Theor Appl Genet (2004) 108:414–422
DOI 10.1007/s00122-003-1450-6

ORIGINAL PAPER

I. Eujayl · M. K. Sledge · L. Wang · G. D. May · K. Chekhovskiy · J. C. Zwonitzer · M. A. R. Mian

Medicago truncatula EST-SSRs reveal cross-species genetic markers for *Medicago* spp.

Received: 31 December 2002 / Accepted: 14 August 2003 / Published online: 16 September 2003

Abstract Expressed sequence tags (ESTs) are important resources for gene discovery and molecular marker development. From over 147,000 ESTs of *Medicago truncatula,* we have identified 4,384 ESTs containing perfect simple sequence repeats (EST-SSR) of di-, tri-, tetra- or pentanucleotides. Six hundred sixteen primer pairs (PPs) were designed and screened over a panel of eight genotypes representing six *Medicago* spp. and subspecies. Nearly, 74% (455) of the PPs produced characteristic SSR bands of expected size length in at least one *Medicago* species. Four hundred six (89%) of these 455 PPs produced SSR bands in all eight genotypes tested. Only 17 PPs were *M. truncatula* -specific. High levels of polymorphism (>70%) were detected for these markers in alfalfa, *M. truncatula*, and other annual medics. About 48% of the reported markers are part of gene transcripts linked to putative functions. Our results indicate that the SSR markers developed from *M. truncatula* ESTs are valuable genetic markers for the *Medicago* genus. These markers will be useful in establishing the genomic relationships of *M. truncatula* to important forage legume crops such as alfalfa and other annual medics.

Electronic Supplementary Material Supplementary material is available in the online version of this article at http://dx.doi.org/10.1007/s00122-003-1450-6.

Communicated by H.C. Becker

I. Eujayl · M. K. Sledge · L. Wang · G. D. May · K. Chekhovskiy · J. C. Zwonitzer · M. A. R. Mian (✉)
The Samuel Roberts Noble Foundation,
2510 Sam Noble Parkway, Ardmore, OK 73401, USA
e-mail: rmmian@noble.org
Tel.: +1-580-2246840
Fax: +1-580-2246802

Present address:
I. Eujayl, Department of Food Science & Human Nutrition,
Washington State University,
E202 Food Quality Building, WWQL, Pullman, WA 99164, USA

Introduction

The co-dominant simple sequence repeat (SSR) or microsatellite markers have become the marker class of choice for molecular mapping and breeding of many plant species. The development of SSR markers from genomic libraries is expensive and inefficient. However, with the availability of large numbers of expressed sequence tags (ESTs) and other DNA sequence data, development of SSR markers through data mining has become a fast, efficient, and low-cost option for many plant species. This is due to the fact that the time-consuming and expensive processes of generating genomic libraries and sequencing of large numbers of clones for finding the SSR containing DNA regions are not needed in this approach. A modest 1% to 5% of the ESTs in various plant species have been found to have SSRs of suitable length (20 nt or more) for marker development (Kantety et al. 2002). It should be possible to find a large number of these SSRs in an organism for which a great number of ESTs have been generated.

The scope of EST-SSR marker development is limited to species for which sequencing databases already exist. An alternative efficient approach for SSR marker development in species that lack large enough sequence databases could be the utilization of SSR markers from related species for which large numbers of SSR markers have been developed. There are a growing number of strategies to utilize databases to cross-reference the biological and genetic information known in one organism to another, such as yeast to mammals (Tugendreich et al. 1993). Moreover, there is evidence that suggests high rates of transferability of SSR markers across species (>50%) within a genus (Peakall et al. 1998; Gaitán-Solís et al. 2002). However, the transferability of genomic SSR markers across genera and beyond is low (White and Powell 1997; Peakall et al. 1998; Roa et al. 2000). Since EST-SSR markers are derived from transcribed regions of the DNA, it is expected that these markers would be more conserved and have a higher rate of transferability than genomic SSR markers (Scott et al. 2000). In contrast, the

conserved nature of EST-SSRs may limit their polymorphism. EST-SSRs also have a high probability of being associated with gene expression and gene function. For example, the *waxy* gene in rice has been found to contain a $(CT)_n$ microsatellite whose length polymorphism is associated with amylose content (Ayers et al. 1997). The functions of many of the SSR containing transcripts can be predicted through homology searches from the sequence databases. The SSR markers developed from DNA sequences with putative functions can be tested for their links to predicted phenotypes.

Many of the worlds forage legumes, including alfalfa (*M. sativa*) belong to the genus *Medicago* (Barnes et al. 1988). This genus is part of the Galegoid phylum that includes a number of important legume crops such as pea, chickpea, clover, and lentil (Doyle et al. 1996). Genetic analysis of many of the crop legumes, including alfalfa, remains difficult due to features such as polyploidy, obligate cross-pollination, large genome sizes, and the lack of efficient methods of transformation (Thoquet et al. 2002). Thus, the establishment of a model system for studying the legume molecular biology is warranted, and *M. truncatula* has emerged as the most likely candidate for such a model plant.

Due to the relative ease of discovering genes and the determination of gene function in model species, comparative genomics has become an important strategy for extending genetic information from model species to more complicated species (Paterson et al. 1995). Until now, comparative genomics efforts have relied heavily on the hybridization based RFLP technique. The resolution of these comparative maps is generally low for determination of microsynteny (Kilian et al. 1997). The application of a PCR-based, co-dominant marker system for comparative genomics would be highly desirable, because such a marker system could increase the efficiency of transferring genetic information across species. The EST-derived markers (e.g., EST-SSR) can be used to cross-reference genes between species for enhancing the resolution of comparative genomics studies and identifying conserved genomic regions among species (Kantety et al. 2002). The availability of SSR markers in most *Medicago* spp. (including alfalfa) and many important food legumes (including pea, lentil, and peanut) remains very low, with the exception of soybean. Until recently, only a handful of linkage maps have been constructed either for diploid or tetraploid alfalfa (Kaló et al. 2000). Also, these maps were constructed mainly with hybridization-based RFLP, or PCR-based dominant markers, e.g., RAPD and AFLP. There are very few SSR markers available for molecular mapping in alfalfa (Diwan et al. 1997). Diwan et al. (2000) reported mapping of 10 SSR loci in diploid and tetraploid alfalfa. They concluded that the co-dominant SSR marker would be a powerful tool for molecular mapping in tetraploid alfalfa as well as in other polyploid species.

