Evidence for evolutionary shifts in the fitness landscape of human complex traits

Lawrence H. Uricchio,[1*†] Hugo C. Kitano,[2*]
Alexander Gusev,[3‡] Noah A. Zaitlen[4,5‡†]

Departments of [1]Biology and [2]Computer Science, Stanford University, Stanford, CA; [3]Dana Farber Cancer Institute, Boston, MA; Departments of [4]Medicine and [5]Bioengineering and Therapeutic Sciences, University of California, San Francisco, CA

[*]These authors contributed equally to this work
[‡]These authors contributed equally to this work
[†]To whom correspondence should be addressed: uricchio@stanford.edu, noah.zaitlen@ucsf.edu

Selection alters human genetic variation, but the evolutionary mechanisms shaping complex traits and the extent of selection's impact on polygenic trait evolution remain largely unknown. Here, we develop a novel polygenic selection inference method (Polygenic Ancestral Selection Test Encompassing Linkage, or PASTEL) relying on GWAS summary data from a single population. We use model-based simulations of complex traits that incorporate human demography, stabilizing selection, and polygenic adaptation to show how shifts in the fitness landscape generate distinct signals in GWAS summary data. Our test retains power for relatively ancient selection events and controls for potential confounding from linkage disequilibrium. We apply PASTEL to nine complex traits, and find evidence for selection acting on five of them (height, BMI, schizophrenia, Crohn's disease, and educational attainment). This study provides evidence that selection modulates the relationship between frequency and effect size of trait-altering alleles for a wide range of traits, and provides a flexible framework for future investigations of selection on complex traits using GWAS data.

Introduction

Natural selection shapes patterns of genetic variation within and between human populations, but the phenotypic targets of selection and the evolutionary mechanisms shaping causal variation for selected traits remain largely unknown. Most studies of selection in humans have focused on classic selective sweeps [1–5], but other selection mechanisms such as stabilizing selection [6], polygenic adaptation [7, 8], and soft sweeps [9] may also play an important role in shaping human diversity. Methods to detect

selection under these more complex models are needed if we are to fulfill the promise of genomics to explain the evolutionary mechanisms driving the distribution of heritable traits in human populations [10].

With the recent proliferation of paired genotype and phenotype data from large human cohorts, it is now feasible to develop and implement statistical tests for polygenic selection in humans. Recently, studies have proposed methods to detect polygenic selection that capitalize on these rich datasets, and have begun to uncover evidence that selection acts on complex traits. Two studies proposed empirical methods that test for an excess of allele frequency differentiation at trait-associated loci [7,11], and showed that selection may have driven increases in the height of northern Europeans. This approach was later extended to a model-based framework that also incorporated environmental variables and was applied to several phenotypes in diverse human populations, providing additional evidence for selection on height and identifying a strong selection signal for skin pigmentation [8]. Recently, a novel haplotype-length-based statistic was introduced [12], and used to provide evidence that selection has acted on several complex traits on very recent timescales, and another study measured the correlation between effect sizes and allele frequencies to provide evidence for selection acting on height and body mass index (BMI) [13].

While these studies provide strong evidence that polygenic selection is an important determinant of human genetic and phenotypic variation for some traits, important questions remain about the evolutionary mechanisms that drive complex trait variation. In particular, most previous studies of selection on human complex traits have focused either on polygenic adaptation [7, 8, 12, 14, 15] or stabilizing selection [13, 16–18], and have ignored the interplay between the two. In stabilizing selection models, all trait-altering alleles are deleterious [6, 19], regardless of their direction of effect, while a common assumption of polygenic adaptation models is that a selection pressure towards increased trait values will induce all trait-increasing alleles to be advantageous [7, 8]. A more natural way to model polygenic adaptation is to view stabilizing selection as a null process, with punctuated changes in the fittest trait value (herein called the "optimal trait value" or "trait optimum") driving brief periods of adaptation [20]. This modeling framework leads to different predictions about the dynamics of polygenic selection after a subtle change in the optimal trait value, because trait-increasing alleles are not generically fitness-increasing after a shift to a higher optimal phenotype value. While models that jointly consider stabilizing and adaptive evolution have received theoretical attention [20, 21], there have been few attempts to use the predictions of these models in empirical research. Hence, we posited that improved integration of our understanding of stabilizing selection and polygenic adaptation may provide new insights into the action of selection on complex traits, and perhaps generate evidence for more subtle evolutionary shifts in the fitness landscapes of complex traits.

In addition to these potential avenues for new conceptual insights, existing methods for polygenic

selection detection have some technical limitations. Methods that use the correlation between effect sizes and allele frequencies achieve their greatest power to detect selection when including rare alleles, but accounting for cryptic confounders is often most difficult for low frequency variants [22]. Moreover, a fundamental assumption of fitting a linear model that relates frequencies to effect sizes is that each observed effect size is independent. Since linkage disequilibrium (LD) drives correlations in effect size between alleles of varying frequencies, this assumption is violated in human genetic data, which is likely to elevate the false positive rate for this style of test. Methods relying on a signal of differentiation between populations [7, 8, 11] can only be applied when multiple populations are available and are useful for detecting recent selection that has occurred post-divergence. When the populations of interest are very recently diverged or data from only a few populations are available, methods of this style may suffer from decreases in statistical power [23]. A state-of-the-art haplotype-based method [12] can be applied within a single population, but requires whole-genome sequencing data and is tuned to detect even more recent selection events. The constraints of these methods may have restricted our understanding of selection on complex traits to very recent time-scales, and perhaps limited our ability to detect subtle selection signals.

Here, we develop a novel and efficient statistical test for the action of selection on complex traits that uses only GWAS summary data and LD information as input. Our method requires estimated effect sizes from only a single population, maintains power when considering only common alleles, and relies on effect size differentiation at derived as compared to ancestral alleles. We use model-based simulations that account for human demography, stabilizing selection, polygenic adaptation, and asymmetric mutation rates for trait-increasing and -decreasing alleles to motivate our method and show that signals of weak polygenic adaptation can persist over relatively long evolutionary times. We then develop empirical tools to control for possible confounding by LD. We apply our method to GWAS summary data for nine phenotypes, and find strong evidence for selection acting on five of them (height, BMI, Crohn's disease, schizophrenia, and educational attainment), and show that four of these signals are suggestive of evolutionary shifts in the fitness landscapes of the traits. We discuss the implications of our findings for human evolutionary history and GWAS of biomedically relevant traits.

Results

Quantitative trait model

We develop a polygenic selection quantitative trait model that maps selection coefficients s to effect sizes β. We suppose that stabilizing selection acts on a trait, and that the fittest value of the trait is ϕ_o (also referred to as the "trait optimum"), such that the fitness f of an individual with trait value ϕ is given by

$$f(\phi) = e^{\frac{-(\phi-\phi_o)^2}{w^2}}, \tag{1}$$

where w is the breadth of the fitness function. We additionally suppose that the trait ϕ has a normal distribution such that

$$P(\phi) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(\phi-\bar{\phi})^2}{\sigma^2}}, \tag{2}$$

where σ is the standard deviation of the fitness distribution and $\bar{\phi}$ is the mean trait value in the population. Under these conditions, it is possible to solve for the per-generation, per-individual selection coefficient s as a function of the above model parameters for causal alleles of effect size β. We calculate s by marginalizing the fitness effect of a new mutation of effect size β across all fitness backgrounds. Our calculation proceeds similarly to previous work [18,21], although we retain the dependence on the current mean phenotype value ($\bar{\phi}$) and the phenotype optimum (ϕ_o), rather than assuming that the phenotype distribution is centered at the optimum. While the full expression for the expected change in allele frequency is provided in the Supplementary Materials, we note that when the trait is at equilibrium such that $\bar{\phi} = \phi_o$,

$$s \approx -\frac{\beta^2}{2\left(\sigma^2 + w^2\right)}, \tag{3}$$

and the expected frequency change δp for an allele at frequency p with effect size β when $\bar{\phi} = \phi_o$ is

$$\mathbb{E}[\delta p] \approx -\frac{\beta^2}{2\left(\sigma^2 + w^2\right)} p(1-p)(-1+2p), \tag{4}$$

which implies that the trait evolves as if underdominant when the mean population trait value is centered close to the optimum.

Eqn. 3 implies that $\beta \propto |s|^{1/2}$. Hence, this model can be directly related to the widely used model of Eyre-Walker [6], which maps selection coefficients to effect sizes β as $\beta \propto |s|^{\tau}$. However, unlike Eyre-Walker's model, this model naturally accommodates shifts the trait optimum because the full expression for s is written as a function of the current optimal value of the trait and the current distribution of the phenotype in the population, which makes it straightforward to link the action of stabilizing selection

(which dominates the selection process when the trait distribution is centered at the optimum ϕ_o) to that of transient polygenic adaptation (which dominates when the trait distribution is not centered at ϕ_o).

In addition to our selection model, we impose a mutation model in which the rates of trait-increasing and -decreasing alleles are not necessarily equal. This aspect of our model captures the biological reality that there may not be exactly equal proportions of fixed sites in the human genome that can either increase or decrease a phenotype. For example, if selection has persistently driven a particular phenotype to larger values, we might expect that fixed sites tend to confer larger phenotype values, and recurrent mutations at these sites will then tend to be trait-decreasing.

Under our model, immediately after a shift in optimum, a portion of the causal alleles will increase fitness (specifically, alleles that are on average fitness-increasing when marginalizing across all phenotype and genotype backgrounds), while the remainder of causal alleles will be fitness-decreasing. The population mean will evolve to move closer to the optimum and eventually equilibrate to the new optimal phenotype value, at which point all trait-altering variable sites will again be fitness-decreasing. Fig. 1A&B provide an illustration of effect of a causal allele on fitness and the impact of a shift in the fitness optimum on the distribution of fitness effects, while a mathematical description of the model is provided in the Supplementary Materials.

A novel polygenic selection statistic

Stabilizing selection on a quantitative trait constrains large effect alleles to low frequencies [6, 24, 25]. Polygenic adaptation, which drives shifts in causal allele frequency depending on both frequency and effect size [8], will also induce mean effect size to vary as a function of allele frequency. Here, we propose a test statistic that captures signals of both relatively ancient polygenic adaptation and long-term stabilizing selection. Our statistic, S_β, detects mismatches in mean effect size (denoted $\overline{\beta}$) between ancestral (*i.e.*, the allele that was present in the ancestral population) and derived alleles (*i.e.*, the mutant allele) of equal minor allele frequency (MAF). Ancestral and derived alleles of the same MAF have dramatically different mean ages [26, 27] – since selection purges large effect alleles rapidly, it will drive the mean effect size of ancestral and derived alleles of equal MAF to be different. Under a neutral model of trait evolution, we expect no such relationship between effect size and frequency. S_β also lends itself to a natural permutation test for significance, and is sensitive to stabilizing selection and polygenic adaptation.

S_β is defined as the sum of the mean difference in effect size between derived and ancestral alleles of equal frequency.

$$S_\beta(x_i, x_f) = \sum_{x=x_i}^{x=x_f} \overline{\beta}_D(x) - \overline{\beta}_A(x), \tag{5}$$

where $\overline{\beta}_D(x)$ is the mean effect size of derived alleles with minor allele frequency x and $\overline{\beta}_A(x)$ is the mean effect size for ancestral alleles. We group alleles into 1% frequency bins such that x_i and x_f are elements of $(0, 0.01, 0.02, ..., 1.0)$. Selection drives $\overline{\beta}_D(x) - \overline{\beta}_A(x)$ to differ from 0, and the sum $S_\beta(x_i, x_f)$ then captures the cumulative deviation from 0. Herein, we refer to this statistic as S_β unless we are specifically indicating the frequency cutoff for a particular calculation. In the next section, we use simulations to show that stabilizing selection and polygenic adaptation can be captured with S_β, and subsequently we develop a permutation-based method that accounts for LD to calculate the significance of the deviation of S_β from 0.

S_β is sensitive to recent and ancient selection

We performed forward simulations of genotypes and phenotypes under a model of European demographic history [28] while incorporating selection on complex traits [6]. We used a weak selection coefficient distribution that was inferred for human conserved noncoding sequences [29], such that the selection coefficients are consistent with patterns of diversity observed in human functional regions. The demographic model is shown in gray in Fig. 1C. We set $\sigma = 1$ (the standard deviation of the trait) and $w = 5$ (the breadth of the fitness function), such that the fitness function is much broader than the trait (*i.e.*, fitness declines only gradually in distance from the trait optimum ϕ_o). In addition to population genetic parameters, our simulations include ancestral misassignment. We supposed that 10% of alleles were assigned incorrect ancestral/derived status for these simulations, and include a thorough investigation of the impact of ancestral state uncertainty and a more detailed description of our simulations in the Supplementary Materials.

We considered four models of selection, first a stabilizing selection model in which the optimal value of the phenotype is unchanged throughout human evolutionary history predating the out-of-Africa event (black line, Fig. 1C), second a polygenic adaptation model in which the optimal trait value ϕ_o increases by 100% relative to the standard deviation of the trait (σ) in the African ancestral population (magenta line, Fig. 1C), third a polygenic adaptation model in which the optimal value increases by 100% at the out-of-Africa event (green line, Fig. 1C), and lastly a model in which a 100% increase in the optimal value occurs at the time of the second bottleneck in the European population (blue line, Fig. 1C). Note that for height, a 100% shift corresponds to a relatively modest change of approximately 2.5 inches [30]. Throughout this section, simulation results represent the mean across 5,000 independent simulations of selection on a polygenic trait. The dashed lines show the optimal value as a function of time, while the solid lines show the mean observed phenotype value in the population. For this set of simulations, we

choose the mutation rate of trait-decreasing alleles to be greater than that of trait-increasing alleles (53% of mutations are trait-decreasing) because this set of parameters produces patterns that are qualitatively similar to human height data, but note that the value of S_β depends on the full suite of population genetic parameters (including mutation rate bias, selection strength, and polygenicity) and can be positive or negative. In the Supplementary Materials, we additionally consider a model with equal trait-increasing and -decreasing mutation rates and show that the results are similar. After a change in the optimal value, the population rapidly adapts to the new environmental conditions (magenta, blue, and green curves, Fig. 1C), similar to the predictions made by similar models of quantitative trait evolution [20].

