








RESEARCH ARTICLE

Collection of native *Theobroma cacao* L. accessions from the Ecuadorian Amazon highlights a hotspot of cocoa diversity

Olivier Fouet^{1,2}  | Rey Gaston Loor Solorzano³  | Bénédicte Rhoné^{1,2}  |
 Cristian Subía³ | Darío Calderón³ | Fabián Fernández³ | Ignacio Sotomayor³ |
 Ronan Rivallan^{1,2} | Kelly Colonges^{1,2}  | Hélène Vignes^{1,2}  |
 Freddy Angamarca³ | Byron Yaguana³ | Pierre Costet⁴ | Xavier Argout^{1,2}  |
 Claire Lanaud^{1,2} 

¹CIRAD, UMR AGAP Institut, Montpellier, France

²UMR AGAP Institut, Univ Montpellier, CIRAD, INRAE, Institut Agro, Montpellier, France

³Instituto Nacional de Investigaciones Agropecuarias, INIAP, Quito, Ecuador

⁴Chocolaterie Valrhona, Tain L'Hermitage, France

Correspondence

Olivier Fouet and Claire Lanaud, CIRAD, UMR AGAP Institut, Montpellier F-34398, France.
 Email: olivier.fouet@cirad.fr and claire.lanaud@cirad.fr

Funding information

Agropolis Fondation; Agence Nationale de la Recherche, Grant/Award Number: ANR-16-IDEX-0006; Valrhona; Montpellier Université d'Excellence

Societal Impact Statement

“Fine flavor” cocoa, known for its superior flavor and aroma, commands a higher price for farmers than “bulk” cocoa produced for market. These fine flavor cocoa varieties make an important contribution to the agricultural sector in Ecuador. However, cocoa diversity is threatened by deforestation. The effective preservation, characterization, and use of cocoa tree diversity are therefore essential to the future development of this market. We carried out participatory collection surveys with local communities in the Ecuadorian Amazon Forest, to evaluate the genetic diversity of native cocoa trees and protect trees as a resource for local communities. Accessing this wealth of diversity will aid farmers to safeguard cocoa against climate change and develop new varieties for market.

Summary

- The aromatic Nacional variety of *Theobroma cacao*, emblematic of Ecuador, is highly sought after by the chocolate industry. The modern Nacional is a hybrid population resulting from genetic admixture that has lost the specificity of the ancestral variety. In the context of progressive forest disappearance, the objective of our study was to collect, safeguard, and evaluate native aromatic cocoa trees from Ecuadorian Amazon areas previously identified as areas of origin of the Nacional variety, as well as those that extend toward the northern Amazon.
- Four collection expeditions were organized in the Ecuadorian Amazon provinces of Zamora-Chinchi, Morona-Santiago, and Pastaza in close collaboration with local communities. A total of 283 native accessions were collected and safeguarded at experimental stations and in local communities. The genetic

Olivier Fouet and Rey Gaston Loor Solorzano contributed equally to this work

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *Plants, People, Planet* published by John Wiley & Sons Ltd on behalf of New Phytologist Foundation.

diversity of the cocoa trees was analyzed by comparison to known genetic groups with a set of 48 simple sequence repeat (SSR) markers.

- This new collection clearly enriches the currently known diversity and improves knowledge of the global genetic structure of cocoa trees. Our results clarify the geographic origin of the Nacional variety in the vicinity of an archeological site that housed a Maya Chinchipe population that consumed cacao 5,000 years ago. In addition, our analyses revealed clues to the origin of Criollo, another ancient variety with a fine flavor.
- These new genetic resources will be used in breeding programs for the varietal improvement of new aromatic cocoa varieties and more globally for the selection of new varieties adapted to environmental changes.

KEYWORDS

aromatic cocoa, diversity hotspot, Ecuadorian Amazon, genetic rescue, Nacional variety, participatory collects, *T. cacao*

1 | INTRODUCTION

Theobroma cacao L. is a diploid ($2n = 2x = 20$) perennial species of the Malvaceae family (Whitlock et al., 2001). The classification of *T. cacao* accessions according to their genetic diversity has evolved significantly over the last century. First, two distinct species, Forastero and Criollo, were proposed by Pittier (1930). Then, Cuatrecasas (1964) suggested two subspecies, ssp. *cacao* and ssp. *sphaerocarpum*. Accessions of *T. cacao* were also classified according to their morphogeographical traits by Cheesman (1944). More recently, a genetic study using simple sequence repeat (SSR) markers allowed a more precise classification into 10 genetic groups (Motamayor et al., 2008).

The cocoa tree is an understory tropical tree native to the upper Amazon basin. This region contains the widest diversity of *Theobroma* species (Thomas et al., 2012). Their distribution extends throughout the Amazon region to Mesoamerica, in part due to past human-mediated dispersal. The history of cocoa expansion remains to be elucidated to better understand the role of human selection in shaping this diversity.

Amazon rainforests have lost more than 20% of their area in the last 60 years (Nobre et al., 2016), and 27% of tree species are threatened according to the IUCN classification. Future projections show that climate change and deforestation could lead to a decline of up to 58% in Amazonian tree species richness. Species could lose an average of 65% of their original environmentally adapted area (Gomes et al., 2019). This continued loss of habitat could lead to an irreversible loss of *T. cacao* genetic diversity, which is essential for farmers and breeders. Indeed, wild *T. cacao* constitutes a reservoir of variability for the varietal improvement of cultivated species. The ability of cacao farming to adapt to changing environments or to more sustainable and resilient farming systems will greatly depend on the preservation of such agrobiodiversity (Satori et al., 2021). Conservation of these native cocoa trees requires access to appropriate land to maintain living collections and long-term accessibility for a sustainable

breeding process. There is an urgent need to rescue, conserve, and characterize these cocoa genetic resources in close collaboration with local communities to make the processes more demand-driven (Kline et al., 2020). A better overall characterization of wild *T. cacao* diversity would help identify the best candidates for the selection of traits such as disease resistance, aroma quality or cadmium accumulation in the bean, for which there is natural variation within the species.

Historically, the economy and development of Ecuador have been closely linked to the production and export of cocoa, particularly based on the aromatic variety Nacional. Ecuador has recently become the top cocoa producer in South America, with 322,000 tons estimated for 2018–2019, or 6.7% of world production (ICCO, 2020). The Nacional variety, also called “arriba” when traditionally grown on the Pacific coast, is classified as a fine cocoa for its floral aromas. Over the last century, the gradual introduction of foreign genetic material resistant to various fungal diseases has led to genetic mixing (Loor et al., 2009) and, consequently, alterations to the Nacional flavor. Extensive monoclonal plantings, particularly the spread of the non-flavored clone CCN51, used for its high productivity (Boza et al., 2014), have led to a partial downgrading of Ecuador's fine cocoa production. The meteoric expansion of this clone now accounts for 70% of Ecuadorian production (Jaimez et al., 2022). The current Nacional plantations called “modern Nacional” correspond to a genetic mix between the ancestral Nacional and the introduced Trinitario (Loor et al., 2009). However, nonintrogressed genotypes, corresponding to representatives of the ancestral Nacional, were identified within this hybrid population. The genetic analysis of a large collection of accessions from the Upper Amazon covering a wide geographic area (Allen & Lass, 1983; Loor et al., 2012; Pound, 1938, 1945) identified the southern part of the Ecuadorian Amazon as the putative region of the ancestral Nacional.

In recent years, several genetic diversity studies have been conducted on Amazonian cocoa trees using SSR or SNP markers, including those in Peru (Arevalo-gardini et al., 2019; Zhang et al., 2006), Brazil

(Sereno et al., 2006), Bolivia (Zhang et al., 2012), Guyana (Lachenaud et al., 2016; Lachenaud & Zhang, 2008), and Colombia (Osorio-Guarín et al., 2017). Several diversity hotspots were identified in the Upper Amazon between northern Bolivia and southern Colombia, particularly in the Ecuadorian Amazon (Thomas et al., 2012).

To enhance and rescue the genetic resources linked to aromatic cocoa in the context of the progressive disappearance of the forest (Figure 1), several expeditions were organized in collaboration with local Amerindian communities to collect native *T. cacao* trees. The diversity and genetic structure of the collected and safeguarded material compared to the global diversity of *T. cacao* species based on microsatellite markers are reported in this study.

