

# Barcoding the major Mediterranean leguminous crops by combining universal chloroplast and nuclear DNA sequence targets

P. Madesis<sup>1\*</sup>, I. Ganopoulos<sup>2\*</sup>, P. Ralli<sup>3</sup> and A. Tsaftaris<sup>1,2</sup>

<sup>1</sup>Institute of Agrobiotechnology, Centre for Research and Technology Hellas, Thessaloniki, Greece

<sup>2</sup>Department of Genetics and Plant Breeding, School of Agriculture, Aristotle University of Thessaloniki, Thessaloniki, Greece

<sup>3</sup>Greek Gene Bank, Agricultural Research Centre of Northern Greece, National Agricultural Research Foundation, Thessaloniki, Greece

\*These authors contributed equally to this study.

Corresponding author: A. Tsaftaris

E-mail: [tsaft@certh.gr](mailto:tsaft@certh.gr)

Genet. Mol. Res. 11 (3): 2548-2558 (2012)

Received October 27, 2011

Accepted May 5, 2012

Published July 10, 2012

DOI <http://dx.doi.org/10.4238/2012.July.10.10>

**ABSTRACT.** The ability to discriminate all species is the ultimate target in barcoding. The Mediterranean basin is a center of origin for legumes and thus they have played a key role in feeding the Mediterranean population. It is also a region with important protected designation of origin and protected geographical indication legumes that provide income in rural areas. We evaluated the use of two chloroplast regions, *trnL* and *rpoCl*, and a nuclear internal transcriber region, ITS2, for their efficiency to barcode the main Mediterranean leguminous crops. Twenty-five legume species were studied. Plant material of pasture and legumes was obtained from the Greek GenBank and the Fodder Crops and Pastures Institute (National Agricultural Research Foundation). DNA was extracted with the Qiagen DNeasy plant mini-kit and PCR amplification was performed using the Kapa Taq DNA polymerase using primers amplifying the chloroplast *trnL* and *rpoCl* regions or the

nuclear region ITS2. PCR products were sequenced and the sequences were aligned using CLUSTAL W. Species identification based on the sequence similarity approach was performed using the GenBank database. In order to evaluate intraspecific and interspecific divergence in legumes we used Molecular Evolutionary Genetics Analysis 5 and for pairwise Kimura 2-parameter distance calculations for all 3 DNA regions (2 chloroplast regions, *trnL* and *rpoC1*, and the nuclear region ITS2). Four tree-based methods (neighbor joining and maximum parsimony, maximum likelihood, and Bayesian inference analyses) were used to exhibit the molecular identification results to represent differences as an unrooted dendrogram. Additionally, the sequence character-based method was used with DnaSP and the information from each site was treated as a character to distinguish the species from one another. The DNA regions *trnL* and ITS2 successfully (100%) discriminated the Mediterranean crop legume species used, while *rpoC1* identified only 72% of them. Furthermore, the use of the *trnL* region enabled the discrimination of even very closely related species, like *Phaseolus lunatus* and *P. coccineus* or *Vicia faba* subsp *major* with *V. faba* subsp *minor*, which are so closely related that even in NCBI they were both referred as *Phaseolus vulgaris* and *V. faba*, respectively. We conclude that *trnL* and ITS2 are efficient DNA barcoding target regions in order to discriminate Mediterranean leguminous crops and provide a reliable and efficient tool for the scientific, agricultural and industrial community.

**Key words:** Barcoding; *trnL*; *rpoC1*; ITS2; Legumes; Mediterranean

## INTRODUCTION

Leguminous crops belong to the Fabaceae family and are second only to cereals in their importance for human nutrition. The Mediterranean basin is a center of origin for legumes (Perrino, 1988). They have played an important role in human nutrition in the region since ancient times, increasing not only the quantity but also the quality of cereal-based food for the population. Legumes constitute the main protein component of the Mediterranean diet (Bromfeild et al., 2001). They are used in the human diet because they are low in calories but rich in flavor and in nutrients such as protein, calcium, fiber, and vitamins. In addition, legumes are rich in soluble fibers, facilitating digestion and preventing cancer (Muñoz de Chávez and Chávez, 1998). Moreover, they are important seed- and forage-producing crops for animal feed. Equally important is the role of legumes to sustainable agriculture, because the agricultural practice of crop rotation with legumes enriches the soil with nitrogen via the symbiosis of legumes with nitrogen-fixing bacteria (Arianoutsou and Thanos, 1996; Sprent, 2001; Velazquez et al., 2010). Legumes are also an excellent source of medicinals, nutraceuticals, tannins, gums, insecticides, resins, varnish, paints, dyes, proteins, and eco-friendly by-products such as soy diesel. Finally, 3 forage crops, *Medicago truncatula*, *Lotus japonicus*, and *Trifolium pratense*, are model legumes for phylogenetic studies and genome sequencing (Graham and Vance, 2003).

