Genetics and Molecular Biology, 28, 3 (suppl), 582-588 (2005)

www.sbg.org.br

Research Article

In silico characterization of microsatellites in *Eucalyptus* spp.: Abundance, length variation and transposon associations

Edenilson Rabello, Adriane Nunes de Souza, Daniel Saito and Siu Mui Tsai

Universidade de São Paulo, Centro de Energia Nuclear na Agricultura, Laboratório de Biologia Celular e Molecular, Piracicaba, SP, Brazil.

Abstract

This study assessed the abundance of microsatellites, or simple sequence repeats (SSR), in 19 *Eucalyptus* EST libraries from FORESTs, containing cDNA sequences from five species: *E. grandis*, *E. globulus*, *E. saligna*, *E. urophylla* and *E. camaldulensis*. Overall, a total of 11,534 SSRs and 8,447 SSR-containing sequences (25.5% of total ESTs) were identified, with an average of 1 SSR/2.5 kb when considering all motifs and 1 SSR/3.1 kb when mononucleotides were not included. Dimeric repeats were the most abundant (41.03%), followed by trimerics (36.11%) and monomerics (19.59%). The most frequent motifs were A/T (87.24%) for monomerics, AG/CT (94.44%) for dimerics, CCG/CGG (37.87%) for trimerics, AAGG/CCTT (18.75%) for tetramerics, AGAGG/CCTCT (14.04%) for pentamerics and ACGGCG/CGCCGT (6.30%) for hexamerics. According to sequence length, Class II or potentially variable markers were the most commonly found, followed by Class III. Two sequences presented high similarity to previously published *Eucalyptus* sequences from the NCBI database, EMBRA_72 and EMBRA_122. Local blastn search for transposons did not reveal the presence of any transposable elements with a cut-off value of 10^{-50}. The large number of microsatellites identified will contribute to the refinement of marker-assisted mapping and to the discovery of novel markers for virtually all genes of economic interest.

Key words: Eucalyptus, EST, microsatellite, simple sequence repeat (SSR), molecular marker.

Accepted: May 28, 2004; Received: May 30, 2005.

Introduction

Trees represent the majority of terrestrial biomass production and the main resource for forestry and wood-processing industries worldwide. Increases in wood productivity and quality have stimulated forest management research and technological advances in timber, pulp and paper, with little contribution from biotechnology. Forest genomics began when expressed sequence tag (EST) projects were initiated in pine (Allona *et al*. 1998) and poplar (Sterky *et al*. 1998), demonstrating the usefulness of EST sequencing, which was later proven to be a cheap and efficient method for finding genes (Bhalerao *et al*. 2003).

E*ucalyptus* is extensively grown in commercial forest plantations in Brazil, mostly established through vegetative propagation based on rooted cuttings (Campinhos and Ikemori 1980). The genetic mapping of species from this genus has been achieved by the use of a variety of marker types, including RAPD (Gan *et al*. 2003), RFLP (Byrne *et al*. 1995), AFLP (Marques *et al*. 1998), isozymes (Byrne *et al*. 1995) and microsatellites (Brondani *et al*. 2002). Also, some QTLs have been located on the genetic maps, which are generally involved in traits of economic interests: vegetative propagation ability (Grattapaglia *et al*. 1995); seedling height and leaf area (Byrne *et al*. 1997a); frost tolerance (Byrne *et al*. 1997b); growth and wood quality (Grattapaglia *et al*. 1996); and monoterpene composition (Sepherd *et al*. 1999).

Parental or individual clone identification by molecular methods has become increasingly important for genetic characterization of *Eucalyptus* spp. Under this new context, the method of choice must allow the design of consistent primer sets for clonal, as well as paternal and maternal identities. Historically, the use of hypervariable probes, designed as a variable number of tandem repeats (VNTRs, Nakamura *et al*. 1987) or minisatellites (Jeffreys *et al*. 1985), used simultaneously to detect multiple *loci* have represented an important step towards higher standards of reliability and reproducibility. Heterozygosities of some minisatellite *loci* can reach values as high as 0.99 (Jeffreys *et al*. 1988). However, it was soon realized that most of these hypervariable *loci* were clustered at proterminal regions (Royle *et al*. 1988) and thus less useful in genetic

Send correspondence to Siu Mui Tsai. Universidade de São Paulo, Centro de Energia Nuclear na Agricultura, Laboratório de Biologia Celular e Molecular, Av. Centenário 303, 13.416-000, Piracicaba, SP, Brazil. Email: tsai@cena.usp.br.

