Commentary

Misestimation of Expected Genetic Differences: A Statistical Note on Some Recent Papers

John Fuerst[1]

1. Department of Biotechnology, University of Maryland, United States

The expected magnitude of phenotypic differences between human populations under genetic drift is often underestimated. This commentary challenges recent claims of minimal differences by addressing statistical weaknesses in Lala & Feldman[1], Gusev[2][3], and Roseman & Bird[4], specifically: the misinterpretation of polygenicity's role in genetic drift, the failure to adjust for diploidy, and the use of non-standard effect size metrics. Using typically reported F_{ST} values and heritabilities, medium to large phenotypic differences are expected under genetic drift across major human biogeographic ancestry groups. Specific phenotypic differences may also be shaped by other evolutionary forces, such as convergent or divergent selection, and environmental factors. By clarifying the mathematical basis for expected differences, this commentary advances the discussion on genetic variance and its implications for human phenotypic diversity.

Correspondence: papers@team.qeios.com — Qeios will forward to the authors

1. Introduction

Under conditions of neutrality, the expected phenotypic variance between populations is approximately proportional to their average genetic variance. Given commonly reported narrow heritabilities for traits within major human biogeographic ancestry groups (e.g., East Asians, Europeans, and Sub-Saharan Africans) and typically reported F_{ST} values among them (e.g., [5]), one would anticipate medium to large phenotypic differences arising under genetic drift

Biologists have frequently underplayed the expected magnitude of these differences, a tendency that some attribute to political considerations. David Reich[6], who runs a major genetics lab at Harvard

University, elaborates on this issue (*p*. 254):

> When asked about the possibility of biological differences among humans, we have tended to obfuscate, making mathematical statements in the spirit of Richard Lewontin about the average difference between individuals from within any one population being around six times greater than the average difference between populations... But this carefully worded formulation is deliberately masking the possibility of substantial average differences in biological traits across populations.

Common assertions such as 'genetic differences among populations are small in comparison to variation within' are not technically incorrect, yet such assertions often mislead. By similar reasoning, what are considered large effects in the social and biomedical sciences could also be dismissed as 'small in comparison to variation within'. Several researchers have moved beyond merely suggesting 'small' differences; they purport to demonstrate small expected differences either theoretically or quantitatively. In this context, recent statements by Lala & Feldman[1], Gusev[2][3], and Roseman and Bird[4] are critically reviewed.

In conservation biology, it is standard practice to compare phenotypic and genetic variance in order to detect signals of selection[7]. This practice involves what are known as Q_{ST}-F_{ST} comparisons, where Q_{ST} measures the additive genetic differentiation in quantitative traits among populations, and F_{ST} measures genetic differentiation based on allele frequencies at genetic loci. Whitlock & Guillaume[8] provide the formula for Q_{ST} in context to diploids:

$$Qst = \frac{\sigma^2_{GB}}{\sigma^2_{GB} + 2\sigma^2_{GW}}, \tag{1}$$

where σ^2_{GB} and σ^2_{GW} are, respectively, the between- and within- group phenotypic variances due to additive genetics.

By rearranging the terms, the formula can be expressed as:

$$\sigma^2_{GB} = \frac{2Qst * \sigma^2_{GW}}{1 - Qst} \tag{2}$$

In this formula, the factor of 2 adjusts for the distribution of genetic variance among diploids, where approximately half of the variance occurs within individuals between homologous chromosomes. Given that F_{ST} and Q_{ST} are equivalent under conditions of neutrality, this formula can be adapted to predict expected phenotypic variance attributable to additive genetic variation as follows:

$$\sigma^2_{GB} = \frac{2Fst * \sigma^2_{GW}}{1 - Fst} \quad (3)$$

This formula can be expressed in terms of within group heritability (h^2w), by noting that

h^2w = $\sigma^2_{GW}/\sigma^2_{PW}$, in which case:

