DE GRUYTER OPEN

Acta Bot. Croat. 76 (1), 103–106, 2017
DOI: 10.1515/botcro-2016-0050

CODEN: ABCRA 25
ISSN 0365-0588
eISSN 1847-8476

Short communication

Microsatellite allele length variations in inter-specific hybrids of *Eucalyptus*

Murugan Sumathi, Yasodha Ramasamy*

Institute of Forest Genetics and Tree Breeding Coimbatore 641 002 Tamilnadu, India

Abstract – The genus *Eucalyptus* encompasses several species of industrial importance. Many of these species have been subjected to genetic characterization using different kinds of DNA markers. More than 1000 microsatellites have been identified from the genome of eucalypts and they are highly amenable for cross species transferability. During cross amplification of microsatellites, homoplasy is reported in many species in which although the allele size might be the same, the sequences are not. Thus, it is essential to ascertain the DNA sequence homology with source and target microsatellite repeats. Accordingly, fifty five alleles from six microsatellite loci (ECc1, ECc2, Eg61, Embra100, Embra1468 and Embra2002) were amplified in two inter-specific hybrid populations (*Eucalyptus tereticornis* × *E. grandis* and *E. tereticornis* × *E. camaldulensis*) and sequenced. The results showed that all the microsatellite loci were amplifying the target repeat types except for the loci Eg61 and Embra2002. The locus Eg61 has target repeat of (CAA)(GAT) but the sampled alleles had either (CAA)(GAT) or (GAT) alone. Similarly, the Embra2002 locus was targeting interrupting repeats of (CCA)..(CCA), but the sequenced alleles had repeats of (CCA) with or without interruption. Nevertheless, the allele size estimated in electrophoresis for hybrids was in conformity with that of the parent alleles. This study suggests the need for validation of the repeat characteristics of microsatellites by sequencing of the alleles particularly in cross species amplification.

Keywords: allele sequences, *Eucalyptus*, inter-specific hybrids, microsatellites, repeat size

Introduction

The species of the genus *Eucalyptus* are widely planted in tropical and temperate regions of the world and are an important commercial tree species for paper pulp, bioenergy and timber production. Several members of *Eucalyptus* have been genetically characterized because of the ease of generating inter-specific crosses and vegetative propagation, its unique biological properties, fast growing nature and comparatively small genome size. Various DNA markers have been applied for the genetic characterization of the eucalyptus genome. Microsatellites or simple sequence repeats (SSRs) are the DNA marker system of choice, for it is highly conserved and easily transferrable across species. Many of these SSRs were shown to be potentially involved in growth, wood quality and stress responsive pathways (Gion et al. 2015). SSR based genetic linkage map has a significant place in eucalyptus genetic improvement because the high cross species transferability led to the integration of linkage maps across different genetic backgrounds. Genetic maps in eucalypts were used in mapping quantitative trait loci (QTL), synteny and collinearity analysis and genome sequence assembly (Hudson et al. 2012, Bartholome et al. 2015).

Recently, Grattapagalia et al. (2014) showed that 84% of the mapped microsatellites had colinearity between linkage map and physical position in the assembled *E. grandis* genome. Although to date about 1200 SSRs have been characterized in different species of eucalypts, only about 535 SSRs have been utilized for genetic mapping studies. Hence, to generate high density genetic map in new combinations of inter-specific hybrids requires the addition of a higher number of SSR loci. So far, the majority of the SSRs mapped (both genomic and expressed sequence tag-derived SSRs) were developed from *E. grandis* and *E. urophylla* and have been transferred to other species. During cross transfer, it is essential to verify the allele variations by sequencing the PCR fragments, wherein the structural variants in the form type of repeats, insertion or deletion generated by the SSRs are accurately detected. In eucalypts, 240 EST-SSRs were confirmed for their sequence homology with the original EST sequences, however, no information on repeat size variations were reported (Zhou et al. 2014). Recently, it was emphasized that the sequence verification

* Corresponding author, e-mail: yasodharaja@yahoo.com

of SSR alleles is critical for interpretation of the SSR polymorphism for genetic analysis (Barthe et al. 2012). Further, size homoplasy is commonly reported among the SSR alleles (Curtu et al. 2004, Javed et al. 2014), hence, information on DNA sequence is essential before the utilization of SSR loci variations. Thus, in the present study, attempts were made to sequence the SSR-PCR amplicons generated in the inter-specific hybrid populations, *E. tereticornis* × *E. camaldulensis* and *E. tereticornis* × *E. grandis*.

