

Database, 2016, 1–9 doi: 10.1093/database/bav090 Database tool



Database tool

# OGDD (*Olive Genetic Diversity Database*): a microsatellite markers' genotypes database of worldwide olive trees for cultivar identification and virgin olive oil traceability

# Rayda Ben Ayed<sup>1,\*</sup>, Hanen Ben Hassen<sup>2</sup>, Karim Ennouri<sup>1</sup>, Riadh Ben Marzoug<sup>1</sup> and Ahmed Rebai<sup>1</sup>

<sup>1</sup>Centre of Biotechnology of Sfax, PB '1177', 3018 Sfax, Tunisia and <sup>2</sup>Laboratory of Physics Mathematics and Applications, Faculty of Sciences of Sfax, Sfax, Tunisia

\*Corresponding author: Tel: +21674871816; Fax: +21674875818; Email: raydabenayed@yahoo.fr

Citation details: Ben Ayed,R., Ben Hassen,H., Ennouri,K., *et al.* OGDD (*Olive Genetic Diversity Database*): a microsatellite markers' genotypes database of worldwide olive trees for cultivar identification and virgin olive oil traceability. *Database* (2015) Vol. 2015: article ID bav090; doi:10.1093/database/bav090

Received 25 April 2015; Revised 13 July 2015; Accepted 24 August 2015

# Abstract

Olive (Olea europaea), whose importance is mainly due to nutritional and health features, is one of the most economically significant oil-producing trees in the Mediterranean region. Unfortunately, the increasing market demand towards virgin olive oil could often result in its adulteration with less expensive oils, which is a serious problem for the public and quality control evaluators of virgin olive oil. Therefore, to avoid frauds, olive cultivar identification and virgin olive oil authentication have become a major issue for the producers and consumers of quality control in the olive chain. Presently, genetic traceability using SSR is the cost effective and powerful marker technique that can be employed to resolve such problems. However, to identify an unknown monovarietal virgin olive oil cultivar, a reference system has become necessary. Thus, an Olive Genetic Diversity Database (OGDD) (http://www.bioinfo-cbs.org/ogdd/) is presented in this work. It is a genetic, morphologic and chemical database of worldwide olive tree and oil having a double function. In fact, besides being a reference system generated for the identification of unkown olive or virgin olive oil cultivars based on their microsatellite allele size(s), it provides users additional morphological and chemical information for each identified cultivar. Currently, OGDD is designed to enable users to easily retrieve and visualize biologically important information (SSR markers, and olive tree and oil characteristics of about 200 cultivars worldwide) using a set of efficient query interfaces and analysis tools. It can be accessed through a web service from any modern programming language using a simple hypertext transfer protocol call. The web site is implemented in java, JavaScript, PHP, HTML and Apache with all major browsers supported.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from https://academic.oup.com/database/article/doi/10.1093/database/bav090/2630090 by guest on 20 May 2024

# Introduction

Olive tree (Olea europaea L.) represents the most important oil producing crop in the Mediterranean basin. Its cultivation covers over eight million hectares of land and about 70% of the produced olive oil is consumed. Olive tree is a diploid species (2n = 46), whose seeds are produced by cross-pollination (1, 2), is able to survive for a long time. It is a glycophytic species that exhibits a big tolerance to drought and salt stresses when it is compared with other fruit trees that are generally salt sensitive.

Olive oil is the oil extracted from *Olea europaea* L. fruit, using only mechanical methods or other physical procedures that do not alter the glyceric structure of the oil, and therefore conserve its vitamins and other natural healthy high-value compounds. The virgin olive oil is recognized to have beneficial effects on health. In fact, it is able to reduce blood pressure and low-density lipoprotein cholesterol. It has also antioxidant and antimicrobial virtues, such as cancer prevention (3, 4).

According to the Food and Agriculture data from the United Nations, the Mediterranean countries produce 90% of World olives, and the main olive producers are Spain, Italy, Greece, Turkey, Tunisia, Morocco, Syria and Portugal (FAO). Olive and olive oil have great commercial and economical importance in the Mediterranean region. In the last few years, the growth of olive production has expanded throughout the world and olive oil consumption all over the world has significantly increased, and consumers are becoming more informed and increasingly aware of the quality of foods they buy and eat.

Virgin olive oils obtained from one genetic variety of olive or from different varieties are called monovarietal or coupage, respectively. Regarding the monovarietal virgin olive oils, they possess specific characteristics that are associated with the olive variety from which they are extracted. In addition, a high occurrence of mislabeling, homonyms and synonyms has been reported in olive (5). Therefore, it is very important to improve or develop new traceability systems allowing easy and accurate cultivars and oils identification to manage properly the rich variability of olive. A well-documented traceability process and a confirming authenticity tool are needed for the control of the quality of the virgin olive oil introduced in the market. In other words, great effort is being made to obtain a unique and unequivocal genetic profile for every cultivar using molecular markers, since major chemical analyses of virgin olive oils from the similar category but from diverse origins have

a limited significance. Indeed, despite the high variability related to the environmental conditions of the different olive groups, their morphological characteristics and the analyses of chemical composition of fatty acid and secondary metabolites are not able to supply reliable results for oil traceability (6–8). For this reason, the genetic identity seems to be the most appropriate technique to identify the variety from which the olive oil under study derives. The use of DNA-based technology in the field of food authenticity, particularly olive oil, is gaining increasing attention.

