Genome Size and DNA Polymorphism in Theobroma cacao

Antonio Figueira, Jules Janick, and Peter Goldsbrough

Department of Horticulture, Purdue University, West Lafayette, IN 47907-1165

Abstract. The size of the haploid genome of Theobroma cacao L. (cacao), estimated using laser flow cytometry, was 0.43 pg. An improved DNA extraction procedure was developed based on isolation of a crude nuclei preparation from leaf tissue that effectively eliminated contamination of the DNA by polysaccharides and produced DNA that was, on average, longer than 50 kb. DNA yields ranged from 2 to 10 µg·g−1 fresh weight of leaf tissue. DNA blot hybridization experiments with a flax (Linum usitatissimum L.) ribosomal DNA probe revealed restriction fragment length polymorphisms between three cacao genotypes. Differences between the DNA of these genotypes were also detected by polymerase chain reaction amplification of polymorphic DNA fragments, using random oligonucleotide primers.

Materials and Methods

Plant material. We used three distinct genotypes: 1) PU 1, a seedling of unknown pedigree from Centro Agronomic Tropical de Investigacion y Ensenanza (CATIE), Turrialba, Costa Rica, with intensely pigmented young leaves and flowers and very red immature pods; 2) PU 8, an open-pollinated seedling of unknown pedigree from, Belize; and 3) PU 50, a seedling of ‘Catongo’, a Brazilian selection obtained from CATIE, Costa Rica, with white cotyledons and albino flowers.

Laser flow cytometry. Healthy young green leaves were collected from maize (Zea mays L.), tobacco (Nicotiana tabacum L.), carnation (Dianthus caryophyllus L.), and cacao, washed with deionized water, blotted dry on paper towels, and weighed. The following procedures were performed in a room at 4°C. Midveins and petioles were removed, and 0.1 g of leaf tissue was placed in a 60 × 150-mm petri dish containing 1 to 2 ml of the lysis buffer described by Galbraith (1990). The leaves were chopped with a single-edged razor blade to release nuclei. The extracts were filtered through a 70-µm nylon screen. The internal standard of chicken red blood cells (20 µl of a phosphate buffered saline solution containing 10 000 cells/ml) was added. Nuclei were stained for 30 min by adding an equal volume of 4′,6-diamidino-2-phenylindole (DAPI) or propidium iodide. Before flow analysis, the samples were filtered through a 30-µm nylon screen. Flow cytometric analysis was performed at the Purdue Univ. Cytometry Laboratories, using an EPICS 753 (Coulter Corp., Hialeah, Fla.) with Laser Innova 90-5 Agron (200 mW light) (Coherent Laser Corp., Palo Alto, Calif.) operating at a wavelength of 488 nm and light power of 200 mW and with 488 Dichroic, 488 Laser-blocking, and 590 Longpass optical filters. Millipore-filtered, deionized water was used as sheath fluid. The data acquisition system employed was the Multiparameter Data Acquisition and Display System (MDADS) and the EPICS Elite workstation software (Coulter Corp.).

DNA extraction method. Young leaves were collected in the greenhouse, washed in deionized water, blotted dry, and weighed. Midveins and petioles were removed before they were frozen in liquid nitrogen. The frozen leaves were ground to a powder using a mortar and pestle in the presence of liquid nitrogen. Ice-cold extraction buffer consisting of 50 mM Tris pH 8.0, 5 mM MgCl2, 0.3 M sucrose, 0.1 M diethyldithiocarbamic acid, and 2% polyvinylpyrrolidone 10,000 was added in a proportion of 10 ml·g−1 fresh tissue. The following steps were performed on ice. The extract was homogenized with a Polytron CH-6010 (Kinematic GmbH, Lucern, Switzerland) four times for 30 to...
45 sec each with 60-sec intervals between cycles. The extract was filtered through two layers of cheesecloth under vacuum, followed by filtration through three layers of Miracloth (Calbiochem Corp., San Diego, Calif.). The filtrate was transferred to centrifuge tubes and spun at 350 × g for 15 min. The supernatant was discarded, and the crude nuclei pellet was resuspended in the smallest possible volume of extraction buffer. An equal volume of lysis buffer (100 mM Tris pH 8.0, 40 mM EDTA, 2% Sarcosyl) was added and the tubes were inverted gently several times. To each milliliter of lysate, 1.04 g of cesium chloride and 0.125 ml of ethidium bromide (10 mg·ml⁻¹) were added. Cesium chloride was dissolved before the addition of the ethidium bromide. The solution was spun for 10 min at 10,000 × g to remove insoluble material, then transferred to tubes for the Beckman VTi65.2 rotor (Palo Alto, Calif.). Tubes were spun at 50,000 rpm, 20°C, for 17 h. DNA was visualized under ultraviolet light, then transferred to a clean tube and centrifuged again at 55,000 rpm, 2°C, for 6 h. The DNA was removed from the second gradient, ethidium bromide was removed by repeated extractions with isopropanol saturated with 20 × SSC (1 × SSC is 0.15 M sodium chloride, 0.015 M sodium citrate, pH 7.0), and the solution was dialyzed overnight at 4°C against TE (10 mM Tris pH 7.5, 1 mM EDTA). Final DNA concentrations were measured spectrophotometrically.

