P-ISSN: 2305-6622; E-ISSN: 2306-3599
International Journal of Agriculture and Biosciences
www.ijagbio.com; editor@ijagbio.com

Research Article

Association Mapping for *Phytophthora* Pod Rot Resistance in a Cacao (*Theobroma cacao* L.) Population Grown in Farmers' Field

Efombagn MIB[1], O Sounigo[2], B Courtois[3], O Fouet[3], M Jeanneau[4], A Lemainque[4], S Pavek[4] and C Lanaud[3]

[1]IRAD PO Box 2123, Yaoundé, Cameroun; [2]CIRAD, UR Bioagresseurs, BP 2572 Elig-Essono, Yaoundé, Cameroun; [3]CIRAD, UMR AGAP, TA A108/3, Avenue Agropolis, 34398 Montpellier, France; CNG, 2 rue Gaston Cremieux, CP 5721, 91057 Evry, France
***Corresponding author:** efombagn@yahoo.fr

Article History: Received: March 23, 2016 Revised: July 24, 2016 Accepted: August 16, 2016

ABSTRACT

Phytophthora pod rot (PPR) caused by the specie *Phytophthora megakarya* is an important disease of cacao tree. Association mapping identified markers linked to PPR resistance in a cacao population of 260 trees planted under high disease pressure in a single plantation in a farmer's field. These cacao trees were derived from both selfing and full-sib progenies. The resistance traits were assessed through field observations of the natural pod attacks of the disease on the trunk (PRTnk) or the canopy (PRCpy) of cacao trees (expressed as the percentage of PPR-infected pods), and the development of the symptoms on the pods two (PT2d) or five days (PT5d) after artificial inoculations. A total of 108 SSRs markers loci covering the 10 chromosomes of the cacao genome were used in the analysis. The percentages of admixture of each genotype, estimated using 17 neutral SSRs markers, were used as co-factor in the analyses, decreasing the proportion of false-positive due to population structure. General and mixed linear models were used to analyze phenotypic data collected over 3 years. For field PPR incidence and artificial pod inoculation tests, a total of 36 and 18 individual marker-trait associations were detected, respectively. The positive and significant correlations found between PT5d and field measurements (PRTnk and PRCpy) explained the fact that both traits were co-localized with PPR scores in 9 marker-trait associations. The results of this study highlight the interests of association mapping to decipher the PPR genetic control and to guide the breeding strategies to produce a sustainable cacao resistance to *Phytophthora* species.

Key words: *Theobroma cacao* L., *Phytopthora* sp, Association mapping, Population structure, SSRs

INTRODUCTION

Theobroma cacao L. is a perennial tree native of South America and grown mainly by smallholders in West Africa and other regions situated in the tropics. It is an economically important cash crop used for chocolate and various manufactured products. Africa contributes about 70% of the total cocoa produced in the world and is facing several agronomic constraints including *Phytophthora* pod rot (PPR): a fungus disease that can cause significant pod losses. Under favorable climatic and environmental conditions, it may lead to yield reduction up to 90% without fungicide treatments in a country such as Cameroon where the species *Phytophthora megakarya* is prevalent (Despréaux *et al*., 1988).

The development of DNA markers and statistical methods to detect markers linked to quantitative trait loci (QTLs) in mapping populations have provided insights into the genetics of PPR and other cacao diseases such as witches' broom. Lanaud *et al*. (1995) published the first linkage map of cacao with ten linkage groups. This map was used for mapping QTLs for resistance to PPR (Lanaud *et al*., 1999a). QTLs that accounted for moderate (17%) to relatively high levels (48%) of resistance to PPR were also detected by Crouzillat *et al*. (2000a, 2000b). However, Flament *et al*. (2001) found no QTL in common across different measurements for PPR resistance, and found results from artificial inoculation data to be poorly related to results based on field resistance.

One problem with QTL analyses is the requirement of large breeding populations (F2, backcrosses or recombinant inbred populations). These populations requires several years between their creation and the obtaining of reliable field data, because of its long and slow growing reproductive cycle, and are impediments due to long juvenile phase (Gutierrez *et al*., 2012).

Cite This Article as: Efombagn MIB, O Sounigo, B Courtois, O Fouet, M Jeanneau, A Lemainque, S Pavek and C Lanaud, 2016. Association mapping for *Phytophthora* pod rot resistance in a cacao (*Theobroma cacao* L.) population grown in farmers' field. Inter J Agri Biosci, 5(6): 329-336. www.ijagbio.com (©2016 IJAB. All rights reserved)

An alternative strategy to QTL study is association mapping, based on linkage disequilibrium (LD), which means non random association of alleles in a population, which can be applied to germplasm collection and commercial cacao plots (Lewontin, 1964). LD enables application of association mapping when no candidate gene is available for the traits under consideration. 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. Large LD blocks are common in cacao breeding populations (Marcano *et al*., 2007). The structuring of a cacao population can generate a pattern of LD that typically associates regions distributed in the whole genome. Cacao breeding which focused only on a few parental genotypes has contributed to accentuate the LD in the specie. Recombination events between previous cacao populations sometimes occurred. Thus, LD was reduced between unlinked genome regions but was maintained within segments hardly affected by recombination. A few number of generations from the ancestor lines of this perennial crop has led to a linkage disequilibrium (LD) as large as 20 cM observed presently in some cacao hybrid populations as Trinitario and the Ecuadorian Nacional hybrid pool (Loor, 2007).

In the first relevant study done by Pugh *et al*. (2004), significant associations were detected between markers and yield components, bean traits, along with QTLs identified in previous studies for some of these traits, assuming the position of anchoring markers in the consensus.

