

Dissecting Genetic Structure in Farmer Selections of *Theobroma Cacao* in the Peruvian Amazon: Implications for on Farm Conservation and Rehabilitation

Dapeng Zhang · Enrique Arevalo Gardini · Lambert A. Motilal · Virupax Baligar · Bryan Bailey · Luiz Zuñiga-Cernades · Carlos E. Arevalo-Arevalo · Lyndel Meinhardt

Received: 5 October 2010 / Accepted: 21 December 2010
© Springer Science+Business Media, LLC (outside the USA) 2011

Abstract Knowledge of genetic diversity in farmers' selections is essential for planning on-farm conservation and rehabilitation. Using 15 microsatellite loci, we analyzed parentage and population structure in 220 farmer selections of cacao from the Huallaga valley in Peruvian Amazon. A high level of allele richness and heterozygosity were detected in these selections. Coordination analysis showed that these farmer selections are mainly comprised of hybrids derived from Trinitario and Upper Amazon Forastero germplasm. Bayesian clustering analysis assigned 54 selections as Trinitario and 166 as Upper Amazon Forastero hybrids. Parentage analysis identified 15 international clones as probable parents for 96 farmer selections, which corresponded to a fraction of the known hybrid families disseminated in this region in the late 1980s. Combined analysis of demographic and molecular data revealed a significant spatial autocorrelation ($r=0.235$; $P=0.006$) at short geographical distances (<5.0 km). This patch-like distribution of spatial heterogeneity suggests a significant "neighborhood effects" in seeds distribution or variety adoption, where closely related hybrid progenies were adopted in the neighboring farms or villages. The

outcomes of this study indicate that in spite of the introgressions of exotic germplasm in the past, Upper Amazon Forastero is still the dominant component in the Huallaga valley of Peru. The high level of on-farm diversity can offer needed variability for participatory selection of superior clones in this low input, small-scale production system, where adaptability to specific environment is more preferred than in a high input, large scale production system.

Keywords Cacao · Farmer selections · Huallaga valley · Peruvian Amazon · Rehabilitation · Spatial genetic correlation

Introduction

Farmer participatory varietal selection refers to processes whereby farmers are systematically involved in selecting lines that are most adaptable to local conditions (Witcombe et al. 1996). Farmer participatory varietal selection has been advocated as an efficient approach to promote on-farm conservation of genetic resources, because farmers are more likely to have continued preferences to use these selections in the local agro-ecosystem (Almekinders and Elings 2001; Bellon 2004; Brush 2000). Participatory selection may be more useful for tropical perennial crops in low input and small scale production systems, where farmers' experience in multi-year field observations can benefit the selection of superior trees, because specific adaption would receive adequate weights as selection criteria (Dias 2001; Eskes 2006; Leakey and Akinnifesi 2007). To use this breeding scheme, the underlining assumption is that high levels of genetic diversity and variability in agronomic traits exist in farmers' fields, which can be explored for the identification

Communicated by Christopher Dick

D. Zhang (✉) · V. Baligar · B. Bailey · L. Meinhardt
USDA ARS PSI SPCL,
10300 Baltimore Avenue, Bldg. 001, Rm. 223, BARC-W,
Beltsville, MD 20705, USA
e-mail: Dapeng.Zhang@ars.usda.gov

E. A. Gardini · L. Zuñiga-Cernades · C. E. Arevalo-Arevalo
Instituto de Cultivos Tropicales (ICT),
Jr. Santa Maria, No. 241, Banda de Shilcayo Tarapoto, Peru

L. A. Motilal
Cocoa Research Unit, The University of the West Indies,
St. Augustine, West Indies, Trinidad and Tobago

of promising cultivars (Bellon 2004; Dias 2001). An improved understanding of the structure of genetic diversity in farmer selections is therefore essential for the application of participatory selection, and to ensure that the indigenous component of genetic diversity is well represented in the “elite” materials that are selected.

Theobroma cacao L., which is native to tropical South America, is in the family Sterculiaceae (alternatively Malvaceae *sensu lato*) (Alverson et al. 1999; Bayer et al. 1999; Cuatrecasas 1964). This species comprises a large number of highly morphologically variable populations, which can all be crossed with one another (Cheesman 1944; Bartley 2005). Although it is commonly accepted that cacao was domesticated in Mesoamerica (De la Cruz et al. 1995; Young 1994), the Upper Amazon, including the rain forest in Peru, has been proposed as the center of origin of this species (Bartley 2005; Cheesman 1944; Cuatrecasas 1964; Pound 1945).

The cultivated cacao has been traditionally subdivided into three main groups: (Criollo, Forastero and Trinitario) although other distinctive groups are recognized (Cheesman 1944; Wood and Lass 2001). Among the three main groups, Criollo cacao was domesticated more than 3,000 years ago primarily for its sweeter pulp by early indigenous tribes in Mesoamerica (Henderson et al. 2007). It was believed that Criollo was the only cacao variety cultivated in Mesoamerica before the arrival of the Europeans (Bartley 2005; Motamayor et al. 2002). The term Forastero refers to introduced foreign trees that are not Criollo. Forastero encompasses a diverse range of populations from South America and each population has a distinctive genetic identity (Bartley 2005; Motamayor et al. 2008). Forastero were brought to the traditional cocoa producing regions including Central America and the Caribbean) in the mid of 18th century when the cacao plantations were devastated by unknown diseases. The resultant hybridization between the Criollo survivors and the introduced Forastero materials generated Trinitario, which was started in Trinidad in the mid of 18th century (Cheesman 1944; Bartley 2005).

