



Cryptic Molecular-Geographical Divergence within *Medicago minima* Revealed by SSR Markers

Zohreh Bagheri¹ · Mostafa Assadi² · Ernest Small³ · Iraj Mehregan¹

Received: 21 July 2021 / Accepted: 10 October 2021 / Published online: 6 November 2021
© Shiraz University 2021

Abstract

Medicago minima (L.) Bartal. is an important rangeland forage and weed, native to Eurasia and Africa but widely distributed elsewhere. Like dozens of its nearest annual *Medicago* species, it is known to be highly uniform due to extreme inbreeding, and no valid infra-specific taxa meriting recognition have yet been proposed. In this study, 67 individuals from 13 natural populations of *M. minima* were collected and subjected to genetic diversity evaluation using 6 simple sequence repeat (SSR). Of the genetic variation parameters, expected heterozygosity (H_e) showed a significant variation from 0.0 in the Meshkinshahr population to 0.337 in the Andimeshk population. The analysis of molecular variance (AMOVA) revealed considerable partitioning of genetic diversity among populations (58%), while intrapopulation diversity accounted for 42%. Moderate to high population genetic differentiation was observed among particular populations, ranging from 0.060 between Marivan and Paveh to 0.562 between Meshkinshahr and Kohdasht populations. A weak total gene flow (0.115) was accompanied by a relatively high PhiPT value (0.617). Structure analysis, principal coordinate analysis, and neighbour-joining analysis, all divided the 13 populations into two genetic groups with significant genetic homogeneity. This discovery of two molecularly distinctive groups of *M. minima* in Iran, with a limited hybrid zone between them, is remarkable and suggests early divergent speciation.

Keywords Genetic variation · *Medicago* · Conservation · Gene flow · Genetic differentiation

1 Introduction

Medicago minima (L.) Bartal. (Fabaceae), a native of Eurasia and Africa has been introduced to much of the world as a minor weed (Small 2011). More importantly, this annual legume is very hardy and highly nutritious and is greatly valued as a rangeland forage (Farag et al. 2007; Judd et al. 2008; Beech et al. 2017). Annual medics (*Medicago* spp.) have significant advantages over perennials, including high growth rate, higher total crude protein,

considerable resistance to pests, vigorous seedlings, and extensive seed banks that enables them to survive harsh environmental conditions and make them essential forage sources (Walsh et al. 2001; Parsamanesh and Sadeghi 2019). Like other legumes, medics are especially friendly to soil due to the nitrogen-fixing ability of symbiotic bacteria housed in their root nodules (Bena et al. 2005; Küster 2013). In the light of the agricultural and environmental importance of annual and perennial species of *Medicago*, many institutions and governments have supported research on several of the *Medicago* species. Ability to tolerate drought, which is increasingly becoming more common, is the key to creating new cultivars that can provide sustainable fodder and forage (Humphries et al. 2020).

Medicago minima (L.) Bartal. is an ascending herb with branches 5–40 cm long, the vegetative organs, calyx, and fruit covered with dense simple and/or glandular hairs. The seeds are 1–2 per coil, kidney-shaped or oblong-ovoid, 1.5 to 2.5 mm long, and about 1.3 mm wide, isolated by a fruit

✉ Iraj Mehregan
Iraj@daad-alumni.de; imehregan@srbiau.ac.ir

¹ Department of Biology, Science and Research Branch, Islamic Azad University, Tehran, Iran

² Agricultural Research Education and Extension Organization (AREEO), Research Institute of Forests and Rangelands, Tehran, Iran

³ Ottawa Research and Development Centre, Science and Technology Branch, Ottawa, ON, Canada

septum, with yellowish to brownish, smooth seed coat, and a radicle approximately half the length of the seed. The diploid chromosome number is $2n = 16$ (Jha and Pal 1992; Small 2011). The ability to withstand the harsh environmental conditions and adaptation to various soil types makes *M. minima* an important potential forage sources in drylands of the world, including the fields and pastures of Iran, where it naturally occurs in northwest, west, and southwest areas (Walsh et al. 2001; Mehregan et al. 2002; Falahati-Anbaran et al. 2006, 2007; Parsamanesh and Sadeghi 2019).

There are very few cultivars of *M. minima* at this time, but given its numerous agronomical advantages and its economic and ecological importance, the germplasm of this species can be exploited for developing more cultivars. This is best done by first clarifying the genetic diversity structure of the species. Molecular markers have many advantages over morphological or cytological ones, as they are not affected by environmental factors (Werlemark et al. 1999; Nybom et al. 2006). Co-dominant simple sequence repeats (SSRs) have advantages over dominant markers such as random amplified polymorphic DNA (RAPD) or inter simple sequence repeat (ISSR) (Liu et al. 2007; Freeland 2020). SSRs are reliable tools to investigate intra and inter-population genetic diversity (Schlötterer et al. 1997; Belicuas et al. 2007; Xu and Crouch 2008; Silva et al. 2012; Bayat et al. 2021; Emami-Tabatabaei et al. 2021).

This study aimed to examine the genetic variability of the species in Iran in the hope of revealing variations that might be useful for breeding new cultivars.

2 Materials and Methods

2.1 Plant Material

Based on the information in Heyn (1984), and the herbaria of research institute forests and rangelands of Iran, a distribution map for *M. minima* in Iran was prepared (Fig. 1). Thirteen populations from various localities in Iran were selected for study (note information in Fig. 1 and Table 1). Employing individual plants separated by at least 20 m, leaf samples were taken from 5 to 6 individuals of *M. minima* in each population. In total, 67 samples were obtained from 13 populations. These were dried on silica gel for molecular studies. Voucher samples are deposited in the Islamic Azad University Herbarium (IAUH) (Table 1).

2.2 DNA Extraction

Total genomic DNA was extracted from silica-gel dried leaf samples using the NucleoSpin® Extract extraction kit

(Machery- Nagel, Germany) and CTAB method of Doyle and Doyle (1990). To assess the DNA qualitatively, 1% agarose gel stained with GelRed® Nucleic Acid Gel Stain (San Francisco Bay, CA, USA) was used. Quantitatively, the DNA was tested using NanoDrop™ 2000 (Thermo Fisher Scientific, Waltham, MA, USA) at 260 nm., and samples were stored at $-20\text{ }^{\circ}\text{C}$.

2.3 Isolation of Microsatellite Loci and Primer Design

100 ng of the extracted genomic DNA was used to generate an Illumina DNA library. After the fragmentation of the DNA, repair of the ends, “a-tailing”, and ligation of the TruSeq adapters, the DNA library was amplified with 8 cycles. The library was selected according to the average size of 670 bp. This size corresponded to an average integrate length of 500 bp. The “Library” was sequenced with 300 bp each in the “paired-end” mode in an Illumina-MiSeq run.

The overlapping “paired-end-reads” were first trimmed at the ends to remove residues of adapter sequences. The quality score was set to at least 20. 91.7% of the 56.9 million trimmed sequences had a Q value of 30%. With the help of the “FLASH” tool, the reads were assembled (Magoč and Salzberg 2011). A subset of the sequences was selected for designing primers based on the following criteria: the type and length of the microsatellite and the position within the sequence; as a result, 817,510 loci with any tandem repeat were selected. In this way, 2,410 loci were detected, which contain repetitive elements with the required criteria.

Of these sequences, 513 could be used to design primer pairs that met the following criteria: (1) PCR-Product 150–350 bp; (2) Primer size 18–22 bp; (3) TM-Value 58–62 °C. Using these criteria, primers designed by Primer3 software (Köressaar et al. 2018). The primer pairs were custom synthesized by Qiagen/Operon Technologies (Alame-da, Calif.). Out of the first 96 primer combinations, only 6 polymorphic loci could be detected.

