Development and characterization of new microsatellites for walnut (*Juglans regia*)

Z.Y. Zhang, J.W. Han, Q. Jin, Y. Wang, X.M. Pang and Y.Y. Li

National Engineering Laboratory for Tree Breeding of the National Development and Reform Commission,
College of Biological Sciences and Technology, Beijing Forestry University, Beijing, China

Corresponding author: Y.Y. Li
E-mail: yingyueli@bjfu.edu.cn

ABSTRACT. The expressed sequence tag (EST) database represents a potentially valuable resource for the development of simple sequence repeat (SSR) markers for use in evolutionary studies. EST-SSRs reveal polymorphisms not only within the source taxon, but in related taxa as well. In this paper, we describe a case study in which the publicly available walnut (*Juglans regia*) EST database was used to develop SSR markers for use in the genetic analysis of the widespread *Juglans nigra* and *Carya cathayensis* and an endangered species *Annamocarya sinensis*. A total of 7262 unigenes, including 1911 contigs and 5351 singletons, were obtained from 13,559 ESTs retrieved from the NCBI database. The 7262 unigenes were further reduced to 706 EST-SSR sequences containing 805 SSR loci. Then, 309 EST-SSR primers were randomly designed, and 77 were identified with five high across-species transferability cross-species: namely, *J. regia*, *J. nigra*, *C. cathayensis*, *Carya dabieshanensis*, and *A. sinensis*. Thirteen highly polymorphic EST-SSRs were further used for genetic analyses in these above five species.

Key words: EST-SSR; *Juglans regia*; *Juglans nigra*; *Carya cathayensis*; *Carya dabieshanensis*; *Annamocarya sinensis*

INTRODUCTION

Microsatellites (simple sequence repeats, or SSRs) consist of tandem repeats of short (1 to 6 bp) nucleotide motifs, and these repetitive stretches are distributed throughout the genome (Toth et al., 2000; Morgante et al., 2002). Owing to their highly polymorphic and co-dominant nature, SSRs have been increasingly used for plant genetic studies (van de Wiel et al., 1999; Simko, 2009).

SSR markers can be developed in several ways: via genomic libraries, enriched genomic libraries, BAC/YAC libraries, and cDNA libraries (Scott, 2001). Unfortunately, the *de novo* development of SSRs is costly and time-consuming, and the primers for polymerase chain reaction (PCR) amplification are frequently species-specific, which means that primers developed in one species cannot be readily applied for other species. In particular, for species that lack commercial importance, the publicly available resources are rare and limited, which further adds difficulty to SSR marker development (Pashley et al., 2006).

One possible solution to those problems might be to develop relatively high-quality SSR primers that could transfer across the border of taxonomic boundaries. Fortunately, expressed sequence tag (EST)-derived SSR primers have been proven to be desirable choices. In recent years, an ever-increasing number of cDNA sequences are being uploaded to the GenBank database, and numerous EST-SSRs have been consequently developed (Bhat et al., 2005). In addition, numerous studies have found that EST-SSRs are significantly more transferable than traditional anonymous counterparts (Leigh et al., 2003; Rungis et al., 2004). Nowadays, more attention has been paid on the potential for using EST-SSRs developed from well-studied species as a tool to facilitate genetic analysis of relatively less-studied species and endangered/rare species (Ellis and Burke, 2007).

Juglans is one of the major genera of the family Juglandaceae, in which several species such as *Juglans regia* and *Juglans nigra* have been relatively well-studied and numerous cultivars have been released. As of July 3, 2012, a total of 74,809 nucleotide sequences, including 1968 nucleotides, 23,505 ESTs, and 49,336 genome survey sequences, have been deposited in GenBank (http://www.ncbi.nlm.nih.gov/nucest/?term=Juglans%20regia). Consequently, an increasing number of molecular markers have been developed. For example, Qi et al. (2010) and Zhang et al. (2010) have developed 35 and 30 highly transferable and polymorphic EST-SSRs by screening the 5025 *J. regia* ESTs, respectively. SSR markers have been widely used for genetic analysis of the progeny of walnut cultivars and lines as a basis for marker-assisted breeding (Pollegioni et al., 2009; Ahmed et al., 2012). However, the limited primers were not enough to create desirable results, especially for narrow kinships, so more polymorphic EST-SSRs are needed.

Carya cathayensis and *Carya dabieshanensis* are important commercial plants, valued for their high-quality timber and edible nuts. A recent few commercial cultivars are only a few breeding and selection cycles removed from their wild progenitors (Guo, 2004). In the past few years, more attention has been paid to the development of molecular markers and their genetic analyses. Grauke et al. (2003) developed 19 *C. cathayensi* genomic SSR markers. Song et al. (2012) built the cDNA library associated with fat metabolism in *C. cathayensis* and developed 34 EST-SSRs.

Annamocarya sinensis is the only species of genus *Annamocarya* that is native to southwestern China and northern Vietnam. It has been listed as a critically endangered species

in China (Gu, 2003). *A. sinensis* is comparatively untapped, and even its taxonomy is a lasting controversy. For some anthropogenic and natural factors, the individual and population situations became worse. Nevertheless, *A. sinensis* showed a high level of genetic diversity, as evaluated on 70 individuals from 3 main populations using 12 EST-SSRs (Zhang et al., 2012).

In the paper, we developed and characterized more EST-SSR primers for *J. regia*. Furthermore, the transferability and polymorphism levels of the primers were tested among five species in three genera of family Juglandaceae.

