Genetic Identity, Ancestry and Parentage in Farmer Selections of Cacao from Aceh, Indonesia Revealed by Single Nucleotide Polymorphism (SNP) Markers

**Lukman · Dapeng Zhang · Agung W. Susilo · Diny Dinarti ·
Bryan Bailey · Sue Mischke · Lyndel W. Meinhardt**

Abstract Cacao (*Theobroma cacao* L.), the source of cocoa powder and cocoa butter used for chocolate, originated in the rainforests of South America. Indonesia is the 3rd largest cocoa producer in the world with an annual cocoa output of 0.84 million tons. The main objective of this study was to examine the genetic background of superior farmer-selected clones from farmers' fields, in order to provide a scientific basis to support cacao rehabilitation in Aceh, Indonesia. The pilot experiment assessed 136 genotypes, including 80 Aceh farmer selections and 56 reference international clones, using 53 single nucleotide polymorphism (SNP) markers. Principle coordinates analysis revealed that the Aceh farmer selections were largely hybrids between Upper Amazon Forastero and Trinitario. Bayesian clustering analysis further specified the germplasm groups of the ancestral contributors. Of the 80 farmer selections, parentage analysis identified 30 parent-offspring relationships contributed by 16 progenitors. Spatial genetic analysis showed no significant global spatial correlation, but local spatial correlation was detected. These results suggest that only a fraction of cacao germplasm from Upper Amazon has been incorporated into farmers' fields, despite the use of diversie progenitors in breeding programs and seed gardens. This relatively narrow genetic background was likely due either to limited access to diverse hybrids by local farmers or to the over-emphasis on selecting for yield and morphological appearance, such as pods and bean size. Based on the results, a subset of farmer selections with diverse genetic background are chosen and evaluated in field trials.

Keywords Cacao Landraces · Chocolate · Conservation · Germplasm · Genetic diversity · Molecular markers · Mesoamerica · *Theobroma cacao* · Tropical tree

Lukman
Faculty of Agriculture, Department of Agronomy, Malikussaleh University, Lhokseumawe 24351, Aceh, Indonesia

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

A. W. Susilo
Indonesian Coffee and Cocoa Research Institute (ICCRI), Jl. PB. Sudirman No. 90, Jember 68118, Indonesia

D. Dinarti
Faculty of Agriculture, Department of Agronomy and Horticulture, Bogor Agricultural University, Bogor 16680, Indonesia

Introduction

Cacao (*Theobroma cacao* L.), a member of the Malvaceae family (Alverson et al. 1999; Bayer et al. 1999), is a tropical forest species native to South America that is cultivated extensively as the source of cocoa butter and cocoa powder for the confectionery industry. Annual worldwide production of cocoa, the product obtained from dried fermented cacao seeds, is four million tons (International Cocoa Organization 2010), of which 90-95% is produced by smallholder farmers in tropical developing countries (ICCO 2010).

Commercial cacao cultivation in Indonesia started in the early 1900s, although the crop was introduced from Venezuela as far back as the 16th century (van Hall 1914; Toxopeus and Geisberger 1983). Presently, Indonesia is the third largest cacao producer in the world after Côte d'Ivoire and Ghana, with an annual cocoa production of 844,626 tons (FAOSTAT 2010). The production regions are mainly distributed in the lowland wet temperate zone in Sulawesi, Sumatra, Nusa Tenggara, Java, Kalimantan, Maluku and Irian Jaya. Among these regions, Aceh is the westernmost province in Indonesia

with 73,000 hectares of cacao and an annual output of 23,840 tons. Cacao production in the Aceh region could potentially be expanded; however, low productivity, high labor cost, and lack of consistency in quality are constraints to growth. The average yield of dry beans in Aceh is only 328 kg/ha, whereas the average yield in the rest of Indonesia is 602 kg/ha (SWISSCONTACT, unpublished document). The use of low-yielding varieties combined with old cacao trees is one of the main factors contributing to low productivity. Diseases and pests, especially black pod disease (*Phytophthora palmivora*) and cacao pod borer (*Conopomorpha cramerella*), are the two main biological constraints for cacao production in Aceh. Vascular streak dieback (VSD - *Oncobasidium theobromae*), which is the number one fungal disease constraint in Sulawesi and Java, has not been found in Aceh.

Cultivated cacao has been traditionally subdivided into three main groups: (Criollo, Forastero and Trinitario) although other distinctive categories were recognized (Cheesman 1944; Wood and Lass 1985). Among the three main groups, Criollo cacao was domesticated more than 3,000 years ago in Mesoamerica (Powis et al. 2011; Henderson et al. 2007; Motamayor et al. 2002). The term Forastero means foreign and refers to introduced trees that are not Criollo. Forastero encompasses a diverse range of South American populations, each having a distinctive genetic identity (Bartley 2005; Motamayor et al. 2008; Zhang et al. 2009, 2012). The devastating impact of witches' broom disease in the 1920s in tropical America led to expeditions to collect disease-resistant germplasm from the Upper Amazon region (Pound 1938; Bartley 2005), which resulted in a substantial number of accessions of wild cacao known as "Upper Amazon Forastero" (UAF). Since 1950, various traits of resistance/tolerance to biotic and abiotic stresses offered by the UAF germplasm have been incorporated in breeding programs worldwide, resulting in improved adaptability in major cacao producing regions (Dias 2001). In Indonesia clones from UAF genetic groups Scavina (SCA), Parinari (PA), Iquitos Mixed Calabacillo (IMC) and Nanay (NA) were used in breeding programs and in various seed gardens. Each genetic group has particular clones offering unique resistance to black pod disease and vascular streak dieback disease and showing variation in other key agronomic traits such as yield potential, bean size and pod shape. Hybrids of these Upper Amazon Forasteros have been introduced from Malaysia since the mid-1970s and have been adopted by Indonesia farmers (Mawardi et al. 1994; Susilo et al. 2011).

Starting in 2009 the Indonesian government launched a national program to improve cacao production and quality. Rejuvenation and rehabilitation using high-yielding, disease-resistant clones was one of the main approaches to increase cacao productivity in Aceh. Supported by the Indonesian government and the international community, initially through the disaster relief and reconstruction effort following the 2004 tsunami, superior clones were propagated through side-grafting and chupon-grafting. In spite of significant progress in rejuvenation and rehabilitation, the number of available clones introduced from outside Aceh was limited. However, a large number of locally grown cacao trees in Aceh provided an alternative source of germplasm for farmer participatory selection.

