Tree Genetics & Genomes (2012) 8:933–940
DOI 10.1007/s11295-012-0481-z

SHORT COMMUNICATION

Development of EST-SSR markers and investigation of genetic relatedness in tung tree

Wei Xu · Qing Yang · Huyin Huai · Aizhong Liu

Received: 2 November 2011 / Revised: 5 January 2012 / Accepted: 30 January 2012 / Published online: 3 March 2012

Abstract The tung tree is an important non-edible oilseed source and consists of two species *Vernicia fordii* and *Vernicia montana* native to southern China and southeast Asia. In this study, the frequency and distribution of simple sequence repeats (SSRs) in expressed sequence tags (ESTs) of *V. fordii* were characterized based on 2,407 available EST sequences from the National Center for Biotechnology Information database. Twenty-two EST-SSR markers were developed by screening the genomic DNAs of six individuals of *V. fordii* from three accessions and 120 individuals of *V. montana* from 30 indigenous *V. montana* accessions collected from different geographic areas in southwestern China and northern Laos. The 22 EST-SSR markers exhibited a moderate level of polymorphism in *V. montana* with an average of 3.36 alleles per locus and PIC of 0.401. Genetic relatedness investigation showed that there was not only distinct genetic differentiation among tung trees but also a distinct geographic pattern among *V. montana* accessions. The current study is the first report of the development of EST-SSRs and genetic relatedness investigation in tung trees. These EST-SSR markers reported here will be valuable resources for future genetic studies, like construction of linkage maps, diversity analysis, quantitative trait locus/association mapping, and molecular breeding of the tung tree. The genetic relatedness identified in *V. montana* would provide potential clues in choosing germplasms in interest as progenitors for cross breeding and variety improvement of *V. montana* in practice.

Keywords EST-SSRs · Genetic diversity · Tung tree · *Vernicia fordii* · *V. montana*

Communicated by Y. Tsumura

Electronic supplementary material The online version of this article (doi:10.1007/s11295-012-0481-z) contains supplementary material, which is available to authorized users.

W. Xu · Q. Yang · A. Liu (✉)
Key Laboratory of Tropical Forest Ecology, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences,
88 Xuefu Road,
Kunming 650223 Yunnan, China
e-mail: liuaizhong@xtbg.ac.cn

W. Xu · H. Huai
College of Bioscience and Biotechnology, Yangzhou University,
Yangzhou 225009 Jiangsu, China

Introduction

The tung tree is an important non-edible oilseed source, and its seed oil is often considered as drying oils used in paints and varnishes, and for polymerization in industry. Among all the vegetable oils, tung oil is distinctive as it majorly consists of *cis*-9, *trans*-11, *trans*-13-a-elaeostearic acid. Recently, studies have demonstrated that tung oil could be excellent feedstock for biodiesel production after blending with other biodiesel (Shang et al. 2010; Chen et al. 2010). Exploiting tung oil as feedstock of biodiesel is drawing great attention in China and South Asia.

There are two species of tung trees: *Vernicia montana* Lour. and *Vernicia fordii* (Hemsl.) Airy Shaw, both native to southern China and southeastern Asia. Taxonomically, *V. montana* is easily distinguished from *V. fordii* by its inflorescence and fruit characters. Physically, *V. montana* is an evergreen tree that grows up to 20 m in height and is usually dioecious with brown bark, glabrous branches, and sparsely elevated lenticels. Its inflorescences are usually monoecious and composed of new leaves. Its fruit is a globular drupe with wrinkled skin that turns from green to yellow upon ripening (Li and Gilbert 2008). Although *V. fordii* has been planted and tung oils have been used for hundreds of years in China, the main germplasm of *V.*

montana is distributed throughout open forests below an altitude of 1,600 m in Southern China, Myanmar, Vietnam, and Thailand. According to our previous field survey and preliminary evaluation, *V. montana* gives much higher yields of tung oil than *V. fordii*. It seems that *V. montana* has higher potential as biodiesel feedstock than *V. fordii*.

To use tung oil from oilseeds of *V. montana* for producing biodiesel, it is crucial to develop high-productivity varieties. The Xishuangbanna Tropical Botanical Garden (XTBG, the Chinese Academy of Sciences) had initiated the program "Collection and Improvement of Germplasm for *V. montana*" since 2006 to develop high-productivity varieties as biodiesel plant. Seeds or seedlings of 30 indigenous accessions were collected from open forests covering the main distribution areas of *V. montana* in China and planted at XTBG. These accessions represent the main resources of *V. montana* distributed in China. Investigating the genetic diversity and relatedness among germplasms is necessary to serve breeding and improvement of varieties. However, to our knowledge, little is known about molecular markers developed for tung trees. In this study, we developed a set of expressed sequence tag-simple sequence repeat (EST-SSR) markers for tung trees based on EST sequences of *V. fordii* and further assessed the genetic diversity and relatedness among tung tree accessions collected at XTBG using SSR markers available. Results reported here will be valuable resources for further breeding and improvement of varieties in tung trees.

