

Diversity analysis of Central Asia and Caucasian lentil (*Lens culinaris* Medik.) germplasm using SSR fingerprinting

Sevda Babayeva · Zeynal Akparov · Mehraj Abbasov · Alamdar Mammadov ·
Mohammad Zaifzadeh · Kenneth Street

Received: 2 August 2008 / Accepted: 26 January 2009 / Published online: 6 March 2009
© Springer Science+Business Media B.V. 2009

Abstract Diversity analysis was performed among 39 cultivated lentil (*Lens culinaris* Medik.) accessions of Central Asia and Caucasian origin using five highly polymorphic microsatellite markers. A total of 33 alleles determined ranging from 3 to 8 per locus. Estimated gene diversity value for 33 loci was 0.66. Genetic similarity indices among 39 accessions ranged from 0.24 to 1.0. Cluster analysis using the unweighted pair group method with arithmetic mean method classified accessions into six major groups at 0.5 similarity coefficient. More than half accessions from Tajikistan formed large cluster. On the other hand, a few accessions from each country showed unique genotypes. Overall, most of the accessions, except ones with closely related origin, were distinguished by the present high quality DNA fingerprinting. This molecular diversity information gives important basis for conservation strategy in gene bank and exotic germplasm introduction in breeding programs in Central Asia and Caucasian countries.

Keywords DNA fingerprinting · Genetic diversity · *Lens culinaris* Medik. · Lentil · SSRs

Introduction

Germplasm fingerprinting gives knowledge of diversity and methods for efficient protection and conservation in genetic resources (Jarvis et al. 2000). Simple sequence repeat (SSR) markers are highly polymorphic, more reproducible and distributed throughout the genome (Joshi and Behera 2006). They are highly variable in the number of repeats in several nucleotide sequences and co-dominantly inherited (Johansson et al. 1992). They are also flexible for uses from the simple gel analysis to large automated detection system. Thus, SSRs are used in the genetic analyses of self-pollinating crops (Jin et al. 2008; Udupa et al. 1999) like legumes.

Lentil (*Lens culinaris* Medik.) is a self-pollinating diploid ($2n = 14$), annual cool season legume crop, that is produced throughout the world and highly valued as a source of food protein. Different types of molecular markers were used to determine allelic diversity of lentil collection (Datta et al. 2007). Pioneering works of Abo-elwafa et al. (1995) and Ford et al. (1997) gave important information for germplasm diversity, but also showed that non-specific PCR based markers could not provide repeatable results in differentiating lentil genotypes.

S. Babayeva (✉) · Z. Akparov · M. Abbasov
Genetic Resources Institute of ANAS, Baku, Azerbaijan
e-mail: seva_genetic@yahoo.com

A. Mammadov
Institute of Botany of ANAS, Baku, Azerbaijan

M. Zaifzadeh
Islamic Azad University – Ardabil Branch, Ardabil, Iran

K. Street
ICARDA, Aleppo, Syria

Hamwiah et al. (2005) develop ca. 80 sequence specific SSR markers in lentil and ca. 30 of them have already been assigned on the genetic map. These mapped SSRs provide a standard tool for fingerprinting lentil germplasm and markers for genetic analysis and breeding.

The objectives of the present study were to estimate the genetic diversity within Central Asia and Caucasian (CAC) lentil germplasm collection which has not been evaluated with an appropriate DNA fingerprinting system. The relationship among the germplasm in this area is discussed using the similarity data from the marker analysis.

Materials and methods

Plant material

A total of 39 lentil accessions of CAC origin obtained from International Centre for Agriculture in Dry Areas (ICARDA) genebank were analyzed. Two of these were improved cultivars and 35 were landraces, whereas categories of two accessions were unknown. According to the classification of Barulina (1930), seed type of 34 accessions was *microsperma* (*L. culinaris* Medik. ssp. *microsperma* (Baumg.) Barulina) (smaller seed) and five was *macrosperma* (*L. culinaris* Medik. ssp. *macrosperma* (Baumg.) Barulina) (larger seed). The list of lentil accessions and their passport information is presented in Table 1.