DNA hybridization. DNA samples (2 µg) were digested for 4 h with 20 units of EcoRI, XbaI (Promega Corp., Madison, Wis.) and BamHI (New England Biolabs, Beverly, Mass.). The digested DNAs were separated on a 0.7% agarose gel using a voltage gradient of 2 V·cm⁻¹ in TAE buffer (40 mM Tris-base, 5 mM sodium acetate, 1 mM EDTA, pH 7.9). The gel was stained with ethidium bromide and DNA transferred to a nitrocellulose membrane (Schleicher & Schuell, Keene, N.H.) according to Southern (1975). The hybridization probe was pBG35, a plasmid containing a flax rDNA repeat (Goldsbrough and Cullis, 1981), and was labelled with ³²P-dCTP by random priming using a Decaprime kit (Ambion, Austin, Texas). The hybridization conditions were as described by Goldsbrough et al. (1990).

Random amplified polymorphic DNA (RAPD). Amplification reactions were performed as described by Williams et al. (1990) with minor modifications. The total reaction volume was 25 µl containing 10 mM Tris-HCl (pH 8.8), 50 mM KCl, 1.5 mM MgCl₂, 0.1% Triton-X, 100 µM each of dATP, dCTP, dGTP and dTTP, 0.2-µM primers (Operon, Alameda, Calif.), 25 ng of genomic DNA from PU 1, PU 50, or PU 8, and 1.5 units of Taq DNA polymerase (Promega Corp.). The primers used in this study had the following sequences: 1) GTGTGACGCAG; 2) CATTGACCACA; 3) GGAGGTGTGT; 4) CATCCTCCCTG; 5) GTAGACCGGT; 6) GATTCGCTCC. Amplification was conducted in a Perkin Elmer Cetus DNA ThermalCycler for 45 cycles of 1 min at 94°C, 1 min at 37°C, and 1 min at 72°C, followed by a final incubation at 72°C for 7 min, using the most rapid transition time between temperatures. Products were analyzed on 1.4% agarose gels run at 6.4 V·cm⁻¹ in TAE.

Results and Discussion

Laser-flow cytometry. The size of the cacao genome had not been previously determined. We adapted the flow cytometry methods of Galbraith (1990) and Michelson et al. (1991) to determine the DNA content of nuclei prepared from cacao leaf tissue. We used three unrelated species with known genome sizes as controls in these experiments (Table 1). The results we obtained for control species are similar to those reported in the literature, indicating the reliability of our method. Using this method, nuclei prepared from cacao leaves produced a single prominent peak corresponding to a nuclear DNA content of 0.85 pg (Fig. 1). We did not detect any additional peaks with lower fluorescence intensity, in particular those representing half the DNA content of the major peak. The absence of a 4C peak (1.7 pg) might be related to the developmental stage of the leaves that were analyzed. Young, light-green, fully expanded leaves were used, and cells in this tissue are likely to be in the postmitotic stage, with the nuclei in the G1 phase. Similarly, Arumuganathan and Earle (1991b) did not observe 4C peak in rice (Oryza sativa L.) or wheat (Triticum aestivum L.). We analyzed nuclei prepared from various cacao genotypes, but no significant differences in genome size were obtained. We concluded, therefore, that the haploid size of T. cacao genome is ≈0.43 pg (equivalent to 415,000 kbp). This genome size is only three times as large as that of the Arabidopsis thaliana (L.) Heynh. (Arumuganathan and Earle, 1991a), considered to have the smallest genome among higher plants.

