

## Development of simple sequence repeat markers and diversity analysis in alfalfa (*Medicago sativa* L.)

Zan Wang · Hongwei Yan · Xinnian Fu ·  
Xuehui Li · Hongwen Gao

Received: 23 August 2012 / Accepted: 17 December 2012 / Published online: 29 December 2012  
© Springer Science+Business Media Dordrecht 2012

**Abstract** Efficient and robust molecular markers are essential for molecular breeding in plant. Compared to dominant and bi-allelic markers, multiple alleles of simple sequence repeat (SSR) markers are particularly informative and superior in genetic linkage map and QTL mapping in autotetraploid species like alfalfa. The objective of this study was to enrich SSR markers directly from alfalfa expressed sequence tags (ESTs). A total of 12,371 alfalfa ESTs were retrieved from the National Center for Biotechnology Information. Total 774 SSR-containing ESTs were identified from 716 ESTs. On average, one SSR was found per 7.7 kb of EST sequences. Tri-nucleotide repeats (48.8 %) was the most abundant motif type, followed by di—(26.1 %), tetra—(11.5 %), penta—(9.7 %), and hexa-nucleotide (3.9 %). One hundred EST–SSR primer pairs were successfully designed and 29 exhibited polymorphism among 28 alfalfa accessions. The allele number per marker ranged from two to 21 with an average of 6.8. The PIC values ranged from 0.195 to 0.896 with an average of 0.608, indicating a high level of polymorphism of the EST–SSR markers. Based on the 29 EST–SSR markers, assessment of genetic diversity was conducted and found that *Medicago sativa* ssp. *sativa* was clearly different from the other subspecies. The high transferability of those EST–SSR markers was also found for relative species.

**Keywords** Alfalfa · EST–SSR markers · Genetic diversity

### Introduction

Alfalfa (*Medicago sativa* L.) is one of the most important forage legumes and widely planted for hay, and pasture in the world. Alfalfa exists at two ploidy levels, diploid ( $2n = 2x = 16$ ) and tetraploid ( $2n = 4x = 32$ ) with a basic chromosome number of eight. Most of alfalfa cultivars are autotetraploid and developed by phenotypic selection. Although many traits such as disease resistances, pest resistances, winter survival, etc., had been successfully improved by the phenotypic selection, it is time consuming. The molecular breeding approaches like marker-assisted selection could enhance efficiency of cultivar development in term of gain per unit cost and time.

Efficient and robust molecular markers are essential for molecular breeding. Simple sequence repeat (SSR) or microsatellite, 1–6 bp iterations of DNA sequences, was found in both coding and noncoding regions [1–3]. Because of co-dominant inheritance, abundance in genomes and high reproducibility, SSR markers have been developed and extensive used in molecular genetic studies for many species. In alfalfa, SSR markers have been broadly applied for population structure and diversity [4], genetic mapping [5, 6], and comparative mapping [7]. For autotetraploid species like alfalfa, up to four different alleles could be scored from one SSR marker in a single genotype. Compared to dominant markers and bi-allelic markers like single nucleotide polymorphism (SNP) markers, multiple alleles of SSR markers are particularly informative and superior in genetic linkage map and QTL analysis for autotetraploid species [8]. Almost the SSR

Z. Wang · H. Yan · X. Fu · H. Gao (✉)  
Institute of Animal Science, Chinese Academy of Agricultural  
Sciences, Beijing 100193, People's Republic of China  
e-mail: hongwengao@hotmail.com

X. Li  
Forage Improvement Division, The Samuel Roberts Noble  
Foundation, 2510 Sam Noble Parkway, Ardmore, OK 73401,  
USA

markers applied in alfalfa were derived from *Medicago truncatula*, a closely relative model species of alfalfa except that 61 polymorphic genomic SSR were developed by He et al. [9]. The expressed sequence tag (EST) derived SSR markers (EST–SSRs) developed from *M. truncatula* showed high level of transferability to alfalfa and other closely related species [10]. Different from *M. truncatula*, alfalfa is perennial. Genome size of alfalfa is about 800–1,000 Mbp, twice of *M. truncatula* [11]. Therefore, development of SSR markers directly from alfalfa could provide more informative markers for genetic studies and breeding applications.

Availability and continuous enrichment of ESTs in alfalfa can be served for development of EST–SSR markers. In total 12,371 ESTs (till 3 May, 2010) are now available for *M. sativa* in the National Center for Biotechnology Information (NCBI). The objects of this study were: (1) to analyze the frequency and distribution of SSRs in the alfalfa ESTs, (2) to develop and characterize of alfalfa EST–SSRs, (3) to assess intra-species genetic diversity and their cross-species/genera transferability.