MATERIAL AND METHODS

Plant materials and DNA extraction

A total of 40 accessions, including five species, were used for SSR primers characterization. Species *J. regia* was represented by eight cultivars: Bokexiang, Xinjufeng, Luguang, Liaoning No.5, Beijing 146, Xiluo, Jingxiang No.1, Jingxiang No.2, whereas *J. nigra* was represented by eight individuals (Black 1-8 as noted), which were propagated by seeds (introduced from America). The above 16 accessions were kindly provided by the germplasm collection of the Beijing Forestry Institute of Pomology. *C. cathayensis*, *C. dabieshanensis*, and *A. sinensis* were respectively represented by eight wild adult individuals collected from April to December 2011.

DNA was extracted from silica-gel-dried young leaves using a modified 2X cetyltrimethylammonium bromide protocol (Doyle and Doyle, 1987). The quality of the DNA was checked using electrophoresis on 1% agarose gel, and the DNA concentration was determined with a spectrophotometer.

SSR identification

A total of 13,559 ESTs of *J. hindsii* x *J. regia* were retrieved from the National Center of Biotechnology Information database on July 3, 2012. The procedure of SSR identification was performed by a PERL script designed by Zhang et al. (2012).

SSR development

Primers (20 to 24 bp) for amplification of SSR-containing fragments (100 to 400 bp) were designed using Online Primer 3, with 56°C as the optimal annealing temperature. The names of the newly developed markers were designed to start with BFU-Jr, representing Beijing Forestry University-*J. regia.*

In addition, to get enough primers for genetic analysis, another 110 SSR markers were also deliberately selected from four different sources: 19 genomic SSR markers (denoted as PM-CIN) were selected from the *C. cathayensis* developed by Grauke et al. (2003); 42 genomic SSRs (denoted as WGA) were from *J. nigra*, including 14 developed by Dangl et al. (2005) and 28 (out of 30, excepted WGA4 and WGA69) developed by Woeste et al. (2002); 37 EST-SSRs (denoted as ZMZ and ZY) were from the *J. regia* developed by Qi et al. (2011); and 12 EST-SSRs (denoted as BJF-As) were from the *A. sinensis* developed by Zhang et al. (2012).

Polymerase chain reaction (PCR) amplification

An M13-tagged sequence (5'-TGTAAAACGACGGCCAGT-3') was added to the 5'-end of the forward primer to allow detection with a universal fluorescent-labeled (FAM, HEX, TAMRA, ROX) M13 primer. PCR amplification, product separation, and gel staining were performed as reported previously by Zhang et al. (2012). The PCR products were resolved using an ABI 3730XL DNA analyzer from Genewiz Biotechnology Co., Ltd. (Beijing, China) and the data were analyzed with the Gene-Marker software (SoftGenetics LLC, Oakwood Ave, PA, USA).

Data analysis

SSR transferability and polymorphism

All SSR markers were initially used to examine the amplification ability and cross-species transferability on five species: *J. regia*, *J. nigra*, *C. cathayensis*, *C. dabieshanensis*, and *A. sinensis*. Each species was represented by three individuals. The presence or absence of PCR products was scored on 2% agarose gels, and was then assigned with a quality score (Q) of between 1 and 5, where 1 denotes a PCR product of expected size with a clear signal and no stuttering; 2 denotes a scorable product accompanied by faint stutter bands; 3 denotes an unscorable ladder of stutter bands or a multilocus amplification product; 4 denotes a weak/unreliable product; and 5 denotes a failed amplification (Leigh et al., 2003; Pashley et al., 2006). Thereafter, only loci with Q = 1 or 2 in all five species were further estimated for the level of polymorphism. These calculations minimized and ignored the possibility of null alleles and missing data.

The level of polymorphism was screened on a panel of 40 accessions, with eight accessions for each of the five species. Consequently, the band results reflected by capillary electrophoresis were classified into three classes according to the transfer criteria: 1, polymorphic; 2, monomorphic; 3, polymorphic but with low product yield or nonspecific amplification (Shepherd et al., 2002).

Genetic diversity

The genetic diversity indices were calculated, including the number of alleles per locus (N_A), effective number of alleles (N_E), observed (H_O) and expected (H_E) heterozygosities, and fixation index (F), using Arlequin 3.11 (Excoffier et al., 2005) and GenAlEx 6.2 (Peakal and Smouse, 2006). The polymorphic information content (PIC) value for each primer was analyzed using Cervus 3.0 (Marshall et al., 1998).

RESULTS AND DISCUSSION

EST database

A total of 13,559 ESTs were downloaded from the NCBI database on July 3, 2012. After removing redundant and junk sequences, 7262 unigenes (including 1911 contigs and

5351 singletons) were assembled. The 7262 unigenes were further reduced to 706 (9.7%) EST-SSR sequences, and 82 sequences containing more than one SSR loci. Additionally, 805 (11.1%) SSR loci were identified, corresponding to one SSR in every 0.17 kb of the ESTs. These results were close to the counterparts of 5025 *J. reiga* ESTs screened by Qi et al. (2011), with 9.7%, 10.75%, and 0.18 kb, respectively.

