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

tRNALeu intron (UAA) of *Ficus carica* L.: genetic diversity and evolutionary patterns

G. Baraket, A.B. Abdelkrim and A. Salhi-Hannachi

Laboratory of Molecular Genetics, Immunology & Biotechnology,
Faculty of Sciences, University Tunis El Manar, El Manar, Tunis, Tunisia

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

Genet. Mol. Res. 14 (2): 3817-3832 (2015)
Received May 22, 2014
Accepted October 29, 2014
Published April 22, 2015
DOI http://dx.doi.org/10.4238/2015.April.22.11

ABSTRACT. Cytoplasmic chloroplast DNA was explored to establish genetic relationships among *Ficus carica* cultivars and elucidate the molecular evolution of the species. The results suggest the occurrence of haplotype and nucleotide diversity. Conserved group I intron sequence motifs were detected and showed a common secondary structure, despite the presence of some mutations on their sequences. The neighbor-joining dendrogram showed a continuous diversity that characterizes local resources. The maximum parsimony tree, with an RI index of 0.507, indicated minimal homoplasy within the data set. Furthermore, our results demonstrate that the *trn*L intron is the seat of numerous substitutions. Herein, new insight on the mechanism involved in the evolution of the *trn*L intron in the fig is presented. From the study, it appears that there is an explicit rejection of the null hypothesis in *F. carica*. A scenario of positive selection and recent expansion of *F. carica* genotypes across Tunisia seems to be retained.

Key words: Chloroplast DNA; *Ficus carica*; Genetic relationships; Molecular evolution; *trn*L UAA intron; Tunisia

INTRODUCTION

The chloroplast genome has become a major focus for studies of plant molecular evolution. This genome has a small size and generally evolves at a relatively slow rate (Clegg et al., 1991), thus making it an ideal system to assess the phylogenetic relationships among genera (Downie et al., 1996; Ben Mustapha et al., 2013), families, or higher order taxonomic ranks (Nickrent and Soltis, 1995). The coding sequences of many genes are interrupted by stretches of noncoding DNA, which are termed introns. Introns are widespread genetic elements found in all major groups of organisms. Since they were discovered, their origin has been debated and different theories have been proposed (Bell-Pedersen et al., 1990). In the conserved *trn*T-*trn*F region of the chloroplast genomes ranging from mosses to seed plants, a group I intron splitting the tRNALeu gene is a stable component of the *trn*T-*trn*L-*trn*F cistron and has been widely used for reconstructing phylogenetic inference at the intra- and interspecific levels (Yu et al., 2007). The intron of *trn*L (UAA) is regularly used and displays a relatively low level of variation in all examinations thus far (Bayer and Starr, 1998). Although coding regions could have phylogenetic potential, it is currently and widely believed that the non-coding regions provide the most practical source of data within the chloroplast genome for phylogenetic inference at lower taxonomic levels (cf. Morton and Clegg, 1993). In plants, this intron has been used in systematic studies (where high resolution is needed) and population genetics (Gielly and Taberlet, 1994; Kita et al., 1995). For example, it has been used to resolve phylogenetic relations when the sequence for the *rbc*L gene showed too little variation (Gielly and Taberlet, 1994). The ability of the group-I introns to catalyze their own splicing is dependent on their highly conserved secondary and tertiary structures. Different group I introns have relatively little sequence similarity, but all share a series of short conserved elements, P, Q, R, and S, known as the catalytic core (Cech et al., 1994).

In this study, we investigated the nucleotide sequence diversity and the mode of evolution of the *trn*L (UAA) chloroplast intron in Tunisian fig accessions. The fig, *Ficus carica* L. (Moraceae), is a fruit tree of antiquity associated with the beginning of horticulture in the Mediterranean Basin (Zohary and Spiegel-Roy, 1975). It was domesticated from a group of diverse spontaneous figs occurring in the south and east of the Mediterranean region sometime in the early Neolithic period (Zohary and Hopf, 1993). In Tunisia, the fig germplasm consists of numerous landraces, is mainly selected by farmers for their fruit qualities, and is maintained in orchards. They have a wide range throughout different eco-geographical areas of the country and are threatened by genetic erosion. In recent years, several studies have focused on the identification and characterization of Tunisian fig cultivars to elaborate a national core collection and preserve these genetic resources (Salhi-Hannachi et al., 2005).

This study was a part of our research activities and aimed to analyze the sequence variation of the *trn*L intron; detect a conserved motif from *trn*L sequences; and reconstruct a genetic tree, detect the footprint of selection, and investigate the evolutionary model of this intron.

MATERIAL AND METHODS

Plant material

Forty-nine cultivars (i.e., 41 female and 8 male trees) of the Tunisian fig (Table 1)

were used in this study. They were collected from 5 regions. The plant material consisted of young leaves sampled from adult trees.

Table 1. Tunisian fig cultivars studied and their geographical origin.

Cultivar	Horticultural classifications	Geographic origin	
		Region	Locality
Soltani 1	Uniferous	Sahel	Ourdanine
Kahli 1	Uniferous		Kalaa Kebira
Hemri	Uniferous		Enfidha
Zidi 1	Uniferous		Mesjed Aissa
Baghali	Uniferous		Mesjed Aissa
Bidhi	Uniferous		Kalaa Kebira
Bither Abiadh 1	Bifèrous		Mesjed Aissa
Besbessi	Bifèrous		Mesjed Aissa
Dchiche Assal	Uniferous		Ghadhabna
Kahli 2	Uniferous		Enfidha
Zidi4	Uniferous		Ghadhabna
Jrani*	Caprificus		Ghadhabna
Assafri*	Caprificus		Ghadhabna
Zidi 3	Uniferous	South West	Tozeur
Zergui	Uniferous		Dégache
Hamri	Uniferous		Dégache
Khadhri	Uniferous		Dégache
Khartoumi	Uniferous		Dégache
Tounsi	Uniferous		Dégache
Wahchi	Bifèrous		Dégache
Chetoui 1	Bifèrous		Dégache
Dhokkar 1*	Caprificus		Gafsa
Dhokkar 2*	Caprificus		Dégache
Sawoudi 1	Uniferous		Gafsa
Gaa Zir	Uniferous		Gafsa
Assal Boudchiche	Uniferous		Gafsa
Sawoudi 2	Uniferous		Dégache
Grichy	Uniferous		Tozeur
Khalt	Uniferous		Tozeur
Khzami	Uniferous		Tozeur
Bither	Bifèrous		Tozeur
khadhouri	Uniferous		Gafsa
Hammouri	Uniferous	South East	Medenine
Widlani	Uniferous		Medenine
Zaghoubi	Uniferous		Medenine
Rogabi	Uniferous		Beni Khedache
Makhbech	Uniferous		Medenine
Tayouri Assfar	Uniferous		Douiret
Dhokkar Zarziz*	Caprificus		Medenine
Bither abiadh 2	Uniferous		Tataouine
Zidi 2	Uniferous	North East	Utique
Dhokkar 4*	Caprificus		Utique
Dhokkar 5*	Caprificus		Raf Raf
Soltani 3	Uniferous		Raf Raf
Chetoui 2	Biferous		Raf Raf
Soltani 2	Uniferous		Mornag
Temri	Uniferous	Kerkenah	Kerkenah
Baghli	Uniferous		Kerkenah
Abiadh	Uniferous		Kerkenah
Dhokkar 3*	Caprificus		Kerkenah

*Male tree.

