

# A PCR based SNPs marker for specific characterization of English walnut (*Juglans regia* L.) cultivars

Loredana F. Ciarmiello · Pasquale Piccirillo ·  
Giovanni Pontecorvo · Antonio De Luca ·  
Ioannis Kafantaris · Pasqualina Woodrow

Received: 3 May 2010 / Accepted: 11 June 2010 / Published online: 25 June 2010  
© Springer Science+Business Media B.V. 2010

**Abstract** English walnut (*Juglans regia* L.) is the most economically important species from all the 21 species belonging to the genus *Juglans* and is an important and healthy food as well as base material for timber industry. The aim of this study was to develop a simple technique for specific characterization of English walnut using DNA method. The first and second internal transcribed spacers (ITS1 and ITS2) as well as the intervening 5.8S coding region of the rRNA gene for 18 cultivars of *J. regia* L. isolated from different geographic origins were characterized. The size of the spacers sequences ranged from 257 to 263 bases for ITS1 and from 217 to 219 bases for ITS2. Variation of GC contents has also been observed and scored as 55–56.7 and 57.1–58.9% for ITS1 and ITS2, respectively. This data exhibited the presence of polymorphism among cultivars. Alignment of the ITS1-5.8S-ITS2 sequences from 18 walnut cultivars showed that there were 244 single nucleotide polymorphisms (SNPs) and 1 short insertion–deletion (indel) at 5' end ITS1. Amplification refractory mutation system strategy was successfully applied to the SNP markers of the ITS1 and ITS2 sequences for the fingerprinting analysis of 17 on 18 walnut cultivars. The prediction of ITS1 and ITS2 RNA secondary structure from each cultivar was improved by detecting key functional elements shared by all sequences in the alignments. Phylogenetic analysis of the ITS1-5.8S-

ITS2 region clearly separated the isolated sequences into two clusters. The results showed that ITS1 and ITS2 region could be used to discriminate these walnut cultivars.

**Keywords** *Juglans regia* L. · rDNA internal transcribed spacer · 5.8S ribosomal RNA · ARMS-PCR · Secondary structure · Single nucleotide polymorphism

## Introduction

English or Persian walnut (*Juglans regia* L.,  $2n = 32$ ), monoic trees, are native to the mountain chain of Central Asia and grows as wild or semi-cultivated trees in a wide area from south-eastern Europe and the Caucasus to Turkey and Iran, through southern portions of the former Soviet Union into China and the eastern Himalayas. The Persian or English walnut is the most economically important species from all the 21 species belonging to the genus *Juglans* and is the only species widely cultivated for its edible nuts. *J. regia* is an important and healthy food as well as base material for timber industry. Recent epidemiologic studies have associated walnut consumption with a reduced incidence of cardiovascular mortality [1].

Nuclear ribosomal genes (rDNA) are constituents of individual 18S-5.8S-28S repeats, tandemly repeated at one or more chromosomal loci. In plant genomes, variations in location and number of rDNA sites are commonly observed and the rDNA site loss or duplication is correlated with polyploidization events [2].

Internal transcribed spacers (ITS1 and ITS2) are part of the 18S-5.8S-28S nuclear ribosomal cistrons. In plants, ITS sequences vary from 500 to 700 bp length in angiosperms to 1,500–3,700 bp length in some gymnosperms [3, 4]. ITS region evolves relatively rapidly and is the popular tool for

L. F. Ciarmiello (✉) · P. Piccirillo · A. De Luca  
C.R.A.—Fruit Tree Research Unit, Via Torrino, 3,  
81100 Caserta, Italy  
e-mail: lorymail80@libero.it

G. Pontecorvo · I. Kafantaris · P. Woodrow  
Department of Life Sciences, University of Naples II,  
Via Vivaldi 43, Caserta, Italy

reconstructing molecular phylogenies at the specific, generic or family levels [5–7]. ITS1, located between the 18S and 5.8S genes, may be especially valuable at the species level and below [8]. ITS2, located between the 5.8S and 28S genes, has proven to contain useful biological information at higher taxonomic levels [9]. ITS sequence data have and may continue to provide insight not only into phylogenetic history but also polyploid ancestry, genome relationships, historical introgression and other evolutionary questions [10, 11]. The ITS regions and 5.8S gene of the nuclear ribosomal DNA are successfully used for investigated phylogenetic and biogeographic relationships within the genus *Juglans* consistent with the known geological history of *Juglans* [12].

It has been observed that the functionality of ITSs is related to specific post transcriptional modification by cleavage of the primary transcript within ITS1 and ITS2 during maturation of the small subunit (SSU), 5.8S, and the large subunit (LSU) ribosomal RNAs. The structural integrity of ITSs is an essential prerequisite for the correct processing of mature rRNA and for the biogenesis of active ribosomal subunits [13]. ITS sequences are subject to evolutionary constraints related to the maintenance of specific secondary structures that provide functionality [14, 15].

It has long been recognized that rDNA belong to multi-gene families and ITS sequences in particular, may be subject to concerted evolution, which occurs primarily through unequal crossing over and gene conversion [16, 17]. Concerted evolutionary mechanisms can homogenize the same sequence type per genome, leaving only species- and clade-specific character-state changes. However, following the hybridization, introgression, polyploidization or ITS pseudogenization events a plant genome may harbour divergent ITS sequence types, recovered from more exhaustive sampling of “ITS clones” from a single genome [1].

Traditional methods to identify walnut cultivars are based on phenotypic observations, but morphological characteristics are often affected by environmental and developmental factors, making their differentiation difficult and sometimes impossible. For this reason, morphological characteristics cannot be used for screening large numbers of walnut samples. Therefore, a simple method of DNA analysis, rather than the traditional authentication methods, is clearly desirable. First described by Newton and colleagues [18], amplification refractory mutations system (ARMS) has become a standard technique that allows the discrimination of alleles at a specific locus differing by as little as 1 bp [19]. The basis of ARMS is that oligonucleotides with a mismatched 3'-residue will not function as primers in PCR under appropriate conditions [18].

The aim of this study was to develop a simple technique for specific authentication of 18 walnut cultivars using

DNA method. In this study, we successfully employed ARMS-PCR to specifically identify of 18 walnut cultivars by exploiting some SNPs in internal transcribed spacers (ITS1 and ITS2 of the 18S-5.8S-28S nuclear ribosomal).

We found a cultivar-specific polymorphism in ITS1 and ITS2 regions and even in 5.8S ribosomal RNA, used in ARMS analysis for cultivar discrimination. Moreover, we explored the ITS1 and ITS2 secondary structures to examine SNPs effects on the ITS1 and ITS2 folds.

## Materials and methods

### Plant materials

A total of 18 cultivars of walnut (*J. regia* L.) were used in this study. Cultivars belonging to distinct geographical areas: Italy, Hungary, France, USA, Spain, Portugal and Greece (Table 1). The investigated plants were taken from the Fruit Tree Research Unit's collection field.