Materials and methods

Plant materials

Based on an extensive field survey, we collected 30 indigenous accessions of *V. montana* from different geographical populations representing main resources of *V. montana* in China (Table 1 of the Electronic supplementary material). The seeds or seedlings of *V. montana* accessions collected were planted at XTBG (21°56′N, 101°15′E, 600 m asl). Four seeds from each accession were sampled to screen polymorphic EST-SSR markers and investigate the genetic relatedness of *V. montana* among accessions. Six samples of *V. fordii* collected from three geographical populations (two for each) were included as a control. The genomic DNA of each individual was isolated and purified from fresh leaves using the Plant Genomic DNA Kit (TianGen, Beijing, China) as described previously (Xu et al. 2011).

EST data mining and SSR identification

There are 2,407 EST sequences of *V. fordii* available from GenBank (http://ww.ncbi.nlm.nih.gov/dbest/) on July 1, 2010. These EST sequences (GO253190-GO253199, GR217756-GR220587, released by Wang et al. in 2009) were downloaded in FASTA format and trimmed to discard the too short sequences (<100 bp) and to shorten the too long sequences (>800 bp) for precluding the inclusion of low-quality sequences (Thiel et al. 2003; Feng et al. 2009; Qiu et al. 2010) or to remove the oligoT and oligoA stretches from the 5′ or 3′ ends of sequences corresponding to the mRNA polyA tail using the EST-TRIMMER software (http://www.pgrc.ipk-gatersleben.de/misa/download/est-trimmer.pl). Further, the redundant ESTs were removed by running the CD-HIT program (Li and Godzik 2006) with a 95% sequence similarity threshold for favoring elimination of redundancy and avoidance of possible overestimation. All SSR loci were identified from the non-redundant EST sequences trimmed by MIcroSAtellite (MISA, http://www.pgrc.ipk-gatersleben.de/misa/). While searching SSRs by the MISA script, the criteria include microsatellites considered to contain motifs that are between two and six nucleotides long, with a minimum of five repeats for dinucleotides and four repeats for trinucleotides, tetranucleotides, pentanucleotides, and hexanucleotides, respectively.

Primer designing, PCR amplification, and visualization of SSR loci

The primer pairs flanking the region of SSRs were designed with PRIMER3 software (http://frodo.wi.mit.edu/primer3) with parameters as follows: product length, 100–300 bp; primer size, 18–24 bp (optimum, 20 bp); and melting temperature between 57 and 63°C (optimum, 60°C). Polymerase chain reaction (PCR) was performed in a 10-μl reaction volume containing 1 μl genomic DNA (10 ng/μl), 1× PCR master mix (TransGen Biotech, Beijing, China), and 0.3 μl of both forward and reverse primers (10 pmol/μl) using an Eppendorf master cycler. The PCR conditions used were as follows: initial denaturation for 5 min at 95°C, followed by 10 cycles of denaturation for 30 s at 94°C, annealing for 45 s at 57–60°C (primer specific), and extension for 30 s at 72°C, subsequently followed by 25 cycles of denaturation for 30 s at 94°C, annealing for 45 s at 55°C, extension for 45 s at 72°C followed by final extension for 10 min at 72°C. After PCR amplification was confirmed on 1.5% agarose gel, PCR products were electrophoresed and separated on 8% polyacrylamide gels (acrylamide/bis-acrylamide, 19:1) using Bio-Rad electrophoresis system (USA). The sizes of PCR products on polyacrylamide gels were visualized by silver staining. The ambiguous counting of PCR products in length were double analyzed on the QIAxcel system (QIAGEN, Irvine, USA) using a 12-capillary QIAxcel DNA High Resolution Cartridge.

Data acquisition and analysis

The amplified bands of PCR resulted in the identification of polymorphic, monomorphic, and nonamplifying primer pairs.

The number and size of amplified bands were recorded as allele number and size. To estimate the informativeness of EST-SSR markers developed using allelic data, the gene diversity (i.e., expected heterozygosity, H_e) and the Polymorphism Information Content (PIC) at each locus were calculated using PIC calculator (http://www.liv.ac.uk/kempsj/pic.html). Inbreeding coefficient (F_{is}) per locus was calculated as the difference between observed (H_o) and expected heterozygosity (Nei and Kumar 2000). To investigate the genetic diversity and relationships among accessions, we used these polymorphic EST-SSR markers identified herein and the genome-based SSR markers developed in our previous study (Xu et al. 2011), and scored the SSR products as band presence (1) and absence (0) of the band, thus generating a binary matrix. The data matrix was analyzed to estimate Nei's genetic distance and construct a dendrogram among accessions using POPGENE 1.31. The data matrix of Nei's genetic distance was transferred to the software MEGA 5.0 (Tamura et al. 2011) to construct the dendrogram with neighbor-joining criteria. Further, we performed a principal coordinate analysis (PCO) for all tested accessions of tung trees based on their genotypes using GenAlEx 6 program (Peakall and Smouse 2006).

