nature ecology & evolution

SUPPLEMENTARY INFORMATION

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

In the format provided by the authors and unedited.

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]

[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

SUPPLEMENTARY INFORMATION

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

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

Affiliations:

[1]Department of Anthropology and Archaeology, University of Calgary, 2500 University Drive NW, Calgary, AB, Canada, T2N 1N4.
[2]Department of Anthropology, University of British Columbia, 6303 NW Marine Drive, Vancouver, BC, Canada, V6T 1Z1.
[3]Department of Nutrition and Department of Environmental Toxicology, West Coast Metabolomics Center, University of California, 3209 Meyer Hall, Davis, CA, USA, 95616.
[4]Gaikwad Steroidomics Laboratory, Davis, CA 95616, USA
[5]CIRAD, UMR AGAP, F-34398 Montpellier, France.
AGAP, Univ Montpellier, CIRAD, INRA, Montpellier SupAgro, Montpellier, France.
[6]Department of Geography and Anthropology, Kennesaw State University, 402 Bartow Avenue, Kennesaw, GA, USA, 30144.
[7]INRA-UMR BIOGECO 69 route d'Arcachon 33610 Cestas, France.
[8]HelixVenture, Mérignac, France.
[9]Instituto Nacional de Investigación Agropecuaria (INIAP) Estación Experimental Tropical Pichilingue Km. 5 vía Quevedo-El Empalme, Cantón Mocache, Provincia de Los Ríos, Ecuador.
[10]INRA, GeT-PlaGe, Genotoul, UAR1209 Castanet-Tolosan, France.
[11]New South Associates, Inc., 6150 East Ponce De Leon Ave, Stone Mountain, GA, 30083.
[12]Ministerio de Cultura y Patrimonio, Ecuador / IRD Ecuador.
[13]Department of Nutrition, University of California, 3209 Meyer Hall, Davis, CA, USA, 95616
[14]Institut de Recherche pour le Développement (IRD), UMR 208 PALOC, MNHN-IRD.

[†]These authors contributed equally to this work

*Correspondence to: MB tmblake@mail.ubc.ca

This section includes Supplementary Information:

1.0 Supplementary Note 1: Archaeological contexts, dates, and samples

The Santa Ana-La Florida (SALF) site is located near the modern town of Palanda, in the Zamora-Chinchipe (ZCH) province of south eastern Ecuador[1,2] (Fig. 1, Supplementary Fig. 1). SALF is in the upper Amazon, at the margins of the headwater of the Mayo Chinchipe basin that flows into the Marañon River, a main tributary of the Amazon River. It covers an area of ~1 ha at the bottom of a steep and narrow river valley at an altitude of 1140 masl, in the warm and humid transition zone, the cloud forest, between the Amazon lowlands and the Andean highlands. The site occupies the flat portion of a river terrace, containing the architectural remains of ~20 buildings structured around a circular sunken plaza (Supplementary Fig. 1, indicated by B).

Two platforms are aligned at the eastern and western ends of the site, each one holding a focal point in the architectural layout. At the western end of the sunken plaza, the slanting terrain has been transformed to create a five-stepped platform that likely dominated all the activities that were carried out in and around the plaza. At the eastern end of the site, an oval shaped artificial mound, of around 900 m^2, stands some four meters above the surrounding inclined part of the terrace adjacent to the river (Supplementary Fig. 1, as indicated by A). The part of the mound that faces the central plaza was the base of a round structure, now called the "temple" due to its peculiar spiral shaped architecture, at the centre of which lies a ceremonial hearth (Supplementary Fig. 1, as indicated by A and inset picture).

Evidence of ceremonial activity has been found in the hearth itself: a cache containing two polished stone bowls, two greenstone mask pendants, and several hundred turquoise beads. Under the stone spiral alignment five tombs were excavated in the body of the platform, but several other tombs were apparently destroyed some twenty years ago when a road was cut through the jungle and a bridge was built over the river (Supplementary Fig. 1). The funerary offerings found in the different tombs include: fine grained polished stone dishes, bowls, and mortars; ceramic vessels of different forms; and a variety of personal adornments made of greenstones, rock crystals and fragments; and beads made of Pacific Ocean seashells.

Although these traits are not unusual in most of the major pre-Columbian societies of the Andes, two peculiarities stand out in the case of the SALF site: its antiquity and its location in the upper Amazon. SALF is a multicomponent site that has at least three major pre-Columbian occupations[1,2], with these defined in terms of archaeological phases throughout the ZCH province:

1. Bracamoro phase, ca. 900 - 1950 AD;
2. Tacana phase, ca. 1800 - 200 BC;
3. Palanda phase, ca. 3500 - 2000 BC.

Transitional phases exist but have not been entirely defined yet; refinements will probably proceed as the survey continues throughout the province. Thirty-two radiocarbon dates support the association with each of the different phases[1-3] (Table 1; Supplementary Table 1). Twenty-eight dates, ranging between 4620 ± 30 to 2210 ± 40 radiocarbon yrs BP (5500 to 2340 cal. yrs

BP, 2 σ Cal.), are associated with the early occupation of the site, and 25 dates place the main occupation between 3860 ± 40 to 3430 ± 40 radiocarbon yrs BP (4410 to 3880 cal. yrs BP, 2 σ Cal.). The dates were obtained from charcoal taken from closed, undisturbed contexts: burnt occupational floors; middens; hearths; charred cooking residues from a ceramic sherd, and burnt materials found in some funerary chambers[1-3] (Table 1; Supplementary Table 1).

We assume that the site started out as a small settlement on the river terrace. A few domestic units constituted a hamlet that eventually grew into a village, organized around a central public space. In time, this space became a sunken plaza encircled by a stone retaining wall and complemented by two platforms aligned on an east-west axis (Supplementary Fig. 1). The architectural layout, the characteristics of the eastern platform, and the meagre amount of residential evidence found throughout the site suggest that SALF eventually became an Early Formative Period ceremonial centre[1,2].

Evidence of the initial settlement was in the lower strata of the terrace. These appear as scattered domestic midden residues (ceramic sherds, charcoal, lithic chips, and organic soil clusters) that were covered by earth and cultural materials of the later occupations. Two such contexts were found under the eastern mound and have been dated by two radiocarbon assays. The oldest date (Beta-312078) was obtained from charred organic cooking residue adhering to the interior of a ceramic sherd (PLD-005) recovered from a midden in the X4(17) unit at a depth between 60 and 80 cm under the modern surface[3] (Table 1; Supplementary Fig. 1; Supplementary Tables 1 and 2). Part of the charred cooking residue was also tested for starch granules[3], theobromine, and aDNA (Sample PLD-005) (Table 1; Supplementary Tables 1 and 2). The other part was used for the assay that dates this evidence at 5450 to 5300 cal. yr BP[3] (Table 1; Supplementary Tables 1 and 2).

The other early settlement date (Beta-197175) was taken from charcoal remains found in a midden in the XII 6 unit, at a depth of 150 cm (Supplementary Fig. 1; Supplementary Tables 1 and 2). The context of this midden is particularly interesting since the organic soil that contains the dated sample had a diagnostic colour and texture that was later identified in other parts of the site, indicated that these areas also dated to the same time. The sherds found in the strata served to build a typological characterization of the early phase Palanda materials[1,2].

A total of 178 samples were tested for either starch grains, theobromine, or aDNA (Supplementary Table 2). Due to sample availability and laboratory schedules, we did not test all samples by all three analytical methods. Four samples that were subjected to all three analytical tests (PLD-001, PLD-002, PLD-003, and PLD-005) were positive for *Theobroma* by each of the different methods and a further 13 were positive for *Theobroma* by at least two analytical methods, with 12 of these dating to the (earliest) Palanda phase (Table 1; Supplementary Table 2).

Some of the most important samples were recovered from ceramic and stone vessels found in the sealed funerary contexts discovered in the body of the eastern platform (Supplementary Fig.1; Supplementary Table 2). Five tombs with funerary objects were excavated; each one was different, but in general all fall around the same time period: 3700 ± 40 radiocarbon yr BP (4400 to 3990 cal. yr BP, 2 σ Cal.) (Table 1, Supplementary Table 1). As we have focused on the results that tested positive for cacao by all three analytical techniques (starch grains,

theobromine, and aDNA), we describe Tombs 2 and 4 below, where the artefacts resulting in samples PLD-001, PLD-002, and PLD-003 originated from. Sample PLD-005, was obtained from the charred residue adhering to the interior of a sherd recovered from a domestic midden, as described above. Samples PLD-072 and PLD-073, important to the discussion of the aDNA results and association of the Curaray/Purus genetic groups in cacao domestication, were obtained from ceramic sherds recovered from the same midden as PLD-005 and are thus contemporaneous (Table 1; Supplementary Table 1 and 2).

Tomb 2 (units XII, 4 and 5) was probably the most important cultural feature discovered at the site. Samples PLD-001 and PLD-002 were obtained from the human effigy stirrup spout bottle and donut-shaped stirrup spout bottle, respectively (Fig. 2, Supplementary Table 1). The tomb was located just a meter to the north of the ceremonial spiral hearth in heart of the "temple" (Supplementary Fig.1, inset). The excavation of this structure revealed at thick burnt floor that hardened and sealed the upper surface of the platform. In the process, a small portion of a stone slab was cleared on the singed

ground; further work revealed a round depression on the burnt surface. Charcoal samples were taken from the upper portion of the burnt floor that sealed the entrance to the tomb. This feature was dug methodically, exposing the mouth of a stone lined circular shaft that was carefully emptied. Three long stone slabs had been vertically placed in the pit, along with earth and other different sized stones. As the exploration of this featured deepened, turquoise beads were found randomly at different depths of the shaft, until the chamber was reached at 198 cm below the ground surface. Offerings and personal adornments began to appear as the sediments covering the deposit were removed. An oval shaped chamber, approximately 2 m in diameter, contained the remains of at least two individuals and was furnished with different types of funerary offerings. Eight ceramic vessels, including the human effigy and donut-shaped stirrup spout bottles, were in the north/northeast section of the chamber, three polished stone bowls were found in the eastern part, and two interesting instruments were found in the southern part of the tomb. These were a bird-shaped stone mortar and an anthropomorphic ceramic lime pot.

The positions of the vessels, the concentrations of personal adornments, and the presence of meagre residues of badly preserved bone suggested the original location of the bodies. The position of the vessels, and of the paraphernalia in general, suggested that the offerings were originally wrapped in textiles or other organic materials that eventually disappeared with time. The displacement of certain elements suggested that the tomb had been reopened at least on a second occasion, with the probable addition of a new funerary package. These actions probably broke some of the vessels, as some fragments were dispersed in different parts of the tomb. Probably the most unexpected find was the presence of several pieces of seashells among some of the ornaments. The seashells were identified as *Strombus* spp. and this implies that interactions with the Pacific coast were a reality at the time.

The tomb and its contents were dated through the assays made from the charcoal of the burnt floor that sealed the entrance of the shaft and from other charcoal materials taken from the interior floor of the funerary chamber. Both radiocarbon assays (Beta-214742 and Beta-197176) (Supplementary Table 1) are consistent and date the tomb to ca. 4400 to 3970 cal. yr BP.

Tomb 4 (units XIV 4 and 5) fit a pattern that has been registered in other instances in the Andean world cosmology: propitiatory burials at the base of walls or buildings. The tomb was made over the geological base of the inclined ancient riverbed. A crude rectangular stone structure had been built to contain the tomb, and this could have induced a better preservation of the remains. Tomb 4 provided evidence for the original position of the body; it had been placed on a large flat boulder at the base of one of the retaining walls that held the eastern part of the platform. Some of the organic remains were better preserved and the position of the corpse could be established. The body was discovered in a flexed foetal position; offerings were placed around and above the deceased. Two *Strombus* spp. fragments were found over the thorax area, among turquoise beads. Four polished stone bowls, including the incised-punctuated stone bowl from which sample PLD-003 was recovered (Fig. 2; Supplementary Table 2), and two ceramic vessels were the main paraphernalia that accompanied the individual. The main part of the burial had probably been placed in a funerary bundle that did not preserve. Charcoal samples were taken from the fill materials found in the makeshift chamber of this tomb. The radiocarbon assay (Beta-261402, 4150 to 3920 cal. yr BP) (Supplementary Table 1) was consistent with dates obtained from the other funerary contexts discovered in the platform.

2.0 Supplementary Note 2: Ancient DNA analyses

2.1 *Modern collection of genetic resources used as reference for aDNA analyses*

2.1.1 Targeted DNA:
A collection of 44 accessions (Supplementary Table 4) representing the diversity of the *T. cacao* species, (including seven accessions native to the Palanda region), and representatives of wild *Theobroma* (7 sp.) and *Herrania* (6 sp.) (a closely related genus) were used to identify SNPs or deletion/insertion (INDEL) specific to *T. cacao* or to its wild relatives in targeted PCR amplifications made with 11 selected mitochondrial and nuclear markers.

2.1.2 Genetic structure of *Theobroma* aDNA mixture:
A collection of modern accessions previously genotyped by GBS (genotyping by sequencing) was used as a reference to evaluate the relative contribution of *T. cacao* or its wild relatives to the aDNA mixtures present in the ceramic residues, after admixture analyses conducted with the STRUCTURE software. This collection includes 52 representatives of *T. cacao* collected in the Zamora Chinchipe region where SALF located[1], 65 accessions from various *T. cacao* genetic groups and wild *Theobroma* (15 accessions from 7 species), and *Herrania* (12 accessions from 9 species) representatives (Supplementary Table 11).

2.2 *Laboratory Environment*

The dedicated laboratory is separated from the main laboratory where PCR amplifications and subsequent analyses took place and is equipped with a high-pressure air system, with continuous filtering of incoming air, daily UV light irradiation, laminar flow hoods with HEPA filters and all surfaces are frequently cleaned with ethanol and Actril Sterilant. Three separate dedicated rooms in the laboratory were used to conduct: 1) DNA isolation; 2) PCR reagent preparation and mixing of PCR reagents; and, 3) DNA mix preparations made for PCR experiments or for aDNA

library constructions. All equipment was DNA-free, and the reagents used for experiments were all new and unopened. The experimenters wore whole-body protective clothing including gloves and shoe protection and took specific precautions during handling to avoid contamination between samples and the introduction of extraneous DNA. In addition, this laboratory had never worked on *Theobroma* species.

2.3 *DNA extraction*

2.3.1 Protocol A: Two types of samples were obtained and analysed: 1) about 250 mg of residue scraped from the interior surface of ceramic sherds (the residue consisting of charred cooking/food residue and ceramic matrix); or 2) small pieces of two rayon swabs that had been rubbed on the interior surface of a ceramic sherd or vessel. The samples were placed in sterile 2 ml tubes containing 2 tungsten balls and 1 ml of extraction buffer,[4] and ground for 1 min x 30 vibrations per min using a tissue lyser apparatus (TissueLyser II, Quiagen).

The tubes were incubated overnight at 50°C with slow stirring.

The solution was centrifuged at 10,000 g for 10 min at room temperature and the supernatant transferred to a Macherey column (Macherey-Nagel NucleoSpin 96 plant II kit) enabling binding of the DNA on a silica membrane. The Macherey protocol was followed from step 4 onwards only, corresponding to the adjustment of the binding conditions. After DNA binding and three washes with PW buffers, DNA was eluted in 100 µl of preheated PE buffer (Macherey protocol).

The DNA solution was then purified with the OneStep™ PCR Inhibitor Removal Kit (Zymo Research), according to the manufacturer's recommendations, and subsequently stored at 20°C.

2.3.2 Protocol B: The (MoBio), which effectively removes PCR inhibitors such as humic acids, was used to extract DNA from about 250 mg of solid residues from scraped sherds or small pieces cut from two rayon swabs. This protocol, normally used to extract DNA from soil, was used (as suggested by E.M. Geigl) because absorption of DNA on soil particles could present similarities with absorption of DNA on ceramic particles. Moreover, it is also adapted to remove PCR inhibitors present in soils, but which are also known to be present in ancient DNA, such as humic acids.