2 | MATERIALS AND METHODS

2.1 | Method of participatory collection expeditions

Four collection expeditions were conducted in 2010, 2013, 2017, and 2019 in the Ecuadorian provinces of Zamora-Chinchipec, Morona-Santiago, and Pastaza. The sampling design of the collection expeditions performed in 2010 and 2013 has already been described by Loor et al. (2015). The last two collection expeditions in 2017 and 2019 (Dataset S1) were conducted following a participatory approach involving the local Amerindian populations in four successive steps (Figures 2 and S1).

2.2 | Plant material

During the last two collection expeditions in 2017 and 2019, 150 new *T. cacao* accessions were collected, bringing the total number of accessions surveyed in the four collection expeditions to 283. These accessions were collected in geographical areas with diverse topology and climate and were grouped into eight distinct zones. Zones 1 to 8 from south to north were representative of the different ecological conditions (Figure 3; Table 1a,b).

To compare the genetic diversity of the collected material with the known genetic diversity of *T. cacao*, representative accessions from the 10 genetic groups previously identified within the species (Motamayor et al., 2008) were included in the genetic analysis; these accessions were referred to as the control population. In addition, accessions originating from northern Peru (referred to as the “Morona” population) and from southern Colombia and the northern part of the Ecuadorian Amazon (referred to as the “Caquetá” population) were also included (Allen, 1988). A total of 418 samples were analyzed in this work.

2.3 | DNA purification and PCR amplification

Leaf samples were stored under vacuum immediately after sampling to avoid degradation during travel. Genomic DNA was extracted from cocoa leaves according to Risterucci et al. (2000) and purified

according to Allegre et al. (2011). A total of 48 SSR markers were selected from genomic DNA (Pugh et al., 2004) or expressed sequence tag sequencing data (Fouet et al., 2011) to study the dissimilarities between individuals. These SSRs are distributed across the 10 cocoa genetic linkage groups (Table S1). PCR amplifications were performed as described by Allegre et al. (2011). The raw data were analyzed with Genemapper 4.0 software (Applied Biosystems). To compare our results to those of other studies, control individuals that matched those used in similar studies were included in the analysis. In this study, 119 control individuals were selected from the study conducted by Loor et al. (2012) in the Ecuadorian Amazon, and 64 individuals were selected from the global analysis of the genetic diversity of *T. cacao* published by Motamayor et al. (2008).

2.4 | Genetic diversity

The average number of alleles per locus (N_a), expected heterozygosity (H_e) (Nei, 1978), observed heterozygosity (H_o), and number of private alleles (PA) specific to each group were calculated with the GenAIEx 6.503 program (Peakall & Smouse, 2012). Allelic richness (A_r) and private allele richness (PA_r) were estimated using a rarefaction approach with the HP-rare program (Kalinowski, 2004, 2005) to correct for unequal sample sizes; a value of g equal to twice the size of the smallest sample for a diploid organism was used (el Mousadiq & Petit, 1996). Unbiased genetic distance (D) or Nei's identity index (Nei, 1972) was calculated with GenAIEx to assess the similarity between all collected and reference groups based on allelic frequencies.

Wright's F -fixation indices were estimated (Weir & Cockerham, 1984). The fixation index (F_{IS}) or inbreeding coefficient was calculated with GENETIX (Belkhir et al., 2004) to estimate the differentiation between individuals within each population. The F_{ST} index was estimated between pairs of populations using GENEPOP web version 4.2 software (Raymond & Rousset, 1995; Rousset, 2008).

2.5 | Genetic structure of the population

Two methods were used to infer the genetic structure of the population investigated in this study. First, a distance model based on the dissimilarity matrix was calculated using the neighborhood assembly method (Perrier et al., 2003), as implemented in DARWIN 6.0.14 software (<https://darwin.cirad.fr/>). The genetic structure inferred from the 135 accessions representative of the *T. cacao* diversity was first evaluated for the concomitance with previously identified genetic groups (Motamayor et al., 2008). Then, a phylogenetic tree with both the collected and reference accessions was constructed using the neighbor-joining method and 500 bootstrap repetitions to assess the uncertainty of the tree structure. Then, we used a Bayesian model implemented in STRUCTURE 2.3.4 software (Pritchard et al., 2000). K groups from 2 to 20 were tested with a burn-in period of 100,000 iterations, 500,000 Markov chain

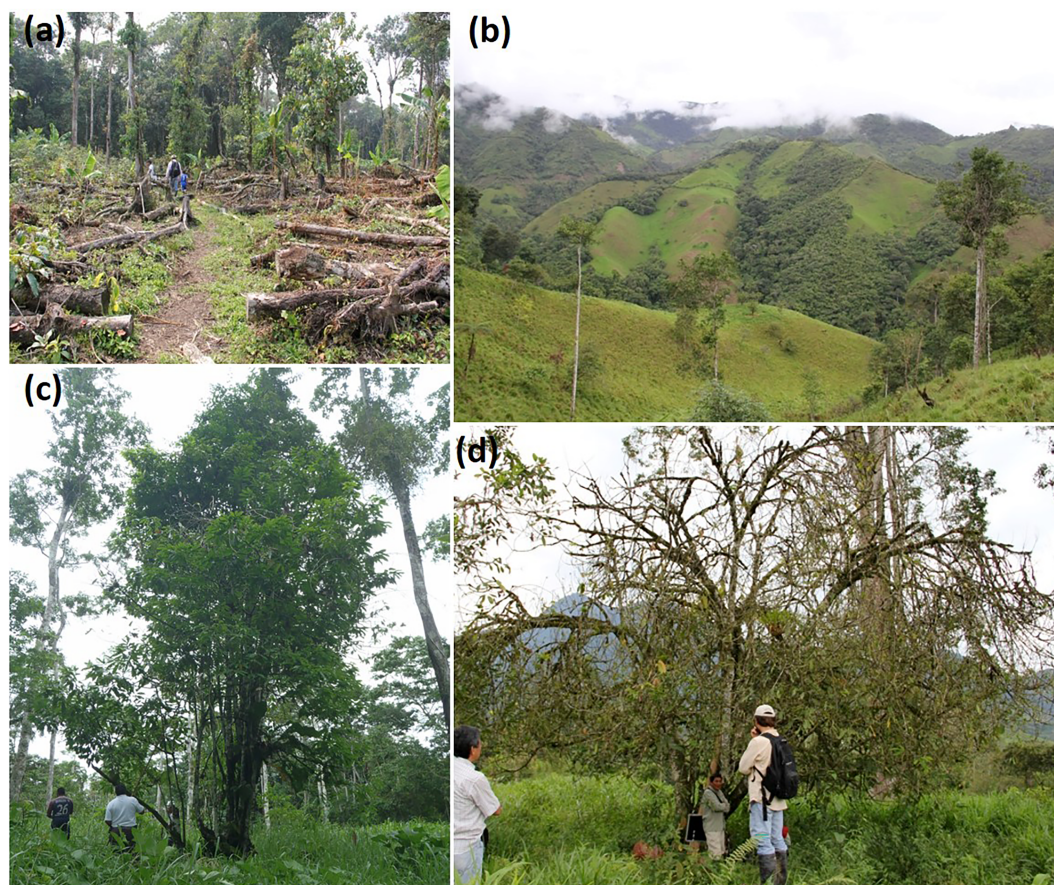


FIGURE 1 Forest degradation and transformation of cocoa habitat that can lead to its loss: (a) cutting of primary forest for the establishment of new crops leading to a direct loss of cocoa trees present; (b) cow pastures a few years after the destruction of the forest where only a few cocoa trees remain; (c) cocoa tree surviving in a recent cow pasture with little or no shade, although it is necessary; (d) almost dead cocoa tree in a pasture under the influence of the Sun. Photograph: (a)–(d) Olivier Fouet

Monte Carlo repetitions, and at least 10 repetitions per K . The optimal value of K , that is, the assumed top structure level, was determined according to the method described by Evanno et al. (2005) using the StructureSelector website (Li & Liu, 2018). This software also incorporates the CLUMPAK program (Kopelman et al., 2015), which combines many features of existing tools for postprocessing STRUCTURE results by incorporating calls to CLUMPP (Jakobsson & Rosenberg, 2007) and DISTRICT (Rosenberg, 2004).

