

RESEARCH ARTICLE

# Genome size, cytogenetic data and transferability of EST-SSRs markers in wild and cultivated species of the genus *Theobroma* L. (Byttnerioideae, Malvaceae)

Rangeline Azevedo da Silva<sup>1,2<sup>¶a</sup></sup>, Gustavo Souza<sup>3</sup>, Lívia Santos Lima Lemos<sup>2<sup>¶b</sup></sup>, Uilson Vanderlei Lopes<sup>2</sup>, Nara Geórgia Ribeiro Braz Patrocínio<sup>1,2</sup>, Rafael Moysés Alves<sup>4</sup>, Lucília Helena Marcellino<sup>5</sup>, Didier Clement<sup>2,6</sup>, Fabienne Micheli<sup>1,6</sup>, Karina Peres Gramacho<sup>2\*</sup>

**1** Universidade Estadual de Santa Cruz, Ilhéus, Bahia, Brazil, **2** Cocoa Research Center, CEPLAC/CEPEC, Itabuna, Bahia, Brazil, **3** Universidade Federal de Pernambuco, Recife, Pernambuco, Brazil, **4** Embrapa Amazônia Oriental, Belém, Pará, Brazil, **5** Embrapa Recursos Genéticos e Biotecnologia, Distrito Federal, Brazil, **6** CIRAD, UMR AGAP, Montpellier, France

<sup>¶a</sup> Current address: Instituto Fundação Oswaldo Cruz, Instituto Oswaldo Cruz, Rio de Janeiro, Rio de Janeiro, Brazil

<sup>¶b</sup> Current address: Universidade Federal do Sul da Bahia, Teixeira de Freitas, Bahia, Brazil.

\* [gramachokp@hotmail.com](mailto:gramachokp@hotmail.com)



**OPEN ACCESS**

**Citation:** da Silva RA, Souza G, Lemos LSL, Lopes UV, Patrocínio NGRB, Alves RM, et al. (2017) Genome size, cytogenetic data and transferability of EST-SSRs markers in wild and cultivated species of the genus *Theobroma* L. (Byttnerioideae, Malvaceae). PLoS ONE 12(2): e0170799. doi:10.1371/journal.pone.0170799

**Editor:** Maoteng Li, Huazhong University of Science and Technology, CHINA

**Received:** September 11, 2016

**Accepted:** January 11, 2017

**Published:** February 10, 2017

**Copyright:** © 2017 da Silva et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** This work was supported by Cocoa Research Center (CEPEC/CEPLAC); Brazilian Agricultural Research Corporation (EMBRAPA); Santa Cruz State University, (UESC), Coordination for the Improvement of Higher Education Personnel (CAPES); FINEP/CNPq. The funder provided support in the form of either research

## Abstract

The genus *Theobroma* comprises several trees species native to the Amazon. *Theobroma cacao* L. plays a key economic role mainly in the chocolate industry. Both cultivated and wild forms are described within the genus. Variations in genome size and chromosome number have been used for prediction purposes including the frequency of interspecific hybridization or inference about evolutionary relationships. In this study, the nuclear DNA content, karyotype and genetic diversity using functional microsatellites (EST-SSR) of seven *Theobroma* species were characterized. The nuclear content of DNA for all analyzed *Theobroma* species was  $1C = \sim 0.46$  pg. These species presented  $2n = 20$  with small chromosomes and only one pair of terminal heterochromatic bands positively stained (CMA<sup>+</sup>/DAPI<sup>-</sup> bands). The small size of *Theobroma* ssp. genomes was equivalent to other Byttnerioideae species, suggesting that the basal lineage of Malvaceae have smaller genomes and that there was an expansion of  $2C$  values in the more specialized family clades. A set of 20 EST-SSR primers were characterized for related species of *Theobroma*, in which 12 loci were polymorphic. The polymorphism information content (PIC) ranged from 0.23 to 0.65, indicating a high level of information per locus. Combined results of flow cytometry, cytogenetic data and EST-SSRs markers will contribute to better describe the species and infer about the evolutionary relationships among *Theobroma* species. In addition, the importance of a core collection for conservation purposes is highlighted.

material/ or salaries for authors KPG, UVL, RMA, and LHM, and CAPES provided scholarship to RAS, but they did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The specific roles of these authors are articulated in the 'author contributions' section.

**Competing interests:** The organization EMBRAPA did not play a role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. Thus, this does not alter our adherence to PLOS ONE policies on sharing data and materials.

## Introduction

The 22 species ascribed to the genus *Theobroma* L. (Malvaceae s.l.) are typically Neotropicals and distributed in the Amazon Tropical Forest. The genus *Theobroma* is monophyletic and a sister group of the *Herrania* genus, but the monophyly of *Theobroma* is weakly supported [1–3]. Molecular systematic studies suggested that the subfamily Sterculioideae (which traditionally included the genus *Theobroma*) is not monophyletic and it is divided into two clades: Byttnerioideae (Byttnerieae, Hermannieae, Lasiopetaleae and Theobromeae) and Sterculioideae *sensu strictu* (e.g., Dombeyeae, Sterculieae) [4, 5].

Nine species of *Theobroma* are present in the Brazilian Amazon, among them—cupuassu (*T. grandiflorum* Schum.) and cacao (*T. cacao* L.). The latter considered the most important species in the genus due to its economic value for providing raw material for production of chocolate and derivatives, cosmetics and pharmaceuticals [1, 6]. *T. grandiflorum*, one of the main tree crops in the Amazon region, has its pulp as the principal product, being used in juices, ice creams, yogurts and cosmetics. Their seeds can be used for *cupulate* production, an alternative to chocolate. [7, 8]. Recent studies also highlight the potential of cupuassu fruit extracts for medicinal use in gastrointestinal treatments [9].

A major limitation of *T. cacao* and *T. grandiflorum* production is witches' broom disease, caused by *Moniliophthora perniciosa*, one of the most devastating diseases of cacao and cupuassu trees [10]. Several studies on the *T. cacao* L. vs. *M. perniciosa* interaction have been carried out to identify genes and proteins involved in mechanisms of fungus pathogenicity and/or plant resistance [11, 12]. Studies on wild *Theobroma* species might contribute to important discoveries regarding genes involved in resistance to various diseases, including witches' broom, which may be useful for cacao breeding programs as well as functional guide for genome sequencing strategies, conservation programs or developing *ex situ* breeding designs [13, 14]. Functional microsatellites derived from expressed sequences tags, EST-SSR, can assist such selection. EST-SSRs have been used for traits characterization, breeding and mapping of quantitative trait loci (QTL's) [15–17]. Coming from coding regions, these markers are more conserved among populations of the same species and congeners, thus, enabling cross-amplification and allowing the characterization of molecular marker sets for species which have not been well characterized genetically [18–22]. Several EST-SSR markers have been identified for a diverse range of crops, such as maize [23] and tomato [24], arboreal crops as coffee [25, 26], cacao [27] and, the first work involving EST-SSR from *T. grandiflorum* [28]. Given their high levels of transferability between species [29, 30], the EST-SSR molecular markers are useful tool for studies of genetic diversity, functional genomics, and comparative mapping between species [31].

Cytogenetics based on chromosome variation and staining with chromomycin A3 (CMA) and 4', 6-diamidino-2-phenylindole (DAPI) [32] is a useful approach to study variations in plants. The fluorochrome CMA binds preferentially to GC-rich DNA sequences [33], whereas DAPI preferentially binds to AT-rich sequences [34]. Chromosome double staining with CMA<sup>+</sup>/DAPI<sup>-</sup> has allowed the identification of AT- and GC-rich heterochromatin fractions in many plant groups [35]. Furthermore, genome size is a relevant information for understanding fundamental mechanisms and processes underlying plant growth, evolutionary, systematic taxonomic and cell biology studies, as well as detection of aneuploidy and apoptosis processes. In addition, it provides significant information to sequencing studies and characterization of novel molecular markers, such as EST-SSR [36–41].

Flow cytometry has been used to address questions related to differences in genome copy number. Previous studies quantified the 1C-Value (haploid DNA content) in four cacao genotypes at approximately 0.45 pg [42–44]. A more detailed comparative analysis of mitotic chromosomes of *T. cacao* and *T. grandiflorum* revealed small chromosomes (~2 μm), with only one

pair of terminal heterochromatic bands, co-localized with the single 45S rDNA site and a single 5S rDNA site in the proximal region of the other chromosome pair [45]. These data suggest that the chromosomes of both species are conserved.

