GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

Cytoplasmic polymorphism and evolutionary history of plum cultivars: Insights from chloroplast DNA sequence variation of *trn*L-*trn*F spacer and aggregated *trn*L intron & *trn*L-*trn*F spacer

S.B. Mustapha, H. Ben Tamarzizt, G. Baraket, D. Abdallah and A. Salhi-Hannachi

Laboratoire de Génétique Moléculaire, Immunologie & Biotechnologie, Faculté des Sciences de Tunis. Université Tunis El Manar, Tunis, Tunisie

Corresponding author: A.S. Hannachi
E-mail: Amel.SalhiHannachi@fsb.rnu.tn

Genet. Mol. Res. 14 (2): 3964-3979 (2015)
Received May 28, 2014
Accepted October 10, 2014
Published April 27, 2015
DOI http://dx.doi.org/10.4238/2015.April.27.11

ABSTRACT. We screened for polymorphisms of the non-coding region of plastid DNA in plum trees. Sequencing data from the *trn*L-*trn*F chloroplast region were used to reveal a pattern of diversity, establish phylogenetic relationships, and test the selection pressure or evolutionary demography scenario for plastome DNA. The size of the non-coding regions varied from 398 to 563 and 865 to 1084 bases pairs for the *trn*L-*trn*F spacer and combined sequences, respectively. The average GC contents were 33.8 and 34.4% in the spacer and pooled sequences, respectively. Genetic distances calculated within the plums were 0.077 and 0.254, on average, for the *trn*L spacer and combined sequences, respectively. The neighbor-joining trees showed clustering relationships among cultivars that were independent of their geographic origins and designations. The neutrality tests and site-frequency spectra

indicated that spacer and pooled sequences fit the neutral theory model at equilibrium between mutation and genetic drift and reject the hypothesis of a recent demographic expansion. The mismatch distribution shows variation patterns, thus providing evidence of an important genetic diversity explained by an excess of intermediate variants that occurred in the sequences analyzed. Further implications of the findings with regard to plum germplasm management and its utilization in breeding programs are also discussed.

Key words: Chloroplast DNA; Plum; Genetic diversity; *trn*L-*trn*F spacer; *trn*L intron

INTRODUCTION

The genus *Prunus*, originating in the northern hemisphere, comprises >400 species of trees and shrubs that typically have 8 chromosomes (Bouhadida et al., 2007). Plums are cultivated for their fruits and consumed both in their fresh and processed forms (Ahmed et al., 2004). Plums consist of >20 species characterized by variations in morphological and pomological traits. Natural hybridization is frequent in the genus *Prunus*. *Prunus domestica* L. (6X), a European plum, is derived from the natural hybridization between *P. spinosa* L. (4X) and *P. cerasifera* Ehrh (2X). As suggested by Okie and Weinberg (1996), the appellation of the Japanese plum was given for the diploid plum *P. salicina* Lindl. (2X). In Tunisia, the germplasm is represented by several plum cultivars, including very old ones of unknown origin. These varieties do not originate from breeding programs and are extended by exchanges between farmers as traditional cultural practices. These cultural practices led to duplicate and misidentifications among trees, which are difficult to distinguish based on morphological and pomological characteristics. In addition, commercial production is usually concerned with 1-2 introduced species recognized for their appreciated fruit quality by consumers, performance, and productivity (e.g., Santa Rosa and Golden Japan cultivars). The study of genetic diversity is essential for the development of efficient conservation strategies for plum trees. Estimating the level of genetic diversity and establishing relationships among cultivated species are important for identifying local gene pools and creating conservation and management strategies. Since the 1980s, variation in the plastid genome has been used to establish genetic relationships among species and is considered a barcode system in plants (Palmer, 1987; Dowling et al., 1990; Clegg and Zurawski, 1992). In fact, chloroplast DNA (cpDNA) has several interesting characteristics such as a non-Mendelian inheritance (Sun, 2002). The genetic information present in the plant chloroplatic DNA is of great importance in population genetics and molecular taxonomy. It is well known that the coding regions of cpDNAs of higher plants are highly conserved across species and genera. As noted by Clegg et al. (1991) and Zurawski and Clegg (1987), the non-coding regions exhibit a higher mutational rate than that of the coding regions. Taberlet et al. (1991) and Demesure et al. (1995) designed a pair of universal primers that amplify the coding and non-coding regions of the plastid genome to generate molecular markers. Several studies demonstrate divergence of chloroplast sequences and illustrate a considerable level of polymorphism across numerous non-coding sequences. Such regions can be explored through direct sequencing or restriction by enzymes to identify useful diversity across taxonomic levels. Moreover, these non-coding regions evolve quickly and offer high levels of variation by ac-

cumulating mutational events that can be used in molecular taxonomic investigations (Taberlet et al., 1991; Ohsako and Ohnishi, 2000, 2001; Yamane et al., 2003; Besnard, 2008).

Herein, we develop cytoplasmic molecular markers to evaluate the level genetic diversity of Tunisian plum resources using sequences of non-coding regions of the *trn*L-*trn*F spacer and combined *trn*L-*trn*F spacer & *trn*L intron sequences of plastid DNA and, thereby, determine if there is a definite correlation between genetic diversity and geographic origin. Analyses and exploration of molecular polymorphisms constitute crucial steps for the development of efficient conservation strategies and improvements in selection programs. In fact, the variation in cytoplasmic DNA became a complementary approach to explain nuclear DNA diversity that has been previously analyzed. Here, we explored the cpDNA variation and the objectives of this study were twofold: i) to give insight into the level and structuring of plum genetic diversity and ii) elucidate the molecular evolutionary history of plums.

MATERIAL AND METHODS

Plant materials

In this study, 23 cultivars of Tunisian plums were collected from 7 regions: 5 regions in northern Tunisia (i.e., Raf-Raf, Ras Djbel, Souinine, Douar Hamouda, and Ghar El Melh), and 2 from Cap Bon and Kairouan (Table 1). This consisted of 20 *P. salicina* cultivated plums (i.e., 17 local and 3 early introduced varieties) and 3 wild plums belonging to *P. insititia*.

Table 1. Plum cultivars investigated in the current study, and their localities and accession numbers.

	Cultivar name	Locality of origin	Ploidy level	Accession No.
Local cultivars	Badri 1	Ras Djebal	Diploid	GU129962
	AwinaSafra (JaponiaSafra)	Raf Raf	Diploid	GU129963
	Jinha	Raf Raf	Diploid	GU129960
	Ain Kounoulia	Raf Raf	Diploid	GU129964
	Cidre	Raf Raf	Diploid	GU129965
	Adam Hmam	Raf Raf	Diploid	GU129966
	Neb Zarouk	Raf Raf	Diploid	GU129967
	Hamda	Raf Raf	Diploid	GU129961
	Awina Hamra Badria	Raf Raf	Diploid	GU129968
	Ain Torkia	Raf Raf	Diploid	GU129951
	Awina Safra Morra	Raf Raf	Diploid	GU129959
	Miski Hamra	Raf Raf	-	GU129970
	Zaghouania	Raf Raf	Diploid	GU129948
	Ain Tastouria	Souinine	Diploid	GU129949
	Miski Kahla	Raf Raf	Diploid	GU129950
	Badri 2	Rass Djebel	Diploid	GU129953
	Hamra Badri	Kairouan	Diploid	GU129955
Introduced cultivars	Golden Japan	Raf Raf	Diploid	GU129969
	Santa Rosa	Raf Raf	Diploid	GU129952
	Stanley	Cap Bon	Diploid	GU129956
Wild plums	Sauvage RafRaf	Raf Raf	-	GU129954
	Chaarawiya	Ghar El Melh	Diploid	GU129957
	Zanou	Douar Hamouda	Diploid	GU129958

