is the *cdf* of a beta-binomial random variable parameterized by the number of chromosomes in the GWA study sample $2n$ and the mutation rate a. As we state in the main text, $\mathbb{P}\{\hat{\beta} = \beta\} = 1 - 2P^{(d)}$ when the detection thresholds are both equal to d. The second variable, $P_1^{(d)}$ arises from moments of the form $\mathbb{E}[\bar{X}\hat{\beta}^2]$, the expectation of the product of the mean genotype in the GWA study sample and the effect estimate. The factor $(i-n)/n$ relates the mean genotype $\bar{X}$ to the allele count D, i.e., $\bar{X} = (D-n)/n$. The remaining terms are less immediately interpretable; their rational is implicit in the derivations presented below and the moments provided in Section 2.8.15.

A form of the polygenic score bias for arbitrary thresholds

The bias of a polygenic score for an individual sampled at time τ in the past and a GWA study conducted at present is,

$$\begin{aligned} bias(\tau) = \mathbb{E}\left[\hat{Y}(\tau) - Y(\tau)\right] &= \mathbb{E}\left[\hat{C}\right] - C + \sum_{\ell=1}^{L} \mathbb{E}\left[X_\ell(\tau)(\hat{\beta}_\ell - \beta_\ell)\right] + \mathbb{E}\left[\epsilon(\tau)\right] \\ &= \sum_{\ell=1}^{L} \beta_\ell \mathbb{E}\left[\bar{X}_\ell\right] - \mathbb{E}\left[\bar{X}_\ell \hat{\beta}_\ell\right] + \mathbb{E}\left[X_\ell(\tau)\hat{\beta}_\ell\right] - \beta_\ell \mathbb{E}\left[X_\ell(\tau)\right]. \end{aligned} \tag{2.64}$$

Further simplification of Eq. (2.64) yields Eq. (2.11). Using the moments derived in Section 2.8.15, we can simplify Eq. (2.11),

$$\begin{aligned} bias_\ell(\tau) &= \beta_\ell \sum_{i=d_{\ell 1}}^{2n-d_{\ell 2}} \binom{2n}{i} \frac{B(a+i, a+2n-i)}{B(a,a)} \left(e^{-a\tau}\left(\frac{i-n}{a+n}\right) - \frac{i-n}{n}\right) \\ &= \beta_\ell \left(e^{-a\tau}\left(\frac{1}{a+n}\right) - \frac{1}{n}\right)\left[n\left(P^{(d_{\ell 1})} - P^{(d_{\ell 2})}\right) - \left(P_4^{(d_{\ell 1})} - P_4^{(d_{\ell 2})}\right)\right] \\ &\approx \beta_\ell \left(e^{-a\tau} - 1\right)\left[\left(P^{(d_{\ell 1})} - P^{(d_{\ell 2})}\right) - \frac{1}{n}\left(P_4^{(d_{\ell 1})} - P_4^{(d_{\ell 2})}\right)\right], \end{aligned} \tag{2.65}$$

where the last line follows for $a \ll n$. As stated in Section 2.4.1, for equal mutation rates and symmetric detection thresholds, $bias(\tau)$ is 0 for all τ.

In Fig. 2.5a, we plot $bias_\ell(\tau)$ in the presence of detection asymmetry for a larger range of mutation rates than presented in the main text, $a \in \{10^{-4}, 10^{-3}, 10^{-2}, 1\}$. In addition, we vary the GWA study sample size over three orders of magnitude, $n = \{10^4, 10^5, 10^6\}$. While the mutation rate $a = 1$ is not biologically plausible—this extreme illustrates features of our model that further illuminate, by contrast, the small mutation rate regime. For example, when $a = 1$ the probability of detecting a locus as significant depends heavily on the GWA study sample size n (Fig. 2.5c). Specifically, for $a = 1$, $P^{(d_\ell)} = \frac{d_\ell}{2n}$ increases linearly with d_ℓ. In contrast, for $a \ll 1$ and modest n, this probability is insensitive to n (Fig. 2.5b).

Specifically, $P^{(d_\ell)} \approx 0.5$ for all values of n and d_ℓ as most of the allele frequencies are very close to, or equal to zero or one, and thus will always elude detection in GWA studies with finite sample sizes. In other words, once n is large enough, varying d_ℓ yields diminishing returns.

In Fig. 2.5a, we set $d_{\ell 1} = 1$ and $d_{\ell 2} = n$ for each sample size to illustrate the effects of an extreme imbalance. For $d_{\ell 2} = n$, positive effect alleles cannot be detected, while $d_{\ell 1} = 1$ implies that a negative effect allele will be detected as long as it is segregating the in the GWA study sample. As $d_{\ell 1} < d_{\ell 2}$, sites where the trait-decreasing allele is at higher frequency in the GWA study ($D_\ell < n$) will be detected more often than sites where the trait-increasing allele is at higher frequency ($D_\ell > n$). This implies that sites where the majority of individuals in the GWA study possess negative effect alleles are more likely to have non-zero effects in the genetic prediction. At the same time, the majority of sites contributing to the estimated intercept $\hat{C}$ will have $\bar{X}_\ell > 0$, and thus, in expectation, $\hat{C} \geq 0$. Thus, at $\tau = 0$, the excess positive contributions to the estimated intercept are tempered by the excess negative contributions to the genetic prediction. As τ increases, $bias_\ell(\tau)$ becomes more positive (Fig. 2.5a). This is because the estimated intercept $\hat{C}$ is constant, whereas, the expected value of the genetic prediction approaches zero with increasing τ. The latter follows from the fact that as τ increases genotype of the ancient sample $X_\ell(\tau)$ becomes independent of average genotype in the GWA study $\bar{X}_\ell$, and its expected value approaches zero. Thus, in the large τ limit, $bias(\tau) = \mathbb{E}[\hat{C}]$, which is positive for $d_1 < d_2$.

Approximating the increase of $bias_\ell(\tau)$, given in Eq. (2.65), for small a and large n,

$$bias_\ell(\tau) \approx bias_\ell(0) + \beta_\ell a \tau \left(P^{(d_{\ell 1})} - P^{(d_{\ell 2})} \right), \tag{2.66}$$

gives us additional insight into these results. In Eq. (2.66), $bias_\ell(0)$ is an exact expression for the $bias_\ell(\tau)$ evaluated at $\tau = 0$; and $P^{(d_{\ell i})}$ is the probability that the allele count in the GWA study, D_ℓ, is less than $d_{\ell i}$ for $i = 1, 2$. Under our assumptions, $P^{(d_{\ell i})}$ is the

cumulative distribution function (*cdf*) of a beta-binomial random variable with $2n$ trials, parameterized by the mutation rate a. At $\tau = 0$, the focal individual is an independent sample from the GWA study population. Thus, the intercept term captures contributes to $bias_\ell(\tau)$ exclusively due to finite sampling. For $\tau > 0$, allelic turnover induces changes in the frequencies of sites not detected in the GWA study (ℓ such that $\hat{\beta}_\ell = 0$), which may contribute to the phenotypic variation of ancient individuals. Thus, the linear term captures additional bias due to finite sampling *and* allelic turnover.

We conclude that increases in $bias_\ell(\tau)$ with τ depend primarily on the difference in the detection probabilities for trait-increasing and decreasing alleles, i.e., $P^{(d_{\ell 1})} - P^{(d_{\ell 2})}$. For small a, this difference is small relative to the (square root) of the additive genetic variance V_A due to the fact that the detection probability is insensitive to the threshold d_ℓ. However, differences in sample size are apparent when the mutation rate is small—with larger sample sizes yielding a larger bias (Fig. 2.5a). This sample size dependency is due to the fact that increased power to detect low frequency alleles with larger n results in a larger difference between the one-sided detection threshold. As a approaches one, the effects of sample size diminish (in log scale). For $a = 1$, the difference in one-sided detection probabilities $P^{(1)} - P^{(n)} = \frac{n-1}{2n}$, which will be close to $\frac{1}{2}$ for modest values of n. In addition, for large a, $bias_\ell(\tau)$ is non-negligible relative to (the square root of) $\mathbb{E}[V_A]$.

Deriving the mean-squared error

Substituting the definitions of $\hat{Y}(\tau)$ and $Y(\tau)$, we can simplify the expression for the mean-squared error (mse),

$$\begin{aligned} mse(\tau) = \mathbb{E}\left[\left(\hat{Y}(\tau) - Y(\tau)\right)^2\right] &= \mathbb{E}\left[\left((\hat{C} - C) + \sum_{\ell=1}^{L} X_\ell(\hat{\beta}_\ell - \beta_\ell) - \epsilon(\tau)\right)^2\right] \\ &= \mathbb{E}\left[\left(\sum_{\ell=1}^{L}(X_\ell(\tau) - \bar{X}_\ell)(\hat{\beta}_\ell - \beta_\ell) + (\bar{\epsilon} - \epsilon(\tau))\right)^2\right] \\ &= \sum_{\ell=1}^{L}\mathbb{E}\left[(X_\ell(\tau) - \bar{X}_\ell)^2(\hat{\beta}_\ell - \beta_\ell)^2\right] + \mathbb{E}\left[(\bar{\epsilon} - \epsilon(\tau))^2\right], \end{aligned} \tag{2.67}$$

where the cross-terms in Eq. (2.67) cancel due to independence between the environmental noise, which has mean 0, and the genotypes. The error term simplifies,

$$\mathbb{E}\left[(\bar{\epsilon} - \epsilon(\tau))^2\right] = \mathbb{E}\left[\bar{\epsilon}^2 - 2\bar{\epsilon}\epsilon(\tau) + (\epsilon(\tau))^2\right] = \left(\frac{n-1}{n}\right)\sigma_e^2. \tag{2.68}$$

When $\sigma_e^2 = 0$, the mse reduces to,

$$\begin{aligned} mse_\ell(\tau) = {} & \mathbb{E}\left[X_\ell^2(\tau)\hat{\beta}_\ell^2\right] - 2\beta_\ell\mathbb{E}\left[X_\ell^2(\tau)\hat{\beta}_\ell\right] + \beta_\ell^2\mathbb{E}\left[X_\ell^2(\tau)\right] \\ & - 2\left(\mathbb{E}\left[X_\ell(\tau)\bar{X}_\ell\hat{\beta}_\ell^2\right] - 2\beta_\ell\mathbb{E}\left[X_\ell(\tau)\bar{X}_\ell\hat{\beta}_\ell\right] + \beta_\ell^2\mathbb{E}\left[X_\ell(\tau)\bar{X}_\ell\right]\right) \\ & + \mathbb{E}\left[\bar{X}_\ell^2\hat{\beta}_\ell^2\right] - 2\beta_\ell\mathbb{E}\left[\bar{X}_\ell^2\hat{\beta}_\ell\right] + \beta_\ell^2\mathbb{E}\left[\bar{X}_\ell^2\right]. \end{aligned} \tag{2.69}$$

For equal detection thresholds $d_{\ell 1} = d_{\ell 2} = d_\ell$, Eq. (2.69) reduces to,

$$mse(\tau) = 2\sum_{\ell=1}^{L}\beta_\ell^2\left[\left(\frac{a+1}{2a+1}\right)P^{(d_\ell)} + P_1^{(d_\ell)} - 2e^{-a\tau}P_2^{(d_\ell)} + \left(\frac{1}{2a+1}\right)e^{-(2a+1)\tau}P_3^{(d_\ell)}\right], \tag{2.70}$$

where $d_\ell = \lceil n\gamma_\ell \rceil$. The change in $mse(\tau)$ is due to the difference between the two exponential terms in Eq. (2.70). From Eq. (2.70), we derive the derivative of $mse(\tau)$,

$$\frac{dmse(\tau)}{d\tau} = 2\sum_{\ell=1}^{L} \beta_\ell^2 \left[2aP_2^{(d_\ell)} e^{-a\tau} - P_3^{(d_\ell)} e^{-(2a+1)\tau} \right], \tag{2.71}$$

which, for small a and τ, is,

$$\frac{dmse(\tau)}{d\tau} \approx 2\sum_{\ell=1}^{L} \beta_\ell^2 \left[2aP_2^{(d_\ell)} - P_3^{(d_\ell)} e^{-\tau} \right] \approx 2a\sum_{\ell=1}^{L} \beta_\ell^2 P^{(d_\ell)} \left(2 - e^{-\tau} \right). \tag{2.72}$$

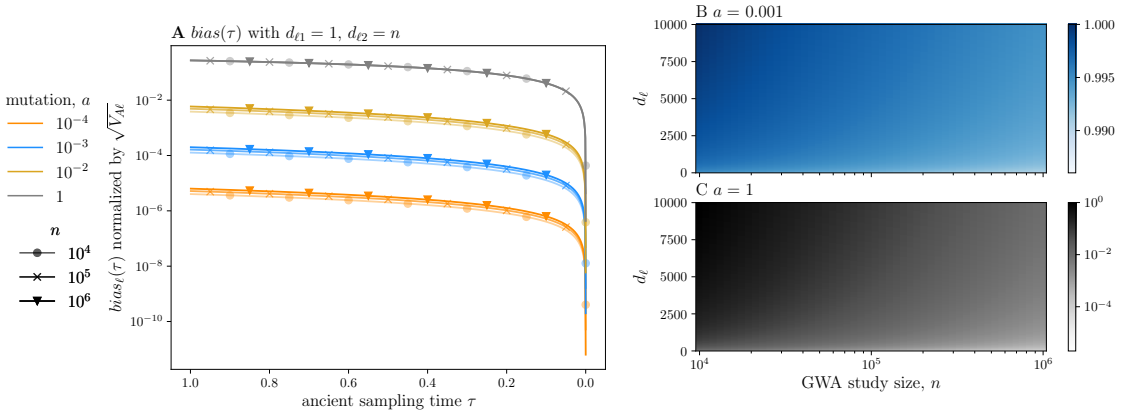


Figure 2.5: **Asymmetry in the detection threshold.** In (A), we plot $bias_\ell(\tau)$ for $a = 10^{-3}$ (orange) and $a = 1$ (blue) across three GWA study sizes, $n = 10^3, 10^4, 10^5$. We set the detection thresholds to $d_1 = 1$ and $d_2 = 100$. In (B) and (C), we plot $2P^{d_\ell}$ as a function of n and the detection threshold d_ℓ for $a = 10^{-3}$ (blue) and $a = 1$ (gray), respectively.

2.8.8 Deriving the expected additive genetic variance

We solve for $\hat{V}_A(\tau)$ in an ancient sample of size n_a and a GWA study sample of size n. Considering a single locus ℓ and conditioning on the ancient and contemporary allele frequencies,

$$\begin{aligned}\hat{V}_{A\ell}(\tau) &= 2\mathbb{E}\left[\mathbb{E}\left[\hat{\beta}_\ell^2\hat{Z}_\ell(\tau)(1-\hat{Z}_\ell(\tau))|Z_\ell, Z_\ell(\tau)\right]\right] \\ &= 2\mathbb{E}\left[\mathbb{E}\left[\hat{Z}_\ell(\tau)(1-\hat{Z}_\ell(\tau))|Z_\ell(\tau)\right]\mathbb{E}\left[\hat{\beta}_\ell^2|Z_\ell\right]\right] \\ &= 2\left(\frac{2n_a-1}{2n_a}\right)\beta_\ell^2\mathbb{E}\left[Z_\ell(\tau)(1-Z_\ell(\tau))\wp(Z_\ell, d_\ell, n)\right],\end{aligned} \tag{2.73}$$

where $\wp(Z_\ell, 2n, d_\ell)$ is the Binomial sampling probability defined in Eq. (2.38). Substituting the spectral representation of the *tdf* yields,

$$\hat{V}_{A\ell}(\tau) = \beta_\ell^2\left(\frac{2n_a-1}{2n_a}\right)\left(\frac{1}{2a+1}\right)\left[a(1-2P^{(d_\ell)}) + 2e^{-(2a+1)\tau}P_3^{(d_\ell)}\right]. \tag{2.74}$$

To compare $\hat{V}_A$ across parameter regimes, we normalize by the expected population additive genetic variance $\mathbb{E}[V_A]$. At stationarity and for mutation rate a,

$$\mathbb{E}\left[V_{A\ell}(\tau)\right] = 2\beta_\ell^2\int_{z_\tau} z_\tau(1-z_\tau)\kappa(z_\tau)dz_\tau = \left(\frac{a}{2a+1}\right)\beta_\ell^2, \tag{2.75}$$

where $\kappa(\cdot)$ is the stationary density given in Eq. (2.21).

For small a, $\hat{V}_A(\tau)$ will change at rate,

$$\frac{d\mathbb{E}[\hat{V}_A(\tau)]}{d\tau} \approx -2\left(\frac{2n_a-1}{2n_a}\right)e^{-\tau}\sum_{\ell=1}^{L}\beta_\ell^2P_3^{(d_\ell)} \approx -2\left(\frac{2n_a-1}{2n_a}\right)e^{-\tau}a\sum_{i=1}^{L}\beta_\ell^2P^{(d_\ell)}, \tag{2.76}$$

where the right hand expression follows from the approximation $P_3^{(d_\ell)} \approx aP^{(d_\ell)}$ (see Section 2.8.11).

2.8.9 Deriving the approximate sample correlation coefficient

In this subsection, we (i) describe the two approximation steps implicit in our definition of $\rho^2(\tau)$ given in Eq. (2.10); (ii) derive an explicit form for $\rho^2(\tau)$; and (iii) derive the approximate decay of relative accuracy $\rho^2(\tau)/\rho^2(0)$.

(i) A practitioner is often interested in the accuracy of their predictor with respect to a particular sample. This sample correlation coefficient (for n_a ancient individuals from τ) is defined as,

$$r(\tau) \coloneqq \frac{Cov[\hat{\boldsymbol{Y}}(\tau), \boldsymbol{Y}(\tau)]}{\sqrt{Var[\hat{\boldsymbol{Y}}]Var[\boldsymbol{Y}]}} = \frac{\sum_{i=1}^{n_a}(\hat{Y}_i(\tau) - \bar{\hat{Y}}(\tau))(Y_i(\tau) - \bar{Y}(\tau))}{\sqrt{\sum_{i=1}^{n_a}(\hat{Y}_i(\tau) - \bar{\hat{Y}}(\tau))^2}\sqrt{\sum_{i=1}^{n_a}(Y_i(\tau) - \bar{Y}(\tau))^2}}, \tag{2.77}$$

where $Cov[\cdot,\cdot]$ and $Var[\cdot]$ are the sample covariance and variance operators, respectively; and $\hat{\boldsymbol{Y}}(\tau), \boldsymbol{Y}(\tau) \in \mathbb{R}^{n_a}$ are the n_a-dimensional vectors of polygenic scores and phenotypes, respectively. Ultimately, we will approximate the expectation of the squared sample correlation coefficient $r^2(\tau)$ with a ratio of expectations,

$$\mathbb{E}\left[r^2(\tau)\right] \approx \frac{\mathbb{E}\left[Cov[\hat{\boldsymbol{Y}}(\tau), \boldsymbol{Y}(\tau)]\right]^2}{\mathbb{E}[Var[\hat{\boldsymbol{Y}}(\tau)]]\mathbb{E}\left[Var[\boldsymbol{Y}(\tau)]\right]}, \tag{2.78}$$

which we defined as $\rho^2(\tau)$ in Eq. (2.10). To arrive at this approximation, we must first approximate the expectation of the ratio in Eq. (2.77) as the ratio of expectations. Second, we must pull the expectation inside the square roots in the denominator. A full investigation of the validity of these steps in general is beyond the scope of the present study. Rather, we validate these approximation steps by simulations of a few parameter regimes of particular interest.

After these two approximation steps, we compute the quantity in Eq. (2.78) exactly under our framework. We take each element of Eq. (2.78) in turn.

(ii) When we plug in our modeling assumptions, the numerator of Eq. (2.77) becomes

$$Cov[\hat{\boldsymbol{Y}}(\tau), \boldsymbol{Y}(\tau)] = \sum_{i=1}^{n_a} \sum_{\ell,\ell'} \hat{\beta}_\ell \beta_{\ell'} (X_{i\ell}(\tau) - \bar{X}_\ell(\tau))(X_{i\ell'}(\tau) - \bar{X}_{\ell'}(\tau)) + Cov[\hat{\boldsymbol{Y}}(\tau), \boldsymbol{\epsilon}], \quad (2.79)$$

due to the linearity of the covariance operator, with $\boldsymbol{\epsilon} \in \mathbb{R}_a^n$ as the vector of environmental effects. In expectation, assuming *iid* loci with equal effects β and *iid* ancient samples,

$$\mathbb{E}\left[Cov[\hat{\boldsymbol{Y}}(\tau), \boldsymbol{Y}(\tau)]\right] = \frac{1}{n_a} \sum_{i=1}^{n_a} \sum_{\ell=1}^{L} \beta_\ell \mathbb{E}\left[\hat{\beta}_\ell (X_{i\ell}(\tau) - \bar{X}_\ell(\tau))^2\right] = L\beta \mathbb{E}\left[\hat{\beta}(X_i(\tau) - \bar{X}(\tau))^2\right]. \quad (2.80)$$

Similarly,

$$\mathbb{E}\left[Var[\hat{\boldsymbol{Y}}(\tau)]\right] = L\mathbb{E}\left[\hat{\beta}^2 (X_i(\tau) - \bar{X}(\tau))^2\right], \quad (2.81)$$

which, under our simple threshold model is equal to the expectation of the covariance given in Eq. (2.80). Finally,

$$\mathbb{E}\left[Var[\boldsymbol{Y}(\tau)]\right] = L\beta^2 \mathbb{E}\left[(X(\tau) - \bar{X}(\tau))^2\right] + \left(\frac{n_a - 1}{n_a}\right) \sigma_e^2. \quad (2.82)$$

All together, our approximation for $r^2(\tau)$ reduces to,

$$\mathbb{E}\left[r^2\right] \approx \frac{L\beta \mathbb{E}\left[\hat{\beta}(X(\tau) - \bar{X}(\tau))^2\right]}{L\beta^2 \mathbb{E}\left[(X(\tau) - \bar{X}(\tau))^2\right] + \left(\frac{n_a - 1}{n_a}\right) \sigma_e^2} = \frac{\mathbb{E}\left[\hat{\beta}(X(\tau) - \bar{X}(\tau))^2\right] / \beta}{\mathbb{E}\left[(X(\tau) - \bar{X}(\tau))^2\right] + \left(\frac{n_a - 1}{n_a}\right) \sigma_{e'}^2}, \quad (2.83)$$

where $\sigma_{e'}^2 = \sigma_e^2 / (L\beta^2)$. All that remains is to solve for the expectations in the numerator and denominator of Eq. (2.83). The first involves both the GWA study and ancient sample

times,

$$\begin{aligned}\mathbb{E}\left[\hat{\beta}(X(\tau)-\bar{X}(\tau))^2\right] &= \mathbb{E}\left[\mathbb{E}\left[\hat{\beta}(X(\tau)-\bar{X}(\tau))^2|Z(0),Z(\tau)\right]\right] \\ &= \beta\mathbb{E}\left[\wp(Z(0))\mathbb{E}\left[(X(\tau)-\bar{X}(\tau))^2|Z(\tau)\right]\right] \\ &= 2\beta\left(\frac{n_a-1}{n_a}\right)\mathbb{E}\left[\wp(Z(0))Z(\tau)(1-Z(\tau))\right],\end{aligned} \tag{2.84}$$

which we recognize as closely related to the expected estimated additive genetic variance, $\hat{V}_A(\tau)$, with $\wp(Z(0))$ defined in Eq. (2.38). The expectation in the denominator is,

$$\mathbb{E}\left[\hat{\beta}(X(\tau)-\bar{X}(\tau))^2\right] = 2\left(\frac{n_a-1}{n_a}\right)\mathbb{E}\left[Z(\tau)(1-Z(\tau))\right] = \left(\frac{n_a-1}{n_a}\right)\left(\frac{a}{2a+1}\right), \tag{2.85}$$

which is equal to the $\mathbb{E}[V_A]$ at stationarity normalized by the squared true effect β^2 and multiplied by the n_a-dependent factor to account for ancient sample size. We then see that our approximation to the expectation of $r^2(\tau)$ is insensitive to the ancient sample size and equal to,

$$\mathbb{E}\left[r^2(\tau)\right] \approx \rho^2(\tau) \coloneqq \frac{2\mathbb{E}\left[\wp(Z(0))Z(\tau)(1-Z(\tau))\right]}{\frac{a}{2a+1}+\sigma^2_{e'}} = \left(\frac{2n_a}{2n_a-1}\right)\frac{\hat{V}_{A\ell}(\tau)/\beta^2}{\frac{a}{2a+1}+\sigma^2_{e'}}, \tag{2.86}$$

where the ancient sample size dependent factor in the rightmost expression cancels with its inverse in $\hat{V}_{A\ell}(\tau)$. Thus, the sample correlation coefficient is proportional to the estimated additive genetic variance, and its derivative is given by,

$$\begin{aligned}\frac{d\mathbb{E}\left[r^2(\tau)\right]}{d\tau} &\approx \frac{d}{d\tau}\left(\frac{2n_a}{2n_a-1}\right)\frac{\hat{V}_{A\ell}(\tau)/\beta^2}{\frac{a}{2a+1}+\sigma^2_{e'}} \\ &\approx \left(\frac{1}{\frac{a}{2a+1}+\sigma^2_{e'}}\right)\left(\frac{2}{2a+1}\right)(-(2a+1))e^{-(2a+1)\tau}P_3^{(d)} \\ &\approx \left(\frac{1}{a+\sigma^2_{e'}}\right)2e^{-\tau}P_3^{(d)} \approx \left(\frac{a}{a+\sigma^2_{e'}}\right)2e^{-\tau}P^{(d)},\end{aligned} \tag{2.87}$$

where the last line follows for $a \ll 1$, as $e^{-(2a+1)\tau} \approx e^{-\tau}$ and $P_3^{(d)} \approx aP^{(d)}$, see Eq. (2.101).

(iii) We show that for small mutation rates, relative accuracy decays at a rate that is independent of the mutation rate a and detection threshold d. For *iid* loci,

$$\rho^2(\tau)/\rho^2(0) = \frac{a(1-2P^{(d)}) + 2e^{-(2a+1)\tau}P_3^{(d)}}{a(1-2P^{(d)}) + 2P_3^{(d)}} \approx \frac{2e^{-\tau}P_3^{(d)}}{2P_3^{(d)}} = e^{-\tau}, \tag{2.88}$$

where, we have claimed that $a(1-2P^{(d)}) \approx 0$ for all $d \in \{1, \ldots, n\}$, and that $2a+1 \approx 1$. If we relax the *iid* assumption, we have,

$$\rho^2(\tau)/\rho^2(0) = \frac{\sum_{\ell=1}^{L} \beta_\ell^2 \left[a(1-2P^{(d_\ell)}) + 2e^{-(2a+1)\tau}P_3^{(d_\ell)}\right]}{\sum_{\ell=1}^{L} \beta_\ell^2 \left[a(1-2P^{(d_\ell)}) + 2P_3^{(d_\ell)}\right]}, \tag{2.89}$$

which could be computed for a given distribution of β_ℓ. For $a \ll 1$, the $a(1-2P^{(d_\ell)})$ terms in Eq. (2.89) *may be* negligible, yielding the same result as Eq. (2.88), which implies that relative accuracy is insensitive to distributional assumptions on β for small a. However, more rigorous theoretical and simulation-based work is required to assess the accuracy of this claim.

2.8.10 Expected accuracy in the UK Biobank

Our theory characterizes ancient polygenic scores in a highly idealized setting. Namely, the population size is constant, allele frequencies evolve neutrally at stationarity, and the estimation of effects coheres with a simple threshold model. In addition, we provide statistics parameterized by a single fixed effect size (although see Section 2.8.9 where we begin to relax this assumption). In practice, many of these assumptions are likely violated. For example, human populations have undergone numerous population size changes, including both bottlenecks and expansions, as well as admixture events (Nielsen et al., 2017). In addition,

many human traits are thought to be under some form of selection, which necessarily alters the allele frequency dynamics of causal loci and neutral loci nearby (e.g. Gazal et al. (2017); O'Connor et al. (2019); Zeng et al. (2021)). And, confounding factors like population structure may still complicate interpretation of the results of GWA studies (Sohail et al., 2019; Berg et al., 2019). Lastly, causal effect sizes of complex traits, e.g. height, vary across loci. This distribution of effects is difficult to characterize, likely with significant mass near zero (e.g. see Zhou et al. (2013)).

In this section, we tackle the last of these complications: that effect sizes are different at each locus, while still retaining the other simplifying assumptions. We model the variation among effect sizes by assuming that each effect is random and *iid*, i.e., independent and drawn from the same probability distribution. We do not attempt to estimate this distribution from the summary statistics, as for example in (Zhou et al., 2013; Moser et al., 2015; Zhang et al., 2018). Instead, we consider several parameterizations of the causal effect size distribution, all with ample mass near zero, including a distribution estimated from GWA study summary statistics in (Zhang et al., 2018).

Using the UK Biobank summary statistics, we estimate the relationship between the minimum allele frequency required to detect a SNP with an effect size β as significant under a particular significance threshold α. In essence, we are replacing our theoretical parameterization of the per-locus detection threshold d_ℓ, Eq. (2.31) in Section 2.8.2, with one derived from data. Then, assuming the population is at equilibrium, we compute the approximate decay in accuracy, as measured by $\rho^2(\tau)$ in Eq. (2.10), for each of the causal distributions and arbitrary ancient sampling times.

Causal effect size distributions. As the causal effect size distribution for human height is unknown, we consider several potential distributions (Fig. 2.6). Namely we model the absolute values of the effect sizes $|\beta|$ as (1) exponential random variables with rate $\lambda \in \{10, 100, 500\}$, referred to as $f_{\exp}(\cdot; \lambda)$; (2) Gamma distributed random variables with shape

parameter $\lambda \in \{10^{-3}, 0.5\}$ and scale parameter equal to one, referred to as $f_\gamma(\cdot; \lambda)$; and (3) a mixture of folded normal distributions estimated from GWA study summary statistics in (Zhang et al., 2018),

$$f_{\text{mix}}(b) = 0.9 \cdot \mathcal{N}^*(b; 0, 1.5439 \cdot 10^{-5}) + 0.1 \cdot \mathcal{N}^*(b; 0, 2.021 \cdot 10^{-4}), \tag{2.90}$$

for some effect size b, where $\mathcal{N}^*(b; 0, \sigma^2)$ denotes the likelihood of effect size b under a folded normal distribution with mean zero and variance σ^2. In all cases, we discretize these distributions over a set $\boldsymbol{\beta}$ of 5000 linearly spaced values in the range $[10^{-4}, 0.1]$. In doing so, we are excluding very large effect mutations— such mutations would be more likely to substantially deviate from neutral dynamics—and mutations with effects indistinguishable from zero.

Estimating the relationship between effect size and the detection threshold. We use the height summary statistics to estimate a function relating effect size to the allele frequency detection threshold, denoted by $g_\alpha(\cdot)$. The function $g_\alpha(b)$ specifies the minimum allele frequency required to detect an effect of size b as non-zero under a given significance threshold α. Here, we use $\alpha = 10^{-8}$ to account for the multiple testing burden imposed by conducting approximately 12 million association tests. We compute the minimum effect size detected as significant (p-value $\leq \alpha$) among SNPs within 250 non-overlapping, log-spaced allele frequency bins, and subsequently interpolate between these minima to specify $g_\alpha(\cdot)$ over the continuous interval $[10^{-3}, 0.5]$. Finally, we "smooth" this function by forcing it to be non-increasing.

Computing the accuracy metrics. An effect size distribution (described above) coupled with our equilibrium assumption specifies the expected additive genetic variance, V_A. For a

population-scaled mutation rate of a,

$$\mathbb{E}[V_A] = \sum_{b \in \boldsymbol{\beta}} \left(\frac{a}{2a+1} \right) b^2 f_\cdot(b), \tag{2.91}$$

where the sum is over all discretized effect sizes b in the set of effect sizes $\boldsymbol{\beta}$; $a/(2a+1)$ is the expected genetic variance at stationarity (Section 2.8.15); and $f_\cdot(\cdot)$ is one of the effect size distributions described above. We can then compute the approximate sample correlation coefficient $\rho^2(\tau)$. In particular, following Eq. (2.19) and ignoring the ancient sample size dependent factor in the GWA study,

$$\rho^2(0) = \frac{\hat{V}_A(0)}{V_A + \sigma_e^2} = \frac{\sum_{b \in \beta} \hat{V}_{Ab}(0) f_\cdot(b)}{\sum_{b \in \beta} \left(\frac{a}{2a+1} \right) b^2 f_\cdot(b) + \sigma_e^2}, \tag{2.92}$$

and following Eq. (2.74),

$$\hat{V}_{Ab}(0) = \mathbb{E}\big[2\hat{\beta}^2 Z(0)(1 - Z(0)) | \beta = b\big] = b^2 \left(\frac{1}{2a+1} \right) \big[a(1 - 2P^{(d_b)}) + 2P_3^{(d_b)} \big], \tag{2.93}$$

where $d_b = \lceil 2n g_\alpha(b) \rceil$ is the allele count threshold derived from the function $g_\alpha(\cdot)$, and with $P^{(\cdot)}$ and $P_3^{(\cdot)}$ defined in Eq. (2.62). Importantly, the denominator of Eq. (2.92) includes non-zero contributions from SNPs that may not achieve genome-wide significance (unequivocally those SNPs with effect sizes to the left of the dashed lines in Fig. 2.6).

When narrow-sense heritability $h^2 = 0.5$, the environmental variance σ_e^2 is equal to V_A. For an arbitrary ancient sampling time τ,

$$\rho^2(\tau) = \frac{\sum_{b \in \beta} \hat{V}_{Ab}(\tau) f_\cdot(b)}{2 \sum_{b \in \beta} \left(\frac{a}{2a+1} \right) b^2 f_\cdot(b)} = \frac{\sum_{b \in \beta} \left[a(1 - 2P^{(d_b)}) + 2e^{-(2a+1)\tau} P_3^{(d_b)} \right] b^2 f_\cdot(b)}{2a \sum_{b \in \beta} b^2 f_\cdot(b)}, \tag{2.94}$$

where d_b, defined in Eq. (2.93), is the empirically estimated allele count threshold corresponding to an effect of size b. We compute Eq. (2.94) with the following parameter values.

While not shown, we can similarly compute the other accuracy metrics (normalized by V_A as we do not know the number of causal sites L).

Notably, Fig. 2.7 shows, not unexpectedly, that the shape of the causal distribution influences accuracy. However, relative accuracy is indistinguishable across the different distributions. Thus, as observed in Section 2.4.4 (and speculated on in Section 2.8.9), relative accuracy appears insensitive to assumptions on the effect size distribution.

It is important to note that several of these distributions yield predicted accuracies for contemporary samples that overestimate the observed prediction accuracy for height in the UK Biobank sample, for example, estimated to be 0.193 in (Wang et al., 2020). This discrepancy suggests that these distributions may not be good approximations to reality. Although, we note that $f_{\text{mix}}(\cdot)$, the distribution estimated in (Zhang et al., 2018) yields the best approximation to observed prediction accuracy in the present day sample.

Nonetheless, many of our simplifying assumptions caution against overinterpretation of these results. In particular, our assumption of neutrality implies that alleles are *iid* irrespective of the magnitudes of their effects; loci may also not be at stationarity. Indeed, large effect alleles are likely to be more deleterious and thus subject to stronger selection relative to small effect alleles (Gazal et al., 2017; Hormozdiari et al., 2018; Zeng et al., 2018a). Nonetheless, the application of our theory in this context provides insight into the relationship between the causal effect size distribution and prediction accuracy. Furthermore, we provide some preliminary evidence that relative accuracy may be more robust to violations of at least some of our assumptions.

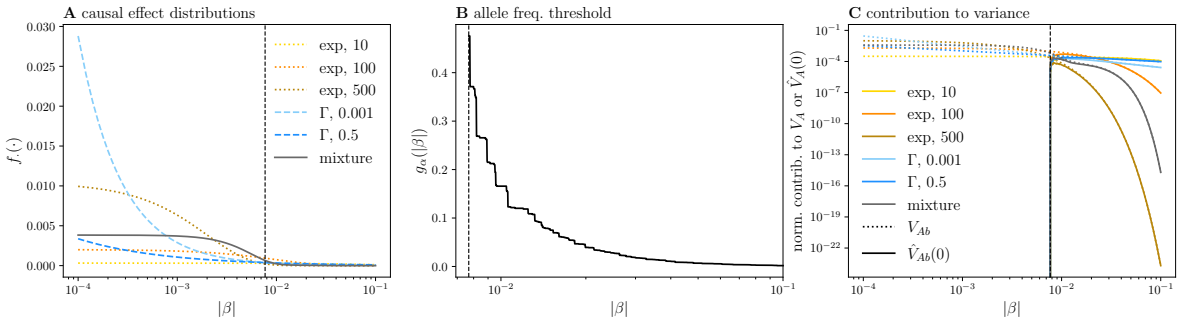


Figure 2.6: **Effect size distribution.** In (A), we plot the six causal distributions described above. The dashed vertical line here, and in (B) and (C), indicates the minimum effect size that can be detected, as specified by $g_\alpha(\cdot)$ of (B). In (B), we find the minimum allele frequency $g_\alpha(\beta)$ required to detect an effect of size β using the significance threshold $\alpha = 10^{-8}$ and a set of 273,671 SNPs with p-values $\leq \alpha$. And, in (C), we plot the relationship between a SNP's effect size and its contribution to the expected additive genetic variance V_{Ab} (dotted lines) or the estimated additive genetic variance at the time of the GWA study $\hat{V}_{Ab}(0)$ (solid lines) for each of the causal distributions. Both values are normalized by the squared effect size β^2 and the expected genetic variance at stationarity, $a/(2a+1)$.

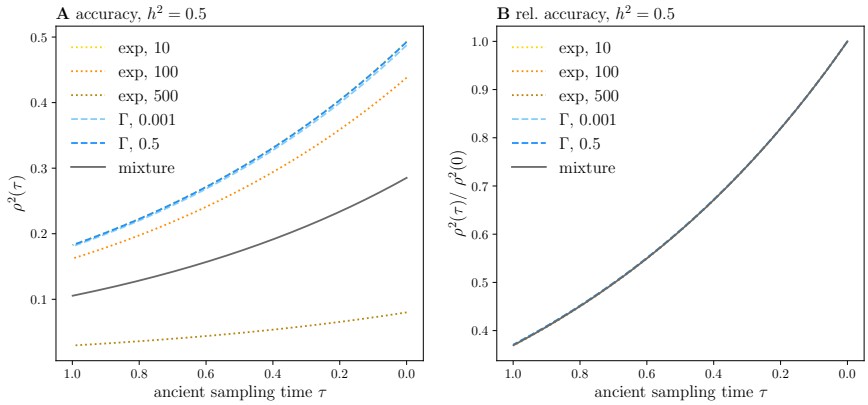


Figure 2.7: **Accuracy and relative accuracy.** In (A), we plot the approximate sample correlation coefficient, or polygenic score accuracy $\rho^2(\tau)$, as a function of ancient sampling time τ. We assume the effect size distributions described above and associated allele frequency thresholds (Fig. 2.6), as well as narrow-sense heritability $h^2 = 0.5$. Relative accuracy is similarly plotted in (B). Note the difference in y-axis limits between panels.

2.8.11 Deriving approximations to the metrics

In the main text, we present several approximations for the initial rate of increase or decrease of the metrics. Here, we show how we arrived at these approximations from the exact forms given in the previous sections. For a given metric, we first compute a first order Taylor series expansion (in τ). We then find the intercept, i.e., the value of the statistic at zero, and the slope. We subsequently make use of the following approximations: (i) $P_1^{(d)} \approx P^{(d)}$; (ii) $P_2^{(d)} \approx P^{(d)}$; and (iii) $P_3^{(d)} \approx aP^{(d)}$.

