Breeding Science 59: 77–86 (2009)

Development of new microsatellite markers and their application in the analysis of genetic diversity in lentils

Aladdin Hamwieh[1)], Sripada M. Udupa[2)], Ashutosh Sarker[1)], Christian Jung[3)] and Michael Baum*[1)]

[1)] *International Center for Agricultural Research in the Dry Areas (ICARDA)*, P.O. Box 5466, Aleppo, Syria
[2)] *ICARDA-INRA Cooperative Research Project, International Center for Agricultural Research in the Dry Areas (ICARDA)*, B.P. 6299, Rabat, Morocco
[3)] *Plant Breeding Institute, Christian-Albrechts University Kiel*, Olshausen Str. 40, D-24098, Germany

This paper reports the development of new microsatellite markers for lentil (*Lens culinaris* subsp. *culinaris*) and their use for genetic diversity analysis of a lentil core collection developed at ICARDA (Aleppo-Syria). Fourteen new markers were developed from microsatellite flanking sequences of a genomic library from a cultivated lentil accession ILL5588. The core collection used comprises 109 accessions from 15 countries representing 57 cultigens (including 18 breeding lines) from 8 countries and 52 wild types of germplasm (*L. culinaris* subsp. *orientalis*, *L. culinaris* subsp. *tomentosus* and *L. culinaris* subsp. *odemensis*) from 11 countries. Total number of alleles detected across all microsatellite loci was 182, with a mean of 13 alleles per locus. The wild accessions were rich in alleles (151 alleles) compared to cultigens (114 alleles). The genetic diversity index for the microsatellite loci in the wild accessions ranged from 0.16 (for locus SSR28 in *L. culinaris* subsp. *odemensis*) to 0.93 (for locus SSR66 in *L. culinaris* subsp. *orientalis*) with a mean of 0.66, while in the cultigens genetic diversity varied between 0.03 (locus SSR28) and 0.87 (locus SSR207) with a mean of 0.65. The cluster analysis indicated two major clusters, mainly one with the cultigens and the other with the wild accessions.

Key Words: lentil, *Lens*, microsatellite, biodiversity.

Introduction

Lentil (*Lens culinaris* Medik. subsp. *culinaris*) is an annual self-pollinated diploid ($2x = 2n = 14$ chromosomes) species, and a highly valued food legume grown extensively in over 35 countries situated in South Asia, Middle East, North Africa, North America and Australia. Worldwide lentil production in 2007 was 3.78 million ton (Mt) from an area of 3.8 million ha., with an average productivity 1019.1 kg/ha (FAOSTAT, 2008). Currently, India leads the world in lentil production and consumption, and Canada and Turkey are the world's largest lentil exporters. Lentil seed has a broad range of uses for human consumption (in soups, stews, salads and vegetarian dishes) and its plant parts, excluding seeds, are used as animal feed around the world, and the different seed and plant types adapted to an increasingly wide range of ecologies.

Lentil originated in the Fertile Crescent of the Near East and later spread to North Africa and eastern Ethiopia, central and southern Europe, North America, Oceania and South Asia (Duke 1981). It is the oldest of the domesticated grain legumes (Bahl *et al.* 1993). In addition to cultivated lentil, the genus *Lens* consists of four wild species: *L. culinaris*, *L. ervoides*, *L. lamottei* and *L. nigricans* (van Oss *et al.* 1997). The species *Lens culinaris* Medikus has three wild subspecies: *L. culinaris* subsp. *orientalis*, *L. culinaris* subsp. *odemensis* and *L. culinaris* subsp. *tomentosus* (Ferguson *et al.* 2000); among these, *L. culinaris* subsp. *orientalis* is considered to be the progenitor of the cultivated lentil. All these species and subspecies are mostly distributed in the Mediterranean region (Ladizinsky *et al.* 1984). The wild gene pool constitutes a large reservoir of genetic diversity, and contains several potential traits for genetic improvement of lentils (Sarker and Erskine 2006).

Genetic diversity and the relationships among different species of *Lens* are of great importance for lentil's genetic resources conservation and breeding potential. Molecular marker studies have made significant contributions to our understanding of genetic diversity and relatedness in various crops (Struss and Plieske 1998, Udupa *et al*. 1999, Abe *et al.* 2003). The main benefit of using molecular markers over morphological markers is that they are good indicators of genetic distances and diversity among accessions because of their selective neutrality (Winter and Kahl 1995).

Over the years, several studies have assessed genetic diversity and phylogenetic relationships in the genus *Lens* using protein and DNA markers. Isozymes, RFLPs, RAPDs and AFLP have all been used in genetic diversity and

Communicated by Donghe Xu
Received June 30, 2008. Accepted January 22, 2009.
*Corresponding author (e-mail: m.baum@cgiar.org)

phylogenetic analyses in the *Lens* genus. Analysis of seed protein polymorphism showed close relationships between the subspecies *culinaris* and *orientalis* in comparison with the other wild species (Hoffman *et al*. 1986). Restriction fragment length polymorphism (RFLP) markers were used to study diversity and phylogenies among different species (Havey and Muehlbauer 1989). Furthermore, random amplified polymorphic DNA (RAPD) and amplified fragment length polymorphism (AFLP) analyses have provided similar conclusions regarding the phylogeny of *Lens* (Abo-elwafa *et al*. 1995, Sharma *et al*. 1995, Sharma *et al*. 1996). However, AFLP analysis detected a higher level of polymorphism than did the RAPD analysis (Collard *et al.* 2005).

Microsatellite markers have been successfully employed in many crops for genetic diversity studies and are useful for a variety of applications in plant genetics and breeding including genetic diversity studies. This is because of their high degree of polymorphism, multiallelic nature, reproducibility, codominant inheritance, relative abundance and good genome coverage (Powell *et al.* 1996). Recently Hamwieh *et al*. (2005) developed microsatellite markers for lentil and used 30 of them to construct the genetic map. In this study, we developed an additional set of 14 microsatellite markers for lentil and used them to estimate genetic diversity and relationships within a core collection of lentil germplasm held at ICARDA.

Material and Methods

Plant material

A core collection of lentil germplasm held at ICARDA was developed by selecting a small set of accessions representing the diversity of the entire collection, based on passport and morphological data. This collection consists of 57 *L. culinaris* subsp. *culinaris* (cultivated lentil), 30 *L. culinaris* Medik. subsp. *orientalis*, 18 *L. culinaris* Medik. subsp. *odemensis*, and four *L. culinaris* subsp. *tomentosus* accessions. These accessions originated from 15 countries, and 18 ICARDA breeding lines were used for this study (Table 1).

DNA extraction

Total DNA was extracted according to the CTAB-method, described by Rogers and Bendich (1985) with minor modifications. The fresh leaf material was collected from a single seedling of each accession.