### Genomic DNA extraction

The extractions were performed in triplicate using young leaves randomly sampled from adult trees and frozen in liquid nitrogen until use. For each cultivar investigated, genomic DNA was isolated using 1 g of young leaves from

**Table 1** English walnut *J. regia* L. cultivars studied with their geographic origin

Cultivar name	Locally origin
Castronovo	Italy
Chiusa	Italy
Franquette	Italy
Freni 2	Italy
Lara	Italy
Marbotte	Italy
Sorrento	Italy
Chico C	USA
Gustine	USA
Hartley	USA
Payne	USA
Vina	USA
Rego	Portugal
Del Carril	Spain
Grand Jefe	Spain
A117	Hungary
FK5	Greece
FM6	Greece

The investigated plants were taken from the Fruit Tree Research Unit's collection field

three individual plants following the protocol described by Doyle and Doyle [20], and purified with the ICRISAT DNA extraction procedure [21] to improve the DNA quality. DNA quality and quantity were determined spectrophotometrically and by electrophoresis on 1% agarose gel stained by ethidium bromide [22].

#### PCR amplification and cloning

ITS1-5.8S-ITS2 region, consisting around 700 bp, were amplified using primers forward 18S (5'-AAGTCGTAACAAGGTTTCCGTA-3') and the reverse 28S (5'-CCCGCTTATTGATATGCTTAAA-3'). The forward and reverse primers, located in the 3' end of 18S rDNA and 5' end of 28S rDNA, respectively, were designed according to the known sequence of *J. regia* (GenBank accession no. AF399876). PCR was performed in a total volume of 50 µl, and the reaction mix consisted of 80–100 ng of template DNA, 200 µM of each dNTP (Roche), 200 µM of each primer and 2.5 U of Fast Start High Fidelity Taq DNA polymerase in 10 mM Tris/HCl pH 8.3 and 2.5 mM MgCl<sub>2</sub> (Roche). A 9800 PerkinElmer thermal cycler (Applied Biosystem) was used to carry out PCR under cycling profile of an initial denaturation step at 94°C for 7 min, followed by 35 cycles consisting of 30 s denaturation at 94°C, 30 s annealing at 58°C and 30 s extension at 72°C. The final extension is 72°C for 7 min. PCR products were gel-purified using the High Pure PCR Product Purification Kit (Roche), in accordance with the manufacturer's instructions, and cloned into pGEM-T Easy vector System II (Promega). Five independent clones

for each cultivar were sequenced in both directions using an automated DNA sequencer (ABI PRISM 310, Applied Biosystems/Perkin-Elmer) with universal SP6 and T7 primers. The sequencing run of both strands was repeated to confirm the sequence data.

#### DNA sequence analysis

Nucleotide sequence data were analyzed and compared to the GenBank-NCBI databases using the BLAST network service (<http://www.ncbi.nlm.nih.gov/BLAST/>). Multiple sequence alignments were performed using AliBee—Multiple Alignment 2.0 ([http://www.genebee.msu.su/services/malign\\_reduced.html](http://www.genebee.msu.su/services/malign_reduced.html)) with Clustal W algorithm [23] from the DDBJ Homology Search system (<http://www.ddbj.nig.ac.jp>).

#### ARMS-PCR strategy

For each investigated population, one or more specific nucleotide positions in ITS1 and ITS2 were identified. Forward primers harbouring at the 3' end the population-specific nucleotide and reverse common primers (located in rDNA 5.8S) for each population were designed. Moreover, primer sequences for the internal control (control 1 and control 2) was constructed. Their respective sequences were depicted in Table 2. Melting temperature (T<sub>m</sub>) calculations were performed with Oligo Calc: Oligonucleotide Properties Calculator server (<http://www.basic.northwestern.edu/biotools/oligocalc.html>).

**Table 2** ARMS-PCR primers

Common primer	Specific primer	Size (bp)	Cultivar
ARMS-PCR ITS1			
5'-agatatccgttccgagagtcg-3'	5'-gcggaagatcattgc <b>actg</b> -3'	300	Sorrento
	5'-ctcgggaagatcattagc <b>acc</b> -3'	300	Chico C
5'-gagagaccgagatccgttg-3'	5'-gggtgtaatgccccctgccc <b>ga</b> -3'	311	Del Carril
5'-tgctgacaccagcagacgt-3'	5'-ccacggcgccccgaaac <b>ggg</b> -3'	222	Franquette
ARMS-PCR ITS2			
5'-atcgatgaagaacgtagcga-3'	5'-ccacacataagaagt <b>taa</b> -3'	167	FK5
5'-tgggtgcggggaagacattg-3'	5'-gcctgacctggggtcg <b>gtt</b> -3'	197	Payne
5'-cgactctcgcaacggatatt-3'	5'-ttcacagcagggcaac <b>ct</b> -3'	335	Marbotte
	5'-agagtcgaggaacgac <b>gc</b> -3'	370	FM6
	5'-gcatgtagagtcgct <b>ttttg</b> -3'	381	Gustine
	5'-gatgtagagtcgagga <b>actg</b> -3'	379	Hartley
	5'-ggtagagtcgagga <b>acta</b> -3'	376	Grand Jefe
	5'-agtcgaggaacgac <b>caatac</b> -3'	371	Freni 2
	5'-gagtcgaggaacgac <b>caaac</b> -3'	372	Rego
	5'-cctgacctggggtcg <b>cta</b> -3'	390	Castronuovo
	5'-cctggggtcgcatgtaga <b>ga</b> -3'	390	Lara
	5'-cctgacctggggtcg <b>gtt</b> -3'	300	Payne
	5'-gcctgacctggggtcg <b>agg</b> -3'	300	Vina

Mutated nucleotides are in boldface

Based on the modified PCR protocol previously described by Newton and colleagues [18], ARMS PCRs were performed in a volume of 50 µl containing approximately 80 ng genomic DNA, 1.5 mM of each dNTP (Roche), 1 µM of each primer and 1U of Taq DNA polymerase in 10 mM Tris/HCl pH 8.3, 2.5 mM MgCl<sub>2</sub> (Roche). In applying ARMS strategy to our analysis, control 1 and control 2 primers were included in all reactions to provide an internal control PCR product. The PCR cycling program consisted of a preliminary denaturation at 94°C for 4 min, followed by 35 cycles of 30 s at 94°C, 45 s at 50 or 66 °C, and 30 s at 72°C with a final extension step for 7 min at 72°C. Then, PCR products were electrophoresed on 2% agarose gels and visualized in presence of 0.5 µg/ml ethidium bromide.

A phylogenetic tree was constructed using Neighbor-Joining algorithm of the phylogeny program of MEGA 3.1.