Thoquet et al. (2002) reported a high level of macrosynteny between *M. truncatula* and diploid alfalfa genomes through comparative mapping. Huguet et al. (2001) reported mapping of EST-SSR markers in their *M. truncatula* mapping population in order to facilitate comparative mapping studies and identification of quantitative trait loci. Danesh et al. (2002) reported the development of 84 EST-SSR markers from the *M. truncatula* EST database. In this study, they also demonstrated the utility of the EST-SSR markers in mapping quantitative trait loci for powdery mildew resistance in *M. truncatula*. Baquerizo-Audiot et al. (2001) developed five SSR primer pairs from *M. truncatula* genomic sequences and four primer pairs from the *M. truncatula* ESTs in GenBank. In addition to *M. truncatula,* they tested the SSR markers on three other *Medicago* spp. (*M. littoralis*, *M. sativa* ssp. *sativa* cv. Coussouls, and *M. sativa* ssp. *sativa* natural population from Spain) and three additional legume species (*Pisum sativum*, *Phaseolus vulgaris*, and *Glycine max*). All primer sets in their study yielded PCR products within the expected size ranges in all of the *Medicago* spp., but outside the *Medicago* genus they failed to produce any PCR product.

The current study was designed to develop a large set of SSR markers from the *M. truncatula* EST database for possible use in legume crops both within and outside the *Medicago* genus.

Materials and methods

EST database access and detection of SSRs

The *M. truncatula* EST sequences used in this study were extracted from NCBI's dbEST (ftp://ftp.ncbi.nih.gov/blast/db/). Approximately 147,000 ESTs were searched for identification of SSRs using the PERL program, Simple Sequence Repeat Identification Tool (SSRIT), downloaded from the Cornell University web site http://www.gramene.org/gramene/searches/ssrtool. The SSRIT program was run on a local computer and the parameters were set for detection of perfect di-, tri-, tetra-, and pentanucleotide motifs with a minimum of 10, 7, 5, and 4 repeats, respectively. The following information was extracted from the SSRIT output and TIGR's *Medicago truncatula* Gene Index (MtGI) using PERL programs – EST GenBank identifiers, TIGR's tentative consensus numbers (TCs), repeat motifs, number of repeats, start- and end positions of the SSRs within the respective ESTs, and gene functional annotations.

Primer design

A subset of the sequences from the SSRs containing ESTs was selected for designing primers based on the following criteria: (1) the start and the end positions of the SSRs should be at least 50 bases from the 5′ and 3′ ends of the sequence, respectively, and (2) only one sequence from each TC was selected from the sets of SSR containing ESTs that had the same motifs. The first criterion was chosen to enhance the chances of designing high quality primers. The second criterion was selected to avoid designing primers for SSRs from redundant sequences. The Primer3 software (freely available at http://www.genome.wi.mit.edu/genome_software/other/primer3.html) was used to design the primers. The major parameters for designing the primers were set as follows: primer length from 18 to 24 with 20 as the optimum, PCR product size from 125 to 300, optimum annealing temperature 60°C, and GC contents from 40% to 70%, with 50% as optimum. The primer pairs were custom synthesized by Qiagen/Operon Technologies (Alameda, Calif.).

Table 1 The list of plant genotypes used in this study

Serial no.	Common name	Ploidy level	Scientific name	Cultivar/accession name
1	Diploid alfalfa	2*n*=2*x*=16	*M. sativa* ssp *coerulea*	PI440501
2		2*n*=2*x*=16	*M. sativa* ssp *sativa*	W2xiso
3		2*n*=2*x*=16	*M. sativa* ssp *sativa*	*PI212798*[a]
4		2*n*=2*x*=16	*M. sativa* ssp *sativa*	*PI206453*
5	Tetraploid alfalfa	2*n*=4*x*=32	*M. sativa* ssp *sativa*	Alfagraze
6		2*n*=4*x*=32	*M. sativa* ssp *sativa*	Amerigraze (702)
7		2*n*=4*x*=32	*M. sativa* ssp *sativa*	Cimmaron –31
8		2*n*=4*x*=32	*M. sativa* ssp *sativa*	CUF101 (95–650)
9		2*n*=4*x*=32	*M. sativa* ssp *falcata*	*Parent*$_1$[b]
10		2*n*=4*x*=32	*M. sativa* ssp *sativa*	*Parent*$_2$ [b]
11		2*n*=4*x*=32	*M. sativa* ssp *sativa*	F_1 (parent$_1$ × parent$_2$)
12		2*n*=4*x*=32	*M. sativa* ssp *sativa*	PI516900
13		2*n*=4*x*=32	*M. sativa* ssp *sativa*	PI464741
14		2*n*=4*x*=32	*M. sativa* ssp *sativa*	–[c]
15		2*n*=4*x*=32	*M. sativa* ssp *sativa*	–
16	Annual medics	2*n*=2*x*=16	*M. polymorpha*	*Armadillo*
17		2*n*=4*x*=16	*M. rugosa*	*Sapo*
18		2*n*=2*x*=16	*M. blancheana*	PI455223
19		2*n*=2*x*=16	*M. truncatula*	*Jemalong A-17*
20		2*n*=2*x*=16	*M. truncatula*	*Borung*
21		2*n*=2*x*=16	*M. truncatula*	*Caliph*
22		2*n*=2*x*=16	*M. truncatula*	DZA-105
23		2*n*=2*x*=16	*M. truncatula*	ESP095
24		2n=2*x*=16	*M. truncatula*	GRC020

