Mol Genet Genomics (2002) 267: 107–114
DOI 10.1007/s00438-002-0643-z

ORIGINAL PAPER

A. Porceddu · E. Albertini · G. Barcaccia · G. Marconi
F.B. Bertoli · F. Veronesi

Development of S-SAP markers based on an LTR-like sequence from *Medicago sativa* L.

Received: 13 November 2001 / Accepted: 18 January 2002 / Published online: 22 February 2002

Abstract The Sequence-Specific Amplification Polymorphism (S-SAP) method, recently derived from the Amplified Fragment Length Polymorphism (AFLP) technique, produces amplified fragments containing a retrotransposon LTR sequence at one end and a host restriction site at the other. We report the application of this procedure to the LTR of the *Tms1* element from *Medicago sativa* L. Genomic dot-blot analysis indicated that *Tms1* LTRs represent about 0.056% of the *M. sativa* genome, corresponding to 16×10^3 copies per haploid genome. An average of 66 markers were amplified for each primer combination. Overall 49 polymorphic fragments were reliably scored and mapped in a F_1 population obtained by crossing diploid *M. falcata* with *M. coerulea*. The utility of the LTR S-SAP markers was higher than that of AFLP or SAMPL (Selective Amplification of Microsatellite Polymorphic Loci) markers. The efficiency index of the LTR S-SAP assay was 28.3, whereas the corresponding values for AFLP and SAMPL markers were 21.1 and 16.7, respectively. The marker index for S-SAP was 13.1, compared to 8.8 for AFLP and 9.5 for SAMPL. Application of the *Tms1* LTR-based S-SAP to double-stranded cDNA resulted in a complex banding pattern, demonstrating the presence of *Tms1* LTRs within exons. As the technique was successfully applied to other species of the genus *Medicago*, it should prove suitable for studying genetic diversity within, and relatedness between, alfalfa species.

Keywords *Medicago* · Long Terminal Repeat (LTR) · Retrotransposons · Sequence-Specific Amplification Polymorphism (S-SAP)

Communicated by R. Hagemann

A. Porceddu[1] · E. Albertini · G. Marconi
F.B. Bertoli · F. Veronesi (✉)
Dipartimento di Biologia Vegetale
e Biotecnologie Agroambientali,
University of Perugia, Borgo XX Giugno 74,
06121, Perugia, Italy
E-mail: veronesi@unipg.it
Tel.: +39-075-5856207
Fax: +39-0755856224

G. Barcaccia
Dipartimento di Agronomia Ambientale
e Produzioni Vegetali, University of Padova,
Agripolis, Via Romea 16, 35020 Legnaro, Padova, Italy

Present address:
[1]CNR Istituto di Ricerche sul
Miglioramento Genetico delle Piante Foraggere,
Via Madonna Alta 130, 06100 Perugia, Italy

Introduction

The retro-element family of transposons is composed of transposable elements that move via an RNA intermediate. LTR-retrotransposons, which are present in high copy numbers in plant genomes, are flanked by long terminal repeat (LTR) sequences (Hull and Will 1989). On the basis of the location of the integrase domain, two major groups of LTR-retrotransposons can be distinguished. In the *Ty1*-copia group, named after the well-studied *Ty* element of *Saccharomyces cerevisiae* and the *copia* element of *Drosophila melanogaster*, the coding sequence for the integrase domain lies 3′ to that for the reverse transcriptase domain, whereas in the *gypsy* group the converse is true (Boeke and Corces 1989).

Elements of both groups have been found in almost every plant species investigated, although copy numbers and degrees of heterogeneity differ widely (Flavell et al. 1992a; Suoniemi et al. 1998). For instance, 30 retrotransposon families, each with an average copy number of two elements, have been found in *Arabidopsis thaliana*, the haploid genome of which corresponds to 0.15 pg of DNA (Kumar and Bennetzen 1999). The total number of retrotransposons in rice, whose genome size is three times that of *Arabidopsis*, is estimated to be about 1,000 (Hirochika et al. 1992), while barley, with a genome size 36 times that of *Arabidopsis*, has 14,000

copies of the *Bare1* element (Vicient et al. 1999). It would seem to follow from these data that there is a correlation between copy number and genome size. However, the *Wis2* family in wheat is composed of only 200 elements, despite the fact that the genome size in this species is three times that of barley, while potato, whose genome size exceeds that of *Arabidopsis* by more than 10-fold, has about 400 retrotransposons (Flavell et al. 1992b). An accurate investigation of the copy numbers of *Ty1*-copia retrotransposons in *Vicia* species indicates that the degree of sequence heterogeneity of elements of the *Ty1*-copia group correlates with their copy number within each genome, but that neither heterogeneity nor copy number is related to the genome size of the host (Pearce et al. 1996).

Another relevant feature of plant retrotransposons is their distribution within the genome. In situ hybridization to metaphase chromosomes has shown that both *Ty1*-copia-like and *gypsy*-like elements are spread throughout the genome. Pearce et al. (1996) have demonstrated that the *Ty1*-copia retrotransposons of *Vicia* are mostly located in euchromatic regions and are far less common in heterochromatic regions. Moreover, computer-assisted database searches, using *Ty1*-copia retrotransposons as query sequence, have revealed that ancient or degenerate retrotransposon sequences are often located close to plant genes (White et al. 1994; Pesole et al. 1997), a finding which suggests that retrotransposons may be involved in the evolution of plant gene structure and expression, supplying genes with regulatory sequences and facilitating gene duplication and/or exon shuffling.

Although all the features of plant retrotransposons mentioned above are of pivotal interest for the development of retrotransposon-based molecular markers, the high copy number of some retrotransposons has hindered their use as multilocus RFLP markers.