The percentage of GC nucleotides in ITS1 and ITS2 from each cultivar was calculated using the GC calculator server ([http://www.genomicsplace.com/gc\\_calc.html](http://www.genomicsplace.com/gc_calc.html)).

The RNA secondary structures for ITS1 and ITS2 were predicted with mfold version 3.2 program (<http://www.bioinfo.rpi.edu/applications/mfold>) available on the World Wide Web [24]. Mfold predicts RNA structures by identifying the suboptimal structures using the free energy optimization methodology at a default temperature of 37°C according to the Zuker dynamic programming algorithm [24]. Prediction of the ITS1 and ITS2 RNA common secondary structures was performed using the Pfold server (<http://www.daimi.au.dk/~compbio/pfold>). The method generates a statistical sample of individual structures from an alignment of RNA sequences using an algorithm based on an explicit evolutionary model and a probabilistic model of structures [25].

#### RNA fold

The Stribo program in Sfold (Statistical Folding and Rational Design of Nucleic Acids) server was used to predict the probable target accessibility sites (loops) for trans-cleaving ribozymes in ITS1 and ITS2 [26]. The probability profiling approach by Ding and Lawrence [26, 27] reveals target sites that are commonly accessible for a large a number of statistically representative structures in the target RNA. This novel approach bypasses the long-standing difficulty in accessibility evaluation due to limited representation of probable structures due to high statistical confidence in predictions. The probability profile for individual bases ( $W = 1$ ) is produced for the region that includes a triplet (the default triplet is GUC) and two flanking sequences of 15 bases each in every site of the selected cleavage triplet.

## Results and Discussion

Isolation and sequence analysis of ITS1-5.8S-ITS2 clones from 18 *J. regia* cultivars

Sequence data have been deposited in the GeneBank Data Library under accession number

bankit1336925	HM049887
bankit1336960	HM049890
bankit1336973	HM049893
bankit1336982	HM049896
bankit1336990	HM049899
bankit1336996	HM049902
bankit1336940	HM049888
bankit1336966	HM049891
bankit1336978	HM049894
bankit1336985	HM049897
bankit1336993	HM049900
bankit1336999	HM049903
bankit1336954	HM049889
bankit1336970	HM049892
bankit1336980	HM049895
bankit1336987	HM049898
bankit1336994	HM049901
bankit1337003	HM049904

We determined and compared the nucleotide sequences of the ITS1-5.8S-ITS2 region obtained from multiple sequence alignment of the five clones for each *J. regia* cultivars belonging to seven geographic areas (Fig. 1; Table 1).

The length of ITS regions in the population investigated was similar to that expected [17, 28], as well as the length compensation between ITS1 and ITS2 is a family-specific trait of Juglandaceae and other families, such as Betulaceae, Scrophulariaceae and Viscaceae [4, 17]. ITS1 sequence varied from 257 to 263 bp, 5.8S region was 164 bp and the ITS2 from 217 to 219 bp (Table 3). The G+C average content of ITS1, 5.8S and ITS2 from the 18 *J. regia* cultivars was 54.3–58.9% (Table 3). It has long been recognized that the G+C content between ITS1 and ITS2 is balanced in Juglandaceae as in most organisms and it has also been suggested that the high content of GC is related to the stability maintenance of the DNA and RNA secondary structures especially in stem-loop [14]. Multiple sequence alignment of the five clones (for each cultivar) from ITS1-5.8S-ITS2 region showed within-cultivar identical sequences. For each cultivar, consensus sequence was used for performed multiple alignment, results are showed in Fig. 1.

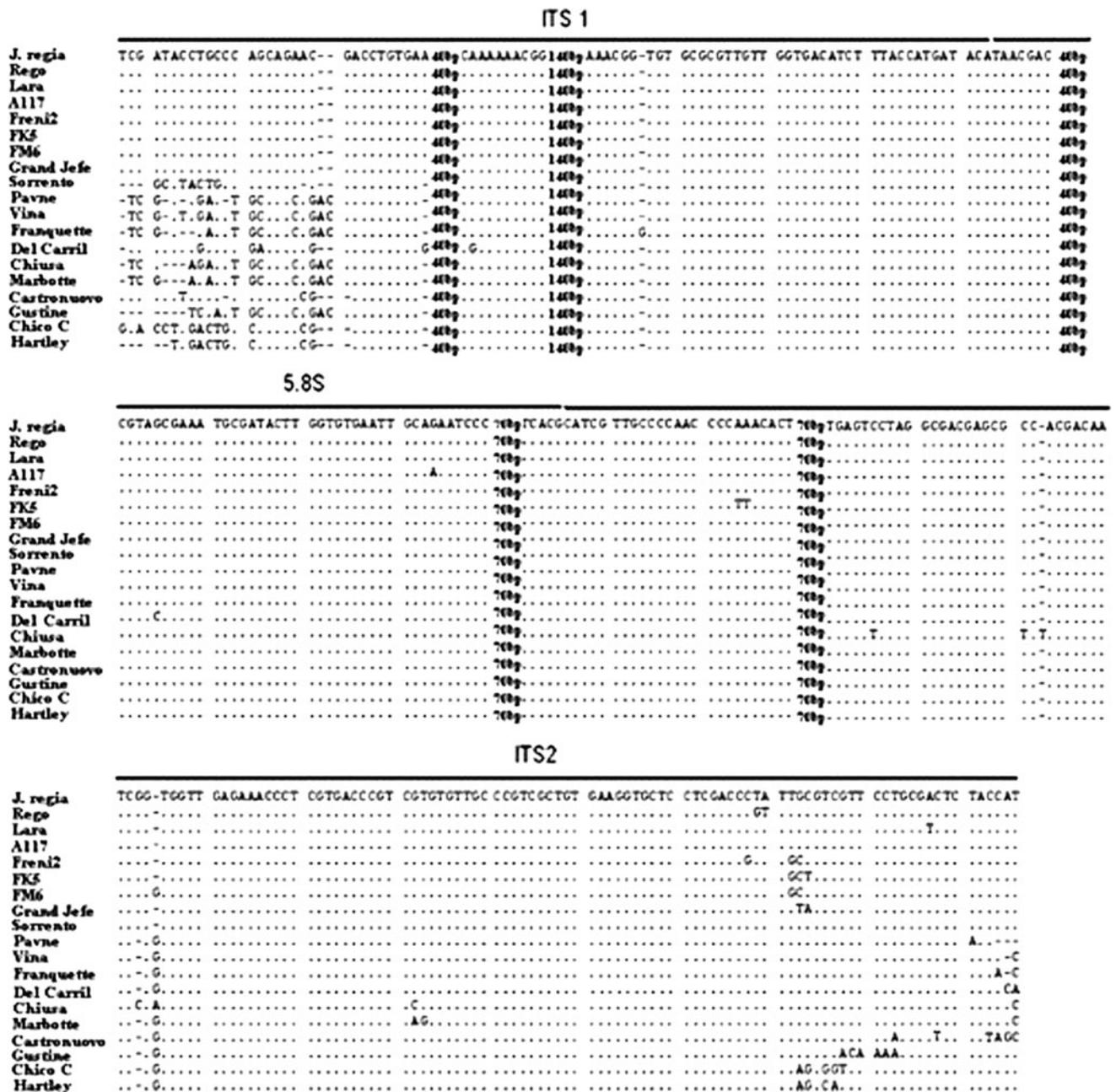


Fig. 1 Multiple sequence alignment of 18S-ITS1–5.8S-ITS2–28S rDNA from 18 walnut cultivars. Names of cultivars are listed on the left. ‘.’ indicate nucleotides from all sequences identical to those of the *J. regia* consensus sequence; ‘-’ indicate deletion. SNPs are indicated to letters

