Mol Gen Genet (1993) 237:123–128

MGG

The genome of the tropical tree *Theobroma cacao* L.

Jennifer A. Couch[1,*], Harold A. Zintel[2,], and Paul J. Fritz[1]**

[1] Intercollege Program in Genetics and Department of Food Science, and
[2] Department of Engineering Science and Mechanics, Pennsylvania State University, University Park, PA 16802 USA

Received April 7, 1992 / Accepted August 20, 1992

Summary. 2C values for angiosperms vary over 2500-fold and a positive correlation exists between C-value and latitude in herbaceous plants. Woody plants differ from herbaceous plants in chromosome size and C-value. In addition, tropical hardwoods have smaller chromosomes than other tropical plants and do not share the correlation of minimum generation time with genome size seen in herbaceous plants. *Theobroma cacao* is a tropical hardwood cultivated for its beans, which are used to make chocolate and cocoa butter. Its cytology is typical of the pantropical and subtropical family Sterculiaceae. Its small chromosomes, single secondary constriction, and lack of C-banding suggest a small genome. The genome size of *T. cacao,* measured by reassociation kinetics, is 2.01×10^8, which is small compared to both temperate and tropical plants previously studied. We also provide data on the melting point, base composition, and relative extent of methylation (at sites most commonly methylated in higher plants), of *T. cacao* DNA.

Key words: *Theobroma cacao* L. – Genome size – Tropical plants – Reassociation kinetics – 5-methylcytosine

Introduction

The genome size of angiosperms varies over 2500-fold (Bennett 1987). Where polyploidy is not involved, most variation in genome size is due to variation in the amount of repetitive DNA (Flavell et al. 1974). In herbaceous plants, DNA content per cell correlates positively with minimum generation time (Bennett 1972), nuclear size and cell size (Price et al. 1973). In addition, DNA content per cell also varies with latitude: generally as latitude increases genome size increases (both within and between genera and families: Bennett et al. 1982; Bennett 1976; Levin and Funderberg 1979; Grime and Mowforth 1982; Bennett 1987; Essad 1988), except in the Antarctic where the cline is reversed (Bennett et al. 1982). Temperate plants are found to have more DNA per diploid genome and larger chromosomes than tropical plants, even though the mean number of chromosomes in the two categories is nearly the same (Levin and Funderberg 1979). Most surveys of DNA content per cell, comparing temperate and tropical plants, have included only herbaceous plants (Bennett 1976; Levin and Funderberg 1979). Woody dicots exhibit smaller genome sizes than herbaceous angiosperms (Grime and Mowforth 1982; Ohri and Khoshoo 1987; Arumugantan and Earle 1991) or gymnosperms (Ohri and Khoshoo 1986). The 2C values for several *Ficus* species, reported by Ohri and Khoshoo (1987), and the genome size of *Hevea brasiliensis,* determined by reassociation kinetics (Chon, personal communication) are smaller than most values reported for tropical plants. Furthermore, the small chromosome sizes seen in most tropical hardwoods suggest that these plants may also have smaller genomes (Mehra 1976).

Tropical hardwoods differ in another respect: the correlation between genome size and minimum generation time (Bennett et al. 1982) does not appear to be true for tropical hardwoods (Ohri an Khoshoo 1987). The correlation between nuclear size and cell size also does not hold, since Mehra and Bawa (1969) have found polyploid series in woody taxa which tolerate different cell-to-nucleus volume ratios.

The vast majority of data collected in these surveys is in the form of DNA content per cell, measured by cytophotometry. Cytophotometry does not indicate genomic complexity and, in addition, cytophotometric data can be severely distorted by other cellular components (Greilhuber 1986). For these reasons, we chose to characterize the genome of *Theobroma cacaco* using reassociation kinetics.

T. cacao is a tropical tree grown for its beans, which are used to make chocolate and cocoa butter. Although

* *Present address:* Department of Botany, University of British Columbia, Vancouver, British Columbia, V6T–1Z4 Canada

** *Present address:* 2406 #2 West 5th Avenue, Vancouver, British Columbia, V6K–1S8 Canada

Correspondence to: J. Couch

it has been cultivated for centuries, breeding programs have met with only limited success (Toxopeus 1969, 1985). Efforts are underway to study *T. cacao* at the molecular level with the goal of improving the plant through genetic engineering and assisting breeders by providing a genomic map and polymorphism studies.