3 | RESULTS

3.1 | Genetic diversity of the population

The intragenetic and intergenetic diversity of the collected and control populations were assessed by different genetic parameters (Table 2). A total of 369 alleles were detected with 48 SSR markers. The number of observed alleles per locus (N_a) varied between 2 and 19, with an average N_a of 7.69 over the whole population. The N_a was found to vary from 4.56 (Zone 7) to 2.92 (Zone 2) across each

collection zone, with the latter having the smallest sample size. The N_a of the whole collected population (6.15) was similar to that of the control set (6.69). After correcting for unequal group size, the high allelic richness ($A_r[g]$) of the overall collected population was confirmed (5.82) and was found to be only 12.5% lower than that of the control set (6.65). Each of the eight areas collected contains between one and six private alleles (PA), which may seem high considering the geographical proximity of the different areas surveyed. These values were comparable to those of the control groups, where PA varied between 0 and 6, except for the Purús (PA = 14) and Caquetá (PA = 15) groups. Considering all collected material, PA accounted for 13.3% of the total 369 alleles. This result could be partly related to unequal sample sizes; private allele richness (PA_r) confirmed the allelic contribution of the collections but with a greater difference between the collection set ($PA_r = 0.85$) and the control set ($PA_r = 1.68$).

The F_{IS} of the eight collected zones (0.324) was lower than that of the control groups (0.683). The majority of the control groups were in Hardy–Weinberg equilibrium with F_{IS} values near zero. However, some negative values showing an excess of heterozygosity were observed, notably for the Iquitos group (−0.360). In contrast, positive values showing an excess of homozygosity were observed for the



FIGURE 2 The four main steps of the participatory data collection with native American communities for the 2017 and 2019 surveys. (1) The authorizations to initiate the project were first obtained by Instituto Nacional de Investigaciones Agropecuarias (INIAP) from the governmental authorities and local authorities of Amerindians composed of the Shuar, Achuar, and Kichwa communities of the Amazon. (2) Contacts were established with local actors and key partners. Seminars in agricultural high schools in the villages located in the targeted regions were organized to involve students, their families, and their extended network in the project. These meetings allowed the sensitization and training of the communities on the preservation of their native *Theobroma cacao* varieties and how to preselect the genetic material of interest on the basis of specific criteria (age of the tree, color, and shape of the pods). (3) Collection expeditions for the preselected material were then organized. Semiripe cuttings, pods (when possible), and leaves were collected for grafting, seedling, and molecular analysis, respectively. Meta-data regarding morphology, sanitary conditions, growth environment, and collection coordinates were registered for each accession. (4) The collected material was propagated at INIAP experimental stations and in partner agricultural high schools located near the collection sites in the localities of El Pangui, San José, Santiago, Taisha, and Kapawi (Figure S1) for agronomical monitoring and inclusion in future participatory evaluation programs. Photograph: (a), (b), (e), (f), (g), (h), and (i) Olivier Fouet and (3c), (d), and (j) Claire Lanaud

Criollo (0.588) and Amelonado (0.753) groups. The F_{IS} values ranged from 0.093 (Zone 3) to 0.309 (Zone 7) across the different collection zones and were thus closer to Hardy–Weinberg equilibrium than some of the control groups.

Differentiation between groups was calculated between pairs of populations using standardized variance (F_{ST}) and showed very significant differentiation ($F_{ST} > 0.25$) between control groups (Dataset S2). The lowest differentiation between control groups was observed between the Iquitos–Amelonado ($F_{ST} = 0.3678$), Iquitos–Nanay ($F_{ST} = 0.3577$), and Iquitos–Marañón ($F_{ST} = 0.3825$) groups, which aligned with the proximity of these groups. Moderate differentiation ($0.15 < F_{ST} < 0.25$) was mainly found between the eight different collection zones, and the lowest values were observed between the zones that were closest geographically (Zones 6 and 7 or Zones 3 and 4). Some of the collection zones showed moderate to low differentiation from the control groups, including Zones 1–3 with the Nacional control group (0.1330, 0.1314, and 0.1517, respectively) and Zones 7 and 8 with the Curaray control group (0.2352 and 0.1984, respectively). Zones 6 and 7, located upstream of the Morona River, showed similarities with the Morona control group (0.1918 and 0.2284, respectively). More surprisingly, Zones 6 and 7 also showed similarities with the Iquitos control group (0.1906 and 0.1771, respectively).

3.2 | Genetic distances between reference groups and collection zones

The genetic distance coefficient (D) was calculated in pairs between the control groups and the eight collection zones (Dataset S2). The lowest genetic distance between the control groups was detected between the Marañón and Guiana groups ($D = 0.274$), followed by the Nacional and Morona groups ($D = 0.295$). The highest genetic distance value was found between the Criollo and Amelonado groups ($D = 2.295$). The genetic distances between the collected zones fluctuated from $D = 0.05$ (between Zones 3 and 4) to $D = 0.482$ (between Zones 1 and 5). Small genetic distances were observed between the Nacional control group and the collected regions located in the south (Zones 1–3), especially with Zone 1 ($D = 0.068$). The Curaray group is closer to the native trees collected in the regions located in the north (Zones 5, 7, and 8).

3.3 | Structure of the population

The collected population was analyzed by two methods to identify its structure.

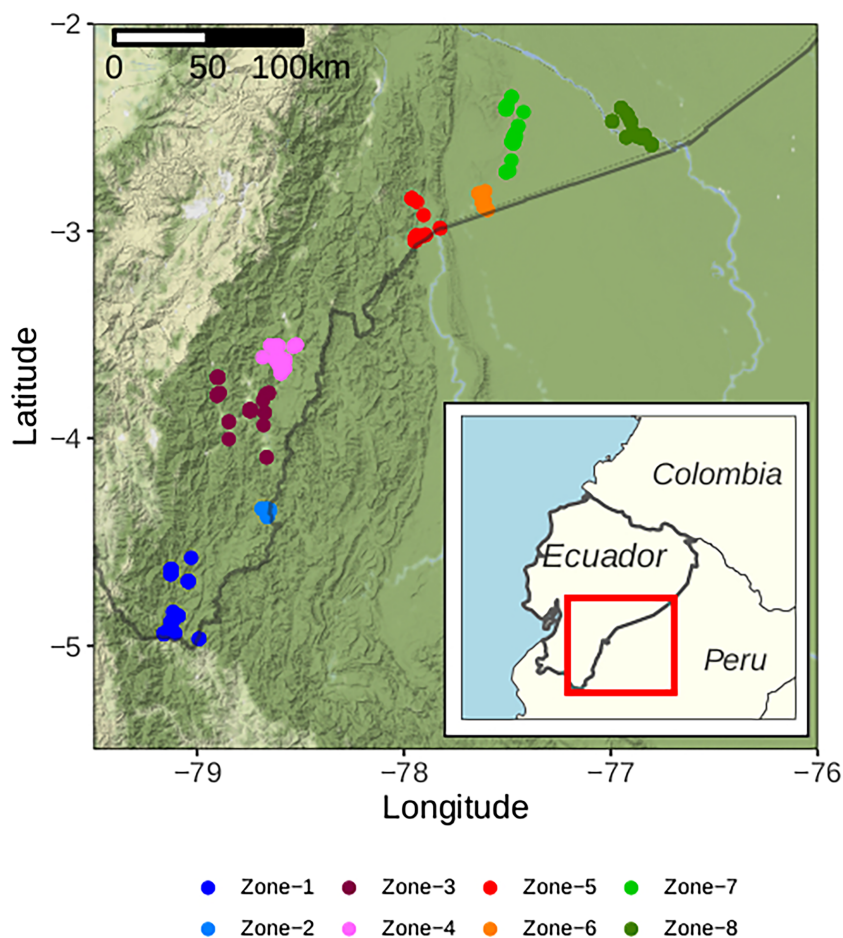


FIGURE 3 Geographic location of the collected accessions of *Theobroma cacao* in the eight zones surveyed in the Ecuadorian Amazon during the four collection expeditions in 2010, 2013, 2017, and 2019

An initial analysis by the neighbor-joining method was performed between the 135 control clones. It was observed that the resulting phylogenetic tree confirmed the structure of the 10 genetic groups known thus far and described by Motamayor et al. (2008). Furthermore, two supplementary groups could be observed and include on the one hand the accessions of from Morona and on the other hand the accessions from Caquetá. In this phylogenetic tree, it was also observed that the Criollo genetic group appears closest to the Caquetá genetic group.