Mediterranean crop legumes are also important to local economies, as some of them, such as “Fava Santorinis” (*Lathyrus clymenum*), are protected designation of origin (PDO; Article 6[2] of Regulation [EC] No. 510/2006) or protected geographical indication products, such as “Fasolia Prespon” (*Phaseolus vulgaris*), or even ornamental plants such as *Lathyrus odoratus*, designations that guarantee an added value to these products.

The identification of the Mediterranean leguminous crops using a method based on their morphological characteristics has proved tricky and even impossible when working with commercial products (Smartt, 1980). The use of a DNA-based method would provide accurate information and facilitate the discrimination of the species. Having a robust and reliable system to discriminate species to secure the quality of certain products is important.

Barcoding is a method of identifying species using short DNA sequences. The definitive goal is to identify a region or a combination of regions capable of discriminating all species (Kress et al., 2005; Hollingsworth et al., 2009; Chen et al., 2010). The barcoding method has been extremely useful in species identification, cryptic species identification, biodiversity studies, forensic analysis, and phylogenetics (Ronning et al., 2005; Ward et al., 2005). In animals, cytochrome oxidase 1 of the mitochondrial cytochrome oxidase 1 gene is the region of choice (Hebert et al., 2003; Hebert et al., 2004). In plants, the most favorable choices are chloroplast DNA regions, as they have been used as a means to identify species (Kress et al., 2005; Hollingsworth et al., 2009). The sequences used, known as DNA barcodes, are usually short (300-800 bp) (Kress et al., 2005). Although chloroplast DNA barcoding is mainly used to identify plant species, its application could be extended to the food industry, evolution studies, and forensics (Ferri et al., 2009). Various regions of the plastid genome have been proposed to serve as DNA barcodes, including the *rbcL*, *matK*, *rpoB*, and *C* genes, the non-coding spacers *atpF-atpH*, *trnH-psbA* and *psbK-psbI trnL-F*, the *trnL* (UAA) intron, and the internal transcribed spacer (ITS) 2 region of nuclear ribosomes. Research groups have their preferred plant barcode regions for study, but no consensus has emerged on the use of a standard region (Kress et al., 2005; Taberlet et al., 2007; Edwards et al., 2008; Hollingsworth et al., 2009; Yao et al., 2010).

Wojciechowski et al. (2004) have used the *matK* region to construct the phylogeny of legumes. Asmussen and Liston (1998) have used *rpoCI* and *psbA trnHGUG ndhF* and their intergenic regions to construct the phylogeny of the genus *Lathyrus*. Kenicer et al. (2005), also working with *Lathyrus*, used *trnL-F*, *trnS-G*, and ITS to study systematics and biogeography. Nuclear sequences have also been used to barcode legume members of the Fabaceae family: Gao et al. (2010) have used the ITS2 region to identify medicinal plants, and Gao and Chen (2009) have investigated the potential of 4 coding chloroplast regions (*rpoB*, *rpoCI*, *rbcL*, and *matK*) and 2 non-coding nuclear regions (ITS1 and ITS2) for use as barcodes of the medicinal plants in Fabaceae by comparing DNA barcoding sequences. The results indicated that the efficiency of amplification for 6 candidate DNA barcodes ranges from 100% (ITS2) to 93% (*matk*) (Gao and Chen, 2009). Gao et al. (2011) have also used *matK* to barcode members of the Fabaceae family.

We evaluated the use of 2 chloroplast regions, *trnL* and *rpoCI*, and a nuclear region, ITS2, for their efficiency in barcoding the main Mediterranean leguminous crops. *trnL* and ITS2 successfully discriminated the species used, whereas *rpoCI* identified only 72% of them.

## MATERIAL AND METHODS

### Plant material

Plant material of pasture and legumes was obtained from the Greek GenBank and the Fodder Crops and Pastures Institute (National Agricultural Research Foundation; Table 1). The species selected were important Greek legumes used as crops for human nutrition or animal feed. Isolation of DNA was performed with 0.1 g seeds or leaves (from 10 plants) prepared as a starting material of fine powder using a Qiagen DNeasy plant mini kit (Qiagen, Crawley, UK) according to manufacturer instructions. The DNA concentration was estimated using standard spectrophotometric methods at 260 and 280 nm ultraviolet lengths with an Eppendorf BioPhotometer. DNA integrity was assessed through gel electrophoresis on a 0.8% agarose gel. Samples were then diluted to 20 ng/ $\mu$ L for experimentation.