Marker analysis

A leaf sample of 500 mg from each accession was ground using liquid nitrogen and DNA was extracted according to CTAB method (Torres et al. 1993) and prepared for PCR amplification. Five SSR markers designed by Hamwiah et al. (2005) were used. These markers mapped genetically as SSR 167 (linkage group (LG)_1), SSR 33 (LG_3), SSR 323 (LG_5) and SSR 156 (LG_7), respectively. Their primer sequences were: SSR 33 (F: CAAGCATGACGCCT ATGAAG and R: CTTTCACTCACTCAACTCTC); SSR 156 (F: GTACATTGAACAGCATCATC and R: CAAATGGGCATGAAAGGAG); SSR 167 (F: CACATATGAAGATTGGTCAC and R: CATTAT GTCTCACACACAC); SSR 199 (F: GTGTGCATG GTGTGTG and R: CCATCCCCCTCTATC) and

SSR 323 (F: AGTGACAACAAAATGTGAGT and R: GTACCTAGTTTCATCATTG). PCR amplifications were performed in a 2720 thermal cycler (Applied Biosystems) with conditions of 3 min at 94°C, followed by 35 cycles of 15 s at 94°C and 15 s (with specific annealing temperature of the primer pair), 30 s at 72°C. PCR products were separated on 8% polyacrylamide gel and their fragment sizes were compared to 50 and 100-bp markers.

Data analysis

Fragments amplified by primer sets were scored for presence or absence which coded as 1 or 0, respectively. A binary matrix was then transformed to genetic similarity (GS) matrix using Dice similarity coefficient (Dice 1945; Nei and Li 1979). GS between pairs of accessions were measured as:

$$GS = \frac{2a}{(2a + b + c)}$$

where a is number of shared fragments, b is number of fragments in line A, and c is number of fragments in line B. Cluster analysis was performed based on unweighted pair group method with arithmetic mean method (UPGMA) using SAHN of software package NTSYS-pc (Rohlf 2000). Gene diversity (heterozygosity) was calculated according to Weir (1990):

$$\text{Gene diversity} = 1 - \sum_{ij}^n p_{ij}^2$$

where P_{ij} is the frequency of the j th pattern for SSR marker i and is summed across n patterns.

Results and discussion

Among the 39 accessions from seven countries representing CAC, a total of 33 bands were identified by 5 SSR markers. The number of alleles for each primer sets ranged from 3 (SSR 199) to 8 (SSR 156, 323). These numbers were high compared to those (3–5 by 10 accessions) in Inder et al. (2008), indicating that present lentil germplasm from CAC region had broader diversity.

The observed allelic frequencies ranged from 0.01 to 0.87, with an average of 0.147. Seventeen alleles (52%) appeared with the frequencies of 0.10 or lower, while only one allele was present in most

Table 1 Accession number, country of origin, collection site, category and seed types of lentil accessions used in this study

