nature ecology & evolution

ARTICLES

https://doi.org/10.1038/s41559-018-0697-x

The use and domestication of *Theobroma cacao* during the mid-Holocene in the upper Amazon

Sonia Zarrillo[1,2,16], Nilesh Gaikwad[3,4,16], Claire Lanaud[5,6,16], Terry Powis[7], Christopher Viot[5,6], Isabelle Lesur[8,9], Olivier Fouet[5,6], Xavier Argout[5,6], Erwan Guichoux[8], Franck Salin[8], Rey Loor Solorzano[10], Olivier Bouchez[11], Hélène Vignes[5,6], Patrick Severts[12], Julio Hurtado[13], Alexandra Yepez[13], Louis Grivetti[14], Michael Blake[2*] and Francisco Valdez[15]

Cacao (*Theobroma cacao* L.) is an important economic crop, yet studies of its domestication history and early uses are limited. Traditionally, cacao is thought to have been first domesticated in Mesoamerica. However, genomic research shows that *T. cacao*'s greatest diversity is in the upper Amazon region of northwest South America, pointing to this region as its centre of origin. Here, we report cacao use identified by three independent lines of archaeological evidence—cacao starch grains, absorbed theobromine residues and ancient DNA—dating from approximately 5,300 years ago recovered from the Santa Ana-La Florida (SALF) site in southeast Ecuador. To our knowledge, these findings constitute the earliest evidence of *T. cacao* use in the Americas and the first unequivocal archaeological example of its pre-Columbian use in South America. They also reveal the upper Amazon region as the oldest centre of cacao domestication yet identified.

Archaeological evidence for the pre-Columbian use of *Theobroma cacao* L. dates as early as approximately 3,900 calibrated years before present (cal. yr BP) in Central America and Mesoamerica and approximately 1,000 cal. yr BP in the southwest of the United States (Fig. 1)[1–12]. However, cacao's history of initial domestication, early uses and eventual spread throughout the Americas remains unclear. Due to the economic and ritual importance of cacao to the pre-Colombian cultures of Mesoamerica, it is thought that some cacao trees, corresponding to the Criollo variety, may have been introduced and transported through Central America and Mexico, accelerating the process of cacao domestication and initiating the first use of chocolate—either as a drink or a food[5]. This long-held idea that *T. cacao* was first domesticated in Mesoamerica, where it was isolated from its wild relatives, has been buoyed by archaeological, ethnohistoric and iconographic evidence[10,13]. However, recent genomic research on *T. cacao* shows that its greatest genetic diversity occurs in the humid forests of the upper Amazon tributaries region, suggesting that its earliest use and initial domestication originated there (Fig. 1)[14–17]. Here we report that *T. cacao* was used in Ecuador by 5,450–5,300 cal. yr BP, predating its earliest known use in Central America and Mexico by approximately 1.5 millennia. Three classes of archaeological evidence: *Theobroma* starch grains, absorbed organic residues, specifically theobromine and ancient DNA (aDNA) of *T. cacao*, as well as of *Theobroma* spp. or *Herrania* spp., recovered from the surfaces of ceramic and stone artefacts, have been identified in ceremonial, mortuary and residential contexts at the archaeological site of SALF[18] in southeast Ecuador. Our findings represent the earliest known use of *T. cacao* in the Americas, the first directly dated archaeological example of its use in South America and support genomic studies indicating this region as a centre of domestication of *T. cacao* from Ecuador[16].

There are currently 22 species of the genus *Theobroma* and 17 species of its wild relative genus *Herrania*, both members of the Malvaceae family[19]. Most of these are native to the upper Amazon tributaries region where *T. cacao* displays its greatest genetic diversity (Fig. 1)[14–17]. These botanical findings, combined with the observed occurrence of only two cultivated *Theobroma* species north of Costa Rica—*T. cacao* L. and *T. bicolor* Humb. & Bonpl.—suggest that *Theobroma* use originated in South America and the domesticated tree later spread or was traded into Mesoamerica[14,16,19]. Ethnographic and ethnohistoric sources document dozens of uses for both *Theobroma* and *Herrania* in South America, including use of the seeds as medicine and food, the pulp eaten fresh, as a juice or as a fermented alcoholic beverage and the bark and leaves for medicinal extractions and infusions[20]. Importantly, sources also indicate that *T. cacao* was in cultivation on the Pacific Coast of Ecuador before the arrival of the Spanish[13,21]. Despite these accounts, and that ancient ceramic vessels from Ecuador and the north coast of Peru include iconographic representations of cacao pods, no unequivocal direct archaeological evidence for the pre-Columbian use of cacao has hitherto been reported for South America[21].

To test the hypothesis that *Theobroma* may have been used anciently in the core region of the plant's wild range and greatest genetic diversity, we looked for evidence at SALF, the earliest known archaeological site belonging to the Mayo-Chinchipe Culture (Fig. 1

[1]Department of Anthropology and Archaeology, University of Calgary, Calgary, Alberta, Canada. [2]Department of Anthropology, University of British Columbia, Vancouver, British Columbia, Canada. [3]Department of Nutrition and Department of Environmental Toxicology, West Coast Metabolomics Center, University of California, Davis, CA, USA. [4]Gaikwad Steroidomics Laboratory, Davis, CA, USA. [5]CIRAD, UMR AGAP, Montpellier, France. [6]AGAP, University Montpellier, CIRAD, INRA, Montpellier SupAgro, Montpellier, France. [7]Department of Geography and Anthropology, Kennesaw State University, Kennesaw, GA, USA. [8]INRA-UMR BIOGECO, Cestas, France. [9]HelixVenture, Mérignac, France. [10]Instituto Nacional de Investigación Agropecuaria Estación Experimental Tropical Pichilingue, Quevedo, Provincia de Los Ríos, Ecuador. [11]INRA, GeT-PlaGe, Genotoul, Castanet-Tolosan, France. [12]New South Associates Inc., Stone Mountain, GA, USA. [13]Ministerio de Cultura y Patrimonio, Ecuador/IRD, Quito, Ecuador. [14]Department of Nutrition, University of California, Davis, CA, USA. [15]Institut de Recherche pour le Développement, UMR 208 PALOC, MNHN-IRD, Marseille, France. [16]These authors contributed equally: Sonia Zarrillo, Nilesh Gaikwad, Claire Lanaud. *e-mail: tmblake@mail.ubc.ca

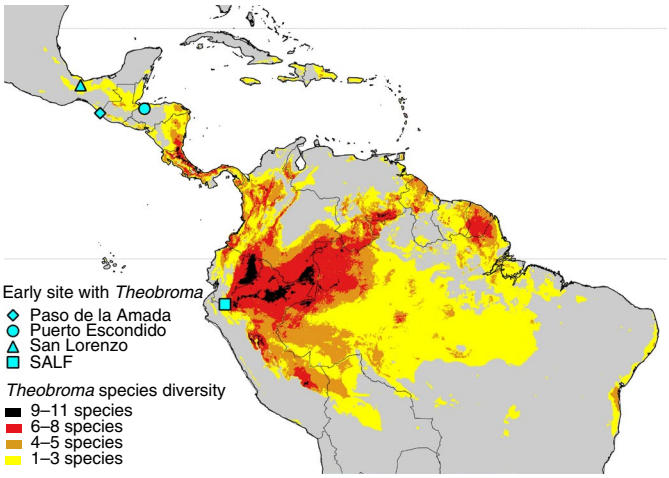


Fig. 1 | Locations of the four archaeological sites in Mexico, Central America and South America with the earliest evidence of *Theobroma* use, as well as the species diversity distributions for the genus *Theobroma*[17,19,58]. The zone surrounding SALF has six known native species of *Theobroma*: *T. bicolor*; *T. sinuosum*; *T. speciosum*; *T. subincanum*; *T. stipulatum*; *T. glaucum*[58]. The species diversity map is derived from ref. [17] and is modelled using species observations extracted from GBIF[58]. We reduced the modelled continuous 1–11 species diversity range of Thomas et al.[17] to four categories: 1–3, 4–5, 6–8 and 9–11 species.

Fig. 2 | Selected artefacts from SALF that tested positive for *T. cacao* and *Theobroma* spp. by aDNA, starch grains and theobromine analyses. **a**, PLD-001 human effigy stirrup-spout bottle. **b**, PLD-002 doughnut-shaped stirrup-spout bottle. **c**, PLD-003 incised-punctuated stone bowl. **d**, Interior of the ceramic sherd PLD-005 where carbon-14 dated charred residues were sampled.

and Supplementary Note 1). SALF was discovered in 2002 and excavated over several seasons by a joint team of Ecuadorian and French archaeologists[18,22]. More than 30 radiocarbon dates from the site place its initial occupation from 5,450 to 3,305 cal. yr BP (Table 1, Supplementary Note 1 and Supplementary Table 1)[18,22,23]. This small village functioned as a local ceremonial centre with the remains of approximately 20 buildings situated around a central sunken plaza (Supplementary Note 1 and Supplementary Fig. 1). At the eastern end of the village, an artificial platform overlooking the Río Valladolid supports a round structure with tombs containing elaborate funerary offerings, including the earliest examples of stirrup-spout bottles in the Americas[22,23] (Fig. 2, Supplementary Note 1 and Supplementary Fig. 1).

