



Optimal population size for RFLP-assisted cultivar identification in alfalfa (*Medicago sativa* L.)

Paola Labombarda, Fulvio Pupilli, Sergio Arcioni

► To cite this version:

Paola Labombarda, Fulvio Pupilli, Sergio Arcioni. Optimal population size for RFLP-assisted cultivar identification in alfalfa (*Medicago sativa* L.). *Agronomie*, 2000, 20 (2), pp.233-240. 10.1051/agro:2000123 . hal-00886006

HAL Id: hal-00886006

<https://hal.science/hal-00886006>

Submitted on 11 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Optimal population size for RFLP-assisted cultivar identification in alfalfa (*Medicago sativa* L.)

Paola LABOMBARDA, Fulvio PUPILLI*, Sergio ARCIONI

Istituto di Ricerche sul Miglioramento Genetico delle Pianta Foraggiere, via Madonna Alta 130, 06128, Perugia, Italy

(Received 4 October 1999; accepted 22 December 1999)

Abstract – The ability of RFLP markers to distinguish between 2 heterogeneous alfalfa ecotypes, “Vogherese” and “Maremmana”, was correlated with the number of plants sampled in either single-plant or bulk analyses. Independent subsample populations of 20 and 50 plants were compared for variance component partitioning and band frequencies in 20-, 50- and 100-plant subpopulations of Maremmana. Homogeneity within 3 independent bulks of the same size as given above was taken as a measure of bulk optimal size. A minimum of 50 plants is required for both single-plant and bulk analyses; however, for a large majority of bands analysed, no significant differences were detected for their frequency among populations of 20, 50 and 100 plants, with the exception of 6 bands out of 48. One ecotype-specific marker was found through bulk analysis. Results are discussed in relation to the improvement of RFLP methodology for cultivar identification in alfalfa.

***Medicago sativa* / RFLP/ cultivar identification**

Résumé – Taille de population optimale pour l'identification des cultivars par RFLP chez la luzerne (*Medicago sativa* L.). La capacité des marqueurs RFLP de distinguer entre deux écotypes hétérogènes de luzerne, « Vogherese » et « Maremmana », a été comparée avec le nombre de plantes composant un échantillon utilisé aussi bien pour l'analyse sur des plantes individuelles que sur des plantes en mélange (« bulk »). La partition des composantes de la variance a été comparée dans des sous-échantillons de populations indépendantes de 20 et 50 plantes, et les fréquences de bandes ont été aussi comparées dans des sous-populations de Maremmana constituées par 20, 50 et 100 plantes. Le degré d'homogénéité révélé à l'intérieur de 3 « bulks » indépendants, de tailles identiques à celles qui viennent d'être mentionnées, a été pris comme mesure de la taille optimale du « bulk ». Un minimum de 50 plantes est nécessaire aussi bien pour l'analyse sur plantes individuelles qu'en « bulk », alors que pour la grande majorité des bandes analysées, aucune différence significative de leur fréquence n'a été relevée entre les populations de 20, 50 et 100 plantes, à l'exception de 6 bandes sur 48. Un marqueur spécifique d'un écotype a été trouvé au moyen de l'analyse « bulk ». Les résultats sont discutés dans l'optique d'une amélioration de la méthodologie RFLP servant à l'identification des cultivars chez la luzerne.

***Medicago sativa* / RFLP / identification des cultivars**

Communicated by Mervyn Humphreys (Aberystwyth, UK)

