

Microsatellites and the Polyploid Guarana Plant: Diversity under a Sea of Alleles

Paula Cristina da Silva Angelo^{1*}, Maria do Perpétuo Socorro Lira²,
Manuella Villar Amado³, Ana Takagaki Yamaguishi⁴, Gilvan Ferreira Silva¹,
Jorge Ivan Rebelo Porto⁵, Spartaco Astofi-Filho⁶, André Luiz Atroch¹

¹Embrapa Western Amazon, Manaus, Brazil

²CNPq Fellowship at Embrapa Western Amazon, Manaus, Brazil

³Federal University of Espírito Santo, Vitória, Brazil

⁴Embrapa Genetic Resources and Biotechnology, Brasília, Brazil

⁵Instituto de Pesquisas da Amazônia, Manaus, Brazil

⁶Federal University of Amazonas, Manaus, Brazil

Email: paula.angelo@embrapa.br

Received 4 April 2014; revised 3 May 2014; accepted 2 June 2014

Copyright © 2014 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Repeat blocks, microsatellites or simple sequence repeats (SSRs) can produce good co-dominant molecular markers for genetic diversity analysis and the determination of self-pollination rates in progenies originating from open pollination of selected genotypes. The enrichment of guarana genomic libraries was underway when it was confirmed that we are working with a complex polyploid species with 210 chromosomes. The probes (CA)₁₂, (CT)₁₂ and (TC)₁₄ were used to finish the enrichment of four libraries for repeat blocks and the screening of a databank of expressed sequence tags (ESTs) from guarana seeded-fruits was accomplished as well. Fifteen clonal cultivars were genotyped with three replicas at 10 out of 27 identified loci using the 59 alleles that passed the reproducibility criterion. A large number of short repeat blocks were identified and this was considered to be a consequence of the recent polyploidization event. However, blocks with eight or more repeats ideal for genotyping were scarce. Annealing of most probes to short blocks by partial complementarity could explain the scarcity of longer blocks in genomic libraries but cannot explain why they were rare in the ESTs. Due to the complexity of the genotypes, alleles were treated as dominant traits. ESTs harboring repeat blocks had the functional annotation renewed. Locus GRN07 is inserted in a homologue of the *MOTHER OF FLOWERING LOCUS T AND TFL1 (MFT)*, in which 3'-UTR displays clear post-transcriptional regulatory features. *MFT* and its variants are probably involved in the determination of seed germination and embryo growth characteristics. Other accessed loci can be involved in plant architecture and defense reactions. It was concluded that the alleles described in the present work can be used to distinguish guarana culti-

*Corresponding author.

How to cite this paper: Angelo, P.C.S., Lira, M.P.S., Amado, M.V., Yamaguishi, A.T., Silva, G.F., Porto, J.I.R., Astofi-Filho, S. and Atroch, A.L. (2014) Microsatellites and the Polyploid Guarana Plant: Diversity under a Sea of Alleles. *Open Journal of Genetics*, 4, 190-201. <http://dx.doi.org/10.4236/ojgen.2014.43020>

vars and possibly to analyze segregation using the progenies of controlled pollinations between divergent genitors. Also, the fingerprints obtained can be added to the morphological and agronomic descriptors of the cultivars.

Keywords

Amazon Rainforest, *Paullinia cupana* var. *sorbilis*, SSR, MFT, Seed Germination

1. Introduction

The guarana plant (*Paullinia cupana* Kunth *sorbilis* (Mart.) Ducke) is a liana native to the Amazon rainforest that acquires the shrub habit when cultivated in open fields. It is included in tribe Paullinieae [1] or Clade X [2] of the Sapindaceae. The extracts prepared with the powder made from guarana roasted seeds have been used as stimulant for centuries by traditional Amazonian people such as the Saterê-Mawé [3]. This and the other biological properties of guarana, e.g. anti-tumorigenic activity, have been the targets for recent investigation [4]-[6].

Over the past 40 years, Embrapa Western Amazon has conducted a breeding program of selected guarana plants aimed at improving productivity and disease resistance. Molecular markers can be important tools in plant breeding programs. They have been used for the identification and distinction (or discrimination) of genotypes and for the quantification of variability in the DNA, which can subsequently be correlated with the divergence of phenotypes. The integration of recombination by breeding, selection, and molecular data analyses leads to a faster gain in results [7]. In perennial species, the assistance of molecular markers (MAS) allows selection to be carried out before the traits of interest are expressed, representing a major incentive to employ molecular techniques. When molecular markers are located within genes of interest the process can be referred to as gene-assisted selection (GAS) [8].

Dominant markers were used for diversity analysis in guarana plants, but dominant markers usually only support approximate assertions about identical individuals. The high degree of multi-allelism and the co-dominant Mendelian inheritance of microsatellites (single sequence repeats or repeat blocks) provide a powerful system for the unique identification of individuals for fingerprinting purposes and parentage testing, particularly when the individuals are expected to be related [9].

We had already initiated the search for repeat blocks when it was definitely stated that guarana plants have 210 chromosomes and are allopolyploids [10]. The advent of the polyploidization is thought to be recent [3] and extremely impacting, taking into account that additional seven species in the genus *Paullinia* have all $2n = 24$ chromosomes [11] and genera on tribe Paullinieae have basal numbers of chromosomes (x) of 7, 10, 11, 12 or 14 [12]. Among complex polyploids, breeding assisted by co-dominant markers has been used, for instance, in sugarcane and wheat, but experimentation has been greatly improved by the identification of the ancestors and/or the knowledge about the contributions of different genomic complements for cultivated genotypes, by the existence of highly saturated linkage maps and, by the access to whole-genome data that were recently made available [13]-[16].

The objectives of the present work were to identify repeat blocks in the genome of guarana plants, to design primers, to amplify and choose the loci displaying reliable variability to use as microsatellite markers and, to validate these loci by genotyping guarana cultivars, which were selected by their valuable agronomic characteristics. We used 10 primer pairs, some of which were located within genes involved in physiological processes such as seed viability, determination of the plant architecture and defense against pathogens. Although we still cannot determine which ancestral genomes contributed to the currently cultivated genotypes of the guarana plant or which are the dosages of the alleles, the fingerprints can be used as descriptive characteristics [9] and to assist selection in progenies of highly contrasting genitors displaying phenotypes determined by divergent allele constitutions in loci controlling interesting agronomical traits.

2. Material and Methods

2.1. Genomic Libraries Enrichment

Four libraries of Sau3AI and Mse I genomic fragments were enriched for repeat blocks using probes (CA)₁₂,

(CT)₁₂, (CA)₁₂ + (CT)₁₂ or (TC)₁₄, according to the procedures described in [17] and modified as in [18] and [19].

2.2. Screening of the Guarana Seeded-Fruits ESTs Databank

The guarana (*Paullinia cupana* Kunth *sorbilis* (Mart.) Ducke) seeded-fruits EST data bank (GenBank accession numbers EC763506-EC778393) was screened for sequences harboring repeat blocks using the *TROLL* routine from the Staden software package [20]. cDNA library preparation and EST sequencing are described in [21].

