Misinterpretations of admixture regression in racial hereditarian research

Kevin Bird[1] & Arslan Zaidi[2,3]

1. Department of Plant Sciences, University of California, Davis
2. Department of Genetics, Cell Biology, and Development, University of Minnesota,
3. Institute of Health Informatics, University of Minnesota

AABA 2024: Undermining the production of race science

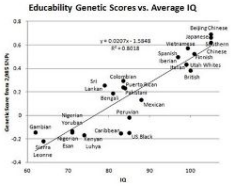
UCDAVIS

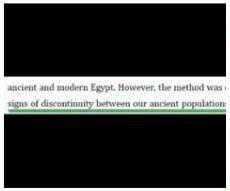
Racist Science in the 21st century

The New York Times

Why White Supremacists Are Chugging Milk (and Why Geneticists Are Alarmed)

Give this article 400



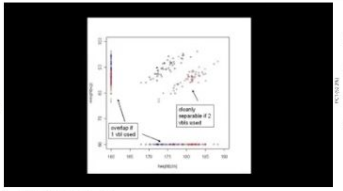


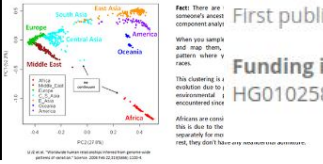
Myth #1: "Race has no biolc

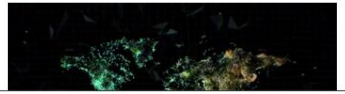
"Race is a social construct"
LOL
Does anyone believe this nonsense?
The truth: Race is biologically real.

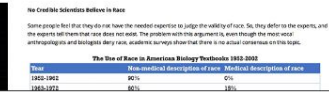
The Cultural Marxist War against Darwinism

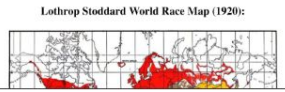
Creationists: evolution is a social construct, not biologically real.
Liberal Creationists: race is a social construct, not biologically real.
Charles Darwin: I'm not a creationist; I actually wrote: "There is, however, no doubt that various races, when carefully compared and measured, differ much from each other...."











AMERICAN JOURNAL OF BIOLOGICAL ANTHROPOLOGY

The Official Journal of the American Association of Biological Anthropologists

RESEARCH ARTICLE

How White nationalists mobilize genetics: From genetic ancestry and human biodiversity to counterscience and metapolitics

Aaron Panofsky, Kushan Dasgupta, Nicole Iturriaga

First published: 28 September 2020 | https://doi.org/10.1002/ajpa.24150 | Citations: 6

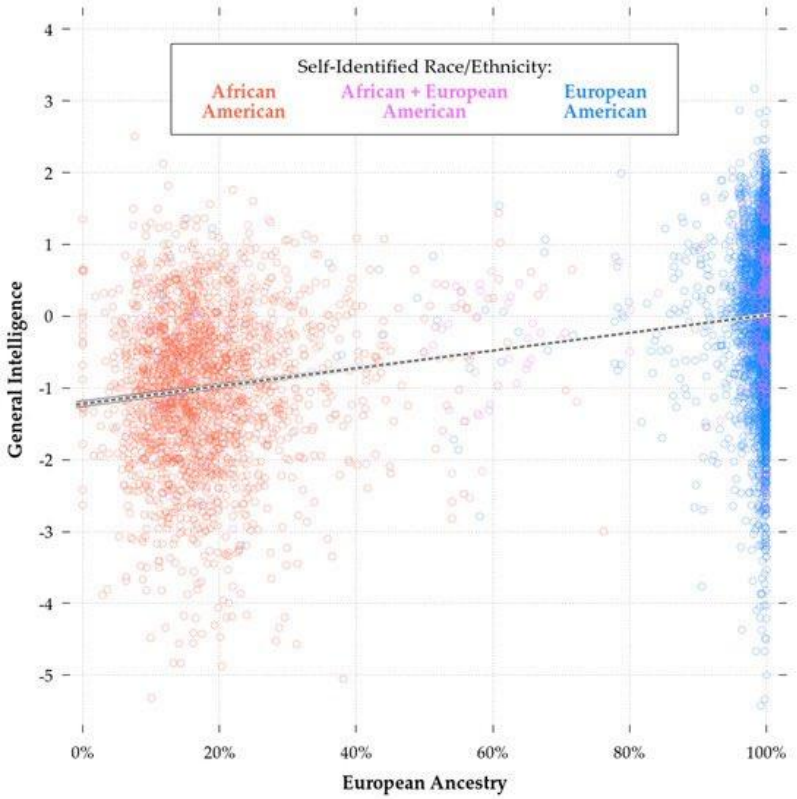
Funding information: National Human Genome Research Institute, Grant/Award Number: R21 HG010258; National Science Foundation, Grant/Award Number: 1322299

UCDAVIS

Racist Science in the 21st century

- **Reframing concepts and standards:** speculating about racial superiority and inferiority is not racism but exploring hypotheses about differences—or, data cannot be racist; repeating correlations is evidence of causation.
- **Challenging academic legitimacy and expertise:** attack academics as ideological to promulgate certainty about genetic causes of race and racial differences.
- **Building a cadre and a toolkit:** ideological foot soldiers armed with diverse race science materials from memes, to IQ classics, to cutting edge population genetics, from shitposting on Reddit, to commenting on New York Times articles, to publishing in scientific journals.

