



Assessment of the legitimacy of cocoa tree hybrids (*Theobroma cacao* L.) resulting from simple crosses using single nucleotide polymorphism (SNP) markers in Côte d'Ivoire

Okayo Sandrine Minakou · Honorine Brigitte Guiraud · Inago Caudou Trébissou · Klotioloma Coulibaly · Tchréwa Stanislas Kouamé · Mathurin Yves Atchi · Bi Firmin Gouré · Attiapo Pepin Assi · Yaya Ouattara · Sidiky Bakayoko · Gnion Mathias Tah

Received: 1 October 2023 / Accepted: 16 July 2024
© The Author(s), under exclusive licence to Springer Nature B.V. 2024

Abstract The legitimacy of 13 hybrid families resulting from crosses via manual pollination techniques was studied using single nucleotide polymorphism (SNP) markers. The aim of this study was to ensure the conformity of offspring resulting from simple crosses. Fifty-one individuals (39 descendants and 12 parent clones) were analysed using 84 SNP markers. The results revealed nonsignificant genetic differentiation between parents and progenies ($P=0.323$). High allelic richness was also revealed, with the average number of alleles per locus ranging from 1.57 to 1.92, with an average of 1.77 in offspring and 2 in parents. Thus, 1% of the total variance was attributed to interpopulation variance, and 99% was attributed to intrapopulation variance. Low genetic differentiation ($F_{ST}<0.05$) and genetic distances ($0.03 < D < 0.09$) were observed between hybrid families F2, F8, F10, F11, F12, F13 and F14 and their respective parents. These results confirm the legitimacy of the hybrids produced and suggest that the manual pollination work carried out at CNRA sites

is reliable for faithfully reproducing hybrid material using known parents.

Keywords Legitimacy · *Theobroma cacao* L. · pollination · SNP · Côte d'Ivoire

Introduction

Theobroma cacao L. is the most important species of the *Theobroma* genus. This species belongs to the Malvaceae family. It is a diploid plant ($2n = 2x = 20$ chromosomes) with a genome size ranging from 411 to 494 Mb (Argout et al. 2011). It is cultivated mainly for its seeds, which, after fermentation and drying, yield merchantable cocoa, a raw material used in chocolate, food, cosmetics and pharmacology (Bruno 2003). Côte d'Ivoire is the world's leading producer of merchantable cocoa, with an annual production of approximately 2,200,000 tonnes (ICCO 2021), or 35% of the world supply. Cocoa farming plays an important role in the Ivorian economy, contributing 15% of gross domestic product (GDP) and 40% of export earnings.

However, cocoa farming faces a number of constraints potentially compromising its sustainability. These include the low level of use of improved planting material (Assiri et al. 2009), high pest pressure and climate change.

In Côte d'Ivoire, only improved planting material hybrids are distributed to farmers. Clonal

O. S. Minakou · T. S. Kouamé · M. Y. Atchi · B. F. Gouré · A. P. Assi · S. Bakayoko
Agroforestry Training and Research Unit, University Jean Lorougnon Guédé, BP 150, Daloa, Côte d'Ivoire

O. S. Minakou · H. B. Guiraud (✉) · I. C. Trébissou · K. Coulibaly · T. S. Kouamé · M. Y. Atchi · B. F. Gouré · A. P. Assi · Y. Ouattara · G. M. Tah
National Agronomic Research Center (CNRA)/Divo Research Station, BP 808, Divo, Côte d'Ivoire
e-mail: brigo2008@yahoo.fr

material is used only at the research station as progenitors for hybrid production. Under these conditions, it is necessary to assess the legitimacy of hybrids tested at experimental sites and distributed to farmers via seed gardens. This study, which focuses on crosses made at the CNRA research station in Divo (Côte d'Ivoire), aims to verify the legitimacy of crosses involved in the production of cocoa hybrid families.

Materials and methods

The study was carried out at the CNRA (National Agronomic Research Center), Divo station (5°50'27.8"N; 5°21' 30.1"W) in Côte d'Ivoire. The station is located approximately 17 km from Divo in the Lôh Djiboua region. The annual rainfall at the site ranges from 1200 mm to 1400 mm/year. The average temperature is 27 °C, and the relative humidity is 85% (Ehounou et al. 2019).

Plant material

The plant material is composed of 13 hybrid families resulting from crosses between 12 cocoa clones. These families, resulting from controlled manual pollination, were planted at the Divo research station in a control plot for the evaluation of drought resilience (Table 1).

Methods

Collection of leaf samples

A total of 57 leaf samples were collected and sent for genotyping to the Chaâbouni Genomics Laboratory (LGC) in England. These samples were taken from three trees per hybrid family and one tree per clone. Eight leaf discs were taken from green leaves of each genotype. These discs were inserted into sampling kits provided by the Chaâbouni Genomics Laboratory (LGC) in England. The kits containing the leaf samples were then sent for genotyping.

Genotyping

Genomic DNA was extracted at the Chaâbouni Genomics Laboratory (LGC) in England following a standardized mixed alkyl trimethyl ammonium bromide (MATAB) protocol as described by Pokou et al. (2009). Genotyping was carried out at LGC Genomics via 99 polymorphic single nucleotide polymorphisms (SNPs) identified on the 10 chromosomes of the cocoa genome (Argout et al. 2008).

Statistical analysis of the genotyping data

The legitimacy of the offspring from single crosses was studied by analysing genetic diversity within and between populations. Genetic diversity parameters were calculated via Genetix V4.05 (Belkhir et al.

