Theor Appl Genet (2006) 112: 977–983
DOI 10.1007/s00122-005-0202-1

ORIGINAL PAPER

S.R. Ellwood · N.K. D' Souza · L.G. Kamphuis
T.I. Burgess · R.M. Nair · R.P. Oliver

SSR analysis of the *Medicago truncatula* SARDI core collection reveals substantial diversity and unusual genotype dispersal throughout the Mediterranean basin

Received: 4 July 2005 / Accepted: 14 December 2005 / Published online: 10 January 2006

Abstract The world's oldest and largest *Medicago truncatula* collection is housed at the South Australian Research and Development Institute (SARDI). We used six simple sequence repeat (SSR) loci to analyse the genetic diversity and relationships between randomly selected individuals from 192 accessions in the core collection. *M. truncatula* is composed of three subspecies (ssp.): ssp. *truncatula*, ssp. *longeaculeata*, and ssp. *tricycla*. Analysis at the level of six SSR loci supports the concept of ssp. *tricycla*, all the samples of which showed unique alleles at two loci. Contingency Chi-squared tests were significant between ssp. *tricycla* and ssp. *truncatula* at four loci, suggesting a barrier to gene flow between these subspecies. In accessions defined as ssp. *longeaculeata*, no unique allelic distribution or diagnostic sizes were observed, suggesting this apparent ssp. is a morphological variant of ssp. *truncatula*. The data also suggest *M. truncatula* that exhibits unusually wide genotype dispersal throughout its native Mediterranean region, possibly due to animal and trade-related movements. Our results showed the collection to be highly diverse, exhibiting an average of 25 SSR alleles per locus, with over 90% of individuals showing discrete genotypes. The rich diversity of the SARDI collection provides an invaluable resource for studying natural allelic variation of *M. truncatula*. To efficiently exploit the variation in the SARDI collection, we have defined a subset of accessions ($n = 61$) that maximises the diversity.

Communicated by F. J. Muehlbauer

S.R. Ellwood (✉) · N.K. D' Souza · L.G. Kamphuis · R.P. Oliver
Australian Centre for Nectrotropic Fungal Pathogens,
Health Sciences, Murdoch University,
6150 Perth, Western Australia
E-mail: sellwood@murdoch.edu.au
Tel.: +61-8-93606323
Fax: +61-8-93606303

T.I. Burgess
School of Biological Sciences, Murdoch University,
6150 Perth, Western Australia

R.M. Nair
South Australian Research and Development Institute,
GPO Box 397, 5001 Adelaide, South Australia

Introduction

Medicago truncatula is an annual diploid species widely regarded as the pre-eminent model species for legume genetics research (Thoquet et al. 2002; Young et al. 2003; Choi et al. 2004; Yan et al. 2004). Legumes belong to a large and diverse family of plants composed of over 650 genera and 18,000 species that are second only to the grasses in agricultural importance. They notably form symbioses with *Rhizobium* bacteria to fix nitrogen, and with mycorrhizal fungi for nutrient acquisition. *Medicago* lies within the 'Galegoid' group of temperate legumes, which includes the genera *Melilotus*, *Trifolium*, *Pisum*, *Vicia*, *Sesbania*, and *Lotus*, and within the Trifolieae, subfamily Papilionoideae (Doyle 2001).

M. truncatula is native to the Mediterranean but has become widely naturalised in other regions of the world following the European migration. In Australia, annual *Medicago* species are used as a forage crop and soil improver, often in cereal–pasture ley-farming systems, in some 50 million ha (Crawford et al. 1989). *M. truncatula* is one of the most widely grown of these species (Pearson et al. 1997). SARDI curates the largest worldwide collection of *Medicago* species, collected since the 1960s. Brown (1989a, b) used several statistical models to suggest that at least 70% of the variation in an entire collection could be represented in a core composed of at least 10% of accessions. Skinner et al. (1999) developed the core collection with representatives from the entire Mediterranean basin by measuring diversity as mean Euclidean distance from within groups of accessions defined by subspecies and geographic origin. The study utilized 5,120 *M. truncatula* accessions. The variation in the core was compared against ranges for 27 agronomic

characteristics such as days to first pod, patterns of anthocyanin leaf markings, direction of pod coil, stature, and yield. In the collection, *M. truncatula* has been further subdivided into three subspecies on the basis of pod characteristics (Heyn 1963); ssp. *truncatula* (5–8 coils, pods longer than wide, spines more or less appressed to the pod surface), ssp. *longeaculeata* (5 or more coils, length of pod equal to or less than its diameter, with spines curved, not appressed to pod surface), and ssp. *tricycla* (2.5–4 coils, length of pod less than its diameter, spines usually not appressed to the pod surface).

