Heritability within groups is uninformative about differences among groups: cases from behavioral, evolutionary, and statistical genetics

Joshua G. Schraiber[1] and Michael D. Edge*[1]

[1]Department of Quantitative and Computational Biology, University of Southern California

February 11, 2024

Abstract

Without the ability to control or randomize environments (or genotypes), it is difficult to determine the degree to which observed phenotypic differences between two groups of individuals are due to genetic vs. environmental differences. However, some have suggested that these concerns may be limited to pathological cases, and methods have appeared that seem to give—directly or indirectly—some support to claims that aggregate heritable variation within groups can be related to heritable variation among groups. We consider three families of approaches: the "between-group heritability" sometimes invoked in behavior genetics, the statistic P_{ST} used in empirical work in evolutionary quantitative genetics, and methods based on variation in ancestry in an admixed population, used in anthropological and statistical genetics. We take up these examples to show mathematically that information on within-group genetic and phenotypic information in the aggregate cannot separate among-group differences into genetic and environmental components, and we provide simulation results that support our claims. We discuss these results in terms of the long-running debate on this topic.

1 Introduction

Understanding the causes and consequences of phenotypic variation is a central goal of many fields of research. Over the last century, researchers in multiple fields have attempted to make a study of the genetic and environmental variation within groups of organisms and use that information to understand the genetic and environmental sources of among-group

*Correspondence: edgem@usc.edu

differences. Such attempts have included suggestions that the explanation of phenotypic differences within groups likely extends to differences among groups (Jensen, 1969), suggestions that observed differences may result from adaptation (for example, Castellani et al., 2023; Salloum et al., 2022; Vitek et al., 2022), and observations that admixed individuals have phenotypes intermediate with respect to members of source populations (Chan et al., 2015; Fernández et al., 2003; Kosoy et al., 2012; Molokhia et al., 2003).

Richard Lewontin critiqued the first of these arguments in a set of well-known thought experiments (1970). In one example, Lewontin imagined two groups of corn plants. Within each group, genotypes vary, but environmental conditions do not. Between groups, there are no average genetic differences, but environmental conditions differ, with one group in more favorable conditions than the other. In this situation, within-group phenotypic variation is genetic (since environmental conditions are the same for all members of the same group), while between-group variation is environmental (since there are no average genetic differences between groups). Lewontin claimed that his thought experiments showed that within-group heritability is irrelevant for determining the causes of among-group differences. Lewontin's argument was a response to Arthur Jensen, who claimed that within-group heritability estimates suggested a genetic explanation for racial differences in scholastic achievement (Jensen, 1969).

In the decades since Lewontin wrote, several researchers have suggested that Lewontin's thought experiment represents a pathological special case. For example, in behavioral genetics, DeFries (1972), replying to both Lewontin and Jensen, built on earlier work by Lush (1949) to propose the "between-group heritability," h^2_B, aiming to quantify the relative influence of genetic and environmental factors on phenotypic differences among groups. DeFries argued that Lewontin's thought experiments correspond to specially picked edge cases and suggested that in other cases, h^2_B provides a useful framework for thinking about group difference.

Here, we show that in contrast to DeFries' suggestion, the rationale underlying Lewontin's thought experiments is general and poses challenges for the interpretation of h^2_B whenever environmental differences among groups are not controlled or understood. In particular, we show that, in line with Lewontin's intuition, the heritabilities of phenotypic differences between groups are not constrained by heritability within groups: perfect knowledge of within-group heritability provides no information about between-group heritability. Crucially, even if the heritability of between-group differences is estimated correctly, it leaves the direction of the genetic and environmental components of phenotypic difference unclear.

We also expand Lewontin's critique to other approaches to similar questions. In evolutionary biology, a statistic called P_{ST} (Brommer, 2011) is used to test whether phenotypic differences among groups of wild organisms are consistent with neutral evolution or instead might be better explained by natural selection. We show that a parameter used in sensitivity analyses involving P_{ST} can be understood as a function of h^2_B. Using our results about h^2_B and their relationship to P_{ST}, we show the sensitivity analysis that is typically under-

taken when using P_{ST} to test for evidence of divergent selection on a phenotype may be conservative, especially if the phenotypic differentiation among groups is small. Moreover, we highlight the difficulty of interpreting Q_{ST} and P_{ST} in the presence of environmental differences between groups, adding to recent arguments by Harpak and Przeworski (2021).

In genetic anthropology and statistical genetics, variation in ancestry among admixed individuals is sometimes used to make claims about the sources of trait variation among groups. For example, correlations between an individuals' phenotype and ancestry proportions may be taken as evidence that an observed phenotypic difference between source populations is genetically caused. As we show and has been mentioned previously as a cautionary point (e.g., Chakraborty & Weiss, 1986; Fernández et al., 2003; Kosoy et al., 2012), the genetic effect of ancestry may be confounded with environmental variation affecting the phenotype. Approaches to heritability estimation based on patterns of local ancestry in admixed populations (Zaitlen et al., 2014)—as opposed to global ancestry proportions—have also been used to calibrate claims about the sources of group differences (Zaidi et al., 2017). We show that the variation due to local ancestry that they leverage has no clear relationship with between-group trait differences.

We discuss our results in terms of their shared origin in gene-environment confounding and relate them to previous claims (Block, 1995; Feldman & Lewontin, 1975; R. Lewontin, 1970). The complications we point out arise under simple additive models, and the situation is only more complicated in models with interactions among genetic or environmental influences. We defer a brief consideration of gene-by-environment interaction to the discussion section.

2 Between-group heritability is unconstrained by within-group heritability and underdetermines genetic and environmental differences among groups

Lush (1949) developed the between-group heritability, h^2_B, to quantify the relative influence of genetic and environmental factors on a trait that differs among groups. Lush's interest was in heritability of differences among families of organisms subject to artificial selection, with the goal of weighting family membership appropriately when selecting organisms for breeding. Lush worked in applications in which many aspects of among-group variation in environment could be controlled or randomized.

Consider a phenotype value Y formed from additive genetic and environmental components, i.e. $Y_i = G_i + E_i$ for individual i. It is standard to define the (narrow-sense) heritability as $h^2 = V_A/V_P$, where the additive genetic variance V_A is the variance of the genetic component (G_i), assuming no epistasis or dominance, and V_P is the variance of the phenotype. If the sample is divided into groups, we can also define variance proportions accounted for by group membership in either traits themselves or in the genetic components of those traits. For example, with respect to the phenotypic variance, we can define

$t = \sigma_B^2/V_P$, with σ_B^2 the between-group component of the phenotypic variance (Falconer & Mackay, 1996). Call the analogous variance proportion for the genetic variance r. Then, when groups are large, the "between-group" heritability is

$$h_B^2 = \frac{rV_A}{tV_P}. \tag{1}$$

Subsequently, DeFries (1972) took up this formalism to argue against the generality of Lewontin's thought experiment. With some algebra, he showed that the between-group heritability is mathematically related to the within-group heritability, h_W^2. Specifically, defining r and t as in the previous paragraph, he showed that

$$h_B^2 \approx h_W^2 \frac{(1-t)r}{(1-r)t}, \tag{2}$$

if the groups are large. DeFries claimed that Lewontin's thought experiments—one described above, and another involving inbred lines reared in different environments—could be explained as pathological cases in which either r or t is equal to 0, and suggested that the relationship in equation (2) could be used to help interpret observed group differences in other cases. He displayed a table showing between-group heritabilities implied by various values of r, conducting a sort of sensitivity analysis.

