Planta (2016) 244:1265–1276
DOI 10.1007/s00425-016-2582-7

ORIGINAL ARTICLE

Association mapping of seed and disease resistance traits in *Theobroma cacao* L.

Lambert A. Motilal[1] · **Dapeng Zhang**[2] · **Sue Mischke**[2] · **Lyndel W. Meinhardt**[2] · **Michel Boccara**[1,3] · **Olivier Fouet**[3] · **Claire Lanaud**[3] · **Pathmanathan Umaharan**[1]

Received: 27 December 2015 / Accepted: 8 August 2016 / Published online: 17 August 2016
© Springer-Verlag Berlin Heidelberg 2016

Abstract

Main conclusion **Microsatellite and single nucleotide polymorphism markers that could be used in marker assisted breeding of cacao were identified for number of filled seeds, black pod resistance and witches' broom disease resistance.**

An association mapping approach was employed to identify markers for seed number and resistance to black pod and witches' broom disease (WBD) in cacao (*Theobroma cacao* L.). Ninety-five microsatellites (SSRs) and 775 single nucleotide polymorphisms (SNPs) were assessed on 483 unique trees in the International Cocoa Genebank Trinidad (ICGT). Linkage disequilibrium (LD) and association mapping studies were conducted to identify markers to tag the phenotypic traits. Decay of LD occurred over an average 9.3 cM for chromosomes 1–9 and 2.5 cM for chromosome 10. Marker/trait associations were generally identified based on general linear models (GLMs) that incorporated principal components from molecular information on relatedness factor. Seven markers (mTcCIR 8, 66, 126, 212; TcSNP368, 697, 1370) on chromosomes 1 and 9 were identified for number of filled seeds (NSEED). A single marker was found for black pod resistance (mTcCIR280) on chromosome 3, whereas six markers on chromosomes 4, 5, 6, 8, and 10 were detected for WBD (mTcCIR91, 183; TcSNP375, 720, 1230 and 1374). It is expected that this association mapping study in cacao would contribute to the knowledge of the genetic determinism of cocoa traits and that the markers identified herein would prove useful in marker assisted breeding of cacao.

Keywords Black pod disease · Cocoa seeds · Linkage disequilibrium · Microsatellite markers · Single nucleotide polymorphism · Witches' broom disease

Abbreviations

DIPP	Fruit resistance to *P. palmivora* from laboratory spray inoculation tests
GLM	General linear model
ICGT	International Cocoa Genebank Trinidad
LD	Linkage disequilibrium
MLM	Mixed linear model
NSEED	Number of filled seeds
QTL	Quantitative trait loci
RWG	Reproductive witches' broom per tree girth
RWY	Reproductive witches' broom per year
SNP	Single nucleotide polymorphism
SSR	Microsatellite
TWG	Total witches' broom per tree girth
WBD	Witches' broom disease

Introduction

Cacao (*Theobroma cacao* L.) is a diploid ($2n = 20$) tree in the family Malvaceae (Alverson et al. 1999; Bayer et al. 1999). It is an economically important plantation crop for

---

✉ Lambert A. Motilal
lambert.motilal@sta.uwi.edu; lamotilal@yahoo.com

[1] Cocoa Research Centre, The University of the West Indies, Sir Frank Stockdale Bldg., St. Augustine, Trinidad 330912, Trinidad and Tobago

[2] USDA/ARS, Beltsville Agricultural Research Center, PSI, SPCL, 10300 Baltimore Avenue, Bldg. 001, Rm. 223, BARC-W, Beltsville, MD 20705, USA

[3] CIRAD-UMR AGAP (Centre de coopération internationale en recherche agronomique pour le développement), Montpellier Cedex 5, France

many tropical countries with the fermented and dried cotyledons (beans) contributing to a multi-billion dollar confectionery industry. The centre of origin and diversity of *Theobroma* is in the Western Upper Amazon (Thomas et al. 2012). Molecular evidence has been used to define ten genetic clusters in cacao (Motamayor et al. 2008). The Refractario group, selected in Ecuador for witches' broom disease (WBD) resistance, was distinct from the established ten clusters except for Nacional members, and represent hybrids between Nacional and Trinitario genotypes (Zhang et al. 2008).

Genetic resources of cacao, maintained in field gene banks as national or universal collections, have been described (Bartley 2005; Zhang et al. 2011; Turnbull and Hadley 2015). Cacao has a characteristic incompatibility system that allows for cross-compatibility among and within populations, cross-compatibility among varieties; and selfing in some varieties (Cope 1962). Hence, wild germplasm can be easily used in breeding programmes (Eskes and Efron 2006), but only a very small fraction of the existing variation has been exploited in breeding programmes (CacaoNet 2012). A long breeding cycle has been a major deterrent in breeding of most tree crops. In cacao, the breeding cycle to develop hybrid varieties with testing for yield was estimated to take at least 12 years (Toxopeus 1985; Kennedy et al. 1987). Successful breeding programmes in Brazil (Lopes et al. 2011) and Trinidad (Maharaj et al. 2011) have released varieties after at least 16 and 60 years, respectively.

Two diseases are presently of economic importance in the Caribbean—black pod disease caused by *Phytophthora palmivora* (Braiser and Griffin 1979) and WBD caused by *Moniliophthora perniciosa* (Aime and Phillips-Mora 2005). The progress in breeding for resistance to these two diseases has been hampered by the use of phenotypic evaluations to identify promising selections rather than the use of molecular screening tools. Crouzillat et al. (2000) showed that, in cacao, the use of molecular markers alone or in combination with phenotypic selection was more effective than only phenotypic evaluation. This molecular breeding approach (marker-assisted selection, marker-aided selection or marker-assisted breeding), uses marker(s) associated with quantitative trait loci (QTL) to tag traits, thereby identifying improved individuals (Collard et al. 2005). Although, marker assisted selection from QTLs has been successful in other crops, inconsistent results and limited transferability over crosses have also been noted (Hospital 2003). Furthermore, classical QTL analyses make little or no use of ancestry information unlike admixture and association mapping methodologies.

Admixture mapping, premised on population differentiation between ancestral populations from two known progenitors, uses the local phenotype–ancestry correlation in recent (<20 generations) admixtures (Shriner 2013). Marcano et al. (2007, 2009) employed admixture mapping to identify several regions and microsatellites (SSRs) linked to various traits including productivity and yield in cacao. Association mapping [association analysis or linkage disequilibrium (LD) mapping], unlike admixture mapping, tests the genotype-phenotype correlation, and is premised on similar allele frequencies across multiple ancestries allowing for fine-scale localization (Buckler and Thornsberry 2002; Shriner 2013). Linkage disequilibrium is the non-random combination of alleles at two or more loci. Spurious or false associations may arise and can be reduced by accounting for population stratification and relatedness (Pritchard 2001; Yu et al. 2006). Markers identified under association mapping are not restricted to a population or populations and the identification is reliant on LD to examine the correlation between phenotypic variation and genetic polymorphisms (Abdurakhmonov and Abdukarimov 2008).

Association mapping started with Bar-Hen et al. (1995) and numerous studies on a wide range of plants have since been reviewed (Abdurakhmonov and Abdukarimov 2008; Zhu et al. 2008; Ingvarsson and Street 2011; Gupta et al. 2014). However, association mapping in cacao is a relatively young field. Pugh (2005) first reported on disequilibrium in a Trinitario population. Stack et al. (2015) used 96 SSR loci and estimated LD in cacao to either rapidly decay (1–2 megabases) or to persist for extended distances (30 megabases) in wild and cultivated diversity groups, respectively. Schnell et al. (2005) employed a trend test to identify SSRs associated with productivity. Motamayor et al. (2013), using association mapping among other techniques, identified a single nucleotide polymorphism (SNP), among the 168 that were studied, that affected gene expression for fruit colour. Recently, da Silva et al. (2016) using 5301 SNPs and 295 individuals identified eight significantly associated SNPs that tagged a major gene on chromosome 4 for self-compatibility. This study was, therefore, undertaken as a model association mapping study, to search for SSR or SNP markers that may be linked to seed number and disease resistance in *T. cacao*.