No	Acc. no	Origin ^a	Place of collection	Longitude	Latitude	Category	Seed type ^b
R1	ILL 85	TJK	–	–	–	Unknown	m
R2	ILL 516	TJK	–	–	–	Cultivar	m
R3	ILL 3395	TKM	–	–	–	Landrace	m
R4	ILL 3396	ARM	Kafan	E46 24	N39 12	Landrace	m
R5	ILL 4875	UZB	–	–	–	Landrace	m
R6	ILL 70167	AZE	Lankaran	E48 49	N38 45	Landrace	m
R7	ILL 70173	ARM	–	–	–	Landrace	M
R8	ILL 70181	TJK	–	–	–	Cultivar	m
R9	ILL 76281	TJK	Dushanbe	E71 40	N37 30	Landrace	m
R10	ILL 123466	UZB	Bukhara	E 64 25	N 39 46	Landrace	m
R11	ILL 123467	KAZ	Chimkent (South Kazakhstan)	E69 35	N42 19	Landrace	m
R12	ILL 123500	UZB	Bukhara	E 64 25	N 39 46	Landrace	m
R13	ILL 123508	TKM	Ashgabat	E58 22	N37 57	Landrace	M
R14	ILL 123525	TJK	–	–	–	Landrace	M
R15	ILL 123527	TJK	–	–	–	Landrace	m
R16	ILL 123534	ARM	Shirak	E46 01	N39 31	Landrace	m
R17	ILL 123568	TJK	–	–	–	Landrace	m
R18	ILL 123571	TJK	–	–	–	Landrace	m
R19	ILL 123576	TJK	–	–	–	Landrace	m
R20	ILL 123577	TJK	–	–	–	Landrace	m
R21	ILL 123579	TJK	–	–	–	Landrace	m
R22	ILL 123584	TJK	–	–	–	Landrace	m
R23	ILL 123585	TJK	–	–	–	Landrace	m
R24	ILL 123587	TJK	–	–	–	Landrace	m
R25	ILL 123593	TJK	–	–	–	Landrace	m
R26	ILL 123594	TJK	–	–	–	Landrace	m
R27	ILL 123595	TJK	–	–	–	Landrace	m
R28	ILL 123596	TJK	–	–	–	Landrace	m
R29	ILL 123599	AZE	Shamaxi	E48 38	N40 38	Landrace	m
R30	ILL 123603	AZE	Lankaran	E48 49	N38 45	Landrace	M
R31	ILL 123613	AZE	–	–	–	Landrace	M
R32	ILL 123618	AZE	Dzhebrail (Dzhafarabad)	E47 02	N39 22	Landrace	m
R33	ILL 123619	GEO	Samegrelo-Zerno Svaneti prov.	E42 43	N43 02	Landrace	m
R34	ILL 123624	TJK	Dushanbe	E70 21	N39 01	Landrace	m
R35	ILL 123626	TJK	Kulyab	E69 46	N37 54	Landrace	m
R36	ILL 123629	TJK	Kulyab	E69 46	N37 54	Landrace	m
R37	ILL 123631	TJK	Kulyab	E69 46	N37 54	Landrace	m
R38	ILL 123634	TJK	Kulyab	E69 46	N37 54	Landrace	m
R39	ILL 123644	TJK	Kulyab	E69 46	N37 54	Unknown	m

^a TJK Tajikistan, ARM Armenia, UZB Uzbekistan, AZE Azerbaijan, KAZ Kazakhstan, TKM Turkmenistan, GEO Georgia

^b m *microsperma*, M *macrosperma*

of lentil accessions with frequency of higher than 0.8. GD (Weir 1990) per locus ranged 0.24 (SSR119) to 0.89 (SSR 33) with the average

0.66. Analysis of genetic diversity within geographical regions found higher diversity index in accessions from Azerbaijan (0.60) compared to

Tajikistan (0.56), Uzbekistan (0.52) and Armenia (0.51).

Pair-wise similarity GS among 39 accessions ranged from 0.30 to 1.0. To visualize genetic relationships among 39 lentil accessions, a dendrogram was constructed based on the UPGMA cluster analysis (Fig. 1). There were six major groups generated at 0.5 similarity coefficients. One landrace from Azerbaijan (R30: cluster VI) and two landraces from Tajikistan (R15: cluster V and R17: cluster VI) showed a clear differentiation from the other accessions. Accession specific alleles of each R30 and R15 in SSR 156 support their unique genotype.

The dendrogram analysis (Fig. 1) revealed that no one cluster (except cluster IB) could be exclusively ascribed to any group a priori defined on the basis of its geographical origin. However, numerous tendencies clearly appeared.

Two main sub-groups were formed (A and B) in the cluster I. Subgroup IA contained five accessions of very diverse origins and without clear tendency. Two of Tajik, one of Uzbek, one of Azeri accessions and the only one studied from Kazakhstan were found in this group. Diversity in the cluster is possibly because of germplasm exchange among breeding centers which contribute to a clustering of accessions from different geographic origins that share common parents. It would be also interesting to note that two Tajik accessions were the closest inside subcluster IA with GS = 0.95.