Admixture regressions as a burgeoning method in race science



Open Access Article

Global Ancestry and Cognitive Ability

by Jordan Lasker [1], Bryan J. Pesta [2,*], John G. R. Fuerst [3] and Emil O. W. Kirkegaard [4]

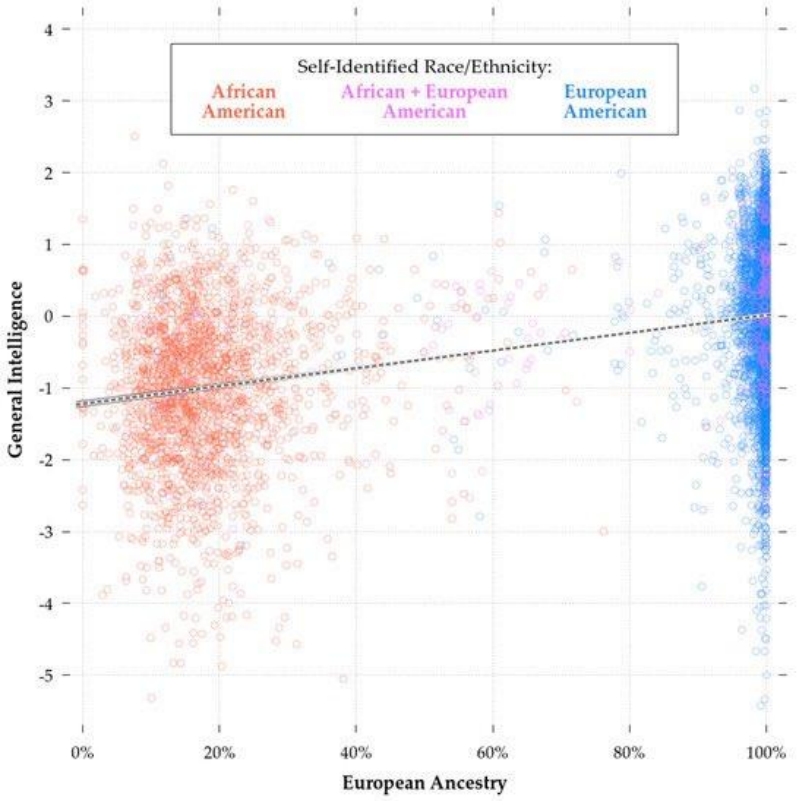
[1] Department of Economics, University of Minnesota Twin Cities, Minneapolis, MN 55455, USA
[2] Department of Management, Cleveland State University, Cleveland, OH 44115, USA
[3] Cleveland State University, Cleveland, OH 44115, USA
[4] Ulster Institute for Social Research, London NW26 9LQ, UK
* Author to whom correspondence should be addressed.

Psych **2019**, *1*(1), 431-459; https://doi.org/10.3390/psych1010034

Received: 8 June 2019 / Revised: 26 August 2019 / Accepted: 28 August 2019 / Published: 30 August 2019

UCDAVIS

Admixture regressions as a burgeoning method in race science

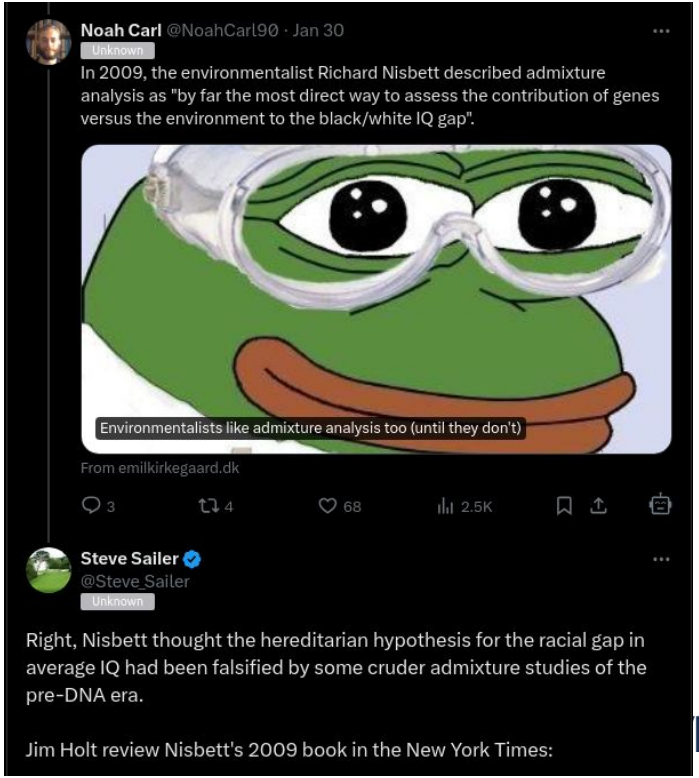


"landmark 2019 study… demonstrating a linkage between intelligence and racial admixture as measured by DNA."
- Steve Sailer, Unz Review Oct 13th, 2022