The same analysis was then conducted with the total dataset of 418 individuals. The overall phylogenetic tree (Figure 4) allowed us to observe that the majority (>70%) of the collected trees were grouped into three large clusters close to the Nacional, Morona, and Curaray control groups. The first cluster comprised accessions collected in the southern zones (Zones 1–3) and accessions from the Nacional and Morona control groups. The second cluster comprised accessions collected in the northern zones (Zones 5–8) and accessions from the Curaray control group. The third cluster comprised accessions collected in Zones 3 and 4. Most of the remaining collected accessions (<30%) were irregularly located closer to other control groups excepted for complex admixed accessions located at the base of the tree that involved more than three genetic groups defined by control populations.

A Bayesian analysis using STRUCTURE software was then used to more precisely determine the ancestry and genetic partition of the collected and control accessions. The top hierarchical level of genetic partitioning was first revealed at $K = 4$ (Figures S2 and S3). This level of hierarchy yielded four major subdivisions of the control groups: (i) Curaray, (ii) Nacional and Morona, (iii) Criollo, and (iv) Amelonado-Nanay-Marañon-IQUITOS-Guiana-Purús and Contamana. At this value of K , the Caquetá control group appeared as an admixed population of the four subdivisions. Further subdivisions could be observed at higher K values, differentiating the following control groups: accessions with Contamana ancestry separated at $K = 7$; accessions with Morona ancestry separated at $K = 8$; and accessions with Amelonado ancestry separated at $K = 11$. The Purús and Caquetá control groups dissociated at $K = 14$. Finally, we did not observe a dissociation between accessions with Nanay and IQUITOS ancestries from those with Marañon and Guiana ancestries, even at higher values of K .

More generally, for the different values of K taken into consideration, the genetic structure of the newly collected accessions varied distinctly from south to north depending on the geographical zone of the collection site. In Zones 1 and 2, the accessions are very close to the Nacional control group, and further north, particularly in Zone 8, the collected trees are closer to the Curaray genetic group (Figure 5). Surprisingly, starting from $K = 5$, accessions collected

TABLE 1 Geographic origin of the *Theobroma cacao* accessions analyzed in the study

(a) Collected populations					
Area	Sample size	Country	Collection year	Geographical origin	Mean altitude in meters (min–max)
Zone 1	49	Ecuador	2013	Palanda and Chinchipe cantons	1,031.5 (724–1,188)
Zone 2	11	Ecuador	2010	Upper Nangaritza, Numpatakaime and Shamatak rivers	918.4 (802–962)
Zone 3	48	Ecuador	2010	Yacuambi, Zamora and lower Nangaritza rivers	932.3 (815–1,085)
Zone 4	60	Ecuador	2010 and 2017	El Pangui canton	934.2 (757–1,171)
Zone 5	22	Ecuador	2017	Santiago river and Yaupi locality	295.0 (245–357)
Zone 6	20	Ecuador	2017	Upper Morona River, San Jose de Morona locality	210.2 (201–223)
Zone 7	34	Ecuador	2019	Panguietza river, Taisha, Tutin Entsa localities	319.3 (210–452)
Zone 8	39	Ecuador	2019	Capahuari and Pastaza rivers	254.8 (224–273)
(b) Control populations					
Group name	Sample size	Country of origin	Geographical origin		
Amelonado	4	Brazil	NA		
Nanay	13	Peru	Nanay river		
Iquitos	12	Peru	Iquitos		
Marañon	10	Peru	Marañon river		
Guiana	11	French Guiana	Camopi and Tanpok rivers		
Purús	9	Brazil	NA		
Contamana	7	Peru	Loreto, Ucayali river		
Morona	7	Peru	Morona		
Nacional	11	Ecuador	Coast provinces		
Curaray	18	Ecuador	Morona-Santiago province		
Caquetá	11	Colombia/Ecuador	Caquetá river/Pastaza and Morona-Santiago province		
Criollo	22	Central América	NA		

Abbreviation: NA, not available.

mostly in Zone 4 and nearby El Pangui formed a distinct genetic group.

4 | DISCUSSION

A total of 283 native trees were collected and characterized with 48 SSR markers to assess their diversity and compare those values to the known genetic diversity of *T. cacao* species. SSRs were chosen for their flexibility in testing small numbers of individuals and for their discriminatory power compared to an equal number of SNP markers (van Inghelandt et al., 2010).

4.1 | High allelic diversity

Our results show that cocoa trees with new alleles and increased diversity were collected during these four surveys, highlighting this hotspot of diversity present in Ecuador. The high number of alleles per locus (N_a), up to 19 (average $N_a = 7.69$), allowed us to assign specific alleles to a genetic group or collection zone. A majority of cocoa

trees collected near the banks of the Morona River upstream of the village of San José de Morona (Zone 6) and further north (Zones 7 and 8) were characterized as being very close to the Curaray genetic group. These Curaray-type accessions were collected at low altitudes (between 200 and 350 m) under very different climatic conditions (cloud cover, high rainfall, and temperature). In contrast, the Nacional-type accessions observed in Zone 1 were collected between 800 and 1,200 m of altitude in mountainous regions (cordilleras). The Curaray and Nacional accessions are genetically very different ($D = 0.695$) and represent two of the four groups constituting the upper hierarchical structure observed in this study. Their presence reflects the high diversity observed over a short geographic distance between the most distant collection areas (approximately 300 km in a straight line). Adaptation to the strong geo-climatic variations of their environments could partially explain their genetic distance. The private alleles of all the collected native trees were found to be very large ($PA = 49$) compared to the control set, although they were well represented by regional groups (Curaray, Nacional, Morona, and Caquetá). Compared to the total alleles present in the control population (320), the private alleles present in the collected populations represent a gain of 15.3% of new alleles.

Group name	N_a	PA	$A_r(g)$	PA_r	H_o	H_e	F_{IS} (SE)
Amelonado	1.48	0	1.48	0.01	0.038	0.168	0.753 (0.059)
Nanay	2.23	4	1.75	0.06	0.234	0.226	-0.053 (0.037)
Marañon	1.88	1	1.69	0.04	0.265	0.236	-0.122 (0.035)
Iquitos	2.25	1	2.23	0.03	0.493	0.347	-0.360 (0.033)
Guiana	1.33	2	1.60	0.06	0.079	0.111	0.217 (0.052)
Contamana	2.08	6	1.95	0.20	0.350	0.314	-0.109 (0.065)
Morona	2.13	4	1.97	0.10	0.328	0.326	-0.022 (0.053)
Nacional	1.79	1	1.53	0.01	0.140	0.177	0.179 (0.060)
Curaray	2.35	1	2.16	0.02	0.173	0.224	0.192 (0.043)
Caquetá	3.19	15	2.45	0.27	0.286	0.410	0.267 (0.042)
Purús	2.17	14	1.93	0.29	0.318	0.307	-0.038 (0.048)
Criollo	1.25	5	1.16	0.18	0.019	0.054	0.588 (0.075)
Zone 1	3.83	2	2.02	0.04	0.253	0.293	0.102 (0.028)
Zone 2	2.92	3	2.35	0.07	0.308	0.388	0.198 (0.053)
Zone 3	3.79	1	2.23	0.02	0.366	0.401	0.093 (0.026)
Zone 4	4.40	1	2.45	0.02	0.397	0.451	0.133 (0.027)
Zone 5	3.67	6	2.66	0.09	0.368	0.507	0.261 (0.035)
Zone 6	4.31	3	3.00	0.03	0.472	0.571	0.166 (0.032)
Zone 7	4.56	4	2.98	0.04	0.388	0.579	0.309 (0.029)
Zone 8	4.21	6	2.39	0.04	0.320	0.406	0.190 (0.034)
Controls ×135	6.69	75	6.65	1.68	0.215	0.648	0.683 (0.015)
Collections ×283	6.15	49	5.82	0.85	0.355	0.535	0.324 (0.018)
Total ×418	7.69	369	NA	NA	0.311	0.603	0.490 (0.013)

Note: Mean number of alleles per locus (N_a), private alleles number (PA), allelic richness ($A_r(g)$), private allelic richness (PA_r), observed heterozygosity (H_o), expected heterozygosity (H_e), and fixation index (F_{IS}) with standard error (SE).

Abbreviation: NA, not available.

Different authors have examined the centers of diversity of *T. cacao* and hypothesized that its zone of origin is located near the Equator (Cheesman, 1944; Motamayor et al., 2008; Pound, 1938, 1945). Taken together, our results confirmed that Ecuador contains a high diversity hotspot for *T. cacao* species, as also previously suggested by Thomas et al. (2012).