### Polymerase chain reaction (PCR) amplification

PCR amplification was performed in a total volume of 25  $\mu$ L in an MJ research thermocycler. The reaction mixture contained 20 ng genomic DNA, 1X PCR buffer, 2.5 mM MgCl<sub>2</sub>, 0.2 mM deoxyribonucleotide triphosphate, 300 nM forward and reverse primers (Supplementary Table 1), and 0.5 U Kapa Taq DNA polymerase (Kapa Biosystems, Foster City, USA). An initial denaturing step of 94°C for 3 min was followed by 30 cycles of 94°C for 20 s, 54°C for 40 s, and 72°C for 20-40 s and a final extension step of 72°C for 2 min.

### Sequence analysis

PCR products were directly sequenced in 2 directions of each fragment with a Big Dye terminator v3.1 Cycle sequencing kit (PE Applied Biosystems, Foster City, CA, USA) in an automated ABI 3730 sequencer (PE Applied Biosystems). The sequences were aligned using the CLUSTAL W program.

### Bioinformatics

Species identification based on the sequence similarity approach was performed using the GenBank database. To test the efficiency of DNA barcoding as a species identification tool, a blind sampling test was conducted in which samples whose identities were known only to the submitting individual were selected and sequenced. We used Molecular Evolutionary Genetics Analysis 5.1 (Tamura et al., 2011) for pairwise Kimura 2-parameter distance calculations for all 3 DNA regions (2 chloroplast regions, *trnL* and *rpoCl*, and the nuclear region ITS2) to evaluate intraspecific and interspecific divergence in legumes. Insertions and deletions (indels) were coded using the simple indel coding method of Simmons and Ochoterena (2000). Four tree-based methods were used to exhibit the molecular identification results to represent differences as an unrooted dendrogram. Neighbor joining and maximum parsimony, maximum likelihood, and Bayesian inference analyses determinations were performed in Phylogenetic Analysis Using Parsimony v. 4.0b10 (Swofford, 2003), PhyML v. 2.4.4 (Guindon and Gascuel, 2003), and MrBayes v. 3.1.2 (Ronquist and Huelsenbeck, 2003), respectively.

Additionally, the sequence character-based method (Rach et al., 2008) was used with DnaSP (Rozas et al., 2003), and the information from each site was treated as a character to distinguish the species from one another.

## RESULTS

Table 1 shows the results of the evaluation of the 3 DNA barcoding marker regions we used in our study. All 25 selected Mediterranean leguminous crop species showed good amplification success. All of the loci analyzed exhibited high PCR success with standard primers, reaching a success rate near 100%. All of the PCR products corresponding to the 3 DNA markers were successfully sequenced, and high-quality bidirectional sequences were obtained. Some sequencing problems were encountered with the ITS2 region for 2 samples (*Medicago lupulina* and *Trifolium repens*; see Table 1), partially owing to a high frequency of mononucleotide repeats that disrupted individual sequencing reads. The ITS2 region also included a partial sequence from nuclear 5.8S ribosomal DNA.

**Table 1.** Evaluation of three DNA barcoding marker regions, two chloroplast regions (*trnL* and *rpoCl*) and a nuclear region ITS2.

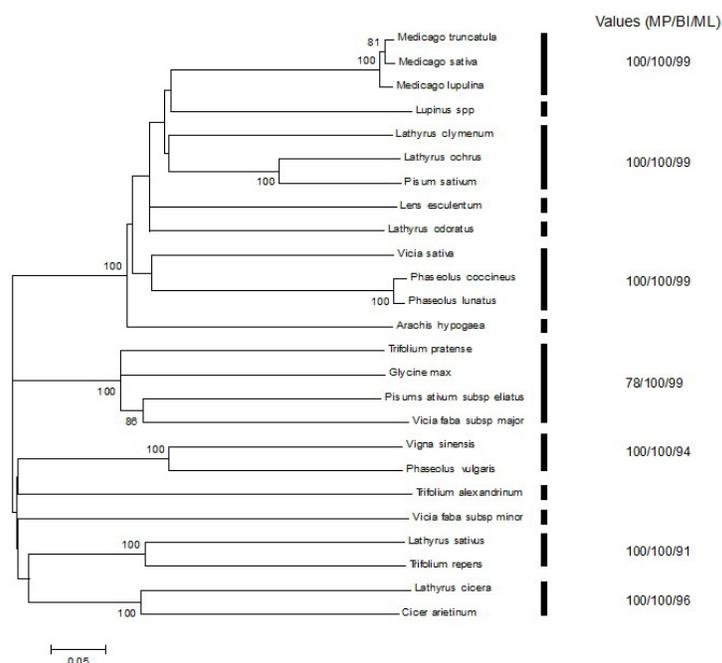
DNA region	<i>rpoCl</i>	<i>trnL</i>	ITS2
Universal ability to primer	Yes	Yes	Yes
Percentage PCR success	100	100	100
Percentage sequencing success	100	100	94
Aligned sequence length (bp)	510	734	510
Average InDel length	2.366	1.965	2.367
No. variable sites	151	83	151
Parsimony-informative sites	112	48	100
Distribution of variable sites	Di and S	I and D	I and D
No. samples species (individuals)	25 (100)	25 (100)	25 (100)
Interspecific distance mean (range), %	1.109 (0-2.176)	1.573 (0-2.676)	1.759 (0-3.275)
Intraspecific distance mean (range), %	-	-	0.006 (0-0.321)
Ability to discriminate	18/25	25/25	25/25
%	72	100	100

Di = dispersive; S = sparse; I = intensive; D = dense.

