ORIGINAL ARTICLE

Parentage analysis and outcrossing patterns in cacao (*Theobroma cacao* L.) farms in CameroonMIB Efombagn¹, O Sounigo², AB Eskes³, JC Motamayor⁴, MJ Manzanares-Dauleux⁵, R Schnell⁶ and S Nyassé¹¹Institute of Agricultural Research for Development, Yaoundé, Cameroon; ²CIRAD-BIOS, UPR31, Montpellier, France; ³Bioversity International, Parc Scientifique Agropolis II, Montpellier, France; ⁴MARS Inc. c/o USDA-ARS-SHRS, Miami, FL, USA;⁵UMR INRA-Agrocampus Rennes APBV, Le Rheu Cedex, France and ⁶USDA-ARS-SHRS, Miami, FL, USA

The present study investigates the parentage of farm accessions in Cameroon using data from 12 microsatellite loci. Bayesian analysis suggests that 25.5% of the 400 farm accessions studied is still closely related to the traditional Amelonado variety called 'German Cocoa' by the farmers. Another 46.3% of the farm accessions were found to be direct descendants (20.8% first-generation (F1) hybrids and 25.5% selfed genotypes) from 24 parental clones used in biclonal seed gardens (BSGs) established in the 1970s in southern and western Cameroon. Furthermore, 28.3% of farm accessions appeared to descend from uncontrolled pollination events in cacao farms, which could be related to a common practice of cacao growers to use seeds collected in their own farm for new plantings. All farm accessions descending from BSG could be individually related through

parentage analysis to the 24 progenitors of the BSG. Only 25% of progenies distributed from BSG corresponded to F1 hybrids combinations originally planned to be released. Significant biparental inbreeding estimates were observed for all 'traditional' farms and for most 'F1 hybrids' farms due to presence of a high proportion of selfed accessions. Biparental inbreeding occurs when plants receive pollen from genetically related neighbors. High levels of outcrossing observed in 'mixed' farms might be explained by the admixture of traditional varieties and BSG progenies. The implications of our finding for management of seed gardens and for further breeding using farm accessions in Cameroon are discussed.

Heredity (2009) **103**, 46–53; doi:10.1038/hdy.2009.30;
published online 1 April 2009

Keywords: cacao; biclonal seed gardens; farm; outcrossing; breeding

Introduction

Cacao (*Theobroma cacao* L.) is a diploid tropical tree belonging to the Malvaceae (Alverson *et al.*, 1999) that displays a preferentially allogamous mating system. Traditionally, cacao has been subdivided into four genetic groups: upper Amazon Forastero (UA), lower Amazon Forastero (LA), Criollo (Cr, anciently cultivated by the Indians in Central America) and Trinitario (Tr, hybrids between LA and Cr types). Cacao breeding largely relied on the creation of hybrids between progenitors belonging to different genetic groups of the species. Hybrid vigor (or heterosis) has been suggested to be the basis of good performance for precocity and yield of many of the selected biparental crosses between UA, LA and Tr or Cr types (Bartley, 1967; Soria, 1978; Reyes, 1979). However, in some countries, intragroup UA hybrids have also been found to be good yielders (Lockwood, 1979) and are recommended for commercial usage.

The attempt to produce hybrid seeds for large-scale distribution to farmers has involved studies on incompatibility reactions of the progenitors of the most

productive hybrid progenies. Self-incompatibility (SI) in cacao was reported by Pound (1932) and its genetics is considered unique among the identified systems in flowering plants. Knight and Rogers (1955) reported that cacao has a sporophytic system controlled by a series of S alleles that show dominant or equal action. Further cytological studies by Cope (1958) revealed, however, that in incompatible combinations, nonfusion of gametes may occur in 25–100% of the ovules, suggesting also a gametophytic control system. SI in cacao is not always complete, because self-pollination of 'self-incompatible' genotypes may sometimes result in a low percentage of fruit setting. Haddon (1961) considered clones to be 'self-incompatible' if fruit setting of self-pollinated flowers is less than 10% and as 'self-compatible' if fruit setting is above 30%. Furthermore, the SI mechanism can be partly overcome by mixing self-compatible and self-incompatible pollen (Glendinning, 1960). This phenomenon has been used in cacao breeding as a tool to produce selfed progenies with self-incompatible genotypes.

In South America, the genetic origin of cacao, wild populations known as Forastero contain predominantly self-incompatible individuals whereas traditional varieties such as Amelonado (an LA type), Cr and Nacional (from Ecuador) are self-compatible (Cheesman, 1944; Haddon, 1961). However, according to Bartley and Cope (1973), Forastero from the Amazon basin and Cr from the Central America segregate both for the self-compatibility (SC) and the SI. This might be explained by the fact that

Correspondence: Dr MIB Efombagn, Institute of Agricultural Research for Development, PO Box 2067 or 2123, Yaoundé, Cameroon.