$$\sigma^2_{GB} = \frac{2Fst * h^2_W}{1 - Fst}\sigma^2_{PW} \quad (4)$$

When between group variance is solely additive genetic, as in our condition when solving for expected genetic differences under environmental equality, then total variance, (σ^2_T), is the sum of between group additive variance, (σ^2_{GB}), and total within group phenotypic variance (σ^2_{pw}):

$$\sigma^2_{T=}\, \sigma^2_{GB} + \sigma^2_{PW} \quad (5)$$

Substituting equation 4 into 5 and rearranging we can restate σ^2_{PW} as:

$$\sigma^2_{PW} = \frac{\sigma^2_T}{1 + \frac{2Fst*h^2_W}{1-Fst}} \quad (6)$$

Substituting equation 6 in equation 4, and setting \(\sigma_{T = \ }^{2}1,\\)we derive a standardized formula:

$$\sigma^2_{GB} = \frac{\frac{2Fst*h^2_W}{1-Fst}}{1 + \frac{2Fst*h^2_W}{1-Fst}} = \frac{2Fst * h^2_W}{1 - Fst + 2Fst * h^2_W} \quad (7)$$

When h^2w = 1.00 and F_{ST} = 1, this formula yields a σ^2_{GB} of 18%, matching the estimate of “only” 18% mentioned by Yair & Coop[9] for these conditions.

Conceptually, F_{ST} values range from 0 (no genetic differentiation, all alleles shared) to 1 (complete genetic differentiation, no alleles shared). However, in practice, F_{ST} values are constrained by within-population heterozygosity[10][11]. As a result, highly variable markers like microsatellites, which typically exhibit high heterozygosity, often produce maximum F_{ST} values significantly below 1. For example, Meirmans & Hedrick[10] note:

> To illustrate this relationship, Fig. 1 gives the joint values of F_{ST} and H_S found in the past 4 years in Molecular Ecology.... Notice that the observed range of FST is always less than H_S and that the range of F_{ST} becomes very small when H_S is large. For example when H_S = 0.9, a value that is commonly encountered for microsatellite markers, the maximum possible value of F_{ST} is 0.1.

For this reason, for Q_{ST}-F_{ST} comparisons, it is commonly recommended to use markers with lower variability, such as SNPs[12], to align Q_{ST} and F_{ST} more closely on the same scale. However, even with SNPs, the maximum F_{ST} value is usually well below 1, which may lead to an underestimation of genetic variance (on a scale from 0 to 1).

2. Critiques

2.1. Lala & Feldman[1].

Lala and Feldman[1] argue against the possibility of ancestry group differences in IQ and scholastic attainment by asserting that "recent human evolution has been dominated by drift rather than selection," and that drift could not lead to large differences in highly polygenic traits. They contend:

> However, if, as the data suggest, intelligence is affected by many genes of small effect, it becomes implausible that IQ differences between socially defined races arose through a process of random genetic drift; this is relevant because analyses of genetic variation show that recent human evolution has been dominated by drift rather than selection (89). The probability that a long sequence of random changes would all go in the same direction, leading to increases in the intelligence of one population and not others, approaches zero.

The argument that non-trivial differences in a trait are implausible under drift is based on a misunderstanding, as noted by Yair & Coop[9]. This misconception contradicts well-established evolutionary theory. Yair and Coop[9] clarify:

> Naively, as trait-increasing alleles underlying a neutral trait are equally likely to drift up or down, one might think that over many loci we expect only a small mean difference between populations. However, the polygenic score is a sum rather than a mean, and so each locus we add into the score is like an additional step in the random walk that two populations take away from each other [99]. We expect the variance among populations, i.e. the average squared difference between population means and the global mean, to be 2VA F_{ST} [11,14].

Edge and Rosenberg[13][14] note that the expected magnitude of differences under drift is independent of polygenicity. If highly polygenic traits cannot diverge under drift and divergent selection is rare, as Lala &

Feldman[1] suggest, how do they explain the substantial anatomical differences across human ancestry groups documented by physical anthropologists?

