MDPI

Article

Geographic Patterns of Genetic Variation among Cacao (*Theobroma cacao* L.) Populations Based on Chloroplast Markers

Helmuth Edisson Nieves-Orduña [1,2], Markus Müller [1], Konstantin V. Krutovsky [1,2,3,4] and Oliver Gailing [1,2,*]

1   Department of Forest Genetics and Forest Tree Breeding, Georg-August University of Göttingen, 37077 Göttingen, Germany; hnieves@gwdg.de (H.E.N.-O.); markus.mueller@forst.uni-goettingen.de (M.M.); konstantin.krutovsky@forst.uni-goettingen.de (K.V.K.)
2   Center for Integrated Breeding Research, Georg-August University of Göttingen, 37075 Göttingen, Germany
3   Laboratory of Forest Genomics, Genome Research and Education Center, Institute of Fundamental Biology and Biotechnology, Siberian Federal University, 660036 Krasnoyarsk, Russia
4   Laboratory of Population Genetics, N.I. Vavilov Institute of General Genetics, Russian Academy of Sciences, 119991 Moscow, Russia
*   Correspondence: ogailin@gwdg.de

Abstract: The cacao tree (*Theobroma cacao* L.) is native to the Amazon basin and widely cultivated in the tropics to produce seeds, the valuable raw material for the chocolate industry. Conservation of cacao genetic resources and their availability for breeding and production programs are vital for securing cacao supply. However, relatively little is still known about the phylogeographic structure of natural cacao populations. We studied the geographic distribution of cpDNA variation in different populations representing natural cacao stands, cacao farms in Ecuador, and breeding populations. We used six earlier published cacao chloroplast microsatellite markers to genotype 233 cacao samples. In total, 23 chloroplast haplotypes were identified. The highest variation of haplotypes was observed in western Amazonia including geographically restricted haplotypes. Two observed haplotypes were widespread across the Amazon basin suggesting long distance seed dispersal from west to east in Amazonia. Most cacao genetic groups identified earlier using nuclear SSRs are associated with specific chloroplast haplotypes. A single haplotype was common in selections representing cacao plantations in west Ecuador and reference Trinitario accessions. Our results can be used to determine the chloroplast diversity of accessions and in combination with phenotypic assessments can help to select geographically distinctive varieties for cacao breeding programs.

Keywords: cacao; chloroplast haplotypes; geographic origin; chocolate; crop dispersal; SSR; microsatellite markers

1. Introduction

Cacao trees (*Theobroma cacao* L.) are native to the Amazon basin, and their valuable seeds are the raw material for the chocolate industry [1,2]. Cultivated by 5–6 million smallholder farmers in Latin America, Africa, and Asia, cacao crops are vitally important for local economies in these continents [3,4]. As an important tropical crop, there is a constant need to develop high-yielding and disease-resistant varieties [5–7]. Thus, conservation and use of cacao genetic diversity are essential not only for sustainable cultivation in producing countries, but also for diverse and growing needs of the chocolate industry and consumers [1,8–11].

Western Amazonia is considered to be a putative center of origin of cacao based on high phenotypic [1,12,13] and genetic [2,14,15] diversity. This part of the Amazon is also a region of high biodiversity [16–20] that in the past served as a forest refugium favoring accumulation of high tree diversity [21,22]. Western Amazonia is also considered a center

of crop domestication [23–26]. Examples of cultivated tree species, whose probable origin of domestication is western Amazonia, include fruit trees, such as the ice cream bean tree (*Inga edulis* Mart.), tree grape (*Pourouma cecropiifolia* Mart.), caimito (*Pouteria caimito* Radlk.), and the Amazon nut tree (*Bertholletia excelsa* Bonpl.) [27]. At least 29 crop species and 38 utilizable palms with probable origin in western Amazonia highlight the importance of the plant genetic resources observed in this part of the Amazon [23,28].

Cacao populations are organized in 10 genetic groups, some of which are called varieties (Amelonado, Contamana, Criollo, Curaray, Guianna, Iquitos, Marañon, Nacional, Nanay, and Purús), spread over the native range in South America [2]. This classification is based on the Bayesian cluster analysis of 96 nuclear SSR markers genotyped in 952 cacao accessions selected from ex situ collections [2].

The cultivated cacao was traditionally classified into Criollo and Forastero (an umbrella designation for the Amazonian populations) varieties and a Trinitario variety, a natural hybrid of Criollo × Forastero [1,12]. However, the contribution of some of the 10 cacao groups to cultivated cacao is relatively small, which is represented mainly by three varieties: Amelonado, Criollo, and Nacional [1,5,8,10,15]. However, during domestication, the Criollo variety accumulated a large proportion of high-frequency deleterious mutations that affect fitness [15], necessitating the use of more diverse material for cacao breeding [5,8,10,15].

One of the approaches to better understand genetic resources available for cacao breeding is to characterize historical patterns of seed dispersal and the origin of cultivated populations using chloroplast (cpDNA) markers, such as chloroplast microsatellites (cpSSRs) [29]. They are polymorphic markers of the chloroplast genome and typically maternally inherited in angiosperms, including cacao, and propagated via seeds. Therefore, they are used preferentially for phylogeographic analysis [30,31]. Nevertheless, a set of standard nuclear SSRs [32–34] and SNPs [35–38] are preferred for cacao identification in germplasm collections and breeding experiments.

Specifically, chloroplast markers are useful to understand geographic variation and dispersal of tropical trees [31,39–42]. For instance, these markers revealed long distance transfer of seeds for plantation establishment in *Dalbergia sissoo* Roxb. [43] and helped to select specific haplotypes to improve site adaptation in cultivated *Pinus armandii* Franch. [44]. In cacao, the historical transfer of reproductive material for cultivation has focused on genotypes exhibiting disease resistance and flavor traits [1].

Nine specific chloroplast microsatellites (cpSSR) are available for cacao genetic analysis [29]. In cacao farms and localities in Trinidad and Tobago, these markers allowed the identification of seven haplotypes among Trinitarian cacao populations [45]. Maternal lineages (seeds) used for cultivar development in Trinidad apparently originated from cacao populations of Central America, Peru, and Venezuela [45].

Ten haplotypes were identified by sequencing the chloroplast intergenic *trn*H-*psb*A spacer region of cultivated cacao trees in Soconusco, southern Mexico [46], where cacao cultivation has been recorded since colonial times [47]. Gutiérrez-López et al. [46] confirmed the introduction of a few mother trees as founder material for Soconusco plantations.

Although these studies did not include a wide range of samples within the natural distribution of cacao populations in the Amazon basin, they validated the use of chloroplast markers for managing cacao genetic resources. Here, we show how chloroplast genetic diversity is geographically distributed in Amazonia, and how it is represented in common cacao genotypes by using cpSSR markers. We hypothesize a center of cpDNA diversity in western Amazonia, considering the high genetic diversity observed in western cacao populations and the recognized origin of the species in this region. Our objectives were to: (i) study the geographic distribution of cpDNA variation in a sample of cacao trees collected in the Amazon basin, and in samples of cultivated cacao from west Ecuador, (ii) identify cpDNA variation in breeding populations, and (iii) determine the correspondence between cacao chloroplast and nuclear DNA variation.

2. Materials and Methods

2.1. Plant Material

Fresh leaves of 233 individual cacao trees were obtained for DNA extraction from the living collection maintained by the International Cocoa Quarantine Centre, Reading University, UK. Leaf samples of 154 cacao trees were selected based on widespread geographic origin from Amazon forests in five countries (Figure 1) (31 locations and 1–28 individuals per location, Table S1). This material corresponds to seeds or budwood collected during different expeditions in 1938–1988, aiming to obtain plant material carrying disease resistance traits and to study the species' ecology and variability [48–52]. Although these plant expeditions took place in Amazonian forests, some collection sites possibly experienced pre-Columbian silvicultural management like ancient cacao cultivation or translocation [26,53,54].
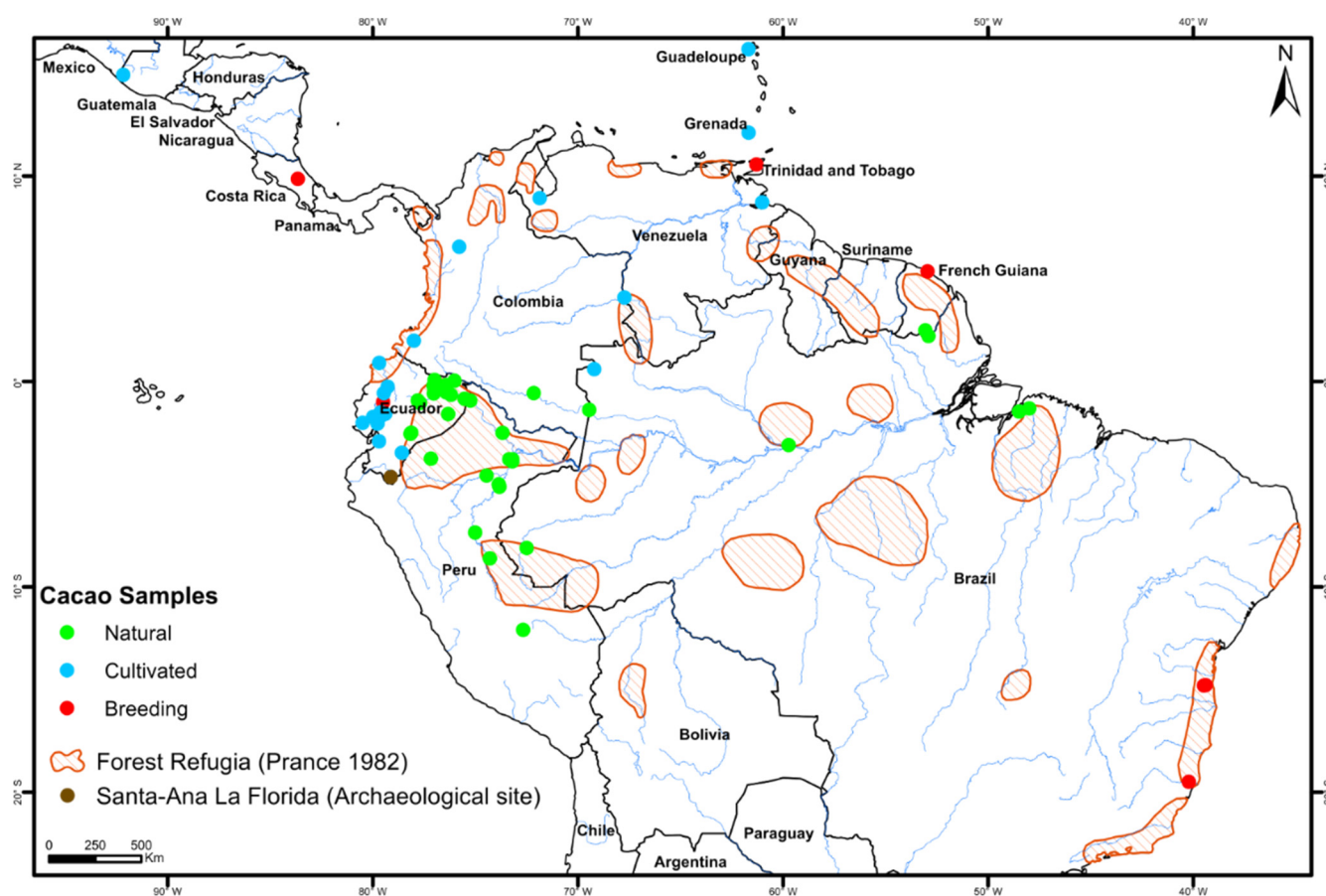

