RESEARCH ARTICLE SUMMARY

HUMAN GENETICS

The genetic legacy of African Americans from Catoctin Furnace

Éadaoin Harney*, Steven Micheletti, Karin S. Bruwelheide, William A. Freyman, Katarzyna Bryc, Ali Akbari, Ethan Jewett, Elizabeth Comer, Henry Louis Gates Jr., Linda Heywood, John Thornton, Roslyn Curry, Samantha Ancona Esselmann, Kathryn G. Barca, Jakob Sedig, Kendra Sirak, Iñigo Olalde, Nicole Adamski, Rebecca Bernardos, Nasreen Broomandkhoshbacht, Matthew Ferry, Lijun Qiu, Kristin Stewardson, J. Noah Workman, Fatma Zalzala, Shop Mallick, Adam Micco, Matthew Mah, Zhao Zhang, 23andMe Research Team, Nadin Rohland, Joanna L. Mountain*†, Douglas W. Owsley*†, David Reich*†

INTRODUCTION: Genetic analysis of historical individuals has the potential to help restore knowledge of people whose stories were omitted from written records. During the late 18th and early 19th centuries, Catoctin Furnace in Maryland relied on a workforce of enslaved individuals to operate the iron furnace and carry out domestic and agricultural tasks. Despite the role that Catoctin Furnace played in early US history (including supplying munitions during the Revolutionary War), relatively little is known about the African Americans who labored there or their descendants compared with the furnace's later, predominantly white workforce.

RATIONALE: We produced genome-wide data for 27 individuals buried in the Catoctin Furnace African American Cemetery and compared them to ~9.3 million consenting research participants genotyped by 23andMe, Inc., to address the following questions: (i) How were the Catoctin individuals related to each other? (ii) What were the sources of their African and European ancestry? (iii) Where in the US do their genetic relatives live today, including t… direct descendants? (iv) What can their genc… reveal about their health?

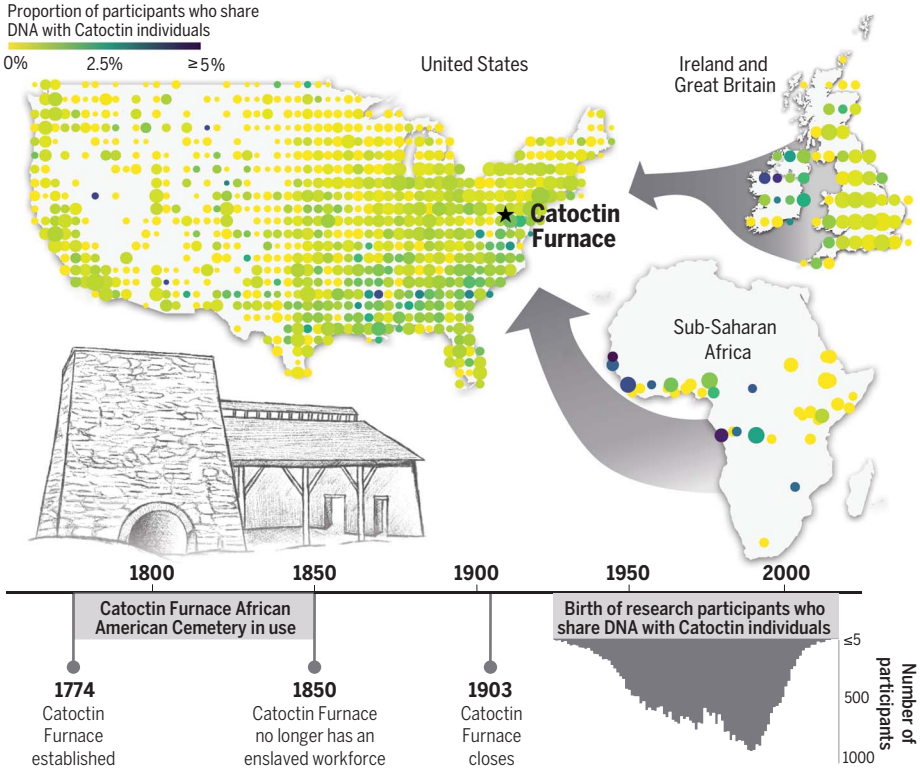
RESULTS: We identified five genetic families, consisting of biological mothers, children, and siblings, among the Catoctin individuals. In most cases, biological family members were buried in close proximity.

All but one of the Catoctin individuals had primarily African ancestry, with variable amounts of European ancestry. To learn more about their ancestry, we developed an approach to detect identical-by-descent segments of the genome shared between the Catoctin individuals and 23andMe research participants. Identical-by-descent segments of DNA are shared by two or more people because they have been inherited from a recent common ancestor. We identified 41,799 close and distant relatives of the Catoctin individuals among 23andMe research participants. Within Africa, we found the highest rates of genetic sharing between Catoctin individuals and research participants who self-identified as belonging to the Wolof or Kongo ethnolinguistic groups. Within Europe, we observed the highest rates of genetic sharing with research participants that have ties to Great Britain and Ireland.

Within the US, participants from the South showed elevated rates of sharing, largely reflecting distant connections to 23andMe research participants with sub-Saharan African ancestry (possibly tracing back to shared common ancestors in Africa). When we considered genetic relatives who share the most identical DNA with the Catoctin individuals, we observed the highest rates of sharing in Maryland, suggesting that at least some descendants stayed in the region after the furnace's transition away from enslaved and paid African American labor.

Finally, we found that some of the Catoctin individuals carried risk factors for sickle cell anemia and glucose-6-phosphate dehydrogenase deficiency, genetic diseases that are common in African Americans today.

CONCLUSION: These results demonstrate the power of joint analysis of DNA from historical individuals and the extremely large datasets generated through direct-to-consumer ancestry testing, and they serve as a model for obtaining direct insights into the genome-wide genetic ancestry of enslaved people in the historical US. ■



Genetic connections to individuals from Catoctin Furnace African American Cemetery in Maryland. A timeline highlighting major events in the history of Catoctin Furnace and the birth years of research participants in the 23andMe cohort, presented alongside maps showing the proportion of 23andMe research participants associated with geographic coordinates in the US, sub-Saharan Africa, and Europe who share genetic connections to the Catoctin individuals.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: eadaoinh@23andme.com (É.H.); jlmountain@gmail.com (J.L.M.); owsleyd@si.edu (D.W.O.); reich@genetics.med.harvard.edu (D.R.)
†These authors contributed equally to this work.
Cite this article as É. Harney *et al.*, *Science* **381**, eade4995 (2023). DOI: 10.1126/science.ade4995

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.ade4995

RESEARCH ARTICLE

HUMAN GENETICS

The genetic legacy of African Americans from Catoctin Furnace

Éadaoin Harney[1,2]*, Steven Micheletti[1], Karin S. Bruwelheide[3], William A. Freyman[1], Katarzyna Bryc[1], Ali Akbari[2,4], Ethan Jewett[1], Elizabeth Comer[5], Henry Louis Gates Jr.[6], Linda Heywood[7], John Thornton[7], Roslyn Curry[1,2], Samantha Ancona Esselmann[1], Kathryn G. Barca[3], Jakob Sedig[2,4], Kendra Sirak[2,4], Iñigo Olalde[2,8,9], Nicole Adamski[4,10], Rebecca Bernardos[4,10], Nasreen Broomandkhoshbacht[4,10], Matthew Ferry[4,10], Lijun Qiu[4,10], Kristin Stewardson[4,10], J. Noah Workman[4,10], Fatma Zalzala[4,10], Shop Mallick[4,10,11], Adam Micco[4,11], Matthew Mah[4,10,11], Zhao Zhang[4], 23andMe Research Team†, Nadin Rohland[4], Joanna L. Mountain[1]*‡, Douglas W. Owsley[3]*‡, David Reich[2,4,10,11]*‡

Few African Americans have been able to trace family lineages back to ancestors who died before the 1870 United States Census, the first in which all Black people were listed by name. We analyzed 27 individuals from Maryland's Catoctin Furnace African American Cemetery (1774–1850), identifying 41,799 genetic relatives among consenting research participants in 23andMe, Inc.'s genetic database. One of the highest concentrations of close relatives is in Maryland, suggesting that descendants of the Catoctin individuals remain in the area. We find that many of the Catoctin individuals derived African ancestry from the Wolof or Kongo groups and European ancestry from Great Britain and Ireland. This study demonstrates the power of joint analysis of historical DNA and large datasets generated through direct-to-consumer ancestry testing.

The vast majority of the ~45 million self-identified Black and/or African American individuals living in the United States descend from ~456,600 enslaved Africans who were forcibly transported to the US from Africa during the transatlantic slave trade between 1501 and 1867 (*1*, *2*). However, African Americans often have little information about these ancestors or their African origins owing to a history of inhumane treatment of the enslaved and their descendants, which included marginalization and the obfuscation of family histories (*3*). Here, we demonstrate that, when combined with genome-wide data from a sufficiently large and diverse genetic database, DNA from historical individuals provides a means for restoring knowledge of familial connections between contemporary peoples and their historical relatives. Specifically, we report on the DNA of enslaved and free African Americans from Catoctin Furnace, Maryland, who lived, worked, died, and were buried there in the late 18th and early 19th centuries.

As early as December 1768, a tract of land was acquired for the purposes of building an iron works at the foot of Catoctin Mountain near present-day Thurmont, Maryland (*4*). The furnace was in blast by 1776, producing pig iron, tools, household items, and munitions used during the Revolutionary War. At least 271 enslaved and an unknown number of free African Americans worked at Catoctin, within and outside the furnace, as ore miners, colliers, forgemen, fillers, teamsters, and woodcutters, as well as in domestic and agrarian roles in the furnace owners' households and plantations (*5*). In the second quarter of the 19th century, the furnace's labor force switched primarily to wage labor and a predominantly white workforce (*6*). Gradually, the contributions of African Americans in this early industrial complex were largely forgotten. The Catoctin Furnace African American Cemetery, near an old ore pit, was excavated in 1979–1980 in advance of highway construction (Fig. 1A) (*7*–*10*). The Maryland State Highway Administration transferred stewardship of the recovered remains of deceased humans to the Smithsonian Institution, where curator J. Lawrence Angel conducted preliminary forensic anthropological investigations (*11*).

The Catoctin Furnace Historical Society, Inc. (CFHS), was initially founded to save the Catoctin Furnace village and its archaeological and architectural heritage from this highway construction (*4*). In recent years, its mission has expanded to include restorative justice, highlighting the critical role that enslaved and free African Americans played in the furnace's history and in the growth of industrial wealth and power in the young United States. In 2015, a grant from the Maryland Heritage Areas Authority supported further scientific analysis of individuals buried in the cemetery. The first phase of the project involved historical documentary research and osteological reanalysis that refined previous assessments of demography and bone and dental pathology, with testing for stable carbon and nitrogen isotopes and trace elements to shed light on the life histories of the individuals (*12*). The project's second phase used DNA to explore their biogeographic ancestries and relationships to one another, details of which were summarized in a technical note (*13*) and are expanded upon here. These data address a critical component of CFHS's mission, pursued jointly with the African American Resources Cultural and Heritage (AARCH) Society of Frederick County, to return knowledge to the African American community and identify descendants of Catoctin's enslaved and free workers.

Advances in ancient DNA (aDNA) technology have made it possible to use genetic data as a tool for restoring knowledge of enslaved and historically marginalized peoples whose stories were often omitted from or disregarded in written records. Studies of the New York African Burial Ground (*14*, *15*), the Anson Street ancestors (*16*, *17*), and others (*18*, *19*) used a combination of anthropological and biomolecular tools to provide insight into the identity and life history of enslaved individuals through the study of their remains. However, the ability of those studies to localize the African origins of these individuals was limited by the exclusive use of mitochondrial DNA (mtDNA) (*20*) and/or reliance on comparisons with data from publicly available reference datasets (*21*, *22*). This study shows how deeper insights into the precise ancestral origins and genetic legacy of enslaved and free African Americans, such as those buried at Catoctin Furnace, can be obtained from genome-wide aDNA data when compared against a reference database containing genetic data from millions of living people, such as the one maintained by 23andMe, Inc.

