Tree Genetics & Genomes (2009) 5:329–337
DOI 10.1007/s11295-008-0185-6

ORIGINAL PAPER

A genomewide admixture mapping study for yield factors and morphological traits in a cultivated cocoa (*Theobroma cacao* L.) population

Maria Marcano · Sonia Morales · Maria Theresa Hoyer · Brigitte Courtois · Ange Marie Risterucci · Olivier Fouet · Tatiana Pugh · Emile Cros · Ventura Gonzalez · Manuel Dagert · Claire Lanaud

Received: 21 May 2008 / Revised: 4 August 2008 / Accepted: 1 September 2008 / Published online: 12 November 2008

Abstract The selection of productive varieties of modern Criollo cocoa, showing fine aromatic qualities in their beans, is of major interest for some producing countries, such as Venezuela. Cultivated populations of Modern Criollo or Trinitario varieties may be suitable for admixture mapping analysis, as large blocks of alleles derived from two identified divergent ancestors, recently admixed, are still preserved, after a few generations of recombination, similar to experimental mapping progenies. Two hundred and fifty-seven individuals from a cultivated population of Modern Criollo were selected and analysed with 92 microsatellite markers distributed along the genome. This population exhibited a wide range of variability for yield factors and morphological features. Population structure analysis identified two main subgroups corresponding to the admixture from the two ancestors Criollo and Forastero. Several significant associations between markers and phenotypic data (yield factors and morphological traits) were identified by a least squares general linear model (GLM) taking into account the population structure and the percentage of admixture of each individual. Results were compared with classical QTL analyses previously reported for other cacao populations. Most markers associated to quantitative traits were very close to QTLs detected formerly for the same traits. Associations were also identified between markers and several qualitative traits including the red pigmentation observed in different organs, mainly associated to common markers in linkage group 4.

Keywords Admixture mapping · *Theobroma cacao* L. · Microsatellite · Quantitative trait loci · Yield · Qualitative traits

Communicated by F. Gmitter

M. Marcano (✉) · S. Morales · M. T. Hoyer · M. Dagert
Facultad de Ciencias, Laboratorio de Genética y Química Celular (GeQuimCel), Universidad de Los Andes (ULA),
La Hechicera, Mérida 5101, Venezuela
e-mail: marseg@ula.ve

M. Marcano · B. Courtois · A. M. Risterucci · O. Fouet · T. Pugh · C. Lanaud
Centre de Coopération Internationale en Recherche Agronomique pour le Développement CIRAD-BIOS. TA 40/03,
UMR DAP (Développement et Amélioration des Plantes).
Equipe "Génome et Sélection",
Av. Agropolis,
34398 Montpellier Cedex 5, France

T. Pugh
Facultad de Agronomía, Departamento de Genética,
Universidad Central de Venezuela,
Av. El Limón,
Maracay 2101 Estado Aragua, Venezuela

E. Cros
CIRAD - Department Perennial Cultures/UPR33,
TA 80/16, Av. Agropolis,
34398 Montpellier Cedex 5, France

V. Gonzalez
Instituto Nacional de Investigaciones Agropecuarias (INIA),
Av. Universidad,
Maracay 2101 Estado Aragua, Venezuela

Introduction

The cacao tree (*Theobroma cacao* L.), native from the basin of the Amazon River in South America (Cheesman 1944), is cultivated in Africa, Asia, America, and Oceania (FAO 2006). First domesticated by the Mayan civilization 2,000 years ago (Paradis 1979), cacao cultivars grown in pre-Columbian times and first disseminated through the Spanish colonies, corresponded to "Criollo" types, which produce "Fine" or "Aromatic" cocoa, with characteristic sensorial properties. Introgressions from Amazonian "Forastero" types into pure Criollo plants started when hybrid trees were dispersed from Trinidad after 1727, when a natural disaster destroyed most of the Criollo plants in this Caribbean island. Remnant individuals of Criollo hybridized with introduced Forastero Amazonian cacao resulting in the development of more vigorous, precocious, and productive hybrids (Trinitario), which were also less susceptible to diseases than pure Criollo trees. "Trinitario" plants were reintroduced in Central and South America where they crossed to Criollo types previously cultivated in these lands, ending in a larger genetic mixing and originating hybrids later regarded as "Modern Criollo" cultivars (Motamayor et al. 2002). "Modern Criollo" progressively replaced pure Criollo plants, but the Criollo aromatic quality traits tend to fade with continuous introgressions from Forastero genotypes (Pittier 1935). "Modern Criollo" cacao, constitutes an extensive group of cultivars currently grown throughout the world with a very limited genetic basis mainly represented by one Criollo and one Forastero ancestors (Motamayor et al. 2003). "Fine aromatic" Criollo cocoa, although accounting for a small contribution to the cocoa world production (FAO 2005), represents an important market "niche" for producing countries such as Venezuela (Eskes and Lanaud 1997).

Criollo and Forastero cacao cultivars are morphologically different. Criollo plants produce pods with typical large and broad seeds, white or very light violet or pink, while Forastero pods contain purple and flattened seeds. The purple pigmentation, besides other seed compounds, naturally developed and enhanced by post harvest treatments, could be the basis for such differential qualities between Criollo and Forastero (Jeanjean 1995). Purple/red pigmentations observed in cocoa seeds and in several other organs (leaves, staminodes, flowers) are due to anthocyanidin pigments (Forsyth and Quesnel 1957), one of the three main subgroups of flavanols, with catechin and procyanidin polymers (Stafford 1990). The antioxidant properties of flavanols, and particularly those from chocolate, have gained attention in the last years for their potential to protect human health against cardiovascular disease and cancer (Borchers et al. 2000; Wollgast and Anklam 2000; Gu et al. 2006). Cocoa is rich in polyphenols particularly in catechins (flavan-3-ols) and procyanidins. Flavanols have also been recognized to play a role in cocoa flavour (Clapperton et al. 1994), contributing to bitterness and astringency, therefore influencing cocoa organoleptic quality (Stark et al. 2005; Counet et al. 2004). They could also interact with the aroma precursors by their ability to complex with proteins, the precursors of cocoa flavour. A more general role in plant disease resistance is also now well recognized (Stafford 1990).

