Theor Appl Genet (1996) 93: 205–214 

D. Crouzillat · E. Lerceteau · V. Petiard · J. Morera
H. Rodriguez · D. Walker · W. Phillips · C. Ronning
R. Schnell · J. Osei · P. Fritz

# *Theobroma cacao* L.: a genetic linkage map and quantitative trait loci analysis

Received: 25 November 1995 / Accepted: 2 February 1996

**Abstract** A genetic linkage map of *Theobroma cacao* (cocoa) has been constructed from 131 backcross trees derived from a cross between a single tree of the variety Catongo and an F1 tree from the cross of Catongo by Pound 12. The map comprises 138 markers: 104 RAPD loci, 32 RFLP loci and two morphologic loci. Ten linkage groups were found which cover 1068 centimorgans (cM). Only six (4%) molecular-marker loci show a significant deviation from the expected 1:1 segregation ratio.The average distance between two adjacent markers is 8.3 cM. The final genome-size estimates based on two-point linkage data ranged from 1078 to 1112 cM for the cocoa genome. This backcross progeny segregates for two apparently single gene loci controlling (1) anthocyanidin synthesis (*Anth*) in seeds, leaves and flowers and (2) self-compatibility (*Autoc*). The *Anth* locus was found to be 25 cM from *Autoc* and two molecular markers co-segregate with *Anth*. The genetic linkage map was used to localize QTLs for early flowering, trunk diameter, jorquette height and ovule number in the $BC_1$ generation using both single-point ANOVA and interval mapping. A minimum number of 2–4 QTLs ($P<0.01$) involved in the genetic expression of the traits studied was detected. Coincident map locations of a QTL for jorquette height and trunk diameter suggests the possibility of pleiotropic effects in cocoa for these traits. The combined estimated effects of the different mapped QTLs explained between 11.2% and 25.8% of the phenotypic variance observed in the $BC_1$ population.

**Key words** Heterozygosity · Molecular markers · Genetic map · Quantitative trait loci · *Theobroma cacao*

Communicated by J. W. Snape

D. Crouzillat (✉) · E. Lerceteau · V. Petiard
Centre de Recherche Nestlé Tours, 101 Avenue Gustave Eiffel, Notre Dame-d'Oé, B.P. 9716, 37097 Tours Cedex 2

J. Morera · H. Rodriguez · D. Walker · W. Phillips · P. Fritz
Centro agronomico tropical de investigacion y ensenanza, CATIE – 7170, Turrialba, Costa Rica

C. Ronning · R. Schnell
United States Department of Agriculture Subtropical Horticulture Research Station, 13601 Old Cutler Road, Miami, FL 33158, USA

J. Osei[1] · P. Fritz
Penn State University, Department of Food Science and the ACRI cocoa molecular biology laboratory. University Park, PA 16802, USA

[1] *Present address*:
University of Ghana, Agriculture Research Station Kade, Ghana

## Introduction

*Theobroma cacao* is one of the major cash crops for a number of developing countries.The annual value of the cocoa crop worldwide has been estimated at 3.6 billion US dollars (Anonymous 1995).The centre of cocoa origin was proposed by Cheesman (1944) to be on the lower eastern equatorial slopes of the Andes. Even though this center of diversity has been questioned (Allen 1988), this area of Peru is considered as a high priority for future genetic conservation (Lockwood and End 1992).

Among the cocoa species used for cultivation, three major types can be differentiated: the Forasteros, the Criollos and the Trinitarios, which are considered as recent hybrids between the Forasteros and the Criollos. This classification is based mainly on morphological and agronomical characters (Engels1992, Bekele 1994). Although genetic diversity exists in cocoa, the genotypes currently used in cocoa production are from a small part of this diversity (Cope 1976) and only a few breeding programs have been developed from this narrow genetic base (Lockwood 1985).

To select for fruit-specific characters in cocoa breeding programs, trees must be maintained until fruiting begins, usually after 2–3 years. This process is both expensive and time-consuming. The ability to select very early in plant development, at the genotype rather than at the phenotype level, could save time and the expense of cultivating trees lacking the traits of interest. The efficiency of breeding

programs could be improved if molecular-based maps were available and markers associated with major QTLs for quality and productivity could be identified. Moreover, the use of QTLs in breeding would permit the differentiation of trees having the same phenotype for a quantitative trait but determined by different genotypes. In this way, the management of additive effects for such a trait can be optimized.

Cocoa is a diploid, mainly cross-pollinated, species (2n=20) and has a small nuclear genome. Using reassociation kinetics, Couch et al. (1993) estimated the size of the haploid cocoa genome to be approximately $2 \times 10^8$ bp. By flow cytometry Earle (1991, personal communication) and Lanaud et al. (1992) reported values of $4.15 \times 10^8$ and $3.88 \times 10^8$ bp respectively. These studies indicate that the cocoa genome is approximately two-times larger than that of *Arabidopsis thaliana*, the smallest plant genome yet reported (Arumuganathan and Earle 1991).

Previously, several reports on genetic variability within *T. cacao* indicated sufficient polymorphism at the isozyme level (Lanaud 1986, Ronning and Schnell 1994, Warren 1994) and at the DNA level with RAPD (Russel et al. 1993; Figueira et al. 1994; N'goran et al. 1994) or RFLP analysis (Laurent et al. 1994) to differentiate and classify cocoa populations. This work demonstrated that even though the level of variability was low, probably due to the narrow genetic base of cocoa germplasm, it is sufficient for phenotypic, RFLP and RAPD mapping.

The goal of the research reported herein was to develop a saturated molecular linkage map of cocoa, and to begin using this map in cocoa breeding experiments, especially for evaluating genetic variation and manipulating production-related QTLs. We report the development of a cocoa molecular linkage map and the identification of QTLs for key traits in a $BC_1$ population.

## Materials and methods

### Mapping population

The map was generated from the analysis of 131 $BC_1$ trees derived from the intraspecific cross of two vegetatively propagated cocoa trees: Catongo and Pound 12.

Catongo is a highly homozygous, self-compatible, white-seeded and white-flowered, lower Amazon Forastero, found near Urucuca, Bahia, Brazil, in 1939 (Wood and Lass 1985). Pound 12 is one of the Nanay origin clones collected as budwood near the headwaters of the Amazon by Pound in 1943. Pound 12 is a self-incompatible, purple-seeded, purple-flowered, upper Amazon Forastero.

A single $F_1$ plant was used as the male parent in a backcross to Catongo (Morera et al. 1991). The 131 $BC_1$ trees were established in 1991 in a 0.25-hectare perennial orchard at CATIE (Centro Agronomico Tropical de Investigacion y Ensenanza) in Turrialba, Costa Rica. The field contained drainage ditches to avoid stress from water logging. Mixed shade trees [guava (*Inga*), poro (*Erythrina*), and banana (*Musa*)] were also included in the orchard. Agronomic management included chupon control, pruning to manage tree architecture, and fertilization according to a yearly schedule. Some of the young trees developed "Mal Rosado" (Pink disease) caused by *Corticum salmonicolor* which was controlled by localized spraying with 0.25% copper sulphate.

