Analysis of Various DNA Barcodes on the Turkish Protected Designation of Origin Apricot “Iğdır Kayısı” (Prunus armeniaca cv. Şalak)

Kaan Hürkan1,a,*

1Department of Agricultural Biotechnology, Faculty of Agriculture, Iğdır University, 76000 Iğdır, Turkey
*Corresponding author

A R T I C L E  I N F O

Identifying the originality and detecting the authentication of the processed and unprocessed commercial food products ensure food safety. Food adulteration of food products with high commercial value by cheap additives could threaten human health. In this study, we generated and tested five DNA barcodes (ITS, LEAFY, matK, rbcL, ycf1) of the Turkish Protected Designation of Origin Apricot “Iğdır Kayısı” (Prunus armeniaca cv. Şalak) with related primer pairs. The generated barcodes were deposited on the GenBank database. The results showed that nuclear originated ITS and LEAFY barcodes discriminated the Prunus species and cultivars better than the plastidial barcodes. Due to plenty of ITS barcodes on the databases, and good results in our study we recommend using ITS to identify Prunus species and cultivars.

A B S T R A C T

Keywords:
DNA barcoding
Apricot
Molecular identification
Food authenticity
Genomics

M A K A L E  B İ L G İ S İ

Araştırma Makalesi

Geliş : 11/05/2020
Kabul : 28/08/2020
Anahtar Kelimeler:
DNA barkodlama
Kayısı
Moleküler tanımlama
Gıda aldatmacası
Genomiks

ÖZ

Türk icatları yapılan işlenmiş veya işlenmemiş gıda ürünleriinin orijinaliliğinin belirlenmesi ve gıda aldatmacalarının belirlenmesinde modern moleküler biyoloji yöntemleri hassas ve kesin sonuç verebilme becerisi verilmiştir. Özellikle yüksek ticari değeri olan gıda ürünlerinin, daha ucuz prensip maddeleri ile hazırlanması, insan sağlığını tehdit edebilir veya tüketiciye tüketilen ürünlerin tüketicinin alınmasına sebep olabilir. Bu çalışmada coğrafi işaretli Iğdır Kayısı’nın (Prunus armeniaca cv. Şalak) beş farklı DNA barkodu (ITS, LEAFY, matK, rbcL ve ycf1) uygun primer çiftleri ile oluşturulmuş ve Sanger dizileme ile oluşturulmuştur. Bu barkodlar GenBank veritabanına aktarılmış ve kayısı tür ve çeşitleri üzerinde ayırt edebildiği incelenmiştir. Elde ettiğimiz sonuçlara göre çekirdek geni çoklu taşıyan ITS ve LEAFY barkodları kayısı tür ve çeşitlerini diğer plastik çoklu taşıyan barkodlara göre daha iyi bir şekilde ayırt etmiştir. Bu nedenle Prunus cinsini içerisindeki türlerin ve kayısı çeşitlerinin moleküler tanımlamalarında ITS barkod bölgesinin kullanılması tavsiye edilmiştir.
Introduction

DNA barcoding is a method using standardised DNA fragments to identify species. It is widely used on biodiversity researches, phylogenetics, population ecology, and forensic analyses since the first quarter of the 2000s (Cheng et al., 2016). Although the method is successfully used for identifying the animal species by sequencing the Cytochrome c oxidase 1 (CO1) gene, no universal barcoding region is implemented for plants, since nucleotide polymorphism rate differs among plant groups (Hollingsworth et al., 2009). Therefore, researchers must assay various DNA barcoding regions for a particular plant group.