4.2 | Population structure

Population structure analyses revealed four main subdivisions by both NJ and Bayesian clustering methods; this structure differentiated the Curaray, Nacional-Morona, and Criollo groups and a cluster containing the other groups formerly called “Forasteros,” including Amelonado, Purús, Guiana, and groups originating from Peru (Nanay, Marañon, Iquitos, and Contamana). This four-group hierarchy was already highlighted in the NJ method analysis performed on the LCT-EEN accessions (Loor et al., 2012) collected in the Amazon between 1980 and 1985 (Allen, 1988). A similar low level of hierarchy ($K = 4$) was also detected for a large collection of *T. cacao* in Colombia (Osorio-Guarín et al., 2017). Furthermore, at a K value of ≥ 10 , it was not possible to separate all 10 of the genetic groups previously

TABLE 2 Genetic diversity parameters revealed by simple sequence repeat (SSR) markers for each collection area and control groups of *Theobroma cacao*

described by Motamayor et al. (2008), suggesting that this reference analysis is not strong enough to be applied to our data even if the control samples had an ancestry value of >0.90 . A possible explanation could be that the number of markers used in our study was not sufficient to discriminate the genetic groups highlighted by Motamayor et al. (2008). However, other genetic groups dissociated before these 10 reference groups in our study. A large number of cocoa trees collected mostly around the village of El Pangui (north of the Zamora-Chinchipe region), mainly located in Zone 4, formed a new group that dissociated at $K = 5$. This new group was formed with an equivalent proportion of Curaray and Nacional ancestry representing the majority of Zone 4 accessions at $K = 4$ (Figure 4). The low number of private alleles ($PA = 1$) found in this Zone 4 and the relatively high observed heterozygosity ($H_o = 0.397$) suggest that the Pangui population seemed to be an admixed population rather than a distinct genetic group. Furthermore, at $K = 8$, accessions from the Morona control group dissociated from accessions of the Nacional control group. In addition, the analysis of the genetic substructure within the neighbor-joining tree separated individuals with Nacional ancestry from individuals with Morona ancestry, suggesting that the Morona control group could be considered a different subpopulation.

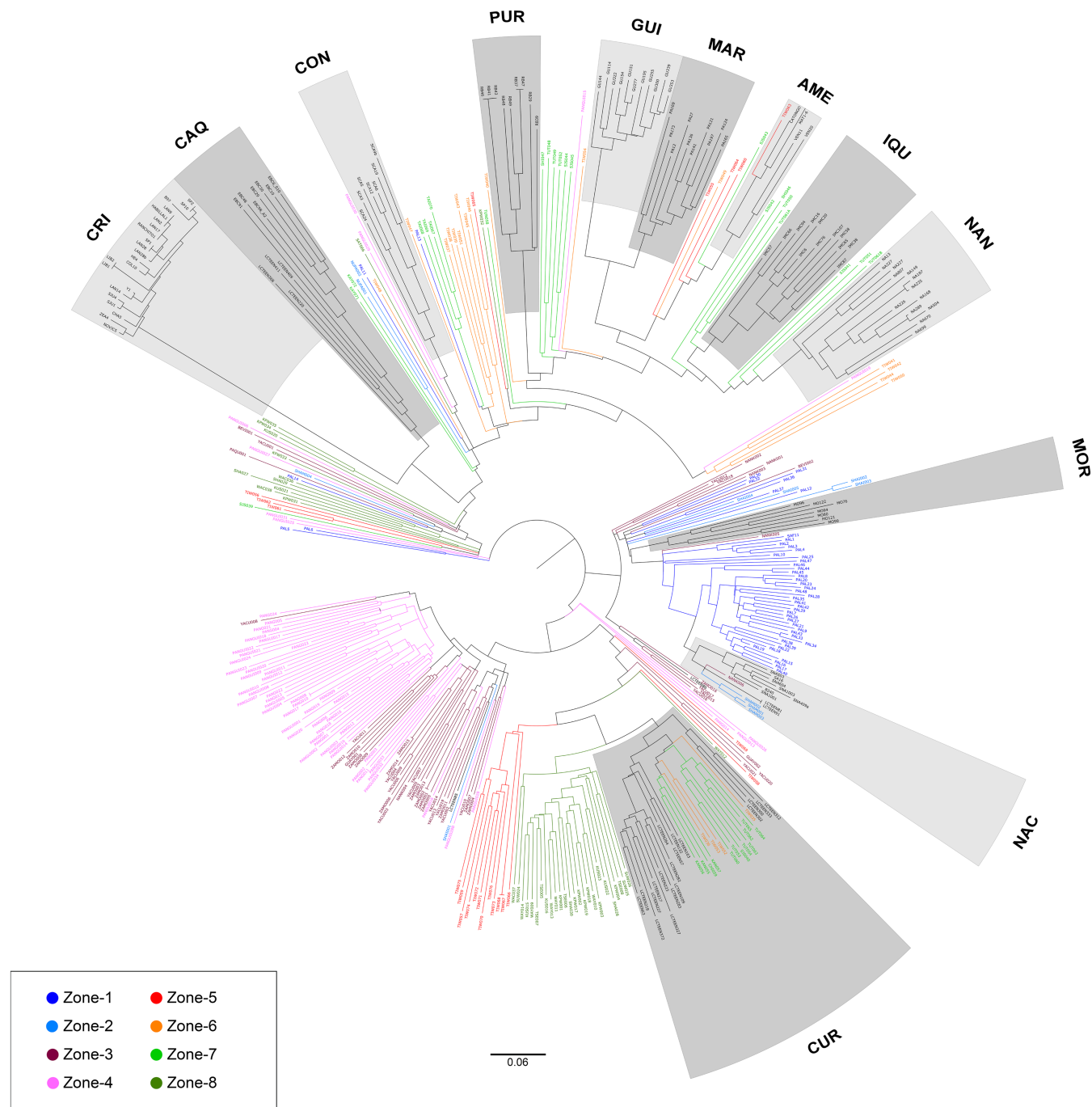


FIGURE 4 Genetic relationships between the collected accessions of *Theobroma cacao* in the eight zones (colors) and the control accessions (black). Dissimilarities between accessions were calculated with DARWIN software using 500 bootstrap replicates, and the unrooted phylogenetic tree was edited and visualized using Figtree software (<http://tree.bio.ed.ac.uk/software/figtree/>). (CRI, Criollo; CAQ, Caquetá; CON, Contamana; PUR, Purús; GUI, Guiana; MAR, Marañon; AME, Amelonado; IQU, Iquitos; NAN, Nanay; MOR, Morona; NAC, Nacional; CUR, Curaray)

Approximately 30% of the cocoa trees collected were not part of the three main clusters around the Nacional–Morona, Curaray, or the accessions collected around El Pangui village and were sometimes very close to the other control groups. The majority of these accessions were collected around villages well served by road connections; they are certainly foreign cocoa trees recently imported like TIW063 (Amelonado) or hybrids between native and foreign cocoa trees.

4.3 | Origin of the Nacional variety

A previous study using paternity analysis performed on a large population of *T. cacao* accessions collected in Amazonia from several countries (Loor et al., 2012) suggested that the genotypes LCT-EEN85, LCT-EEN86, and LCT-EEN91, which were collected in Ecuador close to the Yacuambi and Nangaritza rivers (Allen, 1988),