The cultivation of cacao in Peru did not start until the 18th century, after the European colonization. However, historical evidence indicates that the indigenous people in the Amazon probably utilized the sweet pulp surrounding the seeds in pre-Columbian times (Clement et al. 2010; Dias 2001; Smith 1999). Whether this early utilization pattern has had any impact on the evolution of traditional cacao varieties in Peru is unknown. Nevertheless, it is known that germplasm from Caribbean, Central America and Ecuador was introduced into Peru in the eighteenth century, when a small commercial cacao production started in Peru (Bartley 2005). The cacao genotypes currently

cultivated in Peru consist of traditional varieties and hybrids derived from introduced germplasm. It has been suggested that the genetic background of these genotypes can be partially traced back to the “Pound collection,” which was established in the 1930’s (Hernandez 1991; Arévalo et al. 1999; Pound 1945). This collection was comprised of five natural populations from the Peruvian Amazon including germplasm groups of Iquitos Mixed Calabacillo (IMC), Morona (MO), Nanay (NA), Parinari (PA) and Scavina (SCA) (Pound 1938; 1945; Lockwood and End 1993; Zhang et al. 2009). In the early 1950’s, new foreign clones were introduced into Peru. Included among them were the United Food Company selections from Costa Rica (UF clones), the Imperial College selections from Trinidad (ICS clones), and the selections from Ecuador (EET clones). A breeding program was launched in 1953 by Peruvian national institutions to develop high-yielding and disease-resistant cacao hybrids (Arévalo et al. 1999; UNDCP and UNOPS 1996). Controlled crosses were made between introduced foreign clones and Peruvian Forastero germplasm. A large number of hybrid families were distributed in cacao producing regions, including the Huallaga valley. This was an especially significant effort in the various rehabilitation programs supported by the United Nations Drug Control Programs (UNDCP) in the 1980’s–1990’s (Evans et al. 1998; UNDCP and UNOPS 1996).

Farmer participatory selection and dissemination of new cocoa varieties has been promoted in this region by NGOs, International organizations and local government. The underlining assumption is that large allelic diversity, based on the disseminated hybrid families, exists in farmer’s fields and this diversity can be explored for the identification of superior clones (Eskes 2006). Working with the rural communities through farmer field school, the individual trees with superior yield and disease resistance are identified and collected. These promising clones are being evaluated in replicated trial and the best performers with field resistance to witches’ broom and frosty pod at the various participating locations will be selected (Eskes 2006). These elite farmer selections then will either be recommended for direct adoption on farm (by grafting) or serve as parental clones for breeding.

In this paper we describe a pilot study in which we genotyped 220 farmer selections from Juanhui region in the Huallaga valley, Peru, using microsatellite molecular markers. The first objective of the study was to assess the level of on-farm genetic diversity and infer the ancestry and parentage of these farmer selections. The second objective was to understand how the genetic diversity is spatially distributed in the studied area. The results of the study will elucidate the status of on-farm diversity in the Huallaga valley of Peru. They will also be used for the development

of a rational strategy in participatory varietal selection and deployment in rehabilitation programs.

Results

Genetic Diversity in Farmer Selections

A high level of genotype diversity was observed among the farmer selections from Juanjui. SSR data showed that the 220 farmer selections are all distinctive clones. The farmer selections not only had high level of genotype diversity, but also had a high level of allelic diversity (Table 2). A mean number of alleles of 10.0 was observed in the 220 farmer selections. The expected heterozygosity in the Juanjui farmer selections was 0.741, where as the observed heterozygosity was 0.694 (Table 2). Moreover, there was a negligible inbreeding coefficient ($F=0.058$ per locus; Table 2) in the 220 farmer selections (Fig. 1).

Genetic Relationship and Ancestry Inference

The genetic relationships among the 220 Peruvian farmer selections, as well as the 27 candidate parental clones, were shown by Principal Coordinates Analysis (Fig. 2). The plane of the first two main PCO axes accounted for 53.9% of total variation. Most of the Upper Amazon Forastero clones (e.g. the NA, PA, and Pound clones) scattered among the farmer selections, indicating the Upper Amazon Forastero background in these farmer selections. A small proportion of the farmer selections showed close proximity with the Trinitario cacao (e.g. the ICS clones; Fig. 2).

The contribution of known candidate parental clones to these farmer selections was further illuminated by assignment test (Table 1). Using the admixture model, Bayesian clustering analysis assigned the 247 accessions (220 farmer selections and 27 candidate parents) into three genetically distinct clusters (Fig. 3). The first cluster is comprised of 130 farmer selections that share membership with one group of Peruvian Upper Amazon Forastero clones, including NA 33, PA 7, PA 150, POUND 7, POUND 12, and EET 400. The second cluster consists of 54 selections, which fall in the same cluster with most of the ICS Trinitario and UF clones. The third cluster includes 23 selections assigned in the same cluster with another group of Upper Amazon Forastero clones, including SCA 6 and SCA 12, which originated from the Ucayali river valley of Peru. There are also 13 selections which could not be assigned to any cluster, because their Q value in any one cluster was less than 50%, but their joint Q-value in the two Forastero clusters are above 0.50, thus they are counted as Upper Amazon Forastero. In total, there were 54 farmer selections that were classified as Trinitario type, whereas

166 selections were classified as Upper Amazon Forastero (Table 1; Fig. 3).

Parentage Analysis

Parentage analysis shows that, out of the 27 candidate parents, fifteen were responsible, at 80% confidence level, for the parentage (maternity or paternity) of 96 farmer selections in Juanjui. When the confidence level is raised to 95%, the number of identified parents was reduced to 11 clones that contributed to 56 farmer selections (Table 3). Among the identified parental clones, ICS 1 was by far the largest parentage contributor, which accounted for 34 parent-offspring pairs at 80% confidence. IMC 67 and SCA 6 ranked second and third, with 13 and 10 offspring, respectively, at 80% confidence. These top three clones were found to be responsible for 59.4% of the parent-offspring relationships identified among the international clones, at 80% confidence.

Spatial Genetic Structure

Mantel tests did not detect significant correlation between genetic and geographic distances by ($R_{xy}=0.027$, $P=0.198$). However, the global autocorrelation analysis of the 220 farmer selections found significantly positive correlation within short distance classes (Fig. 4). The correlograms indicated significant positive spatial autocorrelation within a distance of 5 km (Fig. 4). The autocorrelation became erratic after this range.

Significant local spatial genetic structure was also detected by the Two-Dimensional Local Spatial Autocorrelation Analyses (2D LSA). Among the 220 farmer selections, 50% of the local correlation (I_r) values, with the number of nearest neighbors at six ($NN=6$), were significantly positive based on a one-tailed test (Fig. 5). The result is compatible with that generated by global autocorrelation analysis, indicating a patch-like distribution of diversity, where genetically similar individual trees aggregated in the same villages or townships (Fig. 5). Positive values were also found when I_r estimates were based on 4, 8, 10, and 12 nearest neighbors, confirming the strength and consistency of this patch-like pattern of genotype distribution (data not shown).