DNA isolation

Total genomic DNA was purified from frozen young leaves according to the proce-

dure of Dellaporta et al. (1983). The DNA concentration was estimated spectrophotometrically and its integrity was checked by analytical [1% (w/v)] agarose minigel electrophoresis (Sambrook et al., 1989).

PCR amplification, purification, and sequencing

The intron *trn*L (UAA) was amplified using specific primers designed by Taberlet et al. (1991). The procedure followed was that previously described by Baraket et al. (2010). Cycle Sequencing and the Big Dye Terminator Ready Reaction Kit (Applied Biosystems, Foster City, CA, USA) were used.

Sequence analysis

All sequence information has been deposited into the GenBank database (Accession Nos. EU191005-EU191024 and GQ395387-GQ395415). The derived nucleotide sequences were aligned using the ClustalW package in the DAMBE program (Xia, 2000) and analyzed with the MEGA version 5.0 software (Tamura et al., 2011). For each sequence, length and GC content were estimated. The alignment was manually checked, and pairwise sequence divergence between cultivars in the *trn*L intron was calculated according to the maximum composite likelihood (MCL) (Tamura et al., 2011). The resultant distance matrix was then computed to generate phylogenetic trees according to the neighbor-joining (NJ) method (Saitou and Nei, 1987). Positions containing gaps and missing data were eliminated from the dataset (complete deletion option). Support values of the internal branches of the NJ tree were evaluated using the bootstrap method (1000 replicates) (Tamura et al., 2011). The consistency index (CI) and retention index (RI) were calculated (Kluge and Farris 1969). The transition/transversion ratio (ti/tv) was estimated using the following formula

$$R = [A * G * k1 + T * C * k2] / [(A + G) * (T + C)]$$

with *A*, *G*, *C*, *T* as the corresponding frequencies of the 4 nucleotides (Tamura et al., 2011). The number of nucleotide substitutions per site for analysis between the sequences was given. Aligned sequences in the Mega files were analyzed with the DnaSP software version 5.10.01 (Librado and Rozas, 2009) to estimate polymorphism indices. In fact, genetic diversity was quantified by indices of haplotype diversity (H_D) (Nei and Tajima, 1983) and pairwise estimates of nucleotide divergence (P_I) (Jukes and Cantor, 1969). The average of nucleotide differences (K) and minimum number of recombination events (R_M) were also estimated. Selective neutrality was tested by both Tajima's D (Tajima, 1989), and Fu and Li's D* and F* methods (Fu and Li, 1993). Predicted secondary structures of the conserved motif in the *trn*L intron and associated free energy value algorithm were evaluated with the minimum-free energy (MFE) algorithm (Zuker, 2003). Fold predictions were made by the use of mfold version 3.1 (Zuker, 2003) (www.bioinfo.rpi.edu/applications/mfold). Demographic parameters were assessed using the distribution of pairwise sequence differences (mismatch distribution) of Rogers and Harpending (1992)

and site-frequency spectra (distribution of the allelic frequency at a site) of Tajima (1989) using the DnaSP software 5.10.01 (Librado and Rozas, 2009). The genetic relationship of the inferred haplotypes was graphically displayed by NETWORK version 4.5.0.0 (Bandelt et al., 1999).

RESULTS

The amplified fragment showed a length of about 600 bp and corresponds to the tRNA Leu intron (UAA) of the chloroplast DNA. The BLAST results showed the authenticity and identity of sequences and confirmed the *trn*L intron in the genus *Ficus*.

Genetic diversity

*Nucleotide composition and length variation of trn*L *sequences*

The sequences obtained showed variations in their lengths and nucleotide compositions for all fig cultivars. Indeed, the size of the *trn*L intron ranged from 546 bp in the pollinator Dhokkar 3 to 607 bp in the cultivar Khadhri, with an average length of 576.3 bp (Table 2). Sequence analysis of the *trn*L intron showed that the GC content varied from 32.7 (Widlani) to 36.3 (Besbessi), with an average of 34.1 (Table 2). High values of AT bases were consequently found. The AT content varied from 63.6 (Besbessi) to 67.2 (Widlani), with an average of 65.9.

*Nucleotide composition variation and mutational events of trn*L *sequences*

The nucleotide composition of the intron trnL (UAA) showed that the frequencies of the nucleotides making up the sequence were 0.395, 0.278, 0.147, and 0.18 for A, T, C, and G, respectively. The ti/tv ratios were calculated as K1 = 1.248, K2 = 0.841, and R = 0.312 for purine nitrogenous bases, pyrimidine bases, and all bases, respectively. The relatively high content of AT (i.e., 65.9) may partly explain the high proportion of transversions identified.

The different substitutions detected are given in Table 3. An analysis of Table 3 showed that the Tunisian fig TàC and AàG transitions were more frequent than the CàT and GàA transitions. Moreover, transversions were more frequent than transitions at the *trn*L intron (Figure 1).

Multiple alignments of sequences allowed for the establishment of a matrix of 667 characters in the form of 437 conserved sites and 213 variable sites, which were distributed across 130 informative sites and 79 singleton sites. After eliminating gaps, 380 conserved sites and 57 variable sites, which included 32 informative sites and 25 unique sites, were detected. A high level of polymorphism in the intron *trn*L (UAA) of the chloroplast DNA was identified. Indeed, 37 haplotypes were detected among the 49 figs studied. The H_D and nucleotide diversity (π) were estimated at 0.951 and 0.018, respectively, for sequences of the *trn*L (UAA) intron. In addition, K of 8.16 was found, thus showing a large degree of genetic diversity at the intron chloroplast (Table 4).

