Construction of a low-density linkage map of *Theobroma cacao* using random amplified polymorphic DNA markers and an anthocyanin biosynthetic locus[1]

J.K. Osei*, D.B. Furtek**, M.Goodin***, H. Rodriguez****, J. Morera****, R. Lastra****, P.J. Fritz*****

ABSTRACT

Random amplified polymorphic DNA (RAPD) markers were used to construct a low-density linkage map of *Theobroma cacao* L. The segregating population used for map construction derived from a Catongo X (Pound 12 X Catongo) backcross. DNA from 38 backcross trees was amplified with 28 random sequence ten-nucleotide long primers to give 36 polymorphic DNA fragments that segregated in a one-to-one ratio of present (dominant RAPD allele) to absent in the backcross progeny. Eighteen of these RAPD loci were linked in seven groups. When segregation data of anthocyanin biosynthetic alleles, judged from seed, flower, and leaf flush color, were added to the RAPD marker linkage groups, the anthocyanin locus was linked in one of the groups. A linkage map of this group containing four markers, including the anthocyanin biosynthetic locus, is presented.

RESUMEN

Se usaron indicadores ampliados polimórficos al azar del ARN para construir un mapa de baja densidad de ligamiento de *Theobroma cacao* L. La población segregante usada para la elaboración del mapa proviene de un retrocruzamiento de Catongo X (Pound 12 X Catongo). El ARN proveniente de los 38 árboles del retrocruzamiento fue amplificado con 28 secuencias al azar de "primers" largos de de 10 nucleótidos, que dieron 36 fragmentos polimórficos de ARN, que segregaron en una razón de uno por uno de presente (alelo dominante) a ausente en la descendencia del retrocruzamiento. Dieciocho de estos sitios de RAPD están ligados en siete grupos. Cuando los datos de segregación de los alelos de la biosíntesis de la antocianina, examinados de almendras, flores y brotes nuevos coloreados, fueron sumados a los grupos de RAPD marcadores, se ligaron a uno de estos grupos. Se presenta un mapa de ligamientos de estos grupos, que contiene los cuatro grupos de marcadores, incluyendo el de biosíntesis de la antocianina.

INTRODUCTION

Seeds of *T. cacao* L. (2N=20), family Sterculiaceae are second only to coffee as a major cash crop for about 40 developing countries located within 20 degrees of the equator. Fat pressed from dried seeds is used to manufacture chocolate, cosmetics, and pharmaceuticals. The residue after fat extraction is cocoa powder, the source of chocolate flavor. Commercial cocoa varieties have a narrow genetic base and are susceptible to many diseases, including the ubiquitous black pod disease *(Phytophthora* spp.), the South American witches' broom disease *(Crinipellis perniciosa)*, the West African swollen shoot disease (cocoa swollen shoot virus), and the South Pacific vascular streak die-back disease *(Oncobasidium theobromae)* (Wood and Lass 1987).

These diseases and the myriad of insects that feed on cocoa can reduce the world-wide crop by 10 to 30 percent per year. Thus, there is a great need for developing cocoa varieties that are more resistant to these pests. However, because of a long generation time (three to five years), cocoa breeding is slow. Traditional cocoa breeding is generally based on heterosis between parents selected from three major cocoa populations: Upper Amazon Forasteros, Lower Amazon Forasteros, and Trinitarios.

Recurrent selection may be a better approach for breeding superior cocoa varieties if a reliable method for selecting adult characters at the seedling stage can be developed. Such methods are now available in the form of molecular marker-based selection using restriction fragment length polymorphisms (RFLPs) (Botstein *et al.* 1980) or amplification fragment length

1 Received on February 18, 1994.
J. Osei was supported by the University of Ghana and by a fellowship from the African-American Institute, New York. 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, USAID. Major funding was from an endowment provided to Penn State University by the American Cocoa Research Institute (ACRI). ACRI also supported this work through an individual grant to PJF.

