A web-based initiative to accelerate research on African ancestry

Meghan E. Moreno, Kimberly F. McManus, Katarzyna Bryc, Patricia K. Penton, Rebecca Rubenstein, Lisa Vermillion, the 23andMe Research Team, Joanna L. Mountain

23andMe, Inc., Mountain View, CA, USA

Background

The paucity of written records regarding the trans-Atlantic slave trade has resulted in uncertainty for many African Americans regarding their ancestors' origins. The geographic origins or ethnic affiliations of slaves taken from Africa was rarely recorded. More recently, genetic research has excluded participants with African ancestry, which has further limited the understanding of African genetic diversity.

Project structure and goals

In October 2016, 23andMe launched the African Genetics Project, a web-based genetic research study of individuals in the United States with recent African ancestry (Figure 1). The goals were:

1. To improve 23andMe's ability to provide detailed ancestry results for customers with some African origins.
2. To contribute to the global understanding of how people historically migrated throughout Africa and from Africa to the rest of the world.

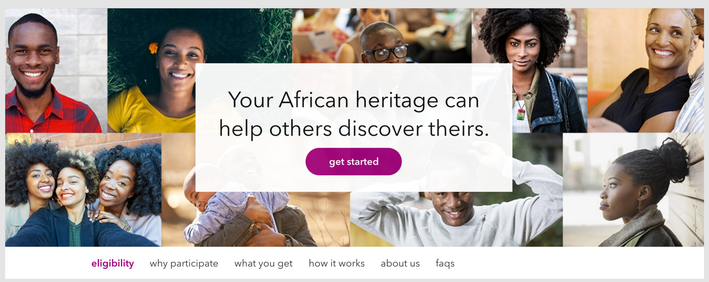

Figure 1. Screenshot of the study landing page that invited readers to learn about and enroll in the African Genetics Project.

Enrollment was open to any adult in the United States whose four grandparents were born in a single high priority country (see "Enrollment and data collection") or were affiliated with a single ethnic group in those countries. Once enrolled, participants were asked to provide a saliva sample for genetic analysis and complete an online survey about their family's ancestry and cultural/ethnic group affiliations.

We selected as high priority those African countries for which 23andMe had data from very few representatives. We focused mostly on West African countries because the majority of slaves brought from Africa to the Americas were taken from this region. We also included Ethiopia, Somalia, and Sudan to improve our reference populations for individuals with more recent African ancestry.

Recruitment

Study recruitment was conducted primarily online. This included organic posts on 23andMe's social media channels, social media ads, a 23andMe blog post, and an article in the online African-American culture magazine, The Root. We also visited local universities to give talks at African Studies centers. Highly targeted Facebook ads had the strongest direct effect on enrollment rates.

Once enrolled, some participants promoted the study by posting about it on blogs, social media, and YouTube. Amena, a Sudanese vlogger, posted a video that mentions this study on her Heart of Africa YouTube channel. To date, this video has been viewed over 81,400 times. Its posting coincided with a spike in enrollments from Sudanese participants that remained consistent through the rest of the study (Figure 2).

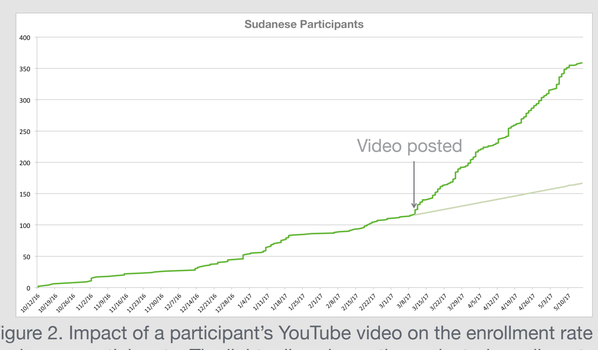

Figure 2. Impact of a participant's YouTube video on the enrollment rate of Sudanese participants. The lighter line shows the projected enrollment rate without this video.

Enrollment and data collection

Participants provided informed consent and participated in the research online, under a protocol approved by the external AAHRPP-accredited IRB, Ethical & Independent Review Services. To screen for eligibility, we asked participants to answer screening questions about their grandparents' birthplaces and any ethnic group affiliations.

Once enrolled, consented participants could choose to take additional surveys on the 23andMe website about a range of topics such as their health and wellness, medications, lifestyle, and environment. Study participants took an average of 5 surveys. Those who completed the study-specific ancestry survey took an average of 7 surveys.

Study participants were genotyped on a custom Illumina Human OmniExpress-24 genotyping chip.

High priority countries	Participants per country
Sudan	360
Somalia	247
Ethiopia	215
Ghana	97
Cameroon	65
Liberia	64
Côte d'Ivoire	30
Senegal	25
Guinea	16
Angola	8
Togo	6
Benin, Burkina Faso, Gabon, Guinea-Bissau, Republic of the Congo	≤5 each

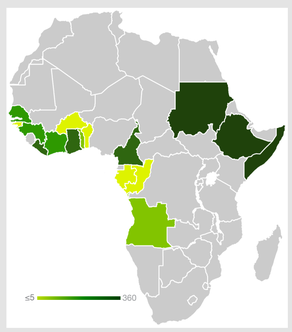

Figure 3. The majority of enrolled participants reported having East African ancestry.