* Correspondence and reprints
f.pupilli@irmgpf.pg.cnr.it

1. Introduction

Cultivar identification is of the utmost importance in plant breeding to preserve the genetic makeup of varieties and to protect breeders' rights [4]. Genetic fingerprints have proved useful for this purpose, and at present the classification of species and cultivars can be based on isozymes [1] and/or on protein extracted from storage organs [11]. However, DNA-based markers, being independent of the influence of the environment and virtually unlimited in number, provide a more reliable tool for strain/cultivar identification. RFLP technology was developed nearly 20 years ago [18], and is now a well established detection method. Further progress in this domain has involved the development of PCR-based systems, namely RAPDs, SSRs and AFLPs. In general, RFLPs are superior in reproducibility but are labour-intensive and less cost-effective compared to PCR-based markers. Among these, AFLPs and SSRs have proved to be sufficiently effective to replace RFLPs for genotype identification [13, 14], while RAPDs are less reproducible [9] due to mismatch annealing [12]. The choice of the appropriate marker system is critical when high efficiency and reproducibility at low cost is the target. This is especially true when heterogeneous populations of outcrossing and sexually reproduced species are considered. Whereas the cultivars of autogamous and vegetatively propagated species can be represented by one or a few individuals, it is difficult to find cultivar-specific markers for outcrossing and seed-propagated species which are generally heterogeneous populations. For the latter, it is of interest to improve well-established molecular marker systems such as RFLPs in order to select the most informative probe/enzyme combinations and the appropriate size of sample populations.

Alfalfa (lucerne, *Medicago sativa*) is one of the most important leguminous forage crops grown in temperate environments. It is an autotetraploid ($2n = 4x = 32$) outcrossing and seed-propagated species. These factors contribute to the genetic complexity of alfalfa at both individual and population levels. Furthermore, in Italy a significant percentage of cultivated alfalfa is represented by local land races or ecotypes, which may have a

wide genetic base. Two general strategies are followed for molecular marker-assisted cultivar identification in alfalfa: i) computing the genetic distance between populations on the basis of either the fraction of shared bands or the differences in band frequencies [2, 5, 8, 10, 15]; and ii) comparison of DNA fingerprints of pooled DNA samples [10, 15, 20]. Taken together, the available data in the literature suggest that distinguishing alfalfa cultivars through molecular markers is possible when they have a narrow genetic base, or are characterized by various degrees of *Medicago falcata* introgressions, or are land races adapted to distant environments. However, studies focussed on the discrimination between commercial ecotypes of closer origin are still lacking. Pupilli et al. (unpubl. obs.) approached this problem by verifying whether the ecotypes had conserved a genetic identity after years of seed multiplication. In this study, RFLP markers were used to investigate whether different seed lots of the same ecotype were homogeneous within the ecotype, but differed from those of another ecotype, using both single plant and bulk strategies. Single plant analysis showed that the above condition was not satisfied, and bulk analysis revealed few cultivar-specific fragments. Population sampling was limited in this study; therefore additional work regarding the optimal number of individual plants is required.

The aim of the present study was to establish whether the ability of RFLPs to distinguish between 2 heterogeneous ecotypes depended on the number of plants analysed. For this purpose we: i) compared the partitioning of the between-versus within-population variance components for genetic diversity in samples of different size, but belonging to the same populations; ii) investigated whether significant differences in band frequencies existed in different-sized subsamples of the same ecotype; and iii) established the optimal size of bulked samples.

2. Materials and methods

2.1. Plant material

The foundation seed of 2 Italian commercial ecotypes, Vogherese and Maremmana, supplied by

the Istituto Sperimentale per le Colture Foraggere of Lodi, Italy, was used in this study. Two different accessions 8, 9 were considered for Vogherese, whereas Maremmana was represented by one accession only 55. Each accession corresponded to a “foundation farm” responsible for the maintenance and multiplication of the foundation seed. Individual plant populations of 20 and 50 plants for each accession plus an additional population of 100 plants for accession 55 only were considered. For bulk preparation, a single trifoliate leaf was collected from 20, 50 or 100 plants for each accession, and 3 replicates consisting of bulks of different plants, were considered for each sampling size.