2.3. Plant Material and DNA Extraction

DNA was extracted from liquid N₂-frozen leaves of 15 of the only guarana cultivars developed to date worldwide, which are maintained at Embrapa Western Amazon (03°06'07"S - 60°01'30"W) and recommended for cultivation in the State of Amazonas (Brazil). The codes used for the identification of these cultivars are 189, 372, 388, 505, 608, 610, 611, 612, 624, 626, 648, 861, 871, 882 and 850. Chemicals and instructions for DNA extraction were provided with the YGP100 kit (BMC Technologies). DNA extracts were quantified using a NanoDrop and examined following resolution in 0.8% agarose gels.

2.4. Development of Primers and PCR Reactions

Primers were designed for single ESTs when repeat blocks were identified in the CONTIGs (sets of singlets aligned to construct the most reliable complete sequence) of the databank. The software *Primer 3* (http://biotools.umassmed.edu/bioapps/primer3_www.cgi) was set to an optimal GC content of 45%, optimal TM of 54°C and optimal length of 19 nucleotides (nt). Initially, the amplicons produced with 27 primer pairs were evaluated in 2.5% agarose gels and/or 6% polyacrylamide gels. Nineteen pairs were selected for the addition of primer tails. Of these 19 pairs, 10 primer pairs that were designed to anneal to the 5'-FAM*TGTAACGACGGCCAGT-3' tail were selected due to better integration of the fluorescence and intensity of the fluorescent amplicon produced. These 10 primer pairs were used for automatic genotyping in an ABI 3100 machine. The polymerase chain reactions (PCRs) included 2.5 mM dNTP mix, 2.5 mM MgCl₂, 0.25 µg/µL bovine serum albumin, 150 nM each primer, 150 nM primer tail*FAM, 15 ng of DNA, 1 U of Taq DNA polymerase and 1 X Taq polymerase buffer (Phonetrria) in a total volume of 13 µL. The PCR machine was programmed to run for 3 min at 95°C, followed by 45 s at 57°C, 30 s at 72°C and 20 s at 95°C for 38 cycles, with a final synthesis step at 72°C for 30 min. The PCR and automatic genotyping assays were replicated three times per locus in each cultivar. The peaks/alleles were automatically identified by their lengths using the *label peaks* tool of the *Genotyper* software (Applied Biosystems) and a ladder labeled with ROX. The *filter labels* tool was set to eliminate peaks differing from 0.1 to 0.8 nt, which was sufficient to preserve and identify alleles that differed from each other by one nucleotide while simultaneously eliminating most artifacts.

2.5. Transferability of Primers for *Litchi chinensis*

The primers reported for lychee [22] were tested on guarana plants with slight modifications in an attempt to determine the best annealing temperature.

2.6. Data Analyses and Statistics

Due to the complex polyploid nature of the species under study, the alleles that were reproduced twice within three replicas were tabled as dominant characters and given a value of "1" for presence and "0" for absence. The power of each allele to distinguish guarana cultivars was evaluated with the software *pars* designating the value "1" for a specific allele per turn and "0" for all the others and applying sufficient turns to assure that the value "1" has been given to each allele at least once. The 59 phylograms produced were compared using the software *treedist* and through visual examination. For the analysis of divergence, 2000 bootstrap permutations were generated using *seqboot*. Peaks that were reproduced twice within three replicas but displayed characteristically very low intensities and/or very irregular areas were maintained in the analysis with the half of the weights given to the other peaks (weight = "1" in half the permutations and weight = "0" in the other half). Matrices of paired distances were produced with *restdist* set to use the *neighbor-joining method* and *site length* = 20. Trees

(phylip files) were produced for each of the 2,000 permutations using the software *neighbor*. These phylip files were submitted to *consense* and to *drawtree* to obtain the cladogram for the 15 cultivars. *Pars*, *treedist*, *seqboot*, *restdist*, *neighbor*, *consense* and *drawtree* are part of the *PHYLIP* package, version 3.6 [23]. The *BLAST* routines and the sequences available in the GenBank (<http://www.ncbi.nlm.nih.gov/>) and the TAIR (www.arabidopsis.org/servlets/Tair) databanks were used to identify and compare orthologues. *BIOEDIT* version 7.0.9 [24] was used for alignments. *Splice Machine* (<http://bioinformatics.psb.ugent.be/webtools/splicemachine>) and *ESEfinder* 3.0 (<http://rulai.cshl.edu/cgi-bin/tools/ESE3/esefinder.cgi>) were used online to attribute probability values for the functionality of splicing sites on *MFT*. The *Softberry* package (<http://linux1.softberry.com/cgi-bin/programs>) trained with data from *A. thaliana* was used to predict the position of regulatory sequences as well as micro RNAs (miRNAs) in the *MFT* 3'-UTR (untranslated region) harboring the microsatellite locus GRN07. The databanks of the *Arabidopsis* Small RNA Project (ASRP, <http://asrp.danforthcenter.org/>) were used to identify small RNAs (smRNAs) predicted for the same untranslated region. *RNAsoft* (www.rnasoft.ca) was used to model the secondary structure for the miRNA predicted for the 3'-UTR of the same coding sequence [25]. Z tests of proportions to compare frequencies of repeat blocks in different sources were performed using *SigmaPlot* version 2.0 ($P \leq 0.5$).

3. Results and Discussion

Among the 688 sequences obtained from the enrichment of the genomic libraries, 28% had perfect blocks with at least four dinucleotide repeats (**Figure 1(a)**). This result indicated an average enrichment of 13.8%. No significant differences in enrichment were verified when the four libraries were compared. Blocks with perfect arrays of four repeats were significantly more frequent than were those with five (13%) or six, seven and eight repeats together.

The data described above differ from those obtained for most plant species using (TC)₁₂₋₁₄ and (AC)₁₂ probes, such as *Caryocar brasiliense* [26], *Ophrys araneola* [27] and oak trees [28]. For *Eucalyptus*, the most highly represented classes of frequency were those with 16 to 20 and 21 to 25 repeats per block, which together accounted for 60% of the blocks most of which (80%) were perfect [29]. Blocks harboring a minimum of 15, 8, and 13 dinucleotide repeats and maxima always above 20 repeats, in perfect and complex compositions, were used for diversity analyses in kiwi [30], coconut [31] and olive tree [32], respectively. In fact, there are indications that loci with higher numbers of repeats per block are more variable than those blocks with few repeats [33]. Therefore, once the screening of the four genomic libraries was completed, we finally assumed that microsatellite loci theoretically ideal to evaluate genetic diversity would be rare in the genome of guarana plants.

The results from the seeded-fruits ESTs followed the same trend. Repeat blocks of dinucleotides were present in 3814 of the 4999 ESTs harboring microsatellites. AG/TC repeats were the most frequent, followed by AC/TG, AT and CG repeats (**Figure 1(b)**). However, the frequency of perfect blocks with eight or more repeats was 0.43% (66 blocks in 15,387 ESTs). This result was not significantly different from 0.29% (2/688) calculated for the genomic libraries. *Arabidopsis*, soybean, wheat, maize and rice exhibit around nine or ten dinucleotide repeats per block as average in transcribed regions [34].