In Fig. 1D, we plot the mean value of the derived allele effect size $\overline{\beta}$ (where we have dropped the subscript D for simplicity) as a function of derived allele frequency at the end of the simulation time-course (*i.e.*, $t = 0$) for each of the selection models. When the trait evolves neutrally, or there is long-term stabilizing selection but no difference in the mutation rates of trait-increasing and trait-decreasing alleles, $\overline{\beta}$ is expected to be 0 in all allele frequency bins (dashed black line). Although stabilizing selection will constrain large effect alleles to low frequency, trait-increasing alleles and trait-decreasing alleles are equally likely to occur when there is no bias in mutation rate, so the mean value of β is 0 at all frequencies. Note that differences in the mutation rate of trait-increasing and -decreasing alleles in the absence of selection will translate the null either upwards (if trait-increasing alleles are more common) or downwards (if trait-decreasing alleles are more common), but will not induce effect size to vary with frequency.

If stabilizing selection acts on the trait and mutation rate differs between trait-decreasing and trait-increasing alleles, $\overline{\beta}$ is negative or very near 0, and increases towards 0 with increasing allele frequency (black lines and points). Shifts in the phenotype optimum to a larger value drive $\overline{\beta}$ to be positive for all but the rarest alleles (magenta, green, and blue lines and points). A recent shift in optimum (in blue) drives a sharp departure from the no-shift model (in black), while a more ancient shift will slowly relax back to the no-shift model (green, magenta) and hence is less differentiated from the no-shift case. Still, with this human-relevant distribution of selection coefficients, even a relatively ancient and modest 100% shift in the optimal phenotype value induces a departure from the stabilizing selection model.

In Fig. 1E, we calculate the value of $S_\beta(0, x)$, *i.e.* the cumulative value of the test statistic as a function of allele frequency x. When no shift in phenotype optimum occurs, mutation rate is biased towards trait-decreasing alleles, and stabilizing selection acts on the trait, $S_\beta(0, x)$ is always negative and departs from the null (*i.e.*, 0 – note that if the mutation rates are not biased then stabilizing selection does not depart from the null as in Fig. S4). When there is an ancient change in phenotype optimum to a larger value, the value of $S_\beta(0, x)$ is strongly positive for all but the lowest allele frequencies. Note that the direction of the departure of $S_\beta(0, x)$ from 0 under models of selection depends on both the

mutation and phenotype optimum parameters – if selection is stabilizing and there is a bias in mutation rate towards trait-decreasing alleles, mean $S_\beta(0, x)$ will be negative or very close to 0 at all frequencies, whereas the opposite is true for a bias towards trait-increasing alleles. A shift towards a larger optimum value of the trait will cause trait-increasing alleles of weak effect to transiently increase in frequency and hence will generally increase the value of S_β, whereas a shift towards a lower optimum will decrease the value of S_β.

We further investigated S_β by comparing it to the correlation coefficient ρ between MAF and β, which was previously used to infer the action of selection [13], and is among the very few published methods used to infer selection on complex traits using only GWAS summary data from a single population (Fig. 2). For this set of simulations, in which we supposed no mutational bias and a 200% shift in ϕ_o relative to σ, ρ is centered near zero and broadly distributed for all but the most recent selection events, and hence has little power to capture relatively ancient polygenic adaptation, while S_β is strongly differentiated from the null even when polygenic adaptation is relatively ancient and relatively modest in magnitude. In Fig. S4, we show that these results hold qualitatively for a more modest 100% shift in ϕ_o.

PASTEL: A permutation-based test for selection

Under the neutral null model, S_β is expected to be 0. If every site in the genome were independent, then we could additionally model the variance of S_β by supposing that causal alleles are drawn from some distribution of known form and summing across the variance induced by each individual marker. Unfortunately, the variance is not straightforward to calculate because it depends on both the distribution of frequencies of putatively causal GWAS alleles (which depends on SNP ascertainment and evolutionary forces such as selection and demography), as well as LD between alleles. To test for a significant departure of S_β from 0, we therefore develop a permutation-based method that accounts for LD and uneven sampling of allele frequencies. Since causal alleles are linked to non-causal alleles in the human genome, a test that does not account for LD will under-estimate the variance in S_β under the neutral null, and will be anti-conservative.

To account for LD, we divide the genome into 1,703 LD blocks, which were previously identified as being approximately independent [31]. For each LD block, we then select a random sign (positive or negative with equal probability), and multiply all the effect sizes in the LD block by this sign. We then recompute S_β on the randomized data. By repeating this procedure, we generate a null distribution for our test statistic S_β. This method maintains the correlations between effect sizes generated by LD, the site frequency spectrum of the sampled alleles, and the joint distribution of the absolute value of effect

size and allele frequency, while breaking any relationship between $\overline{\beta}$ and allele frequency. Note that this is a conservative permutation, because many of the alleles within an LD block are not linked or only weakly linked. We further consider the robustness of our method to population stratification in a subsequent section, which is a persistent potential source of false positives for studies of selection.

To assess significance, we perform a two-tailed test comparing the observed value of S_β to the permutation-based null distribution. We name this test PASTEL, or the Polygenic Ancestral Selection Test Encompassing Linkage.

Signals of selection on human height

Human height has been the focus of numerous selection studies from both anthropological [32–35] and genetic [7,8,36] perspectives, and evidence from both of these fields strongly supports the hypothesis that height is a selected trait. Hence, we begin our data analysis by applying our method to human height GWAS summary data from the GIANT consortium [37], which we take as a positive control to validate our method.

In Fig. 3A, we plot the mean value of β as a function of allele count x in the GIANT data. Note that we have binned alleles by frequency into 100 distinct bins, and the GWAS sample size is much larger than 100. $\overline{\beta}$ is strongly negative for low frequency alleles and increases steadily to a positive value for moderate and high frequency alleles. In the Supplementary Materials, we show that the sharp upward trend in $\overline{\beta}$ at very high derived allele counts is likely to be driven by ancestral state uncertainty, and show that this decreases our power but otherwise does not affect our test.

In Fig. 3B, we plot $S_\beta(0, x)$ for all alleles with effect size estimates in the study [37]. We observe that $S_\beta(0, x)$ is an increasing function of x, in qualitative agreement with the simulation data in Fig. 1D. Lastly, in Fig. 3C, we present the results of LD-preserving permutations of derived and ancestral states with PASTEL. For each of our 2×10^3 permutations of derived/ancestral states, we calculated $S_\beta(0, 1)$ to generate a null distribution for our test statistic. We plot a histogram of this null distribution, while the dashed vertical line represents the observed value of $S_\beta(0, 1)$ in the height summary data. This observed value falls outside the histogram, indicating that the p-value for the test is < 0.0005.

Evidence for selection on human complex traits

We selected 8 additional phenotypes with GWAS summary data available on which to perform our statistical test for selection on complex traits. We prioritized phenotypes with large sample sizes and a large number of GWAS hits, while additionally selecting a wide variety of traits that include body

size [37–39], psychiatric conditions [40], immune-related traits [41], reproductive traits [42], cardiovascular traits [43], and correlates of intelligence [44].

Tab. 1 lists the test statistic value and PASTEL-based p-value for each of the phenotypes. We calculated S_β both for all variants and for common variants only (MAF $> 1\%$ and MAF $> 5\%$). We include the common variant test because ancestral mispolarization has a stronger effect on rare alleles (Fig. S6), and as we show in the Supplementary Materials, incorrect ancestral assignment can decrease the power of our test. Moreover, population structure could potentially confound selection signals at very low frequency (see next section). The full data that correspond to the summary statistics and permutations are plotted in Figs. S7-S15.

After correcting for multiple testing for nine tests (one for each phenotype), we reject the neutral null for Crohn's disease, educational attainment, BMI, height, and schizophrenia, but not BMI-adjusted waist-hip-ratio, global lipid levels, depression, or menopause onset. Three of the statistically significant test statistics are negative (BMI, Crohn's disease, and schizophrenia), which is consistent with models of stabilizing selection where the mutation rate of trait-decreasing alleles exceeds that of trait-increasing alleles, but is also consistent with models in which the trait optimum has moved to a lower value in the species' evolutionary past. However, for both Crohn's disease (Fig. S10) and BMI (Fig. S8), $\overline{\beta}$ is positive at low frequency and negative at high frequency, a pattern which we only observed in simulations that included a decrease in the optimal phenotype value. Similarly, height (Fig. S7) and educational attainment (Fig. S9) have positive values of S_β, and negative $\overline{\beta}$ at low frequency and positive $\overline{\beta}$ at high frequency, a pattern that we only observed in simulations including an evolutionary shift to higher optimal phenotype values.

Potential confounding by population structure

As with nearly all tests for polygenic selection (*e.g.*, [7, 8]), population structure could confound our results. In particular, if a GWAS sample is composed of individuals from two or more populations, and derived alleles have systematically higher frequencies in some populations than others, we might expect that $\overline{\beta}$ is systematically biased and correlated with allele frequency.

While it is very challenging to control for population structure at low frequencies [22], population structure in high frequency alleles has been studied very widely and is very well-controlled with PCA-based methods, which use the PCs as covariates. Each of the GWAS for which we obtained summary data included such a control, and we therefore expect that population structure is very well-controlled for high frequency alleles in each of these studies. We therefore ran our test excluding rare alleles with

MAF under 1%, and we additionally perform a very conservative test excluding all alleles below 5% (see Tab. 1). We find that the selection signals are very robust when including only common alleles for three of the phenotypes (height, BMI, and educational attainment). The signal for schizophrenia is slightly stronger when excluding rare alleles (MAF $< 1\%$), but is somewhat diminished when considering only alleles above 5%. The signal for Crohn's disease becomes somewhat weaker as we increase the allele frequency threshold, indicating that rare alleles at least partially drive the signal for Crohn's – however, all p-values for Crohn's were nominally significant at the $p < 0.05$ level. We conclude that our test retains power even when excluding rare alleles for sufficiently strongly selected phenotypes, and that the putative selection signals discovered herein are unlikely to be driven by population structure.

Discussion

Many studies have suggested that selection shapes human genetic variation [5], and it has long been theorized that it drives the variance in a broad range of human complex phenotypes [4]. Here, we developed a novel test (PASTEL) for polygenic selection that can be applied to GWAS summary data for a single population, can capture relatively ancient selection events, and retains power when applied only to common alleles, for which it is more straightforward to correct for population stratification. We applied our test to GWAS summary data for nine phenotypes, and showed that five of them (educational attainment, height, Crohn's disease, BMI, and schizophrenia) strongly suggest a role for selection in shaping trait variation. It has been previously suggested that height [7, 8, 11, 13] and BMI are under selection [13], but studies have reported both polygenic adaptation and widespread negative selection as possible selection mechanisms. Our results are consistent with a shift towards higher optimal fitness values of educational attainment and height, but lower optimal fitness values for BMI and Crohn's disease risk, suggesting that long-term stabilizing selection on these traits has been punctuated with periods of polygenic adaptation.

If selection acts on biomedically relevant complex traits in humans such as Crohn's disease and schizophrenia, there are strong implications for the future of both medical and evolutionary genomics. In medical genomics, an ongoing debate about the genomic architecture of complex diseases is at the forefront of the field [45]. When strong selection acts on complex traits, it can elevate the role of rare alleles in driving trait variance [46]. If rare alleles contribute a larger fraction of the genetic variance than is expected under neutral models, then very large GWAS that use only array-based genotyping information are very unlikely to be able to capture these signals, and sequence-based studies and powerful rare variant approaches that are robust to evolutionary forces (including those not investigated here, such as partial

recessivity) will be needed [24, 25]. Moreover, recent work has suggested that the over-representation of Europeans in GWAS has limited the effectiveness of estimating polygenic risk scores in other human populations [47]. This is a serious problem for the transfer of genomic research into the clinic, where precision medicine initiatives relying on personal genetic information will only be successful if genetic risk can be accurately inferred in diverse populations. While this effect has been attributed to neutral demographic forces, if selection has driven numerous phenotypes to acclimate to local environmental conditions in ancestral human populations worldwide it could exacerbate this problem dramatically.

In the field of evolutionary genomics, most studies have agreed that the impact of selection is widespread on the human genome, but the evolutionary mechanisms that drive genetic and phenotypic diversity have been widely debated [4, 5]. In our study, we showed that the selection signals for BMI and Crohn's disease are consistent with a bias in mutation rate towards trait-increasing alleles, and a shift to a lower optimal value of the trait, while height and educational attainment are consistent with a mutational bias towards trait-increasing alleles and a shift towards higher values of the trait optimum. The signal for schizophrenia is consistent with both models of an ancestral shift towards a lower optimum and a stabilizing selection only model in which there is a mutational bias towards trait-decreasing alleles. In general, we propose that it is likely for most selected traits that there are unequal numbers of trait-increasing and trait-decreasing sites that can be mutated genome-wide. If past selection events have pushed a selected trait (*e.g.*, height) to ever higher values, then we expect a majority of fixed height-altering sites to be height-increasing. Recurrent mutations at these sites would then tend to be height-decreasing. Incorporating mutational bias into our model allowed us to then distinguish between signals of stabilizing selection alone, or evolutionary shifts in trait optima. However, we note that other models, for example those with no mutational bias and multiple evolutionary shifts in the trait optimum, or purely deleterious fitness costs associated with disease status instead of stabilizing selection [48], could also generate qualitatively similar signals.

Among the nine traits that we tested, we found that five had strong signals of polygenic selection. However, this does not imply that the other four are not under selection. The power of our test depends on the strength of selection, the polygenicity of the trait, the heritability of the trait, and the mutational bias. If a trait is under strong stabilizing selection, but the mutation rate of trait-increasing and -decreasing alleles is exactly equal, then our test has no power. Moreover, if selection is weak, a small number of causal alleles drive variance in the trait, or the trait is only weakly heritable, power is greatly diminished. However, increased sample sizes in GWAS will always increase our power, because the variance on effect size estimates for even weak effect causal alleles decreases dramatically with sample size. When the standard error on $\overline{\beta}$ decreases, the power of our test will increase. Although we showed that population

structure is unlikely to bias our results, uncorrected population structure is always a concern for tests of polygenic selection, and cannot be completely ruled out. An uncorrected bias in the inferred β values due to population structure will make our test anti-conservative. LD score regression on some of our phenotypes is consistent with a small amount of residual uncorrected population structure [49]. However, LD score regression assumes a specific relationship between allele frequencies and effect sizes, and inflation in the relevant test statistic can be driven by either selection or population structure [49].

One strength of our permutation-based approach is that other summary statistics, such as the absolute value of the deviation between ancestral and derived effect sizes, could also easily be applied. Since stabilizing selection strongly increases $\mathbb{E}[\beta^2]$ at low frequency [24], it may prove fruitful to investigate this statistic. In future studies, it will be advantageous to apply other types of summary statistics and compare their relative power, and to use the information in such summary statistics to infer the evolutionary parameters of complex trait selection models such as the timing of shifts in the trait optima and the strength of selection.