Recently, the use of molecular markers, such as RAPDs (random amplified polymorphic DNA) (9), AFLPs (amplified fragment length polymorphisms) (10) and SSR (simple sequence repeats), has been recommended to depict virgin olive oil origin and traceability (6, 7, 11, 12). At present, microsatellites (SSRs) are the most relevant genetic markers used in olive cultivar characterization and virgin olive oil authenticity thanks to their numerous properties. Indeed, SSR markers are multiallelic, codominant, highly polymorphic, widely distributed along the plant genomes and easily amenable to PCR-based analyses. Moreover, they have great reproducibility and are currently the most reliable DNA profiling techniques in forensic investigation (13). Indeed, by carrying out Simple Sequence Repeats (SSRs) markers analysis, we can characterize the genetic profile of monovarietal virgin olive oil by comparing the oil-derived pattern with reference database olive oil cultivars. Thus, to efficiently identify and analyse the unknown commercial virgin olive oils, the development of a database including information about olive cultivars and olive oils based on genetic data sets, particularly, SSR markers, becomes necessary. Few online databases have been developed for olive tree (olea europaea L.), such as istrian olive database (14), which includes morphological and molecular data of some istrian olive trees and OLEA databases (http://www.oleadb.it/) (15). The latter contain data (microsatellite (SSR) loci) of a wide set of olive cultivars and give the possibility to query for cultivars corresponding to a specific data profile or to look for the variety identity when a profile is available. Another database that can be mentioned is Olea EST database (16). Actually, it is a collection of Olea europaea L. and is constructed by first clustering EST reads to produce tentative consensus (TC) sequences and singletons (sESTs). The database annotates and classifies the unique transcripts found according to their biological functions. Other databases such as GLOBAL INVASIVE SPECIES DATABASES (17) remain classical and give simple botanical and biological description of several tree species (such as olea). Among all databases proposed in the literature, only OLEA databases (15) seem to give effective molecular identification of olive cultivars, but they generate only SSR marker size(s) of each olive cultivar without any other information. Therefore, it is necessary to provide public a new available database for identifying an unknown olive tree or oil using SSR markers and providing extended profile description pertaining to the displayed cultivar.

For this purpose, we generated a simple format database of olive species called OGDD, for Olive Genetics Diversity Database, which currently contains many olive tree and oil characteristics (agro-morphological, chemical, genetics (SSR DNA band size(s) ...). The created database also integrates a computer application that provides a supplementary test which compares a user-provided SSR fingerprint. In this paper, we present the OGDD database that is currently implemented on the website (http://www. bioinfo-cbs.org/ogdd/) and which can be considered as a reference system in evaluating data obtained from the analysis of unknown samples and in defining the origin and composition of the virgin olive oil. This database can be used by all researchers and stakeholders interested in olive oil field. In the near future, this database will be a useful platform in the traceability and authenticity of olive oil.

# Database construction and data description

#### Data extraction

The OGDD site allows anyone interested in the olive oil sector to simply and quickly recover genetic profile from

**Table 1.** Data sources of 200 world-wide olive and virgin olive

 oil cultivars included in the OGDD database

Cultivar origin	Source(s)					
	Agro- morphological data	Pomological and chemical data	Genetic data <sup>a</sup>			
Italy	(25)	(25, 26, 27, 28, 29)	(25, 30, 31)			
Spain	(32)	(33)	(31, 34)			
Greece	(32)	(35, 36)	(31, 34)			
Portugal	(32)	(37)	(31, 34)			
France	(38)	(38)	(31)			
Tunisia	(39)	(36, 39, 40, 41, 42, 43)	(6, 7, 34,44)			
Turkey	(32)	(45)	(34)			
Syria	(46)	(46)	(34, 46)			
Morocco	(32)	(47, 48)	(31, 34, 49)			

<sup>a</sup> Up to now the genetic data existing in the OGDD database are the allele sizes of SSR markers.

known world-wide olive cultivars. The data included in the 'OGDD' derive from publications on morphological, biochemical, sensory and molecular olive oils cultivated world-wide (Table 1). For example, the genetic profiles of different olive oil varieties involved with SSR families are as follows: (DCA (18), GAPU (19), UDO (20), EMO (21), IAS-oli (22, 23), IGP (24). Each family marker is composed of 2 (EMO) to 13 (IAS-oli) markers and for each locus, the allele size of both alleles expressed in base pairs is provided. In the case of homozygosis, both values are equal.

Thus, each variety has a range of additional data such as country, region, morphological data represented by four photos (tree, leaf, fruit and kernel), acidity, taste, synonym, bibliographic references, oil content (%), the physicochemical composition and a matching score calculated to identify each new genetic profile. In fact, a user can compare the genetic profiles of the varieties presented in the database with the genetic profile that s/he experimentally determines. S/He can, therefore, identify the origin of her/ his cultivar.

