Identification of Lens Cultivars in Market by Molecular Tools: DNA Barcoding and SSRs

Metin Burak TATLISES

Trakya University

Semra Hasancebi (✉ semrahasancebi@trakya.edu.tr)

Trakya University <https://orcid.org/0000-0003-3898-7413>

Research Article

Keywords: food traceability, DNA barcoding, SSR, Lentil (*Lens culinaris* Medik.)

Posted Date: January 10th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1202399/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Substitution of plant cultivars of high commercial value with a cheaper, lower-quality one is a common fraud committed against consumers and producers. Since it is one of the most widely grown legumes, lentil (*Lens culinaris* Medik.) are a suitable for such frauds. This study aimed to identify lentil cultivars which are registered and authorized in the market in Turkey by using current molecular methods. For this purpose, 26 lentil cultivars were analyzed for 15 SSR markers and 2 DNA barcode loci (*trnH-psbA* and *matK*). A high allele diversity was observed by 12 scorable SSR markers, and the average number of alleles was determined to be 16. One of the important findings was the presence of “cultivar-specific alleles” that can be used to identify each cultivar in the lentil market in Turkey. At least one “cultivar-specific allele” was obtained for each cultivar. The lentil cultivars were analyzed in terms of 2 DNA barcode regions as *trnH-psbA* and *matK*. Sequences that could identify 14 of the 26 cultivars were obtained. While the rate of the intra-species variation for the *trnH-psbA* region was observed to be low, a higher rate was found for *matK*. Nevertheless, it was observed that intra-species discrimination can be made more effective when both loci are used together. We expect that the results of this study, especially the cultivar-specific SSR alleles and DNA barcoding sequence data may be used routinely to identify on production and packaged products that are commercially available in markets.

1. Introduction

In Turkey and in the world in general, the seed sector is developing rapidly. The ever-increasing nutritional demand of the world's population makes it inevitable to develop new cultivars with high quality and yield. This has led to a rapid increase in the number of cultivars in plant species that have economic value. Lentil (*Lens culinaris* Medik.) contains high levels of vegetable protein and is commonly found on our tables. For this reason, lentil is one of the most prevalently consumed legume species that is constantly being developed. Due to the fact that Turkey is the gene center of lentils and has important cultivation areas, there are many different cultivars of lentils in the market. The yield and quality potentials of all lentil cultivars differ from each other. Therefore, it is extremely important to identify the cultivars to minimize cheating/confusions in both the use of cultivars suitable for soil/climatic conditions that are demanded in production and in trade while choosing parents in breeding activities.

All food products must comply with the description provided by the manufacturers or processors with reference to the origin of the ingredients, as well as the identity of the species, breeds or cultivars used. Substitution of plant cultivars of high commercial value with a cheaper, lower-quality one is a common fraud committed against consumers and producers. Such fraud causes confusion in the market, disaffection towards genuine products and price fluctuation. DNA-based methods are more reliable and used for routine analyses to maintain food safety and quality (Martins-Lopes et al. 2013). Additionally, DNA-based methods offer tools to ensure food authenticity and traceability of primary products entering food chains in terms of both fresh and processed food (Böhme et al. 2019)

SSR markers are highly polymorphic markers that have been useful for identifying species and cultivars of both raw materials and processed food. The high level of polymorphism is due to different numbers of repeats in the microsatellite loci that are distributed to the entire genome. Therefore, SSRs are amenable to high throughput genotyping and also a useful tool for paternity analysis, construction of high-density genome maps, marker-assisted selection and for establishing genetic and evolutionary relationships (Kalia et al. 2011).

