Received: 18 July 2020 | Revised: 27 November 2020 | Accepted: 20 December 2020

DOI: 10.1002/ajpa.24216

RESEARCH ARTICLE

American Journal of PHYSICAL ANTHROPOLOGY WILEY

No support for the hereditarian hypothesis of the Black–White achievement gap using polygenic scores and tests for divergent selection

Kevin A. Bird[1,2]

[1]Department of Horticulture, Michigan State University, East Lansing, Michigan, USA

[2]Ecology, Evolutionary Biology and Behavior Program, Michigan State University, East Lansing, Michigan, USA

Correspondence
Kevin A. Bird, Department of Horticulture, Michigan State University, East Lansing, Michigan 48823, USA.
Email: birdkevi@msu.edu

Abstract

Objectives: Debate about the cause of IQ score gaps between Black and White populations has persisted within genetics, anthropology, and psychology. Recently, authors claimed polygenic scores provide evidence that a significant portion of differences in cognitive performance between Black and White populations are caused by genetic differences due to natural selection, the "hereditarian hypothesis." This study aims to show conceptual and methodological flaws of past studies supporting the hereditarian hypothesis.

Materials and methods: Polygenic scores for educational attainment were constructed for African and European samples of the 1000 Genomes Project. Evidence for selection was evaluated using an excess variance test. Education associated variants were further evaluated for signals of selection by testing for excess genetic differentiation (F_{st}). Expected mean difference in IQ for populations was calculated under a neutral evolutionary scenario and contrasted to hereditarian claims.

Results: Tests for selection using polygenic scores failed to find evidence of natural selection when the less biased within-family GWAS effect sizes were used. Tests for selection using F_{st} values did not find evidence of natural selection. Expected mean difference in IQ was substantially smaller than postulated by hereditarians, even under unrealistic assumptions that overestimate genetic contribution.

Conclusion: Given these results, hereditarian claims are not supported in the least. Cognitive performance does not appear to have been under diversifying selection in Europeans and Africans. In the absence of diversifying selection, the best case estimate for genetic contributions to group differences in cognitive performance is substantially smaller than hereditarians claim and is consistent with genetic differences contributing little to the Black–White gap.

KEYWORDS
genomics, hereditarian hypothesis, polygenic selection, race

1 | INTRODUCTION

The debate over the role of genetics in the claimed IQ gap between Black and White populations has been extraordinarily long-lasting. Claims of the putative intellectual inferiority of "the Black race" date back to the roots of colonial expansion and slavery (Saini, 2019) and have been persistent despite the general weakness of evidence in support of the claim. Over the last half-century, several key authors

have forwarded a diverse set of evidence from genetics, anthropology, and psychology to support two primary claims of what has been called the "hereditarian hypothesis": (a) genetic differences between races substantially contribute to differences in cognitive ability, as measured by IQ tests, which are impervious to alterations in the environment (Jensen, 1969a; Jensen, 1985; Rushton & Jensen, 2005); and (b) that the relevant genetic differences between races were produced by divergent natural selection that favored higher cognitive ability outside of Africa (Jensen, 1974; Lynn, 2006; Lynn & Vanhanen, 2006; Rushton, 1996, 2000).

However, these claims have often relied on indirect evidence of genes and their activity, relying greatly on ecological regressions, adoption studies, and admixture correlations. These sources of evidence have been criticized on empirical grounds by previous authors (Centerwall, 1978; Thomas, 2016; Waldman et al., 1994; Wicherts et al., 2010). Additionally, the lack of direct evidence has been noted as a looming issue (Loehlin, 2000), marking a contrast between traits like intelligence and traits likes sickle cell anemia, prostate cancer and end-stage kidney disease, where direct genetic investigation was able to shed light on population differences (Freedman et al., 2018; Freedman & Haiman, 2006; Piel & Patil, 2010).

Recently, behavioral geneticists have claimed major progress in uncovering the genetic basis of cognitive performance. These advances extend on previous research involving twin studies, which allowed only indirect estimation of genetic contributions (Polderman & Benyamin, 2015), by leveraging genome sequencing and large samples of unrelated individuals to perform genome-wide association studies (GWAS) that identify specific genetic variants associated with cognitive performance. The most recent GWAS on educational attainment (EA) and cognitive performance (CP), Lee et al. (2018), identified over 1000 single-nucleotide polymorphisms (SNPs) that are significantly associated with higher EA in a sample of over 1 million people of European ancestry. The combined effects of all associated SNPs in Lee et al. (2018), called a polygenic score (PGS), can explain only between 10% and 13% of variance in EA and 7%–10% of variance in CP. Lee et al. (2018) also estimate effect sizes for associate SNPs at the within-family level by analyzing genetic variance between a large sample of siblings, and compare these effect sizes to those from the main between-family (population level) analysis to further evaluate the role of causal and confounding effects in GWAS results. Beyond expanding the understanding of the genetics of these traits, Lee et al. (2018) also provided valuable community resources in the form of publicly available GWAS summary statistics that provide a full list of EA and CP associated SNPs and the estimated effect size on these traits.

Some authors have recently attempted to use this GWAS to provide support for the hereditarian hypothesis using educational attainment PGS as direct genetic evidence (Lasker et al., 2019; Piffer, 2019). These studies profess to support the first hereditarian claim of substantial genetic contribution to race differences (Lasker et al., 2019) and the second hereditarian claim that divergent natural selection between races contributes to racial phenotypic differences (Piffer, 2019); however, there are key deficiencies in their methodology. Both analyses are affected by systematic biases in between-family polygenic scores applied to people from groups that were not included in the original GWAS. Results from Duncan and Shen (2019) show that across a range of traits, the estimated value for non-European populations using polygenic scores is biased and explains much less variance than in European populations. African ancestry is the most affected, with an average effect size only 50% of the size of the same genetic variants identified in Europeans, and there is even larger misestimation for phenotypic traits related to cognition (Duncan & Shen, 2019). In fact, in Lee et al. the polygenic score accounts for 1.6% of variation in Black individuals in AddHealth compared with 9.2% in White individuals. Additional problems with these analyses exist. For example in Piffer (2019), the correlation of observed national IQ scores and polygenic scores of the 1000 Genomes populations are taken as evidence that differences between populations have been caused by natural selection, even though this is not an accepted test for polygenic selection. In Lasker et al. (2019), many of the variables that the authors attempt to control for in are poorly measured and weaken their conclusions, such as skin color predicted from genotype (Carratto et al., 2019) and insufficient socioeconomic measures that include only parental education, even though income (and in particular permanent income) plays a substantial role in attenuating racial achievement gaps (Rothstein & Wozny, 2013).

The biases of these methods makes it problematic to consider previous findings from Piffer (2019) and Lasker et al. (2019), which apply PGS to the Black–White IQ gap as evidence in support of the genetic hypothesis. The present analysis aims to more rigorously test the two main claims of the hereditarian hypothesis against the null hypotheses that (a) allele frequency differences between Black and White populations are consistent with neutral evolutionary processes, and (b) the genetic basis of differences in EA and CP render similar cognitive abilities between Black and White populations. First, I address shortcomings of previous applications of polygenic score analysis, using recent methods that leverage genomic data and population genetic theory and avoid or remedy many of the known biases of polygenic scores. Contrary to past attempts by Piffer (2019) and Lasker et al. (2019), by accounting for biases in polygenic scores and using formal tests for divergent selection I hereby demonstrate that allele frequency differences are in fact consistent with a neutral evolutionary trajectory. Second, I demonstrate that even when adopting several unrealistic assumptions held by proponents of the hereditarian hypothesis and their own National IQ data (which is heavily biased against African samples), the expected genetic contribution to differences in IQ scores under neutral evolution is insufficient to support the hereditarian position.

For clarification, the terms "White" and "Black" will be used throughout this paper, not as an endorsement of their status as biological categories but to reflect the conceptual framework used by hereditarian authors on this topic. Modern population genetics makes it clear that, even though the people in the samples I used would likely be racialized (in terms of their social identity), these racial categories do not accurately reflect genetic ancestry.