Variations in ITS1 and ITS2 sequences analysed included transition, transversion, substitution and deletion events. The aligned ITS dataset showed 244 variable positions, 168 in ITS1, 76 in ITS2 and 2 in 5.8S region. There was a AC insertion/deletion in the ITS1 region (Fig. 1). In particular, the transitions (60%) predominant over transversions (40%) (Table 4). Transitions from C to T play an important role in the ITS evolution of plants by achieving part of the GC balance recognized in the spacers

[14]. Moreover CpG e CpNpG are the main sites of methylation in plants rDNA.

SNPs are an excellent way to obtain high-quality markers useful for a very quickly identify of a single base differences within the genome [29]. SNPs are the most used molecular markers because they have an high density throughout the genome. They were used for the first time to classify 51 olive cultivars, allowing 49 varieties of olives grown for oil production in Mediterranean area to be

**Table 3** ITS1, 5.8S, ITS2 length and G+C % content for each cultivar analysed

Cultivar	ITS1		5.8S		ITS2	
	Length (bp)	G+C (%)	Length (bp)	G+C (%)	Length (bp)	G+C (%)
Rego	263	55.1	164	54.9	218	58.7
Sorrento	258	55.4	164	54.9	217	58.1
Payne	260	55.8	164	54.9	215	58.1
Hartley	256	55.9	164	54.9	218	57.8
Chico C	261	56.3	164	54.9	218	57.8
Gustine	257	55.6	164	54.9	218	57.3
Freni 2	263	55.1	164	54.9	218	58.7
Vina	262	55.7	164	54.9	217	58.5
Chiusa	260	55.4	164	54.9	219	57.1
Franquette	261	55.6	164	54.9	217	58.1
Marbotte	260	55.4	164	54.9	217	58.5
Lara	263	55.1	164	54.9	218	58.3
Castronuovo	260	55.0	164	54.9	217	57.6
Grand Jefe	263	55.1	164	54.9	218	57.3
A117	263	55.1	164	54.3	218	58.3
Del Carril	263	56.7	164	54.9	218	58.3
FM6	263	55.1	164	54.9	219	58.9
FK5	263	55.1	164	54.9	218	58.3

discriminated and new cultivars identified [30]. ITS sequence was also used to develop molecular marker to distinguish *Tribulus terrestris* L. (Zygophyllaceae) from its adulterants [31]. ITS sequences show a relatively high number of SNPs and are easily isolated by PCR, because they are located between highly conserved sequences (18S and 28S gene) that constitute the target site for the construction of primers. These two characteristics make ITS an important means for reconstructing molecular phylogenies and fingerprinting plants.

#### Relationship between *Juglans* cultivars

A Neighbour-Joining tree based on nucleotide sequence alignment was constructed for ITS1-5.8S-ITS2 whole region of 18 cultivars. Two major clodes of *Juglans* cultivars were recognized (Fig. 2). Clode I included 12 cultivars from Italy, France, Portugal, Spain, Greece, Hungary and USA (Sorrento, Castronuovo, Freni 2, Lara, Rego, Grand Jefe, Del Carril, FK5, FM6, A117, Hartley and Chico C) and *J. regia* (a.n. AF399876). Clode II comprised two french cultivar (Franquette and Marbotte), one italian (Chiusa) and three american cultivar (Vina, Payne, Gustine).

In a few cases, cultivars geographically close or growing in similar pedoclimatic conditions were more similar than populations grown in different geographic areas and under different conditions. This is likely due to migration of *J. regia* between Europe and North America and to certain

phenomena, such as high introgression during plant species domestication that occurs in a pauperization of genetic pool [17], hybridization and polyploidization events [32], and/or incomplete intra- or inter-array homogenization processes [6].

#### SNPs effects on ITS1 and ITS2 secondary structure

Among the molecular markers used for walnut taxonomy, the ITS1-ITS2 of the ribosomal DNA genes are useful for distinguishing among specimens belonging to closely-related species.

Because the structural elements of ITS sequences are essential for the specific cleavage steps during the ribosomal RNA maturation, we analyzed the effects of the SNPs on the ITS1-ITS2 folds and compared the secondary structure from each investigated cultivars.

Hence, the secondary structures of ITS RNA sequences provide additional evolutionary information [33–35] and a rather simple molecular marker, which may be particularly useful when studying closely related species [36]. We relied primarily on the free-energy minimization approach to secondary structure inference, which assumes that the dominant interactions (H-bonding between bases and stacking between adjacent bases) are local and that conformations adopted by RNA are equilibrium, lowest free-energy conformations [37].

However, the thermodynamically optimal structure does not necessarily reflect the in vivo structure, since (largely unknown) interacting factors such as other molecules (e.g.