A low frequency of perfect dinucleotide blocks can be related to large genome sizes, which are generally associated with the accumulation of “junk DNA”. The wheat haploid genome consists of 5600 Mbp, with 43.4 dinucleotide blocks per Mbp in nontranscribed (genomic sequences in general) regions and 61.0 dinucleotide blocks per Mbp in transcribed regions (ESTs). These numbers are lower than those reported for the 125 Mbp genome of *A. thaliana*, which displays 77.7 and 137.0 dinucleotide blocks per Mbp in nontranscribed and transcribed regions, respectively. A generalized increase in genome sizes and the accumulation of “junk DNA” were most likely simultaneous consequences of transposon amplification occurring in the evolutionary eras subsequent to the Tertiary, after the ancient genomic polyploidizations that apparently affected most plant species. Transposons are more efficiently eliminated from transcriptionally active hypomethylated regions of chromatin, where most repeat blocks occur. Therefore, the frequency of repeat blocks would be negatively correlated with that of “junk DNA”, which is abundant in large genomes, but would be positively correlated with that of singlecopy and low-copy-number genes, which are better represented in ESTs than in other regions of the genome [34]. The diploid nuclei of guarana leaves contain 22.8 pg of DNA [10], what lead to an estimate of 22 Gbps and to a expectative to find relatively few repeat blocks. Nevertheless, for the analysis of the seeded-fruits transcriptome,

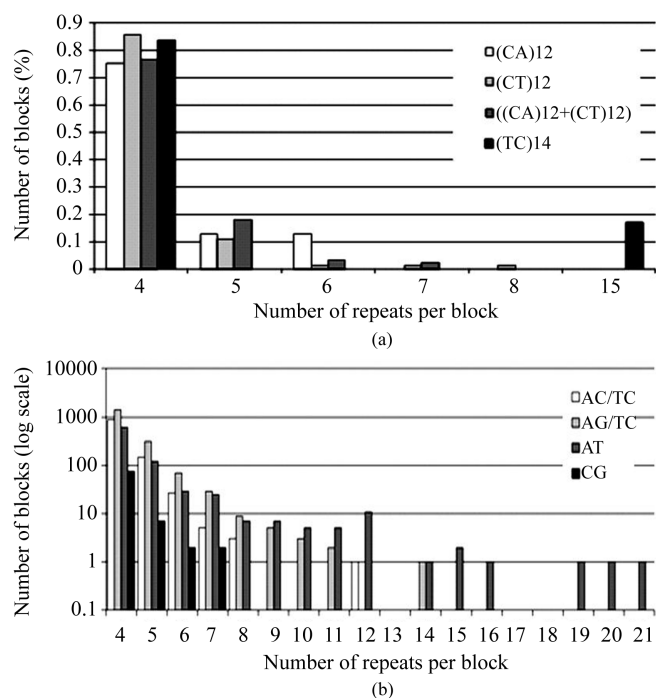


Figure 1. Number of repeat blocks identified in the genomic libraries (a) and in the databank of ESTs from seeded-fruits (b) of *Paullinia cupana* var. *sorbilis*, the guarana plant. Each genomic library in (a) is identified according to the probe used for the enrichment. The percentage of repeat blocks is relative to the number of cloned and sequenced fragments that were complementary to the probes. Repeat blocks identified in ESTs (b) are represented according to their composition. A logarithmic scale was used to facilitate visualization.

15,387 ESTs with average length of 773 nt were accessed [21], and a rough calculation for the presence of 320 blocks with at least four dinucleotide repeats per Mbp is admissible. This number of blocks is 2.3-folds higher than that in *Arabidopsis* and is even more intriguing when one considers that: 1) three-repeats blocks were counted for *Arabidopsis*, while we only counted blocks with four repeats or more for guarana plants, and 2) *Arabidopsis* presents a genome that is 6000-fold smaller than that of guarana. Indeed, our results sounded completely detached from the hypothesis in [34]. However, it must be considered that a recent, possibly 1000 to 2000 years ago, polyploidization event is part of the origin of cultivated guarana plants [3]. Repeat blocks already present in transcribed regions of the genomes of guarana ancestors, which became suddenly replicated on the polyploidization process that followed a very probable interspecific hybridization occurrence [10], must have been equally replicated and came out to the relatively high frequencies we detected in the allopolyploid genome cultivated nowadays.

The accumulation of short repeat blocks in guarana is evident (Figure 1) and it can be inferred that genomic fragments harboring short repeats hybridized so frequently with the probes used for enrichment of the libraries that fragments harboring eight or more repeats had severely reduced chances of isolation. For such a large genome, it can be assumed that re-enrichment trials would not be a solution. However, this is not the reason for the low frequency of blocks harboring eight or more repeats in ESTs, which are accessed using completely different techniques. Theoretically, blocks with very high numbers of repeats are more likely to influence the conservative replication of transcribed regions by unequal crossing or slippage, and this condition can lead to some counter-selection of very long blocks in these regions. Nevertheless, upon examining Figure 1(b) it can be concluded that a considerable number of longer blocks would be necessary to approximate the mean number of repeats per block in guarana ESTs from those nine or ten reported as average for the other species cited above. Whether these results can be related to the recent polyploidization of guarana plants is not clear and we could not yet construct a reasonable hypothesis to explain it.

Genotyping of 10 loci (**Table 1**) for 15 elite clones selected and recommended for cultivation in the State of Amazonas (Brazil) was performed in triplicate as an attempt to increase the reliability of the alleles while eliminating most of the experimental artifacts. Data were arranged according to the guarana cultivar, and peaks that did not contribute to a reproducibility score of “2 in 3” were inventoried as missing data. GRN05 and GRN12 were the loci with the lowest reproducibility scores, as many peaks were present in only one out of three replicas.

Fifty nine alleles were included in the diversity analyses. Reproducible peaks with very low intensities or very irregular areas were given half the weight given to more intense peaks displaying classical approximately triangular areas. Whether these peaks are unique characteristics valuable for the identification of certain cultivars will be verified in future experiments. No two cultivars were identical among the 15 that were analyzed. Bootstrap values above the partitions in the cladogram (**Figure 2**) are related to the efficiency of the alleles to discriminate the different cultivars. Higher bootstrap values indicate that many divergent alleles will be necessary to fractionate the correspondent clade and to identify its members individually. Alleles L12259, L12265, L14208, L17190, L17198, L20235 and L20238 were worthy distinctive characters, as defined by the use of the softwares *pars* and *treedist*.

Due to the allopolyploidy, even more than five reproducible alleles per guarana plant were present at some loci (**Figure 3**). Alleles differing by one or three nucleotides from each other were preserved for diversity analysis, even when observed on loci harboring blocks of dinucleotides (**Figure 3(b)**, allele 217), especially considering the high level of polyploidy of the studied plant species. It has been done before for kiwi [30] and olive

Table 1. Characteristics of the primer pairs used to access ten microsatellite loci and genotype guarana (*Paullinia cupana* var. *sorbilis*) cultivars.