2.4 PCR Amplification and SSR Assay

The 6 SSR primer pairs were tested for PCR amplification to assess the genetic variation in *M. minima* populations (Table 2). PCR mixture composed of 0.4 mM of each forward and reverse primer, 10 μM deoxyribonucleotides, 50 mM KCl, 10 mM TRIS–HCl (pH 8.3), 1.5 mM MgCl₂, 0.01% gelatin, 50–100 ng of DNA, and 1 U Taq DNA (Eujayl et al. 2004). The PCR amplification profile started with 4 min at 94 °C for denaturation, followed by 35 cycles of 30 s at 94 °C, 45 s at the annealing temperature (varied based on the locus), and 60 s at 72 °C, with a final

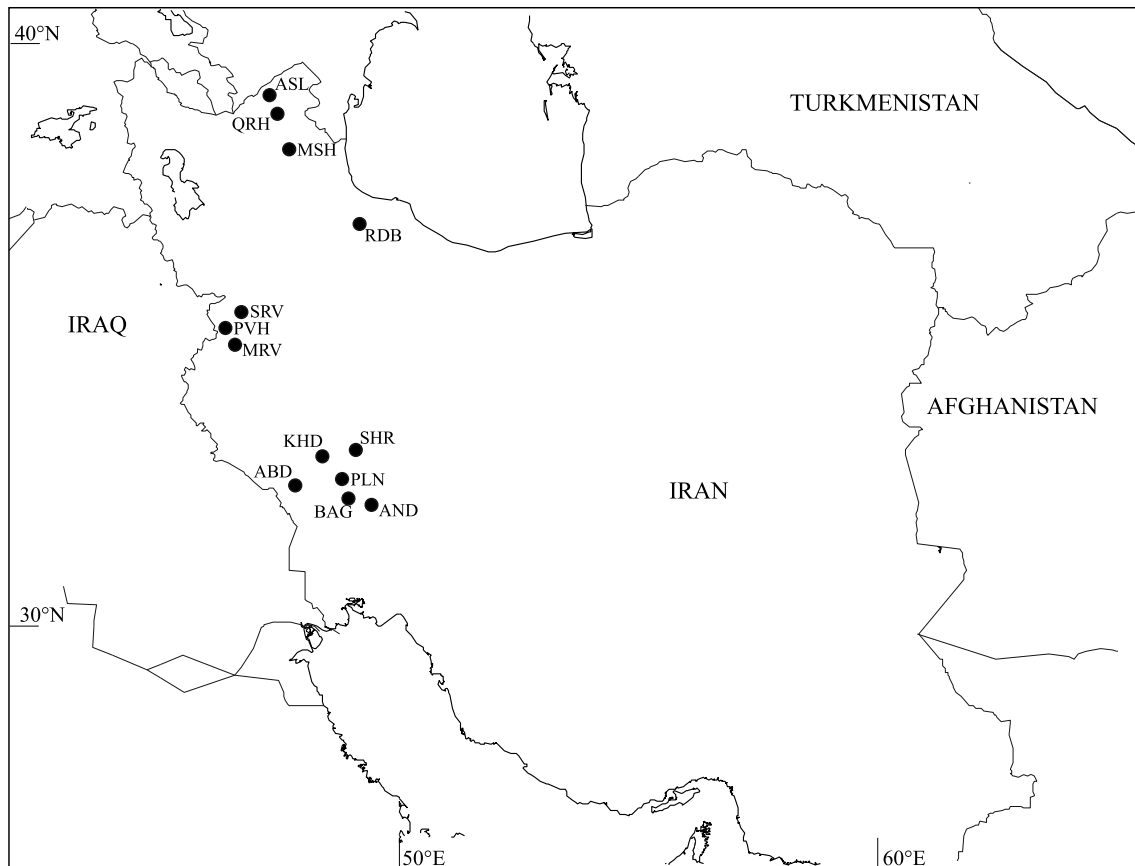


Fig. 1 Map of 13 *M. minima* populations located in northwest, west, and southwest of Iran

extension for 4 min at 72 °C. Two PCR product pools were made out of Min_A2–Min_A9–Min_C12 and Min_H2–Min_A1b–Min_A7b primer pairs. Next, 2 μ L of each pool mixed with 7.75 μ L of HiDi formamide (Applied Biosystems, Foster City, CA, USA) and 0.25 μ L of ROX-500 internal size standard (Applied Biosystems) in the following were injected to a 3730 \times 1 capillary sequencer (Applied Biosystems–Hitachi).

2.5 Data Analysis

The files obtained were aligned utilizing GeneMarker v2.4.2 (GeneMarker, SoftGenetics, State College, Pennsylvania). A signal intensity of at least 5000 was considered for those scored as present. A binary dataset of 1 as presence and 0 as absence created of combination of two primer sets used for analyses.

For each primer, polymorphic information content (PIC) was calculated according to Nei (1973): $PIC = 1 - \sum P_{ij}^2$ where P_{ij} is the frequency of the j th allele for i th locus across all alleles at a locus.

Genetic variation parameters including the number of observed alleles (N_o), the ratio of effective alleles (N_e), the Shannon information index (I), the observed, expected, total

expected, and mean expected heterozygosities (H_o , H_e , H_t , and H_s , respectively), unbiased Nei gene diversity (uH_e), the fixation index (F), the percentage of polymorphic loci ($P\%$), and gene flow (N_m) were estimated using GenAlEx version 6.5 (Peakall and Smouse 2006). F-statistics, including inbreeding coefficient within individuals ($F_{is} = (H_s - H_o)/H_s$); genetic differentiation among populations (F_{st}); F_{it} ($(H_t - \text{Mean } H_o)/H_t$) (Wright 1978) were utilized to estimate genetic differentiation among populations using GenAlEx version 6.5 (Peakall and Smouse 2006). To examine whether the matrix of genetic distances between *M. minima* populations correlated with the matrix of geographic distances between locations, the Mantel test (Mantel 1967) in (Tools for population genetic analyses) TFPGA (Miller 1997) was performed for Nei's unbiased genetic distances matrix and the matrix of geographic distances.

A hierarchical analysis of molecular variance (AMOVA) was carried out using GenAlEx version 6.5 (Peakall and Smouse 2006). A principal coordinate analysis (PcoA) was also carried out. The genetic relationship between populations was assessed by constructing a phylogenetic tree from a genetic distance with the neighbour-joining (NJ) method using POPTREE2 (Takezaki et al. 2010).