Acknowledgments

The authors acknowledge the support of NIGMS grant K12GM088033 and the Stanford IRACDA program (LHU) and thank Noah Rosenberg, Jeremy Berg, Doc Edge, and Graham Coop's lab for helpful comments.

References

[1] Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, Schaffner SF, et al. Detecting recent positive selection in the human genome from haplotype structure. Nature. 2002;419(6909):832–837.

[2] Voight BF, Kudaravalli S, Wen X, Pritchard JK. A map of recent positive selection in the human genome. PLoS Biology. 2006;4(3):e72.

[3] Sabeti PC, Schaffner SF, Fry B, Lohmueller J, Varilly P, Shamovsky O, et al. Positive natural selection in the human lineage. Science. 2006;312(5780):1614–1620.

[4] Hernandez RD, Kelley JL, Elyashiv E, Melton S, Auton A, McVean G, et al. Classic selective sweeps were rare in recent human evolution. Science. 2011;331(6019):920–924.

[5] Enard D, Messer PW, Petrov DA. Genome-wide signals of positive selection in human evolution. Genome Research. 2014;24(6):885–895.

[6] Eyre-Walker A. Genetic architecture of a complex trait and its implications for fitness and genome-wide association studies. Proceedings of the National Academy of Sciences. 2010;107(suppl 1):1752–1756.

[7] Turchin MC, Chiang CW, Palmer CD, Sankararaman S, Reich D, Hirschhorn JN, et al. Evidence of widespread selection on standing variation in Europe at height-associated SNPs. Nature Genetics. 2012;44(9):1015–1019.

[8] Berg JJ, Coop G. A population genetic signal of polygenic adaptation. PLoS Genetics. 2014;10(8):e1004412.

[9] Messer PW, Petrov DA. Population genomics of rapid adaptation by soft selective sweeps. Trends in Ecology & Evolution. 2013;28(11):659–669.

[10] Pritchard JK, Pickrell JK, Coop G. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. Current Biology. 2010;20(4):R208–R215.

[11] Robinson MR, Hemani G, Medina-Gomez C, Mezzavilla M, Esko T, Shakhbazov K, et al. Population genetic differentiation of height and body mass index across Europe. Nature Genetics. 2015;47(11):1357–1362.

[12] Field Y, Boyle EA, Telis N, Gao Z, Gaulton KJ, Golan D, et al. Detection of human adaptation during the past 2000 years. Science. 2016;354(6313):760–764.

[13] Yang J, Bakshi A, Zhu Z, Hemani G, Vinkhuyzen AA, Lee SH, et al. Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. Nature Genetics. 2015;.

[14] Berg JJ, Zhang X, Coop G. Polygenic Adaptation has Impacted Multiple Anthropometric Traits. bioRxiv. 2017; p. 167551.

[15] Racimo F, Berg JJ, Pickrell JK. Detecting polygenic adaptation in admixture graphs. bioRxiv. 2017; p. 146043.

[16] Zeng J, de Vlaming R, Wu Y, Robinson M, Lloyd-Jones L, Yengo L, et al. Widespread signatures of negative selection in the genetic architecture of human complex traits. bioRxiv. 2017; p. 145755.

[17] Ip H, Jansen R, Abdellaoui A, Bartels M, Boomsma DI, Nivard MG. Stratified Linkage Disequilibrium Score Regression reveals enrichment of eQTL effects on complex traits is not tissue specific. bioRxiv. 2017; p. 107482.

[18] Simons YB, Bullaughey K, Hudson RR, Sella G. A model for the genetic architecture of quantitative traits under stabilizing selection. arXiv preprint arXiv:170406707. 2017;.

[19] Gilad Y, Oshlack A, Rifkin SA. Natural selection on gene expression. Trends in Genetics. 2006;22(8):456–461.

[20] Jain K, Stephan W. Rapid adaptation of a polygenic trait after a sudden environmental shift. Genetics. 2017;206(1):389–406.

[21] Barton N. The maintenance of polygenic variation through a balance between mutation and stabilizing selection. Genetics Research. 1986;47(3):209–216.

[22] Mathieson I, McVean G. Differential confounding of rare and common variants in spatially structured populations. Nature Genetics. 2012;44(3):243–246.

[23] Whitlock MC, Guillaume F. Testing for spatially divergent selection: comparing Q ST to F ST. Genetics. 2009;183(3):1055–1063.

[24] Uricchio LH, Zaitlen NA, Ye CJ, Witte JS, Hernandez RD. Selection and explosive growth alter genetic architecture and hamper the detection of causal rare variants. Genome Research. 2016;26(7):863–873.

[25] Sanjak JS, Long AD, Thornton KR. A model of compound heterozygous, loss-of-function alleles is broadly consistent with observations from complex-disease GWAS datasets. PLoS Genetics. 2017;13(1):e1006573.

[26] Maher MC, Uricchio LH, Torgerson DG, Hernandez RD. Population genetics of rare variants and complex diseases. Human Heredity. 2012;74(3-4):118–128.

[27] Fu W, O'connor TD, Jun G, Kang HM, Abecasis G, Leal SM, et al. Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. Nature. 2013;493(7431):216–220.

[28] Gravel S, Henn BM, Gutenkunst RN, Indap AR, Marth GT, Clark AG, et al. Demographic history and rare allele sharing among human populations. Proceedings of the National Academy of Sciences. 2011;108(29):11983–11988.

[29] Torgerson DG, Boyko AR, Hernandez RD, Indap A, Hu X, White TJ, et al. Evolutionary processes acting on candidate cis-regulatory regions in humans inferred from patterns of polymorphism and divergence. PLoS Genetics. 2009;5(8):e1000592.

[30] Fryar CD, Gu Q, Ogden CL, Flegal KM. Anthropometric Reference Data for Children and Adults: United States, 2011-2014. Vital health statistics Series 3, Analytical and epidemiological studies. 2016;(39):1.

[31] Berisa T, Pickrell JK. Approximately independent linkage disequilibrium blocks in human populations. Bioinformatics (Oxford, England). 2016;32(2):283–285.

[32] Pawlowski B, Dunbar RI, Lipowicz A. Evolutionary fitness: tall men have more reproductive success. Nature. 2000;403(6766):156–156.

[33] Nettle D. Height and reproductive success in a cohort of British men. Human Nature. 2002;13(4):473–491.

[34] Nettle D. Women's height, reproductive success and the evolution of sexual dimorphism in modern humans. Proceedings of the Royal Society of London B: Biological Sciences. 2002;269(1503):1919–1923.

[35] Pawlowski B. Variable preferences for sexual dimorphism in height as a strategy for increasing the pool of potential partners in humans. Proceedings of the Royal Society of London B: Biological Sciences. 2003;270(1516):709–712.

[36] Zoledziewska M, Sidore C, Chiang CW, Sanna S, Mulas A, Steri M, et al. Height-reducing variants and selection for short stature in Sardinia. Nature Genetics. 2015;.

[37] Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, et al. Defining the role of common variation in the genomic and biological architecture of adult human height. Nature Genetics. 2014;46(11):1173–1186.

[38] Shungin D, Winkler TW, Croteau-Chonka DC, Ferreira T, Locke AE, Mägi R, et al. New genetic loci link adipose and insulin biology to body fat distribution. Nature. 2015;518(7538):187–196.

[39] Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, et al. Genetic studies of body mass index yield new insights for obesity biology. Nature. 2015;518(7538):197–206.

[40] CDG Psychiatric Genomics Consortium. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. The Lancet. 2013;381(9875):1371–1379.

[41] Franke A, McGovern DP, Barrett JC, Wang K, Radford-Smith GL, Ahmad T, et al. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. Nature Genetics. 2010;42(12):1118–1125.

[42] Day FR, Ruth KS, Thompson DJ, Lunetta KL, Pervjakova N, Chasman DI, et al. Large-scale genomic analyses link reproductive aging to hypothalamic signaling, breast cancer susceptibility and BRCA1-mediated DNA repair. Nature Genetics. 2015;47(11):1294–1303.

[43] Global Lipids Genetics Consortium, et al. Discovery and refinement of loci associated with lipid levels. Nature Genetics. 2013;45(11):1274–1283.

[44] Okbay A, Beauchamp JP, Fontana MA, Lee JJ, Pers TH, Rietveld CA, et al. Genome-wide association study identifies 74 loci associated with educational attainment. Nature. 2016;533(7604):539–542.

[45] Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. Nature. 2009;461(7265):747–753.

[46] Lohmueller KE. The impact of population demography and selection on the genetic architecture of complex traits. PLoS Genetics. 2014;10(5):e1004379.

[47] Martin AR, Gignoux CR, Walters RK, Wojcik GL, Neale BM, Gravel S, et al. Human demographic history impacts genetic risk prediction across diverse populations. The American Journal of Human Genetics. 2017;100(4):635–649.

[48] Charlesworth B. Stabilizing selection, purifying selection, and mutational bias in finite populations. Genetics. 2013;194(4):955–971.

[49] Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, Patterson N, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nature Genetics. 2015;47(3):291–295.

[50] Fisher RA. The correlation between relatives on the supposition of Mendelian inheritance. Earth and Environmental Science Transactions of the Royal Society of Edinburgh. 1919;52(2):399–433.

[51] Hernandez RD. A flexible forward simulator for populations subject to selection and demography. Bioinformatics. 2008;24(23):2786–2787.

[52] Thousand Genomes Project Consortium, et al. A global reference for human genetic variation. Nature. 2015;526(7571):68–74.

Figures & Tables

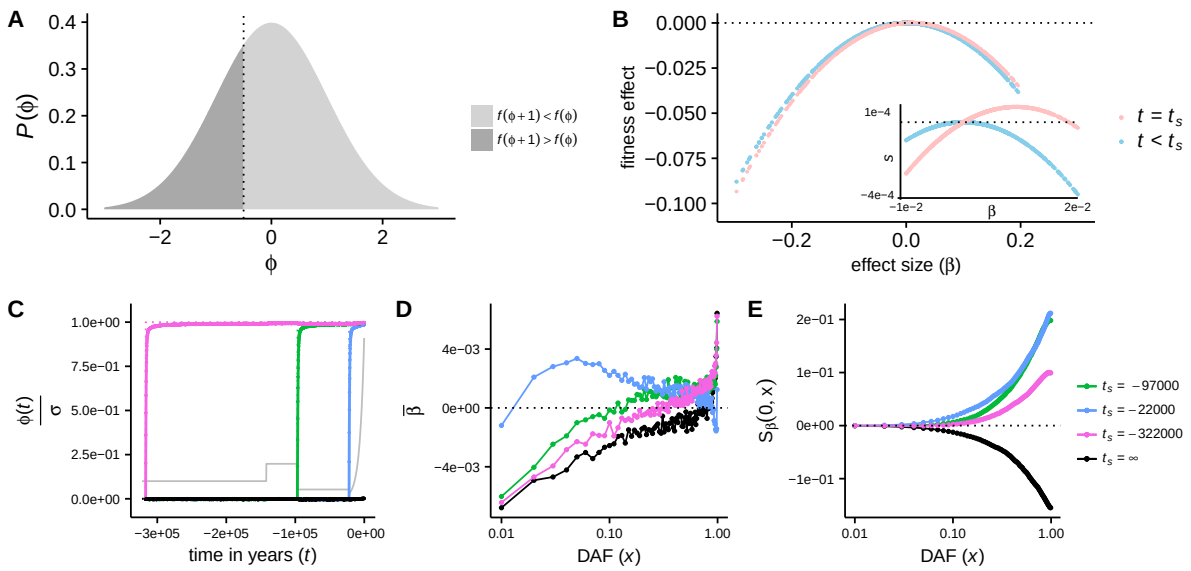

Figure 1: Panels A-B are schematics of our trait model, while C-E show simulation results. A: fitness impact of a $\beta = 1$ mutation. At equilibrium, the trait distribution $P(\phi)$ is symmetric about the optimal value of the phenotype, $\phi = 0$. Individuals with $\phi < {}^{-1}/_{2}$ would increase in fitness ($f(\phi + 1) > f(\phi)$, shaded dark gray) given a mutation with $\beta = 1$, while all others decrease in fitness. B: schematic of the relationship between effect size and fitness effect. At time $t = t_s$ the optimal trait value ϕ_o increases, and trait-decreasing alleles have decreased fitness while trait-increasing alleles have increased fitness. Still, only trait-increasing alleles of small effect are on average fitness-increasing (inset). C: ϕ_o is plotted as a dashed line in black, magenta, blue, and green. ϕ_o changes by 100% relative to the standard deviation (σ) of ϕ coincident with out-of-Africa (green, $t_s = -97000$) and founding of Europe (blue, $t_s = -22000$), in the ancestral African population (magenta, $t_s = -322000$), or remains unchanged (black, $t_s = -\infty$). The demographic model is in gray (not to scale). The solid black, magenta, green, and blue curves correspond to the observed mean of ϕ, which rapidly approaches ϕ_o. D: $\bar{\beta}$ as a function of derived allele frequency (DAF). E: $S_\beta(0, x)$ as a function of DAF. D and E are on a log scale on the x-axis.