Lala and Feldman[1] further argue that the global F_{ST} is "low" on average, suggesting minimal genetic subdivision across human populations and implying little differentiation between socially defined races. However, in conservation biology, a frequently used rule of thumb for interpreting F_{ST} with biallelic markers, based on Sewall Wright's[15] proposal and aligning somewhat with effect size interpretations in social and biomedical sciences[16], is:

- 0.00 to 0.05 = little genetic differentiation
- 0.05 to 0.15 = moderate genetic differentiation
- 0.15 to 0.25 = great genetic differentiation
- 0.25 to 1.00 = very great genetic differentiation

The 1000 Genomes Project Consortium[17] data, cited by Lala and Feldman from Supplementary Information Table 5, provides specific F_{ST} estimates between populations classified as "East Asian Ancestry," "European Ancestry," and "African Ancestry":

	European Ancestry	**African Ancestry**
East Asian Ancestry	.10	.152
European Ancestry		.125

These pairwise F_{ST} values, ranging from 0.10 to 0.152, indicate moderate to great genetic differentiation according to these conservation biology standards. As shown below (Table 1) these magnitudes of F_{ST} differences imply medium-to-large to large expected differences for traits under drift given typically reported kinship-based heritabilities.

2.2. Gusev[2][3]

Gusev[2][3] acknowledges that human populations could diverge substantially in highly polygenic traits under drift but argues that the magnitude of the divergence is equivalent to 1VA F_{ST} -- thus with no adjustment for ploidy. Gusev[2] states:

> In the context of population differences — a focus of the piece in The Atlantic — direct/within-family heritability provides an upper bound on how much a trait can drift between populations under neutrality (see [14] and summary). For educational attainment, for example, we can already calculate that the expected variance between continental populations under neutrality is minuscule: heritability*Fst = 0.04*0.15 = 0.006

Further expanding, in Gusev[3]:

> As a consequence, "*the proportion of heritable variance in the trait attributable to genetic differences between the populations*" (Q_{ST}) is approximately equal to the cross-population (Hudson) F_{ST} (which [13] rederive as $F_{ST,I}$). This quantity also does not depend on polygenicity. *Note*: the relationship is sometimes reported as $2*F_{ST}*h2$, but this is only an approximation for Nei's F_{ST}: Nei's F_{ST} is approximately equal to half of Hudson's F_{ST} when the former is close to 0 or 1 (see [8.6]).

Gusev[2][3] interprets the factor of 2 in this commonly used formula (e.g., [7]) as a correction for Nei's F_{ST}, which is approximately half of Nei's G_{ST} and Hudson's F_{ST} when the number of populations is small. This interpretation is incorrect. First, it would not make sense because F_{ST}~ Q_{ST} comparisons typically involve many populations, under which condition Nei's $F_{ST} \approx$ Nei's G_{ST}. Second, contrary to Gusev's claims (including personal communications, 2025), Nei's G_{ST}, Weir-Cockerham's θ, Wright's F_{ST}, and Hudson's F_{ST} converge under ideal conditions (e.g., large sample sizes, balanced populations, no selection, and similar allele frequency distributions)[18]. Rather, the factor of 2 in the formula represents a diploidy adjustment, as discussed by Edge and Rosenberg[14][13] and Whitlock[19]. It is not an estimator specifc adjustemt. The variance decomposition is illustrated in Figure 1.