To test for the presence of *Theobroma* residues at SALF, we analysed artefacts from the tombs, middens, the sunken plaza floor, construction fill deposits, and a ceremonial hearth (Methods; Supplementary Note 1 and Supplementary Table 2). Artefacts tested from these contexts included whole and partial ceramic bowls, jars, and bottles, as well as stone bowls, mortars and one pestle. Residues from the interior of whole ceramic vessels were also tested. Subsets of these samples were analysed for the presence of diagnostic starch grains, theobromine and aDNA (Methods; Supplementary Notes 1 and 2 and Supplementary Table 2).

Results

Plants produce starch grains to amass energy and they are found in most plant tissues, particularly in storage and reproductive organs such as rhizomes, tubers and seeds. Starch grains are complex carbohydrates and not living biological cells. Thus, they do not contain mitochondrial or nuclear DNA. Starch grains recovered from charred cooking residues adhering to the interior of ceramic sherds, as well as residues from the interior of intact ceramic stirrup-spout bottles and the surfaces of stone artefacts, were the first clues that *Theobroma* was used at SALF[23–25]. Using previously reported methods[23], we analysed a total of 28 residue samples from 19 artefacts (Methods; Supplementary Table 2). After the identification of possible *Theobroma* spp. starch, and because there are few previous studies characterizing *T. cacao* seed starch[23,26–28] and none that describe seed starches of the wild species of *Theobroma* or *Herrania*, our comparative collection was expanded to include samples of modern seeds from both genera (Methods). In our reference samples, we confirmed a starch grain form distinctive to *Theobroma* spp. (Methods; Fig. 3) and found that our *Herrania* seed samples did not contain starch grains, thereby allowing us to confirm the identification of *Theobroma* spp. starches in the archaeological samples. Starch grains characteristic of *Theobroma* spp. were recovered from six of the 19 artefacts (32%), including the charred residues of a ceramic sherd directly dated to 5,450–5,300 cal. yr BP (PLD-005), as well as samples PLD-001, PLD-002 and PLD-003 (Figs. 2 and 3, Table 1, Supplementary Note 1 and Supplementary Tables 1 and 2).

Ancient pottery from SALF was low-fired, under 850 °C and therefore ideal for absorbing and retaining organic compounds on the interior of the vessels (Methods). Stone bowls, mortars and pestles were also porous enough to retain chemical residues. We tested samples for the presence of theobromine, a biochemical compound that is present in mature *T. cacao* seeds but absent in other *Theobroma* and *Herrania* wild species[29]. Using previously published methods[1,12], 81 burr or scraping samples and two residue samples were collected from ceramic artefacts and 98 samples (both dry and liquid) were recovered from stone artefacts (Methods; Supplementary Table 2) and all samples were then analysed by ultra-performance-liquid-chromatography-tandem-mass-spectrometry (UPLC-MS/MS) (Methods). Forty-six (28.6%) of our samples were positive for theobromine: 25 ceramic and 21 stone artefacts, including the oldest sample, PLD-005 and other key samples PLD-001, PLD-002 and PLD-003 (Figs. 2 and 4, Table 1, Supplementary Note 1 and Supplementary Tables 1 and 2).

Table 1 | Presence of *T. cacao* or its wild relatives in archaeological samples from SALF as evidenced by positive results from at least two analytical methods

Sample	Archaeological sample	20 cal. yr BP	Phase	Targeted aDNA amplifications			DNA capture		Theobromine presence (*T. cacao*)	Starch grains presence (*Theobroma* spp.)
				SNP specific to *T. cacao*	SNP specific to wild relatives	SNP specific to *T. cacao* or its wild relatives	*T. cacao* presence	Wild relatives presence		
PLD-001	Human effigy stirrup-spout bottle	4,146 to 3,932	Palanda	XXXX	No	No	Yes	No	Yes	Yes
PLD-002	Doughnut-shaped stirrup-spout bottle	4,090 to 3,980	Palanda	XX	No	No	Yes	Yes	Yes	Yes
PLD-003	Incised-punctuated stone bowl	4,150 to 3,920	Palanda	X	No	X	Not tested	Not tested	Yes	Yes
PLD-005	Charred ceramic residue	5,440 to 5,310[a]	Palanda	No	No	No	Yes	No	Yes	Yes
PLD-009	Stone mortar—bird figure		Palanda	No	X	X	Not tested	Not tested	Yes	Not tested
PLD-010	Stone mortar—human figure		Palanda	No	No	X	Not tested	Not tested	Yes	Not tested
PLD-020	Charred ceramic residue	3,980 to 3,835	Palanda	XX	XX	XXX	Not tested	Not tested	Yes	Not tested
PLD-025	Charred ceramic residue	3,980 to 3,835	Palanda	XXXX	XX	XXXX	Not tested	Not tested	Yes	Not tested
PLD-026	Charred ceramic residue	3,980 to 3,835	Palanda	XX	No	X	Not tested	Not tested	Yes	Not tested
PLD-030	Charred ceramic residue	3,980 to 3,835	Palanda	XX	No	XX	Not tested	Not tested	Yes	Not tested
PLD-038	Charred ceramic residue	3,980 to 3,835	Palanda	X	X	XXX	Not tested	Not tested	Yes	Not tested
PLD-040	Charred ceramic residue	3,980 to 3,835	Palanda	XX	X	XX	Not tested	Not tested	Yes	Not tested
PLD-042	Charred ceramic residue	3,980 to 3,835	Palanda	X	X	XXXX	Not tested	Not tested	Yes	Not tested
PLD-072	Ceramic residue	5,440 to 5,310[b]	Palanda	No	No	No	Yes	No	Yes	No
PLD-073	Ceramic residue	5,440 to 5,310[b]	Palanda	No	No	No	Yes	Yes	Yes	No
PLD-112	Direct collared jar—rim sherd		Palanda	No	X	No	Not tested	Not tested	Yes	Not tested
PLD-131	Incised painted bowl—body sherd	2,340 to 2,120	Tacana	X	No	X	Not tested	Not tested	Yes	Not tested

The positive results are reported for aDNA, theobromine and starch grain analyses. For targeted mitochondrial aDNA sequences, the number of X corresponds to the number of positive results obtained with different markers and repetitions for each archaeological sample. [a]Accelerator mass spectrometry radiocarbon date (Beta-312078) on charred residue containing starch grains, theobromine and aDNA (Supplementary Note 1 and Supplementary Table 1). [b]Ceramic sherds from the same context as PLD-005.

aDNA analysis provides a powerful tool to investigate plant and animal domestication histories[30], although aDNA plant studies are still few[31,32] and lacking samples collected from humid tropical environments. With experimental conditions adapted to scarcity and damage of aDNA due to post-mortem decay and deamination[30], aDNA was successfully extracted from the residues of the ceramic vessels we analysed (Supplementary Table 3). Two strategies were employed to identify aDNA of *T. cacao* and/or its wild relatives (*Theobroma* spp. and *Herrania* spp.) in the mixture of aDNA extracted from pottery residues.

First, markers specific to *T. cacao* or to its wild relatives were selected in two mitochondrial and four unique nuclear sequences (Methods; Supplementary Note 2, Supplementary Figs. 2–4 and Supplementary Table 4). Targeted PCR amplifications were carried out on aDNA extracts and negative controls, sequenced by next generation sequencing (NGS), blasted against the National Center

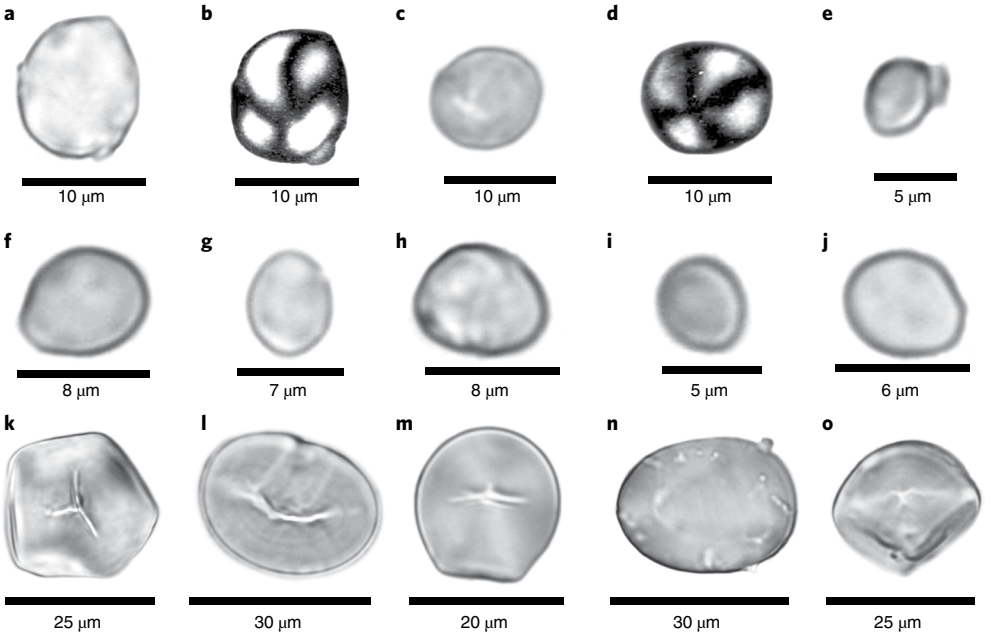


Fig. 3 | Starch grains of *Theobroma* spp. and other economically important plants. a,**b**, Archaeological starch from sample PLD-005, where **b** is the cross-polarized image. **c**–**o**, Images of modern reference starch granules: **c**,**d**, *T. cacao*, where **d** is the cross-polarized image; **e**, *T. angustifolium*; **f**, *T. bicolor*; **g**, *T. gileri*; **h**, *T. glaucum*; **i**, *T. grandiflorum*; **j**, *T. subinacum*; **k**, *Zea mays* (maize); **l**, *Phaseolus vulgaris* (common bean); **m**, *Manihot esculenta* (manioc/cassava); **n**, *Dioscorea trifida* (cush-cush yam/sacha papa) and **o**, *Ipomoea batatas* (sweet potato).

