Integration of EST-CAPS markers into genetic maps of *Eucalyptus urophylla* and *E. tereticornis* and their alignment with *E. grandis* genome sequence

By X. YU[1),2)], Y. GUO[2)], X. ZHANG[2)], F. LI[2)], Q. WENG[2)], M. LI[2)] and S. GAN[1),2),*)]

(Received 22th March 2012)

Abstract

A suite of 91 expressed sequence tag (EST) derived cleaved amplified polymorphic sequence (CAPS) markers were developed and used for enriching the genetic maps of *Eucalyptus urophylla* and *E. tereticornis* built previously based on random amplified polymorphic DNA (RAPD) markers. The EST-CAPS markers were highly similar to original ESTs, with sequence identity ranging from 92.5% to 100.0%. In linkage analysis, 48 and 42 EST-CAPSs were integrated into the genetic maps of *E. urophylla* and *E. tereticornis*, respectively, including 13 shared by both maps, while 14 were unmapped. For *E. urophylla*, the final map had a total length of 1789.5 cM and a mean interval between markers of 9.7 cM, being 284.9 cM larger and 1.3 cM less than those of the prior RAPD map, respectively. For *E. tereticornis*, the final map had a length of 1488.1 cM and a mean interval of 10.3 cM, being 452.4 and 0.2 cM more than the prior map, respectively. All the 77 newly mapped EST-CAPSs found each at least one homologue in the *E. grandis* genome sequence released recently, and conserved synteny and colinearity were observed between *E. grandis* genome and our linkage groups. The enriched maps would provide a set of useful markers for genome analysis, comparative mapping and fine-mapping of important genes located in conserved regions for the important tree genus *Eucalyptus*.

Key words: Expressed sequence tag (EST), cleaved amplified polymorphic sequence (CAPS), genetic map, synteny, colinearity, *Eucalyptus*.

Introduction

The availability of a wealth of expressed sequence tags (ESTs) for many species, in publicly accessible databases in particular, has facilitated the development of sequence tagged site (STS) markers, such as EST-derived simple sequence repeats (EST-SSR), insertion/deletion (EST-indel) and single nucleotide polymorphism (EST-SNP). Besides the advantages of STSs in high reproducibility, specific genomic locality and co-dominant nature, EST-based markers represent genic regions of the genome and are likely to be conserved and transferable across taxa (VARSHNEY et al., 2005; ELLIS and BURKE, 2007), thus providing great potential for comparative mapping, direct gene tagging of quantitative trait loci (QTLs) and functional diversity studies. To date, a large number of EST-based STS markers have been developed in plants, e.g. EST-SSRs and EST-indels in sunflower (*Helianthus annuus* L.; HEESACKER et al., 2008) as well as EST-SNPs in apple (*Malus* × *domestica* Borkh.; CHAGNÉ et al., 2008).

SNPs are the most abundant type of genetic polymorphisms in genomes of most organisms and constitute the commonest class of DNA-based molecular markers for a wide range of applications (CHO et al., 1999; SHERRY et al., 2001; RAFALSKI, 2002; KIM and MISRA, 2007; SLATE et al., 2009). Their biallelism makes them easy to be genotyped with cost- and throughput-flexible assays (JONES et al., 2009), including cleaved amplified polymorphic sequence (CAPS; also known as polymerase-chain-reaction (PCR) based restriction fragment length polymorphism (RFLP), PCR-RFLP). CAPS utilizes restriction enzymes to digest DNA fragments amplified by PCR and displays the polymorphic pattern at one or more restriction sites through agarose gel electrophoresis (KONIECZNY and AUSUBEL, 1993). Though the CAPS method has a limited throughput and is applicable to a limited number of SNPs (KIM and MISRA, 2007), it is one of the most affordable SNP genotyping approaches for common laboratories due to technical simplicity and cost effectiveness (VARSHNEY et al., 2007). CAPS has shown to be capable of detecting polymorphism of PCR products and thus useful for a wide spectrum of studies in plants, such as genetic map construction (e.g. GUJARIA et al., 2011) and candidate gene tagging (e.g. STANIASZEK et al., 2007).

Trees of the genus *Eucalyptus* L'Hérit (Myrtaceae) constitute the most widely planted hardwoods around the world and have a wide range of industrial applications (FAO, 2000). They are usually diploid (2n = 22) with a relatively small genome size (370–700 Mb/C; GRATTAPAGLIA and BRADSHAW, 1994). Thus far, many efforts have been made to develop genetic and genomic resources for the taxon, including genetic linkage maps (e.g. GAN et al., 2003; BRONDANI et al., 2006; NEVES et al., 2011), EST collections (e.g. KELLER et al., 2009) and whole genome sequencing (http://www.phytozome.net/eucalyptus.php). As genetic maps are concerned, functionally anonymous markers such as random amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), RFLP and SSR predominate in

[1)] State Key Laboratory of Tree Genetics and Breeding, Chinese Academy of Forestry, Xiangshan Road, Beijing 100091, People's Republic of China.

[2)] Research Institute of Tropical Forestry, Chinese Academy of Forestry, Longdong, Guangzhou 510520, People's Republic of China.

[*)] Corresponding author: SIMING GAN. Tel: +86 20 87032402, Fax: +86 20 87031622. E-Mail: Siming.Gan@ritf.ac.cn

DOI:10.1515/sg-2012-0031
edited by Thünen Institute of Forest Genetics

the earlier work with a relatively low density. Most recently, however, functional markers based on high-throughput hybridization assays have been incorporated into high-density linkage map construction in *Eucalyptus,* namely, single feature polymorphism (SFP) for *E. grandis* Hill ex Maid. (NEVES et al., 2011) and diversity array technology (DArT) for *E. grandis*, *E. urophylla* S. T. Blake and *E. globulus* Labill. (HUDSON et al., 2012; KULLAN et al., 2012; PETROLI et al., 2012). As such, the addition of functional markers would be helpful for not only increasing map density and coverage but also creating comparative maps. Moreover, with the recent release of *E. grandis* whole genome sequence (http://www.phytozome.net/eucalyptus.php), sequence-tagged functional markers would provide anchor loci to physical maps and whole genome sequences in a map-based gene cloning program.

Here we present the development and sequence validation of 91 EST-CAPS markers in *Eucalyptus* as well as their integration into previously built linkage maps of *E. urophylla* and *E. tereticornis* Smith. We then use the mapped markers to evaluate the extent of genetic/physical synteny and colinearity between the model *E. grandis* genome sequence and our maps. The objectives were to improve the coverage of the previous maps and provide highly transferable markers for utility in comparative mapping among *Eucalyptus* species and physical anchoring of genomic regions of interest.

Materials and Methods

EST sources and primer design

A set of 3134 *Eucalyptus* EST accessions were downloaded from dbEST of GenBank (as of 19 December 2005; http://www.ncbi.nlm.nih.gov/dbEST/) and assembled using PHRAP (http://bozeman.mbt.washington.edu/phrap.docs/phrap.html). All the unigenes were subject to primer design with Primer 3 (ROZEN and SKALETSKY, 2000) following the criteria: primer length 18–22 bp (optimal 20 bp), T_m 55–65 °C (optimal 60 °C, delta T_m < 1 °C), GC content 40–60%, target PCR product 100–500 bp and low probability of dimer or hair-pin loop formation. Primer sets were synthesized by Invitrogen Co. (Shanghai, China).

Plant material, PCR amplification and CAPS genotyping

The mapping population of *E. urophylla* (P_1, UX-30) × *E. tereticornis* (P_2, T4305) was as established previously (82 sibs; GAN et al., 2003). PCR reaction of 10 µL consisted of 1.0 µL 10× buffer (100 mM Tris-HCl pH9.0, 100 mM KCl, 80 mM $(NH_4)_2SO_4$ and 0.5% NP-40), 200 µM each dNTP, 2.5 mM $MgCl_2$ (except for 3.25 mM in *EUCeC86*), 0.5 µM forward primer, 0.5 µM reverse primer, 1 U *Taq* DNA polymerase (Biocolors Technology Co., Shanghai, China) and about 5 ng DNA template. The reaction was amplified in 96-well plates on DNA Engine thermal cycler (Bio-Rad, Hercules, CA, USA) under the following program: 94 °C for 4 min; 35 cycles of 94 °C for 30 s, 56 °C (except for 60 °C in *EUCeC56* and *EUCeC86*) for 30 s, and 72 °C for 1 min; and a final extension at 72 °C for 10 min. The PCR products were checked through electrophoresis in 1.2% agarose gels containing 1 : 20 GoldView (SBS Genetech Co., Beijing, China), and only those each with obviously single amplicon were used for subsequent CAPS experiment.

Twelve enzymes were employed for CAPS restriction digestion, namely, *Hae*III, *Hinf*I, *Msp*I, *Mva*I, *Dra*I, *Pvu*II, *Taq*I, *Rsa*I, *BamH*I, *Pst*I, *Xba*I and *Xho*I (Fermentas International Inc., Burlington, Canada). The digestion reaction (10 µL) consisted of 5.0 µL PCR products, 1.0 µL 10× buffer and 2.5 U enzyme. The digestion was performed at the supplier's recommended temperature for 2–3 h. Digested products were detected by 2.5% agarose gel electrophoresis and sized by comparison to a 100 bp ladder standard (Fermentas International Inc.).