DNA barcoding is another increasingly used molecular technique that analyzes one or few standardized loci for identifying species or cultivars. The mitochondrial gene Cytochrome c oxidase subunit 1 (*cox1* or *COI*) was proposed as a DNA barcode for identification of animal species by Hebert et al. 2003. However, mitochondrial DNA barcode candidates are useless for plants because plant mitochondrial sequences evolve slowly (Mower et al. 2007). Therefore, the attention of researchers is focused on plastid genomes. Plastid genes (*rpoC1*, *rpoB*, *matK*, *rbcl*), plastid intergenic spacers (*atpF-atpH*, *trnH-psbA* and *psbK-psbI*) and the internal transcribed spacer region (*ITS*) have been initially proposed as candidate barcoding loci (Kress et al. 2005; Chase et al. 2007; Fazekas et al. 2008). The CBOL Plant Working Group (2009), recommended a two-locus combination of *rbcl* + *matK* as the core barcode for land plants. Accordingly, it has been reported that the *rbcl*, *matK* and *trnH-psbA* regions have a high level of distinctive characteristics between species when used together (Kress et al. 2007; Group et al. 2009; Hollingsworth et al. 2011). In comparison to other plant barcode loci, *rbcl* has a low mutation frequency. However, it is informative for the intra-species level. *matK* is one of the most rapidly evolving plastid regions and shows high levels of discrimination among angiosperm species (Hilu and Liang 1997; Fazekas et al. 2008). The presence of the highly conserved coding sequences of *trnH-psbA* makes the design of universal primers feasible with a single primer pair likely to amplify nearly all angiosperms (Shaw et al. 2005; Shaw et al. 2007). *trnH-psbA* exhibits the most sequence divergence and has high rates of insertion/deletion (Kress and Erickson 2007).

In this study, we identified registered lentil cultivars in Turkey with two commonly recommended plant DNA barcoding loci (*matK* and *trnH-psbA*) and 15 SSR markers, for ensuring the traceability of lentil cultivars in the market.

2. Material And Methods

2.1. Plant material and DNA isolation

In this study, 26 lentil cultivars that have a production permit in Turkey and one Canadian cultivar which is one of most frequently imported cultivars by Turkey were used (Table 1). Total genomic DNAs (gDNA) were isolated from a single lentil seed by a Plant/Fungi DNA Isolation Kit (Norgen).

Table 1
List of lentil cultivars used in this study

Material No.	Cultivar Name	Type	Material No.	Cultivar Name	Type
1	Çiftçi	R	14	Sazak	R
2	Özbek	R	15	Kayı	G
3	Kafkas	R	16	İpek	R
4	Fırat-87	R	17	Orhas-2019	G
5	Altıntoprak	R	18	Şanlıbey	R
6	Meyveci-2001	G	19	Atacan	R
7	Sultan-1	G	20	Yazlık Yeşil	G
8	Ankara Yeşili	G	21	Emre-20	R
9	Ceren	G	22	Mansur	R
10	Bozok	G	23	Canadian Cultivar	G
11	Gümrah	G	24	Şahan	G
12	Karagül	G	25	Eva-2017	R
13	Yerli Kırmızı	R	26	Yürekli	R
(R: red type, G: green type)					

2.2. SSR Analysis

For genotypic analysis, 15 SSR markers with high PIC values were selected among 149 SSRs which were developed by Andeden et al. (2015) (Table 2). Amplification of each SSR marker locus was carried out by PCR, which was performed in a 20-µL reaction mixture containing 60 ng DNA, 1X PCR buffer, 2.5 mM MgCl₂, 0.2 mM dNTPs, 0.5 µM primers and 1 U Taq polymerase (Invitrogen). Amplification was performed in a T100 Bio-Rad thermal cycler (CA, USA) by following the cycles of 5 min at 94°C for initial denaturation, and 35 times 45 sec at 94°C, 45 sec at 50-65°C (depending on the primer T_m), 1 min at 72°C and 10 min at 72°C for final extension. The amplified PCR products were separated and analyzed by using an AATI Fragment Analyzer System (Advance Analytic, IA, USA). The PCR products were prepared for the capillary system according to the manufacturer's instructions (DNF-905 dsDNA Reagent, IA, USA). Each reaction mixture was diluted at a 1:5 ratio by a dilution buffer, and 24 µL of the mixture was transferred to a 96-well plate. Each well was covered by mineral oil, and electrophoresis was performed by applying 9.0 kV for 80 min. For sizing of the SSR alleles, 1-500 bp DNA ladders were used in each run, and DNA fragment sizes were calculated by the system software ProSize 3.0. Each allelic DNA fragment produced from the SSR loci was scored for statistical analysis.