### Plant DNA extraction

Total genomic DNA was extracted from fresh green mature cocoa leaves. Samples of approximately 25 g of leaves, ground in liquid nitrogen, were homogenized using a polytron for 1 min in 200 ml of extraction buffer at 4°C (0.35 M Sorbitol, 0.1 M Tris-HCl pH 7.5, 0.005 M EDTA) with sodium bisulphite (1 g) as an antioxidant. After filtration through Miracloth the filtrate was centrifuged (20 min, 1000 *g*, 4°C) and the nuclei-enriched pellet re-suspended in 100 ml of extraction buffer. After centrifugation (20 min, 1000 *g*, 4°C), the pellet was re-suspended in 10 ml of extraction buffer with 14 ml of nuclei lysis buffer [0.2 M Tris-HCl pH 8 , 0.005 M EDTA, 2 M NaCl, 2% CTAB (w/v)] with 4.8 ml of 5% Sarkosyl (w/v) and incubated for 60 min at 65°C. The lysate was then extracted with chloroform/isoamyl alcohol (24/1: v/v) and the DNA in the aqueous phase was precipitated with 1 volume of isopropanol overnight at –20°C. The DNA was re-suspended in 5 ml of TE buffer (10 mM Tris-HCl pH 8, 0.1 mM EDTA) and purified with RNAase A (final concentration 200 µg/ml) at 37°C for 30 min, and proteinase K (final concentration 400 µg/ml) at 56°C for 1 h. Proteins were removed with phenol/chloroform/isoamyl alcohol (25/24/1: v/v/v) and the DNA was precipitated with a mixture of 3 M sodium acetate pH 5.2 and cold ethanol (0.1/2: v/v, –20°C). The DNA was hooked on a glass rod, washed in 70% ethanol solution and re-suspended in 400 µl of TE buffer. The purified DNA was quantified at 260 nm using a spectrophotometer.

### Random amplified polymorphic DNA (RAPD) markers

Procedures for RAPD generally follow those described by Williams et al. (1990); conditions were harmonized between the three different laboratories involved in the mapping project (Centre de recherche Nestlé Tours; CATIE; USDA agricultural research station at Miami). Standard RAPD analyses were performed in a reaction volume of 50 µl prepared with 35 ng of cocoa genomic DNA, 25 ng of primer [10-mer oligonucleotides from Operon Technologies, Alameda, California, or from University of British Columbia, except for primers C95 and C96 from cocoa chitinase gene (Snyder 1994)], 100 µM of each dNTP and 2U of *Taq* polymerase (Stratagene or Perkin Elmer-Cetus) made up to a final concentration of 1.5 mM magnesium chloride, and overlaid with 50 µl of mineral oil.

Amplifications were performed in a Braun 60/2 thermocycler programmed for a preliminary 2-min denaturation at 94°C, and 45 cycles including 1 min at 94°C, 1 min at 37°C and 2 min at 72°C, and a final DNA extension cycle at 72°C for 7 min. DNA fragments generated by amplification were separated on a 1.4% agarose gel [0.7% nusieve GTC agarose (FMC) and 0.7% agarose MP (Boerhinger)] followed by staining with ethidium bromide.

Primers were screened against both backcross parents (Catongo and the $F_1$ tree), Pound 12, and six individual backcross trees. When a RAPD band was present in only one parent and in at least one of the six $BC_1$ trees, the parent was classified as potentially heterozygous for that locus (referred to as testcross loci). Bands which were present in only one parent and all the progeny, tentatively classified each parent as homozygous for alternate alleles (referred to as non-segregating loci). A subset of primers that maximized the number of polymorphisms in a testcross configuration was selected and segregation scored in the 131 $BC_1$ progeny. In the case of testcross loci, the presence of a band was scored as "H" (heterozygous) while absence was scored as "A" (homozygous). Those cases in which the presence or absence of DNA bands was unclear were recorded as missing data.

### Restriction fragment length polymorphism DNA (RFLP) markers

Genomic *Pst*I and cDNA clones were screened on Southern blots of parental DNA (Catongo, Pound 12, and the $F_1$ tree) restricted with *Eco*RI, *Eco*RV, *Hin*dIII, *Bgl*II or *Sca*I to identify the best combination of probe and endonuclease digest that detected polymorphism.

Three probes, for the chalcone synthase gene (CHS 600), for the acyl carrier protein gene (ACP 600) and for the cocoa seed protease inhibitor gene (CSP), are from cocoa DNA amplification via PCR using two gene consensus primers obtained from the GCG database (Genetics Computer Group, USA) or in the case of the CSP gene from Spencer and Hodge (1991), Tai et al. (1991). The primers were respectively CATGATGTACCAiCAiGGiTGCTT and CTiGACATGTTiCCATACTC for CHS 600 and AiGAGACiiTiGAGAAAGTiT and CCCATiACiATCTCAACC for ACP 600. A single amplified DNA fragment of 600 bp was obtained in both cases and used as a RFLP probe. The two CSP primers used were ATGAAGACCGCAACAGCCGTAG and TAAAAGAGGCAAGAGGCAAATCA they allowed the amplification of a single DNA fragment of 900 bp.

Five micrograms of DNA were restricted overnight with the appropriate restriction enzyme (10 U/µg) according to the supplier's recommendations. Samples were loaded onto 0.7% agarose gels in TBE buffer. Southern blots were made by alkaline capillary transfer to nylon membranes (Appligène). The DNA probes were labelled by using $^{32}P$ with the Megaprime kit (Amersham) and hybridized on the membranes overnight at 65°C in hybridization buffer (5% w/v SDS, 5SSC, 5Denhardt's solution, 40 µg/ml of heterologous DNA). Post-hybridization treatments consisted of three high-stringency washes (2SSC, 0.1% SDS; 1SSC, 0.1% SDS; 0.2SSC, 0.1% SDS) each at 65°C for 30 min.

Due to the co-dominant nature of RFLP markers the heterozygous status was scored for Catongo, Pound 12, and the $F_1$ tree according to the different RFLP allelic forms detected for each locus studied. All the RFLP probes useful for $BC_1$ mapping detected a test-cross configuration and was scored ("A" or "H") according to the previous RAPD marker analysis.

### Phenotypic markers

The anthocyanidin trait (red leaf flush and red flower staminodes versus pale green flush leaves and white staminodes) was recorded for each $BC_1$ tree.

The self-compatibility status was determined by controlled self pollination of at least ten isolated cocoa flowers per tree. After 10 days the percentage of successful pollination was recorded.

### Analysis of segregation data

Each marker was tested against the expected segregation ratio using a $\chi^2$ goodness of fit and the LINKAGE 1 program (Suiter et al. 1983).

Linkage analysis was performed on a microVAX 3100 using the program MAPMAKER 2.0 (Lander et al. 1987). Kosambi's mapping function (Kosambi 1944) was used to convert the recombination fraction into map distance. The pairwise analysis obtained with MAPMAKER agreed with LINKAGE 1 data and was also used to assign markers to linkage groups with the following criteria – recombination frequency<0.3 and a (LOD)>3.0.

### Estimation of cocoa genetic map length

The genetic map length (G) was estimated from partial linkage data according to Hulbert et al. (1988). This method requires knowledge of the number of locus pairs linked at predetermined LOD scores. The Kosambi function was used to compare the theoretical and experimental cocoa genome size.

