© NIAB 2005
ISSN 1479-2621

Plant Genetic Resources **3**(1); 59–71
DOI: 10.1079/PGR200562

Genetic diversity for seed protein composition in *Medicago truncatula*

C. Le Signor[1]*, K. Gallardo[1], J. M. Prosperi[2], C. Salon[1], L. Quillien[3], R. Thompson[1] and G. Duc[1]

[1]*Unité de recherche en génétique et écophysiologie des légumineuses, BP 86510, F21065 Dijon Cedex,* [2]*UMR-DGPC Station de génétique et d'amélioration des plantes INRA, Domaine de Melgueil, F34130 Mauguio and* [3]*Unité de biochimie et technologie des protéines, INRA, F44316 Nantes Cedex 03, France*

Received 7 September 2004; Accepted 4 January 2005

Abstract

Fifty lines of *Medicago truncatula*, derived from ecotypes or cultivars of diverse geographical origin, were grown under uniform conditions, and variation in seed protein composition and quantity was investigated. One-dimensional electrophoretic profiles revealed 46 major seed polypeptides, of which 26 were polymorphic within the collection. The vicilin/convicilin (7S) and the legumin (11S) type proteins were identified by immunoblotting using antibodies raised against the homologous pea proteins. The polymorphism for the major seed protein classes allowed the clustering of the genotypes into four groups. There was no evidence of clustering according to geographical origin of the lines. However, all lines not belonging to either *M. truncatula* ssp. *truncatula* or ssp. *longispina* were clustered in a single group, demonstrating the value of seed protein profiles in delimiting species boundaries. The Jemalong line group was differentiated early in the dendrogram, and thus represents an ancient clade in the seed diversity of *M. truncatula*. Within-accession variation was investigated for one-dimensional seed profiles, with additional lines obtained from the same ecotypes. As expected for an autogamous species, within-accession variation was low. Seed protein content was highly variable among the 50 lines examined. Lines contrasting for qualitative traits and seed protein content were identified to allow for the genetic determination of these characters.

Keywords: ecotype; electrophoresis; genetic variation; immunoblotting; *Medicago truncatula*; seed protein content

Introduction

Seed storage proteins represent a major component of human protein intake, either directly, or indirectly via their consumption by livestock. Whereas cereal grains contain less than 15% protein, legume seed generally contains more than 20% (Payne, 1983). On a world basis, legume seeds are the second most important protein source after cereals, and provide proteins complementary to those of cereals in terms of essential amino acid content (Perrot, 1995). Seed storage proteins are widely distributed among higher plants, and are synthesized in large quantities during seed development, before becoming sequestered in protein storage vacuoles. Upon germination, their degradation provides a source of nitrogen to support the early growth of the seedling. Storage proteins are composed of a few to many polypeptide species, with individual polypeptides accounting for up to 50% of the total (Croy and Gatehouse, 1985). A specific seed protein comprising over 5% of total protein at maturity, and which is degraded during germination, is assumed to have a storage function. In dicotyledonous plants, seed storage proteins have been classified as either albumin or globulin types, depending on their solubility in water or dilute salt solution. The globulins account for up to 70% of the total seed nitrogen (Croy and Gatehouse, 1985),

* Corresponding author. E-mail: lesignor@epoisses.inra.fr

https://doi.org/10.1079/PGR200562 Published online by Cambridge University Press

and in the legumes they consist of two major protein families, the 7S vicilin/convicilin and the 11S legumin groups, which are differentiated by molecular mass. Both groups are holoproteins composed of regularly assembled subunits (Bewley and Black, 1994). They differ in their digestibility and amino acid content, and therefore their relative abundance affects the nutritional value of the seed (Perrot, 1995).

Recently, we have identified 84 *M. truncatula* proteins with characteristic developmental patterns of accumulation associated with protein deposition during seed development in Jemalong line J5 (Gallardo *et al.*, 2003). The seed of Jemalong is rich in protein (39% of seed dry matter), low in starch (less than 1%), and rich in lipids (13%) and cellulose (10%). Moreover, the proteins are richer in tryptophan (1.5% of the total proteins) than those of either *Vicia faba* or pea (0.9%; Duc, 2004). Genetic diversity for the quantity and quality of legume seed proteins has been investigated in the genera *Pisum* (Schroeder, 1982), *Arachis* (Bertozo and Valls, 2001) and *Vicia* (Potokina and Eggi, 1997), and in *Medicago sativa* (Krochko and Bewley, 2000), but not in the model legume *M. truncatula*. Because this species is closely related to the major legume crops, the genetic determination of its seed quality traits will be relevant for legume crop improvement. In this paper, we describe the seed nitrogen content and protein composition, and their genetic variability, in 50 lines of *M. truncatula*. Using this information, we have been able to identify *M. truncatula* genotypes suitable as parents for the generation of segregating populations useful for the genetic determination of legume seed quality traits.

Material and methods

Plant material

Fifty accessions of the self-pollinating annual species *M. truncatula* were chosen, based on geographical, morphological and isoenzymatic differences, and seed weight (INRA-Montpellier, France) (Table 1):

- Thirty accessions: ecotypes from Mediterranean regions that all belong to *M. truncatula* ssp. *truncatula* or ssp. *Longispina*, except DZA_220 (ssp. *tricycla*), and F11_013 which is a species mixture (*M. truncatula* and *M. littoralis*), the two species being interfertile.
- Ten accessions: cultivars registered in the Australian or French catalogues. Harbinger is derived from *M. littoralis*.
- Ten reference lines: six lines (DZA_315.16, DZA_315.26, DZA_45.5, DZA_45.6, J6 and F83_5.5) are parents of recombinant inbred lines obtained from ecotypes (DZA_315 and DZA_45) or cultivars (Jemalong and F83_5 = Salernes) (INRA-Montpellier). R108_C3 was derived from an Israeli accession (108-1) close to ssp. *tricycla* (J. M. Prosperi, unpublished data), and was selected for *in vitro*

Table 1. Accessions of *Medicago truncatula* used in this study: the genetic type (ecotype, cultivar and line), presence (S, selection) or absence (W, wild) of selection were added as well as the geographical origin

