A novel genome-wide microsatellite resource for species of *Eucalyptus* with linkage-to-physical correspondence on the reference genome sequence

DARIO GRATTAPAGLIA,*,† EVA M. C. MAMANI,* ORZENIL B. SILVA-JUNIOR†,‡
and DANIELLE A. FARIA*

*Laboratório de Genética Vegetal, EMBRAPA Recursos Genéticos e Biotecnologia, PqEB, Brasilia, DF 70770-970, Brazil,
†Programa de Ciências Genômicas e Biotecnologia, Universidade Católica de Brasília, SGAN 916, Brasilia, DF 70790-160, Brazil,
‡Laboratório de Bioinformática, EMBRAPA Recursos Genéticos e Biotecnologia, PqEB, Brasilia, DF 70770-970, Brazil

Abstract

Keystone species in their native ranges, eucalypts, are ecologically and genetically very diverse, growing naturally along extensive latitudinal and altitudinal ranges and variable environments. Besides their ecological importance, eucalypts are also the most widely planted trees for sustainable forestry in the world. We report the development of a novel collection of 535 microsatellites for species of *Eucalyptus*, 494 designed from ESTs and 41 from genomic libraries. A selected subset of 223 was evaluated for individual identification, parentage testing, and ancestral information content in the two most extensively studied species, *Eucalyptus grandis* and *Eucalyptus globulus*. Microsatellites showed high transferability and overlapping allele size range, suggesting they have arisen still in their common ancestor and confirming the extensive genome conservation between these two species. A consensus linkage map with 437 microsatellites, the most comprehensive microsatellite-only genetic map for *Eucalyptus*, was built by assembling segregation data from three mapping populations and anchored to the *Eucalyptus* genome. An overall colinearity between recombination-based and physical positioning of 84% of the mapped microsatellites was observed, with some ordering discrepancies and sporadic locus duplications, consistent with the recently described whole genome duplication events in *Eucalyptus*. The linkage map covered 95.2% of the 605.8-Mbp assembled genome sequence, placing one microsatellite every 1.55 Mbp on average, and an overall estimate of physical to recombination distance of 618 kbp/cM. The genetic parameters estimates together with linkage and physical position data for this large set of microsatellites should assist marker choice for genome-wide population genetics and comparative mapping in *Eucalyptus*.

Keywords: ancestry informative markers, *Eucalyptus globulus*, *Eucalyptus grandis*, microsatellites, simple sequence repeats

Received 5 June 2014; revision received 5 August 2014; accepted 13 August 2014

Introduction

Species of the genus *Eucalyptus* L'Hér. (*Myrtaceae*) are woody perennials native to Australia and adjacent islands. The most recent formal taxonomic revision of eucalypts (Brooker 2000) recognizes some 700 species in 13 subgenera among which the tallest flowering plant *E. regnans* (99.6 m) (Doughty 2000). Eucalypts are keystone species in their native ranges and ecologically very diverse, growing from tropical to temperate latitudes, in high rainfall to arid zones, and from the sea level to alpine altitudes (Pryor & Johnson 1981). *Symphyomyrtus* is the largest subgenus with 474 species, including the 20

or so most broadly used hardwood trees for plantation forestry in the world, among which the tropical *E. grandis*, member of section Latoangulatae, and *E. globulus* in section Maidenaria, stand out (Myburg *et al.* 2007). The wide interspecific diversity and sexual compatibility within subgenus *Symphyomyrtus* have been a major bonus to breeders. Combination of separately evolved gene pools by interspecifc hybridization has resulted in highly adapted hybrid planting material that currently cover large areas of sustainable plantation forestry (Rezende *et al.* 2014).

The unique species diversity and hybrid composition of some natural populations and breeding programmes have posed an additional challenge and an opportunity when developing molecular tools in support of population genetics analyses and breeding. The development

Correspondence: Dario Grattapaglia, Fax: +55 61 33403624; E-mail: dario.grattapaglia@embrapa.br

and validation of molecular markers are usually done across more than one species covering a wider phylogenetic range, in a way that markers can be readily transferred across species. While early types of markers such as RAPD and AFLP did not offer such attributes, microsatellites provided these key features and quickly became the main working tool for a number of applications in genetic analysis of *Eucalyptus*. Microsatellites marker development in *Eucalyptus* has evolved over the years in parallel with the increased availability of sequence information. While the first sets were derived from enriched genomic libraries (Byrne *et al.* 1996; Brondani *et al.* 1998, 2002; Steane *et al.* 2001), in the last few years, microsatellite markers have been designed after mining the increasingly larger EST (Expressed Sequence Tags) collections deposited in sequence databases. Nevertheless, only part of those new markers were effectively wet-lab evaluated for polymorphism in some species (Faria *et al.* 2010, 2011; He *et al.* 2012; Zhou *et al.* 2014), and for none of them, genome-wide linkage and physical mapping location were provided, somewhat limiting their applicability for more comprehensive and refined genome-wide applications now possible with the availability of its reference genome sequence (Myburg *et al.* 2014)

Besides population genetics studies, easily assayable microsatellites have been particularly useful in *Eucalyptus* to connect phenotypic to genotypic variation by linkage mapping and for comparative genomics experiments. Following the first genetic maps built for species of *Eucalyptus* using RAPD and RFLPs, around 26 additional genetic linkage maps were constructed for six different species of *Eucalyptus* (reviewed by Grattapaglia *et al.* 2012). It was the availability of larger sets of transferable microsatellites, however, that allowed such development and significantly enhanced the value of genetic mapping information in the genus. Genetic linkage maps including robust sets of microsatellites were essential resources for comparative genome analysis (Hudson *et al.* 2012b), whole genome assembly (Kullan *et al.* 2012; Petroli *et al.* 2012), QTL mapping (Freeman *et al.* 2009; Gion *et al.* 2011; Bartholome *et al.* 2013) and QTL validation across pedigrees (Mamani *et al.* 2010; Freeman *et al.* 2013). Additionally, dense microsatellite-based genetic maps have provided the necessary framework to map large numbers of markers derived from higher throughput but lower information content technologies such as biallelic single nucleotide polymorphisms (SNPs) (Lima *et al.* 2011) and dominant presence/absence variants genotyped with microarray-based Diversity Array Technology (DArT) (Sansaloni *et al.* 2010; Hudson *et al.* 2012a; Petroli *et al.* 2012) or sequence-based DArT-seq genotyping (Sansaloni *et al.* 2011).

Clearly, while higher throughput marker technologies are becoming available for *Eucalyptus* (Grattapaglia *et al.* 2011; Sansaloni *et al.* 2011), a rich source of laboratory-validated, genetically and physically mapped multi-allelic microsatellites should continue to be a valuable resource for genetic analysis in species of the genus. The objectives of this study therefore were (i) to develop a large set of microsatellite markers for *Eucalyptus,* to be particularly useful across phylogenetically distant species within subgenus *Symphyomyrtus*; (ii) to estimate and compare population genetic and ancestry information content of microsatellites designed from ESTs vs. random shotgun genomic sequences for the two most widely studied *Eucalyptus* species; (iii) to generate a comprehensive microsatellite-only linkage map for *Eucalyptus* aligned to the *Eucalyptus* reference genome sequence to provide linkage-to-physical position for a large number of microsatellites to assist future population genetics, population genomics and comparative mapping studies.