The genetic background of farmer selections from Aceh was presumably very diverse but detailed information is not available. Genetic insight from analysis of local farmer selections will be useful for the testing and selection of promising trees in the on-going rehabilitation and rejuvenation programs in Aceh. Therefore, the main objective of the present study was to identify the genetic identity, ancestry and parentage of the farmer selections. Based on the results of molecular characterization, a subset of these selections can be identified for field evaluation. Moreover, understanding the impact of the introduced international germplasm in farmers' fields, especially the UAF clones that harbor disease-resistance, is highly useful for future use of promising progenitors in breeding and hybrid production through seed gardens in Indonesia

Materials and Methods

Plant Materials and DNA Sample Preparation

A total of 132 farmer selections were identified from six districts, namely Aceh Uara District of Aceh, Bireuen, Pidie Jaya, Aceh Barat Daya (Abdya), Aceh Tanggara (Agara) and Aceh Tamiang (Table 1). The GPS data of the samples was mapped using the computer program PhyloGeoViz (Tsai 2011; Fig. 1). The trees are 15–20 years old and selection was based on productivity and morphological characteristics observed by local farmers. The samples used for DNA fingerprinting profiles were leaves of various ages collected from individual cacao trees on local farms. Two 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. Cacao DNA was extracted from leaf tissue using the DNeasy Plant System (Qiagen Inc., Valencia, CA, USA) according to Saunders et al. (2004). Out of the 132 selections, only 80 yielded DNA of sufficient quality and quantity for genotyping and the other 50 selections were not used in this study due to the inability to acquire suitable DNA from these leaves. To assess the genetic affiliation with known cacao germplasm groups, DNA samples from 56 reference clones were included in this experiment. Selection of the reference clones was based on the criteria that a) they are important international clones; b) they are available in Southeast Asia and c) they have been used as progenitors in seed gardens and/or breeding programs. The genetic identities of these clones are known because of previous DNA fingerprinting of cacao germplasm through an

Table 1 List of 80 Aceh farmer selections and 56 reference clones genotyped using single nucleotide polymorphism (SNP) markers

| Accessions	Source & site	Accessions	Source & site	Accessions	Source & site
ABD2-1	Abdya	BIR2-12	Bireuen	NA 127	Peru
ABD2-2	Abdya	BIR2-13	Bireuen	NA 178	Peru
ABD2-3	Abdya	BIR2-14	Bireuen	NA 191	Peru
ABD2-4	Abdya	BIR2-15	Bireuen	NA 232	Peru
ABD2-6	Abdya	BIR2-16	Bireuen	NA 31	Peru
ABD2-7	Abdya	BIR2-17	Bireuen	NA 32	Peru
ABD2-9	Abdya	BIR2-18	Bireuen	NA 33	Peru
ABD2-10	Abdya	BIR2-19	Bireuen	NA 670	Peru
ABD2-11	Abdya	BIR2-20	Bireuen	NA 702	Peru
ABD2-12	Abdya	PJA1-1	Pidie Jaya	NA 710	Peru
ABD2-13	Abdya	PJA1-4	Pidie Jaya	POUND 7	Peru
ABD2-14	Abdya	PJA1-5	Pidie Jaya	PA 120	Peru
ABD2-15	Abdya	PJA1-7	Pidie Jaya	PA 124	Peru
ABD2-16	Abdya	PJA1-8	Pidie Jaya	PA 137	Peru
ABD2-18	Abdya	PJA1-9	Pidie Jaya	PA 150	Peru
ABD2-19	Abdya	PJA1-10	Pidie Jaya	PA 300	Peru
ABD2-20	Abdya	PJA1-14	Pidie Jaya	PA 310	Peru
AGR2-2	Tenggara	PJA1-15	Pidie Jaya	PA 39	Peru
AGR2-4	Tenggara	PJA1-16	Pidie Jaya	PA 7	Peru
AGR2-5	Tenggara	PJA1-17	Pidie Jaya	EET 397	Ecuador
AGR2-6	Tenggara	PJA1-18	Pidie Jaya	EET 399	Ecuador
AGR2-9	Tenggara	PJA1-19	Pidie Jaya	LCTEEN 241	Ecuador
AGR2-10	Tenggara	PJA1-20	Pidie Jaya	LCTEEN 302	Ecuador
AGR2-11	Tenggara	INT2-4	Pidie Jaya	LCTEEN 412	Ecuador
AGR2-12	Tenggara	INT2-10	Pidie Jaya	ICS 39	Trinidad
AGR2-13	Tenggara	AUT2-13	Pidie Jaya	ICS 60	Trinidad
AGR2-15	Tenggara	TAM2-1	Tamiang	ICS 95	Trinidad
AGR2-16	Tenggara	TAM2-3	Tamiang	ICS 43	Trinidad
AGR2-17	Tenggara	TAM2-5	Tamiang	ICS 63	Trinidad
AGR2-19	Tenggara	TAM2-6	Tamiang	ICS 75	Trinidad
AGR2-22	Tenggara	TAM2-9	Tamiang	UF 667	Costa Rica
AUT2-2	Aceh Utara	TAM2-11	Tamiang	UF 676	Costa Rica
AUT2-6	Aceh Utara	TAM2-14	Tamiang	UF 168	Costa Rica
AUT2-8	Aceh Utara	TAM2-15	Tamiang	UF 221	Costa Rica
AUT2-10	Aceh Utara	IMC 27	Peru	UIT 1	Indonesia
AUT2-11	Aceh Utara	IMC 30	Peru	GC 7	Indonesia
AUT2-12	Aceh Utara	IMC 47	Peru	GS 29	Grenada
BIR2-1	Bireuen	IMC 54	Peru	OC 77	Venezuela
BIR2-2	Bireuen	IMC 6	Peru	SIAL 325	Brazil
BIR2-3	Bireuen	IMC 60	Peru	AMELONADO 15	Brazil
BIR2-4	Bireuen	IMC 67	Peru	AMELONADO 22	Brazil
BIR2-5	Bireuen	IMC 77	Peru	CRIOLLO 13	Costa Rica
BIR2-7	Bireuen	MO 121	Peru	CRIOLLO 22	Trinidad
BIR2-8	Bireuen	MO 96	Peru		
BIR2-9	Bireuen	SCA 12	Peru		
BIR2-10	Bireuen	SCA 6	Peru		
		UCAYALI 70	Peru		
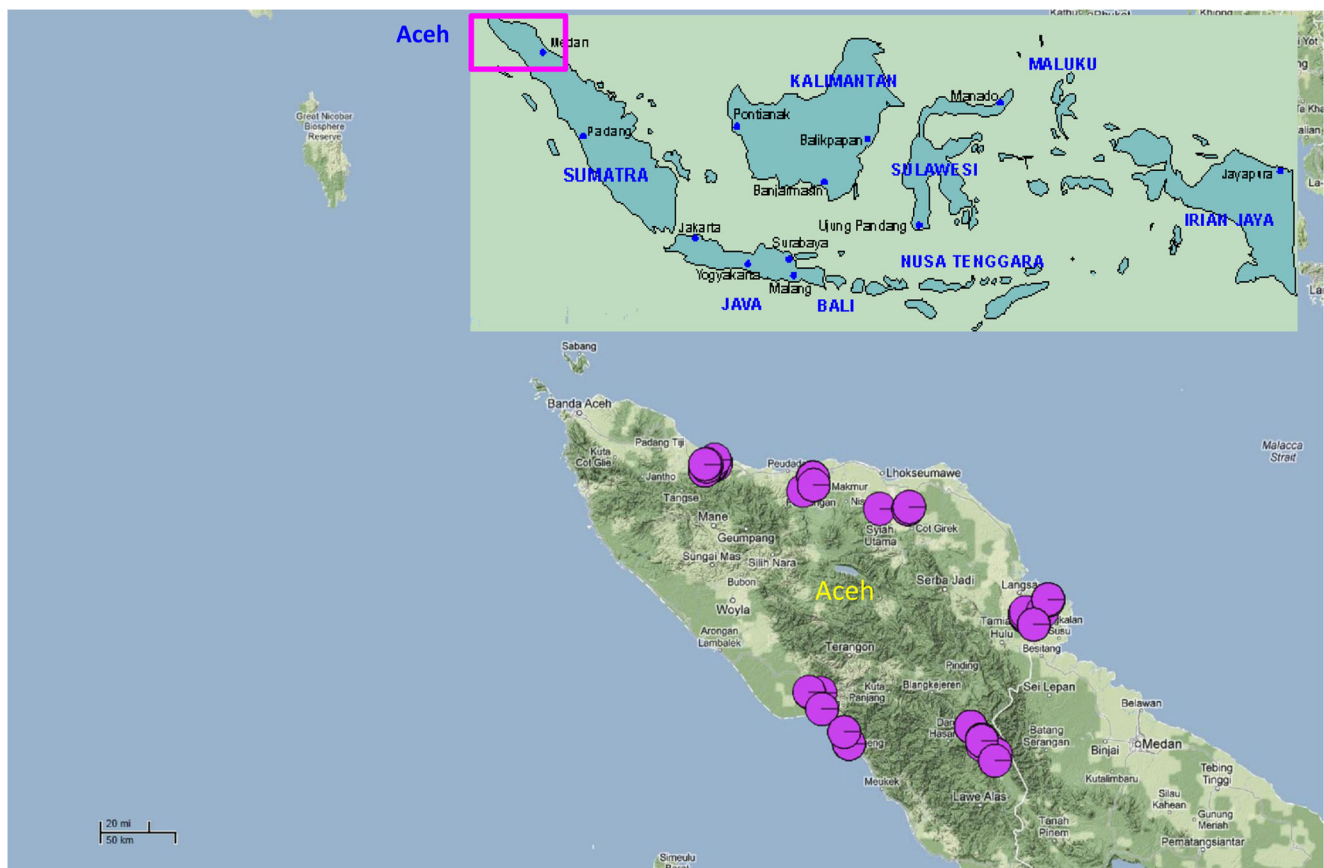
Fig. 1 Geographical sites where farmer selection of cacao were collected in Aceh, Indonesia

