Use of microsatellite markers for molecular characterization of cumin (*Cuminum cyminum* L.) ecotypes

Alireza Bahraminejad[1*] and Ghasem Mohammadinejad[2]

[1]Department of Plant Production, Islamic Azad University, Zarand Branch, Postal code: 7761146595, Zarand, Kerman, Iran.
[2]Horticultural Research Institute, Shahid Bahonar University of Kerman, Postal code: 76169133, Kerman, Iran.
**Corresponding Author, E-mail: a.bahrami@iauzar.ac.ir. Tel: 03424250856. Fax: 03424250855.*

ABSTRACT

In this study, Simple Sequence Repeat markers (SSR) were used to investigate the genetic variation between 49 cumin ecotypes collected from 9 different provinces of Iran. SSR primers Elap1479, Elap040 and Elap1493 showed the highest (89%), while Elap1340 and Elap017 showed the lowest (56%) number of polymorphic bands. Polymorphism information content (PIC) values varied between 0.18 - 0.37. The highest and the lowest PIC value were obtained by Elap017 and Elap1340 (0.37) and Elap040, Elap1493 and Elap1479 (0.18), respectively. Cumin populations of Semnan and Northern-Khorasan showed the highest difference, whereas Kerman and Esfahan populations exhibited the lowest difference. The obtained dendrogram classified all cumin population into three clusters at similarity level of 0.71; the first group was Semnan and Southern-Khorasan populations, the second includes Fars, Kerman, Northern-Khorasan, Khorasan-Razavi, Esfahan, Golestan and the third class consisted of Yazd populations. Based on these clusters, Kerman and Northern-Khorasan showed the closet genetic background and may belong to the same ancestor. The genetic markers used in this study can provide impetus information for tagging of economic traits. It can be concluded that the variability in the Iranian cumin populations have potential important source for cumin breeding objectives.

Keyword: Cumin, Genetic variability, Iranian ecotypes, SSR.

INTRODUCTION

Cumin (2n = 2x = 14) is one of the important members of *Apiaceae* family (*Umbelliferae*) (Bahraminejad *et al*., 2011). Essential oil is extracted from the cumin seeds, which appears similar to caraway seeds (often confused in Europe). Plants of the *Apiaceae* family have been used as food and medicine since humanity earliest written records (Igram, 1997). Cumin is a favorite food spice in Iran, occupying a wide geographical area in different parts of the country. Although it has been known as an ancient medicinal plant, there is no evidence of breeding population of cumin in Iran. Yearly cumin production is usually limited due to the incidence of different abiotic and biotic stresses such as plant diseases caused by *Fusarium oxysporum* and leaf burn (*Alternaria bumsii*) (Kafie and Rashed-Mohasel, 2002). The potential genetic diversity available to conduct conventional breeding for resistance against biotic and abiotic stress factors is be limited in cumin (Champawat and Pathak, 1990). Studies on genetic relationships among cumin ecotypes by means of morphological traits (Avatar *et al*., 1991; Bahraminejad *et al*., 2011), also variation in salinity response (Dhayal *et al*., 1999) as well as yield and growth traits (Baswana *et al*., 1983) have been reported previously. The development of PCR-based DNA markers, such as simple sequence repeat sequences (microsatellites, SSRs), has created the opportunity for wide genetic characterization of genotypes (Martínez-Gómez *et al*., 2007). Bahraminejad *et al*. (2012) reported a high variation between and within Iranian cumin germplasm using phenotypic

Table1. List of 49 studied cumin ecotypes / sub-populations from nine different provinces of Iran.