All flavanols are produced by a common metabolic pathway, and a regulation of flavanols synthesis occurring at the first steps of the metabolic pathway, could impact also on the anthocyanin synthesis. This fact explains the positive correlation found by Cakirer et al. (2002) between anthocyanidin and procyanidin concentrations in most of the samples studied, with lighter seeds containing low levels of flavanol and darker seeds having a higher flavanol concentration. These authors suggested that seed colour could act as a marker for flavanol content in the fresh seeds.

The genetic control of anthocyanin biosynthesis has been extensively studied in several plant species, such as maize, petunia, and Antirrhinum (Coe et al. 1988; Quattrocchio et al. 1993). In cocoa, the genetic control of seed color is poorly understood despite a first work of Wellensiek (1931) suggesting the role of one major dominant gene in the seed pigmentation.

Other morphological traits can be particular of Criollo types when compared to Forastero cultivars, such as fewer seeds per pod, pubescence in young branches, relatively soft fruit shell and light colours in flush leaves and floral structures (Soria 1962). Hybrids between Criollo and Forastero show mixed morphological traits with a wide range in their variation (Braudeau 1970).

The selection of highly productive genotypes, which maintain the aromatic properties of Criollo, is of major interest for Criollo cacao-producing countries. Genomic tools can be useful to provide further information about the genetic determinism of relevant agronomical and quality traits, with potential applications in breeding.

T. cacao is a diploid species, with a small genome of about 390 Mbp (Lanaud et al. 1992). Genetic linkage maps of cacao have been developed. A first highly saturated map of reference comprising 465 codominant markers with an average distance of 1.7 cM between markers is available (Pugh et al. 2004), and recently, a composite linkage map was developed from three crosses between commercial cacao clones (Brown et al. 2008). For the last 10 years, great efforts have been devoted to identify genomic regions implicated in important agronomical traits, such as yield factors (Crouzillat et al. 1996, 2000a; Lanaud et al. 1999; Clément et al. 2003a, b) and

disease resistance (N'Goran et al. 1997; Crouzillat et al. 2000b; Flament et al. 2000; Clément et al. 2003b; Risterucci et al. 2003, Queiroz et al. 2003; Brown et al. 2005, 2007). These studies have been conducted by classical QTL analysis on experimental mapping populations.

However, the lack of large productive progenies from known parental origin is often a major obstacle in perennial plants to carry out classical QTL mapping studies. Association mapping is an alternative approach that has been more recently considered to identify regions of the genome involved in the variation of traits in crop plants. It is based on the nonrandom association of alleles in a population or linkage disequilibrium (LD) (Lewontin 1964). LD determines the configuration of blocks of alleles or haplotypes developed according to the history of the population that will be inherited intact to the next generation. The association mapping strategy may be more precise in natural or cultivated populations, even with wide genetic backgrounds (Kraakman et al. 2004). The increased resolution of association mapping, in comparison with classical QTL studies, is given by the cumulated cycles of recombination at a population level which break associations between a genetic factor determining a quantitative or a qualitative trait and any marker not tightly linked or correlated to it (Jannink and Walsh 2004). A major complication for the use of association mapping strategies in crop plants is due to population structure, which can cause false associations (Thornsberry et al. 2001); this association mapping strategy requires also a tight coverage of the genome with markers.

Admixture mapping, first described by Rife in 1954, can be conducted with a loose genome scan in admixed populations with defined ancestors, similarly to experimental mapping progenies, when not many recombination cycles have taken place, and the frequency of marker alleles contributed by each ancestor is different (Darvasi and Shifman 2005). The admixture mapping approach has proven efficiency for mapping hypertension, a complex trait, in human populations (McKeigue 2005). More recently, this procedure was reported as appropriate to map fruit and seed weight (two important domestication traits) in two "modern Criollo/Trinitario" cacao populations (a germplasm collection and a plantation), after it was demonstrated in these nonstructured populations that large chromosomal fragments (up to 30 cM) were maintained from the ancestors, a situation typical of recent admixture (Marcano et al. 2007).

In this paper, we used an admixture mapping strategy to analyze several cocoa traits, important for breeding, such as yield factors, pigmentation, and other contrasting morphological traits between Criollo and Forastero.

Materials and methods

Plant material

Two hundred and fifty-seven modern Criollo/Trinitario hybrid genotypes were sampled in a plantation located in the south of the Maracaibo Lake, near Mérida in Venezuela. This hybrid population presented a wide range of morphological variations, from the pure Criollo types to Hybrid Trinitario types. All the 257 individuals were evaluated for the quantitative traits and used for the associations analyses. After conducting a structure analysis using the model-based software structure (Pritchard et al. 2000), as described by Marcano et al. (2007), a subset of 197 individuals among the 257, representing a nonstructured group was selected to study marker/qualitative trait associations.

Yield and morphological characterization

A total of seven quantitative traits, and four qualitative traits, were evaluated in pods, and young stems in the selected plants. Some of these traits are yield factors; others are relevant for the distinction of Criollo from Forastero types and for the identification of varieties. Most traits were evaluated following the methodology developed by Engels et al. (1980) for the morphological description of cacao trees, with some adjustments specified in Table 1.