[a] *Italic* indicates the genotype was used in the initial screening panel
[b] *Parent*$_1$ and *Parent*$_2$ are the two alfalfa genotypes used by Dr. Ian Ray for developing two BC_1 populations of tetraploid alfalfa used in this study
[c] Cultivar/accession name not available

Plant materials

A total of 24 genotypes from six species and subspecies were selected to represent the *Medicago* genus (Table 1). These genotypes represent the diploids and the tetraploids, the annuals and the perennials, and the cross-pollinated and the self-pollinated species of *Medicago*. A subset of eight individuals was used for the initial screening of SSR amplification products across alfalfa and annual medics (Table 1). The two diploid alfalfa genotypes used in this panel of eight are from Iran (PI212798) and Turkey (PI206453). The two tetraploid alfalfa genotypes in the panel are the parents of two BC_1 mapping populations developed by Dr. Ian Ray (New Mexico State University, Las Cruces, N.M.). The set of 24 genotypes of the *Medicago* genus was used for estimating the number of SSR alleles, and allele sizes for 39 markers. Twenty-one $F_{2:3}$ families and the two parents of the *M. truncatula* mapping population developed by Dr. Douglas Cook (University of California-Davis, Calif.), were used for testing the inheritance of EST-SSR markers in *M. truncatula.* Ten BC_1 progeny of each of two BC_1 populations developed by Dr. Ian Ray, along with the two parents and the F_1 progeny, were used for testing the inheritance of the markers in tetraploid alfalfa.

DNA isolation, PCR conditions, and amplification of SSR alleles

Approximately 200 mg of tissue from young leaves of each plant was collected in a 2.0 ml Eppendorf tube, immediately frozen in liquid nitrogen and ground to fine powder using a Mixer Mill Type MM 300 (RETSCH, Hann, Germany). The DNA was extracted using Qiagen DNeasy Plant Mini kit (Qiagen, Valencia, Calif.). The DNeasy protocol was used with the following modifications: 500 ul of AP1 buffer, 5 ul of RNase A, 165 ul AP2 buffer, and 90 ul of AE buffer was added to each tube. The DNA concentrations were quantified using a Hoefer Dyna Quant 200 (Amersham Biosciences, Piscataway, N.J.) DNA fluorometer. Twenty nanograms of DNA was used as template for each PCR reaction. The PCR reactions were run under standard conditions for all primers for the screening using one unit of *Taq* DNA polymerase with 10X PCR reaction buffer (Invitrogen Life Technologies, Carlsbad, Calif.) or one unit of AmpliTaq Gold with GeneAmp PCR bufferII (Applied Biosysytems/Roche, Branchburg, N.J.), 3 mM $MgCl_2$, 200 μM of dNTPs, 0.2 mM of each primers in a 10-μl reaction. After 10 min at 95°C, 40 cycles were performed with 50 s at 95°C, 50 s at a temperature between 55°C to 61°C (the optimum annealing temperature for the respective primer pair was used for each PCR reaction), 90 s at 72°C, and a final extension step of 10 min at 72°C. All primers were initially screened using *Taq* DNA polymerase. AmpliTaq Gold was used for SSR primers that amplified non-specific bands from the use of *Taq* DNA polymerase. The PCR products were resolved on 6% polyacrylamide denaturing gels (Gel Mix 6, Invitrogen LifeTechnologies). The gels were silver-stained using Silver Sequence Kit (Promega, Madison, Wis.) for SSR band detection.

Allele scoring and evaluation of polymorphism

The PCR products from Jemalong (A-17) were used as the standard for allele average size and banding morphology (cluster of stutter). The size is reported for the most intensely amplified band for each SSR or the average of the stutter if the intensity was the same using 10-bp DNA ladder (Invitrogen Life Technologies) as the reference point. Null alleles were assigned to genotypes with confirmed no amplification product under the standard conditions.

The polymorphism was determined according to the presence or absence of the SSR locus. The polymorphism information content (PIC) was calculated using the formula

$$\mathrm{PIC} = 1 - \sum_{j=1}^{n} Pij2 \qquad (1)$$

according to Botstein et al. (1980) and Anderson et al. (1993), where *Pij* is the frequency of the *i*th allele for marker *i* and summation extends over *n* alleles.

Determination of genetic distances among *M. truncatula* genotypes

The genetic distances among the six *M. truncatula* genotypes (Table 1) were calculated according to the Nei and Li (1979) similarity definition: $Sij=2a/(2a+b+c)$, where *Sij* is the similarity between two individuals, *I* and *j*; *a* is the number of SSR bands present in both samples *i* and *j*; *b* is the number of SSR bands present in *i* and absent in *j*; and *c* is the number of SSR bands present in *j* and absent in *i*. The distance matrix was used for construction of a dendrogram by the UPGMA method (SAS Institute 1989).

Results

Characteristics of *M. truncatula* EST-SSRs and marker development

We identified 4,384 SSR containing ESTs for di-, tri-, tetra-, and pentanucleotide motifs from 147,000 *M. truncatula* ESTs (Table 2). Approximately 3% of the ESTs had such SSRs. The trinucleotide motifs were the most abundant type of SSRs found in the database (52%), followed by di- (30%), tetra- (9%), and pentanucleotide repeats (9%). The GA motif was the most frequent class of EST-SSRs found (Fig. 1). Unlike the genomic SSRs in soybean (Brown-Guedira et al. 2000) and other legumes, the AT–rich SSRs were low in number. In this data set, only few GC dinucleotide SSRs were detected.

