

Current Biology

The genetic legacy of the 17th-century colonial capital of St. Mary's City

Highlights

- A novel IBD-based framework for identifying anonymous historical remains
- Analysis of 49 St. Mary's colonists reveals 1.3 million living genetic relatives
- Tracing a founder population's origins and descendant migrations across the US
- Possible identification of Maryland's second governor, Thomas Greene, and his family

Authors

Éadaoin Harney, Ethan Jewett, Steven Micheletti, ..., Joanna L. Mountain, Douglas W. Owsley, David Reich

Correspondence

eadaoinh@23andme.com (É.H.),
owsleyd@si.edu (D.W.O.),
reich@genetics.med.harvard.edu (D.R.)

In brief

Harney et al. analyze 49 genomes from 17th-century St. Mary's City, Maryland. Using an IBD-based approach, they identify connections to over 1.3 million living relatives, enabling them to study the founders' ancestral origins, trace descendant migrations across the US, and propose a possible identity for Maryland's second governor.

Article

The genetic legacy of the 17th-century colonial capital of St. Mary's City

Éadaoin Harney,^{1,2,12,*} Ethan Jewett,¹ Steven Micheletti,¹ Roslyn Curry,^{1,2} Karin S. Bruwelheide,³ William A. Freyman,¹ Henry Miller,⁴ Ali Akbari,^{2,5} Kathryn Barca,³ Katarzyna Bryc,¹ Shaeloren Deering,¹ Samantha Ancona Esselmann,¹ Kira Kalkus,¹ Aisling Kearns,⁵ Alexander Moran,¹ Dominique T. Nguyen,¹ Iñigo Olalde,^{2,6,7} Jakob Sedig,^{2,8} Kendra Sirak,^{2,5} Kim Callan,^{5,9} Lora Iliev,^{5,9} Lijun Qiu,^{5,9} J. Noah Workman,^{5,9} Matthew Mah,^{5,9,10} Gregory Soos,^{5,9} 23andMe Research Team,¹ Swapan Mallick,^{5,9,10} Nadin Rohland,^{2,5,10} Joanna L. Mountain,^{1,11} Douglas W. Owsley,^{3,11,*} and David Reich^{2,5,9,10,11,*}

¹23andMe Research Institute, 2555 Park Boulevard, Palo Alto, CA 94306, USA

²Department of Human Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA

³Department of Anthropology, National Museum of Natural History, Smithsonian Institution, Washington, DC 20560, USA

⁴Historic St. Mary's City, St. Mary's City, MD 20686, USA

⁵Department of Genetics, Harvard Medical School, Boston, MA 02115, USA

⁶Department of Zoology and Animal Cell Biology, University of the Basque Country UPV/EHU, 01006 Vitoria-Gasteiz, Spain

⁷Ikerbasque–Basque Foundation of Science, 48009 Bilbao, Spain

⁸Chronicle Heritage, Phoenix, AZ 85004, USA

⁹Howard Hughes Medical Institute, Harvard Medical School, Boston, MA 02115, USA

¹⁰Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA

¹¹Senior author

¹²Lead contact

*Correspondence: eadaoinh@23andme.com (É.H.), owsleyd@si.edu (D.W.O.), reich@genetics.med.harvard.edu (D.R.)

<https://doi.org/10.1016/j.cub.2026.04.046>

SUMMARY

Founded in 1634, St. Mary's City was the first English settlement in the colony of Maryland. Despite existing written records and the ability of many present-day Americans to trace their ancestry to the historic city, substantial gaps remain in our knowledge of this early founder population. To address these gaps, we analyzed the genomes of 49 individuals from 17th-century St. Mary's City to trace their genetic ancestry, examine their enduring legacy, and demonstrate the efficacy of using an identity-by-descent (IBD) approach to link historical individuals to the present. In our analysis, we identified over 1.3 million genetic relatives of the St. Mary's individuals among research participants in the 23andMe Research Institute's database. We found high rates of genetic sharing with participants from western England and Wales, suggesting a likely place of origin for many of the colonial city's earliest inhabitants. Additionally, we observed strong genetic connections with participants from Kentucky, mirroring a recorded post-Revolutionary War migration of Maryland Catholics to that region. By further integrating genealogical information from present-day research participants who share the closest genetic connections to the St. Mary's individuals, we propose possible identities for three sequenced historical St. Mary's City residents, including Thomas Greene, the second governor of the colony of Maryland. This unique case study highlights the power of genetics to restore lost identities and reconstruct historical relationships by tracing geographic signals of ancestry.

INTRODUCTION

Early colonial founder populations have had a lasting impact on the genetic landscape of North America, contributing disproportionately to the biogeographic ancestry of present-day people in the United States (US) and elsewhere across the continent.^{1,2} While historical documents and genetic studies of present-day populations offer partial insights into these early groups, ancient DNA (aDNA) now enables direct investigation of the genomes of historical individuals who made up these founding populations, enabling us to address questions about their biogeographic origins, relationships, and the genetic legacy they left in present-day populations.³

Identical-by-descent (IBD) segments, which are regions of DNA shared between two individuals because they were inherited from a common ancestor, serve as a key tool for studying historical populations.^{4,5} Long IBD segments are evidence of recent genealogical relationships, while shorter segments signal distant shared ancestry.

Here we extend the aDNA IBD framework introduced in Harney et al.³ to study the 17th-century inhabitants of St. Mary's City, the founding English settlement in the colony of Maryland (Figures 1A–1C). Founded in 1634, the settlement drew English and some Irish immigrants of diverse religious backgrounds.⁶ Despite existing primary source documents and the ability of present-day Americans to trace their

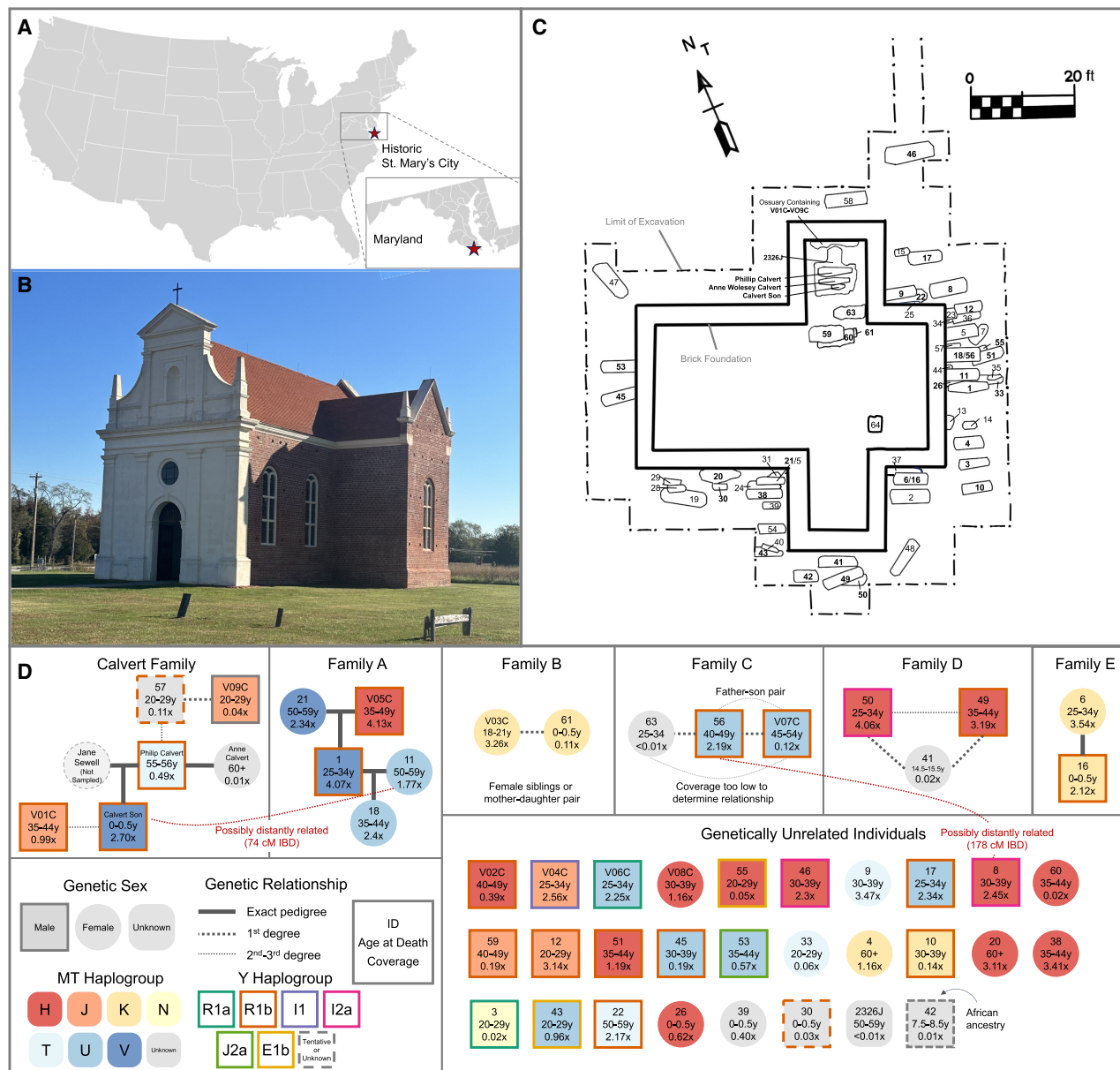


Figure 1. Location, genetic kinship, and haplogroups of St. Mary's individuals

(A) Map showing the location of Historic St. Mary's City in the US, with a panel showing its location within Maryland.

(B) Photo of the reconstructed Brick Chapel at Historic St. Mary's City.

(C) Site map showing the location of excavated burials within and near the Brick Chapel. Burials sampled for DNA are indicated in bold.

(D) Family groupings were identified among the St. Mary's individuals through genetic connections, except in the case of Anne Wolseley Calvert, who is known from historical records to have been the first wife of Philip Calvert and whose biological profile matches the remains buried at his side. Individuals are labeled according to burial ID, estimated age at death, and average chromosomal coverage. Circular and square markers are used to represent genetically female and male individuals, respectively. The fill color of each marker corresponds to the MT haplogroup, and for genetically male individuals, the outline color corresponds to the Y chromosome haplogroup. Genetic relationships shared between individuals are indicated by connector lines. A solid dark gray line indicates the exact pedigree, with first- and second- to third-degree relatives connected by a thick and thin dotted gray line, respectively. More distant relationships detected via IBD sharing that span family groups are connected by a dotted red line.

See also [Data S1](#), [S2](#), [S7A](#), and [S7B](#).

ancestry to St. Mary's City, substantial gaps remain in our knowledge of this early founder population, including aspects of the identity of those buried in the Chapel Field cemetery ([Data S1](#)).

To date, only three individuals interred in the Chapel Field cemetery have been identified with confidence: Philip Calvert; his first wife, Anne Wolseley Calvert; and Philip's infant son with his second wife, Jane Sewell.^{7,8} Even for the prominent

Calvert family, the burial locations of many of its members are unrecorded, including Philip's half-brother, Leonard Calvert (1610–1647), who served as the first governor of Maryland.⁹ Even less is recorded for the remainder of the historic city's ancestral families. Thus, despite its relatively rich historical record, St. Mary's City presents unresolved questions ideally suited to genetic investigation, particularly regarding origins, kinship, and identity.

The ability of aDNA to provide insights into the biogeographic origins of historical individuals and the relationships they share with one another is well established.¹⁰ However, whether genetic data can be used to restore the identities of individuals who have no surviving names in the documentary record is less well explored. Previous aDNA studies that made genealogical connections between historical and living people relied on strong prior hypotheses based on archeological or archival evidence and often could only be applied to questions of very recent descent.^{11,12} By contrast, the IBD-based approach we introduce in this study does not require a candidate identity nor is it limited to the closest genealogical connections. Instead, by searching for IBD between historical individuals and millions of living research participants, we can identify multiple present-day relatives who are genetically connected (likely by 9 or more degrees) to the same historical individuals. Integrating information gathered from their genealogical pedigrees, we use a triangulation approach to converge on possible identities for previously unnamed historical individuals without any initial hypothesis—a capability that significantly expands what genetic analyses can reveal about the past.

We generated genome-wide data from 49 individuals excavated from the Chapel Field cemetery at St. Mary's City and examined IBD connections to 11,524,442 genotyped research participants from 23andMe Research Institute, including more than 14,000 with ancestral ties to the settlement. These comparisons enable us to characterize the biogeographic origins of the St. Mary's individuals, reconstruct family relationships within the cemetery, document their genetic legacy among present-day Americans, and present a case study that proposes possible identities for individuals whose names were otherwise unknown.

These results show how documented historical migrations are reflected in the genomes of present-day Americans and demonstrate that IBD analysis can add to our understanding of even well-studied historical populations. Our analysis of St. Mary's City thus serves as a test case for applying this approach to other, less well-understood sites, offering a framework for reconstructing population history, genealogy, and identity in the past.

RESULTS

Genetic relationships observed among the St. Mary's individuals

From a total of 65 individuals recovered from single interments and a commingled ossuary associated with the Brick Chapel at St. Mary's (c.1636–1730), we selected 50 individuals for genetic analysis. We generated genome-wide data for 49 of these individuals at 1.24 million variable DNA positions, with a median per-individual average chromosomal coverage of 1.16 \times (range: <0.01 \times –4.13 \times) (Data S7A). We identified six genetic families by comparing pairwise mismatch rates in pseudohaploid

genotypes between individuals (Figure 1D). Five genetic families were identified exclusively using genetic connections, while the sixth was identified via a mixture of historical records, osteological analysis, and genetic connections. We use the term genetic family to refer to groups of individuals who share a close genetic relationship. Similarly, we use specific relationship terms (e.g., mother, father, son, and daughter) in the biological sense, acknowledging that these genetic relationships may not directly align with the kin-based relationships or gender identities recognized by the St. Mary's individuals.

Family A is the largest newly identified genetic family, consisting of a genetically female adult, her parents, and her paternal grandparents. Families B–E are each composed of two or three individuals who share first-degree relationships, but the nature of most of these relationships is less understood. For instance, based on the available sequencing coverage, we could not determine whether family B consists of a mother-daughter pair or sisters.

The Calvert family includes three individuals of known identity: Philip Calvert, his first wife Anne Wolseley Calvert, and Philip's infant son, referred to as “the Calvert Son.” We identified three other individuals with genetic ties to Philip Calvert or the Calvert Son. V01C, represented only by a cranial vault interred within the ossuary, shares a second- to third-degree relationship with the Calvert Son, but we detected no relationship with Philip Calvert, suggesting that the two individuals are connected via the infant's mother, Jane Sewell. Burial 57, an individual whose burial contained only lower limb bones (a result of the disturbance of the grave during construction of the Chapel foundation), is a second- to third-degree relative of Philip Calvert and shares a first-degree relationship with V09C, another individual whose cranial vault was interred in the ossuary. While no genetic connection was detected between Philip Calvert and V09C, we cannot exclude the possibility that they were related, given the poor preservation and elevated rate of contamination detected in V09C. In fact, this elevated contamination rate makes it impossible to rule out that V09C and burial 57 are the same individual. We discuss their possible relationship to Philip Calvert further in Data S2.

For the 25 historical individuals with sufficient DNA preservation to enable IBD analysis (i.e., at least 1 \times average chromosomal coverage), the IBD sharing between the individuals supports the inferred relationships (Figure 2; Data S7B). Using IBD sharing patterns, we clarified the relationship between burials 6 and 16 of family E, who we conclude were mother and son. Our IBD analysis also suggested that some of the individuals had more distant relationships to one another. For instance, burials 11 and 18—the mother-daughter pair from family A—share up to 100 centimorgans (cM) of IBD with the Calvert Son, suggesting that they were distant relatives. Similarly, we detected 178 cM of IBD between burial 56 of family C and burial 8.