Each DNA barcoding region has different characteristics due to its origin. The Internal Transcribed Spacer (ITS) region of the nuclear ribosomal cistron is one of the most used DNA barcoding regions for the plants for 25 years (Hürkan, 2017), due to its universality, ease of amplification, good polymorphism rate, and ideal barcode length (~630 bp) (White et al., 1990). The region combines both coding and non-coding sequences which empower the resolution power of the complicated plant groups (Kress et al., 2005). Genomic originated LEAFY gene is responsible for the development of the floral meristem tissue. This highly conserved gene consists of 3 exons and 2 introns (Frohlich and Parker, 2000). The introns have polymorphic sites and this makes the region usable for barcoding and phylogenetic reconstruction on the lower taxonomic levels (Frohlich and Meyerowitz, 1997). The plastidial gene matK (matK) is about ~1500 bp and comes to the forefront due to its balanced conservative/polymorphic characteristics (Hilu and Liang, 1997). The barcoding region has a fast mutation rate, and low transition/transversion rate (Min and Hickey, 1997). The plastidial gene ycf1 (ycf1) is one of the most variable plastidial loci, and this makes it a good barcode option for land plants (Dong et al., 2015). Of 420 tree species, 357 species could be distinguished by ycf1 according to the study of Dong et al. (2015).

A geographical region that is recognised by official rules to produce foods which have special characters is called a Protected Designation of Origin (PDO) (Martelo et al., 2014). Of 420 tree cultivars, 357 species could be distinguished by ycf1 according to the study of Dong et al. (2015).

In this study, we aimed to generate and analyse DNA barcodes ITS, matK, rbcL, ycf1 and LEAFY of the “İğdır Kayısı” (Iğdır Apricot) by the Turkish Patent and Trademark Office (Registration number 385, dated 17 September 2018).

Materials and Methods

Plant Material and DNA Extraction

We collected the fresh leaves of the “İğdır Kayısı” from the Iğdır University, Agricultural Application and Research Centre during their vegetation stage in 2019, and preserved them in silica-gel sachets until DNA extraction.

The DNA extraction was performed according to the modified CTAB (Doyle and Doyle, 1987) protocol. Approximately 10mg silica-gel dried leaf tissue was homogenised in mortar with 2% CTAB buffer (20 mM EDTA (pH 8.0), 100 mM Tris – HCl (pH 8.0) and 1.4 mM NaCl). The homogenate was incubated at 65°C for 1 h, and then centrifuged at 14000 g for 3 min. The supernatant was transferred to a new tube and equal volume of chloroform:isooamyl alcohol (24:1 v/v) was added and mixed vigorously for 1 h. The mixture was centrifuged at 14000 g for 5 min and washed twice with 70% ethanol. The DNA pellet was diluted in 100µl 10 mM TRIS (pH 8.0).

PCR Amplification and Agarose Gel Electrophoresis

In this study, we barcoded five most commonly used DNA barcoding regions of P. armeniaca cv. Şalak, and tested their resolution abilities among other cultivars. We used primer pairs to amplify barcoding regions from the related literature (Table 1). Twenty-five µl of PCR mixture was prepared as 2X Reaction Buffer (w/o Mg²⁺, w/ KCl), 0.1 mM dNTPs, 0.2 µM both primers, 1 U Taq DNA polymerase (Thermo Scientific - USA), 1 mM Mg²⁺, 10 ng total DNA and nuclease-free water. Thermal cycling (Sensosquest Labcycler Gradient, Germany) condition was 95°C 3 min first denaturation, 35 cycles of 95°C 30 s denaturation, 30 s annealing (50°C for ITS, 42°C for LEAFY, 52°C for matK, ycf1 and rblc), 72°C 1 min extension, and thermal cycling was finalised by 72°C 10 min extension step. We validated the PCR products on the agarose gel electrophoresis (3%, 70V for 2 h) by expected product sizes.

Sequencing, Bioinformatics and Genbank Submission

The PCR products were sent to the MedSanTek Laboratory Supplies Trade and Industry Ltd. (Turkey) for purification and two-way Sanger sequencing using the same primers used for PCR.