Figure 1. Geographic distribution of *Theobroma cacao* L. accessions in this study, forest refugial areas during the Pleistocene proposed by Prance (1982), and location of cacao use in Santa-Ana La Florida approximately 5300 years ago.

Samples of 30 individuals of cultivated cacao originated from nine different cacao farms in the coastal valley of Ecuador (1–6 individuals per location, Table S2, Figure 1). This group of germplasm is known as Refractario cacao and includes the progeny of approximately 80 different trees selected in 1937 for their resistance to witches' broom disease [48,52,55,56].

Additionally, leaf samples of 49 clones used for plantations and breeding programs were analyzed. They included the following genotypes: UF (United Fruit); ICS (Imperial Collection Selection); CRU (Cocoa Research Unit); TSH (Trinidad Selected); EET (Estación Experimental Tropical); and VB (Vassoura de Bruxa). The complete list of genotypes and their breeding origin in ten countries from Latin America and the Caribbean are presented in Supplementary Table S3.

Among the 233 samples analyzed, 83 different clones were also included in the analysis of [2] and represented different cacao genetic groups with different geographic origin: Amelonado (8), Contamana (6), Curaray (7), Guiana (2), Iquitos (23), Marañon (23), Nacional (2), and Nanay (12) (Tables S1 and S3).The International Cocoa Germplasm Database (ICGD) [52] provides details about agronomic traits, breeding programs, plant collection expeditions and geographic origin of the 234 accessions analyzed.

2.2. DNA Extraction

To extract DNA, we used about 1 cm^2 of leaf tissue and the DNeasy 96 Plant Kit (Qiagen, Hilden, Germany). DNA was diluted 1:10 before PCR amplification.

2.3. Chloroplast DNA Markers

We tested the amplification and diversity of ten universal chloroplast markers (*ccmps*) [57] in eight samples from distant regions. Additionally, we tested nine chloroplast microsatellite markers developed earlier for *T. cacao* (*CaCrSSRs*) based on the cacao chloroplast genome [29]. Monomorphic amplification products were obtained with all *ccmp* markers, but six out of the nine cacao chloroplast markers showed clear polymorphisms, which were used then to screen the 235 samples. Four of them (*CaCr2*, *CaCr4*, *CaCr5*, and *CaCr8*) represented mononucleotide, one (*CaCr1*)—pentanucleotide, and one—octonucleotide (*CaCr9*) repeats. Although it is difficult to genotype mononucleotide repeats, it is easier to do it for chloroplast SSRs than for nuclear SSRs because they are haploid and their fragment size can be easier determined due to the lack of interference with another allele such as in diploid nuclear SSRs. Different alleles were also verified by running respective samples side-by-side during the same electrophoretic run. When the allele calling was no clear, we did repetitions to ensure a correct allele scoring.

The following six polymorphic chloroplast DNA (cpDNA) markers were amplified by PCR in two multiplexes: (1) *CaCrSSR1*, *CaCrSSR2*, and *CaCrSSR4*, and (2) *CaCrSSR5*, *CaCrSSR8*, and *CaCrSSR9*. A M13 tail (5'-CACGACGTTGTAAAACGAC-3') and a PIG tail (5'-GTTTCTT-3') were attached to the 5' ends of forward and reverse primers, respectively [58].

The PCR reaction mix for each primer in 14 μL volume contained: 1 μL of genomic DNA (about 0.6 ng/μL), 5.7 μL ddH$_2$O; 1.5 μL PCR buffer (10× Buffer B1 from Solis BioDyne, containing Tris–HCl and (NH4)$_2$SO$_4$), 1.5 μL MgCl$_2$ (25 mM), 1 μL dNTP (2.5 mM of each dNTP), 0.2 μL (5 U/μL) HOT FIREPol® Taq Polymerase from Solis BioDyne, and 0.2 μL of each forward primer, 0.5 μL of each reverse primer, and 1 μL of the M13 primer (6-FAM) for the first multiplex, but 0.3 μL and 0.75 μL of the forward and reverse primers for the *CaCrSSR5* marker, respectively, 0.1 μL and 0.25 μL of the forward and reverse primers for the *CaCrSSR8* marker, respectively, 0.2 μL and 0.5 μL of the forward and reverse primers for the *CaCrSSR9* marker, respectively, and 1 μL M13 primer (HEX) for the second multiplex. Concentration of all primers was 5 pM/μL.

The PCR conditions for both multiplexes were 95 °C for 15 min followed by 35 cycles of 95 °C for 1 min, 50 °C for 1 min, 72 °C for 1 min, followed by final 72 °C for 20 min, and a hold at 16 °C.

By using 1.5% agarose gel electrophoresis single bands for all PCR reactions were visualized, and the dilution ratio for the PCR product was determined. PCR products diluted at 1:10 were run on an ABI 3130xl Genetic Analyser (Applied Biosystems, Foster City, CA, USA) with GS 500 ROX used as internal size standard.

2.4. Data Analysis

Allele size analysis was performed using GeneMapper version 4.1 (Applied Biosystems, Foster City, CA, USA). Haplotypes based on all six chloroplast markers and their frequencies were determined using the Haplotype Analysis version 1.05 software [59]. This software was also used to estimate total (H_T) and within population (H_S) haplotypic diversity, genetic differentiation (F_{ST}) between western and eastern natural populations, and between natural populations and cultivated cacao in Ecuador. The haplotype network to visualize the relationships among haplotypes was generated using the Network 5.0.1.1

software [60]. First, a "rdf" file was created in GenAlEx 6.5 [61], and then it was reformatted and saved in Network 5.0.1.1 as an "ych" file and used by this software as input file to generate the haplotype network using the median joining method. The created output file ("out") was saved and used for visualization [60].

3. Results

3.1. Identification of Haplotypes

In total, 26 alleles were identified for the six cacao chloroplast markers in 233 samples. The *CaCrSSR1* marker with a pentanucleotide repeat was the most polymorphic marker with nine alleles. Six alleles were observed at *CaCrSSR5*, five at *CaCrSSR2*, three at *CaCrSSR4*, and two at each *CaCrSSR8* and *CaCrSSR9*. These alleles allowed us to detect 23 haplotypes. Table 1 shows the allele compositions of these haplotypes and their frequencies. Six samples were excluded from the data analysis due to their incomplete genotyping (Tables S1–S3).

Table 1. *Theobroma cacao* L. chloroplast haplotypes and their frequencies observed in different locations in the Amazon, plantations, and breeding populations.

Chloroplast Haplotype	CaCr1	CaCr2	CaCr4	CaCr5	CaCr8	CaCr9	Western Amazonia	Other Locations *	Plantations	Breeding Populations	N	Frequency, %
1	361	243	177	207	309	345	2				2	0.9
2	366	243	177	206	309	345	2	2	4	4	12	5.3
3	366	243	177	207	309	345	1				1	0.4
4	366	244	177	206	309	345			1		1	0.4
5	371	242	178	206	308	345	12			1	13	5.7
6	371	243	177	206	309	345	1	1			2	0.9
7	376	241	177	206	308	345	1				1	0.4
8	376	241	178	211	309	345			1		1	0.4
9	376	242	178	206	308	345	4		1	4	9	3.9
10	376	242	178	211	309	345	4		24	9	37	16.2
11	376	242	178	219	309	345	1				1	0.4
12	376	244	179	216	308	353	2				2	0.9
13	376	245	179	216	308	353	2				2	0.9
14	381	243	177	206	309	345	2				2	0.9
15	381	243	178	214	308	353	1				1	0.4
16	381	243	178	216	308	353	24	34	2	2	62	27.2
17	381	244	178	216	308	353	6	5	1	1	13	5.7
18	386	244	178	216	308	353	1				1	0.4
19	391	244	178	216	308	353	17		1	1	19	8.3
20	406	244	177	207	309	345	6			7	13	5.7
21	416	243	178	216	308	353				1	1	0.4
22	416	244	178	216	308	353	26	2	1	2	31	13.6
23	416	244	178	216	309	353	1				1	0.4

* Central Amazon (Manaus, Brazil) and Eastern Amazon (Belem, Brazil and Camopi, French Guyana). The allele size of each microsatellite marker (*CaCr*) is presented in base pairs. N—number of individuals.