With the advent of AFLP technology (Vos et al. 1995), a new and simple approach to the development of high-multiplex-ratio retrotransposon-based molecular markers (Sequence-Specific Amplification Polymorphisms, S-SAP) has become available (Waugh et al. 1997). It is based on the production of PCR-derived fragments containing a retrotransposon sequence at one end and a flanking host restriction site at the other. The retrotransposon-specific primer in the original S-SAP technique developed by Waugh et al. (1997) in barley was derived from the highly conserved terminus of the *Bare1* LTR. In a subsequent modification of the method, which was applied to various *Pisum* spp., the retrotransposon-specific primer was designed to correspond to the polypurine tract of the retro-element *PDR* (Ellis et al. 1998). Both approaches mainly exploited the variation in the sequences flanking the insertion site.

We report here the adaptation of the S-SAP method to alfalfa (*Medicago sativa* L.), based on the LTR of the retrotransposon *Tms1* (Vegh et al. 1990), and describe its application to both genetic linkage analysis and the elucidation of phylogenetic relationships in the genus *Medicago*. The efficiency with which polymorphisms can be detected by the *Tms1* LTR-derived S-SAP technique was compared with that of SAMPL and AFLP markers. The distribution of *Tms1* LTRs within plant exons was investigated by applying *Tms1* LTR S-SAP to an alfalfa cDNA template.

Materials and methods

Plant materials

An F_1 segregating population, derived from a cross between a *Medicago falcata* (L.) Arcang. (2n = 2x = 16) mutant named PG-F9 and *M. coerulea* (Less.) Schm. (2n = 2x = 16), was used for construction of the linkage map (Barcaccia et al. 1999, 2000). Annual and perennial *Medicago* species (*M. coerulea*, *M. costricta*, *M. falcata* 2x, *M. falcata* 4x, *M. glomerata*, *M. intertexta*, *M. lesinsii*, *M. murex*, *M. muricoleptis*, *M. polymorpha*, *M. praecox*, *M. rigidula*, *M. rugosa*, *M. sativa*) were used to assess LTR sequence distribution within the genus *Medicago*. Genomic DNAs were extracted from single F_1 plants and from pools of 10 plants of annual and perennial *Medicago* species, using the cetyltrimethylammonium bromide (CTAB) procedure (Doyle and Doyle 1990).

S-SAP marker analysis

Total DNA was digested with restriction enzymes, and ligated to adaptors according to the method of Vos et al. (1995), modified as in Barcaccia et al. (1999). Briefly, genomic DNA (500 ng) was digested and ligated for 4 h at 37°C using the enzymes *Eco*RI and *Mse*I (5 U each), 1 U of T4 ligase (Pharmacia Biotech), 50 pmol of *Mse*I adaptor, and 5 pmol of *Eco*RI adaptor in RL buffer (20 mM TRIS-acetate, 20 mM magnesium acetate, 100 mM potassium acetate, 5 mM DTT, 2.5 μg BSA) supplemented with ATP to a final concentration of 10 mM. The template DNA was then pre-amplified in a 20-μl reaction mixture containing 5 μl of ten-fold diluted (digested and ligated) DNA, 75 ng of the primers *Eco*RI + C and *Mse*I + A, PCR buffer (50 mM MgCl, 1.5 mM $MgCl_2$, 10 mM TRIS-HCl), 10 mM dNTPs (Pharmacia Biotech) and 1 U of *Taq* DNA polymerase (Pharmacia Biotech). The cycling conditions were: one cycle of 45 s at 94°C, 30 s at 65°C, 1 min at 72°C, followed by a touch-down profile for the annealing step (13 cycles in which the annealing temperature was decreased at a rate of 0.7°C/cycle), followed by 18 cycles at a constant annealing temperature of 55.9°C, and a final extension step at 72°C for 5 min. Selective restriction fragment amplification was performed with a fluorescence-labelled *Tms1* LTR-derived primer (Vegh et al. 1990) and an unlabelled *Mse*I + 3 primer. Ret1 (5′-CGGTTTTGTGGGGTTGTGTTAGGCCCA-3′, labelled at the 5′ end) was used for mapping experiments and Ret1 or Ret2 (5′-GTTGGCCTGACAATTTGTTTATAA-3′, labelled at the 5′ end) was used for the shift assay (see below). The labelled oligonucleotides were obtained from Genset Oligos. Each 20-μl PCR contained 1% of the pre-amplified DNA, 50 ng of fluorescence-labelled Ret1, 30 ng of unlabelled *Mse*I + 3 primer (the selective bases used were AGA, AAC, AGT, and AGC), 2 μl of PCR buffer (Pharmacia Biotech), 4 mM dNTPs, and 0.4 U of *Taq* DNA polymerase. Amplifications were carried out using the amplification profile described above. After PCR, 8 μl of loading buffer (98% formamide, 2% dextran blue, 0.25 mM EDTA) was added to each sample. Samples were denatured at 90°C for 5 min and then immediately placed on ice. An aliquot (6 μl) of each sample was loaded onto a 6% polyacrylamide gel (60 cm×30 cm×0.4 mm), which had been run for 2 h and 45 min at 80 W. Gels were scanned using the Genomyx LR scanner (Beckman Coulter).

SAMPL markers

SAMPL marker analysis was performed according to Morgante and Vogel (1994). The SAMPL protocol is similar to the S-SAP procedure except for the primer used in the second amplification (primer S1: f-CACACACACACACACTATAT-3′; Genset Oligos). The second amplification uses the fluorescent S1 primer and a standard AFLP *Mse*I primer with three selective nucleotides (AGT, AGC, and AAG). Gel electrophoresis and scanning were carried out as described for the S-SAP analysis.

