

The genetic ancestry of African, Latino, and European Americans across the United States.

Katarzyna Bryc^{1,2}, Eric Y. Durand², David Reich^{1,3,4}, and Joanna L. Mountain²

1. Harvard Medical School, Department of Genetics, Boston, MA, USA 3. Broad Institute of MIT and Harvard, Cambridge, MA, USA
2. 23andMe, Inc., Mountain View, CA, USA 4. Howard Hughes Medical Institute, Boston, MA, USA

Introduction

- Over the past 500 years, North America has been the site of ongoing mixing of Native Americans, European settlers, and Africans brought largely by the Trans-Atlantic slave trade, shaping the early history of what became the United States.
- While US government census surveys have established fine-scale self-described ethnicity at the state and county level, the relationship between genetic ancestry and self-reported ancestry for each region has not been deeply characterized.
- Understanding genetic ancestry of individuals from a self-reported population, and differences in ancestry patterns among regions, can inform medical studies and personalized medical treatment.

Acknowledgments

We thank 23andMe customers who consented to participate in research for enabling this study. We also thank employees of 23andMe who contributed to the development of the infrastructure that made this research possible.

Methods

- All participants were drawn from the customer base of 23andMe, Inc., a consumer personal genomics company. Customers were genotyped on the Illumina HumanOmniExpress+ platform.
- 23andMe customers were invited to fill out web-based questionnaires, including questions on ancestry and ethnicity, on state of birth, and current zip code of residence.
- They were also invited to allow their genetic data and survey responses to be used for research. Only data of customers who signed IRB-approved consent documents were included in our study.
- We generated cohorts of self-reported European American, African American, and Latino individuals from self-reported ethnicity and identity.
- We obtain ancestry estimates from genotype data using a Support Vector Machine-based algorithm that infers population ancestry using reference panels to investigate patterns of Native American, African, and European ancestry.

Discussion

- In spite of present-day high mobility of individuals, the genetic ancestry of present-day individuals recapitulates historical migration events, known settlement patterns, and admixture processes. Our finding that African Americans from Oklahoma carry the highest levels of Native American ancestry traces the Trail of Tears migration; European Americans from the Dakotas carry high levels of Scandinavian ancestry following immigration into the region.
- Our results provide empirical support that, over recent centuries, many individuals with partial African and Native American ancestry have "passed" into the white community^{1,2}.
- Our results undermine the use of cultural labels that group individuals into discrete non-overlapping bins in biomedical contexts.

Genetic ancestry of 5,269 self-reported African Americans from across the US

State and regional ancestry differences

In addition to individual variability in ancestry proportions, we find differences in mean ancestry proportions of African Americans across states and regions of the US.

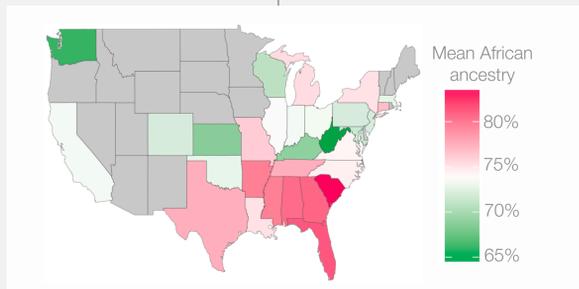


Figure 1. The distribution of mean African ancestry in African Americans across the US. States with fewer than 10 individuals are excluded in gray.

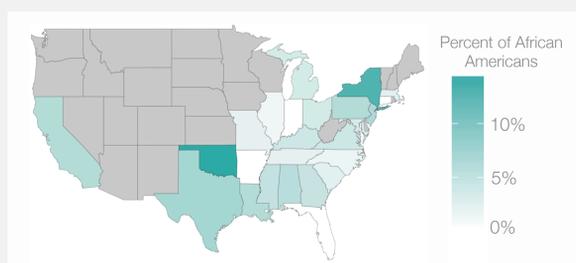


Figure 2. Proportion of African Americans that carry Native American ancestry by state.

As defined as carrying 2% or more Native American ancestry. States with fewer than 20 individuals are excluded in gray.

Figure 3. Estimated ancestral contributions of males and females from different populations to African Americans. Inferred from relative rates of genome-wide and X chromosome ancestry proportions. Scaled to 100 individuals.



Native American ancestry in African Americans

1 in every 20 African Americans carries Native American ancestry, a higher rate than we detected in self-reported European Americans. Over 14% of African Americans from Oklahoma carry at least 2% Native American ancestry, which was the site of contact following the Trail of Tears migration in the 1830's.^{3,4}

Sex bias in ancestry contributions

We find that our estimates of sex bias in ancestry contributions in African Americans support over three times as many male European ancestors as female European ancestors.

Genetic landscape of 8,663 self-reported Latinos

Differences in proportions of Native American ancestry

On average, we estimate that Latinos in the US carry 18.0% Native American ancestry, 65.1% European ancestry, and 6.2% African ancestry. We find the highest levels of estimated Native American ancestry in self-reported Latinos from states in the Southwest, especially those close to Mexico.

Distinct European subpopulation ancestry of Latinos

We estimate that Iberian ancestry composes as much as a third of the European ancestry in Latinos in Florida, New Mexico, and other parts of the Southwest, likely either reflecting early Spanish influence and rule in these regions, or recent immigration from Latin America bearing higher levels of Spanish and Portuguese ancestry.