Results and discussion

Characterization of EST-SSRs in *V. fordii*

Based on the available 2,407 EST sequences of *V. fordii* from GenBank, 969 non-redundant EST sequences were obtained after trimming and eliminating redundancy, representing approximately 0.6 Mb. The search for SSRs in the 969 non-redundant EST sequences revealed that 266 SSRs in 218 ESTs, nearly one SSR in 3.6 unique ESTs (i.e., 22.5% of ESTs contains at least one SSR), 51 ESTs contained more than one SSR, and 16 SSRs were found as compound SSRs. This corresponds to an average distance between SSRs of approximately 2.25 kb (i.e., one SSR per 2.25 kb). To further compare the overall densities of SSRs in *V. forbii* non-redundant EST sequences with other plants reported in the same family Euphorbiaceae, we used the same cutoff criteria as Qiu et al. (2010) used in castor bean, Yadav et al. (2011) used in *Jatropha curcas*, Feng et al. (2009) used in rubber tree, and Raji et al. (2009) used in cassava, and found that the density of EST-SSR is higher than that in *J. curcas*, rubber tree, and cassava, while lower than that in castor bean (data not shown). Thus, the frequency and distribution of EST-SSRs in tung tree are comparable to other members in the family Euphorbiaceae, though the available database for tung trees is limited.

The 266 SSRs identified are comprised of 131 di-, 122 tri-, 8 tetra-, 4 penta-, and 1 hexanucleotide. The frequency of trinucleotide motifs with four repeats was more common (32.3%), followed by dinucleotides of five. The maximum repeat number of SSR motif (GA) reaches to 59. The length of the SSR region (motif length × repeat numbers) varied from 10 to 118 bases. While using the criteria for class I (>20 class I nucleotides in length) and class II SSRs (<20 nucleotides in length) as used in previous studies (Temnykh et al. 2001; Kantety et al. 2002), 60 SSRs represented class I while 206 SSRs represented class II. Among the dinucleotide repeats, the AG/CT motifs showed the most frequency (60.3%), followed by the AT/TA motifs (24.4%) and AC/GT (13.7%). Among the trinucleotide repeats, AAG/CTT motifs were the most common, accounting for 27.8%, followed by ACC/GGT (20.5%) and ACG/CTG (16.4%). Other motifs were identified in insignificant numbers. These characterizations of SSR motifs are consistent with Morgante et al.'s (2002) observation in other plants.

Polymorphism and cross-species transferability of EST-SSR markers in tung trees

All the 266 SSR-containing sequences were used to design primers. Only 170 (63.9%) primer pairs could be designed, including 25 primer pairs flanking class I SSRs and 145 primer pairs flanking class II SSRs. The remaining 96 (36.1%) SSR-containing sequences did not qualify the primer designing criteria such as very short DNA sequences flanking SSR (at least on one side) or the flanking sequences being inappropriate for primer designing. Because class I SSRs usually exhibited higher polymorphism than class II SSRs (Temnykh et al. 2001; Kantety et al. 2002), all 25 class I SSR primer pairs were selected for PCR amplification. In addition, 47 class II SSR primer pairs were randomly selected for PCR amplification. In total, 72 primer pairs were tested. In 56 (77.8%) cases, PCR products could be successfully amplified using genomic DNAs of six *V. fordii* samples, while for 16 primer pairs (22.2%), PCR completely failed, amplified too weakly, or amplified multiple bands, and the 16 primers were excluded from further analyses. If not for the unavoidable inadequate primer design, then a possible reason that those primers failed to be amplified might be because of the introns existing between the two primers (Varshney et al. 2005).

The 56 primer pairs were screened to test their cross-species transferability and characterize their polymorphisms in *V. montana*. Out of the 56 primer pairs, 48 (85.8%), comprised of 13 from class I SSR and 35 from class II SSR, successfully amplified PCR products in *V. montana*. The genotyping data of 48 primers showed that 22 primer pairs were polymorphic and 26 primer pairs were monomorphic within 120 samples. The overall proportion of polymorphic primers was 45.8%, similar to the polymorphic ratio of EST-SSRs in castor bean (41.1%) (Qiu et al. 2010). Out of the 22 polymorphic loci, 11 primers were from class I and class II SSRs, respectively (Table 1). The polymorphic proportions of class I SSR and class II SSR were 84.5% and