3.2. Western Amazonia, a Center of Haplotype Diversity

We observed high cpDNA genetic diversity in western Amazonia, with 19 haplotypes detected in this region in total (Figures 2 and 3). In Peruvian Amazon, 11 haplotypes were observed in three river systems, with an average of 4.6 haplotypes. The highest number of haplotypes was found along the Amazon River at Iquitos, followed by the Marañon and Nanay rivers with haplotypes H22, H16, and H19 being the most frequent, respectively. In addition, five haplotypes were observed in the south and central part of Peru along the Urubamba and Ucayali rivers (Figure 2). Nine haplotypes were observed in Ecuadorian Amazon along the Coca and Napo rivers, with haplotypes H9 and H5 being the most frequent, respectively. Finally, two haplotypes, H22 and H12, were observed in the Colombian Amazon (Figure 3). A high total haplotypic diversity ($H_T = 0.725$) and a large number of haplotypes ($N_h = 19$) were detected in western Amazonia in 28 locations (Figure 2 and Table 2). The within population haplotypic diversity was also high ($H_S = 0.676$) in this area.
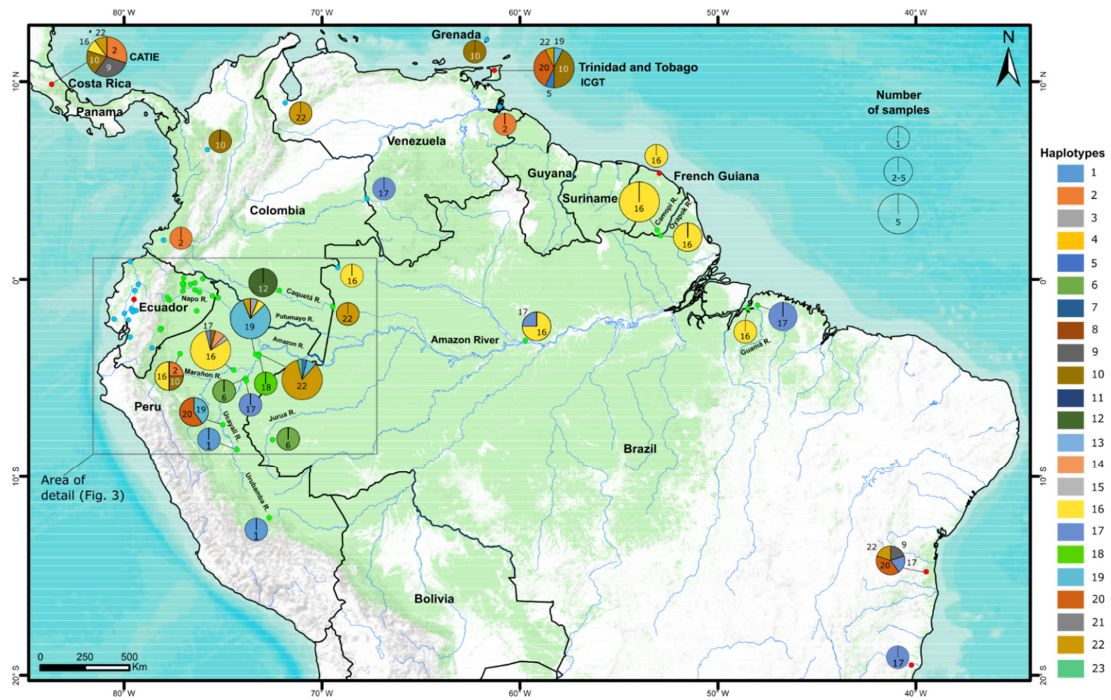
Figure 2. Distribution and frequencies of chloroplast haplotypes among *Theobroma cacao* L. populations in the Amazon basin. The rectangle represents the area of high haplotype diversity in northwestern Amazon (displayed in detail in Figure 3). The green dots represent samples from natural populations, the blue dots plantations, and the red dots accessions in breeding stations (CATIE-Tropical Agricultural Research and Higher Education Center, ICGT-International Cocoa Genebank Trinidad). The size of the circles is proportional to the number of samples per location.
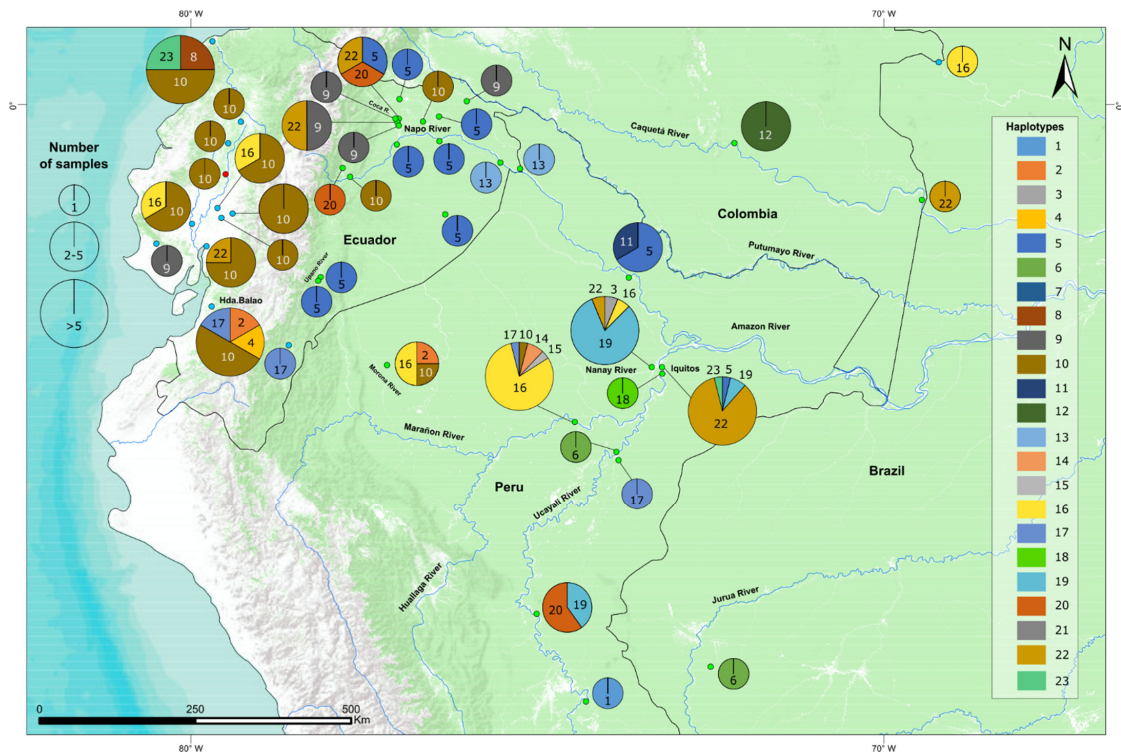

Figure 3. Distribution and frequencies of chloroplast haplotypes among *Theobroma cacao* L. populations in western Amazonia. The green dots represent samples from natural populations, the blue dots plantations, and the red dots accessions in breeding stations. The size of the circles is proportional to the number of samples per location.

Table 2. Chloroplast haplotype diversity within cacao populations in western and eastern Amazonia.

Amazonia	N	N_{Loc}	N_h	H_S	H_T	F_{ST}
Western	115	28	19	0.676	0.725	0.068
Eastern	32	3	2	0.037	0.087	0.572

N—sample size, N_{Loc}—number of locations, N_h—number of haplotypes, H_S—within population haplotypic diversity, H_T—total haplotypic diversity, F_{ST}—differentiation among populations within regions.

In contrast, only two haplotypes (H16 and H17) were found in central and eastern Amazonia (Figure 2). Both the total (H_T = 0.087) and the within population (H_S = 0.037) haplotypic diversity were low in eastern Amazonia (Table 2). When compared, eastern and western cacao populations were highly divergent (F_{ST} = 0.500), since they shared only two haplotypes in this study (Figure 2).

3.3. Geographically Restricted Haplotypes in Western Amazonia

In Peru, seven geographically restricted haplotypes occurred at three locations: the confluence of the Marañon and Amazon rivers (H3, H14, H15, H18, H23), Morona River (H2) and Napo River (H11). In addition, haplotype H1 was locally common in the south of Peru along the Urubamba-Ucayali River (Figure 2).

In Ecuador, the geographically restricted haplotype H13 was observed along the Napo River. Moreover, haplotypes H5 and H9 were common in different locations along the Napo basin at the border with Colombia, and haplotype H5 was common along Upano River. Haplotype 12 was observed in the Colombian Amazon at Caquetá River (Figure 3).