Association mapping approach has been developed in cacao to study the genetic control of seed and fruit traits in a germplasm collection and in a cultivated Cacao population from Venezuela (Marcano *et al*., 2007; Marcano *et al*., 2009).

Using a high density genome-wide scan with SSR loci, it appears possible to identify in cacao, significant linkage disequilibrium and test statistical association between phenotypes and markers, without specific populations derived from controlled crosses and segregation, due to the fact that cacao populations issued from crosses between different genotypes are relatively recent (Marcano *et al*., 2007; Marcano *et al*., 2009). African cacao populations like the one studied in this paper is expected to be similar to that introduced into the continent 200 years ago (Bartley, 2005).

The objective of the present study was to determine the cacao genome regions involved in the control mechanisms of PPR infection. Different testing methods were used to increase the probability of detecting markers – traits associations. These methods include field incidence and artificial inoculations of pods using an association mapping approach carried out in a population of cacao trees composed of a mixture of several selfed and full-sib progenies, released as commercial varieties in Cameroon.

MATERIALS AND METHODS

Plant material and growing conditions

The subset of cacao trees used for the study were selected after a counting of the pods (both ripe and unripe) of3,000 cacao trees planted in a 20-years old commercial cacao plot. The cacao plot is located in the administrative subdivision of Mbankomo, near the city of Yaoundé, in the centre region of Cameroon. The pods were counted at the onset of the main harvest season (end of August), 260 cacao trees with high number of pods were chosen for the study, in order to ensure a reliable counting of the rotten pods.

The rainfall regime in the area of Mbankomo is characterized by two wet and two dry seasons unequally distributed along the year: the longest wet season between August and November with high rainfall (favorable to the PPR infection and its spread) immediately followed by the longest dry season between December and March. The cacao population under study was made of so-called "F1-hybrid" cultivars resulting from open pollinated progeny harvested as pods on parental genotypes planted in seed gardens. These selected trees thus represent a mixture of several selfed and full-sib progenies (Sounigo *et al.*, data no published). The plantation which is shaded by native forest trees did not receive any chemical treatment during the period of the study to allow the expression of the PPR.

Methods

Disease Phenotyping

Data were taken on incidence of PPR for tree trunk (PRTnk), tree canopy (PRCpy), whole tree (PRtree). Inoculation assessment was done at 2days (PT2d) and 5days (PT5d).

- ***Incidence of PPR in the field***

Each of the 260 cacao trees under study was harvested between April 2007 to March 2010, on a weekly basis and both healthy and rotten pods were counted at harvest. Data were collected on the tree truck and the tree canopy of the cacao trees and percentage (%) spoilage were calculated using the formula:

%PPR = number of rotten pods / (number of rotten pods + number of healthy pods) * 100

- ***Symptom scoring after artificial inoculation***

Physiologically mature pods harvested a day earlier were artificially inoculated in the laboratory using the method described by Iwaro *et al*. (2006). This method involves spraying a suspension of zoospores ($3x10^5$ zoospores/ml) from a moderately aggressive isolate of *Phytophthora megakarya* on the pods placed on trays. The trays were kept in dark corner at room temperature for 5 days.

Symptom assessment was made 2 days and 5 days after inoculation, based on a rating perfected by Iwaro *et al*. (2006), taking both the frequency and the surface of the lesions caused by *Phytophthora megakarya* into account.

About 4-20 pods per tree were artificially inoculated with the suspension of zoospores.

Genotyping

The set of selected markers were closed to PPR resistance QTLs identified by a meta QTL analysis and distributed along the ten chromosomes of the cacao genome in a consensus map (Lanaud *et al*., 2009a). DNA was extracted from fresh leaves previously harvested on the 260 selected cacao trees. About 125 SSRs markers

described by Lanaud *et al*. (1999b) and Fouet *et al*. (2011) were used to genotype the corresponding samples. The position of these markers along the cacao linkage groups was presented in a reference map established recently by Allegre *et al*. (2012). After the extraction, DNA fragments were amplified and genotyped in the fluorescent PCR products and analyzed on MegaBACE™ 1000 Sequencer (Amersham Biosciences) as described in Fouet *et al*., 2011). The genotyping was carried out by the National Genotyping Center (CNG) in France following the process described by Fouet *et al*. (2011).

Population structure (Q) and Kinship coefficients (K)

To determine the population structure, the number of sub-populations, was analysed using the software STRUCTURE ver.2.3.3 (Falush *et al*., 2003, 2007; Pritchard *et al*., 2000). To estimate the number of sub-populations among the cacao trees, about 20 neutral markers (not related to QTLs of PPR resistance) were selected among the 125 aforementioned loci. Analyses were run using a burn-in period of 100,000 and Monte Carlo Markov Chain reactions of 200,000. The admixture and correlated option were used for the ancestry and allele frequency models. The criteria used to define the number of subgroups within the studied population, were the position of break point in the L(k) curve and a peak in the Δk distribution. Two (k=2) to twelve (k=12) sub-populations were used in running STRUCTURE. Individuals were assigned to a sub-population only when a cluster membership probability was higher than 60%. The accessions with a coefficient membership lower than 60% (identified in Fig. 1 by the acronym AD = admixed) were not assigned to any of these six sub-populations.

A genetic-distance approach was also used to confirm the number of sub-populations revealed by STRUCTURE. A neighbor-joining tree was built to determine the aggregation of the accessions into clusters using Darwin software version5.0.158 (Perrier and Jacquemoud-Collet, 2006). The kinship matrix (**K**) was also calculated using the MICROSATELLITE ANALYZER (MSA) software (Dieringer *et al*. 2003). MSA is a simple input format with the capacity to analyse large microsatellite data sets. In MSA, the kinship coefficients (DKf) are calculated according to the approach developed by Cavalli-Sforza and Bodmer (1971).