EST re-sequencing and sequence analysis

Re-sequencing was conducted for those ESTs against P_1 and/or P_2 from which CAPS markers were generated. PCR products were directly sequenced using BigDye Terminator Version 3.1 (BDT3.1) on ABI 3130xl genetic analyzer (Applied Biosystems, Foster City, CA) following an optimized protocol (ZHANG et al., 2009). When necessary, PCR products were cloned into pGEM-T Easy vector (Promega Co., Madison, WI, USA), transformed with DH5α (*Escherichia coli* (Migula) Castellani et Chalmers) and sequenced.

The nucleotide sequences generated in re-sequencing were aligned with the original ESTs using DNAMAN version 5.2.2 (Lynnon Biosoft, Quebec, Canada). Within-individual SNPs and correspondence to a specific restriction enzyme were investigated manually. Functional annotation for the ESTs was obtained by BlastX searches against NCBI database of non-redundant protein sequences with a cutoff expectation (*E*) value of 10^{-5}.

Linkage analysis and comparison with E. grandis genome

The EST-CAPS markers were tested for Mendelian inheritance using goodness-of-fit χ^2 test ($P < 0.05$) in SAS/STAT® version 8.0 (SAS institute Inc., Cary, NC) and combined with the previous set of RAPD (GAN et al., 2003) and SSR markers (LI and GAN, 2011). The linkage analysis was performed with Mapmaker 3.0b (LINCOLN et al., 1992) following the same procedure as per LI and GAN (2011), and the designation and orientation of each linkage group established by GAN et al. (2003) were maintained for consistency.

The original EST sequences of the mapped CAPS markers were then aligned in software DNAMAN 5.2.2 (Lynnon Biosoft) with the unigenes mapped as SFP markers in *E. grandis* full map (NEVES et al., 2011). Also, the new genetic maps were compared with the recently released genome sequence of *E. grandis* by aligning marker sequences with BlastN at an *E*-value of 10^{-20} and the least sequence identity of 90.0%, and the results were displayed using GridMap 3.0a program (http://cbr.jic.ac.uk/dicks/software/Grid_Map/). When a marker provided multiple matches, only the best hit with an *E*-value less than one tenth of the next hit was assigned.

DOI:10.1515/sg-2012-0031
edited by Thünen Institute of Forest Genetics

Results and Discussion

Primer design, PCR amplification and EST-CAPS marker development

The 3134 *Eucalyptus* ESTs were assembled into 2404 unigenes, including 227 contigs and 2177 singletons. A total of 1981 primer pairs were successfully designed and synthesized based on the unigenes while primer design was not available for the rest ESTs because of too short or inappropriate sequences.

Effective PCR amplification of an apparently single-locus amplicon was revealed for 1334 (67.3%) out of the 1981 primer pairs synthesized, whereas the remaining 647 (32.7%) failed in amplification, leading to no or very weak product or more than one fragments. PCR products from genomic DNA ranged from 100 to 3500 bp in size (data not shown), averaging at 586 bp. Totally 743 (55.7%) of the effective amplicons were measurably larger (approximately 51–3164 bp) than the corresponding EST length expected, indicating the presence of one or more introns.

Digestion of the effective amplicons resulted in production of a total of 91 (6.8%) EST-CAPS markers, each of which represented at least one fragment being polymorphic between the two parents P_1 and P_2 and segregating among the sibs (*Table 1*). Of the 91 markers, 38 and 38 showed polymorphic alleles unique for P_1 and P_2, respectively, while 15 showed polymorphic alleles for both parents by digestion with one or two restriction enzymes. The 91 EST-CAPS markers were deposited in Probe database of GenBank with IDs 12325920–12326010 (*Table 1*; http://www.ncbi.nlm.nih.gov/probe). Goodness-of-fit χ^2 test indicated that one and three markers showed segregation distortion in P_1 and P_2 ($P \leq 0.05$; *Table 1*), respectively, and were excluded in the subsequent linkage analysis except for *EUCeC6* in P_2 and *EUCeC91* in P_1, which segregated normally in the alternative parent.

The frequency of ESTs converted to CAPS markers is noticeably lower than those previously reported, e.g. 31% in outcrossing conifer suji (*Cryptomeria japonica* D. Don, 32 enzymes; IWATA et al., 2001) and 19.0% (53/279) in self-pollinating legume chickpea (*Cicer arietinum* L., 26 enzymes for an inter-specific mapping population; GUJARIA et al., 2011). This discrepancy could be attributed mainly to the relatively small number of enzymes used herein besides the parental genetic heterozygosity *per se*, implying the great potential of marker development with application of additional restriction enzymes.

Integration of EST-CAPS makers into genetic maps

In linkage analysis, 48 and 42 EST-CAPS markers were incorporated into the prior genetic maps of *E. urophylla* and *E. tereticornis*, respectively (GAN et al., 2003; LI and GAN, 2011), including 13 shared by both maps (*Figure 1*), while 14 (15.4%) could not be linked. For *E. urophylla*, the 48 mapped markers were assigned to 17 previous linkage groups (Eu_LG1–Eu_LG21 except Eu_LG4, Eu_LG13, Eu_LG17 and Eu_LG19), each of which included 1–7 new markers. The final map of *E. urophylla* had a total length of 1789.5 cM and a mean interval between markers of 9.7 cM, which were 284.9 cM larger and 1.3 cM less than those of the RAPD based map (GAN et al., 2003), respectively. For *E. tereticornis*, the 42 markers were allocated to 17 previous linkage groups (Et_LG01–Eu_LG21 except Et_LG10, Et_LG11, Et_LG17 and Et_LG19), each containing 1–6 new markers. The final map of *E. tereticornis* had a length of 1488.1 cM and a mean interval of 10.3 cM, being 452.4 and 0.2 cM more than those of the RAPD map (GAN et al., 2003), respectively.

Together, 37 and 27 EST-CAPS markers were interspersed among the RAPD markers of prior P_1 and P_2 maps, respectively, which would be helpful for reducing the magnitude of the gaps and thus saturating the maps. Meanwhile, one to four EST-CAPS markers were added to the ends of 13 linkage groups (Eu_LG2, Eu_LG12, Eu_LG14, Eu_LG16, Eu_LG20, Eu_LG21, Et_LG7, Et_LG9, Et_LG13, Et_LG15, Et_LG16, Et_LG18 and Et_LG20), extending the length by 1.7 (Eu_LG16) to 72.5 cM (Eu_LG12) each as well as 154.0 and 149.1 cM in total for P_1 and P_2, respectively. Specifically, both ends of Eu_LG12 were extended by a total of 85.8 cM, a substantial increase over the previous group that had been composed of only nine markers covering 53.7 cM. The novel end regions of the genome are shown with grey bars in *Figure 1*. Nevertheless, both maps can not be considered saturated yet with the addition of EST-CAPS markers as the number of linkage groups was still far beyond the cytogenetic chromosome number 11 and many gaps between adjacent markers remained, e.g. nine gaps greater than 30 cM across seven linkage groups (Eu_LG2, Eu_LG5, Eu_LG6, Eu_LG12, Et_LG7, Et_LG8 and Et_LG12).

The 13 EST-CAPS markers in common for the two parental maps help to identify the homologous linkage groups. Seven homologous cases were found between eight linkage groups of P_1 and nine groups of P_2, including three linkage group pairs of one parent homologous to single linkage groups of the alternative parent (Eu_LG2 and Eu_LG10 vs. Et_LG2, Et_LG14 and Et_LG16 vs. Eu_LG9 and Et_LG6 and Et_LG13 vs. Eu_LG12). This will allow attempts to merge the homologous groups. On the other hand, rearrangements with respect to the loci shared by both parental maps were observed in three regions, namely, *EUCeC55–EUCeC57*, *EUCeC63–EUCeC69* and *EUCeC66–EUCeC28*, where two markers linked closely in one parent were separated into independent linkage groups in the alternative parent or vice versa. However, these rearrangements may represent real chromosomal translocation or duplication or be caused by mapping artifact, which need further investigations.

EST-CAPS markers have proven to be efficient in genetic mapping in perennial conifers (HARRY et al., 1998; IWATA et al., 2001) and annual plants (MIURA et al., 2007; GUJARIA et al., 2011). However, their usefulness has not been explored in broad-leaved trees yet. This study represents a valuable attempt in this respect with an important broad-leaved genus *Eucalyptus* and corroborates the advantages of the marker technology. The genetic map of *E. tereticornis* is the first that contained transcript-based genic markers, though a high-density transcript map of *E. urophylla* has been report-

DOI:10.1515/sg-2012-0031
edited by Thünen Institute of Forest Genetics

Table 1. – The 91 EST-CAPS markers developed in this study and their functional annotations. P_1, *E. urophylla*; P_2, *E. tereticornis*.