Isolation of microsatellites

A genomic library of cultivated lentil (ILL 5588) was constructed and approximately 200,000 colonies were screened for six synthetic repeats, namely $(GT)_{10}$, $(GA)_{10}$, $(GC)_{10}$, $(GAA)_8$, $(TA)_{10}$, and $(TAA)_5$ as described previously by Hamwieh *et al*. (2005). The microsatellite motifs were classified as 'simple' and 'compound' as well as 'perfect' and 'imperfect' repeats as described previously (Weber 1990, Hüttel *et al.* 1999, Sibly *et al.* 2001). A microsatellite is referred as 'simple', if a single type of repeat unit repeats several times (e.g., $(CA)_n$; $(TAA)_k$; etc.). On the other hand, a 'compound' microsatellite consists of stretches of more than one type of repeat unit (e.g., $(GA)_n{\cdot}(TA)_k$; $(GT)_k{\cdot}(TAA)_l{\cdot}(TA)_m$, etc.). A 'perfect' microsatellite means that the microsatellite sequence does not contain mutations or interruptions (e.g., $(CA)_n$; $(TAA)_k$; $(CT)_m{\cdot}(GAA)_n$, etc.). An 'imperfect' microsatellite means that the microsatellite sequence does contain mutations or interruptions (e.g. $(CA)_nCC(CA)_m$; $(TA)_kAA(TA)_l{\cdot}(GA)_m$, etc.). In the above examples, subscripts *k*, *l*, *n* and *m* denote number of times the particular microsatellite motif repeats.

Primer design and PCR amplification of microsatellites

Primer pairs were designed close to the microsatellite repeats in the flanking regions as described by Hamwieh *et al*. (2005). PCR amplification reactions of microsatellites were performed in a final volume of 20 µl in the presence of 10 ng of template DNA, 10 pmol of each primer, 0.2 mM of each deoxynucleotide, 1 × PCR buffer (Invitrogen, Carlsbad, CA), and 1 unit *Taq* polymerase (Invitrogen). The forward primer was 5′-labeled with one of three fluorophores (6FAM, NED, or TET). PCR was carried out using a PE 9600 thermocycler (Perkin-Elmer, Foster City, CA). After 3 min at 94°C, 30 cycles were performed with 30 s at 94°C, 30 s at either 52°C, 53°C, 54°C or 55°C (depending on the locus, Table 2), and 1 min at 72°C, followed by final extension step of 5 min at 72°C. Amplified products were detected on a MegaBACE 500 Capillary System (Amersham Pharmacia Biotech, Piscataway, NJ). Samples were prepared by adding 1 µl of diluted PCR products to 9 µl formamide and samples included 1% (v/v) ET-Rox 900 bp DNA size standard (Amersham Bioscience). Microsatellite fragment sizes were estimated using the MegaBACE Genetic Profiler Version 2.0 (Amersham Pharmacia Biotech).

Data analysis

Analyses of microsatellite diversity were conducted at the locus level in a given species. For each locus, we estimated the number of alleles, range of fragment sizes, mean of fragment size and genetic diversity (Nei 1973) using PowerMarker V3.25 (Liu and Muse 2005). Power markers allowed to group the data at several categories or levels. The genetic relationships were further estimated at the levels of accession, subspecies, and countries. The unweighted pair-group method arithmetic average (UPGMA) was used to cluster 109 accessions and neighbor-joining (NJ) methods to cluster the groups at subspecies and countries level. The cluster analysis and bootstrap analysis (with 100 bootstrap samples) were performed using PAST software version 1.62 (Hammer *et al.* 2001). The Principal Coordinate Analysis (PCA) was performed according to Davis (1986) using the PAST software.

Table 1. A core collection of lentil germplasm held at ICARDA

Accession No.	Origin	*L. culinaris* subsp.
ILL5888	Bangladesh	*culinaris*
ILL8006	Bangladesh	*culinaris*
ILL8007	Bangladesh	*culinaris*
ILL10005	ICARDA	*culinaris*
ILL10011	ICARDA	*culinaris*
ILL10012	ICARDA	*culinaris*
ILL8076	ICARDA	*culinaris*
ILL8174	ICARDA	*culinaris*
ILL88527	ICARDA	*culinaris*
ILL9830	ICARDA	*culinaris*
ILL9832	ICARDA	*culinaris*
ILL9837	ICARDA	*culinaris*
ILL9896	ICARDA	*culinaris*
ILL9899	ICARDA	*culinaris*
ILL9907	ICARDA	*culinaris*
ILL9918	ICARDA	*culinaris*
ILL9922	ICARDA	*culinaris*
ILL9924	ICARDA	*culinaris*
ILL9936	ICARDA	*culinaris*
ILL9986	ICARDA	*culinaris*
ILL9995	ICARDA	*culinaris*
ILL2501	India	*culinaris*
ILL2565	India	*culinaris*
ILL2580	India	*culinaris*
ILL2684	India	*culinaris*
ILL3596	India	*culinaris*
ILL3614	India	*culinaris*
ILL7556	India	*culinaris*
ILL7558	India	*culinaris*
ILL7715	India	*culinaris*
ILL7716	India	*culinaris*
ILL8008	India	*culinaris*
L-4147	India	*culinaris*
ILL759	Iran	*culinaris*
ILL5244	Jordan	*culinaris*
ILL5883	Jordan	*culinaris*
ILL7162	Pakistan	*culinaris*
ILL7163	Pakistan	*culinaris*
ILL7164	Pakistan	*culinaris*
ILL8112	Pakistan	*culinaris*
ILL8114	Pakistan	*culinaris*
ILL91517	Pakistan	*culinaris*
ILL7723	Pakistan	*culinaris*
ILL4400	Syria	*culinaris*
ILL4401	Syria	*culinaris*
ILL5597	Syria	*culinaris*
ILL780	Syria	*culinaris*
ILL1878	Turkey	*culinaris*
ILL1880	Turkey	*culinaris*
ILL7155	Turkey	*culinaris*
ILL8143	Turkey	*culinaris*
ILL8146	Turkey	*culinaris*
ILL8611	Turkey	*culinaris*
ILL8612	Turkey	*culinaris*
AKM-302	Turkey	*culinaris*
ILL632	Turkey	*culinaris*
ILL323	Serbia & Montenegro	*culinaris*
ILWL21	Palestine	*odemensis*
ILWL153	Syria	*odemensis*
ILWL164	Syria	*odemensis*
ILWL172	Syria	*odemensis*
ILWL173	Syria	*odemensis*
ILWL174	Syria	*odemensis*
ILWL175	Syria	*odemensis*
ILWL202	Turkey	*odemensis*
ILWL235	Syria	*odemensis*
ILWL238	Syria	*odemensis*
ILWL322	Syria	*odemensis*
ILWL356	Syria	*odemensis*
ILWL357	Syria	*odemensis*
ILWL362	Syria	*odemensis*
ILWL363	Syria	*odemensis*
ILWL404	Syria	*odemensis*
ILWL438	Turkey	*odemensis*
ILWL449	Turkey	*odemensis*
ILWL4	Turkey	*orientalis*
ILWL490	Armenia	*orientalis*
ILWL6	Turkey	*orientalis*
ILWL7	Turkey	*orientalis*
ILWL11	Syria	*orientalis*
ILWL71	Iran	*orientalis*
ILWL72	Cyprus	*orientalis*
ILWL74	Palestine	*orientalis*
ILWL81	Turkey	*orientalis*
ILWL118	Syria	*orientalis*
ILWL145	Syria	*orientalis*
ILWL200	Turkey	*orientalis*
ILWL214	Iran	*orientalis*
ILWL215	Iran	*orientalis*
ILWL220	Turkey	*orientalis*
ILWL228	Syria	*orientalis*
ILWL229	Syria	*orientalis*
ILWL230	Syria	*orientalis*
ILWL231	Syria	*orientalis*
ILWL326	Cyprus	*orientalis*
ILWL341	Jordan	*orientalis*
ILWL344	Syria	*orientalis*
ILWL345	Syria	*orientalis*
ILWL346	Syria	*orientalis*
ILWL372	Uzbekistan	*orientalis*
ILWL373	Uzbekistan	*orientalis*
ILWL374	Uzbekistan	*orientalis*
ILWL376	Tajikistan	*orientalis*
ILWL378	Turkmenistan	*orientalis*
ILWL402	Lebanon	*orientalis*
ILWL90	Turkey	*tomentosus*
ILWL91	Turkey	*tomentosus*
ILWL307	Turkey	*tomentosus*
ILWL308	Turkey	*tomentosus*