Therefore, the objectives of this research were to characterize seven *Theobroma* species using flow cytometry combined with cytogenetic and functional molecular markers from cacao. To our knowledge, this study is pioneer in the analysis of EST-SSRs marker transfer from cacao to the *Theobroma* species.

## Material and methods

### Biological samples

Leaves, seeds, and shoot apices were obtained from cacao accessions at the Cacao Germplasm Bank (CGB) of the Cacao Research Center/Executive Commission of the Cacao Farming Plan—CEPEC/CEPLAC (Ilhéus, Bahia, Brazil). The *T. grandiflorum* genotypes were collected at Embrapa CPATU—Empresa Brasileira de Pesquisa Agropecuária (Belém, Pará, Brazil). Names of the accessions of the individuals used in this study are listed in [Table 1](#).

### Chromosome banding

Root tips obtained from seeds or apical meristems were pre-treated with 0.05% colchicine during 24 h at 10°C and fixed in ethanol:acetic acid (3:1; v/v) for 2–24 h at room temperature and then stored at -20°C. Afterward, the fixed root tips were washed in distilled water and digested in a 2% (w/v) cellulase (Onozuka)/20% (v/v) pectinase (Sigma) solution at 37°C for 90 min. The apical shoots were macerated in a drop of 45% acetic acid and the coverslip removed in liquid nitrogen. For CMA<sup>+</sup>/DAPI<sup>-</sup> bands double staining, the slides were aged for three days, stained with 10 µL of CMA 0.1 mg/mL for 30 min, and restained with 10 µL of DAPI 2g/mL for 60 min [46]. The slides were mounted in glycerol: McIlvaine buffer pH 7.0 (1:1) and aged for three days before analysis in an epifluorescence Leica DMLB microscope. The images were captured with a Cohu CCD video camera using the Leica QFISH software and later edited in Adobe Photoshop CS3 version 10.0.

### Flow cytometry

A suspension of nuclei from young leaves was prepared as described by [47] using Tris-MgCl<sub>2</sub> (WPB buffer). The genome size was estimated using a CyFlow SL flow cytometer (Partec, Görlitz, Germany). For the determination of DNA content, proportionality to fluorescence intensity was assumed and calculated based on at least three different measurements for each individual sample. The histograms were generated in the FloMax (Partec) software using the fluorescence pulse area histogram for analysis. The G1 peak of a diploid *S. lycopersicum* “Stupicke” (1.96 pg DNA content), was used as standard sample, and, set to channel 50 in the 1000 channel histogram. The *S. lycopersicum* seeds were obtained from the Institute of Experimental Botany, Olomouc, Czech Republic [48]. The genome size in base pair for each species and the 1C value of the genome was estimated based on the genome size of *S. lycopersicum*. In order to compare the genome sizes of the different species, a covariance analysis was performed using the GLM procedure of the SAS software (SAS Institute Inc., 2004). The genome size of *S. lycopersicum* was used as covariate to improve the accuracy of the analysis [49].

### DNA extraction

The DNA extraction was performed using healthy leaves in an intermediary stage of maturation with the MATAB protocol [50] slightly modified for *Theobroma* species other than cacao.

**Table 1. Accessions used for characterization, transferability of EST-SSR markers, genome size measurement and cytogenetics analyses.**

Species	Accession	Collection sites	Place of origin	Analysis
<i>T. cacao</i>	P4B	CGB CEPLAC/CEPEC-Ilhéus, Bahia	Peru	EST-SSR
<i>T. cacao</i>	Na33	CGB CEPLAC/CEPEC-Ilhéus, Bahia	Peru	EST-SSR
<i>T. cacao</i>	CCN51	CGB CEPLAC/CEPEC-Ilhéus, Bahia	Ecuador	EST-SSR
<i>T. cacao</i>	Ma16	CGB CEPLAC/CEPEC-Ilhéus, Bahia	Brazil	EST-SSR
<i>T. cacao</i>	MOQ216	CGB CEPLAC/CEPEC-Ilhéus, Bahia	Ecuador	EST-SSR
<i>T. cacao</i>	CSul3	CGB CEPLAC/CEPEC-Ilhéus, Bahia	Brazil	EST-SSR
<i>T. cacao</i>	OC67	CGB CEPLAC/CEPEC-Ilhéus, Bahia	Venezuela	Cytometry/ EST-SSR
<i>T. cacao</i>	EET399	CGB CEPLAC/CEPEC-Ilhéus, Bahia	Ecuador	EST-SSR
<i>T. cacao</i>	TSH1188	CGB CEPLAC/CEPEC-Ilhéus, Bahia	Trinidad/Tobago	EST-SSR
<i>T. cacao</i>	ICS1	CGB CEPLAC/CEPEC-Ilhéus, Bahia	Trinidad/Tobago	Cytometry/ EST-SSR
<i>T. cacao</i>	SPA12	CGB CEPLAC/CEPEC-Ilhéus, Bahia	Colombia	EST-SSR
<i>T. cacao</i>	RB36	CGB CEPLAC/CEPEC-Ilhéus, Bahia	Brazil	EST-SSR
<i>T. cacao</i>	SCA6	CGB CEPLAC/CEPEC-Ilhéus, Bahia	Peru	Cytometry/ EST-SSR
<i>T. cacao</i>	Catongo	CGB CEPLAC/CEPEC-Ilhéus, Bahia	Brazil	EST-SSR
<i>T. cacao</i>	GU171	CGB CEPLAC/CEPEC-Ilhéus, Bahia	French Guiana	EST-SSR
<i>T. cacao</i>	Rosa Maria	CGB CEPLAC/CEPEC-Ilhéus, Bahia	*	EST-SSR
<i>T. cacao</i>	MO20	CGB CEPLAC/CEPEC-Ilhéus, Bahia	Peru	EST-SSR
<i>T. cacao</i>	UF 20	CGB CEPLAC/CEPEC-Ilhéus, Bahia	Costa Rica	Cytometry
<i>T. cacao</i>	UF 667	CGB CEPLAC/CEPEC-Ilhéus, Bahia	Costa Rica	Cytometry
<i>T. cacao</i>	RIM 24	CGB CEPLAC/CEPEC-Ilhéus, Bahia	México	Cytometry
<i>T. cacao</i>	Pentagona	CGB CEPLAC/CEPEC-Ilhéus, Bahia	*	Cytometry
<i>T. cacao</i>	OC 61	CGB CEPLAC/CEPEC-Ilhéus, Bahia	Venezuela	Cytometry
<i>T. cacao</i>	ICS 100	CGB CEPLAC/CEPEC-Ilhéus, Bahia	Trinidad/Tobago	Cytometry
<i>T. cacao</i>	CEPEC 515	CGB CEPLAC/CEPEC-Ilhéus, Bahia	Brazil	Cytometry
<i>T. grandiflorum</i>	C174	EMBRAPA-CPATU-Belém, Pará	Brazil	EST-SSR
<i>T. grandiflorum</i>	C1074	EMBRAPA-CPATU-Belém, Pará	Brazil	EST-SSR
<i>T. grandiflorum</i>	<i>T. grandiflorum</i>	CEPLAC/CEPEC-Ilhéus, Bahia	Brazil	Cytometry
<i>T. bicolor</i>	<i>T. bicolor</i>	CEPLAC/CEPEC-Ilhéus, Bahia	Brazil	Cytometry/ EST-SSR
<i>T. microcarpum</i>	<i>T. microcarpum</i>	CEPLAC/CEPEC-Ilhéus, Bahia	Brazil	Cytometry/ EST-SSR
<i>T. obovatum</i>	<i>T. obovatum</i>	CEPLAC/CEPEC-Ilhéus, Bahia	Brazil	Cytometry/ EST-SSR
<i>T. speciosum</i>	<i>T. speciosum</i>	CEPLAC/CEPEC-Ilhéus, Bahia	Brazil	Cytometry/ EST-SSR
<i>T. subincanum</i>	<i>T. subincanum</i>	CEPLAC/CEPEC-Ilhéus, Bahia	Brazil	Cytometry/ EST-SSR

\*Unknown information

doi:10.1371/journal.pone.0170799.t001

Approximately, 300 mg of leaves were ground using metal beads [51] in the presence of liquid nitrogen. Then, 800 µL of extraction buffer (1.4 M NaCl, 100 mM Tris-HCl pH 8.0, 20 mM EDTA, 10 mM Na<sub>2</sub>SO<sub>3</sub>, 1% PEG 6000, 2% MATAB) preheated at 74 °C were added to the macerate, and incubated for 1 h at 65 °C. After, the sample was cooled at room temperature and 800 µL of chloroform:isoamyl alcohol (24:1, v/v) was added to each sample. The samples were then centrifuged at room temperature for 10 min at 14.000 rpm and the supernatants collected. Afterwards, 700 µL of cold isopropanol was added to the samples and centrifuged for 10 min at 14000 rpm. The pellets were collected and transferred to news 2 mL tubes containing 100 µL of TE (10 mM Tris-HCl, pH 8.0, 1 mM EDTA) in 40 µg/mL RNase. The integrity of the DNA samples was checked on 1% agarose gel stained with 1 ng/µL of gel red. Quantification of the DNA samples was performed using the Picodrop Microliter UV/Vis Spectrophotometer

(Picodrop Limited, UK). The samples had their concentration adjusted to 1 ng DNA/ $\mu$ L and stored at  $-20^{\circ}\text{C}$ .

