COMPARISON BETWEEN DNA-BASED, POMOLOGICAL AND CHEMICAL MARKERS ACCOMPLISHED BY BIOINFORMATIC TOOLS TO DISTINGUISH WITHIN TUNISIAN OLIVE CULTIVARS

R. Ben Ayed1*, K. Ennouri1, H. Ben Hassen2, M. A. Triki3 and A. Rebai1

1 Centre of Biotechnology of Sfax, PB ‘1177’, 3018 Sfax, Tunisia
2 Laboratory of Physics Mathematics and Applications, Faculty of sciences of Sfax, Tunisia.
3 Laboratoire Ressources et Amélioration Génétiques de l’Olivier, du Pistachier et de l’Amandier, Institut de l’Olivier, Tunisia

Received: 22 May 2015 / Accepted: 26 August 2015 / Published online: 1 September 2015

ABSTRACT
The genetic diversity of 16 Tunisian olive cultivars (Olea europaea L.) of known origin sampled from different areas of the country was assessed using genetic markers (6 SSR and 5 SNP markers). Three dendrograms based on cultivar genotypes generated by SSR, SNP and both SSR and SNP markers revealed three clusters which were consistent with the varieties classification according to phenotypic characteristics, but not correlated with the geographic origin. Also, we compared the results obtained with the genetic markers to those obtained with agro-morphological and chemicals data using bioinformatic analyses. This work provides better understanding of the diversity available in Tunisia olive cultivars and supplies an important contribution for olive breeding and olive oil authenticity.

Keywords: Agro-morphological traits; Genetic diversity; PCA; Olive; SNP; SSR; Bioinformatic tools.

Author Correspondence, e-mail: raydabenayed@yahoo.fr
doi: http://dx.doi.org/10.4314/jfas.v7i3.8
1. INTRODUCTION

Olive (*Olea europaea* L.) is the most important oil-producing crop in the Mediterranean region and particularly in Tunisia. Olive tree has allowed a huge patrimony, estimated in more than 1200 cultivars [1]. The width of this germplasm, with a large number of varieties present in the main olive oil producing countries cases some problems for the management of plant material and for its traceability and authenticity in olive oil production. An accurate identification of both the raw material and the derived products is essential for the traceability of extra virgin olive oil and consequently for better marketing. Habitually, the authenticity of virgin olive oil is conventionally assessed by monitoring several components such as, fatty acids, phenols, sterols, tocopherols and volatile compounds. However, the compositional analyses have several limits because chemical composition of virgin olive oil is influenced by environmental and genetic factors. For these reasons, currently the uses of DNA-based markers, which are independent from environmental conditions, have been successfully applied to overcome this problem. Several molecular marker techniques such as random amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), single nucleotide polymorphism (SNP) and simple sequence repeat (SSR), which are independent from environmental conditions, have been successfully applied and are very useful to distinguish within olive cultivars and to study traceability of olive oil [6]. Because of its higher polymorphism and discerning power, SSR markers, also known as microsatellite markers have shown high potential for resolving issues of synonymies, homonymies, and misnaminings. All these characteristics make them ideal markers for applications in analysis of intracultivar variability issues [16], linkage mapping [29], and for characterizing olive germ plasm resources [3; 20; 21; 12]. In addition, several other papers used SSR technologies to study the genetic diversities of olive cultivars [26; 23; 19]. Microsatellites are also very useful markers for paternity analysis [17; 10; 23]. Recently microsatellites have become available and reliable molecular markers for traceability issues to define olive oil origin and to detect the presence of prohibited cultivars [28; 11; 18; 4; 5]. In recent years, a novel class of markers, namely SNP, has emerged as an important tool in genomics and is increasingly being used as molecular markers in various laboratories for diverse applications. The development of this
type of markers needs a high level of genome sequence information. Therefore, only a few SNP markers have been reported in olive since only very few sequence data were available before the year 2010 [22; 24; 7; 9].