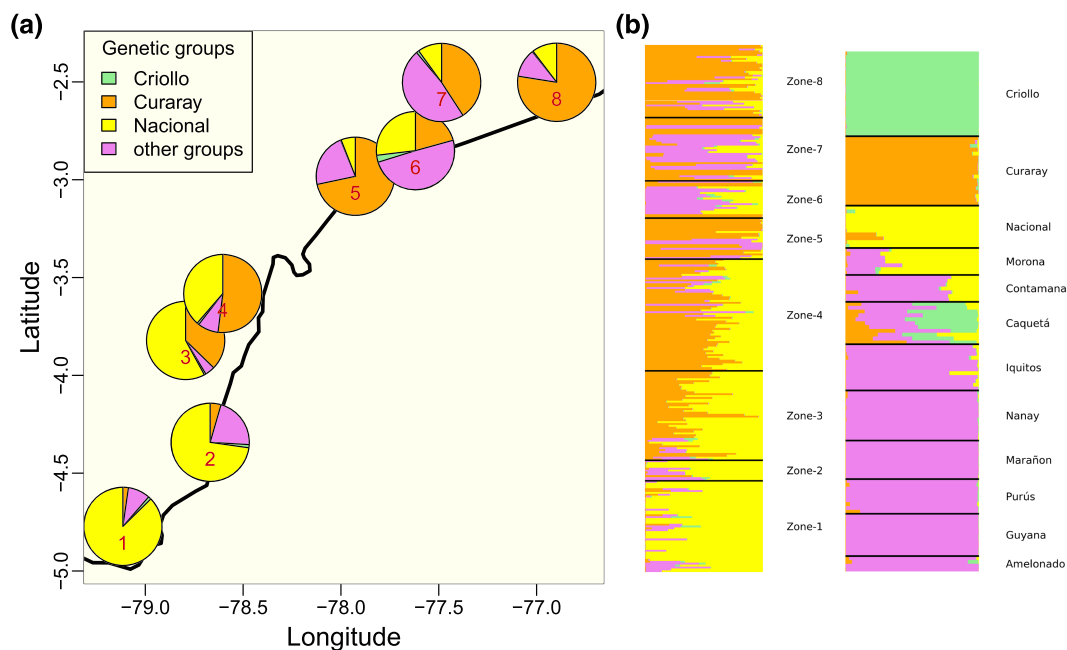


FIGURE 5 Population structure of the 283 collected accessions of *Theobroma cacao* analyzed with simple sequence repeats (SSRs) markers with $K = 4$ groups. (a) Map of the mean ancestry coefficients within the eight collected zones. Numbers in red correspond to the corresponding zones. (b) Population structure barplot of the collected accessions (left panel) and the accessions of the control population. Accessions are shown as horizontal bar colored in proportion to their ancestry within each group. Colors on the bar plot of the collected accessions refer to the control populations

were the closest and most likely representatives of the ancestors of the Nacional variety among the accessions studied. In the vicinity of these rivers (Zone 3) and around the village of Shaïme (Zone 2), we found 19 trees similar to the Nacional variety, including four that were very close genetically to LCTEEN86 and four trees (SHAM001, SHAM002, SHAM003, and NANK006) that were very close to LCTEEN91 and the Nacional control group. However, the vast majority of *T. cacao* accessions close to Nacional were collected further south around the village of Palanda (Zone 1). Over a distance of approximately 50 km around the village of Palanda, 44 of the 49 trees analyzed showed a very strong similarity to each other and to the Nacional control accessions.

An old ceremonial archeological site dating back 5,500 years was discovered in Santa Ana-La Florida near the village of Palanda; this site showed a high degree of societal development at this time (Valdez et al., 2005). A recent study showed the presence of starch grains, theobromine, which is characteristic of *T. cacao*, and ancient cocoa DNA on ceramic remains from this archeological site (Zarrillo et al., 2018). Trees collected around or at the site were very similar to the Nacional group controls. All of these data suggest that the Mayo-Chinchi people could have played a major role in the domestication of the Nacional variety.

Zones 1 and 2, where trees with Nacional ancestry have also been found, are not located in the same watershed but are connected by ancient paths still used by Amerindian communities. Evidence of very old interzone connections across this region (Valdez, 2008) may have facilitated the spread of the Nacional variety from its domestication area (Zone 1) to more northerly regions (Zones 2 and 3).

The hypothesis that the ancestral Nacional trees, once grown on the Pacific coast (Loor et al., 2009), could have originated from the region of Palanda (Zone 1) is supported by different geographical and historical data. Indeed, the Amazon region (Zones 1–3) is close to the Guayas region, where the first Nacional crops were observed on the Pacific coast. Marine shells discovered at the archeological site of Santa Ana-La Florida (Valdez et al., 2005) attest to the connections between Zone 1 and the Pacific coastal regions.

As a result of these surveys, this work has allowed the preservation and collection of genetic material close to the Nacional control group, known for its aromatic qualities and of importance for Ecuadorian cocoa production. The aromatic potential of the collected accessions is currently being analyzed (Colonges, Loor Solorzano, et al., 2022; Colonges, Seguine, et al., 2022). The enrichment of the allelic diversity of the Nacional varieties should provide, within the framework of improvement programs, to have material adapted to the local growing conditions found in the Ecuadorian Amazonian regions surveyed. Participatory breeding programs have been initiated with local communities to promote sustainable family cocoa farming and increase their livelihood.

4.4 | Hypothesis of Criollo domestication

A recent study showed the basal position of the Curaray genetic group in the phylogeny of *T. cacao* (Osorio-Guarín et al., 2017). This result was reinforced by an analysis based on a population

differentiation model derived from the sequencing of 200 individuals representing highly diverse cocoa trees (Cornejo et al., 2017). This latter analysis revealed the Curaray group as the most closely related to the Criollo group, suggesting that the origin of Criollo domestication came from a subset of ancient Curaray individuals.

With the analysis of representatives of a new population, the Caquetá population, our results showed that Criollo is closer to this new group ($D = 0.835$), which could also be visualized in the phylogenetic tree (Figure 4). Among the control groups, the Caquetá group had the highest values of allele richness and private alleles. These observations suggest that the Criollo domestication could have originated from this region or a region located further north, close to the frontier between Colombia and Ecuador, thus representing a hotspot of *T. cacao* diversity.

5 | CONCLUSION

T. cacao diversity is strongly threatened by progressive deforestation and by the rapid emergence of diseases in the context of trade globalization. The four surveys (2010 to 2019) carried out in Ecuadorian Amazonia with the help of Amerindian communities and agricultural colleges constitute enhanced genetic resources that can be used to create new aromatic cocoa varieties adapted to Amazonia. The vast majority of the collected accessions increased the previously known diversity of the Nacional and Curaray genetic groups, and new alleles were discovered. Given their diversity, they could allow the identification of aromatic niches, providing new income for farmers while contributing to the maintenance of *T. cacao* diversity.

This research has clarified the domestication area of the Nacional varieties and gives new clues about the domestication of Criollo, another aromatic variety. In the coming years, safeguarding the diversity of native *T. cacao* should play a major role in varietal improvement for the adaptation of the species, especially regarding climate change.

ACKNOWLEDGMENTS

These surveys were carried out thanks to a collaboration between INIAP and CIRAD. We thank the Agropolis Foundation, MUSE (Montpellier Université d'Excellence), and Valrhona for their financial support of this project. This work, a part of the MUSE Amazcacao project, was publicly funded through ANR (the French National Research Agency) under the "Investissement d'avenir" program with reference (Agence Nationale de la Recherche) ANR-16-IDEX-0006. We are most grateful to the GPTR (Great regional technical platform) of Montpellier core facility for its technical support. We are very grateful to the local populations, mainly the Shuar, Achuar, and Amazonian Kichwa Amerindian communities, as well as to the agricultural colleges of El Pangui, Real Audiencia (San Jose), Raime Roldos (Santiago), Los Angeles (Taisha), and Tuna (Kapawi) for their invaluable help during these surveys.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

R.G.L.S., P.C. and C.L. conceived and coordinated the project; R.G.L.S., C.S., and F.F. established formalities with government authorities and Amerindian communities to enter in the Indian territories; R.G.L.S., C.S., D.C., F.F., and C.L. organized seminars in agricultural colleges; O.F., R.G.L.S., C.S., D.C., F.F., F.A. B.Y., and I.S. carried out the 2017 and 2019 surveys, multiplied the collected material, and established germplasm collections; O.F., R.G.L.S., and C.L. designed the genetic studies; O.F., R.R., K.C., and H.V. carried out the genotyping; O.F., B.R., and X.A. analyzed the data; O.F., R.G.L.S., B.R., X.A., and C.L. wrote the manuscript.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon request.