Of the 616 PPs designed, 455 (74%) produced characteristic SSR bands in at least one of the six *Medicago* spp. tested. The sequences of these PPs along with the SSR marker name, GenBank EST ID, TC name, SSR repeat motif and number, observed SSR alleles size range, annealing temperature, plant spp. with SSR amplification, enzyme used for PCR reaction, and annotated gene function(s) for each EST clone are available as electronic supplementary material. Nearly, 89% (406) of these PPs produced SSR bands in all eight genotypes of the six *Medicago* species and subspecies. Seventeen PPs were *M. truncatula* specific while 16 PPs worked only in *M. truncatula* and alfalfa. Similarly, 14 PPs worked in *M. truncatula* and other annual medics but not in alfalfa.

Table 2 The number and percent of di-, tri-, tetra-, and pentanucleotide SSRs discovered in 147,000 *M. truncatula* ESTs

Repeat type	Total number	Percent of total
Dinucleotide	1307	29.8
Trinucleotide	2259	51.5
Tetranucleotide	402	9.2
Pentanucleotide	416	9.5
Total number	4384	100

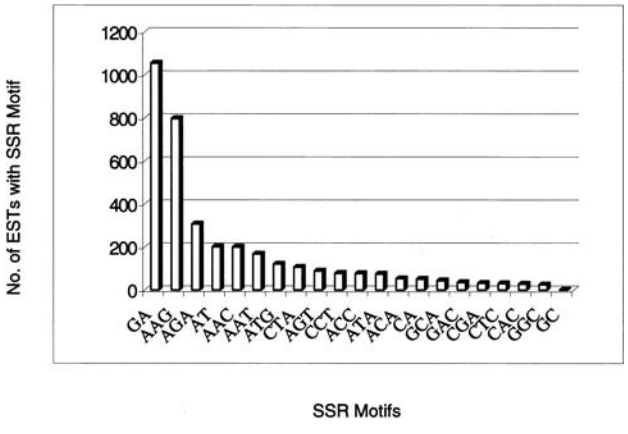

Fig. 1 The distribution of the 21 most abundant SSR motifs in the 4,384 EST-SSRs discovered in this study

Polymorphism of the SSR markers

The SSR markers reported in this study were highly polymorphic in *Medicago* spp. The screening on the panel of eight *Medicago* genotypes revealed that nearly 70% of the markers were polymorphic between the two diploid alfalfa genotypes as well as between the parents of the two BC_1 tetraploid alfalfa mapping populations developed by Dr. Ian Ray. The rate of polymorphism between the two annual medics 'Armadillo' (*M. polymorpha*) and 'Sapo' (*M. rugosa*) were 80%, while ~70% of the markers were polymorphic between the *M. truncatula* genotypes on the panel. The dominant scoring of the SSR bands (scored by the presence or absence of a particular band) yielded an average of 2.5 polymorphic bands per SSR marker between the pair of both the diploid and the tetraploid alfalfa genotypes.

A subset of 39 primer pairs were selected for screening over a larger set of 24 *Medicago* genotypes in order to obtain a better estimate of the number of alleles (bands/cluster of bands) and range of allele sizes of SSR loci amplified by these markers (Table 3). Figure 2 shows the SSR alleles amplified by markers, MtSSRNFAL03 and MtSSRNFBF02, over 24 *Medicago* genotypes as listed in Table 1. The diploid self-pollinated *M. truncatula* genotypes had one homozygous locus each, so did most of the annual medics species that are also self-pollinated diploids. However, the out-crossing diploid and tetraploid alfalfa genotypes had more than one discrete allele in most cases. The number of alleles detected by these markers varied from two to 11 (Table 3). The greatest variation in size range (155–281 bp) of SSR alleles was found for the marker MtSSRNFAW19. The polymorphism information content (PIC) values were calculated based on six *M. truncatula* genotypes for 18 markers (Table 4). The PIC values ranged from 0.28 to 0.83 with the average PIC value of 0.66. The average number of alleles per locus was four.

Genetic relationship among six *M. truncatula* genotypes

Based on 155 SSR markers scored from 39 primer pairs, the genetic distances (GD) between genotypes varied from 0.402 to 0.613 (Table 5). Jemalong A-17 and ESP095 had the minimum GD, while ESP095 and

Table 3 The number and molecular size ranges of alleles of 39 EST-SSR markers tested on 24 genotypes of *Medicago* spp as listed in Table 1

EST GenBank		Number of alleles	Range of alleles size (bp)
ID	NF[a] name		
AL373844	MtSSRNFAL03	8	208–240
AL376369	MtSSRNFAL05	6	160–190
AW559402	MtSSRNFAW01	4	173–185
AW690594	MtSSRNFAW03	5	185–220
AW691356	MtSSRNFAW04	5	181–200
AW694344	MtSSRNFAW05	2	169–176
AW695001	MtSSRNFAW06	4	152–165
AW696663	MtSSRNFAW09	5	142–185
AW774443	MtSSRNFAW11	8	202–240
BE239880	MtSSRNFBE01	8	150–181
BE248330	MtSSRNFBE02	6	117–148
BE323744	MtSSRNFBE04	3	260–280
BE999718	MtSSRNFBE05	2	180–183
BF005053	MtSSRNFBF01	6	160–220
BF636901	MtSSRNFBF02	9	145–192
BF640981	MtSSRNFBF03	3	204–210
BF650454	MtSSRNFBF06	6	176–180
BG451005	MtSSRNFBG01	2	170–176
BG580909	MtSSRNFBG04	4	176–189
AL365892	MtSSRNFAL09	4	194–218
AW685684	MtSSRNFAW16	8	142–174
AW691938	MtSSRNFAW20	4	198–210
BF005053	MtSSRNFBF07	7	166–220
BF005834	MtSSRNFBF08	5	198–210
BF648665	MtSSRNFBF12	5	200–222
BG580745	MtSSRNFBG20	8	180–248
BF518447	MtSSRNFBF19	3	192–204
BF646974	MtSSRNFBF20	8	230–312
BG455405	MtSSRNFBG24	6	145–202
AL378187	MtSSRNFAL18	3	200–210
AL382396	MtSSRNFAL19	6	192–220
AW689945	MtSSRNFAW30	4	134–145
BG457180	MtSSRNFBG29	7	175–200
BF651163	MtSSRNFBI10	6	174–214
AW691701	MtSSRNFAW19	11	155–281
AW736609	MtSSRNFAW23	8	152–182
BE315997	MtSSRNFBE07	9	154–206
BF631700	MtSSRNFBF09	3	304–310
BG587084	MtSSRNFBG15	10	159–216