This paper reports the comparison between compositional and molecular markers (particular SSR and SNP markers) previously studied and analyzed by our research group [4; 5; 6; 7; 8] to investigate genetic diversity and relationships among 16 olive trees from different Tunisian regions. Bioinformatic tools are used to compare the results generated by the different markers. The present work is an assessment of the efficiency of accomplishment between markers in the evaluation of the genetic diversity and relationship among the studied olive trees. To the best of our knowledge, this is the first agro-morphological, chemical and molecular-based study on olive trees.

2. RESULTS AND DISCUSSION

2.1. Comparison Between marker systems: molecular markers and agro-morphological and chemicals characteristics

Matrix of Euclidean distances: Euclidean distances were calculated between all pairs of individual trees based on standardized values of ten quantitative traits and the matrix distance was used for cluster analysis (UPGMA) in SPSS version 13, as shown on figure 1.

![Cluster Analysis Diagram](image)

Fig.1. Classification of olive cultivars studied based on agro-morphological and chemical parameters using clustering by Euclidean distance
Based on the graphical representation of the matrix of Euclidean distances, which lies between the studied varieties based on agro-morphological and chemical characteristics [14], three groups were identified. Group 1 includes four varieties (Chetoui Thibar, Chetoui Siliana, Chetoui Nabeul and Chemlali Tataouine). The first three correspond to the same variety Chetoui, but grown in three different regions.

We firstly noticed that there is a similarity between this group and both group 3 of the dendrogram based on the eight SSR markers (figure 2a) and of the dendrograms based on SSRs combined to SNPs markers (figure 2c) and secondly, with group 2 of the dendrogram found on the five SNPs (figure 2b).

However we noted some differences with the dendrograms generated by molecular markers (SSR, SNP, SNP + SSR), were it was remarked an absence of the variety “Rkhaymi” in the dendrogram based on the agro-morphological and chemical characteristics, and the presence of the variety “Chemlali Tataouine” in group 1. In addition, we observed that the varieties ‘Chemcheli’ and ‘Oueslati’ were common in group 2 of the dendrogram based on the agro-morphological and chemical characteristics (figure 1), and in group 2 of the dendrograms generated by molecular markers SSR and (SNP + SSR). Group 3 of the dendrogram based on the agro-morphological and chemical characteristics (figure 1) showed
a high similarity with the group 1 of the dendrograms generated by molecular markers SSR, SNP and (SNP + SSR) (figure 2). In addition, Table 1 showed that GAPU71B marker had the highest value of PIC (0.790). This result demonstrated that GAPU71B is the most informative SSR marker and it is able to distinguish all the varieties studied. This result is in line with those demonstrated previously by Ben Ayed et al. [5].

Table 1. Characteristics of microsatellites markers used for DNA amplification in the present study and summary of genetic diversity observed by typing 16 samples using 8 simple sequence repeat markers

<table>
<thead>
<tr>
<th>Code</th>
<th>Primer sequence (5’→3’)</th>
<th>Repeat motif</th>
<th>Tm°C</th>
<th>Size of amplicon in bps</th>
<th>Na</th>
<th>PIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>DCA1</td>
<td>F: (FAM) CTCTGAAAATCTACACTCACATCC R: ATGACACAGAAGAAGTGAACATTC</td>
<td>(GA)22</td>
<td>50</td>
<td>222-244</td>
<td>5.0</td>
<td>0.716</td>
</tr>
<tr>
<td></td>
<td>DCA3 F: (FAM) CCAAGCGGAGGTATATGGTTTGC R: TGGCTTTTGCTGTTTGGAGATGGT</td>
<td>(GA)19</td>
<td>50</td>
<td>247-269</td>
<td>3.0</td>
<td>0.615</td>
</tr>
<tr>
<td></td>
<td>DCA4 F: (FAM) TTAACCTTGTGCTTCTCCATATCC R: AGTGACAAAAAGCAAAAGACTAAAAGC</td>
<td>(GA)16</td>
<td>57</td>
<td>208-222</td>
<td>4.0</td>
<td>0.398</td>
</tr>
<tr>
<td></td>
<td>GAPU59 F: (FAM) CCCCCGTCTTGGCTTGGCTAA R: CAAAAGTGCACCTTCTCTGC</td>
<td>(CT)10</td>
<td>50</td>
<td>210-228</td>
<td>4.0</td>
<td>0.646</td>
</tr>
<tr>
<td></td>
<td>GAPU71A F: (FAM) GATCATTAAAAATATTAGAGAGAGAGA R: TCCATCCATGCTGAACCTT</td>
<td>(AG)6</td>
<td>57</td>
<td>117-140</td>
<td>4.0</td>
<td>0.677</td>
</tr>
<tr>
<td></td>
<td>GAPU71B F: (FAM) GATCAAAGGAGAGAGAGAGAGAGA R: ACAACAAAATCCTACGCTTG</td>
<td>(AG)2(AAG)6</td>
<td>50</td>
<td>155-166</td>
<td>6.0</td>
<td>0.790</td>
</tr>
<tr>
<td>UDO09</td>
<td>F: (FAM) TTGATTTCATGCTGACCA R: CATAGGGAAGTGCAGG</td>
<td>(AG)16</td>
<td>57</td>
<td>97-115</td>
<td>3.0</td>
<td>0.525</td>
</tr>
<tr>
<td>UDO12</td>
<td>F: (FAM) TCACCATCCTTAAACTTCACACA R: TCAAGCAATTCCACCGTATG</td>
<td>(GT)10</td>
<td>50</td>
<td>149-179</td>
<td>3.0</td>
<td>0.619</td>
</tr>
<tr>
<td>Totals</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>31</td>
<td>4.986</td>
</tr>
<tr>
<td>Media</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>3.875</td>
<td>0.623</td>
</tr>
</tbody>
</table>