Approximate metrics. For the *bias*, this approach yields,

$$\begin{aligned} bias_\ell(\tau) &\approx \beta_\ell \left[(1 - a\tau) \left(\frac{1}{a+n} \right) - \frac{1}{n} \right] \left[n \left(P^{(d_{\ell 1})} - P^{(d_{\ell 2})} \right) + \left(P_4^{(d_{\ell 2})} - P_4^{(d_{\ell 1})} \right) \right] \\ &\approx bias_\ell(0) + \beta_\ell \cdot a\tau \left(P^{(d_{\ell 1})} - P^{(d_{\ell 2})} \right), \end{aligned} \tag{2.95}$$

where the last line follows from (i) using the approximations noted in the prelude; (ii) ignoring the $P_4^{(d_{\ell i})}$ terms in the slope (which are order $\mathcal{O}(\frac{1}{n})$); (iii) and $\frac{1}{a+n} \approx \frac{1}{n}$. Using the same approach, $mse_\ell(\tau)$ with equal thresholds d_ℓ, becomes,

$$\begin{aligned} mse_\ell(\tau) &\approx 2\beta_\ell^2 \Big[\left(\frac{a+1}{2a+1} \right) P^{(d_\ell)} + P_1^{(d_\ell)} - 2P_2^{(d_\ell)} + \left(\frac{1}{2a+1} \right) P_3^{(d_\ell)} + 2a\tau P_2^{(d_\ell)} - \tau P_3^{(d_\ell)} \Big] \\ &\approx mse_\ell(0) + 2\beta_\ell^2 a\tau P^{(d_\ell)}. \end{aligned} \tag{2.96}$$

And, we can approximate $\hat{V}_A(\tau)$ as,

$$\begin{aligned} \hat{V}_{A\ell}(\tau) &\approx \left(\frac{2n_a - 1}{2n_a} \right) \beta_\ell^2 \left(\frac{1}{2a+1} \right) \left[\left(a(1 - 2P^{(d_\ell)} \right) + \left(2\left(1 - (2a+1)\tau\right) P_3^{(d_\ell)} \right) \right] \\ &\approx \hat{V}_{A\ell}(0) - 2 \left(\frac{2n_a - 1}{2n_a} \right) \beta_\ell^2 a P^{(d_\ell)} \tau. \end{aligned} \tag{2.97}$$

The approximation for the accuracy $\rho^2(\tau)$ follows immediately from Eq. (2.97).

In (A) and (D) of Fig. 2.8, we plot the approximate rate of change $2aP^{(d)}$, for low and high detection thresholds. In addition, in Fig. 2.9, we compare our exact theoretical results to their approximations over a short time scale of $\tau \in [0.2, 0]$. We observe that the approximation fares better for smaller values of d, as well as, for larger n. Below, we show how some of the steps in our approximations are adversely affected by large d and small n.

Approximation error. We quantify the error incurred in the approximation $P_1^{(d)} \approx P_3^{(d)}$ and $P_2^{(d)} \approx P_3^{(d)}$. To do so, we first express $P_1^{(d)}$ and $P_2^{(d)}$ as functions of $P_3^{(d)}$. We refer to the i-th term in quantities specified in Eq. (2.62) as $P_\cdot^i$, dropping the superscript d for succinctness.

$$P_1^i = \left(\frac{i-n}{n}\right)^2 \binom{2n}{i} \frac{B(a+i, a+2n-i)}{B(a,a)} = \left(\frac{i-n}{n}\right)^2 P^i = \left(\frac{i}{n}\right)^2 P^i - 2\left(\frac{in}{n^2}\right) + P^i. \tag{2.98}$$

Thus,

$$P_1^{(d)} = P^{(d)} + \frac{1}{n^2}\sum_{i=0}^{d-1} i^2 P^i - \frac{2}{n}\sum_{i=0}^{d-1} iP^i. \tag{2.99}$$

Note that the summations in Eq. (2.99) are the second and first moments of a beta-binomial random variable truncated at $d-1$, respectively. As long as the two summations are $O(n)$ and $O(1)$, respectively, the error of the approximation will be smaller than $O(\frac{1}{n})$ as both summations are non-negative. We can repeat the same procedure for $P_2^{(d)}$,

$$P_2^{(d)} = \left(\frac{n}{a+n}\right) P^{(d)} + \frac{1}{n(a+n)}\sum_{i=0}^{d-1} i^2 P^i - \left(\frac{2}{a+n}\right)\sum_{i=0}^{d-1} iP^i \approx P_1^{(d)}, \tag{2.100}$$

where the approximation is valid for $a \ll n$, and in this regime, our analysis in the previous paragraph also applies to $P_2^{(d)}$.

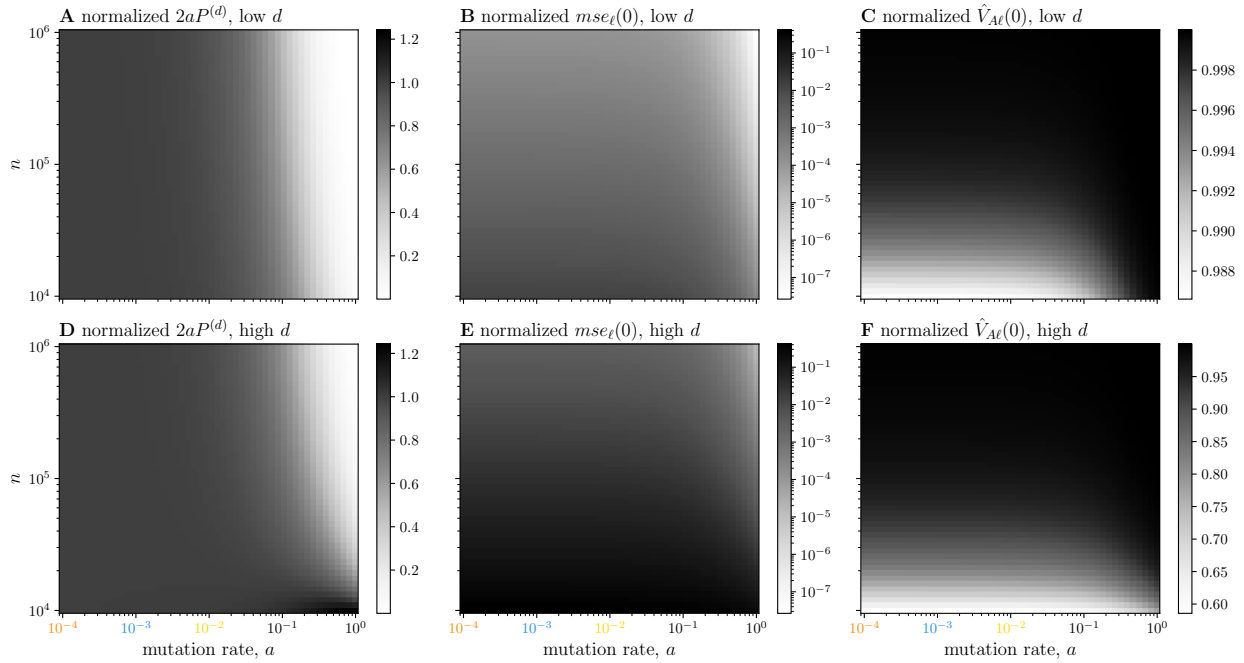


Figure 2.8: **Approximations to the mean-squared error and expected estimated additive genetic variance.** In (A), we plot $2aP^{(d)}$ normalized by $\mathbb{E}[V_A] = \beta^2(\frac{a}{2a+1})$ across a range of mutation rates $a \in \{10^{-4}, \ldots, 1\}$ and GWA study sample sizes n, for a small detection threshold. Here, d is either 132 or 133, corresponding to a squared effect size of $\beta^2 = 0.25$, when the significance threshold is $\alpha = 10^{-8}$ and the phenotypic variance $V_p = 1$. In (B) and (C), we plot the initial $mse_\ell(\tau)$ and $\hat{V}_{A\ell}(\tau)$ (both normalized by the true V_A) for small d. In (D-F), we repeat plots (A-C), except with a higher detection threshold, respectively. The smaller effect size of $\beta^2 = 0.01$ yields thresholds in the range $d \in \{3209, \ldots, 4142\}$, in order of increasing sample size. Note that in contrast to the other pairs of plots, (C) and (F) do not share a scale.

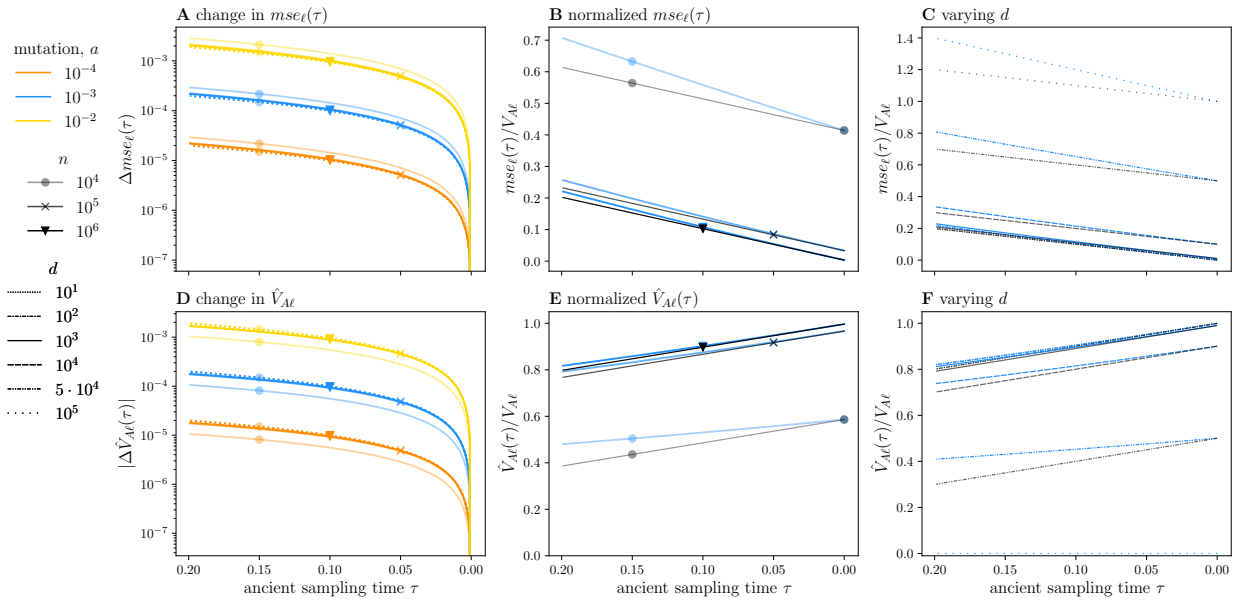


Figure 2.9: **Approximations for the per locus contributions to the mean-squared error and estimated additive genetic variance across sample sizes, mutation rates, and detection thresholds.** This plot is identical to Fig. 2.2 except that we (i) include our approximations to the two statistics, Eq. (2.15) and Eq. (2.17), and (ii) plot our results over a short time frame, $\tau \in [0.2, 0]$. In (A), the approximations are depicted as colored, dotted lines corresponding to each mutation rate and sample size pair. In (B), the approximations are denoted by the same markers and opacity as their blue counterparts. And, in (C), the approximations are provided in black, with line pattern indicating the threshold d.

Finally, we consider the approximation $P_3^{(d)} \approx aP^{(d)}$,

$$\begin{aligned}
P_3^i &= P^i\left(\frac{(2a+1)i(i-2n)+an(2n-1)}{(2a+2n+1)(a+n)}\right)+aP^i-aP^i \\
&= aP^i+P^i\left(\frac{(2a+1)i(i-2n)+an(2n-1)-a(2a+2n+1)(a+n)}{(2a+2n+1)(a+n)}\right) \\
&= aP^i+P^i\left(\frac{(2a+1)i(i-2n)-a(2a^2+4an+a+2n)}{(2a+2n+1)(a+n)}\right) \\
&= aP^i+P^i\left(\frac{(2a+1)i(i-2n)-a(2a+1)(a+2n)}{(2a+2n+1)(a+n)}\right) \\
&= aP^i+(2a+1)P^i\left(\frac{i(i-2n)-a(a+2n)}{(2a+2n+1)(a+n)}\right).
\end{aligned} \tag{2.101}$$

Thus,

$$P_3^{(d)} = aP^{(d)} - \frac{(2a+1)}{(2a+2n+1)(a+n)}\left[a(a+2n)(d-1)P^{(d)}+\sum_{i=0}^{d-1}\binom{2n}{i}\frac{B(a+i,a+2n-i)}{B(a,a)}i(2n-i)\right], \tag{2.102}$$

And,

$$|P_3^{(d)}-aP^{(d)}| \approx \frac{a(d-1)}{n}P^{(d)}+\sum_{i=0}^{d-1}\binom{2n}{i}\frac{B(a+i,a+2n-i)}{B(a,a)}\left(\frac{i}{n}-\frac{i^2}{2n^2}\right) \tag{2.103}$$

where the approximation follows for $a \ll 1$ and $a \ll n$. Thus, $P_3^{(d_\ell)} \approx aP^{(d_\ell)}$ will be a very good approximation when $d \ll n$, but should also hold for modest d as long n is reasonably large. It is possible that when the mutational target is very large, e.g. $O(n)$, the approximation errors may be non-negligible for large enough d. However, large d implies a small β, thereby tempering any approximation errors in practice. An additional benefit of expressing $P_1^{(d)}$, $P_2^{(d)}$, and $P_3^{(d)}$ in terms of $P^{(d)}$ is that we can now take advantage of efficient coding of the beta-binomial probability mass function in the Python module `scipy` to compute analytical results for larger values of d.

Computations for large n and d. For large n, computing the terms in Eq. (2.62), exclud-

ing $P^{(d)}$ (which we compute using `scipy`), becomes computationally prohibitive. However, for large n, we can approximate these quantities as follows. Defining $z = \frac{d}{2n}$ and the incomplete beta function as $I_z(x,y) = \frac{1}{B(x,y)} \int_0^z z^{x-1}(1-z)^{y-1}dz$,

$$
\begin{aligned}
P_1^{(d)} &\approx \left(\frac{1}{n^2}\right) I_z(a+2,a) - \left(\frac{2}{n}\right) I_z(a+1,a) + P^{(d)} \\
P_2^{(d)} &\approx \frac{1}{a+n}\left(\frac{1}{n} I_z(a+2,a) - 2I_z(a+1,a) + nP^{(d)}\right) \\
P_3^{(d)} &\approx \frac{an(2n-1)}{(2a+2n+1)(a+n)} P^{(d)} + \left(\frac{2a+1}{(2a+2n+1)(a+n)}\right)\left(I_z(a+2,a) - 2I_z(a+1,a)\right).
\end{aligned}
\tag{2.104}
$$

These expressions allow us to compute the metrics for a much larger range of n and d values.

2.8.12 Polygenic score bias for recent genic selection

We provide evidence for the claim made in Section 2.5 that, "$bias_\ell(\tau)$ will reach an equilibrium value that depends approximately on the asymmetry of the detection thresholds at the present day, which in turn, depends on both the timing and strength of selection". We treat the simplest case of a detection threshold $d = 1$, i.e., $\hat{\beta} = \beta$ if the locus is variant in the GWA study sample. The time-varying distribution of the allele frequency is $f_t(\cdot)$, and necessarily depends on the timing and strength of selection. For $t > \tau_s$, the time of the onset of selection, $f_t(z) \propto z^{a-1}(1-z)^{a-1}$. For $t \leq \tau_s$, $f_t(z)$ will be skewed toward one and proportional to $\propto e^{\sigma z} z^{a-1}(1-z)^{a-1}$, where $\sigma = 4Ns$ is the population-scaled selection coefficient. For a larger τ_s, $f_t(z)$ for $t \leq \tau_s$ will have more time to shift toward the stationary distribution under selection.

From Eq. (2.11), we have that $bias_\ell(\tau)$, omitting the locus subscript is,

$$\begin{aligned} bias(\tau) &= \mathbb{E}\left[(\bar{X} - X(\tau))(\beta - \hat{\beta})\right] \\ &= \beta\mathbb{E}\left[(\bar{X} - X(\tau))|\hat{\beta} = 0\right]\mathbb{P}\{\hat{\beta} = 0\} \\ &= \beta\left(\mathbb{E}\left[\bar{X}|\hat{\beta} = 0\right] - \mathbb{E}\left[X(\tau)|\hat{\beta} = 0\right]\right)\mathbb{P}\{\hat{\beta} = 0\} \\ &= \beta\left(-\mathbb{P}\{\bar{X} = -1|\hat{\beta} = 0\} + \mathbb{P}\{\bar{X} = +1|\hat{\beta} = 0\} - \mathbb{E}\left[X(\tau)|\hat{\beta} = 0\right]\right)\mathbb{P}\{\hat{\beta} = 0\}. \end{aligned} \tag{2.105}$$

For large τ, $\mathbb{E}\left[X(\tau)|\hat{\beta} = 0\right] \to 0$, such that,

$$bias(\tau) \to \beta\left(\mathbb{P}\{\bar{X} = +1|\hat{\beta} = 0\} - \mathbb{P}\{\bar{X} = -1|\hat{\beta} = 0\}\right)\mathbb{P}\{\hat{\beta} = 0\}, \tag{2.106}$$

which shows that the $bias(\tau)$ will equilibrate at some value that depends on the difference between the + and − detection thresholds as well as the probability that $\hat{\beta} = 0$. This difference, in turn, depends on the time of the onset of selection and the selection coefficient (relative to the mutation rate) itself.

2.8.13 Fixation index and prediction accuracy

The complexity of human population history implies that an ancient sampling time of t years does not readily translate to a coalescent time of $\tau = t/2N$. As a result, it may be difficult to apply our theoretical results in practice. In lieu of an estimated ancient sampling time τ, we instead seek to uncover the relationship between F_{ST} and our various metrics, which may, with some caveats, be more robust to demographic changes. And importantly, F_{ST} is readily measurable from ancient genotypic data.

F_{ST} is defined as the relative difference between within sample and across sample het-

erozygosity (Weir, 1996),

$$F_{\mathrm{ST}} = \frac{\bar{z}(1-\bar{z}) - \overline{z(1-z)}}{\bar{z}(1-\bar{z})}, \tag{2.107}$$

where $\bar{z} = \frac{1}{2}[z(0) + z(\tau)]$ is the average of the allele frequencies in the contemporary and ancient populations, and $\overline{z(1-z)} = \frac{1}{2}[z(0)(1-z(0)) + z(\tau)(1-z(\tau))]$. We can approximate the expectation of F_{ST} by a ratio of expectations and solve under the assumptions of the recurrent mutation model and neutrality,

$$\mathbb{E}[F_{\mathrm{ST}}] = \frac{\mathbb{E}[\bar{z}(1-\bar{z})] - \mathbb{E}[\overline{z(1-z)}]}{\mathbb{E}[\bar{z}(1-\bar{z})]} = \frac{\frac{a+1}{2(2a+1)} - \frac{1}{4}\left(1 + \frac{1}{2a+1}e^{-a\tau}\right)}{1 - \left[\frac{a+1}{2(2a+1)} + \frac{1}{4}\left(1 + \frac{1}{2a+1}e^{-a\tau}\right)\right]}. \tag{2.108}$$

We include derivations of the constituent expectations for completeness,

$$\begin{aligned} \mathbb{E}[\bar{z}(1-\bar{z})] &= \frac{1}{2}\mathbb{E}[z(0) + z(\tau)] - \frac{1}{2}\mathbb{E}[(z(0) + z(\tau))^2] \\ &= \frac{1}{2} - \frac{1}{2}\left(\frac{a+1}{2(2a+1)} + \frac{1}{4} + e^{-a\tau}\frac{1}{4(2a+1)}\right). \end{aligned} \tag{2.109}$$

And,

$$\begin{aligned} \mathbb{E}[\overline{z(1-z)}] &= \frac{1}{2}\mathbb{E}[z(0)(1-z(0)) + z(\tau)(1-z(\tau))] \\ &= \frac{1}{2} - \frac{a+1}{2(2a+1)}. \end{aligned} \tag{2.110}$$

Using the results from Eq. (2.108), in Fig. 2.10, we reproduce Fig. 2.3 with an x-axis of pairwise $\mathrm{F_{ST}}$ between the focal population and the GWA study population instead of ancient sample in time.

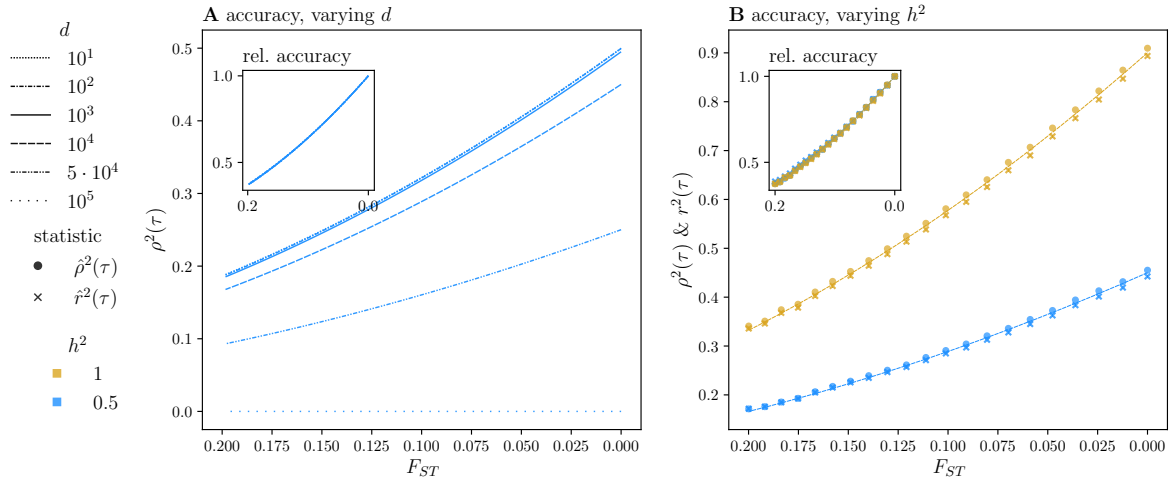


Figure 2.10: **Polygenic score accuracy as a function of F_{ST}.** We reproduce Fig. 2.3 with accuracy and relative accuracy as functions of F_{ST} instead of ancient sampling time τ.

2.8.14 Comparison to the results of Wang et al. 2020

In Section 2.4.4 of the main text, we found that relative accuracy was fairly insensitive to many of the model parameters. This allows us to more readily compare our theoretical results with those of Wang et al. 2020, who also generated predictions for accuracy decay in out-of-sample predictions in humans. In Fig. 2.11, we plot our neutral theory (n = 350,000, d = 1,000, and $a = 10^{-3}$) as a function of pairwise divergence (F_{ST}). Alongside, we plot Wang et al.'s predictions for accuracy reductions in individuals of South Asian (sas), East Asian (eas), and African (afr) ancestries in the UK Biobank relative to a sample of individuals of European (eur) ancestry as a function of observed F_{ST} with eur. In addition, we plot the reductions in accuracy in each ancestry group that were observed in the data set.

In contrast to our theory, Wang et al. take into account the combined effects of (observed) differences in LD and allele frequencies between ancestry groups. Thus, since our theory only accounts for allele frequency changes, it is not surprising that we underestimate the accuracy reductions observed in individuals of African ancestry. Surprisingly, our predictions exceed or approximate those of Wang et al. for sas and eas ancestries—which all underestimate the observed accuracy reductions.

Altogether, these results suggest that accurately predicting out-of-sample accuracy reductions will require more complex modeling of the underlying demographic processes and environmental factors contributing to phenotypic variation. In short, the relationship between F_{ST} and prediction accuracy is not simple.

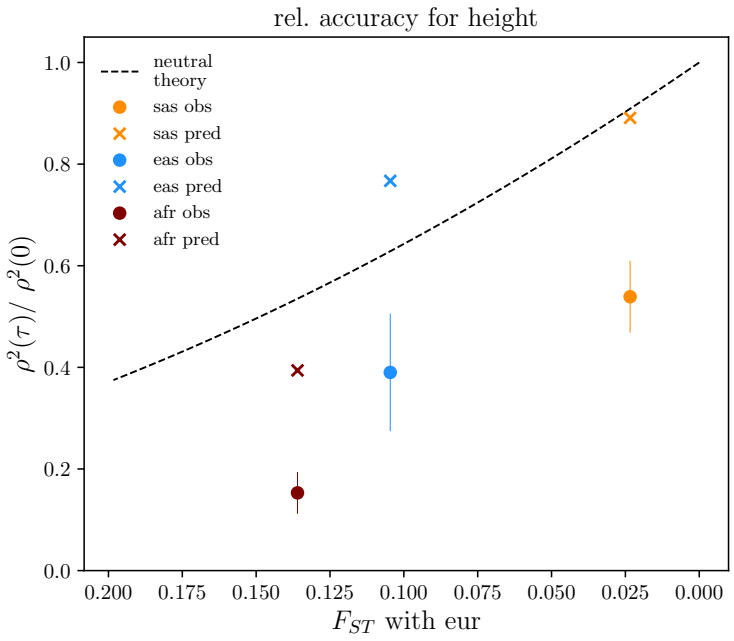


Figure 2.11: **Relative polygenic score accuracy.** We compare our theoretical results (dashed line; $n = 350,000$, $d = 1000$, and $a = 10^{-3}$) for relative accuracy reductions to those of Wang et al. (Wang et al., 2020) for height in individuals of non-European ancestry. Each ancestry group, South Asian (sas), East Asian (eas), and African (afr) is distinguished by color. The x's demarcate Wang et al.'s predictions, while the circles denote the observed accuracy reductions; error bars are 95% confidence intervals. Theoretical values of F_{ST} are computed according to Eq. (2.108); observed values for each ancestry group are from (Wang et al., 2020).

2.8.15 Necessary moments, under neutrality and at stationarity

We provide analytic expressions for the moments which constitute the various metrics under the assumption of equal mutation rates. In addition, we provide simplified expressions when the detection thresholds are equal.

Moments of the population allele frequency and genotype. Because the population is at stationarity, the moments in this subsection are time-invariant. They require integration over the stationary density of the population allele frequency, which is beta-distributed, and in (b) require integration over the Hardy-Weinberg sampling process.

a. *Moments of the population allele frequency:*

$$\mathbb{E}\left[Z_\ell\right] = \tfrac{1}{2},\ \mathbb{E}\left[Z_\ell^2\right] = \tfrac{a+1}{2(2a+1)},\ \text{and}\ \mathbb{E}\left[Z_\ell(1-Z_\ell)\right] = \tfrac{a}{2(2a+1)}.$$

b. *Moments of a genotype:*

$$\mathbb{E}\left[X_{i\ell}\right] = 0 \text{ and thus } \mathbb{V}\left[X_{i\ell}(t)\right] = \mathbb{E}\left[X_{i\ell}^2(t)\right] = \tfrac{a+1}{2a+1}.$$

Moments specific to the GWA study. These moments require integration over the stationary density of the population allele frequency and the sampling probabilities for a sample of n individuals.

a. *Moments of the mean genotype in the GWA study sample:*

$$\mathbb{E}\left[\bar{X}_\ell\right] = 0 \text{ and } \mathbb{E}\left[\bar{X}_\ell^2\right] = \tfrac{1}{n}\left(\tfrac{a+n}{2a+1}\right).$$

b. *Product of the mean genotype in the GWA study sample and the effect estimate:*

$$\begin{aligned}
\mathbb{E}\left[\bar{X}_\ell\hat{\beta}_\ell\right] = \mathbb{E}\left[\mathbb{E}\left[\bar{X}_\ell\hat{\beta}_\ell|\bar{X}_\ell\right]\right] &= \beta_\ell\mathbb{E}\left[\bar{X}_\ell\mathbb{1}_{\{\bar{X}_\ell\in(\gamma-1,1-\gamma)\}}\right] \\
&= \beta_\ell\sum_{i=d_{\ell1}}^{2n-d_{\ell2}}\left(\frac{i-n}{n}\right)\binom{2n}{i}\mathbb{E}\left[Z_\ell^i\left(1-Z_\ell\right)^{2n-i}\right] \\
&= \beta_\ell\sum_{i=d_{\ell1}}^{2n-d_{\ell2}}\left(\frac{i-n}{n}\right)\binom{2n}{i}\frac{B(a+i,b+2n-i)}{B(a,b)}.
\end{aligned} \tag{2.111}$$

And, for $d_{\ell 1} = d_{\ell 2}$,

$$\mathbb{E}\left[\bar{X}_\ell \hat{\beta}_\ell\right] = 0. \tag{2.112}$$

And,

$$\mathbb{E}\left[\bar{X}_\ell^2 \hat{\beta}_\ell\right] = 2\beta_\ell \sum_{i=d_{\ell 1}}^{n-d_{\ell 2}} \left(\frac{i-n}{n}\right)^2 \binom{2n}{i} \frac{B(a+i, a+2n-i)}{B(a,a)} \tag{2.113}$$

For $d_{\ell 1} = d_{\ell 2} = d_\ell$,

$$\begin{aligned} \mathbb{E}\left[\bar{X}_\ell^2 \hat{\beta}_\ell\right] &= \beta_\ell \left(\frac{a+n}{n(2a+1)} - 2 \sum_{i=0}^{d_\ell - 1} \left(\frac{i-n}{n}\right)^2 \binom{2n}{i} \frac{B(a+i, a+2n-i)}{B(a,a)} \right) \\ &= \beta_\ell \left(\frac{a+n}{n(2a+1)} - 2P_1^{(d_\ell)} \right). \end{aligned} \tag{2.114}$$

Under our simple threshold model, the corresponding second moment of $\hat{\beta}_\ell$ is equal to the previous expression multiplied by β_ℓ.

c. *First moment of the mean phenotype in the GWA study sample:* $\mathbb{E}\left[\bar{Y}\right] = 0$.

d. *First moment of the estimated intercept term:*

$$\mathbb{E}\left[\hat{C}\right] = \mathbb{E}\left[\bar{Y}\right] - \sum_{\ell=1}^{L} \mathbb{E}\left[\bar{X}_\ell \hat{\beta}_\ell\right] = C - \sum_{\ell=1}^{L} \beta_\ell \sum_{i=d_{\ell 1}}^{2n-d_{\ell 2}} \left(\frac{i-n}{n}\right) \binom{2n}{i} \frac{B(a+i, a+2n-i)}{B(a,a)}, \tag{2.115}$$

which, for equal detection thresholds equals 0.

Moments involving both the ancient and contemporary genotypes. These moments involve quantities from two time points: the ancient sampling time τ and the GWA study at the present. To compute these moments, we use the spectral representation of the *tdf* (Section 2.8.6).

a. *Product of the first moments of the ancient and contemporary mean genotype:*

$$\begin{aligned}\mathbb{E}\left[X_\ell(\tau)\bar{X}_\ell(0)\right] &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\mathbb{E}\left[X_\ell(\tau)X_{i\ell}(0)|Z_\ell(0),Z_\ell(\tau)\right]\right] \\ &= \mathbb{E}\left[(2Z_\ell-1)(2Z_\ell(\tau)-1)\right] \\ &= \frac{1}{B(a,b)}\sum_{k=0}^{1}\frac{e^{-\lambda_k\tau}}{\langle B_k,B_k\rangle_\pi}\langle 2z-1,B_k\rangle_\pi^2 \\ &= e^{-a\tau}\left(\frac{1}{2a+1}\right).\end{aligned} \tag{2.116}$$

b. *Product of the moments of the ancient and contemporary mean genotypes, and the effect estimate:*

$$\begin{aligned}\mathbb{E}\left[\bar{X}_\ell\hat{\beta}_\ell X_\ell(\tau)\right] &= \sum_{i=d_{\ell 1}}^{2n-d_{\ell 2}}\left(\frac{i-n}{n}\right)\binom{2n}{i}\sum_{k=0}^{1}\frac{e^{-\lambda_k\tau}}{\langle B_k,B_k\rangle_\pi}\langle 2z-1,B_k\rangle_\pi\langle z^i(1-z)^{2n-i},B_k\rangle_\pi \\ &= e^{-a\tau}\beta_\ell\sum_{i=d_\ell}^{2n-d_\ell}\left(\frac{(i-n)^2}{n(a+n)}\right)\binom{2n}{i}\frac{B(a+i,a+2n-i)}{B(a,a)}.\end{aligned} \tag{2.117}$$

For equal detection thresholds, d_ℓ, and using the variables defined in Eq. (2.62), we have,

$$\mathbb{E}\left[\bar{X}_\ell\hat{\beta}_\ell X_\ell(\tau)\right] = \beta_\ell e^{-a\tau}\left(\frac{1}{2a+1}-2P_2^{(d_\ell)}\right). \tag{2.118}$$

The corresponding second moment of $\hat{\beta}_\ell$ is equal to the previous expression multiplied by β_ℓ.

Moments involving the ancient (but not contemporary) genotype. These moments involve the ancient genotype $X_\ell(\tau)$ and the contemporary effect estimate $\hat{\beta}_\ell$, but not the contemporary genotypes.

a. *Product of the first moments of the ancient genotype and the effect estimate:*

$$\begin{aligned}\mathbb{E}\left[X_\ell(\tau)\hat{\beta}_\ell\right] &= \frac{1}{B(a,a)}\sum_{i=d_{\ell 1}}^{2n-d_{\ell 2}}\binom{2n}{i}\sum_{k=0}^{1}\frac{e^{-\lambda_k\tau}}{\langle B_k,B_k\rangle_\pi}\langle z^i(1-z)^{2n-i},B_k\rangle_\pi\langle 2z-1,B_k\rangle_\pi \\ &= e^{-a\tau}\sum_{i=d_{\ell 1}}^{2n-d_{\ell 2}}\binom{2n}{i}\left(\frac{B(a+i,a+2n-i)}{B(a,a)}\right)\left(\frac{i-n}{a+n}\right)=0,\end{aligned} \tag{2.119}$$

for equal thresholds $d_{\ell 1} = d_{\ell 2}$. This result is due to the fact that, in Eq. (2.119), the i-th term is equal to the $(2n-i)$-th term (and the n-th term is 0).

b. *Product of the second moments of the ancient genotype and first moment of the effect estimate:*

$$\begin{aligned}\mathbb{E}\left[X_\ell^2(\tau)\hat{\beta}_\ell\right] &= \frac{\beta_\ell}{B(a,b)}\sum_{i=d}^{2n-d}\binom{2n}{i}\sum_{k=0}^{2}\frac{e^{-\lambda_k\tau}}{\langle B_k,B_k\rangle_\pi}\langle 1-2z+2z^2,B_k\rangle_\pi\langle z^i(1-z)^{2n-i},B_k\rangle_\pi \\ &= \frac{\beta_\ell}{2a+1}\sum_{i=d}^{2n-d}\binom{2n}{i}\frac{B(a+i,a+2n-i)}{B(a,a)} \\ &\qquad\times\left[(a+1)+e^{-(2a+1)\tau}\left(\frac{(2a+1)i(i-2n)+an(2n-1)}{(2a+2n+1)(a+n)}\right)\right].\end{aligned} \tag{2.120}$$

For equal detection thresholds d_ℓ, and Using the terms defined in Eq. (2.62), we can express Eq. (2.120) more succinctly,

$$\mathbb{E}\left[X_\ell^2(\tau)\hat{\beta}_\ell\right] = \frac{\beta_\ell}{2a+1}\left[(a+1)(1-2P^{(d_\ell)})-2e^{-(2a+1)\tau}P_3^{(d_\ell)}\right]. \tag{2.121}$$

The moment $\mathbb{E}\left[X_\ell^2(\tau)\hat{\beta}_\ell^2\right]$ is equal to the previous expression multiplied by β_ℓ.

CHAPTER 3

MITOTIC LOSS OF HETEROZYGOSITY IN *PHYTOPHTHORA CAPSICI*

3.1 Introduction

Mitotic loss of heterozygosity (LOH) refers to the loss of information from one parental chromosome during mitotic growth. Numerous mechanisms can cause LOH, including deletions, mitotic recombination, mitotic gene conversion, and chromosome loss—each producing characteristic genomic signatures (Sui et al., 2020). For example, in *Saccharomyces cerevisiae*, mitotic gene conversion results in short (<15kb) homozygous tracts, whereas mitotic recombination produces homozygous tracts extending from crossover locations to the respective chromosome ends (Sui et al., 2020). Chromosome loss, or aneuploidy, similarly leads to the appearance of homozygosity at all or almost all sites along the chromosome, and is often accompanied by a reduction in sequencing depth. Regardless of the underlying mechanism, mitotic LOH represents a failure of the replication machinery to accurately copy the genome during cell division. It is perhaps not surprising then that LOH occurs frequently in many types of cancer, for which genomic instability is often a defining feature (Cavenee et al., 1983; Beroukhim et al., 2006; Ryland et al., 2015; Steele et al., 2022). However, the observation of LOH across diverse taxa, including chytrids (Rosenblum et al., 2013), yeast (Sui et al., 2020; Ene et al., 2018; Johnson et al., 2021), and oomycetes (Dale et al., 2019; Lamour et al., 2012; Carlson et al., 2017), is less readily explained. Nor can genomic instability explain repeated patterns of LOH within and across cancer types.

Rather, both experimental and theoretical investigations suggest that mitotic LOH may facilitate adaptation in asexually reproducing organisms and cancers (Gerstein et al., 2014; Mandegar and Otto, 2007). The rationale is as follows: When a recessive beneficial mutation is introduced into a diploid, asexually reproducing population, it will exist only as a

heterozygote. Unless another mutation occurs at the same site, the beneficial mutation will exhibit neutral dynamics, and is likely to be lost from the population. However, if mitotic LOH occurs at rates appreciably higher than point mutations, it can substantially increase the probability that the beneficial mutation occurs as a homozygote, and thereby increase the mutation's fixation probability (Mandegar and Otto, 2007). Indeed, mitotic LOH has been shown to facilitate fixation of recessive resistance mutations in heterozygous yeast cells exposed to the fungicide Nystatin (Gerstein et al., 2014), and has been demonstrated to occur early in tumorigenesis at cancer driver genes (Cavenee et al., 1983; Ryland et al., 2015). In addition, the exposure of recessive deleterious mutations via LOH may also facilitate their removal from a population, mitigating the effects of deleterious mutations on fitness.

For sexually reproducing species, mitotic LOH is, of course, not the only source of long homozygous tracts in the genome. Progeny may inherit genomic segments from their parents which share recent ancestry, resulting in runs of homozygosity (ROHs). The frequency and length distributions of these ROHs reflects a population's demographic history. For example, recent consanguinity will produce long ROHs, as large tracts of recent common ancestry will be inherited by offspring without much disruption by recombination (Ringbauer et al., 2021). Similarly, long ROH tracts are more likely to be found in smaller populations due to elevated inbreeding. In species which reproduce both sexually and asexually, inbreeding and mitotic LOH may interact to shape patterns of genomic homozygosity.

To investigate the relative importance of meiotic vs. mitotic processes in determining genomic patterns of homozygosity, we compare the genomic distributions of ROH and mitotic LOH tracts in an oomycete plant pathogen, *Phytophthora capsici*. Genetic investigations suggest that mitotic LOH may be common throughout the genome of *P. capsici* (Lamour et al., 2012; Carlson et al., 2017) and may be characteristic of the genus *Phytophthora* more broadly (Chamnanpunt et al., 2001; Dale et al., 2019). Several aspects of *P. capsici*'s biology make it an ideal organism in which to study ROHs due to both meiotic and mitotic processes.

During an epidemic, *P. capsici* proliferates via mycelial growth and the dispersal of swimming zoospores, each capable of inciting a new infection. Thus, abundant asexual growth provides ample opportunity for somatic mutations. Further, as *P. capsici* is diploid during asexual growth, somatic mutations may include mitotic LOH. In addition, as the germline cells are derived from somatic cell lineages in *P. capsici*, all mutations generated during mitotic cell divisions may be transmitted to offspring generated by sexual reproduction. Frequent sexual reproduction, stimulated by the presence of both mating types, $A1$ and $A2$, produces hardy long-lived oospores, which can survive in the soil for many years, seeding future epidemics (Granke et al., 2012).