LOCUS	REPEAT BLOCK	PRIMERS F AND R	AL	EXP SIZE
GRN01	(CA) ₆ (CA) ₃	AGAACTGGTCCAACCGTCTC and CGTGAAAGGTCTGAGTGAAGC	4	110
GRN04	CA(CCA) ₄ CA(CCA)	CATCCATTGTCACTCTTGG and TGGCATGAGACAATTTGTGG	1	209
GRN05	(TA) ₅ (AT) ₅	GATCAGGTGCCACTCCAAGT and TTGTGTGAGTCCTCAGCCTCT	5	233
GRN07	(TA) ₁₆	GGTTCTTTTCAGCGCAGTTG and GGCATAGAGCACCGAGAGAC	8	249
GRN09	(AT) ₈ C(TA) ₁₂	CTGAATGCTGTCCAAGCA and GCAGCCTTCCCATTTTACC	5	256
GRN12	(TG) ₅ (TA) ₇	TGTCAAACCCCTTTATGTTC and CAATGGTGCCAGTAAATACA	5	244
GRN13	(CT) ₅ (TA) ₇	TCGAAATATATGTGGCATGA and TGCATAAAAATCACCAAATCA	1	252
GTN14	(TC) ₁₁	ATGATCATCAGCATCATGG and TATCCAACCTCAATTCCCAGA	13	191
GRN17	(CT) ₉	TGCGACTGTGAGTGAGCTA and TTGAGATGAACTCAGCCAAC	10	180
GRN20	(CGC) ₅ (CGC) ₃	TGGGAGGAGAGAACCGTACA and GCCTTCCCTTCTATCAAGC	7	222

OBS.: annealing temperature was 57°C for all the primer pairs. EXP SIZE = expected size of the amplicons. AL = alleles passing the reproducibility criterion.

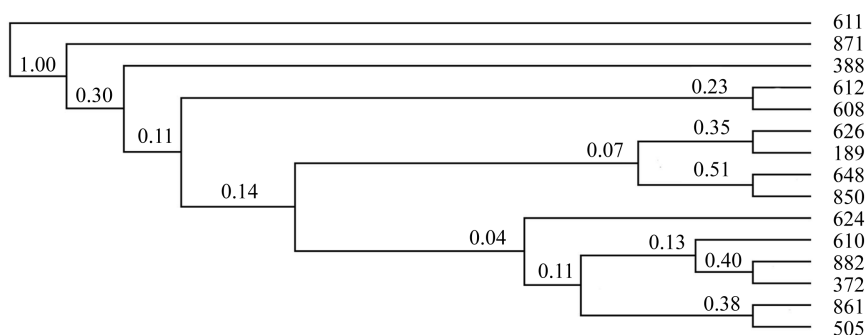


Figure 2. Cultivars of guarana (*Paullinia cupana* var. *sorbilis*) grouped by the neighbor-joining method after 2000 bootstrap permutations. The numbers above partitions represent their frequency in the permuted data sets. Cultivars are identified by their code in the breeding program.

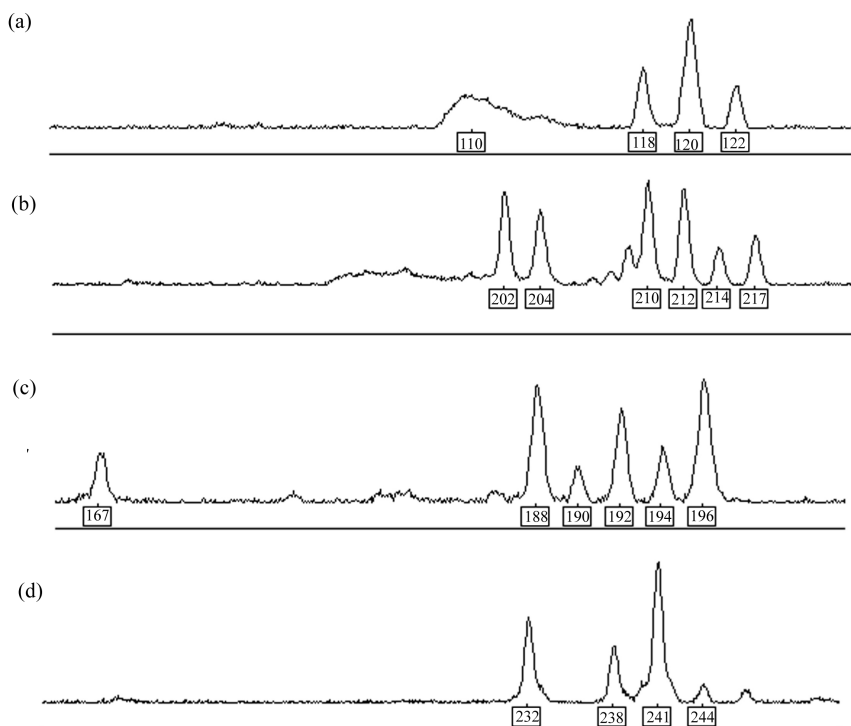


Figure 3. Chromatograms produced by genotyping guarana (*Paullinia cupana* var. *sorbilis*) plants with primers for the microsatellite loci GRN01 (a), GRN14 (b), GRN17 (c) and GRN20 (d).

tree [35]. Besides, different numbers of repeats but also indels in the flanking regions of the blocks were found while transferring primers from *Brassica rapa* to other species of the same genus [36]. So, it is possible that differences of one or three nucleotides in a dinucleotide repeat block can occur, e.g., among homeologous chromosomes. Two exclusive alleles were identified, L14215 and L20230, for cultivars 626 and 608, respectively.

It was not possible to conclude about the number of copies (the dosage) of the alleles. Some work using the areas below the peaks can be useful to accomplish this type of analysis, which depends on information about the ancestors as demonstrated for hemisexual and polyploid *Rosa* L. Sect. *Caninae* D.C. [37] or on the assessment of information about the different dosages of the same alleles in genitors and the resultant hybrid progenies by other means [16]. Two tendencies were observed regarding the number of alleles in each locus: two alleles and multiples of two (**Figure 3(a)**) or three alleles and multiples of three (**Figure 3(b)**) per plant. Loci displaying five alleles were also present (**Figure 3(c)**). To organize the karyogram of the cultivated variety, a set of homologous/homeologous chromosomes was arranged in groups of two pairs (metacentrics and submetacentrics), while another set was arranged in groups of three pairs (acrocentrics) [10]. It is possible that these two tendencies for the occurrence of alleles at different loci are related to the amplification of loci present in each of these two sets of chromosomes and/or in both of them simultaneously.

Seven of the 10 loci used in the present work were identified in the sequences of ESTs included in the guarana seeded-fruits databank [21]. Their orthologues in the GenBank or the conserved domain identified in the orthologues are listed below (**Table 2**).

Locus GRN07 was identified in a CONTIG of ESTs that is homologous to the *MOTHER OF FLOWERING LOCUS T AND TFL1 (MFT)* coding sequence from *Fragaria vesca* (**Table 2**). *MFT* encodes a transcription factor that is expressed in seeds and is responsive to abscisic acid and gibberellins, which are plant growth regulators controlling the time necessary for germination in *Arabidopsis* [38]. The long *MFT* 3'-UTR, consisting of 365 nucleotides (**Figure 4(a)**), displays at least three characteristics of post-transcriptional regulatory regions: it harbors a repeat block (the TA block in the box); it is the target for small RNAs (smRNA shown in red); and it contains 27 tetranucleotides from eight out of nine types of sequences identified as consensus of putative recognition elements (PREs shown in green) for RNA-binding proteins involved in post-transcriptional regulation.

