Selection and Introgression in Humans and Other Animals

By

Sivan Yair

Dissertation

Submitted in partial satisfaction of the requirements for the degree of

Doctor of Philosophy

in

Population Biology

in the

Office of Graduate Studies

of the

University of California

Davis

Approved:

Graham M. Coop, Chair

Michael Turelli

David J. Begun

Committee in Charge

2022

To my family, who has done so much for me.

Contents

Abstract

Selection and Introgression in Humans and Other Animals

Many longstanding questions in evolution concern the major forces that shape genetic variation in natural populations: natural selection, genetic drift, mutation, gene flow, and recombination. Their interaction determines the means by which populations adapt and explains patterns of genetic and phenotypic variation within and among populations. In this dissertation, I use recent developments in population genetics to study the genomic signatures of these outcomes and their implications for inferring the history of adaptation and predicting phenotypes.

In chapters 1 and 2, I focus on the process of adaptive introgression, whereby selection favors the spread of alleles introduced by gene flow from another species. For both of these chapters, I develop methods based on coalescent theory that use recombination as a signal to understand how and when introgressed alleles facilitated adaptation to new environments. In chapter 1, I analyzed both ancient- and present-day DNA samples from humans and Neanderthals, determining genomic regions where Neanderthal introgression provided standing variation for future adaptation in human populations. In chapter 2, I analyzed samples from two sister killifish species in which some populations can survive extremely toxic waters from pollution. In resistant *Fundulus grandis* populations of the Houston ship channel, I identified cases of extremely strong selection acting on rare introgressed variation that was introduced recently from a resistant *Fundulus heteroclitus* population along the Atlantic coast. Together, these chapters uncover the selection pressures acting on populations and highlight that admixture between closely related species can be a source of adaptive variation immediately and in the long term.

In chapter 3, I provide an evolutionary perspective on the limitations of genome-wide association study results to understand complex trait variation in understudied populations.

I model how mutation, stabilizing selection, and genetic drift interact to shape quantitative trait variation and quantify their implications for phenotype prediction across populations when ascertaining underlying genetic variation in a subset of populations.

The combination of theoretical models and analysis of population genomic data in this dissertation provide a rich investigation into the genetics of adaptation and limitations of common approaches in understanding the relationship between genetic and phenotypic variation.

Acknowledgments

I am beyond grateful that I had the opportunity to grow as a scientist in the Coop lab. Thank you to Graham Coop, for helping me transform ideas into exciting projects, your constant encouragement, advice and patience, and for supporting me as I explored different career options and research projects. I have learned an incredible amount from you. I also could not have asked for better labmates, who are great mentors, friends, and role models: Kristin Lee, Anita To, Nancy Chen, Emily Josephs, Erin Calfee, Vince Buffalo, Katie Ferris, Doc Edge, Matt Osmond, Jeff Groh, Pavitra Muralidhar, Carl Veller, Alexis Simon, Silu Wang, and James Kitchens.

The exchanges I've had with other scientists, from conversations in conference venue hallways to collaborative projects, have always energized and inspired me. I thank Michael Turelli, David Begun, Jeff Ross-Ibarra, Chuck Langley, Emilia Huerta-Sanchez, Annie Schmitt, Santiago Ramirez, Andrew Whitehead, Iain Mathieson, and Tim Weaver for their feedback and productive conversations. It was a joy to work with Magdalena Bohutínská (Holcová) on her thesis work on the modes of convergent adaptation across scales of divergence in *Arabidopsis*, as well as Andrew Whitehead, Elias Oziolor, Noah Reid, and Cole Matson on the work on adaptive introgression in *Fundulus heteroclitus* presented in chapter 2. I thank the Ross-Ibarra lab for being my scientific home for a quarter. I am grateful to Laci Gerhart-Barley for guidance on my pedagogy research and for her work helping students become thoughtful educators. Thanks to my fellow students in Population Biology, who set a high standard for scientific achievement and living their values. You all inspire me.

The reviewers and editors of my manuscripts greatly improved by work. I thank Rasmus Nielsen, Sohini Ramachandran, Doc Edge, Noah Rosenberg, Yuval Simons, and five anonymous reviewers for making those contributions. I also thank the journals *Genetics*, *Science*, and *Phil Trans B*, whose material I've used in this dissertation. I am indebted to

Iain Mathieson, who provided the ancient-DNA data used in chapter 1, as well as invaluable advice on imputing ancient-DNA samples and on how to model their population history.

I am grateful to have been a part of the DNA in Society reading group organized by Emily Merchant. My conversations with academics in the Humanities and Social Sciences gave me the tools to think critically about my scientific process and how I communicate my work to others.

I owe so much to my friends and family, who over my time as a graduate student have always looked out for me: my sister Maya, my Imma and Abba, Anita To, Kristin Lee, Michelle Stitzer, Oshiomah Oyageshio, and my cohort-mates Elise Elwood, Fernanda Guizar, Asher Hudson, and Katherine Corn. Finally, I thank Sam Collitt for being my rock and supporting me in more ways than I thought possible.

Chapter 1

The timing of human adaptation from Neanderthal introgression

Sivan Yair[1,2], Kristin M. Lee[3], Graham Coop[1,2]

[1] Center for Population Biology, University of California, Davis

[2] Department of Evolution and Ecology, University of California, Davis

[3] Previous address: Department of Biological Sciences, Columbia University

Abstract

Admixture has the potential to facilitate adaptation by providing alleles that are immediately adaptive in a new environment or by simply increasing the long term reservoir of genetic diversity for future adaptation. A growing number of cases of adaptive introgression are being identified in species across the tree of life, however the timing of selection, and therefore the importance of the different evolutionary roles of admixture, is typically unknown. Here, we investigate the spatio-temporal history of selection favoring Neanderthal-introgressed alleles in modern human populations. Using both ancient and present-day samples of modern humans, we integrate the known demographic history of populations, namely population divergence and migration, with tests for selection. We model how a sweep placed along dif-

ferent branches of an admixture graph acts to modify the variance and covariance in neutral allele frequencies among populations at linked loci. Using a method based on this model of allele frequencies, we study previously identified cases of Neanderthal adaptive introgression. From these, we identify cases in which Neanderthal introgressed alleles were quickly beneficial and other cases in which they persisted at low frequency for some time. For some of the alleles that persisted at low frequency, we show that selection likely independently favored them later on in geographically separated populations. Our work highlights how admixture with ancient hominins has contributed to modern human adaptation and contextualizes observed levels of Neanderthal ancestry in present-day and ancient samples.

1.1 Introduction

Within the last decade, population genomic studies have revealed many cases of hybridization that have led to the introgression of genetic material between diverged populations. While these genetic introductions are often ecologically or developmentally maladaptive, a growing number of studies are providing evidence of adaptive introgression, whereby natural selection favored the spread of introgressed alleles (e.g. Whitney et al., 2006; The Heliconius Genome Consortium, 2012; Jones et al., 2018; Oziolor et al., 2019). These introgressed alleles likely facilitated adaptation to new or changing environments, highlighting admixture as a potentially important source of genetic variation for fitness.

Patterns of archaic introgression in modern humans offer some of the most compelling examples of adaptive introgression (reviewed in Racimo et al., 2015). When a subset of modern humans spread out of Africa, likely in the past hundred thousand years, they encountered a broad range of novel environments, including reduced UV exposure, new pathogen pressures, and colder climates. At roughly the same time that modern humans outside of Africa experienced new conditions, they were also interbreeding with Neanderthals, who had been living in and adapting to these Eurasian environments for hundreds of thousands of years (Hublin,

2009). The early-generation hybrids of Neanderthals and modern humans may have had low fitness due to the accumulation of weakly deleterious alleles in Neanderthals, who had a small effective population size. The low fitness of these early hybrids, combined with the greater efficacy of purifying selection in modern humans, led to widespread selection against deleterious Neanderthal alleles and linked Neanderthal variation (Harris and Nielsen, 2016; Juric et al., 2016). As a result, Neanderthal alleles segregate at around 1-3% frequency genome-wide in present day modern humans, with a depletion in gene rich, regulatory, and low recombination regions (Sankararaman et al., 2014; Schumer et al., 2018; Steinrücken et al., 2018; Petr et al., 2019; Telis et al., 2020).

While, on average, Neanderthal-derived alleles have been selected against, a few alleles persist at high frequency in present-day non-African populations and reflect putative cases of adaptive introgression (Abi-Rached et al., 2011; Sankararaman et al., 2014; Khrameeva et al., 2014; Vernot and Akey, 2014; Dannemann et al., 2016; Deschamps et al., 2016; Gittelman et al., 2016; Racimo et al., 2017; Jagoda et al., 2018; Quach et al., 2016; Sams et al., 2016; Setter et al., 2020). These alleles have been identified based on the characteristic patterns left behind by adaptive introgression, namely high haplotype similarity between Eurasian modern humans and Neanderthals, and the deep divergence of haplotypes within modern human populations. It appears that most sweeps on Neanderthal introgressed alleles were partial, with selected alleles only reaching frequencies around 30 to 60%. Thus, dips in genetic diversity due to adaptive Neanderthal introgression are dampened relative to hard, full sweeps on *de novo* mutations.

Some of the selected Neanderthal alleles contribute to traits that may have been under selection during modern human expansion out of Africa, such as immunity, skin pigmentation, and metabolism (Abi-Rached et al., 2011; Sankararaman et al., 2014; Khrameeva et al., 2014; Vernot and Akey, 2014; Dannemann et al., 2016; Gittelman et al., 2016; Racimo et al., 2017; Jagoda et al., 2018; Quach et al., 2016). Since modern humans mated with

Neanderthals around the same time that they were exposed to new environments, a natural assumption is that Neanderthal variation immediately facilitated modern human adaptation. Alternatively, some of the selected Neanderthal introgressed alleles may have contributed to the reservoir of standing genetic variation that became adaptive much later (Jagoda et al., 2018) as human populations were exposed to and created further novel environments. A recent study found evidence of this phenomenon for a Denisovan haplotype contributing to high altitude adaptation on the Tibetan Plateau (Zhang et al., 2020), and so non-immediate selection on introgressed archaic variation may be more common than previously thought.

Population genetics offers a number of approaches to date the timing of selection on alleles. The first broad category of approaches relies on the hitchhiking signal created in surrounding linked variants as the selected alleles sweep up a chunk of the haplotype on which they arose (or introgressed) (Maynard Smith and Haigh, 1974). Second, ancient DNA now offers an opportunity to assess when these selected haplotypes rose in frequency, and the ancient populations where they first achieved high frequency. Currently, methods that use ancient DNA to investigate the temporal history of selection focus on identifying significant increases in selected allele frequencies over time (e.g. Mathieson et al., 2015; Schraiber et al., 2016; Mathieson and Mathieson, 2018). These time series approaches are powerful because they can provide more direct evidence of selection driving allele frequency change, however they are limited to characterizing cases of selection that began after the oldest sampling time and among sampled populations. For studies of modern human evolution, we are currently restricted to learning about adaptation in mostly European populations and within the last ten thousand years, long after hybridization with Neanderthals.

Here, we leverage the advantages of both temporal sampling and patterns of neutral diversity at linked loci (the hitchhiking effect) to infer the time that Neanderthal introgressed alleles became adaptive. Our investigation of the hitchhiking effect allows us to date selection older than the ancient samples, while ancient populations provide useful reference points

closer to the sweep time. We use coalescent theory to describe patterns of sequence similarity around a selected site when selection favors introgressed variation at different times since admixture. Our predictions relate sequences among ancient and present day populations by incorporating information about their history of divergence and migration. In addition, we utilize information from partial sweeps by modeling this hitchhiking effect among both selected and non-selected haplotypes. Our approach builds on the model-based inference framework introduced in Lee and Coop (2017), which connects predicted coalescent histories to their corresponding probability distribution of population allele frequencies. Thus, we can distinguish among possible selection times by analyzing allele frequency data from ancient and present day populations.

By providing the age of selection favoring Neanderthal alleles in specific regions, we can determine the context within which Neanderthal alleles facilitated modern human adaptation and in turn narrow their potential phenotypic contributions. If selection was immediate, then the Neanderthal variation was useful in the early Eurasian environments that modern humans experienced, possibly because Neanderthals had been adapting to those same conditions for a long period of time. Otherwise, Neanderthal haplotypes were selected during different time periods and in different populations. These latter haplotypes are particularly interesting candidates, as they shed light on the similarities and differences in selection pressures affecting human populations over space and time. Further, they show that admixture between closely related species can provide an important source of standing genetic variation for future adaptation. In this paper, we provide evidence for each of these scenarios in different genomic regions of adaptive introgression.

We begin by introducing the data that we analyzed, followed by an intuitive description of the hitchhiking patterns that we modeled before providing the mathematical details of our model and inference framework. We then show how our method performs on simulations and its results when applied to candidates of adaptive Neanderthal introgression.

1.2 Distribution of adaptive Neanderthal introgressed haplotypes across populations

Choice of population samples We applied our method to a set of eight population samples: one archaic (Neanderthals), three present day modern human, and four ancient modern human. For the Neanderthal sample we used the Vindija *33.19* Neanderthal (n=2, for two chromosomes sampled from one individual) because it is high coverage and most closely related to the introgressing population (Prüfer et al., 2017; Mafessoni et al., 2020). Our results only change slightly when the high coverage Altai Neanderthal sample is included in the Neanderthal population, which we discuss later. The present day populations are from Phase 3 of the 1000 Genomes Project (The 1000 Genomes Project Consortium, 2015): the Yoruba in Ibadan, Nigeria (YRI; n=216), Han Chinese from Beijing (CHB; n=206), and Utah Residents (CEPH) with Northern and Western European Ancestry (CEU; $n = 198$). We chose ancient populations based on previous work about their relationships with present day populations. We included the West Eurasian Upper Paleolithic (EurUP; $n = 12$), who are the oldest samples and basally related to all West Eurasians (mean sampling time $t \approx 34$kya), and three populations known to be ancestral to present day Europeans: the Mesolithic hunter gatherers from western Europe (WHG; $n = 66$; $t \approx 9$kya), the Neolithic Anatolian Farmers (EF; $n = 134$; $t \approx 7$kya), and the Bronze Age Steppe individuals (Steppe; $n = 14$; $t \approx 5$kya). Each population ancestral to present day Europeans was composed of samples with $\geq 90\%$ inferred ancestry corresponding to that population. Inferred ancestry proportions are from Mathieson et al. (2018)'s supervised ADMIXTURE analysis using four clusters with fixed membership corresponding to each of these three ancient populations and Eastern Hunter Gatherers. See File S2 for detailed information on the ancient samples we analyze. One reason for taking this set of ancient populations is that our method requires a known population history, namely the approximate timing and admixture proportions among

populations, as well as the timing of divergence among Neanderthal-admixed populations. The primary admixture graph that we used is shown in Figure 1.1 (Skoglund et al., 2012; Lazaridis et al., 2014; Allentoft et al., 2015; Haak et al., 2015; Mathieson et al., 2015; Fu et al., 2016; Lipson et al., 2017; Sikora et al., 2017; Mathieson et al., 2018). We acknowledge that we take a coarse approach by simplifying human demographic history to a set of ancestral populations, who themselves are products of mixtures of the past.

All ancient modern human samples were typed on a 1240k capture array. Since we needed dense sampling of the neutral loci surrounding the selected site, we imputed genotypes in the ancient modern human samples using Beagle 4.1 (Browning and Browning, 2007, 2016). To ensure imputation accuracy, we selected ancient samples for analysis if they had $\geq$ 1x coverage. This cutoff was chosen because at this point an imputation procedure very similar to ours can correctly recover around 80-95% of an ancient sample's heterozygous sites and preserves properties of the genotypic data such as samples' PCA locations relative to no imputation (Mathieson, 2016). To perform the imputation we followed a procedure similar to that of Mathieson et al. (2015). At the 1240k sites, we computed genotype likelihoods from read counts according to binomial likelihoods of these counts and a small amount of sequencing error. We then ran Beagle twice, first to impute genotypes at just the 1240k sites using the `gl` and `impute=false` arguments, and then to impute the remaining sites using the `gt` and `impute=true` arguments. After each round of imputation we removed polymorphic sites with allelic $r^2 < 0.8$ (a measure of imputation accuracy). These intermediate filtering steps can improve the imputation accuracy of even ultra-low coverage ancient samples (Hui et al., 2020). To account for some of the uncertainty in the imputation procedure, we calculated population allele frequencies by weighting sample genotypes by their posterior probability, rather than using the maximum likelihood genotypes. We imputed each population separately, used the HapMap Project's genetic map (The International HapMap Consortium, 2007), and used a reference panel consisting of all East Asian and European

1000 Genomes populations (as multi-population panels can improve imputation accuracy in untyped populations; Huang et al., 2009). Finally, to assess the impact of imputation uncertainty on our results, we also ran our method on bootstrapped datasets, which we describe after introducing the method.

Choice of Neanderthal introgressed regions We analyzed genomic regions that were previously identified as putative candidates of adaptive introgression by Racimo et al. (2017). We chose regions with signals of adaptive introgression in European populations, as we have more information on their ancestral populations to infer the age of selection. We removed any regions whose introgressed sequences could not be distinguished from Denisovan ancestry (according to Racimo et al., 2017) and any regions on the X chromosome. To ensure we were capturing the full window of potentially selected sites, we extended the 45kb windows identified by Racimo et al. (2017) by 2×10^{-2} cM (20kb if the window has a constant per base pair recombination rate of 10^{-8}) and collapsed those that were overlapping or directly adjacent. This resulted in 36 distinct regions, listed in Table 1.2, out of 50 45kb windows with signals of adaptive Neanderthal introgression in Europeans.

Neanderthal ancestry over time As a first step to understand the temporal history of selection, we investigated levels of Neanderthal ancestry in ancient and present day admixed modern human populations within each of the previously specified regions. These levels were determined from ancestry informative sites, which we identified as bi-allelic sites in modern humans that have one allele fixed in Neanderthals (combined Vindija and Altai, two high coverage Neanderthals) and at less than 5% frequency in Yoruba. We included the Altai Neanderthal sample here to increase our confidence that we identified true fixed differences between Neanderthal and modern human lineages. In Figure 1.2, we show average Neanderthal allele frequencies across all ancestry informative sites with a Neanderthal allele frequency of at least 20% in at least one population. We use this subset of ancestry infor-

mative sites because we are interested in frequencies along the selected haplotype(s), which do not necessarily span the entire length of the previously described windows. If selection quickly favored introgressed alleles, we would at least expect all admixed modern human populations, whether sampled in the past or present, to show high levels of Neanderthal ancestry. However, in a number of these genomic regions, Neanderthal ancestry is almost or completely absent from East Asian and/or some ancient populations (Figure 1.2).

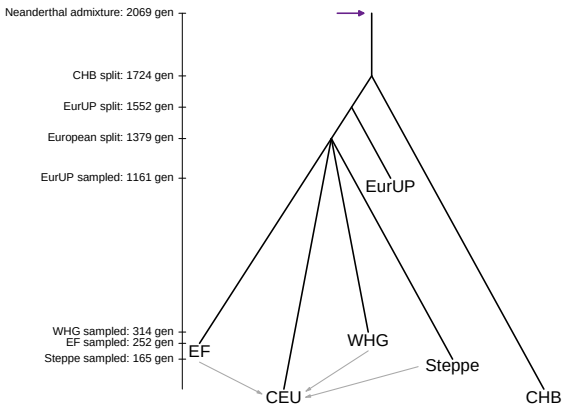

Figure 1.1: The primary admixture graph that we use in our method. Our method requires the divergence times among admixed modern human populations, the time and proportion of admixture with Neanderthals, and the sampling time of each population. It also requires the timing and proportions of admixture among modern human populations, which we describe in detail in Table 1.1. We converted many of these estimates from years to generations assuming a generation time of 29 years (Fenner, 2005). Our method is robust to modifications to this admixture graph, which we discuss later.

1.3 Method to estimate the timing of selection from introgressed haplotypes

1.3.1 Verbal Description

We first describe the intuition behind our models, before laying out the mathematical framework. We focus on how patterns of haplotype similarity among modern human populations

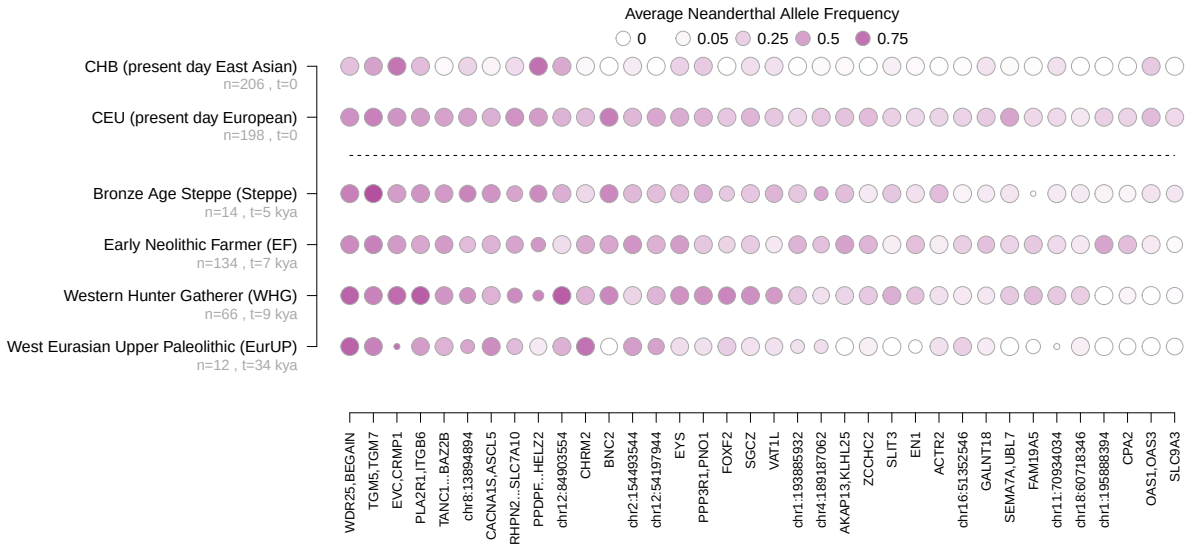

Figure 1.2: Average Neanderthal allele frequency at ancestry informative sites in ancient and present day populations in genomic regions with signatures of adaptive introgression. Darker shades of purple correspond to higher average Neanderthal allele frequencies. The size of points correspond to the number of sites over which the average Neanderthal allele frequency was calculated. Moving left to right, regions have decreasing average Neanderthal allele frequencies among ancient populations.

and Neanderthals change with the start time of selection. Since we describe selection favoring a Neanderthal introgressed allele, the earliest selection can begin is the time of admixture between Neanderthals and modern humans. To accompany our description, in Figure 1.3 we show a cartoon of haplotype diversity and introgressed ancestry under two scenarios: immediate selection and recent selection.

Neanderthal alleles that introgress into modern human populations start on long Neanderthal ancestry tracts (entire chromosomes at the time of admixture) that shorten over time with recombination. Since the selected Neanderthal allele takes its linked variation to high frequency with it, the timing of selection determines the genetic distance over which neutral Neanderthal alleles sweep to high frequency as well. The earlier selection is, the larger

the Neanderthal haplotypes that sweep to high frequency with the selected Neanderthal allele. Conversely, the later selection is, the shorter the Neanderthal haplotypes that sweep to high frequency, such that some haplotypes in modern humans that did not introgress from Neanderthals (which we call 'modern human haplotypes') also rise in frequency.

Because Neanderthal genetic diversity is very low, haplotypes that introgressed from Neanderthals into modern human populations should look almost identical to each other and to haplotypes sampled from Neanderthals. Conversely, Neanderthal haplotypes are relatively distinct from modern human haplotypes compared to differences typically observed among modern human haplotypes. This is because the earliest time that modern human and Neanderthal haplotypes can coalesce is when Neanderthals and modern humans shared a common ancestor, about 16,000 generations ago. Therefore, when selection brings Neanderthal haplotypes to high frequency in those modern human populations where the Neanderthal allele experienced positive selection (hereafter 'selected populations'), selected populations gain unusually high sequence similarity with Neanderthals and unusually low sequence similarity with other modern human populations in which Neanderthal haplotypes are rare in the genomic region affected by the sweep. These patterns of unusually high and low sequence similarity persist over greater genetic distances as selection begins closer to the time of introgression.

Within selected populations, selection increases sequence similarity because the sampled haplotypes descended from one or a few ancestral haplotypes that hitchhiked to high frequency during the sweep. When selection begins earlier, these ancestral haplotypes are mainly one of the few Neanderthal haplotypes introduced by admixture. The later selection begins, the more likely these ancestral haplotypes are modern human haplotypes that became linked to the selected allele prior to the selection onset.

In the cases of adaptive introgression that we investigated, the putative selected Neanderthal allele did not reach fixation, either because the sweep is ongoing and selection is weak,

selection pressures changed before the Neanderthal allele reached fixation, or the phenotypic response to selection was achieved by allele frequency changes at multiple loci. Our models allow for all of these possibilities, where if selection stopped favoring the Neanderthal allele we consider a neutral phase after the sweep. From this phase, we can further distinguish among selection times based on how recombination distributes Neanderthal ancestry among haplotypes within the same population that do and do not carry the selected Neanderthal allele. As the partial sweep finishes, the Neanderthal alleles that hitchhike to higher frequency with the selected allele have a higher chance of recombining onto the background of the non-selected allele. The earlier the onset of selection, the more time post-sweep for these recombination events to occur, and therefore the higher the probability that alleles on the non-selected background descended from Neanderthals.

1.3.2 Model Background

Our aim is to distinguish among the possible scenarios of selection on introgressed variation. Each scenario is defined by the combination of two parameters we aim to infer in our model: the amount of time between Neanderthal admixture and the onset of selection, which we refer to as the waiting time until selection (t_b) and the additive strength of selection favoring the beneficial Neanderthal allele (s). We build on the model-based, statistical approach introduced in Lee and Coop (2017), which uses coalescent theory to describe how different selection scenarios modify the neutral variance and covariance of population allele frequencies surrounding the selected site. This allows us to describe a multivariate normal model of population allele frequencies for each scenario, which serves as a simple approximation for their probability distribution (Weir et al., 2002; Nicholson et al., 2002).

Here, we review the framework laid out by Lee and Coop (2017). We model the change in allele frequency $\Delta x_i = x_i - x_a$ in our sampled populations (i) from their common ancestral population (a) at the root of the tree relating all populations we consider. For neutral alleles,

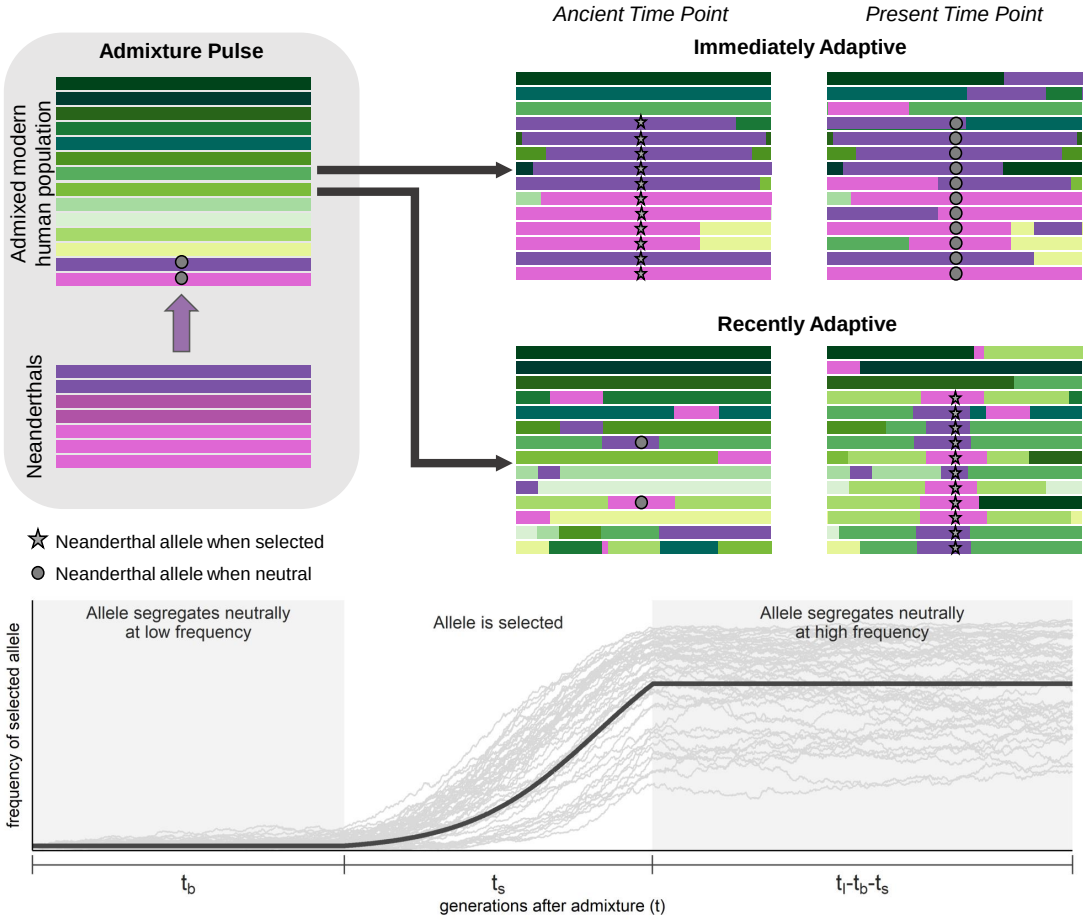

Figure 1.3: (Top) Cartoon of haplotype diversity and ancestry at three time points when selection favors Neanderthal introgressed alleles immediately ($t_b = 0$) and when selection favors Neanderthal introgressed alleles more recently ($t_b > 0$). Our models use a specific value of t_b, but here our purpose is to visualize qualitative differences between early and recent selection. Purple tracts represent Neanderthal introgressed haplotypes, whereas green tracts represent haplotypes that did not descend from Neanderthals. We display Neanderthal ancestry tracts with fewer shades to represent their lower genetic diversity. (Bottom) Example frequency trajectory of a selected Neanderthal introgressed allele when there is a waiting time until selection. The light grey lines represent 50 stochastic trajectories in which the selected allele reached a frequency greater than 20%. The dark grey line represents the trajectory we assume in our models using the same waiting time until selection (t_b), selection coefficient (s), and duration of the sweep phase (t_s) used to generate the stochastic trajectories.

the population allele frequency will on average be the same as the ancestral allele frequency ($\mathrm{E}[\Delta x_i] = 0$) because drift and hitchhiking are direction-less on average. Drift and hitchhiking do however cause an increase in the variance in the change in allele frequencies within a population, and pairs of populations can also covary in their change in allele frequency from the ancestor if they have some shared population history or gene flow, i.e. their changes in allele frequency since the ancestral population are not independent of one another. These effects are captured by the population covariance between populations, i and j, given by

$$\mathrm{Cov}[\Delta x_i, \Delta x_j] = x_a(1 - x_a)f_{ij}, \tag{1.1}$$

where f_{ij} is the probability that the ancestral lineages of an allele sampled in population i and an allele sampled in population j coalesce before reaching the ancestral population (see Lee and Coop, 2017, for details).

We can estimate the neutral probabilities of coalescing among populations from neutral allele frequency data genome-wide. If we are considering k populations in our analysis, we thus have a $k \times k$ matrix, $\mathbf{F}$, that describes probabilities of coalescing within and between these populations. See Appendix 1.8.2 for details on how we estimate this matrix. These estimated neutral probabilities of coalescing allow us to describe our neutral expectations without making any assumptions about the demographic history of populations: they implicitly account for population size, divergence, and migration. We note, however, that while this flexible approach allows for arbitrary relationships among populations, our tree requires a root that our coalescent probabilities and frequency deviations will be relative to. We define this root by placing the two most genetically distant populations on opposite sides of it. This implies that in our models, they have no shared history since the ancestral population, and therefore no chance for coalescence between the ancestral lineages of alleles sampled in each. In our analysis, the Neanderthals and Yoruba are the most distantly related, and therefore are placed on opposite sides of the root, without any subsequent gene flow. This

does not negate the possibility of indirect gene flow between them; we simply define all other relationships relative to theirs.

When we incorporate the effects of selection, we model how probabilities of coalescing change with increasing genetic distance from the selected site. Importantly, all of our predictions converge to our neutral estimates because increasing recombination rates between selected and neutral sites allows for increasing independence of their coalescent history. At a far enough genetic distance, dynamics at the selected site become quickly disassociated with those at the neutral site, such that the probability of coalescing at the neutral site is that of any neutral allele unaffected by linked selection.

1.3.3 Haplotype Partitioning and the Null Model

No identified case of adaptive Neanderthal introgression has resulted in the population fixation of a Neanderthal-derived allele. These partial sweeps weaken the effect of linked selection on surrounding neutral diversity because much of the original genetic diversity prior to selection persists in these populations. In order to increase the power of our approach, we divide admixed populations according to ancestry assignments at the putative selected site. For the 1000 Genomes samples, from which we have phased haplotypes, we create two partitions: one consisting of haplotypes that carry the beneficial Neanderthal allele at the selected site (B), and the other consisting of haplotypes that carry the non-Neanderthal allele at the selected site (b). For the ancient samples, we only have unphased genotype data, and so we divide the population samples into three partitions according to individuals' ancestry genotypes at the selected site: BB, Bb, and bb. In our models, we are concerned with a neutral allele's ancestry background at the selected site when it is sampled. Therefore, haplotype partition B is equivalent to genotype partition BB in that any neutral allele sampled in these partitions is linked to the Neanderthal allele at the selected site. Similarly, haplotype partition b is equivalent to genotype partition bb in that any neutral allele sampled in these partitions

is linked to the non-Neanderthal allele at the selected site. From here on we refer to them in our models as partition B or partition b. Predictions for genotype partition Bb are simply a linear combination of our predictions for the other partitions, which we describe in section 1.9.3.5.

Our haplotype partitioning of the data requires us to modify our null model from that of Lee and Coop (2017), who focused on full sweeps and simply used the genome-wide neutral **F** matrix to parameterize their null model. The neutral probabilities of coalescing that we estimate between any pair of populations (i, j) are essentially averaged over all possible migration histories of our alleles. However, our population partitioning scheme alters the probability of each migration history because we split populations based on ancestry at the partition site. When we create a population that only carries Neanderthal alleles at the partition site (partition B), nearby sites will have a much higher frequency of Neanderthal alleles relative to the full population. Therefore, even under neutrality, the probability that a randomly sampled allele in partition B will coalesce with an allele sampled in Neanderthals is much higher than in the full population case.

Under population partitioning, we can describe our null coalescent probabilities as a function of the genome-wide neutral probabilities of coalescing and the recombination rate (r) between the partition site and neutral site of interest. We first consider the relationship between Neanderthal population n and partition B of population i in which all haplotypes carry a Neanderthal allele at the partition site. We define t_I as the number of generations ago that Neanderthal alleles introgressed into modern human populations. We are interested in the probability that the neutral lineage remains linked to the same Neanderthal allele at the partition site that it was linked to at sampling by the time of admixture. We can approximate this probability as e^{-rt_I}. By remaining linked to the Neanderthal allele at the partition site, we know that it, too, introgressed from Neanderthals and thus will coalesce with the Neanderthal allele with approximately the same probability as any two alleles

sampled from Neanderthals (f_{nn}). If it does recombine out at some point before admixture, it has the population's neutral probability of coalescing with Neanderthals. This neutral probability of coalescing still accounts for the possibility that the neutral allele descended from Neanderthals, when the average Neanderthal-introgressed allele frequency is the initial admixture fraction (g). We define probabilities of coalescing under the null model as $f_{in}^{(N)}$, where superscript (N) refers to predictions under our null model. In all we derive the probability of coalescing between a pair of lineages sampled from partition B of population i and Neanderthal population n to be

$$f_{in}^{(N)} = e^{-rt_I} f_{nn} + (1 - e^{-rt_I}) f_{in}, \text{ if sampled from partition } B. \tag{1.2}$$

In partition b of the same population (i), all alleles are linked to the non-Neanderthal allele at the partition site when sampled. If an allele sampled in this partition never recombines out, we know it cannot have introgressed. Therefore in order to have the chance at coalescing with the ancestral lineage of the sampled Neanderthal allele, it must recombine out before admixture. Forward in time, that means that a Neanderthal allele recombined onto the sequence carrying the non-Neanderthal allele at the partition site. Therefore the ancestral lineage of a neutral allele sampled from partition b of population i has the following probability of coalescing with the neutral lineage sampled from Neanderthals:

$$f_{in}^{(N)} = (1 - e^{-rt_I}) f_{in}, \text{ if sampled from partition } b. \tag{1.3}$$

1.3.4 Selection Model

Selection favoring Neanderthal alleles in modern human populations modifies probabilities of coalescing from neutral expectations because it increases levels of Neanderthal ancestry around the selected site in selected populations. Thus, in the whole (non-partitioned) population, the probability that an allele descended from Neanderthals is much higher than the

original admixture proportion, as we described informally above in our discussion of Figure 1.3.

1.3.4.1 Derivation

In selected populations, we consider three phases in the selected allele frequency trajectory: the neutral phase following admixture in which the Neanderthal-derived variant is at frequency g for t_b generations (neutral phase I), the sweep phase in which the variant rises from frequency g to frequency x_s in time t_s, and the neutral phase between the sweep finish and the present in which the variant remains at frequency x_s (neutral phase II). When the relative fitness advantage of the selected allele is additive, such that heterozygotes have an advantage of s and homozygotes $2s$, the sweep duration

$$t_s = \frac{1}{s} \log \left(\frac{x_s(1-g)}{g(1-x_s)} \right), \text{if } x_s < 1. \tag{1.4}$$

Since the sum of all three phase durations equals the time between the present day and admixture (t_I), the duration of neutral phase II equals $t_I - t_b - t_s$.

As the waiting time until selection t_b increases, and assuming the same t_b for all selected populations, we transition from describing a case in which the selected allele became beneficial in the common ancestor of all Eurasian populations, to the case in which the selected allele became beneficial independently in each selected population. If t_b is short enough such that selection began in the common ancestor of a group of selected populations, we must account for the possibility that this common ancestor also has descendent populations that do not carry the selected allele at high frequency, possibly due to subsequent drift or negative selection. According to t_b, we assign each population with very low frequency of the selected allele into one of two categories: (i) ancestors never selected or (ii) ancestors selected with subsequent loss of the selected allele in this population. We assign the first category if the low frequency population does not share a common ancestor with any selected populations at the time selection starts, i.e. its divergence from all selected populations predates selection.

Otherwise we assign the second category.

Similar to the null model, we condition on ancestry at the putative selected site and consider how recombination modulates ancestry with increasing genetic distance. The most important difference between the null and selection model is neutral phase II: in the selection model, if a neutral lineage recombines, it has a high probability of recombining onto the selected allele's background. Therefore, there is a higher chance that the neutral lineage itself descended from Neanderthals.

Between Neanderthals and selected populations In this section we derive the probabilities of coalescing between the haplotypes in each partition of selected human populations and Neanderthals. When we sample an allele, we know whether or not it is linked to the selected Neanderthal allele based on the partition it belongs to. Conditioning on this ancestry background at sampling, we focus on whether the ancestral lineage is linked to the selected Neanderthal allele when we transition between each phase looking backwards from the present to admixture.

First, we determine the probability that a neutral lineage is linked to the selected Neanderthal allele at the transition point between neutral phase II and the sweep completion (which occurs at time $t_I - t_b - t_s$). If there is never a recombination event between the selected and neutral site during neutral phase II, then the neutral lineage remains associated with its initial ancestry background at the time of the sweep completion. Alternatively, if at least one recombination event occurs, then the final recombination event determines the ancestry background that the neutral lineage is associated with at the time of the sweep completion. The frequency of the selected allele determines the probability that a recombining neutral lineage becomes associated with the selected background. Thus with probability x_s, the neutral lineage becomes associated with the selected Neanderthal allele, and with probability $1 - x_s$ it becomes associated with the non-Neanderthal allele at the selected site.

Therefore, the probability that a neutral lineage sampled from a selected population is linked to the selected Neanderthal allele at the transition point between neutral phase II and the sweep completion is

$$\begin{aligned}
&\Pr(\text{linked to selected allele at sweep finish}) \\
&= \begin{cases} \exp\left(-r(t_I - t_b - t_s)\right) + \left(1 - \exp\left(-r(t_I - t_b - t_s)\right)\right)x_s, \\ \text{if sampled from partition } B. \\ \left(1 - \exp\left(-r(t_I - t_b - t_s)\right)\right)x_s, \\ \text{if sampled from partition } b. \end{cases}
\end{aligned} \tag{1.5}$$

in which an allele sampled from partition b can only be linked if it recombines off of its background.

Second, during the sweep phase, an allele that begins linked to the selected allele will always be associated with that background with probability

$$\begin{aligned}
&\Pr(\text{always linked to selected allele during sweep}) \\
&\approx \exp\left(-r\int_0^{t_s} 1 - r(1 - X(t))dt\right) \\
&= \left(\frac{x(1-g)}{g}\right)^{-\frac{r}{s}},
\end{aligned} \tag{1.6}$$

where $X(t)$ is the frequency of the selected allele in generation t of the sweep, following a deterministic, logistic trajectory. In words, this probability is approximately the product of the probabilities of not recombining onto the non-selected background each generation of the sweep. Similarly, an allele that begins the sweep phase linked to the non-selected allele will always be associated with that background during the sweep with probability

$$\begin{aligned}
&\Pr(\text{always linked to non-selected allele during sweep}) \\
&\approx \left(\frac{1}{x_s}\right)^{-\frac{r}{s}}, \text{ if } x_s < 1.
\end{aligned} \tag{1.7}$$

The above associations persist if the lineage never recombines out of its background during neutral phase I, with approximate probability e^{-rt_b}. So, if a neutral lineage remains linked to the selected allele during the sweep phase and fails to recombine during neutral phase I, it must have descended from Neanderthals and thus coalesces with the lineage sampled from Neanderthals with approximately the same probability as any two lineages sampled from Neanderthals, f_{nn}. If the neutral lineage remains linked to the non-selected allele throughout the sweep phase and fails to recombine during neutral phase I, it definitely did not descend from Neanderthals and thus cannot coalesce with an allele sampled from Neanderthals. If a lineage becomes disassociated with its background at some point during the sweep phase, and/or recombines out at least once during neutral phase I, we assume it has its population's neutral probability of coalescing with Neanderthals. Our approximation ignores the possibility that a lineage linked to the non-selected allele can recombine onto the selected allele's background during the sweep. In total, the ancestral lineage of an allele sampled from an admixed, selected population coalesces with the ancestral lineage of an allele sampled from Neanderthals with probability

$$
\begin{aligned}
&f_{pn}^{(S)} \\
&= \Big[\Pr(\text{linked to selected allele at sweep finish}) \\
&\times \Pr(\text{always linked to selected allele during sweep}) e^{-rt_b} f_{nn} \Big] \\
&+ \Big[\Pr(\text{linked to selected allele at sweep finish}) \\
&\times \big(1 - \Pr(\text{always linked to selected allele during sweep}) e^{-rt_b}\big) \\
&+ \big(1 - \Pr(\text{linked to selected allele at sweep finish})\big) \\
&\times \big(1 - \Pr(\text{always linked to non-selected allele during sweep}) \\
&\times e^{-rt_b}\big) \Big] f_{pn},
\end{aligned}
\tag{1.8}
$$

where the superscript S in $f_{pn}^{(S)}$ refers to our predictions under the selection model, and subscript p denotes any partition of any selected population. In sections 1.9.3.1 and 1.9.3.2 we illustrate predictions for other population relationships under the selection model. We follow by describing modifications under both the null and selection models to incorporate ancient samples (S3.3), migration among admixed modern human populations (S3.4), and genotype partition Bb (S3.5). Our approximations provide a good fit to those obtained by simulations, see Figure 1.4.

1.3.5 Inference

For each model (null or selection) and combination of free parameters, we define a variance-covariance matrix of population allele frequencies for a given distance away from the selected site (Lee and Coop, 2017). We approximate the joint probability distribution of population allele frequencies at a neutral locus (l) as being multivariate normal around the ancestral allele frequency (x_{al}) with covariance equal to $x_{al}(1-x_{al})$ times the modified matrix of coalescent probabilities ($\mathbf{F^{(N)}}$ or $\mathbf{F^{(S)}}$) (Weir et al., 2002; Nicholson et al., 2002; Samanta et al., 2009; Coop et al., 2010). In both the null and selection models, the modified matrices of coalescent probabilities depend on the genetic distance to the selected site (r_l), the genome-wide neutral matrix of coalescent probabilities ($\mathbf{F}$), and all of the admixture graph parameters of interest (A_G), which describe admixture proportions and timing among Neanderthal and admixed modern human populations as well as divergence and sampling times of admixed modern human populations (see Figure 1.1 and Table 1.1). The selection model's $\mathbf{F^{(S)}}$ depends on three additional parameters: the strength of selection (s), the waiting time until selection (t_b), and the final frequency of the selected allele (x_s). Therefore at a neutral polymorphic site with allele frequency data in our populations, we can estimate the probability of the observed population allele frequencies ($\vec{x_l}$) under the selection model as

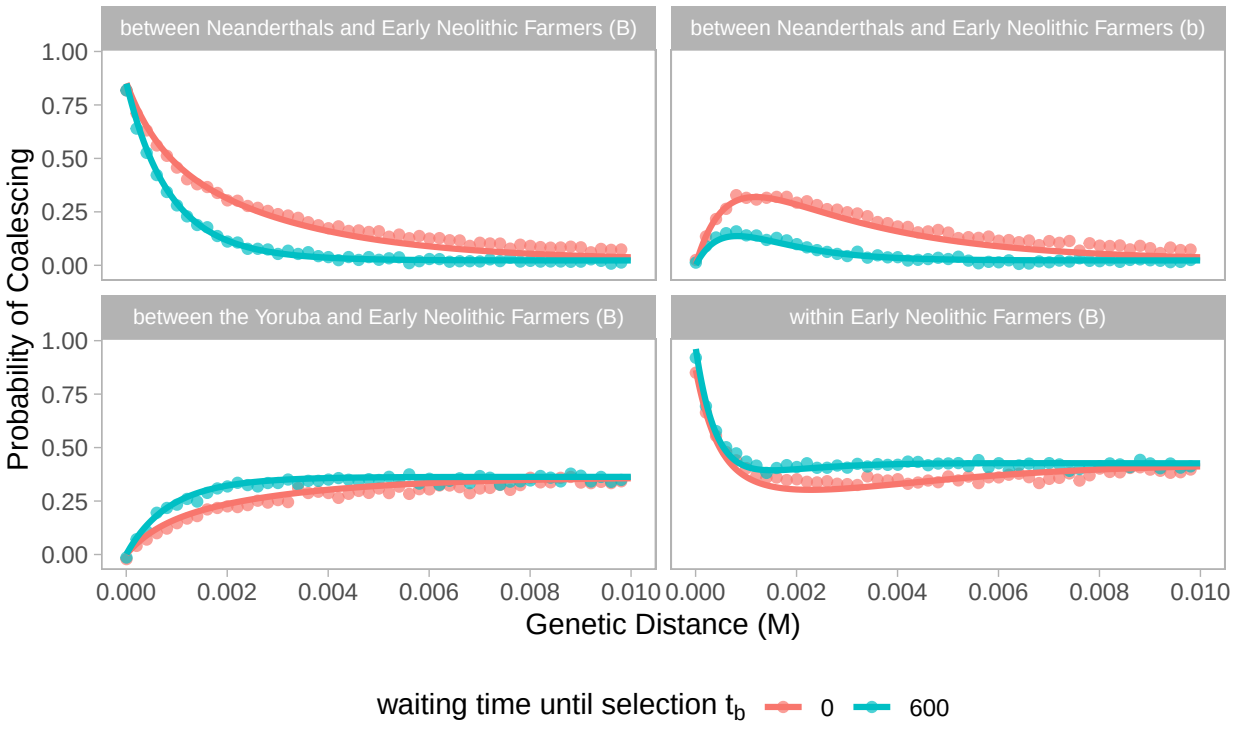

Figure 1.4: Probabilities of coalescing between pairs of alleles at increasing genetic distances from the selected site. Analytical predictions (lines) closely match estimates averaged across 250 simulations (points). Here, we show results for a case in which the Neanderthal derived allele became adaptive in the ancestors of Early Neolithic Farmers immediately upon introduction (t_b=0, red) and after 600 generations (blue). The favorable allele experienced a selective advantage (s) of 0.01 and on average reached a frequency (x_s) of 0.7 before becoming neutral again. The genetic window corresponds to 1 Mb with a constant recombination rate of 10^{-8}. Partition B refers to alleles linked to the beneficial Neanderthal allele, while partition b refers to alleles linked to the non-Neanderthal counterpart at the selected site. We present categories of population relationships in which probabilities of coalescing vary the most with different waiting times until selection. See Figures 1.20 and 1.21 for more population relationships, waiting times until selection, and final frequencies of the selected allele.

$$\begin{aligned}
&\mathrm{P}(\vec{x_l}|r_l, \mathbf{F}, A_G, s, t_b, x_s) \\
&\quad \approx \mathcal{N}\left(\vec{x_l}|x_{al}, x_{al}(1-x_{al})\mathbf{F^{(S)}}(r_l, \mathbf{F}, A_G, s, t_b, x_s)\right),
\end{aligned} \tag{1.9}$$

where the set of parentheses following $\mathbf{F^{(S)}}$ denote the parameters of the function $\mathbf{F^{(S)}}$. When we predict $\mathbf{F^{(S)}}$, we categorize Neanderthal-admixed populations as 'selected' if their frequency of the Neanderthal allele at the selected site is greater than or equal to 0.05. We set x_s to be the average selected Neanderthal allele frequency among all of these putative selected populations. Our method could be extended to allow x_s to vary among populations and to do model choice of the selected populations, however for simplicity we do not pursue those applications here. We estimate the composite likelihood of free parameters s and t_b given the allele frequency data across populations (D) in a window around a candidate selected site by taking the product of all likelihoods calculated at each locus to the left and right of the candidate selected site as follows,

$$\begin{aligned}
\mathcal{L}_{\mathcal{C}}(s, t_b; D) = &\prod_{i=1}^{\mathrm{L_{left}}} \mathrm{P}(\vec{x_i}|r_i, \mathbf{F}, A_G, s, t_b, x_s) \\
&\times \prod_{j=1}^{\mathrm{L_{right}}} \mathrm{P}(\vec{x_j}|r_j, \mathbf{F}, A_G, s, t_b, x_s),
\end{aligned} \tag{1.10}$$

where $\mathrm{L_{left}}$ and $\mathrm{L_{right}}$ represent the number of polymorphic loci to the left and right of the candidate selected site. We calculate composite likelihoods among the proposed parameter combinations of s and t_b shown in Table 1.3.

The composite likelihood under the null model takes the same form as selection; we simply replace $\mathbf{F^{(S)}}$ with $\mathbf{F^{(N)}}$ and remove dependence on s, t_b, and x_s. Each partition site represents a potential selected site and thus differs in its genetic distance to each neutral locus. Therefore, each partition site has its own set of composite likelihoods for the null and selection models. We identify the 'best' partition site for each region by selecting the site whose ratio of its maximum composite likelihood under the selection model to the null

model is greatest, i.e. by maximizing the level of support for selection at this site. From this partition site, we identify the maximum composite likelihood estimates $\hat{t_b}$ and $\hat{s}$ for the region. We note that in our application, these estimates tended to be very similar among partition sites.

Our method focuses on inferring the history of selection from haplotypes alone, and does not incorporate the distribution of the selected allele frequency across populations (as in Racimo et al., 2018; Refoyo-Martínez et al., 2019). We could incorporate into our composite likelihoods the conditional probability of x_s across populations under a given parameterization of the selection model, but here our focus is on the information about the timing of selection contained in haplotype patterns.

We analyze bi-allelic sites that are polymorphic among our set of samples. Since our multivariate normal approximation works best when alleles segregate ancestrally at some intermediate frequency, we remove sites polymorphic in only one population with a minor allele frequency less than 0.01. In practice, we mean centered our observed allele frequencies, which removes dependence on the ancestral allele frequency x_{al}. We provide more detail on mean centering, sample size correction, and implementation in Appendix 1.8.3.

Distinguishing between immediate and non-immediate selection To assess the timing of selection relative to introgression we first distinguish between immediate ($t_b = 0$) and non-immediate ($t_b > 0$) selection. We do so by using the ratio of the maximum composite likelihood under all parameter combinations to the maximum composite likelihood when selection is immediate,

$$\mathrm{CLR}_{t_b>0} = \frac{\sup\limits_{s,t_b} \mathcal{L}_{\mathcal{C}}(s, t_b; D)}{\sup\limits_{s} \mathcal{L}_{\mathcal{C}}(s, t_b = 0; D)}. \tag{1.11}$$

The higher this ratio, the more support we have in favor of an initial neutral period relative to immediate selection. Since composite likelihoods ignore the correlation in allele frequencies across loci due to linkage disequilibrium, we cannot use traditional statistical methods for

model selection. Therefore we rely on simulations and reject the immediate selection case if the upper 97.5th percentile of $\text{CLR}_{t_b>0}$ from immediate selection simulations does not exceed the observed value in a region. In other words, we reject the null hypothesis of immediate selection if we observe a value of $\text{CLR}_{t_b>0}$ high enough to be unlikely if selection were truly immediate. The following sections contain more detail on the simulation procedure used to reject immediate selection.

Validation We ran our method on simulated data to evaluate its performance, using SLiM 3.0 for forward in time simulations (Haller and Messer, 2019). We simulated 2cM (2 Mb) loci with the selected mutation at the center of the locus. We simulated under the demographic history shown in Figure 1.1, along with a divergence time between the Yoruba and Eurasian populations of 2,500 generations and a divergence time between modern humans and Neanderthals of 16,000 generations. The Neanderthal population was simulated with a population size of 3,000, whereas all other populations were simulated with a size of 10,000.

Our SLiM simulations generated tree sequences, onto which we added neutral mutations and calculated population allele frequencies using the same sample sizes as in our real data in Python version 3.7.4 with `msprime` (Kelleher et al., 2016), `tskit` (Kelleher et al., 2018), and `pyslim` (Haller et al., 2019). We ran our method on these data in R version 3.4.4. The method assumed the same demographic history that we simulated and used a neutral **F** estimated from neutral allele frequency data, also produced by SLiM simulations under the same demographic history as the selection simulations. See Appendix 3.8.1 for more simulation details.

Our first goal was to determine our power to detect selection that did not begin immediately, and so we established the significance cutoff using simulations under the null model of immediate selection. Running the method on simulations of immediate selection, we identified a positive correlation between x_s, the average selected allele frequency among all

putative selected populations, and $\text{CLR}_{t_b>0}$, the composite likelihood ratio quantifying the method's support for non-immediate selection relative to immediate selection (Figure 1.9). This relationship between x_s and $\text{CLR}_{t_b>0}$ among our immediate selection simulations does not differ among combinations of the selection coefficient (s) and the duration of the sweep (t_s) (Figure 1.9). Thus, for each simulation of non-immediate selection, we rejected immediate selection if its $\text{CLR}_{t_b>0}$ was greater than the upper 97.5th percentile of $\text{CLR}_{t_b>0}$ from the 500 immediate selection simulations with the closest x_s. We used a similar procedure to determine the significance cutoffs for the real dataset.

Varying the onset time of selection in our simulations, we found that the method's power to reject immediate selection increases with the true waiting time until selection (t_b) and x_s (Figure 1.5). Our method has reasonable power to reject immediate selection when selection begins $>$ 1000 generations after admixture, suggesting that we should correctly identify many of the regions that only recently contributed to adaptation. Overall, s is not well inferred, likely because we deal with old selection events. Therefore, while s appears in our selection model, we concentrate on inferring t_b in our application. The method yields somewhat biased estimates of the waiting time until selection (Figures 1.10–1.13). However, among the simulations in which the method rejected immediate selection, the method does a good job tracking the true waiting time until selection (Figure 1.6). This performance decreases when fewer populations are considered to be selected, but still the method can identify that selection started much later in time. For more details on method performance, see section 1.9.2.1.

1.4 Data Availability

Scripts for imputation, simulations, and the method are provided at `www.github.com/SivanYair/selTime_neanderthal_AI`. The script that computes genotype likelihoods can be found at `www.github.com/mathii/gdc3/blob/master/apulldown.py`. Supplement file

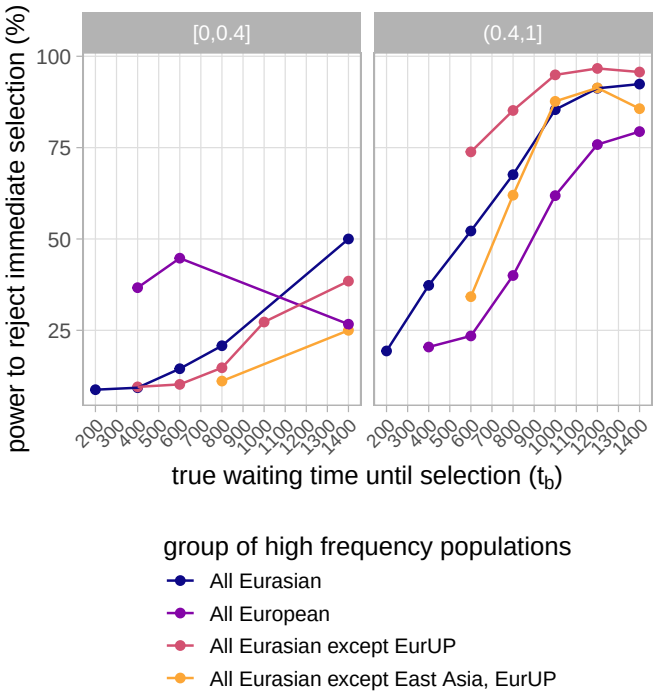

Figure 1.5: Power to reject immediate selection as the true waiting time until selection (t_b) increases. Lines correspond to simulations with different groups of high frequency populations (each with the selected allele frequency greater than 0.05), i.e. groups of populations considered selected. Note that this does not perfectly correspond to groups of populations truly selected in our simulations, and so we have different numbers of observations contributing to each point. Points with fewer than 10 observations were removed. Panels correspond to bins of x_s, the selected allele's average frequency among populations considered selected. Results for $x_s > 0.4$ are qualitatively similar, though we show finer bins in Figure 1.15.

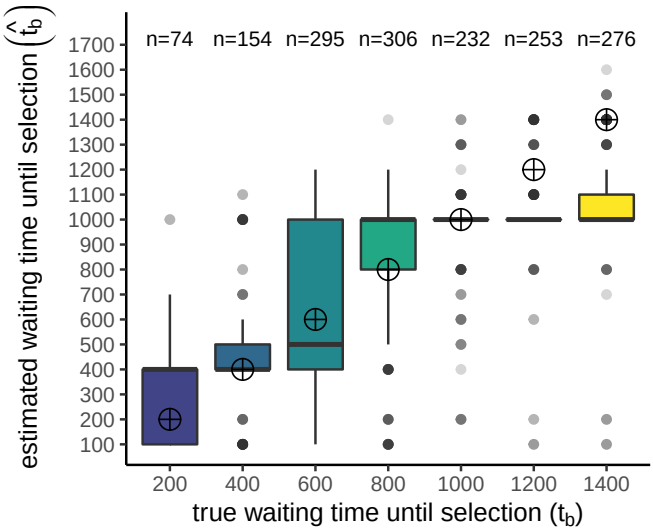

Figure 1.6: Estimated waiting time until selection ($\hat{t_b}$) among simulations with all populations considered selected (each $x_s > 0.05$) and in which the method rejected immediate selection. Target points mark the true waiting time until selection that we aim to estimate. Earlier true waiting times have fewer observations because the method has less power to reject immediate selection in this parameter space. Results for more groups of putative selected populations are shown in Figure 1.14. Estimated waiting times until selection under the immediate selection case can be found in Figure 1.9.

S2 is available on ProQuest.

1.5 Data application to estimate the timing of selection on Neanderthal alleles

Procedure for identifying selected site and parameter estimates For each region, we ran our method on the previously defined windows in addition to 1 cM flanking them on each side, because if selection is immediate with $s = 0.01$ it takes about 1 cM for the signals of selection to almost completely decay. We used the HapMap Project's genetic map to identify the endpoints of genomic analysis windows and recombination rates between each neutral and selected site for the method (note that this means that we need to assume that

recombination rates have been relatively constant over this time period). We excluded sites without Neanderthal allele frequency data from our analysis, as they are less informative of introgression patterns and the method might instead pick up on unrelated signals of selection at these sites.

We estimated the neutral **F** from putative neutral allele frequency data genome wide. We chose regions at least 300 base pairs long, at least 0.4 cM away from a gene, and with a minimum per base pair recombination rate of 0.9×10^{-8} using the 'Neutral Regions Explorer' (Arbiza et al., 2012). This led to 105,786 sites after filtering. While these sites may not be entirely unaffected by linked selection, we can still use them to represent background patterns of genetic diversity from which we distinguish strongly altered patterns caused by adaptive introgression.

