Short note: Development, characterization and cross-amplification of eight EST-derived microsatellites in *Salix*

By X. He[1),*], J. Zheng[1)], M. Serapiglia[2)], L. Smart[2)], S. Shi[1)] and B. Wang[1)]

(Received 4th November 2013)

Abstract

A set of eight simple sequence repeat (SSR) markers were developed from 707 *Salix* expressed sequence tags (ESTs) deposited in GenBank. Each of the EST-SSR amplicons was identical to the original EST, with sequence identity 60.90–96.03% and presence of the expected repeat motifs. Of the eight EST-SSR loci, five were polymorphic among 14 individuals of *S. eriocephala*, with the number of alleles per locus (N_a), observed heterozygosity (H_o), expected heterozygosity (H_e) and polymorphic information content (*PIC*) being 2–7 (mean 4.8), 0.29–0.85 (mean 0.65), 0.25–0.84 (mean 0.65) and 0.21–0.78 (mean 0.58), respectively. High rates of cross-species/genus amplification were also observed within fourteen different species. The primer sequences for the eight EST-SSRs have been deposited in the Probe database of GenBank (IDs Pr031820546 – Pr031820553). The EST-SSRs developed herein would be a valuable addition of functional markers for genetics and breeding applications in a wide range of *Salix* species.

Key words: Salix, Willow, Expressed sequence tag (EST), Simple sequence repeat (SSR); Microsatellites.

SSR markers are very useful for a spectrum of genetic and breeding applications because of their reproducibility, co-dominant inheritance, multi-allelic nature and good genome coverage (Powell et al., 1996). As EST-SSRs are derived from transcribed genomic regions, they are more likely to be conserved and transferable among related organisms. Therefore, EST-SSRs have considerable potentials for comparative mapping and evolutionary studies as well as for assaying the functional diversity of natural populations and germplasm collections (Varshney et al., 2005; Ellis and Burke, 2007). In case of *Salix*, however, only 100 genomic SSRs (7 in Lian et al., 2001; 7 in Stamati et al., 2003; 46 in Barker et al., 2003; 8 in Kikuchi et al., 2005; 5 in Deng et al., 2008; 18 in Hoshikawa et al., 2009; 9 in King et al., 2010) have been reported so far, which are still insufficient for genetic applications as compared with some other woody plants, e.g. over 460 polymorphic EST-SSR markers available for *Eucalyptus* (reviewed by Zhou et al., 2014). Our objectives are to increase the number of EST-SSR markers for *Salix* by mining the current EST resources and evaluate them for polymorphism and cross-species amplification.

A subset of 707 *Salix* EST accessions deposited in dbEST of GenBank (as of 30 June 2013) were downloaded (http://www.ncbi.nlm.nih.gov/dbEST/). The EST sequences were scanned for SSRs with di-, tri-, tetra-, penta- and hexa-nucleotide at a minimum of 6, 5, 4, 4, and 4 repeats respectively using the Perl script Tool (MISA, http://pgrc.ipk-gatersleben.de/misa/). A total of 22 primer pairs based on the 35 SSR-containing unigenes were designed using Primer 3 (Rozen and Skaletsky, 2000) and synthesized by Sangon Co. (Shanghai, China). The remaining 13 SSR-containing unigenes were not available for prime design due to the criteria limitation or inappropriate flanking sequences.

[1)] Department of Tree Genetics and Breeding, Jiangsu Academy of Forestry, Ningdan Road, Nanjing 211153, People's Republic of China.

[2)] Department of Horticulture, Cornell University, New York State Agricultural Experiment Station, 630 West North Street, Geneva, NY 14456, USA.

[*)] Corresponding author: Xudong He. Tel: +86 25 52744080, Fax: +86 25 52741620. E-Mail: hxd_519@163.com