In this investigation, we have used simple sequence repeats (SSRs) or microsatellite markers to examine the level of genetic diversity and patterns of relatedness within the core collection. SSRs are tandem repeats of nucleotide units, commonly di-nucleotide or tri-nucleotide repeats, but also compound and imperfect or interrupted motifs. SSRs are amenable to rapid genotyping, and new approaches using anchored primers with compound SSRs have considerably reduced the cost and effort of development (Hayden et al. 2004). SSRs are codominant, revealing both heterozygotes and homozygotes, while individual SSR loci are also hypervariable, and therefore comparatively small numbers are capable of resolving relatedness (Gustafsson and Lonn 2003; Hammerli and Reusch 2003). Carefully designed SSRs, where the nucleotide repeats are flanked by conserved coding regions, can be applied to different species within a genus (Peakall et al. 1998; Eujayl et al. 2004), and a few reports suggest such markers may function across plant families, notably among the grasses (Gupta et al. 2003; Thiel et al. 2003; Saha et al. 2004; Yu et al. 2004).

The SARDI core collection was constructed without knowledge of the genetic diversity of the accessions. To enable the scientific community to efficiently utilise natural variation for both molecular studies and in phenotyping, we have used the MSTRAT algorithm (Gouesnard et al. 2001), to produce a subset of accessions maximised for allelic richness.

Materials and methods

Plant sampling and DNA isolation

M. truncatula seeds were obtained from the Genetic Resource Centre, SARDI (South Australian Research and Development Institute, Adelaide, South Australia). In total, 192 core collection accessions were sampled (Table 1). The majority of individuals were *M. truncatula* ssp. *truncatula*, while 15 accessions of ssp. *tricycla* and 12 of ssp. *longeaculeata* were included (Heyn 1963). A single three-week-old individual per accession was randomly selected. One hundred milligrams of leaf tissue was frozen in liquid nitrogen and ground into fine powder in a 1.5 ml Eppendorf tube using a mini pestle. The samples were homogenised in 700 ml of extraction buffer (100 mM Tris pH8, 50 mM EDTA pH 8, 500 mM NaCl, 7 mM β-mercaptoethanol, and 2% SDS), incubated at 65°C for 3 min, then combined with 500 mM potassium acetate and held at 4°C for 15 min. Plant debris was spun down at 10,000*g* for 5 min, the homogenate transferred to fresh tubes, and the DNA precipitated with 0.7 vol. isopropanol. Samples were resuspended in 500 µl TE, incubated with 5 µl RNAase (20 mg/ml, 20 min at 37°C), and extracted with an equal volume of phenol/chloroform. DNA in the aqueous phase was precipitated with 1/10 vol. 3 M sodium acetate (pH 5.2) and 0.7 vol. isopropanol, then dissolved in 100 µl TE. DNA concentration was measured on a Perkin Elmer Lambda 25 spectrophotometer.

Microsatellite analysis

We selected six of the most polymorphic microsatellite loci (Table 2) described by Baquerizo-Audiot et al. (2001). Polymerase chain reaction conditions were as described except that the annealing temperatures for each primer pair were optimised on an Eppendorf Mastercycler gradient thermal cycler. Most favourable temperatures were found to be 50°C for MTR52, 51°C for MTSA5 and MTSA6, and 54°C for MTR58, MAA660456, and MAA660538. PCR reactions were assembled with a Beckman Biomek 200 workstation and run on an Applied Biosystems (ABI) GeneAmp 2700. Each forward primer was fluorescently labelled with FAM, HEX, or TET, allowing PCR products to be triplexed during electrophoresis. Samples were combined with ABI GeneScan 500 TAMRA internal size standard and run on a 5% PAGE PLUS denaturing polyacrylamide gel with an ABI Prism 377 DNA sequencer. Allelic sizes were determined with ABI GeneScan software v. 3.1.2. PCR fidelity was verified by an independent reamplification of heterozygous loci.

Statistical resampling of microsatellite loci and linkage disequilibrium

To determine whether the number of microsatellite loci were sufficient to distinguish each individual, the loci were resampled using a plot of genotypic diversity versus the number of loci in Multilocus v. 2.1.1 (Agapow and Burt 2000). Linkage disequilibrium between loci was tested using the programme GENEPOP (Raymond and Rousset 1995).