Among the 805 loci, the most frequent motif was dinucleotide (448 out of 805, 55.6%), followed by tri- (217, 26.9%), tetra- (78, 9.7%), hexa- (41, 5.1%), and penta- (21, 2.6%). In addition, among the dimeric repeats, the AG/CT (314 out of 448) and AT/AT (103/448) were the most common motifs; among the trimeric repeats, AAG/CTT (71/217) and AAT/ATT (32/217) were the most common ones. In contrast to other plants, the motifs AT and AG were also most common in loblolly pine ESTs (Liewlaksaneeyanawin et al., 2004), but the AT motifs were the rarest in rice (Temnykh et al., 2000) and barley (Thiel et al., 2003). For the trimeric repeats, AAG was the most abundant in loblolly pine and *Arabidopsis* (Cardle et al., 2000), whereas GGC was the most abundant in maize (Chin et al., 1996). However, it was difficult and inaccurate to compare the prevalence of motifs even in the same species, owning to the different minimal motif-repeat criteria (Liewlaksaneeyanawin et al., 2004).

SSR markers development

Altogether 321 EST-SSR primer pairs were initially developed for PCR amplification. After checking, all 12 *A. sinensis* EST-SSRs (denoted as BJF-As) previously developed by Zhang et al. (2012) were included in the 321 primers. Thus, here, we actually newly developed 309 EST-SSRs. When all 419 SSR primers (309 newly developed EST-SSRs and 110 additional SSRs selected from four sources) were tested for the ability of PCR amplification, the number of marker-amplified discrete bands (i.e. $Q = 1$ or 2) in *J. regia*, *J. nigra*, *C. cathayensis*, *C. dabieshanensis*, and *A. sinensis* were 257 (61.3%), 264 (63.0%), 199 (47.5%), 183 (43.7%), and 158 (37.7%), respectively (Figure 1). When only the 309 newly developed EST-SSRs were taken into account, the corresponding results were 191 (61.8%), 233 (75.4%), 152 (49.2%), 130 (42.1%), and 120 (38.8%), respectively. The remaining loci produced weak, null, and/or nonspecific amplifications.
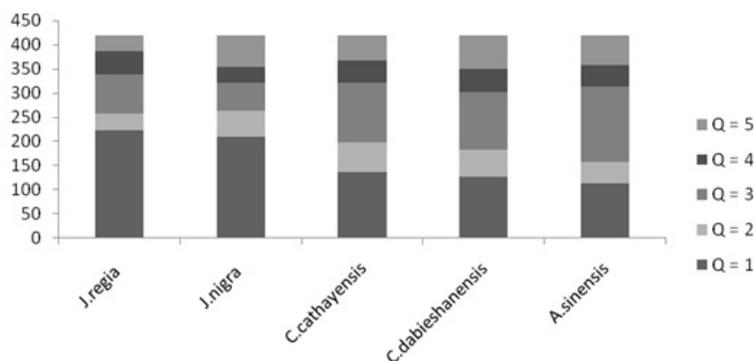

Figure 1. Number of loci with quality score 1 to 5. Quality score (Q) from 1 to 5: 1 denotes a PCR product of expected size with a clear signal and no stuttering; 2 denotes a scorable product accompanied by faint stutter bands; 3 denotes an unscorable ladder of stutter bands or a multilocus amplification product; 4 denotes a weak/unreliable product; 5 denotes a failed amplification.

In general, the success rate for EST-SSR primers was 60 to 90% in various reports (Varshney et al., 2005). In our research, the success rates for *J. regia* (61.8%) and *J. nigra* (75.4%) were in this range, whereas the rates for the other three species were apparently lower. One explanation for this result might be that the transferability success rate reduces as the evolutionary distance between the source taxon and recipient taxa increases. When compared with some studies involving species within a family, this explanation was acceptable. For example, Gutierrez et al. (2005) transferred EST-SSRs derived from *Medicago truncatula* to *Vicia faba*, *Cicer* sp, and *Pisum sativum* (all four species within a subfamily), with a success rate of 43, 39, and 40%, respectively. Our result was in line with the well-proven conclusion that EST-SSRs could transfer across multiple species not only within a genus but within a subfamily or family as well (Ellis and Burke, 2007).

In addition, EST-derived SSRs were generally more transferrable than genomic SSRs. This conclusion was also verified in the study. When only considering the 42 *J. nigra* genomic SSRs, the number of desirable loci (i.e., Q = 1 or 2) in *J. nigra* was 30 (71.4%), which was significantly higher than that in the other four species, namely, 52.3, 30.9, 30.9, and 19.0% in order (data not shown).

SSR polymorphism

A total of 102 SSR markers were identified that amplified clear bands in all five species. Table 1 shows the polymorphic details of 77 newly developed EST-SSRs. Table 2 shows the other 25 from four sources, with 2 from *C. cathayensis*, 4 from *J. nigra*, 10 from *J. regia*, and 9 from *A. sinensis*.

In our study, 42.9% (33 out of 77) of the newly designed EST-SSRs revealed polymorphism among the eight cultivars of *J. regia*, which was much lower than those detected in gerbera (Gong and Deng, 2010) and tall fescue (Saha et al., 2004) EST-SSRs. For the other four species, the results (41.6, 13.0, 15.6, and 27.3% in order) were even lower by contrast. Liewlaksaneeyanawin et al. (2004) reported that EST-SSR markers with trinucleotide repeats were less polymorphic than dinucleotides and other motifs. Our study was consistent with the conclusion. Among the 77 EST-SSRs, 40 (51.9%) were dinucleotides, and 34 (44.2%) were trinucleotides, and the remaining three (3.5%) were two tetranucleotides and one compound motif. Among the 30 (except BFU-Jr243) polymorphic loci, however, 21 (70.0%) were dinucleotides, and 9 (30.0%) were trinucleotides (Tables 1, 2, and 3). Overall, the 77 newly designed markers contained a relatively large percentage of less-polymorphic trinucleotide EST-SSRs. This might be one explanation for the low level of polymorphism. Another possibility might be due to the relatively narrow kinship among the selected walnut cultivars. In this study, *J. regia* was represented by eight cultivars, but four of them (Bokexiang, Xinjufeng, Luguang, and Liaoning No. 5) were originated from Xinjiang, China. The other four cultivars originated from Beijing, China. All eight cultivars were derived from internal crosses, leading to high homozygosity and reduced polymorphism. Moreover, some of them might have common parents in their genetic backgrounds.