The pigmentation of the different organs of the plant was scored with a scale devised according to the actual variability of the sampled individuals, with ascending values, from the lightest (0 or 1), to the darkest tones (6 or 7). In the case of cotyledon pigmentation, we found a wide variety of tones that were first set on six levels (1 to 6). Levels 1 and 2 corresponded to white and light pink, the typical Criollo bean colours. The intermediate 3, and the gradually darker levels 4, 5 and 6, were purple tones commonly found in the beans of hybrids and Forastero trees. Finally, we considered only two classes: class 0 for trees that included levels 1 and 2 (Criollo) in more than 50% of evaluated seeds and class 1 for trees that showed levels 3 to 6 in more than 50% of sampled seeds (hybrids or Forastero). Our reason for this decision was based on the possible plant genotypes present in the studied cacao plantation (white homozygous Criollo -aa- and pigmented heterozygous -ab-) under the assumption of one major and dominant gene determining cotyledons pigmentation, as suggested by Wellensiek (1931). Samples of seeds were derived from open pollination.

A high frequency of self compatible Criollo plants was detected in the plantation, mixed with hybrids which could be self-compatible or not, such that any pod harvested from a hybrid plant (-ab- for this major gene) will contain at least 50% of pigmented seeds, even if only pollinated by Criollo

Table 1 Yield factors and morphological traits evaluated in a cultivated "Modern" Criollo cacao population

Traits	Sampling and evaluation procedures
Quantitative	
Pod number/tree	Total pods produced per tree and per year Mean of 2 years
Fresh bean weight/pod (g)	Determined only from total undamaged pods Mean of 2 years
Number of beans/pod[a]	Estimated from the number of seeds in 400 grams of fresh beans and the mean fresh bean weight per pod
Dry weight of 100 beans (g)	100 beans without pulp were dried 24 h at 105°C
Bean length, width, thickness (cm)	In 100 beans using vernier
Qualitative	
Pod shell rugosity	In 30 pods; mode value; Scale levels of rugosity: 0–1–2–3–4–5
Pubescence in young stems[a]	All flushes of the tree; mode value; scale: 0–1–2–3–4–5–6–7 (absent = 0; high = 7)
Cotyledon pigmentation[a]	Minimum 200 fresh seeds were cut; scale of possible tones of purple and red: 1–2–3–4–5–6 (white = 1, purple = 6) Pigmentation final score: 0:1 or 2 (typical tones of Criollo) in 50% of sampled seeds 1:4 or 5 or 6 in >50% of sampled seeds.
Pod shell colour	Before maturation; scale of possible colours: 1–3–5–7 Light green = 1; dark red = 7.

Sampling and evaluation procedures
[a] Traits used to distinguish Criollo plants

(aa) pollen. If pods are harvested on Criollo trees, homozygous for this locus (aa), a maximum of 50% of pigmented seeds are expected, given the genetic structure of the original plantation where the large majority of plants correspond to pure Criollo or hybrids ranging from Criollo to Trinitario types.

Data analyses

The mean, standard deviation, and coefficient of variation of the data were determined for each quantitative trait. The Shapiro–Wilk normality test was carried out to evaluate the data distribution of each trait, using the univariate procedure of the SAS program version 9 (SAS 2002).

Population structure analysis

The model-based software structure (Pritchard et al. 2000) was used to infer population structure using a burn-in of 100,000, run length of 1,000,000, and ten independent runs. A model with admixture and correlated allele frequencies was chosen. The tested K values (equivalent to the number of subpopulations) ranged from 1 to 5. SSR genotype data for 24 independent microsatellite loci located in all cocoa chromosomes were used for these analyses.

Identification of marker/trait associations

Ninety-two microsatellite markers were selected for the association study. They were distributed at an average distance of 8.5 cM along the ten linkage groups of cacao, according to the cacao reference map (Pugh et al. 2004). Only two alleles, specific of Criollo and Forastero ancestors, were observed for the large majority of loci. For this reason, all the alleles were coded as "Criollo" vs "Forastero" alleles.

Quantitative traits In order to minimize the risk of false positives, we used Trait Analysis by ASSociation, Evolution and Linkage (TASSEL) software (Remington et al. 2001), available at http://www.maizegenetics.net, which integrates population structure as a cofactor in the association analyses (Yu et al. 2006).

The least square solution to the fixed effects general linear model (GLM) (Searle 1987) implemented in TASSEL 2.1 and taking into account the admixture percentages estimated by STRUCTURE, was used to identify associations. The significance of the association between a marker and a trait was determined by the p values.

Multiple testing experiments can generate false-positive associations. To control the false discovery rate (FDR), we determined the p value threshold using the QVALUE software (Storey 2002). This package takes a list of p values resulting from the simultaneous testing of many hypotheses and estimates their q values. The q value of a test (also called false discovery rate) measures the proportion of false positives incurred when that particular test is called significant. Similar to a p value, the q value can be considered as a measure of statistical significance. The q value was calculated from the p values given by TASSEL for the seven considered quantitative traits and ranked them in ascending order. The corresponding p value threshold was given by QVALUE with a cutoff for null hypothesis rejection (q value) of 0.05 to ensure a 5% FDR.

When more than one marker in the same linkage group were related to the same trait, only those separated by more than 20 cM in the reference map were taken as different marker/trait associations, as previous results showed that linkage disequilibrium decayed between markers at such distance in this population (Marcano et al. 2007).

Qualitative traits: in a nonstructured subset of 197 individuals, selected as previously reported in Marcano et al. (2007), a X^2 test was conducted, using the Freq

procedure of the SAS software. For each marker locus, the number of trees of each phenotypic–genotypic class in a contingency table was determined. The differences between observed and expected frequencies were tested under the null hypothesis of no association. A threshold of $p<0.0005$ was chosen to declare a significant marker trait association, based on a significance value of 5% on each individual test at the genome level and considering 92 (markers) tests. The contingency tables varied from 3×2 to $n\times n$, depending on the scoring scales. The degrees of freedom of the X^2 tests varied accordingly.