E-mail: efombagn@yahoo.fr

Received 31 May 2008; revised 14 January 2009; accepted 3 March 2009; published online 1 April 2009

several UA genotypes are partially self-compatibles and the 'Modern' Cr (Tr closed to the Cr) with the Tr's genes of incompatibility are often self-incompatibles.

Management of biconal seed gardens (BSGs) established in many cacao-producing countries between 1960 and 1990 has generally involved the association of a self-incompatible seed parent with a self-compatible or self-incompatible pollen parent in a single plot, with the expectation that true hybrid seeds can be produced by natural pollination. During the 1960s in Cameroon, cacao breeding involved largely the development of new hybrid varieties among Tr and UA parental clones. BSGs were created to produce hybrid seeds for subsequent release to the farmers. In total, 40 BSGs were established, 36 aimed at producing UAxTr (or reciprocal) hybrid progenies and 4 aimed at producing TrxTr hybrid progenies. In most BSGs, the progenitors were planted at a ratio of one pollen to three seed parents. The harvest of pods was to be performed only on the seed trees, based on the assumption that SI is complete. However, studies with isozymes on progenies obtained by open pollination in such BSGs from different countries (including some BSGs from Cameroon) revealed that variable proportions of seeds might result from selfing (Lanaud *et al.*, 1987). The occurrence of selfings through uncontrolled pollinations in BSG progenies may have influenced the efficiency of the breeding programs.

Besides the expected variable degrees of selfing in the BSGs, variable rates of outcrossing between cultivated varieties in farmers' fields may occur because BSG progenies (hybrids) and traditional varieties are often planted in the same fields. The presence of genotypes from different genetic origins has favored the admixture

of genes through natural outcrossing, leading to the appearance of cultivated cacao populations with a complex genetic structure. Recent analyses using a Bayesian model-based method have shown the existence of a high level of admixture for the farm accessions group. The LA genes were most represented (54%), followed by UA (33%) and Cr (7%) genes (Efombagn *et al.*, 2008). The objective of the present study was to investigate the genetic origin of farmers' accessions and to relate the results to the known composition of the BSGs in Cameroon and to the known traditional varieties grown. The specific questions addressed were the following:

- What are the proportions of selfed and outcrossed cacao progenies in farmers' field derived from BSG progenitors?
- What is the pattern of outcrossing in existing cacao farms?
- What are the practical consequences for seed gardens management and breeding?

Materials and methods

Sample collection

Leaf samples were obtained from a total of 400 farms accessions and of the 24 progenitors used in the BSGs in Cameroon (Table 1). Farm accessions were chosen at random during field surveys in different cacao-producing areas of Cameroon, ranging from a latitude of 02°14.199'N to 05°42.924'N, and a longitude of 009°01.430'E to 011°20.885'E. The farm accessions were the same as those analyzed before by Efombagn *et al.*

Table 1 List of BSG progenitors with their compatibility status (SC and SI), as well as their use in seed gardens (P and S) and number of seed gardens in which each progenitor was planted

Parental class	Progenitors group	Progenitor name	Compatibility pattern ^a	Parental use in BSG (pollen or seed)	Number of BSG in which planted
Trinitario	Local	SNK10	SI	P/S	4
		SNK12	SC	P	1
		SNK13	SI	P/S	8
		SNK16	SI	P/S	10
		SNK37	SC	P	2
		SNK48	SI	P/S	6
		SNK64	SC	P/S	3
		SNK109	SI	P/S	5
		SNK413	SI	P	2
		SNK450	SC	P	2
	Introduced	ICS40	SI	P/S	7
		ICS43	SI	S	3
		ICS46	SI	P/S	4
		ICS60	SI	S	1
		ICS95	SC	P/S	8
Upper Amazon (UA)	G1 (UAXUA)	T60/887	SI	P/S	2
		T60/1174	SI	P/S	2
		T79/467	SI	P/S	4
		T79/501	SI	P/S	7
		UPA134	SI	P/S	4
		UPA143	SI	P/S	12
	G0 (original Trinidad introductions)	IMC67	SI	P/S	9
		SCA6 ^b	SI	P/S	4
		SCA12	SI	P/S	2

Abbreviations: P, pollen; S, seed; SC, self-compatible; SI, self-incompatible.

^aThe compatibility status was determined by previous studies carried out in 1960s, before the establishment of seed gardens.

^bThe SCA6 used in BSG in Cameroon is known to be an off-type.

(2008). The farms considered under the study were identified as traditional-based, hybrid-based or mixed traditional and hybrid-based varieties (accessions). Sampling within each farm was carried out according to its size and varietal composition, and the sample sizes varied from 3 to 17 cacao trees per farm.

