

RESEARCH ARTICLE SUMMARY

CROP GENETICS

Historic manioc genomes illuminate maintenance of diversity under long-lived clonal cultivation

Logan Kistler*, Fabio de Oliveira Freitas*, Rafal M. Gutaker, S. Yoshi Maezumi, Jazmín Ramos-Madrigal, Marcelo F. Simon, J. Moises Mendoza F., Sergei V. Drovetski, Hope Loiselle, Eder Jorge de Oliveira, Eduardo Alano Vieira, Luiz Joaquim Castelo Branco Carvalho, Marina Ellis Perez, Audrey T. Lin, Hsiao-Lei Liu, Rachel Miller, Natalia A. S. Przelomska, Aakrosh Ratan, Nathan Wales, Kevin Wann, Shuya Zhang, Magdalena García, Daniela Valenzuela, Francisco Rothhammer, Calogero M. Santoro, Alejandra I. Domic, José M. Capriles, Robin G. Allaby*

INTRODUCTION: Manioc (*Manihot esculenta*), a root crop also called cassava and yuca, is one of the most important staple foods in the world, feeding around a billion people globally. Archaeobotanical and genetic evidence show that it originated in the southwestern Amazon region, and it was widespread throughout the American tropics at the time of European colonization. Manioc's wild progenitor is a short-lived outcrossing perennial species, but manioc is almost exclusively cultivated by vegetative propagation of stem cuttings. It thrives in the fire-shaped ecosystems of seasonally dry forests and edge habitats in South America, where Indigenous farmers first began clonally propagating desirable varieties.

RATIONALE: More than any other group of crop species, our understanding of plant domestication has been shaped by the archaeobotany and genetics of weedy annuals, such as maize and rice. However, many major economic crops are clonal species—e.g., potatoes, yams, sweet potatoes, sugarcane, and others—and we know much less about how clonal cultivation in hu-

man crop fields shapes domestication, population genomics, and genome evolution. In this work, we carry out a broad genomic survey of manioc and its wild relatives sampled from herbaria, living collections, on-farm cultivation, and archaeological sites. Using 573 new and previously published genomes, we aimed to understand how clonal reproduction and selection shape the genomic landscape of manioc. We also integrated crop diversity and interviews with traditional farmers in Brazil's Xingu region to better understand traditional strategies for managing and sustaining crop biodiversity.

RESULTS: We found that manioc sampled from around the Americas spanning >100 years carries almost no geographic population structure. That is, genetic makeup is a negligible predictor of geographic origins. This pattern starkly contrasts with that of its wild progenitor and defies basic expectations of biogeography previously demonstrated in many other crops. Most manioc varieties also have close kinship links with other varieties collected hundreds of kilome-

ters away. These patterns reflect the shift to clonal propagation coupled with human-driven strategies for rapidly spreading clonal lineages over great distances, so that local genetic differentiation is overprinted by cultivation and dispersal practices. Manioc is highly heterozygous, and we observe that offspring are more heterozygous than expected in light of their parents' genotypes. This pattern likely reflects selection for global heterozygosity that suppresses the effects of recessive deleterious mutations under clonality. We found that all pairs of manioc worldwide share substantial fractions of their genomes in large identical-by-descent chromosomal haploblocks, a finding that is normally taken as robust evidence for recent common ancestry. However, we use breeding records to show that even unrelated samples carry far more genomic blocks than can be explained by kinship. Demographic simulations show that the shift to clonality and selection for heterozygosity lead to this unexpected pattern.

CONCLUSION: Clonal crop species account for a major fraction of global food production and caloric intake. Our population genomic treatment of manioc—focusing on traditional varieties, wild relatives, and an exhaustive sampling of farming landscapes in the Americas—establishes some previously unrealized expectations for the interplay of agricultural pressures and genomic outcomes under clonal cultivation. ■

The list of author affiliations is available in the full article online.

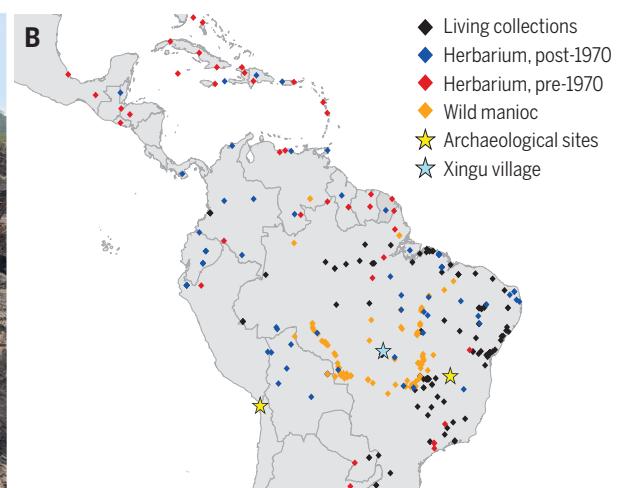
*Corresponding author. Email: kistlerl@si.edu (L.K.); fabio.freitas@embrapa.br (F.D.O.F.); r.g.allaby@warwick.ac.uk (R.G.A.)

Cite this article as L. Kistler et al., *Science* 387, eadq0018 (2025). DOI: 10.1126/science.adq0018

S READ THE FULL ARTICLE AT
<https://doi.org/10.1126/science.adq0018>



Young manioc growing in mounds from stem cuttings, in a newly cleared field that had previously been left to fallow in regrowing forest. (A) Upper Xingu region, Mato Grosso, Brazil. **(B)** New archaeological, herbarium, and modern-day samples sequenced for this study. These 282 samples combined with published data yielded 573 total genomes for analysis.



RESEARCH ARTICLE

CROP GENETICS

Historic manioc genomes illuminate maintenance of diversity under long-lived clonal cultivation

Logan Kistler^{1*}, Fabio de Oliveira Freitas^{2*}, Rafal M. Gutaker³, S. Yoshi Maezumi⁴, Jazmín Ramos-Madrigal⁵, Marcelo F. Simon², J. Moises Mendoza F.⁶, Sergei V. Drovetski⁷, Hope Loiselle^{1,8}, Eder Jorge de Oliveira⁹, Eduardo Alano Vieira¹⁰, Luiz Joaquim Castelo Branco Carvalho², Marina Ellis Perez^{1,11}, Audrey T. Lin¹², Hsiao-Lei Liu¹, Rachel Miller^{1,13}, Natalia A. S. Przelomska^{1,14}, Aakrosh Ratan¹⁵, Nathan Wales¹⁶, Kevin Wann^{1,17}, Shuya Zhang¹¹, Magdalena García^{18,19}, Daniela Valenzuela²⁰, Francisco Rothhammer²¹, Calogero M. Santoro^{19,21}, Alejandra I. Domic^{22,23}, José M. Capriles²³, Robin G. Allaby^{11,*}

Manioc—also called cassava and yuca—is among the world's most important crops, originating in South America in the early Holocene. Domestication for its starchy roots involved a near-total shift from sexual to clonal propagation, and almost all manioc worldwide is now grown from stem cuttings. In this work, we analyze 573 new and published genomes, focusing on traditional varieties from the Americas and wild relatives from herbaria, to reveal the effects of this shift to clonality. We observe kinship over large distances, maintenance of high genetic diversity, intergenerational heterozygosity enrichment, and genomic mosaics of identity-by-descent haploblocks that connect all manioc worldwide. Interviews with Indigenous traditional farmers in the Brazilian Cerrado illuminate how traditional management strategies for sustaining, diversifying, and sharing the gene pool have shaped manioc diversity.

The evolution of domesticated plants and the emergence of agriculture were among the most impactful processes in human history, which laid the groundwork for massive societal reorganization and population growth. Today, a small handful of staple crops is responsible for the bulk of calories consumed by humans worldwide. Among

these key crops is manioc (*Manihot esculenta* Crantz, also commonly called cassava and yuca), a root crop that feeds around a billion people throughout the global tropics (1).

Manioc is the world's seventh leading crop (2) but is primarily grown at small farm scales worldwide (3). All of its tissues, including the starchy roots, are acutely toxic with high levels of cyanogenic compounds (4). The roots and leaves can be lethal to humans if eaten untreated and are traditionally processed through leaching, boiling, or other detoxification strategies (5). Less-toxic “sweet” types are eaten untreated as a vegetable, but “bitter” manioc roots are typically processed into flour using either traditional or industrial methods (6). Despite its toxicity, high starch yield plus its tolerance for drought, pests, and poor soil have made it an important crop species in the tropics.

Previous genetic data point to domestication in southwestern Amazonia from wild *M. esculenta* ssp. *flabellifolia* (called *flabellifolia* hereafter) (7, 8). Native to edge habitats in the Amazon, such as the forest-savanna ecotone, *flabellifolia* thrives in ecological successional areas that experience frequent disturbances, such as fire (9). *Flabellifolia* typically lives up to 10 to 20 years, propagates by outcrossing, and exhibits extreme developmental plasticity (10). Unlike *flabellifolia*, manioc is cultivated almost exclusively by vegetative propagation through stem cuttings—a fundamental anthropogenic shift in the species' life history with the potential to substantially increase the generation times of individual genotypes. Seed-based propagation is difficult (11) but is used sparingly in some traditional set-

tings, especially by taking advantage of volunteer plants from a long-term soil seed bank (12).

Archaeological phytolith evidence for the genus *Manihot* first appears at 10,350 calibrated years before the present (cal yr B.P.) in the seasonally flooded savannas of Llanos de Moxos, Bolivia (13). Although this phytolith type is not diagnostic of crop manioc in the presence of wild *Manihot* (14), these findings may represent early manioc cultivation (13). Full farming societies, however, only appeared around 1500 to 1000 cal yr B.P. in this region (13). Microbotanical analyses from stone tools or stratigraphic sediments show that manioc became widespread throughout and beyond its wild range in the tropical Americas by the mid-Holocene (15, 16), including in northern Peru by ~8500 cal yr B.P. (17); Belize by ~8000 cal yr B.P. (18); Panama by ~7600 to 7400 cal yr B.P. (17, 19); Colombia by ~7000 cal yr B.P. (17, 20); and Tabasco, Mexico, by ~5800 cal yr B.P. (21).