Table 2. Length and nucleotide composition of chloroplast DNA sequences of Tunisian fig (*Ficus carica*) cultivars.

Accession	GenBank accession number	Lenght*	%GC	%AT
Zidi 4	EU191005	588	34.2	65.8
Hammouri	EU191006	589	34.3	65.7
Widlani	EU191007	556	32.7	67.2
Bither	EU191008	588	33.8	66.2
Zaghoubi	EU191009	586	34	66.1
Dhokkar zarzis*	EU191010	588	34	66
Khzami	EU191011	586	33.6	66.4
Tounsi	EU191012	586	34.2	65.8
Zidi 3	EU191013	554	33.4	66.7
Grichy	EU191014	588	34	66
Tayouri Asfar	EU191015	586	34	66
Hamri	EU191016	583	34	66
Bither abiadh 2	EU191017	587	34.2	65.7
Dhokkar2*	EU191018	587	33.9	66.1
Kahli 2	EU191019	587	33.9	66.1
Dchiche Assal	EU191020	586	34.2	65.8
Makhbech	EU191021	587	33.9	66.1
Rogaby	EU191022	584	34.3	65.7
Sawoudi2	EU191023	588	33.6	66.3
Khalt	EU191024	585	33.7	66.4
Soltani1	GQ395387	550	33.7	66.4
Kahli1	GQ395388	593	35.5	64.7
Hemri	GQ395389	549	34	66
Zidi1	GQ395390	553	33.3	66.7
Baghali	GQ395391	553	33.4	66.5
Sawoudi1	GQ395392	584	35.3	64.8
Zidi2	GQ395393	586	34.1	65.9
Temri	GQ395394	551	33.2	66.7
Soltani 2	GQ395395	554	33.4	66.7
Soltani 3	GQ395396	602	35.6	64.5
Gaa Zir	GQ395397	588	35.4	64.6
Baghli	GQ395398	574	34.9	65.2
Bidhi1	GQ395399	554	34.3	65.8
Assal boudchiche	GQ395400	562	32.9	67.1
Khadhouri	GQ395401	594	34.1	65.9
Khadhri	GQ395402	607	35.4	64.9
Khartoumi	GQ395403	560	34.2	65.9
Abiadh	GQ395404	588	34.9	65.2
Bither abiadh1	GQ395405	587	35.1	64.9
Besbessi	GQ395406	594	36.3	63.6
Dhokkar4*	GQ395413	549	33.7	66.3
Dhokkar5*	GQ395414	549	33.4	66.7
Dhokkar3*	GQ395415	546	33.5	66.5
Chetoui1	GQ395408	552	33.3	66.6
Chetoui2	GQ395409	590	34.0	66
Jrani*	GQ395410	595	34.5	65.5
Assafri*	GQ395411	599	34.4	65.6
Dhokkar 1*	GQ395412	557	33.3	66.8
Wahchi	GQ395407	548	33.7	66.3
Moyenne	-	576.3	34.1	65.9

*Male tree, **length: bp.

Table 3. Estimated relative frequencies of nucleotide substitutions in the *trn*L (UAA) intron of chloroplast DNA.

	Original nucleotides			
	A	T	C	G
A	-	*9.05*	*4.77*	**7.31**
T	*12.84*	-	**4.01**	*5.86*
C	*12.84*	**7.61**	-	*5.86*
G	**16.02**	*9.05*	*4.77*	-

Each entry shows the probability of substitution from one base (row) to another base (column) instantaneously. Only the entries within a row should be compared. Rates of different transitional substitutions are shown in bold and those of transversional substitutions are shown in italics.

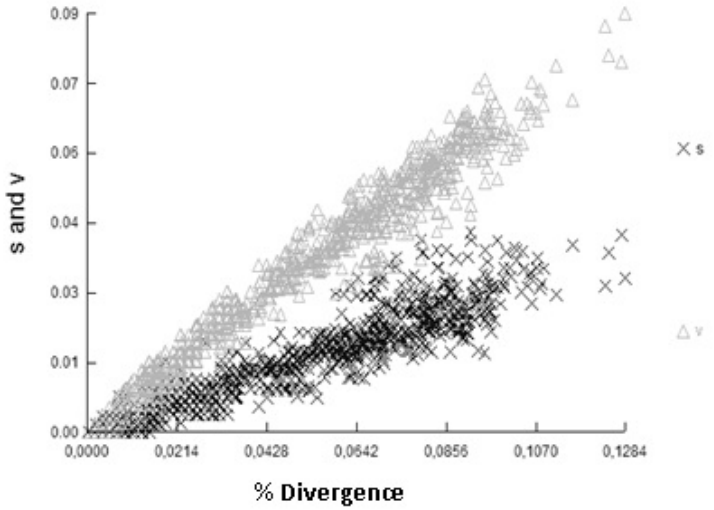

Figure 1. Percent divergence plotted against the number of transitions (S: cross) and transversions (V: triangles) for pairwise comparisons of the *trn*L intron.

Table 4. Sequences polymorphism and neutrality test based on chloroplast DNA of fig cultivars.

	Intron *trn*L (UAA)
N	49
S	57
P_{is}	32
h	37
π (JC)	0.018
H_d	0.951
K	8.16
R_M	4
Tajima's D	-2.081* (P < 0.05)
Fu and Li's D*	-2.15* (NS) (0.10> P > 0.05)
Fu and Li's F*	-2.54* (P < 0.05)
Fu's F_S	-23.48

*Significant.

Conserved group I intron sequence motifs (P, Q, R, and S) and repeat patterns

The alignment of the tRNALeu (UAA) intron revealed great sequence variability. Sequence variation was mostly confined to certain regions that, when the alignment was compared with the secondary structure predictions, were localized in some of the loops or hairpin structures. From *F. carica*, finding these patterns shows that the sequence R (GTGCAGAGACTCAA) was detected within the *trn*L (UAA) intron in all cultivars studied, and the motif S (AAGATAGAGTCC) was observed in most figs, with a deletion of T compared to that reported by Quandt and Stech (2005) (S: AAGAT**T**AGAGTCC). The sequence P: AATTCAGAGAAA was detected in all figs studied, except 4 cultivars whose sequences underwent mutations, including Khartoumi and Hemri 1 (AATT**T**AGAGAAA: C→T substitution), and Khadhouri and Kahli 1 (AATTAGAGAAA: C deletion). Quandt and Stech (2005) showed that the consensus sequences of the P motif in mosses were more

conserved than those of liverworts; their respective sequences were GATTCAGGGAAA and WATTCAGDGAAA. Figure 2 shows the 2D structure of the conserved motifs R, S, and P of the tRNALeu (UAA) intron of Tunisian figs; the secondary structure of the sequences has changed by mutation. In addition, Quandt and Stech (2005) showed that the consensus sequence of the Q motif of bryophytes is RATCCTGAGC. In *F. carica*, the Q motif (i.e., AATCCTGAGC) was detected in all cultivars, except one sequence of the caprifig Dhokkar 1 that exhibited a mutation (i.e., AA**A**CCTGAGC: T→A substitution).