2.2. RFLP analysis

Genomic DNA was extracted basically as described by Saghai-Marooof et al. [16] with some modifications. For the single-plant analysis a trifoliate leaf (weight: 20 ± 2 mg) was ground with liquid nitrogen in 2-mL Eppendorf tubes with a glass pestle and the powder obtained was resuspended in 300 μ L of $2 \times$ CTAB buffer (containing 2% β -mercaptoethanol) pre-warmed to 65 °C. The mixtures were incubated at 65 °C for 30 min under gentle agitation. Five hundred μ L of chloroform/isoamyl alcohol (ratio 24:1) was added to the mixtures, which were agitated in a rotary mixer for 15 min at 45 rpm and at room temperature. After centrifugation at 5 000 g for 7 min at room temperature, the aqueous phase was collected and the nucleic acids were precipitated by adding an equal volume of chilled isopropyl alcohol. The tubes were kept on ice for 5 min and then the DNA was pelleted at 6 000 g for 10 min at 4 °C, vacuum-dried and resuspended in 40 μ L TE buffer (Tris-HCl 1 mM; EDTA 0.1 mM). For DNA extraction from bulked plant material, the total fresh weight was measured and the above procedure was scaled up accordingly. Approximately 1 and 8 μ g of DNA was digested for single plants and bulked samples respectively, with 20 units of *Hind*III (New England Biolabs) according to the supplier's instructions. The resulting fragments were fractionated electrophoretically in 1% agarose gel and blotted onto Hybond N+ membranes (Amersham)

according to the standard capillary procedure [19]. Blots were hybridized with 18 32 P-labelled probes from a partial *Pst*I genomic alfalfa library [3], washed and exposed as reported by Gebhardt et al. [7].

2.3. Statistical analysis

RFLP bands were scored as present (1) or absent (0) on autoradiographs to create a binary matrix. All the monomorphic bands (i.e., bands present in all individuals of each accession) were excluded. An estimate of the genetic distance between each pair of individuals was computed on the basis of the squared euclidean distance of Excoffier et al. [6], defined as:

$$E = e_{xy}^2 = n(1 - 2n_{xy}/2n)$$

where n_{xy} is the total number of bands shared by the 2 plants x and y , and n is the total number of polymorphic bands. The distance matrix obtained was subjected to the analysis of molecular variance (AMOVA) [6] to partition the components of variance of total genetic diversity into between-accessions and among-plants within accessions. The number of permutations carried out irrespective of ecotypes and accessions was set at 100 000 for partitioning the overall variance, and at 5 000 with a significance level of 0.01 (1%), for the pair-wise comparisons of the accessions. All the AMOVA analyses were performed with the ARLEQUIN software version 1.1 [17].

3. Results and discussion

The 18 probes used displayed a total of 127 bands, 13 of which were monomorphic and therefore were not considered. The remaining 114 bands constituted the total data set of the AMOVA analysis. The between- and within-population comparison of variance partitioning with samples of 20 or 50 plants of the same accession is shown in Table I. Although most of the overall variance is accounted for by within-population variability (> 98%), a small but significant contribution to the

Table I. Variance partitioning for 20- (A) and 50-plant populations (B) of accessions 8, 9 and 55 using AMOVA.

Source of variation	D.f.	Sum of squares	Variance components	Variation (%)	P ^a
A					
Among populations	2	40.567	27.37×10^{-2}	1.81	< 0.01
Within populations	57	844.100	14.80877	98.19	< 0.01
Total	59	884.667	15.08247		
B					
Among populations	2	48.680	17.46×10^{-2}	1.11	< 0.01
Within populations	147	2294.640	15.60980	98.89	< 0.01
Total	149	2343.320	15.78440		

^a Probability of having more extreme variance components than those observed by random error.

total variance was given by between-population differences in both the 20- and 50-plant samples. The distinctly higher proportion of within-population versus between-population variability can be attributed to either a large heterogeneity of the ecotypes, or to the relative relatedness of the populations under study that had evolved in similar environments. This contrasts with the results of Crochemore et al. [5], who attributed 50% of the total variance to within-population genetic variability by comparing varieties and land races of distinctly different origins. In any case, the sampling of 20 or 50 plants per population did not greatly affect the proportion of total variability distributed within- and between-populations.