Enrollment breakdown:	
Total Consented Participants	1150
Percent surveyed	67%
Percent genotyped	72%
Percent surveyed and genotyped	59%
Total # of research surveys taken by participants	5,827

Demographics of consented participants:	
Median age	26
Female (%)	65%

Ethnic groups with at least five enrolled participants:
Beja
Danagla
Dinka
Dongola
Ja'alin
Mahas
Mawaleed
North Sudanese
Nuba
Nubian
Shaigiya
Sudanese
Zaghawa

Preliminary data analysis

Over 1100 participants enrolled in only 8 months. We believe the quick enrollment was due to the study's accessibility (it took place entirely online) and because participants were motivated by receiving access to the 23andMe service at no cost. However, there was a risk of enrollment by ineligible individuals who did not understand the eligibility criteria, or who intentionally misrepresented their ancestry to receive free access to 23andMe's service.

We mitigated this risk several ways, primarily through the design of the enrollment screening questions. We asked, separately, where each grandparent was born, and participants chose from a drop-down menu with every country in the world listed. This method appeared to be effective, as we received over 2,000 enrollment attempts from people who were ineligible because they said that one or more of their grandparents were not born in one of the high-priority countries.

Genetic estimates of participants' African ancestry	
Percent of participants with **>90%** African ancestry	90%
Percent of participants with **>80%** African ancestry	96%
Percent of participants with **<75%** African ancestry	3%

The ancestry estimates above show that the rate of fraud was likely low, as very few enrolled participants had less than 75% African ancestry.

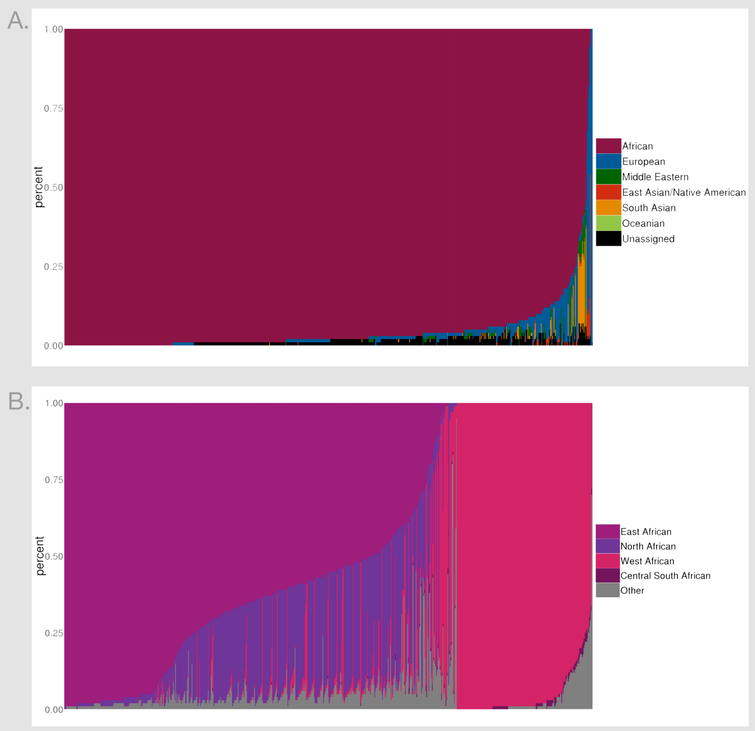

Figure 4. Overview of participants' ancestry. To preserve anonymity, data from 10% of participants were randomly excluded from these plots. (A) Estimates of participants' global ancestry. (B) Estimates of participants' ancestry from East, North, West, and Central/South African regions.

We applied 23andMe's genetic ancestry inference algorithm to data from this cohort. The analysis indicated that study participants primarily have African ancestry, with little ancestry from other continents.

The majority of participants had self-reported ancestry from Ethiopia, Somalia, and Sudan. Our algorithm suggests that these participants have a combination of East African and North African/Middle Eastern ancestry. The plots in Figure 4 raise questions regarding how self-reported ancestry and identity compare with genetically inferred ancestry. DNA from many of the individuals reporting ancestry only from Eastern Africa suggests ancestry also from Northern Africa.

Conclusions

In a relatively short period of time, we enrolled over 1100 participants with recent African ancestry. Preliminary analyses show that the self-reported ancestry information provided at enrollment matches the genetic ancestry results for nearly all participants.

We will continue analyzing the study data with the goals of (1) learning about migrations within Africa and between Africa and the Americas and (2) updating 23andMe's ancestry inference algorithms. In particular we aim to provide more detailed ancestry results for customers with some African ancestry.

The success of this study puts 23andMe in a position to study African populations that are currently underrepresented in genetic research. In addition the study serves as a pilot for recruitment of other underrepresented global populations.

We are committed to mitigating the current disparities in genetics research by extending such research well beyond the current practice of studying peoples of European origin. As such, we plan to continue this research with a new study that includes additional populations from Africa, Asia, and the Pacific.

Acknowledgments

We thank those 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 initiative possible.