Table 1. List of 77 newly developed EST-SSRs for *Juglans regia*, *Juglans nigra*, *Carya cathayensis*, *Carya dabieshanensis* and *Annamocarya sinensis* (EXCEL 1).

Marker	Primer sequence	Repeat length	*J. regia* *J. nigra*	*C. cathayensis* *C. dabieshanensis*	*A. sinensis* -
BFU-Jr10	F: TCTGGAGGGTTTCCATTGTC	$(TGA)_5(GAT)_7$	2	2	2
	R: TGGCTGAGGTGTCTCTTCCT	333	2	2	-
BFU-Jr19	F: CAGAAAGCACGTGAACCAGA	$(TA)_{10}$	1	2	2
	R: AACACTCACGCACCTTCATTT	276	1	1	-
BFU-Jr29	F: CTAGCACTTCCAGCCACACA	$(TC)_{12}$	1	1	1
	R: GGTGGAATCTTGGAGAGCAA	245	1	2	-
BFU-Jr38	F: AGCTCCTCAAGCAAGGCTTA	$(GAT)_{13}$	1	1	2
	R: GTGCATGGAACCACACTCAG	130	2	1	-
BFU-Jr40	F: GAGTTCTGCGGTGTCCATTC	$(TCT)_6$	1	2	2
	R: TGTCAGCGGATTGATCAGAG	147	1	2	-
BFU-Jr41	F: GGTCTGGATCCAATGGTCAC	$(AG)_{13}$	2	3-M	2
	R: TCTAACACTGGGCCAAGACA	383	2	3-M	-
BFU-Jr42	F: CACAGGGCCAAAAGAAACAT	$(GAC)_{18}$	2	2	2
	R: TTAACCACCCTAGCCACGTC	302	2	2	-
BFU-Jr45	F: TTCTGGTCTTCCCAATGGTC	$(TC)_8$	1	2	2
	R: GCCTACCCCCACATTATCCT	195	1	2	-
BFU-Jr46	F: GTAGCGGCGAGATTTCACTC	$(CAC)_6$	2	2	2
	R: GAATTAACTCGGAGGGATGG	196	2	2	-
BFU-Jr48	F: TCACCAAGTCTGCTCCAAAA	$(TC)_6$	2	2	2
	R: GATGGCGAAGCAGTTCTTTC	150	2	2	-
BFU-Jr49	F: ACTCAATTGACGTCCCTCTCT	$(TC)_6$	1	1	1
	R: TCTGCCAACTGAATTGAACG	249	1	1	-
BFU-Jr52	F: GCAACGATCAAGGTCCCTAA	$(CT)_9$	2	2	2
	R: CGGAGACCCAAATCAAAAGA	234	2	2	-
BFU-Jr62	F: AGGGCGAGTAGCCTCTTTGT	$(AGA)_6$	2	2	2
	R: GGGGTGTTTTGCAATGAGAT	140	2	2	-
BFU-Jr67	F: TCGCCAGAGCAGAGTACAGA	$(GA)_{11}$	1	1	1
	R: GTAGCCCAACGAGCAAAGAG	138	1	1	-
BFU-Jr68	F: TGGTATTTCTCCCGATGAGC	$(TC)_6$	1	2	2
	R: AAGGTCGTGGAAGGAGGATT	231	1	2	-
BFU-Jr69	F: CAAGCAGAAACAGTCGGAGA	$(GA)_{15}$	1	2	2
	R: GGGCAGAGAGAGAGCAAAGA	126	3-M	2	-
BFU-Jr70	F: GAGTGTGGAGTCTGCTGCTG	$(AAG)_7$	2	2	2
	R: CGCTCCTCCTCTTCCTTTCT	291	2	2	-
BFU-Jr73	F: CACCCGAGAGAGAAGTGGAG	$(AGA)_5$	2	2	2
	R: GCCTGCTAGTAGGGCACAAG	261	2	2	-
BFU-Jr82	F: TTGCCTCTCTATAGGCTTTTTAGC	$(CT)_7$	1	1	1
	R: TGAGTGTCTGGACAGCAAGG	159	1	1	-
BFU-Jr95	F: GCAACCTTTCCACGTTCATT	$(TC)_8$	2	2	2
	R: TTGAGCACAACAAGCAAAGC	100	2	2	-
BFU-Jr97	F: TCGGGTTTGTGTTTTGTGAA	$(TCT)_5$	2	2	2
	R: AAATGGCCAACTCCACATTC	222	2	2	-
BFU-Jr98	F: AGAGATGGCGAAGTCGAAGA	$(AGAC)_4$	2	2	2
	R: CGCCCCTACTAGGAATTCAA	322	2	2	-
BFU-Jr99	F: GTGCCCTTGGAGACAACACT	$(CT)_8$	1	2	1
	R: AGCTGCTCACCCTCAACAAT	317	2	2	-
BFU-Jr102	F: GTGCCCTAACGTCTGGTGTC	$(TC)_7$	2	2	2
	R: GCATGCCCAGTACCTGATCT	368	2	2	-
BFU-Jr104	F: TGCAGAAGAGGGAGGAGGTA	$(GAA)_7$	2	2	2
	R: ATCCCCAGCTTCATCTTCCT	229	2	2	-
BFU-Jr116	F: GTGGCTTTGCATAAGCAGGT	$(GA)_{13}$	2	2	2
	R: CCTCAAGTTTCTGCCTGGAG	277	2	2	-
BFU-Jr121	F: CGGAAAAGAGAAGCGAAGAA	$(AAG)_7$	2	1	2
	R: TCGACCTTGATCTGGAGCTT	286	2	1	-
BFU-Jr126	F: CCATTGCTTGGGGTTAAAAA	$(CT)_6$	1	1	1
	R: TGAAGAAAAGGCCTCCAAAA	210	1	1	-
BFU-Jr130	F: CTCGCAACTGTGAGGAATGA	$(TA)_7$	1	2	2
	R: AACTGCCAACTGTGCAAATG	160	1	2	-
BFU-Jr132	F: AGGTGGAGGTGGAGGTTATG	$(TGG)_6$	2	2	2
	R: GAACCATCGGTCAAAAAGGA	362	2	2	-