Table 1 List of hybrid families and clones

Clones (parents)	Genetic groups	Hybrid families	Genealogy
IMC67	Iquitos	F1	(UPA402×UF676)
MOQ413	FHA	F2	(UPA409×IFC1)
UPA409	FHA	F3	(UPA608×IFC412)
IFC5	Amelonado	F5	(UPA603×UF667)
ICS46	Criollo	F6	(UPA409×POR)
UF676	Trinitario	F7	(T85/799×IFC15)
ICS1	Trinitario	F8	(SCA6×ICS1)
PA150	Maranon	F10	(T79/501×IFC5)
UPA402	FHA	F11	(IFC720×ICS46)
SCA6	Contamana	F12	(IMC67×IFC1)
T85/799	FHA	F13	(MOQ413×SCA6)
POR	Criollo	F14	(POR×T60/887)
		F15	(PA150×POR)

2004), Fstat v2.9.3.2 (Goudet 1995) and GenAlEx v6.1 (Peakall and Smouse 2012). In addition, an analysis of molecular variance (AMOVA) was also performed via GenAlEx v6.5 software (Peakall and Smouse 2012). For each genotype, several parameters were determined. These indices are the number of alleles per locus (A), the percentage of polymorphic loci (P) at the 95% and 99% thresholds, observed (H_o) and expected (H_e) heterozygosity, total genetic diversity or total heterozygosity (H_t), the genetic differentiation index (G_{ST}), heterozygote deficit within populations (F_{IS}), and interpopulation genetic diversity (D_{ST}). H_t was estimated as the sum of the intrapopulation genetic diversity (H_S) and interpopulation genetic diversity (D_{ST}).

G_{ST} represents the fraction of total genetic diversity corresponding to the genetic difference between populations. This parameter was measured using the formula $G_{ST} = D_{ST} / H_t$ (Nei 1978). Intrapopulation heterozygote deficit and F_{st} [which estimates the proportion of interpopulation genetic diversity to total diversity (Weir and Cockerham 1984)] were also determined. The Mann–Whitney U test was performed via Statistica 7.1 software.

An analysis of molecular variance (AMOVA) was performed to estimate the distribution of diversity within families and parent clones to quantify genetic variation between individuals. AMOVA was also used to determine the F_{ST} (F statistic), which estimates the proportion of interpopulation genetic diversity to total diversity (Weir and Cockerham 1984). A principal coordinate analysis (PCoA) was carried out on 51 individuals to graphically represent each family and clone in a two-dimensional plane. This analysis is based on the dissimilarity matrix between populations. All these analyses were carried out via GenAlEx 6.5 software (Peakall and Smouse 2012).

A phylogenetic analysis was carried out using DARwin v.5 software (Perrier and Jacquemoud-Collet 2006). This analysis involved producing a dendrogram via a Bayesian algorithm that is based on sequential clustering of neighbouring genotypes. The robustness of the tree nodes was tested by applying 1000 bootstraps on the basis of individual repeatability.

Single nucleotide polymorphism (SNP) marker data were subjected to genetic structure assessment of hybrid families and parental clones using STRU CTURE 2.3.4 software (Porrás-Hurtado et al. 2013)

to assign each individual to a genetic group. The optimal number of K groups was determined via the method of Evanno et al. (2005). It was evaluated on the Structure Harvester platform (<http://taylor0.biolo.gy.ucla.edu/structureHarvester/>) (Earl and von Holdt 2012) by applying the admixture model (Lawson et al. 2018). In this analysis, K varied from 1 to 5, with five interactions in the execution of the analysis program. Among the five interactions, those with the highest Ln Pr (X/K) values were selected and represented as a coloured bar chart (Takrama et al. 2014). Next, an analysis probabilistically assigned each individual to a group using a Bayesian algorithm.

The membership coefficient (Q value), which varies between 0 and 1, was used to designate the membership of a family or clone to a genetic group. If (Q value < 0.80) for a specific group, then the family or clone was considered a hybrid, and if (Q value ≥ 0.80) for a specific group, then the individual was considered a parental clone (Li et al. 2021).

Results

Marker polymorphisms

The percentage or rate of marker polymorphisms ($P=95\%$) was 100% for the parent clones. In the progenies, the percentage of marker polymorphisms ranged from 57.14 (UPA402×UF676) to 92.66% (UPA409×IFC1), with an average of 77.29% (Table 2).

Genetic diversity of parent clones and hybrid families

The average number of alleles per locus (A) was 2, and the effective number of alleles (A_e) obtained between parent clones was 1.69. The Shannon diversity index value was 0.57. The expected heterozygosity ($H_e=0.39$) was greater than the observed heterozygosity ($H_o=0.33$). The value of the fixation index (F_{is}) was 0.18.

The average number of effective alleles (A_e) ranged from 1.57 (UPA402×UF676) to 1.92 (UPA409×IFC1), with an average of 1.77. The Shannon diversity index (I) was high for the (UPA608×IFC412) family at 0.54. The expected heterozygosity was lower than the observed heterozygosity ($H_o=0.42$ and $H_e=0.30$). The fixation

Table 2 Percentage of polymorphism in each population

Populations	Polymorphism percentage
Parents	100
Progenies	
(UPA409 × IFC1)	92.86
(T79/501 × IFC5)	89.29
(POR × T50/501)	89.29
(UPA608 × IFC412)	88.10
(SCA6 × ICS1)	82.14
(IMC67 × IFC1)	77.38
(IFC720 × ICS46)	75.00
(UPA409 × POR)	75.00
(T85/799 × IFC15)	71.43
(MOQ413 × SCA6)	71.43
(UPA603 × UF667)	69.05
(PA150 × POR)	66.67
(UPA402 × UF676)	57.14
Mean	77.29

index (F_{is}) ranged from -0.52 (UPA402 × UF676) to -0.16 (POR × T50/501), with an average of -0.34 (Table 3).