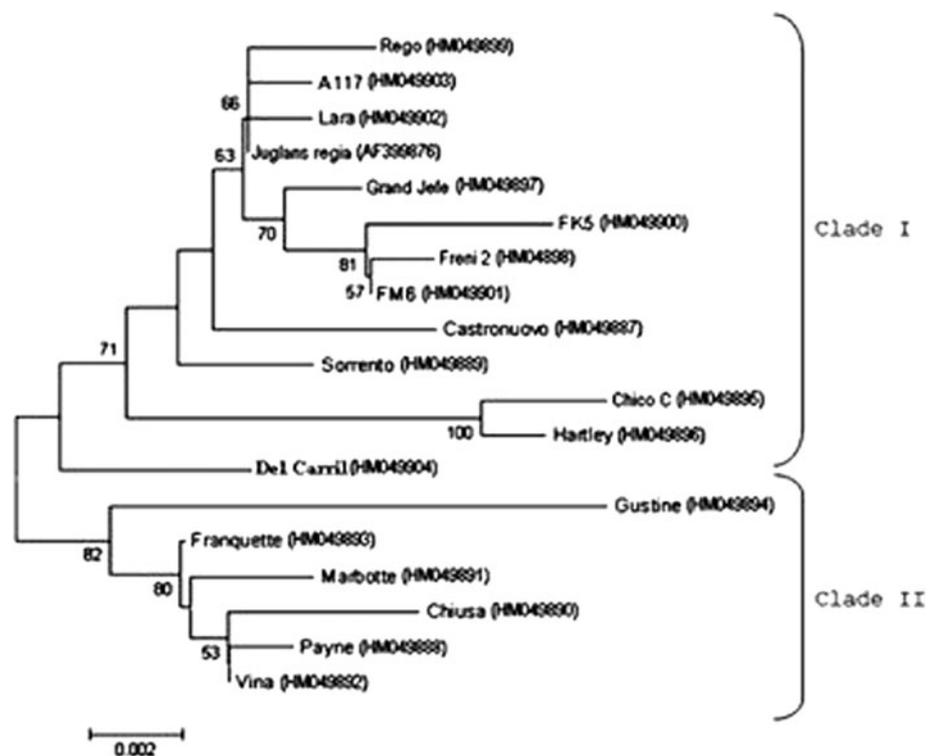
**Table 4** Mutations in ITS1, ITS2 and rDNA 5.8S sequences of 18 walnut cultivars

Cultivar	ITS1		ITS2		Deletion		Insertion		rDNA 5.8S	
	Substitution		Substitution		ITS1	ITS2	ITS1	ITS2	Substitution	
	Transition	Transversion	Transition	Transversion					Transition	Transversion
Castronovo	1 (C–T)	1 (A–C) 1 (C–G)	1 (T–C) 1 (A–G)	3 (T–A) 2 (C–A)	5	1		1		
Chiusa	2 (G–C) 2 (C–T) 1 (G–A)	1 (T–G) 2 (C–A) 1 (C–G)	2 (C–T)	2 (G–C)	5		2	2		
Franquette	2 (C–T) 2 (A–G) 1 (G–A)	2 (G–C) 1 (A–C) 1 (C–G)	1 (T–C)	1 (C–A)	5	2	2			
Freni 2				1 (C–G) 1 (T–G) 1 (G–C) 1 (A–T)						
Lara										
Marbotte	2 (C–T) 2 (A–G) 1 (G–A)	2 (G–C) 1 (C–A) 1 (A–C) 1 (C–G)	1 (G–A) 1 (T–C)	1 (T–G)	5	1	2	1		
Sorrento	1 (A–G) 2 (T–C) 1 (C–T)	1 (C–G) 1 (G–T) 1 (C–A)			4					
Chico C	1 (G–A) 1 (T–C) 1 (C–T) 1 (A–G)	1 (T–G) 1 (C–A) 3 (A–C) 1 (A–T) 3 (C–G) 1 (T–A) 1 (G–C)	1 (G–A)	1 (T–G) 2 (C–G) 1 (G–T)	4	1		1		
Gustine	1 (T–C) 1 (A–G) 2 (C–T)	1 (C–A) 1 (G–C) 1 (A–C) 1 (C–G)	1 (G–A) 1 (T–C)	2 (T–A) 2 (C–A)	8	1	2	1		
Hartley	1 (C–T)	3 (C–G) 1 (A–T) 1 (T–A) 1 (G–C) 1 (A–C)	1 (G–A) 1 (T–C)	1 (C–A) 1 (C–G)	7	1		1		
Payne	2 (A–G) 1 (G–A) 2 (C–T)	2 (G–C) 1 (T–G) 1 (A–C) 1 (C–G)		1 (T–A)	5	4	2	1		
Vina	3 (C–T) 2 (A–G) 1 (G–A)	2 (G–C) 1 (T–G) 1 (A–C) 1 (C–G)	1 (T–C)		3	2	2	1		
Rego				1 (T–G) 1 (A–T)						

**Table 4** continued

Cultivar	ITS1		ITS2		Deletion		Insertion		rDNA 5.8S	
	Substitution		Substitution						Substitution	
	Transition	Transversion	Transition	Transversion	ITS1	ITS2	ITS1	ITS2	Transition	Transversion
Del Carril	3 (A–G) 1 (G–A)	1 (T–G) 1 (C–G)		1 (A–C) 1 (T–A)	1	1		1		1 (G–C)
Grand Jefe				1 (G–T) 1 (C–A)						
A117									1 (G–A)	
FK5			1 (C–T)	1 (G–C) 2 (A–T) 1 (T–G)						
FM6				1 (G–C)				1		

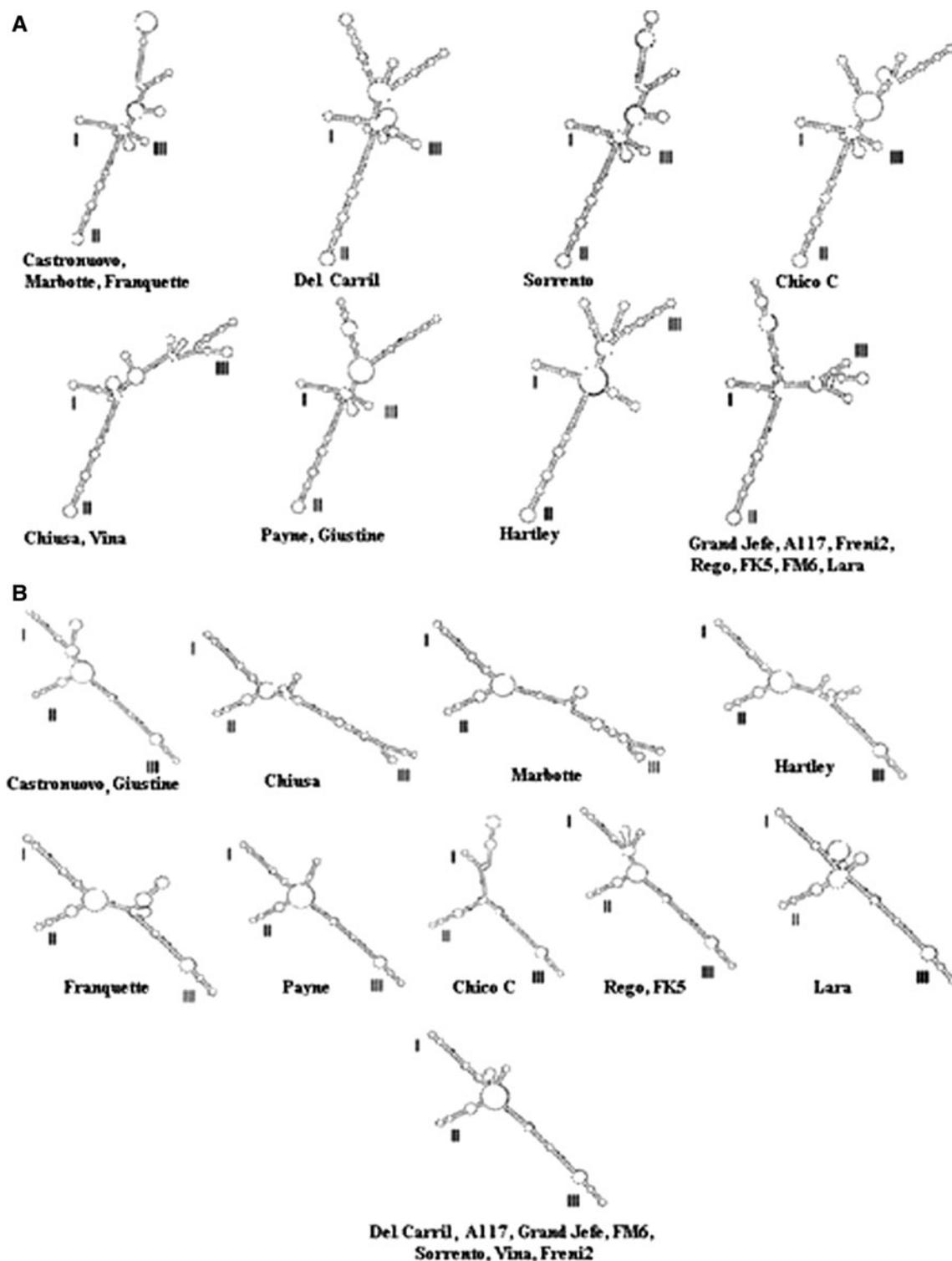
**Fig. 2** Phylogenetic tree depicting the relationship among different *J. regia* cultivars based on rDNA ITS sequences. The tree was inferred using the Neighbor-Joining method. Bootstrap values are shown as percentages at each node based on 500 replicates