DNA extraction, PCR amplification, and sequencing

Total cellular DNA was extracted from each accession following the procedure described by Bernatzky and Tanskley (1986). The extracted DNA was used as a template in the

PCR reaction using the universal primers described by Taberlet et al. (1991). The cpDNA *trn*L-*trn*F spacer and the *trn*L intron were amplified using the primers designed by Taberlet et al. (1991). PCR reactions were conducted using a 25-µL PCR mixture comprised of 200 µL dNTP, 50 µL $MgCl_2$, PCR buffer (10X), 5U Taq polymerase, and 0.5 µM of each primer. PCR programs were organized as follows: one cycle of denaturation at 94°C for 4 min; and 35 cycles of 1 min denaturation at 94°C, 1 min annealing at 94°C, and 2 min elongation at 72°C; and a final elongation at 72°C for 10 min. Weights of DNA fragments obtained by PCR were estimated by electrophoresis on a 1.5% agarose gel using a 1-kb DNA ladder (Invitrogen, USA). Amplified DNA was purified using the PureLink PCR Purification Kit (Invitrogen) and directly sequenced using the 4 capillary DNA Sequencer Genetic Analyser 3130 (Applied Biosystems, USA). Sequences of the *trn*L-*trn*F spacer and the *trn*L intron were submitted for publication in GenBank (Table 1).

Data analysis

Sequences of the spacer and aggregated sequences of the spacer and intron were aligned with DAMBE (Xia, 2000) and analyzed by MEGA version 4.0.2 (Tamura et al., 2007). The lengths, and GC and AT contents for each sequence were assessed. Genetic distances of Tamura-2 were calculated between cultivars. Maximum parsimony (Mp) and neighbor joining (NJ) methods (Saitou and Nei, 1987), performed using the MEGA program (version 4.0.2) (Tamura et al., 2007), were applied to construct genetic relationships among plums.

To estimate the transition/transversion (ti/tv) ratio R, the following formula was used:

$$R = [A \times G \times K1 + T \times C \times K2]/[(A + G) \times (T + C)]$$

where A, G, C, T are the corresponding frequencies of four nucleotides and K1 and K2 the ratio for purine and pyrimidines, respectively. To estimate polymorphisms, the aligned sequences were examined in the Mega files using the DnaSP software version 4.0 (Rozas et al., 2003). Haplotype diversity indices, pairwise estimates of nucleotide divergence (π) (Jukes and Cantor, 1969), and the average of nucleotide differences (K) were calculated. To estimate the robustness of the branches founded in the parsimony analysis, 500 bootstrap replicates were performed (Felsenstein, 1993). Tajima's D (Tajima, 1993), Fu and Li's D and F, and Fu's Fs neutrality tests were estimated to detect signatures of selection or demographic evolutionary scenarios. The raggedness index (r) was also calculated to indicate distributions of stationary/expanding/contracting populations (Harpending et al., 1993). The program DnaSP version 3.5.1 (Rozas and Rozas, 1999) ran these tests and generated the graphs depicting the mismatch distribution or distribution of pairwise differences. The Kimura-two-parameters model was adopted to construct the NJ tree (Kimura, 1980) using 1000 bootstrap replicates. NETWORK version 4.5.0.0 was used to connect and draw the genetic relationships of the inferred haplotypes.

RESULTS

Polymorphisms of the *trn*L-*trn*F spacer

Patterns of sequence variation

Amplification was conducted for all species using the *trn*L-*trn*F primer for the ~500-bp region described by Taberlet et al. (1991). A total of 457 bp of the non-coding spacer

between the *trn*L and *trn*F genes of the cpDNA were sequenced and aligned. DNA sequences varied from 398 bp for the *Prunus* 'Adam Hmam' cultivar to 563 bp for the *Prunus* 'Hamra Badri' cultivar (Table 2). The insertion and deletion (i.e., indel) events explained the observed variation. In addition, the GC and AT contents of the amplified sequences varied from 31.2 to 36.0% and 64.0 to 68.7%, respectively, in the *trn*L-*trn*F spacer. High and low levels of GC contents were detected in the *Prunus* 'Awina Safra Morra' and *Prunus* 'Badri 2' cultivars, respectively. The nucleotide frequencies for A, T, C, and G were 0.332, 0.371, 0.156, and 0.141, respectively. The rates of transition/transversion mutations were calculated for purines ($K1$ = 5.006), pyrimidines ($K2$ = 1.338), and all bases (R = 0.95).

Twenty-two haplotypes were detected, yielding a haplotype diversity of 0.996 and a nucleotide diversity of 0.07193 (Table 3). A high level of polymorphism characterized the sequences of the *trn*L-*trn*F spacer, as confirmed by nucleotide diversity (π) and the segregating site (S) variations (Figure 1a, b). Twenty-two haplotypes were detected among 23 examined plums; 21 were distinctive haplotypes and one was common to 2 trees belonging to *P. salicina* and *P. insititia*. Our results provide evidence for the presence of important genetic diversity in local resources, as demonstrated by the detected variation pattern of the plastid DNA.

Table 2. Variation in length, and GC and AT contents of the *trn*L-*trn*F intergenic spacer and combined sequences of the *trn*L intron & *trn*L-*trn*F spacer.

Cultivar	*trnL-trnF* spacer			Combined sequences *trnL-trnF* & spacer *trnL* intron		
	%AT	%GC	Length	%AT	%GC	Length (bp)
Badri 1	65.9	34.0	514	65.9	34.1	1078
Awina Safra	68.1	31.9	439	66.2	33.9	1001
Jinha	66.5	33.4	415	66.4	33.6	1051
Ain Kounoulia	64.8	35.2	475	65.8	34.2	1037
Cidre	64.2	35.7	473	63.0	37.0	1084
Neb Zarouk	65.2	34.8	477	65.0	35.0	1036
Hamda	65.4	34.6	405	65.0	35.1	965
Awina Hamra Badria	67.2	32.8	427	66.7	33.2	987
Golden Japan	65.6	34.4	506	65.6	34.4	1072
Ain Torkia	65.3	34.6	430	65.3	34.7	1000
Santa Rosa	65.4	34.7	404	65.6	34.4	971
Zaghouania	66.1	33.9	431	65.2	28.8	997
Ain Tastouria	67.4	32.6	445	66.5	33.5	1000
Miski kahla	67.3	32.7	453	66.7	33.4	1013
Sauvage Raf Raf	67.0	33.6	446	65.9	34.1	1042
Hamra Badri	65.6	34.5	563	65.2	34.8	1139
Chaarawiya	67.9	32.1	439	66.8	33.2	1002
Stanley	64.9	35.1	488	63.9	36.1	1041
Adam Hmam	65.4	34.7	398	-	-	-
Miski Hamra	67.2	32.8	516	-	-	-
Badri 2	68.7	31.2	435	-	-	-
Zanou	67.4	32.7	570	-	-	-
Awina Safra Morra	64.0	36.0	494	-	-	-
Average	66.2	33.8	457.3	66.6	34.4	1028.7

Table 3. Polymorphism and neutrality tests parameters calculated for the *trn*L-*trn*F spacer and pooled sequences of the *trn*L intron and *trn*L-*trn*F spacer of plum cultivars.