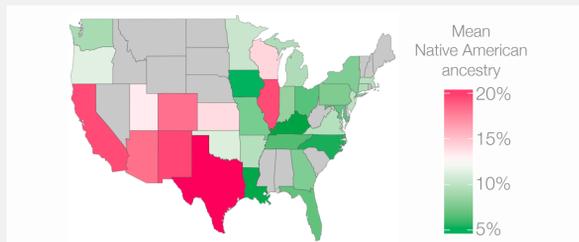


Figure 4. Mean levels of Native American ancestry in Latinos across the US. Highest levels of Native American ancestry are found in Latinos from the Southwest. States with fewer than 10 individuals are excluded in gray.

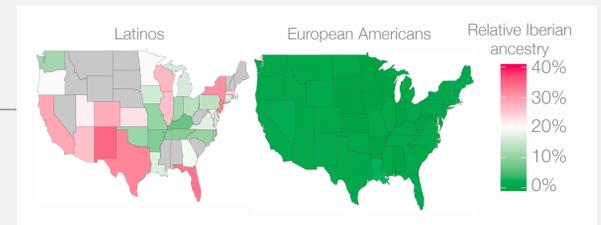


Figure 5. Latinos show proportionately higher levels of Iberian ancestry than European Americans. Out of the total amount of European ancestry per state, the relative proportion of Iberian ancestry in Latinos (left) versus European Americans (right).

Genetic ancestry of 148,789 European Americans

European Americans frequently carry non-European ancestry

We find African ancestry in European Americans:

- 3.5% of European Americans are estimated to carry at least 1% African ancestry
- European Americans with African ancestry comprise as much as 12% of European Americans from Louisiana and South Carolina, and about 1 in 10 individuals in other parts of the South.
- We estimate that 6 million Americans in the US, who self-identify as European, may carry African ancestry.

Differences in composition of European ancestries

Subpopulation ancestry proportions reflect historical migration patterns and coincide with self-identified ancestry.

- 2010 US Census⁵ Norwegian identities co-locate with Scandinavian genetic ancestry in European Americans.
- Genetic British/Irish ancestry in European Americans coincides with self-reported "American" US Census ancestry.

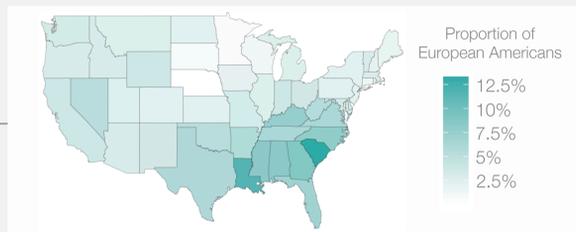
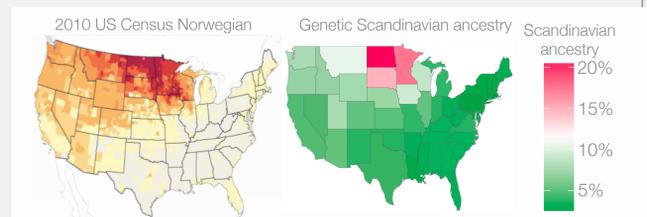


Figure 6. Proportion of European Americans that carry African ancestry. Defined as proportion of European Americans with 1% or more African ancestry.

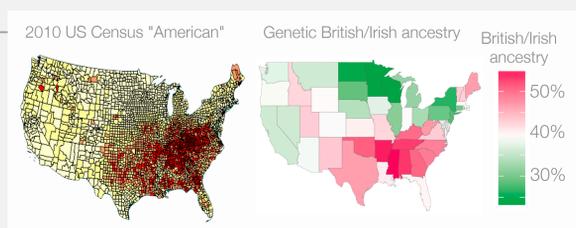


Figure 7. Co-localization of genetic ancestry in European Americans and 2010 US Census self-reported ancestry data.

Self-identity versus genetic ancestry

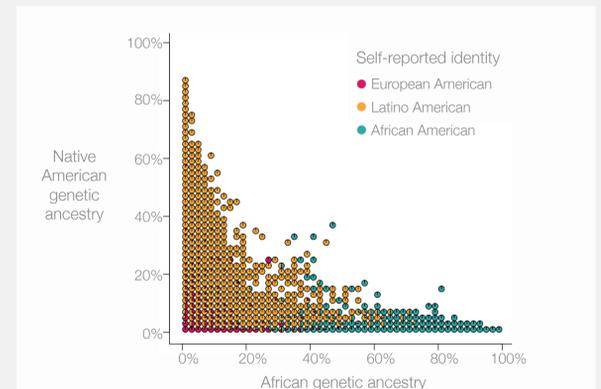


Figure 8. The relationship between genetic ancestry and self-identified ethnicity. The proportion of individuals who self-identify as African American, Latino, and European American for each bin of 2% African and Native American ancestry is shown as a pie chart.

References

- Burma, J. H. (1946). The measurement of negro "passing". *American Journal of Sociology* pp. 18-22.
- Broyard, B. (2007). *One Drop: My Father's Hidden Life—A Story of Race and Family Secrets*. (Hachette Digital, Inc.)
- Wells, S. J. (2004). *After Removal: The Choctaw in Mississippi*. (Univ. Press of Mississippi).
- Green, L. (1978). Choctaw removal was really a "trail of tears". Bishnik, mboucher, University of Minnesota.
- U.S. Census Bureau (2011). 2010 census. U.S. Department of Commerce.