The DNA extraction protocol was applied according to the manufacturer's recommendations. Two sets of extractions were carried out:

1) In the first set, ancient DNA was extracted from 38 archaeological samples using both extraction methods. Several DNA extractions were repeated for each archaeological sample, jointly with four negative controls. These DNA extracts were used to amplify targeted mitochondrial and nuclear fragments, then sequenced by the Illumina MiSeq sequencer.
2) With older archaeological samples PLD-001, PLD-002, PLD-005, PLD-072, and PLD-073, new extractions using only the Powersoil kit were performed. At least two different DNA extracts represented each archaeological sample. A negative control was included in this set of analyses. These DNA extracts were subjected to targeted DNA amplifications of mitochondrial and nuclear fragments, as in the first set of extractions, and then sequenced on a MiSeq sequencer. The aDNA extracts from these older samples

were also used in DNA capture, as described below, and then subjected to a deep NGS sequencing using a HiSeq 3000 Illumina sequencer.

2.3.3 Extraction of DNA from modern accessions: DNA from modern plants were extracted according to Risterucci et al., 2000, to eliminate most of the polysaccharides from DNA extracts.

2.4 Analyses of targeted PCR amplified ancient DNA fragments

2.4.1 Identification of mitochondrial and nuclear DNA fragments suitable for aDNA analyses: Post-mortem DNA damage, in particular depurination and fragmentation, results in a high frequency of small-sized and degraded DNA fragments[5]. Consequently, primers were defined to amplify small-size fragments varying from 78 bp to 115 bp. 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 plant organs were not preserved, and the possible mix of ancient DNA 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, including SNPs specific to *T. cacao* and/or its wild relatives.

Several kinds of unique and repeated sequences were targeted for ancient DNA amplifications (Supplementary Table 4). We designed all primers using the primerBLAST online tool from NCBI (http://www.ncbi.nlm.nih.gov/tools/primer-blast).

All primer pairs were first tested on modern DNA from *T. cacao* and *T. bicolor*, and primers showing good efficiency for amplification (close to two after a qPCR test) and amplifying DNA fragments with Tm values differing between *T. cacao* and its wild relatives were preferentially chosen.

Mitochondrial sequences:
Six primer pairs were selected from mitochondrial gene sequences extracted from cacao transcriptomic data[6] (http://esttik.cirad.fr/):
- Cytochrome C Oxidase subunit 2 (CL294Contig1): MITO6 (position 1641-1740)
- Ribosomal protein S10 (CL1Contig700): MITO25 (position 540-623)
- NADH dehydrogenase subunit 1 (CL1 contig377):
    - MITO71 (position 803-880)
    - NADH80 (position 1228-1309)
- NADH dehydrogenase subunit 2 (CL755contig1):
    - MITO77 (position 221-321)
    - MITO79 (position 320-422)

Nuclear gene sequences:
Primer pairs were defined inside several unique nuclear gene sequences, important for cocoa flavor, and extracted from the cacao Criollo genome sequence[7]:

1. The vicilin gene (Tc04_g024090), is present in one copy in the *T. cacao* genome and encodes seed storage proteins containing aroma precursors. Based on the sequences of the vicilin gene, the phylogenetic relationships of *Theobroma* and *Herrania* species were studied[8]. The alignment of vicilin sequences from 11 *Theobroma* species was carried out with the ClustalW online tool from NCBI and three primer pairs (Vicilin1, 2, and 3) used to amplify sequences with SNPs between *T. cacao* and *T. bicolor* (another species known to have been cultivated in Mesoamerica) were defined.

2. FAT-B gene: cocoa butter (triacylglycerol storage lipids) accounts for 50% of dry cocoa seed weight. The acyl-ACP thioesterase fat B (*FATB*) genes, key enzymes leading to the synthesis of triacylglycerols, are particularly abundant in the *T. cacao* genome[7]. One primer pair (FATB-1) was defined in one unique FAT_B gene: Tc09_g010360.

3. Caffeine synthase gene: the caffeine synthase gene is an N_methyl transferase (3,7-dimethylxanthine methyl transferase) involved in the methylation of theobromine leading to caffeine synthesis. This gene was isolated in *Coffea*. An orthologous mRNA (**BCS1**) (AB096699.1 accession) was found in *T. cacao*, homologous to the Tc10_g001820 gene of the *T. cacao* Criollo genome. A primer pair (THEO2) was defined in this sequence and used to amplify ancient DNA.

4. Dihydroflavonol-4-reductase gene. *T. cacao* seeds contain large amounts of flavonoid polymers (proanthocyanidins) accounting for up to 8% of dry seeds. Among the 96 genes identified in proanthocyanidin synthesis, 18 correspond to DFR genes (dihydroflavonol-4-reductase)[7]. A primer pair (DFR3) was defined in an intronic region of one unique DFR gene (Tc09_g034440) in the *T. cacao* genome.

Definition of primer pairs for amplification of modern Theobroma and Herrania DNA fragments:

To be able to provide complete and high-quality sequences of the targeted DNA fragments mentioned above from modern DNA (using the Sanger sequencing method), new primers enabling the amplification of larger fragments, including the targeted small DNA fragments successfully amplified in aDNA, were defined (Supplementary Table 3) and used to amplify DNA fragments from a collection of genetic resources (Supplementary Table 4).

2.5 *PCR and Quantitative real-time PCR amplifications of targeted mitochondrial and gene fragments*

All aDNA extracts were subjected to both PCR and qPCR amplifications.

2.5.1 PCR: We performed PCR amplifications in 50 µl reactions containing, in an aqueous solution, 3 µl of DNA extract, 0.5 µM of each primer, 200 µM of each dNTP, 2 mM MgCl2, 10 µg Bovine Serum Albumin (Sigma) and either 2.5 U of Diamond Taq (Eurogentec) with Diamond Taq Reaction buffer (Eurogentec) or 2.5 U of AmpliTaq Gold® (Applied Biosystems™) with GeneAmp PCR Gold Buffer (Applied Biosystems™).

Cycling conditions were as follows: Taq polymerase activation at 95°C for 5 min, followed by 10 cycles of 94°C for 45 s, 58°C to 48°C annealing temperature for 45 s and 72°C for 1 min followed by 50 cycles with 48°C as annealing temperature, followed by 10 min at 72°C.

Following PCR, 5 µl of each amplification product was loaded onto a 2% agarose gel with a 1 Kb DNA ladder (Ozyme) used for size evaluation. The gel was run by electrophoresis, stained with ethidium bromide, and visualized under UV light.

2.5.2 Quantitative real-time PCR (qPCR): Quantitative real-time PCR was used as a very sensitive and powerful technique to follow the kinetics of product amplification and observe the melting temperature (Tm) of amplified fragments. Tm is defined as the temperature at which 50% of all molecules of a given DNA sequence are hybridized into double strands, and 50% are present as single strands. DNA sequences can be characterized by their Tm value and even a single point mutation can be revealed in a given sequence. In our experiments, we used this Tm parameter as a control to identify the amplification of *T. cacao* sequences or related sequences from the aDNA extracted from residues recovered from archaeological samples.

Quantitative PCR was performed in a 480 Roche LightCycler, with the following conditions: qPCR was carried out in 96-well plate, each well containing 10 µl of an aqueous solution of: 5 µl of LightCycler 480 SYBR green I master mix; 0.5 µM of each primer; 2 µg of Bovine Serum Albumin (Sigma); and 3 µl of aDNA extract.

The cycling conditions were 1 cycle of Taq activation and initial DNA denaturation at 95°C/5 min, followed by 60 three-segment cycles of amplification with a touchdown during the first 10 cycles: 95°C/20 sec, 58°C decreasing to 48°C/20 sec, 72°C/20 sec where the fluorescence was automatically measured at each of the primer annealing steps.

A melting curve analysis was carried out at the end of the 60 cycles by a denaturation-hybridization cycle (95°C/20 sec, 48°C/1 min). We used the Tm parameter as a control to screen the potential amplification of *T. cacao* sequences or related sequences from the aDNA extracted from ceramic residues and only qPCR amplifications showing the Tm specific to *Theobroma* were selected for further sequencing steps.

2.6 *Construction of libraries with pooled PCR and qPCR products and (NGS) high-throughput sequencing*

2.6.1 Constitution of PCR and qPCR product pools (bulk): After independent qPCR amplifications were carried out with the several markers on the aDNA samples, a selection of qPCR amplification products was made for each marker on the basis of Tm melting curves similar to those of modern *T. cacao* DNA fragments. A selection of amplification products resulting from direct PCR of aDNA was also made on the basis of DNA fragment sizes similar to modern *T. cacao* amplification product size.

All selected PCR and qPCR products were pooled for each archaeological sample. These pools were purified with the QIAquick® Gel Extraction Kit, adapted to keep fragments under 70 bp, according to the manufacturer's instructions, except for the initial step, where 5 volumes of QG buffer were added to the PCR/qPCR pool volume without any migration and excision of agarose gel. Pool DNA concentrations were estimated using an Infinite 200 PRO NanoQuant apparatus (Tecan) and controlled using a high-sensitivity chip (Agilent) for the presence of DNA fragments with sizes between 70 bp and 115 bp.

Three pools of negative controls (N1, N2, and N3) were carried out with the aDNA extractions, and amplifications were also selected for sequencing along with a negative control (N-zero), without any PCR product for library construction, making it possible to check the reliability of index sorting carried out by the Illumina MiSeq sequencing system.

2.6.2 NGS Library construction and sequencing: Four hundred ng of each PCR/qPCR pool was converted into a barcoded sequencing library using the NEXTflex™ Rapid DNA-Seq Kit from Bioo Scientific following the manufacturer's instructions but without DNA size selection or fragmentation steps and with 15 cycles of PCR amplification using the Diamond Taq (Eurogentec) polymerase.

Then, we quantified all libraries using high-sensitivity Agilent chips analysed with a Bio analyser 2100 (Agilent) and real time PCR carried out on a 480 LightCycler (Roche) with the KAPA Library Quantification Kit for Illumina Sequencing Platforms (Kapa Biosystems Ltd, SA). Libraries were then pooled in equimolar ratios before sequencing them simultaneously in a MiSeq Illumina run.

2.6.3 DNA Sequencing: Sequencing was carried out on an Illumina MiSeq sequencing system, located in the CIRAD/AGAP platform, per Illumina instructions, and using a 150-cycle reagent cartridge, enabling paired-end reads of 76 bp each to be carried out.

2.7 *Mapping aDNA sequences against the T. cacao reference sequences*

After NGS sequencing, all the reads were sorted according to their barcode sequences. To improve both the quality and length of Illumina reads, we first used FLASH (Fast Length Adjustment of Short reads)[9] to merge paired-end reads. Duplicate merged reads were then detected by the CD-HIT-DUP[10] tool with no mismatches allowed.

The unique merged reads were then mapped with Bowtie2[11] against the reference sequences of B97-61, a *T. cacao* Criollo genotype[7], with the "end-to-end" read alignment algorithm and the "very-sensitive" preset option. The reference sequences corresponded to PCR products amplified by each primer pair in the modern *T. cacao* B97-61 accession. Finally, sequences that aligned uniquely with the reference sequences were pre-selected for further analysis. The reads matching the target sequences were identified for each marker and each aDNA sample (Supplementary Table 12).

The results of sequencing and mapping of targeted aDNA fragments on the *T. cacao* reference sequences are reported in Supplementary Table 12, and the sequences identified as putative *Theobroma cacao* and/or wild relative species sequences after BLAST against the NCBI-EST database in Supplementary Table 7.

A small number of mapped and selected sequences were also observed in the four controls. One of them (N-zero) was a control linked to the NGS library construction and index sorting. The presence of a small number of *Theobroma* sequences in these controls could be explained by barcode errors, i.e. sequencing errors introduced into barcodes themselves, or low-level post-PCR cross-contamination by some of the most abundant aDNA PCR products.

2.8 *Pre-amplification of libraries submitted to DNA capture*

In order to reach the DNA concentrations required for DNA capture enrichment and HiSeq 3000 sequencing, each library was re-amplified after a size selection made on a 3% agarose gel allowing the elimination of most of the adaptor dimers formed during the first step of library construction: 10 µl of initial library were migrated on a 3% agarose gel, followed by a size selection (150 bp-400 bp). DNA was then purified from the agarose fragment using the

QIAquick® Gel Extraction Kit (QUIAGEN) and re-suspended in 20 µl of EB buffer. Five µl of purified DNA was PCR re-amplified during 2x18 cycles with the AmpliTaq Gold.

The libraries were then quantified by qPCR using the KAPA Library Quantification Kit.

The sequencing of an aliquot of these libraries was performed on an MiSeq Illumina sequencing system using a paired-end read length of 2x150 bp, to try to evaluate the signatures of ancient DNA, and particularly the C-->T and G-->A frequencies observed at the ends of aDNA fragments.

2.9 *Library trimming applied after DNA capture and sequencing*

Illumina paired-end reads were cleaned in a two-step procedure: sequencing adapters were first removed using AdapterRemoval V2.1.7[12] and cutadapt V1.8.3 with default parameters[13]; then polyA-tail was removed.

The paired-reads were mapped on the DNA 121 bp fragments containing the SNP[14] using bwa 0.7.12-r1039 with a minimum base quality of 30 (-q 30) and with the seed option disabled (-l 1000)[15] (Supplementary Table 8). Duplicates were removed using samtool 1.3.1 rmdup function. The resulting bam file was used to make a more stringent mapping on the *Theobroma cacao* whole genome sequence[14].

This mapping was performed with Bowtie2 using the "end to end" algorithm with "very-sensitive" preset option. Due to the potentiality of the presence of DNA damage and to increase sensitivity, one mismatch was allowed (-N 1) during the seed alignment and the function governing the minimum alignment score needed for an alignment to be considered as “valid” was relaxed (--score-min L,-1,-1). Reads that positively aligned the *Theobroma cacao* whole genome sequence were then compared to the NCBI nucleotide collection (NR). We used the BLASTn algorithm (E-value=10^{-5}) from the BLASTall 2.2.26 suite to identify the aDNA reads that successfully aligned with the *T. cacao* sequence containing SNPs, and to select sequences for which *T. cacao* appears as the first and lower hit, compared to other species.

2.10 *Search for ancient DNA signatures*

For all but two samples, only targeted amplifications of repeated mitochondrial sequences were successful, with no amplifications of the targeted nuclear sequences, which conforms to the scarcity of aDNA fragments.

A subset of ancient DNA samples was subjected to supplementary analyses aiming to verify their “ancient” nature. As expected, the recovered sequences exhibited typical signatures of post-mortem DNA damage found in ancient DNA, as follows:

2.10.1 Evaluation of DNA fragmentation due to post-mortem damages: Mostly small DNA fragments remain and can be sequenced as a result of post-mortem depurination[16,17], leading to DNA strand fragmentation. We used the MapDamage 2.0 software[18] to display the fragment size distribution as exemplified in Supplementary Figure 5.

The mean size of aDNA fragments evaluated on 12 samples is 18.4 bp. Some larger size fragments were also observed but with a low frequency, as shown in Supplementary Figure 5. The warm and humid environmental conditions of the SALF archaeological site, located in upper Amazon, probably explain the low mean size of aDNA fragments observed in this study.

2.10.2 Evaluation of misincorporations:

2.10.2.1 *Misincorporations observed at the ends of aDNA fragments*: An enhanced cytosine deamination observed inside aDNA fragments, but mostly in 5' overhanging single strand DNA ends, will translate itself into an excess of cytosine to thymine (C to T) misincorporations at 5' ends of aDNA sequences (and complementary guanine to adenine (G to A) at 3' ends) when PCR amplified. We used MapDamage 2.0 software[18] to evaluate the misincorporations at the ends of the fragments in comparison with flanking reference genome sequences. These misincorporations were observed in our samples as exemplified in Supplementary Figure 6.

2.10.2.2 *Misincorporations observed inside aDNA fragments*:

Misincorporations can also be observed inside the aDNA fragments. The Mito 6 fragment, for example, is well conserved among the various *Theobroma* species with the exception of one SNP that segregates only in *T. cacao*. Two other SNPs present in the Mito 6 fragment allowed differentiation of *Theobroma/Herrania* species from the other species. After amplification of the Mito 6 fragment in 18 archaeological samples and the selection of sequences exhibiting the SNP specific to *Theobroma* species and to *T. cacao*, we observed significant variations in the sequences from the archaeological samples in comparison with the Mito 6 *T. cacao* reference sequence.

We excluded from this analysis the part of the fragment corresponding to the primers and to the two specific SNP motifs used to screen the *Theobroma* sequences.