## Microsatellite loci, amplification by PCR and genotyping

A total of 55 EST-SSR primers previously developed specifically for cacao [15–16, 27] in our laboratory, were synthesized and used to test transferability to cupuassu genotypes. From these, a total of 20 EST-SSR were selected according to the following criteria: a minimum size of 14 bp or originating from genes or proteins involved in mechanisms of plant resistance, or both. Seventeen *T. cacao* accessions representing the diversity of the main genetic *T. cacao* groups were analyzed [17, 52], two *T. grandiflorum* genotypes and five genotypes of wild *Theobroma* (Table 1). These accessions belong to CEPLAC genetic breeding program [17, 21]. After validation on cacao, the markers were tested for transferability on *T. grandiflorum* genotypes 174 (Coari) and 1074, resistant and susceptible to witches' broom disease, respectively. These genotypes are genitors of several progenies in cupuassu breeding program [28] (Table 1).

PCR reactions were carried out with 30 ng of DNA, 0.2 mmol/L of each primer, 2.0 mmol/L of  $\text{MgCl}_2$ , 0.2 mmol/L of each dNTP, 1X buffer and one unit of DNA Taq polymerase (Ludwig Biotecnologia Ltda) in a final volume of 20  $\mu$ L. PCRs were carried out using touchdown protocol with 10 cycles as follow: denaturation at  $94^{\circ}\text{C}$  for 4 min, annealing temperature at  $60\text{--}48^{\circ}\text{C}/56^{\circ}\text{C}$  with  $1^{\circ}\text{C}$  decrease at each cycle, and extension at  $72^{\circ}\text{C}$  for 1 min and 30 s, followed by 30 cycles at  $94^{\circ}\text{C}$  for 30 s,  $48^{\circ}\text{C}$  for one min and 30 s, and final extension of 4 min at  $72^{\circ}\text{C}$ . Polymorphism evaluation was carried out by electrophoresis on 6% denaturing TBE acrylamide gel using as running buffer TBE 1X (89 mmol/L Tris, mmol/L boric acid and 2 mmol/L of EDTA). Microsatellite polymorphism was visualized using the silver staining method [53, 54].

The amplified SSR DNA bands representing different alleles were scored on different genotypes. A principal component analysis, conducted on the allele frequency data, number of allele per locus ( $N_a$ ) and average observed heterozygosity ( $H_o$ ) were determined with the GENETIX software v. 4.05.2 [55]. The polymorphic information content (PIC) was obtained for each locus with the CERVUS software version 3.0 [56, 57], and the genetic distance (D) was calculated according to Nei Genetic Distances [58]. The genetic analyses were carried out considering three groups: *T. cacao*, *T. grandiflorum* and *Theobroma* wild species.

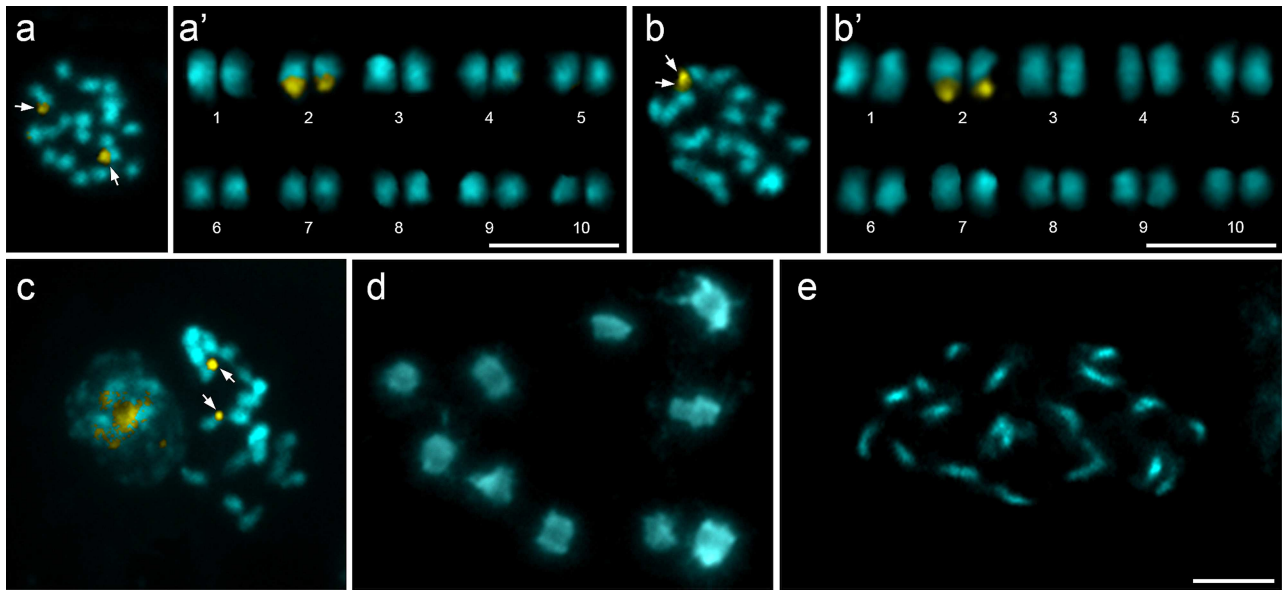
## Results

### Chromosome number, chromosome banding and DNA content

All species showed symmetric karyotypes  $2n = 20$  with small metacentric/submetacentric chromosomes. In all species, a  $\text{CMA}^+/\text{DAPI}^-$  band was present on the terminal region of the long arm of a single chromosome pair. This  $\text{CMA}^+$  band was frequently heteromorphic in size and distended in one or both homologues (Fig 1).

The 1C nuclear DNA content and genome size of the plant species are presented in Table 2, and the histogram of the fluorescence peak are presented in Fig 2.

Nuclear DNA content (1C values) in the studied *Theobroma* species ranged from 0.41 pg in *T. microcarpum* to 0.49 pg in *T. speciosum* and *T. subincanum*, with an average of 0.46 pg (Table 2). Three species (*T. bicolor*, *T. obovatum* and *T. grandiflorum*) showed  $1C = 0.48$ . The ten *T. cacao* genotypes analyzed showed 1C values ranging from 0.44 to 0.47 pg, with an average of 0.45 pg. The coefficient of variation (CV) was low (4.07%), and the estimated genome sizes of the seven *Theobroma* species, after adjustment by the genome size of *S. lycopersicum* were not statistically different by the F test ( $p\text{-value} = 0.975$ ) (S1 Table).



**Fig 1. Cytogenetic analysis of the genus *Theobroma*.** (a-a') metaphase and caryogram of the *T. cacao* ssp. leicocarpum stained with CMA<sup>+</sup> (yellow) and DAPI<sup>-</sup> (blue). (b-b') Metaphase and caryogram of *T. cacao* cv. Scavina 6. (c) metaphase of *T. bicolor* stained with CMA<sup>+</sup>/DAPI<sup>-</sup>. (d) diakinesis of *T. bicolor* showing the 10 pair softbivalent. (e) prometaphase of *T. grandiflorum*, bars = 5 μm. Arrows in a, b, c points to the CMA<sup>+</sup> bands.