mapping for general purposes. Soon after these findings, a new class of polymorphic markers, named microsatellites (Litt and Luty 1989) or simple sequence repeats - SSRs (Tautz 1989) was described. This type of DNA polymorphism could be detected only after PCR amplification of DNA and separation on polyacrylamide gel electrophoresis. All simple sequence repeats with a repeat length of a few base pairs could be considered microsatellites (Wu and Tanksley 1993). In recent years, the use of SSR markers has become the method of choice for applications in forestry industries, because it is a fast and simple technique when compared to AFLPs, RFLPs or isozymes.

Given the interest of the plant genetics community in SSRs as genetic markers, there has been a particular concern in the establishment of methods for rapid identification of robust and informative SSRs linked to genes of agronomic significance. Compared to genome-wide isolation approaches, gene-targeted strategies are more likely to yield SSRs that are relevant to the goals of marker-assisted selection and germplasm assessment. In the former approach, linkage disequilibrium between an SSR and a gene is fortuitous and frequently insufficient for transfer to other germplasm of interest (Cardle *et al*. 2000). For *Eucalyptus* fingerprinting, by using an inter-simple sequence repeat (ISSR) PCR-based enrichment technique for microsatellite-rich regions, primer sets were constructed to amplify mono, di, tri, hexa and nonanucleotide repeats, which were also able to amplify the corresponding microsatellite *loci* from five different *Eucalyptus* spp.: *E. grandis*, *E. nitens*, *E. globulus*, *E. camaldulensis* and *E. urophylla* (Van der Nest *et al*. 2000).

In the search for transposable elements (TEs), two major groups are expected - RNA mediated transposable elements or retroelements, and DNA transposable elements or classical transposons. They are mutagenic agents and their activity in the plant genome may provide high levels of variability, which may be used for genetic fingerprinting, to create novel genes and to modify genetic functions (Bennetzen 2000). Rossi *et al.* (2001) surveyed the TEs from the sugarcane expressed sequence tag (SUCEST) project containing 260,781 sequences and found 276 clones showing homology to previous reported TEs using a stringent cut-off value of e^{-50} or better. More recently, data obtained by Marques *et al*. (2002) and Kirst *et al*. (2005) demonstrated the feasibility of using SSRs for genetic analysis of several commercial *Eucalyptus* species.

This study assessed the abundance of SSRs in the *Eucalyptus* EST-based libraries, by using the recent submission of a large volume of cDNA sequences emerging from the *Eucalyptus* Genome Sequencing Project Consortium (FORESTs) which allowed the estimation of SSR frequency, repeat unit size and classification into three different groups: Class I ≥ 20 pb, Class II = between 11-20 pb, and Class III ≤ 11 pb. Using a local blastn algorithm (BLAST 2.0 - http://www.ncbi.nlm.nih.gov/blast), dispersed repetitive elements were surveyed at the flanking sites of the SSRs and their occurrence evaluated within the *Eucalyptus* EST libraries.

Material and Methods

Sequence data sources

Data were mined from FORESTs - *Eucalyptus* Genome Sequencing Project Consortium, supported by FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo - Brazil) - which contains cDNA sequences from five species of *Eucalyptus*: *E. grandis*, *E. globulus*, *E. saligna*, *E. urophylla* and *E. camaldulensis*. Sequences were obtained from 19 libraries of different plant tissues at different growth stages, under various physiological and stressed conditions (frost, drought, attack of fungal pathogens and insects, boron and phosphorus deficiencies, light/ dark growth). In this study, 17,286 singleton and 15,794 consensus sequences, for a total of 33,080 non-redundant ESTs, were screened for microsatellites or simple sequence repeats (SSRs). Singletons containing more than 550 bp were cut at their 3' end prior to SSR mining, in order to avoid analysis of low-quality bases.

Mining FORESTs database for SSR identification

Mono, di, tri, tetra, penta and hexanucleotide microsatellites were evaluated for their abundance and length distribution. Different SSR motifs were surveyed within the FORESTs database where complementary sequences were considered as belonging to the same class (*e.g.*, AC, CA, TG, GT). The identified SSRs were categorized into three groups based on the length of the repeat units (Class I ≥ 20 bp, Class II = between 11-20 bp, and Class III ≤ 11 bp) (Temnykh *et al*. 2001). Dispersed repetitive elements were surveyed at the flanking sites of the SSRs.