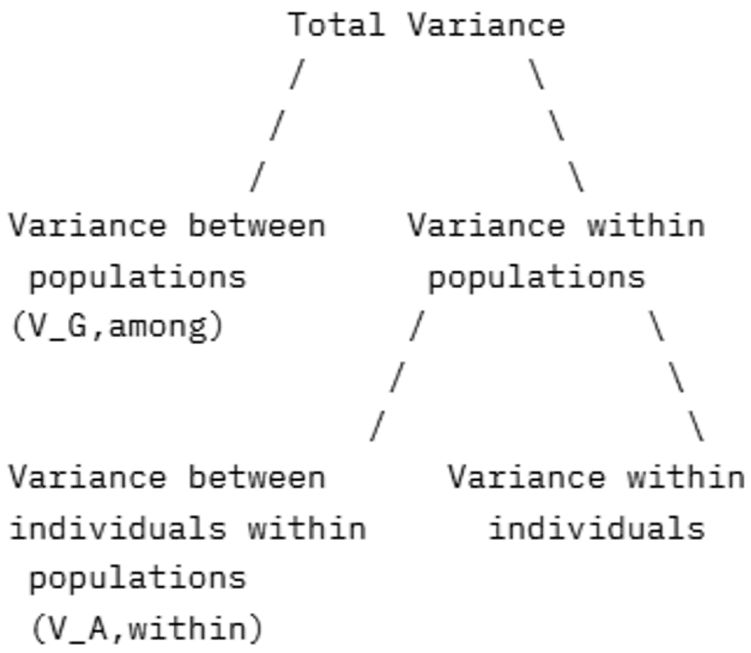

Figure 1. Variance decomposition between populations and between diploid individuals

In diploids, the additive genetic variance within populations (VA, within) reflects differences between individuals, as shown in Figure 1's "Variance between individuals within populations." The factor of two in the QST equation's denominator (VG, among ÷ [VG, among + 2VA, within]) accounts for roughly half the genetic variance among diploids being within individuals, not between them, due to two homologous chromosomes[19]. VA within scales with the additive effects of two alleles per locus, doubling the within-population variance contribution relative to a haploid. Figure 1's "Variance between chromosomes within individuals," part of "Variance within individuals," captures this within-individual variance, which is not part of VA, within. Thus, the factor of 2 in QST adjusts for this by ensuring proper scaling of VA, within for diploids.

2.3. Roseman & Bird[4]

Roseman & Bird[4] state: "We wish to lay particular emphasis on the following point: Under the neutral additive expectation … the expected difference between two lineages as sampled randomly after evolving under random genetic drift is 0." If this is intended to suggest that the expected difference in a specific

trait under drift between two populations is 0, this interpretation is inconsistent with established theory, as also discussed in our commentary on Lala & Feldman[1].

Despite this statement, Roseman & Bird[4] compute σ_B^2 correctly. However, they nonetheless conclude that "unless there is pronounced natural selection acting to differentiate pairs of groups, large pairwise differences between groups would rarely occur" and that "The only way that large amounts of evolutionary divergence among groups in IQ could be reconciled with the FST values estimated using neutrally evolving polymorphisms is if strong natural selection had acted to make the groups diverge from one another."

They base this conclusion on their computed "expected absolute difference between groups". For this statistic, they reference equation 4 in Bird[20] and provide a formula in their appendix 2:

$$E[|\Delta_{i,j}|\ ] = \frac{2\sigma B}{\sqrt{\pi}} = \frac{2\sqrt{\sigma_B^2}}{1.772} \tag{8}$$

where σ_B^2 represents the between-group additive genetic variance.

I refer to this statistic as Bird's *b*. Interpretative claims about the magnitude of differences (e.g., "large pairwise differences"), the use of a related metric in Bird[20], and the stated goal of evaluating claims in the "hereditarian race science literature" where the focus "is on understanding the absolute number of, say, IQ points" strongly suggest that Bird's *b* is treated as equivalent to Cohen's *d*, a standard effect size with established benchmarks for classifying group differences as "large" (i.e., d ≥ 0.8). However, Bird's *b* is not equivalent to Cohen's *d*. The former includes a constant denominator, while the latter accounts for within-group variance, leading to a clear discrepancy. This difference can be demonstrated straightforwardly. The formula for Cohen's *d*, assuming equal variances within groups, is:

$$d = \frac{M_1 - M_2}{\sigma W_{pooled}} = = \frac{M_1 - M_2}{\sqrt{\sigma^2 W_{pooled}}} \tag{9}$$