Materials and methods

Phenotyping

Data was accumulated over 2007–2012 on cacao trees of 483 unique accessions in the International Cocoa Genebank Trinidad (ICGT). Details on the ICGT can be found in Motilal et al. (2012) and references therein. Data was collected from 1 to 6 trees per accession on number of filled seeds (NSEED), resistance of fruit to *P. palmivora*, and field reactions to *M. perniciosa*. Data points were averaged over trees that were verified as genetically identical from SSR data (Motilal et al. 2012, 2013).

To determine NSEED, 3–34 fruits per tree were harvested from the main trunks and primary or secondary branches. The fruits were cracked open and the number of filled seeds was counted in each fruit. Fruit resistance to *P. palmivora* was obtained from in-house data of laboratory spray inoculations tests (DIPP) carried out on intact detached fruits and scored on an eight-point scale (Iwaro et al. 2000). Field WBD observations at the ICGT were made continually for each tree, without removal of infected organs, at least three times over the entire 6-year period and many accessions were visited twice yearly. Tree girths were measured with a tailor tape at 135 cm from soil level. Vegetative brooms, reproductive brooms (floral, cushion) and diseased fruits were counted over the entire tree above soil level. Trees which had zero counts of diseased organs were retained if zero counts were observed over a four-year period. Trees with more than one survey point per year were allowed the maximum observation value for the year. Data on WBD was manipulated to give six phenotypic expressions:

1. Vegetative witches' broom per year.
2. Vegetative witches' broom per tree girth.
3. Reproductive witches' broom per year (RWY).
4. Reproductive witches' broom per tree girth (RWG).
5. Total witches' broom per year.
6. Total witches' broom per tree girth (TWG).

Correlation of phenotypic traits

Correlation analyses on ungrouped data were performed in the program MedCalc v.12.7.7.0 (MedCalc Statistical Software 2013) using Spearman's rank correlation coefficient. Effect sizes were in accordance with Hopkins (2002).

burn-in of 500,000 and 1×10^6 Markov chain-Monte Carlo runs with 20 iterations. Population structure was used as the ancestry covariate in the association mapping analysis. Subsets of markers were, therefore, used to enable complementary analyses without having similar markers in the same ancestry and genotype files within the same run.

Linkage disequilibrium

The consensus SSR/SNP map (Allegre et al. 2012) was obtained from CocoaGenDB (http://cocoagendb.cirad.fr/about.htm). The SSR and SNP data were treated as unphased data and LD was calculated using PowerMarker v3.25 (Liu and Muse 2005) from (1) entire but unordered data sets and (2) ordered markers for each of the ten linkage groups. The multiallelic r^2 parameter for LD of locus pairs was utilised (Hill and Weir 1994). Exact P values were calculated from a maximal iteration of 1×10^6 Markov chain-Monte Carlo runs with the exact test for multilocus association (Zaykin et al. 1995). A scatter plot of LD as multiallelic r^2 vs the map distance of all possible marker pairs was created for each chromosome. A local regression (loess) curve (Cleveland 1979; Cleveland and Devlin 1988) was fitted using the PTS LOESS calculator, a LOESS utility add-in for Excel (Peltier 2009). The map distance coincident with $r^2 = 0.1$ was identified as the decay distance (Abdurakhmonov and Abdukarimov 2008).

Association mapping analysis

Association analysis was configured in TASSEL v4.2.1 (Bradbury et al. 2007). Incorporation of a kinship matrix turned a general linear model (GLM) into a mixed linear model (MLM). Three GLM and four MLM models were employed:

$$
\begin{array}{llll}
\mathrm{GLM1}: Y & = E + A + S_Q t & \mathrm{MLM1}: & Y = E + A + S_K t \\
\mathrm{GLM2}: Y & = E + A + S_M t & \mathrm{MLM2}: & Y = E + A + S_Q t + S_E t \\
\mathrm{GLM3}: Y & = E + A + S_Q t + S_M t & \mathrm{MLM3}: & Y = E + A + S_M t + S_K t \\
 & & \mathrm{MLM4}: & Y = E + A + S_Q t + S_M t + S_K t
\end{array}
$$

Population structure

SSR polymorphisms were obtained as described in Motilal et al. (2009) and SNP polymorphisms were obtained from Illumina GoldenGate Assays as described by Allegre et al. (2012). Population structure was determined independently on unique sets of either SSRs (27 or 52) or SNPs (125 or 281) with STRUCTURE v.2.3.4 (Pritchard et al. 2000) using a

Under this YEAST model, the system information ($S_{\alpha}t$) can be taken from the ancestry information ($S_Q t$), the multivariate analysis based on molecular data ($S_M t$), or the kinship relationship based on molecular data ($S_K t$). The terms E and A refer to the error and allele information, respectively.

The genotype file was filtered to remove alleles with frequency <0.01 and retained for further manipulation. The number of markers used is presented in Table 1. The filtered

Table 1 Synopsis of association mapping strategy in this study

Dataset	# Ancestry markers	Ancestry filter	# Markers for PCA and kinship	# Tassel markers	Models
B_all SSR	None	None	43 (3 PCA axes)	95	GLM2, MLM3
B1 SSR	52	Q10 to Q9	21 (3 PCA axes)	43	GLM1, 2, 3; MLM1, 2, 3, 4
B2 SSR	27	Q10 to Q9	32 (3 PCA axes)	68	GLM1, 2, 3; MLM1, 2, 3, 4
C_all SNP	None	None	432 (3 PCA axes)	775	GLM2, MLM3
E SNP	121	Q11 to Q10	254 (3 PCA axes)	654	GLM2, 3; MLM3, 4
F SNP	281	Q9 to Q8	254 (3 PCA axes)	368	GLM2, 3; MLM3, 4

genotype file was used to create the kinship matrix in TASSEL v4.2.1 (Bradbury et al. 2007). The filtered genotype file was collapsed and markers with >10 % (SSR) or >0.3 % (SNP) missing data were excluded from the un-collapsed filtered genotype file. After removal of these markers, the pruned file was collapsed and missing values were then imputed from unweighted averages of three nearest neighbours using a Manhattan distance. The principal components matrix was created from the repopulated collapsed file using a covariance method, and eigenvectors were retained for three axes. Ancestries based on SNP and SSR datasets were submitted as independent datasets using the best population substructure. Markers used for determination of ancestry were excluded from the corresponding allele file for association mapping analysis, as the ancestry file was used as a covariate. One of the populations was removed prior to analysis. Datasets were joined using the intersect function to minimise the incidence of missing phenotypic values across genotypes or allelic information for phenotypes.

The association mapping strategies (Table 1) included both general linear models (GLMs) and mixed linear models (MLMs). GLMs using a least squares solution (Searle 1987) on trait data were run independently on both SSR and SNP data, using the default settings of 1000 permutations and the permutation test of Anderson and Ter Braak (2003). MLMs were run independently on both SSR and SNP data, using optimum level compression (Yu et al. 2006; Zhang et al. 2010) with estimation of the variance component by Zhang et al. (2010). Sample sizes within the SSR datasets/model combinations ranged from 263 to 369 for NSEED, 179 to 267 for DIPP, and 275 to 374 for WBD reactions. Sample sizes within the SNP datasets/model combinations were 324 for NSEED, 229 for DIPP and 265 for WBD reactions.