The query for SSR was supported by Perl script search module *MISA* (http://pgrc.ipk-gatersleben.de/misa), allowing the identification of perfect and compound microsatellites (Varshney *et al*. 2002). Perfect microsatellites were defined as sequences of ten or more mononucleotide repeats, six or more dinucleotide repeats, five or more tri, tetra, penta and hexanucleotide repeats. Compound microsatellites were considered as those present in the same EST and distant by a maximum of 100 bp. A_n repeats distant by a maximum of 50 bp from the 3' end of sequences were not considered as microsatellites, as they may represent poly-A tails of eucaryotic mRNA. Since the cloning procedure was vector-oriented, there was no need to eliminate poly-T tails from our analyses.

Additional analysis was performed in order to investigate possible matches among the EST-derived SSR sequences identified herein to those from genomic databases. Seventy SSR markers from *E. grandis* and E. *urophylla* (Brondani *et al*. 2002, Brondani *et al*. 1998), 8 from *E.*

sibieri (Glaubitz *et al*. 2001), 8 from *E. nitens* (Byrne *et al*. 1996, http://www.ffp.csiro.au/tigr/molecular/eucmsps.html), 26 from *E. globulus* (http://www.ffp.csiro.au/tigr/molecular/eucmsps.html) and 24 from the NCBI database (http://www.ncbi.nlm.nih.gov) were cross-matched against our results, with a local blastn algorithm.

Searching for transposable elements associated to SSRs

Initially, a possible association between *Eucalyptus* SSRs and dispersed repetitive elements was searched by BLAST analysis, where sequences flanking the SSR motifs were used as queries. Due to the strategy for the *Eucalyptus* genome construction (FORESTs), there was no need for setting simple Perl scripts for semiautomated identification of nonreduntant SSR *loci* (Temnykh *et al*. 2001). TIGR v.2 and REPBASE 8.9 public databases, which gather transposable elements (TEs) sequence data from diverse organisms, were utilized as blastn local databases. Only SSR-containing sequences were used as queries. Positive identification of transposable elements was performed with a maximum expectation value of 10^{-50} to avoid spurious matches (Rossi *et al*. 2001).

Results

Microsatellite frequency, distribution and transposon association

A total of 33,080 EST data representing 29,058,996 bp from the *Eucalyptus* Genome Sequencing Project Consortium (FORESTs) were mined for microsatellites. SSRs were analyzed for abundance, length variation, distribution and transposon associations. In all, 11,534 SSRs and 8,447 SSR-containing sequences (25.5% of total ESTs) were identified, with an average of 1 SSR/ 2.5 kb (or 1 SSR/ 3.1 kb when mononucleotides were not considered) (Table 1). In cereals, including barley, maize, oat, rice, rye and wheat, lower frequencies of SSRs (7-10% of total ESTs) were found from their available genome database (Varshney *et al*. 2002).

The most frequently found motifs were: A/T (87.24%) for monomerics, AG/CT (94.44%) for dimerics, CCG/CGG (37.87%) for trimerics, AAGG/CCTT (18.75%) for tetramerics, AGAGG/CCTCT (14.04%) for pentamerics and ACGGCG/CGCCGT (6.30%) for hexamerics (Figure 1). According to sequence length, Class II or potentially variable markers were the most common (42.36%), followed by Class III (32.84%) (Figure 2). Dimeric repeats were the most abundant (41.03%), followed by trimerics (36.11%) and monomerics (19.59%). The SSRs contained virtually no pentanucleotide repeats (0.49%) (Figure 3). Figure 4 shows the number of SSRs according to the number of repeat units. The number of SSRs

Table 1 - General statistics in the search for SSRs in *Eucalyptus* ESTs from the FORESTs Project (FAPESP-ONSA).

Total number of sequences analyzed	33,080
Total size of examined sequences (bp)	29,058,996
Number of identified SSRs	11,534
Number of SSR-containing sequences	8,447
Number of sequences containing more than one SSR	2,500
Number of compound SSRs[1]	1,823

[1]In the same EST, less than 100 bp apart.