The genetic ancestry of the St. Mary's individuals

We found that the St. Mary's individuals in our study have entirely European-related ancestry, except for burial 42, who has primarily African-related ancestry, with a smaller proportion (25%–30%) of European-related ancestry (Figure S1). This young boy, who was wrapped in a shroud and buried in a gable-lidded coffin, had poor skeletal and DNA preservation and therefore could not be included in subsequent analyses.

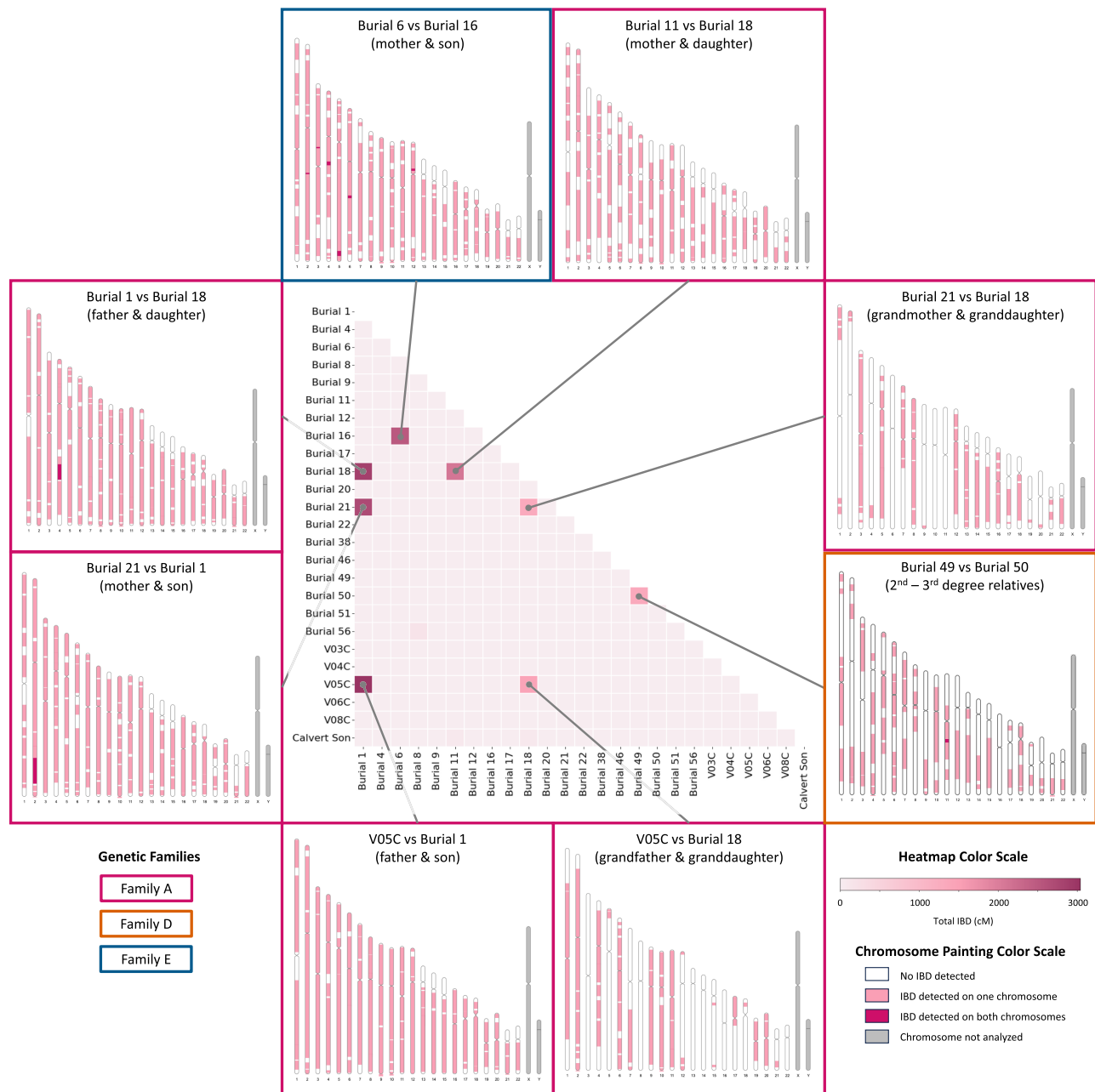


Figure 2. IBD sharing between the St. Mary's individuals

A heatmap showing total IBD shared between all pairs of St. Mary's individuals with at least $1\times$ average chromosomal coverage. Karyotype plots showing the location of IBD segments detected between pairs who share the greatest amount of IBD surround the heatmap. Light pink indicates genomic regions where a single IBD segment was detected, and darker pink marks regions with two overlapping segments. Chromosomes X and Y are shown in gray to indicate that they were not included in the IBD analyses. We note that consecutive short IBD segments within a chromosome likely result from phase switch errors and that the true IBD segments probably span these closely located regions. See also [Data S7B](#).

To explore the biogeographic ancestry of the 25 St. Mary's individuals of European descent whose genetic data could be confidently imputed (i.e., those with $>1\times$ average chromosomal coverage), we searched for IBD connections to 11,524,442 23andMe research participants using the approach established in Harney et al.³ We found that 11.85% of all research participants

share IBD with one or more of the St. Mary's individuals, with up to 150 cM shared with a single St. Mary's individual: burial 1, an American-born, young adult male and member of family A ([Table 1](#); [Figure S2](#); [Data S7C](#) and [S7D](#)). The highest rates of IBD sharing occur among participants from regions where British and Irish ancestries are well represented ([Figures S3](#) and [S4](#); [Data](#)

Table 1. IBD sharing between St. Mary's individuals and the 23andMe participant panel

23andMe participants included in group	No. of 23andMe participants in group	Median total IBD in 23andMe participants with IBD detected (cM) ^a	Maximum total IBD in 23andMe participants with IBD detected (cM) ^a	Percentage of participants that share IBD with St. Mary's individual(s)	Number of participants with any shared IBD ^b	Number of participants with at least 30 cM IBD ^b
All participants	11,524,442	7.3	150	11.85%	1,366,105	9,000
Participants in GB cohort	144,689	7.3	65	24.80%	35,885	60
Participants in IE cohort	11,285	7.6	70	23.70%	2,675	25
Participants in US cohort	2,584,518	7.3	140	15.48%	400,070	2,790
Participants in US cohort with over 99% European ancestry	1,686,066	7.3	140	17.22%	290,345	2,030
Participants in US cohort who answered survey question: any answer	620,260	7.3	140	16.10%	99,890	780
Participants in US cohort who answered survey question: yes	14,814	7.5	140	18.54%	2,745	50
Participants in US cohort who answered survey question: no	342,592	7.3	110	14.85%	50,885	380
Participants in US cohort who answered survey question: not sure	262,854	7.4	120	17.6%	46,255	350

Cohort membership was determined based on grandparent birth location or, in cases where all four grandparents were born in different countries or this information was not available, based on participant birth location. We use the ISO2 country code abbreviations to refer to each country (i.e., GB, Great Britain; IE, Ireland; US, United States). The survey question referred to in the final four rows was “can you trace your relatives back to 17th-century St. Mary's City, Maryland's first colonial settlement?” (possible answers: “yes,” “no,” and “I'm not sure”). To protect the privacy of 23andMe research participants, the following rounding strategies were applied.

^aValues are rounded based on 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

^bCounts are rounded to the nearest 5 (see also [Data S3](#) and [S7C](#))

[S7E](#)), including northwestern Europe, the US, South Africa, and Australia, reflecting distant shared ancestry.

The 25 St. Mary's individuals included in IBD analyses exhibit a high degree of variability in coverage (1.16–4.13 \times). Individuals with less than 2 \times average chromosomal coverage tended to share the fewest IBD connections—an expected consequence of the higher minimum IBD segment length threshold (9 cM) applied to individuals with <2 \times coverage (vs. 6 cM for individuals with 2–5 \times coverage). In [Data S3](#), we explore how variability in coverage impacts IBD sharing rates for historical individuals that were subject to the same segment length filtering thresholds by downsampling higher coverage individuals to 2 \times coverage. We found that although the IBD sharing rate increases with higher coverage, coverage is not the sole driver of differences in IBD sharing patterns observed across historic-era individuals. Instead, differences also appear to reflect real differences in their genetic legacy ([Figure S5](#)).

Genetic connections in Great Britain and Ireland

Within Europe, the St. Mary's individuals exhibit a high rate of sharing with participants from Great Britain and Ireland (26.2%). Using randomization testing, we confirmed that we would be unlikely to observe this frequency of sharing within an identically sized sample of participants randomly selected from across Europe with at least 99% European ancestry (among whom the overall rate of sharing is 13.92%). This pattern,

therefore, likely reflects a historically significant connection to the region ($p < 0.001$) ([Data S7F](#)).

When all 25 St. Mary's individuals were considered together, we observed the highest rates of sharing with participants from Wales and western England, particularly the Western Midlands ([Figures 3A and S5](#); [Data S7G](#)). Comparing the St. Mary's individuals to genetic clusters composed of participants from Great Britain and Ireland corresponding with geography ([Figure 3B](#)), we found that multiple St. Mary's individuals share IBD with cluster 16, a group composed of participants with ties to Wales and surrounding areas in England. Applying the same randomization framework, we found that the frequency of IBD sharing with participants who make up this cluster (51.84%) was far greater than expected given the overall sharing rates observed among participants with at least 99% Northern European ancestry randomly sampled from across Great Britain and Ireland (26.77%; $p < 0.001$). Collectively, these results suggest that the sequenced St. Mary's individuals had ancestral ties to Wales and western England.

The strongest connections to the Welsh genetic cluster (cluster 16) were observed among burials 49, 50, 56, and V03C ([Figures 3B, 3C, and S5](#); [Data S7H](#)). V06C also shares substantial IBD with participants from cluster 16 but with even stronger connections to clusters 14 and 17, composed of participants from the Lancashire and Midlands regions of England,

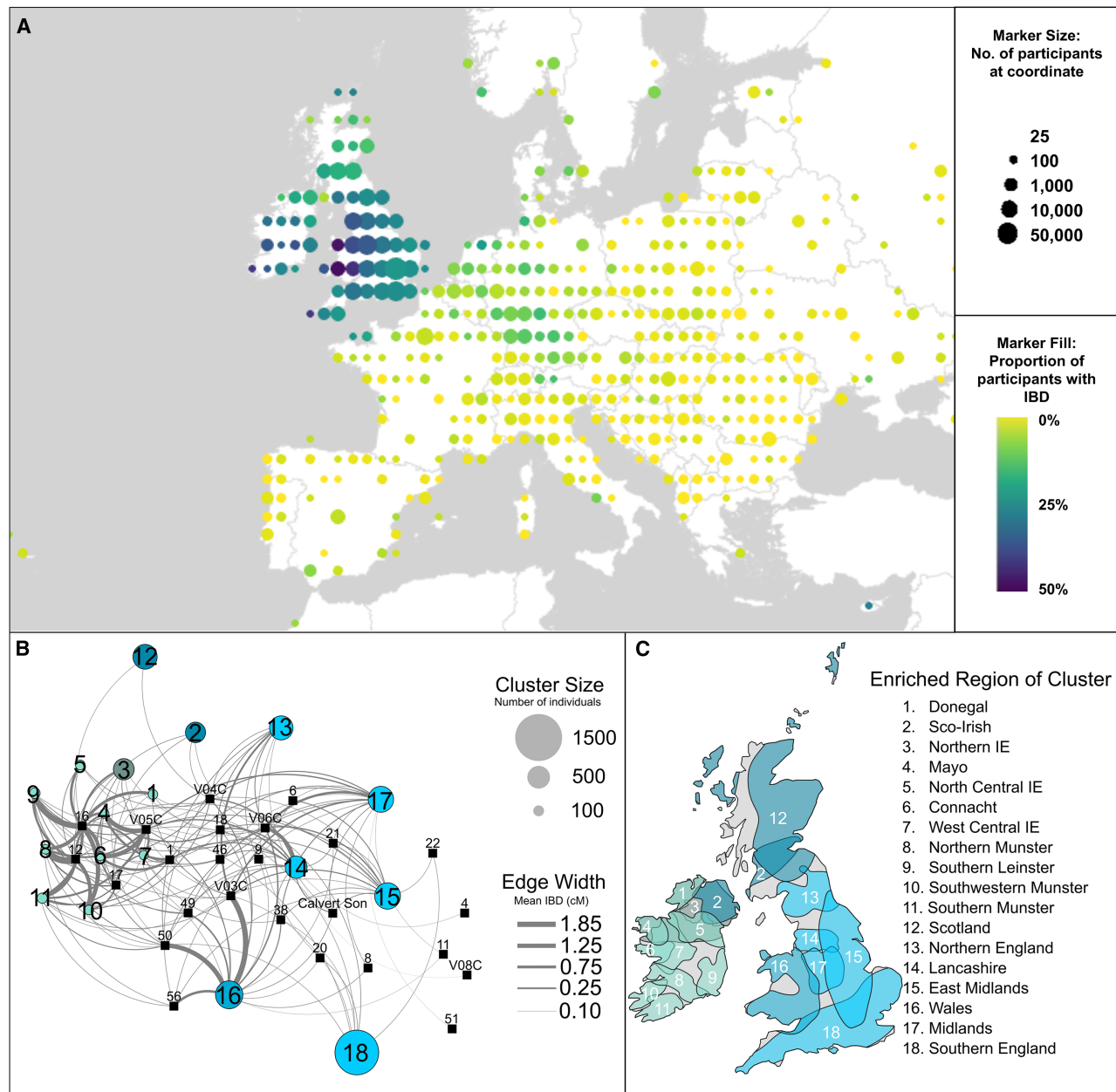


Figure 3. Genetic connections to the St. Mary's individuals among research participants from Great Britain and Ireland

(A) A map of Europe in which marker color indicates the proportion of research participants with at least 99% European ancestry at each geographic coordinate who share IBD with one or more of the St. Mary's individuals. The size of each marker represents the number of participants at the given geographic coordinate (rounded to the nearest integer). To protect participant privacy, we randomly downsampled to include only 80% of participants and only showed results for coordinates with at least 25 associated participants.

(B) IBD network demonstrating St. Mary's individuals' connections to genetic groups in Great Britain and Ireland ($n = 7,872$). Each circle represents a genetic group that was identified using stochastic block modeling¹³ on IBD connections between individuals that have all 4 of their grandparents born in either Great Britain or Ireland. The size of each genetic group is scaled by the total number of individuals assigned to it (ranging from 88 to 1,457). Genetic groups are arranged by the average pairwise IBD sharing between clusters (edges not shown) using a Force Atlas graph layout. St. Mary's individuals, displayed as squares, are projected over the graph network and arranged based on the average IBD that is shared with each cluster (shown as lines; edges smaller than 0.10 cM are not shown).

(C) The geographic distribution of genetic clusters in Great Britain and Ireland. Ranges represent the limits of kernel density estimates based on coordinates of self-reported grandparent locations from individuals in each cluster.

See also [Data S7G](#) and [S7H](#) and [Figures S4](#) and [S5](#).