In this subset of 197 individuals, three genotypic classes were observed for each marker: aa (homozygous Criollo), ab (heterozygous Criollo/Forastero), and bb (homozygous Forastero). However, the "bb" genotypic class was underrepresented in this subset of 197 individuals, generally varying from three to 12 individuals depending on the considered loci. The data from the "bb" genotypic class was not taken into account in the analysis in order to avoid the identification of false associations due to a biased mean related to the low number of individuals used to evaluate this genotypic class.

Results and discussion

Characterization of quantitative traits data

Table 2 presents mean values, measures of dispersion of the quantitative traits evaluated in the population and the Shapiro–Wilk normality test for each quantitative trait. The traits were generally normally distributed in the sampled individuals and highly variable as expected in a cacao hybrid population.

Table 2 Quantitative traits evaluated

	N	Mean	Range	SD	CV	SW
Pod number/tree	252	72	4–188	39.00	54.11	**
Fresh bean weight/pod (g)	257	66..44	18.83–141.8	20.13	30.31	**
Bean number/pod	257	23.31	8.06–48.39	6.93	29.73	**
Dry weight of 100 beans (g)	257	91.9	18.24–164.4	29.94	27.14	
Bean length (cm)	257	2.28	1.61–2.8	0.22	9.8	**
Bean width (cm)	257	1.38	1.05–1.65	0.11	7.9	
Bean thickness (cm)	257	0.97	0.63–1.22	0.1	10.35	*

Means, measures of dispersion and normality test
N number of trees evaluated, *SD* standard deviation, *CV* coefficient of variation, *SW* Shapiro and Wilk normality test with *$P<0.05$ and **$P<0.01$

Population structure

The analysis with STRUCTURE (Pritchard et al. 2000) showed that the model with two subpopulations was the most likely. The percentage of admixture of each individual obtained for $K=2$ was used in the subsequent association analyses.

Identification of marker/trait associations

Quantitative traits Table 3 shows details of the results of significant associations between markers and yield components and bean traits, along with QTLs identified in previous studies for some of these traits, assuming the position of anchoring markers in the consensus map (Pugh et al. 2004).

Numerous marker/trait associations were detected in our target population, some of which (especially for bean traits) were common to QTLs detected in experimental cacao mapping progenies with different genetic backgrounds. The results suggest stability of genomic regions involved in these traits and prove the general efficiency of the implemented admixture mapping approach to detect marker/trait associations.

The identification of regions that were consistently associated or linked to particular traits in genetically diverse populations and through various mapping procedures cumulates support for real QTLs. In order to define efficient selection methods through molecular markers, efforts to map interesting traits is essential, and the admixture approach seems to be efficient for this objective.

Pod number and fresh bean weight per pod

Pod number/tree/year is an important yield component, commonly accepted as particularly advantageous in Forastero type plants. In the target population studied, it ranged from four to 236 pods/tree/year. Our results confirm the Forastero alleles as beneficial for pod number. In addition, a locus (mTcCIR227) suggesting possible heterosis between Criollo and Forastero alleles was also detected.

This is a first report in the evaluation of fresh bean weight/pod. We found eight significant associations with markers, some of which were common to other bean traits; two of these loci were shared with pod number (LG1 and LG2). According to the associations identified, the favourable allele can originate from Criollo or from Forastero ancestors.

Pod number and fresh bean weight/pod are the basic yield factors in cacao; these traits have not been found correlated (data not shown), which open a possibility for simultaneous improvement of both traits through the manipulation of independent associated regions, as those