Table 1. Continued.

Marker	Primer sequence	Repeat length	*J. regia* *J. nigra*	*C. cathayensis* *C. dabieshanensis*	*A. sinensis* -
BFU-Jr144	F: GGAAGGGTCGGAGTTGGTAT	$(TAT)_5$	2	2	2
	R: TGTACCCAAATCACTCACTGAA	310	2	2	-
BFU-Jr150	F: CAACAGTGACACACCCCAAG	$(AAGA)_4$	2	2	2
	TGGGGAAAAGGGAAACTTCT	187	2	2	-
BFU-Jr152	F: CCCAAATTCGAAACCCTAGA	$(CT)_6$	2	2	2
	R: AAGACCGCTGATCCTCTTGA	191	2	2	-
BFU-Jr154	F: GGCATGCAAAGGAAACAAAT	$(TA)_6$	2	2	2
	R: CTGGAAATCATAATCGGTCCA	182	2	2	-
BFU-Jr161	F: CAACGGCTATGATCCCAAGT	$(CT)_{14}$	1	2	1
	R: TCCTGATCCATGCAATCAAA	175	1	2	-
BFU-Jr163	F: ATAGGGTGGTTGGTGTCGAG	$(TAT)_5$	2	2	2
	R: TGAGGTAATCCCAGCTACCAA	191	2	2	-
BFU-Jr167	F: GGGAAGATCTAGTTTTGGGCTA	$(AGG)_5$	1	1	1
	R: CCAACAATGACACCAGTTGC	313	1	1	-
BFU-Jr171	F: GGATCGGCATCACAGAGATT	$(GA)_6$	2	1	1
	R: TCTCCCTCTCCTTCTCCTTG	128	2	1	-
BFU-Jr179	F: AGCCACTGTGGAGGCAATAC	$(TC)_{11}$	1	1	1
	R: ATGGTGGAACCACTTTGGAA	169	1	1	-
BFU-Jr181	F: CAGCTTCAAAAACAGCAGCA	$(CAG)_6$	2	2	2
	R: CTGGTACGCCATTTGGAACT	300	2	2	-
BFU-Jr182	F: AACGCCATAGCTTCATCCAC	$(CT)_6$	2	2	2
	R: AATCGTTTCGGATGAACGTC	138	2	2	-
BFU-Jr183	F: AAGAATAGGGTTGGGCCAGT	$(TC)_{15}$	2	2	2
	R: CAATGCACAACTACTTGACTGC	289	2	2	-
BFU-Jr184	F: CCGACCTCACTCTTGTCTCC	$(TA)_7$	1	2	2
	R: CATTCGCCATGATCCTGTAA	267	1	2	-
BFU-Jr185	F: AGCACATCCCACCTTCTCTC	$(TC)_6(TC)_7$	2	2	2
	R: AGTGACTTGTCGGCCTTTGT	297	2	2	-
BFU-Jr195	F: CCTCCTTGGGTGCTAAATGA	$(TC)_{12}$	1	3-M	2
	R: TTTGCCCAAAAACAACTGTG	220	1	3-M	-
BFU-Jr196	F: CCAAGGCATCCTCCTTAACA	$(TC)_{11}$	1	3-M	1
	R: ATGGCTGGTCGTTTGCTAGT	231	1	3-M	-
BFU-Jr201	F: GGAGAACTTATCCCGCATGA	$(ACC)_7$	2	2	2
	R: GGGACCGGTGGTTTTAGAAT	240	2	2	-
BFU-Jr207	F: CTGGTTTGCTGGTTTTGGTT	$(CT)_{15}$	1	2	2
	R: ATCACCCGTTCGCCATATC	263	1	2	-
BFU-Jr217	F: AGTGGATCCGTGAGGTTGAG	$(TCT)_6$	2	2	2
	R: CGAAACACGCACAAATTCTC	254	2	2	-
BFU-Jr219	F: AACCCAAAGCTTCGCAGATA	$(GA)_7$	1	3-W	1
	R: ATCTCAGCCTGAAAGGCAAA	128	1	3-W	-
BFU-Jr228	F: GGGGCCAACTTTAGGGATAA	$(TAT)_4(ATT)_4$	2	2	2
	R: TTCGCAATGGGAAGGAATTA	259	2	2	-
BFU-Jr230	F: AATCACCACCCTCCAACAAG	$(CCT)_8$	2	1	2
	R: TTTGGAATCCCAAGCTCATC	270	1	1	-
BFU-Jr233	F: TTCCACTTCCTGGACCTCAC	$(CT)_{17}$	1	2	1
	R: ATTTATCCCGATGACCCACA	230	1	2	-
BFU-Jr238	F: TGGGAGAGCTGGTTTTATGC	$(TC)_{18}$	1	2	2
	R: TGAGGTACATCACTGGGAAGG	260	2	2	-
BFU-Jr240	F: CTCCAGAGATGCTCGACTCC	$(GAT)_4(GAC)_4$	2	2	2
	R: GCTCTGAAATTGCGGAACTC	249	2	2	-
BFU-Jr243	F: ACCGAAACTGCACCAAAATC	$(AG)_4(CAA)_4(CACCAG)_4$	1	2	2
	R: ACGGCGCAGTATTTCTGTCT	221	1	2	-
BFU-Jr245	F: GTACTCCATTGCCCACCATC	$(CT)_6(TA)_5$	2	2	2
	R: CAACGCAAAATTCAAACACG	280	2	2	-
BFU-Jr247	F: TTGGGTGCTACCTTGCTTCT	$(TC)_{11}$	1	3-W	2
	R: CAGACTGCAACATGCCAGAT	152	1	3-W	-
BFU-Jr249	F: GCAACGATCAAGGTCCCTAA	$(TC)_9$	2	2	2
	R: CCAAGCACATTGAAACCAAA	105	2	2	-
BFU-Jr250	F: GCATGTTCGTCAAGGGCTAT	$(GAT)_6$	2	2	2
	R: TTCCACGCTATAAAGCACAG	127	2	2	-