* Penn State University, American Cocoa Research Institute Cocoa Molecular Biology Laboratory, Department of Food Science, University Park, PA. 16802, USA; Present address: University of Ghana, Agricultural Research Station, Kade, Ghana.

** Penn State University, American Cocoa Research Institute Cocoa Molecular Biology Laboratory, Department of Food Science, University Park, PA. 16802, USA.

*** Penn State University Department of Plant Pathology, USA.

**** Centro Agronómico Tropical de Investigación y Enseñanza (CATIE), Turrialba, C.R.

***** Present address: CATIE

polymorphisms (AFLPs) and their most common application, random amplified polymorphic DNAs (RAPDs) (Williams *et al.* 1990; Welsh and McClellan (1990). RAPD markers are not codominant, as are RFLP markers, and thus it may be necessary to determine the zygosity at any locus by progeny analysis (Carlson *et al.* 1991). Carlson *et al.* (1991) presented evidence for white spruce and Douglas, fir that RAPD markers segregate appropriately as independent alleles and thus can serve as markers for genetic linkage analysis. The RAPD procedure has the advantage of speed, is less expensive than RFLP analysis and has been used to construct genetic linkage maps for several plants including *Arabidopsis thaliana* (Reiter *et al.* 1992) and loblolly pine (Chaparro *et al.* 1992).

RAPD markers that are tightly linked to genes conferring desirable characters in cocoa can greatly facilitate breeding programs by providing a means to select for desired genotypes at the seedling stage and thereby reduce the cost and time of cocoa breeding. RAPD markers are also valuable for fingerprinting cocoa genotypes to minimize duplications in cocoa collections and to determine the degree of genetic diversity of populations (Wilde *et al.* 1992; Lanaud *et al.* 1992; Lerceteau *et al.* 1992). In this paper we describe the construction of a low-density partial linkage map of *T. cacao* containing 18 RAPD loci and an anthocyanin biosynthetic locus.

MATERIALS AND METHODS

Plant material

Catongo and Pound 12 clones from the CATIE cocoa collection were the primary source of plant material used in this study. Catongo is a white-seeded, self-compatible Lower Amazon Forastero discovered more than 50 years ago in Bahia, Brazil (Wood and Lass 1985); it is moderately susceptible to *P. palmivora* (Phillips and Galindo 1989). Pound 12 is a purple-seeded, self-incompatible Upper Amazon Forastero collected near the headwaters of the Amazon River by Pound in 1943 (Pound 1945); it is moderately resistant to *P. palmivora* (Phillips and Galindo 1989).

The backcross progeny (BC) were derived by pollinating a Catongo clone with an F1 tree originating from a Catongo X Pound 12 cross, i.e., Catongo X (Catongo X Pound 12), at CATIE. Seeds from six pods were planted at CATIE, and seeds from three pods were planted at the ACRI-Penn State University greenhouse. Leaves from both parental clones, the F1 tree, and 10 one-year-old BC saplings were harvested at CATIE, shipped to Penn State, and stored at -80° C until used for DNA isolation. Leaves from 28 one-year-old saplings growing in the ACRI-Penn State greenhouse were harvested and stored under the same conditions until used for DNA isolations.

DNA isolation

Genomic DNA from leaves was isolated according to the method of Couch and Fritz (1990). DNA concentrations were estimated by running aliquots on 0.8% agarose gels and comparing band concentrations with lambda DNA standards after staining with ethidium bromide.

Polymerase chain reaction (PCR)

The PCR was performed basically as described by Williams *et al.* (1990). The 50 μl reaction mixtures contained 10 mM Tris-HCI (pH 8.3 at room temperature), 0.01% gelatin, 50 mM KCI, 1.5 mM MgCl2, 0.2 mM each of four dNTPs, 300 ng of single 10-mer random sequence oligonucleotide primers, 2.5 units *Taq* polymerase (Perkin Elmer Cetus, Norwalk, CT), and 50 ng genomic DNA. The reaction mixture was overlaid with 50 μl mineral oil and incubated in a boiling water bath for 10 min, followed by quick cooling in ice and addition of the *Taq* polymerase. Amplification was performed in a Techne PHC-2 (Princeton, NJ) thermal cycler for 42 cycles with the following temperature conditions: 92° C for 30 sec, 36° C for 1 min, and 72° C for 2 minutes.