*a*: annealing temperature for PCR amplification

*a*: For each locus, the size range in base pairs, the number of alleles (Na) and the polymorphism content information (PIC)

Multivariate analysis using the PCA method: the diversity analysis was performed by the
technique of principal component analysis (PCA) using SPSS program. We selected the 16 Tunisian olive varieties and 10 essentially quantitative parameters of agro-morphological and chemical data (figure 3).

The first two components of the PCA explained 60.9% of the total variation among a subset of 16 Tunisian olive varieties included in the analysis. The first component (axis1) explained 44.7% of the variation, followed by 16.2% for the second components (axis 2). Axis 1 (44.7%) is positively correlated with cholesterol and polyphenol content and negatively correlated with carotene content and the rate of linolenic acid (C18:3). Axis 2 (16.2%) was positively correlated with the acidity of the oil oxidative stability and the content of β-sitosterol and negatively correlated to the rate of unsaponifiable matter. When the olive varieties were studied individually, we showed that primary Axis 1 divides the varieties “Chmlali” (Chmlali Nabeul, Monastir and Chmlali Chemlali Sousse) on the positive side and the varieties “Chmlali Sfax” (Chmlali Chaal, Chmlali Blettech and Chmlali Dokhane) Sidi Bouzid and Chmlali on the negative side. Indeed, the varieties “Chmlali North” (Chmlali Nabeul) and the varieties “Chmlali Center” (Chmlali Monastir and Sousse Chmlali) show a high similarity on physicochemical and organoleptic parameters, when, all “Chmlali”
varieties localized in Sfax region have the same characteristics. Thus, axis 1 separates the varieties' Oueslati', 'Chemcheli' and 'Zarrazi', on the positive side and the varieties' Zalmati', 'Chemlali Tataouine' and 'Rkhaymi' the negative side. Indeed, the varieties 'Oueslati', 'Chemcheli' and 'Zarrazi' have common characteristics especially on the morphology and size of the fruits and their uses as dual purpose (oil and table). For the varieties 'Zalmati' and 'Chemlali Tataouine' (grown in the south of Tunisia and precisely Tataouine) their classification, by the PCA, in the same group along the axis 1 is essentially related to a remarkable similarity of their fruit size parameters, cholesterol and the polyphenol content.

We note from figure 3 that the varieties 'Chetoui' (Chetoui Nabeul, Chetoui Siliana and Chetoui Thibar) are at axis 1. Axis 2 divides the varieties 'Chemlali' (except the variety Chemlali-Tataouine) on one side and the other varieties on the other side (figure 3). This repartition was so consolidated by some specific proprieties of ‘Chemlali’ varieties, as quantitative parameters, also by high acidity and high levels of unsaponifiable matter (associated with high palmitic acid content).