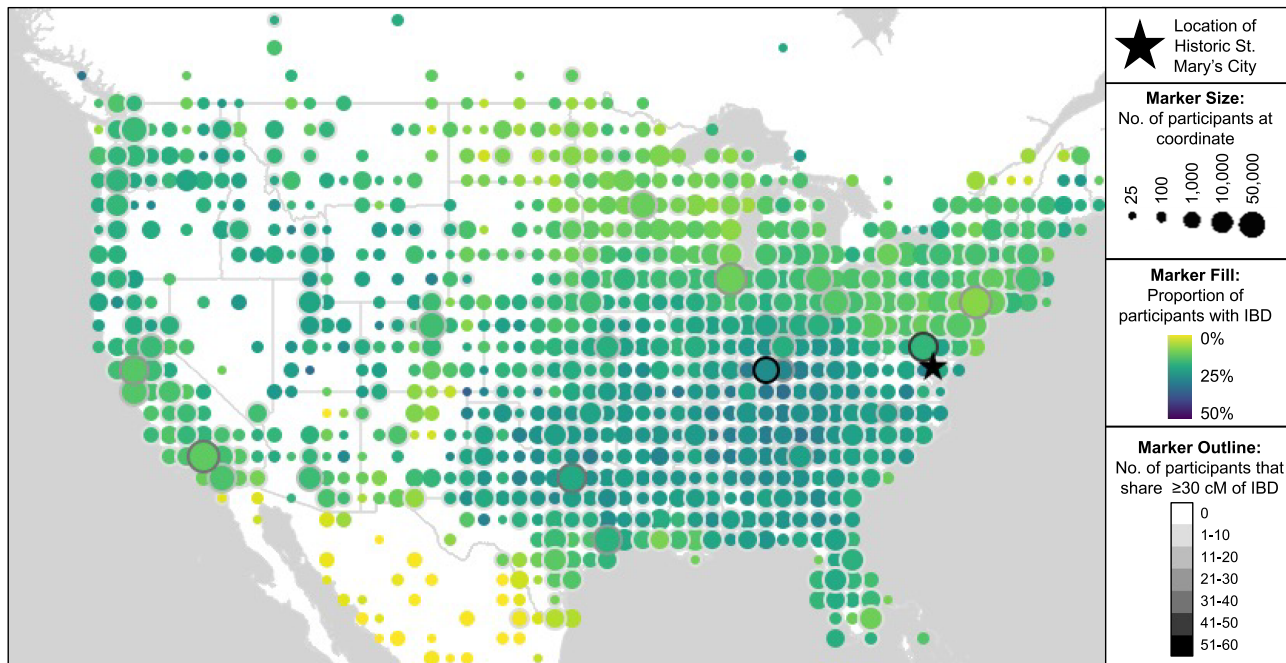


Figure 4. Genetic connections to the St. Mary's individuals among research participants in the US

Marker color indicates the proportion of research participants in the US at each geographic coordinate who share IBD with one or more of the St. Mary's individuals. The size of each marker represents the number of participants at the given geographic coordinate (rounded to the nearest integer). To protect participant privacy, we randomly downsampled to include only 80% of participants and only show results for coordinates with at least 25 associated participants. Marker outline colors indicate the number of participants at each location who share at least 30 cM of IBD with one or more St. Mary's individuals. The star indicates the location of St. Mary's City, Maryland. See also [Data S7E](#) and [Figure S5](#).

respectively. By contrast, five individuals (burials 1, 12, 16, 17, and V05C) show greater sharing with genetic clusters in Ireland than Great Britain, suggesting that this is the source of their ancestry.

Genetic connections in the US

The St. Mary's individuals share IBD with 15.48% of participants in the US ([Table 1](#)), particularly those in southern states ([Figures 4 and S5](#); [Data S7F](#)). We confirmed via randomization testing that there is a significantly higher rate of sharing among participants with 99% European ancestry in southern states than in other US regions (20.1% vs. 17.22%; $p < 0.001$), suggesting that this finding may reflect settlement patterns of individuals of European ancestry in the early US ([Data S7G](#)). This signal is broadly consistent across all St. Mary's individuals, although the intensity of this pattern varies among individuals. Differences in coverage only partially explain this, with the lowest coverage individuals exhibiting the lowest rates of sharing ([Figure S5](#)).

Close relatives

We next studied participants with the strongest genetic connections to the St. Mary's individuals by focusing on pairs of historical individuals and research participants who shared at least 30 cM of IBD (referred to as "close relatives"). Given that most of the St. Mary's individuals died during the 17th century, the most likely predicted relationship between pairs who share 30 cM of IBD on up to 3 segments is less than 17 degrees. Due to the random way in which DNA is inherited, there is a large amount of variation

associated with this prediction, meaning that it cannot be assumed that all participants who share 30 cM of IBD with the St. Mary's individuals are direct descendants. Instead, they could be collateral relatives whose direct ancestors were closely related to the St. Mary's individuals, with a shared common ancestor within a few generations. In fact, as we expect that the St. Mary's individuals have significantly more collateral relatives living today than direct descendants, our approach for identifying genetic relatives may be subject to the "winner's curse," whereby we detect an overabundance of distant relatives who share unusually large amounts of DNA from their distant historical relatives compared with the number of true direct descendants. In [Data S4](#), we discuss IBD-based relationship inference in more detail.

We detected 81 close relatives with ties to Maryland, suggesting that descendants and close collateral relatives of the St. Mary's individuals remain in the region ([Figure 4](#); [Data S7I](#)). However, the largest number of close relatives report ties to Kentucky ($n = 211$), particularly in areas surrounding Louisville, plausibly reflecting a documented migration of families from St. Mary's City and the surrounding area to Kentucky between 1780 and 1820.^{14–17} This signal is particularly evident among connections to burial 56 (part of family C). Applying the randomization testing framework to consider the distribution of close relatives for both Maryland and Kentucky, we would be unlikely to identify as many close relatives among an identically sized, random sample of participants selected from across the US (where we observe a

total of ~2,030 close relatives) after restricting our analysis to participants with at least 99% European ancestry (Data S7F), providing further support for the historical significance of this IBD sharing pattern.

Ancestral ties to 17th-century Historic St. Mary's City

Many people in the US who can trace their ancestry to British colonists report ancestral ties to people who lived in, or were otherwise associated with, 17th-century St. Mary's City. To determine whether people with known ties to the historic city share IBD with the St. Mary's individuals at an elevated rate, we asked research participants to answer the question, "Can you trace your relatives back to 17th-century St. Mary's City, Maryland's first colonial settlement?" (possible answers: "yes," "no," and "I'm not sure"). Of the 620,260 research participants in the US cohort who responded to this question, 14,814 participants answered "yes" (Table 1).

A significantly higher proportion of research participants in the US cohort who answered "yes" to this question share IBD with the St. Mary's individuals than those who answered "no" (18.54% vs. 14.85%) (chi-squared test, $p = 1.62 \times 10^{-34}$, Data S7J). To ensure that this difference was not solely caused by differences in ancestry among respondents of this question, we restricted our analysis to only participants in the US cohort with at least 99% European ancestry, and we again found that "yes" respondents share IBD with the St. Mary's individuals at a significantly ($p = 1.71 \times 10^{-15}$) higher rate than "no" respondents (18.95% vs. 16.12%; Data S7J). These results suggest that the sharing we detect among research participants reflects recent shared ancestry in many cases. A similar pattern appears when considering mean IBD sharing between cohorts (2.33 vs. 1.77 cM; Data S7J). This is expected, as most segments lie near our minimum length threshold, causing patterns in mean sharing to closely track overall sharing rates.

Restoring identity

Next, we explored whether participants' genealogical connections to 17th-century St. Mary's City could be used to identify any of the sequenced St. Mary's individuals (Data S5). We collected self-reported genealogical pedigree information from consenting research participants with some of the strongest genetic connections to the St. Mary's individuals. This work was conducted under a custom research protocol that permitted the analysis of individual-level ancestral connections.

We centered our analysis on burial 56, searching for overlaps in the genealogies of two participants (1 and 2) who shared large amounts of IBD with this individual (48 and 33 cM, respectively). We hypothesized that males appearing in both trees could be candidates for burial 56. We identified three ancestral couples that appeared in both trees (Figures 5A and 5B): Francis Greene Jr. (1694–<1761) and Elizabeth Wheeler (1693–1758), Thomas Mattingly (1623–1664) and Elizabeth McWilliams (1626–1714), and John Shercliffe Sr. (1618–1663) and Anne Spinke (1618–1678). All three couples share a 9th- to 11th-degree direct-descent relationship with both participants, consistent with the maximum likelihood estimates from our genetic pedigree reconstruction (Figure 5C). While random inheritance patterns make predicting exact degrees of relatedness difficult, we prioritized these couples and their close relatives as the most likely candidates. To determine which of

these candidates (if any) could be burial 56, we compared what is known about them from the historical record with the biological profile of burial 56, established by osteological and archeological assessments.

Burial 56 was situated against the eastern edge of the chapel's foundation in a heavily used area of the cemetery. It was an unusual secondary burial, meaning the skeleton had been previously disturbed and redeposited in a wooden box. The burial was commingled, containing remains from more than one individual.⁷ Alongside the primary individual, the box contained leg bones believed to belong to burial 57, a nearby partial grave that was disturbed by the chapel's construction and the later excavation of burial 56. The elements representing burial 56 were carefully sorted based on similarities in age, sex, preservation, and stable isotope analysis, revealing that burial 56 was a male of European ancestry, aged approximately 40–49 years, with a dietary signal that indicated he was either born in Maryland or immigrated there as a young child. Based on the archeological context, he was likely buried between 1667 and 1704 (following the chapel's construction); however, because this was a secondary burial, his actual death date may have been earlier. We therefore considered candidates who died between 1650 and 1704.

Based on this profile, we ruled out John Shercliffe Sr. and Thomas Mattingly as candidates, as both men were born in England and immigrated as adults. While Francis Greene Jr. was born in Maryland, he was ruled out because he lived too recently—while his exact date of death is unknown, his youngest child was born in 1745, indicating that he survived at least 40 years beyond the latest date associated with the cemetery's use. After eliminating these primary candidates, we went on to investigate other members of their family trees based on information available in the FamilySearch database.

The ancestors of John Shercliffe Sr. and Thomas Mattingly (and their wives) could all be ruled out, as they were born in England. Therefore, we focused our analysis on the lineage of Francis Greene Jr. His father, Francis Greene Sr. (1648–1708), died outside our target age and date ranges. However, Francis Greene Sr.'s half-brother, Leonard Greene (c.1636–1688), fits this profile.¹⁸ While his reported age at death of 50–52 is slightly older than the osteological estimate (40–49 years), it is within an acceptable margin of error for skeletal aging.

Providing further support for this identification, Leonard's father, Governor Thomas Greene (1609–1651), fits the profile of V07C. V07C was identified as the father of burial 56 through a combination of genetic and isotopic evidence. Genetics support a father/son relationship, while isotopes indicate that V07C was born in England and emigrated to Maryland, while burial 56 was born and raised in Maryland. Due to the locations of their births, it is far more likely that V07C is burial 56's father than his son. The remains of V07C were redeposited in an ossuary when the chapel was built in 1667, providing an indicator of the latest year he could have died, and his age of death was estimated to be in the range of 40–54 years. Similarly, burial 56 shares a first-degree relationship with burial 63, who fits the profile for Leonard's mother, Anne Cox (1611–1638). We provide more details about the other candidates that we considered in Data S6. However, we believe that these three individuals (Leonard Greene, Anne Cox, and Governor

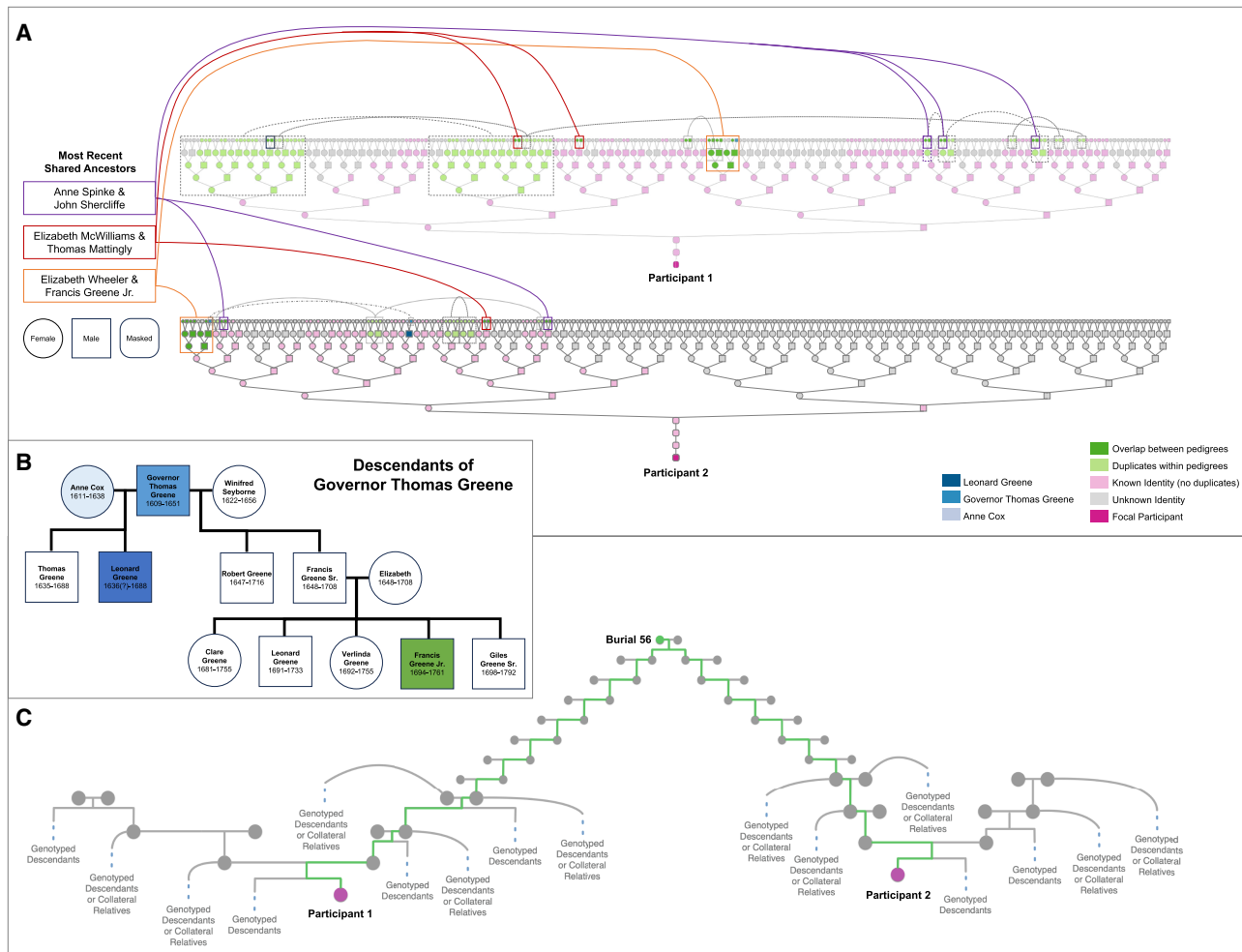


Figure 5. Genealogical and genetic pedigrees connecting focal participants to burial 56

(A) The reconstructed genealogical pedigree of each focal participant who shares over 30 cM of IBD with burial 56 is shown. The pedigrees focus on ancestors of the focal participants who were born prior to 1850 on lineages with reported connections to St. Mary's City (Data S7K and S7L). Ancestors that appear in both pedigrees are shown in dark green (and solid lines lead to the names of these ancestors). Ancestors that appear multiple times within the same pedigree are shown in light green (and duplicate portions of the pedigrees are outlined and connected via a dotted line). Ancestors for whom an identity is known are shown in light pink. The focal participant is shown in dark pink. Individuals whose identity is unknown are shown in gray.

(B) Genealogical tree showing the descendants of Governor Thomas Greene and his wives, Anne Cox and Winifred Seyborne. Only the children of Francis Greene Sr. are shown in order to demonstrate the relationship between Leonard Greene and Francis Greene Jr.

(C) Genetically inferred pedigree for the two focal participants, who are shown in purple. Branches indicate the locations of sets of one or more genotyped descendants or collateral relatives. The genetically inferred position of burial 56 is shown in green, and the lineages leading from each survey participant to this node are shown in green.

See also [Data S4](#), [S5](#), [S6](#), [S7K](#), and [S7L](#).

Thomas Greene) best fit the biological profiles of burials 56, 63, and V07C, respectively.

Interestingly, Leonard Greene is also a direct ancestor of participant 2, thereby making Governor Thomas Greene the direct ancestor of this participant at least two times over. This highlights the complex loops that are common throughout genealogical pedigrees involving the St. Mary's founding population.