Linkage disequilibrium

LD was estimated as squared allele frequency correlation estimates (r) between all pairs of SSRs. For each marker pair, the extent of LD was determined according to Zhao *et al*. (2005). The 108 SSRs markers (selected for association mapping study) with known chromosomal position were used to estimate LD and to measure the significance of r^2 at P values <0.01 for each pair of loci within the same or on different chromosomes. LD decay over genetic distance was investigated by plotting pair-wise r^2 values against genetic distance (cM) between the markers, and displayed graphically using the software GGT 2.0 (Van Berloo, 2008).

Association mapping

The tests of associations between molecular polymorphisms and phenotypes were computed using the software package TASSEL. The simple General Linear Model (GLM) model using the percentage of admixture of each accession (**Q** matrix) and the Mixed Linear Model (MLM), used with both the percentages of admixture of each accession (Q matrix) and kinship coefficient (K matrix) as cofactors to take population structure were used in testing the association. The mixed-model approach of our study follows that of Yu *et al*. (2006). False Discovery Rate (FDR) of q values corresponding to p values obtained in MLM analyses were used to determine the significant associations using the software QVALUE (Storey 2002).

RESULTS

Population structure

Association mapping considers population structure in order to avoid false positive associations. Therefore, the 260 cacao trees were genotyped at 17 SSRs neutral loci mapped on 8 different chromosomes for assessing population structure. The results are as seen in the NJ (Neighbor Joining) tree (Fig. 1). The NJ also clustered the data into six branches, showing the congruence of the results obtained with the two methods at k=6. The color-coded branches supported six sub-populations. Each sub-population is identified by an alpha-numeric number ranging from 1 to 6. The accessions resulted from open-pollinated seeds produced in seed gardens where there was natural mating.

Phenotypic data

Except for the rating of the symptoms on pods five days after inoculation (PT5d), all traits were normally distributed (P>0.01). The highest range of variation among accessions was obtained with the percentage of PPR-infected pods recorded on the whole cacao tree (PRTree) and the smallest variation for PT5d (Table 1).

The result showed significant differences between sub-populations for all traits except PT5d (Table 2). Among traits showing though, PT2d displayed the weakest separation with the smallest percentage of variance explained by the structure.

Positive and significant correlation was observed between PT2d and PT5d (Table 3). Significant positive correlation exists between PT5d and PPR incidence, with coefficient values ranging from 0.16 (between PT5d and PRCpy) to 0.18 (between PT5d and PRTnk). There was no correlation between PT2d and the other traits. All traits related to PPR incidence in the field showed a positive and significant correlation among them, with the highest correlation coefficient recorded between PRCpy and PRTree (0.97; p<0.001).

Linkage disequilibrium

Extent of genome wide LD was evaluated using pairwise comparisons among the 108 SSRs markers. A total of 5,356 LD patterns were identified. No LD ($r^2<$ 0.01) was detected in 2,887 marker pair comparisons. 46% of marker pairs with $r^2\geq$ 0.01, the LD was expressed by r^2 averaged 0.0445 ranging from 0.01 to 0.99. Among these estimates, 121 (5.3%) showed LD scores as r^2higher than 0.20. The r^2 value corresponding to the 95th percentile of distribution was 0.01. Up to 9.6% of the LD parameter r^2 were significant ($P<0.01$) among all the

pairwise comparisons with $r^2 \geq 0.1$, indicative of a validation for further analysis of association mapping. The r^2 values were plotted against the interval genetic distance (Fig. 2). The LD decayed with the increase of genetic distance. The markers that were closely linked (at a genetic distance of 0 cM) possessed the highest LD while the r^2 decreased for the pairs of loci at a genetic distance higher than 20 cM (Fig. 2).

Significant marker-trait associations for pod tests (PT2d and PT5d)

A total of 36 significant markers/trait associations for PPR resistance distributed on 9 chromosomes (C) was detected for PT2d and PT5din this study (Table 4). Among the most represented chromosomes, nine markers were detected on C1, six on C8, five on C6 and four on C4 and C9. None of these markers were found on C7.

PT2d

16 markers in GLM (Generalized Linear Model) and 13 markers in MLM (Mixed Linear Model) were detected as significant. Except C7 and C10, at least one significant marker was detected in the 10 cacao genome studied. In C1 and C6 where the higher number of markers were found, the length of the region where these markers are positioned varies between 18.2 and 27.2 cM for C1 (97.8 cM) and between 7.4 and 38.5 cM for C6 (62.8 cM). In the other chromosomes, the distance between two significant markers was relatively narrow. The marker *mTcCIR024* (located on C9) was the most significant marker (p=6.89-E05) detected in all the pod tests.

PT5d

The proportion of significance was higher compared to PT2d, with 22 markers in GLM and 15 in MLM. About two markers (*mTcCIR301* and *mTcCIR305*) were common in both PT2d and PT5d. In C1, another region made of four other significant markers was detected in PT5d trait, but with a length (38.9 – 90.3 cM) higher than the one previously found in PT2d with the same number of markers. The ratio of markers significant in both GLM and MLM (MLM/GLM) was lower in PT5d (15/22) compared to PT2d (13/16).

Significant marker-trait associations in the field incidence of PPR (PRCpy, PRTnk and PTree)

The amount of significant markers found after field observations of PPR incidence was almost half of the total detected with pod tests. Among these 17 significant markers detected after field tests, nine were similar to PT2d and PT5d (Table 5). These significant markers were distributed on 7 chromosomes out of 10, with ¼ positioned on C4.

