

# Evaluation of Variability in Tunisian *Olea europaea* L. Accessions using Morphological Characters and Computational Approaches

Karim ENNOURI<sup>1,3</sup>, Rayda BEN AYED<sup>1\*</sup>, Sezai ERCISLI<sup>2</sup>,  
Fathi BEN AMAR<sup>3</sup>, Mohamed Ali TRIKI<sup>3</sup>

<sup>1</sup>University of Sfax, Centre of Biotechnology of Sfax, Tunisia; [1karimennouri1@gmail.com](mailto:1karimennouri1@gmail.com); [raydabenayed@yahoo.fr](mailto:raydabenayed@yahoo.fr) (\*corresponding author)

<sup>2</sup>Ataturk University, Faculty of Agriculture, Department of Horticulture, 25240 Erzurum, Turkey; [sercisli@gmail.com](mailto:sercisli@gmail.com)

<sup>3</sup>Institut de l'Olivier, Laboratoire d'Amélioration et Protection des Ressources Génétiques de l'Olivier, Tunisia; [fathibenamar@yahoo.fr](mailto:fathibenamar@yahoo.fr); [trikimali@yahoo.fr](mailto:trikimali@yahoo.fr)

## Abstract

The olive trees (*Olea europaea* L.) have been cultivated for millennia in the Mediterranean basin and its oil has been an important part of human nutrition in the region. In order to distinguish between olive accessions, morphological and biological characters have been widely and commonly used for descriptive purposes and have been used to characterize olive accessions. A comparative study of morphological characters of olive accessions grown in Tunisia was carried out and analyzed using Bayesian Networks (BN) and Principal Components Analysis (PCA). The obtained results showed that averages of fruit and kernel weights were 2.27 grams and 0.41 grams, respectively. Besides, a relatively moderate level of variation (51.22%) being explained by four Principal components. BN revealed that geographical localisation plays a role in the increase of tree habit, size of lenticels and leaf shape. A dendrogram has been carried out in the aim to classify studied olive accessions. We proposed a novel method of analysis based on the three-step scheme, in which first the data set is clustered, then olive tree features are evaluated. The studied accessions can be divided into four main groups by cutting the dendrogram at a similarity value of 0.645. Different relationships are studied and highlighted, and finally the collected features are subjected to a global principal component analysis. Obtained results confirmed that core surface was negatively correlated with geographical location ( $r = -0.52$ ,  $p < 0.05$ ) and maturation period  $r = -0.539$ ,  $p < 0.05$ ). Number of lenticels was positively correlated to lenticels size ( $r = 0.632$ ,  $p < 0.05$ ). Core shape had a negative correlation with fruit shape ( $r = -0.759$ ,  $p < 0.05$ ). On the basis of these findings, this research confirmed that morphological markers are a preliminary tool to characterize olive oil accessions.

**Keywords:** Accession, Bayesian networks, dendrogram, olive, principal analysis component

## Introduction

The cultivated Olive (*Olea europaea* L.) is a long-lived evergreen tree native to the Mediterranean basin (Poljuha *et al.*, 2008); however, it is now found in several countries in Europe, North and South America, and Australia (Therios, 2009). The Mediterranean climate offers ideal growing conditions for olives: a long and hot growing season with a relatively cool winter (Connell, 1994). The leaves of an olive tree are thick, leathery, and oppositely arranged and live for a period from two to three years (Martin, 1994). The leaves are grey-green in colour. Stomata are only located on the lower surfaces of the leaves, and are involved in restricting water loss and protecting the leaves from harmful ultra-violet radiation (Fernandez *et al.*, 1997). The morphology of the olive fruit is interesting. The fruit is a drupe which results from the growth of the ovary with

only one of the ovules developing (Proietti *et al.*, 1999). The size of the olive fruit is variable, even within the same tree, and depends on accession, fruit load, soil fertility, available water and cultural practices (Therios, 2009). Complete fruit development, from anthesis to ripening, lasts between 25-30 weeks, and after the first four weeks the parts of the fruit can be clearly identified. The endocarp consists of the pit and seed. Volatile and phenolic compounds found in the fruit are responsible for the aroma, taste and many of the health benefits associated with olive oil (Mez-Rico *et al.*, 2006). The concentrations of these compounds change depending upon the degree of ripening of the fruit. The quality of pickling olives is greatly affected by the concentrations of organic acids and sugars found in the fruit due to the role of these compounds in the fermentation process (Patumi *et al.*, 1999). It is important to know when maturation has been reached, in order to harvest fruit when these compounds are at optimal levels. The fruit maturation process is characterized by a change in the fruit

color from green to a dark purple or black (Proietti *et al.*, 1999). Each accession has a different intensity of the final colour which indicates ripeness. Olives have great commercial, economic, and social importance in Tunisia, Tunisia is a major producer and exporter of olive oil (Eyres, 2012). Selecting adequate accessions is considered as a crucial requirement because the need to improve efficiency in growing olives and extracting the oil. These accessions have rich genetic diversity, and more synonymy or homonymy were found in Tunisian breeding accessions.

Olive trees can be considered as very complex dynamical systems, showing a remarkable ability of adapting their metabolism to different environment conditions. Many important processes in the olive tree physiology are controlled by dynamic interactions between various endogenous and exogenous parameters. In this context, computational methods such as classification and clustering were successfully used by biologists in order to analyze experimental data. In fact, the most commonly used computational method for analysing olive tree data is clustering. While clustering provides a compact summarisation of the data and might point to functional relationships between clustered varieties, it suffers from the following shortcomings: firstly, clustering is based on a global correlation measure. This obscures relationships that exist over only a subset of the data. Secondly, clustering fails to detect interactions between different varieties from linear correlation. Finally, it is impossible to incorporate additional types of information, such as experimental details.