Table 3 Values of genetic diversity parameters assessed within hybrid families and clones

Populations	A	A_e	I	H_o	H_e	F_{is}
Parents	$2. \pm 0$	1.69 ± 0.02	0.57 ± 0.01	0.33 ± 0.02	0.39 ± 0.01	0.18 ± 0.04
(UPA409 × IFC1)	1.92 ± 0.02	1.64 ± 0.03	0.53 ± 0.02	0.48 ± 0.03	0.36 ± 0.01	-0.29 ± 0.05
(T79/501 × IFC5)	1.89 ± 0.03	1.60 ± 0.03	0.51 ± 0.02	0.45 ± 0.03	0.34 ± 0.01	-0.29 ± 0.05
(POR × T50/501)	1.89 ± 0.03	1.61 ± 0.03	0.51 ± 0.02	0.40 ± 0.03	0.35 ± 0.01	-0.16 ± 0.05
(UPA608 × IFC412)	1.88 ± 0.03	1.67 ± 0.03	0.54 ± 0.02	0.48 ± 0.03	0.37 ± 0.01	-0.29 ± 0.05
(SCA6 × ICS1)	1.82 ± 0.04	1.55 ± 0.03	0.47 ± 0.02	0.42 ± 0.03	0.32 ± 0.01	-0.29 ± 0.04
(IMC67 × IFC1)	1.77 ± 0.04	1.55 ± 0.04	0.45 ± 0.02	0.43 ± 0.03	0.31 ± 0.02	-0.36 ± 0.03
(UPA409 × POR)	1.75 ± 0.04	1.50 ± 0.04	0.42 ± 0.02	0.45 ± 0.04	0.28 ± 0.02	-0.48 ± 0.04
(IFC720 × ICS46)	1.75 ± 0.04	1.56 ± 0.04	0.45 ± 0.03	0.39 ± 0.03	0.31 ± 0.02	-0.24 ± 0.05
(T85/799 × IFC15)	1.71 ± 0.05	1.46 ± 0.04	0.40 ± 0.03	0.39 ± 0.03	0.27 ± 0.02	-0.39 ± 0.04
(MOQ413 × SCA6)	1.71 ± 0.05	1.50 ± 0.04	0.41 ± 0.03	0.41 ± 0.03	0.28 ± 0.02	-0.40 ± 0.04
(UPA603 × UF667)	1.69 ± 0.05	1.48 ± 0.04	0.40 ± 0.03	0.37 ± 0.03	0.27 ± 0.02	-0.32 ± 0.05
(PA150 × POR)	1.66 ± 0.05	1.43 ± 0.04	0.37 ± 0.03	0.37 ± 0.03	0.25 ± 0.02	-0.41 ± 0.04
(UPA402 × UF676)	1.57 ± 0.05	1.42 ± 0.04	0.34 ± 0.03	0.37 ± 0.04	0.23 ± 0.02	-0.52 ± 0.04
Means	1.77 ± 0.04	1.54 ± 0.04	0.44 ± 0.02	0.42 ± 0.03	0.30 ± 0.02	-0.34 ± 0.04

A a number of effective alleles, A_e dark number of effective alleles, I Shannon information index, H_o observed heterozygosity, H_e expected heterozygosity, F_{is} fixation index

Interpopulation genetic diversity

Molecular analysis of variance revealed nonsignificant genetic differentiation between parents and families ($P=0.323$). This analysis revealed that 1% of the total variance was attributed to interpopulation variance and that 99% was attributed to intrapopulation variance (Table 4).

Differentiation between hybrid families and parental clones

Fixation index values between populations ($F_{is}=-0.29$) and for all populations ($F_{it}=-0.13$) indicate excess heterozygosity. Moderate genetic differentiation between the clones and their progeny ($F_{ST}=0.12$) was observed. The value of N_m gene flow per population was 2.8 (Table 5).

Genetic differentiation coefficients (F_{ST}) and genetic distances (D) were calculated for each pair of hybrid families and clones. Analysis of the two matrices revealed significant differentiation (F_{ST}) and genetic distances (D) between populations.

Table 6 presents the genetic differentiation coefficients calculated between families and parents. F_{ST} ranged from 0.035 (between F13 and F15) to 0.145 (between F1 and F3), whereas the genetic distances

Table 4 Distribution of genetic variation according to AMOVA for hybrid families and parents

Sources of variation	DF	SCE	CM	Variance component (%)	Variance percentage	<i>P</i>
Interpopulation (H_s)	13	13.24	1.018	0.008	1	0.323
Intrapopulation (H_i)	37	36.583	0.989	0.989	99	
Total (H_t)	50	49.824		0.997	100	

DF degree of freedom, *SCE* sum of squares of deviations, *CM* mean squares

between families varied from 0.022 (between F13 and F15) to 0.153 (between F1 and F3) (Table 7).

Principal coordinate analysis (PCoA)

The distributions of parent clones (purple) and progeny (black) along Axes 1 and 2 of the principal

Table 5 Mean values of genetic diversity parameters assessed within populations

Genetics diversity parameters between clones and hybrids	Mean \pm standard deviation
F_{is}	-0.29 ± 0.02
F_{it}	-0.13 ± 0.02
F_{ST}	0.12 ± 0
N_m	2.8 ± 0.42

F_{ST} : coefficient of genetic differentiation of the sub-populations, F_{is} : wright fixation index of the sub-population, N_m : gene flow per population, F_{it} : wright fixation index of the total population

component analysis (PCA) are shown in Fig. 1. These two axes contribute 37.29% of the total variability. The distribution of individuals in the 1 and 2 factorial planes reveals two main groups along Axis 1. The first group (**I**) comprises parent clones SCA6, POR, ICS46, UF667, MOQ 413, and ICS1 and their progeny. Group **II** comprises parents IMC67, T85/799, PA150, UPA402, UPA409, and IFC 5 and their progeny. Axis 2 structures each group (**I** and **II**) into four subgroups (**I1**, **I2**, **II1**, **II2**) according to the proximity of descendants and parent clones. Subgroup **I1** comprises descendants close to the SCA6 parent. The remaining parents and descendants of Group **I** are grouped together in subgroup **I2**. Subgroup **II1** includes controls IMC67, UPA409, UPA402, T85/799, and PA150 and their descendants. Subgroup **II2** includes the parent clone IFC5 and its close descendants.