Accession	Germplasm source	Geographical origin
CRE_005	Ecotype W	Crete
GRC_033; GRC_063; GRC_065; GRC_098	Ecotypes W	Greece
DZA_327; DZA_233; DZA_230; DZA_220; DZA_210; DZA_202; DZA_055; DZA_022	Ecotypes W	Algeria
ESP_041; ESP_042; ESP_045; ESP_080; ESP_100; ESP_155; ESP_161; ESP_171	Ecotypes W	Spain
PRT_176; PRT_180	Ecotypes W	Portugal
F20_086; F20_058; F20_047; F20_009; F11_013; F11_008; F11_005	Ecotypes W	France
Salernes	Cultivar S	French catalogue
Borung; Caliph; Cyprus; Jemalong	Cultivars S	Australian catalogue
Parabinga; Sephi; Paraggio; Mogul	Cultivars S	Australian catalogue
Harbinger (*M. littoralis*)	Cultivar S	Australian catalogue
DZA_315.26; DZA_315.16	Ecotype DZA_315 W	Algeria
J6	Jemalong S	Australia
DZA_45.5; DZA_45.6	Ecotype DZA_45 W	Algeria
F83_5.5	Salernes S	France
M192	Cultivar Ghor S	Australia
R108_C3	Line *in vitro*-selected	–
J5	Jemalong S	Australia
TRV25	Mutant of J5 S	–

https://doi.org/10.1079/PGR200562 Published online by Cambridge University Press

regenerability (three cycles) and transformation studies (Hoffman *et al.*, 1997). The non-nodulating non-nitrogen-fixing gamma-irradiation mutant TRV25 derived from J5 (Sagan *et al.*, 1998) was added. Among the 40 ecotypes/cultivars, two mother plants were self-pollinated, one in INRA Dijon (*D* lines) and one in INRA Montpellier (*M* lines), so that each accession was represented by a pair of lines. The *M* lines from ecotypes ESP_080 and F11_008 and cultivar Caliph were infertile, so that only 37 *D*–*M* pairs of lines were analysed.

The *D* lines and reference lines were grown in pouzzolane, an inert medium made of volcanic lava, under greenhouse conditions, with automatic watering with a solution of 3.5:3.1:8.6 N:P:K. The temperature was maintained between 14 and 30°C, with supplementary lighting to provide a 16-h day. Plants were not inoculated with *Sinorhizobium* sp. bacteria, adequate inorganic nitrogen being supplied. Seed yield and protein content were similar for J5 and TRV25, confirming that the nitrogen supply was sufficient. The 37 *M* lines and reference line J6 were multiplied under similar greenhouse conditions. These lines were used only to assess within-accession variability, in comparison to sibling *D* lines produced in Dijon.

Protein extraction

Two independent total protein extractions were conducted using 20 seeds from each of the 87 lines. Whole seed was ground in liquid nitrogen, and 20 mg of powder was suspended in 1.5 ml of extraction buffer (0.125 M Tris–HCl pH 6.7, 2% SDS, 5% glycerol; Schroeder *et al.*, 1982) and stirred for 30 min at room temperature. After centrifugation (14,000 *g* for 10 min), 2% (v/v) mercaptoethanol was added, and the supernatant was incubated at 90°C for 2 min. Aliquots of the supernatant were stored at −80°C. For two-dimensional (2-D) electrophoresis, proteins were extracted in a thiourea/urea lysis buffer (Gallardo *et al.*, 2002b). Protein concentration was estimated according to Bradford (1976).

Electrophoresis procedure

One-dimensional electrophoresis

Proteins (10 μg of protein per well) were fractionated on 1.5 mm SDS polyacrylamide slab gels (Laemmli, 1970). Electrophoresis was performed for 2 h at 50 mA followed by 4 h at 70 mA in Tris–glycine electrode buffer (0.024 M Tris base and 0.192 M glycine, pH 8.3) containing 0.1% SDS. Standard proteins (BioRad, broad range standards) were used to estimate the molecular weight of the protein bands. Two replicate 1-D gels were done from both protein extractions. Gels were stained using silver nitrate (Blum *et al.*, 1987).

Two-dimensional electrophoresis

The 2-D electrophoresis was carried out with total seed protein extracts of line J5, as previously described (Gallardo *et al.*, 2002a). Two replicate 2-D gels were produced from independent extractions using 30 μg protein (analytical gels) or 60 μg protein (preparative gels). Analytical gels were stained with silver nitrate according to Blum *et al.* (1987) and preparative gels were stained with Coomassie Brilliant Blue G-250 according to Mathesius *et al.* (2001). The major spots were excised from preparative gels, digested by trypsin and analysed by MALDI-TOF mass spectrometry (Flanders Interuniversity Institute for Biotechnology, Gent, Belgium).

Western blotting

Following electrophoreses, proteins were transferred onto a polyvinylidene difluoride (PVDF; Amersham Biosciences) membrane using a semi-dry electroblotter (Amersham Biosciences). The membrane was probed with immunoglobulin G (IgG) purified from polyclonal antiserum (rabbit) raised against either pea legumins (1/5000 dilution) or pea vicilins (1/10,000 dilution) (Gueguen *et al.*, 1984). Detection was via peroxidase-linked anti-rabbit antibodies, visualized by chemiluminescence (ECL kit; Amersham).

Nitrogen content

Total percentage nitrogen was measured by the Dumas procedure (Allen, 1974) using a CHN analyser (Carlo-Erba). Six estimations were carried out on individual seeds from each of the 50 lines. Crude protein was calculated using the coefficient %N × 5.10 (J. P. Boutin, personal communication).

Statistical analysis

Analysis of variance and hierarchical clustering analysis were performed with SAS software (SAS Institute Inc., 1990). Similarity indices were calculated according to Brown *et al.* (1979), such that $S = n_{XY}/(n_X + n_Y - n_{XY})$, where n_{XY} is the number of identical sites shared by the two populations, and n_X, n_Y are the numbers of sites in population X and Y, respectively.