Discussion

Genetic Diversity, Parentage and Spatial Pattern in Farmer Selections

The maximum of genotype diversity (each selection has a unique genotype) and high level of heterozygosity in the

Table 1 List of collection sites, number of farmer selections and assigned population membership (K=3) using Bayesian clustering analysis

Name of collecting sites in Juanjui, Peru	Number of farmer selections	Population membership			
		Upper Amazon Forastero (I)	Upper Amazon Forastero (II)	Tritinario	Others
Chambira	7	7	0	0	0
Pachiza	20	19	0	1	0
Aucararca	9	2	0	7	0
Bajo Juñao	5	1	0	2	2
Capirona	4	3	1	0	0
Dos unidos	12	2	3	4	3
Huicungo	11	1	0	8	2
Tingo de Saposoa	17	13	0	4	0
La Libertad	27	1	12	13	1
Ledoy	9	8	0	0	1
Saposoa	12	8	1	3	0
Pajarillo	12	12	0	0	0
Puerto Ríos	8	7	1	0	0
Puerto rico	3	1	2	0	0
Retama	5	1	1	3	0
Sacanche	19	17	1	1	0
Juanjui	14	13	0	1	0
Soledad	11	1	1	5	4
Villaprado	8	8	0	0	0
Zanja Seca	7	5	0	2	0
	220	130	23	54	13

The 220 farmer selections were assigned in two clusters of Upper Amazon Forastero and one cluster of Tritinario based on the Bayesian clustering analysis (Pritchard et al. 2000). Selections with assignment probability (Q value) above the threshold of 0.500 were accepted as a member of that cluster. Selections that do not belong to any particular cluster because the assignment probability to any one of the three clusters was smaller than 0.500 ($Q < 0.500$), but they can be classified as Upper Amazon Forastero based on their combined Q values

220 farmer selections from Juanjui was consistent with the previously reported results in the Huallaga clones from the same region (Zhang et al. 2006b). This result supports the hypothesis that these farmer selections were mainly derived from families of hybrid. Cacao planting materials in the Amazon were traditionally distributed by pods (Dias 2001; Bartley 2005). Propagation by clones was not common until very recently. The segregating hybrids families with diverse background offer a high level of gene and genotype diversity in the cacao fields.

The allele richness (10.0 alleles per locus) in the 220 farmer selections was high, and comparable with the wild populations collected by J.F. Pound from Peruvian Amazon (12.0 alleles per locus in 316 accessions; Zhang et al. 2009). The allele richness was much higher than the cultivated populations in the coast regions of Ecuador (4.96 alleles per locus in 322 accessions; Loor et al. 2008). The expected heterozygosity ($H_e=0.741$; Table 2) is higher than in the cultivated populations in Ecuador ($H_e=0.561$;

Loor et al. 2008) and the five wild cacao populations from Loreto ($H_e=0.628$; Zhang et al. 2009).

The present results suggest that the current on-farm diversity of cacao in Juanjui is largely built on the mixed foundation of Upper Amazon Forastero and Tritinario, as reflected by both PCoA (Fig. 2) and Bayesian clustering analysis (Fig. 3). Upper Amazon Forastero is still a major component in the current on-farm diversity in Juanjui, whereas the Tritinario hybrids only account for 25% of the farmer selections (Table 1; Fig. 3). Parentage analysis supported the outcome of ancestry inference and further showed that only 40% of the parentage (at 80% confidence level) can be explained by the recorded International clones used in the Peruvian cacao breeding program (Arévalo et al. 1999; Hernandez 1991). The remaining parentage was likely from unrecorded parental clones, which may include indigenous germplasm from Huallaga valley. In addition, large numbers of wild populations still exist in this region. Therefore, gene flow from wild populations may also impact the current pattern of on-farm diversity.

Table 2 Allelic diversity of the 15 microsatellite loci scored in the 220 farmer selections from Juanjui, Peru. N, total number of alleles; H_{Obs} , observed heterozygosity; H_{Exp} , expected heterozygosity; PID_sib, probability of identity of siblings

Genebank designation	Accession name ^a	N	H_{Obs}	H_{Exp}	Inbreeding coefficient ^b	PID-sib ^c
TCA16981	mTcCIR7	5	0.596	0.540	-0.105	0.266
TCA16980	mTcCIR6	12	0.784	0.832	0.057	0.049
TCA16995	mTcCIR22	6	0.555	0.568	0.023	0.225
TCA16996	mTcCIR24	8	0.546	0.515	-0.060	0.285
TCA16982	mTcCIR8	9	0.528	0.754	0.301	0.097
TCCTREP	mTcCIR1	9	0.518	0.580	0.106	0.234
TCA16985	mTcCIR11	12	0.615	0.816	0.246	0.058
TCA16986	mTcCIR12	11	0.830	0.804	-0.033	0.065
TCA16988	mTcCIR15	9	0.761	0.805	0.054	0.058
TCA271942	mTcCIR37	13	0.826	0.862	0.042	0.032
TCA271826	mTcCIR33	15	0.748	0.833	0.103	0.048
TCA16991	mTcCIR18	9	0.833	0.794	-0.050	0.069
TCA16998	mTcCIR26	9	0.673	0.744	0.096	0.104
TCA271943	mTcCIR40	13	0.826	0.831	0.006	0.049
TCA271958	mTcCIR60	10	0.771	0.838	0.080	0.046
Mean over 15 loci		10	0.694	0.741	0.058	0.000011 ^d

^a mTcCir=microsatellite *Theobroma cacao* CIRAD (Centre de Coopération Internationale en Recherche Agronomique pour le Développement)

^b Definition of Inbreeding Coefficient follows Wright (1965)

^c PID-sib Probability of identity among siblings follows the definition of Evett and Weir (1988)

^d Accumulated PID-sib as the loci add up, i.e. the PID-sib value of the second locus is the product of PID-sib of the first two loci

Among the international clones tested as candidate parents, Trinitario cacao made a significant contribution, but the contribution was predominantly limited to a small number of the ICS clones, especially ICS 1. The presence of Trinitario ancestry in these selections is consistent with the historical record of using ICS clones as an important progenitor in Peru's cacao breeding program since the 1950's (Arévalo et al. 1999; Evans et al. 1998). In many cases, ICS 1 was crossed with SCA 6 in hopes of incorporating disease resistances (Hernandez 1991). The high frequency of ICS 1 as the identified parental clone in the Juanjui selections also indicates that this clone has a good combinability and breeding value. The SCA 6 clone is a well known source of resistance to several cacao diseases, including witch's broom and black pod disease (Bartley 2005; Pound 1945). Despite the numerous Peruvian research documents showing that SCA 12, PA 150, PA 46, ICS 8, ICS 9 were commonly used parental clones in the same breeding program, no parentage was detected from these clones in the examined farmer selections.