Therefore, we propose an alternative unified approach for data estimation, based on Bayesian Networks (BNs) that take account of hierarchical structure among covariates. A BN, is a probabilistic graphical model tool for describing relationships in a wide variety of domains (Adusei-Poku *et al.*, 2007), including artificial intelligence, decision theory and data fusion (Pourret *et al.*, 2008). When this is done, the BN fully determines the joint probability distribution over all the variables in the domain. As a result, we can answer various probabilistic queries (probabilistic inference) about the domain by conditioning and marginalizing some sets of variables (Koller and Friedman, 2009). The aim of this study was firstly to evaluate the associations of autochthon olive accessions grown in Tunisia on the basis of a combination of morphologic and agronomic parameters by clustering, and also to analyse accurately the type and origin of existing relationships between the studied parameters by Bayesian Networks (BN) and Principal Component Analysis (PCA) methods.

## Materials and Methods

### Plant material

In this investigation, 30 accessions of *Olea europaea* L., representing the diversity of Tunisian olive germplasm, were collected from different local farms. For each olive tree variety, five of 30-40 year-old olive trees were selected. Trees were watered according to usual agricultural practice and were exposed to natural sunlight. All measurements were made between November and early December. Table 1 presented the *Olea europaea* L. accessions and their studied traits. All samples were collected shortly after sunrise, placed within wet paper layers, wrapped in polyethylene bags and transferred immediately to the laboratory.

### Morphological characterisation and clustering analysis of the data

The description of traits was recorded for 3 successive years (2013-2015), from either the collection fields or farms, following indications provided in literature (Ben Ayed *et al.*, 2016). Quantitative traits (fruit and kernel weights) were transformed as ordinal characters using discretization method to minimise environmental effects. This experiment on 21 traits collected from 30 *Olea europaea* L. accessions was carried out in order to highlight the possible connections between different traits and characteristics of accessions using statistical analysis and classification systems. Each accession to select 5 plants of healthy growth was measured. For each tree, morphological observations were made on 20 leaves and on 20 fruits. Maximum length and width of leaves were measured using mm scale and maximum length and diameter of the fruits were measured using a screw gauge. After fruit characterization, the stone was removed and the dimensions were also measured using a screw gauge and subjected to characterization. The morphological study integrated both quantitative and qualitative variables. Regarding tree characteristics, the height and the circumference of both canopy and the trunk were determined. For leaf; the length and width were determined. For flower, all observations on inflorescence and flower were done on 20 flowers at the middle part of fruit bearing branch. Concerning fruit, measures were carried-out on 20 drupes collected uniformly. The fruit polar length, cross-sectional width, weight and length/width ratio were determined. For stone; the polar length, cross-sectional width, weight, length/width ratio, numbers of grooves and flesh to stone ratio were determined. For endocarp, all observations of the endocarp were done on 20 endocarps, well cleaned, taken from the same fruit used for description. Furthermore, other qualitative variables were also recorded according to the methodology for primary characterization of olive varieties (IOOC, 1997; Barranco *et al.*, 2000; Ben Ayed *et al.*, 2016).

### Data analysis dendrogram

Qualitative and quantitative data sets were analysed. Quantitative variables were standardized (mean = 0, variance = 1) for numerical analysis (Manly, 1986). Qualitative data sets were scored as the presence or absence of a character. The data matrix was converted into a matrix of similarity (S) values using Jaccard coefficient (Jaccard, 1908). For a pair of two accessions,  $i$  and  $j$ , this coefficient is calculated as:

$$S_{ij} = n_{ij} / (n_{ij} + n_i + n_j)$$

where  $n_i$  is the number of bands present in accession  $i$  and absent in accession  $j$ ,  $n_j$  is the number of bands present in  $j$  and absent in  $i$ , and  $n_{ij}$  is the number of bands shared by the two accessions;  $i$  and  $j$ . A tree is then inferred using the unweighted pair group method using an arithmetic average (UPGMA) clustering algorithm. All analyses were done using NTSysPc program version 2.1 (Rohlf, 1999). The characteristics of encoding are using the method of rank number coding, of multiple standards of quality performance traits encoded by the classification method. The quantitative traits are not encoded and directly analysed in raw data form in the next operation. Five values are used to calculate the average of the studied quantitative trait.

### Bayesian networks modelling

Bayesian networks (BNs) are a powerful framework for decision support under uncertain knowledge. They come out

Table 1. Localisation of *Olea europaea* L. accessions (Tunisia)