Table 2. Forward and reverse primer sequences and annealing temperature used for amplification of 14 microsatellite loci

Primers	Forward sequence	Reverse sequence	Temperature (°C)	Expected size (bp)	Microsatellite motif
SSR 28	GAGGGCATAAATTCAGATTC	GGACAACGCACATTTGATG	53	383	$(TG)_{15}$
SSR 34-2	CGGCGGATGAAACTAAAG	CATTTCCTTCACAAACCAAC	53	185	$(GT)_{22}$
SSR 46-2	CACTAAACATGGAAAATAGG	CTTATCTTTGTTTGAAGCAA	50	157	$(CA)_8$
SSR 66	GGTAGTGGTGAGGAATGAC	GCATCACTGCAACAGACC	55	253	$(TG)_{10}(AG)_{18}$
SSR 72	CAAACAGTACAAGGAAAGGAG	CTGACTGAGCTGCTTGAAC	55	253	$(AAG)_9$
SSR 90	CCGTGTACACCCCTAC	CGTCTTAAAGAGAGTGACAC	55	181	$(TG)_{11}(AG)_{10}$
SSR 132RN	CCAGAACAAACGTAAACC	CTATCGCATATGAGTGAAC	52	330	$(GT)_9$
SSR 183	GCTCGCATTGGTGAAAC	CATATATAGCAGACCGTG	52	119	$(AC)_8$
SSR 191	GCAAATTTCTTGGTCTACAC	GGGCACAGATTCATAAGG	53	238	$(AC)_{17}(AT)_{13}$
SSR 197	CACCAATCACCAACACAC	GAGCTGTGAAGTCTTATCTG	54	173	$(TA)_{13}$
SSR 202	CAACCTCACTTACCTTAC	GCTCTTTATCATCATTCTAC	52	220	$(CA)_8(CCA)_6(CA)_5$
SSR 207	GAGAGATACGTCAGAGTAG	GATTGTGCTTCGGTGGTTC	55	227	$(CA)_{22}$
SSR 230	CCAACAACAATTCACCATAC	AACATTGTACTGAGAGGTG	53	251	$(CA)_{28}(TA)_8$
SSR 253	GAAGAAGCATTCACGGTG	GAGGGACTACTATATCAG	53	139	$(AC)_8$

Results

Isolation of microsatellites

The screening of around 200,000 genomic clones for $(GT)_{10}$, $(GA)_{10}$, $(GC)_{10}$, $(GAA)_8$, $(TA)_{10}$, and $(TAA)_5$ sequences, resulted 371 clones expected to have the microsatellite. These 371 clones were confirmed as 'positive clones' after two rounds of hybridization using $(GT)_{10}$, $(GA)_{10}$, $(GC)_{10}$, $(TA)_{10}$, $(TAA)_5$ and $(GAA)_8$ as probes. Of these, only 243 clones were sequenced, while the other clones were not sequenced either because of small insert or because of suspected duplicates. The clones were sequenced in either one direction (only 23 clones) or both directions (220 clones sequenced). Upon comparing sequences, we observed that 71.2% of the repeat sequences were simple sequence repeats (SSRs). Among the SSR motifs, dinucleotide repeats were more common than other types with GT/CA repeats at a frequency of 33.1%, followed by TA/AT repeats at 10.2%. Among the trinucleotide repeats, TAA/ATT motifs were most predominant at a frequency of 3.4% (Table 3). From this sequence information, more than 170 primer pairs were designed and used for amplification of the corresponding microsatellite using DNA from the donor line ILL 5588. Out of the 170 primer pairs, 58 pairs (34%) amplified monomorphic fragments and 56 pairs (33%) produced PCR products with polymorphism. The remaining pairs produced either smear or nonspecific amplifications.

Out of 56 polymorphic microsatellite primer pairs (loci), 30 have been previously used to construct a linkage map in lentil (Hamwieh *et al.* 2005). For the present study, 26 polymorphic microsatellite loci for which sequences are not yet published were used. Of these, 12 produced many polymorphic amplicons which were not close to the expected size in ILL 5588 and therefore eliminated from this study. The remaining 14 microsatellite primers which yielded amplicons close to the expected sizes were selected to profile the genetic diversity in the core collection. Nine of these loci contained simple and perfect repeats and the other five contained compound microsatellites repeats.

Table 3. Microsatellite motifs observed in the lentil genomic library

Type*		Microsatellite motif	Number	% Occurrence
Simple	Perfect	CA/GT	57	24.2
		CG/GC	2	0.8
		CT/GA	7	3.0
		CTT/GAA	3	1.3
		AT/TA	21	8.9
		ATT/TAA	7	3.0
		Others types	37	15.7
		Total	134	56.8
	Imperfect	CA/GT	21	8.9
		CG/GC	0	0.0
		CT/GA	1	0.4
		CTT/GAA	0	0.0
		AT/TA	3	1.3
		ATT/TAA	1	0.4
		Others types	8	3.4
		Total	34	14.4
Compound	Perfect		38	16.1
	Imperfect		30	12.7
	Total		68	28.8
Total			236	100

* The microsatellite motifs were classified as simple and compound repeats as well as perfect and imperfect repeats as described previously (Weber 1990, Hüttel *et al.* 1999, Sibly *et al.* 2001).