Selecting associated markers

Probability values were compared to Bonferroni-corrected *P* values (Bonferroni 1936; Dunn 1961) at the 5 % level of significance. Final selection of significantly associated markers employed the following criteria:

1. Present in more than one dataset.
2. Present in at least two models.
3. Most constraining model or dataset chosen from (1) and (2) above.
4. If LD as r^2 is $\geq$0.1, then only one marker was chosen when two markers were significant; when more than two markers were significant, the smallest set of markers to represent the total set was constructed by selecting markers with lowest *P* values, selecting markers significantly associated with more than one trait and selecting markers with at least five observations in the effect size.

Results

Phenotyping

Descriptive statistics of the studied traits are provided in Table 2. Fifteen accessions in the ICGT (CL 10/5, CRU 100, CRUZ 7/8, EET 272, EET 48, ICS 70, IMC 20, IMC 47, MOQ 6/82, NA 168, NA 399, NA 719, PA 46, SCA 6, and SLC 4) were top ranked for DIPP resistance. The top three accessions with the highest average seed numbers were CRU 72 (58.0 ± 1.7), IMC 54 (57.3 ± 2.1), IMC 77 (62.3 ± 0.9) and these accessions had high Iquitos

Table 2 Summary statistics of seed and disease resistance traits over accessions

Trait	*N*	Mean	SD	RSD	Range
NSEED	493	37.9	6.9	0.182	13–62
DIPP	350	5.7	1.97	0.348	1–8
VWY	424	6.15	7.15	1.164	0–56.7
VWG	423	0.62	0.83	1.337	0–5.8
RWY	424	2.07	4.02	1.944	0–33.3
RWG	423	0.18	0.36	1.953	0–2.8
TWY	424	8.20	9.81	1.195	0–72.7
TWG	423	0.80	1.08	1.343	0–7.6

Trait: *NSEED* number of filled seeds, *DIPP Phytophthora* pod rot reaction based on Iwaro et al. (2000), *VWY* vegetative witches' broom per year, *VWG* vegetative witches' broom per tree girth, *RWY* reproductive witches' broom per year, *RWG* reproductive witches' broom per tree girth, *TWY* total witches' broom per year, *TWG* total witches' broom per tree girth, *N* number of samples, *SD* standard deviation, *RSD* relative standard deviation

Table 3 Spearman's rank correlation analyses of fruit, seed and disease incidence traits of *Theobroma cacao* L

	A	B	C	D	E	F	G	H
A	–	343	399	399	399	399	399	399
B	0.01	–	292	292	292	292	292	292
C	0.01	*0.22*	–	424	424	424	424	424
D	−0.08	*0.24*	*0.88*	–	424	424	424	424
E	0.02	0.05	*0.44*	*0.36*	–	424	424	424
F	-0.03	0.10	*0.49*	*0.51*	*0.95*	–	424	424
G	0.01	*0.19*	*0.94*	*0.81*	*0.68*	*0.69*	–	424
H	−0.08	*0.23*	*0.86*	*0.97*	*0.55*	*0.68*	*0.87*	–

Traits assessed were: A, *NSEED* number of filled seeds; B, *DIPP* detached fruit response to spray inoculations with *Phytophthora palmivora*; C, log vegetative witches' broom per year; D, log vegetative witches' broom per tree girth; E, log reproductive witches' broom per year; F, log reproductive witches' broom per tree girth; G, log total witches' broom per year; H, log total witches' broom per tree girth. Values above the diagonal represent sample sizes. Values below the diagonal represent Spearman's rank correlation as calculated in MedCalc v.12.7.7.0 (MedCalc Statistical Software 2013). Cells which are italicized are significant at the 0.1 % level

ancestry. The top three accessions for WBD as TWG or total witches' broom per year were CRU 128, LZ 2 and PA 67 from various ancestral backgrounds. The trait total witches' broom per year, had four other accessions [IMC 50 (Iquitos); MOQ 6/6 (Refractario2); PA 279 and PA 289 (Marañón)] that was coincident with PA 67.

Spearman's rank correlation effects were non-significant between NSEED and disease reactions (Table 3). DIPP was significantly correlated ($P < 0.001$) with low effect (0.2) for vegetative and total WBD reactions. The various forms of derived WBD measurement were, as expected, significantly ($P < 0.001$) correlated and usually of large effect with each other. However, the discrete forms of WBD, vegetative (vegetative witches' broom per year, vegetative witches' broom per tree girth) and reproductive (RWY, RWG) exhibited only a moderate effect for the vegetative vs. reproductive pairwise comparisons.

Linkage disequilibrium

A maximum of 92 SSRs and 762 SNPs (854 loci) were retained for analysis with a maximum of 364,231 pairwise combinations, of which 44,507 (12.2 %) marker pairs were in LD ($r^2 > 0.104$).

There were 85 SSRs and 640 SNPs (725 loci) on the consensus map giving a maximum of 262,450 pairwise combinations, of which 31,605 (12 %) marker pairs were in LD ($r^2 > 0.104$). Linkage disequilibrium was partitioned as intra-chromosomal LD (7192 marker pairs; 22.8 %) and as inter-chromosomal LD (24,413 marker pairs; 77.2 %). Decay of LD occurred over 2.50–12.00 cM at an average of 8.63 cM (Fig. 1; Table 4). Chromosome 10 was strikingly different from the others with the shortest decay length of 2.5 cM. The average decay distance, omitting chromosome 10, was 9.3 cM. The decay of LD for chromosomes 1–9 inclusive was taken as 9.3 cM or 2706, 821 bp and for chromosome 10 as 2.5 cM or 727,640 bp.

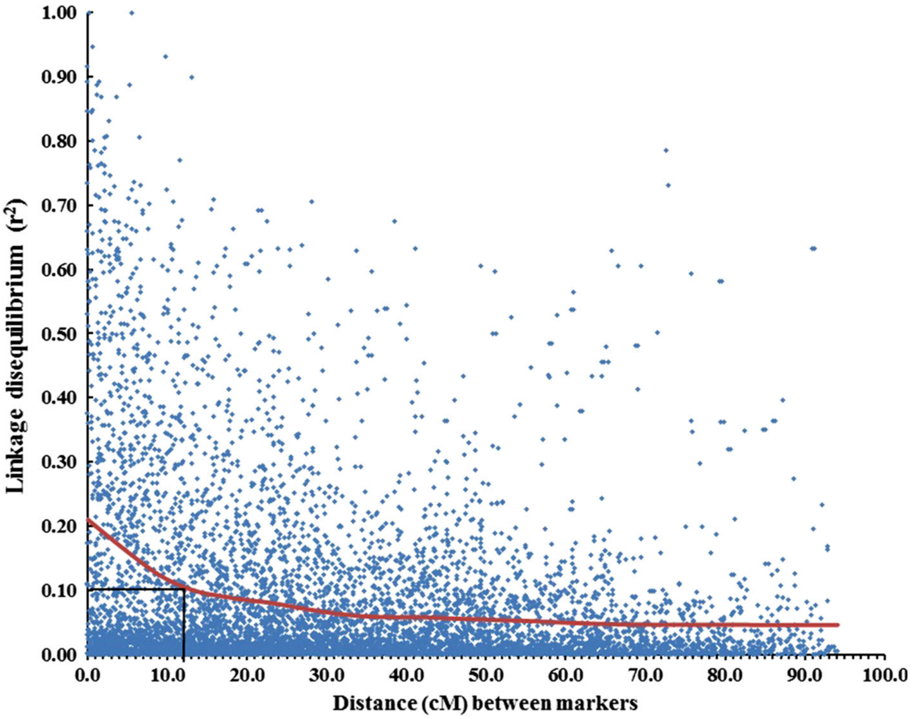

Fig. 1 Decay of linkage disequilibrium with map distance in chromosome 1 of the SSR/SNP consensus map of CocoaGenDb (http://cocoagendb.cirad.fr/). Linkage disequilibrium estimates were obtained in PowerMarker v3.25 (Liu and Muse 2005). Loess curve fitted using the PTS LOESS Calculator (Peltier 2009). Chromosomal map distance at which $r^2 = 0.1$ is indicated

Table 4 Marker coverage and decay of linkage disequilibrium (LD) for the ten chromosomes of cocoa (*Theobroma cacao* L.)