EST-CAPS marker	GenBank EST accession (ProbeDB ID)	Forward primer (5'–3') Reverse primer (5'–3')	Expected size (bp)	Observed PCR size (bp)	Sequence identity with EST	Enzyme	Restriction fragments for P_1 and P_2 (bp)[a]	BlastX for functional annotation at $E \leq 10^{-5}$ [Organism]	*E*-value
EUCeC1	GQ916947 (12325920)	F:GAAGGATGAGGTTGCCAGTG R:AGGCCAACTTTGCTCTTAGGT	203	950	97.7% (84/86)	*Dra*I	P_1:950,**560**,390 P_2:950	4-coumarate:CoA ligase [*Eucalyptus camaldulensis* Dehnh.]	0.0
EUCeC2	ES594550 (12325931)	F:CGCCTACCTCAACCTCCT R:CCGACCCATTGTCATAAA	374	1300	100.0% (227/227)	*Msp*I	P_1:**1300**,970,330 P_2:970,330	60S acidic ribosomal protein PO [*Euphorbia esula* L.]	1e-114
EUCeC3	ES589804 (12325942)	F:GGCTATGGCACCAACACT R:CCCTTGATGGAAACGGAAT	398	570	99.2% (349/352)	*Pvu*II	P_1:570,**310**,260 P_2:570	Alpha-expansin 2 [*Gossypium hirsutum* L.]	2e-139
EUCeC4	CU400578 (12325953)	F:AGCCTCAACACCACCCAC R:GTCCTTGTCCGACTCACC	290	900	98.3% (285/290)	*Msp*I	P_1:900,**700**,200 P_2:900	Catalytic/coenzyme binding protein [*Arabidopsis lyrata* subsp. *lyrata* (L.) O'Kane & Al-Shehbaz]	3e-110
EUCeC5	CD669683 (12325964)	F:GTCTTCAACAACGGGATG R:GCTTTCGGAACAGACTTTA	255	420	98.4% (249/253)	*Mva*I	P_1:420 P_2:420,**300**,120	Flavonoid 3'-O-methyltransferase [*Mentha* x *piperita* L.]	1e-30
EUCeC6	DQ227993 (12325975)	F:ACTTCACTCTTCCTGGACTC R:CTTCCTCCTGTCACCTCC	344	460	93.6% (322/344)	*Msp*I *Mva*I	P_1:460,**380**,80 P_2:460 P_1:220,160,80 P_2:**380***,220,160,80	Sucrose synthase [*Eucalyptus grandis* W. Hill ex Maiden]	0.0
EUCeC7	ES593283, GO248243 (12325986)	F:ACCTGCTTTGACAGTTTATG R:AATTCGTGTTCAGCCTTT	300	1300	99.2% (251/253)	*Taq*I	P_1:600,290,260,150 P_2:**1300**,600,290,260,150	UDP-glucose 4-epimerase [*Solanum tuberosum* L.]	4e-33
EUCeC8	EF392836 (12325997)	F:GCCTATTACGACCCAGCAG R:CGAACAAAGGACAAAAGGAGAT	257	450	98.7% (226/229)	*Hinf*I	P_1:450 P_2:450,**370**,80	RAC-like small GTPase [*Eucalyptus gunnii* Hook. f.]	1e-124
EUCeC9	ES589562 (12326008)	F:TGGACTACAAGCACAACATT R:CACAACCACTAACGCACC	350	630	100.0% (296/296)	*Msp*I	P_1:630,**350**,280 P_2:630	Elongation factor EF-2 [*Medicago truncatula* L.]	3e-107
EUCeC10	CD669367 (12325921)	F:GAGTCAAATCCTTTCCAGC R:TAATAAGCCCCTCGTCCT	252	1050	100.0% (164/164)	*Rsa*I	P_1:**1050**,730,320 P_2:730,320	Cysteine synthase [*Vitis pseudoreticulata* W. T. Wang]	1e-33
EUCeC11	CD669354 (12325922)	F:TTTCGCCAAGGTCAAGTT R:TATTTCTCCCGCCACTCA	236	1300	100.0% (58/58)	*Msp*I, *Dra*I	P_1:**1300**,1100,200 P_2:1100,**670**,430,200	CBL-interacting serine/ threonine-protein kinase 24 [*Arabidopsis thaliana* (L.) Heynh.]	3e-19
EUCeC12	CD669252 (12325923)	F:GTCTTCTTCGGTCGTTAGG R:TTCCCATGAATCCACTTTCT	296	1650	98.5% (262/266)	*Msp*I, *Mva*I	P_1:1250,**950**,300,200,100 P_2:**1350**,1250,200,100	Protein transport protein SEC61 gamma subunit [*Zea mays* L.]	5e-26
EUCeC13	CD669194 (12325924)	F:CCACTGCCTTCTTACGGT R:TAGGGATAACTTTGTGACTTGA	219	540	100.0% (217/217)	*Hae*III	P_1:500,40 P_2:540,**410**,90,40	Polygalacturonase, putative [*Ricinus communis* L.]	4e-55
EUCeC14	CD669162 (12325925)	F:AAGTCCAACCACCGATTA R:CCTCGCATTGCATACTCT	181	610	100.0% (181/181)	*Taq*I	P_1:**520**,320,200,90 P_2:**410**,320,200,90	ATP binding protein, putative [*Ricinus communis*]	1e-61
EUCeC15	CD669144 (12325926)	F:AGACACTGACTTCTGGGATT R:CAATGGTTCGGATGTTAA	170	170	98.8% (168/170)	*Taq*I	P_1:**170**,110,60 P_2:110,60	Tir-nbs-lrr resistance protein [*Populus trichocarpa* Torr. & Gray]	9e-20
EUCeC16	CD669096 (12325927)	F:TCCAGAGGAGGAACAGGC R:AGGACGACCCACGAAGAT	350	460	99.0% (300/303)	*Hae*III	P_1:**340**,230,120 P_2:230,120,110	Sugar transport protein 13 [*Arabidopsis thaliana*]	4e-45
EUCeC17	CD669087 (12325928)	F:GTTGGCTGCGATTACTGT R:TTGGGTCTGAACTTGTGC	170	800	99.3% (133/134)	*Dra*I	P_1:690,110 P_2:690,**400**,290,110	E3 ubiquitin-protein ligase PRT1 [*Medicago truncatula*]	3e-33
EUCeC18	CD669071 (12325929)	F:GCTACGGCTTCGTGCTCT R:CTCGCTATGAAACATACATCCT	194	620	99.0% (192/194)	*Msp*I	P_1:620,**330**,290 P_2:620	Stem-specific protein TSJT1, putative [*Ricinus communis*]	3e-80
EUCeC19	CD669039 (12325930)	F:TGAGCGGTTCGTTTCTTC R:TTCGCCTTGTCAGGGATA	269	1240	92.5% (161/174)	*Msp*I	P_1:**1240**,1000,160,40 P_2:1000,160,40	Heat shock protein 90 [*Camellia sinensis* (L.) Kuntze]	2e-75
EUCeC20	CD669014 (12325932)	F:TGCCTTGACATCCTGAAA R:CTGGTCCATTCCTTTGCT	176	700	100.0% (160/160)	*Mva*I	P_1:700 P_2:700,**480**,220	Ubiquitin-conjugating enzyme E2 36 [*Arabidopsis thaliana*]	8e-51
EUCeC21	CD669011 (12325933)	F:TCAATCATCATCACCTCGTC R:AATCCAGATACTTGTTACCCTC	293	1650	98.1% (158/161)	*Rsa*I	P_1:**1230**,820,240,210,200,180 P_2:820,240,210,200,180	Ornithine aminotransferase [*Vitis vinifera* L.]	2e-18
EUCeC22	CD668992 (12325934)	F:TCGTGCCTCTGCTTCGTT R:GATCGCCACCGCTGCTTT	289	820	99.7% (287/288)	*Pvu*II	P_1:470,350 P_2:**820**,470,350	STRUBBELIG-receptor family 8 protein [*Arabidopsis thaliana*]	8e-17
EUCeC23	CD668984 (12325935)	F:GCCAACCAACGCATCGCAAGA R:TATCAGCGGTGACGCCCAGGAG	246	550	100.0% (245/245)	*Msp*I	P_1:400,150 P_2:400,**380**,150,20	Malate dehydrogenase [*Stevia rebaudiana* Burtoni]	8e-23
EUCeC24	CD668935 (12325936)	F:GGCAGCGGTTCCTCAAAT R:TTCAAGCCCACGGCAAGT	274	700	98.2% (266/271)	*Mva*I	P_1:620,**330**,290,80 P_2:620,80	Aminotransferase family protein [*Populus trichocarpa*]	9e-52
EUCeC25	CD668843 (12325937)	F:ACTACTTGGTGCTTCACA R:CATAGCCTCCTCTACTCC	290	680	98.8% (248/251)	*Dra*I	P_1:680,**560**,120 P_2:680	Ubiquitin-specific protease 14 [*Arabidopsis thaliana*]	2e-51
EUCeC26	CD668764 (12325938)	F:CTGTTGGAGCTGGTCTTC R:GAGCATGTTTGCAGGAGT	263	820	100.0% (206/206)	*Hae*III	P_1:**770**,570,200,50 P_2:570,200,50	Calcium homeostasis regulator CHoR1 [*Solanum tuberosum*]	1e-39
EUCeC27	CD668705 (12325939)	F:GTATCTGCTTCGCCTTCG R:CGAGCACCTTCCCTTTAC	302	1020	100.0% (82/82)	*Hinf*I	P_1:1020 P_2:1020,**700**,320	Serine/threonine-protein kinase AFC2 [*Arabidopsis thaliana*]	5e-19
EUCeC28	CD668668 (12325940)	F:ATCCAAACCTCGTGAAAT R:TCCCAAAGTCCGATAACT	286	550	96.8% (241/249)	*Dra*I	P_1:440,**380**,110,60 P_2:**550**,440,110	Serine/threonine-protein kinase [*Dimocarpus longan* Lour.]	1e-52
EUCeC29	CD668646 (12325941)	F:ATCAGGGAGTCCGACAATG R:AAGCGTTCCAGCAGATTG	338	330	97.0% (292/301)	*Taq*I	P_1:**330**,265,65 P_2:265,65	Polygalacturonase precursor [*Glycine max* (L.) Merr.]	2e-66
EUCeC30	CD668645 (12325943)	F:TACCGTGATTTATTGGAAC R:ATCATCATTTGAAGTGGC	201	1000	100.0% (198/198)	*Msp*I	P_1:1000 P_2:1000,**850**,150	RING finger protein [*Medicago truncatula*]	2e-41
EUCeC31	CD668505 (12325944)	F:AGGCAAGACGCCAAGAA R:GCCGCAGTGTTTTACCGA	224	220	97.7% (173/177)	*Taq*I	P_1:**220**,140,80 P_2:140,80	26S proteasome non-ATPase regulatory subunit, putative [*Ricinus communis*]	5e-45
EUCeC32	CD668425 (12325945)	F:TTTGGACGCTTTGATGATTC R:CCACCGGAGATGTTTGCA	227	490	99.4% (183/184)	*Taq*I	P_1:490,**390**,100 P_2:490	Aquaporin TIP2 [*Malus prunifolia* (Willd.) Borkh.]	2e-27
EUCeC33	CD668409 (12325946)	F:ATCCCGAGGACCACTACT R:GCTCCCAATAACATAACCA	245	1700	100.0% (118/118)	*Taq*I	P_1:**1700**,1300,320,80 P_2:1300,320,80	Actin-related protein 3 [*Nicotiana tabacum* L.]	2e-55
EUCeC34	CD668407 (12325947)	F:TTGCTACCCCAAAGAAATC R:AGACCACCTCACCAAACTG	222	2250	98.3% (170/173)	*Hae*III	P_1:**2250**,1100,920,170,60 P_2:1100,920,170,60	Nuclear acid binding protein, putative [*Ricinus communis*]	1e-33
EUCeC35	CD668394 (12325948)	F:TGAACTTGGCTTGGCTAC R:CAACACCGACAGATGAAAC	185	185	98.6% (136/138)	*Taq*I	P_1:160,25 P_2:**185**,160,25	NADP-dependent malic protein [*Arachis hypogaea* L.]	3e-35