An immediate problem with application of equation (2) is that in the absence of environmental controls, it is impossible to estimate r, the proportion of the variance in the genetic component of the trait accounted for by group membership, a point that DeFries conveyed (DeFries, 1972). Despite this, some researchers in behavior genetics, including DeFries himself, took up DeFries' formalism and argued for its relevance for understanding variation among groups of humans, generally in the context of discussions of racial differences in IQ or scholastic achievement (Plomin & DeFries, 1976; Warne, 2020, 2021). In particular, Jensen (1972a, 1972b, 1975) repeatedly cited the definition of between-group heritability as a linear function of within-group heritability in arguments for the value of heritability estimates for understanding sources of difference among groups. Feldman and Lewontin (1975) criticized the expression for h_B^2 in equation 2 as tautological, in that r is better viewed as a term that simply relates the definitions of h_B^2 and h_W^2 rather than being an independently varying ingredient of a prediction of h_B^2 from h_W^2. Here, we concur with Feldman and Lewontin's interpretation, showing that, despite the appearance of equation (2), the between-group heritability is unconstrained by the within-group heritability. Thus, within-group heritability, on its own, provides no information about between-group heritability (Supplementary Text, section S1.1-S1.2). We also show that even if r is known, h_B^2 still underdetermines the group differences, such that opposing interpretations of among-group variation are possible under the same (known) r and h_B^2.

In the specific case of two groups, the important quantities are δ_G, the difference between the mean genetic values of the two groups, and δ_E, the difference between the mean

environmental values of the two groups. Then, under a standard quantitative-genetic model,

$$h_B^2 = \frac{\delta_G^2}{(\delta_G + \delta_E)^2}. \tag{3}$$

Equation (3) is useful to highlight several properties of between-group heritability that generalize to cases with more than two groups. First, this expression for h_B^2 can be derived without reference to within-group heritability or any quantities that contribute to within-group heritability (section S1.2), indicating that even perfect knowledge of heritability within groups conveys no information about genetic differences between groups without further information or assumptions.

Second, it shows that the between-group heritability is not bounded between 0 and 1: in particular, if the environmental and genetic effects point in different directions, such that the magnitude of the phenotypic deviation, $\delta = \delta_G + \delta_E$ is smaller than the magnitude of the genetic deviation, δ_G, then the between-group heritability will be larger than 1. We illustrate this in Figure 1A, which shows the value of h_B^2 across a range of genetic and environmental deviations. One striking feature of the behavior of h_B^2 is that it rapidly approaches infinity when $\delta_E \approx -\delta_G$, and that larger genetic deviations result in a larger range of environmental deviations over which $h_B^2 > 1$.

Finally, the between-group heritability is uninformative about the *direction* of environmental and genetic differences. For the two-population case, this is reflected in the fact that given knowledge of the phenotypic deviation, δ, and the between-group heritability, there are two possibile directions of genetic effects that are consistent with the data, namely $\delta_G = \pm\delta\sqrt{h_B^2}$. In the Supplementary Text (section S1.3), we show that with more than two groups, there are an infinite number of possible configurations of genetic and environmental deviations that are consistent with a given h_B^2. To give some intuition for this result, we start with an alternative representation of the two-group scenario for groups of equal size (Figure 1B). The two axes indicate genetic deviations of the two group means from the grand mean, $\Delta_{G,1}$ and $\Delta_{G,2}$. Conditional on a given mean phenotypic difference between the groups, δ, obtaining a given value of h_B^2 requires that the sum $\Delta_{G,1}^2 + \Delta_{G,2}^2$ equal a constant, solutions of which are represented by the orange circle. Additionally, because $\Delta_{G,1}$ and $\Delta_{G,2}$ are deviations from the grand mean, they must sum to zero, giving the green line corresponding to $\Delta_{G,1} + \Delta_{G,2} = 0$. The two purple points where the line and circle intersect represent the two configurations of the between-group genetic difference that lead to the same value of h_B^2.

With a larger number of groups, the situation is entirely analogous, only in higher dimension. Figure 1C shows a schematic of the three-group case. The axes again represent the genetic deviations of each group from the grand mean. The combinations of genetic deviations that results in the same h_B^2 (conditional on phenotypic differences) is now an orange sphere, rather than a circle. And the genetic deviations from the grand mean consistent with the condition that all genetic deviations sum to zero fall on the blue plane

(rather than a line). The intersection of the sphere and the plane, shown in purple, gives a circle of genetic deviations on which all have the same h_B^2 while also being consistent as deviations from a grand mean. With more than three groups, the sphere becomes a hypersphere, the plane becomes a hyperplane, and the same basic relationships hold.

Within-group heritabilities, on their own, provide no information about the causes of between-group differences, in part because any configuration of group means for the genetic contributions to the trait is possible under some value of r. The value of r itself can be studied under evolutionary models, and in fact it is closely related to the quantity Q_{ST} studied in evolutionary quantitative genetics, which we consider in the next section.

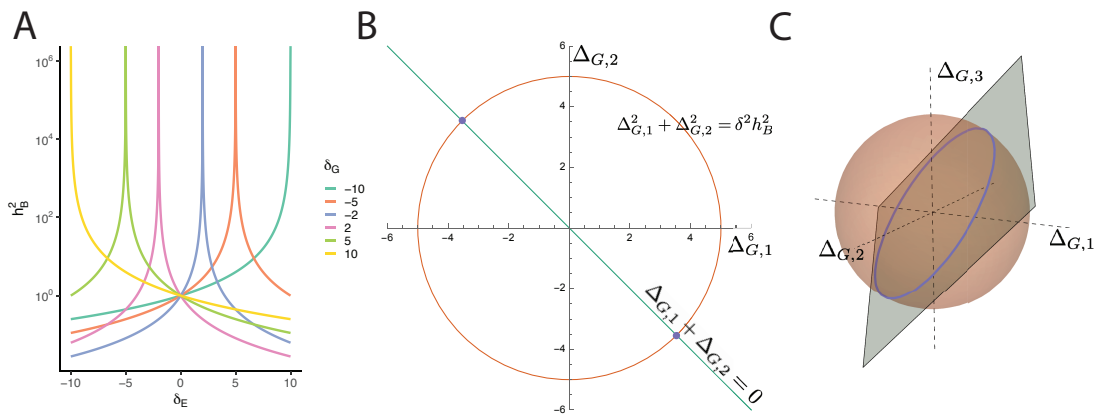

Figure 1: Properties of h_B^2. A) The behavior of h_B^2 in two populations for different values of the genetic deviation δ_G and environmental deviation δ_E. Each curve corresponds to a different genetic deviation. B) Two genetic deviations that are consistent with an observed h_B^2 with two populations. The circle shows all genetic deviations that hold the genetic variance constant, while the line shows all genetic deviations that sum to 0. The intersection of the line and circle shows the genetic deviations that are consistent with a given h_B^2. C) An infinite number of genetic deviations consistent with an observed h_B^2 with three populations. The sphere shows all genetic deviations that hold the genetic variance constant, while the plane shows all genetic deviations that result in the genetic deviations summing to 0. The intersection of the sphere and the plane, indicated by the purple circle, contains the values of genetic deviations from the grand mean that will hold h_B^2 constant.

3 P_{ST} analyses can understate or overstate the possibilities for adaptive differentiation

The statistic Q_{ST} is a measure of the genetic component of trait variation among populations. For diploids, in terms of the variance proportion r discussed in the previous section,

it is equal to $Q_{ST} = r/(2-r)$ (Edge & Rosenberg, 2015). When traits evolve neutrally, Q_{ST} is expected to be equal to F_{ST} at putatively neutral sites, which measures the genetic variation among populations (Whitlock, 1999). Thus, by testing whether Q_{ST} is significantly greater than F_{ST}, it is possible to determine whether there is more genetically explained trait variation among populations than expected under neutral evolution (Leinonen et al., 2013; Prout & Barker, 1993; Spitze, 1993; Whitlock, 2008). Such a pattern is consistent with divergent selection causing trait differences among groups. On the other hand, lower-than-expected values of Q_{ST} may be taken as evidence of stabilizing selection on a shared optimum.

