nature genetics

ARTICLES

https://doi.org/10.1038/s41588-017-0031-6

Reconstructing an African haploid genome from the 18th century

Anuradha Jagadeesan[1,2], Ellen D. Gunnarsdóttir[1], S. Sunna Ebenesersdóttir[1,2], Valdis B. Guðmundsdóttir[1,2], Elisabet Linda Thordardottir[1], Margrét S. Einarsdóttir[1,2], Hákon Jónsson[1], Jean-Michel Dugoujon[3], Cesar Fortes-Lima[3], Florence Migot-Nabias[4,5], Achille Massougbodji[6,7], Gil Bellis[8], Luisa Pereira[9,10,11], Gísli Másson[1], Augustine Kong[1], Kári Stefánsson[1,12]* and Agnar Helgason[1,2]*

A genome is a mosaic of chromosome fragments from ancestors who existed some arbitrary number of generations earlier. Here, we reconstruct the genome of Hans Jonatan (HJ), born in the Caribbean in 1784 to an enslaved African mother and European father. HJ migrated to Iceland in 1802, married and had two children. We genotyped 182 of his 788 descendants using single-nucleotide polymorphism (SNP) chips and whole-genome sequenced (WGS) 20 of them. Using these data, we reconstructed 38% of HJ's maternal genome and inferred that his mother was from the region spanned by Benin, Nigeria and Cameroon.

The Icelandic population was founded by settlers from Scandinavia and the British Isles around 1,100 years ago, and remained relatively isolated until recently[1]. Post-settlement immigration to Iceland was rare, occurring mostly from Denmark and to a lesser degree from other neighboring countries[2]. Records from 1930 to 1980 show that the number of individuals in Iceland from outside of Europe ranged from 73 to 1,322 (0.07% to 0.6% of the population)[2], and it is likely to have been even smaller at the beginning of the nineteenth century. Because of his African ancestry, HJ was an unusual immigrant when he arrived in Iceland in 1802, after departing from Copenhagen. Historical sources indicate that HJ's mother was African and his father European[3]. Accordingly, the chromosomes transmitted by HJ to his two children would have been a mosaic of African-derived and European-derived fragments. As his wife was Icelandic, their children would have inherited African fragments only from HJ. Owing to recombination and Mendelian segregation, we expect to find fewer and smaller fragments from HJ in successive generations of descendants. When chromosome fragments of one ancestor (such as HJ) can be distinguished from those of other ancestors among a group of descendants, his or her genome can be at least partially reconstructed. Thus, assuming that African chromosome fragments can be reliably identified in genotyped descendants of HJ from Iceland and that these individuals have no other recent African ancestors, the reconstruction of HJ's maternal genome amounts to identifying and joining these fragments—like pieces in a jigsaw puzzle. To our knowledge, apart from HJ, there is no evidence of African gene flow to Iceland before the twentieth century. Therefore, we expect African chromosome fragments to be rare in the Icelandic gene pool.

Results

Identifying African fragments in HJ's descendants. A range of 4–8 generations separates HJ from his 182 genotyped descendants (see Fig. 1), of whom 121 are terminal nodes in the resultant genealogical tree and 61 are intermediate nodes (parents, grandparents or great-grandparents of terminal-node descendants). Using HAPMIX[4] and phased source population genotype data from the HapMap project[5], we identified 674 putative African chromosome fragments in the 182 genotyped descendants, ranging in length from 0.78 Mb to 82.98 Mb, with a mean length of 13.7 Mb (234 SNPs per megabase on average) and a combined length of 9,231.8 Mb (Table 1). Because of the relative isolation of Icelanders[6], genuine African fragments in descendants of HJ are likely to originate only from HJ. To verify this assumption, we ran ADMIXTURE[7] and HAPMIX for all our chip-typed Icelanders (151,014 individuals; see Methods). The results reveal that among individuals born between 1880 and 1930, only descendants of HJ have evident (>1%) African ancestry. Among those born after 1930 and not descended from HJ, we identified 320 with evident African ancestry, none of whom were ancestors of HJ's descendants[8].

Among the 674 African fragments in HJ descendants, 64 were excluded because they were found to be shared (at a minimum of 1,000 SNPs, corresponding to 5.5 centimorgans (cM)) with one or more (mean = 215, s.d. = 373) of the 150,832 chip-typed Icelanders who are not descended from HJ. Such fragments are unlikely to be derived from HJ and are probably not truly African. The remaining 610 African fragments were validated using the genealogy of HJ. First, it is expected that the parent of origin of any such fragment (determined via long-range phasing[9]) would be consistent with the genealogical path to HJ. Using this criterion, we deemed 9 of 610

[1]deCODE Genetics/Amgen, Reykjavik, Iceland. [2]Department of Anthropology, University of Iceland, Reykjavik, Iceland. [3]Laboratoire d'Eco-Anthropologie et Ethnobiologie, Equipe d'Anthropologie Evolutive, UMR 7206, Centre National de la Recherche Scientifique (CNRS) et Université Diderot Paris 7, Paris, France. [4]Institut de Recherche pour le Développement, UMR D216 MERIT (Mère et enfant face aux infections tropicales), Paris, France. [5]COMUE Sorbonne Paris Cité, Faculté de Pharmacie, Université Paris Descartes, Paris, France. [6]Centre d'Etude et de Recherche sur le Paludisme Associé à la Grossesse et l'Enfance (CERPAGE), Cotonou, Benin. [7]Laboratoire de Parasitologie, Faculté des Sciences de la Santé, Université d'Abomey-Calavi, Cotonou, Benin. [8]Institut National d'Etudes Démographiques (INED), Paris, France. [9]Instituto de Investigação e Inovação em Saúde (i3S), Universidade do Porto, Porto, Portugal. [10]Instituto de Patologia e Imunologia Molecular da Universidade do Porto (IPATIMUP), Porto, Portugal. [11]Faculdade de Medicina da Universidade do Porto, Porto, Portugal. [12]Faculty of Medicine, University of Iceland, Reykjavik, Iceland. *e-mail: kstefans@decode.is; agnar@decode.is