The Juanjui farmer selections of cacao showed significant spatial structure corresponding to a patch-like distribution in short distances (<5 km), based on townships or villages. This result indicates variation in seed distribution and/or adoption of planting materials at different spatial

ranges. While the previous breeding records indicated that diverse sets of hybrid families were delivered to the entire Huallaga valley, the diversity at local community level appeared small, likely due to the mechanism in seed (in the form of seedlings) distribution. The intensity of local correlation showed a rather regular decay over distance between townships or villages. This pattern could also be the result of different acceptance of particular genotypes by farmers or specific adaptability to local environmental conditions. Sharing productive genotypes has been common among cacao growers around the world and pods from a productive tree would have higher likelihood of being integrated into a neighbor's field. The rehabilitation effort in the mid-1980 may have reinforced this trend, when same hybrid families were delivered to and adopted by farmers in the same communities or villages.

Implication for On-Farm Conservation and Rehabilitation

On-farm conservation is increasingly recognized as an important component of a conservation strategy for crop genetic resources. It supplements the *ex situ* conservation efforts by providing advantages in maintaining the evolution process under the ecological and agricultural system (Bellon 2004; Love and Spaner 2007). For crops cultivated in their

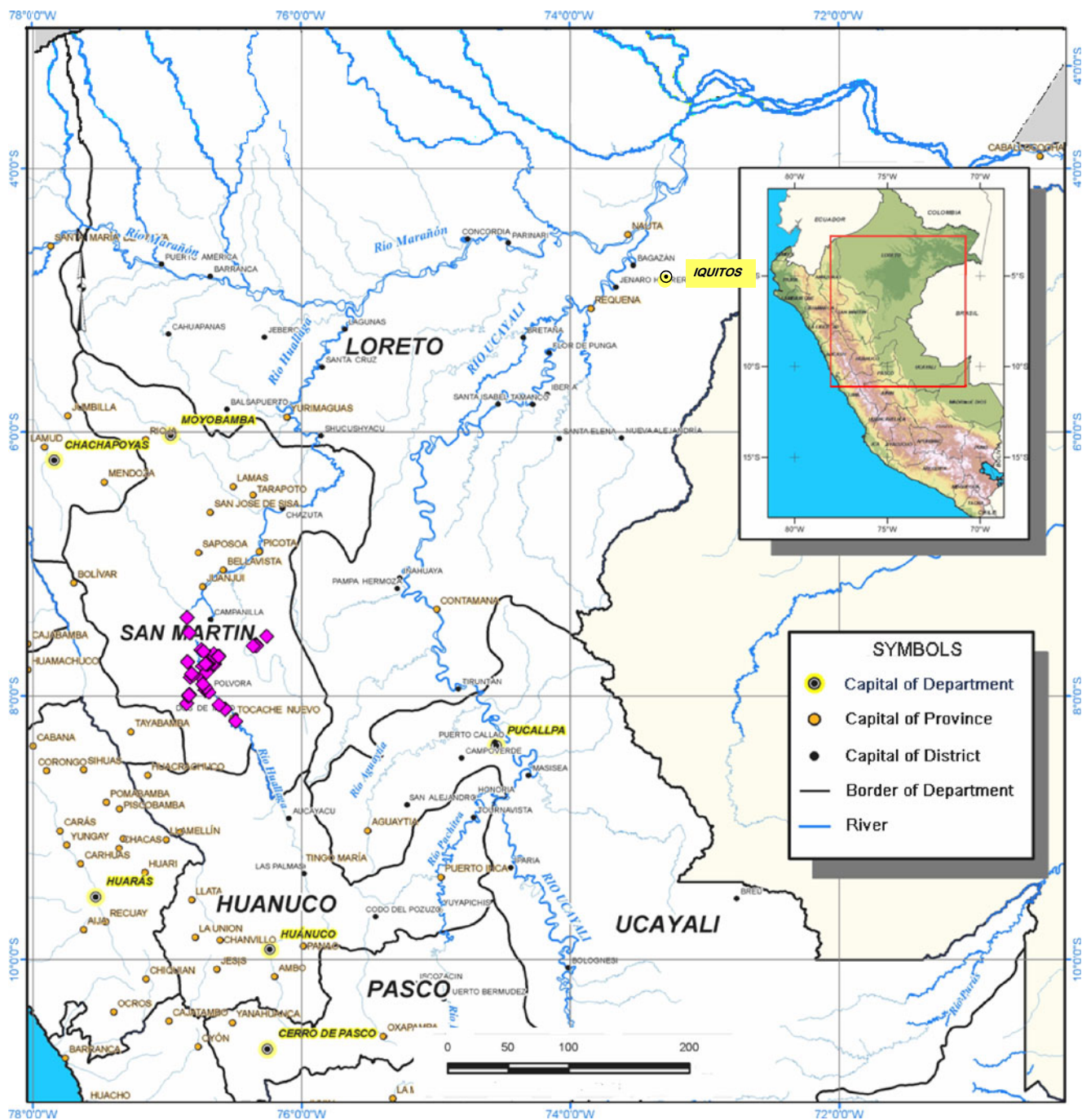
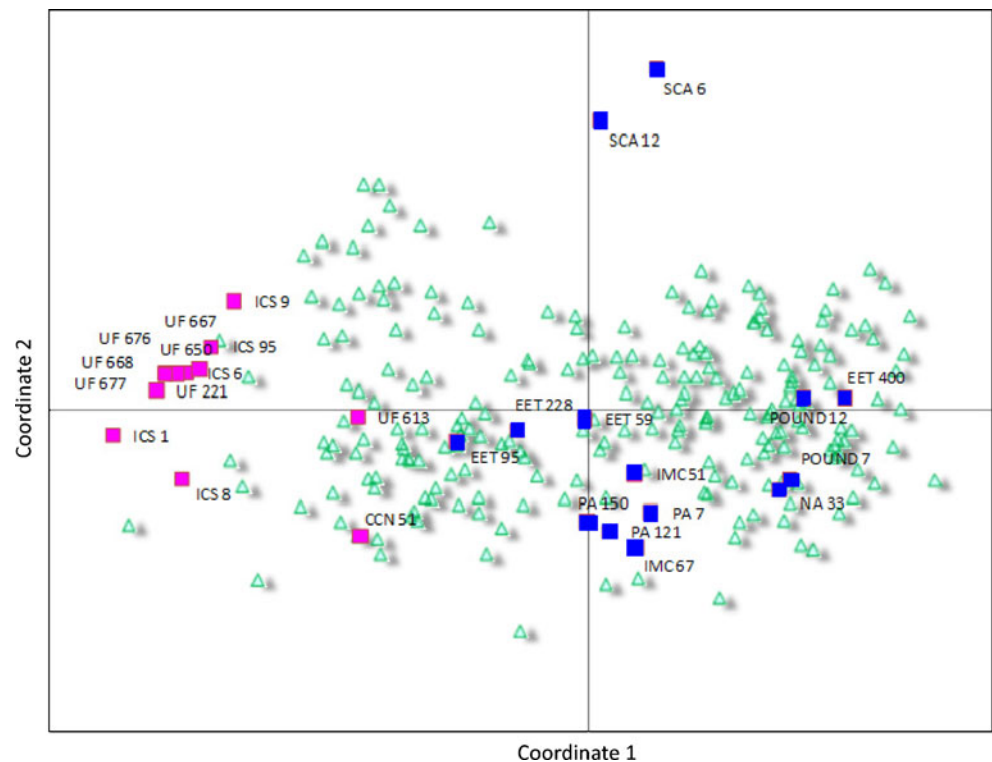