2 | METHODS

2.1 | Data set, quality filtering, and SNP clumping

Phase 3 of the 1000 Genomes project (1000 Genomes Project Consortium, 2015), based on reference-build hg37, was downloaded for autosomal chromosomes. I filtered samples to only include the European superpopulation (EUR) which consists of 503 individuals from CEU (99), TSI (107), FIN (99), GBR (91), and IBS (107) populations and the African superpopulations which consists of 504 individuals from YRI (108), LWK (99), GWD (113), MSL (85), and ESN (99) populations. I filtered variants to remove those with minor-allele-frequency less than 0.01 and those loci that deviate from Hardy–Weinberg Equilibrium at $p < 1 \times 10^{-6}$ independently for each subpopulation, as in Guo et al. (2018), to remove low quality SNPs and genotyping errors. After filtering separately, these populations were merged in with Plink V. 1.9 (Purcell & Neale, 2007) (http://pngu.mgh.harvard.edu/purcell/plink/), and variants with a genotyping rate less than 10% were removed to filter out variants present in only one population. These filtering parameters may introduce biases to the polygenic score estimates, however, removing genotyping errors and missing variants are a vital component of quality filtering (Graffelman et al., 2017). Additionally, because the same procedures were performed on SNPs selected from the GWAS, which are used for the null distribution, there should be no relative bias that would result in a false-positive.

The full reported summary statistics, including physical location, effect size, and standard error, and sample size, from Lee et al. (2018) were downloaded from https://www.thessgac.org/data for both educational attainment and cognitive performance. The educational attainment data consisted of 10,101,242 SNPs associated with educational attainment from independent discovery panels, totaling 766,345 individuals and the cognitive performance data consisted of 10,098,325 SNPs from independent discovery panels, total 257,828 individuals. These discovery panels consisted of 71 total cohorts, only using people of European ancestry. In total these variants account for ~10%–15% of variation in educational attainment and cognitive performance in the discovery panels, while they only explained 1.6% of variance when African-American samples were analyzed. Additionally, 82,609 SNPs from within-family analyses were used for the study. These SNPs were ascertained using four cohorts with sibling data, total 22,135 sibling pairs from the Swedish Twin Registry's Twingene study, the Swedish Twin Registry's Screening Across the Lifespan Twin Youth study, the UK Biobank and the Wisconsin Longitudinal Study. Compared with the between-family effect sizes, within-family effect sizes showed sign concordance (both effect trait in same direction) 56.2% of the time for the most laxed *p*-value filtering, and 65.2% at the strictest *p*-value filtering. Additionally, a within-family regression analysis suggests that within-family parameters are ~40% smaller than the between-family GWAS effect sizes. SNPs with both within- and between family effect sizes are compared in Figure 1.

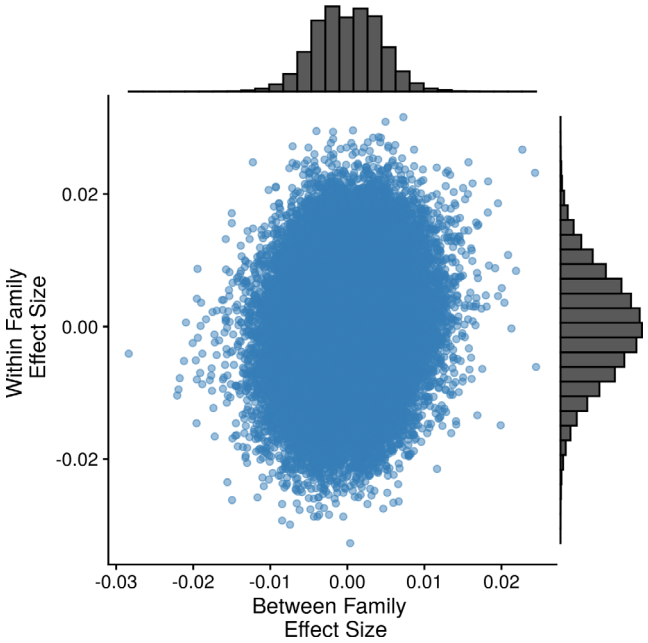

FIGURE 1 Comparison of between-family and within-family effect sizes for variants which have both. Figure generated in R 3.6 0.3 with ggplot2 (Wickham, 2016)

2.2 | Polygenic selection analysis

Polygenic selection analysis used the set of 1,271 genome-wide significant lead SNPs associated with both educational attainment and cognitive performance in a sample of 1,131,881 individuals reported in the supplemental of Lee et al. (2018), which explained a total 3.6% of variance in educational attainment in European individuals. Of the 1,271 SNPs, there were 685 present in the 1000 genomes data sets after filtering that also had both within- and between-family effect sizes. These 685 SNPs were used to calculate population specific polygenic scores (PGS) as:

$$\mathrm{PGS} = \sum_{i=1}^{n} FA_i * eff_i$$

where FA_i is the population frequency of the effect allele for SNP i and eff_i is the SNP specific effect size estimate. The squared difference of the African PGS and European PGS was used to measure the difference between estimated genetic values between populations based on both between—and within-family effect sizes. To see whether the between family PGS difference was larger than expected, an empirical null distribution was calculated by randomly flipping the sign of effect size estimates for 10,000 permutations. The resulting permuted squared population PGS differences were compared with the observed difference, allowing the calculation of an empirical *p*-value based on the proportion of random samples greater than or equal to the observed squared PGS difference.

For within-family effect sizes, I constructed an empirical null distribution of SNPs that were matched to derived allele frequency and LD score quintiles. This was done to control for potential biases from systematic differences in derived allele frequency of GWAS risk alleles between African and non-African populations, which had been identified in previous work (Kim et al., 2018; Lee et al., 2018). First, I made 20 derived allele frequency and LD Score bins for the 685 lead GWAS SNPs from Lee et al. (2018) with within-family effect sizes and derived/ancestral allele status information to the empirical null distribution generated from matched SNPs. Next, I matched the SNPs of each GWAS 10,000 times, and each matched SNP was given the same within-family effect size as its matching GWAS SNP. Finally, I generated polygenic scores by multiplying the GWAS SNP within-family effect size by the matched SNP effect allele frequency in African and European samples and calculated the squared polygenic score difference for the 10,000 matched SNP sets to generate an empirical null distribution. Similar to the Qx analysis of Berg and Coop (2014), significant divergence from the empirical null distribution is taken as evidence of divergent polygenic selection.

2.3 | F_{st} differentiation analysis

I performed LD based clumping in Plink on the 10,101,242 SNPs associated with educational attainment and 10,098,325 SNPs associated with cognitive performance with a p-value cutoff of 5×10^{-6} and r^2 cutoff of 0.01, using a distance threshold of 1 Mb as done in Guo et al. (2018). This resulted in 1259 independent SNPs for educational attainment and 602 independent SNPs for cognitive performance.

F_{st} and LD Scores were calculated for the entire filtered SNP set in vcftools and LDSC, respectively (Bulik-Sullivan et al., 2015; Danecek & Auton, 2011). As done in Guo et al. (2018), SNPs were divided into 20 minor allele frequency (MAF) bins and 20 LDScore bins and 10,000 SNPs were matched for each GWAS SNP of the clumped data sets with the same MAF and LD Score bin, resulting in 10,000 control SNP sets with comparable minor-allele frequency and LD Score. Average F_{st} of the 10,000 control SNP sets were used to create an empirical null distribution, allowing a comparison of the mean F_{st} of the GWAS SNPs with the empirical null distribution with a two-sided one-sample z-test.

2.4 | Predicted phenotypic differences based on trait associated SNP F_{st}

Suppose we granted the implicit assumptions of the hereditarian positions (constant additive effects within and across all possible environments, etc.) In this case only natural selection or genetic drift can contribute to genetic variance that creates phenotypic variation in a population. We can further model and estimate how much phenotypic variance could exist by genetic drift alone and how much phenotypic variance could exist given observed genetic differentiation in trait associated SNPs. This would allow us to see how much genetic variation could contribute to phenotypic variance in the most favorable scenario possible for the hereditarian hypothesis and compare that to actual observed phenotypic variation. This analysis would simultaneously demonstrate the need for natural selection for the hereditarian hypothesis to be tenable, and test the ability of the hereditarian hypothesis to explain phenotypic difference in the most favorable of cases.