Amplified DNA fragments produced in the PCR were separated by gel electrophoresis (5 volts per centimeter) in 0.5 X TBE buffer (25 mM Tris-borate, pH 8.3, 0.5 mM EDTA) using 1.2% agarose containing 0.5 μg per mililiter ethidium bromide. A 123 bp ladder (BRL, Bethesda, MD) was used as a molecular weight marker in each gel.

Primers of arbitrary sequence, but containing at yeast 50% G+C, were obtained from either J. Carlson at the University of British Columbia,

Vancouver, Canada (kits 500 and 600), or from Operon Technologies of Alameda, CA (kit A).

The presence or absence of segregating markers was scored and analyzed using the MAPMAKER computer program (Lander *et al.* 1987). In discussion of the results, use of phrases such as "present in Pound 12" means a band or bands amplified when Pound 12 DNA is used as template in the PCR.

RESULTS AND DISCUSSION

Segregation of RAPD markers in backcross progeny

Primers generating at least one band (dominant RAPD allele) present in Pound 12, but absent in Catongo, were tested singly in the PCR to study segregation of RAPD markers. Most of these bands were present in the F1 and were expected to segregate in a 1:1 ratio of presence or absence in the backcross progeny. Bands present in Pound 12, but absent in the F1, were probably due to heterozygosity in Pound 12 and segregation of the recessive allele in the F1.

Of 220 primers tested on the parents, 33 produced at least one RAPD marker that was present in Pound 12, but absent in Catongo. These markers segregated into distinguishable phenotypic classes of presence or absence of an amplified fragment in the backcross progeny. Eight of the 33 primers produced two segregating markers. In accordance with the MAPMAKER computer program, presence of a marker in a backcross progeny plant was scored as D (not homozygous for allele "A" at this locus), and absence of a marker was scored as B (homozygous for allele "a" at this locus).

Table 1. Segregation patterns of RAPD markers in *T. cacao* backcross population.