While this analysis has enabled us to propose possible identities for three St. Mary's individuals, we caution that there are substantial gaps in the genealogical pedigrees we reconstructed based on the self-reported family histories of the two focal

participants, leaving open the possibility that there are additional, unidentified, overlaps in their trees. Additionally, it is possible that burial 56 may be a more distant collateral relative of the three focal couples we considered.

DISCUSSION

Although St. Mary's City is a relatively well-documented colonial settlement, surviving records provide little information about the vast majority of its residents.⁶ This study therefore highlights the unique power of IBD-based analyses to deepen our

understanding of one of the earliest British colonial communities in North America and to trace the genetic legacy of its founders.

Among the 49 St. Mary's individuals studied, all but one had exclusively European ancestry. The exception, a young boy with majority African ancestry interred among the European-ancestry individuals, following similar burial customs, is a significant finding that warrants additional consideration. His discovery is an important reminder of the diversity present in the Chapel Field cemetery.

For the European-ancestry individuals, we identified five genetic families, at least one of which spans three generations. Further, IBD sharing indicates that many of the St. Mary's individuals included in our study likely shared more distant genetic relationships, involving four or more degrees of separation. Although Philip Calvert, his first wife Anne Wolseley Calvert, and his infant son were tentatively identified by previous archeological, osteological, and historical research, little was known about others who were buried in the Chapel Field cemetery. Here, we identify three additional members of the Calvert extended family, including those whose crania were found in an ossuary pit dating to 1667. The next largest genetic family (family A) may also have been distantly related to the Calvert family.

Historical records document that the *Ark* and the *Dove*, the ships carrying the first St. Mary's colonists, set sail for the Americas from the Isle of Wight off the south coast of England.¹⁹ However, these records provide little information about the exact ancestral origins of the passengers or about the many subsequent vessels that carried immigrants to the colony. We show that many of the St. Mary's individuals shared genetic connections to participants from western England and Wales, suggesting that they were common places of origin for the city's early settlers. Although we cannot exclude the possibility that this pattern is influenced by broader demographic changes that occurred within Great Britain over the last few hundred years, fine-scale regional structure within Great Britain has been shown to persist into the present-day, despite modern mobility, supporting the use of contemporary regional patterns as a proxy for historical ones.²⁰

We also identified a number of St. Mary's individuals who exhibited connections to other regions in Great Britain and Ireland, indicating that the St. Mary's founding population had diverse origins from across the two islands. Notably, two individuals who exhibited strong genetic connections to Ireland (burials 12 and 17) were young adult males who died between 1634 and 1667. Isotopic evidence indicates that both were relatively recent immigrants based on previously established patterns of colonial diet and migration.^{21,22} Neither was buried in a coffin, and only one was with a shroud. Both showed skeletal signs of heavy physical labor and poor health. Although their status of bondage is unknown, these features are consistent with the profile of indentured servants. Although the biomolecular evidence presented here cannot confirm this interpretation, this adds depth to the personal histories of these two men, as many Irish immigrants who arrived in St. Mary's City (and the Chesapeake region more broadly) during the 17th century came as indentured servants.^{23,24}

By examining patterns of IBD sharing with participants across the US, we learn about the lasting genetic impact of the St. Mary's founder population on the demography of the country. We observe elevated rates of distant IBD sharing across the southern US, suggesting that the present-day population of the region

shares strong genetic similarities to the St. Mary's founding population. For most distant genetic connections, it is impossible to infer whether participants descend directly from the St. Mary's individuals or if both descend from common ancestral populations in Great Britain and Ireland. However, the elevated IBD sharing rate observed among research participants who indicated via survey that they have ancestral ties to 17th-century St. Mary's City suggests that many of these distant genetic connections are likely the result of direct descent from individuals included in this analysis (or their close relatives). Notably, the patterns of sharing that we observed appear similar to the distribution of self-reported "American" ancestry in the 2000 US Census, an identity that is commonly adopted by white Americans whose ancestors immigrated to the Americas many generations ago.² This group may therefore be more likely to trace their ancestry back to one or more colonial-era founder populations, like St. Mary's City.

We observed strong geographic patterns of IBD sharing among research participants with the closest genetic connections (≥ 30 cM of shared IBD) to the St. Mary's individuals. These patterns are consistent with the expected distribution of descendants of the St. Mary's founder population throughout the US. For instance, we observe a particularly strong concentration of close relatives of the St. Mary's individuals in Kentucky. This likely reflects a documented migration of Catholic residents from St. Mary's County (where St. Mary's City is located) to Kentucky—particularly to Nelson and Washington (formerly Marion) counties—primarily between 1780 and 1820. Migrating individuals moved in pursuit of economic advantage and religious freedom due to the continued anti-Catholic bias in Maryland during this period.^{14–16} That IBD analysis can detect such a clear signal of this well-documented migration demonstrates the power of this approach.

Finally, we introduced a new framework for identifying otherwise nameless historical individuals by analyzing genealogical connections among research participants who share close genetic links to the same historical genomes. This differs from previous identification efforts involving aDNA, which have typically relied on strong prior hypotheses regarding the identity of specific remains. Using this approach, we propose identities for three St. Mary's individuals: burial 56 (Leonard Greene), burial 63 (Anne Cox), and V07C (Governor Thomas Greene). It is highly probable that these three individuals were interred at the Brick Chapel, as the Greene family lived in St. Mary's City on Greene's Freehold following Thomas Greene's time as governor.¹⁸ While these identifications warrant further verification, the recovery of such a notable figure underscores the limitations of the site's historical records, as the burial locations of even the most prominent citizens were unrecorded. Conversely, our ability to successfully trace these lineages—while many others remain obscure—likely reflects the high status of the Greene family, which ensured the preservation of genealogical records where others were lost. Therefore, the identification of these individuals as members of the Greene family should be interpreted as the most parsimonious assignment rather than a concrete identification, given the potential for missing branches in self-reported trees. Beyond these specific identifications, our results highlight the highly interconnected nature of colonial founder populations, underscoring the complexity of reconstructing the genetic trees that link present-day people to their historic-era relatives.

We anticipate that this combined genetic and genealogical framework will become increasingly effective at more recent

time depths. While this study establishes the method's viability at ~350 years, applying it to more recent historical cases will likely yield higher identification success rates due to the retention of larger IBD segments among direct descendants (and other close collateral relatives) and the availability of denser genealogical records. Future applications of this framework could be further strengthened by incorporating mitochondrial and Y chromosome haplogroup data to independently confirm unbroken maternal or paternal lineages. While this orthogonal line of evidence was outside the scope of our initial institutional review board (IRB) protocol, its inclusion in future studies could provide an even higher degree of certainty when assigning identities. Our findings therefore serve as a proof of concept for a strategy with broad potential to restore lost identities across history.

By generating genome-wide information from 49 individuals from 17th-century St. Mary's City, Maryland, this study demonstrates the power of genetic analyses to localize the biogeographic origins of colonizing populations and trace the subsequent migrations of their descendants. It also highlights the enduring legacy of this colonial founder population in the present-day US population. This unique case study of a relatively well-documented colonial settlement with a large number of excavated and studied historical human remains illustrates how aDNA can expand our knowledge of founder populations and their connections to contemporary communities. Our findings underscore the potential for future applications of this approach to expand our understanding of migration and ancestry at less-well-documented sites. This approach also has applications outside of the field of genetics, especially for historical archeologists, bioarcheologists, and individuals seeking information on their ancestors.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Éadaoin Harney (eadaoih@23andme.com).

Materials availability

This study did not generate any new, unique reagents.

Data and code availability

- All data needed to evaluate the conclusions in the paper are present in the paper and/or the [supplemental information](#).
- The aligned sequences for the two Calvert individuals and the 47 newly reported St. Mary's individuals are available from the European Nucleotide Archive under ENA: PRJEB16735 and PRJEB110565, respectively. Genotype files for pseudo-haploid and phased, imputed versions of the dataset are available at Harvard Dataverse (<https://doi.org/10.7910/DVN/6IHEYQ>).
- There are restrictions to the availability of 23andMe genotype data due to 23andMe informed consent and privacy guidelines, which result in a contractual obligation of 23andMe to customers to not make data publicly available. However, to ensure replicability of these results, 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, [supplemental tables](#) are also 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 research participant must be indistinguishable from at least four other research participants included in the dataset.

- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this work is available from the [lead contact](#) upon request.

CONSORTIA

The members of the 23andMe Research Team are Stella Aslibekyan, Adam Auton, Robert K. Bell, Katelyn Kukar Bond, Zayn Cochinwala, Sayantan Das, Kahsaia de Brito, Emily DelloRusso, Chris Eijsbouts, Sarah L. Elson, Chris German, Julie M. Granka, Barry Hicks, David A. Hinds, Reza Jabal, Aly Khan, Matthew J. Kmieciak, Alan Kwong, Yanyu Liang, Keng-Han Lin, Matthew H. McIntyre, David Poznik, Emily Rios, Shubham Saini, Anjali J. Shastri, Jingchunzi Shi, Suyash Shringarpure, Qiaojuan Jane Su, Vinh Tran, Joyce Y. Tung, Catherine H. Weldon, Peter Wilton, and Wanwan Xu.

ACKNOWLEDGMENTS

We acknowledge the individuals from 17th-century St. Mary's City whose DNA we studied and recognize that they could not directly consent to participate in this study. We also thank the Historic St. Mary's City Commission, who provided consent for this research on their behalf and who provided critical feedback throughout the course of this study. We thank the 23andMe research participants and others who consented to participate in research for enabling this study. We thank Elizabeth Curtis, Kristin Stewardson, and Fatma Zalzal for laboratory work. We thank Iosif Lazaridis for the Y chromosome haplogroup analysis. The aDNA analysis was funded by NIH grant HG012287; John Templeton Foundation grant 61220; a gift from Jean-François Clin; the Allen Discovery Center program, a Paul G. Allen Frontiers Group advised program of the Paul G. Allen Family Foundation; and the Howard Hughes Medical Institute. The pedigree analyses were funded in part by NIH grant R35GM133805. The Rice Family Endowment for Forensic Anthropology and the Kathryn Nell Harrison Foundation provided critical support for Smithsonian Skeletal Biology Program contributions.

AUTHOR CONTRIBUTIONS

Curation of skeletal remains, K.S.B., H.M., K. Barca, and D.W.O.; generation of genetic data (wet lab), A.K., K.C., L.I., L.Q., J.N.W., and N.R.; generation of genetic data (computational), A.A., M.M., G.S., and S. Mallick; data analysis, E.H., E.J., S. Micheletti, and R.C.; interpretation of results, E.H., E.J., S. Micheletti, R.C., K.S.B., W.F., H.M., K. Barca, K. Bryc, S.A.E., I.O., J.S., K.S., J.M., D.W.O., and D.R.; study logistics/computational support, S.D., K.K., A.M., D.N., and 23andMe Research Team; conceptualization and supervision, J.M., D.W.O., and D.R.; writing, E.H., E.J., S. Micheletti, R.C., K.S.B., W.F., H.M., A.A., K. Barca, K. Bryc, S.A.E., J.S., K.S., J.M., D.W.O., and D.R.

DECLARATION OF INTERESTS

E.H., E.J., S. Micheletti, W.A.F., K. Bryc, S.D., S.A.E., K.K., A.M., D.T.N., J.L.M., and members of the 23andMe Research Team are current or former employees of the 23andMe Research Institute (formerly 23andMe, Inc.).

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- [KEY RESOURCES TABLE](#)
- [EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS](#)
 - Community Engagement and Ethics
 - Provenance of historical skeletal remains

- Present-day research participants
- **METHOD DETAILS**
 - Ancient DNA (aDNA) Sequencing
 - Uniparental Haplogroups
 - Publicly available dataset assembly
 - Genetic relatedness
 - ADMIXTURE
 - PCA
 - Imputation
 - Identity by Descent
 - IBD Networks
 - Individual-Level Analyses
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
 - Randomization Tests
 - Chi Squared and *T*-Tests

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.cub.2026.04.046>.

A video abstract is available at <https://doi.org/10.1016/j.cub.2026.04.046#mmc10>.

Received: February 3, 2026

Revised: March 31, 2026

Accepted: April 17, 2026

REFERENCES

1. Gauvin, H., Moreau, C., Lefebvre, J.-F., Laprise, C., Vézina, H., Labuda, D., and Roy-Gagnon, M.-H. (2014). Genome-wide patterns of identity-by-descent sharing in the French Canadian founder population. *Eur. J. Hum. Genet.* 22, 814–821. <https://doi.org/10.1038/ejhg.2013.227>.
2. Bryc, K., Durand, E.Y., Macpherson, J.M., Reich, D., and Mountain, J.L. (2015). The genetic ancestry of African Americans, Latinos, and European Americans across the United States. *Am. J. Hum. Genet.* 96, 37–53. <https://doi.org/10.1016/j.ajhg.2014.11.010>.
3. Harney, É., Micheletti, S., Bruwelheide, K.S., Freyman, W.A., Bryc, K., Akbari, A., Jewett, E., Comer, E., Gates, H.L., Heywood, L., et al. (2023). The genetic legacy of African Americans from Catoctin Furnace. *Science* 381, eade4995. <https://doi.org/10.1126/science.ade4995>.
4. Browning, S.R., and Browning, B.L. (2012). Identity by descent between distant relatives: detection and applications. *Annu. Rev. Genet.* 46, 617–633. <https://doi.org/10.1146/annurev-genet-110711-155534>.
5. Ringbauer, H., Huang, Y., Akbari, A., Mallick, S., Olalde, I., Patterson, N., and Reich, D. (2024). Accurate detection of identity-by-descent segments in human ancient DNA. *Nat. Genet.* 56, 143–151. <https://doi.org/10.1038/s41588-023-01582-w>.
6. Hurry, S.D. (2001). "...once the Metropolis of Maryland": The History and Archaeology of Maryland's First Capital (Historic St. Mary's City Commission).
7. Riordan, T.B. (2000). *Dig a Grave Both Wide and Deep: An Archaeological Investigation of Mortuary Practices in the 17th-Century Cemetery at St. Mary's City, Maryland* (Historic St. Mary's City).
8. Reich, D., Stewardson, K., Lazaridis, I., Mallick, S., Rohland, N., and Owsley, D. (2016). Ancient DNA Analysis of St. Mary's City Lead Coffin Burials⁵ (Reich Laboratory, Harvard Medical School).
9. Newman, H.W. (1984). *The Flowering of the Maryland Palatinate: An Intimate and Objective History of the Province of Maryland to the Overthrow of Proprietary Rule in 1654, with Accounts of Lord Baltimore's Settlement at Avalonxviii* (Genealogical Publishing Com), pp. 359.
10. Orlando, L., Allaby, R., Skoglund, P., Der Sarkissian, C., Stockhammer, P.W., Ávila-Arcos, M.C., Fu, Q., Krause, J., Willerslev, E., Stone, A.C., et al. (2021). Ancient DNA analysis. *Nat. Rev. Methods Primers* 1, 1–26. <https://doi.org/10.1038/s43586-020-00011-0>.
11. Fernandes, D., Sirak, K., Novak, M., Finarelli, J.A., Byrne, J., Connolly, E., Carlsson, J.E.L., Ferretti, E., Pinhasi, R., and Carlsson, J. (2017). The identification of a 1916 Irish rebel: new approach for estimating relatedness from low coverage homozygous genomes. *Sci. Rep.* 7, 41529. <https://doi.org/10.1038/srep41529>.
12. Moltke, I., Korneliusson, T.S., Seguin-Orlando, A., Moreno-Mayar, J.V., LaPointe, E., Billeck, W., and Willerslev, E. (2021). Identifying a living great-grandson of the Lakota Sioux leader Tatanka Iyotake (Sitting Bull). *Sci. Adv.* 7, eabh2013. <https://doi.org/10.1126/sciadv.abh2013>.
13. Micheletti, S.J., Bryc, K., Ancona Esselmann, S.G., Wilton, P.R., and Freyman, W.A. (2025). Bayesian inference of population structure using identity-by-descent-based stochastic block models. Preprint at bioRxiv. <https://doi.org/10.1101/2025.11.26.690810>.
14. Marks, B.E. (1979). *The rage for Kentucky: emigration from St. Mary's County*. In *Geographic Perspectives on Maryland's Past*, R.D. Mitchell, and E.K. Mueller, eds., pp. 108–128.
15. Marks, B.E. (1979). *Economics and Society in a Staple Plantation System*. PhD thesis (St. Mary's).
16. Peden, H.C., Jr. (1991). *Marylanders to Kentucky* (Family Line Publications), pp. 1775–1825.
17. Pesken, L.A. (1996). *A restless generation: migrations of Maryland veterans in the early republic*. *Md. Hist. Mag.* 91, 311–328.
18. Carr, L.G., Men's Career Files MSA SC 5094 Image No: sc5094-1628-5 <https://msa.maryland.gov/megafile/msa/speccol/sc5000/sc5094/001000/001628/html/sc5094-1628-5.html>.
19. Hall, C.C. (1910). *A Briefe Relation of the Voyage unto Maryland, By Father Andrew White, 1634*. In *Narratives of Early Maryland* (Barnes and Noble), pp. 29–45.
20. Leslie, S., Winney, B., Hellenthal, G., Davison, D., Boumertit, A., Day, T., Hutnik, K., Royrvik, E.C., Cunliffe, B., et al.; Wellcome Trust Case Control Consortium 2 (2015). The fine-scale genetic structure of the British population. *Nature* 519, 309–314. <https://doi.org/10.1038/nature14230>.
21. Ubelaker, D.H., and Owsley, D.W. (2003). Isotopic evidence for diet in the seventeenth-century colonial Chesapeake. *Am. Antiquity* 68, 129–139. <https://doi.org/10.2307/3557036>.
22. France, C.A., Owsley, D.W., and Hayek, L.A.C. (2014). Stable isotope indicators of provenance and demographics in 18th and 19th century North Americans. *J. Archaeol. Sci.* 42, 356–366. <https://doi.org/10.1016/j.jas.2013.10.037>.
23. Menard, R.R. (1988). *British migration to the Chesapeake colonies in the seventeenth century*. In *Colonial Chesapeake Society*, L.G. Carr, P.D. Morgan, and J.B. Russo, eds. (University of North Carolina Press), pp. 99–132.
24. *Maryland General Assembly (1704). An Act Imposing a Duty of Twenty Shillings per Poll on Irish Servants to Prevent the Importing Too Great a Number of Irish Papists into This Province* (Maryland Historical Society).
25. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25, 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>.
26. Weissensteiner, H., Pacher, D., Kloss-Brandstätter, A., Forer, L., Specht, G., Bandelt, H.-J., Kronenberg, F., Salas, A., and Schönherr, S. (2016). HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res.* 44, W58–W63. <https://doi.org/10.1093/nar/gkw233>.
27. Rubinacci, S., Ribeiro, D.M., Hofmeister, R.J., and Delaneau, O. (2021). Efficient phasing and imputation of low-coverage sequencing data using large reference panels. *Nat. Genet.* 53, 120–126. <https://doi.org/10.1038/s41588-020-00756-0>.
28. Loh, P.-R., Danecek, P., Palamara, P.F., Fuchsberger, C., A Reshef, Y., K Finucane, H., Schoenherr, S., Forer, L., McCarthy, S., Abecasis, G.R., et al. (2016). Reference-based phasing using the Haplotype Reference Consortium panel. *Nat. Genet.* 48, 1443–1448. <https://doi.org/10.1038/ng.3679>.