Variation at microsatellite loci in cultivated and wild lentil subspecies

All of the primer pairs tested amplified fragments in wild relatives with the PCR program developed for cultivated subspecies; this indicates the conservation of microsatellite primers in related subspecies. Both inter- and intra-specific differences in amplified fragments were found for all the

Table 4. Variation at microsatellite loci in four sub-species of *Lens culinaris*

Locus	*culinaris*		*orientalis*		*tomentosus*		*odemensis*	
	No. of alleles	Genetic diversity	No. of alleles	Genetic diversity	No. of alleles	Genetic diversity	No. of alleles	Genetic diversity
SSR66	14	0.84	19	0.93	6	0.81	12	0.88
SSR28	2	0.03	5	0.29	2	0.22	3	0.16
SSR34-2	9	0.67	6	0.76	3	0.62	6	0.62
SSR46-2	5	0.68	5	0.71	3	0.53	4	0.54
SSR72	5	0.63	5	0.76	3	0.59	5	0.77
SSR90	11	0.74	12	0.80	3	0.66	6	0.71
SSR132RN	8	0.66	11	0.86	4	0.66	8	0.59
SSR183	6	0.64	8	0.77	3	0.63	8	0.76
SSR191	13	0.80	13	0.88	5	0.78	12	0.83
SSR197	14	0.75	16	0.89	4	0.56	12	0.85
SSR202	4	0.26	2	0.22	2	0.38	4	0.53
SSR207	16	0.87	11	0.68	5	0.75	8	0.64
SSR230	13	0.83	11	0.83	4	0.72	5	0.65
SSR253	8	0.70	10	0.82	2	0.38	7	0.67
Total	128		134		49		100	
Mean	9.14	0.65	9.57	0.73	3.5	0.59	7.14	0.66
SD		0.23		0.21		0.17		0.18
CV		35.25		29.18		28.67		27.41

subspecies. The variation at microsatellite loci in all four subspecies is summarized in Table 4.

The total number of alleles detected across all microsatellite loci was 182, with an overall average of 13 alleles per locus. Approximately 83% (151 alleles) of all alleles were observed in the wild accessions and only 62.6% (114 alleles) in the cultivated lentil. The number of alleles varied within the subspecies as well. The highest number of alleles was detected in *L. culinaris* subsp. *orientalis* (134); followed by *L. culinaris* subsp. *culinaris* (128) and *L. culinaris* subsp. *odemensis* (100). The number was lowest in *L. culinaris* subsp. *tomentosus* (49). With the exception of SSR202 and SSR28, all other lentil microsatellite markers showed relatively a high allelic diversity.

The values of genetic diversity for the microsatellite loci in the wild accessions ranged from 0.16 (for locus SSR28 in *L. culinaris* subsp. *odemensis*) to 0.93 (for the locus SSR66 in *L. culinaris* subsp. *orientalis*) with a mean of 0.66, while in the cultivated lentil genetic diversity varied between 0.03 (for locus SSR28) and 0.87 (for locus SSR207) with a mean of 0.65 (Table 4). When gene diversity was compared subspecies wise, the highest mean gene diversity was also observed in *L. culinaris* subsp. *orientalis* (0.73 ± 0.21), followed by *L. culinaris* subsp. *odemensis* (0.66 ± 0.18) and *L. culinaris* subsp. *culinaris* (0.65 ± 0.23) (Table 4). Again, diversity was lowest in *L. culinaris* subsp. *tomentosus* (0.59 ± 0.17) (Table 4).

Genetic relationships

The genetic relationships among the accessions were further studied using cluster analysis based on UPGMA. The dendrogram discriminated between all 109 lentil genotypes, which were grouped into two major clusters denoted as A and B; these contained 51 and 58 accessions respectively, at a bootstrap value of 100% (Fig. 1). Cluster A could be subdivided into four sub-clusters denoted as A1, A2, A3, and A4 (Fig. 1). These clusters comprised 9, 12, 20, and 10 accessions, respectively. The results did not show clear separation of the accessions between *L. culinaris* subsp. *orientalis*, *L. culinaris* subsp. *odemensis*, and *L. culinaris* subsp. *tomentosus*. Among the 51 accessions, 48 were from wild lentil and three were from cultivated lentil (ILL632$_{\text{“Tur”}}$, ILL780$_{\text{“Syr”}}$, and ILL9918$_{\text{“ICARDA”}}$). Only 11 of the 18 *L. culinaris* subsp. *odemensis* accessions clustered together in A3, while the remaining *L. culinaris* subsp. *odemensis* accessions were dispersed among the other clusters. Similarly, of the four accessions from subsp. *tomentosus*, only ILWL91 and ILWL 208 (from Turkey) were grouped together.

Cluster B contained 58 lentil accessions divided into four sub-clusters denoted as B1, B2, B3, and B4; these contained 12, 16, 11, and 19 accessions respectively. Among the 58 Cluster-B accessions, 54 were from cultivated lentil and four were from wild *L. culinaris* subsp. *orientalis* (Fig. 1). 17 of 18 ICARDA breeding lines were clustered in B1, B3 and B4. These lines were clustered together with accessions from India, Pakistan, Syria, Turkey, Jordan, and Serbia. Of the 22 cultivated accessions from South Asia (India, Pakistan and Bangladesh), 19 (86%) were grouped in B1 and B2.

Principal Co-ordinate Analysis (PCA), together with neighbor joining analysis, was conducted based on the accessions' country/origins (Fig. 2). The PCA indicated a strong population substructure with two well-separated clusters. The first cluster could be sub-divided into two smaller