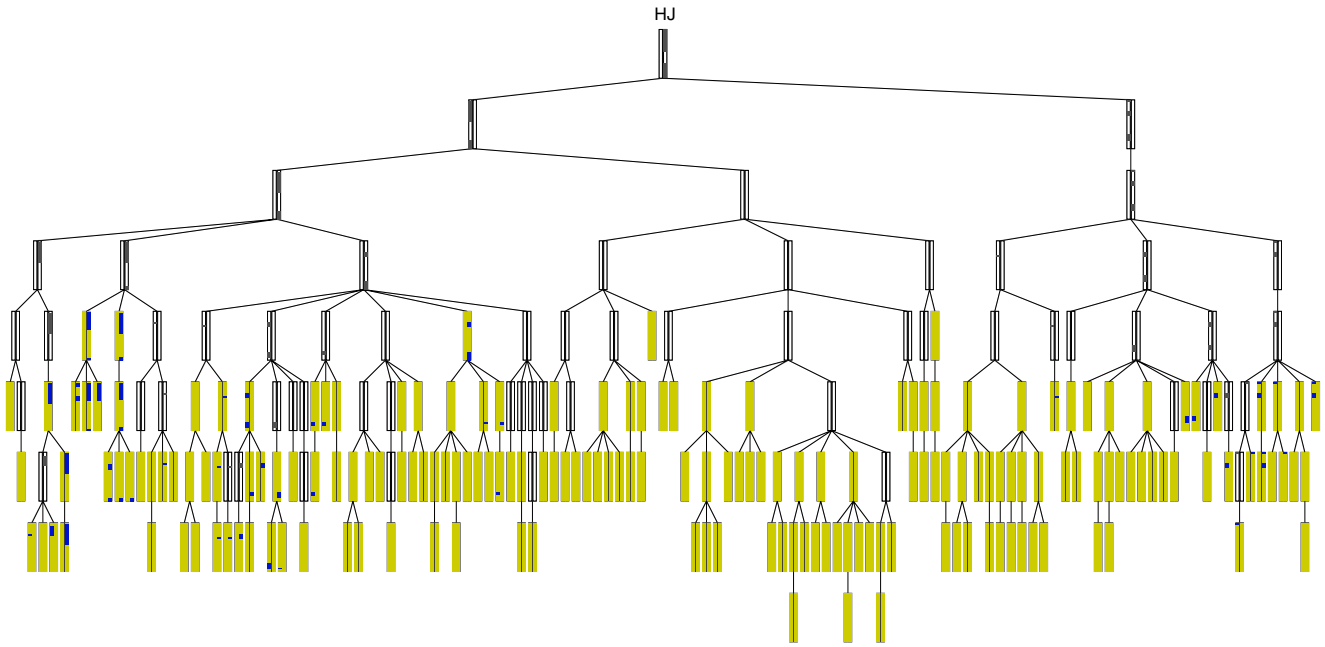

Fig. 1 | Genealogical reconstruction of HJ's maternal chromosome 3. The genealogy connecting HJ to his 182 genotyped descendants (colored chromosomes) via 61 ungenotyped descendants (uncolored chromosomes). Each individual is represented by a pair of chromosomes, with parent–offspring relationships represented by connecting lines. For the genotyped descendants, yellow- and blue-colored fragments indicate European and African ancestry, respectively. Inferred African fragments in ungenotyped ancestors are shown in dark gray. In total, 126.97 (64%) of 198.29 Mb of HJ's maternal chromosome 3 was reconstructed.

Table 1 | Filtering of African fragments identified in genotyped descendants of HJ

	Length of fragments that meet criteria (in Mb)							Length of fragments that do not meet criteria (in Mb)						
Filtering criteria for putative African fragments	***N***	**Mean**	**s.d.**	**Median**	**Range**	**Sum**	**Genome coverage**	***N***	**Mean**	**s.d.**	**Median**	**Range**	**Sum**	**Genome coverage**
None	674	13.7	15.3	6.7	0.8–83	9,231.8	1,165.6	NA	NA	NA	NA	NA	NA	NA
Fragments unique to HJ descendants	610	14.8	15.7	8.4	0.9–83	9,057.5	1,116.3	64[a]	2.7	1.8	2.2	0.8–8	174.3	80.3
Parent of origin consistent with descent from HJ	601	15	15.7	8.5	0.9–83	9,028.5	1,098.1	9[a]	3.2	3.5	2.1	1–12.5	29	27.4
Has genotyped ancestor (GA)	289	13.2	13.7	7.2	0.9–71.3	3,812.1	812.6	312[b]	16.7	17.2	8.9	0.9–83	5,216.5	1,091.9
(i) Fragment also found in GA	288	13.2	13.7	7.2	0.9–71.3	3,810.2	810.7	1[a]	1.9	1.9	1.9	1.9	1.9	1.9
(ii) Fragment African in GA	284	13.4	13.8	7.4	0.9–71.3	3,802.8	808.5	4[a]	1.8	1.6	1.1	0.9–4.2	7.4	7.4
(iii) Fragment has expected parent of origin in GA	282[b]	13.5	13.8	7.5	0.9–71.3	3,793.4	804.6	2[a]	4.7	1.1	4.7	3.9–5.5	9.4	9.4

[a]These fragments were removed from subsequent analyses. [b]These fragments were retained for subsequent analyses. NA, not applicable.

fragments to be incompatible with descent from HJ (see Table 1). This amounts to an error rate of 1.5%, which is reduced to 0.3% after accounting for the fact that false-positive fragments tend to be short (29 of 9,028.5 Mb). Second, in cases where the path lies through another genotyped descendant (parent, grandparent or great-grandparent), a fragment from HJ should fulfill the following

Table 2 | Distribution of African fragment lengths in HJ's descendants by generation

				Summary of fragment lengths (in cM)						Summary of fragment lengths (in Mb)					
Gen.	Total no. of desc.	No. of genotyped desc.	No. of African fragments identified	Mean	s.d.	Median	Range	Sum	Genome coverage	Mean	s.d.	Median	Range	Sum	Genome coverage
4	56	5	45	29.1	27.6	15.9	3.2–108.1	1,307.5	814	21.2	22.2	8.8	0.9–83	955.5	602.3
5	97	44	237	22	19.1	15.9	0.1–98.4	5,205	1393	16.9	16.7	9.5	0.6–71.3	3,995.5	1,044.8
6	244	82	218	17.1	15.4	12.8	0.1–98.4	3,734.2	1140	13.4	13.8	7.2	0.9–71.3	2,918.7	886.7
7	256	48	90	15.8	14.9	12.8	0.1–98.4	1,419.6	651	12.1	12.6	6.8	0.8–71.3	1,091.6	504.5
8	48	3	3	11	3.1	12.8	7.4–12.8	33	17	5	1.7	6	3–6	14.9	9
Total	701	182	593	19.7	18.3	14.1	0.1–108.1	11,699.3	1436	15.1	15.8	8.7	0.6–83	8,976.2	1,090.8

Abbreviations: desc., descendants; gen., generation.

three criteria in the genotyped ancestor: it should (i) be present, (ii) be identified as African and (iii) have a parental origin that is consistent with the genealogical path leading to HJ. Of the remaining 601 putative African fragments, 289 could be traced through a genotyped ancestor on the path to HJ. Of these, one fragment (1.9 Mb) was not present in the genotyped ancestor, four (7.4 Mb) were not identified as African and two (9.4 Mb) had a parental origin inconsistent with the genealogical path leading to HJ. Thus, 282 of 289 fragments passed all criteria in genotyped ancestors, whereas seven (2.42%) were discarded, amounting to an error rate of 0.4% when fragment length is taken into account (18.7 of 3,812 Mb). An assessment of the 12 excluded fragments longer than 4 Mb indicates that they are unlikely to be due to recent African gene flow into Iceland, but rather are misidentified as African by HAPMIX[8]. Overall, our analysis indicates that remaining erroneous fragments are likely to be few and short.

In total, we were left with 594 fragments for subsequent analyses, the 282 aforementioned fragments that passed genealogical filtering and 312 that did not have genotyped ancestors (see Table 1). These fragments were tested again for sharing with the 150,832 Icelanders not descended from HJ, this time with a minimum of 500 loci (corresponding to 2.7 cM), yielding no matches. The mean length of these 594 putative African fragments from HJ is 15.17 Mb, and their combined length of 9,009.9 Mb covers 38% (1,091.99 Mb) of the autosomal genome at least once.

To further evaluate the reliability of these results, we repeated the same procedure of African fragment identification and filtering using HAPMIX with source populations from the 1000 Genomes Project[10] and another local ancestry estimation method, LAMP-LD[11], with both sets of source populations (see Methods). After filtering and trimming, these three analyses yielded 93%, 97% and 96%, respectively, of the African genome recovered by our original run of HAPMIX with the HapMap source populations. Conversely, our original run returned 94%, 98% and 98%, respectively, of the African genome recovered by the additional local ancestry analyses. We therefore conclude that our results are reasonably robust with respect to choice of source populations and local ancestry estimation method.