Table 1 Information of polymorphic primer pairs of EST-SSRs

EST ID	Code	Primer (5′–3′)	Repeat	Size (bp)	Range (bp)	
GR220197	TT42	F:CAACGTTTAGCAGAGAAAGATA R:GCTAAATAGCTTCGTAAAGGTC	$(AG)_6$	153	481–501	Class II
GR218940	TT46	F:GATCTTCTTGATAAGCCATTTC R:AGAGAAGCAACCAGAAGTTTAG	$(AT)_7$	147	147–167	
GR218801	TT65	F:AGAAATTAACTGGTGGAAAGC R:CAGGCAACTTTTCTTTTATCTT	$(CCA)_4$	153	141–156	
GR218312	TT68	F:ATATGGCAGTCTTGAAAATGA R:TTTGAACATACATAGCACATTG	$(AAT)_4$	153	147–162	
GR218445	TT70	F:ACCGTATTGCTAGTGACTTACC R:TCCATATATAGAGCCAATAGCC	$(CGA)_4$	150	147–159	
GR218446	TT71	F:TATTGCTAGTGACTTACCATCG R:TAATTATTCCCACCATTACTCC	$(CAG)_4$	165	162–171	
GR219355	TT62	F:ATAACCCATACAATACCCAAAC R:AGCTGCTATCTTGAAGACTTGT	$(CAC)_5$	150	150–168	
GR220559	TT10	F: CAGCCAAGGGAGTGGTTCTAAT R: CCAGCTAAAGTACCAGAGTTGTGA	$(TGG)_5$	150	144–153	
GR220139	TT14	F:GAATTTTCAACCAGAGGGACAG R:GCAGATACACGAAACCAAGACA	$(GCT)_5$	159	290–305	
GR219355	TT15	F:CCAGTATGATCAAGAGCAGCAG R:TGTACCTGGTGTCACTCATTCC	$(AAC)_5$	161	161–173	
GR218076	TT17	F:CCTAATCCTCCGTCTAGTGGTG R:CAACTGGATTACCTAGCCCAAC	$(GGA)_6$	175	175–274	
GR217944	TT13	F:AGGAGCAACTCAAGGCTAAAGA R:TTGTAACGCCTAGCAAAGGACT	$(CTC)_7$	157	154–160	Class I
GR218992	TT23	F:CACTTGCAGGCGTTGAACTAAT R:AGCGAAGACGAAGATGAGAGTT	$(AGA)_7$	222	222–246	
GR217875	TT16	F:CAACCACAAAGGAGGAAAGAAG R:TTAGGACCCAAAAGGTGCTCTA	$(GAA)_8$	164	164–173	
GR217875	TT31	F:GAAGAATTGCGTGTAGGAAGC R:CTGCGAGGAATGAGCTGGATA	$(TAAT)_5$	218	218–246	
GR217795	TT33	F:ACAAAGTACCACCGTAGCATCA R:CGAATCTTCTTGGGAAGTTGAG	$(TCAT)_5$	242	238–254	
GR220119	TT34	F:CGAGGAAGACAAGAAGGACTTG R:GGCTGCTCTTATGAAAAGATGG	$(CTTTT)_4$	204	209–274	
GR218400	TT36	F:CGATCAAAGGAGGATCAAGGTA R:GGCTGCTCTTATGAAAAGATGG	$(TTTCT)_5$	246	206–251	
GR218271	TT37	F:AGGTGTAGTGGTGGGAATGAGT R:CTCCTCAATGCAAAGCAGTGTA	$(GAAGGA)_4$	229	229–277	
GR218997	TT58	F:AAGCTTTACCAACAAAGAGAAT R:AATGTAGCTCTCATGGAGTCAG	$(AGA)_7$	153	147–156	
GR217946	TT04	F:GAGGCTGGGATTTTCTAAACCT R:TATTGAACAGTCCCAAGGGCTA	$(AT)_{20}$	200	200–260	
GR217915	TT09	F:CCTTGTGTGTGCAATATGAAGC R:GCAACAACTCTTTGTCACGTCT	$(CT)_{11}$	247	239–251	

31.4%, respectively. Within the 48 primer pairs against 120 samples, a total of 52 nonamplifying samples (null alleles, accounting for 0.9%) were identified. The 85.8% cross-species transferability of EST-SSRs between *V. fordii* and *V. montana* and the 0.9% rate of null alleles within *V. montana* showed that the EST-SSR markers developed here have high transferability and considerable low probability of null alleles between tung trees.

In total, 74 alleles were identified, ranging from two to eight each locus, with an average of 3.36 alleles (Table 2). Across 22 loci, observed heterozygosity (H_o) ranged from 0 to 1 (mean=0.177) in *V. montana* and from 0 to 1 (mean=0.258) in *V. fordii*; H_e ranged from 0.150 to 0.853 (mean=0.473) in *V. montana* and from 0 to 0.667 (mean=0.295) in *V. fordii*; F_{is} values ranged from −0.314 to 1 (mean=0.626) in *V. Montana* and from −1.221 to 1 (mean=0.125) in *V. fordii*; PIC values

Table 2 Genetic diversity of *V. montana* and *V. fordii* detected by EST-SSR analysis

Code	*V. montana* germplasms (N=120)				*V. fordii* germplasms (N=6)			
	N_a	H_e/H_o	F_{is}	PIC	N_a	H_e/H_o	F_{is}	PIC
TT42	3	0.233/0.258	−0.107	0.220	1	0.000/0.000	n/a	0.000
TT46	5	0.761/1.000	−0.314	0.725	1	0.000/0.000	n/a	0.000
TT65	2	0.464/0.017	0.963	0.357	2	0.375/0.833	−1.221	0.305
TT68	2	0.462/0.008	0.983	0.355	1	0.000/0.000	n/a	0.000
TT70	2	0.469/0.025	0.947	0.359	2	0.500/0.000	1.000	0.375
TT71	2	0.469/0.008	0.983	0.359	2	0.500/0.000	1.000	0.375
TT62	2	0.460/0.000	1.000	0.354	2	0.278/0.000	1.000	0.239
TT10	3	0.412/0.017	0.959	0.327	2	0.444/0.667	−0.502	0.346
TT14	2	0.461/0.008	0.983	0.355	1	0.000/0.000	n/a	0.000
TT15	4	0.545/0.150	0.725	0.448	1	0.000/0.000	n/a	0.000
TT17	5	0.619/0.092	0.851	0.539	2	0.444/0.667	−0.502	0.346
TT13	3	0.466/0.042	0.910	0.358	3	0.667/0.667	0.000	0.593
TT23	3	0.507/0.058	0.886	0.405	3	0.611/1.000	−0.637	0.536
TT16	3	0.150/0.033	0.780	0.143	2	0.500/0.000	1.000	0.375
TT31	4	0.608/0.275	0.548	0.550	2	0.444/0.333	0.250	0.346
TT33	4	0.586/0.075	0.872	0.504	2	0.500/0.000	1.000	0.375
TT34	3	0.196/0.175	0.107	0.177	1	0.000/0.000	n/a	0.000
TT36	5	0.267/0.100	0.625	0.253	2	0.444/0.000	1.000	0.346
TT37	3	0.330/0.417	−0.264	0.276	2	0.000/0.500	n/a	0.000
TT58	3	0.504/0.108	0.786	0.400	2	0.500/1.000	−1.000	0.375
TT04	8	0.853/1.000	−0.172	0.836	1	0.000/0.000	n/a	0.000
TT09	3	0.591/0.025	0.958	0.524	2	0.278/0.000	1.000	0.239
Mean	3.36	0.473/0.177	0.626	0.401	1.77	0.295/0.258	0.125	0.235

Number of alleles (N_a), expected (H_e) and observed (H_o) heterozygosities, PIC, and inbreeding coefficient (F_{is}) of each marker are shown

ranged from 0.143 to 0.836 (mean=0.401) in *V. montana* and from 0 to 0.593 (mean=0.235) in *V. fordii*, suggesting that the EST-SSR markers developed have a moderate level of polymorphism in tung trees. The overall F_{is} value (0.626) in *V. montana* was significantly higher than zero among all loci, indicating a moderate level of heterozygote deficiency. All the 22 identified markers, except TT42, TT46, TT37, and TT04, contributed to this observed heterozygote shortage (Table 2). This may be due to the limit of our sampling strategy in our current study. Samples used in this investigation were based on 30 geographic gramplasms (not at the population level). Usually, the population subdivision or germplasm differentiation may result in a distinct observed heterozygote shortage. Besides, sampling of four seeds from each individual may contribute to the observed heterozygote shortage. The gene diversity (H_e) reported here in tung trees is comparable to the EST-SSR markers in other members from the same family Euphorbiaceae such as castor bean (Qiu et al. 2010), *J. curcas* (Yadav et al. 2011), cassava (Raji et al. 2009), and rubber tree (Feng et al. 2009). The PIC value of each SSR locus is determined by both the number of alleles and their frequency distribution within the population. This was usually used to assess the informativeness level of markers developed and accordingly defined into high (PIC>0.5), moderate (0.5>PIC>0.25), and low (PIC<0.25) categories (Botstein et al. 1980). According to this criterion, 19 of the 22 EST-SSR markers developed in this study represented moderate or high PIC values, resolving power and marker index of most SSR loci, and only three (TT42, TT65, and TT16) were considered as poor markers in these respects. According to Li et al.'s (2008) investigation, *V. fordii* presents moderate genetic diversity among accessions using inter-simple sequence repeat markers. The low gene diversity (both H_e and PIC) of the 22 EST-SSR markers in *V. fordii* is due to the limited samples tested. Although the main objective of this study is to develop efficient genetic markers and perform the genetic relatedness assessment for *V. montana* accessions collected, we reasonably trust that these EST-SSR markers developed are equivalently useful for the *V. fordii* germplasm.