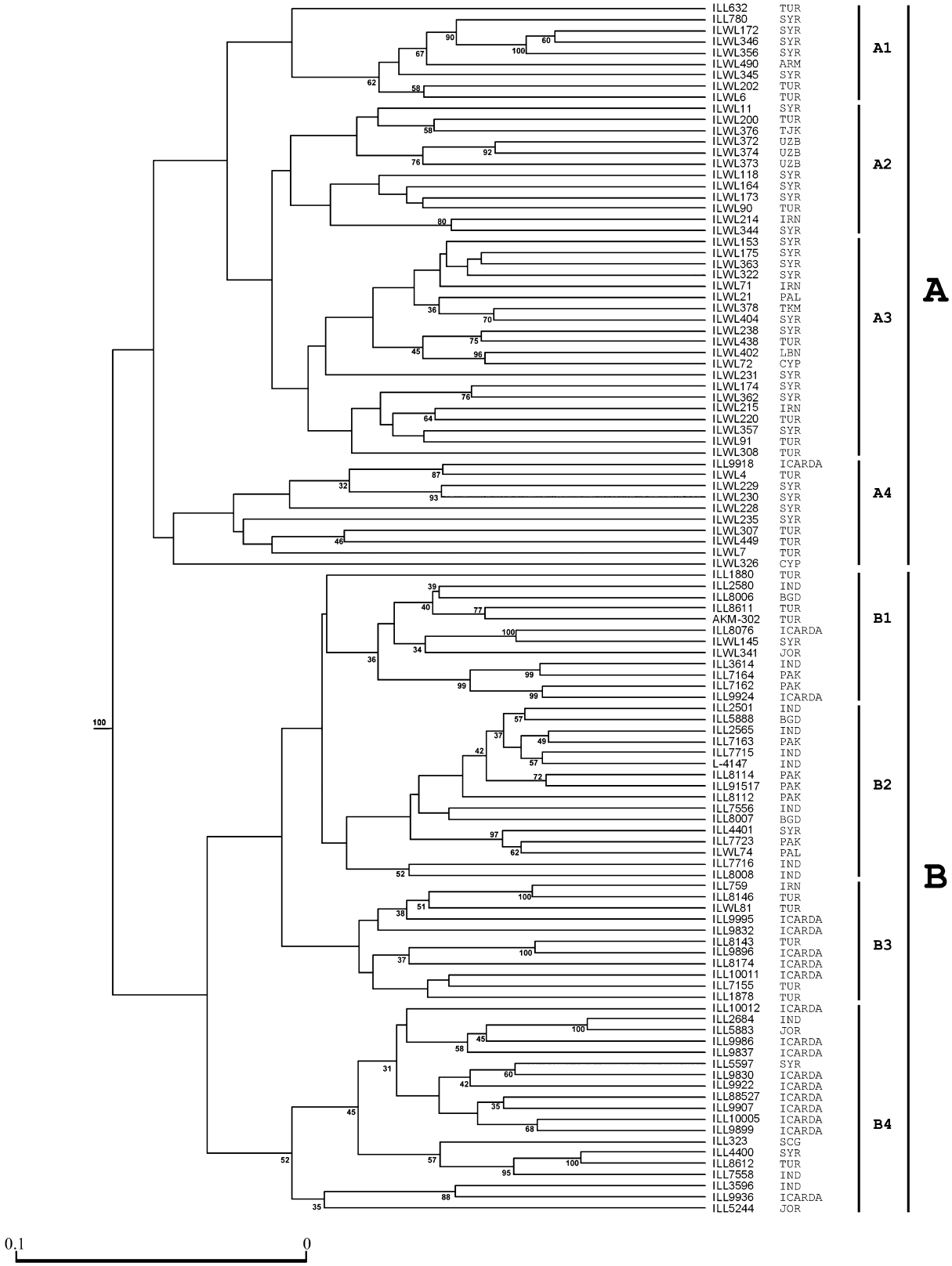

Fig. 1. Cluster analysis of wild and cultivated lentil accessions using 14 microsatellie markers. The groups are denoted on the right side as A or B, and the sub-groups as A1, A2, A3, A4, B1, B2, B3, and B4. The origins of 109 lentil accessions are listed closed to the genotype numbers. Bootstrap values of above 30% are indicated at the nodes. The abbreviations of the countries: Bangladesh (BGD), India (IND), Iran (IRN), Jordan (JOR), Pakistan (PAK), Syria (SYR), Turkey (TUR), Serbia & Montenegro (SCG), Palestine (PAL), Armenia (ARM), Cyprus (CYP), Uzbekistan (UZB), Tajikistan (TJK), Turkmenistan (TKM), Lebanon (LBN).

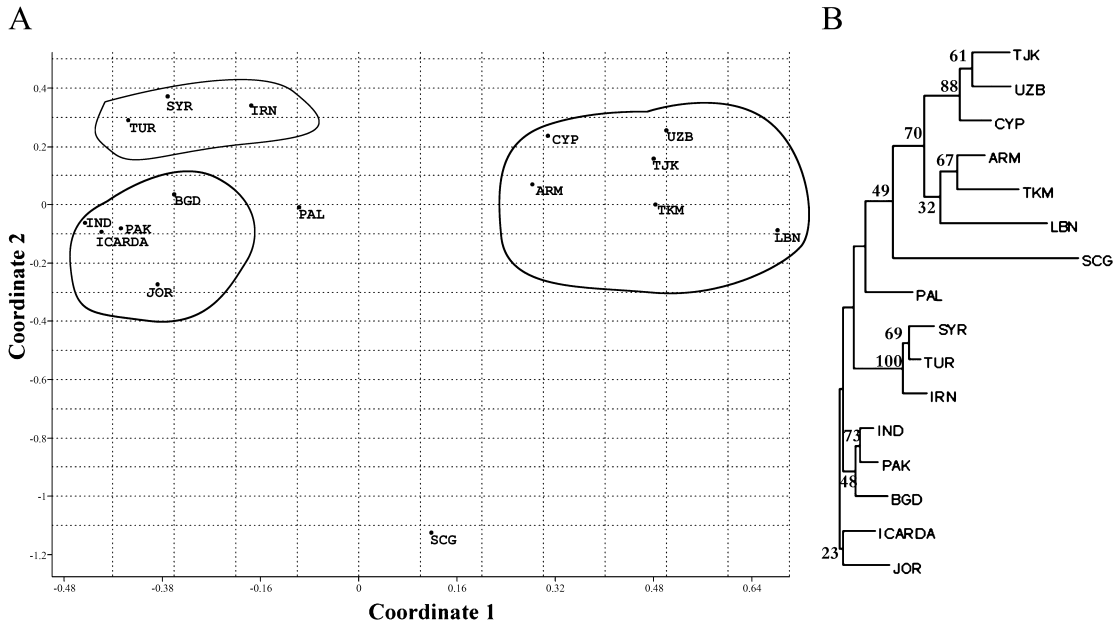

Fig. 2. Genetic relationships of lentil accessions originating from 15 countries of South Asia, Central Asia, and Fertile Crescent, as estimated by Nei's genetic distance (Nei 1973) based on the microsatellite data. (A): Principal co-ordinate analysis. (B): Neighbor joining cluster analysis. Bootstrap values are indicated at the nodes.