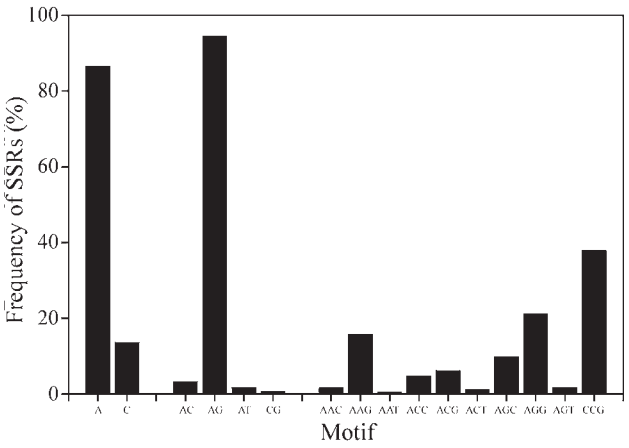

Figure 1 - Frequency of SSRs according to motifs, mined from the FORESTs database (11,534 SSRs).

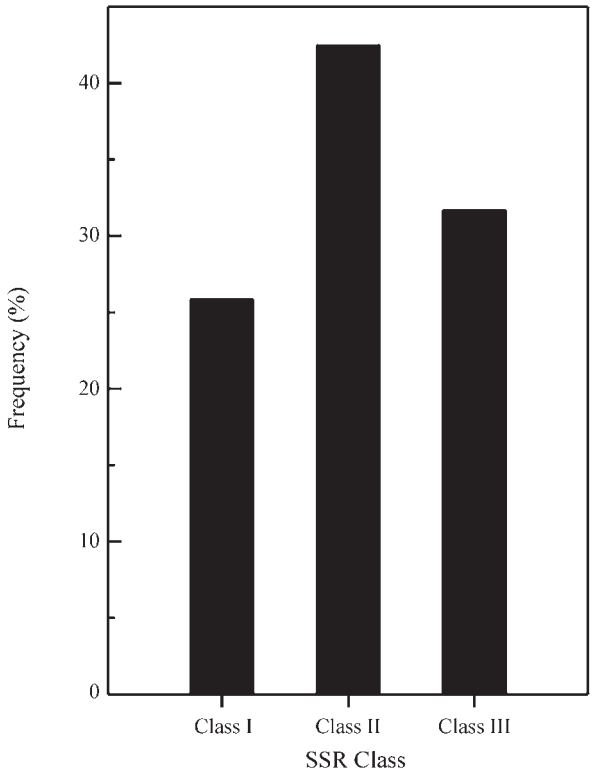

Figure 2 - Distribution of SSRs into classes in *Eucalyptus* spp. from the FORESTs database (11,534 SSRs).

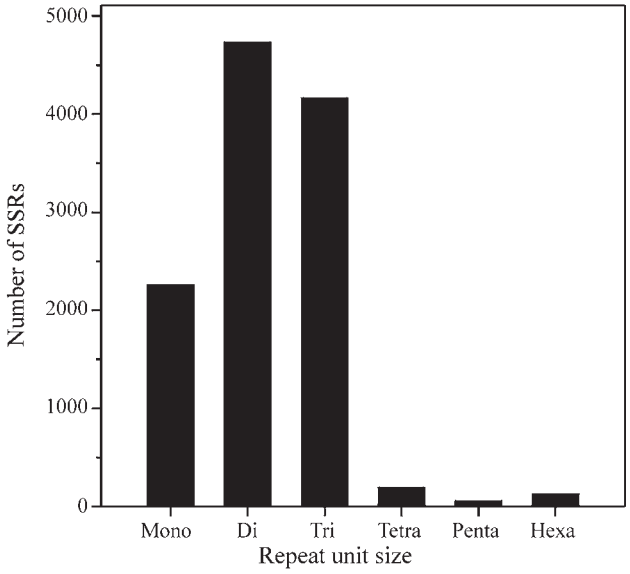

Figure 3 - Number of SSRs according to repeat unit size (motif length) in *Eucalyptus* spp. Di, tri and monomeric repeats are the most frequently found in the FORESTs database.

in each motif length decreases with the increase in number of repeat units, except for mono and dinucleotides.

The cross-matching analysis of the identified SSRs with the published genomic-derived *Eucalyptus* sequences retrieved only two highly similar hits: EMBRA_72 (Expect: 10^{-66}, Identities: 97%) and EMBRA_122 (Expect: 10^{-66}, Identities: 88%).