### Quantitative trait analysis

Individual $BC_1$ trees were monitored for trunk diameter (TD), height of the first jorquette (JH), days to first flower since field planting of seedling (EF), and the number of ovules. Ovules from ten flowers from each $BC_1$ tree were counted under a dissecting microscope in 1994, when the trees were 2 years old. Each trait was tested for normal distribution using Statgraphics software.

A one-way ANOVA using single markers as treatments was used to test for an association between markers and quantitative traits in the $BC_1$. A $P<0.01$ value was used as the threshold for considering the likely presence of a QTL near a marker. The proportion of the total phenotypic variance attributable to each QTL was estimated via linear regression using Statgraphics software.

The total percentage of phenotypic variance explained for each trait by all significant QTLs was calculated with a multiple regression analysis, using the trait as a dependent variable and the previously identified markers, linked to a QTL, as treatments.

Genotypes from the linkage map and quantitative data for each trait were used for input into MAPMAKER/QTL v. 0.9 software to compare with ANOVA results. The $BC_1$ data were analyzed separately according to the trait, using the method of interval mapping (Lander and Botstein 1989) and LOD scores. A LOD score threshold of 1.5 (equivalent to a $P$ value of 0.01 from the ANOVA study) was used to declare the presence of a linked QTL in the interval. With this stringency, and given the number of markers used, a per-chromosome false positive rate of 5% was expected, as estimated numerically by Darvasi et al. (1993).

## Results

### RFLP

Ninety seven of the four-hundred and forty RFLP-assayed probes (22%) revealed polymorphism between Catongo and Pound 12 with at least one of the five restriction enzymes tested. The average polymorphism between the two parents on a per-enzyme basis was 7.5%. The polymorphic RFLPs were mainly detected as single locus types and 66 (68%) could be mapped at the backcross population level using the five restriction enzymes (Fig. 1A). The average number of RFLP allelic forms determined on Catongo and Pound 12 was estimated to be about 2.04 with the 97 RFLP probes and the five endonucleases employed.

A RFLP heterozygosity estimation (Fig. 1A) was obtained with these 97 RFLP probes. Catongo gave the lowest heterozygosity rate (2%). Pound 12 and the $F_1$ tree, with 81% and 66% respectively, appeared as highly heterozygous clones.

### RAPD

Twenty four percent (135/555) of the RAPD primers detected polymorphism between Catongo and Pound 12. Each primer generated between one to nine scorable bands with an average of 3.6. The mean polymorphism per locus was 7.4%.

The RAPD heterozygosity evaluation is more difficult than the RFLP evaluation due to the dominant character of the DNA amplification of RAPDs, and some doubt occurs about the genotypic determination of Catongo and Pound 12 for the differentiation of heterozygous and homozygous loci (Fig. 1B). In all cases, the exact genotype of the $F_1$ tree could be determined, and was estimated to be about 68% heterozygous.

In spite of the dominant inheritance of these molecular markers sufficient information could be obtained from the backcross mapping population to determine the genotype

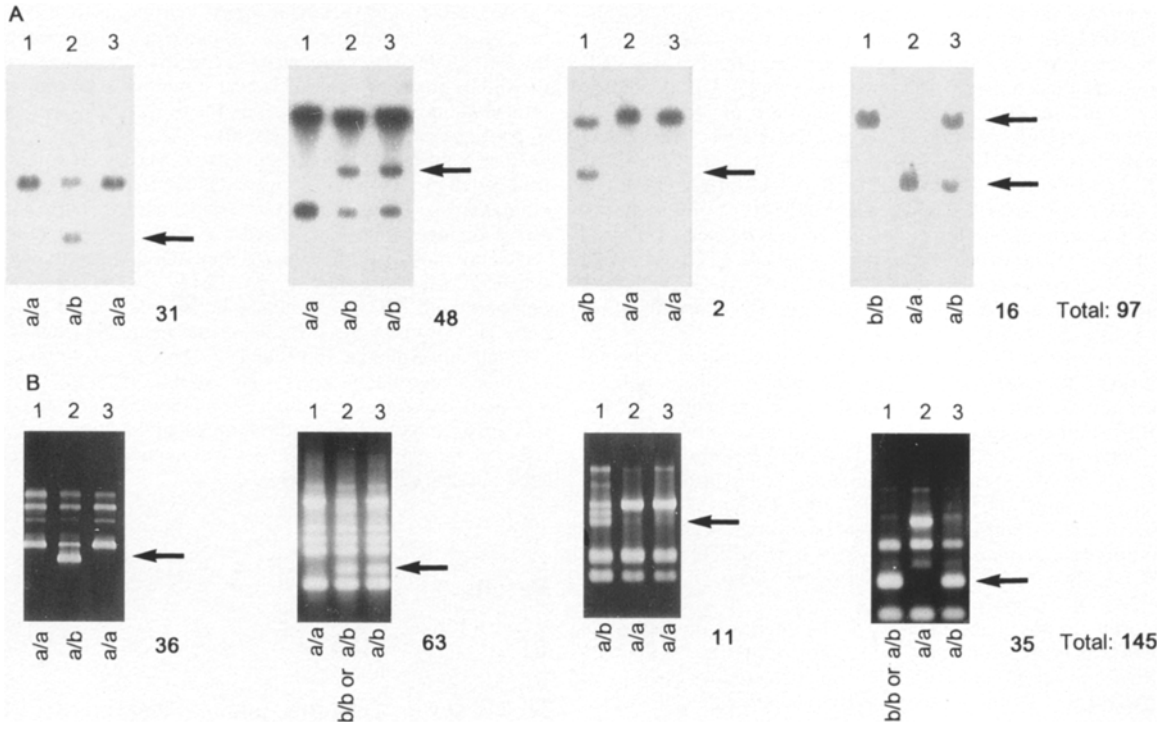


**Fig. 1A,B** RFLP and RAPD heterozygosity estimations on Catongo, Pound12 and the $F_1$ tree used for the backcross. **A** Four RFLP patterns: cDNA probe CCC3, CCG906 *PstI* genomic probe, CCG1082 *PstI* genomic probe, CCG888 *PstI* genomic probe. **B** Four RAPD patterns: AS05 primer, AN11 primer, AK17 primer, primer AN19. The RFLP or RAPD allelic status for Catongo (*1*), Pound 12 (*2*) and the $F_1$ (*3*) is indicated when known. The *arrows* point out the polymorphism of each RFLP or RAPD marker. The total number of RFLP or RAPD loci studied is shown at the right of the figure

of the Catongo parent from the segregating ratios of the 104 RAPD markers mapped in this progeny. Two of the one-hundred and four RAPD loci mapped (2%) were heterozygous in Catongo. This RAPD heterozygosity estimation is close to the RFLP value and suggests that Catongo is a highly homozygous cocoa genotype. The RAPD heterozygosity of Pound 12 was estimated on 55 $F_1$ plants obtained by crossing Catongo and Pound 12. Of 15 RAPD loci tested 9 (60%) are heterozygous (data not shown).

