Genetic Diversity of Fig (*Ficus Carica* L.) Germplasm From The Mediterranean Basin As Revealed by SSR Markers

Athanasios Sclavounos
Hellenic Ministry of Rural Development and Food: Ypourgeio Agrotikes Anaptyxes kai Trophimon
Petros Roussos
Agricultural University of Athens: Geoponiko Panepistemio Athenon
Sotiria Milla
Ionian University: Ionio Panepistemio
Panagiotis Kostas
Hellenic Ministry of Rural Development and Food: Ypourgeio Agrotikes Anaptyxes kai Trophimon
Yiannis Samaras
Ionian University: Ionio Panepistemio
Carlo Pozzi
University of Milano–Bicocca: Universita degli Studi di Milano-Bicocca
Johiruddin Molla
ICRISAT: International Crops Research Institute for the Semi-Arid Tropics
Annapurna Chitikineni
ICRISAT: International Crops Research Institute for the Semi-Arid Tropics
Rajeev K. Varshney
ICRISAT: International Crops Research Institute for the Semi-Arid Tropics
Andreas Voloudakis (avoloud@aua.gr)
Agricultural University of Athens, Greece https://orcid.org/0000-0002-0369-4904

Research Article

Keywords: Ficus carica, genetic diversity, genetic resources, microsatellite, molecular marker, genotyping

Posted Date: December 30th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1140396/v1

Version of Record: A version of this preprint was published at Genetic Resources and Crop Evolution on December 27th, 2022. See the published version at https://doi.org/10.1007/s10722-022-01509-0.

Abstract

Fig (*Ficus carica* L.) tree is cultivated worldwide and is highly appreciated for its fruit, which is consumed fresh or dried, having high nutritional and pharmaceutical value and for these reasons there is an increasing interest for its cultivation. In the present study, an *ex situ* collection of 60 fig accessions (41 indigenous Greek and 19 from other Mediterranean countries) was established and its diversity was analyzed using eight simple sequence repeat (SSR) loci. Greek fig genotypes showed relatively low allelic variation (average number of SSR alleles per locus was 3.3), an excess of heterozygosity (mean He = 0.449 and Ho = 0.537), and extensive outbreeding (mean F index -0.184). Cluster analysis showed that the established fig population exhibited weak genetic structure with the majority of the genetic variation (69%) being present within individual members of the clusters. Both cluster and principal coordinate analysis confirmed that there is no correlation between genetic makeup and geographical origin of the fig accessions. Polymorphism information content (PIC) with an average of 0.398 was reasonably informative. An identification key scheme for fig cultivars that will be useful in cultivar discrimination and intellectual property protection was developed. This work will contribute to a sustainable fig production regionally and worldwide, through the establishment and conservation of a reference fig collection, providing germplasm for future breeding efforts.

Introduction

Fig (*Ficus carica* L.) (2n=26) belongs to the Moraceae family, known worldwide for its fruit and the presence of latex in all plant parts. Figs are eaten fresh or dried, are rich in phenolic antioxidants and nutrients with high fiber content (Vinson et al. 2005), being ideal for the human diet. According to Vavilov (1951), fig's origin is Transcaucasian, whereas other authors place it in southern Arabia or the eastern Mediterranean basin (Stover 2007). From its center of origin, fig spread to the Mediterranean basin, and in the 16th century to the east (Southeast Asia, China, Japan) as well as the New World. Nowadays, the cultivation of fig is worldwide, including Australia and South Africa

In Greece, fig is cultivated almost in the entire country, with three main production centers, namely Attica (Mesogaia region), Evia (Kimi and Taxiarches regions) and Messinia (Peloponnese prefecture). Fig cultivation declined in Greece in the last decades; in 1961 fig was cultivated on 250,000 acres, the production was 169,259 tons, with a yield of 700 kg/acre, whereas in 2011 only 38,000 acres were cultivated, with a total production of 9,400 tons and a yield of 250 kg/acre (FAO stat 2012, www.fao.org).

Fig is threatened by genetic erosion, mainly due to urbanization, monoculture, and introduction of plant material from remote regions. Today, as there is an increasing interest in fresh and dried figs, fig cultivation assumes new perspectives. Since fig is not subjected to intensive breeding (Flaishman et al. 2008) the improvement of fig cultivars should be based on the rich genetic diversity present in established fig populations. The ambiguity in the description of fig cultivars, the lack of official cultivar names and the existence of variants within cultivars hinder proper fig cultivar identification (Perez-Jimenez et al. 2012). Condit (1955) listed more than 700 fig cultivars, with a great deal of confusion in cultivar identification and their relationships still existing.

Plant germplasm characterization, aiming at its conservation, is traditionally carried out using morphological or agronomical traits. These criteria are often variable across years and locations since these phenotypic traits are influenced by the genotype-environment interactions. In contrast, DNA-based data are stable, reliable, and detectable in all tissues regardless of developmental and differentiation stage and are not confounded by environmental, pleiotropic and epistatic effects (Mondini et al. 2009). Molecular markers such as microsatellites (simple sequence repeats, SSRs), RAPDs, ISSRs, RFLPs and others have been used in fingerprinting and assessing genetic diversity in various fig collections (Papadopoulou et al. 2001; Chatti et al. 2010; Perez-Jimenez et al. 2012; Ganopoulos et al. 2015; Boudchicha et al. 2018; Rodolfi et al. 2018). Microsatellites provide adequate resolution of germplasm differences due to their high polymorphism and codominance, are simple, quick, relatively inexpensive, and exhibit high reproducibility among laboratories, and as a result are still used for fingerprinting in plant species (Gupta and Varshney 2000; Mondini et al. 2009).

In the present study, the genetic diversity in fig germplasm was evaluated for an established population in an *ex situ* collection (60 fig accessions, from four Mediterranean countries and the USA) and an identification key for fig cultivars was developed using SSRs. The work was also aimed at assigning each accession to a group based on genetic diversity, thus evaluating the structure of the fig population.

Materials And Methods

Plant material

In the present study the fig population comprised of 60 accessions, with 41 from Greece and 19 from other Mediterranean countries (Table 1). In particular, 33 of them were selected from different regions of Greece, based on their special agronomical, morphological, and fruit quality characteristics, based on information obtained from farmers and authors' personal field inspection. The name of the accession usually represents a geographical qualifier or sometimes it derives from a morphological or fruit quality characteristic. Plants were vegetatively propagated by cuttings and planted in the experimental orchard of the Agricultural University of Athens (AUA). The remaining 27 fig accessions were acquired as trees and planted in the same experimental orchard and used as reference material. Eight of the 27 accessions were Greek fig cultivars analyzed previously using RAPD markers (Papadopoulou et al. 2001). The remaining 19 accessions were internationally well-known fig varieties from other countries (11 from Italy, six from Cyprus, one each from Turkey and the USA) (Suppl. Table 1).