Table 2
List of highly polymorphic SSR markers

No	SSR	Repeat Motifs
1	CULA105	(TA)7(TG)22
2	CULA109	(TG)15A(GA)29
3	CULA211	(GT)23(GA)18
4	CULA308	(TC)20A(CA)6
5	CULA408	(CA)11
6	CULA413B	(AC)14
7	CULB7	(CT)7
8	CULB9	(CT)24
9	CULB107	(CT)28
10	CULB206	(CA)17(CA)6
11	CULB217	(CT)31
12	CULB222	(GA)28
13	CULB310	(TC)18
14	CULB418	(GA)28
15	CULB423	(TC)6

2.3. SSR data analysis

The DNA fragment information observed for each SSR locus in each cultivar by capillary electrophoresis was statistically analyzed using the GenAEx 6.5 program (Peakall and Smouse 2006, 2012). Genetic distance and similarity matrices were created using the DARwin 6.0 program (Perrier and Jacquemoud-Collet 2006), and the genetic relationship between the cultivars was revealed by drawing a phylogenetic tree using the UPGMA method (Sneath and Sokal 1973).

2.4. DNA barcoding

matK and *trnH-psbA* barcode loci in the chloroplast genome were used as the barcode regions in this study. The Universal primers used in the amplification of the barcode regions are given in Table 3. Amplification of the barcode loci was carried out by PCR, which was performed in a 40-μL reaction mixture containing 60 ng gDNA, 1X PCR buffer, 2.5 mM MgCl₂, 0.2 mM dNTPs, 0.5 μM primers and 1 U Taq polymerase (Invitrogen). Amplification was performed following the cycles; 5 min at 94°C for initial denaturation, and 35 times 45 sec at 94°C, 45 sec at 50°C for *matK*, 55°C for *trnH-psbA*, 1 min at 72°C and 10 min at 72°C for final extension. The PCR products were analyzed by gel electrophoresis on 1.5%

agarose gels in 0.5X TBE buffer stained with ethidium bromide and checked under ultraviolet light. DNA fragments of *matK* were cut from the gel and cleaned using the GeneJET Gel Extraction kit (Thermo Scientific), and the purified PCR products were used as a template for sequence analysis.

Table 3
List of DNA barcoding loci and universal primers († HPLC grade purification)

Locus	Primer	Sequence 5' - 3'	Reference
<i>trnH-psbA</i>	psbA3_f	GTT ATG CAT GAA CGT AAT GCT C	23
	trnHf_05r	CGC GCA TGG TGG ATT CAC AAT CC	24
<i>trnH-psbA</i>	†trnH-psbA.F	ACT GCC TTG ATC CAC TTG GC	25
	†trnH-psbA.R	CGA AGC TCC ATC TAC AAA TGG	
<i>matK</i>	matK2.1F	CCT ATC CAT CTG GAA ATC TTA G	10
	matK_5R	GTT CTA GCA CAA GAA AGT CG	
<i>matK</i>	†matK472.F	CCC RTY CAT CTG GAA ATC TTG GTT C	8
	†matK1248.R	GCT RTR ATA ATG AGA AAG ATT TCT GC	26
	†matK-1FKIM.R	AAT ATC CAA ATA CCA AAT CC	Ki-Joong Kim, unpublished

2.5. Barcoding data analysis

The sequences were manually edited and aligned by MEGA X (Kumar et al. 2018). The alignment of each region was performed by ClustalW (Thompson et al. 1994). A phylogenetic tree was created according to the UPGMA method (Sneath and Sokal 1973) to determine the phylogenetic relationships between the cultivars by MEGA X (Kumar et al. 2018). Numbers of groups, as well as group and nucleotide diversity, were determined by using the DnaSP 6 software (Rozas et al. 2017).