We recovered genome-wide aDNA from all 27 of the Catoctin individuals who were selected for sampling, targeting ~1.2 million single-nucleotide polymorphisms (SNPs) using a capture-based approach (*23*–*26*), which we combined with imputation to further increase the amount of available genetic information (table S1). By comparing the DNA of the Catoctin individuals with genotype data from 9,255,493 participants in the 23andMe cohort,

[1]23andMe, Inc., Sunnyvale, CA 94043, USA. [2]Department of Human Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA. [3]Department of Anthropology, National Museum of Natural History, Smithsonian Institution, Washington, DC 20560, USA. [4]Department of Genetics, Harvard Medical School, Boston, MA 02115, USA. [5]Catoctin Furnace Historical Society, Thurmont, MD 21788, USA. [6]Hutchins Center for African and African American Research, Harvard University, Cambridge, MA 02138, USA. [7]African American and Black Diaspora Studies, Department of History, Boston University, Brookline, MA 02446, USA. [8]BIOMICs Research Group, Department of Zoology and Animal Cell Biology, University of the Basque Country UPV/EHU, 01006 Vitoria-Gasteiz, Spain. [9]Ikerbasque–Basque Foundation of Science, 48009 Bilbao, Spain. [10]Howard Hughes Medical Institute, Harvard Medical School, Boston, MA 02115, USA. [11]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA.
*Corresponding author. Email: eadaoinh@23andme.com (É.H.); jlmountain@gmail.com (J.L.M.); owsleyd@si.edu (D.W.O.); reich@genetics.med.harvard.edu (D.R.)
†The members of the 23andMe Research Team are listed at the end of this paper.
‡These authors contributed equally to this work.

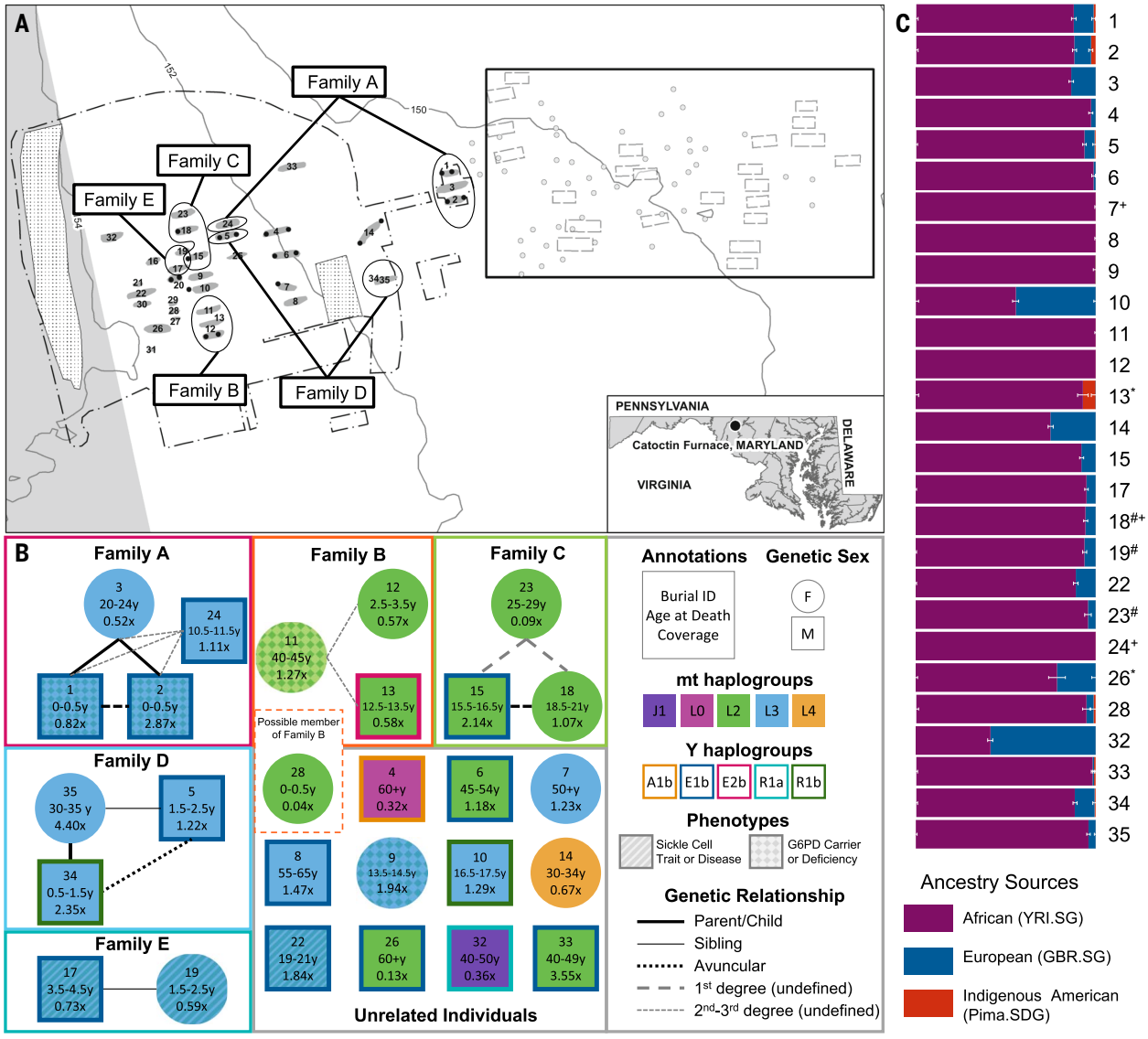


Fig. 1. Burial context, genetic kinship, and ancestry of Catoctin individuals. (**A**) Map showing the location of Catoctin Furnace and burials within the cemetery. Burial locations of the five genetic families are circled. The rectangle in the upper right shows a portion of the cemetery with unexcavated burials identified through ground-penetrating radar. [Map adapted from (*12*), prepared by Robert Wanner] (**B**) Individuals, labeled according to burial ID, are grouped into families on the basis of genetic relationships. Genetic sex, mt haplogroups, and Y haplogroups are indicated by marker shape, fill color, and outline color, respectively. The type of genetic relationship is indicated by connector line style. Marker fill pattern indicates individuals with one or more copies of an allele associated with sickle cell anemia or G6PD deficiency. (**C**) Ancestry proportions assigned to each individual from representative African (YRI), European (GBR), and Indigenous American (Pima) populations drawn from the public dataset according to the qpAdm software. Error bars indicate one standard error. Asterisks (*) indicate cases where damage-restricted data were analyzed. Hash symbols (#) and plus signs (+) indicate models with $P < 0.01$ or ancestry proportion estimates that fall more than three standard errors outside the range of 0 to 1, respectively.

all of whom consented to participate in research, we were able to learn about their biogeographic ancestries and genetic relationships with one another and to provide insight into their genetic legacy by identifying identical-by-descent (IBD) connections with living relatives.

Community engagement and ethical considerations

This research analyzed data from deceased individuals who were unable to directly consent to participate in this study, as well as from millions of research participants (including those genotyped by 23andMe) who actively consented to participate in research. The ties between present-day African Americans, their ancestors within the US, and their ancestors in Africa were forcibly severed by the transatlantic slave trade, the centuries-long institution of slavery, and generational systemic racist practices that have endured after the abolition of race-based slavery (*27*), as illustrated by Frederick Douglass's famous words: "Genealogical trees do not flourish among slaves" (*28*). Our objective is to contribute to the restoration of memories of a past community whose legacy was intentionally obscured and to create an avenue for living people to learn about their ancestors. We followed guidelines for the ethical analysis of the DNA of historical and living people (*29*, *30*), including consultation with stakeholder groups, as emphasized in recent discussions on the future of studies

involving the remains of African Americans (*31*–*35*). On the basis of interactions with stakeholders, we believe there is interest among African Americans and the public to harness aDNA to learn about historical connections to people who lived in the past and to leverage this technology to develop accurate methods to identify genetic relationships, many of which were previously unknown. Equally important is the need to communicate the results of these analyses with descendants and others in a sensitive and accurate manner.

In the case of Catoctin Furnace, the goals of CFHS, developed in partnership with AARCH, include identifying descendants and widening the community of stakeholders. To date, researchers at CFHS have traced the lineages of two African Americans (one enslaved and one free) who labored at Catoctin Furnace to their living descendants using historical documents and genealogical data. Members of CFHS, AARCH, and the recently discovered descendants all expressed support for the use of genetic approaches that could identify a larger descendant community. On the basis of conversations with these stakeholders about their research interests, we sought not only to identify living genetic relatives of the Catoctin individuals but also to conduct analyses to shed light on their life stories, such as identifying family relationships shared between the Catoctin individuals, exploring their African and European origins, and identifying biologically meaningful variants. In collaboration with CFHS and AARCH, we also held a series of public events to directly return research results at various phases of the project (*36*). One of the ways in which stakeholders chose to honor the legacy of enslaved individuals at Catoctin during these events was through the act of reading the names of individuals that could be abstracted from a variety of sources, including land records, probate inventories, church records, diaries, and freedom-seeker ads (although not associated with specific burials) (*5*). We include these names in the supplementary text of the supplementary materials (*36*). Establishing family connections to living people through genetics contributes to remembering and honoring those buried at Catoctin.

Here, we show that the joint analysis of DNA extracted from the remains of deceased humans and millions of living people in a recontactable research cohort (i.e., a cohort in which participants can be asked follow-up questions or receive results), such as the one maintained by 23andMe, makes it possible to recover previously unknown connections between present-day people and historical individuals from sites like Catoctin Furnace. Not all members of the Catoctin stakeholder community have a known genetic connection to Catoctin. Although this study is responsive to community requests to use genetic approaches to identify descendants, future studies applying these methods should be cautious about the danger of contributing to the biologization of notions of identity, as genetic connections represent just one of many ways in which people may feel connected to historical individuals. A full ethics statement is provided in (*36*). Additional ethical considerations involving the coanalysis of aDNA and data from private genetic databases are discussed in (*37*).

Multiple families buried at Catoctin Furnace

Among the 27 Catoctin individuals, we identified five distinct genetic families (labeled A to E) that are primarily composed of mothers, children, and siblings; in this study, the term "genetic family" refers to a group of people who are closely related biologically. Similarly, specific relationship terms (e.g., mother, son, daughter) are used in a biological sense and are based on genetic sex as inferred by the presence or absence of sex chromosomes; thus, they may not reflect the actual kin-based relationships, biological presentation, or gender identities recognized by the Catoctin individuals. Fifteen Catoctin individuals could be assigned to one of the five genetic families, whereas the remaining 12 individuals appear genetically unrelated, except for the individual from burial 28, whose coverage was insufficient to confidently assign to family B. Some unrelated individuals share mitochondrial (mt) or Y-chromosome haplogroups, which may indicate more distant relationships that fall outside the limits of our resolution. We used information about genetic sex and mt and Y haplogroups to further resolve these family groupings (Fig. 1B). Close genetic relatives tended to be buried near one another (Fig. 1A), whereas individuals who were buried separately from their genetic families were typically more distantly related. For instance, family A consists of a mother (burial 3) and two sons (burials 1 and 2), interred side by side, in addition to a second- or third-degree relative (burial 24) who was buried separately and whose exact relationship to the other individuals is unresolved. Although genetic relatedness had a role in burial patterning, other factors, such as temporal context and cultural and religious practices, likely contributed as well.

Sex-biased reproduction

The European ancestry of enslaved African Americans originated largely through a process whereby white men reproduced with Black women through rape. This gender-based sexual violence contributed to the brutal systematic enslavement of African Americans and frequently produced children born into slavery (*22*, *38*). This pattern of behavior, a form of sex-biased admixture, is reflected in the distribution of the mt and Y haplogroups observed among the Catoctin individuals. Three of the 16 Catoctin males have Y haplogroups that are broadly associated with West Eurasian ancestry (fig. S1). These include subclades of the R1a and R1b haplogroups that are common throughout Europe, indicating that their paternal lineages likely trace back to a fully European ancestor. In contrast, only one individual (burial 32) has a European-associated mt haplogroup (J1b1a1a) (fig. S2). This individual is an outlier with respect to ancestry, as they also have a European-associated Y haplogroup (R1a1a1b1a3b) and >50% European ancestry. Among other possible causes, their spatially separated grave, located in the northwestern edge of the burial ground, may reflect their distinct ancestral origins or lack of relatedness within the community represented in the cemetery.