Table 2. Orthologues of the ESTs in the seeded-fruits databank accessed to develop microsatellite markers used for genotyping guarana (*Paullinia cupana* var. *sorbilis*) plants.

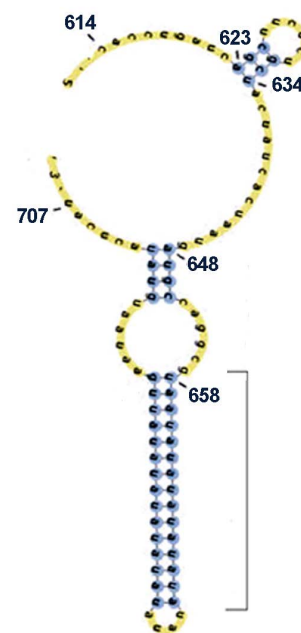
ORTHOLOGUES IN THE GenBank				
LOCUS	CODING SEQUENCE	SPECIES	ID	E-VALUE
GRN07	mother of flowering locus T and TFL1	<i>Fragaria vesca</i>	XP4299541.1	$2 \times e^{-89}$
GRN09	endo-xyloglucan transferase	<i>Gossypium hirsutum</i>	BAA21107.1	$1 \times e^{-136}$
GRN12	cyclase (tryptophan catabolism)	<i>Arabidopsis thaliana</i>	EFH43390.1	$2 \times e^{-39}$
GRN13	gibberellin oxidase	<i>Pisum sativum</i>	ABI64150.1	$4 \times e^{-42}$
GRN14	unnamed protein (TCP transcription factor)	<i>Vitis vinifera</i>	CA048409.1	$8 \times e^{-26}$
GRN17	anfipatic channel (membrane)	<i>Jatropha curcas</i>	ADU56185.1	$6 \times e^{-24}$
GRN20	zinc finger protein C2H2	<i>Ricinus communis</i>	EEF34900.1	$9 \times e^{-49}$

```

AAA TAT ATC TTC TCT CTT GTG TTT TAG CCT CTT TGC TCT CTG TTT GTG TTT GTA ATG 57
GGA TCA ATG GGA ACA CCG GTG GAT CCA TTA GTG GTG GGA AGA ATA ATC GAC GAA GTA 114
      M G T P V D P L V V G R I I D E V 17
GTC GAT TTG TTC CTT AGT GTG AAC ATG ATG GCC TAC TTT AAG AAC AAA CAC GTC 171
V D L F M P S V N M S A Y F K N K H V 36
ACG AAC GGA TGC GAC ATA AAG CCG TCC ATG GCT TCT GAG CCT CCC CGG ATC ACT TTT 228
T N G C D I K P S M A S E P P R I T F 55
TCT GGG AAA CCT GAT GAG CTC TAC ACA CTG ATC ATG ACC GAC CCT GAT GCG CCC AGT 285
S G K P D E L Y T L I M T D P D A P S 74
CCT AGC GAG CCC AGC ATG AGA GAA TGG ATC CAT TGG ATC GTG GTT GAC ATA CCA GGG 342
P S E P S M R E W I H W I V V D I P G 93
GGA GCT AGT CCT TCT CAA GGT ATT GAA ATT CTG CCG TAC ATG GCT CCC AAA CCC CCA 399
G A S P S Q G I E I L P Y M A P K P P 112
ATT GGC ATC CAC CGG TTT ATT CTG ATA CTG TTC AAA CAG AAG TCG AGG CTG GGG ATG 456
I G I H R F I L I L F K Q K S R L G M 131
ATG GAA CCC CCG ATG AAC CGC GCG AAT TTC AAC ACA CGG TTC TTT TCA GCG CAG TTG 513
M E P P M N R A N F N T R F F S A Q L 150
GAC CTG GGC CTT CCA GCG GCA TTA GTG TAT TTC AAC GCA CAT AAA GAG CCA GCA AAC 570
D L G L P A A L V Y F N A H K E P A N 169
AGG CAG TCA CAG GCT AGG AGG ACA TAG TCC CAA AAA TTG AAA GCA CCT GAT CAG CTT 627
R Q S Q A R R T * 177
CTC TGC TAC TAT CAC TAA TGA TGC CAG GCG TAT ATA TAT ATA TAT ATA TAT 684
ATA TAT TG ATA ATG TAT ACT CAT GGA GAG GGG CGG TAT TCT CTC GGT GCT CTA TAC 741
CAT TGA GTT CTC CTG ATC GGG TCG GGT TAT TTG TCC TTA TTT GTG TTT TGG TCT TGT 798
GTT GTC TGT TCG GAT CGT GTT GTC TCT TCG GGT CGA GTC GTG TCA Gac CAA GaT GtA 855
ATC TTT TCT TTG AGT ATG GTG TGT CCT TTG TA TGA ACT TCT AGC TTA TTA AGT TGT 912
ATG TAG GTG AAA CTA GTT TGG TTT CTT TTA TGT ATC TGG TGG TTT CCT CTA aaa aa 968

```

(a)



(b)

Figure 4. (a)—Sequence from the guarana (*Paullinia cupana* var. *sorbilis*) seeded-fruit databank that harbors the TA repeat block (in the box) in microsatellite locus GRN07. The amino acids below the codons constitute the deduced peptide homologous to the mother of flowering locus T and TFL1 transcription factor. Three predicted splicing sites are indicated by black arrow heads. A stop codon is represented by an asterisk. The first six adenines in the poly(A) tail are represented by italicized bold letters. The predicted regulatory motifs in the long 3'-UTR are identified by different colors: a consensus sequence for interaction with the EIN3 transcription factor activated by ethylene is shown in blue; the homologous of a smRNA from *A. thaliana* that also targets the *CYTOSOLIC MALATE DEHYDROGENASE* coding sequence (ASRP and [39]) is displayed in red; 4-mer motives significantly enriched among Putative Recognition Elements (PREs) involved in post-transcriptional “regulons” directed by RNA-binding proteins [40] are displayed in green. Putative polyadenylation signals [41] are presented in bold italicized underlined capital letters and often coincide with other potential regulatory signals. (b)—Secondary structure for the pre-miRNA predicted to be transcribed from nucleotide 614 to 733. The bracket indicates the predicted mature miRNA.

These characteristics were identified through the analysis of untranslated regions in six dicotyledonous families and have presumably been preserved for the last 70 million years, despite they are not protein coding sequences. Corroborating the hypothesis of regulatory power, these elements are preferentially observed in association with major open reading frames that code for transcription factors [40], which is the *MFT* in this study.