Transposable elements associations

Local BLAST search for transposons against TIGR v.2 and REPBASE 8.9 did not reveal the presence of any transposable element with a cut-off value of 10^{-50}. Only nine SSR-containing sequences were associated with 45S rDNA-like sequences, with identities ≥ 91% and expect values ≤ 10^{-87} (Table 2).

Discussion

Over the last decade, the ubiquity of SSRs in eukaryotic genomes and their usefulness as genetic markers has been well established. Microsatellites are simple, tandemly repeated mono to hexanucleotide sequence motifs flanked by unique sequences. They are valuable as genetic markers because they are codominant, detect high levels of allelic diversity, and are easily and economically assayed by PCR. High levels of SSR informativeness have been demonstrated for a variety of plant species and have prompted the initiation of SSR discovery programs for most important crops. Nonetheless, researchers have encountered a number of limitations, such as lack of DNA sequences in the available databases, a perceived low abundance of SSRs (when compared to mammals) and dif-

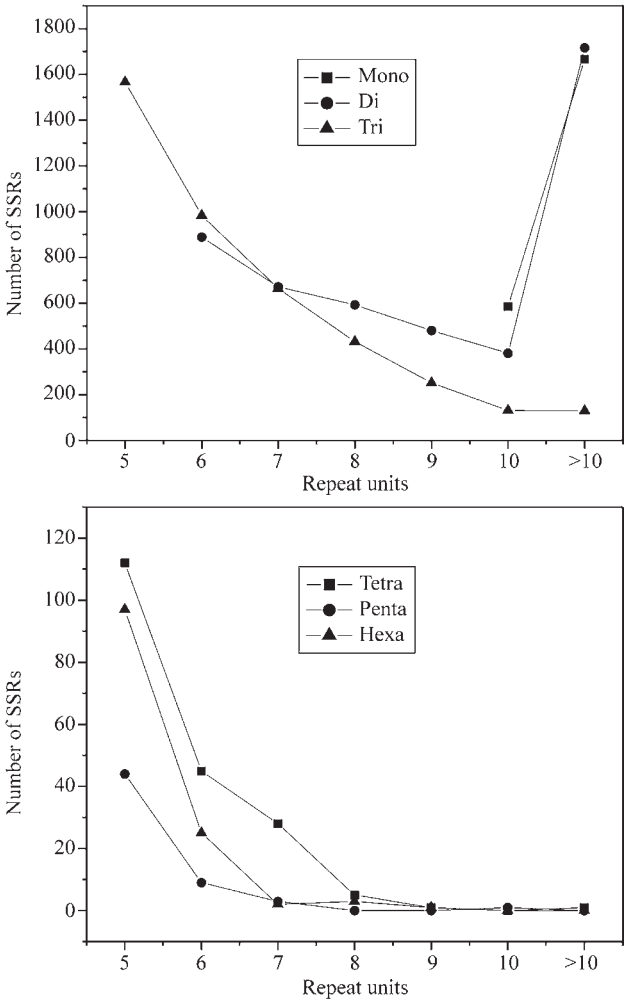

Figure 4 - Number of SSRs according to number of repeat units. The number of microsatellites in a particular repeat unit size decreases with the increasing number of repeat units, except for mono and dimerics.

Table 2 - Repetitive elements associated to microsatellites (SSR) determined in *Eucalyptus* ESTs from FORESTs.

SSR containing sequences	Best hit	Identities (%)	Expect value
EGEQLV2201H05.g	ORSgRGRR00000011 45S rDNA-like	92	0.0
EGUTRT3109D12.g	ORSgRGRR00000011 45S rDNA-like	92	0.0
EGEQLV1201D10.g	ORSgRGRR00000001 45S rDNA-like	94	e-123
EGABST6055D11.g	ORSgRGRR00000001 45S rDNA-like	96	e-136
EGCBLV2219H11.g	ORSgRGRR00000001 45S rDNA-like	98	0.0
EGCBSL5005G05.g	ORSgRGRR00000001 45S rDNA-like	94	3e-90
EGCBSL4038C07.g	ORSgRGRR00000011 45S rDNA-like	91	0.0
EGEZLV2203H01.g	ORSgRGRR00000010 45S rDNA-like	96	0.0
EGMCSL5073G06.g	ORSgRGRR00000001 45S rDNA-like	95	0.0

ferences in the most common types of repeats found (Cardle *et al*. 2000).