## Segregation of markers

Of a total of 138 markers mapped, 104 (75%) were RAPD markers. Two examples of RAPD segregation are shown in Fig. 2. Only four markers (B522, C96, AC01 and U510) show significant deviation from the expected 1:1 segregating ratio ($\chi^2$ respectively of 10.95 $P<0.001$, 10.05 $P<0.005$, 4.52 $P<0.05$ and 4.5 $P<0.05$) with a skewing ratio in favor of Pound-12 alleles, except for the C96 RAPD locus. Ten primers each revealed two different polymorphic bands segregating independently and mapping to different linkage groups.

Only two restriction enzymes (*Hin*dIII and *Eco*RV) were used for mapping. *Hin*dIII showed that 22 (5%) of the 440 probes are polymorphic for Catongo and the $F_1$, while *Eco*RV indicated 18 (4.1%) of them as polymorphic. Using these two enzymes, 32 (7.3%) of the probes revealed RFLP variability. Among the 440 probes, eight were polymorphic with both *Eco*RV and *Hin*dIII. These results show that there was no independence between the two restriction enzymes regarding the RFLP variability they detected ($\chi^2=52$, $P<0.001$) .

The 32 RFLP loci mapped gave the expected segregating ratio (1:1) except for CCG978 and CCG1550 ($\chi^2$ of 4.04 $P<0.05$), with a skewing in favor of the Pound-12 RFLP allele for the two markers.

Absence of anthocyanidin synthesis in leaves, seeds and flowers observed on the Catongo clone did not segregate in the 55 $F_1$ hybrids. The backcross progeny shows a 1:1 segregation ratio ($\chi^2=0$). These results are compatible with a single gene homozygous recessive for Catongo (pale green flush leaves and flowers), homozygous dominant for Pound 12 (red flush leaves and flowers) and heterozygous for the $F_1$ tree used for the backcross with Catongo. Hereafter we refer to this locus as *Anth.*

Catongo is self-compatible while Pound 12 is self-incompatible. Of 47 $F_1$ hybrids tested, 22 were self-compatible and 25 self-incompatible. These data are in agreement with the self-incompatible S system. They could be explained by a single gene locus with homozygous status (Sx/Sx) for Catongo and heterozygous (Sx/Sy) for Pound 12 with the Sy allele dominant with respect to the Sx allele. Hereafter we refer to this locus as *Autoc.*

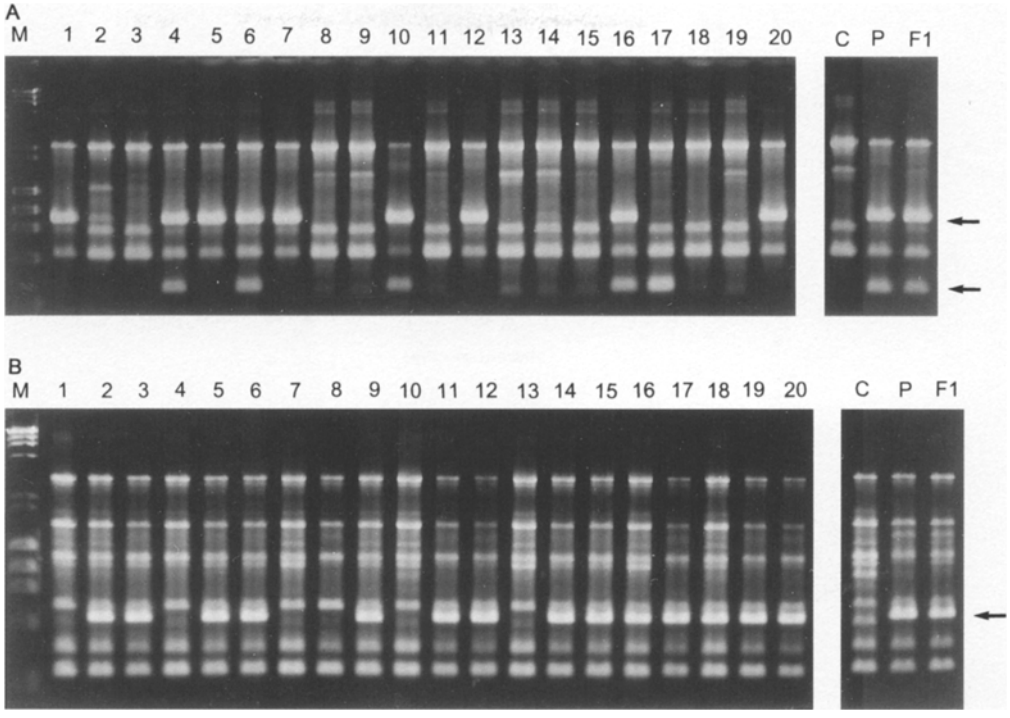


**Fig. 2** RAPD segregation patterns of backcross trees, Catongo (*C*), Pound 12 (*P*) and the $F_1$ tree used for the backcross with Catongo for the primers AM10 (**A**) and L03 (**B**). *Numbers 1–20* represent a sample of 20 segregating backcross trees. *M* indicates molecular size marker (*λ*/*Hin*dIII and *ϕ*X174/*Hae*III). Scorable RAPDs are marked by an *arrow*

## Linkage mapping

One-hundred and thirty eight of the 158 loci segregating in the backcross population could be mapped into ten linkage groups numbered 1–10, in descending order of their length in map units (Fig. 3). This number of linkage groups corresponds to the haploid chromosome number for *T. cacao* (n=10). The 20 other loci were not significantly linked to any other locus (maximum recombination=0.3, LOD=3.0). The linked loci covered 1068 cM of the cocoa genome.

The ten cocoa linkage groups range in size from 138.1 to 65.9 cM and the number of map units per chromosome is correlated with the number of markers per linkage group ($R^2$=0.43), indicating that the loci are randomly distributed over linkage groups. The distribution of interval sizes, expressed in centiMorgans, between adjacent markers on the cocoa genetic map ranged from 0 to 36.9 cM, with an average of 8.3 cM. A majority of the intervals (67%) are below 10 cM. Several markers were found to co-segregate: RAPD markers (B1 M, I1 M, M18B) on linkage group 1, markers (C5M2, AI03, CCC442) on group 7, and the markers (AK08, AX15 and CCG1444, AR06, *Anth*) on group 5. Regions of high marker density could be observed on linkage groups 1, 5, 6 and 7. These regions may represent areas of low recombination frequency which often occur in centromeric areas.

## Genome composition of backcross individuals

Of the 138 mapped loci in the $BC_1$, 25.3% of the alleles were inferred to be of paternal origin (Pound 12), while 74.7% were from the recurrent parent (Catongo). This locus composition of an average backcross population is very close to the theoretical expectation (25%–75%).

Two backcross trees have a relatively low percentage of Catongo alleles (59 and 59.1%). At the other extreme, one $BC_1$ tree contains 91.8% recurrent alleles. This variation of the Catongo genome in backcross individuals could be used for genotypic selection to accelerate introgression of a defined chromosomal segment from Pound 12 into a Catongo-type tree.