3. Results And Discussion

3.1. Results of SSRs

In this study, 26 lentil cultivars were genotyped by 12 SSR markers and 2 DNA barcoding loci. Three SSR markers (CULA413B, CULA107 and CULA109) did not produce scorable PCR products in all cultivars. Therefore, they were not used for genotyping analysis. The alleles of the SSRs were amplified by PCR and analyzed with an AATI Fragment Analyzer. The 12 successful SSR markers were found highly polymorphic, and 172 alleles in total were obtained from 26 cultivars. In the total assessment, the average number of alleles per SSR locus was calculated as 14. The most frequently observed alleles and their frequencies are presented in Table 4. The most polymorphic markers with 19 alleles among 172

were seen in the CULB310 marker. The distribution of the alleles belonging to each SSR marker, the frequencies of the most frequently observed alleles, their loci among the 26 cultivars that were scanned and their PIC values are given in Table 4.

Table 4
Allelic diversity information of SSRs

No	SSR	Number of Alleles	Min. Allele (bp)	Max. Allele (bp)	Max. Observed		PIC
					Allele	Freq.	
1	CULA105	11	142	178	158, 178	0.173	0.868
2	CULA211	17	180	274	258	0.134	0.915
3	CULA308	13	230	294	294	0.211	0.863
4	CULA408	15	144	356	154	0.250	0.837
5	CULB7	16	208	260	216	0.153	0.911
6	CULB9	13	180	220	188	0.115	0.904
7	CULB206	18	202	264	244, 264	0.115	0.928
8	CULB217	9	146	170	146	0.230	0.857
9	CULB222	17	126	184	136	0.192	0.909
10	CULB310	19	266	326	272	0.346	0.836
11	CULB418	9	220	274	224	0.346	0.770
12	CULB423	15	218	266	250, 260	0.115	0.914

Detection of alleles that are unique for each cultivar is one of the significant findings of SSR analysis. With the 12 SSRs used in the scope of this study, unique cultivar-specific alleles were obtained. The Özbek cultivar had 8 unique, cultivar-specific alleles. These alleles, which are especially important for cultivar recognition/determination, can create an alternative solution for problems encountered in both lentil production and lentil seed trade. The obtained cultivar-specific alleles are presented in Table 5. Additionally, a dendrogram (Figure 1) was created using the DARwin 6.0 program to reveal the phylogenetic relationships of the cultivars using the UPGMA method, which was performed based on the genetic similarity and distance between the cultivars according to the report by Sneath and Sokal (1973).

Table 5
List of cultivar-specific alleles

Cultivar	SSR Locus	Allele	Cultivar	SSR Locus	Allele
Çiftçi	CULA308	260	Sultan-1	CULB222	160
Çiftçi	CULB7	240	Sultan-1	CULB310	268
Özbek	CULA211	268	Sultan-1	CULB418	236
Özbek	CULA408	144	Ceren	CULA211	274
Özbek	CULB7	212	Ceren	CULA408	178
Özbek	CULB206	248	Ceren	CULB9	190
Özbek	CULB222	164	Ceren	CULB206	260
Özbek	CULB310	266	Bozok	CULB7	217
Özbek	CULB310	286	Bozok	CULB222	178
Özbek	CULB423	238	Bozok	CULB423	230
Kafkas	CULA105	166	Gümrah	CULB7	260
Kafkas	CULA211	256	Gümrah	CULB222	170
Kafkas	CULA308	282	Gümrah	CULB310	326
Kafkas	CULB9	206	Karagül	CULB222	184
Kafkas	CULB206	216	Yerli Kırmızı	CULB222	130
Kafkas	CULB222	145	Yerli Kırmızı	CULB418	266
Kafkas	CULB418	270	Sazak	CULA105	154
Fırat	CULB7	224	Sazak	CULA308	240
Fırat	CULB206	214	Sazak	CULB222	168
Altıntoprak	CULB7	248	Kayı	CULB7	258
Altıntoprak	CULB206	258	Kayı	CULB206	240
Meyveci-2001	CULA105	142	Kayı	CULB222	180
Meyveci-2001	CULA211	240	Kayı	CULB310	320
Meyveci-2001	CULB222	134	İpek	CULA211	202
Meyveci-2001	CULB310	274	İpek	CULB222	154
Meyveci-2001	CULB310	288	İpek	CULB310	296