Microsatellite analysis

The 12 microsatellite markers used have been mapped by Lanaud *et al.* (1999) and the DNA analyzed was extracted from the leaves of all the plant materials of the study using the method developed by Bhattacharjee *et al.* (2004). PCR reactions were carried out in a 5 µl reaction mix containing 2.5 ng (1 µl) of template DNA, 1 µl of 55 × PCR buffer (10 mM Tris-HCl (pH 8.3), 50 mM KCl), 1 µl of 25 mM MgCl₂, 0.5 µl of each forward and reverse primers, 0.2 µl of 10 mM dNTPs (dATP, dCTP, dGTP, dTTP), 0.1 µl of 5 U Taq polymerase (Bioline, London, UK). Thermal cycling profiles carried out in a Gradient Cycler PTC 200 (MJ Research, Ramsey, MN, USA) consisted of 5 min initial denaturation at 94 °C, 35 cycles of amplification at 94 °C for 30 s, 51 °C annealing for 1 min and 72 °C for 1 min. This was followed by further primer extension at 72 °C for 7 min. PCR reactions were performed on an MJ research Dyad 96 wells PCR. A volume of 0.5 µl of diluted PCR products (1:19) was mixed with 9.5 µl of formamide (PE Applied Biosystems, Foster City, CA, USA) and 0.5 µl ROX-labeled GeneScan-500 size standard (PE Applied Biosystems). The PCR products were denatured at 95 °C for 5 min and fragment analysis was performed on an ABI 3100 Prism Genetic Analyzer (PE Applied Biosystems). The software programs GeneScan 3.7 and GeneMapper 3.5 (PE Applied Biosystems) were used to process and score the data.

Hybridization and parentage analysis

To measure the level of outcrossing and to identify the putative BSG progenitor of each BSG-derived farm accessions, we used the Bayesian approach that consisted of two steps:

1. *Identification of pure and admixed farm accessions:* The purpose of this step was to identify hybrid and nonhybrid samples. The software Structure 2.2 (Pritchard *et al.*, 2000) was run to identify the genetic structure of the whole farm accessions population. The ancestry model with admixture was applied to infer how many clusters or subpopulations (K) were most appropriate to identify the number of potential varietal groups among the farm accessions. Ten independent runs of $K=1-12$ were performed at 100 000 MCMC (Markov chain Monte Carlo) and a postburning simulation length of 500 000. To identify and discard the traditional Amelonado accessions (nonhybrid farm genotypes), we used the option POPFLAG = 1 to assign all the farm accessions to five original Amelonado accessions (controls).
2. *Identification of genealogical classes of farm accessions of hybrid origin:* After discarding the traditional Amelonado genotypes from the farm accessions population, the microsatellite data of all the remaining genotypes (first-generation (F1) hybrids, admixed or selfed genotypes) were subjected along with BSG progenitors to the analysis of hybrid ancestry using the method implemented in the program NewHybrids 1.0

(Anderson and Thomson, 2002). Structure can identify admixtures among any number K of parental populations, whereas NewHybrids assumes that hybrid classes originated after admixture of two parental species or parental populations (genetic groups of putative progenitors in our study) within species. NewHybrids uses an inheritance model defined in terms of genotype frequencies to compute, for each individual, the posterior probability of inclusion in each of the following six classes: P1 (Tr BSG progenitors in our study), P2 (upper Amazon BSG progenitors), selfed P1 or selfed P2 genotypes (S1), F1, F2 and backcross (BC) genotypes between F1 and each of the two genetic groups of progenitors (P1 and P2). NewHybrids was run using the default parameters for the six genotype class frequencies, Jeffreys prior, a burn-in phase of 100 000 steps and 500 000 MCMC sweeps. The five Amelonado controls were subsequently used as a separate parental class to simulate the reality in farmers' field where genotypes may have originated from outcrossing between traditional and BSG-derived progenies.

Parentage analysis was performed on the F1 hybrid accessions identified by NewHybrids. We used the software Cervus 3.0 (Marshall *et al.*, 1998) to detect the putative BSG progenitors of the F1 accessions, as well as of the selfed S1 accessions per candidate progenitor. Cervus uses a simulation program to generate likelihood scores and provides a level of statistical confidence to assign paternity, maternity or both pollen and seed parents. Parent pair analysis used in this study is the common parentage analysis when the candidate pollen and seed parents are known (Marshall *et al.*, 1998). The ln-likelihood ratios were expressed as LOD scores. The most likely progenitor was the one with the highest positive LOD score. The Cervus program uses furthermore the number of mismatching loci between a candidate parent and a putative offspring for the parental assignment. Polymorphic information content (Hearne *et al.*, 1992) and parent pair nonexclusion probability (Jamieson and Taylor, 1997; Marshall *et al.*, 1998) for each candidate progenitor were calculated for each microsatellite used. The nonexclusion probability is the probability of not excluding a single unrelated candidate parent or parent pair from parentage of a given offspring at one locus.