To estimate what phenotypic F_{st} would look like if all among-group differences were genetic, divergence was exclusively attributable to random genetic drift, and no GxE or GxG existed, I use the approach developed by Relethford and Blangero (1990) and Relethford et al. (1997). I also use the country-level IQ data used by Piffer (2015) that match the 1000 genome populations that constitute the AFR and EUR superpopulations. While these National IQ data set have come under harsh criticism recently (Ebbesen, 2020), they are used here because of their common use among hereditarians, for example, Piffer (2015). By using their chosen data the strongest possible version of the hereditarian hypothesis is tested.

We can define the divergence of the means of populations (μ_i) from the grand phenotypic mean of all populations (μ_t) as:

$$c_{ij} = (\mu_i - \mu_t)(\mu_j - \mu_t) \quad (1)$$

If environmental deviation is 0 and all differences between group means is due to genetic effects, the average of all elements c_{ii} ($\sigma^2_B = \sum_{i=1}^{k} w_i c_{ii}$, where w_i is the weight of population i) is our estimate of among population phenotypic variance.

An estimate of F_{st} can be derived by taking the proportion of total variance attributable to variance in genetic effects. To do this we need to have an estimate of the average within population additive genetic variance (σ^2_A), which can be calculated by multiplying phenotypic variance (σ^2_P) by narrow sense heritability (h^2). For narrow sense heritability estimates, I use 0.35, based on several family based analyses (Chipuer et al., 1990; Cloninger et al., 1979; Devlin et al., 1997; Loehlin, 1978; Rao et al., 1982), and 0.5, based Polderman and Benyamin (2015)) and Hill and Arslan (2018). It should be noted these estimates are higher than most genomic-based estimates of heritability and are likely upwardly biased (Holland et al., 2020; Morris et al., 2020; Young et al., 2018) so I also use 0.15 based on Young et al. (2018) and Holland et al. (2020). Using these heritability estimates and the standard deviation of IQ scores, we can then estimate F_{st} as:

$$F_{ST} = \frac{\sum_{i=1}^{k} w_{ii} c_{ii}}{2\sigma^2_A + \sum_{i=1}^{k} w_{ii} c_{ii}} \quad (2)$$

This value represents the expected proportion of genetic variance that accounts for the observed phenotypic variance across these populations.

We can also estimates the amount of between group genetic variance of cognitive performance related SNPs with the F_{st} estimates calculated from Lee et al. (2018) using a rearranged version of Equation (2):

$$\sigma_B^2 = \frac{2F_{ST}\sigma_P^2 h^2}{1-F_{ST}} \quad (3)$$

When we divide the genetic variance of cognitive performance related SNPs by the observed variance in IQ between populations, we get the maximum proportion of between group IQ variance that can be explained by genetic variance between populations. Furthermore, we can estimate the expected value of the absolute differences that is attributable to genetic differences between groups, under the assumption of normality, with Equation (4):

$$E[|x-y|] = 2\sigma_B * \sqrt{\frac{2}{\pi}} \quad (4)$$

It is important to emphasize that this is a rough calculation that fails to include basic statistical elements like bias corrections or uncertainty measures. It also relies on unrealistic assumptions that variance among populations is only attributable to additive genetic effects, that the polygenic effect of these genotypes are not affected by GxG (epistasis) or GxE (gene–environment interaction), and effect sizes will be similar in different populations. These assumptions are unlikely to hold in reality, but violations of these assumptions will tend to overestimate genetic contributions and underestimate environmental contributions. Given the violations of these assumptions, these estimates are best viewed as a rough upper bound rather than a precise value of the genetic contribution to IQ differences under neutrality.

3 | RESULTS

3.1 | Polygenic selection

I sought to test for divergent polygenic selection using an method similar to that employed by Berg and Coop (2014), which should have high power to detect polygenic selection with a low false-positive rate, provided GWAS effect sizes are free from systematic biases (Berg & Harpak, 2019). The squared difference of polygenic scores computed from the 685 reported genome-wide significant GWAS with between-family effect sizes from Lee et al. (2018) was 0.068 (raw difference of −0.261), with the African PGS equal to −0.99, and the European PGS equal to 0.161. I observed only one random polygenic score that exceeded this observed squared difference (Figure 2a). When within-family effect sizes were used, the squared difference in PGS was attenuated nearly 95% to 0.0027 (raw difference of −0.052). When compared with the empirical null distribution, 43.2% of randomly generated polygenic scores had greater than or equal squared polygenic score differences, yielding a nonsignificant empirical *p*-value of 0.432 (Figure 2b).

3.2 | F_{st} differentiation

It is possible that some biases still remain in polygenic scores that artificially inflate the gap between Africans and Europeans. I next use a

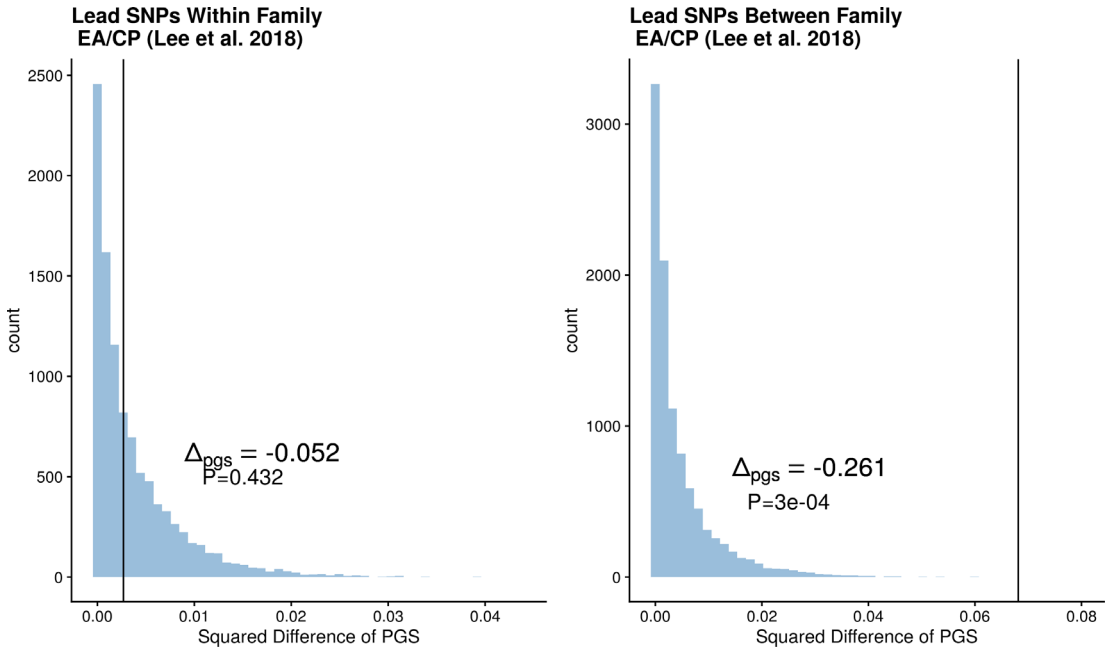

FIGURE 2 Squared polygenic score differences between African and European populations for Lead SNPs based on (a) between family effect sizes, (b) within-family effect sizes with controls for derived allele frequency and LD score reported by Lee et al. (2018). The black line represents the observed squared PGS difference and the histograms represent an empirical null distribution from random shuffling of signs of effect sizes for 10,000 permutations. *p* value is proportion of empirical null distribution greater than or equal to observed squared PGS difference, Δ_{PGS} represents raw PGS difference, negative values indicate $PGS_{AFR} < PGS_{EUR}$. Figure generated in R 3.6.3 with ggplot2 (Wickham, 2016)

method based solely on genetic differentiation at trait associated SNPs, which is free of potential systematic effect size biases, to determine how genetically-differentiated variants associated with cognitive performance and educational attainment are compared with matched sets of non-associated variants. This method not only shows the magnitude of genetic differences in populations, but also serves as an additional test for divergent natural selection (Guo et al., 2018). Based on simulations, Guo et al. (2018) reported maximum power to detect selection with a clumping *p*-value threshold of 5×10^{-6}, I highlight results using that threshold.