Accession	Localisation	Fruit weight (grams)	Kernel weight (grams)
Chemlali Tataouine	Centre and South	0.68±0.08	0.15±0.02
Fougi	Centre and South	1.39±0.11	0.24±0.07
Sahli Gafsa	Centre and South	0.68±0.05	0.21±0.05
Semni Sfax	Centre and South	1.04±0.12	0.30±0.10
Gemri Dhokar	Centre and South	1.89±0.22	0.32±0.08
Chemleli Ontha	Centre and South	0.82±0.06	0.21±0.08
Oueslati	Centre and South	2.48±0.17	0.35±0.04
Tounsi Gafsa	Centre and South	4.06±0.26	0.47±0.11
Gemri	Centre and South	1.23±0.14	0.23±0.03
Chemlali Zarzis	Centre and South	1.18±0.04	0.22±0.05
Chemlali Sfax	Centre and South	1.06±0.16	0.24±0.07
Zalmati	Centre and South	1.49±0.20	0.19±0.06
Jemri Bouchouka	Centre and South	1.89±0.27	0.32±0.04
Chemlali Gafsa	Centre and South	2.44±0.29	0.30±0.09
Zarrazi Injassi	Centre and South	3.26±0.44	0.51±0.13
Dhokar	Centre and South	1.84±0.31	0.26±0.08
Baldi	Centre and South	2.12±0.44	0.27±0.07
Injassi Gafsa	Centre and South	5.05±0.42	0.67±0.08
Fakhari	Centre and South	1.24±0.21	0.35±0.11
Meski	North	5.82±0.77	0.94±0.22
Gerboui	Centre and South	2.58±0.64	0.86±0.13
Besbessi	North	9.6±0.48	1.35±0.42
Toffahi	Centre and South	2.82±0.24	0.34±0.06
Marsaline	North	1.84±0.16	0.47±0.10
Zarrazi	Centre and South	3.26±0.22	0.51±0.15
Chemlali Sousse	Centre and South	0.78±0.15	0.18±0.04
Chemlali Nabeul	Centre and South	0.72±0.11	0.17±0.05
Chetoui	North	1.96±0.27	0.55±0.14
Chemlali	Centre and South	1.06±0.28	0.24±0.08
Rkhaymi	North	1.88±0.34	0.94±0.012

from artificial intelligence studies and constitute one of the most coherent techniques for the acquisition and the modelling of complex systems. They have been applied to a large range of problems, and eventually in biology (Ennouri *et al.*, 2015, 2016). Since the data are discontinuous and experimental data are limited, and it is well known that the application of BN requires a lot of data for the learning and testing procedures, our proposed methodology includes the following three different stages for building model: Data normalization and Construction of Bayesian Network. A sample data from different experiences was normalized from experimental data and a Bayesian network (Pearl, 2000) is generated as following: two nodes  $i$  and  $j$  having a partial correlation are connected by a non-oriented edge. The orientation is determined by a heuristic method based on the following test: If  $B_{ij} = w_{ij} \sigma_i / w_{ji} \sigma_j > 1$ , the arc is then oriented from  $i$  to  $j$  and if  $B_{ij} = w_{ij} \sigma_i / w_{ji} \sigma_j < 1$ , the arc is then oriented from  $j$  to  $i$ . The other edges with  $B_{ij} = w_{ij} \sigma_i / w_{ji} \sigma_j = 1$  remained undirected. The graph with all directed arcs constituted the Bayesian network. It is imperative to note that it does not necessarily include all nodes contained in the network (Opgen-Rhein and Strimmer, 2007). The advantage of Bayesian network is to deduct all parent nodes which are directly dependent on child nodes. R program was used to analyze obtained data.

#### Principal Component Analysis

Principal component analysis (PCA) is a multivariate

statistical technique and it is used by almost all scientific disciplines. PCA analyzes a data table representing observations described by several dependent variables, which are, in general, inter-correlated. Its goal is to extract the important information from the data table and to express this information as a set of new orthogonal variables called principal components. Principal component analysis (PCA) of phenotypic data was performed using OriginPro 9.1. Absolute value of 0.50 was used in the loading matrices to select the traits in a particular principal component (PC). Correlations between variables were calculated with the Pearson correlation coefficients. Pearson's correlation coefficient is a statistical method of quantifying the association between two variables.

#### Results and Discussion

Name accessions, geographical localisation, fruit weight and kernel weight were presented in Table 1. Fruit weight values ranged between 9.6 grams and 0.68 grams, and kernel weight values ranged between 1.35 grams and 0.15 grams. Furthermore, the averages of fruit and kernel weights of studied accessions were 2.27 grams and 0.41 grams, respectively.

#### Genetic diversity and relationship between Tunisian olive accessions

To explain genetic relationships among 30 Tunisian olive accessions, a dendrogram was produced using UPGMA cluster