Climatic and edaphic factors have an important role in the nutritional level of plants. Thus, PDO food products are site-specific. The Iğdır province has 3% (31416 ton) of total apricot production ratio in Turkey and 85% is Şalak type apricot (Altıkat and Temiz, 2019). Şalak type apricot (Prunus armeniaca cv. Şalak), which is cultivated in Iğdır, has been registered as a PDO with the name “İğdır Kayısı” (Iğdır Apricot) by the Turkish Patent and Trademark Office (Registration number 385, dated 17 September 2018).
The raw sequence files were imported to the Geneious R8 (Kearse et al., 2012) software for bioinformatics analysis. We checked each sequence for sequencing quality then trimmed the primer binding regions and low-quality endings with a 5% error probability limit. The forward and reverse reads of each barcoding region were pairwise aligned (Geneious alignment tool, default settings), checked for ambiguities manually, and consensus sequences were generated. We validated each consensus sequence of the barcoding regions on the National Center for Biotechnology Information (NCBI) using the Basic Local Alignment Search Tool (BLASTn) tool. All the coding regions were annotated within the Geneious environment by reference sequences (Accessions NC_043901 and KT803847).

We used the BankIt web-based submission tool to deposit all the generated barcode sequences to the GenBank database. For each barcode, BLASTn was performed and the *Prunus armeniaca* cultivar queries with the identity value above 98% and E value equals 0 were downloaded. We aligned the sequences using the Geneious Alignment Tool with default settings and, used the FastTree 2.1.11 (Price et al., 2010) for the basic phylogenetic reconstructions (Optimized Gamma20 likelihood). We chose the phylogenetically closest organism to the *Prunus* genus as outgroup organism according to the BLASTn results. Hence, we included *Physocarpus capitatus* (AF318748) for ITS, *Lysiphyllum cunninghamii* (KT462063) for LEAFY, *Camellia longissima* for matK (KX216420), *Morus alba* (MOUCPRBCL) for rbcL, and *Pygeum topengii* (KF154931) for ycf1 as outgroups for better phylogenetic resolution.

**Results and Discussion**

We obtained 1400 ng/µl of high quality and non-degraded gDNA with a 260/280 ratio of 1.8, and used it for further PCR reactions. We successfully amplified each barcoding region by PCR with the related primer pairs. After trimming and aligning both directions reads, we verified the barcode lengths as 632 bp for ITS, 160 bp for LEAFY 766 bp for matK, 569 for rbcL, and 796 bp for ycf1. We annotated the coding regions of the barcodes and deposited all the data to GenBank database using the organism name *Prunus armeniaca* cultivar Şalak (Table 2).

We drew cladograms for each DNA barcode to better understand, how DNA barcodes worked for distinguishing the cultivars, where available, or closely related species. Our BLASTn results showed that the ITS region was the most barcoded region for apricot cultivars on the GenBank database; therefore, we were able to see how ITS performed among apricot cultivars (Figure 1). On the ITS tree, all the samples retrieved from the GenBank were apricot cultivars except the *P. capitanus* outgroup. The tree was highly supported statistically according to the FastTree support values on the branches, except Weixin and Shachenhongteke cultivars. The Yinxiangbai cultivar placed as sister to all other cultivars and that was the most distinct cultivar to others. The PDO Şalak separated from all the other cultivars. According to the tree, Şalak placed as sister to Caopixing, but Şalak ITS sequence differs from Caopixing by 8 Single Nucleotide Polymorphisms (SNPs) and one deletion on the sequence. The ITS barcoding region generally separated all of the cultivars retrieved from the GenBank. The ITS barcoding region has 29 in other barcoding regions.

Unlike ITS, there are very few LEAFY sequences of apricot cultivars on the GenBank database. According to our BLASTn results, there is only the Pea-1 apricot cultivar matching the identity value above 80%. Therefore, we could compare the LEAFY sequence of Şalak with the closely related species and genera (Figure 2). The LEAFY tree consisted of two main clades: *Prunus* spp, and phylogenetically close genera (*Rosa, Pyrus, Malus* and *Ziziphus*). This indicates that the LEAFY barcode can separate apricot from the related genera. On the *Prunus* clade, *P. armeniaca* cv Şalak and *Prunus mume* placed as a sister clade to *Prunus* species, but with low FastTree
support value. The cultivar Şalak and _P. mume_ separated from each other with high support by one SNP. Since we did not have enough LEAFY sequences on the database, the only result we obtained was that LEAFY can separate _P. armeniaca_ from closely related species.