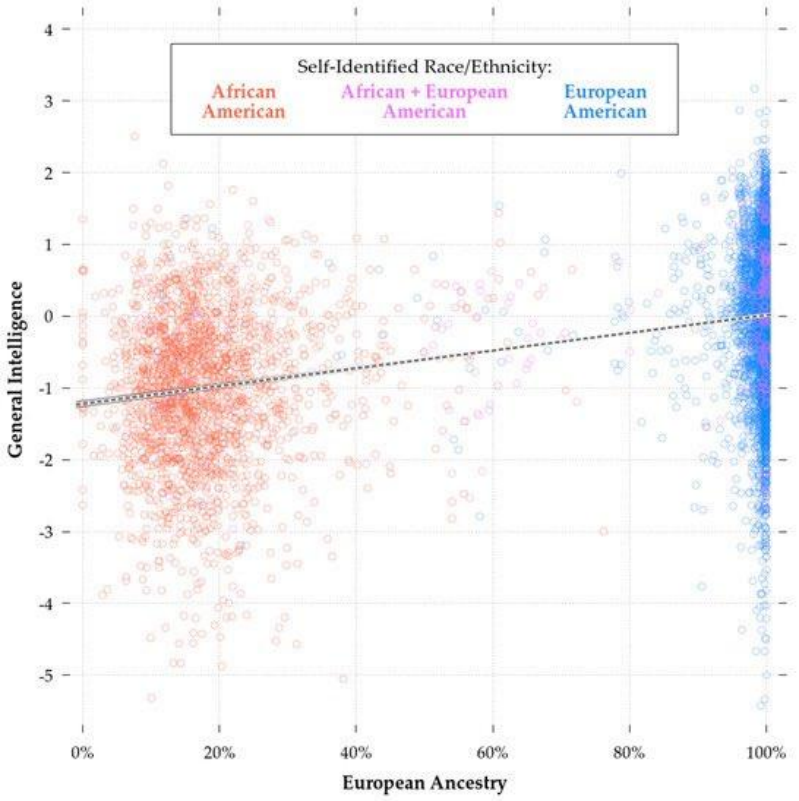
UCDAVIS

Admixture regressions as a burgeoning method in race science

Race science proponents have created their own whig history where everyone agreed admixture correlation was a dispositive method



Admixture regressions as a burgeoning method in race science



Obvious issue with confounding!

UCDAVIS

Problems with admixture regression: confounding

Schraiber and Edge (2024) recently formally derived a formula for expected genetic variation of a trait due to variation in global genetic ancestry

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

Expected trait value of individual with given ancestry proportion

mean genetic contribution to the trait in population 1

Ancestry fraction

Mean difference in genetic value between populations

Function of how environmental effect depends on an individual's global ancestry fraction

UCDAVIS

Problems with admixture regression: confounding

If environmental and genetic effects of global ancestry are confounded we can't recover δ_G and regressions slopes represent some combination of environmental & genetic effects

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

Expected trait value of individual with given ancestry proportion

mean genetic contribution to the trait in population 1

Ancestry fraction

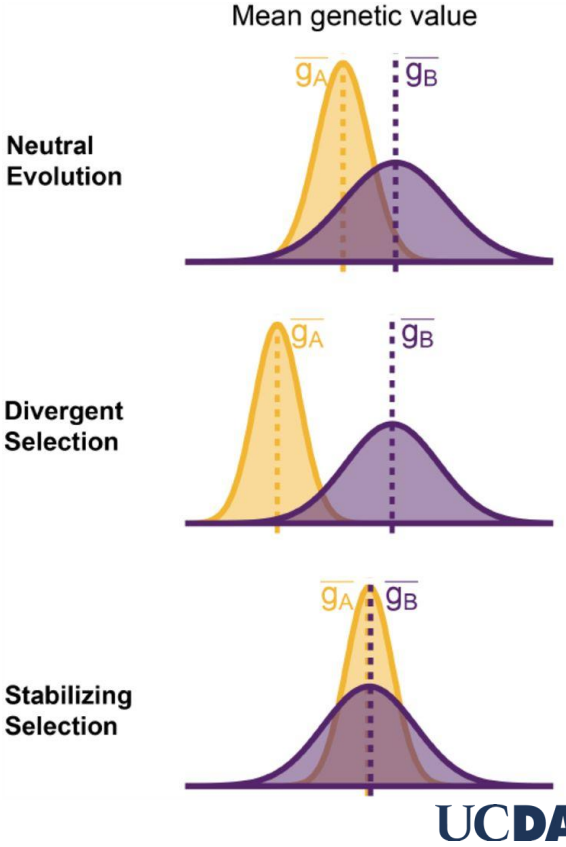
Mean difference in genetic value between populations