Population	Subpopulations / Ecotypes from each province						
Fars	1-Sarvestan	2-Sepidan	3-Sivand	4-Estahban			
Yazd	5-Ardekan	6-Bafq	7-Sadoq	8-Khatam	9-Sadroea		
Golestan	10-Maraveh-Tapeh	11-Aq-Qala	12-Jat	13-Gonbad			
Kerman	14-Baft	15-Bardsir	16-Cha-trood	17-Joopar	18-Kooh-banan	19-Mahan	20-Ravar
	21-Rafsanjan	22-Sirjan	23-Zarand				
Southern-Khorasan	24-Qaen	25-Nah-bandan	26-Birjand	27-Sarayan	28-Darm-ian		
Esfahan	29-Feridan	30-Semi-rom	31-Arides-tan	32-Naien	33-Khan-sar	34-Natanz	
Semnan	35-Shahmirzad	36-Sorkheh	37-Ivanaki	38-Kalateh			
Northen-Kho-rasan	39-Esfarayen	40-Shirvan	41-Bojnord	42-Baneh			
Khorasan-Razavi	43-Gonabad	44-Ferdows	45-Torbat-Heidarieh	46-Torbat-Jam	47-Kash-mar	48-Taybad	49-Bard-sekan

traits and RAPD markers. Microsatellites, alternatively known as simple sequence repeats (SSRs), short tandem repeats (STRs) or simple sequence length polymorphisms (SSLPs), are tandem repeats of sequence units generally less than 5 bp in length (Bruford and Wayne, 1993). These characteristics constitute the ideal markers for application in genetic mapping and genetic map comparisons (Rosa *et al*., 2003). Also, because of the lack of need for basic information about the genome studied and also the speed and high degree of confidence, this method is now used widely for the preparation of polymorphic markers (Chawla, 2003). SSR discovery in species with little or no DNA sequence information usually involves the construction and screening of partial genomic libraries and the sequencing of SSR-positive clones. This is a labor-intensive process, but the screening effort may be greatly reduced by the construction of genomic libraries highly enriched for SSR (Kolliker *et al*., 2001). The enrichment technique enhances the yield of microsatellites from 0.12% with the hybridization technique to 1%. Microsatellites in a perennial *Apiaceae* such as *Eryngium alpinum*, have the potential to provide a new insight on the genetic processes within and between populations (Gaudeul *et al*., 2002).

The present study was undertaken to assess the extent of genetic diversity and relationship among different populations of cumin (*Cuminum cyminum* L.).

MATERIALS AND METHODS

Plant materials

In this study, 49 cumin ecotypes belonging to 9 populations were collected from different provinces of Iran having a wide variation of cumin (Table1).

DNA extraction

Several DNA extraction protocols for DNA isolation were tested and finally a modified method based on CTAB (Krizman *et al*., 2006) was used to extract genomic DNA. Seeds of each ecotype (3 g) were used for DNA extraction. About 30 cumin seeds were soaked in liquid nitrogen, mixed with 6 mL extraction buffer [0.7 M NaCl, 50 mM Tris-HCl (pH 8.0), 10 mM EDTA, 1% CTAB, and 0.3% 2- mercaptoethanol] and incubated at 55°C for 30 minutes, while slowly shaking every 10 minutes. Subsequently, samples were transferred into 6 mL chloroform: Isoamyl alcohol (24:1) and centrifuged at 2400 rpm for 15 min. samples were left to stand for 30 min, then centrifuged at 2400 rpm for 15 min. RNase

Table 2. The microsatellite primers used for this study.

Locus	Microsatellite sequence in library	Primers sequences* (5′–3′)	Product size (bp)	*Ta* (°C)	No. of amplification cycles
Ealp024	(CT)8	F:CACTCCCTCATTTTCA R:CTACCACTTCTCCACT	108-114	50	45
Ealp040	(CT)9	F:CAAAGTCAAACGAAAAGG R:CAAAACGGGGAACATCA	275-295	45	35
Ealp245	(CA)2AA(CA)6	F:CTGAGTTCCATTCTTTTT R:AGGTGGTTGAGGGTTT	222-226	55	30
Ealp741	(CT)7	F:TACACTCCCTCTATCATTC R:ACGGCTTCTCTTCCTGCT	137-151	50	40
Ealp1340	(CT)7	F:TCGTGCCAGTTGTTGTTC R:CAGTAGAGGTAATGCCAGT	274-284	55	40
Ealp017	(CT)6	F:GCTCTCGGCTGTCTTATCTT R:CCGTTATTAGTCGCCTGAGT	110-112	55	35
Ealp1349	(CT)7	F:TCAACAGTAACCGACGACAA R:CGAGGACCCAACCCGAG	217-241	50	40
Ealp035	(CT)3TT(CT	F:CTCCAACTTCGCAAAATCA R:GTATAAACCGCTAAACCCT	101-119	50	40
Ealp1354	6AT(CT) (CT)7	F:CGTGCCCAGTTGTTGTTCC R:CCTAAAGAAAGAGAGAGACTA	205-217	55	30
EalpD268	(CA)10	F:AGCGGTATGAACAAGATGA R:TATATTAGTTGGTTAGGAGA	227-235	50	35
Ealp1493	(TC)4TA(TC)5 (TA)8(TG)5	F:AAAACTGGAACCGCCCCT R:TCTACCCACACATACATCATA	198-206	50	40
EalpD333	(CT)3TT(CT)7 TT(CT)4	F:AGGAGAGAGAAAGTTATGG R:GAGAAGGAAGTGAAAAAGG	251-265	50	35
Ealp1479	(CT)6(CA)4	F:TTTTCTCTGGCGTGCTGCT R:ACTTCAACCTGTGCGTATGT	176-182	55	30