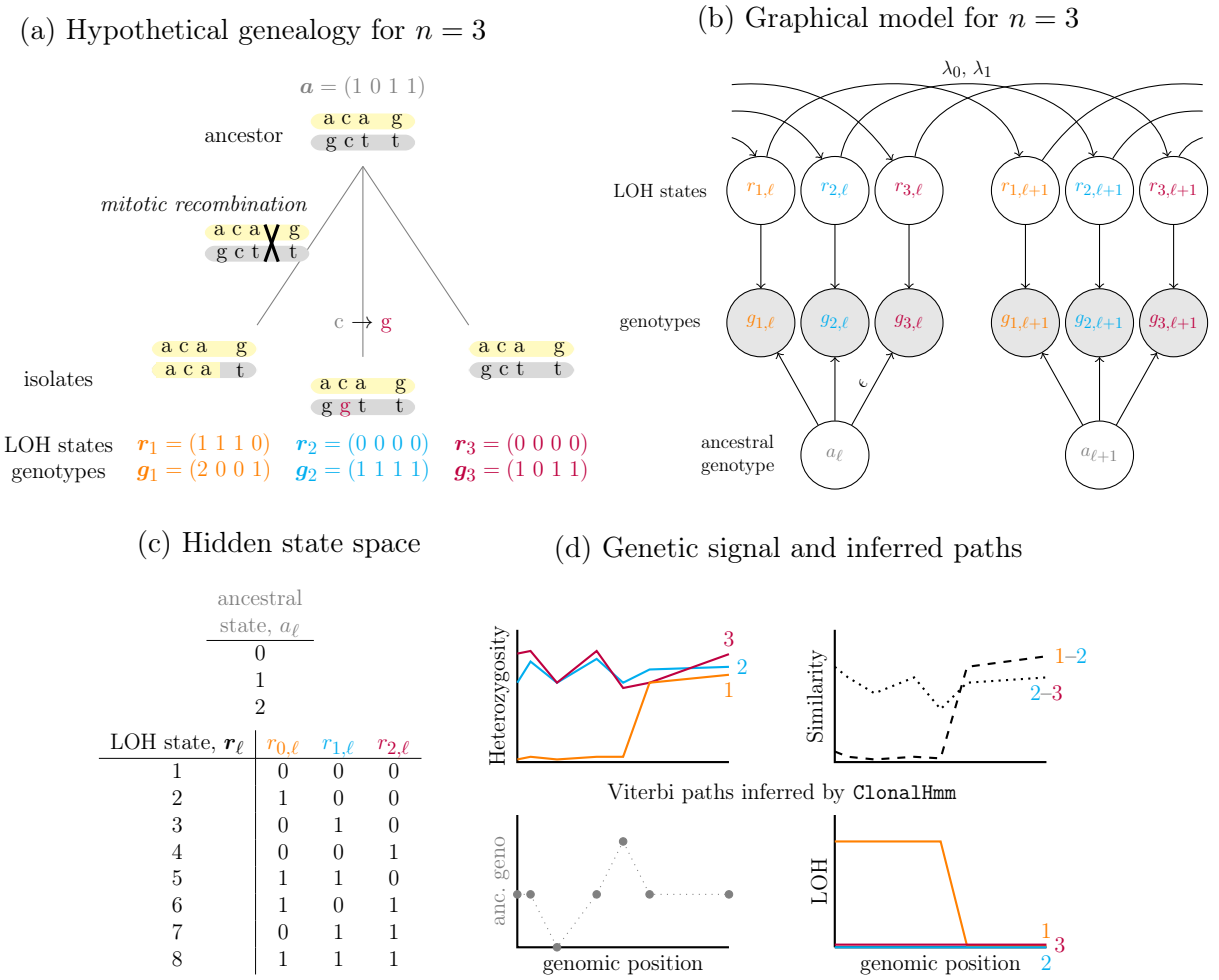
Intriguingly, some of the LOH events reported in *P. capsici* occurred in the genomic regions containing single-nucleotide polymorphisms (SNPs) associated with mating type determination, and were anecdotally linked to mating type switches, predominantly from the $A2$ to $A1$ mating type (Lamour et al., 2012), or mating type discordance within a clonal lineage (Vogel et al., 2021). The directionality of these hypothesized switches is consistent with the current model of mating type inheritance in *P. capsici*, in which the A2 mating type is heterozygous and the A1 mating type is homozygous at the mating type locus (Carlson et al., 2017). In facilitating sexual reproduction, mating type switching would provide an evolutionary advantage where clonal populations are common. While LOH in regions containing mating type associated SNPs is of particular interest, it may be unremarkable in the sense that it is simply a consequence of a genome-wide phenomenon.

To investigate the genomic footprint of mitotic LOH in *P. capsici*, we develop a new inference procedure, `ClonalHmm`, that identifies LOH events simultaneously among members of a single clonal lineage. Our method takes advantage of the fact that genetic variation within a clonal lineage provides information about mutations accrued during mitotic growth to simultaneously infer the ancestral state of the clonal lineage and the LOH states of the isolates within the lineage. This approach, based on a core hidden Markov model (HMM)

framework, shares many features of HMM methods developed to infer runs of homozygosity (ROH) in humans and other species (Narasimhan et al., 2016; Ringbauer et al., 2021) and LOH in cancers (Beroukhim et al., 2006). `ClonalHmm`, however, differs from these methods in jointly analyzing multiple individuals, here isolates within a clonal lineage. Specifically, `ClonalHmm` assumes that the isolates are related by a star-shaped genealogy, with the implication that LOH events occur only on external branches (Fig. 3.1a). Thus, when LOH is modeled as rare, `ClonalHmm` should preferentially identify LOH tracts that have occurred since the common ancestor of the clonal isolates, i.e., where only a subset of the isolates in the clonal lineage exhibit consecutive homozygous genotypes. These tracts will also be marked by elevated genotype mismatches between isolates within the lineage (Fig. 3.1d).

To identify ROHs, we employ an inference procedure that closely follows Narasimhan et al. (2016). In contrast to the LOH tracts identified by `ClonalHmm`, the inferred ROH tracts are distinguished only by homozygosity in excess of Hardy-Weinberg expectations (HWE) with respect to the estimated population allele frequencies (Narasimhan et al., 2016).

We analyzed two publicly available *P. capsici* data sets. The first consists of isolates collected from a field experiment conducted in Geneva, NY over five years, starting in 2009—referred to as the Biparental, or B, population (Dunn et al., 2014; Carlson et al., 2017). This isolated population was founded by inoculating a field of pumpkins with two isolates of opposite mating type, with no subsequent introductions of the pathogen. The data set includes multiple replicates of the parental isolates, including sequences of cultures from multiple time points. Samples from the initial years of the experiment are comprised primarily of F_1, resulting from mating between the parental isolates in the inoculation year, whereas later years also include inbred isolates from subsequent intermating among the progeny (Carlson et al., 2017). The second data set is comprised of 242 isolates collected from 2007-2018 from four regions in New York state, and is referred to as the NY population (Vogel et al., 2021). As it is not possible to know *a priori* whether two infections in the field are caused by the



LOH state, $\boldsymbol{r}_\ell$	$r_{0,\ell}$	$r_{1,\ell}$	$r_{2,\ell}$
1	0	0	0
2	1	0	0
3	0	1	0
4	0	0	1
5	1	1	0
6	1	0	1
7	0	1	1
8	1	1	1

Figure 3.1: **Schematic of the model underlying `ClonalHmm` for n = 3.** (a) `ClonalHmm` assumes that the isolates within a clonal lineage are related by a star-shaped genealogy, implying that mutations only occur on external branches. The yellow and gray bars represent two distinct haplotypes, with letters encoding their allelic states. We represent two possible mutation types: (1) Mitotic recombination in isolate 1, and (2) a point mutation in isolate 2. Mitotic recombination alters the LOH state of individuals 1 (r_1), and results in a depletion of heterozygous genotypes ($g_{1\ell}$) in the affected regions. Where mitotic recombination results in mismatches with the ancestral genotype ($\boldsymbol{a}$) at multiple sites, the effect of the point mutation is localized to a single site. (b) The hidden (ancestral genotype and LOH states) and observed genotypes can be represented in a graphical model. The solid arrows indicate conditional dependencies. While `ClonalHmm` assumes that the ancestral states are mutually independent, one could model correlations between loci. Transitions between states are specified by two rate parameters, λ_0 and λ_1. The error parameter ϵ models both genotyping errors and point mutations. (c) The hidden state space consists of three ancestral states and 2^n combinations of individual LOH states. The LOH states, $\boldsymbol{r}_\ell$, can be mapped to the first 2^n binary numbers. (d) LOH reduces heterozygosity (top left) and increases genotypic mismatches among isolates (top right). `ClonalHmm` uses the Viterbi algorithm to infer the most likely hidden state path, $(\boldsymbol{a}^{(v)}, \boldsymbol{r}^{(v)})$. The "full" hidden state path is readily translated to individual LOH paths $\boldsymbol{r}_i$.

same lineage, both data sets include many clonal lineages (Carlson et al., 2017; Vogel et al., 2021)—ideal for application of `ClonalHmm`. All isolates were sequenced with genotyping-by-sequencing (GBS), a reduced representation sequencing technique (Elshire et al., 2011), and genotyped with respect to a new reference genome consisting of 194 contigs (Shi et al., 2021). We applied `ClonalHmm` and identified ROHs in all identified clonal lineages, allowing us to estimate the contribution of mitotic LOH to patterns of genome-wide homozygosity.

This analysis represents the first large-scale, quantitative investigation of mitotic LOH in *P. capsici*. We find that mitotic LOH affects, on average, approximately 2-3% of an isolate's genome, though LOH spanned more than 20% of the genomes of several isolates.[1] In the B and NY populations, this corresponded to LOH comprising 13 and 9% of all ROHs on average, respectively. These results suggest that while inbreeding is a more potent source of ROHs, mitotic processes make non-negligible contributions to ROHs. In addition, correlations between LOH and ROH incidence across the genome suggest that we may be underestimating the contributions of mitotic LOH to ROHs. Several regions exhibited elevated LOH incidence in both populations, including regions containing mating type associated SNPs, referred to collectively as the MTR. Further inspection revealed that mitotic LOH events in MTR was coincident with mating type discordance within five clonal lineages, including the lineage documented in (Vogel et al., 2021). In all five instances, mitotic LOH events were identified in the $A1$ isolates in the region of peak association.

3.2 A procedure to infer mitotic LOH

We introduce a new HMM framework for inferring mitotic LOH in clonal lineages, referred to as `ClonalHmm`. We provide a brief exposition of the method here, and relegate additional modeling and implementation details to Section 3.6.2.

1. Given that we cannot estimate the number of somatic cell division which separate isolates within a clonal lineage, it is impossible to estimate a rate of mitotic LOH from the analysis performed here.

In `ClonalHmm`, all representatives of the clonal lineage are assumed to be related by a star-shaped genealogy, with the implication that all genotypic variation within the lineage is attributed to private mutations (Fig. 3.1a). The isolates' genotypes provide information about their LOH states and the genotype of the most recent common ancestor (Fig. 3.1a). LOH produces two characteristic genetic signatures in clonal lineages: 1) A local reduction in heterozygosity; and, (2) A local increase in genotype mismatches with the other members of the lineage (Fig. 3.1d). In the absence of LOH and mutation, we expect an isolate's genotype to be identical to that of the ancestor.

We encode individual i's genotype as the number of alternate alleles, $g_{i,\ell} \in \{0, 1, 2\}$, for $i = 1, \ldots, n$ and $\ell = 1, \ldots L$, where n is the number of isolates in the lineage and L is the number of single-nucleotide polymorphisms (SNPs) in the contig or chromosome. We denote the ancestral genotype for site ℓ as $a_\ell \in \{0, 1, 2\}$, and encode the individual LOH states as $\boldsymbol{r}_i \in \{0, 1\}^L$, where $r_{i,\ell} = 1$ indicates that site ℓ is within an LOH tract, and $r_{i,\ell} = 0$, outside of an LOH tract. Restated in our notation, the signatures of LOH are as follows: When $r_{i,\ell} = 0$, we expect that $g_{i,\ell} = a_\ell$ and $g_{i,\ell} = g_{j,\ell}$ for all $j \neq i$ for which $r_{j,\ell} = 0$. In the presence of LOH, $r_{i,\ell} = 1$, we similarly expect that $g_{i,\ell} = a_\ell$ if the ancestral genotype is homozygous, i.e., when $a_\ell \in \{0, 2\}$. However, when $r_{i,\ell} = 1$ and $a_\ell = 1$, we expect that isolate i will be homozygous for either allele, i.e., $g_{i,\ell} \in \{0, 2\}$, and $g_{i,\ell} \neq g_{j,\ell}$ for all $j \neq i$ for which $r_{j,\ell} = 0$ (Fig. 3.1). As LOH may generate either homozygote, isolates may have a genotype mismatch at site ℓ even when both have undergone LOH. Deviations from these expectations are modeled by an error parameter ϵ, which in effect models both genotyping errors and point mutations (Section 3.6.2).

The conditional dependencies among the individual genotypes, LOH states, and ancestral states can be represented in a graphical model (Fig. 3.1b). The LOH states are modeled as independent and identically distributed (*iid*) Markov chains, with transitions between LOH states at adjacent sites governed by an underlying Poisson process with rate parameters λ_1

and λ_0 (Eq. (3.2)). These two rates define the expected lengths of LOH and non-LOH tracts, $1/\lambda_1$ and $1/\lambda_0$, respectively, with $\lambda_0 \ll \lambda_1$ encoding the assumption that LOH is relatively rare. The isolates' genotypes at site ℓ are mutually independent when conditioned on the ancestral states. `ClonalHmm` assumes that the ancestral states are independent from one another. In reality, the a_ℓ may be correlated due to systematic variation in heterozygosity across the genome, for example, due to ROHs in the most recent common ancestor.

As `ClonalHmm` jointly infers the LOH states of all n members of a clonal lineage, the method must keep track of the n-length vector of LOH states at each SNP, $\boldsymbol{r}_\ell \in \{0,1\}^n$ (Fig. 3.1c). The hidden state space then consists of all 2^n possible combinations of individual LOH states and the three possible ancestral genotypes, for a total size of 3×2^n.

`ClonalHmm` uses the Viterbi algorithm (Rabiner, 1989) to infer the most likely hidden state path, consisting of the LOH states of each isolate and the ancestral state of the lineage at each genomic position (Figs. 3.1c and 3.1d). The former is readily collapsed to yield the individual LOH states at each position (Fig. 3.1d, bottom right). In addition, `ClonalHmm` can use the forward-backward procedure (Rabiner, 1989) to compute the marginal posterior probabilities of the hidden states. The marginal posteriors can then be thresholded to yield inferred LOH and ancestral states in a process referred to as posterior thresholding (Section 3.6.2). While both of these algorithms scale exponentially with n, several implementation details allow `ClonalHmm` to be applied to modest lineage sizes ($n \leq 10$).

3.3 Results

3.3.1 Validating the mitotic LOH inference

In order to evaluate the accuracy of `clonalHmm`, we analyzed simulated copy-number neutral mitotic LOH events in clonal lineages consisting of two isolates—the most common lineage size in our data sets (Fig. 3.9). We systematically varied the size of the LOH tracts and

the genotyping error rate. In each simulated replicate, we sampled an isolate from the B population to serve as the ancestral genotype. We conducted 100 simulations for each set of parameters in each of two of the largest contigs (contigs 104 and 141; Section 3.6.6). As results for different simulated error rates were similar, we show only the simulations where $\epsilon = 0.05$, similar to what was estimated from the data.

We assessed the accuracy of `ClonalHmm` for combinations of transition rate parameters (λ_0 and λ_1) spanning several orders of magnitude, and for several realistic genotyping error probabilities for GBS data (Sections 3.6.6 and 3.6.6, and see Carlson et al. (2017)). Accuracy was quantified as the true positive rate, defined as the proportion of simulations in which the inferred LOH tract spanned at least 70% of the simulated tract, and the false positive rate, the proportion of a contig spanned by spuriously inferred LOH tracts, and averaged over the two contigs (Section 3.6.6).

From our simulations, we found that parameter combinations where $\lambda_1 \ll \lambda_0$ had similar true positive rates across different tract lengths and genotyping error rates (Fig. 3.2). In general, and in this regime, smaller values of λ_0 tended to yield higher false positive rates (Figs. 3.2 and 3.10). Differences in the accuracy of the ancestral inference between parameter combinations were less pronounced, where accuracy was defined as the proportion of correctly inferred ancestral genotype states (Fig. 3.3). In all cases, ancestral inference with `ClonalHmm` achieved high accuracy ($\sim 95\%$) and outperformed a majority-rule procedure described in Section 3.6.6.

`ClonalHmm` more readily detected large LOH tracts (Fig. 3.2). This was not surprising given that larger tracts usually contain more SNPs, providing more evidence in favor of LOH. While this general trend has been observed for other HMM-inference procedures (e.g., Ringbauer et al. (2021)), in *P. capsici*, differences in accuracy with respect to tract length may have been further exaggerated by variability in intermarker distances across and within contigs (Fig. 3.11). Indeed, false negative and positive rates were not uniformly distributed

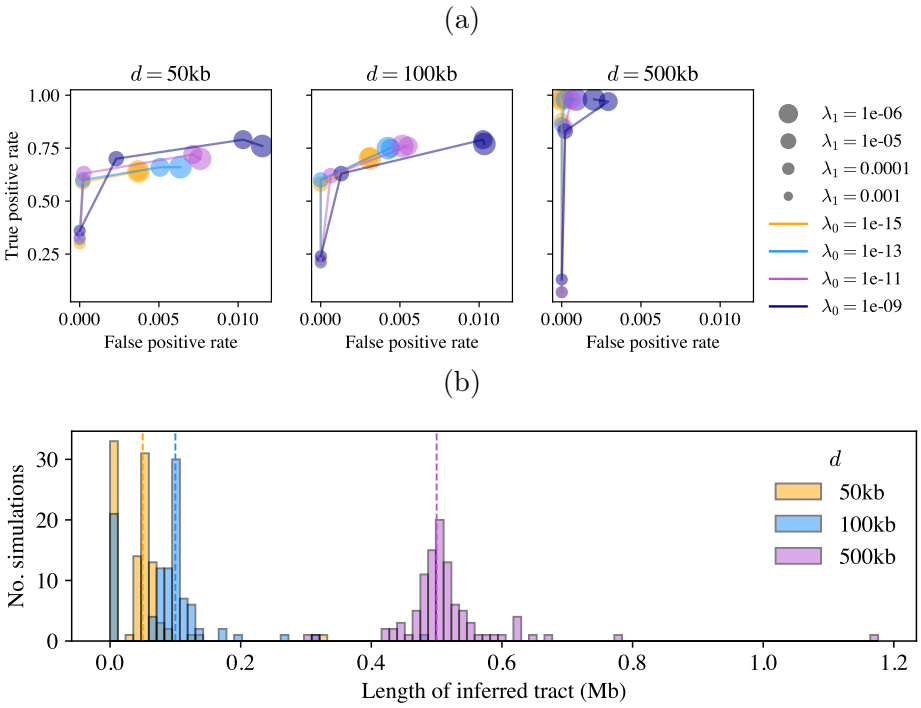


Figure 3.2: **Evaluating the accuracy of `ClonalHmm`.** In (a), we show the accuracy of `ClonalHmm` for for three simulated tract lengths, $d \in \{50\text{kb}, 100\text{kb}, 500\text{kb}\}$, and various inference parameter combinations (color and point size). The simulated error rate is $\epsilon = 0.05$. Accuracy is measured as the false positve rate (x-axis), the genomic length of spuriously inferred LOH tracts normalized by the total contig length, and the true positive rate, the proportion of simulations in which the inferred tract spanned at least 70% of the true LOH tract, averaging over the two simulated contigs (104 and 141) in the Biparental population. Each set of colored points corresponds to a fixed λ_0 value; the size of the points indicates the λ_1 value. In (b), we show the distribution of inferred tracts for each of the three simulated d (colors, vertical dotted lines) for the pair of transition rate parameters ($\lambda_0 = 10^{-13}$, $\lambda_1 = 10^{-5}$) and error rate ($\epsilon = 0.05$) used in our data analysis. Note that the true tract length varied due to variable inter-marker distances in the *P. capsici* genome.

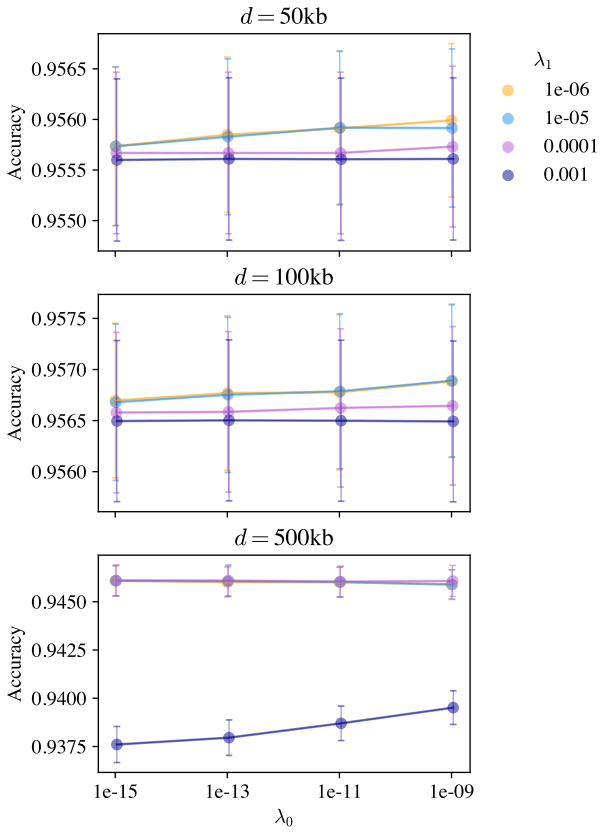


Figure 3.3: **Evaluating the accuracy of ancestral state inference with `clonalHmm`.** The accuracy of ancestral state inference with `ClonalHmm` was defined as the proportion of correctly inferred ancestral genotypes. Each plot (a)-(c) shows the accuracy accuracy for different combinations of transition rate parameters, with λ_0 on the x-axis, and λ_1 indicated by color. The error bars indicate 95% confidence intervals. The simulated error rate was $\epsilon = 0.05$.

across the contig.

To illustrate this, we conducted another set of simulations in which three ancestral genotypes of low, intermediate, and high heterozygosity were selected from each population for several simulated contigs (Section 3.6.6). More false positives occurred in regions of low ancestral heterozygosity (Fig. 3.10), particularly in scenarios where λ_1 was closer in order to λ_0. In addition, more false negatives occurred when the simulated LOH tract fell within a region of low SNP density (Fig. 3.11). In these cases, `clonalHmm` often identified spurious LOH tracts in both members of the clonal lineage or the false positive tract fell within the bounds of a true LOH tract in the other clone. While we cannot address potential false negatives in the data analysis, we do apply a post-processing procedure to exclude putative false positives Section 3.6.4.

We thus selected $\lambda_0 = 10^{-13}$ and $\lambda_1 = 10^{-5}$ for data analysis, and matched the inference error rate to the error rate estimated from the data, letting $\epsilon = 0.05$ (Section 3.6.6).

3.3.2 Application of ClonalHmm to two GBS data sets

We applied `ClonalHmm` to clonal lineages of *P. capsici* in two GBS data sets, where genotypes were called with respect to the SD33 reference genome Shi et al. (2021). Genotypes were inferred with respect to the SD33 genome sequence which consists of 194 contigs spanning 100.5Mb Shi et al. (2021). Clonal lineages were identified using a clustering algorithm described in Section 3.6.1, and ranged in size from one to sixteen isolates (Fig. 3.9). We analyzed all lineages of size $n \geq 2$ and downsampled the larger lineages to a maximum size of eight as violations of the star-shape genealogy assumption become more likely with large n (see Section 3.4). The first data set contained 35 clonal lineages ($n = 111$ isolates) sampled from an experimental population founded by two parents of opposite mating type (Carlson et al., 2017). As there were no subsequent introductions of *P. capsici* to the field experiment, all isolates are descendants of the two founding parents, and are either F_1 or the result of

inbreeding among F_1 and subsequent generations (Carlson et al., 2017). We refer to this data set as the Biparental, or B, population. The second data set consisted of 48 clonal lineages ($n = 161$ isolates) sampled from four regions in New York state (Vogel et al., 2021). We refer to this data set as the New York, or NY, population.

All isolates were either actively maintained via serial transfer, taken out of storage and cultured prior to DNA extraction and sequencing, or both (Carlson et al., 2017; Vogel et al., 2021). Therefore, somatic mutations may have occurred during growth in the laboratory after isolation from the field. Where the same isolate was sequenced at several time points, as is the case for the two parents of the B population, we could unequivocally attribute LOH to mutations during culturing (also see Carlson et al. (2017)). For all other isolates, it was impossible to know whether the identified mitotic LOH events occurred in the field or during passaging in the lab.

As smaller contigs would likely downwardly bias the length distribution of inferred LOH tracts, we excluded contigs smaller than 300kb from the analysis (Fig. 3.8). We also excluded contigs containing fewer than 100 SNPs, as simulations showed that `ClonalHmm` has less power in regions of low SNP density. These filters resulted in 60 and 74 contigs (33,196 and 69,821 SNPs) per data set, spanning approximately 65 and 74Mb, in B and NY, respectively (Fig. 3.8). Lower SNP density in the B population is explained by the fact that, excluding new mutations, all genetic variation is derived from the two founding parents. The NY population, in contrast, is of more diverse provenance.

We conducted several post-processing procedures to address known shortcomings of HMM inference procedures (Ringbauer et al., 2021). Specifically, we merged nearby LOH tracts separated by distances of less than 30kb, and excluded small tracts of less than 20kb (Section 3.6.4). The merging distance was selected based on the distribution of spuriously inferred gaps in simulations, which tended to be smaller than 50kb (Fig. 3.22). In addition, the simulations revealed a problem specific to our application: When ancestral heterozygosity is

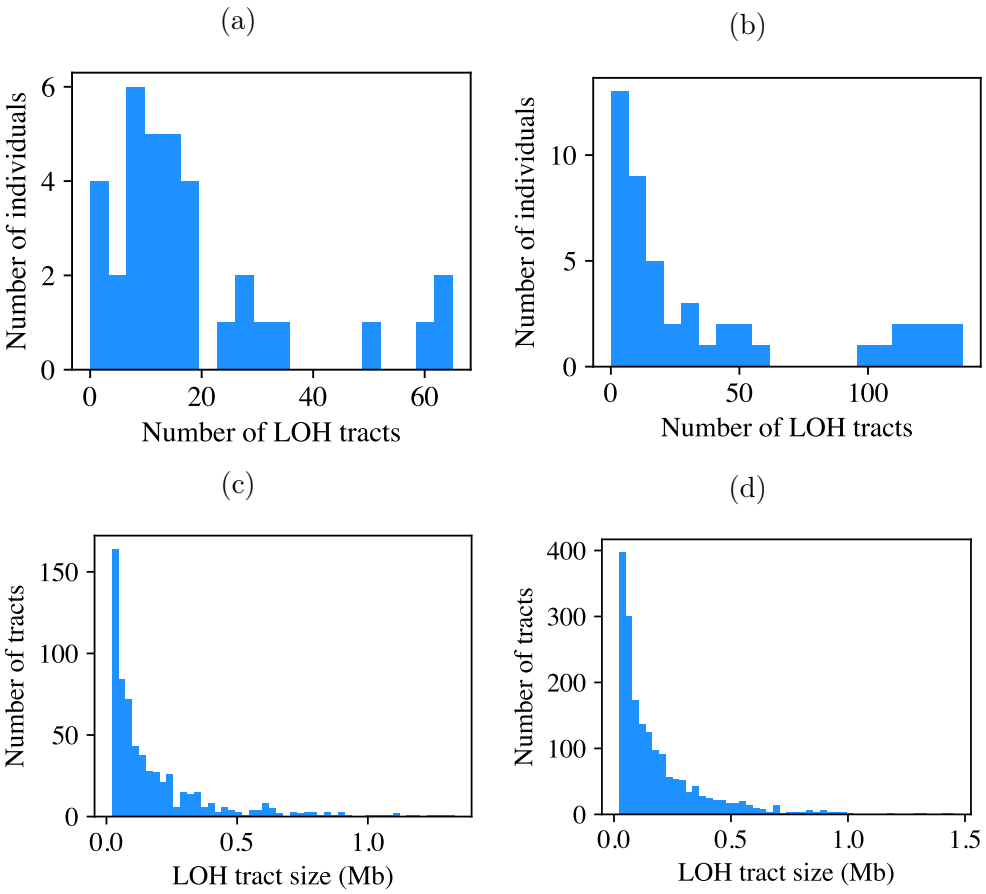


Figure 3.4: **LOH tract incidence and length distributions.** `ClonalHmm` identified fewer LOH tracts per isolate in the (a) Biparental versus (b) New York population, though tract length distributions were similar in both populations (c) and (d), respectively.

low, `ClonalHmm` is more likely to generate false positives, identifying LOH in all members of the clonal lineage (Fig. 3.10). While the ancestral state is unknown in the data analysis—it is in part, the object of our inference—we assumed that regions of low heterozygosity in all members of a clonal lineage are more likely attributable to a homozygous tract in the ancestor, than multiple mitotic LOH events in the clonal isolates. Based on this reasoning, and observation of this phenomenon in our simulations, we corrected for putative false positives by excluding LOH tracts identified in all members of a clonal lineage (Sections 3.3.1 and 3.6.4).

3.3.3 Summaries of mitotic LOH incidence

After post-processing, `ClonalHmm` identified on average 18.3 and 37.9 regions of LOH per isolate (Figs. 3.4a and 3.4b), affecting 1.8 and 3.0% of the genome (1.2 and 2.2Mb) in the B and NY populations, respectively. The maximum inferred LOH was 23.5% and 46.3% of the genome (15.2 and 34.0Mb). While more tracts were identified in NY, the average tract lengths (188kb and 189kb, respectively) and distributions were similar. The tract length distributions appeared to be approximately exponential, with a large number of small tracts (Figs. 3.4c and 3.4d), consistent with our modeling assumptions.[2] Despite the exclusion of small contigs (<300kb), the inferred tract lengths were further constrained by the distribution of contig sizes (Fig. 3.21). As would be expected under a model of uniform LOH incidence across the genome, the number of tracts identified in each contig was correlated with contig size (Fig. 3.23).

To estimate the contribution of mitotic LOH to genome-wide ROH, we identified ROH tracts in each isolate using an ROH-caller similar to (Narasimhan et al., 2016) and described in detail in Section 3.6.3. We used the same rate and error parameters ($\lambda_0 = 10^{-13}$,

2. Note that due to the imposition of a minimum tract size threshold (Section 3.6.4), we artificially reduced the number of observed small tracts. With no minimum size threshold the average numbers of tracts per isolate are 24.7 and 50.7, with average tract sizes of 141 and 144kb in the B and NY populations, respectively. Thresholding had minimal influence on the total amount of genome in LOH.

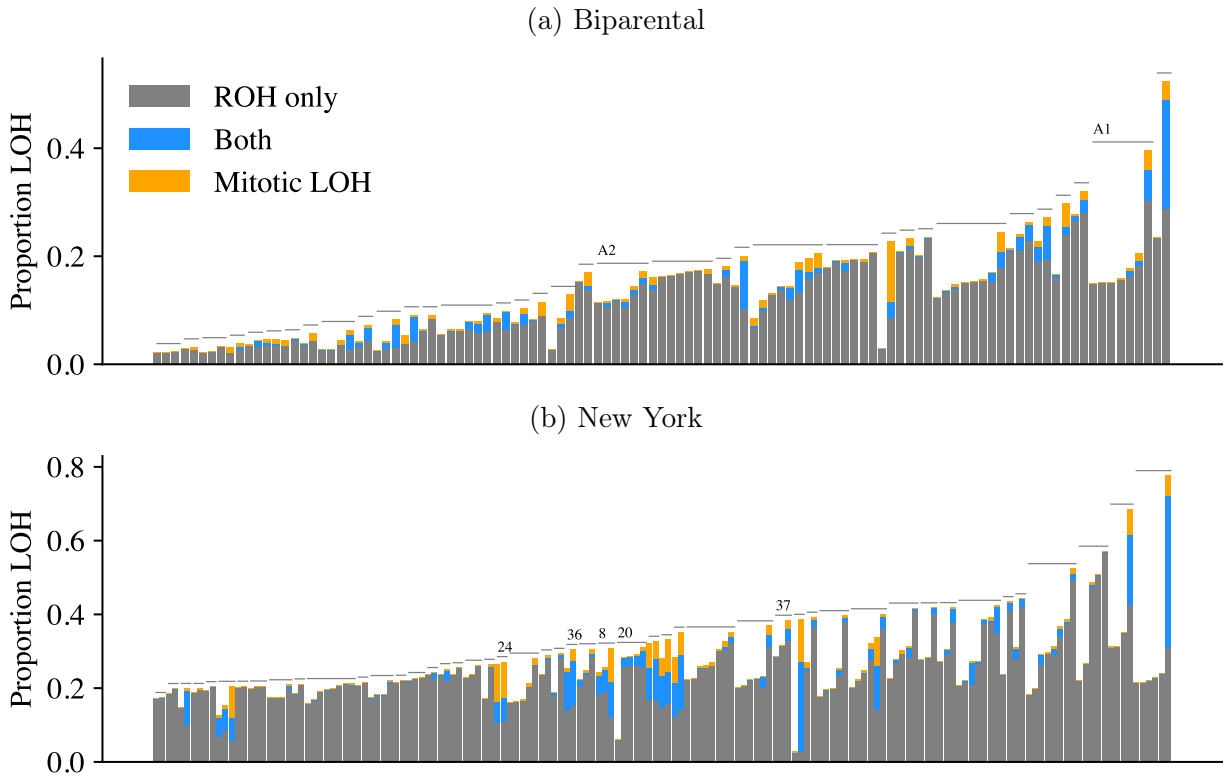


Figure 3.5: **Contribution of mitotic LOH to genome-wide ROH.** Each bar represents an isolate in the (a) Biparental or (b) New York populations analyzed by both `ClonalHmm` and the ROH-caller. Isolates are organized by lineage (denoted by the horizontal black line) and by the total amount of genome inferred to be in ROHs and LOHs. The partitioning of the bar denotes the total length of ROHs identified by both methods (blue) and by each method alone (orange and gray, respectively). In (a), the clonal lineages corresponding to the founding parents are labeled with their mating types, i.e., $A1$ corresponds to the $A1$ parent.

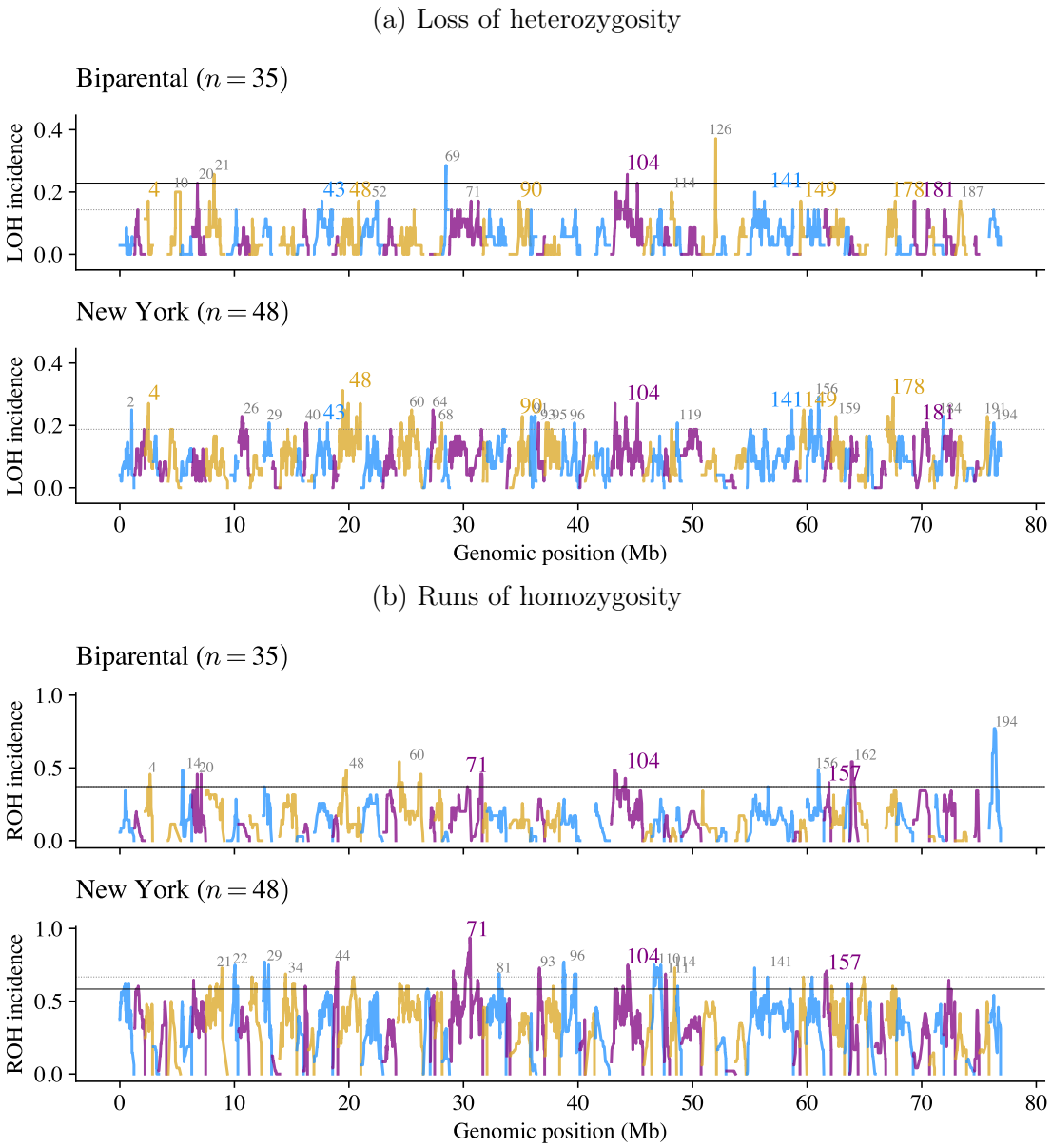


Figure 3.6: **Genome-wide incidence of mitotic LOH and ROH.** Each panel shows the proportion of clonal lineages in which an LOH (a) or ROH (b) was observed at given SNP in at least one member of the lineage, for the Biparental (B, top) and New York (NY, bottom) populations. The SNPs are ordered by contig and genomic position (x-axis), with color alternating with contig. The black solid line, where plotted, indicate the FDR significance thresholds ($\alpha = 0.05$). The dotted line indicates the 95$^{\text{th}}$ percentile, calculated in each population separately. Contigs containing a SNP in the 95$^{\text{th}}$ percentile were labeled in gray when in one population and in color and larger font when in both.

$\lambda_1 = 10^{-5}$, and $\epsilon = 0.05$) so as not to artificially inflate the difference between methods, even though different parameters may more accurately infer ROH. As anticipated, the ROH-caller identified more tracts per individual than `ClonalHmm`—24.8 and 75.8, in B and NY (Fig. 3.27)—with longer average tract lengths, 329 and 245kb, respectively (and see Fig. 3.28). While some tracts were uniquely identified by `ClonalHmm`, many were also identified by the ROH-caller as expected (Fig. 3.5).

On average, 12.9% and 9.8% of genome-wide ROH could be attributed to mitotic LOH in B and NY, respectively (Fig. 3.5). While this result implies that genome-wide patterns of homozygosity at the population scale are dominated by meiotic population genetic processes, such as inbreeding, mitotic LOH may still be a major driver of variation within clonal lineages (Figs. 3.5 and 3.29). In addition, estimates of LOH incidence from `ClonalHmm` are likely conservative for several reasons discussed in Section 3.3.4. 62.1% and 52.7% of the total variance in cumulative LOH tract length (summed over contigs) across individuals was attributed to within lineage differences, in B and NY, respectively. In contrast, 7.2% and 35.25% of variance in cumulative ROH tract length (after excluding mitotic LOH tracts) across individuals was attributed to within lineage variation.[3] In Section 3.7.1, we discuss the implications of these results for population genetic analyses, including the estimation of inbreeding coefficients in *P. capsici*.

In addition, our estimates of the contribution of mitotic LOH to genome-wide ROH are likely conservative as `ClonalHMM` can only detect LOH events that have occurred since the most recent common ancestor of the lineage. ROHs, in contrast, are inferred based on homozygosity in excess of HWE expectations at multiple consecutive sites, and thus represent the cumulative effect of both inbreeding and mitotic LOH. In Section 3.4, we speculate on how these two processes interact to produce genomic patterns of homozygosity.