Function of how environmental effect depends on an individual's global ancestry fraction

UCDAVIS

Expectations of δ_G vary under different evolutionary scenarios

1. Neutral evolution: Differences in mean genetic value are expected to be small (≅ F_{st}), can favor either population
2. Directional selection: Expect larger differences in mean genetic value compared to drift
3. Stabilizing selection: Expect smaller difference in mean genetic value compared to drift



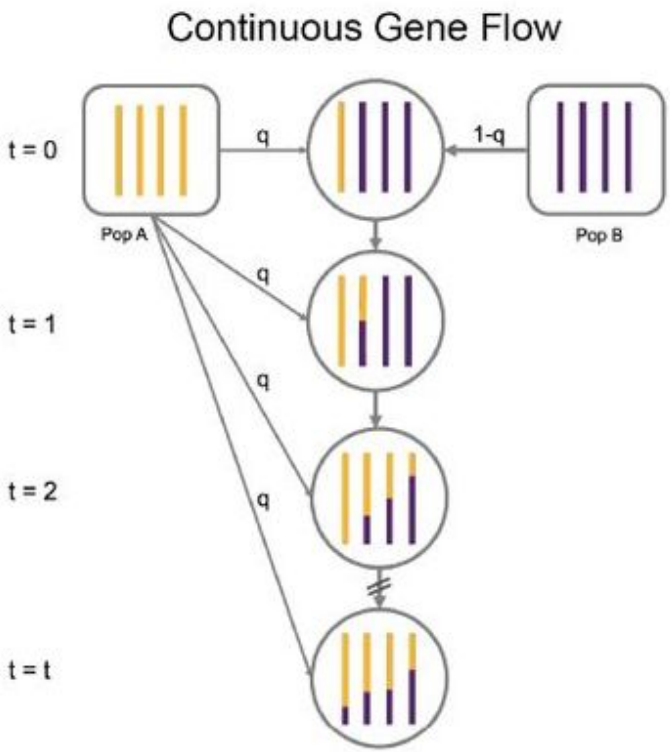
UCDAVIS

Explore limitations and behavior of trait-ancestry correlations

- What is ancestry-trait correlation under different evolutionary scenarios?
- How much can ancestry-trait correlations be affected by environmental confounding?
- How effective is using a proxy variable as a statistical control?

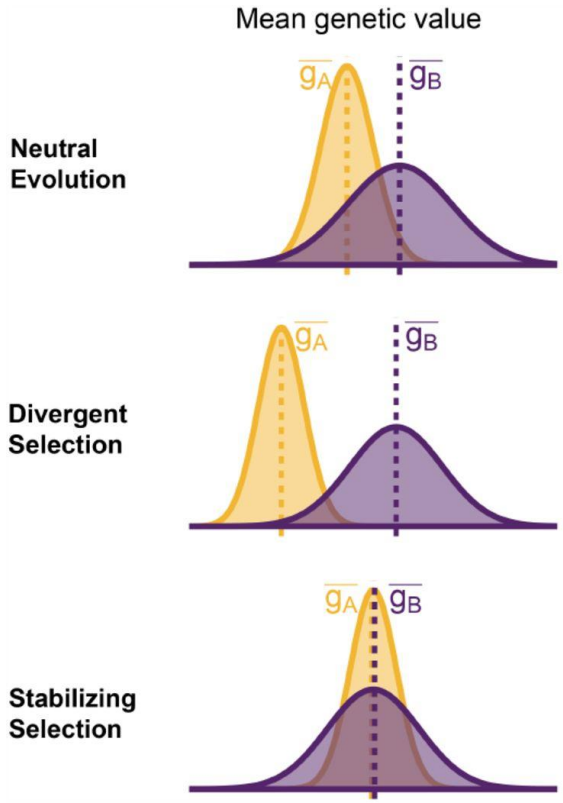
Methods:

- Simulate an admixed population with variable ancestry fractions
  - 10,000 individuals, across 20 generations. Replicated 10 times



Methods:

- Simulate an admixed population with variable ancestry fractions
  - 10,000 individuals, across 20 generations. Replicated 10 times
- Simulate genetic values (polygenic scores) for individuals under different evolutionary scenarios



UCDAVIS

Methods:

Simulate trait values (y) based on :

- g_i - Individual's genetic value
- θ_i - Individual's ancestry fraction
- a - slope between environment and ancestry
- e - random noise