Table 4 The polymorphism information content (PIC) value of 18 markers over six *Medicago truncatula* genotypes

GenBank ID	Marker name	No. of alleles	PIC value
AL374645	MtSSRNFAL04	3	0.50
AL382911	MtSSRNFAL06	2	0.28
AL384296	MtSSRNFAL07	3	0.67
AL366491	MtSSRNFAL08	5	0.78
AW695700	MtSSRNFAW07	5	0.78
AW559402	MtSSRNFAW01	3	0.61
AW691356	MtSSRNFAW04	4	0.67
BE239880	MtSSRNFBE01	4	0.72
BF050553	MtSSRNFBF01	5	0.78
AW695001	MtSSRNFAW06	3	0.61
AL373844	MtSSRNFAL03	4	0.72
AL376369	MtSSRNFAL05	5	0.78
AW696663	MtSSRNFAW09	6	0.83
BF650454	MtSSRNFBF06	3	0.61
BF636901	MtSSRNFBF02	5	0.78
BG451005	MtSSRNFBG01	2	0.44
BG452914	MtSSRNFBG03	3	0.61
BG580909	MtSSRNFBG04	3	0.66
Average	–	3.8	0.66

GRC020 had the maximum GD. The dendrogram generated with the UPGMA clustering method illustrates the closer genetic relationship between Jemalong A-17 and ESP095 compared to other *M. truncatula* genotypes in the study (Fig. 3).

Inheritance of markers in barrel medics and alfalfa

The inheritance of SSR bands was tested using mapping populations of *M. truncatula* and tetraploid alfalfa. The two parents and 21 $F_{2:3}$ families from the *M. truncatula* mapping population developed by Dr. Douglas Cook were used for testing the inheritance of 12 randomly selected polymorphic markers. The SSR alleles were found to be inherited from the parents to the $F_{2:3}$ progeny. Fig. 4 shows the inheritance of SSR alleles from the

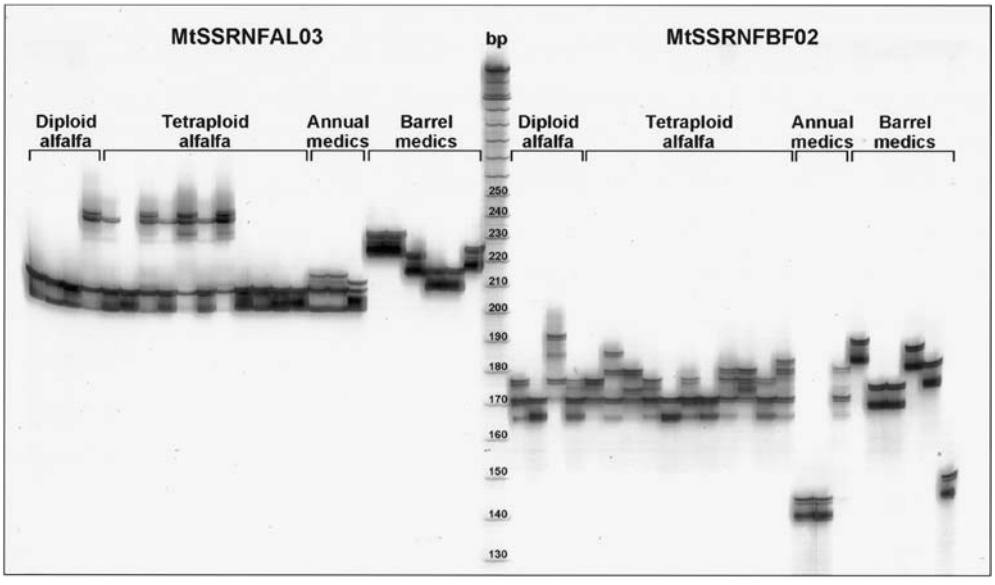

Fig. 2 The silver-stained polyacrylamide gel image of the SSR alleles amplified in 24 genotypes of diploid and tetraploid alfalfa, annual medics, and barrel medics by markers MtSSRNFAL03 and MtSSRNFBF02. The genotypes are in the same order as listed in Table 1. The *center lane* shows the size standard (10 bp DNA ladder)

Table 5 Genetic distances among six *Medicago truncatula* genotypes based on 155 EST-SSR markers

Genotype	Jemalong	Borung	Caliph	DZA105	ESP095
Borung	0.511[a]	–	–	–	–
Caliph	0.558	0.516	–	–	–
DZA105	0.577	0.518	0.577	–	–
ESP095	0.402	0.452	0.480	0.584	–
GRC020	0.608	0.608	0.567	0.579	0.613

[a] The dissimilarity coefficients are based on 155 EST-SSR markers. The dissimilarity coefficients can range from 0, when all markers for the two genotypes are identical, to 1, when all markers for the two genotypes are different

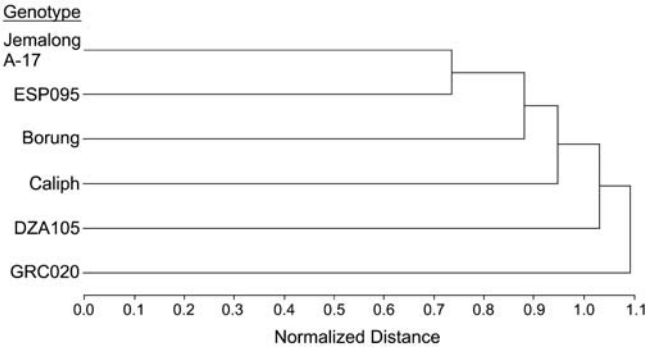

Fig. 3 The UPGMA tree of the six *Medicago truncatula* genotypes based on 155 EST-SSR markers

parents to the progeny amplified by marker MtSSRN-FAW19.