The plastidial barcode _matK_ showed low separating ability for _Prunus_ spp. on the tree (Figure 3). Although the tree consisted of three main clades sister to _Prunus dulcis_ (almond), the _matK_ could not go further to separate at the species level. However, the barcode could separate Şalak and Zhenzhuyou cultivars.

The _rbcL_ barcode separates the tree as two clades: _Prunus consociflora_ and other _Prunus_ species as sister to it (Figure 4). Although the tree has a high supporting value, there is no distinction among most of the _Prunus_ species. The Şalak cultivar grouped with other _Prunus_ species without any distinction on its clade.

The _ycf1_ barcode had the best separating ability among plastidial barcodes. Again, due to the lack of cultivar sequences in the database, we had to compare _Prunus_ species on the _ycf1_ tree (Figure 5). The tree was highly supported since the FastTree supporting value was high throughout the tree. The Şalak cultivar grouped with _Prunus tianschanica_, a type of cherry, with distributing in the central Asian Tianshan Mount, and _Armeniaca zhengheensis_, a type of apricot in China. The Şalak cultivar and _P. zhengheensis_ differ from each other by six SNPs.

---

**Figure 1.** Cladogram resulting from FastTree analysis of ITS data. Numbers on the nodes show the FastTree support values. The Şalak cultivar is marked in bold.

**Figure 2.** Cladogram resulting from FastTree analysis of LEAFY data. Numbers on the nodes show the FastTree support values. The Şalak cultivar is marked in bold.

**Figure 3.** Cladogram resulting from FastTree analysis of _matK_ data. Numbers on the nodes show the FastTree support values. The Şalak cultivar is marked in bold.

**Figure 4.** Cladogram resulting from FastTree analysis of _rbcL_ data. Numbers on the nodes show the FastTree support values. The Şalak cultivar is marked in bold.
In this study, we generated the DNA barcodes of the Turkish PDO *P. armeniaca* cv Şalak by five the most used DNA barcoding regions and evaluated their abilities to distinguish *Prunus* cultivars and species. This is the first DNA barcoding study for a PDO apricot cultivar. We submitted the DNA barcodes to the GenBank with the accession numbers; MT072696, MT090548, MT090550, MT090549, MT120854. The barcodes are available publicly on the database. The ITS region consists of two introns (ITS1 and ITS2) and a 5.8S coding region. This combination brings a good “resolution” power to the barcodes’ both upper and lower taxonomic levels due to ITS1 and ITS2 introns having more polymorphisms than the 5.8S. Moreover, this region has biparental inheritance contrasting to the plastidial regions (Aguilar et al., 1999). Therefore, the genetic information coming from both parents homogenize in this region. This also gives better discrimination success to ITS (Hollingsworth et al., 2011). According to our results, the ITS barcode outperforms the other barcodes by having the better distinguishing ability as mentioned in the literature. The second factor of the success of ITS is the GenBank database has much more ITS sequences belong to *Prunus* species. Thus, we were able to include more *Prunus* cultivars to our dataset. LEAFY consists of three exons and one intron. The exons are highly conserved in flowering plants (Frohlich and Meyerowitz, 1997). Our second nuclear-originated barcoding region LEAFY also showed good discrimination results at the species level since its balanced polymorphism rate by intron-exon combination.

The uniparental inherited DNA barcoding region COI is a standard DNA barcode for the animals and it shows high discriminatory power for animals. Nevertheless, this region is not useful for the plants due to the low rate of nucleotide substitution in the plant mitochondrial genome on the plants (Hollingsworth et al., 2011). Generally, our plastidial DNA barcoding regions showed low discrimination power relative to the ITS. The plastidial ycf1 gene has been used for land plants as a barcoding region (Cheng et al., 2016). It is more variable than other existing plastidial barcodes, and it is called “the most promising plastid DNA barcode” (Dong et al., 2015). The ycf1 barcode was the highest discriminatory barcode among other plastidial barcodes for *Prunus* species according to our results. The CBOL Plant Working Group selected the rbcL and matK genes as the core plant barcodes, and these regions are frequently used for barcoding plants (CBOL Plant Working Group, 2009). However, according to the previous studies, ycf1 is more variable than rbcL and matK barcodes (Oliver et al., 2010; Wolf et al., 2011). Likewise, in our analysis, both the rbcL and matK showed poor discrimination power even at species level comparing to ycf1.