The mean F_{st} of the 1259 independent educational attainment associated SNPs was 0.111, slightly below the mean of the empirical null distribution, which was 0.118. Results from the two-sided one sample z-test yielded a *p*-value of 0.043, rejecting the null hypothesis that the educational attainment GWAS SNPs mean F_{st} is not different from the mean of the null distribution, suggesting that the EA F_{st} was significantly smaller than F_{st} of the control SNPs (Figure 3a). For cognitive performance, the mean F_{st} for the 602 SNPs was 0.112, slightly lower than the null distribution mean of 0.119. The results of the one sample z-test yielded a *p*-value of 0.164, failing to reject the null hypothesis that the cognitive performance GWAS SNPs mean F_{st} is not different from the mean of the null distribution (Figure 3b).

3.3 | Predicted phenotypic differences based on trait associated SNP F_{st}

The previous analysis provides an opportunity to compare variance of trait-associated SNPs to phenotypic data on IQ scores at the national level. By doing so, we estimate the maximum genetic and minimum environmental contribution to among group variance in IQ. We also provide an estimate for how much intergroup variance could be produced by genetic drift alone. Assumptions involved in this analysis typically overestimate genetic contribution, making these estimates an upper bound. These estimates are possible because we would expect the F_{st} value estimated from GWAS identified SNPs to be equal to the phenotypic variance if all among-group variation is due to additive genetic effects. Using the country-level IQ results used by Piffer (2015), the African group mean IQ is 68.4 and the European group mean IQ is 99.2, a pairwise difference of 30.8. This gives us an among-group variance of 327.16 IQ points^2. From Equation (2) we get an estimated phenotypic F_{st} of 0.60, using $h^2 = 0.35$; and 0.51, using $h^2 = 0.5$. Both of these estimates are much larger than the 0.111 estimated from the Lee et al. (2018) cognitive performance GWAS SNPs.

The cognitive performance GWAS SNP F_{st} and Equation (3) yields an estimated among group variance due to genetic effects of 8.5 IQ points^2 using $h^2 = 0.15$, 19.9 IQ points^2 using $h^2 = 0.35$; and 28.4 IQ points^2 using $h^2 = 0.5$. Dividing the phenotypic among-group variance and the among-group variance due to genetic effects shows that assuming $h^2 = 0.15$, a maximum of 3.6% of variance in IQ between Africans and Europeans is attributable to additive genetic variance, a maximum of 8.4% of the variance in IQ is attributable to additive genetic variance at $h^2 = 0.35$, and a maximum of 12% of the variance in IQ at $h^2 = 0.5$. This indicates that well over 85% of IQ variance is environmental. Additional F_{st} values from Lee et al. (2018) lead SNPs, educational attainment SNPs from Lee et al. (2018), and lead SNPs for intelligence from Hill and Marioni (2019) are used to calculate this estimate in Table 1. Finally, using Equation (4), the expected genetic differences from these estimates is, at most, 4.7–8.5 IQ points and could result in higher scores in either Europe or Africa with equal likelihood.

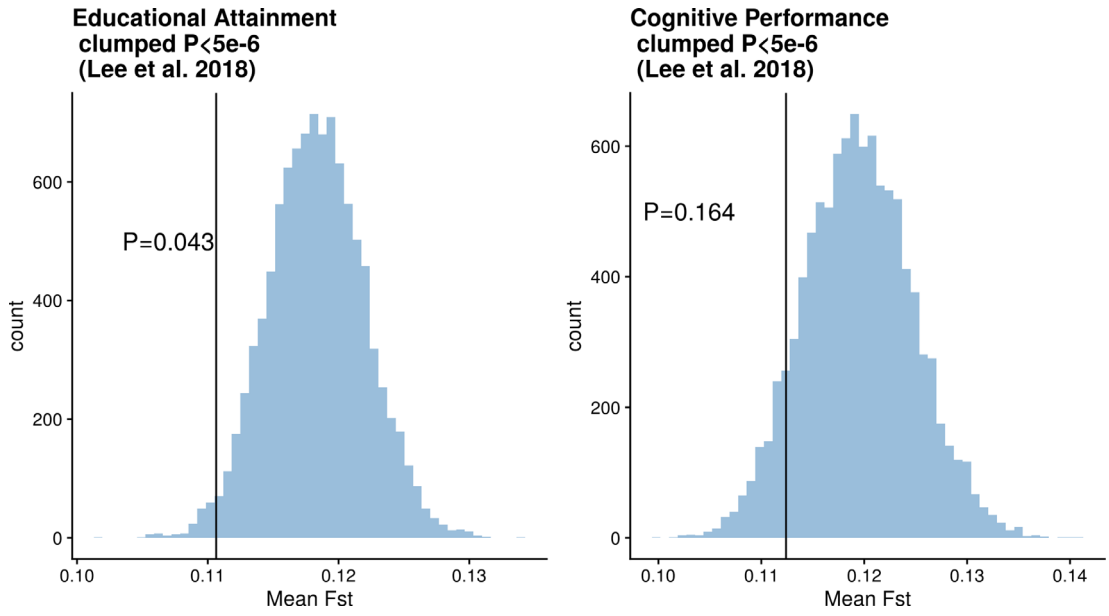

FIGURE 3 Mean F_{st} values of the associated SNPs in 1000G populations against null distribution for (a) educational attainment and (b) cognitive ability. The black line represents the mean F_{st} of the trait-associated SNPs clumped at $p < 5 \times 10^{-6}$. The histogram represents the distribution of mean F_{st} of control SNPs. Empirical *p*-values reported in the figure. Figure generated in R 3.6.3 with ggplot2 (Wickham, 2016)

TABLE 1 Estimated proportion of global IQ variance attributable to additive genetic variation using different calculated F_{st} estimations and heritability estimates from the literature

	Variance attributable to additive genetic variation		
	h^2=0.15	h^2=0.35	h^2=0.5
1301 clumped EA SNPs $p<5 \times 10^{-6}$ (F_{st} = 0.111)	3.60%	8.30%	12.00%
627 clumped cognitive performance SNPs $p<5 \times 10^{-6}$ (F_{st}=0.112)	3.60%	8.40%	12.00%
1021 lead SNPs from Lee et al., 2018 (F_{st}=0.121)	3.90%	9.10%	13.10%
546 lead SNPs from Hill & Marioni, 2019 (F_{st}= 0.126)	4.10%	9.60%	14.00%

4 | DISCUSSION

Before further discussion of the results, several important limitations should be made clear. First, the low genetic and phenotypic variance explained by lead SNPs and the lower sample size of within-family analyses reduces the power for both tests of polygenic selection (CP and EA). While the results presented here are more consistent with neutral evolution rather than divergent natural selection, it is not possible to rule out that data sets with more power could present different results. Additionally, although within-family effect sizes are recommended over between-family effect sizes, if the within-family effect sizes are re-estimated for SNPs ascertained by a between-family GWAS, there is still likely to be some level of confounding from population structure (Cox et al., 2019; Zaidi & Mathieson, 2020). Second, I am not able to account for the complete gap in scores and attainment between groups without full coverage of genetic variation that contributes to group differences. Analyses presented here only partially account for the gap in scores, or make assumptions that the unobserved genetic variants will have similar distribution to the observed genetic variants. Third, the effect sizes used here were estimated for European samples from this meta-analysis and then applied to the independent and more diverse 1000 Genomes Projects samples, which have no corresponding phenotype data. Finally, without environmental measurements and corresponding phenotype data for the 1000 Genomes Project samples, it is not possible to test for gene-by-environment interactions or gene-by-gene interactions. Despite these limitations, this study uses the best current data and methods to test the hereditarian hypothesis using frameworks that have been successful for analyzing traits such as human height (Chen & Chiang, 2020; Cox et al., 2019).