LOH results from copy number neutral mechanisms, e.g., mitotic recombination or gene

3. Not correcting for LOH these numbers were 17.8% and 63.2%, respectively.

conversion, as well as processes that alter copy number, e.g., chromosome loss. To provide preliminary insights into the mechanism of LOH, we compared read depth in and outside of LOH tracts. Sites within LOH tracts tended to have lower read depth than those outside of LOH tracts in both populations (Fig. 3.24). In addition, and not independent of the prior result, the proportion of missing sites was frequently higher in LOH tracts (Fig. 3.25). Higher rates of missing genotypes could arise due to structural variation between haplotypes, e.g., indels, independent of chromosome loss. Alternatively, higher missingness rates in LOH tracts could be a consequence of lower copy number, resulting in fewer average sequencing reads per site and thus a higher probability of missingness. Therefore, these results tentatively suggest that the predominant mechanism(s) underlying LOH in *P. capsici* may alter copy number.[4]However, we caution that several technical factors may have contributed to this result. For one, lower read depth systematically leads to fewer heterozygote genotype calls, a trend observed in both data sets (Fig. 3.16). Systematic heterozygote undercalling would increase the likelihood of an LOH tract due to coincident spurious homozygous genotype calls. If this were the case, we would expect to observe the same relationship between sequencing coverage and inferred ROH tracts, as well as, a correlation between total individual LOH and average sequencing coverage. Yet, differences in average depth and missingness in ROH versus non-ROH tracts were smaller than their respective quantities for LOH (Figs. 3.24 and 3.25), and mitotic LOH correlated only weakly with sequencing coverage (Fig. 3.17). Though more work is needed, these results suggest that differences in read depth and missingness in and outside of LOH tracts may reflect the underlying biological process rather than technical artifacts.

4. The fact that reductions in coverage are less than fifty percent in most isolates also requires explanation. This could be due to the fact that we have averaged over LOH tracts in each isolate. If LOH was due to both copy-neutral and non-copy-neutral processes, this would result in a smaller than fifty percent reduction in coverage. Alternatively, if LOH preferentially occurred in regions with highly divergent haplotypes, with haplotypes distinguished by numerous indels, loss of one haplotype may lead to smaller reductions in coverage than expected under chromosome loss.

3.3.4 Testing for site-specific enrichment

Though mitotic LOH incidence varied across the genome in both the B and NY populations, we only found evidence for site-specific enrichment of mitotic LOH in the B population when LOH incidence across lineages was modeled as binomially distributed with probability given by genome-wide average incidence (Fig. 3.6 and see Section 3.6.5 for methods).

The negative result in the NY population provides some evidence that mitotic LOH may be occurring relatively uniformly across the genome of *P. capsici*. On the other hand, our study may be underpowered to detect significant deviations from the null model for several reasons, enumerated in Section 3.4.

Furthermore, the fact that multiple contigs contained SNPs with elevated LOH incidence—top 95^{th} percentile—in both populations, suggests that shared genomic features may lead to similar LOH outcomes in disparate populations. In addition, LOH incidence at the 24,835 shared SNPs was significantly correlated across populations (Spearman's $\rho = 0.21$, p-value $< 10^{-16}$). ROH incidence was also correlated across populations (shared SNP set, $\rho = 0.43$, p-value $< 10^{-16}$), as well as with LOH incidence in a clone-corrected subset of each population ($\rho = 0.23, .07$, p-values $< 10^{-16}$, respectively).

Mating type and LOH

To identify mating type associated SNPs, we conducted a Fisher's exact test of allele frequency differences between mating types independently in each population (Fig. 3.31, and see Section 3.6.5). After correcting for multiple testing, we identified signficantly associated SNPs spanning several contigs in B and NY, respectively. The localization of mating type determination to several contigs is consistent with previous results (Carlson et al., 2017; Vogel et al., 2021), and may indicate the presence of structural variation or long repeats. We refer to the region defined by the union of significantly associated SNPs in both populations as the mating type region (MTR).

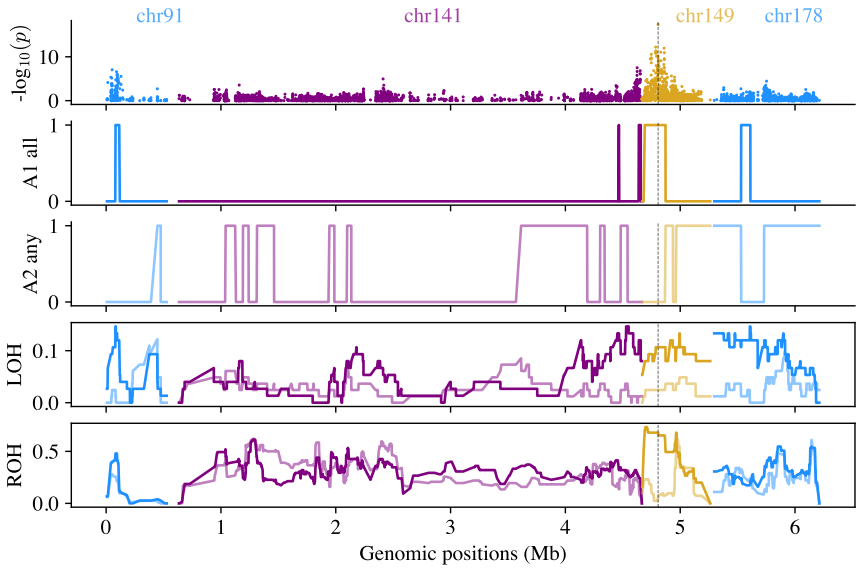


Figure 3.7: **Isolating the mating type region.** In the first row, we show the -$\log_{10}$ transformed p-values from the Fisher's exact test of allele frequencies for the contigs which contained LOH events in all of the $A1$ isolates constituting lineages with mating type discordance. In the second row, we plot the locations of these consensus $A1$ LOH tracts. In the third row, we plot the locations of all LOH tracts observed in any of the $A2$ isolates which belonged to the lineages containing isolates of discordant mating types. In the fourth and fifth rows, we plot the LOH and ROH incidence for $A1$ isolate (dark) and $A2$ isolates (light) in this region, respectively. The dotted vertical line denotes the physical location of the peak SNP (located in contig 149).

While mating type-associated SNPs were not significantly enriched for mitotic LOH (Section 3.3.4), they showed elevated LOH incidence in both populations (see contigs 141 and 149 in Fig. 3.6). Intriguingly, four of the five instances of reported mating type discordance within a clonal lineage—where isolates within a lineage were classified as different mating types—exhibit mitotic LOH in the MTR. In the second row of Fig. 3.7, we show the LOH tracts shared by all $A1$ isolates in these four lineages across the entire genome. These consensus tracts correspond to the peaks in the association analysis (in contigs 91, 141, and 149), and include an additional region in contig 178 (first row, Fig. 3.7). In addition, none of the $A2$ isolates in these lineages exhibit LOH tracts in these regions (third row, Fig. 3.7).

We suspect that the one exception is due to the fact that four of the five isolates are $A1$ in this clonal lineage. The star-shaped genealogy assumption coupled with the assumption that LOH is rare, implies that an inference of four out of five isolates exhibiting LOH is disfavored relative to the ancestor being homozygous. Indeed, these four $A1$ isolates exhibit a depression of heterozygosity in the MTR (bottom lineage in Fig. 3.26).

Furthermore, LOH incidence across all $A1$ lineages is higher in regions of peak association compared to $A2$ lineages (fourth row, Fig. 3.26). The same trend is observed in a more localized region of contig 149 for ROH (fifth row, Fig. 3.26).

Our result of LOH in the MTR coinciding with mating type discordance within a lineage adds to observation of a single instance of this phenomenon in (Vogel et al., 2021). The less stringent clone-correction threshold used here allowed us to identify the additional four instances. (See Section 3.4 for a discussion of the influence of the clone-correction threshold on the results of `ClonalHmm`.) In all instances, the $A1$ isolates within the lineage exhibited mitotic LOH spanning a sub-region of the MTR (Fig. 3.26), consistent with the hypothesis that the $A2$ mating type is conferred by heterozygosity in the mating type region, and $A1$ by homozygosity (Carlson et al., 2017).

3.4 Discussion

We developed an inference procedure, `ClonalHmm`, to identify mitotic LOH events among members of a single clonal lineage. Our method builds upon core hidden Markov model machinery of previous ROH inference procedures (Ringbauer et al., 2021; Narasimhan et al., 2016), but differs from these methods in two key ways: (1) `ClonalHmm` infers LOH jointly among several isolates within a clonal lineage ($2 \leq n \leq 10$); and (2) `ClonalHmm` simultaneously infers the individual LOH states and the genotype of the isolates' most recent common ancestor under the assumption of a star-shaped genealogy. Where ROH inference procedures model ROHs as deviations from Hardy-Weinberg expectations, `ClonalHMM` models deviations from the hidden ancestral genotype. The ancestral genotype, in turn, is inferred by the method. `ClonalHmm` identifies ROHs which are more likely attributable to mitotic processes, i.e., mitotic LOH, as members of a clonal lineage are differentiated only by mitotic cell divisions.

We used this method to infer mitotic LOH events in two large GBS data sets of *P. capsici*: (1) Samples from an isolated, experimental population founded by two parental isolates of opposite mating type, deemed the B population (Carlson et al., 2017), and (2) an assemblage of isolates collected primarily from New York state, the NY population. Altogether, we analyzed 35 clonal lineages in the B population and 48 clonal lineages in the NY population, corresponding to 111 and 174 isolates, respectively, representing the first large-scale, quantitative analysis of LOH in *P. capsici*.

We found that LOH is relatively common, affecting approximately 2-3% of the genome on average. Some isolates exhibited no LOH, while LOH affected more more than 10% of the genome of others. In general, the ROH incidence exceeded that of LOH, with LOH comprising on average 12.95% and 9.77% of ROHs in the B and NY populations, respectively.

While previous studies (Vogel et al., 2021; Dale et al., 2019) suggested that LOH may arise via a copy number neutral mechanism, we find tentative evidence that LOH may alter copy

number, as LOH tracts in many individuals had significantly more missing data and lower read depth than non-LOH tracts. At the same time, we observed only a weak correlation between total inferred LOH and average sequencing coverage per individual. Alternatively, LOH may preferentially affect regions with high haplotype diversity, e.g., with numerous indels or rearrangements, so the loss of a haplotype could manifest in higher missingness and lower read coverage.

We found evidence of significant enrichment for LOH in several genomic regions in the B population, but not in the NY population. There are several reasons why our study may have been underpowered to detect enrichment for LOH. First, in conducting the enrichment test based on lineages rather than isolates, we substantially reduced the sample size. This was a conservative choice to avoid double counting LOH tracts which were inferred in multiple members of a clonal lineage. While LOH may have occurred independently in several isolates within a lineage, coincident inference of LOH within a lineage is more likely due to violations of the star-shaped genealogy assumption. Second, `ClonalHmm` was only able to detect mitotic LOH events that have occurred since the most recent common ancestor of the clonal lineage. Such LOH events will have occurred during growth in the field, prior to sampling, or during growth in the laboratory post-isolation and prior to sequencing. In some cases, both of these time windows may have been small, providing little opportunity for mitotic LOH. Third, the test statistics of nearby markers are highly correlated, thereby exaggerating the multiple testing burden. Fourth, LOH is a relatively rare phenomenon, with a skewed distribution of incidence across individuals (Fig. 3.4). This implies that there were limited opportunities for observing LOH at any given site. Finally, the detection of LOH is limited by ancestral heterozygosity and ROHs; variation among members of a clonal lineage is a prerequisite for detection of mitotic LOH. Thus, elevated ROH in a region constrains the detection of mitotic LOH. Correlation in LOH incidence within and across populations suggests that genomic features may predispose certain regions to LOH. In addition, correlations between LOH and

ROH incidence in both populations suggests that LOH may drive ROH patterns in certain genomic regions.

On the other hand, `ClonalHmm` may be anti-conservative when isolates are spuriously grouped as a clonal lineage. For example, an isolate may be grouped with progeny resulting from self-fertilization, particularly if the isolate had low heterozygosity. To see this, we consider the IBS between an isolate with n_{het} and n_{hom} genotypes. The selfed progeny will have the same genotype as its progenitor at all homozygous sites. At all heterozygous sites, it will have a 50% probability of inheriting the same genotype, and otherwise will be 50% IBS at heterozygous sites. Together, these two facts imply that IBS between the progenitor and selfed progeny will by $.75 \times n_{\text{het}} + n_{\text{hom}}/(n_{\text{het}} + n_{\text{hom}})$. While the threshold used in the identification of clonal lineages was calibrated with respect to known lineages, and did not erroneously group three known selfed progeny with the progenityor lineage, we cannot discount that some of the identified clonal lineages may rather be due to selfing or inbreeding between close relatives.

In addition, `ClonalHmm` suggests a procedure to account for the presence of clonal isolates in data sets of asexually reproducing organisms. Prior to conducting population genetic inference, it is commonplace to perform "clone-correction" in which only a single representative of a clonal lineage is retained for further analysis (e.g., as in Carlson et al. (2017); Vogel et al. (2021) and here Section 3.6.1). Alternatively, one may simply use a majority rule to identify the ancestral genotype. The rationale for clone-correction in *P. capsici* is twofold: (1) When sampling procedures are not methodical with respect to space, it is impossible to know whether the representation of a clonal lineage in a data set reflects its true frequency in the population or biased sampling; and (2) the interpretation of many population genetic analyses relies on the assumption of random mating. As *P. capsici* does not, for the most part, self-fertilize, a clone-corrected data set better approximates a randomly mating population.

Clone-correction is further justified when there is little variation between members of a clonal lineage. As we have shown, this may not always be a good assumption—mitotic LOH can lead to appreciable differences between members of a clonal lineage. Furthermore, mitotic LOH events may have occurred in the laboratory, not in the field. Therefore, failure to account for LOH where possible may unduly render population genetic analyses less informative about pathogen dynamics in the field. In addition, clone-correction, where only a single representative of the lineage is retained, does not fully utilize the genetic information at hand.

As illustrated in our conceptualization of `ClonalHmm` (Fig. 3.1a), representatives of a clonal lineage contain information about their most recent common ancestor, which necessarily lived in the field. Thus, we propose the following procedure: (1) Identify clonal lineages using a clustering procedure, as in Section 3.6.1. (2) Apply `ClonalHmm` to each clonal lineage to infer the posterior probabilities on the ancestral states of the clonal lineage at each SNP. (3) In all population genetic analyses, propagate uncertainty in the genotypic state of a clonal lineage by integrating over its genotype with respect to the posterior probability distribution. Alternatively, one could utilize the Viterbi path for the ancestral states in downstream analyses. We have shown that the Viterbi path achieves high accuracy and outperforms a common-sense majority rule procedure.

We remark that methodology aside, clone-correction necessarily provides a biased picture of the population. For example, if one clonal lineage is extremely successful and spreads throughout much of a farmer's field, it is likely to sexually reproduce more frequently than less successful lineages. Ideally, one could estimate the relative frequencies of distinct clonal lineages from sampling data, and estimate population genetic statistics accordingly.

3.5 Author contributions

We provide author contributions according to Contributor Roles Taxonomy, as in Chapter 2.

Maryn O. Carlson

Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing - original draft, Writing - review & editing

Adam Fine

Methodology, Software, Data visualization, Writing - review & editing

Howard Judelson

Investigation, Writing - review & editing

Matthias Steinrücken

Conceptualization, Funding acquisition, Investigation, Methodology, Supervision, Writing - review & editing

3.6 Supplementary materials and methods

3.6.1 Data sets

We analyzed two publicly available *P. capsici* data sets, genotyped with GBS, a reduced representation sequencing technique (Elshire et al., 2011). The first, the Biparental, or B, population, consists of the raw sequencing reads of 232 isolates sampled from a biparental field population from 2009-2013, and multiple replicates of the founding parents (Dunn et al., 2014; Carlson et al., 2017). The field isolates include putative F_1, and putative inbred isolates from matings that occurred post-inoculation (Carlson et al., 2017). In addition, the data set includes 46 progeny from an *in vitro* cross between the same two parents. The second data set, referred to as the New York, or NY, population, consists of raw GBS sequencing reads for isolates sampled from four New York regions (Western New York, Central New York, Capital District, and Long Island) from 2007-2018 (Vogel et al., 2021).

Our study makes use of two subsets of both of data sets: (1) A clone-corrected data set, which consists of a randomly sampled isolate from each identified clonal lineage (see

Section 3.6.1) and, (2) The set of all isolates which belong to the identified clonal lineages.

All isolates were maintained via active culturing or in storage, or both, and cultured in V8 broth prior to sequencing with GBS (Carlson et al., 2017; Vogel et al., 2021). Thus, passaging in the lab may have introduced somatic mutations, including LOH.

Genotyping

Genotypes were called jointly using the Genome Analysis Toolkit (GATK4) software (Poplin et al., 2018), with SD33 (Shi et al., 2021) as the reference sequence. The 194 contig, 100.5Mb SD33 genome, constructed with both long- and short-read sequencing, improves upon earlier sequencing efforts (Shi et al., 2021). Briefly, after aligning the sequencing reads for each sample to the reference genome (Shi et al., 2021) with `bwa mem` (Li and Durbin, 2009) and conducting base recalibration with GATK4, we used `HaplotypeCaller` and `CombineGVCFs` to jointly call genotypes.

Preliminary filtering. All preliminary filtering was conducted with VCFtools (Danecek et al., 2011). Isolates with more than 20% missing data were removed from the analysis altogether. We subsequently filtered the B and NY isolates separately, removing indels, and retaining only biallelic sites with allele frequencies $0.05 < p_\ell < .95$ and less than 20% missing data in each data set.

Additional filtering. We excluded all contigs containing fewer than 100 SNPs or smaller than 300kb to minimize the splitting of inferred LOH tracts over multiple contigs (Fig. 3.8). In addition, we excluded all sites with fewer than three observed minor allele homozygotes regardless of allele frequency and sample size—for example, if $p_a < 0.5$ then we filter on the frequency of aa homozygotes—in the clone-corrected data sets (see (Carlson et al., 2017) for clone-correction methods). We applied the latter in lieu of standard filtering on deviations from HWE as the unique population structure of the B population immediately implies deviations from HWE. Similarly, the NY isolates likely do not constitute a randomly mating

population. Observed excess heterozygosity can be due to errors in sequencing alignment, for example, if a duplication in the focal isolate is not represented in the reference genome, reads from multiple positions would align to the same locus in the reference genome. The preponderance of this phenomenon in our data is supported by the fact that common alleles with fewer than three observed minor allele homozygotes exhibited consistently higher average read depth (Fig. 3.30).

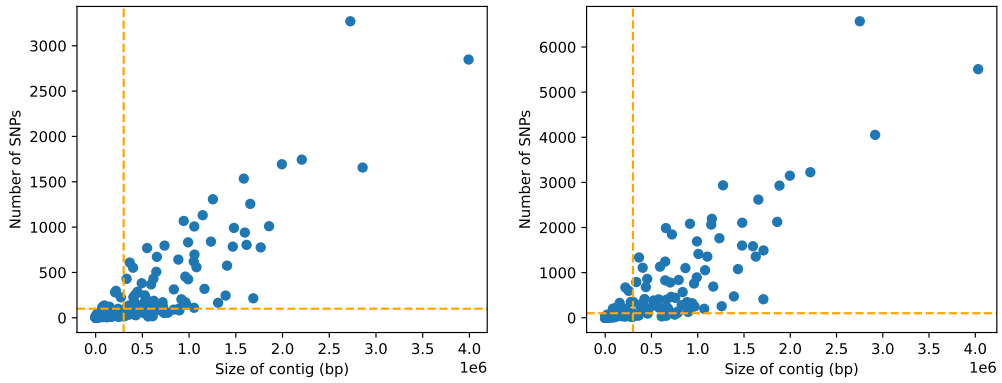
These filters resulted in 63 contigs and 33,196 SNPs spanning approximately 65Mb in the B population, and 74 contigs and 69,821 SNPs spanning approximately 74Mb in the NY population. Lower diversity in the B population is consistent with the fact that, excluding mutation, all genetic diversity was derived from two founding parents (see in particular figures 1A of (Carlson et al., 2017)).

Identification of clonal lineages

We used a procedure based on hierarchical clustering to identify clonal groups. We first computed a relationship matrix where each entry F_{ij} is the proportion of alleles shared identity-by-state (IBS) between isolates i and j,

$$F_{ij} = \frac{1}{2} \sum_{\ell=1}^{L_g} |g_{i\ell} - g_{j\ell}|. \tag{3.1}$$

This relationship matrix implies a distance matrix $D = I \odot F - F$, where I is the identity matrix and $\odot$ is the element-wise product. Hierarchical clustering was performed with `scipy.cluster.hierarchy` using D and the *average* distance method. The resulting tree was cut at a height of 0.13. Incrementing by 0.01, this was the largest height that correctly grouped all representatives of the parental lineages, which include both technical and biological replicates (Carlson et al., 2017), without erroneously including additional isolates.



(a) Biparental (b) New York

Figure 3.8: **Contig sizes and SNP density.** For each data set, we plotted the relationship between contig size and the number of SNPs identified in the contig. The dotted orange lines represent the size and SNP cutoffs of 300kb and 100 SNPs, respectively.

(a) New York (b) Biparental

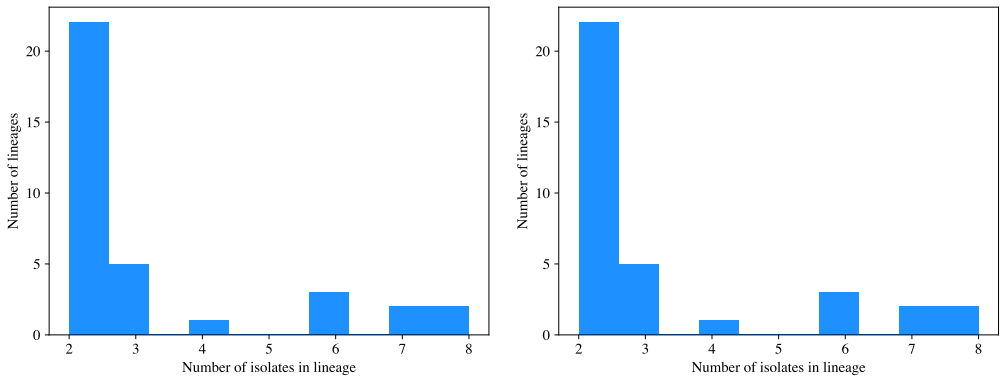


Figure 3.9: **Distribution of lineage sizes**. Histogram of the number of isolates per lineage in the (a) New York and (b) Biparental data sets. Lineages consisting of more than eight isolates were downsampled. The lineage sizes post-downsampling are shown here.

Where the same culture was sequenced twice, we randomly retained one sequencing replicate. Additionally, to obtain a clone-corrected data set, we retained all "unique" isolates and sampled one member of each clonal lineage, resulting in 207 and 118 unique isolates in the B and NY populations, respectively. The 35 and 48 clonal lineages identified in B and NY, respectively, were analyzed by `ClonalHmm`.

3.6.2 A method for inferring mitotic LOH, `ClonalHmm`

`ClonalHmm` simultaneously infers (1) the LOH states at each genotyped site, $\boldsymbol{r}_i \in \{0,1\}^L$, $i = 1, \ldots, n$, for n members of a clonal lineage, and (2) the ancestral genotype, $\boldsymbol{a} \in \{0,1,2\}^L$ of the lineage, where L is the number of SNPs in the contig or chromosome. The isolates' genotypes at all L sites within a contig, $\boldsymbol{g}_i \in \{0,1,2\}^L$, are the observed states, and are modeled as emissions from their LOH states and the ancestral genotype of the clonal lineage (Fig. 3.1b). For each isolate, LOH states are modeled as independent Markov chains with exponential transition rates that increase with the distance between adjacent SNPs. `ClonalHmm` assumes that the isolates are related by a star-shaped genealogy, implying the graphical model of Fig. 3.1b. To fully specify the HMM, we describe the probability distributions governing the emissions and transitions between hidden states.

Transition probabilities. Modeling LOH as a Poisson process along the genome, the transition probabilities between two SNPs ℓ and $\ell + 1$ for LOH state are given by,

$$\mathbb{P}\{r_{i,\ell+1} = s | r_{i\ell} = r, \Theta\} = \begin{cases} (e^{-\lambda_0 d_\ell})^{1-s}(1 - e^{-\lambda_0 d_\ell})^s, & r = 0, \\ (e^{-\lambda_1 d_\ell})^s(1 - e^{-\lambda_1 d_\ell})^{1-s}, & r = 1, \end{cases} \tag{3.2}$$

where λ_0^{-1} and λ_1^{-1} are the expected tract lengths for non-LOH and LOH tracts, respectively; d_ℓ is the recombination or physical distance between sites ℓ and $\ell + 1$, and we have assumed that at most a single transition can occur between sites. In other words, when an isolate is

in a non-LOH state, the "waiting" distance until an LOH event is exponentially distributed with parameter λ_0. Similarly, when an isolate is in an LOH state, the waiting distance to jump out of the LOH state is exponentially distributed with parameter λ_1.

The full transition probability considers the LOH states of all n members of the clonal lineage, i.e., transitions between the n-length vectors of LOH states $\boldsymbol{r}_\ell$ and $\boldsymbol{r}_{\ell+1}$ at the ℓ- and $(\ell+1)$-th sites, respectively. A naive implementation would perform 2^{2n} computations to find the transition probabilities for a fixed d_ℓ and all possible LOH state combinations. However, transitions between some states are equivalent, reducing the number of unique computations. The transition probability between two states is fully specified by the number of transition types, e.g., from $r_{i\ell} = 0$ to $r_{i,\ell+1} = 0$. Thus,

$$\mathbb{P}\{\boldsymbol{r}_{\ell+1} = \boldsymbol{s} | \boldsymbol{r}_\ell = \boldsymbol{r}, \Theta\} = e^{-\lambda_0 d_\ell n_{00}^{(rs)}}(1 - e^{-\lambda_0 d_\ell})^{n_{01}^{(rs)}} e^{-\lambda_1 d_\ell n_{11}^{(rs)}}(1 - e^{-\lambda_1 d_\ell})^{n_{10}^{(rs)}}, \quad (3.3)$$

where $n_{jk}^{(rs)}$, $j, k = 0, 1$, indicate the number of transitions from LOH state j to k in the transition from states $\boldsymbol{r} \in \{0,1\}^n$ to $\boldsymbol{s} \in \{0,1\}^n$.

We assume the ancestral genotypes are independent of each other with prior probabilities given by a contig-wide estimate of the heterozygote frequency within the lineage, i.e., $\hat{\pi}_{Aa} = \frac{1}{L}\sum_{\ell=1}^{L}\sum_{i=1}^{n} \mathbb{1}_{\{g_{i\ell}=1\}}$, with L equal to the number of SNPs in the contig, and $\pi_{AA} = \pi_{aa} = (1-\pi_{Aa})/2$. Thus, the probability of transitioning to a particular ancestral genotype is simply given by its frequency. We assume equal priors on the homozygous states so that `ClonalHmm` is agnostic to the (often arbitrary) allelic encoding. The assumption of independence cannot completely capture the fact that heterozygosity may vary systematically across the genome of the ancestor, due to, for example, ROH in the ancestral genome. However, accounting for the frequencies of the ancestral genotypes in the prior should at least partially mitigate this modeling inaccuracy.

Initial probability. We let π_1 and $1 - \pi_1$ be the initial probabilities of ROH and non-

ROH, respectively. Instead of freely specifying π_1, we assume its value at stationarity, $\pi_1 = \lambda_1^{-1}/(\lambda_0^{-1} + \lambda_1^{-1})$, where λ_0 and λ_1 are the parameters governing the transitions between LOH and non-LOH states introduced above (Eq. (3.2)).

Emission probabilities. The probability of observing the genotype of isolate i at site ℓ, $b_{i\ell}(\cdot)$, depends on the hidden ancestral and the individual's LOH state. We use the error model of (Ringbauer et al., 2021), in which the the true genotype is observed with probability $1 - \epsilon$; with probability ϵ either of the incorrect genotypes is observed with equal probability,

$$b_{i\ell}(g; r, a) \coloneqq \mathbb{P}\{g_{i\ell} = g | r_{i\ell} = r, a_\ell = a, \Theta\} = \begin{cases} g = 0 & \text{w.p. } 1-\epsilon, & a = 0, r \in \{0,1\}, \\ g = 1,2 & \text{w.p. } \epsilon/2, & a = 0, r \in \{0,1\}, \\ g = 2 & \text{w.p. } 1-\epsilon, & a = 2, r \in \{0,1\}, \\ g = 0,1 & \text{w.p. } \epsilon/2, & a = 2, r \in \{0,1\}, \\ g = 0,2 & \text{w.p. } 1/2 - \epsilon/4, & a = 1, r = 1, \\ g = 1 & \text{w.p. } \epsilon/2, & a = 1, r = 1, \\ g = 1 & \text{w.p. } 1-\epsilon, & a = 1, r = 0, \\ g = 0,2 & \text{w.p. } \epsilon/2, & a = 1, r = 0, \end{cases} \tag{3.4}$$

where Θ represents the model parameters, including the genotyping error rate ϵ. While we introduced ϵ as an "error" probability, in practice, ϵ reflects both genotyping errors and point mutations. When $g_{i\ell}$ is missing, we let $b_{i\ell}(\cdot) = 1$.

Identification of LOH. We infer the most likely hidden state path using the Viterbi algorithm (Rabiner, 1989). The Viterbi path is first computed with respect to the full hidden state space, which consists of three ancestral states and 2^n possible LOH states at each locus (Figs. 3.1c and 3.1d), and then collapsed to yield paths for the LOH states of each isolate i, referred to as $\boldsymbol{r}_i^{(v)} \in \{0,1\}^L$ (Fig. 3.1d). While the Viterbi algorithm is appealing for

its simplicity, it only provides a point estimate of the hidden state path. Alternatively, one can estimate and threshold the marginal posterior probabilities of LOH at each site computed using the forward-backward algorithm (Rabiner, 1989), which is also implemented in `ClonalHmm`. These two algorithm scale exponentially with the number of hidden states, $\mathcal{O}(2^n L)$ and $\mathcal{O}(2^{2n} L)$, respectively, limiting application of `ClonalHmm` to modest lineage sizes, $n \leq 10$.

3.6.3 A method for inferring runs of homozygosity

The ROH inference procedure employed here uses deviations from HWE to identify ROH, closely following (Narasimhan et al., 2016). The method assumes that the genotype of an isolate at locus ℓ is sampled from a population with allele frequencies $p_\ell \in [0, 1]$, computed in a reference sample and assumed to be known without error. (Here, the alternate allele simply refers to the "1" allele, and makes no claims whether it is the ancestral or derived state.) We reuse notation and let $\boldsymbol{r}_i \in \{0, 1\}^L$, be the ROH-state of the isolate, where $r_{i\ell} = 1$ now indicates that site ℓ is contained within an ROH, and $r_{i\ell} = 0$, its opposite.

Initial probabilities. The initial probabilities are the same as in `ClonalHmm` (Section 3.2).

Transition probabilities. The transition probabilities are the same as those which govern an individual's LOH states in `ClonalHmm` (Eq. (3.2)). The transition probabilities of Narasimhan et al. can be interpreted as approximations of Eq. (3.2). When $\lambda_0 d_\ell$ and $\lambda_1 d_\ell$ are small—a condition satisfied when genotyping is dense—differences between are negligible.

Emission probabilities. The emissions probabilities, $b_{i\ell}(\cdot)$ depend on the ROH state of the individual and the population allele frequencies. To incorporate genotyping error, we again

follow (Ringbauer et al., 2021),

$$b_\ell^{(p)}(g; r, p) \coloneqq \mathbb{P}\{g_{i\ell} = g | r_{i\ell} = r, p_\ell = p, \Theta\} \tag{3.5}$$

$$= \begin{cases} g = 0 \text{ w.p. } (1-p)^2(1-\frac{3\epsilon}{2}) + \frac{\epsilon}{2}, & r = 0, \\ g = 1 \text{ w.p. } 2p(1-p)(1-\frac{3\epsilon}{2}) + \frac{\epsilon}{2}, & r = 0, \\ g = 2 \text{ w.p. } p^2(1-\frac{3\epsilon}{2}) + \frac{\epsilon}{2}, & r = 0, \\ g = 0 \text{ w.p. } (1-p)(1-\epsilon) + p\frac{\epsilon}{2}, & r = 1, \\ g = 1 \text{ w.p. } \frac{\epsilon}{2}, & r = 1, \\ g = 2 \text{ w.p. } p(1-\epsilon) + (1-p)\frac{\epsilon}{2}, & r = 1. \end{cases} \tag{3.6}$$

In short, we assume HWE when the individual is not in an ROH state. When the individual is in an ROH state, an allele is "copied" from one of the ancestral chromosomes (pertaining to each parent), and thus the emissions probability is given by the allele's population frequency p_ℓ. These probabilities are modified by the error rate: With probability $1-\epsilon$, the genotype is observed without error; With probability ϵ, the genotype is observed as either of the two other genotype states, each with equal probability. If $g_{i\ell}$ is missing, we let $b_\cdot(\cdot) = 1$.

For $p_\ell = a_\ell/2$, Eq. (3.4) is equivalent to the emissions probabilities of the ROH-caller, which are instead functions of the population allele frequencies Eq. (3.5).

3.6.4 Post-processing of LOH tracts

`ClonalHmm` outputs the per-site LOH state for each individual. Here, we describe how to translate the marker-specific LOH states to genomic coordinates. In particular, one must make a decision about where to define the boundary between LOH and non-LOH states. In addition, we describe two post-processing procedures motivated by our simulation results and preliminary data analysis: (1) merging of LOH tracts, and (2) removal of putative false

positive LOH tracts.

Delimiting tract boundaries. The start of the tract is defined by the midpoint between the genomic positions where $r_{i,\ell}^{(v)} = 1$ and $r_{i,\ell-1}^{(v)} = 0$; whereas the end of the tract is defined by the midpoint of the positions where $r_{i,\ell'}^{(v)} = 1$ and $r_{i,\ell'+1}^{(v)} = 0$, and $r_{i,m}^{(v)} = 1$ for all $\ell \leq m \leq \ell'$. When the tract encompasses the first SNP in a contig, we let the boundary of the tract be the maximum of the position of the first marker minus approximately half the average inter-marker distance (1kb) and zero; if the tract contains the last SNP, we let the boundary be the position of the last marker plus 1kb.

Merging tracts. We often observed consecutive, but not contiguous, LOH (or ROH) tracts in our data analysis and simulations (Fig. 3.22), as documented in the application of similar methods previously (Ringbauer et al., 2021). Spurious gaps have been previously attributed to genotyping or assembly errors, structural variation, or low SNP density (Ringbauer et al., 2021). Indeed, in simulations, we observed an elevated false negative rate in regions of low SNP density (Fig. 3.11). Thus, as in (Ringbauer et al., 2021), we follow a similar procedure to (Ralph and Coop, 2013). We merge any two LOH tracts separated by a gap less than 20kb, based on the falsely inferred inter-tract distances observed in simulations.

Identifying putative false positives. Our simulations revealed that `clonalHmm` sometimes identified spurious LOH tracts in regions of low ancestral heterozygosity (Fig. 3.10, and see Section 3.3.1). While this phenomenon was rare for the parameters used in the analysis (Fig. 3.2), for which $\lambda_1 \ll \lambda_0$, we took the conservative approach and excluded any inferred tracts which were identified in all members of a clonal lineage under the assumption that such tracts were more likely attributable to an ROH in the ancestor. This supposition was supported by the fact that the putative false positive tracts identified in the data analysis similarly exhibited a depression in heterozygosity (Fig. 3.10).

Putative false positive tracts were only found in the New York population (Fig. 3.14). In addition, we found that such tracts were identified far more frequently in lineages of size

$n = 2$ (Fig. 3.14), where the difference between the probabilities of LOH versus no-LOH in all members of the lineage is minimized, and never in lineages of $n > 3$ (Fig. 3.14).

A naive approach to remove these putative false positives would simply identify strings of n's in a sequence equal to the sum of the LOH calls of all isolates, i.e., $\sum_{i=1}^{n} \boldsymbol{r}_i^{(v)0}$. However, this procedure would erroneously excise the putative false positive tract in a scenario where the tract resided in a larger LOH tract found in a subset of the lineage. To remove false positive tracts while accounting for this scenario, we use the following iterative procedure: Let the i-th iteration consider the i-th member of the clonal lineage, where the order is arbitrary. We identify the tracts in clone i using the procedure described above. For each identified tract j, we compute the coverage of this tract (with respect to genomic coordinates) in all other members of the clonal lineage. If the coverage in all $n-1$ clones exceeds a threshold of 75%, then the tract is discarded as a putative false positive. Otherwise, if the tract exceeds some predetermined size threshold, here 20kb, it is retained as an inferred LOH tract.

3.6.5 Site-specific analyses

To test for enrichment of LOH at particular sites, we conduct independent analyses in each population at the level of clonal lineage. To do so, we aggregated LOH calls across members of a clonal lineage, to generate a lineage incidence vector, $\boldsymbol{v}_j$, with elements,

$$v_{j,\ell} = \mathbb{1}\Big\{ \sum_{i \in \text{lin} j} r_{i,\ell}^{(v)} > 0 \Big\}, \tag{3.7}$$

where $\mathbb{1}\{\cdot\}$ is the indicator function, the sum is over all members of clonal lineage j, and $\ell = 1, \ldots, L_g$, where L_g is the total number of SNPs in the genome. Thus, $v_{j,\ell} = 1$ if at least one member of a clonal lineage exhibits LOH at site ℓ.

We then followed a procedure similar procedure to (Sui et al., 2020). Under a null model in which LOH is uniformly distributed across sites and lineages, the marginal probability of

SNP ℓ being in an LOH tract in a given individual is equal to the genome-wide probability of LOH $p_0 \coloneqq \mathbb{P}_0\{r_{i\ell} = 1\}$. The LOH incidence at a given site across lineages is binomially distributed, i.e., $\sum_{i=1}^{N_{lin}} V_{i,\ell} \sim \text{Bin}(N_{\text{lin}}, p_0)$. We use the following estimator of this probability,

$$\hat{p}_0 \coloneqq \frac{1}{N_{\text{lin}} L_g} \sum_{\ell=1}^{L_g} \sum_{i=1}^{N_{\text{lin}}} v_{i,\ell}. \tag{3.8}$$

To assess for significant per-site LOH enrichment, we compute p-values under this Binomial null model and correct for multiple testing according to (Benjamini and Hochberg, 1995).

Association study to identify the mating type region

After clone-correction (Section 3.6.1), a Fisher's exact test of allele frequency differences between mating types was conducted separately in each of the two data sets at each segregating site. Association tests were performed previously in both data sets, using LT1534 (Lamour et al., 2012) as the reference sequence (Carlson et al., 2017; Vogel et al., 2021). Multiple testing correction was performed following Benjamini and Hochberg (1995).

3.6.6 Simulations

We used a simulation-based approach to assess the accuracy of `ClonalHmm` in its application to the B and NY data sets. We conducted two sets of simulations. The first set of simulations Section 3.6.6 were conducted with the aims of validating the inference procedure and the inference parameters used in the application of `ClonalHmm` to GBS data. In the second Section 3.6.6, we more systematically studied the consequences of heterozygosity on accuracy.

Validating transition rate parameters

We simulated LOH tracts of three sizes, $d \in \{50\text{kb}, 100\text{kb}, 500\text{kb}\}$, in two large contigs (contigs 104 and 141). In each simulation replicate, we sampled an ancestral genotype

uniformly from the B population. We then imputed missing sites in the ancestral genotype with genotypes sampled according to their observed frequencies in order to preserve the proportion of heterozygous sites.

To simulate LOH, the $\boldsymbol{r}_i$ hidden state vectors, we "inserted" an LOH segment of a pre-specified length and with a uniformly sampled start position in one member of the clonal lineage, and required that the simulated tract contain at least five SNPs and be contained within the contig.

We then simulated the genotypes of the $n = 2$ isolates conditional on the ancestral sequence and each isolate's LOH state, with a genotyping error rate of $\epsilon \in \{.01, .05, .1\}$ (see Eq. (3.4)), and a missingness rate of 0.05 per site. These parameters approximated levels of discordance and missingness observed in the data sets (see Section 3.6.6 and Sections 3.6.6 and 3.6.6). For each combination contig, tract length, and error rate, we conducted $K = 100$ simulations.