## Cocoa genome-length estimation

Map lengths, estimated according to the moments method of Hulbert et al. (1988), were 1081, 1098 and 1112 cM for LOD scores of 3, 4, and 5, respectively. When the equation of Beckmann and Soller (1983) is used and the chromosome ends adjusted (Lange and Boehnke 1982), approximately 80% of a genome should be linked within 10 cM of 130 markers, given a map length of 1200 cM.

The cocoa mapping data are close to these theoretical values with 67% of the cocoa genome linked within 10 cM of the 138 mapped markers. Given a genetic length of 1200 cM, some 214 markers would be expected in order that at least 95% of the genome lie within 10 cM of a marker.

## QTL analysis

The QTL study was performed on a set of four traits which were selected because they represented cocoa characteris-

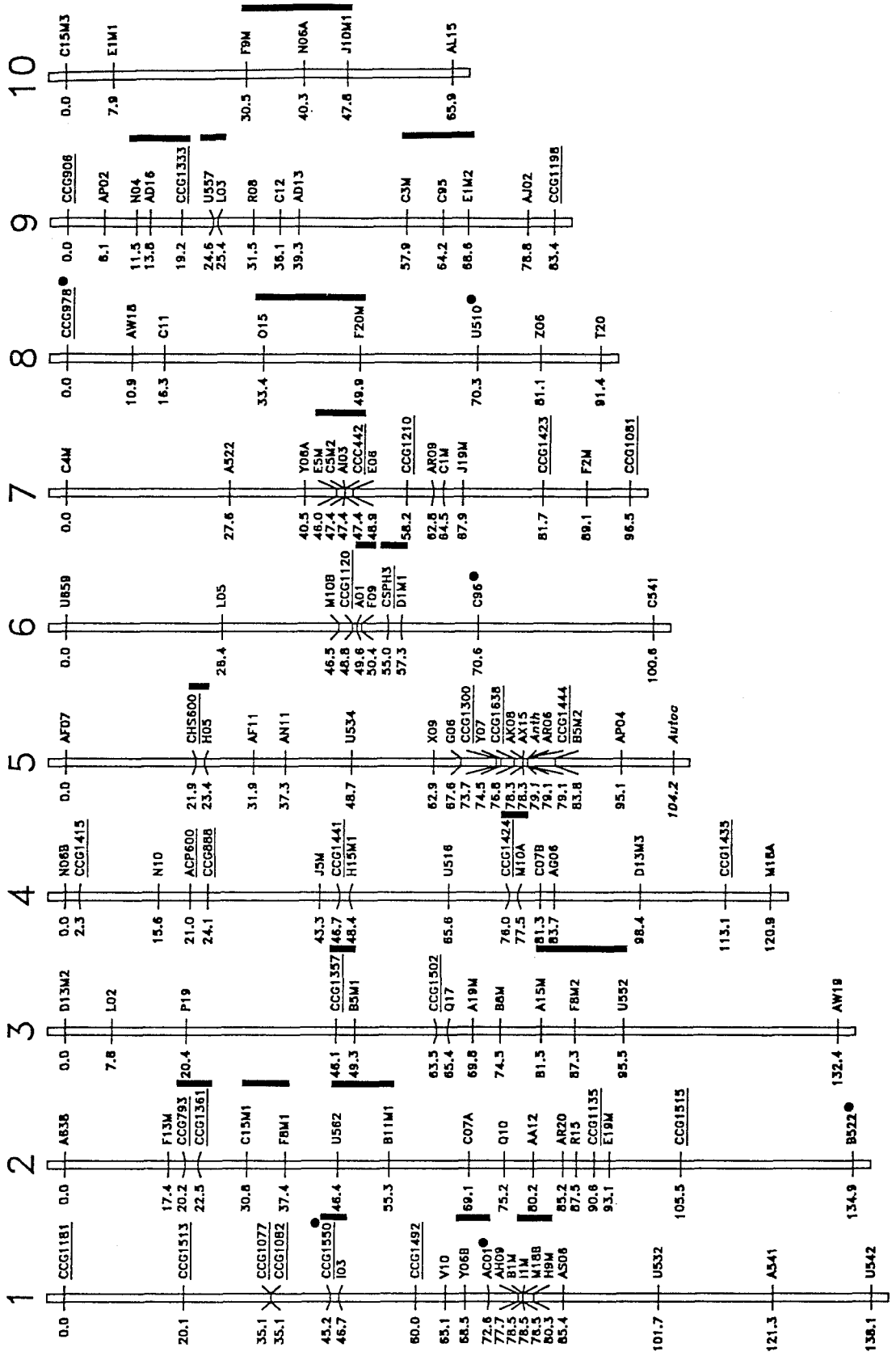

**Fig. 3** Genetic cocoa linkage map derived from the backcross population of 131 plants. Loci (138) were distributed to ten linkage groups (*1–10*). The numbers to the left of each linkage group represent the cumulative map distance in cM (Kosambi function). The best-fit relative locations of all marker loci are shown. Those that could not be ordered unequivocally (Ripples ΔLOD<–2) are identified with a *vertical bracket* extending over the region of possible ambiguity. RFLP loci are *underlined* (32), the two phenotypic loci are in *italics*, and all others are RAPD loci (104). Loci identified by the symbol • displayed distorted segregating ratios

**Table 1** A – homozygote, H – heterozygote, N – number of plants used in the analysis (sum of homozygote and heterozygote trees given in parentheses in columns A and H), $R^2$ – percent phenotypic variance attributable to a marker, LOD – LOD-score values associated with marker loci for the trait, NS – not significant, S – state of the nearest locus favoring the trait. Usually only one marker for a linkage group, associated with the higher percentage of phenotypic variance explained, is shown

**A** QTL data for early flowering habit in cocoa backcross population

| Marker | Days to flower | | Difference | N | F-ratio | $R^2$ | *P*-value | LOD | Linkage group | S |
|---|---|---|---|---|---|---|---|---|---|---|
| | A | H | | | | | | | | |
| AS06 | 823(60) | 879(68) | 56 | 128 | 6.9 | 5.2% | 0.0098 | 1.5 | 1 | A |
| B522 | 900(42) | 823(80) | 77 | 122 | 11.6 | 8.8% | 0.0009 | 2.6 | 2 | H |
| U516 | 883(65) | 811(57) | 72 | 122 | 11.7 | 8.9% | 0.0009 | 2.7 | 4 | H |
| AF07 | 882(55) | 822(68) | 60 | 123 | 7.9 | 6.1% | 0.0057 | 1.9 | 5 | H |

**B** QTL data for trunk diameter in cocoa backcross population

| Marker | Trunk diameter | | Difference | N | F-ratio | $R^2$ | *P*-value | LOD | Linkage group | S |
|---|---|---|---|---|---|---|---|---|---|---|
| | A | H | | | | | | | | |
| CCG1082 | 6.37(67) | 5.59(63) | 0.78 | 130 | 10.1 | 7.3% | 0.0019 | 2.1 | 1 | A |
| CCG1424 | 5.61(65) | 6.34(62) | 0.73 | 127 | 8.2 | 6.1% | 0.0050 | 1.8 | 4 | H |