4.1 | No evidence of divergent natural selection between Africans and Europeans

Claims of natural selection have been a centerpiece of hereditarian explanations of the Black–White IQ gap, likely because selection provides an intuitive answer to why alleles might systematically differ between races. Rushton (2000) and Lynn (2006) put natural selection at the core of their conjecture about why genetic differences might exist between White and Black populations. However, theoretical and empirical support for Rushton's claims have been called into question (Gorey & Cryns, 1995; Cernovsky & Litman, 1993; Cernovsky, 1990; Graves, 2002; Zeitsch and Zietsch & Sidari, 2019). Critics were also quick to point out that Lynn's analyses had a deeply flawed sampling from Africa, resulting at least in a bias against African countries and it showed signs of measurement bias from the IQ tests themselves (Wicherts et al., 2010). The evolutionary reasoning has also been critiqued by research that casts doubt on the validity of the "Cold Winters theory" (MacEachern, 2006; Pesta & Poznanski, 2014; Wicherts et al., 2010). Continuing this legacy of evolutionary explanations for racial theories, recent genomic analyses (Piffer, 2015, 2019) claim to provide strong genetic evidence in support of natural selection using polygenic scores derived from GWAS in European populations.

Here, I provide independent lines of evidence that genetic differences at variants associated with EA and CP are consistent with neutral evolution instead of divergent positive selection. First, the fact that education-and-cognitive-performance-associated alleles do not show more genetic differentiation than control SNPs that are not associated with these traits is demonstrated. Second, I test for polygenic selection using polygenic scores computed from within-family effect sizes that minimize the confounding biases mentioned above (Berg & Harpak, 2019; Sohail & Maier, 2019) and did not find a signal of divergent positive selection. Although there is more noise in within-family effect size estimates, Cox et al. (2019) were able to detect signals of polygenic selection for height in a sample of ancient genomes using within-family effect sizes and between-family effect sizes, which suggests that despite the greater noise in within-family estimates, they are still capable of detecting polygenic selection. Additionally, the results presented here build upon the failure of Guo et al. (2018) to find significant genetic differentiation of a different set of education-associated SNPs compared with control SNPs, and the failure of Racimo et al. (2018) to find evidence of divergent selection for educational-attainment-associated SNPs between African and European populations.

The negative results presented here may differ from Piffer (2015, 2019) because the latter potentially suffer from issues caused by applying between-family polygenic scores across ancestry groups, in particular applying European polygenic scores to African populations (Rosenberg et al.,2019; Novembre & Barton, 2018; Lawson

et al., 2019; Haworth et al., 2019; Freese et al., 2019). One reason for this is that the scores might be biased by a variety of factors, including the nonrandom ways that society is geographically structured (Freese et al., 2019; Haworth et al., 2019; Lawson et al., 2019; Novembre & Barton, 2018; Rosenberg et al., 2019). For instance, Black people in the US, for reasons unrelated to genetics, live in areas with poorer air quality and more exposure to environmental toxins (Manduca & Sampson, 2019). These between-family polygenic scores are also known to be biased by the nonrandom ways that people choose their spouse/partner (assortative mating; Morris et al., 2020); the ways that genes interact with different environments (gene–environment interaction; Cheesman et al., 2020; Selzam & Ritchie, 2019; Mostafavi et al., 2020); or differences in genetic features that genome-wide association studies rely on, which create the illusion of systematic differences between African and non-African populations (linkage disequilibrium, genetic drift, epistasis, and ascertainment bias; Coop, 2019; Martin & Gignoux, 2017; Kim et al., 2018). These kinds of biases have previously led to improper inference of natural selection in even closely related European populations (Berg & Harpak, 2019; Sohail, 2019).

Another source of concern is inflation of polygenic scores' explained variance from indirect genetic effects (Ashbrook et al., 2015), also called genetic nurture (Kong & Thorleifsson, 2018). This aspect of polygenic scores appears to be related to features of the home and prenatal environment of parents who carry a genetic signal that is not passed down directly to the offspring (Armstrong-Carter et al., 2020; Lee et al., 2018; Rosenberg et al., 2019; Selzam & Ritchie, 2019). While these genetic nurture effects correlate with educational attainment, they are not directly passed onto offspring and so do not represent direct genetic effects and will not necessarily produce the same effects across diverse populations.

The substantial attenuation of the signal of polygenic selection when using within-family effect sizes compared with between-family effect sizes suggests that results computed by Piffer (2015, 2019) are not due to signals of polygenic selection, but to biases in applying polygenic scores that have been computed with between-family effect sizes across ancestry groups and a failure to employ a formal test for polygenic selection. Other concerns with these studies are the inclusion of data from Richard Lynn (Lynn & Vanhanen, 2006), which introduces another source of systematic bias in the analysis (Wicherts et al., 2010; Wicherts et al., 2010a). Furthermore, the substantial attenuation of polygenic scores when using within-family effect sizes and accounting for ancestral/derived allele frequency biases in European GWAS SNPs shown here and elsewhere (Morris et al., 2020; Selzam & Ritchie, 2019), suggests that the key results of the education PGS analysis performed in Lasker et al. (2019) may be attributable to systematic bias in between-family effect sizes in addition to issues with control variables.

Critics may argue that the observed attenuation is achieved by eliminating genuine genetic effects, regardless of whether they are indirect or direct. One reason we may not want to consider indirect genetic effects is that it is not clear that they act in the same way between populations, especially between countries with very different levels of economic development. Additionally, it is important to remember that proponents of the genetic hypothesis, typically following Jensen's arguments, claim that the genetic differences are largely fixed, and that environmental interventions are unlikely to close the gap. Indirect genetic effects, however, are often more amenable to environmental interventions (Gage et al., 2016). This is bolstered by the observation that much of the effect-size disparity between within- and between-family polygenic scores of cognitive traits is related to socioeconomic status, which can be changed readily through social policies (Cheesman et al., 2020; Morris et al., 2020). Should proponents of the genetic hypothesis argue that indirect genetic effects nonetheless support their claims, note that such a premise would be inconsistent with maintaining any claims of the fixedness of the phenotypic gap, which is a centerpiece of arguments made by Jensen (1969b); Jensen, 1985), Rushton (1990), and Lynn and Vanhanen (2006). Given the lack of evidence for selection demonstrated here, it is imperative to revise previous speculation about the contribution of natural selection to the global variation in cognitive traits (Winegard et al., 2017; Winegard et al., 2020).

4.2 | The magnitude of the genetic contribution to Black–White differences is predicted to be small

One of the central claims of the hereditarian hypothesis, epitomized by Jensen's work (Jensen, 1969b), is that genetic differences cause a substantial portion of the gap in IQ between Black and White populations and the gap cannot be closed by environmental intervention. The biometric techniques employed by Jensen (1969a) to support the hereditarian hypothesis have been critiqued on several fronts, most notably that heritability estimates provide little insight into the causes of between group differences, that heritability estimates do not inform on the efficacy of environmental interventions, and that heritability is likely inflated due to violations of model assumptions when applied to humans (Kempthorne, 1978). Despite these strong criticisms, direct genetic evidence for or against Jensen's arguments have failed to materialize, bar a recent genomic analysis claiming to provide direct evidence in support of the genetic hypothesis (Jensen, 1969a; Lasker et al., 2019).