## Materials and methods

### Plant materials and DNA extraction

Total 28 accessions from *M. sativa* ssp. *sativa*, *M. sativa* ssp. *falcata*, *M. sativa* ssp. *coerulea*, *M. sativa* ssp. *varia*, *M. sativa* ssp. *hemicycle*, and *M. sativa* ssp. *glomerata* were used to validate EST–SSR markers developed from this study (Table 1). One accession each of *Medicago minima*, *Medicago lupulina*, *Trifolium repens*, and *Melilotus albus* were used to assess the cross-species/genera transferability (Table 1). *T. repens*, and *M. albus* are the two species from different genera but in same tribe with *M. sativa*. Seeds were obtained from the Plant Genetic Resource Conservation Unit of USDA-ARS (Utah, USA) and Institute of Animal Science, Chinese Academy of Agricultural Science (Table 1). All the accessions used in this study were grown in the greenhouse of Institute of Animal Sciences, Chinese Academy of Agricultural Sciences. Total genomic DNA was extracted from young leaves of five plants each accession following the CTAB method [12]. DNA quality and quantity were checked in 1 % agarose gels and Unico UV-2000 Spectrophotometer (Unico, USA), respectively. The working concentration of DNA was adjusted to 50 ng/μL.

### Data mining for EST–SSR

A total of 12,371 *M. sativa* EST sequences were retrieved from the (NCBI, <http://www.ncbi.nlm.nih.gov/dbest/>, 3 May 2010).

The raw EST sequences were processed to remove the 5' or 3' end of poly A or poly T stretches by using the EST-trimmer software (<http://www.pgrc.ipk-gatersleben.de/misa/download/esttrimmer.pl>). After pre-treatment, the ESTs were assembled into larger ESTs using the CAP3 assembler software [13]. The criteria for assembly were an overlap size of 40 bp with 90 % identity. The identification and localization of potential SSRs were carried out using the MISA (<http://www.pgrc.ipk-gatersleben.de/misa>). The criteria for identifying SSRs for all possible combinations of core sequences were 6, 5, 4, 4, and 4 repeats for di-, tri-, tetra-, penta-, and hexanucleotides, respectively. Mononucleotide repeats were ignored because it was difficult to distinguish real mononucleotide repeats from polyadenylation products and single nucleotide stretch errors generated by sequencing.

Primer pairs were designed using the Batch Primer 3 software <http://probes.pw.usda.gov/cgi-bin/batchprimer3/batchprimer3.cgi>. The parameters for primers design were: (1) primer length from 18 to 24 with 20 as the optimum; (2) PCR product size from 100 to 350; (3) annealing temperature from 57 to 63 °C and with an optimum annealing temperature of 60 °C; (4) GC contents from 45 to 55 %, with 50 % as optimum.

### Amplification and detection of SSR alleles

PCR amplifications of genomic DNA was carried out in a 25 μL reaction volume in an ABI 2700 Thermal Cycler (Applied Biosystems, Foster City, CA, USA) containing 2.5 μL 10× PCR buffer (100 mM Tris–HCl pH 8.8 at 25 °C; 500 mM KCl, 0.8 % (v/v) Nonidet), 0.5 μL 10 mM dNTPs, 1 U of Taq DNA polymerase (Sangon, Shanghai, China), 0.5 μL 10 μM of each primer, 2.0 μL 25 mM MgCl<sub>2</sub> and 50 ng of template DNA. The following PCR profile was used: an initial denaturing for 8 min at 95 °C, followed by 10 cycles of 95 °C for 1 min, 60 °C for 30 s, and 72 °C for 45 s; and 20 cycles of 95 °C for 45 s, 55 °C for 30 s, and 72 °C for 45 s; a final extension at 72 °C for 6 min. Fluorescence-labeled primers were synthesized by Sangon (Shanghai, China). An ABI3730xl DNA Analyzer (Applied Biosystems, Foster City, CA, USA) was used to capture amplification products by a fluorescence detection system for SSR markers. Fragment sizes were determined using an internal size standard (LIZ500, ABI, USA), and the outputs were analyzed using GeneMapper software (<http://www.appliedbiosystems.com.cn/>).

### Allele frequency and diversity analysis

The EST–SSR bands were scored as presence (1) or absence (0). The expected heterozygosity (He) was calculated using GenALEX 6 [14]. Polymorphism information content (PIC) was calculated by PIC\_CALC 0.6 (<http://hi>