ORCID

Olivier Fouet  <https://orcid.org/0000-0001-8547-1474>

Rey Gaston Loor Solorzano  <https://orcid.org/0000-0003-3984-8239>

Bénédicte Rhoné  <https://orcid.org/0000-0002-4198-622X>

Kelly Colonges  <https://orcid.org/0000-0001-9944-1399>

Hélène Vignes  <https://orcid.org/0000-0002-8725-2467>

Xavier Argout  <https://orcid.org/0000-0002-0100-5511>

Claire Lanaud  <https://orcid.org/0000-0001-6411-7310>

REFERENCES

- Allegre, M., Argout, X., Boccara, M., Fouet, O., Roguet, Y., Bérard, A., Thévenin, J. M., Chauveau, A., Rivallan, R., & Clement, D. (2011). Discovery and mapping of a new expressed sequence tag-single nucleotide polymorphism and simple sequence repeat panel for large-scale genetic studies and breeding of *Theobroma cacao* L. *DNA Research*, 19, 23–35. <https://doi.org/10.1093/dnares/dsr039>
- Allen, J. B. (1988). Geographical variation and population biology in wild *Theobroma cacao*. Ph.D. thesis, University of Edinburgh.
- Allen, J. B., & Lass, A. R. (1983). Final report phase 1. *Cocoa Growers' Bulletin*, 34, 1–72.
- Arevalo-Gardini, E., Meinhardt, L. W., Zuñiga, L. C., Arevalo-Gardini, J., Motilal, L., & Zhang, D. (2019). Genetic identity and origin of "Piura Porcelana"—A fine-flavored traditional variety of cacao (*Theobroma cacao*) from the Peruvian Amazon. *Tree Genetics & Genomes*, 15, 11. <https://doi.org/10.1007/s11295-019-1316-y>
- Belkhir, K., Borsa, P., Chikhi, L., Raufaste, N., Bonhomme, F. (2004). GENETIX 4.05, Windows TM software for population genetics. Laboratoire génome, populations, interactions, CNRS UMR 5000.
- Boza, E. J., Motamayor, J. C., Amores, F. M., Cedeño-Amador, S., Tondo, C. L., Livingstone, D. S., Schnell, R. J., & Gutiérrez, O. A. (2014). Genetic characterization of the cacao cultivar CCN 51: Its impact and significance on global cacao improvement and production. *Journal of the American Society for Horticultural Science*, 139, 219–229. <https://doi.org/10.21273/JASHS.139.2.219>
- Cheesman, E. E. (1944). Notes on the nomenclature, possible classification and relationships of cocoa populations. *Tropical Agriculture*, 21, 144–159.
- Colonges, K., Loor Solorzano, R. G., Jimenez, J. C., Lahon, M. C., Seguine, E., Calderon, D., Subia, C., Sotomayor, I., Fernández, F., Lebrun, M., Fouet, O., Rhoné, B., Argout, X., Costet, P., Lanaud, C., & Boulanger, R. (2022). Variability and genetic determinants of cocoa

- aromas in trees native to South Ecuadorian Amazonia. In press, accepted for publication. *Plants, People, Planet*.
- Colonges, K., Seguine, E., Saltos, A., Davrieux, F., Minier, J., Jimenez, J. C., Lahon, M. C., Calderon, D., Subia, C., Sotomayor, I., Fernández, F., Fouet, O., Rhoné, B., Argout, X., Lebrun, M., Costet, P., Lanaud, C., Boulanger, R., & Loor Solorzano, R. G. (2022). Diversity and determinants of bitterness, astringency and fat content in cultivated Nacional and native Amazonian *Theobroma cacao* accessions from Ecuador. In press, accepted for publication. *The Plant Journal*.
- Cornejo, O. E., Yee, M. C., Dominguez, V., Andrews, M., Sockell, A., Strandberg, E., Livingstone, D., Stack, C., Romero, A., Umaharan, P., Royaert, S., Tawari, N. R., Pauline, N., Schnell, R., Wilbert, P., Mockaitis, K., Bustamante, C., & Motamayor, J. C. (2017). Genomic insights into the domestication of the chocolate tree, *Theobroma cacao* L. *bioRxiv*. <https://doi.org/10.1101/223438>
- Cuatrecasas, J. (1964). Cacao and its allies. A taxonomic revision of the genus *Theobroma*. *Contributions to the United States National Herbarium*, 35(6), 379–614.
- el Mousadik, A., & Petit, R. J. (1996). High level of genetic differentiation for allelic richness among populations of the argan tree [*Argania spinosa* (L.) Skeels] endemic to Morocco. *Theoretical and Applied Genetics*, 92(7), 832–839. <https://doi.org/10.1007/BF00221895>
- Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, 14(8), 2611–2620. <https://doi.org/10.1111/j.1365-294X.2005.02553.x>
- Fouet, O., Allegre, M., Argout, X., Jeanneau, M., Lemainque, A., Pavek, S., Boland, A., Risterucci, A. M., Loor, G., Tahj, M., Sabau, X., Courtois, B., & Lanaud, C. (2011). Structural characterization and mapping of functional EST-SSR markers in *Theobroma cacao*. *Tree Genetics & Genomes*, 7(4), 799–817. <https://doi.org/10.1007/s11295-011-0375-5>
- Gomes, V. H. F., Vieira, I. C. G., Salomão, R. P., & Steege, H. (2019). Amazonian tree species threatened by deforestation and climate change. *Nature Climate Change*, 9(7), 547–553. <https://doi.org/10.1038/s41558-019-0500-2>
- ICCO. (2020). Quarterly Bulletin of Cocoa Statistics, Vol. XLVI, No. 4, Cocoa year 2019/20 Published: 30-11-2020. <https://www.icco.org/wp-content/uploads/Production-QBCS-XLVI-No-4.pdf>
- Jaimez, R. E., Barragan, L., Fernández-Niño, M., Wessjohann, L. A., Cedeño-García, G., Cantos, I. S., & Arteaga, F. (2022). *Theobroma cacao* L. cultivar CCN 51: A comprehensive review on origin, genetics, sensory properties, production dynamics, and physiological aspects. *PeerJ*, 10, e12676. <https://doi.org/10.7717/peerj.12676>
- Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, 23(14), 1801–1806. <https://doi.org/10.1093/bioinformatics/btm233>
- Kalinowski, S. T. (2004). Counting alleles with rarefaction: Private alleles and hierarchical sampling designs. *Conservation Genetics*, 5(4), 539–543. <https://doi.org/10.1023/B:COGE.0000041021.91777.1a>
- Kalinowski, S. T. (2005). Hp-rare 1.0: A computer program for performing rarefaction on measures of allelic richness. *Molecular Ecology Notes*, 5(1), 187–189. <https://doi.org/10.1111/j.1471-8286.2004.00845.x>
- Kline, K. L., Ramirez, L. F., Sum, C., Lopez-Ridaura, S., & Dale, V. H. (2020). Enhance indigenous agricultural systems to reduce migration. *Nature Sustainability*, 3, 74–76. <https://doi.org/10.1038/s41893-020-0473-1>
- Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak: A program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, 15(5), 1179–1191. <https://doi.org/10.1111/1755-0998.12387>
- Lachenaud, P., Fouet, O., Couturier, C., & Lanaud, C. (2016). Cacao trees (*Theobroma cacao* L.) of the “Guiana” genetic group: “Tanpok 2012” survey. *Genetic Resources and Crop Evolution*, 63, 781–790. <https://doi.org/10.1007/s10722-015-0280-0>
- Lachenaud, P., & Zhang, D. (2008). Genetic diversity and population structure in wild stands of cacao trees (*Theobroma cacao* L.) in French Guiana. *Annals of Forest Science*, 65, 310. <https://doi.org/10.1051/forest:2008011>
- Li, Y. L., & Liu, J. X. (2018). StructureSelector: A web-based software to select and visualize the optimal number of clusters using multiple methods. *Molecular Ecology Resources*, 18, 176–177. <https://doi.org/10.1111/1755-0998.12719>
- Loor, R. G., Fouet, O., Lemainque, A., Pavek, S., Boccara, M., Argout, X., Amores, F., Courtois, B., Risterucci, A. M., & Lanaud, C. (2012). Insight into the wild origin, migration and domestication history of the fine flavour Nacional *Theobroma cacao* L. variety from Ecuador. *PLoS ONE*, 7, e48438. <https://doi.org/10.1371/journal.pone.0048438>
- Loor, R. G., Lachenaud, P., Fouet, O., Argout, X., Peña, G., Macias, J. C., Amores Puyutaxi, F. M., Valdez, F., Hurtado, J., & Lanaud, C. (2015). Rescue of cacao genetic resources related to the nacional variety: Surveys in the Ecuadorian Amazon (2010–2013). *Revista ESPAMCiencia*, 6, 7–15.
- Loor, R. G., Risterucci, A. M., Courtois, B., Fouet, O., Jeanneau, M., Rosenquist, E., Amores, F., Vasco, A., Medina, M., & Lanaud, C. (2009). Tracing the native ancestors of the modern *Theobroma cacao* L. population in Ecuador. *Tree Genetics & Genomes*, 5(3), 421–433. <https://doi.org/10.1007/s11295-008-0196-3>
- Motamayor, J. C., Lachenaud, P., Mota, J. W., Loor, R., Kuhn, D. N., Brown, J. S., & Schnell, R. J. (2008). Geographic and genetic population differentiation of the Amazonian chocolate tree (*Theobroma cacao* L.). *PLoS ONE*, 3, e3311. <https://doi.org/10.1371/journal.pone.0003311>
- Nei, M. (1972). Genetic distance between populations. *The American Naturalist*, 106(949), 283–292. <https://doi.org/10.1086/282771>
- Nei, M. (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, 89(3), 583–590. <https://doi.org/10.1093/genetics/89.3.583>
- Nobre, C. A., Sampaio, G., Borma, L. S., Castilla-Rubio, J. C., Silva, J. S., & Cardoso, M. (2016). Land-use and climate change risks in the Amazon and the need of a novel sustainable development paradigm. *Proceedings of the National Academy of Sciences*, 113(39), 10759–10768. <https://doi.org/10.1073/pnas.1605516113>
- Osorio-Guarín, J. A., Berdugo-Cely, J., Coronado, R. A., Zapata, Y. P., Quintero, C., Gallego-Sánchez, G., & Yockteng, R. (2017). Colombia a source of cacao genetic diversity as revealed by the population structure analysis of germplasm Bank of *Theobroma cacao* L. *Frontiers in Plant Science*, 8. <https://doi.org/10.3389/fpls.2017.01994>
- Peakall, R., & Smouse, P. E. (2012). GenAlEx 6.5: Genetic analysis in excel. Population genetic software for teaching and research—An update. *Bioinformatics (Oxford, England)*, 28, 2537–2539. <https://doi.org/10.1093/bioinformatics/bts460>
- Perrier, X., Flori, A., & Bonnot, F. (2003). Methods of data analysis. *Genetic Diversity of Cultivated Tropical Plants*, 33, 63.
- Pittier, H. (1930). A propos des Cacaoyers spontanés. *Journal Dagriculture Traditionnelle et de Botanique appliquée*, 10, 777–781. <https://doi.org/10.3406/jatba.1930.4906>
- Pound, F. J. (1938). Cacao and witchbroom disease (*Marasmius perniciosus*) of South America. With notes on other species of *Theobroma*. Report by Dr. F. J. Pound on a visit to Ecuador, the Amazon Valley, and Colombia. April 1937–April 1938.
- Pound, F. J. (1945). A note on the cocoa population of South America. In: Report and Proceedings of the 1945 Cocoa Conference. London. 131–133.
- Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155, 945–959. <https://doi.org/10.1093/genetics/155.2.945>
- Pugh, T., Fouet, O., Risterucci, A. M., Brottier, P., Abouladze, M., Deletrez, C., Courtois, B., Clement, D., Larmande, P., & NGoran, J. A. (2004). A new cacao linkage map based on codominant markers: Development and integration of 201 new microsatellite markers.