DOI:10.1515/sg-2012-0031
edited by Thünen Institute of Forest Genetics

Table 1. – Continued.

EST-CAPS marker	GenBank EST accession (ProbeDB ID)	Forward primer (5'–3') Reverse primer (5'–3')	Expected size (bp)	Observed PCR size (bp)	Sequence identity with EST	Enzyme	Restriction fragments for P_1 and P_2 (bp)[a]	BlastX for functional annotation at $E \leq 10^{-5}$ [Organism]	*E*-value
EUCeC36	CD668378 (12325949)	F:AGGAGAACCAAAAGTGATACC R:TGTTACTGTGCCCAAAATG	260	650	97.8% (217/222)	*Pst*I	P_1:400,250 P_2:**650**,400,250	Acetyl-CoA synthetase, putative [*Ricinus communis*]	3e-56
EUCeC37	CD668348 (12325950)	F:CCTTTTGGTGCGGATACT R:ACTTCCTCAAGGGTTCACT	190	390	97.4% (184/189)	*Hae*III	P_1:270,120 P_2:**390**,270,120	Actin related protein [*Populus trichocarpa*]	1e-26
EUCeC38	CD668333 (12325951)	F:ACCGATGCCTACACTTGG R:GGTTGGTTTTCCCTGACT	190	870	100.0% (126/126)	*Mva*I	P_1:800,70 P_2:800,**470**,330,70	Glutathione-dependent formaldehyde dehydrogenase [*Populus trichocarpa*]	7e-79
EUCeC39	CD668290 (12325952)	F:GAGAAATCCAAAGGAGGTG R:CAAAGTGTCAGAATCAACGA	301	510	100.0% (277/277)	*Hinf*I	P_1:**420**,310,110,90 P_2:310,110,90	Endoplasmin-like protein [*Medicago truncatula*]	6e-42
EUCeC40	CD670102 (12325954)	F:ACTGTTTCGCATCTGCAT R:CTTGTTGGAAGTTGTGGC	198	1400	100.0% (75/75)	*Dra*I	P_1:1070,330 P_2:**1400**,1070,330	PRL1 protein [*Arabidopsis thaliana*]	1e-41
EUCeC41	CD670070 (12325955)	F:CTCCTCCTTCCGCTTCTC R:TCTTGGGTGGCAGTTGTG	219	1200	99.2% (121/122)	*Pst*I	P_1:800,300,100 P_2:**1100**,800,300,100	Lactoylglutathione lyase [*Medicago truncatula*]	7e-36
EUCeC42	CD670046 (12325956)	F:CCTGAGGGTGATTGGTTC R:CGCTTGAGGAGGTAGTCG	155	670	97.8% (88/90)	*Hae*III	P_1:**670**,400,270 P_2:400,270	Malate dehydrogenase [*Medicago truncatula*]	1e-56
EUCeC43	CD670044 (12325957)	F:AAGGTGATCTTGGACGAG R:CAGTGGCTGACTTGGGAG	240	940	98.8% (161/163)	*Mva*I	P_1:470 P_2:**940**,470	NMDA receptor-regulated protein [*Medicago truncatula*]	1e-47
EUCeC44	CD670042 (12325958)	F:TTTCCTGAGTTGGCTTTC R:GTATCATTTCCACCACATT	282	1050	99.0% (201/203)	*Xba*I	P_1:1050 P_2:1050,**550**,500	Lon protease-like protein [*Medicago truncatula*]	7e-39
EUCeC45	CD670029 (12325959)	F:GGCAGAGGGATCAGTTTCG R:ACTTGACGCCCGCTTGGAG	172	850	100.0% (169/169)	*Mva*I	P_1:850,**630**,220 P_2:850	Ubiquitin-conjugating enzyme E2 [*Solanum nigrum* L.]	3e-11
EUCeC46	CD669912 (12325960)	F:GCTCAACATGGGCATCAG R:TCAAGTCCCAGCAAGTAGAAT	195	440	100.0% (155/155)	*Rsa*I	P_1:**440**,380,60 P_2:380,60	Auxin-induced protein 5NG4 [*Medicago truncatula*]	4e-45
EUCeC47	CD669872 (12325961)	F:CCGTCAACCTCATCTGCG R:TTTGGCGATGCTTTAGTG	213	870	97.2% (173/178)	*Mva*I	P_1:580,270,20 P_2:**870**,580,270,20	26S proteasome non-ATPase regulatory subunit [*Medicago truncatula*]	1e-85
EUCeC48	CD669845 (12325962)	F:CTCGTGCGAAGTGGGTGTC R:GGCGGACTATTTCTCAAGTGTT	241	930	98.9% (177/179)	*Msp*I	P_1:**930**,720,210 P_2:720,210	Translation initiation factor IF6 [*Arabidopsis thaliana*]	1e-36
EUCeC49	CD669833 (12325963)	F:GCGGACGACGACTACAAT R:CGCCCTCTAACTCAACCA	268	1150	98.5% (199/202)	*Msp*I	P_1:1150 P_2:1150,**750**,400	DNA-directed RNA polymerase IIa [*Nicotiana tabacum*]	2e-49
EUCeC50	CD669819 (12325965)	F:ACTCAGCCACCACTCGTT R:GGAAGAATATGAAGCACCC	363	1000	99.2% (235/237)	*Rsa*I	P_1:1000,**800**,200 P_2:1000	Dynein light chain type 1 family protein [*Arabidopsis lyrata* subsp. *lyrata*]	2e-17
EUCeC51	CD669813 (12325966)	F:GTGGAAAGATAAGGCGAGTG R:TAGCGTGTCTGGTAGATGGT	194	280	99.4% (157/158)	*Pst*I	P_1:140 P_2:**280**,140	Small G protein family protein [*Arabidopsis lyrata* subsp. *lyrata*]	1e-43
EUCeC52	CD669787 (12325967)	F:GGGCTTGGCTCTTTCATG R:ATCCCACCCTTGCTTTCC	234	1800	96.5% (194/201)	*Xho*I	P_1:1800 P_2:1800,**900**	UOS1 [*Medicago truncatula*]	3e-15
EUCeC53	CD669786 (12325968)	F:CCCAAATGGCTAACCTCA R:AACCACAGCCGTCCTTCC	218	750	99.4% (173/174)	*Xba*I	P_1:**750**,500,250 P_2:700,500,**200**	Histone-lysine N-methyltransferase ASHR1 [*Medicago truncatula*]	8e-46