PRCpy

The 11 significant markers detected in GLM were involved in the PPR resistance observed in the canopy of the cacao tree. Six of these markers were located in the C4 region ranging from 8.3to 64.2cM on a total length of 75.5 cM. The markers *mTcCIR431, mTcCIR310*, *mTcCIR444*, *mTcCIR294 and mTcCIR393* were positioned on *C3, CH8, CH8, and C9* respectively. About 3 out of these 11 markers were not significant in MLM.

Table 1: Statistics for PPR traits and percentage of variation of these traits explained by population structure (*k=6)* through multiple linear regression

Trait	Mean	Min	Max	SD	CV(%)	%Var
PT2d	2.02	1.00	5.00	0.63	31.4	5.77***
PT5d	1.03	2.00	8.00	1.30	24.4	1.69ns
PRCpy	49.80	13.03	93.95	15.11	30.4	6.86***
PRTnk	71.44	13.81	98.89	19.02	26.6	4.04*
PRTree	53.61	13.16	92.28	14.83	27.7	7.12***

PT2d: Symptoms scored two days after pod inoculation; PT5d: Symptoms scored two days after pod inoculation; PRCpy: percentage of PPR-infected pods attached to the branches; PRTnk: percentage of PPR-infected pods attached to the trunk; PRTree: percentage of PPR-infected pods attached to the whole cacao tree (branches & trunk); *Significant at 5% level; *** significant at 0.5% level

Table 2: Mean comparison between subpopulations for the accessions assigned to a population (>60% membership in the subpopulation)

Sub Pop	PT2d	PT5d	PRCpy	PRTnk	PRTree
1	1.805 b	0.92 a	45.17 b	58.0 b	51.58b
2	1.96 b	0.99 a	46.18 b	73.26 ab	59.72ab
3	1.89 b	1.17 a	58.44 ab	66.68 ab	62.56ab
4	2.06 b	1.05 a	48.57 b	65.74 ab	57.15ab
5	1.67 b	1.08 a	64.94 a	83.04 a	74.17a
6	2.63 a	1.13 a	51.46 b	66.92 ab	59.19ab
N	175	194	201	200	200

PT2d: Symptoms scored two days after pod inoculation; PT5d: Symptoms scored two days after pod inoculation; PRCpy: percentage of PPR-infected pods attached to the branches; PRTnk: percentage of PPR-infected pods attached to the trunk; PRTree: percentage of PPR-infected pods attached to the whole cacao tree (branches & trunk); *Significant at 5% level; *** significant at 0.5% level

Table 3: Correlation coefficients for all traits scores obtained during field evaluation of PPR incidence and artificial pod inoculation tests

	PT2d	PT5d	PRCpy	PRtnk
PT5d	0.41***			
PRCpy	0.08ns	0.16*		
PRtnk	0.08ns	0.18*	059***	
Prtree	0.07ns	0.17*	0.97***	0.69***

Significant at *p<0.05 and ***p<0.001

PRTnk

The result of the PPR resistance on the tree trunk detected 12 GLM compared to the canopy. C4 had the highest number of chromosomes (4 markers) for field resistance in PRTnk, followed by C8 (3 makers). Markers *mTcCIR446* and *mTcCIR304* on C1 were found significant only to PPR evaluated in the trunk. *mTcCIR434* and *mTcCIR308* also had significant marker trait associations on C2 and C9 respectively.

PTree

About 13 of the 16 markers were significant for the whole cacao tree (*PTree* = canopy + trunk) in GLM, and 8 in MLM. All the significant markers are not always the same between the canopy and the trunk. For example, *mTcCIR446* and *mTcCIR304* in C1were present only in PRTnk. Only four markers of C4 were found on both parts of the cacao tree in GLM and MLM. The marker *mTcCIR394* had the most significant marker (p=1.70-E6) of the field tests during GLM analysis for PTree. In total, PT5d co-localized with field measurement traits in 9 marker-trait associations.

Table 4: Markers showing significant association with PPR symptoms two (PT2d) and five days (PT5d) after pod inoculation