Identification of S-SAP markers in a population of cDNAs

Total RNA was isolated from about 2 g of leaves using TRIZOL (Gibco-BRL); the RNA concentration was determined spectrophotometrically and then adjusted to a final concentration of 1 μg/μl. The poly(A)$^+$ RNA fraction was isolated from 1 mg of total RNA using the mRNA Purification Kit (Amersham Pharmacia Biotech) according to the manufacturer's suggestions. First and second cDNA strands were synthesized following standard protocols (Sambrook et al. 1996). The resulting double-stranded cDNA was purified by extraction with phenol, precipitated with ethanol, and subjected to the LTR S-SAP amplification procedure.

Linkage analysis

Segregation data for LTR S-SAP and SAMPL markers were analyzed together with those of AFLP and RAPD markers previously scored in the same mapping population (Barcaccia et al. 1999) by using Joinmap Version 3.0 (Van Ooijen and Voorrips 2001). For mapping, the "cross pollination" (CP) option was employed, i.e. an F_1 population resulting from a cross between two heterogeneous parents that were, respectively, heterozygous and homozygous at the loci being tested. For the identification of linkage groups with selected markers, the "grouping" module was employed, setting a minimum LOD score of 3.5 and a maximum recombination frequency $r = 0.30$. The map distances, expressed in centimorgans (cM), were calculated using the Kosambi function (Kosambi 1944).

Dot blots

Dot blots were prepared by applying 100-, 200-, 500-, 1000-, and 1500-ng aliquots of genomic DNA samples to membrane filters. The DNA was cross-linked to filters by irradiation with UV light. Herring sperm DNA at 100 ng per dot served as a negative control. PCR products were obtained from diploid *M. coerulaea* and tetraploid *M. sativa* by using a pair of primers based on sequences at the ends of the *Tms1* LTR – Ret3 (5′-GAGTCCCACATCGGTTAGGAGTTGGCCTGACA-3′) and Retant (5′-GAAATTGTGCTTGGGCCTAACACA-3′). The primers were obtained from Invitrogen. Purified PCR products were used as positive controls on each filter. The ^{32}P-labelled probes were synthesized by PCR from purified PCR fragments. Filters were hybridized as in Vicient et al. (1999). All blots were hybridized together, so the probe concentration was identical for all filters. Hybridized filters were washed successively with 2×SSC, 0.1% SDS and 0.1×SSC, 0.1% SDS at 55°C. Bound radioactivity was analyzed using a Packard Instant Imager.

Copy number estimation

Genomic copy number was calculated from the hybridization response of the genomic DNA of *M. coerulea* and *M. sativa* by comparison with the control DNA on the blots as follows. First, the amount of LTR sequence in 1 ng of genomic DNA was calculated using the formula: genomic counts per ng/probe counts per ng. The percentage of the 1C genome represented by LTR sequences was calculated by multiplying the value so obtained by the size of the 1C genome (0.90 pg for *M. coerulea* and 1.75 pg for *M. sativa*). The percentage of the genome represented by the LTRs was finally converted to genomic copy number by multiplying it by the percentage of genome length occupied by one LTR. The data for genome size were obtained from the Kew Gardens website (http://www.rbgkew.org.uk/cval). For each sample the mean of four hybridizations on two separate filters was used for copy number calculations.

Estimation of marker system utility

To compare the efficiency of the LTR S-SAP method with that of other molecular marker systems such as AFLP and SAMPL, an assay efficiency index (Ai) was calculated. This index combines the effective number of alleles identified per locus, calculated as $n_e = 1/\sigma p_i^2$, where p_i is the frequency of the i^{th} marker allele (Kimura and Crow 1964), and the number of the polymorphic bands detected in each assay, computed as $Ai = \sigma n_e / P$, where Σn_e is the total number of effective marker alleles detected over all loci and P is the total number of assays performed (i.e. primers used) for their detection (Pejic et al. 1998).

An additional parameter, the marker index (MI), which is the product of expected heterozygosity and multiplex ratio, was used to evaluate the overall utility of each marker system (Powell et al. 1996). It was calculated as $MI = H_{pl}\beta n$, where H_{pl} is the total genetic diversity computed over all polymorphic loci (Nei 1973), β is the percentage of polymorphic loci and n is the number of loci detected per primer. For dominant marker systems, H_{pl} was determined as $1-\sigma(p_1^2+p_0^2)$ over all loci, where p_1 and p_0 represent present and absent marker alleles, respectively.

Results

Application of S-SAP to *Medicago*, based on the LTR of *Tms1*

An S-SAP protocol for a retrotransposon sequence found in *Medicago* spp. was set up in this study. The procedure named *Tms1* LTR S-SAP is depicted in Fig. 1. The retrotransposon-specific primer was designed on the basis of the terminal region of the LTR sequence of the *Tms1* retrotransposon from *M. sativa* (Vegh et al. 1990). The selective amplification employed only one adaptor-homologous primer, together with a fluorescence-labelled *Tms1* LTR-derived primer.

As LTRs are direct repeats, every LTR-derived primer can prime amplification both into and out from the retrotransposon, generating two types of fragments: those composed of sequences internal to the retrotransposon, and those made up of host sequences flanking the LTR. Samples were amplified adding 0, 1, 2, or 3 selective nucleotides to the *Mse*I primer. When no selective nucleotide was added, internal retrotransposon sequences were mainly seen as intense amplification products on a background smear (data not shown). When selective nucleotides were added, the number of the retrotransposon-derived products was reduced to a level that allowed them to be easily scored in terms of presence/absence (data not shown). As expected, *Eco*RI selective primers produced fewer products than *Mse*I primers, reflecting the lower frequency of *Eco*RI cleavage sites.