Archaeological sediments in the southwestern Amazon show a clear increase in manioc usage around ~7000 cal yr B.P. associated with small-scale, permanent communities growing manioc on Amazonian anthroposols (22) coupled with the intensification of land use and fire activity (23). Several lines of evidence suggest the importance of manioc in later Holocene Amazonian economies, including the presence of *Manihot* phytoliths and pollen in Amazonian anthroposols and nearby lakes, the appearance of manioc processing tools during the late Holocene, and the hundreds of traditional varieties created by Amazonian farmers since their initial cultivation in the early Holocene (9). Microbotanical evidence also shows the widespread use of manioc across the Neotropics and Caribbean region in the middle and late Holocene—for instance, pollen grains recovered from Lake Caraña in the eastern Amazon ~2250 cal yr B.P. (24) and Sierra de Agua, Belize, ~1240 cal yr B.P. (25); phytoliths found at the Real Alto Site, Ecuador, ~4750 cal yr B.P. (14) and White Marl, Jamaica, ~870 cal yr B.P. (26); and root casts from Cerén, El Salvador, ~1230 cal yr B.P. (27, 28). Additionally, manioc starch grains are abundant in several archaeological sites in the coast of northern Chile ~2000 cal yr B.P. (20); La Corona, Guatemala, ~1350 cal yr B.P. (29); the Dominican Republic ~1150 to 450 cal yr B.P. (30); and the central Bahamas ~1250 to 850 cal yr B.P. (31).

In this work, we sequenced new herbarium ($n = 200$; 1890 to 2006) and modern [$n = 80$ from living collections (32)] genomes, capturing the range of wild and cultivated variation in the Americas, plus two archaeological genomes from Peruaçu, eastern Brazil (624 cal yr B.P.), and Arica, Chile (2034 cal yr B.P.) [Fig. 1A and data S1 (32)]. We combined these data with 291 previously sequenced genomes (33, 34) to examine the effects of human intervention on manioc and the resulting impacts on traditional agriculture, population genomics, and

¹Department of Anthropology, National Museum of Natural History, Smithsonian Institution, Washington, DC, USA.

²Embrapa Recursos Genéticos e Biotecnologia, Brasília, DF, Brazil. ³Royal Botanic Gardens, Kew, Richmond, UK.

⁴Department of Coevolution of Land Use and Urbanisation, Max Planck Institute of Geanthropology, Jena, Germany. ⁵Center for Evolutionary Hologenomics, Globe Institute, University of Copenhagen, Copenhagen, Denmark. ⁶Herbario del Oriente Boliviano (USZ), Museo de Historia Natural Noel Kempff Mercado/UAGRM, Santa Cruz, Bolivia. ⁷Laboratories for Analytical Biology, Smithsonian Institution, National Museum of Natural History, Washington, DC, USA. ⁸Department of Anthropology, University of Washington, Seattle, WA, USA.

⁹Embrapa Mandioca e Fruticultura, Nugene, Cruz das Almas, BA, Brazil. ¹⁰Embrapa Cerrados, Planaltina, DF, Brazil. ¹¹School of Life Sciences, University of Warwick, Coventry, UK. ¹²Richard Gilder Graduate School, American Museum of Natural History, New York, NY, USA. ¹³The McMaster Ancient DNA Center, Department of Biology, McMaster University, Hamilton, ON, Canada. ¹⁴School of the Environment and Life Sciences, University of Portsmouth, Portsmouth, UK. ¹⁵Center for Public Health Genomics, University of Virginia, Charlottesville, VA, USA. ¹⁶University of York, BioArCh, Environment Building, Wentworth Way, Heslington, York, UK. ¹⁷Department of Anthropology, Texas A&M University, College Station, TX, USA. ¹⁸Museo Chileno de Arte Precolombino, Santiago, Región Metropolitana, Chile. ¹⁹Millennium Nucleus of Applied Historical Ecology for Arid Forests (AFOREST), Santiago, Chile. ²⁰Departamento de Antropología, Universidad de Tarapacá, Arica, Chile. ²¹Instituto de Alta Investigación, Universidad de Tarapacá, Arica, Chile.

²²Department of Geosciences, The Pennsylvania State University, University Park, PA, USA. ²³Department of Anthropology, The Pennsylvania State University, University Park, PA, USA.

*Corresponding author. Email: kistlerl@si.edu (L.K.); fabio.freitas@embrapa.br (F.D.O.F.); r.g.allaby@warwick.ac.uk (R.G.A.).

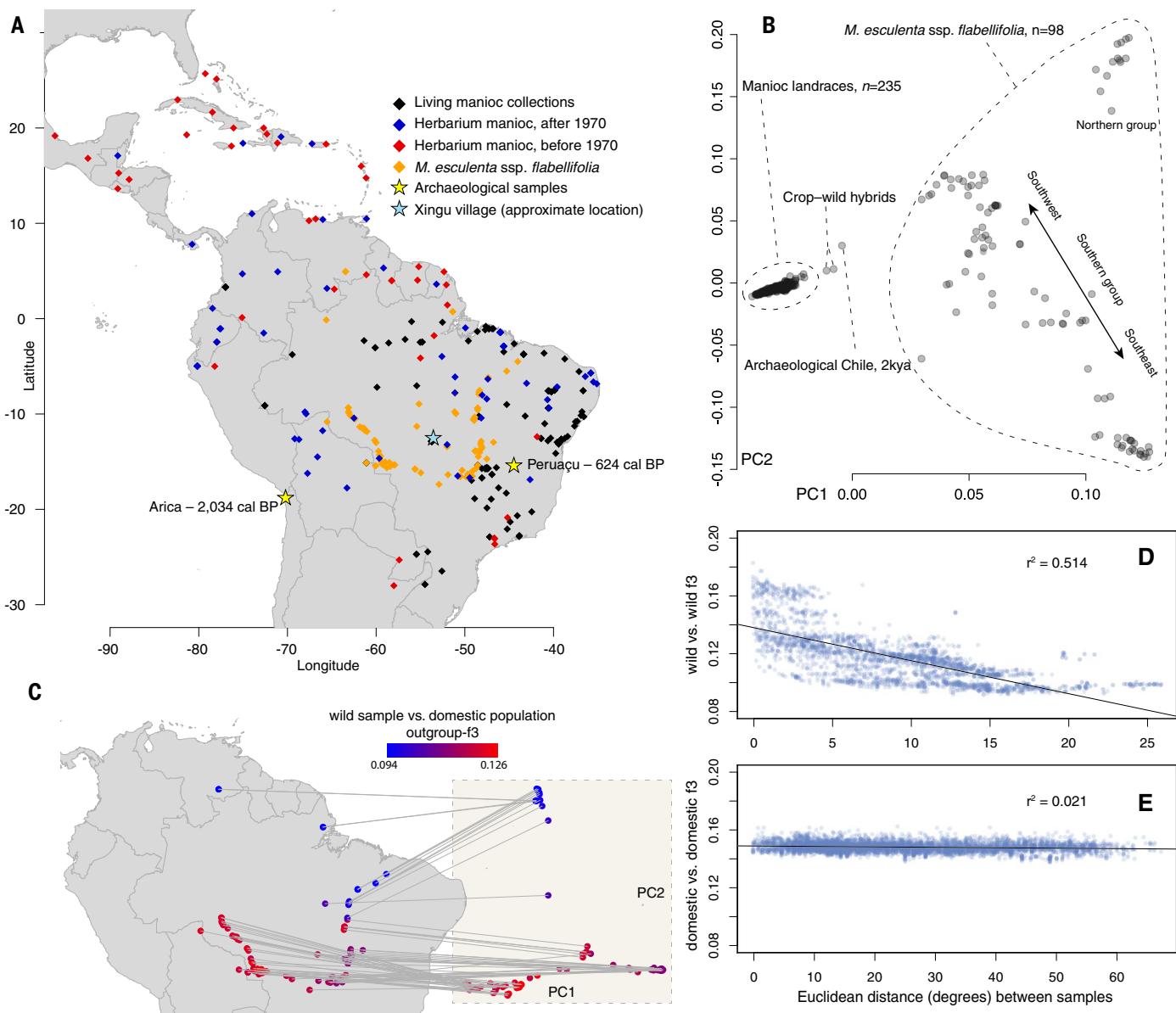


Fig. 1. Sample selection and population structure in manioc and wild relatives. (A) The 282 newly sequenced genomic samples reported in this study, including *flabellifolia*, herbarium specimens, and recent collections from living germplasm banks and active farms. Full sample data are given in data S1. Pre-1970 and post-1970 herbarium specimens are shown separately to indicate the geographic coverage among the older and younger samples; this date is not used as an analytical threshold or otherwise. (B) PCA of landraces, wild *flabellifolia*, and an archaeological sample from Arica, Chile (2034 cal yr B.P.), based on 1.2 million SNPs pruned for linkage. Proportion of variance explained by PC1 = 0.195, PC2 = 0.102. The archaeological sample from Peruacu, Brazil, was determined to be *Manihot glaziovii* and so was not included in the PCA. Samples collected as *M. esculenta* ssp. *peruviana* (data S1) are grouped with

flabellifolia here because previous genomic analysis has shown there to be no difference between these taxa (8). Additionally, some samples reported in data S1 failed the missingness filter for inclusion in the PCA (32) and so are not included here. kya, thousand years ago. (C) Locations of 79 well-provenanced wild samples, with a PCA illuminating their geographic structure and colored by their genetic similarity with manioc based on outgroup-f3 statistics. 423,013 pruned SNPs were used; proportion of variance explained by PC1 = 0.145, PC2 = 0.115. The divergent Atlantic forest clade of wild *Manihot brevibola* and *Manihot macrocarpa* was used as the outgroup population (8). Wild genomes with evidence of manioc admixture are excluded (32). (D and E) Comparison of pairwise geographic distance and shared drift (outgroup-f3) showing geographically structured *flabellifolia* (D) and unstructured manioc (E).

biodiversity. We aimed to interrogate biogeographic structure at continental scales in the wild and under clonality, to assess the persistence and patterning of kinship between clonal lineages through space and time, and in general to probe the distribution of genomic diversity

and structure in a clonal crop species under human influences. In total, we set out to understand how the specific genomic landscape of manioc interfaces with clonal reproduction and selection—different processes from canonical annual crop domestication pathways (35).