Moreover, sequence analysis of the *trn*L intron allowed the detection of repeated sequences for all cultivars, such as $(CT)_4$, $(GA)_4$, $(A)_4T$, $(A)_3$, and $(C)_3$.

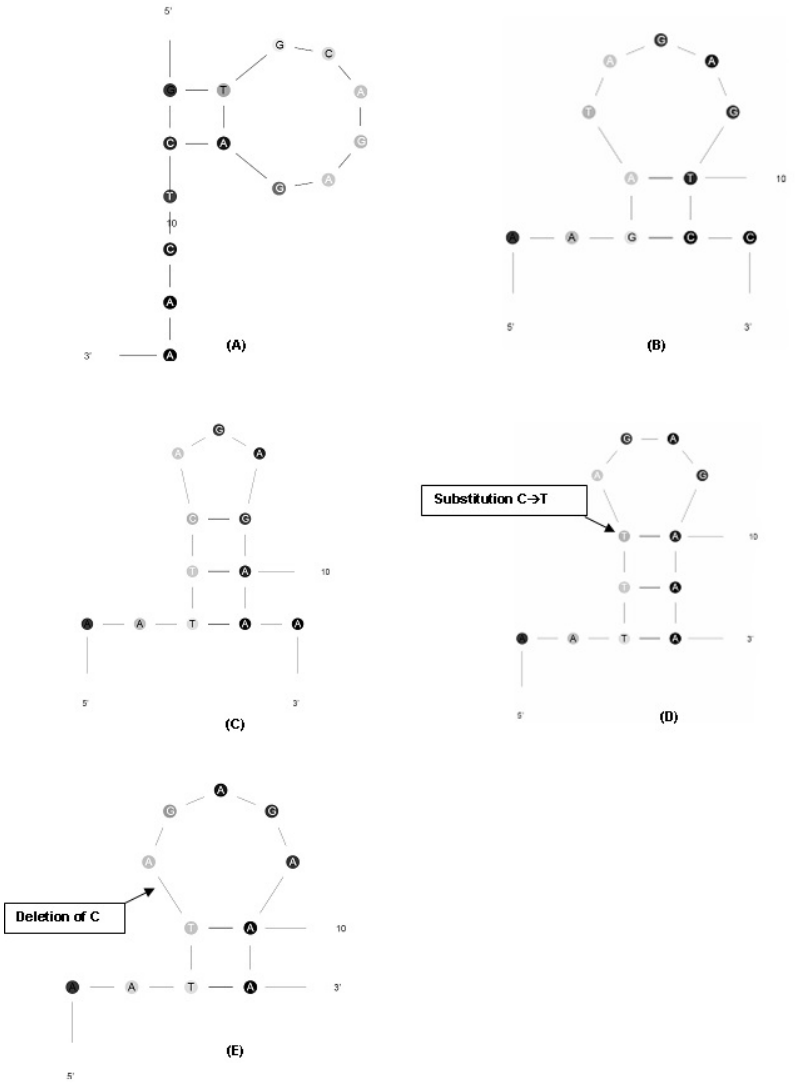

Figure 2. Secondary structures of tge conserved motif of the *trn*L intron: **A.** R motif, **B.** S motif, **C.** P motif, **D.** P motif sequence with C→T substitution, and **E.** P motif sequence with C deletion.

Genetic relationships of *trn*L intron sequences

The genetic distance MCL, based on comparisons of *trn*L intron sequences, showed they ranged from 0.00 to 0.065, with an average of 0.019. No distance (0.00) was observed between the following: Khalt and Sawoudi 2, Makhbech, Hamri, Grichy, Tounsi, Dhokkar Zarzis, Zaghoubi, Zidi 4, Dhokkar 4, and Chetoui 1; Kahli 2 and Dhokkar 2, and Zidi 3; Sawoudi 2 and Hamri, Grichy, Tounsi, Dhokkar Zarzis, Zaghoubi, Zidi 4, Dhokkar 4, and Chetoui 1; Makhbech and Hamri, Grichy, Tounsi, Dhokkar Zarzis, Zaghoubi, Zidi 4, Dhokkar 4, and Chetoui 1; Hamri and Grichy, Tounsi, Dhokkar Zarzis, Zaghoubi, Zidi 4, Dhokkar 4, and Chetoui 1; Dhokkar 2 and Zidi 3; Grichy and Tounsi, Dhokkar Zarzis, Zaghoubi, Zidi 4, Dhokkar 4, and Chetoui 1; Tounsi and Zaghoubi, Zidi 4, Dhokkar 4, and Chetoui 1; Dhokkar Zarzis and Zaghoubi, Zidi 4, Dhokkar 4, and Chetoui 1; Zaghoubi and Zidi 4, Dhokkar 4, and Chetoui 1; Zidi 4 and Dhokkar 4, and Chetoui 1; and Dhokkar 4 and Chetoui 1. On the contrary, the Bither Abiadh 1 and Khadhri accessions were the most divergent at their cpDNA as they showed the highest genetic distance (0.065).

The study identified 329 parsimony trees, and the most parsimonious tree had a length of 153 steps, with a CI of 0.484 and a RI of 0.507. Minimal homoplasy characterized the *trn*L intron sequences of the Tunisian fig tree.

The dendrogram (i.e., NJ) illustrating the genetic relationships between the studied fig shows the presence of 2 groups (Figure 3). The first cluster, marked (I), contained the Bither Abiadh 1, Kahli 1, Chetoui 2, and Soltani 3 varieties and the pollinator Jrani. The second cluster (II) contained all other cultivars and was divided into 2 sub-groups. The first, named (II_1), included the Baghli, Besbessi, Gaa Zir, Sawoudi 1, and Zidi2 varieties and Assafri pollinator; the second, named (II_2), contained the rest of the cultivars. Note that the 8 caprifigs were scattered between the 2 groups.