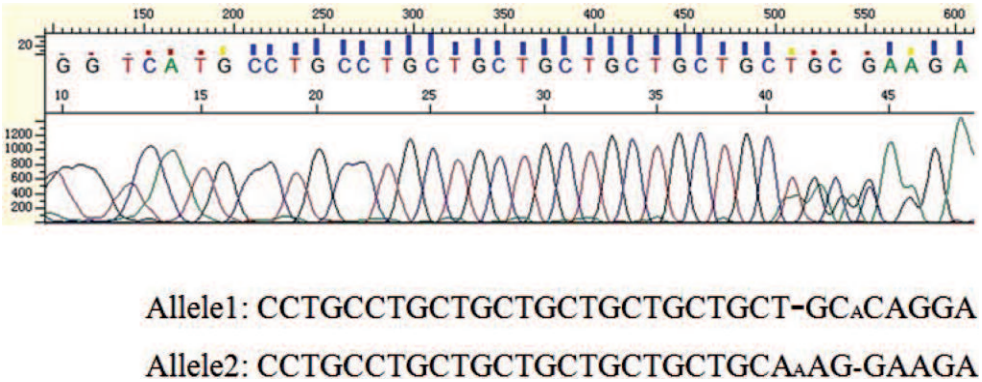

Figure 1. – Partial re-sequencing profile of marker SALeSSR003 with *S. viminalis*. The allele1 and allele2 correspond to $(CTG)_7$ and $(CTG)_6$ among two chromosomes, respectively. The hyphens indicate the INDEL loci within two alleles.

DOI:10.1515/sg-2014-0015
edited by Thünen Institute of Forest Genetics

Table 1. – The sequences of forward and reverse primers, Mg^{2+} concentration, annealing temperature (T_m), repeat motif and sequence identity of the 8 EST-SSRs developed in *Salix*.

EST-SSR locus	NCBI dbEST accession (ProbeDB ID)	Primer sequence (5'–3')	Mg^{2+} (mM)	T_m (°C)	Target size (bp)	Repeat motif in EST	Repeat motif sequenced	Intron	Identity (%)
SALeSSR001	CB185209 Pr031820546	F:TCAACGCACTTGCTTTCCG R:GCGTCCTTGAACCGCATAA	1.5	58	149	$(TC)_8$	$(TC)_{10}$	N	60.90
SALeSSR002	CB185143 Pr031820547	F:GCTCGTCGGAAGAAAGGTG R:GGGAATAGGCGAAGAGGTGT	1.5	58	353	$(ATG)_5$	$(ATG)_5$	Y	94.80
SALeSSR003	JK747690 Pr031820548	F:GGGAATACCCATCAAAGTCG R:GCAAGCCTTCAAACTACCCTC	1.5	58	144	$(CTG)_5$	$(CTG)_{6/7}$	Y	75.00
SALeSSR004	JK747654 Pr031820549	F:ATGAACCCTGGTCACTCTGTC R:ACCTGATTGTCCTTCACCCT	1.5	58	285	$(CCA)_5$	$CCG(CCA)_5CTA$	Y	96.03
SALeSSR005	CB185346 Pr031820550	F:TTGGCGACACTGTTGAGA R:AGGCACATATCCAGGGTC	1.5	58	462	$(AAG)_7$	$(AAG)_2AAC(AAG)_3AAA/(AAG)_7$	N	92.61
SALeSSR006	CB185176 Pr031820551	F:AATCAGCGCCATTCTTG R:TGCCTGTCTGCTCCTTG	1.5	58	237	$(AAG)_6$	$(AAG)_8$	Y	80.84
SALeSSR007	JK747631 Pr031820552	F:AGCGGGTTAAGCAGGAG R:AAATGAAAGCCAGACACG	1.5	58	112	$(CT)_8$	$(CT)_{7/11}$	Y	67.67
SALeSSR008	JK747550 Pr031820553	F:GTCTTCGCCATCTTCTTCC R:CCTCCGTGCAATACCTTG	1.5	58	287	$(TCC)_6$	$(TCC)_{4/5}$	N	94.94

The characters flanking the slash "/" show the allelic variation within the sequenced *S. viminalis* when both alleles were obtained in sequencing. "Y" or "N" indicates if there is an intron included in the amplicons sequenced or not.

An individual of *S. viminalis* was employed to optimize polymerase chain reaction (PCR) with regards to Mg^{2+} concentration and melting temperature (Tm), then to confirm the sequence identity of the PCR product with that of the original EST. PCR reactions of 10 uL consisted of 5.0 uL PCRmix (Sangon Co., Ltd, Shanghai, China), 1.5 mM $MgCl_2$, 10 pM forward primer, 10 pM reverse primer and about 3 ng DNA template. The reactions were performed on a DNA Engine (Bio-Rad, Hercules, CA, USA) under the following program: 94 °C for 4 min; 37 cycles of 94 °C for 30 s, 58 °C for 30 s, and 72 °C for 1.5 min; and a final extension at 72 °C for 10 min. The PCR products were checked for effective amplification through electrophoresis in 1% agarose gels containing 1:20 GoldView (SBS Genetech Co., Beijing, China). Ten primers failed in amplification, leading to no or very weak product or more than one fragments.