Maximising genetic diversity in the SARDI collection

MSTRAT software (Gouesnard et al. 2001) was used to maximise the number of observed alleles at loci in a subset of accessions. The programme uses the M (maximisation) strategy (Schoen and Brown 1993), and an iterative procedure to remove accessions that overlap in

Table 1 A total of 192 *Medicago truncatula* SARDI accessions from which microsatellite genotypes were generated, grouped by geographic region and country of collection

Geographic region	Country	Accession
North Africa	Algeria	Sa 1526, 8601, 8604, 8916, 9137, 9356, 9357, 9642, 9649, 9670, 9851, 9856, 9866, 9876, 14829, 15268, 17498, 21892, 21932
	Libya	Sa 8454, 9048, 9049, 9138, 9820, 14688, 21302, 21362, 21560, 21590, 22922
	Morocco	Sa 8618, 8619, 8623, 8625, 8626, 8638, 8642, 9675, 9888, 10406, 24576, 25654, 25664, 25669, 27774, 27882, 29620, 29625, 29867, 31438, 31442, 31443
	Tunisia	Borung, Sa 3713, 4087, 4586, 7749, 8496, 8935, 9140, 9141, 9434, 9456, 9596, 9693, 9700, 9707, 9710, 9712, 9715, 9720, 9728, 9944, 9970, 10481, 10962, 10964, 17645, 18242, 18395, 18532, 18543, 19957, 23859
Balkans	Bulgaria	Sa 4480, 11954
	Greece	Sa 2084, 3054, 3916, 17562, 27062, 27063, 27137, 27138, 27176
	Romania	Sa 30740
	Yugoslavia	Sa 2162, 4327
Iberian Peninsula	Portugal	Sa 2806, 3047, 3648, 6088, 28375
	Spain	Sa 2826, 9121, 28339
Islands	Canary Islands	Sa 12564
	Corsica	Sa 28889, 28890
	Crete	Sa 2109, 11734
	Cyprus	Cyprus, Sa 2840, 2841, 19995, 19998, 21819, 28064, 28089, 28097, 28099, 28110
	Malta	Sa 1335, 4330
	Madeira	Sa 15951
	Sardinia	Sa 12451, 12455, 25898, 25915, 25926, 25941
	Sicily	Sa 2193, 2203, 2204, 24714, 24968, 25226
Near East	Israel	Sephi, Sa 1489, 1502, 1516, 2715, 2729, 2748, 3116, 3749, 3919, 17590
	Jordan	Sa 4303, 14161, 14163
	Iran	Sa 28645, 28712
	Syria	Sa 22322, 22323
	Turkey	Sa 2820, 9119, 19964
Europe	France	Sa 3562, 4818
	Germany	Sa 1326, 3537, 3569
	Hungary	Sa 3536, 3573
	Italy	Sa 2168, 2218, 27185, 27192
Miscellaneous		Caliph, Mogul, Sa 395, 396, 774, 775, 1306, 1316, 1499, 3308, 4456, 4947, 8105, 8871, 9062, 9295, 11753, 17777, 18346, 18935, 30199, 30203, 31376

'Miscellaneous' comprises lines developed for agriculture, accessions not known to be native to their country of collection, or accessions of unknown origin

their contribution to a core collection. The gain in diversity gained by this method was assessed by graphing a second curve of the richness against a randomly chosen sample of increasing size, and putative core collections ranked by a Nei's diversity index (Nei 1987) as a second criterion of maximisation.

Table 2 Allelic diversity of microsatellite loci among SARDI *Medicago truncatula* ssp. *truncatula* accessions ($n = 177$) and *M. truncatula* ssp. *tricycla* accessions ($n = 15$)

Locus	Repeat motif[a]	Chromosome[b]	Position (cM)[b]	Original publication[a]			This study			
				M. truncatula Gaertner			*M. truncatula* ssp. *truncatula*		*M. truncatula* ssp. *tricycla*	
				n	N_A	Size range (bp)	N_A	Size range (bp)	N_A	Size range (bp)
MTSA5	$(TC)_9$	7	63.9	363	7	215–251[c]	32	212–274	10	216–255
MTSA6	$(TC)_{16}$	5	103.1	374	7	136–164[c]	18	132–168	5	134–150
MTR52	$(AG)_{11}$	5	0	200	4	69	16	50–93	3	50–60
MAA660538	$(AG)_8X_{20}(AG)_{13}X_{22}(AG)_{10}$	–	–	365	6	132–224[c]	42	194–265	4	129–152
MTR58	$(TTG)_5TT(AG)_{12}$	1	47.2	200	4	167	21	147–196	9	167–194
MAA660456	$(TTC)_8$	4	0	357	6	90–132[2]	20	90–141	3	124–130

n number of individuals sampled, N_A number of alleles
[a]Adapted from Baquerizo-Audiot et al. (2001)
[b]Adapted from Kamphuis et al., unpublished data
[c]Adapted from Bonnin et al. (2001)