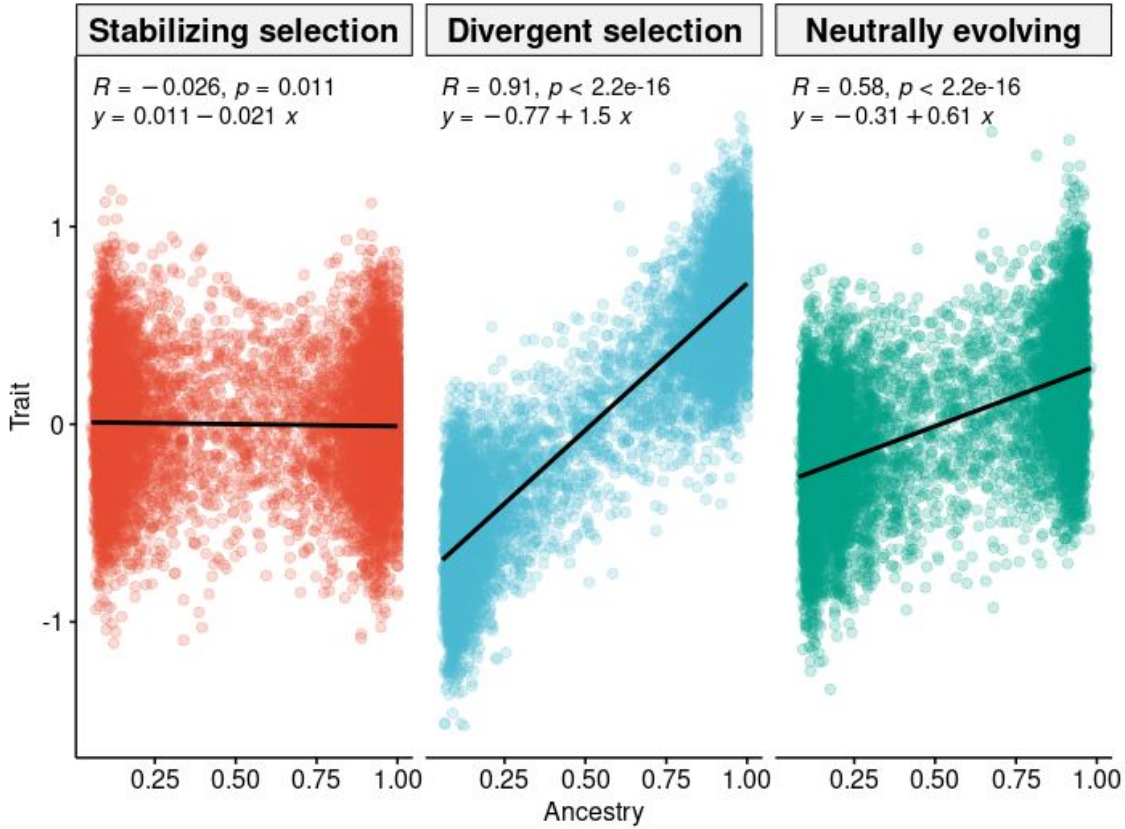
$$y_i = g_i + a*\theta_i + e$$

UCDAVIS

Questions:

- **What is ancestry-trait correlation under different evolutionary scenarios?**
- How much can ancestry-trait correlations be affected by environmental confounding?
- How effective is using a proxy variable as a statistical control?

Selection regime affects ancestry-trait correlations



*All results presented are for generation 20

UCDAVIS

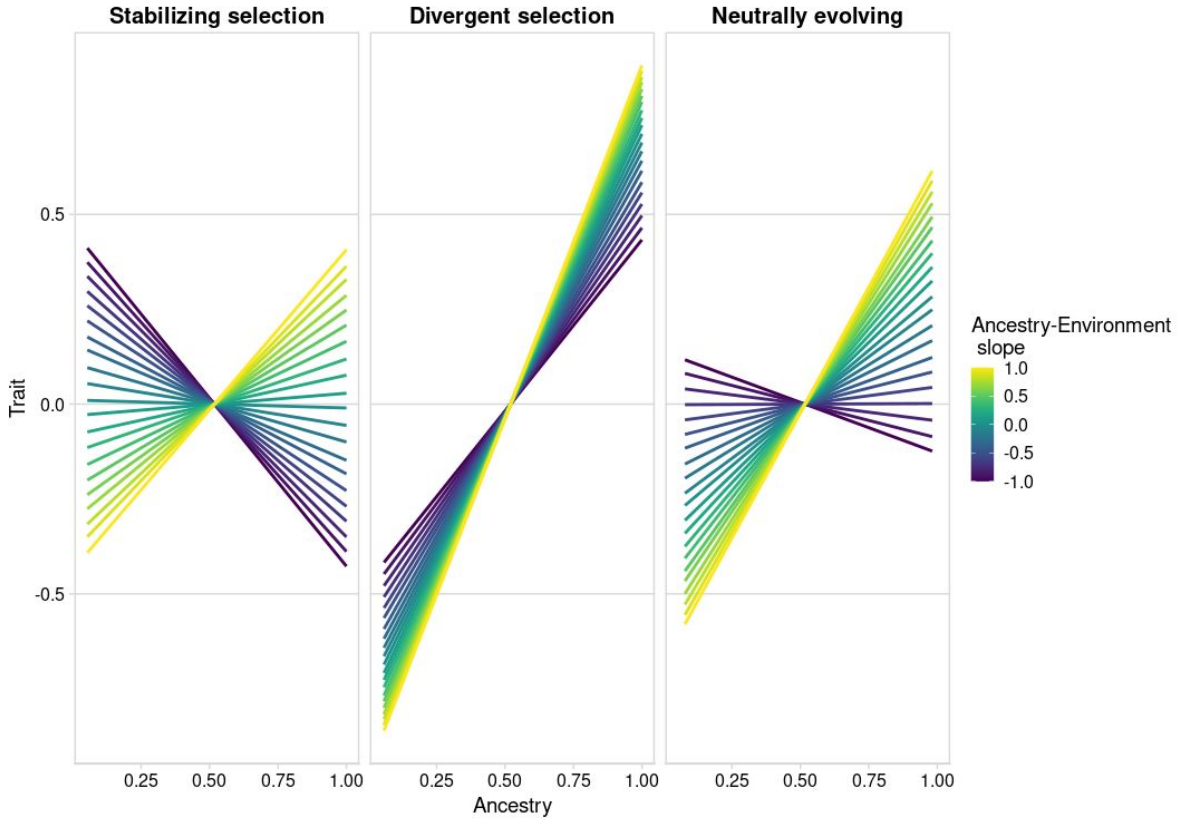
Questions:

- What is ancestry-trait correlation under different evolutionary scenarios?
- **How much can ancestry-trait correlations be affected by environmental confounding?**
- How effective is using a proxy variable as a statistical control?

Effect of environmental confounding?

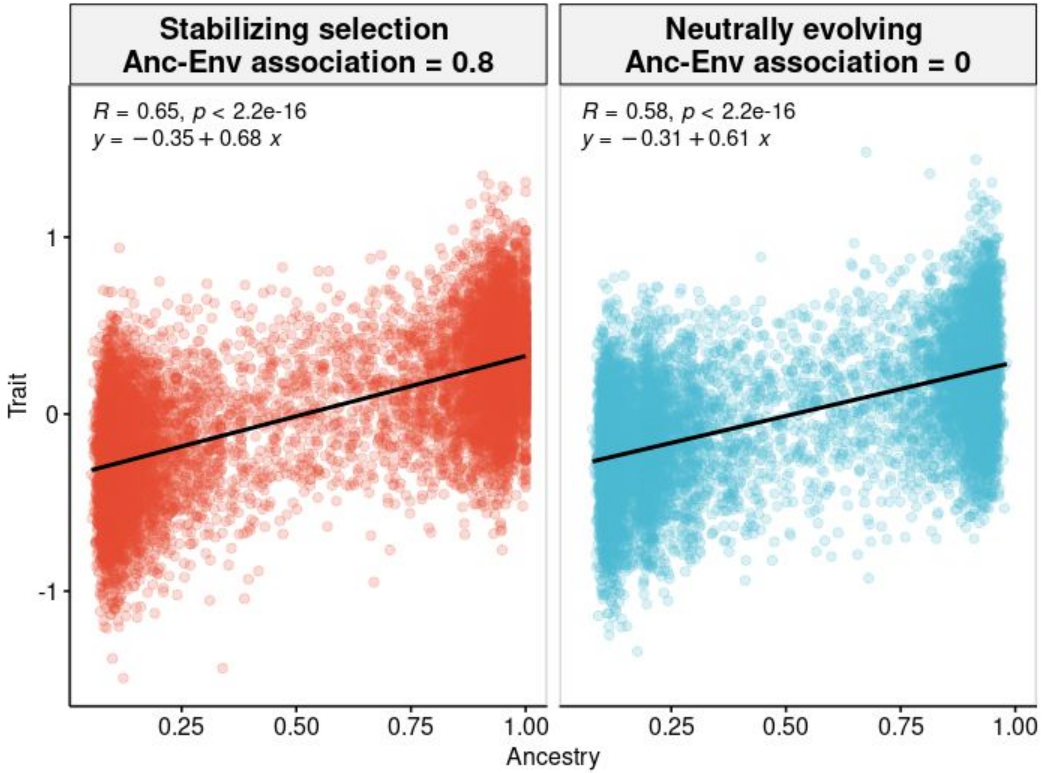
Environmental effects can change magnitude and direction of ancestry-trait correlation