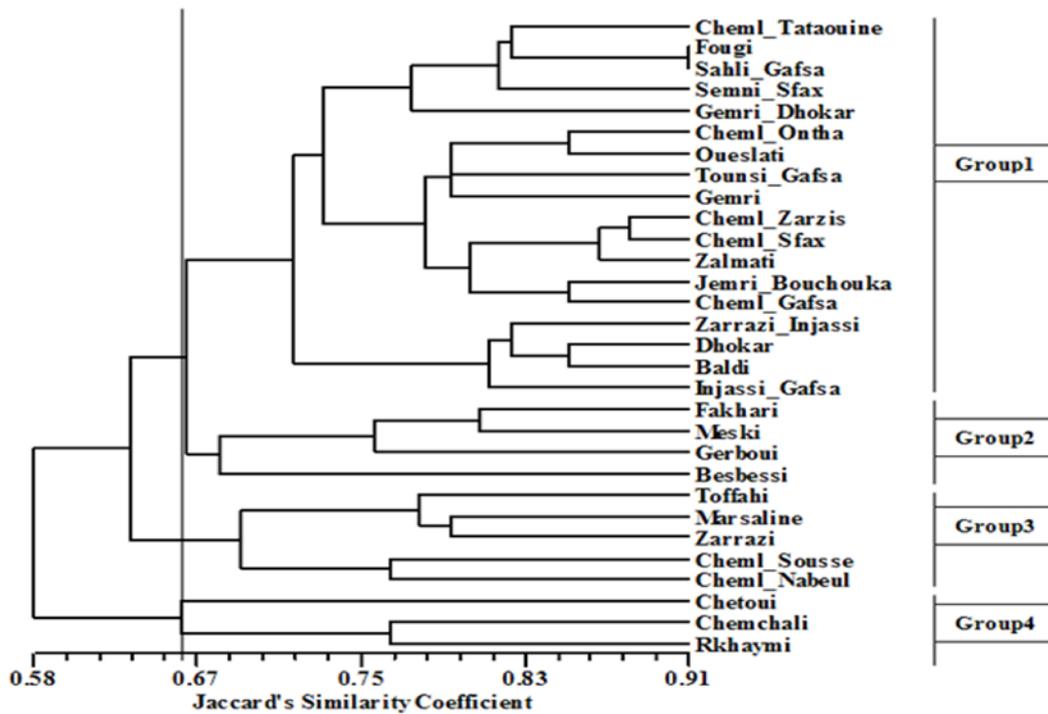


Fig. 1. Dendrogram of 30 Tunisian olive accessions generated by unweighted pair group method using an arithmetic average cluster analysis using Jaccard similarity coefficients from 21 morphological characters

analysis and the Jaccard similarity coefficients over 21 morphological characters (Fig. 1). The accessions studied can be divided into seven main groups by cutting the dendrogram at a similarity value of 0.75. The first group corresponds to the 18 accessions of the top of the dendrogram including accessions originated from center and south region. The second group contains 'Fakhari', 'Meski', 'Gerbou' and 'Besbessi'. The third group contains 5 accessions. The fourth group corresponds to the 3 accessions which also are clonally related. The UPGMA clustering obtained from the agronomic distance matrix showed a rather high variability among the most of Tunisian olive accessions generally clustered according to their geographical origin. For example, North Tunisia accessions were clustered in Group 2, 3 and 4. Accessions that originate from the centre and the south of Tunisia were clustered in Group 1.

Classification of accessions based on their geographic origin was also demonstrated in a larger geographic scale study presented by Belaj *et al.* (2001) with accessions from numerous Mediterranean Basin countries. Moreover, several previous studies demonstrated that the geographic and genetic structure was not exclusively observed among accessions of different countries, but also within accessions of the same country (Claros *et al.*, 2000; Carriero *et al.*, 2002; Ben Ayed *et al.*, 2011, 2015a). Likewise, Sanz-Cortés *et al.* (2001) observed in a specific region of Spain, subclustering according to the geographic origin within that region.

The grouping of the accessions from the same or nearby region suggests a common genetic base and an autochthonous source for these accessions. This result coincides with the hypothesis of autochthonous origin of most of the olive accessions as well as their limited diffusion from their centers of origin (Barranco and Rallo, 2000; Belaj *et al.*, 2001; Besnard *et al.*, 2001). Accession intercrossing and crosses with wild

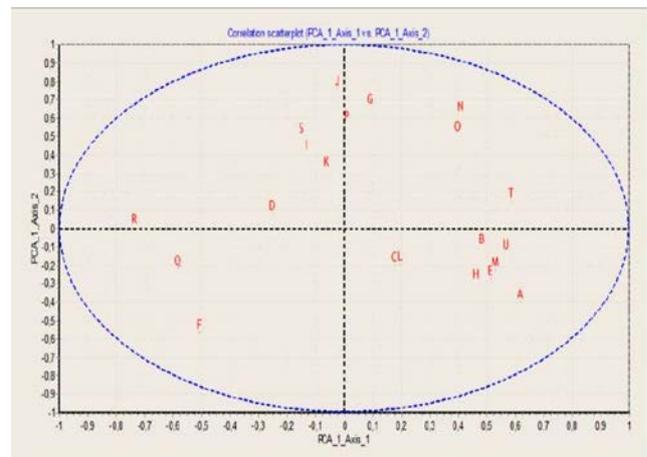


Fig. 2. Circle of correlations based on morphological data in olive varieties. Plot of first two PCs with contributing phenotypic traits (geographical localisation: A, tree habit: B, tree vigor: C, foliage density: D, leaf shape: E, fruit shape: F, symmetry of fruit: G, position of maximum diameter: H, nipple of fruit: I, type of the top of fruit: J, base of fruit: K, number of lenticels of fruit: L, Size of lenticels: M, core form: N, core symmetry: O, top of core: P, base of core: Q, core surface: R, flowering period: S, maturation period: T and oil production level: U)

accessions, along with local selection of outstanding seedlings and subsequent vegetative cloning, could have led to a large number of varieties around their possible original areas of cultivation. Conversely, using DNA-based markers such as SSR and SNP, Ben Ayed *et al.* (2011, 2015a) reported that most of Tunisian olive accessions clustered according to their

fruit size or commercial use (table or oil olive), but no classification was showed based on geographical origin. In addition, they found a comparable grouping pattern among 'Chetoui' and 'Rkhaymi' that were grouped with the Northern large-fruited size accessions.