Variable proportions of African, European, and Indigenous American–related ancestry in Catoctin individuals

All individuals (except the individual from burial 32) were assigned majority African ancestry by qpAdm, using a model designed to estimate each individual's African, European, and Indigenous American ancestry proportions (Fig. 1C and tables S2 and S3). More than one-quarter of individuals ($n = 7$) could be modeled as having no detectable European admixture (i.e., the amount of ancestry assigned to the European source population is within a single standard error of zero). This is in sharp contrast to nearly all present-day self-identifying African Americans, who typically have at least some European-derived ancestry [e.g., there is an average of 24% European ancestry among 23andMe research participants who self-identify as African American (*39*)]. Several individuals could be modeled as having Indigenous American ancestry, but in all cases, estimates were within three standard errors of zero. It is therefore uncertain, on the basis of these analyses alone, whether these estimates represent true Indigenous American ancestry.

We next imputed genotypes across the whole genome for each Catoctin individual using the software GLIMPSE (*40*), with an approach optimized for low-coverage, capture-based aDNA data. To ensure that the imputation process would not bias results, we tested the performance of the templated positional Burrows-Wheeler transform (TPBWT) (*41*) IBD detection tool and 23andMe's Ancestry Composition tool (*42*) on a dataset composed of 48 high-coverage ancient individuals who were down-sampled to varying degrees (*36*). Further, we compared the results of ADMIXTURE analysis (fig. S3), principal components analysis (fig. S4), and qpAdm analysis (fig. S5) using the imputed and nonimputed datasets to ensure that the imputation process did not substantially bias

ancestry estimates. We observed a bias toward estimating excess European-related ancestry in the lowest-coverage individuals when using the imputed dataset relative to the nonimputed dataset; however, this did not appear to substantially affect individuals with >0.5× coverage, which is the group we focused on for subsequent analyses.

Using the imputed dataset, we applied 23andMe's Ancestry Composition tool (*42*) to infer local ancestry along the chromosomes of the Catoctin individuals. Each region of their genomes was assigned to one of six previously determined broad ancestry categories: sub-Saharan African, European, East Asian and Indigenous American, Western Asian and North African, Melanesian, and Central and South Indian (fig. S6 and table S4). These assignments are correlated with qpAdm estimates (fig. S7) and provide support for the identification of Indigenous American ancestry in several Catoctin individuals. For instance, we inferred low levels of Indigenous American ancestry in two brothers (burials 1 and 2) from family A but not in their mother (burial 3) (Fig. 2), suggesting that their unsampled father had some Indigenous American ancestry.

Identity by descent

To learn more about the biogeographic ancestry and genetic legacy of the Catoctin individuals, we searched for IBD segments of the genome—that is, long segments of DNA that are identical in two or more people because they have been inherited from a recent common ancestor. We searched for IBD segments shared between each of the Catoctin individuals and ~9.3 million 23andMe research participants. We identified 55,342 IBD segments shared between the historical Catoctin individuals and 41,799 research participants, ranging up to 60 centimorgans (cM) in length (Table 1, figs. S8 and S9, and tables S5 and S6). We calculated the total IBD shared between each pair of individuals to estimate their most likely genetic relationship; however, we caution that we are likely underestimating the true amount of DNA shared between these individuals, particularly among close relatives (*36*). In Box 1, we discuss how the relationships between Catoctin individuals and research participants with whom they share DNA can be interpreted (table S7), noting that not all present-day individuals who share DNA with Catoctin individuals are direct descendants. In fact, most connections are likely between collateral relatives—relatives who are neither direct ancestors nor descendants of one another but instead both descend from a common ancestor who lived generations before the Catoctin individuals. Further, many of the most distant relatives whom we identified may not share a common ancestor who lived in the Americas. Instead, their connection may trace back to an individual who lived in Africa or Europe before their descendants' arrival in the Americas, either willingly or as part of the transatlantic slave trade.

Genetic connections to present-day Africans

We examined IBD segments shared between Catoctin individuals and members of the African cohort (i.e., 23andMe research participants with ≥95% sub-Saharan African ancestry who indicated that either they or all four of their grandparents were born in Africa) to

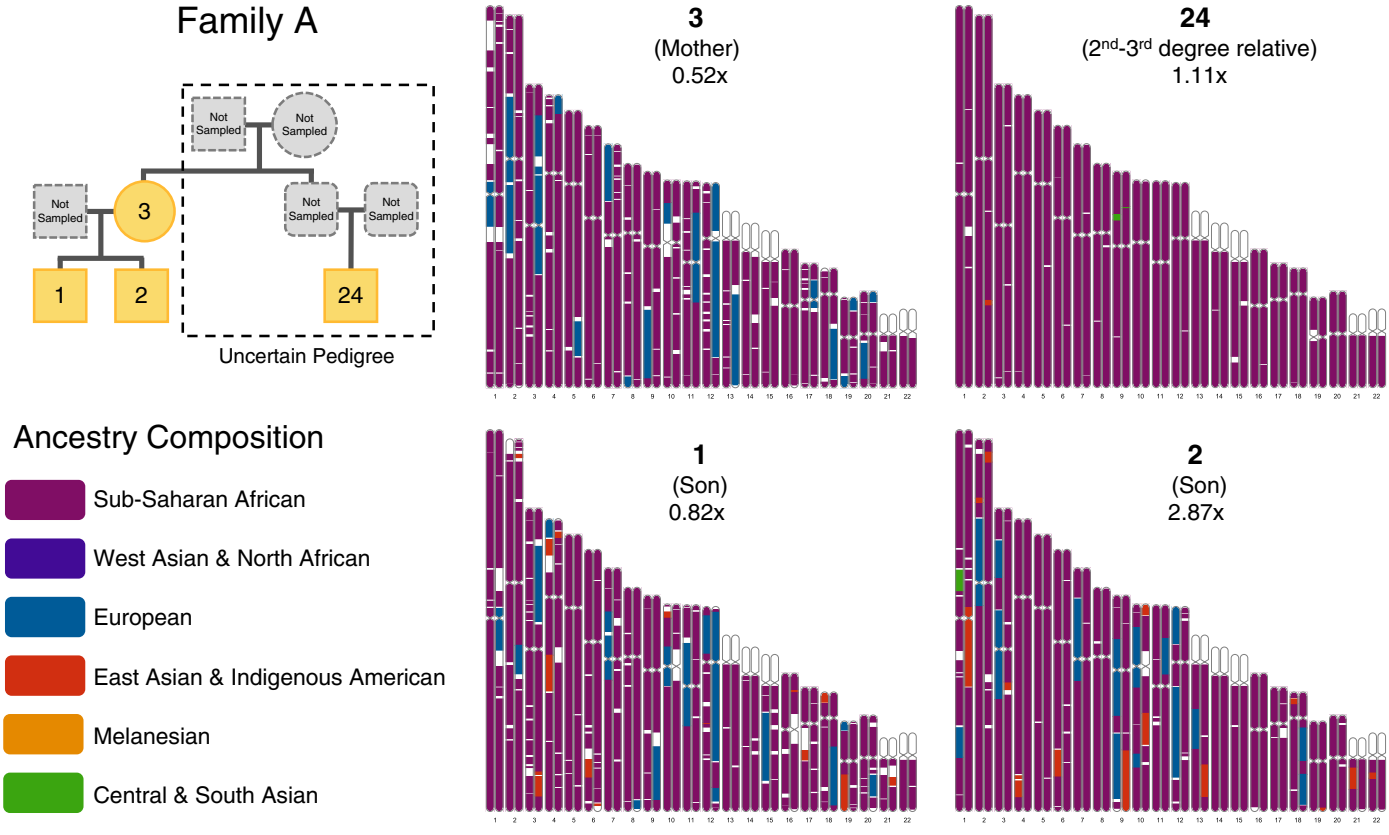


Fig. 2. Ancestry composition chromosome paintings for members of family A. Chromosome paintings demonstrating the biogeographic ancestry assigned to portions of the autosomal chromosomes for four related Catoctin individuals assigned to family A—a mother, two sons, and their second- or third-degree relative. A likely pedigree for family A (top left) describes their relationship to one another, although we note that the true relationship of the individuals from burials 1, 2, and 3 to the individual from burial 24 is uncertain. Across the genome, ancestry is assigned to one of six ancestry components defined using 23andMe reference populations: sub-Saharan African (purple), West Asian and North African (dark blue), European (dark teal), East Asian and Indigenous American (dark orange), Melanesian (light orange), and Central and South Asian (green). Unassigned regions are shown in white.

Table 1. IBD shared between 22 Catoctin individuals and 23andMe participants. Summary statistics describing frequencies and amounts of IBD shared between the 22 Catoctin individuals with >0.5× coverage and subsets of the 23andMe cohort. Participants were included in the US cohort if either (i) all four of their grandparents were born in the US or (ii) they were born in the US and their grandparents' birth location information was either unavailable or their grandparents were born in multiple countries. Similarly, members of the Atlantic African and European cohorts were determined using grandparent or participant birth location, with the additional requirement of ≥95% sub-Saharan African ancestry or ≥99% European, respectively. Results for each Catoctin individual are available in table S5. Values in columns marked with an asterisk are rounded according to the magnitude of IBD sharing as follows: values >100 cM are rounded to the nearest ten, values between 30 and 100 cM are rounded to the nearest five, values between 10 and 30 cM are rounded to the nearest integer, and values <10 cM are rounded to one decimal place.

23andMe participants included in group	Number of 23andMe participants in group	Proportion of 23andMe participants in group that share IBD with Catoctin individuals	Number of IBD segments detected	Median length of IBD segments detected* (cM)	Maximum length of IBD segments detected* (cM)	Median total IBD in 23andMe participants with IBD detected* (cM)	Maximum total IBD shared with 23andMe participants* (cM)
All participants	9,255,493	0.45%	55,342	10	60	9.9	280
Atlantic African cohort	3,304	2.27%	85	7.4	20	7.4	30
European cohort	226,384	0.22%	519	7.6	20	7.6	25
US cohort	2,993,165	0.51%	17,854	9.7	55	9.6	280
Members of US cohort with ≥5% sub-Saharan African ancestry	192,880	4.25%	10,675	11	55	11	280
Members of US cohort with ≥99% European ancestry	1,896,655	0.26%	5,123	8.0	30	8.0	40

identify the geographical regions in Africa with which these present-day people are associated. We observed the highest rates of IBD sharing between the Catoctin individuals and participants linked to Senegal, Gambia, Angola, and the Democratic Republic of the Congo (DRC) (Fig. 4A, fig. S10, and table S8), confirming via randomization testing that we would be unlikely to detect the same number of IBD connections (n = 10) with an identically sized sample (n = 166) of randomly selected individuals from the Atlantic African cohort [i.e., research participants in the African cohort with ties to specific Atlantic African countries, defined in (*36*)] (P < 0.001) (table S9).

Many African ethnolinguistic groups occupy wide geographic ranges that cross present-day national borders. We therefore also determined the amount of IBD each Catoctin individual shared with genetic clusters in Atlantic Africa that correspond to the self-identified ethnolinguistic groups of research participants (Fig. 4B and table S10). Of the 15 Catoctin individuals with detectable IBD connections to Africa, seven share a connection with only a single cluster. Six of these individuals have high proportions of sub-Saharan African ancestry (>90%). In contrast, it is less common for present-day research participants with four grandparents born in the US and ≥50% sub-Saharan African ancestry to have a connection to only a single cluster (*36*). Among the Catoctin individuals with connections to Atlantic Africans, the most commonly observed connections are to genetic groups associated with the Wolof, Mandinka, and Kongo (whose present-day geographic distribution is described in table S11). Overall, the Catoctin individuals share relatively more connections with Senegambian groups, like Wolof and Mandinka, than do research participants with four grandparents born in the US and ≥50% sub-Saharan African ancestry (Fig. 4 and supplementary text section S5).