Table 3 Microsatellite markers associated to quantitative traits

Trait	LG	Marker loci	cM	F	*Q value*	*p value*	Expl.	Means			QTL identified		
							var.%	aa	ab	bb	Progeny	Position	Ref.
Pod number	1	mTcCIR184	5.0	11.1	0.02	0.0003	7.2	60.56	84.36	104.29			
	2	mTcCIR227	53.7	8.05	0.02	0.0004	6.9	62.34	88.04	77.56	S52	43.0–58.0	1
	8	mTcCIR189	46.5	6.3	0.04	0.002	5.6	62.96	76.12	78.96			
Fresh bean	1	mTcCIR184	5.0	8.0	0.01	0.0004	6.7	61.99	55.75	42.01			
weight/pod	1	mTcCIR275	84.2	5.2	0.05	0.006	4.7	60.8	71.8	70.3			
	2	mTcCIR253	57.6	8.2	0.01	0.0004	8.4	63.68	50.19	57.50			
	5	mTcCIR279	12.8	5.0	0.05	0.007	4.2	63.3	66.8	96.5			
	6	mTcCIR182	3.8	6.5	0.03	0,0017	5.3	61.53	75.27	73.03			
	6	mTcCIR193	31.6	5.9	0.04	0.003	4.7	63.14	68.19	86.95			
	9	mTcCIR30	22.0	5.1	0.05	0.008	8.0	65.7	56.2	45.9			
	10	mTcCIR91	49.3	6.8	0.03	0.001	5.4	61.90	57.57	79.55			
N° beans/pod	5	mTcCIR279	12.8	8.5	0.02	0.0003	6.7	22.47	23.84	40.78			
Dry weight	2	mTcCIR253	57.6	9.3	0.008	0.0002	10.6	100.08	82.33	73.16	S52,IMC78	43.0–58.0	1
100 beans	4	mTcCIR115	8.7	8.0	0.008	0.0005	6.9	93.02	108.03	96.80			
	4	mTcCIR57	58.5	5.8	0.03	0.004	5.3	92.62	104.92	93.08	S52,IMC78	60.0–75.0	1
	9	mTcCIR30	22.0	5.9	0.03	0.0004	10.3	97.68	86.13	65.57	P12	23.0–36.0	3
	9	mTcCIR205	57.7	8.2	0.008	0.0004	7.9	95.66	96.46	67.75	S52, UF676	50.0–60.0	1,2
	10	mTcCIR223	72.6	8.8	0.008	0.0002	8.3	93.40	101.08	110.71			
Bean length	1	mTcCIR184	5.0	6.3	0.01	0.002	5.5	2.39	2.31	2.21			
	1	mTcCIR94	27.0	3.4	0.05	0.04	3.0	2.40	2.33	2.24	DR1	27.0–33.7	1
	1	mTcCIR137	53.1	4.4	0.03	0.01	4.5	2.39	2.45	2.07			
	1	mTcCIR194	94.7	3.8	0.04	0.02	3.3	2.37	2.33	2.44			
	2	mTcCIR253	57.6	16.8	<0.0001	<0.0001	16.3	2.43	2.24	2.12	DR1,IMC78,UF676	29.0–52.0	1
	2	mTcCIR11	89.3	4.8	0.02	0.009	4.4	2.39	2.30	2.24	S52	89.6–100.9	1
	3	mTcCIR82	29.5	6.7	0.008	0.002	6.0	2.36	2.42	2.23	IMC78	40.6	1
	4	mTcCIR115	8.7	6.1	0.01	0.003	5.1	2.37	2.46	2.30	DR1	0.00–18.7	1
	4	mTcCIR32	45.9	7.0	0.008	0.001	6.8	2.39	2.53	2.57	DR1,S52,IMC78	43.1–64.1	1
	6	mTcCIR193	31.6	14.4	<0.0001	<0.0001	11.2	2.41	2.56	2.71	DR1,S52,IMC78	0.0–33.9	1
	6	mTcCIR291	59.7	6.8	0.008	0.001	5.6	2.41	2.54	2.59			
	9	mTcCIR30	22.0	7.1	0.008	0.001	11.3	2.42	2.31	2.14			
	9	mTcCIR205	57.7	5.4	0.02	0.005	5.0	2.38	2.34	2.18	S52	40.1–51.7	1
	10	mTcCIR223	72.6	7.67	0.007	<0.0001	6.8	2.37	2.32	2.49			
Bean width	1	mTcCIR184	5.0	3.4	0.05	0.04	3.1	1.41	1.41	1.35	IMC78	18.7–27.3	1
	1	mTcCIR244	57.7	3.1	0.05	0.04	3.3	1.41	1.44	1.37			
	2	mTcCIR253	57.6	5.7	0.03	0.004	6.4	1.44	1.38	1.34	DR1,IMC78,UF676	44.5–46.2	1
	2	mTcCIR73	95.8	3.4	0.05	0.04	2.9	1.42	1.46	1.47	IMC78,S52	78.2–90.0	1
	3	mTcCIR82	29.5	4.2	0.04	0.02	3.9	1.41	1.42	1.34	IMC78	46.2	1
	4	mTcCIR115	8.7	10.7	0.0006	<0.0001	8.6	1.41	1.47	1.35			
	4	mTcCIR32	45.9	4.9	0.03	0.01	5.0	1.42	1.48	1.43			
	5	mTcCIR279	12.8	3.7	0.04	0.025	3.3	1.42	1.46	1.48			
	6	mTcCIR182	3.8	3.2	0.05	0.04	2.8	1.42	1.44	1.46	IMC78	4.1–13.1	1
	8	mTcCIR258	26.3	3.1	0.05	0.04	2.8	1.41	1.45	1.41	DR1,S52	22.7–45.1	1
	9	mTcCIR157	38.1	5.9	0.02	0.003	5.4	1.42	1.39	1.28			
	10	mTcCIR223	72.6	13.5	<0.0001	<0.0001	11.65	1.41	1.35	1.48			
Bean	1	mTcCIR184	5.0	4.9	0.02	0.004	5.4	1.00	1.01	0.96			
thickness	2	mTcCIR268	38.6	5.9	0.01	0.003	5.2	1.01	0.95	0.93			
	3	mTcCIR82	29.5	5.8	0.01	0.004	5.4	1.00	0.99	0.91			
	4	mTcCIR115	8.7	11.1	0.0003	<0.0001	9.0	1.00	1.07	1.03			
	4	mTcCIR57	58.5	6.1	0.01	0.003	5.4	1.00	1.05	1.00	IMC78,DR1,S52	43.1–64.1	1
	9	mTcCIR30	22.0	3.6	0.05	0.03	6.4	1.01	0.98	0.89	DR1,IMC78	22.0–51.0	1
	9	mTcCIR205	57.7	4.4	0.03	0.01	4.2	1.01	1.05	0.97			
	10	mTcCIR31	34.5	5.9	0.01	0.004	12.6	1.02	0.98	1.16			
	10	mTcCIR223	72.6	16.4	<0.0001	<0.0001	14.0	1.00	1.01	1.09			

LG linkage group. cM: Position in the reference map (UPA402×UF676). *F* value, *q* value and *p* value.
Explained variation (%). Means genotypes aa (Criollo homozygous), ab (heterozygous), bb (Forastero homozygous)
QTLs reported: 1 (Clément et al. 2003a, b) 2 (Lanaud et al. 2003) 3 (Crouzillat et al. 2000a, b)

identified in LG1, LG 5, LG6, LG9, and LG10 for fresh bean weight/ pod.

Bean number per pod

Only one marker was associated to this trait in the cultivated population; in other progenies previously studied, seven QTLs had been identified for ovules/ovary number (Clément et al. 2003b), a highly correlated and more stable trait (López et al. 1988). However, no common genome regions were detected in this study, which could be explained by the different genetic origins used as parents in the QTL progenies.

Bean traits

Many associations were detected for the components of bean weight, in the studied population. Results from traditional QTL studies (Clément et al. 2003b; N'Goran et al. 1997) and association analysis in a population from Hawaii in a cacao germplasm collection (Schnell et al. 2005; Pugh 2005), confirm several common regions located in LG 1, LG 2, LG 4, and LG 9 implicated in bean dimensions and weight. From these regions, mTcCIR253 (LG2), mTcCIR115 (LG4), and mTcCIR223 (LG10) explained the highest percentage of phenotypic variation in our target population. The presence of the Criollo allele was favourable in only 68% of the associations identified, suggesting that the Forastero allele may also sometimes be favourable for these bean traits.