Table 1 Locality details and voucher number of the studied *M. minima* populations

Populations	Population code	Altitude	Coordinate	Locality	Province	Voucher number
Rudbar	RDB	500 m	[49° 25.163' E; 36° 50.03' N]	Rudbar, Darestan	Gilan	14,974
Meshkinshahr	MSH	780 m	[47° 39.915' E; 39° 1.362' N]	75 km from Aslanduz to Meshkinshahr	Ardabil	14,929
Garaghaj	QRH	500 m	[47° 39.062' E; 39° 7.803' N]	55 km from Aslanduz to Meshkinshahr Ardebil, Meshkinshahr	Ardabil	14,940
Aslanduz	ASL	220 m	[47° 17.223' E; 39° 21.074' N]	20 km from Aslanduz to Khoda Afarin	East Azerbaijan	15,034
Paveh	PVH	1200 m	[48° 14.951' E; 35° 23.402' N]	20 km from Marivan to Paveh	Kurdistan	15,005
Sarvabad	SRV	1260 m	[46° 21.264' E; 35° 19.023' N]	Sarvabad, Trooper of Sarvabad to Marivan	Kurdistan	15,011
Marivan	MRV	1180 m	[49° 12.799' E; 36° 08.115' N]	30 km from Paveh to Marivan	Kurdistan	15,007
Abdanan	ABD	1000 m	[49° 25.531' E; 33° 0.263' N]	Abdanan towards Kabir Kouh	Ilam	14,973
Poldokhtar	PLN	670 m	[47° 43.433' E; 33° 48.01' N]	10 km from Poldokhtar to Khorramabad	Lorestan	14,931
Andimeshk	AND	310 m	[48° 14.951' E; 35° 23.402' N]	55 km from Andimeshk to Poldokhtar	Lorestan	14,941
Shorabad	SHR	1100 m	[48° 10.286' E; 33° 27.407' N]	Khorramabad, Shorabad	Lorestan	14,981
Baghcheh	BAG	800 m	[48° 14.951' E; 35° 23.402' N]	5 km from Poldokhtar to Andimeshk	Lorestan	15,001
Kohdasht	KHD	1220 m	[48° 15.164' E; 33° 28.917' N]	5 km from Khorramabad to Kohdasht	Lorestan	15,015

Table 2 Description of 6 SSR primers, including sequence of the forward primer and the reverse primer, dye, annealing temperature (T_a), and length of the expected PCR product (bp) for each SSR set used for 13 wild population of *Medicago minima*

Name of the PCR product	Name of primers	Sequence of primers (including labelling agent)	T_m (°C)	Length of the expected PCR product (bp)
Min_A2	Min_A2_for	ACCGTCGCTTCGAGTTTCTA	60,0	296
	Min_A2_rev	Atto550-TCCTTGACCAACAACAGCAG	59,9	
Min_A9	Min_A9_for	CGGAAGTGACGTTAACGGAT	60,0	200
	Min_A9_rev	Hex-CCACATCTGAATTCTAGCCC	58,6	
Min_C12	Min_C12_for	GGTAAACGACCAATCACAAGG	59,3	262
	Min_C12_rev	Fam-GGGAAATATTGGCTTGGACA	59,7	
Min_H2	Min_H2_for	TTGAAAGTTCACAGCAAATCG	59,0	310
	Min_H2_rev	Atto565-TTGACAGAGTTGCAGCATCA	59,1	
Min_A1b	Min_A1b_for	GCTTGCCATAATGTTTGCC	60,5	282
	Min_A1b_rev	Atto550-AAATGCTCTAGAGGGCCACA	59,8	
Min_A7b	Min_A7b_for	TCAGAAGTGATATGCAGCGG	60,0	234
	Min_A7b_rev	Fam-GGTGTGCTTGAGCAATTGA	59,9	

The genetic structure of the studied populations was an analysis based on the Bayesian model and using STRUCTURE Ver. 2.3. (Pritchard et al. 2000) was used. Twenty independent replicate ran for 1–7 K. The values sat for the burn-in period, and the value of MCMC (Markov chain Monte Carlo) were 30,000 and 50,000 times, respectively (Porrás-Hurtado et al. 2013). The actual number of clusters (K) was determined by identifying the delta K value using an online structure harvester (Earl 2012).

3 Results

3.1 SSR Loci Polymorphism

The 6 SSR loci produced a mean of 3.16 alleles per locus (N ; Table 3); among loci, the effective number of alleles (N_e) did not vary significantly ranged from 1.100 in Min_A7B to 1.541 and 1.532 in Min_A2 and Min_H2, respectively. The latter happened to have the highest value for the Shannon information index (I , 0.395), expected, total, and mean expected heterozygosities (H_e , 0.258; H_t , 0.672; H_s , 0.258). Additionally, this locus with 0.285 was indicated to have a higher value for unbiased Nei's gene diversity (uH_e) than the rest of the loci. Locus Min_A9 on the other hand was found to have the highest observed heterozygosity (0.069). A notable variation in terms of polymorphic information contents (PIC) was observed among loci in which locus Min_C12 was found to be the most informative (0.996). Other loci, including Min_A9, Min_A2, and Min_H2, also were significantly informative (above 0.50), however, loci Min_A7b (0.478) and Min_A1b (0.517) were moderately informative, respectively, below and close to 0.50. F-statistics revealed invaluable information on loci in this study where the mean of Wright's fixation index (F) was notably high (0.88) with

the lowest value was in Min_A9 (0.698) and the highest in Min_A7b, Min_C12, and Min_A1b (1.00). Similarly, the lowest values for the inbreeding coefficient within individuals (F_{is}) and F_{it} (0.599 and 0.758) were observed in Min_A9; other loci had significantly higher, close to 1.00 values.

3.2 Genetic Diversity in *M. minima*

Among the 13 studied populations of *M. minima*, the observed mean number of alleles varied from 1 in the Meshkinshahr population to 2.33 in the Marivan population, the former had the value 1 for N_e and the absence of any value for the rest of genetic diversity indices (I , H_o , H_e , uH_e , F , and percentage of polymorphic loci ($P\%$) and the latter also had the highest N_e and I values (1.686 and 0.540, respectively) (Table 4). The population from Garaghaj presented the largest H_o (0.133), whereas H_e and uH_e were the highest in Andimeshk population (0.337 and 0.374, respectively). In the case of Wright's fixation index, populations with negative F (Garaghaj) had the lowest $P\%$ as well (16.67%). However, the increase in F was not consistent with enhancing $P\%$. For example, the population of Rudbar with 50% P had the highest F value (0.841) or Andimeshk population showed 83.33% P while it had the F value of 0.739 (Table 4). Furthermore, the genetic differentiation coefficient (F_{st}) among the 6 loci was generally high, and Min_A7b and Min_A1b with 0.856 and 0.837 observed to be genetically apart from the rest of the loci, and Min_C12 locus had the most negligible value of F_{st} (0.250). Following the pattern observed for F_{st} among loci, gene flow (N_m) was varied where the locus with the lowest F_{st} (Min_C12) observed to have the highest N_m value (0.750), and Min_A7b with the most elevated genetic differentiation showed the least N_m (0.042) (Table 3).

Table 3 Genetic diversity parameters of the 6 SSR loci across 67 *M. minima* individuals

Locus	N	N_a	N_e	I	H_o	H_e	H_t	H_s	uH_e	PIC	F	F_{is}	F_{it}	F_{st}	N_m
Min_A9	2	1.500	1.271	0.263	0.069	0.172	0.286	0.172	0.191	0.956	0.698	0.599	0.758	0.397	0.380
Min_C12	3	1.357	1.224	0.187	0.000	0.120	0.160	0.120	0.133	0.996	1.000	1.000	1.000	0.250	0.750
Min_A2	3	1.714	1.541	0.386	0.043	0.245	0.584	0.245	0.271	0.984	0.855	0.825	0.927	0.581	0.181
Min_A7b	2	1.143	1.100	0.084	0.000	0.057	0.396	0.057	0.063	0.478	1.000	1.000	1.000	0.856	0.042
Min_A1b	2	1.143	1.132	0.096	0.000	0.069	0.420	0.069	0.076	0.517	1.000	1.000	1.000	0.837	0.049
Min_H2	7	1.714	1.532	0.395	0.052	0.258	0.672	0.258	0.285	0.922	0.776	0.797	0.922	0.616	0.156
Mean	3.16	1.42	1.332	0.235	0.027	0.153	0.419	0.153	0.169	0.808	0.888	0.87	0.934	0.589	0.259