Genetic connections to present-day Europeans

Next, we explored the rate of IBD sharing with members of the European cohort (i.e., 23andMe research participants with ≥99% European ancestry who indicated that either they or all four of their grandparents were born in Europe) to learn about the origins of the European-related ancestry that we observed in more than half of the Catoctin individuals. We detect the highest rates of IBD among participants associated with Great Britain and Ireland (Fig. 5A and table S12). Randomization testing confirms that we would be unlikely to detect the same number of IBD connections (n = 467) with an identically sized sample (n = 101,262) of randomly selected individuals from the European cohort (P < 0.001) (table S9). Much of the Irish-related signal is driven by connections to the individual from burial 15, an adolescent male belonging to family C, who, when projected onto a graph network of clusters that correspond to European geography, again shares the most IBD with clusters of participants associated with the northern, western, and southeastern regions of the island of Ireland (Fig. 5B). Multiple other Catoctin individuals (in particular, those from burials 2, 10, and 34) share broader connections with participants from Great Britain and Ireland (Fig. 5B, fig. S11, and table S13).

Distant and close relatives in the United States

When we considered IBD sharing between research participants in the US cohort (i.e., research participants who indicated that either they or all four of their grandparents were born in the US) and each of the Catoctin individuals separately, each historical individual exhibited a distinctive pattern of IBD sharing with respect to geography. However, when we considered the Catoctin individuals together, we observed the highest rates of sharing between Catoctin individuals and research participants from the southern US (Fig. 6A, fig. S12, and table S14). This signal resembles the geographic distribution of 23andMe research participants with sub-Saharan African ancestry in the US (fig. S13) and is therefore plausibly driven by a higher rate of IBD sharing in genomic regions with sub-Saharan African ancestry. To address this source of bias, we

Box 1. Interpreting DNA shared between historical and present-day individuals.

The Catoctin Furnace African American Cemetery was in use between 1774 and 1850 (midpoint: 1812), while the average birth year of 23andMe research participants with genetic connections to Catoctin is 1974 (table S22). The average separation between the death of the historical individuals and birth of present-day genetic relatives is ~162 years (although it is possible that as few as 70 years may have elapsed since the date of death of some of the latest Catoctin burials and the oldest research participants). The average age of reproduction among humans is 28 to 30 years, therefore the most likely number of generations separating the Catoctin individuals from their living relatives is about five to six generations (Fig. 3A). The vast majority of the genetic relatives that we detected are not direct descendants of the Catoctin individuals. Instead, they are collateral relatives (i.e., relatives who are not direct ancestors or descendants) (Fig. 3B). To determine the likelihood that someone shares a direct descendant relationship with a Catoctin individual, we must consider the age of the two individuals and the amount of identical DNA (or IBD) that they share. Relatives who share very little IBD but who lived close together in time (as is the case for most of the connections we identified), are less likely to share a direct descendant-ancestor relationship.

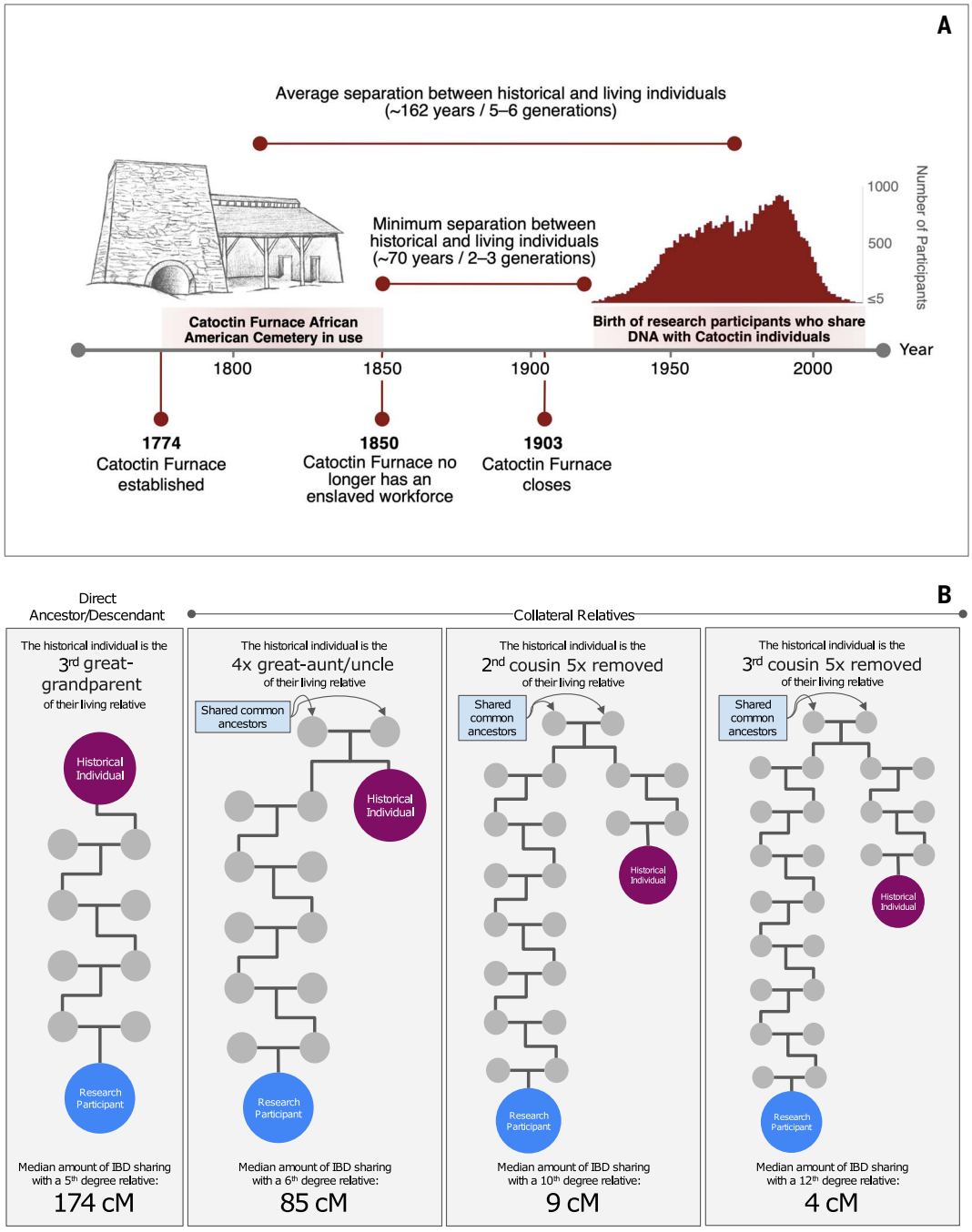


Fig. 3. Examples of chronological and genealogical distance between historical individuals and research participants. (**A**) A timeline showing the years in which the African American Cemetery at Catoctin Furnace was active and a histogram of birth years of research participants who share IBD with Catoctin individuals (table S23). (**B**) Examples of relationships that could be shared between individuals who were born five generations apart, with varying degrees of genetic separation. Median amount of IBD is reported for pairs of present-day and historical individuals (with 2× coverage) (table S7).

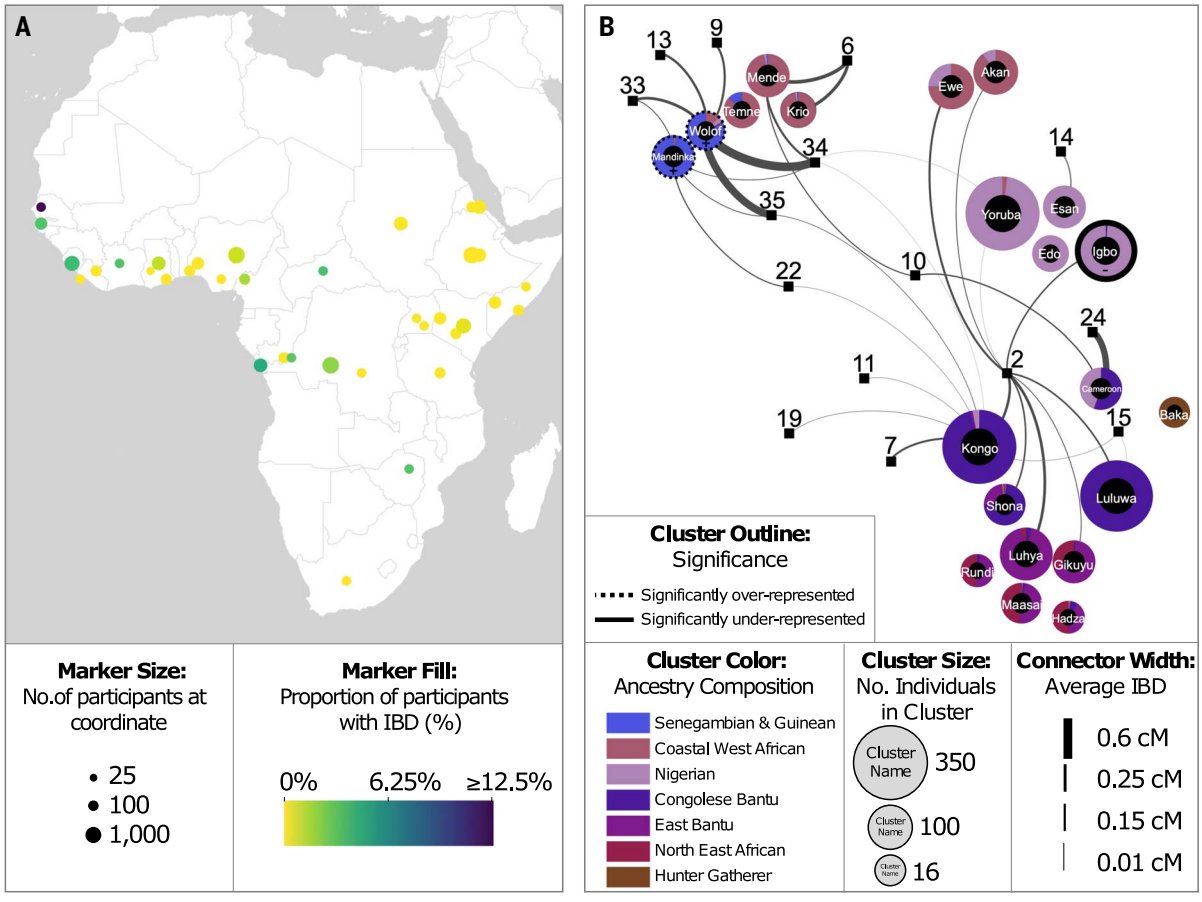


Fig. 4. Genetic connections to the Catoctin individuals among members of the African cohort. (**A**) The proportion of 23andMe participants in the African cohort who share IBD with Catoctin individuals. Geographic coordinates are rounded to the nearest integer, and only coordinates that have at least 25 associated participants after 80% down-sampling are shown. Marker size indicates the number of participants associated with each coordinate, and color indicates the proportion of participants who share IBD with the Catoctin individuals. (**B**) IBD network demonstrating Catoctin individuals' connections to members of the African cohort who share <700 cM of IBD with one another (N = 2807). IBD clusters (represented by circles) are filled according to members' average local African ancestry and arranged by average pairwise IBD sharing between clusters using a Force Atlas graph layout. Catoctin individuals, displayed as squares, are projected on the basis of their average IBD shared with each cluster (shown as lines).

restricted our analysis to research participants in the US cohort with ≥5% sub-Saharan African ancestry (Fig. 6B and table S15). This filtering strategy increases the rate of IBD sharing from 0.45% of all US participants to 4.25% (Table 1).

For participants included in this filtered dataset, we continued to observe elevated rates of IBD sharing with Catoctin in the southern US (including Maryland) (Fig. 6B). We confirmed via a randomization test that we would be unlikely to detect the same number of IBD connections (n = 2034) with an identically sized sample (n = 42,132) of randomly sampled individuals from the US cohort with ≥5% sub-Saharan African ancestry (P < 0.001) (table S9). In contrast, when we filtered to include only participants in the US cohort with ≥99% European ancestry to focus on the genetic legacy of the admixed Catoctin individuals' European ancestors along lineages with little to no African admixture, we did not detect any clear geographic patterns (Fig. 6C and table S12).