doi:10.1371/journal.pone.0170799.g001

### Cross amplification and genetic analyses

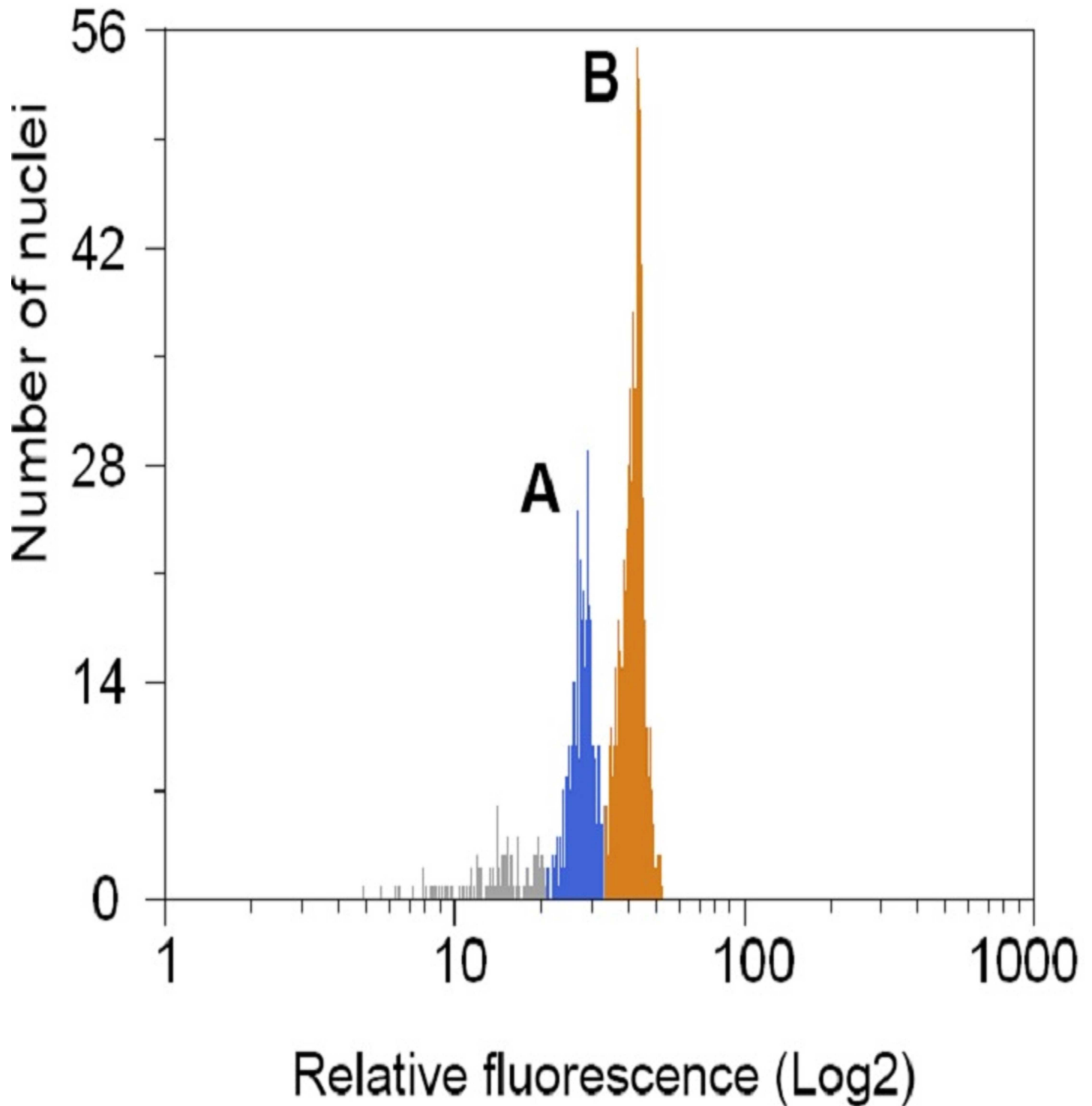
The 20 EST-SSR markers produced informative results about allelic variation among the cultivated and wild species. Twelve of them (60%) were polymorphic and 8 monomorphic (40%) in *T. cacao* individuals. The 20 primer pairs tested showed satisfactory cross-amplification results in 75% of *T. grandiflorum* and *T. subincanum*, 90% in *T. obovatum*, 60% in *T. bicolor* and 35% in *T. speciosum* and *T. microcarpum*. The annealing temperature of the primer pairs ranged from 48–60°C with amplicons varying from 100 to 322 pb (S1 Table).

Considering all 24 sampled individuals and the 20 loci, 16 were polymorphic and 4 were monomorphic. The analysis using the 16 polymorphic EST-SSRs revealed a total of 24 alleles, with an amplitude of size ranging from 155–165 (mEstTcCepec13-3) to 250–260 base pairs (msEstTsh-10) (S2 Table). The number of alleles per locus ranged from two (mEstTcCepec60, mEstTcCepec16-8, mEstTcCepec24, mEstTcCepec13, mEstTcCepec31, mEstTcCepec20-5) to seven (mEstTcCepec16-4), and the polymorphic information content (PIC) of each locus, excluding the monomorphic ones, ranged from 0.08 (mEstTcCepec31) to 0.75 (mEstTcCepec13-4).

**Table 2. Genome size of *Theobroma* species.**

Species	Genome size (1C value)	Number of base pairs
<i>T. cacao</i>	0.4666	456 Mpb
<i>T. grandiflorum</i>	0.4589	448 Mpb
<i>T. microcarpum</i>	0.4624	452 Mpb
<i>T. obovatum</i>	0.4624	452 Mpb
<i>T. speciosum</i>	0.4624	452 Mpb
<i>T. bicolor</i>	0.4624	452 Mpb
<i>T. subincanum</i>	0.4624	452 Mpb
<b>Average</b>	<b>0.46</b>	

doi:10.1371/journal.pone.0170799.t002



**Fig 2. Histogram of relative nuclear DNA content (genome size).** (A) Nuclear DNA content in *T. cacao* cv. ICS100. (B) *S. lycopersicum* cells, included as an internal standard. Nuclei isolated from cocoa leaves were stained with propidium iodide and analyzed by flow cytometry.

doi:10.1371/journal.pone.0170799.g002

The loci were classified either as moderately polymorphic with  $0.25 > PIC > 0.5$  or highly polymorphic with  $PIC > 0.5$  (S2 Table) [53]. The average number of alleles per locus varied among

**Table 3. Genetic diversity estimates for individuals of the *Theobroma* genus.**

Groups	Na	He	H <sub>O</sub>
<i>T. cacao</i>	2.4	0.31	0.12
<i>T. grandiflorum</i>	1.1	0.04	0.00
<i>T. ssp</i>	1.8	0.28	0.06
<b>Average</b>	<b>1.8</b>	<b>0.21</b>	<b>0.06</b>

Na = number of alleles averaged over all loci; H<sub>O</sub> = unbiased observed heterozygosity (Nei, 1978); He = unbiased expected heterozygosity (Nei, 1978).

doi:10.1371/journal.pone.0170799.t003

the *Theobroma* groups, ranging from 2.4 in *T. cacao* to 1.8 in *Theobroma* ssp and 1.1 in *T. grandiflorum* (Table 3).

The genetic distance also varied among the *Theobroma* groups. *Theobroma cacao* group was more genetically distant from the *T. ssp* group (0.444) and less distant from the *T. grandiflorum* group (0.219). *Theobroma grandiflorum* and *T. ssp* groups had a moderate genetic distance of 0.378 (Table 4).

## Discussion

### Genome size and cytogenetics

In general, the genome size among related species is, for the most part, stable and rarely exceeds values above or below 2 to 5% when using techniques commonly suggested for determining nuclear DNA content [59]. This trend was also observed in this study, wherein the genome sizes estimated in this study for species of the *Theobroma* genus (*T. cacao*, *T. grandiflorum*, *T. microcarpum*, *T. speciosum*, *T. bicolor*, *T. subincanum* and *T. obovatum*) did not differ statistically. For most of these species, there are no previous studies about the genome size, except for *T. cacao*. The small differences found in the genome sizes in samples of *T. cacao* have been reported in previous studies [42, 43, 60]. Nevertheless, these differences in measurements of DNA content observed between laboratories cannot be interpreted as interspecific variation, because the measuring tools may present small differences in the alignment along the time. Regarding the variations among species, small differences in measurements can be generated by phenolic compounds, which have antioxidant activity, thus conferring protection against DNA damage to the cells. These compounds are known to inhibit proper DNA staining. In experiments that aim measuring the size of the genome, any failure in the application of the protocol may cause variations in the data obtained, in comparison to homogeneous values [61–64]. However, in this study it was possible to acquire a high resolution of the histograms. Besides replications of samples, other actions were implemented, like use of *S. lycopersicum* genome as control to increase the accuracy of the measurements [65–68]. Besides that, the coefficient of variation was low (4.07%), suggesting a high precision in the comparisons. Therefore, in the present case, the homogeneity of the genome sizes of *Theobroma* species suggests a true uniformity in those sizes.