The cytology of *T. cacao* is typical of the pantropical and subtropical family Sterculiaceae (Mehra 1976). It is a diploid (Glicenstein and Fritz 1989) with 20 small chromosomes (Martinson 1975), which have only one secondary constriction, and do not C-band (Glicenstein and Fritz 1989), all of which suggest a small genome.

In addition to genome size and complexity measurements we provide here data on the melting point, base composition, and extent of methylation at sites most commonly methylated in higher plants (Gruenbaum et al. 1981).

Materials and methods

Biological material. T. cacao leaves were obtained from mature trees, grown from seed, in the ACRI Cocoa Molecular Biological Laboratory greenhouse at the Pennsylvania State University. *Arabidopsis thaliana* strain Columbia was grown froom seed (a gift from Dr. J. Chory). *Escherichia coli* type B, ultrapure grade, and herring sperm DNA, sodium salt, were purchased from Sigma. Raw wheat germ was purchased from a local supplier, Frankferd Farms Foods.

DNA isolation. T. cacao DNA was isolated from immature leaves, by the method of Couch and Fritz (1990). *A. thaliana* DNA and wheat germ DNA were isolated by the methods described by Leutwiler et al. (1984). All DNAs destined for reassociation experiments were extensively purified by two sequential CsCl density centrifugations and by filtration through a 1 m Sephadex G–50 column supported by fragmented glass. DNAs destined for other experiments, were found to be of sufficient purity after only one CsCl density centrifugation step.

Shearing and labelling of DNA. Unlabelled *T. cacao, A. thaliana, E. coli,* and herring sperm DNAs in TE buffer (10 mM TRIS–HCl, 1 mM EDTA pH 8.0), at concentrations of from 40 to 1000 µg/ml, were sonicated for 3 min, 50% cycle at power setting 5, using a Branson Sonic Power Sonifier, model 350 (tip vibrating at 20 kHz). The average single-stranded length was determined by electrophoresis through a 2% agarose gel under denaturing conditions (Sambrook et al. 1989). The gel was neutralized with 1 M TRIS–HCl, pH 7.6 and 1.5 M NaCl, and stained with ethidium bromide (30 min, 0.5 µg/ml ethidium bromide). The fragments averaged 450 bp in length.

Portions of the *T. cacao, A. thaliana* and *E. coli* samples were labelled by nick translation with [^{3}H] or [^{35}S]-dATP, combined with their corresponding unlabelled DNAs and purified through chelating resin (Sigma) columns prepared according to manufacturer's specifications. They were then exchanged into 0.12 M or 0.48 M sodium phosphate buffer (PB), pH 6.8, and concentrated over Centricon 30 filters (Amicon). Concentrations of DNAs were determined by absorbance at 260 nm.

DNA reassociation. Reassociation reactions were performed essentially under the conditions recommended in Britten et al. (1974). Reactions were performed in 0.12 M PB, pH 6.8 or 0.48 M PB, pH 6.8. ^{35}S-labelled *T. cacao* DNA was reassociated with and without an internal control (^{3}H-labelled *E. coli* DNA). DNA concentrations ranged from 1700 µg/ml to 0.5 µg/ml. ^{35}S-labelled *A. thaliana* DNA was run as a parallel control with and without, as internal standard, ^{3}H-labelled *E. coli* DNA. DNA solutions were sealed in capillary tubes or Safelock microfuge tubes (Eppendorf), boiled for 2 min (capillaries) or 10 min (microfuge tubes), and incubated at 60° C for times varying from a few minutes to a few days. Reassociated samples were removed from the 60° C bath and frozen in a dry ice-ethanol bath or directly in the −80° C freezer. Samples were stored at −80° C prior to analysis.

Reassociated samples were separated into single- and double-stranded DNAs over 0.5 ml hydroxyapatite (Calbiochem fast flow or Biorad HTP DNA grade) columns. The columns were run with 0.12 M PB, with and without 0.2% SDS added, at 60° C to collect single-stranded DNA, followed by heating to 100° C to denature and collect DNA that had reassociated. Samples in 0.48 M PB were diluted with water to 0.12 M PB prior to loading. All samples were diluted to 0.5 ml or 1.0 ml with 0.12 M PB and run with 75 µg sonicated herring sperm DNA per column. Aliquots of all fractions eluted were counted in Ecoscint H (National Diagnostics) in a Beckman LS1701 Scintillation counter, utilizing the dual label dpm option or the cpm option (for single label experiments). Calbiochem HAP columns run in 0.2% SDS/0.12 M PB generally gave the best results.