Under the law of total variance, the total variance \(\sigma_{total}^{2}\ \\)is the sum of within-group variance σ_W^2 and between-group variance σ_B^2. The variance between groups can then be expressed as:

$$\sigma_{between}^2 = \frac{n_1(M_1 - M)^2 - n_2(M_2 - M)^2}{n_1 + n_2 - 1} \tag{10}$$

Noting that

$$M_1 - M = M_1 - \frac{n_1 M_1 - n_2 M_2}{n_1 + n_2} \text{ and } M_2 - M = M_1 - \frac{n_1 M_1 - n_2 M_2}{n_1 + n_2} \tag{11}$$

Assuming equal sample sizes, we can substitute and rearrange the terms as:

$$M_1 - M_2 = \sqrt{\sigma_B^2 * 4(\frac{2n-1}{n})} \tag{12}$$

Under the condition of large sample sizes $\frac{2n-1}{n} \approx 1, \ and\ we\ get$:

$$M_1 - M_2 = 2\sqrt{\sigma_B^2} \tag{13}$$

This result corresponds to the numerator, but not denominator, in Bird's equation for Bird's *b*. When variances are the same, we obtain:

$$d \approx \frac{2\sqrt{\sigma_B^2}}{\sqrt{\sigma^2}} = 2\sqrt{\frac{\eta^2}{1-\eta^2}} \tag{14}$$

which is equivalent to converting Cohen's *d* from eta-squared.

It's evident that the denominator in Bird's *b* differs from that in the usual Cohen's *d* formula. Bird's *b* divides by a constant, not by within-group variance. As a result, Bird's *b* underestimates standardized differences in proportion to between-group variance. For example, at σ_B^2 =0.5, Bird's *b* is 0.8, while Cohen's *d* is 2; at σ_B^2 =0.9, Bird's *b* is 1.07, while Cohen's *d* is 6. Thus, at best, Bird's *b* is an idiosyncratic metric, not equivalent to those commonly used in the social and biomedical sciences.

To illustrate, Table 1 presents the expected differences as expressed in Bird's *b* and in Cohen's *d* across various F_{ST} values and heritabilities. Notably, Roseman & Bird's[4] "very small absolute differences" in Bird's *b* turn out to be medium to large sized differences in Cohen's *d*.

Given typically reported kinship-based heritabilities (e.g., [21]) and typical SNP-based F_{ST} values between major human populations, we should anticipate medium to large-sized differences in arbitrary traits under neutral divergence. For example, Roseman & Bird[4] adopt a F_{ST} = 0.12 for SNPs underlying educational and intelligence-related traits based on the results of Bird[20], which compares Europeans to Africans. When h^2 = .35 / .50, Formula 7 yields σ_{GB}^2 = 0.087 / .120, which, with Formula 14, yields *d* = .618 / .739 – a predicted medium-to-large sized effect owing to exclusively additive genetic differences. While factors discussed below might lead us to not expect such pronounced behavioral differences, this magnitude of divergence is consistent with many anthropometric traits.