3.4. Correspondence between Chloroplast and Nuclear DNA Variation

There is an agreement between the chloroplast haplotypes observed in 83 individuals (Tables S1 and S3) included in Motamayor et al. [2] and eight cacao genetic groups. Each group generally has one dominant haplotype and other related haplotypes. Results may suggest that the cacao genetic groups proposed by Motamayor et al. [2] are heterogeneous in terms of the chloroplast variation (Table 3). However, to study in detail patterns of correspondence between cpDNA and nuclear variation in cacao, sampling for cpDNA analysis should ideally include all individuals analyzed by Motamayor et al. [2].

The dominant haplotypes within the genetic groups in the Peruvian Amazon were H22 in Iquitos, H19 in Nanay, H16 in Marañon, and H20 in Contamana in central Peru. In the Ecuadorian Amazon the dominant haplotype was H9 in Curaray and H10 in Nacional, and in central-eastern Amazonia H17 in Amelonado and H16 in Guiana (Table 3).

Four haplotypes were common within nuclear genetic groups: haplotype H16 was observed from west to east in the Amazon basin in Marañon (78%), Amelonado (12%), Iquitos (4%) and Guiana (100%). It is worth noticing that Marañon, Guiana, and Amelonado were also related groups based on nSSRs (Figure 2 in Motamayor et al. [2]). Haplotype H17 occurred together with the closely related H16 in Marañon demonstrating a phylogeographic pattern (Figure 4). Finally, haplotype H19 was shared between distinct clusters Nanay (91%) and Contamana (16%), and haplotype H20 was detected in the distantly related groups Contamana (50%) and Curaray (16%) (Table 3, Figure 2 in Motamayor et al. [2]).

Table 3. Correspondence between nuclear and chloroplast DNA analyses in cacao (*Theobroma cacao* L.). Cacao genetic groups, haplotype frequency and number of samples analyzed in this study and in Motamayor et al. [2].

Genetic Group	Chloroplast Haplotype Frequency	Number of Samples	
		This Study	Motamayor et al. [2]
Nanay	19 (0.91) 3 (0.08)	12	121
Iquitos	22 (0.86) 23 (0.04) 20 (0.04) 16 (0.04)	22	75
Marañon	16 (0.78) 14 (0.08) 17 (0.04) 15 (0.04) 14 (0.04)	23	130
Guiana	16 (1.0)	2	51
Amelonado	17 (0.62) 22 (0.12) 16 (0.12) 5 (0.12)	8	63
Contamana	20 (0.51) 1 (0.33) 19 (0.16)	6	59
Curaray	9 (0.33) 2 (0.33) 13 (0.16) 20 (0.16)	6	87
Nacional	10 (0.50) 6 (0.50)	2	36

3.5. Haplotypes in Ecuadorian Plantations

Eight haplotypes were identified in 29 samples in nine cacao plantations along the coast of Ecuador, with haplotype H10 being the most frequent one (72%) from north to southwest Ecuador (Figure 3). The geographic origin of H10 is associated with three locations in four samples: Napo basin in Ecuador (N = 2), Morona (N = 1) and Marañon (N = 1) rivers in Peru (Figure 3).

Cacao farms in Ecuador also contained haplotype H2 being the second most frequent haplotype in plantations (Table 1, Figure 3). In three locations totaling three samples, H2 was observed along Napo River in Ecuador and Morona River in Peru. However, H2 and H10 may also occur in other Amazonian cacao populations that were not analyzed here.

H4 and H8 were rare haplotypes in Ecuadorian plantations observed only in Hacienda Balao and Vuelta Larga, respectively. These haplotypes were not observed in any natural population studied here (Figures 2 and 3).
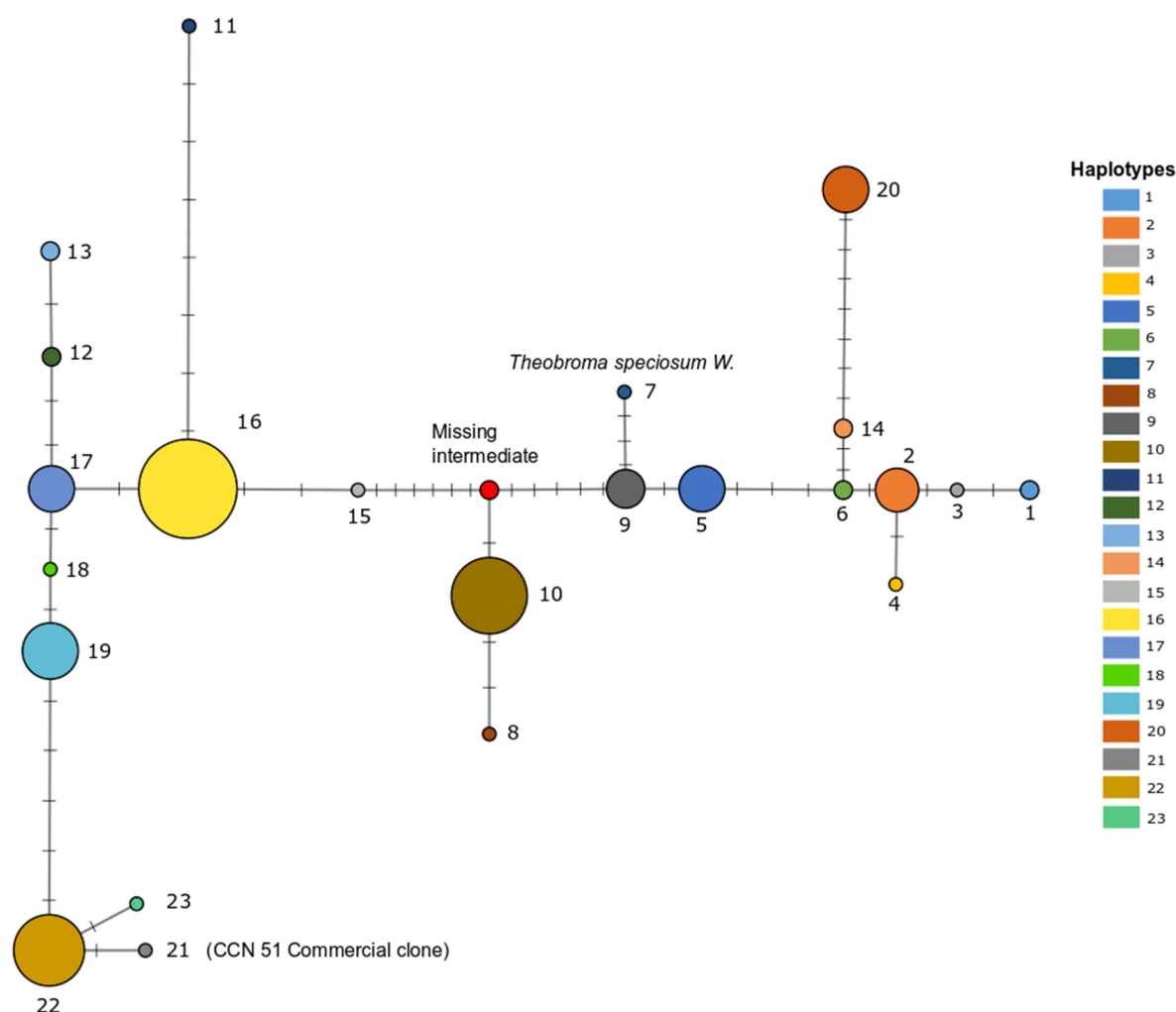
Figure 4. Network of chloroplast haplotypes (1–23) in *Theobroma cacao* L. populations. Each circle represents a specific haplotype, and its size reflects the respective haplotype frequency. Vertical bars in the lines connecting haplotypes indicate hypothetical mutations separating haplotypes. The distance between neighboring vertical lines is the same, but it was condensed in some linages for better space use and visualization of the network.

Low total ($H_T = 0.116$) and within population ($Hs = 0.075$) haplotypic diversity were observed in plantations of western Ecuador due to the high frequency of H10 and relatively low sample size in plantations (N = 29). Differentiation between cultivated and samples collected in Amazonian forest was high ($F_{ST} = 0.500$).

3.6. Haplotypes Observed in Breeding Populations and Cultivars

Cacao breeding has been mostly relying on vegetatively propagated accessions, such as Imperial College Selection (ICS), Iquitos Mixed Calabacillo (IMC), Nanay (NA), Parinari (PA), Pound, Scavina (SCA), and United Fruit (UF) [62–64]. Table 4 presents haplotypes detected in reference accessions and their potential geographic sources. For example, SCA 6, Pound 18, PA 139, and IMC 67 showed H5, H16, H20, and H22 haplotypes, respectively, and the geographic origin of these accessions ranged from central to northern Peru (Figure 2). In addition, cultivars, such as UF 273 (Nacional × Amelonado), ICS, and EET (Nacional × Unknown) revealed the common haplotype H10 observed in Ecuadorian plantations (Figure 3).

Table 4. Chloroplast haplotypes observed in cacao (*Theobroma cacao* L.) breeding populations, tentative geographic origin of haplotypes, source, and breeding program.