* Gaudeul *et al*., (2002).

(1 µg/ml) was added to the supernatants for 2 hours at 55°C, 1 ml of 2-propanol added and the mixture was centrifuged at 2400 rpm for 5 min again. The pellet was washed twice in 70% ethanol, vacuum-dried, and re-suspended in 300 µl TE buffer (10 mM Tris- HCl and 0.1 mM EDTA, pH 8.0) and stored at -20°C. DNA samples were electrophoresed on an agarose gel (3%), were stained with ethidium bromide and observed under UV light in a Gel Documentation system. Varian Carry 50 spectrophotometer was used for DNA concentration determination.

SSR (Simple Sequence Repeat) analysis

Since the SSR markers had not been identified in cumin, 13 pairs of microsatellite primers were used which had been discovered in the Eryngium genus of *Apiaceae* family (Gaudeul *et al*., 2002) (Table 2). Amplification reactions were prepared in a 12 µl volume containing approximately 20 ng template DNA, 0.1 mM of each dNTP (Perkin Elmer), 0.5 µM of each primer, 2 mM $MgCl_2$ (Perkin Elmer), 0.5 U *Taq* Polymerase (Perkin Elmer), and 1× *Taq* Buffer (Perkin Elmer). Amplifications were performed in Gene Amp PCR System 2400 (Perkin Elmer), with the following cycling conditions: 10 min at 95°C, each cycle composed of 30 seconds denaturing at 95°C, 30 seconds for the annealing temperature (Ta), and 30 seconds extension at 72°C, and finally 7 min at 72°C to complete extension. Amplified fragments were then loaded on a 6% Long Ranger poly-

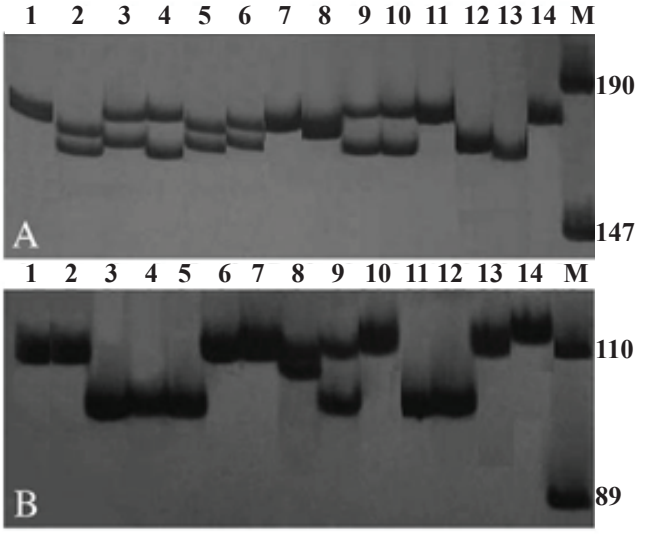

Figure 1. SSR profiles of 14 ecotypes of cumin using primer pair Elap1479 **(A)** and Elap040 **(B)** run on the polyacrylamide gel.