The larger sub-cluster (IB) constituted a very homogenous group of 15 accessions: 14 from Tajikistan and 1 from the Azerbaijan. Of these, two (R8 and R18) and four (R20, R24, R35 and R36) accessions showed identical genotype in the present SSR markers. One part of accessions in subcluster IB were collected from same region Kulyab with collection site E 69°46' and N 37°54' in 1956 and thus have similar genetic backgrounds in the dendrogram. Because of the lack of information on the collection site of other accessions from Tajikistan, it was not possible to reach further conclusions regarding groupings in the dendrogram., anyway this observation might indicate that they share a common origin. The accession from Azerbaijan—R29 with macrosperma seed type, had same SSR 323 pattern with its nearest neighbor in IB subgroup—R23, which led to the grouping of this genotype along with the Tajik accessions. The confused relationships

among them could be due to the small number of loci being used.

Cluster II divided into two sub-groups (IIA and IIB) consisted of accessions representing almost all countries of origin in this study. Genotypes under IIA cluster showed a higher degree of similarity (0.66–0.95). All genotypes from Uzbekistan, except R5, and all genotypes from Armenia, except R4 were in this cluster. Both of Uzbek accessions were coming from Bukhara (E 64°25', N 39°46'), whereas the collection site of R5 was unknown. On the other hand, IIB was formed by only two accessions from Turkmenistan (R13) and Armenia (R4). Another accession from Turkmenistan (R3) also formed an isolated group in cluster III, indicating that accessions from Turkmenistan might have different pedigrees from germplasm in other countries. The remaining sub-cluster III contains accessions from Azerbaijan (R6) and Tajikistan (R9). Considering the exotic germplasm in cluster IV, V and VI, there are some unique germplasm in Azerbaijan and Tajikistan, even many of the accessions from Tajikistan showed high similarities.

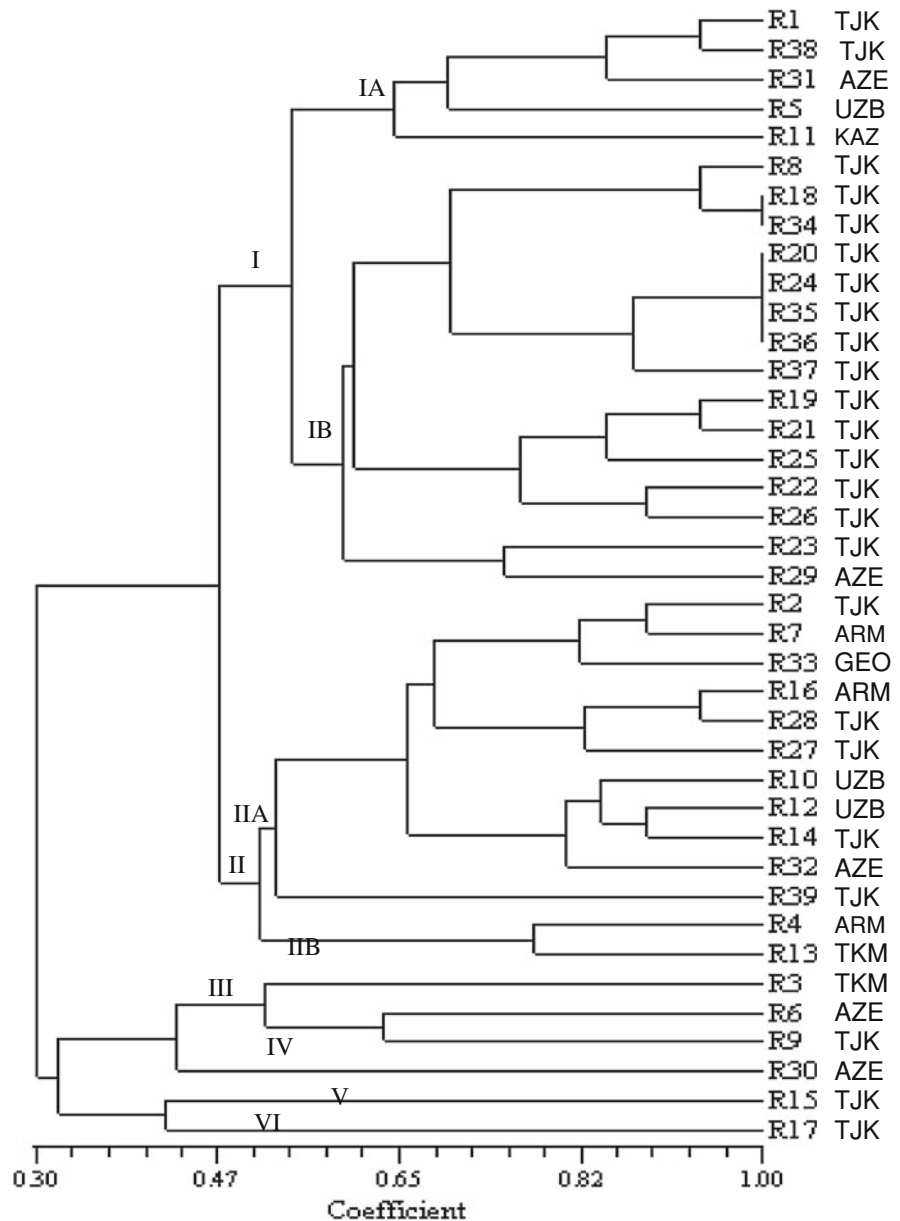
In summary, in the present study, three important results may be underlined. First, high linkage between genetic relations among accessions and their geographical regions of origin was found for accessions from Tajikistan, Uzbekistan and Armenia. Cluster IB was mainly composed of accessions from Tajikistan with one exception. Two out of three landraces from Uzbekistan fall into IIA subcluster, whereas all Armenian genotypes were grouped into II cluster: 2 into IIA and one into IIB.