Qualitative traits

Results of associations of markers with qualitative traits evaluated in pods, young branches and beans are presented in Table 4.

Several qualitative traits are particularly important to distinguish Criollo type plants, as pubescence in young stems, also implicated in insect tolerance, pigmentation of different structures, particularly cotyledons and flowers parts, bean shape, pod shape, pod shell hardness. Some morphological features of pods and pigmentation of the different structures are routinely used to differentiate varieties.

Pigmentation traits

Wellensiek (1931) first stated that the cotyledon pigmentation was determined by one gene with two alleles, the dominant one conferring pigmentation. According to his study, the intensity of the purple colour can be also determined by modifying genes. The range of tones a hybrid cacao tree presents in the beans is an interesting feature that may include not only purple, but red colours. Our sampling and evaluation procedures for this feature (see Table 1) did not finally consider degrees or tones of colours in the beans; therefore, no modifying factors could be detected. However, in the Venezuelan cultivated population, three regions were associated to the pigmentation of different organs (Table 4).

Table 4 Microsatellite markers associated to qualitative morphological traits in a cultivated "Modern" Criollo cacao population

Trait	LG	Marker loci	Position cM	*N*	X^2	*P*
Pubescence in young stems	9	mTcCIR205	59.6	114	10.99	0.0009
Pigmentation						
Bean	1	mTcCIR270	36.9	94	18.10	<0.0001
	4	mTcCIR115	8.5	110	40.25	<0.0001
	10	mTcCIR223	72.4	102	38.07	<0.0001
Pod	4	mTcCIR213	24.1	192	31.77	0.00010
Pod shell rugosity	4	mTcCIR57	59.8	170	32.34	<0.0001

LG linkage group, *Position* position of the marker in the reference map (Pugh et al. 2004), *N* sample size, X^2 Chi^2 value, *P* value.

Engels (1983) observed high correlations in the pigmentation between the different structures of the cocoa plant, including cotyledons. These correlations are in agreement with the co-localization of markers related to the pigmentation of different structures in the cacao plant, especially in a small region of LG4, in which marker mTcCIR115 is located. This sector includes the major locus identified by Crouzillat et al. (1996) implicated in bean colour (anthocyanins) in the Catongo × Pound 12 back cross progeny.

The pattern of biosynthesis of anthocyanins, involving several responsible enzymes, and the variations observed in the population, suggest a more complex genetic system than the one defined by a single gene. This could explain the contribution of several loci to the variation of pigmentation of the cacao plant structures. The association of pigmentation on beans and pods to several regions of the genome, may also explain the differential expression and/or accumulation of the pigmentation that can be observed among these structures in the cocoa plant (Bartley 2005). This is consistent with findings in the anthocyanin pigmentation in other species, such as rice, in which several dispersed genes (structural or regulatory) control this single trait and its variation, according to tissue-specific accumulation (Reddy 1996).

In this paper, we considered an admixture mapping strategy to identify associations between markers and a panel of quantitative and qualitative useful cocoa traits segregating in a modern cacao-cultivated population.

The resolution of association approaches depends on the structure of linkage disequilibrium (LD) within the tested population. The modern Criollo population studied in this work is typically an admixed population which descends from a recent hybridization of two ancestral and contrasting founders: an ancient Criollo and a lower Amazon Forastero, probably admixed during the 18th century. A few generations of recombination has led to long chromosome blocks, as large as 30 cM, conserved from the ancestors (Marcano et al. 2007). Compared to association mapping approaches involving populations with broader genetic origins and higher recombination rates, this situation will facilitate a whole genome scan to detect associations between markers and phenotypic traits, with a fewer markers required to detect associations. For a tree crop species such as *T. cacao*, large controlled populations, adapted for QTL studies, are limited. Admixture mapping offers an alternative strategy to detect genome regions involved in trait variations by using larger admixed populations that already exist and are producing.

In the last decade, most breeding programs lead to the distribution of hybrid clones generally belonging to a limited number of contrasting genetic origins, resulting in admixed populations suitable for such analyses. Using this strategy, the precision of QTL location is improved compared to QTL detected from controlled crosses; a twofold recombination rate was observed in the modern Criollo population studied in this work, compared to the reference genetic map based on a F_1 test cross-progeny (unpublished data). However, the recent admixture of the two ancestors of this modern Criollo population will also limit the power of detection of these associations, preventing precise mapping of genes which could be directly involved in these trait variations, except if the population size is increased. This admixture mapping strategy could be considered as a first step in identifying regions co-localized with QTL, where to assess marker/trait associations targeted more specifically on candidate genes involved in traits variations. Additionally, association mapping conducted at the sequence level in populations resulting from a longer evolution time, such as wild cocoa populations, could confirm the implication of the significant loci in important agronomic traits like disease resistance for which wild populations represent a large resource of different resistance genes.

Numerous marker/traits associations identified in this study, confirm the efficiency of this method to map genome regions involved in cocoa trait variations, and allow the identification of potentially useful markers to conduct marker-assisted selection. Complementary studies are also being conducted with a similar strategy on the genetic determination of cocoa quality traits in order to select new Criollo varieties with higher productivity and good aromatic cocoa flavour.

Acknowledgements We thank the financial support received from FONACIT (*Fondo Nacional para la Ciencia y la Tecnología)*, Venezuela, the *FIRC* (*Fonds Interprofessionnel de la Recherche sur le cacao), the Ministère des Affaires Etrangers*, France and VALRHONA/SOCAOVEN for providing the vegetal material and infrastructure needed for this work. We also thank Dr. Surendra SURUJDEO-MAHARAJ for improving the English writing of this manuscript.