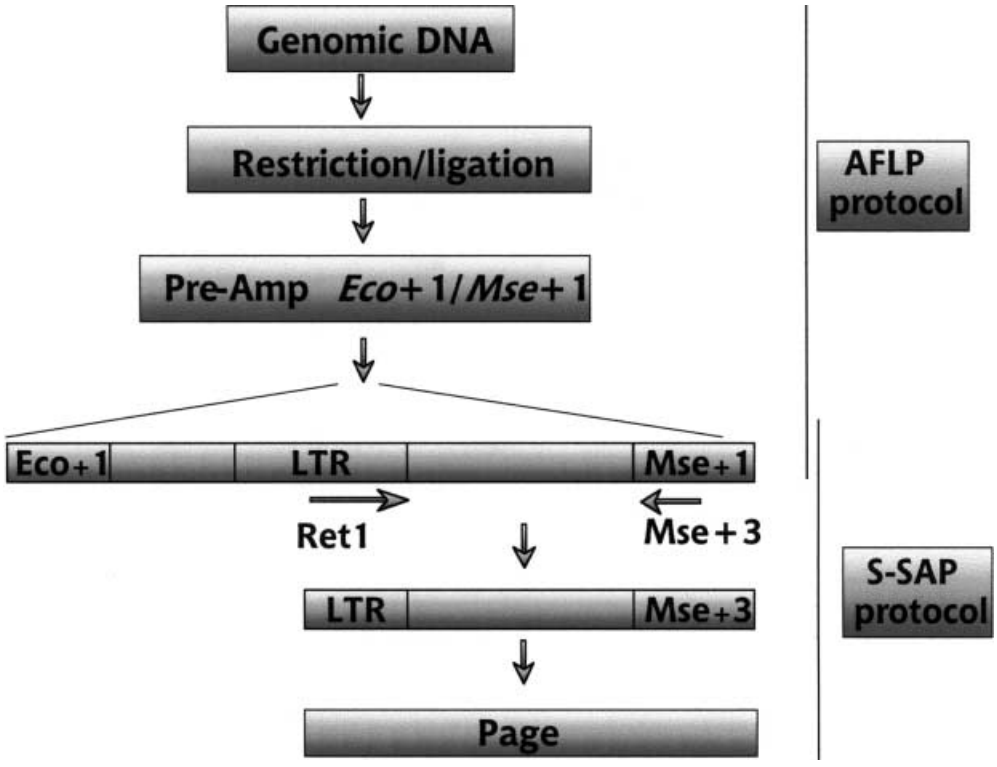


Fig. 1 Schematic representation of the *Tms1* LTR Sequence-Specific Amplification Polymorphism (S-SAP) protocol

The level of polymorphism detected in the F_1 segregating population by the *Tms1* LTR S-SAP procedure was compared with the levels detected using SAMPL and AFLP markers (Table 1). When used in combination with *Mse*I primers (with 3 selective nucleotides), each *Tms1* LTR-derived primer yielded, on average, more products than SAMPL, but less than AFLP, while the percentage of polymorphisms detected was higher than that obtained with either SAMPL or AFLP (Table 1). Figure 2 shows an example of AFLP, S-SAP and SAMPL patterns generated with an F_1 population obtained from the cross *M. coerulea* × M. falcata. Determination of the marker index (MI) and assay efficiency index (Ai) revealed that the utility of the S-SAP marker system was much higher than that of either AFLPs or SAMPLs. MI was 13.1 for S-SAPs, 8.8 and 9.5 for SAMPLs and AFLPs, respectively, whereas Ai was 28.3 for S-SAPs and 21.1 and 16.7 for SAMPLs and AFLPs, respectively. Although the mean expected heterozygosity measured with dominant PCR-derived markers is actually much lower than that scorable with co-dominant single-locus markers, the relative information content of S-SAP, AFLP and SAMPL marker systems was strongly influenced by the higher multiplex-ratio component of these assays.

Localization of *Tms1* LTRs on the genetic map and determination of element copy number

A genetic approach was used to study the distribution of *Tms1* LTRs in the genome. Segregation data for 50 maternal S-SAP markers and 29 SAMPLs, which were polymorphic in the F_1 population, were analyzed alongside an existing framework of other molecular markers (Barcaccia et al. 1999). The final map covered 568.5 cM and consisted of 180 markers (56 AFLPs, 49 S-SAPs, 23 SAMPLs, 49 RAPDs and 3 RFLPs) arranged in 14 linkage groups (with an average length of 40.6 cM and containing 12.9 markers per group). It is worth mentioning that all marker linkages previously reported by Barcaccia et al. (1999) were confirmed in the current map. Linkage groups with more than 6 markers are shown in Fig. 3. S-SAP markers were distributed throughout the linkage groups identified. The RFLP probes Vg2B9 (group 1), Vg2F8 (group 2), and Vg2C2 (group 8) mapped in the current map were also mapped by Tavoletti et al. (1996) in the same maternal genotype PG-F9 and assigned to their groups 1, 2, and 6, respectively, and by Brouwer and Osborn (1999) in tetraploid alfalfa; they refer to the linkage groups that bear these markers as 1, 2 and 3, respectively.