Of the 60,706 bases, across 1,578 reads with Phred quality scores ≥ 20,968 bases, (1.59%) were variants of the *T. cacao* Mito 6 sequence. Supplementary Figure 7 indicates the distribution of these variants.

We identified several kinds of miscodings (transitions and transversions), as classified by Hansen et al.[19] (Supplementary Table 13):

1. As expected, TS1 transitions, characteristic of Taq polymerase errors, were observed over all T/A positions of the sequence, whereas TS2 transitions, characteristic of aDNA degradation, only appeared with any frequency on one C/G nucleotide site (position 46). TV4 transversions were frequent on two C/G nucleotide sites (positions 54 and 58).
2. At five nucleotide positions, 25, 28, 46, 55, and 58, polymorphism frequency appeared to be significantly higher ($p<0.05$ for the observed values relative to a Gaussian distribution of the data). Supplementary Table 14 indicates the observed substitutions and their frequency for these five positions. TS2, the typical aDNA base substitution, was observed with a high frequency at position 46 and a lower frequency at position 55. Miscodings, less frequently cited in published aDNA studies that were observed here, are a TV4 transversion, CG->AT, at positions 55 and 58, which was also observed in ancient DNA by Briggs et al.[20], and a TV3 transversion (AT->CG) at positions 25 and 28—a miscoding which appears infrequently in aDNA publications.
3. The most frequent nucleotide combination at the five nucleotide sites with the highest degree of polymorphism corresponded to the *T. cacao* modern Mito 6 sequence. The variant combinations from *T. cacao,* resulting from BLAST searches in the NCBI EST, did not correspond to known sequences among *Theobroma* or other genera known to exist in the geographic region under study. Consequently, we assume that these sequence variations resulted from misincorporations arising from damage in the ancient *T. cacao* DNA strands or from Taq polymerase errors for TS1.

4. Across all of the misincorporation data analysed in the Mito 6 fragment, we observe a total sequence heterogeneity of around 1.6%. This corresponds to the upper range of sequence heterogeneity for ancient mitochondrial DNA as reported in the review by Binladen et al.[21].
5. The excess of TS2 transitions, observed in our study, is typical of aDNA, and the observation of a notable frequency for TV4 transversions is consistent with previously observed patterns in aDNA studies[20]. In contrast, the higher frequency for AT->CG (TV3) transversions that we observed is less frequently reported in other studies.

2.10.3 Impact of fragment size length on PCR amplification intensity:

We tested the impact of fragment size on PCR intensity. In the case of ancient DNA, PCR amplifications with primers amplifying fragments of an increased length will be effective with a decreased intensity depending on the length of the DNA fragments[22]

In our experiments, we designed primers in the mitochondrial Cytochrome C oxidase subunit 2 gene, amplifying DNA fragments differing in length but having a similar PCR efficiency tested on a modern DNA (from the B97-61 genotype) (Supplementary Table 15): Mito-197: 197 bp; Mito-290: 290 bp; Mito-543: 543 bp.

Ten ancient DNA samples were tested and subjected to qPCR amplifications with the same protocol as previously described above for ancient DNA qPCR amplifications.

The results of PCR amplifications of DNA fragments of different sizes show a decreased PCR amplification level when increasing the length of amplified DNA fragments. Eight out of the 10 ancient DNA samples were no longer amplified with the primer MITO-543, which amplifies a DNA fragment of 543 bp. The remaining two samples displayed a difference of 13 Ct and 11 Ct compared to fragments amplified by primer MITO-197.

The Ct (threshold cycle) results, which are a relative measurement of the concentration of the target in the PCR reaction, are reported for each qPCR amplification reaction in Supplementary Table 16 and visualised in Supplementary Figure 8.

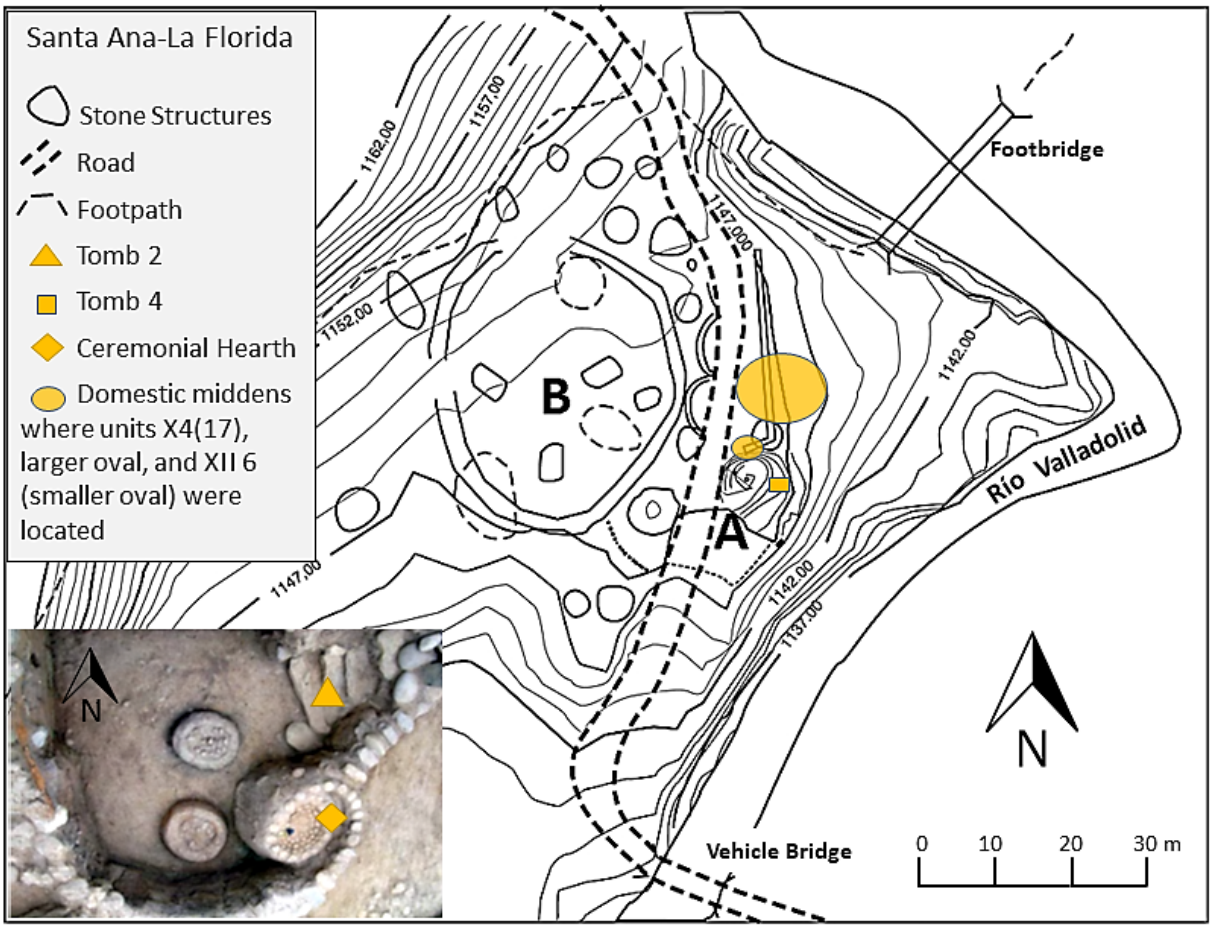
Our results support the hypothesis that the DNA extracted from archaeological samples from the Santa Ana-La Florida site is ancient and not modern. However, it is possible that both the small size of aDNA fragments and the various kinds of misincorporations that we observed could be linked to the adverse preservation conditions in the hot and humid tropics, where hydrolytic and oxidative damage to DNA is known to be intense and can induce greater DNA damage than would be expected in dry and cold environments. Until recently, most ancient DNAs analysed and estimations of the rates of DNA damage have only been carried out on specimens from these latter types of environments. Even so, a growing number of successful ancient DNA studies have been conducted on remains from humid and/or hot environments[23-25]. The existence of ancient DNA in these unfavourable environments may have been enabled by its absorption into mineral layers and organic particles[23]. This was certainly the case for the ancient DNA we found on ceramic residues conserved in the humid environmental conditions of Ecuadorian Amazonia.

3.0 Supplementary Information References

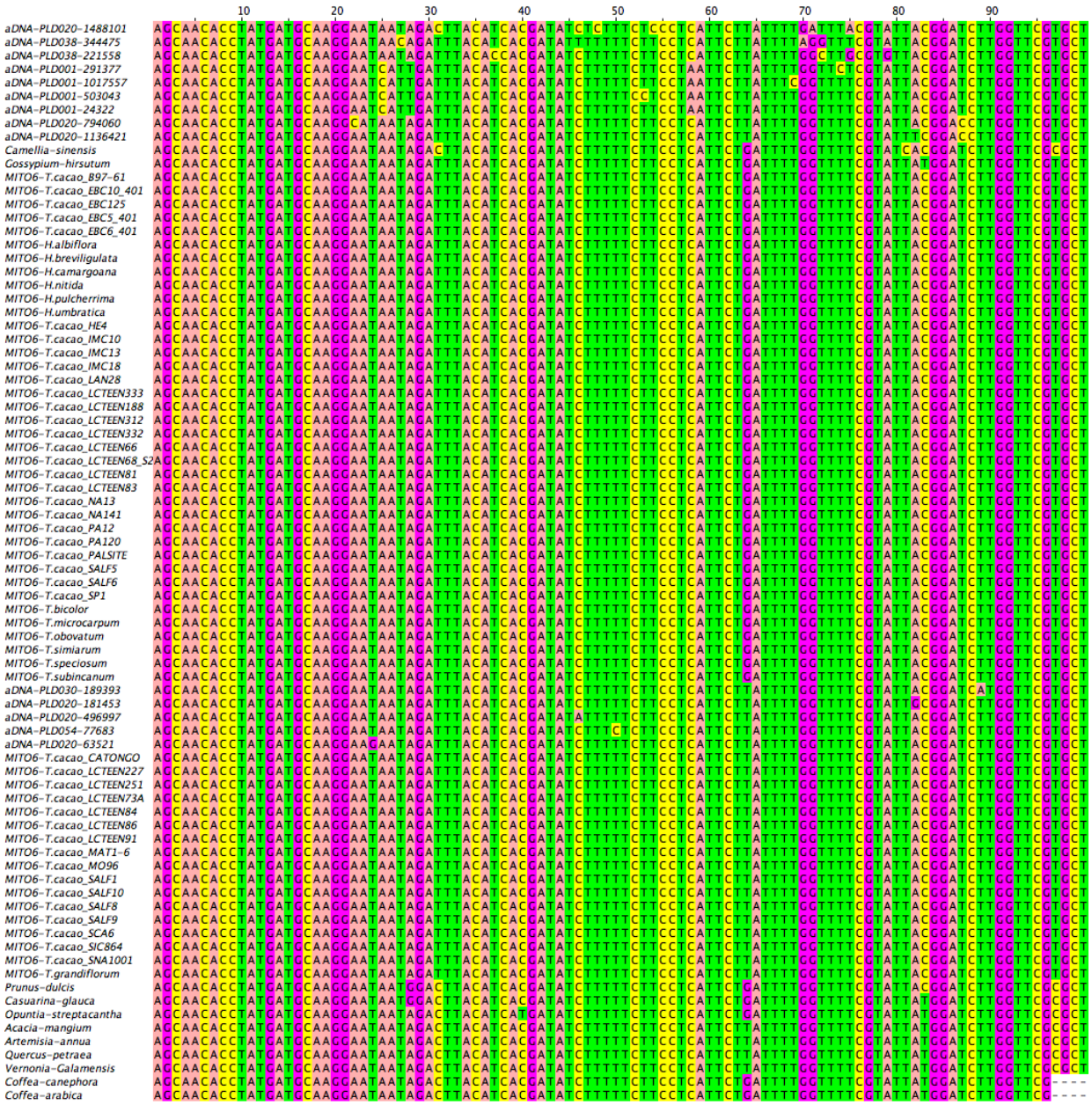
1. Valdez, F. Inter-zonal Relationships in Ecuador. in *The Handbook of South American Archaeology* 865–888 (ed. Silverman, H. & Isbell, W. H.) (Springer, New York, NY, 2008).
2. Valdez, F., Guffroy, J., de Saulieu, G., Hurtado, J. & Yepes, A. Découverte d'un site cérémoniel formatif sur le versant oriental des Andes. *C. R. Palevol.* 4, 369–374 (2005).
3. Zarrillo, S. *Human Adaptation, Food Production, and Cultural Interaction during the Formative Period in Highland Ecuador*. (University of Calgary, 2012). http://dx.doi.org/10.5072/PRISM/10082
4. Cieslak, M. et al. Origin and History of Mitochondrial DNA Lineages in Domestic Horses. *PLOS ONE* 5, e15311 (2010). doi:10.1371/journal.pone.0015311
5. Pääbo, S. et al. Genetic analyses from ancient DNA. *Annu Rev Genet* 38, 645–679 (2004). doi: 10.1146/annurev.genet.37.110801.143214.
6. 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). https://doi.org/10.1186/1471-2164-8-81.
7. Argout, X. et al. The genome of *Theobroma cacao*. *Nature Genetics* 43, 101–108 (2011). doi:10.1038/ng.736 pmid:21186351.
8. B. A. Whitlock, D. A. Baum, Phylogenetic relationships of Theobroma and Herrania (Sterculiaceae) based on sequences of the nuclear gene Vicilin. *Syst. Bot.* **24**, 128 (1999).
9. Magoč, T., Salzberg, S. L. FLASH: Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27, 2957 (2011). doi: 10.1093/bioinformatics/btr507
10. Fu, L., Niu, B., Zhu, Z., Wu, S., Li, W. CD-HIT: accelerated for clustering the next generation sequencing data. *Bioinformatics* 28, 3150 (2012). doi:10.1093/bioinformatics/bts565
11. Langmead, B., Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9, 357 (2012). https://doi.org/10.1038/nmeth.1923.
12. Schubert, M., Lindgreen, S. and Orlando, L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res Notes*, 9, 88. (2016). doi:10.1186/s13104-016-1900-2.
13. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17(1):10-12, (2011). DOI: http://dx.doi.org/10.14806/ej.17.1.200
14. Argout, X. et al. The cacao Criollo genome v2.0: an improved version of the genome for genetic and functional genomic studies. *BMC Genomics* 18, 730 (2017). doi.org/10.1186/s12864-017-4120-9.
15. Li, H., Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* 26, 589 (2010). doi:10.1093/bioinformatics/btp698
16. Pääbo, S., Wilson, A. C. Miocene DNA sequences - a dream come true? *Current Biology* 1, 45 (1991). DOI: https://doi.org/10.1016/0960-9822(91)90125-G
17. Lindahl, T. Instability and decay of the primary structure of DNA. *Nature* 362, 709 (1993). doi:10.1038/362709a0
18. Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F. & Orlando, L. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* 29, 1682–1684 (2013). doi: 10.1093/bioinformatics/btt193.