When we focused on "close relatives" (understood here as pairs of individuals who share at least 30 cM of IBD with a Catoctin individual, reflecting a relationship that is predicted to be ninth degree or closer), we observed particularly pronounced connections to Maryland, identifying a total of 30 close relatives in the state (table S16). A randomization test confirmed that we would be unlikely to identify 30 or more close relatives among an identically sized sample (n = 19,972) of randomly selected individuals (P < 0.001) (table S17). These results suggest that at least some descendants of the Catoctin individuals or their family members remain in the Maryland area.

Next, we considered population substructure among the close relatives of the Catoctin individuals in the US using an IBD network approach. We analyzed 443 close relatives of the Catoctin individuals (≥30 cM of IBD) along with 4385 of those participants' closest relatives who share ≥100 cM of IBD—many of whom we hypothesize could be genetically related to the Catoctin individuals, even if we did not detect IBD sharing between them owing to the high false-negative rate of our approach—and identified 123 familial groups (Fig. 6D and table S18). The positions of these familial clusters within the IBD network appear to be primarily correlated with the relative proportions of European and sub-Saharan African ancestry detected among each cluster's members, likely reflecting the ancestry of the individual IBD segments shared with the Catoctin individuals. In most cases, the familial groups do not appear to be correlated with geography. Notably, however, all members of familial group 36 who provided location information have ties to Maryland, providing further evidence that Catoctin descendants and close relatives remained in or returned to Maryland after emancipation.

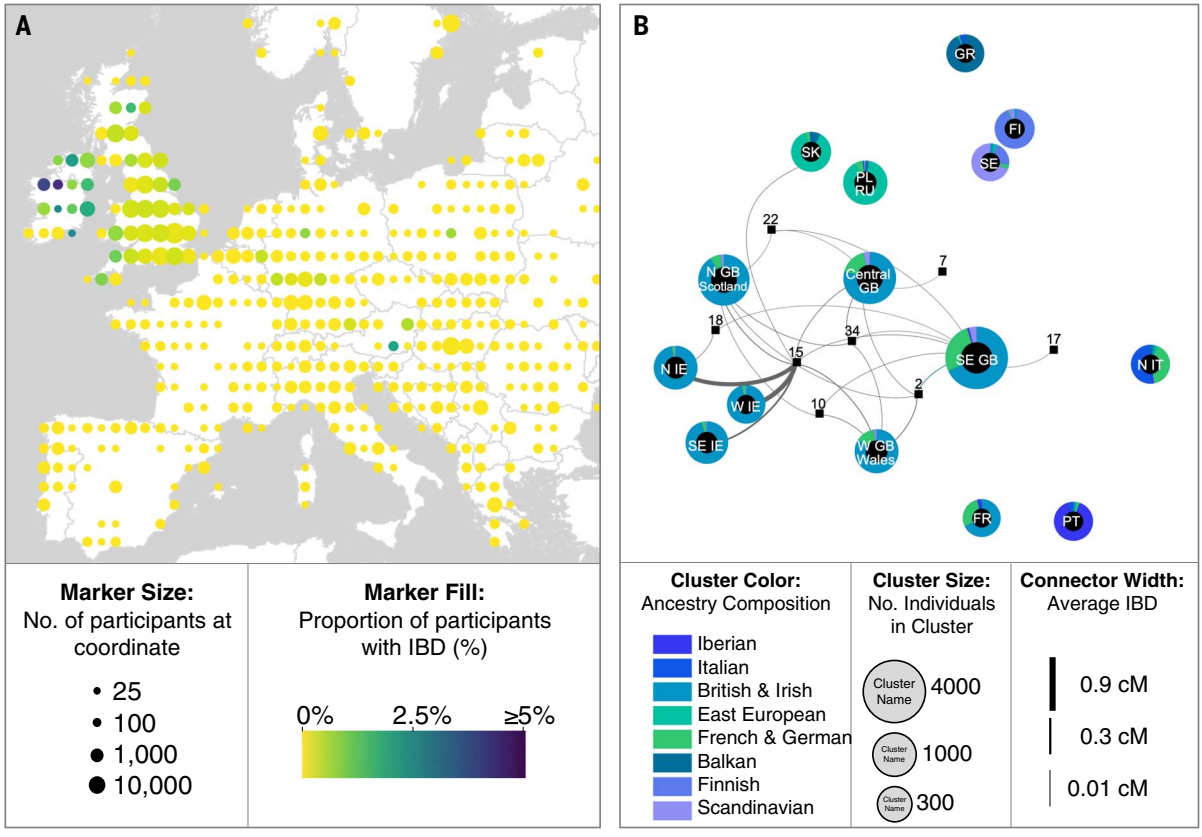


Fig. 5. Genetic connections to the Catoctin individuals among members of the European cohort. (**A**) Proportion of 23andMe participants in the European cohort who share IBD with Catoctin individuals. Geographic coordinates are rounded to the nearest integer, and only coordinates that have at least 25 associated participants after 80% down-sampling are shown. Marker size corresponds to the number of participants associated with each location, while the color indicates the proportion of participants who share IBD with the Catoctin individuals. (**B**) IBD network indicating Catoctin individuals' connections to the 23andMe participants in the European cohort who share <700 cM with one another. Clusters are labeled by the geographic region with which the plurality of cluster members are associated using ISO2 country abbreviations and, when appropriate, prefixes to indicate the cardinal directions. Clusters are arranged by the average pairwise IBD sharing between clusters using a Force Atlas graph layout, with outlines that indicate participants' average local European ancestry. Catoctin individuals, displayed as squares, are projected on the basis of the average IBD shared with each European cluster (shown as lines).

While elevated rates of IBD sharing are particularly evident in the Maryland region, we also identified other regions of the US with an enrichment of Catoctin close relatives. For example, the maximum amount of IBD (280 cM) is observed among a set of individuals from Southern California. This amount of sharing (corresponding to ~4% of the autosomal genome) is consistent with a fifth-degree relationship to the individual from burial 35 on the basis of maximum likelihood predictions (although the actual relationship may differ by a few degrees in either direction). Individuals with this amount of IBD sharing are likely either direct descendants of those buried at Catoctin or direct descendants of very close relatives of the Catoctin individuals (given that the Catoctin individuals likely lived at least five generations before most of the research participants included in this study) (Box 1 and table S7).

To reconstruct pedigrees describing the connections shared between the Catoctin individuals and their closest genetic relatives in the 23andMe cohort, we used a modified version of the tool Bonsai (*36*, *43*). The informed consent process for participation in 23andMe research requires strict protection of research participant anonymity, which means that full pedigrees cannot be published. Instead, we highlight the ways in which 8721 independent pedigrees (636 of which contain more than one research participant) are connected to Catoctin family A (Fig. 7 and table S19). We found no cases where the most likely connection was via a direct descendant of individuals from burials 1, 2, or 24, consistent with our expectations because these individuals died during childhood. The most probable path for most (83%) of the pedigrees we considered connects through an ancestor of the unsampled father of individuals from burials 1 and 2 (referred to as ungenotyped individual f in Fig. 7). We find that pedigrees that include research participants who have at least twice as much Indigenous American ancestry as sub-Saharan African ancestry were significantly more likely to connect through this unsampled individual than through other Catoctin individuals ($P < 10^{-6}$; calculated via a permutation test with 10^6 replicates) (supplementary text S6), consistent with our earlier prediction that this unsampled individual had some Indigenous American ancestry.

Most inferred connections extend upward in pedigrees generated for families A, C, and D (meaning that they connect through an ancestor of one or more members of the family and therefore do not involve a direct-descent relationship) (Fig. 7, fig. S14, and table S19). This is consistent with expectations, as individuals who lived in the past few hundred years are predicted to have far more living collateral relatives than direct descendants. Notably, lineages that extend upward from the Catoctin pedigrees tend to involve research participants who have more European ancestry than those lineages that extend downward. This suggests that a relatively large fraction of the

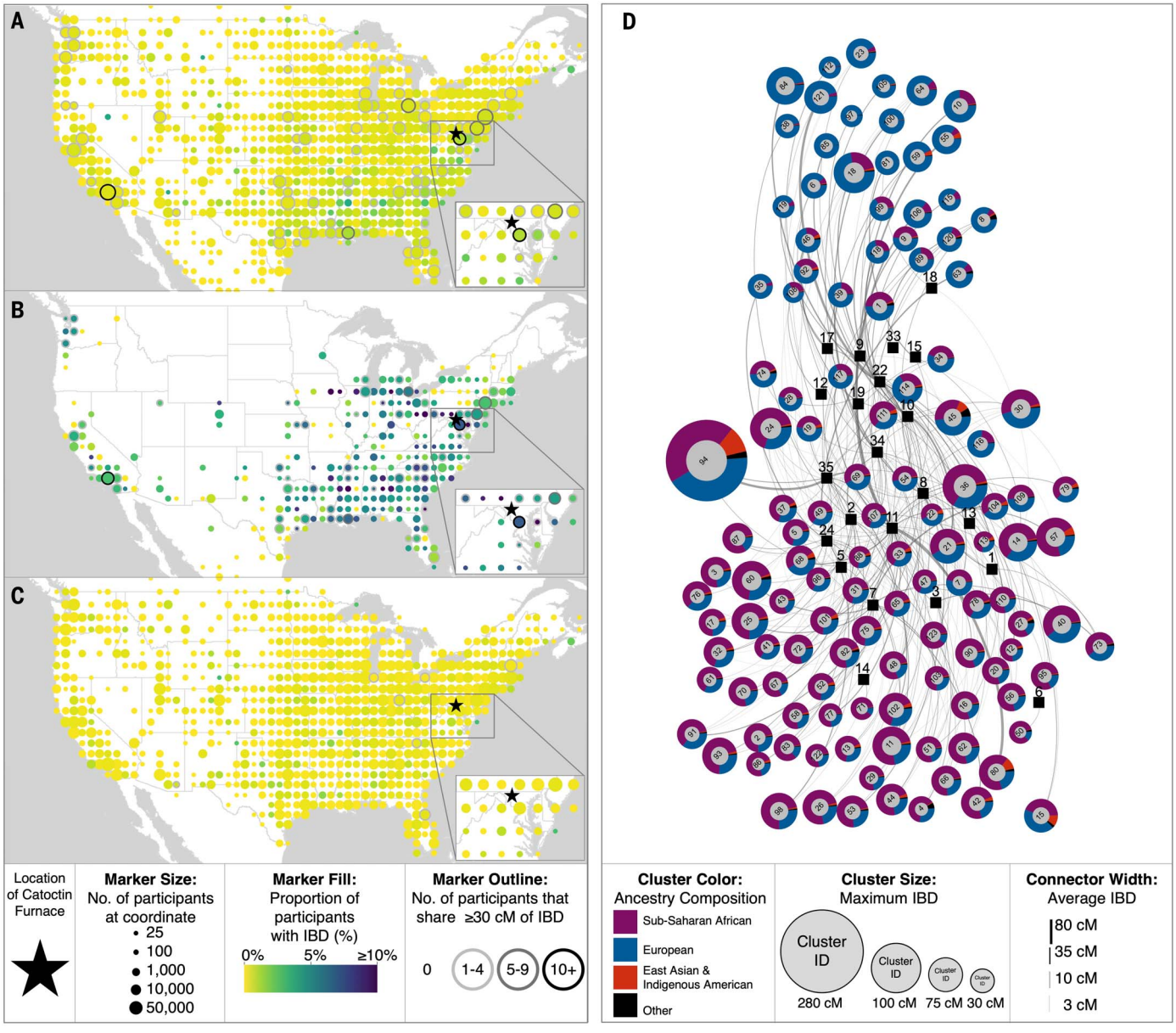


Fig. 6. Geographic distribution of distant and close relatives of the Catoctin individuals among members of the US cohort. (**A**) Proportion of 23andMe research participants in the US cohort who share IBD with Catoctin individuals. Only coordinates representing at least 25 participants after 80% down-sampling are shown. Marker size corresponds to the number of participants associated with each coordinate, while color indicates the proportion of participants with shared IBD. Marker outlines indicate the number of participants at each coordinate who share ≥30 cM of IBD with one or more Catoctin individuals. The same information is shown for (**B**) participants with ≥5% sub-Saharan African ancestry and (**C**) participants with ≥99% European ancestry. (**D**) IBD network of the closest present-day relatives of Catoctin individuals among 23andMe research participants. Circles represent modularity clusters consisting of close Catoctin relatives (sharing ≥30 cM of IBD) along with their relatives (sharing ≥100 cM with a close relative of a Catoctin individual). Clusters are outlined according to their average ancestry and arranged by the average pairwise IBD sharing between clusters using a Force Atlas layout. Catoctin individuals, displayed as squares, are projected on the basis of the average IBD shared with each familial group (shown as lines).

connections that we identified involve ancestors of members of family A with European ancestry, which in part may reflect biases in the 23andMe cohort where European ancestry is overrepresented.