Among the ancestry informative sites in each region, we chose partition sites (which represent potential selected sites) to be all sites with at least two ancient populations with data and a Neanderthal allele frequency $\geq 20\%$ in at least one European, East Asian, or South Asian 1000 Genomes population. For a region, we looped over possible partition sites to run the method on their corresponding dataset, and then selected among these partition sites. Then from the 'best' partition site we selected among parameter estimates of t_b and s. For example, in Figure 1.7 we show the profile composite likelihood surface of t_b at the maximum composite likelihood partition site (relative to the null model at this site) in the region OAS1,OAS3. The peak in this surface corresponds to our estimate of t_b, which in this region we found to be very high. As the relationships among our populations are only partially understood, we ran the method with the Altai Neanderthal sample included in the Neanderthal population (discussed in section 1.9.2.2 and Figure 1.18) and with multiple plausible modifications to the admixture graph in Figure 1.1 (discussed in section 1.9.2.2 and Figure 1.19). To account for the uncertainty in our imputation procedure, in each region we reran the method 40 times on bootstrapped datasets: rather than using the maximum

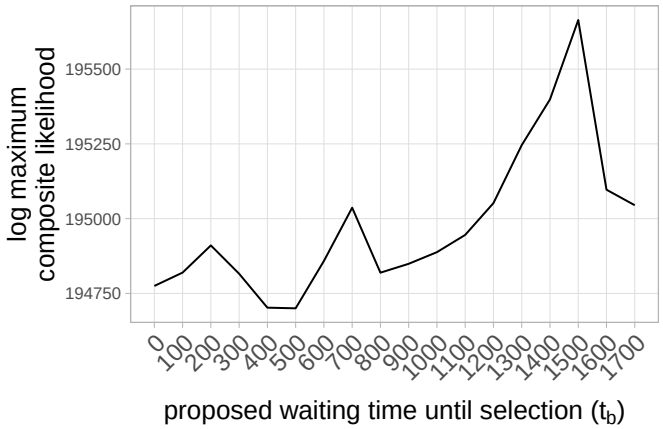

Figure 1.7: Profile composite likelihood surface of the waiting time until selection (t_b) in the region OAS1,OAS3.

likelihood genotype at the partition site to assign an ancient sample to a genotype partition, in each bootstrap run we randomly assigned the sample to one of the partitions according to its posterior genotype probabilities provided by the imputation algorithm (results shown in Figure 1.17).

Procedure for distinguishing between immediate and non-immediate selection Among the regions in which the method estimated a non-immediate waiting time until selection ($\hat{t_b} > 0$), we rejected immediate selection if their $\text{CLR}_{t_b>0}$ exceeded the upper 97.5th percentile of those from the 500 immediate selection simulations with the closest x_s (the average selected allele frequency among all putative selected populations). Recall that we can only identify these significance thresholds with simulations. Since composite likelihoods may be sensitive to the spatial distribution of analyzed sites and their corresponding populations with known allele frequencies, we sampled sites to analyze in the simulated data using the same spatial patterning as the region of interest. Specifically, we binned the genetic positions of analyzed sites into 10^{-3} cM intervals to the left and right of the selected site. For each of these sites, we sampled a site in the simulated data from the same genetic distance bin and

masked its allele frequencies from populations with no data. We used the same immediate selection simulations as in the Validation section, except this time evaluating them over a larger 3 Mb (3 cM) locus .

Of the 36 regions we ran our method on, our method estimated $\hat{t_b} > 0$ in 26, and rejected immediate selection in 17 (Figure 1.8). Four of the regions in which we rejected immediate selection had a standard deviation greater than 100 generations of $\hat{t_b}$ in their bootstraps over the imputation uncertainty: CHRM2, chr1:193885932, chr2:154493544, and chr12:84903554 (Figure 1.17). For the remaining 13 regions in which we have greater confidence in our estimated waiting times, nine of these regions have $\hat{t_b} \leq 700$, whereas the remaining four have extremely high estimates: SEMA7A,UBL7 has $\hat{t_b} = 1300$ and CACNA1S,ASCL5; OAS1,OAS3; and chr4:189187062 have $\hat{t_b} = 1500$.

As with many selection scans, the function of alleles underlying adaptation are obviously unknown. Among the set of regions with a shorter initial period at low frequency, BNC2 is a better-studied candidate of adaptive Neanderthal introgression, where the likely selected Neanderthal variant is associated with lighter skin pigmentation (Dannemann et al., 2017). As for the other regions with early selection, mutations in EYS affect the development and maintenance of photoreceptors in the retina (Collin et al., 2008; Abd El-Aziz et al., 2008; Alfano et al., 2016), proteins encoded by EVC are involved in the regulation of Hedgehog signaling and mutations in this gene affect patterning during development (Ruiz-Perez et al., 2000; Blair et al., 2011; Caparrós-Martín et al., 2013; Pusapati et al., 2014), proteins encoded by CRMP1 may be involved in the development of the nervous system and epithelial sheets (Nakamura et al., 2014; Yu-Kemp et al., 2017), and SLC7A10 contributes to synaptic regulation in the central nervous system (Ehmsen et al., 2016; Palaćin et al., 2016; Mesuret et al., 2018). HELZ2 is involved in regulating the differentiation of adipocytes and could thus play a role in metabolism (Katano-Toki et al., 2013). PLA2R1, ITGB6, and SLIT3 putatively play a role in tumor suppression, however mutations at all of these loci are also

known to influence variation in other phenotypes (Marlow et al., 2008; Hezel et al., 2012; Guo et al., 2013; Bernard and Vindrieux, 2014; Yu et al., 2014; Ng et al., 2018; Yoshikawa and Asaba, 2020). Most research on RHPN2 and PTK6 has focused on the role of their expression in cancer progression (e.g. Zheng et al., 2010; He et al., 2015). The remainder of regions in this category are poorly studied or intergenic and far from genes. As for the recently selected regions, SEMA7A plays a role in both the immune and nervous systems, regulating active T cells in the former and axon growth and formation during development in the latter (Pasterkamp et al., 2003; Suzuki et al., 2007; Gras et al., 2013). Polymorphisms at this locus are also associated with variation in bone mineral density (Koh et al., 2006). UBL7 may play a role in ubiquitin signaling, however little is known about the effects of variation at this locus (Zhang et al., 2017). CACNA1S encodes a protein known to contribute to the structure of calcium channels involved in skeletal muscle contraction (Tanabe et al., 1990; Ptáček et al., 1994; Ertel et al., 2000) and in some cases can influence body heat regulation (Beam et al., 2017). ASCL5 is less well studied, though mutations in this gene have been implicated to affect tooth development (He et al., 2019). OAS1 and OAS3 haplotypes can influence innate immunity, which we discuss in more detail later. Based on what we currently understand about mutations in these genomic regions, the non-immediate benefits provided by Neanderthal alleles may have varied widely and were perhaps associated with different environmental factors.

1.6 Discussion

Here we use patterns of linked selection to develop one of the first model-based methods that infers the time at which Neanderthal introgressed alleles became adaptive. Our approach uses ancient DNA as well as the hitchhiking effect to investigate the temporal history of selection. We directly address partial sweeps, which likely reflect most cases of human adaptation and perhaps adaptive introgression more generally, and use them as a tool to time the onset

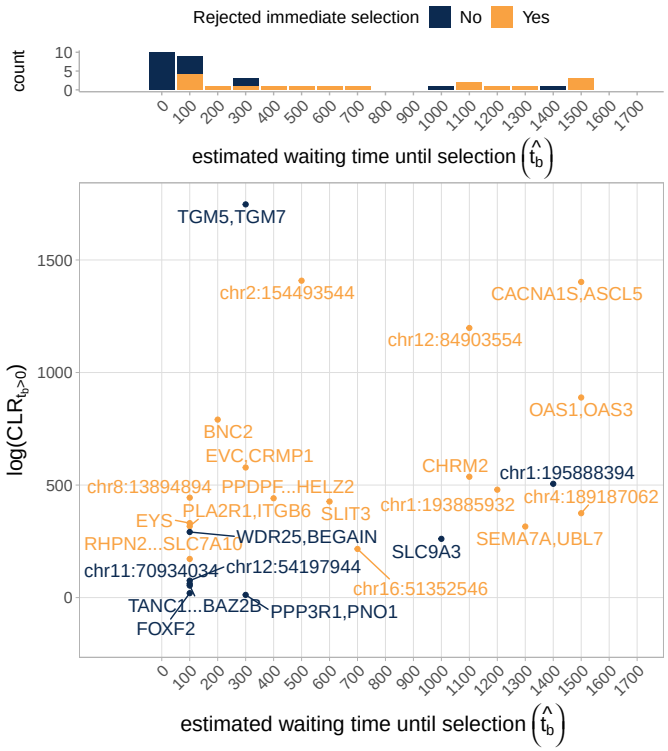

Figure 1.8: (Top) Distribution of estimated waiting times until selection ($\hat{t}_b$) in the application of the method to 36 regions of adaptive introgression. Blue bars represent cases in which the method did not reject immediate selection, and orange bars represent cases in which the method did reject immediate selection. Blue bars are stacked on top of orange bars (they do not overlap). (Bottom) Results among all regions with $\hat{t}_b > 0$. These represent 26 out of 36 cases. Regions in which the method did not reject immediate selection are colored blue, and regions in which the method rejected immediate selection are colored orange (17 cases).

of selection. While we require a known demographic history among Neanderthal-admixed populations, our method is robust to modifications in these assumptions (see section 1.9.2.2 for discussion).

We found that in most regions that we analyzed, we could not reject a model of selection immediately favoring Neanderthal-introgressed alleles upon admixture. In some of these regions, East Asians and/or ancient populations do not contain signatures of adaptive introgression. Obviously some of these cases may represent our lack of power to rule out short onset times from haplotypes alone and so our results should not be taken as indicating that selection began immediately for many Neanderthal haplotypes. In addition, the regions that we ran our method on were identified from signals of unusual sharing with Neanderthals (Racimo et al., 2017), and so we may be missing some cases of recent selection on Neanderthal introgression that dragged up only very short blocks of Neanderthal ancestry. We also limited our analysis to regions with signatures of adaptive introgression in European populations, so there may be cases of recent selection in East Asian or other populations that we did not identify. Among the regions in which we did reject immediate selection, we estimated a wide range of waiting times until selection. From simulations, we found that we can generally classify these cases into earlier ($0 < \hat{t_b} \leq 800$) and more recent ($\hat{t_b} > 800$) selection. These time frames may be coarse, but with our estimates of recent selection being far from the boundary between early and recent selection, the transition between them possibly marks the end of the last glacial period ($\sim$ 11.5 kya). Thus, we are potentially distinguishing between adaptation to conditions during the Upper Paleolithic versus later periods defined by warming, technological innovation, the emergence of farming, and higher population densities.

We modeled that the beneficial allele was at a low frequency for some time followed by the onset of selection due to some environmental change. This initial period could correspond to the Neanderthal allele being neutral or balanced at low frequency. However there are a few

possibilities other than a truly delayed onset of selection that could explain our rejection of immediate selection. First, selection may have immediately favored an allele, but the allele was recessive and thus took a long time to rise in frequency. We chose to model additive selection because we do not have information on dominance, but for recessive alleles we may be timing its rise to intermediate frequency rather than the selection onset (see Jones et al., 2020, for a possible dominance related lag in an introgressed pigmentation allele in snowshoe hares). Given that the method does a poor job estimating the selection coefficient for these old sweeps from haplotype data alone, we likely cannot distinguish among different models of dominance. Second, selection may have immediately favored a beneficial Neanderthal allele, but it was only able to rise in frequency once it recombined off other tightly linked deleterious alleles on its Neanderthal haplotype background. There are two ways that this could lead us to incorrectly infer more recent selection: (*i*) the beneficial allele could have recombined earlier than expected, causing it to rise in frequency close to its introgression time but on a shorter Neanderthal haplotype than we predict under immediate selection, or (*ii*) a time lag between the selection onset and the recombination event that allowed the allele to rise in frequency could cause our timing estimates to mark the recombination event rather than the selection onset. Some studies have described the conditions in which an adaptive allele could introgress and eventually fix given its linked deleterious background upon introduction (e.g. Uecker et al., 2015), with some investigation into the timescale in which recombination would generate a haplotype with a net selective advantage (Sachdeva and Barton, 2018). Finally, the sweep of Neanderthal alleles up in frequency may not be due to a beneficial Neanderthal allele, but instead hitchhiking with a new beneficial allele that arose via mutation more recently than the Neanderthal introgression, by chance on the background of a Neanderthal haplotype. If Neanderthal-introgressed alleles are at $\sim 2\%$ frequency genome-wide, then we should expect about 2% of recent sweeps from new mutation to arise on a Neanderthal-introgressed background and sweep them to high frequency.

We identify two regions as having clear statistical support for very recent selection (OAS1,OAS3 and CACNA1S,ASCL5). These regions have very high $\hat{t_b}$, which our simulations show are very rarely estimated under cases of early selection. We are less convinced of alternative explanations for their late rise in frequency because they would need to be replicated in multiple populations. Specifically, given that the Neanderthal alleles rose in frequency so late, and that in both of these regions Neanderthal ancestry is at high frequency in present day Europeans and East Asians, this rise must have begun after Europeans and East Asians diverged from each other, i.e. independently. The chance that the Neanderthal alleles both recombined off of their deleterious background or both hitchhiked on a sweep in the region are extremely small. Thus, for OAS1,OAS3 and CACNA1S,ASCL5, we have evidence that Neanderthal alleles were independently beneficial in Europeans and East Asians, well after admixture with Neanderthals.

In the CACNA1S,ASCL5 region, the estimated waiting time until selection ($\hat{t_b} = 1500$) implies selection began after the sampling time of the West Eurasian Upper Paleolithic, however this population does carry Neanderthal haplotypes at high frequency in this region. The Neanderthal allele may have drifted away from low frequency in this population after it diverged from the European ancestry populations that we study, which our simulations show happens reasonably often and does not confound our method. Alternatively, selection may have started earlier in the common ancestor of European ancestry populations and more recently in East Asian ancestry populations. We reran our method after removing East Asian samples but our estimates did not change, which could either reflect a selection onset more recent than the West Eurasian Upper Paleolithic were sampled or poor performance of the method in the absence of East Asian samples. The latter possibility was beyond the scope of this paper to investigate, however future applications could investigate the effect of sampling and extend the method to allow for different onset times when selection independently favors alleles. In total, Neanderthal alleles in this region appear to have been independently favored

in the ancestors of East Asians and the ancestors of Europeans, indicating that geographically separated populations experienced similar changes in environmental conditions, possibly at different times.

The selection pressure favoring the OAS1,OAS3 haplotype is thought to be related to innate immunity. Indeed, a recent study found that the Neanderthal haplotype at OAS1 is protective against COVID-19 severity, hospitalization, and susceptibility (Zhou et al., 2020; Zeberg and Pääbo, 2021). It has previously been suggested that this gene contributes to traits under balancing selection because haplotypes at this locus have shown variable expression responses to different Flaviviruses (Sams et al., 2016) and a separate Denisovan haplotype segregates at high frequency at this locus in Melanesians (Mendez et al., 2012). Our recent date of selection, and the fact that genomic samples dated closer to the admixture time do not carry Neanderthal haplotypes in this region, are consistent with a recent onset of positive selection in a longterm cycle of balancing selection. A number of other genes known to influence immunity such as the TLR6-TLR1-TLR10 cluster and HLA class I genes harbor candidates of adaptive introgression, however we did not analyze them as they contain both Neanderthal and Denisovan haplotypes in Eurasian populations (Abi-Rached et al., 2011; Dannemann et al., 2016). The timing of adaptation for the OAS1,OAS3 region is at odds with the hypothesis that the transfer of pathogens from Neanderthals to modern humans necessitated rapid human adaptation using Neanderthal-derived immunity alleles (Enard and Petrov, 2018; Greenbaum et al., 2019). However, it would be interesting to test this idea more generally by attempting to date the start of adaptation for more of these introgressed, putative immunity-related haplotypes.

Our investigation contributes to a growing model-based understanding of the genomic patterns of adaptive introgression. Setter et al. (2020) introduced VolcanoFinder, a method that identifies cases of adaptive introgression by probing the recipient population's patterns of heterozygosity around the selected site. Using expected coalescence times and assuming

fixation of the selected allele, they predict very low diversity in a small window around the selected site and very high levels of diversity at intermediate distances. Our predicted within-population probabilities of coalescing characterize similar patterns when combining results from both partitions of a selected population (weighted by their frequency in the population, i.e. the frequency of the selected allele). However, the characteristic volcano pattern would disappear as the frequency of the selected allele decreases or waiting time until selection increases. Our work demonstrates that both of these situations are frequent, and thus the continued difficulty of developing relaxed tests for adaptive introgression while avoiding false positives. Shchur et al. (2020) also use a coalescent approach, in their case to model how adaptive introgression affects the distribution of introgression tract lengths. Fixing the time until introgression and admixture proportion, they show how stronger selection increases the introgression tract lengths. While we do not explicitly model tract lengths, these patterns are apparent in our coalescent predictions.

We implemented a flexible inference framework that could be further modified to investigate the spatio-temporal history of adaptation in modern humans and other systems, whether via introgression or derived mutations. Future applications could distinguish among groups of truly selected populations, allow the frequency of the selected allele to vary among populations, or to distinguish between drift and selection. Our models and estimation method for the timing of adaptation via introgression can be generalized to other organismal systems. In our current application we assume that the introgressed lineage coalesces with the Neanderthal lineage due to the Neanderthal population's low effective population size, as this allows us to side step modeling the sweep in Neanderthals. However, in other systems the donor population cannot be assumed to have a very low effective population size, but it should be simple to include the extra terms modeling the sweep-induced increase in coalescence between the recipient and donor populations close to the selected site. The application of methods like this to other species would allow a more general understanding of how often

introgression is rapidly favored versus supplying genetic variation for future adaptation.

Using a combination of ancient DNA and haplotype based timing, we documented spatially and temporally varying selection on Neanderthal alleles. Our results provide evidence that admixture between diverged populations can be a source of genetic variation for adaptation in the long term. They also allow us to better understand the historical and geographic contexts within which selection favored Neanderthal introgression. As experimental work begins to identify the specific alleles and phenotypes potentially selected on, we can more fully flesh out how interbreeding with archaic populations tens of thousands of years ago has shaped the evolution of modern human populations.

1.7 Acknowledgements

We thank Iain Mathieson for helpful feedback and for sharing genotype likelihoods from the ancient samples that we included in our analysis. We also thank members of the Coop lab for helpful discussions and Erin Calfee, Chuck Langley, Pavitra Muralidhar, Matthew Osmond, Anita To, Michael Turelli, and Carl Veller for comments on earlier drafts of our manuscript. We thank Rasmus Nielsen and three anonymous reviewers for valuable feedback on earlier drafts as well. This work used the Extreme Science and Engineering Discovery Environment (XSEDE) computing resource PSC Bridges through allocation TG-MCB200074 to SY. This work was supported by the National Science Foundation (IOS-1353380 and PGR-1546719) and the National Institutes of Health (GM108779 and GM136290 to GC and GM121372 to Molly Przeworski).

1.8 Appendix

1.8.1 Simulation Details

1.8.1.1 Selection Simulations

Keeping demography constant, we simulated different combinations of t_b, s, t_s, and set of selected populations (Tables 1.4 and 1.5). After the divergence between the ancestors of Neanderthals and modern humans, we added a mutation with $s = 0.025$ to a single Neanderthal haplotype at the selected site. It became neutral before admixture with modern humans. We restarted any runs in which the selected mutation did not reach fixation in the Neanderthal population prior to admixture, was not segregating in all selected populations at the onset of selection, was lost from any selected population during the sweep, was not at greater than 20% frequency in at least one selected population at the sweep finish, or was not at greater than 20% frequency in either the European or East Asian population sampled in the present day. The 20% frequency cut off was motivated by the same criteria with which adaptive introgression candidates were identified (Racimo et al., 2017).

As frequency trajectories in SLiM are stochastic, we chose combinations of s and t_s that would target different final frequencies (x_s) under a deterministic trajectory. For each s and target x_s, we calculated t_s according to equation 1.4 and assuming that the frequency at the sweep start is the admixture proportion we used in simulations ($g = 0.02$). However, because our simulations condition on the selected mutation segregating in all selected populations at the selection onset, the average starting frequency in our simulations tended to be greater than g, and increased the later selection began. Therefore, our chosen t_s on average led to rather high x_s in cases of late selection, even when we targeted $x_s = 0.2$. Alternative approaches to target a certain low x_s in our simulations would have led to many simulations where a target x_s was reached due to genetic drift rather than natural selection. Those approaches are inappropriate for us, as we focus on understanding the history of natural

selection, rather than distinguishing between genetic drift and natural selection.

Of all polymorphic sites that were not filtered, we randomly down-sampled the data set to 12,000 sites, which is close to the number of sites analyzed for data in our application with loci of the same size. As described in the main text, we used a different sampling approach for simulations generated to identify the $\text{CLR}_{t_b>0}$ boundary for the regions we analyzed in our application.

1.8.1.2 Neutral Simulations

We generated simulations under neutrality using the same demographic history as the selection simulations so that we could estimate the "genome-wide" neutral $\mathbf{F}$. We simulated 2000 independent trees by separately recording the tree sequences of 1 bp loci and subsequently overlaying multiple mutations. From these 2000 independent "loci" we estimated the neutral $\mathbf{F}$ from 19,323 sites after filtering.

1.8.2 Estimating probabilities of coalescing from allele frequencies

Using allele frequencies, we obtain unbiased estimates of $\mathbf{F}$, our genome-wide neutral probabilities of coalescing. Rearranging the equation providing the covariance of the change in population allele frequencies from the ancestral population, the probability a pair of ancestral lineages coalesce before the root is

$$f_{ij} = \frac{\text{Cov}[\Delta x_i, \Delta x_j]}{x_a(1-x_a)}. \tag{1.12}$$

Since $\Delta x_i = x_i - x_a$, the above covariance equals $\mathbb{E}_l[x_{il}x_{jl}] - \mathbb{E}_l[x_{al}^2]$. In our scripts we estimate f_{ij} using an equivalent expression that directly averages over the two possible alleles at a locus,

$$f_{ij} = 1 - \frac{\mathbb{E}_l[x_{il}(1-x_{jl}) + x_{jl}(1-x_{il})]}{\mathbb{E}_l[2x_{al}(1-x_{al})]}. \tag{1.13}$$

The numerator of the fraction represents the probability of sampling different alleles from population i and population j, which we estimate by taking its average over all loci. When

i and j correspond to the populations that root the tree, this average serves as the unbiased estimator of the denominator, $2x_a(1-x_a)$. When $i=j$, such that the numerator represents the expected heterozygosity within a population, we must account for how allele frequencies change when we sample without replacement, i.e. we must remove the finite sampling bias. If n_{il} represents the sample size of population i at locus l, then when $i=j$,

$$\mathbb{E}_l[x_{il}(1-x_{jl})+x_{jl}(1-x_{il})] = \frac{1}{L}\sum_{l=1}^{L}\frac{n_{il}}{n_{il}-1}(2x_{il}(1-x_{il})). \tag{1.14}$$

When validating that our model predictions match the probabilities of coalescing in our selection simulations, such as in Figure 1.4, we repeated the above procedure for sets of loci binned by genetic distance to the selected site.

1.8.3 Inference Details

When we calculate likelihoods of population allele frequencies, we need to account for the finite sampling bias in our estimates. Previously, we described the variance in the true population allele frequency due to genetic drift, however there is an additional variance in our estimates due to sampling. Since the count of a certain allele type in a sample is binomially distributed according to the true allele frequency, then with sample size n_{il} the variance in allele frequency due to sampling is expected to be approximately $x_a(1-x_a)\frac{1}{n_{il}}$. Thus the total variance in the population allele frequency is $x_a(1-x_a)(f_{ii}+\frac{1}{n_{il}})$. Therefore, when we calculate the probability of population allele frequencies at a locus, we first modify $\mathbf{F^{(S)}}$ by adding $\frac{1}{n_{il}}$ along the diagonal for each population i.

Since we do not know the ancestral allele frequency at each locus, we approximate it with the mean across our sampled population allele frequencies,

$$\bar{x_l} = \frac{1}{k_l}\sum_{i=1}^{k_l} x_{il}, \tag{1.15}$$

where k_l is the number of populations with allele frequency data at locus l. We acknowledge that in our case $\bar{x_l}$ will be biased toward European allele frequencies. With allele frequencies

now distributed relative to each other, we accordingly modify the population allele frequencies and covariance matrix by mean-centering them. The new population allele frequencies therefore represent deviations from their mean, and a negative covariance among pairs of populations implies that we predict their allele frequencies to be on opposite sides of this mean. As $\bar{x}_l$ uses information from all populations, we lose a degree of freedom and thus drop information from a single population. Our resulting mean-centered allele frequencies are

$$\vec{x_l}' = \mathbf{T}_l \vec{x_l} \tag{1.16}$$

and our mean-centered covariance matrix is

$$\mathbf{F}^{(\mathbf{S})'} = \mathbf{T}_l \mathbf{F}^{(\mathbf{S})} \mathbf{T}_l^{\mathrm{T}}, \tag{1.17}$$

where $\mathbf{T}_l$ is the $k_l - 1$ by k_l matrix with $k_l - \frac{1}{k_l}$ on the diagonal and $-\frac{1}{k_l}$ elsewhere. Therefore we calculate the probability of mean-centered allele frequencies at a locus as

$$\begin{aligned} &\mathrm{P}(\vec{x_l}'|r_l, \mathbf{F}, A_G, s, t_b, x_s) \\ &\quad \approx \mathcal{N}\left(\vec{x_l}'|0, \bar{x}_l(1-\bar{x}_l)\mathbf{F}^{(\mathbf{S})'}(r_l, \mathbf{F}, A_G, s, t_b, x_s)\right). \end{aligned} \tag{1.18}$$

We take the same approach when calculating probabilities under the null model.

To allow for computational efficiency, we first bin each site's absolute genetic distances to the selected site into 10^{-3} cM intervals. Each bin's midpoint value is then used as the representative recombination rate to predict $\mathbf{F}^{(\mathbf{N})}$ and $\mathbf{F}^{(\mathbf{S})}$ for each parameter combination. When we calculate the probability of allele frequencies at each site, we use its bin's representative $\mathbf{F}$ and remove the rows and columns corresponding to unsampled populations at the site prior to sample size correction and mean centering.

1.9 Supplementary Information

1.9.1 Supplementary Tables

Population	Migration Time (generations ago)	Admixture Proportion
EF	224	0.52
Steppe	155	0.36
WHG	226	0.12

Table 1.1: Timing and proportion of admixture between listed ancient populations and CEU (a European-ancestry population). Admixture proportions were determined from Haak *et al.* (2015). Simulations use adjusted admixture proportions so that CEU has the listed admixture proportions following the final admixture event. Our models assume a single time for all admixture among European-ancestry populations, which we define as the average of the migration times listed here.

Chrom.	Start to End Positions (hg19)	Genes	Plot Label
1	193772128 - 193999736		chr1:193885932
1	195854615 - 195922174		chr1:195888394
1	201038832 - 201102553	CACNA1S, ASCL5	CACNA1S,ASCL5
2	65395511 - 65533682	ACTR2	ACTR2
2	68347097 - 68507952	PPP3R1, PNO1	PPP3R1,PNO1
2	119536920 - 119693701	EN1	EN1

2	154435207 - 154551882		chr2:154493544
2	159794421 - 160204173	TANC1, WDSUB1, BAZ2B	TANC1...BAZ2B
2	160896628 - 161078020	PLA2R1, ITGB6	PLA2R1,ITGB6
4	5828060 - 5920067	EVC, CRMP1	EVC,CRMP1
4	189157618 - 189216507		chr4:189187062
5	507500 - 653905	SLC9A3	SLC9A3
5	167971758 - 168166025	SLIT3	SLIT3
6	1396838 - 1451472	FOXF2	FOXF2
6	66231951 - 66532520	EYS	EYS
7	129861223 - 129929228	CPA2	CPA2
7	136570934 - 136667889	CHRM2	CHRM2
8	13846885 - 13942904		chr8:13894894
8	13946965 - 14174597	SGCZ	SGCZ
9	16596898 - 16845056	BNC2	BNC2
11	11552692 - 11625808	GALNT18	GALNT18
11	70891472 - 70976597		chr11:70934034
12	54151324 - 54244563		chr12:54197944

12	84808525 - 84998584		chr12:84903554
12	113337379 - 113427536	OAS1, OAS3	OAS1,OAS3
14	100944952 - 101071031	WDR25, BEGAIN	WDR25,BEGAIN
15	43432954 - 44053859	TGM5, TGM7	TGM5,TGM7
15	74702840 - 75030779	SEMA7A, UBL7	SEMA7A,UBL7
15	85908189 - 86341507	AKAP13, KLHL25	AKAP13,KLHL25
16	51306915 - 51398178		chr16:51352546
16	77959647 - 78071863	VAT1L	VAT1L
18	60124891 - 60268585	ZCCHC2	ZCCHC2
18	60658804 - 60777888		chr18:60718346
19	33510019 - 33763643	RHPN2, GPATCH1, WDR88, LRP3, SLC7A10	RHPN2...SLC7A10
20	62156560 - 62210005	PPDPF, PTK6, HELZ2	PPDPF...HELZ2
22	49077496 - 49144645	FAM19A5	FAM19A5

Table 1.2: Putative regions of adaptive introgression that we analyzed. Genes correspond to those listed with signals of adaptive introgression in Racimo et al. (2017). We label intergenic regions by their chromosome and midpoint position.

Waiting time until selection (t_b)	**Selected human populations**	**Combinations of s and t_s (rows in Table 1.5)**
0	All Eurasian	1, 2, 3, 4, 5
200	All Eurasian	7, 8, 9
400	All Eurasian	7, 8, 9
	All European	6, 7, 8, 9
600	All Eurasian	7, 8, 9
	All European	6, 7, 8, 9
	All Eurasian except EurUP	6, 7, 8, 9
800	All Eurasian	7, 8, 9
	All Eurasian except EurUP	6, 7, 8, 9
	All Eurasian except East Asia, EurUP	6, 7, 8, 9
1000	All Eurasian except EurUP	6, 7, 8, 9
	All Eurasian except East Asia, EurUP	6, 7, 8, 9
1200	All Eurasian except EurUP	6, 7, 8, 9
	All Eurasian except East Asia, EurUP	6, 7, 8, 9
1400	All Eurasian except EurUP	6, 7, 8, 9

	All Eurasian except East Asia, EurUP	6, 7, 8, 9

Table 1.4: Simulated combinations of waiting time until selection (t_b), group of selected populations, selection coefficient s, and sweep phase duration (t_s). Combinations of s and t_s are labeled by numbers in Table 1.5.

Proposed selection coefficients (s)	Proposed waiting times until selection (t_b)
0.002, 0.005, 0.01, 0.015, 0.02, 0.025, 0.03, 0.05	0, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700

Table 1.3: Proposed parameters of s and t_b, each combination of which we calculated composite likelihoods. For a given run of the method, we did not evaluate a combination of s and t_b if the sweep would not complete (deterministically) before the present day, i.e. if the time until selection is recent enough and/or the sweep time t_s (which also depends on x_s) is long enough such that $t_b + t_s > t_i$.

	Selection Coefficient (s)	Sweep Phase Duration (t_s)	Target x_s	# Runs
1	0.005	501	0.20	700
2	0.005	697	0.40	700
3	0.010	349	0.40	700
4	0.010	430	0.60	700
5	0.010	528	0.80	700
6	0.005	501	0.20	100
7	0.005	655	0.35	200
8	0.010	327	0.35	200
9	0.010	474	0.70	100

Table 1.5: Combinations of s and t_s used in selection simulations used to evaluate the method's performance. Target x_s corresponds to the average frequency the selected allele would reach at the end of the sweep phase if it started the sweep at frequency $g = 0.02$. The last column (# Runs) provides the number of replicates simulated for a provided waiting time until selection and group of populations experiencing selection.

1.9.2 Method Details

1.9.2.1 Method Validation

The method's power to reject immediate selection increases with the true waiting time until selection (t_b) and the average frequency of the selected allele among putative selected populations (x_s) because they each lead to later estimates of t_b, which correspond to higher $\text{CLR}_{t_b>0}$ (Figures 1.9-1.13). Among the simulations in which our method rejects immediate selection, it reliably identifies late selection when $t_b > 800$. Notably, it can tell that selection is recent despite the selected allele having drifted to high frequency earlier in time. For example, the West Eurasian Upper Paleolithic (EurUP) were sampled 908 generations after admixture, such that whenever $t_b > 900$, allele frequencies in that population can only reach higher frequencies because of genetic drift. Despite these high frequency cases when the true $t_b > 900$, the method can tell that selection started after the EurUP were sampled.

When East Asians are the only Eurasian population that does not carry the allele at high frequency, the method is biased toward inferring earlier selection when $t_b < 800$ and does poorly when $t_b = 800$. When EurUP is the only population that does not carry the allele at high frequency, the method is biased toward inferring later selection when $t_b < 800$.

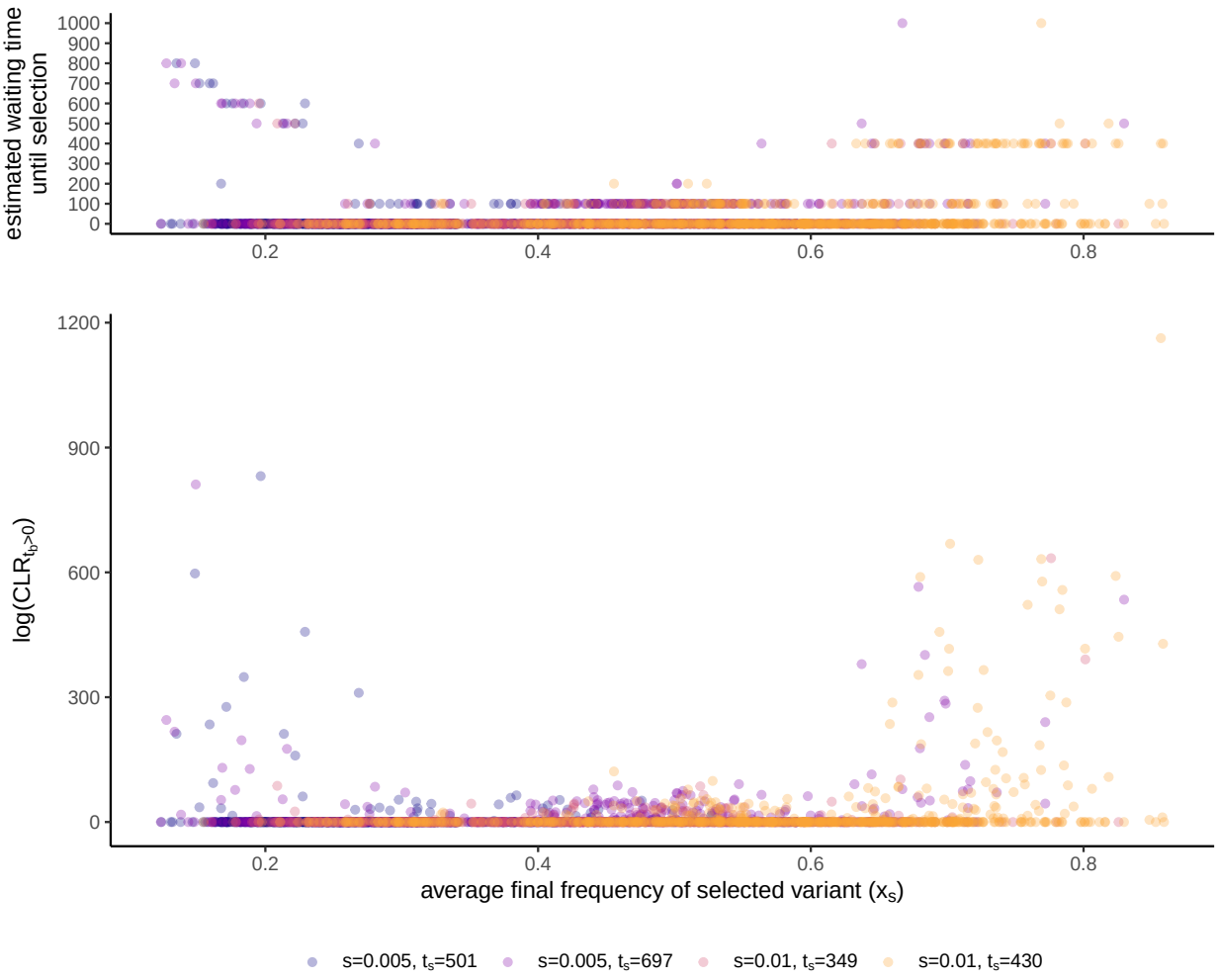

Figure 1.9: (Top) Relationship between the average final frequency of the selected variant (x_s) and the estimated waiting time until selection ($\hat{t_b}$) among immediate selection simulations. The method estimated immediate selection ($\hat{t_b} = 0$) in 85% of these simulations. (Bottom) Relationship between x_s and the log of the composite likelihood ratio between non-immediate and immediate selection, $\text{CLR}_{t_b>0}$, among immediate selection simulations. $\log \text{CLR}_{t_b>0}$ equals zero for the vast majority of low to intermediate x_s because at these lower frequencies the method is more biased toward inferring immediate selection. The relationship between x_s and $\log \text{CLR}_{t_b>0}$ does not differ among combinations of s and t_s. See Table 1.5 for targeted x_s corresponding to each combination of s and t_s.

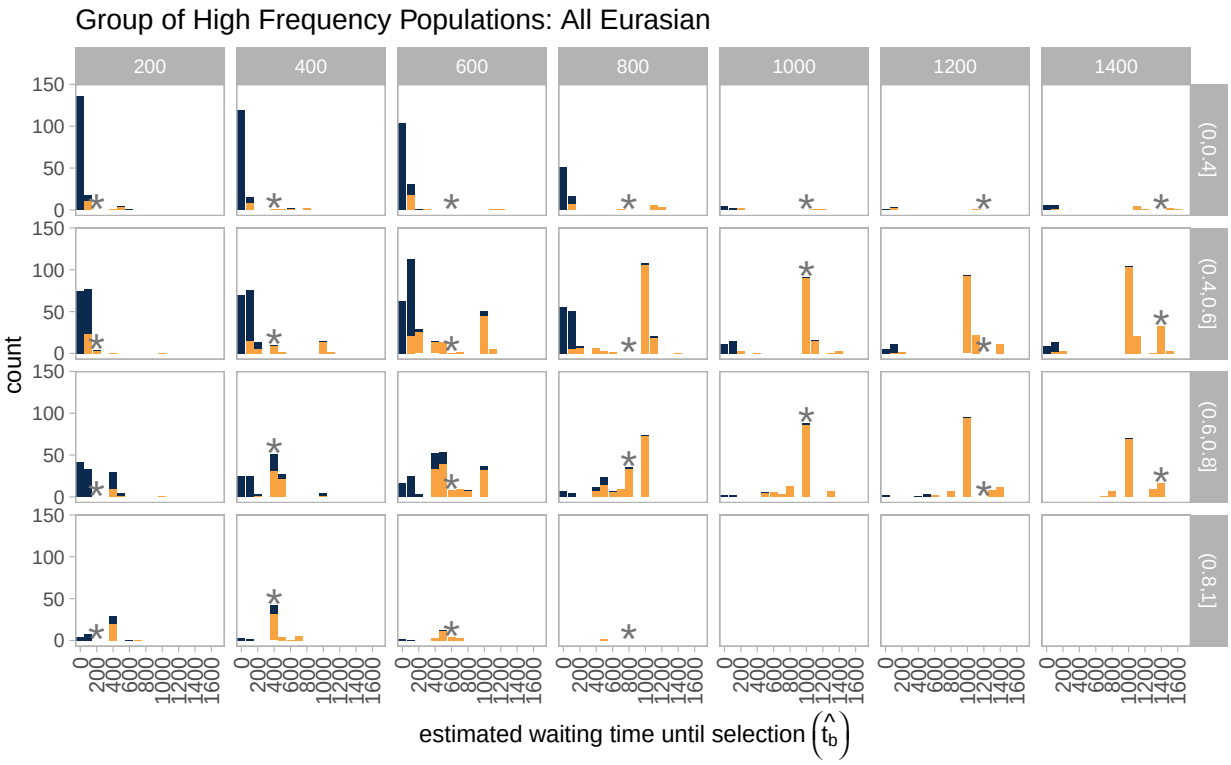

Figure 1.10: Distribution of estimated waiting times until selection among simulations in which the method considers all Eurasian populations to be selected. Blue bars represent cases in which the method did not reject immediate selection, and orange bars represent cases in which the method did reject immediate selection. Blue bars are stacked on top of orange bars (they do not overlap). Results are shown for each combination of true waiting times until selection (columns) and binned average final frequencies of the selected variant (x_s, rows). Asterisks mark the estimated waiting time until selection that matches the truth.

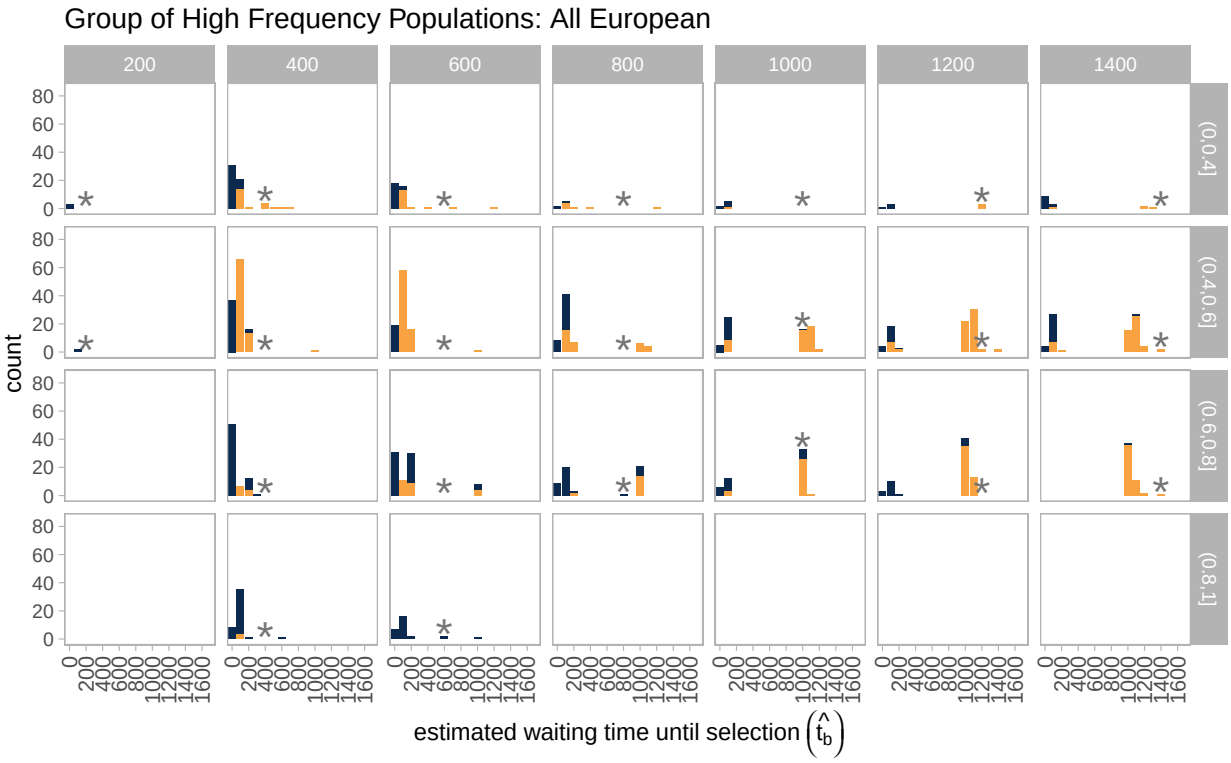

Figure 1.11: Distribution of estimated waiting times until selection among simulations in which the method considers all European populations (including the West Eurasian Upper Paleolithic) to be selected. Blue bars represent cases in which the method did not reject immediate selection, and orange bars represent cases in which the method did reject immediate selection. Blue bars are stacked on top of orange bars (they do not overlap). Results are shown for each combination of true waiting times until selection (columns) and binned average final frequencies of the selected variant (x_s, rows). Asterisks mark the estimated waiting time until selection that matches the truth.

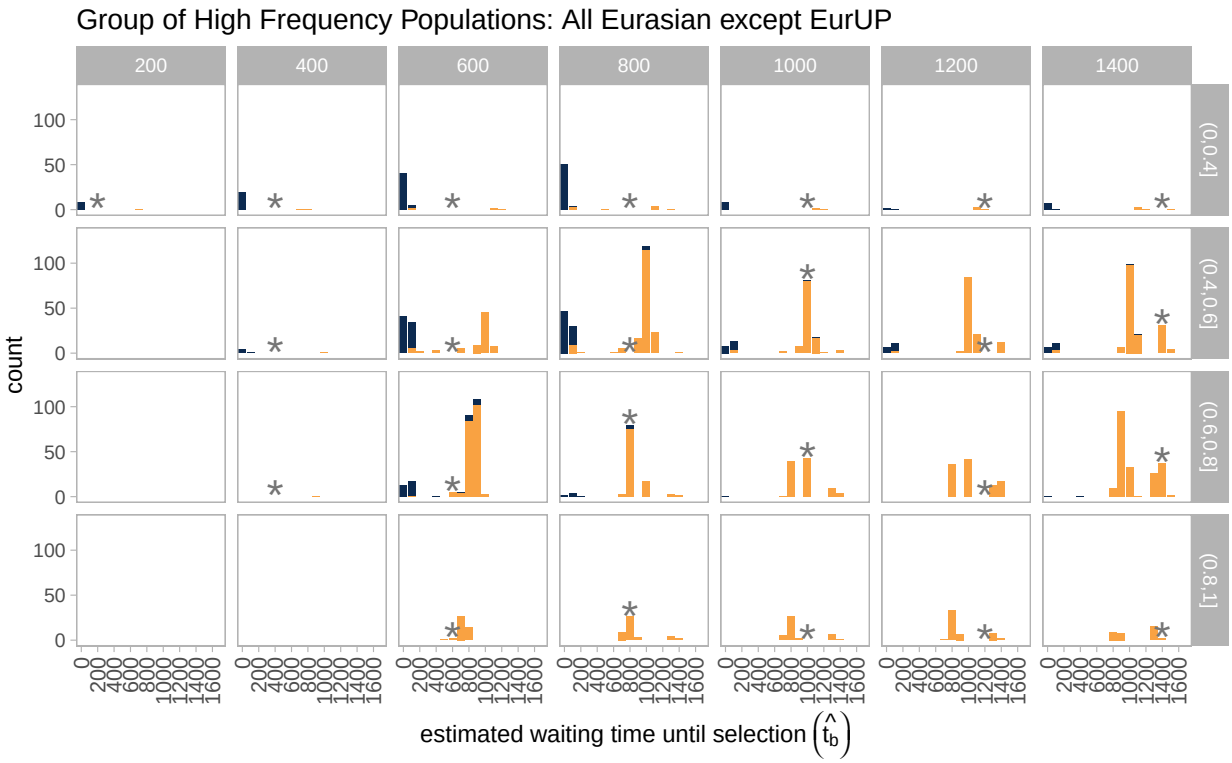

Figure 1.12: Distribution of estimated waiting times until selection among simulations in which the method considers all Eurasian populations except the West Eurasian Upper Paleolithic to be selected. Blue bars represent cases in which the method did not reject immediate selection, and orange bars represent cases in which the method did reject immediate selection. Blue bars are stacked on top of orange bars (they do not overlap). Results are shown for each combination of true waiting times until selection (columns) and binned average final frequencies of the selected variant (x_s, rows). Asterisks mark the estimated waiting time until selection that matches the truth.

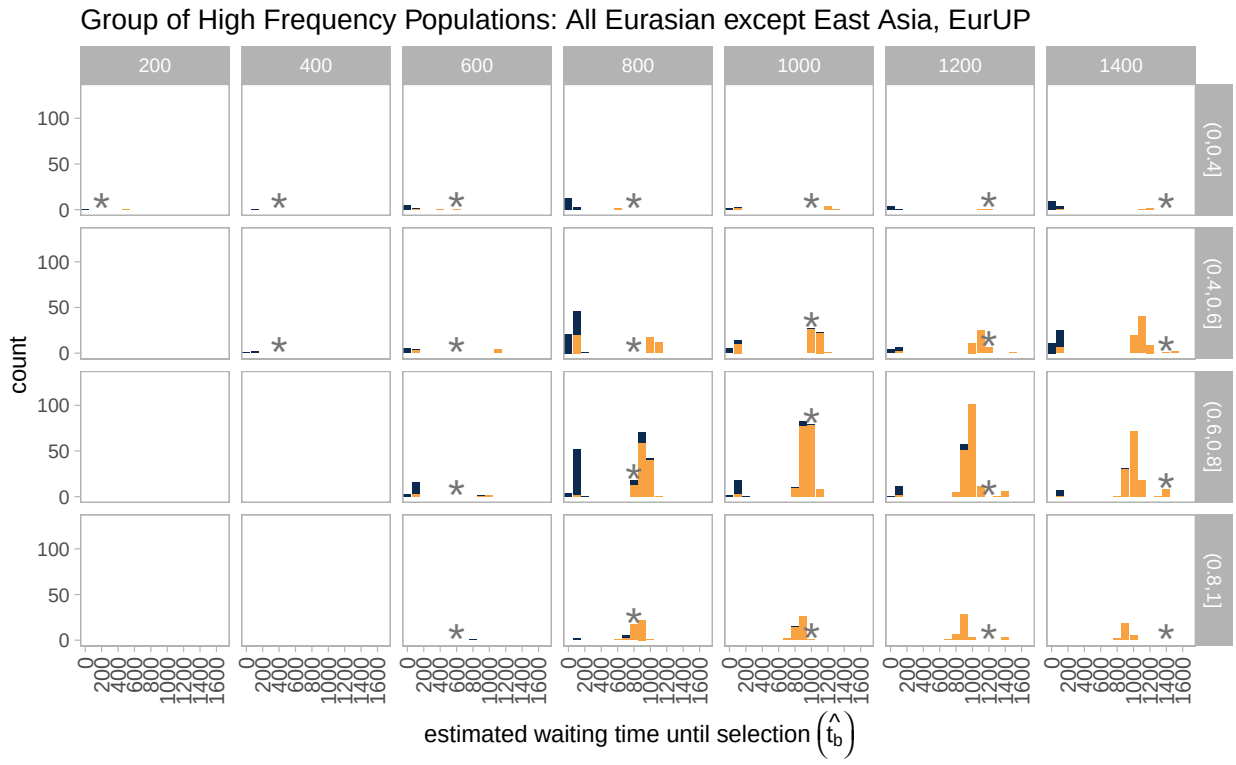

Figure 1.13: Distribution of estimated waiting times until selection among simulations in which the method considers all Eurasian populations except the West Eurasian Upper Paleolithic and East Asians (represented by CHB) to be selected. Blue bars represent cases in which the method did not reject immediate selection, and orange bars represent cases in which the method did reject immediate selection. Blue bars are stacked on top of orange bars (they do not overlap). Results are shown for each combination of true waiting times until selection (columns) and binned average final frequencies of the selected variant (x_s, rows). Asterisks mark the estimated waiting time until selection that matches the truth.

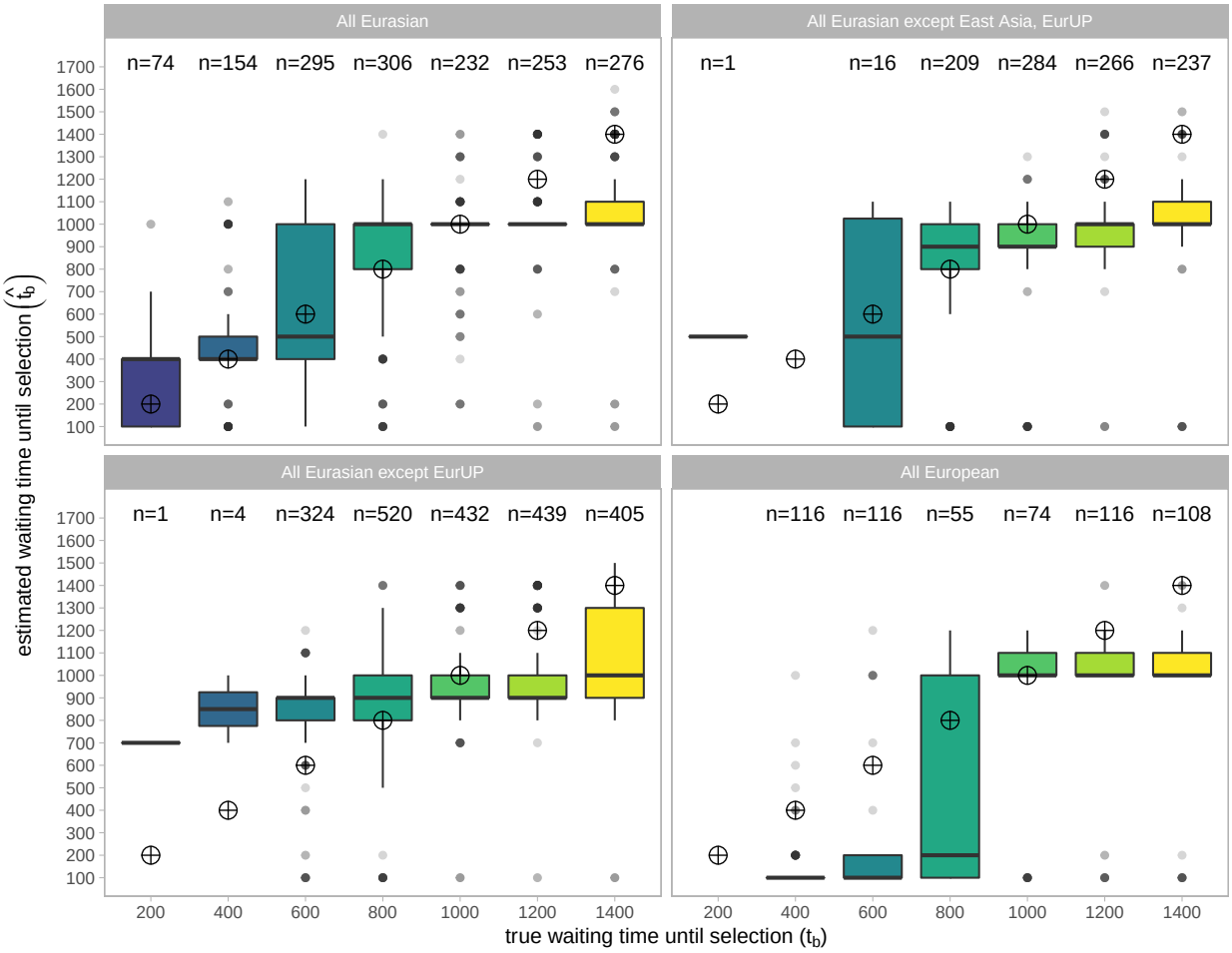

Figure 1.14: Estimated waiting time until selection ($\hat{t_b}$) among simulations in which the method rejected immediate selection. Target points mark the true waiting time until selection that we aim to estimate. Earlier true waiting times have fewer observations because the method has less power to reject immediate selection in this parameter space. Each panel corresponds to a different group of populations that the method considers selected.

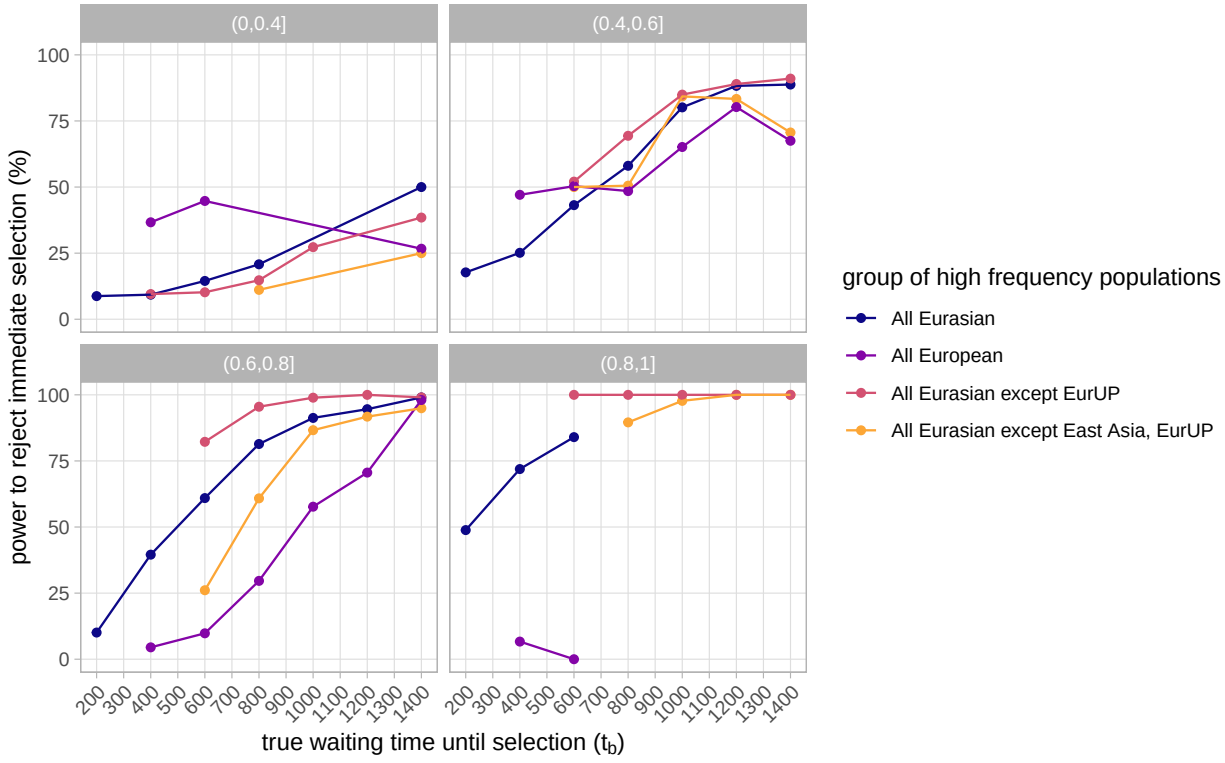

Figure 1.15: Power to reject immediate selection as the true waiting time until selection (t_b) increases. Lines correspond to simulations with different groups of populations considered selected. Points with fewer than 10 observations were removed. Panels correspond to bins of x_s, the selected allele's average frequency among populations considered selected.

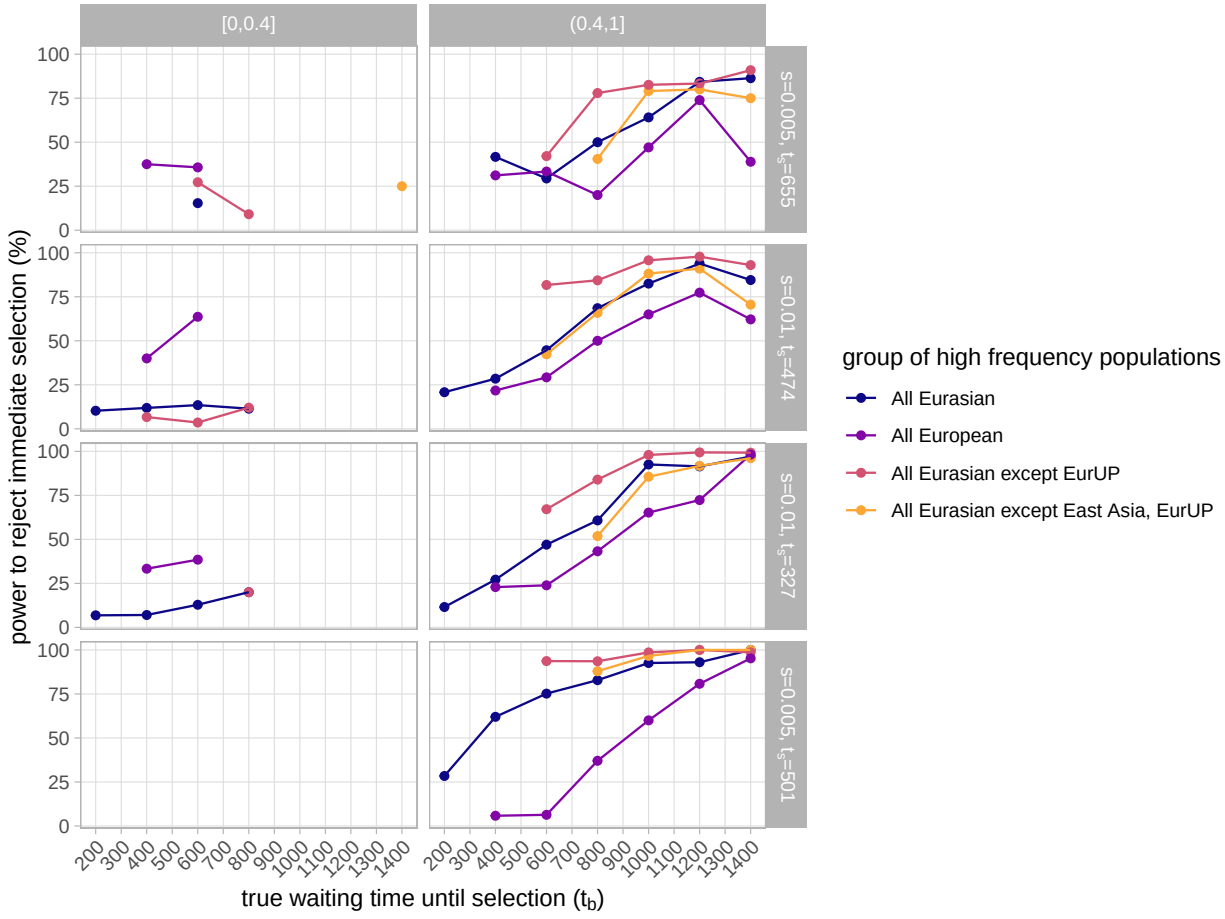

Figure 1.16: Power to reject immediate selection as the true waiting time until selection (t_b) increases. Lines correspond to simulations with different groups of populations considered selected. Points with fewer than 10 observations were removed. Columns correspond to bins of x_s, the selected allele's average frequency among populations considered selected, and rows correspond to combinations of s and t_s used in simulations.