Genetic distance

Genetic distance between accessions based on microsatellite data was calculated in two ways, $\delta\mu^2$ and D_{AD}. $\delta\mu^2$ is the square of the mean distance between two populations and was developed specifically for the microsatellite data, assuming single-step mutations (Goldstein et al. 1995). Like $\delta\mu^2$, D_{AD} is based on absolute distance, but with this statistic, the distances are not squared. Data were represented as total nucleotide length of the locus. The distance between accessions was calculated using the program MICROSAT (Minch et al. 1995) and neighbour-joining trees were constructed in MEGA 2.1 (Kumar et al. 2001).

Genetic differentiation of subspecies

The accessions in this study do not represent randomly sampled natural populations, but rather admixtures, and therefore parameters commonly used in population genetics, such as differentiation between each subspecies at each locus (F_{ST}), and indeed population analyses in themselves, cannot be applied. Hence, contingency Chi-squared tests were used, since this approach does not make inferences on the mating system. For each accession, a data matrix of multistate characters was compiled by assigning a different letter to each allele at each of the six polymorphic loci (e.g. ABDCGD). The frequency of each allele at each locus for complete and clone corrected populations was calculated, and allele diversity determined, using the program POPGENE (Yeh et al. 1999) and the equation $H = 1 - \sum x_k^2$, where x_k is the frequency of the kth allele (Nei 1973). Chi-squared tests for differences in allele frequencies were calculated for each locus across clone corrected populations (Workman and Niswander 1970).

Results

Allele diversity and heterozygosity

The six SSR loci were successfully amplified in all samples, with no null alleles. The SARDI core collection is highly polymorphic, with a total of 149 SSR alleles detected among the 192 SARDI *M. truncatula* Gaertner accessions (mean = 24.83 per locus). One hundred and seventy-five accessions showed unique genotypes. To test whether the six loci were sufficient to distinguish 192 individuals, statistical resampling showed just two loci was the minimum number required to describe the diversity. Allele size ranges and numbers per loci are given in Table 2. Five of the loci used in this study have been mapped and are unlinked (Table 2). Linkage equilibrium tests were conducted to determine whether the remaining locus, MAA660538, was uninformative.