DNA isolation

Plant DNA was isolated from fig leaves of all 60 accessions using the CTAB method (Murray and Thompson 1980). The DNA concentration was estimated spectrophotometrically and its integrity was evaluated by electrophoresis on 0.8% agarose gel followed by ethidium bromide staining. DNA suitability as PCR

template was checked by PCR reaction using primers for the ITS (Internal Transcribed Spacer) locus, following the methodology described by Roy et al. (2010).

SSR performance and data analysis

Eight SSR markers, namely MFC1 to MFC8, developed by Khadari et al. (2001) for fig, were used in this study. DNA amplification reactions were carried out in a total volume of 25 μl containing 0.5 mM of each PCR primer, 200 mM of each deoxynucleotide triphosphate, 1.5 mM of $MgCl_2$, 1 U of *Taq* DNA polymerase (New England Biolabs, USA) and 50 ng of template DNA (Khadari et al. 2001), using a PTC-200/A100 thermocycler (BioRad, USA). No DNA negative control reactions were performed.

PCR products were resolved using a 12% non-denaturing PAGE in a 20x20 gel (Biorad Protean II, USA) at 60 V for 20 min, followed by 180 V for 5 hrs. Gels were subsequently stained with ethidium bromide and photographed under UV, with photos digitized for further analysis. A DNA ladder (50 bp GeneRuler, Thomas Scientific, USA) was loaded in three wells in each gel, in asymmetric locations, to avoid gel orientation problems when scoring and assist in allele size determination (Hoffman and Amos 2005).

The amplified bands per SSR were scored for each fig accession, using GelAnalyzer (2010a) (www.GelAnalyzer.com). Only gels/lanes with unambiguous band patterns, after background subtraction using GelAnalyzer, were considered for allele assignment. A band was accepted, when the corresponding fluorescence intensity value, from the digitized photos, was >10 fluorescent units. Bands with a fluorescence intensity < 35% (i.e. stutter bands) of the main fluorescence intensity value were filtered out following previous recommendations (Ewen et al., 2000; UPOV/INF/17/1, 2010). The microsatellite alleles were sized using a standard curve generated for each gel, employing the known molecular size DNAs of the DNA ladder. Since PCR products of a fig accession per microsatellite were electrophoresed two to four times (each time in a different gel), the size of an allele was estimated several times, with the mean value recorded as the allele size. The allele sizes, for each microsatellite and accession, were recorded in an Excel spreadsheet (Microsoft Inc., Redmond, USA) producing thus a data matrix for data storage and further processing. SSR allele size data were binned using Flexibin (Amos et al., 2007), following the methodology described by Ghosh et al. (1997). A final correction of the allele's size was done by visual gel inspection as previously reported (Pompanon et al., 2005; Hoffman and Amos, 2005).

Based on the SSR allele size data, genetic variability parameters were computed for all the 60 fig accessions originating from Greece and other Mediterranean countries using GenAlEx v.6.5 (Peakall and Smouse, 2012). Genetic diversity estimates such as the number of alleles per locus (Na), number of effective alleles per locus (Ne), observed (Ho) and expected heterozygosity (He) and fixation index (F) were computed. Deviations from Hardy-Weinberg equilibrium (HWE) at individual loci were tested using GENEROP 3.1 (Rousset, 2008) (Table 1). In addition, the polymorphism information content (PIC) value per locus was estimated using PICCalc software (Nagy et al., 2012).

Estimation of genotyping error

MicroChecker v.2.2.3 (Van Oosterhout et al. 2004) was used to statistically estimate the percentage of null (nonamplified alleles due to nucleotidic changes in flanking sequences of the SSR) alleles per SSR, which is the main non-technical contributor to the genotyping error.

In order to estimate genotyping error, a subset (20%) of the fig population was reanalyzed (Pompanon et al. 2005) following the same methodology, except that a different PCR thermocycler (Eppendorf Mastercycler Gradient 5341, USA), DNA polymerase (Phusion® High-Fidelity, NEB, USA), and a new fig DNA preparation were used. In addition, to strengthen the reliability of the obtained results: a) independent random PCRs, for each of the eight SSRs, were conducted again for approximately 20% of the fig accessions and were re-genotyped, b) for each PCR, PCR samples were re-electrophorized two to four times in different gels, and c) the allele sizes were scored twice by two different persons (Hoffman and Amos 2005).

Cluster analysis

In order to depict the genetic relationships among accessions of the established fig population the FreeTree program (Pavlicek et al. 1999) was employed. The binary data for all the eight SSR loci were used to compute the Nei and Li distance (Nei and Li 1979). The obtained matrix was subjected to cluster analysis (CA) followed by phylogenetic tree construction employing the UPGMA method. Bootstrap method with 500 resampled datasets was used to test the reliability of each interior branch of the phylogenetic tree.

A DARwin file with extension ".DIS" stores the dissimilarity lower semi-matrix (without the diagonal) as computed by the software. Dissimilarity re-sampling done with 10,000 bootstrap and each semi-matrix is successively recorded at the end of the file. Dissimilarity based cluster analysis was performed and dendrogram was done following the Weighted Neighbor-Joining (WNJ) method with the help of DARwin v6 (Perrier and Jacquemoud-Collet 2006). The final cluster of WNJ method confirmed by 10000 bootstraps. In the dendrogram, the scale defined the edge length.

Analysis of molecular variance (AMOVA) was carried out by GenAlEx, with the analysis based on groups as revealed by the above CA analysis using the binary data for all SSR loci. Multivariate relationships among 60 fig accessions, grouped according to their country of origin, were examined with principal coordinate analysis (PCoA) using GenAlEx and binary data from all fig accession for all SSR loci.

Population structure

Possible population structure was investigated using a model-based Bayesian procedure implemented in the software STRUCTURE v2.3.2 (Pritchard et al. 2000). The analysis was carried out using a burning period of 10,000 iterations and a run length of 200,000 MCMC replications. We tested a continuous series of K, from 1 to 8, in 10 independent runs. We did not introduce prior knowledge about the population of origin, and assumed correlated allele frequencies and admixture. For selecting the optimal value of K, DK values (Evanno et al. 2005) were calculated using the Structure Harvester software (Earl and vonHoldt 2012).

Establishment of an identification key for fig

In order to establish an identification (Id) key for fig, the methodology of Tessier et al. (1999) was followed. According to this method two parameters were estimated; the confusion probability C*j* and the discriminating power D*j* (*Dj* = 1 - *Cj*). These parameters were computed by the frequencies of allelic pattern revealed per SSR locus in the established fig population.

To determine the sequence and the number of SSRs that are sufficient to establish an identification key for fig: a) an alphabetic character was given to each SSR allele, with fig accessions exhibiting the same SSR allele pattern obtaining the same alphabetic character designation, b) SSR loci were sorted by descending order of D*j*, c) fig accessions with the same alphabetic designation per SSR were pooled together in the same group, and d) the total number of expected and the observed non-differentiated fig pairs of comparisons were determined after the completion of grouping. MFC6, regardless of its high D*j* (0.85), was not included in the Id key due to the lack of banding pattern in 12 out 60 of fig accessions.