Diversity analysis of farm accessions subgroups

Diversity parameters were computed on the different farm accessions subgroups (as indicated in Table 2) generated from the hybridization and parentage analysis. Gene diversity (GD) and F-statistics (F_{ST} and F_{IS}) as defined by Weir and Cockerham (1984) of each subgroup were calculated using FSTAT software (Goudet, 1995). Due to significant variation in sampling size between farm and reference accession groups (AGs), the average GD (=expected heterozygosity), allelic richness (AR; Leberg, 2002) were estimated using the rarefaction approach implemented in the software HP-RARE 1.0 (Kalinowski, 2005). Rarefaction approach was used to standardize the mean number of alleles per locus or the number of private alleles for a given AG, based on the number of genes (alleles) present in the smallest sample (AG) of accessions present in the study.

Table 2 Distribution of the farm accessions according to their genealogical classes into Amelonado, BSG progenies (S1, F1) and complex genotypes (F2 and BC), and analysis of genetic diversity for each subgroup

Type of farm accessions	Genetic structure	Number (% of accessions)	Number of alleles	Diversity parameters				Total (accessions)
				GD	F _{IS}	AR	F _{ST}	
Traditional	Amelo-nado	102 (25.5)	58	0.20	0.22	2.58	0.20	102
BSG F1 hybrids	UAxTr or TrxUA	37 (9.3)	74	0.60	0.16	5.04	0.05	83
	TrxTr	31 (7.8)	79	0.57	0.12	5.35	0.06	
	UAxUA	15 (3.8)	66	0.62	0.12	5.27	0.07	
BSG selfed (S1) genotypes	Tr	59 (14.8)	59	0.44	0.14	3.80	0.07	102
	UA	43 (10.8)	82	0.63	0.10	5.34	0.09	
Others: F2+BC ^a	—	113 (28.3)	93	0.55	0.09	5.18	0.04	113
All		400 (100)	113	0.50	0.15	(MSZ = 26) ^b	—	400

Abbreviations: AR, allelic richness; BC, backcross; BSG, biclonal seed garden; F_{IS}, inbreeding coefficient; F_{ST}, coefficient of differentiation; GD, gene diversity; Tr, Trinitario; UA, upper Amazon Forastero.

^aFarm accessions issued from seed derived from natural pollination (taken from pods collected by the farmers in their own or neighboring farms).

^bMSZ, value of the minimum sample size considered to estimate the allelic richness with the rarefaction approach.

Mating system analysis at the level of cacao farms

To investigate the level of outcrossing in farmers' field, we analyzed farm accessions from 15 cacao farms identified with at least 10 sampled accessions. The mating system within each farm was analyzed based on a mixed mating model (Ritland and Jain, 1981). A multilocus mating system implemented in the software MLTR (Ritland, 2002) based on the algorithm of Ritland and Jain (1981) provided estimates of single-locus (*ts*) and multilocus (*tm*) population outcrossing rates. The probability model underlying multilocus estimation of mating system assumes *n* unlinked loci. In the mixed mating model used, standard errors for the outcrossing estimates were based on the 1000 bootstrap replications. Biparental inbreeding (*tm*–*ts*) within each selected farm was then estimated (Shaw *et al.*, 1980). Biparental inbreeding may appear during selfings or when mating occurs between relatives. In most cacao farms in Cameroon, a large proportion of trees derive from seeds collected within farms. Some cacao trees produce selfed seeds, and neighboring cacao trees may be genetically related, suggesting the existence of potential biparental inbreeding within the corresponding farms.

Results

Identification of the genealogical structure of farm accessions

The value *K* = 3 maximized the posterior probability of the farm accessions data according to the Structure program. We found that one of the three clusters was composed of 102 genotypes that were designated as traditional accessions (according to the Amelonado reference genotypes), with an individual posterior probability ≥ 0.70 (Table 2). In the farm accessions varietal structure, this group was therefore identified as Amelonado (Table 2). The genotypes of all the remaining farm accessions were more or less admixed, and their genealogical classes in relation to the BSG progenitors used in Cameroon were determined by NewHybrids. Results suggest that 83 out of 400 farm accessions (20.75% of the total population) were composed of F1

hybrids between BSG progenitors, 102 (25.5%) were selfed progenies from BSG progenitors, whereas 113 (28.25%) accessions belonged to either F2 or BC generations. Because the probabilities of assignment to F2 or BC were rather similar, these were considered as one group (identified as F2 + BC in Table 2). F2 + BC material was found as the likely product of natural pollination in farmers' field between F1 hybrids or between F1 hybrids and traditional genotypes. Within the F1 group, three subgroups based on the likely BSG progenitors were detected as shown in Table 2. Out of the 83 F1 genotypes identified, 37 (45% of the total F1 population) arose from crosses between UA and Tr progenitors.