19. 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). https://doi.org/10.1093/oxfordjournals.molbev.a003800
20. Briggs, A. W., Stenzel, U., Johnson, P. L. F., Green, R. E., Kelso, J., Prüfer, K., Meyer, M., Krause, J., Ronan, M. T., Lachmann, M., Pääbo, S. Patterns of damage in genomic DNA sequences from a Neandertal. *Proc Natl Acad Sci USA* 104, 14616-14621 (2007). DOI: 10.1073/pnas.0704665104
21. Binladen, J., Wiuf, C., Gilbert, M. T. P., Bunce, M., Barnett, R., Larson, G., Greenwood, A. D., Haile, J., Ho, S. Y. W., Hansen A. J., Willerslev, E. Assessing the Fidelity of Ancient DNA Sequences Amplified from Nuclear Genes. *Genetics* 172, 733 (2006). https://doi.org/10.1534/genetics.105.049718
22. Cooper, A. & Poinar H. N. Ancient DNA: do it right or not at all. *Science* 289, 1139 (2000).
23. Coolen, M. J. L., Overmann, J. 217 000-year-old DNA sequences of green sulphur bacteria in Mediterranean sapropels and their implications for the reconstruction of the paleoenvironment. *Environ. Microbiol.* 9, 238 (2007). DOI:10.1111/j.1462-2920.2006.01134.x
24. Hansson, M. C., Foley, B. P. Ancient DNA fragments inside Classical Greek amphoras reveal cargo of 2400-year-old shipwreck. *J. Arch. Sci*. 35, 1169 (2008). https://doi.org/10.1016/j.jas.2007.08.009
25. Anderson-Carpenter, L. L., McLachlan, J. S., Jackson, S. T., Kuch, M., Lumibao, C. Y., Poinar, H. N., Ancient DNA from lake sediments: Bridging the gap between paleoecology and genetics. *BMC Evol. Biol.* 11, 30 (2011). https://doi.org/10.1186/1471-2148-11-30

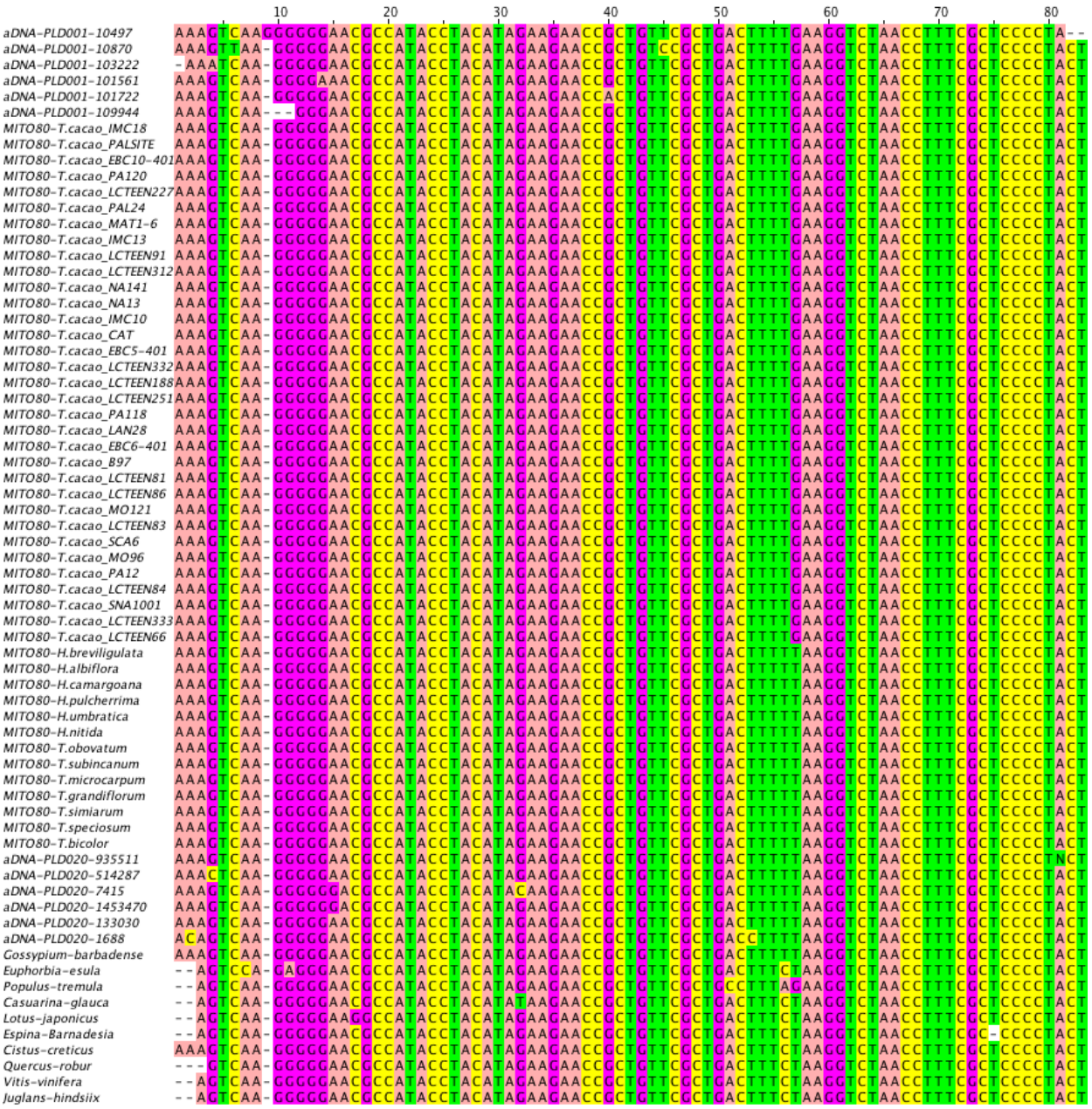
4.0 Supplementary Figures



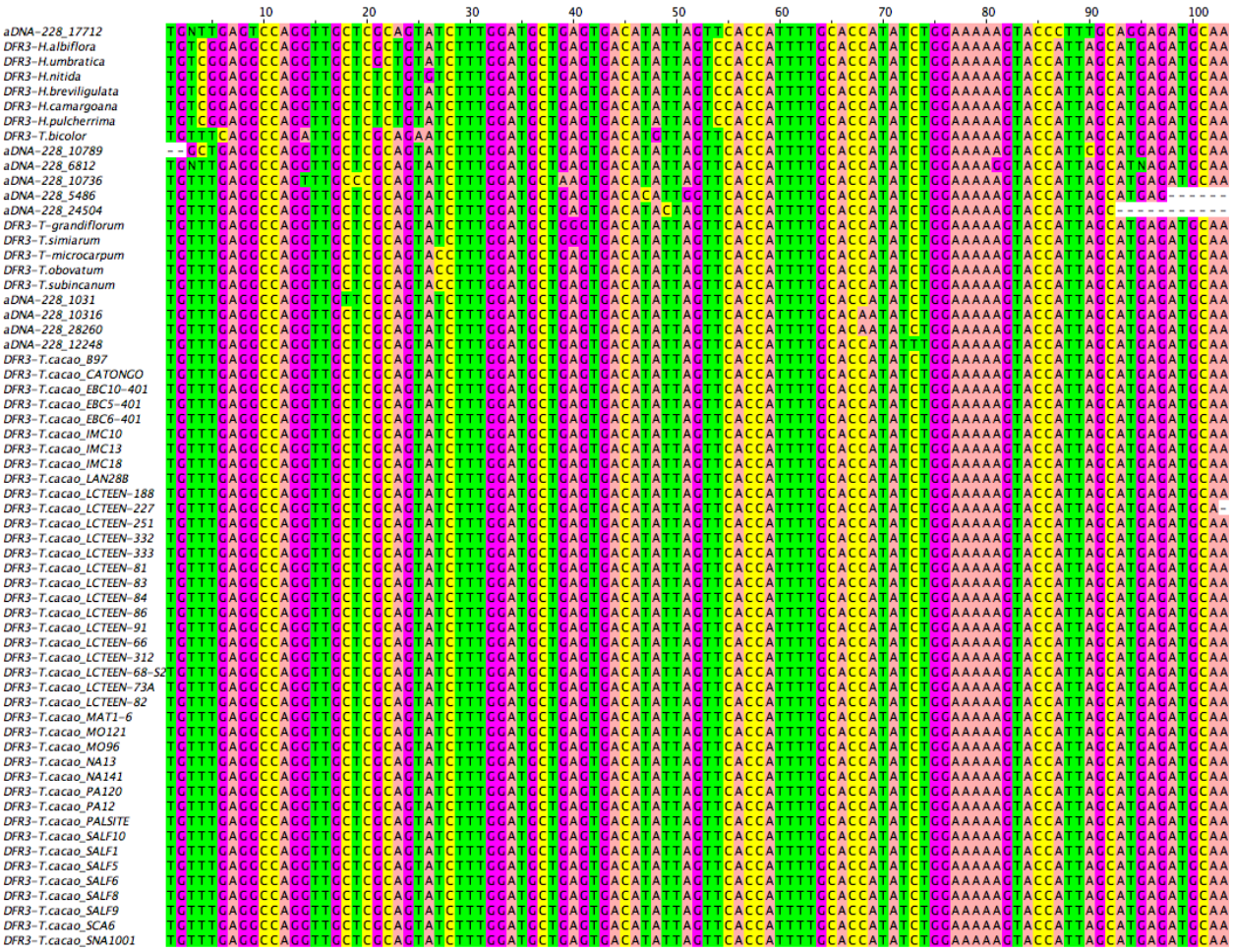
Supplementary Figure 1 | Map of the SALF site showing major stone constructions, important features, and topography[1,3]. Location A shows where the terrace was modified with retaining walls to form an oval platform. The highest portion, just above "A" may have functioned as a temple, particularly as tombs with human burials were located here. Note the spiral orientation of the retaining walls and how they converge in the centre where a ceremonial hearth is located, as also shown in the inset picture. Location B is the centre of the circular sunken plaza.



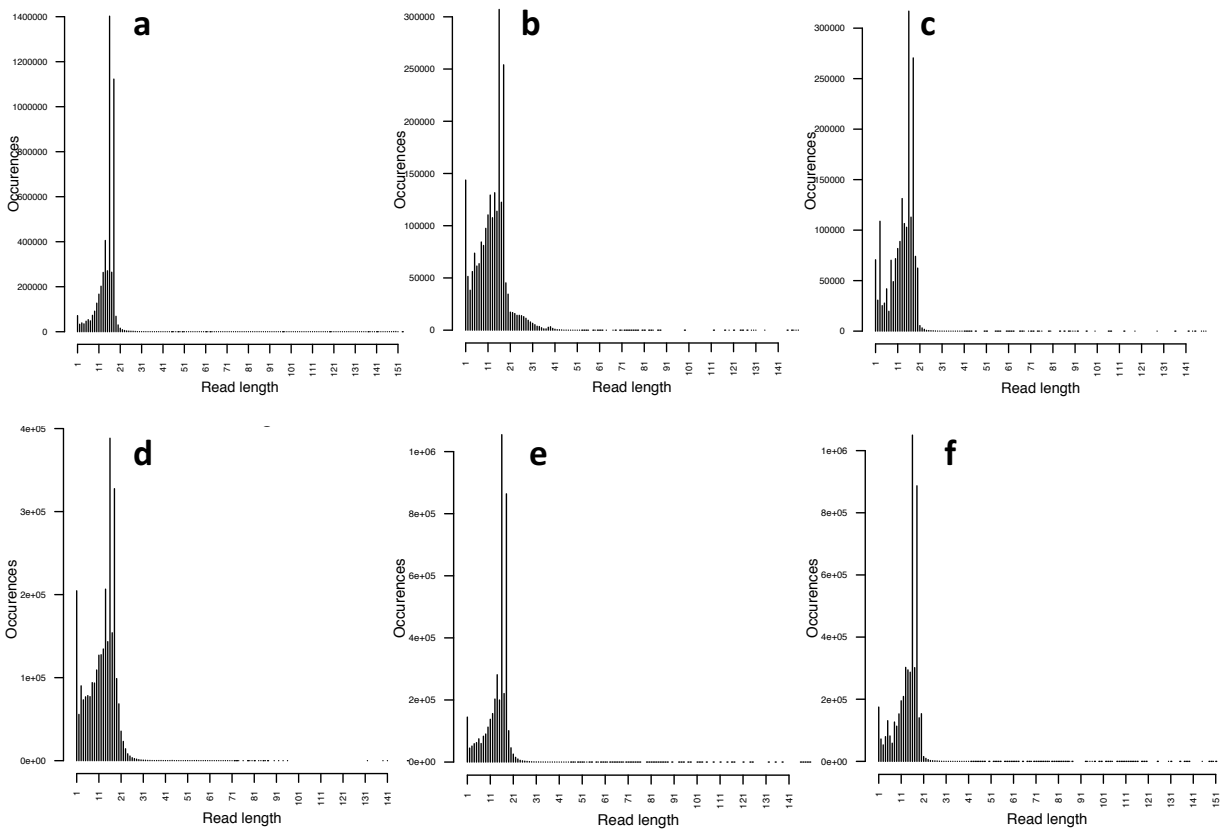
Supplementary Figure 2 | Alignment of MITO 6 sequences from modern accesions and some aDNA sequences.



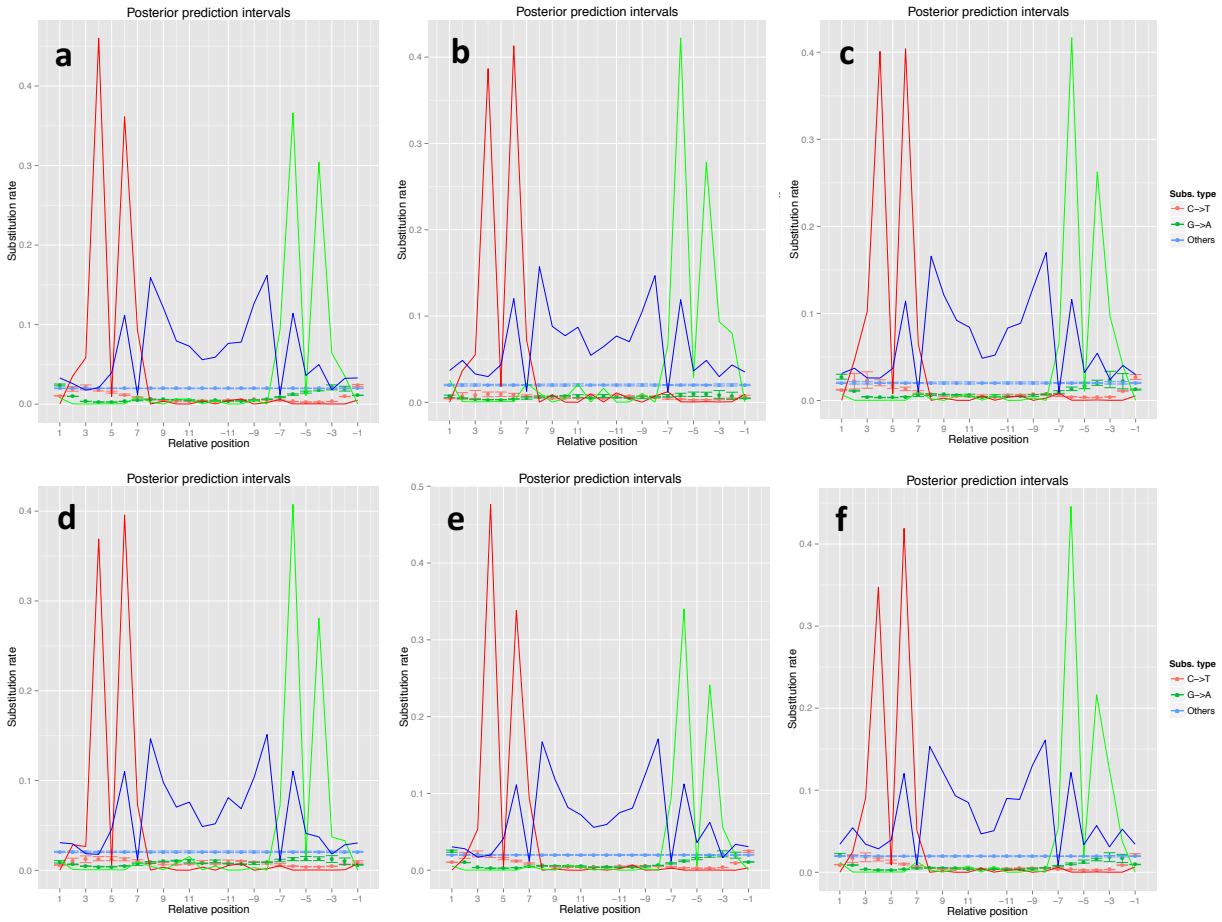
Supplementary Figure 3 | Alignment of MITO 80 sequences from modern accessions and some aDNA sequences.