Material and methods

Microsatellite discovery and primer design

Microsatellites (2–6 base SSR motifs) were mined from a multispecies collection of 88 000 5′-end EST sequences as described earlier (Faria *et al.* 2010). Primer pairs were designed targeting primers 18–21 bases long, amplicon sizes between 80 and 350 bp and annealing temperature of 60°C. Genomic microsatellites were mined from a shotgun genomic library built and sequenced as part of the Genolytpus project (Grattapaglia 2004). Regular non-labelled primer pairs were synthesized for an initial screening for amplification success, interspecific transferability and polymorphism in agarose and polyacrylamide gels electrophoresis. For a selected set of primer pairs following screening, the forward primer of each pair was resynthesized labelled with different fluorochromes (HEX, 6-FAM, NED) to allow precision genotyping.

Microsatellite screening, genotyping and characterization

Initial microsatellite screening was carried out using 12 unrelated trees of five different species (*E. grandis*, *E. urophylla*, *E. globulus*, *E. camaldulensis* and *E. dunnii*) and 2 hybrids (*E. dunnii* × *E. grandis* and *E. urophylla* × *E. globulus*). PCR parameters, agarose and polyacrylamide gel electrophoresis conditions were described earlier (Brondani *et al.* 1998). Genotyping of selected microsatellites was then carried out by semi-automatic fluorescence detection in an ABI 3100XL genetic analyser as described earlier, using multiplexed

systems with at least three loci per PCR (Faria *et al.* 2010, 2011), using a home-made ROX labelled size standard (Brondani & Grattapaglia 2001). Data were collected under dye set D spectral calibration using Genescan and analysed with Genotyper (Applied Biosystems, Foster City, CA). A set of 223 selected microsatellites, 182 EST and 41 genomic, were characterized for population and parentage testing parameters (PE, probability of paternity exclusion and PI, probability of identity) using population samples of *E. grandis* (Section *Latoangulatae*) ($n = 16$) and *E. globulus* (Section *Maidenaria*) ($n = 16$), as described earlier (Faria *et al.* 2011). Species differentiation at the locus level was estimated based on allele identity (Wright's Fixation index F_{ST}) and allele size variation under the stepwise-mutation model (R_{ST}). Arlequin 3.5 (Excoffier & Lischer 2010) and its function 'locus-by-locus AMOVA' for genotypic data was used to estimate F_{ST} and test its significance by permuting individual genotypes among populations. SPAGeDi 1.4 (Hardy & Vekemans 2002) was used to estimate R_{ST} and its significance by permuting individuals' genotypes between the two populations. A false discovery rate (q value ≤ 0.05) was used to declare microsatellite loci displaying significant genetic differentiation between the two species. Separate analyses were carried out for EST and genomic microsatellites.

Individual and consensus genetic map construction

Genotyping and genetic mapping of microsatellite markers were carried out on two mapping populations of 188 F1 individuals each. One population was derived from a cross between a female *E. grandis* (clone G38) and a male *E. urophylla* (clone U15) (G38 × U15 pedigree), while the other population had been generated from a cross between two hybrids parents, a female (*E. dunnii* × *E. grandis*) and a male (*E. urophylla* × *E. globulus*) (DG × UGL pedigree). Polymorphic microsatellites between the two parents in each mapping population were genotyped and data tested for adherence to the expected Mendelian segregation ratios, either a 1:1 when only one of the parents was heterozygous, 1:2:1 following a phase-unknown F2 configuration with both parents equally heterozygous for the same genotype, or in a fully informative 1:1:1:1 configuration with three or four different alleles segregating from both heterozygous parents. Linkage analysis was carried out using JoinMap v3.0 (Van Ooijen & Voorrips 2001). Markers were assigned to linkage groups by applying the maximum-likelihood algorithm of JoinMap with population type CP; initial grouping at LOD > 15.0 followed by some individual markers fitted at LOD 10, recombination fraction ≤ 0.4; ripple value=1; jump in goodness-of-fit threshold (the normalized difference in goodness-of-fit

chi-square before and after adding a locus) equal to 5 under a Kosambi mapping function. Marker ordering with JoinMap was carried out by simulated annealing, fitting all markers to the linkage map by the third and final round of JoinMap so as to provide map position for all segregating microsatellites. Linkage group numbering followed the established consensus (Brondani *et al.* 2006), which is also the numbering adopted for the 11 main chromosome scaffolds of the reference genome sequence. Data gathered in these two mapping populations were then combined with the existing reference *E. grandis* × *E. urophylla* linkage map data (Brondani *et al.* 2006), to construct a single consensus genetic linkage map. This multipedigree consensus map was built based on common markers using JoinMap by estimating mean recombination frequencies and combined LOD scores. For each pair of loci, the (virtual) numbers of recombinant and nonrecombinant gametes in each individual population are calculated from the estimated recombination frequencies and corresponding LOD scores. The total numbers of recombinant and nonrecombinant gametes over all populations were then calculated by totalling the numbers of the individual populations (Van Ooijen & Voorrips 2001). Map figures were constructed using Mapchart v2.2 (Voorrips 2002).

Positioning of microsatellites and alignment of the linkage map onto the Eucalyptus reference genome sequence

To assess the consistency between the ordering of marker loci in the linkage map and their physical position in the genome, to provide a genomic location for the mapped microsatellites and to estimate the effective genome coverage provided by the map, all the available original cDNA or genomic sequences where from the microsatellites were derived were aligned onto the masked *Eucalyptus grandis* genome sequence (version 1.0 available in Phytozome 9.1 – http://www.phytozome.net/) using BLASTN. For some microsatellites, the primers were used because the original sequences were not available. Annotation of the genomic region corresponding to the microsatellite was provided using BLASTX reporting the best hit. When more than one hit was observed, possibly corresponding to duplicated regions in the genome, we reported the hit with the lowest e-value.

Results

Microsatellite marker development

Data mining of 22 298 EST unigenes resulted in 1765 sequences matching the simple sequence repeat search criteria. In the EST sequences 54% of the SSRs were

trinucleotide repeats, followed by dinucleotides (24%). Of the 7395 shotgun genomic sequences analysed, 807 contained microsatellites matching the criteria, with dinucleotides as the most frequent repeats (48%) followed by trinucleotides (20%). Primers could be designed for 1244 (70.5%) of the 1765 microsatellite sequences identified in the ESTs and for only 178 (22%) of the 807 shotgun genomic sequences (Table 1). From the 1244 primer pairs designed for the EST loci, 759 were selected for screening prioritizing those microsatellites that displayed the largest number of tandemly repeated units. Following primer screening, 494 (65%) amplified discrete amplicons showing length polymorphism among *Eucalyptus* species in agarose gels. The complete source sequence of each one of these 494 EST loci was submitted to GenBank, and the corresponding primers sequences, motif, expected amplicon size (bp) and GenBank Accession nos are provided (Table S1, Supporting information). Among the 494 EST microsatellite loci, 182 displaying the clearest polymorphism in the screening step were selected and fluorochrome-labelled forward primers synthesized for downstream use in genetic mapping and population analysis. For the microsatellites from shotgun genomic sequences, all 178 primer pairs were tested, 111 amplified single amplicons, 53 (47.7%)