1.	534/615	BBBDBBDBDBBBDBD-D--DDDDBBDBDDBBBDDDBB
2.	516/307	BDBBBDDBDDBDDDDDBBBDDD-DDDDDDDDDBDBB-
3.	516/492	BBDDBBBBDBBBBBDDBDDDD-DDDDDDDBB
4.	522/184	-BDBBDBDDDBDDDBBDDBBBDDBDBBBDBD
5.	522/553	BBBBDBDBDBBDBDBBBBDBDBDDDBDDDDDDD—
6.	532/861	BDBBDBDDDDDBDDDDBDDDDDB-BBBDDDDBDBDB–
7.	537/738	BBBBBDDDDBDBDDDBBDDBDBBDBBBDDDBBBB—
8	541/984	DBBDDDBDDDDBBDBBBBBDDBBDBBBD-DBB-B–B
9.	541/861	BDBBBBDBDDBDDDDDBBBDBDBBBDBD-DDD-B–B
10.	546/307	BBBDBBB-BDDBBBD-DDDDDDBBB-BDBBB——
11	551/738	DBDDDDDBDDDBDDB-B–BBB–BBBDBBDD-BBBB
12	550/615	DBBDBDDBBDDDBBB-BD-DDDD-DBBBBDBD-DD-D
13.	562/615	DBBDDD-DDDBDDBD-DB-BDDDBBDDD-BBD-DB-DB
14.	552/400	BBBDBD-BBDBBBBB-BD-BDDDBDBBBDDDBDBBD-
15.	557/615	BDBBBB-DDDBDDDB-BB-DBDDDDDDD-DDD-BD-BD
16.	553/615	DBDDDD-DDDDBDBB-BB-DBBDDBBBB-BBB-BD-BB
17.	553/117	BBBBBB-DDDDDBBD-DD-DBDDDDDDD-BDD-DB-BD
18.	659/117	DDBBBB-BBDDDBBD-DD-BDBBBBDBD-DBD-D–BB
19.	638/300	DBD-DB-BDBBDDDD-DB-DBDDBB-DB-DBB-DB–DD
20.	575/984	DBD-DB-BDBBDDBD-BB-BBBBBB-DB-DBB-D–BD
21	564/615	DDD-DD-BBBDDDDD-DD-DDBBBD-BB-BBD-B–BB
22.	608/300	DDD-DB-BBBBBDBD-BD-DBBBDB-BB-D-B-B–B-
23.	608/984	BBD-DD-BDBBBBDB-BD-DBDDBD-BB-B-B-B–B-
24.	627/117	BDD-DB-DDBBDD-D-BD-DDDDDD-DB-BBB-BB-B-
25.	627/615	DBB-BD-DDDDBB-D-BD-DBDDBB-B———
26	525/861	————BD-DB-BDDDBB-DD-DBB-D–DD
27.	691/984	DDD-DB-BDBDBDDD-BB-DBDDBB-DB-BDD-BD-DB
28.	690/227	BDB-DB–B-DBBBB-BB-BDBBDD-BD-BBD-D–BD
29.	686/117	BBB-BD-DBBDDBBB-DD-B-DDDD-DD-DDD-D–D
30.	667/100	BDB-BD-DDDDBDBD-DD-B-BBB–BB—D-D–B
31	668/861	BBB-BB—BDD——BDDDDB–D-DBB—DD
32.	OA4/984	DDB-BDDBBBBDBBB-DD-DBBBDB–D-DBD-D–D
33.	OA4/615	BBB-BDDDDBBBDDD-BD-BBDBDB–D-DBB-B–B
34.	A10/400	BBB-BDDDDBDB-DD-BD-BDBBDB—DBB-BB–
35.	OA9/861	DBD-B-DBBDD–D–BB-BDBBDD-BD-BBD-DD-D
36	638/984	BDB-DB-DBBDBBBB-BB-BDBBDD-BD-BBD-DD-B
37.	ACY/LOC	BBBDBBBBBBBBBBBDDDDDDDBBBBBBDBDBBBBDBBB

Table 1 shows the segregation patterns of the RAPD markers in the backcross progeny. Dashed lines indicate missing data, also allowed in the MAPMAKER program. RAPD loci are designated by their respective primer number and the size in base pairs of the amplified DNA fragment. Each letter or dash represents an individual tree in the backcross population. For example, the first row shows that when primer number 534 was used in the PCR with DNA from the 38 trees, a 615 base pair amplified fragment (D) was observed in 16 instances, no band of this length (B) was seen in 18 instances, and in 4 instances no data for this primer were collected. Five of the 41 markers tested deviated from the expected 1:1 ratio of present:absent by Chi square analysis and were not used for linkage analysis. Also included in Table 1 (designated ACY/LOC) is the segregation

pattern of the trees used in this study for seed, flower, and leaf flush color.

Linkage groups of the RAPD loci

The MAPMAKER program places markers into linkage groups based on log-likelihood (LOD) scores and recombination fractions. LOD is the logarithm (base 10) of the ratio of probabilities that (a) an observed segregation pattern arose from linkage between two loci given a recombination fraction, 0, greater than 0.5, compared to (b) the segregation pattern that arose from unlinked loci. The recombination fraction between two loci is that which maximizes the LOD score. In the absence of linkage, the LOD score is zero, meaning there is equal probability that an observed segregation pattern is due to linked or unlinked loci. If the LOD score is 2, the odds are 100 times greater that two loci are linked than unlinked. Linkage is considered proved if the LOD score is equal to or greater than 3.