*Principal Component Analysis*

Principal component analysis (PCA) on the basis of phenotypic data on fruit morphology and colour traits identified four principal components (PCs) explaining more than 51% of the total variation (Table 2). The circle of correlations was presented in the Fig. 2. In a correlation circle, each measured variable is shown as a vector, which signals the combined strength of the relationships between the measured variable and two PCs (vector length) and whether these relationships are positive or negative (vector direction). The angle between two vectors signals the degree of correlation between two measured variables. A right angle indicates that two variables are completely uncorrelated; zero or 180 degrees between two variables indicates complete positive or negative correlation. Correlation circles allow for graphical examination of the relationships among indices, and the consistency of these relationships among olive groups (Ben Ayed *et al.*, 2015a, 2015b). The Eigenvalues, contribution and cumulative contribution of studied characters were summarized in Table

2. The first PC had high loadings for core surface "R", geographical localisation "A", base of core "Q", maturation period "T", oil production level "U", Size of lenticels "M", fruit shape "F", and leaf shape "E", which explained together 18% of the total phenotypic variation. Core surface "R", fruit shape "F" and base of core "Q" were negatively correlated whereas

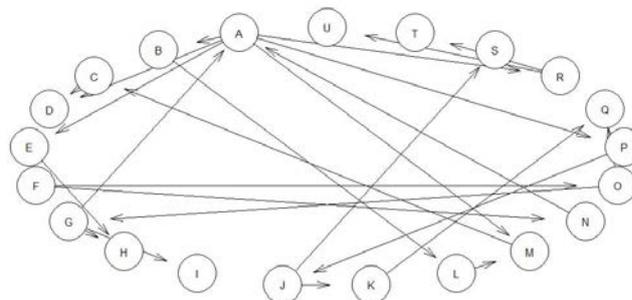


Fig. 3. Bayesian networks connecting related variables (geographical localisation: A, tree habit: B, tree vigor: C, foliage density: D, leaf shape: E, fruit shape: F, symmetry of fruit: G, position of maximum diameter: H, nipple of fruit: I, type of the top of fruit: J, base of fruit: K, number of lenticels of fruit: L, Size of lenticels: M, core form: N, core symmetry: O, top of core: P, base of core: Q, core surface: R, flowering period: S, maturation period: T and oil production level: U).

Table 2. Eigenvalues, contribution and cumulative contribution of principal components of 21 characters (Geographical localisation: A, tree habit: B, tree vigor: C, foliage density: D, leaf shape: E, fruit shape: F, symmetry of fruit: G, position of maximum diameter: H, nipple of fruit: I, type of the top of fruit: J, base of fruit: K, number of lenticels of fruit: L, Size of lenticels: M, core form: N, core symmetry: O, top of core: P, base of core: Q, core surface: R, flowering period: S, maturation period: T and oil production level: U)

Attribute	Axis1		Axis2		Axis3		Axis4	
	Corr.	% (Tot. %)						
R	-0.734	54% (54%)	0.013	0% (54%)	0.208	4% (58%)	-0.037	0% (58%)
A	0.618	38% (38%)	-0.396	16% (54%)	-0.066	0% (54%)	0.280	8% (62%)
T	0.586	34% (34%)	0.151	2% (37%)	-0.254	6% (43%)	-0.229	5% (48%)
Q	-0.582	34% (34%)	-0.214	5% (38%)	-0.109	1% (40%)	-0.370	14% (53%)
U	0.568	32% (32%)	-0.128	2% (34%)	-0.166	3% (37%)	-0.096	1% (38%)
M	0.532	28% (28%)	-0.224	5% (33%)	0.691	48% (81%)	0.096	1% (82%)
E	0.514	26% (26%)	-0.269	7% (34%)	0.410	17% (50%)	-0.011	0% (51%)
F	-0.508	26% (26%)	-0.563	32% (57%)	0.272	7% (65%)	-0.104	1% (66%)
J	-0.024	0% (0%)	0.761	58% (58%)	0.151	2% (60%)	-0.077	1% (61%)
G	0.092	1% (1%)	0.666	44% (45%)	-0.008	0% (45%)	-0.282	8% (53%)
N	0.409	17% (17%)	0.623	39% (56%)	-0.260	7% (62%)	0.225	5% (67%)
P	0.010	0% (0%)	0.569	32% (32%)	0.421	18% (50%)	-0.154	2% (52%)
O	0.398	16% (16%)	0.516	27% (42%)	0.077	1% (43%)	-0.188	4% (47%)
S	-0.147	2% (2%)	0.506	26% (28%)	0.140	2% (30%)	0.207	4% (34%)
L	0.195	4% (4%)	-0.194	4% (8%)	0.704	50% (57%)	-0.189	4% (61%)
K	-0.063	0% (0%)	0.325	11% (11%)	0.512	26% (37%)	-0.063	0% (38%)
C	0.176	3% (3%)	-0.198	4% (7%)	-0.088	1% (8%)	-0.697	49% (56%)
I	-0.131	2% (2%)	0.417	17% (19%)	0.042	0% (19%)	0.245	6% (25%)
D	-0.252	6% (6%)	0.083	1% (7%)	0.236	6% (13%)	0.432	19% (31%)
B	0.484	23% (23%)	-0.096	1% (24%)	0.018	0% (24%)	-0.413	17% (41%)
H	0.462	21% (21%)	-0.289	8% (30%)	-0.009	0% (30%)	0.455	21% (50%)
Var. Expl.	3.689	18% (18%)	3.394	16% (34%)	1.988	9% (43%)	1.687	8% (51%)

Table 3: Correlation coefficients among all morphological traits in olive varieties