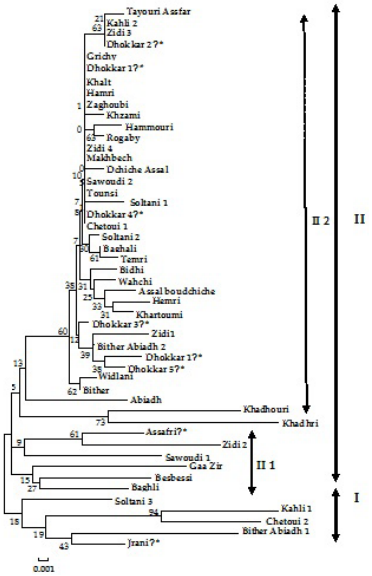

Figure 3. Neighbor-joining tree based on haplotypes of the *trn*L (UAA) intron of the chloroplast DNA of *Ficus carica* cultivars. Numbers at nodes indicate the bootstrap values.

Molecular evolution

Tajima's and Fu and Li's tests

Statistical tests of evolutionary neutrality Tajima's D (Tajima, 1989), and Fu and Li's (Fu and li, 1993) were performed using the aligned sequences. Our results yielded significant negative values obtained for both tests (Figure 4). The deviation from neutrality selectively highlighted by Tajima and Fu and Li tests, is explained by an excess of rare mutations, as identified in singleton in the sequences studied. Our results reflect the actions of positive selection and population expansion, as evidenced by the values of the test of Fu and Li (D*: -2.15*, 0.10 > P > 0.05; F*: -2, 54*, P < 0.05) for the *trn*L UAA intron of the cpDNA (Table 4; Figure 4).

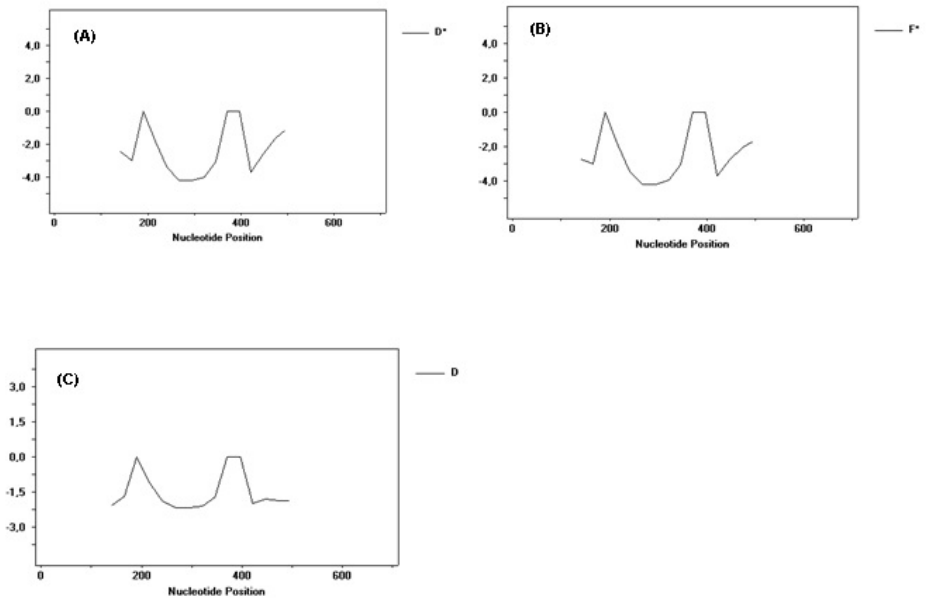

Figure 4. Base pair sliding window showing values of the Fu and Li (1993) [**A.** D* and **B.** F*] and Tajima (1989) D (**C**). Indices estimated for the *trn*L intron of cpDNA in *Ficus carica* L.

To clarify the cause of the deviation from the neutrality statistics, Fu was estimated. This quantity is known to be very powerful in detecting the deviation from neutrality and testing population growth and recent enlargement of the sample. The analysis of Table 4 shows strong and significant negative values for this parameter: Fu's F_S = -23.48 for the *trn*L intron.

The Fu and Li parameters and statistics of neutrality Fu reject neutrality in the *trn*L intron analyzed and suggest the action of a recent demographic expansion of the Tunisian fig. Our suggestions are also confirmed by the R2 index (Ramos-Onsins and Rosas, 2002) calculated, which returned a low value of 0.0603 for the *trn*L (UAA) intron for cpDNA.

Spectra variations of π and segregating sites (S) for the sequences studied were used to locate the sites affected by selection and clarify the scope of this selection (Figure 5). Sites at 200 and 400 bp were identified as the sites where selection operates in the *trn*L intron of the chloroplast DNA.

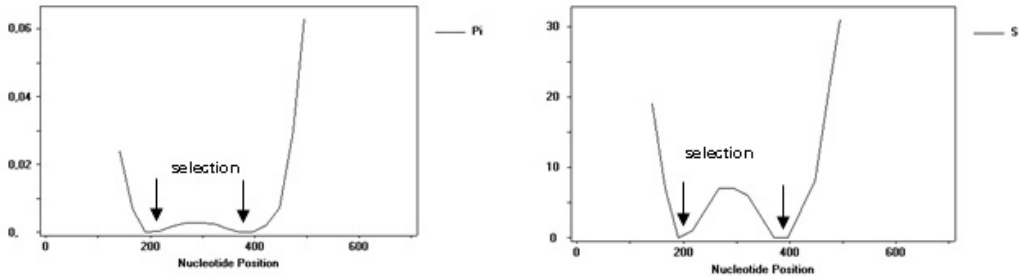

Figure 5. Hundred base-pair sliding window of chloroplast DNA *trn*L intron for *Ficus carica* L. Variability of the nucleotide diversity (P_I) and the segregating sites (S). Arrows indicate footprints of selection.

The empirical distribution of the P_I and segregating sites among pairs of individuals deviated from selective neutrality and appear to be multi-modal for the intron (Figure 6).