According to Grime and Mowforth (1982) and Price (1988), tropical plants tend to have smaller genomes than those of plants that originated in temperature regions. Grime and Mowforth (1982) suggested that selection on genome size might be caused by a differential effect of temperature on cell division and cell expansion. Growth and development through cell division, as opposed to cell expansion, are favored at higher temperatures, as observed in cacao (Machado and Hardwick, 1988). This process may select for a shorter mitotic cycle and, therefore, for cells that are smaller and have smaller genomes than those of plants from temperate regions.

Further evidence for a small genome size in cacao comes from cytogenetic studies. The chromosome number of cacao has been determined to be 2 n = 20 (Munoz, 1948). Cytogenetic analysis was complicated by the small size of cacao chromosomes, ranging from 1.25 to 2.85 µm in length (Martison, 1975). Small chromosome size is in agreement with the relatively low nuclear DNA content of cacao. Additionally, cacao chromosomes were not susceptible to C-band staining (Glicenstein and Fritz, 1989), which is specific for heterochromatic regions. This result suggests that the cacao genome may contain a relatively low amount of repetitive DNA, similar to other plants that have a small genome (Flavell, 1980). For molecular genetic studies on cacao, its small genome will facilitate the isolation of genomic DNA sequences that encode specific genes.

DNA extraction method. To further characterize the cacao genome, we attempted to isolate DNA from cacao using several procedures. Standard methods of DNA isolation from cacao, using phenol extraction and alcohol precipitation, were unsuitable for this species as noted previously (Couch and Fritz, 1990; Febres, 1990). The ubiquitous presence of large amounts of polysaccharides (Figueira et al., 1991) that co-purify with the DNA, and the presence of polyphenolic compounds, make it very difficult to isolate DNA that is suitable for molecular analysis. DNA prepared from cacao by conventional methods could not be digested by restriction enzymes, and it migrated abnormally when separated by agarose gel electrophoresis. Couch and Fritz (1990) described a method that was designed to overcome the difficulties of isolating DNA from cacao. In our hands, however, this procedure did not give good results and was considered overly complex, requiring the use of a refractometer to adjust the cesium chloride concentration before buoyant density centrifugation. The method described here includes the use of diethylthiocarbamic acid and polyvinylpyrrolidone in the extraction buffer to reduce the production of polyphenol com-
Fig. 1. Determination of nuclear DNA content in cacao. Nuclei isolated from cacao leaves were stained with propidium iodide and analyzed by flow cytometry. Chicken red blood cells (CRBC) were included as an internal standard.

Table 1. Nuclear DNA content of various species determined by flow cytometry.

<table>
<thead>
<tr>
<th>Species</th>
<th>Nuclear DNA content (pg)</th>
<th>Reported 2C values (pg)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Dianthus caryophyllus</em></td>
<td>1.23 ± 0.05</td>
<td>1.16</td>
<td>Harkins &amp; Galbraith, personal communication</td>
</tr>
<tr>
<td><em>Nicotiana tabacum</em></td>
<td>9.11 ± 0.22</td>
<td>9.67</td>
<td>Galbraith et al., 1983</td>
</tr>
<tr>
<td><em>Zea mays</em></td>
<td>6.55 ± 0.25</td>
<td>5.37</td>
<td>Michaelson et al., 1991</td>
</tr>
<tr>
<td><em>Theobroma cacao</em></td>
<td>0.85 ± 0.07</td>
<td>--</td>
<td>Galbraith et al., 1983</td>
</tr>
</tbody>
</table>

pounds, as described by Couch and Fritz (1990). Our method is also based on the isolation of a crude nuclei preparation, which is critical as it separates the DNA from polysaccharides. The preparation of cesium chloride ethidium bromide gradients has also been simplified. This protocol has proven to be reliable and gives yields of 2 to 10 µg DNA/g leaf tissue. Although this is not a high yield, the DNA is of good quality, having an average size >50 kb, and is readily digested by several restriction enzymes (Fig. 2). The availability of fresh leaf tissue is not critical, as the method has been used successfully with leaf tissue that had been stored at −70°C for >6 months. The protocol is also suitable for isolating DNA from related *Theobroma* and *Herrania* spp.