	R	A	T	Q	U	M	E	F	J	G	N	P	O	S	L	K	C	I	D	B	H
R	1.000																				
A	-0.520	1.000																			
T	-0.539	0.160	1.000																		
Q	0.329	-0.206	-0.418	1.000																	
U	-0.452	0.284	0.372	-0.299	1.000																
M	-0.302	0.378	0.051	-0.284	0.174	1.000															
E	-0.206	0.496	0.149	-0.115	0.121	0.466	1.000														
F	0.237	-0.096	-0.249	0.323	-0.232	0.006	-0.069	1.000													
J	0.007	-0.355	-0.049	-0.036	-0.003	-0.097	-0.136	-0.309	1.000												
G	-0.034	-0.164	0.172	-0.060	-0.005	-0.162	-0.061	-0.320	0.498	1.000											
N	-0.236	0.138	0.205	-0.288	0.073	-0.021	-0.037	-0.759	0.354	0.389	1.000										
P	0.089	-0.276	0.053	-0.139	-0.062	0.081	0.025	-0.172	0.468	0.335	0.188	1.000									
O	-0.314	0.057	0.281	-0.213	-0.044	0.181	0.118	-0.461	0.320	0.350	0.313	0.241	1.000								
S	0.085	-0.295	0.126	-0.196	-0.221	0.018	-0.190	-0.065	0.373	0.171	0.254	0.122	0.090	1.000							
L	0.087	-0.043	-0.023	-0.170	0.147	0.632	0.305	0.193	-0.044	0.099	-0.199	0.119	-0.133	-0.058	1.000						
K	0.064	-0.044	-0.129	0.018	-0.096	0.194	0.112	-0.040	0.277	0.052	0.061	0.374	0.268	0.191	0.011	1.000					
C	-0.055	0.033	0.148	0.179	0.209	-0.044	0.160	0.066	-0.048	0.028	-0.129	-0.095	0.135	-0.274	0.052	0.009	1.000				
I	0.119	-0.214	-0.116	-0.148	-0.082	-0.220	-0.113	-0.107	0.347	0.287	0.163	0.201	0.050	0.046	0.031	-0.029	-0.171	1.000			
D	0.140	-0.092	-0.194	-0.101	-0.207	0.009	-0.059	0.131	0.029	-0.106	-0.115	-0.006	0.089	0.121	-0.096	0.156	-0.141	0.198	1.000		
B	-0.235	0.093	0.266	-0.240	0.126	0.265	0.191	-0.185	-0.038	-0.030	0.067	-0.041	0.208	-0.167	0.218	-0.183	0.351	-0.067	-0.225	1.000	
H	-0.217	0.411	-0.050	-0.329	0.245	0.294	0.335	-0.147	-0.054	-0.248	0.138	-0.156	-0.102	-0.254	0.068	-0.261	0.007	0.053	-0.048	0.115	1.000

localisation “A”, maturation period “T”, oil production level “U”, size of lenticels “M” and leaf shape “E” are positively correlated. The second PC had significant loadings for type of the top of fruit “J”, symmetry of fruit “G”, core form “N”, top of core “P”, core symmetry “O”, flowering period “S” and fruit shape “F” and parameter 2 explained 16% of the total phenotypic variation. The highest positive loading is attributed for type of the top of fruit “J” and negative for fruit shape “F”. The third PC explained 9 % of the total phenotypic variation. Finally, the fourth component explained 8% of the total phenotypic variation. This was a relatively moderate level of variation (51.22%) being explained by only four PCs using morphological data. Eigenvalue is the variance explained by the PCs. This variance can also be presented as proportion of the total phenotypic variance, which is 100% in total.

*Bayesian networks modelling*

BNs are directed acyclic graphs composed by nodes (variables of the problem) and arcs that encode conditional probabilistic independencies between the nodes. These graphical models are very attractive for their aptitude to explain probabilistic interactions connecting variables.

In fact, they have proven to capture causal relationships between variables and they can show excellent forecast accuracy even with relatively small sample data sizes (Benson, 2015). To achieve the mentioned objectives, Bayesian networks modelling were used. We considered 21 nodes as represented in Fig. 3. Correlation coefficients among all morphological traits in olive accessions are presented in Table 3. Determinations of physical and mechanical properties of agricultural products are very important factors in the design of processing, grading, transporting and other agricultural

machinery (Altuntaş *et al.*, 2007). Moreover, the shape and the size of the product are the most important physical properties (Altuntaş *et al.*, 2005). In our study, base of core “Q” has a double connection: “Q” is negatively related with core symmetry “O” and positively with base of fruit “K”. Core symmetry “O” and core form “N” were negatively influenced simultaneously by fruit shape “F” and core symmetry “O” which influenced directly symmetry of fruit “G”. El-Soaly (2008) found that the olive fruit length, diameter and weight were directly proportional to its pit for the investigated varieties.

Geographical localization “A” plays a role in the increase of tree habit “B”, Size of lenticels “M” and leaf shape “E”. In fact, olive leaves are affected by geographical region (Al-Rimawi *et al.*, 2014). However, localization “A” is involved in the decrease of foliage density “D”, core surface “R” and top of core “P”. It has been demonstrated that pit development is a continuous and progressive process and length of this period can vary according to water status (Rapaport *et al.*, 2004). Several studies have shown that climatic factors such as temperature and precipitation, closely related with geographical location, have an effect on plant physiology (Pannelli *et al.*, 1994; Ryan *et al.*, 1998). Furthermore, symmetry of fruit “G” has an active function in enhancement of nipple of fruit “I” and opposite effect on position of maximum diameter “H”. Moreover, position of maximum diameter “H” is affected by leaf shape “E”. The top of core “P” influenced positively the type of fruit top “J” and the base of fruit “K” is indirectly connected with the top of core “P” by the type of the top of fruit “J”. Moreover, core surface “R” has simultaneous associations: in fact, “R” is negatively connected with maturation period “T” and oil