Diversity in different farm accessions subgroups

The mean number of alleles (*N*) and GD recorded in traditional accessions (*N* = 58; GD = 0.20), were lower than the ones found in outcrossed (*N* = 66–74; GD = 0.57–0.62) and selfed (*N* = 59–82; GD = 0.44–0.63) groups derived from the BSG progenitors (Table 2). The inbreeding coefficient (*F*_{IS}) for the traditional accessions group (Amelonado) was also higher (*F*_{IS} = 0.22) compared to the other groups (*F*_{IS} = 0.10–0.12 for outcrossed and *F*_{IS} = 0.12–0.14 for selfed BSG accessions). AR varied from 2.58 for traditional accessions to 5.04–5.35 for outcrossed and 3.80–5.34 for selfed BSG accessions. The traditional group was more genetically differentiated from the other subgroups (*F*_{ST} = 0.20), whereas this genetic differentiation was relatively low among the BSG subgroups (*F*_{ST} = 0.05–0.09).

Parentage analysis

All the F1 genotypes were subjected to a parentage analysis to detect the contribution of different BSG progenitors in the released material. Nonexclusion probabilities varied from 0.14 to 0.82 across the 12 loci of the study, and most of the null allele frequencies of the same loci were lower than 0.05 (Table 3). The genotypes were assigned to candidate BSG progenitors based on the highest positive LOD score and on the number of mismatched alleles registered (Table 4). The Tr progenitor SNK450 was the most represented with 33 putative

Table 3 Diversity, parental pair nonexclusion probability and null allele frequencies estimated for the 12 SSR loci in the parentage analysis of the F1 hybrids accessions

Locus	No. of alleles	Expected heterozygosity	PIC	Nonexclusion probability	Null allele frequency estimates
mTcCIR 3	15	0.84	0.82	0.14	0.02
mTcCIR 6	9	0.53	0.48	0.536	0.03
mTcCIR 9	12	0.43	0.41	0.57	0.05
mTcCIR 12	11	0.8	0.77	0.2	0.01
mTcCIR 15	13	0.77	0.73	0.25	0.02
mTcCIR 17	4	0.29	0.27	0.73	0.04
mTcCIR 18	8	0.61	0.58	0.37	0.04
mTcCIR 19	13	0.62	0.59	0.37	0.00
mTcCIR 21	10	0.57	0.54	0.42	0.00
mTcCIR 24	7	0.2	0.19	0.8	0.00
mTcCIR 25	13	0.79	0.76	0.18	0.06
mTcCIR 26	10	0.67	0.62	0.36	0.14

Table 4 Parentage and self-compatibility analyses

Genetic group of progenitors	Candidate progenitor genotype	Assigned parentage		Total (farm accessions)
		Outcrossed (F1) ^a	Selfed (S1)	
Trinitarios (Tr)	SNK10	4	10	14
	SNK12 ^b	11	8	19
	SNK13	6	1	7
	SNK16	9	3	12
	SNK37 ^b	5	6	11
	SNK48	0	0	0
	SNK64 ^b	3	0	3
	SNK109	6	3	9
	SNK413	6	1	7
	SNK450 ^b	22	11	33
	ICS40	10	2	12
	ICS43	8	6	14
	ICS46	6	6	12
	ICS60	2	1	3
	ICS95 ^b	1	1	2
Upper Amazon (UA)	IMC67	9	3	12
	SCA6	0	0	0
	SCA12	1	2	3
	T60/887	6	0	6
	T60/1174	3	2	5
	T79/467	13	13	26
	T79/501	10	13	23
	UPA134	8	4	12
	UPA143	17	6	23
	Total	166	102	268

Assignment of farm accessions to the most-likely seed gardens progenitor, based on the higher positive LOD score and the number of loci mismatching observed.

^aEach outcrossed progeny has two parents identified.

^bSelf-compatible clones.

offspring, followed by T79/501 and SNK450. The BSG progenitors SNK48 and 'SCA6' were not found as parents of any accession (Table 4). For 45% of the F1 hybrid genotypes, the two parents identified did not correspond to the known combinations of the BSG progenitors (data not shown). These are therefore derived from unwanted hybridizations in the BSG.

Outcrossing estimates in BSG progenies and farmers' fields

The total number of selfed progenies was estimated for each of the 24 BSG progenitors (Table 4). Selfed offspring was detected for 20 out of 24 progenitors, including both SC and SI progenitors. The proportion of selfed individuals for all SC progenitors together was similar to that of the group of SI progenitors.

The difference between *tm* and *ts* (*tm*–*ts*) provides an estimate of biparental inbreeding within each farm. Positive *tm*–*ts* values implied the existence of biparental inbreeding within the farm. The *tm*–*ts* values in all 'traditional' farms were positive (Table 5), suggesting that biparental inbreeding has occurred within these farms. In the 'F1 hybrids' farms, *tm*–*ts* values ranged from –0.26 to 0.36, with high values being associated with a large number of selfed BSG individuals and low values with a large number of F2+BC individuals. The 'mixed' farms recorded low to average biparental inbreeding, with *tm*–*ts* values ranging from –0.51 to 0.19.

Discussion

In earlier work, the genetic diversity of farm populations was compared to that of gene bank and reference accessions (Efombagn *et al.*, 2008). In the present study, the genetic composition of the same farm accessions has been compared to the BSG progenitors and to the traditional Amelonado variety (named 'German Cocoa' by farmers). The results throw lights on the outcrossing patterns observed in BSG and in farmers' fields. They also help to explain the performance of different varieties in farmers' fields and stress the need to improve seed garden management.