Alleles observed N , Number of different alleles N_a , Number of effective alleles N_e , Shannon information index I , Observed heterozygosity H_o , Expected heterozygosity H_e , Total expected heterozygosity H_t , Mean expected heterozygosity H_s , Unbiased Nei's gene diversity uH_e , Polymorphic information content PIC , Wright's fixation index F , Inbreeding coefficient within individuals F_{is} , Genetic differentiation among populations F_{st} , Fit and gene flow N_m

Table 4 Genetic diversity parameters of 13 studied populations of *M. minima* using 6 SSR markers

Pop	N	N_a	N_e	I	H_o	H_e	uH_e	F	%P
RDB	5	1.50	1.278	0.269	0.033	0.177	0.196	0.841	50.00
MSH	5	1.00	1.000	0.000	0.000	0.000	0.000	0.000	0.00
QRH	5	1.17	1.154	0.112	0.133	0.080	0.089	-0.667	16.67
ASL	5	1.67	1.411	0.344	0.100	0.217	0.241	0.603	50.00
PVH	5	1.67	1.390	0.352	0.067	0.233	0.259	0.603	66.67
SRV	5	1.50	1.450	0.288	0.133	0.187	0.207	0.271	66.67
MRV	6	2.33	1.686	0.540	0.111	0.306	0.333	0.705	66.67
ABD	5	2.00	1.676	0.501	0.100	0.313	0.348	0.729	66.67
PLN	5	1.67	1.314	0.334	0.000	0.213	0.237	1.000	66.67
AND	5	2.17	1.609	0.544	0.100	0.337	0.374	0.739	83.33
SHR	5	1.83	1.595	0.461	0.067	0.307	0.341	0.807	66.67
BAG	6	1.83	1.557	0.434	0.056	0.278	0.303	0.833	66.67
KHD	5	1.67	1.507	0.410	0.100	0.283	0.315	0.643	66.67
Mean	–	1.69	1.432	0.353	0.076	0.225	0.249	0.546	56.41

Number of individuals N , Number of different alleles N_a , number of effective alleles N_e , Shannon information index I , Observed heterozygosity H_o , Expected heterozygosity H_e , Unbiased Nei's gene diversity uH_e , Polymorphic information content PIC , Wright's fixation index F , Polymorphism percentage %P. For an explanation of population symbols, see Table 1

3.3 Population Genetic Structure of *M. minima*

The Bayesian-based model method was utilized to investigate the structure of 67 individuals of *M. minima*. To identify the optimum number of genetic groups of the investigated populations, delta K harvested using the Evanno method was implemented on STRUCTURE HARVESTER website. The K number was determined to be two (Fig. 2A). The membership of each individual in

shaping the population structure is presented in Fig. 2B. Each vertical line represents an individual, and those with a probability of 0.75 counted as homogenous and the lower as an admixture. The results indicated the presence of 47 individuals in the “red cluster” and 21 in the “green” without heterogeneity, while only 4 individuals showed genetic admixture. Only three populations, Abdanan, Baghcheh, and Marivan, clustered into the second group showed genetic admixture.

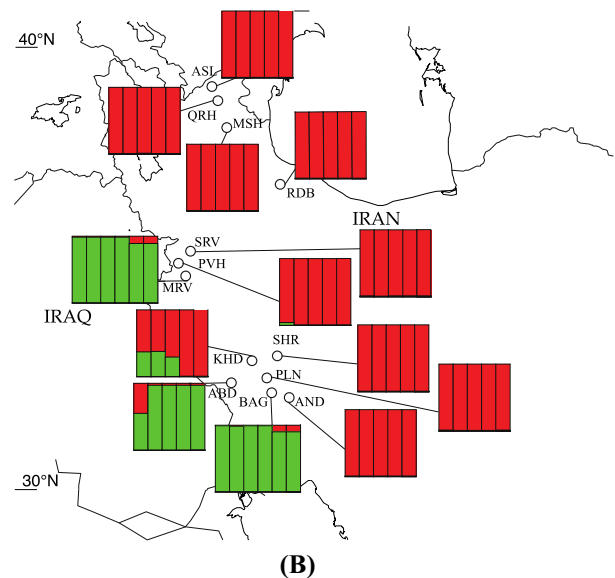
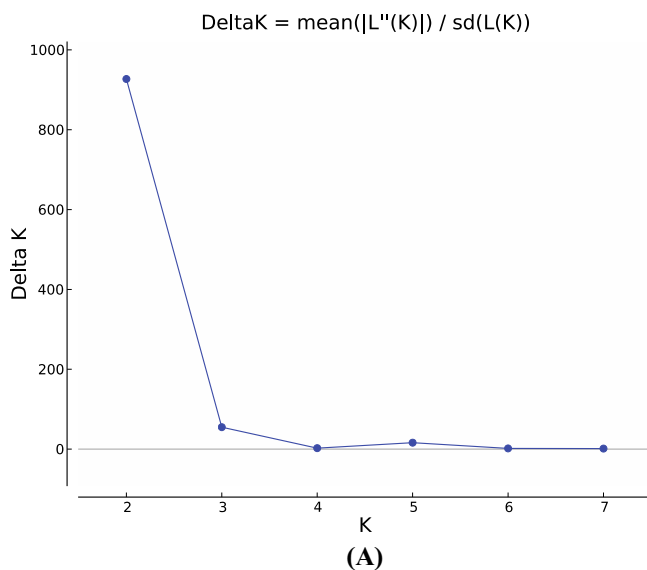


Fig. 2 The outcomes of STRUCTURE analysis for 67 *M. minima* individuals based on SSR markers data. (A, left) Delta K value detecting the optimum number of genetic groups (K, 2). (B, right) The

map of populations and the proportion of alleles shared by individuals in each population

To study correlation between Nei's unbiased genetic distance and the matrix of geographic distances of 14 populations of *M. minima*, the Mantel test was utilized to reveal the significant correlation ($R = 0.70$, $P = 0.002$) between these two matrixes. The principal coordinate analysis (PCoA) divided the population into two groups, similar to the structure analysis. However, the first group (smaller cluster) harboured geographically closed populations. Still, it did not wholly follow the structure analysis pattern since it contained Marivan, Sarvabad, Abdanan, and Kohdasht. The second group encompassed the rest of the populations (Fig. 3). Furthermore, the neighbour-joining clustering based on genetic distance confirmed the number of genetic groups (2) in structure analysis and PCoA, however in terms of the populations contained, it was only consistent with PCoA result and populations of Marivan, Sarvabad, Abdanan, and Kohdasht grouped similar to NJ result. The remaining 10 populations placed in the second cluster inside the cluster the population grouping mainly indicate significant geographical affinity, for instance, Roudbar and Meshkinshahr, or Poldokhtar and Andimeshk grouped, however populations like Garaghaj from north grouped with Baghcheh from the south (Fig. 4).

To carry out assessments of among-population variation, AMOVA and pairwise F_{st} analysis were used. AMOVA analysis (Table 5) revealed the distribution of total genetic diversity where the share of among populations (58%) was higher than within-population (42%) from overall genetic diversity. In agreement with F_{st} , AMOVA showed the

presence of a strong genetic differentiation among populations (P , 0.001, $\Phi_{PT} = 0.617$, Table 5). The overall gene flow was minimal (0.155). Moreover, the pairwise F_{st} test ranged from 0.060 between Marivan and Paveh populations to 0.562 between Meshkinshahr and Kohdasht populations (Table 6). Thus the genetic results roughly parallel the geographical distances between populations.