**C** QTL for jorquette height in cocoa backcross population

| Marker | Jorquette height | | Difference | N | F-ratio | $R^2$ | *P*-value | LOD | Linkage group | S |
|---|---|---|---|---|---|---|---|---|---|---|
| | A | H | | | | | | | | |
| CCG1082 | 105.6(66) | 91.9(63) | 13.7 | 129 | 12.1 | 8.7% | 0.0007 | 2.6 | 1 | A |
| A541 | 98.7(28) | 115(28) | 16.3 | 56 | 7.1 | 11.7% | 0.0100 | NS | 1 | H |

**D** QTL for ovule number in cocoa backcross population

| Marker | Ovule number | | Difference | N | F-ratio | $R^2$ | *P*-value | LOD | Linkage group | S |
|---|---|---|---|---|---|---|---|---|---|---|
| | A | H | | | | | | | | |
| CCG1135 | 50.1(45) | 48.4(48) | 1.7 | 93 | 14.9 | 14.1% | 0.0002 | 3.1 | 2 | A |
| P19 | 48.4(45) | 50.1(48) | 1.7 | 93 | 14.3 | 13.6% | 0.0003 | 3.3 | 3 | H |

tics of agronomic significance. These traits were: early flowering (EF), trunk diameter (TD), jorquette height (JH) and ovule number (OV). Analyses of variance revealed that 62 (11.2%) of the 552 marker-locus quantitative trait comparisons were significant at the 5% probability level, 3.5% of the comparisons were significant at 1%, and 1% was significant at 0.1%.

Marker-trait associations significant at $P<0.01$ are listed in Table 1. A comparison of ANOVA probability levels with the LOD score values, for a number of backcross individuals of at least 90, show that for these data a 1% probability level corresponds to a LOD score of 1.5.

Four QTLs were detected for the early flowering trait, and two QTLs each for trunk diameter, jorquette height and for ovule number (Table 2). These QTLs are located on linkage groups 1 to 5. Some of them seem to be associated. A common QTL is detected for trunk diameter and jorquette height on linkage group 1 near RFLP marker CCG1082. These data are corroborated by a linear regression analysis between trunk diameter and jorquette height values, with $R^2=0.27$, $P<0.0001$ (Table 2). The data suggest a partial pleiotropic relationship for these traits.

The magnitudes of gene effects are represented here according to the portion of the total phenotypic variation in a quantitative trait that is explained ($R^2$) by the marker locus, as described by Edwards et al. (1987). The $R^2$ values among the QTL varied from 5.2% (for RAPD locus AS06 associated with the early flowering trait) to 14.1 (CCG1135 RFLP locus correlation with ovule number). Multiple regression analyses performed on the four traits studied allowed the genetic explanation of, respectively, only 11.2% (trunk diameter) to higher than 25.8% (early flowering) of the phenotypic variances (Table 3).

Of the ten QTLs detected for the four studied traits, six of them are favored by a heterozygous locus and four by a homozygous status.

**Table 2** Correlation coefficients ($R^2$) among the four $BC_1$ traits studied

| Trait | Early flowering | Trunk diameter | Jorquette height | Ovule number |
|---|---|---|---|---|
| Early flowering | | | | |
| Trunk diameter | 24.6 **** | | | |
| Jorquette height | 17.7 **** | 27.1 **** | | |
| Ovule number | 0.7 | 0.1 | 0.2 | |

**** $P<0.0001$

**Table 3** Summary of QTL data for early flowering, trunk diameter, jorquette height, and ovule number

| Trait | Number of QTLs | QTLs on linkage groups | % Phenotypic variance[a] |
|---|---|---|---|
| Early flowering | 4 | 1,2,4,5 | 25.8 |
| Trunk diameter | 2 | 1,4 | 11.2 |
| Jorquette height | 2 | 1 | 16.5 |
| Ovule number | 2 | 2,3 | 23.8 |

[a] Percent phenotypic variance explained for each trait was performed with a multiple regression analysis

## Discussion

### Heterozygosity evaluation

The level of DNA polymorphism detected by the RAPD and RFLP techniques in *T. cacao* allowed the estimation of heterozygosity in parental cocoa clones and the construction of a medium-density linkage map using $BC_1$ progeny.

High heterozygosity in woody species has been referenced in conifers, using RAPD (Carlson et al. 1991) and RFLP (Devey et al. 1991) techniques, and in peach (Rajapakse et al. 1995). The use of RAPD markers does not appear to be a problem when it is used for mapping highly heterozygous outcrossed species. This is probably due to the large number of loci found to be in a testcross configuration between the two parents (Grattapaglia et al. 1992). The average number of RFLP alleles found in the present study was about 2.04, and does not represent any advantage in comparison with the two (presence versus absence) identified by RAPD markers. Only when a RAPD locus is heterozygous in both parents does the fragment-present phenotype represent an ambiguous genotype in the progeny. However, due to the low heterozygosity in the Catongo clone we did not find this to be a problem in our cocoa mapping experiment.

In this study, the number of polymorphic loci found to be heterozygous for Catongo is 2%, for Pound 12 60 to 81%, and for the $F_1$ tree used for the backcross 66–68%. These values are clearly different due perhaps in part to the self-incompatibility status favoring outcrossing. It is worth noting that Catongo is a near-isogenic line created by continuous selfing in its native environment, and so can serve a unique role in cocoa breeding programs.

Similar studies in apple mapping (Hemmat et al. 1994) found a high heterozygosity level for both parents [65.5% (268/409) and 44% (180/409)]. In *Pinus* interspecific mapping, the parental heterozygosity values were about 56% (167/298) and 44% (131/298) (Kubisiak et al. 1995).

These results suggest that by taking advantage of the high level of heterozygosity often present in this long-lived, generally self-incompatible, species, mapping experiments could be performed with an $F_1$ cocoa population.

### Linkage-map construction

The cocoa genetic linkage map was based firstly on RAPD markers, which provided a fast and efficient way to construct a genomic map, and secondly on RFLP markers to allow a better reliability of this map for future cocoa QTL study.

The RAPD markers were chosen on the basis of repeatability and inheritance patterns. These RAPD polymorphisms should be valuable genetic markers for further mapping studies provided that care is taken to use identical conditions for carrying out the PCR protocols, as noted by Penner et al. (1993). A RAPD analysis performed on a subset of the same $BC_1$ cocoa progeny (Ronning et al. 1995) allowed us to successfully place some of the previous segregating RAPD markers on this linkage map.

Cocoa trees are generally genetically heterozygous and highly outcrossed; thus anonymous markers as RAPDs can be readily detected despite their dominant pattern of inheritance. RAPD markers are considered as an efficient first step towards establishing a genomic map for previously unstudied species (Tulsieran et al. 1992; Nelson et al. 1993; Grattapaglia and Sederoff 1994).

In cocoa, the complex ovarian incompatibility system (S) has been studied extensively with the general agreement that most results can be explained by a single locus with several alleles (Knight and Rogers 1955), although some other data (Cope 1962) show the presence of three independent loci. More recently, Warren et al. (1995) found that two isozymes were linked to one locus controlling cocoa compatibility status. In the present study, we demonstrate that for the analysed $BC_1$ cross only one locus is involved in this phenotype determination, and the genetic linkage between *Autoc* and *Anth* loci could be useful as a predictor of self-compatibility since this trait seems associated with yield (Lockwood 1977).