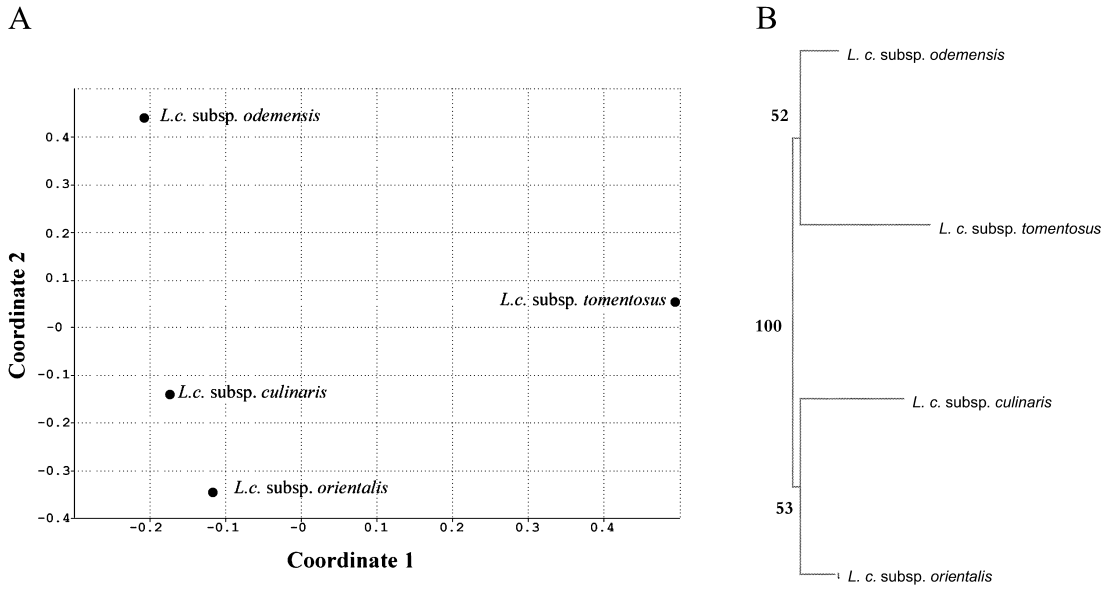

Fig. 3. Genetic relationships (as estimated by Nei' genetic distance) among the four *Lens* taxa based on 14 microsatellite markers analysis. (A) Principal co-ordinate analysis. (B) Neighbor joining cluster analysis. Bootstrap values are indicated at the nodes.

clusters separating Iran, Syria and Turkey from Jordan, India, Pakistan, Bangladesh and ICARDA breeding lines. The second cluster consisted of accessions from Armenia, Uzbekistan, Tajikistan, Turkmenistan, Cyprus, and Lebanon (Fig. 2).

The PCA of accessions based on the subspecies indicated that subspecies *culinaris* is more closely with subspecies *orientalis* than subspecies *tomentosus* and subspecies *odemensis* (Fig. 3).

Discussion

Isolation of microsatellites

The screening of 200,000 clones for $(GT)_{10}$, $(GA)_{10}$, $(GC)_{10}$, $(TA)_{10}$, $(TAA)_5$ and $(GAA)_8$ identified 371 clones (0.19%) likely to contain microsatellites. This is nearly 15 times less than the frequency achieved in wheat (2.8%; Röder *et al.* 1995) and about five times less than that achieved in chickpea (0.93%; Hüttel *et al.* 1999). This difference between the frequency of clones likely to contain

microsatellites in lentil and other species could be due to a property of lentil genome itself or due to various factors, such as the number and type of microsatellite motifs used for screening, insert size of the library, and the abundance of the microsatellites in the genome of each species.

Of the 371 lentil clones with putative microsatellite motifs, only 243 clones (65.5%) were used for sequencing. The remaining clones (34.5%) were rejected based on insert size analysis, because of the presence of more than one insert, or the possibility that they represented duplicate clones. After sequencing, 173 (71.2%) clones possessed microsatellites, and the remaining 70 clones (28.8%) were 'lost' because of the absence of a microsatellite. In several previous studies, inevitably some clones were sequenced that did not contain a microsatellite; in those studies, the proportions of not containing a microsatellite were, for example, 42% in *Anisantha sterilis* (Green *et al.* 2000), 33% in rapeseed (Kresovich *et al.* 1995), and 23% in olive (Carriero *et al.* 2002).

More of the repeats in lentil were dinucleotide rather than trinucleotide repeats. This finding agrees with earlier reports that dinucleotide repeats are relatively abundant in many plant species including wheat (Wang *et al.* 1994, Ma *et al.* 1996, Bryan *et al.* 1997). Our results indicated CA/GT repeats might be a major type of microsatellite in the lentil genome. The CA/GT repeat is one of the most frequently occurring microsatellites in humans and many other mammals (Toth *et al.* 2000) and also in some plant libraries, such as wheat (Varshney *et al.* 2000) and *Pinus radiata* (Smith and Devey 1994); however, that repeat is less frequent in many other plants (Lagercrantz *et al.* 1993).

Variation at microsatellite loci in cultivated and wild lentil species

We used the PCR conditions standardized for cultivated lentils successfully to amplify microsatellites in the wild relatives. Amplification with the same set of primers should be possible only if these species are closely related. Several studies have used homologous microsatellite flanking sequences developed in one species for diversity analysis in closely related species (Udupa *et al.* 1999, Choumane *et al.* 2000, Brondani *et al.* 2003, Adonina *et al.* 2005).

By assessing a worldwide collection of four subspecies of the genus *Lens* for a set of 14 microsatellite loci, we revealed a large amount of genetic variability among and within the subspecies. This result is not surprising considering the sampling scale and the class of markers we used. Microsatellite markers are known to show high mutation rates (Udupa and Baum 2001), and are thus expected to reveal a high degree of polymorphism (Udupa *et al.* 1999). We did, however, find large variation among microsatellite marker loci for both gene diversity and allelic numbers, with the exception of two loci (SSR28 and SSR202) which gave less than four alleles for the species tested.

Genetic relationships

The cluster analysis of 109 accessions indicated two major clusters at a bootstrap value of 100%. About 92% of wild accessions were clustered together in cluster A, and were clearly separated from the cultivated lentil. Similarly, 93% of cultivated lentil accessions clustered in the major cluster B. Four of 30 accessions (13.3%) from *L. culinaris* subsp. *orientalis* clustered with the cultivated germplasm, while none of the other wild subspecies were clustered with the cultivated germplasm. The clustering of some of the wild and cultivated accessions could be due to the occurrence of intermediate morphological forms particularly in the areas of the Fertile Crescent, resulting from out-crossing between wild and cultivated forms (gene introgression), as previously observed by Ferguson *et al*. (1998b).

L. culinaris subsp. *tomentosus* and *L. culinaris* subsp. *odemensis* are genetically distant from cultivated lentil, and are therefore unlikely to be the progenitors of cultivated lentil. The high degree of genetic differentiation at microsatellite loci and their hitch-hiking to the adaptive traits could be responsible for this level of the differences between wild and cultivated accessions. Genetic drift due to genetic bottlenecks and fluctuations in the population sizes of cultivated and wild species might have also contributed to the overall genetic differentiation in wild and cultivated lentil populations. A similar strong differentiation between cultivated and wild relatives has been observed in pigeonpea and pearl millet (Mariac *et al.* 2006, Yang *et al.* 2006).