The proportion of African ancestry in HJ. We next sought to verify the claim that HJ's maternal and paternal haploid genomes were African and European in origin, respectively. Conditioning on the genealogy that links the 182 genotyped descendants to HJ, we simulated the expected autosomal fraction covered by African fragments, assuming 25%, 50% and 100% African ancestry of HJ. The observed fraction of 38% was within the expected distribution for 50% African ancestry (mean = 0.42, 95% confidence interval (CI) = 0.32–0.53). In contrast, the 95% quantiles for both 25% African ancestry (mean = 0.21, 95% CI = 0.13–0.29) and 100% African ancestry (mean = 0.74, 95% CI = 0.65–0.83) did not include our empirical value. In addition, we used WGS data from a sixth-generation patrilineal descendant of HJ to determine that his Y chromosome belongs to haplogroup I2a2a3a2 (according to

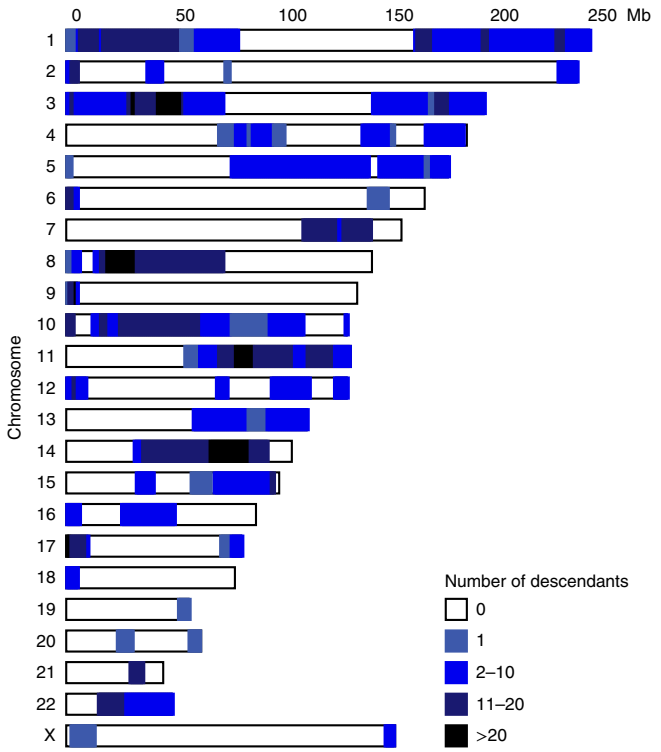

Fig. 2 | HJ's reconstructed genome. HJ's reconstructed maternal genome is made up of the African fragments identified in his 182 descendants that survived the multiple filtering criteria. After joining overlapping fragments from different descendants, we were left with 50 fragments, which amounted to 1,104 Mb of a haploid genome. These fragments are shown colored in shades of blue reflecting the number of descendants carrying each one.

the ISOGG 2013 classification system), which has a frequency of 0.15% in Iceland. Haplogroup I is found mostly in Europe and is essentially absent from African populations[12], thus supporting the story that HJ's father was European. Further support for the African ancestry of HJ's mother comes from the observation of two African fragments (lengths 10.21 and 3.86 Mb) found on the X chromosome of a female fifth-generation descendant linked to HJ through an alternating chain of daughter–son pairs that is consistent with X chromosome transmission from HJ's mother.

Allelic consistency of African fragments. Assuming that only one of each pair of autosomal chromosomes from HJ is African, it follows that overlapping African fragments identified in the 182 descendants should carry identical alleles (with the exception of rare subsequent mutation events and genotyping or phasing errors). Of the 594 African fragments, 7 (35.29 Mb) did not overlap with any other. The remaining 587 fragments overlapped with at least one other, forming 4,674 unique pairs and providing an intersection of 52,252.94 Mb. The mean mismatch per megabase across these pairs was 0.45. We excluded a 4.85-Mb fragment, carried by a seventh-generation descendant (who did not have a genotyped ancestor), that showed inconsistent alleles across its entire length, on comparing with two other overlapping fragments. The remaining 593 fragments had a combined length of 9,005.05, of which 33.58 Mb was trimmed from 163 fragments at their ends so as to exclude discrepant positions; the mean mismatch rate per megabase after trimming the ends of fragments was 0.05. After this final round of filtering, we were left with 593 putative African fragments from HJ, amounting to a total of 8,971.47 Mb, covering 1,090.77 of 2,824.18 Mb (38%) of the autosomal genome at least once (see Table 2). We were also able to retrieve 14.07 Mb (9%) of HJ's X chromosome. Figure 1 shows an example of the distribution of African fragments covering 126.97 of 198.29 Mb of chromosome 3 in the context of the genealogy. As expected, owing to the cumulative impact of recombination and Mendelian segregation, the African fragments become shorter with an increasing number of generations from HJ (Table 2). Figure 2 shows the locations of the reconstructed African fragments for all 23 chromosomes from HJ's maternal genome.

WGS data from HJ descendants. To more fully determine the allelic state of the African fragments derived from HJ, we examined WGS data from 20 of the 182 genotyped descendants. Typically, an individual will share most parts of each chromosome identical by descent (IBD), or very close to IBD, with several others from the same population due to recent common ancestry. The situation is somewhat different in the case of gene flow from a distantly related population, such as the introduction of chromosome fragments from HJ's African maternal genome to the Icelandic gene pool. Here, the degree of mutational divergence from other chromosomes in the receiving population will be much greater, reflecting more time back to the most recent common ancestor. Specifically, the African fragments in descendants of HJ are expected to carry a high density of alleles absent from other Icelanders. To evaluate this, we estimated the proportion of alleles per kilobase found in at least one HJ descendant with WGS data, but not in the 13,882 Icelanders with WGS data who are not descendants of HJ (the HJ singleton rate) (Fig. 3a). The mean HJ singleton rate for non-African fragments was 5.5×10^{-4} (s.d. $=6.2\times10^{-4}$), whereas the mean for African fragments was 7.2×10^{-2} (s.d. $=1.9\times10^{-2}$), a difference that is significant (t-test $P<2.2\times10^{-6}$) (Fig. 3d).

The origin of HJ's maternal genome. Between the seventeenth and nineteenth centuries, millions of Africans were forcibly transported mainly from the west coast of Africa to the Americas and the Caribbean[13]. From 1700 to 1790, the Danes transported ~45,000 Africans to St. Thomas, St. John and St. Croix (now known collectively as the US Virgin Islands)[13]. Either HJ's mother, Emilia Regina (born around 1760[3]), or her parents are likely to have been part of this group. With about 38% of HJ's autosomal maternal genome reconstructed, it is possible to recover otherwise lost information about their history and geographical origin. A multidimensional scaling (MDS) analysis plot, based on pairwise identity-by-state (IBS) values and a principal-component analysis (PCA)[14], places HJ's reconstructed maternal genome decisively within West Africa (see Fig. 4a), when compared to a range of populations from Africa, Europe and Asia (based on 90,078 overlapping loci from Illumina chips). Strong statistical support for this interpretation is provided by a direct assessment of a pairwise measure of genetic similarity (IBS weighted by allele frequency (wIBS); see Methods) between HJ's maternal genome and individuals from the different reference populations[15] (see Fig. 5).

We used the same approach to narrow down the origin of HJ's reconstructed genome within Africa, albeit constrained by the rather limited availability of reference genotype data[16–22]. Figure 4b shows an MDS plot of reference individuals from Eastern, Southern and Western Africa, in which HJ's reconstructed maternal genome clearly clusters with populations from Benin, Nigeria, Cameroon and Gabon

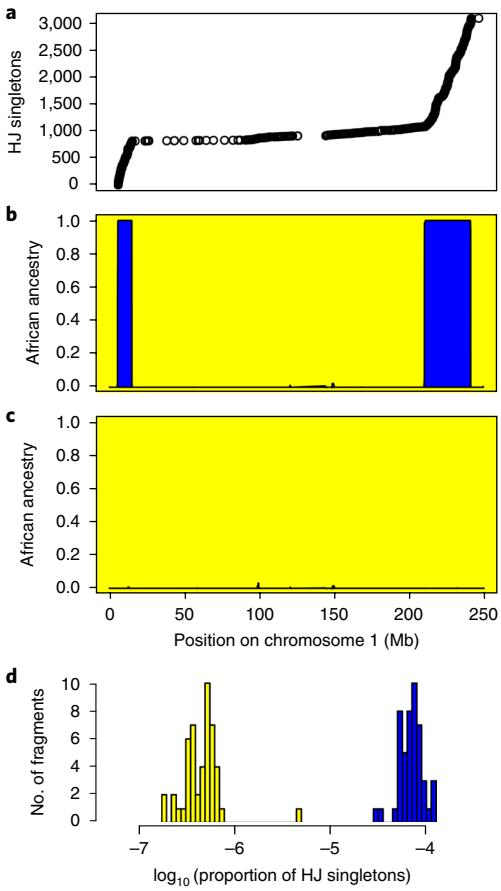

Fig. 3 | HJ singleton alleles in descendants with WGS data. a, Cumulative count of alleles from chromosome 1 that are not found in the 13,882 WGS Icelanders not descended from HJ for one descendant of HJ (HJ singleton alleles). **b**,**c**, Ancestry results, as inferred by HAPMIX, for the paternal and maternal chromosomes of the same HJ descendant, where blue represents African ancestry and yellow represents European ancestry. **d**, Distribution of the proportion of HJ singleton alleles for all filtered African fragments (blue) identified in the 20 descendants with WGS data against matched non-African fragments (yellow) selected from the same set of individuals (see Methods).