While considering the criteria for class I and class II SSRs, we separately analyzed the allele number and PIC value from 11 class I and 11 class II SSR loci in tung trees. It was shown that 42 (mean=3.82) and 32 (mean=2.91) alleles were identified from classes I and II SSR loci with an average PIC of 0.42 (for class I) and 0.38 (for class II), respectively, indicating class I SSR markers are significantly more polymorphic than the class II SSR markers (N_a and PIC both $p<0.01$, 2-sample t test). This result could provide

a clue in choosing efficiently SSR markers for further marker development in tung trees.

Hitherto, little work has been done on the development and application of SSR markers in tung tree genetic and breeding studies. To our knowledge, this is the first report on the development of EST-SSRs in tung trees. The current investigation, however, is based on the limited EST sequence sources. To obtain more gene-based SSR markers for the tung tree developing more EST sequence sources is necessary.

Genetic relatedness among accessions

In total, 42 SSR markers resulting in 154 allele loci, including 22 EST-SSR markers developed herein and the 20 genomic SSR markers (resulting in 80 allele loci) developed in our previous report (Xu et al. 2011), were used to test the genotypes of 120 *V. montana* samples representing 30 accessions and six *V. fordii* samples representing three accessions. These samples were used as study materials in this study, primarily for the development and characterization of EST-SSR markers and investigation of genetic relatedness among *V. montana* accessions collected. Samples were not selected for revealing the extent of genetic diversity at the population level. The PCO based on these genotypic data clearly showed that the two species, *V. fordii* and *V. montana*, were genetically distinct, and there was an obvious genetic differentiation among *V. montana* accessions as well (Fig. 1). Axis one separated *V. fordii* from *V. montana* and accounted for 49.5% of the variation. Axis two separated and generated three genetically differentiated groups (I, II, and III) of *V. montana* accessions and accounted for 36.3% of the variation.

The dendrogram constructed using neighbor-joining criteria revealed that the two species, *V. fordii* and *V. montana*, were genetically distinct. Within *V. montana* accessions, the dendrogram classified all the 30 accessions into three distinct clusters: clusters I, II, and III, with 12 accessions nested in cluster I, 14 accessions nested in cluster II, and 4 accessions nested cluster III (Fig. 2). Genetic differentiation resulting from variation among *V. montana* accessions is significant with F_{st}=0.327, p<0.01. These results are consistent with the results of PCO analysis, suggesting that there is a distinct genetic differentiation among *V. montana* accessions. Correspondingly, clusters I, II, and III are consistent with groups I, II, and III identified in PCO analysis. According to the medium genetic diversity and the significant genetic differentiation within *V. montana*, the gene pool of *V. montana* should not be narrow, favoring the breeding and variety improvement. Inspection of the geographic distribution of *V. montana* accessions revealed that the four accessions nested in cluster III were restricted in southwestern Yunnan, the accessions nested in cluster II were distributed from the central to eastern areas, and the accessions nested in cluster I were distributed across the western to central areas in southwestern China (Fig. 3). The distribution of accessions in cluster I overlapped with the accessions in cluster II in the central areas, and with the accessions in cluster III in the western areas, suggesting that a geographic