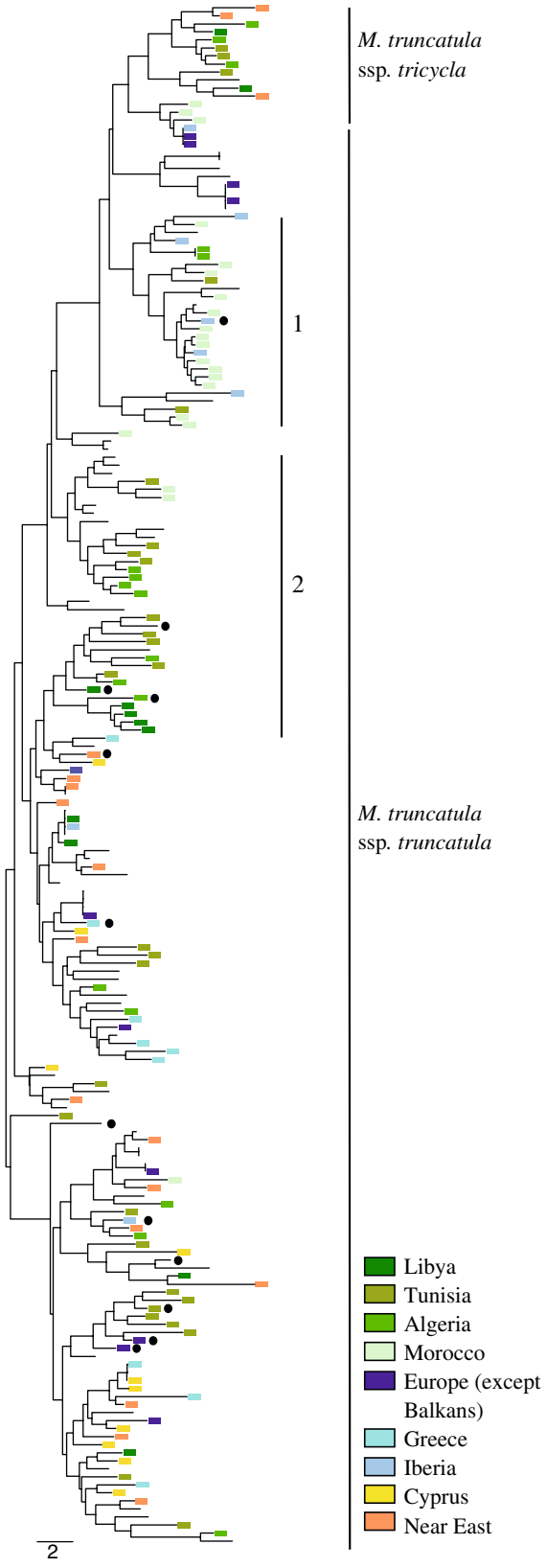

Fig. 1 Neighbour-joining tree based on *Medicago* SSR allele sizes. Individuals assigned as *M. truncatula* ssp. *longeaculeata* are highlighted (*filled circle*). Group 1 represents clades mainly composed of accessions from Iberia and Morocco. Group 2 represents clades mainly composed of accessions from North Africa

Table 3 Gene diversity (*H*) and contingency χ^2 tests for the differences in allele frequencies at six SSR loci across clone-corrected populations of *M. truncatula* ssp. *longeaculeata*, ssp. *tricycla*, and ssp. *truncatula*

Locus	Gene diversity (*H*)			A		B	
	longeaculeata	*tricycla*	*truncatula*	χ^2	*df*	χ^2	*df*
MTSA5	0.88 (9)	0.89 (10)	0.94 (31)	51.5	62	24.8	31
MTSA6	0.72 (5)	0.68 (5)	0.92 (18)	62.2**	34	16.5	17
MTR52	0.74 (4)	0.44 (3)	0.87 (16)	106.3***	30	6.8	15
MAA538	0.84 (8)	0.62 (5)	0.95 (36)	220.0***	80	42.3	35
MTR58	0.86 (8)	0.85 (8)	0.82 (20)	122.3***	40	34.9*	19
MAA456	0.76 (6)	0.61 (3)	0.83 (17)	155.1***	36	7.6	16
n	12	15	165				
Mean	0.80	0.68	0.89				

The number of alleles is given in parentheses and *n* defines the number of individuals in each subspecies
A denotes contingency χ^2 tests across the three sub-species, and *B* denotes contingency χ^2 tests between ssp. *longeaculeata* and ssp. *truncatula*

These were highly significant ($P < 0.01$) for the null hypothesis that the collection is in linkage equilibrium at any pair of loci.

Accessions identified as *M. truncatula* ssp. *truncatula* and ssp. *tricycla* were clearly distinguished by alleles of different size ranges at the locus MAA660538 (194–265 and 90–141 bp, respectively, Table 2). At the locus MAA660456, the size ranges of these two subspecies overlapped (129–152 and 124–130 bp, respectively) but were distinguished by a distinctive 1 bp phase shift in allele size. Accessions assigned to *M. truncatula* ssp. *longeaculeata* showed no clear separation in allele size or composition compared to ssp. *truncatula*. Indeed ssp. *longeaculeata* accessions were randomly distributed in trees generated by both neighbour-joining (Fig. 1) and UPGMA (not shown) analysis in MEGA 2.1. Contingency Chi-squared tests confirm these results (Table 3), with no significant differences in allele frequencies between ssp. *longeaculeata* and ssp. *truncatula*, but significant differences across the three subspecies, which is due to the inclusion of ssp. *tricycla*. Selfing rates across loci, based on the observed heterozygosity, were 97.6–99.3% for ssp. *truncatula*, and 74.5–100% for ssp. *tricycla*. The higher range of values for ssp. *tricycla* is based on a sample size of just 15, which was influenced by a single accession sample heterozygous at five of the loci.

Genotypic clones

The SARDI core collection contained eight groups of clones, or individuals with identical genotypes across the six loci tested (Table 4). Although possessing distinctive characters, these clones are typified by very similar morphologies. A feature of most groups is that members were collected from very different, often physically isolated locations. For example, Group 5 individuals show extreme physical and distance isolation (Israel, Spain and Libya), and Group 8 individuals are notable in their physical isolation on islands (Cyprus, Greece and Corsica).