#### Data collection and classification

The OGDD data are organized into a database that gathers each type of data separately. The following information is collected for each olive cultivar variety such as the country of origin, the growing region, the biological characteristics of the variety (pollination, climate requirements and resistance to diseases, pests and climates), morphological characteristics (tree, leaf, fruit and nucleus), biochemical characteristics (acidity, oil content, polyphenols, tocopherols, pigments (chlorophylles, carotenes), ...), organoleptic characteristics (taste and aroma) and molecular data particularly microsatellite markers (SSR) (Figure 1). For the microsatellite markers, the retrieved data are their name and the groups to which they belong and their allele size (base pairs). OGDD updates are provided manually yearly by the bioinformatics group of the Centre of Biotechnology of Sfax, Tunisia, by checking the SSR markers that are newly obtained by research teams. The current status of the molecular markers data included in OGDD database contains only the SSR markers. In a following step, other molecular markers such as SNP markers will be integrated. Updates include records and complete information concerning Olea europaea species.

## Database construction and implementation

OGDD is a browser-independent Web database built using MySQL (version 5.1.41) as the Database Management System. The OGDD database web interface was constructed using HTML, JavaScript and PHP as programming

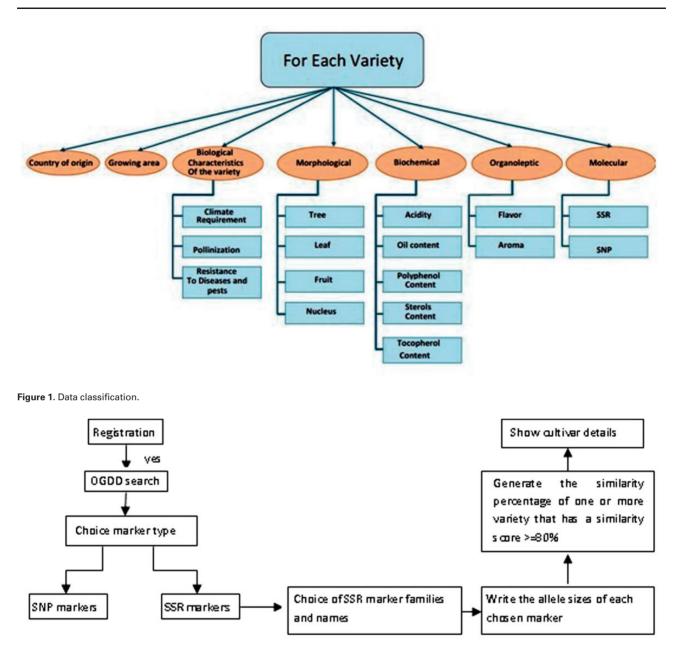


Figure 2. A schematic architecture of OGDD database. Solid arrows mean that the user can choose the analysis path according to results obtained at each previous step. Dashed lines mean that user can return to the OGDD search if he didn't find the required SSR.

languages. The database was hosted on Apache (version 2.2.14) web server with a Linux operating system. The web tool is compatible with all major browsers including FireFox, Safari, Chrome and Internet Explorer.

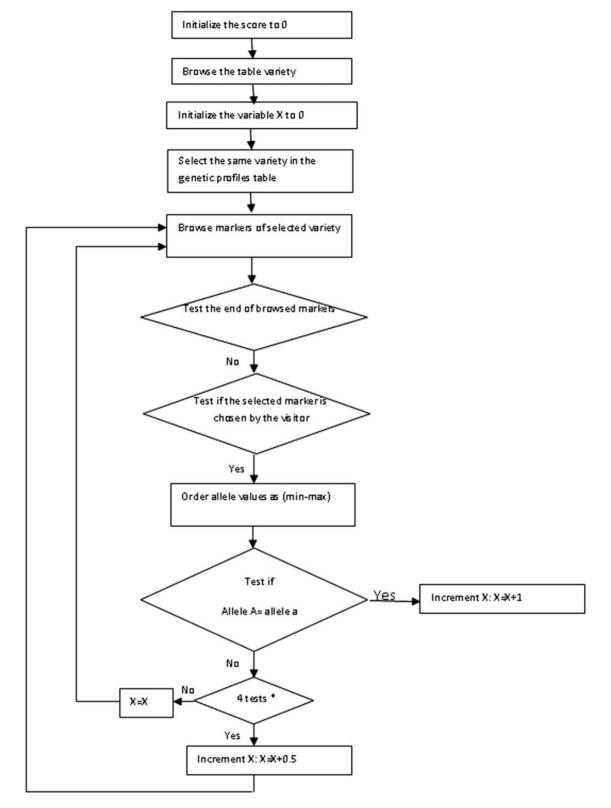
The architecture of OGDD is shown in Figure 2. The data in OGDD can be used to compute the similarity of an unknown virgin olive oil variety based on the SSR information and compare with those existing in the database.

#### Database use and access

The major goal for OGDD is to develop an approach toward identification, traceability and authenticity of virgin olive oil to protect the interest of both consumers and producers. OGDD database can be accessed from any computer with web-access, just requiring the registration of the user.