Even though plant SSRs can be about 10 times less frequent than those found in humans, the screening of large numbers of clones and the development of selective SSR enrichment techniques have proven to be advantageous techniques for plant geneticists (Cardle *et al*. 2000). Results from screening a rice genomic library suggest that there are about 5,700-10,000 microsatellites, with the relative frequency of different repeats decreasing with increasing size of the motif (McCouch *et al*. 1997). Our data have shown a high number of SSRs - 11,534 out of a total of 33,080 FOREST data representing 29,058,996 bp, as well as 8,447 SSR-containing sequences (25.5% of total ESTs), with an average of 1 SSR/2.5 kb or 1 SSR/3.1 kb (excluding mononucleotides), which is about four times (1 SSR/14 kb) that found for Arabidopsis (Cardle *et al*. 2000) and about twice (1 SSR/ 6.0 kb) that found for cereals (Varshney *et al.* 2002).

Motif A/T was found to be more abundant than C/G in exons in all the taxa studied by Tóth *et al*. (2000), which is in agreement with our data. Moreover, the high percentage of the AG/CT motif is in accordance to a previous study conducted in SSR-enriched genome libraries from two *Eucalyptus* species - *E. urophylla* and *E. grandis* (Brondani *et al*. 1998) and in cereal species ESTs (Varshney *et al*. 2002). Among the trimerics, motif CCG/CGG was the most abundant, a result also obtained by Varshney *et al*. (2002). Moreover, 79.50% of trinucleotides were represented by GC-rich motifs (containing ≥ 2G and/or C), suggesting that they may be associated with genes (Temnyhk *et al*. 2001). Along with CCG/CGG trinucleotides, GGA/TTC, CCT/AGG, GAA/TTC and CCG/GGC can form hairpin-like structures, which may stabilize them and allow them to escape from repair mechanisms (Tóth *et al*. 2000, Li *et al*. 2002). They are, therefore, expected to be more frequent. In fact they represent 79.30% the SSRs found. As for tetra and hexanucleotide repeats, there was a noticeable proximity among the frequencies of the first and second most abundant motifs (data not shown). When repeat unit sizes were analyzed, dinucleotides were the most abundant, a result that agrees with Cardle *et al*. (2000) in a study on *Arabdopsis*, but differs from that of Varshney *et al*. (2002), who found trinucleotides as the most frequent in cereals, followed by dinucleotides.

It remains unknown why certain repeat motifs are more common than others, or the reason they vary so much among or even within taxa. For example, the fungi species *P. chrysosporium* and *U. maydis* have A_n frequencies of 35 and 70%, respectively (Lim *et al*. 2004). Furthermore, SSR motifs, abundance, and mutation rates are different among species, with a wide range of genetic properties (Cruz *et al*. 2005).

The division of microsatellites into classes represents their potential as molecular markers. Class I repeats are highly polymorphic, class II are less variable, and class has a mutation potential similar to most unique sequences (Temnyhk *et al*. 2001). Class II represented 42.36% of all SSRs found and it is the most common within the repeat unit sizes in which it appears (mono to tetranucleotides). Although class I represented only about one fourth of all microssatellites, they should be the starting point for the design of molecular markers as they are the most polymorphic.

Two different patterns were observed when comparing the number of motif lengths to the number of repeat units. While there are well-defined decaying curves for tri to hexameric motifs, this tendency was not observed for mono and dimerics, which is in agreement with the results of Varshney *et al*. (2002).

Only two SSR sequences (EMBRA_72 and EMBRA_122) were identified by searching the available *Eucalyptus* genomic-derived SSR databases. This is probably due to the fact that microsatellites from these databases may be located in noncoding regions, or that these databases are still reduced.

In contrast to a similar study conducted in sugarcane (Rossi *et al*. 2001), we could not detect any relationships between the SSR-containing sequences and TEs at a cut off value of 10^{-50}. This bias may be due to differences in the total number of ESTs analyzed, which was almost 10 times lower in the present investigation. Also, we used only SSR-containing sequences in our analysis, which may also have contributed to a lower SSR-TE correlation rate.