Accession	Source	Haplotype	Tentative Geographic Origin	Breeding Program Origin
CC 137	Cacao Center	2	Morona River, Peru	CATIE
* UF 168	United Fruit selections			UF Company, Costa Rica
NA 399	Nanay	3	Nanay River, Peru	ICG,T
ICS 68	Imperial Collection Selection	5	Western Amazonia	ICG,T
POUND 18	Pound Collections		Iquitos, Peru	ICG,T
CC 252	Cacao Center	9	The northeast of Ecuador (Coca and San Miguel rivers)	CATIE
PMCT 93	Programa Mejoramiento de Cultivos Tropicales			CATIE
* UF 712	United Fruit selections			UF Company, Costa Rica
* UF 273	United Fruit selections	10	The northeast of Ecuador (Coca River) and north of Peru (Morona and Marañon rivers)	UF Company, Costa Rica
* ICS 5, 15, 42, 46, 48, 63	Imperial Collection Selection			ICG,T
EET 19, 95	Estación Experimental Tropical			EET, Pichilingue, Ecuador
PA 150, 169	Parinari	16	Marañon River, Peru	ICG,T
SCA 9,10	Scavina	19	Ucayali River, Peru	ICG,T
NA 33,	Nanay		Nanay River, Peru	ICG,T
* CRU 100	Cocoa Research Unit		Nanay River, Peru	ICG,T
* ICS 35, 41	Imperial Collection Selection	20	Western Amazonia	ICG,T
* TSA 654, 656,792	Trinidad Selected Amazons			ICG,T
SCA 11, 12, 6	Scavina		Ucayali River, Peru	ICG,T
* ICS 10	Imperial Collection Selection	22	Western Amazonia	ICG,T
IMC 47, 60, 67	Iquitos Mixed Calabacillo		Iquitos, Peru	ICG,T

Accession name refers to an established accession name accepted at national and international level. * Trinitario cultivars; CATIE: Centro Agronómico Tropical de Investigación y Enseñanza, Costa Rica; ICGT: International Cocoa Genebank, Trinidad. Agronomic details (yield, disease resistance, and favorable traits) of these cacao clones are available in The International Cocoa Germplasm Database (ICGD) [52].

The cacao cultivar CCN 51 was characterized by the unique haplotype H21. This haplotype is separated by one mutational step from H22 in the mononucleotide motif repeat of *CaCrSSR2* (Figure 4). H22 was also found in the International Cacao Collection in Catie, including the Matina accession, which is also a common cultivar. Another unique and closely related haplotype is H23, which was observed along with haplotype 22 in Iquitos, Peru. H22 and H23 formed a separate cluster (lineage) in the haplotype network (Figure 4).

H10 was the most common haplotype (42%) in a set of Trinitario cultivars followed by H20 (35%) and other less frequent haplotypes H5, H19, and H22 (Table 4).

4. Discussion

4.1. Distribution of Haplotype Diversity

High haplotype diversity was observed in western Amazonia (Figure 3). This region represents a hot spot of cacao diversity [10,65], where cacao populations have a high haplotype diversity in the river systems of Marañon, Amazon (Iquitos), Nanay, and Uyacali in Peru, and along the Coca and Napo river systems in Ecuador (Figures 2 and 3).

Our results support the hypothesis of decreasing cacao diversity from the western part to the eastern part of the range. The same differentiation pattern from west to east was observed in the Amazon basin for 200 cacao accessions representing ten genetic groups based on whole genome sequencing and multidimensional scaling analysis [15]. Thus, the distribution of haplotype diversity is consistent with the suggested origin of the species in western Amazonia [1,2,12,13].

Our results support the hypothesis of western Amazonia being considered as a forest refugial area [21,66] and a center of crop genetic diversity [26,67], where cacao populations could have been restricted to temporarily isolated refugia during the Pleistocene (Figure 1). After the Pleistocene, suitable habitats during the Holocene allowed migration of cacao populations by human dispersal from western to central Amazonia along the Amazon

River [65]. Cacao populations may have experienced selection in this center of crop diversity to some degree by domestication for fruit pulp and later were dispersed through the basin by pre-Columbian human expansion [24,26]. Western Amazonia may also be a center of domestication which is reflected in the presence of predominant haplotypes H10, H20, and H22 both in breeding populations and plantations in western Amazonia (Figure 2 and Table 3).

Haplotypes that were observed only in western Amazonia (H12 in Colombia, H13 in Ecuador, and H11, H3, H2, H14, H15, H18, and H23 in Peru) probably are associated with cacao populations that experienced isolation in forest refugia when the Amazon forest was reduced during the Pleistocene (Figure 1) [21,22,65]. Another possible explanation assumes in situ occurrence due to edaphic adaptation to rich soils of refugial areas, which is common in tree species of Amazonia [68,69].

Although the cacao samples studied here were distributed across the Amazon basin, they mainly represented the Upper Amazon (Figure 1). Thus, it is important to analyze additional cacao populations within the species range that could reveal new and distinctive haplotypes. For example, a recent forest survey of 1170 plots in Amazonia identified cacao as a common tree in the basin; specifically, cacao populations were observed in southwestern and southern Amazonia, and French Guyana [27,70]. Pre-Columbian human societies likely enriched these forest areas with cacao trees brought from the center of cacao diversity in western Amazonia [27].

The occurrence of only a single haplotype H16 in French Guiana could indicate human meditated dispersal of a few cacao individuals to this region. However, additional evidence to support human-mediated distribution of cacao in Amazonia is needed using more extensive sampling, specific studies such as radiocarbon and stable isotope analyses, and archaeological and ecological surveys looking at past human interactions with tropical tree species [71,72].

4.2. Chloroplast Haplotypes Match Genetic Groups Based on nSSRs

Haplotype diversity and distribution observed in our study partly mirrored the genetic clusters of the dendrogram based on 96 nSSRs (Figure 2 in [2]). For example, the Marañon, Guiana and Amelonado cluster is represented by related haplotypes 16 and 17 in the haplotype network (Table 4, Figure 4). Moreover, the cluster comprising Nanay and Iquitos is represented by the related haplotypes 19 and 22, and the closely related clusters Curaray and Nacional by the related haplotypes H10 and H9 in our haplotype network (Table 4, Figure 4).

The presence of specific haplotypes in different geographic regions and within different genetic groups suggest particular patterns of seed dispersal. For example, haplotype H20 was observed in Curaray (north of Ecuador), Iquitos (north of Peru), and Contamana (central Peru) groups (Figure 3, Table 4). Similarly, haplotype H19 was observed in the distantly related Contamana (central Peru) and Nanay (north of Peru) groups (Figure 3, Table 4). This may be explained by the presence of temporarily separated forest refugia in western Amazonia [21,67] where cacao populations were likely restricted during the Pleistocene [65]. Later, during the Holocene cacao populations experienced habitat expansion when the climate became wetter and warmer, favorable for wide distribution in the Amazon basin [65]. Cacao habitat expansion aided by human meditated dispersal during the Holocene times may explain ample seed dispersal observed among cacao groups [27,65].

Marañon, Guiana and Amelonado clustered together in the nSSR-based dendrogram [2]; these populations share H16, and Amelonado and Marañon also have the related haplotype H17 in common (Table 4). These haplotypes reveal a wide distribution from western to eastern Amazonia (Figure 2) and likely experienced human dispersal in the basin. Indeed, current abundance of cacao populations in south-southwestern Amazon and French Guyana was explained partially by pre-Columbian human dispersal [27].

Cacao populations in French Guyana probably originated from local forest refugia during the Pleistocene–Holocene epochs [73]. However, an alternative explanation of their

origin may be human meditated seed dispersal from western to eastern Amazon [14,27,65]. The observed wide distribution of haplotypes of H16 and H17 in the Amazon basin supports this hypothesis (Figures 2 and 3). However, to clarify the origin of Amelonado and Guiana groups more extensive sampling across the species' geographic range is needed accompanied by historical records of cacao dispersal in the Amazon. However, difficulties may arise to identify the hypothetical origin of native Amazonian crops due to forest expansion during the Holocene [24].

In the case of the Nacional group, two main natural populations with H10 were observed: one in the Ecuadorian Amazon (Napo River) and another in the Peruvian Amazon (Morona River) (Figure 2). We consider that additional cpDNA analyses in Ecuadorian populations are needed to provide more evidence about the origin of the Nacional variety. However, our results are in agreement with [2] who found 13 individuals from the Morona River clustered in the Nacional group. Furthermore, based on the nSSR genetic analysis of 65 wild individuals widespread in the Amazonian forest of Ecuador and eight individuals from the Morona River in Peru, [74] suggested an ancestral origin of the Nacional group in the Southern region of Ecuadorian Amazon close to Morona River.

The distribution of H10 suggests a pattern of seed dispersal among cacao populations in western Amazonia, which supports the hypothesis of a shared origin of this haplotype from Ecuadorian and Peruvian populations and later their introduction to the coast of Ecuador, where they were established in pure plantations. Furthermore, chloroplast analysis of Chuncho, a cacao population from Peru that resembles the Nacional fine-flavor cacao [75], would help to trace the geographic origin of fine-flavor cacao to the south of Peru as suggested by [75].

4.3. Chloroplast Haplotypes and Domesticated Cacao

Combined analyses of nuclear and cpDNA variation can help to unravel the history of cacao domestication and to narrow down the origin of domesticated varieties. For example, the Criollo variety (not included in the present study) was domesticated from a fraction of the ancestral Curaray population about 3600 years ago [15]. Within the Curaray population we observed haplotypes H9, H13, and H20 in northern Ecuador (Figure 3 and Table 4). CpDNA analyses of the Criollo variety could reveal its geographic origin within the distribution range of Curaray. Indeed, the geographic origin of haplotype H9 observed with high frequency in the Curaray group (Figure 4) possibly suggests that trees from the Ecuador-Colombia border served as a source of plant material during the domestication of Criollo cacao (Figure 3).

Further questions arise such as why the domestication of cacao started with the Ecuadorian Curaray group and not with other populations. It was probably because its sweet pulp was used for the elaboration of fermented beverages, which may also have favored its selection and dispersal out of the Ecuadorian Amazon [24,74,76].