4 Discussion

4.1 SSR Loci Polymorphism

In this study, the 6 SSR loci produced a mean of 3.16 alleles per locus, which is relatively low, compared to those of 6.57 for the Moroccan *M. minima* (Zitouna et al. 2014), 3.5 for the Iranian *M. polymorpha* (Emami-Tabatabaei et al. 2021), and 6.5 for the Iranian *M. rigidula* complex (Bayat et al. 2021). The mean PIC for the Iranian *M. minima* was 0.808, compared to 0.66 for *Medicago truncatula* Gaertner (Eujayl et al. 2004), 0.923 for *M. polymorpha* (Emami-Tabatabaei et al. 2021), and 0.89 for the *M. rigidula* complex (Bayat et al. 2021). Dominant markers produce lower PIC contents for *Medicago sativa* L. (ISSR, Rashidi et al. 2013; AFLP, Keivani et al. 2010). Therefore, these SSR loci proved to be highly informative in investigating the genetic diversity of *M. minima* populations.

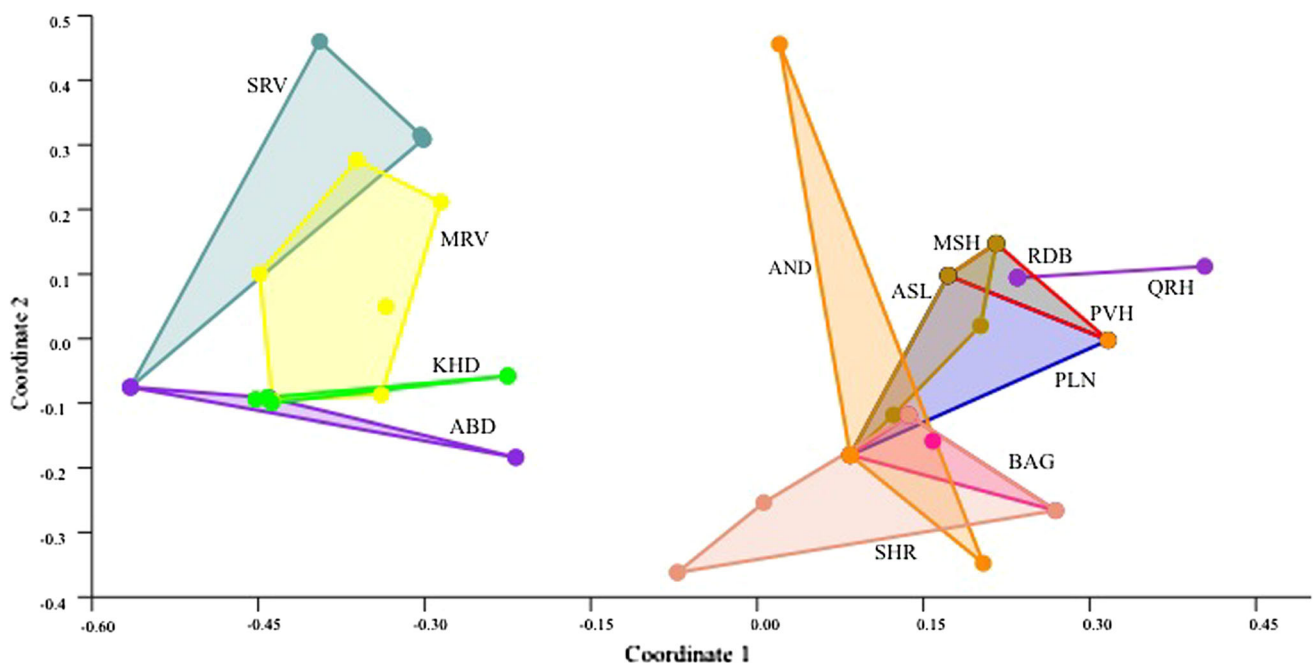


Fig. 3 Principal coordinates analysis (PCoA) for 13 populations of *M. minima*

Fig. 4 SSR-based phylogenetic tree showing the relationship of *M. minima* populations in Iran

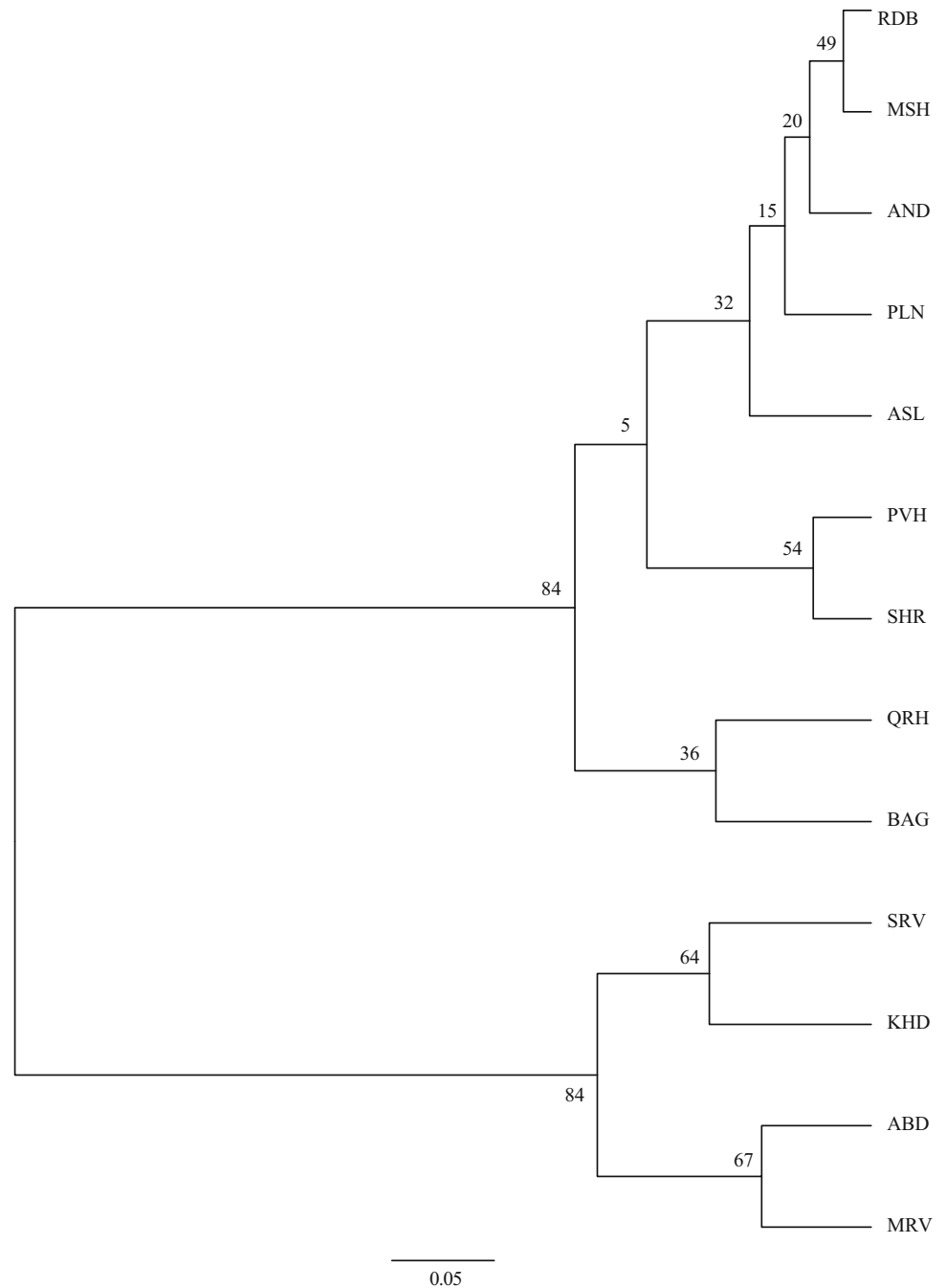


Table 5 Summary of the Analysis of molecular variance (AMOVA) results for 13 populations of *M. minima*