The inheritance of 200 polymorphic markers was tested using 10 BC_1 progeny of each of the two BC_1 populations of tetraploid alfalfa developed by Dr. Ian Ray. The polymorphic SSR bands could be tracked from the parents to the F_1 and to the BC_1 progeny (Fig. 5). Nearly, 40% of the polymorphic bands segregated in a 1:1 ratio (presence vs absence of a band) in both BC_1 populations. These markers will be very useful for linkage map construction in these BC_1 populations by using the single dose polymorphism approach (Wu et al. 1992).

Discussion

Discovery of SSRs in ESTs and primer development

The availability of the large number of *Medicago* ESTs was the incentive to search for only perfect SSRs. The total number of the perfect SSRs (4,384) identified in this study is an indication of the feasibility of searching into EST database of organisms to design SSR markers that have the characteristics of ideal markers for genetic studies. The EST-SSRs information revealed in this study is in agreement with findings by other researchers (Cho et al. 2000; Kantety et al. 2002). Kantety et al. (2002) searched 262,631 ESTs from five different grass (rice, maize, wheat, barley, and sorghum) databases for SSRs (di-, tri-, and tetranucleotide motifs with a minimum repeat length of 18 bp) and found that 3.2% ESTs contained SSRs. The proportion of di-, tri-, and tetranucleotide motifs found in that study was 19%, 72%, and

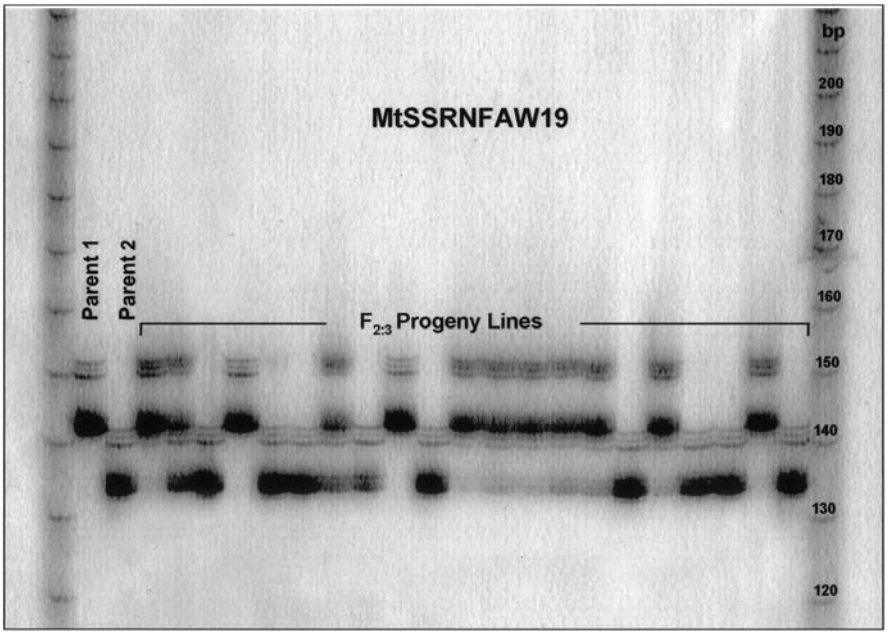

Fig. 4 The silver-stained polyacrylamide gel image showing inheritance of SSR alleles amplified by marker MtSSRN-FAW19 in the $F_{2:3}$ families of a *Medicago truncatula* mapping population developed by Dr. Cook (University of California, Davis, Calif.). The far *left-* and *right- lanes* show the size standard (10-bp DNA ladder)

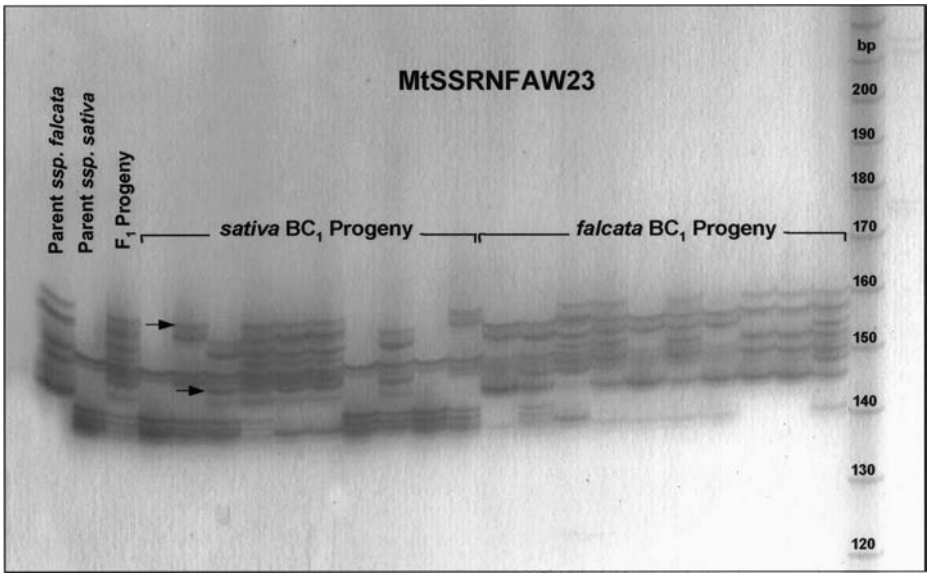

Fig. 5 The silver-stained polyacrylamide gel image showing the inheritance of SSR alleles amplified by marker MtSSRNFAW23 in the parents, the F_1, and 10 BC_1 progeny of two tetraploid mapping populations developed by Dr. Ian Ray (New Mexico State University, Las Cruces, N.M.). The far *right-lane* shows the size standard (10 bp ladder). The *arrows* indicate two of the SSR alleles segregating in a 1:1 ratio among the BC_1 progeny of the *M. sativa* × F_1 cross

9%, respectively. We selected relatively long SSRs ($\geq$20 bp) for marker development, because they are known to be more polymorphic and have greater amplitudes of variation (Weber 1990; Cho et al. 2000; Temnykh et al. 2000). A preference was given to design primers from ESTs with putative functions revealed by the TIGR blast results. Nearly, 48% of the ESTs from which the primers were designed were linked to putative functions (the blast results are available at the supplementary material).