4.4. Haplotype Diversity in Plantations and Breeding Populations

The high haplotype diversity observed in natural populations contrasts with the low diversity observed in cacao plantations (Table 2). We observed a high contribution of haplotype H10 in both selections made in cacao plantations in the west of Ecuador and breeding populations such as the ICS clones of Trinidad (Table 3). In addition, H10 was also observed in Colombia and Grenada, providing further evidence for introduction events of this haplotype to other countries.

Cacao cultivation in coastal Ecuador has been recorded since the 17th century, and cacao genotypes named Nacional were selected and cultivated because of their special chocolate flavor [1]. Currently, new sources of genetic material are required to improve flavor, increase yield and add disease resistance traits to the Nacional cultivar [74]. We suggest that the four wild provenances associated with haplotype H10 in western Amazonia of Peru and Ecuador could be screened for agronomic traits to test their use in breeding programs, potentially adding useful traits to Nacional cultivars.

The presence of six haplotypes, with haplotype H10 being the most frequent (42%), in 15 reference Trinitario accessions suggests seed introduction events occurring probably from northeast of Ecuador and north of Peru to Trinidad (Table 3). Similar to our results, five different haplotypes were observed in 21 different reference Trinitario accessions based on nine cacao cpSSRs markers [45].

The river basins associated with geographically restricted haplotypes (Figures 2 and 3) could serve as source of specific adaptations and potential new traits for cacao breeding programs, but only if these samples carry favorable agronomic traits. Examples of these areas include the Napo basin of Ecuador (H5, H9, and H13), the Caquetá River (H12) in Colombia, and Ucayali River in south and central Peru (H1).

5. Conclusions

Western Amazonia contains high haplotype diversity areas in Peru and Ecuador. Populations with at least two common haplotypes may have been dispersed by humans from this center of cacao genetic diversity to new suitable habitats following a west to east route of migration in the Amazon basin. The western Amazonia has high value for cacao conservation considering its haplotype diversity and presence of geographically unique haplotypes, which were not observed in breeding populations and cultivars yet and could be evaluated for agronomic traits in support of cacao improvement.

The cacao cpDNA haplotypes observed here can be used to determine the chloroplast diversity of accessions and select distinctive haplotypes in cacao breeding programs. Additionally, the map of cpDNA haplotypes can be used to verify the geographic origin of planting material at finer geographic scale and point to potential areas for new collections. Craft chocolate makers can use this reference map to create new markets for cacao products based on geographic origin.

6. Outlook

New collections in Amazonia will expand the availability of cacao genetic resources and help us better understand the patterns of geographic genetic variation and dispersal. However, rapid land use changes and gold mining in the Amazon forest are threatening the natural habitat of wild cacao populations and eroding its genetic diversity [8,70,77]. Likewise, examples of documented cacao natural populations that may reveal new haplotypes are Upper Orinoco [78], Caquetá River in Colombian Amazon and Ecuadorian Amazon [51], Santiago and Morona rivers in northern Peru [79], Chuncho in southern Peru [75], Beni River in Bolivia [80], the Brazilian Amazon [14,81,82], southwestern and southern Amazonia [27,83], and Amazonian basin [2]. Furthermore, the analysis of additional samples like those held in international (CATIE and ICGT) and national cacao collections (see [52]) likely will reveal additional chloroplast diversity.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/d13060249/s1, Table S1: Clone name, ICQC (The International Cocoa Quarantine Centre) accession number, geographic origin, samples per location, genetic group according to Motamayor et al. 2008, and chloroplasts haplotypes observed in a sample of *Theobroma cacao* L. accessions. The haplotype column shows the order of the markers *CaCr1, CaCr2, CaCr4, CaCr5, CaCr8* and *CaCr9*, and the fragment sizes are given in base pairs (bp). Details about donor collections, geographic location of collections expeditions and agronomic traits of these cacao clones are available on the ICGD (The International Cocoa Germplasm Database) website [52]. Table S2: Clone name, ICQC accession number, locations, and chloroplast haplotypes of *Theobroma cacao* L. in nine farms of the pacific coast of Ecuador. The haplotype column shows the order of the markers *CaCr1, CaCr2, CaCr4, CaCr5, CaCr8* and *CaCr9*, and the fragment sizes are given in base pairs (bp). Collection and agronomic details of these cacao clones are available on the ICGD website. Table S3: Clones names, ICQC accession number, origin of samples, genetic group (Motamayor et al. 2008), and chloroplasts haplotypes observed in a sample of *Theobroma cacao* L. genotypes used for cultivation or breeding. The haplotype column shows the order of the markers *CaCr1, CaCr2, CaCr4, CaCr5, CaCr8* and *CaCr9*, and the

fragment sizes are given in base pairs (bp). Agronomic details of these cacao clones (yield, disease resistance, and favorable traits) and breeding details are available on the ICGD website.

References

1. Bartley, B.G.D. *The Genetic Diversity of Cacao and Its Utilization*; CABI Pub: Cambridge MA, USA, 2005; ISBN 0851996191.
2. Motamayor, J.C.; Lachenaud, P.; da Silva, E.; Mota, J.W.; Loor, R.; Kuhn, D.N.; Brown, J.S.; Schnell, R.J. Geographic and genetic population differentiation of the Amazonian chocolate tree (Theobroma cacao L). *PLoS ONE* **2008**, *3*, e3311. [CrossRef]
3. Rice, R.A.; Greenberg, R. Cacao Cultivation and the Conservation of Biological Diversity. *Ambio* **2000**, *29*, 167–173. [CrossRef]
4. Beg, M.S.; Ahmad, S.; Jan, K.; Bashir, K. Status, supply chain and processing of cocoa—A review. *Trends Food Sci. Technol.* **2017**, *66*, 108–116. [CrossRef]
5. Bennett, A.B. Out of the Amazon: Theobroma cacao enters the genomic era. *Trends Plant Sci.* **2003**, *8*, 561–563. [CrossRef] [PubMed]
6. Goenaga, R.; Irizarry, H.; Irish, B. TARS Series of Cacao Germplasm Selections. *HortScience* **2009**, *44*, 826–827. [CrossRef]
7. Phillips-Mora, W.; Arciniegas-Leal, A.; Mata-Quirós, A.; Motamayor-Arias, J.C. *Catalogue of Cacao Clones: Selected by Catie for Commercial Plantings*; CATIE: Turrialba, Costa Rica, 2013; ISBN 9789977575902.
8. Zhang, D.; Figueira, A.; Motilal, L.; Lachenaud, P.; Meinhardt, L.W. Theobroma. In *Wild Crop Relatives: Genomic and Breeding Resources*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 277–296.
9. CacaoNet. *A Global Strategy for the Conservation and Use of Cacao Genetic Resources, as the Foundation for a Sustainable Cocoa Economy*; Bioversity International: Montpellier, France, 2012.
10. Zhang, D.; Motilal, L. Origin, Dispersal, and Current Global Distribution of Cacao Genetic Diversity. In *Cacao Diseases*; Bailey, B.A., Meinhardt, L.W., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 3–31. ISBN 978-3-319-24787-8.
11. Laliberté, B.; End, M.; Cryer, N.; Daymond, A.; Engels, J.; Eskes, A.B.; Gilmour, M.; Lachenaud, P.; Phillips-Mora, W.; Turnbull, C.; et al. Conserving and exploiting cocoa genetic resources: The key challenges. In *Achieving Sustainable Cultivation of Oil Palm*; Burleigh Dodds Science Publishing: Cambridge, UK, 2018; Volume 1, pp. 19–46.
12. Cheesman, E.E. Notes on the nomenclature, classification and possible relationships of cacao populations. *Trop. Agric.* **1944**, *21*, 144–159.
13. Cuatrecasas, J. Cacao and its allies: A taxonomic revision of the genus Theobroma. *Syst. Plant Stud.* **1964**, *35*, 379–614.
14. Sereno, M.L.; Albuquerque, P.S.B.; Vencovsky, R.; Figueira, A. Genetic Diversity and Natural Population Structure of Cacao (*Theobroma cacao* L.) from the Brazilian Amazon Evaluated by Microsatellite Markers. *Conserv. Genet.* **2006**, *7*, 13–24. [CrossRef]
15. Cornejo, O.E.; Yee, M.-C.; Dominguez, V.; Andrews, M.; Sockell, A.; Strandberg, E.; Livingstone, D., III; Stack, C.; Romero, A.; Umaharan, P.; et al. Population genomic analyses of the chocolate tree, Theobroma cacao L., provide insights into its domestication process. *Commun. Biol.* **2018**, *1*, 1–12. [CrossRef] [PubMed]
16. Tuomisto, H.; Ruokolainen, K.; Kalliola, R.; Linna, A.; Danjoy, W.; Rodriguez, Z. Dissecting Amazonian Biodiversity. *Science* **1995**, *269*, 63–66. [CrossRef]
17. Myers, N.; Mittermeier, R.A.; Mittermeier, C.G.; Da Fonseca, G.A.B.; Kent, J. Biodiversity hotspots for conservation priorities. *Nature* **2000**, *403*, 853–858. [CrossRef] [PubMed]

18. Ter Steege, H.; Pitman, N.; Sabatier, D.; Castellanos, H.; Van Der Hout, P.; Daly, D.C.; Silveira, M.; Phillips, O.; Vasquez, R.; Van Andel, T.; et al. A spatial model of tree α-diversity and tree density for the Amazon. *Biodivers. Conserv.* **2003**, *12*, 2255–2277. [CrossRef]

19. Bass, M.S.; Finer, M.; Jenkins, C.N.; Kreft, H.; Cisneros-Heredia, D.F.; McCracken, S.F.; Pitman, N.C.A.; English, P.H.; Swing, K.; Villa, G.; et al. Global Conservation Significance of Ecuador's Yasuní National Park. *PLoS ONE* **2010**, *5*, e8767. [CrossRef]

20. Garcia-Davila, C.; Gomero, D.A.; Renno, J.-F.; Soria, R.D.; Pizango, G.H.; Llampazo, G.F.; Castro-Ruiz, D.; De Loayza, E.M.; Chavez, C.A.; Mader, M.; et al. Molecular evidence for three genetic species of Dipteryx in the Peruvian Amazon. *Genetics* **2019**, *148*, 1–11. [CrossRef] [PubMed]

21. Prance, G. Forest refuges: Evidence from woody angiosperms. In *Biological Diversification in the Tropics*; Prance, G., Ed.; Columbia University Press: New York, NY, USA, 1982.