F_{ST}	h^2	σ^2_{GB}	Cohen's *d*	Bird's *b*	Interpretation
0.05	0.05	0.005	0.145	0.08	Small
0.05	0.2	0.021	0.290	0.16	Small to Medium
0.05	0.35	0.036	0.384	0.22	Small to Medium
0.05	0.5	0.050	0.459	0.26	Small to Medium
0.05	0.65	0.064	0.523	0.3	Medium
0.1	0.05	0.011	0.211	0.12	Small
0.1	0.2	0.043	0.422	0.24	Small to Medium
0.1	0.35	0.072	0.558	0.31	Medium to Large
0.1	0.5	0.100	0.667	0.38	Medium to Large
0.1	0.65	0.126	0.760	0.43	Medium to Large
0.15	0.05	0.017	0.266	0.15	Small to Medium
0.15	0.2	0.066	0.531	0.3	Medium
0.15	0.35	0.110	0.703	0.4	Medium to Large
0.15	0.5	0.150	0.840	0.47	Large
0.15	0.65	0.187	0.958	0.54	Large
0.2	0.05	0.024	0.316	0.18	Small to Medium
0.2	0.2	0.091	0.632	0.36	Medium to Large
0.2	0.35	0.149	0.837	0.47	Large
0.2	0.5	0.200	1.000	0.56	Large
0.2	0.65	0.245	1.140	0.64	Large

Table 1. Relation between F_{ST}, h^2, σ^2_{GB}, Cohen's *d*, and Bird's *b*, along with the typical interpretation of the effect sizes

3. Conclusion

There is a tendency among some biologists to downplay the magnitude of genetic differences between human populations, which may reflect an effort to avoid conclusions similar to those the young Franz Boas drew:

> It does not seem probable that the minds of races which show variations in their anatomical structure should act in exactly the same manner. Differences of structure must be accompanied by differences of function, physiological as well as psychological; and, as we found clear evidence of difference in structure between the races, so we must anticipate that differences in mental characteristics will be found.[22].

Lala & Feldman[1], Gusev[2][3], and Roseman and Bird[4] are notable in this context for their theoretical quantitative claims, which extend beyond carefully worded qualitative statements. Lala & Feldman[1] assert that differences in polygenic traits under genetic drift cannot be substantial. Gusev[2] [3] acknowledges the widely accepted relationship between expected phenotypic variance and genetic variance but fails to account for the necessary adjustment for diploidy. While Roseman and Bird[4] do incorporate this adjustment, they introduce an effect-size-like statistic — denoted here as Bird's *b* — that does not align with commonly used metrics in the social and biomedical sciences, where established interpretive guidelines are available.

While these authors focus on academic achievement and intelligence, the argument is more general: medium to large differences between human ancestry groups can arise without pronounced natural selection. The evolutionary default is not zero phenotypic difference but rather differences proportional to neutral genetic divergence, adjusted for ploidy.

Although this commentary emphasizes statistical expectations under neutral genetic drift, it is acknowledged that complex behavioral traits like intelligence may exhibit smaller-than-expected differences due to factors such as low narrow-sense heritability or stabilizing/convergent selection across populations. For example, personality traits—unlike many anthropometric differences—often show only small differences between ancestry groups within the same country (e.g., [23]).

Understanding these expected differences can illuminate the evolution and genetic structure of such traits. For instance, the consistently small personality differences, even between the most genetically

distant ancestry groups, pose an intriguing question for future research—an area currently obscured by misunderstandings about the expected magnitude of genetic differences under neutrality.