international initiative (Zhang et al. 2006). The majority of the accessions are maintained in the international cacao genebanks at the University of West Indies, Trinidad and Tobago and Centro Agronómico Tropical de Investigación y Enseñanza (CATIE) in Turrialba, Costa Rica (Bocarra and Zhang 2006; Zhang et al. 2009; Motilal et al. 2010).

SNP Markers and Genotyping

Fifty-three SNP markers were selected from 1,560 candidate SNPs developed from cDNA sequences from a wide range of cacao organs (Allegre et al. 2011; Argout et al. 2008). The selection was based on the level of polymorphism and their distribution across the ten chromosomes in cacao. SNP genotyping was performed at the Human Genetics Division Genotyping Core facility, Washington University, St. Louis, using MALDI-TOF mass spectrometry (product of Sequenom Inc.; http://hg.wustl.edu/info/Sequenom_description.html).

Data Analysis

Key descriptive statistics for measuring informativeness of the 53 SNP markers were calculated, including minor allele frequency, observed heterozygosity, expected heterozygosity,

and probability of identity (Evett and Weir 1998; Waits et al. 2001). The program GenAlEx 6.0 (Peakall and Smouse 2006) was used for computation. For clone or duplicate identification, pair-wise multi-locus matching was applied among individual varieties and the reference clones, using the same program. Accessions with different names that were fully matched at the genotyped SNP loci were declared duplicates or synonymous accessions.

Multi-variant analysis was used to assess the relationship among the individual farmer varieties and their relationships with reference clones from international genebanks. 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 then used to perform Principal Coordinates Analysis (PCoA), based on the pair-wise distance matrix. Both distance and covariance were standardized.

For 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 80 farmer selections were analyzed, together with 51 reference clones that potentially had made ancestral contribution to these farmer selections. The reference clones of ancient Criollo and Amelonado were excluded in this analysis. The number of

clusters (K-value) was set from 2 to 10 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 10 was chosen and presented as bar plots. Q-value was used to present the ancestral contribution (membership) from each germplasm group. Accessions possessing ≥25 % membership (Q-value) in a given cluster were considered as receiving a significant ancestry contribution from that cluster (genetic group). Accessions possessing ≥75 % membership were considered to be a member of that cluster. Accessions possessing >25 % but <75 % membership were considered as hybrids of two (or more) clusters.

Parentage analysis was applied to assess the direct parentage contribution from the putative parental clones to these farmer selections. The reference clones, the majority of which had been used as progenitors in seed gardens or breeding programs in Indonesia, were used as candidate parents, whereas the farmer selections were used as offsprings. 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 60 % of candidate parents were sampled, 90 % of the loci were typed. 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).