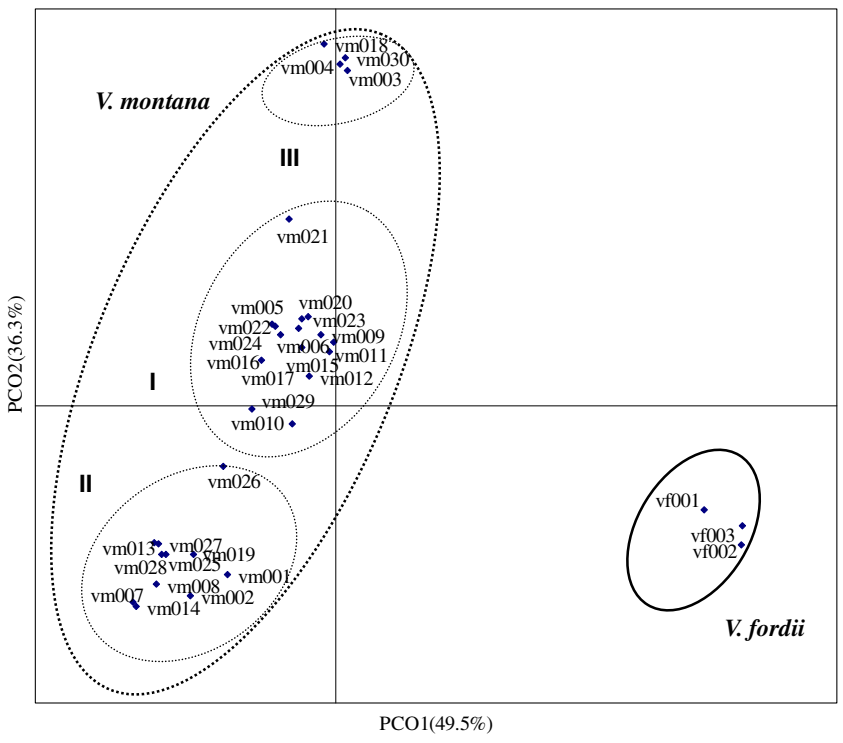

Fig. 1 PCO of the 30 *V. montana* accessions and 3 *V. fordii* accessions based on genotypic data from 42 SSR markers

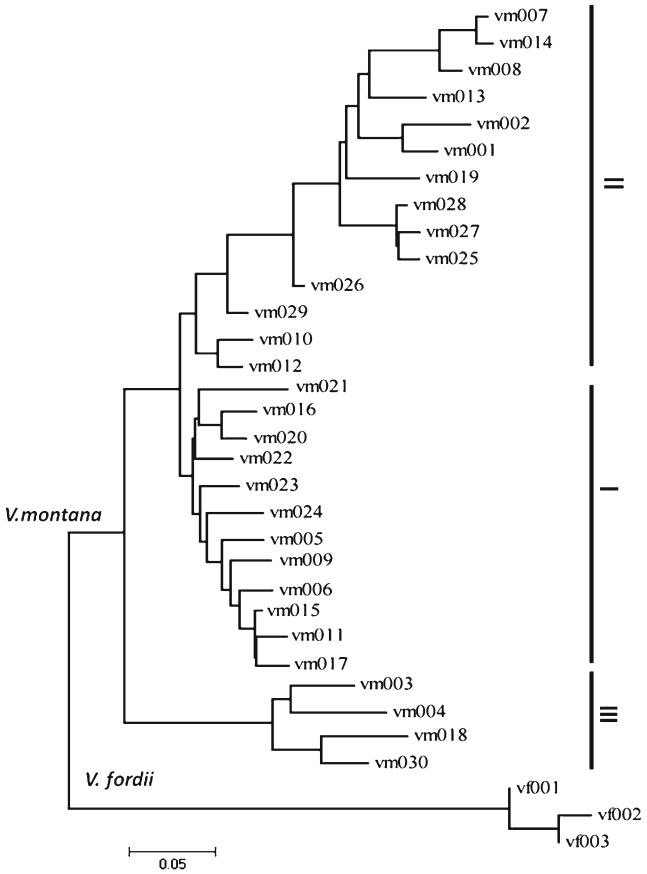

Fig. 2 Genetic relatedness among 30 accessions of *V. montana* and 3 *V. fordii* accessions based on 42 polymorphic SSRs using neighbor-joining criteria

distribution pattern exists from southwestern Yunnan, cross central to eastern areas in southern China. These clusters identified may affect the geographic differentiation of accessions, involving the diversification and spread of species *V. montana* in history. However, it is difficult to discuss the cause of the geographic differentiation of accessions biogeographically in this study due to a lack of knowledge on the genetic differentiation at the population level, speciation of tung trees, and fossil evidences. The genetic differentiation identified among *V. montana* accessions would provide potential clues in choosing germplasms in interest as progenitors for cross breeding and variety improvement of *V. montana* in practice.

Conclusions

In this study, we reported 22 novel EST-SSR markers which represented a valuable resource for future genetic studies in tung trees, like construction of linkage maps, genetic diversity analysis at the population level, quantitative trait locus/association mapping, and molecular breeding of the tung tree. The EST-SSR markers exhibited a moderate level of polymorphism in tung trees and were used to establish genetic relatedness among 30 accessions of *V. montana* and three accessions of *V. fordii*. The genetic differentiation identified among accessions would favor breeding and variety improvement of tung trees in practice.