**DNA hybridization.** We hypothesized that variation in repeated DNA sequences could be used to estimate genetic diversity in genotypes of *T. cacao* and in related species. Restriction enzyme digests of DNA from three cacao genotypes were hybridized with a rDNA probe from flax (Fig. 3, Table 2). In XbaI digests, this probe hybridized predominantly to fragments of 10.0 kb in PU 8, 10.8 kb in PU 50, and 9.3 and 10.0 kb in PU 1 (the two bands are discernible on shorter exposures of the blot to X-ray film). These fragments likely contain the majority of full-length rDNA repeat units in these genotypes. Additional minor bands hybridized in XbaI digests of each DNA. In EcoRI digests, a fragment of 4.0 kb hybridized in all genotypes tested. Similarly, fragments of 1.2, 1.5, and 2.9 kb hybridized in BamHI digests of each of the three genotypes. The restriction enzyme fragments that are common to each of the genotypes are probably derived from highly conserved sequences that encode the rRNAs. The remaining bands that hybridized likely include the more variable spacer sequences present in rDNA genes. Polymorphisms between genotypes are clearly present in these fragments. In PU 1, the EcoRI fragments of 5.2 and 5.8 kb are consistent with the two size classes of rDNA repeats produced by XbaI. However, a BamHI fragment corresponding to the smaller rDNA repeat in PU 1 was not detected. This discrepancy needs to be resolved. In addition to the presence of rDNA RFLPs, it is clear that there is also variation in rRNA gene copy number, with PU 50 (‘Catongo’) having significantly less rDNA than either PU 1 or PU 8. This is not due to errors in DNA loading, as shown by the photograph of the stained gel (Fig. 2), in which the polymorphic full-length rDNA repeats in XbaI digests can be discerned. The presence of RFLPs in cacao rDNA suggests that sufficient DNA sequence diversity exists in cacao genotypes to enable a genetic map of RFLP markers to be constructed for this species.

**Random amplified polymorphic DNA (RAPD).** Recently, a method has been described for identification of genetic poly-

![Fig. 2. Restriction enzyme digests of cacao DNA. Two micrograms of DNA isolated from PU 1 (A), PU 50 (B), and PU 8 (C) were digested with EcoRI, BamHI, or XbaI as indicated. The DNA samples were separated by electrophoresis in a 0.7% agarose gel, stained with ethidium bromide, and photographed under ultraviolet illumination. The positions of DNA size standards are shown on the left.](image-url)
Fig. 3. Detection of rDNA sequences in cacao-DNA. Two-microgram samples of DNA from PU 1 (A), PU 50 (B), and PU 8 (C) were digested with EcoRI, BamHI, or XbaI, separated by electrophoresis in a 0.7% agarose gel, and transferred to a nitrocellulose membrane. pBG35, a plasmid containing a flax rDNA repeat, was labeled with 32P-dCTP and hybridized to the membrane. An autoradiogram of the membrane after hybridization and washing is shown. The positions of DNA size standards are shown on the left.

Table 2. Sizes of restriction enzyme fragments that hybridized to the flax rDNA probe.

<table>
<thead>
<tr>
<th>Restriction enzyme</th>
<th>Genotype</th>
<th>Fragment size (kb)</th>
<th>Estimated rDNA repeat length</th>
</tr>
</thead>
<tbody>
<tr>
<td>EcoRI</td>
<td>PU 1</td>
<td>5.8/5.2*</td>
<td>9.8/9.2*</td>
</tr>
<tr>
<td></td>
<td>PU 50</td>
<td>6.4</td>
<td>10.4</td>
</tr>
<tr>
<td></td>
<td>PU 8</td>
<td>6.0</td>
<td>10.0</td>
</tr>
<tr>
<td>BamHI</td>
<td>PU 1</td>
<td>5.2/2</td>
<td>10.8/2</td>
</tr>
<tr>
<td></td>
<td>PU 50</td>
<td>5.8</td>
<td>11.4</td>
</tr>
<tr>
<td></td>
<td>PU 8</td>
<td>5.3</td>
<td>10.9</td>
</tr>
<tr>
<td>XbaI</td>
<td>PU 1</td>
<td>10.0/9.3*</td>
<td>10.0/9.3*</td>
</tr>
<tr>
<td></td>
<td>PU 50</td>
<td>10.8</td>
<td>10.8</td>
</tr>
<tr>
<td></td>
<td>PU 8</td>
<td>10.0</td>
<td>10.0</td>
</tr>
</tbody>
</table>