https://doi.org/10.1079/PGR200562 Published online by Cambridge University Press

Results

Strategy adopted to analyse genetic variation in seed protein composition

Eight, 10 and 12% polyacrylamide gels were compared for effectiveness in resolving proteins from *M. truncatula* seeds. The former resolved proteins from 30 to 150 kDa, whereas the latter was optimal in the range 15–75 kDa. A total of 46 abundant protein bands were resolved using a combination of 8 and 12% gels, of which 26 were polymorphic. The 10% gels resolved all the polymorphic bands and were therefore used for immunoblotting (Fig. 1). To analyse globulin polymorphism, eight profiles showing all detectable polymorphisms were immunoblotted with pea vicilin or pea legumin antibodies (Fig. 2). High molecular weight (HMW) convicilin/vicilin polypeptides were revealed by the pea anti-vicilin antibodies (Fig. 1, arrows a and b), and acidic (α) legumin chains were detected by the pea anti-legumin antibodies (Fig. 1, arrows c and d). The major polymorphic proteins detected by silver staining corresponded to the 7S or 11S globulins. The immunodetection results were then confirmed by mass spectrometry. Because certain bands detected by anti-legumin and anti-vicilin antibodies overlapped in 1-D separations, proteins were resolved by 2-D electrophoresis prior to mass spectrometry analyses (Fig. 3). To confirm the correspondence between the 1-D and the 2-D patterns on the same gel, the same protein extract was loaded in a parallel well prior to the second dimension run. The 2-D protein pattern was immunoblotted with antibodies raised against vicilins and legumins. Ten abundant

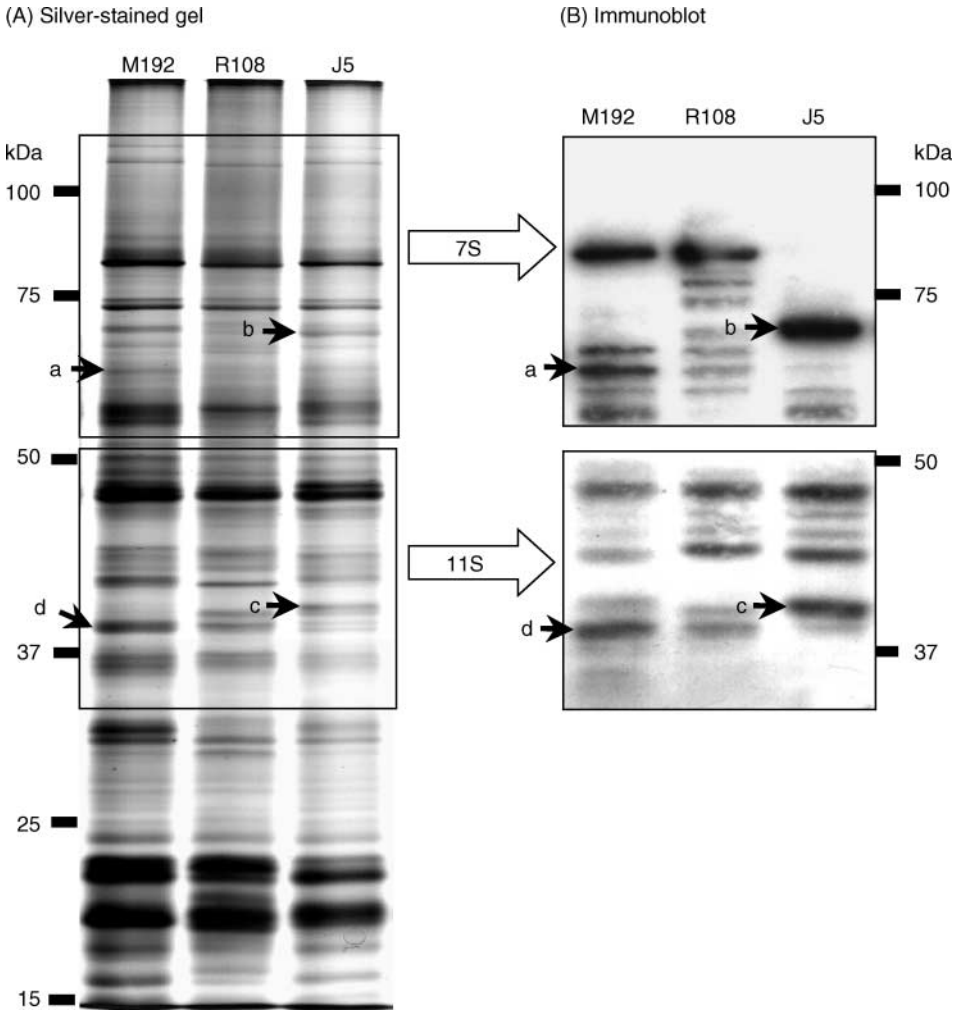


Fig. 1. Polymorphism in the protein profiles from mature seeds of three *Medicago truncatula* lines (M192, R108_C3 and Jemalong J5). (A) 1-D SDS–PAGE separation, in 10% acrylamide gels under reducing conditions. (B) Western-blotted 1-D gels. Arrows indicate four polymorphic bands identified by immunoblotting as vicilin/convicilin (a, b) and legumin (c, d).