Ankara Yeşili	CULB222	162	İpek	CULB423	222
Orhas 2019	CULA308	258	Mansur	CULA211	264
Orhas 2019	CULA408	168	Mansur	CULB7	250
Orhas 2019	CULA408	324	Canadian V.	CULA308	252
Orhas 2019	CULB9	220	Canadian V.	CULA408	158
Orhas 2019	CULB206	250	Canadian V.	CULB206	204
Orhas 2019	CULB423	252	Şahan	CULA308	232
Şanlıbey	CULA308	230	Şahan	CULA408	150
Şanlıbey	CULB310	306	Şahan	CULB206	202
Atacan	CULA308	234	Şahan	CULB418	268
Atacan	CULB7	254	Şahan	CULB423	266
Atacan	CULB9	198	Eva 2017	CULB7	246
Atacan	CULB310	300	Eva 2017	CULB206	238
Yazlıkyeşil	CULA211	200	Yürekli	CULB206	206
Emre-20	CULB206	234	Yürekli	CULB423	258
Emre-20	CULB423	242			

3.2 DNA barcoding

In addition to SSRs, 2 barcode regions were used for genotyping the lentil cultivars. Both 2 barcode loci were successfully reproduced by PCR in 26 lentil cultivars with all primer pairs (Table 3). However, the sequence analysis of the PCR products did not show the same success. Nevertheless, the *trnH-psbA* and *matK* loci were successfully sequenced with high purity and different universal primers (Table 3) after cloning.

In this study, **trnH-psbA_F/trnH-psbA_R** was the most successful primer pair for both amplification and sequencing. The amplified *trnH-psbA* barcode locus was ~350 bp, and it was directly sequenced and compared among 26 cultivars. For the *matK* barcode locus, the **matK472.F/matK-1FKIM.R** primer pair produced 1000 bp DNA fragments which were successfully sequenced by cloning. *matK* sequences were compared among 26 cultivars by the MEGA software. The number of groups, group diversity and nucleotide diversity results of the sequences belonging to the *matK* and *trnH-psbA* loci are given at Table 6. According to the results, the highest group number for the cultivars was obtained with the combination of the *trnH-psbA* + *matK* regions. A phylogenetic tree was constructed with the sequencing data of the *trnH-psbA* + *matK* regions for the analyzed lentil cultivars (Figure 2). Each branch of the phylogenetic tree

is associated with groups. As demonstrated in Figure 2, the Çiftçi-Kafkas-Karagül, Şanlıbey-Emre 20, Sazak-Yürekli-Sultan 1-Şahan and Özbek-Ceren-Kayı cultivars were in the same branch as a result of a similar sequence. On the other hand, other red and green cultivars had unique sequences, and each cultivar fell into a different branch. The dendrogram based on sequence data (Figure 2) proved that the combination of *trnH-psbA* + *matK* could discriminate 14 lentil cultivars in the market.