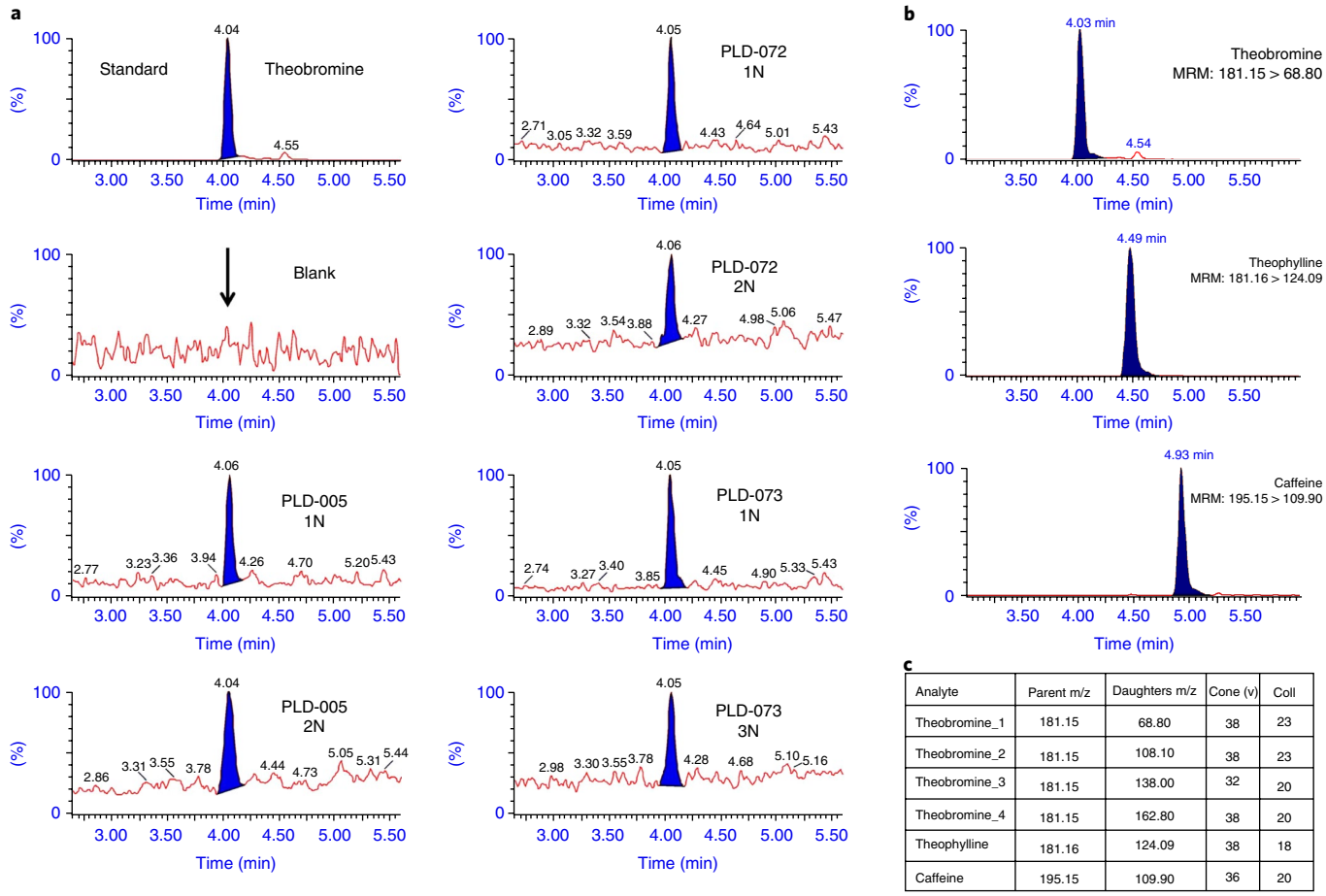
for Biotechnology Information (NCBI) international database and compared to modern reference sequences from *Theobroma*, *Herrania* (Supplementary Table 5) and to at least the ten closest species identified by BLASTn for each of the mitochondrial markers (Methods; Supplementary Note 2 and Supplementary Table 6). Among the 42 archaeological items analysed, ancient *Theobroma* or *Herrania* DNA amplifications were observed in mitochondrial fragments for 20 (47.6%) of them and 14 (35%) showed sequences specific to *Theobroma cacao*. PCR amplifications of unique nuclear sequences were only observed in one archaeological item and for only one nuclear fragment (DFR3) out of the four analysed (Table 1 and Supplementary Table 7).

In a second step, DNA capture[33] was carried out on a negative control and on five archaeological samples including two ceramic stirrup-spout bottles found in tombs, PLD-001 and PLD-002, dating to 4,070–3,830 cal. yr BP and 4,220–3,970 cal. yr BP, respectively, and three ceramic sherds, PLD-005, PLD-072 and PLD-073 dating to 5,450–5,300 cal. yr BP (Methods; Fig. 2, Table 1, Supplementary Notes 1 and 2, Supplementary Fig. 1 and Supplementary Tables 1 and 2). DNA fragments containing single nucleotide polymorphisms (SNPs) segregating in the *T. cacao* species, identified by genotyping by sequencing and corresponding to about 70% of gene fragments, were targeted.

After NGS deep sequencing of amplified captured DNA fragments, successive mapping of aDNA sequences on targeted *T. cacao* fragments (Supplementary Table 8) and on the whole cacao genome sequence[34], and a final screening by BLAST on the NCBI/nr database, aDNA sequences specific to *T. cacao* or to its wild relatives were not observed in the negative control but were observed in each analysed archaeological sample (Supplementary Tables 9 and 10).

Ceramic residues can contain a mixture of DNA from several possible species or varieties. To estimate their relative contributions to the DNA mixture extracted from each archaeological sample, we used a Bayesian model-based clustering method[35] to identify distinct genetic groups in a population and to assign individuals to genetic groups or identify admixed individuals. Nineteen SNPs, polymorphic between *Theobroma* and *Herrania* relative species or within the *T. cacao* species were identified in the selected aDNA sequences specific to *Theobroma*. These SNPs clearly differentiate *T. cacao* from the wild relative species, as well as several *T. cacao* genetic groups in a modern reference collection (Fig. 5 and Supplementary Table 11), which includes representatives of the several *T. cacao* genetic groups present in South America as well as *T. cacao* individuals native to the Zamora-Chinchipe province, where SALF is located. If we consider the $K=3$ to $K=5$ inferred cluster/population (Fig. 5), the presence of both wild relative species and *T. cacao* were observed in samples PLD-073 and PLD-002, whereas three other samples, PLD-001, PLD-005 and PLD-072, revealed SNP alleles specific only to *T. cacao*. For the $K=5$ inferred cluster/population, the Peruvian and Ecuadorian populations were differentiated, except for some Contamana individuals similar to Purus and Curaray individuals with these markers. Based on this analysis, the *T. cacao* ceramic residues of all archaeological samples analysed were similar to individuals from the Purus, Curaray and Contamana genetic groups, as well as to several individuals collected from Zamora-Chinchipe province.

To further support the antiquity of the DNA extracted from ceramic residues, typical signatures of post-mortem DNA damage were observed: (1) preferential (nearly unique) PCR amplifications of repeated mitochondrial sequences compared to nuclear sequences (Supplementary Tables 7 and 12); (2) high aDNA fragmentation with a small mean fragment size of 18.4 base pairs (bp), which is probably due to the humid tropical environmental conditions (Supplementary Fig. 5); (3) misincorporations typical to aDNA as already described[36] (Supplementary Table 13) observed at the ends of aDNA fragments (Supplementary Fig. 6); (4) misincorporations inside aDNA fragments (Supplementary Fig. 7 and Supplementary Table 14) and (5) a decreased PCR amplification intensity when increasing the length of amplified DNA fragments (Supplementary Fig. 8 and Supplementary Tables 15 and 16).



c

Analyte	Parent m/z	Daughters m/z	Cone (v)	Coll
Theobromine_1	181.15	68.80	38	23
Theobromine_2	181.15	108.10	38	23
Theobromine_3	181.15	138.00	32	20
Theobromine_4	181.15	162.80	38	20
Theophylline	181.16	124.09	38	18
Caffeine	195.15	109.90	36	20

Fig. 4 | UPLC-MS/MS chromatograms, multiple reaction monitoring transitions and method parameters. a, Chromatograms illustrating standard theobromine, blank and representative SALF samples confirming the presence of theobromine. **b**, Chromatograms of theobromine, theophylline and caffeine showing that these biomolecules are clearly resolved from one another and can be profiled with high accuracy. **c**, Multiple reaction monitoring (MRM) transitions and method parameters for theobromine, theophylline and caffeine analysis by UPLC-MS/MS.