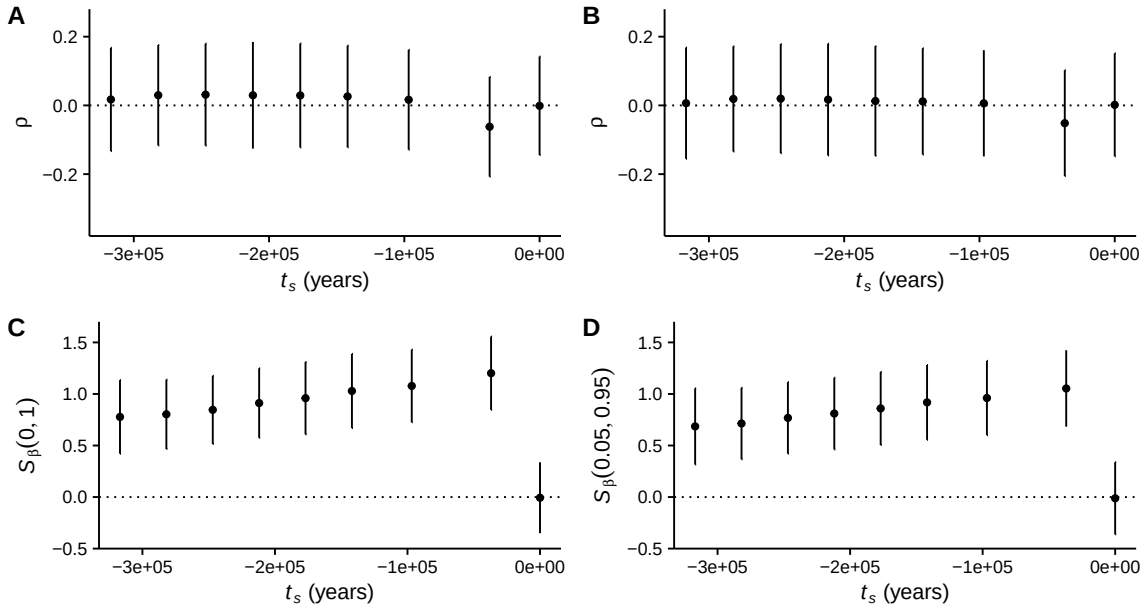

Figure 2: A comparison between the correlation coefficient (ρ) and our selection summary statistic (S_β), computed on all variants (A & C) or only common variants with frequencies greater than 5% (B & D). The points represent the mean over 1,000 simulations, while the bars represent the standard deviation. For each data point, a shift of 200% in the optimal phenotype value occurs at the corresponding time on the x-axis. The neutral null is plotted as a horizontal dashed line.

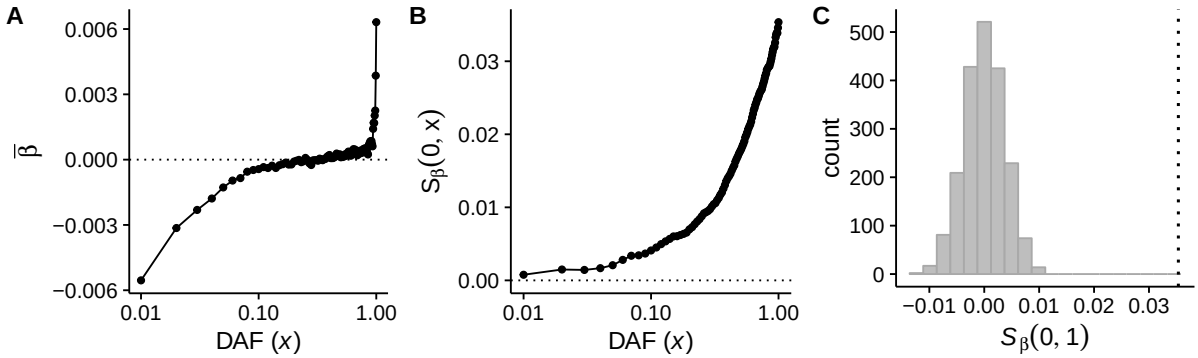

Figure 3: In A, we plot the mean value of β as a function of derived allele frequency x for GWAS summary data from the GIANT project for the height phenotype. In B, we plot $S_\beta(0, x)$ for the same data. C is the neutral null distribution of $S_\beta(0, 1)$ obtained by permutations, and the vertical dashed line indicates the observed value of $S_\beta(0, 1)$ in the GIANT height data.

Phenotype	$S_\beta(0,1)$	p-value	$S_\beta(0.01,0.99)$	p-value	$S_\beta(0.05,0.95)$	p-value
Height	0.0354	$<$ **0.0005**	0.03339	$<$ **0.0005**	0.0305	$<$ **0.0005**
BMI	-0.0113	**0.0015**	-0.0117	$<$ **0.0005**	-0.00981	$<$ **0.0005**
WHR-BMI	0.00272	0.61	-0.00105	0.835	0.00248	0.5205
Education	0.00863	$<$ **0.0005**	0.00863	$<$ **0.0005**	0.00662	$<$ **0.0005**
GLL	0.00309	0.519	-0.00317	0.0185	-0.00211	0.0055
Crohn's disease	-0.0487	**0.0025**	-0.0361	0.014	-0.0291	0.028
Menopause onset	-0.0121	0.3555	-0.0106	0.3715	-0.0228	0.054
Depression	-0.0187	0.356	-0.0177	0.008	-0.0111	0.036
Schizophrenia	-0.0622	0.0075	-0.0488	**0.0045**	-0.026	0.0815

Table 1: Test-statistics and p-values corresponding to GWAS for nine phenotypes that we hypothesized may be under selection. The first and second columns include all alleles, while the third and fourth columns include only alleles that have MAF $> 1\%$, and the final two columns include only alleles with MAF $> 5\%$. Tests that pass a multiple testing correction are bolded. *BMI: body mass index, WHR-BMI: waist-hip ration adjusted for body mass index, GLL: global lipid levels*

Supplementary Materials

A model of quantitative traits under selection

Biological motivation

We develop a quantitative trait model that is qualitatively similar to that of Eyre-Walker at equilibrium [6], but additionally accommodates other realistic features of complex traits, including asymmetric mutation rates for trait-increasing and trait-decreasing alleles and changes in the optimal trait value over time.

Biologically, it is likely that there are an unequal number of fixed bi-allelic sites genome-wide that can increase or decrease a phenotype relative to its current value. For example, if past selection events have driven the phenotype to ever larger values, we might expect that the majority of trait-increasing alleles have already been fixed by positive selection in the evolutionary past, and that further recurrent mutations at these fixed sites would therefore decrease the phenotype.

Temporal variation in the optimal value of selected traits may also be important to consider in evolutionary models of complex traits, as such changes may be an evolutionary mechanism for polygenic adaptation [20]. When a population's environment is altered, perhaps by migration, a change in climate, or the elimination/introduction of competing species, it is likely that selection pressures on phenotypes will also change. If the population persists long enough in this new environment (say on the order of $2N$ generations for a population of size N), it is expected that the phenotype mean will approach the new phenotype optimum, and the population will again be at equilibrium. In the intervening time, while the population is out of equilibrium, allele frequencies will shift as a function of their effect on the phenotype. Our goal is to capture selection's impact on the causal variation during this out-of-equilibrium time period.

Mathematical motivation

In Eyre-Walker's model [6], the effect size β of a causal allele is determined by its selection coefficient s according to the equation $\beta = \delta|s|^{\tau}$, where τ is an exponent that controls the shape of the fitness-phenotype relationship and δ is a random sign. Our model retains features of this model while introducing some modifications.

Asymmetries in mutation rate are therefore easily accommodated by biasing the probability that the random sign δ is positive or negative. In our simulations, we suppose that 47% of sites are trait-increasing while the remainder are trait-decreasing (because this choice generates patterns that are qualitatively similar to the human height data), but the model accommodates any level of asymmetry by simply modifying this parameter.

Accommodating changes in the optimal value of the phenotype is more challenging, and requires adopting some weak assumptions. First, we assume that the phenotype is normally distributed. This is a weak assumption because many phenotypes, such as height, have an empirical distribution that is well-approximated by a normal distribution, and because Fisher showed that polygenic traits under very general assumptions are expected to achieve a normal distribution [50]. Our simulation model could accommodate any phenotype distribution, but some *a priori* assumption about the distribution must be made in our framework. Next, we suppose that the fitness function f is symmetric with respect to the optimal phenotype value ϕ_o, such that $f(\phi_o + \beta) = f(\phi_o - \beta)$, and that fitness decreases monotonically with distance from ϕ_o. While many functions have this property, we additionally suppose that the fitness function f is Gaussian such that $f(\phi) = e^{\frac{-(\phi-\phi_o)^2}{w^2}}$.

Recasting the selection coefficient in terms of the trait distribution

To calculate the time dependent selection coefficient $s(t)$ of a site with effect size β, we first develop some results that allow us to recast the selection coefficient s at equilibrium as a function of the proportion of the population in which an allele of effect size β is fitness-increasing. Since the trait is normally distributed, the probability that an individual has phenotype value ϕ, $P(\phi)$, is given by $P(\phi) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(\phi-\bar{\phi})^2}{\sigma^2}}$. To calculate the selection coefficient for an allele with effect size β, we then marginalize across all trait backgrounds and possible genotypes in the population to obtain the mean fitness effect, since the fitness effect varies across trait backgrounds according to $f(\phi)$. The three possible genotypes for a causal variable allele are denoted as aa (homozygous ancestral), ad (heterozygous), and dd (homozygous derived). For each genotype, we calculate the fitness w of an allele at frequency p as

$$w_{aa} = \int_{-\infty}^{\infty} f(\phi) P(\phi + 2p\beta) d\phi, \tag{6}$$

$$w_{ad} = \int_{-\infty}^{\infty} f(\phi + \beta) P(\phi + 2p\beta) d\phi, \tag{7}$$

$$w_{dd} = \int_{-\infty}^{\infty} f(\phi + 2\beta) P(\phi + 2p\beta) d\phi. \tag{8}$$

substituting in $P(\phi)$ and $f(\phi)$, these integrals can be solved to obtain $w_{aa} \propto e^{\frac{-(2p\beta+\phi_o-\bar{\phi})^2}{2(w^2+\sigma^2)}}$, $w_{ad} \propto e^{\frac{-((-1+2p)\beta+\phi_o-\bar{\phi})^2}{2(w^2+\sigma^2)}}$, and $w_{dd} \propto e^{\frac{-(2(-1+p)\beta+\phi_o-\bar{\phi})^2}{2(w^2+\sigma^2)}}$. We then can solve for the expected change in frequency $\mathbb{E}[\delta p]$ at time of an allele at frequency p as

$$\mathbb{E}[\delta p] = \frac{p(1-p)w_{ad} + p^2 w_{dd}}{2p(1-p)w_{ad} + p^2 w_{dd} + (1-p)^2 w_{aa}} - p, \tag{9}$$

Taking a series expansion about $\beta = 0$, we find that

$$\mathbb{E}[\delta p] \approx \frac{\beta\left(\phi_o - \bar{\phi}\right) p\,(1-p)}{w^2 + \sigma^2} - \frac{\beta^2\left(p(1-p)(2p-1)\left(w^2 + \sigma^2 - \left(\bar{\phi} - \phi_o\right)^2\right)\right)}{2(w^2+\sigma^2)^2} \tag{10}$$

When $\bar{\phi} = \phi_o$, alleles evolve as if underdominant because $\mathbb{E}[\delta p] \approx -\frac{\beta^2}{2(\sigma^2+w^2)} p(1-p)(-1+2p)$. When $\bar{\phi} \neq \phi_o$, alleles may either be expected to increase in fitness (*i.e.*, be transiently positively selected) or decrease in frequency, depending on their effect size and the current mean population trait value $\bar{\phi}$. In our Wright-Fisher simulations, we replace the expected frequency change with eqn. 10 and track the current mean population trait value as a time-dependent quantity (see next section for more details).

Validation of simulations

We developed a custom simulator of our model in Python. Our simulator accommodates changes in population size, including explosive growth and bottlenecks, as well as any arbitrary distribution of selection coefficients. We calculate the distribution of effect sizes β given a distribution on s using eqn. 3. We perform a burn-in period of $5N$ generations for a simulated population of size N, during which we suppose that the population is close to equilibrium and hence s does not vary each generation. After the first burn-in, we perform an additional $5N$ generations of burn-in during which we recalculate $\mathbb{E}[\delta p]$ as a function of β, $\bar{\phi}$, and ϕ_o in each generation with eqn. 10. To calculate the current value of $\bar{\phi}$, we sum over all j causal alleles, such that $\bar{\phi} = \sum_j 2q_j\beta_j$, where q_j is the frequency of a site with effect size β_j.

For simulations of polygenic adaptation, at some time t_s during the simulations, we reset ϕ_o to a new value, which induces the trait distribution to be out-of-equilibrium with respect to the fitness function. During the out-of-equilibrium period, a portion of the causal sites will be positively selected (specifically those that are fitness increasing when marginalizing across all trait backgrounds, as described in the previous section), while the remaining sites will be fitness-decreasing. Each generation, we recalculate δp based on the current configuration of $\bar{\phi}$ and ϕ_o.

To validate our simulator, we simulated a complex selection and demographic model using previously published models of European demographic history [28] and selection on human conserved elements [29] in SFS_CODE [51] and compared the SFS_CODE frequency spectrum to the frequency spectrum in our simulations. When no shift in the trait optimum occurs, selection coefficients in our model have the same expected value as in the standard Wright-Fisher model with underdominance, so the frequency spectra should be similar. However, since underdominance is not straightforward to simulate in SFS_CODE, for this set of validation simulations we replaced the underdominance term in our simulations with the standard genic selection model. Results of these simulations are plotted in Fig. S2. We observe good agreement between the two spectra overall, although our model results in a slight over-representation of rare alleles and a slight under-representation of common alleles relative to SFS_CODE. Note that there is weak LD in the SFS_CODE simulations and that our selection model differs slightly, so we do not expect perfect agreement. For the SFS_CODE simulations, we used the following command line: `sfs_code 1 500 -N 1000 -n 100 -A -t 0.001 -r 0.0 -TE 0.405479 -Td 0 P 0 1.982738 -Td 0.265753 P 0 0.128575 -Td 0.342466 P 0 0.554541 -Tg 0.342466 P 0 55.48 -L 100 150 -l g 0.5 R -a N R -W 2 0 1 1 0.0415 0.00515625 -s <random_seed> -o <out>`

For all of the simulations presented in the manuscript, the curves represent the mean over 5,000 independent simulations unless otherwise stated. We used $N = 1,000$ (where N is the ancestral population size) and rescaled the mutation and selection coefficient parameters to accommodate this smaller population size. The simulation code was implemented in Python and numpy, and will be freely available and posted on Github. For all of our simulations, we used the selection coefficient distribution from [29] which has $\mathbb{E}[2Ns] = -8$, and the following parameters unless otherwise stated: $\theta = 4N\mu = 0.001$, $\theta_- = 4N\mu_- = 0.001 * 0.53$ (*i.e.*, the mutation rate for alleles that are trait-decreasing – trait-increasing alleles make up the remainder of the mutations), 10^6 causal target causal sites (*i.e.*, the number of causal alleles that can affect the phenotype – note that not all sites are mutated at any given time in a particular simulation), $w = 5$, $\sigma = 1$.

The behavior of S_β in simulation

We performed simulations under a suite of models, including those with equal mutation rates for trait-increasing and trait-decreasing alleles, unequal mutation rates, stabilizing selection, and polygenic adaptation. In the main text we focused on the models with unequal mutation rates for trait-increasing and -decreasing alleles because these models result in patterns similar to those observed in the empirical height data. Here, we consider models with equal mutation rates for trait-increasing and -decreasing alleles.