Variations in the number of TA repeats of the block inside the microsatellite locus GRN07 could influence transcript maturation or processing and be related to differences in the time necessary for seed germination. Changes in the 3' portion of the TA block could result in the emergence or disruption of proximal/distal polyadenylation signals, which are relatively free variations on the themes UUUGUA and AAUGAA. The efficiency of polyadenylation, the length and the age of poly(A) tails are directly related to the stability and translatability of mRNAs [41] [42]. On the other hand, the same TA repeat block displays a predicted potential to function as a TATA box (*Softberry-TSSP*) and to promote the transcription of smRNAs located downstream. By their turn, interactions with smRNAs could repress the expression of *MFT* in the nucleus, impairing polyadenylation, or in the cytosol, attaching to the mRNA and inhibiting translation, as reported for the *SPL3* gene in *A. thaliana* [43]. The coordinated repression of *MFT* and *CYTOSOLIC MALATE DEHYDROGENASE* by a single small RNA (smRNA in red in **Figure 4(a)**) in the nucleus or the cytosol could lead to the prevalence of sucrose synthesis and accumulation over glycolysis while the fruit is functioning as a photosynthate sink and the embryo is dormant. *MFT* regulation by ethylene (regulatory element in blue in **Figure 4(a)**) could also be part of the sequential programs of gene expression concurring for fruit maturation and dehiscence, embryo dormancy and seed germination [44]. Hypothetically, annealed smRNAs could open the chromatin facilitating transcription by RNA polymerase II and keep it open, even when the major ORF has been methylated and entered the heterochromatin. Finally, there is at least one predicted pre-miRNA (*Softberry-findmirna*) in the *MFT* 3'-UTR, which could generate 21 or 24 nucleotides long mature miRNAs with the sequence 5'-UGCCAGGCGUAAUAUAUAUAU(AUA)-3' (**Figure 4(b)**). These elements are frequently related to disease defense or transposon silencing.

Locus GRN13 is inserted in a homologue of *GIBBERELLIN OXIDASE* from *Pisum sativum* (**Table 2**) and was found to be monomorphic, with a single allele—L13236—identified for all guarana cultivars. Because this same single allele was observed in the 15 cultivars selected for productivity from different backgrounds it is possibly the *OXIDASE* allelic variant that better contributes to the maintenance of a proper concentration of gibberellins in dormant embryos [44], avoiding viviparity. Locus GRN14 includes the repeat block located in the 3'-UTR of the homologue of a *Vitis vinifera* transcription factor harboring the TCP domain, which can be related to the architecture of the branches and leaves, flowering time, defense reactions and seed viability [45]-[48]. Locus GRN20 lays in a homologue of a zinc finger transcription factor, which is member of a very large C2H2 protein family whose members are also related to the onset and development of the fruit (www.arabidopsis.org/servlets/Tair). In barley, nine-repeats blocks are in loci related to flowering time. The segregation pattern of two of these blocks combined explained 15% of the variability of this phenotypic characteristic when the plants were evaluated under short days. One of the loci—HVM40—is a complex repeat block consisting of (GA)₆(GT)₄(GA)₇, while the other is composed of ATs [49] [50], such as GRN07 and part of GRN13 from the guarana plants. Combinations among other of those loci in barley explained more than 30% of the variability in the time necessary for flowering [49] [50].

The primers developed for *Litchi chinensis* [22], another Sapindaceae from tribe Nephelieae, were tested for transferability. Only LMLY5, LMLY7 and LMLY8 produced amplicons from guarana plant DNA, although the results (not shown) were not reliably reproducible. This was interpreted as a consequence of the evolutionary distance between Nephelieae and Paullinieae. Accordingly, just four out of the 12 primer pairs have previously been successfully applied for longan (*Euphoria longan*) [22], which is included in the tribe Nephelieae as well.

4. Conclusion

In conclusion, dinucleotide blocks harboring eight or more repeats are less frequent in the genome of guarana plants compared with most plant species. The primer set reported in the present work can provide descriptive characteristics and can be used to distinguish between the genotypes of the 15 guarana (*Paullinia cupana* var. *sorbilis*) analyzed cultivars. Some of the primer pairs described here, such as GRN07 and GRN14, can also be useful in future research on seed viability, fruit setting time, the uniformity of fruit maturation and/or plant/in-florescence architecture and defense against pathogens using progenies arising from controlled pollination between phenotypic divergent genitors.

Acknowledgements

We are thankful to FAPEAM (grant # 9242003) and EMBRAPA for the financial support and to CNPq for the

fellowship to MPSSL while she was working in this project. We thank Dr. Izeni P. Farias for the collaboration to the cloning and sequencing trails that took place at the Federal University of Amazonas (Manaus-AM) and Dr. Vânia C. R. Azevedo and Dr. Peter W. Inglis for the collaboration to the cloning and sequencing trails that took place at Embrapa Genetic Resources and Biotechnology (Brasília-DF). We thank Cleiton Fantin, Waleska Gravena and Jeferson C. da Cruz for the help to clone and sequence genomic libraries at the Federal University of Amazonas (Manaus-AM); Natália D. M. Carvalho and Susan K. B. Soares at Embrapa Western Amazon (Manaus-AM) and Alexandra M. B. de Souza at INPA (Manaus-AM) for the contribution to the visual curation of data produced by sequencing genomic libraries and using *TROLL* for the EST databank.