Discussion

Ethnographic and ethnohistoric accounts describe different culturally important uses of cacao in South America versus Central and Mesoamerica[1–6,10–13,20,21,23]. In pre-Columbian Mesoamerica, cacao seeds were dried, ground and made into a variety of gruels and drinks (but not solid chocolate) by combining various ingredients and flavourings. Cacao was economically and politically important as the seeds were a key commodity in trade, were used as tribute payment and as currency and cacao was cultivated in large plantations. The various chocolate drinks, and especially the frothy foam created from them, held a prominent role in rituals, feasts and consumption by the elite, and specialized vessels were used in their preparation, storage and serving. Cacao trees, pods, seeds and drinks were also intimately associated with their myths and gods. Finally, our knowledge of the importance of cacao to pre-Columbian Mesoamerican cultures is not limited to the ethnographic and ethnohistoric records, as cacao features prominently in their texts: in their books, especially the Mayan Dresden and Madrid codices, on ceramic vessels and symbolized in Aztec carvings and paintings[5].

As the pre-Columbian cultures of South America had no known formal writing system, we must rely on ethnographic and ethnohistoric sources, as well as depictions on ceramics and carved in stone, to elucidate cacao's uses and importance. Domesticated cacao and its wild relatives' seeds, pulp and pods are consumed as food by many present-day Indigenous cultures in Ecuador[20,23]. The leaves and bark are used for medicinal extractions and infusions, and the trunks used in construction[20,23]. The pulp is eaten as an invigorating snack, made into a fresh juice and fermented to make an alcoholic beverage used in rituals and celebrations[20,23]. In addition to *T. cacao*, *T. grandiflorum* and *T. bicolor* are cultivated in Ecuador, and are all used in similar ways by Indigenous peoples, as are the other wild relative *Theobroma* and *Herrania* species[23]. These practises are not unique to cacao, as many plants have similar uses. While cacao seemingly did not play as paramount a role in the economic or spiritual life of South American cultures at the time of European contact as it did in Mesoamerica, it is possible that its importance waned over time. The disparity in cacao's prestige between the two continents, noted in early European accounts, has undoubtedly contributed to the long-held belief that cacao was initially domesticated in Mesoamerica.

Because macrobotanical remains are rarely preserved in humid tropical environments and any singular test may not be positive due to differential degradation of starch grains, organic residues and DNA, we used three independent methods to test for *T. cacao* and its wild relatives' presence at SALF. Our study is the first to use aDNA and starch grain analysis, together with theobromine, to test for cacao in archaeological contexts. We conservatively highlight samples PLD-001, PLD-002, PLD-003 and PLD-005 (Fig. 2 and Table 1), as these tested positive for cacao by all three methods—a further 13 samples tested positive by two methods (Table 1). We also directly

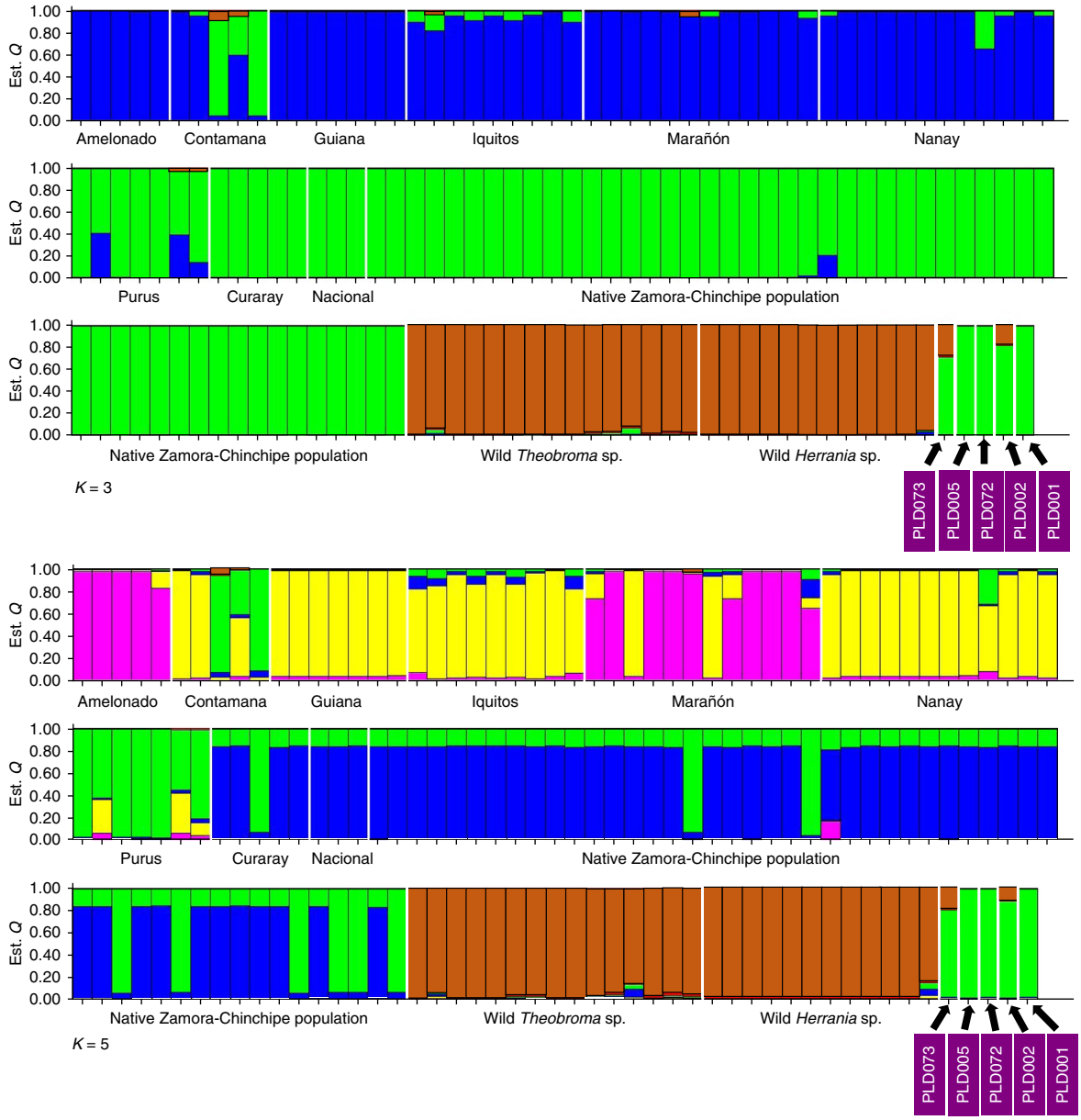


Fig. 5 | Population structure of modern individuals and aDNA mixtures extracted from ceramic residues. *Y* axis: plot of relative estimates of *Q* (estimated membership coefficient for each individual and in each cluster/population) for $K=3$ and $K=5$ inferred populations, estimated with 19 SNP markers. Each individual is represented by a single vertical line broken into *K* coloured segments, with lengths proportional to each of the *K* inferred clusters. Modern representatives of several cacao genetic groups, including trees prospected in the Zamora-Chinchipe region of SALF, are present in these graphics as well as aDNA extracted from artefacts' samples (PLD-001, PLD-002, PLD-005, PLD-072 and PLD-073).

dated the charred organic residue sample PLD-005 by accelerator mass spectrometry radiocarbon dating to 5,450–5,300 cal. yr BP (Methods; Table 1, Supplementary Note 1 and Supplementary Table 1) to erase any doubt about the antiquity of cacao at SALF. Our data show, by a convergence of evidence, a stable and continuous use of *T. cacao* and its wild relatives by the Mayo-Chinchipe people in several contexts, including ritual and domestic settings, from at least approximately 5,300–2,100 cal. yr BP (Table 1, Supplementary Note 1 and Supplementary Table 1). Their association with both domestic contexts and in mortuary offerings in the tombs of presumably high-status individuals indicates that these plants very likely served as an important source of food, drink, medicine, stimulants and for use in ceremonies at SALF.

Theobroma cacao diversity has recently been classified into ten genetic groups, with three main 'old domesticated' varieties. The geographic locale and the approximate dates of their origin are: the Criollo variety, cultivated in Mexico for approximately 3,900 years[2,14]; the Nacional variety, cultivated in Ecuador for at least 500 years[16] and the Amelonado variety, originating from Brazil and cultivated for nearly 300 years[37]. The Zamora-Chinchipe province has already been shown as the centre of domestication of the fine flavour cacao Nacional variety of Ecuador[16]. However, the origin of the aDNA from ceramic residues seems closer to the Purus and Curaray groups than to the Nacional variety, suggesting that the Nacional variety was domesticated more recently than the *Theobroma* trees used between approximately 5,300 and 3,700 cal. yr BP in the SALF region.

Representatives of the ancestral Criollo variety introduced in Central America, and cultivated by Olmec and Maya people, can be identified in the current cacao population cultivated in

Central America[14]. These trees, highly homozygous, constitute one of the ancestors of the present hybrid Criollo population (named Trinitario) and are also closer to the Purus or Curaray genetic groups[16], as is the aDNA from SALF. Modern *T. cacao* trees belonging to both genetic groups are presently found in the Zamora-Chinchipe region where SALF is located, and these trees extend north to the Colombian Amazon region. These findings highlight the important role the Curaray/Purus cacao groups in the upper Amazon may have played in cacao domestication events occurring at different times, but also raise questions regarding their dispersal from Amazonia to both the Pacific coast of Ecuador and to Central America.