Fig. 1 Sampling sites of farmer selections in the Juanjui region in the upper Huallaga valley on the eastern slopes of the Andes. Administratively, Juanjui belongs to the Province of Mariscal Cáceres in the Department of San Martín

center of origin, such as cocoa Peruvian Amazon, a high level of indigenous intraspecific variation is often confounded with introduced germplasm or modern varieties. In recent years, this region experienced replacement of indigenous germplasm by the introduction of exotic varieties with reduced diversity. In extreme cases, a few highly productive varieties were grafted in the whole plantation. In Juanjui and Tingo María, it was estimated that 50% of the production is

from Trinitario types, including CCN 51 (The Peruvian Ministry of Agriculture 2000). The present study in Juanjui, however, indicates that in spite of the introductions of exotic germplasm over the years, Upper Amazon Forastero germplasm is still a major component of the current on-farm diversity. The elite farmer selections provide a good basis for on-farm conservation. Evaluation and distribution of the elite farmer selections, with the support of the on-going rehabilita-

Fig. 2 PCoA plot of 247 cacao accessions, including 220 farmer selections from Juanjui, Peru and 27 candidate parental clones used by the Peruvian breeding program in the 1980s. First axis=35.6% of total information and the second=18.3%



tion program, is an economically viable approach to conserve these germplasm *in situ*.

The present study showed that a high level of intraspecific allelic diversity exists in farmer's fields in the Huallaga valley in Peru. This high level of diversity can be explored for participatory selection of superior clones in this low input, small-scale production system, where adaptability to specific environment is more preferred than in a high input, large scale production system. Furthermore, the detected local spatial correlation suggested that the varietal deployment at the community level needs further improvement. Currently a large number of individual trees with superior traits have been identified and collected from rural communities through farmer field school. Genetic insight from analysis of these

selections, as provided by the present study, will be useful for the identification and deployment of clones with diverse genetic background aiming for durable diseases resistances in this region.

Materials and Methods

Site of Investigation

Juanjui is located in the upper Huallaga valley on the eastern slopes of the Andes, Administratively, it belongs to the Province of Mariscal Cáceres, in the Department of San Martín (Fig. 1). In the past two decades this region has been

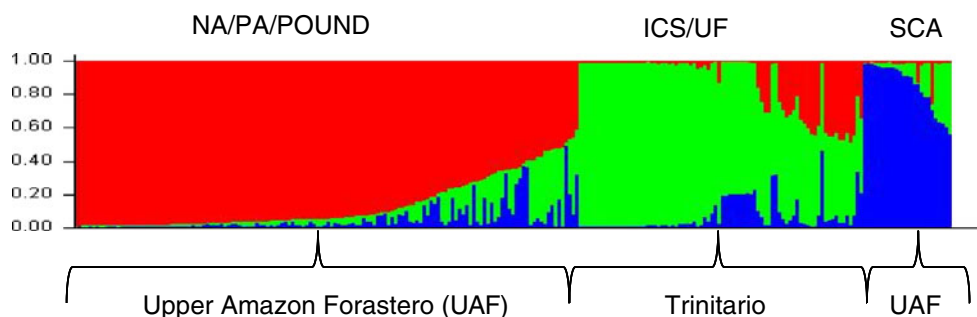


Fig. 3 Inferred clusters of 247 cacao accessions, including 220 farmer selections from Juanjui, Peru and 27 candidate parental clones used by the Peruvian breeding program in the 1980s using STRUCTURE. The most probable clusters were obtained at $K=3$, where K is the potential number of genetic clusters that may exist in the overall sample of

individuals. Each vertical line represents one individual multilocus genotype. Individuals with multiple colors have admixed genotypes from multiple clusters. Each color represents the most likely ancestry of the cluster from which the genotype or partial genotype was derived. Clusters of individuals are represented by colors

Table 3 Likelihood assignment of parentage in the 220 Juanjui farmers selections based on twenty-seven candidate parental clones

Candidate parents	Source of samples	Country of origin	Number of assigned parent-offspring pairs	
			>80% Confidence	>95% Confidence
EET 51	CATIE, Costa Rica	Ecuador	7	6
EET 228	CATIE, Costa Rica	Ecuador	2	1
EET 400	CRU, Trinidad	Ecuador	4	4
EET 59	CATIE, Costa Rica	Ecuador	0	0
EET 95	CATIE, Costa Rica	Ecuador	1	1
ICS 1	CRU, Trinidad	Trinidad	34	8
ICS 6	CRU, Trinidad	Trinidad	9	7
ICS 8	CRU, Trinidad	Trinidad	0	0
ICS 9	CRU, Trinidad	Trinidad	0	0
ICS 95	CRU, Trinidad	Trinidad	4	4
IMC 67	CRU, Trinidad	Peru	13	10
IMC 51	CRU, Trinidad	Peru	0	0
NA 33	CRU, Trinidad	Peru	3	2
PA 7	CRU, Trinidad	Peru	5	4
PA 121	CRU, Trinidad	Peru	0	0
PA 150	CRU, Trinidad	Peru	0	0
POUND 12	CRU, Trinidad	Peru	1	0
POUND 7	CRU, Trinidad	Peru	1	0
SCA 12	CRU, Trinidad	Peru	0	0
SCA 6	CRU, Trinidad	Peru	10	9
UF 221	CATIE, Costa Rica	Costa Rica	1	0
UF 613	CRU, Trinidad	Costa Rica	1	0
UF 650	CATIE, Costa Rica	Costa Rica	0	0
UF 667	CATIE, Costa Rica	Costa Rica	0	0
UF 677	CATIE, Costa Rica	Costa Rica	0	0
UF 668	CATIE, Costa Rica	Costa Rica	0	0
UF 676	CATIE, Costa Rica	Costa Rica	0	0
Total			96	56