To verify that the fragments detected indeed result primarily from the amplification of transposon-flanking sequences, S-SAP analysis was performed on an F_1 individual using either the Ret2 or the Ret1 primer. The mobility shift that occurs when the Ret1 primer is used instead of the Ret2 primer can be predicted precisely from the LTR sequence (Vegh et al. 1990). When Ret2 is used as primer, the PCR products are expected to be 50 bp longer than those obtained using Ret1 as primer, but the patterns should be similar (data not shown). About 80% (65 out of 82) of the fragments amplified by

Table 1 Descriptive statistics of LTR S-SAP, SAMPL and AFLP markers as estimated using data for a diploid F_1 segregating population

Marker type	Number of primer combinations	Average number of bands per assay (±SE)	Percentage poly morphic bands (±SE)	MI[a]	Ai[b]
S-SAP	4	66.5 ± 6.9	42.0 ± 6.9	13.1	28.3
SAMPL	4	50.0 ± 3.4	36.6 ± 3.3	8.8	21.1
AFLP	7	70.0 ± 5.6	28.9 ± 5.1	9.5	16.7

[a] MI, marker index; Ai, assay efficiency index (see Materials and methods for definitions)

the Ret1 primer were also observed in the amplification pattern obtained using the Ret2 primer, 13% (11 out of 82) of the bands was found only in the Ret1 profile and 7% (6 out of 82) only in the Ret2 reaction. Thus about 80% of the amplified fragments were derived from genomic regions flanking the conserved *Tms1* LTR sequence. The remaining 20% of the products can be explained as arising from modified LTR sequences or as non-specific amplifications.

Tms1 LTR copy numbers per haploid genome equivalent in *M. sativa* and *M. coerulea* were also determined. Genomic dot blots were hybridized with radiolabelled LTR probes obtained by PCR amplification using Ret3 and Retant primers on total genomic DNA of each species. *M. coerulea* was estimated to contain about 8.028 (±0.23) × 10^3 LTR copies per haploid genome equivalent. LTR copy number in *M. sativa* was about double that, being equivalent to 16.110 (±1.02) × 10^3 per 1C. Note that *M. sativa* is 2n = 4x, while *M. coerulea* is 2n = 2x.

Some *Tms* LTRs are associated with exons

A computer-assisted database search using the *Ty1*-copia retrotransposon as query sequence revealed that ancient or degenerate retrotransposon sequences are often located in close proximity to plant genes. As little is known about the frequency of retrotransposons within plant exons, S-SAP was carried out on cDNA from *Medicago* leaves. Figure 4a shows a typical LTR derived S-SAP gel obtained from a cDNA preparation. On average 20 bands were visualized for each primer combination.

Potential use of *Tms1* LTR S-SAP for the study of diversity in *Medicago*

We also tested the potential of the *Tms1* LTR S-SAP technique to other species in the genus *Medicago*. Pools of genomic DNAs, isolated from 10 plants of

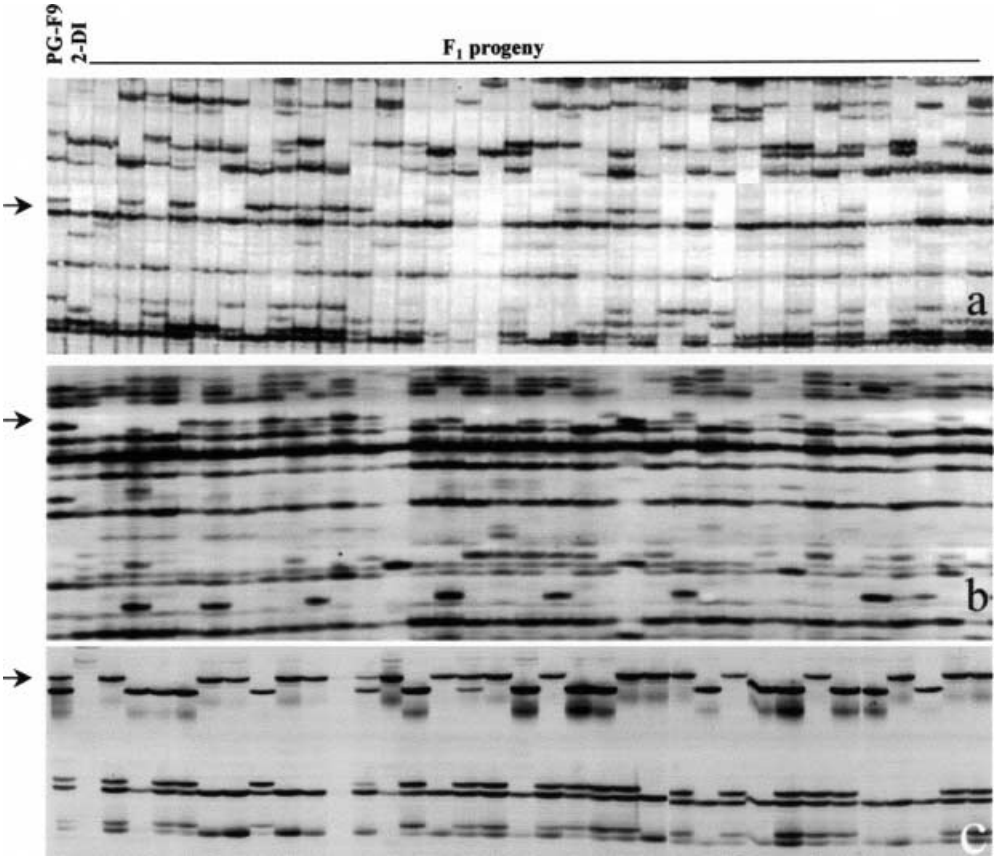


Fig. 2a–c AFLP (**a**), S-SAP (**b**) and SAMPL (**c**) fingerprints. The *arrows* indicate examples of markers that are polymorphic between the parents and segregate in the F_1 progeny