Evolution parameters	*trnL-trnF* spacer	Combined sequences *trnL* intron & *trnL-trnF* spacer
N	23	18
S	67	334
π	0.07193	0.17214
H	22	18
H_D	0.996	1.000
K	22.372	130.137
Tajima's *D*	-1.07967 (P > 0.10)	0.14399 (P > 0.10)
Fu and Li's D*	-0.68494 (P > 0.10)	0.28605 (P > 0.10)
Fu and Li's F*	-0.94861 (P > 0.10)	0.28385 (P > 0.10)
Fu's Fs statistic	-6.214 (P = 0.002)	-0.734 (P = 0.324)

N: number of accessions (sequences); *S*: number of polymorphic sites; *H*: number of haplotypes, π: nucleotide diversity, H_d: haplotype diversity, *K*: average of pairwise nucleotide differences, P_{is}: parsimony informative sites; Tajima's *D*, Fu and Li's *D** and Fu and Li's F*: neutrality tests.

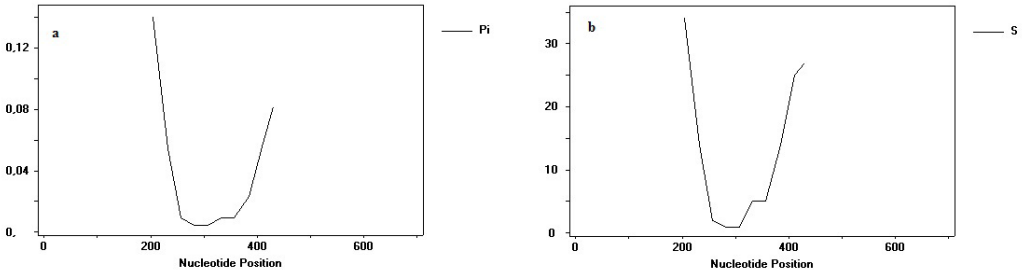

Figure 1. Hundred base-pair sliding window of the chloroplast DNA (cpDNA) non-coding region for *Prunus* spp. Variability in nucleotide diversity (π) (**a**) and segregating sites (*S*) (**b**) for the *trn*L-*trn*F spacer.

Genetic relationships

The cpDNA data set for the broad investigation of plums consisted of 649 aligned nucleotide positions (including sites with gaps/missing data), with 308 conserved sites and 330 variable positions, 144 of which were parsimony informative. The NI and Mp methods were applied to construct phylogenetic relationships among plums. The Mp analysis showed moderate consistency (CI = 0,493) and retention (RI = 0,483) indices. Bootstrap values supporting each branch are given at the branch nodes (Figure 2). This result indicates minimal homoplasy within the data set. The genetic distances among the accessions ranged from 0.003 to 0.14 (average 0.077). The minimum distance was observed between *Prunus* 'Badri 2' and *Prunus* 'Miski Hamra' and showed important similarities between the sequences; the maximum value was obtained between the *Prunus* 'Sauvage RafRaf' and *Prunus* 'Ain kounoulia' varieties, supporting the great dissimilarity among their cpDNA intergenic spacers. All remaining plums displayed different and intermediate levels of similarity. Figure 3 shows the NJ dendrogram derived from the *trn*L-*trn*F spacer sequences. In this tree, 2 major clusters were distinguished. The first one (designated I) was composed of *Prunus* 'Hamra Badri', *Prunus* 'Cidre', *Prunus* 'Ain Kounoulia', and *Prunus* 'Golden Japan'. The other plum trees were grouped in the second cluster (designated II), which displayed secondary branches labeled II-1 and II-2. The first subgroup (II-1) was composed of the local cultivars *Prunus* 'Jinha', *Prunus* 'Ain Torkia',

Prunus 'Awina Safra Morra', *Prunus* 'Hamda', and *Prunus* 'Adam Hmam'; and the introduced cultivar *Prunus* 'Santa Rosa'. The phylogenetic tree showed a cultivars distribution that occurs independently of the geographic origin of trees and a genetic closeness between local and introduced plums. The NJ tree exhibited strong clustering relationships, with >50% bootstrapping values for the majority of nodes.

To build a plausible scenario for the molecular history of the chloroplast genome, demographic expansion was tested using different approaches. First, Tajima's *D*, and Fu and Li's neutrality tests (Tajima, 1993) were calculated to check for selective neutrality. The selective neutrality tests showed that Tajima's *D*, and Fu and Li's tests were negative in the sample, whereas their respective probabilities did not reach statistical significance (Table 3). Consequently, the estimated neutrality tests accepted the neutrality assumption. Therefore, Tajima's *D*, and Fu and Li's tests suggest an excess of singletons throughout the sample and indicate a negative *D* value, which would suggest population expansion. The mismatch distribution showed a multimodal pattern (Figure 4a). The low value of the Harpending raggedness index ($r = 0.0075$, P = 0.000) confirmed the satisfactory fit to a unimodal shape. Secondly, Fu's *Fs* test (Fu, 1997) was also estimated. The Fu's *Fs* test value was negative and significant at the 2% level (Table 3). The site-frequency spectra suggest, therefore, a deficit of singleton and an excess of intermediate variants in the whole sample (Figure 4b).

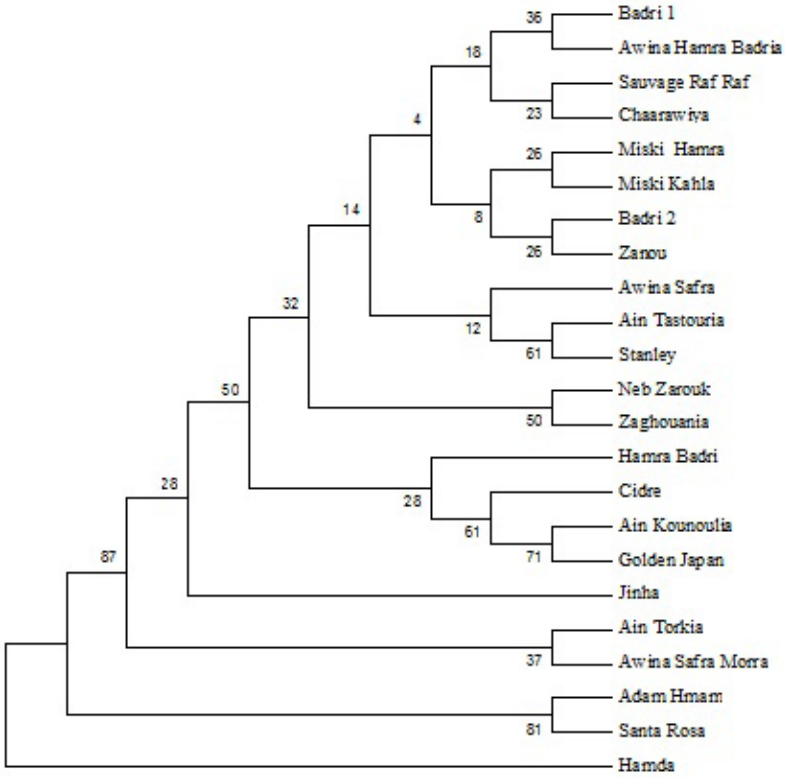

Figure 2. A maximum parsimony dendrogram derived from the *trn*L-*trn*F spacer. This single most parsimonious tree was 128 steps (CI = 0.493; RI = 0.483).

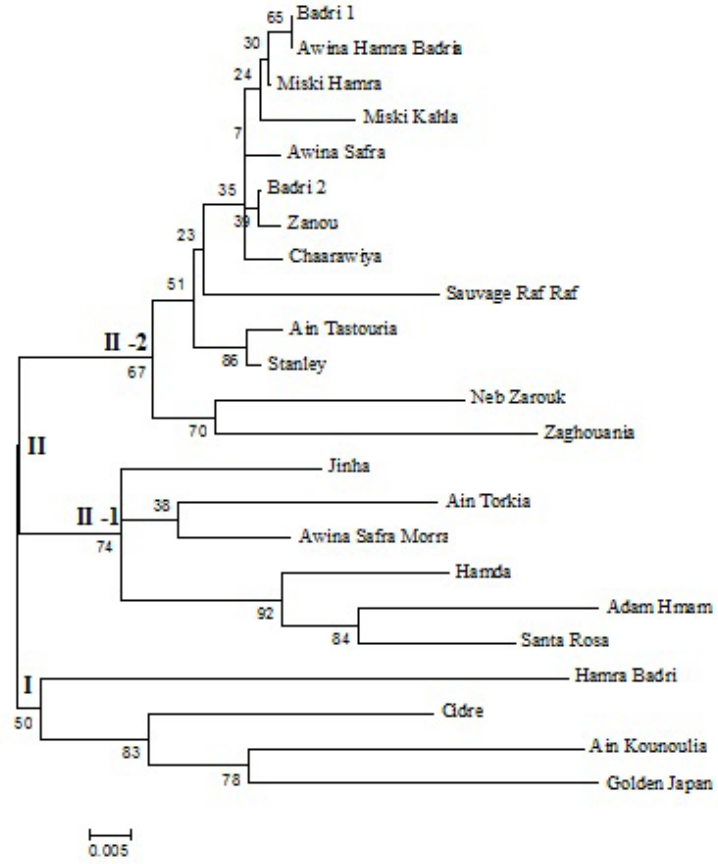

Figure 3. Neighbor-joining (NJ) tree based on haplotypes of the *trn*L-*trn*F spacer of the chloroplast DNA of plum cultivars. Numbers at the nodes indicate bootstrap values.

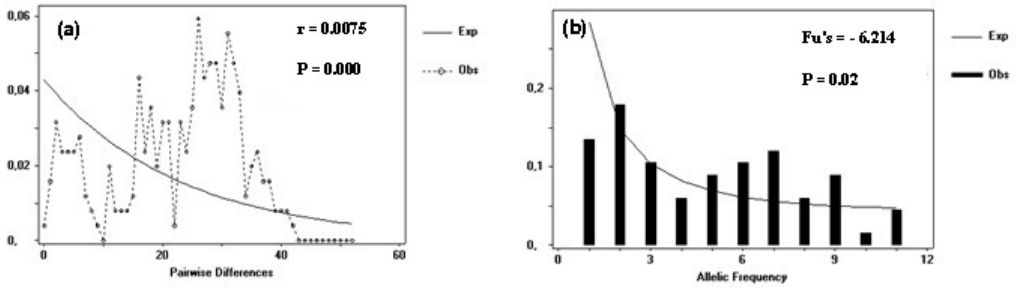

Figure 4. (**a**) Mismatch distribution of cpDNA sequences of plums based on pairwise nucleotide differences in the *trn*L-*trn*F spacer. Solid lines in the curves indicate the expected distribution under expansion, and dotted lines indicate the observed distribution under population expansion. The raggedness statistics and corresponding P values are given. (**b**) Site-frequency spectra of the cpDNA sequences in plums. Solid lines in the site-frequency spectra indicate the expected distributions under neutrality and at equilibrium. Fu's *F*s statistics and corresponding P values are given.

Haplotype distributions

The relationships among the 22 cpDNA haplotypes detected for the *trn*L-*trn*F spacer are presented in a minimum spanning network (Figure 5) that present a good topology of linked haplotype groups based on detected mutational events that occurred throughout their history. Haplotype 1 was represented by the cultivars *Prunus* 'Badri 1' and *Prunus* 'Awina Hamra Badria'. The other haplotypes detected from the cpDNA suggest that each cultivar is unique. The shaft connecting the haplotypes obtained from plum accessions shows the beginning of a star-like structure. The topology of the network appears to be complex and indicates the presence of some ancient haplotypes and the recent expansion of the others. In fact, long opposed branches are shown by several haplotypes such as H6, H21, H3, and H16, and short connections between the ancestral haplotype H1 and some others haplotypes (e.g., H18, H14, H19, and H20) (Figure 5).

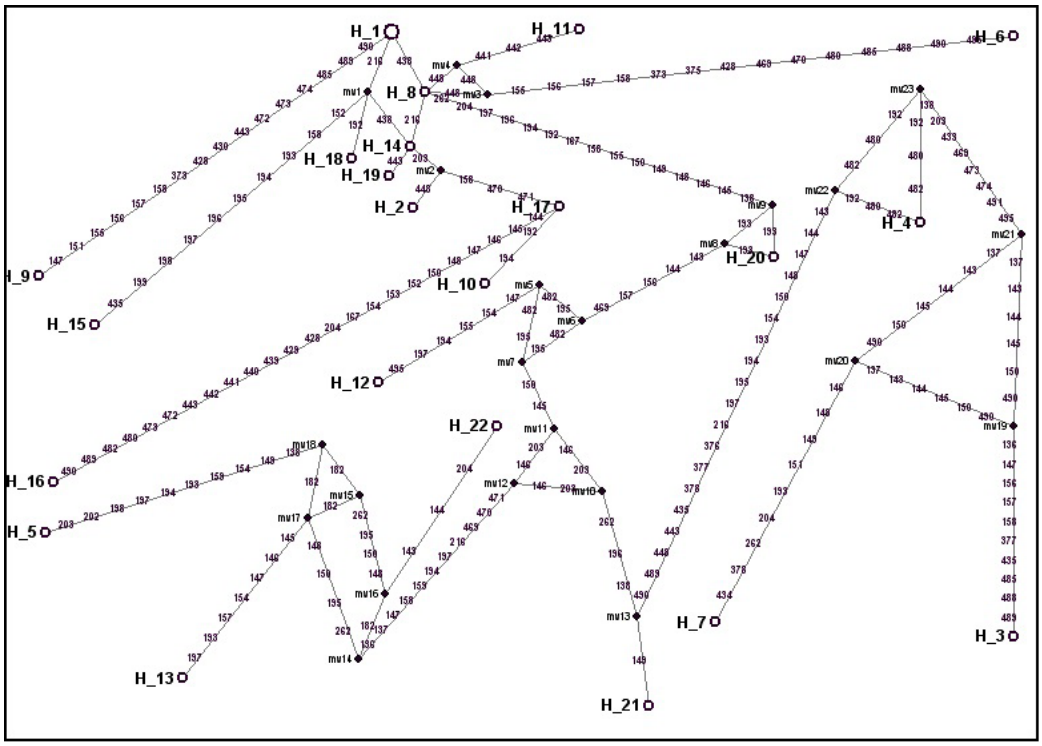

Figure 5. Minimum spanning network of the *trn*L-*trn*F spacer haplotypes of Tunisian plum cultivars.

Combined sequences of *trn*L-*trn*F region

Sequences analyses

All sequences obtained were aligned to analyze polymorphisms in plum cultivars. The lengths of the *trn*L(UAA)-*trn*F(GAA) ranged from 965 bp for *Prunus* 'Hamda' to 1139 bp for *Prunus* 'Hamra Badri'. The G+C contents ranged from 28.8 to 37, with a mean value of 34.4,

and the A+T content was 66.6, on average, ranging from 63 to 66.8 in the pooled sequences (Table 2). The nucleotide frequencies were 0.347, 0.336, 0.151, and 0.166 (G) for A, T/U, C, and G, respectively. The transition/transversion rate ratios were *K1* = 1.027 and *K2* = 0.919 for purines and pyrimidines, respectively. For all bases, the transition/transversion rate was R = 0.305. Eighteen haplotypes were detected. The values calculated of the nucleotide (π) and haplotype (H_D) diversities were 0.1721 and 1.000, respectively (Table 3). The high haplotypic and nucleotide diversities characterized the non-coding sequences of the cpDNA. Nucleotide variability characterized the sequences of the *trn*L-*trn*F region, as demonstrated by nucleotide diversity (π) and segregating site (*S*) variations (Figure 6 a,b).

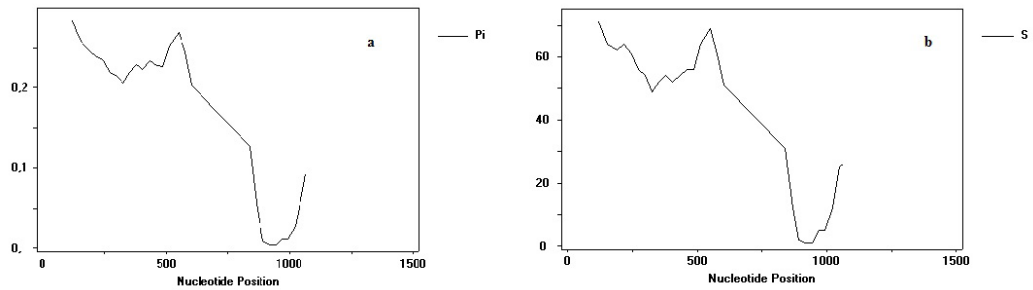

Figure 6. Hundred base-pair sliding window of the cpDNA non-coding region for *Prunus* spp. Variability of π (**a**) and *S* (**b**) for the combined sequences.

Genetic relationships

The combined cpDNA data set of plum cultivars consisted of 1290 aligned nucleotide positions, with 709 variable positions, 459 of which were parsimony informative. The Mp analysis suggested high consistency (CI = 0.732) and retention (RI = 0.871) indices, and, when compared to the *trn*L-*trn*F spacer, indicated low homoplasy within the dataset (Figure 7). Genetic relationships among the Tunisian plums were established using sequences of the region *trn*L (UAA)-*trn*F (GAA). The pairwise distance was calculated and ranged from 0.030 to 0.491, with a mean of 0.254. The minimum distance was observed between *Prunus* 'Miski Kahla' and *Prunus* 'Neb Zarouk' and revealed important similarities between the cultivars; the maximum distance was found to exist between *Prunus* 'Chaarawiya' and *Prunus* 'Hamra Badri', thus, showing their dissimilarity. The NJ dendrogram illustrated 2 main groups of plums (Figure 8). The first cluster was composed of *Prunus* 'Awina Safra', *Prunus* 'Stanley', *Prunus* 'Chaarawiya', *Prunus* 'Badri 1', and *Prunus* 'Zaghouania'. All other plums were grouped into the second cluster. The introduced cultivars *Prunus* 'Santa Rosa' and *Prunus* 'Golden Japan' did not differ from the Tunisian plums, and the classification of trees was made independent of their geographic origin or name.

Tests of neutrality

Tests of neutrality were used to evaluate the sequences of the *trn*L-*trn*F spacer and the pooled sequences of *trn*L intron and *trn*L-*trn*F spacer. To detect departure from neutral-

ity, complementary statistics Tajima's *D*, and Fu and Li's *D* and *F* were tested. These quantities showed a statistically non-significant deviation from selective neutrality. Table 3 illustrates the statistically non-significant and positive values for the adopted tests (i.e., 0.14399, 0.28605, and 0.28385, for Tajima's *D*, and Fu and Li's *D* and *F*, respectively). These results reflect an excess of intermediate variants with regard to the expectations of neutrality. Consequently, Tajima's, and Fu and Li's tests were positive in their combined sequences, and the estimated neutrality tests accepted the neutrality assumption in the total sample. Therefore, Tajima's, and Fu and Li's tests suggested a deficit of singletons, which is in accordance with the *trn*L-*trn*F spacer. The Fu's *Fs* statistic showed a low negative and non-significant value of -0.734 (Table 3). Figure 9 shows a multimodal pattern of pairwise nucleotide differences distribution. The Harpending raggedness index was uniformly low, confirming the satisfactory fit of the data to a unimodal distribution ($r = 0.0119$, P = 0.000). The Fu's *Fs* test was negative and not significant at the 2% level (Table 3). Therefore, the site-frequency spectra (Figure 9b) imply a deficit of singletons and an excess of intermediate variants throughout the whole sample.

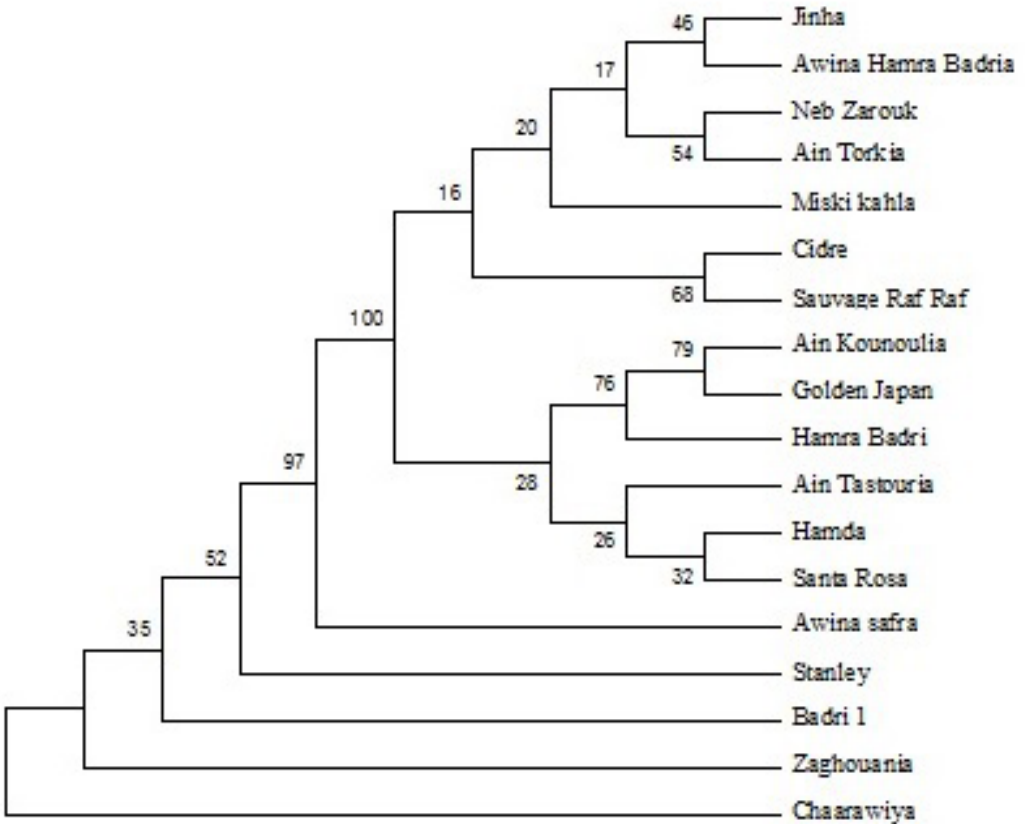

Figure 7. Strict consensus tree (CI = 0.732; RI = 0.871) obtained from the pooled *trnL* intron and *trnL-trnF* spacer data from Tunisian plum cultivars. Numbers below the nodes indicate bootstrap values.

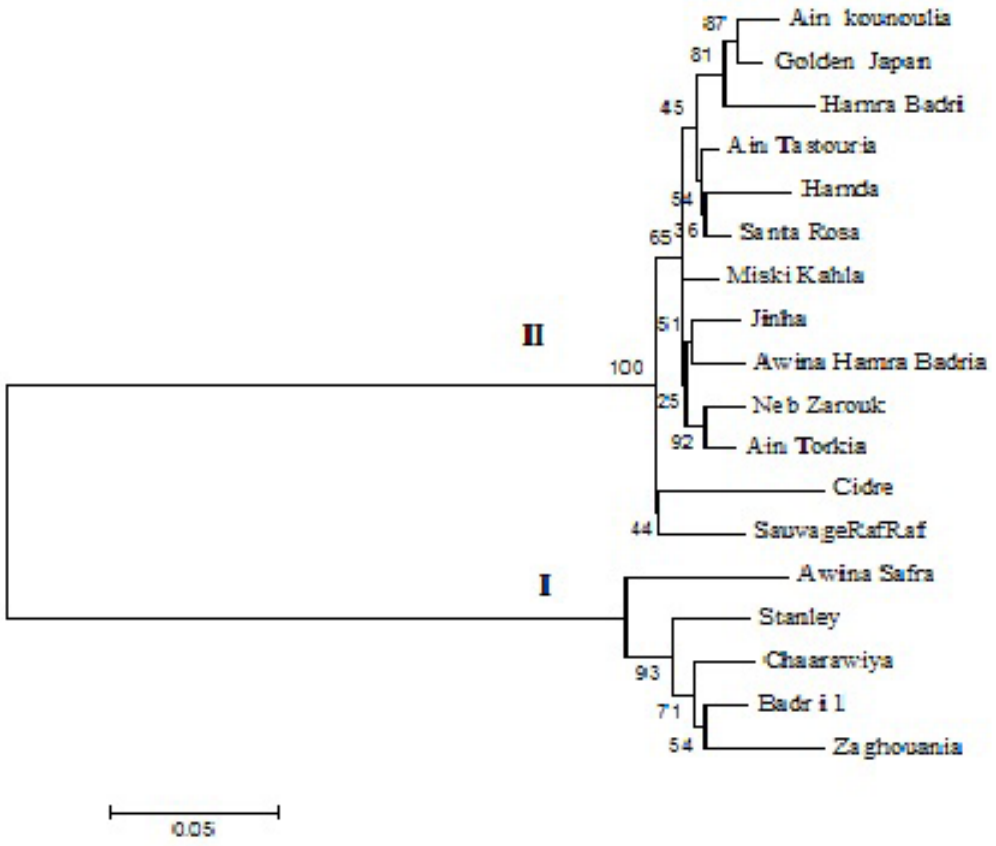

Figure 8. Neighbor-joining tree of haplotypes of the combined sequences of the *trnL-trnF* region (*trn*L intron and *trn*L-*trn*F spacer) of the plastid DNA of Tunisian plum cultivars. Numbers at nodes indicate bootstrap values.

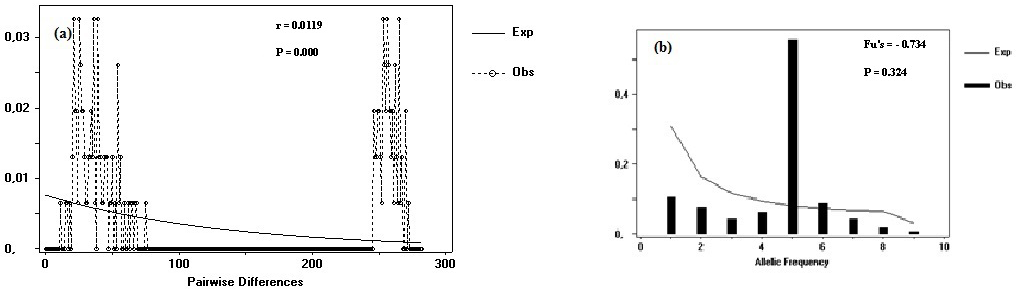

Figure 9. (**a**) Mismatch distribution of cpDNA sequences of plums based on pairwise nucleotide differences in the combined intron *trn*L and *trn*L-*trn*F spacer. Solid lines in the curves indicate the expected distribution under expansion, and dotted lines indicate the observed distribution under population expansion. The raggedness statistics and corresponding P values are given. (**b**) Site-frequency spectra of the cpDNA sequences in plums. Covered in this study for the region *trn*L-*trn*F of plum cultivars. Solid lines in the site-frequency spectra indicate the expected distributions under neutrality and at equilibrium. Fu's *F*s statistics and corresponding P values are given.

Network analysis

The 18 chlorotypes identified were connected in a minimum-length-spanning tree (Figure 10). Two main clusters of haplotypes were noted, with H14 and H1 as potential ancestral haplotypes. The group related to haplotype H1 included sequences from *Prunus* 'Awina Safra', *Prunus* 'Stanley', *Prunus* 'Chaarawiya', *Prunus* 'Badri 1', and *Prunus* 'Zaghouania'. The group connected to haplotype H14 comprised the other accessions. The network organization suggests the presence of 2 gene pools that are representative of the genetic diversity and may be candidates for use in conservation programs (Figure 10).

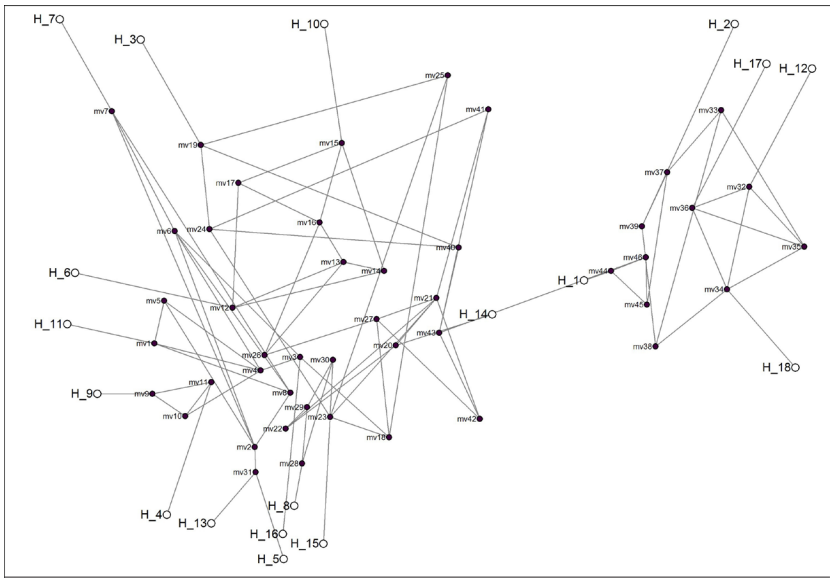

Figure 10. Minimum spanning network of the *trn*L-*trn*F region (*trn*L intron and *trn*L-*trn*F spacer) of Tunisian plums.

DISCUSSION

In the present study, we inspected the plastome genome to reveal useful molecular markers and elucidate the genetic diversity of plum cultivars. The nucleotide sequence of the chloroplast *trn*L-*trn*F intergenic spacer and combined sequences of the spacer and intron at the *trn*L-*trn*F region were investigated. The AT content (64-68.7%) was estimated to be similar to those calculated in angiosperm (64.5-68.7%) by Bakker et al. (2000) and were comparable to those identified for 19 other cpDNA non-coding regions in Poaceae (Morton and Clegg, 1995; Morton et al., 1997) and the *atp*B-*rbc*L intergenic spacer region in Rubiaceae (Manen and Natali, 1995). The cpDNA variation in 23 accessions revealed 22 *trn*L-*trn*F spacer haplotypes of the cpDNA, suggesting a high level of Tunisian plum diversity. In addition, the nucleotide diversity of 0.07193 was higher than the value observed in *Ficus carica* (Baraket et al., 2009) for the same region. It is also important to note that the *trn*L intron is highly diversified in comparison to the *trn*L-*trn*F spacer in the same plum varieties. In fact, nucleotide diversity of the *trn*L intron is 0.28324, as reported by Ben Mustapha et al. (2013). Moreover, our result shows that the *trn*L-*trn*F plum spacer is more variable than the spacer of *Aegilops speltoides*, as deter-

mined by Miyashita et al. (1994) based on restriction site analyses ($\pi = 0.00083$) for the *atp*B-*rbc*L and *psb*B-*psb*N-*psb*H-*pet*B-*pet*D-*rpo*A regions ($\pi = 0.00018$), and *F. leptopodum* for the *trn*K intron ($\pi = 0.2810^{-3}$) (Ohsako and Ohnishi, 2001). Contrarily, Shaw et al. (2005) reported that the *trn*S-*trn*G-*trn*G, *trn*C-*ycf*6-*psb*M, *trn*D-*trn*T, *trn*T-*trn*L, and *rpo*B-*trn*C regions in the chloroplast genome were more variable than the *trn*L-*trn*F spacer. It appears that the *trn*L-*trn*F spacer evolves about 3.24 times faster than the *trn*L intron. In fact, the overall transition/transversion rates (*R*) were 0.293 (Ben Mustapha et al., 2013) and 0.95 for the *trn*L intron and the *trn*L-*trn*F spacer, respectively. It has also been observed that the transition/transversion rates for purines (*K1*) and pyrimidines (*K2*) were 5.006 and 1.338, respectively, when calculated for the *trn*L-*trn*F spacer, which were higher than those calculated for the *trn*L intron (*K1* = 0.803 and *K2* = 0.971) (Ben Mustapha et al., 2013). The non-coding region exhibited higher mutational rates than the coding regions and can be exploited to increase the usefulness and resolution of the cpDNA, and the level of taxonomic resolution of the cpDNA, thus estimating divergence among closely related species even at the intra-specific level (Taberlet et al., 1991). Plum genetic diversity and its distribution is not only allied to the incidence of refugial areas and ecogeographic and bioclimatic conditions, but also human practices that cause dispersal of trees. The observed inconsistency of the identified haplotypes may be generated by human diffusion or hybridization. The sampling was rather careful to award confidence in the larger patterns revealed in the data. Since plums have undoubtedly been spread by humans, we are less inclined to focus on the less frequent chlorotypes and their distribution. Moreover, the low rate of homoplasy provided by the parsimonious approaches to assessing cpDNA variation implied a rapid and radial mode of speciation for plums. The current phylogenetic study demonstrated that the distribution of plum varieties is independent of their geographical origin and locality or their relatedness to introduced cultivars. A similar result was observed in Tunisian plums using other molecular markers (i.e., RAPD). In fact, Santa Rosa and Golden Japan did not significantly differ from the local plum cultivars. This assumption may signify a potential common origin between all plum varieties, as suggested by Ben Tamarzizt et al. (2009). In the same contest, AFLP markers appear to be more informative than cytoplasmic markers when exploring genetic diversity and establishing relatedness between Turkish plum cultivars (Ilgin et al., 2009). In addition, historical populations were elucidated via renowned selective neutrality tests and mismatch distribution. The neutrality tests used did not support the hypothesis that plums endured a population bottleneck or expansion. No statistically significant values were calculated for Tajima's *D,* and Fu and Li parameters throughout the data sample set. Moreover, the profiles of the mismatch distributions were irregularly bimodal or multimodal, suggesting that distribution was relatively stable in the past. In our study, non-significant values of Tajima's *D*, and Fu and Li tests and the shape of the multimodal peaks of the mismatch distribution pointed out that the plum cultivars have undergone demographic equilibrium. Additionally, extremely high haplotype and nucleotide diversities ($H_D = 0.996$, $\pi = 0.071$) obtained for the *trn*L-*trn*F spacer and profile of the mismatch distribution illustrate an old population of constant size and show that plums share closely related chlorotypes. Moreover, it is important to provide another scenario for the plum plastome sequence evolution and discuss the negative and significant value of Fu's *Fs* test (i.e., -6.214, P = 0.002), which suggested that expansion or negative selection occurred in the spacer. In this case, the negative and nonsignificant Tajima's *D*, and Fu and Li quantities are likely artifacts of population demographics and do not necessarily support neutrality. The topology of the haplotype network based on the *trn*L-*trn*F spacer or combined data (i.e., *trn*L intron and *trn*L-*trn*F spacer)

suggests an adaptive divergence of plum species to a variety of living habitats and potential refuges during the Quaternary Ice Age.

Our finding is in agreement with others based on molecular markers and, in part, consistent with clustering by the use of morphological traits that are polygenic and influenced by severe selection pressure and environmental effects, which lead to differences between trees of the same variety. As a result, we advise that the chloroplast genome is a source of molecular markers in the Tunisian plum germplasm, which can be used for identification and descriptions of genetic diversity and history. It is evident that the sample dataset should be enlarged to implicate more spontaneous species, provide deeper insight into the genetic polymorphisms present in this compartment, and elucidate the precise molecular evolutionary scenario. Investigation is currently in progress to describe the genetic resources and design rational conservation strategies for plum species. Herein, the *trn*L-*trn*F region of the chloroplast genome revealed powerful markers that can be used to investigate cytoplasm genetic diversity and construct a molecular database for cultivated and spontaneous plum species for the development and improvement of conservation programs.

ACKNOWLEDGMENTS

Research supported by grants from the Tunisian Ministère de l'Enseignement Supérieur et de la Recherche Scientifique.

REFERENCES

Ahmed J, Shivhare US and Raghavan GSV (2004). Rheological characteristics and kinetics of color degradation of green chilli puree. *J. Food Eng.* 218: 525-528.

Bakker FT, Culham A, Gomez-Martinez R, Carvalho J, et al. **(**2000). Patterns of nucleotide substitution in angiosperm cpDNA trnL (UAA)-*trn*F (GAA) regions. *Mol. Biol. Evol.* 17: 1146-1155.

Baraket G, Saddoud O, Chatti K, Mars M, et al. (2009). Chloroplast DNA analysis in Tunisian fig cultivars (*Ficus carica* L.): Sequence variations of the *trn*L-*trn*F intergenic spacer. *Biochem Syst Ecol.* 36: 828-835.

Ben Mustapha S, Ben Tamarzizt H, Baraket G, Abdallah D, et al. (2013). A signature of balancing selection in the plastid *trn*L UAA intron and investigation of the phylogenetic relationships among Tunisian plum species (*Prunus* spp). *Sci. Hortic.* 151: 51-56.

Ben Tamarzizt H, Baraket G, Ben Mustapha S, Marrakchi M, et al. (2009). Genetic relatedness among Tunisian plum cultivars by random amplified polymorphic DNA analysis and evaluation of phenotypic characters. *Sci. Hortic.* 121: 440-446.

Bernatzky R and Tanskley SD (1986). Genetics of actin-related sequences in tomato. *Theor. Appl. Genet.* 72: 314-321.

Besnard G (2008). Chloroplast DNA variations in Mediterranean olive. *J. Hortic. Sci. Biotechnol.* 83: 51-54.

Bouhadida M, Martin JP, Eremin G, Pinochet J, et al. (2007). Chloroplast DNA diversity in *Prunus* and its implication on genetic relationships. *J. Am. Soc. Hortic. Sci.* 132: 670-679.

Clegg MT and Zurawski G (1992). Chloroplast DNA and the study of plant phylogeny: present status and future prospects. In: Molecular systematics of plants (Soltis PS, Soltis DE and Doylle JJ, eds.). Chapman and Hall, New York, 1-13.

Clegg MT, Learn GH and Golenberg EM (1991). Molecular evolution of chloroplast DNA. Chapter 7 In: Evolution at the Molecular level. (Selander RK, Clark AG and Whittam TS, eds.) Sinauer, Sunderland, MA, 135-149.

Demesure B, Sodzi N and Petit RJ (1995). A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. *Mol. Ecol.* 4: 129-131.

Dowling TE, Moritz C and Palmer JD (1990). Nucleic acids II: restriction site analysis. In: Molecular systematics (Hillis DM and Moritz C, eds.). Sinauer Assoc., Inc., Suderland, Massachusetts, 250-315.

Felsenstein J (1993). Phylogeny inference package (PHYLIP). Version 3.5. University of Washington, Seattle.

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

Harpending HC, Sherry ST, Rogers AR and Stoneking M (1993). The genetic structure of ancient human populations. *Curr. Anthropol.* 34: 483-496.

Ilgin M, Kafkas S and Ercisli S (2009). Molecular characterization of plum cultivars by AFLP markers. *Biotechnol. Biotechnol. Equip.* 1189-1193.

Jukes TH and Cantor CR (1969). Evolution of protein molecules. In: Mammalian protein metabolism (Munro HN, ed.). Academic Press, New York, 21-132.

Kimura M (1980). A simple method for estimating evolutionary rate of base substitution through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16: 111-120.

Manen JF and Natali A (1995). Comparison of the evolution of ribulose-1, 5-biphosphate carboxylase (rbcL) and *atp*B-*rbc*L noncoding spacer sequences in a recent plant group, the tribe Rubieae (Rubiaceae). *J. Mol. Evol.* 41: 920-927.

Miyashita NT, Mori N and Tsumewaki K (1994). Molecular variation in chloroplast DNA regions on ancestral species of wheat. *Genetics* 137: 883-889.

Morton BR and Clegg MT (1995). Neighboring base composition is strongly correlated with base substitution bias in a region of the chloroplast genome. *J. Mol. Evol.* 41: 597-603.

Morton BR, Oberholzer VM and Clegg MT (1997). The influence of specific neighboring bases on substitution bias in noncoding regions of the plant chloroplast genome. *J. Mol. Evol.* 45: 227-231.

Okie WR and Weinberger JH (1996). Plums. In: Fruit breeding, Vol. I: Tree and tropical fruits (Janick J and Moore JN, eds.). Wiley, New York, 559-607.

Ohsako T and Ohnishi O (2000). Intra- and interspecific phylogeny of wild Fagopyrum (Polygonaceae) species based on nucleotide sequences of noncoding regions in chloroplast DNA. *Am. J. Bot.* 87: 573-582.

Ohsako T and Ohnishi O (2001). Nucleotide sequence variation of the chloroplast trnK/matK region in two wild Fagopyrum (Polygonaceae) species, *F. leptopodium* and *F. statice*. *Genes Genet. Syst.* 76: 39-46.

Palmer JD (1987). Chloroplast DNA evolution and biosystematic uses of chloroplast DNA variation. *Am Nat.* 130: S6-S29.

Rozas J and Rozas R (1999). DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* 15: 174-175.

Rozas J, Sánchez-Delbarrio JC, Messeguer X and Rozas R (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19: 2496-2497.

Saitou N and Nei M (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4: 405-425.

Shaw J, Lickey EB, Beck JT, Farmer SB, et al. (2005). The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *Am. J. Bot.* 92: 142-166.

Sun G (2002). Interspecific polymorphism at non-coding regions of chloroplast, mitochondrial DNA and rRNA IGS region in *Elymus* species. *Heriditas* 137: 119-124.

Taberlet PL, Gielly G, Pautou G and Bouvet J (1991). Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Mol. Biol.* 17: 1105-1109.

Tajima F (1993). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123: 585-595.

Tamura K, Dudley J, Nei M and Kumar S (2007). MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24: 1596-1599.

Xia X (2000). Data Analysis in Molecular Biology and Evolution. Kluwer Academic Publishers, Boston.

Yamane K, Yasui Y and Ohnishi O (2003). Intraspecific cpDNA variations of diploid and tetraploid perennial buckwheat, *Fagopyrum cymosum* (Polygonaceae). *Am. J. Bot.* 90: 339-346.

Zurawski G and Clegg M (1987). Evolution of higher plant chloroplast DNA encoded genes: implications of structures functions and phylogenetic studies. *Ann. Rev. Plant Physiol*. 38: 391-418.