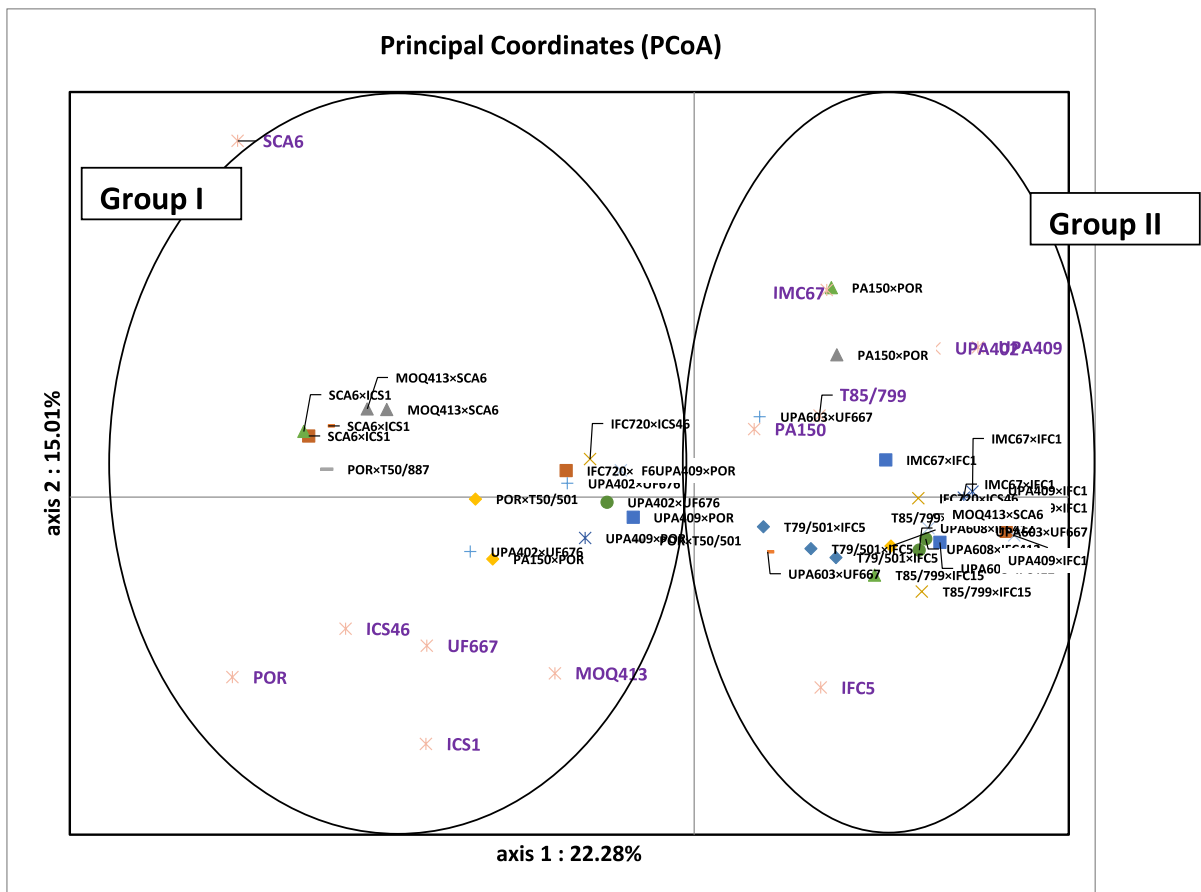
Table 6 Matrix of F_{ST} genetic differentiation coefficients calculated between families and parents

	F1	F2	F3	F5	F6	F7	F8	F10	F11	F12	F13	F14	F15
F1													
F2	0.096												
F3	0.145	0.055											
F5	0.077	0.067	0.114										
F6	0.092	0.077	0.114	0.067									
F7	0.090	0.070	0.115	0.068	0.061								
F8	0.114	0.074	0.102	0.105	0.087	0.086							
F10	0.092	0.044	0.075	0.064	0.066	0.060	0.079						
F11	0.088	0.057	0.096	0.074	0.081	0.074	0.061	0.063					
F12	0.099	0.068	0.109	0.073	0.089	0.084	0.071	0.081	0.051				
F13	0.097	0.064	0.106	0.059	0.052	0.042	0.095	0.059	0.071	0.089			
F14	0.105	0.045	0.058	0.069	0.070	0.085	0.084	0.046	0.074	0.088	0.061		
F15	0.102	0.063	0.121	0.062	0.048	0.047	0.099	0.063	0.076	0.090	0.035	0.075	

Bold mean that difference between family and parent is significant

Table 7 Matrix of genetic distances (*D*) calculated between families and parents

	F1	F2	F3	F5	F6	F7	F8	F10	F11	F12	F13	F14	F15
F2	0.085												
F3	0.153	0.066											
F5	0.051	0.059	0.121										
F6	0.073	0.079	0.133	0.052									
F7	0.067	0.063	0.125	0.051	0.048								
F8	0.105	0.089	0.126	0.106	0.089	0.088							
F10	0.076	0.051	0.090	0.054	0.061	0.048	0.093						
F11	0.068	0.051	0.104	0.058	0.076	0.064	0.060	0.060					
F12	0.084	0.072	0.128	0.060	0.090	0.078	0.076	0.089	0.048				
F13	0.072	0.056	0.113	0.042	0.037	0.029	0.100	0.049	0.058	0.086			
F14	0.098	0.051	0.069	0.061	0.067	0.084	0.098	0.049	0.073	0.098	0.053		
F15	0.075	0.050	0.126	0.043	0.032	0.034	0.101	0.049	0.061	0.082	0.022	0.064	

**Fig. 1** Projection of *Theobroma cacao* hybrid families and parent clones in plane 1–2 of the Principal Coordinate Analysis (PCoA)

Phylogenetic relationships

The hierarchical ascending classification (HAC) performed on the basis of the averages of descriptive characters produces a dendrogram showing two sets of molecular diversity. An examination of the phylogenetic tree revealed 2 major genetic groups whose branches were supported by probabilities ranging from 3 to 98%.