References

- [1] Harrington, M.G., Edwards, K.J., Johnson, S.A., Chase, M.W. and Gadek, P.A. (2005) Phylogenetic Inference in Sapindaceae *sensu lato* Using Plastid mat K and rbcL DNA Sequences. *Systematic Botany*, **30**, 366-382. <http://dx.doi.org/10.1600/0363644054223549>
- [2] Buerki, S., Forest, F., Acevedo-Rodríguez, P., Callmender, M.W., Nylander, J.A., Harrington, M., *et al.* (2009) Plastid and Nuclear DNA Markers Reveal Intricate Relationships at Subfamilial and Tribal Levels in the Soapberry Family (Sapindaceae). *Molecular Phylogenetics and Evolution*, **51**, 238-258. <http://dx.doi.org/10.1016/j.ympev.2009.01.012>
- [3] Atroch, A.L., Nascimento-Filho, F.J., Angelo, P.C.S., Freitas, D.V., Sousa, N.R., Resende, M.D.V. and Clement, C. (2012) Domestication and Breeding of the Guarana Tree. In: Borém, A., Lopes, M.T.G., Clement, C.R. and Noda, H., Eds., *Domestication and Breeding: Amazonian Species*, Federal University of Viçosa, Viçosa, 333-360.
- [4] Bittencourt, L., Machado, D., Machado, M., Santos, G.D., Algarve, T., Marinowic, D., *et al.* (2013) The Protective Effects of Guaraná Extract (*Paullinia cupana*) on Fibroblast NIH-3T3 Cells Exposed to Sodium Nitroprusside. *Food and Chemical Toxicology*, **53**, 119-125. <http://dx.doi.org/10.1016/j.fct.2012.11.041>
- [5] Krewer, C.D.C., Ribeiro, E.E., Ribeiro, E.A.M., Moresco, R.N., Rocha, M.I.U.M., Montagner, G.F.F.S., *et al.* (2011) Habitual Intake of Guaraná and Metabolic Morbidities: An Epidemiological Study of an Elderly Amazonian Population. *Phytotherapy Research*, **25**, 1367-1374.
- [6] Portella, R.D.L., Barcelos, R.P., Rosa, E.J.F., Ribeiro, E.E., Cruz, I.B.M., Suleiman, L., *et al.* (2013) Guaraná (*Paullinia cupana* Kunth) Effects on LDL Oxidation in Elderly People: An *in Vitro* and *in Vivo* Study. *Lipids in Health and Disease*, **12**, 2-9. <http://dx.doi.org/10.1186/1476-511X-12-12>
- [7] Ferreira, M.E. and Grattapaglia, D. (1996) *Introdução ao uso de marcadores moleculares em análise genética*. EMBRAPA-CENARGEN, Brasília-DF, Brazil.
- [8] Ruane, J. and Sonnino, A. (2007) Marker-Assisted Selection as a Tool for Genetic Improvement of Crops, Livestock, Forestry and Fish in Developing Countries: An Overview of the Issues. In: Guimarães, E.P., Ruane, J., Scherf, B.D., Sonnino, A. and Dargie, J.D., Eds., *Marker-Assisted Selection, Current Status and Future Perspectives in Crops, Livestock, Forestry and Fish*, Food and Agriculture Organization of the United Nations, Rome, 4-13.
- [9] Grattapaglia, D. (2007) Marker-assisted selection in *Eucalyptus*. In: Guimarães, E.P., Ruane, J., Scherf, B.D., Sonnino, A. and Dargie, J.D., Eds., *Marker-Assisted Selection, Current Status and Future Perspectives in Crops, Livestock, Forestry and Fish*, Food and Agriculture Organization of the United Nations, Rome, 252-281.
- [10] Freitas, D.V., Carvalho, C.R., Nascimento-Filho, F.J. and Astolfi-Filho, S. (2007) Karyotype with 210 Chromosomes in Guarana (*Paullinia cupana* "sorbilis"). *Journal of Plant Research*, **120**, 399-404. <http://dx.doi.org/10.1007/s10265-007-0073-4>
- [11] Solís-Neffa, V.G. and Ferrucci, M.S. (2001) Karyotype Analysis of Some Paullineae Species (Sapindaceae). *Caryologia*, **54**, 371-376. <http://dx.doi.org/10.1080/00087114.2001.10589248>
- [12] Ferrucci, M.S. and Solís-Neffa, V.G. (1997) Citotaxonomia de Sapindaceae Sudamericanas. *Boletín da Sociedad Argentina de Botánica*, **33**, 77-83.
- [13] Nie, L., Han, Z., Lu, L., Yao, Y., Sun, Q. and Ni, Z. (2008) Genomic and Genic Sequence Variation in Synthetic Hexaploid Wheat (AABBDD) as Compared to Their Parental Species. *Progress in Natural Science*, **18**, 533-538. <http://dx.doi.org/10.1016/j.pnsc.2007.09.006>
- [14] Figueira, T.R.E.S., Okura, V., Silva, F.R.D., Silva, M.J.D., Kudrna, D., Ammiraju, J.S., *et al.* (2012) A BAC Library of the SP80-3280 Sugarcane Variety (*Saccharum* sp.) and Its Inferred Microsynteny with the Sorghum Genome. *BMC Research Notes*, **5**, 185. <http://dx.doi.org/10.1186/1756-0500-5-185>
- [15] Pastina, M.M., Malosetti, M., Gazaffi, R., Mollinari, M., Margarido, G.R.A., Oliveira, K.M., *et al.* (2012) A Mixed Model QTL Analysis for Sugarcane Multiple-Harvest-Location Trial Data. *Theoretical and Applied Genetics*, **124**, 835-849. <http://dx.doi.org/10.1007/s00122-011-1748-8>
- [16] Garcia, A., Mollinari, M., Marconi, T., Serang, O., Silva, R., Vieira, M., *et al.* (2013) SNP Genotyping Allows an In-

- Depth Characterization of the Genome of Sugarcane and Other Complex Autopolyploids. *Scientific Reports*, **3**, 3399. <http://dx.doi.org/10.1038/srep03399>
- [17] Tenzer, I., Ivanissevich, S.D., Morgan, M. and Gessler, C. (1999) Identification of Microsatellite Markers and Their Application to Population Genetics of *Venturia inaequalis*. *Phytopathology*, **89**, 748-753. <http://dx.doi.org/10.1094/PHYTO.1999.89.9.748>
- [18] Gautschi, O., Tschopp, S., Olie, R.A., Leech, S.H., Simões-Wust, A.P., Ziegler, A., *et al.* (2001) Activity of a Novel Bcl-2/Bcl-xL-Bispecific Antisense Oligonucleotide against Tumors of Diverse Histologic Origins. *JNCI Journal of the National Cancer Institute*, **93**, 463-471. <http://dx.doi.org/10.1093/jnci/93.6.463>
- [19] Farias, I.P., Hrbek, T., Brinkmann, H., Sampaio, I. and Mayer, A. (2003) Characterization and Isolation of DNA Microsatellite Primers for *Arapaima gigas*, an Economically Important but Severely Over-Exploited Fish Species of the Amazon Basin. *Molecular Ecology Notes*, **3**, 128-130. <http://dx.doi.org/10.1046/j.1471-8286.2003.00375.x>
- [20] Martins, W., de Sousa, D., Proite, K., Guimarães, P., Moretzsohn, M. and Bertioli, D. (2006) New Softwares for Automated Microsatellite Marker Development. *Nucleic Acids Research*, **34**, e31. <http://dx.doi.org/10.1093/nar/gnj030>
- [21] Angelo, P.C.S., Nunes-Silva, C., Brígido, M., Azevedo, J., Assunção, E., Sousa, A., *et al.* (2008) Guarana (*Paullinia cupana* var. *sorbilis*), an Anciently Consumed Stimulant from the Amazon Rainforest: The Seeded-Fruit Transcriptome. *Plant Cell Reports*, **27**, 117-124. <http://dx.doi.org/10.1007/s00299-007-0456-y>
- [22] Viruel, M.A. and Hormaza, J.L. (2004) Development, Characterization and Variability Analysis of Microsatellites in Lychee (*Litchi chinensis* Sonn., Sapindaceae). *Theoretical and Applied Genetics*, **108**, 896-902. <http://dx.doi.org/10.1007/s00122-003-1497-4>
- [23] Felsenstein, J. (2004) PHYLIP (Phylogeny Inference Package) Version 3.6. Department of Genome Sciences, University of Washington, Seattle.
- [24] Hall, T.A. (1999) BioEdit: A User-Friendly Biological Sequence Alignment Editor and Analysis Program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, **41**, 95-98.
- [25] Andronescu, M.S., Pop, C. and Condon, A.E. (2010) Improved Free Energy Parameters for RNA Pseudoknotted Secondary Structure Prediction. *RNA*, **16**, 26-42. <http://dx.doi.org/10.1261/rna.1689910>
- [26] Collevatti, R.G., Brondani, R.V. and Grattapaglia, D. (1999) Development and Characterization of Microsatellite Markers for Genetic Analysis of a Brazilian Endangered Tree Species *Caryocar brasiliense*. *Heredity*, **83**, 748-756.
- [27] Soliva, M., Gautschi, B., Salzmann, C., Tenzer, I. and Widmer, A. (2000) Isolation and Characterization of Microsatellite loci in the Orchid *Ophrys araneola* (Orchidaceae) and a Test of Cross-Species Amplification. *Molecular Ecology*, **9**, 2178-2179. <http://dx.doi.org/10.1046/j.1365-294X.2000.105313.x>
- [28] Hodgetts, R.B., Aleksyuk, M.A., Brown, A., Clarke, C., MacDonald, E., Nadeem, S., *et al.* (2001) Development of Microsatellite Markers for White Spruce (*Picea glauca*) and Related Species. *Theoretical and Applied Genetics*, **102**, 1252-1258. <http://dx.doi.org/10.1007/s00122-001-0546-0>
- [29] Brondani, R.P.V., Brondani, C., Tarchini, R. and Grattapaglia D. (1998) Development, Characterization and Mapping of Microsatellite Markers in *Eucalyptus grandis* and *E. urophylla*. *Theoretical and Applied Genetics*, **97**, 816-827. <http://dx.doi.org/10.1007/s001220050961>
- [30] Huang, W.G., Cipriani, G., Morgante, M. and Testolin, R. (1998) Microsatellite DNA in *Actinidia chinensis*: Isolation, Characterization and Homology in Related Species. *Theoretical and Applied Genetics*, **97**, 1269-1278. <http://dx.doi.org/10.1007/s001220051019>
- [31] Perera, L., Russel, J.R., Provan, J. and Powell, W. (1999) Identification and Characterization of Microsatellite loci in Coconut (*Cocos nucifera* L.) and the Analysis of Coconut Population in Sri Lanka. *Molecular Ecology*, **8**, 344-346.
- [32] Carriero, F., Fontanazza, G., Cellini, F. and Giorio, G. (2002) Identification of Simple Sequence Repeats (SSR) in Olive (*Olea europaea* L.). *Theoretical and Applied Genetics*, **104**, 301-307. <http://dx.doi.org/10.1007/s001220100691>
- [33] Felsenstein, J. (2004) Inferring Phylogenies. Sinauer Associates, Sunderland.
- [34] Morgante, M., Hanafey, M. and Powell, W. (2002) Microsatellites Are Preferentially Associated with Nonrepetitive DNA in Plant Genomes. *Nature Genetics*, **30**, 194-200. <http://dx.doi.org/10.1038/ng822>
- [35] Rallo, P., Dorado, G. and Martín, A. (2000) Development of Simple Sequence Repeats (SSRs) in Olive Tree (*Olea europaea* L.). *Theoretical and Applied Genetics*, **101**, 984-989. <http://dx.doi.org/10.1007/s001220051571>
- [36] Ramchiary, N., Nguyen, V.D., Li, X., Hong, C.P., Dhandapani, V., Choi, S.R., *et al.* (2011) Genic Microsatellite Markers in *Brassica rapa*: Development, Characterization, Mapping, and Their Utility in Other Cultivated and Wild *Brassica* Relatives. *DNA Research*, **18**, 305-320. <http://dx.doi.org/10.1093/dnares/dsr017>
- [37] Nybom, H., Esselink, G., Werlemark, G. and Vosman, B. (2004) Microsatellite DNA Marker Inheritance Indicates Preferential Pairing between Two Highly Homologous Genomes in Polyploid and Hemisexual Dog-Roses, *Rosa* L. Sect. *Caninae* DC. *Heredity*, **92**, 139-150. <http://dx.doi.org/10.1038/sj.hdy.6800332>