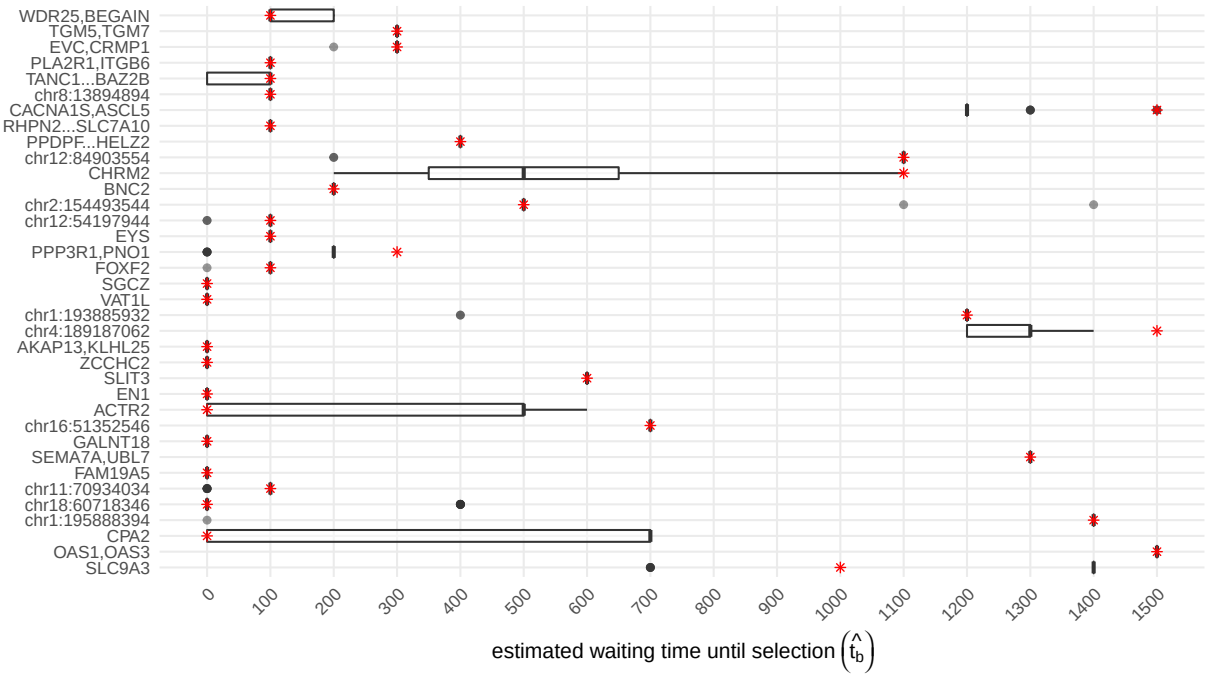

Figure 1.17: Each region's estimated waiting time until selection ($\hat{t_b}$) when different sets of ancient samples were randomly assigned to each population's genotype partitions according to their posterior genotype probabilities from imputation. Boxplots show the distribution under 40 resampled datasets. The red asterisks mark each region's $\hat{t_b}$ under the dataset in which ancient partitions are assigned based on the maximum likelihood genotypes. Regions are presented in the same order as in Figure 1.2.

1.9.2.2 Results under different datasets and demographic assumptions

Inclusion of the Altai Neanderthal Sample Population structure among Neanderthals and among Neanderthal-introgressed haplotypes could lead to erroneous results if we do not include the true introgressing populations in our analysis. To investigate our method's sensitivity to our choice of Neanderthal population, we compared our results between using only the Vindija sample and using both the Vindija and Altai samples to calculate allele frequencies for the Neanderthal population. We made a single comparison for each region at the 'best' partition site whose maximum composite likelihood under the selection model

was greatest relative to the null model. In the vast majority of regions, our estimated waiting times until selection ($\hat{t_b}$) did not change, and if they did they only moved within 200 generations (Figure 1.18). Our estimates changed greatly in two regions, CHRM2 and chr11:70934034, both of which we have identified elsewhere as difficult to infer their timing of selection onset (see the above and below sections). Our results are likely not sensitive to specification of the Neanderthal population because of their low effective population sizes leading to high probabilities of coalescing before the common ancestor of Neanderthals and modern humans. When the Neanderthal population is composed of Vindija alone, we estimate that the probability of coalescing $f_{nn} = 0.95$. For a population composed of Altai and Vindija (n=4), we estimate $f_{nn} = 0.91$. Therefore, the ancestral lineages of an allele sampled from the Altai Neanderthal and an allele sampled from the Vindija Neanderthal still have an extremely high probability of coalescing. Since in our models we mainly track whether sampled lineages descended from Neanderthals, such that those lineages coalesce while segregating in a Neanderthal population, the Neanderthal source population has a negligible effect on our predictions.

Sensitivity to Admixture Graphs In our application, we ran our method on all partition sites while assuming the admixture graph specified in Figure 1.1, which we call option A. To evaluate the sensitivity of our estimates to these assumptions, we re-ran the method using three other graphs and the dataset of the highest maximum composite likelihood partition site relative to the null model under option A. Each of these additional graphs contain a single modification from the original. In option B, we changed the time of admixture with Neanderthals from 2068 generations in the past (60kya with a generation time of 29 years) to 1988 generations (55kya with a generation time of 29 years). In option C, we did not model present day Europeans as a mixture of the Bronze Age Steppe, Early Neolithic Farmers, and Western Hunter Gatherers. While there is strong evidence that present day Europeans

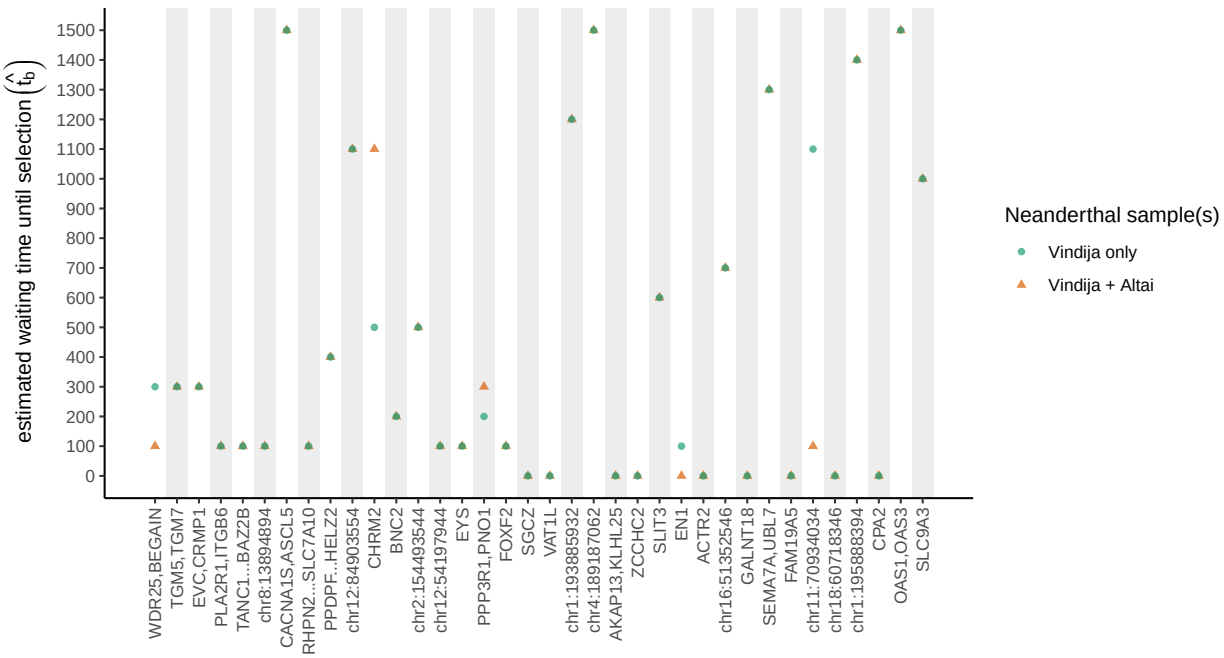

Figure 1.18: Each region's estimated waiting time until selection $\hat{t_b}$ under different sets of samples comprising the Neanderthal population. Regions are presented in the same order as in Figure 1.2.

are indeed a mixture of these three populations, we chose an extreme option to generally demonstrate the insensitivity of our method to specifications of admixture among human populations. In option D, we modified the divergence time among present day Europeans, the Bronze Age Steppe, Early Neolithic Farmers, and Western Hunter Gatherers to be 333 generations (10k years with a generation time of 29 years) later (from 40kya to 30kya). Each region's timing estimates when assuming each admixture graph are shown in Figure 1.19.

Overall, estimated waiting times until selection for the majority of regions did not change with each admixture graph, or changed very little. In particular, our modification of the divergence time among European populations (option D) did not change the estimated waiting time until selection from option A with the exception of 3 regions (PPPDF..HELZ2;

PPP3R1,PNO1; and chr11:7093403). Generally, estimates more often changed when the admixture time with Neanderthals shifted to be more recent. Among those regions in which there was a change in the estimate from option A when using option C, the estimated waiting time until selection was one or two hundred generations shorter, therefore corresponding to a very similar fixed time that selection began. Thus, changing the admixture time resulted in consistency either in the waiting time until selection or how recently selection began. However, three regions showed large inconsistencies in their estimates as we changed the admixture time, transitioning from very early to very late selection: chr12:54197944, FAM19A5, and chr16:51352546. In the former two regions, we did not reject immediate selection under option A. In chr12:54197944, we would potentially reject immediate selection if we simulated immediate selection under the new admixture graph, and therefore caution that the history of selection in this region is unclear. As for FAM19A5, the composite likelihood ratio between non-immediate and immediate selection remains low such that we would likely still reject immediate selection. In chr16:51352546, we previously rejected immediate selection and likely still would, however we caution that our estimated waiting time until selection is not informative in this region. When we remove any migration among modern human populations, the regions with a large change (> 300 generations) are chr12:84903554, chr1:193885932, ACTR2, SEMA7A,UBL7, and CPA2. Based on the composite likelihood ratio between non-immediate and immediate selection, we would likely continue to reject immediate selection in chr12:84903554, chr1:193885932, and SEMA7A,UBL7. In each of those regions the removal of migration is the only change that alters results, and in each case moves our estimates from recent to early selection. These earlier estimates may be unrealistic given that we have strong evidence for migration among European ancestry populations. It is unclear whether we would continue not to reject immediate selection in ACTR2 and CPA2, though we note these two regions also had wide variation in their estimates when we re-partitioned ancient populations (Figure 1.17), and so we refrain from making any claims

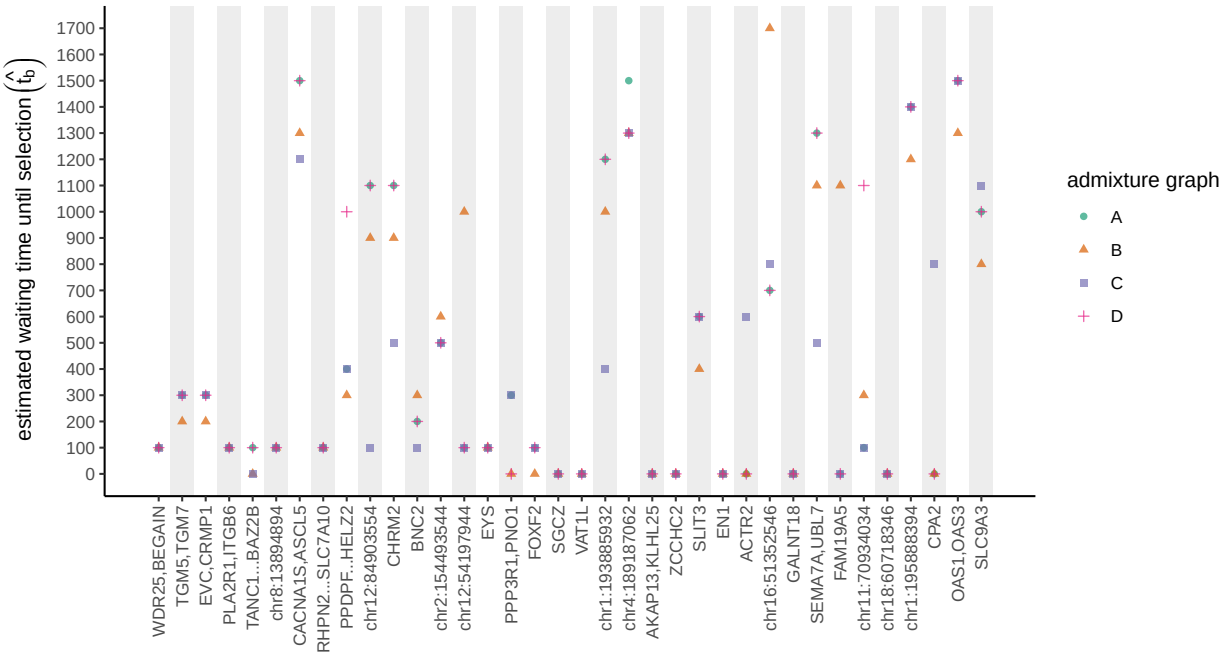

Figure 1.19: Each region's estimated waiting time until selection $\hat{t_b}$ under different admixture graphs. Each admixture graph specified by a letter is introduced in section 1.9.2.2. Regions are presented in the same order as in Figure 1.2.

about the timing of selection in these regions.

1.9.3 Model Derivation

1.9.3.1 Between selected populations that diverged during neutral phase I

Here, we illustrate another example of how we predict probabilities of coalescing between different categories of populations. We focus in depth on the probability of coalescing between pairs of selected populations who diverged from each other before the sweep started, i.e. during neutral phase I. Afterward, we describe how we change these predictions among other categorical pairs of populations. As a reminder, we consider three phases of the selected allele frequency trajectory: during neutral phase I the selected allele segregates at frequency g for t_b generations, during the sweep phase the selected allele rises from frequency g to

frequency x_s over t_s generations, and during neutral phase II the selected allele segregates at frequency x_s until the present day. Thus neutral phase II lasts a duration of $t_I - t_b - t_s$ generations, where t_I is the time between introgression and the present.

If a pair of populations $(i,\ j)$ diverged from each other d_{ij} generations in the past, and if this divergence occurred during neutral phase I, then they share a common ancestor for $t_I - d_{ij}$ generations before the time of introgression. Since the duration of neutral phase I is t_b, we also know that these populations were isolated for the last $t_b - t_I + d_{ij}$ generations of neutral phase I. To get the probability that the ancestral lineages of an allele sampled in population i and an allele sampled in population j coalesce, we consider the ancestry background at the selected site that each ancestral lineage is associated with at each phase transition. In the main text, we described the probabilities that a lineage is linked to the selected or non-selected background at the transition between neutral phase II and the sweep completion (Equation 1.5), in addition to the probability that a lineage remains strictly associated with this ancestry background for the duration of the sweep phase (Equation 1.6). With probability $\exp\left(-(t_b - t_I + d_{ij})\right)$ a lineage does not recombine out of the background that it was associated with at the transition between the sweep start and neutral phase I, until the time at which the two populations share a common ancestor in neutral phase I.

We begin by describing the probability that the pair of lineages coalesce conditional on them both being linked to the selected allele at the transition between neutral phase II and the sweep finish. We say a lineage is "strictly associated" with the sweep if it never recombines out of that background during the sweep phase and does not recombine at all during the preceding generations of neutral phase I when the lineages are in separate populations, such that

$$\begin{aligned}\Pr(\text{strictly associated with sweep}) = \Pr(&\text{always linked to selected allele during sweep})\\ &\times \exp\big(-r(t_b - t_I + d_{ij})\big).\end{aligned} \tag{1.19}$$

If both lineages remain strictly associated with the sweep, then we consider the possibility that they coalesce because they are associated with a background that is now relatively rare during neutral phase I. This could be due to their descent from the same introgressed allele, or the same non-introgressed allele that became associated with the selected background before the sweep began. As these lineages can only coalesce if they remain on the same background, we treat coalescence and recombination as competing Poisson processes, with recombination occurring at rate $2r$ and coalescence at rate $\frac{1}{2N_e g}$, where N_e is the mean effective population size of populations $(i,\ j)$ and g is the neutral admixture proportion, or similarly the frequency of the selected background during neutral phase I. The probability that neither of these events occur before the time of introgression is

$$\Pr(\text{no event}) = \exp\left(-(t_I - d_{ij})\left(2r + \frac{1}{2N_e g}\right)\right). \tag{1.20}$$

In that case, both lineages are still associated with the selected background at the time of introgression and thus must have descended from Neanderthals, so they coalesce with probability f_{nn} (the probability a pair of alleles sampled from Neanderthals coalesce). If at least one coalescence or recombination event does occur, we care about the outcome of the first event. The relative probability of coalescence is $\frac{1}{1+4N_e gr}$ and of recombination is $\frac{4N_e gr}{1+4N_e gr}$. If the event is recombination of one of the lineages, then we condition on whether the other lineage recombines at some point before introgression. With approximate probability $\exp\left(-r(t_I - d_{ij})\right)$ it does not recombine, and therefore remains associated with the selected background such that it descended from Neanderthals. Thus the lineages can only coalesce if the lineage that did recombine also descended from Neanderthals, with probability g. If both lineages recombined during this shared portion of neutral phase I, then they coalesce with neutral probability f_{ij}, which already accounts for whether they are each introgressed or not. We use a similar logic for the remainder of cases that involve one or no lineages remaining strictly associated with the selected background before the pair are in the same

population, such that

$$\begin{aligned}
&\Pr(\text{coalesce} \mid \text{both linked to selected allele at sweep finish}) = \\
&\quad \Pr(\text{strictly associated with sweep})^2 \Bigg[\Pr(\text{no event}) f_{nn} + \big(1 - \Pr(\text{no event})\big) \times \\
&\quad \Bigg(\frac{1}{1 + 4N_e gr} + \frac{4N_e gr}{1 + 4N_e gr} \big(\exp\left(-r(t_I - d_{ij})\right) g f_{nn} + \\
&\quad \left(1 - \exp\left(-r(t_I - d_{ij})\right)\right) f_{ij} \big) \Bigg) \Bigg] + \\
&\quad 2 \Pr(\text{strictly associated with sweep}) \big(1 - \Pr(\text{strictly associated with sweep})\big) \times \\
&\quad \Big[\exp\left(-r(t_I - d_{ij})\right) g f_{nn} + \left(1 - \exp\left(-r(t_I - d_{ij})\right)\right) f_{ij} \Big] + \\
&\quad \big(1 - \Pr(\text{strictly associated with sweep})\big)^2 f_{ij}.
\end{aligned} \tag{1.21}$$

Next we describe the probability that the pair of lineages coalesce conditional on one being associated with the selected allele and the other being associated with the non-selected allele at the transition between neutral phase II and the sweep finish. Since it is relatively rare for the lineage linked to the non-selected allele to switch to the background of the selected allele during the sweep phase, we do not consider the possibility of increased coalescence during neutral phase I. Therefore, we are simply interested in the ancestry background of each lineage at the time of introgression. If neither of them become disassociated with their backgrounds during the sweep phase and do not recombine before the time of introgression, then we know that one allele descended from Neanderthals while the other did not, such that they cannot coalesce. If the lineage originally associated with the non-selected background remains associated with this background during the sweep and does not recombine before introgression, whereas the other lineage does disassociate at some point, they can only coalesce if the disassociating allele did not descend from Neanderthals, with probability $1 - g$. At this

point, we need to slightly modify their probability of coalescing from the neutral estimate because we are dealing with a case in which we exclude descent from Neanderthals as a possibility. The probability that two lineages sampled from Neanderthal admixed populations coalesce, conditional on them both not descending from Neanderthals, is

$$\Pr(\text{coalesce} \mid \text{both non-Neanderthal}) = \frac{f_{ij} - g^2 f_{nn}}{(1-g)^2}, \tag{1.22}$$

where g is the neutral admixture proportion. We have previously described our predictions when one of the lineages descends from Neanderthals due to their consistent association with the selected background while the other lineage may not maintain a consistent association, in addition to when neither lineage maintains a consistent association. Therefore we obtain the following with conditional probabilities ordered in the same manner we discussed them,

$$\begin{aligned}
&\Pr(\text{coalesce} \mid \text{linked to different allele types at sweep finish}) = \\
&\qquad \Pr(\text{always linked to non-selected allele during sweep}) e^{-r t_b} \\
&\qquad \times \left(1 - \Pr(\text{always linked to non-selected allele during sweep}) e^{-r t_b}\right)(1-g) \\
&\qquad \times \Pr(\text{coalesce} \mid \text{both non-Neanderthal}) + \\
&\qquad \Pr(\text{always linked to selected allele during sweep}) e^{-r t_b} \\
&\qquad \times \left(1 - \Pr(\text{always linked to non-selected allele during sweep}) e^{-r t_b}\right) g f_{nn} + \\
&\qquad \left(1 - \Pr(\text{always linked to non-selected allele during sweep}) e^{-r t_b}\right) \\
&\qquad \times \left(1 - \Pr(\text{always linked to selected allele during sweep}) e^{-r t_b}\right) f_{ij}
\end{aligned} \tag{1.23}$$

Finally, we describe the probability that the pair of lineages coalesce conditional on both being associated with the non-selected allele at the transition between neutral phase II and the sweep finish. We again condition on their background associations at the time of introgression. If both lineages never disassociate from this background during the sweep and

never recombine during neutral phase I, then neither descended from Neanderthals and they coalesce with a slightly higher probability than neutrality as we previously described. If this happens with just one of the lineages, then the other lineage must not have descended from Neanderthals for them to coalesce, again with a slightly higher probability than neutrality. In the remainder of cases, the lineages coalesce neutrally. These cases are summarized in the following,

$$\begin{aligned}
&\Pr(\text{coalesce} \mid \text{both linked to non-selected allele at sweep finish}) = \\
&\qquad \big(\Pr(\text{always linked to non-selected allele during sweep})e^{-rt_b}\big)^2 \times \\
&\qquad \Pr(\text{coalesce} \mid \text{both non-Neanderthal}) + \\
&\qquad \Pr(\text{always linked to non-selected allele during sweep})e^{-rt_b} \\
&\qquad \times \big(1 - \Pr(\text{always linked to non-selected allele during sweep}e^{-rt_b})\big)(1-g) \\
&\qquad \times \Pr(\text{coalesce} \mid \text{both non-Neanderthal}) + \\
&\qquad \big(1 - \Pr(\text{always linked to non-selected allele during sweep}e^{-rt_b})\big)^2 f_{ij}.
\end{aligned} \tag{1.24}$$

Together, the full probability that a pair of lineages coalesce before the root is

$$\begin{aligned}
f^{(S)}_{p_i p_j} = &\Pr(\text{linked to selected allele at sweep finish})^2 \times \\
&\Pr(\text{coalesce} \mid \text{both linked to selected allele at sweep finish}) + \\
&2\Pr(\text{linked to selected allele at sweep finish}) \times \\
&\big(1 - \Pr(\text{linked to selected allele at sweep finish})\big) \\
&\times \Pr(\text{coalesce} \mid \text{linked to different allele types at sweep finish}) + \\
&\big(1 - \Pr(\text{linked to selected allele at sweep finish})\big)^2 \\
&\times \Pr(\text{coalesce} \mid \text{both linked to non-selected allele at sweep finish}),
\end{aligned} \tag{1.25}$$

where subscript p_i refers to any partition of population i and subscript p_j refers to any

partition of population j.

1.9.3.2 Modifications under other scenarios among pairs of selected and non-selected populations

If the pair of selected populations diverged during the sweep phase, we modify our above predictions so that the lineages are in the same population for all of neutral phase I by substituting any term $t_I - d_{ij}$ with t_b (these terms represent the time during neutral phase I in which the pair of lineages are segregating in the same population). If the pair of selected populations are either the same population or diverged from each other during neutral phase II (when the selected allele segregates at frequency x_s), and if the pair of lineages are linked to the same allele type at the most recent time they share a common ancestor, we consider the additional possibility that they coalesce during neutral phase II because of their associations with a subpopulation of haplotypes. We account for this using the same Poisson processes described above, except during neutral phase II the frequency of the selected background is x_s and the frequency of the non-selected background is $1 - x_s$. Recall that in cases of earlier selection, it is possible that the common ancestor of a set of selected populations also has descendant populations that do not carry the allele at high frequency. For those populations, our models assume the selected allele has the same frequency trajectory as other selected populations until the time it diverges from them, after which we switch the frequency of the selected allele in this population to be its sampled frequency. This influences the probability that this population's ancestral lineages were linked to the selected allele at the time its sweep phase begins.

In non-selected Neanderthal admixed populations, we track whether a lineage recombines out of its initial ancestry association, specified by its partition, over the entire time to introgression. If it loses this association, the only way it can have a non-neutral probability of coalescing with another lineage from a similarly non-selected population would be if that other lineage never recombined out of its initial background. As for a lineage from a selected

population, they would have a non-neutral probability of coalescing if the lineage from the selected population maintained an association with a selected or non-selected background by the time of introgression. In the non-selected population, the probability that a lineage could have descended from Neanderthals, conditional on it recombining out of its background at sampling, is g throughout the whole time to introgression. We show how our predictions change with t_b and x_s in Figure 1.20 and 1.21.

1.9.3.3 Modifications if a population was sampled in the past

If a population was sampled in the past, we modify the probabilities that its sampled lineages are linked to each background according to the number of generations allotted to a given phase. For example, consider a population i that was sampled τ_i generations ago, and carries the selected allele at high frequency such that in our application we consider it selected. If its sampling time falls within neutral phase II, then to calculate the probability that a lineage sampled from partition B of population i is linked to the selected background at the transition between neutral phase II and the sweep completion we simply reduce the amount of time for recombination in neutral phase II by τ_i. If the provided waiting time until selection t_b and selection coefficient s make it so that population i was instead sampled during the sweep phase, then at the transition between neutral phase II and the sweep completion, the sampled lineages are linked to the background denoted by their partition with probability 1. Then, we calculate that lineage's probabilities of remaining associated with either background during the sweep phase as if the selected allele reached the lower frequency expected under the provided s and the allotted time for the sweep, $t_I - t_b - \tau_i$. If population i was sampled during neutral phase I, then we do not consider it a selected population, and it has $t_I - \tau_i$ generations for possible recombination by the introgression time.

1.9.3.4 Incorporating migration among Neanderthal admixed human populations

In our application, most admixture graphs we use make CEU (the European ancestry representative population) a mixture of ancient populations WHG, EF, and Steppe. In our models, we modify predictions for partition B of CEU, assuming that ancient populations contributed only haplotypes with selected alleles to CEU. We thus consider donor populations to be only those contributing ancient populations with a partition B in the dataset. We set a single migration time t_m to be the average migration time from all human donor populations (see Table 1.1). With probability $\exp(-rt_m)$ a lineage sampled from partition B of CEU does not recombine before the migration time. In this case, the probability this lineage coalesces with a lineage sampled from any other partition p_j is an average of the probability that each donor population's partition A coalesces with the lineage from p_j, weighted by the relative admixture proportions of each of these donor populations. If the lineage does recombine by the time of migration, the probability it coalesces with a lineage from p_j is that of partition b of CEU. To predict coalescent probabilities for a pair of alleles sampled within CEU's partition B, we average over results from the possible migration histories of both lineages. We acknowledge that our approach to treat migration does not reflect a realistic coalescent history. Other options did not allow our predictions to converge to CEU's neutral probabilities of coalescing with increasing genetic distance from the selected site.

1.9.3.5 Predictions for genotype partition *Bb*

After making modifications for migration, we make predictions for all Bb partitions. The probability a lineage sampled from this partition in population i coalesces with any other partition from any other population p_j is the mean of this prediction for population i's partition B and partition b. This reflects that half of the time we are sampling a neutral allele that is linked to the selected allele, whereas in the remainder of cases the neutral allele is linked to the non-selected allele. The probability of coalescing between partition Bb of

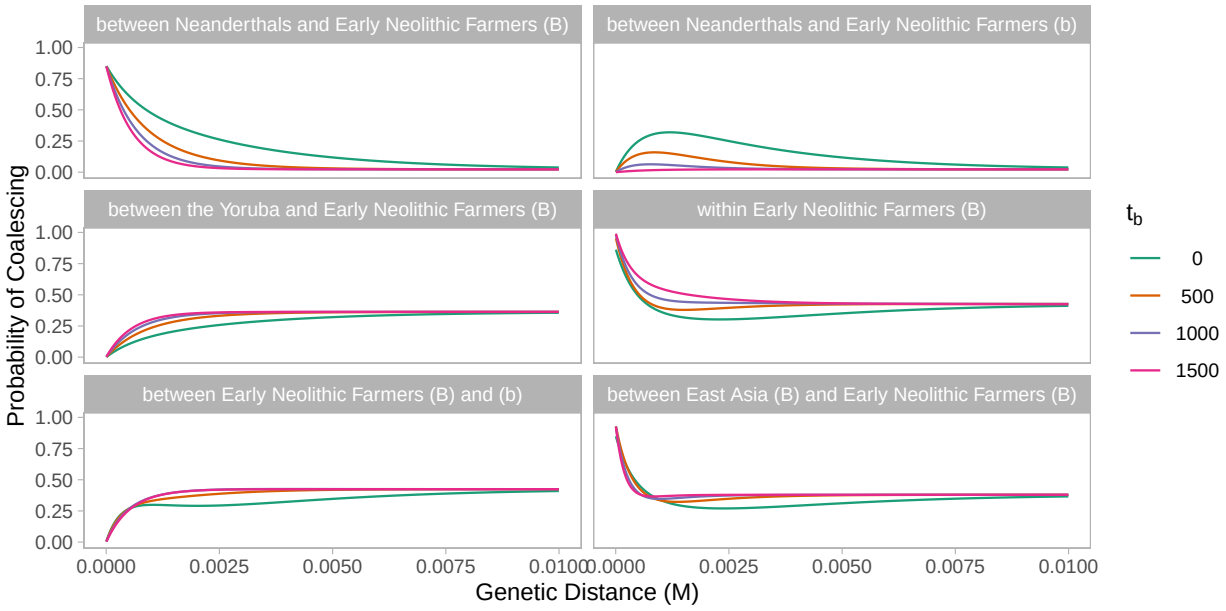

Figure 1.20: Model predictions under different values of t_b (the waiting time until selection) when $s = 0.01$ and $x_s = 0.7$.

population i and partition Bb of population j is the mean of the probability of coalescing of all four combinations of partitions A and a in populations i and j, in which there is a 50% chance either partition is sampled in either population. For probabilities of coalescing between lineages that were both sampled from partition Bb of population i, we take a weighted average of the probabilities of coalescing within partition B, within partition b, and between partition B and partition b, all partitions corresponding to population i. The first two cases occur with probability 0.25 and the last case occurs with probability 0.5.

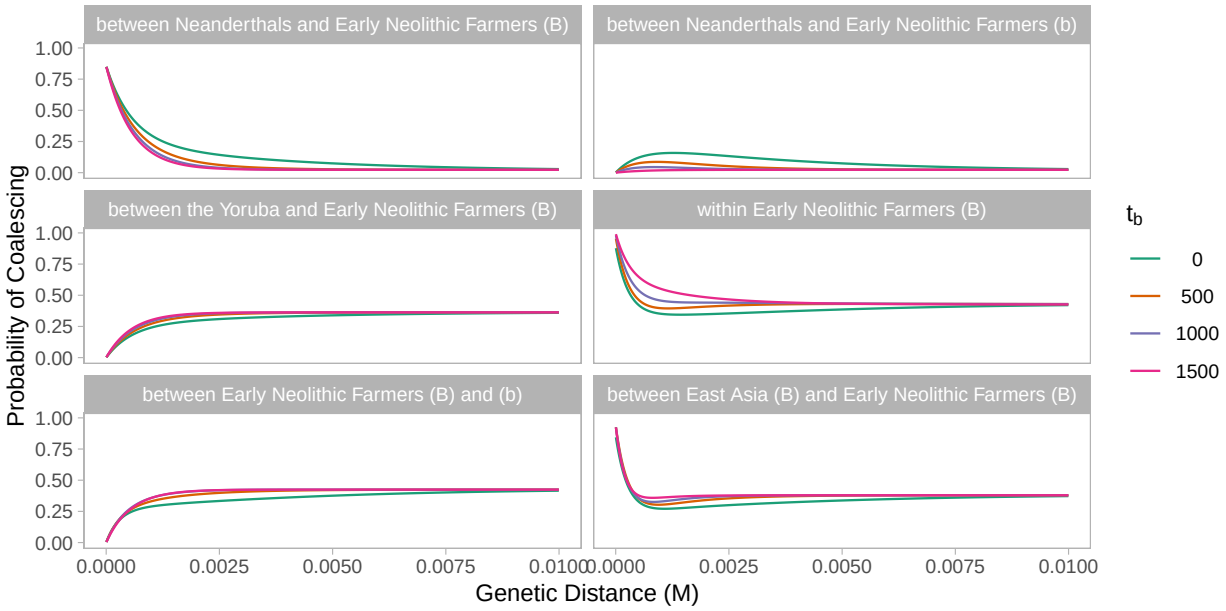

Figure 1.21: Model predictions under different values of t_b (the waiting time until selection) when $s = 0.01$ and $x_s = 0.3$. Note that predictions of t_b become less distinguishable when x_s decreases (see Figure 1.20 for comparison to $x_s = 0.7$).

Chapter 2

Adaptive introgression enables evolutionary rescue from extreme environmental pollution

Elias M. Oziolor[1,2,3], Noah M. Reid[4], Sivan Yair[2], Kristin M. Lee[2], Sarah Guberman VerPloeg[1], Peter C. Bruns[1], Joseph R. Shaw[5], Andrew Whitehead[2,3], Cole W. Matson[1]

[1] Department of Environmental Science, Center for Reservoir and Aquatic Systems Research (CRASR), and Institute for Biomedical Studies, Baylor University, Waco, TX 76798

[2] Center for Population Biology, University of California, Davis, CA 95616

[3] Department of Environmental Toxicology and Coastal and Marine Sciences Institute, University of California, Davis, CA 95616

[4] Department of Molecular and Cell Biology, University of Connecticut, Storrs, CT, 06269

[5] O'Neill School of Public and Environmental Affairs, Indiana University, Bloomington, IN, 47405

Statement of Contribution

This chapter resulted from a collaboration among the above authors. My contribution to this work was to extend and run a model-based inference method that was used to formally test for introgression and estimate the timing and strength of selection. To include the full story, the Main Text describes the work of all authors whereas the Supplement describes only my contribution.

Abstract

Radical environmental change that provokes population decline can impose constraints on the sources of genetic variation enabling evolutionary rescue. Adaptive toxicant resistance has rapidly evolved in Gulf killifish (Fundulus grandis) occupying polluted habitats. We show that resistance scales with pollution level and negatively correlates with inducibility of aryl hydrocarbon receptor (AHR) signaling. Loci with the strongest signatures of recent selection harbor genes regulating AHR signaling. Two of these loci introgressed recently (18-34 generations ago) from Atlantic killifish (F. heteroclitus). One introgressed locus contains a deletion in AHR conferring a large adaptive advantage (s = 0.8). Given limited migration of killifish, recent adaptive introgression is likely mediated by human-assisted transport. We suggest migratory inter-species connectivity as an important source of adaptive variation during extreme environmental change.

2.1 Main Text

Human alterations of the environment can be swift and severe, and thereby result in population declines and elevated selective pressures. Adaptation requires genetic variation from mutations, standing variation, migration or interspecific hybridization (introgression). However, declining population size coupled with immediate threats to fitness may constrain the sources of genetic variation that enable evolutionary rescue (Bell and Gonzalez, 2009; Messer

et al., 2016); new mutations arise slowly, standing variation may be insufficient, habitat fragmentation may limit migration, and hybridization may be rare and deleterious. Little is known of the relative importance of these sources of variation in the context of evolutionary rescue.

The Gulf killifish (Fundulus grandis) is common in coastal estuaries along the northern Gulf of Mexico. They occupy the Houston Ship Channel, much of which is heavily polluted with halogenated and polycyclic aromatic hydrocarbons (HAHs and PAHs) resulting from over 60 years of industrial activity (Yeager et al., 2007), forming a gradient of contamination through Galveston Bay (Oziolor et al., 2018) (Figure 2.1A). Many HAHs cause cardiac deformities in developing vertebrates that directly impair fitness (Hicken et al., 2011; Whitehead et al., 2012). Nevertheless, *F. grandis* persist at these sites and are resistant to the normally lethal effects of these toxicants (Oziolor et al., 2014). Here we examine the molecular underpinnings of apparently evolved resistance to extreme pollution among populations densely sampled across the pollution gradient, using experimental and population genomic approaches. We integrate prior work on *F. grandis*' sister taxon, the Atlantic killifish (F. heteroclitus), which has also adapted to similar chemicals (Reid et al., 2016), to uncover the evolutionary history of key adaptive variants.

We sampled fish from 12 sites spanning the pollution gradient (Figure 2.1A), spawned them separately in the lab, and exposed their embryos to a range of polychlorinated biphenyl 126 concentrations (PCB126; a model HAH). At the population level, resistance to toxicity, measured by rates of cardiac teratogenesis, is correlated with pollution along this gradient, where fish from the most polluted sites are resistant to concentrations of HAHs 1,000 times higher than normally teratogenic levels (Figure 2.1B, 2.1C). Resistance is retained through at least two generations in a clean environment (Figure 2.4) and is intermediate in hybrids from crosses between sensitive and resistant populations (Oziolor et al., 2014), indicating a genetic basis.

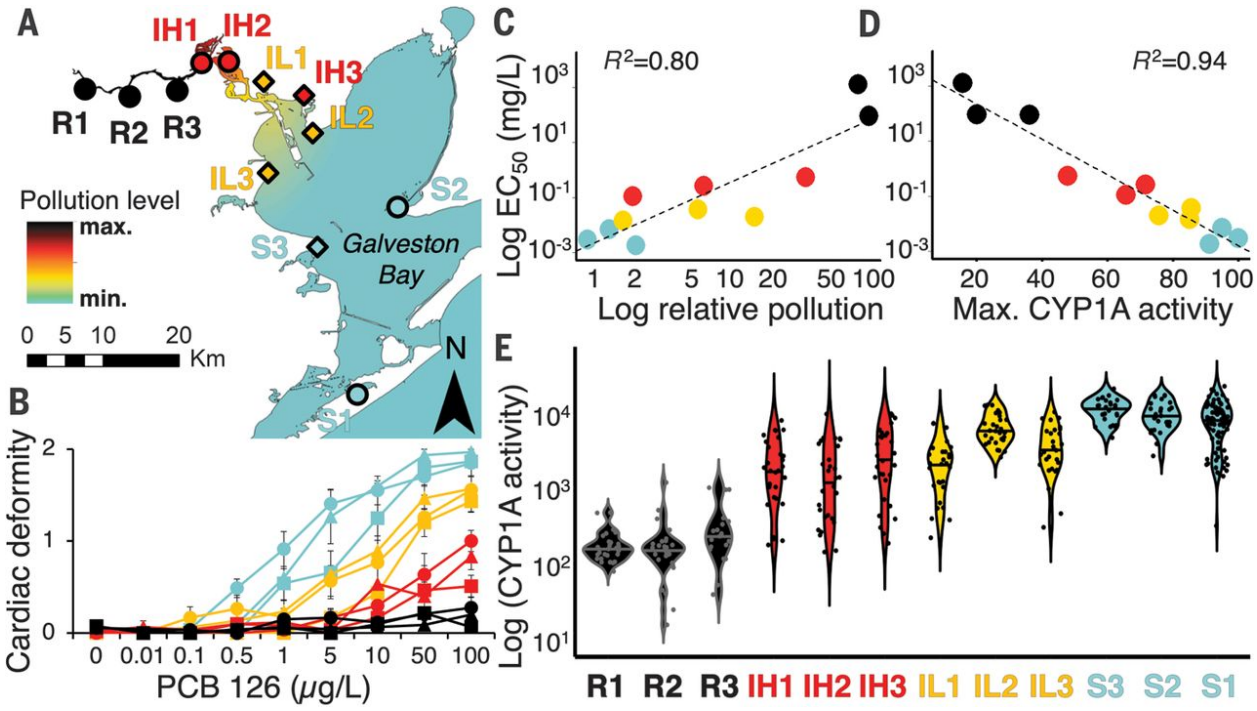

Figure 2.1: Variation in sensitivity to pollution among *F. grandis* populations distributed along a steep pollution gradient in Galveston Bay (USA). (**A**) Pollution gradient is scaled by color, from low (blue) to high (black). Populations include resistant (black, R1 to R3), intermediate-high resistance (red, IH1 to IH3), intermediate-low resistance (gold, IL1 to IL3), and sensitive (blue, S1 to S3). Genomics data were collected for populations denoted with circles. (**B**) Population variation in cardiac deformities in embryos exposed to PCB126 (error bars indicate standard error of the mean). Population variation in sensitivity to PCB-induced cardiac deformities (log median effective concentration EC_{50}) correlates with (**C**) habitat pollution and (**D**) AHR pathway inducibility (CYP1A activation by PCB126). (**E**) PCB-induced CYP1A activity varies among individuals and populations.

AHR signaling mediates HAH-induced cardiac teratogenesis (Clark et al., 2010). We compared AHR pathway function (CYP1A activity in response to PCB126 exposure (Nacci et al., 1998)) among populations across the pollution gradient (Figure 2.5). Similar to resistance, desensitization of the AHR pathway scales with levels of pollution (Figure 2.1D, 2.6). Within intermediate populations, inter-individual variation in AHR inducibility ranges between sensitive and resistant populations (Figure 2.1E). Together, our experimental results suggest that desensitization of AHR to HAHs underlies resistance in *F. grandis*.

To understand the genetic basis of this phenotype, we sequenced whole genomes of 24-49 fish ($\sim$ 0.6x coverage) from three resistant (R1, R2, R3), two intermediate (IH1, IH2), and two sensitive (S1, S2) populations. We searched for signatures of selection that co-vary with pollution through pairwise genetic differentiation (F_{ST}), population branch statistic (PBS) (Yi et al., 2010), and differences in nucleotide diversity (π) between polluted and sensitive populations in 20 kb windows. Neighboring 20 kb windows were merged into regions, which were ranked by their size and overall divergence. We defined the 1% most differentiated windows as outliers for each polluted population.

Genome-wide differentiation between populations is low (pairwise F_{ST} = 0.002-0.029; Figure 2.7) and genetic structure is consistent with geography (Figure 2.2A). Compared to sensitive populations, resistant populations have reduced nucleotide diversity and elevated Tajima's D (Figure 2.2B, 2.8), consistent with a recent reduction in population size. Evolved resistance (Figure 2.1) accompanied by strong recent selection (Figure 2.2C, 2.9) and population decline (Figure 2.2B) is consistent with evolutionary rescue in polluted populations (Bell and Gonzalez, 2009).

Top-ranked signatures of selection are shared across all pollution-adapted populations (Figure 2.2C). We detected 88 shared outlier regions distributed across 23 of 24 chromosomes. Differentiation in shared regions is higher between resistant and sensitive populations than between intermediate and sensitive populations (Figure 2.2D), and intermediate populations

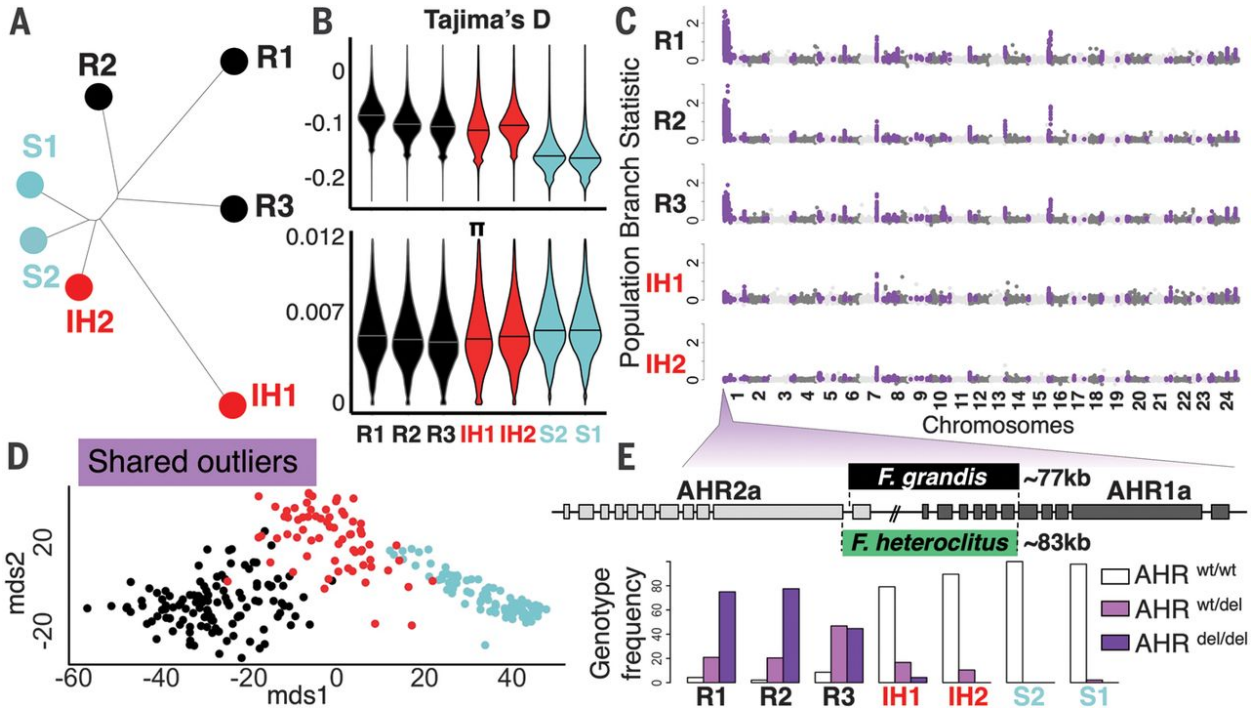

Figure 2.2: Genetic variation within and among populations. (**A**) Neighbor-joining tree (from genome-wide data) indicates population clustering along the pollution gradient. (**B**) Consistent with a demographic bottleneck, populations from polluted sites have an excess of rare variants (Tajima's D, top panel) and reduced nucleotide diversity genome wide (π, bottom panel) compared with clean site populations. (**C**) Genetic differentiation for each resistant and intermediate population (five panels) compared with both sensitive populations (population branch statistic). Background divergence is low, and purple indicates outlier regions (1% most divergent) that are shared among all adapted populations. (**D**) Multidimensional scaling (mds) of genetic variation for shared outlier regions clusters individuals and populations by resistance level. (**E**) A deletion in *F. grandis* ($\sim$ 77 kb spanning AHR2a and AHR1a on chromosome 1) has swept to high frequency in resistant populations (wt, wild type; del, deletion).

show no consistent signatures of selection relative to both resistant and sensitive populations (Figure 2.10). We conclude that intermediate populations are distinguished from resistant populations not by selection on different loci, but by weaker selection at the same loci.

Genomic regions showing the strongest signatures of selection harbor genes encoding key regulators of the AHR signaling pathway. AHR1a and AHR2a are centered in the most highly ranked shared outlier region (chromosome 1; Figure 2.2C, 2.11). Two paralogs of aryl hydrocarbon receptor nuclear translocator (ARNT), which dimerize with AHR to activate transcription (Figure 2.5), are within the second- and third-ranked shared outlier regions on chromosomes 8 and 10, respectively (Figure 2.2C, 2.12-2.14). Aryl hydrocarbon receptor interacting protein (AIP) regulates nuclear translocation of pollutant-activated AHR and is within the R1-R2 shared outlier region on chromosome 2 (Figure 2.2C, 2.15). Signatures of selection in multiple AHR pathway elements, coupled with additional selection signatures throughout the genome, indicate that AHR pathway modification is an important component of polygenic adaptation to the fitness challenges present in this urban estuary.

The top ranked outlier region spans $\sim$ 2 Mb and encompasses the tandem paralogs AHR1a and AHR2a (Figure 2.2C). Spanning these genes, we detect a 77 kb deletion at high frequency in resistant populations (85%, 83%, and 68%, in R1, R2, and R3, respectively), and moderate frequency in intermediate populations (25% and 5%, in IH1 and IH2, respectively), and in only one heterozygous individual from a sensitive population (Figure 2.2E, 2.16, 2.17). Given that the AHR pathway is profoundly desensitized in resistant populations (Figure 2.1D, 2.1E, 2.6), moderately desensitized in intermediate populations, and its experimental knockdown is protective of toxicity (Clark et al., 2010), we propose that this mutation is important for the adaptive phenotype. Five unique deletions in AHR genes have now been associated with rapidly evolved resistance to HAHs across three species of wild fish (Reid et al., 2016; Wirgin et al., 2011). Consistent with the "less-is-more" hypothesis (Olson, 1999), this suggests that rapid adaptation via disabling mutations may be common. We

speculate that sweeps of such large effect loci may quickly recover fitness following extreme environmental change, while smaller effect loci with reduced negative pleiotropy advance more slowly and perhaps eventually supersede the large effect loci.

We analyzed whole genome sequences from *F. grandis* and *F. heteroclitus* to contrast variants associated with resistance. In resistant *F. grandis*, the region containing the AHR deletion on chromosome 1, as well as the region from chromosome 10 encompassing ARNT, was more similar to *F. heteroclitus* haplotypes than to other *F. grandis* haplotypes (Fig 3A, 3B, S11), which could result from incomplete lineage sorting or admixture. The species divergence time elsewhere in the genome predated the divergence of the selected haplotypes, suggesting admixture as the source (Fig 3C).

To formally test for introgression and estimate the timing and strength of selection, we extended a coalescent theory-based inference method that distinguishes among modes of adaptation (Lee and Coop, 2017). Introgression of the deletion-bearing haplotype was much more likely than incomplete lineage sorting (Figure 2.3E, 2.18-2.20), consistent with divergence time estimates (Figure 2.3C). We inferred a large fitness advantage for individuals carrying the AHR deletion haplotype (selection coefficient = 0.8), and very recent gene flow (16 generations before onset of selection; Figure 2.3E) with the sweep occurring over 18 generations. This model estimates introgression happened 34 generations before sampling. As we are unlikely to have sampled the true *F. heteroclitus* source population (resistance is widespread in this species), this time may be even shorter. For the adaptive ARNT region we infer similar timing of gene flow (36-37 generations before sampling) and strong selective advantage (s=0.55) (Figure 2.21-2.22). These alleles clearly have a large effect on fitness and so likely explain a large proportion of the genetic variance in pollution tolerance. We conclude that recent introgression from a few successful *F. heteroclitus* migrants (Figure 2.20D, 2.22D) provided crucial genetic variation to rescue *F. grandis* populations from rapid environmental change. We speculate that sweeps of such large effect loci may quickly re-

cover fitness following extreme environmental change, while smaller effect loci with reduced negative pleiotropy advance more slowly and perhaps eventually supersede the large effect loci.

We evaluated the footprint of introgression in *F. grandis* genomes by scanning for genomic segments (> 400 kb) showing evidence of *F. heteroclitus* ancestry. These ancestry blocks collapse into 15 discrete regions totaling 70 Mb on 9 chromosomes (Figures 2.3A, 2.23, 2.24). Most large introgressed blocks are rare (11 of the 15 regions found in ¡ 5 individuals) and are found exclusively in polluted populations (Figure 2.3D). Over 30 Mb of the 40 Mb comprising chromosome 1 showed evidence of introgression in at least 1 individual, representing *F. heteroclitus* ancestry that hitchhiked with the deletion sweep. Mean haplotype block size also suggests that the introgression event was recent (65-155 generations before sampling). A recent discrete hybridization event occurred in the Houston Ship Channel, introducing highly advantageous adaptive haplotypes from *F. heteroclitus*, and leaving behind evidence of *F. heteroclitus* ancestry scattered across the genome. Since *F. heteroclitus* have small home ranges (Teo and Able, 2003) and the nearest populations are in Florida, ¿2500 km away, human-mediated transport is the likely mechanism of introduction (Morim et al., 2019), possibly through ballast water transfer (Wonham et al., 2000) or baitfish transport (Kilian et al., 2012).

The importance of hybridization in conservation biology is contentious (Allendorf et al., 2001; Hamilton and Miller, 2016). Recipient populations are at risk for negative ecological interactions with invaders, introduction of locally maladaptive or deleterious alleles, or genetic incompatibilities (Uecker et al., 2015; Simberloff et al., 2013). We propose that *F. heteroclitus* introgression into Galveston Bay was sufficiently rare to preclude extensive accumulation of deleterious variation in *F. grandis*, and that the adaptive advantage afforded by introgressed loci was sufficient to overcome maladaptation imposed by linked loci. Though a growing body of work shows that ancient hybridization can contribute adaptive

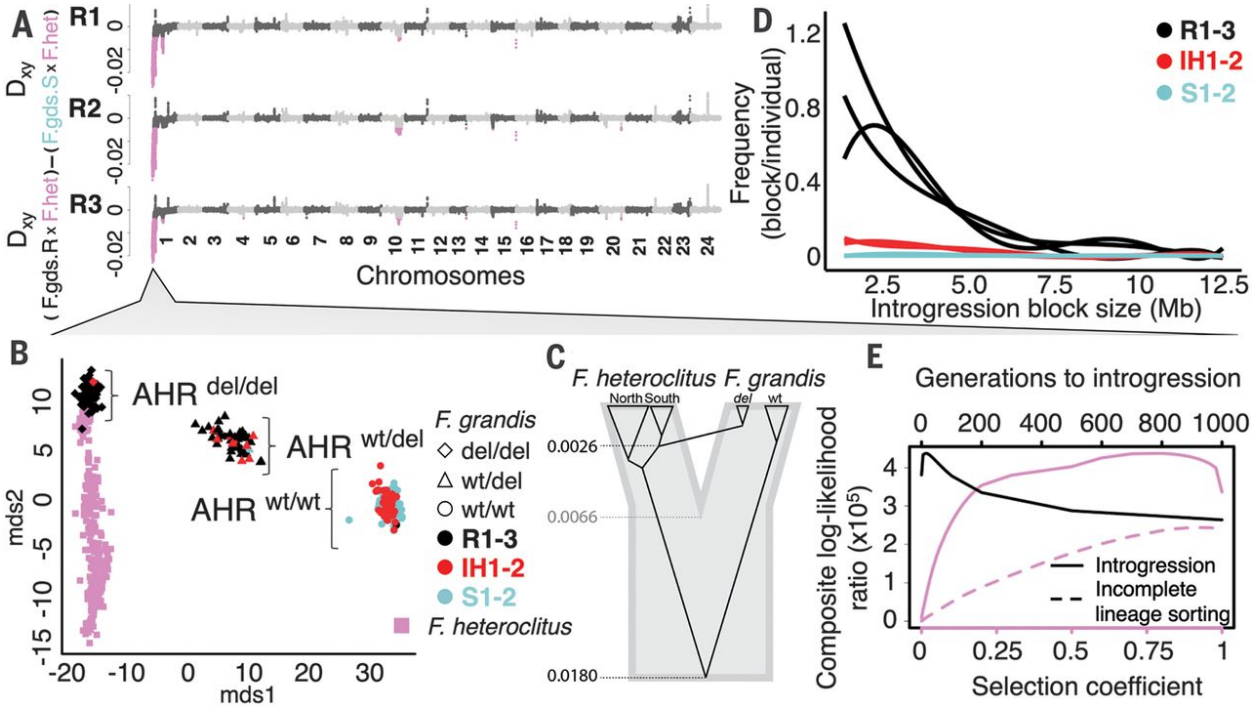

Figure 2.3: Evidence for adaptive introgression from *F. heteroclitus*. (**A**) Sequence similarity (D_{XY}) calculated between each resistant *F. grandis* population and *F. heteroclitus*, contrasted with similarity between a sensitive *F. grandis* population and *F. heteroclitus*. Negative values indicate similarity with *F. heteroclitus* that is higher in resistant than sensitive *F. grandis* populations (pink). (**B**) Genetic variation flanking the AHR deletion is more similar between resistant *F. grandis* and *F. heteroclitus* than between resistant and sensitive *F. grandis*, suggesting introgression. (**C**) Mean time to coalescence of 1 Mb surrounding the AHR deletion between *F. grandis* and southern *F. heteroclitus* is 0.0026 mutations per site, and between intact *F. grandis* alleles and all of *F. heteroclitus* is 0.018. Interspecific coalescence times that predate population divergence (0.0066) is evidence of gene flow. (**D**) Large introgression blocks are restricted to resistant populations. (**E**) Maximum composite log-likelihood ratios under introgression and incomplete lineage-sorting models, as a function of strength of selection (pink, bottom axis) and time between introgression and onset of selection (top axis) for the AHR deletion region, relative to a neutral model. Introgression is likely the source of the adaptive haplotype in *F. grandis*, which rapidly swept because of large selective advantage ($s = 0.8$).

genetic variation (Huerta-Sánchez et al., 2014; Jones et al., 2018, e.g.,), our work shows that hybridization can provide variation crucial for adaptation following swift and extreme environmental change.

2.2 Supplement

The supplementary material included here concerns the model-based inference of introgression, DMC-MYAdIDAS (**DMC** **M**odified for **Y**oung **A**daptive **I**ntrogression **D**istinguished from **A**llele **S**tanding)

2.2.1 Method overview

In two genomic regions, we confirmed and characterized adaptive introgression using and extending the inference method DMC (`https://github.com/kristinmlee/dmc`) (14) that aims to distinguish among types of convergent adaptation. This composite likelihood-based approach utilizes genomic data to detect cases of convergent adaptation at the genetic level and the loci involved, taking advantage of the signature that selection has on linked neutral variation. We compared the composite likelihoods estimated under various modes of convergent adaptation, including those due to gene flow and selection on shared ancestral standing variation. Here, we extended this method by *i)* modifying the migration model to include staggered sweeps, *ii)* allowing for extremely strong selection regimes, and *iii)* allowing for differences in effective population sizes among populations. We also slightly modified this method to fit patterns observed to be unique to each region. In our analysis for chromosome 1, we assigned the selected site to be the deletion midpoint instead of making it a free parameter. In our analysis for chromosome 10, we modified the models to exclude selection in the source population of the putatively introgressed haplotype. Details for these extensions can be found in Section 4.4. The R code and documentation for our approach are provided in `https://github.com/kristinmlee/dmc/myAdidas`.

We ran this method on two regions that show the greatest signals of adaptive introgression in the dxy analysis (Fig 3A): the AHR deletion on chromosome 1 and the ARNT region on chromosome 10. To investigate the origin of the putatively introgressed haplotypes in *F. grandis*, we distinguished between models in which the deletion was found in both *F. grandis* and *F. heteroclitus* due to incomplete lineage sorting or introgression. These models are consistent with the evidence that the AHR deletion in resistant *F. grandis* lies on a haplotype most closely related to the resistant *F. heteroclitus* population sampled from the Elizabeth River (h_{T4}) and that, for an *F. heteroclitus* population with whole-genome resequencing data (8), h_{T4} is geographically closest to the species range of *F. grandis*. DMC can be used to dissect the mode of convergence between any group of populations and controls for background population structure. Here, we ran the method using population allele frequency data from three pairs of resistant and sensitive populations, respectively: *i)* h_{T2} and h_{S2} – from the northern clade of *F. heteroclitus*, *ii)* h_{T4} and hS4 – from the southern clade of *F. heteroclitus*, and *iii)* g_{R2} and g_{S2} – from *F. grandis*. We chose g_{R2} as the representative for resistant *F. grandis* because it has the AHR deletion at high frequency (85%) and the highest number of individuals sampled among the resistant *F. grandis* populations. The *F. grandis* g_{S2} population was used as the representative sensitive population because no deletions were sampled in it. In the analysis for the AHR deletion, for each resistant and sensitive population used for inference, we included only individuals that were either homozygous for a deletion (resistant) or homozygous for intact haplotypes (sensitive). In the analysis of the ARNT region, we included only *F. grandis* individuals in g_{R2} that had an introgressed haplotype (at 30-45% frequency), and only individuals in g_{R2} that did not have an introgressed haplotype. Haplotypes were estimated with a custom script (`https://github.com/eoziolor/fgfh_post/blob/master/scripts/combined.Rmd`), assigning individuals to groups of introgressed vs. non-introgressed haplotype in this region based on genetic similarity.

We tested four models (Figure 2.18) to further investigate the similar haplotypes observed

between h_{T4} and g_{R2}. First, we assumed a model of selection on ancestral standing variation (Figure 2.18A), (here referred to as the incomplete lineage sorting model). This model specifies that the selected variant was standing at very low frequency in the ancestor of the two adapted populations (here the two species) and remained standing after the species split until it recently swept independently in both populations. The other three models invoke migration of the selected variant from h_{T4} to g_{R2}. First, the standing variant source model (see Appendix A.4 in Lee and Coop, 2017) in which the selected variant was standing at low frequency in h_{T4} and then was introduced to g_{R2} via migration (Figure 2.18B). The selected variant may still be standing for some time post-migration before it sweeps in both populations when the selective pressure begins. Next, we assumed the migration model (Figure 2.18C), in which the selected variant migrates from h_{T4} to g_{R2} during the sweep in h_{T4} and immediately begins to sweep in g_{R2}. This model assumes that migration occurs during the sweep and may be more relevant for populations that are more closely related and experience constant rates of gene flow. We call this case the concurrent sweeps model since both populations are sweeping at once. Lastly, we specified a new model of migration (Figure 2.18D) that is a modification of the concurrent sweeps model and may be more relevant for our distantly related populations. Since contemporary migration from the Atlantic coast to the Gulf coast is extremely rare, we allowed the sweep to finish in h_{T4} before migration to g_{R2}. In this case, we also allow for some amount of "standing time" (a free parameter) between when migration ended and the sweep began in g_{R2}. This may be relevant if g_{R2} does not immediately experience a large selective advantage of having the migrant haplotype (i.e. pollution has not started yet or is weaker). Additionally, while h_{T4} is our southernmost sampled *F. heteroclitus* population, it is likely that the deletion actually migrated from a different southern-clade population of *F. heteroclitus*. The standing time can then represent some time that allows for more independent recombination to occur between h_{T4} and g_{R2} that may be a better fit to the data if the haplotypes in h_{T4} and g_{R2} look slightly different

at the edges. We refer to this model as the staggered sweeps model and outline the details below. Here, the standing variant source, concurrent sweeps, and staggered sweeps models are all cases of adaptive introgression that we aimed to distinguish between.