Weaker effect under divergent selection



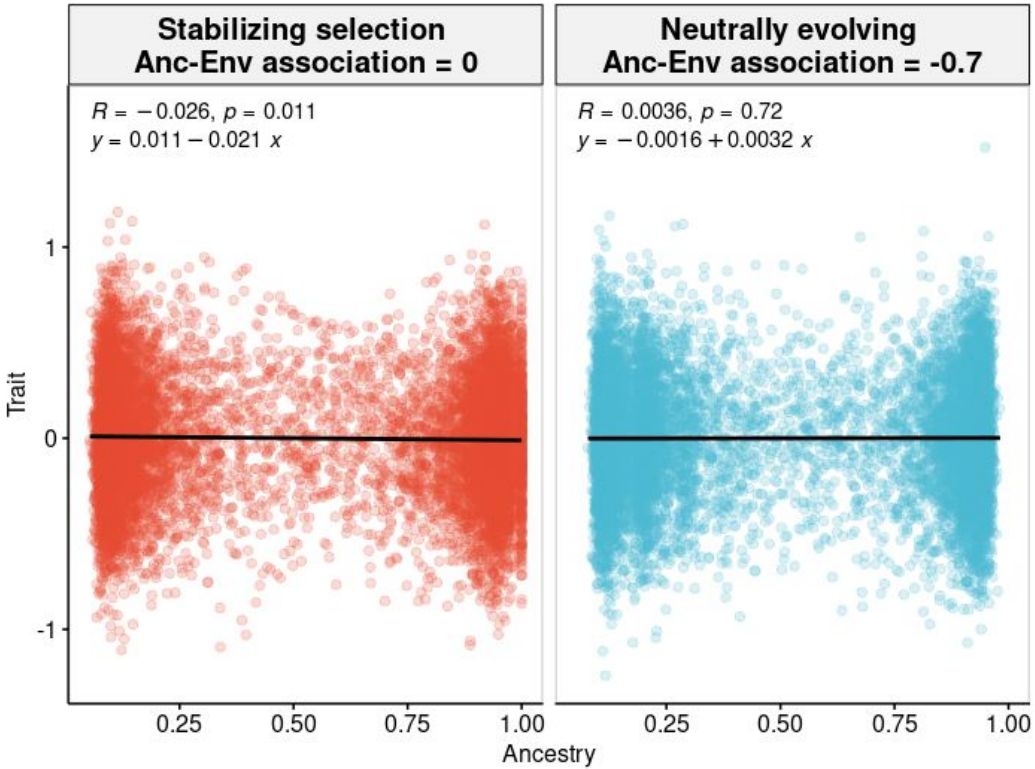
Multiple scenarios produce identical ancestry-trait correlations

Ancestry-trait correlation that is entirely driven by environmental effects (left) is indistinguishable from an ancestry-trait correlation driven by neutral genetic differences and no environmental effects (right)



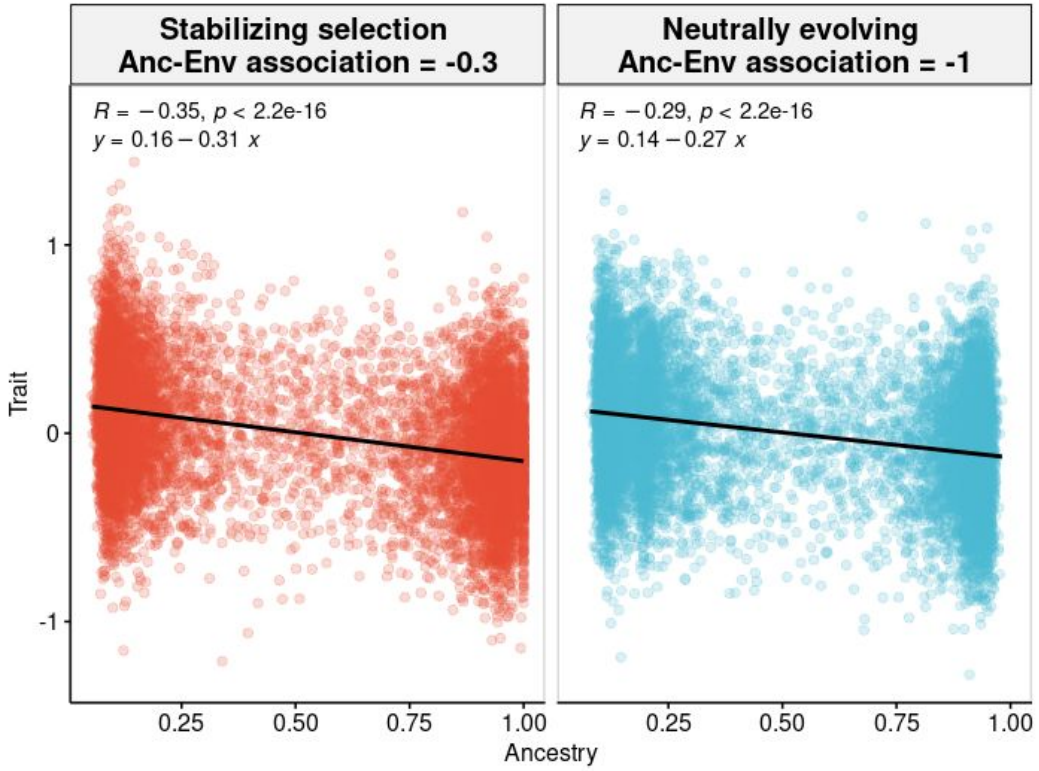
Multiple scenarios produce identical ancestry-trait correlations

A true Ancestry-trait correlation of zero (left) is indistinguishable from a case where a true positive ancestry-trait correlation exists but is counteracted by a negative environmental effect (right)



Multiple scenarios produce identical ancestry-trait correlations

Environmental effects of different strength can create nearly identical negative slopes in cases where no correlation exists (left) or when the correlation should be positive (right)



Questions:

- What is ancestry-trait correlation under different evolutionary scenarios?
- How much can ancestry-trait correlations be affected by environmental confounding?
- **How effective is using a proxy variable as a statistical control?**

How effective is using a proxy variable as a statistical control?