In a recent review based on computational and experimental characterization of physically clustered SSRs in plants, the type and frequency of SSRs in plant genomes were investigated using the expanding quantity of DNA sequence data deposited in public databases (Cardle *et al*. 2000). For example, 306 genomic DNA sequences longer than 10 kb and 36,199 EST sequences were searched in *Arabidopsis* for all possible mono- to pentanucleotide repeats, with an average of 1 SSR for every 6.04 kb in the genomic DNA, decreasing to one every 14 kb in ESTs. Similar frequencies were also found in other plant species, although higher SSR frequencies associated to *Eucalyptus* ESTs were observed in the present study, when compared to different cereal or naturally-occurring tree species. On the basis of these findings and the previous data from other authors, we can conclude that there is a good potential for using the present approach for the targeted isolation of single or multiple, physically clustered SSRs linked to any *Eucalyptus* gene that has been mapped using DNA-based markers. Further mining within the available databases will be needed if unique primer pairs for *Eucalyptus* spp. are requested for genetic discrimination.

Acknowlegements

Data from this work were mined from FOREST database, supported by FAPESP-ONSA. The group collabo-

rated on the FOREST sequencing genome from AEG program - Agricultural and Environmental Genomes (Proc. 00/10168-6). We gratefully acknowledge CNPq/CAPES for fellowships granted to all authors of this study.

References

Allona I, Quinn M, Shoop E, Swope K, St. Cyr S, Carlis J, Riedl J, Retzel E, Campbell MM, Sederoff R and Whetten RW (1998) Analysis of xylem formation in pine by cDNA sequencing. Proc Natl Acad Sci USA 95:9693-9698.

Bhalerao R, Nilsson O and Sandberg G (2003) Out of the woods: Forest biotechnology enters the genomic era. Curr Opin Biotechnol 14:206-213.

Bennetzen JL (2000) Transposable element contributions to plant gene and genome evolution. Plant Mol Biol 42:251-69.

Brondani RPV, Brondani C and Grattapaglia D (2002) Towards a genus-wide reference linkage map for *Eucalyptus* based exclusively on highly informative microsatellite markers. Mol Genet Genomics 267:338-347.

Brondani RPV, Brondani C, Tarchini R and Grattapaglia D (1998) Development, characterization and mapping of microsattelite markers in *Eucalyptus grandis* and *E. urophylla*. Theor Appl Genet 97:816-827.

Byrne M, Marques-Garcia MI, Uren T, Smith DS and Moran GF (1996) Conservation and genetic diversity of microsatellite *loci* in the genus *Eucalyptus*. Aust J Bot 44:331-341.

Byrne M, Murrell JC, Allen B and Moran GF (1995) An integrated genetic linkage map for eucalyptus using RFLP, RAPD and isozyme markers. Theor Appl Genet 91:869-875.

Byrne M, Murrell JC, Owen JV, Kriedemann P, Williams ER and Moran GF (1997a) Identification and mode of action of quantitative trait *loci* affecting seedling height and leaf area in *Eucalyptus nitens*. Theor Appl Genet 94:674-681.

Byrne M, Murrell JC, Owen JV, Williams ER and Moran GF (1997b) Mapping of quantitative trait *loci* influencing frost tolerance in *Eucalyptus nitens*. Theor Appl Genet 95:975-979.

Campinhos E and Ikemori Y (1980) Mass production of *Eucalyptus* spp. by rooting cuttings. *In*: IUFRO Symp. Genet. Improvement and Productivity of Fast-Growing Trees, São Paulo, Brazil, pp 60-67.

Cardle L, Ramsay L and Milbourne D (2000) Computational and experimental characterization of physically clustered simple sequence repeats in plants. Genetics 156:847-854.

Cruz F, Pérez M and Presa P (2005) Distribution and abundance of microsatellites in the genome of bivalves. Gene 346:241-247.

Gan S, Shi J, Li M, Wu K, Wu J and Bai J (2003) Moderate-density molecular maps of *Eucalyptus urophylla* S.T. Blake and *E. tereticornis* Smith genomes based on RAPD markers. Genetica 118:59-67.

Glaubitz JC, Emebiri LC and Moran GF (2001) Dinucleotide microsatellites from *Eucalyptus sieberi*: Inheritance, diversity and improved scoring of single-base differences. Genome 44:1041-1045

Grattapaglia D, Bertolucci FLG, Penchel R and Sederoff RR (1996) Genetic mapping of quantitative trait *loci* controlling growth and wood quality traits in Eucalyptus grandis using a maternal half-sib family and RAPD markers. Genetics 144:1205-1214.

Grattapaglia D, Bertolucci FL and Sederoff RR (1995) Genetic mapping of QTLs controlling vegetative propagation in *Eucalyptus grandis* and *E. urophylla* using a pseudo-testcross strategy and RAPD markers. Theor Appl Genet 90:933-947.