For the analysis of the AHR deletion and ARNT region, we used SNPs from the first 15 Mb of chromosome 1 and last 16 Mb of chromosome 10, respectively. We assumed a recombination rate of 1.48 x 10-8 M/bp/gen, estimated from a linear fit in our regions of interest from the *F. heteroclitus* recombination map (EBI BioStudies Accession S-BSST163). Genome-wide SNP data from regions not showing signs of selection and only individuals included in the respective analysis (565,528 SNPs for chromosome 1 analysis and 594,086 SNPs for chromosome 10 analysis) were used to estimate the neutral coancestry matrix, following Lee and Coop (2017), to summarize neutral population structure. We used SNPs that had data across all populations, effectively excluding the deletion region from our analysis of chromosome 1. Since the approach utilizes signatures in neutral loci flanking the putatively selected site (their population allele frequencies and the recombination distance from the selected site), there is no effect of removing the deletion apart from fixing the location of the selected site, which is normally a free parameter in the models. We expand on this in Section 4.4.3. The parameters in which we tested our models on for each genomic region are shown in Tables 2.1 (AHR) and 2.2 (ARNT).

2.2.2 AHR Deletion Analysis

The profile-likelihood surfaces for each parameter are shown for each model in Figures S16 and S17. The standing variant source and staggered sweeps models consistently have higher likelihoods than the concurrent sweeps model. All models have their highest likelihoods at extremely strong selection pressures, between 60% and 92% reduction in fitness for individuals homozygous for intact haplotypes compared to individuals homozygous for deletions (Figure 2.19). The models that include a frequency of the deletion haplotype, the standing

variant source and staggered sweeps models, both have maximum composite-likelihood estimates of fraction of migrants in in g_{R2} that imply the initial deletion haplotype was very rare (Figure 2.20D). The inferred standing time between introgression and selection is 18 generations in the standing variant source model and 16 generations in the staggered sweeps model (Figure 2.20B). We consider this an upper bound because the inferred standing time may have actually been shorter had we sampled a more southern *F. heteroclitus* population. Under the incomplete lineage sorting model, the inferred minimum age of the deletion is 10,000 generations (Figure 2.20C). This time specifies the time between the splitting of h_{T4} and g_{R2} and the onset of selection. Overall, the standing variant source and staggered sweeps models have the highest likelihoods. We are unable to distinguish between them (Figure 2.20A). This is not surprising because these two models leave very similar signatures on haplotype patterns. Even though we are not able to distinguish between whether the deletion had already swept in h_{T4} or not prior to migration, the staggered sweeps model is more plausible. In the standing variant source model, the deletion haplotype migrated into *F. grandis* from *F. heteroclitus* where it was standing at frequency 2x10-6. Given the high estimates for the strength of selection, it is not unlikely for the deletion haplotype to sweep from a very low initial frequency. However, with such rare migration between the species, the probability of that deletion haplotype migrating would be very low. It is more probable to have arrived from a source population where the haplotype is at very high frequency as is the case in the staggered sweeps model.

2.2.3 ARNT Region Analysis

We ran separate analyses for a sweep that finishes at both 30% and 45% frequency, the range of the observed putatively introgressed haplotype frequency. For both final frequencies, the profile-likelihood surfaces for each parameter are almost identical. This is because when the sweep ends at intermediate frequencies, the same selection coefficient will lead to a very

similar number of generations that the sweep takes to reach each final frequency. Our method uses this sweep duration to describe patterns of haplotype similarity among populations, and therefore the results for each final frequency are indistinguishable. We therefore present results for a sweep that ends at 45% frequency and note when results differ for a sweep that ends at 30% frequency.

The results from the ARNT region are largely consistent with the results from the AHR deletion region. Like the AHR deletion region, the adaptive introgression models have much higher likelihoods than the incomplete lineage sorting model (Figure 2.21). In addition, the standing variant source and staggered sweeps models consistently have higher likelihoods than the concurrent sweeps model. All models have their highest likelihoods at strong selection pressures, though weaker than in the AHR deletion region (Figure 2.22A). The standing variant source and staggered sweeps models both have the highest likelihood for a 55% reduction in fitness for individuals without any copies of the selected variant. This is consistent with a sweep that lasts 30-31 generations, the lower bound of this estimate corresponding to a final frequency of 30% and the upper bound, 45%. Inferred migration rates in the concurrent sweeps model are the same as in the AHR deletion region, at approximately 10-5 proportion migrants per generation. For the models that include the fraction of migrants in g_{R2}, the standing variant source and staggered sweeps models, the former has a maximum composite-likelihood estimate at 5x10-6 and the latter at 10-6 (Figure S19D). The inferred standing time between introgression and selection is 10 generations in the standing variant source model and 6 generations in the staggered sweeps model (Figure 2.22B). Thus, introgression was extremely recent and occurred approximately 36-37 generations ago in the staggered sweeps model (upon inclusion of an estimate of the duration of the sweep). This is consistent with results from the AHR deletion region, in which introgression was inferred to occur 34 generations ago. Under the incomplete lineage sorting model, the inferred minimum age of the deletion is 5000 generations, shorter than in the AHR deletion region (Figure

2.22C). This time specifies the time between the splitting of h_{T4} and g_{R2} and the onset of selection.

2.2.4 Details of modifications to DMC

We begin with a brief description of the model framework to provide some background for this section. The inference method, DMC, uses coalescent theory to model how a sweep acts to modify the variance and covariance in mean-centered neutral allele frequencies among populations at linked loci, relative to neutral population structure (14). The variance in these allele frequencies at a locus within a population can be modeled as a function of the ancestral allele frequency at the locus and the probability that a pair of alleles sampled within a population coalesce before reaching the ancestral population. The ancestral population is the common ancestor of all populations that are considered. The covariance of allele frequencies between populations at a locus is the same as the variance within a population except it instead considers the probability that a pair of alleles, one sampled from each of the two populations of interest, coalesce before reaching the ancestral population. The ancestral allele frequency used to estimate the variances and covariances at a locus is the same for each model; it is the probabilities of coalescing that differ among models and allow us to distinguish among and characterize scenarios of convergent adaptation. The modifications and new models that we introduce all inform these probabilities of coalescing.

The models that we outline are concerned with defining a matrix of probabilities of coalescing due to a sweep, $\mathbf{F^{(s)}}$, which is a function of the neutral probabilities of coalescing due to population structure ($\mathbf{F}$) the recombination distance between the locus of interest and the selected site, and free parameters such as the location of the selected site, the strength of selection, and others specific to the models. The rows and columns of $\mathbf{F^{(s)}}$ and $\mathbf{F}$ represent the different populations we consider.

2.2.4.1 Modification of the probability that a lineage does not recombine out of the sweep (y)

Previously, the probability that a lineage stays linked to the selected variant during the sweep used approximations that assumed the strength of selection was very low ($s \ll 1$). However, the length of haplotype similarity within resistant *F. grandis* populations implies that selection was extremely strong and thus violates these assumptions. Instead of using this analytical approximation, we used numerical analysis as outlined by Charlesworth and Charlesworth 2010 (37). We calculated the probability that a lineage stays linked to the selected variant during the sweep by taking the product of the probabilities that the lineage does not recombine out each generation, which change according to the frequency of the deletion. To get the frequency of the deletion each generation, we simulated a deterministic sweep from starting frequency value $1/2N_e$ or low standing frequency (depending on the model) to frequency 0.85 in AHR and 0.3 or 0.45 in ARNT. We used an additive model of selection in which the homozygote of the deletion has fitness 1, the heterozygote $1 - s/2$, and the homozygote of the intact haplotype $1 - s$. Thus,

$$y = \prod_{i=1}^{\tau} \left[1 - r(1 - x_i)\right], \tag{2.1}$$

where τ is the number of generations the sweep lasts, and x_i is the frequency of the selected variant in generation i.

2.2.4.2 Allowing for different effective population sizes between species

Our two species have different effective population sizes, with *F. heteroclitus* being larger than *F. grandis*. Two terms in the model are parameterized by the effective population size: *i)* the probability that a lineage does not recombine out of the sweep (y) and *ii)* the rates of coalescence and recombination during standing phase. While the written model assumes that these terms don't vary according to effective population size, in practice we estimate and apply them separately for each population. Effective population sizes for each species

were inferred from estimates of π, using the Atlantic herring mutation rate (38). There are few estimates of mutation rate existing in fish and this mutation rate is on the same order of magnitude, and within the confidence interval of other known rates in fishes, such as the estimated cichlid mutation rate (39). More importantly, the mutation rate itself is not a parameter in these models, and the effective population sizes estimated from it play a small role in them. The models are robust to effective population sizes changing within an order of magnitude. The effective population size for *F. heteroclitus* is 2.25×10^6 and for *F. grandis* is 8.00×10^5.

2.2.4.3 Assignment of selected site to be the AHR deletion

In the analysis of the AHR deletion on chromosome 1, we used the midpoint of the deletion breakpoints that are furthest from each other and therefore did not vary the position of the selected site in the inference. We defined the breakpoints as specified in Figure S14. Therefore, we set the locus experiencing selection to be 762.684 kb and assumed normal recombination outside the breakpoints as we did not consider SNPs falling inside our defined breakpoints. Additionally, this allowed us to utilize only the SNPs with reads in all populations.

2.2.4.4 Staggered Sweeps Model

Prelude The resistant *F. heteroclitus* population from which the haplotype migrated is called the 'source' (j) and the resistant *F. grandis* population that received migrants is called the 'recipient' (i). In this model, there is a sweep in the source. After the sweep finishes, there is migration from the source to the recipient. We specify migration as a single pulse of admixture (lasts one generation), in which a proportion g of alleles in the recipient population are migrants from the source. After some standing time t_m, there is a sweep at that locus in the now resistant *F. grandis* recipient population (i).

The sections below describe the modified probability of coalescing within or between

pairs of populations due to migration and selection. Since we know g should be very low, we can assume that the probability a lineage in the recipient population that is not a migrant coalesces with lineages in other populations or lineages in its own population is the same as the neutral probability of coalescing estimated from the data.

The following terms are used multiple times in the model specification and are therefore described here:

- r is the recombination rate between the selected site and the neutral site of interest in a single generation
- $y = \prod_{i=1}^{\tau} [1 - r(1 - x_i)]$ is the probability that a lineage does not recombine out of the sweep (i.e. remains linked to the selected variant). τ is the number of generations the sweep lasts, and x_i is the frequency of the selected variant in generation i.
- e^{-rt_m} is the probability a lineage does not recombine out of its background during the standing time (t_m) between migration and the sweep
- p_m is the probability a lineage sampled from the recipient (resistant) population migrates back to the source population. $p_m = y\,(e^{-rt_m} + (1 - e^{-rt_m})g) + (1 - y)g$. A lineage does not migrate back with probability $1-p_m = y(1-e^{-rt_m})(1-g)+(1-y)(1-g)$.

Below, each heading describes independent probabilities. Nested ones are conditional on the statement of the overarching subsection heading.

Within resistant, source *F. heteroclitus* population (j) We can treat this case the same as independent sweep model (14).

$$f_{jj}^{(s)} = y^2 + (1 - y^2) f_{jj} \tag{2.2}$$

Within recipient population (i) First we define a term z that is used multiple times in this case. Conditional on both lineages recombining out at some point before migration,

the probability that they coalesce is $z = g^2 f_{jj}^{(s)} + 2g(1-g)f_{ij} + (1-g)^2 f_{ii}$. Here, whether the lineages coalesce or not depends on whether they still migrate back or not.

- Neither lineage recombines out of the sweep with probability y^2. Now, they enter standing phase that lasts time t_m before migration.
  - A coalescence or recombination event occurs before migration with probability $w = 1 - e^{-(\frac{1}{2Ng}+2r)t_m}$. This event is coalescence with probability $\frac{1}{1+4Ngr}$. Otherwise, one of the lineages recombines off of the introgressed 'haplotype' with probability $\frac{4Ngr}{1+4Ngr}$. We call the lineage that recombined off L1, and the other lineage that is still linked L2. In the case that the first event is recombination:
    - L2 does not recombine out with probability e^{-rt_m}, meaning it must migrate back to the source population (j). The lineage that did recombine out also migrates back with probability g or does not migrate back with probability $1-g$. Thus they coalesce with probability $gf_{jj}^{(s)} + (1-g)f_{ij}$.
    - L2 also recombines out with probability $1 - e^{-rt_m}$. Conditional on this, the lineages coalesce with probability z (defined earlier).
  - A coalescence or recombination event does not occur before migration with probability $1-w$. Therefore both lineages migrate back to the source population, and coalesce there with probability $f_{jj}^{(s)}$.
- Both lineages recombine out of the sweep with probability $(1-y)^2$. Conditional on this, the lineages coalesce with probability z (defined earlier).
- One lineage recombines out of the sweep while the other doesn't with probability $2y(1-y)$. We call the lineage that recombined out L1, and the other lineage that is still linked L2.

– L2 does not recombine out with probability e^{-rt_m}, meaning it must migrate back to the source population (j). The lineage that did recombine out also migrates back with probability g or does not migrate back with probability $1-g$. Thus they coalesce with probability $gf_{jj}^{(s)} + (1-g)f_{ij}$.

– L2 recombines out during standing phase with probability $1-e^{-rt_m}$. Conditional on this, the lineages coalesce with probability z (defined earlier).

Altogether:

$$\begin{aligned} f_{ii}^{(s)} =& y^2 \left[w \left(\frac{1}{1+4Ngr} + \frac{4Ngr}{1+4Ngr} (e^{-rt_m}(gf_{jj}^{(s)} + (1-g)f_{ij}) + (1-e^{-rt_m})z) \right) + (1-w)f_{jj}^{(s)} \right] \\ &+ (1-y)^2 z + 2y(1-y) \left[e^{-rt_m}(gf_{jj}^{(s)} + (1-g)f_{ij}) + (1-e^{-rt_m})z \right] \end{aligned} \tag{2.3}$$

Between recipient (i) and source (j) Both the recipient and the source experience a sweep. To get the probability that one lineage sampled from each coalesce, we condition on whether the lineage from i migrates to j or not (p_m and $1-p_m$, respectively).

$$f_{ij}^{(s)} = p_m f_{jj}^{(s)} + (1-p_m)f_{ij} \tag{2.4}$$

Between recipient (i) and non-source populations (k) The probability that a lineage sampled from i (recipient) and a lineage sampled from k (not source) coalesce depends on whether the lineage in i migrates to j (source population).

$$f_{ik}^{(s)} = p_m f_{jk} + (1-p_m)f_{ik} \tag{2.5}$$

2.2.4.5 Modification of models to exclude selection in the source population of the putatively introgressed haplotype

In the ARNT region of chromosome 10, there is no signal of selection in h_{T4} or any other *F. heteroclitus* population we included in our model. Therefore, the following describes the

probability of coalescing for the terms that are altered in each model. For all models, the probability of coalescing within the source population (j) is no longer modified by selection, and is rather the neutral probability of coalescing:

$$f_{jj}^{(s)} = f_{jj} \tag{2.6}$$

In the incomplete lineage sorting model, there is no migration modeled between populations. Therefore, the probability of coalescing for every pair of lineages sampled from different populations is the neutral probability estimated for each population pair.

The probability of coalescing within the recipient population (g_{R2}) changes for both the staggered sweeps and concurrent sweeps models. This is because lineages sampled in the recipient population that migrate back to the source population were previously considered to experience another sweep in the source population. The change we make in the staggered sweeps model is more straightforward: replace all $f_{jj}^{(s)}$ terms with f_{jj} (where j is the source population). The following describes the modifications we make to the concurrent sweeps model:

As described in Lee and Coop (2017), Q describes the probability that two lineages sampled at the selected site coalesce in the recipient population (i), such that a single lineage migrates back to the source population (j). $1 - Q$ is therefore the probability that two lineages sampled at the selected site do not coalesce before migrating back to the source population. We can condition on these two scenarios to get the probability of coalescing for a pair of lineages sampled at a neutral site a given distance away from the selected site.

$$\begin{aligned} f_{ii}^{(s)} &= Q\left(y^2 * 1 + (1-y)^2 f_{ii} + 2y(1-y)f_{ij}\right) + (1-Q)\left(y^2 f_{jj} + (1-y)^2 f_{ii} + 2y(1-y)f_{ij}\right) \\ &= y^2\left(Q + (1-Q)f_{jj}\right) + (1-y^2)f_{ii} + 2y(1-y)f_{ij} \end{aligned} \tag{2.7}$$

For all adaptive introgression models, the probability of coalescing between the recipient

population (g_{R2}) and all other populations depends on whether the lineage sampled in the recipient population migrates back to the source populations. This probability of migrating to the source (p_m) varies by model and is described below:

For the standing variant source and staggered sweeps models, v is the probability that a lineage does not recombine out of the selected background during standing phase, if it didn't recombine out during the sweep. As a reminder, t_m is the amount of standing time between migration and the sweep and g is the standing frequency of the selected variant.

$$v = e^{-r*t_m} \tag{2.8}$$

Standing variant source and staggered sweeps:

$$p_m = y(v + (1 - v)g) + (1 - y)g \tag{2.9}$$

Concurrent sweeps:

$$p_m = y \tag{2.10}$$

According to the different probabilities of migrating to the source for each model, the probability of coalescing between the source and recipient populations for all adaptive introgression models is as follows:

$$f_{ik}^{(s)} = p_m f_{jk} + (1 - p_m) f_{ik} \tag{2.11}$$

2.2.5 Figures and Tables

Parameter	Model(s) that include it	Values

Strength of selection (s)	All	0.001, 0.005, 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.08, 0.1, 0.12, 0.14, 0.16, 0.18, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.82, 0.84, 0.86, 0.88, 0.9, 0.92, 0.94, 0.96, 0.98, 1
Standing time between migration and selection (t_m)	Standing variant source, Staggered sweeps	0, 5, 10, 15, 16, 18, 20, 22, 24, 26, 28, 30, 35, 40, 60, 80, 100, 200, 500, 1000, 5000, 10^7
Minimum age of standing variant / Standing time between species split and selection (t)	Incomplete lineage sorting	3000, 4000, 5000, 7500, 9000, 10000, 25000, 50000, 10^5, 10^6, 10^7, 10^8
Frequency of variant before selection / fraction of migrants in R2 (g)	Standing variant source, Staggered sweeps, Incomplete lineage sorting	$6.878 \times 10^{-7} (= 1/2N_e)$, $8.5 * 10^{-7}$, 10^{-6}, $2 * 10^{-6}$, 5×10^{-6}, 8×10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}
Migration rate per generation (m)	Concurrent sweeps	10^{-5},10^{-4}, 5×10^{-4}, 0.001, 0.005, 0.01, 0.1, 0.3

Table 2.1: Parameters that coalescent-based models were tested on in the AHR deletion region.

Parameter	Model(s) that include it	Values

Position of selected site	All	18000003, 18572796, 19145589, 19718382, 20291175, 20863968, 21436761, 22009554, 22582347, 23155140, 23727933, 24300726, 24873519, 25446312, 26019105, 26591897, 26735096, 26878294, 27021492, 27164690, 27355621, 27546552, 27737483, 28310276, 28883069, 29455862, 30028655, 30601448, 31174241, 31747034, 32319827, 32892620, 33465413, 34038206, 34610999
Strength of selection (s)	All	0.001, 0.005, 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.08, 0.1, 0.12, 0.14, 0.16, 0.18, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.82, 0.84, 0.86, 0.88, 0.9, 0.92, 0.94, 0.96, 0.98, 1

Standing time between migration and selection (t_m)	Standing variant source, Staggered sweeps	0, 5, 10, 15, 16, 18, 20, 22, 24, 26, 28, 30, 35, 40, 60, 80, 100, 200, 500, 1000, 5000, 10^7
Minimum age of standing variant / Standing time between species split and selection (t)	Incomplete lineage sorting	3000, 4000, 5000, 7500, 9000, 10000, 25000, 50000, 10^5, 10^6, 10^7, 10^8
Frequency of variant before selection / fraction of migrants in R2 (g)	Standing variant source, Staggered sweeps, Incomplete lineage sorting	$6.878 \times 10^{-7} (= 1/2N_e)$, $8.5 * 10^{-7}$, 10^{-6}, $2 * 10^{-6}$, 5×10^{-6}, 8×10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}
Migration rate per generation (m)	Concurrent sweeps	10^{-5},10^{-4}, 5×10^{-4}, 0.001, 0.005, 0.01, 0.1, 0.3

Table 2.2: Parameters that coalescent-based models were tested on in the ARNT deletion region.

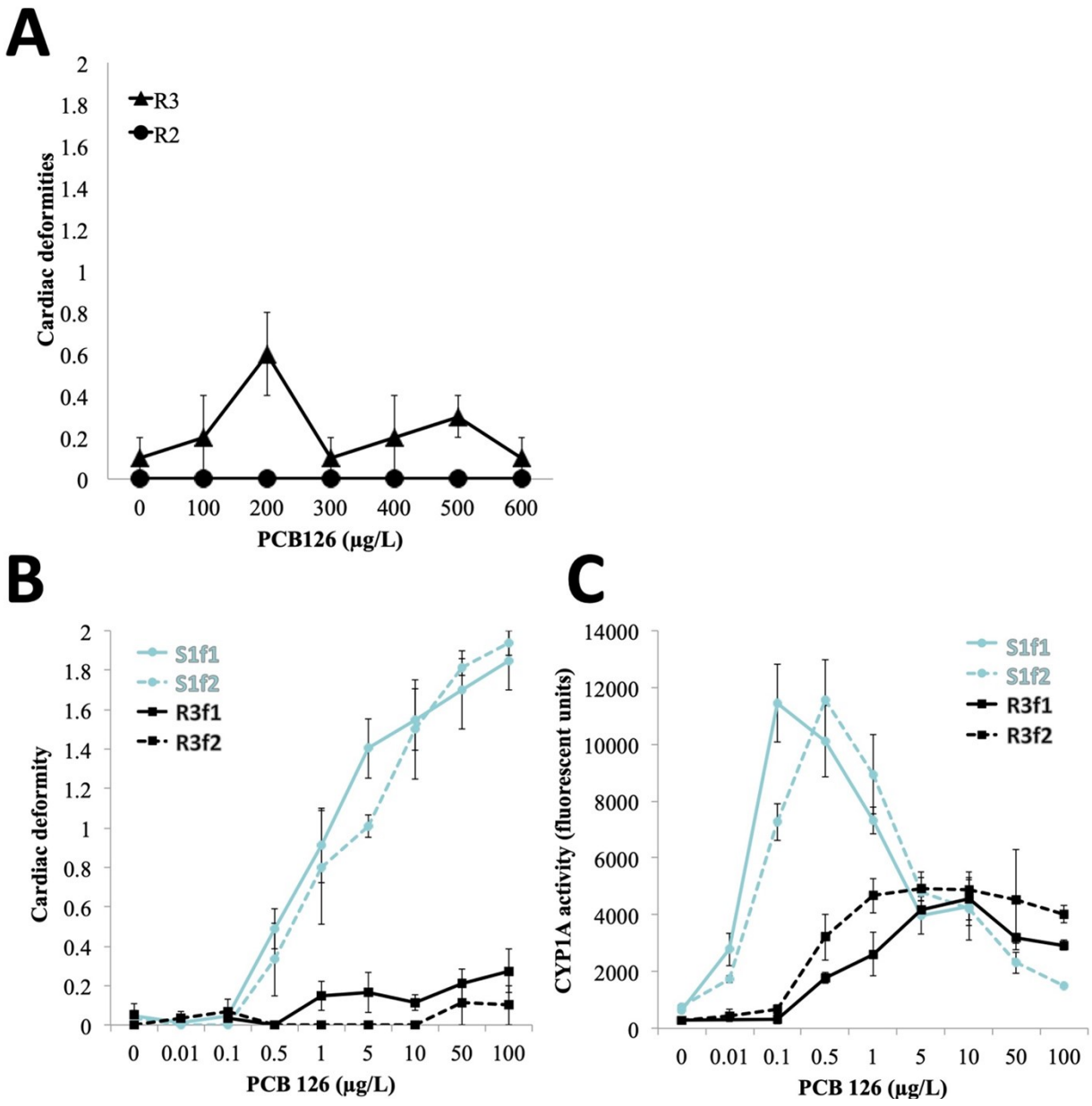

Figure 2.4: (Contined on the following page.)

Figure 2.4: Resistance to high levels of contamination is inherited to two generations out of the stressor environment. A) The most resistant populations of *F. grandis* (e.g., R2 and R3) can withstand PCB126-induced cardiac teratogenesis at concentrations more than 1200x higher than those that induce induces cardiac teratogenesis in reference sensitive populations (B). Populations (e.g., R3 and S1) vary in resistance to PCB126-induced cardiac teratogenesis (B) and CYP1A activation (C), and these differences are retained between first generation (f1) and second generation (f2) animals raised in a common clean environment, indicating that population differences in resistance are heritable.

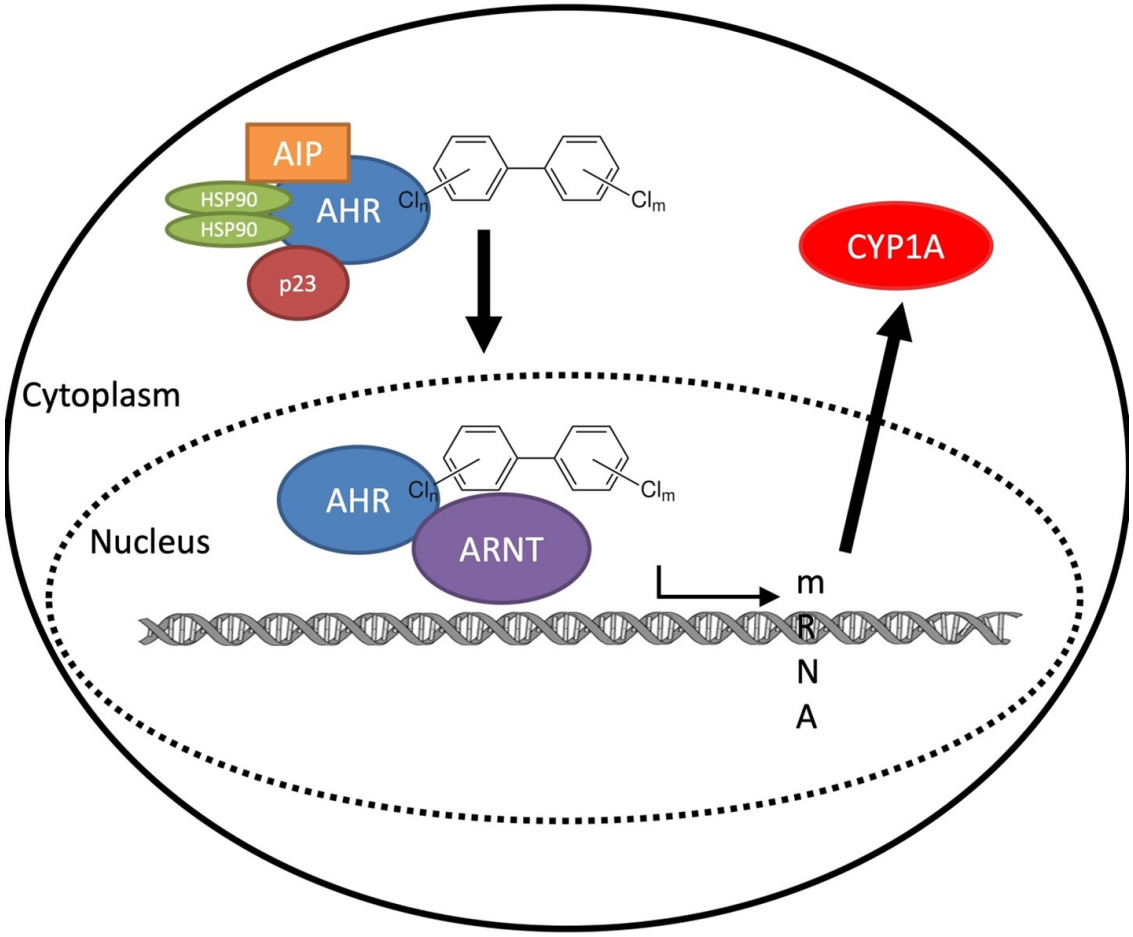

Figure 2.5: Diagram of the aryl hydrocarbon receptor (AHR) signaling pathway. In the cytoplasm, AHR is associated with proteins heat shock protein 90 (HSP90), aryl hydrocarbon receptor interacting protein (AIP, also known as XAP2), and HSP90 co-chaperone (p23). When entering the cell, halogenated aromatic hydrocarbons (HAHs) act as ligands that bind to, and activate, AHR which then translocates to the nucleus. In the nucleus, ligand-activated AHR forms a complex with the aryl hydrocarbon receptor nuclear translocator (ARNT). This complex binds xenobiotic response elements (XREs) and activates transcription of a battery of genes. One of the most strongly up-regulated genes within this battery is cytochrome P450 1A (CYP1A).

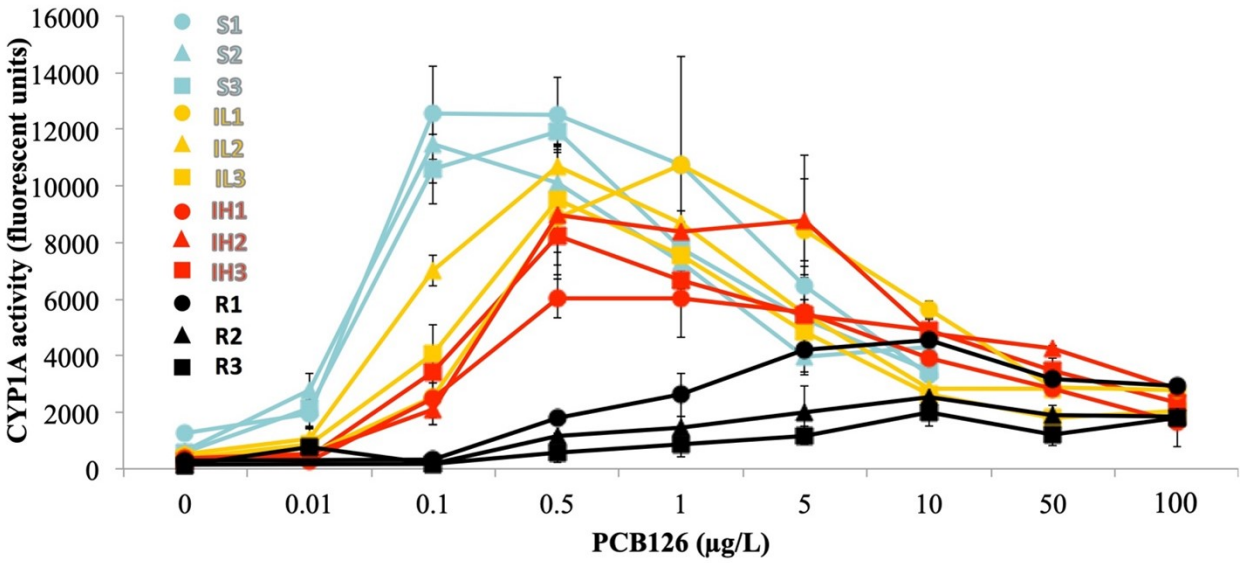

Figure 2.6: Multi-population dose-response curves of CYP1A activity (relative fluorescent units, EROD assay) following embryonic exposure to PCB126. Fish from polluted sites (black curves) were resistant to the dose-dependent induction of the AHR pathway observed in fish from clean sites (blue curves), whereas fish from sites of intermediate pollution exhibited intermediate levels of CYP1A induction. The de-sensitization of the AHR pathway in fish from polluted sites is characterized by a lower total expression, and therefore activity, of CYP1A, as well as activation at higher concentrations of PCB126. Variation among populations in maximal CYP1A induction and the dose that causes significant up-regulation is scaled by level of pollution.

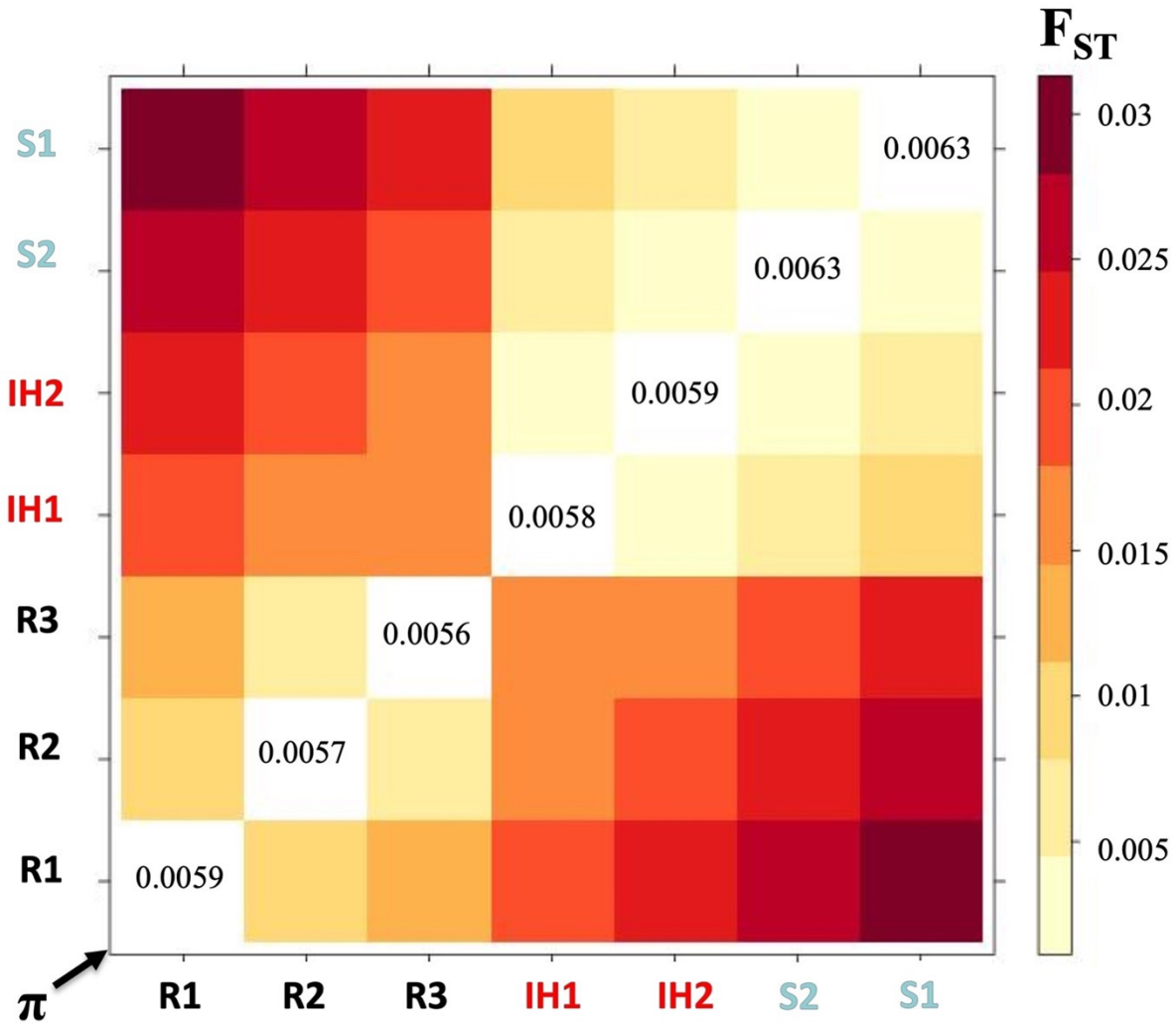

Figure 2.7: Genome-wide differentiation (F_{ST}) between all pairs of populations (colored cells), including nucleotide diversity for each population (π; diagonal). Genome-wide differentiation between populations tends to be low (< 0.03), whereas nucleotide diversity within populations is high. Nucleotide diversity tends to be lower in polluted site populations (red and black population labels) compared to clean site populations (blue population labels).

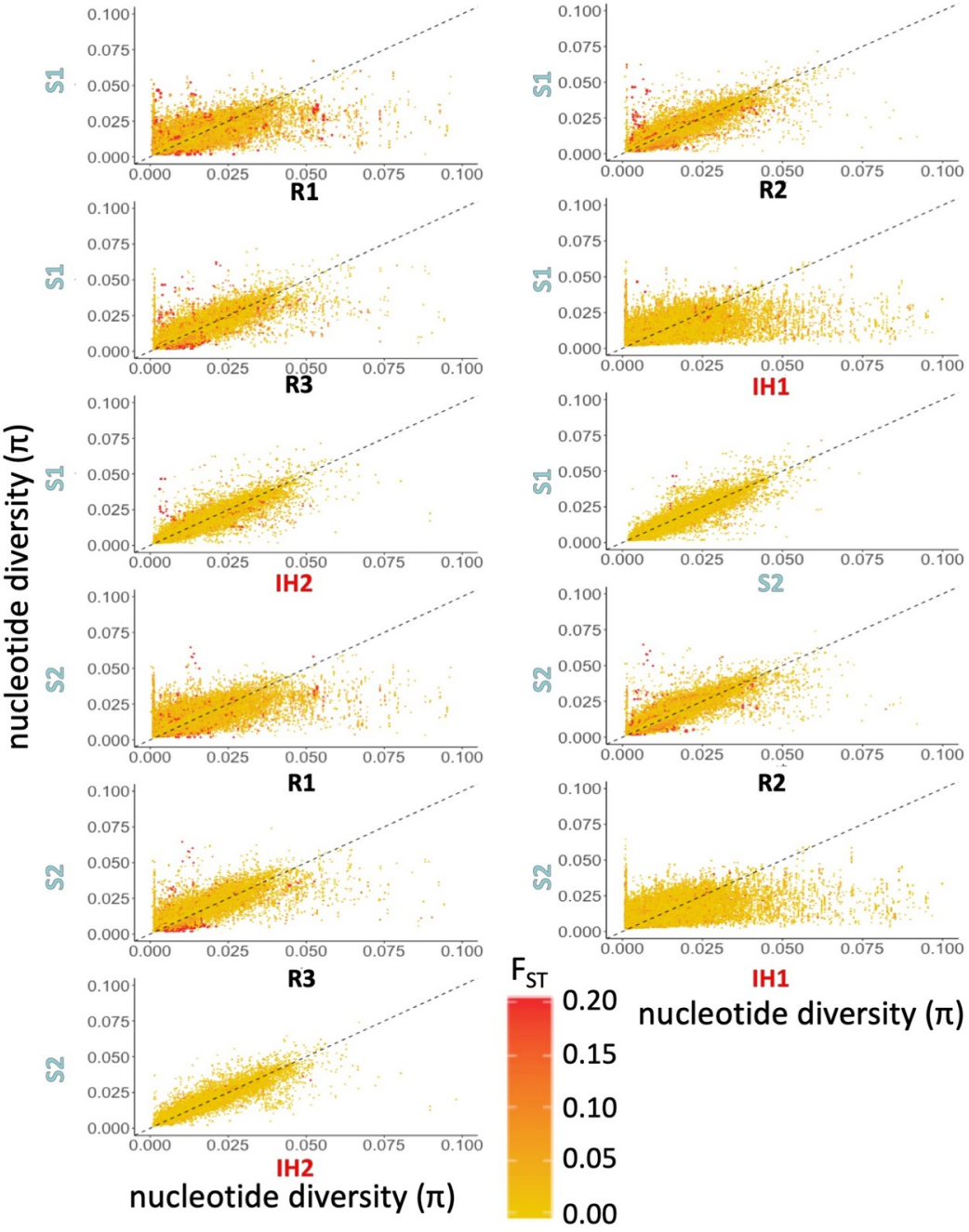

Figure 2.8: (Continued on the following page.)

Figure 2.8: Genome wide nucleotide diversity (π; x and y axes) between pairs of polluted (red and black labels) and clean site (blue labels) populations. Genome-wide nucleotide diversity tends to be reduced in pollution adapted populations (skew to below 1:1 dotted line). Loci most differentiated (F_{ST}) from sensitive populations (red dots) tend to coincide with regions of low π in pollution adapted populations compared to sensitive populations. F_{ST} and π are calculated in 5kb sliding windows (1kb slide).

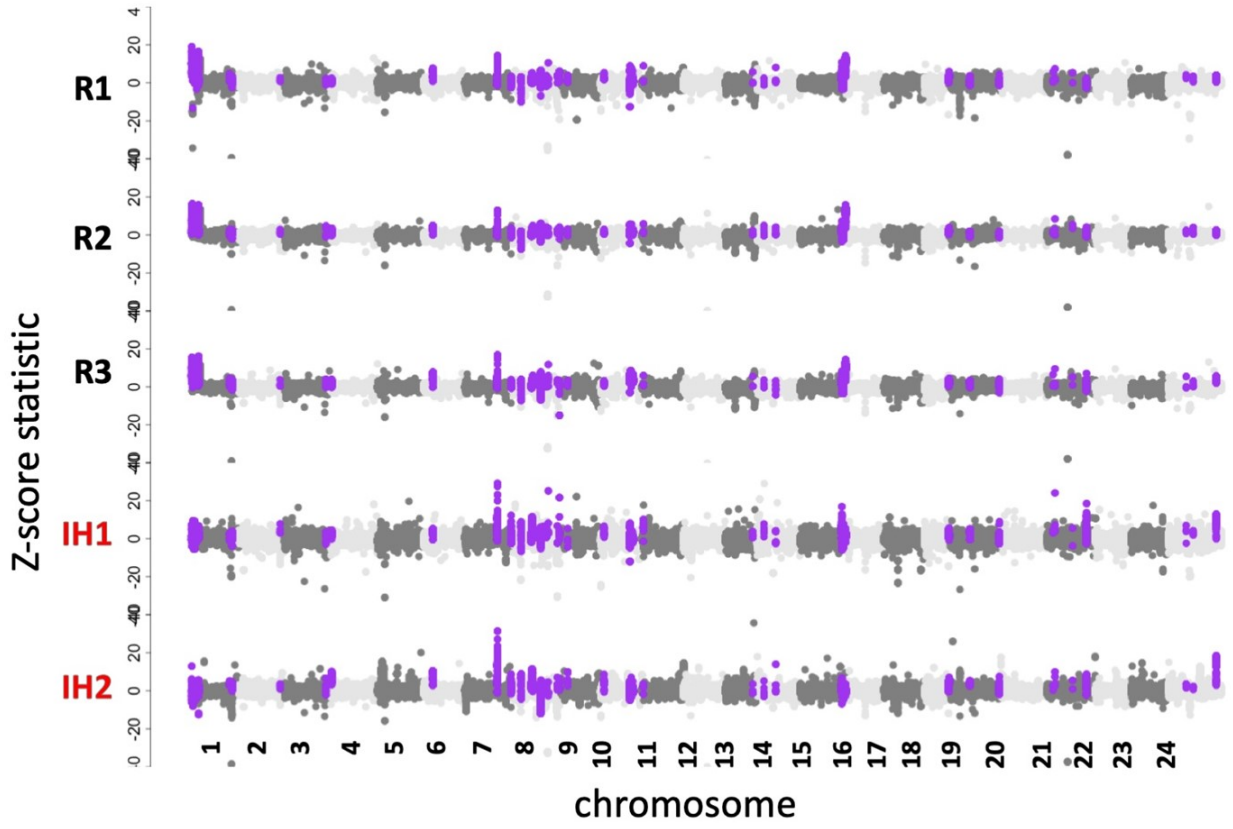

Figure 2.9: Genome-wide signatures of selection for each polluted site population compared to both reference clean site populations. Y-axis is the z-score statistic that combines two signals of natural selection: high divergence (F_{ST}) and low nucleotide diversity (π) in the polluted site population relative to clean site populations, which are each calculated in 20kb sliding windows (1kb slide). Outliers (1% highest z-score) are colored for each population. Those colored purple are outliers that are shared across all polluted site populations. Outliers colored black and red are outliers that are shared only among resistant or intermediate populations, respectively.

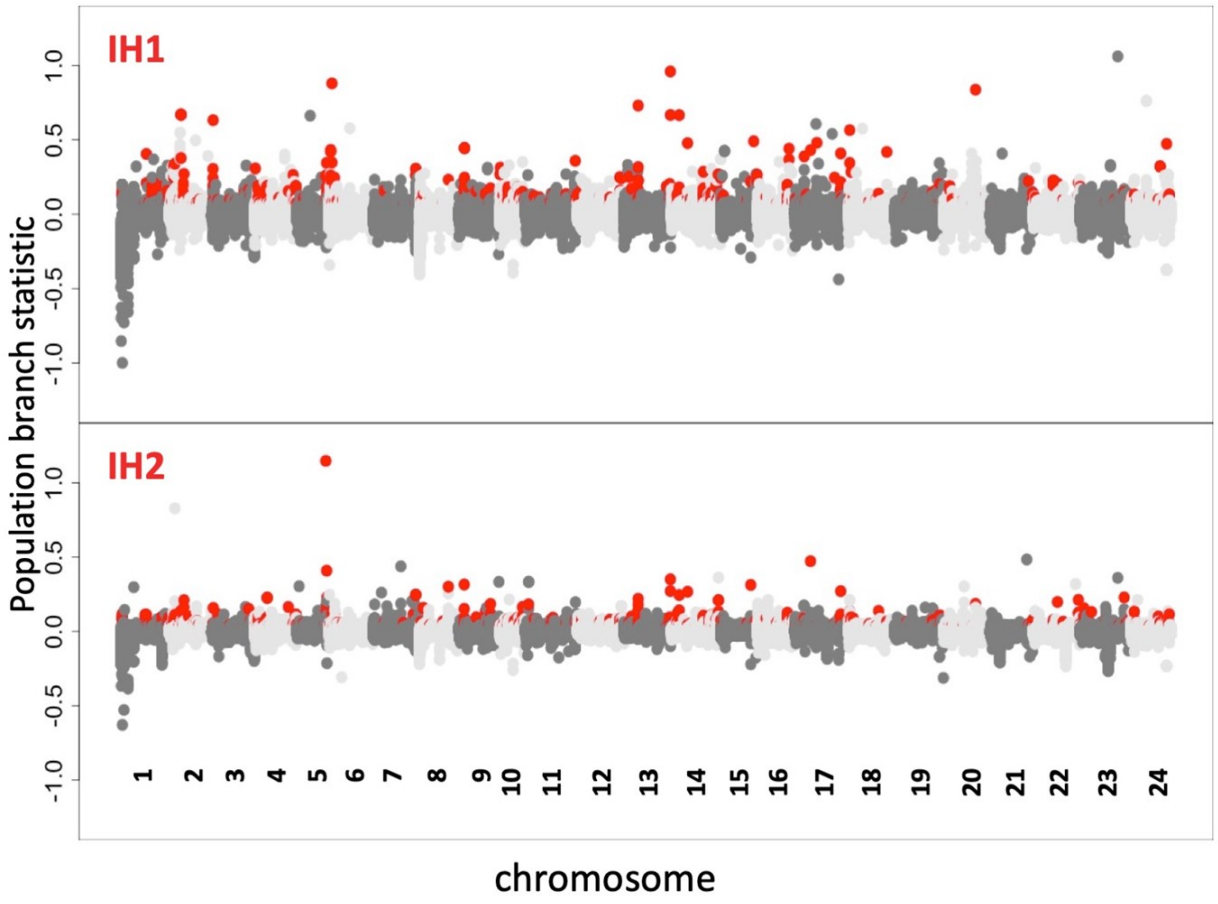

Figure 2.10: Differentiation between each intermediate population (IH1, IH2) and both resistant and sensitive populations (R2 and S1; population branch statistic) indicate no strong signatures of selection. Outliers (highlighted in red) are the 1% most highly differentiated regions (5kb windows), and indicate no consistent or highly differentiated regions that are unique to intermediate populations.

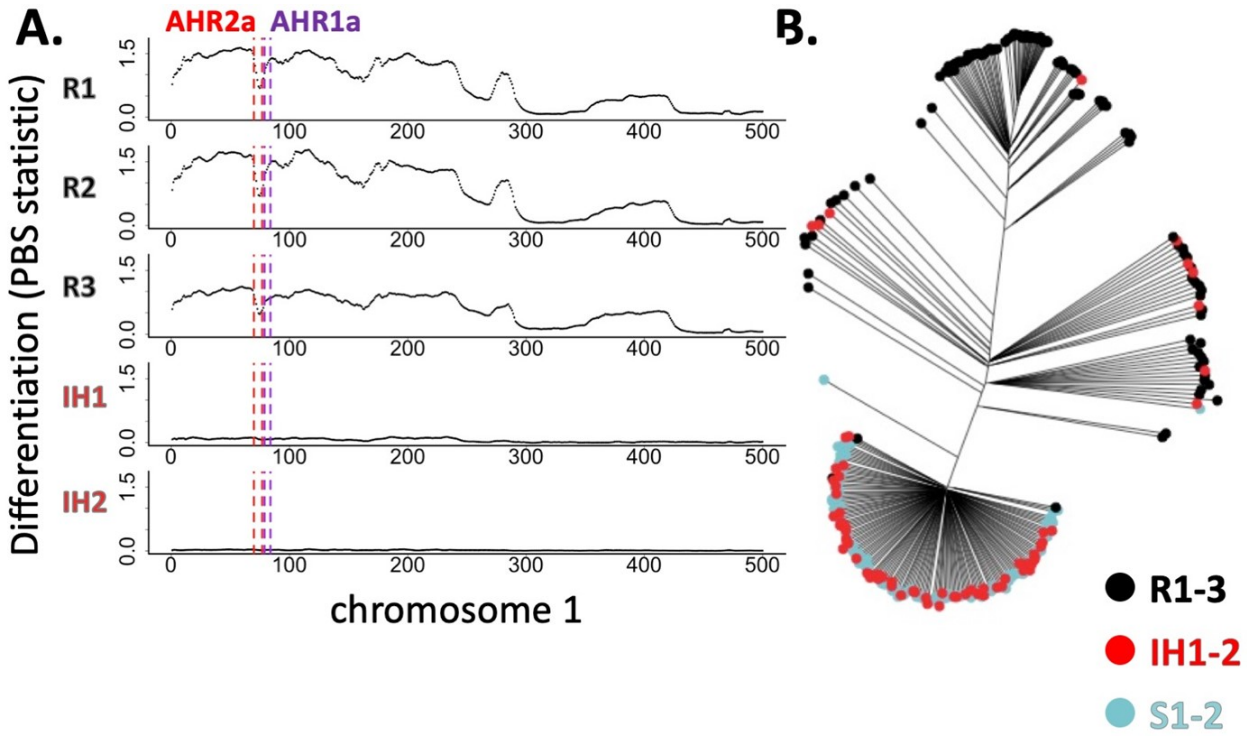

Figure 2.11: Patterns of differentiation surrounding key genes in the aryl hydrocarbon receptor (AHR) signaling pathway reveal signatures of selection in polluted site populations. Highlighted genes include tandem paralogs of the aryl hydrocarbon receptor (AHR), a xenobiotic-activated transcription factor. A) Population divergence measured by population branch statistic between each adapted population and both sensitive reference populations. B) Neighbor-joining tree groups individuals with deletion haplotype together (top), heterozygotes for AHR deletion in the middle, and wild type haplotypes together (bottom).

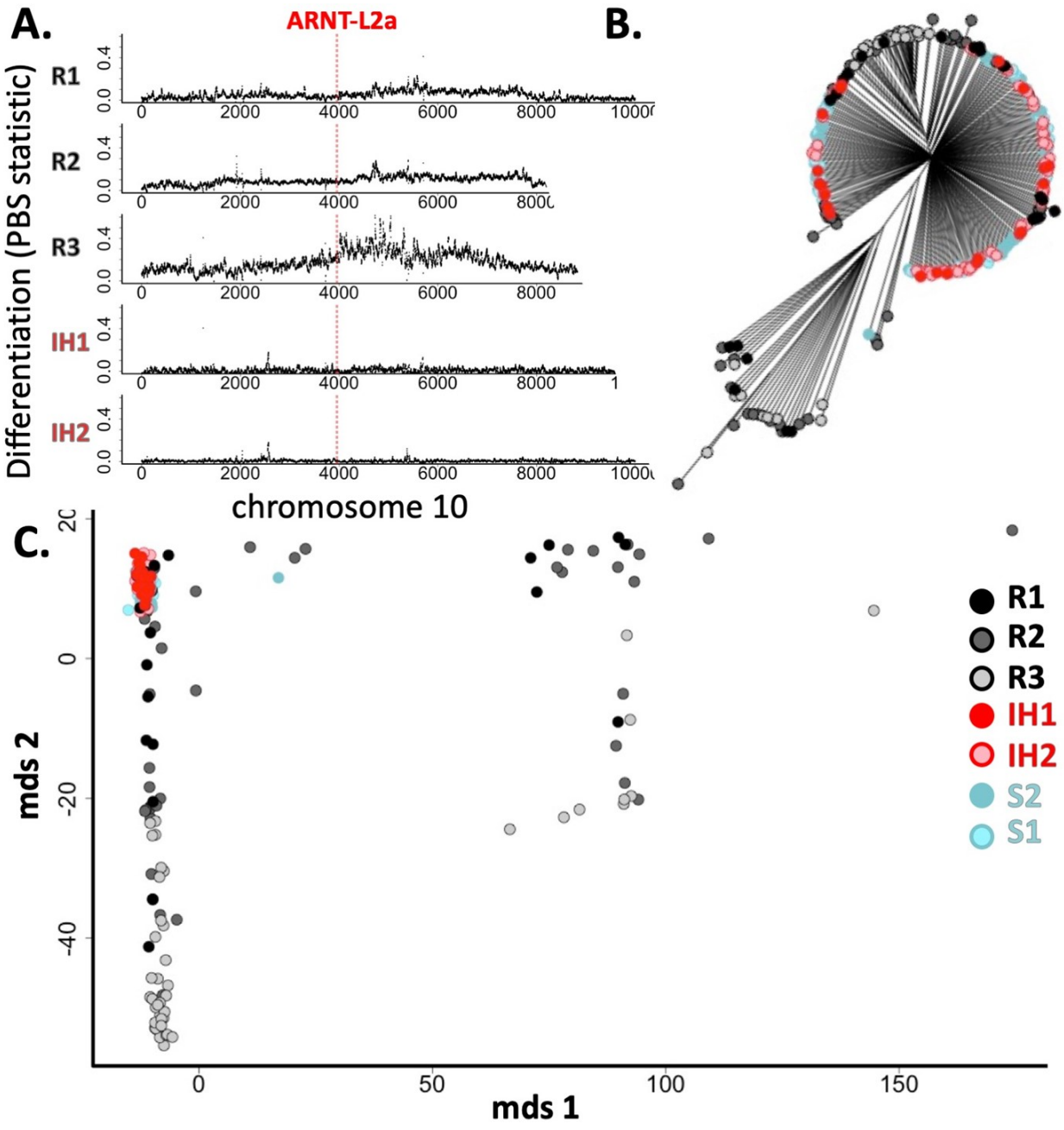

Figure 2.12: (Continued on the following page.)

Figure 2.12: Patterns of differentiation surrounding key genes in the aryl hydrocarbon receptor (AHR) signaling pathway reveal signatures of selection in polluted site populations. Highlighted gene is the aryl hydrocarbon receptor nuclear translocator-like 2a (ARNT-L2a), which is the nuclear dimerization partner of AHR, required for activation of the xenobiotic response pathway. A) Population divergence measured by population branch statistic between each adapted population and both sensitive reference populations. The ARNT gene is located within or near the region of high differentiation. Neighbor-joining tree (B) and MDS plot (C) reveal clustering of haplotypes from sensitive and resistant populations.

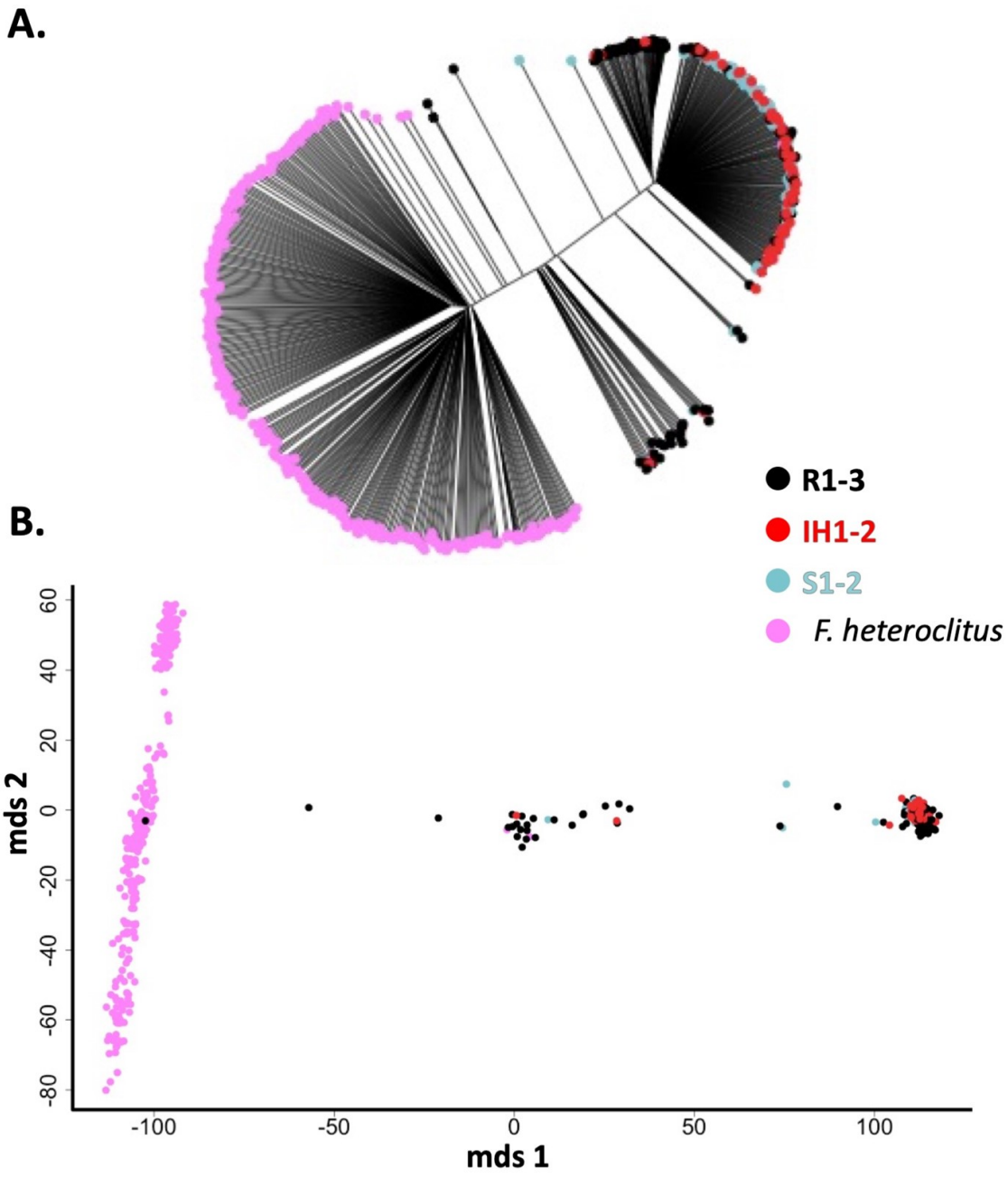

Figure 2.13: (Continued on the following page.)

Figure 2.13: Genetic differentiation encompassing a key AHR binding partner, AHR nuclear translocator-like 2a (ARNT-L2a) in the aryl hydrocarbon receptor (AHR) signaling pathway in individuals from resistant populations reveal higher similarity to *F. heteroclitus* than reference *F. grandis*. (A) A neighbor joining tree and (C) MDS plot reveal clustering of haplotypes suggesting an introgression pattern, later confirmed in population genetic models.