Chrom[a]	Consensus length (cM)	Length covered (cM)	# markers in study; (# per cM)	# Marker pairs (Mprs)	# Mprs with $r^2 > 0.104$	Length (cM) at LD of $r^2 = 0.1$
1	94.1	94.1	122 (1.30)	7381	1745	12.00
2	101.1	98.5	88 (0.89)	3828	750	6.50
3	76.9	75.4	98 (1.30)	4753	1168	10.00
4	64.2	62.6	77 (1.23)	2926	648	8.50
5	78.1	77.6	83 (1.07)	3403	699	9.50
6	64.0	59.0	39 (0.66)	741	189	9.50
7	52.6	51.4	30 (0.58)	436	95	9.25
8	59.2	52.3	46 (0.88)	1035	243	7.50
9	100.9	94.7	112 (1.18)	6216	1595	11.00
10	59.5	57.6	30 (0.52)	435	60	2.50
Total over all chromosomes	750.6	723.2	725 (1.00)	31,154	7192	86.25
Average per chromosome	75.1	72.3	73 (1.01)	3115	719	8.63

[a] Chromosome from the SSR/SNP consensus map of CocoaGenDb (http://cocoagendb.cirad.fr/). LD estimates were obtained in PowerMarker v3.25 (Liu and Muse 2005). Only markers in the consensus map were used

Association mapping

The majority of the indicated markers were obtained under GLM rather than MLM models. Examples of the Manhattan and quantile–quantile plots are given in Fig. 2. A minimal set of 7 SSRs and 7 SNPs (Table 5) were selected as associated markers. Associations were usually on a one-to-one basis except for mTcCIR91 which was associated with logRWG, logRWY and logTWG and for TcSNP1230 which was associated with both logRWY and log total witches' broom per year. The adjusted Bonferroni threshold for $\alpha = 0.05$ was 5.26×10^{-4} and 6.45×10^{-5} for all of the SSR and SNP loci, respectively. Applying a threshold of 5×10^{-5} and 5×10^{-6} for SSR and SNP loci identified one locus (mTcCIR126) that was strongly associated under a GLM2 model that tagged NSEED (Table 5). The SSR and SNP markers explained between 4.24–7.53 and 4.95–11.01 % of the variation, respectively (Table 5). Similar marker and genotypic effect were returned for related traits. This was observed for the microsatellite loci mTcCIR91 (logRWG, logRWY, logTWG; 188/188 low disease incidence, 188/190 high disease incidence) and the SNP loci TcSNP1230 (logRWY, log total witches broom per year; –C low disease incidence, AA high disease incidence).

Discussion

This study described three main traits (NSEEDF, DIPP and WBD) over 483 cacao accessions and identified 7 SSRs and 7 SNPs that could be used to tag these traits (Table 5). The markers that were identified represented the best minimum number of associated markers since correlated markers (LD as multiallelic $r^2 > 0.1$, markers within decay distance) and imprecise markers (less than five observations in effect size) were discarded. Furthermore, complementary datasets were used to generate ancestry files to allow for the testing of every locus under a 1 % filter. A 5 % filter (Ingvarsson and Street 2011; Gupta et al. 2014) or 10 % filter (Li et al. 2011) can be employed but this may increase the incidence of missing data. Hence the chance of discarding a marker due to low counts in effect sizes may be increased. The possibility of missing a true low frequency marker linked to a trait is also increased (Gupta et al. 2014). In this study, a 1 % filter presumably was valid as common significant markers for a given trait were found.

The number of associated markers depended on the dataset, trait and model employed; with GLM having more associated markers than MLM models. The latter has been shown to be more effective in controlling spurious association with false positives that are due to population structure and relatedness (Yu et al. 2006). Using MLM models detected only three markers (NSEED/TcSNP697, TcSNP1370; and logRWG/TcSNP720). Marcano et al. (2009) found nearly twice the number of associated SSRs from a similar number of loci and traits. The higher discovery rate in the aforementioned studies may be explained by their use of a GLM model with *q*-value (Storey 2002) as opposed to Bonferroni testing that was employed in this study. Multiple testing is prone to false negatives due to increased stringency by Bonferroni correction and false positives from lowered stringency of *q*-value application (Gupta et al. 2014). The Bonferroni testing used in this study would have led to the discovery of more conservative markers.

The markers presented in Table 5 were likely true positive trait tags because they were common across datasets; were common across models; population structure was variably accounted for by ancestry, kinship and/or principal coordinate analyses; and one marker was highly

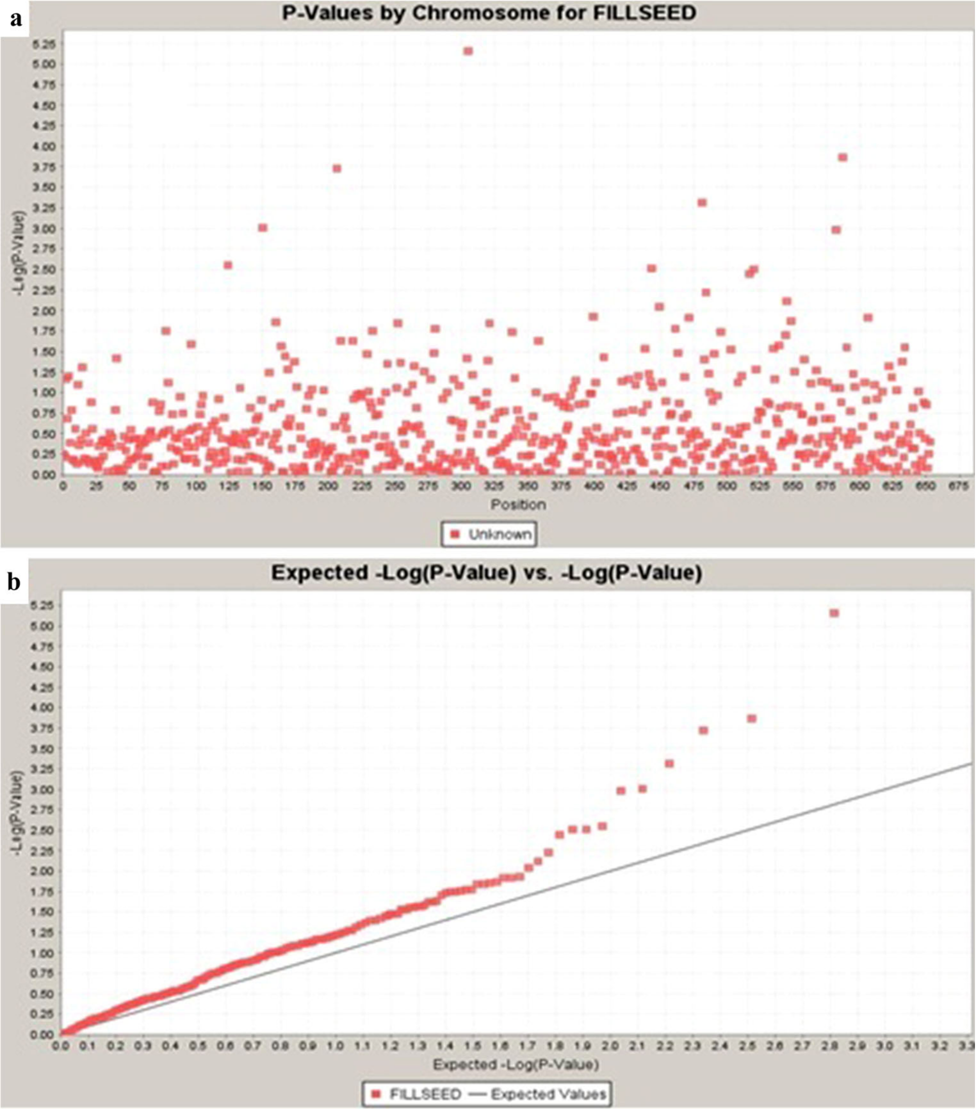

Fig. 2 Manhattan (**a**) and QQ (**b**) plots for SNP markers analysed for number of filled seeds using an MLM2 association mapping approach. MLM2: $Y = E + A + S_{Q}t + S_{K}t$ where Y trait value, E error, A allele information, $S_{Q}t$ population ancestry matrix, $S_{K}t$ kinship matrix. One marker (TcSNP697; highest in each plot) is significantly associated with the number of seeds. Probability estimated at 6.99×10^{-6}. Bonferroni correction for $\alpha = 0.05$ for 654 markers in this dataset $= 7.65 \times 10^{-5}$