The new data presented in this work include 19 manioc lineages contributed by a Waurá Indigenous village, Ulupuwene, in the upper Xingu region of the Brazilian Cerrado. Ulupuwene consists of 10 extended households—around 150 people—and people there cultivate ~50 ha

of bitter manioc using traditional methods as a primary staple alongside tree fruits, river fish, and other wild and cultivated foods. The Waurá community are an Arawak-speaking Indigenous group that make up 1 of around 50 sovereign communities in the 2.6-million-ha Xingu region of Mato Grosso and Pará states. Between 2018 and 2023, we spoke with these traditional farmers about the loss and acquisition of manioc varieties over time, management of diversity, and decision-making about cultivation. The genetic snapshot of all manioc grown in one village can combine with farmer insights to illuminate the maintenance and circulation of crop genetic variation through traditional farming.

Crop-wild population structure, geographic patterning, and kinship among clones

Initial exploration of the genomic dataset using principal components analysis (PCA) revealed a geographically structured wild population (Fig. 1, B and C). Longitude and latitude predict the majority of the first [coefficient of determination ($R^2 = 0.670$, $P < 2 \times 10^{-16}$] and second ($R^2 = 0.781$, $P < 2 \times 10^{-16}$) genomic principal components (PCs) in wild accessions, and geographic distance strongly predicts shared genetic drift (outgroup- f_3 statistics, $R^2 = 0.514$, $P < 2 \times 10^{-16}$; Fig. 1D). Wild samples segregate into a divergent northern population in the northern Amazon and Atlantic forest regions and a southern population forming a genetic-geographic cline along the Amazon-Cerrado ecotone (Fig. 1, B and C). Model-based clustering using ADMIXTURE (36) further stratifies this southern cline into east and west population groups (32) (Fig. 1, B and C, and fig. S3). Manioc most resembles the southwestern wild population (Fig. 1C), consistent with the archaeological evidence for domestication and intensification described above.

In contrast to *flabellifolia*, all manioc landraces (regionally cultivated traditional varieties) are tightly clustered in the PCA with no evidence of population structure or geographic patterning (Fig. 1B). Geographic distance is correlated with genetic distance across the Americas ($P < 2 \times 10^{-16}$), but the relationship is extremely weak ($R^2 = 0.021$; Fig. 1E). Manioc also contains extensive gene flow from wild species other than *flabellifolia*, which highlights the importance of wild-to-crop admixture in traditional agriculture (8, 34). We observed non-*esculenta* ancestry in 59.1% of manioc genomes, originating from seven major *Manihot* clades (8) and covering 2.1% of the genome on average when present (data S1). Large chromosomal blocks of non-*esculenta* ancestry are frequently shared by multiple lineages, again with little evidence for spatial patterning (data S2). Among the 21 introgressed blocks shared by at least two landraces in the Americas, only four are spatially autocorrelated [Moran's I , significance threshold $P < 0.05$ (32)].

In total, manioc's wild progenitor is highly structured through isolation by distance, but by contrast, manioc is almost entirely without geographic structure. In other major food and fiber crops, such as rice (37), maize (38), flax (39), sunflower (40), and soy (41), a strong population structure generally reflects the biogeographic distribution and dispersal history. In many crops, a limited number of elite improved cultivars have come to agronomically dominate global production in these species under present-day industrialized agriculture. However, the underlying geographic structure among widely sampled varieties continues to reflect the longer-term processes playing out in crop fields. For example, maize landraces in the Americas are highly spatially correlated despite complex population structure and dispersal history and a massive modern-day production regime [outgroup- f_3 versus distance, $R^2 = 0.233$, $P < 2 \times 10^{-16}$ (32); fig. S1]. Manioc defies this basic expectation that geographic distribution underlies genetic structure.

The lack of geographic structure indicates that the pace of genetic drift, local adaptation, and crop breeding is insufficient to drive local differentiation under traditional cultivation. Instead, cultural mechanisms for rapid dispersal of clones seem to drive diffuse geographic-genetic relationships under long generation times. For example, according to Xingu farmers, marriage between members of different villages typically involves a wife relocating to a husband's household and bringing vegetative stock of her family's manioc varieties. Ad hoc exchange of manioc varieties between villages is also common, and these interactions can rapidly disperse high-quality manioc clones into widespread usage over large regions. Some Indigenous groups also reportedly plant gardens of root crops, including manioc, away from their homes to provide reliable sustenance during extended hunting trips. These gardens provide additional opportunities for the spread of clones and reinforce the long-term soil seed bank, which later contributes to originating new lineages.

The human-driven increase in generation times under clonality, combined with cultural mechanisms for rapid dispersal, created opportunities for close kinship spanning vast cultural and geographic areas under traditional agriculture. Using ngsRelate 2—a method based on allele frequencies that accommodates inbreeding (42)—we find that 82% of all landraces from the Americas have a second-degree or closer relative collected elsewhere (mean distance, 1470 km; SD, 1334 km; data S3 and fig. S8). This set includes 25 varieties with clones grown independently in another region (mean, 1550 km; SD, 1352 km; data S3 and fig. S8). In one case, a lineage collected in the northern Amazon in 1961 (NYBG_Pires_50815) is the parent to five other lineages collected across nearly a century from the Americas, including

a Cuban sample from 1904 (Fig. 2B). A living parent collected several decades after its offspring illustrates the extreme increase in clonal lineage longevity and the effectiveness of human-mediated dispersal.

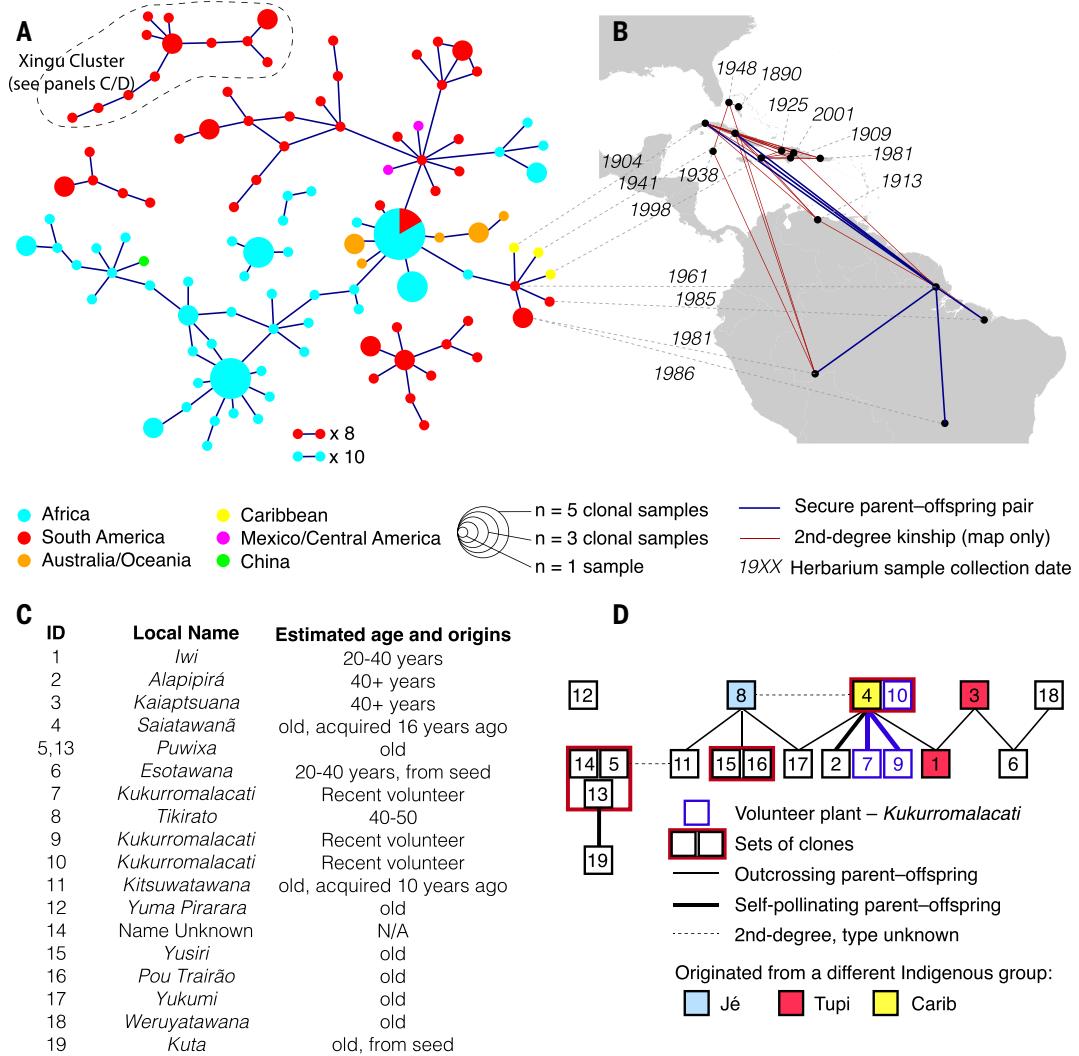
New lineages originate in key geographic and cultural regions

In South America, 21.3% of manioc landraces have no close relatives grown at a different site (data S3). However, among 17 landraces from the Caribbean, Central America, and Mexico, all share at least one relative grown elsewhere ($P = 0.011$ under the binomial distribution compared with South America). In other words, long-distance kinship is ubiquitous throughout the Americas but is more consistent outside the core region of South American cultivation.