High genetic diversity was observed within the accessions of *L. culinaris* subsp. *orientalis*, of which about 80% are from the Fertile Crescent (the north and south of Syria, Turkey, Iran, Jordan, Lebanon, Cyprus, and Palestine). Ferguson and Robertson (1996) found that *L. culinaris* subsp. *orientalis* possessed the high diversity based on morphological characters compared to isozymes markers. Ferguson *et al*. (1998a) also reported high genetic diversity (based on RAPD analysis) in *L. culinaris* subsp. *orientalis* accessions from Jordan, Palestine, and southern Syria, with slightly lower diversity in those accessions from northwest Syria and southeast Turkey.

The accessions of *L. culinaris* subsp. *odemensis* in this study collected from Syria, Turkey, and Palestine were distributed widely within cluster A, indicating high genetic diversity at microsatellite loci. Previously, Ferguson *et al*. (1998a) reported high genetic diversity for this species only in Sweida province of Southern Syria.

Out of the 22 accessions originating from South Asia (India, Pakistan and Bangladesh), 14 (64%) were clustered in sub-cluster B2, while the other accessions were distributed across the B1 and B4 clusters. The PCA based on geographic origin also showed that accessions from South Asian countries clustered together, providing molecular evidence for the clustering based on phenotypic variation observed previously for the South Asian accessions (Rahman and Sarker 1993, Erskine *et al.* 1998).

Although the accessions from Syria, Turkey, and Iran showed high genetic diversity, they were found to be closely related according to their origins (Fig. 2). These countries

are a part of the Fertile Crescent, where lentil was first domesticated. A previous study that mapped the geographical distribution of genetic variation in lentil from Syria and Turkey using RAPD markers also revealed areas of high and unique diversity (Ferguson *et al.* 1998a).

Our grouping of accessions based on microsatellite variation confirmed that the *orientalis* subspecies is closer to the *culinaris* subspecies than to the *tomentosus* and *odemensis* subspecies and did not show whether *tomentosus* subspecies is closer to the *culinaris* subspecies. However, Ferguson *et al*. (2000) re-classified *Lens* taxa, indicating that the *tomentosus* subspecies is closer to the *culinaris* subspecies than to the *odemensis* subspecies.

To meet the need for co-dominant markers in lentil, we identified additional microsatellite primers, and used them for the first time to study the genetic diversity within wild and cultivated lentils. The results indicated strong differentiation between wild and cultivated lentils, which were clearly separated by the analyses into two gene pools. Our results show a better resolution in classification of the *Lens* taxa by using microsatellite markers.

Acknowledgements

The research of the ICARDA scientists was supported by grants from the German Federal Ministry of Economic Cooperation and Development (BMZ, Bonn, Germany), the CGIAR's Generation Challenge Program, and the Global Center of Excellence for Dryland Science, Tottori, Japan.

Literature Cited

Abe, J., D.H. Xu, Y. Suzuki, A. Kanazawa and Y. Shimamoto (2003) Soybean germplasm pools in Asia revealed by nuclear SSRs. Theor. Appl. Genet. 106: 445–453.

Abo-elwafa, A., K. Murai and T. Shimada (1995) Intra- and interspecific variations in *Lens* revealed by RAPD markers. Theor. Appl. Genet. 90: 335–340.

Adonina, I.G., E.A. Salina, E.G. Pestsova and M.S. Röder (2005) Transferability of wheat microsatellites to diploid *Aegilops* species and determination of chromosomal localizations of microsatellites in the S genome. Genome 48: 959–970.

Bahl, P.N., S. Lal and B.M. Sharma (1993) An overview of the production and problems in southeast Asia. *In*: Erskine, W. and M.C. Saxena (eds.) Lentil in South Asia. Proceedings of the seminar on lentils in South Asia, ICARDA, Aleppo, Syria, pp. 1–10.

Brondani, C., P.H. Rangel, T.C. Borba and R.P. Brondani (2003) Transferability of microsatellite and sequence tagged site markers in *Oryza* species. Hereditas 138: 187–192.

Bryan, G.J., A.J. Collins, P. Stephenson, A. Orry, J.B. Smith and M.D. Gale (1997) Isolation and characterization of microsatellites from hexaploid bread wheat. Theor. Appl. Genet. 94: 557–563.

Carriero, F., G. Fontanazza, F. Cellini and G. Giorio (2002) Identification of simple sequence repeats (SSRs) in olive (*Olea europaea* L.). Theor. Appl. Genet. 104: 301–307.

Choumane, W., P. Winter, F. Weigand and G. Kahl (2000) Conservation and variability of sequence-tagged microsatellite sites (STMSs) from chickpea (*Cicer arietinum* L.) within the genus *Cicer*. Theor. Appl. Genet. 101: 269–278.

Collard, B.C.Y., R.A. Grams, W.D. Bovill, C.D. Percy, R. Jolley, A. Lehmensiek, G. Wildermuth and M.W. Sutherland (2005) Development of molecular markers for crown rot resistance in wheat: mapping of QTL for seedling resistance in a "2-49" X "Janz" population. Plant Breeding 124: 532–537.

Davis, J.C. (1986) Statistics and Data Analysis in Geology. John Wiley & Sons, New York.

Duke, J.A. (1981) Handbook of Legumes of World Economic Importance. Plenum Press, New York, pp. 52–57.

Erskine, W., S. Chandra, M. Chaudhry, I.A. Malik, A. Sarker, B. Sharma, M. Tufail and M.C. Tyagi (1998) A bottleneck in lentil: Widening its genetic base in South Asia. Euphytica 101: 207–211.

FAOSTAT database (2008) http://faostat.fao.org/

Ferguson, M.E. and L.D. Robertson (1996) Genetic diversity and taxonomic relationships within the genus *Lens* as revealed by allozyme polymorphism. Euphytica 91: 163–172.

Ferguson, M.E., B.V. Ford-Lloyd, L.D. Robertson, N. Maxted and H.J. Newbury (1998a) Mapping the geographical distribution of genetic variation in the genus *Lens* for the enhanced conservation of plant genetic diversity. Mol. Ecol. 7: 1743–1755.

Ferguson, M.E., H.J. Newbury, N. Maxted, B.V. Ford-Lloyd and L.D. Robertson (1998b) Population genetic structure in Lens taxa revealed by isozyme and RAPD analysis. Genetic Resources and Crop Evolution. 45: 549–559.