These results demonstrate the power of our IBD-based approach to identify connections between historical and present-day individuals. In the future, by obtaining additional informed consent from research participants, it may be possible to present more complex pedigrees that include direct descendants of historical individuals using this approach.

Biologically meaningful variants

We considered sites in the Catoctin individuals' genomes that might shed light on their physical traits and health (table S20 and fig. S15). However, we caution that these results are based on low-coverage data and that further work is required to conclusively infer genotypes at these positions.

For three Catoctin individuals, we identified copies of only the causal A allele at the genetic position *rs334/i3003137*, which is associated with protection against malaria in the heterozygous form and sickle cell disease—a red

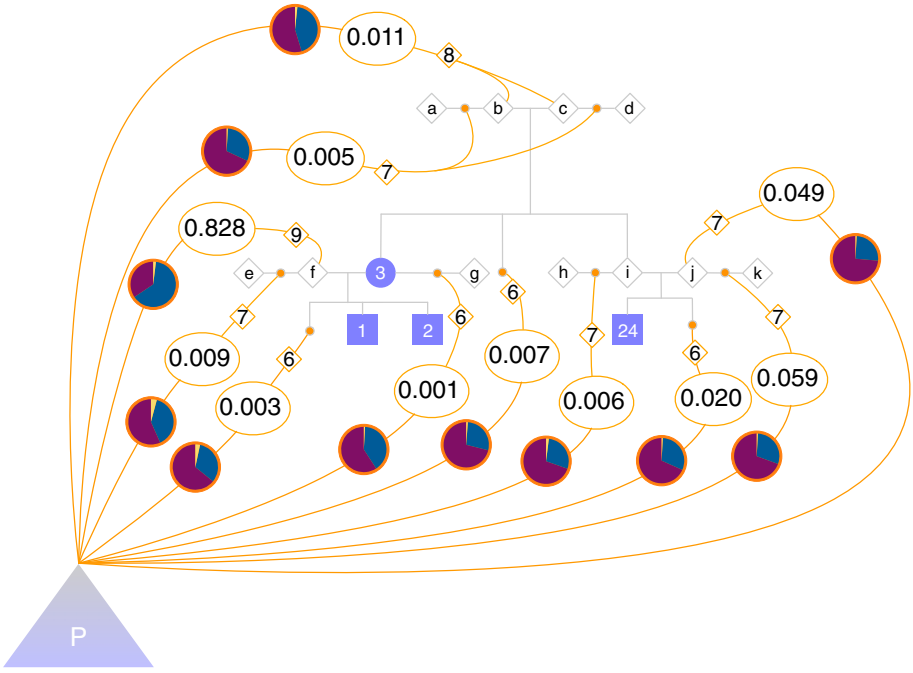


Fig. 7. Connections between Catoctin family A and modern pedigrees. The pedigree for Catoctin family A is shown with blue-shaded individuals connected by gray lines. Open gray diamonds indicate ungenotyped individuals, some of whom must have existed (individuals b, c, f, i, and j), whereas others (individuals a, d, e, g, h, and k) may have existed. The large triangle represents all present-day pedigrees composed of 23andMe research participants, and the probability distribution of how these pedigrees connect to the historical pedigree is inferred. Orange dots indicate all possible points of connection of present-day pedigrees to the historical pedigree. Numbers in ovals give the percentage of present-day pedigrees whose most likely connection was to a given point on the historical pedigree. Numbers in orange diamonds indicate the average degree of a lineage connecting to a particular point. Pie charts show the average European (blue), sub-Saharan African (purple), and Indigenous American (yellow) ancestry (normalized to sum to one) of individuals in pedigrees whose most likely point of connection was through the respective lineage leading to the present day.

blood cell disorder that causes pain and increased susceptibility to infection and early mortality (*44*)—in individuals with two copies of the allele. Although we only have limited coverage at this position for each of these three individuals (range: 1 to 5× coverage), making it impossible to confidently determine their genotypes, we note that two individuals (burials 17 and 19) are siblings who died during early childhood. Sickle cell disease, if they had it, may have contributed to their early mortality.

We also observed four individuals with at least one copy of the causal T allele at another genetic position (*rs1050828*) that provides protection against malaria in genetically female individuals who carry one copy of this allele on only one of their X chromosomes. In genetic males, who are hemizygous, and in genetic females who are homozygous for the T allele at this position, it is associated with glucose-6-phosphate dehydrogenase (G6PD) deficiency (*45*). Both variants occur at elevated rates in populations with ancestry from sub-Saharan Africa, where malaria has historically been endemic (*46*). Understanding how the frequency of these alleles may have changed in African American populations over time is a question of great interest (*47*, *48*), which may be informed by aDNA data as more historical African Americans are sequenced.

Discussion

This study demonstrates the power of genetic analyses to uncover previously unknown information about the family structure and ancestry of historical individuals and to connect them with living relatives. For Catoctin Furnace, this research is a critical step toward identifying a larger descendant community, one of the main goals expressed by stakeholders from CFHS, AARCH, and the two Catoctin descendant families recently identified using historical documents and genealogical data. Using the approach introduced here to identify genetic connections between historical individuals and their present-day relatives, researchers now have a powerful means to reconstruct these relationships through DNA analysis.

At least 15 of the Catoctin individuals can be clustered into five genetic families, providing insight into the social structure of African Americans at this early industrial site. Historical records suggest that iron works enslavers often kept families together to benefit from a sustained knowledge transfer between generations. It was also believed that this practice minimized the likelihood of revolt or escape caused by family separation (*49*–*51*). Most of the genetic connections we identified were between first-degree relatives, usually mothers and children. No fathers were identified, nor did we find families represented by three or more generations. There were at least 11 individuals who appeared unrelated to others buried in the cemetery. These results may therefore indicate that, in practice, families did not remain together at this iron works. One possible explanation is that partners may have been sought outside of the Catoctin village. However, additional consideration must be given to how the types of genetic relationships we observed may be biased by incomplete excavation combined with burial patterning at the site, which was likely influenced by personal choice and/or imposed religious practices. For instance, from firsthand accounts recorded in journals and diaries, we know that some African American funerals at Catoctin were conducted by Moravian ministers (*52*, *53*). The Moravians prescribed burial protocols in which families were not interred together but instead were buried in "choirs," separated by marital status, age, and gender (*54*). This tradition introduces the possibility that married men were buried elsewhere in the cemetery and therefore were not sampled as part of this study (*54*). It was only through the coanalysis of aDNA and archival Moravian diaries that this distinctive aspect of burial and demography at Catoctin could be considered.

The genetically inherited conditions identified in this study (sickle cell anemia and G6PD deficiency) provide additional insight into the health and well-being of the Catoctin African Americans. Although these results are tentative and require more refined analyses, they offer insight and potential future avenues by which to explore the remains of deceased humans recovered from archaeological contexts.

At the population level, the Catoctin individuals have diverse ancestry, with clear genetic links to Africa, Europe, and the Americas. Most individuals have primarily sub-Saharan African ancestry, with the strongest ties to present-day peoples in Senegambia and West Central Africa, a region that primarily encompasses present-day Angola and the DRC. These findings accord with historical records that show that slave ships originating from Senegambia (particularly before the end of the 18th century) and West Central Africa accounted for the highest disembarkation rates in Maryland over the course of the transatlantic slave trade, whereas lower disembarkation rates were recorded from intervening coastal

regions and the Caribbean (table S21) (*55*). However, records of the intra-American slave trade, which was responsible for the arrival of many enslaved African Americans to Maryland, are more limited, making it challenging to infer what the most likely major sources of African ancestry in Maryland might have been on the basis of historical records alone. According to the transatlantic slave trade database, ships departing from Senegambia and West Central Africa accounted for 15.8 and 20.8%, respectively, of overall disembarkations in North America (*55*). Notably, the most common sources of African ancestry among the Catoctin individuals do not align with larger trends in the greater Chesapeake region, where the highest rate of disembarkation was from the Bight of Biafra, home to people of Igbo and Yoruba ancestry (*55*).

The Catoctin individuals were less likely to have genetic connections to multiple distinct African ethnolinguistic groups than research participants with substantial (≥50%) sub-Saharan African ancestry and long-standing ties to the US (i.e., four grandparents born in the US), likely reflecting admixing between individuals with diverse African ancestries that occurred among enslaved African Americans and their descendants over time, after forced migration to the Americas (*22*, *56*).

Given the overrepresentation of research participants with European (particularly British and Irish) ancestry in the 23andMe cohort, we have even more power to localize the European ancestry of the Catoctin individuals. Eight individuals exhibit connections to Great Britain and/or Ireland. Differing frequencies of European-associated maternal and paternal haplogroups observed at Catoctin indicate that their European ancestry was likely introduced through a sex-biased process almost certainly driven by rape of enslaved women as part of the gender-based sexual violence inherent in the US's system of chattel slavery (*38*). We also observed a single European-associated mt haplogroup (in an individual who we estimate to have >50% European ancestry). This may be one of numerous examples present in the historical record of enslaved or free Black men having children with white women (often indentured servants) (*57*).

Another objective of this study was to explore the possibility of identifying direct descendants using DNA. We identified 41,799 research participants with genetic connections to the Catoctin individuals. In many cases, it was possible to construct detailed genetic pedigrees that link 23andMe research participants to the historical individuals from Catoctin, either as direct descendants or, most commonly, as collateral relatives with a shared common ancestor. Although most of these connections are distant, we identified >500 relatives who share ≥30 cM of IBD, reflecting a maximum likelihood estimate of ninth-degree relationship or closer.

It has been suggested that Catoctin's enslaved workers were sold and transported to more-southern states when the furnace transitioned to white wage labor (*4*). During the early 19th century, large numbers of enslaved individuals were sold from mid-Atlantic states and transported south (*58*). Although we observe strong connections to Maryland that suggest that this was unlikely to be the fate of Catoctin's entire enslaved population, we do observe small clusters of close relatives throughout the US, including in the South.

The methodological approach we present in this study can be applied to the remains of deceased humans from other sites and contexts, offering a new scientific tool for individuals and descendant communities seeking greater knowledge of their ancestors, as well as archaeologists, bioarchaeologists, historians, and genealogists. Museums and universities that steward the remains of deceased humans now have an additional means by which to identify these individuals and potentially link them to biological descendants. This work can currently only be done in partnership with organizations with access to massive, genetically diverse, recontactable research cohorts, such as those maintained by genetic-ancestry companies. These partnerships will require conversations on the ethical implications and consequences of this work, particularly how to avoid reinforcing the biologization of identity (*37*). For researchers, this study represents an advance in scientific methodology, but the impact may be even greater for those seeking connections to their past.

Materials and methods summary

We sampled aDNA from the temporal bones of 27 Catoctin individuals, using a minimally destructive cranial-based drilling approach when sampling from intact skulls (*59*). We generated double-stranded, partially uracil-DNA glycosylase–treated DNA libraries (*60*–*64*). Before sequencing, we enriched the libraries for DNA aligning to the mitochondrial genome and 1.2 million positions in the nuclear genome using a capture-based approach (*23*–*26*). After bioinformatic processing, all 27 DNA samples were deemed suitable for analysis, although the data for two individuals were subjected to damage restriction because of slightly elevated mitochondrial (*26*) or X-chromosome (*65*) contamination rates. For each individual, we inferred genetic sex (*66*) and uniparental haplogroups (*67*) and estimated the proportion of African, European, and Indigenous American–related ancestry by comparing against publicly available datasets (*68*–*71*). Diploid genotypes were imputed using GLIMPSE (*72*), with the 1000 Genomes Project phase 3 dataset (*68*) as a reference panel. After filtering and re-phasing the data (*73*), we searched for IBD (*41*) between the imputed Catoctin individuals and 9,255,493 participants (elsewhere referred to as the 23andMe cohort) who had been genotyped by 23andMe, Inc., a consumer personal genetics company, and provided informed consent to participate in research by 28 July 2020. Summary statistics were generated describing these IBD connections for cohorts of research participants that were created on the basis of each research participant's genetic ancestry (as determined by the tool Ancestry Composition) and answers to 23andMe survey questions about their birth location and grandparents' birth locations. Genetic pedigrees were reconstructed using a modified version of the Bonsai pedigree inference algorithm (*43*). Finally, we counted the number of different DNA sequences that overlap biologically meaningful positions in the genome for each Catoctin individual.