Maximising genetic diversity in the SARDI collection

The M strategy has been shown to be particularly effective in retaining genetic richness for characters not used in sampling the collection, when accessions come from populations with restricted gene flow or from predominantly selfing species (Bataillon et al. 1996; McKhann et al. 2004). Based on the allelic data from all six loci in ssp. *truncatula* and ssp. *longeaculeata* ($n = 177$), 61 was the

Table 4 Identity and native country of collection for accessions with the same genotype across six SSR loci

Group	Accession	Native country	Morphological traits that differ between members of the same group[a]
1	SA774	–	3
	SA 775	–	
2	SA 1316	–	2, 3
	SA 3537	–	
3	SA 1326	–	1, 3, 5, 6
	SA 1335	Malta	
	SA 3569	–	
	SA 4480	Bulgaria	
4	SA 3573	–	1, 2, 3
	SA 4947	–	
	SA 9062	–	
	SA 30203		
5	SA 3919	Israel[b]	1, 2, 3, 7
	SA 9121	Spain[b]	
	SA 9295	–	
	SA 21590	Libya[b]	
6	SA 14303	Jordan	Nil
	SA 17777	–	
7	SA 14829	Algeria[b]	2
	SA 15268	Algeria[b]	
8	SA 21819	Cyprus[b]	1, 3, 7
	SA 27063	Greece[b]	
	SA 28890	Corsica[b]	

The table includes seven informative morphological traits distinguishing these group members
[a]1, No. flowers per floret; 2, growth habit; 3, internode length; 4, leaf mark colour; 5, leaf mark position, 6, leaf mark shape; 7, pod coil direction
[b]Denotes accessions with defined longitude and latitude collection coordinates

minimum number of accessions required to capture the diversity of the collection.

Geographic distribution of genotypes

Examination of genotypes of ssp. *truncatula* present in discrete regions, including glacial refugia, and islands, revealed few clear relationships between genotype and geographic origin (Fig. 1). Notably, several accessions from Morocco and the Iberian peninsula share a common branch in the neighbour-joining tree (Group 1, Fig. 1), and North Africa possesses several distinct clades and subclusters (Group 2, Fig. 1).

Discussion

The SARDI core collection is highly polymorphic, exhibiting an average 0.126 alleles per locus per individual. The two most polymorphic loci, MTSA5 and MAA660538, show large differences in allele lengths; 62 and 136 nucleotides, respectively. The SARDI core collection is unsuitable for a direct comparison with natural variation because the initial collection was not systematically sampled (Skinner et al. 1999). The only comparable data for *M. truncatula*, based on population sampling from Aude, France, showed an average of 0.018 alleles per locus per individual (Baquerizo-Audiot et al. 2001; Bonnin et al. 2001).

The genotypic data has been used to construct a refined core of accessions using the M method in MSTRAT. McKhann et al. (2004) suggested that, for highly inbreeding species, identifying population genetic boundaries is challenging, as neighbouring locations can be as differentiated as more distant locations. Maximising allelic richness on a set of unlinked markers should, therefore, ensure that the most divergent representatives are selected, even when using a reduced number of markers. In this situation, the M method efficiently captures the diversity throughout the genome, as well as phenotypic variation not linked to loci used to sample a collection. In the SARDI core, the capture of morphological diversity by the M method is complete, in part because of the diversity of this collection producing a relatively large inner core, relative to the number of phenotypic classes (data not shown).

The collection contains almost 8% clones based on the SSR data. This demonstrates a drawback of using physical characteristics to exclude genetic uniformity, given that the selected characteristic(s) may not represent the true diversity within a species. Studies of natural populations under selection for characteristics such as disease resistance have shown high levels of variation over limited distances (1 m^2), with little additional variation over much larger distances (Bevan et al. 1993a, b). Thus, allelic diversity and frequency for a given trait may not necessarily be representative of genome-wide diversity, potentially resulting from simple allelic differences even when samples are gathered from geographically disparate locations.