production level “U”. The number of lenticels of fruit “L” influenced directly size of lenticels “M” and “M” has a retroactive effect on tree vigor “C”. Indeed, tree vigor has been defined as the overall physiological condition or “health” of a tree in a given environment (Wargo, 1978). Physiological condition of the tree will determine the tree’s response to defoliation, but it is difficult to measure (Kozłowski, 1969). Difficult environmental conditions such as drought, late spring frosts, ice storms, excessive moisture, competition, cutting, slash disposal, and other disturbances can cause stress on trees that influences and changes overall tree vigor. Trees can recover from stresses over time. Moreover, our findings are in concordance with several studies. In fact, De La Rosa *et al.* (2007) and Tous *et al.* (2010) demonstrated that ‘Arbequina’, a low vigor accession, is currently considered the most important accession for super high-density olive groves. Lenticels are lens-shaped macroscopic openings that occur on the surfaces of roots, shoots, some fruits (Kuo-Huang and Hung, 1995). Lenticels are essential to the plant, since they control gaseous exchange for photosynthesis, respiration and transpiration in the absence of stomata (Mauseth, 1988). The gross anatomy of mature lenticels in many plants has been described, but only a few published reports are referring to the development of lenticels (Jacob *et al.*, 1989).

## Conclusions

In conclusion, this research indicated that morphological markers are a preliminary tool to characterize olive oil accessions. In fact, this study has revealed that the distribution of Tunisian olive accessions is based on geographic origin (North, Center and South). The use of DNA-based markers such as SSR and SNP will be suitable to confirm our findings and provide automated tools for olive accessions identification, characterization and classification. Since the clustering of Tunisian olive accessions is a complicate task, we have need the combination of several marker systems to provide a more complete understanding of the diversity of available Tunisian olive tree accessions and the way in which it can be best used for olive breeding.

## Acknowledgements

The authors are grateful to Mr. Mohamed Ennouri and Mr. Hichem Ennouri, who were helpful and offered assistance and guidance.

## References

Adusei-Poku K, Van den Brink GJ, Zucchini W (2007). Implementing a Bayesian network for foreign exchange settlement: a case study in operational risk management. *Journal of Operational Risk* 2(2):101-107.

Al-Rimawi F, Odeh I, Bisher A, Abbadi J, Qabbajeh M (2014). Effect of geographical region and harvesting date on antioxidant activity, phenolic and flavonoid content of olive leaves. *Journal of Food and Nutrition Research* 2(12):925-930.

Altuntaş E., Özgöz E, Taşer ÖF (2005). Some physical properties of fenugreek (*Trigonella foenum-graceum* L.) seeds. *Journal of Food Engineering* 71(1):37-43.

Altuntaş E, Yildiz M (2007). Effect of moisture content on some physical properties of Faba bean grain. *Journal of Food Engineering* 78:174-183.

Barranco D, Cimato A, Fiorino P, Rallo L, Touzani A, Castaneda C, Serafini F, Trujillo I (2000). *World Catalogue of Olive Varieties*. International Olive Oil Council, Madrid, Spain.

Barranco D, Rallo L (2000). Olive cultivars in Spain. *HortTechnology* 10(1):107-110.

Belaj A, Trujillo I, De La Rosa R, Rallo L, Gimenez MJ (2001). Polymorphism and discriminating capacity of randomly amplified polymorphic markers in an olive germplasm bank. *Journal of the American Society for Horticultural Science* 126(1):64-71.

Ben Ayed R, Ben Hassen H, Ennouri K, Rebai A (2016). Genetic markers analyses and bioinformatic approaches to distinguish between olive tree (*Olea europaea* L.) cultivars. *Interdisciplinary Sciences: Computational Life Sciences* 8(4):366.

Ben Ayed R, Ennouri K, Ben Hassen H, Rebai A (2015a). Molecular phylogeny to specify *Zalmati* and *Chemlali Tataouine* Tunisian olive cultivars. *Journal of New Sciences* 18(6):689-694.

Ben Ayed R, Ennouri K, Ben Hassen H, Triki MA, Rebai A (2015b). Comparison between DNA-based, pomological and chemical markers accomplished by bioinformatic tools to distinguish within tunisian olive cultivars. *Journal of Fundamental and Applied Sciences* 7(3):408-421.

Ben Ayed R, Grati-Kamoun N, Sans-Grout C, Moreau F, Rebai A (2012). Characterization and authenticity of virgin olive oil (*Olea europaea* L.) cultivars by microsatellite markers. *European Food Research and Technology* 234:263-271.

Benson M (2015). *Bayesian Networks Handbook*. M L Books International, New Delhi.

Besnard G, Baradat P, Berville A (2001). Genetic relationships in the olive (*Olea europaea* L.) reflect multilocal selection of cultivars. *Theoretical and Applied Genetics* 102:251-258.

Carriero F, Fontanazza G, Cellini F, Giorio G (2002). Identification of simple sequence repeats (SSRs) in olive (*Olea europaea* L.) *Theoretical and Applied Genetics* 104:301-307.

Claros MG, Crespillo R, Aguilar ML, Canovas FM (2000). DNA fingerprinting and classification of geographically related genotypes of olive-tree (*Olea europaea* L.). *Euphytica* 116(2):131-142.