References

Bartley BGD (2005) The genetic diversity of cacao and its utilization. CAB, UK, p 341

Borchers AT, Keen CL, Hannum SM, Gershwin ME (2000) Cocoa and chocolate: composition, bioavailability and health implications. J Med Food 3(2):77–105

Braudeau J (1970) El cacao. Técnicas agrícolas y producciones tropicales. Colección Agricultura Tropical. Ed. Blume, Barcelona, España

Brown JS, Schnell RJ, Motamayor JC, Lopes U, Kuhn DN, Borrone JW (2005) Resistance gene mapping for witches' broom disease in *Theobroma cacao* L. in an F2 population using SSR markers and candidate genes. J Amer Soc Hort Sci 130:366–373

Brown JS, Phillips-Mora W, Power EJ, Krol CA, Cervantes Martinez C, Motamayor JC, Schnell RJ II (2007) Mapping QTL for resistance to frosty pod and black pod diseases, and for horticultural traits in *Theobroma cacao* L. Crop Sci 47:1851–1858

Brown JS, Sautter RT, Tondo CT, Borrone JW, Kuhn DN, Motamayor JC, Schnell RJ II (2008) A composite linkage map from the combination of three crosses made from commercial clones of cacao, *T. cacao* L. Tropical Plant Biology 1(2):120–130

Cakirer MS, Ziegler GR, Guiltinan MJ, Jones AD (2002) Fresh bean colour as an indicator of chocolate flavour potential. In: Le Quere JL, Etievant PX (eds) Flavour research at the dawn of the twenty-first century. Proceedings of the 10th Weurman Flavour Research Symposium. Intercept, London, pp 540–543

Cheesman EE (1944) Notes on the nomenclature, classification and possible relationship of cacao populations. Trop Agric 21:144–159

Clapperton J, Yow S, Chan J, Lim D, Lockwood R, Romanczyk L, Hammerstone J (1994) The contribution of genotype to cocoa (*Theobroma cacao* L.) flavor. Trop Agric 71(4):303–308

Clément D, Risterucci AM, Motamayor JC, N'Goran J, Lanaud C (2003a) Mapping QTL for yield components, vigour and resistance to *Phytophthora palmivora* in *Theobroma cacao* L. Genome 46:204–212

Clément D, Risterucci AM, Motamayor JC, N'Goran J, Lanaud C (2003b) Mapping quantitative trait loci for bean traits and ovule number in *Theobroma cacao* L. Genome 46:103–111

Coe EH, NeuVer MG, Hoisington DA (1988) The genetics of corn. In: Sprague GF, Dudley JW (eds) Corn and corn improvement. 3rd edn. ASA, CSSA, SSSA, Madison, pp 81–258

Counet C, Ouwerx C, Rosoux D, Collin S (2004) Relationship between procyanidin and flavor contents of cocoa liquors from different origins. J Agric Food Chem 52:6243–6249

Crouzillat D, Lercetau E, Pétiard V, Morera J, Rodríguez H, Walker D, Phillips W, Ronning C, Schnell R, Osei J, Fritz P (1996) *Theobroma cacao* L.: a genetic linkage map and quantitative trait loci analysis. Theor Appl Genet 93:205–214

Crouzillat D, Ménard B, Mora A, Phillips W, Fritz PJ, Pétiard V (2000a) Quantitative trait loci analysis in *Theobroma cacao* L. using molecular markers. Euphytica 114:13–23

Crouzillat D, Phillips W, Fritz PJ, Pétiard V (2000b) Quantitative trait loci analysis in *Theobroma cacao* L. using molecular markers. Inheritance of polygenic resistance to *Phytophthora palmivora* in two related cacao populations. Euphytica 114:25–36