**Table 4. Genetic distance matrix between the three groups of individuals of the genus *Theobroma*.**

Groups	<i>T. cacao</i>	<i>T. grandiflorum</i>	<i>T. ssp</i>
<i>T. cacao</i>	0.000	–	–
<i>T. grandiflorum</i>	0.219	0.000	–
<i>T. ssp</i>	0.444	0.378	0.000

doi:10.1371/journal.pone.0170799.t004



Variations in genome size of angiosperms are wide, varying from  $1C = 0.06$  pg in *Genlisea margaretae* to  $1C = 152.23$  pg in *Paris japonica*, with an extensive variation occurring within the genus [68, 69]. Some studies have revealed that chromosome length ranged from 2.00 to 1.19  $\mu\text{m}$  for *T. cacao* and from 2.21 to 1.15  $\mu\text{m}$  for *T. grandiflorum*, with most chromosome pairs similar in morphology and size, corroborating with our findings [45]. Plant species native from tropical forests tend to have a reduced genome size, which could be influenced by differences in temperature affecting cellular division and expansion [70]. The process of cellular division and growth are favored by elevated temperatures as seen in cacao and other species of the genus [71]. This process can select a more short mitotic cycle and consequently small cells and small genomes compared to plants of temperate regions [42].

Studies show that probably the ancestral genome size in Angiosperms was reduced, and that along the evolutionary scale DNA content suffered increases due to polyploidization events and self-replicating DNA elements. The reduced genome sizes observed in the studied species may have also a phylogenetic signal. The clade Byttnerioideae (which includes *Theobroma* and *Herrania*) presents small genomes ( $\sim 0.43$  pg), as well as it is related to the Grewioideae ( $\sim 0.48$  pg) clade [72–74]. Byttnerioideae and Grewioideae form the most basal group of Malvaceae, suggesting that the small genomes are a plesiomorphic condition in the family. In contrast, the most specialized clades within Malvaceae, Sterculioideae ( $\sim 1.944$  pg), Bombacoideae (1.97 pg) and Malvoideae (1.72 pg) have much larger genomes, suggesting that there was a trend of expansion of the genome size in these lineages [3]. In some cases, this increase in the genome can be influenced by polyploidy events, such as for Bombacoideae [1, 75].

The small genomes in these species studied here may be correlated with the karyotype stability in the genus. Our data support the karyotypes already described for other species in the genus, that it is a taxon predominantly diploid ( $2n = 20$ ), with small and symmetric chromosomes [44, 76]. The only differences reported in the literature are detectable cytological heterochromatic bands observed on the centromeric/ pericentromeric regions of all 20 chromosomes of cacao after C-banding stained with either Giemsa or DAPI, whereas never being detected in *T. grandiflorum*. The presence of only a couple of bands  $\text{CMA}^+/\text{DAPI}^-$  in all species, and in different varieties of cacao also confirms the karyotype correlations among them and are in agreement with previous analyses [45].

## Characterization and transferability of EST-SSR markers

Adoption of genomic approaches to crop improvement studies and preservation of wild tree species is severely limited by the lack of sufficient molecular markers. EST-SSRs are codominant, highly reproducible and polymorphic markers. With these characteristics, they have been used favorably for population genetic analyses and genetic mapping in several species.

Rate of polymorphism in genomic SSRs is generally high in comparison to SSR from ESTs [77, 78]. However, EST-SSR shows some advantages, such as higher frequency in the genome and their link to interesting agronomical traits [79]. From the 20 EST-SSR tested here, 60% generated polymorphic loci. In previous studies lower rates were found, when 32 EST-SSRs associated with resistance to *M. perniciosa* were tested only 26.7% were polymorphic, nevertheless the individuals used in the present study were more heterogeneous [15].

Regarding the genetic diversity found in the sampled individuals, we noted that the highest polymorphic information content (PIC) was in the *T. cacao* group, showing the potential coverage of these markers. It is worth mentioning that it was not the aim of this study to characterize the genetic structure of populations in the *Theobroma* genus, but to test the cross-amplification to wild species of *Theobroma* genus and to *T. grandiflorum*. One of the advantages of EST-SSR markers is the fact they are easily transferable between species of the same genus, since they are

derived from commonly conserved transcribed regions of DNA, directly decreasing in the costs of molecular studies in wild species [80–82]. In the present work, the transferability rate of these EST-SSRs markers among the *Theobroma* species ranged from 35% to 90% among samples. Therefore, based on the premise that EST-SSRs are highly conserved in congeners, the percent of functional primers to other species of the genus decreased with the increase in genetic distance, that is, *T. microcarpum* from the section Telmatocarpus and *T. speciosum* from the section Oreanthes showed the lowest rates of primer transferability (35%). In a work with wheat, it was reported that this is an expected result due to the insertion of introns in correlated species [83] that modify the target sequence, however, with rates of amplification normally higher than 50% [31].

Microsatellite markers are especially useful to characterize the genetic diversity and kinship between species, due the high polymorphism and number of alleles per locus [84]. *Theobroma grandiflorum* and *T. obovatum* species belong to section Glossopetalum, considered the most ancestral section of the genus, while *T. cacao* (section *Theobroma*) suggests a distant evolutionary relationship between these species. Nevertheless, there have been reports of hybrids between *T. cacao* and *T. grandiflorum* and *T. grandiflorum* with *T. obovatum* meaning closer relationship between those species [3, 85].

The first sets of molecular markers reported for wild species of *Theobroma* were RAPD (Random Amplified Polymorphic DNA) used to elucidate phylogenetic relationships between species. Patterns of RAPD showed intra and interspecific polymorphism and some bands of similar lengths between species classified in the same section or correspondents [86].

Previous studies found a transferability rate of SRR markers from *T. cacao* to *T. grandiflorum* of 60.4%, showing similarity between species and highlighting the possibility of using those markers in association mapping and breeding [87]. In the present study, the transferability rate of EST-SSRs from *T. cacao* to *T. grandiflorum* was slightly higher, around 70%, as expected, considering that EST-SSRs come from conserved sequences. Regarding *T. obovatum*, the rate of cross-amplification was 90%. Until now, there is no previous report of SSR or EST-SSR markers to this species. These cross-amplification rates are in agreement with previous studies that have shown that the proportion of cross-amplification studies involving polymorphic SSR loci within the same plant genus ranged from 20% to 100% [88].

The constancy of the genome size of these *Theobroma* species and the high rate of EST-SSR cross-amplification supports our conclusion that the areas bordering the microsatellites in the studied species are conserved enough to allow cross-amplification. Additionally, suggesting that these species have been conserved along the evolution regarding their genomic sequences and number and size of chromosomes, which are relatively small compared to most angiosperms. This information can also be confirmed by the amount of non-coding repetitive DNA, composed of transposable elements, satellite DNA, introns and pseudo genes, as seen for the *T. cacao* genome [89–91]. Preservation of the sequence depends mainly on the evolutionary relationship between the species of origin. Thus, the more diverse the taxon, the less successful the cross-amplification will be. Data from this research allows inferences that these *Theobroma* species have a certain genomic homology.

The EST-SSR markers evaluated in this study showed a considerable transferability rate into six related *Theobroma* species, therefore, these markers are important to closely monitor the genetic variability. In fact, this is the first report of EST-SSR molecular markers used in wild species of the *Theobroma* genus, thus, representing a new set of primers highly informative that can be used in studies of transferability, paternity, and genetic flow of Amazon species. They can also help in studies aiming better strategies of biodiversity conservation and studies of disease resistance genes in *T. grandiflorum* [92].

## Supporting information

**S1 Table. Analysis of covariance for the genome size measured in seven species of *Theobroma*.**

(DOCX)

**S2 Table. Characteristics of 20 EST-SSRs markers used on the present work.** Shown for each primer pair are the previously published forward and reverse sequence, repeat type, allelic amplitude (bp), number of alleles (Na) and polymorphic information content (PIC) for species of *Theobroma*.