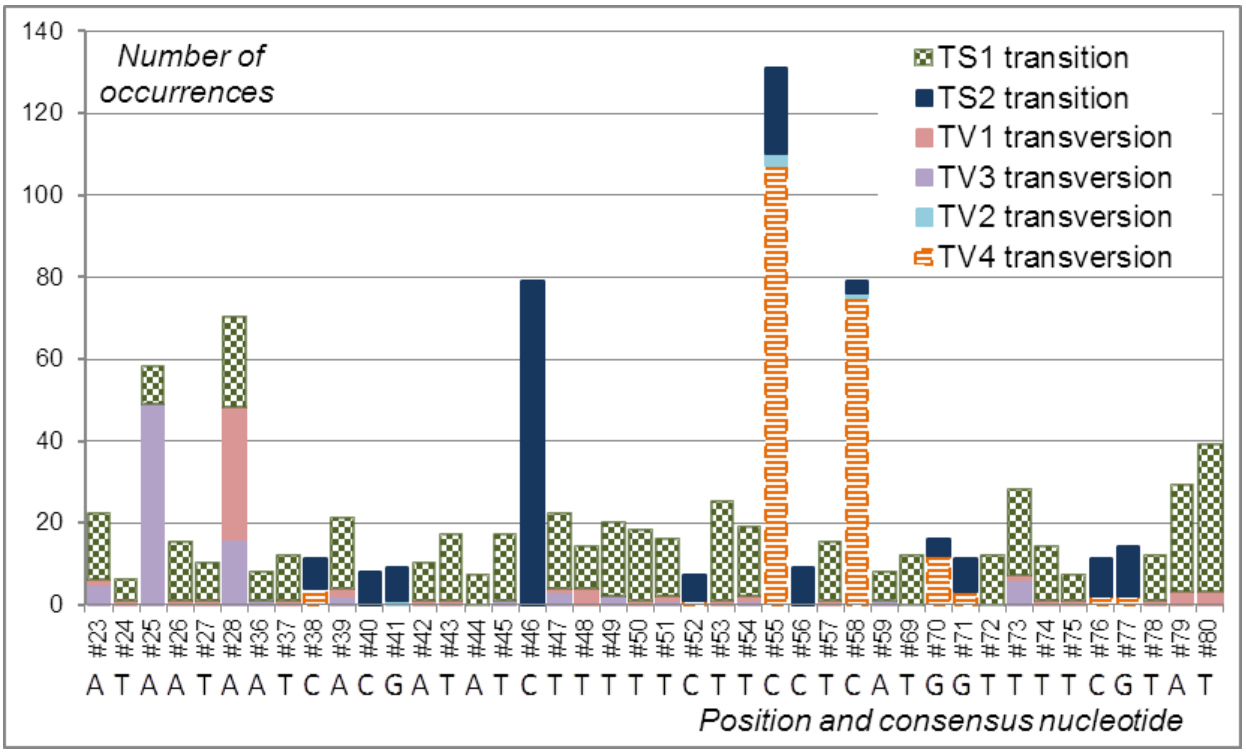
Supplementary Figure 4 | Alignment of DFR3 sequences from modern accessions and some aDNA sequences.



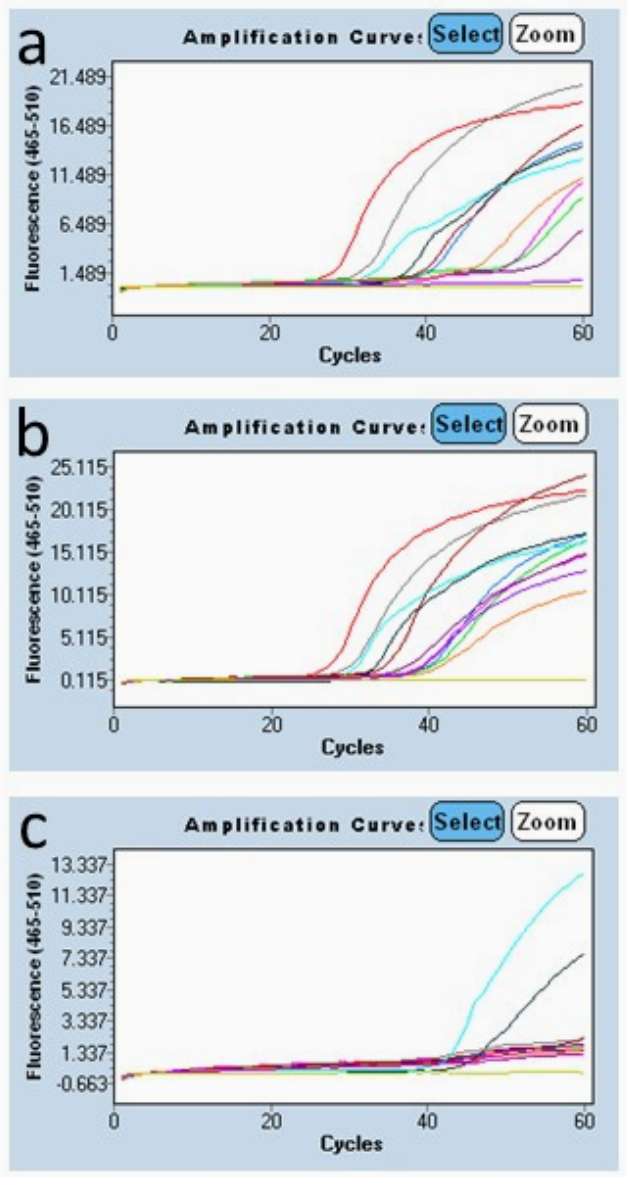
Supplementary Figure 5 | Single-end read length distributions for some archaeological samples. Histograms for 6 different aDNA extracts from 4 archaeological samples are provided by Mapdamage V2.0 after sequencing subsets of aDNA libraries before DNA capture.
The samples PLDXXX/(aDNA extract) are as follows: a: PLD_001/(324 extract); b: PLD_005/(307 extract); c: PLD_005/(319 extract); d: PLD_072/(313 extract); e: PLD_073/(301 extract); f: PLD_073/(302 extract).



Supplementary Figure 6 | DNA damage patterns for some ancient SALF archaeological samples revealed by MapDamage V2.0 (MCMC histograms). Example of cytosine deamination damage patterns observed at 3-prime ends (G->A) (in green) and 5-prime ends (C->T) (in red) of nuclear aDNA fragments selected after mapping the reads from hiseq 3000 sequencing on the cacao genome. Frequency of C->T and G->A substitutions are reported according to the base position along the aDNA fragments. Sequences analysis of six different aDNA extracts from 4 archaeological samples (PLD_XXX) are reported in this figure: a: PLD_001/(324 extract); b: PLD_005/(307 extract); c: PLD_005/(319 extract); d: PLD_072/(313 extract); e: PLD_073/(301 extract); f: PLD_073/(302 extract).



Supplementary Figure 7 | Occurrences of the different types of nucleotide substitutions over the *T. cacao* Mito6 reads. Five nucleotide sites showed significantly higher polymorphisms according to Peirce's criterion. Out of these five sites, substitutions of the TS1 type typical of Taq polymerase errors were very dominant. (TS1: AT ->GC; TS2: CG >TA; TV1: AT ->TA; TV2: CG ->GC; TV3: AT ->CG; TV4: CG ->AT)



Supplementary Figure 8 | qPCR amplification curves of three DNA fragments of different sizes from ten aDNA samples. The qPCR amplifications were carried out on a Lightcycler 480 apparatus (Roche) with 3 primer pairs amplifying DNA fragments differing in size: (a) 197 bp; (b) 290 bp; and (c) 543 bp. The qPCR amplifications were carried out using ten aDNA samples, with the DNA extract number and colour of the amplification curve indicated in brackets: PLD-054 (283-orange); PLD-001 (79-red); PLD-020 (258-light blue); PLD-025 (183-purple); PLD-025 (212-brown); PLD-038 (61-dark blue); PLD-042 (113-green); PLD-042 (265-black); PLD-010 (122-pink); and PLD-050 (168-grey).

5.0 Supplementary Tables

Supplementary Table 3 | List of archaeological samples analysed for aDNA and the corresponding number of extractions made for each one. Two types of aDNA analyses were carried out: 1. targeted PCR and qPCR made directly on aDNA extracts; or 2. PCR amplifications following DNA capture of genomic DNA fragments containing SNPs. Archaeological samples reported in Table 1 (positive for *T. cacao* aDNA and theobromine or starch analyses) are in bold.

Archaeological sample	object	2δ Cal yr (BP)	Phase	extraction n°	Extraction method	Methodology used for aDNA analyses				Archaeological sample	object	2δ Cal yr (BP)	Phase	extraction n°	Extraction method	Methodology used for aDNA analyses			
						Targeted PCR and qPCR	Miseq sequencing	amplifications after DNA capture	Hiseq 3000 sequencing							Targeted PCR and qPCR	Miseq sequencing	amplifications after DNA capture	Hiseq 3000 sequencing
PLD-001	**human effigy stirrup spout bottle**	**4146 to 3932**	**Palanda**	**52**	**A**	**X**	**X**			PLD-045	ceramic sherd/burnt earth	4095 to 4060	Palanda	103	A	X	X		
PLD-001				**66**	**B**	**X**	**X**			PLD-045				104	A	X	X		
PLD-001				**79**	**B**	**X**	**X**			PLD-046	ceramic sherd seed probable	3980 to 3835	Palanda	116	A	X	X		
PLD-001				**141**	**A**	**X**	**X**			PLD-046				266	B	X	X		
PLD-001				**171**	**B**	**X**	**X**			PLD-047	stone mortar cacao pod shaped		Palanda	135	A	X	X		
PLD-001				**255**	**B**	**X**	**X**			PLD-047				165	B	X	X		
PLD-001				**285**	**A**	**X**	**X**			PLD-048	stone mortar bird shaped		Palanda	136	A	X	X		
PLD-001				**323**	**B**	**X**	**X**	**X**		PLD-048				166	B	X	X		
PLD-001				**324**	**B**			**X**		PLD-050	stone mortar anthropomorphic shaped		Palanda	138	A	X	X		
PLD-002	**donut shaped stirrup spout bottle**	**4090 to 3980**	**Palanda**	**63**	**A**	**X**	**X**			PLD-050				168	B	X	X		
PLD-002				**77**	**B**	**X**	**X**			PLD-051	stone mortar spoon shaped		Palanda	139	A	X	X		
PLD-002				**170**	**B**	**X**	**X**			PLD-051				169	B	X	X		
PLD-002				**254**	**B**	**X**	**X**			PLD-052	ceramic sherd - channelled		Palanda	251	B	X	X		
PLD-002				**284**	**A**	**X**	**X**			PLD-052				281	A	X	X		
PLD-002				**321**	**B**	**X**		**X**		PLD-053	ceramic sherd punctuate		Palanda	252	B	X	X		
PLD-002				**322**	**B**	**X**	**X**	**X**	**no DNA**	PLD-053				282	A	X	X		
PLD-003	**incised-punctuated stone bowl**	**4150 to 3920**	**Palanda**	**81**	**B**	**X**	**X**			PLD-054	stirrup handle ceramic fragment		Palanda	74	B	X	X		
PLD-003				**236**	**B**	**X**	**X**			PLD-054				172	B	X	X		
PLD-003				**243**	**B**	**X**	**X**			PLD-054				253	B	X	X		
PLD-003				**257**	**B**	**X**	**X**			PLD-054				283	A	X	X		
PLD-003				**273**	**A**	**X**	**X**			**PLD-072**	**ceramic residue**	**5440 to 5310**	**Palanda**	**311**	**B**	**X**	**X**	**X**	**X**
PLD-003				**287**	**A**	**X**	**X**			**PLD-072**				**312**	**B**	**X**	**X**	**X**	**X**
PLD-003				**323**	**B**	**X**	**X**			**PLD-072**				**313**	**B**	**X**	**X**	**X**	**no DNA**
PLD-004	red jasper incised stone bowl	4090 to 3980	Palanda	272	A	X	X			**PLD-072**				**314**	**B**	**X**		**X**	**no DNA**
PLD-005	**charred ceramic residue**	**5440 to 5310**	**Palanda**	**306**	**B**	**X**	**X**	**X**	**X**	**PLD-073**	**ceramic residue**	**5440 to 5310**	**Palanda**	**301**	**B**	**X**	**X**	**X**	**X**
PLD-005				**307**	**B**	**X**	**X**	**X**	**X**	**PLD-073**				**302**	**B**	**X**	**X**	**X**	**X**
PLD-005				**308**	**B**	**X**	**X**	**X**	**no DNA**	**PLD-073**				**303**	**B**	**X**	**X**	**X**	**no DNA**
PLD-005				**309**	**B**	**X**	**X**	**X**	**no DNA**	**PLD-073**				**304**	**B**	**X**		**X**	**no DNA**
PLD-005				**310**	**B**	**X**	**X**	**X**	**no DNA**	**PLD-073**				**305**	**B**	**X**	**X**	**X**	**no DNA**
PLD-005				**319**	**B**	**X**	**X**	**X**	**X**	PLD-079	stirrup gourd - ceramic	4146 to 3932	Palanda	75	B	X	X		
PLD-006	squash stirrup spout bottle		Palanda	78	B	X	X			PLD-079				202	A	X	X		
PLD-006				234	B	X	X			PLD-079				232	B	X	X		
PLD-006				240	B	X	X			PLD-079				237	B	X	X		
PLD-007	stone bowl		Palanda	118	A	X	X			PLD-079				267	A	X	X		
PLD-008	stone mortar bird figure	4150 to 3920	Palanda	119	A	X	X			PLD-081	stirrup ceramic flask	4146 to 3932	Palanda	203	A	X	X		
PLD-008				149	B	X	X			PLD-081				233	B	X	X		
PLD-009	**stone mortar bird figure**		**Palanda**	**121**	**A**	**X**	**X**			PLD-088	constricted neck pot	4090 to 3980	Palanda	80	B	X	X		
PLD-009				**151**	**B**	**X**	**X**			PLD-088				158	B	X	X		
PLD-010	**stone mortar human figure**		**Palanda**	**122**	**A**	**X**	**X**			PLD-088				235	B	X	X		
PLD-010				**152**	**B**	**X**	**X**			PLD-088				241	B	X	X		
PLD-011	stone bowl fragment	4150 to 3920	Palanda	129	A	X	X			PLD-088				249	B	X	X		
PLD-020	**charred ceramic residue**	**3980 to 3835**	**Palanda**	**91**	**A**	**X**	**X**			PLD-088				271	A	X	X		
PLD-020				**258**	**B**	**X**	**X**			PLD-088				279	A	X	X		
PLD-020				**288**	**A**	**X**	**X**			PLD-097	stone dipper	4146 to 3932	Palanda	117	A	X	X		
PLD-025	**charred ceramic residue**	**3980 to 3835**	**Palanda**	**96**	**A**	**X**	**X**			PLD-097				147	B	X	X		
PLD-025				**97**	**A**	**X**	**X**			PLD-097				256	B	X	X		
PLD-025				**183**	**A**	**X**	**X**			PLD-097				274	A	X	X		
PLD-025				**212**	**B**	**X**	**X**			PLD-097				286	A	X	X		
PLD-025				**259**	**B**	**X**	**X**			PLD-104	rim carinated ceramic bowl	4090 to 3980	Palanda	245	B	X	X		
PLD-025				**289**	**A**	**X**	**X**			PLD-104				275	A	X	X		
PLD-026	**charred ceramic residue**	**3980 to 3835**	**Palanda**	**58**	**A**	**X**	**X**			**PLD-112**	**straight collared jar rim**		**Palanda**	**125**	**A**	**X**	**X**		
PLD-026				**82**	**B**	**X**	**X**			**PLD-112**				**155**	**B**	**X**	**X**		
PLD-026				**260**	**B**	**X**	**X**			**PLD-112**				**246**	**B**	**X**	**X**		
PLD-026				**290**	**A**	**X**	**X**			**PLD-112**				**276**	**A**	**X**	**X**		
PLD-028	charred ceramic residue	3980 to 3835	Palanda	99	A	X	X			PLD-116	carinated body fragment punctuated decoration		Palanda	250	B	X	X		
PLD-028				216	B	X	X			PLD-116				280	A	X	X		
PLD-029	ceramic sherd	3980 to 3835	Palanda	188	A	X	X			PLD-123	ceramic body fragment	2348 to 2185	Tacana	127	A	X	X		
PLD-029				218	B	X	X			PLD-123				157	B	X	X		
PLD-030	**charred ceramic residue**	**3980 to 3835**	**Palanda**	**59**	**A**	**X**	**X**			PLD-123				248	B	X	X		
PLD-030				**71**	**B**	**X**	**X**			PLD-123				278	A	X	X		
PLD-030				**261**	**B**	**X**	**X**			PLD-130	bowl base sherd slipped	2340 to 2120	Tacana	123	A	X	X		
PLD-030				**291**	**A**	**X**	**X**			PLD-130				153	B	X	X		
PLD-037	seed charcoal	3980 to 3835	Palanda	109	A	X	X			**PLD-131**	**bowl sherd incised painted**	**2340 to 2120**	**Tacana**	**126**	**A**	**X**	**X**		
PLD-038	**charred ceramic residue**	**3980 to 3835**	**Palanda**	**61**	**A**	**X**	**X**			**PLD-131**				**156**	**B**	**X**	**X**		
PLD-038				**73**	**B**	**X**	**X**			**PLD-131**				**247**	**B**	**X**	**X**		
PLD-038				**263**	**B**	**X**	**X**			**PLD-131**				**277**	**A**	**X**	**X**		
PLD-038				**293**	**A**	**X**	**X**			control				N-zero		X	X		
PLD-040	**charred ceramic residue**	**3980 to 3835**	**Palanda**	**112**	**A**	**X**	**X**			control				N1	A	X	X		
PLD-040				**197**	**A**	**X**	**X**			control				N2	B	X	X		
PLD-040				**227**	**B**	**X**	**X**			control				N3	A	X	X		
PLD-040				**264**	**B**	**X**	**X**			control				315	B			X	no DNA
PLD-040				**294**	**A**	**X**	**X**			control				316	B			X	X
PLD-042	**charred ceramic residue**	**3980 to 3835**	**Palanda**	**113**	**A**	**X**	**X**			control				317	B			X	no DNA
PLD-042				**228**	**B**	**X**	**X**			control				318	B			X	no DNA
PLD-042				**265**	**B**	**X**	**X**												