Results

Genetic diversity among fig genotypes revealed by SSR markers

In the present study, 60 fig accessions were examined using eight microsatellites (SSRs), designated MFC1 to MFC8 (Khadari et al. 2001). All figs revealed PCR banding pattern for seven out of the eight SSRs (Suppl. Fig. 1). For MFC6, no banding pattern (no PCR products) was observed for 12 out of 60 fig accessions, namely seven from Greece (Acc. No 118, 124, 131, 139, 140, 142, 146), three from Italy (Acc. No 251, 252, 253), one from Cyprus (Acc. No 236) (Suppl. Fig. 1), and one from Turkey (Acc. No 230). PCR reactions for the 12 above-mentioned fig accessions were repeated confirming the initial observation.

For each fig accession showing a banding pattern, one allele (homozygous individual) or two alleles (heterozygous individual) were identified and their size was estimated. For fig accessions from Greece, a total of 28 alleles were revealed for the eight SSRs studied, with an average of 3.5 alleles per SSR locus. In fig accessions from other countries, two more alleles were revealed, both in MFC1, namely the 198 bp in Columbra nera (Acc. No 253), originating from Italy, and the 170 bp in Blank Klirou (Acc. No 210), White Klirou (Acc. No 236), Winter Blank Cyprus (Acc. No 220) and White Prodromou (Acc. No 233), all originating from Cyprus. The number of alleles per locus (Na) varied between 2 (for MFC4, MFC5, and MFC8) to 6 (for MFC3) (Table 1).

Alleles with a frequency < 5% were designated as rare alleles (Table 1). In particular, allele 192 bp (Acc. No 109) of MFC1, 158 bp (Acc. No 145) of MFC2, 248 bp (Acc. No 127) and 290 bp (Acc. No 161) of MFC6 was observed only once, in fig accessions originating from Greece. In addition, allele 194 bp was shared between two fig accessions from Greece (Acc. No 117 and 118) and one from Italy (Acc. No 244). Furthermore, allele 126 bp (Acc. No 118 and 131) and 154 bp (Acc. No 142 and 155) were identified as two rare alleles for MFC3. Lastly, alleles 170 bp and 198 bp of MFC1 were observed only in fig accessions originating from Cyprus and Italy, respectively.

For each SSR locus, several genetic parameters were computed, with the results shown in Table 1. The observed heterozygosity for the fig accessions originating from Greece using seven SSRs (MFC6 was not included) ranged from 0.220 in MFC8 to 0.902 in MFC4 (with mean value 0.537) and the expected heterozygosity ranged from 0.195 in MFC8 to 0.604 in MFC2 (with mean value 0.449). For each SSR, except MFC1 and MFC3, the observed heterozygosity was higher than the expected. Consequently, most of the F values for the SSR loci are negative (average value -0.184) indicating a tendency for heterozygosity selection. The above data indicated that allele frequencies did not significantly deviate from the Hardy-Weinberg equilibrium (P>0.05) (Table 1).

MFC6 was not included in the determination of the mean value of the genetic parameters because in 7 out of 41 the fig accessions from Greece no banding pattern was observed, and genetic variability parameters were estimated only for the remaining 34 fig accessions. Nevertheless, MFC6 is a highly informative locus, possessing five alleles, He and Ho of 0.676 and 0.622, respectively (Table 1) and a PIC value of 0.584 (Table 2).

The MicroCheker results indicated that the presence of null alleles was found to be statistically nonsignificant, with their frequencies per SSR ranging from 0 to 3.53%. No new alleles appeared for any of the SSRs and the genotyping procedure was repeated for approximately 20% of fig accessions. The genotyping error is non-significant since it was found that allele size shift of two base pairs was observed only in 3% of the cases, indicating that the methodology used is reproducible.

Genetic relationship among fig genotypes - Population structure of fig germplasm collection

The genetic relationship among the fig genotypes is presented in a dendrogram produced using UPGMA (Suppl. Fig. 2). Fig accessions formed two groups, designated as group I and II. Group I included all fig accessions, except two, both originating from Greece, and which constituted group II. Group I could be separated into four subgroups, namely A, B, C, and D. Moreover, subgroup A was further divided into two major clusters, AI and AII, including 21 and 19 fig accessions, respectively. Based on the above clustering, AMOVA analysis indicated that 69% of the total population variation could be accounted from within-subgroup individuals and 31% from among subgroup difference. The robustness of the branches was evaluated by bootstrap analysis. The first node was significantly supported (100%) by bootstrap analysis, indicating that the two groups (I and II) could clearly cluster. In contrast, the rest of the early nodes were not significantly supported by bootstrap analysis indicating that subgrouping, clusters (AI and AII) and lower clustering present a weak separation.

From the dendrogram, it became obvious that fig accessions from different countries or from different regions of Greece were mixed into the same subgroup, group or cluster (Suppl. Fig. 2). PCoA analysis supported this observation since the percentage of variability that could be explained by grouping figs by geographical origin (country) was only 45.28% of the total variation in the established fig population (Fig. 1).

In cluster AII, Maurosykia (Acc. No 108) and Zakynthos (Acc. No 160) were pooled together; the branch was supported by significant bootstrap (86%), suggesting that they are two representatives of the same genotype and could be synonymous. However, it should be noted that these two fig accessions are

morphologically different (Ntanos et al. 2015) and were collected from remote regions of Greece, namely an island in Ionian Sea (southwestern Greece) and Macedonia prefecture (northern Greece), respectively, and thus their close genetic relationship should be investigated further. The second instance of a very close genetic relationship, which was supported by a 65% bootstrap value, is that of the San Pietro (Acc. No 213) and Dottato (Acc. No 234) from Italy. Both are worldwide well-known varieties, which produce fruit twice per year, and differ in some morphological characteristics, such as fruit color (Ntanos et al. 2015). Their close genetic relationship was also shown in a previous study using RAPD (Papadopoulou et al. 2001). On the other hand, Maurosykia (Acc. No 108), originating from Macedonia prefecture, and Maurosykia 19 (Acc. No 155), originating from Eleia prefecture, are two fig accessions possessing the same name but originate from two remote regions of Greece (northern and southern Greece) and were found to be genetically different. The only caprifig accession included in the population, namely Ag. Anargyri (Acc. No 124), was grouped in a separate group (C) together with a fig accession from Greece (Acc. No 132, Tzavelas Black Large), which is gynodioecious producing edible figs. To further confirm the results described in Fig. 2, a WNJ analysis was produced that is reported in Fig. 3 and Suppl. Fig. 2.