Critical LOD for assignment at >80% and >95% confidence are 1.68 and 4.56 respectively

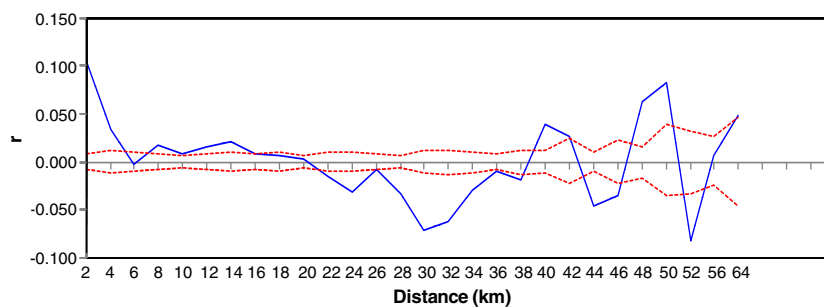
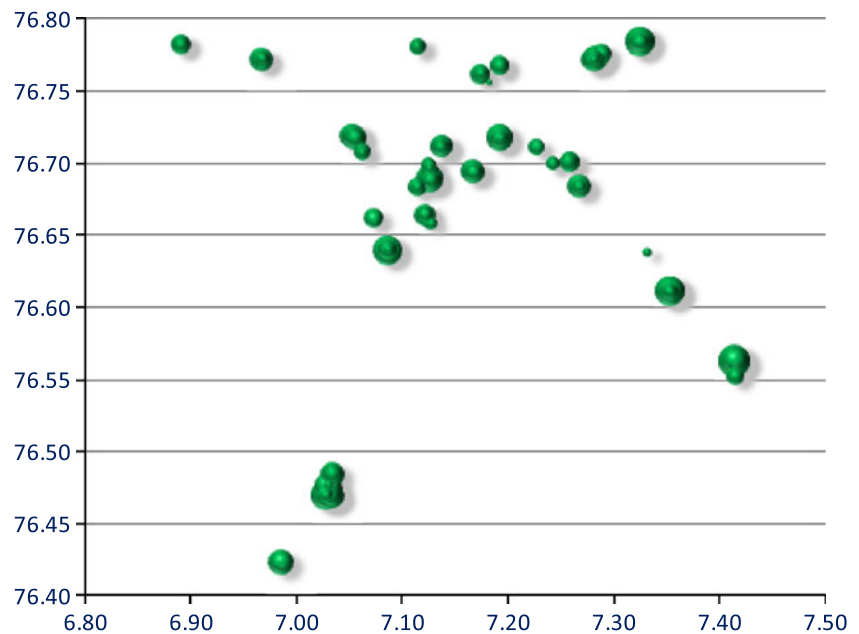


Fig. 4 Global spatial autocorrelation of the 220 farmer selections of cacao from Juanjui, Peru based on a dataset of 15 SSR loci. The Y axis is the spatial genetic correlation coefficient (r) computed as a function of distance (Peakall and Smouse 2006). The dashed lines

represent the 95% confidence interval around the null hypothesis of no spatial structure (or random distribution) and the error bars represent the bootstrapped 95% confidence around r . Distance classes of 2 km were used

Fig. 5 Bubble plots of two-dimensional local spatial auto-correlation analyses of 220 farmer selections in Juanjui, Peru. The plot shows the entire study area with green dots indicating the geographic coordinates of each territory sampled. Bubbles surround territories with positive I_r values that fell within the 5% tails of the permuted distribution. The size of the bubble is proportional to the magnitude of I_r . In these plots significantly positive I_r values range from 0.156 to 0.413. Calculations of I_r were based on sampling six nearest neighbors



the center of illicit coca production in Peru. Although various promising crops are being promoted for alternative development, cacao is the priority crop for local farmers because of its market opportunity. Cacao is primarily produced by small scale farmers and majority of the farms are smaller than two hectares (Ministerio de Agricultura, Peru 2006)

Plant Materials

A total of 220 farmers' selections were sampled from fields with full-grown cacao trees in Juanjui (Fig. 1; Table 1). The selection was coordinated by "Instituto de Cultivos Tropicales" (ICT), a non-profit and non-governmental organization promoting the sustainable agricultural development in the Peruvian Amazon region. Farmer participatory approaches were applied in identifying these selections. The trees are about 20 years old. Selection was based on years of observation for productivity and disease resistance by the local farmers.

To assess the genetic composition of the farmer selections and their inter-relationship with the candidate parental clones, we included 27 accessions that had been used as parental clones in the previous cacao breeding program in Peru (Arevalo, personal communication). These clones also served as reference genotypes in data analysis (Table 1).

The samples used for DNA fingerprinting profiles were leaves of various ages collected from individual cacao trees on local farms. Each sampled branch was tagged for

potential revisiting. Major agronomic data were recorded, following the protocol developed for the ICT farmer field school training program. Four healthy young leaves were collected from each tree, and the samples were air dried and sent to the USDA Beltsville Agricultural Research Center, Maryland, USA for genotyping.

DNA was extracted, according to the manufacturer's recommendation, with the DNeasy Plant System (Qiagen Inc., Valencia, CA, USA). Modifications were made to impede covalent reactions with the high level of endogenous phenolics in the cacao leaf samples (Saunders et al. 2004; Zhang et al. 2006a).

SSR Analysis

The development and selection of the cacao microsatellite markers has been previously described (Lanaud et al. 1999; Saunders et al. 2004). PCR was performed as described in Saunders et al. (2004), using commercial hot-start PCR supermixes that had been fortified by the addition of 30 U of the respective hot-start Taq DNA polymerase (Invitrogen Platinum Taq, Carlsbad, CA; Eppendorf HotMaster Taq, Brinkman, Westbury, NY) to each ml of the supermix. The amplified PCR products were separated by capillary electrophoresis as previously described (Saunders et al. 2004) using a CEQ™ 8000 genetic analysis system (Beckman Coulter Inc.). Data analysis was performed using the CEQ™ 8000 Fragment Analysis software version 9.0 according to manufacturer's recommendations (Beckman Coulter Inc). Sizes of SSR (Simple Sequence Repeat) fragments were

automatically calculated to two decimal places by the CEQ™8000 Genetic Analysis System. Allele-calling was performed using the CEQ™8000 binning wizard software (CEQ™8000 software version 9.0, Beckman Coulter Inc.), and edited based on the bin list using a SAS program (SAS, 1999).