displayed polymorphism in the screening step, and eventually 41 were selected for further characterization using fluorescence detection. Original sequences, primers, motif and Accession nos for these markers are also provided (Table S1, Supporting information). From a total of 535 microsatellites reported in this study (494 EST and 41 genomic), 22 are based on hexanucleotide repeats, 13 on pentanucleotides, 43 on tetranucleotides, 255 on trinucleotides and 201 on dinucleotides. All hexa, penta and tetranucleotide repeat microsatellites were EST derived but three tetranucleotides, which were genomic derived. Annotation of the *Eucalyptus* EST microsatellites following BLASTX against the *Eucalyptus grandis* reference genome (http://www.phytozome.net/ Eucalyptus.php – accessed in March, 2014) is also provided (Table S2, Supporting information). For 16 EST microsatellites, no hit was found to any predicted gene model, and for 49, a hit was obtained, but no functional annotation exists for this locus yet. Therefore, of the 494 EST microsatellites, 437 of them could allow mapping a known predicted gene model in future mapping experiments. For 21 of these microsatellites, best hits with e-values of 0 were seen to one of the additional small still unanchored scaffolds of the genome assembly (e.g. EMBRA872 and EMBRA898) (Table S2, Supporting

Table 1 Statistics of microsatellite development steps and summary of genetic information content of 223 (182 EST and 41 Genomic) newly developed microsatellites in two phylogenetically contrasting *Eucalyptus* species (see Tables S3 and S4, Supporting information for individual locus estimates)

Microsatellite development step		EST microsatellites		Genomic microsatellites
Total number of sequences mined		22 298		7395
Number of sequences containing microsatellites		1765		807
Microsatellites for which primers could be designed		1244		178
Primers pairs screened		759		178
Primer pairs amplifying discrete amplicons		494		111
Microsatellites selected for characterization		182		41

Parameters for polymorphic microsatellites	*Eucalyptus grandis*	*Eucalyptus globulus*	*Eucalyptus grandis*	*Eucalyptus globulus*
No. markers	167	165	40	41
Average no. alleles	4.9 ± 2.3	5.2 ± 2.5	7.6 ± 3.7	6.4 ± 2.8
Average H_{obs}	0.514 ± 0.246	0.514 ± 0.266	0.623 ± 0.226	0.475 ± 0.227
Average H_{esp}	0.620 ± 0.197	0.619 ± 0.215	0.738 ± 0.196	0.680 ± 0.207
Average PIC	0.555 ± 0.195	0.552 ± 0.208	0.681 ± 0.199	0.620 ± 0.206
Average PE	0.390 ± 0.188	0.389 ± 0.193	0.530 ± 0.210	0.454 ± 0.202
Average PI	0.241 ± 0.188	0.365 ± 0.254	0.150 ± 0.174	0.193 ± 0.176
No. markers in HWE (*q* value ≤ 0.05)	136	137	34	22
Range of single locus F_{ST}	0–0.937		0.018–0.743	
No. markers with significant F_{ST} (*q* value ≤ 0.05)	139 of 177 (79%)		41 of 41 (100%)	
Range of single locus R_{ST}	−0.104 to 0.992		−0.041 to 0.940	
No. markers with significant R_{ST} (q value ≤ 0.05)	95 of 177 (54%)		22 of 41 (54%)	
No. markers with significant F_{ST} and R_{ST} (q value ≤ 0.05)	94		22	

information). The identities between the original microsatellite containing sequence and the genome sequence were generally high, with only 20 loci displaying values below 80% possibly resulting from higher rates of sequence polymorphism between the reference genome and the original tree where from the microsatellite containing sequence was isolated.

Microsatellite characterization

Over 90% of the 182 EST microsatellites tested were confirmed polymorphic in each species studied, 167 in *E. grandis* and 165 in *E. globulus,* and 155 were simultaneously polymorphic in the two species (Table 1). In the population samples analysed, 10 microsatellites were monomorphic in *E. grandis*, and 12 in *E. globulus*, while only five markers were monomorphic in both species samples (Table S3, Supporting information). All tested genomic derived microsatellites were polymorphic in *E. globulus* and 40 of 41 in *E. grandis*. More than 80% of the EST markers were in HWE in both species. As for the genomic markers, 85% were in HWE in *E. grandis* but only 54% in *E. globulus.* Deviation from HWE was always due to a lower than expected heterozygosity. The average number of alleles was similar in the two species for the EST markers, but higher for the genomic markers in *E. grandis* when compared to *E. globulus*. Genomic markers had between 23 and 55% more alleles in both species when compared to those from ESTs. The higher hypervariability of genomic microsatellites resulted in higher heterozygosities, both observed and expected, higher information content for genetic mapping (PIC), and higher power of parentage exclusion and individual discrimination, as revealed by higher estimates of PE and lower PI (Table S4, Supporting information). Significant genetic differentiation between the two species was seen at the majority of the EST microsatellites (139 of 177; 79%) and at all 41 genomic loci based on allele identity (F_{ST}), while only at 54% of the loci when based on allele size variation (R_{ST}), both for EST (95 in 177) and genomic markers (22 in 41). Significant differentiation measured both by F_{ST} and R_{ST} was seen at 94 EST and 22 genomic microsatellites.

The allele size range for all EST microsatellites was essentially overlapping between the two species with only 11 of the 182 loci displaying a nonoverlapping size range accompanied by a high and significant estimate of R_{ST} (EMBRA904, EMBRA922, EMBRA925, EMBRA997, EMBRA1307, EMBRA1362, EMBRA1363, EMBRA1428, EMBRA1578, EMBRA1757 and EMBRA1973). A similar proportion (3 in 41) of genomic markers had such a behaviour (EMBRA653, EMBRA691 and EMBRA695) (Tables S3 and S4, Supporting information). The frequencies of null alleles at all loci were estimated under the individual inbreeding model by which an estimate of null allele frequency is considered significantly different from zero only when the locus deviates from HWE expectations and not based on its absolute estimated value. Of the 31 EST loci deviating from HWE, 28 had null allele frequency higher than 0.1 in *E. grandis* and 23 of 28 in *E. globulus*. All six genomic markers deviating from HWE had null allele frequencies greater than 0.1 in *E. grandis* and 18 of 19 in *E. globulus*. Seven EST loci (EMBRA1063, EMBRA1367, EMBRA1469, EMBRA1474, EMBRA1625, EMBRA1679 and EMBRA1945) and three genomic loci (EMBRA640, EMBRA705 and EMBRA710) had such features in both species. These loci are likely more subject to the occurrence of null alleles, which could complicate genetic analysis especially in parentage testing.