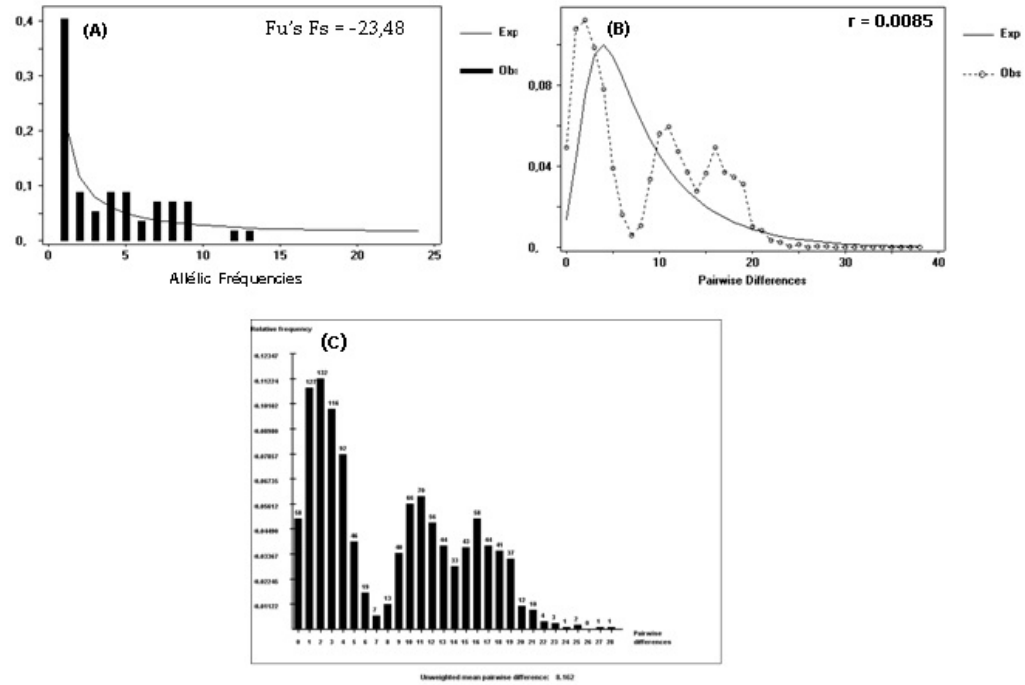

Figure 6. Site-frequency spectra. Solid lines in the site-frequency spectra indicate the expected distributions under neutrality and at equilibrium. Fu's F_S statistics and corresponding P values are given (**A**). Mismatch distribution based on pairwise nucleotide differences in the *trn*L intron. Solid 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 value are given (**B**). The distribution of the pairwise number of nucleotide differences and segregating sites among pairs of individuals of the chloroplast DNA sequences in *Ficus carica* L. covered in this study for the *trn*L intron. Empirical distribution is represented by vertical bars (**C**).

Haplotype distributions

The haplotype network based on the 49 sequences of the *trn*L intron of the cpDNA (Figure 7) showed the presence of 37 haplotypes, a clear demographic expansion, and a local evolution of figs from an ancestral haplotype. This is underlined by the star-like shape structure of the network at the turn of the founder haplotype H20 represented by Chetoui 1, Dhokkar 4, Zidi 4, Zaghoubi, Dhokkar 2, Tounsi, Grichy, Hamri, Makhbech, Sawoudi 2, and Khalt. H20 seems to be the ancestor of most chlorotypes formed during the evolution of the fig (Figure 7). The network structure is more complex with many divergent haplotypes (i.e., with one or 2 mutational steps) of the ancestral haplotype H20, and the branches are strongly divergent (i.e., many mutational steps). This seems to be consistent with the negative and significant values of Tajima's D, and the evidence of history and ancient origin of the fig in Tunisia.

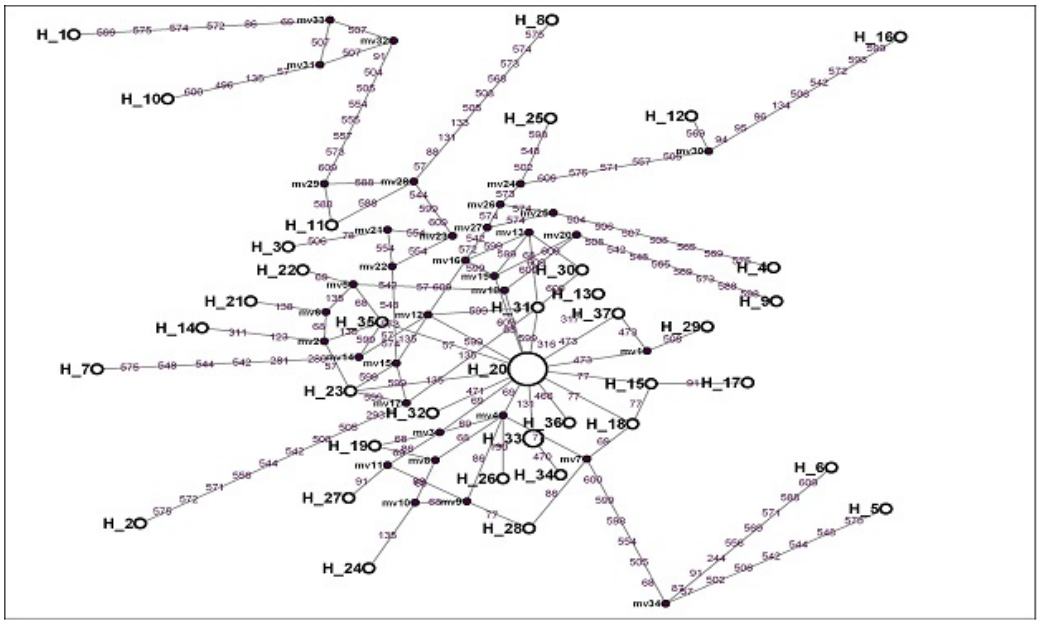

Figure 7. Network of haplotypes inferred from the *trn*L sequences. Nodes are proportional to haplotype frequencies, and branch lengths are proportional to the number of mutations. H1: Kahli 1; H10: Chetoui 2; H7: Abiadh; H2: Sawoudi 1; H6: Khadhri; H5: Khadhouri; H16: Zidi 2; H8: Bither Abiadh 1.

DISCUSSION

In this study, the *trn*L intron of chloroplast DNA was used to study the genetic diversity and molecular evolution of cultivated figs, the establishment of relationships, and elaboration on a scenario for the origin of the Tunisian fig germplasm. The sequence analysis showed variations in both the lengths and nucleotide compositions for all fig cultivars. The average size of the intron *trn*L (UAA) of the fig tree (576.3) is consistent with the size range recorded in several angiosperm taxa. Indeed, the size varies from 324 bp in *Calycanthus floridus* to 615 bp in *Orontium aquaticum* (Borsch et al., 2003). For Dipterocarpaceae species, the size of the intron *trn*L varies from 458 to 509 bp (Kajita et al., 1998). The length of this same

region of the cpDNA was also reported for *Triticum aestivum* (587 bp), *Hordeum vulgare* (555 bp), *Panicum virgatum* L. (557 bp) (Missaoui et al., 2006), and *Prunus* spp (574 bp) (Ben Mustapha et al., 2013). Among the gymnosperm lines, the size of this intron is 447 ± 2.26 bp in *Taxus*, 450 bp in *Pseudotaxus*, 447 bp in *Austrotaxus*, 477 bp in *Amentotaxus*, and 461-472 bp in *Cephalotaxus* (Hao et al., 2009). The GC content in the *trn*L intron of the Tunisian fig (GC = 34) is similar to the results obtained for *Prunus* spp (33.8%) (Ben Mustapha et al., 2013). The *Torreya* and *Cephalotaxus* gymnosperm lines exhibit GC base percentages of 35.12 and 35.19%, respectively (Hao et al., 2009). High values of the AT bases (65.9) were consequently found. The same result was obtained for switch grass by Missaoui et al. (2006) and Ben Mustapha et al. (2013) in *Prunus* spp.

The frequencies of the nucleotides were estimated and a similar composition has been reported for the *trn*L intron of cpDNA among Dipterocarpaceae species with frequencies of 0.40, 0.27, 0.15, and 0.17 for A, T, C, and G, respectively (Kajita et al., 1998); and in *Prunus* spp with frequencies of 0.34, 0.323, 0.163, and 0.174 for A, T, C, and G, respectively (Ben Mustapha et al., 2013). The ti/tv ratios were calculated (R = 0.312) for all bases. The relatively high content of AT (65.9) may partly explain the high proportion of transversions identified. Similar values were obtained in *Taxus* (R = 0.496), *Amentotaxus* (R = 0.384) (Hao et al., 2009), and *Prunus* spp (R = 0.293) (Ben Mustapha et al., 2013).

The *trn*L intron was considered a mosaic structure of conserved elements (internal guide sequence P, Q, R, and S) and common secondary structure elements, which are essential for correct splicing (Cech, 1990). In cases where variation in single nucleotides occurs, the secondary structure of the conserved motif of the intron is often retained. This can be seen in both cases where a change in one position is accompanied by a change in the base pairing strand so that different sequences either have a G:C or A:T/U base pair in this position, as well as where base pairing of the G:T/U type allows changes to occur on one strand without disturbing the base pairing structure. Most of the positions of the mutations in base pairing regions were also labeled to indicate the retained structure among the different sequences (Figure 2). There are two possible explanations for this: 1) there may be a requirement for base pairing to retain the structure needed for autocatalysis of the intron, which would give a selection pressure for compensatory changes, and 2) the tendency for a higher mutation rate in unpaired or mispaired bases on structures formed by single-stranded DNA during transcription, for example (Wright, 2000).

Moreover, sequence analysis of the *trn*L intron allowed the detection of repeated sequences for all cultivars such as $(CT)_4$, $(GA)_4$, $(A)_4T$, $(A)_3$, and $(C)_3$. These repeated patterns have also been reported in bryophytes at the same intron for chloroplast DNA (Quandt and Stech, 2005).

Genetic relationships of the *trn*L intron sequences show continuous diversity that characterizes local resources. The maximum parsimony tree, with an RI index of 0.507, indicated minimal homoplasy within the data set. Furthermore, our results demonstrate that the *trn*L intron is the seat of numerous substitutions.

Neutrality tests reflect the action of positive selection and population expansion or, as evidenced by the values of the test of Fu and Li (i.e., D*: -2.15*, $0.10 > P > 0.05$; F*: -2.54*, $P < 0.05$) for the *trn*L UAA intron of cpDNA. Sites at 200 and 400 bp are indicated as the sites where selection operates in the *trn*L intron of chloroplast DNA.

Note that the network analysis of different non-coding regions of ribosomal DNA (ITS: ITS1, 5.8S, ITS2) (Baraket et al., 2013), the *trn*L-*trn*F intergenic spacer of cpDNA and the *trn*L intron showed that the cultivar Hamri belongs to all ancestral haplotypes detected. This haplotype seems to be the ancestor of all nuclear and cytoplasmic haplotypes of the Tuni-

sian fig. Part of this study involves the identification of non-coding regions subjected to cytoplasmic selection as an evolutionary force. The search for the molecular signature of selection is made in the sequences of the *trn*L intron of cpDNA. The positive natural selection detected in the target intron was used to better understand the evolutionary past of the species and identify important functional genetic variants. Indeed, the selection leash traces of DNA sequences and fingerprints were detected when the genetic variability of a region was different from that expected under the hypothesis of selective neutrality. Our results show that intraspecific variation is obtained by analyzing the *trn*L intron of cpDNA. A deviation from neutrality was detected and explained by positive selection and/or demographic expansion of the Tunisian fig studied. In addition, cytoplasmic markers are very effective in building an evolutionary scenario of the case. The results indicate that positive selection and demographic expansion have contributed to the trends of nucleotide diversity and haplotype structure. In fact, according to Gillespie (2000) and Lagercrantz et al. (2002), each site in the genome is affected by selection and variation in certain genes and is more likely affected by natural selection.

CONCLUSIONS

Our study shows that the fig tree (*F. carica* L.) is one of the oldest domesticated fruit species and a good model for the identification of positive selection or selective sweep in the chloroplast genome. A great diversity and high differentiation chlorotypes were observed in the local germplasm. Accessions studied can be identified as a significant development unit and provide a rational basis for the identification of candidate units for conservation, as was reported by Andrianoelina et al. (2006).

ACKNOWLEDGMENT

Research supported by grants from the Tunisian Ministère de l'Enseignement Supérieur et de la Recherche Scientifique et de la Technologie projet Lab B02.

REFERENCES

Andrianoelina Andrianaivo O, Rakotondraoelina H, Ramamonjisoa L, Maley J, et al. (2006). Genetic diversity of *Dalbergia monticola* (Fabaceae) an endangered tree species in the fragmented oriental forest of Madagascar. *Biodivers. Conserv.* 15: 1109-1128.

Bandelt HJ, Forster P and Röhl A (1999). Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* 16: 37-48.

Baraket G, Ben Abdelkrim A, Chatti K, Saddoud O, et al. (2010). Molecular evolution of chloroplast DNA in fig (*Ficus carica* L.): Footprints of sweep selection and recent expansion. *Bioch. Syst. Ecol.* 38: 563-575.

Baraket G, Ben Abdelkarim A, Mars M and Salhi-Hannachi A (2013). Genetic Diversity and Molecular evolution of the Internal Transcribed Spacer (ITSs) of nuclear ribosomal DNA in the Tunisian fig cultivars (*Ficus carica* L; Moracaea). *Bioch. Syst. Ecol*. 48: 20-33.

Bayer RJ and Starr JR (1998). Tribal phylogeny of the Asteraceae based on two non-coding chloroplast sequences, the *trn*L intron and *trn*L-*trn*F intergenic spacer. *Ann. Missouri Bot. Gard*. 85: 242-256.

Bell-Pedersen D, Quirk S, Clyman J and Belfort M (1990). Intron mobility in phage T4 is dependent upon a distinctive class of endonucleases and independent of DNA sequences encoding the intron core: mechanistic and evolutionary implications. *Nucleic Acids Res.* 18: 3763-3770.

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.

Borsch T, Hilu KW, Quandt D, Wilde V, et al. (2003). Noncoding plastid *trn*T-*trn*F sequences reveal a well resolved phylogeny of basal angiosperms. *J. Evol. Biol.* 16: 558-576.

Cech TR (1990). Self-splicing of group I introns. *Annu. Rev. Biochem*. 59: 543-568.

Cech TR, Damberger SH and Gutell RR (1994). Representation of the secondary and tertiary structure of group I introns. *Nat. Struct. Biol*. 1: 273-280.

Clegg MT, Learn GH and Golenberg EM (1991). Molecular Evolution of Chloroplast DNA. In: Evolution at the Molecular Level (Selander RK, Clark AG and Whittam TS, eds.). Sinauer Publishers, Sunderland, 135-149.

Dellaporta SL, Wood J and Hicks JB (1983). A plant DNA minipreparation: Version II. *Plant Mol. Biol. Rep*. 1: 19-21.

Downie SR, Katz-Downie DS and Cho KJ (1996). Phylogenetic analysis of apiaceae subfamily Apioideae using nucleotide sequences from the chloroplast *rpo*C1 intron. *Mol. Phylogenet. Evol*. 6: 1-18.

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

Gielly L and Taberlet P (1994). The use of chloroplast DNA to resolve plant phylogenies: noncoding versus *rbc*L sequences. *Mol. Biol. Evol*. 11: 769-777.

Gillespie JH (2000). Genetic drift in an infinite population: the pseudohitchhiking model. *Genetics* 155: 909-919.

Hao da C, Huang BL, Chen SL and Mu J (2009). Evolution of the chloroplast *trn*L-*trn*F region in the gymnosperm lineages Taxaceae and Cephalotaxaceae. *Biochem. Genet*. 47: 351-369.

Jukes TH and Cantor CR (1969). Evolution of protein molecules. In: Mammalian protein metabolism (Munroled HN, ed.). Academy Press, New York, 31-132.

Kajita T, Kamiya K, Nakamura K, Tachida H, et al. (1998). Molecular phylogeny of Dipetrocarpaceae in Southeast Asia based on nucleotide sequences of matK, *trn*L intron, and *trn*L-*trn*F intergenic spacer region in chloroplast DNA. *Mol. Phylogenet. Evol*. 10: 202-209.

Kita Y, Ueda K and Kadota Y (1995). Molecular phylogeny and evaluation of the Asian *Aconitum* subgenus *Aconitum* (Ranunculaceae). *J. Plant Res*. 108: 429-442.

Kluge AG and Farris JS (1969). Quantitative phylogenetics and the evolution of anurans. *Syst. Zool.* 18: 1-32.

Lagercrantz U, Kruskopf Osterberg M and Lascoux M (2002). Sequence variation and haplotype structure at the putative flowering-time locus COL1 of *Brassica nigra*. *Mol. Biol. Evol*. 19: 1474-1482.

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

Missaoui AM, Paterson AH and Bouton JH (2006). Molecular markers for the classification of switchgrass (*Panicum virgatum* L.) germplasm and to assess genetic diversity in three synthetic switchgrass populations. *Genet. Res. Crop Evol*. 53: 1291-1302.

Morton BR and Clegg MT (1993). A chloroplast DNA mutational hotspot and gene conversion in a noncoding region near *rbc*L in the grass family (Poaceae). *Curr. Genet*. 24: 357-365.

Nei M and Tajima F (1983). Maximum likelihood estimation of the number of nucleotide substitutions from restriction sites data. *Genetics* 105: 207-217.

Nickrent DL and Soltis DE (1995). A comparison of angiosperm phylogenies from nuclear 18S rDNA and *rbc*L sequences. *Ann. Missouri Bot. Gard.* 82: 208-234.

Quandt D and Stech M (2005). Molecular evolution of the *trn*L (UAA) intron in bryophytes. *Mol. Phylogenet. Evol*. 36: 429-443.

Ramos-Onsins SE and Rozas J (2002). Statistical proprieties of new neutrality tests against population growth. *Mol. Biol. Evol*. 19: 2092-2100.

Rogers AR and Harpending H (1992). Population growth makes waves in the distribution of pairwise genetic differences. *Mol. Biol. Evol*. 9: 552-569.

Saitou N and Nei M (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol*. 4: 406-425.

Salhi-Hannachi A, Chatti K, Mars M, Marrakchi M, et al. (2005). Comparative analysis of genetic diversity in two collections figs cultivars based on random amplified polymorphic DNA and inter simple sequence repeats fingerprints. *Genet. Res. Crop Evol*. 52: 563-573.

Sambrook J, Fritsch EF and Maniatis T (1989). Molecular cloning: a laboratory manual. 2nd edn. Cold Spring Harbor, New York.

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

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

Tamura K, Peterson D, Peterson N, Stecher G, et al. (2011). MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol*. 28: 2731-2739.

Wright BE (2000). A biochemical mechanism for nonrandom mutations and evolution. *J. Bact*. 182: 2993-3001.
Xia X (2000). Data Analysis in Molecular Biology and Evolution. Kluwer Academic Publishers, Boston, 276.
Yu Z, Chen L, Qi-Jia W, Yu W, et al. (2007). GISSD: Group I intron sequence and structure database. *Nucleic Acids Res*. 36: D31-37.
Zohary D and Spiegel-Roy P (1975). Beginnings of fruit growing in the old world. *Science* 187: 319-327.
Zohary D and Hopf M (1993). Domestication of Plants in the Old World. Clarendon Press, Oxford.
Zuker M (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res***.** 31: 3406-3415.