Calculations and curve fitting. C_0t values were calculated by multiplying the concentration (moles of nucleotide/liter) by time in sec. The percentage single-stranded DNA (%SS) was calculated by dividing the total dpm (or cpm) eluted as SS DNA by the total dpm (or cpm) eluted as SS and DS. All data sets were plotted. The final curves used are fit to the two best *T. cacao* data sets (least scatter) merged, and to the best single *A. thaliana* and *E. coli* data sets.

The %SS and C_0t values were fit to the equation

$$\frac{C}{C_0} = \frac{f_1}{1+k_1C_0t} + \frac{f_2}{1+k_2C_0t} \ldots + \frac{f_n}{1+k_nC_0t} + F$$

where C is concentration of single stranded DNA when the reassociation was ended (time t), C_0 is the original DNA concentration, and k_1–k_n are the second-order rate constants, f_1–f_n are the fractions represented by the individual kinetic components and F is the fraction of the original sample that was excluded from reassociation.

The results were analyzed by the Levenberg-Marquardt method, an algorithm which minimizes the rms of the residues, a method similar to the method of Pearson et al. (1977). The following assumptions are

made: firstly, the error in %SS is far greater than error in C_0t and secondly, the deviation of the data is independent of C_0t. A Monte Carlo technique is then used to estimate 95% confidence limits on the parameters of the curve fit due solely to measurement errors. These residues are then randomly added to the original curve fit to produce a new data set which is again fit with a curve (bootstrapping). By repeating this process 1000 times a distribution of values for the parameters of the curve is obtained. Upper and lower confidence limits are then estimated by sorting each parameter in descending order and recording the 25th and 975th values.

The complexities are calculated for each curve fit from k_3 (the reassociation constant from the single copy component of the curve), where

$$\frac{1}{k} = C_0t_{\frac{1}{2}}.$$

The genome size is calculated by comparing the unknown genome size in the following ratio to the known *E. coli* genome size and its $Cot_{1/2}$ under identical conditions.

$$\frac{C_0t_{\frac{1}{2}}\ E.\ coli}{C_0t_{\frac{1}{2}}\ \text{Cocoa}} = \frac{\text{Complexity } E.\ coli}{\text{Complexity Cocoa}}.$$

For eukaryotic DNA, the $Cot_{1/2}$ is the C_0t value at which half of the last component, the single copy component, would reassociate if it were the only component reassociating. This is computed by the following equation

$$C_0t_{\frac{1}{2}}\ \text{Cocoa} = \left(\frac{1}{k_3}\right)\left(\frac{f_3}{1-F}\right).$$

The complexity estimates are also calculated 1000 times and sorted according to size with the 25th and 975th taken as our 95% confidence limits.

Determination of melting point and nucleotide composition. Whole (unsheared) *T. cacao* DNA (12.5 µg/ml) in 0.12 M PB was heat-denatured at a rate of 1.5° C/min in a sealed cuvette in a Beckman DU Series 60 Spectrophotometer equipped with an auto-6-sampler water-jacketed cuvette holder. The temperature was monitored with a digital thermometer probe inserted into a separate cuvette containing only 0.12 M PB. The relative absorbance at 260 nm was plotted against temperature. The melting temperature (T_m) is the point at which one half of the DNA has denatured, as measured by hyperchromicity at 260 nm. *E. coli* DNA was denatured simultaneously in a separate cuvette as a control. The base composition, %(G+C), was calculated from T_m using the equation of Felsenfeld (1971): $\%(G+C) = 2.44\ (T_m - 69.35)$.

Determination of relative extent of methylation. Whole (unsheared) *T. cacao* DNA was restinction digested with two pairs of isoschizomers: *Hpa*II/*Msp*I and *Eco*RII/*Bst*NI. *Hpa*II and *Msp*I cleave the site CCGG; only *Msp*I will cleave the site if the internal C is methylated. *Eco*RII and *Bst*NI cleave the site CCA/TGG; only BstNI will cleave the site if the internal C is methylated. Wheat germ DNA was digested simultaneously as a control: wheat germ DNA is methylated at 82% of the CCGG sites and at 90% of the CCA/TGG sites (Gruenbaum et al. 1981). All digests were fractionated on a 1% agarose gel and stained with ethidium bromide (Sambrook et al. 1989).