#### Algorithm development and score calculating

To help the users know and identity the origin of the virgin olive oil genetic profile they have, we propose the diagram of the implemented algorithm that identifies the unknown variety based on the data available in the database by computing the probabilities of homology with the unknown virgin olive oil variety user input (Figure 3). The result is displayed as a percentage of homology.





The graphical depiction of the algorithm is given in Figure 2. The scoring computation will proceed with four major steps. First, it does not only use an intermediate variable X for the exact calculating score(s), but also initializes

the scores range to zero for all cultivars existing in the present database. Second, for each variety it initializes the variable X to zero, selects the genetic profile of every variety and checks all markers of this variety already chosen by the

(a) Authenticati	on		(b) C	hoice of m	arker		
Olive Genetic Diversit	and company		Genetic Dive	rsity Data	base	80	
18	ASSAULT PLATES THE PLATE PLATE		1	BELECT MARK	EMB		
				Ill prosts	Mathan Care	424 W2224	-
	Exercuted			ED BEANS ED BEANS	Datives Res Deters Datives	Alle Book	
	Annual of the Annual Annua		n	100 (Mar)	Beaus Geo	in Dear	
				GLANFETRA Manfetta	Deservices D	nina Elevie	uria.
c) Allele size		~					
ic Diversity Database			V V		Ц		
		ty Database	lie Diver	sity Data	V	The second	
ENTER THE DNA PRO	FILE TO IDENTIFY	HELP - DISCLAMED		1-1-1		Page 1	
r Janut	EN	TER THE DNA PROFILE		NO HELP DISCLONED	She -		
Number of p	-	inersh		ENTER THE DN	A PROFILE TO IS	DENTIFY	
DEADS 232 DEADS 754		tumber of p		[ beauth			
0.07¥716 [144	- Tex	1448 <b>247</b>		marker at p			
	(Rent.)	x6439 [196	200	DCA03		. [243	
		117 (117	Rasset	00409 [183 04PUTIk [124		183	
			_	and the last	Rea		[00]
) Variety identification	17						
	etic Diversity	Database	100				
		AND A PARTY PROPERTY AND	and the second se		11		
c Diversity Databa	ise	OFFICIAMEN	E.	- D-			
E PARIN V	ise A A		si	ity Data	base	10	
E PARIN V	ise	ARETY CAN BE AS:		ity Datal	M DECEMPANT	10	
E PARIN V				and and	M DECEMPANT	10	
	DE AS:	RETY CAN BE AS:	Pichates Likejantos	and and	ter-	10	
		ROFILE 15:	Pichates Likejantos	HELP, DECLAINER	ter-	10	
THE VARIETY CAN	DE AS:	ARETY CAN BE AS:	Panatina Languestoi Th	HELP, DECLAINER	ter-	Attendes	Details
THE VARIETY CAN	BE AS:	ARETY CAN BE AS:	Permise Langestor 19 19	HER OSCIANCE	AN DE AS:	Arbeeuine	Details
THE VARIETY CAN	BE AS: Cnittee Details	COFILE 18: 1230/01 - 2130/01 1240/0- 1240/0 1440/0- 1440/0	Permise Langestor 19 19	HE VARIETY C	AN DE AS:	Artequise	Details
THE VARIETY CAN	BE AS: Cnitter Details 2019-0 2019-0 12240	COFILE 18: 1230/01 - 2130/01 1240/0- 1240/0 1440/0- 1440/0	Pennina Languesta: Destains 11 11 11 11 11 11 11 11 11 11 11 11 11	HELD, DOCUMER HE VARIETY C 10% HE PROFILE IN	AN DE AS:	Attenits	Details
THE VARIETY CAN	BE AS: Cnitter Details 2019-0 2019-0 12240	COFILE 18: 1230/01 - 2130/01 1240/0- 1240/0 1440/0- 1440/0	Contains Langueebol 11 11 11 11 11 11 11 11 11 11 11 11 11	HE VARIETY C	AN DE AS:	Attente	Detais
THE VARIETY CAN THE VARIETY CAN TOPS THE PROFILE IS: CASS 0796 CARS 1796 CARS 1796 CARS 1796 CARS 1796 CARS 1796	BE AS: Cnitter Details 2019-0 2019-0 12240	COFILE 18: 1230/01 - 2130/01 1240/0- 1240/0 1440/0- 1440/0	Automatical Contraction of Contracti	HE VARIETY C	AN DE A6: b: b: b: - 24090 pa - 24090	Ateaus	Details
THE VARIETY CAN 10% THE VARIETY CAN 10% THE PROFILE IS: SCAIS 51796 SCAIS 11960 SCAIS 11960 Modig Ince as	BE AS: Cnitter Details 2019-0 2019-0 12240	COFILE 18: 1230/01 - 2130/01 1240/0- 1240/0 1440/0- 1440/0	Automatical Contraction of Contracti	HE VARIETY C HE VARIETY C 10% HE PROFILE H CA03 228 CA03 103 APU715 124	AN DE AS: B: ps - 24596 ps - 144pb	2000 - Attendes	Details
THE VARIETY CAN THE VARIETY CAN TOPS THE PROFILE IS: SCARS STORE SCARS STORE	BE AS: Cnitter Details 2019-0 2019-0 12240	COFILE 18: 1230/01 - 2130/01 1240/0- 1240/0 1440/0- 1440/0	Automatical Contraction of Contracti	HE VARIETY C HE VARIETY C 10% HE PROFILE H CA03 228 CA03 103 APU715 124	AN DE AS: B: ps - 24596 ps - 144pb	. Arteouile	Details