**Table 1** Germplasm accessions used in this study

Code	Accession ID	Species	Chromosome	Type	Country of origin
1	ZXY01	<i>M. sativa</i> ssp. <i>sativa</i>	32	Cultivar	China
2	PI 230783	<i>M. sativa</i> ssp. <i>sativa</i>	32	Cultivar	India
3	PI 183262	<i>M. sativa</i> ssp. <i>sativa</i>	32	Uncertain	Saudi Arabia
4	PI 170551	<i>M. sativa</i> ssp. <i>sativa</i>	32	Cultivar	Turkey
5	PI 243224	<i>M. sativa</i> ssp. <i>sativa</i>	32	Cultivated material	Iran
6	PI 430552	<i>M. sativa</i> ssp. <i>sativa</i>	32	Cultivar	Kazakhstan
7	PI 215671	<i>M. sativa</i> ssp. <i>sativa</i>	32	Cultivar	Sweden
8	PI 258756	<i>M. sativa</i> ssp. <i>sativa</i>	32	Cultivar	Russia
9	PI 202731	<i>M. sativa</i> ssp. <i>sativa</i>	32	Landrace	Spain
10	PI 206574	<i>M. sativa</i> ssp. <i>sativa</i>	32	Uncertain	Greece
11	PI 418217	<i>M. sativa</i> ssp. <i>sativa</i>	32	Cultivar	France
12	PI 207775	<i>M. sativa</i> ssp. <i>sativa</i>	32	Cultivar	UK
13	PI 199273	<i>M. sativa</i> ssp. <i>sativa</i>	32	Landrace	Poland
14	PI 435317	<i>M. sativa</i> ssp. <i>sativa</i>	32	Cultivar	Canada
15	PI 405068	<i>M. sativa</i> ssp. <i>sativa</i>	32	Cultivar	USA
16	PI 422516	<i>M. sativa</i> ssp. <i>sativa</i>	32	Landrace	Venezuela
17	PI 434600	<i>M. sativa</i> ssp. <i>sativa</i>	32	Cultivar	Argentina
18	PI 172188	<i>M. sativa</i> ssp. <i>sativa</i>	32	Cultivar	Peru
19	PI 445676	<i>M. sativa</i> ssp. <i>sativa</i>	32	Cultivar	Australia
20	PI 435231	<i>M. sativa</i> ssp. <i>sativa</i>	32	Cultivated material	Egypt
21	PI 516770	<i>M. sativa</i> ssp. <i>sativa</i>	32	Landrace	Morocco
22	PI 199279	<i>M. sativa</i> ssp. <i>sativa</i>	32	Cultivar	South Africa
23	4P450	<i>M. sativa</i> ssp. <i>sativa</i>	32	Wild material	Tanzania
24	8P4998	<i>M. sativa</i> ssp. <i>falcata</i>	32	Wild material	Russia
25	6P1639	<i>M. sativa</i> ssp. <i>coerulea</i>	16	Wild material	Russia
26	6P2111	<i>M. sativa</i> ssp. <i>varia</i>	32	Wild material	Kazakhstan
27	8P4818	<i>M. sativa</i> ssp. <i>hemicycla</i>	16	Wild material	Armenia
28	6P2513	<i>M. sativa</i> ssp. <i>glomerata</i>	32	Wild material	Russia
29	7A4	<i>M. minima</i>	16	Wild material	Turkmenistan
30	A1519	<i>M. lupulina</i>	16	Wild material	China
31	7P3603	<i>Trifolium repens</i>	32	Wild material	Estonia
32	7P3621	<i>Melilotus albus</i>	14	Wild material	Russia

Accession 2–22 from Plant Genetic Resource Conservation Unit of USDA-ARS (Utah, USA) and others from Institute of Animal Science, Chinese Academy of Agricultural Sciences (Beijing, China)

[baidu.com/luansheng1229/item/306815126d58e3a4feded5a4](http://baidu.com/luansheng1229/item/306815126d58e3a4feded5a4)) according to Botstein et al. [15]. A dendrogram was constructed based on Jaccard's similarity coefficient using the unweighted pair group method with arithmetic average (UP-GMA) with the SAHN module of NTSYS-pc [16].

## Results

### Frequency and distribution of alfalfa EST–SSRs

A total of 11,732 ESTs with an average length of 545 bp were obtained after pre-treatment analysis of the 12,371 alfalfa ESTs retrieved from NCBI. This represented approximately 5.96 Mb of alfalfa genomic. In total, 4,913 potential Unigenes with a mean length of 604 bp including 1,478 contigs and 3,435 singletons were generated. Of the

11,732 ESTs, 716 were identified containing SSRs. Of the 716 SSR-containing ESTs, 54 (7.5 %) contained two or more SSRs, and 39 (5.4 %) presented in compound formation. Total 774 SSRs were identified from the 716 unique ESTs. On an average, one SSR was identified per 7.7 kb. Of the 774 EST–SSRs, di-, tri-, tetra-, penta-, and hexanucleotide SSRs account for 26.1, 48.8, 11.5, 9.7, and 3.9 %, respectively (Table 2). As shown in Table 3, SSR length was mostly distributed from 12 to 20 bp, accounting for 81.4 % of total SSRs, followed by 21–30 bp length range (141 SSRs, 18.2 %). A total of 68 SSR motifs were identified. The di-, tri-, tetra-, penta- and hexanucleotide repeats had 3, 10, 19, 18 and 18 types, respectively (data not show). The most abundant type was the AG/CT repeats (17.2 %) followed by AAG/CTT repeats (15.1 %), ACC/GGT repeats (7.2 %), ATC/ATG repeats (7.1 %), AAC/GTT repeats (6.9 %), AC/GT (6.6 %), AGC/CTG repeats

(6.3 %), and AAAAC/GTTTT repeats (5.8 %). The remaining motifs presented a frequency of 27.8 %. It is interesting that there is no CG/GC repeats.