22. van der Hammen, T.; Hooghiemstra, H. Neogene and Quaternary history of vegetation, climate, and plant diversity in Amazonia. *Quat. Sci. Rev.* **2000**, *19*, 725–742. [CrossRef]

23. Clement, C. 1492 and the loss of amazonian crop genetic resources. Crop Biogeography at contact. *Econ. Bot.* **1999**, *53*, 203–216. [CrossRef]

24. Clement, C.R.; De Cristo-Araújo, M.; D'Eeckenbrugge, G.C.; Pereira, A.A.; Picanço-Rodrigues, D. Origin and Domestication of Native Amazonian Crops. *Divers* **2010**, *2*, 72–106. [CrossRef]

25. Meyer, R.S.; DuVal, A.E.; Jensen, H.R. Patterns and processes in crop domestication: An historical review and quantitative analysis of 203 global food crops. *New Phytol.* **2012**, *196*, 29–48. [CrossRef]

26. Clement, C.R.; Denevan, W.M.; Heckenberger, M.J.; Junqueira, A.B.; Neves, E.G.; Teixeira, W.G.; Woods, W.I. The domestication of Amazonia before European conquest. *Proc. R. Soc. B Boil. Sci.* **2015**, *282*, 20150813. [CrossRef]

27. Levis, C.; Costa, F.R.C.; Bongers, F.; Peña-Claros, M.; Clement, C.R.; Junqueira, A.B.; Neves, E.G.; Tamanaha, E.K.; Figueiredo, F.O.G.; Salomão, R.P.; et al. Persistent effects of pre-Columbian plant domestication on Amazonian forest composition. *Science* **2017**, *355*, 925–931. [CrossRef]

28. Paniagua-Zambrana, N.; Byg, A.; Svenning, J.-C.; Moraes, M.; Grandez, C.; Balslev, H. Diversity of palm uses in the western Amazon. *Biodivers. Conserv.* **2007**, *16*, 2771–2787. [CrossRef]

29. Yang, J.Y.; Motilal, L.A.; Dempewolf, H.; Maharaj, K.; Cronk, Q.C.B. Chloroplast microsatellite primers for cacao (Theobroma cacao) and other Malvaceae. *Am. J. Bot.* **2011**, *98*, e372–e374. [CrossRef] [PubMed]

30. Weising, K.; Nybom, H.; Wolff, K.; Kahl, G. *DNA Fingerprinting in Plants: Principles, Methods, and Applications*, 2nd ed.; Taylor & Francis Group: Boca Raton, FL, USA, 2005; ISBN 0849314887.

31. Lemes, M.R.; Dick, C.W.; Navarro, C.; Lowe, A.J.; Cavers, S.; Gribel, R. Chloroplast DNA Microsatellites Reveal Contrasting Phylogeographic Structure in Mahogany (Swietenia macrophylla King, Meliaceae) from Amazonia and Central America. *Trop. Plant Biol.* **2010**, *3*, 40–49. [CrossRef]

32. Saunders, J.A.; Mischke, S.; Leamy, E.A.; Hemeida, A.A. Selection of international molecular standards for DNA fingerprinting of Theobroma cacao. *Theor. Appl. Genet.* **2004**, *110*, 41–47. [CrossRef]

33. Irish, B.M.; Goenaga, R.; Zhang, D.; Schnell, R.; Brown, J.S.; Motamayor, J.C. Microsatellite Fingerprinting of the USDA-ARS Tropical Agriculture Research Station Cacao (*Theobroma cacao* L.) Germplasm Collection. *Crop. Sci.* **2010**, *50*, 656–667. [CrossRef]

34. Motilal, L.A.; Zhang, D.; Umaharan, P.; Mischke, S.; Pinney, S.; Meinhardt, L.W. Microsatellite fingerprinting in the International Cocoa Genebank, Trinidad: Accession and plot homogeneity information for germplasm management. *Plant Genet. Resour.* **2011**, *9*, 430–438. [CrossRef]

35. Jemmy, T.; Ji, K.; Lyndel, M.; Sue, M.; Stephen, Y.O.; Francis, K.P.; Dapeng, Z. Verification of genetic identity of introduced cacao germplasm in Ghana using single nucleotide polymorphism (SNP) markers. *Afr. J. Biotechnol.* **2014**, *13*, 2127–2136. [CrossRef]

36. Zhang, D.; Susilo, A.W.; Dinarti, D.; Bailey, B.; Mischke, S.; Meinhardt, L.W. Genetic Identity, Ancestry and Parentage in Farmer Selections of Cacao from Aceh, Indonesia Revealed by Single Nucleotide Polymorphism (SNP) Markers. *Trop. Plant Biol.* **2014**, *7*, 133–143. [CrossRef]

37. Padi, F.K.; Ofori, A.; Takrama, J.; Djan, E.; Opoku, S.Y.; Dadzie, A.M.; Bhattacharjee, R.; Motamayor, J.C.; Zhang, D. The impact of SNP fingerprinting and parentage analysis on the effectiveness of variety recommendations in cacao. *Tree Genet. Genomes* **2015**, *11*, 1–14. [CrossRef]

38. Ji, K.; Zhang, D.; Motilal, L.A.; Boccara, M.; Lachenaud, P.; Meinhardt, L.W. Genetic diversity and parentage in farmer varieties of cacao (*Theobroma cacao* L.) from Honduras and Nicaragua as revealed by single nucleotide polymorphism (SNP) markers. *Genet. Resour. Crop. Evol.* **2013**, *60*, 441–453. [CrossRef]

39. Hamilton, M.B. Tropical tree gene flow and seed dispersal. *Nat. Cell Biol.* **1999**, *401*, 129–130. [CrossRef]

40. Caron, H.; Dumas, S.; Marque, G.; Messier, C.; Bandou, E.; Petit, R.J.; Kremer, A. Spatial and temporal distribution of chloroplast DNA polymorphism in a tropical tree species. *Mol. Ecol.* **2000**, *9*, 1089–1098. [CrossRef]

41. Cavers, S.; Navarro, C.; Lowe, A.J. Chloroplast DNA phylogeography reveals colonization history of a Neotropical tree, Cedrela odorata L., in Mesoamerica. *Mol. Ecol.* **2003**, *12*, 1451–1460. [CrossRef]

42. Dick, C.W. Phylogeography and Population Structure of Tropical Trees. *Trop. Plant Biol.* **2010**, *3*, 1–3. [CrossRef]

43. Pandey, M.; Gailing, O.; Leinemann, L.; Finkeldey, R. Molecular markers provide evidence for long-distance planting material transfer during plantation establishment of Dalbergia sissoo Roxb. in Nepal. *Ann. For. Sci.* **2004**, *61*, 603–606. [CrossRef]

44. Jia, Y.; Milne, R.I.; Zhu, J.; Gao, L.; Zhu, G.; Zhao, G.; Liu, J.; Li, Z. Evolutionary legacy of a forest plantation tree species (Pinus armandii): Implications for widespread afforestation. *Evol. Appl.* **2020**, *13*, 2646–2662. [CrossRef] [PubMed]

45. Yang, J.Y.; Scascitelli, M.; Motilal, L.A.; Sveinsson, S.; Engels, J.M.M.; Kane, N.C.; Dempewolf, H.; Zhang, D.; Maharaj, K.; Cronk, Q.C.B. Complex origin of Trinitario-type Theobroma cacao (Malvaceae) from Trinidad and Tobago revealed using plastid genomics. *Tree Genet. Genomes* **2013**, *9*, 829–840. [CrossRef]

46. Gutiérrez-López, N.; Ovando-Medina, I.; Salvador-Figueroa, M.; Molina-Freaner, F.; Arrazate, C.H.A.; Vázquez-Ovando, A. Unique haplotypes of cacao trees as revealed bytrnH-psbAchloroplast DNA. *PeerJ* **2016**, *4*, e1855. [CrossRef]

47. Coe, S.D.; Coe, M.D. *The True History of Chocolate*, 3rd ed.; Thames & Hudson: London, UK, 2013; ISBN 9780500290682.

48. Pound, F.J. Cacao and witches' broom disease (Marasmius perniciosus) of South America with notes on other species of Theobroma. *Arch. Cocoa Res.* **1938**, *1*, 26–71.

49. Vello, F.; Madeiros, A. Expedicao botanica a Amazonica brasileira. Cacau Atualidades (Brasil). *Cacau Atual.* **1965**, *2*, 47–51.

50. Sallée, B. *Compte-Rendu de Prospection Dans les POPULATIONS Naturelles de Theobroma Cacao L. du Haut Camopi (Guyane Française), 14/4 au 12/5/1987*; Camopi: Haut Camopi, French Guiana, 1987.