Jeffreys AJ, Royle NJ, Wilson V and Wong Z (1988) Spontaneous mutation rates to new length alleles at tandem-repetitive hypervariable *loci* in human DNA. Nature 332:278-281.

Jeffreys AJ, Wilson V and Thein SL (1985) Hypervariable "minisatellite" regions in human DNA. Nature 314:67-73.

Kirst M, Cordeiro CM, Rezende GDSP and Grattapaglia D (2005) Power of microsatellite markers for fingerprinting and parentage analysis in *Eucalyptus grandis* breeding populations. J Hered 96:161-166.

Li Y-C, Korol AB, Fahima T, Beiles A and Nevo E (2002) Microsatellites: Genomic distribution, putative functions and mutational mechanisms: A review. Mol Ecol 11:2453-2465.

Lim S, Notley-McRobb L, Lim M and Carter DA (2004) A comparison of the nature and abundance of microsatellites in 14 fungal genomes. Fungal Genet Biol 41:1025-1036.

Litt M and Luty JA (1989) A hypervariable microsatellite revealed by *in vitro* amplification of a dinucleotide repeat within the cardiac muscle actin gene. Am J Hum Genet 44:397-401.

Marques CM, Araújo JA, Ferreira JG, Whetten R, O'Malley DM, LI B-H and Sederoff R (1998) AFLP genetic maps of *Eucalyptus globulus* and *E. terticornis*. Theoret Appl Genet 96:727-737.

Marques CM, Brondani RPV, Grattapaglia D and Sederoff R (2002) Conservation and synteny of SSR *loci* and QTLs for vegetative propagation in four *Eucalyptus* species. J Hered 105:474-478.

McCouch SR, Chen X, Panaud O, Temnykh S, Xu Y, Cho YG, Huang N, Ishii T and Blair M (1997) Microsatellite marker development, mapping and applications in rice genetics and breeding. Plant Mol Biol 35:89-99.

Nakamura Y, Leppert M, O'Connell P, Wolff R, Holm T, Culver M, Martin C, Fujimoto E, Hoff M and Kumlin E (1987) Variable number of tandem repeats (VNRT) markers for human genome mapping. Science 235:1616-1622.

Rossi M, Araujo PG and Van Sluys MA (2001) Survey of transposable elements in sugarcane expressed sequence tags (ESTs). Genet Mol Biol 24:147-154.

Royle NJ, Clarkson RE, Wong Z and Jeffreys AJ (1988) Clustering of hypervariable minisatellites in the proterminal regions of human autosomes. Genomics 3:352-360.

Sepherd M, Chaparro JX and Teasdale R (1999) Genetic mapping of monoterpene composition in an interspecific eucalypt hybrid. Theor Appl Genet 99:1207-1215.

Sterky F, Regan S, Karlsson J, Hertzberg M, Rohde A, Holmberg A, Amini B, Bhalerao R, Larsson M, Villarroel R, Van Montagu M, Sandberg G, Olsson O, Teeri TT, Boerjan W, Gustafsson P, Uhlen M, Sundberg B and Lundeberg J (1998) Gene discovery in the wood-forming tissues of poplar: Analysis of 5,692 expressed sequence tags. Proc Natl Acad Sci USA 95:13330-5.

Tautz D (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucl Acids Res 17:6463-6471.

Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinhour S and McCouch S (2001) Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): Frequency, length variation, transposon associations, and genetic marker potential. Genome Res 11:1441-1452.

Tóth G, Gáspari Z and Jurka J (2000) Microsatellites in different eukaryotic genomes: Survey and analysis. Genome Res 10:967-981.

Van der Nest MA, Steenkamp ET, Wingfield BD and Wingfield MJ (2000) Development of simple sequence repeat (SSR) markers in *Eucalyptus* from amplified inter-simple sequence repeats (ISSR). Pl Breeding 119:433-436.

Varshney RK, Thiel T, Stein N, Langridge P and Graner A (2002) *In silico* analysis on frequency and distribution of microsatellites in ESTs of some cereal species. Cell Mol Biol Lett 7:537-546.

Wu K-S and Tanksley SD (1993) Abundance, polymorphism and genetic mapping of microsatellites in rice. Mol Gen Genet 241:225-235.

Associate Editor: Marie Anne Van Sluys