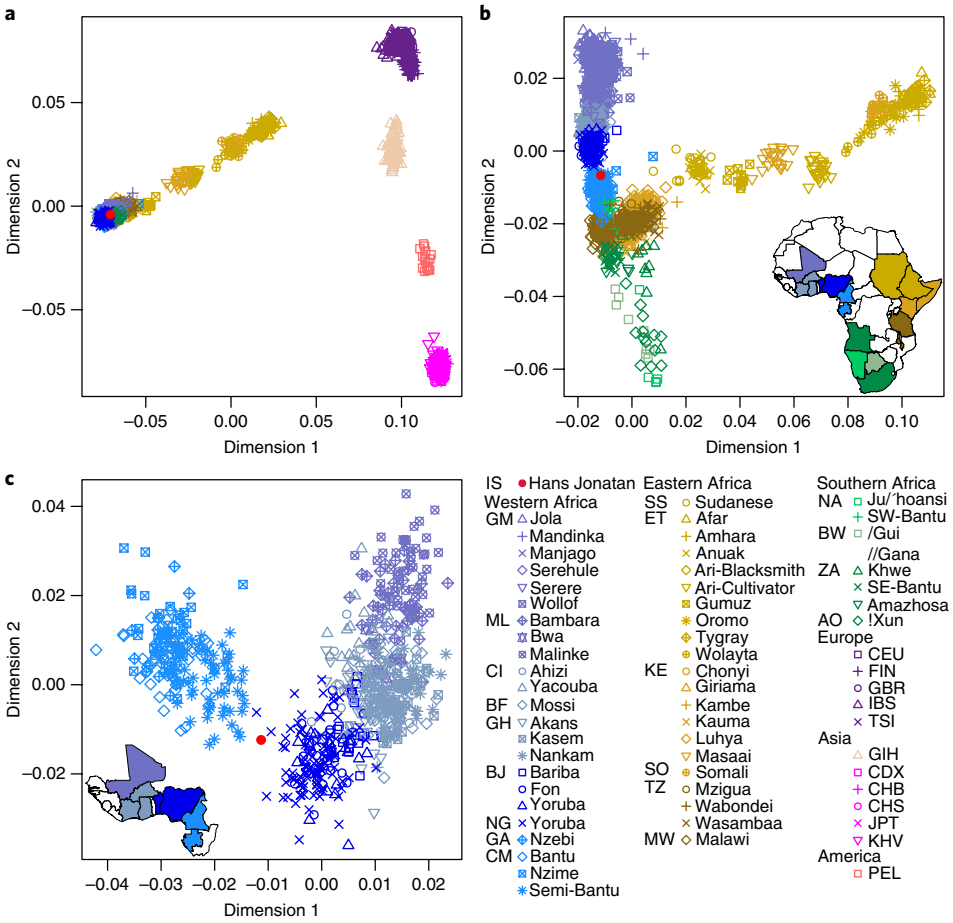

Fig. 4 | Multidimensional scaling (MDS) using worldwide reference populations. a–c, Plots showing the first two dimensions of an MDS analysis with HJ's reconstructed maternal genome along with samples from contemporary reference population from Africa, Europe, Asia and the Americas (stress value = 0.8; **a**), only African reference populations (stress value = 0.94; **b**) and only West African reference populations (stress value = 0.96; **c**).

(based on 90,078 overlapping loci from Illumina chips). Figure 4c shows an MDS plot of HJ's maternal genome with just West African populations (excluding the distantly related populations from Gambia), which indicates that the closest relationship is to populations from Nigeria and Cameroon. Analogous results were obtained for an independent set of reference populations typed on Affymetrix chips[23]. Furthermore, the position of HJ's maternal genome was the same when only the fragments identified by HAPMIX using both the HapMap and 1000 Genomes reference panels were used[24].

We used the pairwise wIBS measure to assess the relationship between HJ's maternal genome and the African reference populations (Fig. 6a). This indicated that HJ's mother was most related to populations from Benin, Cameroon and Nigeria, with the greatest value, of 0.186 (95% CI 0.1861–0.1866), observed for the Yoruba from Benin. Using a series of permutation tests of wIBS values between HJ's mother and individuals from reference population pairs, we were able to determine that she was significantly more related to the Yoruba from Benin than to 4 of the 16 other populations from West Africa (Nzime from Cameroon, Malinke from Mali, Bambara from Mali and Kasem from Ghana, $P < 0.05/16$; see also Fig. 6a). More resolution can be obtained through a pairwise measure of genetic similarity reflecting the proportion of a genome shared in IBD fragments longer than 0.5 cM. For this measure, the closest relationship was again observed with the Yoruba from Benin (0.0148, 95% CI 0.0142–0.0154); permutation tests revealed this value to be significantly greater than those for 10 of the 16 other populations from West Africa ($P < 0.05/16$; see Fig. 6b). The six

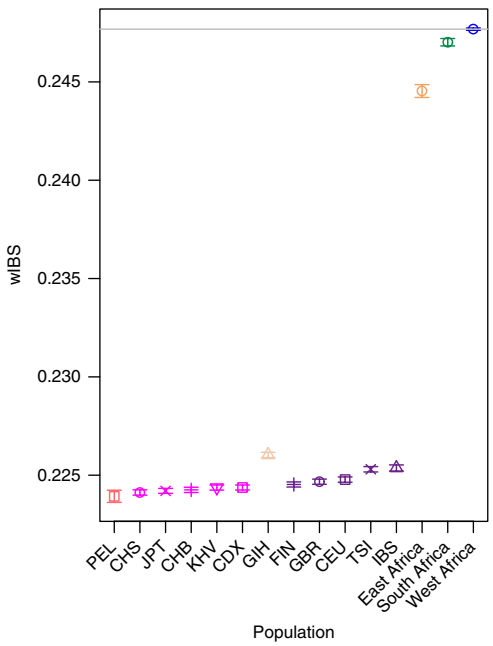

Fig. 5 | wIBS using worldwide reference populations. Plots showing the mean and 95% CIs (bars) for wIBS values between HJ and worldwide reference populations.