Table 3. Number of alleles and polymorphism information content (PIC) values of SSR markers for 49 cumin ecotypes.

Primers	Major allele frequency	Allele No.	PIC
Elap024	0.78	2.00	0.29
Elap040	0.89	2.00	0.18
Elap245	0.67	2.00	0.35
Elap741	0.78	2.00	0.29
Elap1340	0.56	2.00	0.37
Elap017	0.56	2.00	0.37
Elap1349	0.67	2.00	0.35
Elap035	0.67	2.00	0.35
ElapD268	0.67	2.00	0.35
Elap1493	0.89	2.00	0.18
ElapD333	0.78	2.00	0.29
Elap1479	0.89	2.000	0.18

acrylamide gel and electrophoresis was carried out for 3h on an automated sequencer ABI 377 (Perkin Elmer). Microsatellite patterns were visualized with genotyper 2.0 (Perkin Elmer). For subsequent statistical analysis, polymorphic bands amplified by SSR markers were scored as present (1) or absent (0). The generated data matrices were subjected to statistical analysis using the NTSYS-pc2 and Power Marker-v.3.25 analytical software (Applied Biostatistics, Setauket, USA) (Rohlf, 2000). Genetic similarities for SSR data were calculated by using the Dice similarity index, according to Nei and Li (1979). Dendrograms were constructed using the unweighted pair-group method with arithmetic averages (UPGMA) and the Gene Alex ver. 6.

RESULTS AND DISCUSSION

Molecular variability among cumin populations using SSR markers

This investigation was conducted using 12 polymorphic microsatellite markers. In total, a range of 108 to 284 bp bands were observed among cumin populations. Meanwhile the microsatellites markers Elap1479, Elap040 and Elap1493 showed the highest (89%), while Elap1340 and Elap017 showed the lowest number of polymorphic bands (56%, Figure 1). Polymorphic Information Content (PIC) values were obtained for 12 markers based on 49 ecotypes. The amount of PIC values varied from 0.18 - 0.37 and the highest level of PIC (0.37), was observed in microsatellites markers Elap1340 and Elap017, while the lowest PIC (0.18) was observed in microsatellite primers Elap040, Elap1493 and Elap1479 (Table 3). Cumin populations of Semnan and Northern-Khorasan showed the highest differences with eight polymorphic bands, whereas Kerman and Esfahan populations showed the lowest difference, Elap741 marker showed the highest polymorphic bands in Fars and Yazd populations, while in the rest of populations, Elap1340 marker exhibited the highest polymorphism. Cluster analysis revealed that the studied cumin populations were categorized into three classes at 0.71 level of similarity. The first group was Semnan and Southern-Khorasan populations and the second included Fars, Kerman, Northern-Khorasan, Khorasan-Razavi, Esfahan and Golestan. It is evident that most populations were placed in the second class, which consisted of five subclasses: the first subclass was Fars, second was Kerman and Northern-Khorasan, third included Khorasan-Razavi, fourth included Esfahan and fifth included Golestan populations. The third class also included Yazd populations. Based on these clusters, it can be concluded that Kerman and Northern-Khorasan, had the closest genetic background and may have the same ancestor (Figure 2). Average gene diversity based on Nie and Shannon was 0.37 and 0.54, respectively. Dendrogram grouping and credibility was