Second, the dendrogram revealed the existence of isolated accessions that were genetically distinct in several countries: the Azeri accession in cluster IV, the Tajik ones in clusters V and VI.

Third, the germplasm from Azerbaijan proved to be a group of high diversity, as revealed by the low mean pairwise genetic similarities. Accessions from this country are scattered over four clusters. The high variability among Azeri accessions is possibly result of different morphological groups (Two *macrosperma* and three *microsperma*) as well as diverse provinces covered (between 38°45' and 40°38' in the North, and 47°02' and 48°49' in the East) (see Table 1).

No differentiation along Kazakhstan and Georgia could be studied because of the low sample size in these countries. Other authors (Lelley et al. 2000)

Fig. 1 A dendrogram showing the genetic similarity of 39 Central Asia and Caucasian origin lentil genotypes based on unweighted pair group method with arithmetic mean method using NTSYS software package



have encountered similar problems in the study of genetic distance.

Our results confirmed that the microsatellites could be powerful tools for study between cultivar diversity in lentil. The resolution of fingerprinting by current five SSRs was rather high for the 39 lentil germplasm used in the study. Except for some closely related germplasm with same collection place in Tajikistan, most of the germplasm were distinguished only by five reactions of PCR. A reason might be carefully chosen SSR markers with different linkage groups and their higher

number of alleles. Therefore, more attention should be focused on the establishment of a genotype database using convenient and effective indicators such as SSR, and hence monitoring the dynamic change in the gene pools and as a reference for germplasm management and breeding strategies. The fewer number of PCR can be multiplexed by using the fluorescent detection for automatic fingerprinting.

Since the germplasm information of lentil germplasm in CAC countries were not enough, further comparison of genomic constitution as well as the

trait evaluation will be necessary in the future. Information on current levels of genetic diversity of germplasm at gene bank is essential for designing appropriate strategies for future conservation. Especially, the DNA fingerprinting information gives important strategies for genebank conservation in a country, e.g. Azerbaijan, and germplasm introduction from outside of the country in the breeding program. According to the results of the present study, the high genetic differentiation among accessions from Azerbaijan suggests that this gene pool should be well represented by more samples/accessions *ex situ*.

Acknowledgements We thank Drs. Durna Aliyeva and Zarifa Suleymanova from Institute of Botany of ANAS for highly skilled technical assistance. Many thanks to ICARDA for providing seeds and its financial support for competition of research work. The authors are grateful to Dr. Kazuhiro Sato, Okayama University for critical reading of the manuscript.