However, to estimate Q_{ST} requires arduous common-garden experiments that eliminate environmental confounds. Such common-garden experiments may be difficult or impossible to carry out in some organisms. Thus, it is common in the literature to use a version of Q_{ST} calculated from phenotypes measured on wild organisms rather than organisms raised in a common garden (for example, Castellani et al., 2023; Salloum et al., 2022; Vitek et al., 2022). This statistic computed from observations of wild organisms, often labeled P_{ST}, is a function of the between-population phenotypic variance, σ_B^2, the within-population phenotypic variance, σ_W^2, as well as parameters a and b, which are interpreted as determining the fraction of between- and within-population phenotypic variance that is due to genetics, respectively. For diploids,

$$\begin{aligned} P_{ST} &= \frac{a\sigma_B^2}{a\sigma_B^2 + 2b\sigma_W^2} \\ &= \frac{c\sigma_B^2}{c\sigma_B^2 + 2\sigma_W^2} \end{aligned}$$

where the parameter $c = a/b$. Because σ_B^2 and σ_W^2 can be estimated from purely phenotypic data but c cannot, it is common practice in the literature to explore values of c between 0 and 1, seemingly motivated by a suggestion that the genetic component of between-population differences must be proportionally smaller than the genetic component of within-population differences. A well appreciated point is that if the choice of c is too large, then P_{ST} analysis can be anticonservative, leading to spurious inferences of divergent selection. In effect, environmental differences are mistaken for genetic ones, leading to a false rejection of neutral evolution. In the Methods, we show that the coefficients for P_{ST} are $a = h_B^2$, the between-group heritability, and $b = h_W^2$, the within-group heritability, so that the correct value of c to obtain $P_{ST} = Q_{ST}$ is

$$c^* = \frac{h_B^2}{h_W^2}. \tag{4}$$

As shown above, h_B^2 is unconstrained by any measure of genetic and environmental variability within groups. Moreover, because h_B^2 is not constrained to be between 0 and 1, a

sensitivity analysis that explores values of c between 0 and 1 may lead to extremely conservative tests for divergent selection. Finally, because both Q_{ST} and P_{ST} depend on the variance of between-population means, any evolutionary scenarios that produce identical h^2_B will also produce identical Q_{ST} and P_{ST}, indicating that knowledge of P_{ST} or Q_{ST} is not sufficient for disentangling the directions of environmental and genetic effects.

4 Simulated example: evolution in the face of an environmental and optimum shift

To demonstrate the impact of our results on evolutionary analyses, we performed population-genetic simulations using SLiM (Haller & Messer, 2019). In particular, we developed a model in which very different underlying evolutionary forces result in identical h^2_B and Q_{ST}, indicating that these statistics are insufficient for characterizing evolutionary history and may be misleading in some circumstances. The scenario we considered is outlined in Figure 2 panels A and D are and inspired by a recent perspective by Harpak and Przeworski (2021): a population evolves with a trait under stabilizing selection until a population split. In one of the populations, the environment and stabilizing selection remain constant; however, in the other population, the environment shifts, along with the optimum trait value. We explore two scenarios: one in which the environment shifts in the same direction as the fitness optimum for the trait ("concordant", panel A), and one in which they shift in opposite directions ("discordant", panel D). We use the same optimum shift in both cases, so that the simulations result in a phenotypic difference of $\delta = 1$, and set the environmental shift to maintain $h^2_B = 9$ at equilibrium.

Figure 2 panels B and E show the evolution of the difference in phenotypic and genotypic values between the two populations over the course of adaptation. In both directions of environmental shift, the phenotypic differences converge to $\delta_P = 1$, consistent with identical optima shifts in both situations. However, the genetic differences display opposite dynamics, depending on the direction of the environmental shift. If the environmental shift is in the same direction as the optimum shift (panel B), then the adapting population will have a mean genetic value that is less than that of the population that did not experience an environmental shift; in the other scenario (panel E), the adapting population will have a mean genetic contribution to the trait that is greater than the population that did not experience a shift. Thus, to compare the impact of stabilizing selection with environmental change, both the common-garden genetic values *and* the *in situ* phenotypic values are necessary.

Figure 2 panels C and F show that, $h^2_B \approx 9$ at equilibrium by construction. However, the two scenarios approach that value in different ways. When the environment and optimum shift in the same direction, the average approach to equilibrium is monotonic and smooth (panel C). However, when the optimum and environment shift in opposite directions, h^2_B becomes large when the phenotypic difference between the populations is near 0, and comes

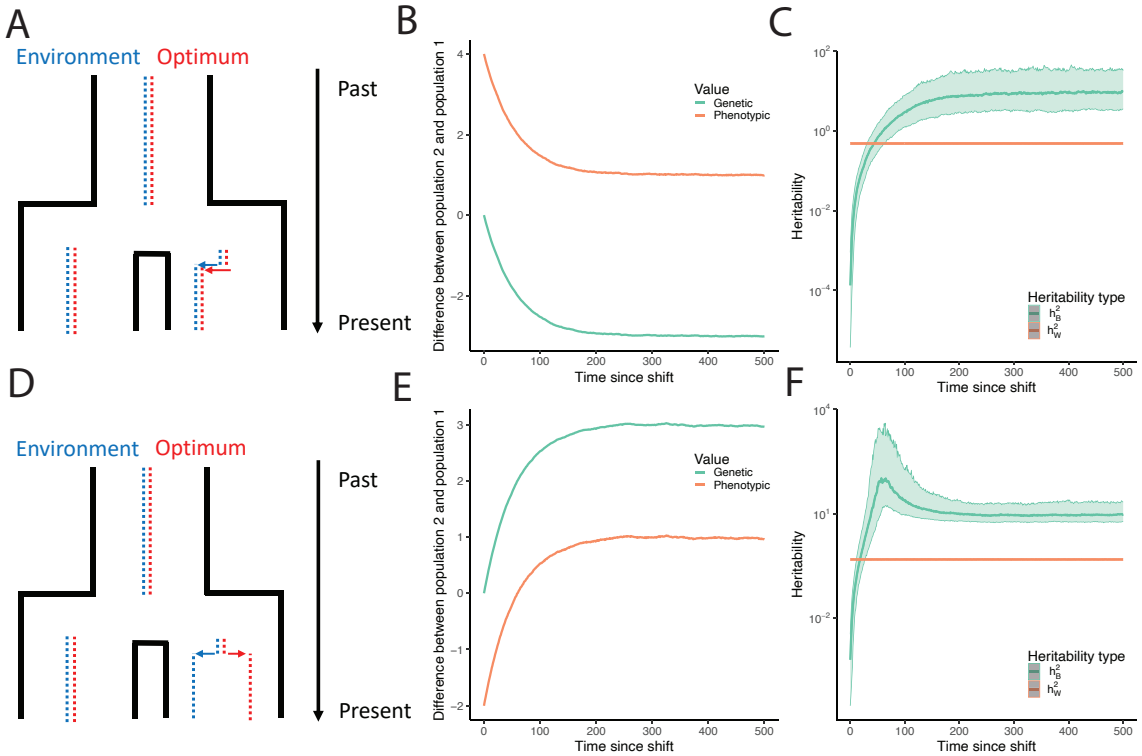

Figure 2: Simulation of a evolutionary scenarios. A, D) Description of the evolutionary scenarios. A population evolves until it splits into two subpopulations of equal size. At the time of split, one of the subpopulations experiences both an environmental shift (in blue) and an optimum shift (red). The environmental shift may either be in the same direction as the optimum shift (concordant, panel A) or in an opposite direction (discordant, panel D). B, E) Evolution of genetic and phenotypic values over time. The horizontal axis shows the number of generations since the split, and the vertical axis shows the value of either the genetic difference between the two populations or the phenotypic difference between the two populations. Concordant shifts are shown in Panel B, discordant shifts in panel E. C, F) Evolution of between- and within-population heritability. The horizontal axis indicates time since the population split, and the vertical axis shows the value of different summaries of phenotypic variation, h^2_B and h^2_W; median values are shown with a solid line, and the shaded band indicates the range of values between the 10th and 90th percentiles. Concordant shifts are shown in panel C, discordant shifts in panel F.

back down from infinity to settle at the equilibrium value (panel F). It is also much noisier across simulation replicates, as seen by the middle 80% of the distribution spanning 3 orders of magnitude. This is mostly due to the denominator being near zero for a large portion of the simulations (Figure S1). On the other hand, the long-term expected value of h^2_W is constant across the simulations and identical regardless of the direction of the environmental shift, as expected, because within-population variability will not change when there is an optimum shift. As emphasized in the previous section, knowledge of h^2_B at any time during adaptation is not sufficient to characterize the direction in which selective forces are acting.