In Fig. S3A, we plot the European demographic model that we simulated [28], as well as the mean phenotype value across 1,000 independent simulations. As in the main text, the three colors correspond to a shift of 100% in the optimal phenotype value at the out-of-Africa bottleneck (green), an ancient 100% change in the optimal phenotype value (magenta), and long-term stabilizing selection with no change in the optimal value of the phenotype (black). All three models result in qualitatively very similar patterns for the mean phenotype value as those simulated in the main text (Fig. 1C).

In contrast to the mutation bias models, when mutation rates are equal for trait-increasing and -decreasing alleles, stabilizing selection does not induce a relationship between allele frequency and effect size (black lines/points, Fig. S3B), and S_β has no power. However, polygenic adaptation again results in elevating the effect sizes relative to the stabilizing selection model and the neutral null (Fig. S3B). And

in qualitative agreement with the results in Fig. 1C, S_β is highly differentiated from the null and has a similar form to the model with unequal mutation rates.

Comparing S_β to ρ

We compared S_β to ρ, the Pearson correlation coefficient between effect size and allele frequency. ρ was previously used to infer selection acting on polygenic traits [13]. We simulated polygenic adaptation as a 100% shift (note that the same model with a 200% shift is plotted in the main text in Fig. 2) in the optimal value of a quantitative trait over a range of time spans, from more than 300,000 years ago to 22,000 years ago (the latter date corresponding to the bottleneck at the founding of Europe). As in the main text, we chose a demographic model that was inferred from European genomic sequences [28] and a distribution of selection coefficients inferred from human conserved noncoding regions [29]. We calculated ρ and S_β at the end of each simulation (which corresponds to the present day) for 1,000 independent simulations at each time point. In keeping with the simulations in the previous section, for this set of simulations we set the mutation rate of trait-increasing and -decreasing alleles to be equal.

In Fig. S4, we plot the distribution of S_β and ρ, where the points represent the mean and the bars represent the standard deviation over the 1,000 simulations. ρ is not substantially different from 0 for all but the most recent selection events regardless of whether it is computed on all variants (A) or only common variants (C, $> 5\%$), implying that it would not be able to differentiate between neutral evolution and polygenic selection for ancient selection events. In contrast, S_β is elevated relative to the null for all time points. This signal is not driven by rare variants, because common variants (D) show a very similar signal to all variants (C).

We note that in this set of simulations, we have used 10^6 causal sites per simulation, only a portion of which are expected to be variable in any given simulation. If the number of causal sites is much larger, then both ρ and S_β are both likely to have increased power. Moreover, power depends on the full suite of population genetic parameters, including selection strength and the demographic model, because these parameters affect the distribution of effect sizes and allele frequencies. Hence, our results in this section should be interpreted as a statement of relative power under a constrained set of model assumptions, and not a general statement about the time-scales and absolute power of each of these styles of statistical tests.

The relationship between MAF and $\overline{\beta}$

We simulated a wide variety of parameter ranges, and reported three of the simulated models in this study. As discussed in the main text, these models include a stabilizing selection model in which the optimal value of the phenotype is unchanged throughout recent human evolutionary history (black line), a polygenic adaptation model in which the optimal value increases by 100% in the African ancestral population (magenta line), at the out-of-Africa event (green line), or at the time of the second bottleneck in the European population (blue line). For this set of simulations, we included a bias in mutation rate towards trait-decreasing alleles (53% of new mutations are trait-decreasing).

While the relationship between derived/ancestral allele frequencies and effect sizes is not commonly explored in most complex trait studies, the relationship between minor allele frequency (MAF) and $\overline{\beta}$ is somewhat more familiar [13]. We therefore explored this relationship in our simulated data as well. In Fig. S5, we plot this relationship for the three selection models simulated in the main text, namely stabilizing selection with a bias in mutation rate towards trait-decreasing alleles (black), the same model plus a relatively ancient shift in the trait optimum corresponding with the out-of-Africa bottleneck (green), a more recent shift in the trait optimum corresponding to the founding of the European population (blue), and an ancient shift (magenta). We observe that a downwards mutation bias and stabilizing selection results in a positive correlation between effect size and MAF, while mean effect sizes are generally negative or near 0. Adding shifts in the trait optimum maintains this overall picture, but shifts the mean effect

sizes upwards to positive values at moderate to high allele frequencies. When the shift is very recent (blue), a negative correlation between effect size and allele frequency is induced.

The impact of ancestral uncertainty on S_β

Ancestral states are not directly observed in genomic data, and are typically inferred by comparing human sequences to those of an out-group. The underlying assumption is that the allelic state in the out-group represents the most likely ancestral state for the allele. While this approach correctly assigns the ancestral state for the majority of alleles, it does not account for recurrent fixation events at a single site, leading to some rate of ancestral misassignment. Since our method to detect selection compares ancestral and derived alleles, we sought to understand how ancestral misassignment would impact our inferences.

Ancestral misassignment will tend to decrease the absolute value of S_β within each frequency bin, and hence decrease our power. To see this, suppose that for a given minor allele frequency x, we misassign ancestral state with probability $1-p$, and $\Psi(x)$ is the number of derived alleles observed at frequency x. The expected value of the the test statistic within this bin is then

$$\mathbb{E}[S_\beta(x)] = \left(p\Psi(x)\overline{\beta}_D + (1-p)\Psi(1-x)\overline{\beta}_A\right) - \left(p\Psi(1-x)\overline{\beta}_A + (1-p)\Psi(x)\overline{\beta}_D\right) \tag{11}$$

Note that if $p = 0.5$, we are randomly guessing at ancestral states, and $\mathbb{E}[S_\beta(x)] = 0$. If $p < 0.5$, then the absolute value of $\mathbb{E}[S_\beta(x)]$ is strictly less than it's true value (*i.e.*, the value that would be observed with no ancestral uncertainty).

While we expect that ancestral uncertainty then only serves to make our analyses more conservative, we also sought to understand its potential impact on the observed relationship between allele frequency and effect size. Because $\Psi(x)$ is generally much larger than $\Psi(1-x)$ for $x < 0.5$ (*i.e.*, there are more rare alleles than common alleles in genomic sequencing data), we expect that the impact of ancestral uncertainty will be greatest at very high derived allele frequencies (*e.g.*, $x > 0.9$). We fit a linear model by regressing the mean effect size on allele frequency for the observed height data for derived alleles in the frequency range 40-60%, and extrapolated this curve out to 100% frequency, supposing that frequencies near 50% were only modestly impacted by ancestral uncertainty. We then simulated the impact of ancestral uncertainty at various levels from $p = 1\%$ to $p = 10\%$. At $p = 10\%$ uncertainty we see a striking resemblance between our simulated data and the observed data. We conclude that moderate levels of ancestral uncertainty are likely responsible for the "S" shaped curve that we observe for many of our phenotypes.

Aggregating phenotype data

We obtained GWAS summary data from several published studies for nine different phenotypes as discussed in the main text [37–44]. We obtained allele frequencies for each allele in each study from the GWAS summary data in the corresponding study. To calculate S_β for each study, we first needed to polarize all the alleles by their derived/ancestral status. We obtained inferred derived/ancestral states for each allele from the 1000 Genomes project [52]. Our permutation method requires that each allele then be assigned to a LD block within the genome [31]. To assign each of these alleles to LD blocks, we also used the 1000 Genomes data to obtain the genomic coordinates for each allele. Alleles for which we could not assign states or LD blocks were excluded.

Calculating S_β and implementing PASTEL

We developed custom software in Python, which is freely available upon request and will be posted on Github. To calculate S_β, we group alleles into 1% frequency bins – without performing this grouping, many high frequency bins would have very few alleles and very noisy estimates of $\overline{\beta}$. We selected 1%

because it strikes a balance between obtaining low standard errors on $\overline{\beta}$ and finely parsing the allele frequency space, but we note that other choices of bin size could potentially improve our power.

To calculate p-values, we perform a two-tailed test by comparing our test-statistic to the empirical null distribution that we obtain from our permutations. We performed 2,000 permutations as described in the main text for each phenotype.

Supplemental Figures

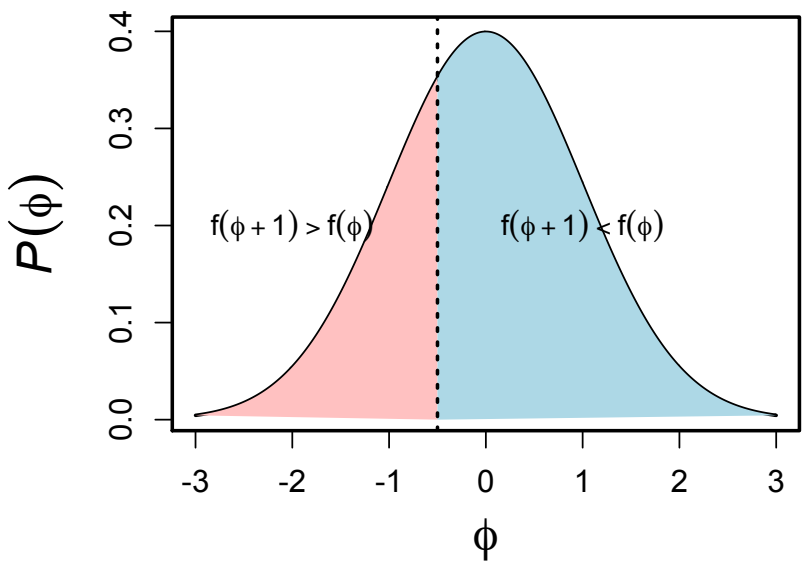

Figure S1: A plot of the distribution of a hypothetical trait under stabilizing selection. Suppose that a large effect allele with $\beta = 1$ is mutated in a random individual in the population. If this allele falls on an individual with a trait value to the right of $-1/2$, the allele will be fitness-decreasing, assuming that the fitness function is symmetric, centered at the mean trait value, and monotonically decreasing away from its optimum. If it falls to the left of $-1/2$, it will be fitness-increasing.

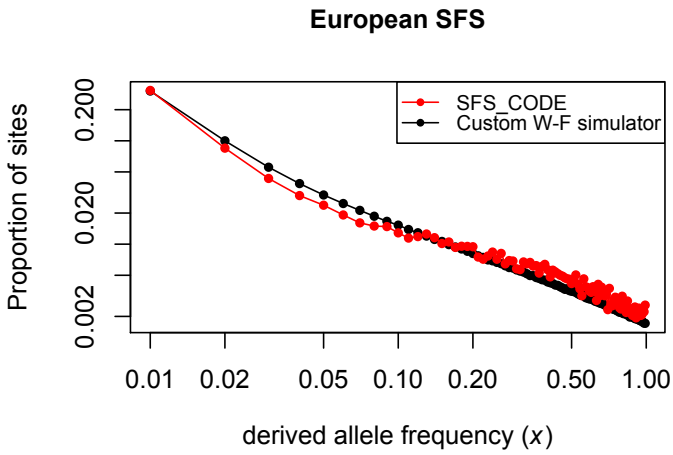

Figure S2: A comparison of simulated frequency spectra for a complex demographic/selection model for our Wright-Fisher simulator (black) and SFS_CODE (red).

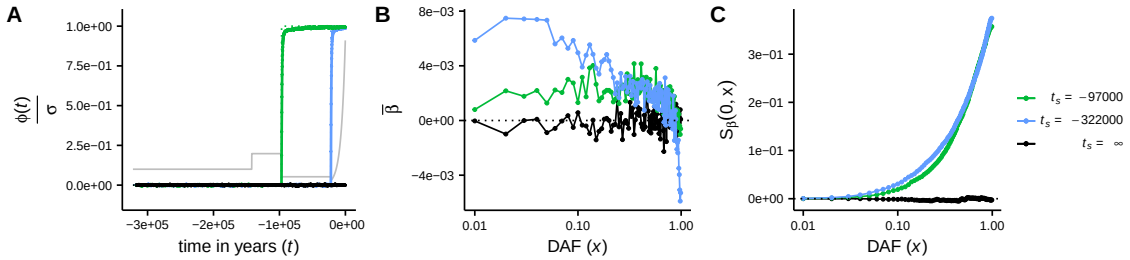

Figure S3: In A, we plot the demographic history of Europe as inferred from human genomic sequences [28] in gray (not to scale). The optimal value of a selected phenotype (ϕ_o) is plotted as a dashed line in black, blue, and green (the black dashed line overlaps with the population mean and is not visible). In the green and blue dashed lines, the optimal value of the phenotype changes by 100% relative to the standard deviation of the phenotype (σ) coincident with bottleneck events corresponding to out-of-Africa (green, $t = -97000$ years ago) or in the ancestral African population (magenta, $t = -322000$). In the solid black, green, and magenta curves we plot the observed value of the mean phenotype, which dynamically approaches the optimal phenotype value. In B, we plot the mean value of the effect size at the end of the simulation ($t = 0$) for each of the models presented in A. In C, we plot $S_\beta(0, x)$ for each model in A. In contrast to Fig. 1, in this plot the mutation rates of trait-increasing and trait-decreasing alleles are equal.

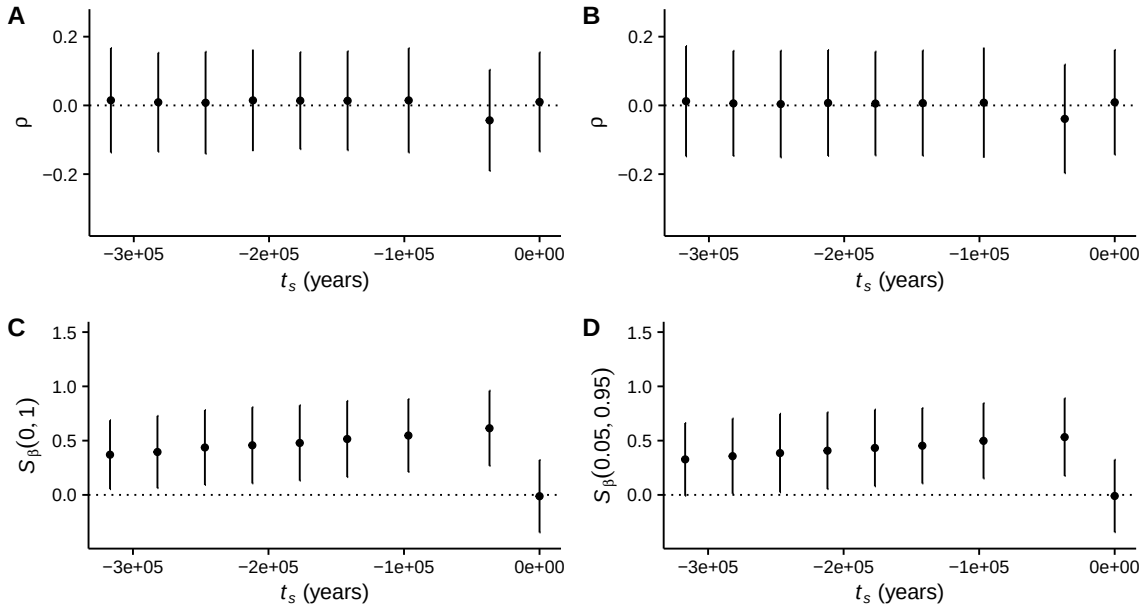

Figure S4: A comparison between the correlation coefficient (ρ) and our selection summary statistic (S_β), computed on all variants (A & C) or only common variants with frequencies greater than 5% (B & D). The points represent the mean over 1,000 simulations, while the bars represent the standard deviation. For each data point, a shift of 100% in the optimal phenotype value occurs at the corresponding time on the x-axis. The neutral null is plotted as a horizontal dashed line.

.

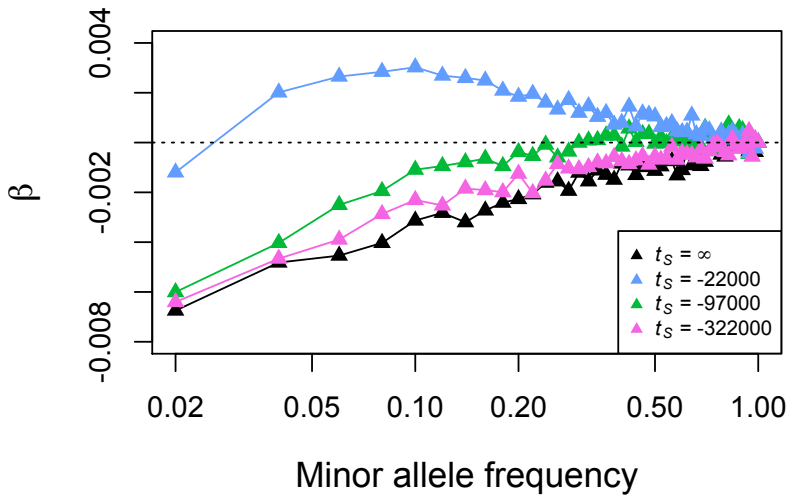

Figure S5: The value of mean β as a function of MAF instead of derived allele frequency for simulations of stabilizing selection with time variable optimum phenotype value. The black curve corresponds to no shift in the optimum, while green and blue correspond to shifts at the out-of-Africa bottleneck (97,000 years ago) and founding of Europe (22,000 years ago), respectively. The magenta curve corresponds to an ancient shift in optimum occurring in the African ancestral population, prior to the out-of-Africa event.

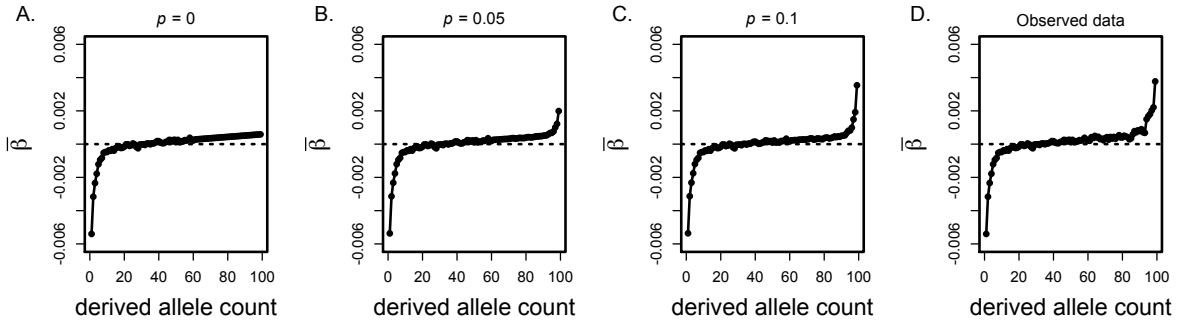

Figure S6: The impact of ancestral uncertainty on the observed value of mean β as a function of derived allele count x. Panel A shows results corresponding to no uncertainty ($p = 0$), in B $p = 0.05$, in C $p = 0.1$, and D shows the observed height data.

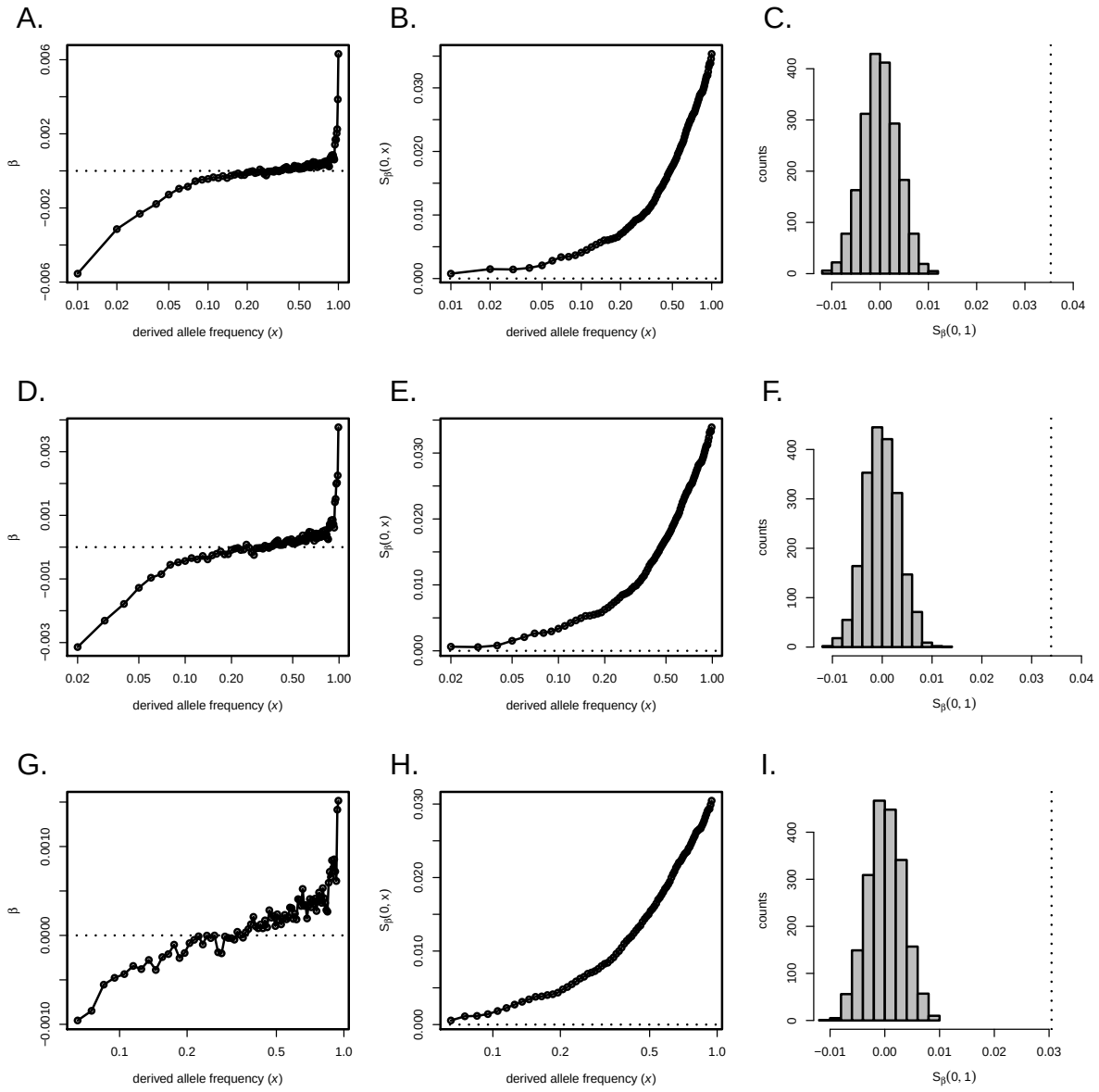

Figure S7: S_β for height. The panels in the left column show the relationship between allele frequency and $\overline{\beta}$, the middle column displays the cumulative value of $S_\beta(x_i, x_f)$, and the right columns show the null distribution of S_β given by our permutation test, PASTEL. Panels A-C correspond to $x_i = 0$ and $x_f = 1$, panels D-F correspond to $x_i = 0.01$ and $x_f = 0.99$, and panels G-I correspond to $x_i = 0.05$ and $x_f = 0.95$

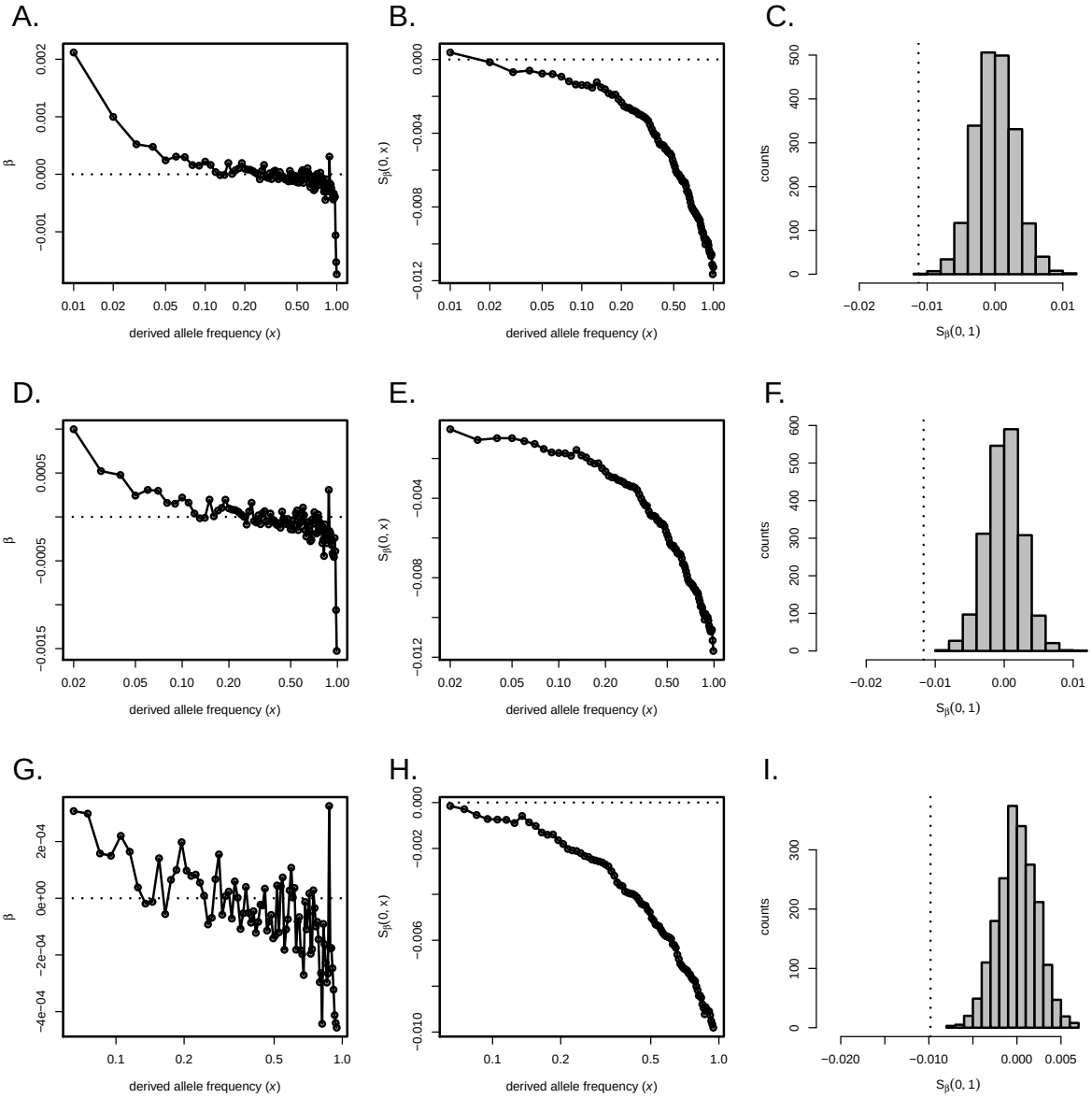

Figure S8: S_β for BMI. The panels in the left column show the relationship between allele frequency and $\overline{\beta}$, the middle column displays the cumulative value of $S_\beta(x_i, x_f)$, and the right columns show the null distribution of S_β given by our permutation test, PASTEL. Panels A-C correspond to $x_i = 0$ and $x_f = 1$, panels D-F correspond to $x_i = 0.01$ and $x_f = 0.99$, and panels G-I correspond to $x_i = 0.05$ and $x_f = 0.95$

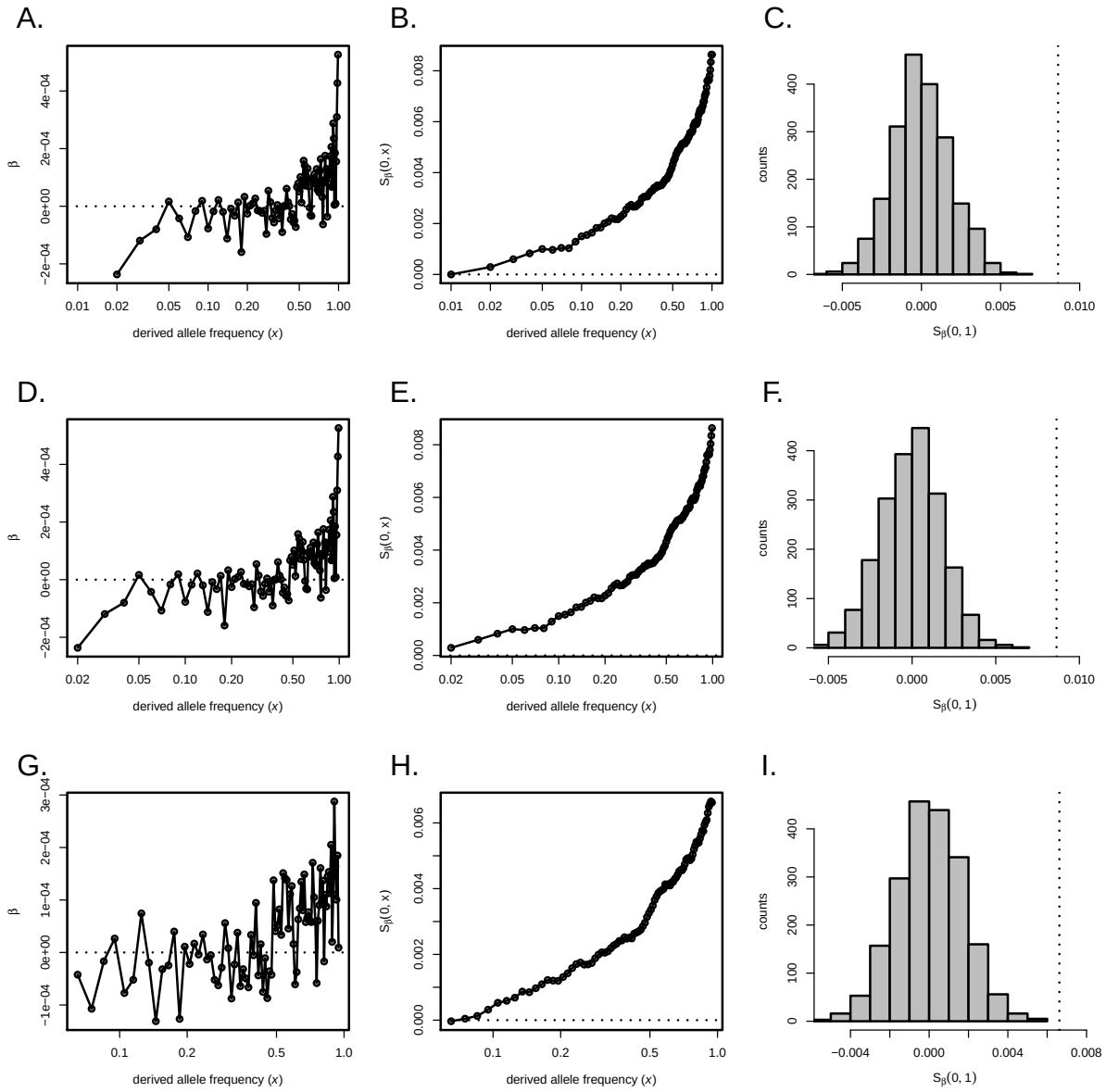

Figure S9: S_β for educational attainment. The panels in the left column show the relationship between allele frequency and $\overline{\beta}$, the middle column displays the cumulative value of $S_\beta(x_i, x_f)$, and the right columns show the null distribution of S_β given by our permutation test, PASTEL. Panels A-C correspond to $x_i = 0$ and $x_f = 1$, panels D-F correspond to $x_i = 0.01$ and $x_f = 0.99$, and panels G-I correspond to $x_i = 0.05$ and $x_f = 0.95$. Note that the results for A-C are the same as D-F because no alleles under 1% were included in the summary data for this study.

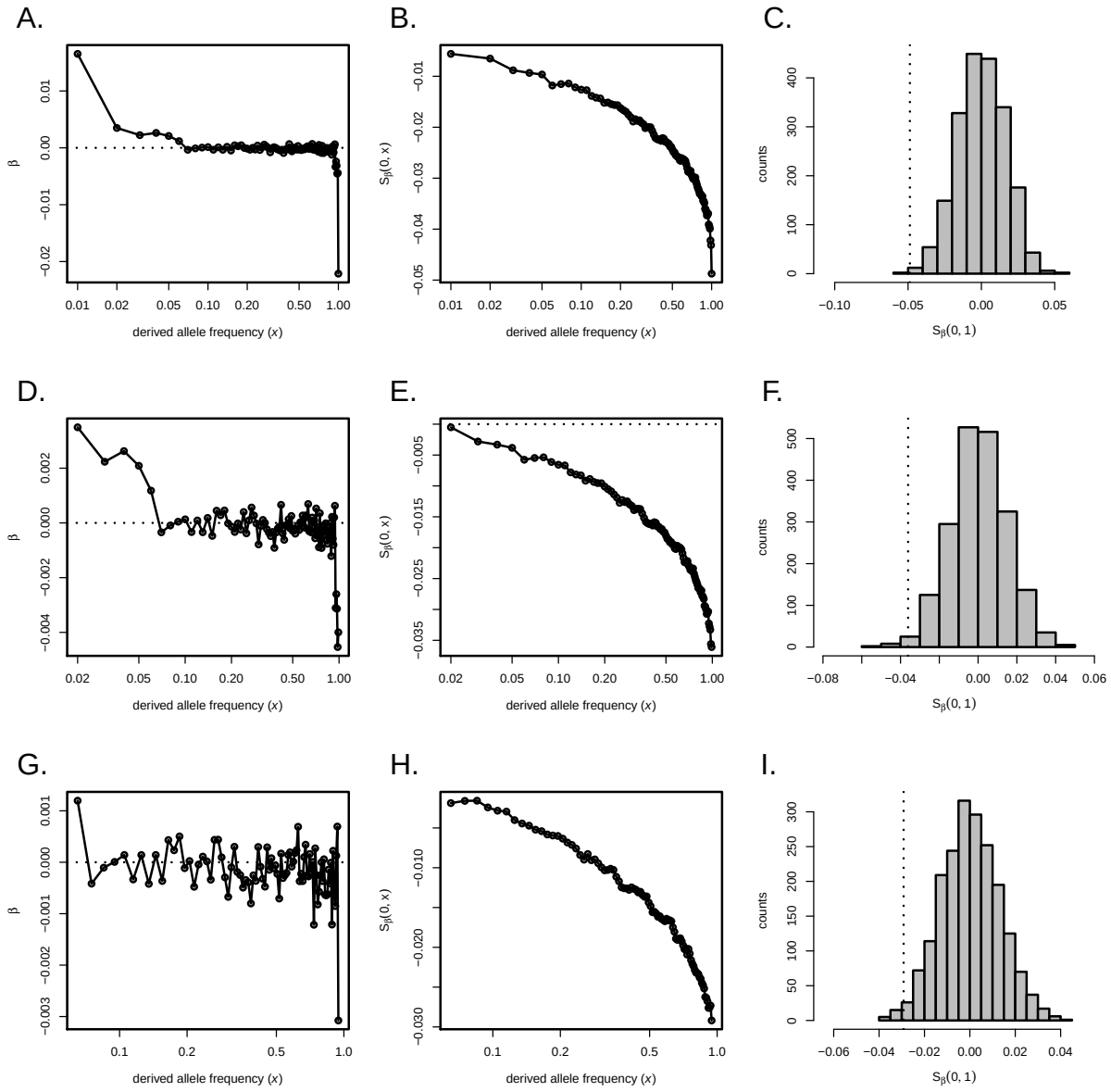

Figure S10: S_β for Crohn's disease. The panels in the left column show the relationship between allele frequency and $\overline{\beta}$, the middle column displays the cumulative value of $S_\beta(x_i, x_f)$, and the right columns show the null distribution of S_β given by our permutation test, PASTEL. Panels A-C correspond to $x_i = 0$ and $x_f = 1$, panels D-F correspond to $x_i = 0.01$ and $x_f = 0.99$, and panels G-I correspond to $x_i = 0.05$ and $x_f = 0.95$

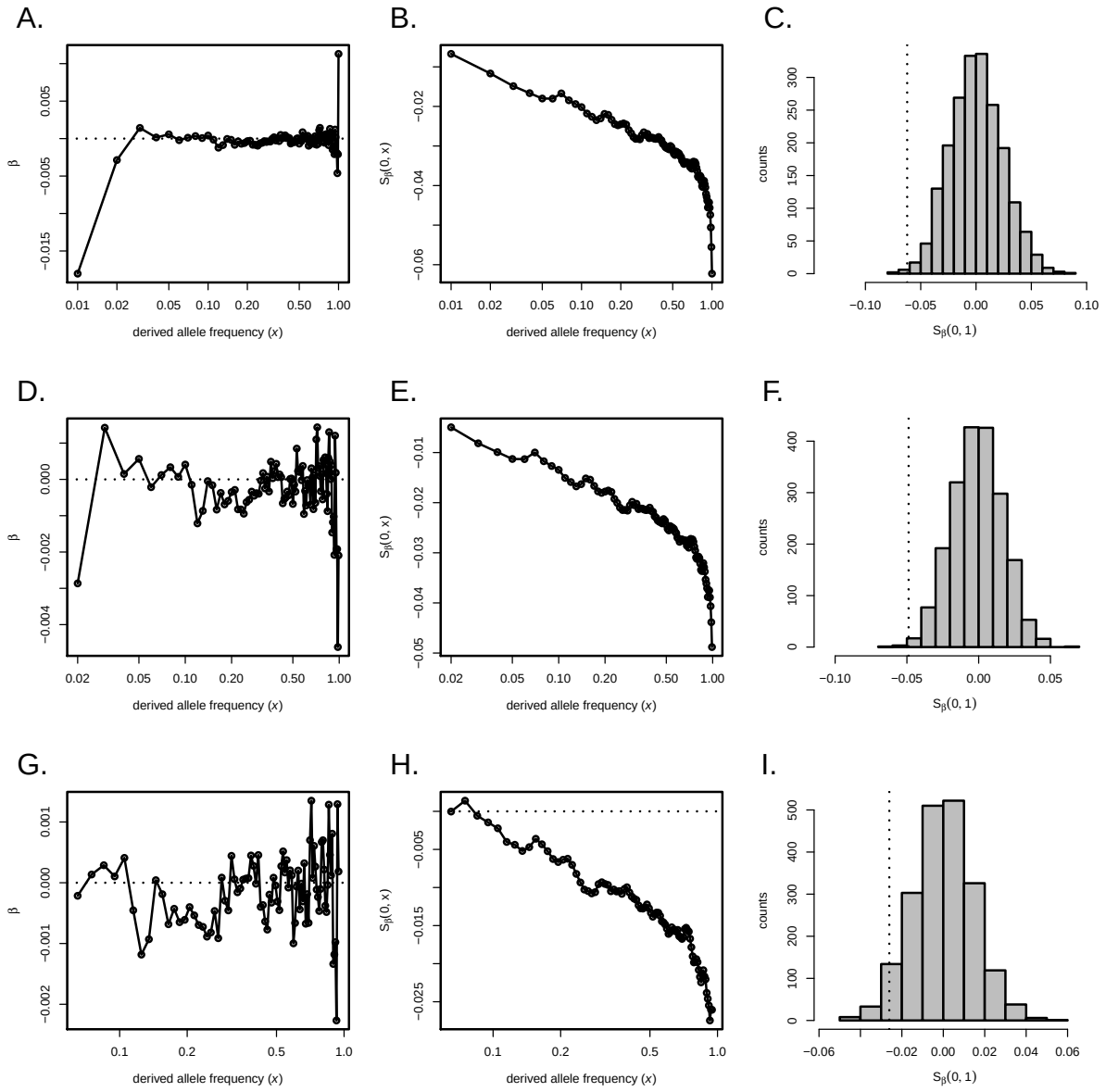

Figure S11: S_β for schizophrenia. The panels in the left column show the relationship between allele frequency and $\overline{\beta}$, the middle column displays the cumulative value of $S_\beta(x_i, x_f)$, and the right columns show the null distribution of S_β given by our permutation test, PASTEL. Panels A-C correspond to $x_i = 0$ and $x_f = 1$, panels D-F correspond to $x_i = 0.01$ and $x_f = 0.99$, and panels G-I correspond to $x_i = 0.05$ and $x_f = 0.95$

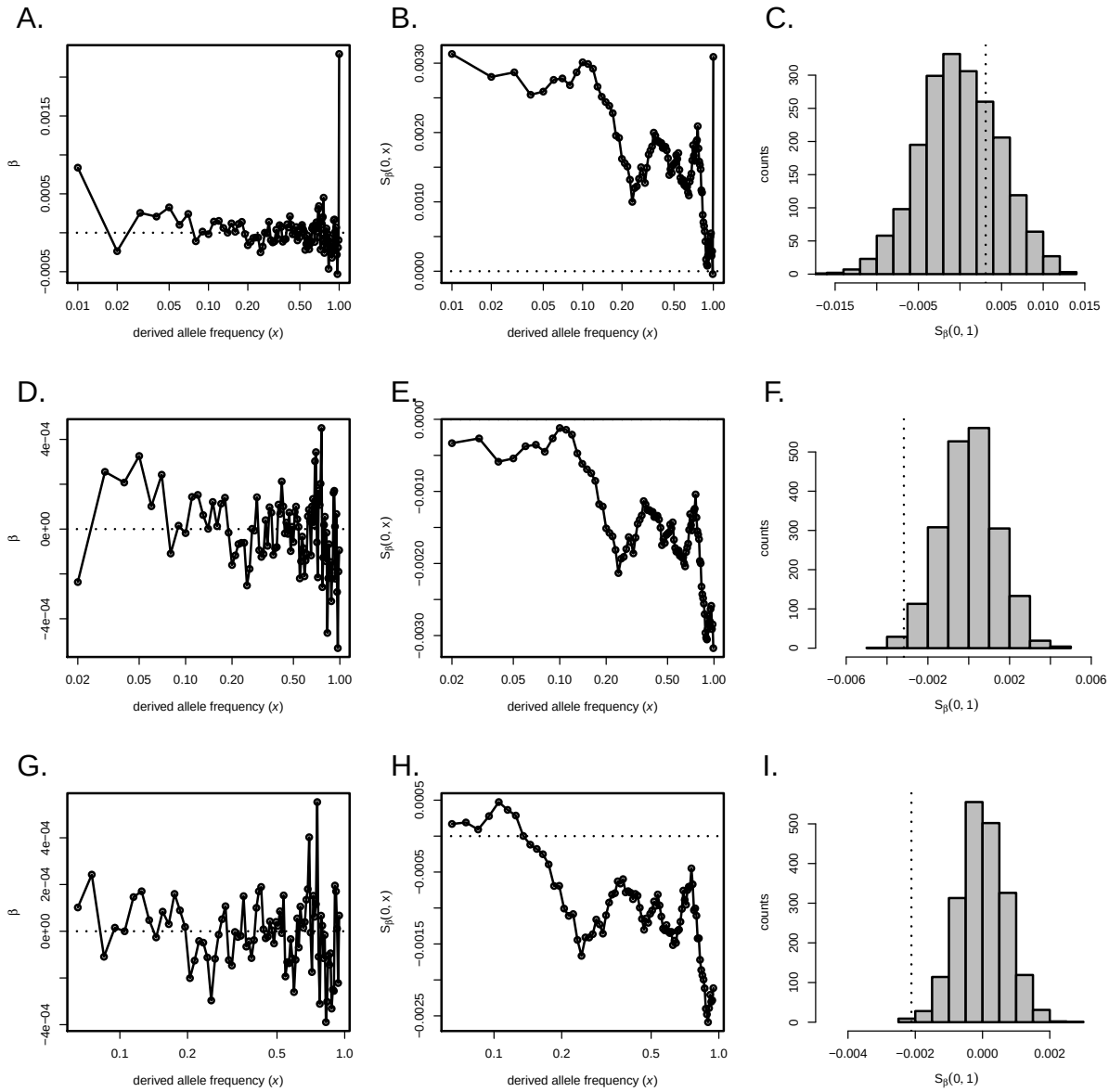

Figure S12: S_β for global lipid levels. The panels in the left column show the relationship between allele frequency and $\overline{\beta}$, the middle column displays the cumulative value of $S_\beta(x_i, x_f)$, and the right columns show the null distribution of S_β given by our permutation test, PASTEL. Panels A-C correspond to $x_i = 0$ and $x_f = 1$, panels D-F correspond to $x_i = 0.01$ and $x_f = 0.99$, and panels G-I correspond to $x_i = 0.05$ and $x_f = 0.95$

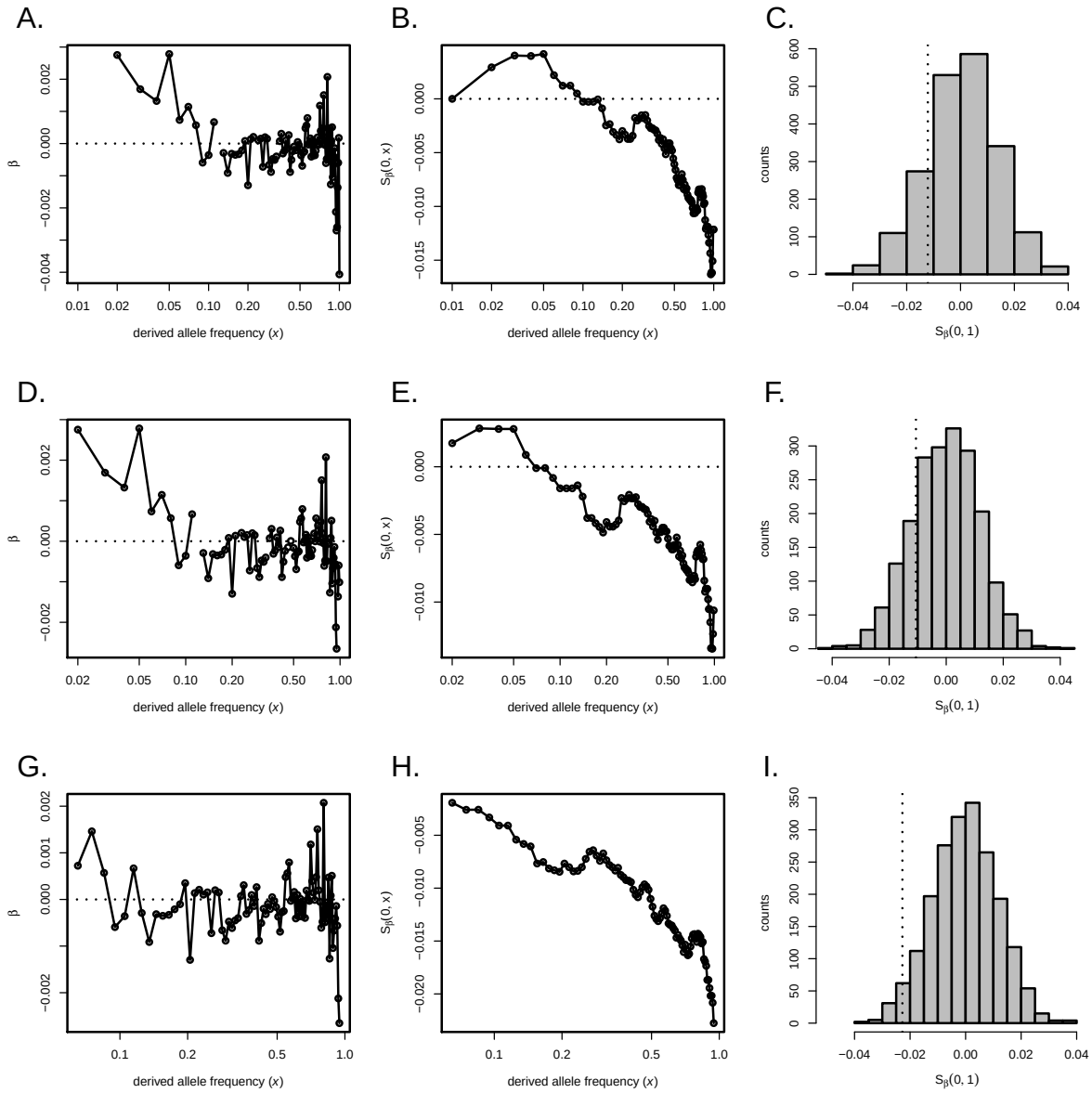

Figure S13: S_β for menopause onset. The panels in the left column show the relationship between allele frequency and $\overline{\beta}$, the middle column displays the cumulative value of $S_\beta(x_i, x_f)$, and the right columns show the null distribution of S_β given by our permutation test, PASTEL. Panels A-C correspond to $x_i = 0$ and $x_f = 1$, panels D-F correspond to $x_i = 0.01$ and $x_f = 0.99$, and panels G-I correspond to $x_i = 0.05$ and $x_f = 0.95$

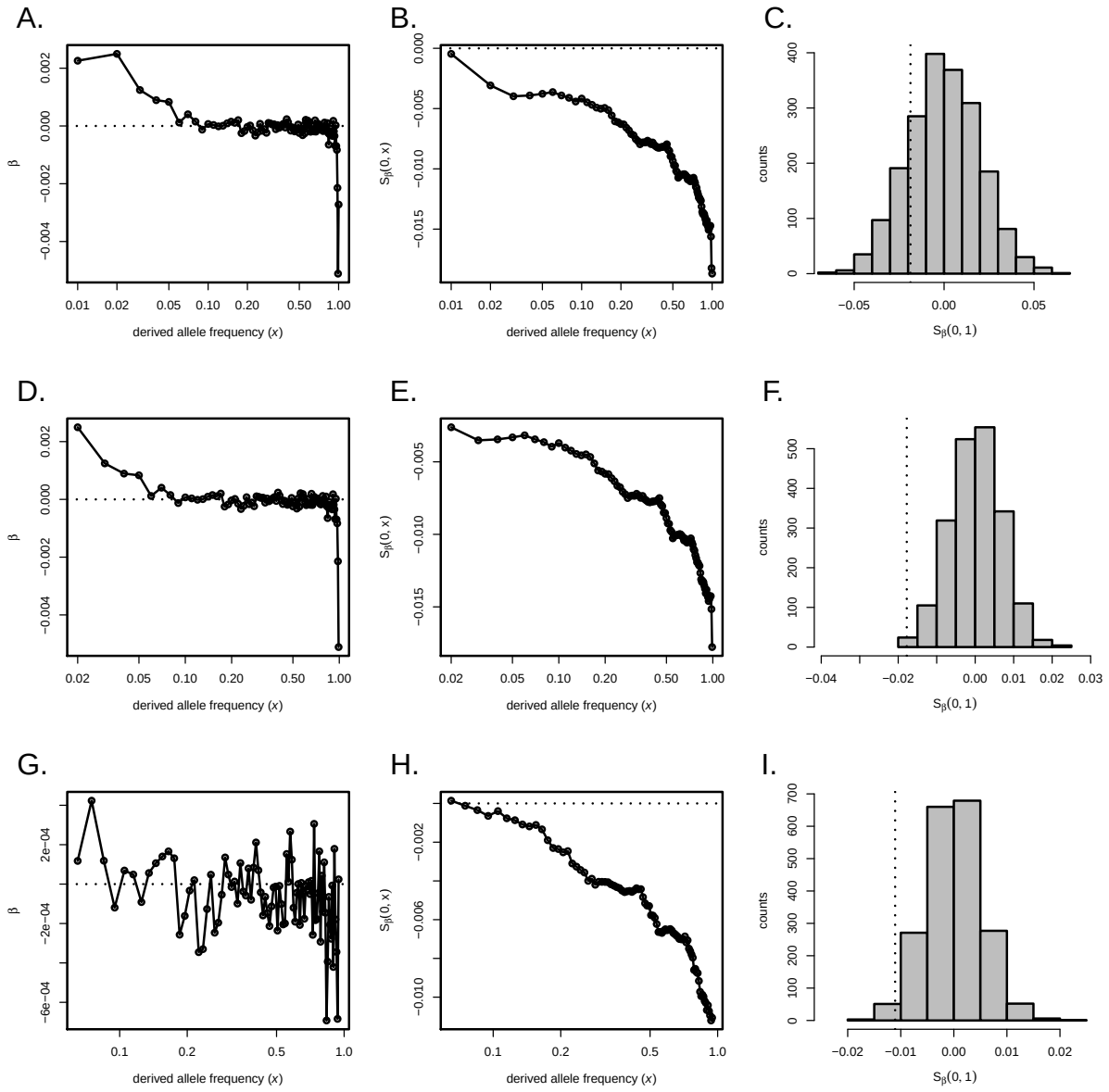

Figure S14: S_β for major depression. The panels in the left column show the relationship between allele frequency and $\overline{\beta}$, the middle column displays the cumulative value of $S_\beta(x_i, x_f)$, and the right columns show the null distribution of S_β given by our permutation test, PASTEL. Panels A-C correspond to $x_i = 0$ and $x_f = 1$, panels D-F correspond to $x_i = 0.01$ and $x_f = 0.99$, and panels G-I correspond to $x_i = 0.05$ and $x_f = 0.95$

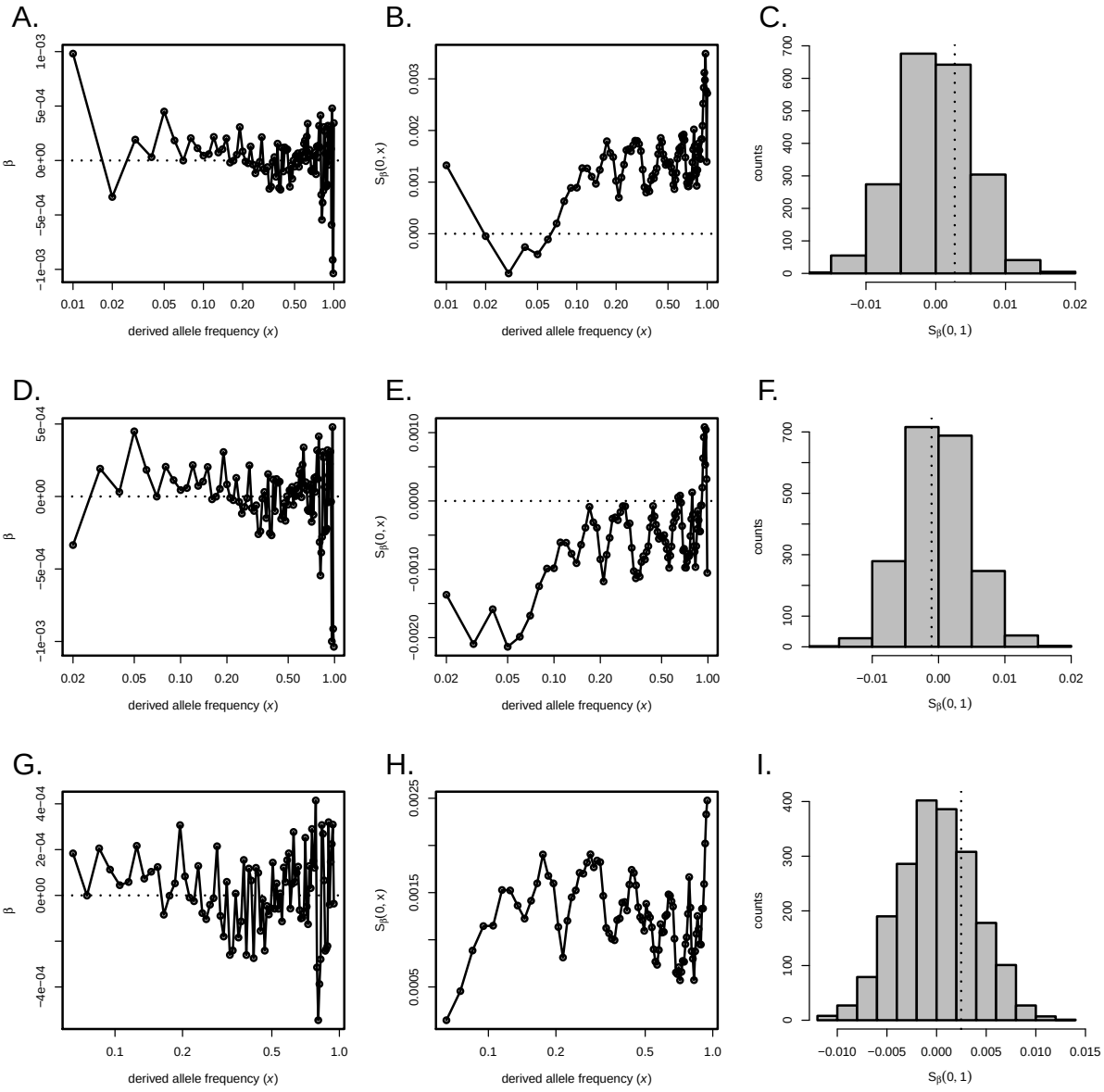

Figure S15: S_β for waist-hip ratio adjusted for BMI. The panels in the left column show the relationship between allele frequency and $\overline{\beta}$, the middle column displays the cumulative value of $S_\beta(x_i, x_f)$, and the right columns show the null distribution of S_β given by our permutation test, PASTEL. Panels A-C correspond to $x_i = 0$ and $x_f = 1$, panels D-F correspond to $x_i = 0.01$ and $x_f = 0.99$, and panels G-I correspond to $x_i = 0.05$ and $x_f = 0.95$