Ferguson, M.E., N. Maxted, H.V. Slangeren and L.D. Robertson (2000) A reassessment of the taxonomy of *Lens* Mill. (Leguminosae, Papilionoidae, Vicieae). Bot. J. Lin. Soc. 133: 41–59.

Green, J.M., K.J. Edwards, S.L. Usher, J.H. Barker, E.J. Marshall, R.J. Froud-Williams and A. Karp (2000) Microsatellites for Barren Brome (*Anisantha sterilis*). Mol. Ecol. 9: 2195–2197.

Hammer, Ø., D.A.T. Harper and P.D. Ryan (2001) PAST: Palaeontological Statistics software package for education and data analysis. Palaeontologia Electronica 4: 9.

Hamwieh, A., S.M. Udupa, W. Choumane, A. Sarker, F. Dreyer, C. Jung and M. Baum (2005) A genetic linkage map of *Lens* spp. based on microsatellite and AFLP markers and the localization of fusarium vascular wilt resistance. Theor. Appl. Genet. 110: 669–677.

Havey, M.J. and F.J. Muehlbauer (1989) Variability for restriction fragment lengths and phylogenies in lentils. Theor. Appl. Genet. 77: 839–843.

Hoffman, D.L., D.E. Soltis, F.J. Muehlbauer and G. Ladizinsky (1986) Isozyme polymorphism in Lens (*Leguminosae*). Syst. Bot. 11: 392–402.

Hüttel, B., P. Winter, K. Weising, W. Choumane, F. Weigand and G. Kahl (1999) Sequence-tagged microsatellite site markers for chickpea (*Cicer arietinum* L.). Genome 42: 210–217.

Kresovich, S., A.K. Szewc-McFadden, S.M. Bliek and J.R. McFerson (1995) Abundance and characterization of simple-sequence repeats (SSRs) isolated from a size-fractionated genomic library of *Brassica napus* L. (rapeseed). Theor. Appl. Genet. 91: 206–211.

Ladizinsky, D., D. Braun and F.J. Muehlbauer (1984) The biological species of the genus *Lens*. Bot. Gaz. 145: 235–261.

Lagercrantz, U., H. Ellegren and L. Andersson (1993) The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. Nucl. Acid Res. 21: 1111–1115.

Liu, K. and S.V. Muse (2005) PowerMarker: Integrated analysis environment for genetic marker data. Bioinformatics 21: 2128–2129.

Ma, Z.Q., M. Röder and M.E. Sorrells (1996) Frequencies and sequence characteristics of di-, tri- and tetra-nucleotide microsatellites in wheat. Genome 39: 123–130.

Mariac, C., V. Luong, I. Kapran, A. Mamadou, F. Sagnard, M. Deu, J. Chantereau, B. Gérard, J. Ndjeunga, G. Bezancon, J. Pham and Y. Vigouroux (2006) Diversity of wild and cultivated pearl millet accessions (*Pennisetum glaucum* L.) in Niger assessed by microsatellite markers. Theor. Appl. Genet. 114: 49–58.

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

Powell, W., G.C. Mackray and J. Provan (1996) Polymorphism revealed by simple sequence repeats. Trends in Plant Science 1: 215–222.

Rahman, M.M. and A. Sarker (1993) Genetic resources and breeding of lentil in Bangladesh. Lentils in South Asia. Proceedings of the Seminar on Lentils in South Asia, 11–15 March 1991. vi+236 pp.

Röder, M.S., J. Plaschke, S.U. König, A. Börner, M.E. Sorrells, S.D. Tanksley and M.W. Ganal (1995) Abundance, variability and chromosomal location of microsatellites in wheat. Theor. Appl. Genet. 246: 327–333.

Rogers, S.O. and A.J. Bendich (1985) Extraction of DNA from milligram amounts of fresh, herbarium and mummified plant tissues. Plant Mol. Biol. 5: 69–76.

Sarker, A. and W. Erskine (2006) Recent progress in the ancient lentil. Journal of Agricultural Science 100: 19–29.

Sharma, S.K., I.K. Dawson and R. Waugh (1995) Relationship between cultivated and wild lentils revealed by RAPD analysis. Theor. Appl. Genet. 91: 647–654.

Sharma, S.K., M.R. Knox and T.H.N. Ellis (1996) AFLP analysis of the diversity and phylogeny of *Lens* and its comparison with RAPD analysis. Theor. Appl. Genet. 93: 751–758.

Sibly, R.M., J.C. Whittaker and M. Talbot (2001) A maximum-likelihood approach to fitting equilibrium models of microsatellite evolution. Mol. Biol. Evol. 18: 413–417.

Smith, D. and M.E. Devey (1994) Occurrence and inheritance of microsatellites in *Pinus radiata*. Genome 37: 977–983.

Struss, D. and J. Plieske (1998) The use of microsatellite markers for detection of genetic diversity in barley populations. Theor. Appl. Genet. 97: 308–315.

Toth, G., Z. Gaspari and J. Jurka (2000) Microsatellites in different eukaryotic genomes: survey and analysis. Genome Res. 10: 967–981.

Udupa, S.M., L.D. Robertson, F. Weigand, M. Baum and G. Kahl (1999) Allelic variation at $(TAA)_n$ microsatellite loci in a world collection of chickpea (*Cicer arietinum* L.) germplasm. Mol. Gen. Genet. 261: 354–363.

Udupa, S.M. and M. Baum (2001) High mutation rate and mutational bias at $(TAA)_n$ microsatellite loci in chickpea (*Cicer arietinum* L.). Mol. Genet. Genomics 265: 1097-103.

van Oss, H., Y. Aron and G. Ladizinsky (1997) Chloroplast DNA variation and evolution in the genus *Lens* Mill. Theor. Appl. Genet. 94: 452–457.

Varshney, R.K., A. Kumar, H.S. Balyan, J.K. Roy, M. Prasad and P.K. Gupta (2000) Characterization of microsatellites and development of chromosome specific STMS markers in bread wheat. Plant Mol. Biol. Rep. 18: 5–16.

Wang, Z., J.L. Weber, G. Zhong and S.D. Tanksley (1994) Survey of plant short tandem DNA repeats. Theor. Appl. Genet. 88: 1–6.

Weber, J.L. (1990) Informativeness of human $(dC\text{-}dA)_n$ $(dG\text{-}dT)_n$ polymorphisms. Genomics 7: 524–530.

Winter, P. and G. Kahl (1995) Molecular marker technologies for crop improvement. World J. Microbiol. Biotechnol. 11: 449–460.

Yang, S., W. Pang, G. Ash, J. Harper, J. Carling, P. Wenzl, E. Huttner, X. Zong and A. Kilian (2006) Low level of genetic diversity in cultivated Pigeonpea compared to its wild relatives is revealed by diversity arrays technology. Theor. Appl. Genet. 113: 585–595.