significant being tenfold lower than that indicated by Bonferroni adjustment. The reliability of the present results was also supported by the parallel of non-significant correlations of NSEED with disease reactions and lack of common markers for these traits.

In cacao, LD was found to decay within 9.3 cM for chromosomes one to nine. This result was twofold to threefold lower than that of Marcano et al. (2007) but closer to the 15 cM reported by Lanaud (2004). Recently, Stack et al. (2015) estimated LD in cacao at 1–2 megabases in wild groups and at 30 megabases in ‘cultivated’ groups. This transforms to 0.29–0.58 and 8.7 cM, respectively. However, Stack et al. (2015) did not estimate LD by linkage group and this study did not separate accessions into wild and cultivated. Notwithstanding this, the study employed a substantial number of Refractario and Amelonado accessions which would position the LD towards the higher value of, and in congruency with Stack et al. (2015). Plants that are autogamous often exhibit LD over a 1000-fold greater length than allogamous species, and the number of markers needed is substantially more in allogamous plants (Abdurakhmonov and Abdukarimov 2008; Ingvarsson and Street 2011). Association mapping in cacao could be successful in the ICGT with moderate numbers of markers, as the LD was, at most, moderate in size.

The combined phenotypic variance for the various traits, explained by the markers in this study across multiple QTLs for NSEED, was higher than that for blotch resistance in barley (Zhou and Steffenson 2013) but similar to the ranges reported for melon fruit traits (Tomason et al., 2013) and seed size and shape in soybean (Niu et al. 2013). The missing phenotypic variance has been partially attributed to unidentified minor alleles, as well as, to epistatic effects (Zhou and Steffenson 2013).

Corroboration of these marker-trait associations was also obtained from previous admixture, association

Table 5 Selected loci significantly associated with cacao (*Theobroma cacao* L.) phenotypic traits in the International Cocoa Genebank, Trinidad

Trait	Locus[a]	Chrom[b]	Position (cM)	Dataset/model[c]	%Var(P)[d]	Effects	
						Low	High
NSEED	TcSNP1370	1	23.4	F/MLM3	6.45	TT	TG
	mTcCIR66	1	24.8	B_all/GLM2	4.97		284/302
	mTcCIR126	**9**	**9.7**	**B_all/GLM2**	**6.44**		**208/208**
	mTcCIR8	9	48.5	B_all/GLM2	4.76		295/295
	mTcCIR212	9	61.3	B_all/GLM2	4.86		188/192
	TcSNP697	9	65.1	E/MLM4	8.68		CT
	TcSNP368	9	77.1	E/GLM3	4.95		CG
DIPP	mTcCIR280	3	42.30	B1/GLM3	7.53	99/99	91/91
LRWG	TcSNP720	5	77.2	C_all/MLM3	11.01	−T	AA
	mTcCIR91	10	38.9	B2/GLM3	4.88	188/188	188/190
LRWY	TcSNP1230	8	37.9	C_all/GLM2	8.01	AA	AC
	mTcCIR91	10	38.9	B2/GLM3	4.92	188/188	188/190
LVWG	TcSNP375	8	33.2	C_all/GLM2	8.44	A−	GG
LVWY	TcSNP1374	6	39.4	F/GLM3	7.23	AA	GG
LTWG	mTcCIR183	4	24.1	D1/GLM3	4.24	357/357	357/359
	mTcCIR91	10	38.9	D2/GLM3	4.50	188/188	188/190
LTWY	TcSNP1230	8	37.9	E/GLM2	8.24	AA	CC

NSEED number of filled seeds, *DIPP* detached fruit response to spray inoculations with *Phytophthora palmivora*, *LRWG* log reproductive witches' broom per tree girth, *LRWY* log reproductive witches' broom per year, *LVWG* log vegetative witches' broom per tree girth, *LVWY* log vegetative witches' broom per year, *LTWG* log total witches' broom per tree girth, *LTWY* log total witches' broom per year

[a] Entries with $P \leq 5 \times 10^{-5}$ bolded (SSR)

[b] Chromosome and map position from SSR/SNP consensus map of CocoaGenDb (http://cocoagendb.cirad.fr/)

[c] Datasets as in Table, significant markers identified using Bonferroni correction (Bonferroni 1936; Dunn 1961); Models are general linear models (GLM) or mixed linear models (MLM); GLM1: $Y = E + A + S_{Q}t$; GLM2: $Y = E + A + S_{M}t$; GLM3: $Y = E + A + S_{Q}t + S_{M}t$; MLM1: $Y = E + A + S_{K}t$; MLM2: $Y = E + A + S_{Q}t + S_{K}t$; MLM3: $Y = E + A + S_{M}t + S_{K}t$; MLM4: $Y = E + A + S_{Q}t + S_{M}t + S_{K}t$

[d] Percentage of phenotypic variation explained

mapping QTL studies in cacao that identified similar genic regions. The SSR locus mTcCIR212, which was associated with NSEED, was also found to be associated with productivity (Schnell et al. 2005). The number of seeds was significantly associated with mTcCIR8, which was located near to acid phosphatase and mTcCIR124. The former has been linked to compatibility (Warren et al. 1995) and the latter was associated with self-compatibility (Royaert et al. 2011) in cacao. Three genes have been implicated in the compatibility system of cacao (Cope 1962). A major gene for self-compatibility was located on the proximal end of chromosome 4 in an association mapping approach (da Silva et al. 2016) and QTL analyses (Yamada et al. 2010; Royaert et al. 2011). However, although several markers within this area were available, markers associated with NSEED were absent from chromosome 4 in this study. Nevertheless, four unlinked markers for NSEED on chromosomes 1 and 9 were found in this study (Table 5; $r^2 < 0.1$; mTcCIR66/TcSNP1370, mTcCIR126/TcSNP697 and mTcCIR212/TcSNP368 were linked). Three of these were approximately 20 cM away from two additional SNPs that were associated with self-compatibility by da Silva et al. (2016) at 2369,472 bp (chromosome 1; TcSNP1370, mTcCIR66) and 26,855,878 bp (chromosome 9; TcSNP368). On chromosome 1, the marker mTcCIR66 was within 5 and 10 cM of flanking markers, for number of ovules (Clément et al. 2003a, b). Flanking pairs of markers that were not in LD for NSEED (mTcCIR126/mTcCIR8, mTcCIR8/mTcCIR212, TcSNP697/TcSNP368, and mTcCIR212/TcSNP697; all on chromosome 9) were also obtained. The pair mTcCIR212/TcSNP697 had the shortest (~4 cM) flanked distance. These markers, especially those within 20 cM distance of each other, represent good choices of being able to reliably identify superior individuals with large NSEED when used in conjunction with each other.

The absence of common markers that were associated between DIPP and that for WBD suggested that these two diseases may utilise different resistance mechanisms. However, the marker mTcCIR280 for DIPP did not coincide with any of the QTLs for *Phytophthora* resistance reported earlier [Schnell et al. 2007a, b; Lanaud et al. 2009 (and references therein); Kuhn et al. 2012] or with the genes identified by Legavre et al. (2015). Such non-agreement may be the result of the imprecise scoring for DIPP in the screening method (Iwaro et al. 2000) or the differential mechanisms (pre- and post-penetration) of resistance that have been suggested. Inconsistent WBD trait data was similarly obtained, with none of the seven accessions having the highest levels of WBD resistance (CRU 128, IMC 50, LZ 2, MOQ 6/6, PA 67, PA 279, PA 289) being represented in a core collection of the ICGT (Sounigo et al. 2006). Similarly, only 10 of the 25 accessions in common with the study of Thevenin et al. (2005) had similar WBD disease reactions in the ICGT. In contrast to the previous authors, this study utilised WBD data that was collected over a longer time (6 years), which may explain the incongruous results. Furthermore, observations of Thevenin et al. (2005) were made on only three branches, whereas in this study, the entire canopy and trunk of the tree was observed. Recording symptoms over the whole tree for a longer period of time would have reduced the probability of escapes. Trees that had branches with only one or a few brooms, many brooms on one branch, or brooms on spatially distant branches were prone to escape in Thevenin et al. (2005).

In contrast, the markers for resistance to *M. perniciosa* were supported by QTLs reported in the literature and by gene function. The SSR marker mTcCIR91 (chromosome 10) was significantly associated with logRWG, logRWY and logTWG and was adjacent to the NBS-like sequences N172/F12 (1 cM distant) and NSCA6/A1 (2.5 cM distant) of Lanaud et al. (2004). The SSR marker mTcCIR183, significantly associated with LTWG, was 0.3 cM distant from the *Pto*-like sequence PT172/B9 (Lanaud et al. 2004) on chromosome 4. The SNP marker TcSNP1230, significantly associated with logRWY and log total witches' broom per year, had a strong blast hit to the protein disulphide isomerase. Protein disulphide isomerase belongs to the thioredoxin superfamily and is involved in regulatory and stress responses (Lepistö et al. 2009). An in silico study by Lima et al. (2009) also found putative SNPs for WBD linked to thioredoxin among other proteins. None of the SSR tags of Brown et al. (2005) and Santos et al. (2007), and common to this study were retained as associated markers.

In this study, WBD resistance in the ICGT was assessed as two components: vegetative and reproductive. Common markers for these traits were absent, suggesting that different genes may be involved in the resistance mechanisms for witches' broom disease on reproductive and vegetative organs. This has not been previously established in the literature, although Suarez-Capello et al. (2006) noted that some cultivars tended to produce either more vegetative or more cushion brooms. Resistance to WBD was found to be oligogenic in this study and was influenced at six possible areas. This was congruent with the results of Brown et al. (2005); Faleiro et al. (2006) and Leal et al. (2007); Santos et al. (2007); Gesteira et al. (2007); Lima et al. (2008, 2009) and Silva et al. (2014) who found 2, 5, 6, 9, 11 and 4 genic regions, respectively, that were involved in the resistance pathway. Dantas Neto et al. (2005) also postulated an oligogenic mode of resistance to WBD. Although, several authors have suggested a polygenic mode of inheritance (Simmonds 1994; Pires et al. 1999; Faleiro et al. 2002), they may have based their definition of polygenic on the additive effects of several genes rather than the small and equal effects of hundreds of genes. In addition, da Hora Junior et al. (2012) identified 103 upregulated genes, but did not specify whether all 103 were upregulated only in the cultivar resistant to WBD. Overall, it would appear that the current results favour the involvement of an oligogenic system with additive effects in the resistance towards *M. perniciosa*.

Conclusion

Under an association mapping approach, DNA tags that could be utilised in marker assisted breeding of cacao were identified for number of filled seeds (four SSRs and three SNPs), black pod resistance (one SSR) and witches' broom disease resistance (two SSRs and four SNPs) in cacao were identified. The absence of common DNA tags for number of seeds and disease resistance indicate that multiple loci are required to identify cacao plants that pyramid these traits.

Author contribution statement LAM, DZ and PU designed the study. LAM, DZ, SM and LWM were responsible for the SSR data while MB, OF and CL were responsible for the SNP data. LAM collected the field data and performed the statistical analyses. All authors read, reviewed and approved the manuscript.

Acknowledgments This study was partially funded by a grant from the Government of Trinidad and Tobago Research Development Fund and from The UWI-Trinidad and Tobago Research and Development Impact Fund from The University of the West Indies. The use of black pod resistance data from the Cocoa Research Unit (now Cocoa Research Centre) is gratefully acknowledged. The inclusion of some witches' broom disease evaluations from a data set managed by Mrs. Romina Umaharan is gratefully acknowledged. Field evaluations were made possible by several persons including Mr. Leon Ali, Ms. Zainab Ali, Ms. Sarah Bharath, Ms. Rena K. Kalloo, Ms. Carelene Lakhan, Mr. Gangadeen Ramdhanie, Ms. Cassandra Shaw, Mr. Valmiki Singh and Mr. Eusebius Solozano.

References

Abdurakhmonov IY, Abdukarimov A (2008) Application of association mapping to understanding the genetic diversity of plant germplasm resources. Int J Plant Genom. doi:10.1155/2008/574927

Aime MC, Phillips-Mora W (2005) The causal agents of witches' broom and frosty pod rot of cacao (chocolate, *Theobroma cacao*) form a new lineage of Marasmiaceae. Mycologia 97:1012–1022

Allegre M, Argout X, Boccara M et al (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 Res 19:23–35. doi:10.1093/dnares/dsr039

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

Anderson MJ, Ter Braak CJF (2003) Permutations tests for multifactorial analysis of variance. J Stat Comput Simul 73:85–113

Bar-Hen A, Charcosset A, Bourgoin M, Guiard J (1995) Relationship between genetic markers and morphological traits in a maize inbred line collection. Euphytica 84:145–154

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

Bayer C, Fay MF, De Bruijn PY et al (1999) Support for an expanded family concept of Malvaceae within a recircumscribed order Malvales: a combined analysis of plastid *atp*B and *rbc*L DNA sequences. Bot J Linn Soc 129:267–303

Bonferroni CE (1936) Teoria statistica delle classi e calcolo delle probabilità, Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23(19):2633–2635. doi:10.1093/bioinformatics/btm308

Braiser CM, Griffin MJ (1979) Taxonomy of *Phytophthora palmivora* on cocoa. Trans Br Mycol Soc 72:111–143

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 a F2 population using SSR markers and candidate genes. J Am Soc Hort Sci 130:366–373

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

CacaoNet (2012) A global strategy for the conservation and use of cacao genetic resources, as the foundation for a sustainable cocoa economy (B. Laliberté, compiler). Bioversity International, Montpellier

Clément D, Risterucci AM, Motamayor JC, N'Goran J, Lanaud C (2003a) Mapping QTL for yield components, vigor 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

Cleveland WS (1979) Robust locally weighted regression and smoothing scatterplots. J Am Stat Assoc 74:829–836

Cleveland WS, Devlin SJ (1988) Locally weighted regression: an approach to regression analysis by local fitting. J Am Stat Assoc 83:596–610

Collard BCY, Jahufer MZZ, Brouwer JB, Pang ECK (2005) An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. Euphytica 142:169–196. doi:10.1007/s10681-005-1681-5

Cope FW (1962) The mechanism of pollen incompatibility in *Theobroma cacao* L. Heredity 17:157–182

Crouzillat D, Menard B, Mora A, Phillips W, Petiard V (2000) Quantitative trait analysis in *Theobroma cacao* using molecular markers. Yield QTL detection and stability over 15 years. Euphytica 114:13–23

da Hora Junior BT, Poloni JF, Lopes MA et al (2012) Transcriptomics and systems biology analysis in identification of specific pathways involved in cacao resistance and susceptibility to witches' broom disease. Mol BioSyst. doi:10.1039/c2mb05421c

da Silva MR, Clément D, Gramacho KP, Monteiro WR, Argout X, Lanaud C, Lopes U et al (2016) Genome-wide association mapping of sexual incompatibility genes in cacao (*Theobroma cacao* L.). Tree Genet Genom 12:62. doi:10.1007/s11295-016-1012-0

Dantas Neto A, Corrêa RX, Monteiro WR, Luz EDMN, Gramacho KP, Lopes UV (2005) Caracterização de uma população de cacaueiro para mapeamento de genes de resistência à vassoura-de-bruxa e podridãoparda. Fitopatol Bras 30:380–386

Dunn OJ (1961) Multiple comparisons among means. J Am Stat Assoc 56: 52–64 http://sci2s.ugr.es/keel/pdf/algorithm/articulo/1961-Bonferroni_Dunn-JASA.pdf. Accessed 20 May 2016

Eskes AB, Efron Y (2006) Global approaches to cocoa germplasm utilization and conservation. CFC, ICCO, IPGRI, Rome

Faleiro FG, Resende MLV, Niella GR et al (2002) Resistência horizontal/vertical e agressividade/virulência no patossistema *Theobroma cacao* × *Crinipellis perniciosa*. Agrotrópica 14:121–126

Faleiro FG, Quieroz VT, Lopes UV et al (2006) Mapping QTLs for witches' broom (*Crinipellis perniciosa*) resistance in cacao (*Theobroma cacao* L.). Euphytica 149:227–235

Gesteira AS, Micheli F, Carels N et al (2007) Comparative analysis of expressed genes from cacao meristems infected by *Moniliophthora perniciosa*. Ann Bot 100:129–140. doi:10.1093/aob/mcm092

Gupta PK, Kulwal PL, Jaiswal V (2014) Association mapping in crop plants: opportunities and challenges. Adv Genet 85:109–147

Hill WG, Weir BS (1994) Maximum-likelihood estimation of gene location by linkage disequilibrium. Am J Hum Genet 54:705–714

Hopkins WG (2002) A scale of magnitude for effect statistics. In: A new view of statistics. http://www.sportsci.org/resource/stats/index.html. Accessed 20 May 2016

Hospital F (2003) Marker-assisted breeding. In: Newbury HJ (ed) Plant molecular breeding. Blackwell, Oxford, pp 30–59

Ingvarsson PK, Street NR (2011) Association genetics of complex traits in plants. New Phytol 189:909–922. doi:10.1111/j.1469-8137.2010.03593.x

Iwaro AD, Sreenivasan TN, Butler DR, Umaharan P (2000) Rapid screening for *Phytophthora* pod rot resistance by means of detached pod inoculation. In: Eskes AB, Engels JMM, Lass RA (eds) Working procedures for cocoa germplasm evaluation and selection. Proceedings of the CFC/ICCO/IPGRI project workshop, 1–6 Feb 1998. International Plant Genetic Resources Institute, Rome, pp 109–113

Kennedy AJ, Lockwood G, Mossu G, Simmonds NW, Tan GY (1987) Cocoa breeding: past, present and future. Cocoa Grow Bull 38:5–22

Kuhn DV, Livingstone D III, Main D et al (2012) Identification and mapping of conserved ortholog set (COS) II sequences of cacao and their conversion to SNP markers for marker-assisted selection in *Theobroma cacao* and comparative genomics studies. Tree Genet Genom 8:97–111. doi:10.1007/s11295-011-0424-0

Lanaud C (2004) Genetic origin and movement of cocoa varieties: consequences on genome analyses and modern breeding strategies. American Association for the Advancement of Science Annual Conference, 14 Feb, Seattle, Washington

Lanaud C, Risterucci A-M, Pieretti I, N'Goran JAK, Fargeas D (2004) Characterisation and genetic mapping of resistance and defence gene analogs in cocoa (*Theobroma cacao* L.). Mol Breed 13:211–227

Lanaud C, Fouet O, Clement D et al (2009) A meta-QTL analysis of disease resistance traits of *Theobroma cacao* L. Mol Breed 24:361–374. doi:10.1007/s11032-009-9297-4

Leal GA, Albuquerque PSB, Figueira A (2007) Genes differentially expressed in *Theobroma cacao* associated with resistance to witches' broom disease caused by *Crinipellis perniciosa*. Mol Plant Pathol 8(3):279–292

Legavre T, Ducamp M, Sabau X et al (2015) Identification of *Theobroma cacao* genes differentially expressed during *Phytophthora megakarya* infection. Physiol Mol Plant Pathol 92:1–13. doi:10.1016/j.pmpp.2015.08.005

Lepistö A, Kangasjärvi S, Luomala E-M et al (2009) Chloroplast NADPH-thioredoxin reductase interacts with photoperiodic development in *Arabidopsis*. Plant Physiol 149:1261–1276

Li X, Wei Y, Moore KJ et al (2011) Association mapping of biomass yield and stem composition in a tetraploid alfalfa breeding population. Plant Genome 4:24–35. doi:10.3835/plantgenome2010.09.0022

Lima LS, Gramacho KP, Gesteira AS et al (2008) Characterization of microsatellites from cacao–*Moniliophthora perniciosa* interaction expressed sequence tags. Mol Breed 22:315–318. doi:10.1007/s11032-008-9174-6

Lima LS, Gramacho KP, Carels N et al (2009) Single nucleotide polymorphisms from *Theobroma cacao* expressed sequence tags associated with witches' broom disease in cacao. Genet Mol Res 8(3):799–808

Liu K, Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21:2128–2129. doi: 10.1093/bioinformatics/bti282, V 3.25 originally at http://www.powermarker.net. Accessed 7 Nov 2012. Now hosted at http://statgen.ncsu.edu/powermarker/

Lopes UV, Monteiro WR, Pires JL, Clement D, Yamada MM, Gramacho KP (2011) Cacao breeding in Bahia, Brazil—strategies and results. Crop Breed Appl Biot S1:73–81

Maharaj K, Maharaj P, Bekele FL et al (2011) Trinidad selected hybrids: an investigation of the phenotypic and agro-economic traits of 20 selected cacao cultivars. Trop Agric (Trinidad) 88:175–185

Marcano M, Pugh T, Cros E et al (2007) Adding value to cocoa (*Theobroma cacao* L.) germplasm information with domestication history and admixture mapping. Theor Appl Genet 114:877–884

Marcano M, Morales S, Hoyer MT et al (2009) A genomewide admixture mapping study for yield factors and morphological traits in a cultivated cocoa (*Theobroma cacao* L.) population. Tree Genet Genom 5:329–337. doi:10.1007/s11295-008-0185-6

MedCalc Statistical Software (2013) MedCalc Software beta version 12.7.7. http://www.medcalc.org. Accessed 20 May 2016

Motamayor JC, Lachneaud P, da Silva e Mota JW et al (2008) Geographic and genetic population differentiation of the Amazonian chocolate tree (*Theobroma cacao* L). PLoS One 3(10):e3311. doi:10.1371/journal.pone.0003311

Motamayor JC, Mockaitis K, Schmutz J et al (2013) The genome sequence of the most widely cultivated cacao type and its use to identify candidate genes regulating pod colour. Genome Biol 14:r53. doi:10.1186/gb-2013-14-6-r53

Motilal LA, Zhang D, Umaharan P, Mischke S, Boccara M, Pinney S (2009) Increasing accuracy and throughput in large-scale microsatellite fingerprinting of cacao field germplasm collections. Trop Plant Biol 2:23–37. doi:10.1007/s12042-008-9016-z