The first group comprised six (red) clones (UF676, ICS46, POR, ICS1, MOQ413 and SCA6) and the following 6 progeny: (UPA 402×UF676); (UPA 409×POR); (POR×T60/887); (SCA6×ICS1); (MOQ 413×SCA6); and (IFC 720×ICS 46).

The second group was made up of six (red) progenitors (IFC5, IMC67, UPA402, T85/799 and

PA150) and the following 8 offspring: (UPA608 X IFC412), (UPA409 X IFC1), (IMC67 X IFC1), (T85/799 X IFC15), (T79/501 X IFC5), and (PA150 X POR) (UPA603 X UF667) (Fig. 2).

Genetic structuring of parental clones and families

The results of the Bayesian analyses used to determine the number of clusters within parents and descendants are shown in Fig. 3. Analysis of the figure revealed the existence of two major genetic groups (clusters) (K=2).

Figure 4 shows the genetic structure of the two populations (parents and offspring). It shows the contribution of each cluster to the constitution of the individuals. These clusters are associated with several control genotypes. Cluster 1 (red) is associated

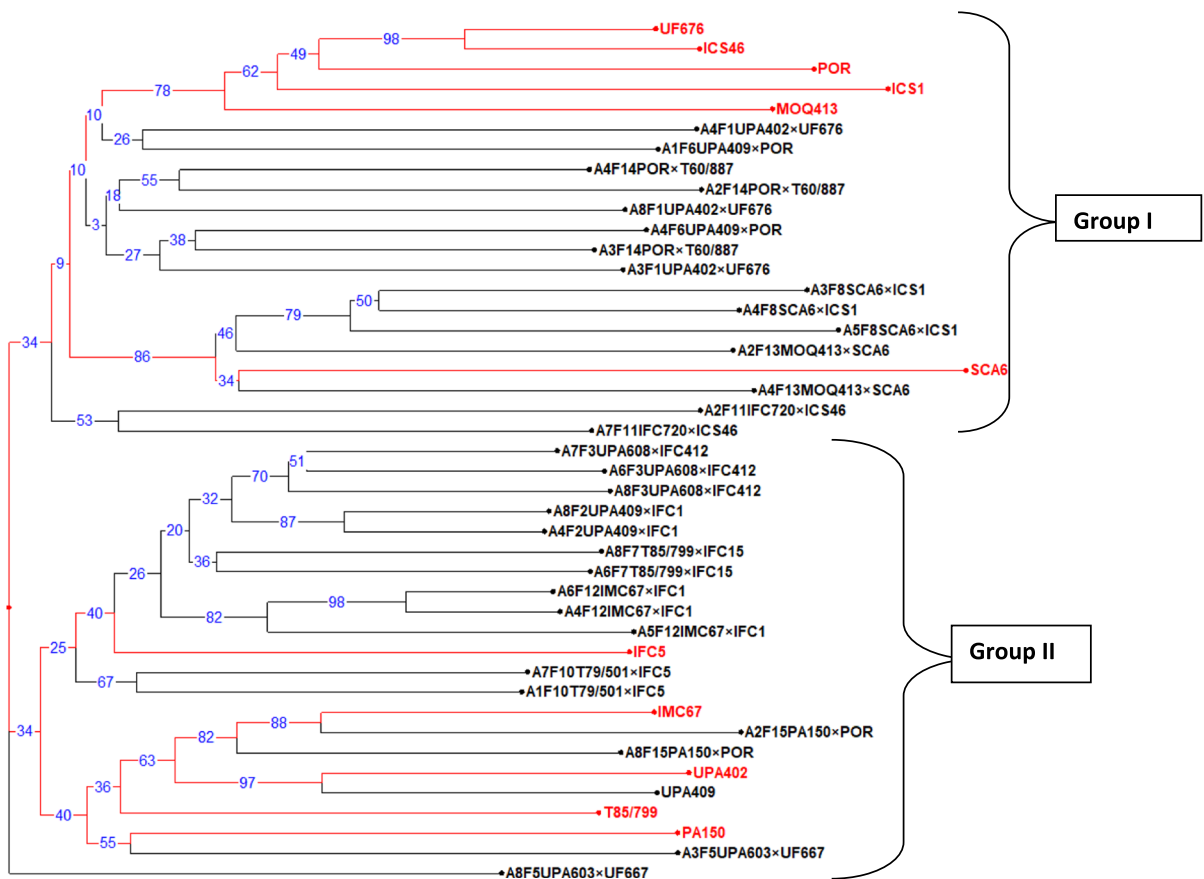


Fig. 2 Dendrogram presenting the hierarchical ascending classification (HAC) of hybrid families and parents of *cocoa trees* using 84 single nucleotide polymorphism (SNP) markers

Fig. 3 Graph showing the values of the Delta K statistic, allowing 2 to be considered the optimum number