Supplementary Table 4 | Primer pairs defined in mitochondrial and nuclear genes used to amplify ancient DNA and modern accessions.

primer name	gene/sequence name	sequence name	sequence reference	5'---> 3' forward primer	Tm forw.	reverse primer	Tm rev.	product length (bp)	positions of amplified fragment in the sequence	
									start	end
for aDNA amplifications										
MITO 6	Cytochrome C oxydase subunit 2	ESTTik/CL294Contig1	Argout et al. (2008)	AGCAACACCTATGATGCAAGGA	60.03	AGCACGAACCAAGATCCGTA	59.11	**100**	1641	1740
MITO 25	Ribosomal protein S10	ESTTik/CL1Contig700	Argout et al. (2008)	GAGCGGGGTCGAGGTAAG	59.19	CCAGTTGGTTGAGGAGAAGTAAG	58.68	**84**	540	623
MITO 71	NADH dehydrogenase subunit 1	ESTTik/CL1 contig377	Argout et al. (2008)	GGCCCTGGTCGGTCA	57.45	GCTCTTGATCTGAGGTGCGA	59.83	**78**	803	880
MITO 77	NADH dehydrogenase subunit 2	ESTTik/CL755Contig1	Argout et al. (2008)	CTGAGAACAAAGGAGAGGGAT	56.03	GATTCTATATGAACCAATGGATCGT	56.07	**101**	221	321
MITO 79	NADH dehydrogenase subunit 2	ESTTik/CL755Contig1	Argout et al. (2008)	TCCATGTCCTAGGTGTATCA	57.58	GCTGTTAGTTAGTACCCCCTT	56.72	**103**	320	422
MITO 80	NADH dehydrogenase subunit 1	ESTTik/CL1 contig377	Argout et al. (2008)	AAAGTCAAGGGGGAACGCCA	62.00	AGTAGGGGAGCGAAAGGTTAGA	60.29	**82**	1228	1309
Vicilin1	vicilin	Tc04_g024090	Argout et al. (2011)	CCAAGACAACCAAGAGAAGCTAA	58.61	AAGTAGAAAACCTCATATTTGCCAG	57.60	**81**	877	957
Vicilin2	vicilin	Tc04_g024090	Argout et al. (2011)	GTCTTGGTTAGTCTATTGACCCTGT	60.05	CGGTCTCAAGAACTTCATAGCTG	59.14	**115**	996	1110
Vicilin3	vicilin	Tc04_g024090	Argout et al. (2011)	GAGCAATAAGCCAACAAGCTAC	57.98	GGAGTAGACAGGCGATTGG	57.33	**89**	1322	1410
THEO2	caffeine synthase	Tc10_g001820	Argout et al. (2011)	AGTGAACAAGGAGAGGGTTCTTTC	58.18	TCCGGTTCCGTCCATATACCT	60.13	**81**	1524	1604
DFR3	Dihydroflavonol-4-reductase	Tc09_g034440	Argout et al. (2011)	TGTTTGAGGCCAGGTTGCT	59.77	TTGCATCTCATGCTAATGGTACTT	58.56	**103**	1295	1397
FATB-1	acyl-ACP thioesterase fat B	Tc09_g010360	Argout et al. (2011)	AGAAACAGATCAGGCCACAAGA	59.63	TGGCTAGGAAGATATAGAAACAGG	57.33	**77**	2623	2699
for modern DNA amplifications										
NADH-sub1	NADH dehydrogenase subunit 1	ESTTik/CL1 contig377	Argout et al. (2008)	GGGCACAAAGCTGTTGAATG	58.21	TTTCATTAAGTCAAGTAGGCGAACC	59.59	**995**		
NADH-sub2	NADH dehydrogenase subunit 2	ESTTik/CL755Contig1	Argout et al. (2008)	GGCAAACGCCTGAATAATCCT	58.98	CTTCAATCCTTATTCCTTATCCCCA	58.15	**755**		
MITO6-P2	Cytochrome C oxydase subunit 2	ESTTik/CL294Contig1	Argout et al. (2008)	TTTCCATTCTCCCATTGATTCCT	57.79	CAACCACTCTATTGTCCACTTCT	58.10	**543**		
MITO25-P2	Ribosomal protein S10	ESTTik/CL1Contig700	Argout et al. (2008)	TCCGCCACCAATCAAACCA	59.85	ATCCTTGCGTTGACCTTTTCC	59.11	**575**		
MITO35-P2	NADH dehydrogenase subunit 1	ESTTik/CL1 contig377	Argout et al. (2008)	ATGGGTTCGCCTACTTGACT	58.72	GCATAGGGGACCATCCTTTTG	58.69	**516**		
DFR3-P	Dihydroflavonol-4-reductase	Tc09_g034440	Argout et al. (2011)	AAGGTTTGGATGTGGTGGTG	58,3	CCTTCCGGTTGCTGATGAGTT	60.61	**862**		
VICI2-P	vicilin	Tc04_g024090	Argout et al. (2011)	CCACTCAAGGGCATCAACGA	60.32	TGGTTGGAGTAGACAGGCGA	60.54	**844**		

Supplementary Table 5 | Origin of modern accessions used to compare targeted PCR amplified aDNA sequences. The targeted aDNA sequences were compared by BLAST with a collection of modern DNA sequences established, for each marker, from accessions covering the diversity of the *T. cacao* species (44 accessions), including representatives of 7 wild *Theobroma* species and 6 wild *Herrania* species.

Species	Accession	Country of origin	Origin	Type
T. cacao	B 97-61	Belize	Maya Mountains	old domesticated
T. cacao	CATONGO	Brazil	Itabuna	old domesticated
T. cacao	EBC1 0_401	Colombia	Caqueta River	native
T. cacao	EBC1 25_59	Colombia	Caqueta River	native
T. cacao	EBC 5_401	Colombia	Caqueta River	native
T. cacao	EBC 6_401	Colombia	Caqueta River	native
T. cacao	HE 4	Mexico	Tachira	old domesticated
T. cacao	IMC 10	Peru	Iquitos	native
T. cacao	IMC 13	Peru	Iquitos	native
T. cacao	IMC 18	Peru	Iquitos	native
T. cacao	Lan 28	Mexico	Lacandona	old domesticated
T. cacao	LCTEEN 188	Ecuador	Putumayo River	native
T. cacao	LCTEEN 227	Ecuador	Villano River	native
T. cacao	LCTEEN 251	Ecuador	Curaray River	native
T. cacao	LCTEEN 312	Ecuador	Upano River	native
T. cacao	LCTEEN 332	Ecuador	Upano River	native
T. cacao	LCTEEN 333	Ecuador	Upano River	native
T. cacao	LCTEEN 66	Ecuador	Zamora River	native
T. cacao	LCTEEN 68_S2	Ecuador	Zamora River	native
T. cacao	LCTEEN 73A	Ecuador	Nangaritza River	native
T. cacao	LCTEEN 81	Ecuador	Zamora River	native
T. cacao	LCTEEN 82	Ecuador	Zamora River	native
T. cacao	LCTEEN 83	Ecuador	Zamora River	native
T. cacao	LCTEEN 84	Ecuador	Zamora River	native
T. cacao	LCTEEN 86	Ecuador	Yacuambi River	native
T. cacao	LCTEEN 91	Ecuador	Nangaritza River	native
T. cacao	Matina 1-6	Brazil	Costa Rica Selection	old domesticated
T. cacao	MO 121	Peru	Morona River	native
T. cacao	MO 96	Peru	Morona River	native
T. cacao	NA 13	Peru	Nanay River	native
T. cacao	NA 141	Peru	Nanay River	native
T. cacao	PA 12	Peru	Parinari	native
T. cacao	PA 120	Peru	Parinari	native
T. cacao	PALSITE	Ecuador	Palanda area	native
T. cacao	SALF1	Ecuador	Palanda area	native
T. cacao	SALF10	Ecuador	Palanda area	native
T. cacao	SALF 5	Ecuador	Palanda area	native
T. cacao	SALF 6	Ecuador	Palanda area	native
T. cacao	SALF 8	Ecuador	Palanda area	native
T. cacao	SALF 9	Ecuador	Palanda area	native
T. cacao	Sca 6	Peru	Ucayali River-Contamana	native
T. cacao	SIC8 64	Brazil	Itabuna	old domesticated
T. cacao	SNA 1001	Ecuador	Pichilinge	old domesticated
T. cacao	SP 1	Venezuela	Julia	cultivated
T. bicolor	Tr6	Costa rica	Limon	native
T. grandiflorum	CR	Colombia	Caqueta River	wild
T. microcarpum	CR	Colombia	Caqueta River	wild
T. obovatum	CR	Colombia	Caqueta River	wild
T. speciosum	Tr1	Brazil	Belem	wild
T. subincanum	Tr1	(Catie collection)	information not available	wild
T. simiarum	Tr1	Costa rica	Limon	native
H.albiflora	CR	Colombia	UWI Campus	wild
H. breviligulata	CR	Colombia	Putumayo River	wild
H. camargoana	CR	Colombia	Negro River	wild
H. nitida	CR	Colombia	Caqueta River	wild
H. pulcherrima	CR	Colombia	Lloro	wild
H. umbratica	CR	Colombia	Catatumbo River	wild

Supplementary Table 6 | Homologous sequences of the closest species. For each marker, the homologous sequences of the closest species, out of the *Theobroma* or *Herrania* genus, were searched after BLASTn of the reference *T. cacao* sequence against the NCBI EST database. No homology was found for DFR3 marker with sequences from the EST or the nucleotide collections of the NCBI databases. For the other two markers, the closest species and accession number of corresponding sequences are indicated.

Marker	*species*	accession number
MITO 6	*Gossypium hirsutum*	gb|DW493071.1|
MITO 6	*Camellia sinensis*	dbj|FS945805.1|
MITO 6	*Prunus dulcis*	gb|BQ641073.1|
MITO 6	*Coffea canephora*	gb|GW346002.1|
MITO 6	*Coffea arabica*	gb|BQ448875.1|
MITO 6	*Opuntia streptacantha*	gb|HO058767.1|
MITO 6	*Casuarina glauca*	emb|FQ325045.1|
MITO 6	*Acacia mangium*	dbj|FS584083.1|
MITO 6	*Artemisia annua*	gb|GW330314.1|
MITO 6	*Quercus petraea*	gb|GW330314.1|
MITO 6	*Vernonia Galamensis*	gb|GO614975.1|
MITO 80	*Gossypium barbadense*	gb|JK792079.1|
MITO 80	*Cistus creticus*	gb|FF404083.1|
MITO 80	*Quercus robur*	emb|FR632144.1|
MITO 80	*Vitis vinifera*	gb|GR904856.1|
MITO 80	*Juglans hindsii x Juglans regia*	gb|EL903615.1|
MITO 80	*Populus tremula*	gb|BU890203.1|
MITO 80	*Casuarina glauca*	emb|FQ375866.1|
MITO 80	*Lotus japonicus*	gb|GO038784.1|
MITO 80	*Espina Barnadesia*	gb|GE535219.1|
MITO 80	*Euphorbia esula*	gb|DV137640.1|

Supplementary Table 7 | Number of specific sequences observed after targeted amplifications of aDNA extracts from the archaeological samples with two mitochondrial markers and one nuclear marker. The sequence specificity is indicated in the first line for each marker, and the number of selected clusters/sequences is indicated for each aDNA extract and archaeological sample. Four negative controls were included in the analysis. The N-zero sample corresponded to a control of the final NGS procedure used to check for acceptable sorting of barcodes specific to each DNA extract.

marker type		mitochondrial		mitochondrial		nuclear				mitochondrial		mitochondrial		nuclear	
Archeol. sample	extraction N°	MITO6	MITO6	Mito80	Mito80	DFR3	Theobromine presence	Archeol. sample	extraction N°	MITO6	MITO6	Mito80	Mito80	DFR3	Theobromine presence
		SNP specific to *T. cacao*	SNP specific to *T. cacao* or its wild relatives	SNP specific to *T. cacao*	SNP specific to Wild relatives	SNP specific to *T. cacao* or its wild relatives				SNP specific to *T. cacao*	SNP specific to *T. cacao* or its wild relatives	SNP specific to *T. cacao*	SNP specific to Wild relatives	SNP specific to *T. cacao* or its wild relatives	
PLD-001	255	0	0	0	0	0	YES	PLD-030	59	3/4	8/12	0	0	0	YES
PLD-001	285	0	2/2	0	0	0		PLD-030	71	5/5	12/15	0	3/5	0	
PLD-001	52	3/4	0	0	0	0		PLD-030	261	17/38	20/44	0	1/1	0	
PLD-001	66	17/34	1/1	2/2	0	0		PLD-030	291	52/112	97/296	0	3/3	0	
PLD-001	79	48/169	4/4	0	0	0									
PLD-001	141	1/1	0	0	0	0		PLD-038	263	7/8	17/28	3/3	1/1	0	YES
PLD-001	171	184/540	5/5	1/1	1/1	0		PLD-038	293	5/5	9/10	0	1/1	0	
PLD-001	323	3/10	0	287/528	5/6	0		PLD-038	61	66/91	131/247	6/7	36/72	0	
								PLD-038	73	6/7	44/69	0	3/5	0	
PLD-002	254	2/2	0	1/1	1/1	0	YES								
PLD-002	284	0	0	262/477	0	0		PLD-040	294	27/39	30/44	0	1/1	0	YES
PLD-002	63	1/1	8/11	0	1/1	0		PLD-040	112	13/18	25/42	0	3/3	0	
PLD-002	77	8/25	7/10	1/1	1/1	0		PLD-040	197	2/2	2/3	0	0	0	
PLD-002	170	5/5	4/4	3/3	0	0		PLD-040	227	0	2/2	3/3	10/18	0	
PLD-002	322	10/10	5/6	2/3	0	0		PLD-040	264	3/3	3/7	0	1/1	0	
PLD-003	257	1/1	0	1/1	1/1	0	YES	PLD-042	228	6/6	13/24	0	4/6	60/168	YES
PLD-003	287	0	5/10	2/7	0	0		PLD-042	265	8/10	20/30	0	10/19	0	
PLD-003	81	43/123	60/152	0	1/1	0		PLD-042	113	185/484	4608/6809	0	2/2	0	
PLD-003	236	0	0	1/1	0	0									
PLD-003	243	1/1	3/3	0	0	0		PLD-046	116	4/5	0	103/142	0	0	YES
PLD-003	273	0	0	0	7/7	0		PLD-046	266	1/1	2/2	0	0	0	
PLD-009	151	1/1	45/99	2/2	24/35	0	YES	PLD-052	251	5/7	9/12	4/9	8/16	0	
								PLD-052	281	2/2	13/20	2/2	17/27	0	
PLD-010	122	4/5	20/50	1/1	1/1	0	YES								
PLD-010	152	0	1/1	1/4	0	0		PLD-053	252	0	0	4/9	1/1	0	
								PLD-053	282	2/2	2/3	52/106	0	0	
PLD-011	129	2/2	18/24	0	1/1	0									
								PLD-079	202	1/1	0	0	0	0	
PLD-020	91	942/2593	172/460	1/1	14/31	1/1	YES	PLD-079	232	0	0	2/2	0	0	
PLD-020	258	12/21	25/42	0	11/25	0		PLD-079	237	1/1	1/1	2/4	0	0	
PLD-020	288	1/1	15/20	0	0	0		PLD-079	267	0	0	1/1	0	0	
								PLD-079	75	2/2	29/114	0	0	0	
PLD-025	212	11/13	33/50	0	5/5	0	YES								
PLD-025	96	30/123	53/176	0	0	0		PLD-088	80	5/5	251/749	0	1/1	0	
PLD-025	183	16/35	33/56	0	10/42	0		PLD-088	235	0	2/3	0	0	0	
PLD-025	213	0	0	0	0	0		PLD-088	241	0	6/7	1/1	1/1	0	
PLD-025	97	23/31	228/798	1/1	10/13	0		PLD-088	271	12/23	0	0	0	0	
PLD-025	259	1/1	1/1	1/1		0		PLD-088	158	0	0	30/33	47/124	0	
PLD-025	289	0	6/6	0	4/6	0		PLD-088	249	0	0	0	0	0	
								PLD-088	279	8/8	19/30	0	2/2	0	
PLD-026	58	8/12	9/10	0	3/3	0	YES								
PLD-026	82	16/17	133/359	0	4/4	0		PLD-112	125	0	0	0	0	0	YES
PLD-026	260	0	3/4	0	1/1	0		PLD-112	155	0	0	1/1	14/34	0	
PLD-026	290	1/2	7/7	0	2/3	0		PLD-112	246	0	0	0	0	0	
								PLD-112	276	1/1	1/1	0	0	0	
N-zéro	NGS Contr.	6/14	4/4	0	0	0									
N1	Control	0	0	1/1	0	0		PLD-131	126	37/171	37/89	0	0	0	YES
N2	Control	3/3	11/13	1/1	4/5	0		PLD-131	156	3/4	1/1	1/1	0	0	
N3	Control	0	0	1/1	0	0		PLD-131	247	0	1/1	0	0	0	
N4	Control	0	0	1/1	0	0		PLD-131	277	0	0	0	0	0	

Supplementary Table 8 | NGS sequencing after capture. Number of reads produced after the Hiseq3000 sequencing.