CIRCOOLOGY 722 BELP CORLABOL THE VARIETY CAN 100% THE PROFILE IS: DCASS 0796 DCASS 07	BE AS: Cnitter Details 2019-0 2019-0 12240	COFILE 18: 1230/01 - 2130/01 1240/0- 1240/0 1440/0- 1440/0	Permissi Languestic Languestic In In In In In In In In In In In In In	HELP OSCUMER HE VARIETY C 10% HE PROFILE IN CANY 22% CANY 10% APUTID 124 Modify New	AN DE AS: B: ph - 24090 ph - 15300 ph - 144p0 - 1644p0	Artenite	Detais
CREASE STATE	Decate Cristee 20160 exh Cristee Cristee	COFILE 18: 1230/01 - 2130/01 1240/0- 1240/0 1440/0- 1440/0	Automatical Contraction of Contracti	HE VARIETY C HE VARIETY C 10% HE PROFILE H CA03 274 163 APDR10 124 Modily Here Arbert	AN DE AS: B: ps - 24050 ps - 11360 ps - 1160 ps - 11	- Arteaura	Details
CIERCOOLOGY 722 PELP CONCUMENT THE VARIETY CAN 100% THE PROFILE IS: SCASS 04796 DCASS 04796 CARPATION 11720 Modely New AN CONCENSION NAME CONCENSION NAME	Chetter Chetter 2016-0	ARETY CAN BE AS:		HE VARIETY C HE VARIETY C 10% HE PROFILE II CA03 274 163 ABUUTO 122 Modily New CA03 CA	AN DE AS: B: ps - 24096 ps - 14420 ps - 14420 eranyth summa s	Attenue	Details
CONSTRUCTION FOR THE VARIETY CAN THE VARIETY CAN TOTA THE PROFILE IS: DCANS 17400 CANVITO 11220 Modify Down	Contrait Con	COFILE IS:	Parameter Langeesse Langeesse Til Second Content Accession name Accession name Accession name Accession name Accession name Accession name Accession name Accession name	HE VARIETY C	AN DE AS: B: ps - 24096 ps - 14420 ps - 14420 eranyth summa s	Attendes	Details
CONSTRUCTION FOR THE VARIETY CAN THE VARIETY CAN TOTA THE PROFILE IS: DCANS 17400 CANVITO 11220 Modify Down	Chatter Chatter 2016-0 2016-0 2020-0 arch Chatter	ARETY CAN BE AS:	Penning Langeston Langeston Langeston Second Control C	HE VARIETY C HE VARIETY C 10% HE PROFILE II CLA93 22% CLA93 124 Modely Name Autor Name	AN DE AS: B: ps - 24096 ps - 14420 ps - 14420 eranyth summa s		}
CORRECTOR FAIL FELD CONCLARED THE VARIETY CAN 100% THE PROFILE IS: DCASS 04796 DCASS 04796 CARETOR 11296 Model Corrector 11296 Model Corrector 11296 Model Corrector 11296 Model Corrector 11296 Corrector 11296 Co	IBE AS: Context 2019-0 1222-0 arch Context	COFILE IS:	PENNING Independent Independen	HE VARIETY C HE VARIETY C 10% HE PROFILE II CLA93 22% CLA93 124 Modely Name Autor Name	AN DE AS: b: b: ps - 24596 ps - 14456 ps - 14556 ps - 145566 ps - 145566	{	}
CARE AND	Contrait Contra	ARETY CAN BE AS:	PENNING Independent Independen	HE VARIETY C HE VARIETY C 10% HE PROFILE II CLA93 22% CLA93 124 Modely Name Autor Name	AN DE AS: b: b: ps - 24596 ps - 14456 ps - 14556 ps - 145566 ps - 145566	{	}
CONCOMIC DATA	IBE AS: Context 2019-0 1222-0 arch Context	COFILE IS: 23240 - 233pe Pade - 234pe Pade - 234pe Pad	PERSONAL PARTY OF CONTRACT OF	HE VARIETY C HE VARIETY C 10% HE PROFILE II CLA93 22% CLA93 124 Modely Name Autor Name	AN DE AS: b: b: ps - 24596 ps - 14456 ps - 14556 ps - 145566 ps - 145566	{	}
CCEREDON RAME	Contrait Contra	ARETY CAN BE AS:	ACCESSION NAME Annues MORPHOLOGICAL DAT Nervers MORPHOLOGICAL DAT Survey Morephological DAT Survey Morephological DAT Survey Morephological DAT Survey Morephological DAT Survey Morephological DAT Survey Morephological DAT	HE VARIETY C HE VARIETY C 10% HE PROFILE II CLA93 22% CLA93 124 Modely Name Autor Name	AN DE AS: b: b: ps - 24596 ps - 14456 ps - 14556 ps - 145566 ps - 145566	{	}
CCEREDON HAME	Contrait Contra	ARETY CAN BE AS:	ACCESSION NAME Accession Access	HE VARIETY C HE VARIETY C 10% HE PROFILE II CLA93 22% CLA93 124 Modely Name Autor Name	AN DE AS: b: b: ps - 24596 ps - 14456 ps - 14556 ps - 145566 ps - 145566	{	}
	Contrait Contra	ARETY CAN BE AS:	ACCESSION NAME ACCESSION NAME	HE VARIETY C HE VARIETY C 10% HE PROFILE II CLA93 22% CLA93 124 Modely Name Autor Name	AN DE AS: b: b: ps - 24596 ps - 14456 ps - 14556 ps - 145566 ps - 1455666 ps - 1455666 ps - 1455666 ps - 1455666 ps -	{	}
Details	Contrait Contra	ARETY CAN BE AS:	ACCESSION NAME Accession Access	HE VARIETY C HE VARIETY C 10% HE PROFILE II CLA93 22% CLA93 124 Modely Name Autor Name	AN DE AS: b: b: ps - 24596 ps - 14456 ps - 14556 ps - 145566 ps - 1455666 ps - 1455666 ps - 1455666 ps - 1455666 ps -	{	Detals.