When the segregation data of the RAPD markers shown in Table 1 was analyzed in the MAPMAKER program at a LOD score of 3.0, and recombination fraction 0.25, 17 markers were arranged in 7 linkage groups, 3 with 3 members in the group, and 4 with 2 members in the group (Table 2). Nineteen markers were unlinked. Ultimately, when more data are available, 10 linkage groups, representing the haploid chromosome number for *T. cacao*, are expected.

Table 2. Linkage groups.

Linkage group	RAPD markers in group
A	1, 10, 18, 37
B	7, 33, 34
C	8, 11, 16
D	9, 15
E	14, 23
F	19, 20
G	28, 36

The number of markers (N) needed to construct a linkage map covering the entire genome is given by the relation: N = log (1-P)/ log (1-2 c/k), where P is the proportion of the genome covered by the markers, c is the distance in centimorgans (cM) between the markers, and k is the cumulative map distance (Beckman and Soller 1983). The cumulative map distance for most crop plants is near 1000 cM, and markers should be spaced 20 cM apart to enhance the chances of being linked to quantitative trait loci. Thus, assuming equal spacing, some 75 markers are needed to ensure that any locus will have a 95% probability of being within 20 cM of at least one marker. For 10 cM spacing, 150 markers will be needed, and for 5 Cm, 300 markers. Clearly, the 17 markers reported here represent only a beginning.

Mapping of a *T. cacao* anthocyanin biosynthetic locus

Of 329 seeds obtained from the 9 backcross pods, 172 were white, and 157 purple, demonstrating that segregation of seed color in the backcross progeny fits a 1:1 ratio of purple to white seeds. In addition, several pods from F2 intercrosses revealed a 3:1 ratio of purple to white seeds. From these observations, we conclude that the white-seeded trait in Catongo is conferred by a single recessive allele, and that Pound 12 has a dominant wild-type allele at an anthocyanin biosynthetic locus. To map this locus, backcross progeny raised from purple seeds were considered to be heterozygous at the locus and were scored the same as the dominant RAPD alleles.

Progeny raised from white seeds were considered to be homozygous recessive and scored the same as the recessive RAPD alleles. When these data were combined with the RAPD marker segregation data, the anthocyanin locus was linked in one of the groups. A map of this group is shown in Fig. 1. The cumulative map distance for this group is about 50 cM, probably representing roughly half of one chromosome.

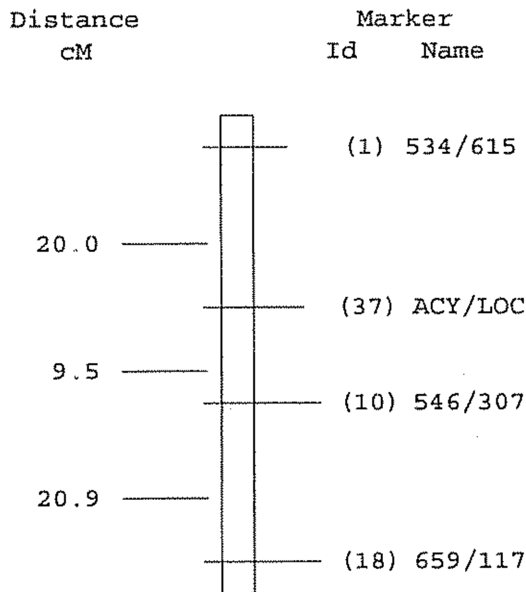

Fig. 1. Group a linkage map.

Work to complete the cacao linkage map is continuing in a collaboration between CATIE, the FRANCERECO laboratory in Notre Dame D'Oe, France, and the USDA Subtropical Horticultural Research Station in Miami, Florida, USA. The completed map, in addition to RAPD markers and single gene traits, will also include markers for multigene coded traits, the so-called quantitative trait loci or QTLs.

LITERATURE CITED

BECKMAN, J.S.; SOLLER, M. 1983. Restriction fragment length polymorphisms in genetic improvement: Methodologies, mapping, and costs. Theoretical Applied Genetics 67:35-43.