https://doi.org/10.1079/PGR200562 Published online by Cambridge University Press

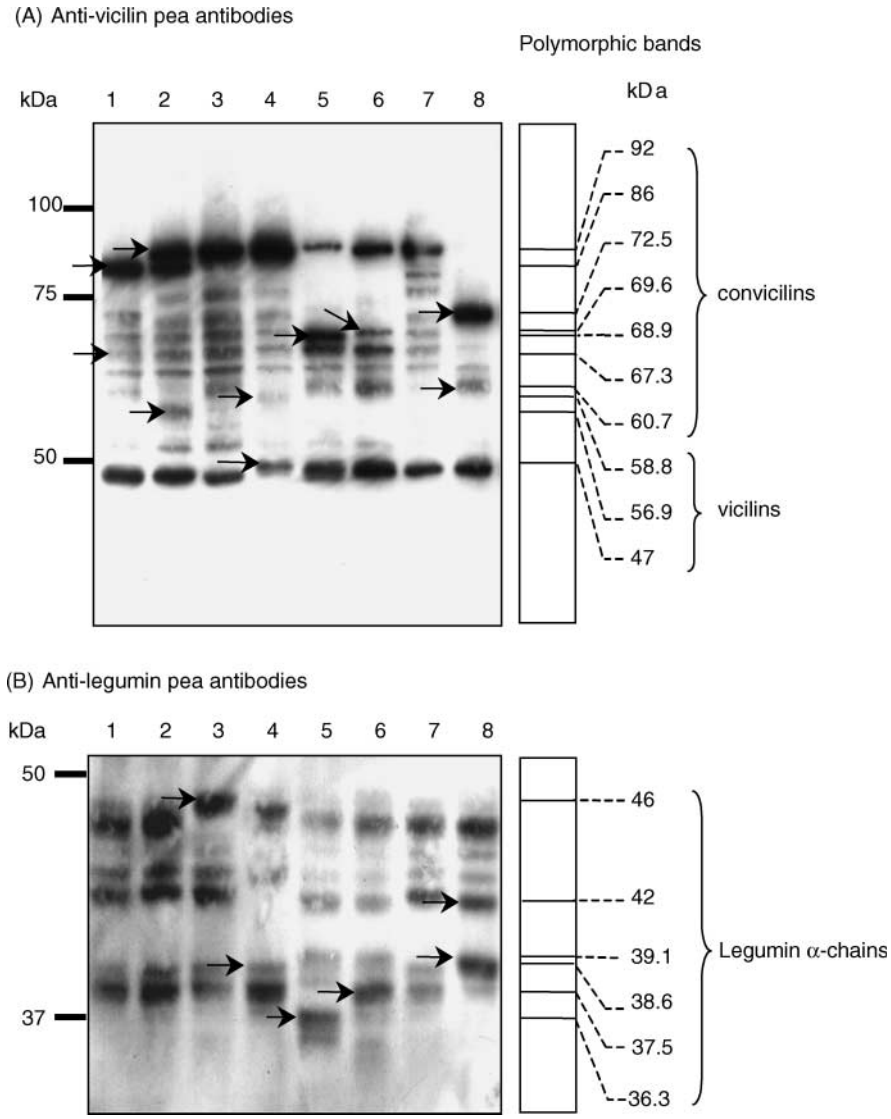


Fig. 2. Detection of legumin-like and vicilin-like storage proteins from protein extracts of mature *Medicago truncatula* seeds. Proteins were separated by 1-D SDS–PAGE in 10% polyacrylamide gels, and immunoblotted with anti-pea vicilin antibodies (A) and anti-pea legumin antibodies (B). Lane 1, GRC_065-*D*; lane 2, DZA_220-*D*; lane 3, F11_013-*D*; lane 4, Harbinger-*D*; lane 5, DZA_45.6; lane 6, M192; lane 7, R108_C3; lane 8, J5. Polymorphic bands indicated by arrows in the left panel are represented in graphic form in the right panel.

spots of sizes identical to the major polymorphic bands detected on the 1D gels were analysed by mass spectrometry. One protein spot of 73 kDa was identified as convicilin and three, of about 46 kDa, were as vicilins. Several polypeptides corresponded to α-chains (ranging in size from 37 to 45 kDa) or β-chains (~23 kDa) of legumin. This confirmed the specificity of the antibodies toward the 7S (vicilin and convicilin) and 11S (legumin) globulins (Fig. 3). Furthermore, the correspondence between the 1-D and the 2-D profiles on the same gel allowed the identification of 16 out of 26 of the major protein bands in the 1-D profiles as 7S or 11S globulins (Figs 2 and 3; Table 2).

Analysis of variability in seed protein quality between lines

Thirty-one distinct seed protein profiles were identified among the 1-D patterns of the 50 lines produced in Dijon. A few profiles were common to more than one line, but at least 25 profiles were line-specific. Of the 26 polymorphic bands, five were present only in single lines, and eight were present in five lines or fewer (Table 2), i.e. half of the observed polymorphism was represented by rare alleles. A hierarchical clustering analysis was performed based on the 26 polymorphic bands