Results

Genotype and Gene Diversity in Farmer's Fields

A high level of genotype diversity was observed among the 80 farmer selections from Aceh. SNP data showed that each of the 80 farmer selections is genetically distinctive and there were no duplicates (or clones) among the selections. Moreover, none of the 80 trees matched with the introduced international or local clones used as reference in this study. The farmer selections not only had high genotype diversity, but also had a high level of heterozygosity (Table 2). The observed and expected heterozygosity in the Aceh farmer selections are 0.349 (SE=0.018) and 0.372 (SE=0.016)

Table 2 Sample Size, Information Index, Observed Heterozygosity, Expected Heterozygosity, and Fixation Index in Aceh farmer selections and reference clones of cacao

Pop		N	I	Ho	He	F
Aceh farmer selections	Mean	80	0.553	0.349	0.372	0.066
	SE		0.018	0.018	0.016	0.031
Reference clones	Mean	56	0.592	0.320	0.404	0.216
	SE		0.014	0.015	0.012	0.030

respectively, which is comparable with those in the reference international clones of diverse origin. The inbreeding coefficient is only 0.066 (SE=0.031 among the farmer selections (Table 2). suggesting that all the farmer selections were derived from hybrid seed families, and clone propagation was rare in Aceh. The result is also compatible with the local farmers' practice of generating seedlings from the superior trees on their own farms. It is known that cloning and grafting were not commonly used by farmers until recently introduced 2 years ago. Cacao pods from a productive tree had a higher chance than budwood of being a mode of dispersal, because seeds have a better chance of survival.

Principle Coordinate Analysis

The genetic relationships among the farmer selections and reference clones were shown by Principal Coordinates Analysis (Fig. 2). The plane of the first three main PCO axes accounted for 31.90, 25.0 and 15.7 % of total variation, respectively. The 60 reference clones were clustered in six groups, which matched well with their known classification of cacao germplasm groups. The first group is the ancient Criollo, which is highly differentiated from the rest of the accessions. The second group is a combination of wild trees that were originally from Iquitos, Peru (Pound 1938). Among the NANAY and IMC clones commonly used as parental clones for seed gardens and for breeding in Southeast Asia, were NA 31, NA 32, NA 33 and IMC 67. The third group is also comprised of wild Peruvian trees, however these are from the Ucayali River and include SCA 6 and SCA 12. These are the most widely used sources of disease resistance, particularly for black pod, witches' broom and vascular streak dieback diseases. The fourth group is exclusive to the Parinari clones and includes PA 300, PA 150, PA 37 and PA 7, some of the most commonly used parental clones in Southeast Asia. The fifth group is the Amelonado group, which originated in the lower Amazon. The final group consists of all Trinitario clones, represented by several important progenitors used in the seed gardens in Indonesia. These progenitors included well-known reference Trinitario varieties from Trinidad (GC 7, GS 29, ICS 39, ICS 43, ICS 60, ICS 95), Costa Rica (UF 168, UF
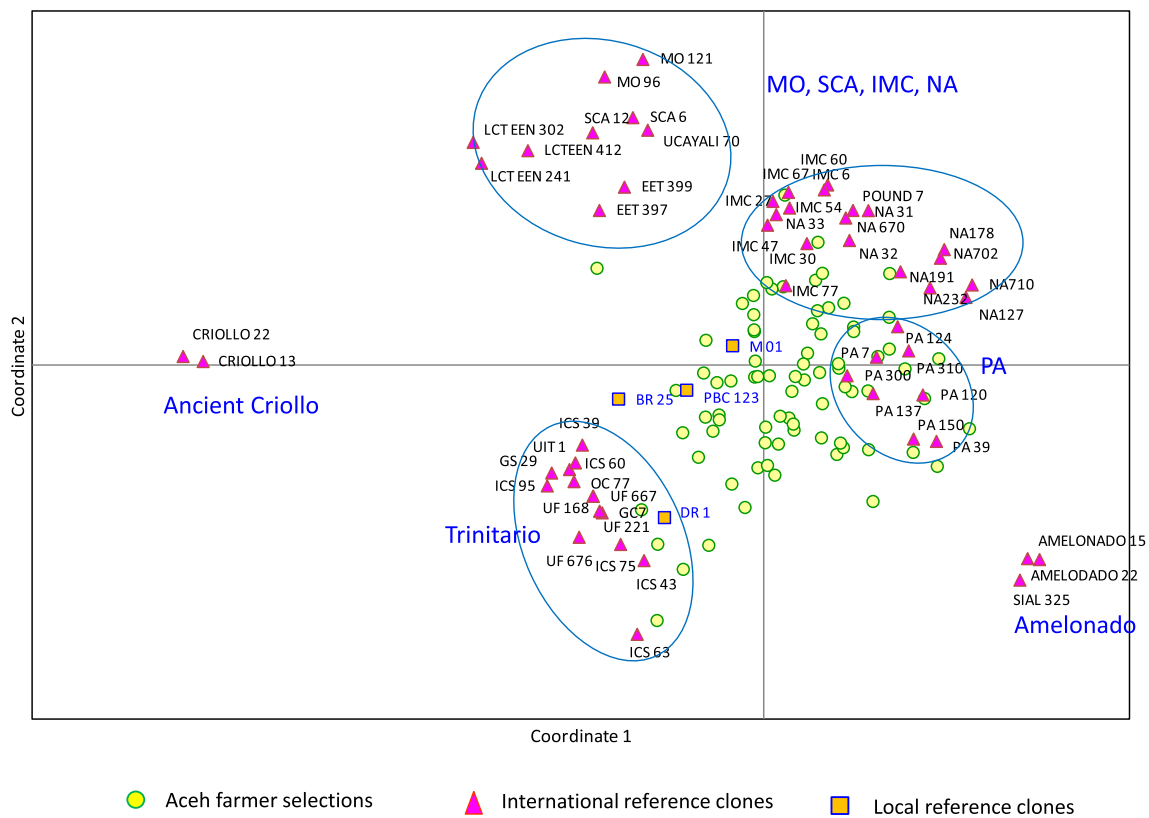
Fig. 2 PCoA plot of 136 cacao accessions, including 80 accessions of Aceh farmer selections and 56 reference accessions of international and local cacao clones. First axis=31.9 % of total information and the second=25.0 %

221, UF 667, UF 676) and Venezuela (OC 77). The majority of the Aceh farmer selections were distributed among the reference groups, indicating their background as hybrids of these groups. A substantial number of the farmer selections overlapped with the Parinari, IMC and NA, as well as the Trinitario group, suggesting their close genetic relationships (Fig. 2).

Ancestry Inference and Parentage Analysis

The result of Bayesian cluster analysis largely agreed with the distance-based multivariate analysis. Based on the value of ΔK (Evanno et al. 2005), the 51 reference accessions were grouped into four most probable clusters representing the four main clusters in the same pattern as in PCoA (Fig. 3), which includes the Trinitario, IMC & NA, MO & SCA and Parinari (Fig. 3). On average, the four clusters have a coefficient of membership (Q-value) of 0.912. A Q value of 0 corresponds to an individual of purely exogenous origin, whereas a value of 1 is purely from a home cluster (Fig. 3).