However, when the populations underwent pairwise comparison, significant differences were detected between the 2 different population sizes (Tab. II). When 20 plants per population were considered, the percentage of between-population variability compared to the total variance was significant only when accessions 8 and 55 were compared, whereas this source of variation was not significant for the other comparisons irrespective of whether they belonged to the same ecotype or not. When 50 plants were sampled, accession 55 of the Maremmana ecotype differed significantly from accessions 8 and 9 belonging to the Vogherese ecotype. The lack of significance attributable to between-population variability of accession 8 versus accession 9, already noted when comparing 20-plant populations, was confirmed.

Again, the between-population variability values in our study were remarkably lower than that reported by Crochemore et al. [5] in their variety pairwise comparisons for the reason given above.

To investigate whether the detected frequency of a band varied with the size of the population, the frequency of the same band was compared in samples of 20, 50 and 100 plants of accession 55. A χ^2 test of heterogeneity was carried out considering 2 phenotypic classes (presence/absence) for each fragment and data; these were pooled, when necessary, to maintain the expected class frequency at about 5. A subset of 8 probes, selected on the basis of their high polymorphism, revealed 48 informative polymorphic bands. Among these, only 6 showed significant overall heterogeneity indicating the presence of differences in their frequencies

Table II. Percentage and significance (*) at $P < 0.01$ of the between-accession variance compared to total variance for each pairwise comparison of the accessions considered in 20- (A) and 50-plant populations (B).

	8	9
A		
9	1.185	
55	2.254*	1.995
B		
9	0.728	
55	1.084*	1.504*

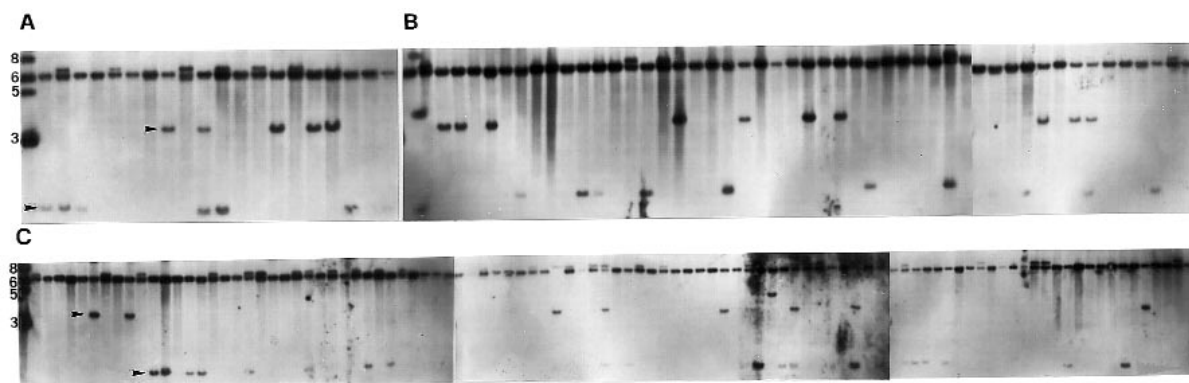
Table III. Chi-square analysis of heterogeneity of band frequencies detected among 20-, 50- and 100-plant subpopulations of accession 55.

	D.f.	1A11 ₂	1A11 ₅	1B4 ₃	2G5 ₁	2G11 ₂	2G11 ₄
Total heterogeneity	2	8.75*	11.78**	10.97**	7.06*	7.21*	6.64*
20 vs. 50	1	5.17*	5.63*	9.79**	4.17*	4.32*	0.21 ^{ns}
(20 + 50) vs. 100	1	3.58 ^{ns}	6.15*	1.18 ^{ns}	2.89 ^{ns}	2.89 ^{ns}	6.43*

*Significant for $0.01 < P < 0.05$; **significant for $P < 0.01$; ns: not significant.