Results

Melting point and base composition of T. cacao *DNA*

The melting temperature of *T. cacao* DNA was found to be 84.1° C ± 0.5° C. No hyperchromicity was observed at temperatures lower than the denaturation transition, indicating that no RNA was presen. *E. coli* DNA run as an adjacent control had a melting temperature of ca. 93° C, the exact melting point could not be determined because the equipment used only reached 95° C, which is not high enough to complete a melting curve for *E. coli*. The base composition of *T. cacao* is 36% (G+C) which is within the range for angiosperms, 35.6 to 41.4% G+C (Leutwiler et al. 1984).

Relative extent of methylation

The relative extent of cytosine methylation can be seen from the relative levels of digestion by methylation-sensitive and -insensitive isoschizomers. Fig. 1 shows the

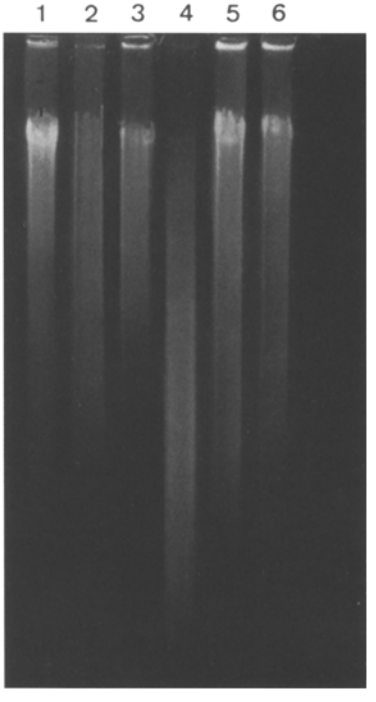

Fig. 1. Extent of methylation of DNA at the site CCA/TGG detected by restriction enzyme digestion. The isoschizomers *Bst*NI and *Eco*RII cleave this site; only *Bst*NI will cleave if the internal cytosine is methylated. DNA was purified and digested as described in Materials and methods. Wheat DNA, which is 90% methylated at this site, is used as a control. Lane 1, undigested *Theobroma cacao* DNA; lanes 2 and 3, *T. cacao* DNA digested with *Bst*NI and *Eco*RII, respectively. Lanes 4 and 5 contain wheat DNA digested with *Bst*NI and *Eco*RII, respectively; lane 6, undigested wheat DNA

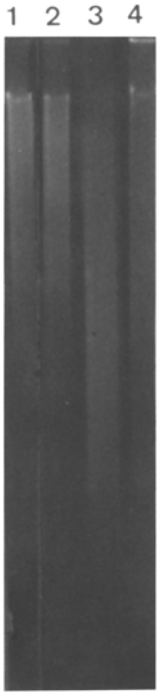

Fig. 2. Extent of methylation of DNA at the site CCGG detected by restriction enzyme digestion. The isoschizomers *Hpa*II and *Msp*I cleave this site, but only *Msp*I will cleave if the internal cytosine is methylated. Wheat DNA, which is methylated at 82% of these sites, is used as a control. Lanes 1 and 2, *Theobroma cacao* DNA digested with *Hpa*II and *Msp*I, respectively. Lanes 3 and 4 contain wheat DNA digested with *Msp*I and *Hpa*II respectively. DNAs used in the digestions in Fig. 1 and 2 were from the same batches

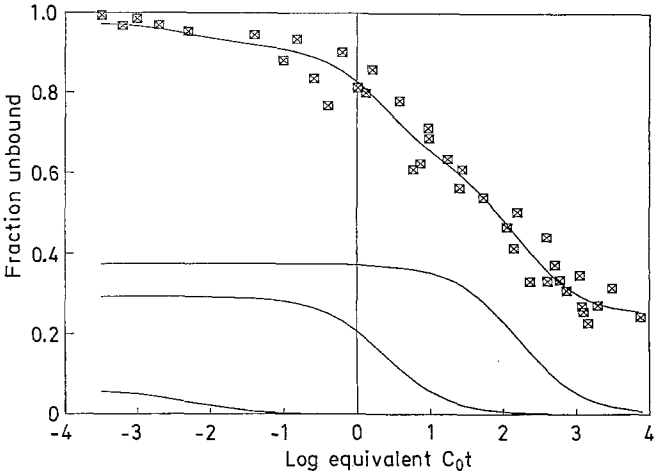

Fig. 3. Reassociation kinetics of *Theobroma cacao* DNA. DNA fragments of 450 bp were denatured, reassociated to the indicated equivalent C_0t values, and fractionated by hydroxyapatite as described in Materials and methods. The upper curve represents a nonlinear least squares fit for the three second-order components, as described in Materials and methods. The lower curves represent computer calculated second-order solutions for each individual component