- [38] Xi, W., Liu, C., Hou, X. and Yu, H. (2010) *MOTHER OF FT AND TFL1* Regulates Seed Germination through a Negative Feedback Loop Modulating ABA Signaling in *Arabidopsis*. *The Plant Cell*, **22**, 1733-1748. <http://dx.doi.org/10.1105/tpc.109.073072>
- [39] Lee, T.F., Gurazada, S.G.R., Zhai, J., Li, S., Simon, S.A., Matzke, M.A., *et al.* (2012) RNA Polymerase V-Dependent Small RNAs in *Arabidopsis* Originate from Small, Intergenic Loci Including Most SINE Repeats. *Epigenetics*, **7**, 781-795.
- [40] Vaughn, J.N., Ellingson, S.R., Mignone, F. and Arnim, A.V. (2012) Known and Novel Post-Transcriptional Regulatory Sequences Are Conserved across Plant Families. *RNA*, **18**, 368-384. <http://dx.doi.org/10.1261/rna.031179.111>
- [41] Rothnie, H.M. (1996) Plant mRNA 3'-End Formation. *Plant Molecular Biology*, **32**, 43-61. <http://dx.doi.org/10.1007/BF00039376>
- [42] Grierson, D. and Covey, S.N. (1988) Structure and Expression of Nuclear Genes. In: Grierson, D. and Covey, S.N., Eds., *Plant Molecular Biology*, Blackie & Son-Chapman & Hall, New York, 22-46. http://dx.doi.org/10.1007/978-1-4615-3666-6_2
- [43] Gandikota, M., Birkenbihl, R.P., Höhmann, S., Cardon, G.H., Saedler, H. and Huijser, P. (2007) The MiRNA156/157 Recognition Element in the 3'-UTR of the *Arabidopsis* SBP Box Gene *SPL3* Prevents Early Flowering by Translational Inhibition in Seedlings. *The Plant Journal*, **49**, 683-693. <http://dx.doi.org/10.1111/j.1365-313X.2006.02983.x>
- [44] Taiz, L. and Zeiger, E. (2010) Gibberellins: Regulators of Plant Height and Seed Germination. In: Taiz, L. and Zeiger, E., Eds., *Plant Physiology*, Sinauer Associates, Sunderland, 583-619.
- [45] Cubas, P., Lauter, N., Doebley, J. and Coen, E. (1999) The TCP Domain: A Motif Found in Proteins Regulating Plant Growth and Development. *The Plant Journal*, **18**, 215-222. <http://dx.doi.org/10.1046/j.1365-313X.1999.00444.x>
- [46] Martín-Trillo, M. and Cubas, P. (2010) TCP Genes: A Family Snapshot Ten Years Later. *Trends in Plant Science*, **15**, 13-39. <http://dx.doi.org/10.1016/j.tplants.2009.11.003>
- [47] Danisman, S., van der Wal, F., Dhondt, S., Waites, R., de Folter, S., Bimbo, A., *et al.* (2012) *Arabidopsis* Class I and Class II TCP Transcription Factors Regulate Jasmonic Acid Metabolism and Leaf Development Antagonistically. *Plant Physiology*, **159**, 1511-1523. <http://dx.doi.org/10.1104/pp.112.200303>
- [48] Balsemão-Pires, E., Andrade, L.R. and Sachetto-Martins, G. (2013) Functional Study of TCP23 in *Arabidopsis thaliana* during Plant Development. *Plant Physiology and Biochemistry*, **67**, 120-125. <http://dx.doi.org/10.1016/j.plaphy.2013.03.009>
- [49] Liu, Z.W., Biyashev, R. and Saghai-Marouf, M. (1996) Development of Simple Sequence Repeat DNA Markers and Their Integration into a Barley Linkage Map. *Theoretical and Applied Genetics*, **93**, 869-876. <http://dx.doi.org/10.1007/BF00224088>
- [50] Ivandic, V., Hackett, C.A., Nevo, E., Keith, R., Thomas, W.T.B. and Forster, B.P. (2002) Analysis of Simple Sequence Repeats (SSRs) in Wild Barley from the Fertile Crescent: Associations with Ecology, Geography and Flowering Time. *Plant Molecular Biology*, **48**, 511-527. <http://dx.doi.org/10.1023/A:1014875800036>