Contra the genetic hypothesis, the results provided here, using evolutionary genetic techniques, are consistent with genetic differences in variants associated with cognitive performance contributing an insignificant portion of the African-European IQ gap. When the between group variance attributable to trait-associated SNPs is compared with the observed phenotypic between-group variance, over 85% of the between-group variance in IQ is not attributable to additive genetic effects, where at most 4.7–8.5 IQ points could be attributed to such genetic effects. Importantly, this model assumes that the genetic contribution is completely additive, and no gene–environment interactions exist even though both gene–gene (GxG) and gene–environment (GxE) interactions can influence traits. This means that the manner in which a genotype relates to a phenotype, and the magnitude of the genetic contribution, could differ between populations

with different allele frequencies or environments. If the contributions of SNPs can differ between populations due to GxG and GxE, then the effects of a variant on a trait can vary in direction and magnitude within and between populations. This variability could obscure the evolutionary trajectories of populations and lead to erroneous conclusions about patterns of selection and genetic contribution to phenotypic differences if GxG and/or GxE are not accounted for (Coop, 2019; Rosenberg et al., 2019); as in the case of the Alzheimer linked APOE variants (Rajabli & Feliciano, 2018; Tang et al., 1998).

Hence, the mean-expected-difference provided here is likely an overestimate and should be thought of as the maximum mean difference attributable to genetic variation due to genetic drift. It is also important to note that the direction of the mean difference could favor Africans or Europeans with equal likelihood. These genetic estimates fall far short of the observed 30.8 IQ gap between African and European populations. Whatever value the genetic contribution takes, it is important to remember that the effects of gene-by-gene and gene-by-environment interactions will likely reduce this value in reality because the interactive nature of these traits can overwhelm the contribution of genetic drift. Furthermore, there is no reason to believe the genetic contribution to trait differences is immutable to environmental interventions given the known presence of gene–environment interplay in educational attainment.

Even in the scenarios most favorable to the genetic hypothesis, these results are far less than the claimed contribution of hereditarians like Jensen, Rushton, and Lynn. Considering that the assumptions of this model violate core principles of modern population genetics (such as no gene–environment interplay, gene–gene interactions, and similar allelic effect across populations), there is little reason to expect the genetic contribution to be this large in reality. In fact, it is still possible that the genetic contribution to the IQ gap is zero. Above all, these results are consistent with genetic differences having an insignificant effect on the IQ and attainment gap.

4.3 | Where next for the debate on group differences

It might be tempting for supporters of the genetic hypothesis and some neutral readers to think that a positive result from these analyses would be strong support for the genetic hypothesis. I would caution against such a conclusion. As discussed above, these analyses use unsound assumptions about the additivity of genetic effects and the similarity of their behavior between populations. These assumptions are known to be false—for example, in a complex trait like life span, extensive context-dependent allelic effects across several environments and between sexes has been demonstrated in *Drosophila melanogaster* (Huang et al., 2020). Similarly, there is evidence of context dependence of polygenic scores in humans of the same ancestry group, especially for traits like educational attainment (Mostafavi et al., 2020). These model violations are most likely to lead to overestimation of the additive genetic contribution to phenotypic variance. Consequently, should subsequent analyses find larger values, these should not be taken as definitive proof that genetic contributions are that large. Extensive follow-up of how model violations impact estimation would be required to make a determination of any potential genetic contribution. Furthermore such evidence would require extensive functional validation (Flint & Ideker, 2019; Gallagher & Chen-Plotkin, 2018) to demonstrate that causal variants exist in both populations and have similar effects in each, though evidence of such kind would be consistent with the genetic hypothesis and stronger than current evidence. However, without the ability to rule out effects of appropriate environmental interventions, this would still be insufficient to show a fixed genetic difference between populations.

Furthermore, although future studies may claim to show natural selection has acted on these populations, caution is needed in interpreting the impact that selected variants may have on phenotypic differences. Others have noted the difficulty of moving from past selection on variants associated with a trait to substantial genetic contribution to group differences without additional analyses (Harpak & Przeworski, 2020; Rosenberg et al., 2019). The difficulty in distinguishing between divergent positive selection and the stabilizing selection of adjusting to a new environment while leaving the trait optimally unchanged presents another reason for caution in interpretation (Harpak & Przeworski, 2020). The consequence of this context dependency and the ambiguous implications of putative positive results is that providing strong support for the genetic hypothesis is exceedingly difficult, and perhaps impossible without the direct experimentation ordinarily permissible in studies in models systems or in agricultural settings. The ability to do massive reciprocal transplant experiments in plants with replicated genotypes and tight control over the environment is extremely powerful in plant settings (Lowry et al., 2019; Price et al., 2018). In humans, however, a conclusive experiment would need to control for prenatal and early developmental heterogeneity (the subtle and pervasive psychosocial and social effects that contribute to group differences), in order to compare performance in the same or across a range of environments to overcome limitations of observational studies. Such an experiment is of course not moral or ethical. This major hindrance means that while the analyses performed in this study may not find evidence supporting the hereditarian hypothesis, positive results from similar studies are not nearly so dispositive.

5 | CONCLUSION

This analysis employs the current most reliable methods given concerns over cross-population polygenic score analysis. It demonstrates that patterns of genetic differences between African and European populations in the 1000 Genomes Project data set is consistent with neutral evolution and insignificant genetic contribution to the Black–White IQ gap. In other words, the patterns observed in this study can be explained without appealing to the core tenets of the hereditarian hypothesis. Claims made by proponents of the hereditarian hypothesis and recent analyses using polygenic scores (Lasker et al., 2019;

Piffer, 2019) are stronger than what current evidence shows. Given the results presented here, a minimal, insufficiently dispositive requirement for verifying the hereditarian hypothesis is not met. The claims for large, immutable group differences in intelligence and educational attainment are not supported in the least by these analyses.

ACKNOWLEDGMENTS

I would like to thank Michael Edge and Jeremy Berg for helpful methodological suggestions and comments on the manuscript, Charles Roseman for helpful comments and substantial guidance with phenotypic Fst derivations, and Aysu Okbay for sharing within-family effect sizes from Lee et al. (2018). I also thank Cathryn Townsend, Dan Weissman, James Lingford, Jedidiah Carlson, Jonathan Kaplan, Matias Kaplan, Ronald, and twitter users @Rasmansa, @Sillyolyou, @pp0196 for helpful discussion and comments on the manuscript. I am supported by NSF-GRFP DGE-142487.1.

AUTHOR CONTRIBUTIONS

Kevin Bird: Conceptualization; data curation; formal analysis; investigation; methodology; visualization; writing-original draft; writing-review and editing.

CONFLICT OF INTEREST

The author declares this research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

DATA AVAILABILITY STATEMENT

Data and code to reproduce core analyses are available at https://doi.org/10.5281/zenodo.4284020.

ORCID

Kevin A. Bird https://orcid.org/0000-0002-3174-3646

REFERENCES

1000 Genomes Project Consortium. (2015). A global reference for human genetic variation. *Nature*, *526*(7571), 68–74.

Armstrong-Carter, E., Trejo, S., Hill, L. J., Crossley, K. L., Mason, D., & Domingue, B. W. (2020). The earliest origins of genetic nurture: The prenatal environment mediates the association between maternal genetics and child development. *Psychological Science*, *31*(7), 781–791.

Ashbrook, D. G., Gini, B., & Hager, R. (2015). Genetic variation in offspring indirectly influences the quality of maternal behaviour in mice. *eLife*, *4*, e11814.

Berg, J. J., & Coop, G. A. (2014). Population genetic signal of polygenic adaptation. *PLoS Genetics*, *10*, e1004412.

Berg, J. J., & Harpak, A. (2019). Reduced signal for polygenic adaptation of height in UKbiobank. *eLife*, *8*, e39725.

Bulik-Sullivan, B. K., Loh, P. R., Finucane, H. K., Ripke, S., Yang, J., Patterson, N., Daly, M. J., Price, A. L., & Neale, B. M. (2015). LD score regression distinguishes confounding from Polygenicity in genome-wide association studies. *Nature Genetics*, *47*, 291–295.

Carratto, T. M. T., Carratto, T. M. T., Marcorin, L., Debortoli, G., Silva, G. V., Fracasso, N. C. A., Oliveira, M. L. G., Pereira, A. L. E., Silva, A. B. C., Donadi, E. A., Simões, A. L., Castelli, E. C., Norton, H. L., Parra, E. J., & Mendes-Junior, C. T. (2019). Evaluation of the HIrisPlex-S system in a Brazilian population sample. *Forensic Science International: Genetics Supplement Series*, *7*, 794–796.