This result also reveals that ancestry in these farmer selections is mainly hybrid. Among the 80 farmer selections, only 11 genotypes can be classified as single ancestral origin (Q-value ≥0.75), of which six were Trinitario and five were Parinari. The remaining sixty-nine selections

were classified as interpopulation hybrids, which showed combinations of ancestry, ranging from two to four reference groups (Figs. 2 and 3). The Parinari germplasm group made the largest ancestral contribution to these farmer selections; 34 selections were found to have a significant PA ancestry (Q-value ≥25 %). The second largest contribution to these farmer selections was from the Trinitario group, with significant proportion (Q-value ≥25 %) of Trinitario ancestry detected in 22 farmer selections. In contrast, contributions from the MO/SCA and IMC/NA groups were much smaller. Significant ancestry from the IMC & NA group was found in 12 selections and the ancestry from the MO & SCA group was only found in five farmer selections (Fig. 4). This result is fully compatible with the previous result of PCoA and further demonstrates that a majority of the Aceh farmer selections are hybrids of Trinitario and Upper Amazon Forastero progenitors, especially those from Parinari group.

Parentage analysis shows that 15 known parental clones were responsible, at 80 % confidence level, for the maternity (or paternity) of 30 farmer selections in Aceh. When the confidence level is raised to 95 %, the number of identified parents is reduced to 13, which were shared by 16 farmer selections (Table 3). Among the 30 identified parent-offspring relationships, ten were associated with Trinitario parents, seven with IMC &
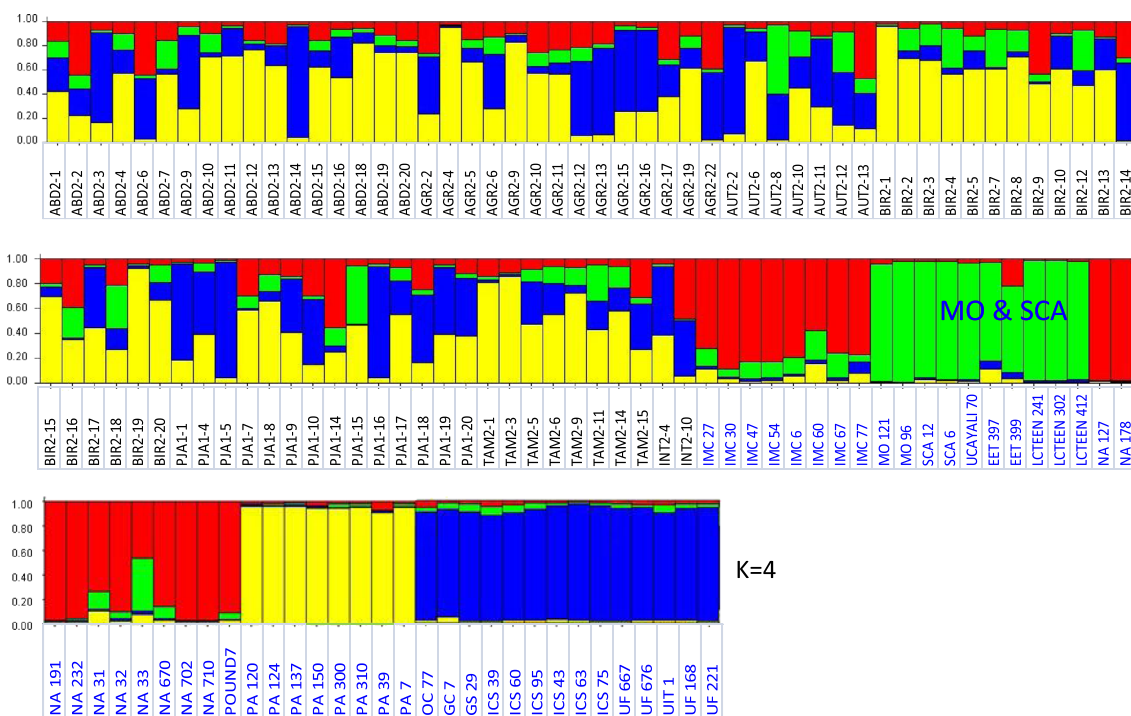
Fig. 3 Inferred clusters in the Aceh farmer selections and reference clones using STRUCTURE, where K is the potential number of genetic clusters that may exist in the overall sample of individuals. Each vertical line represents one individual multilocus genotype. Individuals with multiple colors have 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

NA, two with MO & SCA, and 11 with the Parinari group. The result of parentage analysis fully agreed with that of the Bayesian assignment test.

Fig. 4 Comparison of a number of farmer selections assigned to the four different ancestral clusters according to the result of Bayesian clustering analysis (IMC & Nanay, MO & SCA, Trinitario, Parinari). Q-value was used to present the ancestral contribution (membership) from each germplasm group. Accessions possessing ≥25 % membership (Q-value) in a given cluster were considered as a significant ancestry contribution to that cluster (genetic group). Accessions possessing ≥75 % membership were considered as a member of that cluster. Accessions possessing >25 % but <75 % membership were considered as hybrids of two (or more) clusters

Spatial Pattern and Autocorrelation

Mantel tests did not detect significant correlation between genetic and geographic distances (R_{xy}=0.027, P=0.298). Nor did global autocorrelation analysis of the 80 farmer selections find significant positive correlation. However, using the method of Two-Dimensional Local Spatial Autocorrelation Analyses (2D LSA), significant local correlations (lr) were detected in 19 farmer selections, based on a one-tailed test. Significantly positive local correlation ranged from 0.107 to 0.262 (Fig. 5). The result indicated that a small fraction of genetically similar individual trees aggregated in the same farm or communities.
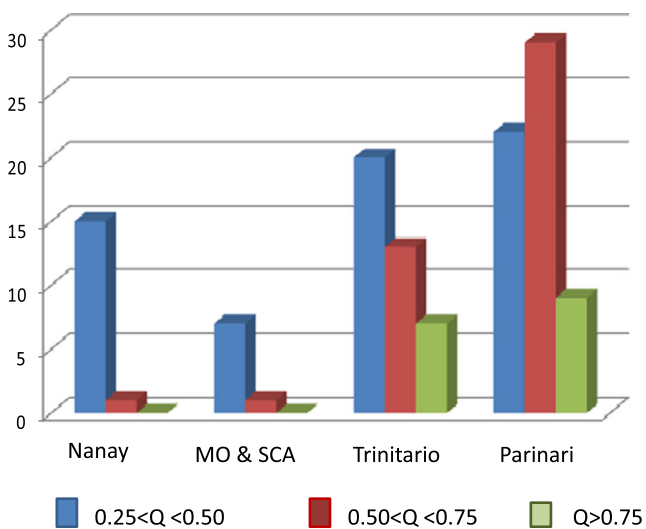
Discussion