Table 6
Number of groups, group diversity and nucleotide diversity of *matK*, *trnH-psbA* and *matK* + *trnH-psbA* sequences

	<i>trnH-psbA</i>	<i>matK</i>	<i>trnH-psbA+matK</i>
Number of Groups	7	14	18
Group Diversity	0.680	0.917	0.960
Nucleotide Diversity	0.00734	0.00292	0.00364

4. Conclusion

In conclusion, some lentil cultivars which are used commercially in Turkey were identified with the SSR marker and DNA barcoding methods. As a result of the SSR analysis, "cultivar-specific" alleles that can be used for identification of 26 lentil cultivars were obtained. As a result of the DNA Barcoding process, sequencing data that could identify 14 lentil cultivars were found. It was observed that the *matK* locus was more successful in intra-species identification, but the combined usage of the barcoding loci would increase the success in the discrimination of cultivars. The sequence and SSR data of the Canadian cultivar can be used as an identification tool when necessary, during import. Furthermore, the data on the Turkish cultivars that are reported in this study will protect the manufacturer, the consumer and even the vendor. Still, the difficulties in sequence analysis and the need for cloning for the *matK* locus in this study are thought-provoking for the effective use of the DNA barcoding method as a tool to identify foods.

Declarations

Declarations

Conflicts of interest

No conflicts of interests or competing interests.

Ethical approval

Not applicable.

Funding

This research was financially supported by Scientific and Technological Research Council of Turkey (TUBITAK). (Project No. #2180172)

Author contributions

SH & MBT contributed to the design of the study. Seed collection and laboratory analysis were performed by MBT. Data processing and analysis were performed by SH & MBT. MBT wrote the first draft of the manuscript. SH edited it and approved the final draft.

Acknowledgements

The authors wish to thank Tarla Bitkileri Merkez Araştırma Enstitüsü Müdürlüğü (Ankara), GAP Uluslararası Tarımsal Araştırma ve Eğitim Merkezi Müdürlüğü (Diyarbakır), Geçit Kuşağı Tarımsal Araştırma Enstitüsü Müdürlüğü (Eskişehir), Tarım İşletmeleri Genel Müdürlüğü (Ankara) for providing us the plant material.

References

1. Andeden EE, Baloch FS, Çakır E, et al (2015) Development, characterization and mapping of microsatellite markers for lentil (*Lens culinaris* Medik.). *Plant Breeding* 134:589–598. <https://doi.org/10.1111/pbr.12296>
2. Böhme K, Calo-Mata P, Barros-Velázquez J, Ortea I (2019) Review of Recent DNA-Based Methods for Main Food-Authentication Topics. *Journal of Agricultural and Food Chemistry* 67:3854–3864. <https://doi.org/10.1021/acs.jafc.8b07016>
3. Chase MW, Cowan RS, Hollingsworth PM, et al (2007) A proposal for a standardised protocol to barcode all land plants. *TAXON* 56:295–299. <https://doi.org/10.1002/tax.562004>
4. Fazekas AJ, Burgess KS, Kesanakurti PR, et al (2008) Multiple Multilocus DNA Barcodes from the Plastid Genome Discriminate Plant Species Equally Well. *PLoS ONE* 3:e2802. <https://doi.org/10.1371/journal.pone.0002802>
5. Hamilton MB, (1999) Four primer pairs for the amplification of chloroplast intergenic regions with intraspecific variation. *Molecular ecology*, 8(3), 521–523.
6. Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications through DNA barcodes. *Proceedings Biological sciences* 270:313–21. <https://doi.org/10.1098/rspb.2002.2218>
7. Hilu KW, Liang (1997) The matK gene: sequence variation and application in plant systematics. *American Journal of Botany* 84:830–839. <https://doi.org/10.2307/2445819>
8. Hollingsworth PM, Forrest LL, Spouge JL, et al (2009) A DNA barcode for land plants. *Proceedings of the National Academy of Sciences* 106:12794–12797. <https://doi.org/10.1073/pnas.0905845106>