Archaeological samples	extraction n°	Consensus sequences R1-R2	# of selected sequences after mapping	# of unique clusters after mapping
PLD001	323	22368706	8650	142
PLD001	324	49948330	264515	1522
PLD002	321	69876275	222643	327
PLD005	306	51652769	64995	448
PLD005	307	52136028	44758	583
PLD005	319	57172773	40618	600
PLD072	311	62166924	83215	446
PLD072	312	27980503	76258	139
PLD073	301	58906955	25699	411
PLD073	302	74769919	25642	555
control	316	2236362	1927	46

Supplementary Table 9 | Characteristics of sequences containing SNPs specific to *Theobroma*. Selected after DNA capture, mapping on the *T. cacao* genome, and BLASTed against the NCBI-nr database. The extraction number is noted after the sample name.

SNP name	CH	position (V2)	archaeological sample	Nb sequences	length	e-value aligned sequence
3248626|F|0--8	1	28459204	PLD005-306	17302	74	4,00E-11
3249023|F|0--49	1	28459258	PLD005-306	17302	74	4,00E-11
7358640|F|0—15	1	35319403	PLD073-301	9	130	5,00E-58
			PLD073-302	11	130	2,00E-57
			PLD005-306	6	130	5,00E-58
			PLD005-307	9	130	5,00E-58
			PLD072-311	9	130	2,00E-57
			PLD005-319	30	130	5,00E-58
			PLD002-321	2	130	5,00E-58
			PLD001-324	5	130	5,00E-58
3249476|F|0—68	5	8186853	PLD073-302	482	115	7,00E-50
			PLD005-306	540	115	2,00E-50
			PLD005-307	553	115	7,00E-50
			PLD072-311	691	115	2,00E-50
			PLD002-321	141	115	3,00E-48
			PLD001-324	819	115	2,00E-50
3249476|F|0—13	5	8186908	PLD073-302	482	115	7,00E-50
			PLD005-306	540	115	2,00E-50
			PLD005-307	553	115	7,00E-50
			PLD072-311	691	115	2,00E-50
			PLD002-321	141	115	3,00E-48
			PLD001-324	819	115	2,00E-50
3241303|F|0—36	5	34722614	PLD072-311	15	152	2,00E-52
			PLD073-301	59	155	5,00E-66
			PLD073-302	94	155	2,00E-65
			PLD005-306	1718	155	1,00E-67
			PLD005-307	1743	155	2,00E-65
			PLD072-311	2665	155	1,00E-67
			PLD002-321	417	155	1,00E-67
			PLD001-324	1892	155	1,00E-67
3223225|F|0—12	5	38151505	PLD073-302	2	71	3,00E-25
			PLD002-321	3	71	3,00E-25
4744853|F|0—12	7	121814	PLD005-307	45	103	7,00E-44
			PLD072-311	7	103	7,00E-44
			PLD005-319	48	103	7,00E-44
			PLD002-321	19	103	2,00E-43
			PLD001-324	70	103	7,00E-44
3228603|F|0—47	7	429218	PLD005-306	1	112	4,00E-42
			PLD005-307	5	112	4,00E-42
			PLD005-319	1	112	4,00E-42
			PLD072-311	2	112	4,00E-42
			PLD072-312	3	112	4,00E-42
			PLD002-321	5	112	4,00E-42
			PLD001-323	1	112	4,00E-42
7364548|F|0—32	7	21023660	PLD005-307	5	178	7,00E-71
			PLD073-301	1	178	1,00E-75
			PLD072-311	2	178	5,00E-73
			PLD001-324	1	178	1,00E-75
3230278|F|0—35	7	21023663	PLD005-307	5	178	7,00E-71
			PLD073-301	1	178	1,00E-75
			PLD072-311	2	178	5,00E-73
			PLD001-324	1	178	1,00E-75
3230278|F|0—44	7	21023672	PLD005-307	5	178	7,00E-71
			PLD073-301	1	178	1,00E-75
			PLD072-311	2	178	5,00E-73
			PLD001-324	1	178	1,00E-75
3672232|F|0—7	9	709720	PLD005-306	8	79	1,00E-27
			PLD005-307	11	79	1,00E-27
			PLD073-311	6	79	1,00E-27
			PLD005-319	12	79	1,00E-27
			PLD002-321	4	79	1,00E-27
7364264|F|0—14	9	709727	PLD005-306	8	79	1,00E-27
			PLD005-307	11	79	1,00E-27
			PLD072-311	6	79	1,00E-27
			PLD005-319	12	79	1,00E-27
			PLD002-321	4	79	1,00E-27
3672232|F|0—23	9	709736	PLD005-306	8	79	1,00E-27
			PLD005-307	11	79	1,00E-27
			PLD072-311	6	79	1,00E-27
			PLD005-319	12	79	1,00E-27
			PLD002-321	4	79	1,00E-27
4739729|F|0—46	9	2108892	PLD001-324	2	59	4,00E-21
4739729|F|0—31	9	2108907	PLD001-324	2	59	4,00E-21
4739729|F|0—28	9	2108910	PLD001-324	2	59	4,00E-21
3225480|F|0—10	9	8165487	PLD005-306	1	100	2,00E-34
			PLD002-321	1	100	2,00E-34

Supplementary Table 11 | Modern *Theobroma* and *Herrania* accessions used to evaluate the contribution of *T. cacao* or wild relative species to the mixture of aDNA extracted from the ceramic residues of five archaeological samples (PLD-001, PLD-002, PLD-005, PLD-072, PLD-073).

Species	Accession	Country of origin	Origin	genetic group	*Species*	Accession	Country of origin	Origin	Genetic group
T. cacao	Catongo	Brazil	Brazil	Amelonado	*T. cacao*	NANK 4	Ecuador	Zamora Chinchipe province	Nacional
T. cacao	IFC 5	Côte d'Ivoire	Côte d'Ivoire	Amelonado	*T. cacao*	NANK 5	Ecuador	Zamora Chinchipe province	Nacional
T. cacao	Matina 1-6	Costa Rica	Brazil	Amelonado	*T. cacao*	NANK 6	Ecuador	Zamora Chinchipe province	Nacional
T. cacao	VEN 4	Venezuela	Venezuela	Amelonado	*T. cacao*	PANG 1	Ecuador	Zamora Chinchipe province	Nacional
T. cacao	VEN 20	Venezuela	Venezuela	Amelonado	*T. cacao*	PANG 10	Ecuador	Zamora Chinchipe province	Nacional
T. cacao	SCA 3	Peru	Rio Ucayali	Contamana	*T. cacao*	PANG 12	Ecuador	Zamora Chinchipe province	Nacional
T. cacao	SCA 9	Peru	Rio Ucayali	Contamana	*T. cacao*	PANG 13	Ecuador	Zamora Chinchipe province	Nacional
T. cacao	SCA 24	Peru	Rio Ucayali	Contamana	*T. cacao*	PANG 14	Ecuador	Zamora Chinchipe province	Nacional
T. cacao	SCA 12	Peru	Rio Ucayali	Contamana	*T. cacao*	PANG 15	Ecuador	Zamora Chinchipe province	Nacional
T. cacao	SCA 6	Peru	Rio Ucayali	Contamana	*T. cacao*	PANG 16	Ecuador	Zamora Chinchipe province	Nacional
T. cacao	GU 143B	France	French Guiana	Guiana	*T. cacao*	PANG 18	Ecuador	Zamora Chinchipe province	Nacional
T. cacao	GU 156B	France	French Guiana	Guiana	*T. cacao*	PANG 19	Ecuador	Zamora Chinchipe province	Nacional
T. cacao	GU 175V	France	French Guiana	Guiana	*T. cacao*	PANG 2	Ecuador	Zamora Chinchipe province	Nacional
T. cacao	GU 195V	France	French Guiana	Guiana	*T. cacao*	PANG 20	Ecuador	Zamora Chinchipe province	Nacional
T. cacao	GU 225B	France	French Guiana	Guiana	*T. cacao*	PANG 21	Ecuador	Zamora Chinchipe province	Nacional
T. cacao	GU 230B	France	French Guiana	Guiana	*T. cacao*	PANG 23	Ecuador	Zamora Chinchipe province	Nacional
T. cacao	GU 255V	France	French Guiana	Guiana	*T. cacao*	PANG 24	Ecuador	Zamora Chinchipe province	Nacional
T. cacao	IMC 107	Peru	Iquitos	Iquitos	*T. cacao*	PANG 5	Ecuador	Zamora Chinchipe province	Nacional
T. cacao	IMC 31	Peru	Iquitos	Iquitos	*T. cacao*	PANG 6	Ecuador	Zamora Chinchipe province	Nacional
T. cacao	IMC 38	Peru	Iquitos	Iquitos	*T. cacao*	PANG 7	Ecuador	Zamora Chinchipe province	Nacional
T. cacao	IMC 48	Peru	Iquitos	Iquitos	*T. cacao*	PAQ 1	Ecuador	Zamora Chinchipe province	Nacional
T. cacao	IMC 50	Peru	Iquitos	Iquitos	*T. cacao*	YACU 1	Ecuador	Zamora Chinchipe province	Nacional
T. cacao	IMC 65	Peru	Iquitos	Iquitos	*T. cacao*	YACU 11	Ecuador	Zamora Chinchipe province	Nacional
T. cacao	IMC 67	Peru	Iquitos	Iquitos	*T. cacao*	YACU 12	Ecuador	Zamora Chinchipe province	Nacional
T. cacao	IMC 71	Peru	Iquitos	Iquitos	*T. cacao*	YACU 15	Ecuador	Zamora Chinchipe province	Nacional
T. cacao	IMC 76	Peru	Iquitos	Iquitos	*T. cacao*	YACU 16	Ecuador	Zamora Chinchipe province	Nacional
T. cacao	PA 125	Peru	Parinari	Marañón	*T. cacao*	YACU 17	Ecuador	Zamora Chinchipe province	Nacional
T. cacao	PA 126	Peru	Parinari	Marañón	*T. cacao*	YACU 18	Ecuador	Zamora Chinchipe province	Nacional
T. cacao	PA 169	Peru	Parinari	Marañón	*T. cacao*	YACU 19	Ecuador	Zamora Chinchipe province	Nacional
T. cacao	PA 191	Peru	Parinari	Marañón	*T. cacao*	YACU 2	Ecuador	Zamora Chinchipe province	Nacional
T. cacao	PA 279	Peru	Parinari	Marañón	*T. cacao*	YACU 3	Ecuador	Zamora Chinchipe province	Nacional
T. cacao	PA 293	Peru	Parinari	Marañón	*T. cacao*	YACU 5	Ecuador	Zamora Chinchipe province	Nacional
T. cacao	PA 296	Peru	Parinari	Marañón	*T. cacao*	YACU 6	Ecuador	Zamora Chinchipe province	Nacional
T. cacao	PA 30	Peru	Parinari	Marañón	*T. cacao*	YACU 7	Ecuador	Zamora Chinchipe province	Nacional
T. cacao	PA 301	Peru	Parinari	Marañón	*T. cacao*	YACU 9	Ecuador	Zamora Chinchipe province	Nacional
T. cacao	PA 303	Peru	Parinari	Marañón	*T. cacao*	ZAMO 1	Ecuador	Zamora Chinchipe province	Nacional
T. cacao	PA 150	Peru	Parinari	Marañón	*T. cacao*	ZAMO 10	Ecuador	Zamora Chinchipe province	Nacional
T. cacao	PA 272	Peru	Parinari	Marañón	*T. cacao*	ZAMO 11	Ecuador	Zamora Chinchipe province	Nacional
T. cacao	NA 127	Peru	Rio Nanay	Nanay	*T. cacao*	ZAMO 12	Ecuador	Zamora Chinchipe province	Nacional
T. cacao	NA 184	Peru	Rio Nanay	Nanay	*T. cacao*	ZAMO 13	Ecuador	Zamora Chinchipe province	Nacional
T. cacao	NA 191	Peru	Rio Nanay	Nanay	*T. cacao*	ZAMO 14	Ecuador	Zamora Chinchipe province	Nacional
T. cacao	NA 435	Peru	Rio Nanay	Nanay	*T. cacao*	ZAMO 15	Ecuador	Zamora Chinchipe province	Nacional
T. cacao	NA 672	Peru	Rio Nanay	Nanay	*T. cacao*	ZAMO 16	Ecuador	Zamora Chinchipe province	Nacional
T. cacao	NA 678	Peru	Rio Nanay	Nanay	*T. cacao*	ZAMO 3	Ecuador	Zamora Chinchipe province	Nacional
T. cacao	NA 697	Peru	Rio Nanay	Nanay	*T. cacao*	ZAMO 4	Ecuador	Zamora Chinchipe province	Nacional
T. cacao	NA 218	Peru	Rio Nanay	Nanay	*T. cacao*	ZAMO 5	Ecuador	Zamora Chinchipe province	Nacional
T. cacao	Na 32	Peru	Rio Nanay	Nanay	*T. cacao*	ZAMO 6	Ecuador	Zamora Chinchipe province	Nacional
T. cacao	NA 33	Peru	Rio Nanay	Nanay	*T. cacao*	ZAMO 7	Ecuador	Zamora Chinchipe province	Nacional
T. cacao	NA4 6	Peru	Rio Nanay	Nanay	*T. cacao*	ZAMO 8	Ecuador	Zamora Chinchipe province	Nacional
T. cacao	P 7	Peru	Rio Nanay	Nanay	*T. cacao*	ZAMO 9	Ecuador	Zamora Chinchipe province	Nacional
T. cacao	EBC 10	Colombia	Caqueta River	Purus	*H. breviligulata*	CR	Colombia	Rio Putumayo	wild relative
T. cacao	EBC 125	Colombia	Caqueta River	Purus	*H. nitida(2)*	CR -TD	Colombia	Rio Caqueta	wild relative
T. cacao	EBC 5	Colombia	Caqueta River	Purus	*H. umbratica*	CR	Colombia	Rio Catatumbo	wild relative
T. cacao	LCTEEN 362	Ecuador	Rio Conambo	Purus	*H. albiflora (2)*	CR -TD	Colombia	N/A	wild relative
T. cacao	LCTEEN 368	Ecuador	Rio Conambo	Purus	*H. balaensis (2)*	CR -TD	CRC/Catie collections	N/A	wild relative
T. cacao	LCTEEN 411	Ecuador	Rio Aguarico	Purus	*H. kanukensis*	TD	CRC collection-Trinidad	N/A	wild relative
T. cacao	LCTEEN 413	Ecuador	Rio Aguarico	Purus	*H. cuatrecasana*	CR	(CATIE collection)	N/A	wild relative
T. cacao	LCTEEN 163 A	Ecuador	Coca	Curaray	*H. nycterodendron*	CR	(CATIE collection)	N/A	wild relative
T. cacao	LCTEEN 327	Ecuador	Rio San Miguel	Curaray	*H. purpurea*	CR	(CATIE collection)	N/A	wild relative
T. cacao	LCTEEN 333	Ecuador	Rio Upano	Curaray	*T. speciosum (3)*	Tr1 -Tr2-TD	(CATIE collection)	N/A	wild relative
T. cacao	LCTEEN 57	Ecuador	Loreto	Curaray	*T. simiarum (2)*	Tr1 -Tr2	COSTA RICA	Limon	wild relative
T. cacao	LCTEEN 83	Ecuador	Rio Zamora	Curaray	*T. Mammosum*	CR	(CATIE collection)	N/A	wild relative
T. cacao	MO 121	Peru	Rio Morona	Nacional	*T. subincanum (3)*	Tr1 - Tr2 -TD	(CATIE collection)	N/A	wild relative
T. cacao	SNA 1003	Ecuador	Pichilingue	Nacional	*T. bicolor (2)*	Tr6 -Tr7	Costa Rica (CR)	Limon	wild relative
T. cacao	SNA 604	Ecuador	Pichilingue	Nacional	*T. grandiflorum (2)*	CR -TD	Colombia	Rio Caqueta	wild relative
T. cacao	NANK 1	Ecuador	Zamora Chinchipe province	Nacional	*T. microcarpum (2)*	CR -TD	Colombia	Rio Caqueta	wild relative
T. cacao	NANK 2	Ecuador	Zamora Chinchipe province	Nacional					