https://doi.org/10.1079/PGR200562 Published online by Cambridge University Press

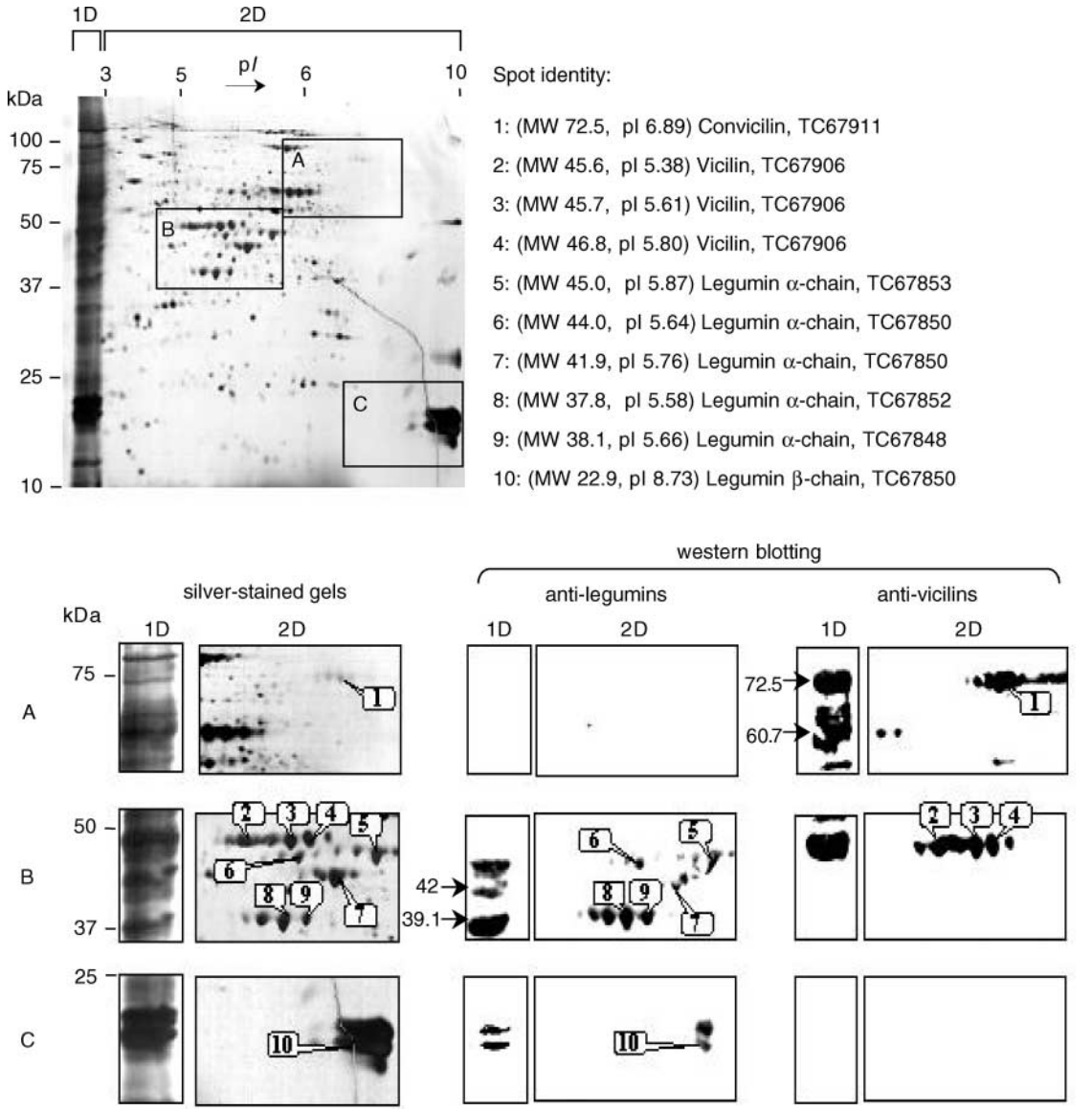


Fig. 3. Characterization of legumin-like and vicilin-like storage proteins from mature seed extracts of *Medicago truncatula* line Jemalong. Proteins were separated by 2-D gel electrophoresis. Major spots (1–10) were identified by MALDI-TOF mass spectrometry. These spots were detected by the anti-vicilin or anti-legumin antibodies and corresponded to the major bands on the 1-D separation. Theoretical masses, pI and Tag Consensus (TC) numbers are indicated for each of the proteins identified. Arrows indicate polymorphic bands (see Fig. 2).

using the SAS 'Cluster' procedure and Ward's minimum variance method (Fig. 1A; Table 2). The resulting dendrogram (Fig. 4) generated four main groups. Group I was represented by four lines including R108_C3 and Harbinger (*M. littoralis*); group II by 10 lines, including M192 and F83_5.5; group III by 18 lines, including all the Algerian parents of RILs; and group IV by 18 lines, including the five Jemalong lines clustered in the same subgroup.

The 13 rare alleles were all in group I lines. The divergence between groups I and II, and groups III and IV was due mainly to protein bands at 92 kDa (HMW convicilin-like; Fig. 2A) and the doublet α-subunits of legumin (37.5 and 36.3 kDa; Fig. 2B). The 69.6 kDa convicilin-like polypeptide (Fig. 2A) was specific to group II. Groups III and IV were differentiated by the presence or absence of the 37.5 and 36.3 kDa α-subunits of legumin. The Jemalong lines within group IV formed a distinct subgroup in which all entries possessed the same profile, characterized by the presence of a specific HMW convicilin (72.5 kDa) and the absence of the common convicilins at 86 and 67.3 kDa.

Similarity indices (Brown *et al.*, 1979) on the entire profile of 46 bands provided a global view of the real polymorphism within the sample (total matrix not

https://doi.org/10.1079/PGR200562 Published online by Cambridge University Press

https://doi.org/10.1079/PGR200562 Published online by Cambridge University Press

Table 2. Schematic representation of the 1-D SDS–PAGE profiles of the 50 lines grown in Dijon

1-D band size (kDa)	CRE_ 005- *D*	ESP_ 171- *D*	F20_ 009- *D*	Cyprus- *D*	Parabinga- *D*	PRT_ 180- *D*	ESP_ 161- *D*	GRC_ 065- *D*	GRC_ 098- *D*	F11_ 005- *D*	DZA_ 327- *D*	DZA_ 202- *D*	PRT_ 176- *D*	ESP_ 042- *D*	Jemalong- *D*	TRV25	J6	J5
161	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
155	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
151	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1
92.0[a]	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0
86.0[a]	1	1	1	1	1	1	1	1	0	0	1	1	0	1	0	0	0	0
74.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
72.5[a]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1
69.6[a]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
68.9[a]	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0
67.3[a]	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0
60.7[a]	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
58.8[a]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
56.9[a]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
54.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
47.0[a]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
46.0[b]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
42.0[b]	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
39.1[b]	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1
38.6[b]	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
37.5[b]	1	1	1	1	1	0	1	1	1	1	1	1	0	1	0	0	0	0
36.3[b]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21.5	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1
21.1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
20.3	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1
19.6	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0
19.2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	◄																	►
										IV								

1-D band size (kDa)	GRC_033-D	ESP_045-D	ESP_080-D	DZA_233-D	DZA_230-D	F20_086-D	F20_058-D	F20_047-D	Sephi-D	GRC_063-D	DZA_210-D	DZA_055-D	DZA_022-D	DZA_45-6	DZA_45-5	Caliph-D	DZA_315-26	DZA_315-16
161	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1
155	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
151	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0
92.0[a]	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1
86.0[a]	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	1	0	0
74.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
72.5[a]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
69.6[a]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
68.9[a]	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
67.3[a]	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1
60.7[a]	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
58.8[a]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
56.9[a]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
54.5	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
47.0[a]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
46.0[b]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
42.0[b]	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
39.1[b]	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
38.6[b]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
37.5[b]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
36.3[b]	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
21.5	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
21.1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
20.3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
19.6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19.2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
										III								

https://doi.org/10.1079/PGR200562 Published online by Cambridge University Press

1-D band size (kDa)	ESP_ 041-*D*	Salernes-*D*	F83_ 5-5	ESP_ 100-*D*	ESP_ 155-*D*	Borung-*D*	Mogul-*D*	F11_ 008-*D*	Paraggio-*D*	M192	DZA_ 220-*D*	R108	F11_ 013-*D*	Harbinger-*D*
161	0	0	0	0	0	0	0	0	0	0	0	0	0	1
155	0	0	0	0	0	0	0	0	0	0	0	0	1	0
151	1	1	1	1	1	1	1	1	1	1	1	1	0	0
92.0[a]	1	1	1	1	1	1	1	1	1	1	1	1	1	1
86.0[a]	0	0	0	0	0	0	0	0	0	0	1	0	0	0
74.5	0	0	0	0	0	0	0	0	0	0	0	0	0	1
72.5[a]	0	0	0	0	0	0	0	0	0	0	0	0	0	0
69.6[a]	1	1	1	1	1	1	1	1	1	1	0	0	0	0
68.9[a]	0	0	0	0	0	0	0	0	0	0	0	0	0	0
67.3[a]	1	1	1	1	1	1	1	1	1	1	1	1	1	1
60.7[a]	1	1	1	1	1	1	1	1	1	1	0	0	1	0
58.8[a]	0	0	0	0	0	0	0	0	0	0	0	0	0	1
56.9[a]	0	0	0	0	0	0	0	0	0	0	1	0	0	0
54.5	0	0	0	0	0	0	0	0	0	0	0	0	0	1
47.0[a]	0	0	0	0	0	0	0	0	0	0	0	0	0	1
46.0[b]	0	0	0	0	0	0	0	0	0	0	0	0	1	0
42.0[b]	1	1	1	1	1	1	1	1	0	1	1	1	1	0
39.1[b]	1	1	1	1	1	1	1	1	1	1	0	0	0	0
38.6[b]	0	0	0	0	0	0	0	0	0	0	1	1	1	1
37.5[b]	0	0	0	1	1	1	1	1	1	1	1	1	1	1
36.3[b]	0	1	1	0	0	0	0	0	0	0	0	0	0	0
21.5	1	1	1	1	1	1	1	1	0	0	1	1	1	0
21.1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
20.3	1	1	1	1	1	1	1	0	1	1	1	1	1	1
19.6	0	0	0	0	0	0	0	1	0	0	0	0	0	0
19.2	0	0	0	0	0	0	0	0	0	0	1	1	1	1
	◄									►	◄			►
						II							I	