More specifically, we obtained 100 *rpoCl*, 100 *trnL*, and 100 ITS2 sequences from 25 leguminous species, resulting in 300 new sequences (GenBank accession Nos. appear in Supplementary Table 2). With regard to universality of primers and success of sequence amplification, the proportion at each of the 3 regions was 100% for the chloroplast regions *trnL* and *rpoCl* and 94% for the nuclear region ITS2 (see Table 1). The *trnL* matrix aligned sequence length was 734 bp; the distribution of the 48 parsimony informative sites and 83 variable sites was intensive and dense across the matrix, with included indels (average length, 1.965). The *rpoCl* matrix had 510 bp and indels with an average length 2.366 bp; the distribution of 112 parsimony informative sites and 151 variable sites was dispersive and sparse across the matrix. For the ITS2 matrix, the aligned sequence length was 510 bp; the distribution of 100 parsimony informative sites and 151 variable sites was intensive and dense across the matrix, and there were indels with an average length of 2.367 bp.

The mean sequence divergences in legume species were 1.173% (*trnL*), 1.109% (*rpoCl*), and 1.759% (ITS2). For ITS2, the mean value of genetic distance between species

was more than 10 times that of the intraspecific distance (0.006%). The discriminating power of the 3 markers at the species level was 72% (*rpoC1*), 100% (*trnL*), and 100% (ITS2). In general, ITS2 ranked first in divergence values among species, followed by *trnL* (see Table 1). On the contrary, *rpoC1* sequences showed lower sequence divergence. Figure 1 shows a graphical representation of the differences found among species as an unrooted dendrogram.



**Figure 1.** Unrooted dendrogram (NJ) based on the *trnL* sequence matrix for 25 Mediterranean crop legume species. The support values (bootstrap or Bayesian posterior probabilities, in percentage) with three different tree-based methods [maximum parsimony (MP), Bayesian inference analysis (BI), maximum likelihood (ML)].

Using the *trnL* region, we searched the National Center for Biotechnology Information database using the nucleotide Basic Local Alignment Search Tool (nBLAST). All sequences showed 99-100% sequence identity with corresponding sequences in the database, with the exception of *Phaseolus lunatus*, *Phaseolus coccineus*, *Vicia faba ssp major*, and *V. faba ssp minor* (Table 2).

The chloroplast *trnL* barcoding region allowed the discrimination of all 25 crop legumes, but the *rpoC1* region allowed the discrimination of only 78% of the species. Moreover, the nuclear ITS2 region also allowed the discrimination of the legume species used in this study but had only 94% sequence success. Thus, the combined use of at least 2 of the barcoding regions allowed the discrimination of all Mediterranean crop legumes used in this study.

Table 3 lists the specific informative and discriminatory sites. *P. lunatus* has 2 diagnostic sites, 234 G and 288 G, whereas *P. coccineus* has 288 C and 327 C. *Medicago sativa* has 2 informative sites, 289 A and 307 C, and *M. truncatula* has one at 289 A. *Pisum sativum ssp eliatum*, which is closely related to *P. sativum*, has one informative site at 303 A. Another interesting informative site occurs between *V. faba ssp major* and *V. faba ssp minor* at 539 G.

**Table 2.** Summary of identification based on each species consensus barcoded sequence using BLASTN search from GenBank.