9. Hollingsworth PM, Graham SW, Little DP (2011) Choosing and Using a Plant DNA Barcode. *PLoS ONE* 6:e19254. <https://doi.org/10.1371/journal.pone.0019254>
10. Kalia RK, Rai MK, Kalia S, et al (2010) Microsatellite markers: an overview of the recent progress in plants. *Euphytica* 177:309–334. <https://doi.org/10.1007/s10681-010-0286-9>
11. Kress WJ, Erickson DL (2007) A Two-Locus Global DNA Barcode for Land Plants: The Coding *rbcl* Gene Complements the Non-Coding *trnH-psbA* Spacer Region. *PLoS ONE* 2:e508. <https://doi.org/10.1371/journal.pone.0000508>
12. Kress WJ, Wurdack KJ, Zimmer EA, et al (2005) Use of DNA barcodes to identify flowering plants. *Proceedings of the National Academy of Sciences of the United States of America* 102:8369–8374. <https://doi.org/10.1073/pnas.0503123102>
13. Kumar S, Stecher G, Li M, et al (2018) MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular Biology and Evolution* 35:1547–1549. <https://doi.org/10.1093/molbev/msy096>
14. Martins-Lopes P, S G, Pereira L, Guedes-Pinto H (2013) Molecular markers for food traceability. *Food Technology and Biotechnology* 51:198–207
15. Mower JP, Touzet P, Gummow JS, et al (2007) Extensive variation in synonymous substitution rates in mitochondrial genes of seed plants. *BMC Evolutionary Biology* 7:135. <https://doi.org/10.1186/1471-2148-7-135>
16. PEAKALL R, SMOUSE PE (2006) genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* 6:288–295. <https://doi.org/10.1111/j.1471-8286.2005.01155.x>
17. Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* 28:2537–2539. <https://doi.org/10.1093/bioinformatics/bts460>
18. Perrier X, Jacquemoud-Collet JP (2006) DARwin - Dissimilarity Analysis and Representation for Windows. In: *darwin.cirad.fr*. <http://darwin.cirad.fr/>. Accessed 7 Dec 2021
19. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, et al (2017) DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. *Molecular Biology and Evolution* 34:3299–3302. <https://doi.org/10.1093/molbev/msx248>
20. Sang T, Crawford DJ, Stuessy TF (1997) Chloroplast DNA phylogeny, reticulate evolution, and biogeography of *Paeonia*
21. (*Paeoniaceae*). *American Journal of Botany* 84:1120–1136. <https://doi.org/10.2307/2446155>
22. Schlee D, Sneath PHA, Sokal RR, Freeman WH (1975) Numerical Taxonomy. The Principles and Practice of Numerical Classification. *Systematic Zoology* 24:263. <https://doi.org/10.2307/2412767>
23. Shaw J, Lickey EB, Beck JT, et al (2005) The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *American Journal of Botany* 92:142–166. <https://doi.org/10.3732/ajb.92.1.142>

24. Shaw J, Lickey EB, Schilling EE, Small RL (2007) Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. *American Journal of Botany* 94:275–288. <https://doi.org/10.3732/ajb.94.3.275>
25. Tate J, Simpson B (2003) Paraphyly of *Tarasa* (Malvaceae) and Diverse Origins of the Polyploid Species. *Systematic Botany* 28:723-. <https://doi.org/10.1043/0264.1>
26. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* 22:4673–4680. <https://doi.org/10.1093/nar/22.22.4673>
27. YU J, XUE J-H, ZHOU S-L (2011) New universal matK primers for DNA barcoding angiosperms. *Journal of Systematics and Evolution* 49:176–181. <https://doi.org/10.1111/j.1759-6831.2011.00134.x>

Figures

Figure 1

UPGMA dendrogram based on dissimilarity matrix obtained from DARwin 6.0 software (R: red lentil cultivars, G: green lentil cultivars)

Figure 2

UPGMA dendrogram based on sequencing data on combination of *trnH-psbA* + *matK* regions (R: red lentil cultivars, G: green lentil cultivars)

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [S1.pdf](#)
- [S2.pdf](#)