proteins) or tertiary structure constraints may occur. Consequently, common structural elements of rRNA transcripts should not be exclusively reconstructed by energy optimization (provided by mFOLD), but also by homologizing internal regions of sequences [‘phylogenetic comparative method’] by mutual comparison (i.e. plausibility of thermodynamically optimal and suboptimal hypotheses). Predicted ITS1 structures of 18 *J. regia* cultivars are given in Fig. 3. The stems (double stranded paired regions) are assumed to stabilize RNA secondary structures [38]. Comparing secondary structure ITS1 sequences showed

that all cultivars had a common base topology, a different number of loops, due to the sequence variability found at the 5′ end ITS1, which generates the assessment in overall structure of ITS1 folds. These results suggest that the mutations that affect ITS1 secondary structures of different walnut cultivars are not simply accumulated due to random mutation and have evolved for functional selection in ribosome biogenesis to assess target accessibility, an essential pre-requisite for the correct RNA processing.

All sequences contain the following ITS1 motif: ‘GGCG-CGGTCT-GCGCCAAGGAA’ which corresponds



**Fig. 3** ITS1 and ITS2 transcript secondary structures. Major helices are labeled I ± III for ITS1 (a) and ITS2 (b), respectively

to the published conserved angiosperm ITS1 motif: ‘GGCG-(4–7n)–GYGYCAAGGAA’ (where Y = C or T) by Liu and Schardl [39] (Table 5). In previous studies on many flowering plants this characteristic sequence has been reported in the middle of ITS1 and this sequence is

presumed as a recognition site for processing of a primary transcript into the structural rRNA [39]. The highly conserved specific region of ITS1 which was found in our study, as well as for example in the *Asteraceae* [40] and in the *Rosaceae* genus *Rosa* [41], supports its conservative

**Table 5** Conserved ITS1 and ITS2 motifs in *J. regia*

Region	ITS1 motif 1	ITS2 motif 4	ITS2 motif 5	ITS2 motif 6
Liu and Schardl (1994)	'GGCG-(4–7n) –GYGYCAAGGAA'	'TGGT'		
Hershkovitz and Zimmer (1996)		'NNH-N-HRRYNNNAYGGTGGTWNNN'	'NYGYNGYN'	'RCYY'
cDNA consensus sequence	'GGCG-CGGTCT-GCGCCAAGGAA'	'GCGCCACGACAATCGGTGGTTGAGA'	'GTGTTGCC'	'GCTC'

*Note:* Positions of *J. regia* consensus sequences corresponding to the conserved angiosperm motifs are marked bold. The IUPAC code (K = G or T, M = A or C, R = A or G, S = C or G, W = A or T, Y = C or T, B = C or G or T, D = A or G or T, H = A or C or T, V = A or C or G, N = A or C or G or T) was used to present different nucleotides at the same position

<sup>a</sup> The consensus sequences of ITS1 motifs were done manually according to the published conserved angiosperm ITS1 motif of Liu and Schardl [39], ITS2 motifs were done manually according to the published alignment of Hershkovitz and Zimmer ([42], C1–C6 motifs, p. 2859). Consensus sequence positions of cDNA and putative gene sequences, where deletions were observed in one or more sequences are in *brackets*

status and also its important role in secondary structure forming.

A general secondary structural model could be set up, the main features of which are: (1) three helices highly conserved; (2) helix I containing in the loop the UGUA-AUG motif; (3) helix II that is longer than other two and highly conserved containing a CUCCUCGUGUG motif; (4) a helix III containing GGAAAC motif (Fig. 3). In Payne, Castronuovo, Franquette, Sorrento, Marbotte, Hartley, Gustine and Chico C is present an additional helix from the helix II and III that present the CGGUCUG motif, absent in Del Carril, Grand Jefe, A117, Freni 2, Rego, FK5, FM6, Lara, Chiusa and Vina. In addition, Castronuovo, Franquette, Marbotte, Chiusa, Vina, Sorrento, Grand Jefe, A117, Freni2, Rego, FK5, FM6, Lara present a further helix with the UAAACAAGG motif (Fig. 3).

The secondary structure of the walnut ITS2 cDNA sequences mainly correspond to the common secondary structure (helices I–IV) in angiosperms [15, 40, 42]. According to the position of the stem-loop regions of both spacers among angiosperms, it has been proposed that these act as a scaffold for the processing of the coding regions [39].

In the case of ITS2, distinct hallmarks of a core structure have been shown. These are (1) four helices with (2) helix I present the CUUAUG motif in the loop (3) as well as CUUCUG motif in the helix II with a pyrimidine (U–U) mismatch which appears to play the most important role [15] and (4) helix III as the longest and containing an UGAGAA motif in the loop. A substitution from GT to CT and AG was observed in cultivars Chiusa and Marbotte occurring the absence of the motif UGAGAA in the loop of helix III.