We can conclude from the above results that there is a positive correlation between the agro-morphological and chemical olive oils characteristics using sophisticated bioinformatics analysis and results revealed by SSR, SNP and (SNP + SSR) genetic molecular markers. Our results are similar to those found by Grati Kamoun et al. [14] in their studies of the genetic diversity of the Tunisian olive oils using AFLP markers with sizes of fruit (morphological character). However, Belaj et al. [2], studying the genetic variability of Spain olive cultivars using allozyme and RAPD markers, showed a moderate correlation with the fruit size. Therefore, the difference between the morphological markers of fruit and oil chemical and molecular markers (mainly SSR markers previously studied) is showed. This discordance could be related to small number of genetic markers which were scrutinized, so they do not cover all genomic regions responsible for variability of agro-morphological and chemical characters; additively to the effect of environmental factors.

2.2. Multi-marker analysis to study diversity between Tunisian olive cultivars

To classify the varieties studied on all genetic, morpho-agronomy and chemicals information that we have, we calculated the Gower distance, which is an index of distance to the
combination of quantitative and qualitative variables. This calculation was performed by using R language. Based on the generated matrix we have rebuilt a global dendrogram (figure 4).

![Global dendrogram performed by using R language using all genetic, morpho-agronomy and chemicals information of olive oil Tunisian cultivars](image)

According to the dendrogram, we can identify three groups of varieties as follows: group 1 contains Chem_Blett, Chem_Chàal, Chem_Nab, Chet_Thibar and Rkhaymi. The second group contains Chemch, Chem_Dok, Chem_Sous, Chet_Nab and Zarrazi. The third group includes Chem_Mon, Chem_SB, Chem_Tat, Chet_Silia, Oueslati and Zalmati.

We note here that the Tunisian studied varieties do not have a structure according to geographical area, size and variety of olives. Indeed, in agreement with the results of Grati-Kamoun et al. [14], Rekik et al. [23] and Taamali et al. [27], we could show that the pomological and chemical criteria used for the varietal identification and the determination of the genetic variability of Tunisian olive oil wealth may be insufficient. These chemical and agro-morphological criteria allowed us to make a first classification of different varieties with different characters by multivariate analysis (PCA method). However, the study of Euclidian distance could not allow discriminating cultivars. Unlike dendrograms obtained by genetic
data (SSR and SNP), we found a classification of the studied varieties according to the average weight of fruit. This confirms the hypothesis that the molecular markers are powerful for distinguishing cultivars of the species *Olea europaea* and complementing pomological and chemical analyzes. Although, there was sufficient variability to discriminate all the Tunisian olive tree by means of both morphological, agronomical and chemicals data, the dendrogram based on SSR+SNP markers did identify clearer structure of tree grouping, suggesting that DNA markers are more informative in depicting genetic relationships. Each marker system measures different aspects of this genetic variability and this may explain the lack of consistency in genetic diversity and relationship studies. However, in spite of the powerfulness of SSR markers to detect genetic variability and genetic relationships, they should not be seen as a substitute of the traditional agro-morphological descriptors. All these marker systems should be considered as complimentary tools to provide a more complete understanding of the diversity available in Tunisia olive cultivars and the way in which it can be best used for olive breeding and olive oil traceability and authenticity.

3. EXPERIMENTAL

3.1. Plant Material

A total of sixteen Tunisian olive tree cultivars were chosen and used. They were selected from different geographical regions of Tunisia. For each cultivar, 2 trees were used, and from each tree, DNA was extracted from young leaves [5].

3.2. DNA Extraction

DNA was separately extracted from leaves using the CTAB methods described by Rekik *et al.* [23] and an additional purification was introduced, consisting in washing and eluting once with the QIAamp DNA stool (Qiagen) to eliminate contaminant molecules and generate a high quality DNA for specific, reproducible and consistent amplifications [4].