Finally, we examined P_{ST} calculations and their impact on the sensitivity of a test to reject neutral trait evolution in these scenarios. In each simulation, we computed P_{ST} across an array of values of c and also computed the true Q_{ST}. Figures 3A and C show the behavior of the statistics and the test of neutral evolution, respectively, when the the environment and optimum trait value shift in the same direction. Here, the behavior of P_{ST} for a given value of c is monotonic, but at different points during the course of adaptation, the same value of c can produce either a conservative or anti-conservative estimate of Q_{ST} from P_{ST}. Early in adaptation, there should be very little genetic signal of adaptation, and Q_{ST} rejects the null hypothesis approximately 5% of the time at the 5% significance threshold, showing that the test is well calibrated. However, because the environmental shift results in large phenotypic differences, every P_{ST} test initially results in high false positive detection of adaptation. Subsequently, as adaptation proceeds, Q_{ST} detects adaptation with high power, although it loses power after approximately 300 generations, because the neutral divergence between populations increases substantially. However, for values of $c < 1$, as typically explored in the literature, P_{ST} quickly becomes smaller than Q_{ST}, and tests using P_{ST} fail to reject the null hypothesis significantly earlier into adaptation. Figures 3B and D shows that when the environment and optimum shift in opposite directions, the behavior of Q_{ST} remains monotonic, but the behavior of P_{ST} for a given c becomes non-monotonic, reaching a minimum when the phenotypic difference between populations is small and then increasing again. This results in complicated behavior of the P_{ST} test in comparison with the Q_{ST} test, in which P_{ST} can go from anti-conservative to conservative, and then back to anti-conservative. Finally, the trajectories of Q_{ST} are identical regardless of the direction of the environmental shift, in line with the fact that knowledge of Q_{ST} is not informative of the direction of environmental and genetic shifts.

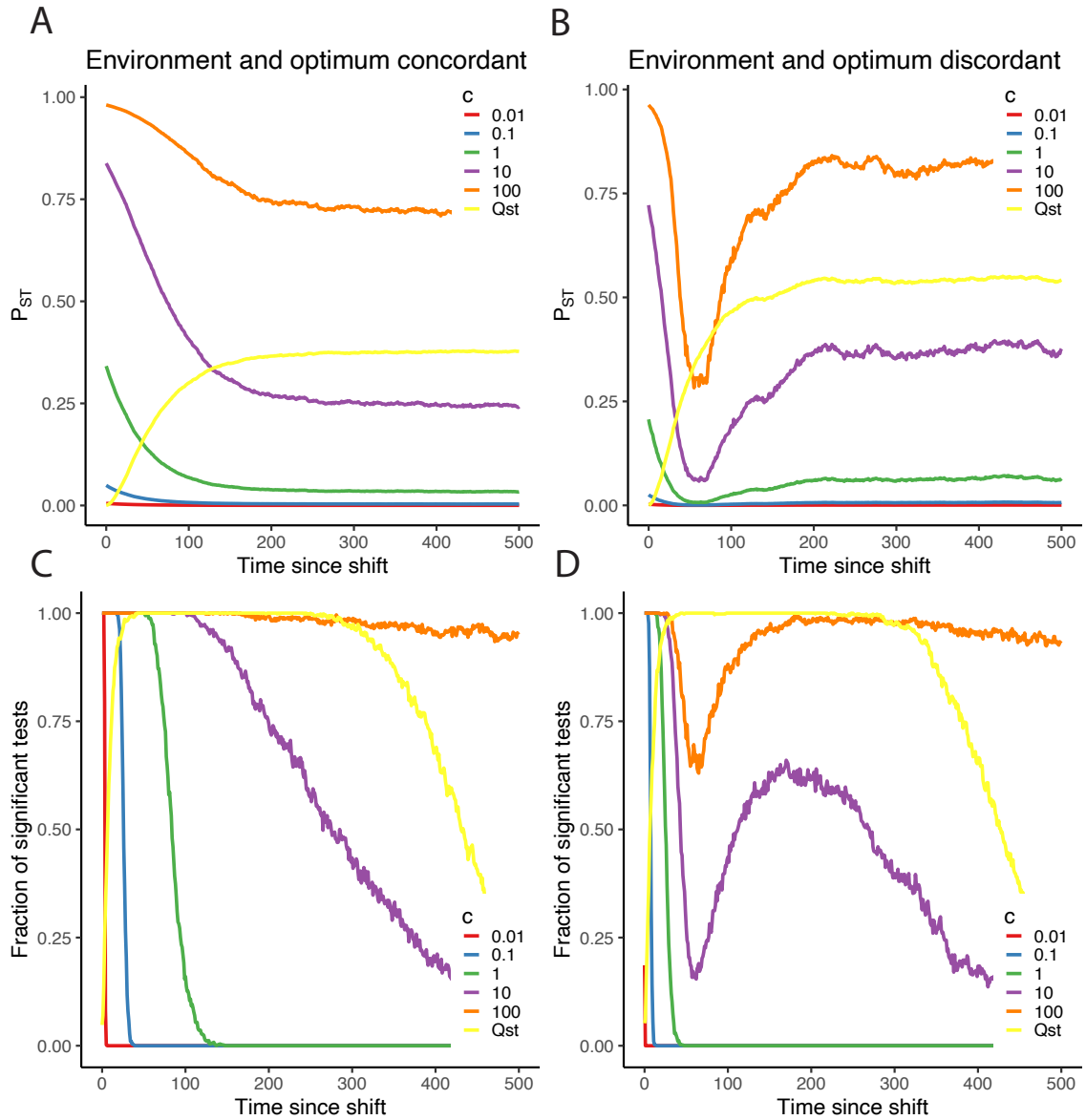

Figure 3: Evolution of P_{ST} and Q_{ST}. In the top row, the horizontal axis shows the time since the population split, and the vertical axis shows the value of either Q_{ST} or P_{ST}, with the different colors indicating different values for the choice of c in the P_{ST} formula. In the bottom row, the horizontal axis shows the time since the population split, and the vertical axis shows the fraction of tests in which the null hypothesis of neutral evolution is rejected at the 5% level because $Q_{ST} > F_{ST}$. The colors are the same as the top row. A, C) The environment and optimum shifts are concordant. B, D) The environment and optimum shifts are discordant.

5 Trait correlations with global ancestry or genome-wide local-ancestry sharing do not reveal the causes of between-group difference

5.1 Global ancestry

We have seen that efforts toward quantifying the genetic contribution to differentiation among groups can be confounded by environmental variation. Some have suggested that admixed individuals present opportunities to resolve this difficulty and estimate the genetic component of between-group differences. One suggested approach is to examine the correlation between trait values and global ancestry proportions among admixed individuals (Chakraborty & Weiss, 1986; Chan et al., 2015; Fernández et al., 2003; Kosoy et al., 2012; Molokhia et al., 2003; Rushton & Jensen, 2005). If trait value is positively correlated with the fraction of global ancestry from source population A, then that correlation might be taken as evidence suggesting population A is genetically predisposed to larger trait values than other source populations contributing to the admixed population.

However, this approach depends strongly on assumptions about the relationship between ancestry and environmental effects. We show in the supplementary text (section S.2.1.3) that the expected phenotype of an individual with ancestry fraction θ_i from source population 2 in a two-source admixture is

$$\mathbb{E}(Y_i \mid \theta_i) = \mu_1 + \theta_i \delta_G + h_E(\theta),$$

where μ_i is the mean genetic contribution to the trait in population i and $h_E(\theta)$ is a function that expresses how the conditional expectation of the environmental effect depends on an individual's global ancestry fraction. If the environmental and genetic effects of global ancestry are confounded in the admixed population (i.e., if the correlation of Y and $h_E(\theta)$ is nonzero), then it may be impossible to recover δ_G. As a simple example, consider a case in which individuals with $\theta = 0$ experience environmental effects with expectation E_1, individuals with $\theta = 1$ experience environmental effects with expectation E_2, and there is a linear gradient between those extremes for individuals of intermediate admixture fraction. If we again define the difference between mean environmental effects to be, $\delta_E = E_2 - E_1$, then $h_E(\theta) = E_1 + \theta\delta_E$, and

$$\mathbb{E}(Y_i \mid \theta) = \mu_1 + E_1 + \theta(\delta_G + \delta_E).$$

Under this formulation, it would be possible to use linear regression to estimate the slope (in terms of ancestry proportion θ) and intercept given a sample of admixed individuals with measured admixture fractions and phenotypes, but the resulting slope would provide information only about the sum of the genetic and environmental effects of global ancestry, not their relative contribution or even their direction.

A similar approach might be applied to admixed siblings, where the correlation with trait value is computed only with respect to variation in realized global ancestry between

full siblings, who have the same expected global ancestry proportions on the autosomes. Limiting to within-sibship comparisons would remove some confounds of global ancestry proportion, but it would not exclude environmentally mediated phenotypic differences caused by variation in global ancestry. For example, in African Americans, colorism (Dixon & Telles, 2017) might lead to systematically different treatment as a function of ancestry fraction, even within sets of full siblings.

5.2 Local-ancestry heritability

Although variation in global ancestry may be confounded with environmental variation, one might imagine that variation in the genome-wide sharing of local ancestry segments could provide a way forward in identifying the genetic vs. environmental basis of among-group difference.

Zaitlen et al. (2014) proposed a novel estimator of heritability applicable to individuals from admixed populations. The estimator works on principles similar to SNP heritability estimators (Yang et al., 2017), such as implemented in GCTA (Yang et al., 2011). SNP heritability methods work by measuring the degree to which aggregate sharing of alleles at genotyped SNPs among people who are not closely related is correlated with similarity on a phenotype. A strength of SNP heritability methods is that they avoid potential biases stemming from environmental influences shared among closely related people. A weakness is that they cannot inform about contributions to heritability from variants that are not well tagged by SNPs included in the genetic relatedness matrix, in particular rare variants. In local-ancestry heritability estimates, rather than examining shared SNP genotypes, one examines the sharing of segments of local ancestry in an admixed population. For example, consider pairs of individuals from an admixed population formed from two source populations, where both members of all pairs have 50% of their ancestry from each source population. If a trait is heritable, then a pair whose genetic segments from source population 1 fall in exactly the same locations might be expected to be slightly more phenotypically similar than another pair who share no local ancestry segments. Heritability estimates based on local ancestry offer the potential to address one of the main issues facing SNP heritability estimates, since segments of local ancestry tag both common and rare variants. The resulting estimate has also been taken as a proxy for c in P_{ST} analysis (Zaidi et al., 2017), though Zaitlen and colleagues (2014) made no claims about revealing the causes of phenotypic differences among source populations.

Despite the promise of the approach, heritability estimates based on local ancestry face their own challenges. Specifically, directional or stabilizing selection in the history of the source populations can lead to upward or downward biases in heritability estimates, respectively (Huang et al., 2023). In the supplementary text (section S2), we consider this problem and describe a related one that, in a pathological worst case, can lead to heritability estimates near 0 when the true heritability is 1. This problem is most pronounced when the trait under consideration has experienced selection in the source populations, a scenario

that Zaitlen et al. (2014) recognized as potentially problematic for their estimator.

In the Supplementary Text (section S2), we decompose the genetic variation for a phenotype in a population having experienced several generations of admixture between two source populations into a component due to variation in global ancestry, a component due to the placement of local ancestry segments conditional on global ancestry, and a component due to randomness in genotype conditional on local ancestry. The variance component principally used by local-ancestry heritability estimation is the variance due to placement of local ancestry segments, conditional on global ancestry,

$$(\mathbb{E}(\theta)(1-\mathbb{E}(\theta)) - \mathrm{Var}(\theta))\frac{(2\ell)^2}{2\ell-1}\mathrm{var}(\beta_j(p_{2j}-p_{1j})). \tag{5}$$

Here, θ is a random variable representing global ancestry fraction from source population 2, ℓ is the number of (unlinked) loci considered, β_j is the effect size of an effect allele at locus j, p_{1j} and p_{2j} are frequencies of the effect allele in populations 1 and 2, and var() indicates the population variance.

Under this model, the difference in genetic means between the source populations is

$$\begin{aligned}\delta_G &= 2\sum_j \beta_j(p_{2j}-p_{1j}) \\ &= 2\ell\overline{\beta_j(p_{2j}-p_{1j})},\end{aligned}$$

where the overbar denotes the population mean. However, the expression in (5) does not provide a means to estimate δ_G, since the average of the quantity $\beta_j(p_{2j}-p_{1j})$ does not appear in equation (5); only its variance does. The mean value of $\beta_j(p_{2j}-p_{1j})$ is not constrained by the variance of the $\beta_j(p_{2j}-p_{1j})$ values. The approach of Zaitlen and colleagues (2014) can be applied with or without a fixed effect for global ancestry fraction. However, comparing these results does not reveal the genetic component of the between-group difference if there may also be environmental effects associated with ancestry fraction, for the reasons discussed in the previous subsection (and see section S2.3).

6 Discussion

We showed that knowledge of the heritability of within-population differences does not constrain the heritability of between-population differences at all, as discussed by Lewontin (1970). Our results establish Lewontin's interpretation generally.

Moreover, even if the between-group heritability were estimable, it is consistent with potentially infinitely many configurations of genetic differences among populations and cannot be used to infer the direction(s) of group differences in genetic contributions to a trait. This second point is especially important. Evolutionary models may be able to identify plausible ranges, under given assumptions, for the variance proportion of the genetic component of the trait accounted for by group membership, r, since r is a function

of Q_{ST}, which is studied in evolutionary quantitative genetics. However, even if r (and thus between-group heritability, which has r as a key ingredient) can be constrained somewhat, the causes of group difference remain ambiguous without a clear understanding of environmental differences among groups and their relevance for the trait. This point was also anticipated by a verbal argument, this one by philosopher Ned Block (1995).

Although our analysis of h_B^2 was partially anticipated by Lewontin, Block, and others (Kempthorne, 1978; R. C. Lewontin, 1974, 1970; Mackenzie, 1980; Turkheimer, 1991), the existence of h_B^2 has been invoked for decades, up to the present, to suggest that within-group heritability is informative about among-group differences (Jensen, 1972a, 1972b, 1975; Plomin & DeFries, 1976; Warne, 2020, 2021). As one example, Warne (2021) writes "The existence of the equation [relating h_B^2 to h_W^2] alone shows that the claim that 'the genetic basis of the difference between two populations bears no logical or empirical relation to the heritability within populations' (Lewontin, 1970, p. 7) is incorrect." Our work clarifies that Lewontin's claim is correct in general and not only in the special cases he used for illustration. The relationship between h_W^2 and h_B^2 is rather like the relationship between a random variable's expectation and its variance. One can write the variance as a function of the expectation (i.e. $\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$), but unless further assumptions are made, knowledge of the expectation alone gives no information about the variance. Roseman & Bird (2023) recently arrived at a similar position, concluding that "there is no relationship between h_B^2 and h_W^2 under an additive genetic model...unless evolutionarily explicit models are used to relate them."

These findings have important implications not just for quantifying the genetic component of among-group differences, but also for methods of detecting genetic adaptation using phenotypic data from multiple populations. As we showed, the phenotypic analogue of Q_{ST}, called P_{ST}, is not constrained by within-population heritability. In fact, because the between-population heritability is not bounded from above by 1, common sensitivity analyses, which assume that the between-population component of genetic differentiation is smaller than the within-population component, can be very conservative.P_{ST} has been criticized (Brommer, 2011; Pujol et al., 2008) and is often treated cautiously as a result. However, to our knowledge, the fact that c can be much larger than 1 is not widely understood, nor is the fact that it can achieve very large values in realistic evolutionary scenarios.

The same fundamental problem that plagues h_B^2 and P_{ST} also prevents the separation of genetic and environmental components of differences among source populations via study of ancestry variation in admixed populations. Variation among admixed people in global ancestry proportions may be associated with both genetic and environmental effects, leading to confounding. Variation among admixed people in the sharing of local ancestry segments provides information that is sometimes useful in heritability estimation, but the resulting estimates do not provide a way forward for estimating differences in the genetic contribution to the trait among source populations, since local-ancestry heritability estimates depend principally on a variance component that is not closely related to the genetic

contribution to differences among groups.

All of the phenomena we have discussed so far arise under the simplest possible models, in which phenotypes arise via additive influence of genotype and environment. In the presence of gene-by-environment interaction, the role of genetic factors in explaining group differences itself depends on the environment in which the groups exist. Even careful control of the environment then provides only a partial picture, since it tells us only about the genetic contribution to group difference in the environment studied, which may well change or reverse in a different one (R. C. Lewontin, 1974). For example, Chakraborty & Weiss (1986) studied diabetes incidence in admixed populations with varying proportions of European or Native American ancestries. The increased incidence of diabetes with more Native American ancestries that they observed ran against what seemed to them the most obvious environmental risk factor, the degree to which a group's diet was westernized. However, as they noted, "earlier prevalence studies did not reveal a high occurrence of [diabetes] in pre-World War II populations of Amerindians" suggesting the inadequacy of a purely genetic model, and a potential role for gene-environment interaction in explaining the observed group differences.

Clearly, our results do not prevent the identification of specific genetic or environmental factors that might contribute to group differences, in the sense that, if their influence were removed, the magnitude or direction of a group difference might change. For example, in domestic dogs and related canid species, a specific variant in *IGF-1* appears to be an important factor in explaining body-size differences (Plassais et al., 2022). This variant is associated with body size within and across species and breeds, in a broad set of different genetic backgrounds and different environments, and it plausibly has a regulatory function that relates it to body size. On the environmental side, exposure to known toxins such as air pollution can lead to health problems such as asthma or some cancers, and exposure is sometimes stratified by race or socioeconomic status, plausibly contributing to health disparities (Evans & Kantrowitz, 2002; Liu et al., 2021). The problems we identify arise when simplistic assumptions are made about aggregate contributions of genetic or environmental differences to group differences, particularly when our understanding of the sources of variation in the phenotype are not understood. In some cases, a single causal factor might be identified, such as a genotype that causes a Mendelian disease, although even such seemingly straightforward cases may be more complicated and susceptible to environmental modification, as in the case of phenylketonuria (Scriver, 2007). But when the causes of variation in a phenotype are poorly understood and most phenotypic variation is unexplained, projection on the basis of known factors is particularly difficult. For example, in humans, attempts to explain group differences in terms of polygenic scores—predicted trait values calculated from genotypes, which are typically noisy predictions—are fraught with difficulties (Coop, 2019; Rosenberg et al., 2018).

When it is possible, the best solution to the problems we point out is careful control of the environment. Common-garden observations support the original uses of h_B^2 by Lush (1949) and applications of Q_{ST}. Departures from the common garden are unavoidable for

studies of humans and of some wild organisms. At the same time, if the environment is not controlled, then the sources of among-group differences become ambiguous, and this ambiguity cannot be resolved by study of within-group variation in aggregate.

If control of the environment is not possible, it may sometimes be possible to observe that a phenotypic difference is persistent across a broad set of environments that vary in all dimensions that might be relevant for the trait. With respect to the methods discussed here, they may allow for firmer conclusions if the environmental sources of trait variation are understood and measured or modeled, even if they cannot be controlled. In practice, it is often difficult to make such conclusions—the relevant environmental variables are unknown, and it is difficult to tell whether they have been explored. In fact, Lewontin (1970) contained a third thought experiment to this effect. In the same case as described in the introduction, with genetic variation within groups but no genetic variation between groups raised in different soils, Lewontin asks us to to imagine a chemist making a survey of the soils. The chemist finds a large difference in nitrate concentrations, but correcting the difference only partially lessens the phenotypic difference between the groups, leading researchers to conclude incorrectly that the remaining phenotypic difference is genetic. In fact, a much subtler difference in zinc levels is "the real culprit." This thought experiment was inspired by the long delay in the discovery of the importance of trace minerals for plant development, "because ordinary laboratory glassware will leach out enough of many trace elements to let plants grow normally." Lewontin asked rhetorically, "Should educational psychologists study plant physiology?" We might ask today how we can come to understand the traits we care about well enough to distinguish nitrate cases from zinc cases.

7 Methods

7.1 A linear model of quantitative traits within and between groups

We model the phenotype of an individual j in population i as the sum of genetic and environmental components, G and ϵ, respectively,

$$Y_{ij} = G_i + \tilde{G}_{ij} + \epsilon_i + \tilde{\epsilon}_{ij},$$

where $\tilde{G}_{ij} = G_{ij} - G_i$ and $\tilde{\epsilon}_{ij} = \epsilon_{ij} - \epsilon_i$ are the genetic and environmental deviations, respectively, of individual j from the mean genetic and environmental values for group i. This model is applicable to a trait regardless of genetic architecture, as long as there are no gene-by-environment interactions.

7.2 Computing the between- and within-group heritability

In terms of the linear model formulation, the definition of the between-group heritability is

$$h_B^2 = \frac{\mathrm{Var}(G_i)}{\mathrm{Var}(G_i + \epsilon_i)}$$

and the within-group heritability is

$$h_W^2 = \frac{\mathrm{Var}(\tilde{G}_{ij})}{\mathrm{Var}(\tilde{G}_{ij} + \tilde{\epsilon}_{ij})}.$$

In the Supplementary Text (section S1.2), we show that with these definitions, one can recover the relationship between h_B^2 and h_W^2 derived by DeFries (1972).

7.3 The parameters used in P_{ST}

In terms of the linear model formulation, for diploids, the definition of Q_{ST} is

$$Q_{ST} = \frac{\mathrm{Var}(G_i)}{\mathrm{Var}(G_i) + 2\mathrm{Var}(\tilde{G}_{ij})},$$

and the definition of P_{ST} is

$$P_{ST} = \frac{a\mathrm{Var}(G_i + \epsilon_i)}{a\mathrm{Var}(G_i + \epsilon_i) + 2b\mathrm{Var}(\tilde{G}_{ij} + \tilde{\epsilon}_{ij})}.$$

In terms of the variance proportions t and r in DeFries' expression for between-group heritability, it follows that $Q_{ST} = r/(2 - r)$.

For $P_{ST} = Q_{ST}$ we need

$$\begin{aligned} a\mathrm{Var}(G_i + \epsilon_i) &= \mathrm{Var}(G_i) \\ \implies a &= \frac{\mathrm{Var}(G_i)}{\mathrm{Var}(G_i + \epsilon_i)} \\ &= h_B^2 \end{aligned}$$

and similarly

$$\begin{aligned} b\mathrm{Var}(\tilde{G}_{ij} + \tilde{\epsilon}_{ij}) &= \mathrm{Var}(\tilde{G}_{ij}) \\ \implies b &= \frac{\mathrm{Var}(\tilde{G}_{ij})}{\mathrm{Var}(\tilde{G}_{ij} + \tilde{\epsilon}_{ij})} \\ &= h_W^2. \end{aligned}$$

7.4 Stabilizing selection simulations

We simulated phenotypes under selection using SLiM (Haller & Messer, 2019). In all simulations, an initial population of 500 diploid individuals with genomes of 1,000,000 positions, with a mutation rate of 10^{-7} per generation and a recombination rate of 10^{-5} (to approximate free recombination). New mutations have an effect on the trait drawn from a Normal($\mu = 0$,$\sigma^2 = 0.005$) distribution. Environmental effects are Gaussian with

Scenario	Concordant	Discordant
N	500	500
Mutation rate	10^{-7}	10^{-7}
Recombination rate	10^{-5}	10^{-5}
Mutation distribution	N(0,.005)	N(0,.005)
Baseline environment distribution	N(0,2)	N(0,2)
Baseline selection kernel	N(0,100)	N(0,100)
Time of population split	20,000	20,000
Optimum shift	+1	+1
Environment shift	+4	-2

Table 1: Simulation parameters

variance 2. We impose Gaussian stabilizing selection with an optimum at 0 and variance of 100. After 20,000 generations, the populations split into two populations of 20,000 diploid individuals. We then consider two scenarios. In both scenarios, one subpopulation maintains an optimum at 0, while the other experiences an optimum shift to an optimum at 1. In one scenario, the environment shifts to a mean of -2, in the opposite direction of the optimum shift. In the other scenario, the environment shifts to +4, in the same direction as the optimum shift. The simulations are run for 500 generations after the population split. See Table 7.4 for a summary of simulation parameters. For each scenario, we conducted 500 replicate simulations.

7.5 Testing for adaptation using the Lewontin-Krakauer distribution

To test for adaptation in our simulations by comparing P_{ST} or Q_{ST} to F_{ST}, we first computed ratio-of-averages genome-wide F_{ST} (as implemented in `SLiM`'s `calc_Fst` function) on the basis of neutral markers that were not associated with the trait in the simulations. The Lewontin-Krakauer (R. C. Lewontin & Krakauer, 1973) test uses the statistic

$$T_{LK} = \frac{n-1}{F_{ST}} Q_{ST},$$

which has χ^2 distribution with $n-1$ degrees of freedom, where n is the number of demes (in our case, $n = 2$). For the tests using P_{ST}, we substituted P_{ST} for Q_{ST} in the definition of the Lewontin-Krakauer test statistic. We note that we computed the variance in group means without using Bessel's correction, as shown by Weaver (2016) to be the correct formula for Nei's F_{ST} (Nei, 1973).

8 Code availability

Code used in this paper can be found at `https://github.com/Schraiber/heritability`.

Acknowledgments

We thank K. Bird, G. Coop, A. Gusev, A. Harpak, M. Pennell, and M. Przeworski for helpful comments on an earlier draft of the manuscript, and members of the Edge, Mooney, and Pennell labs for helpful conversations. We are also grateful for the helpful comments of the anonymous reviewers, which improved the manuscript. We acknowledge support from NIH grant R35GM137758 to MDE.

References

Block, N. (1995). How heritability misleads about race. *Cognition*, *56*(2), 99–128. https://doi.org/10.1016/0010-0277(95)00678-R

Brommer, J. E. (2011). Whither Pst? The approximation of Qst by Pst in evolutionary and conservation biology [Publisher: John Wiley & Sons, Ltd]. *Journal of Evolutionary Biology*, *24*(6), 1160–1168. https://doi.org/10.1111/j.1420-9101.2011.02268.x https://doi.org/10.1111/j.1420-9101.2011.02268.x

Castellani, M. B., Dalla Vecchia, A., Bolpagni, R., Natale, R., Piaser, E., Lastrucci, L., Coppi, A., & Villa, P. (2023). Genetic drift versus natural selection affecting the evolution of spectral and functional traits of two key macrophytes: Phragmites australis and nuphar lutea. *Freshwater Biology*, *n/a*(n/a). https://doi.org/https://doi.org/10.1111/fwb.14162

Chakraborty, R., & Weiss, K. M. (1986). Frequencies of complex diseases in hybrid populations. *American Journal of Physical Anthropology*, *70*(4), 489–503. https://doi.org/https://doi.org/10.1002/ajpa.1330700408

Chan, Y., Salem, R. M., Hsu, Y.-H. H., McMahon, G., Pers, T. H., Vedantam, S., Esko, T., Guo, M. H., Lim, E. T., Franke, L., Smith, G. D., Strachan, D. P., & Hirschhorn, J. N. (2015). Genome-wide analysis of body proportion classifies height-associated variants by mechanism of action and implicates genes important for skeletal development. *The American Journal of Human Genetics*, *96*(5), 695–708. https://doi.org/https://doi.org/10.1016/j.ajhg.2015.02.018

Coop, G. (2019). Reading tea leaves? polygenic scores and differences in traits among groups.

DeFries, J. C. (1972). Quantitative aspects of genetics and environment in the determination of behavior. In *Genetics, environment and behavior: Implications for educational policy.* (pp. xx, 324–xx, 324). Academic Press. https://doi.org/10.1016/B978-0-12-233450-4.50009-4

Dixon, A. R., & Telles, E. E. (2017). Skin color and colorism: Global research, concepts, and measurement. *Annual Review of Sociology*, *43*(1), 405–424. https://doi.org/10.1146/annurev-soc-060116-053315

Edge, M. D., & Rosenberg, N. A. (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.

Evans, G. W., & Kantrowitz, E. (2002). Socioeconomic status and health: The potential role of environmental risk exposure [PMID: 11910065]. *Annual Review of Public Health*, *23*(1), 303–331. https://doi.org/10.1146/annurev.publhealth.23.112001.112349

Falconer, D., & Mackay, T. (1996). *Quantitative genetics* (4th edition). Pearson Prentice-Hall.

Feldman, M. W., & Lewontin, R. C. (1975). The heritability hang-up. *Science*, *190*(4220), 1163–1168. https://doi.org/10.1126/science.1198102

Fernández, J. R., Shriver, M. D., Beasley, T. M., Rafla-Demetrious, N., Parra, E., Albu, J., Nicklas, B., Ryan, A. S., McKeigue, P. M., Hoggart, C. L., Weinsier, R. L., & Allison, D. B. (2003). Association of african genetic admixture with resting metabolic rate and obesity among women. *Obesity Research*, *11*(7), 904–911. https://doi.org/https://doi.org/10.1038/oby.2003.124

Haller, B. C., & Messer, P. W. (2019). SLiM 3: Forward Genetic Simulations Beyond the Wright–Fisher Model. *Molecular Biology and Evolution*, *36*(3), 632–637. https://doi.org/10.1093/molbev/msy228

Harpak, A., & Przeworski, M. (2021). The evolution of group differences in changing environments. *PLOS Biology*, *19*(1), 1–14. https://doi.org/10.1371/journal.pbio.3001072

Huang, J., Basu, S., Shriver, M. D., & Zaidi, A. A. (2023). Interpreting snp heritability in admixed populations. *bioRxiv*. https://doi.org/10.1101/2023.08.04.551959

Jensen, A. R. (1969). How much can we boost IQ and scholastic achievement? [Place: US Publisher: Harvard Education Publishing Group]. *Harvard Educational Review*, *39*(1), 1–123. https://doi.org/10.17763/haer.39.1.l3u15956627424k7

Jensen, A. R. (1972a). Interpretation of heritability. [Place: US Publisher: American Psychological Association]. *American Psychologist*, *27*, 973–975. https://doi.org/10.1037/h0038049

Jensen still confused about h2b here. claims a monotonic relationship between h2w and probability that h2b is large (or something like that)?

Jensen, A. R. (1972b). The iq controversy: A reply to Layzer. *Cognition*, *1*(4), 427–452. https://doi.org/10.1016/0010-0277(72)90008-X

Unlike his discussion from 3 years later, Jensen is pretty confused here and thinks that h2B seems to show Layzer (and Lewontin) are wrong

Jensen, A. R. (1975). The meaning of heritability in the behavioral sciences [Publisher: Routledge]. *Educational Psychologist*, *11*(3), 171–183. https://doi.org/10.1080/00461527509529142

doi: 10.1080/00461527509529142 Jensen has an analysis of hˆ2B that is pretty sensible and notes that r is inestimable, and so the formula is empirically useless. (He doesn't note that even if you knew r, you are still hosed)

Kempthorne, O. (1978). A Biometrics Invited Paper: Logical, Epistemological and Statistical Aspects of Nature-Nurture Data Interpretation [Publisher: [Wiley, International Biometric Society]]. *Biometrics*, *34*(1), 1–23. https://doi.org/10.2307/2529584

Kosoy, R., Qi, L., Nassir, R., Garcia, L., Allison, M., Shigeta, R., Robbins, J., & Seldin, M. F. (2012). Relationship between hypertension and admixture in post-menopausal African American and Hispanic American women. *Journal of Human Hypertension*, *26*(6), 365–373. https://doi.org/10.1038/jhh.2011.52

Leinonen, T., McCairns, R. J. S., O'Hara, R. B., & Merilä, J. (2013). QST–FST comparisons: Evolutionary and ecological insights from genomic heterogeneity. *Nature Reviews Genetics*, *14*(3), 179–190. https://doi.org/10.1038/nrg3395

Lewontin, R. C. (1974). The analysis of variance and the analysis of causes. *American Journal of Human Genetics*, *26*(3).