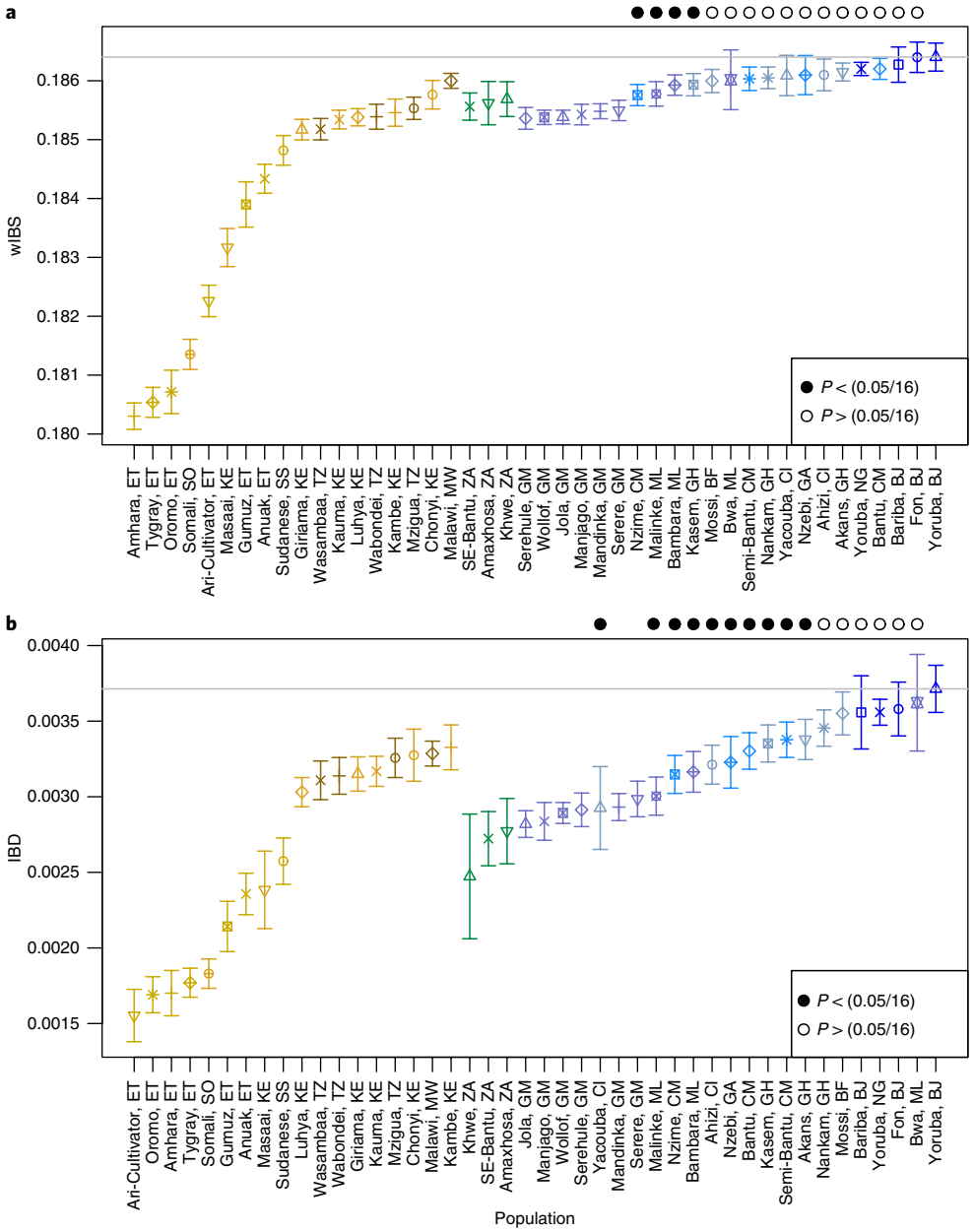

Fig. 6 | wIBS and IBD with African reference populations. a,**b**, Plots showing the mean and 95% CIs for wIBS (**a**) and IBD values (**b**) calculated between HJ's reconstructed maternal genome and African reference populations based on more than ten genotyped individuals.

populations that were not significantly less related to HJ's mother than the Yoruba from Benin were the Nankam from Ghana, Yoruba from Nigeria, Mossi from Burkina-Faso, Fon from Benin, Bariba from Benin and Bwa from Mali. Although these results, in conjunction with the MDS plot in Fig. 4, do not pinpoint the source population of HJ's mother within West Africa, they do narrow down the region to Benin, Nigeria or Cameroon.

Discussion

To our knowledge, this study demonstrates the first use of genotype data from contemporary individuals, along with information about their genealogical relationships, to reconstruct a sizeable portion of the genome from a single ancestor born more than 200 years ago. Ancestor genome reconstruction of this kind can be viewed as a virtual ancient DNA study, whereby genotype information is retrieved from a long-dead individual without the need for DNA samples from physical remains. On the basis of the reference data and 38% of HJ's reconstructed maternal genome available to us, his mother's origin was most likely in one or more groups from Benin, Nigeria or Cameroon. A more precise assignment of origin will require both larger sample sizes and genotype data from a wider range of populations from this part of Africa. We note that such reference data would be valuable for many other purposes.

The reconstruction of HJ's maternal genome was unusually tractable because chromosome fragments from recent African ancestors are very rare in the Icelandic gene pool. When a relatively diverged genome is introduced into an otherwise homogeneous gene pool, ancestry estimation techniques can be used to identify introgressed fragments[25]. However, in principle, IBD detection methods can be applied to reconstruct admixed as well as non-admixed genomes, given extensive genealogical and genotype data. In the latter case, the primary challenge is to assign fragments found in genotyped descendants to the correct ancestors.

An implicit form of ancestor reconstruction, albeit not based on IBD detection, is already used in genome-wide association studies, in which alleles are imputed from genotyped individuals to ungenotyped ancestors with phenotype information up to two generations back[9,26]. As family-based high-resolution genotype data accumulate for specific populations, we believe that more explicit and systematic IBD-based efforts to reconstruct genomes of ancestors will be feasible. One obvious application is to study individuals of historical interest, such as HJ. Another could be to extend medical genetic research in cases where valuable phenotype data is available for ancestors who can no longer be sampled for DNA and directly genotyped. Finally, we note that previous studies have used limited and implicit ancestor reconstruction to study recombination[27] and mutation and gene conversion[28]. A more systematic and direct application of ancestor genome reconstruction in populations with extensive genotype and genealogy data would provide even greater power to study these important evolutionary forces.

There are two key constraints on the generation depth of ancestors who can be subjected to genome reconstruction. The first is the practical limitation of the diminishing magnitude and accuracy of genealogical information. The second is that as we go further back in time, the probability that an ancestor has contributed any given chromosome fragment to his or her descendants rapidly approaches zero[29]. Thus, after only ten generations, the expected genetic contribution of an ancestor to a single descendant is only 3.48 cM ($2^{-10}R$, where $R = 35.67$ M)[30]. Given the sparsity of independent lines of descent from ancestors in typical genealogies, it follows that a relatively small fraction of an ancestor's genome is expected to survive in tenth-generation descendants.

Nonetheless, the example of HJ shows that with extensive genealogical records, genotype data and divergent ancestry, genome reconstruction of an ancestor who died almost 200 years ago is relatively straightforward. Furthermore, it represents an uplifting story of a resourceful refugee from the Danish transatlantic slave trade, who was able to prosper in a culturally homogeneous and insular community of early-nineteenth-century Icelanders.

Methods

Methods, including statements of data availability and any associated accession codes and references, are available at https://doi.org/10.1038/s41588-017-0031-6.

Received: 1 March 2017; Accepted: 18 December 2017;
Published online: 15 January 2018