We analyzed these simulations with all pairwise combinations of rate parameters, $\lambda_0 \in \{10^{-15}, 10^{-13}, 10^{-11}, 10^{-9}\}$ and $\lambda_1 \in \{10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}\}$, and the same three ϵ values used in the simulations. Thus each simulation was analyzed by $4 \times 4 \times 3$ parameter combinations.

We considered a true positive to be any case where the inferred tract covered at least 70% of the simulated tract's length, where tract boundaries were delimited according to Section 3.6.4. We similarly delimited the boundaries of false positives.

We measured the accuracy of the ancestral state inference by the statistic

$$\rho_{\text{anc}}^{(v)} = \frac{1}{L} \sum_{\ell=1}^{L} \mathbb{1}_{\{a_\ell^{(v)} = a_\ell\}}, \tag{3.9}$$

where $a_\ell^{(v)}$ is the ancestral state inferred by the Viterbi algorithm and a_ℓ is the true ancestral state. We compared this accuracy to that of the posterior probabilities on the an-

cestral states computed using the forward-backward procedure (Rabiner, 1989), $\rho_{\text{anc}}^{(fb)} = \frac{1}{L}\sum_{\ell=1}^{L}\sum_{j=0}^{2}\gamma_{\ell j}\mathbb{1}_{\{a_\ell=j\}}$, where $\gamma_{\ell j} = \mathbb{P}\{a_\ell = j|D,\Theta\}$.

Effects of heterozygosity and SNP density on accuracy

We conducted simulations as above, however, instead of randomly sampling each ancestral genotype, we selected three ancestral genotypes from each population on the basis of heterozygosity in each of the simulated contigs. Specifically, we sampled isolates in the 10^{th}, 50^{th}, and 90^{th} percentiles for two statistics, average contig heterozygosity, and a measure of switching between homozygous genotypes, defined as, $1/(L-1)\sum_{\ell=1}^{L-1}|\mathbb{1}_{g_{i,\ell+1}=1} - \mathbb{1}_{g_{i,\ell}=1}|$, where L is the number of sites in the contig. We simulated 25 replicates per ancestor.

We then considered the false positive rate at each SNP along the contig, for each ancestral genotype. We compared this local false positive rate between the parameters with the most false positives considered above, $\lambda_0 = 10^{-9}$ and $\lambda_1 = 10^{-6}$, to the parameters ultimately used in the analysis, $\lambda_0 = 10^{-13}$ and $\lambda_1 = 10^{-5}$. In Fig. 3.10, we show these results for several representative contigs in each population, where the smoothed heterozygosity of each SNP was the proportion of heterozygous sites in a 50kb window centered at the focal SNP.

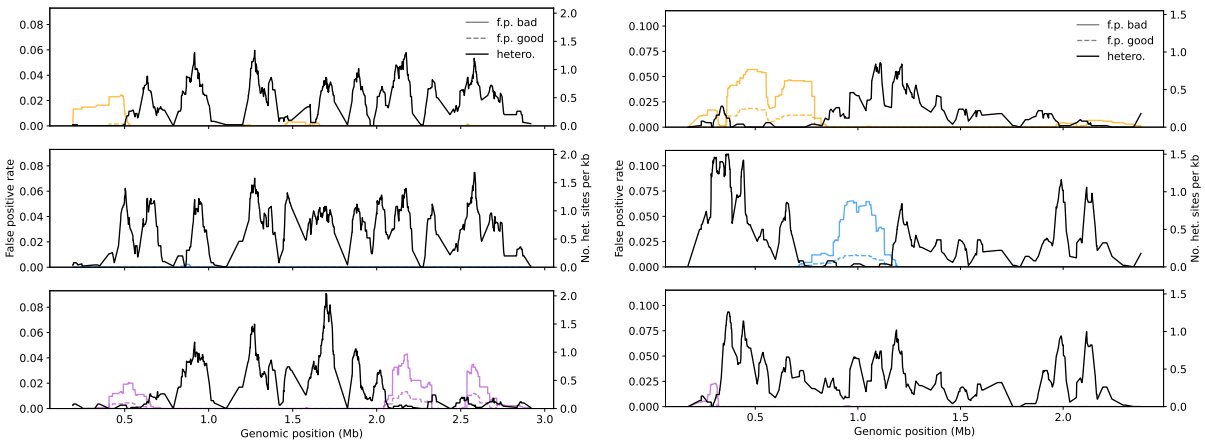
In addition, we considered the relationship between the false negative rate and SNP density, where SNP density associated with each SNP was defined as the proportion of SNPs in the 50kb window centered at the focal SNP.

Estimating the error and missingness rates from data

`ClonalHmm` treats sites with missing data as uninformative. To make the simulations more realistic, we set a per-site missing probability, p_m. This implies that the number of missing sites per simulated L-length genotype will be $\text{Bin}(L, p_m)$. Further, the expected number of missing sites in any sample of two isolates—where either or both isolates can be missing a genotype call at a "missing" site—will be $L(2p_m(1-p_m) + p_m^2) = L(1-(1-p_m)^2)$ To

(a) Biparental, contig 104

(b) Biparental, contig 60



(c) New York, contig 104

(d) New York, contig 71

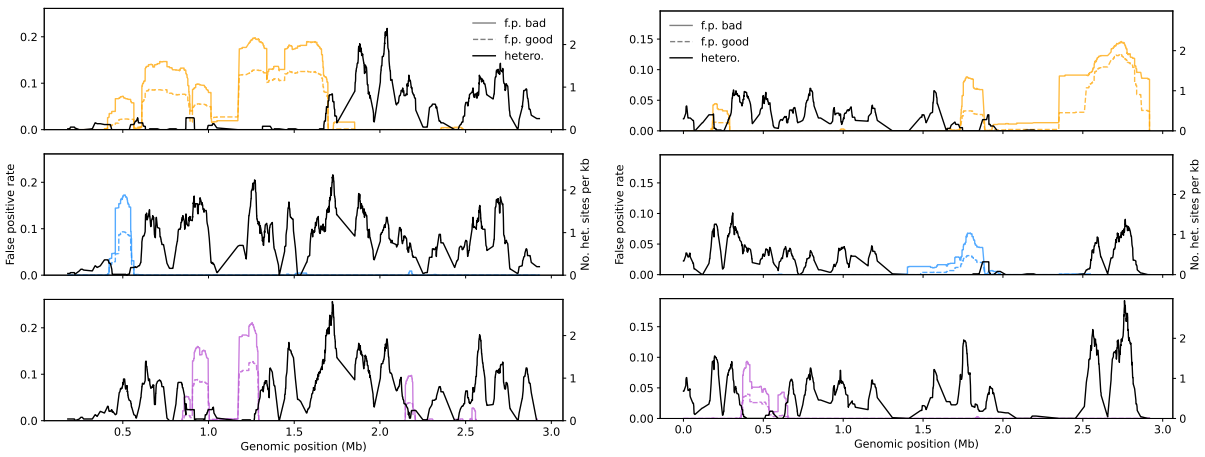


Figure 3.10: **False positives and local heterozygosity**. False positives were more common in regions of low ancestral heterozygosity. Here, we show the false positive rate for the parameters with the most false positives ($\lambda_0 = 10^{-9}$ and $\lambda_1 = 10^{-13}$, "f.p. bad") and analysis ($\lambda_0 = 10^{-13}$ and $\lambda_1 = 10^{-5}$, "f.p. good") parameters (y-axis), separated by ancestor (row), with respect to genomic position in several contigs. Smoothed heterozygosity is indicated by the solid black line.

estimate p_m from data, we averaged over the per-contig proportion of missing sites for sub-samples of size two from each clonal lineage (Section 3.6.6). In the Biparental and New York populations, our estimates of p_m were approximately $\hat{p} = 0.06$ and 0.05, respectively.

The parameter ϵ in the emissions probabilities Eq. (3.4) models genotyping error (and mutation). To estimate ϵ from data, we averaged over genotype mismatches between the same sub-samples (Section 3.6.6). This quantity $\bar{d}$ is an outcome of ϵ and p_m in our model, $1-\bar{d} = (1-p_m)^2(1-\epsilon)^2$, i.e., the probability that neither isolate has a missing genotype nor an error (nor mutation). This equation suggests an estimate of ϵ, $\hat{\epsilon} = 1-(1-\bar{d})^{\frac{1}{2}}(1-\hat{p}_m)^{-1}$, which yielded $\hat{\epsilon} = 0.09$ and 0.05 for the Biparental and NY populations, respectively.

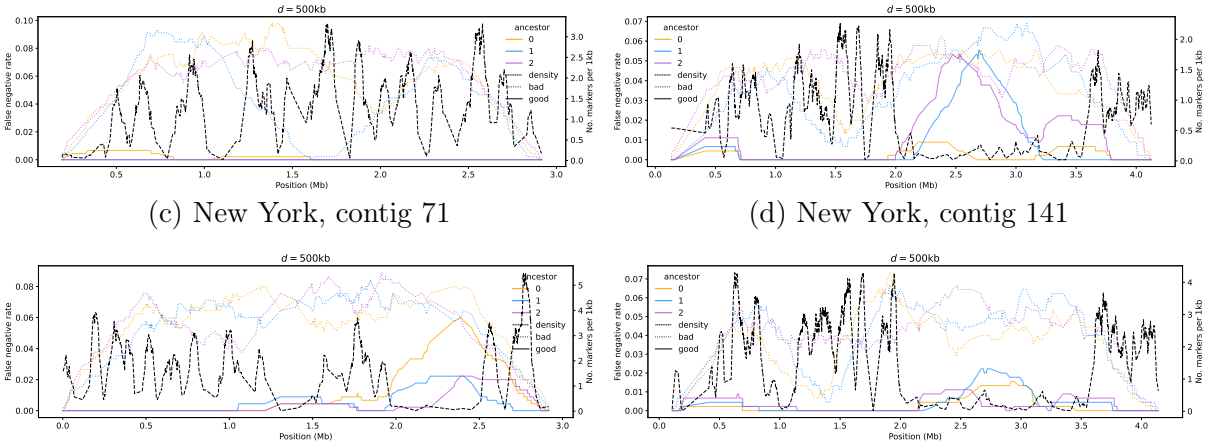
Effects of local SNP density and heterozygosity on accuracy. To further dissect the effects of local genomic characteristics, namely SNP density and heterozygosity, on the accuracy of `clonalHMM`, we considered true and false positive rate as functions of these two features (Figs. 3.10 and 3.11). In particular, we computed a smoothed per-site heterozygosity statistic, h_ℓ, which is the average of all SNPs within a 20kb SNP window (with the focal SNP at center). Similarly, we let h'_ℓ be the SNP density within a 10 kb window, with the focal SNP ℓ at its center.

In general, parameter combinations where $\lambda_1 \leq 10^{-5}$ had lower false negative rates and higher false positive rates than for larger λ_1 values (Fig. 3.2). The expected length of an LOH tract is inversely related to λ_1, thus larger values of λ_1 will perform worse when the simulated tract length is longer (Fig. 3.2). At the same time, smaller values of λ_1 sometimes yielded higher false positive rates, an effect exacerbated by larger values of λ_0.

Our simulation study revealed that `ClonalHmm` sometimes identified false positive tracts when ancestral heterozygosity was low (Section 3.3.1). These false positives were additionally marked by the fact that all members of the clonal lineage had an LOH tract spanning the false positive tract. Though simulations showed that this effect was minimal for the parameters used in our analysis (Section 3.3.1), we nonetheless excluded any LOH tracts identified in

(a) Biparental, contig 104

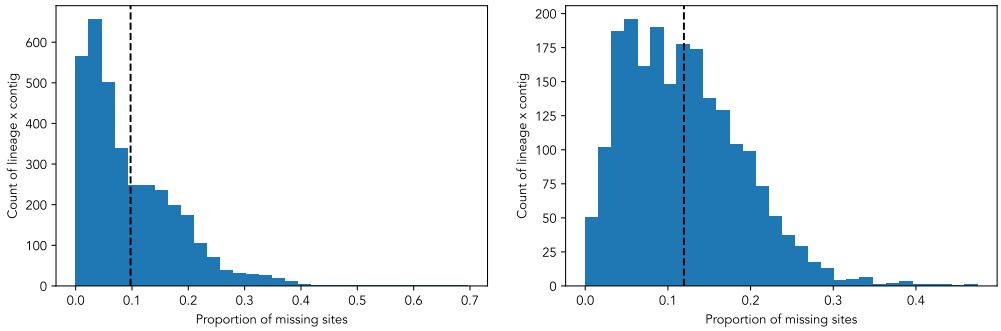
(b) Biparental, contig 141



(c) New York, contig 71

(d) New York, contig 141

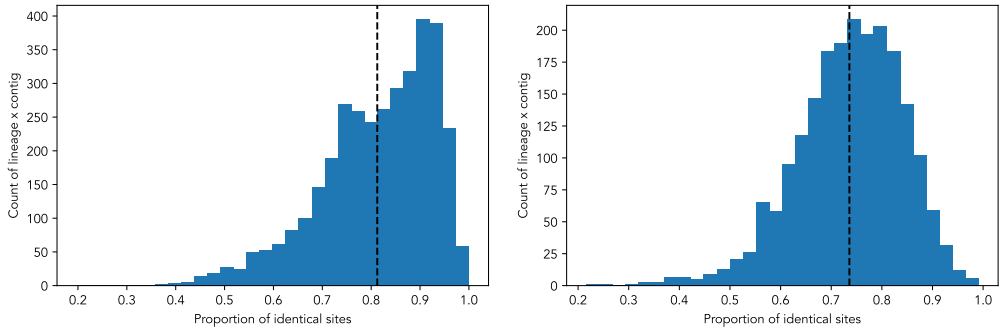
Figure 3.11: **Marker density and false negatives**. False negative rates were plotted for simulations from three different ancestral genotypes in three different contigs (a-d) as a function of smoothed marker density when the simulated tract was approximately 500kb. The dotted lines show the false negative rates for the parameters with the most false negatives ($\lambda_0 = 10^{-15}$ and $\lambda_1 = 10^{-4}$, "bad"), and the parameters used in the data analysis ($\lambda_0 = 10^{-13}$ and $\lambda_1 = 10^{-5}$, "good"). In contrast to the simulations presented in Section 3.3.1, simulated LOH tracts were only required to contain one SNP.



(a) New York

(b) Biparental

Figure 3.12: **Proportion of missing sites in sub-samples of clonal lineages.** Proportion of missing sites in sub-samples ($n = 2$) of clonal lineages in the (a) New York and (b) Biparental data sets.



(a) New York (b) Biparental

Figure 3.13: **Proportion of identical sites in sub-samples of clonal lineages**. Proportion of identical sites, i.e., non-missing and identical genotypes, in sub-samples ($n = 2$) of the clonal lineages in the (a) New York and (b) Biparental data sets.

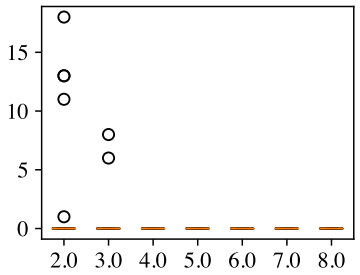


Figure 3.14: **False positives in lineages of different sizes.** A boxplot of the number of putative false positives observed in lineages of different sizes in the New York population. No putative false positives were inferred in the Biparental population.

all members of a clonal lineage from our data analysis.

3.7 Supplementary results and text

3.7.1 Inbreeding estimators

Inbreeding increases the probability that an individual inherits genomic segments identity-by-descent, thereby increasing the probability of ROH incidence. For this reason, the proportion of an individual's genome in ROH has been used as a measure of an individual's degree of inbreeding (McQuillan et al., 2008). Indeed, genomic ROH and the inbreeding coefficient F_{IS}, defined as in (Carlson et al., 2017), are highly correlated in our data set ($R^2 > .7$ in both populations, Fig. 3.17). We find a marked increase in ROH in the B population in the later years of sampling (Fig. 3.15), consistent with the previously reported increase in F_{IS} with year (Carlson et al., 2017). This trend is explained by the fact that in the early years of the experiment, F_1 from the original cross dominated, whereas inbred isolates resulting from inter-mating among the F_1 were more common in later years (Carlson et al., 2017). This correlation persisted after excluding identified LOH tracts (Fig. 3.17).

Furthermore, ROH better recapitulates the demographic structure of the B population relative to $\mathrm{F_{IS}}$, originally described in (Carlson et al., 2017). This is likely explained by the fact that after the initial filter steps (Section 3.6.1), we do not conduct any individual and site-specific genotype filtering, whereas, in (Carlson et al., 2017), all genotypes with read depths lower than five were set to missing. Thus, Fig. 3.15a reflects differences in the degree of inbreeding as well as variation in sequencing coverage across isolates. ROH appears to be more robust to variation in sequence coverage in the B population ($R^2 = -0.05$, $p = .60$). In contrast, in NY, we observe a strong negative correlation between the two measures (Fig. 3.17b). The range of sequencing coverage in NY is larger than in the B population. At high read depths, heterozygote overcalling may result in fewer inferred ROH tracts. In

addition, high read depth may be a consequence of ploidy variation, for which imbalanced allele ratios at heterozygote SNPs has provided evidence (Carlson et al., 2017; Vogel et al., 2021).

In contrast to ROH, in the B population, incidence of mitotic LOH was similar across years (Fig. 3.15). The result suggests that differences in the amount of mitotic LOH among lineages was not driven primarily by the amount of time that an isolate had been in storage and/or maintained in the laboratory—otherwise we would have observed a negative correlation between sampling year and cumulative LOH. Nonetheless, the detection of mitotic LOH in sequential cultures of the A1 and A2 parental isolates *does* demonstrate that mitotic LOH indeed occurs in the course of repeated passaging in the lab (Figs. 3.18 and 3.19), and is consistent with observations reported in (Carlson et al., 2017). LOH appears to have predominantly occurred in one of the A1 isolate sequenced in 2014, and affected approximately 6Mb of the 64Mb analyzed (Fig. 3.18b). In addition, the fact that this isolate exhibits far more ROH than its counterparts (Fig. 3.15b), suggests that `ClonalHmm` provides a conservative estimate of LOH incidence, and that more than 15Mb may have undergone LOH in the lab. Interestingly, this culture preceded those of (c-g) of Fig. 3.18 by approximately eight months, suggesting that substantial genotypic variation can be generated (in subcultures) and lost within short time periods of culturing in the lab.

Inferred LOH was less widespread in the A2 parental replicates (Fig. 3.19). However, we did recover an approximately .6Mb region which had undergone LOH in the course of the experiment, with heterozygosity in the older two replicates, and inferred LOH in the latter four (Fig. 3.20), consistent with an LOH event reported in (Carlson et al., 2017).

Carlson et al. (2017) estimated that approximately 2% of the genome was affected by LOH based on errors in Mendelian segregation in F_1. As most of these tracts were already homozygous in the oldest sequenced parental isolates (but see Fig. 3.20, they would not be detected in our analysis.

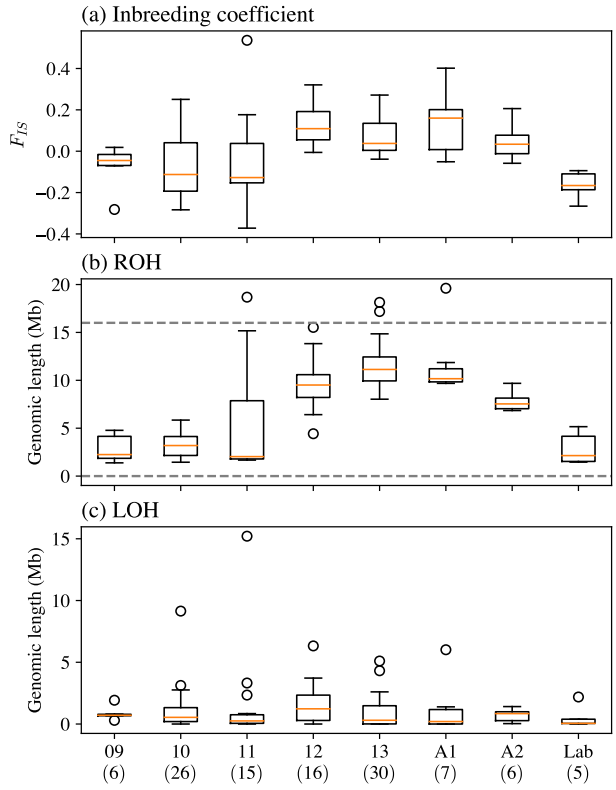


Figure 3.15: **Loss of heterozygosity incidence with year**. In (a), we plot the inbreeding coefficient F_{IS} for each isolate, separating the sampling years (2009-2013), replicates of the A1 and A2 parents, and members of the *in vitro* cross. In (b) and (c), we similarly plot the cumulative ROH (after removing LOH tracts) and LOH tract lengths, respectively.

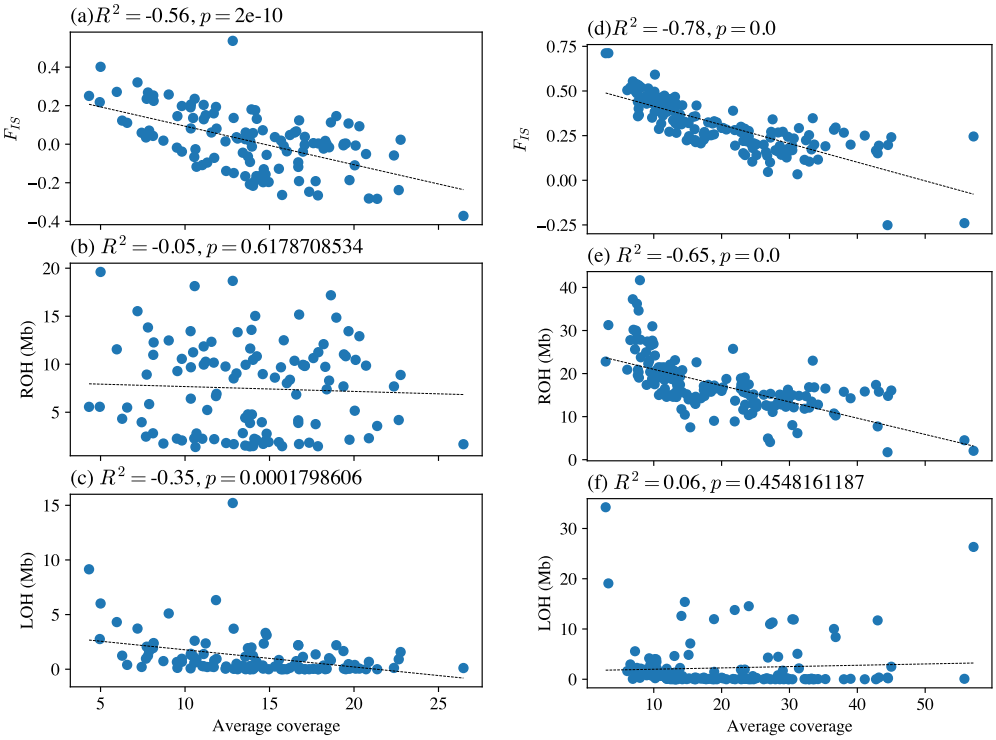


Figure 3.16: **Sequencing coverage and LOH incidence**. In (a-c), we plot the inbreeding coefficient F_{IS}, total ROH length (after excluding LOH), and LOH as a function of average sequencing coverage, for all isolates analyzed in the Biparental population. In (d-f), we plot the same quantities for the New York population.

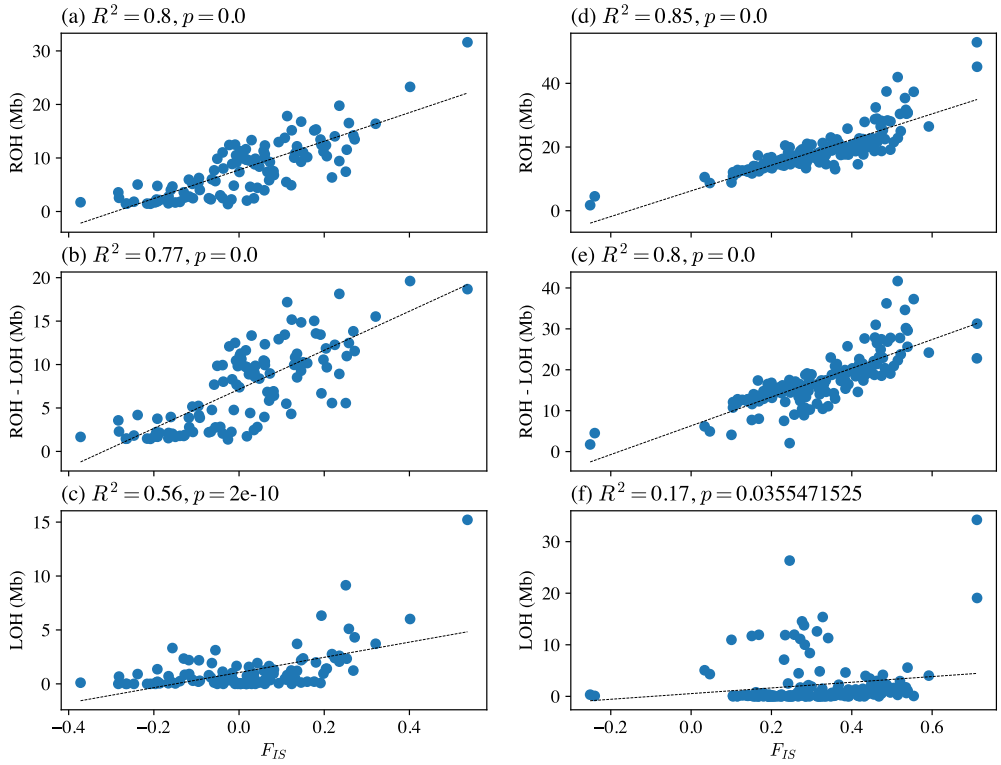


Figure 3.17: **ROH and LOH incidence and the inbreeding coefficient**. The same as Fig. 3.16, except that here we have plotted (a,d) all ROH, (b,e) ROH excluding LOH, and (c,f) LOH as a function of F_{IS} for the Biparental and NY populations, respectively.

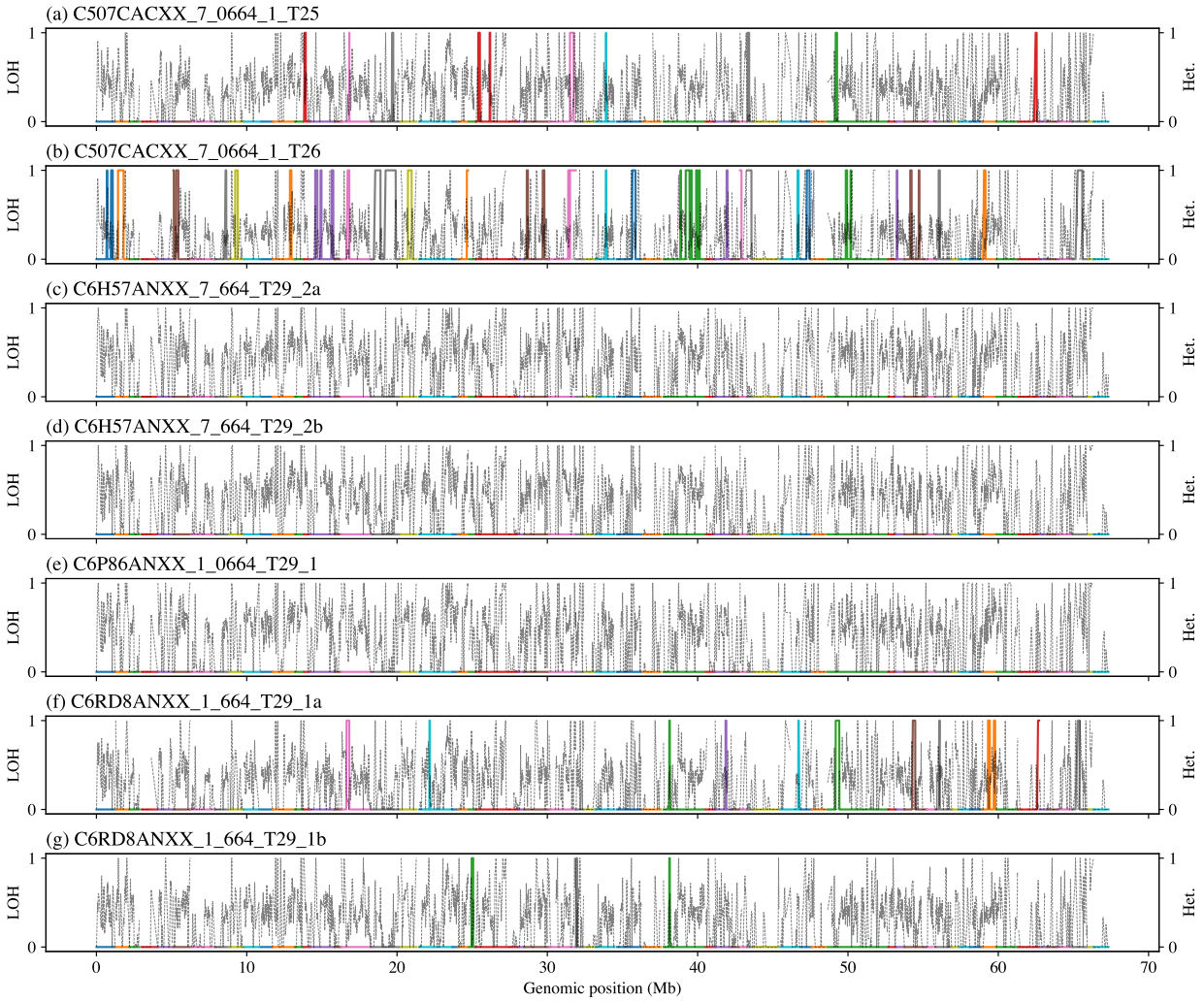


Figure 3.18: **Mitotic LOH in replicates of the A1 parent**. Cultures of the A1 parental isolate were sequenced at several time points in the course of the experiment (a-g). For a description of the replicates see Table S2 of (Carlson et al., 2017). The inferred LOH state of each isolate (y-axis, left) is plotted with respect to genomic position (x-axis), ordered by contigs (alternating colors). The dotted line represents a smoothed measure of heterozygosity (y-axis, right).

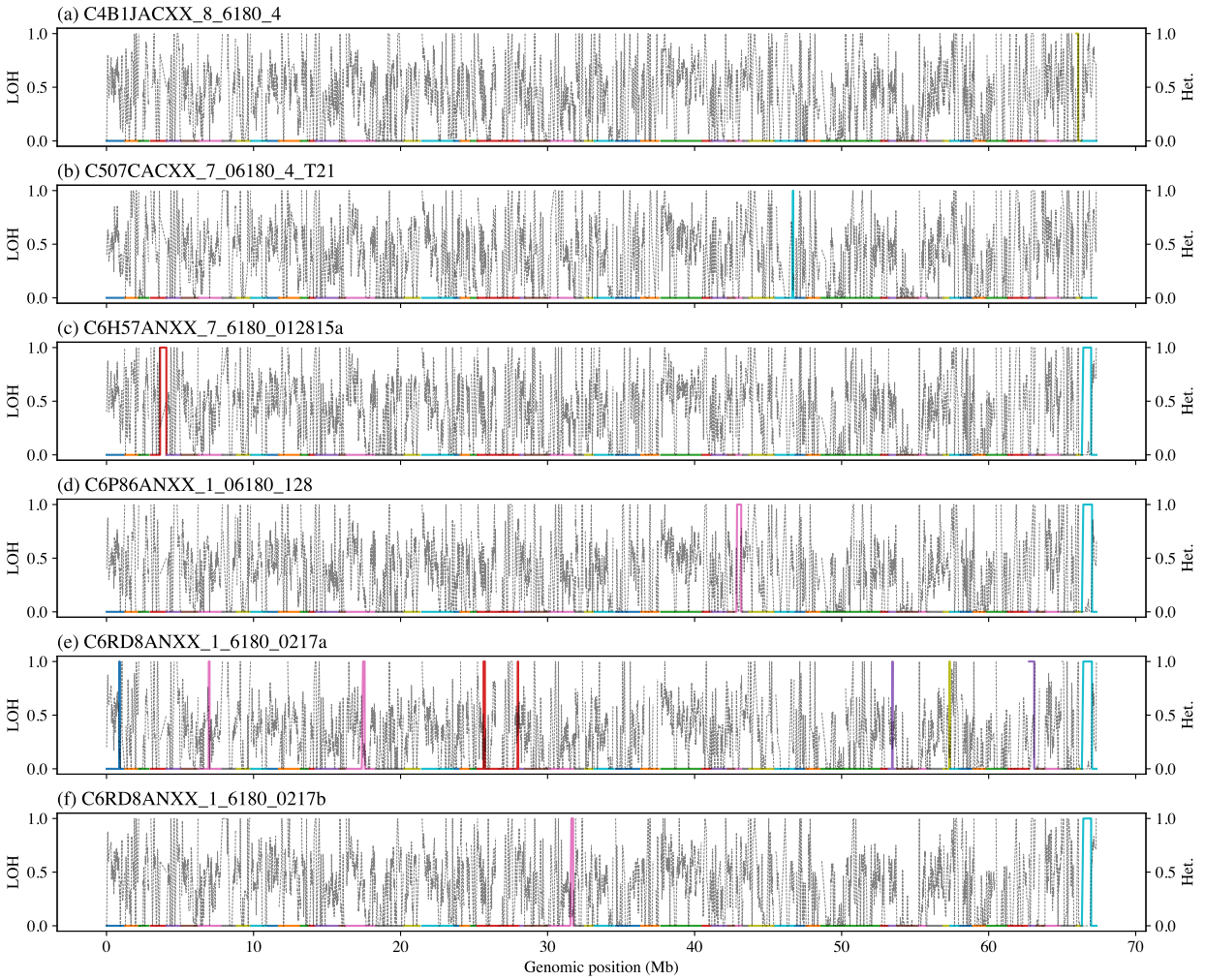


Figure 3.19: **Mitotic LOH in replicates of the A2 parent**. All is identical to Fig. 3.18, except here, replicates of the A2 parent are shown.

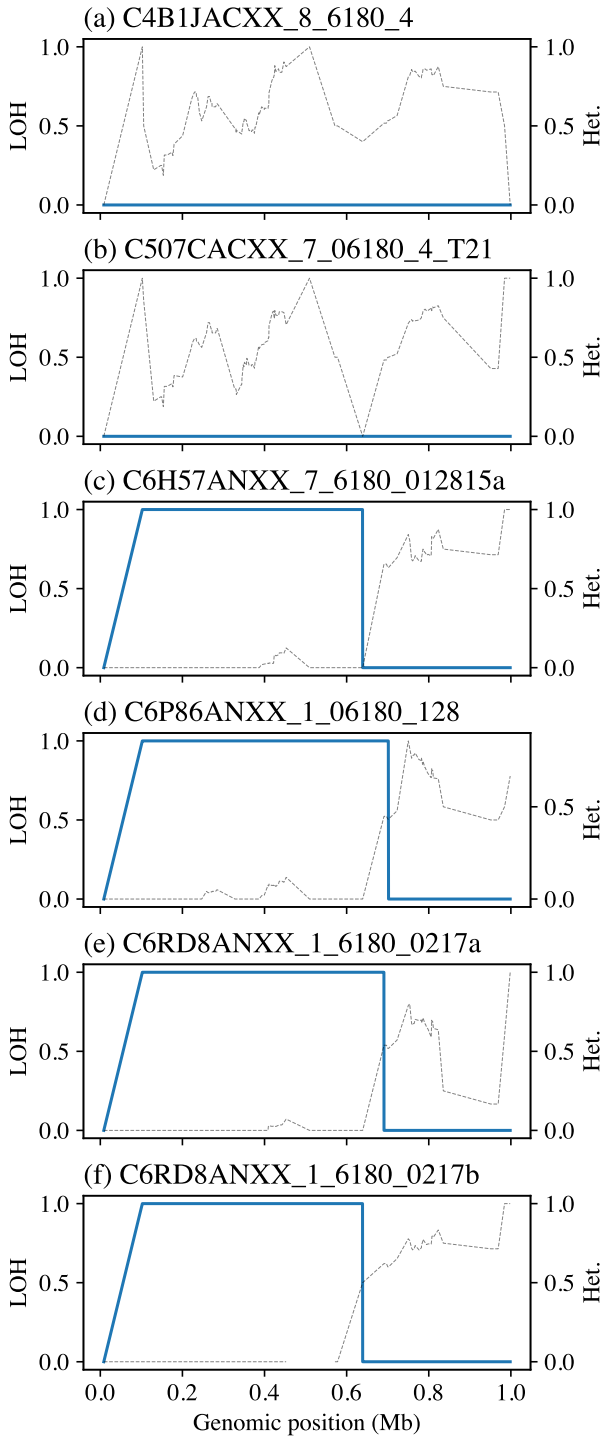


Figure 3.20: **Mitotic LOH among replicates of the A2 parent in contig 194**. All is identical to Fig. 3.19, except here, only results for contig 194 are shown.

3.7.2 An approximate EM algorithm

Expectation-maximization (EM) is a widely used iterative optimization procedure. When applied to HMMs, it is also referred to as the Baum-Welch algorithm (Rabiner, 1989). The EM is guaranteed to identify a local maximum , and in certain special cases to identify the global maximum .

Each iteration consists of two steps, an expectation (E) step and a maximization (M) step. In the E step, the expectation of the log-likelihood is taken with respect to the current parameter values. In the M step, the parameters are maximized with respect to this expectation.

Ideally, we would use EM to optimize the transition rate parameters λ_0 and λ_1 Eq. (3.2), and the error parameter ϵ Eq. (3.4). At first glance, the model of `ClonalHmm` appears well-suited to optimization with EM. However, as we will show, the form of the transition probabilities Eq. (3.2) precludes computation of an analytic form for the parameter updates. This limitation is not unique to `ClonalHmm`. As we will show, it arises in other HMM models where the transition probabilities depend on the genetic distance between markers.

3.7.3 Deriving EM updates

To illustrate the challenges that the model of `ClonalHmm` poses for optimization with EM, we present a derivation of the updates.

We refer to the set of parameters as Θ, and the values of these paramters at the t-th generation as $\Theta^{(t)}$.