BOTSTEIN, D.R.; WHITE, L.; SKOLNICK, M.H.; DAVIS, R.W. 1980. Construction of a genetic map in using restriction fragment length polymorphisms. American Journal of Human Genetics 32:314-331.

CARLSON, J.E.; TULSIERAM, L.K.; GLAUBITZ, J.C.; LUK, V.; KAUFFELDT, C.; RUTLEDGE, R. 1991. Segregation of random amplified DNA markers in F1 progeny of conifers. Theoretical Applied Genetics 83:194-200.

CHAPARRO, J.; WILCOX, P.; GRATTAPAGLIA, D.; O'MALLEY, D.; MCCORD, S.; SEDEROFF, R.; MCINTYRE, L.; WHETTEN, R. 1992. Genetic mapping of pine using RAPD markers: Construction of a 191 marker map and development of half sib genetic analysis. In Miami Winter Symposium: Advances in Gene Technology: Feeding the World in the 21st Century. Miami, Fla.

COUCH, J.A.; ZINTEL, H.A.; FRITZ, P.J. 1993. The genome of the tropical tree *Theobroma cacao* L. Molec. Gen. Genet. 237:123-128.

LANAUD, C.; LAURENT, V.; GORAN, J.N.; RISTERUCCI, A.M.; BOUET, A.; SOUNIGO, O. 1992. Assessment of the genetic diversity of cocoa using biochemical and molecular markers at CIRAD. In International Workshop Conservation, Characterization, and Utilization of Cocoa Genetic Resources for the 21st Century (1992, Port-of-Spain, Tri.). Proceedings.

LANDER, E.S.; GREEN, P.; ABRAHAMSON, J.; BARLOW, A.; DALY, M.J.; LINCOLN, S.E.; NEWBURG, L. 1987. MAPMAKER: An interactive computer package for construction of primary genetic maps of experimental and natural populations. Genomics 1:185-199

LERCETEAU, E.; CROUZILLAT, D.; PETIARD, V. 1992. Use of random amplified polymorphic DNA (RAPD) and restriction fragment length polymorphism (RFLP) to evaluate genetic variability within the *Theobroma* genus. In International Workshop Conservation, Characterization, and Utilization of Genetic Resources of Cocoa for the 21 st Century (1992, Port-of-Spain, Tri.). Proceedings.

PHILLIPS, W.; GALINDO, J.J. 1989. Método de inoculación y evaluación de la resistencia a *Phytophthora palmivora* en frutos de cacao (*Theobroma cacao*). Turrialba (C.R.) 39:488-497.

POUND, F. 1945. A note on the cocoa population of South America. In Research Conference (1945, London). Colonial 192:95-97.

REITER, R.S.; WILLIAMS, J.; FELDMAN, K.A.; RAFALSKI, J.A.; TINGEY, S.V.; SCOLNIK, P.A. 1992. Global and local mapping in *Arabidopsis thaliana* by using recombinant inbred lines and random amplified polymorphic DNAs. Proceedings of the National Academy of Science 89:1477-1481.

WELSH, J.; MCCLELLAND, M. 1990. Fingerprinting genomes using PCR with arbitrary primers. Nuclear Acids Research 18:7213-7218

WILDE, J.R.; WAUGH, R.; POWELL, W. 1992. Genetic fingerprinting of *Theobroma* clones using randomly amplified polymorphic DNA markers. Theoretical Aplied Genetics 83:871-877.

WILLIAMS, J.G.K.; KUBELIK, A.R.; LIVAK, K.J.; RAFALSKI, J.A.; TINGEY, S.V. 1990. DNA polymorphisms amplified by arbitrary primers are useful genetic markers. Nuclar Acids Research 18:6531-6535.

WOOD, G.A.R.; LASS, R.A. 1987. Cocoa. 4a ed. Longmans. New York.