Species studied	Species identification ( <i>trnL</i> )	% Maximum identity
<i>Lathyrus sativus</i>	<i>Lathyrus sativus</i>	100
<i>Lathyrus clymenum</i>	<i>Lathyrus clymenum</i>	99
<i>Lathyrus cicera</i>	<i>Lathyrus cicera</i>	100
<i>Lathyrus ochrus</i>	<i>Lathyrus ochrus</i>	99
<i>Lathyrus odoratus</i>	<i>Lathyrus odoratus</i>	99
<i>Pisum sativum</i> subsp <i>eliatius</i>	<i>Pisum sativum</i>	100
<i>Pisum sativum</i>	<i>Pisum sativum</i>	100
<i>Lupinus</i> spp	<i>Lupinus albus</i>	100
<i>Vicia faba</i> subsp <i>major</i> *	<i>Vicia faba</i>	100
<i>Vicia faba</i> subsp <i>minor</i> *	<i>Vicia faba</i>	100
<i>Cicer arietinum</i>	<i>Cicer arietinum</i>	100
<i>Glycine max</i>	<i>Glycine max</i>	99
<i>Lens esculentum</i>	<i>Lens esculentum</i>	100
<i>Phaseolus coccineus</i>	<i>Phaseolus coccineus</i>	100
<i>Vigna sinensis</i>	<i>Vigna sinensis</i>	100
<i>Phaseolus vulgaris</i>	<i>Phaseolus vulgaris</i>	100
<i>Phaseolus lunatus</i> *	<i>Phaseolus vulgaris</i>	100
<i>Vicia sativa</i>	<i>Vicia sativa</i>	99
<i>Trifolium pratense</i>	<i>Trifolium pratense</i>	100
<i>Trifolium repens</i>	<i>Trifolium repens</i>	100
<i>Trifolium alexandrinum</i>	<i>Trifolium alexandrinum</i>	99
<i>Medicago truncatula</i>	<i>Medicago truncatula</i>	100
<i>Medicago lupulina</i>	<i>Medicago lupulina</i>	100
<i>Medicago sativa</i>	<i>Medicago sativa</i>	99
<i>Arachis hypogaea</i>	<i>Arachis hypogaea</i>	100

**Table 3.** Character based DNA database for Mediterranean crop legume species from the *trnL* region.

Species	Position															
	93	129	154	165	187	234	288	289	292	294	303	307	327	330	339	539
<i>Arachis hypogaea</i>	A	C	C	G	G	G	C	G	T	T	C	T	T	C	G	A
<i>Cicer arietinum</i>	T	T	T	G	T	A	C	T	C	T	G	T	T	T	T	-
<i>Glycine max</i>	C	C	T	A	C	G	C	C	T	A	G	C	T	T	G	G
<i>Lathyrus cicera</i>	T	C	T	G	T	A	C	T	T	T	G	T	T	T	T	A
<i>Lathyrus clymenum</i>	A	C	T	G	T	A	C	C	T	A	G	T	T	T	T	A
<i>Lathyrus ochrus</i>	T	T	T	G	T	A	C	C	T	A	G	T	T	T	T	A
<i>Lathyrus odoratus</i>	T	C	T	G	C	G	C	T	T	T	G	T	T	T	T	G
<i>Lathyrus sativus</i>	T	C	T	G	T	A	C	T	T	T	G	C	T	T	T	G
<i>Lens esculentum</i>	T	C	T	G	T	A	C	T	C	T	C	T	T	T	T	G
<i>Lupinus</i> spp	C	T	C	G	T	G	T	T	C	C	C	C	T	T	T	G
<i>Medicago lupulina</i>	T	T	T	G	T	A	C	T	T	T	G	T	T	C	T	A
<i>Medicago sativa</i>	T	T	T	G	T	A	C	A	T	T	G	C	T	C	T	G
<i>Medicago truncatula</i>	T	T	T	G	T	A	C	A	T	T	G	T	T	C	T	G
<i>Phaseolus coccineus</i>	C	T	T	A	C	A	C	G	T	C	G	C	C	T	A	A
<i>Phaseolus lunatus</i>	C	T	T	A	C	G	G	G	T	C	G	C	T	T	A	A
<i>Phaseolus vulgaris</i>	C	T	T	A	C	A	G	G	T	C	G	C	T	T	A	A
<i>Pisum sativum</i>	T	C	T	G	T	A	T	T	T	T	G	T	T	T	T	A
<i>Pisum sativum</i> subsp <i>eliatius</i>	T	C	T	G	T	A	T	T	T	T	A	T	T	T	T	G
<i>Trifolium alexandrinum</i>	T	T	T	G	T	A	C	T	T	T	G	T	T	T	T	A
<i>Trifolium pratense</i>	T	T	T	G	C	A	C	T	T	T	G	T	T	C	G	A
<i>Trifolium repens</i>	T	T	C	G	C	A	C	T	T	T	G	T	T	C	T	A
<i>Vicia faba</i> subsp <i>major</i>	T	T	T	G	T	A	C	T	T	T	G	T	T	T	T	-
<i>Vicia faba</i> subsp <i>minor</i>	T	T	T	G	T	A	C	T	T	T	G	T	T	T	T	G
<i>Vicia sativa</i>	T	T	T	G	T	A	C	T	T	T	A	T	T	T	T	-
<i>Vigna unguiculata</i>	C	T	T	G	C	G	G	T	T	T	G	T	C	T	G	G

Character states (nucleotides) at 16 selected positions (ranging from 93 to 539) are shown. Bold letters show important diagnostic character sites. (-) = indel site.

## DISCUSSION

Information on plastid genome sequences is of great importance for barcoding land plants. Sequence information plays a key role in identifying and discriminating species, thus allowing the certification of commercial and, especially, PDO products. Barcoding regions should be universal, should have very few or no ambiguous calls bi-directionally, and should allow the discrimination of most species. Many regions of the plastid genome have been used for DNA barcoding to date, usually in combinations of 2 or 3 (Kress et al., 2005; Chase et al., 2007; Hollingsworth et al., 2009, 2011).