Marker	Chromosome	Position on the reference map (cM)	PT2d			PT5d		
			p		R^2	p		R^2
			Q	Q+K		Q	Q+K	
mTcCIR015	1	18.21	4.77E-04	1.74E-03	9.44E-02	-	-	-
mTcCIR419	1	18.73	2.86E-03	8.33E-04	5.55E-02	-	-	-
mTcCIR331	1	26.76	2.03E-03	1.27E-03	6.77E-02	-	-	-
mTcCIR416	1	27.2	9.37E-03	1.35E-02	3.23E-02	-	-	-
mTcCIR426	1	38.97	-	-	-	1.99E-02	2.19E-02	2.34E-02
mTcCIR304	1	60.9	-	-	-	2.50E-02	2.13E-02	2.34E-02
mTcCIR342	1	77.01	-	-	-	1.48E-04	2.31E-04	5.96E-02
mTcCIR333	1	84.96	-	-	-	5.82E-03	2.58E-02	2.11E-02
mTcCIR022	1	90.3	-	-	-	3.17E-02	-	-
mTcCIR434	2	5.38	-	-	-	4.80E-02	-	-
mTcCIR430	2	22.62	2.16E-02	4.25E-02	4.78E-02	-	-	-
mTcCIR411	2	33.94	-	-	-	1.99E-02	-	-
mTcCIR379	2	39.32	-	-	-	4.61E-02	-	-
mTcCIR431	3	46.65	2.13E-02	2.18E-02	2.39E-02	-	-	-
mTcCIR369	3	47.7	-	-	-	1.71E-02	2.25E-02	4.08E-02
mTcCIR402	4	10.23	3.51E-03	1.66E-03	4.71E-02	-	-	-
mTcCIR018	4	19.21	-	-	-	4.05E-02	2.35E-02	4.31E-02
mTcCIR420	5	29.67	4.95E-02	-	-	-	-	-
mTcCIR006	6	0	-	-	-	3.85E-02	-	-
mTcCIR439	6	7.41	4.54E-02	-	-	-	-	-
mTcCIR413	6	25.16	2.28E-02	2.32E-02	3.06E-02	-	-	-
mTcCIR301	6	37.38	4.98E-04	1.33E-03	6.85E-02	1.93E-02	1.08E-02	3.31E-02
mTcCIR398	6	38.51	4.49E-02	-	-	-	-	-
mTcCIR444	8	20.52	-	-	-	5.36E-03	1.51E-02	6.44E-02
mTcCIR310	8	20.9	-	-	-	2.71E-03	4.40E-03	7.48E-02
mTcCIR026	8	35.36	-	-	-	6.97E-03	1.74E-03	5.72E-02
mTcCIR382	8	38.07	-	-	-	3.28E-02	3.87E-02	5.57E-02
mTcCIR348	8	46.64	5.69E-03	3.01E-03	4.46E-02	-	-	-
mTcCIR391	8	48.91	4.03E-02	2.46E-02	2.48E-02	-	-	-
mTcCIR305	9	30.03	4.26E-03	1.64E-02	2.70E-02	1.21E-02	7.11E-03	3.10E-02
mTcCIR024	9	30.35	6.89E-05	3.78E-04	5.44E-02	-	-	-
mTcCIR429	9	68.45	-	-	-	2.65E-02	-	-
mTcCIR445	9	73.65	-	-	-	1.32E-02	1.31E-02	4.81E-02
mTcCIR383	10	14.7	-	-	-	4.35E-02	-	-
mTcCIR388	10	48.25	-	-	-	2.57E-02	4.72E-02	6.21E-02
mTcCIR377*	*	-	-	-	-	3.03E-02	2.05E-02	2.35E-02

*no mapped.

Table 5: Markers showing significant association with PPR incidence in the field at the levels of the canopy (PRCpy), the trunk (PRTnk) as well as on the whole tree (PRTree)

Marker	Chromo-some	Putative position (cM)	PRCpy			PRTnk			PRTree		
			p		R^2	p		R^2	p		R^2
			Q	Q+K		Q	Q+K		Q	Q+K	
mTcCIR446	1	47.44	-	-	-	1.92E-02	-	-	-	-	-
mTcCIR304	1	60.9	-	-	-	8.84E-03	4.06E-02	2.40E-02	-	-	-
mTcCIR434	2	5.38	-	-	-	2.15E-02	2.46E-02	4.81E-02	3.62E-02	-	-
mTcCIR431	3	46.65	3.15E-02	4.01E-02	2.10E-02	-	-	-	3.27E-02	-	-
mTcCIR355	4	8.38	9.35E-03	2.95E-02	5.31E-02	-	-	-	8.47E-03	3.06E-02	5.45E-02
mTcCIR402	4	10.23	6.78E-05	4.40E-03	3.79E-02	1.76E-04	2.47E-02	2.63E-02	3.37E-05	2.26E-03	4.39E-02
mTcCIR018	4	19.21	1.19E-03	1.06E-02	6.18E-02	3.81E-02	4.38E-02	3.51E-02	1.33E-03	1.17E-02	6.01E-02
mTcCIR394	4	19.74	1.99E-06	2.29E-04	1.08E-01	2.60E-03	3.06E-04	7.02E-02	1.70E-06	2.01E-04	1.08E-01
mTcCIR359	4	54.03	5.60E-03	7.79E-03	4.96E-02	1.36E-02	4.18E-03	5.85E-02	3.27E-03	4.18E-03	5.54E-02
mTcCIR344	4	64.28	1.50E-02	-	-	-	-	-	1.76E-02	-	-
mTcCIR301	6	37.38	-	-	-	1.16E-02	8.80E-03	2.79E-02	-	-	-
mTcCIR444	8	20.52	4.36E-02	-	-	2.12E-02	4.40E-02	2.05E-02	1.71E-02	2.72E-02	4.78E-02
mTcCIR310	8	20.9	2.69E-02	-	-	8.48E-03	-	-	1.02E-02	1.71E-02	5.04E-02
mTcCIR342	8	46.64	-	-	-	4.19E-02	1.30E-02	4.00E-02	-	-	-
mTcCIR308	9	11.0	-	-	-	3.38E-02	1.00E-02	5.14E-02	4.53E-02	-	-
mTcCIR294	9	16.05	3.83E-02	-	-	-	-	-	4.61E-02	-	-
mTcCIR393	9	34.57	6.94E-03	1.82E-02	3.41E-02	-	-	-	8.39E-03	1.98E-02	3.24E-02

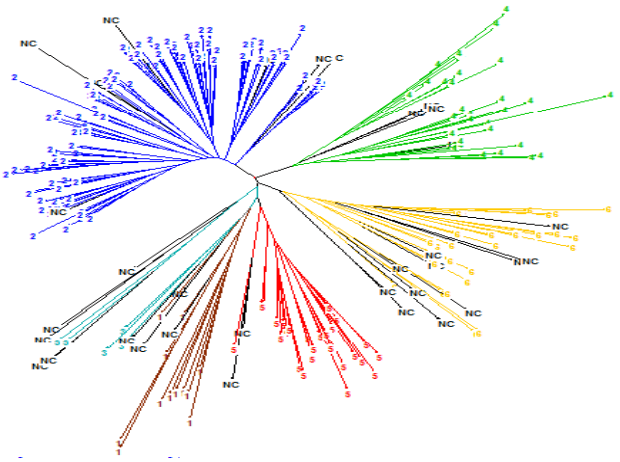

Fig. 1: NJ tree of cacao accessions based on 17 SSR markers. Projections of assignments based on Structure results (k=6 and 60% memberships in each cluster; NC= admixed accessions with no memberships above the 60% threshold on the NJ tree).