Single nucleotide polymorphisms (SNPs) are the most abundant class of polymorphisms in plant genomes (Buckler and Thornsberry 2002). Compared with SSR markers, the assays of SNPs can be done without requiring DNA separation by size, and therefore can be automated in high-throughput assay formats. The diallelic nature of SNPs offers a much lower error rate in allele calling and raises the consistency in allele calling across laboratories. While SNP markers have been widely used in plant varietal identification in many other

Table 3 Likelihood assignment of 30 parent-offspring pairs in the Pound collection, based on 51 candidate parental clones

#	Parent	Offspring	LOD	Location	Farm
1	GS 29	ABD2-14	8.2	Abdya	Azman
2	ICS 39	BIR2-14	8.87	Bireuen	M. Sulaiman
3	ICS 39	ABD2-9	4.2	Abdya	Akmal Syah
4	ICS 63	ABD2-3	5.23	Abdya	Yusran
5	IMC 47	AUT2-13	3.89	Aceh Utara	Muslim
6	IMC 54	PJA1-14	4.84	Pidie Jaya	Nurlaila
7	IMC 67	BIR2-18	6.33	Bireuen	M. Sulaiman
8	IMC 67	ABD2-6	2.48	Abdya	Yusran
9	IMC 67	AGR2-22	2.07	Tenggara	Hasyim Syah
10	NA 670	BIR2-16	2.91	Bireuen	M. Sulaiman
11	NA 702	BIR2-9	4.14	Bireuen	Syarifuddin
12	OC 77	PJA1-1	3.13	Pidie Jaya	Amri
13	PA 124	PJA1-7	7.44	Pidie Jaya	M Ali
14	PA 124	BIR2-2	4.63	Bireuen	M. Sulaiman
15	PA 124	AGR2-5	3.72	Tenggara	Muhammad Yani
16	PA 124	BIR2-8	3.3	Bireuen	Syarifuddin
17	PA 124	ABD2-10	3.21	Abdya	Akmal Syah
18	PA 137	BIR2-3	5.64	Bireuen	Syarifuddin
19	PA 300	BIR2-7	2.9	Bireuen	Usman AR
20	PA 310	BIR2-1	11.06	Bireuen	M. Sulaiman
21	PA 310	BIR2-20	3.24	Bireuen	Ibrahim Umar
22	PA 39	ABD2-11	6.96	Abdya	Marjohan
23	PA 7	BIR2-19	2.58	Bireuen	M. Sulaiman
24	SCA 12	AUT2-8	11.18	Aceh Utara	Muslim
25	SCA 6	PJA1-15	9.19	Pidie Jaya	Nurlaila
26	UF 221	PJA1-5	11.36	Pidie Jaya	A Gani
27	UF 221	AUT2-2	5.24	Aceh Utara	Muslim
28	UF221	PJA1-16	2.7	Pidie Jaya	Badrol Munir
29	UF 676	INT2-10	7.02	Pidie Jaya	Nurlaila
30	UF 676	AGR2-15	3.21	Tenggara	Hasyim Syah

*Critical LOD (the natural logarithm of the likelihood) ratios for assignment of maternity are 4.98 at >95 % confidence and 2.04 at >80 % confidence

crops, the efficacy of using SNP markers for genotype identification and diversity assessment remains to be investigated in cacao. The present study shows that high-throughput genotyping using a small set of SNP markers is highly cost-effective and useful for a broad range of research and field applications, including identification of mislabeled accessions, parentage and sibship analysis, and examination of on-farm diversity.

Cacao germplasm used by Aceh farmers today can be traced back to several different points of introduction. The earliest introduction started in the 17th century during the Dutch colonial period. The latest introductions of improved materials, includes the government-assisted importation from seed gardens and seed growers outside of Aceh. In addition to
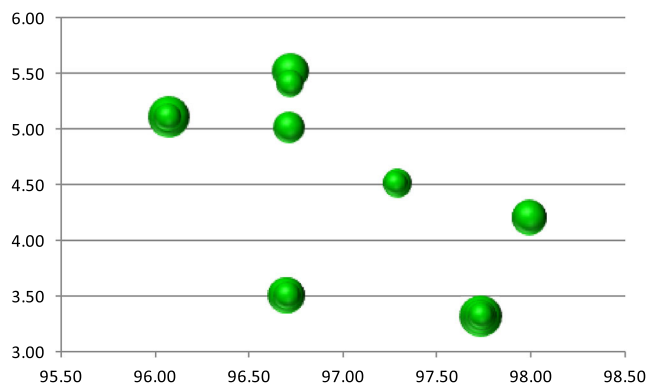

Fig. 5 Bubble plots of two-dimensional local spatial autocorrelation analyses of farmer selections in Aceh, Peru. The plot shows the entire study area, with Green dots indicating the geographic coordinates of each territory sampled. Bubbles surround territories with positive lr values that fell within the 5 % tails of the permuted distribution. The size of the bubble is proportional to the magnitude of lr. In these plots significantly positive lr values range from 0.107 to 0.262. Calculations of lr were based on sampling five nearest neighbors

the direct adoption of introduced planting materials, Aceh farmers also made selections from their own nurseries by using seeds from the best trees on their farms. The seedlings were usually raised and evaluated in the nursery and selections were subsequently planted in the fields.