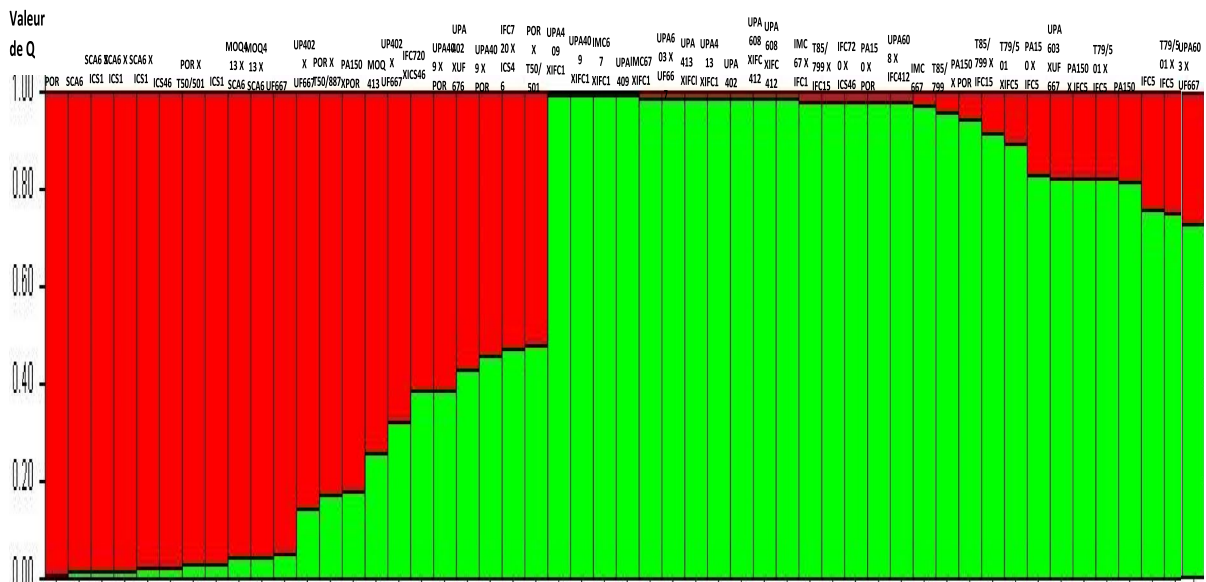
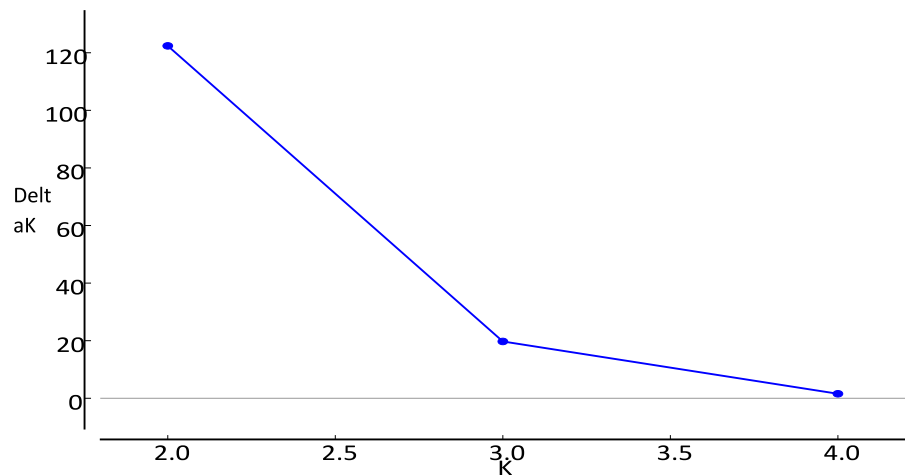


Fig. 4 Genetic structure of hybrid families and parent clones of cocoa trees by the Bayesian method using 84 single nucleotide polymorphism (SNP) loci. Each color represents a genetic

group, and each bar is a clone or hybrid family with the probability of belonging to a genetic group ranging from 0 to 1 (Q value)

with the parent clones POR, SCA6, ICS46, ICS1 and UF667. Cluster 2 (green) includes the parent clones (UPA 402, UPA409, IMC 67, T85/799, IFC5, and PA150). Analysis of the figure revealed that 16 hybrids had a strong genetic contribution from Cluster 1 (red). On the other hand, 23 hybrid offspring strongly contributed to the genetic constitution of Cluster 2 (green).

Discussion

The study of the legitimacy of a few hybrid families resulting from simple crosses is an approach that can be used to ensure conformity between parents and progenies resulting from manual pollination. In this study, 84 informative SNP markers were used to assess the genetic conformity of 39 offspring from 12 parent clones. Indeed, of the more than 1,000 SNP markers identified on chromosomes 1 to 10 of the

cocoa genome by Argout et al. (2008), only 99 are commonly used for genotyping research because of their high level of polymorphism and discriminatory power.

The results of our study confirmed the polymorphism of the markers used. Indeed, the SNP markers were polymorphic for the entirety of the two populations studied (parents and descendants). Our results are in line with those of Ji et al. (2013), whose work observed marker polymorphisms by assessing the genetic diversity and relatedness of cocoa varieties from Honduras and Nicaragua via 70 SNP markers.

In addition, nonsignificant genetic differentiation between parents and offspring ($P=0.323$) was revealed. This finding indicates that there is significant similarity between parent clones and their offspring. Similar results ($P=0.285$) were reported in a study of genetic diversity and relatedness between cocoa trees from Bogua and Utcubamba in northern Peru, in which 192 SNP markers were used (Danilo et al. 2022).

The average number of alleles per locus ranged from 1.57 to 1.92, with an average of 1.77 in the offspring and 2 in the parents. The results revealed high allelic richness in both populations. Numerous investigations on the diversity of *T. cacao* have also indicated significant allelic richness in this species. Indeed, diversity research on cocoa populations carried out by Danilo et al. (2022) reported averages of 1.52 and 1.59 alleles per population, respectively.

This allelic richness could also be attributed to the gametophytic self-incompatibility system characteristic of *T. cacao* (Royaert et al. 2011; Lanaud et al. 2017). This high allelic richness is also an asset for conservation strategies (Bataillon et al. 1996).