51. Allen, J.B. Geographical Variation and Population Biology in Wild Theobroma Cacao. Ph.D. Thesis, The University of Edinburgh, Edinburgh, UK, 1988.

52. Turnbull, C.; Hadley, P. International Cocoa Germplasm Database (ICGD). Available online: http://www.icgd.reading.ac.uk (accessed on 24 November 2019).

53. Peters, C. PreColumbian silviculture and indigenous management of neotropical forests. In *Imperfect Balance: Landscape Transformations in the Precolumbian Americas*; Lentz, D., Ed.; Columbia University Press: New York, NY, USA, 2000; pp. 203–224. ISBN 0-231-11157-6.

54. Barlow, J.; Gardner, T.A.; Lees, A.; Parry, L.; Peres, C.A. How pristine are tropical forests? An ecological perspective on the pre-Columbian human footprint in Amazonia and implications for contemporary conservation. *Biol. Conserv.* **2012**, *151*, 45–49. [CrossRef]

55. Pound, F.J. *First Report on the Selection of Cacao Trees for Resistance to Witches' Broom Disease (Marasmius perniciosa): Report on a Recent Visit to the Amazon Territory of Peru, September 1942–February 1943*; International Cocoa Genebank: Centeno, Trinidad and Tobago, 1943.

56. Zhang, D.; Boccara, M.; Motilal, L.; Butler, D.R.; Umaharan, P.; Mischke, S.; Meinhardt, L. Microsatellite variation and population structure in the "Refractario" cacao of Ecuador. *Conserv. Genet.* **2007**, *9*, 327–337. [CrossRef]

57. Weising, K.; Gardner, R.C. A set of conserved PCR primers for the analysis of simple sequence repeat polymorphisms in chloroplast genomes of dicotyledonous angiosperms. *Genome* **1999**, *42*, 9–19. [CrossRef] [PubMed]

58. Kubisiak, T.L.; Nelson, C.D.; Staton, M.E.; Zhebentyayeva, T.; Smith, C.; Olukolu, B.A.; Fang, G.-C.; Hebard, F.V.; Anagnostakis, S.; Wheeler, N.; et al. A transcriptome-based genetic map of Chinese chestnut (Castanea mollissima) and identification of regions of segmental homology with peach (Prunus persica). *Tree Genet. Genomes* **2013**, *9*, 557–571. [CrossRef]

59. Eliades, N.; Eliades, D. *Haplotype Analysis; Forest Genetics and Forest Tree Breeding, Georg-August*; University Goettingen: Goettingen, Germany, 2009.

60. Bandelt, H.J.; Forster, P.; Rohl, A. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* **1999**, *16*, 37–48. [CrossRef]

61. Peakall, R.; Smouse, P.E. GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research—An update. *Bioinformatics* **2012**, *28*, 2537–2539. [CrossRef] [PubMed]

62. Bekele, F.; Phillips-Mora, W. Cacao (*Theobroma cacao* L.) Breeding. In *Advances in Plant Breeding Strategies: Industrial and Food Crops*; Al-Khayri, J.M., Jain, S.M., Johnson, D.V., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 409–487. ISBN 978-3-030-23264-1.

63. Lopes, U.V.; Monteiro, W.R.; Pires, J.L.; Clement, D.; Yamada, M.M.; Gramacho, K.P. Cacao breeding in Bahia, Brazil: Strategies and results. *Theor. Appl. Genet.* **2011**, *11*, 73–81. [CrossRef]

64. Dias, L.A.D.S. The contributions of breeding. In *Genetic Improvement of Cacao*; Dias, L.A.D.S., Ed.; Food and Agriculture Organization of the United Nations: Rome, Italy, 2001.

65. Thomas, E.; Van Zonneveld, M.; Loo, J.; Hodgkin, T.; Galluzzi, G.; Van Etten, J. Present Spatial Diversity Patterns of Theobroma cacao L. in the Neotropics Reflect Genetic Differentiation in Pleistocene Refugia Followed by Human-Influenced Dispersal. *PLoS ONE* **2012**, *7*, e47676. [CrossRef] [PubMed]

66. Haffer, J.; Prance, G. Climatic forcing of evolution in Amazonia during the Cenozoic: On the refuge theory of biotic differ-entiation. *Limnol. Oecologia Reg. Syst. Fluminis Amazon.* **2001**, *16*, 579–607.

67. Clement, C.R. A Center of Crop Genetic Diversity in Western Amazonia. *Biosciences* **1989**, *39*, 624–631. [CrossRef]

68. Gentry, A.H. Neotropical Floristic Diversity: Phytogeographical Connections Between Central and South America, Pleistocene Climatic Fluctuations, or an Accident of the Andean Orogeny? *Ann. Mo. Bot. Gard.* **1982**, *69*, 557. [CrossRef]

69. Gentry, A.H. Speciation in tropical forests. In *Tropical Forests: Botanical Dynamics, Speciation and Diversity: Symposium: Papers*; Holm-Nielsen, L.B., Nielsen, I.C., Balslev, H., Eds.; Elsevier: Amsterdam, The Netherlands, 1989; pp. 113–134. ISBN 9780123535504.

70. Ter Steege, H.; Pitman, N.C.A.; Sabatier, D.; Baraloto, C.; Salomão, R.P.; Guevara, J.E.; Phillips, O.L.; Castilho, C.V.; Magnusson, W.E.; Molino, J.-F.; et al. Hyperdominance in the Amazonian Tree Flora. *Science* **2013**, *342*, 1243092. [CrossRef]

71. Levis, C.; Flores, B.M.; Moreira, P.A.; Luize, B.G.; Alves, R.; Franco-Moraes, J.; Lins, J.; Konings, E.; Peña-Claros, M.; Bongers, F.; et al. How People Domesticated Amazonian Forests. *Front. Ecol. Evol.* **2018**, *5*, 15. [CrossRef]

72. Caetano-Andrade, V.L.; Clement, C.; Weigel, D.; Trumbore, S.; Boivin, N.; Schöngart, J.; Roberts, P. Tropical Trees as Time Capsules of Anthropogenic Activity. *Trends Plant Sci.* **2020**, *25*, 369–380. [CrossRef]

73. Lachenaud, P.; Zhang, D. Genetic diversity and population structure in wild stands of cacao trees (Theobroma cacaoL.) in French Guiana. *Ann. For. Sci.* **2008**, *65*, 310. [CrossRef]

74. Solórzano, R.G.L.; Fouet, O.; Lemainque, A.; Pavek, S.; Boccara, M.; Argout, X.; Amores, F.; Courtois, B.; Risterucci, A.M.; Lanaud, C. Insight into the Wild Origin, Migration and Domestication History of the Fine Flavour Nacional Theobroma cacao L. Variety from Ecuador. *PLoS ONE* **2012**, *7*, e48438. [CrossRef]

75. Eskes, A.; Rodriguez, C.; Cruz-Condori, D.; Seguine, E.; Garcia-Carrion, L.; Lachenaud, P. Large genetic diversity for fi-ne-flavor traits unveiled in cacao (*Theobroma cacao* L.) with special attention to the native Chuncho variety in Cusco, Peru. *Agrotrópica* **2018**, *30*, 157–174. [CrossRef]

76. Zarrillo, S.; Gaikwad, N.; Lanaud, C.; Powis, T.; Viot, C.; Lesur, I.; Fouet, O.; Argout, X.; Guichoux, E.; Salin, F.; et al. The use and domestication of Theobroma cacao during the mid-Holocene in the upper Amazon. *Nat. Ecol. Evol.* **2018**, *2*, 1879–1888. [CrossRef] [PubMed]

77. Sierra, R. Dynamics and patterns of deforestation in the western Amazon: The Napo deforestation front, 1986–1996. *Appl. Geogr.* **2000**, *20*, 1–16. [CrossRef]

78. Patiño, V.M. *Plantas Cultivadas y Animales Domésticos en América Equinoccial*, 1st ed.; Imprenta Departamental: Prado, Colombia, 1963.

79. Arevalo-Gardini, E.; Meinhardt, L.W.; Zuñiga, L.C.; Arévalo-Gardni, J.; Motilal, L.; Zhang, D. Genetic identity and origin of "Piura Porcelana"—A fine-flavored traditional variety of cacao (Theoborma cacao) from the Peruvian Amazon. *Tree Genet. Genomes* **2019**, *15*, 11. [CrossRef]

80. Zhang, D.; Martínez, W.J.; Johnson, E.S.; Somarriba, E.; Phillips-Mora, W.; Astorga, C.; Mischke, S.; Meinhardt, L.W. Genetic diversity and spatial structure in a new distinct *Theobroma cacao* L. population in Bolivia. *Genet. Resour. Crop. Evol.* **2012**, *59*, 239–252. [CrossRef]

81. Dias, L.A.D.S.; Barriga, J.P.; Kageyama, P.Y.; De Almeida, C.M.V.C. Variation and its distribution in wild cacao populations from the Brazilian Amazon. *Braz. Arch. Biol. Technol.* **2003**, *46*, 507–514. [CrossRef]

82. Almeida, C.; Neto, F.; Silva, P.; Medeiros, J.; Lima, A. Genetic resources of cacao in Rondônia: Historical overview, origin and inventory (in Portuguese). *Agrotrópica* **2015**, *27*, 93–124.

83. Wittmann, F.; Schongart, J.; Montero, J.C.; Motzer, T.; Junk, W.J.; Piedade, M.T.F.; Queiroz, H.L.; Worbes, M. Tree species composition and diversity gradients in white-water forests across the Amazon Basin. *J. Biogeogr.* **2006**, *33*, 1334–1347. [CrossRef]