Source	df	Sum of square	Mean of square	Variance component	Percentage of variation	PhiPT	N _m
Among Pops	12	110.736	9.228	0.790	58%		
Within Pops	54	58.817	1.089	0.507	42%		
Total	66	169.553		1.297	100%	0.617	0.155

Table 6 Pairwise Genetic Differentiation Index (F_{st}) between the 13 populations of *M. minima*

Pop	RDB	MSH	QRH	ASL	PVH	SRV	MRV	ABD	PLN	AND	SHR	BAG	KHD
RDB	0.000												
MSH	0.133	0.000											
QRH	0.151	0.429	0.000										
ASL	0.188	0.377	0.396	0.000									
PVH	0.150	0.283	0.297	0.022	0.000								
SRV	0.134	0.277	0.328	0.115	0.095	0.000							
MRV	0.098	0.241	0.236	0.066	0.060	0.101	0.000						
ABD	0.139	0.238	0.217	0.141	0.127	0.113	0.087	0.000					
PLN	0.447	0.500	0.509	0.442	0.429	0.431	0.423	0.405	0.000				
AND	0.242	0.354	0.366	0.267	0.230	0.292	0.218	0.287	0.185	0.000			
SHR	0.218	0.406	0.251	0.203	0.156	0.281	0.163	0.214	0.379	0.231	0.000		
BAG	0.403	0.481	0.499	0.362	0.329	0.375	0.362	0.381	0.109	0.116	0.316	0.000	
KHD	0.348	0.562	0.535	0.304	0.281	0.334	0.262	0.296	0.211	0.131	0.336	0.138	0.000

For an explanation of population symbols, see Table 1

4.2 Genetic Diversity of *M. minima*

There is only one previous report on the genetic diversity of *M. minima* in Iran: using 6 SSR markers, an average genetic diversity of 0.13 was reported (Falahati-Anbaran et al. 2006); by contrast, in this study, a significantly higher level of genetic diversity (0.213) was observed. The previous study was based only on 5 samples of *M. minima* in one location, whereas we included 67 individuals from 13 populations. In the study of various species of *Medicago*, Zitouna et al. (2014) reported high average values for the majority of genetic diversity parameters in *M. minima*, including N_a (6.571), N_e (4.487), %P (85.71%), H_o (0.330), H_e (0.671), however, the F value in our study was slightly higher (0.673 vs. 0.536). These differences can be attributed to the difference in sampling (Powell et al. 1996) and the number of molecular markers (Ferrer et al. 2004); additionally, differences in population sizes are another possible reason. Frequently, associations between genetic diversity in wild plant species and geographical distribution, size of the population size, life history parameters, and the breeding system have been found (Hamrick and Godt 1996; Young and Boyle 2000; Ferrer et al. 2004; Li et al. 2018). A higher level of genetic diversity might be expected for species with significant adaptability in harsh environments (Gadissa et al. 2018). However, the average level of genetic diversity of *M. minima* populations was shown by diversity indices H_o and H_e (0.076 and 0.225, respectively).

Genetic diversity is highly influenced by life form and breeding system. Outcrossing plant species tend to have higher genetic diversity and a lower differentiation (more

homogeneity) among populations (Hamrick and Godt 1996; Young et al. 2002). Inbreeding and outbreeding are present in different species of *Medicago*, with majority of annual medics being inbreeders (selfers) (Small 2011). The fact that *M. minima* is a self-pollinating species to a large extent explains the low genetic diversity and significant F_{st} among populations that were accompanied by low overall gene flow ($0.115 < 1$). Gene flow greater than 1.0 is considered strong (Slatkin 1985). By comparison, in wild populations of *M. sativa*, a strongly outcrossing species, expected heterozygosity and PIC content were 0.680 and 0.639, respectively (Qiang et al. 2015). Annual self-pollinating *Medicago* species have often been found to have a much lower level of expected heterozygosity: *Medicago rugosa* Desr. (0.0), *Medicago scutellata* Mill. (0.0), *M. truncatula* Gaertner. (0.067), *Medicago orbicularis* (L.) Bartal. (0.067), *M. minima* (0.13), *Medicago rigidula* (L.) All. (0.10) (Falahati-Anbaran et al. 2006). Nevertheless, Moroccan annual *Medicago* species, including those cited above as well as *M. murex* Willd. and *M. polymorpha* L. were observed to have high values for expected heterozygosity (Zitouna et al. 2014) similar to *M. sativa* (Qiang et al. 2015). In two similar studies to ours, Bayat et al. (2021) and Emami-Tabatabaei et al. (2021) reported expected heterozygosities of 0.373 and 0.185, respectively. It appears that various factors affecting the genetic diversity pattern among wild plant populations of *Medicago*, and perhaps the size of populations is critical: in small populations the chance for mating is significantly higher than larger populations (Moritsuka et al. 2017).

4.3 Genetic Differentiation and Population Structure

Annual predominately inbreeding populations sometimes have high levels of differentiation and variation among populations (Culley and Wolfe 2001; Honnay and Jacquemyn 2007; Stöcklin et al. 2009). Indeed, for *M. minima* the AMOVA test indicated the significant distribution of genetic diversity among populations (58%), while 42% was within populations of *M. minima*. By comparison, in wild populations of *Cicer* an ISSR-based study estimated the partitioning of total genetic diversity was 48.90% among the population and 51.10% intrapopulation (Rajesh et al. 2003). In an investigation of *M. sativa* using SSR, the outcrossing breeding system of *M. sativa* expectedly led to a distribution of about 97.6% of genetic diversity within the population (Falahati-Anbaran et al. 2007), much higher than self-pollinated annual *M. minima*.

Gene flow may predominantly affect the genetic differentiation among wild plant populations (Schaal et al. 1998). The overall value of N_m observed in this study (0.115) indicates the occurrence of genetic drift as N_m above 1 facilitates this possibility (Kang and Chung 1997). Also, the PhiPT value of 0.617 confirms the low gene flow among populations. It reflects the high genetic differentiation, similar to the observation of low PhiPT values among the Iranian *M. rigidula* populations which revealed the absence of or weak genetic differentiation (Bayat et al. 2021). Moreover, F -statistics revealed substantial genetic differentiation among individuals ($F_{is} = 0.87$) and among populations ($F_{st} = 0.589$, $p = 0.001$). This seems to reflect significant weak historical gene flow was insufficient to prohibit genetic drift and as a result, diversity among populations increases while decreasing within populations.

Pairwise genetic differentiation coefficient (F_{st}) had a range of 0.060 to 0.562 (Table 6), indicating a moderate to high population genetic differentiation. The highest genetic differentiation value of 0.562 was between Meshkinshahr, and Kohdasht populations, and the minimum was found between Marivan and Paveh. The physical distance between the first two populations is 953 km and for the second two is 114 km, which was consistent with the generalization that increase in distance can restrict the degree of gene flow between populations (Li et al. 2018). However, northwest populations generally had higher F_{st} with populations in the southwest.

4.4 Evolution of Two Molecularly Distinctive Geographical Groups in Iran

The structure analysis indicated the presence of two clusters as the optimum genetic group for *M. minima*

populations. Consistently, similar results were produced from both NJ and PCoA analyses. The studied populations appeared to fall into two groups. The groups' composition was the same between NJ and PCoA, whereas the otherwise observed was observed with those two tests and structure analysis. A moderate polarity among population grouping with these tests was observed because populations from northwest and west mainly grouped, while those from the southwest were placed in a group. It seems that the distance between populations played a considerable role in generating such a pattern for populations of *M. minima*. The Mantel test also indicated the presence of a close link between geographical distance and genetic differentiation between *M. minima* populations in this study, similar to Wang et al. (2008) report, which indicated the critical role of physical distance in genetic differentiation among populations of *Chondrus crispus* Stackhouse.