Only polymorphic bands from 19 to 161 kDa are shown: 1 corresponds to band presence and 0 to absence of band. Genotypes were sorted according to the similarity of their profiles (order of the dendrogram shown in Fig. 4). The groups are indicated at the bottom of the table.
[a]Protein bands that were revealed after immunoblotting the 1-D gels with antibodies directed against vicilins/convicilins (see Fig. 2).
[b]Protein bands that were revealed after immunoblotting the 1-D gels with antibodies directed against legumins (see Fig. 2).

https://doi.org/10.1079/PGR200562 Published online by Cambridge University Press

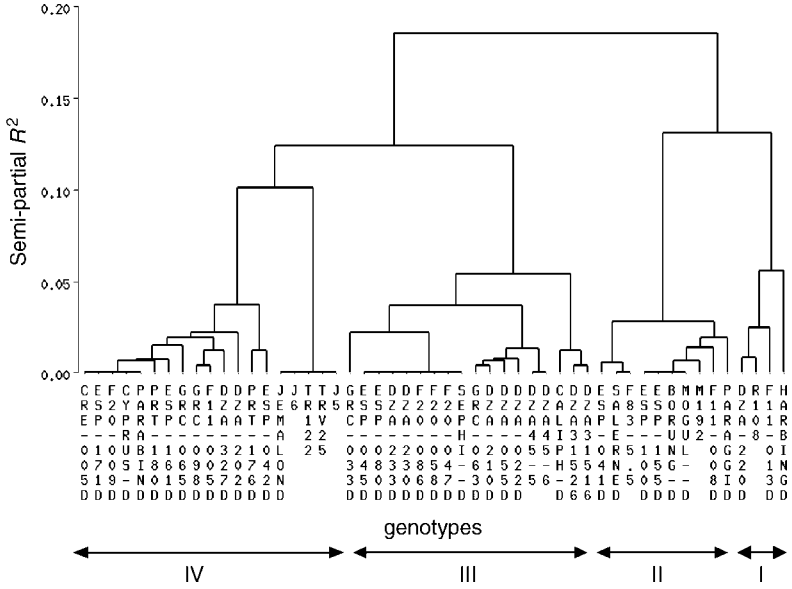


Fig. 4. Dendrogram of the 1-D profiles of the 50 *Medicago truncatula* lines grown in Dijon using Ward's minimum variance method.

shown). Within groups II, III and IV, mean values of similarity were above 90%. Between these three groups, mean similarities were 85–87%, indicating uniformity between the profiles of all these lines. Lower similarity values were obtained for group I with a mean within-group similarity of 81% and mean inter-group (between I and other groups) similarities ranging from 71 to 75%. Harbinger-*D*, belonging to *M. littoralis*, scored the lowest values in the similarity matrix ranging from 55 to 77% for a mean value of 63%. Thus the 50 lines evidently do not cluster according to their geographical origin. For example, Algerian lines were found in groups I, III and IV, and Spanish lines in group II, III and IV.

Variability analysis for seed protein quality within accessions

Within-accession variability was studied by comparing profiles between the *D*–*M* pairs of lines. Differences were observed in 14 protein bands (from 155 to 19.2 kDa) and 15 pairs of lines (14 pairs from ecotypes and one pair from a cultivar) (Table 3). Between one and three changes occurred in 14 pairs of line profiles, although one pair of lines (from ecotype F11_013) showed eight differences between the two profiles. Variation appeared most frequently for the 86 kDa (HMW convicilin), and the 37.5 and 36.3 kDa (α-subunits of legumin) proteins. The profile of line F11_013-*M* included a unique 21.8 kDa band. This ecotype was selected from a mixed population of *M. truncatula* and *M. littoralis*. As F11_013-*D* and F11_013-*M* both belong to group I, and show a similarity of 82%, they probably represent progeny of hybrids between the two species. From the hierarchical classification of the 37 pairs of lines, the resulting dendrogram revealed four cases where the two *D*–*M* lines were assigned to different groups (distance greater than 0.12). The maximum distance between *D* and *M* lines was observed for the three ecotypes ESP_171, GRC_063, DZA_055, and the cultivar Mogul which showed polymorphism for the most discriminatory protein bands, the two α-subunits of legumins (at 37.5 and 36.3 kDa). A difference in the 86 kDa protein band resulted in a maximum distance between *D* and *M* lines of 0.05.

Variability for total seed protein content

Substantial variation, fitting a Gaussian distribution, was observed for protein content among the 50 lines (Fig. 5). The Jemalong reference lines were close to the median for protein content. Selected lines (denoted S in Table 1) displayed lower protein content than those derived from wild ecotypes (denoted W in Table 1). French ecotypes showed the lowest mean values for nitrogen content, whereas Greek ecotypes showed the highest values.