- Theoretical and Applied Genetics*, 108(6), 1151–1161. <https://doi.org/10.1007/s00122-003-1533-4>
- Raymond, M., & Rousset, F. (1995). GENEPOP (version 1.2): Population genetic software for exact tests and ecumenicism. *Journal of Heredity*, 86, 248–249. <https://doi.org/10.1093/oxfordjournals.jhered.a111573>
- Risterucci, A. M., Grivet, L., NGoran, J. A. K., Pieretti, I., Flament, M. H., & Lanaud, C. (2000). A high-density linkage map of *Theobroma cacao* L. *Theoretical and Applied Genetics*, 101(5–6), 948–955. <https://doi.org/10.1007/s001220051566>
- Rosenberg, N. A. (2004). DISTRUCT: A program for the graphical display of population structure. *Molecular Ecology Resources*, 4, 137–138. <https://doi.org/10.1007/s001220051566>
- Rousset, F. (2008). genepop'007: A complete re-implementation of the genepop software for windows and Linux. *Molecular Ecology Resources*, 8, 103–106. <https://doi.org/10.1111/j.1471-8286.2007.01931.x>
- Satori, D., Tovar, C., Faruk, A., Hammond Hunt, E., Muller, G., Cockel, C., Kühn, N., Leitch, I. J., Lulekal, E., Pereira, L., Ryan, P., Willis, K., & Pironon, S. (2021). Prioritising crop wild relatives to enhance agricultural resilience in sub-Saharan Africa under climate change. *Plants, People, Planet*, 4, 269–282. <https://doi.org/10.1002/ppp3.10247>
- Sereno, M. L., Albuquerque, P. S. B., Vencovsky, R., & Figueira, A. (2006). Genetic diversity and natural population structure of cacao (*Theobroma cacao* L.) from the Brazilian Amazon evaluated by microsatellite markers. *Conservation Genetics*, 7, 13–24. <https://doi.org/10.1007/s10592-005-7568-0>
- Thomas, E., Zonneveld, M., Loo, J., Hodgkin, T., Galluzzi, G., & Etten, J. (2012). Present spatial diversity patterns of *Theobroma cacao* L. in the Neotropics reflect genetic differentiation in Pleistocene Refugia followed by human-influenced dispersal. *PLoS ONE*, 7, e47676. <https://doi.org/10.1371/journal.pone.0047676>
- Valdez, F. (2008). Mayo Chinchipe. La nouvelle frontière. *Les Nouvelles de l'archéologie*, (111-112), 53–58. <https://doi.org/10.4000/nda.288>
- Valdez, F., Guffroy, J., de Saulieu, G., Hurtado, J., & Yepes, A. (2005). Découverte dun site cérémoniel formatif sur le versant oriental des Andes. *Comptes Rendus Palevol*, 4, 369–374. <https://doi.org/10.1016/j.crpv.2005.02.005>
- van Inghelandt, D., Melchinger, A. E., Lebreton, C., & Stich, B. (2010). Population structure and genetic diversity in a commercial maize breeding program assessed with SSR and SNP markers. *Theoretical and Applied Genetics*, 120(7), 1289–1299. <https://doi.org/10.1007/s00122-009-1256-2>
- Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, 38, 1358–1370.
- Whitlock, B. A., Bayer, C., & Baum, D. A. (2001). Phylogenetic relationships and floral evolution of the Byttnerioideae (“Sterculiaceae” or Malvaceae s.l.) based on sequences of the chloroplast gene, *ndhF*. *Systematic Botany*, 26, 420–437. <https://doi.org/10.1043/0363-6445-26.2.420>
- Zarrillo, S., Gaikwad, N., Lanaud, C., Powis, T., Viot, C., Lesur, I., Fouet, O., Argout, X., Guichoux, E., Salin, F., Loor, R. G., Bouchez, O., Vignes, H., Severt, P., Hurtado, J., Yopez, A., Grivetti, L., Blake, M., & Valdez, F. (2018). The use and domestication of *Theobroma cacao* during the mid-Holocene in the upper Amazon. *Nature Ecology & Evolution*, 2(12), 1879–1888. <https://doi.org/10.1038/s41559-018-0697-x>
- Zhang, D., Arevalo-Gardini, E., Mischke, S., Zúñiga-Cernades, L., Barreto-Chavez, A., & Aguila, J. A. D. (2006). Genetic diversity and structure of managed and semi-natural populations of cocoa (*Theobroma cacao*) in the Huallaga and Ucayali valleys of Peru. *Annals of Botany*, 98(3), 647–655. <https://doi.org/10.1093/aob/mcl146>
- Zhang, D., Martínez, W. J., Johnson, E. S., Somarriba, E., Phillips-Mora, W., Astorga, C., Mischke, S., & Meinhardt, L. W. (2012). Genetic diversity and spatial structure in a new distinct *Theobroma cacao* L. population in Bolivia. *Genetic Resources and Crop Evolution*, 59, 239–252. <https://doi.org/10.1007/s10722-011-9680-y>

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Fouet, O., Loor Solorzano, R. G., Rhoné, B., Subía, C., Calderón, D., Fernández, F., Sotomayor, I., Rivallan, R., Colonges, K., Vignes, H., Angamarca, F., Yaguana, B., Costet, P., Argout, X., & Lanaud, C. (2022). Collection of native *Theobroma cacao* L. accessions from the Ecuadorian Amazon highlights a hotspot of cocoa diversity. *Plants, People, Planet*, 4(6), 605–617. <https://doi.org/10.1002/ppp3.10282>