(DOC)

## Acknowledgments

R.A. da Silva acknowledges CAPES/EMBRAPA (Coordination for the Improvement of Higher Education Personnel, Brasilia, Brazil—Projects Geneaçu—genetic basis to assist the cupuassu tree breeding), for supporting her with a research assistantship during her MS programme. The authors also thank the molecular plant pathology laboratory personnel Dr. Rogério Ferreira Mercês and Rodrigo Ganem for their assistance in accomplishing this study, and Dr. Raúl Valle (CEPLAC, Brazil) and Dr. Claudia Fortes Ferreira (EMBRAPA, Brazil) for English language revision. K.P.G. and F.M. were supported by research fellowship Pq-1 from CNPq,

## Author Contributions

**Conceptualization:** KPG RAS GS LSSL.

**Data curation:** RAS KPG.

**Formal analysis:** RAS GS LSSL UVL NGRBP.

**Funding acquisition:** GS KPG FM.

**Investigation:** RAS GS LSSL.

**Methodology:** RAS GS LSSL.

**Project administration:** KPG FM.

**Resources:** RMA DC FM GS.

**Software:** RAS KPG GS.

**Supervision:** KPG FM GS.

**Validation:** RAS GS LSSL.

**Visualization:** KPG RAS LSSL.

**Writing – original draft:** RAS KPG GS UVL FM LHM.

**Writing – review & editing:** KPG LSSL GS.

## References

1. Cuatrecasas JA. Taxonomic revision of the genus *Theobroma*. *Contr. U. S. Natl Herb.* 1964; 35(6): 379–607.
2. Whitlock BA, Baum DA. Phylogenetic relationships of *Theobroma* and *Herrania* (Sterculiaceae) based on sequences of the nuclear gene Vicilin. *Systematic Botanic.* 1999; 24: 128–138.

3. Silva CRS, Figueira AVO. Phylogenetic analysis of *Theobroma* (Sterculiaceae) based on Kunitz-like trypsin inhibitor sequences. *Plant Systematics and Evolution*. 2005; 250: 93–104.
4. Alverson WS, Whitlock BA, Nyffeler R, Bayer C, Baum DA. Phylogenetic analysis of the core Malvales based on sequences of *ndhF*. *American Journal of Botany*. 1999; 86: 1474–1486. PMID: [10523287](#)
5. Bayer CMF, Fay AY, De Bruijn V, Savolainen CM, Morton K, Kubitzki, et al. Support for an expanded family concept of Malvaceae within circumscribed order Malvales: a combined analysis of plastid *atpB* and *rbcl* DNA sequences. *Botanical Journal of the Linnean Society*. 1999; 129: 267–303.
6. Roiaini M, Seyed HM, Jinap S, Norhayati H. Effect of extraction methods on yield, oxidative value, phytochemicals and antioxidant content of cocoa butter. *International Food Research Journal*. 2016; 23(1): 47–54
7. Lannes SCS, Medeiros ML, Amaral RL. Formulação de "chocolate" de cupuaçu e reologia do produto líquido. *Revista Brasileira de Ciências Farmacêuticas*. 2002; 38(4):463–469.
8. Pinet M, Castell-Auví A, Genovese MI, Serrano J, Casanova A, Blay M, et al. Antioxidant effects of proanthocyanidin-rich natural extracts from grape seed and cupuassu on gastrointestinal mucosa. *Science and Food and Agriculture*. 2016; 96(1): 178–182.
9. Costa MP, Frasao BS, Rodrigues BL, Silva ACO, Conte-Junior CA. Effect of different fat replacers on the physicochemical and instrumental analysis of low-fat cupuassu goat milk yogurts. *Journal of Dairy Research*. 2016; 83(4): 493–496. doi: [10.1017/S0022029916000674](#) PMID: [27845025](#)
10. Aime MC, Phillips-Mora W. The causal agents of witches broom and frosty pod rot of cacao (chocolate, *Theobroma cacao*) form a new lineage of Marasmiaceae. *Mycologia*. 2005; 97(5): 1012–1022. PMID: [16596953](#)
11. Gesteira AS, Micheli F, Carels N, Da Silva AC, Gramacho KP, Schuster I, et al. Comparative analysis of expressed genes from cacao meristems infected by *Moniliophthora perniciosa*. *Annals of Botany*. 2007; 100(1): 129–140. doi: [10.1093/aob/mcm092](#) PMID: [17557832](#)
12. Teixeira PJ, Thomazella DP, Reis O, Do Prado PF, Do Rio MC, Fiorin GL, et al. High-resolution transcript profiling of the atypical biotrophic interaction between *Theobroma cacao* and the fungal pathogen *Moniliophthora perniciosa*. *Plant Cell*. 2014; 26(11): 4245–4269. doi: [10.1105/tpc.114.130807](#) PMID: [25371547](#)
13. Kerr WE, Clement CR. Práticas agrícolas de consequências genéticas que possibilitaram aos índios da Amazônia uma melhor adaptação às condições ecológicas da região. *Acta Amazônica*. 1980; 10(2): 251–261.
14. Faleiro FG, Junqueira NTV, Braga MF. Maracujá: germoplasma e melhoramento genético. Planaltina, DF: Embrapa Cerrados. 2005; 670p.
15. Lemos, LSL. Desenvolvimento de Marcadores EST-SSR e SNPs para mapeamento genético de novas fontes de resistência do cacaueiro à vassoura-de-bruxa. D.Sc. Thesis, Universidade Estadual de Santa Cruz, Ilhéus. 2010.
16. Lima LS, Gramacho KP, Gesteira AS, Lopes UV, Gaiotto F, Zaidan HA, et al. Characterization of microsatellites from cacao-*Moniliophthora perniciosa* interaction expressed sequence tags. *Molecular Breeding*. 2008; 22:315–318.
17. Silva SDVM Luz EDMN, Pires JL Yamada MM, Santos Filho LP. Parent selection for cocoa resistance to witches'-broom. *Pesquisa Agropecuária Brasileira*. 2010; 45(7): 680–685.
18. Hagraas AA, Kishii M, Sato K, Tanaka H, Tsujimoto H. Extended application of barley EST markers for the analysis of alien chromosomes added to wheat genetic background. *Breeding Science*. 2005; 55: 335–341.
19. Ayala-Navarrete L, Bariana HS, Singh RP, Gibson JM, Mechanicos AA, Larkin PJ. Trigenomic chromosomes by recombination of *Thinopyrum intermedium* and *Th. ponticum* translocations in wheat. *Theoretical and Applied Genetics*. 2007; 116(1): 63–75. doi: [10.1007/s00122-007-0647-5](#) PMID: [17906848](#)
20. Haq SU, Kumar P, Singh RK, Verma KS, Bhatt R, Sharma M, et al. Assessment of Functional EST-SSR Markers (Sugarcane) in Cross-Species Transferability, Genetic Diversity among Poaceae Plants, and Bulk Segregation Analysis. *Genetic Research International*. 2016; 2016: 1–16.
21. Rowland LJ, Dhanaraj AL, Polashock JJ, Arora R. Utility of blueberry derived EST-PCR primers in related Ericaceae species. *HortScience*. 2003; 38(7): 1428–1432.
22. Sargent DJ, Rys A, Nier S, Simpson DW, Tobutt KR. The development and mapping of functional markers in *Fragaria* and their transferability and potential for mapping in other genera. *Theoretical and Applied Genetics*. 2007; 114(2): 373–384. doi: [10.1007/s00122-006-0441-9](#) PMID: [17091261](#)
23. Galvão KS, Ramos HC, Santos PH, Entringer GC, Vettorazzi JC, Pereira MG. Functional molecular markers (EST-SSR) in the full-sib reciprocal recurrent selection program of maize (*Zea mays* L.). *Genetic Molecular Research*. 2015; 14(3):7344–55.