Discussion

This study was conducted to investigate intra-specific variation in protein subunit composition and protein content in seeds of *M. truncatula*, and to examine possible associations between storage protein polymorphism, total protein content, and geographical origin. Since all

https://doi.org/10.1079/PGR200562 Published online by Cambridge University Press

https://doi.org/10.1079/PGR200562 Published online by Cambridge University Press

Table 3. 1-D protein profiles of *M* lines

	ESP_171-*M*	PRT_180-*M*	DZA_327-*M*	DZA_202-*M*	PRT_176-*M*	GRC_033-*M*	ESP_080-*M*	DZA_233-*M*
155	0	0	0	**0**	0	0	0	0
151	1	1	1	**1**	1	1	1	1
86.0	**0**	**0**	**0**	1	**1**	**0**	0	**0**
54.5	0	0	0	0	0	0	0	0
42.0	1	1	1	1	1	**1**	1	1
39.1	1	1	1	1	1	1	1	1
38.6	0	0	0	0	0	0	0	0
37.5	**0**	**1**	1	1	0	0	0	0
36.3	**1**	0	0	0	0	1	0	1
21.8	0	0	0	0	0	0	0	0
21.5	1	1	1	1	1	**1**	1	1
20.3	1	1	1	1	0	1	1	1
19.6	0	0	0	0	1	0	0	0
19.2	0	0	0	0	0	0	0	0

	GRC_063-*M*	DZA_210-*M*	DZA_055-*M*	ESP_041-*M*	ESP_100-*M*	Mogul-*M*	F11_013-*M*
155	0	0	0	0	0	0	**0**
151	1	1	1	1	1	1	**1**
86.0	**1**	**1**	0	**1**	0	0	0
54.5	0	0	**0**	0	0	0	0
42.0	1	1	1	1	1	1	**0**
39.1	1	1	1	1	1	1	**1**
38.6	0	0	0	0	0	0	**0**
37.5	**1**	0	**1**	0	1	**0**	1
36.3	**0**	1	**0**	0	0	**1**	0
21.8	0	0	0	0	0	0	**1**
21.5	1	1	1	1	1	1	**0**
20.3	1	1	1	1	**0**	1	1
19.6	0	0	0	0	**1**	0	0
19.2	0	0	0	0	0	0	**0**

Only profiles showing polymorphism with *D* lines are shown (15 profiles). Differences in profiles are indicated in bold: 0 in line *M* corresponding to 1 in line *D* and vice versa.

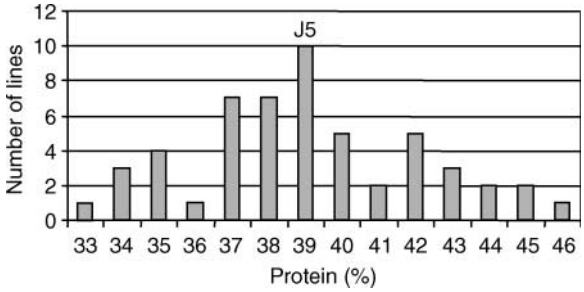


Fig. 5. Distribution of 50 *Medicago truncatula* lines according to seed protein content. The median position of line Jemalong J5 is indicated.

lines were grown under similar environmental conditions, the variation in these traits was genetically determined. As the reference line J6 and most selected cultivar lines showed the same protein profiles in the two experimental locations, the few differences observed between representatives of ecotypes grown at the two locations were attributed to genetic variation within the seed sample of the ecotype.

Genetic variability for seed protein profiles

Comparative analysis of seed protein profiles has been a powerful tool for species identification, for clarifying taxonomic and evolutionary problems, and for studying genetic diversity (Ladizinsky and Hymowitz, 1979). The seed protein profiles analysed in the present study showed a high degree of similarity among different accessions within the species *M. truncatula* (Fig. 1A). Previous studies at the intra-specific level have also reported that, despite morphological and ecological differences, genotypes possessed very similar protein profiles. Potokina and Eggi (1997), investigating the intraspecific variability of *Vicia sativa* and *V. angustifolia* protein profiles, found only minor variation within the *V. sativa* profiles. In *Pisum sativum*, Mishra *et al.* (2000) report similarity indices ranging from 53 to 100% for seed protein profiles. In *Medicago sativa*, Krochko and Bewley (2000) observed very similar 1-D and 2-D seed protein patterns among 27 varieties, consistent with our findings for *M. truncatula*.

The comparative analysis of 1-D profiles and immunoblots allowed the identification of the major seed storage proteins. The acidic subunits of legumin ranged in size from 36 to 46 kDa, the basic chains of legumin ranged from ~23 to 24 kDa, the vicilins from 59 to 47 kDa, and the convicilins from 60.7 to 92 kDa, values similar to those reported for *Medicago sativa* (Krochko and Bewley, 1988). Despite overall similarity in the profile of major *M. truncatula* seed proteins, we found extensive polymorphism in more minor proteins, acidic legumin polypeptides, certain HMW vicilins and other poorly characterized seed proteins (Fig. 2; Table 2). Interestingly, Krochko and Bewley (2000) have reported similar results for *M. sativa*. The minor differences in our protein profiles allowed successful hierarchical clustering of the 50 lines into well-separated groups. All *M. truncatula* ssp. *tricyla* (DZA_220-*D*, R108_C3) and *M. littoralis* (Harbinger-*D*) lines, as well as a probable hybrid between the two species (F11_013-*D*) clustered in one group, demonstrating the utility of seed protein profiles to delimit species boundaries. Indeed, the similarity indices allowed a straightforward discrimination between *M. littoralis* and *M. truncatula* (the lowest values were obtained for Harbinger-*D* and -*M* with a mean S of 63%), despite their phylogenetic proximity according to rDNA spacer comparisons (Bena *et al.*, 1998). Furthermore, the dendrogram showed that the differentiation of the Jemalong group represented an early event in genetic diversity within *M. truncatula*. Consequently, selection of lines distinct from the Jemalong reference for genetic analyses of seed characters should be straightforward.

Intra-accession polymorphism for seed protein patterns has been seldom investigated. Our results demonstrated a low level of polymorphism within accessions of *M. truncatula*, compared to the inter-accession level, as expected for a self-pollinating species, and in agreement with other legume seed protein data (Bertozo and Valls, 2001).

Interrelation between seed protein profiles and protein content

The present study provides evidence for extensive genetic variability in seed protein content in the model legume *M. truncatula*, consistent with high genetic variability for seed protein content in other legumes, such as pea (Schroeder, 1982; Bastianelli *et al.*, 1998). Significant variation in protein content was detected between groups of lines from different geographical origins. More selected lines had lower protein content than wild lines, suggesting that selection for higher forage yield in cultivars may have a negative effect on seed protein content. While variation in protein content was partially explicable on the basis of geographical origin, this was not the case for polypeptide composition, similar to the situation in *V. sativa* (Potokina and Eggi, 1997). In addition, no correlation between nitrogen content and protein profile was observed. Consequently, lines with high or low values of nitrogen content could be selected within each group of protein profiles, suggesting that the loci responsible for protein and N content do not correspond to the major storage protein genes. From these results, contrasting pairs of lines have been selected for

https://doi.org/10.1079/PGR200562 Published online by Cambridge University Press

further study of the genetic control of seed protein content and of protein quality.