Figure 4. A workflow to identify and visualize a monovarietal olive oil using OGDD database. (a) Authentication page where user needs to login, (b) Choice of marker page with selection of desired markers, (c) Allele size page specifying the size of the allele. (d) Variety identification, displaying the resulting monovarietal olive oil cultivar, (e) Details: displaying characteristics (morphological, agronomical, physicochemical data) of the identified cultivar.

user: if yes, the user must enter the size of the markers of the unknown cultivar. Third, the scoring computation compares allele sizes entered to the database with those of all varieties existing in the OGDD for the same marker. The X value is incremented according to the allele sizes of the existing varieties in OGDD. In fact, it is 1 if the two allele sizes are equal, 0.5 when only one allele size is equal and 0 when the two allele sizes are different. Finally, it calculates the score after the comparison of SSR markers. The score by this method depends on both similarities between the allele sizes and the number of SSR markers.

Calculate the score: Score = X/N

- X: sum of the elementary scores for each marker
- N: number of markers typed by the user

Select the varieties that have a score  $\geq$  0.8.

If there is no variety that satisfies this condition, the result will be: 'Not found variety in this genetic profile'.

## OGDD web site interrogation

Once the parameters of the database are defined and the data are grouped, the database is created physically with php/MySQL.

Querying the database needs a registration for a user account. The administrator has the option to accept or reject the user access permission. In addition, the administrator must authenticate to access the session and perform the functions described in this page. Once accepted, the user must log, check the boxes to select markers, enter the marker numbers and the allele sizes. The result is a homology probability of the unknown variety in question with the varieties of the database. In addition, for each displayed result, we have a link to the details of the variety, namely, the morphology of the tree, leaves and fruit, biochemical characteristics, oil organoleptic quality and allele sizes of all SSR markers existing until now in the database. Such a result is illustrated in Figure 4 following an interrogation case of our database. Finally, the user can edit or modify his/her search.

#### Case examples of unknown varieties

Based on the algorithm described above, we have developed software for data basing and managing SSR DNA fingerprint profiles. A query profile can be compared with all fingerprint profiles included in the database, resulting in a profile list formed at decreasing similarity percentages. Only the results with similarity levels higher than 80% are displayed. Three profile samples (SSR fingerprints of three unknown cultivars) have been separately tested in the OGDD database. The objective was to identify with accuracy these three varieties from their SSR fingerprint profiles. Therefore, for each unknown cultivar, we selected three markers (DCA03, DCA09, GAPU71b). Allele marker sizes were respectively included as follows: 232/253, 194/194, 144/144 for the first sample; 247/255, 194/206, 117/122 for the second sample and 228/240, 183/183, 124/144 for the third sample. For each examined cultivar, the displayed result showed 100% similarity to *Chetoui, Picholine Languedoc* and *Arbequina* cultivar, respectively. Clicking on details, the database displayed characteristics (such as, morphological, agronomical, physicochemical data) for each resulting monovarietal virgin olive oil cultivar (Figure 4).

## Conclusion

OGDD, which is a new and comprehensive database, has been developed focusing on the identification and authentication of olive plant. Compared with other few existing databases for olive species, OGDD has its own specific features and advantages. In fact, it provides a reference system in evaluating the data obtained from the analysis of unknown samples and in defining the varietal composition of the virgin olive oil. It also contains a great deal of information for each cultivar identified by the user. This database can be used by all researchers and stakeholders interested in olive cultivar identification and virgin olive oil authentication. Up to now, OGDD initializes the search only by SSR markers. However, its next update extends the query by adding other molecular markers, such as, SNP and includes more information of other olive cultivars.

#### Acknowledgements

The authors thank Mariem Masmoudi and Farah Kallel for their contribution to the web interface design, and Leila Mahfoudhi, English teacher at the Sfax Faculty of Science, for proofreading the manuscript.

Conflict of interest. None declared.