Although neighbour-joining analysis shows evidence of geographic clustering in North Africa, across the Mediterranean basin there are no clear associations between genotype and origin. This may be explained by the nature of the SARDI core, in that it represents an admixture of accessions selected for diversity rather than a systematically sampled collection. However, it is notable that the Iberian peninsula and Morocco share similar accessions, that Greek accessions are associated with those of North Africa and Cyprus, while unusual genotype dispersal is seen in the origin of some clones which are characterised by their geographic and physical isolation. Sheep are one of the most common animals around the Mediterranean basin, and in Australia, *Medicago* breeders select against spiny pods because they reduce the value of wool (Lunney 1983). A likely explanation, therefore, for the unusual patterns of genotype dispersal is seed transport along well-established shipping routes involving such animals. Clearly these results would indicate the need for carefully designed phylogeographic experiments to determine the underlying processes.

Our analysis agrees with the concept of *M. truncatula* ssp. *tricycla* but not with ssp. *longeaculeata* (Heyn 1963). Neighbour-joining analysis placed ssp. *tricycla* into a discrete group; however, ssp. *longeaculeata* individuals were randomly distributed throughout the tree, suggesting this apparent subspecies is a morphological variant, and further emphasises that a morphological characteristic may not be indicative of genetic variation. Loci MAA660538 and MAA660456 contain unique alleles in ssp. *tricycla* that differ in their size range and composition to ssp. *truncatula*. Significant differences in allele frequency were pronounced at only four loci, suggesting that the regions of the genome related to the phenotype of ssp. *tricycla* are essentially intact but the gene flow occurs between these subspecies. Although a recent divergence between ssp. *truncatula* and ssp. *tricycla* cannot be excluded as an explanation, the skewed allele distribution may also be explained by interspecific hybridisation. *M. littoralis* appears to share similar alleles to ssp. *tricycla* at MAA660538 and MAA660456 (Baquerizo-Audiot et al. 2001). Morphologically, ssp. *tricycla* is intermediate in form between *M. littoralis* and *M. truncatula* ssp. *truncatula*, and gene flow has been reported between these species (Small and Jomphe 1989). There is no published quantitative data on hybrid fertility between ssp. *truncatula* and ssp. *tricycla*, but SARDI researchers report F_1 progeny show poor early vigour and a reduction in fertility, which may provide a barrier to frequent hybridisation. In contrast, ssp. *truncatula* and ssp. *longeaculeata* crosses are compatible.

This is the first major SSR study examining a *M. truncatula* collection, and illustrates the positive and negative aspects of assembling such collections. The subspecies and origin-based methodology used by Skinner et al. (1999) has produced a highly diverse core, but exhibits genotypic redundancy based on the six loci

examined, and, as an admixture of accessions, is limited in the phylogeographic inferences that can be drawn. This study will enable researchers to make informed polymorphic crosses, and by maximising diversity in the collection into a refined core, researchers can more efficiently exploit this diversity.

Acknowledgements This research was funded by the Grains Research and Development Corporation of Australia (GRDC), and the Department of Education, Science and Training (DEST). We would like to thank Steve Hughes, the curator of Australian Medicago Genetic Resources Centre, for providing seed. The research was conducted at the State Agricultural Biotechnology Centre, Murdoch University, Western Australia.

References

Agapow P-M, Burt A (2000) MultiLocus v1.2.2. Imperial College, London

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

Bataillon TM, David JL, Schoen DJ (1996) Neutral genetic markers and conservation genetics: simulated germplasm collections. Genetics 144:409–417

Bevan J, Crute I, Clarke D (1993a) Diversity and variation in expression of resistance to *Erysiphe fischeri* in *Senecio vulgaris*. Plant Pathol 42:647–653

Bevan J, Crute I, Clarke D (1993b) Resistance to *Erysiphe fischeri* in two populations of *Senecio vulgaris*. Plant Pathol 42:636–646

Bonnin I, Ronfort J, Wozniak F, Olivieri I (2001) Spatial effects and rare outcrossing events in *Medicago truncatula* (Fabaceae). Mol Ecol 10:1371–1383

Brown AHD (1989a) The case for core collections. In: Brown AHD, Frankel OH, Marshall DR, Williams JT (eds) The use of plant genetic resources. Cambridge University Press, Cambridge, pp 136–156

Brown AHD (1989b) Core collections: a practical approach to genetic resources management. Genome 31:818–824

Choi H-K, Kim D, Uhm T, Limpens E, Lim H, Mun J-H, Kalo P, Penmetsa RV, Seres A, Kulikova O, Roe BA, Bisseling T, Kiss GB, Cook DR (2004) A sequence-based genetic map of *Medicago truncatula* and comparison of marker colinearity with *M. sativa*. Genetics 166:1463–1502

Crawford EJ, Lake AWH, Boyce KG (1989) Breeding annual *Medicago* species for semi-arid conditions in southern Australia. Adv Agron 42:399–437