In this regard, artefacts from SALF also reveal linkages to the Pacific coast. Funeral offerings from Tomb 2 included a stirrup-spout bottle (Fig. 2, Supplementary Note 1, Supplementary Fig. 1 and Supplementary Table 2), from which interior residue sample PLD-001 was obtained. Strikingly, the human effigy stirrup-spout bottle (Fig. 2) shows a human face emerging from a *Spondylus* bivalve (*Spondylus crassisquama* Lamark, 1819, formerly known as *S. princeps* Broderip, 1833), found only along the far northwest coast of Peru to Baja California[38]. Moreover, other artefacts found in tombs at SALF included shell fragments and beads made of *Strombus* spp. shell, which also originates along the Pacific coast[18] (Supplementary Note 1). It is evident then that the people of the Mayo-Chinchipe culture were in contact with groups on the Pacific coast, either through direct or indirect trade of goods[18,22]. These exchanges undoubtedly also involved culturally important plants[22,39].

By at least 5,300 cal. yr BP, the Mayo-Chinchipe people were using wild species at the same time as domesticated cacao, as shown by the presence of *Herrania* spp. and/or wild *Theobroma* spp. in some samples at SALF. As work continues to discover other archaeological sites of the Mayo-Chinchipe and other cultures in the upper Amazon region, even earlier examples of cacao domestication may yet be found. Our study suggests that the Mayo-Chinchipe people were involved in cacao domestication for several millennia, while continuing to use other related species.

Our findings support the hypothesis that *Theobroma cacao* was domesticated in South America at least 1,500 cal. yr BP before it was moved into Central America and Mesoamerica. This research constitutes the earliest evidence of *T. cacao* use in the Americas and reveals the upper Amazon region as the oldest centre of cacao domestication yet identified. Future research will focus on filling the gaps between the upper Amazon region, the Pacific coast and both Central and Mesoamerica in cacao's complex domestication history, hybridization, uses and the mechanisms of its spread. This adds one more example to a growing body of evidence for the early exchange of socially and economically important plants, technologies and ideas between the two continents, as well as the importance of the varied environments on the margins of the Amazon region in plant domestication[39,40].

Methods

Starch grain comparative samples and archaeological samples. *Theobroma cacao*, *T. angustifolium*, *T. bicolor*, *T. gileri*, *T. glaucum*, *T. grandiflorum*, *T. subincanum*, as well as *Herrania umbricata* and *Herrania* spp. whole-pod fresh samples were obtained from The United States Department of Agriculture, Agricultural Research Services, Plant Genetic Resources Conservation Unit and prepared for analysis by dissecting the pods, removing the seed and drying all parts in a food dehydrator set to <40 °C. Once dry, the seeds were gently ground and a small amount was suspended in 50/50 glycerine/distilled water, mounted on sealed microscope slides and the starch grains characterized. Comparative slides were also prepared from the dried fruits' mesocarp and endocarp (together, the 'flesh' or 'cortex'), as well as the pulp of each of the samples.

To guard against cross-contamination, all probes, knives, mortars and pestles, microscope slides, cover slips and any other instruments used were washed and then boiled in vinegar (5% acetic acid) in a pressure cooker to gelatinize/destroy starch granules and were not used across samples; this same method was used to clean all containers, instruments and microscope slides and cover slips for the archaeological samples, with separate sets of instruments used for the comparative samples verses archaeological samples. Comparative collection plant and slide preparations and archaeological sample processing were performed in separate rooms to avoid the possibility of contaminating the archaeological samples with modern starches. Microscope slides were viewed under ×400 magnification (Zeiss Axioskop 2) with plane and cross-polarized light to identify and confirm diagnostic forms and attributes of the starch grains.

Although not abundant, starch composes 4.5–7% (average 5.3%) of cocoa seeds[26]. Starch grains, lipid vacuoles and protein bodies are found in protein/lipid cells, the more abundant of the two types of parenchyma storage cells in cacao seed cotyledons; cells containing polyphenolic compounds form the other type[41,42]. Reichert[26] described the singular starch grains of dry seed cotyledons as 'rounded-angular', frequently with pressure facets and one curved surface (~8 μm), and compound grains with two to four, usually equal, granula. Schmieder and Keeney[27] highlight the compound form of cacao starch, elliptical and spherical forms were also found (2–12.5 μm), while Pagán Jiménez[28] characterizes cacao seed starch grains as being dominated by circular and oval shapes ranging in size from 3.3 to 10.5 μm (median 6.12 ± 2.1 μm). In our modern comparative samples, we did not find starch grains in any of the pods' flesh/cortex or pulp, nor were starch grains observed in seeds from three separate pods of *Herrania umbratica* and one *Herrania* spp. pod. All the *Theobroma* species seeds contained starch and these results allowed us to confirm the identification of *Theobroma* spp. starches in the archaeological samples.

Theobroma spp. seed starch granules can be characterized into three broad categories (based on 100 count)[23], and we consider the ovoid/teardrop shape (with narrower distal end) to be diagnostic for *Theobroma* spp. (Fig. 3):

- Spherical or sub-spherical (somewhat polygonal) with closed or slightly open centric or slightly eccentric hila (30%, 3.8–8 μm).
- Small compound granules composed of two or three truncated-spherical granula, with closed or slightly open centric or slightly eccentric hila (48%, 3.8–10 μm)
- Ovoid (teardrop-shaped) starches with eccentric hila located at the broad proximal end of the granule, distal end often pointed or very narrow in comparison to proximal end, closed or slightly open hila occasionally marked with one or more short fissures radiating laterally and/or towards the distal end and with a sharp extinction cross where the arms may bend or curve (22%, 6.3–12.5 μm).

We obtained and analysed 28 unique samples from 19 artefacts (Supplementary Note 1 and Supplementary Table 2), including ceramic sherds, intact ceramic bottles and stone bowls. The methods used to recover charred and organic residues from the different types of artefacts are in part described in the section 'Sampling stone artefacts and charred residues from ceramics for theobromine analysis'. The methods used to recover starch grains from the samples and analyse the results have been previously described in detail[23]. Techniques and methods used to avoid and test for modern starch grain contamination, including the extensive use of control samples, have also been previously reported[23]. A sample was considered 'positive' only when five or more starch grains were observed[23].

Theobromine analysis. We collected 181 separate samples from approximately 100 artefacts, including whole and partial ceramic vessels, stone bowls, stone mortar and one stone pestle from the SALF site (Supplementary Note 1 and Supplementary Table 2). Permission to analyse the SALF artefacts was granted by The Ecuadorian National Institute of Cultural Heritage (INPC), where the artefacts are presently stored and where the ceramic burr sampling was undertaken.

Sampling ceramics for theobromine analysis. SALF pottery is generally low-fired, typically under 850 °C and therefore ideal for absorbing and retaining organic compounds. Because there were no visible organics adhering to the interior of these pottery fabrics, except for samples from charred residues adhering to the interior of ceramic sherds that were sampled for starch grain analysis, as described below[23], chemical extraction techniques were necessary for confirmation. The interior surface of each ceramic vessel or sherd was lightly scraped using a new piece of fine-grained sandpaper to remove any substances that may have permeated the vessel wall. A burr from each sample, in the range 1–6 g, was captured on a new sheet of multipurpose white paper and the material funnelled into clean, previously unused collection vials and immediately sealed. New sheets of sandpaper and multipurpose white paper were used for each sample collected. This method was rigorously upheld throughout the collection process to eliminate potential cross-contamination of sample materials and to prevent contamination by modern theobromine residues[43]. Following collection, sealed vials were sent to the Gaikwad Laboratory at the University of California Davis, for analysis.

Sampling stone artefacts and charred residues from ceramics for theobromine analysis. Stone and ceramic artefacts were sampled in a laboratory dedicated to extraction of residues from archaeological samples at the Universidad San Francisco de Quito, Riobamba, Ecuador. To avoid contamination by modern starches and organic residues (that is, theobromine and modern cacao DNA), no food was allowed in the laboratory. The walls, lab benches, floor and surfaces

were washed with dilute bleach and then rinsed with water prior to any sampling procedures and the artefacts were sampled as quickly as possible with steps taken, as described below, to reduce exposure to air[23,43]. The laboratory does not have heating/cooling ventilation ducts, the windows were sealed and the door kept closed during all work. The tools were placed into individual, new and cleaned (boiled in 5% acetic acid in a pressure cooker, as described in the section 'Starch grain comparative samples and archaeological samples') containers and then placed onto a rack in an ultrasonic bath (without heat), loosely covered with aluminium foil and sonicated for 30 min. The sonicated samples were then rinsed into new, cleaned (as above) and labelled centrifuge tubes. These wet samples were centrifuged at 3,000 r.p.m. for 5 min and the supernatant decanted, leaving a 'residue' pellet. The centrifuge tubes were set in a rack, covered loosely with new aluminium foil and allowed to dry before sealing with the centrifuge tube caps, placed in new, labelled plastic ziplock bags and transported to Calgary. All ceramic sherds sampled for adhering charred residues were washed with distilled water and allowed to dry (loosely covered with aluminium foil) prior to sampling. This step was performed to remove adhering sediment and decrease the chance that starch granules and other modern residues present in the sediment might remain on the charred material. A separate clean (as described above) stainless steel dental pick was used for each sherd to gently scrape the charred residues onto new aluminium foil squares. The charred residues were then transferred to new, cleaned and labelled sample containers, placed in new, labelled ziplock bags and transported to Calgary[23]. Subsets of some of these samples were analysed for theobromine at the Gaikwad Laboratory at the University of California Davis and for aDNA as detailed below.