EST-SSRs polymorphism and cross-species transferability

The high level of polymorphism of these SSR markers in alfalfa and annual medics imply that these markers will be highly effective for molecular study of these crops. The high level of polymorphism within the diploid and tetraploid alfalfa may be partly due to the fact that each of the genotypes was distinctly different. One of the two diploid alfalfa genotypes is from Iran (PI212798) and the other one is from Turkey (PI206453). Two of the tetraploid genotypes belong to different sub-species of alfalfa (Table 1). Similarly, the two annual medics, Armadillo and Sapo, belong to two different species. However, the marker polymorphism between each pair is meaningful, because they can potentially be crossed to produce segregating populations. The polymorphism between 'Jemalong' and 'Borung', as well as the PIC value from six *M. truncatula* genotypes, indicate the great potential of these markers for molecular studies in *M. truncatula*. We have also demonstrated that these markers could be used for determination of genetic diversity in *M. truncatula*.

In general, the EST-derived SSR markers have been reported to have lower rate of polymorphism compared to the SSR markers derived from genomic libraries in rice and wheat (Cho et al. 2000; Eujayl et al. 2002). The higher level of EST-SSR polymorphism observed in this study is also in part due to the greater level of inherent genetic variation that exist within these *Medicago* spp., compared to highly domesticated self-pollinating cereal crops, e.g., rice and wheat.

Our target was to develop a set of transferable SSR markers from the *M. truncatula* EST database that will be useful for important crop species of the *Medicago* genus, particularly alfalfa. The high level of transferability (89%) of these markers across various *Medicago* spp. indicates that, in fact, these markers should be highly useful in alfalfa, annual medics, and perhaps other crops within the *Medicago* genus. These results were expected, because the EST-SSRs are derived from transcribed regions of the DNA that generally are more conserved across species, and such SSR markers may have a higher rate of transferability than SSRs derived from non-transcribed regions (Scott et al. 2000). Also, all genotypes in the panel belong to genus *Medicago*, and high rates of transferability of SSR loci across species (>50%) within a genus have been documented earlier (Peakall et al. 1998; Gaitán-Solís et al. 2002). These markers with cross species amplification of SSR loci can be valuable assets for comparative mapping among *Medicago* spp. (Laurent et al. 2000; Rebeiz and Lewin 2000; Kantety et al. 2002). The 17 primer pairs that produced SSR products only in the two *M. truncatula* genotypes indicate that these are species-specific markers for *M. truncatula* and are not conserved across species.

Utility and deployment of EST-SSR markers

The usefulness of *M. truncatula* EST-SSR markers in construction of linkage maps, comparative mapping, and QTL discovery in *M. truncatula* has been demonstrated (Danesh et al. 2002; Huguet et al. 2001). We have tested

the transferability of such markers to a number of legume species within the genus *Medicago*. The high level of transferability and polymorphism of these markers for a number of important *Medicago* species, including alfalfa, are very important attributes of these markers. The 455 markers developed in the current study are different from the *M. truncatula* EST-SSR markers previously reported by Dr. N.D. Young (University of Minnesota, Minn.) and Dr. T. Huguet (CNRS-INRA, France, [personal communication 2002]). Thus, these markers provide a valuable asset for linkage mapping, comparative genomics, gene and QTL identification, and marker-assisted selection for plants belonging to the *Medicago* genus.

The *M. truncatula* EST-SSR markers may also be used in establishing the genomic relationship of the model legume *M. truncatula* to important legume crops such as alfalfa. The mapping of these markers in *M. truncatula* and alfalfa (for example) will help reveal the extent of the macro- and micro-synteny between the two species. The establishment of such syntenic relationships will facilitate the discovery of orthologous genes and the flow of genetic information between the model species and the crop species of interest. With the whole genome sequencing effort under way for *M. truncatula*, the ability to transfer genetic information across species from this model species will become a very important aspect of legume genomics. Also, many of these EST-SSR markers are located on *M. truncatula* genes with known functions, and the functions of other genes with such SSRs may also become available in the future. The knowledge about functions of genes with such markers offers the opportunity to investigate the consequences of polymorphism in the markers on gene functions.

Acknowledgements We thank Drs. Douglas Cook and Ian Ray for the plant materials of their respective mapping populations used in this study. This research work was funded by The Samuel Roberts Noble Foundation.

References

Anderson JA, Churchill GA, Sutrique JE, Tanksley SD, Sorrells ME (1993) Optimizing parental selection for genetic linkage maps. Genome 36:181–186

Ayers NM, McClung AM, Larkin PD, Bligh HFJ, Jones CA, Park WD (1997) Microsatellite and single nucleotide polymorphism differentiate apparent amylose classes in an extended pedigree of US rice germplasm. Theor Appl Genet 94:773–781

Baquerizo-Audiot E, Desplanque B, Prosperi JM, Santoni S (2001) Characterization of microsatellite loci in the diploid legume *Medicago truncatula* (barrel medic). Mol Ecol Notes 1:1–3

Barnes DK, Goplen BP, Baylor JE (1988) Highlights in the USA and Canada. In: Hanson AA, Barnes DK, Hill RR (eds) Alfalfa and alfalfa improvement. Agronomy Monograph No. 29, American Society of Agronomy – Crop Science Society of America– Soil Science Society of America, Madison, Wis., USA, pp1–24