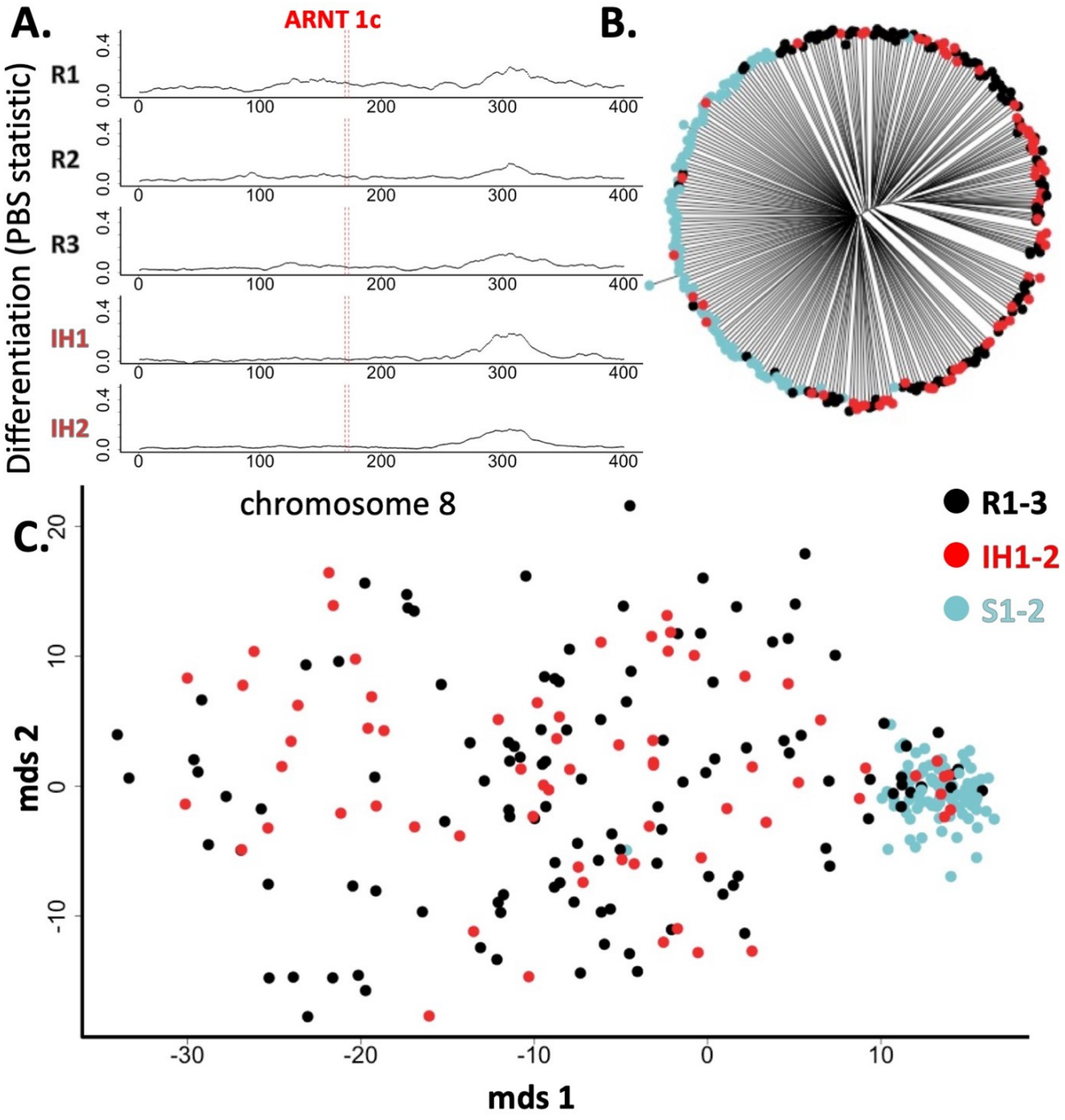

Figure 2.14: (Continued on the following page.)

Figure 2.14: Patterns of differentiation surrounding key genes in the aryl hydrocarbon receptor (AHR) signaling pathway reveal signatures of selection in polluted site populations. Highlighted gene is the aryl hydrocarbon receptor nuclear translocator 1c (ARNT), which is the nuclear dimerization partner of AHR, required for initiation of the xenobiotic response pathway. A) Population divergence measured by population branch statistic between each adapted population and both sensitive reference populations. The ARNT gene is located within the region of high differentiation. Neighbor-joining tree (B) and MDS plot (C) reveal clustering of haplotypes from sensitive and resistant populations.

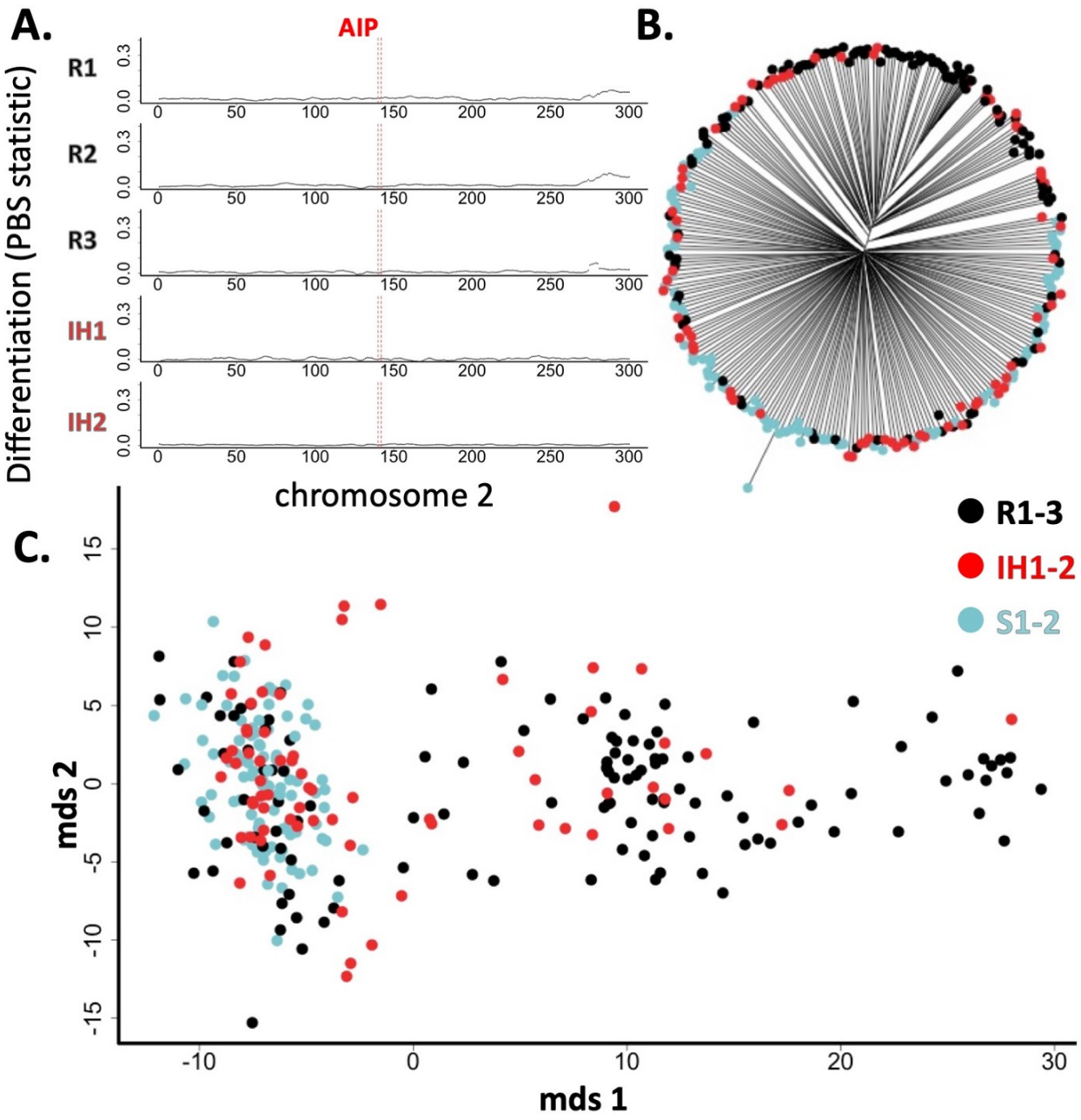

Figure 2.15: (Continued on the following page.)

Figure 2.15: Patterns of differentiation surrounding key genes in the aryl hydrocarbon receptor (AHR) signaling pathway reveal signatures of selection in polluted site populations. Highlighted gene is the aryl hydrocarbon receptor interacting protein (AIP, also known as XAP2) which regulates AHR signaling. A) Population divergence measured by population branch statistic between each adapted population and both sensitive reference populations. The AIP gene is located within the region of high differentiation. Neighbor-joining tree (B) and MDS plot (C) reveal clustering of haplotypes from sensitive and resistant populations.

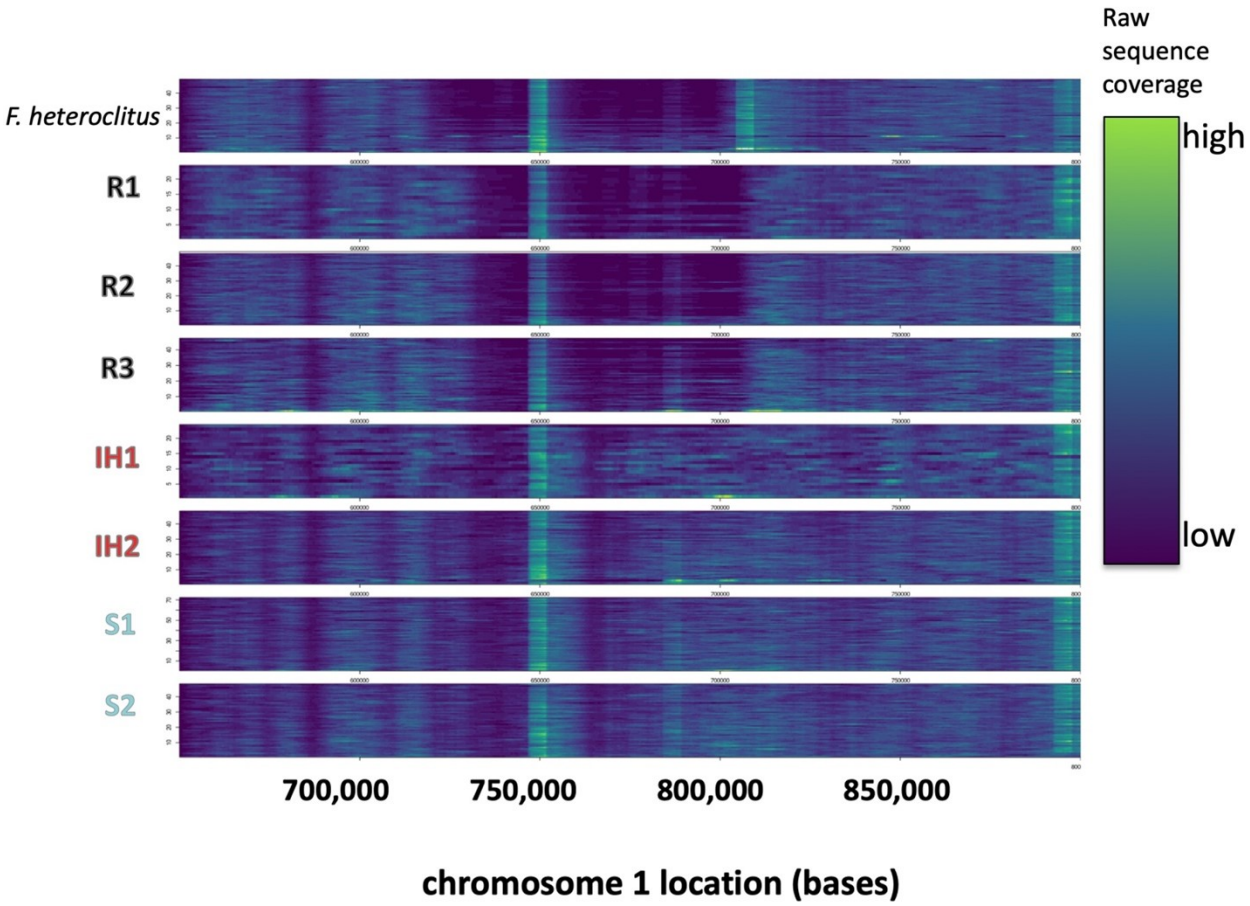

Figure 2.16: A deletion, spanning the last exon of AHR2a and first 6 exons of AHR1a, is present at high frequency in resistant populations (R1, R2) and at low frequency in intermediate populations (IH1, IH2). Each horizontal line represents a single individual and the intensity of green is amount of raw sequence coverage over the represented region. Dark coloring indicates regions where few or no reads mapped to the reference genome from that individual. Though deletions are found in the same genomic region in both *F. heteroclitus* (resistant population Elizabeth River, top panel) and in *F. grandis* polluted populations, the deletion break points appear to be different between *F. heteroclitus* and *F. grandis*.

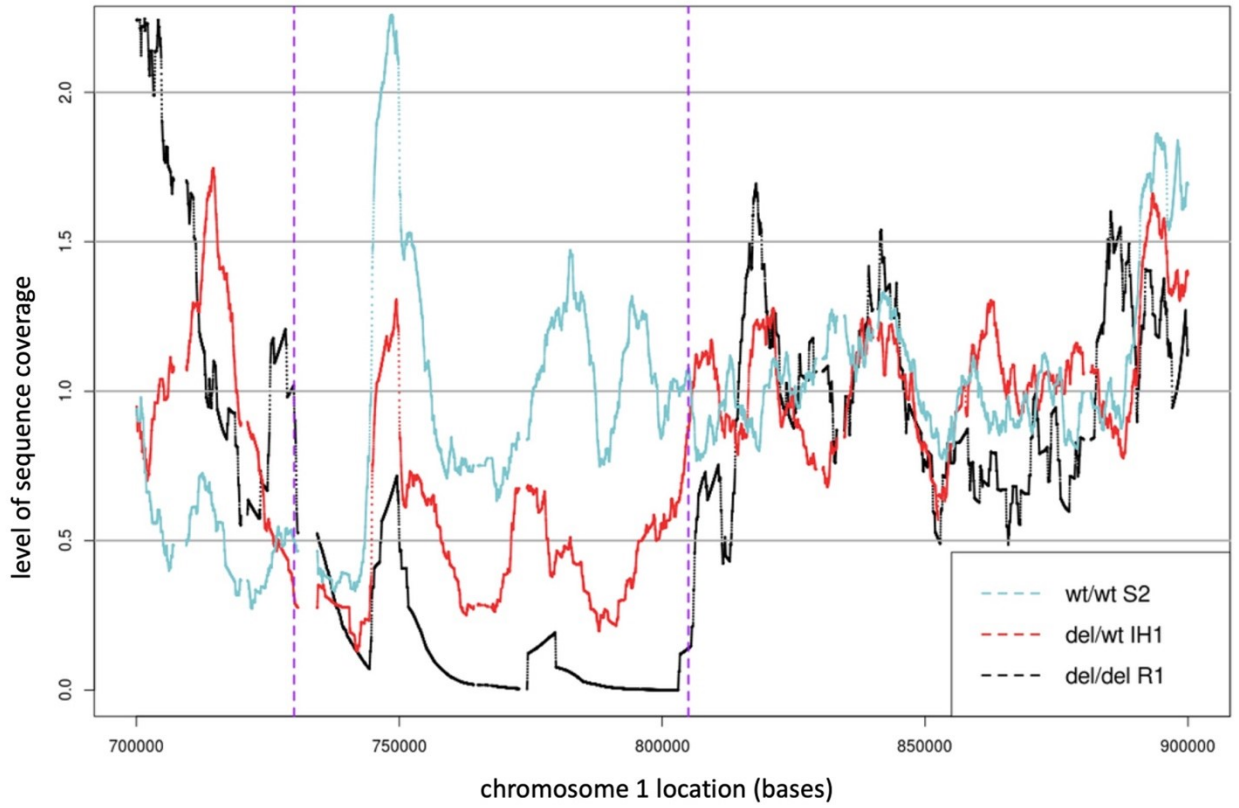

Figure 2.17: The deletion that encompasses portions of genes AHR1a and AHR2a on chromosome 1 is apparent from plots of smoothed sequence coverage from individuals that represent the three genotypic variants (homozygous wildtype, homozygous deletion, heterozygote). The blue line is from an individual from a reference sensitive population (S2) that does not carry the deletion (homozygous wildtype), where sequence coverage is relatively continuous across the region. The black line is from an individual from a resistant population (R1) that is homozygous for the deletion, where sequence coverage drops very low (effectively zero). The red line is from an individual from an intermediate population (IH1) that is heterozygous for the deletion, where reduced sequence coverage is intermediate. Purple dashed lines bracket the most continuous portion of the deletion, which is preceded by a region of likely high mis-mapping, which is common for deletion breakpoints.

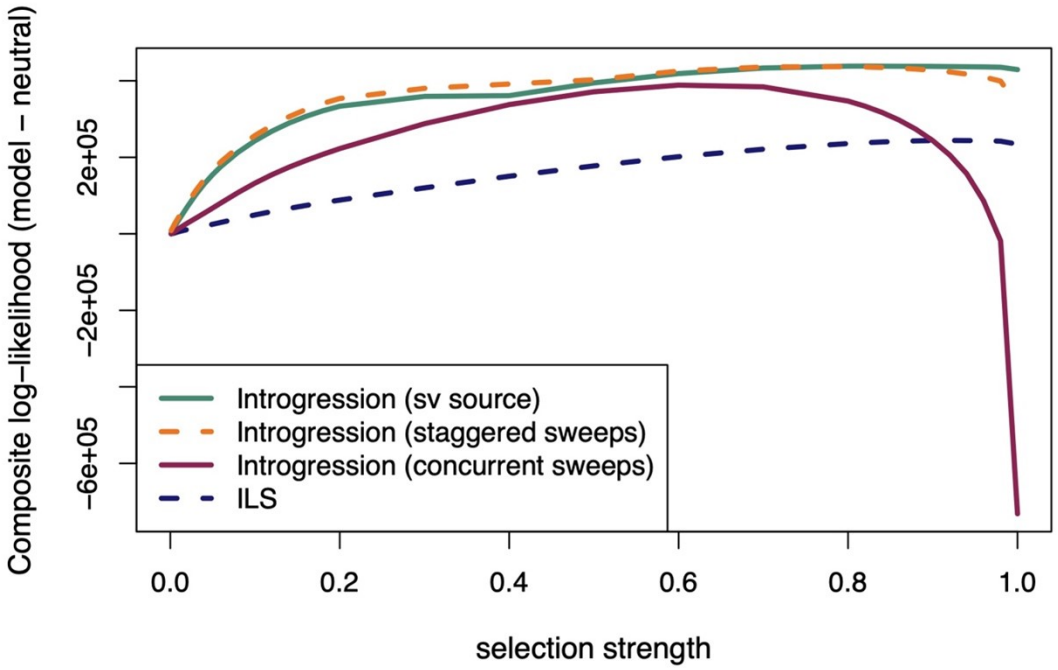

Figure 2.18: Models tested on three pairs of sensitive (black) and resistant (red) populations (one from *F. grandis* and two from *F. heteroclitus* representing Northern and Southern clade). We allowed for selection in hT4 and gR2 for the analysis of the AHR deletion region (chromosome 1) and selection only in gR2 for the ARNT region (chromosome 10). A) Incomplete lineage sorting (incomplete lineage sorting) model: selected variant arose (orange star) in ancestor of *F. heteroclitus* and *F. grandis*, and stood until the onset of selection (blue triangles). B) Standing variant source model: selected variant arose in hT4 (could be along ancestral branch between hS4 and hT4 in this cartoon as well). Haplotype was introduced into gR2 and may have stood for some time in both populations (free parameter) prior to selection. C) Concurrent sweeps model: deletion haplotype arose in hT4 and migrated into gR2 during the sweep in hT4 and began to sweep immediately. D) Staggered sweeps model: Selected variant arose in hT4 and was immediately selected. After fixation, the haplotype migrated into gR2, where it may have stood for some time (free parameter) prior to the onset of selection in gR2.

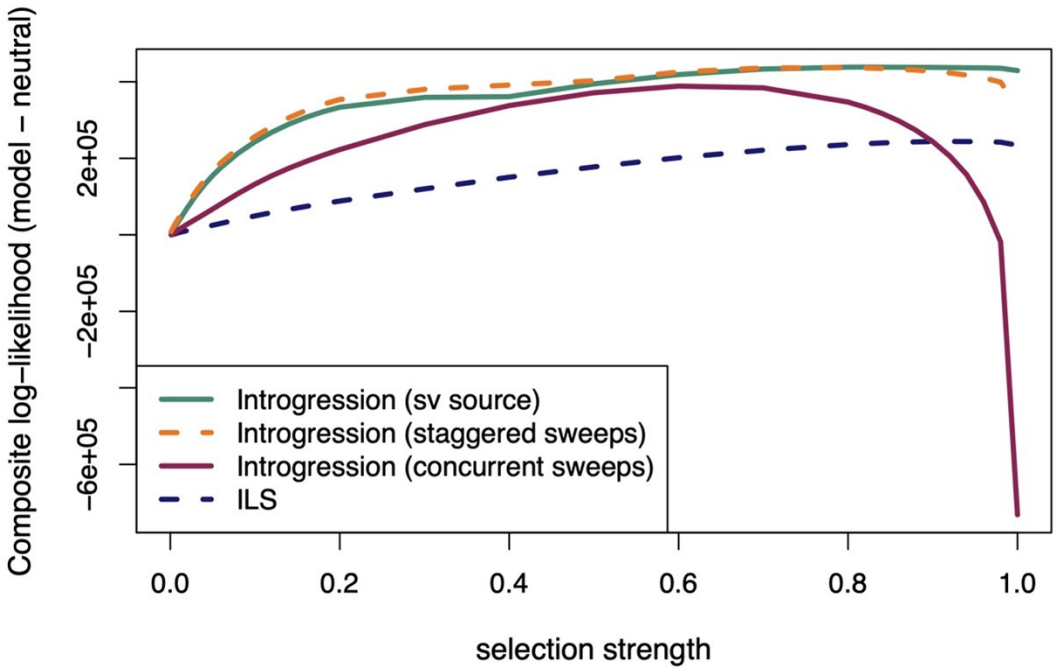

Figure 2.19: Likelihood results from AHR deletion region (chromosome 1). Composite log-likelihood ratio of given model relative to the neutral model of no selection as a function of the selection coefficient (s). The standing variant source and staggered sweeps models have the highest maximum likelihood. For the standing variant source model, the maximum-composite-likelihood estimate (MCLE) is at $s = 0.82$. For the staggered sweeps model, the MCLE is at $s = 0.8$.

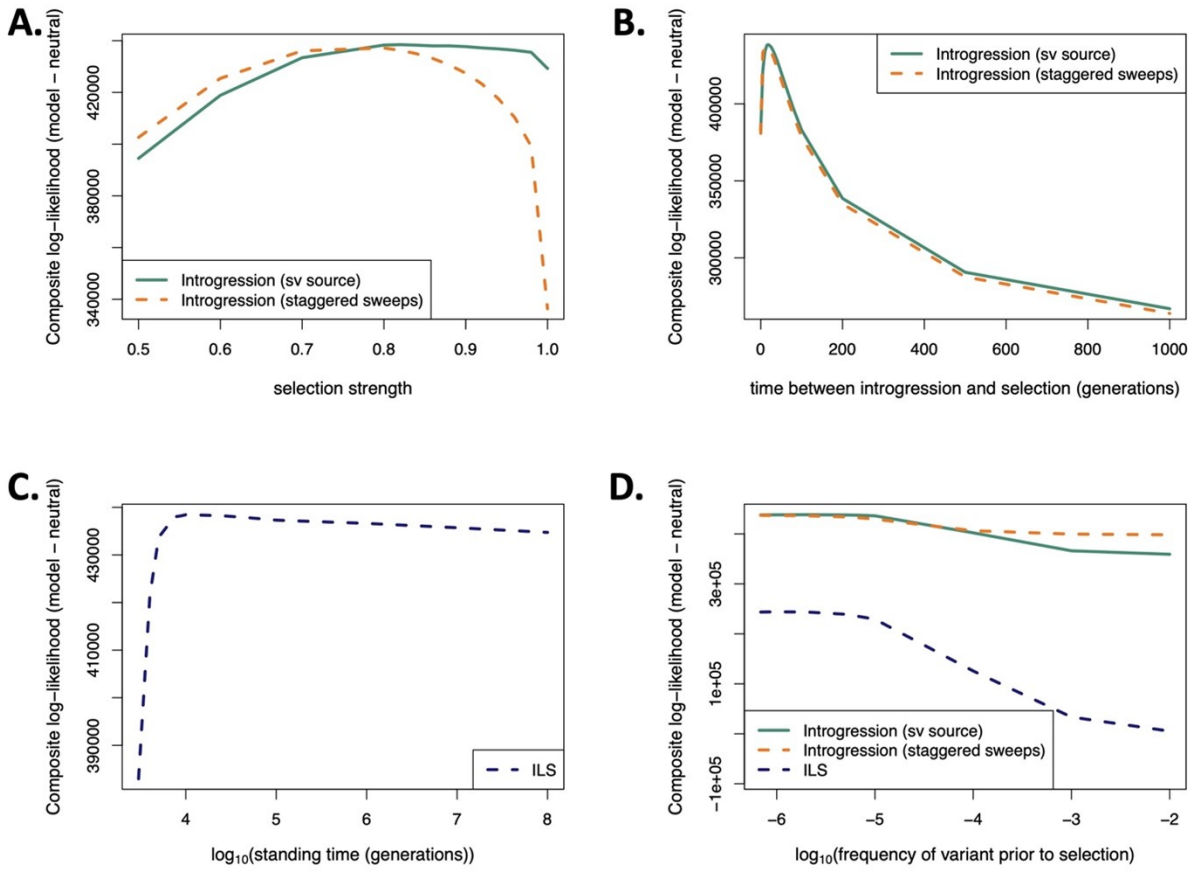

Figure 2.20: (Continued on the following page.)

Figure 2.20: Likelihood results from AHR deletion region (chromosome 1). Composite log-likelihood ratio (CLR) of given model relative to the neutral model of no selection as a function of various parameters. A) CLR as function of selection coefficient zoomed in for the top two models. B) CLR as a function of time between introgression and selection for models that include this parameter. The standing variant source model has a maximum-composite-likelihood estimate (MCLE) of 18 generations while the staggered sweeps model has a MCLE of 16 generations. C) CLR as a function of the duration of the standing time in the incomplete lineage sorting model. This time represents the minimum age of the standing variant and specifically is the time (in generations) between the splitting of hT4 and gR2 and the onset of selection. The MCLE of this time is 10,000 generations but there is little change in likelihood as the parameter increases past this value. D) CLR as a function of the frequency of variant prior to selection. In the top two introgression models, this represents the frequency of the allele upon introgression. For the incomplete lineage sorting model, this parameter represents the frequency at which the variant is standing. We find the MCLE for the staggered sweeps model is $1/2N_e = 6.8x10^{-7}$, where N_e is the effective population size estimate for *F. grandis*. The MCLE for the standing variant source model is 2×10^{-6} and for the incomplete lineage sorting model is 10^{-6}.

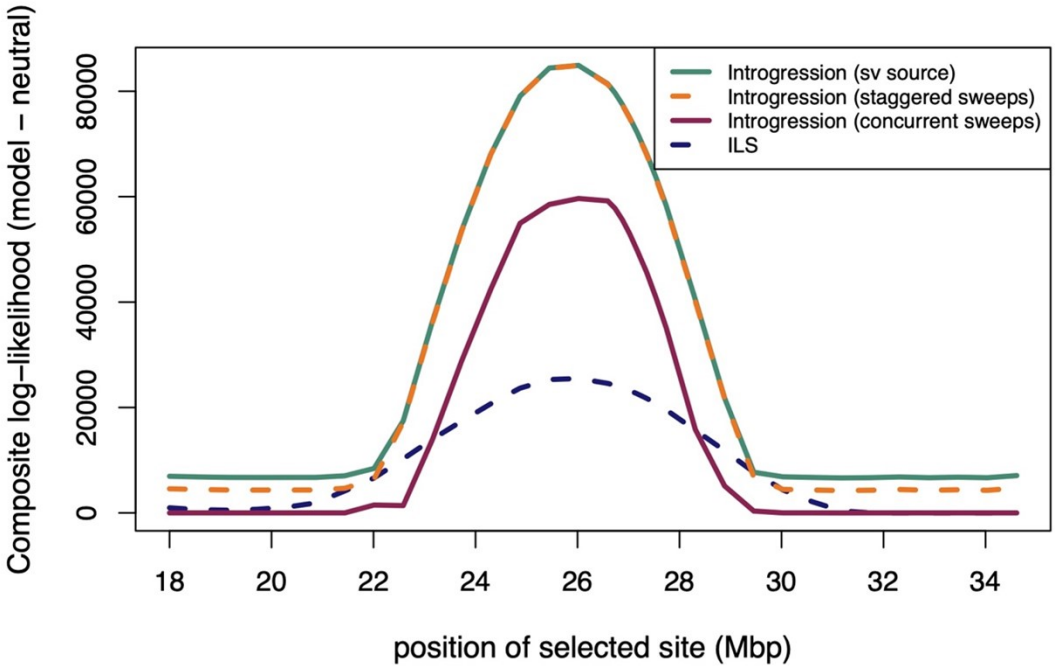

Figure 2.21: Likelihood results from ARNT region (chromosome 10). Composite log-likelihood ratio of given model relative to the neutral model of no selection as a function of the position of the selected site. All models have their maximum-composite-likelihood estimate (MCLE) at 26019105 bp, or 2 Mb downstream of the ARNT region (24 Mbp).

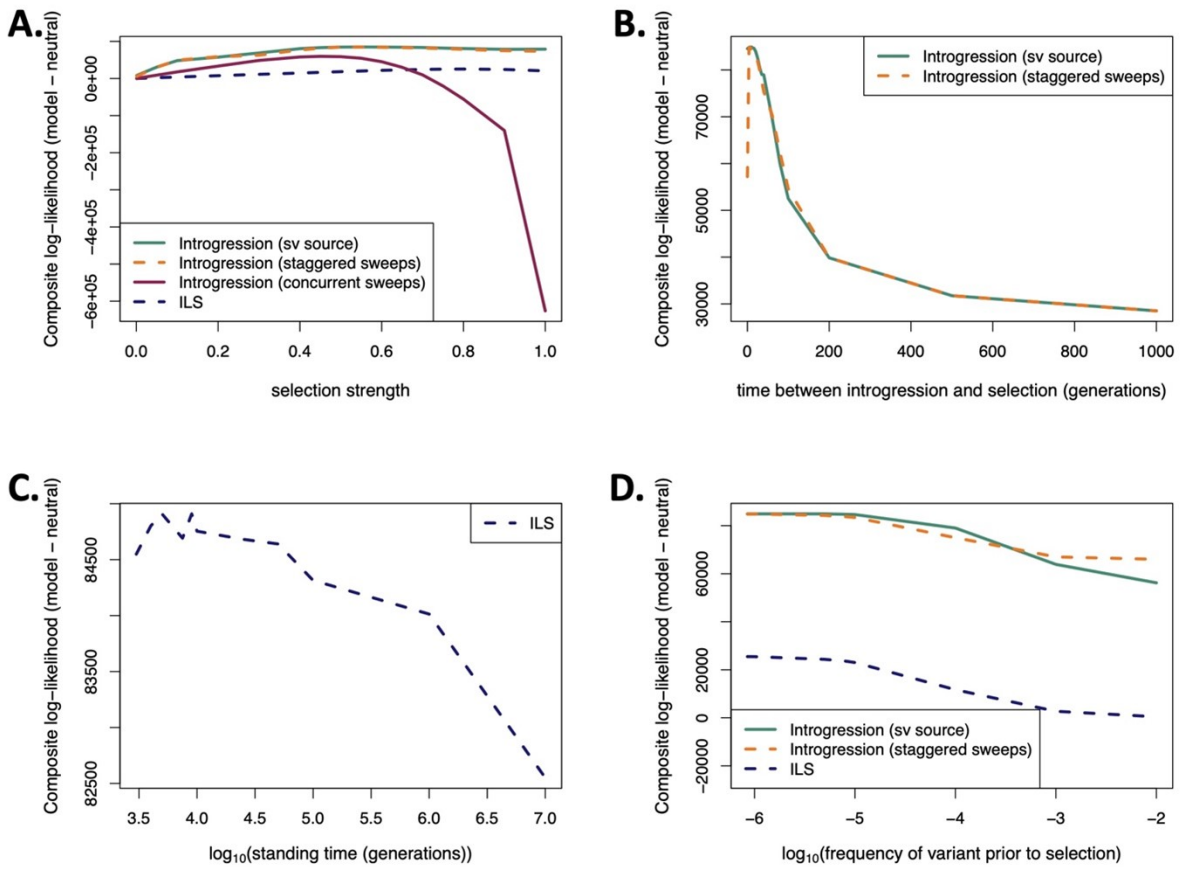

Figure 2.22: (Continued on the following page.)

Figure 2.22: Likelihood results from ARNT region (chromosome 10). Composite log-likelihood ratio (CLR) of given model relative to the neutral model of no selection as a function of various parameters. A) CLR as function of selection coefficient for all models. The maximum-composite-likelihood estimate (MCLE) for the standing variant source and staggered sweeps model is $s = 0.55$, for the incomplete lineage sorting model is $s = 0.8$, and for the concurrent sweeps model is $s = 0.45$. B) CLR as a function of time between introgression and selection for models that include this parameter. The standing variant source model has an MCLE of 10 generations and the staggered sweeps model has an MCLE of 6 generations. C) CLR as a function of the duration of the standing time in the incomplete lineage sorting model. This time represents the minimum age of the standing variant and specifically is the time (in generations) between the splitting of hT4 and gR2 and the onset of selection. The MCLE of this time is 9000 generations. D) CLR as a function of the frequency of variant prior to selection. In the top two introgression models, this represents the frequency of the allele upon introgression. The standing variant source model has an MCLE of 5×10^{-6} and the staggered sweeps model has an MCLE of 10^{-6}. For the incomplete lineage sorting model, this parameter represents the frequency at which the variant is standing. This model's MCLE is 8.5×10^{-7}.

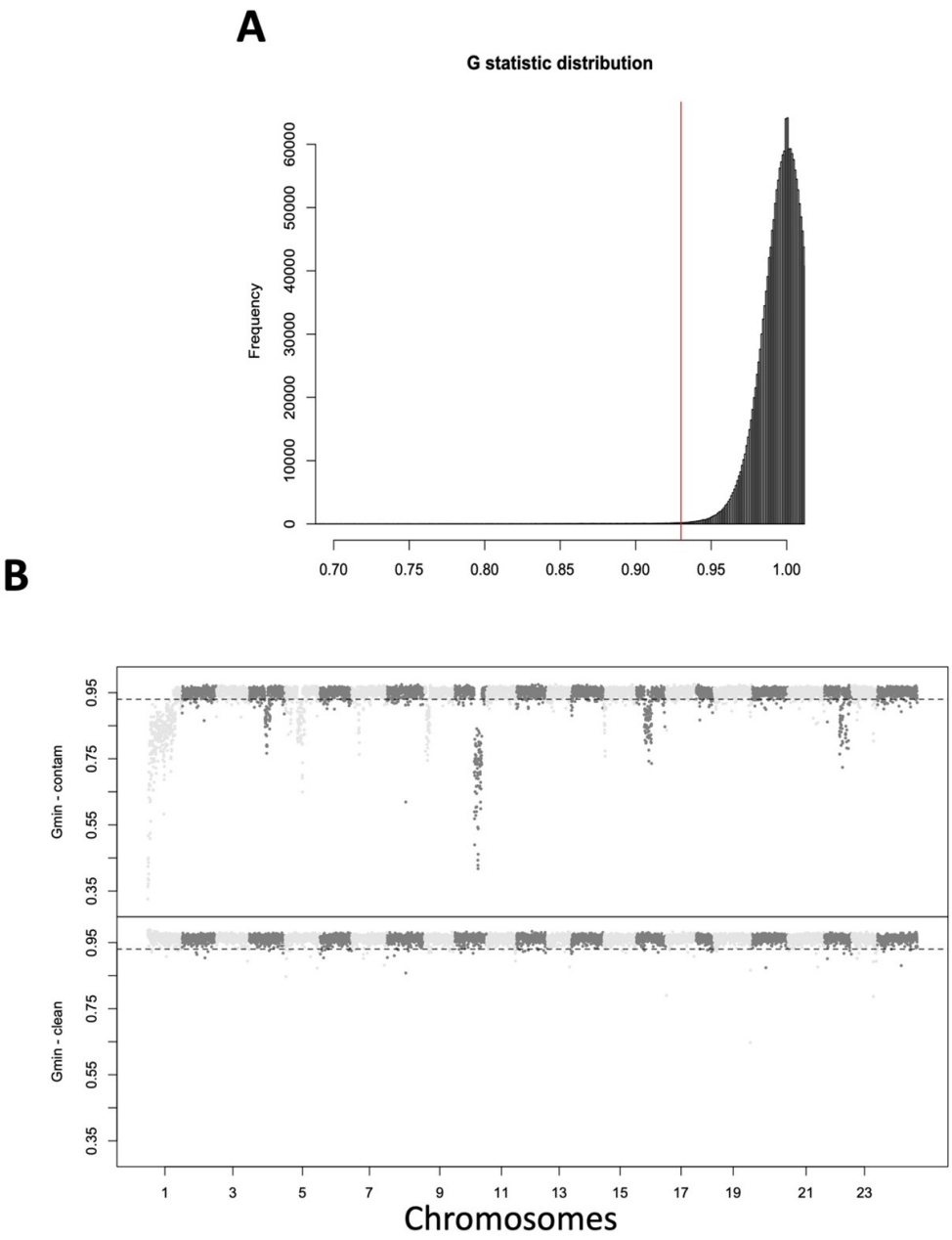

Figure 2.23: (Continued on the following page.)

Figure 2.23: Modified Gmin statistic reveals regions of introgression present among polluted sites. A) Histogram of window scores with selected outlier threshold, resulting in 0.6% of windows being labeled as introgressed. B) Genome-wide distribution of modified Gmin statistic (100kb windows) for 5 resistant populations (top panel) and 2 sensitive populations (bottom panel). Large regions of introgressed *F. heteroclitus* ancestry are restricted to resistant populations located in the Houston Ship Channel.

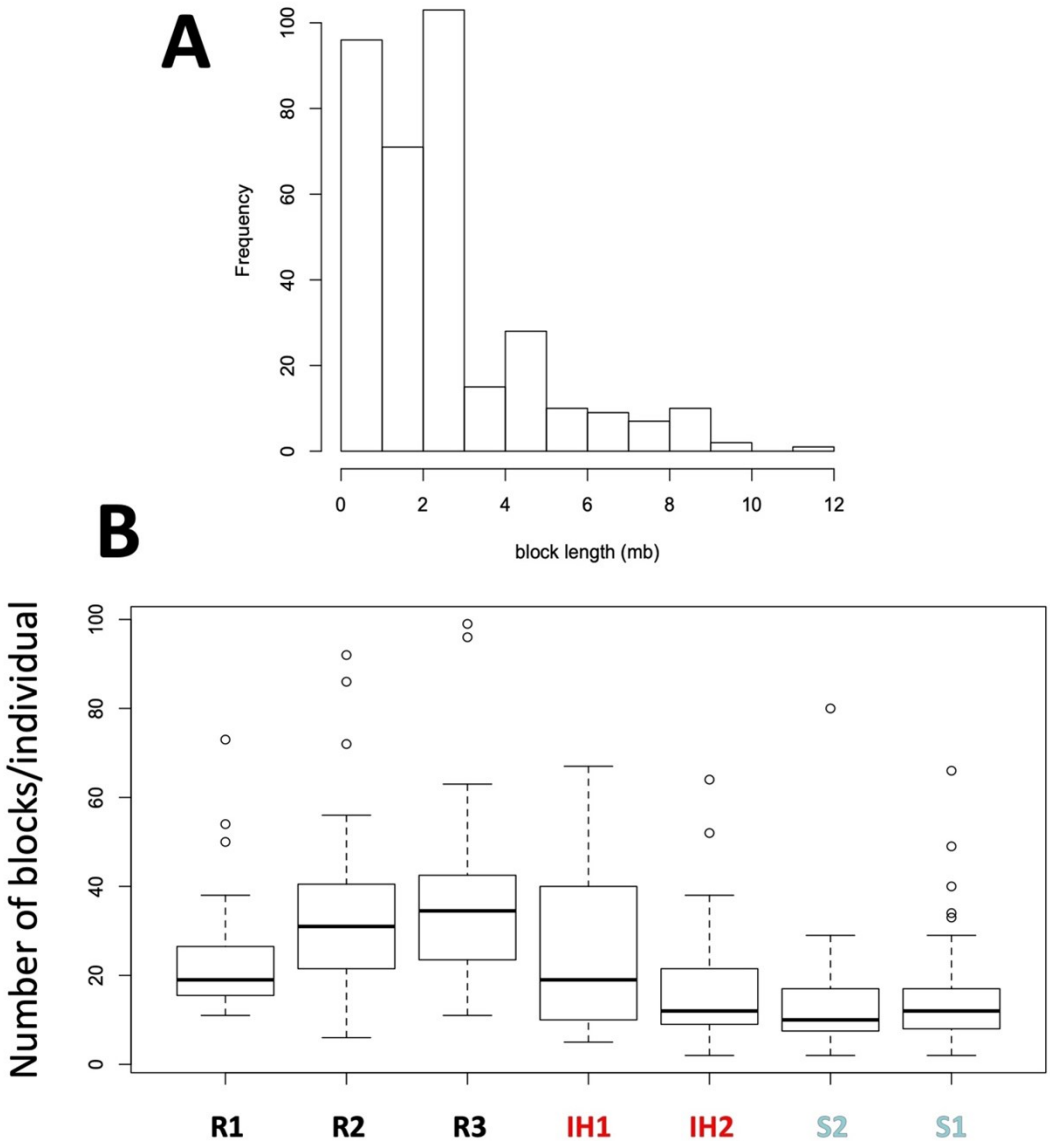

Figure 2.24: The distribution of introgression blocks. A) Length distribution of introgression blocks $\geq$ 400kb across all *F. grandis* individuals. Most introgression blocks were $<$ 3mb, with few larger blocks found in few individuals. B) Number of introgression blocks $\geq$ 100kb per individual, by population. The largest number of introgression blocks are found in R3, which suggests that R3 may closest among sampled populations to the location of the initial hybridization event between *F. grandis* and *F. heteroclitus* that introduced beneficial alleles.

Chapter 3

Population differentiation of polygenic score predictions under stabilizing selection

Sivan Yair[1,2], Graham Coop[1,2]

[1] Center for Population Biology, University of California, Davis

[2] Department of Evolution and Ecology, University of California, Davis

Abstract

Given the many small-effect loci uncovered by genome-wide association studies (GWAS), polygenic scores have become central to genomic medicine, and have found application in diverse settings including evolutionary studies of adaptation. Despite their promise, polygenic scores have been found to suffer from limited portability across human populations. This at first seems in conflict with the observation that most common genetic variation is shared among populations. We investigate one potential cause of this discrepancy: stabilizing selection on complex traits. Counter-intuitively, while stabilizing selection constrains phenotypic evolution, it accelerates the loss and fixation of alleles underlying trait variation

within populations (GWAS loci). Thus even when populations share an optimum phenotype, stabilizing selection erodes the variance contributed by their shared GWAS loci, such that predictions from GWAS in one population explain less of the phenotypic variation in another. We develop theory to quantify how stabilizing selection is expected to reduce the prediction accuracy of polygenic scores in populations not represented in GWAS samples. In addition, we find that polygenic scores can substantially overstate average genetic differences of phenotypes among populations. We emphasize stabilizing selection around a common optimum as a useful null model to connect patterns of allele frequency and polygenic score differentiation.

3.1 Introduction

Lewontin's foundational early work found that most common genetic variation is shared among human populations (Lewontin, 1972). Thus it strongly refuted the view that genetic variation is partitioned among mostly invariable populations and has become a classic work discrediting discrete human races. Lewontin's finding is reflected by low estimates of F_{ST} among populations (Li et al., 2008), i.e. only a small proportion ($\sim 10\%$) of the total allelic variance is attributable to differences in frequency between populations (a finding that has been replicated for many different marker types genome-wide; Li et al., 2008; Barbujani et al., 1997; Jorde et al., 2000; Rosenberg et al., 2002; Conrad et al., 2006; Bergström et al., 2020). While some loci that are highly differentiated among human populations have been uncovered (e.g. underlying loci in skin pigmentation and infectious disease immunity; see Fan et al., 2016, for a review), there are relatively few such strongly selected loci (Coop et al., 2009; Hernandez et al., 2011).

Lewontin's results also have implications for our *a priori* expectations of the partitioning of phenotypic variation within and among human populations. For phenotypes evolving neutrally in diploids, we expect the proportion of additive genetic variance attributable to

among-population differences to be $\approx 2F_{ST}$ when all individuals are measured in a common set of environments (Wright, 1951; Lande, 1976; Rogers and Harpending, 1983; Lande, 1992; Whitlock, 1999; Edge and Rosenberg, 2015). Thus in humans, we expect only $\sim$18% of additive genetic variance to be due to among population differences when measured in a common set of environments. Compared to this null, phenotypes subject to divergent selection among populations are expected to be over-dispersed, while phenotypes subject to stabilizing selection with the same selective optimum are expected to be less differentiated among populations. To distinguish among potential contributors to genetic differences among populations, researchers often turn to settings like common gardens in an effort to eliminate among-population environmental variation. However, it is not feasible to measure human phenotypes in a common environment (a major drawback of studies that investigated population level phenotypic variation; Relethford and Lees, 1982; Chakraborty, 1990). Thus for the majority of complex traits, we do not know the role of genetics, let alone natural selection, in explaining phenotypic differences among human populations.

These questions have received renewed interest in human genetics due to genome-wide association studies (GWAS), which have found common variation associated with many phenotypes within populations. GWAS have revealed that most phenotypes are highly polygenic within populations (Loh et al., 2015; Shi et al., 2016; Boyle et al., 2017) and confirmed that much of the genetic variance is additive (reviewed in Hill et al., 2008). These observations have motivated phenotype prediction through additive genetic values, the additive contribution of polymorphisms to phenotypic differences among individuals. One common approach to predict an individual's additive genetic value is based on a polygenic score, the sum across trait-associated loci of genotypes weighted by their estimated effects. Predictions based on polygenic scores are being explored in a number of clinical settings and more generally as a tool for understanding the genetic basis of disease and phenotypic variability. However, the generalizability of polygenic scores across populations is concerning because GWAS samples

are strongly biased towards European populations and studies of other populations are much smaller (Li and Keating, 2014; Popejoy and Fullerton, 2016; Macarthur et al., 2017; Martin et al., 2019). There is wide agreement that these portability issues must be addressed so that the future clinical use of polygenic scores does not further compound inequalities in healthcare (Li and Keating, 2014; Popejoy and Fullerton, 2016; Martin et al., 2019).

Currently, for most complex traits, polygenic scores poorly predict additive genetic values, and therefore phenotypes. This issue stands even in samples closely related to the GWAS population, because polygenic scores aggregate across many loci with slightly mis-estimated effects. These problems increase as we move to populations that are genetically and environmentally more distant to the GWAS populations. For one, the associated loci are usually not the causal loci underlying trait variation; instead they tag the effects of linked causal sites. The interpretation of the effect size of an associated variant can be tricky because of: (i) linkage disequilibrium (LD), whereby it absorbs the effects at correlated causal sites (Vilhjálmsson et al., 2015; Horikoshi et al., 2017; Lam et al., 2019; Wang et al., 2020; Weissbrod et al., 2021); (ii) population stratification, whereby it absorbs the effects of covarying environments (Haworth et al., 2019; Kerminen et al., 2019; Sakaue et al., 2020; Trochet et al., 2021; Isshiki et al., 2021); and (iii) gene-by-environment (GxE) or gene-by-gene (GxG) interactions, whereby its estimate is averaged over the interacting environmental contexts or genetic backgrounds in the sample (Horikoshi et al., 2017; Brown et al., 2016; Coram et al., 2017; Adhikari et al., 2019; Bentley et al., 2019; Galinsky et al., 2019; Grinde et al., 2019; Veturi et al., 2019; Wojcik et al., 2019; Mostafavi et al., 2020; Mathieson, 2021; Patel et al., 2021). As all of these factors can and will vary across populations, the effect sizes of alleles will differ among them and so polygenic scores will have lower prediction accuracy, i.e. imperfect portability, across populations.

Even with perfectly estimated effects at the causal loci with significant trait associations, the prediction accuracy of polygenic scores will be limited because GWAS only identify loci

with common alleles that contribute enough variance to exceed some significance threshold determined by the sample size. Therefore, an allele that is rare in the GWAS sample but common elsewhere will not be discovered. This would lead to a greater reduction in the phenotypic variance accounted for, or prediction accuracy, in populations not represented in the GWAS sample (hereafter 'unrepresented populations'; Wang et al., 2020; Bentley et al., 2019; Wojcik et al., 2019; Martin et al., 2017,?; Curtis, 2018; Kim et al., 2018; Conti et al., 2021). Indeed, many variants contributing to trait variation in European GWAS samples are not at a high enough frequency to be detected in other populations, suggesting different sets of polymorphisms contribute to the trait variance in different populations (Liu et al., 2015; Durvasula and Lohmueller, 2021). Genetic differentiation likely contributes to the reduction in the prediction accuracy of polygenic scores in unrepresented populations, as groups with increasing genetic distance from GWAS samples experience a greater loss in prediction accuracy (Scutari et al., 2016; Bitarello and Mathieson, 2020; Cavazos and Witte, 2021; Privé et al., 2022).

The factors that reduce the utility of polygenic scores for individual level prediction may also complicate the interpretation of average polygenic score differences across populations. Such issues arise in studies of adaptation that use polygenic scores to assess the contribution of selection to the genetic basis of phenotypic differentiation among human populations. An early application of this approach found polygenic signals of selection on height within Europe (Turchin et al., 2012; Berg and Coop, 2014), where polygenic scores were over-differentiated among populations compared to the neutral prediction based on F_{ST}. Importantly, the null distribution of this test of neutrality at the level of the phenotype does not rely on accurate polygenic scores and so changes in LD, GxG, and GxE should not cause false signals. However, the results are very sensitive to slight biases in estimated effect sizes due to population structure, and indeed the signal of polygenic selection on height turned out to be almost entirely due to stratification (Berg et al., 2019; Sohail et al., 2019; Refoyo-

Martínez et al., 2020). More generally, the imperfect portability of polygenic scores, and the fact that the mean environmental contribution to phenotypes can vary greatly between populations, raises concerns about over-interpreting differences in mean polygenic scores as genetic differences in the average phenotype among populations (Berg and Coop, 2014; Novembre and Barton, 2018; Coop, 2019; Rosenberg et al., 2019; Harpak and Przeworski, 2021).

The consequence of genetic differentiation on the low generalizability of GWAS results for phenotype prediction may appear to contradict Lewontin's observation of minor allele frequency differences between populations. However, allele frequency differentiation for complex traits under selection can occur at a rate faster than drift. This would lead to noisy estimates of additive genetic values, and so reduced prediction accuracy of polygenic scores, for unrepresented populations. In order to understand the impact of this turnover on portability, we need models of allele frequency differentiation that are informed by plausible forms of natural selection on complex traits.

Stabilizing selection with a constant fitness optimum is a sensible null model for the evolution of complex traits. Indeed, studies of its influence on allelic dynamics have set a conceptual foundation for interpreting and designing GWAS within populations (Simons et al., 2018). Under stabilizing selection, intermediate trait values have the highest fitness, with decreasing fitness with distance from that optimum (Figure 3.1A). Many quantitative traits have been shown to experience stabilizing selection in humans (e.g. Sanjak et al., 2018) as well as across many other species (e.g. Bumpus, 1899; Kingsolver et al., 2001; de Villemereuil et al., 2020). Stabilizing selection also contributes to the lack of variation in morphological traits within and between closely related species and the morphological constancy of traits in the fossil record (Gingerich, 1983; Weber, 1990; Barton and Keightley, 2002; Hill and Kirkpatrick, 2010; Houle et al., 2017).

Here, we show how stabilizing selection on complex traits reduces portability and in-

creases the chance of false signals of directional polygenic adaptation, despite low overall genetic differentiation and no genetically-based trait differentiation among populations. Under parameter ranges estimated from empirical studies, we combine simulations of stabilizing selection on complex traits with analytical models of its effect on genetic differentiation to investigate how stabilizing selection drives these results. We focus only on scenarios in which the trait optimum is shared among populations, leading to levels of trait differentiation among populations much smaller than neutrality. In our baseline scenario, stabilizing selection occurs on a single additive trait, where alleles have the same effect in different populations (reflecting no GxG or GxE interactions and no pleiotropy; we relax these assumptions later).

We study differentiation between a pair of populations, either an ancestral and descendant population or a pair of contemporary populations, in which the results of a GWAS in one of those populations is used to make trait predictions in the other. In doing so we provide a polygenic score perspective on earlier investigations into the relationship between population structure, stabilizing selection, and quantitative trait variation (e.g. Cohan, 1984; Narain and Chakraborty, 1987; Lande, 1991; Goldstein and Holsinger, 1992; Latta, 1998; Le Corre and Kremer, 2003). We show how these factors reduce the prediction accuracy of polygenic scores and can readily lead to patterns of polygenic score differentiation rife with the potential for misinterpretation.

3.2 Model Background

An individual's additive genetic value G_i is the sum of the additive effects of all alleles they carry,

$$G_i = \sum_l a_l g_{il}, \tag{3.1}$$

where a_l represents the additive effect of an allele relative to another at locus l and g_l is the number of copies of that allele carried by the individual. All of these loci denoted

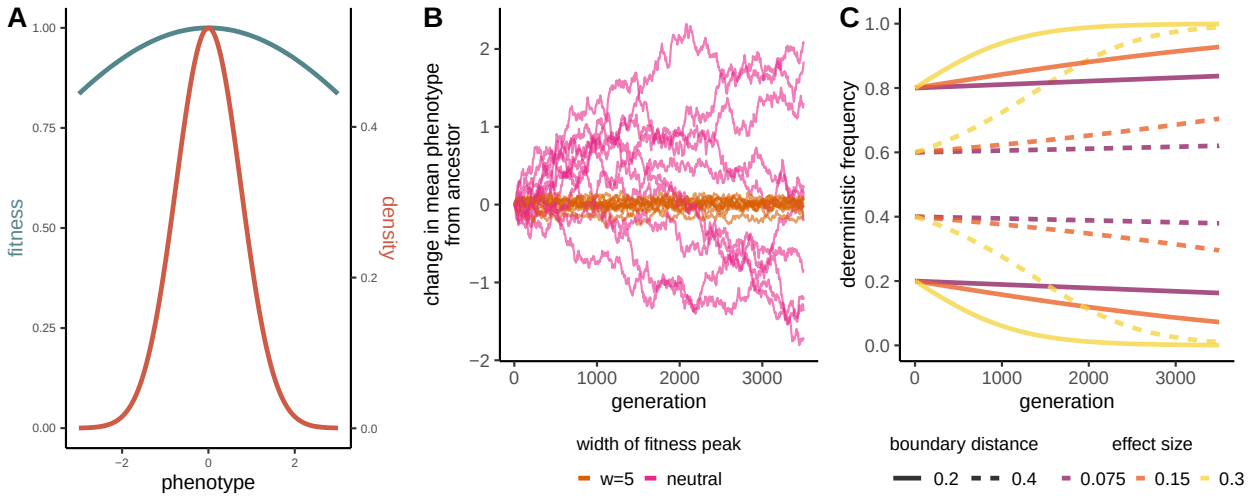

Figure 3.1: **Stabilizing selection to a constant phenotypic optimum constrains the evolution of mean phenotypes through underdominant selection at the locus-level. A)** Gaussian fitness function for phenotypic stabilizing selection (teal) and resulting density of simulated phenotypes in the population (coral). Lines correspond to the intermediate strength of stabilizing selection that we simulated ($w = 5$, where w determines the width of the fitness peak, see Section 3.2). Note how stabilizing selection keeps the genetic variance of the population (V_P) small compared to the width of the selection peak (w). **B)** Change in mean phenotype from the ancestor (generation 0) under stabilizing selection (orange) and neutral phenotypic evolution (pink). Each line corresponds to a single simulation; here 10 simulations are shown for each strength of selection. **C)** Deterministic allele frequency trajectories, assuming an infinite population size, based on the underdominant model, at a locus that contributes to the variance of a trait under stabilizing selection (here $w = 5$). The trajectory differs according to effect size and starting frequency. Note the symmetry for starting frequencies that are the same distance to their closest boundaries.

by l are polymorphic within a specified population from which the individual was drawn. Note that the additive genetic value does not represent an absolute measure of an individual's phenotype, and instead represents the additive contribution of the polymorphisms they carry to their deviation from their population's mean.

Under a constant selective environment, stabilizing selection keeps the population mean phenotype close to the optimum and decreases the phenotypic variance in the population because individuals on both tails of the distribution have lower fitness (Figure 3.1A,B; Lande, 1976; Wright, 1935; Lande and Arnold, 1983). To understand the process by which stabilizing selection reduces the phenotypic variance, we focus on the additive genetic variance,

$$V_A = \mathrm{Var}(G) = \underbrace{\sum_{l} a_l^2 \mathrm{Var}(g_l)}_{\text{genic variance}} + \underbrace{\sum_{l \neq l'} a_l a_{l'} \mathrm{Cov}(g_l, g_{l'})}_{\text{LD contribution}}, \tag{3.2}$$

in which the first component refers to the additive *genic* variance (V_a) and the second component accounts for the contribution of linkage disequilibrium (LD) among loci that contribute to the variance. In the short term, stabilizing selection reduces the phenotypic variance by generating negative LD between like-effect alleles, thereby limiting extreme phenotypes (known as the Bulmer effect; Bulmer, 1971). The additive genetic variance quickly reaches an equilibrium reflecting a balance between selection producing negative LD and recombination and chromosome segregation breaking up that generated LD (Turelli and Barton, 1994).

The long-term genetic response to stabilizing selection is driven by a reduction in the additive genic variance, in particular the variance in genotypes at a locus ($\mathrm{Var}(g_l)$; the expected heterozygosity). Yet if the genetic basis of trait variation was truly infinitesimal, i.e. made up of loci of infinitely small effect, only genetic drift would erode variation in the long term. However, while the loci discovered by GWAS contribute very small effects, these effects are not infinitesimally small and so they can be directly acted on by selection (Simons et al., 2018; Sella and Barton, 2019). Selection at the individual loci underlying trait

variation trait is in many cases well approximated by underdominant selection, in which the more common allele fixes and the minor allele is lost (Figure 3.2A; Wright, 1935; Robertson, 1956). Due to underdominant selection at the locus level, stabilizing selection removes polymorphisms at a faster rate than neutrality with selection coefficients proportional to their squared effect sizes (see Simons et al., 2018; Sella and Barton, 2019; Hayward and Sella, 2021, for recent applications of such models to understand GWAS variation within populations). Meanwhile, under moderate strengths of stabilizing selection and a constant environment, the population mean phenotype stays very close to the optimum through rapid, small fluctuations at many individual loci.

We note that the full analysis of models of stabilizing selection is challenging because changes in allele frequencies and covariances must be tracked over many loci (e.g. Turelli and Barton, 1994, 1990). Indeed, when we simulated physically linked loci with low recombination, we found that selection on an allele was weaker than we predicted due to the persistence of selection-generated LD between alleles with opposing effects (the Bulmer effect; 3.6). To understand the long-term turnover in the genetic basis of trait variation, from here on we focus on the additive genic variance, V_a, under the assumption that loci contributing to the variance are physically unlinked.

We begin by thinking of an ancestral population at equilibrium and the loss of ancestral phenotypic variance over time. When the population is at mutation-drift-selection equilibrium, stabilizing selection and genetic drift remove variation such that as time goes on less and less of the variance in the descendant population is contributed by ancestral polymorphisms. Instead, the variance contributed by new mutations that are private to the descendant population will increase, replenishing what was removed by drift and selection.

We assume a model of Gaussian stabilizing selection where the width of the fitness peak is determined by w (w^2 is equivalent to V_S in other models of stabilizing selection, e.g. Turelli, 1984), which will approximate any symmetric quadratic stabilizing selection model when the

mean trait is close to its optimum. We also assume thousands of unlinked loci contributing to trait variation and explore dynamics resulting from three different mutation effect size distributions (Figure 3.2C); see Appendix 3.8.1 for simulation details.

We can predict the reduction in additive genic variance contributed by ancestral variants (anc) in the descendant population (desc) after t generations for alleles with effect size a using a common approximation for the per generation loss (e.g. Keightley and Hill, 1988),

$$\frac{V_{a[\text{anc, desc}]}(a,t)}{V_a(a,0)} = (1 - 1/2N)^t \times (1 - a^2/4(w^2+V_P))^t \tag{3.3a}$$

$$\approx \exp\left(\frac{-t}{2N}\left(1 + \frac{2Na^2}{4w^2}\right)\right) \tag{3.3b}$$

$$\approx \exp\left(-F_{ST}\left(1 + \frac{S}{4}\right)\right), \tag{3.3c}$$

where V_P is the total population trait variance and $V_P \ll w^2$, $S = 2Na^2/w^2$ is the population-scaled selection coefficient of the allele, and $F_{ST} \approx t/2N$ for neutral polymorphisms. On the right-hand side of Equation 3.3a, the first bracketed term is the per generation reduction due to drift and the second term is the reduction due to stabilizing selection. Looking at the exponent in Equation 3.3c, we see that the decay of the variance (heterozygosity) contributed by alleles depends on F_{ST}, but will be increased for alleles whose effect sizes are large enough such that their population scaled selection is appreciable ($S > 1$). The above equation offers good intuition, however in the remainder of the main text we show results from a diffusion approximation that we developed that is better for our purposes (extending from Simons et al., 2018, see Appendix 3.8.3).

The reduction in additive genic variance contributed by a particular polymorphism with effect size a is the same as the reduction in heterozygosity at that site. Selection causes a stronger reduction when the fitness peak is narrower (i.e. when w is smaller) and when the allele's effect is larger (Figure 3.2A; see 3.7 for the approximation in Equation 3.3a). We

can average the reduction in heterozygosity across all sites, weighting by the distribution of effect sizes and genic variance contributed by a given effect size, to predict the total remaining variance and thus total reduction in V_a over time (Figure 3.2B; Equation 3.22). In the example shown in Figure 3.2, this total reduction tends to be weaker than what we see for the largest effect polymorphisms, because most sites that contribute to the variance are of small effect. This is because (i) under our mutational distribution most alleles have small effects, and (ii) under stabilizing selection, the equilibrium distribution of observed effects is more narrow than the distribution of mutation effects.