E step. At the $(t+1)$-th iteration, the expected value of the log likelihood is taken with respect to the distribution of the hidden variables conditional on the observed genotypes for the n members of the clonal lineage, $\boldsymbol{G} \in \{0,1,2\}^{n \times L}$, and the current parameter values

$\Theta^{(t)}$. Denoting the matrix of individual LOH states $\boldsymbol{R} \in \{0,1\}^{n\times L}$,

$$\begin{aligned} Q(\Theta|\Theta^{(t)}) &\coloneqq \mathbb{E}_{\boldsymbol{R},\boldsymbol{a}|\boldsymbol{G},\Theta^{(t)}}\left[\log\left(\mathbb{P}\{\boldsymbol{R},\boldsymbol{a},\boldsymbol{G}|\Theta\}\right)\right] \\ &= \sum_{\boldsymbol{R},\boldsymbol{a}} \log\left(\mathbb{P}\{\boldsymbol{G}|\boldsymbol{R},\boldsymbol{a},\Theta\}\mathbb{P}\{\boldsymbol{R},\boldsymbol{a}|\Theta\}\right)\mathbb{P}\{\boldsymbol{R},\boldsymbol{a}|\boldsymbol{G},\Theta^{(t)}\}, \end{aligned} \tag{3.10}$$

where the sum is over all possible hidden states. Denoting the posterior probability of the hidden states with respect to the current parameter values as $\gamma^{(t)}(\boldsymbol{R},\boldsymbol{a}) \coloneqq \mathbb{P}\{\boldsymbol{R},\boldsymbol{a}|\boldsymbol{G},\Theta^{(t)}\}$, we have,

$$\begin{aligned} Q(\Theta|\Theta^{(t)}) &= \sum_{\boldsymbol{R},\boldsymbol{a}} \gamma^{(t)}(\boldsymbol{R},\boldsymbol{a}) \\ &\qquad \times \log\left(\mathbb{P}\{\boldsymbol{a}|\Theta\}\prod_{i=1}^{n}\mathbb{P}\{r_{i1}|\Theta\}\prod_{\ell=2}^{L}\mathbb{P}\{g_{i\ell}|r_{i\ell},a_\ell,\Theta\}\mathbb{P}\{r_{i\ell}|r_{i,\ell-1},\Theta\}\right) \\ &= \sum_{\boldsymbol{R},\boldsymbol{a}} \gamma^{(t)}(\boldsymbol{R},\boldsymbol{a})\Big[\log(\mathbb{P}\{\boldsymbol{a}|\Theta\})+ \\ &\qquad \sum_{i=1}^{n}\log\left(\mathbb{P}\{r_{i1}|\Theta\}\right)+\sum_{\ell=2}^{L}\log\left(\mathbb{P}\{g_{i\ell}|r_{i\ell},a_\ell,\Theta\}\right)+\log\left(\mathbb{P}\{r_{i\ell}|r_{i,\ell-1},\Theta\}\right)\Big]. \end{aligned} \tag{3.11}$$

M step. At the t-th iteration, the expectation defined in Eq. (3.10) is maximized to find the parameter updates,

$$\Theta^{(t+1)} = \text{argmax}_{\Theta} Q(\Theta|\Theta^{(t)}). \tag{3.12}$$

To optimize λ_0, we differentiate Eq. (3.11),

$$\begin{aligned}\frac{\partial Q(\Theta|\Theta^{(t)})}{\partial \lambda_0} &= \sum_{\boldsymbol{R},\boldsymbol{a}} \gamma^{(t)}(\boldsymbol{R},\boldsymbol{a}) \sum_{\ell=2}^{L} \frac{\partial}{\partial \lambda_0} \log(\mathbb{P}\{\boldsymbol{r}_{i\ell}|\boldsymbol{r}_{i,\ell-1},\Theta\}) \\ &= \sum_{\boldsymbol{R},\boldsymbol{a}} \gamma^{(t)}(\boldsymbol{R},\boldsymbol{a}) \sum_{i=1}^{n} \sum_{\ell=2}^{L} d_\ell \left((1-e^{-\lambda_0 d_\ell})^{-1} e^{-\lambda_0 d_\ell} \mathbb{1}_{i\ell}\{0\to 1\} - \mathbb{1}_{i\ell}\{0\to 0\} \right),\end{aligned} \tag{3.13}$$

and then set the above equation equal to 0,

$$\sum_{\ell=2}^{L} d_\ell \sum_{i=1}^{n} \gamma_{i,\ell}^{(t)}(0\to 0) = \sum_{\ell=2}^{L} d_\ell \sum_{i=1}^{n} (1-e^{-\lambda_0 d_\ell})^{-1} e^{-\lambda_0 d_\ell} \gamma_{i,\ell}^{(t)}(0\to 1), \tag{3.14}$$

where we introduced the notation $\gamma_{i,\ell}^{(t)}(r\to s) \coloneqq \mathbb{P}\{r_{i,\ell+1}=s|r_{i,\ell+1}=r,\boldsymbol{G},\Theta^{(t)}\}$.

If we use the transition probabilities of Narasimhan et al. (2016), we arrive at a similar expression,

$$\sum_{\ell=2}^{L} d_\ell \sum_{i=1}^{n} \gamma_{i,\ell}^{(t)}(0\to 0) = \sum_{\ell=2}^{L} d_\ell \sum_{i=1}^{n} \frac{\lambda_0 d_\ell}{1-\lambda_0 d_\ell} \gamma_{i,\ell}^{(t)}(0\to 1). \tag{3.15}$$

These expressions cannot be readily simplified to obtain analytic solutions for the M-step, and thus solving the M-step in this context would require a numerical procedure.

Per-site rate parameters. The problem simplifies significantly if we allow the rate to vary by site, in the spirit of a similar derivation presented in Scheet and Stephens (2006) in the context of haplotype phasing. Though per-site rate estimation is likely not suited to our application—a clonal lineage only consists of a handful of individuals, and mitotic LOH is rare—the derivation illustrates one way of circumventing the issue evident in our derivations

above. Letting $\lambda_{0\ell}$ be the rate parameter governing transitions between sites ℓ and $\ell+1$,

$$\begin{aligned} 0 &= \frac{\partial Q(\Theta|\Theta^{(t)})}{\partial \lambda_{0\ell}} \\ \frac{\sum_{i=1}^{n} \gamma_{i,\ell}^{(t)}(0 \to 0)}{\sum_{i=1}^{n} \gamma_{i,\ell}^{(t)}(0 \to 1)} &= (1 - e^{-\lambda_{0\ell} d_\ell})^{-1} e^{-\lambda_{0\ell} d_\ell} = (e^{\lambda_{0\ell} d_\ell} - 1)^{-1}. \end{aligned} \tag{3.16}$$

And, thus,

$$\lambda_{0,\ell}^{(t+1)} = \frac{1}{d_\ell} \log \left[\frac{\sum_{i=1}^{n} \gamma_{i,\ell}^{(t)}(0 \to 1)}{\sum_{i=1}^{n} \gamma_{i,\ell}^{(t)}(0 \to 0)} + 1 \right], \tag{3.17}$$

which closely resembles Equation C3 of Scheet and Stephens (2006).

3.7.4 Viterbi training

In Narasimhan et al. (2016), the authors state that they used a Viterbi training procedure to optimize the transition rate parameters. In Viterbi training, the parameters are optimized with respect to the Viterbi path, computed with respect to the current parameter values, rather than the full posterior distribution, as in the full EM. While Viterbi training does not fundamentally simplify the optimization problem, it would reduce the computational complexity of the EM algorithm by a factor of 2^n.

In this case, the M step takes the form,

$$\sum_{i=1}^{n} \sum_{\ell=1}^{L} \mathbb{1}\{r_{i,\ell+1}^{(v,t)} = 0, r_{i,\ell}^{(v,t)} = 0\} = \sum_{i=1}^{n} \sum_{\ell=1}^{L} \frac{\lambda_0 d_\ell}{1 - \lambda_0 d_\ell} \mathbb{1}\{r_{i,\ell+1}^{(v,t)} = 0, r_{i,\ell}^{(v,t)} = 1\}, \tag{3.18}$$

where $r_{i,\ell}^{(v,t)} \in \{0, 1\}$ are the inferred Viterbi LOH states at the t-th iteration for individual i. Thus, the M step presents a numerical optimization problem almost identical to that of Eq. (3.15).

3.7.5 An alternative error model

In the main text, we employed an error model that assigns equal weight to the incorrect genotypes, i.e., if the true genotype is 0, then an erroneous genotype call is equally likely to be 1 as it is 2. The probabilities of these two errors likely differ and depend on the total read count at the site. Modeling read counts rather than genotype calls provides may therefore more accurate emissions probabilities.

Let $(g_{i\ell}^{(1)}, g_{i\ell}^{(2)})$ be the read counts for the reference (A_1) and alternate (A_2) alleles at locus ℓ in individual i. Instead of treating the genotype call as the emission from the HMM, we model the read counts and integrate over the possible genotypes. We assume that conditional on the true genotype, $g_{i\ell}^*$, the read counts are binomially distributed with some error rate, $\tilde{\epsilon}$. In contrast to the ϵ defined in Eq. (3.4), $\tilde{\epsilon}$ reflects errors in the sequencing reads specifically, and,

$$\mathbb{P}\{(g_{i\ell}^{(1)}, g_{i\ell}^{(2)}) = (g_1, g_2) | g_{i\ell} = g\} = \binom{g_1 + g_2}{g_1} \begin{cases} (1-\tilde{\epsilon})^{g_1}(\tilde{\epsilon})^{g_2} & g = 0 \\ (.5)^{g_1}(.5)^{g_2} & g = 1 \\ (\tilde{\epsilon})^{g_1}(1-\tilde{\epsilon})^{g_2} & g = 2 \end{cases} . \tag{3.19}$$

Thus, integrating over the possible genotypic states, the emissions probabilities become,

$$\begin{aligned} \mathbb{P}\{(g_{i\ell}^{(1)}, g_{i\ell}^{(2)}) &= (g_1, g_2) | a_\ell = a, r_{i\ell} = r\} \\ &= \sum_{g=0}^{2} \mathbb{P}\{(g_{i\ell}^{(1)}, g_{i\ell}^{(2)}) = (g_1, g_2) | g_\ell = g\} \mathbb{P}\{g_\ell = g | a_\ell = a, r_{i\ell} = r\}, \end{aligned} \tag{3.20}$$

where the first term in each summand is given be Eq. (3.19), and the second is given by Eq. (3.4). In this context, the error rate ϵ of Eq. (3.4) can be interpreted more narrowly as the mutation rate, while $\tilde{\epsilon}$ represents sequencing errors alone.

When read counts are sufficiently high and genotypes are thus called with high confidence,

modeling the read counts explicitly may not be necessary.

3.8 Supplementary figures

In this section, we provide additional figures to support the observations in the main text.

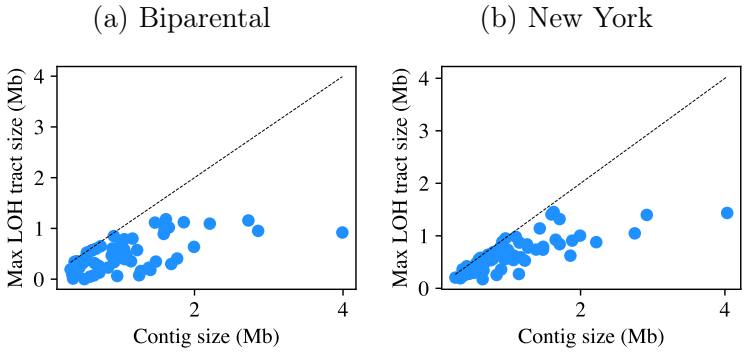


Figure 3.21: **Contig size and inferred LOH tract sizes.** In (a) and (b), we show the maximum inferred tract length as a function of the size of the contig (Mb), for the Biparental and New York populations, respectively. The dotted line indicates corresponds to identity.

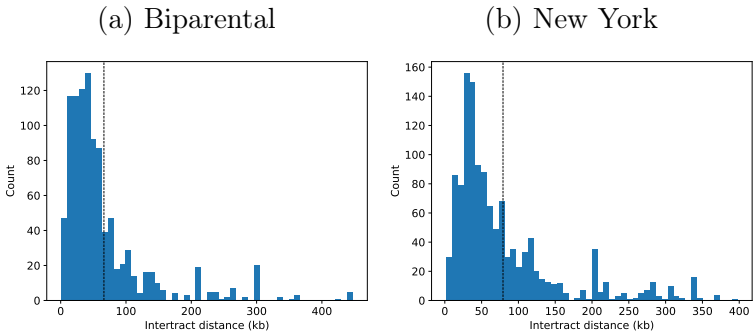


Figure 3.22: **Spurious gaps between inferred LOH tracts.** We plot the distributions of spurious gaps between inferred LOH tracts in simulated data for the (a) Biparental and (b) New York populations, in contig 141.

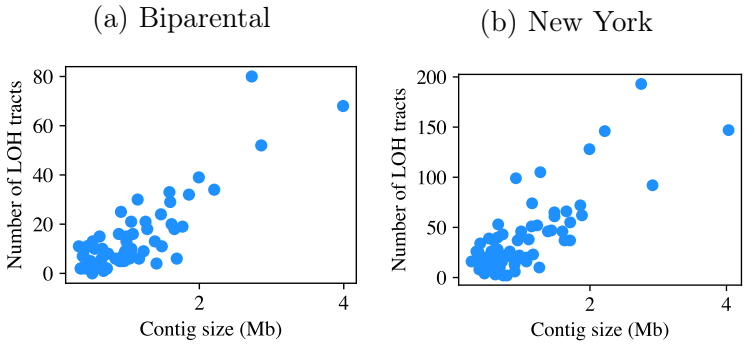


Figure 3.23: **Contig size and number of inferred LOH tracts.** In (a) and (b), we show the number of inferred LOH tracts as a function of the size of the contig (Mb), for the Biparental and New York populations, respectively. The dotted line corresponds to identity.

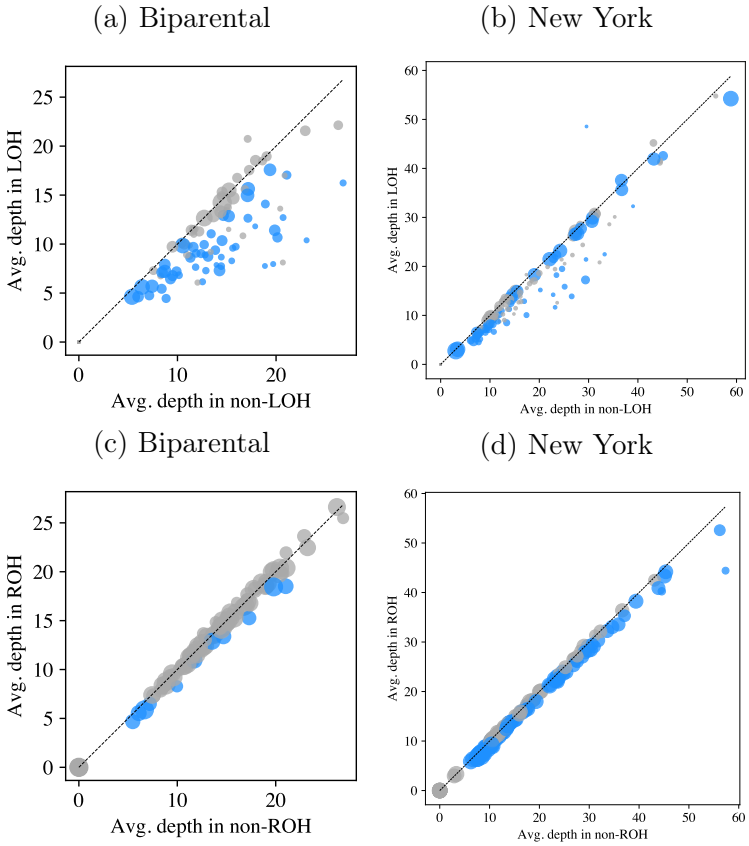


Figure 3.24: **Read depth inside and outside of LOH tracts**. In (a) and (b), we show average read depth inside of inferred LOH tracts as a function of the same quantity outside of inferred LOH tracts, for each isolate analyzed in the Biparental (B) and New York (NY) populations. The size of the dot is a function of the total genomic length in inferred LOH tracts. The color indicates whether the average depths were significantly different (blue) in a two-sample T-test after multiple testing correction, or not (gray). In (c) and (d), we show the same statistics for ROH tracts, respectively. The dotted lines correspond to identity.

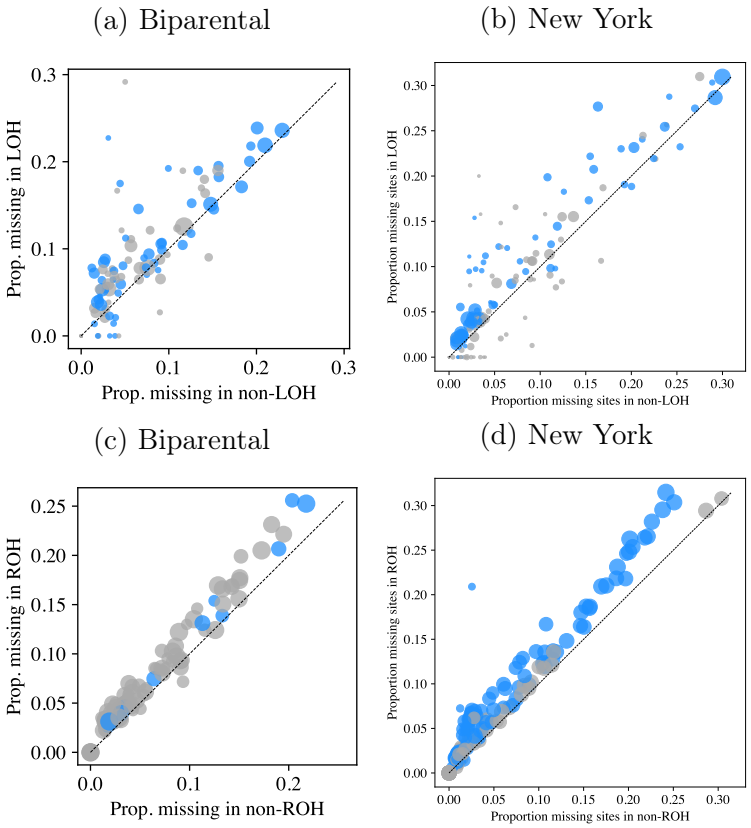


Figure 3.25: **Missing data inside and outside of LOH tracts.** We reproduce Fig. 3.24, except here, the proportion of missing sites is plotted instead of average read depth. The points are colored with respect to the significance test of Fig. 3.24.

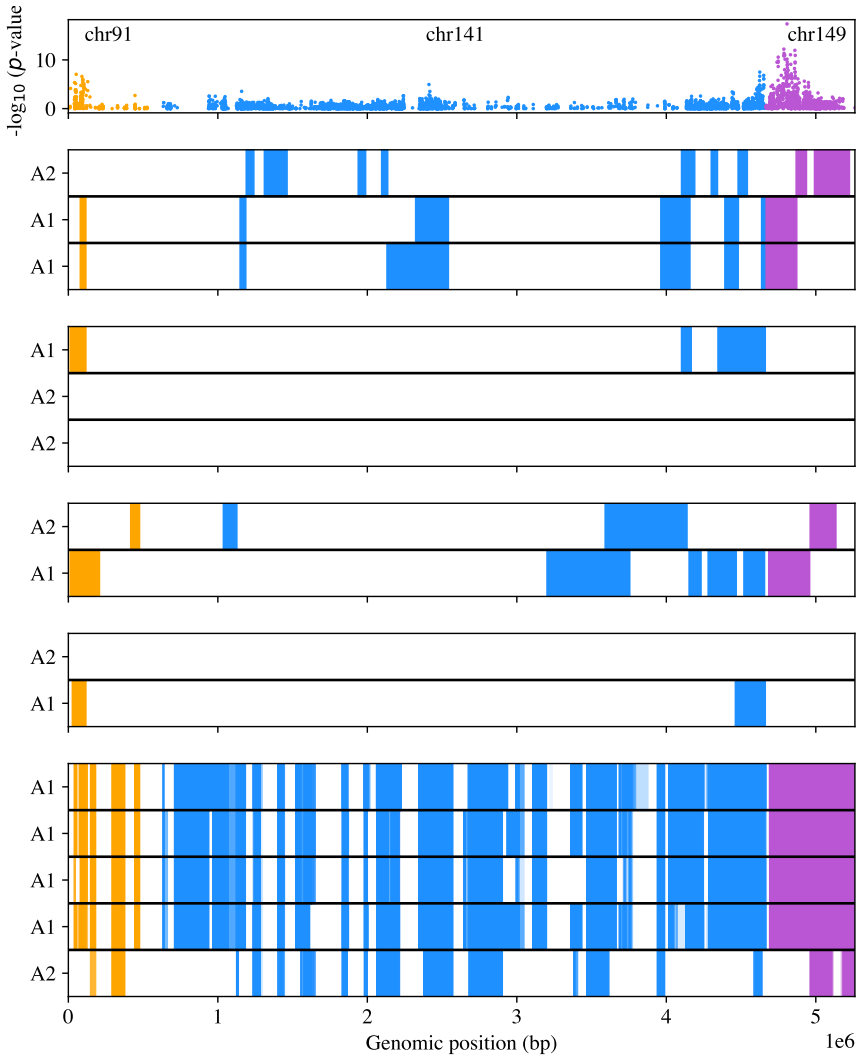


Figure 3.26: **LOH in the mating type region and mating type discordance.** All lineages in which both mating types were observed exhibited LOH in the mating type region (MTR). In (a), we show the results of a Fisher's exact test of allele frequency differences between mating types in the New York (NY) population. The -$\log_{10}$ (p-value) is plotted as a function of genomic position in the three contigs comprising the MTR, contigs 91, 141, and, 149. Panels (b)-(f) each correspond to a lineage with discordant mating types, with mating type indicated along the y-axis. In (b)-(e), the shaded regions denote LOH tracts inferred by `ClonalHmm`. In (f), as `ClonalHmm` did not infer LOH tracts in the A1 isolates, the intensity of the color reflects a smoothed measure of homozygosity, with darker color corresponding to more homozygous regions.

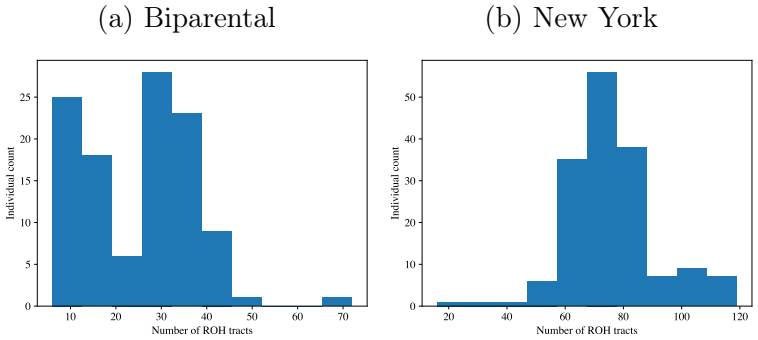


Figure 3.27: **Number of runs of homozygosity per individual.** A histogram of the number of inferred runs of homozygosity (ROHs) in (a) Biparental and (b) New York isolates.

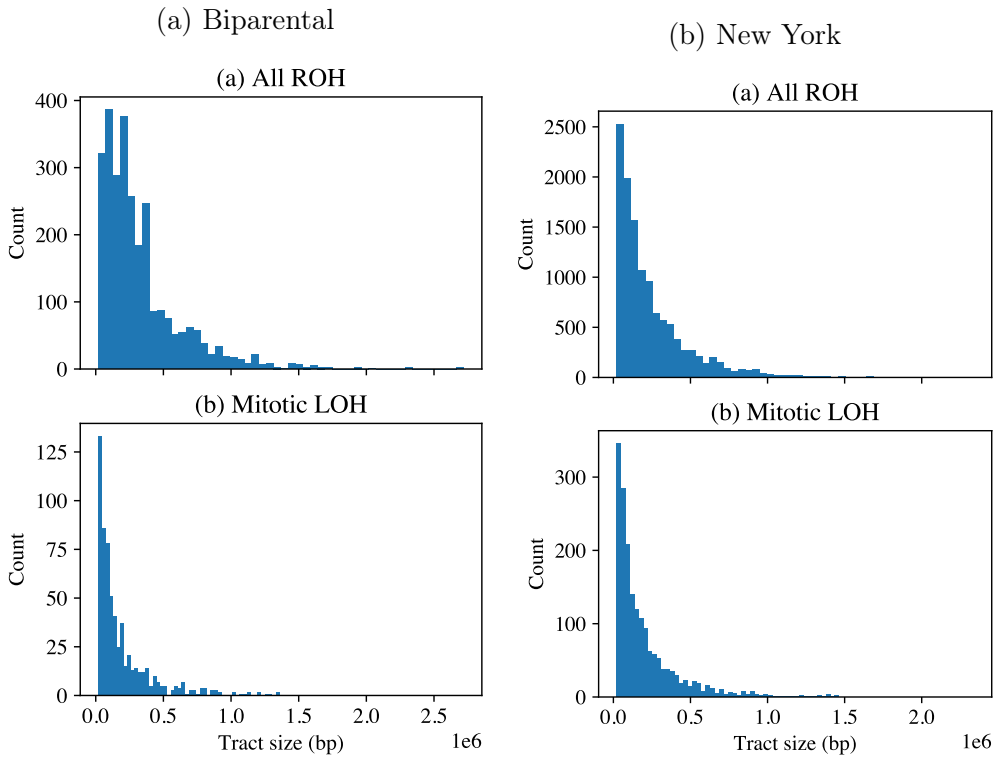


Figure 3.28: **ROH and LOH cumulative tract size distributions.** Histograms of ROH tract lengths (a,c) and LOH tract lengths (b,d) in the Biparental and New York populations, respectively.

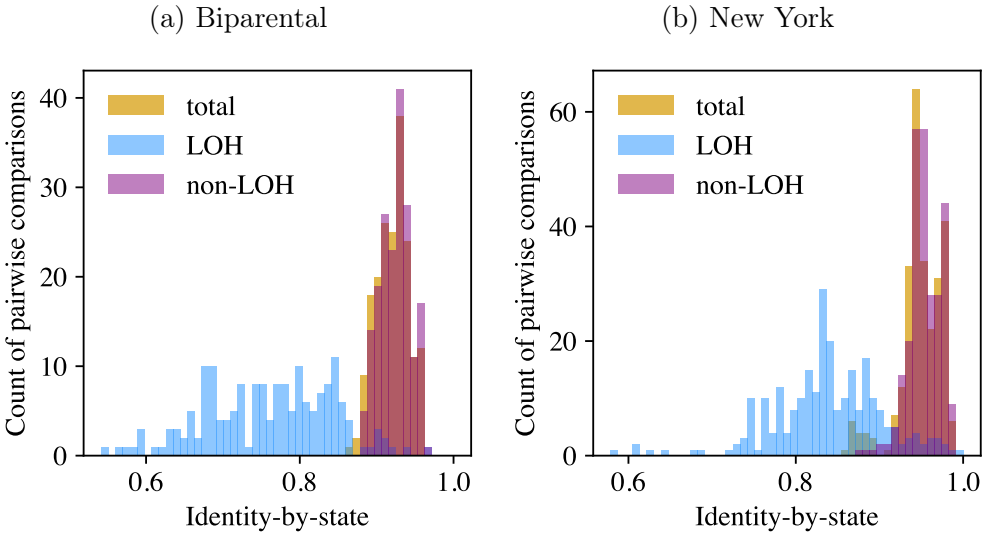


Figure 3.29: **Identity-by-state in LOH tracts.** Histograms of identity-by-state (IBS) computed for all pairwise comparisons of isolates within lineages in the (a) Biparental and (b) New York populations. The blue histograms summarize IBS non-shared LOH tracts in both isolates, and the orange histograms summarize IBS computed with respect to the portion of SNPs not affected by LOH in either isolate.

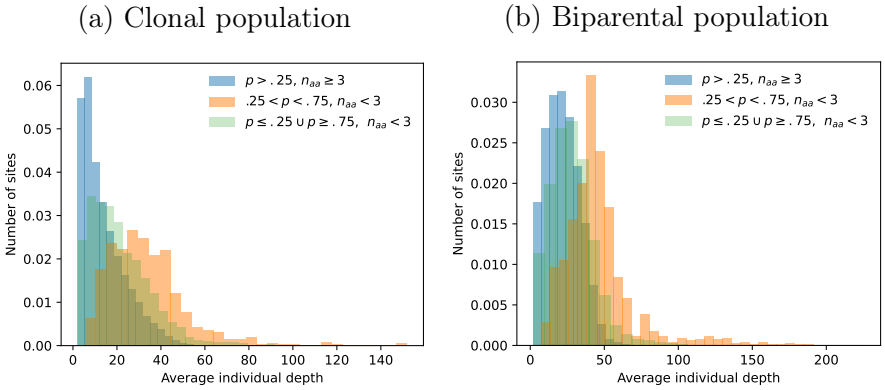


Figure 3.30: **Excess heterozygosity and allele frequency.** Sites with few minor allele homozygotes (orange and green) tend to have higher read depth than sites with more than three observed minor allele homozygote (blue) in both the NY (a) and Biparental (b) data sets. This trend suggests that duplications in the sampled population not represented in the reference genome may be responsible for erroneous heterozygote calls. The effect is most pronounced for moderate frequency alleles, as observing a minor allele homozygote is less probable for low frequency sites. This implies that our filter on minor allele heterozygosity may erroneously remove low frequency sites.

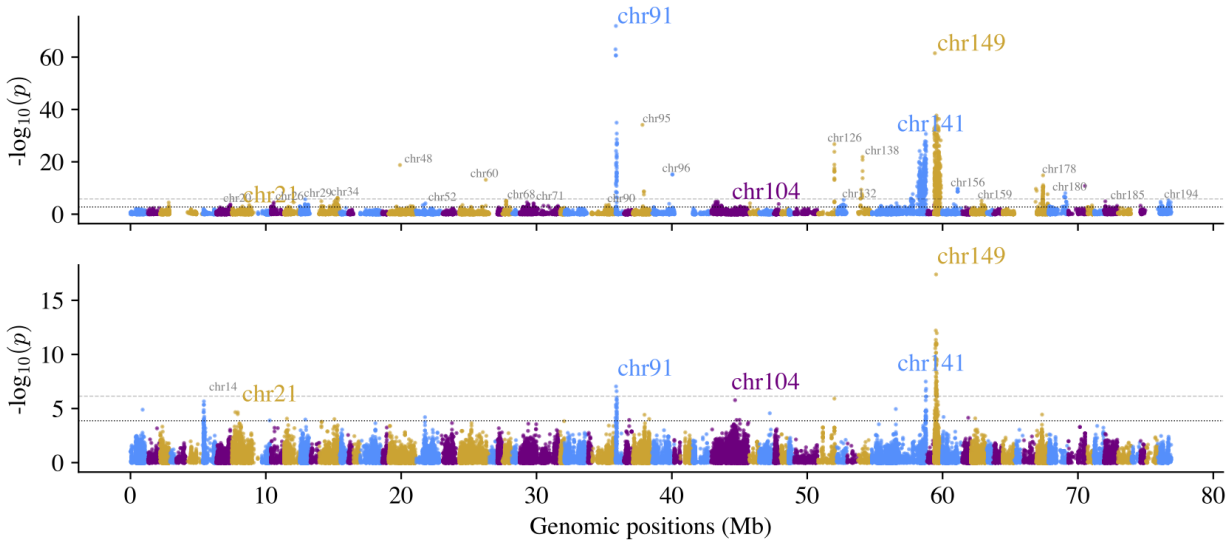


Figure 3.31: **Allele frequency differences between mating types.** The -$\log_{10}$ transformed p-values of a Fisher's exact test of allele frequency differences performed in the (a) Biparental and (b) New York populations. The gray dashed lines correspond to the Bonferroni thresholds ($\alpha = 0.05$) and the black dotted lines to the false-discovery rate (FDR) thresholds ($\alpha = 0.05$), computed separately in each population. When contig contained at least one SNP exceeding the FDR threshold in one population it was labeled in gray, small text, when it exceeded the thresholds in both populations, it was labeled in large, colored text.

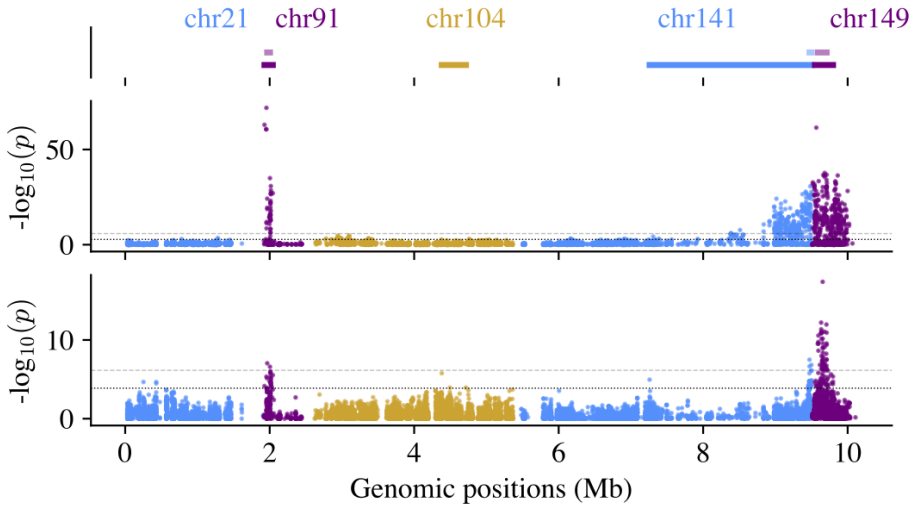


Figure 3.32: **Allele frequency differences in the mating type region.** The same as Fig. 3.31, except only the contigs containing significantly associated SNPs in both populations are shown. The top track shows the region containing significantly associated SNPs in the New York (top) and Biparental (bottom) populations.

CHAPTER 4

CONCLUSION

4.1 Polygenic score accuracy

In Chapter 2, I investigated the effects of allelic turnover on the accuracy of out-of-sample polygenic scores. To do so, I developed a generative model for polygenic scores for individuals sampled in the past, i.e., ancient individuals, or for individuals from a population that diverged from the GWA study population at some point in the past. Under the model assumptions used therein, I showed that these two scenarios were analogous. That is, the problem of predicting the phenotypes of individuals with distant ancestry is fundamentally analogous to predicting the phenotypes of ancient individuals.

In this work, I showed that allelic turnover alone can explain large reductions in prediction accuracy over human relevant time scales. Specifically, my model predicts an approximately 50% reduction in accuracy due to allelic turnover in individuals of African ancestry. In practice, the observed relative accuracy in individuals of African ancestry, as well as other ancestry groups, is frequently lower than my predictions. This suggests that other factors may be contributing substantially to accuracy reductions. For example, decay in linkage disequilibrium between the genotyped and casual sites in the focal (predicted) population and differences in environmental conditions may both contribute to accuracy reductions. Future work may investigate the consequences of these two features, and others, on polygenic score accuracy.

Importantly, this work suggests that analysis of a time series of polygenic scores, for example, from ancient samples, should take into account the fact that the statistical properties of a polygenic score depend on when the focal individual was sampled. Under neutrality, the *mse* increases and the estimated additive genetic variance, $\hat{V}_A$, decreases as a function of the sampling time. My work provides explicit expressions for these quantities. However,

these expressions rely on assumptions about the population genetic dynamics and trait architecture. For example, to compute *mse* as a function of the ancient sampling time, one must make an assumption about the mutational target size, i.e., how many sites potentially contribute to trait variance. Further, more work is required to compute these statistics *conditional* on the results of a GWA study.

In addition, the simulations described in Chapter 2 show that even weak selection can result in a biased polygenic score. This result suggests that estimates of selection coefficients from polygenic score time-series may be inflated with respect to the true strength of selection. Future work may investigate how to correct inference procedures for this inherent bias.

In conclusion, Chapter 2 demonstrates that a relatively simple model can yield significant insights into the statistical properties of polygenic scores.

4.2 Mitotic loss of heterozygosity

In Chapter 3, I developed a procedure to infer mitotic loss of heterozygosity (LOH) from the genotypes of multiple representatives of a clonal lineage. This method takes advantage of the fact that all genetic variation among members of a clonal lineage can be attributed to mitotic processes to infer both LOH events and the genotype of the most recent common ancestor. This method, referred to as `ClonalHmm`, was applied to two data sets of the oomycete plant pathogen *Phytophthora capsici*. Our results demonstrated that mitotic LOH is common among clonal lineages as well as variable within lineages. In addition, mitotic LOH incidence was widespread across the genome, with some evidence of elevated incidence in particular genomic regions, including several regions containing mating type associated SNPs.

While `ClonalHmm` is an effective method for identifying mitotic LOH, several features of the data sets limit our interpretation of the empirical results. Foremost, the isolates within a lineage are separated by variable and unknown numbers of cell divisions. Thus, the ob-

servation of more LOH in one member of a clonal lineage relative to another cannot be unequivocally attributed to a higher LOH rate as the isolate may have simply experienced more cell divisions. To overcome this limitation more precise measurements of growth coupled with genotyping are required. Similarly, except in cases where an isolate was genotyped at multiple time points, the observed mitotic LOH events cannot be attributed to growth in the lab or field. (Though, if a subset of n_s isolates, where $n > n_s > 1$, within a lineage exhibit the same LOH event, it is more parsimonious for the LOH event to have occurred in the field, prior to sampling.) Different (or diminished) environmental pressures in the lab may also affect LOH rates.

Future theoretical work may investigate the consequences of mitotic LOH on population genetic dynamics. As discussed, mitotic LOH may provide an evolutionary advantage for asexually reproducing organisms. In particular, mitotic recombination may uniquely increase the efficacy of selection. In mitotic recombination, the reciprocal products of recombination are inherited by distinct daughter cells. In a radially expanding culture, this would place all three possible genotypes in close proximity, facilitating competition between them.

The fact that *P. capsici* also reproduces sexually slightly complicates this picture. Indeed, the spores resulting from sexual reproduction, oospores, are the only propagules which can overwinter in regions with cold winters. Sexual reproduction may mitigate reductions in heterozygosity due to LOH if genetically divergent isolates reproduce, but would accelerate reductions in heterozygosity if mating occurred between related isolates. Future work may investigate the joint effects of mitotic LOH and sexual reproduction on the population genetic dynamics of *P. capsici*.

In systematically identifying mitotic LOH events within clonal lineage, Chapter 3 provokes more questions than it answers. Careful experimental and theoretical work may help us to begin to answer these questions.

REFERENCES

Hussein Al-Asadi, Kushal Dey, John Novembre, and Matthew Stephens. Inference and visualization of DNA damage patterns using a grade of membership model. *Bioinformatics*, pages 1–7, 2018. doi:10.1093/bioinformatics/bty779.

Theresa Albrecht, Valentin Wimmer, Hans Jürgen Auinger, Malena Erbe, Carsten Knaak, Milena Ouzunova, Henner Simianer, and Chris Carolin Schön. Genome-based prediction of testcross values in maize. *Theoretical and Applied Genetics*, 123(2):339–350, jul 2011. ISSN 00405752. doi:10.1007/s00122-011-1587-7. URL `https://link.springer.com/article/10.1007/s00122-011-1587-7`.

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

Jonathan P. Beauchamp. Genetic evidence for natural selection in humans in the contemporary United States. *Proceedings of the National Academy of Sciences of the United States of America*, 113(28):7774–7779, jul 2016. ISSN 10916490. doi:10.1073/pnas.1600398113. URL `https://www.pnas.org/doi/abs/10.1073/pnas.1600398113`.

Yoav Benjamini and Yosef Hochberg. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1):289–300, jan 1995. ISSN 2517-6161. doi:10.1111/j.2517-6161.1995.tb02031.x. URL `https://onlinelibrary.wiley.com/doi/full/10.1111/j.2517-6161.1995.tb02031.xhttps://onlinelibrary.wiley.com/doi/abs/10.1111/j.2517-6161.1995.tb02031.xhttps://rss.onlinelibrary.wiley.com/doi/10.1111/j.2517-6161.1995.tb02031.x`.

Jeremy J. Berg and Graham Coop. A population genetic signal of polygenic adaptation. *PLoS Genetics*, 10(8):e1004412, aug 2014. doi:10.1371/journal.pgen.1004412.

Jeremy J Berg, Arbel Harpak, Nasa Sinnott-Armstrong, Anja Moltke Joergensen, Hakhamanesh Mostafavi, Yair Field, Evan August Boyle, Xinjun Zhang, Fernando Racimo, Jonathan K Pritchard, and Graham Coop. Reduced signal for polygenic adaptation of height in UK Biobank. *eLife*, 8, 2019. doi:10.7554/elife.39725. URL `https://doi.org/10.7554/eLife.39725.001`.

Rameen Beroukhim, Ming Lin, Yuhyun Park, Ke Hao, Xiaojun Zhao, Levi A Garraway, Edward A Fox, Ephraim P Hochberg, Ingo K Mellinghoff, Matthias D Hofer, Aurelien Descazeaud, Mark A Rubin, Matthew Meyerson, Wing Hung Wong, William R Sellers, and Cheng Li. Inferring loss-of-heterozygosity from unpaired tumors using high-density oligonucleotide snp arrays. *PLOS Computational Biology*, 2(5):1–10, 05 2006. doi:10.1371/journal.pcbi.0020041. URL `https://doi.org/10.1371/journal.pcbi.0020041`.

Bárbara D. Bitarello and Iain Mathieson. Polygenic scores for height in admixed populations. *G3: Genes, Genomes, Genetics*, 10(11):4027–4036, nov 2020. ISSN 21601836. doi:10.1534/g3.120.401658. URL https://doi.org/10.1534/g3.120.401658.

Jonathan P Bollback, Thomas L York, and Rasmus Nielsen. Estimation of $2N_e s$ from temporal allele frequency data. *Genetics*, 179(1):497–502, 2008.

JH Bowers, GC Papavizas, SA Johnston, et al. Effect of soil temperature and soil-water matric potential on the survival of phytophthora capsici in natural soil. *Plant disease*, 74 (10):771–777, 1990.

Adam R Boyko, Scott H Williamson, Amit R Indap, Jeremiah D Degenhardt, Ryan D Hernandez, Kirk E. Lohmueller, Mark D. Adams, Steffen Schmidt, John J. Sninsky, Shamil R. Sunyaev, Thomas J. White, Rasmus Nielsen, Andrew G. Clark, and Carlos D. Bustamante. Assessing the evolutionary impact of amino acid mutations in the human genome. *PLoS Genetics*, 4(5):1000083, 2008a. ISSN 15537390. doi:10.1371/journal.pgen.1000083. URL www.plosgenetics.org.