REFERENCES AND NOTES

1. D. Eltis, D. Richardson, *Atlas of the Transatlantic Slave Trade* (Yale Univ. Press, 2010).
2. US Census Bureau QuickFacts; https://www.census.gov/quickfacts/fact/table/US/RHI225219#RHI225219.
3. H. A. Williams, *Help Me to Find My People: The African American Search for Family Lost in Slavery* (Univ. of North Carolina Press, 2012).
4. E. Y. Anderson, *Catoctin Furnace: Portrait of an Iron Making Village* (Arcadia Publishing, 2013).
5. Access database of confirmed and potential members of the Catoctin Furnace community, compiled by donated labor for the Catoctin Furnace Historical Society, Elizabeth Anderson Comer, primary manager (Catoctin Furnace Historical Society, 2015).
6. Frederick County Account of Sales Liber GME 9, Folio 353-361, List of names, p. 355.
7. R. A. Thomas, G. S. Mellin, T. Payne, S. Burnston, J. P. McCarthy, "Archaeological Investigations at Catoctin Furnace, Frederick County, Maryland" (Mid-Atlantic Archaeological Research, Inc., 1980).
8. A. Townsend, R. Thomas, K. Orr, "Specifications and Research Designs. The Catoctin Furnace Archaeological Mitigation Project. U.S. Route 15 from Putman Road to Route 77" (Prepared for the State Highway Administration, Baltimore, 1979).
9. K. Orr, R. Orr, "Interim Report of the Catoctin Furnace Archaeological Mitigation Project Contract F522-152-770" (Prepared for the Maryland State Highway Administration, Baltimore, 1980).
10. S. Burnston, R. A. Thomas, "Archaeology Data Recovery at Catoctin Furnace Cemetery" (Prepared for Orr and Son, Consulting Archaeologists, 1981).
11. J. O. Kelley, J. L.Angel, The workers of Catoctin Furnace. *Maryland Archeol.* **19**, 2–17 (1983).
12. K. S. Bruwelheide *et al.*, Restoring identity to people and place: Reanalysis of human skeletal remains from a cemetery at Catoctin Furnace, Maryland. *Hist. Archaeol.* **54**, 110–137 (2020). doi: 10.1007/s41636-019-00214-7
13. É. Harney *et al.*, Technical Report on Ancient DNA analysis of 27 African Americans from Catoctin Furnace, Maryland. bioRxiv [Preprint] 2022.06.12.495320 (2022); https://doi.org/10.1101/2022.06.12.495320.
14. M. L. Blakey, L. M. Rankin-Hill, "The skeletal biology of the New York African Burial Ground" (Howard Univ. Press in association with the US General Services Administration, 2009).
15. C. K. Clinton, F. L. C. Jackson, Historical overview, current research, and emerging bioethical guidelines in researching the New York African burial ground. *Am. J. Phys. Anthropol.* **175**, 339–349 (2021). doi: 10.1002/ajpa.24171; pmid: 33247601
16. R. E. Fleskes *et al.*, Ancestry, health, and lived experiences of enslaved Africans in 18th century Charleston: An osteobiographical analysis. *Am. J. Phys. Anthropol.* **175**, 3–24 (2021). doi: 10.1002/ajpa.24149; pmid: 33022107

17. R. E. Fleskes *et al*., Community-engaged ancient DNA project reveals diverse origins of 18th-century African descendants in Charleston, South Carolina. *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2201620120 (2023). doi: 10.1073/pnas.2201620120; pmid: 36623185
18. R. Barquera *et al*., Origin and health status of first-generation Africans from early Colonial Mexico. *Curr. Biol.* **30**, 2078–2091.e11 (2020). doi: 10.1016/j.cub.2020.04.002; pmid: 32359431
19. R. E. Fleskes *et al*., Ancient DNA and bioarchaeological perspectives on European and African diversity and relationships on the colonial Delaware frontier. *Am. J. Phys. Anthropol.* **170**, 232–245 (2019). doi: 10.1002/ajpa.23887; pmid: 31270812
20. A. Salas *et al*., The African diaspora: Mitochondrial DNA and the Atlantic slave trade. *Am. J. Hum. Genet.* **74**, 454–465 (2004). doi: 10.1086/382194; pmid: 14872407
21. A. R. Bentley, S. Callier, C. N. Rotimi, Diversity and inclusion in genomic research: Why the uneven progress? *J. Community Genet.* **8**, 255–266 (2017). doi: 10.1007/s12687-017-0316-6; pmid: 28770442
22. S. J. Micheletti *et al*., Genetic consequences of the transatlantic slave trade in the Americas. *Am. J. Hum. Genet.* **107**, 265–277 (2020). doi: 10.1016/j.ajhg.2020.06.012; pmid: 32707084
23. I. Mathieson *et al*., Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499–503 (2015). doi: 10.1038/nature16152; pmid: 26595274
24. W. Haak *et al*., Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207–211 (2015). doi: 10.1038/nature14317; pmid: 25731166
25. Q. Fu *et al*., An early modern human from Romania with a recent Neanderthal ancestor. *Nature* **524**, 216–219 (2015). doi: 10.1038/nature14558; pmid: 26098372
26. Q. Fu *et al*., A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr. Biol.* **23**, 553–559 (2013). doi: 10.1016/j.cub.2013.02.044; pmid: 23523248
27. H. L. Gates, *Life Upon These Shores: Looking at African American History, 1513–2008* (Alfred A. Knopf, 2011).
28. F. Douglass, *My Bondage and My Freedom* (Oxford Univ. Press, 2019).
29. J. K. Wagner *et al*., Fostering responsible research on ancient DNA. *Am. J. Hum. Genet.* **107**, 183–195 (2020). doi: 10.1016/j.ajhg.2020.06.017; pmid: 32763189
30. S. Alpaslan-Roodenberg *et al*., Ethics of DNA research on human remains: Five globally applicable guidelines. *Nature* **599**, 41–46 (2021). doi: 10.1038/s41586-021-04008-x; pmid: 34671160
31. J. Dunnavant, D. Justinvil, C. Colwell, Craft an African American Graves Protection and Repatriation Act. *Nature* **593**, 337–340 (2021). doi: 10.1038/d41586-021-01320-4; pmid: 34012089
32. F. Jackson, C. Clinton, J. Graves, Genomics of African American remains—limits must not compound inequity. *Nature* **596**, 35–35 (2021). doi: 10.1038/d41586-021-02129-x pmid: 34345046
33. J. Benn Torres, Anthropological perspectives on genomic data, genetic ancestry, and race. *Am. J. Phys. Anthropol.* **171** (suppl. 70), 74–86 (2020). doi: 10.1002/ajpa.23979; pmid: 31837009
34. A. M. Lans, Decolonize this collection: Integrating black feminism and art to re-examine human skeletal remains in museums. *Fem. Anthropol.* **2**, 130–142 (2021). doi: 10.1002/fea2.12027
35. R. M. Austin *et al*., Remembering St. Louis individual-structural violence and acute bacterial infections in a historical anatomical collection. *Commun. Biol.* **5**, 1050 (2022). doi: 10.1038/s42003-022-03890-z; pmid: 36192528
36. Detailed information is provided in the supplementary materials.
37. É. Harney *et al*., Ethical considerations when co-analyzing ancient DNA and data from private genetic databases. *Am. J. Human Genet.* (2023). doi: 10.1016/j.ajhg.2023.06.011
38. E. E. Baptist, "Cuffy," "fancy maids," and "one-eyed men": Rape, commodification, and the domestic slave trade in the United States. *Am. Hist. Rev.* **106**, 1619–1650 (2001). doi: 10.2307/2692741
39. K. Bryc, E. Y. Durand, J. M. Macpherson, D. Reich, J. L. Mountain, The genetic ancestry of African Americans, Latinos, and European Americans across the United States. *Am. J. Hum. Genet.* **96**, 37–53 (2015). doi: 10.1016/j.ajhg.2014.11.010; pmid: 25529636
40. R. Hui, E. D'Atanasio, L. M. Cassidy, C. L. Scheib, T. Kivisild, Evaluating genotype imputation pipeline for ultra-low coverage ancient genomes. *Sci. Rep.* **10**, 18542 (2020). doi: 10.1038/s41598-020-75387-w; pmid: 33122697
41. W. A. Freyman *et al*., Fast and robust identity-by-descent inference with the templated positional Burrows–Wheeler transform. *Mol. Biol. Evol.* **38**, 2131–2151 (2021). doi: 10.1093/molbev/msaa328; pmid: 33355662
42. E. Y. Durand *et al*., A scalable pipeline for local ancestry inference using tens of thousands of reference haplotypes. bioRxiv 2021.01.19.427308 [Preprint] (2021); https://doi.org/10.1101/2021.01.19.427308.
43. E. M. Jewett, K. F. McManus, W. A. Freyman, A. Auton; 23andMe Research Team, Bonsai: An efficient method for inferring large human pedigrees from genotype data. *Am. J. Hum. Genet.* **108**, 2052–2070 (2021). doi: 10.1016/j.ajhg.2021.09.013; pmid: 34739834
44. L. Pauling, H. A. Itano, S. J. Singer, I. C. Wells, Sickle cell anemia a molecular disease. *Science* **110**, 543–548 (1949). doi: 10.1126/science.110.2865.543; pmid: 15395398
45. E. Beutler, G6PD deficiency. *Blood* **84**, 3613–3636 (1994). doi: 10.1182/blood.V84.11.3613.bloodjournal84113613; pmid: 7949118
46. K. Esoh, A. Wonkam, Evolutionary history of sickle-cell mutation: Implications for global genetic medicine. *Hum. Mol. Genet.* **30**, R119–R128 (2021). doi: 10.1093/hmg/ddab004; pmid: 33461216
47. W. Jin *et al*., Genome-wide detection of natural selection in African Americans pre- and post-admixture. *Genome Res.* **22**, 519–527 (2012). doi: 10.1101/gr.124784.111; pmid: 22128132
48. G. Bhatia *et al*., Genome-wide scan of 29,141 African Americans finds no evidence of directional selection since admixture. *Am. J. Hum. Genet.* **95**, 437–444 (2014). doi: 10.1016/j.ajhg.2014.08.011; pmid: 25242497
49. C. B. Dew, David Ross and the Oxford Iron Works: A study of industrial slavery in the early nineteenth-century South. *William Mary Q.* **31**, 189–224 (1974). doi: 10.2307/1920909
50. C. B. Dew, *Bond of Iron: Master and Slave at Buffalo Forge* (Norton, 1995).
51. R. L. Lewis, *Coal, Iron, and Slaves: Industrial Slavery in Maryland and Virginia, 1715–1865* (Greenwood Press, 1979).
52. Moravian Missionary Letters – 1766-1787, translated by Edward Quinter (Moravian Archives, Bethlehem, Pennsylvania, 2021–2022).
53. Diaries of Moravian clergy – 1798-1850, translated by Edward Quinter (Moravian Archives, Bethlehem, Pennsylvania, 2021–2022).
54. S. P. Gordon, Fishing for a few: Moravians on the Eighteenth-Century Pennsylvania Frontier. *Pa. Hist.* **88**, 319–350 (2021). doi: 10.5325/pennhistory.88.3.0319
55. D. Eltis, "A Brief Overview of the Trans-Atlantic Slave Trade," Slave Voyages: The Trans-Atlantic Slave Trade Database; https://slavevoyages.org/.
56. J. K. Thornton, *A Cultural History of the Atlantic World, 1250–1820* (Cambridge Univ. Press, 2012).
57. L. M. Heywood, J. K. Thornton, *Central Africans, Atlantic Creoles, and the Foundation of the Americas, 1585–1660* (Cambridge Univ. Press, 2007).
58. S. E. G. Chapelle, J. H. Baker, D. R. Esslinger, W. H. Ridgway, J. B. Russo, C. B. Schulz, G. A. Stiverson, *Maryland: A History of Its People* (Johns Hopkins Univ. Press, 1986).
59. K. A. Sirak *et al*., A minimally-invasive method for sampling human petrous bones from the cranial base for ancient DNA analysis. *Biotechniques* **62**, 283–289 (2017). doi: 10.2144/000114558; pmid: 28625158
60. J. Dabney *et al*., Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 15758–15763 (2013). doi: 10.1073/pnas.1314445110; pmid: 24019490
61. P. Korlević *et al*., Reducing microbial and human contamination in DNA extractions from ancient bones and teeth. *Biotechniques* **59**, 87–93 (2015). doi: 10.2144/000114320; pmid: 26260087
62. N. Rohland, I. Glocke, A. Aximu-Petri, M. Meyer, Extraction of highly degraded DNA from ancient bones, teeth and sediments for high-throughput sequencing. *Nat. Protoc.* **13**, 2447–2461 (2018). doi: 10.1038/s41596-018-0050-5; pmid: 30323185
63. N. Rohland, E. Harney, S. Mallick, S. Nordenfelt, D. Reich, Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. *Philos. Trans. R. Soc. London Ser. B* **370**, 20130624 (2015). doi: 10.1098/rstb.2013.0624; pmid: 25487342
64. M.-T. Gansauge, A. Aximu-Petri, S. Nagel, M. Meyer, Manual and automated preparation of single-stranded DNA libraries for the sequencing of DNA from ancient biological remains and other sources of highly degraded DNA. *Nat. Protoc.* **15**, 2279–2300 (2020). doi: 10.1038/s41596-020-0338-0; pmid: 32612278
65. T. S. Korneliussen, A. Albrechtsen, R. Nielsen, ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* **15**, 356 (2014). doi: 10.1186/s12859-014-0356-4; pmid: 25420514
66. P. Skoglund, J. Storå, A. Götherström, M. Jakobsson, Accurate sex identification of ancient human remains using DNA shotgun sequencing. *J. Archaeol. Sci.* **40**, 4477–4482 (2013). doi: 10.1016/j.jas.2013.07.004
67. H. Weissensteiner *et al*., HaploGrep 2: Mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res.* **44**, W58–W63 (2016). doi: 10.1093/nar/gkw233; pmid: 27084951
68. The 1000 Genomes Project Consortium, A global reference for human genetic variation. *Nature* **526**, 68–74 (2015). doi: 10.1038/nature15393; pmid: 26432245
69. A. Bergström *et al*., Insights into human genetic variation and population history from 929 diverse genomes. *Science* **367**, eaay5012 (2020). doi: 10.1126/science.aay5012; pmid: 32193295
70. S. Mallick *et al*., The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* **538**, 201–206 (2016). doi: 10.1038/nature18964; pmid: 27654912
71. S. Mallick *et al*., The Allen Ancient DNA Resource (AADR): A curated compendium of ancient human genomes. bioRxiv 2023.04.06.535797 [Preprint] (2023); https://doi.org/10.1101/2023.04.06.535797.
72. S. Rubinacci, D. M. Ribeiro, R. J. Hofmeister, O. Delaneau, Efficient phasing and imputation of low-coverage sequencing data using large reference panels. *Nat. Genet.* **53**, 120–126 (2021). doi: 10.1038/s41588-020-00756-0; pmid: 33414550
73. P.-R. Loh, P. F. Palamara, A. L. Price, Fast and accurate long-range phasing in a UK Biobank cohort. *Nat. Genet.* **48**, 811–816 (2016). doi: 10.1038/ng.3571; pmid: 27270109