The current result showed that a high level of genotype diversity exists in farmers' fields in Aceh, which is primarily due to the common practice of using different hybrid families as planting materials. The large number of hybrid trees in farmers' fields can be explored for participatory selection of superior clones in this low input, small-scale production system. The present results also showed that the current on-farm diversity of cacao in Aceh is largely built on the mixed foundation of Upper Amazon Forastero and Trinitario, as reflected by both Bayesian clustering analysis (Fig. 3) and PCoA (Fig. 2). However, among the three Upper Amazon Forastero germplasm groups revealed in this study, only one (Parinari) has made a significant impact in terms of ancestral or parental contribution to the Aceh farmer selections. Seventy-eight percent of the tested farmer selections can be classified as having significant Parinari ancestry (Q value >0.25). In contrast, the other two UAF groups, NA/IMC and MO/SCA/Ucayali, have made a much smaller ancestral or parental contribution. Only 12 and 15 % of the farmer selections have significant parentage from NA/IMC and MO/SCA respectively. These results showed that, in spite of the common use of parental clones from NA/IMC and MO/SCA germplasm groups in Indonesia's seed gardens and breeding programs, these germplasm groups were not well incorporated into the Aceh farmer selections. This disparity could be due to farmers' limited access to diverse planting materials. It is also possible that the local farmers' strong preference for large pods and large bean size may have affected the selection outcome. Like the Parinari group, the wild germplasm of

NA/IMC and MO/SCA were collected from the Peruvian Amazon in the 1930s. These genotypes are known for their resistance to diseases and adaptability in marginal production conditions. However, these clones don't have large pods and large beans. The PA group has more appealing pod and bean size attributes because this group was likely derived from introduced and selected material from Brazil (Bartley 2005). This preference for the exterior attractiveness of the pods and beans may undermine efforts to incorporate diverse disease resistances into the cacao plantings in the Aceh region of Indonesia.

No significant spatial structure was detected by the Mantel test and global spatial correlation among the farmer selections. However significant local spatial correlation was observed, indicating that farmers in the same region tend to adopt similar planting materials. Since records on previous cacao seed distribution are not available, we do not know if hybrid families were delivered to farmers in the same region or if local environmentally selected clones, with similar genetic backgrounds, were shared in the region. Sharing productive genotypes has been common among cacao growers in Indonesia, and pods from productive trees usually have a higher likelihood of being integrated into a neighbor's field. The recent rehabilitation effort may have reinforced this trend, where same hybrid families were delivered to and adopted by farmers in the same communities or villages.

Parentage analysis showed that only 30 % of the farmer selections could directly trace their parentage to the parental clones that were used by the Indonesia cacao breeding program and seed gardens. It's likely that a majority of the Aceh farmer selections were not the "F1" (first generation) hybrids directly taken from either seed gardens or breeding programs. These farmer selections were more likely the second or third generation hybrids, which were derived from more than one recombination of the parental clones. In fact, of the 80 selections, only five showed a two-way hybrid genotype (Fig. 3). Their genotypes were mainly formed by two large memberships (Q-values >40 %). The remaining farmer selections were composed of three or more small memberships (<30 %). These results indicate significant effort by Aceh farmers to select superior clones based on the pods of the "F1" hybrids released from the government-run seed gardens. However, experimental backstopping would be needed to confirm the agronomic performance, especially disease resistance, before these accessions can be recommended for multiplication and distribution in Aceh.

Currently, there is a strong demand for superior cacao clones in Aceh. There are still about 120,000 hectares of land suitable for expansion of cacao production, and the goal is to become the largest cacao producer in Sumatra by 2020. Black pod disease caused by *Phytophthora palmivora* (Butl.) has been one of the main production constraints. In some farms the losses due to black pod can be as high as 70 %. Although chemical controls have been developed to reduce yield losses, use of resistant genotypes would be more cost-effective for Aceh farmers. Among the sources of resistance available in Indonesia, progenies from SCA 6 are an important source for selecting resistant clones. SCA 6 has been used in almost every polyclone seed gardens in Indonesia, including Java, North Sumatra and Southeast Sulawesi. However, SCA 6 is also well known for its small pod and bean size, which may have led to the lack of adoption of these genotypes by farmers. Based on the SNP genotypes generated in the present study, farmer selections with diverse UAF background should be selected and included in formal field trials.

Given the Indonesian farmers' strong preference for these morphological traits, biparental crosses between SCA 6 and clones with large pod and bean size (e.g. THSA clones) should be pursued. Other resistant parental clones with larger pod and bean sizes, such as UF 273, UF 712 and CC 253, should be imported to Indonesia and used as parental clones in the seed gardens. In addition to black pod disease, VSD is another important fungal disease in all major cacao producing regions in Indonesia, although this disease has not yet been a problem in Aceh. In Sulawesi, resistant clones have been recommended for clonal propagation using embryogenesis. In view of the potential of VSD to be problematic in Aceh, use of VSD-resistant clones selected in Java and Sulawesi should be recommended to provide a complementary source of superior clones.

Farmer varieties of crop species often represent the preferable combination of both natural evolution and human intervention (Bellon 2004). Farmers' fields not only provide a natural laboratory that allow the crop landraces to continually generate new variations; they also allow for the inclusion of farmer-selected agronomic traits (Eskes 2006). However, farmers can make their decision based on only a limited range of genetic diversity and on usually infrequent, informal observational results on the farm. Their observations can often be misled due to bias from to micro-environmental effects, since formal experimental trials are not carried out. Technical support from researchers is therefore essential to improve efficiency of participatory selection. With the available molecular data, information regarding genetic identity, ancestry and parentage will be incorporated with the phenotypic and agronomic data for these farmer selections. Based on this information, a subset of these selections will be chosen. Formal field trials will be carried out in disease-infected areas to confirm the agronomic performance before these selections are distributed.

In summary, we carried out a pilot experiment to assess on-farm genetic diversity in Aceh, Indonesia using SNP-based DNA fingerprinting technology. Genetic identity, ancestry, parentage and spatial pattern were analyzed in 80 farmer selections together with 60 reference clones. Our results confirmed the hybrid nature (Trinitario x UAF) in most of the

farmer selections and quantified the ancestral/parentage contribution from specific UAF germplasm groups. We show that despite the high level of genotype diversity and heterozygosity, the overall genetic background in these farmer selections is relatively narrow, due to limited incorporation of diverse germplasm groups from the Upper Amazon. Access of farmers to diverse planting materials is needed to improve the on-farm diversity in Aceh. Our results also provide explicit genotypic profiles based on SNPs, thereby enabling the selection of a subset of farmer selections with diverse genetic background for field evaluation. The combined information of agronomic performance and molecular characterization then will be used to select the best clones for propagation and dissemination. To our knowledge, this is the first application of SNP markers to assess on-farm diversity in cacao. Information generated through this pilot experiment provides a scientific basis for rapid identification of productive trees for rehabilitation and rejuvenation of cacao farms in Aceh.