Adam R Boyko et al. Assessing the evolutionary impact of amino acid mutations in the human genome. *PLoS Genetics*, 4(5):1000083, 2008b.

Evan A. Boyle, Yang I. Li, and Jonathan K. Pritchard. An expanded view of complex traits: From polygenic to omnigenic. *Cell*, 169(7):1177–1186, 2017a. ISSN 0092-8674. doi:https://doi.org/10.1016/j.cell.2017.05.038. URL https://www.sciencedirect.com/science/article/pii/S0092867417306293.

Evan A Boyle, Yang I Li, and Jonathan K Pritchard. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell*, 169(7):1177–1186, 2017b. ISSN 10974172. doi:10.1016/j.cell.2017.05.038. URL http://dx.doi.org/10.1016/j.cell.2017.05.038.

Brendan Bulik-Sullivan, Po Ru Loh, Hilary K. Finucane, Stephan Ripke, Jian Yang, Nick Patterson, Mark J. Daly, Alkes L. Price, Benjamin M. Neale, Aiden Corvin, James T.R. Walters, Kai How Farh, Peter A. Holmans, Phil Lee, David A. Collier, Hailiang Huang, Tune H. Pers, Ingrid Agartz, Esben Agerbo, Margot Albus, Madeline Alexander, Farooq Amin, Silviu A. Bacanu, Martin Begemann, Richard A. Belliveau, Judit Bene, Sarah E. Bergen, Elizabeth Bevilacqua, Tim B. Bigdeli, Donald W. Black, Richard Bruggeman, Nancy G. Buccola, Randy L. Buckner, William Byerley, Wiepke Cahn, Guiqing Cai, Murray J. Cairns, Dominique Campion, Rita M. Cantor, Vaughan J. Carr, Noa Carrera, Stanley V. Catts, Kimberly D. Chambert, Raymond C.K. Chan, Ronald Y.L. Chen, Eric Y.H. Chen, Wei Cheng, Eric F.C. Cheung, Siow Ann Chong, C. Robert Cloninger, David Cohen, Nadine Cohen, Paul Cormican, Nick Craddock, Benedicto Crespo-Facorro, James J. Crowley, David Curtis, Michael Davidson, Kenneth L. Davis, Franziska Degenhardt, Jurgen Del Favero, Lynn E. DeLisi, Ditte Demontis, Dimitris Dikeos, Timothy Dinan, Srdjan Djurovic, Gary Donohoe, Elodie Drapeau, Jubao Duan, Frank Dudbridge, Naser Durmishi, Peter Eichhammer, Johan Eriksson, Valentina Escott-Price, Lau-

rent Essioux, Ayman H. Fanous, Martilias S. Farrell, Josef Frank, Lude Franke, Robert Freedman, Nelson B. Freimer, Marion Friedl, Joseph I. Friedman, Menachem Fromer, Giulio Genovese, Lyudmila Georgieva, Elliot S. Gershon, Ina Giegling, Paola Giusti-Rodríguez, Stephanie Godard, Jacqueline I. Goldstein, Vera Golimbet, Srihari Gopal, Jacob Gratten, Lieuwe De Haan, Christian Hammer, Marian L. Hamshere, Mark Hansen, Thomas Hansen, Vahram Haroutunian, Annette M. Hartmann, Frans A. Henskens, Stefan Herms, Joel N. Hirschhorn, Per Hoffmann, Andrea Hofman, Mads V. Hollegaard, David M. Hougaard, Masashi Ikeda, Inge Joa, Antonio Juliá, René S. Kahn, Luba Kalaydjieva, Sena Karachanak-Yankova, Juha Karjalainen, David Kavanagh, Matthew C. Keller, Brian J. Kelly, James L. Kennedy, Andrey Khrunin, Yunjung Kim, Janis Klovins, James A. Knowles, Bettina Konte, Vaidutis Kucinskas, Zita Ausrele Kucinskiene, Hana Kuzelova-Ptackova, Anna K. Kähler, Claudine Laurent, Jimmy Lee Chee Keong, S. Hong Lee, Sophie E. Legge, Bernard Lerer, Miaoxin Li, Tao Li, Kung Yee Liang, Jeffrey Lieberman, Svetlana Limborska, Carmel M. Loughland, Jan Lubinski, Jouko Lönnqvist, Milan Macek, Patrik K.E. Magnusson, Brion S. Maher, Wolfgang Maier, Jacques Mallet, Sara Marsal, Manuel Mattheisen, Morten Mattingsdal, Robert W. McCarley, Colm McDonald, Andrew M. McIntosh, Sandra Meier, Carin J. Meijer, Bela Melegh, Ingrid Melle, Raquelle I. Mesholam-Gately, Andres Metspalu, Patricia T. Michie, Lili Milani, Vihra Milanova, Younes Mokrab, Derek W. Morris, Ole Mors, Kieran C. Murphy, Robin M. Murray, Inez Myin-Germeys, Bertram Müller-Myhsok, Mari Nelis, Igor Nenadic, Deborah A. Nertney, Gerald Nestadt, Kristin K. Nicodemus, Liene Nikitina-Zake, Laura Nisenbaum, Annelie Nordin, Eadbhard O'Callaghan, Colm O'Dushlaine, F. Anthony O'Neill, Sang Yun Oh, Ann Olincy, Line Olsen, Jim Van Os, Christos Pantelis, George N. Papadimitriou, Sergi Papiol, Elena Parkhomenko, Michele T. Pato, Tiina Paunio, Milica Pejovic-Milovancevic, Diana O. Perkins, Olli Pietiläinen, Jonathan Pimm, Andrew J. Pocklington, John Powell, Ann E. Pulver, Shaun M. Purcell, Digby Quested, Henrik B. Rasmussen, Abraham Reichenberg, Mark A. Reimers, Alexander L. Richards, Joshua L. Roffman, Panos Roussos, Douglas M. Ruderfer, Veikko Salomaa, Alan R. Sanders, Ulrich Schall, Christian R. Schubert, Thomas G. Schulze, Sibylle G. Schwab, Edward M. Scolnick, Rodney J. Scott, Larry J. Seidman, Jianxin Shi, Engilbert Sigurdsson, Teimuraz Silagadze, Jeremy M. Silverman, Kang Sim, Petr Slominsky, Jordan W. Smoller, Hon Cheong So, Chris C.A. Spencer, Eli A. Stahl, Hreinn Stefansson, Stacy Steinberg, Elisabeth Stogmann, Richard E. Straub, Eric Strengman, Jana Strohmaier, T. Scott Stroup, Mythily Subramaniam, Jaana Suvisaari, Dragan M. Svrakic, Jin P. Szatkiewicz, Erik Söderman, Srinivas Thirumalai, Draga Toncheva, Paul A. Tooney, Sarah Tosato, Juha Veijola, John Waddington, Dermot Walsh, Dai Wang, Qiang Wang, Bradley T. Webb, Mark Weiser, Dieter B. Wildenauer, Nigel M. Williams, Stephanie Williams, Stephanie H. Witt, Aaron R. Wolen, Emily H.M. Wong, Brandon K. Wormley, Jing Qin Wu, Hualin Simon Xi, Clement C. Zai, Xuebin Zheng, Fritz Zimprich, Naomi R. Wray, Kari Stefansson, Peter M. Visscher, Rolf Adolfsson, Ole A. Andreassen, Douglas H.R. Blackwood, Elvira Bramon, Joseph D. Buxbaum, Anders D. Børglum, Sven Cichon, Ariel Darvasi, Enrico Domenici, Hannelore Ehrenreich, Tõnu Esko, Pablo V. Gejman, Michael Gill, Hugh Gurling, Christina M. Hultman, Nakao Iwata, Assen V. Jablensky, Erik G. Jönsson, Kenneth S. Kendler, George Kirov,

Jo Knight, Todd Lencz, Douglas F. Levinson, Qingqin S. Li, Jianjun Liu, Anil K. Malhotra, Steven A. McCarroll, Andrew McQuillin, Jennifer L. Moran, Preben B. Mortensen, Bryan J. Mowry, Markus M. Nöthen, Roel A. Ophoff, Michael J. Owen, Aarno Palotie, Carlos N. Pato, Tracey L. Petryshen, Danielle Posthuma, Marcella Rietschel, Brien P. Riley, Dan Rujescu, Pak C. Sham, Pamela Sklar, David St Clair, Daniel R. Weinberger, Jens R. Wendland, Thomas Werge, Patrick F. Sullivan, and Michael C. O'Donovan. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics*, 47(3):291–295, feb 2015. ISSN 15461718. doi:10.1038/ng.3211. URL `https://www.nature.com/articles/ng.3211`.

Clare Bycroft, Colin Freeman, Desislava Petkova, Gavin Band, Lloyd T Elliott, Kevin Sharp, Allan Motyer, Damjan Vukcevic, Olivier Delaneau, Jared O'Connell, Adrian Cortes, Samantha Welsh, Alan Young, Mark Effingham, Gil McVean, Stephen Leslie, Naomi Allen, Peter Donnelly, and Jonathan Marchini. The UK Biobank resource with deep phenotyping and genomic data. *Nature*, 562(7726):203–209, 2018. ISSN 14764687. doi:10.1038/s41586-018-0579-z. URL `https://doi.org/10.1038/s41586-018-0579-z`.

Christopher S. Carlson, Tara C. Matise, Kari E. North, Christopher A. Haiman, Megan D. Fesinmeyer, Steven Buyske, Fredrick R. Schumacher, Ulrike Peters, Nora Franceschini, Marylyn D. Ritchie, David J. Duggan, Kylee L. Spencer, Logan Dumitrescu, Charles B. Eaton, Fridtjof Thomas, Alicia Young, Cara Carty, Gerardo Heiss, Loic Le Marchand, Dana C. Crawford, Lucia A. Hindorff, and Charles L. Kooperberg. Generalization and Dilution of Association Results from European GWAS in Populations of Non-European Ancestry: The PAGE Study. *PLoS Biology*, 11(9):e1001661, 2013. ISSN 15457885. doi:10.1371/journal.pbio.1001661. URL `https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1001661`.

Maryn O. Carlson, Elodie Gazave, Michael A. Gore, and Christine D. Smart. Temporal genetic dynamics of an experimental, biparental field population of phytophthora capsici. *Frontiers in Genetics*, 8, 2017. ISSN 1664-8021. doi:10.3389/fgene.2017.00026. URL `https://www.frontiersin.org/articles/10.3389/fgene.2017.00026`.

Maryn O. Carlson, Daniel P. Rice, Jeremy J. Berg, and Matthias Steinrücken. Polygenic score accuracy in ancient samples: Quantifying the effects of allelic turnover. *PLoS Genetics*, 18 (5):e1010170, may 2022. ISSN 15537404. doi:10.1371/journal.pgen.1010170. URL `https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010170`.

W. K. Cavenee, T. P. Dryja, R. A. Phillips, W. F. Benedict, R. Godbout, B. L. Gallie, A. L. Murphree, L. C. Strong, and R. L. White. Expression of recessive alleles by chromosomal mechanisms in retinoblastoma. *Nature*, 305(5937):779–784, 1983. ISSN 00280836. doi:10.1038/305779a0. URL `https://www.nature.com/articles/305779a0`.

Ranajit Chakraborty and K. M. Weiss. Admixture as a tool for finding linked genes and detecting that difference from allelic association between loci. *Proceedings of the National Academy of Sciences*, 85(23):9119–9123, 1988.

Jureerat Chamnanpunt, Wei Xing Shan, and Brett M. Tyler. High frequency mitotic gene conversion in genetic hybrids of the oomycete Phytophthora sojae. *Proceedings of the National Academy of Sciences of the United States of America*, 98(25):14530–14535, dec 2001. ISSN 00278424. doi:10.1073/pnas.251464498. URL https://www.pnas.org/doi/abs/10.1073/pnas.251464498.

Yingleong Chan, Elaine T. Lim, Niina Sandholm, Sophie R. Wang, Amy Jayne McKnight, Stephan Ripke, Mark J. Daly, Benjamin M. Neale, Rany M. Salem, and Joel N. Hirschhorn. An excess of risk-increasing low-frequency variants can be a signal of polygenic inheritance in complex diseases. *American Journal of Human Genetics*, 94(3):437–452, mar 2014. ISSN 00029297. doi:10.1016/j.ajhg.2014.02.006. URL http://dx.doi.org/10.1016/j.ajhg.2014.02.006.

B. Charlesworth, M T Morgan, and D Charlesworth. The effect of deleterious mutations on neutral molecular variation, 1993. ISSN 00166731. URL https://academic.oup.com/genetics/article/134/4/1289/6011194.

Luis Miguel Chevin and Frédéric Hospital. Selective sweep at a quantitative trait locus in the presence of background genetic variation. *Genetics*, 180(3):1645–1660, 2008. ISSN 00166731. doi:10.1534/genetics.108.093351.

Alec J Coffman, Ping Hsun Hsieh, Simon Gravel, and Ryan N Gutenkunst. Computationally efficient composite likelihood statistics for demographic inference. *Molecular Biology and Evolution*, 33(2):591–593, 2016. ISSN 15371719. doi:10.1093/molbev/msv255. URL https://academic.oup.com/mbe/article-abstract/33/2/591/2579696.

Laura L Colbran, Eric R Gamazon, Dan Zhou, Patrick Evans, Nancy J Cox, and John A Capra. Inferred divergent gene regulation in archaic hominins reveals potential phenotypic differences. *Nature Ecology and Evolution*, 3(11):1598–1606, 2019. ISSN 2397334X. doi:10.1038/s41559-019-0996-x. URL https://doi.org/10.1038/s41559-019-0996-x.

Graham Coop, Joseph K Pickrell, John Novembre, Sridhar Kudaravalli, Jun Li, Devin Absher, Richard M. Myers, Luigi Luca Cavalli-Sforza, Marcus W. Feldman, and Jonathan K. Pritchard. The role of geography in human adaptation. *PLoS Genetics*, 5(6):1000500, 2009. ISSN 15537390. doi:10.1371/journal.pgen.1000500.

Samantha L Cox, Christopher B Ruff, Robert M Maier, and Iain Mathieson. Genetic contributions to variation in human stature in prehistoric Europe. *Proceedings of the National Academy of Sciences of the United States of America*, 116(43):21484–21492, 2019. ISSN 10916490. doi:10.1073/pnas.1910606116. URL www.pnas.org/cgi/doi/10.1073/pnas.1910606116.

Samantha L Cox, Hannah Moots, Jay T Stock, Andrej Shbat, Bárbara D Bitarello, Wolfgang Haak, Eva Rosenstock, Christopher B Ruff, and Iain Mathieson. Predicting skeletal stature using ancient DNA. *bioRxiv*, page 2021.03.31.437877, 2021. doi:10.1101/2021.03.31.437877. URL https://doi.org/10.1101/2021.03.

31.437877http://biorxiv.org/content/early/2021/03/31/2021.03.31.437877.abstract.

Hans D Daetwyler, Beatriz Villanueva, and John A Woolliams. Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS ONE*, 3(10), 2008. ISSN 19326203. doi:10.1371/journal.pone.0003395. URL www.plosone.org.

Angela L. Dale, Nicolas Feau, Sydney E. Everhart, Braham Dhillon, Barbara Wong, Julie Sheppard, Guillaume J. Bilodeau, Avneet Brar, Javier F. Tabima, Danyu Shen, Clive M. Brasier, Brett M. Tyler, Niklaus J. Grünwald, and Richard C. Hamelin. Mitotic recombination and rapid genome evolution in the invasive forest pathogen phytophthora ramorum. *mBio*, 10(2), mar 2019. ISSN 21507511. doi:10.1128/mBio.02452-18. URL https://journals.asm.org/doi/10.1128/mBio.02452-18.

Petr Danecek, Adam Auton, Goncalo Abecasis, Cornelis A. Albers, Eric Banks, Mark A. DePristo, Robert E. Handsaker, Gerton Lunter, Gabor T. Marth, Stephen T. Sherry, Gilean McVean, Richard Durbin, and 1000 Genomes Project Analysis Group. The variant call format and VCFtools. *Bioinformatics*, 27(15):2156–2158, 06 2011. ISSN 1367-4803. doi:10.1093/bioinformatics/btr330. URL https://doi.org/10.1093/bioinformatics/btr330.

Gustavo de los Campos, John M. Hickey, Ricardo Pong-Wong, Hans D. Daetwyler, and Mario P.L. Calus. Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics*, 193(2):327–345, feb 2013a. ISSN 00166731. doi:10.1534/genetics.112.143313.

Gustavo de los Campos, Ana I Vazquez, Rohan Fernando, Yann C Klimentidis, and Daniel Sorensen. Prediction of Complex Human Traits Using the Genomic Best Linear Unbiased Predictor. *PLoS Genetics*, 9(7), 2013b. ISSN 15537390. doi:10.1371/journal.pgen.1003608. URL www.plosgenetics.org.

A. P.W. De Roos, B J Hayes, R J Spelman, and M E Goddard. Linkage disequilibrium and persistence of phase in Holstein-Friesian, Jersey and Angus cattle. *Genetics*, 179(3): 1503–1512, 2008. ISSN 00166731. doi:10.1534/genetics.107.084301.

A. P.W. De Roos, B J Hayes, and M E Goddard. Reliability of genomic predictions across multiple populations. *Genetics*, 183(4):1545–1553, 2009. ISSN 00166731. doi:10.1534/genetics.109.104935. URL https://academic.oup.com/genetics/article/183/4/1545/6063064.

L Duncan, H Shen, B Gelaye, J Meijsen, K Ressler, M Feldman, R Peterson, and B Domingue. Analysis of polygenic risk score usage and performance in diverse human populations. *Nature Communications*, 10(1), 2019. ISSN 20411723. doi:10.1038/s41467-019-11112-0. URL https://doi.org/10.1038/s41467-019-11112-0.

Amara R. Dunn, Stephen R. Bruening, Niklaus J. Grünwald, and Christine D. Smart. Evolution of an experimental population of phytophthora capsici in the field. *Phytopathology*®,

104(10):1107–1117, 2014. doi:10.1094/PHYTO-12-13-0346-R. URL https://doi.org/10.1094/PHYTO-12-13-0346-R. PMID: 24702666.

Richard Durrett. *Probability Models for DNA Sequence Evolution*. Springer-Verlag, New York, 2 edition, 2008. ISBN 978-0-387-78168-6. doi:10.1007/978-0-387-78168-6.

Arun Durvasula and Kirk E. Lohmueller. Negative selection on complex traits limits phenotype prediction accuracy between populations. *American Journal of Human Genetics*, 108(4):620–631, apr 2021. ISSN 15376605. doi:10.1016/j.ajhg.2021.02.013.

Michael D. Edge and Graham Coop. Reconstructing the history of polygenic scores using coalescent trees. *Genetics*, 211:235–262, January 2019.

Robert J. Elshire, Jeffrey C. Glaubitz, Qi Sun, Jesse A. Poland, Ken Kawamoto, Edward S. Buckler, and Sharon E. Mitchell. A robust, simple genotyping-by-sequencing (gbs) approach for high diversity species. *PLOS ONE*, 6(5):1–10, 05 2011. doi:10.1371/journal.pone.0019379. URL https://doi.org/10.1371/journal.pone.0019379.

Eyal Elyashiv, Shmuel Sattath, Tina T. Hu, Alon Strutsovsky, Graham McVicker, Peter Andolfatto, Graham Coop, and Guy Sella. A Genomic Map of the Effects of Linked Selection in Drosophila. *PLoS Genetics*, 12(8):e1006130, aug 2016. ISSN 15537404. doi:10.1371/journal.pgen.1006130. URL https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1006130.

Iuliana V. Ene, Rhys A. Farrer, Matthew P. Hirakawa, Kennedy Agwamba, Christina A. Cuomo, and Richard J. Bennett. Global analysis of mutations driving microevolution of a heterozygous diploid fungal pathogen. *Proceedings of the National Academy of Sciences of the United States of America*, 115(37):E8688–E8697, 2018. ISSN 10916490. doi:10.1073/pnas.1806002115.

M. Erbe, B. J. Hayes, L. K. Matukumalli, S. Goswami, P. J. Bowman, C. M. Reich, B. A. Mason, and M. E. Goddard. Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. *Journal of Dairy Science*, 95(7):4114–4129, jul 2012. ISSN 00220302. doi:10.3168/jds.2011-5019.

Alison Etheridge. *Some Mathematical Models from Population Genetics: École d'Été de Probabilités de Saint-Flour XXXIX-2009*. Springer Heidelberg, 2012.

Warren J Ewens. *Mathematical Population Genetics I: Theoretical Introduction*. Springer-Verlag, New York, 2004.

L Excoffier, T Hofer, and M Foll. Detecting loci under selection in a hierarchically structured population. *Heredity*, 103(4):285–298, 2009. ISSN 0018067X. doi:10.1038/hdy.2009.74. URL www.nature.com/hdy.

Adam Eyre-Walker. Genetic architecture of a complex trait and its implications for fitness and genome-wide association studies. *Proceedings of the National Academy of Sciences of the United States of America*, 107(SUPPL. 1):1752–1756, jan 2010. ISSN 10916490. doi:10.1073/pnas.0906182107. URL www.pnas.org/cgi/doi/10.1073/pnas.0906182107.

Adam Eyre-Walker, Megan Woolfit, and Ted Phelps. The distribution of fitness effects of new deleterious amino acid mutations in humans. *Genetics*, 173(2):891–900, 2006. ISSN 00166731. doi:10.1534/genetics.106.057570. URL https://academic.oup.com/genetics/article/173/2/891/6061608.

Shaohua Fan, Matthew E B Hansen, Yancy Lo, and Sarah A Tishkoff. Going global by adapting local: A review of recent human adaptation. *Science*, 354(6308):54–59, 2016.

Yair Field et al. Detection of human adaptation during the past 2000 years. *Science*, 354 (6313):760–764, 2016.

Hilary K. Finucane, Brendan Bulik-Sullivan, Alexander Gusev, Gosia Trynka, Yakir Reshef, Po Ru Loh, Verneri Anttila, Han Xu, Chongzhi Zang, Kyle Farh, Stephan Ripke, Felix R. Day, Shaun Purcell, Eli Stahl, Sara Lindstrom, John R.B. Perry, Yukinori Okada, Soumya Raychaudhuri, Mark J. Daly, Nick Patterson, Benjamin M. Neale, and Alkes L. Price. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nature Genetics*, 47(11):1228–1235, sep 2015. ISSN 15461718. doi:10.1038/ng.3404. URL https://www.nature.com/articles/ng.3404.

Eric Friedlander and Matthias Steinrücken. A numerical framework for genetic hitchhiking in populations of variable size. *bioRxiv*, page 2021.03.25.437048, 2021. doi:10.1101/2021.03.25.437048.

Matteo Fumagalli, Manuela Sironi, Uberto Pozzoli, Anna Ferrer-Admettla, Linda Pattini, and Rasmus Nielsen. Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution. *PLoS Genetics*, 7(11):1002355, 2011. ISSN 15537390. doi:10.1371/journal.pgen.1002355.

Kevin J. Galinsky, Yakir A. Reshef, Hilary K. Finucane, Po Ru Loh, Noah Zaitlen, Nick J. Patterson, Brielin C. Brown, and Alkes L. Price. Estimating cross-population genetic correlations of causal effect sizes. *Genetic Epidemiology*, 43(2):180–188, mar 2019. ISSN 10982272. doi:10.1002/gepi.22173.

Steven Gazal, Hilary K Finucane, Nicholas A Furlotte, Po Ru Loh, Pier Francesco Palamara, Xuanyao Liu, Armin Schoech, Brendan Bulik-Sullivan, Benjamin M Neale, Alexander Gusev, and Alkes L Price. Linkage disequilibrium-dependent architecture of human complex traits shows action of negative selection. *Nature Genetics*, 49(10):1421–1427, 2017. ISSN 15461718. doi:10.1038/ng.3954.

A. C. Gerstein, A. Kuzmin, and S. P. Otto. Loss-of-heterozygosity facilitates passage through Haldane's sieve for *Saccharomyces cerevisiae* undergoing adaptation. *Nature Communications*, 5, 2014. ISSN 20411723. doi:10.1038/ncomms4819.

Benjamin H Good. Linkage disequilibrium between rare mutations. *bioRxiv*, 2021. doi:10.1101/2020.12.10.420042. URL https://doi.org/10.1101/2020.12.10.420042.

Leah L. Granke, Lina Quesada-Ocampo, Kurt Lamour, and Mary K. Hausbeck. Advances in research on Phytophthora capsici on vegetable crops in the United States. *Plant Disease*, 96(11):1588–1600, oct 2012. ISSN 01912917. doi:10.1094/PDIS-02-12-0211-FE. URL https://apsjournals.apsnet.org/doi/10.1094/PDIS-02-12-0211-FE.

Robert C Griffiths and Dario Spano. Diffusion processes and coalescent trees. *arXiv*, 2010. URL http://arxiv.org/abs/1003.4650.

Zhigang Guo, Dominic M. Tucker, Christopher J. Basten, Harish Gandhi, Elhan Ersoz, Baohong Guo, Zhanyou Xu, Daolong Wang, and Gilles Gay. The impact of population structure on genomic prediction in stratified populations. *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik*, 127(3):749–762, jan 2014. ISSN 14322242. doi:10.1007/s00122-013-2255-x. URL https://link.springer.com/article/10.1007/s00122-013-2255-x.

Ryan N Gutenkunst, Ryan D Hernandez, Scott H Williamson, and Carlos D Bustamante. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genetics*, 5(10):1000695, 2009a. ISSN 15537390. doi:10.1371/journal.pgen.1000695. URL www.plosgenetics.org.

Ryan N. Gutenkunst, Ryan D. Hernandez, Scott H. Williamson, and Carlos D. Bustamante. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genetics*, 5(10):e1000695, oct 2009b. ISSN 15537390. doi:10.1371/journal.pgen.1000695. URL https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1000695.

D. Habier, R. L. Fernando, and J. C.M. Dekkers. The impact of genetic relationship information on genome-assisted breeding values. *Genetics*, 177(4):2389–2397, 2007. ISSN 00166731. doi:10.1534/genetics.107.081190.

Benjamin C Haller and Philipp W Messer. SLiM 3: Forward Genetic Simulations Beyond the Wright-Fisher Model. *Molecular Biology and Evolution*, 36(3):632–637, 2019. ISSN 15371719. doi:10.1093/molbev/msy228. URL https://academic.oup.com/mbe/article-abstract/36/3/632/5229931.

Martha T Hamblin, Edward S Buckler, and Jean Luc Jannink. Population genetics of genomics-based crop improvement methods. *Trends in Genetics*, 27(3):98–106, 2011. ISSN 01689525. doi:10.1016/j.tig.2010.12.003.

Arbel Harpak and Molly Przeworski. The evolution of group differences in changing environments. *PLoS Biology*, 19(1), jan 2021. ISSN 15457885. doi:10.1371/journal.pbio.3001072.

Kelley Harris and Rasmus Nielsen. The genetic cost of neanderthal introgression. *Genetics*, 203(2):881–891, 2016.

Mary K. Hausbeck and Kurt H. Lamour. Phytophthora capsici on vegetable crops: Research progress and management challenges, feb 2004. ISSN 01912917. URL https://apsjournals.apsnet.org/doi/10.1094/PDIS.2004.88.12.1292.

Laura Katharine Hayward and Guy Sella. Polygenic adaptation after a sudden change in environment. *eLife*, 11:e66697, sep 2022. ISSN 2050-084X. doi:10.7554/eLife.66697. URL https://doi.org/10.7554/eLife.66697.

Elliot L. Heffner, Aaron J. Lorenz, Jean Luc Jannink, and Mark E. Sorrells. Plant breeding with Genomic selection: Gain per unit time and cost. *Crop Science*, 50 (5):1681–1690, sep 2010. ISSN 0011183X. doi:10.2135/cropsci2009.11.0662. URL https://onlinelibrary.wiley.com/doi/full/10.2135/cropsci2009.11.0662https://onlinelibrary.wiley.com/doi/abs/10.2135/cropsci2009.11.0662https://acsess.onlinelibrary.wiley.com/doi/10.2135/cropsci2009.11.0662.

Farhad Hormozdiari, Steven Gazal, Bryce Van De Geijn, Hilary K Finucane, Chelsea J.T. Ju, Po Ru Loh, Armin Schoech, Yakir Reshef, Xuanyao Liu, Luke O'connor, Alexander Gusev, Eleazar Eskin, and Alkes L Price. Leveraging molecular quantitative trait loci to understand the genetic architecture of diseases and complex traits. *Nature Genetics*, 50(7):1041–1047, 2018. ISSN 15461718. doi:10.1038/s41588-018-0148-2. URL https://doi.org/10.1038/s41588-018-0148-2.

Jian Hu, Sandesh Shrestha, Yuxin Zhou, Joann Mudge, Xili Liu, and Kurt Lamour. Dynamic extreme aneuploidy (dea) in the vegetable pathogen phytophthora capsici and the potential for rapid asexual evolution. *PLOS ONE*, 15(1):1–15, 01 2020. doi:10.1371/journal.pone.0227250. URL https://doi.org/10.1371/journal.pone.0227250.

Richard R Hudson and N. L. Kaplan. Deleterious background selection with recombination. *Genetics*, 141(4):1605–1617, 1995. ISSN 00166731. doi:10.1093/genetics/141.4.1605. URL https://academic.oup.com/genetics/article/141/4/1605/6061970.

Jon Hulvey, Jacque Young, Ledare Finley, and Kurt Lamour. Loss of heterozygosity in Phytophthora capsici after N-ethyl-nitrosourea mutagenesis. *Mycologia*, 102(1):27–32, jan 2010. ISSN 00275514. doi:10.3852/09-102. URL https://www.tandfonline.com/action/journalInformation?journalCode=umyc20.

Julio Isidro, Jean Luc Jannink, Deniz Akdemir, Jesse Poland, Nicolas Heslot, and Mark E. Sorrells. Training set optimization under population structure in genomic selection. *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik*, 128(1):145–158, jan 2015. ISSN 14322242. doi:10.1007/s00122-014-2418-4. URL https://link.springer.com/article/10.1007/s00122-014-2418-4.

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

Milo S. Johnson, Shreyas Gopalakrishnan, Juhee Goyal, Megan E. Dillingham, Christopher W. Bakerlee, Parris T. Humphrey, Tanush Jagdish, Elizabeth R. Jerison, Katya Kosheleva, Katherine R. Lawrence, Jiseon Min, Alief Moulana, Angela M. Phillips, Julia C. Piper, Ramya Purkanti, Artur Rego-Costa, Michael J. McDonald, Alex N.Nguyen Ba, and Michael M. Desai. Phenotypic and molecular evolution across 10,000 generations in laboratory budding yeast populations. *eLife*, 10:1–28, 2021. ISSN 2050084X. doi:10.7554/eLife.63910.

Parul Johri, Brian Charlesworth, and Jeffrey D. Jensen. Toward an evolutionarily appropriate null model: Jointly inferring demography and purifying selection. *Genetics*, 215(1):173–192, may 2020. ISSN 19432631. doi:10.1534/genetics.119.303002. URL `https://academic.oup.com/genetics/article/215/1/173/5930426`.

Julien Jouganous, Will Long, Aaron P Ragsdale, and Simon Gravel. Inferring the joint demographic history of multiple populations: Beyond the diffusion approximation. *Genetics*, 206(3):1549–1567, 2017. ISSN 19432631. doi:10.1534/genetics.117.200493. URL `https://doi.org/10.1534/genetics.117.200493`.

Masahiro Kanai, Masato Akiyama, Atsushi Takahashi, Nana Matoba, Yukihide Momozawa, Masashi Ikeda, Nakao Iwata, Shiro Ikegawa, Makoto Hirata, Koichi Matsuda, Michiaki Kubo, Yukinori Okada, and Yoichiro Kamatani. Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nature Genetics*, 50(3):390–400, feb 2018. ISSN 15461718. doi:10.1038/s41588-018-0047-6. URL `https://www.nature.com/articles/s41588-018-0047-6`.

S Karlin and H Taylor. *A Second Course in Stochastic Processes*. Academic Press, San Diego, 1981.

Jerome Kelleher, Alison M Etheridge, and Gilean McVean. Efficient Coalescent Simulation and Genealogical Analysis for Large Sample Sizes. *PLoS Computational Biology*, 12: 1004842, 2016.

Bernard Y. Kim, Christian D. Huber, and Kirk E. Lohmueller. Inference of the distribution of selection coefficients for new nonsynonymous mutations using large samples. *Genetics*, 206(1):345–361, may 2017. ISSN 19432631. doi:10.1534/genetics.116.197145. URL `https://academic.oup.com/genetics/article/206/1/345/6064197`.

Yuseob Kim and Wolfgang Stephan. Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genetics*, 160(2):765–777, 2002.

Motoo Kimura. Solution of a Process of Random Genetic Drift with a Continuous Model. *Proceedings of the National Academy of Sciences*, 41(3):144–150, 1955.

Evan M. Koch and Shamil R. Sunyaev. Maintenance of Complex Trait Variation: Classic Theory and Modern Data. *Frontiers in Genetics*, 12(November):1–14, 2021. ISSN 16648021. doi:10.3389/fgene.2021.763363.

K. H. Lamour and M. K. Hausbeck. Effect of crop rotation on the survival of Phytophthora capsici in Michigan. *Plant Disease*, 87(7):841–845, feb 2003. ISSN 01912917. doi:10.1094/PDIS.2003.87.7.841. URL https://apsjournals.apsnet.org/doi/10.1094/PDIS.2003.87.7.841.

Kurt H. Lamour, Joann Mudge, Daniel Gobena, Oscar P. Hurtado-Gonzales, Jeremy Schmutz, Alan Kuo, Neil A. Miller, Brandon J. Rice, Sylvain Raffaele, Liliana M. Cano, Arvind K. Bharti, Ryan S. Donahoo, Sabra Finley, Edgar Huitema, Jon Hulvey, Darren Platt, Asaf Salamov, Alon Savidor, Rahul Sharma, Remco Stam, Dylan Storey, Marco Thines, Joe Win, Brian J. Haas, Darrell L. Dinwiddie, Jerry Jenkins, James R. Knight, Jason P. Affourtit, Cliff S. Han, Olga Chertkov, Erika A. Lindquist, Chris Detter, Igor V. Grigoriev, Sophien Kamoun, and Stephen F. Kingsmore. Genome sequencing and mapping reveal loss of heterozygosity as a mechanism for rapid adaptation in the vegetable pathogen phytophthora capsici. *Molecular Plant-Microbe Interactions®*, 25(10): 1350–1360, 2012. doi:10.1094/MPMI-02-12-0028-R. URL https://doi.org/10.1094/MPMI-02-12-0028-R. PMID: 22712506.

Iosif Lazaridis, Dani Nadel, Gary Rollefson, Deborah C Merrett, Nadin Rohland, Swapan Mallick, Daniel Fernandes, Mario Novak, Beatriz Gamarra, Kendra Sirak, Sarah Connell, Kristin Stewardson, Eadaoin Harney, Qiaomei Fu, Gloria Gonzalez-Fortes, Eppie R Jones, Songül Alpaslan Roodenberg, György Lengyel, Fanny Bocquentin, Boris Gasparian, Janet M Monge, Michael Gregg, Vered Eshed, Ahuva Sivan Mizrahi, Christopher Meiklejohn, Fokke Gerritsen, Luminita Bejenaru, Matthias Blüher, Archie Campbell, Gianpiero Cavalleri, David Comas, Philippe Froguel, Edmund Gilbert, Shona M Kerr, Peter Kovacs, Johannes Krause, Darren McGettigan, Michael Merrigan, D. Andrew Merriwether, Seamus O'Reilly, Martin B Richards, Ornella Semino, Michel Shamoon-Pour, Gheorghe Stefanescu, Michael Stumvoll, Anke Tönjes, Antonio Torroni, James F Wilson, Loic Yengo, Nelli A Hovhannisyan, Nick Patterson, Ron Pinhasi, and David Reich. Genomic insights into the origin of farming in the ancient Near East. *Nature*, 536(7617):419–424, 2016. ISSN 14764687. doi:10.1038/nature19310. URL https://www-nature-com.proxy.uchicago.edu/articles/nature19310.pdf.

Han Li and Peter Ralph. Local PCA Shows How the Effect of Population Structure Differs Along the Genome. *Genetics*, *Early online*, 2018. doi:10.1534/genetics.118.301747.

Heng Li and Richard Durbin. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, 25(14):1754–1760, 05 2009. ISSN 1367-4803. doi:10.1093/bioinformatics/btp324. URL https://doi.org/10.1093/bioinformatics/btp324.

Yichen Liu, Xiaowei Mao, Johannes Krause, and Qiaomei Fu. Insights into human history from the first decade of ancient human genomics. *Science*, 373(6562):1479–1484, sep 2021. ISSN 10959203. doi:10.1126/science.abi8202. URL https://www-science-org.proxy.uchicago.edu/doi/abs/10.1126/science.abi8202.

Po-Ru Loh, Gleb Kichaev, Steven Gazal, Armin P. Schoech, and Alkes L. Price. Mixed-model association for biobank-scale datasets. *Nature Genetics*, 50(7):906–908, 2018. doi:10.1038/s41588-018-0144-6.

A. J. Lorenz, K. P. Smith, and J. L. Jannink. Potential and optimization of genomic selection for Fusarium head blight resistance in six-row barley. *Crop Science*, 52(4):1609–1621, jul 2012. ISSN 0011183X. doi:10.2135/cropsci2011.09.0503.

Sergio Lukic and Jody Hey. Demographic inference using spectral methods on SNP data, with an analysis of the human out-of-Africa expansion. *Genetics*, 192(2):619–639, 2012. ISSN 19432631. doi:10.1534/genetics.112.141846. URL `http://www.genetics.org/lookup/suppl/`.

Sergio Lukić, Jody Hey, and Kevin Chen. Non-equilibrium allele frequency spectra via spectral methods. *Theoretical Population Biology*, 79(4):203–219, 2011.

Michael Lynch and Bruce Walsh. *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, 1st edition edition, 1998.

Jianzhong Ma and Christopher I Amos. Investigation of inversion polymorphisms in the human genome using principal components analysis. *PLoS ONE*, 7(7):40224, 2012.

Anna Sapfo Malaspinas, Orestis Malaspinas, Steven N Evans, and Montgomery Slatkin. Estimating allele age and selection coefficient from time-serial data. *Genetics*, 192(2): 599–607, 2012.

Nicholas Mancuso, Nadin Rohland, Kristin A. Rand, Arti Tandon, Alexander Allen, Dominique Quinque, Swapan Mallick, Heng Li, Alex Stram, Xin Sheng, Zsofia Kote-Jarai, Douglas F. Easton, Rosalind A. Eeles, Loic Le Marchand, Alex Lubwama, Daniel Stram, Stephen Watya, David V. Conti, Brian Henderson, Christopher A. Haiman, Bogdan Pasaniuc, and David Reich. The contribution of rare variation to prostate cancer heritability. *Nature Genetics*, 48(1):30–35, nov 2015. ISSN 15461718. doi:10.1038/ng.3446. URL `https://www.nature.com/articles/ng.3446`.

Mohammad A. Mandegar and Sarah P. Otto. Mitotic recombination counteracts the benefits of genetic segregation. *Proceedings of the Royal Society B: Biological Sciences*, 274(1615): 1301–1307, 2007. ISSN 14712970. doi:10.1098/rspb.2007.0056.

Teri A Manolio et al. Finding the missing heritability of complex diseases. *Nature*, 461, 2009.