Composition of farm accessions

Bayesian analysis suggests that 25.5% of the 400 farm accessions studied is closely related to the traditional Amelonado variety called 'German Cocoa' by the farmers. Another 46.3% of the farm population (20.8% outcrossed and 25.5% selfed) were found to be descendants from 24 parental clones used in BSGs established in the 1970s in southern and in western Cameroon. Furthermore, 28.3% of the farm accessions appeared to descent from uncontrolled pollination events in cacao farms, which is related to a common practice of cacao growers to use seeds collected in their own farm for new plantings. Diversity parameters revealed that the rather 'pure' traditional accessions subgroup was less diverse compared to BSG-derived accessions. Therefore, the existing diversity in cacao farms has not been reduced by the selection process carried out in Cameroon, but rather increased by the release of BSG material in farmers' field. However, the AR and GD between outcrossed and selfed BSG accessions from UA origin did not differ significantly, indicating high levels of heterozygosity of the UA progenitors used in seed gardens.

Expected and observed crossing patterns in BSGs

Hybrid cacao seed production in BSGs started in many countries in the 1960s and 1970s. In some countries hybrid seed has always been produced by hand

Table 5 Outcrossing rates in 15 cacao farms studied and number of accessions in different genealogical classes

Type of material ^a	Farm class age ^b	<i>tm</i> – <i>ts</i> (s.d.) ^c	Number of farm accessions according to their genealogical classes				<i>N</i> ^d
			Traditional	BSG Selfed (S1)	BSG F1 Hybrids	Others: F2+BC	
Traditional	D	0.32 (0.05)	9	0	0	1	10
	C	0.05 (0.12)	5	0	0	5	10
	D	0.01 (0.4)	6	0	0	4	10
	C	0.39 (0.00)	12	0	0	0	12
	Total		32	0	0	10	42
F1 hybrids	B	–0.26 (0.11)	0	6	4	7	17
	C	–0.08 (0.09)	0	5	2	5	12
	A	0.21 (0.08)	1	9	0	0	10
	A	0.06 (0.12)	1	8	1	0	10
	A	0.36 (0.23)	2	7	0	1	10
	Total		4	35	7	13	59
Mixed	A	–0.19 (0.11)	4	5	0	3	12
	B	–0.51 (0.19)	0	2	1	7	10
	B	–0.28 (0.00)	1	2	1	6	10
	D	0.06 (0.45)	1	3	1	5	10
	C	0.19 (0.13)	4	4	1	2	11
	A	–0.24 (0.28)	2	1	0	7	10
	Total		12	17	4	30	63

^aAccording to the information received from the farmers.

^bA, 11–20 years; B, 21–30 years; C, 30–50 years; D, > 50 years.

^c*tm*, multilocus population outcrossing rate; *ts*, single locus population outcrossing rate; *tm*–*ts*, biparental inbreeding. The values between brackets represent the standard deviation.

^d*N*, number of accessions studied.

pollinations, whereas in other countries (such as Cameroon) the strategy has been to harvest open-pollinated pods on the self-incompatible parental clone in the BSGs. The latter strategy was based on the assumption that self-incompatible genotypes would only yield cross-pollinated progenies. However, different rates of selfing have subsequently been observed when allo-pollen is mixed with auto-pollen (Lanaud *et al.*, 1987), and such may happen with natural pollination in BSGs. Consequently hand pollinations should have been adopted in all hybrid seed production programs, but due to lack of financial resources related to low cocoa prices in the 1990s and early 2000s such measures have often been adopted with delay. In Cameroon, hand pollinations were only adopted to produce hybrid seed recently (2005). The data analyzed in the present study mainly refer to trees that derived from seed gardens before hand pollinations became adopted in the BSGs.

Despite the fact that 19 of the 24 BSG progenitors are considered as self-incompatible, 55% of the BSG-derived individuals appear to be the result of self-pollinations. This can be explained by the fact that SI of cacao can be overcome by mixing self-compatible with self-incompatible pollen (Glendinning, 1960), such as may be expected from insect pollinations in BSGs. Lanaud *et al.* (1987) have shown that selfing rates for eight open-pollinated progenitors of BSG in Cameroon varied between 0 and 89% and obtained 5–82% selfed seeds after hand pollination of self-incompatible progenitors with mixtures of self- and allo-pollen. These large variations suggest that SI is more easily overcome with some genotypes than with others. Furthermore, the level of SC may vary according to the genotype and can also be affected by seasonal variations (Voelcker, 1938). Physiological reasons of SI may be caused by modifications of biochemical contents of cacao flowers

(Haseinstein and Zavada, 2001) as well as morphological changes in floral parts. Several factors may therefore interact to favor selfing. Therefore, uncontrolled pollination (as occurred in BSGs in Cameroon) is not an efficient method for hybrid-seed production.