DNA was quantified by Hoechst H33258 dye incorporation detected by spectrofluorometer (Tecan GENIOS Plus) and dilution series of lamda DNA (D150A Promega) were used as calibration standard. Genomic DNA was undiluted in TE buffer (10 mM Tris-HCl pH 8,1 mM EDTA pH 8) at 20 °C.
3.3. Microsatellite markers

Six microsatellites (SSRs) markers were used in this study. Two markers (DCA1, DCA3), three markers (GAPU59, GAPU71A, GAPU71B) and one marker (UDO12) were selected for their high polymorphism in many olive cultivars and mainly in Tunisian cultivars [4; 5]. PCRs and capillary sequencer PCRs were performed in a 15 µL volume consisting of 10 ng of olive oil genomic DNA or 20 ng of young leaves genomic DNA, 2 mM MgCl2, 0.05 mM of each dNTP, 0.1 µM of forward primer (forward primer was labeled with FAM fluorescent dye), 0.4 µM of reverse primer, 0.5 unit of Go Taq (Go Taq Flexi DNA polymerase, Promega), 1*buffer Go Taq; performing PCR amplifications on a thermal cycler verity (Applied Biosystems, 96 well) for: DCA1, DCA3, GAPU59, GAPU71B primers at: 95 °C for 5 min for 1 cycle, 95 °C for 30 s, 50 °C for 45 s, 72 °C for 45 s for 35 cycles, then 72 °C for 10 min and for GAPU71A and UDO12 primers at: 95 °C for 5 min for 1 cycle, 95 °C for 30 s, 57 °C for 45 s, 72 °C for 45 s for 35 cycles, then 72 °C for 10 min. 5 µL of PCR products was mixed with 0.3 µL of (marqueur 420pb) and 14.7 µL deionised H2O, centrifuged at 2000 rpm for 1 min, denaturated at 94 °C for 3 min, cooled in ice and analysed on a (3130XL Genetic analyser of Applied Biosystems) capillary sequencer.

3.4. Genotyping SNP markers

SNP SOD (insertion/deletion type) was genotyped by a simple polymerase chain reaction and agarose gel electrophoresis. The other four SNPs (FAD2.1, FAD2.3, ANTHO3 and CALC) were genotyped by a polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) method [7]. The PCR product (171 bp) of the SNP (ANTHO3) was digested by MspI restriction enzyme (Fermentas, LIFE SCIENCES) at 37°C overnight. This restriction enzyme recognizes the sequence AA/GG. The G-allele carrying PCR product is cleaved once by the enzyme generating two fragments (64 and 107 bp). The PCR product (476 bp) of the SNP (CALC) was digested by BstZI restriction enzyme (Promega) at 50°C overnight. This restriction enzyme recognizes the sequence CC/GG. The C-allele carrying PCR product is cleaved once by the enzyme leading to two fragments (316-160 bp). The two other SNPs (FAD2.1 and FAD2.3) were analyzed using PCR-RFLP. The PCR product (241bp) of the SNPs (FAD2.1) and (240 bp) of the SNP (FAD2.3) were digested by BamHI restriction
enzyme (Fermentas, LIFE SCIENCES) and Alw26I, respectively, at 37°C overnight. The sizes of the restriction fragments of PCR product were 224 and 17 bp, 130 and 110 bp for CC genotype of FAD2.1 SNP and FAD2.3 SNP, respectively. All digestion products were separated by electrophoresis on 3% NuSieve etidium bromide-stained agarose gels and visualized under UV light.

3.5. Data analysis

The alleles detected for each SNP or/and SSR were recorded and a binary data matrix is established with (1) for presence of bands (1) (each allele representing a band) and (0) for absence of bands (each allele representing a band). Similarity matrices were generated using the SIMQUAL sub-program of NTSYS-PC software [25]. Similarity coefficients were used for cluster analysis of varieties using the SAHN sub-program of NTSYS-PC software and dendrograms were inferred using an arithmetic average (UPGMA) clustering algorithm. Genetic diversity among cultivars based on agro-morphological and chemicals data was examined using the Euclidean distance matrix and the principal components analysis (PCA). This analysis was carried out by using the computer program, SPSS (version 13.0). Global dendrogram was performed using R language based on all genetic, morpho-agronomic and chemical informations of olive oil Tunisian cultivars.

4. ACKNOWLEDGEMENTS

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

5. REFERENCES


How to cite this article: