Scientia Horticulturae 245 (2019) 99–106

Contents lists available at ScienceDirect

Scientia Horticulturae

journal homepage: www.elsevier.com/locate/scihorti

Molecular diversity and phylogeny of Tunisian *Prunus armeniaca* L. by evaluating three candidate barcodes of the chloroplast genome

Mohamed Amine Batnini, Hedia Bourguiba, Neila Trifi-Farah, Lamia Krichen*

Université Tunis El Manar, Faculté des Sciences de Tunis, Laboratoire de Génétique Moléculaire, Immunologie et Biotechnologie, Campus Universitaire El Manar, 2092, Tunisia

ARTICLE INFO

Keywords:
Prunus armeniaca L.
Chloroplastic DNA
Genetic diversity
DNA barcodes
Tunisia

ABSTRACT

Three intergenic spacers (*atp*B-*rbc*L, *trn*L-*trn*F and *rps*16-*trn*Q) of chloroplast DNA (cpDNA) considered as candidate DNA barcode markers were used on a total of 42 Tunisian apricot accessions to assess the genetic diversity and phylogenetic structure. A total of 702 variable sites, including 208 singleton variable sites and 492 parsimony informative sites were obtained from the combined cpDNA sequences (2333 bp). Forty-two haplotypes were identified based on the combined fragments. The values of nucleotide diversity (P_i), and haplotype diversity (H_d) were 0.049 and 1, respectively. No statistical significance was detected in Tajima's *D* and Fu and Li's tests. The accession 1 F ('Bedri') from Kairouan was considered as one of the ancient haplotypes. Four genetic groups were identified in a neighbour-joining tree that are independent from the geographic origin and the mode of propagation of the studied accessions. Overall, the analysis of data sequence revealed that *rps*16-*trn*Q region was the most suitable candidate barcode and gave an overall variation at the interspecies level. This study provides required information to develop strategies of conservation.

1. Introduction

The genus *Prunus* (Rosacea, Amygdaloideae) is an economically important genus subdivided into 400 species of trees and shrubs (Maynard et al., 1991), mainly from the Northern Hemisphere. Taxonomists used morphological features of flowers and drupes to identify the species within the section. To the lay mind, those features are difficult to grasp. To solve phylogenetic problems at the species level, or to distinguish species using DNA sequences, it became necessary to identify regions of the genome with very high evolutionary rates. In fact, a greater availability of these regions will increase the ability to resolve such identification problems.

Apricot (*Prunus armeniaca* L.) is a stone fruit species, native from China and Central Asia (Vavilov, 1992). In Tunisia, several studies were assessed to characterize and estimate the morphological variability as well as the genetic diversity and structure of autochthonous traditional apricot germplasm using different nuclear DNA markers such as AFLP (Krichen et al., 2008) and microsatellites (Bourguiba et al., 2010a,b). The obtained results revealed the existence of two apricot gene pools, related to the mode of propagation of apricot as grafted and seed-propagated accessions, sharing the same genetic diversity (Bourguiba et al., 2010b; Krichen et al., 2014). Grafted-propagated accessions were encountered in all apricot regions of culture from the north to the south of Tunisia, while seed-propagated ones were specific to the oasis agroecosystems and locally named 'Bargoug'.

Compared to nuclear DNA, the chloroplast genome could provide valuable information for species identification, phylogeny and population genetic analyses due to its uniparental inheritance, nearly neutral evolution, low evolutionary rate, and absence of recombination (Hewitt, 2001). Furthermore, chloroplastic DNA (cpDNA) is able to detect the genetic structure and evolutionary events even with a smaller effective population size when compared to nuclear genome (Petit et al., 2003). The use of many coding regions, introns, and intergenic spacers of the cpDNA were described to study the genetic diversity and phylogeography of different plant species like pears (Liu et al., 2013; Zhong et al., 2014), date palm (Rhouma-Chatti et al., 2014), and Eurasian plums (Reales et al., 2010).