Table 1. Reassociation parameters for *Theobroma cacao* DNA and *Arabidopsis thaliana* DNA

T. cacao	Component	Fraction	k	Complexity	Average no. of repeats
Nominal	Highly repetitive	(0.080)	(200)	(4.88×10^2)	(32703)
High	Middle repetitive	0.536	1.18	5.54×10^6	164
Nominal	mid. rep.	0.391	0.420	1.14×10^6	68
Low	mid. rep.	0.266	0.129	2.73×10^5	29
High	Single copy	0.615	0.014	2.67×10^8	1
Nominal	Single copy	0.498	0.0061	1.00×10^8	1
Low	Single copy	0.338	0.0018	4.34×10^7	1
A. thaliana					
Nominal	High rep.	(0.073)	(8.0)	(1.12×10^4)	(1177)
High	Middle rep.	0.414	32.05	2.52×10^5	5039
Nominal	mid. rep.	0.333	7.96	5.12×10^4	1172
Low	mid. rep.	0.252	2.09	9.97 ± 10^3	304
High	Single copy	0.525	0.0219	3.64×10^8	1
Nominal	Single copy	0.450	0.0068	8.10×10^7	1
Low	Single copy	0.373	0.0014	1.82 ± 10^7	1

Reassociation parameters were determined by the techniques described in Materials and methods. Values are given as nominal with high and low representing the 95% confidence limits. Values in parentheses were fixed due to insufficient data because the initial rate of reassociation is too rapid

digestion patterns of *T. cacao* and wheat DNAs by the isoschizomers *Msp*I and *Hpa*II. There is very little difference in the lengths of the *T. cacao* fragments obtained after digestion with either enzyme, compared to the large differences observed for wheat. The same is true of the *Bst*NI and *Eco*RII digests shown in Fig. 2. It should be noted that none of the four enzymes tested digested *T. cacao* well. However this same DNA preparation was readily digestible with other restriction endonucleases (data not shown). We conclude that *T. cacao* is not heavily methylated at either of the sites typically methylated in higher plants.

Reassociation kinetics

T. cacao reassociation data was fit with 2, 3 and 4 components. The root mean square error decreased from 0.0058 to 0.0043 as n increased from 2 to 3. A further increase in n did not decrease the rms error and also resulted in negative f values which are not possible in this model. Thus, the genome is best modeled as consisting of 3 components. Since there is no significant change with more than 3 components, the 3-component fit is used. Table 1 shows the complexities, fractions, reassociation constants and repetition frequencies for the three com-

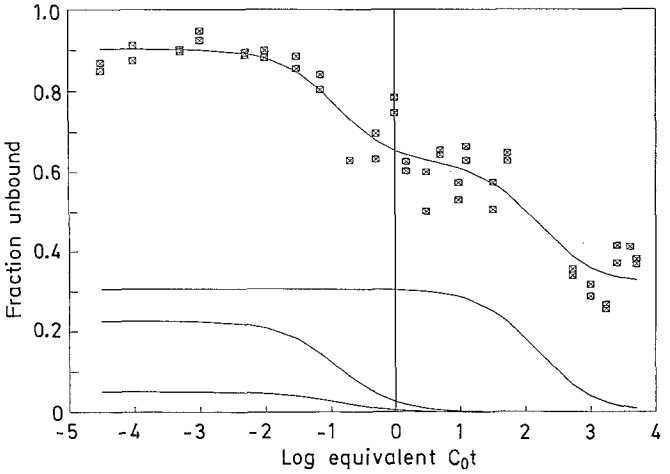

Fig. 4. Reassociation kinetics of *Arabidopsis thaliana* DNA. *A. thalian* DNA was treated in the same manner as *Theobroma cacao* DNA, as described in Materials and methods. The upper curve represents a nonlinear least squares curve fit for the three second-order components and the lower curves represent the computer calculated solutions for each individual component