Stephanie Marciniak, Christina M Bergey, Ana Maria Silva, Agata Hałuszko, Mirosław Furmanek, Barbara Veselka, Petr Velemínský, Giuseppe Vercellotti, Joachim Wahl, Gunita ZariÅĘa, Cristina Longhi, Jan KoláÅŹ, Rafael Garrido-Pena, Raúl Flores-Fernández, Ana M Herrero-Corral, Angela Simalcsik, Werner Müller, Alison Sheridan, ŽydrÅńnÄŮ MiliauskienÄŮ, Rimantas Jankauskas, Vyacheslav Moiseyev, Kitti

Köhler, Ágnes Király, Beatriz Gamarra, Olivia Cheronet, Vajk Szeverényi, Viktoria Kiss, Tamás Szeniczey, Krisztián Kiss, Zsuzsanna K Zoffmann, Judit Koós, Magdolna Hellebrandt, László Domboróczki, Cristian Virag, Mario Novak, David Reich, Tamás Hajdu, Noreen von Cramon-Taubadel, Ron Pinhasi, and George H Perry. An integrative skeletal and paleogenomic analysis of prehistoric stature variation suggests relatively reduced health for early European farmers. *bioRxiv*, 53 (9):2021.03.31.437881, 2021. ISSN 1098-6596. doi:10.1101/2021.03.31.437881. URL https://doi.org/10.1101/2021.03.31.437881http://biorxiv.org/content/early/2021/03/31/2021.03.31.437881.abstract.

Alicia R Martin, Masahiro Kanai, Yoichiro Kamatani, Yukinori Okada, Benjamin M Neale, and Mark J Daly. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nature Genetics*, 51(4):584–591, 2019. ISSN 15461718. doi:10.1038/s41588-019-0379-x. URL https://doi.org/10.1038/s41588-019-0379-x.

Alicia R Martin et al. Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. *American Journal of Human Genetics*, 100:635–649, 2017.

Rui Martiniano, Lara M Cassidy, Ros Ó'Maoldúin, Russell McLaughlin, Nuno M Silva, Licinio Manco, Daniel Fidalgo, Tania Pereira, Maria J Coelho, Miguel Serra, Joachim Burger, Rui Parreira, Elena Moran, Antonio C Valera, Eduardo Porfirio, Rui Boaventura, Ana M Silva, and Daniel G Bradley. The population genomics of archaeological transition in west Iberia: Investigation of ancient substructure using imputation and haplotype-based methods. *PLoS Genetics*, 13(7), 2017. ISSN 15537404. doi:10.1371/journal.pgen.1006852. URL https://doi.org/10.1371/journal.pgen.1006852.

Iain Mathieson. Human adaptation over the past 40,000 years. *Current Opinion in Genetics and Development*, 62:97–104, jun 2020. ISSN 18790380. doi:10.1016/j.gde.2020.06.003. URL https://doi.org/10.1016/j.gde.2020.06.003.

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

Iain Mathieson, Iosif Lazaridis, Nadin Rohland, Swapan Mallick, Nick Patterson, Songül Alpaslan Roodenberg, Eadaoin Harney, Kristin Stewardson, Daniel Fernandes, Mario Novak, Kendra Sirak, Cristina Gamba, Eppie R. Jones, Bastien Llamas, Stanislav Dryomov, Joseph Pickrell, Juan Luís Arsuaga, JoséMaría Bermúdez de Castro, Eudald Carbonell, Fokke Gerritsen, Aleksandr Khokhlov, Pavel Kuznetsov, Marina Lozano, Harald Meller, Oleg Mochalov, Vyacheslav Moiseyev, Manuel A. Rojo Guerra, Jacob Roodenberg, Josep Maria Vergès, Johannes Krause, Alan Cooper, Kurt W. Alt, Dorcas Brown, David Anthony, Carles Lalueza-Fox, Wolfgang Haak, Ron Pinhasi, and David Reich. Genome-wide patterns of selection in 230 ancient eurasians. *Nature*, 528(7583):499–503, 2015. doi:10.1038/nature16152. URL https://doi.org/10.1038/nature16152.

Sara Mathieson and Iain Mathieson. FADS1 and the timing of human adaptation to agriculture. *Molecular Biology and Evolution*, 35(12):2957–2970, 2018.

Ruth McQuillan, Anne-Louise Leutenegger, Rehab Abdel-Rahman, Christopher S Franklin, Marijana Pericic, Lovorka Barac-Lauc, Nina Smolej-Narancic, Branka Janicijevic, Ozren Polasek, Albert Tenesa, Andrew K Macleod, Susan M Farrington, Pavao Rudan, Caroline Hayward, Veronique Vitart, Igor Rudan, Sarah H Wild, Malcolm G Dunlop, Alan F Wright, Harry Campbell, and James F Wilson. Runs of homozygosity in european populations. *Am J Hum Genet*, 83(3):359–372, Sep 2008. ISSN 1537-6605 (Electronic); 0002-9297 (Print); 0002-9297 (Linking). doi:10.1016/j.ajhg.2008.08.007.

Gil McVean. A genealogical interpretation of principal components analysis. *PLoS Genetics*, 5(10):1000686, 2009.

Graham McVicker, David Gordon, Colleen Davis, and Phil Green. Widespread genomic signatures of natural selection in hominid evolution. *PLoS Genetics*, 5(5):1000471, 2009. ISSN 15537390. doi:10.1371/journal.pgen.1000471. URL `www.plosgenetics.org`.

T. H. Meuwissen, B. J. Hayes, and M. E. Goddard. Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 157(4):1819–1829, 2001. ISSN 00166731. doi:10.1093/genetics/157.4.1819.

Gerhard Moser, Sang Hong Lee, Ben J. Hayes, Michael E. Goddard, Naomi R. Wray, and Peter M. Visscher. Simultaneous Discovery, Estimation and Prediction Analysis of Complex Traits Using a Bayesian Mixture Model. *PLoS Genetics*, 11(4):e1004969, apr 2015. ISSN 15537404. doi:10.1371/journal.pgen.1004969. URL `https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1004969`.

Hakhamanesh Mostafavi, Arbel Harpak, Ipsita Agarwal, Dalton Conley, Jonathan K. Pritchard, and Molly Przeworski. Variable prediction accuracy of polygenic scores within an ancestry group. *eLife*, 9, jan 2020. ISSN 2050084X. doi:10.7554/eLife.48376.

David A Murphy, Eyal Elyashiv, Guy Amster, and Guy Sella. Broad-scale variation in human genetic diversity levels is predicted by purifying selection on coding and non-coding elements. *eLife*, 11:e76065, 2022. doi:10.7554/eLife.76065.

Vagheesh Narasimhan, Petr Danecek, Aylwyn Scally, Yali Xue, Chris Tyler-Smith, and Richard Durbin. BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data. *Bioinformatics*, 32(11):1749–1751, 01 2016. ISSN 1367-4803. doi:10.1093/bioinformatics/btw044. URL `https://doi.org/10.1093/bioinformatics/btw044`.

Caitlin A. Nichols, William J. Gibson, Meredith S. Brown, Jack A. Kosmicki, John P. Busanovich, Hope Wei, Laura M. Urbanski, Naomi Curimjee, Ashton C. Berger, Galen F. Gao, Andrew D. Cherniack, Sirano Dhe-Paganon, Brenton R. Paolella, and Rameen Beroukhim. Loss of heterozygosity of essential genes represents a widespread class of potential cancer vulnerabilities. *Nature Communications 2020 11:1*, 11(1):1–14, may 2020. ISSN 2041-1723. doi:10.1038/s41467-020-16399-y. URL `https://www.nature.com/articles/s41467-020-16399-y`.

Rasmus Nielsen, Joshua M Akey, Mattias Jakobsson, Jonathan K Pritchard, Sarah Tishkoff, and Eske Willerslev. Tracing the peopling of the world through genomics. *Nature*, 541 (7637):302–310, 2017.

Magnus Nordborg, Brian Charlesworth, and Deborah Charlesworth. The Effect of Recombination on the Speed of Evolution. *Genetical Research*, 67(2):159–174, 1996. doi:doi: 10.1017/s0016672300033619.

John Novembre and Matthew Stephens. Interpreting principal component analyses of spatial population genetic variation. *Nature Genetics*, 40(5):646–649, 2008.

Luke J. O'Connor, Armin P Schoech, Farhad Hormozdiari, Steven Gazal, Nick Patterson, and Alkes L Price. Extreme Polygenicity of Complex Traits Is Explained by Negative Selection. *American Journal of Human Genetics*, 105(3):456–476, 2019. ISSN 15376605. doi:10.1016/j.ajhg.2019.07.003. URL `https://doi.org/10.1016/j.ajhg.2019.07.003.`

Diego Ortega-Del Vecchyo, Kirk E Lohmueller, and John Novembre. Haplotype-based inference of the distribution of fitness effects. *Genetics*, 220(4), 2022. ISSN 19432631. doi:10.1093/genetics/iyac002. URL `https://doi.org/10.1093/genetics/iyac002`.

Nick Patterson, Priya Moorjani, Yontao Luo, Swapan Mallick, Nadin Rohland, Yiping Zhan, Teri Genschoreck, Teresa Webster, and David Reich. Ancient admixture in human history. *Genetics*, 192(3):1065–1093, 2012.

Nick Patterson et al. Methods for High-Density Admixture Mapping of Disease Genes. *The American Journal of Human Genetics*, 74:979–1000, 2004.

Joseph K Pickrell and Jonathan K Pritchard. Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. *PLoS Genetics*, 8(11):1002967, 2012.

Ryan Poplin, Valentin Ruano-Rubio, Mark A. DePristo, Tim J. Fennell, Mauricio O. Carneiro, Geraldine A. Van der Auwera, David E. Kling, Laura D. Gauthier, Ami Levy-Moonshine, David Roazen, Khalid Shakir, Joel Thibault, Sheila Chandran, Chris Whelan, Monkol Lek, Stacey Gabriel, Mark J Daly, Ben Neale, Daniel G. MacArthur, and Eric Banks. Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv*, 2018. doi:10.1101/201178. URL `https://www.biorxiv.org/content/early/2018/07/24/201178`.

Alkes L Price, Nick J Patterson, Robert M Plenge, Michael E Weinblatt, Nancy A Shadick, and David Reich. Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, 38(8):904–909, 2006. ISSN 10614036. doi:10.1038/ng1847. URL `http://www.nature.com/naturegenetics`.

Alkes L Price et al. Sensitive detection of chromosomal segments of distinct ancestry in admixed populations. *PLoS Genetics*, 5:1000519, 2009.

George R Price. Selection and Covariance. *Nature*, 227:520–521, 1970.

Jonathan K Pritchard and Molly Przeworski. Linkage Disequilibrium in Humans: Models and Data. *The American Journal of Human Genetics*, 69(1):1–14, 2001.

Jonathan K. Pritchard, Joseph K. Pickrell, and Graham Coop. The genetics of human adaptation: Hard sweeps, soft sweeps, and polygenic adaptation. *Current Biology*, 20(4): R208–R215, 2010. ISSN 0960-9822. doi:https://doi.org/10.1016/j.cub.2009.11.055. URL https://www.sciencedirect.com/science/article/pii/S0960982209020703.

Shaun Purcell, Benjamin Neale, Kathe Todd-Brown, Lori Thomas, Manuel A R Ferreira, David Bender, Julian Maller, Pamela Sklar, Paul I W de Bakker, Mark J Daly, and Pak C Sham. Plink: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*, 81(3):559–575, Sep 2007. ISSN 0002-9297 (Print); 1537-6605 (Electronic); 0002-9297 (Linking). doi:10.1086/519795.

Lluis Quintana-Murci. Human Immunology through the Lens of Evolutionary Genetics. *Cell*, 177(1):184–199, 2019.

Lawrence R Rabiner. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.

Fernando Racimo, Jeremy J Berg, and Joseph K Pickrell. Detecting polygenic adaptation in admixture graphs. *Genetics*, 208(4):1565–1584, 2018a.

Fernando Racimo, Jeremy J Berg, and Joseph K Pickrell. Detecting polygenic adaptation in admixture graphs. *Genetics*, 208(4):1565–1584, 2018b. ISSN 19432631. doi:10.1534/genetics.117.300489. URL https://doi.org/10.1534/genetics.117.300489.

Aaron P Ragsdale and Ryan N Gutenkunst. Inferring demographic history using two-locus statistics. *Genetics*, 206(2):1037–1048, 2017.

Aaron P. Ragsdale, Dominic Nelson, Simon Gravel, and Jerome Kelleher. Lessons Learned from Bugs in Models of Human History. *American Journal of Human Genetics*, 107 (4):583–588, oct 2020. ISSN 15376605. doi:10.1016/j.ajhg.2020.08.017. URL https://doi.org/10.1016/j.ajhg.2020.08.017.

Peter Ralph and Graham Coop. The geography of recent genetic ancestry across europe. *PLOS Biology*, 11(5):1–20, 05 2013. doi:10.1371/journal.pbio.1001555. URL https://doi.org/10.1371/journal.pbio.1001555.

David Reich. *Who We Are and How We Got Here: Ancient DNA and the New Science of the Human Past*. Pantheon Books, Oxford University Press, New York, 2018.

Harald Ringbauer, John Novembre, and Matthias Steinrücken. Parental relatedness through time revealed by runs of homozygosity in ancient dna. *Nature Communications*, 12 (1):5425, 2021. doi:10.1038/s41467-021-25289-w. URL https://doi.org/10.1038/s41467-021-25289-w.

Erica Bree Rosenblum, Timothy Y. James, Kelly R. Zamudio, Thomas J. Poorten, Dan Ilut, David Rodriguez, Jonathan M. Eastman, Katy Richards-Hrdlicka, Suzanne Joneson, Thomas S. Jenkinson, Joyce E. Longcore, Gabriela Parra Olea, Luís Felipe Toledo, Maria Luz Arellano, Edgar M. Medina, Silvia Restrepo, Sandra Victoria Flechas, Lee Berger, Cheryl J. Briggs, and Jason E. Stajich. Complex history of the amphibian-killing chytrid fungus revealed with genome resequencing data. *Proceedings of the National Academy of Sciences of the United States of America*, 110(23):9385–9390, jun 2013. ISSN 00278424. doi:10.1073/pnas.1300130110. URL `https://www.pnas.org/doi/abs/10.1073/pnas.1300130110`.

Georgina L. Ryland, Maria A. Doyle, David Goode, Samantha E. Boyle, David Y.H. Choong, Simone M. Rowley, Jason Li, David Dl Bowtell, Richard W. Tothill, Ian G. Campbell, and Kylie L. Gorringe. Loss of heterozygosity: What is it good for? *BMC Medical Genomics*, 8(1):1–12, aug 2015. ISSN 17558794. doi:10.1186/s12920-015-0123-z. URL `https://bmcmedgenomics.biomedcentral.com/articles/10.1186/s12920-015-0123-z`.

Pardis C. Sabeti, David E. Reich, John M. Higgins, Haninah Z.P. Levine, Daniel J. Richter, Stephen F. Schaffner, Stacey B. Gabriel, Jill V. Platko, Nick J. Patterson, Gavin J. McDonald, Hans C. Ackerman, Sarah J. Campbell, David Altshuler, Richard Cooper, Dominic Kwiatkowski, Ryk Ward, and Eric S. Lander. Detecting recent positive selection in the human genome from haplotype structure. *Nature*, 419(6909):832–837, 2002. ISSN 00280836. doi:10.1038/nature01140.

Jaleal S. Sanjak, Julia Sidorenko, Matthew R. Robinson, Kevin R. Thornton, and Peter M. Visscher. Evidence of directional and stabilizing selection in contemporary humans. *Proceedings of the National Academy of Sciences of the United States of America*, 115(1):151–156, jan 2018. ISSN 10916490. doi:10.1073/pnas.1707227114. URL `https://www.pnas.org/doi/abs/10.1073/pnas.1707227114`.

Stanley A Sawyer and Daniel L Hartl. Population Genetics of Polymorphism and Divergence. *Genetics*, 132:1161–1176, 1992. URL `https://academic.oup.com/genetics/article/132/4/1161/6009179`.

Paul Scheet and Matthew Stephens. A fast and flexible statistical model for large-scale population genotype data: Applications to inferring missing genotypes and haplotypic phase. *American Journal of Human Genetics*, 78(4):629–644, apr 2006. ISSN 00029297. doi:10.1086/502802. URL `http://www.cell.com/article/S000292970763701X/fulltexthttp://www.cell.com/article/S000292970763701X/abstracthttps://www.cell.com/ajhg/abstract/S0002-9297(07)63701-X`.

Armin P. Schoech, Daniel M. Jordan, Po Ru Loh, Steven Gazal, Luke J. O'Connor, Daniel J. Balick, Pier F. Palamara, Hilary K. Finucane, Shamil R. Sunyaev, and Alkes L. Price. Quantification of frequency-dependent genetic architectures in 25 UK Biobank traits reveals action of negative selection. *Nature Communications*, 10(1):1–10, feb 2019. ISSN 20411723. doi:10.1038/s41467-019-08424-6. URL `https://www.nature.com/articles/s41467-019-08424-6`.

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

Guy Sella and Nicholas H. Barton. Thinking about the Evolution of Complex Traits in the Era of Genome-Wide Association Studies. *Annual Review of Genomics and Human Genetics*, 20:461–493, aug 2019. ISSN 1545293X. doi:10.1146/annurev-genom-083115-022316. URL https://www.annualreviews.org/doi/abs/10.1146/annurev-genom-083115-022316.

Pak C Sham and Shaun M Purcell. Statistical power and significance testing in large-scale genetic studies. *Nature Reviews Genetics*, 15(5):335–346, 2014. ISSN 14710064. doi:10.1038/nrg3706. URL www.nature.com/reviews/genetics.

Jinxia Shi, Wenwu Ye, Dongfang Ma, Junliang Yin, Zhichao Zhang, Yuanchao Wang, and Yongli Qiao. Improved whole-genome sequence of phytophthora capsici generated by long-read sequencing. *Molecular Plant-Microbe Interactions®*, 34(7): 866–869, 2021. doi:10.1094/MPMI-12-20-0356-A. URL https://doi.org/10.1094/MPMI-12-20-0356-A. PMID: 33720746.

Yuval B Simons, Kevin Bullaughey, Richard R Hudson, and Guy Sella. A population genetic interpretation of GWAS findings for human quantitative traits. *PLoS Biology*, 16, 2018. doi:10.1371/journal.pbio.2002985.

Montgomery Slatkin and Fernando Racimo. Ancient DNA and human history. *Proceedings of the National Academy of Sciences*, 113(23):6380–6387, 2016.

Mashaal Sohail, Robert M Maier, Andrea Ganna, Alex Bloemendal, Alicia R Martin, Michael C Turchin, Charleston WK Chiang, Joel Hirschhorn, Mark J Daly, Nick Patterson, Benjamin Neale, Iain Mathieson, David Reich, and Shamil R Sunyaev. Polygenic adaptation on height is overestimated due to uncorrected stratification in genome-wide association studies. *eLife*, 8:1–17, 2019. doi:10.7554/elife.39702.

Yun S Song and Matthias Steinrücken. A simple method for finding explicit analytic transition densities of diffusion processes with general diploid selection. *Genetics*, 190(3): 1117–1129, 2012. doi:10.1534/genetics.111.136929.

Doug Speed, Gibran Hemani, Michael R. Johnson, and David J. Balding. Improved heritability estimation from genome-wide SNPs. *American Journal of Human Genetics*, 91(6):1011–1021, dec 2012. ISSN 00029297. doi:10.1016/j.ajhg.2012.10.010. URL http://www.cell.com/article/S0002929712005332/fulltexthttp://www.cell.com/article/S0002929712005332/abstracthttps://www.cell.com/ajhg/abstract/S0002-9297(12)00533-2.

Chris C.A. Spencer, Zhan Su, Peter Donnelly, and Jonathan Marchini. Designing genome-wide association studies: Sample size, power, imputation, and the choice of genotyping chip. *PLoS Genetics*, 5(5):1000477, 2009. ISSN 15537390. doi:10.1371/journal.pgen.1000477. URL www.plosgenetics.org.

Christopher D. Steele, Ammal Abbasi, S. M.Ashiqul Islam, Amy L. Bowes, Azhar Khandekar, Kerstin Haase, Shadi Hames-Fathi, Dolapo Ajayi, Annelien Verfaillie, Pawan Dhami, Alex McLatchie, Matt Lechner, Nicholas Light, Adam Shlien, David Malkin, Andrew Feber, Paula Proszek, Tom Lesluyes, Fredrik Mertens, Adrienne M. Flanagan, Maxime Tarabichi, Peter Van Loo, Ludmil B. Alexandrov, and Nischalan Pillay. Signatures of copy number alterations in human cancer. *Nature*, 606(7916):984–991, jun 2022. ISSN 14764687. doi:10.1038/s41586-022-04738-6. URL `https://www.nature.com/articles/s41586-022-04738-6`.

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

Matthias Steinrücken, Ethan M Jewett, and Yun S Song. SpectralTDF: Transition densities of diffusion processes with time-varying selection parameters, mutation rates and effective population sizes. *Bioinformatics*, 32(5):795–797, 2016. ISSN 14602059. doi:10.1093/bioinformatics/btv627. URL `https://academic.oup.com/bioinformatics/article-abstract/32/5/795/1743415`.

Matthias Steinrücken, Jeffrey P Spence, John A Kamm, Emilia Wieczorek, and Yun S. Song. Model-based detection and analysis of introgressed neanderthal ancestry in modern humans. *Molecular Ecology*, 27:3873–3888, 2018.

Matthew Stephens. False discovery rates: A new deal. *Biostatistics*, 18(2):275–294, apr 2017. ISSN 14684357. doi:10.1093/biostatistics/kxw041. URL `https://academic.oup.com/biostatistics/article/18/2/275/2557030`.

Aaron J Stern, Leo Speidel, Noah A Zaitlen, and Rasmus Nielsen. Disentangling selection on genetically correlated polygenic traits via whole-genome genealogies. *American Journal of Human Genetics*, 108(2):219–239, 2021. ISSN 15376605. doi:10.1016/j.ajhg.2020.12.005. URL `https://doi.org/10.1016/j.ajhg.2020.12.005.`

Aravind Subramanian, Pablo Tamayo, Vamsi K Mootha, Sayan Mukherjee, Benjamin L Ebert, Michael A Gillette, Amanda Paulovich, Scott L Pomeroy, Todd R Golub, Eric S Lander, and Jill P Mesirov. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*, 102(43): 15545–15550, Oct 2005. doi:10.1073/pnas.0506580102.

Yang Sui, Lei Qi, Jian-Kun Wu, Xue-Ping Wen, Xing-Xing Tang, Zhong-Jun Ma, Xue-Chang Wu, Ke Zhang, Robert J. Kokoska, Dao-Qiong Zheng, and Thomas D. Petes. Genome-wide mapping of spontaneous genetic alterations in diploid yeast cells. *Proceedings of the National Academy of Sciences*, 117(45):28191–28200, 2020. doi:10.1073/pnas.2018633117. URL `https://www.pnas.org/doi/abs/10.1073/pnas.2018633117`.

Kelly Swarts, Rafal M. Gutaker, Bruce Benz, Michael Blake, Robert Bukowski, James Holland, Melissa Kruse-Peeples, Nicholas Lepak, Lynda Prim, M. Cinta Romay, Jeffrey

Ross-Ibarra, Jose De Jesus Sanchez-Gonzalez, Chris Schmidt, Verena J. Schuenemann, Johannes Krause, R. G. Matson, Detlef Weigel, Edward S. Buckler, and Hernán A. Burbano. Genomic estimation of complex traits reveals ancient maize adaptation to temperate North America. *Science*, 357(6350):512–515, aug 2017. ISSN 10959203. doi:10.1126/science.aam9425. URL `http://science.sciencemag.org/`.

Bjarni J. Vilhjálmsson, Jian Yang, Hilary K. Finucane, Alexander Gusev, Sara Lindström, Stephan Ripke, Giulio Genovese, Po Ru Loh, Gaurav Bhatia, Ron Do, Tristan Hayeck, Hong Hee Won, Benjamin M. Neale, Aiden Corvin, James T.R. Walters, Kai How Farh, Peter A. Holmans, Phil Lee, Brendan Bulik-Sullivan, David A. Collier, Hailiang Huang, Tune H. Pers, Ingrid Agartz, Esben Agerbo, Margot Albus, Madeline Alexander, Farooq Amin, Silviu A. Bacanu, Martin Begemann, Richard A. Belliveau, Judit Bene, Sarah E. Bergen, Elizabeth Bevilacqua, Tim B. Bigdeli, Donald W. Black, Richard Bruggeman, Nancy G. Buccola, Randy L. Buckner, William Byerley, Wiepke Cahn, Guiqing Cai, Dominique Campion, Rita M. Cantor, Vaughan J. Carr, Noa Carrera, Stanley V. Catts, Kimberly D. Chambert, Raymond C.K. Chan, Ronald Y.L. Chen, Eric Y.H. Chen, Wei Cheng, Eric F.C. Cheung, Siow Ann Chong, C. Robert Cloninger, David Cohen, Nadine Cohen, Paul Cormican, Nick Craddock, James J. Crowley, David Curtis, Michael Davidson, Kenneth L. Davis, Franziska Degenhardt, Jurgen Del Favero, Lynn E. Delisi, Ditte Demontis, Dimitris Dikeos, Timothy Dinan, Srdjan Djurovic, Gary Donohoe, Elodie Drapeau, Jubao Duan, Frank Dudbridge, Naser Durmishi, Peter Eichhammer, Johan Eriksson, Valentina Escott-Price, Laurent Essioux, Ayman H. Fanous, Martilias S. Farrell, Josef Frank, Lude Franke, Robert Freedman, Nelson B. Freimer, Marion Friedl, Joseph I. Friedman, Menachem Fromer, Lyudmila Georgieva, Elliot S. Gershon, Ina Giegling, Paola Giusti-Rodrguez, Stephanie Godard, Jacqueline I. Goldstein, Vera Golimbet, Srihari Gopal, Jacob Gratten, Jakob Grove, Lieuwe De Haan, Christian Hammer, Marian L. Hamshere, Mark Hansen, Thomas Hansen, Vahram Haroutunian, Annette M. Hartmann, Frans A. Henskens, Stefan Herms, Joel N. Hirschhorn, Per Hoffmann, Andrea Hofman, Mads V. Hollegaard, David M. Hougaard, Masashi Ikeda, Inge Joa, Antonio Julia, Rene S. Kahn, Luba Kalaydjieva, Sena Karachanak-Yankova, Juha Karjalainen, David Kavanagh, Matthew C. Keller, Brian J. Kelly, James L. Kennedy, Andrey Khrunin, Yunjung Kim, Janis Klovins, James A. Knowles, Bettina Konte, Vaidutis Kucinskas, Zita Ausrele Kucinskiene, Hana Kuzelova-Ptackova, Anna K. Kahler, Claudine Laurent, Jimmy Lee Chee Keong, S. Hong Lee, Sophie E. Legge, Bernard Lerer, Miaoxin Li, Tao Li, Kung Yee Liang, Jeffrey Lieberman, Svetlana Limborska, Carmel M. Loughland, Jan Lubinski, Jouko Lnnqvist, Milan Macek, Patrik K.E. Magnusson, Brion S. Maher, Wolfgang Maier, Jacques Mallet, Sara Marsal, Manuel Mattheisen, Morten Mattingsdal, Robert W. McCarley, Colm McDonald, Andrew M. McIntosh, Sandra Meier, Carin J. Meijer, Bela Melegh, Ingrid Melle, Raquelle I. Mesholam-Gately, Andres Metspalu, Patricia T. Michie, Lili Milani, Vihra Milanova, Younes Mokrab, Derek W. Morris, Ole Mors, Preben B. Mortensen, Kieran C. Murphy, Robin M. Murray, Inez Myin-Germeys, Bertram Mller-Myhsok, Mari Nelis, Igor Nenadic, Deborah A. Nertney, Gerald Nestadt, Kristin K. Nicodemus, Liene Nikitina-Zake, Laura Nisenbaum, Annelie Nordin, Eadbhard O'Callaghan, Colm O'Dushlaine, F. Anthony O'Neill, Sang Yun Oh, Ann Olincy, Line Olsen, Jim

Van Os, Christos Pantelis, George N. Papadimitriou, Sergi Papiol, Elena Parkhomenko, Michele T. Pato, Tiina Paunio, Milica Pejovic-Milovancevic, Diana O. Perkins, Olli Pietilinen, Jonathan Pimm, Andrew J. Pocklington, John Powell, Alkes Price, Ann E. Pulver, Shaun M. Purcell, Digby Quested, Henrik B. Rasmussen, Abraham Reichenberg, Mark A. Reimers, Alexander L. Richards, Joshua L. Roffman, Panos Roussos, Douglas M. Ruderfer, Veikko Salomaa, Alan R. Sanders, Ulrich Schall, Christian R. Schubert, Thomas G. Schulze, Sibylle G. Schwab, Edward M. Scolnick, Rodney J. Scott, Larry J. Seidman, Jianxin Shi, Engilbert Sigurdsson, Teimuraz Silagadze, Jeremy M. Silverman, Kang Sim, Petr Slominsky, Jordan W. Smoller, Hon Cheong So, Chris C.A. Spencer, Eli A. Stahl, Hreinn Stefansson, Stacy Steinberg, Elisabeth Stogmann, Richard E. Straub, Eric Strengman, Jana Strohmaier, T. Scott Stroup, Mythily Subramaniam, Jaana Suvisaari, Dragan M. Svrakic, Jin P. Szatkiewicz, Erik Sderman, Srinivas Thirumalai, Draga Toncheva, Paul A. Tooney, Sarah Tosato, Juha Veijola, John Waddington, Dermot Walsh, Dai Wang, Qiang Wang, Bradley T. Webb, Mark Weiser, Dieter B. Wildenauer, Nigel M. Williams, Stephanie Williams, Stephanie H. Witt, Aaron R. Wolen, Emily H.M. Wong, Brandon K. Wormley, Jing Qin Wu, Hualin Simon Xi, Clement C. Zai, Xuebin Zheng, Fritz Zimprich, Naomi R. Wray, Kari Stefansson, Rolf Adolfsson, Ole A. Andreassen, Peter M. Visscher, Douglas H.R. Blackwood, Elvira Bramon, Joseph D. Buxbaum, Anders D. Børglum, Sven Cichon, Ariel Darvasi, Enrico Domenici, Hannelore Ehrenreich, Tonu Esko, Pablo V. Gejman, Michael Gill, Hugh Gurling, Christina M. Hultman, Nakao Iwata, Assen V. Jablensky, Erik G. Jonsson, Kenneth S. Kendler, George Kirov, Jo Knight, Todd Lencz, Douglas F. Levinson, Qingqin S. Li, Jianjun Liu, Anil K. Malhotra, Steven A. McCarroll, Andrew McQuillin, Jennifer L. Moran, Bryan J. Mowry, Markus M. Nthen, Roel A. Ophoff, Michael J. Owen, Aarno Palotie, Carlos N. Pato, Tracey L. Petryshen, Danielle Posthuma, Marcella Rietschel, Brien P. Riley, Dan Rujescu, Pak C. Sham, Pamela Sklar, David St. Clair, Daniel R. Weinberger, Jens R. Wendland, Thomas Werge, Mark J. Daly, Patrick F. Sullivan, Michael C. O'Donovan, Peter Kraft, David J. Hunter, Muriel Adank, Habibul Ahsan, Kristiina Aittomäki, Laura Baglietto, Sonja Berndt, Carl Blomquist, Federico Canzian, Jenny Chang-Claude, Stephen J. Chanock, Laura Crisponi, Kamila Czene, Norbert Dahmen, Isabel Dos Santos Silva, Douglas Easton, A. Heather Eliassen, Jonine Figueroa, Olivia Fletcher, Montserrat Garcia-Closas, Mia M. Gaudet, Lorna Gibson, Christopher A. Haiman, Per Hall, Aditi Hazra, Rebecca Hein, Brian E. Henderson, John L. Hopper, Astrid Irwanto, Mattias Johansson, Rudolf Kaaks, Muhammad G. Kibriya, Peter Lichtner, Eiliv Lund, Enes Makalic, Alfons Meindl, Hanne Meijers-Heijboer, Bertram Müller-Myhsok, Taru A. Muranen, Heli Nevanlinna, Petra H. Peeters, Julian Peto, Ross L. Prentice, Nazneen Rahman, María José Sánchez, Daniel F. Schmidt, Rita K. Schmutzler, Melissa C. Southey, Rulla Tamimi, Ruth Travis, Clare Turnbull, Andre G. Uitterlinden, Rob B. Van Der Luijt, Quinten Waisfisz, Zhaoming Wang, Alice S. Whittemore, Rose Yang, Wei Zheng, Sekar Kathiresan, Michele Pato, Carlos Pato, Eli Stahl, Noah Zaitlen, Bogdan Pasaniuc, Eimear E. Kenny, Mikkel H. Schierup, Philip De Jager, Nikolaos A. Patsopoulos, Steve McCarroll, Mark Daly, Shaun Purcell, Daniel Chasman, Benjamin Neale, Michael Goddard, Nick Patterson, and Alkes L. Price. Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *American Journal of Hu-*

man Genetics, 97(4):576–592, 2015. ISSN 15376605. doi:10.1016/j.ajhg.2015.09.001. URL http://dx.doi.org/10.1016/j.ajhg.2015.09.001.

Gregory Vogel, Michael A. Gore, and Christine D. Smart. Genome-wide association study in new york phytophthora capsici isolates reveals loci involved in mating type and mefenoxam sensitivity. *Phytopathology*, 111(1):204–216, jan 2021. ISSN 19437684. doi:10.1094/PHYTO-04-20-0112-FI. URL https://apsjournals.apsnet.org/doi/10.1094/PHYTO-04-20-0112-FI.

Roman Voronka and Joseph B. Keller. Asymptotic analysis of stochastic models in population genetics. *Mathematical Biosciences*, 25(3-4):331–362, 1975. ISSN 00255564. doi:10.1016/0025-5564(75)90010-3.

Joshua M Wang, Richard J Bennett, and Matthew Z Anderson. The Genome of the Human Pathogen Candida albicans Is Shaped by Mutation and Cryptic Sexual Recombination. *mBio*, 9(5), 2018. doi:10.1128/mbio.01205-18.

Ying Wang, Jing Guo, Guiyan Ni, Jian Yang, Peter M Visscher, and Loic Yengo. Theoretical and empirical quantification of the accuracy of polygenic scores in ancestry divergent populations. *Nature Communications*, 11(1), 2020. ISSN 20411723. doi:10.1038/s41467-020-17719-y. URL https://doi.org/10.1038/s41467-020-17719-y.

B. S. Weir. *Genetic Data Analysis II.* Sinauer Associates, Inc., Sunderland, MA, 1996.

Michael C. Whitlock. Evolutionary inference from QST, apr 2008. ISSN 09621083. URL https://onlinelibrary.wiley.com/doi/full/10.1111/j.1365-294X.2008.03712.xhttps://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-294X.2008.03712.xhttps://onlinelibrary.wiley.com/doi/10.1111/j.1365-294X.2008.03712.x.

Katherine Wilkins and Thomas LaFramboise. Losing balance: Hardy–Weinberg disequilibrium as a marker for recurrent loss-of-heterozygosity in cancer. *Human Molecular Genetics*, 20(24):4831–4839, 09 2011. ISSN 0964-6906. doi:10.1093/hmg/ddr422. URL https://doi.org/10.1093/hmg/ddr422.

Scott H. Williamson, Ryan Hernandez, Adi Fledel-Alon, Lan Zhu, Rasmus Nielsen, and Carlos D. Bustamante. Simultaneous inference of selection and population growth from patterns of variation in the human genome. *Proceedings of the National Academy of Sciences of the United States of America*, 102(22):7882–7887, may 2005. ISSN 00278424. doi:10.1073/pnas.0502300102. URL https://www.pnas.org/doi/abs/10.1073/pnas.0502300102.

Vanessa S Windhausen, Gary N Atlin, John M Hickey, Jose Crossa, Jean Luc Jannink, Mark E Sorrells, Babu Raman, Jill E Cairns, Amsal Tarekegne, Kassa Semagn, Yoseph Beyene, Pichet Grudloyma, Frank Technow, Christian Riedelsheimer, and Albrecht E Melchinger. Effectiveness of genomic prediction of maize hybrid performance in different

breeding populations and environments. *G3: Genes, Genomes, Genetics*, 2(11):1427–1436, 2012. ISSN 21601836. doi:10.1534/g3.112.003699. URL http://www.g3journal.org/lookup/suppl/doi:10.1534/g3.112.003699/-/DC1.

Naomi R Wray, Michael E Goddard, and Peter M Visscher. Prediction of individual genetic risk to disease from genome-wide association studies. *Genome research*, 17(10):1520–1528, 2007.

Naomi R. Wray, Jian Yang, Ben J. Hayes, Alkes L. Price, Michael E. Goddard, and Peter M. Visscher. Pitfalls of predicting complex traits from SNPs. *Nature Reviews Genetics*, 14 (7):507–515, 2013. ISSN 14710056. doi:10.1038/nrg3457.

Sivan Yair and Graham Coop. Population differentiation of polygenic score predictions under stabilizing selection. *bioRxiv*, page 2021.09.10.459833, 9 2021. doi:10.1101/2021.09.10.459833. URL https://www.biorxiv.org/content/10.1101/2021.09.10.459833v1https://www.biorxiv.org/content/10.1101/2021.09.10.459833v1.abstract.

Sivan Yair and Graham Coop. Population differentiation of polygenic score predictions under stabilizing selection. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 377(1852), 2022. ISSN 14712970. doi:10.1098/rstb.2020.0416. URL https://royalsocietypublishing.org/doi/10.1098/rstb.2020.0416.

Jian Yang, Beben Benyamin, Brian P. McEvoy, Scott Gordon, Anjali K Henders, Dale R Nyholt, Pamela A Madden, Andrew C Heath, Nicholas G Martin, Grant W Montgomery, Michael E Goddard, and Peter M Visscher. Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics*, 42(7):565–569, 2010. ISSN 10614036. doi:10.1038/ng.608.

Jianming Yu, Gael Pressoir, William H Briggs, Irie Vroh Bi, Masanori Yamasaki, John F Doebley, Michael D. McMullen, Brandon S Gaut, Dahlia M Nielsen, James B Holland, Stephen Kresovich, and Edward S Buckler. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics*, 38(2):203–208, 2006. ISSN 10614036. doi:10.1038/ng1702. URL http://www.nature.com/naturegenetics.

Jian Zeng, Ronald De Vlaming, Yang Wu, Matthew R Robinson, Luke R Lloyd-Jones, Loic Yengo, Chloe X Yap, Angli Xue, Julia Sidorenko, Allan F. McRae, Joseph E Powell, Grant W Montgomery, Andres Metspalu, Tonu Esko, Greg Gibson, Naomi R Wray, Peter M Visscher, and Jian Yang. Signatures of negative selection in the genetic architecture of human complex traits. *Nature Genetics*, 50(5):746–753, 2018a. ISSN 15461718. doi:10.1038/s41588-018-0101-4. URL https://doi.org/10.1038/s41588-018-0101-4.

Jian Zeng, Ronald De Vlaming, Yang Wu, Matthew R. Robinson, Luke R. Lloyd-Jones, Loic Yengo, Chloe X. Yap, Angli Xue, Julia Sidorenko, Allan F. McRae, Joseph E. Powell, Grant W. Montgomery, Andres Metspalu, Tonu Esko, Greg Gibson, Naomi R. Wray,

Peter M. Visscher, and Jian Yang. Signatures of negative selection in the genetic architecture of human complex traits. *Nature Genetics*, 50(5):746–753, apr 2018b. ISSN 15461718. doi:10.1038/s41588-018-0101-4. URL https://www.nature.com/articles/s41588-018-0101-4.

Jian Zeng, Angli Xue, Longda Jiang, Luke R Lloyd-Jones, Yang Wu, Huanwei Wang, Zhili Zheng, Loic Yengo, Kathryn E Kemper, Michael E Goddard, Naomi R Wray, Peter M Visscher, and Jian Yang. Widespread signatures of natural selection across human complex traits and functional genomic categories. *Nature Communications*, 12(1), 2021. ISSN 20411723. doi:10.1038/s41467-021-21446-3. URL https://doi.org/10.1038/s41467-021-21446-3.

Yan Zhang, Guanghao Qi, Ju Hyun Park, and Nilanjan Chatterjee. Estimation of complex effect-size distributions using summary-level statistics from genome-wide association studies across 32 complex traits. *Nature Genetics*, 50(9):1318–1326, 2018. ISSN 15461718. doi:10.1038/s41588-018-0193-x. URL https://doi.org/10.1038/s41588-018-0193-x.

Xiang Zhou, Peter Carbonetto, and Matthew Stephens. Polygenic Modeling with Bayesian Sparse Linear Mixed Models. *PLoS Genetics*, 9(2):e1003264, feb 2013. ISSN 15537390. doi:10.1371/journal.pgen.1003264. URL https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1003264.