#### Polymorphic analysis and transferability of EST–SSR markers

A total of 100 primer pairs were designed. The remaining ESTs did not quality for primer design as the flanking the SSRs were too short (generally < 40 nucleotides) or inability to match the criteria for primer design. The 100 EST–SSR primer pairs were used to screen a panel of 28 alfalfa accessions (Table 1). Total 30 primer pairs were able to produce clear and expected size of amplicons (Table 4). The remaining 70 EST–SSR primer pairs either had no amplification products or produced a number of faint bands indicative of non-specific amplifications or gave larger and smaller amplicons than the expected size. The details of these 30 EST–SSR primer pairs are available in Table 4.

Of the 30 EST–SSR markers, 29 were polymorphic among the 28 alfalfa accessions. A total of 198 alleles were scored from the 29 polymorphic EST–SSR markers (Table 4). The number of alleles per marker varied from two (MsEST109) to 21 (MsEST66) with average of 6.8 alleles per marker (Table 4). The PIC ranged from 0.195 for MsEST84 to 0.896 for MsEST109 with an average value of 0.608 (Table 4). The expected heterozygosity ( $H_e$ ) varied from 0.068 for MsEST79 to 0.442 for MsEST43 with an average of 0.207 (Table 4).

To assess cross-species/genera transferability, 30 alfalfa EST–SSRs markers were also used to screen the four accessions from *M. minima*, *M. lupulina*, *T. repens*, and *M. albus*. The transferability rates of these EST–SSRs ranged from 100 % in *M. minima*, followed by 83.3 % in *M. lupulina*, 70 % in *M. albus*, and 63.3 % in *T. repens* (Table 5).

**Table 2** Summary of EST–SSR searching results

Searching items	Numbers
Total number of sequences examined	11,732
Total size of examined sequences (bp)	5,960,476
Total number of identified SSRs	774
Number of SSR containing sequences	716
Number of sequences containing more than 1 SSR	54
Number of SSRs present in compound formation	39
Di-nucleotide	202
Tri-nucleotide	378
Tetra-nucleotide	89
Penta-nucleotide	75
Hexa-nucleotide	30

**Table 3** Length distribution of EST–SSRs based on the number of repeat units

Repeat number	Di-	Tri-	Tetra-	Penta-	Hexa-	Total
4	0	0	69	62	26	157
5	0	218	15	9	4	246
6	74	101	3	4	0	182
7	35	33	1	0	0	69
8	26	12	1	0	0	39
9	19	10	0	0	0	29
10	11	2	0	0	0	13
11	3	2	0	0	0	5
12	9	0	0	0	0	9
13	7	0	0	0	0	7
14	5	0	0	0	0	5
≥15	13	0	0	0	0	13
Total	202	378	89	75	30	774

#### Genetic diversity analysis

The Jaccard's coefficients for the 28 alfalfa accessions were calculated based on the 29 EST–SSR markers. The lowest genetic similarity coefficients value (0.238) was observed between *M. sativa* ssp. *coerulea* (6P1639) and *M. sativa* ssp. *varia* (6P2111). A dendrogram was constructed based on the estimated Jaccard's coefficients by a total of 198 polymorphic bands (Fig. 1). The value of cophenetic correlation coefficient is 0.92, indicating a very good fit between the dendrogram cluster and the original similarity matrix. All the 23 accessions of *M. sativa* ssp. *sativa* were clustered together and clearly separated from other five subspecies (Fig. 1). Among the 23 *M. sativa* ssp. *sativa* accessions, the dendrograms did not show clear clustering pattern of geographically closer accessions in the present study indicating that the association between genetic similarity and geographical distance was less significant (Fig. 1). ZXY01 (a cultivar from China) and PI 445767 (a cultivar from Australia) are similar to each other, but different from others. Sixteen accessions grouped into the biggest cluster which including almost accessions from Europe, four accession from Africa, two accessions from Asia, two from South America and one from North America. Other five accessions one from USA, one from Russia, one from Argentina, two from Asia (Kazakhstan and Saudi Arabia) separated from each other. Among the different ploid level, two diploid alfalfa (6P1639 and 8P4818) are clearly different from other tetraploid alfalfa genotypes.