The form of the decay in ancestral variance strongly depends on the distribution of mutation effects. Consider a case in which a higher proportion of introduced mutations are strongly selected due to their large effects ($S > 10$). Some of these alleles are still capable of drifting to intermediate frequencies, and so a higher proportion of the ancestral variance will be contributed by larger effect polymorphisms (Simons et al., 2018). Therefore stabilizing selection will cause a steeper reduction in the ancestral variance (Figure 3.2C). Moreover, if most of the other mutations are nearly neutral, the early and steep decline in ancestral variance will be followed by a decline more consistent with neutrality (yellow lines in Figure 3.2C; see Simons et al., 2018, for a discussion of selection regimes). The extent to which ancestral variance is depleted determines the amount of shared additive genetic variance between diverging populations and thus the portability of polygenic scores.

3.3 Accuracy of polygenic score predictions

To understand the effect of stabilizing selection and drift on the prediction accuracy of polygenic scores in isolation from other sources of bias, we make the simplifying assumption that GWAS identify associations between polymorphisms and trait variation only at causal loci. For a GWAS within a population to identify a locus as being associated with the trait, the locus has to be polymorphic in that population and its phenotypic association has to

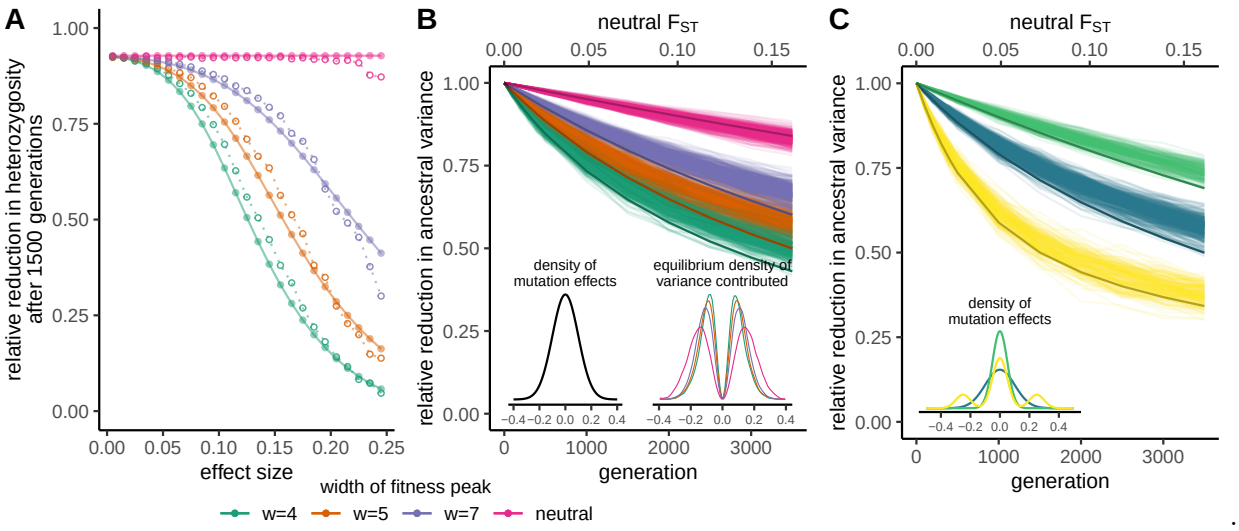

Figure 3.2: **The effect of stabilizing selection, distribution of effect sizes, and GWAS ascertainment scheme on the portability of a GWAS to another population.** **A)** Reduction in heterozygosity at loci that contributed to the variance 1500 generations ago in a population with $N_e = 10,000$. Each open point represents the midpoint of the effect size bin of width 0.01, within which we averaged heterozygosity from 200 simulations. Each filled point represents our analytical predictions for that midpoint. **B)** Reduction over time in the total variance contributed by polymorphisms in an ancestral population. Lighter lines show results from 200 simulations and darker lines show analytical predictions. The axis at the top shows the expected neutral F_{ST} between the ancestral and descendant population for the time scales of divergence shown on the bottom axis. The inset on the left shows the distribution from which mutation effect sizes are drawn. For the results shown we used a normal distribution with standard deviation of 0.1. The inset on the right shows the equilibrium density of variance contributed from each effect size. As the width of the fitness peak increases, and thus stabilizing selection weakens, mutations with large effects can drift to higher frequencies and contribute a greater proportion of the trait's variance. **C)** The reduction over time in the total variance contributed by polymorphisms in an ancestral population for three different mutation effect size distributions, holding the strength of phenotypic stabilizing selection constant ($w = 5$). The inset shows the three mutation effect size distributions: two normal (green and blue) and a mixture of normals to produce a heavy-tailed distribution (yellow). The blue lines correspond to the same distribution used to produce sub-figure B, thus replicating the orange lines in sub-figure B. Lighter lines show results from simulations (100 each for yellow and green) and darker lines show analytical predictions. Predictions in **B** and **C** are based on the diffusion approximation, see Equations 3.22 and 3.23.

achieve some level of statistical significance (i.e. contribute above some level of variance to the trait). For those causal loci with significant associations, we also assume their effects are estimated perfectly and for the moment that these true effects do not vary within the sample, population the sample was drawn from, or across populations (i.e. populations experience the same set of environments). We also assume their effects are strictly additive. At the trait-associated loci, one can sum the additive effects of all alleles an individual carries at a predefined set of markers to form a polygenic score.

We are interested in the reduction in prediction accuracy for a population not represented in the GWAS sample, relative to the prediction accuracy of those represented populations. To explore this we consider the genetic differentiation between a pair of populations, A and B, in which the GWAS sample is drawn from population A but not population B. When the effects of alleles do not vary between populations, we can quantify the reduction in phenotype prediction accuracy (r^2) of polygenic scores, compared to using additive genetic values, constructed for any population as

$$\frac{r_S^2}{r_G^2} = \frac{\mathrm{Var}(S)}{\mathrm{Var}(G)}, \tag{3.4}$$

where S_i is an individual's polygenic score and G_i is their additive genetic value. Since an individual's polygenic score is only part of their additive genetic value when a GWAS estimates true effects, this reduction can be understood as the proportion of the total additive genetic variance, or proportion of the heritability, explained by GWAS-significant sites (Appendix 3.8.2). The ratio of r_S^2/r_G^2 in population B to population A quantifies the reduction in prediction accuracy due to a lack of GWAS representation for population B. If these populations experience the same selective environment, we expect the same $\mathrm{Var}(G)$ for each population, so this reduction due to a lack of representation would simply be the ratio of the variances explained by polygenic scores.

Note that our definition of the prediction accuracy in population B is the squared cor-

relation of the deviation of an individual's polygenic score with its deviation in phenotype, where both of these deviations are with respect to the mean phenotype in population B. This definition matches typical polygenic score practices where predictions are statements about the departure of an individual from their ancestry group's mean genetic value, rather than a prediction of their departure from the mean genetic value of the GWAS population. However, this definition of prediction accuracy does not include discrepancies from the evolution of mean polygenic scores and phenotypes among populations, meaning it does not account for a systematic shift in polygenic scores between populations. We turn to this point in Section 3.4.

3.3.1 Ascertainment of all causal loci in GWAS sample

We begin with the simplified case in which all causal loci that are polymorphic in population A have been identified, such that polygenic scores equal the additive genetic values in population A. As polygenic scores from ancient DNA are being used to investigate the phenotypic diversity of ancient human populations (Irving-Pease et al., 2021), we first consider the prediction accuracy in a population ancestral to population A of a polygenic score constructed using variants found in population A. The reduction in polygenic score prediction accuracy for this ancestral population is approximately the reduction in variance contributed by that ancestral population to the present (Figure 3.2B,C; 3.9; see Appendix 3.8.3.2 for an explanation). While ancient individuals were likely not drawn from populations directly ancestral to present-day populations, we should observe the same general patterns with genetic differentiation between the ancient and present-day population (see also Carlson et al., 2021, for approximations of this decline).

For the remainder of this article we consider A and B to be contemporary populations; we assume for simplicity that they split from a common ancestral population without subsequent gene flow. In Figure 3.3A, we use simulations and analytical predictions to show how

the prediction accuracy in population B decreases with increasing time since its common ancestor with population A. The span of neutral genetic differentiation was chosen to reflect a scale along which various human populations could fall. We see slightly weaker reductions in prediction accuracy with time when the variance of the mutation effect size distribution is quartered and stronger reductions when the mutation effect size distribution has a heavy tail (Figure 3.3B; 3.10). All variance-contributing loci were ascertained in population A, so the only reason the full genic variance in population B was not captured is because private polymorphisms contribute to the phenotypic variance in each population. At the time of divergence between the pair of populations, they entirely share their genetic basis of trait variation. Then as stabilizing selection and drift remove polymorphisms (at equilibrium), new mutations replenish them at different sites in each population, leading to the same total variance but different genetic bases of that variance (assuming a very large mutational target). Without gene flow between the pair of populations, the polymorphisms that arose since their common ancestor will remain private to each population and will thus not contribute to the variance in the other population. Therefore the polymorphisms in population A that also contribute to the variance in population B must be ancestrally shared polymorphisms, specifically at those loci in which ancestral polymorphisms were not removed by drift or selection in either descendant population. Note that these shared polymorphisms on average contribute less variance in the descendant populations than in their ancestor. Thus we see a loss in the additive genetic variance shared between populations over time due to a reduction in both the number of shared ancestral polymorphisms and the variance they each contribute. In Appendix 3.8.3.2 we describe our analytical predictions for this process, which match well to simulations.

3.3.2 Ascertainment of subset of causal loci in GWAS sample

To more realistically explore the impact of GWAS ascertainment, we explore power- and frequency-based variant discovery. In the main text, we focus on a case in which a GWAS uncovered the top 5% of variance-contributing polymorphisms in population A. Under our choice of effect size distribution and strengths of stabilizing selection, these polymorphisms explain just under 50% of the additive genic variance in population A (similar to that of height in Europeans; Yang et al., 2010; Wood et al., 2014; Yang et al., 2015). We also consider a case in which a GWAS uncovered all causal loci with a minor allele frequency (MAF) that exceeds 1%. As expected, under any ascertainment scheme we observe a substantial drop in the variance explained in population B compared to the case in which all polymorphisms in population A are ascertained (Figure 3.3B). If this reduction in variance for each population is the same, then the decline in prediction accuracy in population B *relative* to population A over time will be similar to the case when all polymorphisms in population A are ascertained. We find that this is approximately the case for a Gaussian mutation effect size distribution but not for the heavy-tail effect size distribution that we simulate (Figure 3.3C). These differences arise from differences in the strength of selection on and variance contributed by ancestrally shared polymorphisms; see Supplement 3.9.2 for details. We present results under GxE, pleiotropy, and directional selection in Supplement 3.9.3.

3.4 Difference in polygenic means among populations

In the previous section we described how stabilizing selection and ascertainment in the GWAS sample can reduce the prediction accuracy of individual genetic values, but what are the consequences for the mean polygenic score of populations? The mean polygenic score of the population is twice the sum of population allele frequencies weighted by effect sizes. If a trait is neutrally evolving, the loci contributing to its variation are just like other neutrally evolving loci, and so differences among populations in their mean polygenic score

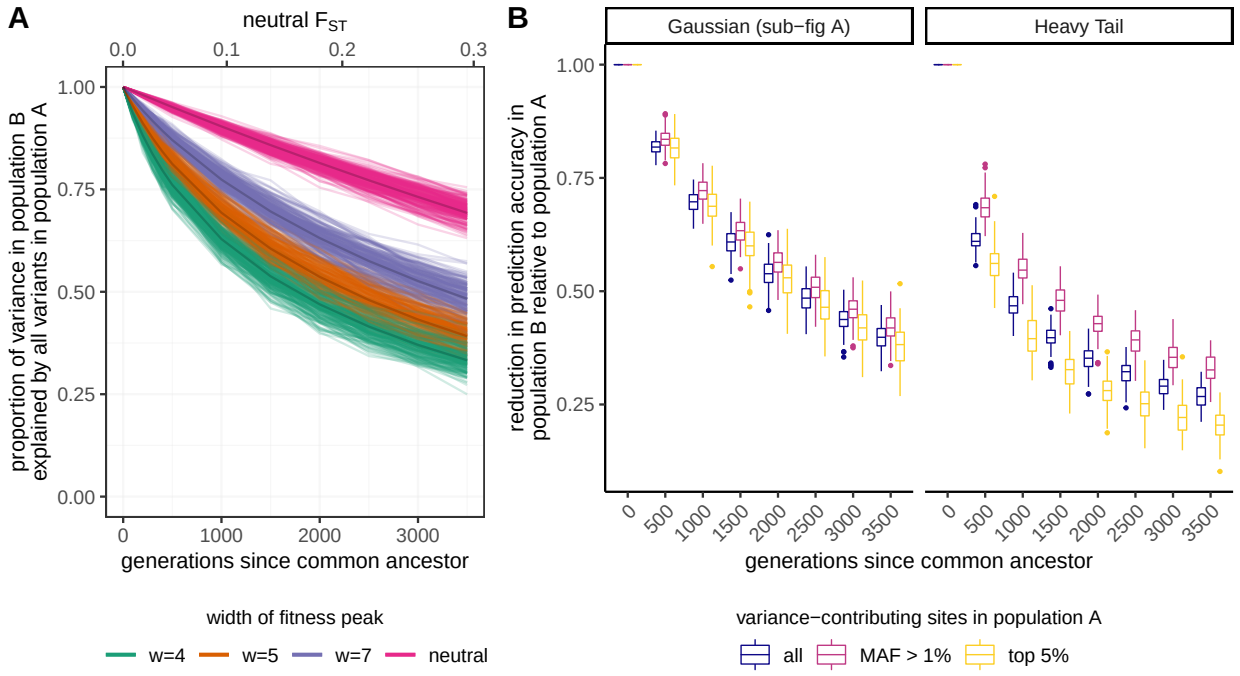

Figure 3.3: **Reduction of polygenic score prediction accuracy in unrepresented population B relative to contemporary, represented population A increases with the strength of stabilizing selection, density of large-effect mutations, and power-based GWAS discovery schemes. A)** Reduction of prediction accuracy in population B, when all variance-contributing polymorphisms in population A were ascertained, with increasing time since the common ancestor of populations A and B. Lighter lines show results from simulations and darker lines show results from analytical predictions. **B)** Distribution of the reduction in prediction accuracy for population B relative to population A from simulations with $w = 5$ when different sets of variance-contributing sites in population A were ascertained. Results are shown for two mutation effect size distributions. The Gaussian distribution corresponds to the blue distribution and the Heavy Tail to the yellow distribution in the inset of Figure 3.1C. The same Gaussian distribution was used to produce results in sub-figure A.

just reflect a weighted sum of neutral allele frequency differences. Naively, as trait increasing alleles underlying a neutral trait are equally likely to drift up or down, one might think that over many loci we expect only a small mean difference between populations. However, the polygenic score is a sum rather than a mean, and so each locus we add into the score is like an additional step in the random walk that two populations take away from each other (Chakraborty and Nei, 1982). We expect the variance among populations, i.e. the average squared difference between population means and the global mean, to be $2V_A F_{ST}$ (Wright, 1951; Lande, 1992). We first explore the differentiation of mean polygenic scores under a constant optimum and constant environment, and then relax these assumptions.

3.4.1 Stabilizing selection to a constant optimum

Under a constant selective environment, stabilizing selection keeps population mean phenotypes close to their optimum in the face of genetic drift and mutation, such that the difference in mean phenotypes among populations with the same optimum should be minor relative to neutral expectations (even accounting for the lower genetic variance within populations under stabilizing selection; Figure 3.4A). This reduction in the divergence of the mean phenotype between populations reflects the fact that if trait-increasing alleles accidentally drift up in frequency, thus pushing the population mean away from its selective optimum, trait-decreasing alleles are subject to directional selection in their favor (and vice versa).

To explore population differences in mean polygenic scores, we can use the polygenic score equivalent of $Q_{ST}/2F_{ST}$ (Prout and Barker, 1993; Spitze, 1993), where Q_{ST} measures the proportion of total variance in additive genetic values attributable to among-population differences. Under neutrality, for strictly additive traits we expect Q_{ST} to equal $2F_{ST}$ estimated from neutral polymorphisms (Lande, 1992; Whitlock, 1999), and their ratio should be χ^2-distributed with degrees of freedom equal to one fewer than the number of populations (an extension of the Lewontin-Krakauer test; Lewontin and Krakauer, 1973). The polygenic

score analog is Q_X (Berg and Coop, 2014), where additive genetic values are substituted by polygenic scores. For our pair of populations, our Q_X statistic is

$$Q_X = \frac{(Z_A - Z_B)^2}{4V_a F_{ST}}, \tag{3.5}$$

where Z is the mean polygenic score in the population denoted in the subscript. When population mean phenotypes are over-dispersed relative to neutral expectations, Q_X will be larger than 1, potentially resulting in a statistically significant p-value under the null distribution. When population means are under-dispersed relative to neutral expectations, as we would expect for traits under stabilizing selection with the same optimum across populations, Q_X will be much smaller than 1, and the p-values under the null will be large.

Stabilizing selection to the same optimum tightly constrains the difference in mean additive genetic values between populations, i.e. the difference in mean polygenic scores using all of the variation, to be much lower than the difference for neutral traits even after standardizing for the lower overall levels of variation (Figure 3.4A). This leads to a distribution of Q_X that is skewed toward lower values than the neutral χ^2-distribution (3.17). However, when we ascertain the top 5% of variance-contributing polymorphisms in population A, the level of standardized polygenic score differentiation under stabilizing selection becomes more similar to neutral levels (Figure 3.4B) with a comparable level of false positive signals of adaptive differentiation as in the neutral case (3.17). While this result is more specific to our choice of mutation effect size distribution and strengths of stabilizing selection, we can generally conclude that for cases of stabilizing selection with limited ascertainment, estimates of standardized mean polygenic score differences will approach and perhaps exceed what we observe under neutrality. Note that under the neutral case, the distribution of standardized mean polygenic score differences stays about the same between ascertainment levels, whether ascertaining all polymorphisms from both populations or just the top 5% in population A. This is because under neutrality, all alleles exhibit the same behavior and do not evolve in

coordination with one another. Thus stabilizing selection, combined with incomplete and asymmetric ascertainment, causes the inflation of estimated mean standardized population differences above that seen for the underlying genetic values.

Since under stabilizing selection to the same optimum we expect only small differences between populations in their mean genetic value, the disparity between the (true) differences in mean genetic values and (estimated) differences in mean polygenic scores represents the mean difference contributed by non-ascertained polymorphisms. Intuitively this occurs because if population B has a larger value of an ascertained polygenic score than population A, then the ascertained trait-increasing alleles have by chance drifted up in population B (compared to A) and this imbalance will have induced directional selection for the rest of the trait-increasing alleles to decrease their frequency to keep the population close to the optimum (under high polygenicity). In line with our expectations, we find that the mean polygenic score differences calculated from ascertained sites and mean polygenic score differences calculated from non-ascertained sites are close to opposite one another (Figure 3.4B). The countervailing effect of the non-ascertained loci is noisier when stabilizing selection is weaker, because with weaker selection population means can drift further from their optimum. These results confirm that mean polygenic scores calculated from all polymorphisms (i.e. additive genetic values) should closely match between populations experiencing stronger and similar selection pressures and highlight how incomplete ascertainment can by chance generate misleading differences between them.

3.4.2 Adapting to a changing optimum or environment

The combination of changes in the stabilizing selection regime and incomplete ascertainment of causal polymorphisms can also generate misleading differences in mean polygenic scores and signals of differential selection among populations. When the optimal phenotype shifts from its ancestral value equally in both descendant populations, this directional selection

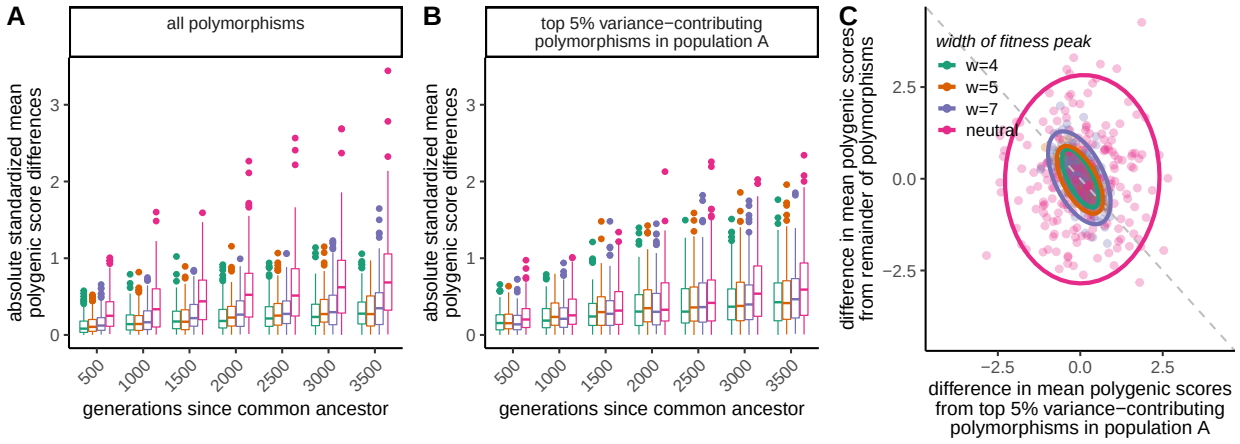

Figure 3.4: **Polygenic scores can overestimate differences in additive genetic values between populations for traits under stabilizing selection. A)** Absolute standardized mean polygenic score differences between populations A and B, $\frac{|Z_A - Z_B|}{\sqrt{V_a}}$, either when all polymorphisms in both populations were ascertained, or when the top 5% of variance-contributing polymorphisms in population A were ascertained. This measure has the same interpretation as Q_X; it equals $\sqrt{4Q_X F_{ST}}$. **B)** Partitioning of mean polygenic score differences between the ascertained and non-ascertained set of polymorphisms. Ascertained polymorphisms are from the top 5% of variance-contributing sites in population A. Points represent results for a single simulation. Ellipses denote the 95% confidence interval. The dashed grey line is the line of exactly opposing effects.

generates responses in the true mean genetic values in each population that track each other extremely closely. However, because our polygenic score based on ascertained SNPs explains less of the variance in population B than population A, we capture less of its response to selection and thus artificially generate a difference between populations in their mean polygenic scores (Figure 3.5A). This shift in the mean polygenic score between populations, as well as a decrease in the total variance explained, can push the distribution of Q_X towards larger values than the neutral case (Figure 3.5B). The chances of getting a p-value below a significance threshold of 0.05 are highest when stabilizing selection is strongest ($w = 4$; up to 30% chance depending on time since the optimum shift) but tend to be greater than neutrality for the strengths of selection we investigated (3.11). These signals of polygenic

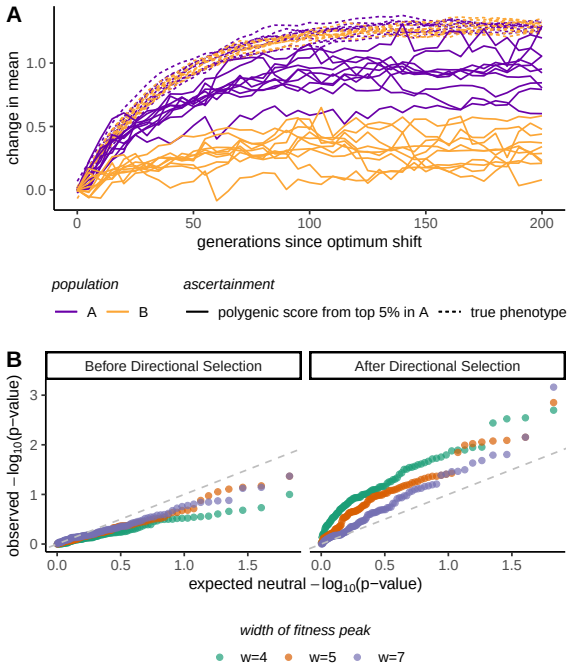

Figure 3.5: **Polygenic scores can generate false signals of adaptive differentiation when populations adapt to the same optimum shift. A)** Change in mean phenotype or polygenic score over time since the optimum shifted 2 standard deviations of the phenotype distribution. True phenotypes are different from additive genetic values in that they account for the contribution of substitutions. **B)** Quantile-quantile plot of observed p-values against expected p-values under neutrality (uniform distribution). The dashed line shows equality; points that lie above this line indicate that the observed distribution has a higher density of low p-values than the neutral distribution, and points that lie below it indicate the opposite.

selection are consistent with the idea that Q_X informs us about directional selection on polygenic scores, as selection has driven an increase in the polygenic score of A relative to B. However, they also highlight the very incomplete picture that we obtain about selection and genetic differences in the phenotype.

3.5 Discussion

Our work builds on population genetics theory of stabilizing selection to investigate why polygenic scores perform poorly in populations not represented in GWAS samples ('unrepresented populations') and how they can be misleading about average genetic differences among populations. Specifically, we provide a theoretical foundation to understand how stabilizing selection on complex traits impacts the population differentiation of polygenic scores without a difference in environments or trait optima. Negative selection has been invoked as an explanation for low portability, consistent with large effect alleles discovered by GWAS being rare in the population(Wang et al., 2020; Durvasula and Lohmueller, 2021). Here we explore these predictions under a model of stabilizing selection that has the advantage of offering clear-cut connections between an allele's phenotypic and fitness effects. With this approach to modeling allele frequency dynamics, we obtained a detailed understanding of how a lack of population representation in GWAS can reduce the utility of polygenic scores in explaining phenotypic variation within and among populations. We discuss the dynamics contributing to low portability, considerations for those interested in explaining variation beyond the GWAS sample, and necessary discretion when using polygenic scores to understand group differences.

3.5.1 Dynamics contributing to low portability

While stabilizing selection around the same optimum among populations reduces phenotypic differentiation relative to neutrality, somewhat counter-intuitively it increases genetic differentiation at trait-influencing loci compared to neutral polymorphisms. Genetic differentiation increases because stabilizing selection drives the turnover of polymorphisms that contribute to trait variation within populations; thus polymorphisms that are ancestrally shared between populations are eventually lost and replaced by private mutations. The consequences of this differentiation at large effect QTLs and few loci have been explored by

various authors (Latta, 1998; Le Corre and Kremer, 2003; Kremer and Le Corre, 2012), and here we investigated its implications on GWAS results and downstream analyses. Over time, the sets of polymorphisms that contribute substantial variance in each population diverge, such that stabilizing selection reduces the additive genic variance in one population that can be explained by polymorphisms in another. Thus polygenic scores constructed from polymorphisms ascertained in a GWAS will have reduced prediction accuracy in unrepresented populations compared to represented ones. The prediction accuracy for unrepresented populations decreases with increasing strengths of stabilizing selection and time since the common ancestor with represented populations. Occasional fluctuations in the fitness optima would negligibly influence our portability results, because such directional selection would cause very minor shifts in frequency over the short time scales we consider (Hayward and Sella, 2021).

The distribution of mutation effect sizes is critical to understanding the effects of stabilizing selection on the turnover of genetic variation among populations. With increasing weight towards larger mutation effects, stabilizing selection leads to a faster reduction in the shared additive genetic variance, and thus polygenic score portability. In addition, the decline in prediction accuracy for unrepresented populations relative to represented ones will depend on frequency- and power-based ascertainment schemes; since polymorphisms shared between populations will more likely be of small effect and common, the discovery of polymorphisms based on minor allele frequency will lead to a weaker decline than discovery based on variance contributed. Conversely, if the trait was truly infinitesimal, stabilizing selection would not contribute to the loss of heterozygosity, such that the sharing of genetic variance would be well predicted by genetic drift, or neutral F_{ST}. While the loci mapped by GWAS are often of small effect, and the traits highly polygenic, we know that their effects are not infinitesimal, as loci must make a reasonable contribution to the variance to be discovered by a GWAS (see Simons et al., 2018, for a detailed population genetic model).

While we do not consider migration between represented and unrepresented populations here, the general decline in portability with increasing genetic differentiation between groups should hold, though the exact prediction will differ. Pleiotropy and GxE complicate the predictions of stabilizing selection for portability, which we discuss in Supplement 3.9.3. Differences in the environment among populations can lead to changes in the effect sizes of alleles (GxE), which reduces the portability because (i) effect sizes will be only partially correlated between populations and (ii) stabilizing selection will purge more of the ancestral variation shared between populations since GxE increases the overall trait variance when the interacting environment changes. Pleiotropy, in our implementation, leads to the opposite effect. Holding the average strength of selection on all alleles constant, when an allele can independently affect multiple traits under stabilizing selection, the correlation between its effect size on the trait of interest and its selection coefficient weakens. Thus stabilizing selection purges less of the shared variation and causes a weaker reduction in portability.

3.5.2 Considerations for applications of GWAS results across groups

Like much of the recent work on portability, our work emphasizes the reduction in prediction accuracy for populations not represented in GWAS (Lam et al., 2019; Weissbrod et al., 2021; Martin et al., 2017; Márquez-Luna, 2017; Duncan et al., 2019). A natural conclusion is that polygenic predictions that work well across populations will require GWAS across a range of diverse ancestries, in line with other calls to reduce Euro-centrism in GWAS (Li and Keating, 2014; Popejoy and Fullerton, 2016; Martin et al., 2019). In addition to changes in allele frequency and linkage disequilibrium, the relationship of polygenic scores to phenotypes will vary across populations due to variation in genetic effects, assortative mating, and differences in GxG and GxE. Indeed the prediction accuracy of polygenic scores for some traits was recently shown to be quite variable across different groups within an ancestry, suggesting that

GxE and environmental variation are quite prevalent for some traits (Mostafavi et al., 2020). In addition, associated variants on the same haplotype were found to have differing effects in European-Americans and admixed African-Americans, suggesting genetic or environmental interactions modify additive effect sizes across groups (Patel et al., 2021). Thus, we caution that a better understanding of the portability of polygenic scores across populations also requires a stronger understanding of the causes of variation in prediction accuracy within populations.

With the rise of ancient DNA sequencing, GWAS from contemporary populations have also been used to construct polygenic scores for ancient individuals (reviewed in Irving-Pease et al., 2021). These scores have been used to provide a window into the phenotypic diversity of past populations (Mathieson et al., 2015; Berens et al., 2017; Martiniano et al., 2017) and to disentangle genetic and environmental contributors to temporal phenotypic variation (e.g. at the Neolithic transition; Cox et al., 2019, 2022; Marciniak et al., 2021). Such studies are most convincing when there are relevant phenotypic measurements on at least some ancient individuals and polygenic prediction accuracies can be judged. However, investigators will often not have this luxury, leaving unclear the insight these approaches can provide. Some studies using ancient DNA have identified reduced rates of disease alleles in the past compared to present-day populations. While we have focused here on quantitative traits, rather than disease traits, we caution that purifying selection against risk alleles will lead modern day populations to systematically underrepresent the diversity of disease alleles in the past (Berens et al., 2017; Aris-Brosou, 2019; Esteller-Cucala et al., 2020; Simonti and Lachance, 2021).

3.5.3 Misinterpretations of group differences based on polygenic scores

When using polygenic scores constructed from the ascertained set of polymorphisms, we increase the possibility of generating misleading signals of differentiation between populations. Stabilizing selection to a constant optimum alone does not generate more false signals of directional selection than what we expect under the neutral evolution of complex traits. However, the standardized difference between populations will be systematically over-estimated with ascertained polygenic scores. We see this result because with stabilizing selection, the unascertained portion of the variance tends to act exactly counter to the trend seen in the ascertained portion. Misleading signals of adaptive polygenic differentiation can also be generated when the stabilizing selection regime shifts in the same way in each population (such that there is still minor phenotypic differentiation between them) as the ascertained polymorphisms only capture a shift in the ascertainment population. This issue arises because stabilizing selection lowers the proportion of the additive genic variance explained in the unrepresented population and so we capture a lower proportion of that population's response to directional selection. Thus this issue of missing the parallel adaptive response across ancestries can be expected in many situations with imperfect portability. Such signals of polygenic adaptation can be useful as Q_X is correctly detecting that directional selection has acted on the genetic variation along the branch leading to the represented population, but the signal is very open to the misinterpretation that the unrepresented population has not also responded to the same selection pressures.

Many traits have likely experienced a mixture of stabilizing selection and bursts of directional selection across human history. Even if the populations share the same phenotypic optimum, if an environmental change systematically shifts one population away from this optimum, there would be directional polygenic adaptation to move that population back towards the optimum, resulting in a difference in polygenic scores but no difference in the mean

phenotypes between populations (Harpak and Przeworski, 2021). Therefore, under current ascertainment schemes and pervasive stabilizing selection, the difference in mean polygenic scores among populations provides very unreliable information about the potential role of selection in generating phenotypic differences among populations.

A polygenic score is a prediction of how an individual's phenotype is expected to deviate away from the sample mean given their genotype at some (large) number of polymorphisms. Sometimes they are quoted as absolute values, but that is always based on an empirical phenotypic mean. The mean polygenic score in a population cannot inform us of the mean phenotype in a population, even if it was constructed from all polymorphisms and averaged across all individuals. This is because these scores are based on polymorphic genotypes alone whereas the absolute phenotype of an individual represents the end product of the entire genome and environment played out through a vast number of developmental processes. Yet it is easy to fall into the trap of believing that a difference in polygenic scores between groups is a strong statement about the difference in the mean phenotype of those groups who also differ in a myriad of environmental and cultural factors (a related set of issues are present in epidemology; Rose, 2001). Thus while the field of human genetics is increasing its power to predict phenotypic variance among individuals within groups, it remains a poor guide to the causes of phenotypic variance among groups with greater environmental and genetic differentiation (Lewontin, 1974; Feldman and Lewontin, 1975).

3.6 Data Availability

Code used to generate simulations and process output can be found at `https://github.com/SivanYair/SLiMsims_StabilizingSelection`.

3.7 Acknowledgements

We thank Molly Przeworski, Guy Sella, Michael Turelli, Vince Buffalo, Carl Veller, Sohini Ramachandran, Doc Edge, Noah Rosenberg, Yuval Simons and two anonymous reviewers for their valuable comments on earlier drafts of our manuscript. We also thank the members of the Coop lab for helpful discussions. Funding was provided by the National Institute of General Medical Sciences of the National Institutes of Health (NIH R01 GM108779 and R35 GM136290, awarded to GC). This work used the Extreme Science and Engineering Discovery Environment (XSEDE) computing resource PSC Bridges-2 through allocation TG-MCB200074 to SY and TG-BIO210159 to GC.

3.8 Appendix

3.8.1 Simulation Details

In SLiM version 3.6 (Haller and Messer, 2019), we simulated stabilizing selection on additive traits and divergence between a pair of populations with the same optimal phenotype (set at zero). Each simulation consisted of a constant population size of 10,000 diploid individuals and genomes comprising 4000 1bp quantitative trait loci and 1000 1bp neutral loci, with free recombination between loci. Mutations arose at rate 2×10^{-6} per base pair and for quantitative trait loci their effect size was randomly assigned from a normal distribution or a mixture of normal distributions. We used a Gaussian fitness function setting the fitness of an individual to

$$W(\vec{r}) = \exp\left(-\frac{r^2}{2w^2}\right), \tag{3.6}$$

where $r = ||\vec{r}||$ is the phenotypic distance from the optimum of an individual with a vector of phenotypes $\vec{r}$ and w quantifies the width of the fitness peak. The form of Gaussian selection is consistent with any form of quadratic selection if the population mean is near the optimum (see section 3.8.1.2). We simulated quantitative traits under neutrality ($w = \infty$) and under

varying widths of the fitness peak, w, in which larger values of w indicate weaker stabilizing selection. We chose strengths of stabilizing selection and an effect size distribution that correspond to observations in human populations (see section 3.8.1.1 and 3.8.1.2 for details). We calculated an individual's fitness each generation based on equation 3.6, where their phenotype was the sum of the effects of alleles they carried at both variable and fixed sites.

We burned in each simulation for 60,000 ($6N$) generations before splitting the population into two for 3500 generations. We recorded mutation effect sizes and frequencies in each population every 100 generations for the first 500 generations, and then every 500 generations. To reduce computation times, groups of 10 simulation replicates shared the first 40,000 generations of their burn-in. The remaining 20,000 generations provided more than enough time for a complete turnover in the genetic basis of trait variation, such that simulations within a group were effectively independent. We simulated 100 or 200 replicates of each parameter combination.

3.8.1.1 Choice of parameter values for strength of selection on the trait

We simulated under four different strengths of stabilizing selection (w), chosen to range from no selection (neutral trait) to the highest strength of stabilizing selection estimated from the UK BioBank (Sanjak et al., 2018). These estimates were determined from regressions of relative fitness (based on estimates of lifetime reproductive success) on squared standard normal phenotypes, r^2, where half of the coefficient of the quadratic term is the quadratic selection gradient, γ, in which negative values imply stabilizing selection, and the intercept is the mean fitness in the population, $\bar{w}$. The authors report estimates of γ from their regressions, and so to get strengths of stabilizing selection (w) under our slightly different fitness model, we make the following approximations to fitness. At equilibrium, the distribution of phenotypes are closely centered around the optimal phenotype relative to the width of the fitness gradient, and therefore we assume that mean fitness $\bar{w} \approx 1$. Additionally assuming

that the mean phenotype in the population is the optimal phenotype, we can then approximate that the quadratic fitness function that is equivalent to ours when $\gamma = -\frac{1}{w^2}$, because $\frac{\gamma}{2}r^2+1 \approx \exp(\frac{\gamma r^2}{2})$. Based on incremental γ values of 0, -0.02, -0.04, and -0.06 (the maximum estimated), we use the following values of the strength of stabilizing selection on the trait: $w = \infty$ (neutral trait), $w = 7$, $w = 5$, and $w = 4$.

3.8.1.2 Choice of effect size distribution

Effect sizes of new mutations were drawn from a normal distribution or mixture of normal distributions with mean zero. Provided that the strengths of selection we chose correspond to selection on standard normal phenotypes, we used a standard deviation that would result in a genetic variance close to 1 at equilibrium. However, we note that the size of the mutational target provides a simple inflation factor to the variance and cancels out when calculating the reduction in prediction accuracy; the mutation effect size distribution that we use is more important for determining the distribution of selection coefficients. See equation 3.18 for the solution to the equilibrium genic variance. Since the strength of selection helps determine the equilibrium genic variance, by changing the variance of effect sizes of incoming mutations, we used the weakest strength of selection to determine the variance of this effect size distribution ($w = 7$). For a Gaussian distribution of effect sizes, this leads to a solution for the standard deviation (σ_a) of 0.1. This distribution generates mutations that are mostly nearly neutral or weakly selected. To explore how different selection regimes lead to different declines in prediction accuracy, we also simulated mutation effect size distributions that involve a higher density of nearly neutral mutations or a higher density of strongly selected mutations. We used a Gaussian distribution with $\sigma_a = 0.05$ for the former and a mixture of Gaussians for the latter, which we call the "heavy tail" distribution. The heavy tail distribution generates mutations from a Gaussian distribution with $\sigma_a = 0.05$ that has mean 0 with probability 0.65 and mean ± 0.25 with probability 0.175. We chose these specifications for the heavy tail

distribution in order to match the median absolute deviation of selection coefficients for the standard Gaussian with $\sigma_a = 0.1$.

3.8.1.3 Extension: Directional Selection

We extended our baseline stabilizing selection scenario for directional selection by imposing a positive optimum shift of either one or two standard deviations of the phenotypic distribution. The optimum shifted in the same way 2500 generations after the pair of populations diverged. We started directional selection simulations from the states of the baseline scenario that were recorded at the time of the optimum shift. To calculate the extent of the optimum shift in each simulation, we averaged the phenotypic standard deviation of each population in that simulation. Simulation details and results for other extensions can be found in Supplement 3.9.3.

3.8.2 Relative prediction accuracy of polygenic scores

The prediction accuracy of a polygenic score is the the proportion of the phenotypic variance explained by that polygenic score (the squared correlation R^2). We write out the correlation between an individual's polygenic score constructed from unlinked polymorphisms discovered by GWAS, S_i, and the true additive genetic value, G_i, in a particular target population. We assume that a GWAS discovers associations only at causal loci, and that at these loci with significant associations, it estimates their true effects. There are L_s loci with significant associations in the study and a remaining L_r that contribute to trait variation in a particular

target population. Therefore an individual's predicted and true genetic values are

$$\begin{aligned} S_i &= \sum_{l=1}^{L_s} a_l g_{il} \\ G_i &= \sum_{l=1}^{L_s+L_r} a_l g_{il} \\ &= S_i + G_i^{(r)}, \text{where } G_i^{(r)} = \sum_{l=L_s+1}^{L_s+L_r} a_l g_{il}, \end{aligned} \tag{3.7}$$

in which their true genetic value is a sum of their predicted genetic value and genetic value contributed by the remainder of sites contributing to trait variation. Because the predicted genetic value is a portion of the true genetic value, the covariance between them is the variance in predicted genetic values,

$$\begin{aligned} \text{Cov}(S, G) &= \text{Cov}(S, S + G_i^{(r)}) \\ &= \text{Cov}(S, S) \\ &= \text{Var}(S) \end{aligned} \tag{3.8}$$

Thus the correlation between the predicted and true genetic values is

$$\text{Corr}(S, G) = \frac{\text{Var}(S)}{\sqrt{\text{Var}(S) \times \text{Var}(G)}} = \sqrt{\frac{\text{Var}(S)}{\text{Var}(G)}}, \tag{3.9}$$

meaning that the correlation between the predicted and true additive genetic values is the square root of the proportion of the trait variance explained by SNPs ascertained by the GWAS, or that the variance in true genetic values explained by polygenic scores (r^2) is the proportion of the variance explained by the polymorphisms used to construct the polygenic scores. This value also represents the reduction in the prediction accuracy of polygenic scores compared to additive genetic values, when predicting true additive phenotypes (with contributions from both genetics and environment). We denote the additive environmental contribution to the phenotype in an individual as E_i. The prediction accuracy of an

individual's full additive phenotype ($G_i + E_i$) when using polygenic scores is

$$\begin{aligned} r_S^2 &= \frac{\mathrm{Cov}^2(S, G+E)}{\mathrm{Var}(S)\mathrm{Var}(G+E)} \\ &= \frac{\mathrm{Var}^2(S)}{\mathrm{Var}(S)\mathrm{Var}(G+E)} \\ &= \frac{\mathrm{Var}(S)}{\mathrm{Var}(G+E)} \end{aligned} \tag{3.10}$$

and when using additive genetic values is

$$\begin{aligned} r_G^2 &= \frac{\mathrm{Cov}^2(G, G+E)}{\mathrm{Var}(G)\mathrm{Var}(G+E)} \\ &= \frac{\mathrm{Var}(G)}{\mathrm{Var}(G+E)}, \end{aligned} \tag{3.11}$$

such that the reduction of prediction accuracy when using polygenic scores instead of additive genetic values is

$$\frac{r_S^2}{r_G^2} = \frac{\mathrm{Var}(S)}{\mathrm{Var}(G)}. \tag{3.12}$$

Thus r_S^2/r_G^2 represents the prediction accuracy of polygenic scores for additive genetic values, as well as the reduction in prediction accuracy for full additive phenotypes due to incomplete ascertainment.

3.8.2.1 Modification for gene-by-environment interactions

While we assume the effects of a causal allele are perfectly estimated in the GWAS sample, these effects may differ in populations that experience different environments (GxE). Thus for populations not represented in the GWAS, the correlation between polygenic scores and true additive genetic values will be lower than without GxE. Imagine that we use the set of effect sizes from GWAS population A ($a_{l,A}$) to construct polygenic scores for population B where the true effect sizes ($a_{l,B}$) differ. The covariance of the polygenic scores with true additive genetic values is then

$$\mathrm{Cov}(S, G) = \sum_l a_{l,A} a_{l,B} \mathrm{Var}(g_l). \tag{3.13}$$

Therefore the correlation between polygenic scores and additive genetic values is

$$\begin{aligned}\mathrm{Corr}(S,G) &= \frac{\sum_l a_{l,A}a_{l,B}\mathrm{Var}(g_l)}{\sqrt{Var(S)Var(G)}}, \text{where} \\ \mathrm{Var}(S) &= \sum_{l=1}^{L_s} a_{l,A}^2\mathrm{Var}(g_l), \text{and} \\ \mathrm{Var}(G) &= \sum_{l=1}^{L_s+L_r} a_{l,B}^2\mathrm{Var}(g_l).\end{aligned} \tag{3.14}$$

For a neutrally evolving trait, effect sizes and heterozygosities are independent and so

$$\mathrm{Corr}(S,G) = \mathrm{Corr}(a_{l,A},a_{l,B})\frac{\sum_l \mathrm{Var}(g_l)}{\sqrt{\left(\sum_{l=1}^{L_s}\mathrm{Var}(g_l)\right)\left(\sum_{l=1}^{L_s+L_r}\mathrm{Var}(g_l)\right)}} \tag{3.15}$$

such that the neutral drop in prediction accuracy due to GxE is simply the decrease from 1 of the correlation of effect sizes between A and B.

3.8.3 Modeling Details

3.8.3.1 Background

Assuming Gaussian stabilizing selection and that the population mean stays close to its optimum value, selection at an individual locus is well described by a model of underdominance where the per-generation change in an allele's frequency x can be described by its mean and variance,

$$\begin{aligned}\mathrm{E}(\Delta x) &\approx sx(1-x)(x-1/2) \\ \mathrm{Var}(\Delta x) &\approx \frac{x(1-x)}{2N},\end{aligned} \tag{3.16}$$

where $s = \frac{a^2}{w^2}$ is the the selection coefficient, with a being the additive effect of the allele with frequency x relative to the alternate allele, and N is the diploid population size. This approximation further assumes that the phenotypic variance is much smaller than the width of the stabilizing selection ($\sigma^2 \ll w$).

Assuming an infinite sites model, the mean time that a mutation spends in a certain frequency interval before fixation or loss (at a single site) multiplied by its population mutation rate is equivalent to the expected number of alleles segregating at those frequencies at a single time point (see Sawyer and Hartl, 1992). The mutation rate of an allele with effect a is $2NU\Pr(a)$, where U is the per-generation mutation rate for the entire mutational target, and $\Pr(a)$ is the proportion of mutations with effect a (with density $f(a \mid \sigma_a)$, where σ_a is the standard deviation of the mutation effect size distribution). The mean time τ spent in frequency interval $(x,\ x+dx)$ for a new mutation with effect a can be solved using the Green's function (see equation A18 of Simons et al., 2018), such that

$$\tau(x \mid a, w) = \frac{\sqrt{\frac{N\pi}{2s}} \exp\left(2Ns(x-\frac{1}{2})^2\right) \left(\operatorname{erf}(\gamma) + \operatorname{erf}(\gamma(1-2x))\right) \left(\operatorname{erf}(\gamma) - \operatorname{erf}\left(\gamma\left(1-\frac{1}{N}\right)\right)\right)}{\operatorname{erf}(\gamma)\, x(1-x)} \tag{3.17}$$

where erf is the error function and $\gamma = \sqrt{\frac{Ns}{2}}$. Thus we can solve for the expected total additive genic variance at equilibrium by summing over the heterozygosities contributed by all sites,

$$\mathrm{E}(V_a|w, \sigma_a) \approx 2NU \int_{-\infty}^{\infty} \int_{1/2N}^{1-1/2N} 2a^2 x(1-x)\tau(x \mid a, w) f(a \mid \sigma_a) dx da, \tag{3.18}$$

3.8.3.2 Results

In our models we focus on divergence without gene flow between a pair of populations, in which case they only share polymorphisms that arose in their common ancestor and were not lost in either population. To predict the loss of ancestral heterozygosity and probability that the descendant populations share polymorphisms, and thus describe the proportion of additive genic variance in one population explained by polymorphisms identified by a GWAS in the other, we consider the frequency trajectory of ancestrally segregating variants. To fully calculate these quantities we would need the diffusion transition density with underdominant selection to calculate the distribution of the frequency in the current day given the ancestral

frequency, but we can build simple approximations of the combined effects of selection and drift. If we assume that the frequency of our allele x is close to 0 such that $1 - x \approx 1$, then we can approximate its expected per-generation change (eqn (3.16)) as

$$\mathrm{E}(\Delta x) \approx sx(x - 1/2), \tag{3.19}$$

such that its deterministic frequency trajectory follows the logistic function from starting frequency x_0,

$$\mathrm{E}(x_t \mid x_0, a, w) \approx \frac{1/2}{1 - (\exp(st/2) \times (1 - 1/2x_0))}, \text{where } s = \frac{a^2}{w^2} \tag{3.20}$$

To consider the effects of genetic drift on the frequency trajectory, we assume that x_t is normally distributed around its deterministic frequency trajectory as follows,

$$g(x_t \mid x_0, a, w) \sim \mathcal{N}\left(\mathrm{E}(x_t \mid x_0, a, w),\ x_0(1 - x_0)\frac{t}{2N}\right). \tag{3.21}$$

Reduction in variance due to drift and selection We can predict the additive genic variance contributed by all ancestral polymorphisms (anc) to the descendent (desc) after t generations by averaging over their effect sizes, starting frequencies, and trajectories,

$$\begin{aligned}\mathrm{E}(V_{a[\text{anc, desc}]}(t, w, \sigma_a)) = 2NU \int_{-\infty}^{\infty} \Bigg[\int_{1/2N}^{1-1/2N} \Bigg[\int_0^1 \underbrace{2a^2 x_t(1 - x_t)}_{\text{var. after time } t} \underbrace{g(x_t \mid x_0, a, w)}_{\substack{\text{prob. density of} \\ x_t \text{ given } x_0 \text{ and } a}} dx_t \Bigg] \\ \times \underbrace{\tau(x_0 \mid a, w)}_{\substack{\text{mean time} \\ \text{at freq. } x_0 \\ \text{given effect } a}} dx_0 \Bigg] \underbrace{f(a \mid \sigma_a)}_{\substack{\text{prob. density} \\ \text{of } a}} da.\end{aligned} \tag{3.22}$$

The total reduction in the additive genetic variance contributed by ancestral polymorphisms after time t is thus given by $\frac{\mathrm{E}(V_{a[\text{anc, desc}]}(t,w,\sigma_a))}{\mathrm{E}(V_a(w,\sigma_a))}$.

To obtain the reduction in heterozygosity used in Figure 3.2A, we use the ratio of expected heterozygosities for the midpoints of the effect size bins evaluated. The equations we use

have a similar structure to those where we predict the total variance explained by ancestral polymorphisms at time 0 and t, except we do not integrate over all effect sizes and remove a^2 to calculate the heterozygosity instead of variance,

$$\frac{H_{t[\text{anc, desc}]}(w,a)}{H_0(w,a)} = \frac{\int_{1/2N}^{1-1/2N}\left[\int_0^1 2x_t(1-x_t)g(x_t \mid x_0,a,w)dx_t\right]\tau(x_0 \mid a,w)dx_0}{\int_{1/2N}^{1-1/2N} 2x_0(1-x_0)\tau(x_0 \mid a,w)dx_0,} \tag{3.23}$$

where H_t refers to the heterozygosity at time t.

Variance in ancestor explained by polymorphisms in descendant To understand the ancestral variance attributable to present-day polymorphisms, we need to think of the properties of present-day alleles backward in time. For alleles with effect sizes large enough for selection to act strongly, selection (usually) constrains them from reaching appreciable frequencies and selection acts approximately like additive selection against the allele (eqn (3.19)). Conditional on the present-day frequency, the distribution of a deleterious, additive allele's trajectory backward in time is the same as the process forward in time, such that $g(x_t|x_0,a,w) = g(x_0|x_t,a,w)$ (Maruyama and Kimura, 1975). Therefore, we can write the expected additive genic variance in the ancestral population explained by variation segregating in the descendant population, in which they have diverged for t generations, as

$$\begin{aligned}\mathrm{E}(V_{a[\text{desc, anc}]}(t,w,\sigma_a)) = 2NU\int_{-\infty}^{\infty}\Bigg[\int_{1/2N}^{1-1/2N}\Bigg[\int_0^1 \underbrace{2a^2x_0(1-x_0)}_{\text{var. } t \text{ gen. ago}}\underbrace{g(x_0 \mid x_t,a,w)}_{\substack{\text{prob. density of}\\ x_0 \text{ given } x_t \text{ and } a}}dx_0\Bigg] \\ \times \underbrace{\tau(x_t \mid a,w)}_{\substack{\text{mean time}\\ \text{at freq. } x_t\\ \text{given effect } a}}dx_t\Bigg]\underbrace{f(a \mid \sigma_a)}_{\substack{\text{prob. density}\\ \text{of } a}}da,\end{aligned} \tag{3.24}$$

which is identical to the additive genic variance contributed by all ancestral polymorphisms to the descendent (with just a flip in the subscripts, see equation (3.22)). Thus the decay in the additive genic variance segregating in an ancestral population to the present day is

the same as loss in prediction accuracy when using present day polymorphisms to predict phenotypes in the ancestral population.

Variance in population B explained by polymorphisms in population A When the population in which we make predictions is contemporary with the population represented in the GWAS, we again consider the variance explained by all polymorphisms that contribute to the variance in the prediction population. In our model of population divergence, these sites that contribute to the variance in the prediction population are sites that were segregating in the common ancestor that remain segregating in both the GWAS and prediction population. We first account for the variance contributed by ancestrally segregating sites in the descendant population in which we make predictions (B), as explained in equation (3.22)), and then for the possibility that each site is segregating, and thus ascertained, in the population from which the GWAS sample was drawn (A),

$$\mathrm{E}(V_{a[\mathrm{A,\ B}]}(a,t,w,\sigma_a)) = 2NU \int_{-\infty}^{\infty} \Bigg[\int_{1/2N}^{1-1/2N} \underbrace{\Bigg[\int_0^1 2a^2 x_{t,B}(1-x_{t,B}) g(x_{t,B} \mid x_0, a, w) dx_{t,B} \Bigg]}_{\substack{\text{expected variance after time } t \text{ explained by} \\ \text{ancestral variants starting at freq. } x_0 \text{ with effect } a}} \times \underbrace{\Pr(0 < x_{t,A} < 1 \mid x_0, a, w)}_{\substack{\text{prob. of ancestral variant} \\ \text{starting at freq. } x_0 \text{ with effect } a \\ \text{segregating in pop. A}}} \underbrace{\tau(x_0 \mid a, w)}_{\substack{\text{mean time} \\ \text{at freq. } x_0 \\ \text{given effect } a}} dx_0 \Bigg] \times \underbrace{f(a \mid \sigma_a)}_{\substack{\text{prob. density} \\ \text{of } a}} da, \tag{3.25}$$

where x_0 is the frequency of an allele in the common ancestor of populations A and B, $x_{t,A}$ and $x_{t,B}$ are its frequencies after time t in each of those descendant populations respectively, and the probability that the ancestral variant is segregating in population A for ascertainment

is given by,

$$\Pr(0 < x_{t,A} < 1 \mid x_0, a, w) = \int_{1/2N}^{1-1/2N} g(x_{t,A} \mid x_0, a, w) d_{x_{t,A}}. \tag{3.26}$$

3.9 Supplementary Text and Figures

3.9.1 Figures referenced in the main text

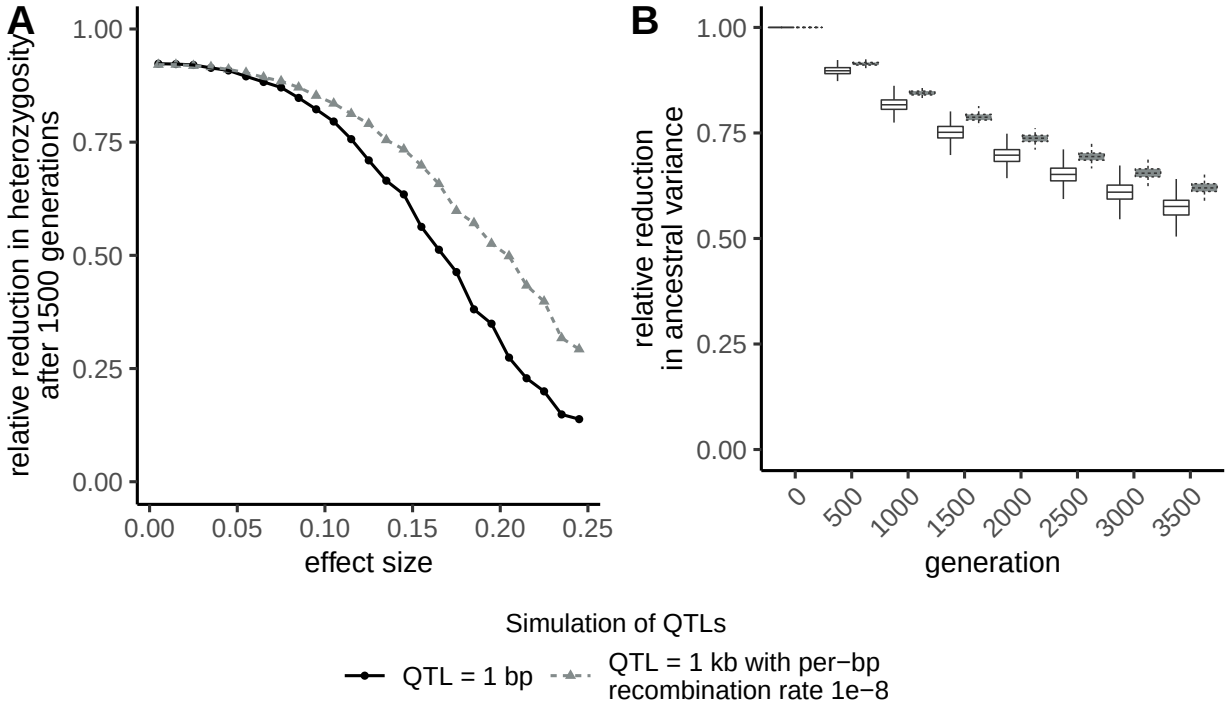

Figure 3.6: Comparison of **A)** reduction in heterozygosity at loci that contributed to the variance 1500 generations ago within effect size bins of width 0.01 and **B)** total reduction in ancestral variance over time for simulations that differed on the size of QTLs. QTLs were either 1bp (main text results) or 1kb with low within-QTL recombination. Simulations with 1kb QTL consisted of genomes comprising 3000 QTL, with free recombination between QTLs and a per base pair recombination rate of 10^{-8} within a QTL. Mutations arose at rate 10^{-8} per base pair. All other simulation details match the baseline stabilizing selection scenario introduced in Appendix 3.8.1. The genome-wide mutation rate (U) differs for each simulation type, however under the assumptions of our models this should not influence the reduction in heterozygosity or variance.

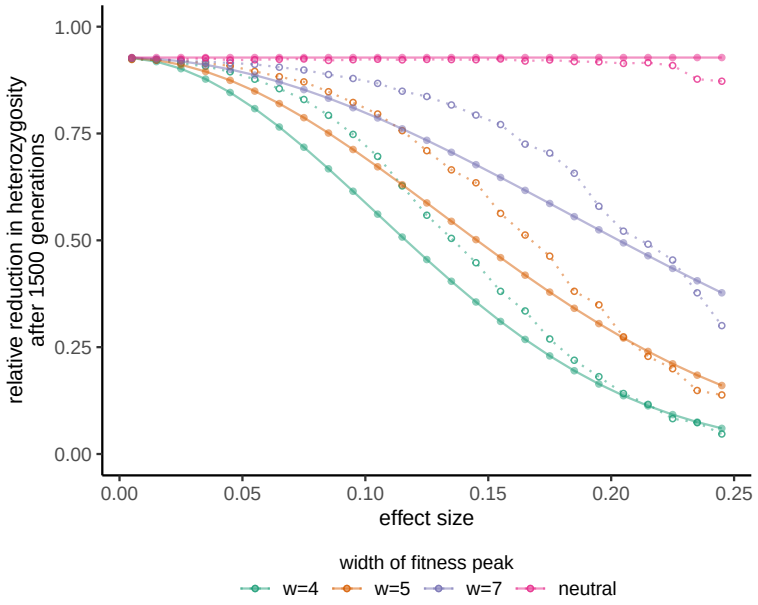

Figure 3.7: Reduction in heterozygosity at loci that contributed to the variance 1500 generations ago. Each open point, connected by a dashed line, represents the midpoint of the effect size bin of width 0.01 within which we averaged heterozygosity from 200 simulations. Each filled point, connected by a solid line, represents our analytical predictions for that midpoint using the approximation in Equation 3.3. These predictions over-estimate the reduction in heterozygosity, whereas the predictions from the diffusion approximation provide a much better fit (Figure 3.2).