In conclusions, marker selection is very important as a reliable identification, saving laboratory consumables and, of course, saving time. Our results supported the literature for *Prunus* species. The nuclear-originated barcodes performed better than the plastidial barcodes. We recommend using nuclear-originated barcodes such as ITS and LEAFY to identify *Prunus* species and cultivars.

Acknowledgement

We would like to thank the journal editors and anonymous referees for their improvements in the article. We also thank Cathy Seither for the language proof and Molecular Biologist (PhD student) Aybükę Erol for her help on laboratory studies. This study was financially supported by Iğdır University, Scientific Research Coordination Unit; Project number: 2019-FBE-A17.

References

Aguilar JF, Rossello JA, Feliner, GN. 1999. Nuclear ribosomal DNA (nrDNA) concerted evolution in natural and artificial hybrids of Armeria (Plumbaginaceae). Molecular Ecology 8: 1341–1346

Altıkat S, Temiz Ş. 2019. İğdır Ili Kayısı Çeşitlerinin Fiziko-Mekanik ve Bazı Kimyasal Özellikleri. Yüzüncü Yıl Üniversitesi Tarım Bilimleri Dergisi, 29 (3), 373–381. doi: 10.29133/yyutbd.521570

Anonymous, 2007. http://sgh.tarim.gov.tr/Proje_Yontemleri/Masterplanlar/masterplan/masterplan.htm. [Accessed 30 March 2020]

Batmini MA, Bourguiba H, Trifi-Farah N, Krichen L. 2019. Molecular diversity and phylogeny of Tunisian Prunus armeniaca L. by evaluating three candidate barcodes of the chloroplast genome. Scientia Horticulturae (Amsterdam) 245:99–106. https://doi.org/10.1016/j.scienta.2018.09.071

CBOL Plant Working Group. 2009. A DNA barcode for land plants. Proceedings of the National Academy of Sciences U.S.A. 106: 12794–12797.

Cheng T, Xu C, Lei L, et al. 2016. Barcoding the kingdom Plantae: new PCR primers for ITS regions of plants with improved universality and specificity. Molecular Ecology Resources 16:138–149. https://doi.org/10.1111/1755-0998.12438

Dong W, Xu C, Li C, Sun J, Zuo Y, Shi S, Cheng T, Guo J, and Zhou S. 2015. ycf1, the most promising plastid DNA barcode of land plants. Scientific Reports, 5(1): 8348. https://doi.org/10.1038/srep08348

Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin, 19, 15–87. https://doi.org/10.2307/4119796

Figure 5. Cladogram resulting from FastTree analysis of ycf1 data. Numbers on the nodes show the FastTree support values. The Şalak cultivar is marked in bold.
Frohlich MW, Meyerowitz EM. 1997. The Search for Flower Homoeotic Gene Homologs in Basal Angiosperms and Gnetales: A Potential New Source of Data on the Evolutionary Origin of Flowers. International Journal of Plant Sciences 158:S131–S142.

Frohlich MW, Parker DS. 2000. The Mostly Male Theory of Flower Evolutionary Origins: From Genes to Fossils. Systematic Botany 25:155–170.

Gulyás G, Sramkó G, Molnár VA, Rudnóy S, Illyés Z, Balázs T, Bratek Z. 2005. Nuclear ribosomal DNA ITS paralogs as evidence of recent interspecific hybridization in the genus Ophrys (Orchidaceae). Acta Biologica Cracoviensia Series Botanica, 47, 61–67.

Hasebe M, Omori T, Nakazawa M, Sano T, Kato M, Iwatsuki K. 1997. The rbcL gene sequences provide evidence for the evolutionary lineages of leptosporangiate ferns. Proceedings of the National Academy of Sciences, 94(12): 5730–5734. https://doi.org/10.1073/pnas.94.12.5730

Hilu K, Liang H. 1997. The matK Gene: Sequence Variation and Application in Plant Systematics. American Journal of Botany, 84: 830–839.