among the 3 subpopulations (Tab. III). Chi-square partitioning showed that the significant heterogeneity was mainly due to the comparison of the 20- versus 50-plant subpopulations, with the exception of bands 1A11₅ and 2G11₄. Curiously enough, the latter band did not show significant differences in its frequency when the two 20- and 50-plant subpopulations were compared with each other, but it did when the 2 pooled populations were compared against the 100-plant subpopulation. To calibrate the system, the appearance (+) or absence (-) of the bands analysed in 100-plant bulks of the same accession was used as control (data not shown). The band 2G11₄, showing a high frequency in 20- and 50-plant populations and a low frequency in 100-plant populations, was clearly detectable in the bulks and this indicated that different fragment frequencies do not necessarily

mean different gene (allele) frequencies. This band is around 3 500 bp long and its relative quantity in the 20-, 50- and 100-plant populations of accession 55 is shown in Figures 1A, B and C respectively. It can be seen that the intensity of this fragment varies among the different plants, probably reflecting a different allele contribution as a consequence of the tetrasomic inheritance of alfalfa. Therefore we can assume that the appearance of the bands in the bulks depends on the dosage, reflecting the high number of alleles bulked, even if band frequency is relatively low. This indicates that at least for those fragments which were of low frequency across alfalfa populations, the allelic dosage can prevail over band frequency, as already suggested by Pupilli et al. [15]. Since the total allelic contribution for each band cannot be established a priori, these bands might bias the real genetic distance

**Figure 1.** Hybridizing banding pattern of the probe 2G11 with the *Hind*III-digested DNA of 20- (panel A), 50- (panel B) and 100-plant (panel C) populations of accession 55 (Maremmiana). Arrowheads indicate the band 2G11₄ (higher molecular weight) and 2G11₅. Molecular weight markers (first lane on the left) are expressed in kbp.

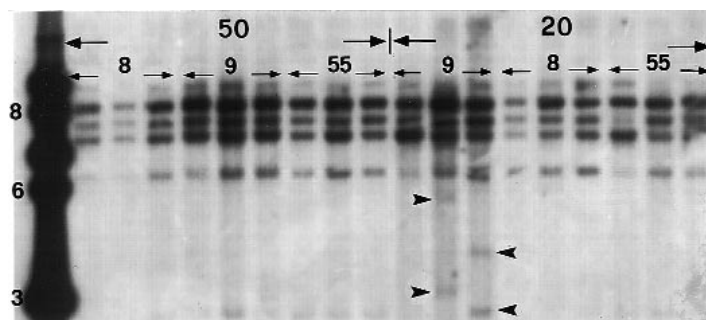


Figure 2. Hybridizing banding pattern of the probe 1B4 with the *Hind*III-digested DNA of 20- and 50-plant bulks (in triplicate) of the 3 accessions 8, 9 (Vogherese) and 55 (Maremmna). The arrowheads indicate polymorphic bands among 20 plant bulks of the same accession. Molecular weight markers (first lane on the left) are expressed in kbp.

among populations when this is estimated on the basis of band frequencies. However, in our results the effect of such bands was negligible; and in most cases the appearance of a band in a bulk depended on its frequency in the population. As an example, the band with lower molecular weight indicated by the arrowhead in Figure 1 showed a frequency above 2G11₄ and it was detected in the bulk despite a different allelic contribution from each plant in the population.

To establish the optimal size of the bulked samples, we compared variability for RFLPs among 3 independent bulks (replicates) each composed of 20, 50 and 100 plants of the same accession. In a survey of 12 probes showing 96 bands in total, 5 showed 11 polymorphic bands among the three 20-plant bulks only. One of these probes is illustrated in Figure 2, in which the arrowheads indicate polymorphic bands among three 20-plant bulks of accession 9, but not in the corresponding 50-plant bulks. In no cases were differences between the 50- and the corresponding 100-plant bulks detected. As expected, such polymorphisms were mainly detected for bands of minor intensity since, as above mentioned, these bands occurred at a low intensity within the populations and therefore were strongly affected by the sampling effect.