Within helix III, all sequences contain three conserved motifs: 'GCGCCACGACAATCGGTGGTTGAGA' which corresponds to the published conserved angiosperm ITS1 motif 4 'NNH-N-HRRYNNNAYGGTGGTWNNN'; in addition 'GTGTTGCC' which corresponds to motif 5

'NYGYNGYN' and finally, 'GCTC' corresponding to motif 6 'RCYY' [39]. These motifs playing an important role in the stem-loop formation of the secondary structure of ITS2.

A complex network of interactions take place to assemble the pre-rRNA structural features directly involved in processing steps into a relatively compact structure. The identification of common structural elements indicates equivalent functionality of the corresponding molecules [43].

Subtle secondary structural motifs may participate in the ITS excision process, spatially positioned by the conserved framework of helices.

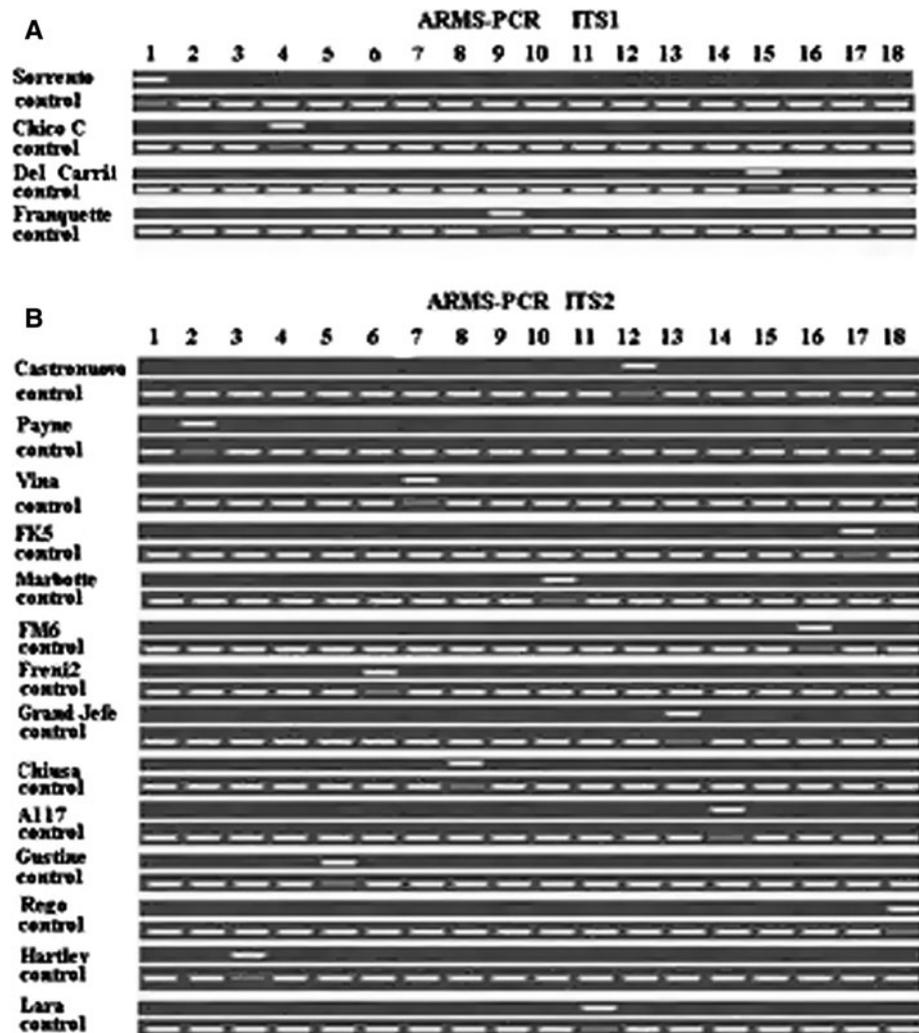
The biological function of an RNA molecule is determined by its secondary structure and the sequence variations that contribute to differences between species are those that preserve the RNA functions [44]. The probability profiling approach revealed that the interior loops represent the targeting preferential sites, whereas the exterior loops are target sites of difficult accessibility. Our results suggest that the ITS folds assessed by the SNPs do not affect the target accessibility, an essential pre-requisite for the correct RNA processing. Critical changes in the rRNA folding pattern due to evolutionary sequence variation in the ITS regions may have an important role on the kinetics of precursor rRNA formation for the efficient functioning of the rDNA cluster [44].

#### ARMS-PCR

Molecular characterization of *Juglans* cultivars was performed via ARMS-PCR. As shown in Fig. 1 there were 7 variable sites in ITS1 and 12 in ITS2.

The primers designed to differentiate the 18 *J. regia* cultivars were highly specific, as only the targeted cultivar was amplified by the corresponding primers. The 90 bp internal control was generated in all PCR amplifications whereas no product was observed when the cultivar-specific primers was applied to other cultivars. We designed specific primers that contain at their 3' ends the specific

**Fig. 4** ARMS-PCR strategy based on the SNPs. Lanes 1–18 amplification PCR products for ITS1 (a) and ITS2 (b) and 90 bp internal control obtained from the genomes of Sorrento, Payne, Hartley, Chico C, Gustine, Freni 2, Vina, Chiusa, Franquette, Marbotte, Lara, Castronuovo, Grand Jefe, A117, Del Carril, FM6, FK5, Rego



nucleotide, that match with the SNP which we used in conjunction with a single common primer (Table 2). Specific primers 3'-OH terminal mismatches (which are purine/purine, pyrimidine/pyrimidine and purine/ pyrimidine type mismatches) were refractory to extension by the DNA polymerase lacking in 3'-exonucleolytic proofreading activity [18], thus probably prohibiting hybridization of the specific primer on template DNA from other cultivars.

In particular, the C insertion located next to CpC residues at position 126 in the ITS2 region from Chiusa may be the result of C residue duplication due to the slippage replication events and short mismatch during replication [45]. A substitution from TA to GC at positions 7 and 8 in the ITS1 region are fixed in Payne, Gustine, Chico C and Hartley population-specific single point sites. The primers designed allow to differentiate seventeen of 18 *J. regia* cultivars (Fig. 4).

The SNPs, can provide useful molecular markers for the development of an assay that could discriminate

different *J. regia* cultivars by using primers that hybridize with the specific nucleotide substitutions. Our results demonstrated the feasibility and usefulness of the ARMS-PCR for differentiation of seventeen on 18 *J. regia* cultivars, suggesting that the ITS specific variation sequence could be used as a molecular marker in cultivar identification and discrimination. Molecular characterization can be performed using other molecular markers such as RAPD, ISSR, RFLP and SSR. These methods are useful for revealing genetic polymorphisms among different walnut cultivars, but RAPD and ISSR are not suitable for cultivar identification due to their lack of reproducibility. RFLP is a laborious test, because it requires treating amplified DNA with restriction enzyme and the maintenance of probes clones. The drawback of SSR is that the size differences between the products amplified from each allele are usually small, complicating reliable scoring by standard agarose gel electrophoresis (Fig. 4).