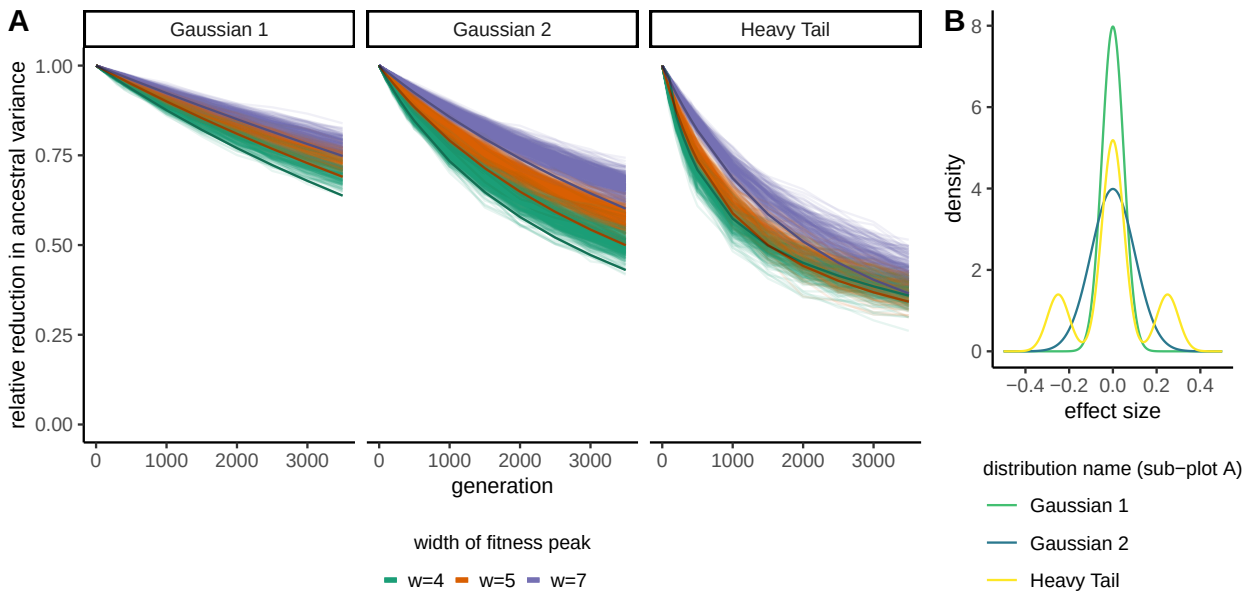

Figure 3.8: Reduction over time in the total variance contributed by polymorphisms in an ancestral population. Lighter lines show results from simulations and darker lines show analytical predictions based on the diffusion approximation. The results from "Gaussian 2" are the same as presented in Figure 3.2B.

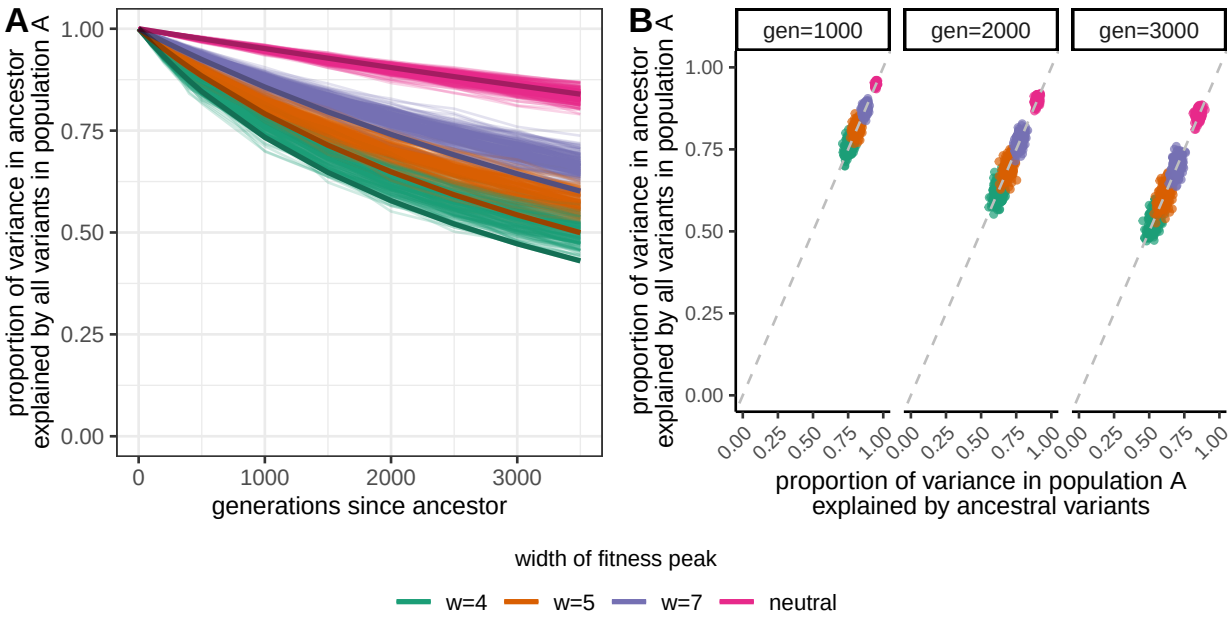

Figure 3.9: A) Reduction in prediction accuracy in population ancestral to population A with increasing divergence time between them. Lighter lines show results from simulations and darker lines show analytical predictions. B) At particular time depths of divergence, the reduction in prediction accuracy in ancestral population (y-axis) is about the same as the reduction in ancestral variance (x-axis). Points represent results from a single simulation. The dashed line is the line of equality.

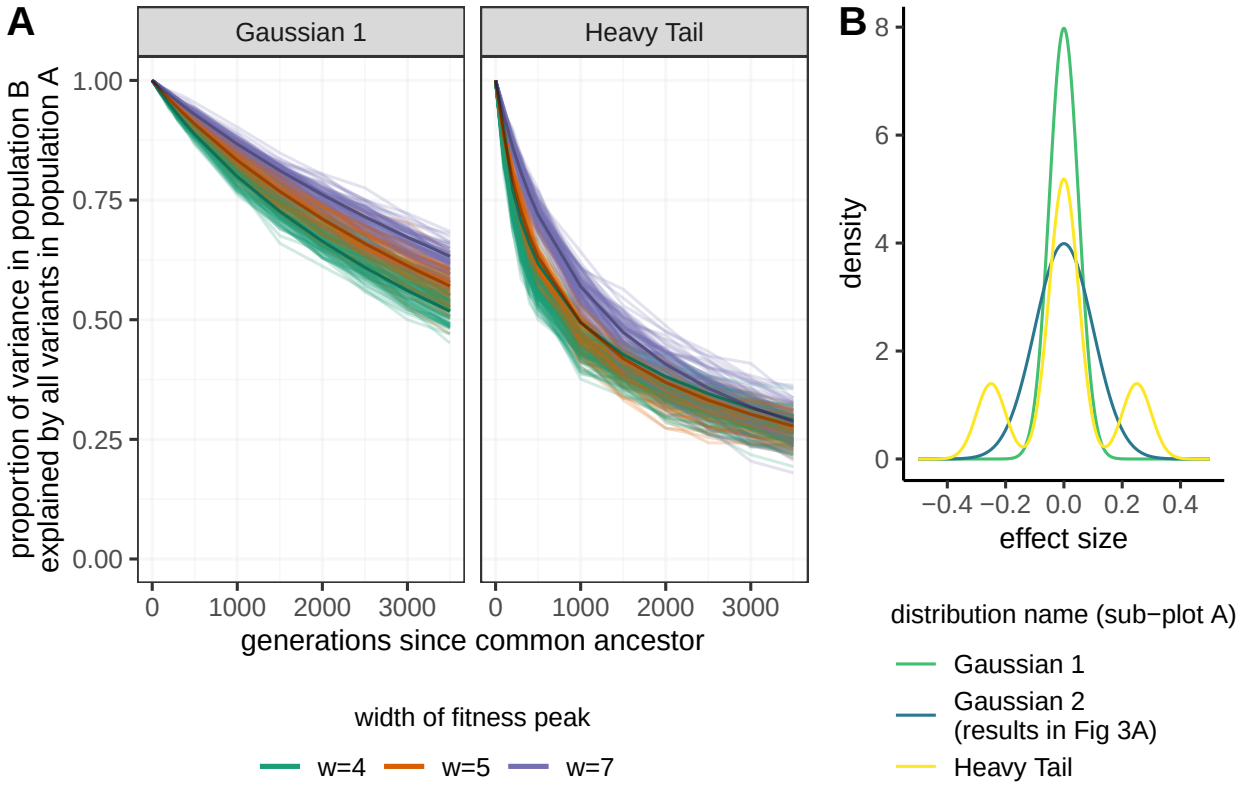

Figure 3.10: **A)** Reduction in prediction accuracy in population B when ascertaining all polymorphisms in population A, under different mutation effect size distributions than shown in the main text. Light lines show results from simulations, and dark lines show results from analytical predictions. **B)** mutation effect size distributions correspond to what was used to produce results from sub-figure A. Results from the distribution "Gaussian 2" are shown in the main text Figure 3.3A.

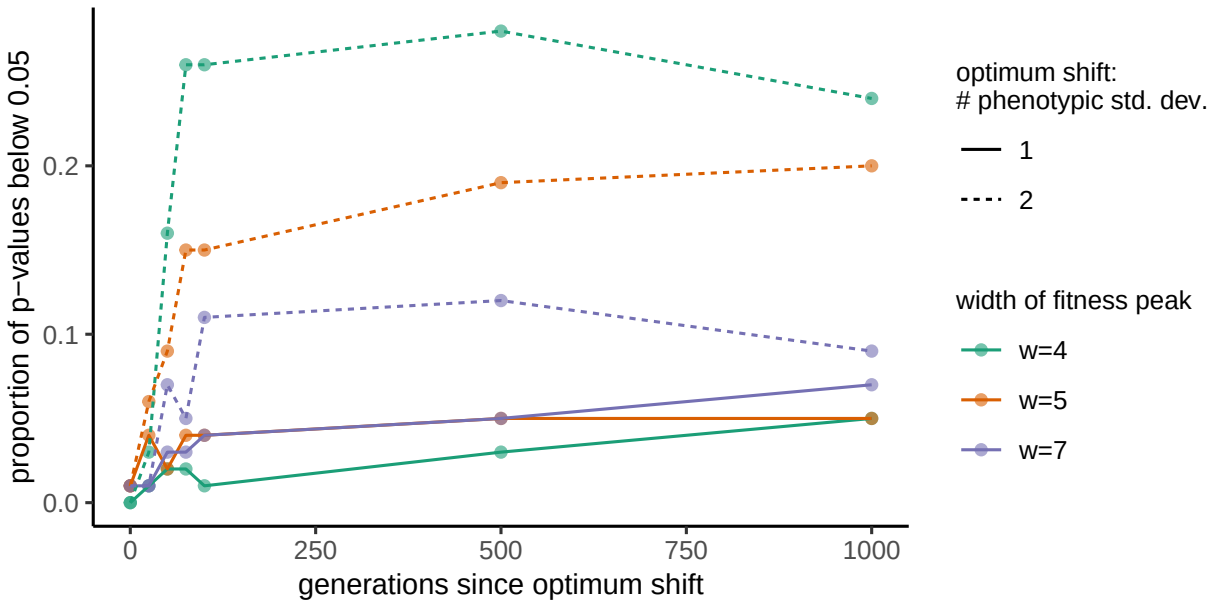

Figure 3.11: Under the directional selection scenario, proportion of p-values below the significance threshold of 0.05 with time since the shift in the optimum (the shift began at 2500 generations of divergence between populations A and B). Each point is averaged across 100 simulations. Solid lines connect points corresponding to an optimum shift of one standard deviation of the phenotypic distribution, while dashed lines connect points corresponding to an optimum shift of two standard deviations. Following the optimum shift, we recorded results every 25 generations for the first 100 generations, and then every 500 generations.

3.9.2 Decline in prediction accuracy for population B relative to population A under various ascertainment schemes and mutational effect size distributions

Here we compare the decline in prediction accuracy for population B relative to population A among various ascertainment schemes within population A. Given that any set of polymorphisms discovered within population A is a subset of all polymorphisms in population A, we use the case where all polymorphisms in A were ascertained as the point of comparison for other ascertainment schemes. We simulate cases where populations A and B experienced the same selective environment, such that they have the same expected total additive genic variance at equilibrium. Therefore the prediction accuracy in population B relative to population A is simply the ratio of the variances explained by the GWAS-significant sites in each population. This relative decline will not change when moving from all polymorphisms in A ascertained to a particular subset if the reduction in variance explained for population A and population B is the same. We find that this is approximately the case for a Gaussian mutation effect size distribution containing mostly nearly neutral and weakly selected mutations (Figure 3.3C; 3.13). However, the heavy-tailed effect size distribution, which consists of a higher density of strongly selected mutations, leads to different patterns of relative decline for population B when compared to the case when all polymorphisms in A are ascertained. This is because in the heavy-tailed case, most ancestral polymorphisms that are shared between populations are nearly neutral and segregate at intermediate frequencies 3.12. Thus when moving from ascertaining all polymorphisms in A to just those with MAF>1%, population B experiences less of a reduction in variance explained than population A. However, few of those shared polymorphisms explain much of the variance in population A, and thus population B experiences more of a reduction in variance explained when ascertaining the polymorphisms in population A based on variance explained.

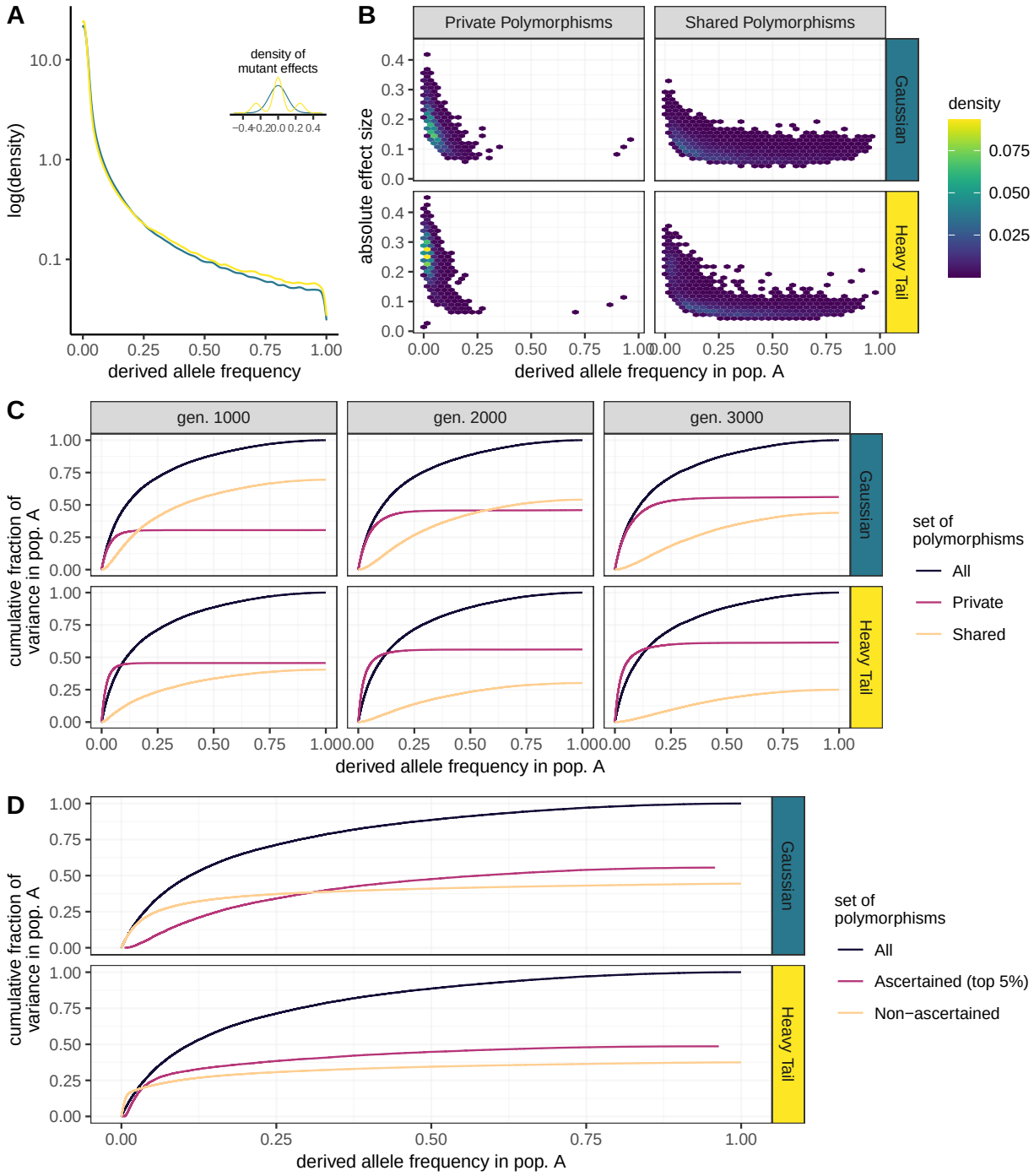

Figure 3.12: (Continued on the following page.)

Figure 3.12: Descriptions of the joint distribution of allele frequencies and effect sizes within a population. **A)** Derived allele frequency spectrum under two mutation effect size distributions shown in the inset, whose colors are used to reference them in the panels of B-D. **B)** Joint density of derived allele frequencies and effect sizes in population A, separated for polymorphisms private to A and shared with B. Note that some private polymorphisms may be polymorphisms in the common ancestor of A and B that persisted in A but were subsequently lost in B. **C)** Cumulative fraction of the total variance in population A with increasing derived allele frequencies in population A for different generations of divergence between population B (columns) and two different mutation effect size distributions (rows). Line colors refer to all polymorphisms in population A, polymorphisms private to population A, and polymorphisms shared between populations A and B. **D)** Cumulative fraction of the total variance in population A with increasing derived allele frequencies in population A for two different mutation effect size distributions (rows). Line colors refer to all polymorphisms in population A, the top 5% of variance-contributing polymorphisms in population A, and the remaining non-ascertained polymorphisms (bottom 95%).

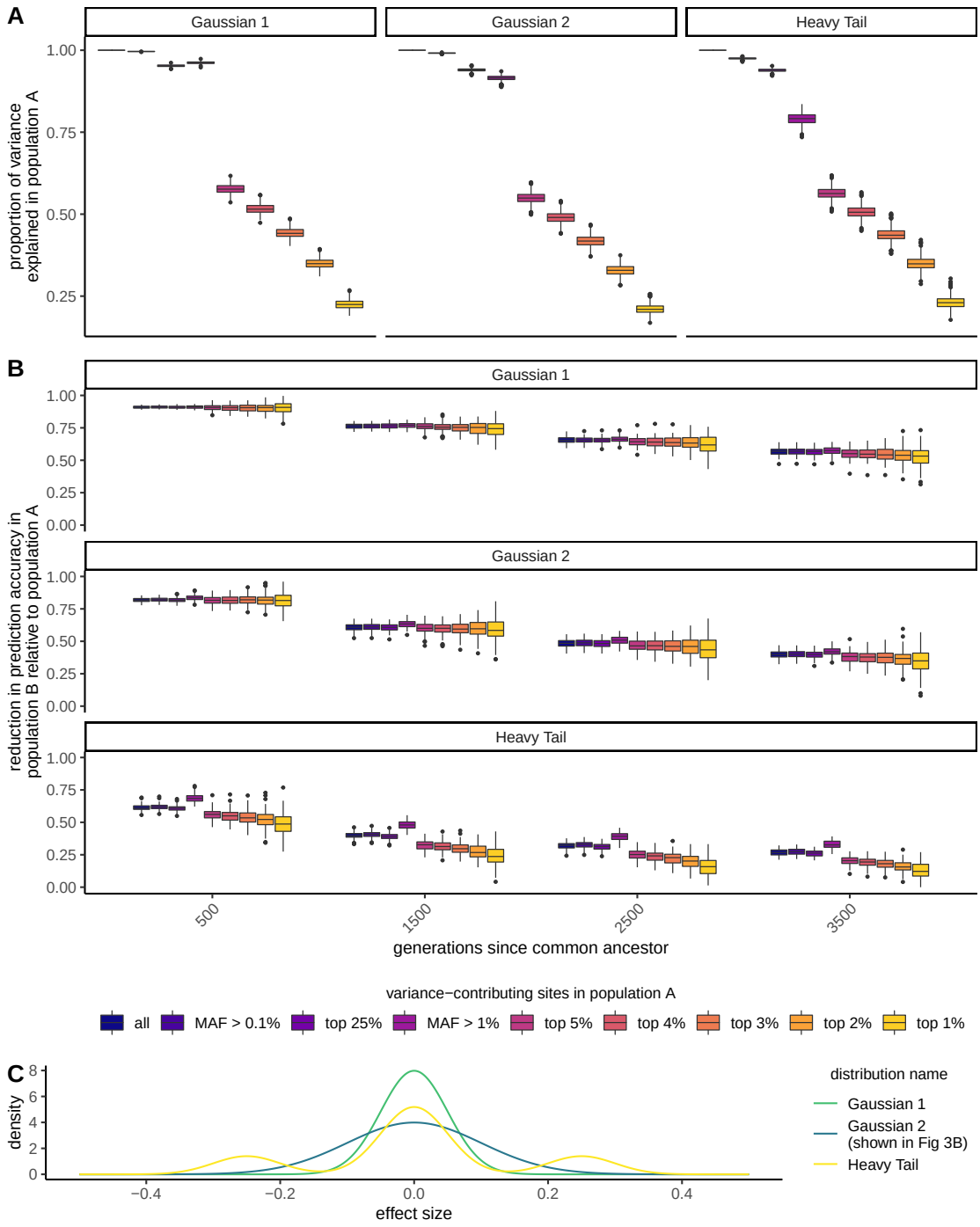

Figure 3.13: (Continued on the following page.)

Figure 3.13: **A)** Proportion of the variance explained in population A when ascertaining different sets of variance-contributing sites in population A. The proportion of the variance explained is the same as the reduction in prediction accuracy for population A when ascertaining a particular set of sites instead of all variance-contributing sites. Facets show results from three different mutation effect size distributions. **B)** Reduction in prediction accuracy for population B *relative* to population A under different ascertainment schemes in population A and mutation effect size distributions. **C)** mutation effect size distributions that were used to produce results in sub-figures **A** and **B**.

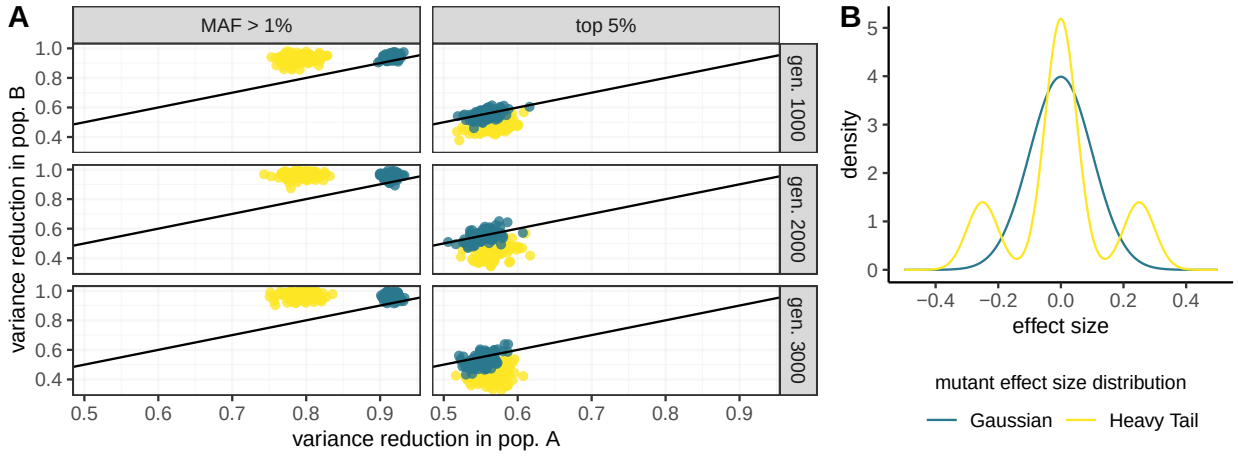

Figure 3.14: **A)** The relative reduction in the variance explained for populations A and B when switching from ascertaining all variance-contributing polymorphisms in population A to either those polymorphisms with MAF>1% or the top 5% of variance-contributing polymorphisms. Points show results from simulations, and their colors correspond to the mutation effect size distributions in **(B)**. The black line is the line of equality. Points that fall along this line indicate the same variance reduction in both populations, and thus a similar decline in prediction accuracy for this ascertainment scheme compared to when all polymorphisms in population A are ascertained. Points above the line indicate that population B experiences less of a reduction in variance explained, and points below it indicate that population B experiences more of a reduction in variance explained, compared to population A.

3.9.3 Extensions to the baseline stabilizing selection scenario

3.9.4 Directional Selection

Simulation details for directional selection are provided in Appendix 3.8.1.3. We also discuss in the main text how the same optimum shift among populations can lead to false signals of adaptive differentiation.

Accuracy of polygenic score predictions The observed relative reduction in prediction accuracy for the baseline scenario (main text) is slightly weaker than under directional selection for the optimum shifts we considered (up to two standard deviations of the trait distribution; 3.15A).

3.9.5 Gene-by-environment interactions (GxE)

Simulation Details We simulated GxE using the same framework as the baseline scenario, except we drew the effect sizes of mutations from a multivariate normal distribution with three dimensions representing the ancestral burn-in population and its two descendant populations. The covariance matrix was determined according to the standard deviation of the effect size distribution and correlation of effect sizes provided (either 0.9 or 0.95). Phenotypes and thus fitness were calculated using the effect size assigned for the particular population in which they were evaluated. This means that the environments of the descendant populations changed at their time of divergence from the ancestral population, causing the effect sizes in each descendant population to be perturbed away from the ancestral effect sizes by a small normal deviate. These environmental changes do not generate a difference in the mean environment nor the optimum phenotype. The effect size correlations we chose are higher than most estimates, though GxG and differences in LD likely play a role as well (Liu et al., 2015; Galinsky et al., 2019; Lam et al., 2019; Veturi et al., 2019).

Accuracy of polygenic score predictions Gene by environment interactions (GxE) combined with environmental differences between the GWAS sample and unrepresented pop-

ulation result in lower prediction accuracies in the unrepresented population, relative to the baseline scenario we discuss in the main text (3.15). Two factors contribute to why we find lower prediction accuracies under GxE. First, the effect sizes estimated by the GWAS in population A are only partially correlated with the true effects in population B. Thus, noise is introduced to the predictions when using the effect sizes from A for polygenic prediction in B. In Appendix 3.8.2.1 we describe how we modify the equation for the reduction in prediction accuracy when the effect sizes between the GWAS sample and the prediction population are only partially correlated. Second, the slight changes in effect sizes, due to environmental changes experienced by each descendant population after they shared a common ancestor, transiently inflate the variance in a population relative to equilibrium conditions, leading to a faster removal of ancestral polymorphisms and thus fewer shared polymorphisms between the population pair. Specifically, at equilibrium, most common alleles have small effects. After an environmental change, many of these previously near-zero effect alleles tend to shift toward larger effect sizes (3.16). The accumulation of these minor shifts at many sites leads to an increase in the total variance within the descendant populations, and these changes are not correlated between them, so prediction accuracy is reduced.

Difference in polygenic means among populations Under GxE, we also find higher chances of false signals of adaptation, even though the optimum never changes and is shared among populations. The chances of these false signals decay with time since the common ancestor of the population pair (3.18).

3.9.6 Pleiotropy

Simulation Details We simulated pleiotropy using the same framework as the baseline scenario. A mutation's effect on each trait was independently drawn from the same normal distribution. An individual's fitness was calculated using their Euclidean distance from the optimum, based on their distance in each phenotypic dimension. For this extension we

simulated $n = 5$ and $n = 10$ traits (the baseline scenario is $n = 1$). The selection coefficient that an allele experiences is approximately

$$s \approx \frac{\sum_{i=1}^{n} a_i^2}{w^2}, \tag{3.27}$$

where a_i denotes the allele's effect on trait i (Simons et al., 2018). Thus when we sample the effect of an allele on each trait independently and use the same distribution, with more pleiotropy there would always be stronger selection, where

$$\mathrm{E}(s) = \frac{n\mathrm{Var}(a_i)}{w^2}, \tag{3.28}$$

such that as n increases there would be a decrease in the amount of shared additive genetic variance, and thus portability, between populations. However, we were interested in investigating how the distribution of an allele's effects on multiple traits impacted the amount of shared additive genetic variance between populations, in isolation from the effects of overall stronger selection on alleles. Therefore, we standardized the fitness function so that the average selection coefficient would be the same across pleiotropy levels (n). To achieve the same expectation across simulations, we multiplied w^2 by n in the fitness function. While the average selection coefficient is the same for a particular w across n, its distribution differs according to the number of traits influenced,

$$s \sim \frac{\mathrm{Var(a_i)}}{nw^2}\chi_n^2. \tag{3.29}$$

Accuracy of polygenic score predictions Under this model of pleiotropy, we find less of a reduction in the prediction accuracy of ascertained variation compared to our baseline scenario. This is because the effect of the mutation on the trait of interest becomes more weakly correlated with its total selection coefficient. Specifically, common alleles can have larger effects on the trait of interest but by chance experience weaker selection due to small effects on the other selected traits. Therefore, the initial variance explained by a large

effect polymorphism will be higher in the ancestor and there will be less of a per-generation reduction of variance at this locus. Together, more large effect and common alleles will be shared between populations, increasing the amount of additive genetic variance that polymorphisms ascertained in one population can explain in another. Thus pleiotropy, for a constant average strength of selection, reduces the rate at which prediction accuracy declines.

Difference in polygenic means among populations Under pleiotropy, we see similar, though slightly lower, chances for false signals of adaptation as under the baseline stabilizing selection scenario (3.19).

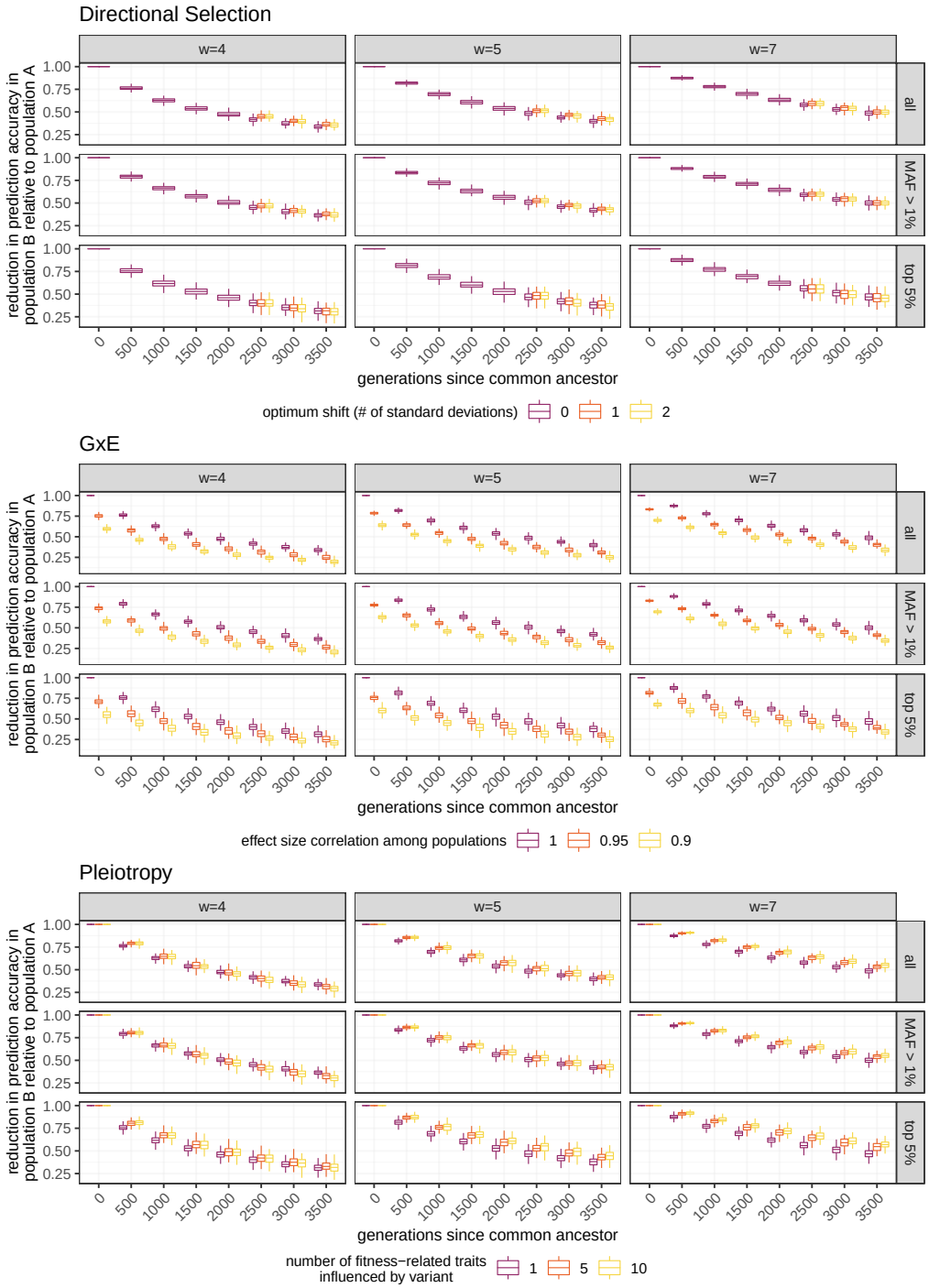

Figure 3.15: (Continued on the following page.)

Figure 3.15: Comparison of the baseline stabilizing scenario and each extension (labelled by title) for the reduction in prediction accuracy in population B. Boxplots show the distribution of observations from 100 simulations, and are colored by the parameter values that vary in a particular extension. Results from the baseline scenario are always represented in purple (the first parameter value in each legend). Each column represents a different width of the fitness peak, and each row represents the set of polymorphisms ascertained in population A (*top*: all, *bottom*: top 5% based on variance contributed).

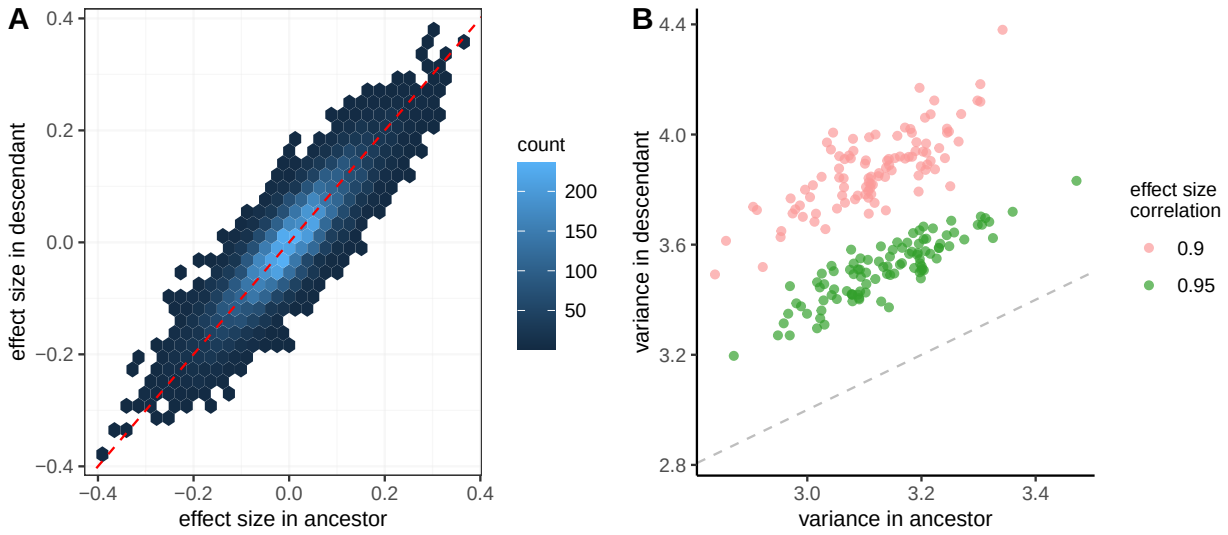

Figure 3.16: Shift of effect sizes and within-population variances when there's GxE and populations A and B begin to experience different environments from their common ancestor at the time of divergence. **A)** Changes in effect sizes from the ancestral population (x-axis) to one of its descendant populations (y-axis) for each variance-contributing polymorphism. Colors represent the number of sites in a particular bin (outlined by hexagon). Results are shown for a single simulation with $w = 5$ and effect size correlation of 0.9. **B)** At the time of divergence/environmental shift, the variance in the descendant population (y-axis) increases relative to the variance in the ancestral population (x-axis) due to the shift in effect sizes. Each point represents a single simulation. Dashed lines in A and B represent lines of equality.

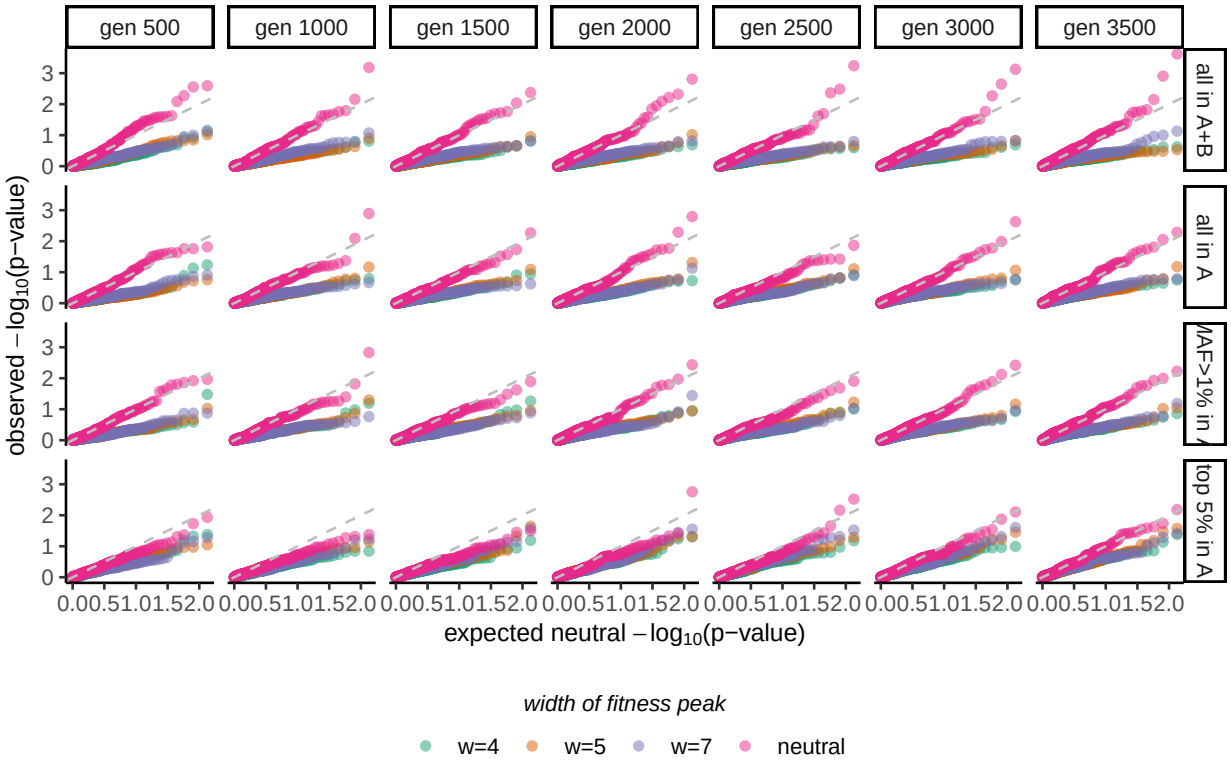

Figure 3.17: Quantile-quantile plot of observed p-values for Q_X under the **baseline** case against expected p-values under neutrality (uniform distribution). The dashed line shows equality; points that lie above this line indicate that the observed distribution has a higher density of low p-values than the neutral distribution, and points that lie below it indicate the opposite. Columns show results from every 500 generations of divergence between populations A and B and rows show results from different sets of ascertained polymorphims (*top:* all polymorphisms from both populations, *top-middle:* all polymorphisms in A, *bottom-middle:* polymorphisms with MAF>1% in A, *bottom:* top 5% of variance-contributing polymorphisms in A).

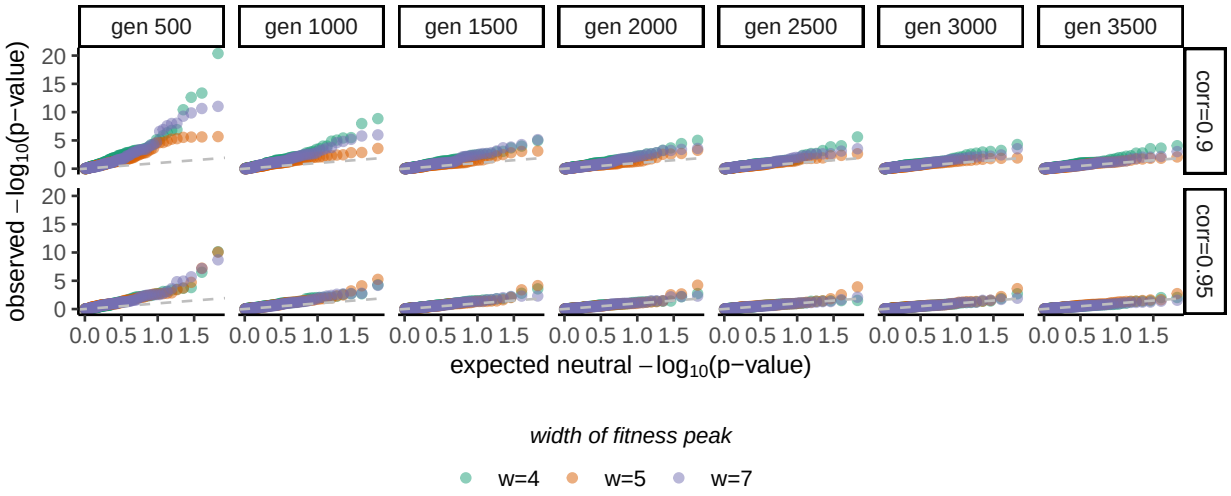

Figure 3.18: Quantile-quantile plot of observed p-values for Q_X under **GxE** against expected p-values under neutrality (uniform distribution). The dashed line shows equality; points that lie above this line indicate that the observed distribution has a higher density of low p-values than the neutral distribution, and points that lie below it indicate the opposite. Columns show results from every 500 generations of divergence between populations A and B and rows show results from each simulated correlation of effect sizes among populations.

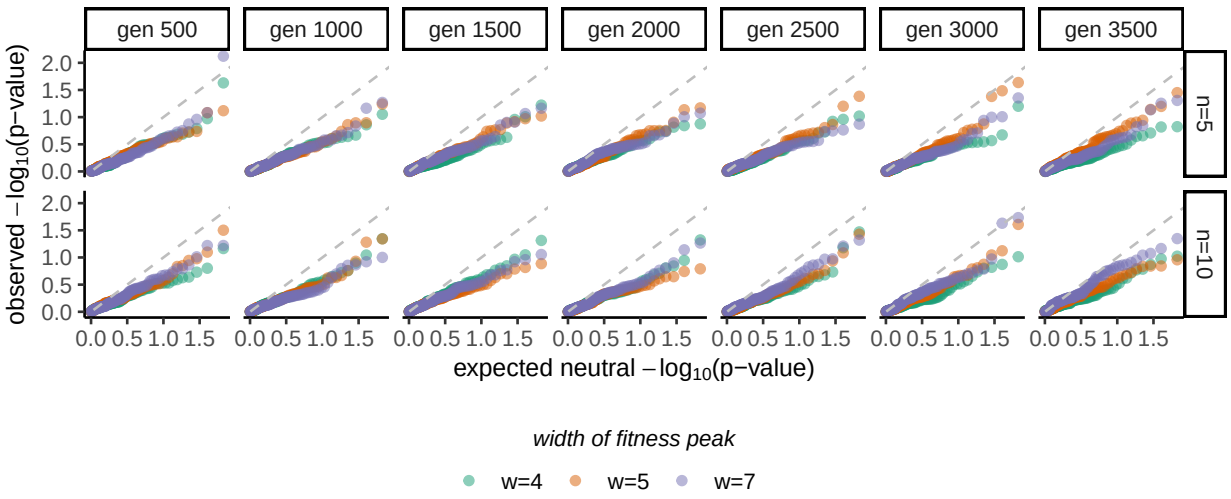

Figure 3.19: Quantile-quantile plot of observed p-values for Q_X under **pleiotropy** against expected p-values under neutrality (uniform distribution). The dashed line shows equality; points that lie above this line indicate that the observed distribution has a higher density of low p-values than the neutral distribution, and points that lie below it indicate the opposite. Columns show results from every 500 generations of divergence between populations A and B and rows show results from each simulated number of fitness-related traits (n) influenced by a mutation.

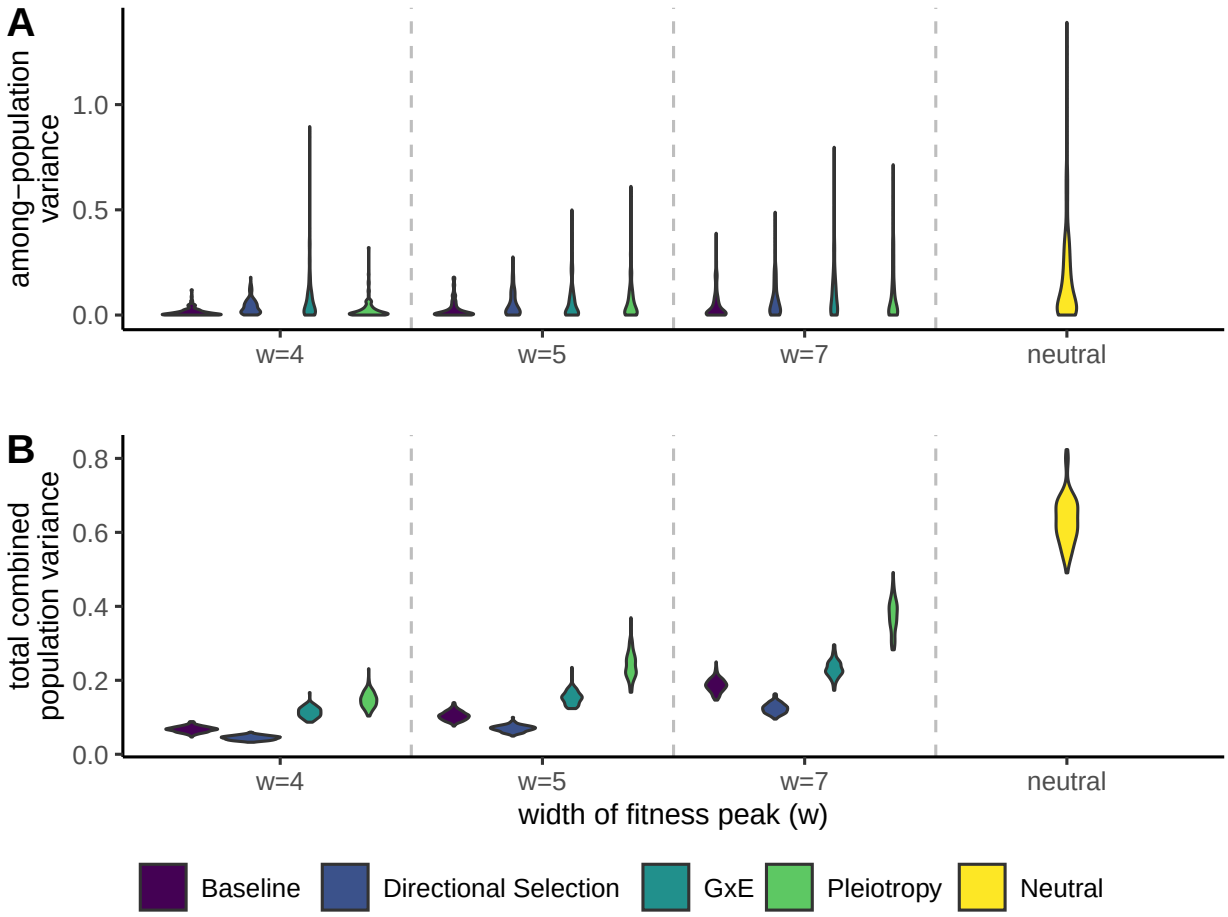

Figure 3.20: Components of Q_{ST} when estimated from the top 1% of variance-contributing sites in population A. Results for GxE are shown for the effect size correlation 0.9 and results for directional selection are shown for an optimum shift of two phenotypic standard deviations. **A)** The estimated mean difference in polygenic scores. The larger this value the larger the among-population variance, which is the numerator of Q_{ST}. **B)** The total variance of the populations combined, which is the denominator of Q_{ST}.

References

Abd El-Aziz, M. M., I. Barragan, C. A. O'Driscoll, L. Goodstadt, E. Prigmore, S. Borrego, M. Mena, J. I. Pieras, M. F. El-Ashry, L. A. Safieh, A. Shah, M. E. Cheetham, N. P. Carter, C. Chakarova, C. P. Ponting, S. S. Bhattacharya, and G. Antinolo (2008). EYS, encoding an ortholog of Drosophila spacemaker, is mutated in autosomal recessive retinitis pigmentosa. *Nature Genetics 40*(11), 1285–1287.

Abi-Rached, L., M. Jobin, S. Kulkarni, A. McWhinnie, K. Dalva, L. Gragert, F. Babrzadeh, B. Gharizadeh, M. Luo, F. A. Plummer, J. Kimani, M. Carrington, D. Middleton, R. Rajalingam, M. Beksac, S. G. E. Marsh, M. Maiers, L. A. Guethlein, S. Tavoularis, A.-M. Little, R. E. Green, P. J. Norman, and P. Parham (2011). The Shaping of Modern Human Immune Systems by Multiregional Admixture with Archaic Humans. *Science 334*, 89–95.

Adhikari, K., J. Mendoza-Revilla, A. Sohail, M. Fuentes-Guajardo, J. Lampert, J. C. Chacón-Duque, M. Hurtado, V. Villegas, V. Granja, V. Acuña-Alonzo, C. Jaramillo, W. Arias, R. B. Lozano, P. Everardo, J. Gómez-Valdés, H. Villamil-Ramírez, C. C. Silva de Cerqueira, T. Hunemeier, V. Ramallo, L. Schuler-Faccini, F. M. Salzano, R. Gonzalez-José, M. C. Bortolini, S. Canizales-Quinteros, C. Gallo, G. Poletti, G. Bedoya, F. Rothhammer, D. J. Tobin, M. Fumagalli, D. Balding, and A. Ruiz-Linares (2019). A GWAS in Latin Americans highlights the convergent evolution of lighter skin pigmentation in Eurasia. *Nature Communications 10*(358).

Alfano, G., P. M. Kruczek, A. Z. Shah, B. Kramarz, G. Jeffery, A. C. Zelhof, and S. S. Bhattacharya (2016). EYS is a protein associated with the ciliary axoneme in rods and cones. *PLoS ONE 11*(11), 1–20.

Allendorf, F. W., R. F. Leary, P. Spruell, and J. K. Wenburg (2001). The problems with hybrids: setting conservation guidelines. *Trends in ecology & evolution 16*(11), 613–622.

Allentoft, M. E., M. Sikora, K.-G. Sjögren, S. Rasmussen, M. Rasmussen, J. Stenderup, P. B. Damgaard, H. Schroeder, T. Ahlström, L. Vinner, A.-S. Malaspinas, A. Margaryan, T. Higham, D. Chivall, N. Lynnerup, L. Harvig, J. Baron, P. D. Casa, P. Dąbrowski, P. R. Duffy, A. V. Ebel, A. Epimakhov, K. Frei, M. Furmanek, T. Gralak, A. Gromov, S. Gronkiewicz, G. Grupe, T. Hajdu, R. Jarysz, V. Khartanovich, A. Khokhlov, V. Kiss, J. Kolář, A. Kriiska, I. Lasak, C. Longhi, G. McGlynn, A. Merkevicius, I. Merkyte, M. Metspalu, R. Mkrtchyan, V. Moiseyev, L. Paja, G. Pálfi, D. Pokutta, Ł. Pospieszny, T. D. Price, L. Saag, M. Sablin, N. Shishlina, V. Smrčka, V. I. Soenov, V. Szeverényi, G. Tóth, S. V. Trifanova, L. Varul, M. Vicze, L. Yepiskoposyan, V. Zhitenev, L. Orlando, T. Sicheritz-Pontén, S. Brunak, R. Nielsen, K. Kristiansen, and E. Willerslev (2015). Population genomics of Bronze Age Eurasia. *Nature 522*, 167–172.

Arbiza, L., E. Zhong, and A. Keinan (2012). NRE: A tool for exploring neutral loci in the human genome. *BMC Bioinformatics 13*(301).

Aris-Brosou, S. (2019). Direct Evidence of an Increasing Mutational Load in Humans. *Molecular Biology and Evolution 36*(12), 2823–2829.

Barbujani, G., A. Magagni, E. Minch, and L. L. Cavalli-Sforza (1997). An apportionment of human dna diversity. *Proceedings of the National Academy of Sciences 94*(9), 4516–4519.

Barton, N. H. and P. D. Keightley (2002). Understanding quantitative genetic variation. *Nature Reviews Genetics 3*(1), 11–21.

Beam, T. A., E. F. Loudermilk, and D. F. Kisor (2017). Pharmacogenetics and pathophysiology of CACNA1S mutations in malignant hyperthermia. *Physiological Genomics 49*, 81–87.

Bell, G. and A. Gonzalez (2009). Evolutionary rescue can prevent extinction following environmental change. *Ecology letters 12*(9), 942–948.

Bentley, A. R., Y. J. Sung, M. R. Brown, T. W. Winkler, A. T. Kraja, I. Ntalla, and E. Al. (2019). Multi-ancestry genome-wide gene–smoking interaction study of 387,272 individuals identifies new loci associated with serum lipids. *Nature Genetics 51*(April), 636–648.

Berens, A. J., T. L. Cooper, and J. Lachance (2017). The genomic health of ancient hominins. *Human Biology 89*(1), 7–19.

Berg, J. J. and G. Coop (2014). A Population Genetic Signal of Polygenic Adaptation. *PLoS Genetics 10*(8), e1004412.

Berg, J. J., A. Harpak, N. Sinnott-Armstrong, A. M. Joergensen, H. Mostafavi, Y. Field, E. A. Boyle, X. Zhang, F. Racimo, J. K. Pritchard, and G. Coop (2019). Reduced signal for polygenic adaptation of height in UK Biobank. *eLife 8*, e39725.

Bergström, A., S. A. McCarthy, R. Hui, M. A. Almarri, Q. Ayub, P. Danecek, Y. Chen, S. Felkel, P. Hallast, J. Kamm, et al. (2020). Insights into human genetic variation and population history from 929 diverse genomes. *Science 367*(6484), eaay5012.

Bernard, D. and D. Vindrieux (2014). PLA2R1: Expression and function in cancer. *Biochimica et Biophysica Acta 1846*(1), 40–44.

Bitarello, B. D. and I. Mathieson (2020). Polygenic scores for height in admixed populations. *G3: Genes, Genomes, Genetics 10*(11), 4027–4036.

Blair, H. J., S. Tompson, Y. N. Liu, J. Campbell, K. MacArthur, C. P. Ponting, V. L. Ruiz-Perez, and J. A. Goodship (2011). Evc2 is a positive modulator of Hedgehog signalling that interacts with Evc at the cilia membrane and is also found in the nucleus. *BMC Biology 9*(14), 1–13.

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

Brown, B. C., A. Genetic, E. Network, T. Diabetes, C. J. Ye, A. L. Price, and N. Zaitlen (2016). Transethnic Genetic-Correlation Estimates from Summary Statistics. *The American Journal of Human Genetics 99*(1), 76–88.

Browning, B. L. and S. R. Browning (2016). Genotype Imputation with Millions of Reference Samples. *American Journal of Human Genetics 98*, 116–126.

Browning, S. R. and B. L. Browning (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *American Journal of Human Genetics 81*, 1084–1097.

Bulmer, M. (1971). The Effect of Selection on Genetic Variability. *The American Naturalist 105*(943), 201–211.

Bumpus, H. (1899). *The Elimination of the Unfit as Illustrated by the Introduced Sparrow, Passer Domesticus: (a Fourth Contribution to the Study of Variation)*. Biological lectures delivered at the Marine Biological Laboratory of Wood's Hole. Gin.

Caparrós-Martín, J. A., M. Valencia, E. Reytor, M. Pacheco, M. Fernandez, A. Perez-Aytes, E. Gean, P. Lapunzina, H. Peters, J. A. Goodship, and V. L. Ruiz-Perez (2013). The ciliary EVC/EVC2 complex interacts with Smo and controls Hedgehog pathway activity in chondrocytes by regulating Sufu/Gli3 dissociation and Gli3 trafficking in primary cilia. *Human Molecular Genetics 22*(1), 124–139.

Carlson, M. O., D. P. Rice, J. J. Berg, and M. Steinrücken (2021). Polygenic score accuracy in ancient samples: quantifying the effects of allelic turnover. *bioRxiv*.

Cavazos, T. B. and J. S. Witte (2021). Inclusion of variants discovered from diverse populations improves polygenic risk score transferability. *Human Genetics and Genomics Advances 2*(1), 100017.

Chakraborty, R. (1990). Quantitative Traits in Relation to Population Structure: Why and How Are They Used and What Do They Imply? *Human Biology 62*(1), 147–162.

Chakraborty, R. and M. Nei (1982). Genetic differentiation of quantitative characters between populations or species I: Mutation and random genetic drift. *Genetical Research 39*, 303–314.

Clark, B. W., C. W. Matson, D. Jung, and R. T. Di Giulio (2010). Ahr2 mediates cardiac teratogenesis of polycyclic aromatic hydrocarbons and pcb-126 in atlantic killifish (fundulus heteroclitus). *Aquatic Toxicology 99*(2), 232–240.

Cohan, F. M. (1984). Can uniform selection retard random genetic divergence between isolated conspecific populations? *Evolution 38*(3), 495–504.

Collin, R. W., K. W. Littink, B. J. Klevering, L. I. van den Born, R. K. Koenekoop, M. N. Zonneveld, E. A. Blokland, T. M. Strom, C. B. Hoyng, A. I. den Hollander, and F. P. Cremers (2008). Identification of a 2 Mb human ortholog of Drosophila eyes shut/spacemaker that is mutated in patients with Retinitis Pigmentosa. *The American Journal of Human Genetics 83*(5), 594–603.

Conrad, D. F., M. Jakobsson, G. Coop, X. Wen, J. D. Wall, N. A. Rosenberg, and J. K. Pritchard (2006). A worldwide survey of haplotype variation and linkage disequilibrium in the human genome. *Nature Genetics 38*(11), 1251–1260.

Conti, D. V., B. F. Darst, L. C. Moss, E. J. Saunders, X. Sheng, A. Chou, F. R. Schumacher, A. A. A. Olama, S. Benlloch, T. Dadaev, M. N. Brook, A. Sahimi, T. J. Hoffmann, A. Takahashi, K. Matsuda, Y. Momozawa, M. Fujita, K. Muir, A. Lophatananon, P. Wan, L. Le Marchand, L. R. Wilkens, V. L. Stevens, S. M. Gapstur, B. D. Carter, J. Schleutker, T. L. Tammela, C. Sipeky, A. Auvinen, G. G. Giles, M. C. Southey, R. J. MacInnis, C. Cybulski, D. Wokołorczyk, J. Lubiński, D. E. Neal, J. L. Donovan, F. C. Hamdy, R. M. Martin, B. G. Nordestgaard, S. F. Nielsen, M. Weischer, S. E. Bojesen, M. A. Røder, P. Iversen, J. Batra, S. Chambers, L. Moya, L. Horvath, J. A. Clements, W. Tilley, G. P. Risbridger, H. Gronberg, M. Aly, R. Szulkin, M. Eklund, T. Nordström, N. Pashayan, A. M. Dunning, M. Ghoussaini, R. C. Travis, T. J. Key, E. Riboli, J. Y. Park, T. A. Sellers, H. Y. Lin, D. Albanes, S. J. Weinstein, L. A. Mucci, E. Giovannucci, S. Lindstrom, P. Kraft, D. J. Hunter, K. L. Penney, C. Turman, C. M. Tangen, P. J. Goodman, I. M. Thompson, R. J. Hamilton, N. E. Fleshner, A. Finelli, M. É. Parent, J. L. Stanford, E. A. Ostrander, M. S. Geybels, S. Koutros, L. E. Freeman, M. Stampfer, A. Wolk, N. Håkansson, G. L. Andriole, R. N. Hoover, M. J. Machiela, K. D. Sørensen, M. Borre, W. J. Blot, W. Zheng, E. D. Yeboah, J. E. Mensah, Y. J. Lu, H. W. Zhang, N. Feng, X. Mao, Y. Wu, S. C. Zhao, Z. Sun, S. N. Thibodeau, S. K. McDonnell, D. J. Schaid, C. M. West, N. Burnet, G. Barnett, C. Maier, T. Schnoeller, M. Luedeke, A. S. Kibel, B. F. Drake, O. Cussenot, G. Cancel-Tassin, F. Menegaux, T. Truong, Y. A. Koudou, E. M. John,

E. M. Grindedal, L. Maehle, K. T. Khaw, S. A. Ingles, M. C. Stern, A. Vega, A. Gómez-Caamaño, L. Fachal, B. S. Rosenstein, S. L. Kerns, H. Ostrer, M. R. Teixeira, P. Paulo, A. Brandão, S. Watya, A. Lubwama, J. T. Bensen, E. T. Fontham, J. Mohler, J. A. Taylor, M. Kogevinas, J. Llorca, G. Castaño-Vinyals, L. Cannon-Albright, C. C. Teerlink, C. D. Huff, S. S. Strom, L. Multigner, P. Blanchet, L. Brureau, R. Kaneva, C. Slavov, V. Mitev, R. J. Leach, B. Weaver, H. Brenner, K. Cuk, B. Holleczek, K. U. Saum, E. A. Klein, A. W. Hsing, R. A. Kittles, A. B. Murphy, C. J. Logothetis, J. Kim, S. L. Neuhausen, L. Steele, Y. C. Ding, W. B. Isaacs, B. Nemesure, A. J. Hennis, J. Carpten, H. Pandha, A. Michael, K. De Ruyck, G. De Meerleer, P. Ost, J. Xu, A. Razack, J. Lim, S. H. Teo, L. F. Newcomb, D. W. Lin, J. H. Fowke, C. Neslund-Dudas, B. A. Rybicki, M. Gamulin, D. Lessel, T. Kulis, N. Usmani, S. Singhal, M. Parliament, F. Claessens, S. Joniau, T. Van den Broeck, M. Gago-Dominguez, J. E. Castelao, M. E. Martinez, S. Larkin, P. A. Townsend, C. Aukim-Hastie, W. S. Bush, M. C. Aldrich, D. C. Crawford, S. Srivastava, J. C. Cullen, G. Petrovics, G. Casey, M. J. Roobol, G. Jenster, R. H. van Schaik, J. J. Hu, M. Sanderson, R. Varma, R. McKean-Cowdin, M. Torres, N. Mancuso, S. I. Berndt, S. K. Van Den Eeden, D. F. Easton, S. J. Chanock, M. B. Cook, F. Wiklund, H. Nakagawa, J. S. Witte, R. A. Eeles, Z. Kote-Jarai, and C. A. Haiman (2021). Trans-ancestry genome-wide association meta-analysis of prostate cancer identifies new susceptibility loci and informs genetic risk prediction. *Nature Genetics 53*(1), 65–75.