Supplementary Table 12 | Statistics on NGS (next-generation sequencing) of PCR and qPCR pool libraries. After pooling of PCR and qPCR products amplified with the 3 markers from each DNA extract, 151 aDNA libraries and 4 negative controls were sequenced using an Illumina MiSeq sequencing system. The number of total and overlapped reads, as well as the number of aDNA sequences mapped on modern *T. cacao* sequences, is given for each DNA extract and each marker. The extraction procedure (A: Macherey procedure; B: Powersoil procedure – see Supplementary Note 2) is indicated for each DNA extract.

Archaeological sample	extraction n°	Extraction method	number of paired end reads	number of overlapped paired end reads	number of unique clusters	number of clusters and sequences mapped on MITO6/B97-61	number of clusters and sequences mapped on MITO80/B97-61	number of clusters and sequences mapped on DFR3/B97-61
PLD-001	52	A	270854	243336	10885	612/4627	5/10	0
PLD-001	66	B	589013	523387	17740	5532/26092	14/39	0
PLD-001	79	B	2158453	1952265	64163	15505/89754	30/79	0
PLD-001	141	A	127482	111541	7905	212/373	4/17	0
PLD-001	171	B	769035	711059	50513	707/1851	13/26	0
PLD-001	255	B	178238	137932	14319	56/252	6/9	0
PLD-001	285	A	208889	151424	27797	230/615	7/12	0
PLD-002	63	B	134927	98103	8567	1254/4324	80/105	0
PLD-002	77	B	364340	314259	13270	3862/25061	202/988	0
PLD-002	170	B	121444	110690	14913	490/1246	10/19	0
PLD-002	254	B	653492	443872	22141	239/947	97/138	0
PLD-002	284	A	526016	405046	26094	103/416	379/789	0
PLD-002	322	B				1825/3248	2/2	0
PLD-003	81	B	305444	267922	12059	233/501	17/23	0
PLD-003	236	B	561013	482147	11493	88/293	19/26	0
PLD-003	243	B	87543	67984	6301	15/16	1/2	0
PLD-003	257	B	65293	47309	5912	142/455	82/127	0
PLD-003	273	A	282343	232294	18126	54/258	229/450	0
PLD-003	287	A	406036	315299	27017	227/1653	19/45	0
PLD-003	323	B				2295/4354	295/537	0
PLD-004	272	A	334580	268000	27571	15/27	5/5	0
PLD-005	306	B				384/901	7/9	0
PLD-005	307	B				24/61	7/14	0
PLD-005	308	B				409/1057	5/7	0
PLD-005	309	B				6/22	5/9	0
PLD-005	310	B				309/753	3/3	0
PLD-005	319	B				26/36	1/1	0
PLD-006	78	B	41019	35511	1895	14/20	4/7	0
PLD-006	234	B	65840	54513	3571	116/692	5/8	0
PLD-006	240	B	20838	17321	2690	53/151	2/3	0
PLD-007	118	A	111093	80836	14224	464/1141	302/415	0
PLD-008	119	A	17844	15065	3322	20/25	378/4051	0
PLD-008	149	B	641143	543566	38054	44/136	5/13	0
PLD-009	121	A	152446	113540	14481	252/613	17/47	0
PLD-009	151	B	168536	140336	14838	99/208	1413/7130	0
PLD-010	122	A	1044672	573991	37596	739/1731	30/79	0
PLD-010	152	B	615666	531119	27827	191/378	16/45	0
PLD-011	129	A	129234	108748	8835	416/2976	5/5	0
PLD-020	91	A	2331654	1693464	67824	35236/233532	85/166	1/1
PLD-020	258	B	33467	23164	5952	145/356	139/342	0
PLD-020	288	A	60906	32527	5202	96/208	14/26	0
PLD-025	96	A	130445	98031	8730	220/558	40/98	0
PLD-025	97	A	402296	310851	19579	11560/59240	626/3905	0
PLD-025	183	A	11835	8864	1514	150/323	43/112	0
PLD-025	212	B	342796	277350	8849	181/438	29/56	0
PLD-025	259	B	5267	1528	517	61/237	26/75	0
PLD-025	289	A	3932	1053	479	54/146	14/18	0
PLD-026	58	A	21589	16260	2260	91/131	26/43	0
PLD-026	82	B	271710	236304	23057	7431/34950	21/28	0
PLD-026	260	B	4927	1919	716	34/48	410/4	0
PLD-026	290	A	97716	61976	7349	28/45	9/15	0
PLD-028	99	A	5092387	4530583	165436	96/221	15/41	0
PLD-028	216	B	84243	61701	6679	3/8	138/182	5/8
PLD-029	188	A	388740	341153	17164	49/132	5/9	0
PLD-029	218	B	114459	59328	4953	21/49	4/4	0
PLD-030	59	A	147645	125065	10375	92/252	11/22	0
PLD-030	71	B	300831	214560	14368	106/179	443/1228	0
PLD-030	261	B	2993	1167	450	232/763	7/13	0
PLD-030	291	A	316564	264865	15917	395/865	38/57	0
PLD-037	109	A	108331	87195	5641	23/39	831/10904	0
PLD-038	61	A	2212050	1326124	116022	2060/4400	351/11988	0
PLD-038	73	B	126485	99275	8802	140/196	18/27	0
PLD-038	263	B	3443	1345	640	106/174	16/24	0
PLD-038	293	A	74201	48341	6325	85/162	44/152	0
PLD-040	112	A	130458	78666	3505	129/206	10/88	0
PLD-040	197	A	647019	506267	45876	54/159	9/17	0
PLD-040	227	B	92539	47097	5169	106/304	1801/21115	0
PLD-040	264	B	60457	28919	5283	36/64	12/17	0
PLD-040	294	A	178657	117213	6594	406/731	37/82	0
PLD-042	113	A	695628	589965	28373	6454/70994	28/38	0
PLD-042	228	B	180753	30079	8205	336/1719	67/252	60/168
PLD-042	265	B	46834	17513	3497	79/130	41/88	0
PLD-045	103	A	159214	126366	7312	37/56	2/3	0
PLD-045	104	A	190249	158572	6953	49/71	14/57	0
PLD-046	116	A	192061	165214	12193	33/71	108/152	0
PLD-046	266	B	88739	54298	7003	58/141	2/3	0
PLD-047	135	A	128462	99618	10037	452/709	1465/13391	0
PLD-047	165	B	283820	184843	22231	1237/2731	51/110	0
PLD-048	136	A	60742	49959	9287	829/1290	874/7509	0
PLD-048	166	B	159708	131993	16278	6835/26507	5/19	0
PLD-050	138	A	734857	560380	48520	5421/25266	5358/27353	0
PLD-050	168	B	675298	518432	30913	2008/6560	970/4805	0
PLD-051	139	A	48399	42616	6189	2017/7264	799/15027	0
PLD-051	169	B	160348	144331	19349	3281/11329	3/5	0
PLD-052	251	B	20225	15486	2581	174/365	41/80	0
PLD-052	281	A	282683	243396	19759	73/132	948/2427	0
PLD-053	252	B	570803	388679	21608	50/149	227/1525	0
PLD-053	282	A	5330	3204	880	90/362	111/220	0
PLD-054	74	B	357135	244978	19858	2835/26014	78/88	0
PLD-054	172	B	2535	1372	509	42/101	10/37	0
PLD-054	253	B	988331	770435	62917	115/390	141/313	0
PLD-054	283	A	369384	216685	19795	229/947	25/41	0
PLD-072	311	B				1012/2673	958/2160	0
PLD-072	312	B				1246/3788	6/6	0
PLD-072	313	B				1053/3060	781/2001	0
PLD-073	301	B				992/2076	3/4	0
PLD-073	302	B				590/1948	3/3	0
PLD-073	303	B				396/859	3/7	0
PLD-073	305	B				1166/2666	3/3	0
PLD-079	75	B	440069	364575	19955	4076/21792	13/46	0
PLD-079	202	A	106982	94043	3103	13/20	2/3	0
PLD-079	232	B	284575	139257	10961	123/509	9/29	0
PLD-079	237	B	344480	183760	10407	54/148	14/35	0
PLD-079	267	A	485611	361036	21677	153/400	75/304	0
PLD-081	203	A	168000	148204	5082	256/619	13/48	0
PLD-081	233	B	182692	151830	7884	91/417	28/4	0
PLD-088	80	B	1984060	1847401	53069	20674/127214	25/54	0
PLD-088	158	B	390247	324686	23756	56/102	3516/27609	0
PLD-088	235	B	412562	346897	7232	274/1259	8/15	0
PLD-088	241	B	235166	172288	11899	688/4913	12/28	0
PLD-088	249	B	123642	72466	6366	77/205	40/140	0
PLD-088	271	A	256497	103773	7125	121/202	21/90	0
PLD-088	279	A	13201	10939	2206	141/227	12/19	0
PLD-097	117	A	1622692	1354586	60327	469/2289	34/105	0
PLD-097	147	B	266417	230709	26610	433/1386	7/18	0
PLD-097	256	B	79614	60516	4461	39/184	3/3	0
PLD-097	274	A	595828	504094	34077	179/472	190/355	0
PLD-097	286	A	1876611	1476711	46317	493/2267	23/63	0
PLD-104	245	B	243938	157521	10351	264/675	128/1029	0
PLD-104	275	A	132611	104861	12268	39/106	34/121	0
PLD-112	125	A	193783	164340	18598	17/30	2/4	0
PLD-112	155	B	300093	238495	17660	58/98	2654/17138	0
PLD-112	246	B	386330	222980	17994	29/79	16/35	0
PLD-112	276	A	652190	486627	24766	154/334	23/50	0
PLD-116	250	B	3312	1304	4616	187/437	7/10	1/1
PLD-116	280	A	3855	1898	685	1/5	1/1	0
PLD-123	127	A	1107628	840658	65979	101/309	28/57	0
PLD-123	157	B	354217	272677	39070	602/1549	21/46	0
PLD-123	248	B	106818	69306	6572	20/55	49/447	0
PLD-123	278	A	235006	155783	15092	260/1286	17/66	0
PLD-130	123	A	313632	260316	8750	137/243	36/143	0
PLD-130	153	B	276634	219113	14114	85/318	5/6	0
PLD-131	126	A	169863	121654	11845	1088/2787	9/27	0
PLD-131	156	B	762847	621920	56366	237/622	23/29	0
PLD-131	247	B	68215	46489	4806	30/91	6/10	0
PLD-131	277	A	619605	346201	13877	189/645	11/53	0
control	N-zero		30799	1392	624	33/70	5/7	0
control	N1	A	105847	78214	4998	21/30	4/4	0
control	N2	B	97215	16271	2163	56/92	118/363	0
control	N3	A	303684	259821	13929	123	9	0

Supplementary Table 13 | Classification of miscoding lesions. Adapted from Hansen et al.[19]. Transition (TS), transversion (TV).

	Base substitution	Abbreviated name	Occurrence
Miscoding lesions originally derived from A and T nucleotides	AT ->GC	TS1	Most frequent error from Taq polymerase during PCR amplification of DNA[22]
	AT ->TA	TV1	Low frequencies in published aDNA studies
	AT ->CG	TV3	Low frequencies in published aDNA studies
Miscoding lesions originally derived from C and G nucleotides	CG ->TA	TS2	Most frequent miscodings encountered in previous aDNA studies[23-25]
	CG ->GC	TV2	Very infrequent in other aDNA studies
	CG ->AT	TV4	Some notable frequency in certain ancient DNAs[17]

Supplementary Table 14 | Base substitution types and observed rates (%) at the five positions with significantly higher polymorphisms.

Nucleotidic position	#25	#28	#46	#55	#58
Base substitution type and rate (%)	TV3 3.2%	TV1 2.1%	TS2 5.5%	TV4 7.4%	TV4 5.1%
		TS1 1.5%		TS2 1.4%	
				TV2 0.2%	

Supplementary Table 15 | Primer pairs defined to show evidence of ancient DNA fragmentation. Three primer pairs having a similar PCR efficiency were used to amplify DNA fragments differing for their size (197bp, 290 bp, and 543 bp). These primers allowed us to assess the decreased amplification level of ancient DNA samples when associated with an increased size of the amplified fragments, which is a characteristic of ancient DNA samples.

primer name	PCR efficiency with modern DNA	gene/sequence name	sequence name	sequence reference	forward primer	Tm forw.	reverse primer	Tm rev.	product length	start	end
MITO-197	**1,91**	Cytochrome C oxydase subunit 2	ESTTik/CL294Contig1	Argout et al. (2008)	TCCTTGTGATGCAGCGGAAC	60.96	CATGAACAATCCTTTGCGGGA	59.18	197	1596	1792
MITO-290	**1,94**	Cytochrome C oxydase subunit 2	ESTTik/CL294Contig1	Argout et al. (2008)	CCTTGTGATGCAGCGGAAC	59.50	TCGTCCATTGAGTATAAGAGAGCA	59.11	290	1597	1886
MITO-543	**1,94**	Cytochrome C oxydase subunit 2	ESTTik/CL294Contig1	Argout et al. (2008)	TTTCCATTCTCCCATTGATTCCT	57.79	CAACCACTCTATTGTCCACTTCT	58.10	543	1526	2068

Supplementary Table 16 | Ct values obtained from qPCR amplifications of three DNA fragments of different sizes carried out from ten aDNA samples. The qPCR amplifications were carried out on a Lightcycler 480 apparatus (Roche) with 3 primer pairs amplifying DNA differing in size (197 bp, 290 bp, and 543 bp) from ten aDNA samples. The Ct (threshold cycle) results, a relative measurement of the concentration of the target in the PCR reaction, are reported for each qPCR amplification reaction.

Sample name	DNA extraction number	Ct values		
		MITO-197	MITO-290	MITO-543
PLD-054	283	39,52	45,88	-
PLD-001	79	25,94	28,54	-
PLD-020	258	29,84	33,82	44,25
PLD-025	183	35,44	52,87	-
PLD-025	212	32,64	40,14	-
PLD-038	61	38,44	40,97	-
PLD-042	113	38,22	48,23	-
PLD-042	265	32,33	37,89	49,35
PLD-010	122	37,32	49,92	-
PLD-050	168	28,55	32,28	-