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

Brown-Guedira GL, Thomson JA, Nelson RL, Warburton ML (2000) Evaluation of genetic diversity of soybean introductions and North American ancestors using RAPD and SSR markers. Crop Sci 40:815–823

Cho YG, Ishii T, Temnykh S, Chen X, Lipovich L, McCouch SR, Parl WD, Ayers N, Cartinhour S (2000) Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). Theor Appl Genet 100:713–722

Danesh D, Shekhawat NS, Cardinet G, Thoquet P, Mudge J, Penuela S, Kim D, Kiss G, Choi H, Limpens E, Zeyen R, Huguet T, Cook DR, Young ND (2002) Integrated microsatellite marker mapping and powdery mildew resistance in *Medicago truncatula* (abstract). In: Plant and Animal Genome X. The International Conference on the Status of Plant and Animal Genome Research, 12–16 January 2002, San Diego, Calif., pp 196

Diwan N, Bhaqwat AA, Bauchan GB, Cregan PB (1997) Simple sequence repeats DNA markers in alfalfa and perennial and annual *Medicago* species. Genome 40:887–895

Diwan N, Bouton JH, Kochert G, Cregan PB (2000) Mapping of simple sequence repeat (SSR) DNA markers in diploid and tetraploid alfalfa. Theor Appl Genet 101:165–172

Doyle JJ, Doyle JL, Ballenger JA, Palmer JD (1996) The distribution and phylogenetic significance of a 50-kb chloroplast DNA inversion in the flowering plant family Leguminosae. Mol Phylogenet Evol 5:429–438

Eujayl I, Sorrells ME, Wolters P, Baum M, Powell W (2002) Isolation of EST- derived microsatellite markers for genotyping the A and B genomes of wheat. Theor Appl Genet 104:399–407

Gaitán-Solís E, Duque MC, Edwards KJ, Tohme J (2002) Microsatellite repeats in common bean (*Phaseolus vulgaris*): isolation, characterization, and cross-species amplification in *Phaseolus* ssp. Crop Sci 42:2128–2136

Huguet T, Thoquet P, Gherardi M, Kereszt A, Ane JM, Vilotte L, Cardinet G, Baquerizo E, Santoni S, Prosperi JM (2001) The molecular linkage map of the model legume *Medicago truncatula*: a tool for legume genome comparison and gene mapping. In: 4th Workshop on *Medicago truncatula*. 7–10 July 2001, University of Wisconsin-Madison, Wis., pp 51

Kaló P, Endre G, Zimnyi L, Csanadi G, Kiss GB (2000) Construction of an improved linkage map of diploid alfalfa (*Medicago sativa*). Theor Appl Genet 100:641–657

Kantety RV, Rota ML, 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

Kilian A, Chen J, Han F, Steffenson B, Kleinhofs A (1997) Towards map-based cloning of the barley stem rust resistance genes *Rpg1* and *rpg4* using rice as an intergenomic cloning vehicle. Plant Mol Biol 35:187–195

Laurent P, Elduque C, Hayes H, Saunier K, Eggen A, Leveziel H (2000) Assignment of 60 human ESTs in cattle. Mammal Genome 11:748–754

Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci 76:5269–5273

Paterson AH, Lin YR, Li S, Schertz KF, Doebley JF, Pinson SRM, Liu SC, Stansel JW, Irvine JE (1995) Convergent domestication of cereal crops by independent mutations at corresponding genetic loci. Science 269:1714–1717

Peakall R, Gilmore S, Keys W, Morgante M, Rafalski A (1998) Cross species amplification of soybean (*Glycine max*) simple sequence repeat (SSRs) within the genus and other legume genera: implication for transferability of SSRs in plants. Mol Biol Evol 15:1275–1287

Rebeiz M, Lewin HA (2000) COMPASS of 47,787 cattle ESTs. Anim Biotechnol 11:175–241

Roa AC, Chavarriaga-Aguirre P, Duque MC, Maya MM, Bonierbale MW, Iglesias C, Tohme J (2000) Cross-species amplification of cassava (*Manihot esculenta*) (Euphorbiaceae) microsatellites: allelic polymorphism and degree of relationship. Am J Bot 87:1647–1655

SAS Institute (1989) SAS/STAT user's guide, version 6, 4th edn, vol 1. SAS Institute, Cary, N.C.

Scott KD, Eggler P, Seaton G, Rossetto M, Ablett EM, Lee LS, Henry RJ (2000) Analysis of SSRs derived from grape ESTs. Theor Appl Genet 100:723–726

Temnykh S, Park WD, Ayers N, Cartinhour S, Hauck N, Lipovich L, Cho YG, Ishii T, McCouch SR (2000) Mapping and genome organization of microsatellite sequences in rice (*Oryza sativa* L.). Theor Appl Genet 100:697–712

Thoquet P, Ghérardi M, Journet E, Kereszt A, Ané JM, Prosperi JM, Huguet T (2002) The molecular genetic linkage map of the model legume *Medicago truncatula*: an essential tool for comparative legume genomics and the isolation of argonomically important genes. BMC Plant Biology 2:1–19

Tugendreich S, Boguski M, Seldini MS, Hieter P (1993) Linking yeast genetics to mammalian genomes: identification and mapping of the human homolog of CDC27 via the expressed sequence tag (EST) database. Proc Natl Acad Sci 90:10031–10035

Weber JL (1990) Informativeness of human (dC-dA)n (dG-dT)n polymorphisms. Genomics 7:524–530

White G, Powell W (1997) Isolation and characterization of microsatellite loci in *Swietenia humilis* (Meliaceae): an endangered tropical hardwood species. Mol Ecol 6:851–860

Wu KK, Burnquist W, Sorrells ME, Tew TL, Moore PH, Tanksley SD (1992) The detection and estimation of linkage in polyploids using single dose restriction fragments. Theor Appl Genet 83:294–300