Materials and methods

Plant material and DNA isolation

The two inter-specific hybrid populations used in this study were *E. tereticornis* (Et86) × *E. grandis* (Eg9) and *E. tereticornis* (Et217) × *E. camaldulensis* (Ec17). These species belong to the subgenus (*Symphyomyrtus*) and the section Exsertaria (*E. tereticornis* and *E. camaldulensis*) and Latoangulatae (*E. grandis*). The number indicated for each species represents the parent clone identity. The hybrid populations were generated by controlled pollination between selected parents. F1 seeds generated from controlled pollination were established in a field trial at Panampally, Kerala. Total genomic DNA was extracted using DNA isolation kit (Qiagen Ltd, UK) from the freshly collected leaves of two individual hybrid plants in each population. The DNA of parents was isolated from leaf samples of trees employed in controlled pollination. DNA quality and quantity was assessed by electrophoresis on 0.8% agarose gel with Lambda DNA (Bangalore Genei, Bangalore, India) as the standard and spectrophotometry (NanoDrop 8000; Thermo Scientific, Wilmington, DE, USA).

SSR amplification

Three hundred and twenty SSRs (Sumathi et al. 2014) were tested between the parents of two hybrid populations for the selection of polymorphic loci. For sequence analysis and repeat motif identification of the parents and hybrids, 6 loci (4 genomic SSRs – ECc1, ECc2, Eg61, EMBRA100, and 2 EST-SSRs -EMBRA1468, EMBRA2002) were randomly selected to represent different repeat motif types and source species such as *E. camaldulensis* (da Silva et al. 2009) and *E. globulus* (Thamarus et al. 2002) and *E. grandis* (Brondani et al. 2006, Faria et al. 2010). Details on allele size range, annealing temperature and repeat motif and NCBI accession number of each SSR loci are presented in Tab. 1.

SSR amplification was carried on 40 µl of reaction containing 4 µl of 10× buffer (100 mM Tris–HCl (pH 8.3), 500 mM KCl and 15 mM $MgCl_2$), 125 µM of dNTPs, 0.4 µM of each forward and reverse primer, 4 U of Taq DNA polymerase, and 20 ng of template DNA of 2 parents and 2 F1 individuals of the hybrid populations. The PCR amplification was carried out for initial denaturation 5 min at 94 °C for 1 cycle and 30 cycles of 1 min at 94 °C, 30 s at the primer specific annealing temperature (Tab. 1), 1 min at 72 °C, final extension for 15 min at 72 °C. PCR products were size separated using 5% denaturing polyacrylamide (PAGE) gels of size 21×50 cm (Sequi-Gen GT System, Bio-Rad, USA) containing 7 M urea and 1× Tris-borate-EDTA buffer. A DNA ladder of 50 bp (MBI, Fermentas, USA) was included during electrophoresis of samples in the first and last well of the electrophoresis system. The gels were run at 61 W constant power for 3 h and visualized by silver staining.

Sequencing of PCR fragments and sequence analysis

To isolate the specific bands of SSRs for sequencing, the PCR amplified products were separated on a 3.5% Metaphor agarose gel (Cambrex Bio Science, Rockland, ME, USA) in 1× TBE buffer and visualized with ethidium bromide staining. Only the best amplified bands for each SSR primer combination was removed from the gel and purified using QIAquick gel extraction kit (Qiagen, Crawley, UK). These DNA fragments were sequenced at Eurofins Biotechnologies Pvt Ltd, Bangalore, India using standard procedures on an ABI3700 DNA analyser (Applied Biosystems, USA) by employing the specific forward or reverse SSR primer.

The sequences obtained were checked for their homology with the SSR source sequence in NCBI and the genome sequence of *Eucalyptus grandis* in Phytozome (https://phytozome.jgi.doe.gov/pz/portal.html). All the DNA sequences of SSR locus along with the source SSR sequence and *E. grandis* genomic region harbouring SSR were aligned using Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/) to identify the structural variations in microsatellites.

Results and discussion

Microsatellites are well known to generate high polymorphism across species due to their mutational characteristics (Bhargava and Fuentes 2010). Cross transferability of these markers across genera and species enabled the use of microsatellites, where the DNA sequence is not available.

Tab. 1. Information on the six microsatellite loci used in this study.