Data Analysis

Summary statistics for each marker locus were computed using PowerMarker v. 3.0 (Liu and Muse 2005). For the purpose of this present study, we define genotype diversity as the percentage of unique genotypes in the farmer selections. Gene diversity was defined by the allele richness and expected heterozygosity. To compare the allele richness between the farmer selections and the five wild populations from Loreto, Peru (316 genotypes in total), we estimated unbiased allelic richness using the rarefaction approach as implemented in the computer program HP-Rare (Kalinowski 2005).

To assess the relationship among the individual selections, pair-wise Euclidean distance was computed for every pair of accessions using the genetic distance procedure in GenAlEx 6.0 (Peakall and Smouse 2006). The same program was used to perform Principal Coordinates Analysis (PCoA), based on the pair wise distance matrix. Both distance and covariance were standardized. Twenty-seven international accessions that have served as parental clones in the cacao breeding history of Peru were included in the Principal Coordinates Analysis.

For the analysis of population structure and inference of admixed ancestry (hybrids or ancestral forms), we used a model-based clustering method implemented in the software program STRUCTURE (Pritchard et al. 2000). The number of clusters (K -value) was set from 2 to 11 and the analysis was carried out without assuming any prior information about the genetic group or geographic origin of the samples.

Ten independent runs were assessed for each fixed number of clusters (K). The ΔK value was computed to detect the most probable number of clusters (Evanno et al. 2005). The run with the highest $\ln Pr(X|K)$ value of the 10 was chosen and presented as bar plots. Accessions possessing >50% membership (Q -value) in their own cluster were considered as a member of that cluster (genetic group).

Parentage analysis was applied to verify the parentage contribution of the twenty seven cacao clones that were used in production of hybrids in the Huallaga valley during the 1980s. A likelihood-based method implemented in the program CERVUS 3.0 (Marshall et al. 1998; Kalinowski et al. 2007) was used for computation. For each parent–offspring pair, the natural logarithm of the likelihood ratio

(LOD score) was calculated. Critical LOD scores were determined for the assignment of parentage to a group of individuals without knowing the maternity or paternity. Simulations were run for 10,000 cycles, assuming that 80% of candidate parents were sampled and a total of 80% of loci were typed, with a typing error rate of 1%. The most probable single mother (or father) for each offspring was identified on the basis of the critical difference in LOD scores (Δ) between the most likely and next most likely candidate parent at greater than 95% or 80% confidence (Marshall et al. 1998; Kalinowski et al. 2007).

To assess if there is spatial genetic structure (SGS), we tested both global and local SGS using the spatial procedure implemented in GenAlEx 6.0 (Peakall and Smouse 2006). For the test of global SGS, the significance of autocorrelation coefficient (r) was tested by constructing a classic 2-tailed 95% CI around the null hypothesis of no SGS ($r=0$) by performing 999 random permutations of genotypes among geographic locations (Peakall and Smouse 2006). For the test of local SGS, the 2-dimensional local spatial analysis (Smouse and Peakall 1999; Peakall and Smouse 2006) was applied. Permutation testing and a 1-tailed test (to detect only positive SGS) at $p=0.1$ was used. The 2D LSA was computed for four, six, eight, and 10 nearest neighbors, but we present only the results for six nearest neighbors, considering the average number of selections is approximately five from each farm.

Acknowledgements We thank Stephen Pinney and Eric Tillson for their contributions to the SSR genotyping; Tulio Pisco-Rojas, Henry Yalta, Katerine Rengifo, Janet Gonzales, Cecilia Ortiz and Kennet Reategui for their technical support in the taking samples from fields. We also thank Lucinda Vela Vargas for her help in the data organization and all the farmers that participated in the identification of trees and on farm interview. Special thanks are due to Drs. Sue Mischke, Ainong Shi and two anonymous reviewers who reviewed the manuscript and made critical suggestions for revision. This work was supported in part by the INCAGRO/MINAG/PERU project and NAS-Peru.

References

- Almekinders CJM, Elings A (2001) Collaboration of farmers and breeders: participatory crop improvement in perspective. *Euphytica* 122:425–438
- Alverson WS, Whitlock BA, Nyffler R et al (1999) Phylogeny of the core Malvales: evidence from *ndhF* sequence data. *Am J Bot* 86:1474–1486
- Arévalo E, García L, Krauss U, et al. (1999) Mejoramiento genético para el control de enfermedades del cacao en el Perú. In: Proceeding of the International Workshop on the Contribution of Disease Resistance to Cocoa Variety Improvement, 24th–26th, November, 1996. Bahía, Brazil: INGENIC, 127–133
- Bartley BGD (2005) The genetic diversity of cacao and its utilization. CABI International, Wallingford, Oxfordshire, OX10, 8DE, UK: CABI Publishing