5 Conclusions

This investigation is the first onto examine genetic variation of *M. minima* populations using SSR markers. Remarkably limited or absent genetic diversity within 13 populations was observed. Nevertheless, substantial genetic differentiation was observed among the populations. This pattern is probably due to the breeding system and physical distance separating the populations. This study is a model for additional examination of *M. minima* employing larger numbers of populations and SSRs. The data presented are valuable for conservation and breeding projects. The most significant finding was the discovery of two geographically differentiated groups based on SSR markers. These have not yet been carefully analysed for possible corresponding morphological differences, but they suggest very early speciation.

Authors' Contributions Conceptualization: IM, ES; Data curation: ZB and IM; Formal analysis: ZB and IM; Investigation: ZB; Methodology: IM, ZB, and MA; Project administration: IM; Resources: ZB and IM; Software: ZB and IM; Supervision: IM and MA; Validation: IM, MA, and ES; Visualization: ZB; Writing—original draft: ZB, ES and IM; Writing—review and editing: IM, ES. All authors read and approved the final manuscript.

Funding Authors did not receive any external support for this study.

Data Availability The data that support the findings of this study are available from the corresponding author, IM, upon reasonable request.

Declarations

Conflict of interests The authors declare that they have no conflict of interest.

Ethical Approval Not applicable since this study did not involve human participants or material.

References

- Bayat M, Assadi M, Small E, Mehregan I (2021) Molecular studies of Iranian populations support the morphology-based taxonomic separation of *Medicago rigidula* and *M. rigiduloides*. *Phytotaxa* 518(4):281–299
- Beech E, Rivers M, Oldfield S, Smith P (2017) GlobalTreeSearch: the first complete global database of tree species and country distributions. *J Sustain* for 36:454–489. <https://doi.org/10.1080/10549811.2017.1310049>
- Belicuas PR, Guimarães CT, Paiva LV, Duarte JM, Maluf WR, Paiva E (2007) Androgenetic haploids and SSR markers as tools for the development of tropical maize hybrids. *Euphytica* 156:95–102. <https://doi.org/10.1007/s10681-007-9356-z>
- Bena G, Lyet A, Huguet T, Olivieri I (2005) *Medicago*-Sinorhizobium symbiotic specificity evolution and the geographic expansion of *Medicago*. *J Evol Biol* 18:1547–1558
- Culley TM, Wolfe AD (2001) Population genetic structure of the cleistogamous plant species *Viola pubescens* Aiton (Violaceae), as indicated by allozyme and ISSR molecular markers. *Heredity* 86:545–556. <https://doi.org/10.1046/j.1365-2540.2001.00875.x>
- Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. *Focus* 12:13–15
- Earl DA (2012) Structure harvester: a website and program for visualizing structure output and implementing the Evanno method. *Conserv Genet Resour* 4:359–361. <https://doi.org/10.1007/s12686-011-9548-7>
- Emami-Tabatabaei SS, Small E, Assadi M, Dehshiri MM, Mehregan I (2021) Genetic variation among Iranian *Medicago polymorpha* L. populations based on SSR markers. *Genet Resour Crop Evol* 68:1411–1424. <https://doi.org/10.1007/s10722-020-01071-7>
- 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):422. <https://doi.org/10.1007/s00122-003-1450-6>
- Falahati-Anbaran M, Habashi A, Esfahany M, Mohammadi S, Ghareyazie B (2006) Study of genetic diversity and relationships of diploid and tetraploid annual medics using microsatellite markers. *J Sci Tech Agri Nat Resour* 10:349–359
- Falahati-Anbaran M, Habashi A, Esfahany M, Mohammadi S, Ghareyazie B (2007) Population genetic structure based on SSR markers in alfalfa (*Medicago sativa* L.) from various regions contiguous to the centres of origin of the species. *J Genet* 86:59–63. <https://doi.org/10.1007/s12041-007-0008-9>
- Farang MA, Huhman DV, Lei Z, Sumner LW (2007) Metabolic profiling and systematic identification of flavonoids and isoflavonoids in roots and cell suspension cultures of *Medicago truncatula* using HPLC–UV–ESI–MS and GC–MS. *J Phytochem* 68:342–354. <https://doi.org/10.1016/j.phytochem.2006.10.023>
- Ferrer MM, Eguarte LE, Montaña C (2004) Genetic structure and outcrossing rates in *Flourensia cernua* (Asteraceae) growing at different densities in the South-western Chihuahuan Desert. *Ann Bot* 94:419–426. <https://doi.org/10.1093/aob/mch159>
- Freeland JR (2020) *Molecular ecology*, 3rd edn. Wiley-Blackwell, London, p 363
- Gadissa F, Tesfaye K, Dagne K, Geleta M (2018) Genetic diversity and population structure analyses of *Plectranthus edulis* (Vatke) Agnew collections from diverse agro-ecologies in Ethiopia using newly developed EST-SSRs marker system. *BMC Genet* 19:92. <https://doi.org/10.1186/s12863-018-0682-z>
- Hamrick JL, Godt MW (1996) Effects of life history traits on genetic diversity in plant species. *Philos Trans R Soc Lond B Biol Sci* 351:1291–1298. <https://doi.org/10.1098/rstb.1996.0112>
- Heyn C (1984) *Medicago*. In: Rechinger KH (ed) *Flora Iranica*. Akademische Druck- u. Verlagsanstalt, Graz, pp 253–271
- Honnay O, Jacquemyn H (2007) Susceptibility of common and rare plant species to the genetic consequences of habitat fragmentation. *Conserv Biol* 21:823–831. <https://doi.org/10.1111/j.1523-1739.2006.00646.x>
- Humphries AW, Ovalle C, Hughes S, del Pozo A, Inostroza L, Barahona V, Yu L, Yerzhanova S, Rowe T, Hill J, Meirman G, Abayev S, Brummer EC, Peck DM, Toktarbekova S, Kalibayev B, Espinoza S, Ivelic-Saez J, Bingham E, Small E, Kilian B (2020) Characterization and pre-breeding of diverse alfalfa wild relatives originating from drought-stressed environments. *Crop Sci* 61:69–88
- Jha S, Pal A (1992) Seed structure in *Medicago* L. *Flora* 187:209–218. [https://doi.org/10.1016/S0367-2530\(17\)32224-7](https://doi.org/10.1016/S0367-2530(17)32224-7)
- Judd W, Campbell C, Kellogg E, Stevens P, Donoghue M (2008) *Plant systematics: a phylogenetic approach*. Sinauer Associates, Massachusetts
- Kang SS, Chung MG (1997) Genetic variation and population structure in Korean endemic species: IV. *Hemerocallis hakunensis* (Liliaceae). *J Plant Res* 110:209–217. <https://doi.org/10.1007/BF02509309>
- Keivani M, Ramezani-pour S, Soltanloo H, Choukan R, Naghavi M, Ranjbar M (2010) Genetic diversity assessment of alfalfa (*Medicago sativa* L.) populations using AFLP markers. *Aust J Crop Sci* 4:491
- Köressaar T, Lepamets M, Kaplinski L, Raime K, Andreson R, Remm M (2018) Primer3_masker: integrating masking of template sequence with primer design software. *Bioinformatics* 34:1937–1938. <https://doi.org/10.1093/bioinformatics/bty036>
- Küster H (2013) *Medicago truncatula*. In: Maloy S, Hughes K (eds) *Brenner's encyclopedia of genetics*, 2nd edn. Academic Press, pp 335–337
- Li X, Li M, Hou L, Zhang Z, Pang X, Li Y (2018) De novo transcriptome assembly and population genetic analyses for an endangered Chinese endemic *Acer miaotaiense* (Aceraceae). *Genes* 9:378. <https://doi.org/10.3390/genes9080378>
- Liu ZP, Liu GS, Yang QC (2007) A novel statistical method for assessing SSR variation in autotetraploid alfalfa (*Medicago sativa* L.). *Genet Mol Biol* 30(2):385–391
- Magoč T, Salzberg SL (2011) FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27:2957–2963. <https://doi.org/10.1093/bioinformatics/btr507>
- Mantel N (1967) The detection of disease clustering and a generalized regression approach. *Cancer Res* 27:209–220
- Mehregan I, Rahiminejad MR, Azizian D (2002) A taxonomic revision of the genus *Medicago* in Iran. *Iran J Bot* 9:207–221
- Miller MP (1997) Tools for population genetic analysis (TFPGA) 1.3: a Windows program for the analysis of allozyme and molecular population genetic data. Computer software distributed by the author
- Moritsuka E et al (2017) Genetic variation and population structure of a threatened timber tree *Dalbergia cochinchinensis* in Cambodia. *Tree Genet* 13:115. <https://doi.org/10.1007/s11295-017-1199-8>
- Nei M (1973) Analysis of gene diversity in subdivided populations. *Proc Nat Acad Sci* 70:3321–3323. <https://doi.org/10.1073/pnas.70.12.3321>
- Nybom H, Esselink G, Werlemark G, Leus L, Vosman B (2006) Unique genomic configuration revealed by microsatellite DNA in polyploid dogroses. *Rosa Sect Caninae J Evol Biol* 19:635–648. <https://doi.org/10.1111/j.1420-9101.2005.01010.x>
- Parsamanesh S, Sadeghi H (2019) The phytoremediation effect of *Medicago scutellata* (L.) Mill. on soils under Cd–water stress: a