24. Zhou R, Wu Z, Jiang FL, Liang M. Comparison of gSSR and EST-SSR markers for analyzing genetic variability among tomato cultivars (*Solanum lycopersicum* L.). *Genetic Molecular Research*. 2015; 14(4):13184–13194. October.26.14.
25. Poncet V, Rondeau M, Tranchant C, Cayrel A, Hamon S, De Kochko A, et al. SSR mining in coffee tree EST databases: potential use of EST-SSRs as markers for the *Coffea* genus. *Molecular and Genetics Genomics*. 2006; 276(5): 436–449.
26. Ogutu C, Fang T, Yan L, Wang L, Huang L, Wang X, et al. Characterization and utilization of microsatellites in the *Coffea canephora* genome to assess genetic association between wild species in Kenya and cultivated coffee. *Tree Genetics & Genomes*. 2016; 12; 54.
27. Lemos, LSL. Identificação de polimorfismo em ESTs da interação cacau-*Moniliophthora perniciosa*. M. Sc. Thesis, Universidade Estadual de Santa Cruz, Ilhéus. 2007.
28. Santos LF, Fregapani RM, Falc LL, Togawa RC, Costa MMC, Lopes UV, et al. First microsatellite markers developed from cupuassu ESTs: duplication in diversity analysis and cross-species transferability to cacao. *PLoS ONE*. 2016; 11(3):1–19.
29. Gupta S, Prasad M. Development and characterization of genic SSR markers in *Medicago truncatula* and their transferability in leguminous and non-leguminous species. *Genome*. 2009; 52(9): 761–771. doi: [10.1139/g09-051](https://doi.org/10.1139/g09-051) PMID: [19935924](https://pubmed.ncbi.nlm.nih.gov/19935924/)
30. Luro FL, Costantino G, Terol J, Argout X, Allario T, Wincker P, et al. Transferability of the EST-SSRs developed on *Nules clementine* (*Citrus clementina* Hort ex Tan) to other *Citrus* species and their effectiveness for genetic mapping. *BMC Genomics*. 2008; 9: 287–299. doi: [10.1186/1471-2164-9-287](https://doi.org/10.1186/1471-2164-9-287) PMID: [18558001](https://pubmed.ncbi.nlm.nih.gov/18558001/)
31. Varshney RK, Graner A, Sorrells ME. Genic microsatellite markers in plants: features and applications. *Trends Biotechnology* 2005; 23(1):48–55.
32. Felix LP, Guerra M. Variation in chromosome number and the basic number of subfamily Epidendroideae (Orchidaceae). *Botanical Journal of the Linnean Society*. 2010; 163(2): 234–278.
33. Hou M, Robinson H, Gao Y, Wang AH. Crystal structure of the  $[Mg^{2+}(\text{chromomycin } A_3)_2]\text{-d}(\text{TTGGCCAA})_2$  complex reveals GGCCn binding specificity of the drug dimmer chelated by a metal ion. *Nucleic Acids Research*. 2004; 32(7): 2214–2222. doi: [10.1093/nar/gkh549](https://doi.org/10.1093/nar/gkh549) PMID: [15107489](https://pubmed.ncbi.nlm.nih.gov/15107489/)
34. Kapuscinski J. DAPI: A DNA-specific fluorescent probe. *Biotechnic Histochemistry*. 1995; 70(5): 220–233. PMID: [8580206](https://pubmed.ncbi.nlm.nih.gov/8580206/)
35. Guerra M. Patterns of heterochromatin distribution in plant chromosomes. *Genetic Molecular Biology*. 2000; 23(4):1029–1041.
36. Loureiro J, Travníček P, Rauchová J, Urfus T, Vít P, Štech M, et al. The use of flow cytometry in the bio-systematics, ecology and population biology of homoploid plants. *Preslia*. 2010; 82: 3–21.
37. Warren A, Crampton JM. The *Aedes aegypti* genome: complexity and organization. *Genetical Research*. 1991; 58(3): 225–232. PMID: [1802804](https://pubmed.ncbi.nlm.nih.gov/1802804/)
38. Kawara S, Takata M, Takeara K. High frequency of DNA aneuploidy detected by DNA flow cytometry in Brown's disease. *J. Dermatological Science*. 1999; 21(1): 23–26.
39. Vermes I, Haanen C, Reutelingsperger C. Flow cytometry of apoptotic cell death. *Journal of Immunological Methods*. 2000; 243(1): 167–190.
40. Ochatt SJ. Flow cytometry in plant breeding. *Cytometry*. 2008; 73(7): 581–598. doi: [10.1002/cyto.a.20562](https://doi.org/10.1002/cyto.a.20562) PMID: [18431774](https://pubmed.ncbi.nlm.nih.gov/18431774/)
41. Gregory TR. The C-value enigma in plants and animals: a review of parallels and an appeal for partnership. *Annals of Botany*. 2005; 95(1): 133–146. doi: [10.1093/aob/mci009](https://doi.org/10.1093/aob/mci009) PMID: [15596463](https://pubmed.ncbi.nlm.nih.gov/15596463/)
42. Figueira AVO, Janick J, Goldsbrough P. Genome size and DNA polymorphism in *Theobroma cacao*. *Journal of the American Society for Horticultural Science*. 1992; 117(4): 673–677.
43. Argout X, Salse J, Aury J, Guiltinan M, Droc G, Gouzy J, et al. The genome of *Theobroma cacao*. *Nature Genetics*. 2011; 43: 101–108. doi: [10.1038/ng.736](https://doi.org/10.1038/ng.736) PMID: [21186351](https://pubmed.ncbi.nlm.nih.gov/21186351/)
44. Muñoz, JMO. Estudios cromosómicos en el género *Theobroma*. IICA, Turrialba. 1948.
45. Dantas LG, Guerra M. Chromatin differentiation between *Theobroma cacao* L. and *T. grandiflorum* Schum. *Genetic and Molecular Biology*. 2010; 33: 94–98.
46. Barros e Silva AE, Guerra M. The meaning of DAPI bands observed after C-banding and FISH procedures. *Biotechnic Histochemistry*. 2010; 85(2): 115–125. doi: [10.1080/10520290903149596](https://doi.org/10.1080/10520290903149596) PMID: [19657781](https://pubmed.ncbi.nlm.nih.gov/19657781/)
47. Loureiro J, Rodriguez E, Dolezel J, Santos C. Two new nuclear isolation buffers for plant DNA flow cytometry: a test with 37 species. *Annals of Botany*. 2007; 100(4): 875–888. doi: [10.1093/aob/mcm152](https://doi.org/10.1093/aob/mcm152) PMID: [17684025](https://pubmed.ncbi.nlm.nih.gov/17684025/)

48. Dolezel J, Sgorbati S, Lucretti S. Comparison of three DNA fluorochromes for flow cytometric estimation of nuclear DNA content in plants. *Physiologia Plantarum*. 1992; 85(4): 625–631.
49. SAS Institute Inc. 2004. *SAS/STAT® 9.1 User's Guide*. Cary, NC: SAS Institute Inc.
50. Risterucci AM, Grivet L, Goran JAK, Pieretti I, Flament MH, Lanaud C. A high-density linkage map of *Theobroma cacao* L. *Theoretical and Applied Genetics*. 2000; 101(5): 948–955.
51. Santos RMF, Lopes UV, Clement D, Pires JL, Lima EM, Messias TB, et al. A protocol for large scale genomic DNA isolation for cacao genetics analysis. *African Journal of Biotechnology*. 2014; 13(7): 814–820.
52. Marita JM, Nienhuis J, Pires JL, Aitken WM. Analysis of genetic diversity in *Theobroma cacao* with emphasis on witches' broom disease resistance. *Crop Science*. 2001; 41: 1305–1316.
53. Creste S, Tulmann-Neto A, Figueira A. Detection of simple sequence repeat polymorphisms in denaturing polyacrylamide sequencing gels by silver staining. *Plant Molecular Biological Report*. 2001; 19(4): 299–306.
54. Gramacho KP, Moreira RFC, Lemos LSL, Lima ES, Clement D. Obtenção de marcadores microssatélites para genotipagem e análise genética de *Moniliophthora perniciosa* em gel corado com prata. *Ilhéus: CEPLAC- Boletim Técnico*. 2009; 196.
55. Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F. GENETIX 4.05, logiciels sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier (France). (1996–2004); <http://kimura.univ-montp2.fr/genetix/>.
56. Botstein D, White RL, Skolnick M, Davis RW. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *The American Journal Human Genetics*. 1980; 32(3): 314–331.
57. Kalinowski ST, Taper M, Marshall TC. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular Ecology*. 2007; 16(5): 1099–1106. doi: [10.1111/j.1365-294X.2007.03089.x](https://doi.org/10.1111/j.1365-294X.2007.03089.x) PMID: [17305863](https://pubmed.ncbi.nlm.nih.gov/17305863/)
58. Nei M. Genetic distance between populations. *The American Naturalist*. 1972; 106(949): 283–292.
59. Bennett MD, Bhandol P, Leitch IJ. Nuclear DNA amounts in angiosperms and their modern uses—807 new estimates. *Annals of Botany*. 2000; 86(4): 859–909.
60. Dolezel J, Greilhuber J, Lucretti S, Meister A, Lysak MA, Nardi L, et al. Plant genome size estimation by flow cytometry: Inter-laboratory comparison. *Annals of Botany*. 1998; 82(1): 17–26.
61. Wollgast J, Anklam E. Review on polyphenols in *Theobroma cacao*: changes in composition during the manufacture of chocolate and methodology for identification and quantification. *Food Research International*. 2000; 33(6): 423–447.
62. Noirot M, Barre P, Duperray C, Louarn J, Hamon S. Effects of caffeine and chlorogenic acid on propidium iodide accessibility to DNA: Consequences on genome size evaluation in coffee tree. *Annals of Botany*. 2003; 92(2): 259–264. doi: [10.1093/aob/mcg139](https://doi.org/10.1093/aob/mcg139) PMID: [12876189](https://pubmed.ncbi.nlm.nih.gov/12876189/)
63. Efraim P, Tucci ML, Pezoa-Garcia NH, Haddad R, Eberlin MN. Teores de compostos fenólicos de sementes de cacauero de diferentes genótipos. *Brazilian Journal Food Technology*. 2006; 9:229–236.
64. Greilhuber J. Severely distorted Feulgen DNA amounts in *Pinus* (Coniferophytina) after non additive fixations as a result of meristematic self-tanning with vacuole contents. *Canadian Journal of Genetics and Cytology*. 1986; 28:409–415.
65. Greilhuber J. Critical reassessment of DNA content variation in plants. In: Brandham PE, editor. *Kew Chromosome Conference III.b*. London: HMSO. 1988: 39–50.
66. Noirot M, Barre P, Louarn J, Duperray C, Hamon S. Nucleus-cytosol interactions: a source of stoichiometric error in flow cytometric estimation of nuclear DNA content in plants. *Annals of Botany*. 2000; 86(2): 309–316.
67. Noirot M, Barre P, Duperray C, Hamon S, De Kochko A. Investigation on the causes of stoichiometric error in genome size estimation using heat experiments: consequences on data interpretation. *Annals of Botany*. 2005; 95(1): 111–118. doi: [10.1093/aob/mci006](https://doi.org/10.1093/aob/mci006) PMID: [15596460](https://pubmed.ncbi.nlm.nih.gov/15596460/)
68. Greilhuber J, Borsch T, Müller K, Worberg A, Porembski S, Barthlott W. Smallest angiosperm genomes found in Lentibulariaceae, with chromosomes of bacterial size. *Plant Biol*. 2006; 8(6): 770–777. doi: [10.1055/s-2006-924101](https://doi.org/10.1055/s-2006-924101) PMID: [17203433](https://pubmed.ncbi.nlm.nih.gov/17203433/)
69. Pellicer J, Garcia S, Canela MÁ, Garnatje T, Korobkov AA, Twibell JD, et al. Genome size dynamics in *Artemisia* L. (Asteraceae): following the track of polyploidy. *Plant Biology*. 2010; 12(5):820–830. doi: [10.1111/j.1438-8677.2009.00268.x](https://doi.org/10.1111/j.1438-8677.2009.00268.x) PMID: [20701707](https://pubmed.ncbi.nlm.nih.gov/20701707/)
70. Grime JP, Mowforth MA. Variation in genome size an ecological interpretation. *Nature*. 1982; 299: 151–153.