The capability of a DNA region to discriminate all species is the ultimate test. We used the chloroplast *trnL* (Taberlet et al., 2007) and *rpoCI* regions and the nuclear ITS2 region to barcode the main Mediterranean leguminous crop species. The *trnL* barcoding region successfully discriminated all crop species used. This region has been used extensively because it incorporates all of the desirable characteristics of a barcoding region and includes a small region of the P6 loop, which is very useful, especially for small DNA fragments (degraded DNA, ancient DNA, etc.) (Taberlet et al., 2007; Hollingsworth et al., 2011). The *trnL* intron is the only group I intron in the chloroplast genome and has a conserved secondary structure, especially at the regions that flank the *trnL* exons (Bakker et al., 2000). The central part of the intron includes the P6 and P8 stem-loop regions, which account for most of the sequence variation. The *trnL* region has been successfully exploited in barcoding and food authentication (Taberlet et al., 2007; Spaniolas et al., 2010). Although this barcoding region successfully discriminated all 25 legume crop species, 2 samples (*M. lupulina* and *T. repens*; see Table 1) did not perform as well as others in the sequence reaction, possibly owing to the presence of a high frequency of mononucleotide repeats that disrupt individual sequencing reads.

The ITS region has been proposed by others as a suitable barcode (Kress et al., 2005). This region can be problematic in some species, however, owing to gene duplication and the presence of paralogues and pseudogenes (Chase et al., 2007). When *rpoCI* was used alone, it discriminated only 72% of the species, which might be expected owing to the generally low discrimination efficiency of the locus (Hollingsworth et al., 2011). Gao and Chen (2009) have investigated the potential of 4 coding chloroplast regions (*rpoB*, *rpoCI*, *rbcl*, and *matK*), and 2 noncoding nuclear regions (ITS, ITS2) as barcodes for the medicinal plants in the Fabaceae family by comparing DNA barcoding sequences. The results indicated that the efficiency of amplification for the 6 candidate DNA barcodes ranged from 100% (ITS2) to 93% (*matK*).

The sequence data obtained for *trnL* and ITS2 were used to obtain a dendrographic representation of the differences between the species used. Although it is not a phylogenetic tree in the strict sense, it can still be used to discriminate both *P. coccineus* and *P. lunatus* and *V. faba* and *V. sativa*. The placement of the other species used in this study in the phylogenetic tree cannot be directly compared with placements made in other studies owing to the methods used to construct the tree. Of interest, however, is that when we searched the National Center for Biotechnology Information database with *V. faba* ssp *major* and *minor*, we obtained 100% identity with *V. faba*. The same applied for *P. lunatus*, which was identified as *P. vulgaris*. This result may occur because the sequences in the database have not been annotated or the sequences of the specific species are absent from the database. It could also be because the sequence is 100% identical and not in the full length of the sequence used to interrogate the database but in the longest stretch of it.

In addition, *trnL* discriminated *Lathyrus* species but, surprisingly, also discriminated *V. faba* ssp *major* from *V. faba* ssp *minor*, which are considered to be very closely related. Our analysis also shows that *P. sativum* can be discriminated as *P. sativum* ssp *eliatum* and *L. odoratus* (ornamental *Pisum*). *Pisum* species have been placed between *Vicia* and *Lathyrus* species in accordance with Smykal et al. (2010), although in our study, *Pisum* species were found to be closely related to those of *Lathyrus*. The presence of indels in *trnL*, *rpoCl*, and ITS2 is a prerequisite for barcoding species. Indels in general increase the discriminatory power of a region and in the case of *V. faba* ssp *major* and *minor*, can discriminate even species considered identical (Kress et al., 2005).

Mediterranean crop legumes have an important role in human diets, animal agriculture, and protection of soil (enriching it either with nitrogen as plants or as a green manure and thus contributing to sustainable agriculture). Barcoding Mediterranean legumes offers important information to the scientific community and the productive sector because it is a useful tool for the identification of these crops. Moreover, as legumes are used in various processed products, which could include PDO or protected geographical indication species or products, the use of barcoding, especially of *trnL*, could yield valuable information on product content and enable the identification of product composition. A single locus, *trnL* or ITS, is suitable to barcode all of the main Mediterranean crop legumes, whereas *rpoCl* cannot be used alone for this purpose.