EUCeC54	CD669766 (12325969)	F:TGGCGTATGGAGAAGGAT R:GCTTGGTGGTGATGAAAA	176	800	99.4% (171/172)	*Hae*III	P_1:**800** P_2:630,170	NAD(P)-linked oxidoreductase domain-containing protein [*Arabidopsis thaliana*]	5e-36
EUCeC55	CD669738 (12325970)	F:TTGGTAACTGTGGGTCTGG R:GTTTGATTTCCTCCGTTTT	325	540	96.9% (280/289)	*Taq*I	P_1:**330**,270,130,80 P_2:270,**200**,130,80	Auxin/indole-3-acetic acid 3 [*Solanum tuberosum*]	4e-55
EUCeC56	CD669735 (12325971)	F:AGAGGCTGATGAGGGATT R:CACTGTCGGAGGTTTATTT	191	650	97.9% (185/189)	*Taq*I	P_1:530,120 P_2:530,**380**,150,120	Ubiquitin conjugating protein [*Avicennia marina* (Forsk.) Vierh.]	4e-65
EUCeC57	CD669604 (12325972)	F:CACATCAAGAATCAAGGGTA R:AACAACAACGGCACCAAA	289	1170	98.9% (86/87)	*Rsa*I	P_1:**1120**,850,270,50 P_2:850,620,**300**,270,200,50	Cytidine deaminase [*Paspalum simplex* Morong ex Britton]	8e-13
EUCeC58	CD669486 (12325973)	F:TGGCACCGAACAAGAAGT R:GCAACATTGGCTAGTTGATTTA	161	340	100.0% (145/145)	*Taq*I	P_1:200,70 P_2:**270**,200,70	Ring finger and CHY zinc finger domain-containing protein 1 [*Arabidopsis thaliana*]	3e-22
EUCeC59	CD669466 (12325974)	F:CCGCTTCTTTACTCCTCC R:CGTTCAGACATTCCCTCC	167	270	97.8% (131/134)	*Taq*I	P_1:200,**130**,70 P_2:**270**,200,70	Alcohol dehydrogenase [*Citrus* x *paradisi* Macfad.]	3e-29
EUCeC60	CD669443 (12325976)	F:GGTGGAATCAAGATGGAA R:AAGACACCAGGGAAGGAA	187	1450	100.0% (91/91)	*Hae*III	P_1:800,650 P_2:**1450**,800,650	Hypersensitive-induced response protein 1 [*Arabidopsis thaliana*]	4e-56
EUCeC61	CD669421 (12325977)	F:AAGAGCATTGTCGAGGAC R:AACAGTTCAGCAGTAGCG	226	950	98.2% (222/226)	*Mva*I	P_1:**950**,580,370 P_2:580,370	Receptor for activated C kinase 1B [*Arabidopsis thaliana*]	1e-36
EUCeC62	CD668192 (12325978)	F:CTACCTTGGCTGCTCTGA R:TGAACCCTCGTAACTTGTC	231	390	99.0% (191/193)	*Rsa*I	P_1:390,**290**,100 P_2:390	Hydroxypyruvate reductase [*Bruguiera gymnorhiza* (L.) Lamk.]	1e-62
EUCeC63	CD668142 (12325979)	F:CCTCCTTCGCATAAACCC R:CTTCAACAGATTACCCGTCAA	177	750	98.9% (87/88)	*Mva*I, *Msp*I	P_1:**650**,450,200,100 P_2:450,**300**,200,150,100	U1 snRNP-interacting 70 kDa protein [*Triticum aestivum* L.]	2e-19
EUCeC64	CD668141 (12325980)	F:GCTGTAGATTCTTTCGTCG R:CACCCACATTCACATTCC	339	1210	100.0% (251/251)	*Rsa*I	P_1:**930**,700,280,230 P_2:700,280,230	Superoxide dismutase [*Fagus sylvatica* L.]	1e-50
EUCeC65	CD668126 (12325981)	F:TCAACGGTCCCTTGTCTT R:TTGCCTTTGTATGTAATGGTCT	405	1400	98.3% (58/59)	*Msp*I	P_1:900,400,100 P_2:**1400**,900,400,100	Protein regulator of cytokinesis [*Medicago truncatula*]	1e-27
EUCeC66	CD668103 (12325982)	F:CGATTACCGTGTCTTCCT R:TTCCCTTCTTAGTGTCCTG	185	910	99.0% (102/103)	*Mva*I, *Pvu*II	P_1:770,**140**,120,20 P_2:**790**,770,120,20	Inorganic pyrophosphatase [*Zea mays*]	4e-47
EUCeC67	CD668099 (12325983)	F:AGCCCAGTCAAAGTTCAT R:AAGCCTATCATTGCGAGT	373	700	100.0% (334/334)	*Rsa*I	P_1:**700**,520,180 P_2:520,180	Early light-inducible protein [*Populus trichocarpa*]	1e-29
EUCeC68	CD668095 (12325984)	F:CACAGACAATCGCCAGTA R:GGAGTTGATAAGCCTTCG	225	1350	100.0% (106/106)	*Msp*I	P_1:1350 P_2:1350,**1020**,230,100	Enolase-phosphatase E1 [*Arabidopsis thaliana*]	5e-28
EUCeC69	CB968051 (12325985)	F:CGTCGCTGTTAGGGTTAG R:CAGATTGCCTCAATTTCC	189	1150	100.0% (185/185)	*Msp*I	P_1:920,**730**,420,230 P_2:920,420,**280**,230,220	No significant match	–
EUCeC70	CB967984 (12325987)	F:ACATAGAATACGATACCTCCTT R:TGTCAACTTCAGCCAAAT	231	310	100.0% (120/120)	*Taq*I	P_1:310 P_2:310,**280**,30	Peptidylprolyl isomerase [*Arabidopsis lyrata* subsp. *lyrata*]	5e-54
EUCeC71	CB967974 (12325988)	F:AGATGCCACTGTTGAGGT R:ATCAGGACAGGGTATGGA	213	610	99.4% (168/169)	*Taq*I	P_1:570,40 P_2:570,**310**,260,40	Glycerophosphodiesterase-like protein [*Medicago truncatula*]	5e-27

DOI:10.1515/sg-2012-0031
edited by Thünen Institute of Forest Genetics

Table 1. – Continued.

EST-CAPS marker	GenBank EST accession (ProbeDB ID)	Forward primer (5'–3') Reverse primer (5'–3')	Expected size (bp)	Observed PCR size (bp)	Sequence identity with EST	Enzyme	Restriction fragments for P_1 and P_2 (bp)[a]	BlastX for functional annotation at $E \leq 10^{-5}$ [Organism]	*E*-value
EUCeC72	CB967973 (12325989)	F:CTGGAGCTTATGGGTGAA R:CAGGAGTAAATCCCGACT	258	830	99.0% (189/191)	*Msp*I	P_1:830,**620**,210 P_2:830	tRNA-binding arm; t-snare [*Medicago truncatula*]	1e-09
EUCeC73	CB967967 (12325990)	F:TGGATCACAGGCTTGGTA R:TGAAGGCATAGGAGGTTC	176	730	98.9% (174/176)	*Msp*I	P_1:520,210 P_2:**730***,520,210	Serine/threonine-protein kinase, putative [*Ricinus communis*]	2e-64
EUCeC74	CB967873 (12325991)	F:TGGATTTCGGTGACTTGA R:TGAATAGGCAGATGACAGTAG	392	2050	99.4% (343/345)	*Msp*I	P_1:**2050**,1600,450 P_2:1350,450,250	COP9 signalosome complex subunit [*Medicago truncatula*]	1e-58
EUCeC75	CB967775 (12325992)	F:TCCCAATCTGCGATGCTC R:ACCCAAGAAACGGACACT	203	1280	NA	*Taq*I, *Rsa*I	P_1:540,**440**,320,220,200,100 P_2:**760**,540,320,220,200	Glycogen synthase kinase [*Glycine max*]	8e-96
EUCeC76	CB967772 (12325993)	F:TCCACGGTGAAGTTTGCT R:CCTTGAGGCTTTGATTGA	218	850	100.0% (67/67)	*Taq*I	P_1:850 P_2:850,**600***,250	GroES-like protein [*Arachis hypogaea*]	4e-47
EUCeC77	CB967731 (12325994)	F:CATTGCCAAGAAAGTTGT R:AAGCCATTGATACGAGAC	188	900	99.3% (150/151)	*Mva*I	P_1:**900**,580,320,50 P_2:580,320,50	Ubiquitin-specific protease 6 [*Arabidopsis lyrata* subsp. *lyrata*]	1e-69
EUCeC78	CB967728 (12325995)	F:ACGAGTTAGACGCTATGA R:TCCCGAACTATGTATGTG	231	800	99.6% (230/231)	*Mva*I	P_1:800,**400** P_2:800	Cyclic nucleotide gated channel 9 [*Arabidopsis thaliana*]	1e-88
EUCeC79	CB967631 (12325996)	F:CCTCACAACTTCAGCCATTT R:TTCTTCAGCACGCCGATA	168	2450	NA	*Msp*I	P_1:1200,650,350,200,50 P_2:**1400**,1200,650,350,200,50	DET1- and DDB1-associated protein 1 [*Arabidopsis thaliana*]	4e-34
EUCeC80	CB967583 (12325998)	F:AGCGGTATTACGGGTTCG R:TTTGCTCCTGATCTGTGC	158	1300	100.0% (43/43)	*Msp*I	P_1:1300 P_2:1300,**700**,600	Hypoxia-responsive family protein [*Arabidopsis lyrata*]	3e-30
EUCeC81	CB967573 (12325999)	F:ACGGACAGCATTCAAGAT R:ATTTCACTGGCGTTCTCA	228	1580	98.3% (113/115)	*Msp*I	P_1:1580,1050,**530** P_2:1580	Anionic peroxidase [*Petroselinum crispum* (Mill.) Hill]	4e-42
EUCeC82	CB967513 (12326000)	F:AAGGCTGAGCAGATGAGA R:GGAAGGATGGGTTAGTGA	169	520	97.0% (164/169)	*Taq*I	P_1:520 P_2:520,**320**,200	Ubiquitin-associated (UBA)/TS-N domain-containing protein [*Arabidopsis thaliana*]	1e-51
EUCeC83	AJ627837 (12326001)	F:CTGGTTTACACCGTGTTCTC R:TAGTTGGTTGGGTTGCTG	323	680	99.7% (322/323)	*Rsa*I	P_1:**530**,460,80,70 P_2:460,80,70	Aquaporin-like protein [*Petunia* x *hybrida* Vilm.]	8e-81
EUCeC84	AJ627817 (12326002)	F:GCGGGAGAACTTTGAGAA R:CATAGCGGCGGTAGCAGA	223	330	100.0% (223/223)	*Xba*I	P_1:330 P_2:330,**190**,140	NADP-dependent malic enzyme 3 [*Hydrilla verticillata* L. f.]	2e-28
EUCeC85	AJ627805 (12326003)	F:GGGGATAAGCAAAGAGGT R:ACACCAGGTTCAGGCTCT	174	270	99.4% (173/174)	*Taq*I	P_1:220,50 P_2:**270**,220,50	Hsp90 co-chaperone AHA1, putative [*Ricinus communis*]	1e-60
EUCeC86	AJ627739 (12326004)	F:GGCACTTTCAGGCACCAC R:ATCGCATCGTATCGGGAG	252	260	98.8% (249/252)	*Taq*I	P_1:170,60,30 P_2:**260**,170,60,30	Transposon protein CACTA, En/Spm sub-class [*Zea mays*]	4e-17
EUCeC87	CB009738 (12326005)	F:GTCGGAACAACAAGAAGC R:TCGTCGCAGAATTACACC	340	430	99.0% (298/301)	*Pst*I	P_1:330,100 P_2:**430**,330,100	Small nuclear ribonucleoprotein-like protein [*Medicago truncatula*]	2e-40
EUCeC88	CD669233 (12326006)	F:CAACAGGAATAACCAGAAGC R:TGGGTGTTACTTGATCGGCA	236	340	99.5% (198/199)	*Dra*I	P_1:**340**,200,140 P_2:200,140	Hypothetical protein, predicted [*Vitis vinifera*]	5e-06
EUCeC89	CD668837 (12326007)	F:ATGATGCACAAGCGTATGGT R:CTCACAGATCCGCTTAATG	325	500	99.2% (280/282)	*Hae*III	P_1:270,**230**,170,60 P_2:**330**,270,170,60	AGO1-1 [*Nicotiana benthamiana* Domin.]	6e-45
EUCeC90	CD668835 (12326009)	F:GAAACCGCTGTCATGGAAGT R:TTCCCCGTTTTCATCAGG	306	750	100.0% (305/305)	*Msp*I	P_1:750,**490**,240,20 P_2:750	No significant match	–
EUCeC91	CD668626 (12326010)	F:CTCAACCCAAACAAGCAG R:GACCCCAACGCTCTTCATGA	472	480	99.5% (437/439)	*Mva*I	P_1:230,**170***,120,80,50 P_2:**350**,230,120,80,50	Cytochrome P450 [*Medicago truncatula*]	3e-36

[a] The underlined bold number denotes the approximate size of the fragment scored for segregation among the mapping population.
* Segregation distortion at 0.05 significance level (four markers *EUCeC6*, *EUCeC73*, *EUCeC76* and *EUCeC91*).
NA, not available due to re-sequencing of the intron(s) only (two markers *EUCeC75* and *EUCeC79*).

ed most recently (KULLAN et al., 2012). As genic markers tend to be transferable across taxa (VARSHNEY et al., 2005; ELLIS and BURKE, 2007), the EST-CAPS markers mapped herein may be potential anchor loci for comparative genome analysis and QTL mapping in a genus context. Moreover, the technical simplicity, cost efficiency and co-dominance nature of the marker type may help to extend readily the technology to other eucalypts in common laboratories, which will facilitate the rapid accumulation of genome mapping data and strengthen the understanding of genome evolution in *Eucalyptus*.

Sequence validation and comparative analysis with E. grandis genome

Because non-specific priming can lead to unwanted PCR products (CHA and THILLY, 1995; KUNKEL and BEBENEK, 2000), it is necessary to validate the sequence fidelity of the mapped markers to the original ESTs. Sequence alignment revealed that 104 out of the 106 amplicons (76 markers for P_1 or P_2 plus 15 for both parents) were highly similar to the original ESTs (*Table 1*), with sequence identity ranging from 92.5% to 100.0%, whereas the other two (*EUCeC75* with P_1 and *EUCeC79* with P_2) could not be aligned because of sequencing of the intron regions. Also, the restriction sites were identified from the amplicon sequences, with nine exceptions (*EUCeC4*, *EUCeC14*, *EUCeC19*, *EUCeC34*, *EUCeC40*, *EUCeC41*, *EUCeC52*, *EUCeC61* and *EUCeC74*) which produced only one allelic sequence with the digestion site(s) traceable (the former seven) or not (the latter two; genomic sequences available upon request). In addition, BlastX searches against NCBI database of non-redundant protein sequences showed that 88 (96.7%) of these 91 ESTs were homologous to known genes and 1 (1.1%) corresponded to a hypothetical protein at a threshold of *E*-value less than 10^{-5} while only 2 (2.2%) produced no significant match (*Table 1*).

Only four EST markers (*EUCeC34*, *EUCeC54*, *EUCeC69* and *EUCeC80*) were in common with the *E. grandis* reference map (NEVES et al., 2011), with sequence identity more than 98.0% (data not shown). Even though the small number of common markers