References

- Abo-elwafa A, Murai K, Shimada T (1995) Intra- and inter-specific variations in *Lens* revealed by RAPD markers. *Theor Appl Genet* 90:335–340. doi:[10.1007/BF00221974](https://doi.org/10.1007/BF00221974)
- Barulina HI (1930) Lentils of the USSR and other countries. *Bull Appl Bot Plant Breed Leningrad Suppl* 40:1–319
- Datta S, Sourabh B, Kumari J, Kumar Sh (2007) Molecular diversity analysis of lentil genotypes. *Pulses newsletter*. Indian Institute of Pulses Research, Kanpur. Vol 18(2), 2 pp
- Dice LR (1945) Measures of the amount of ecologic association between species. *Ecology* 26:297–302. doi:[10.2307/1932409](https://doi.org/10.2307/1932409)
- Ford R, Pang ECK, Taylor PWJ (1997) Diversity analysis and species identification in *Lens* using PCR generated markers. *Euphytica* 96:247–255. doi:[10.1023/A:1003097600701](https://doi.org/10.1023/A:1003097600701)
- Hamwiah A, Udupa SM, Choumane W, Sarker A, Dreyer F, Jung C, Baum M (2005) A genetic linkage map of *Lens* sp. based on microsatellite and AFLP markers and localization of fusarium vascular wilt resistance. *Theor Appl Genet* 110:669–677. doi:[10.1007/s00122-004-1892-5](https://doi.org/10.1007/s00122-004-1892-5)
- Inder P, Materne M, Taylor PWJ, Ford R (2008) Genotyping elite genotypes within the Australian lentil breeding program with lentil-specific sequenced tagged microsatellite site (STMS) markers. *Aust J Agric Res* 59:222–225. doi:[10.1071/AR07188](https://doi.org/10.1071/AR07188)
- Jarvis D, Sthapit B, Sears L (2000) Conserving agricultural biodiversity in situ: a scientific basis for sustainable agriculture. International Plant Genetic Resources Institute, Rome
- Jin L, Jian-Ping G, Dong-Xu X, Xiao-Yan Z, Jing G, Xu-Xiao Z (2008) Analysis of genetic diversity and population structure in lentil (*Lens culinaris* Medik.) germplasm by SSR markers. *Acta Agron Sin* 34(11):1901–1909
- Johansson M, Ellegren H, Andersson L (1992) Cloning and characterization of highly polymorphic porcine microsatellites. *J Hered* 83:196–198
- Joshi RK, Behera L (2006) Identification and differentiation of indigenous non-Basmati aromatic rice genotypes of India using microsatellite markers. *Afr J Biotechnol* 6(4):348–354
- Lelley T, Stachel M, Grausgruber H, Vollmann J (2000) Analysis of relationships between *Aegilops tauschii* and the D genome of wheat utilizing microsatellites. *Genome* 43:661–668. doi:[10.1139/gen-43-4-661](https://doi.org/10.1139/gen-43-4-661)
- Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc Natl Acad Sci USA* 76:5269–5273. doi:[10.1073/pnas.76.10.5269](https://doi.org/10.1073/pnas.76.10.5269)
- Rohlf FJ (2000) NTSYS-pc numerical taxonomy and multivariate analysis system, version 2.1. Exeter Publ, New York
- Torres AM, Weeden NF, Martin A (1993) Linkage among isozyme, RFLP and RAPD markers in *Vicia faba*. *Theor Appl Genet* 85:937–945. doi:[10.1007/BF00215032](https://doi.org/10.1007/BF00215032)
- Udupa SM, Robertson LD, Weigand F, Baum M, Kahl G (1999) Allelic variation at (TAA)_n microsatellite loci in a world collection of chickpea (*Cicer arietinum* L.) germplasm. *Mol Gen Genet* 261:354–363. doi:[10.1007/s004380050976](https://doi.org/10.1007/s004380050976)
- Weir BS (1990) Genetic-data analysis methods for discrete genetic data. Sinauer Assoc Inc., Sunderland