Doyle J (2001) Leguminosae. In: Brenner SMJ (ed) Encyclopaedia of genetics. Academic, San Diego, pp 1081–1085

Eujayl I, Sledge M, Wang L, May G, Chekhovskiy K, Zwonitzer J, Mian M (2004) *Medicago truncatula* EST-SSRs reveal cross-species genetic markers for *Medicago* spp. Theor Appl Genet 108:414–422

Goldstein D, Linares A, Cavalli-Sforza L, Feldman M (1995) Genetic absolute dating based on microsatellites and the origin of modern humans. Proc Natl Acad Sci USA 92:6723–6727

Gouesnard B, Bataillon TM, Decoux G, Rozale C, Schoen DJ, David JL (2001) MSTRAT: an algorithm for building germ plasm core collections by maximizing allelic or phenotypic richness. J Hered 92:93–94

Gupta P, Rustgi S, Sharma S, Singh R, Kumar N, Balyan H (2003) Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. Mol Genet Genomics 270:315–323

Gustafsson S, Lonn M (2003) Genetic differentiation and habitat preference of flowering-time variants within *Gymnadenia conopsea*. Heredity 91:284–292

Hammerli A, Reusch T (2003) Genetic neighbourhood of clone structures in eelgrass meadows quantified by spatial autocorrelation of microsatellite markers. Heredity 91:448–455

Hayden M, Stephenson P, Logojan A, Khatkar D, Rogers C, Koebner R, Snape J, Sharp P (2004) A new approach to extending the wheat marker pool by anchored PCR amplification of compound SSRs. Theor Appl Genet 108:733–742

Heyn C (1963) The annual species of *Medicago*. Scripta Hierosolymitana 12:1–154

Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: molecular evolutionary genetics analysis software. Bioinformatics 17:1244–1245

Lunney HWM (1983) Vegetable fault in Australian wool: Classification, consequences, and economic loss. J Aust Inst Agric Sci 49:207–211

McKhann HI, Camilleri C, Berard A, Bataillon T, David JL, Reboud X, Le Corre V, Caloustian C, Gut IG, Brunel D (2004) Nested core collections maximizing genetic diversity in *Arabidopsis thaliana*. Plant J 38:193–202

Minch E, Ruiz-Linares A, Goldstein D (1995) Microsat 1.4, a computer program for the calculation of genetic distance from microsatellite data. Stanford University Medical Centre, Stanford

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

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

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

Pearson CJ, Brown R, Collins WJ, Archer KA, Wood MS, Petersen C, Bootle B (1997) An Australian temperate pastures database. Aust J Agric Res 48:453–466

Raymond M, Rousset F (1995) Genepop version 1.2: population genetics software for exact tests and ecumenicism. J Hered 86:248–249

Saha MC, Mian MAR, Eujayl I, Zwonitzer JC, Wang L, May GD (2004) Tall fescue EST-SSR markers with transferability across several grass species. Theor Appl Genet 109:783–791

Schoen D, Brown A (1993) Conservation of allelic richness in wild crop relatives is aided by assessment of genetic markers. Proc Natl Acad Sci USA 90:10623–10627

Skinner D, Bauchan G, Auricht G, Hughes S (1999) A method for the efficient management and utilization of large germplasm collections. Crop Sci 39:1237–1242

Small E, Jomphe M (1989) A synopsis of the genus *Medicago* (Leguminosae). Can J Bot 67:3260–3294

Thiel T, Michalek W, Varshney R, Graner A (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theor Appl Genet 106:411–422

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

Workman P, Niswander J (1970) Population studies on southwestern Indian tribes. II. Local genetic differentiation in the Papago. Am J Hum Genet 22:24–49

Yan H, Mudge J, Kim D, Shoemaker R, Cook D, Young N (2004) Comparative physical mapping reveals features of microsynteny between *Glycine max*, *Medicago truncatula*, and *Arabidopsis thaliana*. Genome 47:141–55

Yeh F, Yang R, Boyle T (1999) POPGENE version 1.31. Microsoft Windows-based software for population genetics analysis. University of Alberta and Centre for International Forestry Research, Alberta

Young N, Mudge J, Ellis T (2003) Legume genomes: more than peas in a pod. Curr Opin Plant Biol 6:199–204

Yu J-K, La Rota M, Kantety RV, Sorrells ME (2004) EST derived SSR markers for comparative mapping in wheat and rice. Mol Genet Genomics 271:742–751