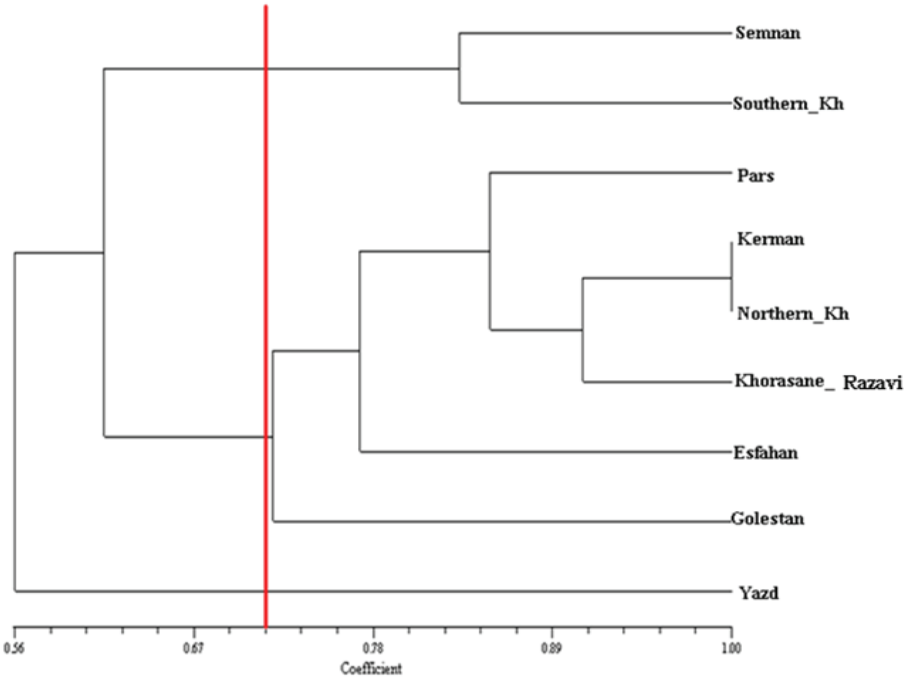

Figure 2. Dendrogram of 49 studied cumin ecotypes based on SSR markers according to the un-weighted pair group mean algorithm (UPGMA) with the Nei's similarity index.

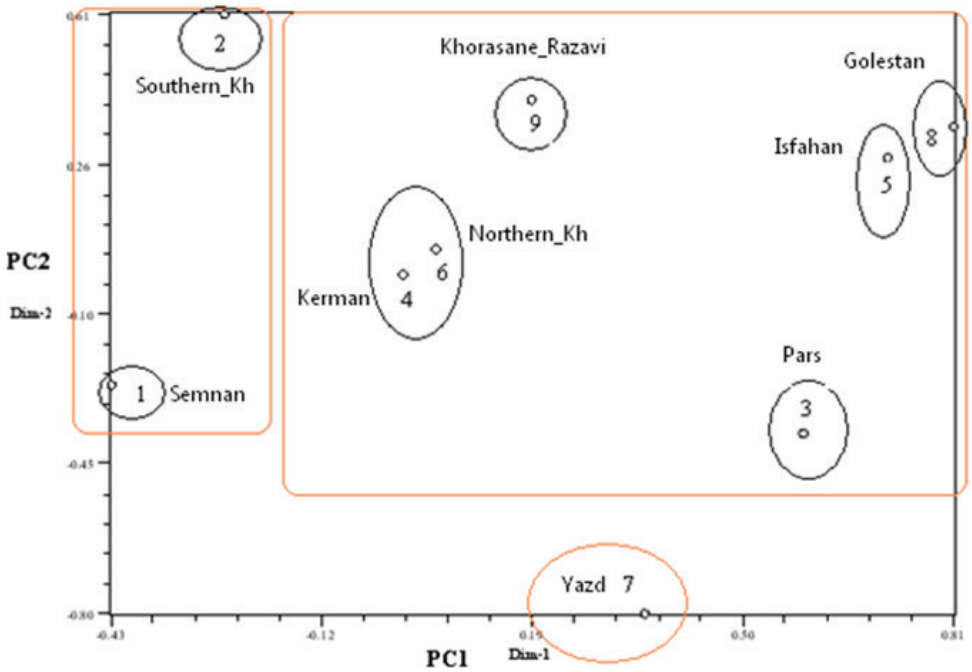

Figure 3. Biplot of nine cumin growing provinces by the first two principal components.