Sample	SSR code	Repeat motif	Annealing temperature (ºC)			Allele size	Linkage group	Reference	NCBI Accession No.
			Original	Optimized Et86 × Eg9	Optimized Et217 × Ec17				
1	ECc1	(TC)18	60–45	56	58	157–192	1	da Silva et al. 2009	GQ302860.1
2	ECc2	(GA)9TATA(GA)6	60–46	57	57	238–255	5	da Silva et al. 2009	GQ302861.1
3	Eg61	(GAA)9(GAT)6	64	56	61	302–316	2//7	Thamarus et al. 2002	EU699745.1
4	EMBRA100	(CT)26	56	58	56	195	1	Brondani et al. 2006	BV682837.2
6	EMBRA1468	(TC)12	60	57	59	181	8	Faria et al. 2010	GF101903.1
5	EMBRA2002	(CCA)8…(CCA)8	60	56	64	263	1	Faria et al. 2010	GF101907.1

Tab. 2. Details on allele size and repeat type for the six microsatellite loci sequenced from the parents and their hybrids of eucalypt species (alleles in bold letters were sequenced). Asterisk (*) denotes sequence from which the SSRs are reported.

Sample	Plant identity	ECc1		ECc2		Eg61		Embra 100		Embra 1468		Embra 2002	
		Allele size	Repeat type	Allele size	Repeat type	Allele size	Repeat type	Allele size	Repeat type	Allele size	Repeat type	Allele size	Repeat type
1	*E. grandis*	–	(TC)18	–	(GA)18	–	(GAA)6 (GAT)8	–	(CT)26	–	(TC)23	–	(CCA)7…(CCA)8
2	Source sequence*	–	(TC)18	–	(GA)9TATA(GA)6	–	(GAA)9 (GAT)9	–	(CT)19	–	(TC)12	–	(CCA)8…(CCA)8
3	Et86	**247**/**297**	(TC)16/(TC)25	241/**247**	(GA)18	**322**/328	(GAT)6	253/**269**	(CT)25	206/**218**	(TC)20	**263**/272	(CCA)3
4	Eg9	**263**/**275**	(TC)17/(TC)19	**238**/238	(GA)16	325/**338**	(GAT)7	**238**/247	(CT)20	**218**/225	(TC)21	275/**278**	(CCA)6 (CCA)7
5	EtEgH1	**275**/**297**	(TC)25/(TC)19	**238**/241	(GA)17	322/**338**	(TAT)3	247/**269**	(CT)23	**218**/225	(TC)22	**263**/278	(CCA)4
6	EtEgH2	**247**/**275**	(TC)16/(TC)17	**238**/247	(GA)17	322/**338**	(AAT)4	**238**/269	(CT)18	**206**/218	(TC)18	**263**/275	(CCA)3
7	Et217	**263**/297	(TC)19	235/**250**	(GA)18	**325**/325	(GAA)6 (GAT)8	253/**269**	(CT)18	**225**/225	(TC)18	**263**/268	(CCA)4
8	Ec17	**263**/263	(TC)19	**238**/238	(GA)15	**313**/325	(GAT)5	**238**/244	(CT)29	206/**218**	(TC)13	**278**/278	(CCA)9
9	EtEcH1	**263**/297	(TC)19	**250**/238	(GA)19	**313**/325	(GAT)5	244/**253**	(CT)20	**225**/218	(TC)18	**263**/263	(CCA)4
10	EtEcH2	**263**/263	(TC)19	235/**238**	(GA)17	**313**/325	(GAT)5	238/**253**	(CT)22	225/**218**	(TC)13	**268**/268	(CCA)7

Hence, it is essential to confirm the repeat characteristics of SSRs to utilize them as markers for DNA fingerprinting, genetic linkage map generation and quantitative trait loci identification. In order to evaluate the variation in repeat region of SSR alleles, six SSR loci were sequenced in two different inter-specific hybrid populations of *Eucalyptus*. The lengths of the sequenced fragments varied between 206–338 bp.

Out of 87 bands (alleles), fifty five bands showing high intensity in the agarose gel were selected for sequencing and the DNA sequences were analyzed to estimate the degree of sequence divergence among equally sized alleles, repeat number variations and repeat composition. The information on allele size, repeat motif for the sequenced alleles is presented in Tab. 2. The SSR locus, ECc1 (GQ302860.1) was developed from *E. camaldulensis* targeting (TC) repeats which had amplified same type of repeats in all the samples. The parents Et86 and Eg9 generated four different alleles (247 and 297 bp; and 263 and 275 bp, respectively) for the ECc1 locus while their hybrids generated allelic combination of 275/297 bp (H1) and 247/263 bp (H2). Sequencing of all the four alleles showed repeat number variations from 16 to 25 (Tab. 2) and no direct relationship between repeat number and allele size could be established, because the allele size variation could be due to the variations in the flanking regions also. The parent Et217 amplified two different alleles of size 263 and 297 bp and the parent Ec17 amplified a homozygous allele of size 263 bp. Sequencing was done only for the 263 bp allele from all the four individuals and all the sequences showed same repeat number. The 263 bp allele sequence of parent Et217 showed 19 (TC) repeats followed by an insertion of one (GT) repeat and 6 (TC) repeats. Such insertions or interruptions within the motif sequence were documented in *Vitis* (Masi et al. 2004), *Citrus* (Barkley et al. 2009) and rice (Dong et al. 2013). Sequence data on microsatellite variation in birds indicated that size homoplasy exists with no sequence similarity (Anmarkrud et al. 2008). In tree species, Barthe et al. (2014) showed the high amount of variations in flanking regions. The similarity of repeat regions across the genome sequence of *E. grandis* in Phytozome, the SSR source sequence and the allele sequences for all the loci analyzed are given in On-line Suppl. Material.

The locus ECc2 produced heterozygous alleles in parents Et86 (241/247 bp) and Et217 (235/250 bp) and homozygous alleles in parents Eg9 and Ec17 (238 bp). In the original *E. camaldulensis* sequence of ECc2 (GQ302861.1), GA repeat was interrupted with TATA and was not observed in any of the sequences analyzed including the genome sequence of *E. grandis*. Instead, all the sequences had (GA) repeats in uninterrupted form and the number varied from 16–19. Similarity across the sequences was also very low except in the repeat region. The Eg61 SSR locus generated heterozygous alleles in both the parents (ET86 and Eg9). Original microsatellite sequence of Eg61 (EU699745.1) and *E. grandis* genome sequence showed compound SSR (GAA) (GAT) repeats. The parents Et86 and Eg9 showed only (GAT) repeats and their hybrid individuals showed (TAT) and (AAT) repeats and this could be attributed towards the error during PCR amplification or sequencing.

The Et217 parent had (GAA and (GAT) motifs and Ec17 and the hybrids showed only (GAT) motifs.

The SSR loci Embra100 (BV682837.2) was present in the *E. grandis* gene sequence *Protein Transparent Testa12* (Eucgr.A02047), which had 18 repeat units of the target motif (CT). This locus was heterozygous in both the crosses analyzed. Alleles from the individuals sequenced showed (CT) repeats varying from 18 to 25. The EST-SSR locus Embra1468 (GF101903.1) showed similarity with the gene sequence *Acetyl-CoA C-acetyltransferase / Acetoacetyl-CoA thiolase* (Eucgr.H00849). It had heterozygous alleles in Et86, Eg 9 and Ec17 whereas the Et217 showed homozygous alleles. The original sequence consisted of 12 repeat units of TC and all the alleles analyzed had 13–22 repeat units. The SSR locus Embra2002 (GF101907.1) sequence showed similarity with *Auxin responsive protein (Auxin_inducible)* of *E. grandis* (Eucgr.A01420). Except Ec17 all the parents had heterozygous alleles and the target repeat of (CCA) was present in varying repeat numbers in interrupted and uninterrupted form (Tab. 2).

This study demonstrated the successful cross species amplification of several eucalypt SSR loci within the genus. Microsatellite allele sequencing results of parents and their hybrids show almost same repeat type, corroborating the earlier results in eucalypt species (Ochieng et al. 2007, He et al. 2012). Variations in flanking regions and repeat numbers reflect the occurrence of mutational processes in both the repeat and flanking regions. Similar observations were made in *Vitis* (Masi et al. 2004) and *Shorea* (Javed et al. 2014) and insisting the requirement of information on underlying DNA sequence. However, with the availability of genome sequence of *E. grandis*, the primer sequence similarity search can be applied to find the primer binding regions. Loci with more than one similar site may be avoided to limit any discrepancy during genotyping. Although microsatellites are proven to be highly informative markers and continue to be used along with NGS based markers, though they throw up lot of challenges during genotyping and allele calling in the form of homoplasy, which demands enough care while employing the microsatellites for linkage mapping and QTL studies.

Acknowledgements

This research was supported by a research grant from the Department of Biotechnology (DBT), New Delhi, India.

References

Anmarkrud J. A., Kleven O., Bachmann L., Lifjeld J. T., 2008: Microsatellite evolution: Mutations, sequence variation, and homoplasy in the hypervariable avian microsatellite locus HrU10. BMC Evolutionary Biology 8, 138.