This pattern suggests greater genetic turnover through seed-based propagation in the Amazon and Cerrado, with more consistent reliance on long-lived clones elsewhere. This pattern likely stems from differences in long-term ecological context and traditional farming practices, especially the degree to which new genetic crosses are recognized and used. In particular, traditional manioc agriculture in much of South America includes strategies for producing new crosses from soil seed banks in rotating forest-field systems. Seed-borne volunteer plants from these seed banks can be selected by farmers after clearing fields last occupied 15 to 20 years prior and also from within active crop fields (43, 44). In some areas, farmers group high-quality volunteer plants into the closest-matching existing landraces (45). Alternatively, the Waurá farmers that we interviewed carefully monitor the horticultural and food characteristics of volunteer plants over multiple growing seasons to decide whether to scale up cultivation as a new variety. One farmer at Ulupuwene cultivates a “*kukurro* house” (46)—a traditional method of encouraging volunteer crosses by coplanting several varieties separated by a forest border from the main cropping areas. *Kukurro* references a cosmological tradition linked with manioc cultivation, and *kukurromalacati* refers to seed-borne volunteer plants. Reproduction by seed is highly managed through these cultural practices, with seed-borne plants playing pivotal roles in traditional agrobiodiversity.

By contrast, seed-based propagation is essentially nonexistent in traditional farming outside of South America. The rotating forest-field agricultural strategy in the Amazon and Cerrado has continually built up a soil seed bank, creating a long-term repository of genetic diversity that can be integrated back into cultivation through new crosses, mediated by traditional strategies for curating volunteer plants, such as those described above. However, manioc agriculture in the Caribbean and Mesoamerica has not traditionally included a focus on volunteer crosses

Fig. 2. Widespread kinship in manioc genomes. (A) Kinship network of 137 secure parent-offspring pairs involving 148 individual manioc samples from worldwide germplasm excluding clonal replicates, including 58 from the Americas. (B) Collection sites and dates of a kinship network spanning >100 years and connecting the Amazon region with the Caribbean. Kinship levels up to conservative second-degree relatives are added here. (C and D) Descriptions (C) and detailed pedigree (D) inferred from 16 cultivated manioc varieties and 3 volunteer plants under observation collected from Ulupuwene.



grown from seed, which are less vigorous and productive early in their growth cycle compared with crops grown from stem cuttings.

As a result, introduced manioc varieties may be functionally “locked in” for long-term clonal propagation, without traditional farming practices to promote outcrossing and curate new genetic lineages and without the forest-field seed bank that recycles genetic diversity through crop fields. When a suite of varieties is introduced to a new region without a background gene pool and a history of cultivation, they may therefore tend to persist clonally for longer than in regions where outcrossing plays a role, thus sustaining long-lasting close kinship linkages to geographically distant relatives. Literature review revealed no evidence for post-colonial movements of manioc from South America into the Caribbean. It is plausible that close kinship ties between Amazonian and Caribbean herbarium specimens could stem from precolonial dispersals of manioc by Indigenous farmers followed by centuries of continuous clonal propagation.

Within Ulupuwene, we again observe abundant interrelatedness (Fig. 2, C and D; table S1; and data S3). All manioc in the village are toxic bitter varieties. They are sorted into high-yielding starchy types used for staple flatbreads and less starchy types used for a sweet drink after boiling in addition to some starch yield. The dominant lineage in the village—*saiatawaná*, or sample 4, a starchy type grown by all households—originated from a Carib-speaking village around 16 years ago through marriage. This variety is the parent of another outcrossed type (sample 1, *iwi*) obtained from a Tupi village an estimated 20 to 40 years ago and used by only two households. In addition, sample 4 is the sole parent of a self-pollinated variety obtained from another Arawak village that is estimated to be at least 40 years old (sample 2, *alapipirá*).

At the time of genetic sample collection, farmers also recognized three *kukurromalacati* (volunteer) plants under observation as similar to the dominant type. One of these (sample 10; Fig. 2, C and D) proved to be genetically identical to the dominant lineage—a not-yet-recognized

clone that likely sprouted from discarded stem trimmings at the time of planting. The other two volunteers were self-pollinated offspring of sample 4 (samples 7 and 9; Fig. 2, C and D), one of which was attractive to the farmers because it had characteristics of a previously lost variety.

In the 5 years since this genetic sampling, all three of these *kukurromalacati* volunteers were rejected for further cultivation. This snapshot of all genetic variation grown in one cultural context illustrates how long-term successful lineages are deliberately trialed and established and less suitable types are abandoned. As a management strategy, this process originates new genetic lineages at a rate of only a few per decade after stringent evaluation according to cultural preferences and horticultural characteristics. In this way, the gene pool is extremely deliberately managed, with explicit goals, such as starch yield and water stress tolerance, shaping decisions about whether to invest effort in new varieties. The clonal propagation of manioc confers a level of genetic control that is not

possible in seed-borne crops. Balancing out this process is the occasional loss or abandonment of other varieties, with farmers reporting 17 named lineages that they grew in living memory but are no longer used.

Global selection for heterozygosity maintains high genetic diversity

Despite manioc's lack of genetic structure and abundant interrelatedness, it retains high genetic diversity. At single-nucleotide polymorphism (SNP) sites masked to exclude non-*esculenta* introgression, manioc retains 95.8% average pairwise nucleotide diversity (32) compared with the entire wild gene pool and 97.9% compared with the southern population (Fig. 3A). Manioc is 8% more diverse than the southwestern population segment most closely tied to domestication, likely because of a history of deliberate maintenance of diverse varieties in the crop gene pool under traditional cultivation. Manioc is also highly heterozygous, and previous research has demonstrated that farmers' selection of robust volunteer plants effectively selected individuals with high genomic heterozygosity (45, 47). Under random mating, individual heterozygosity should approximate average pairwise nucleotide diversity in the population over a given set of genomic positions. In our dataset, however, we observe that genome-wide heterozygosity exceeds intersample diversity in 96.1% of manioc genomes ($n = 259$ genomes with at least 10x median coverage: heterozygosity mean = 0.204, SD = 0.035; mean pairwise diversity = 0.140, SD = 0.019; Fig. 3B and data S6). The remaining 10 low-heterozygosity genomes belong to highly inbred lineages—the threefold selfed reference genome (34), three selfed lineages from Ulupuwene (table S2), the backcrossed lineage BGM1130 (table S2), and five other previously published genomes (33). The average pairwise genetic distance between two chromosomes of the same plant is 46.7% greater than the difference between different individuals (Fig. 3B), which demonstrates an extreme enrichment of heterozygosity in manioc.

Furthermore, heterozygosity is not simply the result of accumulated somatic mutations in long-lived clones. In parent-offspring pairs at sites where a known parent is heterozygous, the heterozygous parent has an equal probability of contributing either allele to the offspring, which leads to a 50% chance of heterozygosity in the offspring even with no knowledge of the other parent. Thus, we expect 50% of the parent's heterozygous sites to remain heterozygous in the offspring under neutrality. However, in 67% of secure pairs where the parent and offspring can be established (data S3 and table S2), we observe >50% heterozygosity at these sites (mean, 52.0% heterozygosity; SD, 4.7%; data S7). Using central limit theorem to determine the probability distribution of mean population heterozygosity based on this sample

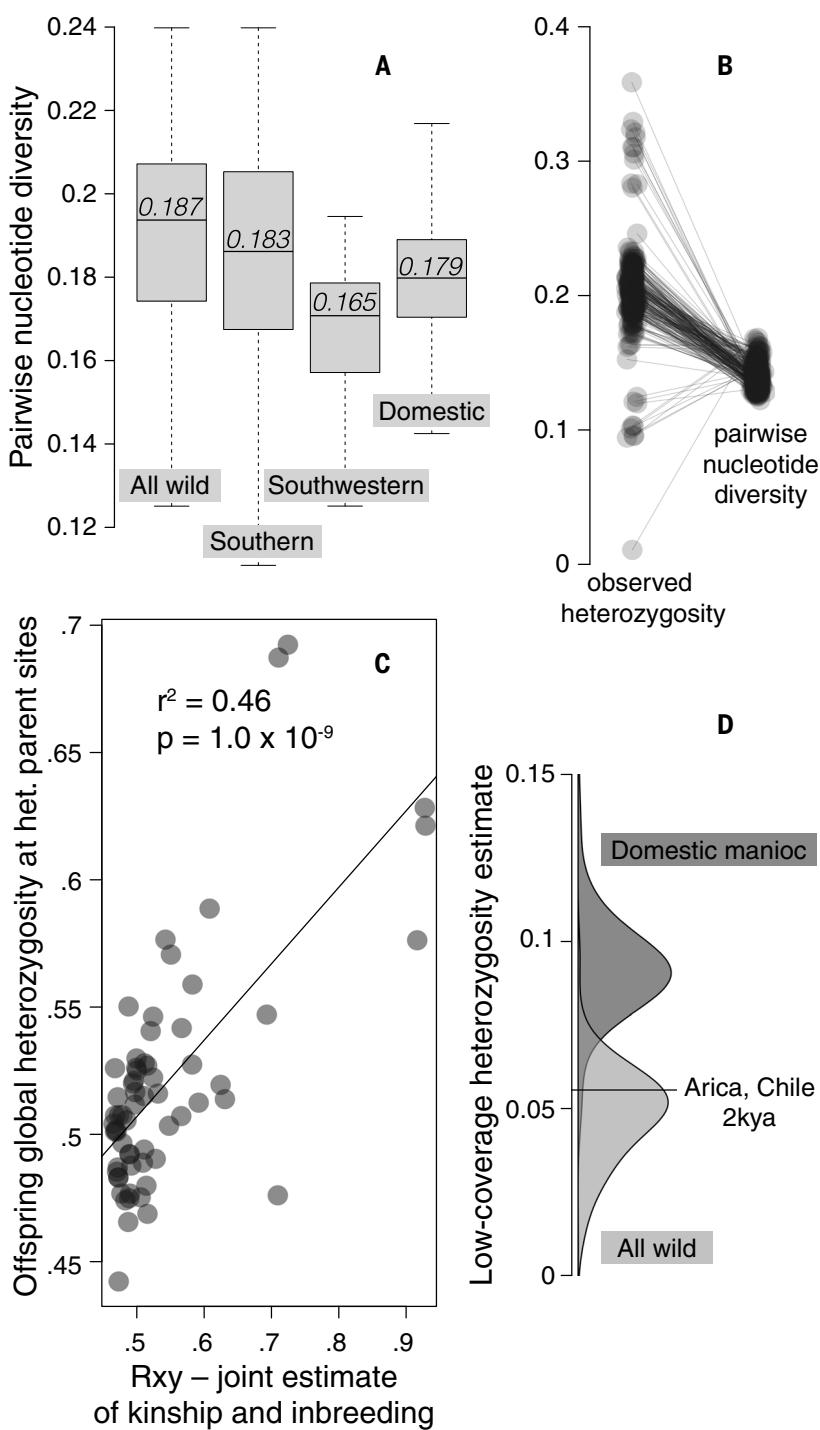


Fig. 3. High diversity and heterozygosity enrichment in manioc. (A) Pairwise nucleotide diversity at variant sites in wild population segments compared with manioc, which retains 95.7% of average nucleotide diversity compared with all *flabellifolia*. Outliers are not shown. (B) Comparison of heterozygosity and pairwise nucleotide diversity in 259 manioc genomes with sufficient coverage for diploid genotype calling. Pairs of points show a sample's global heterozygosity (left) and average pairwise nucleotide diversity compared with all other samples. This comparison reveals that the difference between two chromosomes in the same individual is 47% greater than the difference between two individuals, on average. The tight clustering in the right-hand side reinforces an unstructured population. (C) R_{xy} jointly estimates kinship and inbreeding coefficients of two genomes, with values of >0.5 indicating an inbreeding background in parent-offspring pairs (32). Here, inbreeding strongly predicts enrichment of heterozygosity in offspring at sites where parents are heterozygous. (D) Low-coverage estimates of global transversion heterozygosity in wild and domestic populations compared with an ~2000-year-old archaeological specimen from coastal Chile.

($n = 63$), we establish that the observed sample mean is significantly higher than the expected 50% ($P = 3.8 \times 10^{-4}$). We further examined the robustness of this result with a permutation test, again yielding a significant enrichment of >50% ($P = 4.2 \times 10^{-4}$; fig. S7) (32). In total, offspring are more heterozygous on average than expected by chance based on their parents' genotypes, which suggests greatest viability and long-term survival among the more heterozygous lineages.

This heterozygosity enrichment is greatest from self-pollinating single parents, where inbred offspring reached an extreme 69.2% heterozygosity at heterozygous parent sites (data S7). Inbreeding correlates strongly with heterozygosity enrichment across lineages (Pearson's $R^2 = 0.46$, $P = 1.0 \times 10^{-9}$; Fig. 3C). Homozygosity in inbred lineages appears to increase the importance of maintaining heterozygosity where possible. Previous work has shown that manioc's clonal propagation leads to a large number of deleterious recessive alleles that are not easily purged through recombination (33). By definition, inbreeding drives these numerous potentially deleterious alleles into runs of homozygosity, which leads to an increased importance of propping up heterozygosity throughout the genome. The outcome mimics nonspecific associative overdominance at a genome-wide scale, maintaining nucleotide diversity by systematically retaining high global heterozygosity to mask potentially deleterious variants.

Similar to other crops (48–50), manioc lineages have accumulated a substantial recessive mutation load—deleterious mutations confined to heterozygous genotypes and masking their harmful effects (33). Unlike in seed crops, the shift to clonality has meant that these variants are not easily purged through recombination and purifying selection (33). The many-fold increase in the somatic life span of manioc under human care has therefore allowed for the accumulation of fraught genomic landscapes. Although extant lineages currently remain viable, these deleterious variants threaten offspring viability through chromosome segregation into homozygous states. Additionally, self-pollination is prevented in the wild by male and female flowering asynchronously on the same plant. Growing fields of clones circumvents this strategy because normal variation in flowering time easily allows pollination to occur between genetically identical plants. Thus, the shift to clonality led to higher levels of inbreeding while simultaneously allowing the accumulation of mutations that most acutely affect inbred offspring. The resulting global enrichment of heterozygosity under traditional cultivation is likely a response to these outcomes, providing resilience by masking harmful mutations within heterozygous states where possible. Under this model, it is likely that other clonal crops

(potatoes, yams, sweet potatoes, grafted tree fruits, etc.) may show similar patterns of heterozygosity enrichment.

One archaeological sample from Arica in coastal Chile (20), directly dated to ~2 thousand years ago, falls between the landrace cluster and wild varieties in the PCA. This sample shows no sign of *flabellifolia* admixture compared with other manioc [F_4 Z-score = 0.71 (32)], and its shared drift with *flabellifolia* (outgroup- $\beta = 0.116$) is consistent with all other manioc (mean, 0.115; SD, 0.002). Therefore, it apparently derives from the same fundamental gene pool of domestication without additional gene flow from the wild. Using a conservative strategy optimized for low coverage (32), we estimate that the ancient sample's heterozygosity is in the range of that of *flabellifolia* but is more than three standard deviations below the mean of manioc (Fig. 3D). A sparse macrobotanical record for archaeological manioc unfortunately prevents a detailed time series genomic analysis. However, based on this sample, this finding suggests that the intensity of selection for heterozygote advantage may have increased in the past 2000 years, likely in response to the ongoing outcomes of the near-total shift to clonal cultivation.

We cannot entirely rule out the possibility that ancient manioc had much lower population diversity, which is reflected in the Arica sample's low heterozygosity. However, this would have required a severe bottleneck and rapid recovery, which seems inconsistent with high nucleotide diversity and heterozygosity in present-day manioc. The most parsimonious model is that nucleotide diversity in manioc has remained on par with that of its wild counterpart but that heterozygosity selection has intensified more recently.

Mosaics of identity-by-descent haploblocks connect all living manioc

Using only high-coverage genomes, including breeding lineages and varieties from outside the Americas, we scanned for shared haploblocks of identity-by-descent (IBD) to further investigate kinship (32). We observed an extreme pattern of IBD haploblock sharing: All pairs of present-day manioc share sufficient IBD blocks to suggest a median second-degree relationship, vastly greater than kinship estimates obtained using ngsRelate 2 (Fig. 4A and data S3). These shared IBD haploblocks are distributed throughout the entire genome, and a median 67 individual haploblocks of mean 3.2 mega-base pairs each are shared by any two genomes (fig. S9). In one extreme case, the published KBH-2006/18 breeding line developed recently in Kenya (34) shares sufficient IBD haploblocks with other varieties that it appears to be a first-degree relative—full sibling or parent-offspring—with an implausible 73.7% of all manioc sampled worldwide. Using known pedigrees from breeding records, we established

a nearest possible kinship level between 11,768 pairs of published genomes included in our dataset (32). Comparing these records with IBD results, we observed that these pairs of genomes share IBD far in excess of what could be conservatively attributed to kinship (Fig. 4B). Additionally, noninbred parent-offspring pairs are expected to have exactly one chromosome copy in common across the genome. However, validated parent-offspring pairs share a mean 14.2% of their genomes in IBD blocks across both chromosome copies (IBD2; SD, 6.3%), again showing IBD in excess of kinship.

IBD haploblock sharing is widely considered a robust standard of kinship inference (51). However, pairs of demonstrably unrelated manioc genomes from around the world share high levels of intact IBD blocks. To explain this pattern, we hypothesized that each genome could represent a mosaic of long-lived haplotypes in circulation. Pairs of genomes are therefore united by IBD on the basis of their randomly shared blocks from many ancestors as opposed to uniform IBD through a recent common ancestor. Heterozygosity is strongly associated with the degree of nonkinship IBD sharing (Fig. 4C), which suggests that this IBD mosaicism is reinforced by the traditional selection practices that increase heterozygosity. We hypothesize that the shift to clonal propagation in diverse lineages, plus selection for the heterozygote advantage, could create the observed patterns of nonkinship IBD mosaic sharing from multiple distant ancestors.

We simulated scenarios to test for the effects of clonality on nonkinship IBD accumulation (32). We found that modeling a simple shift from 100% sexual to 99% clonal propagation 5000 years in the past is sufficient to overestimate IBD-based kinship in the majority of individuals, so that the average unrelated pair shares IBD consistent with a second-degree relationship (half siblings, grandparent-grandchild, and avuncular relationships; Fig. 4D). Simulating selection on a suite of genome-wide variants—for example, domestication-linked alleles—does not drive further IBD. However, slight positive selection on genome-wide heterozygosity boosts IBD enrichment to the equivalent of first-degree kinship. This effect is true independent of demographic parameters, including population size and bottlenecks (fig. S10) (32). Although these simulations do not capture any of the cultural complexity around manioc dispersal and selection, these results suggest that the switch to clonality and culturally mediated heterozygosity selection are sufficient to explain the ubiquitous nonkinship IBD that unites all manioc.

Finally, to probe the antiquity of this pattern, we examined 3658 genome-wide short tandem repeats (STRs) within observed IBD regions uniting pairs of genomes (32). Because of their rapid mutation rate (52), STRs can offer greater

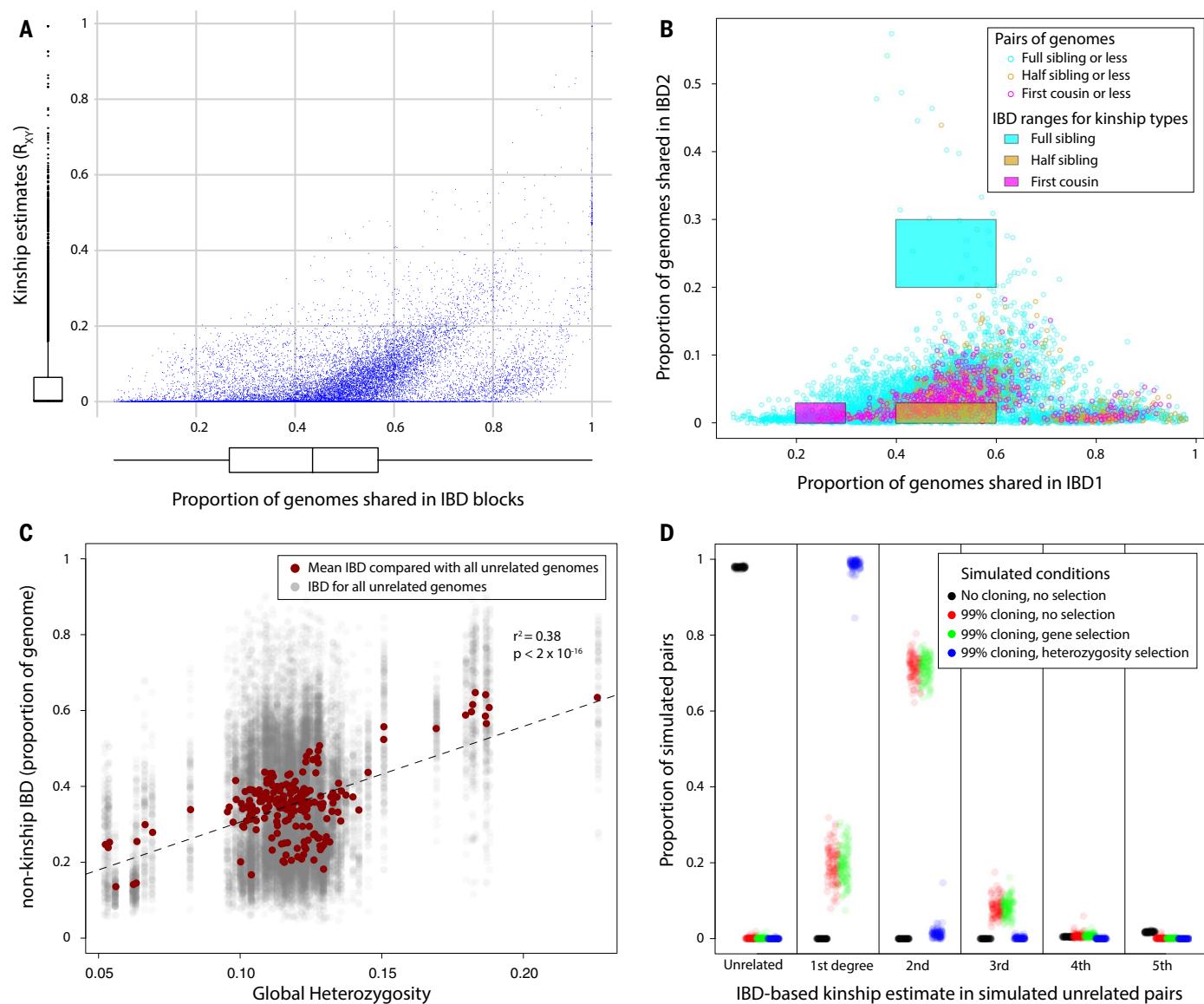


Fig. 4. All manioc lineages share large genomic tracts of IBD. (A) All-versus-all comparison of global manioc varieties with replicate clones removed, comparing the fraction of the genome shared in IBD blocks covering at least one chromosome, with kinship estimated using allele frequency. In first-degree relatives with no inbreeding, $R_{xy} = 0.5$. In samples analyzed here, median observed $R_{xy} = 0.003$, consistent with a low level of inbreeding and no recent kinship between most pairs of genomes. However, median observed IBD fraction = 0.44, consistent with a second-degree relative. The y-axis values are derived from ngsRelate kinship results, given in data S3, shown in fig. S8, and discussed above regarding global and regional kinship (Fig. 2). The x-axis values are from IBIS analysis, given in data S3. (B) Breeding records establish a conservative closest possible kinship level between pairs of published genomes with pedigree information (32).

IBD sharing exceeding what these kinship levels prescribe illustrates that the ubiquitous IBD observed cannot be explained by kinship. IBD ranges are shown with 20% above and below expectations, and an arbitrary 0 to 0.03 IBD2 range is added to half sibling and first cousin for visualization—no IBD2 is expected with these relationships. (C) The greater genomic heterozygosity that a sample carries, the more IBD blocks it shares with unrelated ($R_{xy} \leq 0.01$) individuals. Each sample's heterozygosity is shown with the mean of IBD shared with all other genomes (red) plus all individual comparisons (gray). (D) In 100 replicate simulated populations, a simple shift to cloning is sufficient to drive high levels of nonkinship IBD. For each of the four simulated conditions with a constant population size, the y axis shows the proportion of unrelated sampled pairs ($R_{xy} \leq 0.01$) assigned to kinship categories based on IBD blocks.

insights compared with SNP-based methods into the timing of events on Holocene time-scales, such as the shaping and distribution of crop genomes. We used the average square distance (ASD) between STR genotypes, which scales logarithmically with evolutionary time (53), and can therefore establish conservative lower boundaries on the average common ancestry of IBD blocks uniting two unrelated individuals. In 75 secure parent-offspring pairs (table S2), we observe a mean ASD of 6.6×10^{-4} in their shared blocks, with 37 of these pairs carrying no new STR mutations. In unrelated pairs sharing IBD mosaics covering 38.7 to 89.7% of their genomes, mean ASD in shared IBD blocks is 29-fold greater on average than

parent-offspring pairs (internal 95% range, 5.7- to 79.7-fold greater). Sufficient data do not exist for a robust calibration of a genomic STR-based molecular clock. However, breeding records and traditional farmer feedback reveal that parent-offspring pairs with a minimum common ancestry between 10 and 40 years ago (32) show very little new STR mutation.

This finding broadly suggests that shared IBD blocks have likely been circulating for millennia.

Globally, many crops are propagated vegetatively, and the associated tendency toward genomic deterioration is well known. The unexpected pattern of diversity in manioc highlights the important role of traditional management strategies that have shielded clonal populations from risks introduced through thousands of years of propagation.

Materials and methods summary

We collected whole-genome sequencing data from 282 manioc and wild relative samples, from herbarium specimens, archaeological tissues, and living collections (data S1). We collected an additional 291 previously sequenced genomic datasets, for a combined 573 manioc and wild relative genomes (data S1). Raw read data underwent quality and adapter trimming using AdapterRemoval 2 (54) and were mapped to the version 6 *Manihot esculenta* reference genome using the Burrows Wheeler Aligner (55). MapDamage 2 (56) was used to quantify ancient DNA-specific cytosine deamination in archaeological genomes, which was then mitigated through base quality rescaling. Following previous procedures for manioc, we removed reads that could also be mapped to a “bait” set of plastid and repeat sequences (33). We used VarScan (57) to permissively detect candidate variant positions independently in 729 genomes—the 571 nonarchaeological manioc and wild relative genomes analyzed in this work, plus 130 wild *Manihot* species previously reported (8) that we used only for introgression analysis. Among variant sites discovered in at least three genomes and falling between 0.1 and 0.9 coverage quantiles, we discovered 17,113,083 candidate SNP positions within the uniquely mappable fraction of the reference genome. Using these sites, we generated (i) pseudohaplotype base calls for introgression analysis, PCA, ADMIXTURE, f-statistics, and nucleotide diversity; (ii) diploid genome calls (>10x genomes only) for IBD detection, heterozygosity, and pairwise nucleotide diversity; and (iii) genotype likelihoods for kinship inference. We also used MISA (58) and HipSTR (59) to characterize genome-wide STR variants for genetic distance estimation within IBD blocks. We used ngsRelate2 (42) for kinship estimation in the presence of inbreeding; IBIS (60) for IBD block detection; ADMIXTURE (36) for model-based clustering; and PLINK 1.9 (61) for PCA, data filtration tasks (e.g., minor allele frequency and missingness filters), and computation of heterozygosity and nucleotide diversity. Previously published scripts were used for f-statistic computation (38, 62). Statistical tests reported throughout were done using R (63). Visualizations were carried out in R, with nonquantitative stylistic adjustments made using Adobe

Illustrator. We used SLiM 3 (64) to simulate demographic scenarios to test for the effects of a shift to clonality and various selection regimes on the accumulation of nonkinship IBD blocks. In Ulupuwene, we carried out informal interviews over multiple visits under approvals from local authorities, the Brazilian Fundação Nacional do Índio, and the Smithsonian’s Institutional Review Board. The full materials and methods are provided in the supplementary materials (32).