References

Allegre M, Argout X, Boccara M, Fouet O, Roguet Y, Bérard A, Thévenin JM, Chauveau A, Rivallan R, Clement D, Courtois B, Gramacho K, Boland-Augé A, Tahi M, Umaharan P, Brunel D, Lanaud C (2011) Discovery and mapping of a new expressed sequence tag-single nucleotide polymorphism and simple sequence repeat panel for large-scale genetic studies and breeding of *Theobroma cacao* L. DNA Res. doi:10.1093/dnares/dsr039

Alverson WS, Whitlock BA, Nyffeler R, Bayer C, Baum DA (1999) Phylogeny of the core Malvales: evidence from ndhF sequence data. Am J Bot 86:1474–1486

Argout X, Fouet O, Wincker P et al (2008) Towards the understanding of the cocoa transcriptome: production and analysis of an exhaustive dataset of ESTs of *Theobroma cacao* generated from various tissues and under various conditions. BMC Genomics 9:512–530. doi:10.1186/1471-2164-9-512

Bartley BGD (2005) The genetic diversity of cacao and its utilization. CAB International, CABI Publishing, Wallingforde

Bayer C, Fay MF, De Bruijn AY, Savolainen V, Morton CM, Kubitzki K, Alverson WS, Chase MW (1999) Support for an expanded family concept of Malvaceae within a recircumscribed 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

Boccara M, Zhang D (2006) Progress in resolving identity issues among the Parinari accessions held in Trinidad: the contribution of the collaborative USDA/CRU project. In: CRU Annu Rep 2005.

Cacao Research Unit, The University of the West Indies, St. Augustine, Trinidad and Tobago

Buckler ES, Thornsberry J (2002) Plant molecular diversity and applications to genomics. Curr Opin Plant Biol 5:107–111

Cheesman EE (1944) Notes on the nomenclature, classification and possible relationships of cocoa populations. Trop Agr 21:144–159

Dias LAS (2001) Origin and distribution of *Theobroma cacao* L: a new scenario. In: Dias LAS (ed) Genetic improvement of cacao. FAO. http://ecoport.org/ep?SearchType=earticleView&earticleId=197&page=−2. Accessed 17 September 2012

Eskes A (2006) Collaborative and participatory approaches to cocoa germplasm evaluation and selection. In: Denamany G, Lamin K, Ling A, Maisin N, Ahmad AC, Saripah B, Nuraziawati MY (eds) Sustainable cocoa economy through increase in productivity, efficiency and quality: proceedings of 4th Malaysian international cocoa conference (MICC), Kuala Lumpur, Malaysia, 18th -19th july 2005. Malaysian Cocoa Board, Kota Kinabalu, pp 77–88

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. MolEcol 14:2611–2620. doi:10.1111/j.1365-294X.2005.02553.x

Evett IW, Weir BS (1998) Interpreting DNA evidence: statistical genetics for forensic scientists. Sinauer, Sunderland

FAOSTAT (2010) Food and Agricultural commodities production. Food and Agricultural Organization of the U N, Rome.http://www.fao.org/economic/ess/ess-trade/en/

Henderson JS, Joyce RA, Hall GR, Hurst WJ, McGovern PE (2007) Chemical and archaeological evidence for the earliest cacao beverages. Proc Natl Acad Sci U S A 104:18937–18940

International Cocoa Organization (ICCO) (2010) The world cocoa economy: Past and present. London, UK

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

Marshall TC, Slate J, Kruuk LEB, Pemberton JM (1998) Statistical confidence for likelihood-based paternity inference in natural populations. Mol Ecol 7:639–655

Mawardi S, Winarno H, Suhendy D (1994) The present status of cocoa breeding at ICCRI: results and future programmes. In: End MJ, Eskes AB, Lee MT, Lockwood G (eds) Proceedings of international workshop on cocoa breeding strategies, vol 1994, Kuala Lumpur Oct., pp 81–87

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, Wallace J, Loor G, Kuhn DN, Brown JS, Schnell RJ (2008) Geographic and genetic population differentiation of the amazonian chocolate tree. PLoS One. doi:10.1007/s12042-008-9011-4

Motilal L, Zhang D, Umaharan P, Mischke S, Mooledhar V, Meinhardt LW (2010) The relic Criollo cacao in Belize - genetic diversity and relationship with Trinitario and other cacao clones held in the International Cocoa Genebank, Trinidad. Plant Genet Resour: Charact Util 8:106–115

Peakall R, Smouse PE (2006) Genalex 6: genetic analysis in excel. Population genetic software for teaching and research. Mol Ecol Notes 6:288–295

Pound FJ (1938) Cacao and witches' broom disease (Marasmius perniciosus) of South America. Arch Cacao Res 1:20–72

Powis TG, Cyphers A, Gaikwad NW, Grivetti L, Cheong K (2011) Cacao use and the San Lorenzo Olmec. Proc Natl Acad Sci U S A 108:8595–8600

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure from multilocus genotype data. Genetics 155:945–959

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

Susilo A, Zhang D, Motalal L, Meinhardt LW (2011) Assessing genetic diversity in Java fine-flavor cocoa (*Theobroma cacao* L.) germplasm by simple sequence repeat (SSR) markers. Trop Agric 55: 84–92

Toxopeus H, Geisberger (1983) History of cocoa and cocoa research in Indonesia. In: Toxopeus H, Wessel PC, Larson RE (eds) Archives of cocoa research vol 2. American Cocoa Research Institute, Washington; International Office of Cocoa and Chocolate, Brussels, pp 7–34

Tsai YHE (2011) PhyloGeoViz: a web-based program that visualizes genetic data on maps. Mol Ecol Resour 11:557–561

Van Hall CJJ (1914) Cocoa. Macmillan, London http://biodiversitylibrary.org/page/22927228. Accessed 14 September 2012

Waits LP, Luikart G, Taberlet P (2001) Estimating the probability of identity among genotypes in natural populations: cautions and guidelines. Mol Ecol 10:249–256

Wood GAR, Lass RA (1985) Cocoa. Longman, London

Zhang D, Mischke S, Goenaga R, Hemeida AA, Saunders JA (2006) Accuracy and reliability of high-throughput microsatellite genotyping for cacao clone identification. Crop Sci 46:2084–2092

Zhang D, Boccara M, Motilal L, Mischke S, Johnson ES, Butler D, Bailey BA, Meinhardt LW (2009) Molecular characterization of an earliest cacao (*Theobroma cacao* L.) collection from Peruvian Amazon using microsatellite DNA markers. Tree Genet Genomes 5:595–607. doi:10.1007/s11295-009-0212-2

Zhang D, July-Martínez W, Johnson ES, Somarrib E, Phillips-Mora W, Astorga C, Mischke S, Meinhardt LW (2012) Genetic diversity and spatial structure in a new distinct Theobroma cacao L. population in Bolivia. Genet Resour Crop Evol 59:239–252