Genotypes of the figs originating from Greece

Allele patterns, per SSR locus, in fig accessions, revealed different genotypes, ranging between two for MFC4, MFC5 and MFC8 to nine for MFC3, as shown in Table 2. Combining the data for all eight SSR markers indicated the presence of 40 different genotypes for the 41 fig accessions originating from Greece (Suppl. Table 1). Only two fig accessions, namely Zakynthos (Acc. No160) and Maurosykia (Acc. No 108), could not be distinguished genotypically; but could be phenotypically distinguished. Similarly, 18 out of 19 fig accessions from other countries revealed different genotypes. Only Dottato (Acc. No 234) and San Pietro (Acc. No 213), both from Italy, exhibit the same genotypic pattern in the present analysis.

The genetic structure of the fig tree cultivars was investigated by a Bayesian-based population assignment analysis using STRUCTURE software (Pritchard et al. 2000). The aim is to group different individuals into K different clusters based on genetic similarity, so that the individuals within a cluster are more similar to each other than to individuals outside the cluster. In our case, we do not have a clear prior hypothesis for the number of clusters to expect, and thus the clustering algorithm was run several times with different values of K. Each run gave a different probability reflecting the fit of the clusters to the data, assuming Hardy–Weinberg equilibrium. The K value that gives the best fit to the data is the best estimate of the true number of clusters, or populations. Our results show a clear maximum for ΔK at K = 3 (Fig. 2A). As shown in Fig. 2B, most cultivars were divided into three sub-populations.

Identification key for figs

In this study, an identification (Id) key was generated based on the discriminating power (D*j*) of each SSR. The SSR loci were hierarchically ordered according to their D*j* values. MFC3 was placed first, as it is the most discriminative among the seven SSRs (MFC6 was not included at this stage). As a result, the Id key produced was: MFC3-MFC1-MFC2-MFC7-MFC8-MFC5. MFC4, having the smallest D*j*, was left out from the Id key since it was not able to differentiate further figs (Table 2). The sixty fig accessions in this study produced 1770 pairs of comparisons. Using the above Id key, only 12 from the 1770 pairs were not differentiated; 54 different fig genotypes were differentiated in the 60 fig accessions.

MFC6 was not included initially in the Id key due to the lack of banding pattern in 12 out of 60 fig accessions, regardless of its high D*j* (0.85). Based on this D*j* value, MFC6 would have been ranked first in the Id key, but this may introduce ambiguity due to the lack of the locus in 12 accessions. Therefore, we propose to introduce MFC6 at the end of the Id key (step 6, Table 2), where most figs are already differentiated. Using the MFC6 data, it was possible to further discriminate six fig accessions of the 12 non-differentiated pairs. If the absence of banding pattern for MFC6 is considered a genetic characteristic, then four more pairs could also be differentiated. As a result, from the 1770 pairs of fig accessions only two pairs could not be differentiated (Table 3); therefore, 58 different fig genotypes could be distinguished among the 60 fig accessions. In summary, the proposed identification key for figs is as follows: MFC3-MFC1-MFC2-MFC7-MFC8-MFC5-MFC6. MFC4 did not reduce the number of non-differentiated pairs (Table 2) and thus was not included in the Id key.

Discussion

Genetic erosion threatens the local fig varieties and landraces and intensifies the need to protect them. For this reason, an *ex situ* fig collection was established in the experimental orchard of the Agricultural University of Athens. This comprised mainly of uncharacterized Greek fig accessions, along with few partly characterized Greek varieties (Papadopoulou et al. 2001) along with some well-known foreign cultivars. It is well known, that an *ex situ* collection where landraces are preserved without the accompanying information on their characteristics and registration of their region of origin would be of limited value. The information describing plant germplasm could be used for identification purposes, recognition of deficiencies of the collection and planning future efforts to strategically enrich it with new plant material. Such information includes morphological, agronomical features, biochemical and molecular data. In the present study, the genetic characterization of the established fig population was described using SSR markers. Such investigations with plant genetic resources are a prerequisite for breeding crops in order to face new challenges, including climate change. Data obtained were also used to propose an identification key scheme.

In a number of genetic studies using microsatellites, genotyping errors that are due to null alleles (nonamplified alleles), DNA degradation and low DNA concentrations are increasingly recognized as important factors that could render the conclusions doubtful (Hoffman and Amos 2005). Especially null alleles that usually result from changes in flanking region sequence of the SSR could alter the estimation of the genetic parameters of the population under study. In the present study, the frequency of null alleles was statistically estimated per SSR and it ranged from 0 to 3.53%. This range is considered non-significant since frequencies of 5-8% introduce only a small bias in the genetic parameters investigated (Chapuis and Estoup 2007). Moreover, re-genotyping 20% of the members of the population studied here reinforced the reliability of the results.

The wide genetic basis of our *ex situ* germplasm collection was assessed by SSRs. PCoA analysis indicated that nearly all individuals in the established fig collection are genetically different. Only two fig pairs were found to be genetically identical: Maurosykia (Acc. No 108) and Zakynthos (Acc. No 160); San Pietro (Acc. No 213) and Dottato (Acc. No 235). However, it is noteworthy that the figs in the above pairs differ morphologically (Ntanos et al. 2015) indicating that several methods need to be employed in studying germplasm. The close genetic relationship of San Pietro (Acc. No 213) and Dottato (Acc. No 235) was also reported previously (Papadopoulou et al. 2001). Based on the dendrogram constructed the established fig population has weak genetic relationships, being organized in two main genetic groups, with the major group (group A) divided further into four subgroups, with the largest of them separated in two clusters (Fig. 2). AMOVA analysis confirmed the weak genetic relationship, as the 69% of genetic variability (differences) existed within groups.

Assuming that the 41 fig landraces from Greece (Suppl. Table 1) comprise a representative sample of the country's fig population, the genetic data set could be used to compare them with other *ex situ* fig collections from other countries. The genetic parameters, which quantified variability, were estimated using seven microsatellites (MFC6 not included). Concerning the population genetics statistics, our results agree with data published for other fig populations and SSR combinations (Khadari et al. 2003; Giraldo et al. 2008; Achtak et al. 2009; Aradhya et al. 2010; Caliskan et al. 2012; Perez-Jimenez et al. 2012; Boudchicha et al. 2018; Rodolfi et al. 2018). However, Ganopoulos et al. (2015), using 90 fig accessions from Greece and other countries and seven microsatellites, observed higher values of Na (13) and He (0.747). Such difference in genetic results could be due to SSR selection, origin of the fig germplasm, number of accessions studied and technical variation. Klekowski (1988) stated that clonally propagated perennial species are known to carry a relatively high genetic load and tend to exhibit high heterozygosity as a mechanism to overcome the deleterious effects of recessive mutations. Moreover, the weak genetic differences of fig populations, as shown in the present study, has been reported by other authors worldwide (Saddoud et al. 2007; Aradhya et al. 2010; Ganopoulos et al. 2015; Boudchicha et al. 2018). Furthermore, Aradhya et al. (2010) commented that the weak genetic differences of fig are probably due to the fact that genetic variability exists across different fig types due to mutation-recombination process existing in fig. In the present study, negative values for the mean fixation index (-0.184) supported the presence of high heterozygosity and outbreeding, confirming the above-mentioned observations.