Continued on next page

Table 1. Continued.

Marker	Primer sequence	Repeat length	*J. regia* *J. nigra*	*C. cathayensis* *C. dabieshanensis*	*A. sinensis* -
BFU-Jr254	F: GCCGGAAAAATCAGAATTGA	$(TA)_{11}$	2	2	2
	R: TTCAACGTCAACTTGGCATC	260	2	2	-
BFU-Jr256	F: CCTTTCCATTCACTTTTGCTG	$(CT)_{19}$	2	2	1
	R: CCAGCACCAAGGCTAAGAAG	150	1	1	-
BFU-Jr263	F: ACCGTCTACGTCACGGGTTC	$(AAG)_8$	2	2	1
	R: CTGTTGCTGTTGTCGGTGAT	151	3-M	2	-
BFU-Jr275	F: CATCTGCCCCCAACTTCTAA	$(CT)_{12}$	2	2	2
	R: CGTTAGCATCTTCTGCCACA	284	2	2	-
BFU-Jr276	F: TCGTGTGATTCTGGGTGTGT	$(CT)_{14}$	2	2	2
	R: GCCTCATTGGTTGAATTGCT	183	2	2	-
BFU-Jr277	F: TATTCACCCGGAGGTTTCAG	$(GAT)_{10}$	1	3-M	1
	R: CCGAAGCCAGTCGAGTTATC	250	1	3-M	-
BFU-Jr284	F: GCATGTTCGTCAAGGGCTAT	$(GAT)_6$	2	2	2
	R: TTCCACGCTATGAAGCACAG	126	2	2	-
BFU-Jr286	F: TGAACCCATGCTTCTCCTTC	$(AT)_9$	2	2	1
	R: CAATGGGTGACTCGCAAATA	247	1	1	-
BFU-Jr288	F: CCTCCACTGCTACAGCCAAA	$(AAC)_7$	1	2	2
	R: GCTTTTGAGACAGGGAGTGC	180	1	2	-
BFU-Jr292	F: CAAAGAAGACCCAAGAACATAGC	$(AGA)_8$	2	1	2
	R: GAACCAGTGAGGCAGAAAGC	112	1	1	-
BFU-Jr298	F: CTCGTCATGAACCACCTCCT	$(CAC)_6$	1	2	2
	R: GAACATCCTTGCTAGCTTTGG	251	1	2	-
BFU-Jr301	F: CAGCTGCTTCTGGTTTTGTG	$(CAG)_7$	1	2	1
	R: GACCAATTGGAGGTGGCTTA	225	1	2	-
BFU-Jr306	F: TGCAGTAGCCATAGCTCGAA	$(AAG)_6$	2	2	2
	R: GAAGGCTGCTCCAAGAACAC	140	2	2	-
BFU-Jr307	F: GCTTCCTCCCAGAAAACTTG	$(TCT)_7$	1	2	2
	R: ACCCATTTTGATTTCGTCCA	111	1	2	-
BFU-Jr319	F: CCTGTCACGTCGAGGAGTTC	$(CAC)_8$	1	2	1
	R: CATCAAACCCTAAAGCACCAA	136	1	2	-
BFU-Jr321	F: TGGTATGAGGAAGCAGTTGAAA	$(TCA)_9$	2	2	2
	R: TCACTATCACGTTGCCTTCG	175	2	2	-

Quality class of SSRs. 1 = Polymorphic, 2 = monomorphic, 3 = poor amplification, M multiple alleles, W weak amplification.

Table 2. List of 25 additional SSR markers for *Juglans regia*, *Juglans nigra*, *Carya cathayensis*, *Carya dabieshanensis* and *Annamocarya sinensis*.

Locus	*J. regia*	*J. nigra*	*C. cathayensis*	*C. dabieshanensis*	*A. sinensis*
PM-CIN20	2	2	2	2	2
PM-CIN22	2	2	2	2	2
WGA9	1	1	2	2	2
WGA24	2	2	2	2	2
WGA69	1	1	3-W	3-W	3-W
WGA74	2	2	2	2	2
ZMZ1	2	2	2	2	2
ZMZ3	1	2	2	2	2
ZMZ4	1	1	2	2	2
ZMZ7	2	2	1	1	2
ZMZ11	1	1	2	1	2
ZMZ15	2	2	2	2	3-W
ZMZ21	2	2	1	2	2
ZMZ34	2	2	2	2	2
ZMZ39	2	1	2	2	2
ZY1	1	1	3-M	3-W	2

Table 2. Continued.

Locus	*J. regia*	*J. nigra*	*C. cathayensis*	*C. dabieshanensis*	*A. sinensis*
BJF-AS4	2	2	2	2	2
BJF-AS6	2	2	2	2	2
BJF-AS16	1	2	2	1	1
BJF-AS57	1	1	2	2	1
BJF-AS65	2	2	2	2	1
BJF-AS107	1	1	2	2	1
BJF-AS170	1	1	1	2	1
BJF-AS180	1	1	2	2	1
BJF-AS192	1	1	2	2	1

Quality class of SSRs. 1 = Polymorphic, 2 = monomorphic, 3 = poor amplification, M multiple alleles, W weak amplification.

Table 3. Summary of the polymorphism of the 102 SSRs selected.

Species	*J. regia*	*J. nigra*	*C. cathayensis*	*C. dabieshanensis*	*A. sinensis*
Class-1	33 (9)	32 (9)	10 (5)	12 (5)	21 (3)
Class-2	44 (16)	43 (16)	61 (18)	59 (18)	56 (20)
Class-3	0 (0)	2 (0)	6 (2)	6 (2)	0 (2)
Total	77 (25)	77 (25)	77 (25)	77 (25)	77 (25)

Quality class of SSRs. 1 = Polymorphic; 2 = monomorphic; 3 = poor amplification.

Genetic analysis

In total, 13 polymorphic EST-SSR markers were randomly selected to estimate the genetic diversity among the 40 accessions. A total of 104 alleles were generated, ranging from 5 to 11, with an average of 8 per marker. The mean number of effective alleles was 4.887, ranging from 2.415 to 7.711. The other genetic diversity parameters are shown in Table 4, all demonstrating a high level of polymorphism among the accessions.

Table 4. Polymorphism levels of 13 selected EST-SSRs.

Locus	N_A	N_E	H_O	H_E	PIC	F
BFU-Jr 19	9	5.915	0.475	0.831	0.812	0.428
BFU-Jr 29	11	7.711	0.550	0.870	0.857	0.368
BFU-Jr 49	9	6.026	0.700	0.834	0.817	0.161
BFU-Jr 67	9	4.651	0.600	0.785	0.758	0.236
BFU-Jr 82	9	6.751	0.750	0.852	0.834	0.120
BFU-Jr 161	10	3.448	0.275	0.710	0.689	0.613
BFU-Jr 179	5	2.766	0.225	0.638	0.577	0.648
BFU-Jr 230	5	2.415	0.525	0.586	0.546	0.104
BFU-Jr 286	7	4.342	0.325	0.770	0.744	0.578
BFU-Jr 319	5	3.236	0.275	0.691	0.649	0.602
BJF-AS16	9	6.262	0.325	0.840	0.820	0.613
BJF-AS 170	8	5.604	0.525	0.822	0.799	0.361
BJF-AS 192	8	4.408	0.250	0.773	0.748	0.677
Mean	8.000	4.887	0.446	0.769	0.742	0.424
SE	0.543	0.454	0.049	0.024	0.025	0.059

N_A = number of observed alleles; N_E = number of expected alleles; H_O = observed heterozygosity; H_E = expected heterozygosity; PIC = polymorphic information content; F = fixation index.

In summary, herein we reported 77 new EST-SSR markers, which have shown a high success rate in PCR amplification and a high level of molecular polymorphism. These markers

should provide a powerful tool in *Juglans* and *Carya* breeding and genetic studies. Additionally, utilization of these markers for the elucidation of *A. sinensis* taxonomy is in progress.

ACKNOWLEDGMENTS

REFERENCES

Ahmed N, Mir JI, Mir RR, Rather NA, et al. (2012). SSR and RAPD analysis of genetic diversity in walnut (*Juglans regia* L.) genotypes from Jammu and Kashmir, India. *Physiol. Mol. Biol. Plants* 18: 149-160.

Bhat PR, Krishnakumar V, Hendre PS, Rajendrakumar P, et al. (2005). Identification and characterization of expressed sequence tags-derived simple sequence repeats markers from robusta coffee variety 'CxR' (an interspecific hybrid of *Coffea canephora* x *Coffea congensis*). *Mol. Ecol. Notes* 5: 80-83.

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

Chin EC, Senior ML, Shu H and Smith JS (1996). Maize simple repetitive DNA sequences: abundance and allele variation. *Genome* 39: 866-873.

Dangl GS, Woeste K, Aradhya MK, Koehmstedt A, et al. (2005). Characterization of 14 microsatellite markers for genetic analysis and cultivar identification of walnut. *J. Am. Soc. Hort. Sci.* 130: 348-354.