29. Patterson, N., Price, A.L., and Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genet.* 2, e190. <https://doi.org/10.1371/journal.pgen.0020190>.
30. Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655–1664. <https://doi.org/10.1101/gr.094052.109>.
31. Freyman, W.A., McManus, K.F., Shringarpure, S.S., Jewett, E.M., Bryc, K., 23 and Me Research Team, and Auton, A. (2021). Fast and robust identity-by-descent inference with the templated positional Burrows-Wheeler transform. *Mol. Biol. Evol.* 38, 2131–2151. <https://doi.org/10.1093/molbev/msaa328>.
32. Jewett, E.M., McManus, K.F., Freyman, W.A., 23andMe Research Team, and Auton, A. (2021). Bonsai: An efficient method for inferring large human pedigrees from genotype data. *Am. J. Hum. Genet.* 108, 2052–2070. <https://doi.org/10.1016/j.ajhg.2021.09.013>.
33. Alpaslan-Roodenberg, S., Anthony, D., Babiker, H., Bánffy, E., Booth, T., Capone, P., Deshpande-Mukherjee, A., Eisenmann, S., Fehren-Schmitz, L., Frachetti, M., et al. (2021). Ethics of DNA research on human remains: five globally applicable guidelines. *Nature* 599, 41–46. <https://doi.org/10.1038/s41586-021-04008-x>.
34. Wagner, J.K., Colwell, C., Claw, K.G., Stone, A.C., Bolnick, D.A., Hawks, J., Brothers, K.B., and Garrison, N.A. (2020). Fostering responsible research on ancient DNA. *Am. J. Hum. Genet.* 107, 183–195. <https://doi.org/10.1016/j.ajhg.2020.06.017>.
35. Harney, É., Sirak, K., Sedig, J., Micheletti, S., Curry, R., Ancona Esselmann, S., and Reich, D. (2023). Ethical considerations when co-analyzing ancient DNA and data from private genetic databases. *Am. J. Hum. Genet.* 110, 1447–1453. <https://doi.org/10.1016/j.ajhg.2023.06.011>.
36. 1000 Genomes Project Consortium, Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., et al. (2015). A global reference for human genetic variation. *Nature* 526, 68–74. <https://doi.org/10.1038/nature15393>.
37. Durand, E.Y., Do, C.B., Wilton, P.R., Mountain, J.L., Auton, A., Poznik, G.D., and Macpherson, J.M. (2021). A scalable pipeline for local ancestry inference using tens of thousands of reference haplotypes. Preprint at bioRxiv. <https://doi.org/10.1101/2021.01.19.427308>.
38. Sirak, K.A., Fernandes, D.M., Cheronet, O., Novak, M., Gamarra, B., Balassa, T., Bernert, Z., Cséki, A., Dani, J., Gallina, J.Z., et al. (2017). A minimally-invasive method for sampling human petrous bones from the cranial base for ancient DNA analysis. *BioTechniques* 62, 283–289. <https://doi.org/10.2144/000114558>.
39. Rohland, N., Glocke, I., Aximu-Petri, A., and Meyer, M. (2018). Extraction of highly degraded DNA from ancient bones, teeth and sediments for high-throughput sequencing. *Nat. Protoc.* 13, 2447–2461. <https://doi.org/10.1038/s41596-018-0050-5>.
40. Rohland, N., Harney, E., Mallick, S., Nordenfelt, S., and Reich, D. (2015). Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. *Philos. Trans. R. Soc. Lond., B Biol. Sci.* 370, 20130624. <https://doi.org/10.1098/rstb.2013.0624>.
41. Gansauge, M.-T., Aximu-Petri, A., Nagel, S., and Meyer, M. (2020). 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. <https://doi.org/10.1038/s41596-020-0338-0>.
42. Fu, Q., Mittnik, A., Johnson, P.L.F., Bos, K., Lari, M., Bollongino, R., Sun, C., Giemsch, L., Schmitz, R., Burger, J., et al. (2013). A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr. Biol.* 23, 553–559. <https://doi.org/10.1016/j.cub.2013.02.044>.
43. Fu, Q., Hajdinjak, M., Moldovan, O.T., Constantin, S., Mallick, S., Skoglund, P., Patterson, N., Rohland, N., Lazaridis, I., Nickel, B., et al. (2015). An early modern human from Romania with a recent Neanderthal ancestor. *Nature* 524, 216–219. <https://doi.org/10.1038/nature14558>.
44. Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B., Brandt, G., Nordenfelt, S., Harney, E., Stewardson, K., et al. (2015). Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* 522, 207–211. <https://doi.org/10.1038/nature14317>.
45. Mathieson, I., Alpaslan-Roodenberg, S., Posth, C., Szécsényi-Nagy, A., Rohland, N., Mallick, S., Olalde, I., Broomandkoshbacht, N., Candilio, F., Cheronet, O., et al. (2018). The genomic history of southeastern Europe. *Nature* 555, 197–203. <https://doi.org/10.1038/nature25778>.
46. Behar, D.M., Van Oven, M., Rosset, S., Metspalu, M., Loogväli, E.-L., Silva, N.M., Kivisild, T., Torroni, A., and Villems, R. (2012). A “Copernican” reassessment of the human mitochondrial DNA tree from its root. *Am. J. Hum. Genet.* 90, 675–684. <https://doi.org/10.1016/j.ajhg.2012.03.002>.
47. Skoglund, P., Storå, J., Götherström, A., and Jakobsson, M. (2013). Accurate sex identification of ancient human remains using DNA shotgun sequencing. *J. Archaeol. Sci.* 40, 4477–4482. <https://doi.org/10.1016/j.jas.2013.07.004>.
48. Olalde, I., Mallick, S., Patterson, N., Rohland, N., Villalba-Mouco, V., Silva, M., Dülías, K., Edwards, C.J., Gandini, F., Pala, M., et al. (2019). The genomic history of the Iberian Peninsula over the past 8000 years. *Science* 363, 1230–1234. <https://doi.org/10.1126/science.aav4040>.
49. Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., and Reich, D. (2012). Ancient admixture in human history. *Genetics* 192, 1065–1093. <https://doi.org/10.1534/genetics.112.145037>.
50. Pickrell, J.K., Patterson, N., Barbieri, C., Berthold, F., Gerlach, L., Güldemann, T., Kure, B., Mpoloka, S.W., Nakagawa, H., Naumann, C., et al. (2012). The genetic prehistory of southern Africa. *Nat. Commun.* 3, 1143. <https://doi.org/10.1038/ncomms2140>.
51. Lazaridis, I., Patterson, N., Mittnik, A., Renaud, G., Mallick, S., Kirsanow, K., Sudmant, P.H., Schraiber, J.G., Castellano, S., Lipson, M., et al. (2014). Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* 513, 409–413. <https://doi.org/10.1038/nature13673>.
52. Lazaridis, I., Nadel, D., Rollefson, G., Merrett, D.C., Rohland, N., Mallick, S., Fernandes, D., Novak, M., Gamarra, B., Sirak, K., et al. (2016). Genomic insights into the origin of farming in the ancient Near East. *Nature* 536, 419–424. <https://doi.org/10.1038/nature19310>.
53. Biagini, S.A., Solé-Morata, N., Matisoo-Smith, E., Zalloua, P., Comas, D., and Calafell, F. (2019). People from Ibiza: an unexpected isolate in the Western Mediterranean. *Eur. J. Hum. Genet.* 27, 941–951. <https://doi.org/10.1038/s41431-019-0361-1>.
54. Flegontov, P., Altınışık, N.E., Changmai, P., Rohland, N., Mallick, S., Adamski, N., Bolnick, D.A., Broomandkoshbacht, N., Candilio, F., Culleton, B.J., et al. (2019). Palaeo-Eskimo genetic ancestry and the peopling of Chukotka and North America. *Nature* 570, 236–240. <https://doi.org/10.1038/s41586-019-1251-y>.
55. Jeong, C., Balanovsky, O., Lukianova, E., Kahbatkyzy, N., Flegontov, P., Zaporozhchenko, V., Immel, A., Wang, C.-C., Ixan, O., Khussainova, E., et al. (2019). The genetic history of admixture across inner Eurasia. *Nat. Ecol. Evol.* 3, 966–976. <https://doi.org/10.1038/s41559-019-0878-2>.
56. Bergström, A., McCarthy, S.A., Hui, R., Almarri, M.A., Ayub, Q., Danecek, P., Chen, Y., Felkel, S., Hallast, P., Kamm, J., et al. (2020). Insights into human genetic variation and population history from 929 diverse genomes. *Science* 367, eaay5012. <https://doi.org/10.1126/science.aay5012>.
57. Wang, C.C., Yeh, H.Y., Popov, A.N., Zhang, H.Q., Matsumura, H., Sirak, K., Cheronet, O., Kovalev, A., Rohland, N., Kim, A.M., et al. (2021). Genomic insights into the formation of human populations in East Asia. *Nature* 597, 413–419. <https://doi.org/10.1038/s41586-021-03336-2>.
58. Mallick, S., Micco, A., Mah, M., Ringbauer, H., Lazaridis, I., Olalde, I., Patterson, N., and Reich, D. (2023). The Allen Ancient DNA Resource (AADR): A curated compendium of ancient human genomes. *Sci. Data* 11. <https://doi.org/10.1101/2023.04.06.535797>.
59. Mallick, S., and Reich, D. (2024). The Allen Ancient DNA Resource (AADR): A Curated Compendium of Ancient Human Genomes. Version V9 Data Release (Harvard Dataverse). <https://doi.org/10.7910/DVN/FFIDCW>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
2 Calvert Individuals (Phillip & Son)	Reich et al. ⁸	N/A
47 newly reported St. Mary's individuals	This paper	N/A
Chemicals, peptides, and recombinant proteins		
23 HI-RPM hybridization buffer	Agilent Technologies	5190-0403
Herculase II fusion DNA polymerase	Agilent Technologies	600679
Pfu turbo Cx Hotstart DNA polymerase	Agilent Technologies	600412
50% PEG 8000	Anatrace	OPTIMIZE-82 100 ML
0.5 M EDTA pH 8.0	BioExpress	E177
Sera-Mag SpeedBead CarboxylateModified [E3] magnetic particles	Cytiva Life Sciences	65152105050250
Silica magnetic beads	G-Biosciences	786–916
10 3 T4 RNA ligase buffer	New England Biolabs	B0216L
Bst DNA Polymerase2.0, large frag.	New England Biolabs	M0537
UGI	New England Biolabs	M0281
USER enzyme	New England Biolabs	M5505
Buffer PB	QIAGEN	19066
Buffer PE concentrate	QIAGEN	19065
1 M Tris-HCl pH 8.0	Sigma Aldrich	AM9856
1 M NaOH	Sigma Aldrich	71463
20% SDS	Sigma Aldrich	5030
3 M Sodium Acetate (pH 5.2)	Sigma Aldrich	S7899
5 M NaCl	Sigma Aldrich	S5150
Ethanol	Sigma Aldrich	E7023
Guanidine hydrochloride	Sigma Aldrich	G3272
Isopropanol	Sigma Aldrich	650447
PEG-8000	Sigma Aldrich	89510
Proteinase K	Sigma Aldrich	P6556
Tween-20	Sigma Aldrich	P9416
Water	Sigma Aldrich	W4502
103 Buffer Tango	Thermo Fisher Scientific	BY5
503 Denhardt's solution	Thermo Fisher Scientific	750018
AccuPrime Pfx Polymerase (2.5 U/ul)	Thermo Fisher Scientific	12344032
ATP	Thermo Fisher Scientific	R0441
dNTP Mix	Thermo Fisher Scientific	R1121
Dyna MyOne Streptavidin C1 beads	Thermo Fisher Scientific	65002
FastAP (1 U/mL)	Thermo Fisher Scientific	EF0651
GeneAmp 103 PCR Gold Buffer	Thermo Fisher Scientific	4379874
Human Cot-1 DNA	Thermo Fisher Scientific	15279011
Klenow Fragment (10 U/mL)	Thermo Fisher Scientific	EP0052
Maxima Probe qPCR 2xMM	Thermo Fisher Scientific	K0233
Maxima SYBR Green kit	Thermo Fisher Scientific	K0251
Maxima SYBR Green kit	Thermo Fisher Scientific	K0253
Salmon sperm DNA	Thermo Fisher Scientific	15632-011
SSC Buffer (203)	Thermo Fisher Scientific	AM9770