Acknowledgements

We are very grateful to Françoise Moussy and Florence Cassecuelle for their valuable technical assistance. Thanks to Estelle Carteret and Pascal Perney for greenhouse production (Experimental Unit of Glasshouses, INRA-Dijon F). This work was supported by the INRA-ATS (Action Transversale Structurante) programme on *Medicago truncatula*.

References

Allen SE (1974) *Chemical Analysis of Ecological Materials*. New York: John Wiley and Sons.

Bastianelli D, Grosjean F, Peyronnet C, Duparque M and Régnier JM (1998) Feeding value of pea (*Pisum sativum* L.). 1. Chemical composition of different categories of pea. *Animal Science* 67: 609–619.

Bena G, Prosperi JM, Lejeune B and Olivieri I (1998) Evolution of annual species of the genus *Medicago*: a molecular phylogenetic approach. *Molecular Phylogenetics and Evolution* 9: 552–559.

Bertozo MR and Valls JFM (2001) Seed storage protein electrophoresis in *Arachis pintoi* and *A. repens* (Leguminosae) for evaluating genetic diversity. *Genetic Ressources and Crop Evolution* 48: 121–130.

Bewley JD and Black M (1994) *Seeds: Physiology of Development and Germination*, 2nd edn. New York: Plenum Press.

Blum H, Beier H and Gross HJ (1987) Improved silver staining of plant proteins, RNA and DNA in polyacrylamide gels. *Electrophoresis* 8: 93–99.

Bradford MM (1976) A rapid and sensitive method for quantitation of microgram quantities of protein utilising the principle of protein-dye binding. *Analytical Biochemistry* 72: 248–254.

Brown WM, George M and Wilson AC (1979) Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Science USA* 76: 1967–1971.

Croy RRD and Gatehouse JA (1985) Genetic engineering of seed proteins: current and potential application. In: Dodds JH (ed.) *Plant Genetic Engineering*. Cambridge: Cambridge University Press, pp. 143–268.

Duc G (2004) Seed composition of *Medicago truncatula* (line J5), compared to other seed legumes. *Proceedings of the Fifth European Conference on Grain Legumes with the Second International Conference on Legume Genomics and Genetics*, AEP-ICLGG 2004, 7–11 June, Dijon, France, p. 404.

Gallardo K, Job C, Groot SPC, Puype M, Demol H, Vandekerckhove J and Job D (2002a) Importance of methionine biosynthesis for Arabidopsis seed germination and seedling growth. *Plant Physiology* 126: 238–247.

Gallardo K, Job C, Groot SPC, Puype M, Demol H, Vandekerckhove J and Job D (2002b) Proteomics of Arabidopsis seed germination. A comparative study of wild-type and gibberellin-deficient seeds. *Plant Physiology* 129: 823–837.

Gallardo K, Le Signor C, Vandekerckhove J, Thompson RD and Burstin J (2003) Proteomics of *Medicago truncatula* seed development establishes the time frame of diverse metabolic processes related to reserve accumulation. *Plant Physiology* 133: 1–19.

Gueguen J, Vu AT and Schaeffer F (1984) Large scale purification and characterization of pea globulins. *Journal of the Science of Food and Agriculture* 35: 1024–1033.

Hoffman B, Trinh TH, Leung J, Kondorosi A and Kondorosi E (1997) A new *Medicago truncatula* line with superior in vitro regeneration, transformation, and symbiotic properties isolated through cell culture selection. *Molecular Plant–Microbe Interactions* 10: 307–315.

Krochko JE and Bewley JD (1988) Use of electrophoretic techniques in determining the composition of seed storage proteins in alfafa. *Electrophoresis* 9: 751–763.

Krochko JE and Bewley JD (2000) Seed storage proteins in cultivars and subspecies of alfafa (*Medicago sativa*). *Seed Science Research* 10: 423–434.

Ladizinsky G and Hymowitz T (1979) Seed protein electrophoresis in taxonomic and evolutionary studies. *Theoretical and Applied Genetics* 54: 145–151.

Laemmli UK (1970) Cleavage of structural protein during the assembly of the head of bacteriophage T4. *Nature* 227: 680–685.

Mathesius U, Keijzers G, Natera SH, Weinman JJ, Djordjevic MA and Rolfe BG (2001) Establishment of a root proteome reference map for the model legume *Medicago truncatula* using the expressed sequence tag database for peptide mass fingerprinting. *Proteomics* 1: 1424–1440.

Mishra SK, Sharma B, Dasgupta SK and Mathur N (2000) Variations in electrophoretic patterns of storage seed proteins in pea (*Pisum sativum*). *Annals of Agricultural Research* 21: 405–409.

Payne PI (1983) Breeding for protein quantity and protein quality in seed crops. In: Daussant J, Mossé J and Vaughan J (eds) *Seed Proteins, Annual Proceedings of the Phytochemical Society of Europe No. 20*. London: Academic Press, pp. 223–253.

Perrot C (1995) Les protéines de pois: de leur fonction dans la graine à leur utilisation en alimentation animale. *INRA Productions Animales* 8: 151–164.

Potokina E and Eggi E (1997) Intraspecific diversity of *Vicia sativa* L. and *Vicia angustifolia* Reichard deduced from seed protein electrophoresis. *FABIS Newsletter* 40: 13–17.

Sagan M, De Larembergue H, Morandi D, *et al.* (1998) Genetic analysis of symbiosis mutants in *Medicago trucatula*. In: Emerich C (ed.) *Biological Nitrogen Fixation for the 21st Century*. Dordrecht: Kluwer Academic Publishers, pp. 317–318.

SAS Institute Inc. (1990) *SAS/STAT User's Guide*, Version 6, 4th edn. Cary, NC: SAS Institute, pp. 893–997.

Schroeder HE (1982) Quantitative studies on the cotyledonary proteins in the genus Pisum. *Journal of the Science of Food and Agriculture* 33: 623–633.

https://doi.org/10.1079/PGR200562 Published online by Cambridge University Press