Determination of the optimal population size for cultivar identification in outcrossing heterogeneous species is crucial if a lower labour cost is the target. Large-scale experiments, saving space in the greenhouse and cutting down the cost of DNA manipulations, could render RFLPs competitive with PCR-based markers such as AFLPs and SSRs. On the basis of our results we can conclude that

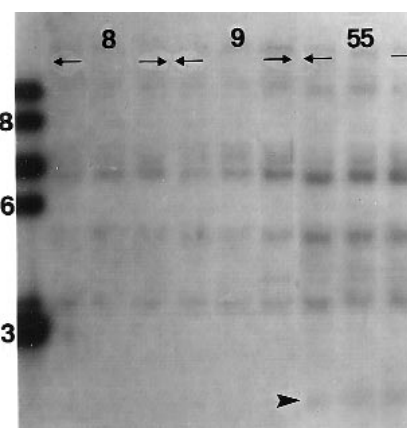


Figure 3. Hybridizing banding pattern of the probe 1C3 with the *Hind*III-digested DNA of 50-plant bulks (in triplicate) of the 3 accessions 8, 9 (Vogherese) and 55 (Maremmna). The arrowhead indicates a Maremmana-specific band. Molecular weight markers (first lane on the left) are expressed in kbp.

ecotype identification in alfalfa either by bulk analysis or pairwise comparisons of individual plants requires a minimum of 50 individuals. Out of the 12 probes used for bulk analysis, one designated 1C3, revealed a fragment (arrowhead, Fig. 3) that is present in all three 50-plant bulks of accession 55 and that is always absent in the corresponding bulks of accessions 8 and 9. This fragment could constitute an example of a Maremmana-specific marker, absent in Vogherese.

Since in general no dramatic differences in band frequency were detected among 20-, 50- and 100-plant samples, we can state that 20 plants are sufficient to represent the genetic variability of an

alfalfa ecotype when cultivar identification is based on this parameter. Gherardi et al. [8] reported 40 plants as the optimal population size for cultivar discrimination in alfalfa. The same authors stated that because of the dominant nature of the markers, RAPD analysis was unlikely to provide a useful tool to measure the real genetic distance between populations; and also that the occurrence of bands of identical size but unknown sequences can underestimate the genetic distance between individuals. Therefore, as stated by the same authors, improved co-dominant markers could be more interesting, and our results seem to confirm this assumption in alfalfa cultivar identification. Italian alfalfa ecotypes may cover large geographic regions, and the problem arises of whether they still conform to the definition of "ecotype". The results presented here indicate that the two ecotypes "Maremmana" and "Vogherese" can be distinguished on the basis of either single plant or bulk analysis. Since these two ecotypes have evolved in similar environments, they probably share a common gene pool. Nevertheless, RFLP analysis has sufficient resolution to detect hidden differences. Therefore good discrimination is expected in the entire set of Italian ecotypes, which in many cases are adapted to very different environments.

Acknowledgements: The authors thank Prof. M.D. Hayward and Dr. P. Rotili for their constructive criticism and advice. This research was supported by a grant from the Ministero delle Politiche Agricole in the context of the project "Caratterizzazione di ecotipi di erba medica iscritti al registro varietale".