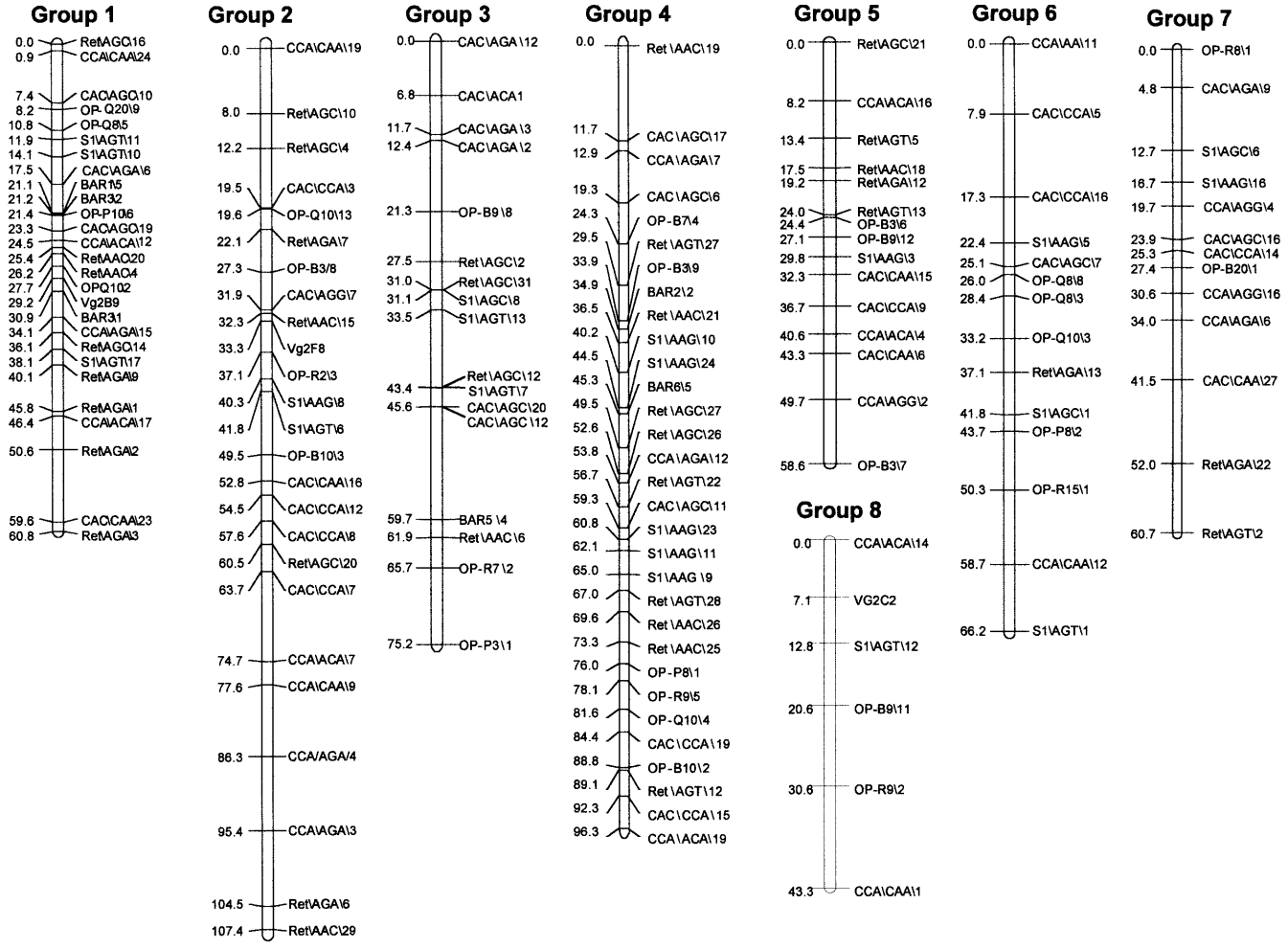


Fig. 3 Integrated AFLP, S-SAP, SAMPL, RAPD and RFLP map of the *M. falcata* mutant PG-F9. Map distances in centimorgans (cM) are given on the *left* of each group, while marker names are given on the *right*

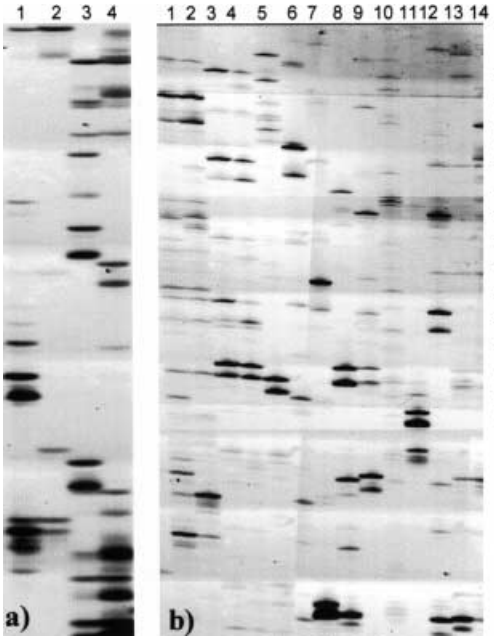
annual and perennial *Medicago* species, were subjected to S-SAP amplification. Every species generated a complex amplification pattern with fragments ranging in length from 50 to 500 bp. Fig. 4b shows the pattern generated using the Ret1 and *Mse*I + AGT primer combination, together with a list of the species analyzed, and also gives additional information on chromosome numbers and average numbers of *Tms1* LTR-derived bands. Means and standard errors were calculated for data sets obtained from reactions carried out with Ret1 and *Mse*I specific primers with three selective nucleotides.

Discussion

Retrotransposons are short mobile DNA elements that are thought to have played an important role in the shaping of eukaryotic genomes. Thus, it has been proposed that nuclear spliceosomal introns, retroviruses, and even telomerases may all have evolved from retrotransposons (Eickbush 1999).