DOI:10.1515/sg-2012-0031
edited by Thünen Institute of Forest Genetics

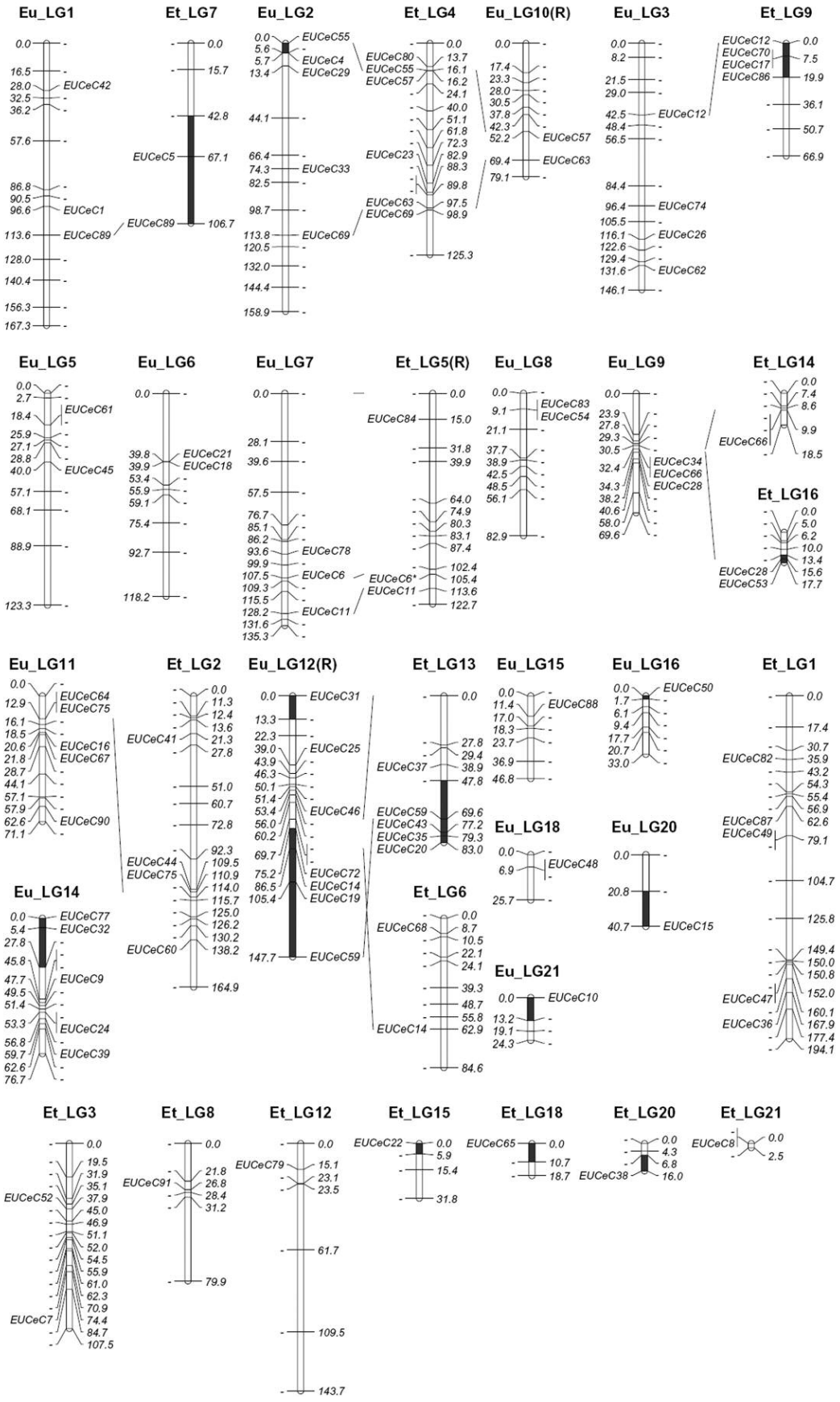

Figure 1. – Integration of EST-CAPS markers into the RAPD based genetic maps of *E. urophylla* and *E. tereticornis* constructed previously (GAN et al., 2003). The cumulative distances in centiMorgans (Kosambi) are given at the left of the bar for *E. urophylla* (Eu) and the right for *E. tereticornis* (Et) linkage groups. Grey bars show the novel end regions of the genome. The asterisk indicates segregation distortion (marker *EUCeC6*; $P < 0.05$).

DOI:10.1515/sg-2012-0031
edited by Thünen Institute of Forest Genetics

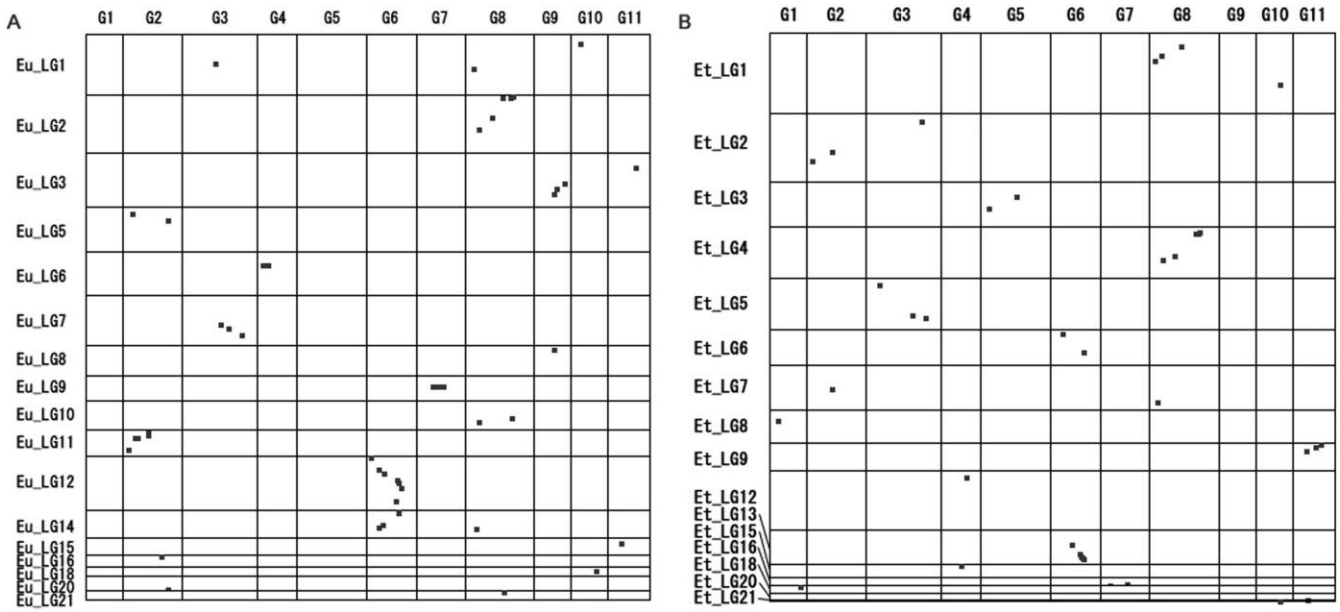

Figure 2. – Syntenic relationship between the linkage groups and *E. grandis* main genome scaffolds. A, *E. urophylla*; B, *E. tereticornis*. Loci showing homology between the two genomes at an *E*-value of 10^{-20} and the least sequence identity 90.0% are indicated by dots.

made detailed comparative analysis difficult, they helped to anchor the homologous linkage groups, namely, our Eu_LG9, Eu_LG2, Et_LG4 and Eu_LG8 to LG7, LG8, LG8 and LG9 of NEVES et al. (2011), respectively.

In BlastN homology searching, all the 77 mapped CAPS markers found each at least one homologous locus in *E. grandis* genome sequence (identity from 90.6% to 100% and *E*-value less than $1e^{-59}$; data not shown), including 76 corresponding to the 11 main scaffolds assembled and one (*EUCeC44*) to the small scaffold 17. Seven markers (*EUCeC36*, *EUCeC54*, *EUCeC66*, *EUCeC67*, *EUCeC69*, *EUCeC78* and *EUCeC81*) had multiple homologues in the *E. grandis* genome. Such markers were deemed unsuitable for comparative mapping (NELSON et al., 2006) and removed from further comparative analysis. In total, there were 45 and 38 correspondences in *E. urophylla* and *E. tereticornis*, respectively, as compared with *E. grandis* genome. *Figure 2* provides a global view of similarity between our linkage groups and *E. grandis* main scaffolds based on the homologous loci.

Extensive syntenic relationship was observed between *E. grandis* genome sequence and either of our linkage maps. For example, the majority of markers from a linkage group of *E. urophylla* or *E. tereticornis* found their matches within the same scalffold of *E. grandis* genome. However, non-syntenic markers were observed in Eu_LG1, Eu_LG3, Eu_LG14, Et_LG1, Et_LG2 and Et_LG7, each of which corresponded to two or three *E. grandis* scalffolds (*Figure 2*). Occurrence of non-synteny was also found between *E. grandis* genome assembly and genetic maps of *E. urophylla*, *E. grandis* and *E. globulus* (HUDSON et al., 2012). Though no sufficient supportive evidence is available, non-synteny between genomes may be resulted from mapping errors or genomic mechanisms, e.g. translocation (transposable element activity) and duplication (NELSON et al., 2010; HUDSON et al., 2012). Moreover, colinearity was also observed between *E. grandis* genome and our maps. In total, as compared to *E. grandis* genome, seven and six colinear blocks sharing the order of at least three loci were identified in *E. urophylla* and *E. tereticornis* genetic maps, respectively, each containing 3–6 loci (data not shown). Such colinearity could form the basis for marker-assisted selection and be used to direct the fine-mapping of important genes located in conserved regions (NELSON et al., 2006).

Acknowledgments

This work was financially supported by grants from the National Natural Science Foundation of China (31070592), the Guangdong Natural Science Foundation (10151052001000000) and the Ministry of Science and Technology of China (2011AA100202 and 2012CB114505). The authors are grateful to KUNMING WU and JUYING WU for valuable cooperation in plant material establishment and maintenance.

References

BRONDANI, R. P. V., E. R. WILLIAMS, C. BRONDANI and D. GRATTAPAGLIA (2006): A microsatellite-based consensus linkage map for species of *Eucalyptus* and a novel set of 230 microsatellite markers for the genus. BMC Plant Biol **6**: 20.

CHA, R. S. and W. G. THILLY (1995): Specificity, efficiency, and fidelity of PCR, pp. 37–52. *In:* PCR primer: a laboratory manual, edited by C. W. DIEFFENBACH and G. S. DVEKSLER, Cold Spring Harbor Press, New York.

CHAGNÉ, D., K. GASIC, R. N. CROWHURST, Y. HAN, H. C. BASSETT, D. R. BOWATTE, T. J. LAWRENCE, E. H. A. RIKKERINK, S. E. GARDINER and S. S. KORBAN (2008): Development of a set of SNP markers present in expressed genes of the apple. Genomics **92**: 353–358.