Twelve PCR products were purified in 96-well plates through ethanol-NaAc precipitation, ethanol cleaning and 1×TE dissolution, and sequencing was performed using BigDye Terminator Version 3.1 (BDT3.1) on ABI 3730xl genetic analyzer (Applied Biosystems, Foster City, CA) according to the manufacture's instructions. Finally, 10 amplicons were sequenced successfully with

Table 2. – Polymorphisms, cross-species amplification and functional annotations of the eight EST-SSRs.

EST-SSR locus	Polymorphisms in *S. eriocephala*					Cross-species amplification[a]	SSR position	EST putative function at $E \leq 10^{-5}$ [Organism]	BlastX *E*-value
	N_a	*ASR*	H_o	H_e	*PIC*				
SALeSSR001	4	152-160	0.85	0.61	0.53	*Svi Sca Ser Sci*	CDS	predicted protein	5E-69
SALeSSR002	1	483	NC	NC	NC	*Ssu Svi Sca Sin Seri Ser Sci Car Sch Sma Sko Sni Sal Sba*	CDS	FK506-binding protein	9E-60
SALeSSR003	1	244	NC	NC	NC	*Ssu Svi Sca Sin Seri Ser Sci Car Sch Sma Sko Sni Sal Sba*	CDS	putative B-type response regulator 13	3E-70
SALeSSR004	1	372	NC	NC	NC	*Ssu Svi Sca Sin Seri Ser Sci Car Sch Sma Sko Sni Sal Sba*	CDS	predicted protein	2E-45
SALeSSR005	5	450-465	0.57	0.78	0.71	*Ssu Svi Sca Sin Seri Ser Sci Car Sch Sma Sko Sni Sal Sba*	3'UTR	anthocyanidin synthase, partial	3E-75
SALeSSR006	2	339-354	0.29	0.25	0.21	*Ssu Svi Sca Sin Seri Ser Sci Car Sch Sma Sko Sni Sal Sba*	5'UTR	light-harvesting complex I protein	4E-99
SALeSSR007	7	202-220	0.75	0.84	0.78	*Ssu Svi Sca Sin Seri Ser Sci Car Sch Sma Sko Sni Sal Sba*	CDS	uncharacterized protein	3E-06
SALeSSR008	6	286-304	0.79	0.76	0.68	*Ssu Svi Sca Sin Seri Ser Sci Car Sch Sma Sko Sni Sal Sba*	5'UTR	predicted protein	1E-24
Mean[b]	4.8		0.65	0.65	0.58				

N_a number of alleles per locus, *ASR* allele size range, *PIC* polymorphic information content, *NC* not calculated.

[a] Ssu, *S. suchowensis*; Svi, *S. viminalis*; Sca, *S. caprea*; Sin, *S. integra*; Seri, *S. eriocephala*; Ser, *S. erioclada*; Sci, *S. cinerea*; Car, *Chosenia. arbutifolia*; Sch, *S. chaenomeloides*; Sma, *S. matsudana*; Sko, *S. koreensis*; Sni, *S. nigra*; Sal, *S. alba*; Sba, *S. babylonica*.

[b] The 3 monomorphic EST-SSRs were excluded for calculation of mean values.

DOI:10.1515/sg-2014-0015
edited by Thünen Institute of Forest Genetics

primer sequences and approximate length confirmed. The other two amplicons failed in re-sequencing, as the tandem repeats were close to the binding site of sequencing primer. *Figure 1* shows partial re-sequencing profile of *S. viminalis* for marker SALeSSR003.

Sequence alignment using DNAMAN 5.2.2 (Lynnon Biosoft, Point-Claire, Quebec, Canada) revealed that 8 out of the 10 amplicons sequenced were identical to the original ESTs with a range from 60.90% to 96.03% (*Table 1*). The remaining two were considered to be non-specific amplification. Moreover, in terms of SSR position by using ORF Finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html), five were found in coding sequence (CDS), two in 5'-untranslated region (5'UTR) and one in 3'UTR (*Table 2*). Furthermore, BlastX searches against NCBI non-redundant protein database under the default settings showed that 5 of the 8 EST-SSRs were homologous to known genes and 3 corresponded to predicted or hypothetical proteins (E≤10-5; *Table 2*).