Motilal LA, Zhang D, Umaharan P et al (2012) Elucidation of genetic identity and population structure of cacao germplasm within an international cacao genebank. Plant Genet Resour 10:232–241. doi:10.1017/S1479262112000305

Motilal LA, Zhang D, Mischke S, Meinhardt LW, Umaharan P (2013) Microsatellite-aided detection of genetic redundancy improves management of the International Cocoa Genebank, Trinidad. Tree Genet Genom 9:1395–1411. doi:10.1007/s11295-013-0645-5

Niu Y, Xu Y, Liu X-F et al (2013) Association mapping for seed size and shape traits in soybean cultivars. Mol Breed 31:785–794. doi:10.1007/s11032-012-9833-5

Peltier J (2009) PTS LOESS calculator. A LOESS utility for Excel. http://peltiertech.com/loess-utility-awesome-update/. Accessed 20 May 2016

Pires JL, Monteiro WR, Luz EDMN et al (1999) Cocoa breeding for witches' broom resistance at CEPEC, Bahia, Brazil. In: Proc. Intl. Workshop on the Contribution of Disease Resistance to Cocoa Variety Improvement, Salvador, Bahia, Brazil. Ingenic, UK, pp 91–101

Pritchard J (2001) Deconstructing maize population structure. Nat Genet 28:203–204

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

Pugh T (2005) Etude des déséquilibres de liaison dans une collection de cacaoyers (*Theobroma cacao* L.) appartenant au groupe Criollo/Trinitario et application au marquage génétique des caractères d'intérêt. Soutenue le 11 janvier 2005. ENSAM—ED biologie intégrative, Montpellier

Royaert S, Phillips-Mora W, Leal AMA et al (2011) Identification of marker-trait associations for self-compatibility in a segregating mapping population of *Theobroma cacao* L. Tree Genet Genomes 7:1159–1168. doi:10.1007/s11295-011-0403-5

Santos RMF, Lopes UV, Bahia RC, Machado RCR, Ahnert D, Corrêa RX (2007) Marcadores microssatélites relacionados com a resistência à vassoura-de-bruxa do cacaueiro. Pesqui Agropecu Bras 42:1137–1142

Schnell RJ, Olano CT, Brown JS et al (2005) Retrospective determination of the parental population of superior cacao (*Theobroma cacao* L.) seedlings and association of microsatellite alleles with productivity. J Am Soc Hortic Sci 130:181–190

Schnell RJ, Brown JS, Kuhn DN et al (2007a) Current challenges of tropical tree crop improvement: integrating genomics into an applied cacao breeding program. Acta Hortic 738:129–144

Schnell RJ, Kuhn DN, Brown JS et al (2007b) Development of a marker assisted selection program for cacao. Phytopathology 97:1664–1669

Searle SR (1987) Linear models for unbalanced data. Wiley, New York

Shriner D (2013) Overview of admixture mapping. Curr Protoc Hum Genet. doi:10.1002/0471142905.hg0123s76

Silva DV, Araújo IS, Branco SMJ et al (2014) Analysis of resistance to witches' broom disease (*Moniliophthora perniciosa*) in flower cushions of *Theobroma cacao* in a segregating population. Plant Pathol 63:1264–1271. doi:10.1111/ppa.12204

Simmonds NW (1994) Horizontal resistance to cocoa diseases. Cocoa Grow Bull 47:42–53

Sounigo O, Bekele FL, Iwaro AD, et al (2006) Description of cocoa clones proposed for the "CFC/ICCO/IPGRI Project Collection". In: Eskes AB, Efron Y (eds) Global approaches to cocoa germplasm utilization and conservation. Final report of the CFC/ICCO/IPGRI project on "Cocoa Germplasm Utilization and Conservation: a Global Approach" (1998–2004). CFC, Amsterdam; ICCO, UK; IPGRI, Rome, pp 67–81

Stack JC, Royaert S, Gutiérrez O, Nagai C, Holanda ISA, Schnell R, Motamayor JC (2015) Assessing microsatellite linkage disequilibrium in wild, cultivated, and mapping populations of *Theobroma cacao* L. and its impact on association mapping. Tree Genet Genomes 11:19. doi:10.1007/s11295-015-0839-0

Storey JD (2002) A direct approach to false discovery rates. J R Stat Soc Ser B Stat Methodol 64:479–498

Suarez-Capello C, Delgado R, del Pozo P et al (2006) Witches' broom resistance screening of seedlings and clones in Ecuador: some comparisons between natural infection in the field and artificial inoculations. In: Eskes AB, Efron Y (eds) Global approaches to cocoa germplasm utilization and conservation. Final report of the CFC/ICCO/IPGRI project on "Cocoa Germplasm Utilization and Conservation: a Global Approach" (1998–2004). CFC, Amsterdam; ICCO, UK; IPGRI, Rome, pp 151–157

Thevenin JM, Umaharan R, Surujdeo-Maharaj S, Latchman B, Cilas C, Butler DR (2005) Relationships between black pod and witches'-broom diseases in *Theobroma cacao*. Phytopathology 95:1301–1307

Thomas E, van Zonneveld M, Loo J, Hodgkin T, Galluzzi G et al (2012) Present spatial diversity patterns of *Theobroma cacao* L. in the Neotropics reflect genetic differentiation in Pleistocene refugia followed by human-influenced dispersal. PLoS One 7(10):e47676. doi:10.1371/journal.pone.0047676

Tomason Y, Nimmakayala P, Levi A, Reddy UK (2013) Map-based molecular diversity, linkage disequilibrium and association mapping of fruit traits in melon. Mol Breed 31:829–841. doi:10.1007/s11032-013-9837-9

Toxopeus H (1985) Planting material. In: Wood GAR, Lass RA (eds) Cocoa, 4th edn. Longman, London, pp 80–92

Turnbull CJ, Hadley P (2015) International Cocoa Germplasm Database (ICGD). (Online Database). CRA Ltd./NYSE Liffe/ University of Reading, UK. http://www.icgd.reading.ac.uk. Accessed 20 May 2016

Warren J, Misir S, Kalai (1995) Isozyme markers for self-compatibility and yield in *Theobroma cacao* (cacao). Heredity 74:354–356

Yamada MM, Faleiro FG, Clement D, Lopes UV, Pires JL, Melo GRP (2010) Relationship between molecular markers and incompatibility in *Theobroma cacao* L. Agrótropica 22:71–74

Yu J, Pressoir G, Briggs WH et al (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38:203–208. doi:10.1038/ng1702

Zaykin D, Zhivotovsky L, Weir BS (1995) Exact tests for association between alleles at arbitrary numbers of loci. Genetica 96:169–178

Zhang D, Boccara M, Motilal L et al (2008) Microsatellite variation and population structure in the "Refractario" cacao of Ecuador. Conserv Genet 9:327–337. doi:10.1007/s10592-007-9345-8

Zhang Z, Ersoz E, Lai C-Q et al (2010) Mixed linear model approach adapted for genome wide association studies. Nat Genet 42:355–360. doi:10.1038/ng.546

Zhang D, Figueira A, Motilal L, Lachenaud P, Meinhardt LW (2011) Theobroma. In: Kole C (ed) Wild crop relatives: genomic and breeding resources. Plantation and ornamental crops. Springer, Berlin, pp 277–296

Zhou H, Steffenson B (2013) Genome-wide association mapping reveals genetic architecture of durable spot blotch resistance in US barley breeding germplasm. Mol Breed 32:139–154. doi:10.1007/s11032-013-9858-4

Zhu C, Gore M, Buckler ES, Yu J (2008) Status and prospects of association mapping in plants. Plant Genome 1:5–20. doi:10.3835/plantgenome2008.02.0089