Cocoa has ten chromosome pairs (Munoz 1948; Glicenstein and Fritz 1989). The framework linkage map from this cocoa $BC_1$ population has a total map distance of 1068 cM. The 138 marker loci appeared to provide nearly complete coverage of the cocoa genomic map, since the method of Hulbert et al. (1988) provided an estimate of total map length ranging from 1081 to 1112 cM, close to the observed experimental map length. With an average content per haploid genome of $4 \times 10^8$ bp (Figueira et al. 1992; Lanaud et al. 1992) it could be estimated that 1 cM equals approximately 375 kb. However, this is only an average

value which depends on the portion of the map considered (Tanksley et al. 1992).

Recent cocoa mapping experiments using two heterozygous clones were performed by Lanaud et al. (1995). The resulting cocoa genetic map comprises 193 loci and covers 759 cM with a 3.9-cM average distance between two markers. Despite being of a high density, the size of this map appears to be smaller than the one studied here. This could be due to the high number of RFLP probes (83%) used compared to 23% for our map.

## QTL mapping in a $BC_1$ population

In this cocoa $BC_1$ population we were able to map four putative QTLs affecting early flowering and two each affecting trunk diameter, jorquette height and ovule number. In total, the QTLs which could be mapped in this $BC_1$ cross accounted only for a small part of the phenotypic variance (11.2–25.8%). There are a number of reasons why no more than 25.8% of the total phenotypic variance for any trait was found in this study, including the fact that the medium-density cocoa map has several large gaps that could contain genes related to these traits. Also, there may be significant interactions between genotypes and environmental factors which contribute to regulating the expression of the traits (Paterson et al. 1991). However, it seems most likely that the traits considered here are among the most complex determining the character of a woody species such as cocoa.

Trunk diameter and jorquette height traits appear to be partially correlated. A RFLP marker is significantly associated with both traits in this investigation. The genetic control of these traits has not yet been shown to be distinct, thus leaving open the possibility of either linkage or pleiotropy as the cause of their correlation. A separation of QTL effects for the two traits on linkage group 1 may be accomplished by fine-mapping approaches (Paterson et al. 1990).

ANOVA and interval-mapping analyses were compared for the cocoa QTL study. The significance thresholds used ($P$=0.01 for ANOVA and LOD 1.5 for Mapmaker QTL) were generally comparable with one exception for the association of RAPD marker A541 on linkage group 1, which was declared significant by ANOVA and out of range by the interval-mapping study. Similarity between regression analysis and interval-mapping analysis in QTL mapping has been observed previously (Stuber et al. 1992; Doebley and Stec 1993). Darvasi et al. (1993) estimated numerically that, at a marker spacing of 10 cM and 11 markers per chromosome, a LOD score of 1.53 corresponds to a per-marker type-I error rate of 0.0084 and ensures a 0.05 per-chromosome type-I error. With an infinite number of markers, the LOD threshold would have to increase to 1.96. These estimates were obtained under a backcross model. In our study, marker spacings along the cocoa map are about 8.3 cM and the average number of markers per chromosome is about 14. Therefore, the stringency adopted to declare a QTL in this study seems satisfactory since Doerge (1995) showed a clear relationship between the ANOVA and interval-mapping methods.

In conclusion, results from this investigation demonstrate the effectiveness of using a $BC_1$ cocoa population in QTL analysis. The QTL management, for the four traits considered, can be performed through selection assisted by DNA markers. In this way, the long duration of cocoa breeding work could be significantly reduced and the management of important cocoa traits, such as yield or disease resistance, could be facilitated.

**Acknowledgments** The research was supported in part under grant number DPE-55542-G-SS-7034, program in science and technology cooperation, office of the science advisor, US Agency for international development, and by an endowment to Penn State university by the American Cocoa Institute (ACRI). Estelle Lerceteau was supported by a grant from the Ministère de l'Enseignement supérieur et de la Recherche. Joseph Osei was supported by the University of Ghana and by a fellowship from the African-American Institute, New York. The authors are grateful to Prof. Steven Tanksley of Cornell University and Dr. Alain Deshayes for valued counsel, to Ms. Laurence Bellanger and Mr. Michel Rigoreau from CRN Tours for their contribution to the molecular mapping, to Mr. Antonio Alfaro at CATIE for superior field work, and to Mrs. Lydia Fritz for counting 20 000 very small ovules.

## References

Allen JB (1988) Geographical variation and population biology in wild *Theobroma cacao*. PhD thesis, University of Edinburgh

Anonymous (1995) Quarterly bulletin of cocoa statistics, 21:29. ICCO, London

Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. Plant Mol Biol Rep 9:208–219

Beckmann JS, Soller M (1983) Restriction fragment length polymorphisms in genetic improvement: methodologies, mapping and costs. Theor Appl Genet 67:35–43

Bekele FL, Kennedy AJ, Mc David C, Lauckner FB, Bekele I (1994) Numerical taxonomic studies on cacao (*Theobroma cacao* L.) in Trinidad. Euphytica 75:234–240

Carlson JE, Tulsieram LK, Glaubitz JC, Luk VWK, Kauffeldt C, Rutledge R (1991) Segregation of random amplified DNA markers in $F_1$ progeny of conifers. Theor Appl Genet 83:194–200

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

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

Cope FW (1976) *Theobroma cacao* (Sterculiaceae). In: Simmonds NW (ed) Evolution of crop plants. Longman, London, pp 285–289

Couch J, Zintel HA, Fritz P (1993) The genome of the tropical tree *Theobroma cacao* L. Mol Gen Genet 237:123–128

Darvasi A, Weinreb A, Minke V, Weller JI, Soller M (1993) Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. Genetics 134:943–951

Devey ME, Jermstad KD, Tauer CG, Neale DB (1991) Inheritance of RFLP loci in a loblolly pine three-generation pedigree. Theor Appl Genet 83:238–242

Doebley J, Stec A (1993) Inheritance of the morphological differences between maize and teosinte: comparison of results for two F2 populations. Genetics 134:559–570

Doerge RW (1995) The relationship between the LOD score and the analysis of variance F-statistic when detecting QTLs using single markers. Theor Appl Genet 90:980–981

Edwards MD, Stuber CW, Wendel JF (1987) Molecular-marker-facilitated investigations of quantitative trait loci in maize. I. Numbers, genomic distribution and types of gene action. Genetics 116:113–125

Engels JM (1992) The use of botanical descriptors for cocoa characterization: CATIE experiences. In: International workshop on conservation, characterisation and utilisation of cocoa genetic resources in the 21st century. Port-of-Spain, Trinidad. 13–17th september 1992. The Cocoa research Unit, The University of the West Indies, pp 69–76

Figueira A, Janick J, Goldsbrough P (1992) Genome size and DNA polymorphism in *Theobroma cacao*. J Am Soc Hort Sci 117:673–677

Figueira A, Janick J, Morris L, Goldsbrough P (1994) Re-examining the classification of *Theobroma cacao* L. using molecular markers. J Am Soc Hort Sci 119:1073–1082

Glicenstein L, Fritz P (1989) Ploidy level in *Theobroma cacao*. J Hered 80:464–467