Members of the *Ty*-copia and *gypsy*-like families of retrotransposons have been found in almost every species investigated (Flavell et al. 1992a; Suoniemi et al. 1998). However, only a few species have been shown to contain active retrotransposon elements. In fact, in most cases retrotransposon sequences are present as remnants of full-length elements scattered throughout the genome. In barley, for example, the most abundant retrotransposon sequences are LTRs (Vicient et al. 1999), followed by reverse trascriptase and integrase domains. The technique we developed for revealing polymorphisms in the regions flanking retrotransposon *Tms1* LTR sequences in *Medicago* spp. produced complex amplification patterns, indicating that the copy number of *Tms1* LTR sequences in the *Medicago* genome is high. Our estimates of the copy numbers of *Tms1* LTRs in *M. coerulea* (2x) and *M. sativa* (4x), a diploid and a tetraploid species of *Medicago*, correspond to 8×10^3 and 16×10^3, respectively.

The potential application of *Tms1* LTR-derived markers to the study of diversity patterns among different *Medicago* species was also investigated. Each species, when tested with the *Tms1* LTR S-SAP



	Medicago species	Chrom. number	N°. bands average
1	*M. murex*	2n=14*	58.5±6.4
2	*M. lesinsii*	2n=16*	59.0±9.8
3	*M. rugosa*	2n=30	40.2±5.0
4	*M. rigidula*	2n=14*	35.7±4.1
5	*M. polymorpha*	2n=14*	36.7±7.4
6	*M. costricta*	2n=14*	33.0±1.6
7	*M. praecox*	2n=14*	38.0±2.3
8	*M. sativa*	2n=32	33.7±5.0
9	*M. glomerata*	2n=16*	28.2±3.2
10	*M. intertexta*	2n=16*	35.7±4.4
11	*M. muricoleptis*	2n=16*	32.7±4.1
12	*M. coerulea*	2n=16*	41.0±5.6
13	*M. falcata 2x*	2n=16*	45.0±7.1
14	*M. falcata 4x*	2n=32	39.7±7.3

*=diploid species,

Fig. 4 LTR S-SAP gel generated by two different primer combinations on cDNAs obtained from two *M. sativa* genotypes. Lanes 1 and 2, Ret1/*Mse*I + AGA; lanes 3 and 4, Ret1/*Mse*-I + AGT (**a**). LTR S-SAP technique applied to a set of *Medicago* species. The sources of the DNAs used for each lane are listed on the right, together with data for relative chromosome number and average number of bands. The amplification products were obtained using the primer combination Ret1/Mse + AGT (**b**)

protocol, produced large numbers of amplified products. The *Medicago* genus is particularly amenable to investigation of the mechanism that controls retrotranposon copy numbers, owing to the range of ploidy levels represented. In *Hordeum* spp. homologous recombination between LTRs is thought to reduce the complement of functional retrotransposons within the genome, providing a sort of "return ticket from genomic obesity". A similar mechanism seems to operate in yeast, since out of about 330 LTRs found, only 50 were part of full-length retrotransposons, the rest being *solo* or individual LTRs (Vicient et al. 1999). Our in\vestigations on *M. coerulea* and *M. sativa* suggest that in these two species there is a good correlation between copy number and ploidy level. Further investigations are needed before we can conclude that this finding holds for the genus as a whole.

By using a PCR-based approach we demonstrated that not every band produced can be considered as being derived from a conserved LTR sequence. There is more than one possible explanation for this phenomenon. Firstly, some of the amplification products may result from intervening *Mse*I sites in the region between the two LTR primers. However, scanning of the sequence between the LTR primers for potential *Mse*I sites reveals that there are only two potential sites which could form an *Mse*I site as a result of point mutations. Unless we accept the hypothesis that the nucleotide substitution rate is particularly high within LTR regions, it is unlikely that this explanation could account for most of the sequence heterogeneity found. Another possibility is that LTRs represent hot spots for recombination. The recent finding of an excess of *solo* LTR structures in barley, probably the result of intrachromosomal homologous recombinations between LTRs (Vicient et al. 1999), is consistent with this hypothesis.

The ability to map markers with confidence is a function of the reproducibility of the protocol and the ability to score segregating bands unambiguously. In this regard the *Tms1* LTR S-SAP markers were highly robust, as 80.3% of them could be mapped with high LOD scores.

We compared the genetic distribution and the polymorphism level of LTR, AFLP and SAMPL markers. The number of bands for LTR-derived S-SAP was higher than for SAMPL but lower than for AFLP, while the percentage polymorphism was greater. As there were no significant clusters of LTR markers, even with respect to SAMPL and AFLP markers, this type of marker provides good genome coverage. The distribution of the marker relative to gene sequences has important implications for the general applicability and utility of any particular marker class (Kumar and Bennetzen 1999). By performing the S-SAP protocol on a cDNA synthesized from total RNA extracted from young *Medicago* leaves, we demonstrated the presence of the *Tms1* LTR sequence within plant exons. This finding may be useful for generating markers linked to agronomically important traits (Kumar and Bennetzen 1999).

Our ongoing experiments are aimed at identifying bands originating from LTRs of intact retrotransposons. The issue is of particular interest since retrotransposons carry relatively strong promoter-enhancer elements (Pouteau et al. 1991; Hirochika et al. 1996) and the presence of a large number of retrotransposons in a genome could have a marked influence on the transcription of their flanking regions.

Acknowledgements The authors wish to thank Dr. G. B. Kiss of the Institute of Genetics at the Biological Research Centre of the Hungarian Academy of Sciences (Szeged, Hungary) and Prof. F. Salamini of the Max Planck Institute for Plant Breeding Research (Cologne, Germany) for critical reading of the manuscript and

helpful suggestions. Thanks are also due to Prof. E. Falistocco of the University of Perugia for providing seeds of different *Medicago* species. The research was funded by the Italian Ministry for Universities, Research, Science and Technology, Project "Characterization of mutations affecting sporogenesis and gametogenesis in *Medicago* spp." (Project Leader: Prof. F. Veronesi).