*Indicates the restriction enzyme fragments in PU 1 that gives rise to the shorter rDNA repeat in this genotype. The expected BamHI fragment of ~4.8 kb in PU 1 was not present.

Fig. 4. DNA polymorphisms between cacao genotypes detected by RAPD analysis. Samples (25 ng) of DNA from PU 1 (A), PU 50 (B), and PU 8 (C) were used as templates for polymerase chain reaction (PCR) amplification with six oligonucleotides primers (1 to 6); the sequences of the primers are given in Material and Methods. The PCR products were separated by electrophoresis in a 1.4% agarose gel, stained with ethidium bromide, and photographed. Lane M contains DNA size standards, and the size of these markers are shown on the left.

morphisms by polymerase chain reaction (PCR) amplification of DNA fragments using random primers (Welsh and McClelland, 1990; Williams et al., 1990). We have used this method to determine if such polymorphisms can be detected in DNA isolated from the three cacao genotypes investigated by us. RAPD reaction conditions were optimized in preliminary experiments (data not shown). The effects of varying the following conditions were examined: concentrations of template DNA, oligonucleotide primers, and Mg. Use of >50 ng of cacao DNA in 25 µl reaction produced no amplified DNA bands, whereas products were consistently detected using from 5 to 50 ng of DNA. We concluded that 25 ng (1 ng·µl⁻¹) of cacao DNA was suitable for the RAPD assay, as recommended for other species. Use of primers below a concentration of 0.2 µm gave little amplification. Increasing the Mg concentration to >1.5 mM produced an overall decrease in the number and abundance of amplified DNA bands. We have also shown that identical PCR products are obtained when independent DNA preparations from the same cacao genotype are used as templates.

Most of the primers that we tested produced few amplified fragments that were detected by gel electrophoresis and staining with ethidium bromide (Fig. 4). Typically, one or two predominant fragments were amplified, as well as some minor fragments. For some primers (e.g., 3 and 6) the pattern of amplified DNAs was identical for each genotype. However, other primers produced polymorphisms, detected by amplification of DNA fragments that varied in size depending on the cacao genotype used. Using primers 1, 2, and 5, genotypes PU 50 and PU 8 produced very similar amplified products, whereas genotype PU 1 produced a major band that was not present in either PU 50 or PU 8. With primer 4, the sizes of the amplified DNA products were the same in all three genotypes, but the highest molecular weight band of ~1.7 kb was consistently more intense in PU 8 than PU 1 or PU 50. One explanation for this result is that the amplified 1.7-kb fragment in PU 50 presents a polymorphism that co-migrates with a minor fragment common to all genotypes.

On the basis of this preliminary analysis, it appears that genotypes PU 50 and PU 8 are more similar to each other than to PU 1. Out of a preliminary screen of 20 primers, every primer produced amplified DNA products. Of these, six primers produced unambiguous polymorphisms with two to six polymorphic bands. As only three genotypes have been tested in this initial screen, more polymorphisms maybe detected using more cacao genotypes. Less time-consuming methods of DNA isolation can now be evaluated for their applicability to RAPD analysis.

It has been demonstrated in other species that RAPD markers are inherited as simple dominant Mendelian genes and have been used to develop genetic linkage maps (Welsh and McClelland, 1990; Williams et al., 1990). RAPD markers have also been identified that are linked to genes for disease resistance in tomato (Lycopersicon esculentum Mill.) (Martin et al., 1991). The results presented here indicate that the RAPD tech...
nique may be a useful tool for analysis of genetic diversity in cacao and identification of molecular markers that are linked to traits of agronomic importance.

**Literature Cited**