ponents of *T. cacao* and *A. thaliana* under our conditions. Values are shown as nominal with high and low values indicating the 95% confidence limits. Figs. 3 and 4 show the reassociation curve fits for *T. cacao* and *A. thaliana,* respectively. Values in parentheses were fixed due to insufficient data (reassociation is too fast initially). Our estimated genome size for *A. thaliana* of 8.73×10^7 (with 95% confidence limits of 1.96×10^7 and 3.93×10^8) agrees well with the value of 7×10^7 reported by Leutwiler et al. (1984). *A. thaliana* was chosen as a control for the reassociation because it is an angiosperm with a genome size in the range expected for *T. cacao*. The *T. cacao* genome size is 2.01×10^8 with confidence limits of 7.95×10^7 and 6.77×10^8.

Discussion

The melting temperature and %(G+C) of *T. cacao* fall within standard range for angiosperms. These parameters will be useful in the continuing molecular characterization of the *T. cacao* genome.

The low degree of methylation found at the two sites analyzed in *T. cacao* is unusual for an angiosperm. It is possible that *T. cacao,* like *A. thaliana* an angiosperm with a very small amount of highly repetitive DNA, has a low amount of cytosine methylation due to its low level of repetitive DNA. Alternatively *T. cacao* DNA may be methylated at sites other than CCGG or CCA/TGG, which would make it atypical, with respect to other higher plants studied. In either case, *T. cacao* is unusual in its lack of methylation at the sites tested.

The most interesting result presented is the genome size: the *T. cacao* genome size of 2.01×10^8 bp is consistent with earlier estimates (Couch and McHenry, unpublished data) and the data of Figuera (personal communication) who calculated the genome size to be 0.43 pg per haploid nucleus by flow cytometric measurements. In addition, the low level of repetitive DNA is expected from the lack of chromosome C-banding and the single secondary constriction per genome. While it is possible that some of the middle repetitive DNA detected is due to chloroplast DNA contamination, we consider it unlikely to be a major component for the following reasons. Firstly, the DNA isolation technique enriches for nuclei (Couch and Fritz 1990). Secondly, there are only three chloroplasts per cell in *T. cacao* mesophyll (Baker et al. 1975), and the yield of chloroplast DNA per gram of leaf, using standard techniques, is extremely low (Yeoh et al. 1989).

The *T. cacao* genome is small compared to those of other plants studied, both temperate and tropical. However, since the majority of tropical plants for which 2C values are known are herbaceous and since tropical hardwoods are known to possess smaller chromosomes (Mehra 1976) which generally correlate with DNA content (Levin and Funderberg 1979), the *T. cacao* results may be part of a general trend and not an anomaly among tropical hardwoods.

In addition, the correlation between minimum generation time and genome size (Bennett 1972) observed for herbaceous plants does not appear to be true for woody dicots (Ohri and Khoshoo 1987). *T. cacao,* with a very small genome, has a lengthy minimum generation time of 2–7 years, where the 2-year value is for a man-made cross (Osei, personal communication).

Clearly more studies of tropical hardwood genomes need to be performed before any general conclusions can be drawn. But, it appears that tropical hardwoods are different from both temperate plants and tropical herbaceous plants in the size and complexity of their genomes. Also, some of the correlations observed in herbaceous plants between cell size and DNA content and between genome size and minimum generation time appear not to apply to them.

In addition to the data on *T. cacao* we have introduced a new variation on the method of fitting curves to reassociation data. The method used in this paper differs from the method of Pearson et al. (1977) in that the confidence limits on the curve fit parameters are more accurately evaluated. Confidence limits can be approximated with the Pearson et al. (1977) method if a normal distribution is assumed. The 95% confidence limits are then taken as approximately plus and minus 2 standard deviations. Often this type of approach will result in parameters with negative lower limits which are not realistic. In fact the distribution of the *T. cacao* residues is leptokurtic and slightly skewed. This, combined with the non-linear nature of the curve that is used to fit the data, results in variations in the parameters which are not normal. The technique presented here is simple and takes little time when run on a personal computer. A copy of the program is available on request from J. Couch.

Acknowledgements. The authors wish to thank Dr. H. Kriebel for his generous gift of the hydroxyapatite and helpful suggestions and Dr. J. Kamalay for helpful discussions. This work supported in part by an endowment from the American Cocoa Research Institute.

References

Arumugantan K, Earle ED (1991) Nuclear DNA content of some important plant species. Plant Mol Biol Reporter 9:208–218

Baker NR, Hardwick K, Jones P (1975) Biochemical and physiological aspects of leaf development in cocoa *(Theobroma cacao)* II. Development of chloroplast ultrastructure and carotenoids. New Phytologist 75:513–518

Bennett MD (1972) Nuclear DNA content and minimum generation time in herbaceous plants. Proc R Soc Lond Ser B 181:109–135

Bennett MD (1976) DNA amount, latitude, and crop plant distribution. Environ and Exp Botany 16:93–108

Bennett MD (1987) Ordered disposition of parental genomes and individual chromosomes in reconstructed plant nuclei, and their implications. Somatic Cell Mol Genet 13:463–466

Bennett MD, Smith JB, Smith RI (1982) DNA amounts of angiosperms from the Antarctic and South Georgia. Environ Exp Botany 22:307–318

Britten RJ, Graham DE, Neufeld BR (1974) Analysis of repeating DNA sequences by reassociation. Methods Enzymol 29E:383–405 Academic Press, New York

Couch JA, Fritz PJ (1990) Isolation of DNA from plants high in polyphenolics. Plant Mol Biol Reporter 8:8–12

Essad S (1988) Mise en évidence de variations saltatoires de l'ADN nucléaire dans et entre les espéces du genre Medicago L. Genome 30:825–833

Felsenfeld G (1971) Analysis of temperature-dependent absorption spectra of nucleic acids. In: Cantoni GL, Davies DR (eds) Procedures of nucleic acids research, vol 2. Harper and Row, New York, pp 233–244

Flavell RB, Bennett MD, Smith JB, Smith DB (1974) Genome size and the proportion of repeated nucleotide sequence DNA in plants. Biochem Genet 12:257–269

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

Greilhuber J (1985) Severely distorted Feulgen-DNA amounts in *Pinus* (Coniferophytina) after nonadditive fixations as a result of meristematic self-tanning with vacuole contents. Can J Genet Cytol 28:409–415

Grime JP, Mowforth MA (1982) Variation in genome size – an ecological interpretation. Nature 299:151–153

Gruenbaum Y, Naveh-Many T, Cedar H, Razin A (1981) Sequence specificity of methylation in higher plant DNA. Nature 292:860–862

Leutwiler LS, Hough-Evans BR, Meyerowitz EM (1984) The DNA of *Arabidopsis thaliana*. Mol Gen Genet 194:15–23

Levin DA, Funderberg SW (1979) Genome size in angiosperms: temperature versus tropical species. Amer Nat 114:784–795

Martinson VA (1975) Cytological studies of diploid and tetraploid *Theobroma cacao*. Genetica 45:341–348

Mehra PN (1976) Cytology of Himalayan hardwoods. Sree Saraswaty Press Ltd, Calcutta (India)

Mehra PN, Bawa KS (1969) Chromosomal evolution in tropical hardwoods. Evolution 23:466–481

Ohri D, Khoshoo TN (1986) Genome size in gymnosperms. Plant Syst Evol 153:119–132

Ohri D, Khoshoo TN (1987) Nuclear DNA content in the genus *Ficus* (Moraceae). Plant Syst Evol 156:1–4

Pearson WR, Davidson EH, Britten RJ (1977) A program for least squares analysis of reassociation and hybridization data. Nucleic Acids Res 4:1727–1735

Price HJ, Sparrow AH, Nauman AF (1973) Correlations between nuclear volume, cell volume and DNA content in meristematic cells of herbaceous angiosperms. Proc Roy Soc Lond Ser B 202:517–525

Sambrook J, Fritsch Ef, Maniatis T (1989) Molecular cloning: a laboratory manual, 2nd edn, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York

Toxopeus H (1969) Breeding. In: Ferwerda FP, Wits F (eds) Cacao. Outlines of perennial crop breeding in the tropics, pp 92–109

Toxopeus H (1985) Planting material. In: Wood GAR, Lass RA (eds) Cocoa. 4th edn. Longman Scientific and Technical Publishers, Essex (England), pp 80–92

Yeoh HH, Chung DK, Fritz PJ (1989) *Theobroma cacao* chloroplast DNA: isolation, molecular cloning and characterization. Cafe, Cacao and The XXXIV:173–178

Communicated by M. Meyerowitz