#### Discussion

In the present study, 716 (6.1 %) SSR-containing ESTs were identified from a total of 11,732 alfalfa ESTs. This

**Table 4** Characterization of 30 *Medicago sativa* EST–SSR markers

Locus name	Genbank ID	SSR motif	Primer sequence(5'–3')	T <sub>m</sub>	Expected size(bp)	Allele no.	Size range(bp)	He	PIC
MsEST1	50319233	(TAT)9	F: ACGGTGGTGTGTTATTGCT R: CCATATGCAACAGCCCTTAG	58.3	185	4	176–188	0.177	0.43
MsEST9	157976168	(TCT)9	F: ACAACATGGCTTCAACACCT R: TCACCCACCAAATCAGAATC	58.3	274	8	257–277	0.329	0.79
MsEST15	157975881	(ATG)8	F: AGACACCCCGACAGATACAA R: CACACTCTTCATTTCGCACTG	58	212	3	205–221	0.102	0.36
MsEST18	50318415	(TC)10	F: GGGATTACCACCATACAAA R: ATGTTGTCCAACCTCGGTGAC	58.4	235	8	219–241	0.245	0.73
MsEST26	50321265	(GAGAA)5	F: GGGGGTGGATAGTTAGCTC R: ACCCTATCATGCCAGTTGAA	58	227	1	215	0	0
MsEST28	50317321	(AAG)7	F: GGGATCTCACAACTTCCAAA R: ATTCTTCATCAGTGGGGTGA	57.9	222	8	206–231	0.332	0.81
MsEST30	50318027	(CATAA)4	F: GAGGATCAAACCTACCAACTCA R: TAGCTTAGGTGGCACACTCC	58	197	3	187–197	0.191	0.32
MsEST43	50317502	(TAAT)5	F: GGGATTCCCTTCAAACACAT R: GAGGGAAGCTCTGACACAAA	58	207	4	147–205	0.442	0.7
MsEST48	50319014	(TAGT)5	F: AAGACTTTGGAACCATCGTG R: TAACATCCCGTTCTGTGGTT	57.9	209	5	172–213	0.09	0.23
MsEST52	50319086	(CAT)7	F: CACCAATCTCAGGTGTGTGA R: GGATCAGCTTCAAACCAAG	57.4	129	6	88–130	0.232	0.63
MsEST53	50317832	(TGC)7	F: GGATCGTATCTGCTGCGAC R: GTTCCAAACGTTGGAGTGAC	58	111	7	101–119	0.16	0.7
MsEST61	157976664	(AG)12	F: GGACACAACACCTCACACAT R: TGGCTTCTTCCTTATCATCG	57.9	128	7	117–132	0.186	0.69
MsEST63	50318069	(ATG)11	F: AAGGTCAATGAAACTGCTGC R: TGCACCTCAAATAGCACAGC	57.7	220	8	209–227	0.2	0.71
MsEST66	7581455	(ACC)7	F: CATCAATGCTACTTTCCTAGCC R: CCCAAAAGCACTTCACTACC	57.3	150	2	147,153	0.226	0.36
MsEST79	157976299	(AAACGA)4	F: GTCACACAGTTTTGGTCTGCT R: AAAACGAAACCACTCAGACG	57.9	308	5	292–313	0.068	0.38
MsEST83	157975732	(GCA)8	F: ATTACGCGACCATCATCATC R: CTAACAGAGAGTTTCCCGCA	58.1	339	6	329–344	0.098	0.55
MsEST84	157975252	(ACC)7	F: TCATCAGCAACATCATCGAC R: GTCGGAGACATGAACCTTGG	58.1	182	2	176,182	0.069	0.2
MsEST88	157975003	(AGA)8	F: CTCTGAAGGAGATGGCAAGA R: AAACAGAAGTGGAGCGAGAA	57.7	277	11	255–308	0.097	0.68
MsEST92	157974580	(TCTTTC)4	F: GCATTGCACCATAAACCCCTA R: CACTGCATCATCATCATCGT	58	249	7	252–276	0.257	0.72
MsEST94	157974449	(CGTG)7	F: AACTCCAAACTCAGAACC R: TCCTCGTTCTCGAACTTGTC	58	178	6	139–180	0.294	0.74
MsEST99	157973731	(TCT)9	F: CGTCACTTCAACATCAGAA R: TAGTTGGACTTGCTTCCCTG	58	248	9	242–260	0.267	0.8
MsEST100	136141817	(TCT)7	F: TCGAACCCCTTGGTCACTATC R: CTGAAGTTCAACGGAAGGAA	57.9	112	5	108–120	0.307	0.68
MsEST102	81485488	(AGAA)6	F: CATTAGCAGGGGAAAACAGA R: CTTCCCTCCCTTGCTTTCTTT	57.7	221	4	198–217	0.137	0.32
MsEST107	50321576	(AGTGA)5	F: TTGAAAACCTCAGTCTTGCCC R: AGAGCGTAATCACGAATTG	57.8	147	5	137–168	0.108	0.51
MsEST109	50321181	(GTT)7-(ACATT)4	F: CTGTCCTTCTCTGCAACGC R: GCGTATCCAACAACCCTTT	57.6	320	21	282–342	0.153	0.9
MsEST110	50321113	(AAAATC)4	F: GATTGATTCGGTTACCGTTG R: GTAATCATCGTTCGGCTCTGT	57.8	155	9	143–191	0.281	0.81
MsEST128	50318223	(TTC)7	F: GGTAACGCCACAGAGTTC R: GTTGTCCATTGGAGTTGGAG	58	203	5	191–210	0.234	0.62