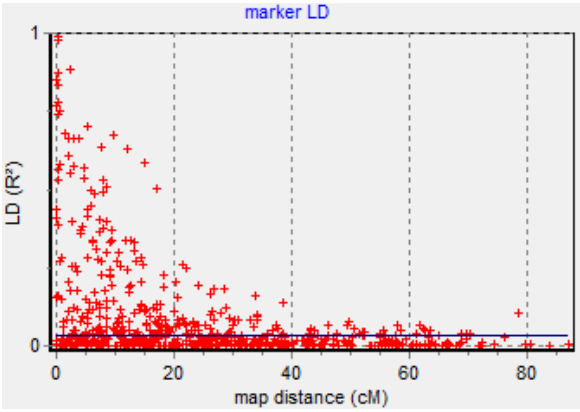

Fig. 2: LD (r^2) in the hybrid cacao population as a function of the genetic map distance between the SSRs markers. The *blue line* represents the second-degree LOESS smoothing.

DISCUSSION

Association mapping is gaining gradual importance in cacao breeding as an alternative method to overcome limitations of QTLs mapping in multi parental populations. To elucidate the genetic basis of PPR resistance, associations between the structure of a cultivated population and more than a hundred of SSR markers were analyzed. These QTLs are linked to different types of PPR resistance traits identified during artificial inoculation of pods as well as the disease rating on-farm.

Population structure

Generally, cacao populations display diverse structure in most cacao farms in Cameroon, and the genetic diversity is relatively high in plantations of hybrid cultivars (Efombagn *et al*., 2008). In association studies, the success of an association mapping can be influenced by the underlying population structure (Buckler and Thornsberry, 2002). The accuracy of the estimation of the percentage of admixtures and the kinship matrix is consequently important for the results of variance analysis. The percentages of admixture explained largely the trait variability, so an effect of population structure on the PPR traits variance was expected.

The unified mixed model-model method used in the present study has been reported to reduce the type I error in association mapping (Yu *et al*., 2006). Therefore, it has been possible to remove more structure effect, as shown by comparison between GLM and MLM analyses. Similar studies carried out recently on several cultivated species including wheat (Kulwal *et al*., 2012) and sorghum (Figueiredo *et al*., 2010) have shown the contribution of kinship in the control of the population structure. However, to get robust estimates of kinship, many more markers than our hundreds of SSRs are required (Yu *et al*., 2009).

Linkage disequilibrium

The extent of LD in crops is highly determined by the reproductive biology of the plant. Outbreeders like cacao have generally less LD than inbreeders. The resolution of association mapping depends on the LD genotyped markers and the causative polymorphism. LD decay is a good predictor of what is expected, given the present of marker density. In cocoa, LD can span long genetic distances along chromosome regions, as expected in populations derived from recent admixture Marcano *et al*. (2007). Our study revealed that the highest LD was detected between the markers pairs closely linked in the physical map positions. Results of the same LD analysis conducted by Marcano *et al*. (2007) with microsatellite markers indicated that loose genome coverage, with markers spread every 10-15 cM, is sufficient to identify chromosome regions of the cacao trait variations.

The analysis of the population structure in our study shows that the genotypes of the studied plantation are mixed. The practical consequence is the widening of the diversity on a single plot which leads to genome-wide association mapping study. Generally, diverse genetic material contributes to minimize the linkage disequilibrium (LD) and increase the probability to detect a large number of QTLs for the same trait. However, LD which results from the presence of subgroups in a diversified population with different allelic frequencies can affect the estimates. Association mapping depends on the choice of map taken to represent LD (Morton, 2005). The decrease of LD caused by recombination is useful for precision of mapping. Selection and mating system may also cause spurious association between markers and the scored traits within the mixed population of our study which underwent such processes. Knowing the amount and the structure of the LD in such population has greatly facilitated the association mapping study to identify markers linked to various PPR resistant traits.

Marker-Trait association

The results of the artificial inoculation showed that five associations were more likely found in C1, C6 and C8 for PT2d and PT5d. Previous mapping studies for PPR resistance carried out in cacao populations revealed the presence of QTLs identified during inoculations of leaves and pods as well as measurement of the percentage of diseased fruits in planted cacao trees (Lanaud *et al*., 1999, 2004; Risterucci *et al*., 2003). Only CH4 registered more than five significant associations with Q and Q+K in field

observation of PPR incidence. Lanaud *et al*. (2009) have detected about 13 QTLs for PPR resistance in C1 and C4. The presence of 'hot spot' regions corresponding to a large number of QTLs related to PPR resistance in these chromosomes tallied with the relative high number of significant associations identified on the same chromosome groups in our study. However, several minor genes localized along the ten linkage groups of the cacao genome, are not always involved in the same PPR resistance trait. Therefore, some marker-trait associations were only detected in inoculated pods where as other ones are found in the field incidence of the disease. For example, Flament *et al*. (2001) performed artificial inoculations on leaves and pods. These tests were weakly correlated with the pod rot rate in the field. QTLs of resistance were detected but none were common between the three traits measured. Similar specialization patterns of QTLs for disease resistance were previously observed in other cacao diseases such as Witches' Broom (Brown *et al*., 2005) and Frosty pod (Brown *et al*., 2007). It shows that some cacao vegetative traits were correlated to percentage of PPR-infected organs. The QTLs of PPR resistance and their closed markers of a hybrid population analyzed can have direct and practical application through association mapping. This is because many of their alleles will be represented in the association mapping panel as shown by Breseghello and Sorrels (2006).