REFERENCES AND NOTES

- W. G. Adebayo, Cassava production in Africa: A panel analysis of the drivers and trends. *Helyon* **9**, e19939 (2023). doi: 10.1016/j.helyon.2023.e19939; pmid: 37809559
- Food and Agriculture Organization of the United Nations. FAOSTAT: Crops and livestock products (2023); <https://www.fao.org/faostat/en/#data/QCL/visualize>.
- R. Howeler, L. Utaladio, G. Thomas, *Save and Grow: Cassava: A Guide to Sustainable Production Intensification* (Food and Agriculture Organization of the United Nations, 2013).
- L. Rival, D. McKey, Domestication and diversity in manioc (*Manihot esculenta* Crantz ssp. *esculenta*, Euphorbiaceae). *Curr. Anthropol.* **49**, 1119–1128 (2008). doi: 10.1086/593119
- J. Mkumbira et al., Classification of cassava into ‘bitter’ and ‘cool’ in Malawi: From farmers’ perception to characterisation by molecular markers. *Euphytica* **132**, 7–22 (2003). doi: 10.1023/A:1024619327587
- M. M. Nye, The mis-measure of manioc (*Manihot esculenta*, Euphorbiaceae). *Econ. Bot.* **45**, 47–57 (1991). doi: 10.1007/BF02860049
- K. M. Olsen, B. A. Schaal, Evidence on the origin of cassava: Phylogeography of *Manihot esculenta*. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 5586–5591 (1999). doi: 10.1073/pnas.96.10.5586; pmid: 10318928
- M. F. Simon et al., Phylogenomic analysis points to a South American origin of *Manihot* and illuminates the primary gene pool of cassava. *New Phytol.* **233**, 534–545 (2022). doi: 10.1111/nph.17743; pmid: 34537964
- J. Iriarte et al., The origins of Amazonian landscapes: Plant cultivation, domestication and the spread of food production in tropical South America. *Quat. Sci. Rev.* **248**, 106582 (2020). doi: 10.1016/j.quascirev.2020.106582
- L. Ménard, D. McKey, G. S. Mühlen, B. Clair, N. P. Rowe, The evolutionary fate of phenotypic plasticity and functional traits under domestication in manioc: Changes in stem biomechanics and the appearance of stem brittleness. *PLOS ONE* **8**, e7427 (2013). doi: 10.1371/journal.pone.007427; pmid: 24023960
- R. H. Ellis, E. H. Roberts, Germination of stored cassava seed at constant and alternating temperatures. *Ann. Bot.* **44**, 677–684 (1979). doi: 10.1093/oxfordjournals.aob.a085782
- B. Pujol et al., Germination ecology of cassava (*Manihot esculenta* Crantz, Euphorbiaceae) in traditional agroecosystems: Seed and seedling biology of a vegetatively propagated domesticated plant. *Econ. Bot.* **56**, 366–379 (2002). doi: 10.1663/0013-0001(2002)056[0366:GEOME]2.0.CO;2
- U. Lombardo et al., Early Holocene crop cultivation and landscape modification in Amazonia. *Nature* **581**, 190–193 (2020). doi: 10.1038/s41586-020-2162-7; pmid: 32404996
- K. Chandler-Ezell, D. M. Pearsall, J. A. Zeidler, Root and tuber phytoliths and starch grains document manioc (*Manihot esculenta*), arrowroot (*Maranita arundinacea*), and llérén (*Calathea* sp.) at the Real Alto site, Ecuador. *Econ. Bot.* **60**, 103–120 (2006). doi: 10.1663/0013-0001(2006)60[103:RATPAS]2.0.CO;2
- A. Alves-Pereira et al., Selective signatures and high genome-wide diversity in traditional Brazilian manioc (*Manihot esculenta* Crantz) varieties. *Sci. Rep.* **12**, 1268 (2022). doi: 10.1038/s41598-022-05160-8; pmid: 35075210
- C. Isendahl, The domestication and early spread of manioc (*Manihot Esculenta* Crantz): A brief synthesis. *Lat. Am. Antiq.* **22**, 452–468 (2011). doi: 10.7183/1045-6635.22.4.452
- D. R. Piperno, The origins of plant cultivation and domestication in the New World Tropics: patterns, process, and new developments. *Curr. Anthropol.* **52**, S453–S470 (2011). doi: 10.1086/659998
- R. M. Rosengwig, D. M. Pearsall, M. A. Masson, B. J. Culleton, D. J. Kennett, Archaic period settlement and subsistence in the Maya lowlands: New starch grain and lithic data from Freshwater Creek, Belize. *J. Archaeol. Sci.* **41**, 308–321 (2014). doi: 10.1016/j.jas.2013.07.034
- R. Dickau, A. J. Ranere, R. G. Cooke, Starch grain evidence for the preceramic dispersals of maize and root crops into tropical dry and humid forests of Panama. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 3651–3656 (2007). doi: 10.1073/pnas.0611605104; pmid: 17360697
- J. M. Capriles et al., Pre-Columbian cultivation of vegetatively propagated and fruit tree tropical crops in the Atacama Desert. *Front. Ecol. Evol.* **10**, 993630 (2022). doi: 10.3389/fevo.2022.993630
- K. O. Pope et al., Origin and environmental setting of ancient agriculture in the lowlands of Mesoamerica. *Science* **292**, 1370–1373 (2001). doi: 10.1126/science.292.5520.1370; pmid: 11359011
- M. Arroyo-Kalin, Slash-burn-and-churn: Landscape history and crop cultivation in pre-Columbian Amazonia. *Quat. Int.* **249**, 4–18 (2012). doi: 10.1016/j.quaint.2011.08.004
- S. Mora Camargo, *Cultivars, Anthropic Soils, and Stability: A Preliminary Report of Archaeological Research in Aracarua, Colombian Amazonia*, no. 2, Latin American Archaeology Reports (Univ. Pittsburgh, 1991).
- S. Y. Maezumi et al., New insights from pre-Columbian land use and fire management in Amazonian dark earth forests. *Front. Ecol. Evol.* **6**, 111 (2018). doi: 10.3389/fevo.2018.00111
- N. Dunning, J. G. Jones, T. Beach, S. Luzzader-Beach, in *Heterarchy, Political Economy, and the Ancient Maya: The Three Rivers Region of the East-Central Yucatan Peninsula*, V. Scarborough, F. Valdez Jr., N. P. Dunning, Eds. (Univ. Arizona Press, 2003), pp. 14–24.
- S. Elliott et al., The legacy of 1300 years of land use in Jamaica. *J. Island Coast. Archaeol.* **19**, 312–343 (2022). doi: 10.1080/15564894.2022.2078448
- P. Sheets et al., Ancient Manioc Agriculture South of the Cerén Village, El Salvador. *Lat. Am. Antiq.* **23**, 259–281 (2012). doi: 10.7183/1045-6635.23.3.259
- V. Slotten, D. Lentz, P. Sheets, Landscape management and polyculture in the ancient gardens and fields at Joya de Cerén, El Salvador. *J. Anthropol. Archaeol.* **59**, 101191 (2020). doi: 10.1016/j.jaa.2020.101191
- C. Cagnato, J. M. Ponce, Ancient Maya manioc (*Manihot esculenta* Crantz) consumption: Starch grain evidence from late to terminal classic (8th–9th century CE) occupation at La Corona, northwestern Petén, Guatemala. *J. Archaeol. Sci. Rep.* **16**, 276–286 (2017). doi: 10.1016/j.jasrep.2017.09.035
- A. J. Ciocaflo, P. T. Sinelli, C. L. Hofman, Late Precolumbian Culinary Practices: Starch Analysis on Griddles from the Northern Caribbean. *J. Archaeol. Method Theory* **26**, 1632–1664 (2019). doi: 10.1007/s10816-019-09421-1
- M. J. Berman, D. M. Pearsall, Crop Dispersal and Lucayan Tool Use: Investigating the Creation of Transported Landscapes in the Central Bahamas through Starch Grain, Phytolith, Macrobotanical, and Artifact Studies. *J. Field Archaeol.* **45**, 355–371 (2020). doi: 10.1080/00934690.2020.1740958
- The full materials and methods are available in the supplementary materials online.
- P. Ramu et al., Cassava haplotype map highlights fixation of deleterious mutations during clonal propagation. *Nat. Genet.* **49**, 959–963 (2017). doi: 10.1038/ng.3845; pmid: 28416819
- J. V. Bredeson et al., Sequencing wild and cultivated cassava and related species reveals extensive interspecific hybridization and genetic diversity. *Nat. Biotechnol.* **34**, 562–570 (2016). doi: 10.1038/nbt.3535; pmid: 27088722
- D. Q. Fuller, T. Denham, in *The Convergent Evolution of Agriculture in Humans and Insects*, T. R. Schultz, R. Gawne, P. N. Peregrine, Eds., Vienna Series in Theoretical Biology (MIT Press, 2022), pp. 177–208.
- D. H. Alexander, J. Novembre, K. Lange, Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009). doi: 10.1101/gr.094052.109; pmid: 19648217
- R. M. Gutaker et al., Genomic history and ecology of the geographic spread of rice. *Nat. Plants* **6**, 492–502 (2020). doi: 10.1038/s41477-020-0659-6; pmid: 32415291
- L. Kistler et al., Multiproxy evidence highlights a complex evolutionary legacy of maize in South America. *Science* **362**, 1309–1313 (2018). doi: 10.1126/science.aav0207; pmid: 30545889
- R. M. Gutaker et al., Flax latitudinal adaptation at *LuTFL1* altered architecture and promoted fiber production. *Sci. Rep.* **9**, 976 (2019). doi: 10.1038/s41598-018-37086-5; pmid: 30700760
- J. R. Mandel, J. M. Dechaine, L. F. Marek, J. M. Burke, Genetic diversity and population structure in cultivated sunflower and a comparison to its wild progenitor, *Helianthus*