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
T4 DNA Ligase	Thermo Fisher Scientific	EL0012
T4 DNA Ligase, HC (30U/mL)	Thermo Fisher Scientific	EL0013
T4 DNA Polymerase	Thermo Fisher Scientific	EP0062
T4 Polynucleotide Kinase	Thermo Fisher Scientific	EK0032
23 HI-RPM hybridization buffer	Agilent Technologies	5190-0403
2% Sodium Hypochlorite Solution	Millipore Sigma	Cat# XX0637-76
Acetic Acid, Glacial (TraceMetal Grade)	Fisher Chemical	Cat# A507-P212
7 M HNO ₃ (Optima)	Fisher Chemical	Cat# A467-2
6 M HCL (TraceMetal Grade)	Fisher Chemical	Cat# A508-4
0.05 M HNO ₃ (Optima)	Fisher Chemical	Cat# A467-2
30% H ₂ O ₂ (GR ACS Grade)	Millipore Sigma	Cat# HX0635-2
0.1 M CH ₃ COOH (GR ACS Grade)	Millipore Sigma	Cat# AX0073-6

Critical commercial assays

HiSeq X Ten Reagent Kit v2.5	Illumina	FC-501-2521
NextSeq 500/550 High Output Kit v2.5	Illumina	Cat.# 20024906

Deposited data

Sequencing data from 2 Calvert Individuals (Phillip & Son)	Reich et al. ⁸	ENA: PRJEB16735
Sequencing data from 47 newly reported St. Mary's individuals	This paper	ENA: PRJEB110565
Genotype data from all 49 St. Mary's individuals (including pseudohaploid genotypes for all individual, and imputed and phased genotypes for the 29 highest coverage individuals)	This paper	Harvard Dataverse: https://doi.org/10.7910/DVN/6IHEYQ

Software and algorithms

aDNA-tools	https://github.com/DReichLab/ADNA-Tools	N/A
BWA (v0.6.1)	Li and Durbin ²⁵	N/A
HaploGrep2	Weissensteiner et al. ²⁶	N/A
GLIMPSE (v1.0.0)	Rubinacci et al. ²⁷	N/A
EAGLE (v2.4.1)	Loh et al. ²⁸	N/A
Bcftools (v1.14)	Li and Durbin ²⁵	N/A
SAMtools (v1.22)	Li and Durbin ²⁵	N/A
Admixtools (v6)	https://github.com/dReichLab/AdmixTools	N/A
Smartpca (v18700)	Patterson et al. ²⁹	N/A
Admixture (v1.3.0)	Alexander et al. ³⁰	N/A
TPBWT	Freyman et al. ³¹ ; https://github.com/23andme/phasedibd	N/A
Bonsai	Jewett et al. ³² ; https://github.com/23andMe/bonsaitree/tree/main/bonsaitree/v3	N/A

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Community Engagement and Ethics

This study was made possible by the millions of research participants who actively consented to participate in genetic research (including those from 23andMe and others whose data has been released as part of publicly available datasets). Additionally, we recognize the 49 historic-era individuals from Historic St. Mary's City, who could not directly consent to participate in this study. This research was conducted following all applicable ethical guidelines for the study of historical genomes^{33–35} and was conducted under the auspices of the HSMC, who were consulted over the course of the study's design and execution. The skeletal remains from the 49 St. Mary's City individuals who were sampled as part of this analysis were reinterred with a religious ceremony inside the

reconstructed Brick Chapel on September 20th, 2025. HSMC has consulted with the descendant community of the early colonists (The Society of the Ark and the Dove) since the beginning of the Chapel project and remains actively engaged in the research effort. Additionally, HSMC shared findings with representatives from several local groups, including the Unified Committee on Afro-American Contributions, National Association for the Advancement of Colored Peoples, and St. Mary's College of Maryland.

Provenance of historical skeletal remains

The skeletons of 65 individuals were documented from excavated single burials within and surrounding the Great Brick Chapel and an ossuary of remains displaced during Chapel construction. Fifty of the 65 individuals were sampled for aDNA analysis as authorized by HSMC, including Philip Calvert, Anne Wolseley Calvert and the Calvert Son. All remains were reinterred at the Chapel site in 2025, in a climate controlled vault within the reconstructed structure and will be accessible for future scientific studies.

Present-day research participants

We compared the genomes of the St. Mary's individuals to data from 11,524,442 research participants who were genotyped by April 26th, 2023 by 23andMe Research Institute, formerly 23andMe Inc. All research participants provided informed consent and answered relevant research questions online via a protocol approved by the external AAHRPP-accredited IRB, Ethical & Independent Review Services (E&I Review), which has since been renamed Salus IRB (<https://www.versiticlinicaltrials.org/salusirb>). Research participants were included in the analysis on the basis of consent status, which was checked at the time data analyses were initiated. We also included data from the 1000 Genomes Project³⁶ and the People of the British Isles Project²⁰ in these comparisons (Data S7M). When possible, we report results for IBD analyses performed on these datasets in the supplementary materials in order to enable replication of the study results.

We assigned each research participant to a specific geographic location based on either self-reported [1] birth location of all four grandparents (when available and when all four grandparents shared the same birth location) or [2] participant birth location. In cases where participant birth location and grandparental birth location were the same, and where participant birth location provided more fine grained resolution, we favored participant birth location. Otherwise, we favored grandparental birth locations.

We assigned research participants into a variety of cohorts based on their associated location and in some cases, their genetic ancestry, as assigned by the tool Ancestry Composition.³⁷ For instance, the US, GB and IE cohorts are comprised of all participants associated with the US, Great Britain and Ireland, respectively, while the European cohort is comprised of all participants associated with the following European countries (as indicated by ISO2 country codes: AL, AD, AM, AT, BY, BE, BA, BG, CH, CY, CZ, DE, DK, EE, ES, FO, FI, FR, GB, GE, GI, GR, HU, HR, IE, IS, IT, LI, LT, LU, LV, MC, MD, ME, MK, MT, NO, NL, PL, PT, RO, RS, RU, SE, SI, SK, SM, TR, UA, VA, and XK) with at least 99% European ancestry.

METHOD DETAILS

Ancient DNA (aDNA) Sequencing

The remains of Philip Calvert and his infant son were sampled for aDNA in 2015 and the results of this sequencing analysis were described in a report⁸ that accompanied the public release of their genome-wide data. For all other St. Mary's individuals, we sampled bone powder from the petrous portion of the temporal bone using a minimally destructive cranial base drilling approach.³⁸ Sampling occurred in 2018 and 2022 following curatorial guidelines in place at the time.

aDNA sequencing was performed in dedicated aDNA facilities at Harvard Medical School, which were designed to minimize the likelihood of contamination.¹⁰ We followed published protocols to extract aDNA from ~37 mg of bone powder,³⁹ creating dual-bar-coded, double-stranded, partially uracil-DNA glycosylase (UDG) treated libraries.^{40,41} We performed targeted enrichment capture to maximize the proportion of sequences that align to ~1.24 million (1240k) single nucleotide polymorphisms (SNPs) in the nuclear genome in addition to the entire mitochondrial genome.^{42–45} After adding unique indexing barcodes (7 base pairs) to both ends of each DNA molecule, we sequenced the enriched libraries (along with a small amount of the unenriched libraries, to be used for quality assessment) for 2x10¹ cycles on an Illumina HiSeqX10 or 2x76 cycles on an Illumina NextSeq500 instrument, with 2x7 cycles for indexing.

After sequencing, we used custom software (<https://github.com/DReichLab/ADNA-Tools>) to trim the molecular adapters and barcodes from the ends of each sequenced read, and then merged paired-end reads—requiring 15 base pair overlap with no more than three mismatches in low quality bases (<20) or 1 mismatch in high quality bases (≥20). Next, we used samse in BWA (v0.6.1)²⁵ to map to the mitochondrial consensus sequence (RSRS)⁴⁶ and the human reference genome (hg19; https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000001405.13/). After aligning, we removed duplicates—defined as reads with identical start and end positions, orientation, and barcodes—retaining the highest quality sequence among the duplicates.

To determine whether the sequenced DNA was likely to be authentic to the St. Mary's individuals, we considered 3 standard criteria for aDNA authenticity. (1) We assessed the cytosine-to-thymine substitution rate at the 5' end of each molecule, requiring a rate of at least 3% to be considered authentic aDNA. (2) We assessed the mitochondrial match-to-consensus rate, requiring a minimum match rate of 95%. (3) To exclude individuals with evidence of contamination derived from a source of the opposite genetic sex, we required that the number of reads aligning to the Y chromosome divided by the total number of reads aligning to either the X or Y chromosome fall either below 2% (consistent with a genetically female individual) or above 33% (consistent with a genetically male individual).⁴⁷ (4) For genetically male individuals, we estimated a contamination rate based on the polymorphism rate on the X chromosome (which

should have no variation in genetically male individuals who are expected to only have one) and considered a rate over 3% to be indicative of substantial contamination.

We flagged four low-coverage individuals as potentially contaminated based on their mitochondrial match-to-consensus rate. Additionally, we flagged an additional five individuals as having very low coverage ($<0.1\times$). We report individual level results for these samples, but excluded them from all combined analyses of the St. Mary's individuals (along with all other St. Mary's individuals with average chromosomal coverage below $1\times$).

Uniparental Haplogroups

For all individuals, we determined mitochondrial haplogroups using haplogrep2²⁶ and Phylotree version 17, by considering all reads that aligned to the RSRS mitochondrial genome with $\text{MAPQ} \geq 30$ and base quality ≥ 20 . For genetically male individuals, we called Y-chromosome haplogroups based on the most derived mutation identified among reads aligning to the Y-chromosome with $\text{MAPQ} \geq 30$ and base quality ≥ 30 , using the nomenclature defined by the International Society of Genetic Genealogy (ISOGG) (<http://www.isogg.org>) version 14.76 (updated April 2019).

Publicly available dataset assembly

To compare the St. Mary's dataset to other publicly available datasets, we created a pseudo-haploid dataset by randomly sampling a single sequence that aligned to each of the 1240k SNP positions targeted during enrichment capture and assigning a diploid genotype based on the allele observed at that position. We merged this pseudo-haploid dataset with data from the Allen Ancient DNA Resource (Version 9)^{36,48–59} creating two merged datasets with information at 1,233,013 SNPs (the “1240k dataset”) and 584,131 SNPs (the “Human Origins dataset”), respectively.

Genetic relatedness

We considered patterns of allele sharing between pairs of individuals to identify 1st–3rd-degree relative pairs, using the approach introduced in Olalde et al.⁴⁸ Additionally, we considered IBD shared between pairs of individuals (using the approach described below for comparisons of a single St. Mary's individual to 23andMe research participants) to identify more distant relationships among the St. Mary's individuals.

ADMIXTURE

Using the clustering tool ADMIXTURE³⁰ we assigned the ancestry of the St. Mary's individuals and four present-day populations (YRI.DG, CEU.DG, Pima.DG, ASW.DG) to one or more of three theoretical ancestral populations, which broadly correspond to African, European and Indigenous American ancestry (Figure S1A). Using the 1240k dataset, we pruned SNPs in linkage disequilibrium from the dataset using the parameters—`indep-pairwise 200 25 0.4`, resulting in a dataset with 462,447 SNPs. Then we ran ten replicates of ADMIXTURE with $k=3$, and reported the results of the replicate with the highest likelihood.

PCA

To understand the biogeographic ancestry of the St. Mary's individuals, we performed principal components analysis (PCA) using the tool smartpca,²⁹ projecting the St. Mary's individuals onto PCA plots generated using the Human Origins dataset with the following parameters: `lsqproject: YES` and `shrinkmode: YES` (Figures S1B–S1D). First, to explore their continental-level ancestry, we projected them onto a PCA generated with 162 populations from around the world. Next to localize the ancestry of the 49 European-related St. Mary's individuals, we projected them onto a PCA generated with 60 populations from Europe and the Near East. Finally, to assess whether the single African-related St. Mary's individual had any European or Indigenous American and/or East Asian ancestry, we projected all of the St. Mary's individuals onto a PCA generated from four populations with West African (YRI.DG), European (CEU.DG), African American (ASW.DG) and East Asian (CHB.DG) ancestry.

Imputation

We generated diploid genotype calls at all biallelic SNP positions for all of the St. Mary's individuals with at least $1\times$ average chromosomal coverage using bcftools mpileup.²⁵ We then performed imputation with the tool GLIMPSE (v1.0.0),²⁷ using the phase 3 1000 Genomes reference dataset,³⁷ filtering out genotype calls with an estimated maximum genotype posterior below 95%. We used EAGLE²⁸ to rephase the imputed dataset using a reference panel of either 691,759 research participants genotyped at 454,507 SNPs on the version 1–4 23andMe genotyping platform or 706,995 research participants genotyped at 541,948 SNPs on the version 5 23andMe genotyping platform, with default settings and optional parameters `–allowRefAltSwap` and `–noImpMissing`.

Identity by Descent

We used the templated positional Burrows–Wheeler transform (TPBWT) IBD detection tool³¹ to identify IBD segments shared between the imputed St. Mary's dataset and 23andMe research participants. We used default parameters, including `use_phase_correction = True`, with `L_m = 300` (requiring that matching subsegments span at least 300 SNPs to be considered part of an IBD segment) and `missing_site_threshold = 10` (allowing up to 10 consecutive missing SNPs within a segment). While TPBWT outputs

segments as short as 3.0 cM by default, we applied minimum length filters to reduce false positives. Specifically, we retained only segments ≥ 6 cM for individuals with 2–5x average chromosomal coverage, and segments ≥ 9 cM for individuals with 1–2x coverage (Data S3).³

We generated summary statistics that group research participants based on associated location and in some cases, ancestry, to explore geographic patterns of IBD sharing (Data S7C–S7E and S7G–S7I; Figures S2 and S3). To ensure participant anonymity, we rounded the maximum total IBD shared and maximum IBD segment lengths to the nearest digit when IBD sharing is < 10 cM, 0 digits when IBD sharing is between 10–30 cM, to the nearest 5 when IBD sharing is between 30–100 cM, and to the nearest 10 when IBD sharing is over 100 cM. All reported averages involve a minimum of 5 participants unless otherwise stated.

For US maps, we report results for participants in the US cohort who reported county-level geographic information (Data S7E). In the European maps, we report results for participants in the European cohort who reported state-level geographic information (Data S7G; Figure S4). To ensure research participant anonymity, we only report coordinate level results for geographic coordinates that have at least 25 associated participants. Since in the majority of cases fewer than five research participants at each coordinate share IBD with one or more of the St. Mary's individuals, we randomly downsampled the research participant panel and reported results for 80% of research participants, enabling us to report counts fewer than 5 while abiding by 23andMe privacy protection requirements.

In comparisons between pairs of St. Mary's individuals, we used the lower coverage individual when determining the minimum length threshold (Data S7B). Only comparisons where both individuals have over 1x coverage are shown.

IBD Networks

To identify St. Mary's individuals' distant and recent connections to genetic groups (Data S7H), we first performed community detection on contemporary participants using planted partition stochastic block models (SBMs).¹³ SBMs are generative graph models that partition individuals into communities called blocks, based on the total amount of IBD shared with each other with the probability of an edge existing between individuals depending on block membership. We applied this method to individuals who reported all four of their grandparents were born in either Great Britain or Ireland and had $\geq 99\%$ northern European ancestry ($n = 7,872$). Individuals included in this cohort were filtered so that no two individuals share ≥ 700 cM. Individuals were clustered with SBMs using a minimum IBD segment length of 5.5 cM and were required to have a $\geq 99\%$ probability of being assigned to a block to be retained. After identifying genetic clusters across Great Britain and Ireland, we determined the average amount of IBD each St. Mary's individual shares with each of the groups. We used these data to display each St. Mary's individual's connections to each group in the form of a graph layout. To achieve this, we first arranged groups within each of the cohorts, using the Force Atlas layout. Force Atlas is an algorithm that situates groups (or nodes) in a graph using a physical “magnetic” model. In this case, groups with more IBD sharing will be attracted to one another and groups with less IBD sharing are repelled. Force Atlas runs until balance between repulsion and attraction is achieved, essentially illustrating the structure of groups via their IBD sharing. After Force Atlas was run on each of the participant cohorts, we independently ran Force Atlas between each St. Mary's individual and the pre-arranged graph of participants, projecting St. Mary's individuals onto the structure of each cohort, thus illustrating where St. Mary's individuals physically fell into the structure of each cohort.

Individual-Level Analyses

The subjects of the individual-level pedigree-based analyses provided additional informed consent and shared information about their genealogical connections to 17th-century St. Mary's City online via a protocol approved by the external AAHRPP-accredited IRB, described above. Genealogical information was shared via a secure online folder to which only the participants and the lead author (E.H.) had access. The lead author then used this information to reconstruct genealogical trees for each of the consenting participants, showing lineages of the tree that could be traced to St. Mary's City and surrounding areas. Publicly available genealogical databases (e.g., [FamilySearch.org](https://www.familysearch.org)) were used to validate and add to the trees when possible.

To protect participant privacy, the lead author created masked versions of the pedigrees that excluded personal information—including name, age, sex, and birth and death location—about all individuals included in the trees who were born after 1850. The masked reconstructed pedigrees were then exported from the secure online environment to be used in subsequent analyses and included in publication, in accordance with the IRB approved protocol. To further protect participant privacy, when describing the IBD connections detected between participants whose data were included in these analyses we shared only the number and length of IBD segments shared, while masking the exact genomic location of these segments.

We searched for overlapping names in the reconstructed genealogical trees of participants who shared over 30 cM of IBD to the same sequenced St. Mary's individuals. Individuals from these parts of the genealogical trees of the same sex as the sequenced individuals were considered to be potential candidates for their identities. For all candidates, we considered whether the age of the burial (estimated based on the position of the burial, relative to the foundations of the Brick Chapel), grave fill content, shaft orientation, and the age at death of the sequenced individuals (estimated based on morphological analysis) was consistent with the reported date and age of death of the individuals named in the genealogical trees. In cases where a match seemed plausible, we expanded our analysis to also consider whether any of the sequenced individuals had genetic relatives who could be associated with individuals named in the genealogical trees.

QUANTIFICATION AND STATISTICAL ANALYSIS

Randomization Tests

We used randomization tests to determine whether observed patterns of IBD sharing were statistically significant or if they could instead be explained by random chance. For each randomization test, we counted the number of participants who shared a given amount of IBD (either any amount of IBD or at least 30 cM) with one or more of the St. Mary's individuals from a given region of interest and compared this value to the number of participants with IBD sharing in a random sample of participants that was equal in size to the number of participants from the region of interest (Data S7F). We sampled participants from either the US or European cohorts, restricting the cohorts to participants with at least 99% European ancestry. We performed 1000 replicates of each randomization test and calculated *p-values* by determining the proportion of replicates where the observed number of participants with IBD sharing exceeded that of the actual region of interest. We considered all *p-values* <0.01 to be statistically significant, reflecting a Bonferroni correction for the five independent tests performed (0.05/5).

Chi Squared and T-Tests

We performed chi square tests to compare the proportion of participants who share IBD with the St. Mary's individuals between those who answered "Yes" and "No" to the question "Can you trace your relatives back to 17th-century St. Mary's City, Maryland's first colonial settlement?" (Data S7J). To ensure these differences were not solely driven by broad population structure, we applied this test to both the full US cohort and a subset restricted to participants with at least 99% European ancestry. Additionally, we used independent t-tests to evaluate differences in the mean amount of shared IBD (in cM) between the two response cohorts. We considered *p-values* <0.05 to be statistically significant.

Current Biology, Volume 36

Supplemental Information

The genetic legacy of the 17th-century colonial capital of St. Mary's City

Éadaoin Harney, Ethan Jewett, Steven Micheletti, Roslyn Curry, Karin S. Bruwelheide, William A. Freyman, Henry Miller, Ali Akbari, Kathryn Barca, Katarzyna Bryc, Shaeloren Deering, Samantha Ancona Esselmann, Kira Kalkus, Aisling Kearns, Alexander Moran, Dominique T. Nguyen, Iñigo Olalde, Jakob Sedig, Kendra Sirak, Kim Callan, Lora Iliev, Lijun Qiu, J. Noah Workman, Matthew Mah, Gregory Soos, 23andMe Research Team, Swapan Mallick, Nadin Rohland, Joanna L. Mountain, Douglas W. Owsley, and David Reich

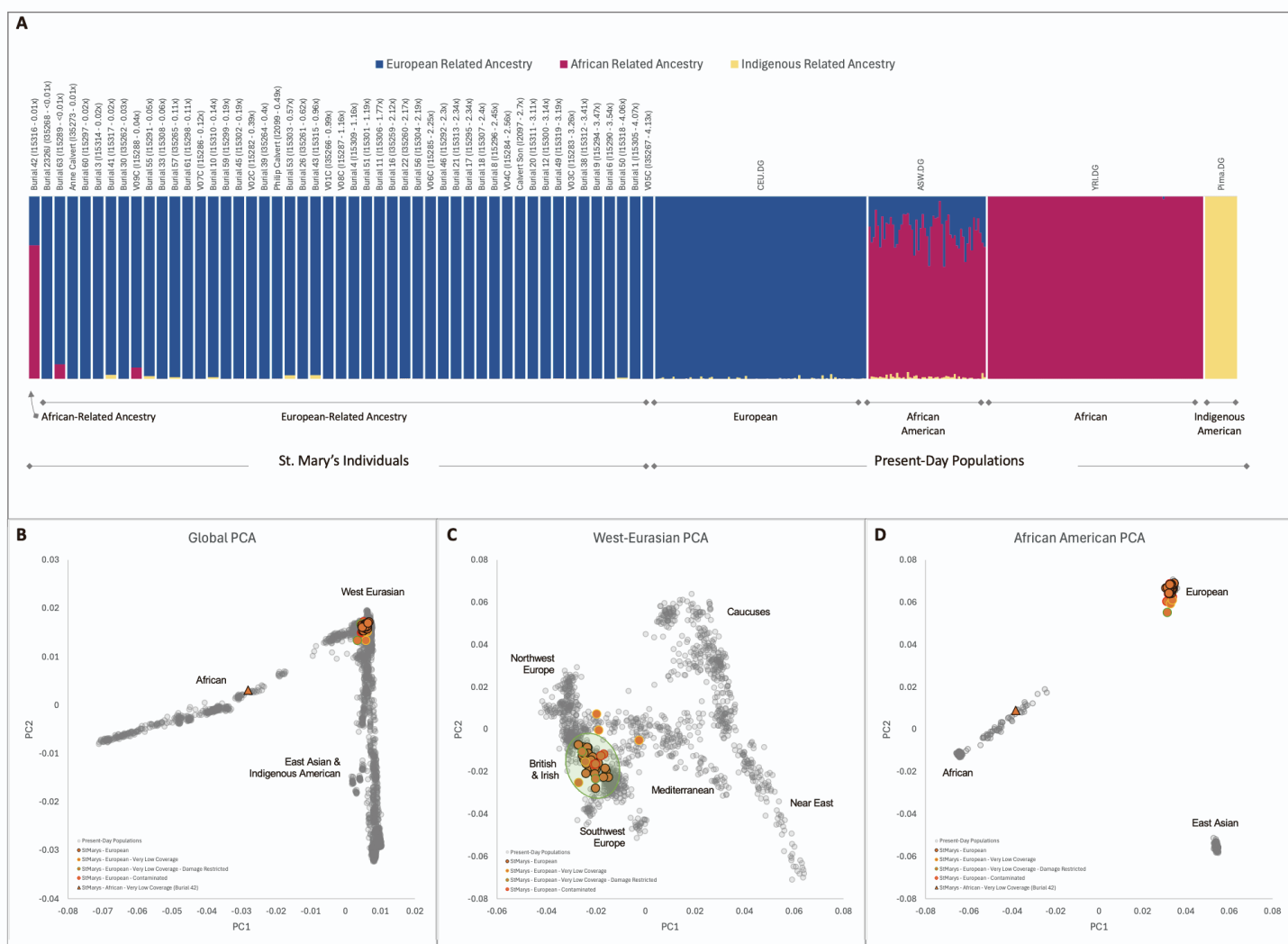
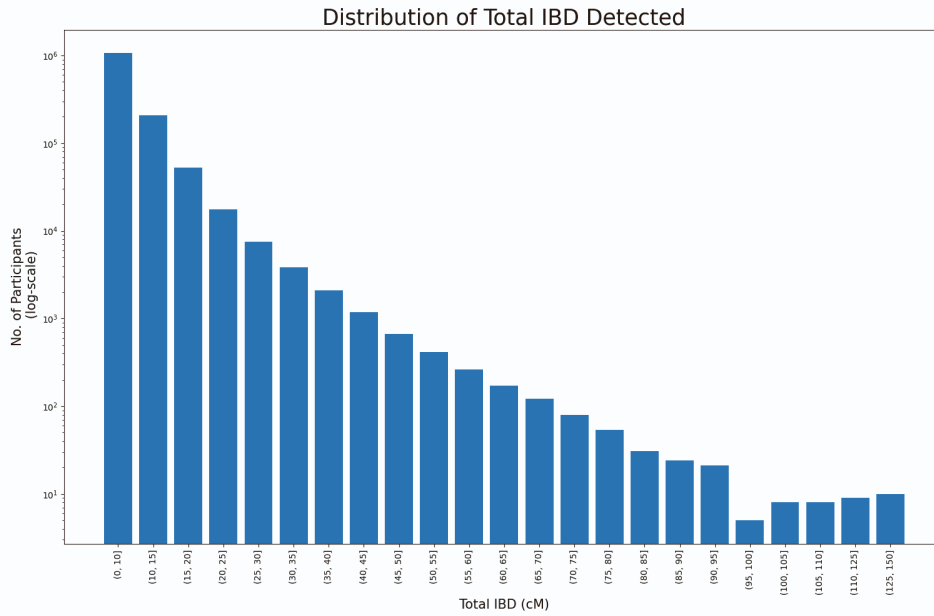


Figure S1. St. Mary's Individuals Ancestry Analysis. Related to STAR Methods. (A) ADMIXTURE analysis of the St. Mary's individuals with present-day individuals with European, African American, African and Indigenous American ancestry ($k=3$). The St. Mary's individuals are ordered from lowest to highest coverage, except individual I15316, who has substantial African ancestry and is shown on the far left. (B) Global PCA. The St. Mary's individuals projected onto a PCA generated with present-day individuals from across the globe. The approximate location of present-day populations with West Eurasian, African and East Asian and Indigenous American ancestry are indicated. (C) West-Eurasian PCA. The St. Mary's individuals projected onto a PCA generated with present-day individuals from Europe and the Near East. The green circle indicates the approximate location where present-day individuals from Great Britain and Ireland plot on the PCA. (D) African American PCA. The St. Mary's individuals projected onto a PCA generated with present-day individuals from representative populations with European (CEU.DG), African (YRI.DG), African American (ASW.DG) and East Asian (CHB.DG; not shown) ancestry.

A



B

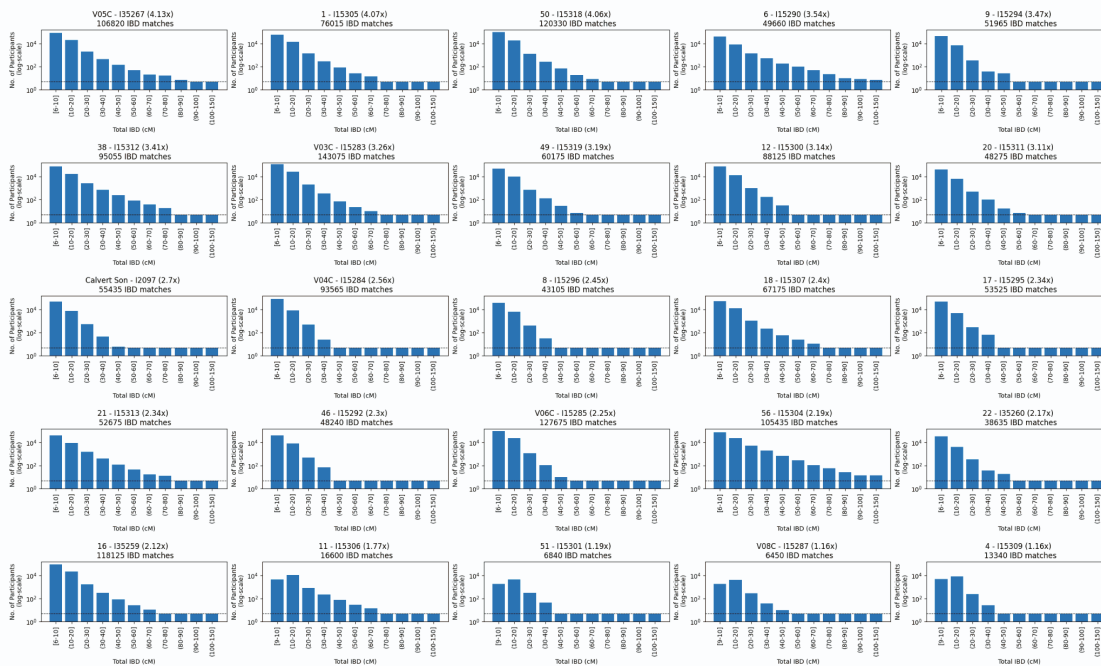


Figure S2. Total IBD sharing with St. Mary’s individuals. Related to STAR Methods. (A) Histograms show the number of research participants that share a given amount of total IBD with any St. Mary’s individual. In cases where research participants share IBD with multiple St. Mary’s individuals, the largest amount of total IBD is reported. Bin sizes were selected to include a minimum of 5 research participants. (B) Histograms show the number of research participants that share a given amount of total IBD with each St. Mary’s individual. To maintain research participant anonymity, bins with five or fewer associated research participants are reported as 5, indicated by the black, dotted horizontal line. The minimum IBD segment size considered for historical individuals with coverage of greater than or equal to 2x is 6 cM and 9 cM for historical individuals with coverage less than 2x. The number of IBD matches reported in each figure title is rounded to the nearest 5.