References

1. Helgason, A. et al. mtDNa and the islands of the North Atlantic: estimating the proportions of Norse and Gaelic ancestry. *Am. J. Hum. Genet.* **68**, 723–737 (2001).
2. Jónsson, G. & Magnússon, M. S. *Hagskinna: Icelandic Historical Statistics* (Statistics Iceland, 1997).
3. Palsson, G. *The Man Who Stole Himself: The Slave Odyssey of Hans Jonathan* (University of Chicago Press, 2016).
4. Price, A. L. et al. Sensitive detection of chromosomal segments of distinct ancestry in admixed populations. *PLoS Genet.* **5**, e1000519 (2009).
5. International HapMap Consortium. The International HapMap Project. *Nature* **426**, 789–796 (2003).
6. Helgason, A., Nicholson, G., Stefánsson, K. & Donnelly, P. A reassessment of genetic diversity in Icelanders: strong evidence from multiple loci for relative homogeneity caused by genetic drift. *Ann. Hum. Genet.* **67**, 281–297 (2003).
7. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
8. Jagadeesan, A. et al. Linked open-access data supporting this article, including but not limited to the data items separately cited as other references https://doi.org/10.6084/m9.figshare.5640985 (2017).
9. Gudbjartsson, D. F. et al. Large-scale whole-genome sequencing of the Icelandic population. *Nat. Genet.* **47**, 435–444 (2015).
10. 1000 Genomes Project Consortium et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
11. Baran, Y. et al. Fast and accurate inference of local ancestry in Latino populations. *Bioinformatics* **28**, 1359–1367 (2012).
12. Karafet, T. M. et al. New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res.* **18**, 830–838 (2008).
13. Eltis, D. & Richardson, D. The Trans-Atlantic Slave Trade Database http://www.slavevoyages.org/ (2010).
14. Jagadeesan, A. et al. PCA plots with African reference populations https://doi.org/10.6084/m9.figshare.5640937 (2017).
15. Lawson, D. J. & Falush, D. Population identification using genetic data. *Annu. Rev. Genomics Hum. Genet.* **13**, 337–361 (2012).
16. Busby, G. B. et al. Admixture into and within sub-Saharan Africa. *eLife* **5**, e15266 (2016).
17. Bryc, K. et al. Genome-wide patterns of population structure and admixture in West Africans and African Americans. *Proc. Natl. Acad. Sci. USA* **107**, 786–791 (2010).
18. Gurdasani, D. et al. The African Genome Variation Project shapes medical genetics in Africa. *Nature* **517**, 327–332 (2015).
19. Pagani, L. et al. Ethiopian genetic diversity reveals linguistic stratification and complex influences on the Ethiopian gene pool. *Am. J. Hum. Genet.* **91**, 83–96 (2012).
20. Patin, E. et al. The impact of agricultural emergence on the genetic history of African rainforest hunter-gatherers and agriculturalists. *Nat. Commun.* **5**, 3163 (2014).
21. Schlebusch, C. M. et al. Genomic variation in seven Khoe-San groups reveals adaptation and complex African history. *Science* **338**, 374–379 (2012).
22. Tishkoff, S. A. et al. The genetic structure and history of Africans and African Americans. *Science* **324**, 1035–1044 (2009).
23. Jagadeesan, A. et al. MDS plots with African reference populations (Affymetrix) https://doi.org/10.6084/m9.figshare.5640934 (2017).
24. Jagadeesan, A. et al. MDS/PCA plots within West Africa https://doi.org/10.6084/m9.figshare.5640931 (2017)
25. Sankararaman, S. et al. The genomic landscape of Neanderthal ancestry in present-day humans. *Nature* **507**, 354–357 (2014).
26. Kong, A. et al. Detection of sharing by descent, long-range phasing and haplotype imputation. *Nat. Genet.* **40**, 1068–1075 (2008).
27. Hinch, A. G. et al. The landscape of recombination in African Americans. *Nature* **476**, 170–175 (2011).
28. Palamara, P. F. et al. Leveraging distant relatedness to quantify human mutation and gene-conversion rates. *Am. J. Hum. Genet.* **97**, 775–789 (2015).
29. Barton, N. H. & Etheridge, A. M. The relation between reproductive value and genetic contribution. *Genetics* **188**, 953–973 (2011).
30. Kong, A. et al. A high-resolution recombination map of the human genome. *Nat. Genet.* **31**, 241–247 (2002).

Acknowledgements

A.J., H.J. and C.F.-L. were funded by the EUROTAST Marie Curie Framework Programme 7 Initial Training Network 290344). We thank all the members of EUROTAST network for their helpful comments and suggestions. S.S.E. and V.B.G. received grants from The Research Fund of University of Iceland for doctoral studies. M.S.E. received a grant from the Icelandic Research Fund (163428-051).We thank E. Soumonni, a historian whose advice guided the recruitment of Beninese individuals, and J.-P. Chippaux (CERPAGE, Cotonou, Benin) for his help with local authorities.

Author contributions

A.J., K.S. and A.H. planned and directed the research. A.J. and A.H. analyzed the data, with A.K., E.D.G., S.S.E., V.B.G., E.L.T., M.S.E., H.J. and G.M. providing assistance with particular tasks. A.J., K.S. and A.H. wrote the paper. J.-M.D., C.F.-L., F.M.-N., A.M., G.B. and L.P. provided data for reference samples from African populations and helped with population structure analysis.

Competing interests

Some authors affiliated with deCODE Genetics are employed by the company, which is owned by Amgen, Inc.: A.J., E.D.G., S.S.E., V.B.G., E.L.T., H.J., G.M., A.K., K.S. and A.H.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41588-017-0031-6.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to K.S. or A.H.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Methods

Genotype data. The genotype data used in this study are from previously genotyped and long-range-phased samples at deCODE genetics, typed on different Illumina microarray single-nucleotide polymorphism (SNP) chips, typed for >300,000 markers[26]. At the time of analysis, 151,014 Icelandic samples had been genotyped on SNP chips and subjected to the long-range phasing algorithm[9]. These data were gathered as part of ongoing research studies at deCODE. Permission was obtained from the Data Protection Commission of Iceland and National Bioethics Committee of Iceland to obtain the encrypted identifier of HJ. The study was carried out only with encrypted keys for these samples, using the encrypted version of the genealogical database. Using the encrypted identifier of HJ and the genealogical database, we determined that 182 of the full set of 151,014 chip-typed Icelandic samples were recorded as his descendants. We will refer to these encrypted keys in the paper with a custom naming scheme, wherein each sample is referred to by its generation and a unique number, with both being separated by an underscore character. These samples are our primary data set on which we have carried most of our analysis. All samples were typed on two main chip families from Illumina, the Human Hap series (with a common set of 277,523 SNPs) and the Omni series (with a common set of 495,280 SNPs). A set of 142,080 SNPs were common on these two chip families. Using the information from long shared haplotypes within genotyped Icelanders, variants that have not been directly typed have been imputed to individuals with other markers typed and are phased[26]. The long-range-phased marker set consisted of 630,724 markers. The majority of the descendants (144 of 182), typed on the Omni family, have been imputed for a maximum of 143,132 markers, and the rest (38 of 182), typed on the Human Hap family, have been imputed for a maximum of 358,600 markers. To hinder identification of HJ descendants, we only analyzed genealogies of the 182 genotyped descendants, and we do not report this genealogy with the sex of individuals displayed. The primary data set spans five generations of HJ descendants, starting from his great-great grandchildren (fourth generation) and four subsequent generations. The data from the remainder of the 150,832 chip-typed Icelanders available at deCODE were used to evaluate haplotype sharing with the 182 descendants of HJ.

Relatedness among HJ's descendants. We used a version of the genealogical database which had been modified to reflect concordance between genotypes and genealogical relationships of chip-typed Icelanders. We verified whether the genotype data of HJ's descendants were consistent with the genealogical relationships as recorded in the database, by estimating the genome shared identical by descent (IBD) between each pair of parent–offspring, grandparent–grandchildren, great-grandparent–great-grandchildren, full siblings, half siblings, first cousins, half-first cousins, uncle/aunt–niece/nephew, half-uncle/aunt–niece/nephew. To calculate estimates of genome-wide IBD proportions, we initially used the following thresholds to identify fragments with consecutive identical alleles shared by pairs of HJ descendants. The minimum length of such fragments was required to be 5 cM. At least 1,000 loci with matching alleles were required, with no mismatches tolerated. We found no inconsistencies when comparing the proportion of the genome shared between pairs of HJ descendants in such IBD fragments, with the expectation based on the reported genealogy[31].