Despite chloroplast genome conservation, structural alterations such as inversions, translocations, deletions (gaps) and insertions found in hypervariable regions of cpDNA like *rps*16-*trn*Q region, evolved at a faster rate than other regions. These variations have numerous important implications in deeper insights into molecular phylogeny, population genetics and conservation aspects, as reported in *Prunus pseudocerasus* (Chen et al., 2013a) and *Pyrus* genus (Chang et al., 2017).

The term 'DNA barcode' for global species identification was first coined by Hebert et al. (2003) and has gained worldwide attention in

* Corresponding author.
E-mail address: lamia.krichen@issbat.utm.tn (L. Krichen).

https://doi.org/10.1016/j.scienta.2018.09.071
Received 21 May 2018; Received in revised form 24 September 2018; Accepted 26 September 2018

the scientific community (Schindel and Miller, 2005). DNA barcoding involved the sequencing of short segments of the chloroplastic or mitochondrial genome and compared the results with orthologous reference sequences available in public database such as BOLD (www.boldsystem.org) and GenBank (www.ncbi.nlm.nih.gov/genbank). This approach is based on the analysis of the nucleotide variability existing within standard regions of the genome that are particularly informative for the identification of species (Ward et al., 2005). In fact, the DNA barcode sequence contained enough unique information, in terms of single nucleotide polymorphisms (SNPs) and insertion/deletions (InDels), shared among individuals of a species with slight variations, but specifically associated to one species with a unique haplotype. Some regions of the chloroplast genome, such as *mat*K, *atp*B, *rbc*L, and *trn*H-*psb*A, were relied upon heavily for development of candidate markers for plant DNA barcoding. Thus, DNA barcoding has proved to be successful in the species identification process of a large number of plants such as medicinal plants (Chen et al., 2010; Christina and Annamalai, 2014), *Schima* woody genus (Yu et al., 2017), and Fabaceae (Gao et al., 2010). However, a lack of barcoding method was noted for *Prunus* genus.

In this study, the barcode method involving three chloroplastic markers, *atp*B-*rbc*L, *trn*L-*trn*F and *rps*16-*trn*Q was used, in order to: (*i*) investigate the genetic diversity and structure, (*ii*) assess phylogenetic analysis of apricot species in Tunisia, and (*iii*) discuss the efficiency of the use of the three candidate barcodes.

2. Materials and methods

2.1. Plant materials

The present study focused on a total of 42 Tunisian apricot accessions, including 34 domesticated cultivars propagated by grafting and 8 spontaneous accessions propagated by seeds, specific to the oasian agroecosystems and locally named 'Bargougs'. Each accession corresponded to one genotype. The graft-reproduced genotypes constituted commercial cultivars.

All the selected accessions belonged to the local core collection established by Krichen et al. (2012) in order to maximize the local diversity according to their morphological characteristics and molecular information.

Samples covered the principal areas of apricot culture as Testour, Ras Jbel, Kairouan, Sfax, Gabes, Jerba, Gafsa, Tozeur, Tameghza and Midess, as well as all the different bioclimatic levels of Tunisia (Fig. 1). Samples consisted on fresh leaves from these plants which were stored at −20 °C until DNA isolation.

Furthermore, additional sequences belonging to the *Prunus* genus obtained from the NCBI GenBank were used for comparatives studies.

2.2. DNA isolation

Genomic DNA was extracted with a modified cetyl trimethyl ammonium bromide (CTAB) method as described by Doyle and Doyle (1987) and subjected to 1.2% agarose gel electrophoresis for quality examination. Then, DNA concentration was estimated using a Thermo Scientific NanoDrop™ spectrophotometer. The isolated DNA was diluted to a concentration of 25 ng.µl^{-1} (working solution) and stored at −20 °C for subsequent analyses.

2.3. PCR amplification and sequencing

The universal barcode forward and reverse primers were used for polymerase chain reaction (PCR) amplification of *atp*B-rbcL, trnL-trnF and rps16-trnQ regions of the cpDNA as mentioned in Table 1 (Chen et al., 2013b).