Based on the constructed dendrogram (Fig. 2) no correlation between the genetic make-up of figs of the established population and their origin was observed. This was also supported by PCoA clustering, where varieties from other countries were grouped together with Greek ones. Group II included only two fig accessions, both originating from Greece. The inability of this dataset to group figs based on their origin was confirmed by PCoA analysis as well as UPGMA. Further analysis and characterization need to be performed for two apparently identical Greek fig accessions (Zakynthos [Acc. No 160] and Maurosykia [Acc. No 108]). Fig accessions with different geographical origin clustered together and fig accessions with the same geographical origin did not cluster. This is evident when considering that, of the seven fig accessions originating from the island of Syros, only four were clustered to group AI. It becomes clear that the geographic origin could not be the main criterion for fig classification; this was also concluded by other studies (Giraldo et al. 2008; Aradhya et al. 2010; Boudchicha et al. 2018). Only Ikegami et al. (2009), using a population of eight Japanese and 11 foreign accessions, proposed that fig accessions from Japan have no genetic relationship to figs from other countries. Many authors attribute the absence of grouping of fig trees according to their origin, to the vegetative propagation of fig, which allows an easy exchange of plant material among different geographical regions (Ciraldo et al. 2008; Aradhya et al. 2010).

The only caprifig (Acc. No 124) included in this study, was clustered together with fig Tzavellas black large (Acc. No 132), that produces edible fruits. According to Machado et al. (2001), this could be explained by the fact that caprifigs did not significantly diverge from the common fig groups, and agrees with the monoecious origin of *Ficus* that evolved into two gynodioecious forms.

The development of a reliable tool for cultivar identification is essential to ensure plant material identity for registration, cultivar protection, and germplasm management. In the present study, we proposed an identification key scheme using six microsatellites that could discriminate 58 out of 60 genotypes of fig accessions in the established *ex situ* collection (98% discrimination). A study on morphological characteristics of the fig accessions indicated that the accessions non-differentiated by molecular analysis could be differentiated in fruit and fruit flesh color (Ntanos et al. 2015). The latter suggests that a complementary approach including molecular and morphological traits may be essential to ensure plant material identity as suggested by others (i.e. Oukabli and Khadari 2005).

Conclusion

Sixty fig accessions (41 indigenous Greek and 19 from other Mediterranean countries) established an *ex situ* collection at the Agricultural University of Athens and its diversity was analyzed using eight SSR markers. Allelic analysis of the Greek fig genotypes showed relatively low allelic variation, an excess of heterozygosity, and extensive outbreeding. Cluster analysis showed that the established fig population exhibited a weak genetic structure. Both cluster and principal coordinate analysis confirmed that there is no correlation between genetic makeup and geographical origin of the fig accessions. All the above made a step towards the characterization of the germplasm collection. An identification key scheme for fig cultivars that will be useful in cultivar discrimination and intellectual property protection was developed.

Declarations

Acknowledgements

The authors would like to thank Prof. Mike Roose for critically reviewing of the manuscript and Dr. Alessandra Stella for insights in Structure results.

Ethics declarations

Conflict of interest

The authors declare no conflicts of interest.

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Achtak H, Oukabli A, Ater M, Santoni S, Khadari B (2009) Microsatellite markers as reliable tools for fig cultivar identification. J Am Soc Hortic Sci 134:624–631
2. Amos WJ, Hoffman I, Frodsham A, Zhang L, Best Sand Hill AVS (2007) Automated binning of microsatellite alleles: problems and solutions. Mol Ecol Notes 7:10–14. https://doi.org/10.1111/j.1471-8286.2006.01560.x
3. Aradhya MK, Stover ED, Velasco D, Koehmstedt A (2010) Genetic structure and differentiation in cultivated fig (*Ficus carica* L.). Genetica 138:681–694. https://doi: 10.1007/s10709-010-9442-3
4. Boudchicha RH, Hormaza JI, Benbouza H (2018) Diversity analysis and genetic relationships among local Algerian fig cultivars (*Ficus carica* l.) using SSR markers. South African Journal of Botany 116:207–215. https://doi.org/10.1016/j.sajb.2018.03.015
5. Caliskan O, Polat AA, Celikkol P, Bakir M (2012) Molecular characterization of autochthonous Turkish fig accessions. Spanish Journal of Agricultural Research 10:130–140. doi: http://dx.doi.org/10.5424/sjar/2012101-094-11
6. Chapuis M-P, Estoup A (2007) Microsatellite null alleles and estimation of population differentiation. Molecular Biology and Evolution 24:621–631. https://doi.org/10.1093 /molbev/msl191
7. Chatti K, Baraket G, Ben Abdelkrim A, Saddoud O, Mars M, Trifi M, Hannachi S, A.C (2010) Development of molecular tools for characterization and genetic diversity analysis in Tunisian fig (*Ficus carica*) cultivars. Biochem Genet 48:789–806. https://doi.org/ 10.1007/s10528-010-9360-1
8. Condit IJ (1955) Fig varieties: a monograph. Hilgardia 23:323–538
9. Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources 4:359–361. doi: 10.1007/s12686-011-9548-7
10. Ewen KR, Bahlo M, Treloar SA, Levinson DF, Mowry B, Barlow JW, Foote SJ (2000) Identification and analysis of error types in high-throughput genotyping. Am J Hum Genet 67:727–736. https://doi.org/10.1086/303048
11. Flaishman M, Radov V, Stover ED (2008) The fig: Botany, Horticulture, and Breeding, in: Jules Janick, (Eds.), Horticultural Reviews Volume 34, John Wiley & Sons Ed., Inc., pp. 113-196
12. Ganopoulos I, Xanthopoulou A, Molassiotis A, Karagiannis E, Mousiadis T, Katsaris P, Aravanopoulos F, Tsaftaris A, Kalivas A, Madesis P (2015) Mediterranean basin Ficus carica L.: from genetic diversity and structure to authentication of a protected designation of origin cultivar using microsatellite markers. Trees 29:1959–1971. https://doi.org/10.1007 /s00468-015-1276-2
13. Ghosh S, Karanjawala ZE, Hauser ER, Ally D, Knapp JI, Rayaman JB, Musick A, Tannenbaum J, Te C, Shapiro S, Eldridge W, Musick T, Martin C, Smith JR, Carpten JD, Brownstein MJ, Powell JI, Whiten R, Chines P, Nylund SJ, Magnuson VL, Boehnke M, Collins FS (1997) Methods for precise sizing, automated binning of alleles, and reduction of error rates in large-scale genotyping using fluorescently labeled dinucleotide markers. FUSION (Finland–US Investigation of NIDDM Genetics) Study Group. Genome Res 7:165–178. https://doi.org/10.1101/gr.7.2.165
14. Giraldo E, Viruel MA, Lopez-Corrales M, Hormaza JI (2005) Characterisation and cross-species transferability of microsatellites in the common fig tree (*Ficus carica* L.). Journal of Horticultural Science and Biotechnology 80:217–224. https://doi.org/10.1080/14620316.2005.11511920
15. Gupta PK, Varshney RK (2000) The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. Euphytica 113:163–185. https://doi.org/10.1023/A:1003910819967
16. Hoffman JI, Amos W (2005) Microsatellite genotyping errors: detection approaches, common sources and consequences for paternal exclusion. Mol Ecol 14:599–612. https://doi.org/10.1111/j.1365-294X.2004.02419.x
17. Ikegami H, Nogata H, Hirashima K, Awamura M, Nakahara T (2009) Analysis of genetic diversity among European and Asian Fig Varieties (*Ficus carica L.*) using ISSR, RAPD and SSR markers. Genetic Resources and Crop Evolution. 56:201–209. https://doi.org/10.1007/s10722-008-9355-5
18. Khadari B, Hochu I, Bouzid L, Roger JP, Kjellberg F (2003) The use of microsatellite markers for identification and genetic diversity evaluation of the fig collection in CBNMP. Acta Hort 605:77–86. https://doi.org/10.17660/ActaHortic.2003.605.10
19. Khadari B, Hochu I, Santoni S, Kjellberg F (2001) Identification and characterization of microsatellite loci in the common fig (*Ficus carica* L.) and representative species of the genus *Ficus*. Mol Ecol Notes 1:191–193. https://doi.org/10.1046/j.1471-8278.2001.00072.x
20. Kislev EM, Hartmann A, Bar-Yosef O (2006) Early domesticated Fig in the Jordan Valley. Science 312:1372–1374. https://doi.org/10.1126/science.1125910
21. Klekowski EJ (1988) Genetic load and its causes in long-lived plants. Trees 2:195–203
22. Machado CA, Jousselin E, Kjellberg F, Compton SG, Herre EA (2001) Phylogenetic relationships, historical biogeography and character evolution of fig pollinating wasps. Proceedings of the Royal Society B: Biological Sciences 268, 7–10. doi: 10.1098/rspb.2000.1418
23. Mondini L, Noorani A, Pagnotta MA (2009) Assessing plant genetic diversity by molecular tool. Diversity 1:19–35. https://doi.org/10.3390/d1010019
24. Murray MG, Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res 8:4321–4325. https://doi.org/10.1093/nar/8.19.4321

25. Nagy S, Poczai P, Cernak I, Gorji AM, Hegedus G, Taller J (2012) PICcalc: An online program to calculate polymorphic information content for molecular genetic studies. Biochem Genet 50:670–672. https://doi.org/10.1007/s10528-012-9509-1
26. Nei M, Li W-H (1979) Genetics Mathematical model for studying genetic variation in terms of restriction endonucleases (molecular evolution/mitochondrial DNA/nucleotide diversity). Proceedings of the National Academy of Sciences USA 76, 5269-5273. https://doi.org/10.1073/pnas.76.10.5269
27. Ntanos EK, Roussos PA, Sklavounos A, Voloudakis A, Assimakoroulou A (2015) Phenotypic characteristics of various fig genotypes. 27th Conference of the Greek Society for Horticultural Science, Volos
28. Oukabli A, Khadari B (2005) Characterization des varieties polyclonales marocaines de figuiers, *Ficus carica* L. Fruits 60:47–54. https://doi.org/10.1051/fruits:2005012
29. Papadopoulou K, Ehaliotis C, Tourna M, Kastanis P, Karydis I, Zervakis G (2002) Genetic relatedness among dioecious *Ficus carica* L. cultivars by random amplified polymorphic DNA analysis, and evaluation of agronomic and morphological characters. Genetica 114:183–194. https://doi.org/10.1023/A:1015126319534
30. Pavlicek A, Hrda S, Flegr J (1999) Free-Tree freeware program for construction of phylogenetic trees on the basis of distance data and bootstrap/jackknife analysis of the tree robustness. Application in the RAPD analysis of genus Frenkelia. Folia Biol (Praha) 45:97–99
31. Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics 28:2537–2539. https://doi.org/10.1111/j.1471-8286.2005.01155.x
32. Perez-Jiminez M, Lopez B, Dorado G, Pujadas-Salva A, Guzman G, Hernandez P (2012) Analysis of genetic diversity of southern Spain fig tree (*Ficus carica* L.) and reference materials as a tool for breeding and conservation. Hereditas 149:108–113. https://doi.org/10.1111/j.1601-5223.2012.02154.x
33. Pompanon F, Bonin A, Bellemain E, Taberlet P (2005) Genotyping errors: causes, consequences and solutions. Nat Rev Genet 6:847–859. https://doi.org/10.1038/nrg1707
34. Rodolfi M, Ganino T, Chiancone B, Petrucelli R (2018) Identification and characterization of Italian common figs (*Ficus carica*) using nuclear microsatellite markers. Genet Resour Crop Evol 65:1337–1348. https://doi.org/10.1007/s10722-018-0617-6
35. Rousset F (2008) Genepop'007: a complete re-implementation of the genepop software for Windows and Linux. Molecular Ecology Resources 8:103–106. https://doi.org/10.1111/j.1471-8286.2007.01931.x
36. Roy S, Tyagi A, Shukla V, Kumar A, Singh UM, Chaudhary LB, Datt B, Bag SK, Singh PK, Nair NK, Husain NT, Tuli R (2010) Universal plant DNA barcode loci may not work in complex groups: A Case Study with Indian Berberis species. PLoS ONE 5:e13674. https://doi.org/10.1371/journal.pone.0013674
37. Saddoud O, Chatti K, Salhi-Hannachi A, Mars M, Rhouma A, Marrakchim M, Trifi M (2007) Genetic diversity of Tunisian figs (*Ficus carica* L.) as revealed by nuclear microsatellites. Hereditas 144:149–157. https://doi.org/10.1111/j.2007.0018-0661.01967.x
38. Stover E, Aradhya M, Ferguson L, Crisosto CH (2007) The Fig: Overview of an ancient fruit. HortScience 42:1083–1087
39. Tessier C, David J, This P, Boursiquot JM, Charrier A (1999) Optimization of the choice of molecular markers for varietal identification in *Vitis vinifera* L. Theor Appl Genet 98:171–177. https://doi.org/10.1007/s001220051054
40. UPOV/INF/17/1 (2010) Guidelines for DNA-Profiling: Molecular Marker Selection and database construction ("BMT Guidelines")
41. Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) Micro-checker: software for identifying and correcting genotyping errors in microsatellite data. Mol Ecol Notes 4:535–538. https://doi.org/10.1111/j.1471-8286.2004.00684.x
42. Vavilov NI (1951) The origin, variation, immunity and breeding of cultivated plants. The Ronald Press Company, NY
43. Vinson JA, Zubik L, Bose P, Samman N, Proch J (2005) Dried Fruits: Excellent in vitro and in vivo Antioxidants. J Am Coll Nutr 24:44–50. https://doi.org/10.1080/07315724.2005.10719442

Tables

Table 1. Genetic parameters per SSR marker for the fig accessions originating from Greece.