Connell JH (1994). History and scope of the olive industry. In: Ferguson L, Sibbett GS, Martin GC (Eds). *Olive Production Manual*. University of California, Division of Agriculture and Natural Resources, Berkeley, California pp 1-9.

De la Rosa R, León L, Guerrero N, Rallo L, Barranco D (2007). Preliminary results of an olive cultivar trial at high density. *Australian Journal of Agricultural Research* 58:392-395.

El-Soaly IS (2008). Some physical and mechanical proprieties of olive fruits. *Misir Journal of Agricultural Engineering* 25(3):899-909.

Ennouri K, Ben Ayed R, Ben Hassen H, Mazzarello M, Ottaviani E (2015). Experimental design and Bayesian networks for enhancement of delta-endotoxin production by *Bacillus thuringiensis*. *Acta Microbiologica et Immunologica Hungarica* 62(4):381-394.

- Ennouri K, Ben Ayed R, Mazzarello M, Ottaviani E, Hertelli F, Azzouz H (2016). Classical and Bayesian predictions applied to *Bacillus* toxin production. 3 Biotech 6(2):206.
- Eyres L (2012). Oils and fats news. Food New Zealand 12(1):22-23.
- Fernandez JE, Moreno F, Giron IF, Blazquez OM (1997). Stomatal control of water use in olive tree leaves. Plant and Soil 190:179-192.
- IOOC (1997). Primary characterization methodology of olive cultivars. Project of conservation, characterization collect and use of olive genetic resources. European Community, Olive International Council.
- Jaccard P (1908). Nouvelles recherches sur la distribution florale. Bulletin de la Société Vaudoise des Sciences Naturelles 44:223-270.
- Jacob A, Lehmann H, Stelzer R (1989). Development and structure of the lenticels from Beech-trees (*Fagus sylvatica* f. *purpurea* Ait.) Flora 183:417-427.
- Koller D, Friedman N (2009). Probabilistic graphical models: principles and techniques. MIT Press, Cambridge.
- Kozłowski TT (1969). Tree physiology and forest pests. Journal of Forestry 67:118-123.
- Kuo-Huang LL, Hung LF (1995). The formation of lenticels on the branches of *Ficus microcarpa* L. f. Taiwania 40(2):139-150.
- Manly FJ (1986). Multivariate statistical methods: A primer. Chapman and Hall, London-New York.
- Mantel N (1976). The detection of disease clustering and generalized regression approach. Cancer Research 27:209-220.
- Martin GC (1994). Botany of the olive. In: Ferguson L, Sibbett GS, Martin GC (Eds). Olive Production Manual. University of California, Division of Agriculture and Natural Resources, Berkeley, California pp 19-21.
- Mauseth JD (1988). Plant anatomy. The Benjamin Cummings Publishing Company Inc, California, USA.
- Mez-Rico A, Fregapane G, LaGreca M, Salvador MD (2006). Phenolic and volatile compounds of extra virgin olive oil (*Olea europaea* L. Cv. Cornicabra) with regard to fruit ripening and irrigation management. Journal of Agricultural and Food Chemistry 54:7130-7136.
- Opgen-Rhein R, Strimmer K (2007). From correlation to causation networks: a simple approximate learning algorithm and its application to high-dimensional plant gene expression data. BMC Systems Biology 1:1-37.
- Pannelli G, Servili M, Selvaggini R, Baldioli M, Montedoro GF (1994). Effect of agronomic and seasonal factors on olive (*Olea europaea* L.) production and on the qualitative characteristics of the oil. Acta Horticulturae 356:239-243.
- Patumi M, Giorio P, Sorrentino G, Morelli G, D'Andria R, Fontanazza G (1999). Yield and oil quality of intensively trained trees of three cultivars of olive (*Olea europaea* L.) under different irrigation regimes. Journal of Horticultural Science and Biotechnology 74:729-737.
- Pearl J (2000). Causality: models, reasoning, and inference. Cambridge University Press, Cambridge.
- Poljuha D, Sladonja B, Setic E, Milotic A, Bandelj D, Jakse J, Javornik B (2008). DNA fingerprinting of olive varieties in Istria (Croatia) by microsatellite markers. Scientia Horticulturae 115:223-230.
- Pourret O, Naim P, Marcot B (2008). Bayesian networks: a practical guide to applications. John Wiley and Sons Ltd, Chichester, England.
- Proietti P, Famiani F, Tombesi A (1999). Gas exchange in olive fruit. Photosynthetica 36(3):423-432.
- Rapaport HF, Costagli G, Gucci R (2004). The effect of water deficit during early fruit development on olive fruit morphogenesis. Journal of the American Society for Horticultural Science 129:121-127.
- Rohlf FJ (1999) NTSYS-pc. Numerical taxonomy and multivariate analysis system. Version 2.1. Exeter Software, Setauket.
- Ryan D, Robards K, Lavee S (1998). Evolution de la qualité de l'huile d'olive. Olivae 72:23-41.
- Sanz-Cortes F, Badenes ML, Paz S, Iniguez A, Llacer G (2001). Molecular characterization of olive cultivars using RAPD markers. Journal of the American Society for Horticultural Science 126(1):7-12.
- Therios I (2009). Olives (Crop Production Science in Horticulture). CAB International, Oxfordshire, UK.
- Tous J, Romero A, Hermoso JF (2010). New trends in olive orchard design for continuous mechanical harvesting. Advances in Horticultural Science 24:43-52.
- Wargo PM (1978). Insects have defoliated my tree-now what's going to happen? Journal of Arboriculture 4:169-175.