71. Machado RCR, Hardwick K. Does carbohydrate availability control flush growth in cocoa? In: Proc. 10th International Cocoa Research Conference. Proceedings, Santo Domingo: Dominican Republic. 1998: 151–157.
72. Sarkar D, Kundu A, Saha A, Mondal NA, Sinha MK, Mahapatra BS. First nuclear DNA amounts in diploid ( $2n = 2x = 14$ ) *Corchorus* spp. by flow cytometry: genome sizes in the cultivated jute species (*C. capsularis* L. and *C. olitorius* L.) are ~300% smaller than the reported estimate of 1100–1350 Mb. *Caryologia*. 2011; 64(2):147–153.
73. Benor S, Fuchs J, Blattner FR. Genome size variation in *Corchorus olitorius* (Malvaceae s.l.) and its correlation with elevation and phenotypic traits. *Genome*. 2011; 54(7):575–585. doi: [10.1139/G11-021](https://doi.org/10.1139/G11-021) PMID: [21745142](https://pubmed.ncbi.nlm.nih.gov/21745142/)
74. Samad MA, Kabir G, Islam AS. Interphase nuclear structure and heterochromatin in two species of *Corchorus* and their F1 hybrid. *Cytologia*. 1992; 57(1): 21–25.
75. Marinho RC, Mendes-Rodrigues C, Bonetti AM, Oliveira PE. Pollen and stomata morphometrics and polyploidy in *Eriotheca* (Malvaceae-Bombacoideae). *Plant Biology*. 2014; 16(2): 508–511. doi: [10.1111/plb.12135](https://doi.org/10.1111/plb.12135) PMID: [24341784](https://pubmed.ncbi.nlm.nih.gov/24341784/)
76. Glicenstein LJ, Fritz PJ. Ploidy level in *Theobroma cacao* L. *J Hered*. 1989; 80(6):464–467.
77. Zhang M, Mao W, Zhang G, Wu F. Development and characterization of polymorphic EST-SSR and genomic SSR markers for Tibetan annual wild barley. *PLoS One*. 2014; 9(4): e94881. doi: [10.1371/journal.pone.0094881](https://doi.org/10.1371/journal.pone.0094881) PMID: [24736399](https://pubmed.ncbi.nlm.nih.gov/24736399/)
78. Sullivan AR, Lind JF, McCleary TS, Romero-Severson J, Gailing O. Development and characterization of genomic and gene based microsatellite markers in North American red oak species. *Plant Molecular Biology Reporter*. 2013; 31(1): 231–239.
79. Morgante M, Hanafey M, Powell W. Microsatellites are preferentially associated with non repetitive DNA in plant genomes. *Nature Genetics*. 2002; 30:194–200. doi: [10.1038/ng822](https://doi.org/10.1038/ng822) PMID: [11799393](https://pubmed.ncbi.nlm.nih.gov/11799393/)
80. Gupta PK, Rustgi S. Molecular markers from the transcribed/expressed region of the genome in higher plants. *Functional and Integrative Genomics*. 2004; 4(3):139–162. doi: [10.1007/s10142-004-0107-0](https://doi.org/10.1007/s10142-004-0107-0) PMID: [15095058](https://pubmed.ncbi.nlm.nih.gov/15095058/)
81. Chistiakov DA, Hellemans B, Volckaert FA. Microsatellites and their genomic distribution, evolution, function and applications: a review with special reference to fish genetics. *Aquaculture*. 2006; 255: 1–29.
82. Mantello CC, Suzuki FI, Souza LM, Gonçalves PS, Souza AP. Microsatellite marker development for the rubber tree (*Hevea brasiliensis*): characterization and cross-amplification in wild *Hevea* species. *BMC Research Notes*. 2012; 5:329. doi: [10.1186/1756-0500-5-329](https://doi.org/10.1186/1756-0500-5-329) PMID: [22731927](https://pubmed.ncbi.nlm.nih.gov/22731927/)
83. Tang J, Gao L, Cao Y, Jia J. Homologous analysis of SSR-ESTs and transferability of wheat SSR-EST markers across barley, rice and maize. *Euphytica*. 2006; 151:87–93, 2006.
84. Alves RM, Silva CRS, Silva MSC, Silva DCS, Sebbenn AM. Diversidade genética em coleções amazônicas de germoplasma de cupuaçuzeiro [*Theobroma grandiflorum* (Willd. ex Spreng.) Schum.]. *Revista Brasileira de Fruticultura*. 2013; 35(3): 818–828.
85. Venturieri G, Venturieri C. Calogênese do híbrido *Theobroma grandiflorum* x *T. obovatum* (Sterculiaceae). *Acta amazônica*. 2004; 34(4):507–511.
86. Silva FCO, Neto EF, Kodama KR, Figueira A. Avaliação das relações genéticas entre genótipos de cacauzeiro (*Theobroma cacao* L.) contrastantes para reação à vassoura-de-bruxa através de marcadores RAPD. *Genetics and Molecular Biology*. 1998; 21: 205.
87. Alves RM, Sebbenn AM, Artero AS, Figueira A. Microsatellite loci transferability from *Theobroma cacao* to *Theobroma grandiflorum*. *Molecular Ecology Resources*. 2006; 6: 583–586.
88. Oliveira GAF, Pádua JG, Costa JL, Jesus ON, Carvalho FM, Oliveira EJ. Cross-species amplification of microsatellite loci developed for *Passiflora edulis* Sims. in related *Passiflora* Species. *Brazilian Archives of Biology and Technology*. 2013; 56(5):785–792.
89. Galbraith D, Lambert G, Macas J, Dolezel J. Analysis of nuclear DNA content and ploidy in higher plants. *Current Protocols in cytometry*. 2002:7.6.1–7.6.22.
90. Flavell RB, Odell M, Thompson WF. Regulation of cytosine methylation in ribosomal DNA and nucleus organizer expression in wheat. *Journal Molecular Biology*. 1998; 204(3): 523–534.
91. Bennett MD, Leitch IJ. *Genome size evolution in plants*. San Diego. Elsevier Inc. 2005: 89–162.
92. Avise JC. Perspective: conservation genetics enters the genomics era. *Conservation Genetics*. 2010; 11(2):665–669.