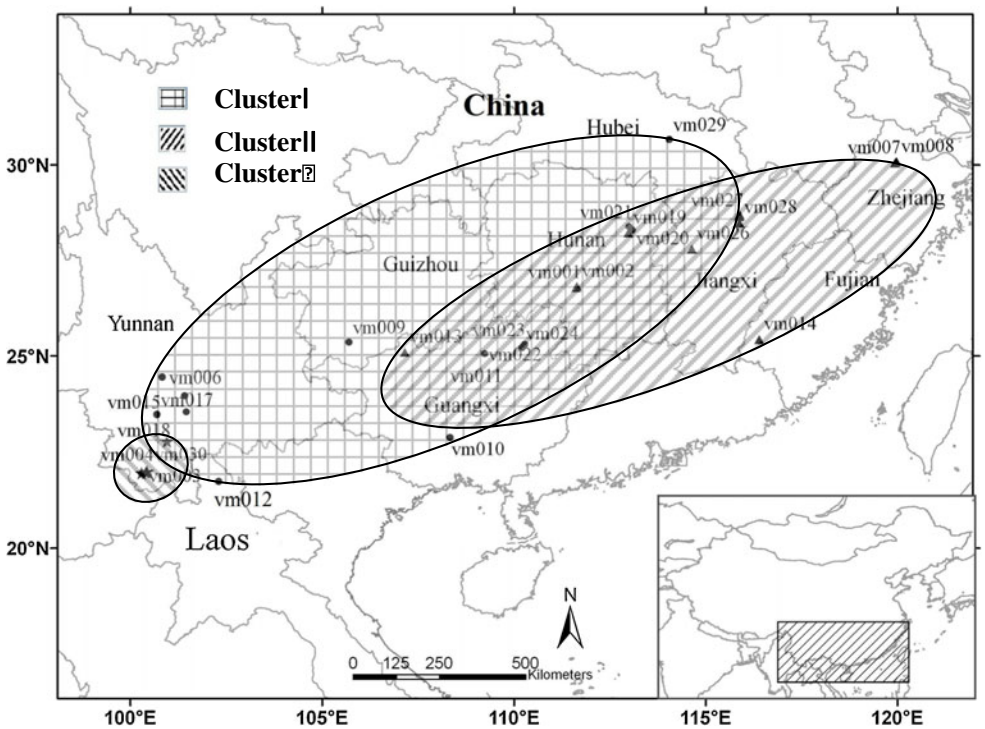

Fig. 3 Geographical distribution and genetic clusters of 30 *V. montana* accessions tested

Acknowledgments This research was jointly supported by the "100 Talents" Program of the Chinese Academy of Sciences (to AL) and research programs of the Yunnan Natural Science and Technology (2009FXS0021).

References

Botstein D, White RL, Skolnick M et al (1980) Construction of a genetic-linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32:314–331

Chen YH, Chen JH, Chang CY, Chang CC (2010) Biodiesel production from tung (*Vernicia montana*) oil and its blending properties in different fatty acid compositions. Bioresour Technol 101:9521–9526

Feng SP, Li WG, Huang HS, Wang JY, Wu YT (2009) Development, characterization and cross-species/genera transferability of EST-SSR markers for rubber tree (*Hevea brasiliensis*). Mol Breed 23:85–97

Kantety RV, La Rota M, Matthews DE, Sorrells ME (2002) Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. Plant Mol Biol 48:501–510

Li B, Gilbert MG (2008) The genus *Vernicia* in Euphorbiaceae. In: Wu CY, Raven PH (eds) Flora of China. Science Press, Beijing; Missouri Botanical Garden Press, St. Louis, vol. 11, pp 266–267

Li WZ, Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22:1658–1659

Li P, Zhang X, Chen Y, Lu G, Zhou G, Wang Y (2008) Genetic diversity and germplasm resource research on tung tree (*Vernicia fordii*) cultivars, investigated by inter-simple sequence repeats. Afr J Biotechnol 7:1054–1059

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

Nei M, Kumar S (2000) Molecular evolution and phylogenetics. Oxford University Press, Oxford

Peakall R, Smouse P (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Notes 6:288–295

Qiu L, Yang C, Tan B, Yang J, Liu AZ (2010) Exploiting EST databases for the development and characterization of EST-SSR markers in castor (*Ricinus communis* L.). BMC Plant Biol 10:278

Raji AA, Anderson VJ, Kolade AO, Ugwu DC, Dixon GA, Ingelbrecht LI (2009) Gene-based microsatellites for cassava (*Manihot esculenta* Crantz): prevalence, polymorphisms, and cross-taxa utility. BMC Plant Biol 9:118

Shang Q, Jiang W, Lu H (2010) Properties of tung oil biodiesel and its blends with $0^{\#}$ diesel. Bioresour Technol 101:826–828

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731–2739

Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinhour S, McCouch S (2001) Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. Genome Res 11:1441–1452

Thiel T, Michalek W, Varshney RK, 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

Varshney RK, Sigmund R, Börner A, Korzum V, Stein N, Sorrells ME, Langridge P, Graner A (2005) Interspecific transferability and comparative mapping of barley EST-SSR markers in wheat, rye and rice. Plant Sci 168:195–202

Xu W, Yang Q, Huai H, Liu AZ (2011) Microsatellite maker development in Tung trees (*Vernicia montana* and *V. fordii*, Euphorbiaceae). Am J Bot 98:e226–e228

Yadav HK, Ranjan A, Asif MH et al (2011) EST-derived SSR markers in *Jatropha curcas* L.: development, characterization, polymorphism, and transferability across the species/genera. Tree Genet Genomes 7:207–219