Hollingsworth PM, Forrest LL, Spouge J.L, Hajibabaei M, Erickson DL, Fazekas AJ, Graham SW, James KE, Kim K-J, Kress WJ, Schneider H, van Alphen-Stahl J, Barrett SCH, van den Berg C, Bogarin D, … Little DP. 2009. A DNA barcode for land plants. Proceedings of the National Academy of Sciences, 106(31): 12794–12797. https://doi.org/10.1073/pnas.0905845106

Hollingsworth PM, Graham SW, and Little DP. 2011. Choosing and Using a Plant DNA Barcode. PLoS ONE, 6(5): e19254. https://doi.org/10.1371/journal.pone.0019254

Hürkan K. 2017. Karasal bitiklerde DNA barkodlama: Bazı DNA barkod bölgelerinin incelenmesi. International Journal of Innovative Approaches in Science Research, 1(1): 57–67. https://doi.org/10.29329/iijaar.2017.99.6

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meinjies P, Drummond A. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics, 28(12): 1647–1649. https://doi.org/10.1093/bioinformatics/bts199

Kibar H, Kibar B, Sürmen M. 2014. Sicaklık ve Yağış Değişiminin Iğdır Ilinde Bitkisel Ürün Deseni Üzerine Etkileri. Adnan Menderes Üniversitesi Ziraat Fakültesi Dergisi, 11(1): 11-24. Retrieved from https://dergipark.org.tr/en/pub/aduziraat/issue/26419/278139

Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH. 2005. Use of DNA barcodes to identify flowering plants. Proceedings of the National Academy of Sciences of the United States of America, 102(23): 8369–8374. http://www.pnas.org/content/102/23/8369.abstract

Li X, Yang Y, Henry RJ, Rossetto M, Wang Y, Chen S. 2015. Plant DNA barcoding: from gene to genome. Biological Reviews, 90(1): 157–166. https://doi.org/10.1111/brv.12104

Martelo-Vidal, MJ, Vázquez M. 2016. Advances in ultraviolet and visible light spectroscopy for food authenticity testing. In Advances in Food Authenticity Testing. https://doi.org/10.1016/B978-0-08-100220-9.00003-5

Min XJ, Hickey DA. 2007. BARCODING: Assessing the effect of varying sequence length on DNA barcoding of fungi. Molecular Ecology Notes, 7(3): 365–373. https://doi.org/10.1111/j.1471-322.2007.01698.x

Price MN, Dehal PS, Arkin AP. 2010. FastTree 2 - Approximately Maximum-Likelihood Trees for Large Alignments. PLoS ONE, 5(3):e9490. doi:10.1371/journal.pone.0009490.

Sramkó G, Attila MV, Hawkins JA, Bateman RM. 2014. Molecular phylogeny and evolutionary history of the Eurasian orchid genus Himantoglossum s.l. (Orchidaceae). Annals of Botany, 114(8): 1609–1626. https://doi.org/10.1093/aob/mcu179

White TJ, Bruns TD, Lee S Taylor JW. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In M. A. Innis, D. H. Gelfand, J. J. Sninsky, T. J. White (Eds.), PCR protocols: A guide to methods and applications (pp. 315–322). Academic Press.

Yu J, Xue J-H, Zhou S-L. 2011. New universal matK primers for DNA barcoding angiosperms. Journal of Systematics and Evolution, 49(3): 176–181. https://doi.org/10.1111/j.1759-683x.2011.00134.x

Yu LX, Yan B. 2013. Development of universal primers for isolating fragments of the LEAFY gene. Genetics and Molecular Research, 12(2): 1777-1780.

Wolf PG, Der JP, Duffy AM, Davidson JB, Grusz AL, Pryer KM. 2011. The evolution of chloroplast genes and genomes in ferns. Plant Molecular Biology, 76(3-5): 251-261.