Grattapaglia D, Sederoff RR (1994) Genetic linkage maps of *Eucalyptus grandis* and *E. urophylla* using a pseudo-testcross mapping strategy and RAPD markers. Genetics 137:1121–1137

Grattapaglia D, Chaparro J, Wilcox P, McCord S, Werner D, Amerson H, McKeand S, Bridgewater F, Whetten R, O'Malley D, Sederoff R (1992) Mapping in woody plants with RAPD markers: application to breeding in forestry and horticulture. Proc Symp Appl RAPD Technol to Plant Breed, November 1, 1992, Minneapolis, Minnesota, pp 37–40

Hemmat M, Weeden NF, Manganaris AG, Lawson DM (1994) Molecular marker linkage map for apple. J Hered 85:4–11

Hulbert SH, Ilott TW, Legg EJ, Lincoln SE, Lander ES, Michelmore RW (1988) Genetic analysis of the fungus, *Bremia lactucae*, using restriction fragment length polymorphisms. Genetics 120: 947–958

Knight R, Rodgers HH (1955) Incompatibility in *Theobroma cacao*. Heredity 9:69–77

Kosambi DD (1944) The estimation of map distances from recombination values. Ann Eugen 12:172–175

Kubisiak TL, Nelson CD, Nance WL, Stine M (1995) RAPD linkage mapping in a longleaf pine×slash pine $F_1$ family. Theor Appl Genet 90:1119–1127

Lanaud C (1986) Utilisation des marqueurs enzymatiques pour l'étude génétique du cacaoyer: *Theobroma cacao* L. II. Etude du polymorphisme de six systèmes enzymatiques. Café Cacao Thé 30:271–276

Lanaud C, Hamon P, Duperray C (1992) Estimation of nuclear DNA content of *Theobroma cacao* L. by flow cytometry. Café Cacao Thé 36:3–8

Lanaud C, Risterucci AM, N' Goran AKJ, Clement D, Flament MH, Laurent V, Falque M (1995) A genetic linkage map of *Theobroma cacao* L. Theor Appl Genet 91:987–993

Lander ES, Botstein D (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121: 185–199

Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newburg L (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174–181

Lange K, Boehnke M (1982) How many polymorphic genes will it take to span the human genome? Am J Hum Genet 34:842–845

Laurent V, Risterucci AM, Lanaud C (1994) Genetic diversity in cocoa revealed by cDNA probes. Theor Appl Genet 88:193–198

Lockwood G (1977) Studies on the effects of cross-incompatibility on the cropping of cocoa in Ghana. J Hort Sci 52:113–126

Lockwood G (1985) Genetic resources of the cocoa plant. Span 28:14–16

Lockwood G, End M (1992) History, technique and future needs for cocoa collecting. In: International workshop on conservation, characterisation and utilisation of cocoa genetic resources in the 21st century. Port-of-Spain, Trinidad, 13–17th September 1992. The Cocoa research Unit, The University of the West Indies, pp 1–14

Morera J, Paredes A, Mora A (1991) Germoplasma de cacao en el CATIE entre 1947 y 1991. In: CATIE/IICA.a II: generacion y transferencia de technologia. San Jose, Costa Rica. pp 49

Munoz JMO (1948) Estudios cromosomicos en el genero *Theobroma* L. MAg thesis, CATIE/IICA, Turrialba, Costa Rica

Nelson CD, Nance WL, Doudrick RL (1993) A partial genetic linkage map of slash pine (*Pinus elliottii* Engelm var. *elliottii*) based on random amplified polymorphic DNAs. Theor Appl Genet 87:145–151

N'goran JAK, Laurent V, Risterucci AM, Lanaud C (1994) Comparative genetic diversity studies of *Theobroma cacao* L. using RFLP and RAPD markers. Heredity 73:589–597

Paterson AH, De Verna JW, Lanini B, Tanksley SD (1990) Fine mapping of quantitative trait loci using selected overlapping recombinant chromosomes, in an interspecies cross of tomato. Genetics 124:735–742

Paterson AH, Damon S, Hewitt JD, Zamir D, Rabinowitch HD, Lincoln SE, Lander SE, Tanksley SD (1991) Mendelian factors underlying quantitative traits in tomato: comparison across species, generations, and environments. Genetics 127:181–197

Penner GA, Bush A, Wise R, Kim W, Domier L, Kasha K, Laroche A, Scoles G, Molnar SL, Fedak G (1993) Reproducibility of random amplified DNA, RAPD analysis among laboratories. PCR Methods Applic 2:341–345

Rajapakse S, Belthoff LE, He G, Estager AE, Scorza R, Verde I, Ballard RE, Baird WV, Callahan A, Monet R, Abbot AG (1995) Genetic linkage mapping in peach using morphological, RFLP and RAPD markers. Theor Appl Genet 90:503–510

Ronning CM, Schnell RJ (1994) Allozyme diversity in a collection of *Theobroma cacao* L. J Hered 85:291–295

Ronning CM, Schnell RJ, Kuhn DN (1995) Inheritance of random amplified polymorphic DNA (RAPD) markers in *Theobroma cacao* L. J Am Soc Hort Sci 120:681–686

Russel JR, Hosein F, Johnson E, Waugh R, Powell W (1993) Genetic differentiation of cocoa (*Theobroma cacao* L.) populations revealed by RAPD analysis. Mol Ecol 2:89–97

Snyder T (1994) Isolation and characterization of a genomic chitinase clone from *Theobroma cacao*. PhD thesis, The Pennsylvania State University

Spencer ME, Hodge R (1991) Cloning and sequencing of the cDNA encoding the major albumin of *Theobroma cacao*. Identification of the protein as a member of the Kunitz protease inhibitor family. Planta 183:528–535

Stuber CW, Lincoln SE, Wolff DW, Helentjaris T, Lander ES (1992) Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers. Genetics 132:823–839

Suiter K, Wendel JF, Case JS (1983) Linkage-1 a Pascal computer program for the detection and analysis of genetic linkage. J Hered 74:203–204

Tanksley SD, Ganal MW, Prince JP, de Vincente MC, Bonierbale MW, Broun P, Fulton TM, Giovanonni JJ, Grandillo S, Martin GB, Messeguer R, Miller JC, Miller L, Paterson AH, Pineda O, Röder MS, Wing RA, Wu W, Young ND (1992) High-density molecular-linkage maps of the tomato and potato genomes: biological inferences and pratical applications. Genetics 132:1141–1160

Tai H, Mc Henry L, Fritz P, Furtek D (1991) Nucleic-acid sequence of a 21-kDa cocoa seed protein gene with homology to the soybean trypsin inhibitor (Kunitz) family of protease inhibitors. Plant Mol Biol 16:913–915

Tulsieram LK, Glaubitz JC, Kiss G, Carlson JE (1992) Single-tree genetic linkage analysis in conifers using haploid DNA from megagametophytes. Bio/Technol 10:686–690

Warren J (1994) Isozyme variation in a number of populations of *Theobroma cacao* L. obtained through various sampling regimes. Euphytica 72:121–126

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

Williams J, Kubelik A, Livak K, Rafalski J, Tingey S (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res 18:6531–6535

Wood G, Lass R (1985) Cocoa, 4th edn, Longman Scientific and Technical, Essex, England