Darvasi A, Shifman S (2005) The beauty of admixture. Nat Genet 37 (2):118–120
Engels JMM (1983) A systematic description of cacao clones. III. Relationships between clones, between characteristics and some consequences for the cacao breeding. Euphytica 32:719–733
Engels JMM, Bartley BGD, Enríquez GA (1980) Cacao descriptors, their states and modus operandi. Turrialba 30(2):211–218
Eskes A, Lanaud C (1997) Cocoa. In: Charrier A (ed) Tropical plant breeding. Montpellier, France, pp 78–105
FAO. Anuario Estadístico 2005–2006. http://www.fao.org.statistics/yearbook/
Flament MH, Kébé I, Clément D, Pieretti I, Risterucci AM, N'Goran JAK, Cilas C, Despéux D, Lanaud C (2000) Genetic mapping of resistance factors to *Phytophthora palmivora* in cocoa. Genome 44:79–85
Forsyth WCG, Quesnel VC (1957) Cacao polyphenolic substances 4. The anthocyanin pigments. Biochemistry Journal 65:177–179
Gu L, House SE, Wu X, Ou B, Prior RL (2006) Procyanidin and catechin contents and antioxidant capacity of cocoa and chocolate products. J Agric Food Chem 54:4057–4061
Jannink JL, Walsh B (2004) Association mapping in plant populations. In: Kang MS (ed) Quantitative genetics, genomics and plant breeding. CAB International, pp 59–68
Jeanjean N (1995) Influence du génotype, de la fermentation et de la torréfaction sur le développement de l'arôme cacao. Rôle des précurseurs d'arôme. Thèse de Doctorat, Université de Montpellier II, Montpellier, France, 200 p
Kraakman ATW, Niks RE, Van der Berg PMMM, Stam P, Van Eeuwijk FA (2004) Linkage disequilibrium and mapping of yield and yield stability in modern spring barley cultivars. Genetics 168:435–446
Lanaud C, Hamon PC, Duperray C (1992) Estimation of nuclear DNA content of *Theobroma cacao* L. by flow cytometry. Café, Cacao, Thé 36:3–8
Lanaud C, Kébé I, Risterucci AM, Clément D, N'Goran JKA, Grivet L, Tahi M, Cilas C, Pieretti I, Eskes A, Despréaux D (1999) Mapping quantitative trait loci (QTL) for resistance to *Phytophthora palmivora* in *T. cacao* L. 12th International Cocoa Research Conference, November 17–23, Salvador, Bahía, Brazil, pp 99–105
Lanaud C, Boult E, Clapperton J, N'Goran JKA, Cros E, Chapelin M, Clément D, Petithugenin P (2003) Identification of QTLs related to fat content, seed size an sensorial traits in *Theobroma cacao* L. 14th International Cocoa Conference, Accra, Ghana
Lewontin R (1964) The interaction of selection and linkage. I. General considerations; heterotic models. Genetics 49:49–67
López O, Enríquez GA, Soria J (1988) Herencia del número de óvulos por ovario en *Theobroma cacao* L. Turrialba 38(3):163–167
Marcano M, Pugh T, Cros E, Morales S, Portillo Páez E, Courtois B, Glaszmann J, Engels M, Phillips W, Astorga C, Risterucci AM, Fouet O, González V, Rosenberg K, Vallat I, Dagert M, Lanaud C (2007) Adding value to cocoa (*Theobroma cacao* L) germplasm information with domestication history and admixture mapping. Theor Appl Gen 114:877–884
McKeigue P (2005) Prospects for admixture mapping of complex traits. Review article. Am J Hum Genet 76:1–7
Motamayor JC, Risterucci AM, López 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, Risterucci AM, Heat M, Lanaud C (2003) Cacao domestication II: progenitor germplasm of the *Trinitario cacao* cultivar. Heredity 91:322–330
N'Goran JAK, Risterucci AM, Clément D, Sounigo O, Lorieux M, Lanaud C (1997) Identification of quantitative trait loci (QTL) in *Theobroma cacao* L. Agron Afr 9:55–63
Paradis L (1979) Le cacao précolombien: monnaie d'échange et breuvage des dieux. J Agric Tradit Bot Appl 26:3–4
Pittier H (1935) Degeneration of cacao through natural hybridization. J Heredity 26(10):385–390
Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959
Pugh T (2005) Etude du déséquilibre de liaison chez le cacaoyer appartenant aux groupes Criollo/Trinitario. Application au marquage génétique d'intérêt pour la sélection. Thèse Doctorat, Ecole National Supérieur d'Agonomie, Montpellier, France, 107 p
Pugh T, Fouet O, Risterucci AM, Brottier P, Deletrez C, Courtois B, Clément D, Lamande P, N'Goran JAK, Lanaud C (2004) A new codominant markers based cocoa linkage map: Development and integration of 201 new microsatellites markers. Theor Appl Genet 108:1151–1161
Quattrocchio F, Wing JF, Leppen H, Mol J, Koes RE (1993) Regulatory genes controlling anthocyanin pigmentation are functionally conserved among plant species and have distinct sets of target genes. Plant Cell 5:1497–1512
Queiroz VT, Guimaraes CT, Anhert D, Schuster I, Daher RT, Pereira MG, Miranda VRM, Loguercio LL, Barros EG, Moreira MA (2003) Identification of a major QTL in cocoa (*Theobroma cacao* L) associated with resistance to witches' broom disease. Plant Breed 122:268–272
Reddy AR (1996) IRRI International Rice research Institute. Rice genetics III. Proc. of the 3rd International Rice Genetics Symposium. Philippines, 16–20 Oct
Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, Kresovich S, Goodman MM, Buckler ES (2001) Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc Natl Acad Sci 20:11479–11484
Rife DC (1954) Populations of hybrid origin as source material for the detection of linkage. Am J Hum Gen 6:26–33
Risterucci AM, Paulin D, Ducamp M, N'Goran JAK, Lanaud C (2003) Identification of QTLs related to cocoa resistance to three species of *Phytophthora*. Theor Appl Genet 108:168–174
SAS Institute Inc. (2002) Version 9. SAS Institute Inc., Cary, North Caroline, USA
Schnell RJ, Olano CT, Brown JS, Meerow AW, Cervantes-Martínez C (2005) Retrospective determination of the parental population of superior cacao (*Theobroma cacao* L) seedlings and association of microsatellite alleles with productivity. J Amer Soc Hort Sci 130 (2):181–190
Searle SR (1987) Linear models for unbalanced data. Wiley, New York
Soria J (1962) "Porcelana" cacao of Venezuela. Cacao 7(4):7–9 Costa Rica
Stafford HA (1990) Flavonoid metabolism. CRC, Boca Raton, FL, p 317
Stark T, Bareuther S, Hofmann T (2005) Sensory-guided decomposition of roasted cocoa nibs (*Theobroma cacao* L) and structure determination of taste-active polyphenols. J Agric Food Chem 53:5407–5418
Storey JD (2002) A direct approach to false discovery rates. J R Stat Soc Ser B 64:479–498
Thornsberry J, Goodman M, Doebley J, Kresovich S, Nielsen D, Buckler E IV (2001) *Dwarf8* polymorphisms associate with variation in flowering time. Nat Genet 28:286–289
Wellensiek SJ (1931) The genetics of cotyledon colour of cocoa as a basis for quality selection. Translated by H. Toxopeus from Archief voor de Koffiecultuur in Nederlandsch-Indië (Buitenzorg, Java)
Wollgast J, Anklam E (2000) Polyphenols in chocolate: is there a contribution to human health? Food Res Int 33(6):449–454
Yu J, Pressoir G, Briggs W, Vroh Bi I, Yamasaki M, Doebley JF, McMullen MD, Gaut B, Nielsen DM, Holland JB, Kresovich S, Buckler E (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38:203–208