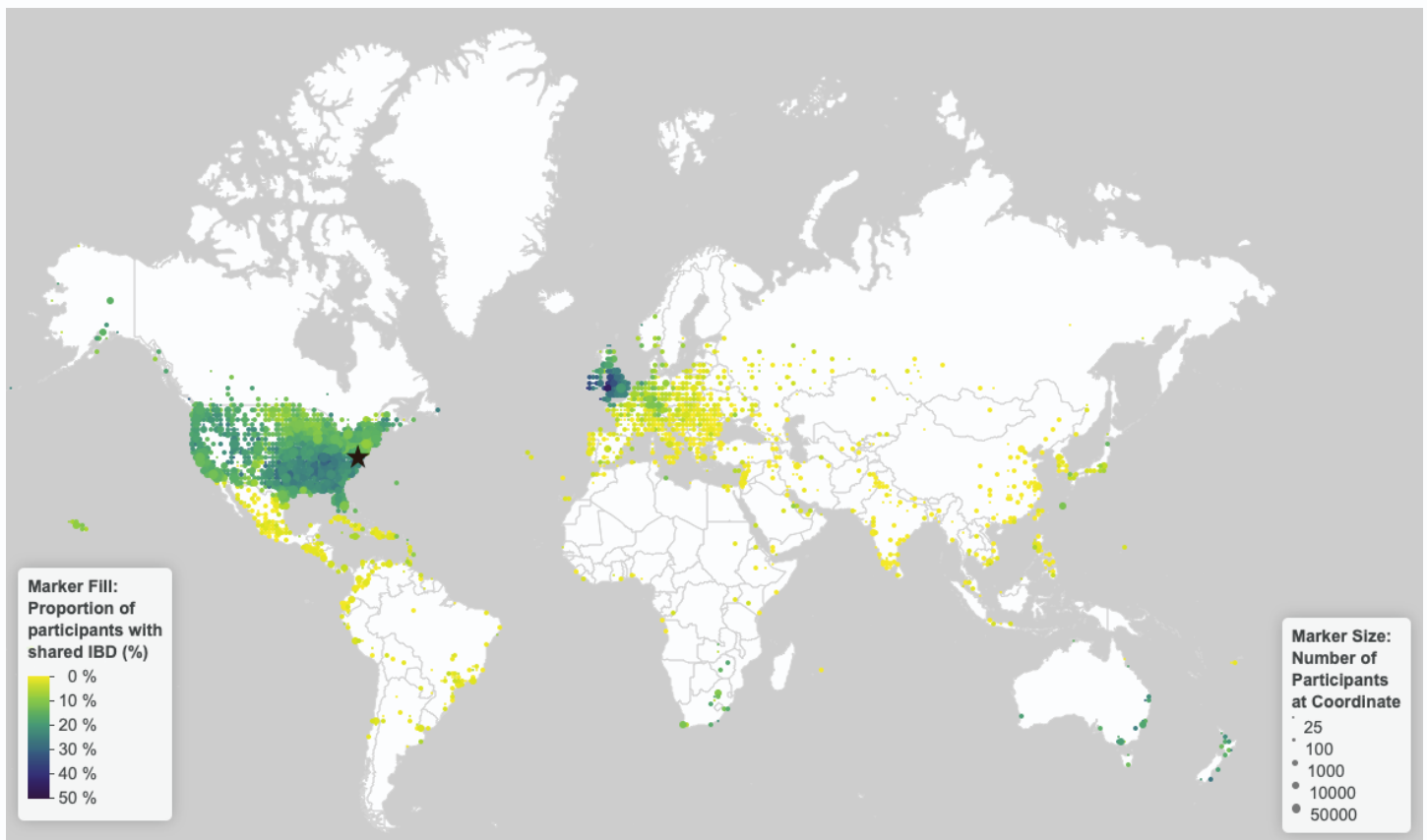


Figure S3. Global Geographic distribution of IBD sharing with St. Mary's individuals. Related to STAR Methods. Marker color indicates the proportion of research participants at each geographic coordinate who share IBD with one or more of the St. Mary's individuals. The size of each marker represents the number of participants at the given geographic coordinate (rounded to the nearest integer). To protect participant privacy, we randomly downsampled to include only 80% of participants and only showed results for coordinates with at least 25 associated participants. The star indicates the location of St. Mary's City, Maryland.

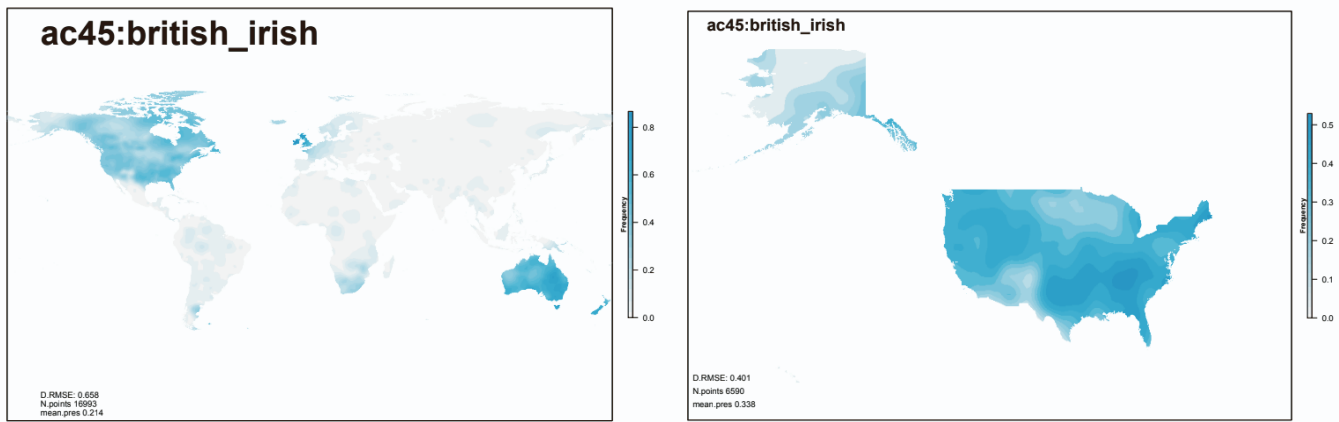


Figure S4. Geographic distribution of 23andMe research participants with British & Irish ancestry. Related to Figure 3. Maps show the predicted British & Irish ancestry proportions (left) globally and (right) in the US using the grandparent birth location of 23andMe research participants. We used Kriging interpolation of the mean proportion of assigned British & Irish ancestry at each geographic coordinate to generate the prediction surfaces.

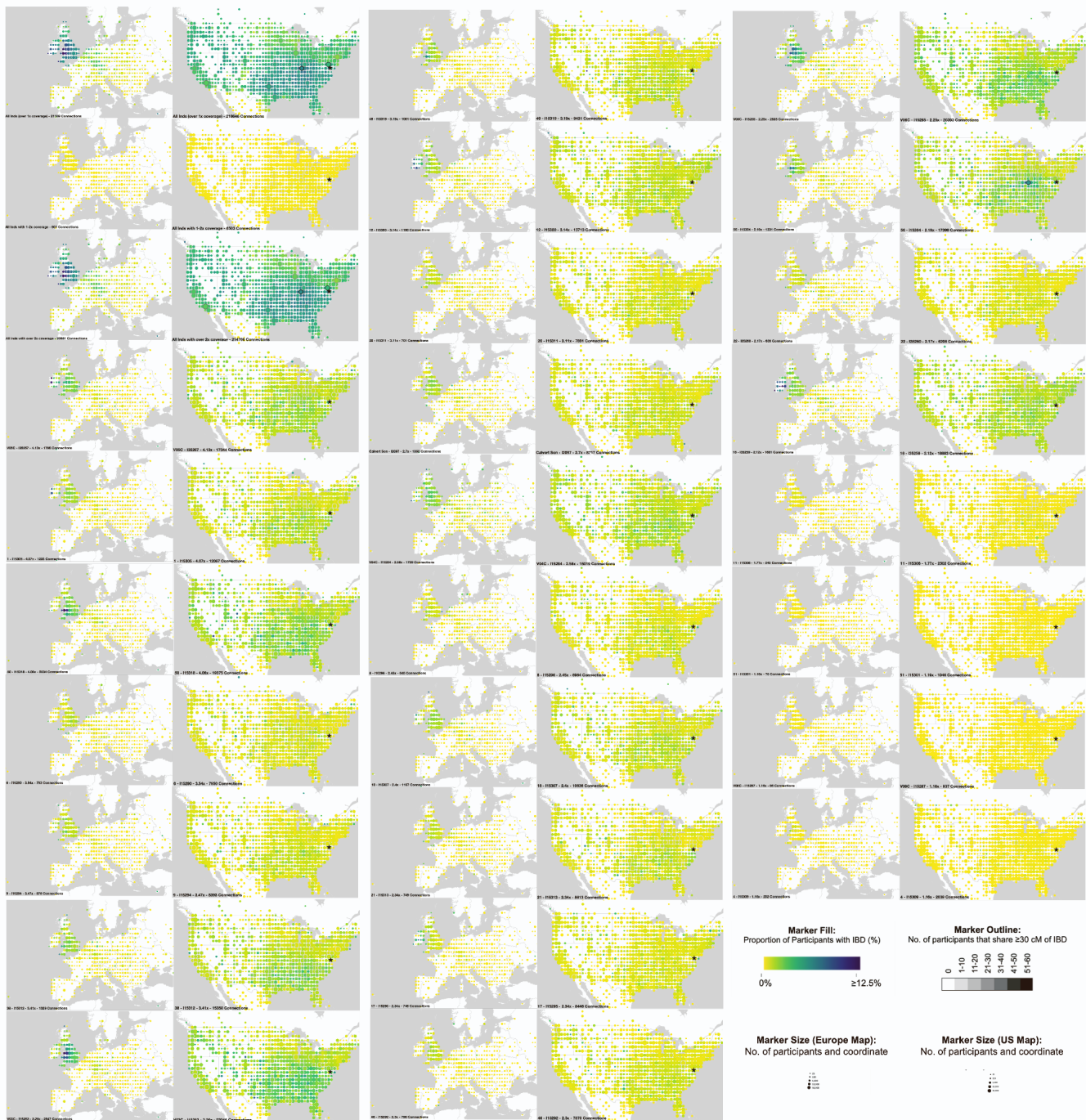


Figure S5. Geographic distribution of IBD sharing with St. Mary's individuals. Related to Figures 3-4. For each St. Mary's individual with over 1x coverage (grouped by coverage and displayed separately), we show the proportion (as indicated by the color of the marker) of research participants at each geographic coordinate who share any amount of IBD. The size of each marker represents the number of participants at the given geographic coordinate (rounded to the nearest integer). To protect participant privacy, we randomly downsampled to include only 80% of participants and only showed results for coordinates with at least 25 associated participants. The total number of participants who share IBD with the St. Mary's individual who are associated with geographic coordinates included in the image is indicated in the figure label, along with the individual ID and their average chromosomal coverage. For each St. Mary's individual we report IBD sharing with (left) research participants with at least 99% European ancestry throughout western and central Europe and (right) research participants in the US. The color of the marker outlines indicates the number of participants at each location who share at least 30 cM of IBD with the St. Mary's individual. The figure title (bottom left of each image) indicates how many research participants shared IBD with (at least one of) the St. Mary's individuals shown.

Supplemental references

- S1. Hurry, S.D. (2001). --once the Metropolis of Maryland": The History and Archaeology of Maryland's First Capital (Historic St. Mary's City Commission).
- S2. Horn, J. (1979). Servant Immigration to the Chesapeake in the Seventeenth Century. In *The Chesapeake in the Seventeenth Century: Essays on Anglo-American Society and Politics*. Edited by Thad W. Tate and David L. Ammerman. (University of North Carolina Press), pp. 51–95.
- S3. Horn, J. (1994). *Adapting to a New World English Society in the Seventeenth-Century Chesapeake* (University of North Carolina Press).
- S4. Menard, R. (1988). British Migration to the Chesapeake Colonies in the Seventeenth Century. In *Colonial Chesapeake Society*, Edited by Lois Green Carr, Philip D. Morgan and Jean B. Russo. (University of North Carolina Press), pp. 99–132.
- S5. Hall, C.C. (1910). A Relation of Maryland. In *Narratives of Early Maryland* (Barnes and Noble), pp. 70–77.
- S6. Carr, L.G., Menard, R., and Peddicord, L. (1984). *Maryland At the Beginning* (Maryland State Archives).
- S7. Miller, H.M., and Hurry, S.D. (2021). The Archaeology of the Maryland Indians at St. Mary's City and the Interaction of Cultures. In *Unearthing St. Mary's City Fifty Years of Archaeology at Maryland's First Capital* (University Press of Florida), pp. 95–112.
- S8. Seib, R., and Rountree, H.C. (2014). *Indians of Southern Maryland* (Maryland Historical Society).
- S9. Bruwelheide, K.S., Miller, H.M., Owsley, D.W., and Riordan, T.B. (2021). The Lead Coffins of St. Mary's: Burials of the Elite in the Early Chesapeake. In *Unearthing St. Mary's City: Fifty Years of Archaeology at Maryland's First Capital* (University Press of Florida).
- S10. Riordan, T.B., Hurry, S.D., and Miller, H.M. (1995). "A Good Brick Chappell": The Archaeology of the c.1667 Catholic Chapel at St. Mary's City. (Department of Research, Historic St. Mary's City).
- S11. Riordan, T.B. (2000). *Dig a Grave Both Wide and Deep: An Archaeological Investigation of Mortuary Practices in the 17th-Century Cemetery at St. Mary's City, Maryland* (Historic St. Mary's City).
- S12. Newman, H.W. (1984). *The Flowering of the Maryland Palatinate: An Intimate and Objective History of the Province of Maryland to the Overthrow of Proprietary Rule in 1654, with Accounts of Lord Baltimore's Settlement at Avalon* (Genealogical Publishing Com).
- S13. Harney, É., Micheletti, S., Bruwelheide, K.S., Freyman, W.A., Bryc, K., Akbari, A., Jewett, E., Comer, E., Gates, H.L., Heywood, L., et al. (2023). The genetic legacy of African Americans from Catoctin Furnace. *Science* 381, eade4995. <https://doi.org/10.1126/science.ade4995>.
- S14. Micheletti, S.J., Bryc, K., Ancona Esselmann, S.G., Wilton, P.R., and Freyman, W.A. (2025). Bayesian inference of population structure using identity-by-descent-based stochastic block models. Preprint, <https://doi.org/https://doi.org/10.1101/2025.11.26.690810>
<https://doi.org/https://doi.org/10.1101/2025.11.26.690810>.
- S15. Huff, C.D., Witherspoon, D.J., Simonson, T.S., Xing, J., Watkins, W.S., Zhang, Y., Tuohy, T.M., Neklason, D.W., Burt, R.W., Guthery, S.L., et al. (2011). Maximum-likelihood estimation of recent shared ancestry (ERSA). *Genome Res.* 21, 768–774. <https://doi.org/10.1101/gr.115972.110>.

- S16. Jewett, E.M., and Team, the 23andMe R. (2024). Correcting model misspecification in relationship estimates. Preprint at bioRxiv, <https://doi.org/10.1101/2024.05.13.594005>
<https://doi.org/10.1101/2024.05.13.594005>.
- S17. Thomas, A., Skolnick, M.H., and Lewis, C.M. (1994). Genomic mismatch scanning in pedigrees. *Mathematical Medicine and Biology: A Journal of the IMA* 11, 1–16. <https://doi.org/10.1093/imammb/11.1.1>.
- S18. Jewett, E.M., McManus, K.F., Freyman, W.A., and Auton, A. (2021). Bonsai: An efficient method for inferring large human pedigrees from genotype data. *The American Journal of Human Genetics* 108, 2052–2070. <https://doi.org/10.1016/j.ajhg.2021.09.013>.
- S19. Henn, B.M., Hon, L., Macpherson, J.M., Eriksson, N., Saxonov, S., Pe'er, I., and Mountain, J.L. (2012). Cryptic Distant Relatives Are Common in Both Isolated and Cosmopolitan Genetic Samples. *PLOS ONE* 7, e34267. <https://doi.org/10.1371/journal.pone.0034267>.
- S20. Seidman, D.N., Shenoy, S.A., Kim, M., Babu, R., Woods, I.G., Dyer, T.D., Lehman, D.M., Curran, J.E., Duggirala, R., Blangero, J., et al. (2020). Rapid, Phase-free Detection of Long Identity-by-Descent Segments Enables Effective Relationship Classification. *The American Journal of Human Genetics* 106, 453–466. <https://doi.org/10.1016/j.ajhg.2020.02.012>.
- S21. Freyman, W.A., McManus, K.F., Shringarpure, S.S., Jewett, E.M., Bryc, K., the 23andMe Research Team, and Auton, A. (2020). Fast and robust identity-by-descent inference with the templated positional Burrows-Wheeler transform. *Molecular Biology and Evolution*. <https://doi.org/10.1093/molbev/msaa328>.
- S22. Dr. Lois Green Carr, Men's Career Files MSA SC 5094 Image No: sc5094-1628-5
<https://msa.maryland.gov/megafile/msa/speccol/sc5000/sc5094/001000/001628/html/sc5094-1628-5.html>.
- S23. Archives of Maryland, Volume 0010, Page 0088 - Judicial and Testamentary Business of the Provincial Court, 1649/50-1657
<https://msa.maryland.gov/megafile/msa/speccol/sc2900/sc2908/000001/000010/html/am10--88.html>.
- S24. Newman, H.W. (1984). *The Flowering of the Maryland Palatinate: An Intimate and Objective History of the Province of Maryland to the Overthrow of Proprietary Rule in 1654, with Accounts of Lord Baltimore's Settlement at Avalon* (Genealogical Publishing Com).
- S25. Carr, L.G., Menard, R.R., and Walsh, L.S. (2017). *Robert Cole's World: Agriculture and Society in Early Maryland* (UNC Press Books).