PCR was performed with a PTC200 thermal cycler (Applied Biosystems, FosterCity, CA, USA) in a 25-µL reaction mixture containing the following components: 25 ng of DNA template, 0.2 mM of each dNTP, 0.2 mM of each primer (forward and reverse), 2 mM of $MgCl_2$, and 1 U of Taq DNA polymerase. PCR amplification conditions for each marker were provided in Table 1.

PCR products were examined electrophoretically using 1.5% agarose gel and purified from agarose gel using the QIAquick PCR purification kit (Qiagen, Germany) following the protocols provided by the manufacturer. The purified PCR products were sequenced directly by an ABI 3130 automated capillary sequencer (Applied Biosystems, Foster City, CA, USA) using Big Dye Terminator v.3.1 sequencing kit (Applied Biosystems, Foster City, CA, USA).

2.4. Data analyses

Chloroplast DNA regions were adjusted manually, aligned using software ClustalX v1.81 (Thompson et al., 1997) and then analyzed by MEGA software package v6.06 (http://www.me-gasoftware.net/index.php). Chloroplast DNA fragments from the three regions *atp*B-rbcL, trnL-trnF and rps16-trnQ were combined using software ClustalX v1.81 for further analyses.

Different parameters of genetic diversity including haplotype number (h), haplotype diversity (H_d), nucleotide diversity (P_i), variable sites number (V_s), singleton variable sites number (S_s), and parsimony informative sites number (P_s) were calculated based on each cpDNA region and combined regions using software DnaSP v5.10.01 (Librado and Rozas, 2009).

To estimate the population growth and to test the deviation from the panmictic equilibrium, neutrality tests, estimated by Tajima's D parameter (Tajima, 1989) and the F and D indices (Fu and Li, 1993), were also performed using software DnaSP v5.10.01 (Librado and Rozas, 2009). Statistical significance was estimated with coalescent simulations as implemented. Mismatch distribution was also estimated using DnaSP v5.10.01.

A neighbor-joining (NJ) phylogenetic tree was constructed based on the combined matrix of the three studied intergenic sequences: *atp*B-*rbc*L, *trn*L-*trn*F and *rps*16-*trn*Q using DARwin v5.0.158 software (Perrier and Jacquemoud-Collet, 2006). In fact, NJ method was considered to be the simplest and the most convenient for barcoding studies. The NJ tree was generated with 1000 bootstrap replications and the obtained tree with the highest log likelihood is shown.

Considering the haplotype spatial structure, several networks were constructed using the median-joining (MJ) method (Bandelet et al., 1999) by Network v4.6.13 software (http://www.fluxus-engineering.com/) derived from each cpDNA region and combined regions.

3. Results

3.1. Polymorphism of chloroplast DNA

The three cpDNA intergenic spacers *atp*B-*rbc*L, *trn*L-*trn*F and *rps*16-*trn*Q, were amplified and sequenced independently. All three barcodes presented a maximum success rate (100%) for PCR and sequencing. A BLASTN search with the above sequences showed that *atp*B-*rbc*L, *trn*L-*trn*F and *rps*16-*trn*Q sequences exhibited strong homology, as 99% of identity, with already deposited sequences of other *Prunus* species.

After visual inspection and adjustment, the characters of the three barcoding regions were submitted to the statistic methods and the obtained results were listed in Table 2. The aligned sequences length was 955 bp for *rps*16-*trn*Q, 729 bp for *atp*B-*rbc*L and 479 bp for *trn*L-*trn*F. By analyzing nucleotide composition, a higher level of A/T bases was detected for the three sequences, while the GC content varied from 29.6% to 35.8%. The intergenic spacer region *rps*16-*trn*Q was the highest polymorphic region with 320 variable sites among a total of 955 sites (33.5%), in which 309 were parsimony-informative sites and 63 were singleton variable sites. Regarding to *trn*F-*trn*L and *atp*B-*rbc*L regions, 126 (26.3%) and 215 (29.5%) variable sites were obtained respectively,