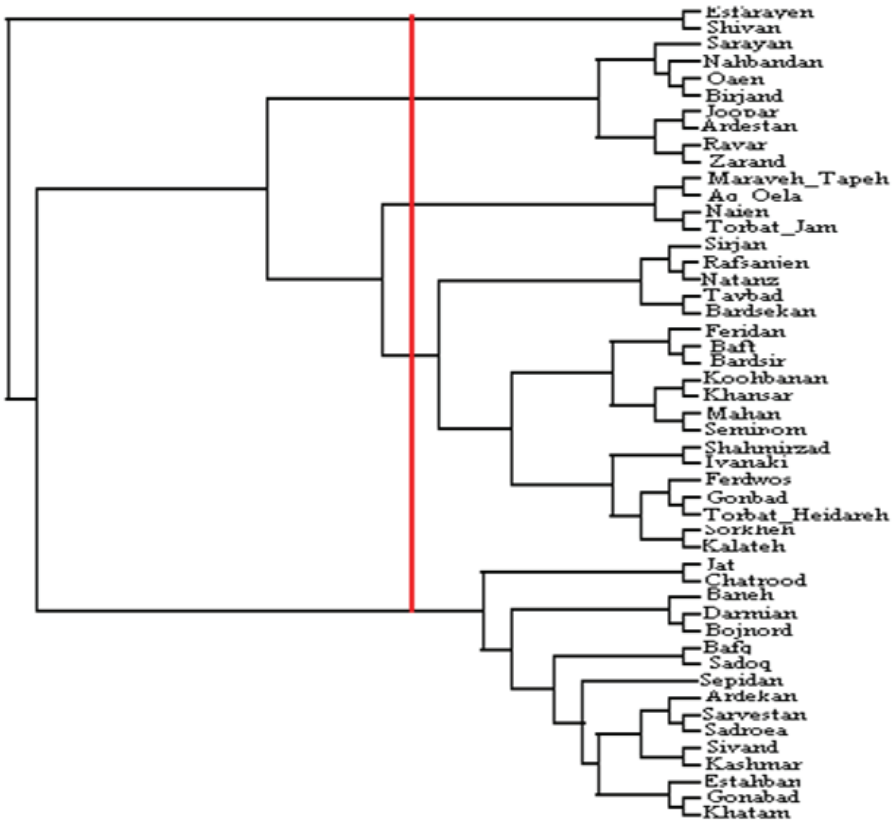

Figure 4. Dendrogram of 49 cumin ecotypes based on SSR markers according to the un-weighted pair group mean algorithm (UPGMA) with the Nei's similarity index.

confirmed by the high cophentic coefficient (r = 0.88). Results of pattern analysis based on the first two main principal component as well as cluster results wereshowed in Figure 3. According to this biplot, the second group had the highest number of populations, containing five subclasses, namely Fars, Kerman, Northern-Khorasan, Khorasan-Razavi and Esfahan; while Golestan and Yazd populations were classified in separate groups. The first three components which had the highest amount of variance justification could explain 85% of the total amount of the variation. Based on the obtained results, cluster and pattern analysis showed that yazd and Semnan populations had the farthest distance and can produce powerful hybrid lines in breeding strategies. On the other hand in breeding strategies to make high yielding populations as the composite varieties, those populations showing a far distances can be crossed and used for future breeding programs.

Molecular variability between the subpopulation (within population) with SSR markers

Kerman population with eleven polymorphic SSR markers possessed the highest intra-variation; followed by Golestan and Esfahan having nine, Khorasan-Razavi having eight, Northern-Khorasan with seven and Fars, Yazd and Southern-Khorasan with six and Semnan with five polymorphic SSR markers. Due to its cross pollination nature it is expected to have high seed yield for populations which has more variability. the dendrogram clasified 49 ecotypes in to five classes, the first class was ecotypes of Esfarayen and Shirvan and the second class consisted of two subclasses with Sarayan, Nahbandan, Qaen, Birjand as the first subclass and Joopar, Ardestan, Ravar and Zarand as the second subclass. The third subclass contained Maraveh-tapeh, Aq-Qala, Naien and Torbat_Jam. The fourth subclass had the highest number of ecotypes including two subclasses: first subclass

was Sirjan, Rafsanjan, Natanz, Taybad, Bardsekan and the second subclass included Feridan, Baft, Bardsir, Koohbanan, Khansar, Mahan, Semirom, Shahmirzad, Ivanaki, Ferdows, Gonbad, Torbat-Heidareh, Sorkheh and Kalateh. The fifth subclass comprised of three subclasses: the first subclass Jat, Chatrood and the second class Baneh, Darmian, Bojnord and the third subclass included Sepidan, Ardestan, Sarayan, Sadroea, Sivand, Kashmar, Estahban, Gonbad and Khatam (Figure 4).