Barthe S., Gugerli F., Barkley N. A., Maggia L., Cardi C., Scotti I., 2012: Always look on both sides: phylogenetic information conveyed by simple sequence repeat allele sequences. PLoS One 7:e40699.

Barkley N. A., Krueger R. R., Federici C. T., Roose M. L., 2009: What phylogeny and gene genealogy analyses reveal about homoplasy in citrus microsatellite alleles. Plant Systematics and Evolution 282, 71

Bartholome J., Mandrou E., Mabiala A., Jenkins J., Nabihoudine I., Klopp C., Schmutz J., Plomion C., Gion J. M., 2015: High-resolution genetic maps of *Eucalyptus* improve *Eucalyptus grandis* genome assembly. New Phytologist 206, 1283–1296.

Bhargava A., Fuentes F. F., 2010: Mutational dynamics of microsatellites. Molecular Biotechnology 44, 250–266.

Brondani R. P. V., Williams E. R., Brondani C., Grattapaglia D., 2006: A microsatellite based consensus linkage map for species of *Eucalyptus* and a novel set of 230 microsatellite markers for the genus. BMC Plant Biology 6, 20.

Curtu A. L., Finkeldey R., Gailing O., 2004: Comparative sequencing of a microsatellite locus reveals size homoplasy within and between European oak species (*Quercus* spp.). Plant Molecular Biology Reporter 22, 339–346.

da Silva J. M., de Sousa A. C. B., de Souza A. P., Mori E. S., Freitas M. L. M., Sebbenn A. M., Demoraes M. L. T., 2009: Development and characterization of 14 microsatellite loci from an enriched genomic library of *Eucalyptus camaldulensis* Dehnh. Conservation and Genetic Resources 1, 465–469.

Dong Z., Wang H., Dong Y., Wang Y., Liu W., Miao G., Lin X., Wang D., Liu B., 2013: Extensive Microsatellite Variation in Rice Induced by Introgression from Wild Rice (*Zizania latifolia* Griseb.). PLoS One 8:e62317.

Faria D. A., Mamani E. M., Pappas M. R., Pappas G. J., Grattapaglia D. A., 2010: Selected set of EST-derived microsatellites, polymorphic and transferable across 6 species of *Eucalyptus*. Journal of Heredity 101, 512–520.

Gion J. M., Chaumeil P., Plomion C., 2015: EucaMaps: linking genetic maps and associated QTLs to the *Eucalyptus grandis* genome. Tree Genetics and Genomes 11, 1–7.

Grattapaglia D., Mamani E., Silva-Junior O. B., Faria D. A., 2014: A novel genome-wide microsatellite resource for species of *Eucalyptus* with linkage-to-physical correspondence on the reference genome sequence. Molecular Ecology Resources 15, 437–448.

He X., Wang Y., Li F., Weng Q., Li M., Xu L. A., Gan S., 2012: Development of 198 novels EST- derived microsatellites in *Eucalyptus* Myrtaceae. American Journal of Botany 99, 134–148.

Hudson C. J., Kullan A. R. K., Freeman J. S., Faria D., Grattapaglia D., Kilian A., Myburg A., Potts B. M., Vaillancourt R. E., 2012: High synteny and colinearity among *Eucalyptus* genomes revealed by high-density comparative genetic mapping. Tree Genetics and Genomes 8, 339–352.

Javed A. M., Cannon, CH. H., Wickneswari R., 2014: Microsatellite DNA Markers in *Shorea platyclados* (Dipterocarpaceae): genetic diversity, size homoplasy and mother trees. Journal of Forest Science 60, 18–27.

Masi E., Vignani R., Sensi E., Scali M., Cresti M., Biselli M., 2004: Sequence analysis of a microsatellite and its flanking regions in intraspecific hybrids of grapevine (*Vitis vinifera* L.). Vitis 43, 69–73.

Ochieng J. W., Steane D. A., Ladiges P. Y., Baverstock P. R., Henry R. J., Shepherd M., 2007: Microsatellites retain phylogenetic signals across genera in eucalypts (Myrtaceae). Genetics and Molecular Biology 30, 1125–1134.

Thamarus K., Groom K., Murrell J., Byrne M., Moran G., 2002: A genetic linkage map of *Eucalyptus globulus* with candidate loci for wood, fibre and floral traits. Theoretical and Applied Genetics 104, 379–387.

Zhou C., He X., Li F., Weng Q., Yu X., Wang Y., Li M., Shi J., Gan S., 2014: Development of 240 novel EST-SSRs in *Eucalyptus* L'Hérit. Molecular Breeding 33, 221–225.