The present association mapping experiment was carry out under a single location which is however subjected to frequent seasonal variation (two wet and two dry seasons). This seasonal variation could play key role on determining the significance of a marker-trait association. Similar QTLs and marker-trait stability across various environmental conditions was shown by Kulwal *et al*. (2012) in association mapping experiment for pre-harvest sprouting resistance in wheat. Ndoumbè *et al*. (2002) have demonstrated that differences in the epidemiology feature of PPR might occur in a single cacao plantation in Cameroon, basically between the lower stage of the cacao tree (Trunk) and the upper stage (canopy). Even if a significant correlation were found between PRCpy and PRTnk traits, several marker-trait associations were not the same between the trunk and the canopy.

Acknowledgements

We are grateful for the technical support of the personnel of the Cacao Program of the Institute of Agricultural Research (IRAD), Cameroon and the UMR-AGAP of the Centre of International Cooperation in Agronomic Research (CIRAD), France. We also thank René, the owner of the cacao plantation used for the study. This work was financially supported by Mars, Inc.

REFERENCES

Allègre M, X Argout, M Boccara, O Fouet, Y Roguet, A Bérard, JM Thévenin, A Chauveau, R Rivallan, D Clement, B Courtois, K Gramacho, A Boland Augé, M Tahi, P Umaharan, D Brunel andC Lanaud, 2012. Discovery and mapping of a new expressed sequence tag-single nucleotide polymorphism and simple sequence repeat panel for large-scale genetic studies and breeding of *Theobroma cacao* L. DNA Research, 19: 23-35.

Bartley BGD, 2005. The genetic diversity of cocoa and its utilization. CABI Publishing (ed), Wallingford, UK. 341p.

Breseghello F andME Sorrells, 2006. Association analysis as a strategy for improvement of quantitative traits in plants. Crop Sci, 46: 1323–1330.

Brown SJ, RJ Schnell, JC Motamayor, U Lopes, D Kuhn andJ Borrone, 2005. Resistance gene mapping for witches' broom disease in *Theobroma cacao* L. in an F2 population using SSR markers and candidates genes. J Amer Soc Hort Sci, 130: 366-373.

Brown SJ, W Phillips-Mora, EJ Power, C Krol, MC Cervantes, JC Motamayor andRJ Schnell, 2007. Mapping QTLs for resistance to frosty pod and black pod diseases and horticultural traits in *Theobrome cacao* L. Crop Sci, 47: 1851-1858.

Buckler ES andJM Thornsberry, 2002. Plant molecular diversity and application to genomics. Curr Opin Plant Biol, 5: 107-111.

Cavalli-Sforza LL andWF Bodmer, 1971. The genetics of human populations, W.H. Freeman and Company, NY.

Crouzillat D, B Menard, A Mora, W Phillips andV Petiard, 2000a. Quantitative traits analysis in *Theobroma cacao*using molecular markers. Yield QTL detection and stability over 15 years. Euphytica, 114: 13-23.

Crouzillat D, W Phillips, P Fritw andV Petiard, 2000b. Quantitative traits analysis in *Theobroma cacao* using molecular markers. Inheritance of polygenic resistance to *Phytophthora palmivora* in two related cacao populations. Euphytica, 114: 23-26.

Despréaux D, D Cambrony, D Clément, S Nyassé andM Partiot, 1988. Etude de la pourriture brune des cabosses du cacaoyer au Cameroun: définition des nouvelles méthodes de lutte. Xth International Conference on Cacao Research, Dominican Republic, pp: 407-412.

Dieringer D andC Schlötterer, 2003. Microsatellite analyser (MSA): a platform independent analysis tool for large microsatellite data sets. Mol Ecol Notes, 3: 167-169.

Efombagn MIB, JC Motamayor, O Sounigo, AB Eskes, S Nyassé, C Cilas, R Schnell, MJ Manzanares-Dauleux andM Kolesnikova-Allen, 2008. Genetic diversity and structure of farm and genebank accessions of cacao (*Theobroma cacao* L.) in Cameroon revealed by microsatellite markers. Tree Genetics & Genome, 4: 821-831.

Falush D, M Stephens andJK Pritchard, 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics, 164:1567–1587.

Falush D, M Stephens andJK Pritchard, 2007. Inference of population structure using multilocus genotype data: dominant markers and null alleles. Mol Ecol Notes, 7: 574–578.

Figueiredo LFA, B Sine, J Chantereau, C Mestres, G Fliedel, JF Rami, JC Glaszmann, M Deau and Courtois B, 2010. Variability of grain quality in sorghum: association with polymorphism in *sh*2, *bt*2,

*Sss*I, *Ae*1, *Wx* and *O2*. Theor Appl Genet, 121: 1171-1185.

Flament MH, I Kebe, D Clément, I Pieretti, AM Risterucci, JAK N'Goran, C Cilas, D Despréaux and Lanaud C, 2001. Genetic mapping of resistance factors to *Phytophthora palmivora* in cacao. Genome, 44: 79-85.

Fouet O, M Allègre, X Argout, M Jeanneau, A Lemainque, S Pavek, A Boland, AM Risterucci, G Loor, GM Tahi, X Sabau, B Courtois andC Lanaud, 2011. Structural characterization and mapping of functional EST-SSR markers in *Theobroma cacao*. Tree Genetics & Genomes, 7: 799-817.

Gutierrez OA, C Nagai, R Heinig, EJ Boza, CL Tondo, JC Motamayor and RJ Schnell, 2012. Association mapping in cacao. Annual International Plant & Animal Genome Conference. XX Annual International Plant & Animal Genome Conference, January 14-18, 2012, San Diego, CA.

Iwaro AD, JM Thevenin, DR Butler andAB Eskes, 2006. Usefulness of the Detached Pod Test for Assessment of Cacao Resistance to *Phytophthora* Pod Rot. Eur J Plant Path, 113: 173-182.

Kulwal P, G Ishikawa, D Benscher, Z Feng, LX Yu, A Jadhav, S Mehetre andME Sorrells, 2012. Association mapping for pre-harvest sprouting resistance in white winter wheat. Theor Appl Genet, 125: 793–805.

Lanaud C, AM Risterucci, JAK N'Goran, D Clement, MH Flament, V Laurent andM Flaque, 1995. A genetic linkage map of *Theobroma cacao*L. Theor Appl Genet, 91: 987-993.

Lanaud C, I Kebe, AM Risterucci, D Clement, JAK N'Goran, L Grivet, M Tahi, C Cilas, I Pieretti, AB Eskes and D Despreaux, 1999a. Mapping quantitative trait loci (QTL) for resistance to *Phytophthora palmivora* in *T.cacao* L. Proc Intl Cacao Res Conf, 12: 99-105.

Lanaud C, AM Risterucci, I Pieretti, M Falque, A Bouet and PJL Lagoda, 1999b. Isolation and characterization of microsatellites in *Theobroma cacao* L. Mol Ecol, 8: 2141–2152.

Lanaud C, D Clement, MH Flament, AM Risterucci, I Kebe, S Nyasse, O Sounigo, L Motilal, JM Thevenin, D Paulin, M Ducamp, JAKN'Goran, D Fargeas and C Cilas, 2004. Genetic mapping of quantitative trait loci for black pod resistance in cacao. In: Cilas C, Despreaux D (eds) Improvement of cacao tree resistance to *Phytophthora* diseases. Collection repères, Cirad, pp: 147–164.

Lanaud C, O Fouet, D Clement, M Boccara, AM Risterucci, S Surejdeo-Maharaj, T Legavre, X Argout, 2009. A meta-QTL analysis of disease resistance traits of *Theobroma cacao* L. Mol Breeding, 24: 361-374.

Lewontin RC, 1964. The interaction of selection and linkage. I. General considerations; heterotic models. Genetics, 49: 49–67.

Loor RG, 2007. Domestication of the Nacional *Theobroma cacao* L. variety from Ecuador: An approach to identify the native Nacional variety and its wild ancestors. Ph D Thesis, ENSAM, Montpellier, France, 133p.

Marcano M, T Pugh, E Cros, S Morales, PEA Portillo, B Courtois, JC Glaszmann, JMM Engels, W Phillips, C Astorga, AM Risterucci, O Fouet, V González, K Rosenberg, I Vallat, M Dagert and C Lanaud, 2007. Adding value to cacao (*Theobroma cacao* L.) germplasm information with domestication history and admixture mapping. Ther Appl Genet, 114: 877-884.

Marcano M, S Morales, MT Hoyer, B Courtois, AM Risterucci, O Fouet, T Pugh, E Cros, V Gonzalez, M Dagert and C Lanaud, 2009. A genome wide admixture mapping study for yield factors and morphological traits in a cultivated cacao (*Theobroma cacao* L.) population. Tree Genetics & Genomes, 5: 329-337.

Morton NE, 2005. Linkage disequilibrium maps and association mapping. J Clin Invest, 115: 1425-1430.

Ndoumbè-Nkeng M, 2002. Incidence des facteurs agro-écologiques sur l'épidémiologie de la pourriture brune des fruits du cacaoyer au Cameroun: contribution à la mise en place d'un modèle d'avertissement agricole. Paris, France: INAPG, PhD thesis.

Perrier X and JP Jacquemoud-Collet, 2006.DARwin software: http://darwin.cirad.fr/darwin

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

Pugh T, O Fouet, AM Risterucci, P Brottier, C Deletrez, B Courtois, D Clément, P Lamande, JAK N'Goran and C Lanaud, 2004. A new codominant markers based cocoa linkage map: Development and integration of 201 new microsatellites markers. Theor Appl Genet, 108: 1151–1161

Risterucci AM, D Paulin, M Ducamp, JAK N'Goran and C Lanaud, 2003. Identification of QTLs related to cacao resistance to three species of *Phytophthora*. Theor Appl Genet, 108: 168-174.

Storey JD, 2002. A direct approach to false discovery rates. Journal of the Royal Statistical Society, Series B, 64: 479-498.

Van BERLOO R, 2008. GGT 2.0: Versatile Software for Visualization and Analysis of Genetic Data. J Hered, 99: 232 -236.

Yu J, G Pressoir, WH Briggs, Bi I Vroh and M Yamasaki, 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet, 38: 203–208.

Yu J, Z Zhang, C Zhu, DA Tabanao, G Pressoir, MR Tuinstra, S Kresovich, RJ Todhunter and ES Buckler, 2009. Simulation appraisal of adequacy of number of background markers for relationship estimation in association mapping. Plant Genome, 2: 63-77.

Zhao H, D Nettleton, M Soller and JCM Dekkers, 2005. Evaluation of linkage disequilibrium measures between multi-allelic markers as predictors of linkage disequilibrium between markers and QTL. Genet Res Camb, 86: 77-87.