Doyle JJ and Doyle JL (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* 19: 11-15.

Ellis JR and Burke JM (2007). EST-SSRs as a resource for population genetic analyses. *Heredity* 99: 125-132.

Excoffier L, Laval G and Schneider S (2005). Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol. Bioinform. Online* 1: 47-50.

Gong L and Deng Z (2010). EST-SSR markers for gerbera (*Gerbera hybrida*). *Mol. Breed.* 26: 125-132.

Grauke LJ, Iqbal MJ, Reddy AS and Thompson TE (2003). Developing microsatellite DNA markers in pecan. *J. Am. Soc. Hort. Sci.* 128: 374-380.

Gu YC (2003). Status quo of China's state priority protected wild plants. *Cent. South For. Invent. Plann.* 22: 1-7.

Guo CY (2004). Ecological and Genetic Diversity Studies on *Carya cathayensis* and *C. dabieshanensis*. PhD thesis, Nanjing Forestry University, Nanjing.

Gutierrez MV, Vaz Patto MC, Huguet T, Cubero JI, et al. (2005). Cross-species amplification of *Medicago truncatula* microsatellites across three major pulse crops. *Theor. Appl. Genet.* 110: 1210-1217.

Leigh F, Lea V, Law J, Wolters P, et al. (2003). Assessment of EST- and genomic microsatellite markers for variety discrimination and genetic diversity studies in wheat. *Euphytica* 133: 359-366.

Leigh F, Kalendar R, Lea V, Lee D, et al. (2003). Comparison of the utility of barley retrotransposon families for genetic analysis by molecular marker techniques. *Mol. Genet. Genomics* 269: 464-474.

Liewlaksaneeyanawin C, Ritland CE, El-Kassaby YA and Ritland K (2004). Single-copy, species-transferable microsatellite markers developed from loblolly pine ESTs. *Theor. Appl. Genet.* 109: 361-369.

Marshall TC, Slate J, Kruuk LE and Pemberton JM (1998). Statistical confidence for likelihood-based paternity inference in natural populations. *Mol. Ecol.* 7: 639-655.

Morgante M, Hanafey M and Powell W (2002). Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. *Nat. Genet.* 30: 194-200.

Pashley CH, Ellis JR, McCauley DE and Burke JM (2006). EST databases as a source for molecular markers: lessons from *Helianthus*. *J. Hered.* 97: 381-388.

Peakal, R and Smouse PE (2006). GenAlEx 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Ecol. Notes* 6: 288-295.

Pollegioni P, Woeste K, Mugnozza GS and Malvolti ME (2009). Retrospective identification of hybridogenic walnut plants by SSR fingerprinting and parentage analysis. *Mol. Breed.* 24: 321-335.

Qi JX, Hao YB, Zhu Y, Wu CL, et al. (2011). Studies on germplasm of Juglans by EST-SSR markers. *Acta Hortic. Sin.* 38: 441-448.

Rungis D, Berube Y, Zhang J, Ralph S, et al. (2004). Robust simple sequence repeat markers for spruce (*Picea* spp.) from expressed sequence tags. *Theor. Appl. Genet.* 109: 1283-1294.

Saha MC, Mian MAR, Eujayl I, Zwonitzer JC, et al. (2004). Tall fescue EST-SSR markers with transferability across several grass species. *TAG Theor. Appl. Genet.* 109: 783-791.

Scott KD (2001). Plant Genotyping: The DNA Fingerprinting of Plants. In: Microsatellites Derived from ESTs, and their Comparison with Those Derived by other Methods (Henry RJ, ed.). CABI, Oxon, 225-237.

Shepherd M, Cross M, Maguire L, Dieters J, et al. (2002). Transpecific microsatellites for hard pines. *Theor. Appl. Genet.* 104: 819-827.

Simko I (2009). Development of EST-SSR markers for the study of population structure in lettuce (*Lactuca sativa* L.). *J. Hered.* 100: 256-262.

Song S, Huang YZ, Dong LM, Huang YJ, et al. (2012). Simple sequence repeat (SSR) markers in *Carya cathayensis*. *J. Zhejing A F Univ.* 29: 626-633.

Temnykh S, Park WD, Ayres N, Cartinhour S, et al. (2000). Mapping and genome organization of microsatellite sequences in rice (*Oryza sativa* L.). *Theor. Appl. Genet.* 100: 697-712.

Thiel T, Michalek W, Varshney RK and Graner A (2003). Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor. Appl. Genet.* 106: 411-422.

Toth G, Gaspari Z and Jurka J (2000). Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Res.* 10: 967-981.

van de Wiel C, Arens P and Vosman B (1999). Microsatellite retrieval in lettuce (*Lactuca sativa* L.). *Genome* 42: 139-149.

Varshney RK, Graner A and Sorrells ME (2005). Genic microsatellite markers in plants: features and applications. *Trends Biotechnol.* 23: 48-55.

Woeste K, Burns R, Rhodes O and Michler C (2002). Thirty polymorphic nuclear microsatellite loci from black walnut. *J. Hered.* 93: 58-60.

Zhang R, Zhu AD, Wang XJ, Yu J, et al. (2010). Development of *Juglans regia* SSR markers by data mining of the EST database. *Plant Mol. Biol. Rep.* 28: 646-653.

Zhang ZY, Pang XM, Han JW, Wang Y, et al. (2012). Conservation genetics of *Annamocarya sinensis* (Dode) Leroy, an endangered endemic species. *Genet. Mol. Res.* 12: 3965-3974.