References

1. a, b, c, d, e, f, g, h, i *Lala KN, Feldman MW (2024). "Genes, Culture, and Scientific Racism." Proc Natl Acad Sci U S A. **121**(48):e2322874121.*
2. a, b, c, d, e, f, g, h *Gusev A (2024). "No, Intelligence Is Not Like Height." The Infinitesimal. https://theinfinitesimal.substack.com/p/no-intelligence-is-not-like-height.*
3. a, b, c, d, e, f, g, h *Gusev A (2024). "A Molecular Genetics Perspective on the Heritability of Human Behavior and Group Differences." Gusev Lab. http://gusevlab.org/projects/hsq/hsq.pdf.*
4. a, b, c, d, e, f, g, h, i *Roseman CC, Bird KA (2023). "Between Group Heritability and the Status of Hereditarianism as an Evolutionary Science." BioRxiv. 2023-12.*
5. ^ *Bhatia G, Patterson N, Sankararaman S, Price AL (2013). "Estimating and Interpreting FST: The Impact of Rare Variants." Genome Res. **23**(9):1514–1521.*
6. ^ *Reich D (2018). Who We Are and How We Got Here: Ancient DNA and the New Science of the Human Past. New York: Pantheon Books.*
7. a, b *Leinonen T, McCairns RS, O'hara RB, Merilä J (2013). "Q ST–F ST Comparisons: Evolutionary and Ecological Insights From Genomic Heterogeneity." Nat Rev Genet. **14**(3):179–190.*
8. ^ *Whitlock MC, Guillaume F (2009). "Testing for Spatially Divergent Selection: Comparing Q ST to F ST." Genetics. **183**(3):1055–1063.*
9. a, b, c *Yair S, Coop G (2022). "Population Differentiation of Polygenic Score Predictions Under Stabilizing Selection." Philos Trans R Soc B. **377**(1852):20200416.*
10. a, b *Meirmans PG, Hedrick PW (2011). "Assessing Population Structure: FST and Related Measures." Mol Ecol Resour. **11**(1):5–18.*
11. ^ *Alcala N, Rosenberg NA (2017). "Mathematical Constraints on F ST: Biallelic Markers in Arbitrarily Many Populations." Genetics. **206**(3):1581–1600.*
12. ^ *Edelaar PIM, Burraco P, Gomez-Mestre IVAN (2011). "Comparisons Between QST and FST—How Wrong Have We Been?" Mol Ecol. **20**(23):4830–4839.*
13. a, b, c *Edge MD, Rosenberg NA (2015). "Implications of the Apportionment of Human Genetic Diversity for the Apportionment of Human Phenotypic Diversity." Stud Hist Philos Sci Part C: Stud Hist Philos Biol Biomed*

Sci. ***52****:32–45.*

14. a, b, c*Edge MD, Rosenberg NA (2015). "A General Model of the Relationship Between the Apportionment of Human Genetic Diversity and the Apportionment of Human Phenotypic Diversity." Hum Biol.* ***87****(4):313.*
15. ^*Wright S (1978). Evolution and Genetics of Populations, Volume 4. Chicago, IL: University of Chicago.*
16. ^*Cohen J (1988). Statistical Power Analysis for the Behavioral Sciences. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates.*
17. ^*1000 Genomes Project Consortium (2015). "A Global Reference for Human Genetic Variation." Nature.* ***526*** *(7571):68.*
18. ^*de Jong MJ, van Oosterhout C, Hoelzel AR, Janke A (2024). "Calculating and Interpreting F ST in the Genomics Era." bioRxiv. 2024-09.*
19. a, b*Whitlock MC (2008). "Evolutionary Inference From QST." Mol Ecol.* ***17****(8):1885–1896.*
20. a, b, c*Bird KA (2021). "No Support for the Hereditarian Hypothesis of the Black–White Achievement Gap Using Polygenic Scores and Tests for Divergent Selection." Am J Phys Anthropol.* ***175****(2):465–476.*
21. ^*Polderman TJ, Benyamin B, De Leeuw CA, Sullivan PF, Van Bochoven A, Visscher PM, Posthuma D (2015). "Meta-Analysis of the Heritability of Human Traits Based on Fifty Years of Twin Studies." Nat Genet.* ***47****(7):702–709.*
22. ^*Boas F (2023). The Mind of Primitive Man: A Course of Lectures Delivered Before the Lowell Institute, Boston, Mass., and the National University of Mexico, 1910-1911. Project Gutenberg. https://www.gutenberg.org/ebooks/71630.*
23. ^*Foldes HJ, Duehr EE, Ones DS (2008). "Group Differences in Personality: Meta-Analyses Comparing Five US Racial Groups." Pers Psychol.* ***61****(3):579–616.*

Declarations

Funding: No specific funding was received for this work.

Potential competing interests: No potential competing interests to declare.