**Table 4** continued

Locus name	Genbank ID	SSR motif	Primer sequence(5′–3′)	T <sub>m</sub>	Expected size(bp)	Allele no.	Size range(bp)	He	PIC
MsEST130	50317942	(TC)13	F: GACAACCTCAAACCGTCCATC R: GATCGGAAGAGGATTTGGT	58	143	15	130–162	0.19	0.86
MsEST137	50317397	(CT)14	F: GGGGCGACTCTTCATATTTT R: GAGAGGATGAAGAGGGAAGC	58	154	10	140–159	0.224	0.79
MsEST143	50317066	(GAA)7	F: TCGAAACGGTTACCAAAGAG R: TGTTGTTGATGATGACGAGG	58	208	5	199–210	0.315	0.66

**Table 5** Cross-species/genera transferability of 30 *M. sativa* EST–SSR markers

Locus name	<i>M. minima</i>	<i>M. lupulina</i>	<i>Trifolium repens</i>	<i>Melilotus albus</i>
MsEST1	176,182	194	176	179
MsEST9	262	238,244	242,251	254
MsEST15	205	199	205	202
MsEST18	205,221	188,197	197	197
MsEST26	215	215	215	215
MsEST28	209,221	206,221	206,212	212,218
MsEST30	197	188	197	197
MsEST43	194,201	192,194	194,198	190
MsEST48	208	149	NA	NA
MsEST52	118,130	116,124	NA	NA
MsEST53	107,110	107	107	99,104
MsEST61	121	121	124	121
MsEST63	215,218	NA	NA	NA
MsEST66	140,147	142	NA	122
MsEST79	297	287	NA	NA
MsEST83	338	335	NA	NA
MsEST84	182	180	184	184
MsEST88	294	296,300	296,300	292
MsEST92	272,276	269,272	269	266
MsEST94	164,180	160	144	157
MsEST99	250,252	NA	NA	210
MsEST100	111,114	108	108,114	103
MsEST102	205,217	220	217	212
MsEST107	141,148	148	148	141
MsEST109	310,333	NA	NA	NA
MsEST110	143,150	NA	NA	NA
MsEST128	210	203,206	210	200
MsEST130	138,140	NA	NA	130
MsEST137	146,153	153	NA	NA
MsEST143	201,204	199	204	NA
Transferability (%)	100	83.3	63.3	70