Microsatellite	N	Allele range	Allele size	Genotypes	Na	Ne	Ho	He	F	P-val
MFC1	40	170-206	186, 192, 202*	8	3	1.807	0.350	0.447	0.216	0.104 (ns)
MFC2	41	158-194	158 , 164, 174, 186, 194	7	5	2.524	0.780	0.604	-0.293	0.995 (ns)
MFC3	41	126-154	126, 128, 130, 132, 138, 154	9	6	2.282	0.537	0.562	0.045	0.093 (ns)
MFC4	41	204-230	206, **230**	2	2	1.981	0.902	0.495	-0.822	1.000 (ns)
MFC5	41	132-142	134, **140**	2	2	1.303	0.268	0.232	-0.155	1.000 (ns)
MFC7	41	148-172	152, 160, 168	6	3	2.576	0.707	0.612	-0.156	0.373 (ns)
MFC8	41	175-207	**177**, 205	2	2	1.243	0.220	0.195	-0.123	1.000 (ns)
Average					**3.3**	1.959	**0.537**	**0.449**	**-0.184**	
MFC6**	34	284-336	284, 290, 304, 316, 328	8 **	5	2.645	0.676	0.622	-0.088	0.499 (ns)

N: Number of fig accessions

Na: Alleles per locus

Ne: Effective alleles per locus

Ho: Observed heterozygosity

He: Expected heterozygosity

ns = not significant

Bold allele size indicates monomorphic allele,

Underlined allele size indicates rare allele

*In reference fig accession from foreign countries, two more different alleles were revealed, both in MFC1, 198 bp and 170 bp.

** For MFC6, seven out of 41 fig accessions, originating from Greece, did not reveal any banding pattern. Genetic parameter estimated only for the fig accessions that revealed banding pattern.

Table 2
Construction of an Identification (Id) key for fig population and relative parameters.

									No of genotypes differentiated	Accessions Not differentiat	
SSR Locus	**No Gen.***	**PIC****	**Cj**	**Dj**	**C_L**	**D_L**	**Steps**	**Locus combination**		Exp.	Ol
MFC3	9	0.546	0.21	**0,79**	0.22	0.78	1	MFC3	9	371.7	36
MFC1	8	0.435	0.29	**0,71**	0.31	0.69	2	MFC3+MFC1	16	107.8	10
MFC2	7	0.511	0.37	**0,63**	0.38	0.62	3	MFC3+MFC1+MFC2	39	39.9	40
MFC7	6	0.525	0.4	**0,6**	0.41	0.59	4	MFC3+MFC1+MFC2+MFC7	49	15.9	22
MFC8	2	0.214	0.59	**0,41**	0.59	0.41	5	MFC3+MFC1+MFC2+MFC7+MFC8	53	9.4	13
MFC5	2	0.185	0.64	**0,36**	0.64	0.36	6	MFC3+MFC1+MFC2+MFC7+MFC8+MFC5	54	6.0	12
MFC4†	2	0.368	0.74	**0,26**	0.74	0.26	7	MFC3+MFC1+MFC2+MFC7+MFC8+MFC5+MFC4	54	4.5	12
MFC6	8	0.584	0.15	**0,85**	0.17	0.83	8	MFC3+MFC1+MFC2+MFC7+MFC8+MFC5+MFC4+MFC6	58	0.7	2

where,

Cj: confusion probability,

C_L: confusion probability when sample is very large ($N \to \infty$)

Dj: discriminating power,

D_L: discriminating power when sample is very large ($N \to \infty$).

Mathematic relationship is C*j* + D*j* =1.

Exp.: Expected

Obs.: Observed

* No Gen: Number of genotypes per locus in all 60 fig accessions.

** PIC: polymorphism information content based on allele frequencies per SSR locus

† MFC4 is not included in the Identification key since it does not differentiate additional pairs.

Table 3 is not available with this version.