Coop, G. (2019). Reading tea leaves? polygenic scores and differences in traits among groups. *arXiv preprint arXiv:1909.00892*.

Coop, G., J. K. Pickrell, J. Novembre, S. Kudaravalli, J. Li, R. M. Myers, L. L. Cavalli-sforza, M. W. Feldman, and J. K. Pritchard (2009). The Role of Geography in Human Adaptation. *PLoS Genetics 5*(6), e1000500.

Coop, G., D. Witonsky, A. Di Rienzo, and J. K. Pritchard (2010). Using environmental correlations to identify loci underlying local adaptation. *Genetics 185*, 1411–1423.

Coram, M. A., H. Fang, S. I. Candille, T. L. Assimes, and H. Tang (2017). Leveraging Multi-ethnic Evidence for Risk Assessment of Quantitative Traits in Minority Populations. *The American Journal of Human Genetics 101*(2), 218–226.

Cox, S. L., H. M. Moots, J. T. Stock, A. Shbat, B. D. Bitarello, N. Nicklisch, K. W. Alt, W. Haak, E. Rosenstock, C. B. Ruff, and I. Mathieson (2022). Predicting skeletal stature using ancient dna. *American Journal of Biological Anthropology 177*(1), 162–174.

Cox, S. L., C. B. Ruff, R. M. Maier, and I. Mathieson (2019). Genetic contributions to variation in human stature in prehistoric Europe. *Proceedings of the National Academy of Sciences 116*(43), 201910606.

Curtis, D. (2018). Polygenic risk score for schizophrenia is more strongly associated with ancestry than with schizophrenia. *Psychiatric Genetics 28*(5), 85–89.

Dannemann, M., A. M. Andrés, and J. Kelso (2016). Introgression of Neandertal- and Denisovan-like Haplotypes Contributes to Adaptive Variation in Human Toll-like Receptors. *The American Journal of Human Genetics 98*, 22–33.

Dannemann, M., K. Prüfer, and J. Kelso (2017). Functional implications of Neandertal introgression in modern humans. *Genome Biology 18*(61).

de Villemereuil, P., A. Charmantier, D. Arlt, P. Bize, P. Brekke, L. Brouwer, A. Cockburn, S. D. Côté, F. Stephen Dobson, S. R. Evans, M. Festa-Bianchet, M. Gamelon, S. Hamel, J. Hegelbach, K. Jerstad, B. Kempenaers, L. E. Kruuk, J. Kumpula, T. Kvalnes, A. G. McAdam, S. Eryn McFarlane, M. B. Morrissey, T. Pärt, J. M. Pemberton, A. Qvarnström, O. W. Røstad, J. Schroeder, J. C. Senar, B. C. Sheldon, M. van de Pol, M. E. Visser, N. T. Wheelwright, J. Tufto, and L. M. Chevin (2020). Fluctuating optimum and temporally variable selection on breeding date in birds and mammals. *Proceedings of the National Academy of Sciences 117*(50), 31969–31978.

Deschamps, M., G. Laval, M. Fagny, Y. Itan, L. Abel, J.-L. Casanova, E. Patin, and L. Quintana-Murci (2016, jan). Genomic Signatures of Selective Pressures and Introgression from Archaic Hominins at Human Innate Immunity Genes. *American Journal of Human Genetics 98*, 5–21.

Duncan, L., H. Shen, B. Gelaye, J. Meijsen, K. Ressler, M. Feldman, R. Peterson, and B. Domingue (2019). Analysis of polygenic risk score usage and performance in diverse human populations. *Nature Communications 10*.

Durvasula, A. and K. E. Lohmueller (2021). Negative selection on complex traits limits phenotype prediction accuracy between populations. *The American Journal of Human Genetics 108*(4), 620–631.

Edge, M. D. and N. A. Rosenberg (2015). A general model of the relationship between the apportionment of human genetic diversity and the apportionment of human phenotypic diversity. *Human Biology 87*(4), 313–337.

Ehmsen, J. T., Y. Liu, Y. Wang, N. Paladugu, A. E. Johnson, J. D. Rothstein, S. Du Lac, M. P. Mattson, and A. Höke (2016). The astrocytic transporter SLC7A10 (Asc-1) mediates glycinergic inhibition of spinal cord motor neurons. *Scientific Reports 6*(July), 1–13.

Enard, D. and D. A. Petrov (2018). Evidence that RNA Viruses Drove Adaptive Introgression between Neanderthals and Modern Humans. *Cell 175*, 360–371.

Ertel, E. A., K. P. Campbell, M. M. Harpold, F. Hoffman, Y. Mori, E. Perez-Reyes, A. Schwartz, T. P. Snutch, T. Tanabe, L. Birnbaumer, R. W. Tsien, and W. A. Catterall (2000). Nomenclature of Voltage-Gated Calcium Channels Voltage-gated. *Neuron 25*, 533–535.

Esteller-Cucala, P., I. Maceda, A. D. Børglum, D. Demontis, S. V. Faraone, B. Cormand, and O. Lao (2020). Genomic analysis of the natural history of attention-deficit/hyperactivity disorder using Neanderthal and ancient Homo sapiens samples. *Scientific Reports 10*(1).

Fan, S., M. E. Hansen, Y. Lo, and S. A. Tishkoff (2016). Going global by adapting local: A review of recent human adaptation. *Science 354*(6308), 54–59.

Feldman, M. W. and R. C. Lewontin (1975). The Heritability Hang-up. *Science 190*(4220), 1163–1168.

Fenner, J. N. (2005). Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. *American Journal of Physical Anthropology 128*, 415–423.

Fu, Q., C. Posth, M. Hajdinjak, M. Petr, S. Mallick, D. Fernandes, A. Furtwängler, W. Haak, M. Meyer, A. Mittnik, B. Nickel, A. Peltzer, N. Rohland, V. Slon, S. Talamo, I. Lazaridis, M. Lipson, I. Mathieson, S. Schiffels, P. Skoglund, A. P. Derevianko, N. Drozdov, V. Slavinsky, A. Tsybankov, R. G. Cremonesi, F. Mallegni, B. Gély, E. Vacca, M. R. G. Morales, L. G. Straus, C. Neugebauer-M, S. Pääbo, and D. Reich (2016). The genetic history of Ice Age Europe. *Nature 534*, 200–205.

Galinsky, K. J., Y. A. Reshef, H. K. Finucane, P. R. Loh, N. Zaitlen, N. J. Patterson, B. C. Brown, and A. L. Price (2019). Estimating cross-population genetic correlations of causal effect sizes. *Genetic Epidemiology 43*(2), 180–188.

Gingerich, P. D. (1983). Rates of evolution: Effects of time and temporal scaling. *Science 222*(4620), 159–161.

Gittelman, R. M., J. G. Schraiber, B. Vernot, C. Mikacenic, M. M. Wurfel, J. M. Akey Correspondence, and J. M. Akey (2016). Archaic Hominin Admixture Facilitated Adaptation to Out-of-Africa Environments. *Current Biology 26*, 3375–3382.

Goldstein, D. B. and K. E. Holsinger (1992). Maintenance of polygenic variation in spatially structured populations: roles for local mating and genetic redundancy. *Evolution 46*(2), 412–429.

Gras, C., B. Eiz-Vesper, A. Seltsam, S. Immenschuh, R. Blasczyk, and C. Figueiredo (2013). Semaphorin 7A protein variants differentially regulate T-cell activity. *Transfusion 53*(2), 270–283.

Greenbaum, G., W. M. Getz, N. A. Rosenberg, M. W. Feldman, E. Hovers, and O. Kolodny (2019). Disease transmission and introgression can explain the long-lasting contact zone of modern humans and Neanderthals. *Nature Communications 10*.

Grinde, K. E., T. A. Thornton, K. H. K. Chan, and A. P. Reiner (2019). Generalizing polygenic risk scores from Europeans to Hispanics / Latinos. *Genetic Epidemiology 43*(April 2018), 50–62.

Guo, H., G. Du, L. Wang, D. Wang, L. Hu, and Y. Huang (2013). Integrin alpha v beta 6 contributes to maintaining corneal epithelial barrier function. *Cell Biology International 37*(6), 593–599.

Haak, W., I. Lazaridis, N. Patterson, N. Rohland, S. Mallick, B. Llamas, G. Brandt, S. Nordenfelt, E. Harney, K. Stewardson, Q. Fu, A. Mittnik, E. Bánffy, C. Economou, M. Francken, S. Friederich, R. G. Pena, F. Hallgren, V. Khartanovich, A. Khokhlov, M. Kunst, P. Kuznetsov, H. Meller, O. Mochalov, V. Moiseyev, N. Nicklisch, S. L. Pichler, R. Risch, M. A. Rojo Guerra, C. Roth, A. Szécsényi-Nagy, J. Wahl, M. Meyer, J. Krause, D. Brown, D. Anthony, A. Cooper, K. W. Alt, and D. Reich (2015). Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature 522*, 207–211.

Haller, B. C., J. Galloway, J. Kelleher, P. W. Messer, and P. L. Ralph (2019). Tree-sequence recording in SLiM opens new horizons for forward-time simulation of whole genomes. *Molecular Ecology Resources 19*, 552–566.

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

Hamilton, J. A. and J. M. Miller (2016). Adaptive introgression as a resource for management and genetic conservation in a changing climate. *Conservation Biology 30*(1), 33–41.

Harpak, A. and M. Przeworski (2021). The evolution of group differences in changing environments. *PLoS Biology 19*(1), e3001072.

Harris, K. and R. Nielsen (2016). The Genetic Cost of Neanderthal Introgression. *Genetics 203*, 881–891.

Haworth, S., R. Mitchell, L. Corbin, K. H. Wade, T. Dudding, A. Budu-Aggrey, D. Carslake, G. Hemani, L. Paternoster, G. D. Smith, N. Davies, D. J. Lawson, and N. J. Timpson (2019). Apparent latent structure within the UK Biobank sample has implications for epidemiological analysis. *Nature Communications 10*(333).

Hayward, L. K. and G. Sella (2021). Polygenic adaptation after a sudden change in environment. *bioRxiv*.

He, B., Y. Chiba, H. Li, S. de Vega, K. Tanaka, K. Yoshizaki, M. Ishijima, K. Yuasa, M. Ishikawa, C. Rhodes, K. Sakai, P. Zhang, S. Fukumoto, X. Zhou, and Y. Yamada (2019). Identification of the Novel Tooth-Specific Transcription Factor AmeloD. *Journal of Dental Research 98*(2), 234–241.

He, D., L. Ma, R. Feng, L. Zhang, Y. Jiang, Y. Zhang, and G. Liu (2015). Analyzing large-scale samples highlights significant association between rs10411210 polymorphism and colorectal cancer. *Biomedicine and Pharmacotherapy 74*, 164–168.

Hernandez, R. D., J. L. Kelley, E. Elyashiv, S. C. Melton, A. Auton, G. McVean, . G. Project, G. Sella, and M. Przeworski (2011). Classic Selective Sweeps Were Rare in Recent Human Evolution. *Science 331*(February), 920–924.

Hezel, A. F., V. Deshpande, S. M. Zimmerman, G. Contino, B. Alagesan, M. R. O'Dell, L. B. Rivera, J. Harper, S. Lonning, R. A. Brekken, and N. Bardeesy (2012). TGF-β and αvβ6 integrin act in a common pathway to suppress pancreatic cancer progression. *Cancer Research 72*(18), 4840–4845.

Hicken, C. E., T. L. Linbo, D. H. Baldwin, M. L. Willis, M. S. Myers, L. Holland, M. Larsen, M. S. Stekoll, S. D. Rice, T. K. Collier, et al. (2011). Sublethal exposure to crude oil during embryonic development alters cardiac morphology and reduces aerobic capacity in adult fish. *Proceedings of the National Academy of Sciences 108*(17), 7086–7090.

Hill, W. G., M. E. Goddard, and P. M. Visscher (2008, 02). Data and theory point to mainly additive genetic variance for complex traits. *PLOS Genetics 4*(2), 1–10.

Hill, W. G. and M. Kirkpatrick (2010). What animal breeding has taught us about evolution. *Annual Review of Ecology, Evolution, and Systematics 41*, 1–19.

Horikoshi, M., T. Sofer, A. Mahajan, H. Kitajima, N. Franceschini, M. I. Mccarthy, K. Consortium, T.-g. Consortium, and A. P. Morris (2017). Trans-ethnic meta-regression of genome-wide association studies accounting for ancestry increases power for discovery and improves fine-mapping resolution. *Human Molecular Genetics 26*(18), 3639–3650.

Houle, D., G. H. Bolstad, K. Van Der Linde, and T. F. Hansen (2017). Mutation predicts 40 million years of fly wing evolution. *Nature 548*(7668), 447–450.

Huang, L., Y. Li, A. B. Singleton, J. A. Hardy, G. Abecasis, N. A. Rosenberg, and P. Scheet (2009). Genotype-imputation accuracy across worldwide human populations. *American Journal of Human Genetics 84*, 235–250.

Hublin, J. J. (2009). The origin of Neandertals. *Proceedings of the National Academy of Sciences 106*(38), 16022–16027.

Huerta-Sánchez, E., X. Jin, Z. Bianba, B. M. Peter, N. Vinckenbosch, Y. Liang, X. Yi, M. He, M. Somel, P. Ni, et al. (2014). Altitude adaptation in tibetans caused by introgression of denisovan-like dna. *Nature 512*(7513), 194–197.

Hui, R., E. D'Atanasio, L. M. Cassidy, C. L. Scheib, and T. Kivisild (2020). Evaluating genotype imputation pipeline for ultra-low coverage ancient genomes. *Scientific Reports 10*, 1–8.

Irving-Pease, E. K., R. Muktupavela, M. Dannemann, and F. Racimo (2021). Quantitative Human Paleogenetics: what can ancient DNA tell us about complex trait evolution? *Frontiers in Genetics 12*(August), 703541.

Isshiki, M., Y. Watanabe, and J. Ohashi (2021). Geographic variation in the polygenic score of height in Japan. *Human Genetics 140*(7), 1097–1108.

Jagoda, E., D. J. Lawson, J. D. Wall, D. Lambert, C. Muller, M. Westaway, M. Leavesley, T. D. Capellini, M. M. Lahr, P. Gerbault, M. G. Thomas, A. B. Migliano, E. Willerslev, M. Metspalu, and L. Pagani (2018). Disentangling immediate adaptive introgression from selection on standing introgressed variation in humans. *Molecular Biology and Evolution 35*(3), 623–630.

Jones, M. R., L. S. Mills, P. C. Alves, C. C. M, J. M. Alves, D. J. R. Lafferty, F. M. Jiggins, J. D. Jensen, J. Melo-Ferreira, and J. M. Good (2018). Adaptive introgression underlies polymorphic seasonal camouflage in snowshoe hares. *Science 360*(6395), 1355–1358.

Jones, M. R., L. S. Mills, J. D. Jensen, and J. M. Good (2020). The origin and spread of locally adaptive seasonal camouflage in snowshoe hares. *The American Naturalist 196*.

Jorde, L. B., W. S. Watkins, M. J. Bamshad, M. Dixon, C. Ricker, M. Seielstad, and M. A. Batzer (2000). The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and y-chromosome data. *The American Journal of Human Genetics 66*(3), 979–988.

Juric, I., S. Aeschbacher, and G. Coop (2016). The Strength of Selection against Neanderthal Introgression. *PLOS Genetics 12*(11).

Katano-Toki, A., T. Satoh, T. Tomaru, S. Yoshino, T. Ishizuka, S. Ishii, A. Ozawa, N. Shibusawa, T. Tsuchiya, T. Saito, H. Shimizu, K. Hashimoto, S. Okada, M. Yamada, and M. Mori (2013). THRAP3 interacts with HELZ2 and plays a novel role in adipocyte differentiation. *Molecular Endocrinology 27*(5), 769–780.

Keightley, P. D. and W. G. Hill (1988). Quantitative genetic variability maintained by mutation-stabilizing selection balance in finite populations. *Genetical Research 52*(1), 33–43.

Kelleher, J., A. M. Etheridge, and G. McVean (2016). Efficient Coalescent Simulation and Genealogical Analysis for Large Sample Sizes. *PLoS Computational Biology 12*(5), 1–22.

Kelleher, J., K. R. Thornton, J. Ashander, and P. L. Ralph (2018). Efficient pedigree recording for fast population genetics simulation. *PLoS Computational Biology 14*(11), 1–21.

Kerminen, S., A. R. Martin, J. Koskela, S. E. Ruotsalainen, A. S. Havulinna, I. Surakka, A. Palotie, M. Perola, V. Salomaa, M. J. Daly, S. Ripatti, and M. Pirinen (2019). Geographic Variation and Bias in the Polygenic Scores of Complex Diseases and Traits in Finland. *American Journal of Human Genetics 104*(6), 1169–1181.

Khrameeva, E. E., K. Bozek, L. He, Z. Yan, X. Jiang, Y. Wei, K. Tang, M. S. Gelfand, K. Prufer, J. Kelso, S. Paabo, P. Giavalisco, M. Lachmann, and P. Khaitovich (2014). Neanderthal ancestry drives evolution of lipid catabolism in contemporary Europeans. *Nature Communications 5*(3584).

Kilian, J. V., R. J. Klauda, S. Widman, M. Kashiwagi, R. Bourquin, S. Weglein, and J. Schuster (2012). An assessment of a bait industry and angler behavior as a vector of invasive species. *Biological Invasions 14*(7), 1469–1481.

Kim, M. S., K. P. Patel, A. K. Teng, A. J. Berens, and J. Lachance (2018). How genetic disease risks can be misestimated across global populations. *Genome Biology 19*(179).

Kingsolver, J. G., H. E. Hoekstra, J. M. Hoekstra, D. Berrigan, S. N. Vignieri, C. Hill, A. Hoang, P. Gibert, and P. Beerli (2001). The strength of phenotypic selection in natural populations. *The American Naturalist 157*(3), 245–261.

Koh, J. M., B. Oh, J. Y. Lee, J. K. Lee, K. Kimm, G. S. Kim, B. L. Park, H. S. Cheong, H. D. Shin, J. M. Hong, T. H. Kim, E. K. Park, and S. Y. Kim (2006). Association study

of semaphorin 7a (sema7a) polymorphisms with bone mineral density and fracture risk in postmenopausal Korean women. *Journal of Human Genetics 51*(2), 112–117.

Kremer, A. and V. Le Corre (2012). Decoupling of differentiation between traits and their underlying genes in response to divergent selection. *Heredity 108*(4), 375–385.

Lam, M., C.-y. Chen, Z. Li, A. R. Martin, and J. Bryois (2019). Comparative genetic architectures of schizophrenia in East Asian and European populations. *Nature Genetics 51*(December).

Lande, R. (1976). Natural selection and random genetic drift in phenotypic evolution. *Evolution 30*, 314–334.

Lande, R. (1991). Isolation by distance in a quantitative trait. *Genetics 128*(2), 443–452.

Lande, R. (1992). Neutral Theory of Quantitative Genetic Variance in an Island Model with Local Extinction and Colonization. *46*(2), 381–389.

Lande, R. and S. J. Arnold (1983). The measurement of selection on correlated characters. *Evolution 37*, 1210–1226.

Latta, R. G. (1998). Differentiation of allelic frequencies at quantitative trait loci affecting locally adaptive traits. *American Naturalist 151*(3), 283–292.

Lazaridis, I., N. Patterson, A. Mittnik, G. Renaud, S. Mallick, K. Kirsanow, P. H. Sudmant, J. G. Schraiber, S. Castellano, M. Lipson, B. Berger, C. Economou, R. Bollongino, Q. Fu, K. I. Bos, S. Nordenfelt, H. Li, C. De Filippo, K. Prüfer, S. Sawyer, C. Posth, W. Haak, F. Hallgren, E. Fornander, N. Rohland, D. Delsate, M. Francken, J. M. Guinet, J. Wahl, G. Ayodo, H. A. Babiker, G. Bailliet, E. Balanovska, O. Balanovsky, R. Barrantes, G. Bedoya, H. Ben-Ami, J. Bene, F. Berrada, C. M. Bravi, F. Brisighelli, G. B. Busby, F. Cali, M. Churnosov, D. E. Cole, D. Corach, L. Damba, G. Van Driem, S. Dryomov, J. M. Dugoujon, S. A. Fedorova, I. Gallego Romero, M. Gubina, M. Hammer, B. M. Henn, T. Hervig, U. Hodoglugil, A. R. Jha, S. Karachanak-Yankova, R. Khusainova, E. Khusnutdinova, R. Kittles, T. Kivisild, W. Klitz, V. Kučinskas, A. Kushniarevich, L. Laredj, S. Litvinov, T. Loukidis, R. W. Mahley, B. Melegh, E. Metspalu, J. Molina, J. Mountain, K. Näkkäläjärvi, D. Nesheva, T. Nyambo, L. Osipova, J. Parik, F. Platonov, O. Posukh, V. Romano, F. Rothhammer, I. Rudan, R. Ruizbakiev, H. Sahakyan, A. Sajantila, A. Salas, E. B. Starikovskaya, A. Tarekegn, D. Toncheva, S. Turdikulova, I. Uktveryte, O. Utevska, R. Vasquez, M. Villena, M. Voevoda, C. A. Winkler, L. Yepiskoposyan, P. Zalloua, T. Zemunik, A. Cooper, C. Capelli, M. G. Thomas, A. Ruiz-Linares, S. A. Tishkoff, L. Singh, K. Thangaraj, R. Villems, D. Comas, R. Sukernik, M. Metspalu, M. Meyer,

E. E. Eichler, J. Burger, M. Slatkin, S. Pääbo, J. Kelso, D. Reich, and J. Krause (2014). Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature 513*(7518), 409–413.

Le Corre, V. and A. Kremer (2003). Genetic variability at neutral markers, quantitative trait land trait in a subdivided population under selection. *Genetics 1219*(July), 1205–1219.

Lee, K. M. and G. Coop (2017). Distinguishing Among Modes of Convergent Adaptation Using Population Genomic Data. *Genetics 207*, 1591–1619.

Lewontin, R. C. (1972). The Apportionment of Human Diversity. In T. Dobzhansky, M. K. Hecht, and W. C. Steere (Eds.), *Evolutionary Biology*, Chapter 14, pp. 381–398. New York: Appleton-Century-Crofts.

Lewontin, R. C. (1974). The analysis of variance and the analysis of causes. *American Journal of Human Genetics 26*, 400–411.

Lewontin, R. C. and J. Krakauer (1973). Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics 74*(1), 175–195.

Li, J. Z., D. M. Absher, H. Tang, A. M. Southwick, A. M. Casto, S. Ramachandran, H. M. Cann, G. S. Barsh, M. Feldman, L. L. Cavalli-Sforza, et al. (2008). Worldwide human relationships inferred from genome-wide patterns of variation. *Science 319*(5866), 1100–1104.

Li, Y. R. and B. J. Keating (2014). Trans-ethnic genome-wide association studies: advantages and challenges of mapping in diverse populations. *Genome Medicine 6*.

Lipson, M., A. Szécsényi-Nagy, S. Mallick, A. Pósa, B. Stégmár, V. Keerl, N. Rohland, K. Stewardson, M. Ferry, M. Michel, J. Oppenheimer, N. Broomandkhoshbacht, E. Harney, S. Nordenfelt, B. Llamas, G. Mende, K. Köhler, K. Oross, M. Bondár, T. Marton, A. Osztás, J. Jakucs, T. Paluch, F. Horváth, P. Csengeri, J. Koós, K. Sebők, A. Anders, P. Raczky, J. Regenye, J. P. Barna, S. Fábián, G. Serlegi, Z. Toldi, E. G. Nagy, A. Mujika-Alustiza, C. A. Fernández, J. J. Echevarría, R. Bollongino, J. Orschiedt, K. Schierhold, H. Meller, A. Cooper, J. Burger, E. Bánffy, K. W. Alt, and D. Reich (2017). Parallel palaeogenomic transects reveal complex genetic history of early European farmers. *Nature 551*.

Liu, J. Z., S. Van Sommeren, H. Huang, S. C. Ng, R. Alberts, A. Takahashi, S. Ripke, J. C. Lee, L. Jostins, T. Shah, S. Abedian, J. H. Cheon, J. Cho, N. E. Daryani, L. Franke, Y. Fuyuno, A. Hart, R. C. Juyal, G. Juyal, W. H. Kim, A. P. Morris, H. Poustchi, W. G.

Newman, V. Midha, T. R. Orchard, H. Vahedi, A. Sood, J. J. Sung, R. Malekzadeh, H. J. Westra, K. Yamazaki, S. K. Yang, J. C. Barrett, A. Franke, B. Z. Alizadeh, M. Parkes, B. K. Thelma, M. J. Daly, M. Kubo, C. A. Anderson, and R. K. Weersma (2015). Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nature Genetics 47*(9), 979–986.

Loh, P. R., G. Bhatia, A. Gusev, H. K. Finucane, B. K. Bulik-Sullivan, S. J. Pollack, T. R. De Candia, S. H. Lee, N. R. Wray, K. S. Kendler, M. C. O'Donovan, B. M. Neale, N. Patterson, and A. L. Price (2015). Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. *Nature Genetics 47*(12), 1385–1392.

Macarthur, J., E. Bowler, M. Cerezo, L. Gil, P. Hall, E. Hastings, H. Junkins, A. Mcmahon, A. Milano, J. Morales, Z. M. Pendlington, D. Welter, T. Burdett, L. Hindorff, P. Flicek, F. Cunningham, and H. Parkinson (2017). The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucelic Acids Research 45*, 896–901.

Mafessoni, F., S. Grote, C. de Filippo, V. Slon, K. A. Kolobova, B. Viola, S. V. Markin, M. Chintalapati, S. Peyrégne, L. Skov, P. Skoglund, A. I. Krivoshapkin, A. P. Derevianko, M. Meyer, J. Kelso, B. Peter, K. Prüfer, and S. Pääbo (2020). A high-coverage Neandertal genome from Chagyrskaya Cave. *Proceedings of the National Academy of Sciences 117*(26), 15132–15136.

Marciniak, S., C. M. Bergey, A. M. Silva, A. Hałuszko, M. Furmanek, B. Veselka, P. Velemínský, G. Vercellotti, J. Wahl, G. Zarina, C. Longhi, J. Kolář, R. Garrido-Pena, R. Flores-Fernández, A. M. Herrero-Corral, A. Simalcsik, W. Müller, A. Sheridan, Ž. Miliauskiene, R. Jankauskas, V. Moiseyev, K. Köhler, Á. Király, B. Gamarra, O. Cheronet, V. Szeverényi, V. Kiss, T. Szeniczey, K. Kiss, Z. K. Zoffmann, J. Koós, M. Hellebrandt, L. Domboróczki, C. Virag, M. Novak, D. Reich, T. Hajdu, N. von Cramon-Taubadel, R. Pinhasi, and G. H. Perry (2021). An integrative skeletal and paleogenomic analysis of prehistoric stature variation suggests relatively reduced health for early european farmers. *bioRxiv*.

Marlow, R., P. Strickland, S. L. Ji, X. Wu, M. PeBenito, M. Binnewies, E. K. Le, A. Moran, H. Macias, R. D. Cardiff, S. Sukumar, and L. Hinck (2008). SLITs suppress tumor growth in vivo by silencing Sdf1/Cxcr4 within breast epithelium. *Cancer Research 68*(19), 7819–7827.

Márquez-Luna, C. (2017). Multiethnic polygenic risk scores improve risk prediction in diverse populations. *Genetic Epidemiology 41*(April), 811–823.

Martin, A. R., C. R. Gignoux, R. K. Walters, G. L. Wojcik, B. M. Neale, S. Gravel, M. J. Daly, and C. D. Bustamante (2017). Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. *The American Journal of Human Genetics 100*(4), 635–649.

Martin, A. R., M. Kanai, Y. Kamatani, Y. Okada, B. M. Neale, and M. J. Daly (2019). Clinical use of current polygenic risk scores may exacerbate health disparities. *Nature Genetics 51*(4), 584–591.

Martin, A. R., M. Lin, J. M. Granka, J. W. Myrick, X. Liu, A. Sockell, E. G. Atkinson, C. J. Werely, M. Möller, M. S. Sandhu, D. M. Kingsley, E. G. Hoal, X. Liu, M. J. Daly, M. W. Feldman, C. R. Gignoux, C. D. Bustamante, and B. M. Henn (2017). An Unexpectedly Complex Architecture for Skin Pigmentation in Africans. *Cell 171*, 1340–1353.

Martiniano, R., L. M. Cassidy, R. Ó'Maoldúin, R. McLaughlin, N. M. Silva, L. Manco, D. Fidalgo, T. Pereira, M. J. Coelho, M. Serra, J. Burger, R. Parreira, E. Moran, A. C. Valera, E. Porfirio, R. Boaventura, A. M. Silva, and D. G. Bradley (2017). The population genomics of archaeological transition in west Iberia: Investigation of ancient substructure using imputation and haplotype-based methods. *PLoS Genetics 13*(7), 1–24.

Maruyama, T. and M. Kimura (1975). Moments for Sum of an Arbitrary Function of Gene Frequency along a Stochastic Path of Gene Frequency Change. *Proceedings of the National Academy of Sciences 72*(4), 1602–1604.

Mathieson, I. (2016, Nov). Imputation for ancient data.

Mathieson, I. (2021). The omnigenic model and polygenic prediction of complex traits. *The American Journal of Human Genetics 108*, 1–6.

Mathieson, I., S. Alpaslan-Roodenberg, C. Posth, A. Szécsényi-Nagy, N. Rohland, S. Mallick, I. Olalde, N. Broomandkhoshbacht, F. Candilio, O. Cheronet, D. Fernandes, M. Ferry, B. Gamarra, G. González Fortes, W. Haak, E. Harney, E. Jones, D. Keating, B. Krause-Kyora, I. Kucukkalipci, M. Michel, A. Mittnik, K. Nägele, M. Novak, J. Oppenheimer, N. Patterson, S. Pfrengle, K. Sirak, K. Stewardson, S. Vai, S. Alexandrov, K. W. Alt, R. Andreescu, D. Antonovic, A. Ash, N. Atanassova, K. Bacvarov, M. Balázs Gusztáv, H. Bocherens, M. Bolus, A. Boroneant, Y. Boyadzhiev, A. Budnik, J. Burmaz, S. Chohadzhiev, N. J. Conard, R. Cottiaux, M. C˘ uka, C. Cupillard, D. G. Drucker, N. Elenski, M. Francken, B. Galabova, G. Ganetsovski, B. Gély, T. Hajdu, V. Handzhyiska, K. Harvati, T. Higham, S. Iliev, I. Jankovic, I. Karavanic, D. J. Kennett, D. Komšo, A. Kozak, D. Labuda, M. Lari, C. Lazar, M. Leppek, K. Leshtakov, D. Lo Vetro, D. Los, I. Lozanov, M. Malina, F. Martini, K. McSweeney, H. Meller,

M. Mendušic, P. Mirea, V. Moiseyev, V. Petrova, T. Douglas Price, A. Simalcsik, L. Sineo, M. Šlaus, V. Slavchev, P. Stanev, A. Starovic, T. Szeniczey, S. Talamo, M. Teschler-Nicola, C. Thevenet, I. Valchev, F. Valentin, S. Vasilyev, F. Veljanovska, S. Venelinova, E. Veselovskaya, B. Viola, C. Virag, J. Zaninovic, S. Zäuner, P. W. Stockhammer, G. Catalano, R. Krauß, D. Caramelli, G. Zarin, B. Gaydarska, M. Lillie, A. G. Nikitin, I. Potekhina, A. Papathanasiou, D. Boric, C. Bonsall, J. Krause, R. Pinhasi, and D. Reich (2018). The genomic history of southeastern Europe. *Nature 555*, 197–203.

Mathieson, I., I. Lazaridis, N. Rohland, S. Mallick, N. Patterson, S. A. Roodenberg, E. Harney, K. Stewardson, D. Fernandes, M. Novak, K. Sirak, C. Gamba, E. R. Jones, B. Llamas, S. Dryomov, J. Pickrell, J. L. Arsuaga, J. M. B. de Castro, E. Carbonell, F. Gerritsen, A. Khokhlov, P. Kuznetsov, M. Lozano, H. Meller, O. Mochalov, V. Moiseyev, M. A. R. Guerra, J. Roodenberg, J. M. Vergès, J. Krause, A. Cooper, K. W. Alt, D. Brown, D. Anthony, C. Lalueza-Fox, W. Haak, R. Pinhasi, and D. Reich (2015). Genome-wide patterns of selection in 230 ancient Eurasians. *Nature 528*, 499–503.

Mathieson, S. and I. Mathieson (2018). FADS1 and the Timing of Human Adaptation to Agriculture. *Molecular Biology and Evolution 35*(12), 2957–2970.

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

Mendez, F. L., J. C. Watkins, and M. F. Hammer (2012). Global Genetic Variation at OAS1 Provides Evidence of Archaic Admixture in Melanesian Populations. *Molecular Biology and Evolution 29*(6), 1513–1520.

Messer, P. W., S. P. Ellner, and N. G. Hairston Jr (2016). Can population genetics adapt to rapid evolution? *Trends in Genetics 32*(7), 408–418.

Mesuret, G., S. Khabbazzadeh, A. M. Bischoff, H. Safory, H. Wolosker, and S. Hülsmann (2018). A neuronal role of the Alanine-Serine-Cysteine-1 transporter (SLC7A10, Asc-1) for glycine inhibitory transmission and respiratory pattern. *Scientific Reports 8*(1), 1–8.

Morim, T., G. R. Bigg, P. M. Madeira, J. Palma, D. D. Duvernell, E. Gisbert, R. L. Cunha, and R. Castilho (2019). Invasion genetics of the mummichog (fundulus heteroclitus): recent anthropogenic introduction in iberia. *PeerJ 7*, e6155.

Mostafavi, H., A. Harpak, I. Agarwal, D. Conley, J. K. Pritchard, and M. Przeworski (2020). Variable prediction accuracy of polygenic scores within an ancestry group. *eLife 9*, e48376.

Nacci, D., L. Coiro, A. Kuhn, D. Champlin, W. Munns Jr, J. Specker, and K. Cooper (1998). Nondestructive indicator of ethoxyresorufin-o-deethylase activity in embryonic fish. *Environmental Toxicology and Chemistry: An International Journal 17*(12), 2481–2486.

Nakamura, F., K. Kumeta, T. Hida, T. Isono, Y. Nakayama, E. Kuramata-Matsuoka, N. Yamashita, Y. Uchida, K. I. Ogura, K. Gengyo-Ando, S. Mitani, T. Ogino, and Y. Goshima (2014). Amino- and carboxyl-terminal domains of Filamin-A interact with CRMP1 to mediate Sema3A signalling. *Nature Communications 5*(5325).

Narain, P. and R. Chakraborty (1987). Genetic differentiation of quantitative characters between populations or species II: Optimal selection in infinite populations. *Heredity 59*, 199–212.

Ng, L., A. K. Chow, J. H. Man, T. C. Yau, T. M. Wan, D. N. Iyer, V. H. Kwan, R. T. Poon, R. W. Pang, and W. L. Law (2018). Suppression of Slit3 induces tumor proliferation and chemoresistance in hepatocellular carcinoma through activation of GSK3β/β-catenin pathway. *BMC Cancer 18*(1), 1–13.

Nicholson, G., A. V. Smith, F. Jónsson, Ó. Gústafsson, K. Stefánsson, and P. Donnelly (2002). Assessing population differentiation and isolation from single-nucleotide polymorphism data. *Journal of the Royal Statistical Society B: Statistical Methodology 64*(4), 695–715.

Novembre, J. and N. H. Barton (2018). Tread lightly interpreting polygenic tests of selection. *Genetics 208*(4), 1351–1355.

Olson, M. V. (1999). When less is more: gene loss as an engine of evolutionary change. *The American Journal of Human Genetics 64*(1), 18–23.

Oziolor, E. M., J. N. Apell, Z. C. Winfield, J. A. Back, S. Usenko, and C. W. Matson (2018). Polychlorinated biphenyl (pcb) contamination in galveston bay, texas: Comparing concentrations and profiles in sediments, passive samplers, and fish. *Environmental Pollution 236*, 609–618.

Oziolor, E. M., E. Bigorgne, L. Aguilar, S. Usenko, and C. W. Matson (2014). Evolved resistance to pcb-and pah-induced cardiac teratogenesis, and reduced cyp1a activity in gulf killifish (fundulus grandis) populations from the houston ship channel, texas. *Aquatic Toxicology 150*, 210–219.

Oziolor, E. M., N. M. Reid, S. Yair, K. M. Lee, S. Guberman VerPloeg, P. C. Bruns, J. R. Shaw, A. Whitehead, and C. W. Matson (2019). Adaptive introgression enables evolutionary rescue from extreme environmental pollution. *Science 364*, 455–457.

Palaćin, M., E. Errasti-Murugarren, and A. Rosell (2016). Heteromeric amino acid transporters. In search of the molecular bases of transport cycle mechanisms. *Biochemical Society Transactions 44*(3), 745–752.

Pasterkamp, R. J., J. J. Peschon, M. K. Spriggs, and A. L. Kolodkin (2003). Semaphorin 7A promotes axon outgrowth through integrins and MAPKs. *Nature 424*(6947), 398–405.

Patel, R. A., S. A. Musharoff, J. P. Spence, H. Pimentel, C. Tcheandjieu, H. Mostafavi, N. Sinnott-Armstrong, S. L. Clarke, C. J. Smith, V. M. V. Program, P. P. Durda, K. D. Taylor, R. Tracy, Y. Liu, C. W. Johnson, F. Aguet, K. G. Ardlie, S. Gabriel, J. Smith, D. A. Nickerson, S. S. Rich, J. I. Rotter, P. S. Tsao, T. L. Assimes, and J. K. Pritchard (2021). Effect sizes of causal variants for gene expression and complex traits differ between populations. *bioRxiv*.

Petr, M., S. Pääbo, J. Kelso, and B. Vernot (2019). Limits of long-term selection against Neandertal introgression. *Proceedings of the National Academy of Sciences 116*(5), 1639–1644.

Popejoy, A. B. and S. M. Fullerton (2016, oct). Genomics is failing on diversity. *Nature 538*, 161–164.

Privé, F., H. Aschard, S. Carmi, L. Folkersen, C. Hoggart, P. F. O'Reilly, and B. J. Vilhjálmsson (2022). Portability of 245 polygenic scores when derived from the uk biobank and applied to 9 ancestry groups from the same cohort. *The American Journal of Human Genetics 109*(1), 12–23.

Prout, T. and J. S. F. Barker (1993). F Statistics in Drosophila buzzatii: Selection, Population Size and Inbreeding. *Genetics 134*, 369–375.

Prüfer, K., C. D. Filippo, S. Grote, F. Mafessoni, P. Korlevi, M. Hajdinjak, B. Vernot, L. Skov, P. Hsieh, S. Peyrégne, D. Reher, C. Hopfe, T. Maricic, Q. Fu, C. Theunert, R. Rogers, P. Skoglund, M. Dannemann, B. J. Nelson, F. M. Key, P. Rudan, Ž. Ku, L. V. Golovanova, V. B. Doronichev, N. Patterson, D. Reich, E. E. Eichler, M. Slatkin, M. H. Schierup, A. Andrés, J. Kelso, M. Meyer, and S. Pääbo (2017). A high-coverage Neandertal genome from Vindija Cave in Croatia. *Science 1887*.

Ptáček, L. J., R. Tawil, R. C. Griggs, A. G. Engel, R. B. Layzer, H. Kwieciński, P. G. McManis, L. Santiago, M. Moore, G. Fouad, P. Bradley, and M. F. Leppert (1994). Dihydropyridine receptor mutations cause Hypokalemic Periodic Paralysis. *Cell 77*, 863–868.

Pusapati, G. V., C. E. Hughes, K. V. Dorn, D. Zhang, P. Sugianto, L. Aravind, and R. Rohatgi (2014). EFCAB7 and IQCE regulate Hedgehog signaling by tethering the EVC-EVC2 complex to the base of primary cilia. *Developmental Cell 28*, 483–496.

Quach, H., M. Rotival, J. Pothlichet, J. Kelso, M. L. Albert, L. Quintana-Murci, H. lè ne Quach, M. Rotival, Y.-H. Eddie Loh, M. Dannemann, N. Zidane, G. Laval, E. Patin, C. Harmant, M. Lopez, M. Deschamps, N. Naffakh, D. Duffy, A. Coen, G. Leroux-Roels, F. Clé ment, A. Boland, and J.-F. ois Deleuze (2016). Genetic Adaptation and Neandertal Admixture Shaped the Immune System of Human Populations. *Cell 167*, 643–656.

Racimo, F., J. J. Berg, and J. K. Pickrell (2018). Detecting Polygenic Adaptation in Admixture Graphs. *Genetics 208*, 1565–1584.

Racimo, F., D. Marnetto, and E. Huerta-Sánchez (2017). Signatures of archaic adaptive introgression in present-day human populations. *Molecular Biology and Evolution 34*(2), 296–317.

Racimo, F., S. Sankararaman, R. Nielsen, and E. Huerta-Sánchez (2015). Evidence for archaic adaptive introgression in humans. *Nature Reviews Genetics 16*, 359–371.

Refoyo-Martínez, A., R. R. Da Fonseca, K. Halldórsdóttir, E. Árnason, T. Mailund, and F. Racimo (2019). Identifying loci under positive selection in complex population histories. *Genome Research 29*, 1506–1520.

Refoyo-Martínez, A., S. Liu, A. M. Jørgensen, X. Jin, A. Albrechtsen, A. R. Martin, and F. Racimo (2020). How robust are cross-population signatures of polygenic adaptation in humans? *bioRxiv*, 1–66.

Reid, N. M., D. A. Proestou, B. W. Clark, W. C. Warren, J. K. Colbourne, J. R. Shaw, S. I. Karchner, M. E. Hahn, D. Nacci, M. F. Oleksiak, et al. (2016). The genomic landscape of rapid repeated evolutionary adaptation to toxic pollution in wild fish. *Science 354*(6317), 1305–1308.

Relethford, J. H. and F. C. Lees (1982). The use of quantitative traits in the study of human population structure. *American Journal of Physical Anthropology 25*, 113–132.

Robertson, A. (1956). The effect of selection against extreme deviants based on deviation or on homozygosis. *Journal of Genetics 54*(2), 236–248.

Rogers, A. R. and H. C. Harpending (1983). Population structure and quantitative characters. *Genetics 105*(4), 985–1002.

Rose, G. (2001). Sick individuals and sick populations. *International Journal of Epidemiology 30*, 396–427–432.

Rosenberg, N. A., M. D. Edge, J. K. Pritchard, and M. W. Feldman (2019). Interpreting polygenic scores, polygenic adaptation, and human phenotypic differences. *Evolution, medicine, and public health 2019*(1), 26–34.

Rosenberg, N. A., J. K. Pritchard, J. L. Weber, H. M. Cann, K. K. Kidd, L. A. Zhivotovsky, and M. W. Feldman (2002). Genetic structure of human populations. *Science 298*(5602), 2381–2385.

Ruiz-Perez, V. L., S. E. Ide, T. M. Strom, B. Lorenz, D. Wilson, K. Woods, L. King, C. Francomano, P. Freisinger, S. Spranger, B. Marino, B. Dallapiccola, M. Wright, T. Meitinger, M. H. Polymeropoulos, and J. Goodship (2000). Mutations in a new gene in Ellis-van Creveld syndrome and Weyers acrodental dysostosis. *Nature Genetics 24*, 283–286.

Sachdeva, H. and N. H. Barton (2018). Replicability of introgression under linked, polygenic selection. *Genetics 210*, 1411–1427.

Sakaue, S., M. Kanai, Y. Tanigawa, J. Karjalainen, M. Kurki, S. Koshiba, A. Narita, T. Konuma, K. Yamamoto, M. Akiyama, K. Ishigaki, A. Suzuki, K. Suzuki, W. Obara, K. Yamaji, K. Takahashi, S. Asai, Y. Takahashi, T. Suzuki, N. Shinozaki, H. Yamaguchi, S. Minami, S. Murayama, K. Yoshimori, S. Nagayama, D. Obata, M. Higashiyama, A. Masumoto, Y. Koretsune, K. Ito, C. Terao, T. Yamauchi, I. Komuro, T. Kadowaki, G. Tamiya, M. Yamamoto, Y. Nakamura, M. Kubo, Y. Murakami, K. Yamamoto, Y. Kamatani, A. Palotie, M. Rivas, M. Daly, K. Matsuda, and Y. Okada (2020). A global atlas of genetic associations of 220 deep phenotypes. *medRxiv*, 1–52.

Samanta, S., Y.-J. Li, and B. S. Weir (2009). Drawing Inferences about the Coancestry Coefficient. *Theoretical Population Biology 75*(4), 312–319.

Sams, A. J., A. Dumaine, Y. Nédélec, V. Yotova, C. Alfieri, J. E. Tanner, P. W. Messer, and L. B. Barreiro (2016). Adaptively introgressed Neandertal haplotype at the OAS locus functionally impacts innate immune responses in humans. *Genome Biology 17*(246), 1–15.

Sanjak, J. S., J. Sidorenko, M. R. Robinson, K. R. Thornton, and P. M. Visscher (2018). Evidence of directional and stabilizing selection in contemporary humans. *Proceedings of the National Academy of Sciences 115*(1), 151–156.

Sankararaman, S., S. Mallick, M. Dannemann, K. Prüfer, J. Kelso, S. Pääbo, N. Patterson, and D. Reich (2014, jan). The genomic landscape of Neanderthal ancestry in present-day humans. *Nature 507*, 354–357.

Sawyer, S. A. and D. L. Hartl (1992). Population Genetics of Polymorphism and Divergence. *Genetics 132*(1), 1161–1176.

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

Schumer, M., C. Xu, D. L. Powell, A. Durvasula, L. Skov, C. Holland, J. C. Blazier, and S. Sankararaman (2018). Natural selection interacts with recombination to shape the evolution of hybrid genomes. *Science 360*, 656–660.

Scutari, M., I. Mackay, and D. Balding (2016, 09). Using genetic distance to infer the accuracy of genomic prediction. *PLOS Genetics 12*(9), 1–19.

Sella, G. and N. H. Barton (2019). Thinking about the evolution of complex traits in the era of genome-wide association studies. *Annual Review of Genomics and Human Genetics 20*(1), 461–493.

Setter, D., S. Mousset, X. Cheng, R. Nielsen, M. DeGiorgio, and J. Hermisson (2020). VolcanoFinder: Genomic scans for adaptive introgression. *PLoS Genetics 16*(6).

Shchur, V., J. Svedberg, P. Medina, R. Corbett-Detig, and R. Nielsen (2020). On the distribution of tract lengths during adaptive introgression. *G3*.

Shi, H., G. Kichaev, and B. Pasaniuc (2016). Contrasting the Genetic Architecture of 30 Complex Traits from Summary Association Data. *American Journal of Human Genetics 99*, 139–153.

Sikora, M., A. Seguin-orlando, V. C. Sousa, A. Albrechtsen, T. Korneliussen, A. Ko, S. Rasmussen, I. Dupanloup, P. R. Nigst, M. D. Bosch, G. Renaud, and M. E. Allentoft (2017). Ancient genomes show social and reproductive behavior of early Upper Paleolithic foragers. *Science 358*, 659–662.

Simberloff, D., J.-L. Martin, P. Genovesi, V. Maris, D. A. Wardle, J. Aronson, F. Courchamp, B. Galil, E. García-Berthou, M. Pascal, et al. (2013). Impacts of biological invasions: what's what and the way forward. *Trends in ecology & evolution 28*(1), 58–66.

Simons, Y. B., K. Bullaughey, R. R. Hudson, and G. Sella (2018). A population genetic interpretation of GWAS findings for human quantitative traits. *PLoS Biology 16*(3).

Simonti, C. N. and J. Lachance (2021). Ancient DNA reveals that few GWAS loci have been strongly selected during recent human history. *bioRxiv*.

Skoglund, P., H. Malmström, M. Raghavan, J. Storå, P. Hall, E. Willerslev, M. T. P. Gilbert, A. Götherström, and M. Jakobsson (2012). Origins and genetic legacy of neolithic farmers and hunter-gatherers in Europe. *Science 336*(6080), 466–469.

Sohail, M., R. M. Maier, A. Ganna, A. Bloemendal, A. R. Martin, M. C. Turchin, C. W. Chiang, J. Hirschhorn, M. J. Daly, N. Patterson, B. Neale, I. Mathieson, D. Reich, and S. R. Sunyaev (2019). Polygenic adaptation on height is overestimated due to uncorrected stratification in genome-wide association studies. *eLife 8*, e39702.

Spitze, K. (1993). Population structure in Daphnia obtusa: Quantitative genetic and allozymic variation. *Genetics 135*(2), 367–374.

Steinrücken, M., J. P. Spence, J. A. Kamm, E. Wieczorek, and Y. S. Song (2018). Model-based detection and analysis of introgressed Neanderthal ancestry in modern humans. *Molecular Ecology 27*, 3873–3888.

Suzuki, K., T. Okuno, M. Yamamoto, R. J. Pasterkamp, N. Takegahara, H. Takamatsu, T. Kitao, J. Takagi, P. D. Rennert, A. L. Kolodkin, A. Kumanogoh, and H. Kikutani (2007). Semaphorin 7A initiates T-cell-mediated inflammatory responses through $\alpha 1\beta 1$ integrin. *Nature 446*(7136), 680–684.

Tanabe, T., K. G. Beam, B. A. Adams, T. Niidome, and S. Numa (1990). Regions of the skeletal muscle dihydropyridine receptor critical for excitation- contraction coupling. *Nature 346*, 567–569.

Telis, N., R. Aguilar, and K. Harris (2020). Selection against archaic hominin genetic variation in regulatory regions. *Nature Ecology and Evolution*.

Teo, S. and K. Able (2003). Growth and production of the mummichog (fundulus heteroclitus) in a restored salt marsh. *Estuaries 26*(1), 51–63.

The 1000 Genomes Project Consortium (2015). A global reference for human genetic variation. *Nature 526*, 68–74.

The Heliconius Genome Consortium (2012). Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature 487*, 94–98.

The International HapMap Consortium (2007). A second generation human haplotype map of over 3.1 million SNPs. *Nature 449*, 851–861.

Trochet, H., J. Pelletier, R. Tadros, and J. Hussin (2021). Comparison of polygenic risk scores for heart disease highlights obstacles to overcome for clinical use. *bioRxiv*.

Turchin, M. C., C. W. Chiang, C. D. Palmer, S. Sankararaman, D. Reich, and J. N. Hirschhorn (2012). Evidence of widespread selection on standing variation in Europe at height-associated SNPs. *Nature Genetics 44*(9), 1015–1019.

Turelli, M. (1984). Heritable genetic variation via mutation-selection balance: Lerch's zeta meets the abdominal bristle. *Theoretical Population Biology 25*(2), 138–193.

Turelli, M. and N. Barton (1990). Dynamics of polygenic characters under selection. *Theoretical Population Biology 38*(1), 1–57.

Turelli, M. and N. H. Barton (1994, 11). Genetic and statistical analyses of strong selection on polygenic traits: what, me normal? *Genetics 138*(3), 913–941.

Uecker, H., D. Setter, and J. Hermisson (2015). Adaptive gene introgression after secondary contact. *Journal of Mathematical Biology 70*, 1523–1580.

Vernot, B. and J. M. Akey (2014). Resurrecting Surviving Neandertal Lineages from Modern Human Genomes. *Science 343*, 1017–1021.

Veturi, Y., N. Yi, W. Huang, and A. I. Vazquez (2019). Modeling Heterogeneity in the Genetic Architecture of Ethnically Diverse Groups Using Random Effect Interaction Models. *Genetics 211*(April), 1395–1407.

Vilhjálmsson, B. J., J. Yang, H. K. Finucane, A. Gusev, S. Ripke, G. Genovese, P.-r. Loh, G. Bhatia, S. Lindstro, R. Do, N. Zaitlen, B. Pasaniuc, G. Belbin, E. E. Kenny, M. H. Schierup, S. Purcell, D. Chasman, B. Neale, and M. Goddard (2015). Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *American Journal of Human Genetics 97*, 576–592.

Wang, Y., J. Guo, G. Ni, J. Yang, P. M. Visscher, and L. Yengo (2020). Theoretical and empirical quantification of the accuracy of polygenic scores in ancestry divergent populations. *Nature Communications 11*(1), 1–9.

Weber, K. E. (1990). Selection on wing allometry in Drosophila melanogaster. *Genetics 126*(4), 975–989.

Weir, B. S., N. Carolina, and W. G. Hill (2002). Estimating F Statistics. *Annual Review of Genetics 36*, 721–750.

Weissbrod, O., M. Kanai, H. Shi, S. Gazal, W. Peyrot, A. Khera, Y. Okada, T. Biobank, J. Project, A. Martin, H. Finucane, and A. L. Price (2021). Leveraging fine-mapping and non-European training data to improve trans-ethnic polygenic risk scores. *medRxiv*.

Whitehead, A., W. Pilcher, D. Champlin, and D. Nacci (2012). Common mechanism underlies repeated evolution of extreme pollution tolerance. *Proceedings of the Royal Society B: Biological Sciences 279*(1728), 427–433.

Whitlock, M. C. (1999). Neutral additive genetic variance in a metapopulation. *Genetics Research 74*(3), 215–221.

Whitney, K. D., R. A. Randell, and L. H. Rieseberg (2006). Adaptive Introgression of Herbivore Resistance Traits in the Weedy Sunflower Helianthus annuus. *The American Naturalist 167*, 794–807.

Wirgin, I., N. K. Roy, M. Loftus, R. C. Chambers, D. G. Franks, and M. E. Hahn (2011). Mechanistic basis of resistance to pcbs in atlantic tomcod from the hudson river. *Science 331*(6022), 1322–1325.

Wojcik, G. L., M. Graff, K. K. Nishimura, R. Tao, J. Haessler, C. R. Gignoux, H. M. Highland, Y. M. Patel, E. P. Sorokin, C. L. Avery, G. M. Belbin, S. A. Bien, I. Cheng, S. Cullina, C. J. Hodonsky, Y. Hu, L. M. Huckins, J. Jeff, A. E. Justice, J. M. Kocarnik, U. Lim, B. M. Lin, Y. Lu, S. C. Nelson, S. S. L. Park, H. Poisner, M. H. Preuss, M. A. Richard, C. Schurmann, V. W. Setiawan, A. Sockell, K. Vahi, M. Verbanck, A. Vishnu, R. W. Walker, K. L. Young, N. Zubair, V. Acuña-Alonso, J. L. Ambite, K. C. Barnes, E. Boerwinkle, E. P. Bottinger, C. D. Bustamante, C. Caberto, S. Canizales-Quinteros, M. P. Conomos, E. Deelman, R. Do, K. Doheny, L. Fernández-Rhodes, M. Fornage, B. Hailu, G. Heiss, B. M. Henn, L. A. Hindorff, R. D. Jackson, C. A. Laurie, C. C. Laurie, Y. Li, D. Y. Lin, A. Moreno-Estrada, G. Nadkarni, P. J. Norman, L. C. Pooler, A. P. Reiner, J. Romm, C. Sabatti, K. Sandoval, X. Sheng, E. A. Stahl, D. O. Stram, T. A. Thornton, C. L. Wassel, L. R. Wilkens, C. A. Winkler, S. Yoneyama, S. Buyske, C. A. Haiman, C. Kooperberg, L. Le Marchand, R. J. Loos, T. C. Matise, K. E. North, U. Peters, E. E. Kenny, and C. S. Carlson (2019). Genetic analyses of diverse populations improves discovery for complex traits. *Nature 570*(7762), 514–518.

Wonham, M. J., J. T. Carlton, G. M. Ruiz, and L. D. Smith (2000). Fish and ships: relating dispersal frequency to success in biological invasions. *Marine Biology 136*(6), 1111–1121.

Wood, A. R., T. Esko, J. Yang, S. Vedantam, T. H. Pers, and S. et al. Gustafsson (2014). Wood, Andrew R Esko, Tonu Yang, Jian Vedantam S, Pers TH, Gustafsson S, et al. *Nature Genetics 46*(11).

Wright, S. (1935). Evolution in populations in approximate equilibrium. *Journal of Genetics 30*(2), 257–266.

Wright, S. (1951). The Genetical Structure of Populations. *Annals of Eugenics 15*(4), 323–354.

Yang, J., A. Bakshi, Z. Zhu, G. Hemani, A. A. E. Vinkhuyzen, S. H. Lee, M. R. Robinson, J. R. B. Perry, I. M. Nolte, J. V. V. Vliet-ostaptchouk, H. Snieder, T. Lifelines, C. Study, T. Esko, L. Milani, R. Mägi, A. Metspalu, A. Hamsten, P. K. E. Magnusson, N. L. Pedersen, E. Ingelsson, N. Soranzo, and M. C. Keller (2015). Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nature Genetics 47*(10).

Yang, J., B. Benyamin, B. P. Mcevoy, S. Gordon, A. K. Henders, D. R. Nyholt, P. A. Madden, A. C. Heath, N. G. Martin, G. W. Montgomery, M. E. Goddard, and P. M. Visscher (2010). Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics 42*(7).

Yeager, K. M., P. H. Santschi, H. S. Rifai, M. P. Suarez, R. Brinkmeyer, C.-C. Hung, K. J. Schindler, M. J. Andres, and E. A. Weaver (2007). Dioxin chronology and fluxes in sediments of the houston ship channel, texas: influences of non-steady-state sediment transport and total organic carbon. *Environmental science & technology 41*(15), 5291–5298.

Yi, X., Y. Liang, E. Huerta-Sanchez, X. Jin, Z. X. P. Cuo, J. E. Pool, X. Xu, H. Jiang, N. Vinckenbosch, T. S. Korneliussen, et al. (2010). Sequencing of 50 human exomes reveals adaptation to high altitude. *science 329*(5987), 75–78.

Yoshikawa, M. and K. Asaba (2020). Single-nucleotide polymorphism rs4664308 in PLA2R1 gene is associated with the risk of idiopathic membranous nephropathy: a meta-analysis. *Scientific Reports 10*(1), 1–8.

Yu, Y., S. Chen, G. F. Lu, Y. Wu, L. Mo, Z. Q. Liu, P. Y. Zheng, Z. Liu, and P. C. Yang (2014). Alphavbeta6 is required in maintaining the intestinal epithelial barrier function. *Cell Biology International 38*(6), 777–781.

Yu-Kemp, H. C., J. P. Kemp, and W. M. Brieher (2017). CRMP-1 enhances EVL-mediated actin elongation to build lamellipodia and the actin cortex. *Journal of Cell Biology 216*(8), 2463–2479.

Zeberg, H. and S. Pääbo (2021). A genomic region associated with protection against severe COVID-19 is inherited from Neandertals. *Proceedings of the National Academy of Sciences 118*(9), e2026309118.

Zhang, X., A. H. Smits, G. B. van Tilburg, P. W. Jansen, M. M. Makowski, H. Ovaa, and M. Vermeulen (2017). An Interaction Landscape of Ubiquitin Signaling. *Molecular Cell 65*(5), 941–955.

Zhang, X., K. Witt, A. Ko, K. Yuan, S. Xu, R. Nielsen, and E. Huerta-Sanchez (2020). The history and evolution of the denisovan-EPAS1 haplotype in tibetans. *bioRxiv*, 1–21.

Zheng, Y., M. Peng, Z. Wang, J. M. Asara, and A. L. Tyner (2010). Protein Tyrosine Kinase 6 Directly Phosphorylates AKT and Promotes AKT Activation in Response to Epidermal Growth Factor. *Molecular and Cellular Biology 30*(17), 4280–4292.

Zhou, S., G. Butler-Laporte, T. Nakanishi, D. Morrison, J. Afilalo, M. Afilalo, L. Laurent, M. Pietzner, N. Kerrison, K. Zhao, E. Brunet-Ratnasingham, D. Henry, N. Kimchi, Z. Afrasiabi, N. Rezk, M. Bouab, L. Petitjean, C. Guzman, X. Xue, C. Tselios, B. Vulesevic, O. Adeleye, T. Abdullah, N. Almamlouk, Y. Chen, M. Chassé, M. Durand, M. Pollak, C. Paterson, H. Zeberg, J. Normark, R. Frithiof, M. Lipcsey, M. Hultström, C. M. T. Greenwood, C. Langenberg, E. Thysell, V. Mooser, V. Forgetta, D. E. Kaufmann, and J. B. Richards (2020). A Neanderthal OAS1 isoform Protects Against COVID-19 Susceptibility and Severity: Results from Mendelian Randomization and Case-Control Studies. *medRxiv*, 2020.10.13.20212092.