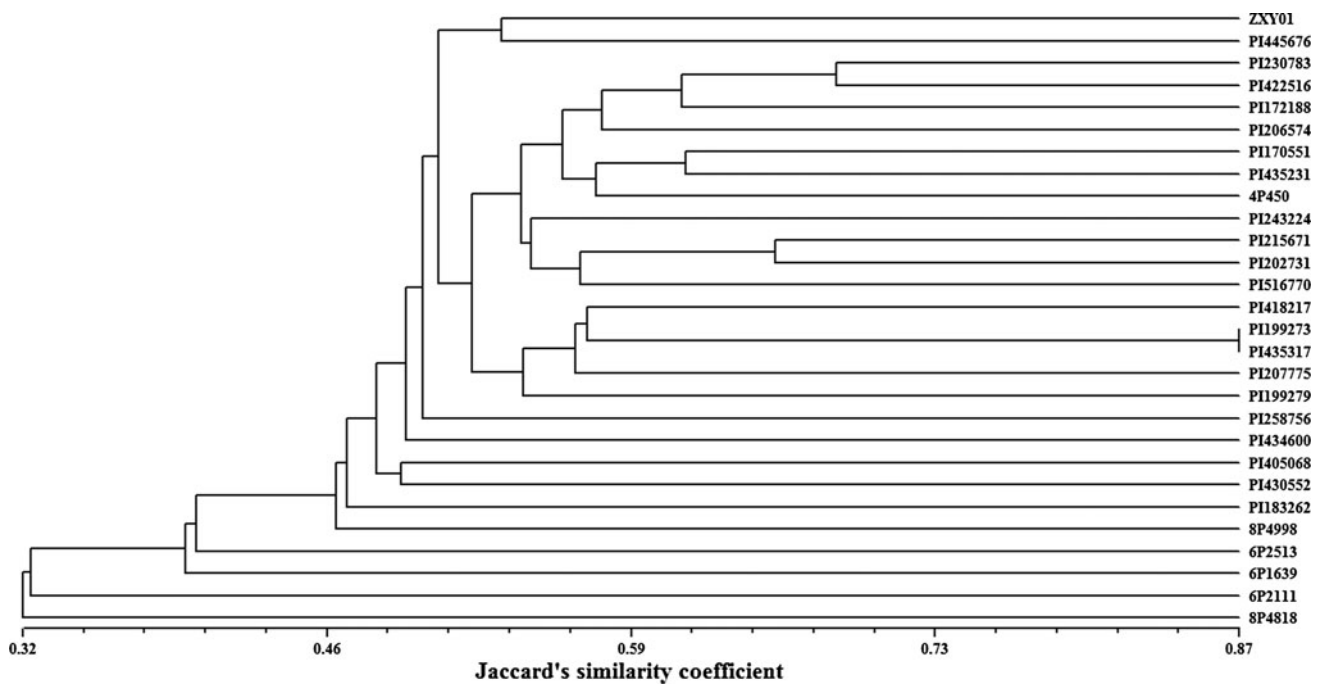
NA no amplification

result indicated that abundance of SSRs for alfalfa ESTs was higher than that for *M. truncatula* (3.0 %) [10]. Total 774 EST–SSR markers were identified from the 716 SSR-containing ESTs. The frequency of alfalfa EST–SSRs was about one EST–SSR marker per 7.7 kb. This was similar to the frequency reported in *Coffea* (7.73 kb) [17], peanut (7.3 kb) [18], sweet potato (7.1 kb) [19], cassava (7.0 kb) [20] but higher than that lotus (13 kb) [21]. A lower frequency was found in rubber tree (2.25 kb) [22], pepper (3.8 kb) [23], and in tea (3.5 kb) [24]. However, a direct comparison of abundance estimation and frequency occurrence of SSR in different reports is difficult due to the fact that the estimates were dependent on the SSR search criteria, the size of the dataset, the database mining tools and the EST sequence redundancy [25].

In previous studies, di- and tri-nucleotide repeats were generally the dominant motif found in many species. Di-nucleotide repeats were the dominant repeat in EST–SSR in cassava [20], rubber tree [22], tea [24], coffee [26], and physic nut [27]. Tri-nucleotide repeats were found the most abundant repeat motif in EST–SSR in peanut [18], sweet potato [19], chickpea [28], and castor bean [29]. In the present study, tri-nucleotide repeats (48.8 %) was the most abundant motif and followed by di-nucleotide (26.1 %). AG/CT (17.2 %) and AAG/CTT (15.1 %) were the most frequent di- and trinucleotide motif types, respectively, in this study. Similar results have been reported in many dicotyledonous species [20–22, 28, 29]. In the present study, there is no GC dinucleotide SSRs were detected, which agree with other reports that no GC dinucleotide SSRs were found [10, 22, 26, 30].

Total 29 polymorphic EST–SSR markers were scored for 28 alfalfa accessions. The PIC values across the 28 accessions ranged from 0.195 to 0.896, and most of the markers (72.4 %), indicating a high level of polymorphism. EST–SSRs derived from *M. truncatula* showed high level of transferability to relative species including *M. sativa* [10]. Similarly, the EST–SSRs derived from alfalfa also showed high level of transferability in this study.





**Fig. 1** The UPGMA tree of the 28 alfalfa accessions based on 29 EST-SSR markers

## Conclusion

In this study, we developed EST-SSR markers for alfalfa using *M. sativa* EST database. The alfalfa EST-SSR markers showed high level of polymorphism and were highly transferable across a number of distantly related species. As the enrichment of alfalfa ESTs, more genome wide EST-SSR markers could be developed. These EST-SSR markers will facilitate marker-trait association, QTL mapping, and genetic diversity analysis.

**Acknowledgments** This research was funded by a Grant from National Key Technology R&D Program (2011BAD17B01) of People's Republic of China and the Fundamental Scientific Research Fund from Chinese Academy of Agricultural Sciences of People's Republic of China (2011cj-15).