Sample preparation. Quantities of burr (17–379 mg) were extracted from each sample vial using stainless steel spatulas that were cleaned after each procedure to prevent potential cross-contamination during the analytical phase. Samples were incubated with 200 µl milli-Q water at 80 °C for 30 min. After incubation, samples were vortexed and centrifuged. The resulting sediment from each sample was removed and the supernatant filtered using 5 kD membrane filters. Filtrates were transferred to vials for UPLC-MS/MS analysis. Modern pottery samples were extracted and included in the analysis as controls.

UPLC-MS/MS analysis. Theobromine reference standard was purchased from Sigma-Aldrich. All solvents were HPLC grade and all other chemicals used were of the highest grade available. Theobromine stock solution of 1 mg ml^{-1} concentration was prepared in de-ionized water and was stored at 0 °C. Subsequently, a Xevo-TQ triple quadruple mass spectrometer (Waters) recorded mass spectrometry and tandem-mass spectrometry spectra using Electro Spray Ionization in positive ion mode, capillary voltage of 3.0 kV, extractor cone voltage of 3 V, sample cone voltage of 32 V and detector voltage of 500 V. Cone gas flow was set at 50 l h^{-1} and desolvation gas flow was maintained at 600 l h^{-1}. Source temperature and desolvation temperatures were set at 150 °C and 500 °C, respectively. The collision energy was varied from 16 to 26 to optimize four different daughter ions and a multiple reaction monitoring transition 181.15 > 68.80 was used in the analysis (Fig. 4). The acquisition range was 20–300 Da.

UPLC-MS/MS analyses of all the samplers were conducted using a Waters Acquity UPLC system connected with Xevo-TQ triple quadruple mass spectrometer. Analytical separations on the UPLC system were conducted using an Acquity UPLC C18 1.6 µm column (2 × 150 mm^2) at a flow rate of 0.15 ml min^{-1}. The gradient was started with 100% A (0.1% formic acid in H_2O) and 0% B (0.1% formic acid in CH_3CN), after 0.1 min changed to 80% A over 1 min and then 45% A over 5 min, followed by 20% A in 2 min. Finally, over 2 min, it was changed to 0% A, then after 13 min it was changed to the original 100% A over 1 min, resulting in a total separation time of 13 min. The elutions from the UPLC column were introduced to the mass spectrometer and resulting data were analysed and processed using MassLynx 4.1 software. Pure theobromine was used to optimize the UPLC conditions before analysis. The UPLC-MS/MS analyses of the SALF samples clearly showed the presence of a peak at 4.05 min (Fig. 4) that matched well with the standard for theobromine (Fig. 4, Standard Theobromine), whereas no peaks were observed at ~4.05 min in control samples (Fig. 4, Blank). The theobromine standard was run before the first sample, in middle of the run and after the last sample to prevent errors due to matrix effect and day-to-day instrument variations. In addition, immediately after the initial standard and before the first sample, two spiked samples were run to calibrate for the drift in the retention time due to the matrix effect. After standard and spiked sample runs, two blanks were injected to wash the injector and remove carry over effect.

Ancient DNA analysis. *Modern plant materials used for ancient DNA comparison.* Several collections of modern accessions of *T. cacao* and its wild relatives were used to compare aDNA sequences (Supplementary Tables 5 and 11). Indeed, it is known that wild relative species, and particularly *T. bicolor*, may also have been used for consumption as cacao[23,44].

Laboratory environment. We took adequate precautions to prevent contamination by modern, undamaged DNA, which would have been preferentially amplified if present in the DNA extracts. All pre-PCR experiments were conducted under sterile conditions in a dedicated laboratory (located at UMR BIOGECO-INRA Pierroton, France) routinely used for DNA analyses on ancient, rare and degraded plant DNA (Supplementary Note 2). All post-PCR experiments were carried out at the CIRAD genomics laboratory (Montpellier, France).

Ancient DNA extraction and purification. No conserved cacao organs or cacao seed fragments were found on the ceramics. Traces of charred cooking/food residues adhering to the interior of the ceramics sherds were collected (as described below and similar to the Methods section 'Sampling Stone Artefacts and Charred Residues from Ceramics for Theobromine Analysis'), as well as some of the underlying ceramic surface (matrix). The existence of aDNA in ceramic sherds may have been enabled by its absorption into the ceramics' matrices, as was certainly also the case for the theobromine remains found in many of the ceramics sampled by burring. The first step for sampling each ceramic sherd was to scrape them with a sterile scalpel after having removed, if necessary, the superficial soil layers, with the sampled residue rapidly placed in 2 ml sterile tubes. Samples from ceramic surfaces were also collected using rayon swabs (Copan-Italia) saturated with extraction buffer (0.1 M Tris, 0.45 M EDTA and 0.25 mg ml^{-1} proteinase K)[45]. The swabs were wiped against the inner walls of the ceramics and stored in their sterile box at 4 °C until returning to Europe, where all samples were stored at −20 °C pending DNA extraction. About ten swabs were collected per ceramic item. Two different methods, based on aDNA binding on silica membranes, were tested to extract aDNA jointly with negative controls (Supplementary Note 2).

Analyses of targeted PCR amplified ancient DNA fragments. *Identification of mitochondrial and nuclear DNA fragments suitable for aDNA analyses.* Primers were defined to amplify small-size fragments varying from 78 to 115 bp, adapted to post-mortem DNA damage[36,46,47]. The scarcity of aDNA in the archaeological samples also meant that repeated sequences had to be preferentially targeted to optimize the aDNA amplifications.

The sequencing and reconstitution of entire mitochondrial genome sequences from our aDNA samples was not possible in these experiments. Indeed, specific archaeological *T. cacao* organs were not available and the possible mix of aDNA from different species present in ceramic residues, associated with the high conservation of mitochondrial sequences between species, prevented the reconstitution of unique ancient *T. cacao* mitochondrial genomes. For this reason, we opted, in a first step, for the amplification of several targeted sequences from the mitochondrial and nuclear genomes (taken as control), including SNPs specific to *T. cacao* and/or its wild relatives.

Mitochondrial sequences. Several mitochondrial gene sequences were extracted from cacao transcriptomic data (http://esttik.cirad.fr/)[48] (Supplementary Note 2 and Supplementary Table 4).

Only two mitochondrial sequences for which the first hit was *Theobroma* after BLAST against the NCBI/EST database were conserved for aDNA analysis:

- MITO6: Cytochrome C Oxidase subunit 2 (CL294 Contig1): (position 1641–1740)
- MITO80: NADH dehydrogenase subunit 1 (CL1 Contig377): (position 1228–1309)

We designed all primers using the primerBLAST online tool from NCBI (http://www.ncbi.nlm.nih.gov/tools/primer-blast)[49].

Nuclear gene sequences. Primer pairs were defined inside several unique nuclear gene sequences, from the cacao Criollo genome sequence[34]:

- Vicilin gene (Tc04_g024090)
- FAT-B gene (Tc09_g010360.) acyl-ACP thioesterase fat B (*FATB*) genes
- Caffeine synthase gene (Tc10_g001820)
- Dihydroflavonol-4-reductase gene (Tc09_g034440)

To be able to provide complete and high-quality sequences of the targeted DNA fragments mentioned above from modern DNA of a collection of genetic resources (Supplementary Table 5) used as references, new primers enabling the amplification of larger fragments were defined (Supplementary Table 4).

Ancient DNA sequences were amplified by PCR or quantitative real-time PCR amplifications (Supplementary Note 2) and libraries were constructed with purified pooled PCR and quantitative PCR products made for each sample and subjected to (NGS) high-throughput sequencing (Supplementary Table 12).

Sequence analyses of amplified targeted DNA fragments. After sequencing, library trimming and mapping on the reference Criollo sequences (Supplementary Note 2), a selection of putative *T. cacao* sequences among the set of pre-selected aDNA sequences was carried out.

The identification of *T. cacao* sequences among the whole set of aligned aDNA sequences was difficult for several reasons:

1. The pottery residues contained a mixture of DNA from different species (plants, bacteria and so on), and due to the general high conservation of mitochondrial sequences between species, SNPs that are specific to *T. cacao* or related wild *Theobroma* species were not frequent and had to be identified for the detection of *T. cacao* sequences.

2. The DNA damage that impacts aDNA (modifying the resulting sequences), changes the way analyses of sequence similarity with modern cacao sequences must be carried out.

To overcome these problems the following strategy was chosen:

1. A collection of modern present-day DNA sequences was constituted for each marker, (Supplementary Figs. 2–4 and Supplementary Table 5) and sequences from the ten closest species to *T. cacao* were selected from the NCBI *EST* database (Supplementary Table 6).

2. For each marker, the sequences of this modern collection were aligned with the multiple sequence aligner ClustalW2 (http://www.ebi.ac.uk/Tools/msa/clustalw2/). To better highlight sequence polymorphisms, alignment was visualized with Jalview 2.8 software[50]. SNP or INDEL (insertion/deletion) alleles specific to *T. cacao* sequences or to *Theobroma*/*Herrania* sequences were identified.

3. The pre-selected aDNA sequences were then BLASTed against the NCBI *EST* database. The aDNA sequences for which the highest e-value corresponded to *T. cacao* sequences were then considered as putative *Theobroma* sequences.

The NCBI EST database does not include sequences from the wild *Theobroma*/*Herrania* relative species amplified by our markers. So, after a BLAST search against the NCBI database, even if the first BLAST hit was *Theobroma cacao*, only the search of SNP specific to the different *Theobroma*/*Herrania* species and *Theobroma cacao* allowed their origin to be distinguished. In most cases a second BLAST search of the aDNA sequences was made against the collection of sequences from modern *Theobroma*/*Herrania* accessions, allowing us to distinguish the closest *Theobroma* accessions.