Figures

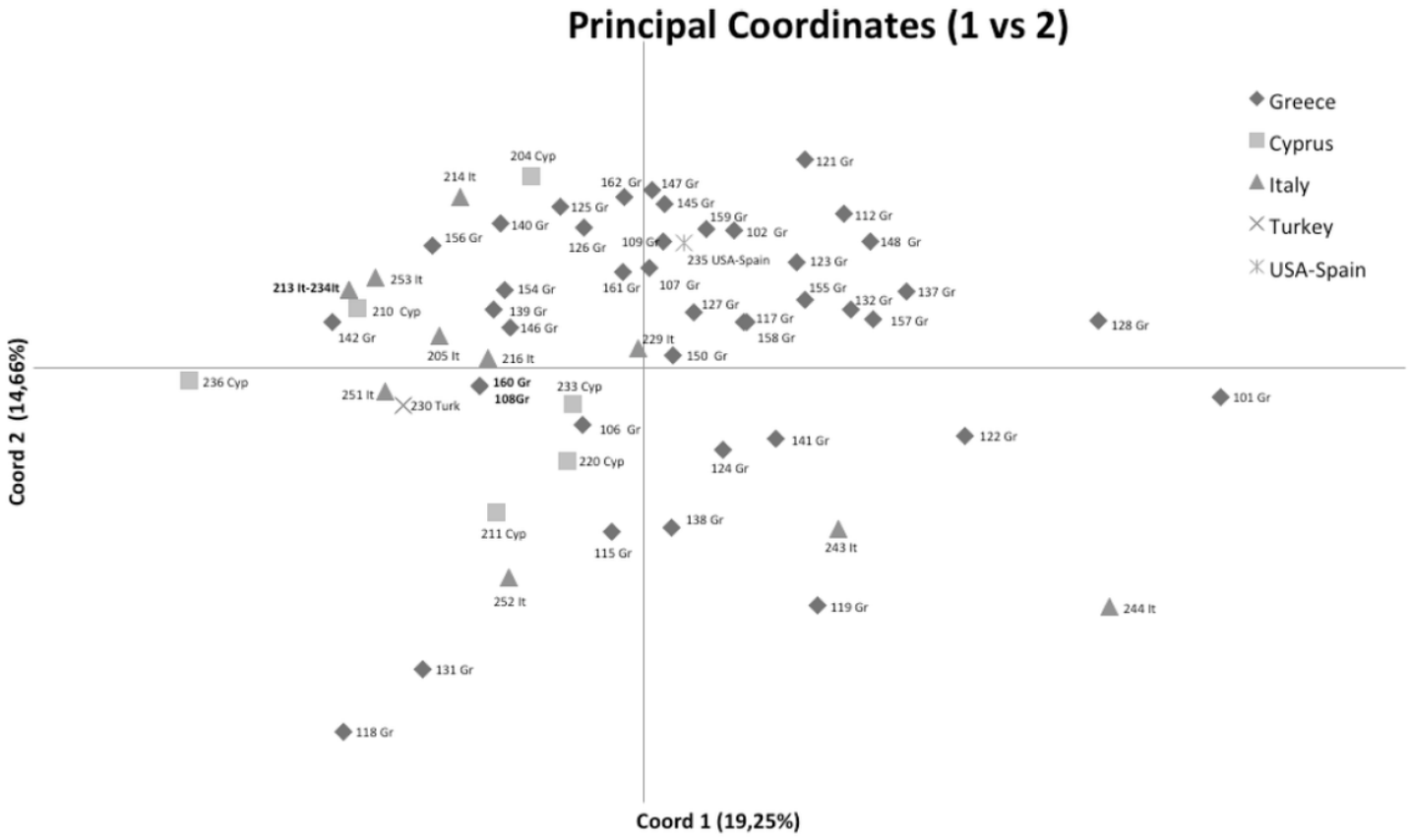

Figure 1

Principal coordinate analysis (PCoA) of the 60 fig accessions based on eight SSR markers. Numbers are shown in Table 1.

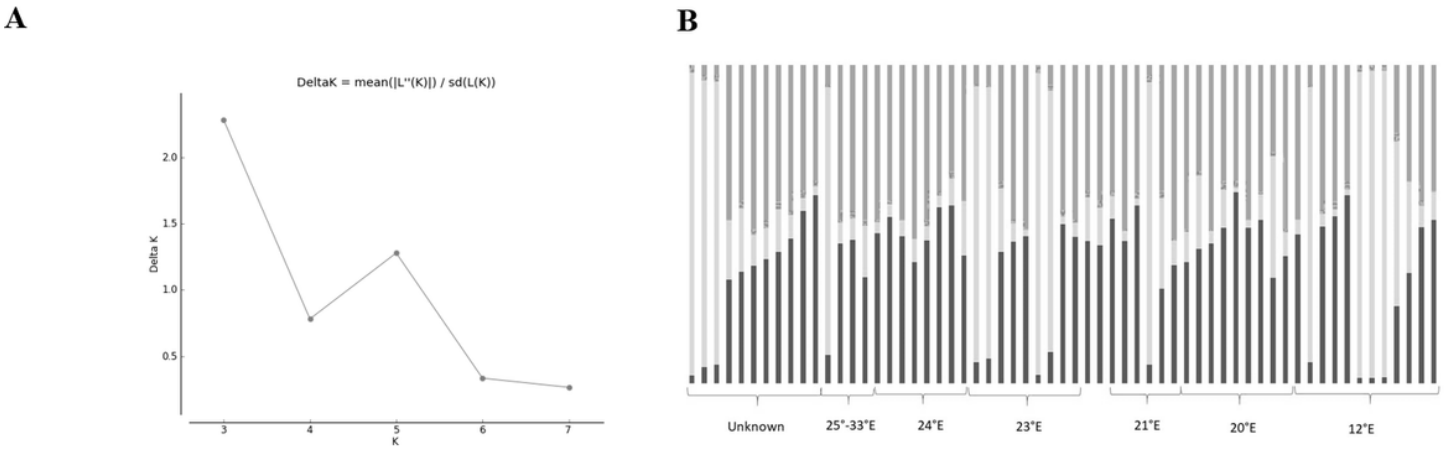

Figure 2

Prediction of sub-populations in a collection of 60 fig genotypes via STRUCTURE analysis. (**A)** Estimation of the number of populations K=3 by calculating delta K values, (**B**) STRUCTURE plot depicting sub-populations. Inferred population structure for K=3. Each individual is represented by bars partitioned into K segments representing the membership fraction in K clusters. The bars are arranged based on geographical eastern approximate coordinate of the accessions (right: most western accession, Italy; left: most eastern accession, Cyprus).

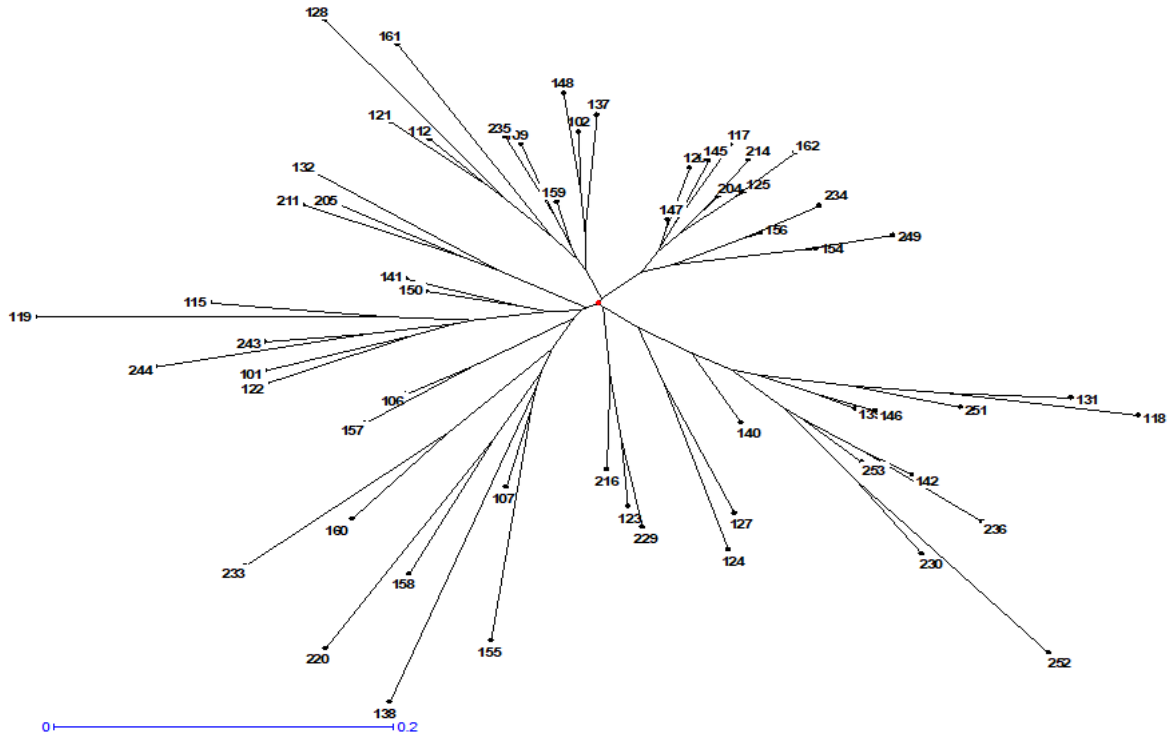

Figure 3

Dendrogram showing genetic relationship of 60 Fig accessions based on SSR data. The dendrogram constructed in DARwin v6.0.15 using the WNJ method (confirmed by 10000 bootstraps). The scale indicates the edge length.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Sup.Fig.1.docx
- Sup.Fig.2hiRes2.png
- Sup.Table1.docx