#### References

- Doveri,S., Donal,M. and Lee, D. (2006) Non concordance between genetics profiles of olive oil and fruit: a cautionary note to the use of DNA markers for provenance testing. *J. Agric. Food Chem.*, 54, 9221–9226.
- Besnard,G., Khadari,B., Villemur,P. et al. (2000) Cytoplasmic male sterility in the olive (Olea europaea L.). Theor. Appl. Genet., 100, 1018–1024.
- Elloumi, J., Ben-Ayed, R. and Aifa, S. (2012) An overview of olive oil biomolecules. *Curr. Biotechnol.*, 1, 115–124.

- Owen, R., Giacosa, A., Hull E., *et al.* (2000) Olive-oil consumption and health: the possible role of antioxidants. *Lancet Oncol.*, 1, 107–121.
- 5. Barranco, D. and Rallo, L. (2000) Olive cultivars in Spain. *HortTechnology*, **10**, 107–110.
- Ben-Ayed,R., Grati-Kamoun,N., Moreau,F. *et al.* (2009) Comparative study of microsatellite profiles of DNA from oil and leaves of two Tunisian olive cultivars. *Eur. Food Res. Technol.*, 229, 757–762.
- Ben-Ayed,R., Grati-Kamoun,N., Sans-grout,C. *et al.* (2012) Characterization and authenticity of virgin olive oil (*Olea europaea* L.) cultivars by microsatellite markers. *Eur. Food Res. Technol.*, 234, 263–271.
- Ben-Ayed,R., Grati-Kamoun,N. and Rebai,A. (2013) An overview of the authentication of olive tree and oil. *Compr. Rev. Food Sci. F.*, 12, 218–227.
- Busconi,M., Foroni,C., Corradi,M. *et al.* (2003) DNA extraction from olive oil and its use in the identification of the production cultivar. *Food Chem.*, 83, 127–134.
- Pafundo, S., Agrimonti, C. and Marmiroli, N. (2005) Traceability of plant contribution in olive oil by amplified fragment length polymorphisms. J. Agric. Food Chem., 53, 6995–7002.
- 11. Testolin, R. and Lain, O. (2005) DNA extraction from olive oil and PCR amplification of microsatellite markers. *Food Sci.*, 70, 108–112.
- Ben-Ayed,R., Sans-grout,C., Moreau,F. *et al.* (2014) Genetic similarity among Tunisian olive cultivars and two unknown feral olive trees estimated through SSR markers. *Biochem. Genet.*, 52, 258–268.
- Jobling, M.A. and Gill, P. (2004) Encoded evidence: DNA in forensic analysis. *Nat. Rev. Genet.*, 5, 739–751.
- 14. Istrian Olive Database (2005) http://www.iptpo.hr/iod/
- 15. Olea databases (2008) http://www.oleadb.it/
- 16. Olea EST database (2010) http://140.164.45.140/oleaestdb/
- 17. GLOBAL INVASIVE SPECIES DATABASES: http://www.issg. org/database/welcome/
- Sefc,K.M., Lopes,M.S., Mendoncxa,D. *et al.* (2000) Identification of microsatellite loci in olive (*Olea europaea*) and their characterization in Italian and Iberian olive trees. *Mol. Ecol.*, 9, 1171–1173.
- 19. Carriero,F., Fontanazza,G., Cellini, F. *et al.* (2002) Identification of simple sequence repeats (SSRs) in olive (*Olea europaea* L.). *Theor. Appl. Genet.*, **104**, 301–307.
- Cipriani,G., Marrazzo,M.T., Marconi,R. *et al.* (2002) Microsatellite markers isolated in olive (*Olea europaea* L.) are suitable for individual fingerprinting and reveal polymorphism within ancient cultivars. *Theor. Appl. Genet.*, 104, 223–228.
- 21. De La Rosa,R., James,C.M. and Tobutt,K.R. (2002) Isolation and characterization of polymorphic microsatellites in olive (*Olea europaea* L.) and their transferability to other genera in the Oleaceae. *Mol. Ecol. Notes*, **2**, 265–267.
- Diaz, A., Martin, A., Rallo, P. et al. (2006) Self-incompatibility of 'Arbequina' and 'Picual' olive assessed by SSR markers. J. Am. Soc. Hortic. Sci., 131, 250–255.
- Rallo,P., Dorado,G. and Martin, A. (2000) Development of simple sequence repeats (SSRs) in olive tree (*Olea europaea* L.). *Theor. Appl. Genet.*, 101, 984–989.