In our study, we further found about 45% of F1 progenies derived from progenitors planted in different BSGs. The possible hypothesis that might explain this situation is either the transfer of pollen from one BSG to another or the errors occurred during field planting of progenitors as recently observed in several BSGs. Furthermore, cacao pods produced by SC clones used as progenitors were not supposed to be harvested for seedlings production and distribution. Nevertheless selfed farm accessions were found as deriving from SC progenitors as shown in Table 4, and this shows that the process and rules for seeds production were not followed up rigorously.

Outcrossing patterns in cacao farms

Our study shows that some accessions of the farm population are derived from natural hybridization between genotypes of various origins (F2 + BC genotypes). This can be explained by the general practice of farmers to collect cacao seeds on-farm to produce seedlings for new plantings or replanting.

The three types of farms (Table 5) studied were defined according to farmers' knowledge. The results on outcrossing and on genealogical groups present in these farms show that the farmers had identified quite correctly the type of material in their farms. High biparental inbreeding (*tm*–*ts* values) was correlated with high proportions of selfed (S1) or traditional genotypes (Amelonado) in the farms studied. Traditional farms appear to be still rather pure. However, the

presence of some F2 + BC genotypes in three farms suggest that the farmers have possibly used outcrossed seeds from neighboring farms for infilling. With regard to the hybrid farms, it is striking to note that in the younger farms (below 20 years) there were very few F1 trees identified. This suggests that the BSG descendants distributed over the last 20 years may have contained even more selfed seeds than the BSG descendants distributed before. The oldest 'F1 hybrids' show also more F2 + BC accessions, which indicate the relative high degree of outcrossing in these farms. This result also demonstrates clearly that some 'hybrid farms' do contain very few pure F1 hybrids. In the mixed farms, the F2 + BC trees were predominating, showing a very high level of outcrossing.

Implications for breeding and seed gardens management
Approximately 26% of the 400 farm accessions were classified as Amelonado. This traditional variety seems therefore still to be well presented in farmers' fields. The GD (expected heterozygosity) of the traditional Amelonado variety (0.20) is similar to that of the reference genotypes used in the study, which are the same as used by Efombagn *et al.* (2008). This suggests that the group of traditional farm accessions as identified by Structure is still rather pure. However, due to a certain level of introgression from other genomes into the traditional Amelonado background (Efombagn *et al.*, 2008), it cannot be excluded that selections for improved agronomic traits will be possible within this group. The highest levels of GD were found in the hybrid (F1) and in the F2 + BC progenies. The latter group has most likely been generated through natural pollination in farmers' fields. This widely diverse group should offer good opportunities for further selection and breeding.

The parentage analysis indicates that 46% of the farm accessions analyzed descends from BSG progenitors, but only 45% of these appear to be the result of outcrossing. Furthermore, the identified parentage of 45% of the F1 hybrid genotypes did not correspond to the BSG combinations established in Cameroon. This may be due to mislabeling in the seed gardens or to uncontrolled pollinations between progenitors of neighboring BSGs. Consequently, only 25% of the seeds that have been distributed from the BSGs represents the hybrid cross combinations that were originally intended to be distributed to the farmers. The predominance of unselected progenies distributed from BSGs may well explain why the performance of the so-called 'hybrid' varieties has not been satisfactory in farmers' fields. First, the high percentage (55%) of selfed Tr and UA genotypes among the BSG progenies may cause significant inbreeding effects on yield, as has been demonstrated in other African countries when comparing yield of outcrossed and selfed UA (N'Goran *et al.*, 2003) and Tr progenitors (Toxopeus, 1972). Second, the high percentage (49%) of Tr genotypes in the BSG progenies, which as a genetic group are generally considered to be highly susceptible to diseases and pests, may explain the complaints made by farmers that the 'hybrid' varieties are more susceptible to *Phytophthora* pod rot and dieback than the traditional varieties.

The results stress the need for strict control of pollinations in cocoa seed gardens in Cameroon. Only when this is correctly accomplished, the farmers may change their views on the value of hybrid varieties as compared to the traditional Amelonado variety.

Acknowledgements

This work was part of the PhD thesis of the first author. This study was partially funded by the West Africa Cocoa Diversity Project (IITA/USAID), the CFC/ICCO/Bioversity project entitled 'Cacao Productivity and Quality Improvement, a Participatory Approach' and Mars Inc. We acknowledge the Institute of Agricultural Research for Development (IRAD), Cameroon, for authorizing the publication of this article.