Fourteen species, including seven shrub willow and seven tree willow, were sampled for cross-species/genus amplification analyses (*Table 2*). Meanwhile, another 14 *S. eriocephala* individuals collected from the United States were used for polymorphism estimation. The PCR reaction was the same as above except that 10 pmol fluorescein-12-dUTP (Fermentas International Inc., Burlington, Ontario, Canada) was incorporated. The PCR products (1 µL) were diluted 1:10.5 with loading buffer (9.34 µL deionized formamide and 0.16 µL GeneScan 500LIZ) and then detected on ABI 3730xl using GeneMapper 4.0 (Applied Biosystems). Of the eight EST-SSR loci scored, five were polymorphic among the 14 individuals of *S. eriocephala* tested. Polymorphism estimation in MSA software (Dieringer and Schlötterer, 2003) indicated that the number of alleles per locus (N_a), observed heterozygosity (H_o), expected heterozygosity (H_e) and polymorphic information content (*PIC*) were 2–7 (mean 4.8), 0.29–0.85 (mean 0.65), 0.25–0.84 (mean 0.65) and 0.21–0.78 (mean 0.58), respectively (*Table 2*). High rates of cross-species/genus amplification were observed since all the EST-SSRs amplified a PCR product within seven shrub willow and seven tree willow genotypes except SALeSSR001, for which there was only observed a successful amplification from the four shrub willow genotypes (*Table 2*).

In conclusion, this report demonstrates several advantages of re-sequencing in developing SSR markers from publicly accessible databases. High polymorphisms and cross-species/subgenus amplification were characterized for most of the EST-SSR loci. Thus, the markers developed in this study would be a valuable addition of functional markers for genetic applications in *Salix*.

Acknowledgments

This work was financially supported by grants from the Introduction of Advanced International Programme of Forestry Science and Technology (2011-4-38), Jiangsu Provincial Natural Science Foundation (BK2011871) and the National Natural Science Foundation of China (31300556). The authors are grateful to Chunfen Han and Hongling Wang for plant material maintenance. Special thanks are also due to Dr. Siming Gan and anonymous reviewers for their valuable comments on the manuscript.

References

Barker, J. H. A., A. Pahlich, S. Trybush, K. J. Edwards and A. Karp (2003): Microsatellite markers for diverse *Salix* species. Mol Ecol Notes **3**: 4–6.

Deng, Y. Y., L. A. Xu, B. Zhang, Q. Zhuge, M. R. Huang and M. X. Wang (2008): Development of genomic based microsatellites from *Salix caprea* and SSR polymorphic loci identification in *Salix* species. Mol Plant Breed **1**: 89–94.

Dieringer, D. and C. Schlötterer (2003): Microsatellite analyzer (MSA): a platform independent analysis tool for large microsatellite data sets. Mol Ecol Notes **3**: 167–169.

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

Hoshikawa, T., S. Kikuchi, T. Nagamitsu and N. Tomaru (2009): Eighteen microsatellite loci in *Salix arbutifolia* (Salicaceae) and cross-species amplification in *Salix* and *Populus* species. Mol Ecol Resour **9**: 1202–1205

Kikuchi, S., W. Suzuki, N. Ban, A. Kanazashi and H. Yoshimaru (2005): Characterization of eight polymorphic microsatellites in endangered willow *Salix hukaoana*. Mol Ecol Notes **5**: 869–870.

King, R. A., S. L. Harris, A. Karp and J. H. A. Barker (2010): Characterisation and inheritance of nuclear microsatellite loci for use in population studies of the allotetraploid *Salix alba – Salix fragilis* complex. Tree Genet Genomes **6**: 247–258.

Lian, C., K. Nara, H. Nakaya, Z. Zhou, B. Wu, N. Miyashita and T. Hogetsu (2001): Development of microsatellite markers in polyploid *Salix reinii*. Mol Ecol Notes **1**: 160–161.

Powell, W., G. C. Machray and J. Provan (1996): Polymorphisms revealed by simple sequence repeats. Trends Plant Sci **1**: 215–222.

Rozen, S. and H. J. Skaletsky (2000): Primer3 on the WWW for general users and for biologist programmers. *In:* Bioinformatics methods and protocols: methods in molecular biology, eds Krawetz, S. and S. Misener. Humana Press, Totowa, NJ. 365–386.

Stamati, K., S. Blackie, J. W. S. Brown and J. Russell (2003): A set of polymorphic SSR loci for subarctic willow (*Salix lanata*, *S. lapponum* and *S. herbacea*). Mol Ecol Notes **3**: 280–282.

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

Zhou, C. P., X. D. He, F. G. Li, Q. J. Weng, X. L. Yu, Y. Wang, M. Li, J. S. Shi and S. M. Gan (2014): Development of 240 novel EST-SSRs in *Eucalyptus* L'Herit. Mol Breeding **33**: 221–225.

DOI:10.1515/sg-2014-0015
edited by Thünen Institute of Forest Genetics