- Busconi, M., Sebastiani, L. and Corrado, F. (2006) Development of SCAR markers for germplasm characterisation in olive tree (*Olea europaea* L.). *Mol. Breed.*, 17, 59–68.
- Muzzalupo,I. (2012) Olive germplasm—Italian catalogue of olive varieties. http://dx.doi.org/10.5772/3577.
- Brescia, M.A., Alviti, G., Liuzzi, V. *et al.* (2003) Chemometrics classification of olive cultivars based on compositional data of oils. J. Am. Oil Chem. Soc., 80, 945–950.
- Mannina,L., Dugo,G., Salvo,F. *et al.* (2003) Study of the cultivar-composition relationship in Sicilian olive oils by GC, NMR, and statistical methods. *J. Agric. Food Chem.*, 51, 120–127.
- Nagy,K., Bongiorno,D., Avellone,G. *et al.* (2005) Highperformance liquid chromatography-mass spectrometry-based chemometric characterization of olive oils. *J. Chromatogr. A*, 1078, 90–97.
- 29. Cerretani,L., Bendini,A., Del Caro,A. *et al.* (2006) Preliminary characterisation of virgin olive oils obtained from different cultivars in Sardinia. *Eur. Food Res. Technol.*, **222**, 354–361.
- Muzzalupo, I., Stefanizzi, F. and Perri, E. (2009) Evaluation of olives cultivated in southern Italy by simple sequence repeat markers. *HortScience*, 44, 582–588.
- Doveri,S., Sabino Gil,F., Díaz,A. *et al.* (2008) Standardization of a set of microsatellite markers for use in cultivar identification studies in olive (*Olea europaea* L.). *Sci. Hortic.*, **116**, 367– 373.
- 32. World Catalogue of Olive Varieties (International Olive Council, 2000).
- Garcia, A., Brenes, M., Garcia, P. et al. (2003) Phenolic content of commercial olive oils. Eur. Food Res. Technol., 216, 520–525.
- Diaz, A., Martin, A., Rallo, P. et al. (2006) Self-incompatibility of 'Arbequina' and 'Picual' olive assessed by SSR markers. J. Am. Soc. Hort. Sci., 131, 250–255.
- 35. Alves, M.R., Cunha, S.C., Amaral, J.S. *et al.* (2005) Classification of PDO olive oils on the basis of their sterol composition by multivariate analysis. *Anal. Chim. Acta*, 549, 166–178.
- Manai,H., Mahjoub-Haddada,F., Oueslati,I. *et al.* (2008) Characterization of monovarietal virgin olive oils from six crossing varieties. *Sci. Hortic.*, 115, 252–260.
- Cunha,S.C., Amaral,J.S., Fernandes,J.O. *et al.* (2006) Quantification of tocopherols and tocotrienols in Portuguese olive oils using HPLC with three different detection systems. *J. Agric. Food Chem.*, 54, 3351–3356.
- Moutier, N., Pinatel, C., Martre, A. *et al.* (2004) Identification et caractérisation des variétés d'olivier cultivées en France (tome1). *Naturalia Publications, Turriers*, p. 248.
- Khlif, M. and Grati, N. (2001) Caractérisation technologique des variétés d'olivier cultivées en Tunisie. *Ezzaitouna* ISSN: 0330-6828.
- Issaoui, M., Flamini, G., Brahmi, F. *et al.* (2010) Effect of the growing area conditions on differentiation between Chemlali and Chétoui olive oils. *Food Chem.*, 119, 220–225.
- Baccouri,O., Cerretani,L., Bendini,A. *et al.* (2007) Preliminary chemical characterization of Tunisian monovarietal virgin olive oils and comparison with Sicilian ones. *Eur. J. Lipid Sci. Technol.*, 109, 1208–1217.
- 42. Guerfel,M., Ben Mansour,M., Ouni,Y. *et al.* (2012) Triacylglycerols composition and volatile compounds of virgin

olive oil from Chemlali cultivar: comparison among different planting densities. *SciWorld J.*, 2012, ID 354019.

- 43. Tena, N., Lazzez, A., Aparicio-Ruiz, R. *et al.* (2007) Volatile compounds characterizing Tunisian Chemlali and Chétoui virgin olive oils. *J. Agric. Food Chem.*, 55, 7852–7858.
- Taamalli, W., Geuna, F., Banfi, R. *et al.* (2006) Agronomic and molecular analyses for the characterisation of accessions in Tunisian olive germplasm collections. *J. Biotechnol.*, 9, 467–81.
- Dıraman,H. (2010) Characterization by chemometry of the most important domestic and foreign olive cultivars from the National Olive Collection Orchard of Turkey. *Grasas Aceites*, 61, 341– 351.
- Jibara, G., Ashtar, S., Jawhar, A. *et al.* (2007) Oil quality and morphological, phenological, bio-agronomical and molecular characterization of Syrian *Olea europaea* L. germplasm. In: Di

Terlizzi, B., Dragotta, A. and Jamal, M. (eds.). *Syrian national strategic plan for olive oil quality: final report. Bari : CIHEAM.* pp. 85–94.

- 47. Essiari, M., Zouhair, R. and Chimi, H. (2014) Contribution to the study of the typical characteristics of the virgin olive oils produced in the region of Sais (Morocco). *Olivae*, **119**, 8–21.
- 48. Haddam,M., Chimi,H., El-Antari,A. *et al.* (2014) Physicochemical characterisation and oxidative stability of olive oils produced from the 'Picholine marocaine', 'Haouzia', 'Koroneiki' and 'Arbequina' varieties in the central olive growing region of Morocco (Chaouia-Ouardigha) *Olivae*, **119**, 22–34.
- 49. Khadari,B., Charafi,J., Moukhli,A. *et al.* (2008) Substantial genetic diversity in cultivated Moroccan olive despite a single major cultivar: a paradoxical situation evidenced by the use of SSR loci. *Tree Genet. Genomes*, 4, 213–221.