Linkage map construction

Of the 223 newly developed microsatellites for which labelled primers were synthesized, 178 were polymorphic between the parents of the two mapping populations and were mapped, while 45 loci were not informative. Of the 178 newly mapped microsatellites, 142 were from ESTs and 36 from genomic sequences. Mapping statistics for the DGxUGL and the G38xU15 pedigree maps are provided (Table S5, Supporting information). For the G44xU28 pedigree, the original segregation data (Brondani *et al.* 2006) were used to build a new version of a linkage map. A consensus map was then built by combining each corresponding linkage group of the three populations. A map with 437 microsatellites (295 genomic SSRs and 142 EST-SSRs) was constructed, covering 1065 cM with an average distance between adjacent markers equal to 2.6 cM (Table S5, Supporting information Fig. 1 and Data S1, Supporting information). Of the 437 linkage mapped microsatellites, seven revealed two separately segregating loci mapped to different linkage groups (EMBRA22, EMBRA91, EMBRA135, EMBRA157, EMBRA195, EMBRA205 and ES140). The two loci for each EMBRA marker were labelled with an additional subscript 'a' and 'b', where 'a' was assigned to the locus mapped to the same linkage group as originally defined in the microsatellite reference map (Brondani *et al.* 2006), unless the physical mapping position indicated otherwise. The final linkage map therefore had 423 microsatellites positioned at a unique locus and 7 at two independent loci (Fig. 1).

Alignment of the microsatellite linkage map to the reference genome sequence

Of the 437 mapped microsatellite loci, 414 (94.7%) were physically assigned to the *Eucalyptus grandis* reference genome assembly, therefore allowing successful
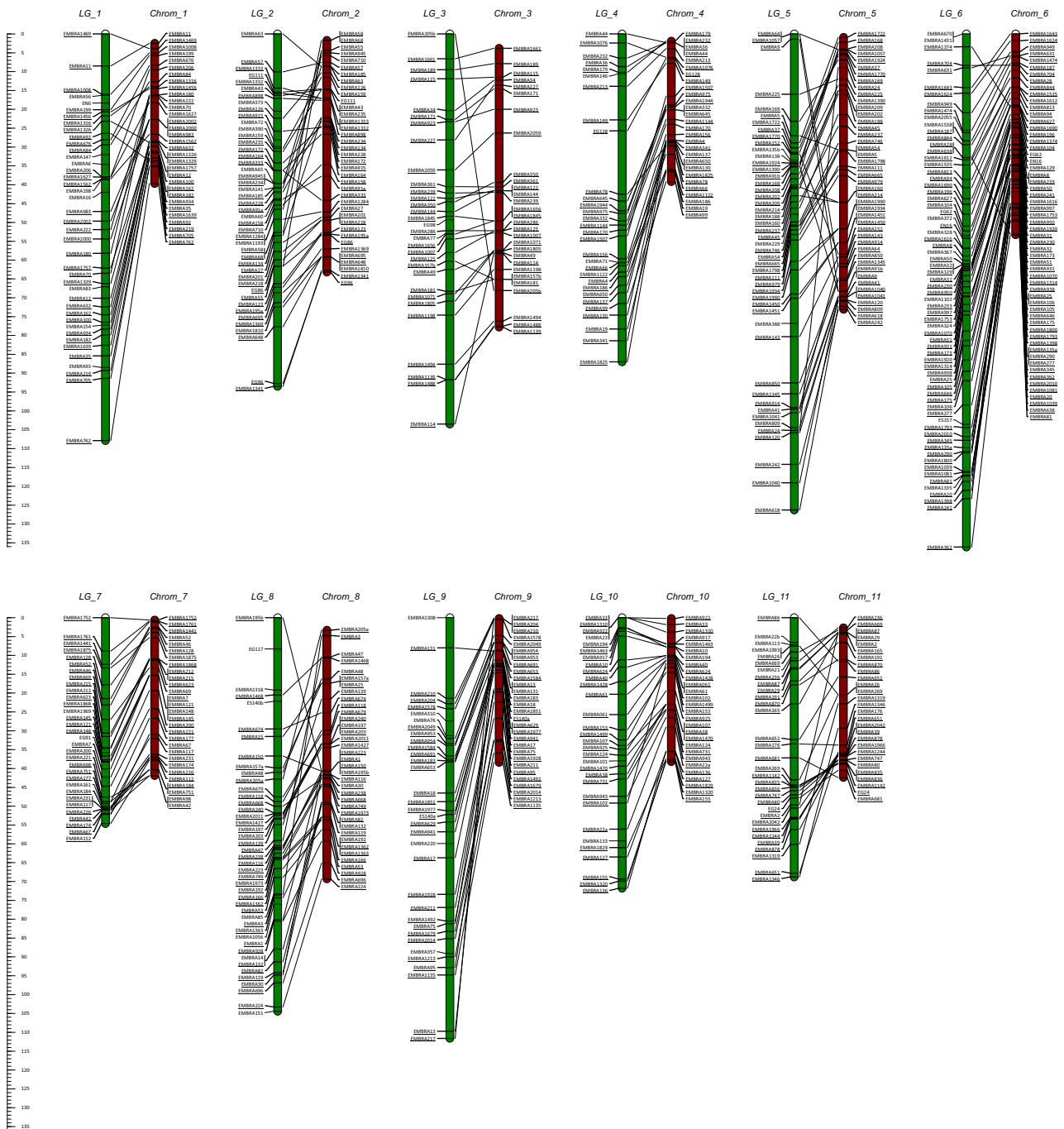
Fig. 1 Alignment of the multipedigree consensus linkage map (green bars) to the *Eucalyptus grandis* reference genome sequence (red bars). Microsatellites that were both linkage and physically mapped consistently to the chromosome scaffold corresponding to the linkage group number are connected by lines and are underlined. The scale on the left corresponds simultaneously to centiMorgan (cM) distances for the linkage map and to Mbp of sequence for the chromosome scaffolds.

alignment of the linkage map to the genome sequence (Fig. 1 and Supporting Information Data S1). Of the 414 loci, 380 were physically assigned to a chromosome scaffold whose number matched the linkage group number while 57 did not. These 57 loci were classified as follows;

(i) 27 were physically assigned to a chromosome numbered differently than the linkage group number to which it was mapped; these 27 loci likely represent duplicated genomic regions in the genome; (ii) 5 were physically assigned to one of the 4941 additional small

scaffolds numbered beyond e 11 main chromosome scaffolds; these small scaffolds likely correspond to the unassembled alternative haplotype at the same locus; (iii) 25 microsatellites could not be physically assigned to any chromosome scaffold, noting, however, that for 23 of them the original sequence was not available and the alignment could only be attempted using the primer sequences (e.g. microsatellites named EG, EN and ES Byrne *et al.* 1996) (Table S6, Supporting information). Of the 380 loci assigned to the 11 chromosomes, 376 loci had a unique physical position and four microsatellites, for which a second locus was linkage mapped, were physically mapped to a second position still on one of the 11 main chromosome scaffolds (EMBRA91, EMBRA157, EMBRA195 and EMBRA205). Finally, 28 of the 380 loci consistently assigned to the 11 linkage groups and corresponding chromosomes, additionally physically mapped to one of 4941 small scaffolds, likely corresponding to the unassembled alternative haplotype at the same locus. Jointly, these 28 markers therefore provided evidence to locate 21 small scaffolds, totalling 1.27 Mbp of still unanchored sequence, to the 11 main chromosomes of the genome assembly (Table S7, Supporting information). Overall, the linear ordering of the microsatellites along the linkage map agreed well with their relative physical position in the genome, although some sparse order inconsistencies were observed mainly on chromosomes 5 and 8 (Fig. 1). The alignment of 380 microsatellites that were both linkage and physically mapped to the genome sequence spanned 576.4 Mbp (95.2%) of the 605.8-Mbp genome sequence assembled into the 11 chromosomes, with a microsatellite physically mapped, on average, every 1.55 ± 0.4 Mbp. The ratio between physical sequence length in Mbp and genetic distance in cM, calculated by dividing the average Mbp intermarker distance by the average cM intermarker distance per chromosome, varied from 0.405 Mbp/cM for chromosome 9 to 0.841 Mbp/cM for chromosome 7 with an estimated overall average of 0.618 Mbp/cM (Table 2).

Discussion

In this work, we have provided a novel genomic resource for species of *Eucalyptus*, consisting of 535 microsatellites, 494 designed from ESTs and 41 from shotgun genomic libraries. Of the 494 EST microsatellites, 453 are novel, and 41 had been previously reported (Faria *et al.* 2010, 2011), but no attempt was made to linkage map them. Of this set of markers, 223 were characterized for transferability and population parameters in two contrasting *Eucalyptus* species of subgenus *Symphyomyrtus*. Additionally 178 of them were linkage mapped on a consensus map built by compiling segregation and

Table 2 Mapping statistics of the multipedigree consensus linkage map and its alignment to the 11 chromosome scaffolds of the *Eucalyptus grandis* reference genome sequence

LG/Chromosome	1	2	3	4	5	6	7	8	9	10	11	Total	Mean
No. of Linkage mapped markers	41	45	31	30	48	69	32	44	34	31	32	437	39.7 ± 11.7
No. of EST-SSRs	14	10	12	7	17	30	6	9	15	11	11	142	12.9 ± 6.5
No. of Genomic SSR	27	35	19	23	31	39	26	35	19	20	21	295	26.8 ± 7.2
Total size (cM)	107.9	93.5	103.5	87.0	126.3	136.1	54.6	104.4	111.6	71.8	68.8	1065.3	96.8 ± 24.8
Intermarker distance (cM)	2.69	2.12	3.45	3.00	2.69	2.00	1.76	2.43	3.38	2.39	2.22	–	2.56 ± 0.5
No. of Physically mapped markers	32.0	38.0	29	28	44	58	29	37	29	28	28	380	34.55 ± 9.4
Physical distance covered (Mbp)	39.7	63.2	77.81	39.21	73.1	53.26	41.91	69.3	38.4	38.23	42.37	576.4	–
Intermarker distance (Mbp)	1.2	1.7	2.64	1.38	1.68	0.93	1.48	1.83	1.37	1.41	1.47	–	1.55 ± 0.4
Ratio Mbp/cM	0.446	0.783	0.765	0.460	0.625	0.465	0.841	0.753	0.405	0.590	0.662	–	0.618 ± 0.156

linkage mapping data of microsatellites in three independent pedigrees. This consensus map, with 437 microsatellite markers, is the most comprehensive microsatellite-only genetic map for *Eucalyptus* and for any forest tree species to date. Finally, this linkage map was aligned to the *Eucalyptus grandis* genome sequence providing information on the genome-wide distribution and physical position of 380 microsatellites, including already widely used ones (Brondani *et al.* 1998, 2006) and 178 new ones developed in this study.

Microsatellite hypervariability and conservation across Eucalyptus species

A total of 223 microsatellites were effectively characterized for polymorphism in two phylogenetically contrasting *Eucalyptus* species. Experience from our and others' transferability experiments (Kirst *et al.* 1997; Steane *et al.* 2001; Shepherd *et al.* 2006; Faria *et al.* 2010; He *et al.* 2012; Zhou *et al.* 2014), however, strongly points to the fact that more than 80% of the 494 new markers should transfer well and be polymorphic across a wide spectrum of *Eucalyptus* species and even related genera. This should be the case particularly for the 453 markers designed from more conserved transcribed sequences, reminding that all the microsatellites were most likely derived from the 5′ untranslated region of the EST, given that they were designed from 5′ sequenced ESTs and that simple sequence repeats in exonic regions would hardly be hypervariable. Of the 182 characterized EST microsatellites, 155 were simultaneously polymorphic in the two species considered, 22 were polymorphic in only one of them and only five were monomorphic in the population samples analysed, although these could still prove polymorphic in larger samples sizes. These results clearly corroborate the high level of microsatellite hypervariability and sequence diversity observed for *Eucalyptus* species. A significantly higher average number of alleles was seen for the genomic microsatellites when compared to those designed from EST sequences, both in *E. grandis* ($p = 4.6 \times 10^{-5}$) and *E. globulus* ($p = 0.0138$) consistent with theoretical expectations. Although this had been suggested earlier based on subjective assessment between studies (Faria *et al.* 2010), no experimental comparison was ever carried out by genotyping a large number of loci in the same individuals like it was done in this study.

The high level of microsatellite transferability and the overlapping allele size ranges observed for 209 of the 223 characterized microsatellites substantiate the conservation and genome synteny between these two species as observed by comparative mapping (Hudson *et al.* 2012b). These results also suggest that these microsatellites most likely evolved in the common

ancestor of the two *Eucalyptus* species representative of two most divergent sections within subgenus *Symphyomyrtus*, therefore pre-dating species radiation, is consistent with molecular dating estimates suggesting a relatively recent radiation of *Symphyomyrtus* species taking place between 10 and 36 MYA (Crisp *et al.* 2004). Nevertheless, one should not overlook that complex mutational patterns and high levels of allele size homoplasy are common occurrences in microsatellites, resulting from the replication slippage-based mutation process, so that alleles identical by size may not be identical by descent (Lia *et al.* 2007). The microsatellite resource presented in this study can thus be very useful for genetic analysis within and across the two and other closely related species of *Symphyomyrtus*, as the same allelic size range can be expected, a feature that facilitates microsatellite multiplexing. However, the unknown patterns of sequence mutation and size homoplasy do not support their use for interspecific phylogenetic inferences.

While the vast majority (> 80%) of the EST-derived microsatellites were in HWE in both species, the proportion of genomic microsatellites in HWE was considerably smaller in *E. globulus* (22 in 41; 54%) than *E. grandis* (34 in 40; 75%) (Table S3, Supporting information). As expected, a coincidence was seen between loci that deviated from HWE expectations due to heterozygosity deficiency and loci with increased estimates of null allele frequencies. PCR primers for these markers are possibly located on genomic regions harbouring relatively frequent sequence polymorphisms that cause amplification failure resulting in null alleles and apparent homozygous genotypes. Two issues should be considered in this comparison. First, these results are coherent with expectations that microsatellites whose primers are designed on more conserved genic regions should be less prone to the occurrence of null alleles due to sequence polymorphism in the priming sites when compared to microsatellites designed from nongenic and therefore more polymorphic genomic regions. Second, since the genomic shotgun sequences used for primer design of the genomic microsatellites were from *E. grandis*, the higher proportion of microsatellites not fitting to HWE when analysed in *E. globulus* can be explained by interspecific sequence polymorphism causing null alleles. Although simulation studies had found that the incidence of null alleles does not significantly impact overall estimates of exclusion probabilities, the use of such markers in experimental studies may lead to considerable errors in individual parentage assignments, notably the declaration of mistaken parentage exclusions (Dakin & Avise 2004). Our results therefore indicate that while most EST-derived microsatellites would be less subject

to null alleles, and as such could be confidently used across species, this is not the case for microsatellites designed from genomic sequences. Out of those microsatellites that deviated from HWE, eleven from EST sequences (EMBRA870, EMBRA1056, EMBRA1063, EMBRA1319, EMBRA1329, EMBRA1367, EMBRA1469, EMBRA1474, EMBRA1625, EMBRA1679 and EM-BRA1945) and 3 from genomic sequences (EMBRA640, EMBRA705 and EMBRA710) did so in both species. These are definitely markers that should be used with caution or possibly avoided for parentage testing. It is important to note that due to the relatively high levels of nucleotide diversity in the *Eucalyptus* genome, it is likely that most microsatellites will suffer to some extent from the occurrence of null alleles requiring adequate treatment of genotype data such as allowing a certain level of parent-offspring mismatches (Slavov *et al.* 2005) when establishing threshold LOD scores to assign parentage.

Genetic divergence at microsatellites and ancestry informative markers

Significant genetic divergence was seen at most microsatellites consistent with the taxonomically established separation of the two species. After applying a false discovery rate (q value ≤0.05), more EST microsatellites displaying genetic divergence were detected based on F_{ST} (139) than R_{ST} (95), and 94 loci revealed significant divergence by both measures (Table 2). Both measures of differentiation when applied to microsatellite data are known to have limitations. While the assumption of low mutation rate used to estimate F_{ST} does not apply to microsatellites, it is also true that the mutational processes at such loci are known to deviate more or less strongly from an ideal stepwise- or generalized mutation model, which may lead to biased estimates of R_{ST} (Hardy *et al.* 2003). Nevertheless, comparisons of F_{ST} and R_{ST} values on microsatellite data have been used as a means to infer the relative importance of mutation vs. migration rates within species. When genetic differentiation is high, a higher R_{ST} relative to F_{ST} is generally found, indicating that mutation explains more of the differentiation (Hardy *et al.* 2003). For 72 of the 94 (77%), EST microsatellites showing significant differentiation between the two species by both measures, the estimate of R_{ST} was nominally larger than F_{ST} (Table S3, Supporting information). Similar trend was observed for the genomic microsatellites. This result is entirely consistent with the unambiguous differentiation between the two species, including the considerable genome size difference between *E. grandis* (640 Mbp) and *E. globulus* (530 Mbp) (Grattapaglia & Bradshaw 1994), recently found

to be derived from more dynamic genome size evolution beyond standard transposable element activity (J. Tibbits unpublished data).

No difference in the proportion of microsatellites showing significant F_{ST} or R_{ST} was seen among the 11 chromosomes. This proportion varied between 0.73 and 1.00 and a linear regression of the number of markers displaying significant differentiation on the total number of markers resulted in an R^2 of 0.948 with no clear outlier linkage group. Given the sparse nature of the data in hand, with large differences in the total numbers of markers assayed per linkage group, we really do not have enough data to speculate any further about genome-wide genetic differentiation between the two species. Our objective was only to provide information about the markers themselves, not on the underlying genomic segment. This will require much higher density genotyping along the genome. From the practical standpoint, the interspecific genetic differentiation captured by the newly developed microsatellites allows one to select a set of ancestry informative markers (AIM) to be used for genome-wide ancestry estimation in hybrids derived from these two species. The estimate of F_{ST} has been shown to be strongly correlated with the first principal component in a PCA and therefore an adequate metric to select AIMs (McVean 2009; Sawler *et al.* 2013). From the 94 EST microsatellites displaying significant F_{ST} and R_{ST}, it is possible to select, for example, a set of 31 loci in HWE in both species, mapping to 7 of the 11 chromosomes having an F_{ST} > 0.3 and R_{ST} > 0.7, to be used for genomic ancestry determination in hybrids of these two species. Similar sets of AIM could be identified once genotype data for population samples of other species are collected for these microsatellites.

Linkage-to-physical mapping of microsatellites

Extensive colinearity between the genome of *Eucalyptus* species belonging to subgenus *Symphyomyrtus* had been suggested since the first linkage mapping experiments that used small sets of transferable microsatellites (Marques *et al.* 2002). Recently reported higher density linkage maps based on DArT markers confirmed and expanded those same findings at a much higher resolution (Hudson *et al.* 2012a,b; Petroli *et al.* 2012). This fact allowed us to confidently integrate linkage map data from the three independent mapping populations into a microsatellite-only consensus linkage map. It is noteworthy, however, that despite the large number of DArT markers mapped to a recently reported reference linkage map (Hudson *et al.* 2012a), only 81 DArT markers out of 3880 (~2%) could actually be mapped to two or more of the seven independent mapping populations used to generate the consensus map. The key resource that effec-

tively allowed the construction of the consensus map was a set of 213 microsatellites mapped, out of which 193 came from the original reference microsatellite map published earlier (Brondani *et al.* 2006). These microsatellites provided locus bridges among the several mapping pedigrees and by consequence among the different sets of DArT markers mapped in each one of them. These results are very illustrative of the limitations that such high throughput but dominantly inherited markers have for comparative mapping purposes, while at the same time demonstrate the value of multi-allelic transferable microsatellites providing fully informative segregation configurations in multiple pedigrees. SNP markers, although codominant, are biallelic as well and therefore also limited for genetic mapping especially in single-generation pedigrees where markers segregating 1:2:1 in phase-unknown configurations cannot be mapped. In fact, we saw that of 768 SNPs, only around 30% segregated in two mapping populations in a pseudo-testcross configuration, providing no clear advantage over microsatellites besides the faster data generation (Lima *et al.* 2011). We therefore are confident that the novel microsatellite resource presented in this work will be highly valuable for any future comparative genome mapping study in eucalypts.

Recently, we provided linkage-to-physical data for linkage mapped DArT markers, but microsatellites were not contemplated (Petroli *et al.* 2012). This is therefore the first study in *Eucalyptus* where microsatellites were both linkage and physically mapped to the reference genome. An overall colinearity between recombination-based and physical positioning was observed for 84% of the mapped microsatellites with 61 of 380 markers displaying inverted orders (Fig. 1). For each linkage group, we counted the number of markers that once removed would make the remaining ones be in sequential order in relation to the physical assembly. It should be pointed out, however, that our intention was not to evaluate whether the linkage mapping order matched the estimated physical order in the genome. Rather, our objective was simply to check whether the microsatellite assignment based on linkage mapping matched the assignment based on sequence alignment. This is so because we are aware of two basic limitations in trying to do this: (i) the current assembly of the Eucalyptus genome might contain local mis-assemblies; (ii) our microsatellite data set would not provide sufficient marker density to try to tackle local inconsistencies in any appropriate way. Furthermore, order inconsistencies seen can generally be attributed to various sources of genotyping inaccuracies, missing data and artefacts (Hackett & Broadfoot 2003) and not to any major structural variation. Locus duplications along the same chromosome and in independent chromosomes could, however,

explain, respectively, inverted orders between linkage and physical positions, and nonsyntenic markers, that is markers that linkage map to one chromosome but are physically located to a different one. This would happen when only one of the two repeated loci segregates and gets linkage mapped while the best sequence alignment to the genome is declared to the alternative locus. In fact, although the majority of the microsatellites (380 of 437; 87%) could be genetically and physically located to a single position, different patterns of locus duplications both from the linkage and physical mapping perspectives were also seen for the remaining microsatellites (detailed in Results). These results are in line with previous observations in linkage mapping studies (Hudson *et al.* 2012b; Petroli *et al.* 2012) and provide additional support to the recent findings that the *Eucalyptus* genome underwent a lineage-specific whole genome duplication event and has the largest proportion of tandem duplications in any plant genome (Myburg *et al.* 2014)

The alignment of the linkage map to the genome sequence showed that the microsatellites grant a comprehensive coverage (95.2%) of the 605.8 assembled genome sequence, place, on average, one microsatellite every 1.55 Mbp and converge to an overall estimate of physical to recombination distance of 618 kbp/cM (Table 2), higher than the previous estimate of 513.4 kbp/cM based on high density DArT marker mapping (Petroli *et al.* 2012). Likewise, this mapping experiment also provides evidences in support of the completeness of the *Eucalyptus* reference genome sequence. Only 25 microsatellites could not be physically mapped at the thresholds used for alignment, but for 23 of them, we did not have the full sequence and therefore the poor alignment may explain this result. Additionally only 21 of the 4941 small unanchored scaffolds totalling 1.27 of the 85.4 Mbp of sequence were captured by the mapped microsatellites. As suggested earlier (Petroli *et al.* 2012), these small scaffolds possibly correspond to the alternative haplotypes at the same loci and not to additional unrepresented sequence in the assembly.

Conclusion

We reported a novel and comprehensive microsatellite resource for *Eucalyptus*, including detailed information on linkage map and physical genome location, as well as polymorphism, information content and genetic divergence for several hundred markers in the two most widely studied species of the genus. The map and physical information provided should enhance population genetics and genomics studies by allowing one to choose informative microsatellites covering the whole *Eucalyptus* genome or targeting specific genomic segments, depending on the application, while avoiding violations of the

premise of independent segregation when that is needed. In view of the generally high transferability of microsatellites, this resource should be valuable for comparative genomic and population studies involving additional *Eucalyptus* species. Although high-throughput sequence-based SNP and PAV (presence absence variant) marker assays are increasingly becoming accessible, the simple Mendelian inheritance, high multiallelism and low running cost of microsatellites still make them a very useful and accessible tool for fast and precise genetic analysis in *Eucalyptus*.

Acknowledgements

References

Bartholome J, Salmon F, Vigneron P *et al.* (2013) Plasticity of primary and secondary growth dynamics in *Eucalyptus* hybrids: a quantitative genetics and QTL mapping perspective. *BMC Plant Biology*, **13**, 120.

Brondani RP, Grattapaglia D (2001) Cost-effective method to synthesize a fluorescent internal DNA standard for automated fragment sizing. *BioTechniques*, **31**, 793–795, 798, 800.

Brondani RPV, Brondani C, Tarchini R, Grattapaglia D (1998) Development, characterization and mapping of microsatellite markers in *Eucalyptus grandis* and E-urophylla. *Theoretical and Applied Genetics*, **97**, 816–827.

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

Brondani RPV, Williams ER, 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.

Brooker MIH (2000) A new classification of the genus *Eucalyptus* L'Her. (Myrtaceae). *Australian Systematic Botany*, **13**, 79–148.

Byrne M, Marquezgarcia MI, Uren T, Smith DS, Moran GF (1996) Conservation and genetic diversity of microsatellite loci in the genus *Eucalyptus*. *Australian Journal of Botany*, **44**, 331–341.

Crisp M, Cook L, Steane D (2004) Radiation of the Australian flora: what can comparsions of molecular phylogenies across multiple taxa tell us about evolution of diversity in present-day communities? *Philosophical Transactions of the Royal Society of London – Series B: Biological Sciences*, **359**, 1551–1571.

Dakin EE, Avise JC (2004) Microsatellite null alleles in parentage analysis. *Heredity*, **93**, 504–509.

Doughty RW (2000) *The Eucalyptus. A Natural and Commercial History of the gum Tree*. The Johns Hopkins University Press, Baltimore and London.

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564–567.

Faria DA, Mamani EMC, Pappas MR, Pappas GJ, Grattapaglia D (2010) A Selected Set of EST-Derived Microsatellites, Polymorphic and Transferable across 6 Species of *Eucalyptus*. *Journal of Heredity*, **101**, 512–520.

Faria DA, Mamani EMC, Pappas GJ, Grattapaglia D (2011) Genotyping systems for *Eucalyptus* based on tetra-, penta-, and hexanucleotide repeat EST microsatellites and their use for individual fingerprinting and assignment tests. *Tree Genetics & Genomes*, **7**, 63–77.

Freeman JS, Whittock SP, Potts BM, Vaillancourt RE (2009) QTL influencing growth and wood properties in *Eucalyptus globulus*. *Tree Genetics & Genomes*, **5**, 713–722.

Freeman JS, Potts BM, Downes GM *et al.* (2013) Stability of quantitative trait loci for growth and wood properties across multiple pedigrees and environments in *Eucalyptus globulus*. *New Phytologist*, **198**, 1121–1134.

Gion JM, Carouche A, Deweer S *et al.* (2011) Comprehensive genetic dissection of wood properties in a widely-grown tropical tree: *Eucalyptus*. *BMC Genomics*, **12**, 301.

Grattapaglia D (2004) Integrating genomics into *Eucalyptus* breeding. *Genetics and Molecular Research*, **3**, 369–379.

Grattapaglia D, Bradshaw HD (1994) Nuclear-DNA content of commercially important *Eucalyptus* species and hybrids. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere*, **24**, 1074–1078.

Grattapaglia D, Silva OB, Kirst M *et al.* (2011) High-throughput SNP genotyping in the highly heterozygous genome of Eucalyptus: assay success, polymorphism and transferability across species. *BMC Plant Biology*, **11**, 65.

Grattapaglia D, Vaillancourt RE, Shepherd M *et al.* (2012) Progress in Myrtaceae genetics and genomics: Eucalyptus as the pivotal genus. *Tree Genetics & Genomes*, **8**, 463–508.

Hackett CA, Broadfoot LB (2003) Effects of genotyping errors, missing values and segregation distortion in molecular marker data on the construction of linkage maps. *Heredity*, **90**, 33–38.

Hardy OJ, Vekemans X (2002) SPAGEDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*, **2**, 618–620.

Hardy OJ, Charbonnel N, Freville H, Heuertz M (2003) Microsatellite allele sizes: a simple test to assess their significance on genetic differentiation. *Genetics*, **163**, 1467–1482.