Our work also revealed a high contribution of intrapopulation genetic diversity ($H_s=99\%$) to total genetic diversity ($H_t=100\%$), with an interpopulation diversity of 1%. Because of their high diversity, parents and offspring from single crosses could constitute a valuable reservoir of genetic resources for many selection criteria (Ouédraogo et al. 2005). As a result, this plant material can be used in breeding programs aimed at genetic improvement of the cocoa tree for resistance to induced water deficit, for example. Indeed, the high diversity observed within these two populations could facilitate their integration within the recurrent and reciprocal selection program currently underway in Côte d'Ivoire (Pokou et al. 2009).

The expected heterozygosity (H_e) was 0.30 for the progeny population and 0.39 for the parental population. These values are close to those obtained by Gopaulchan et al. (2020) ($H_e=0.32$), whose work focused on the genetic diversity of cocoa (*Theobroma cacao* L.) in Dominica via 180 SNP markers. On the other hand, these values are higher than the results from Padi et al. (2015) ($H_e=0.24$), whose work on the utility of SNP fingerprints from 64 loci examined the diversity, mislabelling and parentage of 2,551 trees from six seed fields and hybrid progeny plots and farmer accessions in Ghana.

The fixation index values ranged from -0.52 to -0.16 for the hybrid progenies and 0.18 for the parental clones. These values are less than unity, indicating excess heterozygosity in these populations (Wright 1965).

Low genetic differentiation ($F_{ST}<0.05$) and genetic distances ($0.03<D<0.09$) were observed between hybrid families F2, F8, F10, F11, F12, F13 and F14 and their respective parents. Thus, the F_{ST} values and genetic distances (D) confirmed the proximity between these clones and their respective progenies. These results are similar to those of Jaime et al. (2017), who reported low genetic distances ($0.06<D<0.07$) between cocoa trees from Chucho and Beni in Colombia, indicating strong similarity between these two populations.

In addition, principal component analysis (PCA) and hierarchical ascending classification (HAC) revealed clustering of the parental clones SCA6, PA150, POR, IFC5 and IMC67 with their progenies. The clones used in this trial would therefore be genetically close to their progeny. This proximity can be explained by the closely related genetic basis of the parental material. Similar results were reported by Koffi et al. (2022) in coconut palms. These authors explained that genetic differences or similarities in coconut trees (parents and offspring) are favoured by the contribution of genes from the male progenitor.

STRUCTURE analysis of hybrid families and clones revealed that families F9 (PA150 X IFC5), F15 (PA150 X POR), F12 (IMC67 X IFC1), F14 (POR X T50/501), F11 (IFC720 X ICS46) and F8 (SCA6 X ICS1), and F13 (MOQ413 X SCA6) have membership rates (Q values) below 0.80 with the Maranon (PA150), Iquitos (IMC67), Criollo (POR and ICS46) and Contamana (SCA6) genetic groups. Membership rates less than 0.80 indicate that these individuals are

offspring of different genetic groups. Membership rates less than 0.80 were reported in a study comparing traditional Madagascar varieties with their progenitors (Li et al. 2021). In this study, varieties with Q values less than 0.80 were considered to have a strong resemblance to their progenitors.

These results indicate that the seed production technique in cocoa seed fields adopted by the CNRA is reliable in guaranteeing the expected performance of these hybrids. This work highlights the great level of satisfaction expressed by cocoa growers regarding the performance of hybrids distributed through CNRA seed fields.

Conclusion

The data obtained in this work, aiming at the legitimacy of hybrids, revealed little differentiation between the hybrids and the parents for seven hybrid families in addition to a high degree of allelic richness.

Therefore, these results constitute an asset for the large-scale dissemination of hybrid families presenting traits of agronomic and/or technological interest and for the conservation of these populations.

Acknowledgements The authors would like to thank the National Agronomic Research Center (CNRA) for providing plant material and technical facilities. The authors also thank the Cocoa Board in Côte d'Ivoire and FIRCA for funding the activities carried out in the framework of this work.

Author contributions Okayo Sandrine Minakou, Honorine Brigitte Guiraud, Inago Caudou Trébissou and Gnion Mathias Tahi—wrote the main manuscript text and statistical analysis. All authors reviewed the manuscript.

Funding The authors have not disclosed any funding.

Declarations

Conflict of interest The authors declare no conflict of interest.