Whenever very large numbers of aDNA sequences were selected after the mapping procedure following the NGS sequencing (as for MITO6 amplifications), a first screening based on the specific *Theobroma cacao* 7 bp SNP motif was applied before BLASTn (associated with a Phred quality score ≥ 20)[51], to reduce the number of BLASTn that needed to be carried out.

The selection of aDNA sequences, based on *T. cacao* specific SNP alleles, when they existed, helped to reinforce the hypothesis of the *T. cacao* origin of some sequences. However, it also excluded part of the putative aDNA *T. cacao* sequences for which a SNP allele was shared between *T. cacao* and its wild relatives. The presence of a specific *T. cacao* allele also depended on the allelic diversity that existed in this Amazonian region some 5,000 years ago.

Analyses of nuclear aDNA fragments homologous to *T. cacao* after DNA capture. *Library construction and sequencing.* Twelve Illumina compatible sequencing libraries were constructed from five archaeological samples: PLD-001, PLD-002, PLD-005, PLD-072 and PLD-073, as well as a negative control, using the NEXTflexTM Rapid DNA-Sequencing Kit (Bioo Scientific) according to the protocol indicated in the kit and pre-amplified to reach the DNA capture experimental requirements (Supplementary Note 2).

DNA capture. DNA capture has recently been proposed as a way to enrich the aDNA pool of sequences[33] and may target a larger number of sequences from diverse genome regions, including unique nuclear sequences.

A total of 4,847 unique nuclear *Theobroma cacao* sequences containing SNP sites were targeted for DNA capture to study the relatedness of aDNA sequences with modern accessions. These SNPs were identified by genotyping by sequencing on a collection of *Theobroma* and *Herrania* genetic resources (http://tropgenedb.cirad.fr/tropgene/)[52] and are located in genes for about 70% of them. Sequence fragments of 121 bp with a central position of the SNP were extracted from the *T. cacao* genome V2.0[53].

For each fragment, four 60 bp probes were designed according to Haak et al.[54] and Cruz-Dávalos et al.[55], with two probes per targeted SNP, containing each possible SNP allele in a central position and two other probes flanking the SNP. Between 300 ng and 500 ng of re-amplified aDNA were used for each library for the capture.

For target enrichment, we used a custom-designed Mybaits sequence capture kit (V 3.02) for targeted high-throughput sequencing, provided by the Microarray company and carried out the protocol indicated in the kit, except the hybridization step where an annealing temperature of 55 °C was applied as recommended by Cruz-Dávalos et al.[55] and during 48 h.

Following hybridization with RNA probes, 18 cycles of post capture re-amplification were performed, followed by a purification and sizing step made using ×1.2 to ×1.25 AMPure beads (Agencourt Ampure XP-Beckman Coulter Inc.) to eliminate the primer dimers (120 bp) before sequencing the samples by two lanes of HiSeq 3000 sequencing.

Selection of aDNA fragments homologous to T. cacao fragments containing SNPs. Sequences shorter than 30 bp were discarded from the analyses. Sequences specific to *Theobroma* were selected after library trimming and mapping successively on the 121 bp targeted DNA fragments and on the whole cacao genome sequence and BLASTed against the NCBI/nr database (Supplementary Tables 9 and 10).

Genetic structure analyses of aDNA mixtures. SNP alleles were then identified for each selected aDNA sequence and each locus; the corresponding selected SNP can be found in the Tropgene-DB (http://tropgenedb.cirad.fr/tropgene/)[52], as well as the genotyping, with the same SNP panel, of a collection of modern genetic resources used as reference (Supplementary Table 11).

We used a Bayesian model-based clustering method, implemented in the STRUCTURE software V2.3.4 (ref. [35]), to identify distinct genetic groups in a population and to assign individuals to genetic groups or identify admixed individuals. We applied this model to try to evaluate the genetic composition (*T. cacao* or its wild relatives) of the mixture of aDNA extracted from ceramic residues in comparison with the collection of modern accessions from *T. cacao* and its wild relatives. STRUCTURE was run under an admixture model, using a burn-in initial period of 100,000, a run length of 500,000 steps and ten independent runs for each sample where K values equalled from two to six.

Search for ancient DNA signatures. Ancient DNA is characterized by serious post-mortem damages[36] leading to both specific nucleotide misincorporation when amplified by PCR (Supplementary Table 13) and predominantly short-sized DNA fragments. Several analyses were conducted to evaluate the signatures of aDNA.

A part of the aDNA libraries were sequenced before DNA capture and the reads were independently mapped on the reference genome[34] using bwa 0.7.12-r1039 with a minimum base quality of 30 (-q 30) and with the seed option disabled (-l 1000)[56]. Sam files were converted to bam files, then sorted and indexed using Samtools 1.1. For each sample, the forward and reverse bam files were merged. The resulting bam file was used as input for mapDamage 2.0[57] to estimate DNA fragmentation and misincorporations observed at the ends of aDNA fragments. Misincorporations observed inside aDNA fragments were also analysed in a 100 bp mitochondrial fragment (Supplementary Note 2).

Data availability

NCBI GenBank accession numbers of the *Theobroma* mitochondrial reference sequences are MF462389, MF462390 and MF462396 to MF462398. Examples of PCR amplified mitochondrial ancient DNA sequences identified as *T. cacao* sequences are reported in Supplementary Figs. 2–4. All results on ancient DNA sequences obtained after DNA capture and containing SNPs are reported in Supplementary Information; corresponding SNPs from the collection of modern accessions, used as controls, are reported in the Tropgene database (http://tropgenedb.cirad.fr/tropgene/)[52]. Additional data that support the findings of this study are available from the corresponding author on reasonable request.

Received: 31 March 2018; Accepted: 14 September 2018;
Published online: 29 October 2018

References

1. Powis, T. G., Cyphers, A., Gaikwad, N. W., Grivetti, L. & Cheong, K. Cacao use and the San Lorenzo Olmec. *Proc. Natl Acad. Sci. USA* **108**, 8595–8600 (2011).
2. Powis, T. G. et al. Oldest chocolate in the New World. *Antiquity* **81** (2007); https://www.antiquity.ac.uk/projgall/powis314/
3. Henderson, J. S., Joyce, R. A., Hall, G. R., Hurst, W. J. & McGovern, P. E. Chemical and archaeological evidence for the earliest cacao beverages. *Proc. Natl Acad. Sci. USA* **104**, 18937–18940 (2007).
4. Sheets, P. D. *The Ceren Site: An Ancient Village Buried by Volcanic Ash in Central America* 2nd edn (Wadsworth Publishing, Belmont, 2005).
5. Coe, S. D. & Coe, M. D. *The True History of Chocolate* 3rd edn (Thames and Hudson, London, 2013).
6. Stuart, D. in *Chocolate in Mesoamerica: A Cultural History of Cacao* (ed. McNeil, C. L.) 184–201 (Univ. Press of Florida, Gainesville, 2009).
7. Crown, P. L. & Hurst, W. J. Evidence of cacao use in the Prehispanic American Southwest. *Proc. Natl Acad. Sci. USA* **106**, 2110–2113 (2009).
8. Washburn, D. K., Washburn, W. N. & Shipkova, P. A. Cacao consumption during the 8th century at Alkali Ridge, southeastern Utah. *J. Arch. Sci.* **40**, 2007–2013 (2013).
9. Washburn, D. K., Washburn, W. N., Shipkova, P. A. & Pelleymounter, M. A. Chemical analysis of cacao residues in archaeological ceramics from North America: considerations of contamination, sample size and systematic controls. *J. Arch. Sci.* **50**, 191–207 (2014).
10. Grivetti, L. H. & Shapiro, H.-Y. (eds) *Chocolate, History, Culture, and Heritage* (John Wiley & Sons, Hoboken, 2009).
11. Hurst, W. J., Tarka, S. M. Jr, Powis, T. G., Valdez, F.Jr & Hester, T. R. Cacao usage by the earliest Maya civilization. *Nature* **418**, 289–290 (2002).
12. Powis, T. G., Valdez, F., Hester, T. R., Hurst, W. J. & Tarka, S. M. Spouted vessels and cacao use among the Preclassic Maya. *Lat. Am. Antiq.* **13**, 85–106 (2002).
13. McNeil, C. L. *Chocolate in Mesoamerica: A Cultural History of Cacao* (Univ. Press of Florida, Gainesville, 2009).
14. Motamayor, J. C. et al. Cacao domestication I: the origin of the cacao cultivated by the Mayas. *Heredity* **89**, 380–386 (2002).
15. Motamayor, J. C. et al. Geographic and genetic population differentiation of the Amazonian chocolate tree (*Theobroma cacao* L.). *PLoS ONE* **3**, e3311 (2008).