Centerwall, B. S. (1978). The use of racial admixture as evidence in intelligence research: A critique. *Human Genetics*, *45*, 237–238.

Cernovsky, Z. Z. (1990). Race and brain weight: A note on J. P. Rushton's conclusions. *Psychological Reports*, *66*, 337–338.

Cernovsky, Z. Z., & Litman, L. C. (1993). Re-analyses of J.P. Rushton's crime data. *Canadian Journal of Criminology*, *35*, 31–36.

Cheesman, R., Hunjan, A., Coleman, J. R. I., Ahmadzadeh, Y., Plomin, R., McAdams, T. A., Eley, T. C., & Breen, G. (2020). Comparison of adopted and nonadopted individuals reveals gene–environment interplay for education in the UKbiobank. *Psychological Science*, *31*(5), 582–591.

Chen, M., & Chiang, C. W. (2020). Allele frequency differentiation at height-associated SNPs among continental human populations. *Preprint bioRxiv* doi:https://doi.org/10.1101/2020.09.28.317552

Chipuer, H. M., Rovine, M. J., & Plomin, R. (1990). LISREL modeling: Genetic and environmental influences on IQ revisited. *Intelligence*, *14*, 11–29.

Cloninger, C. R., Rice, J., & Reich, T. (1979). Multifactorial inheritance with cultural transmission and assortative mating. II. A general model of combined polygenic and cultural inheritance. *American Journal of Human Genetics*, *31*, 176–198.

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

Cox, S. L., Ruff, C. B., Maier, R. M., & Mathieson, I. (2019). Genetic contributions to variation in human stature in prehistoric Europe. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(43), 21484–21492.

Danecek, P., & Auton, A. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*, 2156–2158.

Devlin, B., Daniels, M., & Roeder, K. (1997). The heritability of IQ. *Nature*, *388*, 468–471.

Duncan, L., & Shen, H. (2019). Analysis of polygenic risk score usage and performance in diverse human populations. *Nature Communications*, *10*, 3328.

Ebbesen, C. L. (2020). Flawed estimates of cognitive ability in Clark et al. Psychological Science, 2020. *Preprint at PsyArxiv* https://doi.org/10.31234/osf.io/tzr8c

Flint, J., & Ideker, T. (2019). The great hairball gambit. *PLoS Genetics*, *15*, e1008519.

Freedman, B. I., Limou, S., Ma, L., & Kopp, J. B. (2018). APOL1-associated nephropathy: A key contributor to racial disparities in CKD. *American Journal of Kidney Diseases*, *72*, S8–S16.

Freedman, M. L., & Haiman, C. A. (2006). Admixture mapping identifies 8q24 as a prostate cancer risk locus in African-American men. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 14068–14073.

Freese, J., Domingue, B., Trejo, S., Sicinski, K. & Herd, P. (2019) Problems with a causal interpretation of polygenic score differences between Jewish and non-Jewish respondents in the Wisconsin longitudinal study. Preprint SocArxiv doi:https://doi.org/10.31235/osf.io/eh9tq.

Gage, S. H., Smith, G. D., Ware, J. J., Flint, J., & Munafò, M. R. (2016). G = E: What GWAS can tell us about the environment. *PLoS Genetics*, *12*, e1005765.

Gallagher, M. D., & Chen-Plotkin, A. S. (2018). The post-GWAS era: From association to function. *American Journal of Human Genetics*, *102*, 717–730.

Gorey, K. M., & Cryns, A. G. (1995). Lack of racial differences in behavior: A quantitative replication of Rushton's (1988) review and an independent meta-analysis. *Personality and Individual Differences*, *19*, 345–353.

Graffelman, J., Jain, D., & Weir, B. A. (2017). Genome-wide study of hardy–Weinberg equilibrium with next generation sequence data. *Human Genetics*, *136*, 727–741.

Graves, J. L. (2002). What a tangled web he weaves. *Anthropological Theory*, *2*, 131–154.

Guo, J., Guo, J., Wu, Y., Zhu, Z., Zheng, Z., Trzaskowski, M., Zeng, J., Robinson, M. R., Visscher, P. M., & Yang, J. (2018). Global genetic differentiation of complex traits shaped by natural selection in humans. *Nature Communications*, *9*, 1–9.

Harpak, A., & Przeworski, M. (2020). The evolution of group differences in changing environments. *Preprint arXiv* arXiv:2010.03048.

Haworth, S., Mitchell, R., Corbin, L., Wade, K. H., Dudding, T., Budu-Aggrey, A., Carslake, D., Hemani, G., Paternoster, L., Smith, G. D., Davies, N., Lawson, D. J., & Timpson, N. J. (2019). Apparent latent structure within the UKbiobank sample has implications for epidemiological analysis. *Nature Communications*, *10*, 1–9.

Hill, W. D., & Arslan, R. C. (2018). Genomic analysis of family data reveals additional genetic effects on intelligence and personality. *Molecular Psychiatry*, *23*, 2347–2362.

Hill, W. D., & Marioni, R. E. (2019). A combined analysis of genetically correlated traits identifies 187 loci and a role for neurogenesis and myelination in intelligence. *Molecular Psychiatry*, *24*, 169–181.

Holland, D., Frei, O., Desikan, R., Fan, C. C., Shadrin, A. A., Smeland, O. B., Sundar, V. S., Thompson, P., Andreassen, O. A., & Dale, A. M. (2020). Beyond SNP heritability: Polygenicity and discoverability of phenotypes estimated with a univariate Gaussian mixture model. *PLoS Genetics*, *16*(5), e1008612.

Huang, W., Campbell, T., Carbone, M. A., Jones, W. E., Unselt, D., Anholt, R. R. H., & Mackay, T. F. C. (2020). Context-dependent genetic architecture of drosophila life span. *PLoS Biology*, *18*(3), e3000645.

Jensen, A. R. (1969a). How much can we boost IQ and scholastic achievement. *Harvard Educational Review*, *39*, 1–123.

Jensen, A. R. (1969b). Reducing the heredity-environment uncertainty: A reply. *Harvard Educational Review*, *39*, 449–483.

Jensen, A. R. (1974) Race and Mental Ability. in *Symposium of the institute. of Biology on 'Racial Variation in Man'* (Royal Geographical Society).

Jensen, A. R. (1985). Compensatory education and the theory of intelligence. *The Phi Delta Kappa*, *66*, 554–558.

Kempthorne, O. (1978). A biometrics invited paper: Logical, epistemological and statistical aspects of nature-nurture data interpretation. *Biometrics*, *34*, 1–23.

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

Kong, A., & Thorleifsson, G. (2018). The nature of nurture: Effects of parental genotypes. *Science*, *359*, 424–428.

Lasker, J., Pesta, B. J., Fuerst, J. G. R., & Kirkegaard, E. O. W. (2019). Global ancestry and cognitive ability. *Psychiatry*, *1*, 431–459.

Lawson, D. J., Davies, N. M., Haworth, S., Ashraf, B., Howe, L., Crawford, A., Hemani, G., Smith, G. D., & Timpson, N. J. (2019). Is population structure in the genetic biobank era irrelevant, a challenge, or an opportunity? *Human Genetics*, *139*(1), 23–41.

Lee, J. J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., Nguyen-Viet, T. A., Bowers, P., Sidorenko, J., Karlsson Linnér, R., Fontana, M. A., Kundu, T., Lee, C., Li, H., Li, R., Royer, R., Timshel, P. N., Walters, R. K., Willoughby, E. A., … Cesarini, D. (2018). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature Genetics*, *50*, 1112–1121.

Loehlin, J. C. (1978). Heredity-environment analyses of Jencks's IQ correlations. *Behavior Genetics*, *8*, 415–436.