## REFERENCES

- Arianoutsou M and Thanos CA (1996). Legumes in the fire-prone Mediterranean regions: an example from Greece. *Int. J. Wildland* 6: 77-82.
- Asmussen C and Liston A (1998). Chloroplast DNA characters, phylogeny, and classification of *Lathyrus* (Fabaceae). *Am. J. Bot.* 85: 387.
- Bakker FT, Culham A, Gomez-Martinez R, Carvalho J, et al. (2000). Patterns of nucleotide substitution in angiosperm cpDNA *trnL* (UAA)-*trnF* (GAA) regions. *Mol. Biol. Evol.* 17: 1146-1155.
- Bromfeild S, Butler G and Barran LR (2001). Temporal effect on the composition of a population of *Sinorhizobium meliloti* associated with *Medicago sativa* and *Medicago alba*. *Can. J. Microbiol.* 47: 567-573.
- Chase MW, Cowan RS, Hollingsworth PM and van den Berg C (2007). A proposal for a standardised protocol to barcode all land plants. *Taxon* 56: 295-299.
- Chen S, Yao H, Han J, Liu C, et al. (2010). Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. *PLoS One* 5: e8613.
- Edwards D, Horn A, Taylor D and Savolainen V (2008). DNA barcoding of a large genus, *Aspalathus* L. (Fabaceae). *Taxon* 57: 1317-1314.
- Ferri G, Alu M, Corradini B and Beduschi G (2009). Forensic botany: species identification of botanical trace evidence using a multigene barcoding approach. *Int. J. Legal Med.* 123: 395-401.
- Gao T and Chen SL (2009). Authentication of the medicinal plants in Fabaceae by DNA barcoding technique. *Planta Med.* 75: 417.
- Gao T, Yao H, Song J, Liu C, et al. (2010). Identification of medicinal plants in the family Fabaceae using a potential DNA barcode ITS2. *J. Ethnopharmacol.* 130: 116-121.
- Gao T, Sun Z, Yao H, Song J, et al. (2011). Identification of Fabaceae plants using the DNA barcode matK. *Planta Med.* 77: 92-94.
- Graham PH and Vance CP (2003). Legumes: importance and constraints to greater use. *Plant Physiol.* 131: 872-877.
- Guindon S and Gascuel O (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52: 696-704.
- Hebert PD, Cywinska A, Ball SL and deWaard JR (2003). Biological identifications through DNA barcodes. *Proc. Biol. Sci.* 270: 313-321.
- Hebert PD, Penton EH, Burns JM, Janzen DH, et al. (2004). Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proc. Natl. Acad. Sci. U. S. A.* 101: 14812-14817.

- Hollingsworth PM, Forrest LL, Spouge JL and Hajibabaei M (2009). A DNA barcode for land plants. *Proc. Nat. Acad. Sci. U. S. A.* 106: 12794-12797.
- Hollingsworth PM, Graham SW and Little DP (2011). Choosing and using a plant DNA barcode. *PLoS One* 6: e19254.
- Kenicer GJ, Kajita T, Pennington RT and Murata J (2005). Systematics and biogeography of *Lathyrus* (Leguminosae) based on internal transcribed spacer and cpDNA sequence data. *Am. J. Bot.* 92: 1199-1209.
- Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, et al. (2005). Use of DNA barcodes to identify flowering plants. *Proc. Natl. Acad. Sci. U. S. A.* 102: 8369-8374.
- Muñoz de Chávez M and Chávez A (1998). Diet that prevents cancer: recommendations from the American Institute for cancer research. *Int. J. Cancer Suppl.* 11: 85-89.
- Perrino P (1988). The diversity in Vavilov's mediterranean gene center. *Kulturpflanze* 36: 85-105.
- Rach J, DeSalle R, Sarkar IN, Schierwater B, et al. (2008). Character-based DNA barcoding allows discrimination of genera, species and populations in Odonata. *Proc. R. Soc. Lond. B Biol. Sci.* 275: 237-247.
- Ronning SB, Rudi K, Berdal KG and Holst-Jensen A (2005). Differentiation of important and closely related cereal plant species (Poaceae) in food by hybridization to an oligonucleotide array. *J. Agric. Food Chem.* 53: 8874-8880.
- Ronquist F and Huelsenbeck JP (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572-1574.
- Rozas J, Sánchez-DelBarrio JC, Messeguer X and Rozas R (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19: 2496-2497.
- Simmons MP and Ochoterena H (2000). Gaps as characters in sequence-based phylogenetic analyses. *Syst. Biol.* 49: 369-381.
- Smartt J (1980). Evolution and evolutionary problems in food legumes. *Econom. Bot.* 34: 219-235.
- Smykal P, Kenicer G, Flavell AJ and Corader J (2010). Phylogeny, phylogeography and genetic diversity of *Pisum* genus. *Plant Genet. Res.* 9: 4-18.
- Spaniolas S, Bazakos C, Spano T and Zoghby C (2010). The potential of plastid *trnL* (UAA) intron polymorphisms for the identification of the botanical origin of plant oils. *Food Chem.* 122: 850-856.
- Sprent JI (2001). Nodulation in Legumes. Kew Publishing, Richmond.
- Swofford DL (2003). PAUP\*. Phylogenetic Analysis Using Parsimony (\*and Other Methods). Version 4. Sinauer Associates Inc. Publishers, Sunderland.
- Taberlet P, Coissac E, Pompanon F and Gielly L (2007). Power and limitations of the chloroplast *trnL* (UAA) intron for plant DNA barcoding. *Nucleic Acids Res.* 35: e14.
- Tamura K, Peterson D, Peterson N and Stecher G (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28: 2731-2739.
- Velazquez ER, Silva L and Alvaro P (2010). Legumes: a healthy and ecological source of flavonoids. *Curr. Nutr. Food Sci.* 6: 109-144.
- Ward J, Peakall R, Gilmore SR and Robertson J (2005). A molecular identification system for grasses: a novel technology for forensic botany. *Forensic Sci. Int.* 152: 121-131.
- Wojciechowski MF, Lavin M and Sanderson MJ (2004). A phylogeny of legumes (Leguminosae) based on analysis of the plastid matK gene resolves many well-supported subclades within the family. *Am. J. Bot.* 91: 1846-1862.
- Yao H, Song J, Liu C, Luo K, et al. (2010). Use of ITS2 region as the universal DNA barcode for plants and animals. *PLoS One* 5.