## References

- Levinson G, Gutman GA (1987) Slipped-strand mispairing—a major mechanism for DNA-sequence evolution. *Mol Biol Evol* 4:203–221
- Powell W, Machray G, Provan J (1996) Polymorphism revealed by simple sequence repeats. *Trends Plant Sci* 1:215–222
- Toth G, Gaspari Z, Jurka J (2000) Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Res* 10:967–981
- Sakiroglu M, Doyle JJ, Brummer EC (2010) Inferring population structure and genetic diversity of broad range of wild diploid alfalfa (*Medicago sativa* L.) accessions using SSR markers. *Theor Appl Genet* 121(3):403–415
- Julier B, Flajoulot S, Barre P, Cardinet G, Santoni S, Huguet T, Huyghe C (2003) Construction of two genetic linkage maps in cultivated tetraploid alfalfa (*Medicago sativa*) using microsatellite and AFLP markers. *BMC Plant Biol* 3:9
- Sledge MK, Ray IM, Jiang G (2005) An expressed sequence tag SSR map of tetraploid alfalfa (*Medicago sativa* L.). *Theor Appl Genet* 111:980–992
- Choi HK, Kim D, Uhm T, Limpens E, Lim H, Mun JH, Kalo P, Penmetsa RV, Seres A, Kulikova O, Roe BA, Bissling T, Kiss GB, Cook DR (2004) A sequence-based genetic map of *Medicago truncatula* and comparison of marker colinearity with *M. sativa*. *Genetics* 166:1463–1502
- Luo ZW, Hackett CA, Bradshaw JE, McNicol JW, Milbourne D (2001) Construction of a genetic linkage map in tetraploid species using molecular markers. *Genetics* 157:1369–1385
- He C, Xia ZL, Campbell TA, Bauchan GR (2009) Development and characterization of SSR markers and their use to assess genetic relationships among alfalfa germplasm. *Crop Sci* 49:2176–2186
- Eujayl I, Sledge MK, Wang L, May GD, Chekhovskiy K, Zwonitzer JC, Mian MAR (2004) *Medicago truncatula* EST-SSRs reveal cross-species genetic markers for *Medicago* spp. *Theor Appl Genet* 108:414–422
- Blondon F, Marie D, Brown S, Kondorosi A (1994) Genome size and base composition in *Medicago sativa* and *M. truncatula* species. *Genome* 37:264–270
- Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull* 19:11–15
- Huang X, Madan A (1999) CAP3: a DNA sequence assembly program. *Genome Res* 9:868–877
- Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in excel. population genetic software for teaching and research. *Mol Ecol Notes* 6:288–295
- Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet* 32:314–331
- Rohlf FJ (2000) NTSYS-PC version 2.10s numerical taxonomy and multivariate analysis system. Exeter publications, Setauket
- Poncet V, Rondeau M, Tranchant C, Cayrel A, Hamon S, de Kochko A, Hamon P (2006) SSR mining in coffee tree EST

- databases: potential use of EST–SSRs as markers for the *Coffea* genus. *Mol Genet Genomics* 276:436–449
18. Liang X, Chen X, Hong Y, Liu H, Zhou G, Li S, Guo B (2009) Utility of EST-derived SSR in cultivated peanut (*Arachis hypogaea* L.) and *Arachis* wild species. *BMC Plant Biol* 9:35
  19. Wang ZY, Li J, Luo ZX, Huang LF, Chen XL, Fang BP, Li YJ, Chen JY, Zhang XJ (2011) Characterization and development of EST-derived SSR markers in cultivated sweet potato (*Ipomoea batatas*). *BMC Plant Biol* 11:139
  20. Raji AA, Anderson JV, Kolade OA, Ugwu CD, Dixon AG, Ingelbrecht IL (2009) Gene-based microsatellites for cassava (*Manihot esculenta*): prevalence, polymorphisms, and cross-taxa utility. *BMC Plant Biol* 9:118
  21. Pan L, Xia Q, Quan Z, Liu H, Ke W, Ding Y (2010) Development of novel EST–SSRs from sacred lotus (*Nelumbo nucifera* Gaertn) and their utilization for the genetic diversity analysis of *N. nucifera*. *J Hered* 101:71–82
  22. 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(1):85–97
  23. Yi G, Lee JM, Lee S, Choi D, Kim BD (2006) Exploitation of pepper EST–SSRs and an SSR-based linkage map. *Theor Appl Genet* 114:113–130
  24. Sharma RK, Bhardwaj P, Negi R, Mohapatra T, Ahuja PS (2009) Identification, characterization and utilization of unigene derived microsatellite markers in tea (*Camellia sinensis* L.). *BMC Plant Biol* 9:53
  25. Varshney RK, Graner A, Sorrells ME (2005) Genic microsatellite markers: features and applications. *Trends Biotechnol* 23:48–55
  26. Aggarwal RK, Hendre PS, Varshney RK, Bhat PR, Singh L, Krishnakumar V (2007) Identification, characterization and utilization of EST-derived genic microsatellite markers for genome analyses of coffee and related species. *Theor Appl Genet* 114(2):359–372
  27. Yadav HK, Ranjan A, Asif MH, Mantri S, Sawant SV, Tuli R (2011) EST-derived SSR markers in *Jatropha curcas* L.: development, characterization, polymorphism, and transferability across the species/genera. *Tree Genet Genomes* 7:207–219
  28. Choudhary S, Sethy NK, Shokeen B, Bhatia S (2009) Development of chickpea EST–SSR markers and analysis of allelic variation across related species. *Theor Appl Genet* 118:591–608
  29. Qiu LJ, Yang C, Tian B, Yang JB, Liu AH (2010) Exploiting EST databases for the development and characterization of EST–SSR markers in castor bean (*Ricinus communis* L.). *BMC Plant Biol* 10:278
  30. Gupta S, Prasad M (2009) Development and characterization of genic SSR markers in *Medicago truncatula* and their transferability in leguminous and non-leguminous species. *Genome* 52:761–771