TÜBİTAK

Turkish Journal of Botany

http://journals.tubitak.gov.tr/botany/

Research Article

Turk J Bot
(2019) 43: 737-748
© TÜBİTAK
doi:10.3906/bot-1810-44

Genetic Diversity, Phylogeography and Population gene flow of Tunisian *Pistacia vera* L.

Sarra CHOULAK[1,*], Zined MARZOUK[1], Khaled CHATTİ[1], Soumaya RHOUMA-CHATTİ[1], Rim OUNİ[2], Noureddine CHATTİ[1]

[1]Laboratory of Genetics, Biodiversity and Bioresources Valorisation, Higher Institute of Biotechnology of Monastir, University of Monastir, Monastir, Tunisia

[2]Research Unit on Agrobiodiversity (UR13AGR05), Higher Agronomic Institute (ISA), IRESA-University of Sousse, Chott-Mariem, Tunisia

Received: 18.10.2018 • **Accepted/Published Online:** 19.04.2019 • **Final Version:** 21.11.2019

Abstract: The aim of this paper was to determine how the pistachio trees evolve, and the evolving factors that influence pistachio population establishment. For that, we investigated pistachio genetic structure from some regions of Tunisia, by sequencing two noncoding chloroplastic regions (*trn*L (UAA) and *trn*L-*trn*F intergenic spacer). We found strong genetic diversity among groups, with the absence of high differentiation between population pairs. A deep phylogeographical break separated two major clusters: "El-Guetar and "Gafsa/Sidi-Bouzid. This conclusion is proved by the haplotype networks, the phylogenetic trees, and the molecular variance analysis. Different interpretations were proposed to explain this cytoplasm dimorphism, based on the molecular evolution and demographic history analysis: (1) the domestication events, which are very important to understand the variability between the prospective areas, (2) the gene flow between them, a process that occurs both in time and in space through pollen and seeds, and strongly interacts with the local farming systems, (3) the geographical barriers that exist, which limit gene flow transfer and make particular climatic conditions of the El-Guetar oasis. Given that, the genetic diversity study within Tunisian pistachio cultivars is very useful to contribute to the national management effort for the improvement and conservation of pistachio genetic resources. Moreover, in 2017, Tunisia ranked among the top 10 countries in global pistachio production. Thus, studying the diversity of Tunisian pistachio can make an important impact on global production.

Key words: Chloroplast DNA, gene flow, phylogeography, *Pistacia vera* L., *trn*L-F markers, Tunisia

1. Introduction

Pistacia vera L. (2n = 30) is a dioecious and wind-pollinated member of the family Anacardiaceae. It is the only cultivated and commercialized species in the genus *Pistacia* (Zohary, 1996). The pistachio tree in Tunisia is an ancient crop, particularly in semiarid and arid zones (Mlika, 1980). It is found in the north (Ariana, Jendouba, Bizerte and Le Kef), the center (Mahdia, Monastir, Kairouan, Kasserine and Sidi Bouzid), and in the oases of southern Tunisia (Gafsa and El-Guetar). Pistachio cultivation is an important economic activity in Tunisia; the national production in 2017 reached 3637 t (Faostat, 2017).

Pistachio trees are currently threatened by genetic erosion particularly due to the extension of monovarietal orchards using the main variety in Tunisia: "*Mateur*". It should be noted that the local variety "*Sfax*" has recently been reintroduced in the region of Gafsa, after being grown in California (Abdedaeim, 2015). Therefore, it is imperative to establish strategies for conservation of pistachio cultivars, with the first step being the identification and characterization of the remaining traditional accessions (Abdedaeim, 2015). Phenotypic diversity has been studied using morphological characters derived from the pistachio descriptor list (IPGRI, 1997) by Ghrab et al. (2012) and Chatti et al. (2017). In addition, inter-simple sequence repeat (ISSR), sequence-related amplified polymorphism (SRAP), and chloroplastic markers were utilized to investigate the genetic variability, population structure, and population's differentiation of this species in Tunisia (Farès et al., 2009; Choulak et al., 2015; Guenni et al., 2016). In the world, many studies have described the use of the random amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), ISSR, SRAP methods, chloroplastic markers, and microsatellite marker (Katsiotis et al., 2003; Golan-Goldhirsh et al., 2004; Baghizadeh et al., 2010; Talebi et al., 2012; Ziya-Motalebipour et al., 2016) to identify *Pistacia vera* L. cultivars. These studies confirmed the efficiency of the molecular markers used to evaluate the genetic diversity within the studied genotypes. Many

* Correspondence: choulak.sarra@gmail.com

results proved that chloroplastic noncoding DNA regions, like the intergenic spacer between the *trn*L (UAA) 3'exon and the *trn*F (GAA) and the *trn*L-F gene can be used to answer questions concerning the relationships at intra- and interspecific levels (Choulak et al., 2015; Talebi et al., 2016). In fact, the noncoding regions provide the most practical source of data for phylogenetic inference at lower taxonomic levels (Morton and Clegg, 1993).

In this study, we took into account the eco-geographical distribution of populations and the use of molecular markers to estimate intra- and inter-population diversity, to evaluate gene flow, and to characterize the individuals and the Tunisian pistachio populations genetically and with precision.

2. Materials and methods

2.1. Sample collection

Fields visits were made from April 2012 to July 2013 in the central southern traditional areas of pistachio culture in Tunisia (Gafsa and El-Guetar oasis).

We also included accessions from Sidi-Bouzid, obtained from the pistachio germplasm collection maintained at the Regional Center for Agricultural Research of Sidi-Bouzid. Overall, thirty-nine accessions, including 30 female and 9 male trees (Dhokkars) and 2 *P. atlantica* (Battoum) were used in the study (Table 1).

2.2. DNA isolation, PCR reactions, and DNA sequencing

Total genomic DNA was extracted, from frozen leaves of single adult trees, by means of a modified MATAB method (Risterucci et al., 2000) and the QIAGEN DNeasy Plant Mini Kit (Qiagen, Venlo, the Netherlands). Before extraction, leaves were ground in liquid nitrogen using a ball mill (type MM2; Retsch, Haan, Germany). The DNA was resuspended in molecular biology water after isopropanol evaporation. A spectrophotometric measurement and electrophoresis checking of the extracted nucleic acid were performed (Sambrook et al., 1989). Primers and the PCR protocol were previously described by Choulak et al. (2015, 2017). Before sequencing, the PCR reactions were purified using the Wizard SV Gel PCR Clean-up system Kit (Promega, WI, USA).

The *trn*L-*trn*F spacer sequences were separately aligned and studied. Next, the two matrices (spacer and intron) were combined for the final analysis.

The *trn*L-F marker sequences of all the 39 annotated accessions were submitted to NCBI GenBank (accession numbers: MK654683-MK654721).

2.3. Sequence analysis

2.3.1. Genetic diversity and phylogenetic analyses

Chloroplast DNA sequences were confirmed by the BLAST database (Altschul et al., 1997). Sequence alignment was performed with the ClustalW program executed in Bioedit

Table 1. Label, sex, and origin of pistachio cultivars studied.

Geographic origin	Cultivars label	Sex	Species
	GT1	♀	*P. vera*
	GT2	♀	*P. vera*
	GT3	♂	*P. vera*
	GT4	♂	*P. vera*
	GT5	♀	*P. vera*
	GT6	♂	*P. vera*
	GT 7	♀	*P. vera*
	GT 8	♀	*P. vera*
	GT 9	♂	*P. vera*
El Guetar (Tunisia)	GT10	♀	*P. vera*
	GT11	♀	*P. vera*
	GT12	♀	*P. vera*
	GT13	♀	*P. vera*
	GT14	♀	*P. vera*
	GT15	♂	*P. vera*
	GT16	♀	*P. vera*
	GT17	♀	*P. vera*
	GT18	♀	*P. vera*
	GT19	♀	*P. vera*
	GT20	♀	*P. vera*
	GF1	♂	*P. vera*
	GF2	♂	*P. vera*
Gafsa (Tunisia)	GF8	♀	*P. vera*
	GF9	♀	*P. vera*
	GF10	♀	*P. vera*
	Battoum1		*P. atlantica*
	SB1	♀	*P. vera*
	SB2	♀	*P. vera*
	SB3	♀	*P. vera*
	SB4	♀	*P. vera*
	SB5	♂	*P. vera*
	SB6	♂	*P. vera*
Sidi Bouzid (Tunisia)	SB7	♀	*P. vera*
	SB9	♀	*P. vera*
	SB10	♀	*P. vera*
	Matteur1	♀	*P. vera*
	Mtatteur2	♀	*P. vera*
	Irani	♀	*P. vera*
	Battoum2		*P. atlantica*
USA	*Pell*	♀	*P. vera*
	Wen	♀	*P. vera*
Palestine	*Golan*	♀	*P. vera*

software (Hall, 1999). Lengths and polymorphisms of the sequences were estimated by the MEGA and DNAsp programs (Tamura et al., 2013; Librado and Rozas, 2009).

The genetic relationships between varieties and the evolutionary trees were evaluated using the neighbor-joining (NJ) and Bayesian inference (BI) reconstructions (Saitou and Nei, 1987; Ronquist et al., 2012). The methods were applied using MEGA v6.06 and MrBayes v. 3.2.2 software. Three published *trn*L-*trn*F sequences from two American varieties (GenBank accessions: KP055540 (Weeks et al., 2014) and EF193139 (Yi et al., 2008)), and a single accession from Palestine (GenBank accession: AY677204 (Yi et al., 2008)) were included in our analysis. Two sequences of *Cotinu scoggygria Scop.* (GenBank accessions: KF600601 and KF600602) were used as an outgroup.

The distribution of pairwise sequence differences between *P. vera* populations was considered according to the correction model based on the Bayesian Information Criterion (BIC) test implemented in jModelTest 0.1.1 (Posada, 2008). The BIC analysis detects the best-fit substitution model for the *trn*L-*trn*F sequences. In addition, the relationships among the different haplotypes were graphically traced by the NETWORK software (Bandelt et al., 1999).

2.3.2. Demographic histories

By means of selective neutrality tests, we checked the hypothesis of the mutation/drift equilibrium for a supposedly neutral polymorphism. Tajima's D (Tajima, 1989), Fu's Fs (Fu, 1997), and Fu and Li's (1993) statistical tests were conducted to verify this hypothesis, using DNAsp program. Complementary to these tests, the R2 statistic, as a detector of population growth, was conducted based on the differences between the number of singleton mutations and the average of the nucleotide difference (Ramos-Onsins and Rozas, 2002). To test the population expansion, we examined the distribution of the observed number of differences between pairs of haplotypes by "*mismatch distribution*" (Rogers and Harpending, 1992).

2.3.3. Genetic differentiation and gene flow

Molecular variance analysis (*AMOVA*) (Excoffier et al., 1992) was executed by Arlequin (Excoffier and Lischer, 2010) to evaluate the genetic variation within and among populations of the *P. vera* species. Additionally, we measured genetic differentiation between populations using Wright's F-statistics (Wright, 1931). The Arlequin (Excoffier and Lischer, 2010) program was, also, used to calculate the F_{ST} (Excoffier et al., 1992) and the Φ_{ST} estimators. Based on these parameters (Φ_{ST} and F_{ST}), the pairwise comparisons between populations were represented by the ordinal multidimensional scaling (MDS) analyses using XLSTAT (AddinSoft, 2007). The program *IMa2* (Hey, 2010) was used to characterize the gene flow between groups in both directions. To do this, we performed a series of analyses using sequence data of the *trn*L-F marker. Five simulations were carried out; each comprising a step of a burning of 10,000-iteration chains followed by 100,000 data collection iterations. These five simulations were then combined and analyzed.

3. Results

3.1. Genetic diversity of Tunisian pistachio

The *trn*L-*trn*F intergenic spacer PCR fragments were about 421 bp. The percentage of GC was 37% to 39.5% with an average of 38.1%. The AT percentage was 60.5% to 63%. Transitional / transversional ratio (*R*) was low and equal to 0.689 (Table 2). The multiple sequence alignment showed 29 highly variable sites (22 informative sites and 7 unique sites) (Table 2). The haplotype diversity (*Hd*) recorded a very high value (0.953), as well as the nucleotide diversity (*Pi*) (0.013) (Table 2).

On the other hand, the length of the combined region (*trn*L (UAA) and *trn*L-*trn*F spacer) was 927 pb for all cultivars. The combined sequences revealed 70 polymorphic sites and defined 32 haplotypes. Among the 70 variable sites, 35 were parsimoniously informative and 35 were singletons sites. Throughout the combined sequences, 77 mutations were detected. The indices of genetic diversity were calculated. The means of the haplotype (*Hd*) and nucleotide (*Pi*) diversity were higher than for *trn*L (UAA) (Choulak et al., 2015) and the *trn*L-*trn*F spacer taken separately. In fact, these values were 0.983 ± 0.014 and 0.011 ± 0.001, respectively. The ratio *R* was equal to 0.647 (Table 2).

3.2. Haplotype networks, phylogeographic analyses

Phylogenetic trees and haplotype networks testified considerable phylogeographical structures for both chloroplastic regions. Sequences of *Cotinus coggygria* and foreign varieties of *P. vera* were used to root phylogenetic reconstructions. *Pistacia atlantica* (*Battoum*) sequences, which were already sequenced, were only used in evolutionary trees.

The chloroplast DNA sequences made well-resolved and well-supported phylogenetic trees (*BI* / *NJ*) (Figure 1). The phylogeographical structure, with *NJ* and *BI* methods, was very comparable between the intergenic spacer and the combined sequences, in particular: (1) Local varieties were distributed in two main clusters; El-Guetar and Gafsa/Sidi-Bousid, (2) '*Mateur*' varieties (1 and 2) were grouped in a distinct subgroup, (3) The variety '*Irani*' was clearly discarded from local ones, (4) Foreign varieties of *P. vera* were grouped in a separate cluster, showing a strong divergence from autochthonous varieties.

Using the NJ and BI methods, we detected a genetic structure according to the geographical origin of pistachio trees, sustained by high nodes values (Figure 1). In

Table 2. Summary of polymorphism of sequences for *trn*L-*trn*F spacer and the combined data used in this study among the Tunisian pistachio accessions.

	***trn*L-*trn*F spacer**	**Combined: *trn*L-*trn*F spacer and *trn*L intron**
Number of sequences	37	37
Alignment length (bp)	421	927
Conserved sites	392	857
Variable sites	29	70
Parsimony informative characters	22	35
Singleton variable sites	7	35
Total number of mutations	36	77
Number of haplotypes (H)	22	32
Haplotype diversity (*Hd*)	0.953	0.983
Nucleotide diversity (*Pi*)	0.013	0.011
Average of pairwise differences (*k*)	5.556	11.167
Transition/transversion rate ratios for purines (*k1*)	0.489	1.702
Transition/transversion rate ratios for pyrimidines (*k2*)	2.820	2.505
Transition/transversionbias (*R*)	0.689	0.647

addition, *Pistacia atlantica* (*Battoum*) haplogroups formed distinct clusters and are very well supported. Moreover, a haplotype network by the parsimony method was produced for *P. vera* populations (Figure 2). Each disc constitutes a haplotype, and the numbers correspond to mutational steps between haplotypes. Using *trn*L-F sequences, the haplotype network (Figure 2) revealed two major haplogroups (I and II) separated by five mutations. The first haplogroup (I) displayed a star-like pattern and assembled El-Guetar's accessions. This pattern is indicative of a population expansion during the recent history of the Tunisian pistachio (Ray et al., 2003). In contrast, the haplogroup II was heterogeneous and could be divided into four groups of haplotypes: IIa represented only the accessions of Gafsa, IIb reassembled Sidi-Bouzid accessions, IIc presented the two '*Mateur*' varieties, and IId represented by *Irani* haplotype. Several missing haplotypes were noticed in this chloroplastic network, corresponding to median vectors or *mv* (black points in Figure 2). The haplotypic network confirms the subdivision of the *P. vera* species according to their geographical distribution.

3.3. Neutrality tests and demographic histories

Selective neutrality analyses showed that Tajima's and Fu's Fs statistical tests were negative and insignificant for each population, which favors the hypothesis of a population in a stable demographic state (Table 3). These observations confirmed that the Tunisian pistachio population has not undergone a demographic expansion in the past. In addition, low significant values of the Harpending's raggedness index (*rg*) and *R2* statistical test were recorded. These analyses support the suggestion of a demographically stable population.

Population size changes or "mismatch distributions" were constructed for the three different localities. The "mismatch distribution" is usually multimodal when we analyze stable populations. On the contrary, it is unimodal for populations that have experienced population expansion. For both localities (Gafsa and Sidi-Bouzid), the multimodal curves indicated a demographic stability during a long period of their evolutionary history and until today (Figure 3). On the other hand, a case of population expansion had been suggested for El-Guetar (Figure 3); this was indicated by unimodal mismatch distribution. The results obtained for El-Guetar appeared contrasted according to the tests considered. Indeed, Fu's Fs and Tajima's tests were insignificant indicating demographic stability, whereas *R2* parameter was significant. Overall, the Tunisian pistachio has a stable population.

3.4. Population differentiation

Using the combined sequences, AMOVA was tested to detect population differentiation. AMOVA of Tunisian pistachio revealed that 48.48% (Φ_{CT} = 0.608, P < 0.05) of the genetic variation was observed among the two groups suggested by the haplotype networks (El-Guetar vs. Gafsa/Sidi Bouzid). Intrapopulation and interpopulation variances within these groups were also significant (Table 4). Although the maximum variance was associated at the intrapopulation level for two new tested groups (South

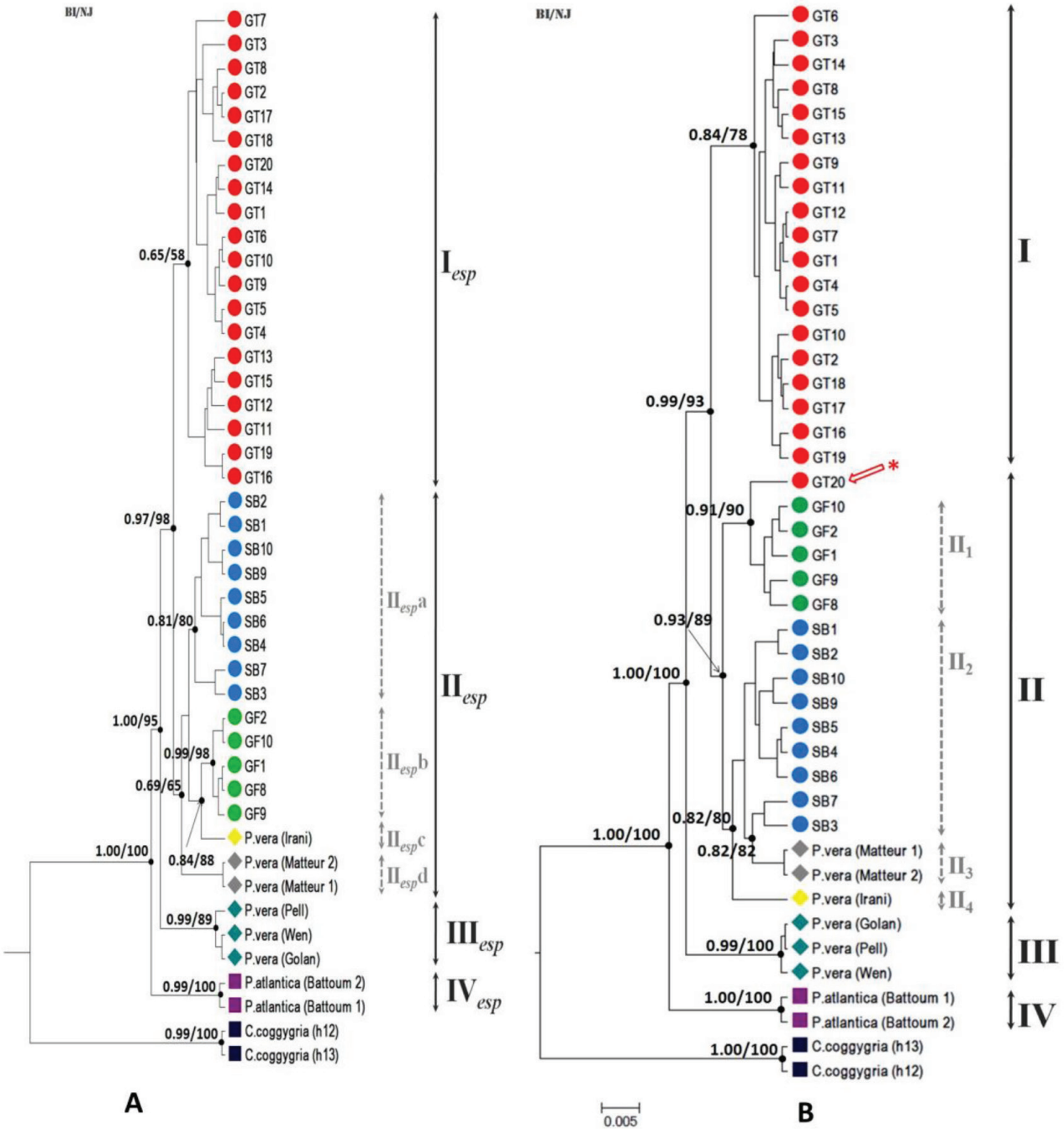

Figure 1. Phylogenetic reconstruction among *trn*L-*trn*F spacer haplotypes (A) and the combined *trn*L-F region sequences (B).

vs. Central Tunisia), the variance associated with genetic differentiation was only 15.22% (Φ_{CT} = 0.072, p < 0.05) (Table 4). This value was significant but was three times lower than the one found in the first pool. The entire pairwise comparisons of populations based on F_{ST} and Φ_{ST} were significant (Table A1). The F_{ST} values were very low between population pairs and do not exceed 0.049. The F_{ST} between Gafsa and Sidi-Bouzid did not differ from zero; indicating that these two populations were not well differentiated. Conversely, intermediate genetic values were recorded between Gafsa and El-Guetar and between El-Guetar and Sidi-Bouzid. Φ_{ST} values (0.408–0.723) were greater than those of F_{ST}. The highest values were obtained between El-Guetar/Gafsa and El-Guetar/Sidi-Bouzid. The Φ_{ST} in Gafsa and Sidi-Bouzid appeared small compared to others comparisons. Gafsa and Sidi-Bouzid were not very differentiated between them but they were genetically distant from the El-Guetar population. To illustrate the

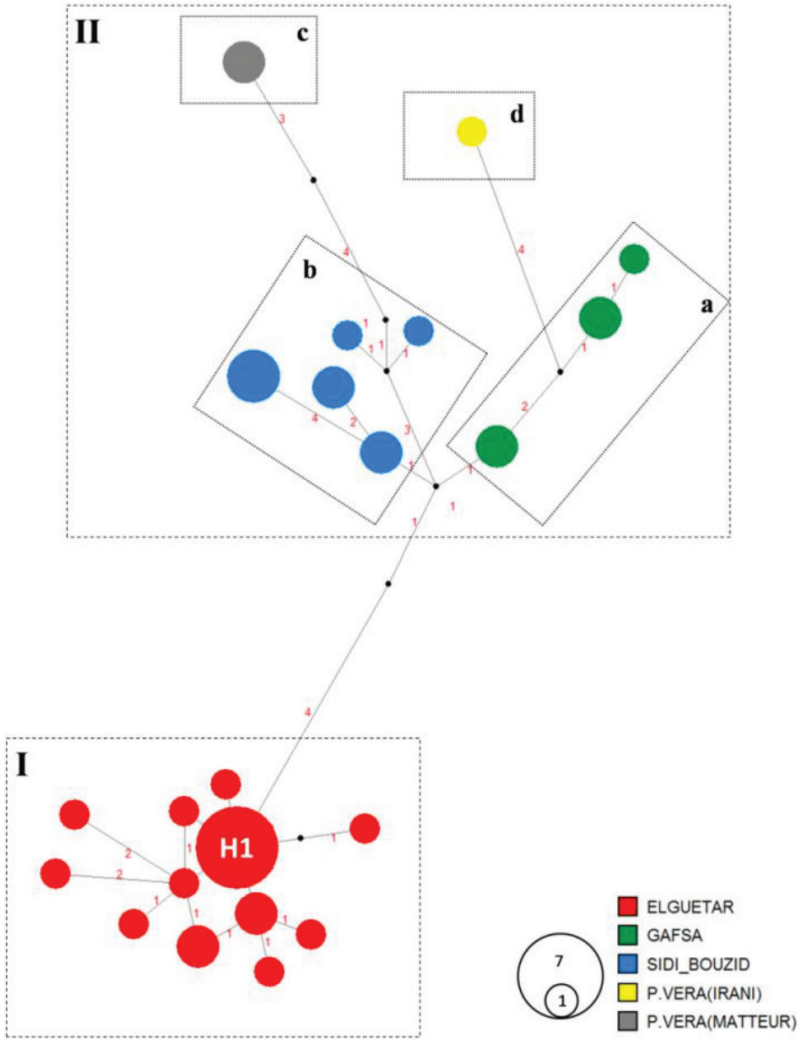

Figure 2. Median-joining network of the haplotypes inferred from the *trn*L-*trn*F marker. Nodes are proportional to haplotypes frequencies and branch lengths are proportional to the number of mutations.

Table 3. Demographic parameters and neutrality. Statistical significance: * P < 0.05.

***trn*L-*trn*F spacer**								
Geographic origin	*D*	*P*	*Fs*	*P*	*R2*	*P*	*Rg*	*P*
El-Guetar	−1.72	0.54	−3.01	0.51	0.08	0.00*	0.02	0.12
Gafsa	−1.17	0.58	−2.23	0.39	0.15	0.37	0.05	0.54
Sidi-Bouzid	−1.77	0.53	−3.82	0.49	0.10	0.00*	0.09	0.68
Global	−1.71	0.09	−6.02	0.00*	0.06	0.19	0.01	0.01*
Combined: *trn*L-*trn*F spacer and *trn*L intron								
El-Guetar	−1.72	0.51	−3.38	0.45	0.09	0.00*	0.02	0.00*
Gafsa	−1.66	0.52	−3.55	0.43	0.09	0.00*	0.02	0.00*
Sidi-Bouzid	−1.71	0.54	−3.55	0.46	0.10	0.00*	0.02	0.00*
Global	−1.74	0.52	−3.24	0.42	0.10	0.00*	0.02	0.00*

Tajima test: (D), Fu's Fs test: (Fs), Ramos-Onsins and Rozas test: (R2), Probability (P), raggedness index (*rg*).

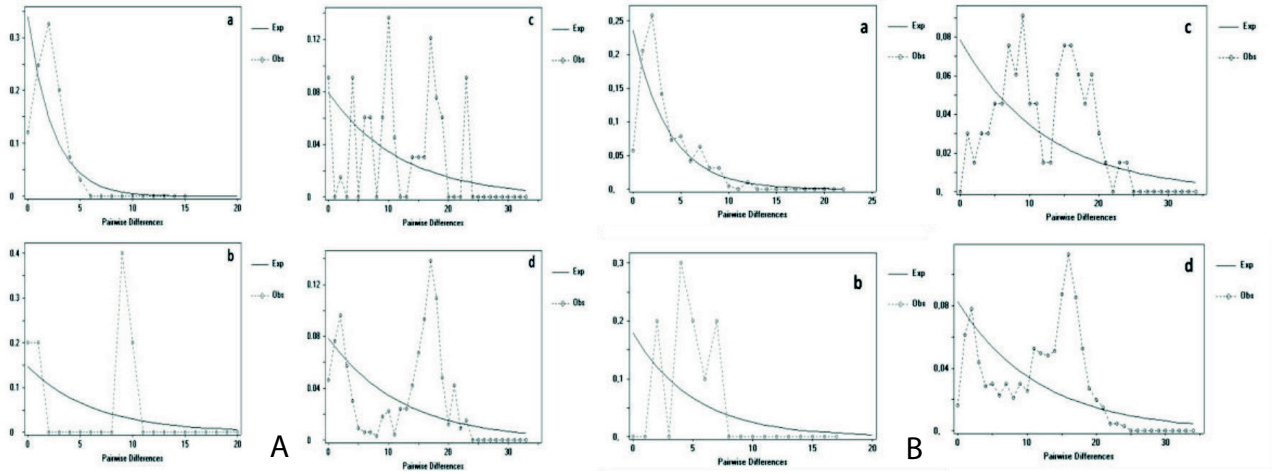

Figure 3. Pairwise mismatch distribution among *trn*L-*trn*F spacer (A) and the combined *trn*L-F region sequences (B). **a**: El-Guetar, **b**: Gafsa, **c**: Sidi-Bouzid, and **d**: Mismatch Global.

Table 4. AMOVA of the pistachio populations, * P < 0.05.

AMOVA groups	**Fixation index**	**Percentage of variance (%)**		
		Intergroup	Among populations within groups	Intrapopulation
El-Guetar vs. Gafsa and Sidi-Bouzid (groups inferred by the haplotype networks)	Φ_{SC}= 0.608 Φ_{ST}= 0.688 Φ_{CT}= 0.203	48.48*	20.36*	31.16*
El-Guetar and Gafsa vs. Sidi-Bouzid (South vs. Center of Tunisia)	Φ_{SC}= 0.662 Φ_{ST}= 0.686 Φ_{CT}= 0.072	15.22*	41.43*	43.35*

degree of haplotypic differences between populations and geographic areas, MDS analysis was performed on Φ_{ST} and F_{ST} parameters (Figure A1). Clearly, axis 1 (x-axis) separated El-Guetar population from the haplogroup formed by Gafsa and Sidi-Bouzid. MDS showed a significant genetic differentiation, compatible with AMOVA grouping (El-Guetar vs. Gafsa and Sidi-Bouzid).

3.5. Gene flow in Tunisian pistachio

The isolation with migration rates analyses between localities, using *IMa2* (Figure 4, Table A2), revealed that the best-supported model was bidirectional gene flow ($2Nm_{1-->2}$ = 3.87 and $2Nm_{2-->1}$ = 3.73) between Gafsa and Sidi-Bouzid. The gene flow rate was lower from El-Guetar to Gafsa ($2Nm_{1-->2}$ = 1.42) and from El-Guetar to Sidi-Bouzid ($2Nm_{2-->1}$ = 1.69). Gene flow can hinder genetic diversity by inhibiting genetic drift and natural selection from keeping local genetic differences.

4. Discussion and conclusion

The cytoplasmic genome is highly conserved for size, sequence, and order. The chloroplastic genome is widely used for interspecific genetic diversity studies, given the high conservation of its structure within a species (Palmer, 1986). This conclusion has been tested in 125 species, with regulatory sequences (Jansen et al., 2011) and intergenic spacers (Saski et al., 2007). However, some regions of the chloroplastic genome show extremely high polymorphism in several species; the case of the *rpl*22 sequences (Jansen et al., 2011), the *atp*B-*rbc*L, *mat*K, *rlp*16, *trn*L-*trn*F, and *trn*H-*trn*K regions (Reales et al., 2010; Batnini et al., 2014).

In this context, our study investigated the development of the *trn*L-*trn*F intergenic spacer and the combined *trn*L-F to disclose polymorphism in *Pistacia vera* L. The intergenic spacer *trn*L-*trn*F had an average length of 421 bp for all studied accessions. Comparable sizes have been reported in the genus *Pinus* and *Ficus* species (426–471 pb and 430–474 bp, respectively) (Chen et al., 2002; Baraket et al., 2009). Chloroplast combined sequences (*trn*L inton and *trn*L-*trn*F intergenic spacer) were examined in multiple studies which showed variable sizes according to the species. The sequencing generated an alignment of 927 bp. Numerous Angiosperms taxa registered similar size such

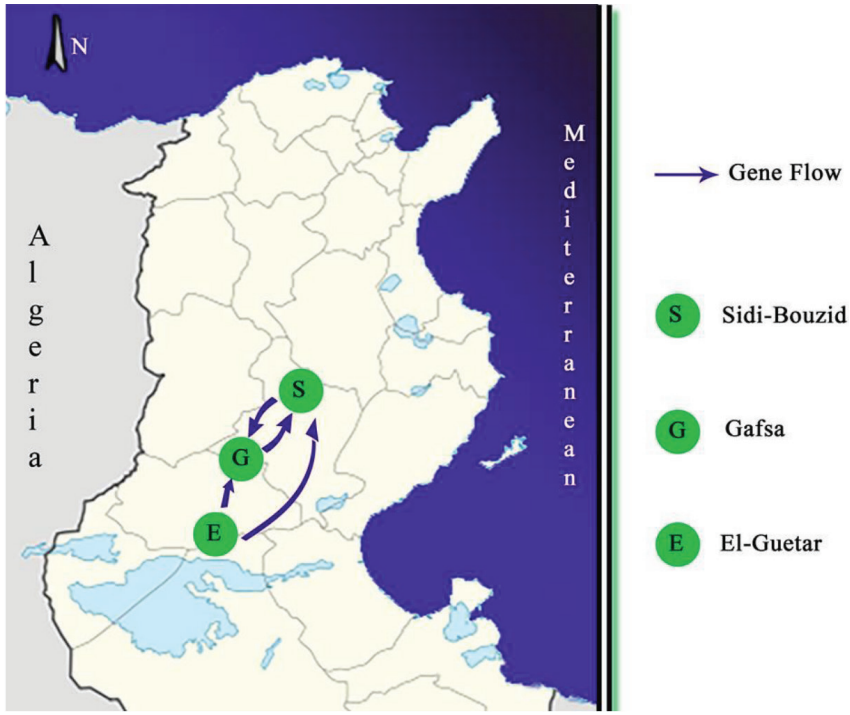

Figure 4. Graphic representation of gene flows between different populations.

as *Pinus* (866 bp), *Araucaria* (961 bp), and *Ceratophyllum* (971 bp) (Borsch et al., 2003). The transition/transversion (ti/tv) ratio has been commonly computed in genetic studies; this ratio is helpful to better understand the patterns of cpDNA sequence evolution. However, an inverse relationship existed between the genetic distance and the observed value of R (ti/tv). The ratio R was high at low rates of divergences, whereas it becomes lower for strong divergences (Purvis and Bromham, 1997). The calculated ti/tv ratios for the chloroplastic markers did not exceed 0.7. This conclusion agrees with those of several studies of angiosperm species, which had ti/tv ratios of less than 1 (Bakker et al., 2000; Baraket et al., 2009; Batnini et al., 2014; Benabid et al., 2014).

Species are usually subdivided into populations in which the allelic and genotypic frequencies differ from one region to another. This variation and its origin were the objects of population genetics studies (Ramanatha-Rao and Hodgkin, 2002). Cytoplasmic markers, relative to nuclear markers, usually showed a significant geographic distribution in both plants and animals (Ennos, 1994). Pistachio trees, generally encountered in arid and semiarid climates of Tunisia, shaped a distinct pattern that was further differentiated according to geographic origin. Phylogenetic trees, the AMOVA, as well as haplotypic networks, advanced this conclusion. In fact, the studied populations were always subdivided into two major clusters; "El-Guetar" and "Gafsa/Sidi-Bouzid". The cluster Gafsa/Sidi-Bouzid was heterogeneous, forming distinct geographical subgroups. Guenni et al. (2016) have observed similar results, where El-Guetar's varieties were often differentiated from other populations. On the other hand, the different accessions of *P. atlantica* formed a clade in the *trn*L-F data, which is consistent with previous analyses: plastid and nuclear DNA data sets all suggested that *P. atlantica* formed a monophyletic group with *P. vera* (Yi et al., 2008; Xie et al., 2014). Our results indicating that 'Wen' and 'Golan' (from the USA) varieties are most closely related to germplasm from Asia (Pell) than Tunisian ones. A similarity data obtained using the RAPD analysis confirmed this classification (Hormaza et al., 1994). Hormaza et al. (1994) related this finding to the hypothesis that the American cultivars derive from seeds originating from the Caspian/Caucasus region and carried to California by immigrants from that area.

AMOVA, applied for the two groups inferred by the haplotype networks (El-Guetar vs. Gafsa-Sidi/Bouzid), showed that the majority of the variance was at the intergroup field. Despite that, the two megapopulations: El-Guetar and Gafsa/Sidi-Bouzid showed large intrapopulation diversity. This confirmed a substantial interindividual diversity of the studied pistachios. Similar results have already been reported in Iranian pistachio nuts (Pazouki et al., 2010) and also in Tunisian pistachio trees using nuclear markers (SRAP) (Guenni et al., 2016).

Moreover, the Φ_{ST} and F_{ST} genetic distance matrices per pair of populations, as well as MDS illustration, were

in line with phylogenetic analyses and the AMOVA. Two significantly differentiated haplogroup classes (**I** and **II**) were inferred by the used cpDNA regions. Different interpretations could be proposed to explain this conclusion. Population polymorphisms were generally correlated with the accumulated mutations by an advantageous selection (Castric et al., 2008), whereas the DNA regions, selected for this analysis, were a neutral molecular marker (intron and spacer). Thus, the observed dimorphism cannot be the result of a natural selection. Moreover, in this research, Tajima's D test was negative and nonsignificant. In fact, negative Tajima's D values were more correlated to demographic factors (like the population expansion) than other selective events (Beck et al., 2008).

Secondly, the cytoplasm dimorphism could be related to the breeding methods practiced in horticulture, which was a major key to determine the differentiation between the prospected areas (Ramanatha-Rao and Hodgkin, 2002). Vegetative reproduction allows, in allogamous plant species, to maintain elite genotypes of agronomic interest for the next generation and to avoid the disadvantages of Mendelian segregation associated with sexual reproduction (Zohary and Hopf, 2000). Consequently, the existences of multiple domestication events caused, certainly, the structuring of populations in numerous plant species (McKey et al., 2010). Another explication of the chloroplast dimorphism is factors of evolution, like the gene flow migratory model. Within a species, genetic diversity generally has geographic variation (Ramanatha-Rao and Hodgkin, 2002). This variation is the result of the equilibrium between evolutionary forces tending to generate a divergence between populations and those tending to create a genetic homogeneity. The forces creating homogenization are gamete or variety movements, these factors contribute to gene flow between populations.

Gene flow estimates maintain this observation, the number of migrants per generation *2Nm* value was elevated among these populations (Sidi-Bouzid and Gafsa), signifying high connectivity between them. High *2Nm* estimates were generally related to high levels of genetic variability (Wade and McCauley, 1988). The prospected geographic areas involved a very interesting gene flow model, which was an important mechanism for transferring genetic diversity among populations. The proximity on the regional scale of Gafsa and Sidi-Bouzid may explain this conclusion. Nevertheless, genetic diversity was higher than genetic differentiation. The F_{ST} and Φ_{ST} values supported this connection; these estimators were negligible among Sidi-Bouzid and Gafsa.

Gene flow is a process that occurs both in time and in space, through pollen, seeds, and other propagules which strongly interacts with the local farming systems (Hamrick et al., 1993). Theoretically, in *Pistacia vera* as angiosperm species, markers of chloroplast genome are only maternal transmission markers. Consequently, seeds are implicated in this gene transfer between populations (Heredia and Ellstrand, 2014), while pollen grains have no role in the observed migration pattern.

Practically, in cultivated Tunisian pistachio, seeds are rarely exchanged among growing regions, for the reason of the large use of vegetative propagation (Maggs, 1973). In these localities, female varieties are exchanged in order to provide a guarantee of production in unfavorable growing areas.

In contrast, 2*Nm* values were lower between population's pairs El-Guetar/Sidi-Bouzid and El-Guetar/Gafsa. In fact, gene flow was generally restricted by geographic barriers. El-Guetar oasis had a specific geographic location and geomorphologic limits, which make it a relatively isolated area. The Mountain chain 'Orbata' (a national park of 5700 ha and 1165 m) is located in the east of the Gafsa city, between El-Guetar and Gafsa farms. In a natural way, this geographic factor could be a crucial barrier to the long-distance dispersal of propagules (pollen, seeds). For *P. vera*, natural pollination is most likely exclusively anemophilic, honey bees play no role in pollination due to the absence of nectar in female flowers (Zuang et al., 1988). However, it has been noted that some Diptera insects sometimes contribute to the pollination of this species (Evreinoff, 1955). This physical barrier could avoid gene exchange among areas situated on both sides of the mountains. Indeed, the oasis of El-Guetar is structured around two geomorphologic compartments; the mountainous terrain of the Orbata range and Chott El-Guetar on an area of 3730 ha. The oasis benefits of particular climatic conditions with regard to the direction of dominant winds, since it is naturally protected in the North by the Jebel Orbata. Thus, El-Guettar Pistachio belongs to the group of mountain oases. They are closely associated with specific geology, geomorphology, and water supply with the accumulation of salts in the environment (Hachicha and Ben Aissa, 2014). The Chott acts as an evaporative system of various salts, including gypsum, which is scattered by the wind while the more soluble salts (NaCl, $MgSO_4$, Na_2SO_4) concentrate on the soil. These salts, during massive rains, were invading the lower parts of the oasis (Job, 1992).

The null hypothesis of molecular evolution (Kimura, 1968), and specifically, the neutral theory gives precise predictions about patterns and structure of sequence variation expected under the null hypothesis. This hypothesis suggested that genetic polymorphisms may not always have effects on phenotypes. The *Tajima' D* statistics, Fu's Fs and *R2*, supported the hypothesis of a population in a stable demographic state. Mismatch distributions

were multimodal with significant values of the raggedness index (*rg*). The findings were coherent with the suggestion of the demographic stability (Rogers and Harpending, 1992) of the Tunisian pistachio population. The absence of demographic expansion signatures seems to be mainly due to the large divergence between inferred haplogroups. Mismatch distribution and haplotype network shape (star-like pattern), for El-Guetar, propose a recent population expansion or a selective sweep in this locality. Therefore, an excess of singletons (Fu and Li, 1993; Tajima, 1989) and of haplotypes (Fu, 1997) can be the cause of a demographic expansion.

The pistachio tree is one of the symbols of horticulture and landscapes of the Mediterranean basin. The rich heritage of pistachio in Tunisia proves the great potential for a local selection. In an ecosystem already modified by human activity, many conservation strategies must be applied. These molecular tools will significantly promote breeding programs, genetic conservation, and management of the species.

Acknowledgment
This work was supported by grants from the Tunisian Ministry of Higher Education and Scientific Research.

References

Abdedaeim S (2015). Plan d'action de la stratégie de développement durable des oasis en Tunisie. Ministère de l'environnement et de développement durable Direction générale de l'environnement et de la qualité de la vie. p. 47 (in French).

Addinsoft (2007). XLSTAT, Analyse de données et statistique avec MS Excel. AddinSoft, New York, USA (in French).

Altschul SF, Thomas LM, Alejandro AS, Jinghui Z, Zheng Z et al. (1997). Gapped BLAST and PSI-BLAST a new generation of protein database search programs. Nucleic Acids Research 25: 3389-3402.

Baghizadeh A, Noroozi SH, Jalali-Javaran M (2010). Study on genetic diversity of some Iranian Pistachio (*Pistacia vera* L.) cultivars using random amplified polymorphic DNA (RAPD), inter sequence repeat (ISSR) and simple sequence repeat (SSR) markers: A comparative study. African Journal of Biotechnology 9(45): 7632-7640.

Bakker FT, Culham A, Gomez Martinez R, Carvalho J, Compton J et al. (2000). Patterns of nucleotide substitution in angiosperm cpDNA *trn*L (UAA)-*trn*F (GAA) regions. Molecular Biology and Evolution 17: 1146-1155.

Bandelt HJ, Forster P, Rohl A (1999). Median-joining networks for inferring intraspecific phylogenies. Molecular Biology and Evolution 16: 37-48.

Baraket G, Olfa S, Chatti K, MarsM, Marrakchi M, Trifi M et al. (2009). Chloroplast DNA analysis in Tunisian fig cultivars (*Ficus carica* L.): sequence variations of the *trn*L-*trn*F intergenic spacer. Biochemical Systematics and Ecology 36(11): 828-835.

Batnini MA, Bourguiba H, Trifi-Farah N, Krichen L (2014). Étude de la diversité de l'ADN chloroplastique chez l'abricotier (*Prunus armeniaca* L.). Revue des Régions Arides - Numéro Spécial - n° 35 (3/2014) - Actes du 4ème Meeting International Aridoculture et Cultures Oasisennes: Gestion des Ressources et Applications Biotechnologiques en Aridoculture et Cultures Sahariennes : perspectives pour un développement durable des zones arides 17-19/12/2013 (in French).

Beck JB, Schmuths H, Schaal BA (2008). Native range genetic variation in *Arabidopsis thaliana* is strongly geographically structured and reflects Pleistocene glacial dynamics. Molecular Ecology 17: 902-915.

Benabid A (2000). Flore et Écosystèmes Du Maroc: Évaluation et préservation de la biodiversité, ed. Paris: Ibis Press (in French).

Borsch T, Hilu KW, Quandt D, Wilde V, Neinhuis C et al. (2003). Non-coding plastid *trn*T-*trn*F sequences reveal a well resolved phylogeny of basal angiosperms. Journal of Evolutionary Biology 16: 558-576.

Castric V, Bechsgaard J, Schierup MH, Vekemans X (2008). Repeated adaptive introgression at a gene under multiallelic balancing selection. PLoS Genetics 4(8): e1000168.

Chatti K, Choulak S, Guenni K, Salhi-Hannachi A (2017). Genetic diversity analysis using morphological parameters in Tunisian Pistachio (*Pistacia vera* L.). Journal of Research in Biological Sciences 2: 29-34.

Chen J, Tauer CG, Huang Y (2002). Paternal chloroplast inheritance patterns in pine hybrids detected with *trn*L-*trn*F intergenic region polymorphism. Theoretical and Applied Genetics 104: 1307-1311.

Choulak S, Chatti K, Marzouk Z, Chatti, N (2017). Genetic differentiation and gene flow of some Tunisian pistachio (*Pistacia vera* L.) varieties using chloroplastic DNA. Journal of Research in Biological Sciences 2: 35-41.

Choulak S, Rhouma-Chatti S, Marzouk Z, Said K, Chatti N, Chatti K (2015). Chloroplast DNA analysis of Tunisian pistachio (*Pistacia vera* L.): Sequence variations of the intron *trn*L (UAA). Scientia Horticulturae191: 57-64.

Ennos R (1994). Estimating the relative rates of pollen and seed migration among plant populations. Heredity 72: 250-259.

Evreinoff VA (1955). Le Pistachier. Étude pomologique. Journal d'agriculture tropicale et de botanique appliquée 2(7): 387-415 (in French).

Excoffier L, Smouse P, Quattro J (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. Genetics 131: 479-491.

Excoffier L, Lischer HEL (2010). Arlequin suite version 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources 10: 564-567.

Faostat (2017). http://www.fao.org/faostat/en/data/QC.

Farès K, Guasmi L, Touil T, Ferchichi A (2009). Genetic diversity of pistachio tree using inter-simple sequence repeat markers ISSR supported by morphological and chemical marker. Biotechnology 8(1): 24-34.

Fu YX (1997). Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics 147: 915-925.

Fu YX, Li WH (1993). Statistical tests of neutrality of mutations. Genetics 133: 693-709.

Ghrab M, Zribi F, Ben Mimoun M, Rhouma A (2012). Morphological investigation of genetic diversity of pistachio (*Pistacia vera* L.) germplasm in arid land of Tunisia. Plant Ecology and Evolution 145(3): 363-372.

Golan-Goldhirsh A, Barazani O, Wang ZS, Khadkal DK, Saunders JA et al. Genetic relationships among Mediterranean *Pistacia* species evaluated by RAPD and AFLP markers. Plant Systematics and Evolution 246: 9-18.

Guenni K, Aouadi M, Chatti K, Salhi-Hannachi A (2016). Analysis of genetic diversity of Tunisian pistachio (*Pistacia vera* L.) using sequence-related amplified polymorphism (SRAP) markers. Genetics and Molecular Research 15: 166-175.

Hachicha M, Ben Aissa I (2014). Managing salinity in Tunisian oases. Life Science Journal 8(9): 775-782.

Hall TA (1999). Bioedit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/ NT. Nucleic Acids Symposium Series 41: 95-98.

Hamrick JL, Murawski DA, Nason JD (1993). The influence of seed dispersal mechanisms on the genetic structure of tropical tree populations. In: Fleming TH, Estrada A. (eds) Frugivory and seed dispersal: ecological and evolutionary aspects. Advances in vegetation science, vol 15. Springer, Dordrecht.

Heredia SM, Ellstrand NC (2014). Novel seed protection in the recently evolved invasive, California wild radish, a hybrid *Raphanus* sp. (Brassicaceae). American journal of botany 101(12): 2043-2051.

Hey J (2010). Isolation with migration models for more than two populations. Molecular Biology and Evolution 27: 905.

Hormaza JI, Dollo L, Polito VS (1994). Determination of relatedness and geographical movements of *Pistacia vera* (Pistachio; Anacardiaceae) germplasm by RAPD analysis. Economic Botany 48(4): 349-358.

IPGRI (1997). Descriptors for Pistachio (*Pistacia vera* L.). International Plant Genetic Resources Institute, Rome, Italy.

Jansen RK, Saski C, Lee SB, Hansen AK, Daniell H (2011). Complete Plastid Genome Sequences of three Rosids (*Castanea, Prunus, Theobroma*): Evidence for at least two independent transfers of *rpl*22 to the Nucleus. Molecular Biology and Evolution 28(1): 835-847.

Job JO (1992). Les sols salés de l'oasis d'El Guettar. Thèse de doctorat. Université de Montpellier II, 150 p (in French).

Katsiotis A, Hagidimitriou M, Drossoul A, Pontikis C, Loukas M (2003). Genetic relationships among species and cultivars of *Pistacia* using RAPDs and AFLPs. Euphytica 132: 279-286.

Kimura M (1968). Evolutionary rate at the molecular level. Nature 217 (5129): 624-626.

Librado P, Rozas J (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25: 1451-1452.

Maggs DH (1973). Genetic resources in pistachio. Plant Genetic Resources Newsletter 29: 7-15.

McKey D, Elias M, Pujol B, Duputie A (2010). The evolutionary ecology of clonally propagated domesticated plants. New Phytologist 186: 318-332.

Mlika M (1980). Contribution à l'étude du pistachier en Tunisie : Choix de variétés mâles et femelles à floraison synchrone-Anatomie des fleurs. Mémoire de fin d'études du cycle de spécialisation INAT, Tunisie (in French).

Morton BR, Clegg MT (1993). A chloroplast DNA mutational hotspots and gene conversion in a non-coding region near *rbc*l in the grass family (Poaceae). Current Genetics 24: 357-365.

Palmer JD, Stein DB (1986). Conservation of chloroplast genome structure among vascular plants. Current Genetics 10: 823-833.

Pazouki L, Mardi M, Salehi-Shanjani P, Hagidimitriou M, Pirseyedi SM et al. (2010). Genetic diversity and relationships among *Pistacia* species and cultivars. Conservation Genetics 11: 311-318.

Posada D (2008). jModel Test: phylogenetic model averaging. Molecular Biology and Evolution 25(7): 1253-1256.

Purvis A, Bromham L (1997). Estimating the transition / transversion ratio from independent pairwise comparison with an assumed phylogeny. Journal of Molecular Evolution 44: 112-119.

Ramanatha-Rao V, Hodgkin T (2002). Genetic diversity and conservation and utilization of plant genetic resources. Plant Cell, Tissue and Organ Culture 68: 1-19.

Ramos-Onsins SE, Rozas J (2002). Statistical properties of new neutrality tests against population growth. Molecular Biology and Evolution 19: 2092-2100.

Ray N, Currat M, Excoffier L (2003). Intra-Deme molecular diversity in Spatially expanding populations. Molecular Biololgy and Evolution 20: 76-86.

Reales A, Sargent DJ, Tobutt KR, Rivera D (2010). Phylogenetics of Eurasian plums, *Prunus* L. section *Prunus* (Rosaceae), according to coding and non-coding chloroplast DNA sequences. Tree Genetics and Genomes 6: 37-45.

Risterucci AM, Grivet L, N'goran J, Pieretti I, Flament MH, Lanaud C (2000). A high-density linkage map of *Theobroma cacao* **L.** Theoretical and Applied Genetics 101(5)**:** 948-955.

Rogers AR, Harpending H (1992). Population growth makes waves in the distribution of pairwise genetic differences. Molecular Biology and Evolution 9: 552-569.

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A et al. (2012). MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. Systematic Biology 61(3): 539-542.

Saitou N, Nei M (1987). The Neighbor-Joining method: A new method for reconstructing phylogenetic Trees. *Molecular Biology and Evolution* **4**(4)**:** 406-425

Sambrook J, Frithsch EF, Maniatis T (1989). Molecular Cloning: A Laboratory Manual. 2nd ed. New York, NY, USA: Cold Spring Harbor Laboratory.

Saski C, Lee SB, Fjellheim S, Guda C, Jansen RK et al. (2007). Complete chloroplast genome sequences of *Hordeum vulgare*, *Sorghum bicolor* and *Agrostis stolonifera* and comparative analyses with other grass genomes. Theoretical and Applied Genetics 115: 571-590.

Tajima F (1989). Statistical testing for the neutral mutation hypothesis by DNA polymorphism. Genetics 123: 585-595.

Talebi M, Kazemi M, Sayed-Tabatabaei BE (2012). Molecular diversity and phylogenetic relationships of *Pistacia vera*, *Pistacia atlantica* subsp. *mutica* and *Pistacia khinjuk* using SRAP markers. Biochemical Systematics and Ecology 44: 179-185.

Talebi M, Akbari M, Zamani M, Sayed-Tabatabaei BE (2016). Molecular polymorphism in *Pistacia vera* L. using non-coding regions of chloroplast DNA. Journal of Genetic Engineering and Biotechnology 14: 31-37.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. Molecular Biology and Evolution 30: 2725-2729.

Wade MJ, McCauley DE (1988). Extinction and recolonization: Their effects on the genetic differentiation of local populations. Evolution 42: 995-1005.

Weeks A, Zapata F, Pell SK, Daly DC, Mitchell JD et al. (2014). To move or to evolve contrasting patterns of intercontinental connectivity and climatic niche evolution in "Terebinthaceae" (Anacardiaceae and Burseraceae). Frontiers in Genetics 5: 409.

Wright S (1931). Evolution in mendelian populations. Genetics 16: 97-159.

Xie L, Yang ZY, Wen J, Li DZ, Yi TS (2014). Biogeographic history of *Pistacia* (Anacardiaceae), emphasizing the evolution of the Madrean-Tethyan and the eastern Asian-Tethyan disjunctions. Molecular Phylogenetics and Evolution 77: 136-146.

Yi T, Wen J, Golan-Goldhirsh A, Parfitt DE (2008). Phylogenetics and reticulate evolution in*Pistacia* (Anacardiaceae). American Journal of Botany 95: 241-251.

Ziya-Motalebipour E, Kafkas S, Khodaeiaminjan M, Çoban N, Gözel H (2016). Genome survey of pistachio (*Pistacia vera* L.) by next generation sequencing: Development of novel SSR markers and genetic diversity in *Pistacia* species. BMC Genomics 17(1): 998.

Zohary D (1996). The genus *Pistacia* L. In Taxonomy, distribution, conservation and uses of *Pistacia* genetic resources, Padulosi S, Caruso T, Barone E (eds) (IPGRI, Palermo), pp. 1-11.

Zohary D, Hopf M (2000). Domestication of Plants in the Old World. 3rd ed. Oxford, UK : Oxford University Press.

Zuang H, Barret P, Beau C (1988). Nouvelles espèces fruitières. CTIFL, Paris, p. 182 (in French).

Table A1. F_{ST} (above the diagonal) and Φ_{ST} (below the diagonal) between different populations, * P < 0.05.

	El-Guetar	Gafsa	Sidi-Bouzid
El-Guetar	-	0.033*	0.049*
Gafsa	0.723*	-	0.000*
Sidi-Bouzid	0.699*	0.408*	-

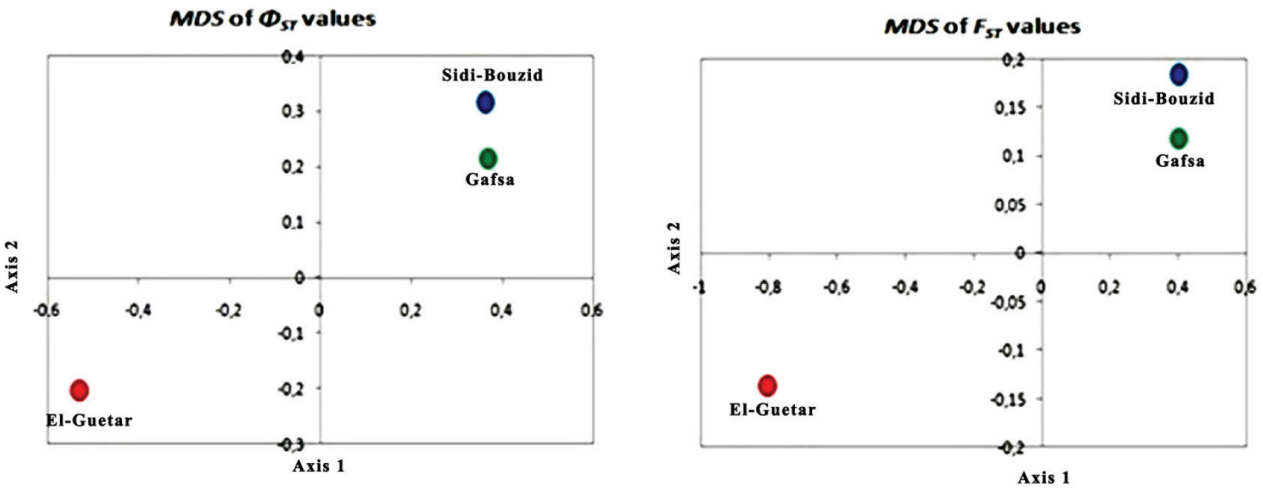

Figure A1. Nonparametric multidimensional scaling analyses (MDS) based on F_{ST} and Φ_{ST} distances between populations for *trn*L-F marker.

Table A2. Isolation template settings with migration from the combined dataset. Confidence intervals are given in parentheses. Ne: effective population size; 2NM: migration parameter in number of gene copies per generation.

Population 1	Population 2	Ne_1	Ne_2	$2NM_{1-->2}$	$2NM_{2-->1}$
El-Guetar	Sidi-Bouzid	2607	322	1.42	0
				(0–8)	(0–1)
El-Guetar	Gafsa	1178	411	1.69	0
				(0–10)	(0–1)
Sidi-Bouzid	Gafsa	1392	668	3.87	3.73
				(0–33)	(0–26)