Ancestry analysis. HAPMIX was used to estimate local ancestry in the 151,014 chip-typed Icelandic samples (including those from the 182 descendants of HJ) in batches of 100 samples. Phased genotypes from European and African reference populations were obtained from HapMap Phase 3, release II[5]. They were from 69 individuals from the HapMap CEU sample and 63 individuals from the HapMap YRI sample. These samples had genotypes for 1,361,297 loci (converted to NCBI Build 38 coordinates) that had been phased using the IMPUTE algorithm[32], of which 553,333 loci overlapped with those available for the genotyped descendants of HJ. Using HAPMIX in haploid mode with default settings[4], we estimated probabilities of CEU ancestry at every locus in each phased haploid genome of the 151,014 genotyped Icelandic samples. This was made possible by long-range phasing of the Icelandic samples, which provides accurate phase of alleles across entire chromosomes and also provides information about the parental origin of alleles[9]. Any allele with a probability of CEU ancestry (anc_{EUR}) < 0.7 in the haploid mode was considered as potentially African in origin. Contiguous alleles that passed this threshold were combined into fragments for which we recorded the physical and cM positions of the first and last loci, to give us estimated breakpoints of African fragments. The cM positions were obtained from HapMap Phase 3, release II[5], recombination rates (converted to NCBI Build 38 coordinates). Fragments where none of the alleles had $anc_{EUR} = 0$ were subsequently removed.

For the 182 descendants we also ran HAPMIX and LAMP-LD with 1000 Genomes Phase 3 haplotypes. The European reference panel consisted of 503 samples from England and Scotland [GBR], Toscana from Italy [TSI], Spain [IBS], Finns [FIN], and CEPH (US Utah residents with ancestry from northern and western Europe) [CEU]. The African reference panel consisted of 504 samples from Esan in Nigeria [ESN], Gambian in Western Division–Mandinka [GWD], Luhya in Webuye, Kenya [LWK], Mende in Sierra Leone [MSL] and Yoruba in Ibadan, Nigeria [YRI]. We retrieved phased alleles for 1,343,587 markers for these reference samples to use with both HAPMIX and LAMP-LD.

Finally, we ran ADMIXTURE (unsupervised clustering into three ancestral populations, $K = 3$) for all 151,014 chip-typed Icelandic samples in batches of 100 samples, with the CEU (representing North European ancestry), YRI (representing West African ancestry) and CHB (representing East Asian ancestry) HapMap samples.

Haplotype-sharing analysis with genotyped Icelanders not descended from HJ. We carried out an IBD haplotype-sharing analysis to assess if any of the fragments identified as African in HJ's descendants could be found in Icelanders who are not recorded as descendants of HJ. Any putative African fragment identified in HJ's genotyped descendants that was fully shared with at least one of the 150,832 Icelanders who are not recorded as descendants of HJ in the genealogical database was assumed not to be derived from HJ and excluded. In this analysis, a fragment was considered to be shared when at least 1,000 consecutive SNPs (corresponding an average of 5.5 cM) were identical (with no constraint on missing genotypes). After applying genealogical filters, we repeated the haplotype-sharing analysis using a threshold of 500 SNPs, corresponding to an average of 2.7 cM.

Validation of African fragments using genealogical data. The genealogy allowed us to determine for each genotyped descendant which of his or her two chromosomes (paternal or maternal) is expected to carry African fragments, thus allowing us to identify African fragments likely to be derived from HJ (assuming correct assignment of parent of origin to phased alleles). Each of the fragments was annotated with a binary variable (Y/N) depending on whether its parent of origin was consistent with the recorded genealogical path linking the descendant to HJ. In cases where another genotyped descendant lay on such a path to HJ (i.e. a parent, grandparent or great-grandparent), we evaluated the following three criteria: (i) whether the African fragment was present in the genotyped ancestor, (ii) whether it was identified as African in the genotyped ancestor and (iii) whether the parental origin of the fragment in the genotyped ancestor was consistent with the genealogical path leading to HJ. We recorded this information in another binary variable (Y/N). We excluded fragments that did not meet our filtering criteria[33].

Chromosome drop simulation. Simulations were carried out to estimate how much of HJ's genome was expected to have been retained in the 182 genotyped descendants, conditional on the genealogy linking them. We simulated 22 of his autosomal chromosomes with sex-averaged cM lengths of 278, 263, 224, 213, 203, 193, 186, 170, 168, 179, 159, 172, 127, 117, 131, 134, 129, 119, 107, 108, 62 and 73[5]. Recombination along the chromosome was modeled as a Poisson process, with a mean number of recombination events given by the length of chromosomes in morgans. For each autosomal chromosome and each generation of HJ descendants, recombination and Mendelian sampling was performed to produce a chromosome to be transmitted from parent to offspring. We define a maternal and paternal founder chromosome, and keep track of their fate among the descendants. For the 50% African ancestry simulation, we only keep track of the 22 maternal founder chromosomes from HJ. For the 100% simulation, we keep track of both maternal and paternal founder chromosomes. For the 25% simulation, we keep track of either the maternal or paternal founder chromosomes from HJ's mother (thereby simulating one extra generation). These simulations were repeated 10,000 times, yielding null distributions for the expected proportion of genome transmitted to genotyped descendants if HJ was 25%, 50% or 100% African.

Allelic consistency between overlapping African fragments. Based on the expectation that HJ was half African, it follows that he would have transmitted only a single African allele to his descendants at any given position in the genome. Accordingly, barring errors in genotyping, phasing or ancestry estimation, putative African fragments from HJ in genotyped descendants should carry identical alleles when they span overlapping positions. To evaluate this, we examined all pairs of overlapping fragments. The intersecting region in each pair was assessed in windows of 30 SNPs, which were recorded as discrepant if one or more alleles differed. For each fragment, we counted the number of instances it was in an overlapping pair where discrepancies were observed for more than half of the windows. For each pair involved in at least one such major discrepancy, we removed the fragment involved in a greater number of major discrepancies (updating the counts for other fragments after removal).

After removing a single fragment responsible for major discrepancies, we then sought to trim 163 putative African fragments with minor discrepancies at the ends due to inaccurate breakpoint assignments. In what follows, the procedure for trimming the start of fragments is described (an analogous procedure was used to trim the stop positions of fragments). For overlapping fragment pairs with discrepancies in the first three windows of the intersecting region, the fragment with a start position closest to the start of the intersecting region was selected for trimming. A new start position for this fragment was defined as the position after the last discrepancy in the first three windows. When a fragment was involved with multiple end-discrepancy pairs, the new start position was defined as the position after the last discrepancy across all pairs. The end-trimming procedure described above was repeated until further end discrepancies were found.

We performed two stages of filtering, one based on genealogy and the other based on pairwise analysis. We counted the number of discrepant positions per chromosome after these two filtering stages.

Whole-genome sequence data. At the time of analysis, 13,902 Icelanders had WGS data (>20× autosomal sequence depth) at deCODE Genetics[9], aligned and mapped to NCBI Build 38 of the human genome. Of these, 20 were descendants of HJ. The average number of reads per autosomal position in these individuals ranged from 26.7 to 55.8, with a mean of 38.9. The proportion of autosomal positions covered by reads was ≥99.69% in all individuals. We used the WGS genotypes from these 20 descendants of HJ to assess the imputation error rate by comparing them with imputed SNP genotypes in the long range phasing set. Of the 9,064,004 imputed alleles (at 4,532,002 loci) across the 20 descendants, 1,089 were discordant, which yields an error rate of 0.01%.

For each of the 20 HJ descendants with WGS data, we recorded the position of each locus where he or she carried an allele that was not found in any of the other 13,882 Icelanders with WGS data. The spatial distribution of these HJ descendant singleton alleles was examined to validate putative African fragments that were identified using the microarray SNP genotype data.

WGS data from the 20 HJ descendants was further used to determine the allelic state of African fragments in comparative analyses aimed at identifying the geographical origin of HJ's mother. In this case, we made use of parent-of-origin information to select the correct alleles[9,34].

African and world reference population genotype data. We created two main reference data sets consisting of microarray SNP genotype data from 74 different African populations, with the aim of identifying the most likely source population for HJ's mother. For the first reference data set, we merged genotype data from several previous studies, including 73 individuals from Gabon and Cameroon, from Patin et al., typed on Omni1 SNP array[20], and 4,216 individuals from publicly released genotyped data sets from the Malaria Gen Network[16], of which 2,504 individuals were from previously published data sets[10,19,21,35]. To our combined data set from previous studies, we merged data sets collected by our collaborators. Six different populations from West Africa, i.e. from Benin, Ivory Coast and Mali, were typed on Omni5 SNP array[36]. We applied quality checks, removed 2 individuals due to missing genotypes and removed 300 related individuals. We ran ADMIXTURE[7] to filter out individuals with recent European or Asian ancestry. The reference Africa Illumina data set contained 1,877 samples with 90,048 markers, when merged with overlapping SNPs from HJ. We also added representative samples from European, Asian and American populations from the 1000 Genomes Project to our reference Africa Illumina data set to create a worldwide data set[37].