DOI:10.1515/sg-2012-0031
edited by Thünen Institute of Forest Genetics

CHO, R. J., M. MINDRINOS, D. R. RICHARDS, R. J. SAPOLSKY, M. ANDERSON, E. DRENKARD, J. DEWDNEY, N. FEDERSPIEL, A. THEOLOGIS, W. H. YANG, E. HUBBELL, M. AU, E. Y. CHUNG, D. LASHKARI, B. Lemieux, C. DEAN, R. J. LIPSHUTZ, F. M. AUSUBEL, R. W. DAVIS and P. J. OEFNER (1999): Genome-wide mapping with biallelic markers in *Arabidopsis thaliana*. Nat Genet **23**: 203–207.

ELLIS, J. R. and J. M. BURKE (2007): EST-SSRs as a resource for population genetic analyses. Heredity **99**: 125–132.

FAO (2000): Global forest resources assessment 2000 – main report. FAO forestry paper No. 140. FAO, Rome.

GAN, S., J. SHI, M. LI, K. WU, J. WU and J. BAI (2003): Moderate-density molecular maps of *Eucalyptus urophylla* S. T. Blake and *E. tereticornis* Smith genomes based on RAPD markers. Genetica **118**: 59–67.

GRATTAPAGLIA, D. and H. D. BRADSHAW JR (1994): Nuclear DNA content of commercially important *Eucalyptus* species and hybrids. Can J For Res **24**: 1074–1078.

GUJARIA, N., A. KUMAR, P. DAUTHAL, A. DUBEY, P. HIREMATH, A. B. PRAKASH, A. FARMER, M. BHIDE, T. SHAH, P. M. GAUR, H. D. UPADHYAYA, S. BHATIA, D. R. COOK, G. D. MAY, J. K. VARSHNEY (2011): Development and use of genic molecular markers (GMMs) for construction of a transcript map of chickpea (*Cicer arietinum* L.). Theor Appl Genet **122**: 1577–1589.

HARRY, D. E., B. TEMESGEN and D. B. NEALE (1998): Codominant PCR-based markers for *Pinus taeda* developed from mapped cDNA clones. Theor Appl Genet **97**: 327–336.

HEESACKER, A., V. K. KISHORE, W. GAO, S. TANG, J. M. KOLKMAN, A. GINGLE, M. MATVIENKO, A. KOZIK, R. M. MICHELMORE, Z. LAI, L. H. RIESEBERG and S. J. KNAPP (2008): SSRs and INDELs mined from the sunflower EST database: abundance, polymorphisms, and cross-taxa utility. Theor Appl Genet **117**: 1021–1029.

HUDSON, C. J., A. R. K. KULLAN, J. S. FREEMAN, D. A. FARIA, D. GRATTAPAGLIA, A. KILIAN, A. A. MYBURG, B. M. POTTS and R. E. VAILLANCOURT (2012): High synteny and colinearity among *Eucalyptus* genomes revealed by high-density comparative genetic mapping. Tree Genet Genomes **8**: 339–352.

IWATA, H., T. UJINO-IHARA, K. YOSHIMURA, K. NAGASAKA, Y. MUKAI and Y. TSUMURA (2001): Cleaved amplified polymorphic sequence markers in sugi, *Cryptomeria japonica* D. Don, and their locations on a linkage map. Theor Appl Genet **103**: 881–895.

JONES, N., H. OUGHAM, H. THOMAS and I. PAŠAKINSKIENĖ (2009): Markers and mapping revisited: finding your gene. New Phytol **183**: 935–966.

KELLER, G., T. MARCHAL, H. SANCLEMENTE, M. NAVARRO, N. LADOUCE, P. WINCKER, A. COULOUX, C. TEULIÈRES and C. MARQUE (2009): Development and functional annotation of an 11,303-EST collection from *Eucalyptus* for studies of cold tolerance. Tree Genet Genomes **5**: 317–327.

KIM, S. and A. MISRA (2007): SNP Genotyping: technologies and biomedical applications. Annu Rev Biomed Eng **9**: 7.1–7.32.

KONIECZNY, A. and F. M. AUSUBEL (1993): A procedure for mapping *Arabidopsis* mutations using co-dominant ecotype-specific PCR-based markers. Plant J **4**: 403–410.

KULLAN, A. R. K., M. M. DYK VAN, N. JONES, A. KANZLER, A. BAYLEY and A. A. MYBURG (2012): High-density genetic linkage maps with over 2,400 sequence-anchored DArT markers for genetic dissection in an F_2 pseudo-backcross of *Eucalyptus grandis* × *E. urophylla*. Tree Genet Genomes **8**: 163–175.

KUNKEL, T. A. and K. BEBENIK (2000): DNA replication fidelity. Annu Rev Biochem **69**: 497–529.

LI, F. and S. GAN (2011): An optimised protocol for fluorescent-dUTP based SSR genotyping and its application to genetic mapping in *Eucalyptus*. Silvae Genet **60**: 18–25.

LINCOLN, S., M. DALY and E. LANDER (1992): Constructing genetic maps with MAPMAKER/EXP 3.0. Whitehead Institute technical report, 3rd edition. Whitehead Institute, Cambridge, Massachusetts.

MIURA, Y., M. HIRATA and M. FUJIMORI (2007): Mapping of EST-derived CAPS markers in Italian ryegrass (*Lolium multiflorum* Lam.). Plant Breed **126**: 353–360.

NELSON, M. N., P. M. MOOLHUIJZEN, J. G. BOERSMA, M. CHUDY, K. LESNIEWSKA, M. BELLGARD, R. P. OLIVER, W. ŚWIĘCICKI, B. WOLKO, W. A. COWLING and S. R. ELLWOOD (2010): Aligning a new reference genetic map of *Lupinus angustifolius* with the genome sequence of the model legume, *Lotus japonicus*. DNA Res **17**: 73–83.

NELSON, M. N., H. T. T. PHAN, S. R. ELLWOOD, P. M. MOOLHUIJZEN, J. HANE, A. WILLIAMS, C. E. O'LONE, J. FOSU-NYARKO, M. SCOBIE, M. CAKIR, M. G. K. JONES, M. BELLGARD, M. KSIĄŻKIEWICZ, B. WOLKO, S. J. BARKER, R. P. OLIVER AND W. A. COWLING (2006): The first gene-based map of *Lupinus angustifolius* L. – location of domestication genes and conserved synteny with *Medicago truncatula*. Theor Appl Genet **113**: 225–238.

NEVES, L. G., E. M. C. MAMANI, A. C. ALFENAS, M. KIRST and D. GRATTAPAGLIA (2011): A high-density transcript linkage map with 1,845 expressed genes positioned by microarray-based Single Feature Polymorphisms (SFP) in *Eucalyptus*. BMC Genomics **12**: 189.

PETROLI, C. D., C. P. SANSALONI, J. CARLING, D. A. STEANE, R. E. VAILLANCOURT, A. A. MYBURG, O. B. SILVA, G. J. PAPPAS, A. KILIAN and D. GRATTAPAGLIA (2012): Genomic characterization of DArT markers based on high-density linkage analysis and physical mapping to the *Eucalyptus* genome. PLoS One **7**: e44684.

RAFALSKI, A. (2002): Applications of single nucleotide polymorphisms in crop genetics. Curr Opin Plant Biol **5**: 94–100.

ROZEN, S. and H. J. SKALETSKY (2000): Primer3 on the WWW for general users and for biologist programmers, pp. 365–386. *In:* Bioinformatics methods and protocols: methods in molecular biology, edited by S. KRAWETZ and S. MISENER, Humana Press, Totowa, New Jersey.

SHERRY, S. T., M.-H. WARD, M. KHOLODOV, J. BAKER, L. PHAN, E. M. SMIGIELSKI and K. SIROTKIN (2001): dbSNP: the NCBI database of genetic variation. Nucleic Acids Res **29**: 308–311.

SLATE, J., J. GRATTEN, D. BERALDI, J. STAPLEY, M. HALE and J. M. PEMBERTON (2009): Gene mapping in the wild with SNPs: guidelines and future directions. Genetica **136**: 97–107.

STANIASZEK, M., E. U. KOZIK and W. MARCZEWSKI (2007): A CAPS marker $TAO1_{902}$ diagnostic for the *I-2* gene conferring resistance to *Fusarium oxysporum* f. sp. *lycopersici* race 2 in tomato. Plant Breed **126**: 331–333.

VARSHNEY, R. K., A. GRANER and M. E. SORRELLS (2005): Genic microsatellite markers in plants: features and applications. Trends Biotechnol **23**: 48–55.

VARSHNEY, R. K., S. NAYAK, B. JAYASHREE, K. ESHWAR, H. D. UPADHYAYA and D. A. HOISINGTON (2007): Development of cost-effective SNP assays for chickpea genome analysis and breeding. SAT eJournal **3**: 1–3.

ZHANG, X., F. LI, Y. WANG, L. XU, M. LI and S. GAN (2009): An optimized protocol for sequencing EST-PCR products in *Eucalyptus*. Genomics Appl Biol **28**: 535–543.

DOI:10.1515/sg-2012-0031
edited by Thünen Institute of Forest Genetics