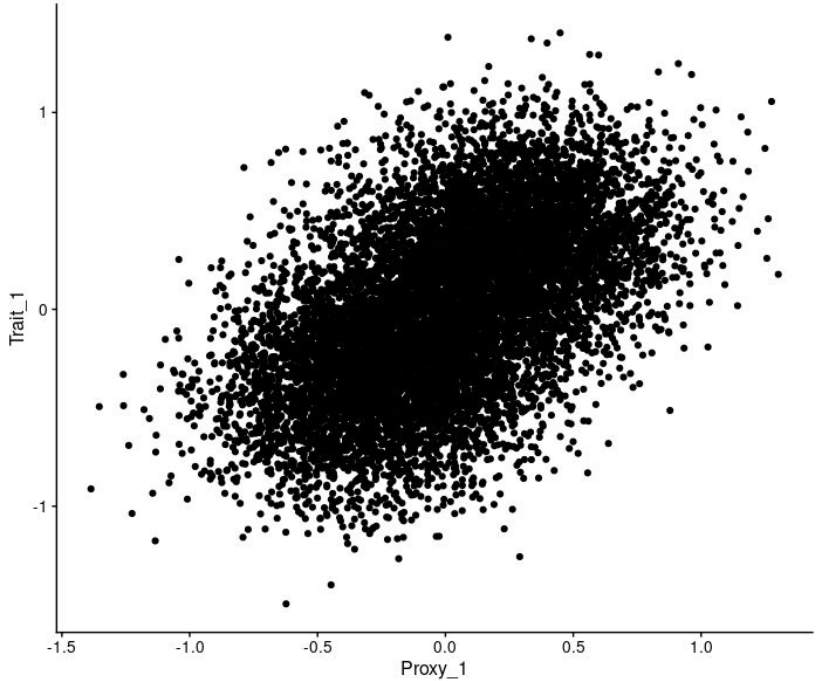
- Racial hereditarian researchers often include variables related to socioeconomic status (parental education, self-reported household income, etc)

- Claim the failure of such controls to fully attenuate ancestry-trait associations is evidence there is a true genetic effect

How effective is using a proxy variable as a statistical control?

Simulate a noisy proxy trait that is partially correlated (r=0.5) with actual environmental effects.

Include proxy trait as covariate in scenario where a strong ancestry-trait correlation is completely spurious.



UCDAVIS

Statistical controls only partially attenuate spurious correlations

Regression model without control

```
lm(formula = Trait_1 ~ Ancestry, data = .)

Residuals:
     Min       1Q   Median       3Q      Max
-1.16284 -0.20275 -0.00116  0.20038  1.10251

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.436678   0.005032  -86.78   <2e-16 ***
Ancestry     0.843741   0.007792  108.28   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3009 on 9998 degrees of freedom
Multiple R-squared:  0.5398,    Adjusted R-squared:  0.5397
F-statistic: 1.173e+04 on 1 and 9998 DF,  p-value: < 2.2e-16
```

Statistical controls only partially attenuate spurious correlations

Regression model without control

```
lm(formula = Trait_1 ~ Ancestry, data = .)

Residuals:
     Min       1Q   Median       3Q      Max
-1.16284 -0.20275 -0.00116  0.20038  1.10251

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.436678   0.005032  -86.78   <2e-16 ***
Ancestry     0.843741   0.007792  108.28   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3009 on 9998 degrees of freedom
Multiple R-squared:  0.5398,    Adjusted R-squared:  0.5397
F-statistic: 1.173e+04 on 1 and 9998 DF,  p-value: < 2.2e-16
```

Regression model with control

```
lm(formula = Trait_1 ~ Ancestry + Proxy_1, data = .)

Residuals:
     Min       1Q   Median       3Q      Max
-1.11126 -0.20082 -0.00159  0.20089  1.07666

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.391343   0.005917  -66.14   <2e-16 ***
Ancestry     0.756145   0.009876   76.56   <2e-16 ***
Proxy_1      0.138260   0.009733   14.21   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.298 on 9997 degrees of freedom
Multiple R-squared:  0.5489,    Adjusted R-squared:  0.5488
F-statistic:  6081 on 2 and 9997 DF,  p-value: < 2.2e-16
```

UCDAVIS

Conclusions

- Ancestry-environment correlations are uninterpretable without additional knowledge of evolutionary dynamics and environmental effects.
- Environmental variables that only partially capture true environmental effects will be unable to fully attenuate even completely spurious ancestry-trait correlations
    - Showing that ancestry-trait associations survive controls for SES can’t serve as evidence that a true genetic effect exists.