## SUPPLEMENTARY MATERIAL

**Supplementary Table 1.** Primers used in this study.

Region	Primer	Reference
<i>trnL</i> -F	5'-CGAAATCGGTAGACGCTACG-3'	Taberlet et al., 2007
<i>trnL</i> -R	5'-GGGGATAGAGGGACTTGAAC-3'	Taberlet et al., 2007
<i>rpoC1</i> -F	5'-GGCAAAGAGGGAAGATTTTCG-3'	Hollingsworth et al., 2009
<i>rpoC1</i> -R	5'-CCATAAGCATATCTTGAGTTGG-3'	Hollingsworth et al., 2009
ITS2-F	5'-ATGCGATACCTGGTGTGAAT-3'	Chen et al., 2010
ITS2-R	5'-GACGCTTCTCCAGACTACAAT-3'	Chen et al., 2010

**Supplementary Table 2.** GenBank accession Nos. of samples subjected to the species identification test using the ITS2, *rpoC1* and *trnL* regions.

Species	Source	ITS2 Genbank accession	<i>rpoC1</i> Genbank accession	<i>trnL</i> Genbank accession
<i>Lathyrus clymenum</i>	NAGREF	JN617185	JN617135	JN617161
<i>Lathyrus cicera</i>	NAGREF	JN617186	JN617136	JN617162
<i>Lathyrus sativus</i>	NAGREF	JN617187	JN617137	JN617160
<i>Lathyrus ochrus</i>	NAGREF	JN617188	JN617138	JN617163
<i>Pisum sativum</i> subsp <i>eliatum</i>	NAGREF	JN617189	JN617139	JN617164
<i>Pisum sativum</i>	NAGREF	JN617190	JN617140	JN617165
<i>Lupinus</i> spp	NAGREF	JN617191	JN617141	JN617166
<i>Vicia faba</i> subsp <i>major</i>	NAGREF	JN617192	JN617142	JN617167
<i>Vicia faba</i> subsp <i>minor</i>	NAGREF	JN617193	JN617143	JN617168
<i>Cicer arietinum</i>	NAGREF	JN617194	JN617144	JN617169
<i>Glycine max</i>	NAGREF	JN617195	JN617145	JN617170
<i>Lens esculentum</i>	NAGREF	JN617196	JN617146	JN617171
<i>Phaseolus coccineus</i>	NAGREF	JN617197	JN617147	JN617172
<i>Vigna sinensis</i>	NAGREF	JN617198	JN617148	JN617173
<i>Phaseolus vulgaris</i>	NAGREF	JN617199	JN617149	JN617174
<i>Phaseolus lunatus</i>	NAGREF	JN617200	JN617150	JN617175
<i>Vicia sativa</i>	NAGREF	JN617201	JN617151	JN617176
<i>Lathyrus odoratus</i>	NAGREF	JN617202	JN617152	JN617177
<i>Trifolium pratense</i>	NAGREF	JN617203	JN617153	JN617178
<i>Trifolium repens</i>	NAGREF	JN617204	JN617154	JN617179
<i>Trifolium alexandrinum</i>	NAGREF	JN617205	JN617155	JN617180
<i>Medicago truncatula</i>	NAGREF	JN617206	JN617156	JN617181
<i>Medicago lupulina</i>	NAGREF	JN617207	JN617157	JN617182
<i>Medicago sativa</i>	NAGREF	JN617208	JN617158	JN617183
<i>Arachis hypogaea</i>	NAGREF	JN617209	JN617159	JN617184