ACKNOWLEDGMENTS

The author would like to thank Dr. Mihdzar Abdul Kadir from Department of Agriculture Technology, University Putra Malaysia for his guidance and Mr. Alagi Bah for his valuable comments and editorial assistance.

REFERENCES

Avatar R., Dashora S. L., Sharma R. K., and Sharma M. M. (1991). Analysis of genetic divergence in cumin (*Cumin cyminum* L.). *Indian Journal of Genetics and Plant Breeding*, 51: 289-291.

Bahraminejad A., Mohammadi-Nejad G., and Abdul Khadir M. (2011). Study of genetic diversity of cumin (*Cumin cyminum* L.) based on phenotypic characteristics. *Australian Journal of Crop Science*, 5: 301-307.

Bahraminejad A., Mohammadi-Nejad G., Abdul Khadir M., and Bin Yusop M. (2012). Molecular diversity of cumin (*Cuminum cyminum* L.) using RAPD markers. *Australian Journal of Crop Science*, 6: 194-199.

Baswana K. S., Pandita M. L., and Malik Y. S. (1983). Genetic variability studies in cumin (*Cuminum cyminum* L.). *Haryana Agricultural University Journal of Research*, 13: 596-598.

Bruford M.W., and Wayane R. K. (1993). Microsatellites and their application to population genetic studies. *Current Genetics*, 3: 939-943.

Champawat R. S., and Pathak V. N. (1990). Field screening of cumin germplasm against *Fusarium oxysporum* f. sp. *cumini*. *Indian Journal Arecanut Spices*, 13: 142.

Chawla H. S. (2003). Plant Biotechnology: A Practical Ap proach. Science Publishers, USA.

Dhayal L. S., Bhargava S. C., and Mahala S. C., (1999). Studies on variability in cumin (*Cuminum cyminum* L.) on normal and saline soil. *Journal of Spices and Aromatic Crops*, 8: 197-199.

Gaudeul M., Naciri Y., Guathier P., and Pompanon F. (2002). Isolation and characterization of microsatellites in a perennial *Apiaceae*, *Eryngium alpinum* L. *Molecular Ecology*, 2:107-109.

Igram C., (1997). Cumin Oil as a Plant and Spice, self-published, Buffalo Grove, Illinois, Lavabre, Marcel, Aromatherapy Workbook, Healing Arts Press.

Kafie M., and Rashed-Mohasel M. H. (2002). Effect of number of weed control, row spacing and density on growth and yield of cumin. *Journal of Agricultural Science and Technology*, 60:151-154.

Kolliker R., Jones E. S., Drayton M. C., Dupal M. P., and Forster J. W. (2001). Development and characterisation of simple sequence repeat (SSR) markers for white clover (*Trifolium repens* L.). *Theoretical and Applied Genetics*, 102: 416-424.

Krizman M., Jakse J., Baricevic D., Javornik B., and Prosek M. (2006). Robust CTAB-activated charcoal protocol for plant DNA extraction. *Acta agriculturae Slovenica*, 87: 427-433.

Martínez-Gómez P., Sánchez-Pérez R., Dicenta F., Howad W., Arus P., and Gradziel T. M. (2007). Almonds. In: Kole, C.R. (Ed.), Genome Mapping and Molecular Breeding. Fruits & Nuts, vol. 4 (pp. 229-242). Springer, Heidelberg, Berlin, New York, Tokyo.

Nei M., and Li W. H. (1979). Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, 87: 583-590.

Rohlf F. J. (2000). NTSYS-pc Numerical taxonomy and multivariate analysis system, Version 2.1 (New York: Exeter) Owners manue.

Rosa R. L., Angiolillo A., Guerrero C., Pellegrini M., Rallo L., Besnard G., Bervillé A., Martin A., and Baldoni L. (2003). A first linkage map of olive (*Oleae uropaea* L.) cultivars using RAPD, AFLP, RFLP and SSR markers. *Theoretical and Applied Genetics*, 106: 1273-1282.