- annuus* L. *Theor. Appl. Genet.* **123**, 693–704 (2011). doi: [10.1007/s00122-011-1619-3](https://doi.org/10.1007/s00122-011-1619-3); pmid: [21638000](https://pubmed.ncbi.nlm.nih.gov/21638000/)
41. N. Bandillo *et al.*, A Population Structure and Genome-Wide Association Analysis on the USDA Soybean Germplasm Collection. *Plant Genome* **8**, plantgenome2015.04.0024 (2015). doi: [10.3835/plantgenome2015.04.0024](https://doi.org/10.3835/plantgenome2015.04.0024); pmid: [33228276](https://pubmed.ncbi.nlm.nih.gov/33228276/)
42. K. Hangjai, I. Moltke, P. A. Andersen, A. Manica, T. S. Korneliussen, Fast and accurate relatedness estimation from high-throughput sequencing data in the presence of inbreeding. *Gigascience* **8**, giz034 (2019). doi: [10.1093/gigascience/giz034](https://doi.org/10.1093/gigascience/giz034); pmid: [31042285](https://pubmed.ncbi.nlm.nih.gov/31042285/)
43. J. B. M. Sambatti, P. S. Martins, A. Ando, Folk taxonomy and evolutionary dynamics of cassava: A case study in Ubatuba, Brazil. *Econ. Bot.* **55**, 93–105 (2001). doi: [10.1007/BF02864549](https://doi.org/10.1007/BF02864549)
44. B. Pujol, F. Renoux, M. Elias, L. Rival, D. McKey, The unappreciated ecology of landrace populations: Conservation consequences of soil seed banks in Cassava. *Biol. Conserv.* **136**, 541–551 (2007). doi: [10.1016/j.biocon.2006.12.025](https://doi.org/10.1016/j.biocon.2006.12.025)
45. M. Elias, G. S. Mühlen, D. McKey, A. C. Roa, J. Tohme, Genetic diversity of traditional South American landraces of cassava (*Manihot esculenta* Crantz): An analysis using microsatellites. *Econ. Bot.* **58**, 242–256 (2004). doi: [10.1663/0013-0001\(2004\)058\[0242:GDOTSA\]2.0.CO;2](https://doi.org/10.1663/0013-0001(2004)058[0242:GDOTSA]2.0.CO;2)
46. F. d. O. Freitas, S. B. B. d. C. Zarur, Casa do Kukurro – Tradição cultural e sua influência na amplificação e manutenção da diversidade em uma roça. *Rev. Museu Arqueologia Etnologia* **17**, 381–389 (2007). doi: [10.11606/issn.2448-1750.revmae.2007.89806](https://doi.org/10.11606/issn.2448-1750.revmae.2007.89806)
47. B. Pujol, P. David, D. McKey, Microevolution in agricultural environments: How a traditional Amerindian farming practice favours heterozygosity in cassava (*Manihot esculenta* Crantz, Euphorbiaceae). *Ecol. Lett.* **8**, 138–147 (2005). doi: [10.1111/j.1461-0248.2004.00708.x](https://doi.org/10.1111/j.1461-0248.2004.00708.x)
48. S. Renaud, L. H. Rieseberg, The accumulation of deleterious mutations as a consequence of domestication and improvement in sunflowers and other composite crops. *Mol. Biol. Evol.* **32**, 2273–2283 (2015). doi: [10.1093/molbev/msv106](https://doi.org/10.1093/molbev/msv106); pmid: [25939650](https://pubmed.ncbi.nlm.nih.gov/25939650/)
49. Q. Liu, Y. Zhou, P. L. Morrell, B. S. Gaut, Deleterious variants in Asian rice and the potential cost of domestication. *Mol. Biol. Evol.* **34**, 908–924 (2017). doi: [10.1093/molbev/msw296](https://doi.org/10.1093/molbev/msw296); pmid: [28087781](https://pubmed.ncbi.nlm.nih.gov/28087781/)
50. L. Wang *et al.*, The interplay of demography and selection during maize domestication and expansion. *Genome Biol.* **18**, 215 (2017). doi: [10.1186/s13059-017-1346-4](https://doi.org/10.1186/s13059-017-1346-4); pmid: [29132403](https://pubmed.ncbi.nlm.nih.gov/29132403/)
51. M. D. Ramstetter *et al.*, Benchmarking relatedness inference methods with genome-wide data from thousands of relatives. *Genetics* **207**, 75–82 (2017). doi: [10.1534/genetics.117.1122](https://doi.org/10.1534/genetics.117.1122); pmid: [28739658](https://pubmed.ncbi.nlm.nih.gov/28739658/)
52. J. X. Sun, J. C. Mullikin, N. Patterson, D. E. Reich, Microsatellites are molecular clocks that support accurate inferences about history. *Mol. Biol. Evol.* **26**, 1017–1027 (2009). doi: [10.1093/molbev/msp025](https://doi.org/10.1093/molbev/msp025); pmid: [19221007](https://pubmed.ncbi.nlm.nih.gov/19221007/)
53. J. X. Sun *et al.*, A direct characterization of human mutation based on microsatellites. *Nat. Genet.* **44**, 1161–1165 (2012). doi: [10.1038/ng.2398](https://doi.org/10.1038/ng.2398); pmid: [22922873](https://pubmed.ncbi.nlm.nih.gov/22922873/)
54. M. Schubert, S. Lindgreen, L. Orlando, AdapterRemoval v2: Rapid adapter trimming, identification, and read merging. *BMC Res. Notes* **9**, 88 (2016). doi: [10.1186/s13104-016-1900-2](https://doi.org/10.1186/s13104-016-1900-2); pmid: [26868221](https://pubmed.ncbi.nlm.nih.gov/26868221/)
55. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009). doi: [10.1093/bioinformatics/btp324](https://doi.org/10.1093/bioinformatics/btp324); pmid: [19451168](https://pubmed.ncbi.nlm.nih.gov/19451168/)
56. H. Jónsson, A. Ginolhac, M. Schubert, P. L. F. Johnson, L. Orlando, mapDamage2.0: Fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**, 1682–1684 (2013). doi: [10.1093/bioinformatics/btt193](https://doi.org/10.1093/bioinformatics/btt193); pmid: [23613487](https://pubmed.ncbi.nlm.nih.gov/23613487/)
57. D. C. Koboldt *et al.*, VarScan: Variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics* **25**, 2283–2285 (2009). doi: [10.1093/bioinformatics/btp373](https://doi.org/10.1093/bioinformatics/btp373); pmid: [19542151](https://pubmed.ncbi.nlm.nih.gov/19542151/)
58. T. Thiel, W. Michalek, R. K. Varshney, A. Graner, Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor. Appl. Genet.* **106**, 411–422 (2003). doi: [10.1007/s00122-002-1031-0](https://doi.org/10.1007/s00122-002-1031-0); pmid: [12589540](https://pubmed.ncbi.nlm.nih.gov/12589540/)
59. T. Willems *et al.*, Genome-wide profiling of heritable and *de novo* STR variations. *Nat. Methods* **14**, 590–592 (2017). doi: [10.1038/nmeth.4267](https://doi.org/10.1038/nmeth.4267); pmid: [28436466](https://pubmed.ncbi.nlm.nih.gov/28436466/)
60. D. N. Seidman *et al.*, Rapid, Phase-free Detection of Long Identity-by-Descent Segments Enables Effective Relationship Classification. *Am. J. Hum. Genet.* **106**, 453–466 (2020). doi: [10.1016/j.ajhg.2020.02.012](https://doi.org/10.1016/j.ajhg.2020.02.012); pmid: [32197076](https://pubmed.ncbi.nlm.nih.gov/32197076/)
61. C. C. Chang *et al.*, Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* **4**, s13742-015-0047-8 (2015). doi: [10.1186/s13742-015-0047-8](https://doi.org/10.1186/s13742-015-0047-8); pmid: [25722852](https://pubmed.ncbi.nlm.nih.gov/25722852/)
62. L. Kistler *et al.*, Data from: Multi-proxy evidence highlights a complex evolutionary legacy of maize in South America, *Dryad* (2018). doi: [10.5061/dryad.70t85k2](https://doi.org/10.5061/dryad.70t85k2)
63. R Development Core Team, R: A language and environment for statistical computing (R Foundation for Statistical Computing, 2016). <http://www.R-project.org/>.
64. B. C. Haller, P. W. Messer, SLIM 3: Forward Genetic Simulations Beyond the Wright-Fisher Model. *Mol. Biol. Evol.* **36**, 632–637 (2019). doi: [10.1093/molbev/msy228](https://doi.org/10.1093/molbev/msy228); pmid: [30517680](https://pubmed.ncbi.nlm.nih.gov/30517680/)

ACKNOWLEDGMENTS

We are deeply grateful to the Waurá community in Xingu for welcoming us into their sovereign homeland and for sharing their knowledge and traditions. Computations performed for this paper were conducted on the Smithsonian Institution High Performance Cluster (<https://researchcomputing.si.edu/high-performance-computing-cluster>; doi:10.25572/SIHPC). Portions of the laboratory work were conducted in and with the support of the Laboratories for Analytical Biology (LAB) facilities of the Smithsonian's National Museum of Natural History. Authorization for sequencing Brazilian species was provided by SISGEN no. A725DDC. Many thanks to D. Piperno, J. Ross-Ibarra, and N. Mueller for helpful insights on the manuscript. **Funding:** Funding for the research was provided by a Smithsonian Institution grant to L.K. Fieldwork was funded by the National Geographic Society (NGS-61704R-19 to L.K., F.d.O.F., and R.G.A.). **Author contributions:** Study conceptualization and design: L.K., F.d.O.F., R.G.A.; Sample acquisition: L.K., F.d.O.F., M.F.S., J.M.M.F., E.J.d.O., L.J.C.B.C., E.A.V., M.G., D.V., F.R., C.M.S., J.M.C., R.G.A.; Genomic data collection: L.K., F.d.O.F., H.L., S.V.D., E.J.d.O., N.W., S.Z.; Genomic data analysis: L.K. with input from F.d.O.F., R.M.G., S.Y.M., J.R.-M., N.A.S.P., A.T.L., H.-L.L., A.R., and R.G.A.; Manioc lineage breeding record and literature surveys: L.K., R.M., K.W., M.E.P.; Demographic simulations: R.M.G., L.K.; Ethnographic interview: F.d.O.F., L.K., R.G.A.; Visualization: L.K.; Funding acquisition: L.K., F.d.O.F., R.G.A.; Writing – original draft: L.K. with contributions from F.d.O.F., S.Y.M., R.M.G., and R.G.A.; Writing – review & editing: All authors. **Competing interests:** All authors declare that there are no competing interests.

Data and materials availability: All raw genomic sequence data generated are freely available through National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) Bioproject no. PRJNA1167259, with sample accession numbers for all new and published data provided in data S1. Please contact the corresponding authors for any other postprocessed data formats.

License information: Copyright © 2025 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. <https://www.science.org/about/science-licenses-journal-article-reuse>

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adq0018

Materials and Methods

Figs. S1 to S10

Tables S1 and S2

References (65–100)

MDAR Reproducibility Checklist

Data S1 to S7

Submitted 24 April 2024; accepted 12 December 2024

10.1126/science.adq0018