To create our second reference data set based on Affymetrix chips, we merged genotypes from four different studies: 30 individuals from 6 different African populations typed on 500 K from HGDP[38]; 203 individuals, primarily from West and Central Africa typed on 500 K, from Bryc et al.[17]; 50 individuals from Xing et al.[39] from West Africa typed on 6.0; and lastly 254 individuals from West African populations typed on 6.0 chips from the 1000 Genomes Project[10]. The reference data set was merged with overlapping SNPs from HJ with 689 samples with 50,551 markers. We randomly sampled one allele from reference African individuals to match HJ's haploid genome while carrying out multidimensional scaling analysis. Metric multidimensional scaling was performed with the R command cmdscale. For a pair of individuals i and j, if d_{ij} represents the actual genetic distance and $\hat{d}_{ij}$ and represents the predicted distance based on the MDS model, stress was calculated as:

$$\text{stress} = \sqrt{\frac{\sum (dij - \hat{dij})^2}{\sum dij^2}}$$

Calculation of pairwise wIBS and IBD values. We used diploid genotypes from reference African individuals when calculating wIBS and IBD proportions, from populations where sample sizes were greater than 10. In the weighted IBS (wIBS) method, alleles are weighted according to their frequencies. For example if *a0* and *a1* are two alleles at a locus, weight for *a0* can be defined as

$$Wt_{a0} = \frac{1 - \text{Frq}_{a0}^2}{\text{Frq}_{a0} * \text{Frq}_{a1}}$$

A weight for *a1* can also be similarly defined. The weighted IBS score between two individuals p and q, across N overlapping SNPs, ignoring loci where either p or q have a missing genotype, is calculated as

$$\text{wIBS}(p,q) = \sum_{i=0}^{N} \frac{\text{no . of shared copy}(a0) * Wt_{a0} + \text{no . of shared copy}(a1) * Wt_{a1}}{2 * Max(Wt_{a0}, Wt_{a1})}$$

This has an effect of up-weighting the IBS scores when rare alleles are shared. To estimate the proportion of the genome shared IBD between pairs of individuals, we used the following thresholds to identify IBD fragments shared between HJ and the different African populations: (i) the minimum length of IBD fragments was 0.5 cM, (ii) the minimum number of loci with matching alleles within the IBD fragment was 40, (iii) the maximum number of mismatches tolerated within an IBD fragment was 0 and (iv) the maximum number of loci with missing genotypes in the IBD fragment was 5.

Permutation tests for pairwise wIBS and IBD values. In order to test whether the population yielding the greatest mean wIBS or IBD with HJ's reconstructed African genome (population A) was significantly greater than that observed for other reference populations, we performed population pairwise permutation tests. In each instance of this test, where B is another reference population, the wIBS or IBD values between HJ and each member of reference population A and B were shuffled 100,000 times to obtain the null distribution for the difference in mean wIBS or IBD values with HJ (A-B), under the assumption of no difference between them. For a single test, HJ was deemed to be more closely related to population A than B, when 5% or fewer values from the null distribution were greater than the observed (non-permuted) difference between A and B. These P values were further Bonferroni adjusted for the number of tests performed.

Life Sciences Reporting Summary. A Life Sciences Reporting Summary is available.

Data availability. Linked open-access data supporting this article, including but not limited to the data items separately cited as references, are available in figshare at https://doi.org/10.6084/m9.figshare.c.3940516 Icelandic law and the regulations of the Icelandic Data Protection Authority prohibit the release of individual-level and personally identifying data (as described[9]). Access to genotype, WGS or genealogical data is only possible onsite at deCODE Genetics. The previously unpublished Illumina microarray genotype data from 118 individuals from the following populations have been submitted to the European Genome-phenome Archive (EGA) under accession EGAD00010001283 and can be requested from J.M.D. (jean-michel.dugoujon@univ-tlse3.fr) or C.F.L. (cesar@eurotast.eu): Yoruba, Fon and Bariba from Benin, Ahizi and Yacouba from Ivory Coast and Bwa from Mali.

References

31. Jagadeesan, A. et al. Relatedness among HJ's descendants https://doi.org/10.6084/m9.figshare.5640982 (2018).
32. Howie, B. N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* **5**, e1000529 (2009).
33. Jagadeesan, A. et al. Filtering HAPMIX inferred African fragments https://doi.org/10.6084/m9.figshare.5640868 (2018).
34. Jonsson, T. et al. A mutation in APP protects against Alzheimer's disease and age-related cognitive decline. *Nature* **488**, 96–99 (2012).
35. Petersen, D. C. et al. Complex patterns of genomic admixture within southern Africa. *PLoS Genet.* **9**, e1003309 (2013).
36. Fortes-Lima, C. et al. Genetic population study of Y-chromosome markers in Benin and Ivory Coast ethnic groups. *Forensic Sci. Int. Genet.* **19**, 232–237 (2015).
37. Jagadeesan, A. et al. Reference populations https://doi.org/10.6084/m9.figshare.5640796 (2018).
38. López Herráez, D. et al. Genetic variation and recent positive selection in worldwide human populations: evidence from nearly 1 million SNPs. *PLoS One* **4**, e7888 (2009).
39. Xing, J. et al. Toward a more uniform sampling of human genetic diversity: a survey of worldwide populations by high-density genotyping. *Genomics* **96**, 199–210 (2010).

nature research

Corresponding Author: Agnar Helgason

Date: Sep 1, 2017

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work we publish. This form is published with all life science papers and is intended to promote consistency and transparency in reporting. All life sciences submissions use this form; while some list items might not apply to an individual manuscript, all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

   Describe how sample size was determined.

   Sample sizes were determined by availability of genotyped individuals at DeCODE (Hans Jonatan desdendants and other Icelanders). For reference populations samples sizes were determined by publicly available data.

2. Data exclusions

   Describe any data exclusions.

   No data were excluded from the analysis

3. Replication

   Describe whether the experimental findings were reliably reproduced.

   Our core experiments were repeated with a different local ancestry estimation method and with a different set of reference populations.

4. Randomization

   Describe how samples/organisms/participants were allocated into experimental groups.

   NA

5. Blinding

   Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

   NA

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

   For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or the Methods section if additional space is needed).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒	A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly.
☐	☒	A statement indicating how many times each experiment was replicated
☐	☒	The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒	A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒	The test results (e.g. *p* values) given as exact values whenever possible and with confidence intervals noted
☐	☒	A summary of the descriptive statistics, including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒	Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.	We provide a description of all the softwares used for analyses in the methods section of the manuscript.

For all studies, we encourage code deposition in a community repository (e.g. GitHub). Authors must make computer code available to editors and reviewers upon request. The *Nature Methods* guidance for providing algorithms and software for publication may be useful for any submission.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.	No unique materials were used.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	No antibodies were used.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	No eukaryotic cell lines were used
b. Describe the method of cell line authentication used.	No eukaryotic cell lines were used
c. Report whether the cell lines were tested for mycoplasma contamination.	No eukaryotic cell lines were used
d. If any of the cell lines used in the paper are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	No commonly misidentified cell lines were used

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.	No animals were used.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.	No medically relevant phenotype data from the human research participants was analyzed in this study. Given the sensitive nature of the study involving a single individual and his descendants, we withheld information on sex, age, birth-year and death-year in order to prevent the identifiability of the research participants (explained in the manuscript). More detailed information about the data used for our study can be found in the Gudbjartsson et al. 2015 Nature Genetics paper. We refer to this paper in our manuscript.