References

Barcaccia G, Albertini E, Tavoletti S, Falcinelli M, Veronesi F (1999) AFLP fingerprinting in *Medicago* spp.: its development and application in linkage mapping. Plant Breed 118:335–340

Barcaccia G, Albertini E, Rosellini D, Tavoletti S, Veronesi F (2000) Inheritance and mapping of 2n egg production in diploid alfalfa. Genome 43:528–537

Boeke JD, Corces VG (1989) Transcription and reverse transcription of retrotransposons. Annu Rev Microbiol 43:403–434

Brouwer DJ, Osborn TC (1999) A molecular marker linkage map of tetraploid alfalfa (*Medicago sativa* L.). Theor Appl Genet 99:1194–1200

Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. Focus 12:13–15

Eickbush T (1999) Exon shuffling retrospect. Science 283:1465–1466

Ellis TH, Poyser SJ, Knox MR, Vershinin AV, Ambrose MJ (1998) Polymorphism of insertion sites of *Ty1*-copia class retrotransposons and its use for linkage and diversity analysis in pea. Mol Gen Genet 260:9–19

Flavell AJ, Dunbar E, Anderson R, Pearce SR, Hartley R, Kumar A (1992a) *Ty1*-copia group retrotransposons are ubiquitous and heterogeneous in higher plants. Nucleic Acids Res 20:3639–3644

Flavell AJ, Smith DB, Kumar A (1992b) Extreme heterogeneity of *Ty*-copia group retrotransposons in plants. Mol Gen Genet 231:233–242

Hirochika H, Fukuchi A, Kikuchi F (1992) Retrotransposon families in rice. Mol Gen Genet 233:209–216

Hirochika H, Otsuki H, Yoshikawa M, Otsuki Y, Sugimoto K, Takeda S (1996) Autonomous transposition of the tobacco retrotransposon *Tto1* in rice. Plant Cell 8:725–734

Hull R, Will H (1989) Molecular biology of viral and nonviral retroelements. Trends Genet 5:357–359

Kimura M, Crow JF (1964) The number of alleles that can be maintained in a finite population. Genetics 49:725–738

Kosambi DD (1944) The estimation of map distances from recombination values. Ann Eugen 12:172–175

Kumar A, Bennetzen JL (1999) Plant retrotransposons. Ann Rev Genet 33:479–532

Morgante M, Vogel J (1994) Compound microsatellite primers for the detection of genetic polymorphisms. US Patent Application No. 08/326456

Nei M (1973) Analysis of gene diversity in subdivided populations. Proc Natl Acad Sci USA 70:3321–3323

Pearce SR, Harrison G, Li D, Heslop-Harrison J, Kumar A, Flavell AJ (1996) The *Ty1-copia* group retrotransposons in *Vicia* species: copy number, sequence heterogeneity and chromosomal localisation. Mol Gen Genet 250:305–315

Pejic I, Ajmone-Marsan P, Morgante M, Kozumplick V, Castiglioni P, Taramino G, Motto M (1998) Comparative analysis of genetic similarity among maize inbred lines detected by RFLPs, RAPDs, SSRs, and AFLPs. Theor Appl Genet 97:1248–1255

Pesole G, Liumi S, Grillo G, Saccone C (1997) Structural and compositional features of untranslated regions of eukaryotic mRNAs. Gene 205:95–102

Pouteau S, Huttner E, Grandbastien MA, Caboche M (1991) Specific expression of the tobacco *Tnt1* retrotransposon in protoplasts. EMBO J 10:1911–1918

Powell W, Machray GC, Provan J (1996) Polymorphism revealed by simple sequence repeats. Trends Plant Sci 1:215–222

Sambrook J, Fritsh EF, Maniatis T (1989) Molecular cloning: a laboratory manual (2nd edn). Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Suoniemi A, Tanskanen J, Schulman AH (1998) *Gypsy*-like retrotransposons are widespread in the plant kingdom. Plant J 13:699–705

Tavoletti S, Veronesi F, Osborn TC (1996) RFLP linkage map of an alfalfa meiotic mutant based on an F1 population. J Heredity 87:167–170

Van Ooijen JW, Voorrips RE (2001) JoinMap® Version 3.0, Software for the calculation of genetic linkage maps. Plant Research International, Wageningen, The Netherlands

Vegh Z, Vincze E, Kadirov R, Toth G, Kiss GB (1990) The nucleotide sequence of a nodule-specific gene, Nms-25 of *Medicago sativa*: its primary evolution via exon-shuffling and retrotransposon-mediated DNA rearrangements. Plant Mol Biol 15:295–306

Vicient CM, Suoniemi A, Anamthawat-Jonsson K, Tanskanen J, Beharav A, Nevo E, Schulman AH (1999) Retrotransposon *Bare-1* and its role in genome evolution in the genus *Hordeum*. Plant Cell 11:1769–1784

Vos P, Hogers R, Bleeker M, Reijans M, Van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23:4407–4414

Waugh R, McLean K, Flavell AJ, Pearce SR, Kumar A, Thomas BB, Powell W (1997) Genetic distribution of *Bare-1*-like retrotransposable elements in the barley genome revealed by sequence-specific amplification polymorphisms (S-SAP). Mol Gen Genet 253:687–694

White SE, Habera LF, Wessler SR (1994) Retrotransposons in the flanking regions of normal plant genes: a role for *copia*-like elements in the evolution of gene structure and expression. Proc Natl Acad Sci USA 91:11792–11796