- good choice for contaminated dry lands. *Environ Sci Poll Res* 26:29065–29073. <https://doi.org/10.1007/s11356-019-05989-z>
- Peakall R, Smouse PE (2006) Genalex 6: genetic analysis in excel. Population genetic software for teaching and research. *Mol Ecol Notes* 6:288–295. <https://doi.org/10.1111/j.1471-8286.2005.01155.x>
- Porrás-Hurtado L, Ruiz Y, Santos C, Phillips C, Carracedo Á, Lareu M (2013) An overview of structure: applications, parameter settings, and supporting software. *Front Genet* 4:98. <https://doi.org/10.3389/fgene.2013.00098>
- Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S, Rafalski A (1996) The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Mol Breed* 2:225–238. <https://doi.org/10.1007/BF00564200>
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959
- Qiang H, Chen Z, Zhang Z, Wang X, Gao H, Wang Z (2015) Molecular diversity and population structure of a worldwide collection of cultivated tetraploid alfalfa (*Medicago sativa* subsp. *sativa* L.) germplasm as revealed by microsatellite markers. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0124592>
- Rajesh P, Sant V, Gupta V, Muehlbauer F, Ranjekar P (2003) Genetic relationships among annual and perennial wild species of *Cicer* using inter simple sequence repeat (ISSR) polymorphism. *Euphytica* 129:15–23. <https://doi.org/10.1023/A:1021567821141>
- Rashidi M, Farshadfar M, Safari H, Shirvani H (2013) Utility of ISSR molecule marker in examine of genetic diversity 17 genotypes of perennial alfalfa (*Medicago sativa*). *J Nov Appl Sci* 2:969–973
- Schaal B, Hayworth D, Olsen KM, Rauscher J, Smith W (1998) Phylogeographic studies in plants: problems and prospects. *Mol Ecol* 7:465–474. <https://doi.org/10.1046/j.1365-294x.1998.00318.x>
- Schlötterer C, Vogl C, Tautz D (1997) Polymorphism and locus-specific effects on polymorphism at microsatellite loci in natural *Drosophila melanogaster* populations. *Genetics* 146:309–320. <https://doi.org/10.1046/j.1365-294x.1998.00382.x>
- Silva DC, Dos Santos J, de Souza Barbosa GV, Almeida C (2012) DNA fingerprinting based on simple sequence repeat (SSR) markers in sugarcane clones from the breeding program RIDESA. *Afr J Biotechnol*, 11:4722–4728. <https://doi.org/10.5897/AJB11.2614>
- Slatkin M (1985) Rare alleles as indicators of gene flow. *Evolution* 39:53–65. <https://doi.org/10.1111/j.1558-5646.1985.tb04079.x>
- Small E (2011) Alfalfa and relatives: Evolution and classification of *Medicago*. NRC Research Press, Ottawa
- Stöcklin J, Kuss P, Pluess AR (2009) Genetic diversity, phenotypic variation and local adaptation in the alpine landscape: case studies with alpine plant species. *Bot Helv* 119:125–133. <https://doi.org/10.1007/s00035-009-0065-1>
- Takezaki N, Nei M, Tamura K (2010) POPTREE2: Software for constructing population trees from allele frequency data and computing other population statistics with Windows interface. *Mol Biol Evol* 27:747–752. <https://doi.org/10.1093/molbev/msp312>
- Walsh MJ, Delaney RH, Groose RW, Krall JM (2001) Performance of annual medic species (*Medicago* spp.) in Southeastern Wyoming. *Agron J* 93:1249–1256. <https://doi.org/10.2134/agronj2001.1249>
- Wang X, Zhao F, Hu Z, Critchley AT, Morrell SL, Duan D (2008) Inter-simple sequence repeat (ISSR) analysis of genetic variation of *Chondrus crispus* populations from North Atlantic. *Aquat Bot* 88:154–159. <https://doi.org/10.1016/j.aquabot.2007.10.001>
- Werlemark G, Uggla M, Nybom H (1999) Morphological and RAPD markers show a highly skewed distribution in a pair of reciprocal crosses between hemisexual dogrose species *Rosa Sect Caninae*. *Theor Appl Genet* 98:557–563. <https://doi.org/10.1007/s001220051104>
- Wright S (1978) The relation of livestock breeding to theories of evolution. *J Anim Sci* 46:1192–1200. <https://doi.org/10.2527/jas1978.4651192x>
- Xu Y, Crouch JH (2008) Marker-assisted selection in plant breeding: from publications to practice. *Crop Sci* 48:391–407. <https://doi.org/10.2135/cropsci2007.04.0191>
- Young AG, Boyle TJ (2000) Forest fragmentation. In: Young A, Boshier D, Boyle T (eds) *Forest Conservation Genetics: principles practice*. CABI Publishing, Wallingford, pp 123–134
- Young A, Hill J, Murray B, Peakall R (2002) Breeding system, genetic diversity and clonal structure in the sub-alpine forb *Rutidosis leiolepis* F. Muell. (Asteraceae). *Biol Conserv* 106:71–78. [https://doi.org/10.1016/S0006-3207\(01\)00230-0](https://doi.org/10.1016/S0006-3207(01)00230-0)
- Zitouna N, Marghali S, Gharbi M, Haddioui A, Trifi-Farah N (2014) Sequence divergence of microsatellites for phylogeographic assessment of Moroccan *Medicago* species. *Genet Mol Res* 13:1548–1562. <https://doi.org/10.4238/2014.March.12.7>