16. Loor Solorzano, R. G. et al. Insight into the wild origin, migration and domestication history of the fine flavour Nacional *Theobroma cacao* L. variety from Ecuador. *PLoS ONE* **7**, e48438 (2012).
17. Thomas, E. et al. Present spatial diversity patterns of *Theobroma cacao* L. in the neotropics reflect genetic differentiation in Pleistocene refugia followed by human-influenced dispersal. *PLoS ONE* **7**, e47676 (2012).
18. Valdez, F. in *The Handbook of South American Archaeology* (eds Silverman, H. & Isbell, W. H.) 865–888 (Springer, New York, 2008).
19. Cuatrecasas, J. Cacao and its allies, a taxonomic revision of the genus.*Theobroma. Contr. US Natl Herb.* **35**, 379–614 (1964).
20. Walker, T. in *Chocolate: History, Culture, and Heritage* (eds Grivetti, L. E. & Shapiro, H.-Y.) 543–558 (John Wiley & Sons, Hoboken, 2009).
21. Bletter, N. & Daly, D. C. in *Chocolate in Mesoamerica: A Cultural History of Cacao* (ed. McNeil, C. L.) 31–68 (Univ. Press of Florida, Gainesville, 2009).
22. Valdez, F., Guffroy, J., de Saulieu, G., Hurtado, J. & Yepes, A. Découverte d'un site cérémoniel formatif sur le versant oriental des Andes. *C. R. Palevol.* **4**, 369–374 (2005).
23. Zarrillo, S. *Human Adaptation, Food Production, and Cultural Interaction During the Formative Period in Highland Ecuador.* PhD dissertation, Univ. Calgary (2012).
24. Lanaud, C., Loor, R. G., Zarrilo, S. & Valdez, F. Origen de la domesticacion del cacao y su uso temprano en el Ecuador. *Nuestro Patrimonio* **12**, 12–14 (2012).
25. Zarrillo, S. & Valdez, F. in *Arqueologia Amazonica: las Civilizaciones Ocultas del Bosque Tropical* (ed. Valdez, F.) 147–171 (Abya-Yala, Ecuador, 2013).
26. Reichert, E. T. *The Differentiation and Specificity of Starches in Relation to Genera, Species, Etc.* (Carnegie Institution of Washington, Washington DC, 1913).
27. Schmieder, R. L. & Keeney, G. Characterization and quantification of starch in cocoa beans and chocolate products. *J. Food. Sci.* **45**, 555–557 (1980).
28. Pagán Jiménez, J. R. *Almidones: Guía de Material Comparativo Moderno del Ecuador Para los Estudios Paleoetnobotánicos en el Neotrópico* (Aspha, Buenos Aires, 2015).
29. Hammerstone, J. F., Romanczyk, L. J. & Martin Aitken, W. Purine alkaloid distribution with *Herrania* and *Theobroma*. *Phytochemistry* **35**, 1237–1240 (1994).
30. Pääbo, S. et al. Genetic analyses from ancient DNA. *Annu. Rev. Genet.* **38**, 645–679 (2004).
31. Parducci, L. et al. Molecular- and pollen-based vegetation analysis in lake sediments from central Scandinavia. *Mol. Ecol.* **22**, 3511–3524 (2013).
32. Willerslev, E. et al. Fifty thousand years of Arctic vegetation and megafaunal diet. *Nature* **506**, 47–51 (2014).
33. Carpenter, M. L. et al. Pulling out the 1%: whole-genome capture for the targeted enrichment of ancient DNA sequencing libraries. *Am. J. Hum. Genet.* **93**, 852–864 (2013).
34. Argout, X. et al. The genome of *Theobroma cacao*. *Nat. Genet.* **43**, 101–108 (2011).
35. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959 (2000).
36. Hansen, A., Willerslev, E., Wiuf, C., Mourier, T. & Arctander, P. Statistical evidence for miscoding lesions in ancient DNA templates. *Mol. Biol. Evol.* **18**, 262–265 (2001).
37. Silva, P. & Oliveira-Cardoso, A. Historico das introduçoes de cacaueiro (*Theobroma cacao* L.) no Reconcavo da Bahia, Brasil. *Revista Theobroma* **10**, 135–140 (1980).
38. Mackensen, A. K., Brey, T., Bock, C. & Luna, S. *Spondylus crassisquama* Lamarck, 1819 as a microecosystem and the effects of associated macrofauna on its shell integrity: isles of biodiversity or sleeping with the enemy? *Mar. Biodivers.* **42**, 443–451 (2012).
39. Piperno, D. R. & Pearsall, D. M. *The Origins of Agriculture in the Lowland Neotropics* (Academic Press, New York, 1998).
40. Hilbert, L. et al. Evidence for mid-Holocene rice domestication in the Americas. *Nat. Ecol. Evol.* **1**, 1693–1698 (2017).
41. Biehl, B. Veränderungen der subcellulären struktur in keimblättern von kakaosamen (*Theobroma cacao* L.) während der fermentation und troeknung. *Z. Lebensm. Unters. Forch.* **153**, 137–150 (1973).
42. Redgwell, R. J. & Hansen, C. E. Isolation and characterisation of cell wall polysaccharides from cocoa *(Theobroma cacao* L.) beans. *Planta* **210**, 823–830 (2000).
43. King, A., Powis, T. G., Cheong, K. & Gaikwad, N. Cautionary tales on the identification of caffeinated beverages in North America. *J. Archaeol. Sci.* **85**, 30–40 (2017).
44. Kufer, J. & McNeil, C. L. in *Chocolate in Mesoamerica: A Cultural History of Cacao* (ed. McNeil, C. L.) 384–407 (Univ. Press of Florida, Gainesville, 2009).
45. Cieslak, M. et al. Origin and history of mitochondrial DNA lineages in domestic horses. *PLoS ONE* **5**, e15311 (2010).
46. Pääbo, S. Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. *Proc. Natl Acad. Sci. USA* **86**, 1939–1943 (1989).
47. Dabney, J., Meyer, M. & Pääbo, S. in *DNA Repair, Mutagenesis, and Other Responses to DNA Damage* (eds Friedberg, E. C., Elledge, S. J., Lehmann, A. R., Lindahl, T. & Muzi-Falconi, M.) 19–26 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 2013).
48. Argout, X. et al. Towards the understanding of the cocoa transcriptome: production and analysis of an exhaustive dataset of ESTs of *Theobroma cacao* L. generated from various tissues and under various conditions. *BMC Genomics* **9**, 512 (2008).
49. Ye, J. et al. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinform.* **13**, 134 (2012).
50. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G J. Jalview version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).
51. Ewing, B., Hillier, L., Wendl, M. C. & Green, P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* **8**, 175–185 (1998).
52. Hamelin, C., Sempere, G., Jouffe, V. & Ruiz, M. TropGeneDB, the multi-tropical crop information system updated and extended. *Nucleic Acids Res.* **41**, D1172–D1175 (2013).
53. Argout, X. et al. The cacao Criollo genomev2.0: an improved version of the genome for genetic and functional genomic studies. *BMC Genomics* **18**, 730 (2017).
54. Haak, W. et al. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207–211 (2015).
55. Cruz-Dávalos, D. I. et al. Experimental conditions improving in-solution target enrichment for ancient DNA. *Mol. Ecol. Resour.* **17**, 508–522 (2017).
56. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
57. Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F. & Orlando, L. mapDamage 2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**, 1682–1684 (2013).
58. *GBIF Backbone Taxonomy* (GBIF Secretariat, 2017); https://doi.org/10.15468/39omei

Acknowledgements

S.Z.'s research was funded by a Social Sciences and Humanities Research Council of Canada Doctoral Fellowship and an Honorary Isaak Walton Killam Memorial Scholarship, as well as the Martha Biggar Anders Memorial Award (2008, 2009, 2011), a Graduate Research Scholarship, a Faculty of Graduate Studies Travel Award, a Graduate Studies Scholarship and a Graduate Student Association Professional Development Grant from the University of Calgary. We thank the Universidad San Francisco de Quito, Riobamba and the University of British Columbia, Okanagan for use of laboratory facilities for some of S.Z.'s research. Funding for this project was provided by a Hampton Research Grant (no. F11-00878) from the University of British Columbia. Funding for archaeological excavations at SALF was provided by IRD. This project is supported by Agropolis Fondation under the reference ID 1202-029 through the Investissements d'avenir programme (Labex Agro: no. ANR-10-LABX-0001-01). The Hiseq sequencing activities were made in collaboration with the GeT platform, a partner of the National Infrastructure France Génomique, with thanks for support by the Commissariat aux Grands Investissements (no. ANR-10-INBS-0009). We are grateful to P. Lachenaud for providing genotyping by sequencing data on Guina group, to H. Kucera for assisting with Biochemical analyses, to N. Waber for preparing Fig. 1 and M. Berard for assistance in preparing Figs. 2 and 3. We also thank The United States Department of Agriculture, Agricultural Research Services, Plant Genetic Resources Conservation Unit for generously supplying *Theobroma* and *Herrania* pods and the Centro Agronómico Tropical de Investigación y Enseñanza (CATIE—Costa Rica) and the Cocoa Research Center, University of West Indies (CRC, Trinidad and Tobago) for providing *Theobroma* and *Herrania* leaves for modern DNA analyses.

Author contributions

F.V., M.B., S.Z., T.P., N.G. and C.L. designed the research. F.V., J.H., A.Y. and S.Z. performed excavations at Santa Ana-La Florida. S.Z. designed starch investigation methods and performed starch granule analysis. T.P. and P.S. sampled artefacts for theobromine analysis. N.G. designed mass spectrometry analysis, performed UPLC-MS/MS analyses and processed and analysed mass spectrometry data. L.G. provided general input to the overall project. C.L., C.V., I.L., O.F., X.A., E.G., F.S. and R.L.S., performed aDNA experiments and analyses. H.V. and O.B. performed NGS aDNA sequencing. S.Z., C.L., N.G., T.P., M.B. and F.V. led the writing of the paper with inputs from all other authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41559-018-0697-x.

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

Correspondence and requests for materials should be addressed to M.B.

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