Lewontin, R. (1970). Race and Intelligence [Publisher: Routledge]. *Bulletin of the Atomic Scientists*, *26*(3), 2–8. https://doi.org/10.1080/00963402.1970.11457774
doi: 10.1080/00963402.1970.11457774

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

Liu, J., Clark, L. P., Bechle, M. J., Hajat, A., Kim, S.-Y., Robinson, A. L., Sheppard, L., Szpiro, A. A., & Marshall, J. D. (2021). Disparities in air pollution exposure in the united states by race/ethnicity and income, 1990–2010. *Environmental Health Perspectives*, *129*(12), 127005. https://doi.org/10.1289/EHP8584

Lush, J. L. (1949). HERITABILITY OF QUANTITATIVE CHARACTERS IN FARM ANIMALS [Publisher: John Wiley & Sons, Ltd]. *Hereditas*, *35*(S1), 356–375. https://doi.org/10.1111/j.1601-5223.1949.tb03347.x
https://doi.org/10.1111/j.1601-5223.1949.tb03347.x

Mackenzie, B. (1980). Hypothesized genetic racial differences in IQ: A criticism of three proposed lines of evidence. [Place: Germany Publisher: Springer]. *Behavior Genetics*, *10*, 225–234. https://doi.org/10.1007/BF01066273
discusses Jensen's invocation of h^2b but is a bit confused itself

Molokhia, M., Hoggart, C., Patrick, A. L., Shriver, M., Parra, E., Ye, J., Silman, A. J., & McKeigue, P. M. (2003). Relation of risk of systemic lupus erythematosus to west African admixture in a Caribbean population. *Human Genetics*, *112*(3), 310–318. https://doi.org/10.1007/s00439-002-0883-3

Nei, M. (1973). Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences*, *70*(12), 3321–3323. https://doi.org/10.1073/pnas.70.12.3321

Plassais, J., vonHoldt, B. M., Parker, H. G., Carmagnini, A., Dubos, N., Papa, I., Bevant, K., Derrien, T., Hennelly, L. M., Whitaker, D. T., Harris, A. C., Hogan, A. N., Huson, H. J., Zaibert, V. F., Linderholm, A., Haile, J., Fest, T., Habib, B., Sacks, B. N., … Ostrander, E. A. (2022). Natural and human-driven selection of a single

non-coding body size variant in ancient and modern canids. *Current Biology*, *32*(4), 889–897.e9. https://doi.org/https://doi.org/10.1016/j.cub.2021.12.036

Plomin, R., & DeFries, J. C. (1976). Heritability of IQ [Publisher: American Association for the Advancement of Science]. *Science*, *194*(4260), 10–12. https://doi.org/10.1126/science.194.4260.10

doi: 10.1126/science.194.4260.10 Attacks Feldman & Lewontin's discussion of h2b

Prout, T., & Barker, J. S. (1993). F statistics in Drosophila buzzatii: Selection, population size and inbreeding. *Genetics*, *134*(1), 369–375. https://doi.org/10.1093/genetics/134.1.369

Pujol, B., Wilson, A. J., Ross, R. I. C., & Pannell, J. R. (2008). Are QST–FST comparisons for natural populations meaningful? [Publisher: John Wiley & Sons, Ltd]. *Molecular Ecology*, *17*(22), 4782–4785. https://doi.org/10.1111/j.1365-294X.2008.03958.x https://doi.org/10.1111/j.1365-294X.2008.03958.x

Roseman, C. C., & Bird, K. A. (2023). Between group heritability and the status of hereditarianism as an evolutionary science. *bioRxiv*. https://doi.org/10.1101/2023.12.18.572247

Rosenberg, N. A., Edge, M. D., Pritchard, J. K., & Feldman, M. W. (2018). Interpreting polygenic scores, polygenic adaptation, and human phenotypic differences. *Evolution, Medicine, and Public Health*, *2019*(1), 26–34. https://doi.org/10.1093/emph/eoy036

Rushton, J. P., & Jensen, A. R. (2005). Thirty years of research on race differences in cognitive ability. [Place: US Publisher: American Psychological Association]. *Psychology, Public Policy, and Law*, *11*(2), 235–294. https://doi.org/10.1037/1076-8971.11.2.235

Salloum, P. M., Lavery, S. D., de Villemereuil, P., & Santure, A. W. (2022). Local adaptation in shell shape traits of a brooding chiton with strong population genomic differentiation. *Evolution*, *77*(1), 210–220. https://doi.org/10.1093/evolut/qpac011

Scriver, C. R. (2007). The pah gene, phenylketonuria, and a paradigm shift. *Human Mutation*, *28*(9), 831–845. https://doi.org/https://doi.org/10.1002/humu.20526

Spitze, K. (1993). Population structure in Daphnia obtusa: Quantitative genetic and allozymic variation. *Genetics*, *135*(2), 367–374. https://doi.org/10.1093/genetics/135.2.367

Turkheimer, E. (1991). Individual and group differences in adoption studies of IQ. [Place: US Publisher: American Psychological Association]. *Psychological Bulletin*, *110*, 392–405. https://doi.org/10.1037/0033-2909.110.3.392

Vitek, N. S., McDaniel, S. F., & Bloch, J. I. (2022). Microevolutionary variation in molar morphology of Onychomys leucogaster decoupled from genetic structure. *Evolution*, *76*(9), 2032–2048. https://doi.org/10.1111/evo.14576

Warne, R. T. (2020). *In the know: Debunking 35 myths about human intelligence*. Cambridge University Press.

Warne, R. T. (2021). Between-Group Mean Differences in Intelligence in the United States Are ¿0% Genetically Caused: Five Converging Lines of Evidence. *The American Journal of Psychology*, *134*(4), 480–501. https://doi.org/10.5406/amerjpsyc.134.4.0479

Weaver, T. D. (2016). Estimators for qst and coalescence times. *Ecology and evolution*, *6*(21), 7783–7786.

Whitlock, M. C. (1999). Neutral additive genetic variance in a metapopulation. *Genetics Research*, *74*(3), 215–221. https://doi.org/10.1017/S0016672399004127

Whitlock, M. C. (2008). Evolutionary inference from QST [Publisher: John Wiley & Sons, Ltd]. *Molecular Ecology*, *17*(8), 1885–1896. https://doi.org/10.1111/j.1365-294X.2008.03712.x https://doi.org/10.1111/j.1365-294X.2008.03712.x

Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2011). Gcta: A tool for genome-wide complex trait analysis. *The American Journal of Human Genetics*, *88*(1), 76–82. https://doi.org/https://doi.org/10.1016/j.ajhg.2010.11.011

Yang, J., Zeng, J., Goddard, M. E., Wray, N. R., & Visscher, P. M. (2017). Concepts, estimation and interpretation of SNP-based heritability. *Nature Genetics*, *49*(9), 1304–1310. https://doi.org/10.1038/ng.3941

Zaidi, A. A., Mattern, B. C., Claes, P., McEcoy, B., Hughes, C., & Shriver, M. D. (2017). Investigating the case of human nose shape and climate adaptation. *PLOS Genetics*, *13*(3), 1–31. https://doi.org/10.1371/journal.pgen.1006616

Zaitlen, N., Pasaniuc, B., Sankararaman, S., Bhatia, G., Zhang, J., Gusev, A., Young, T., Tandon, A., Pollack, S., Vilhjálmsson, B. J., Assimes, T. L., Berndt, S. I., Blot, W. J., Chanock, S., Franceschini, N., Goodman, P. G., He, J., Hennis, A. J. M., Hsing, A., … Price, A. L. (2014). Leveraging population admixture to characterize the heritability of complex traits. *Nature Genetics*, *46*(12), 1356–1362. https://doi.org/10.1038/ng.3139