He XD, Wang Y, Li FG *et al.* (2012) Development of 198 novel Est-Derived microsatellites in *Eucalyptus* (Myrtaceae). *American Journal of Botany*, **99**, E134–E148.

Hudson CJ, Freeman JS, Kullan ARK *et al.* (2012a) A reference linkage map for *Eucalyptus*. *BMC Genomics*, **13**, 240.

Hudson CJ, Kullan ARK, Freeman JS *et al.* (2012b) High synteny and colinearity among Eucalyptus genomes revealed by high-density comparative genetic mapping. *Tree Genetics & Genomes*, **8**, 339–352.

Kirst M, Brondani RPV, Brondani C, Grattapaglia D (1997) Screening of designed primer pairs for recovery of microsatellite markers and their transferability among species of *Eucalyptus*. In: *IUFRO Conference on Eucalyptus Genetics and Silviculture* (ed. EMBRAPA), pp. 167–171. EMBRAPA CNPF, Salvador, BA, Brazil.

Kullan ARK, van Dyk MM, Jones N *et al.* (2012) High-density genetic linkage maps with over 2,400 sequence-anchored DArT markers for genetic dissection in an F2 pseudo-backcross of *Eucalyptus grandis* × E. urophylla. *Tree Genetics & Genomes*, **8**, 163–175.

Lia VV, Bracco M, Gottlieb AM, Poggio L, Confalonieri VA (2007) Complex mutational patterns and size homoplasy at maize microsatellite loci. *Theoretical and Applied Genetics*, **115**, 981–991.

Lima B, Silva-Junior O, Faria D *et al.* (2011) Assessment of SNPs for linkage mapping in *Eucalyptus*: construction of a consensus SNP/microsatellite map from two unrelated pedigrees. *BMC Proceedings*, **5**, P31.

Mamani EMC, Bueno NW, Faria DA *et al.* (2010) Positioning of the major locus for *Puccinia psidii* rust resistance (Ppr1) on the *Eucalyptus* reference map and its validation across unrelated pedigrees. *Tree Genetics & Genomes*, **6**, 953–962.

Marques CM, Brondani RPV, Grattapaglia D, Sederoff R (2002) Conservation and synteny of SSR loci and QTLs for vegetative propagation in four *Eucalyptus* species. *Theoretical and Applied Genetics*, **105**, 474–478.

McVean G (2009) A genealogical interpretation of principal components analysis. *Plos Genetics*, **5**, e1000686.

Myburg AA, Potts BM, Marques CM *et al.* (2007) Eucalyptus. In: *Genome Mapping and Molecular Breeding in Plants – Forest Trees* (ed Kole C), pp. 115–160. Springer, Germany.

Myburg AA, Grattapaglia D, Tuskan GA *et al.* (2014) The genome of *Eucalyptus grandis*. *Nature*, **510**, 356–362.

Petroli CD, Sansaloni CP, Carling J *et al.* (2012) Genomic characterization of DArT markers based on high-density linkage analysis and physical mapping to the *Eucalyptus* genome. *PLoS One*, **7**, e44684.

Pryor LD, Johnson LAS (1981) *Eucalyptus*, the universal Australian. In: *Ecological Biogeography of Australia* (ed. Keast A), pp. 499–536. W. Junk, The Hague.

Rezende GDSP, Resende MDV, Assis TF (2014) *Eucalyptus* breeding for clonal forestry. In: *Challenges and Opportunities for the World's Forests in the 21st Century* (ed. Fenning T), pp. 393–424. Springer Science+Business Media, Dordrecht.

Sansaloni CP, Petroli CD, Carling J *et al.* (2010) A high-density Diversity Arrays Technology (DArT) microarray for genome-wide genotyping in *Eucalyptus*. *Plant Methods*, **6**, 16.

Sansaloni C, Petroli C, Jaccoud D *et al.* (2011) Diversity Arrays Technology (DArT) and next-generation sequencing combined: genome-wide, high throughput, highly informative genotyping for molecular breeding of *Eucalyptus*. *BMC Proceedings*, **5**, P54.

Sawler J, Reisch B, Aradhya MK *et al.* (2013) Genomics assisted ancestry deconvolution in grape. *PLoS One*, **8**, e80791.

Shepherd M, Kasem S, Lee D, Henry R (2006) Construction of microsatellite linkage maps for Corymbia. *Silvae Genetica*, **55**, 228–238.

Slavov GT, Howe GT, Gyaourova AV, Birkes DS, Adams WT (2005) Estimating pollen flow using SSR markers and paternity exclusion: accounting for mistyping. *Molecular Ecology*, **14**, 3109–3121.

Steane DA, Vaillancourt RE, Russell J *et al.* (2001) Development and characterisation of microsatellite loci in *Eucalyptus globulus* (Myrtaceae). *Silvae Genetica*, **50**, 89–91.

Van Ooijen JW, Voorrips RE (2001) *JoinMap 3.0 Software for the Calculation of Genetic Linkage Maps*. Plant Research International, Wageningen, The Netherlands.

Voorrips RE (2002) MapChart: software for the graphical presentation of linkage maps and QTLs. *Journal of Heredity*, **93**, 77–78.

Zhou CP, He XD, Li FG *et al.* (2014) Development of 240 novel EST-SSRs in *Eucalyptus* L'Herit. *Molecular Breeding*, **33**, 221–225.

D.G. conceived the development of this microsatellite resource. E.M.C.M. and D.A.F. performed the experiments. O.B.S. Jr. performed bioinformatics analyses. D.A.F., E.M.C.M. and D.G. analysed the microsatellite data. and D.G. wrote the manuscript.

Data accessibility

GenBank Accession nos for all microsatellites are provided in Table S1 (Supporting information). All raw microsatellite genotype data are provided as Table S8 (Supporting information). Genetic linkage positions in centiMorgans and physical mapping coordinates in base pairs in the *Eucalyptus grandis* reference genome are provided in Supporting information Data S1.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. Basic properties of the microsatellite markers developed in the study.

Table S2. Annotation results for the 494 EST-derived microsatellites by BLASTX against the *Eucalyptus grandis* reference genome version 1.1.

Table S3. Descriptive population statistics of 182 newly developed EST microsatellites estimated from population samples of two flagship *Eucalyptus* species, *E. grandis* and *E. globulus*.

Table S4. Descriptive population statistics of 41 newly developed genomic microsatellites estimated from population samples of two flagship *Eucalyptus* species, *E. grandis* and *E. globulus*.

Table S5. Summary for the two single pedigree linkage maps and the consensus linkage map built with data from the two pedigrees plus the previously published reference map of Brondani *et al.* (2006).

Table S6. Microsatellite markers physically mapped to a different chromosome of the *Eucalyptus* reference genome than the expected one based on linkage group number, or physically unmapped.

Table S7. Microsatellite markers that mapped genetically and physically to the 11 main chromosomes and physically to one of the still unanchored scaffolds of the *Eucalyptus* genome assembly.

Table S8. Raw microsatellite genotype data for the 223 microsatellites characterized in the two population samples of *Eucalyptus grandis* (GRA) and *E. globulus* (GLO).

Data S1. Mapchart data file of the linkage map and its alignment to *Eucalyptus* genome sequence.