Loehlin, J. C. (2000). Group differences in intelligence. In *Handbook of Intelligence* (pp. 176–194). New York: Cambridge University Press. https://doi.org/10.1017/cbo9780511807947.010

Lowry, D. B., Lovell, J. T., Zhang, L., Bonnette, J., Fay, P. A., Mitchell, R. B., Lloyd-Reilley, J., Boe, A. R., Wu, Y., Rouquette, F. M., Jr., Wynia, R. L., Weng, X., Behrman, K. D., Healey, A., Barry, K., Lipzen, A., Bauer, D., Sharma, A., Jenkins, J., … Juenger, T. E. (2019). QTL× environment interactions underlie adaptive divergence in switchgrass across a large latitudinal gradient. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(26), 12933–12941.

Lynch, M. (1990). The rate of morphological evolution in mammals from the standpoint of the neutral expectation. *The American Naturalist*, *136*, 727–741.

Lynn, R. (2006) *Race Differences in Intelligence: An Evolutionary Analysis*. Augusta, GA: Washington Summit Publishers.

Lynn, R. & Vanhanen, T. (2006) *IQ and Global Inequality*. Augusta, GA: Washington Summit Publishers.

MacEachern, S. (2006). Africanist archaeology and ancient IQ: Racial science and cultural evolution in the twenty-first century. *World Archaeology*, *38*(1), 72–92.

Manduca, R., & Sampson, R. J. (2019). Punishing and toxic neighborhood environments independently predict the intergenerational social mobility of black and white children. *Proceedings of the National Academy of Sciences of the United States of America*, *116*, 7772–7777.

Martin, A. R., & Gignoux, C. R. (2017). Human demographic history impacts genetic risk prediction across diverse populations. *American Journal of Human Genetics*, *100*, 635–649.

Morris, T. T., Davies, N. M., Hemani, G., & Smith, G. D. (2020). Population phenomena inflate genetic associations of complex social traits. *Science Advances*, *6*, eaay0328.

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

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

Pesta, B. J., & Poznanski, P. J. (2014). Only in America: Cold winters theory, race, IQ and well-being. *Intelligence*, *46*, 271–274.

Piel, F. B., & Patil, A. P. (2010). Global distribution of the sickle cell gene and geographical confirmation of the malaria hypothesis. *Nature Communications*, *1*, 104.

Piffer, D. (2015). A review of intelligence GWAS hits: Their relationship to country IQ and the issue of spatial autocorrelation. *Intelligence*, *53*, 43–50.

Piffer, D. (2019). Evidence for recent polygenic selection on educational attainment and intelligence inferred from GWAS hits: A replication of previous findings using recent data. *Psychiatry*, *1*, 55–75.

Polderman, T. J. C., & Benyamin, B. (2015). Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nature Genetics*, *47*, 702–709.

Price, N., Moyers, B. T., Lopez, L., Lasky, J. R., Monroe, J. G., Mullen, J. L., Oakley, C. G., Lin, J., Ågren, J., Schrider, D. R., Kern, A. D., & McKay, J. K. (2018). Combining population genomics and fitness QTLs to identify the genetics of local adaptation in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(19), 5028–5033.

Purcell, S., & Neale, B. (2007). PLINK: A toolset for whole-genome association and population-based linkage analysis. *American Journal of Human Genetics*, *81*, 559–575.

Racimo, F., Berg, J. J., & Pickrell, J. K. (2018). Detecting polygenic adaptation in admixture graphs. *Genetics*, *208*, 1565–1584.

Rajabli, F., & Feliciano, B. E. (2018). Ancestral origin of ApoE ε4 Alzheimer disease risk in Puerto Rican and African American populations. *PLoS Genetics*, *14*, e1007791.

Rao, D. C., Morton, N. E., Lalouel, J. M., & Lew, R. (1982). Path analysis under generalized assortative mating. II. American I.Q. *Genetical Research*, *39*, 187–198.

Relethford, J. H., & Blangero, J. (1990). Detection of differential gene flow from patterns of quantitative variation. *Human Biology*, *62*, 5–25.

Relethford, J. H., Crawford, M. H., & Blangero, J. (1997). Genetic drift and gene flow in post-famine Ireland. *Human Biology*, *69*, 443–465.

Rosenberg, N. A., Edge, M. D., Pritchard, J. K., & Feldman, M. W. (2019). Interpreting polygenic scores, polygenic adaptation, and human

phenotypic differences. *Evolution, Medicine, and Public Health*, *2019*, 26–34.

Rothstein, J., & Wozny, N. (2013). Permanent income and the black-white test score gap. *Journal of Human Resources*, *48*(3), 510–544.

Rushton, J. P. (1990). Race differences and r/K theory. *Ethology and Sociobiology*, *11*, 131–140.

Rushton, J. P. (1996) *Race, evolution, and behavior: A life history perspective*. (Transaction Publishing).

Rushton, J. P. (2000) *Race, Evolution, and Behavior: A Life History Perspective*. (Transaction Publishing).

Rushton, J. P., & Jensen, A. R. (2005). Thirty years of research on race differences in cognitive ability. *Psychology, Public Policy, and Law*, *11*, 235–294.

Saini, A. Superior: the return of race science. (2019). Boston, MA: Beacon Press.

Selzam, S., & Ritchie, S. J. (2019). Comparing within- and between-family polygenic score prediction. *American Journal of Human Genetics*, *105*, 351–363.

Sohail, M., & Maier, R. M. (2019). Polygenic adaptation on height is overestimated due to uncorrected stratification in genome-wide association studies. *eLife*, *8*, e39702.

Tang, M. X., Stern, Y., Marder, K., Bell, K., Gurland, B., Lantigua, R., Andrews, H., Feng, L., Tycko, B., & Mayeux, R. (1998). The APOE-epsilon4 allele and the risk of Alzheimer disease among African Americans, whites, and Hispanics. *Journal of the American Medical Association*, *279*, 751–755.

Thomas, D. (2016). Racial IQ differences among transracial adoptees: Fact or Artifact? *Journal of Intelligence*, *5*, 1.

Waldman, I. D., Weinberg, R. A., & Scarr, S. (1994). Racial-group differences in IQ in the Minnesota transracial adoption study: A reply to Levin and Lynn. *Intelligence*, *19*, 29–44.

Wicherts, J. M., Borsboom, D., & Dolan, C. V. (2010). Evolution, brain size, and the national IQ of peoples around 3000 years B.C. *Personality and Individual Differences*, *48*, 104–106.

Wicherts, J. M., Borsboom, D., & Dolan, C. V. (2010a). Why national IQs do not support evolutionary theories of intelligence. *Personality and Individual Differences*, *48*, 91–96.

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag. https://ggplot2.tidyverse.org

Winegard, B., Winegard, B., & Anomaly, J. (2020). Dodging Darwin: Race, evolution, and the hereditarian hypothesis. *Personality and Individual Differences*, *160*, 109915.

Winegard, B., Winegard, B., & Boutwell, B. (2017). Human biological and psychological diversity. *Evolutionary Psychological Science*, *3*, 159–180.

Young, A. I., Frigge, M. L., Gudbjartsson, D. F., Thorleifsson, G., Bjornsdottir, G., Sulem, P., Masson, G., Thorsteinsdottir, U., Stefansson, K., & Kong, A. (2018). Relatedness disequilibrium regression estimates heritability without environmental bias. *Nature Genetics*, *50*, 1304–1310.

Zaidi, A. A., & Mathieson, I. (2020). Demographic history mediates the effect of stratification on polygenic scores. *eLife*, *9*, e61548.

Zietsch, B. P., & Sidari, M. J. A. (2019). A critique of life history approaches to human trait covariation. *Evolution and Human Behavior*, *41*, 527–535. https://doi.org/10.1016/j.evolhumbehav.2019.05.007

How to cite this article: Bird KA. No support for the hereditarian hypothesis of the Black–White achievement gap using polygenic scores and tests for divergent selection. *Am J Phys Anthropol*. 2021;1–12. https://doi.org/10.1002/ajpa.24216