- Bayer C, Fay MF, De Bruijn PY et al (1999) Support for an expanded family concept of Malvaceae within a circumscribed order Malvales: a combined analysis of plastid *atpB* and *rbcL* DNA sequences. *Bot J Linn Soc* 129:267–303
- Bellon MR (2004) Conceptualizing interventions to support on-farm genetic resource conservation. *World Dev* 32:159–172
- Brush SB (2000) The issues of in situ conservation of crop genetic resources. In: Brush S (ed) *Genes in the field*. IDRC/IPGRI/Lewis Publishers, Ottawa, 300 pp
- Cheesman EE (1944) Notes on the nomenclature, classification and possible relationships of cocoa populations. *Trop Agric* 21:144–159
- Clement CR, de Cristo-Araújo M, d'Eeckenbrugge GC et al (2010) Origin and domestication of native Amazonian crops. *Diversity* 2:72–106
- Cuatrecasas J (1964) Cacao and its allies. A taxonomic revision of the genus *Theobroma*. Contributions from the United States National Herbarium, vol 35. Smithsonian Institution Press, Washington, DC, pp 375–614
- De la Cruz M, Whitkus R, Gomez-Pompa A et al (1995) Origins of cacao cultivation. *Nature* 375:542–543
- Dias LAS (2001) Origin and distribution of *Theobroma cacao* L: a new scenario. In: Dias LAS (ed) *Genetic Improvement of Cacao*. Available at: <http://ecoport.org/ep?SearchType=earticleView&earticleId=197&page=2>
- Eskes AB (2006) Collaborative and participatory approaches to cocoa germplasm evaluation and selection. In: Denamany G, Lamin KH, Albert L, Maisin N, Ahmad AC, Saripah B, Yazik NM (ed) *Sustainable cocoa economy through increase in productivity, efficiency and quality: Proceedings of 4th Malaysian International Cocoa Conference (MICC) 2005*. Kuala Lumpur, Malaysia, 18th–19th July 2005. Kota Kinabalu: MCB, pp. 77–88
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol Ecol* 14:2611–2620
- Evans HC, Krauss U, Rutz RR et al (1998) Cocoa in Peru. *Cocoa Grow Bull* 51:7–22
- Evvett IW, Weir BS (1998) *Interpreting DNA Evidence*. Sunderland, MA: Sinauer Associates, Inc
- Henderson JS, Joyce RA, Hall GR, Hurst WJ (2007) Chemical and archaeological evidence for the earliest cacao beverages. *Proc Natl Acad Sci USA*, 10.1073
- Hernandez TA (1991) Cacao-Sistemas de Producción en la Amazonía Peruana, Programa de Promoción Agroindustrial AD/PER/459 UNFDAC-PNUD/OSP, Tingo María
- Kalinowski ST (2005) HP-Rare: a computer program for performing rarefaction on measures of allelic diversity. *Mol Ecol Notes* 5:187–189
- Kalinowski ST, Taper ML, Marshall TC (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Mol Ecol* 16:1099–1006
- Lanaud C, Risterucci AM, Pieretti I et al (1999) Isolation and characterization of microsatellites in *Theobroma cacao* L. *Mol Ecol* 8:2141–2143
- Leakey RRB, Akinnifesi FK (2007) towards a domestication strategy for indigenous fruit trees: clonal propagation, selection and the conservation and use of genetic resources. In: *Indigenous Fruit Trees in the Tropics: Domestication, Utilization and Commercialization*. In: Akinnifesi FK, Leakey RRB, Ajayi OC, Silési G, Tchoundjeu Z, Matakala P, Kwesiga FR (eds) *World Agroforestry Centre: Nairobi*. CAB, Wallingford
- Liu K, Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* 21:2128–2129
- Lockwood C, End M (1993) History, technique and future needs for cacao collection. In: *Proceedings of the International Workshop on Conservation, Characterization and Utilization of Cocoa Genetic Resources in the 21st Century*. Port-of-Spain, Trinidad and Tobago: The University of the West Indies, Cocoa Research Unit, 1–14
- Loor RG, Risterucci AM, Fouet O et al (2008) Tracing the native ancestors of the modern *Theobroma cacao* L. population in Ecuador. *Tree Genetics & Genomes* 5:1614–2942
- Love B, Spaner D (2007) Agrobiodiversity: its value, measurement, and conservation in the context of sustainable Agriculture. *Sustain Agric* 31:53–82
- Marshall TC, Slate J, Kruuk LEB et al (1998) Statistical confidence for likelihood-based paternity inference in natural populations. *Mol Ecol* 7:639–655
- Ministerio de Agricultura, Peru (2006) *Manual del Cultivo del Cacao Programa para el Desarrollo de la Amazonia*. Lima, Peru
- Motamayor JC, Lopez PA, Ortiz CF, Moreno A, Lanaud C (2002) Cacao domestication I: the origin of the cacao cultivated by the Mayas. *Heredity* 89:380–386
- Motamayor JC, Lachenaud P, de Silva e Mota JW, Loor R, Kuhn DN, Brown JS, Schnell RJ (2008) Geographic and genetic population differentiation of the Amazonian chocolate tree (*Theobroma cacao* L). *PLoS ONE* 3(10):e3311
- Peakall R, Smouse PE (2006) Genalex 6.0: genetic analysis in Excel. Population genetic software for teaching and research. *Mol Ecol Notes* 6(1):288–295
- Pound FJ (1938) Cacao and witchbroom disease (*Marasmius perniciosus*) of South America. *Arch Cocoa Res* 1:20–72
- Pound FJ (1945) A note on the cocoa population of South America. In: *Report and Proceedings of the 1945 Cocoa Conference*. London: 131–133
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure from multilocus genotype data. *Genetics* 155:945–959
- SAS (1999) SAS Version 8.02: SAS/STAT Software: changes and enhancements through Release 8.02. SAS Institute Inc., Cary, NC
- Saunders JA, Mischke S, Leamy EA, Hemeida AA (2004) Selection of international molecular standards for DNA fingerprinting of *Theobroma cacao*. *Theor Appl Genet* 110:41–47
- Smith NJH (1999) *The Amazon River Forest: a natural history of plants, animals, and people*. Oxford University Press, New York
- Smouse PE, Peakall R (1999) Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. *Heredity* 82:561–573
- United Nations Office on Drugs and Crime (UNDCP) & The United Nations Office for Project Services (UNOPS) (1996) *Program report: Semilleros y Bancos de Germoplasma de Cacao, Perú*
- Witcombe JR, Joshi A, Joshi KD, Sthapit BR (1996) Farmer participatory crop improvement. In: *Varietal selection and breeding methods and their impact on biodiversity*. *Exp Agric* 32:445–460
- Wood GAR, Lass RA (eds) (2001) *Cocoa, tropical agricultural series, 4th edn*. Longman Scientific and Technical and John Wiles & Sons, Inc, London, England and New York
- Wright S (1965) The interpretation of population structure by *F*-statistics with special regard to systems of mating. *Evolution* 19:395–420
- Young AM (1994) *The chocolate tree: a natural history of cacao*. Smithsonian Institution Press, Washington
- Zhang D, Mischke S, Goenaga R et al (2006a) Accuracy and reliability of high-throughput microsatellite genotyping for cacao clone identification. *Crop Sci* 46:2084–2092
- Zhang D, Arevalo-Gardini E, Mischke S et al (2006b) Genetic diversity and structure of managed and semi-natural populations of cacao (*Theobroma cacao*) in the Huallaga and Ucayali valleys of Peru. *Ann Bot* 98:647–655
- Zhang D, Boccarda M, Motilal L et al (2009) Molecular characterization of an earliest cacao (*Theobroma cacao* L.) collection from Upper Amazon using microsatellite DNA markers. *Tree Genetics and Genomes* 5:595–607