References

- [1] Bailey D.C., Isozymic variation and plant breeders rights, in: Tanksley S.D., Orton T.J. (Eds.), *Isozymes in plant genetics and breeding*. Part A, Elsevier, Amsterdam, 1983, pp. 425–440.
- [2] Brummer E.C., Kochert G., Bouton J.H., RFLP variation in diploid and tetraploid alfalfa, *Theor. Appl. Genet.* 83 (1991) 89–96.
- [3] Businelli S., Pupilli F., Damiani F., Arcioni S., Selection of homologous probes for the characterisation of *Medicago* genotypes through RFLP analysis, *J. Genet. Breed.* 47 (1993) 333–340.
- [4] Cooke R.J., Varietal identification of crop plants, in: Skerrett J.H., Appels R. (Eds.), *New Diagnostic in Crop Science, Biotechnology in Agriculture* Vol. 13, CAB International, Wallingford Oxon, UK, 1995, pp. 33–63.
- [5] Crochemore M.L., Huyghe C., Kerlan M.C., Durand F., Julier B., Partitioning and distribution of RAPD variation in a set of populations of the *Medicago sativa* complex, *Agronomie* 16 (1996) 421–432.
- [6] Excoffier L., Smouse P.E., Quattro J.M., Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data, *Genetics* 131 (1992) 479–491.
- [7] Gebhardt C., Ritter E., Debener T., Schachtschabel U., Walkemeier B., Uhrig H., Salamini F., RFLP analysis and linkage mapping in *Solanum tuberosum*, *Theor. Appl. Genet.* 78 (1989) 65–75.
- [8] Gherardi M., Mangin B., Goffinet B., Bonnet D., Huguet T., A method to measure genetic distance between allogamous populations of alfalfa (*Medicago sativa*) using RAPD molecular markers, *Theor. Appl. Genet.* 96 (1998) 406–412.
- [9] Jones C.J., Edwards K.J., Castaglione S., Winfield M.O., Sala F., van de Wiel C. et al., Reproducibility testing of RAPD, AFLP and SSR, markers in plants by a network of European laboratories, *Mol. Breed.* 3 (1997) 381–390.
- [10] Kidwell K.K., Austin D.F., Osborn T.C., RFLP evaluation of nine *Medicago* accessions representing the original germplasm sources for North American alfalfa cultivars, *Crop Sci.* 34 (1994) 230–236.
- [11] Ladizinsky G., Hymowitz T., Seed production electrophoresis in taxonomic and evolutionary studies, *Theor. Appl. Genet.* 54 (1979) 145–151.
- [12] Neale D.B., Harry D.E., Genetic mapping in forest trees: RFLPs, RAPDs and beyond, *Ag. Biotech. News Inf.* 6 (1994) 107N–114N.
- [13] Pejic I., Ajmone-Marsan P., Morgante M., Kozumplick V., Castiglioni P., Taramino G. et al., Comparative analysis of genetic similarity among maize inbred lines detected by RFLPs, RAPDs, SSRs, and AFLPs, *Theor. Appl. Genet.* 97 (1998) 1248–1255.
- [14] Powell W., Morgante M., Andre C., Hanafey M., Vogel J., Tingey S., Rafalsky A., The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) mark-

ers for germplasm analysis, *Mol. Breed.* 2 (1996) 225–238.

[15] Pupilli F., Businelli S., Paolocci F., Scotti C., Damiani F., Arcioni S., Extent of RFLP variability in tetraploid populations of alfalfa, *Medicago sativa*, *Plant Breed.* 115 (1996) 106–112.

[16] Saghai-Marooif M.A., Soliman K.M., Jorgensen R.A., Allard R.W., Ribosomal DNA spacer-length polymorphism in barley: Mendelian inheritance, chromosomal location and population dynamics, *Proc. Natl. Acad. Sci. USA* 81 (1984) 8014–8018.

[17] Schneider S., Kueffer J.M., Roessli D., Excoffier L., Arlequin ver. 1.1: A software for popula-

tion genetic data analysis, Genetics and Biometry Laboratory, Univ. Geneva, Switzerland (1997).

[18] Soller M., Beckmann J.S., Genetic polymorphism in varietal identification and genetic improvement, *Theor. Appl. Genet.* 67 (1983) 25–33.

[19] Southern E., Detection of specific sequences among DNA fragments separated by gel electrophoresis, *J. Mol. Biol.* 98 (1975) 508–517.

[20] Yu K., Paulus K.P., Rapid estimation of genetic relatedness among heterogeneous populations of alfalfa by random amplification of bulked genomic DNA samples, *Theor. Appl. Genet.* 86 (1993) 788–794.