References

- Argout X, Fouet O, Wincker P, Gramacho K, Legavre T, Sabau X, Risterucci A, Da Silva C, Cascardo J, Allegre M, Kuhn D, Verica J, Courtois B, Loor G, Babin R, Sounigo O, Ducamp M, Guiltinan M, Ruiz M, Alemanno L, Machado R, Phillips W, Schnell R, Gilmour M, Rosenquist E, Butler D, Maximova S, Lanaud C (2008) Towards the understanding of the cocoa transcriptome: production and analysis of an exhaustive dataset of ESTs of *Theobroma cacao* L. generated from various tissues and under various conditions. *BMC Genomics* 9:512
- Argout X, Sale J, Aury JM, Gwiltinan MJ, Droc G, Gouzy J (2011) The genome of *Theobroma cacao*. *Nat Genet* 43:101–108
- Assiri AA, Yoro GR, Dehevels O, Kebe BI, Keli ZI, Adiko A, Assa A (2009) Les caractéristiques agronomiques des vergers de cacaoyers (*Theobroma cacao* L.) en Côte d'Ivoire. *J Anim Plant Sci* 2(1):55–56
- Bataillon TM, David JL, Schoen DJ (1996) Neutral genetic markers and conservation genetics: simulated germplasm collections. *Genetics* 144:409–447
- Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (2004) GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Science Open Laboratoire Genome, Populations, Interactions; CNRS UMR 5000; Université Montpellier II, Montpellier (France). <http://www.genetix.univmontp2.fr/genetix/intro.htm>
- Bruno (2003) De la fève ivoirienne de cacao à la plaquette française de chocolat noir: transmission des prix, partage de la valeur et politique de concurrence Nord/Sud. Montpellier: CIRAD-AMIS, N° de rapport : CIRAD-AMIS N°36, N° de rapport: CP_SIC 1602, p 62
- Danilo EB, Lambert AM, Martha SC, Amrita M, Manuel O (2022) Genetic diversity and population structure of fine aroma cacao (*Theobroma cacao* L.) from north Peru revealed by single nucleotide polymorphism (SNP) markers. *Front Écol Evol*. <https://doi.org/10.3389/fevo.2022.895056>
- Earl DA, Von Holdt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour* 4:359–361
- Ehounou JN, Kouamé B, Tahi GM, Kassin KE, Kotaix JA, Dékoula SC, Guy YF, Kouadio KH, N'guessan EBA, Soro N (2019) Global warming influence on major seasonal and intra-seasonal rainfall indicators for sustainable cocoa production in West-central Côte d'Ivoire. *J Exp Agric* 40(5):1–10
- 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
- Gopaulchan D, Motilal LA, Kalloo RK, Mahabir A, Moses M, Joseph F (2020) Genetic diversity and ancestry of cacao (*Theobroma cacao* L.) in Dominica revealed by single nucleotide polymorphism markers. *Genome* 63:583–595
- Goudet J (1995) FSTAT (1.2): a computer program to calculate F-statistics. *J Hered* 86(6):485–486
- ICCO (2021) ICCO quarterly bulletin of cocoa statistics. Vol XLVII, No 2, Cocoa year 2020/21
- Jaime A, Osorio-Guarín JBC, Roberto AC, Yeny PZ, Constanza Q, Gerardo GS, Roxana Y (2017) Colombia a source of cacao genetic diversity as revealed by the population structure analysis of germplasm bank of *Theobroma cacao* L. *Front Plant Sci* 8:1994
- Ji K, Zhang D, Motilal LA, Boccara M, Lachenaud P (2013) Genetic diversity and parentage in farmer varieties of cacao (*Theobroma cacao* L.) from Honduras and

- Nicaragua as revealed by single nucleotide polymorphism (SNP) markers. *Genet Resour Crop Evol* 60:441–453
- Koffi EZ, Thiémélé DEF, Daramcoum WAMP, Yao SDM, Konan KJ, Sié RS, Diarrassouba N (2022) Évaluation de la diversité moléculaire de quelques descendances hybrides F1 NJM x GVT de cocotier (*Cocos nucifera* L.) en Côte d'Ivoire à l'aide de marqueurs microsatellites. *Int J Innov Appl Stud* 36(1):54–66
- Lanaud C, Fouet O, Legavre T, Lopes U, Sounigo O, Eyango MC, Mermaz B, Da Silva MR, Loor Solorzano RG, Argout X (2017) Deciphering the *Theobroma cacao* self-incompatibility system: from genomics to diagnostic markers for self-compatibility. *J Exp Bot* 68(17):4775–4790
- Lawson JD, Van Dorp L, Falush D (2018) A tutorial on how not to over-interpret STRUCTURE and admixture bar plots. *Nat Commun* 9(3258):1–11
- Li Y, Zhang D, Motilal LA, Lachenaud P, Mischke S, Meinhardt LW (2021) Traditional varieties of cacao (*Theobroma cacao*) in Madagascar: their origin and dispersal revealed by SNP markers. *Beverage Plant Res* 1:4
- Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 89(583–590):1978
- Ouédraogo M, Maquet M, Baudoin JP (2005) Etude comparative de la diversité et de la structure génétique de populations sauvages de *Phaseolus lanatus* L. à l'aide des marqueurs enzymatiques et microsatellites. *J Appl Biosci* 43:2947–2957
- Padi FK, Ofori A, Takrama J, Djan E, Opoku SY (2015) The impact of SNP fingerprinting and parentage analysis on the effectiveness of variety recommendations in cacao. *Tree Genet Genomes* 11:44
- Peakall R, Smouse P (2012) GenA1Ex 6.5: genetic analysis in excel. Population genetic software for teaching and research—an update. *Bioinformatics* 28:2537–2539
- Perrier X, Jacquemoud-Collet JP (2006) DARwin Software. <http://darwin.cirad.fr/darwin>
- Pokou ND, N'Goran JAK, Lachenaud P, Eskes AB, Motamayor JC, Schnell R, Kolesnikova-Allen M, Clément D, Sangaré A (2009) Recurrent selection of cocoa populations in Côte d'Ivoire: comparative genetic diversity between the first and second cycles. *Plant Breed* 128:514–520
- Porrás-Hurtado L, Ruiz Y, Santos C, Phillips C, Carracedo Á, Lareu MV (2013) An overview of STRUCTURE: applications, parameter settings, and supporting software. *Front Genet* 4(98):1–13
- Royaert S, Phillips-Mora W, Leal A, Cariaga K, Steven BJN, Kuhn D, Schnell R, Motamayor JC (2011) Identification of marker-trait associations for self-compatibility in a segregating mapping population of *Theobroma cacao* L. *Tree Genet Genomes* 7:1159–1168
- Takrama J, Dadzie AM, Opoku FK, Padi FK, Adomako B, Asuampomah Y, Livingstone DS, Motamayor JC, Schnell RJ, Kuhn RJ (2014) Application de la technologie des marqueurs SNP dans le programme de sélection du cacao au Ghana. *Afr Crop Sci J* 20:67–75
- Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution* 38:1358–1370
- Wright S (1965) The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution* 19:395–420
- Zhang D, Mischke S, Johnson E, Phillips-Mora W, Meinhardt L (2011) Molecular characterization of an international cacao collection using microsatellite markers. *Tree Genet Genomes* 5:1–10

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.