ACKNOWLEDGMENTS

We acknowledge the individuals from Catoctin Furnace whose DNA we studied and recognize that they could not directly consent to participate in this study. We hope to honor the people of Catoctin and their contributions by telling their story and highlighting their ongoing legacy in the form of descendant communities. The genetic research was authorized by the Smithsonian Institution's National Museum of Natural History Department of Anthropology Collections Advisory Committee. For nearly 50 years, the CFHS has been at the forefront of stewardship, research, and interpretation of the African American cemetery located in Catoctin village. The CFHS has been a driving force in the effort to link the remains of deceased individuals who are yet to be fully identified to a descendant community, with the ultimate goal of using genealogical and historical research as well as DNA to establish definitive genealogies. Inquiries from those interested in learning more about the history of Catoctin Furnace, and from possible descendants interested in exploring their connections to the people who worked at the site, are welcome to contact CFHS at https://catoctinfurnace.org/. We thank the CFHS; the AARCH Society of Frederick, Maryland; and members of descendant families for support of this research. We thank the 23andMe research participants and others who consented to participate in research for making this study possible. **Funding:** The aDNA analysis was funded by NIH grant HG012287, by John Templeton Foundation grant 61220, and by the Howard Hughes Medical Institute. The pedigree analysis was funded in part by NIH grant R35GM133805. In 2016, a $14,000 Maryland Heritage Areas Authority grant was awarded to CFHS to fund forensic research for the African American cemetery. The Rice Family Endowment for Forensic Anthropology provided critical support for Smithsonian Skeletal Biology Program contributions. **Author contributions:** D.R., J.L.M., and D.W.O. conceived of and supervised the study. K.S.B., K.G.B., and D.W.O. completed the osteological analysis spearheaded by E.C. N.A., R.B., N.B., M.F., L.Q., K.St., J.N.W., F.Z., and N.R. generated the data via wet laboratory analyses. A.A., S.Ma., A.M., and M.M. generated the data via bioinformatic analyses. Z.Z. and the 23andMe Research Team provided additional computational support. É.H., S.Mi., E.J., and I.O. analyzed the genetic data. É.H., S.Mi., K.S.B., W.A.F., K.B., E.J., E.C., H.L.G., L.H., J.T., R.C., S.A.E., K.G.B., J.S., K.Si., I.O., J.L.M., D.W.O., and D.R. contributed to the interpretation of results. É.H. wrote the manuscript, with substantial input from S.Mi., K.S.B., W.A.F., K.B., A.A., E.J., E.C., H.L.G., L.H., J.T., R.C., S.A.E., K.G.B., J.S., K.Si., J.L.M., D.W.O., and D.R. and feedback from all coauthors. **Competing interests:** É.H., S.Mi., W.A.F., K.B., E.J., H.L.G., S.A.E., and J.L.M. and members of the 23andMe Research Team are employees of and hold stock, stock options, or both in 23andMe. The remaining authors declare no competing interests.

Data and materials availability: Unaligned and aligned sequences for the 27 Catoctin individuals were originally reported in Harney *et al*. (*13*) and are available from the European Nucleotide Archive under accession number PRJEB52230. Genotype files for pseudo-haploid and imputed versions of the dataset are available at https://reich.hms.harvard.edu/datasets. There are restrictions to the availability of 23andMe genotype data owing to 23andMe consent and privacy guidelines. 23andMe agrees that the publication coauthors will rerun the comparison of historical genomic data against customer genetic data upon request by other academic and nonprofit researchers on reasonable terms to enable the results of the research activities to be replicated for at least 7 years after publication or for as long as the coauthors are employed by, or otherwise affiliated with, 23andMe in a capacity that allows them to rerun the analysis. Wherever possible, supplementary tables were included that report the summary statistics that were used to create figures that involved 23andMe datasets. Unless another comparable anonymizing approach was specified, these summary statistics were generated with the requirement that, in all reported results, any 23andMe research participant must be indistinguishable from at least four other research participants included in the dataset. When meaningful, we repeated analyses performed on the 23andMe dataset using only the 1000 Genomes Project and/or African American imputation panel datasets and reported these results.

Members of the 23andMe Research Team
Stella Aslibekyan, Adam Auton, Elizabeth Babalola, Robert K. Bell, Jessica Bielenberg, Emily Bullis, Daniella Coker, Gabriel Cuellar Partida, Devika Dhamija, Sayantan Das, Sarah L. Elson, Teresa Filshtein, Kipper Fletez-Brant, Pierre Fontanillas, Karl Heilbron, Barry Hicks, David A. Hinds, Yunxuan Jiang, Katelyn Kukar, Keng-Han Lin, Maya Lowe, Jey McCreight, Matthew H. McIntyre, Meghan E. Moreno, Priyanka Nandakumar, Elizabeth S. Noblin, Jared O'Connell, Aaron A. Petrakovitz, G. David Poznik, Morgan Schumacher, Anjali J. Shastri, Janie F. Shelton, Jingchunzi Shi, Suyash Shringarpure, Vinh Tran, Joyce Y. Tung, Xin Wang, Wei Wang, Catherine H. Weldon, Peter Wilton, Alejandro Hernandez, Corinna D. Wong, Christophe Toukam Tchakouté, Alison Fitch, Alexandra Reynoso, Julie M. Granka, Qiaojuan Jane Su, Alan Kwong, Nicholas Eriksson, Dominique T. Nguyen, ianca A. Llamas, Susana A. Tat

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.ade4995
Materials and Methods
Supplementary Text
Figs. S1 to S16
Tables S1 to S24, S1.1, S3.2 to S3.7, and S4.1
References (*74*–*124*)
MDAR Reproducibility Checklist

Submitted 22 August 2022; accepted 20 June 2023
10.1126/science.ade4995

The genetic legacy of African Americans from Catoctin Furnace

adaoin Harney, Steven Micheletti, Karin S. Bruwelheide, William A. Freyman, Katarzyna Bryc, Ali Akbari, Ethan Jewett, Elizabeth Comer, Henry Louis Gates, Jr., Linda Heywood, John Thornton, Roslyn Curry, Samantha Ancona Esselmann, Kathryn G. Barca, Jakob Sedig, Kendra Sirak, Iigo Olalde, Nicole Adamski, Rebecca Bernardos, Nasreen Broomandkhoshbacht, Matthew Ferry, Lijun Qiu, Kristin Stewardson, J. Noah Workman, Fatma Zalzala, Shop Mallick, Adam Micco, Matthew Mah, Zhao Zhang, 23andMe Research Team, Nadin Rohland, Joanna L. Mountain, Douglas W. Owsley, David Reich, Stella Aslibekyan, Adam Auton, Elizabeth Babalola, Robert K. Bell, Jessica Bielenberg, Emily Bullis, Daniella Coker, Gabriel Cuellar Partida, Devika Dhamija, Sayantan Das, Sarah L. Elson, Teresa Filshtein, Kipper Fletez-Brant, Pierre Fontanillas, Karl Heilbron, Barry Hicks, David A. Hinds, Yunxuan Jiang, Katelyn Kukar, Keng-Han Lin, Maya Lowe, Jey McCreight, Matthew H. McIntyre, Meghan E. Moreno, Priyanka Nandakumar, Elizabeth S. Noblin, Jared OConnell, Aaron A. Petrakovitz, G. David Poznik, Morgan Schumacher, Anjali J. Shastri, Janie F. Shelton, Jingchunzi Shi, Suyash Shringarpure, Vinh Tran, Joyce Y. Tung, Xin Wang, Wei Wang, Catherine H. Weldon, Peter Wilton, Alejandro Hernandez, Corinna D. Wong, Christophe Toukam Tchakout, Alison Fitch, Alexandra Reynoso, Julie M. Granka, Qiaojuan Jane Su, Alan Kwong, Nicholas Eriksson, Dominique T. Nguyen, Bianca A. Llamas, and Susana A. Tat

Science, **381** (6657), eade4995.
DOI: 10.1126/science.ade4995

Editor's summary

Catoctin Furnace operated in Maryland in the 18th and 19th centuries using both free and enslaved African American labor. Harney *et al*. analyzed DNA from 27 individuals from an African American cemetery that was excavated 40 years ago during highway construction (see the Perspective by Jackson). The authors found genetic evidence of biological family groups, modern-day African populations with whom they may have shared ancestry, and even possible distant relatives in the United States through identity-by-descent comparisons with consenting customers of 23andMe. This study demonstrates that when studied responsibly with input from stakeholders, long-buried DNA can be used to uncover obfuscated or forgotten histories of marginalized individuals. —Corinne Simonti