References

- 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.
- Anderson EC, Thomson EA (2002). A model-based method for identifying species hybrids using multilocus genetic data. *Genetics* **160**: 1217–1229.
- Bartley BGD (1967). Twenty years of cacao breeding at the imperial college of Tropical Agriculture, Trinidad. II International Cocoa Research Conference, Salvador-Itabuna, Brsil, pp 29–34.
- Bartley BGD, Cope FW (1973). Practical aspects of self-incompatibility in *Theobroma cacao* L. In: Moav R (ed). *Agricultural Genetics—Selected Topics*. Wiley and Sons: New-York, pp 109–134.
- Bhattacharjee R, Kolesnikova-Allen M, Aikpokpodion P, Taiwo S, Ingelbrecht I (2004). An improved semi-automated rapid method of extracting genomic DNA for molecular marker analysis in cacao, *Theobroma cacao* L. *Plant Mol Biol Report* **22**: 435a–435h.
- Cheesman EE (1944). Notes on the nomenclature, classification and possible relationships of cocoa populations. *Trop Agric* **21**: 144–159.
- Cope FW (1958). Incompatibility in *Theobroma cacao*. *Nature* **181**: 279.
- Efombagn MIB, Motamayor JC, Sounigo O, Eskes AB, Nyass S, Cilas C *et al.* (2008). Genetic diversity and structure of farm and GenBank accessions of cacao (*Theobroma cacao* L.) in Cameroon revealed by microsatellite markers. *Tree Genet Genome* **4**: 821–831.
- Glendinning DR (1960). Selfing of self-incompatible cocoa. *Nature* **187**: 170.
- Goudet J (1995). FSTAT 1.2: a computer program to calculate F-statistics. *J Hered* **86**: 485–486.
- Haddon A (1961). Variety trials of seedlings cacao in Malaya. *Malayan Agric J* **43**: 169–232.
- Haseinstein KH, Zavada MS (2001). Auxin modification of the incompatibility response in *Theobroma cacao*. *Physiologia Plantarum* **112**: 113–118.
- Hearne CM, Ghosh S, Todd JA (1992). Microsatellites for linkage analysis of genetic traits. *Trends Genet* **8**: 288–294.
- Jamieson A, Taylor CS (1997). Comparisons of three probability formulae for parentage exclusion. *Anim Genet* **28**: 397–400.
- Kalinowski ST (2005). HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. *Mol Ecol Notes* **5**: 187–189.
- Knight R, Rogers HH (1955). Incompatibility in *Theobroma cacao* L.. *Heredity* **9**: 69–77.
- Lanaud C, Sounigo O, Amefia YK, Paulin D, Lachenaud P, Clment D (1987). Nouvelles donnes sur le fonctionnement

- du système d'incompatibilité du cacaoyer et ses conséquences pour la sélection. *Café, Cacao, Thé* **31**: 267–277.
- Lanaud C, Risterucci AM, Piretti I, Falque M, Bouet A, Lagoda PJJ (1999). Isolation and characterization of microsatellites in *Theobroma cacao* L. *Mol Ecol* **8**: 2141–2152.
- Leberg PL (2002). Estimating allelic richness: effects on sample sizes and bottlenecks. *Mol Ecol* **11**: 2445–2449.
- Lockwood G (1979). Cocoa breeding in Ghana with reference to swollen shoot disease. VII International Cocoa Research Conference, Douala, Cameroon, pp 407–413.
- Marshall TC, Slate J, Kruuk LEB, Pemberton JM (1998). Statistical confidence for likelihood-based paternity inference in natural populations. *Mol Ecol* **7**: 639–655.
- N'Goran JAK, Sounigo O, Clement D (2003). Comparaison des variétés utilisées par les cacaoculteurs ivoiriens. XIV International Cocoa Research Conference, Accra, Ghana, October 2003, pp 95–102.
- Pound FJ (1932). Studies on fruitfulness in cacao, I and II. First Annual Report on Cacao Research (1931), pp 24–28.
- Pritchard JK, Stephens M, Donnelly P (2000). Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.
- Reyes EH (1979). Mejoramiento genético del cacao en Venezuela. VII International Cocoa Research Conference, Douala, Cameroon, pp 513–518.
- Ritland K (2002). Extensions of models for the estimation of mating systems using n independent loci. *Heredity* **88**: 221–228.
- Ritland K, Jain S (1981). A model for the estimation of outcrossing rate and gene frequencies using n independent loci. *Heredity* **47**: 35–52.
- Shaw DV, Kahler AL, Allard RW (1980). A multilocus estimator of mating system parameters in plant populations. *Proc Natl Acad Sci USA* **78**: 1298–1302.
- Soria VJ (1978). The breeding of cacao (*Theobroma cacao* L.). *Trop Agric Res Ser* **11**: 161–168.
- Toxopeus H (1972). Cocoa breeding: a consequence of mating system, heterosis and population structure. Cocoa and Coconuts Conference in Malaysia, Kuala Lumpur, Malaysia. Incorporated Society of Planters, pp. 3–12.
- Voelcker OJ (1938). Self incompatibility of cacao II. Seventh Annual Report on Cocoa Research (Trinidad), pp 2–5.
- Weir BS, Cockerham CC (1984). Estimating F-statistics for the analysis of population structure. *Evolution* **38**: 1358–1370.