In recent years, with increasing interest in the plant systematic community for resource improvement and conservation, several studies have been undertaken for new marker-assisted breeding development [46], molecular authentication of the medical plants using PCR-RFLP and ARMS strategies [47–51], and genetic assessment based on ITS sequence polymorphism from commercially important cultivars and species of authentic geographical localization [51–53]. Because the variety/species identification is an important link between the conservation and utilization of a commercially important plant genetic resource such as *J. regia*, new and fast molecular authentication methods for the cultivar discrimination are needed. ARMS strategy is routinely performed by molecular geneticists for new disease-associated mutation identification with good genotyping accuracy [18, 54, 55].

This strategy may therefore be a good tool for the comparative genetic studies of agronomically important cultivars and identification of the unnamed local cultivars, and can also provide useful information on the uniqueness of the accessions from germplasm collection.

**Acknowledgments** The authors are grateful to Joseph Sepe (University of Naples II, Caserta, Italy), for improving the manuscript.

## References

- Anderson KJ, Teuber SS, Gobeille A, Cremin P, Waterhouse AL, Steinberg FM (2001) Walnut polyphenolics inhibit in vitro human plasma and LDL oxidation<sup>1, 2</sup>. *J Nutr* 131:2837–2842
- Mishima M, Ohmido N, Fukui K, Yahara T (2002) Trends in site-number change of rDNA loci during polyploid evolution in *Sanguisorba* (Rosaceae). *Chromosoma* 110:550–558
- Bobola MS, Smith DE, Klein AS (1992) Five major nuclear ribosomal repeats represent a large and variable fraction of the genomic DNA of *Picea rubens* and *P. Mariana*. *Mol Biol Evol* 9:125–137
- Baldwin BG, Sanderson MJ, Porter JM, Wojciechowski MF, Campbell CS, Donoghue MJ (1995) The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. *Ann Mo Bot Gard* 82:247–277
- Baldwin BG (1992) Phylogenetic utility of the internal transcribed spacers of nuclear ribosomal DNA in plants: an example from the Compositae. *Mol Phylogenet Evol* 1:3–16
- Álvarez I, Wendel JF (2003) Ribosomal ITS sequences and plant phylogenetic inference. *Mol Phylogenet Evol* 29:417–434
- Tippery NP, Les DH (2008) Phylogenetic analysis of the internal transcribed spacer (ITS) region in *Menyanthaceae* using predicted secondary structure. *Mol Phylogenet Evol* 49:526–537
- Soltis PS, Kuzoff RK (1993) ITS sequence variation within and among populations of *Lomatium grayi* and *L. laevigatum* (Umbelliferae). *Mol Phylogenet Evol* 2:166–170
- Dassanayake RS, Gunawardene YINS, Nissanka BDD, De Silva K (2008) ITS-2 secondary structures and phylogeny of *Anopheles culicifacies* species. *Bioinformation* 2(10):456–460
- Bailey CD, Hughes CE, Harris SA, Carr TG (2002) Characterization of angiosperm nrDNA polymorphism, paralogy, and pseudogenes. *Mol Phylogenet Evol* 29:435–455
- Wissemann V (2002) Molecular evidence for allopolyploid origin of the *Rosa canina* complex (Rosaceae, Rosoideae). *J Appl Bot* 76:176–178
- Stanford AM, Harden R, Parks CR (2000) Phylogeny and biogeography of *Juglans* (Juglandaceae) based on *matk* and its sequence data. *Am J Bot* 87(6):872–882
- Musters W, Boon K, Van Der Sande CAEM, Van Heerikhuizen H, Planta RJ (1990) Functional analysis of transcribed spacers of yeast ribosomal DNA. *EMBO J* 9:3989–3996
- Torres RA, Ganal M, Hembleden V (1990) GC balance in the internal transcribed spacers ITS1 and ITS2 of nuclear ribosomal RNA genes. *J Mol Evol* 30:170–181
- Mai JC, Coleman AW (1997) The internal transcribed spacers 2 exhibits a common secondary structure in green algae and flowering plants. *J Mol Evol* 44:258–271
- Dover G (1982) Molecular drive: a cohesive mode of species evolution. *Nature* 299:111–117
- Jobst J, King K, Hembleden V (1998) Molecular evolution of the internal transcribed spacers (ITS1 and ITS2) and phylogenetic relationships among species of the family Cucurbitaceae. *Mol Phylogenet Evol* 9:204–219
- Newton CR, Graham A, Heptinstall LE, Powell SJ, Summers C, Kalsheker N, Smith JC, Markham AF (1989) Analysis of any point mutation DNA. The amplification refractory mutation system (ARMS). *Nucleic Acids Res* 17:2503–2516
- Stirling D (2003) Multiplex amplification refractory mutation system for the 20 detection of prothrombotic polymorphisms. In: Bartlett JMS et al (eds) PCR 21 protocols. Humana, New Jersey, p 323
- Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull* 19:11–15
- Sharma KK, Lavanya M, Anjaiah V (2000) A method for isolation and purification of peanut genomic DNA suitable for analytical applications. *Plant Mol Biol Rep* 18:393a–393h
- Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual, 2nd edn. Cold Spring Harbour Laboratory, Cold Spring Harbour, NY, USA
- Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673–4680
- Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 31:3406–3415
- Knudsen B, Hein J (2003) Pfold: RNA secondary structure prediction using stochastic context-free grammars. *Nucleic Acids Res* 31:3423–3428
- Ding Y, Chan CY, Lawrence CE (2004) Sfold web server for statistical folding and rational design of nucleic acids. *Nucleic Acids Res* 32:W135
- Gruber AR, Lorenz R, Bernhart SH, Neuböck R, Ivo L, Hofacker IL (2008) The Vienna RNA websuite. *Nucleic Acids Res* 36(2):W70–W74
- Garcia-Mas J, Monforte AJ, Arús P (2004) Phylogenetic relationships among *Cucumis* species based on the ribosomal internal transcribed spacer sequence and microsatellite markers. *Plant Syst Evol* 248:191–203
- Cooper DN, Smith BA, Cooke H, Niemann S, Schmidtke J (1985) An estimate of unique sequence heterozygosity in the human genome. *Hum Genet* 69:201–205
- Consolandi C, Palmieri L, Doveri S, Maestri E, Marmiroli N, Reale S, Lee D, Baldoni L, Tosti N, Severgnini M, De Bellis G, Castiglioni B (2007) Olive variety identification by ligation detection reaction in a universal array format. *J Biotechnol* 129:565–574

31. Balasubramani SP, Murugan R, Ravikumar K, Venkatasubramanian P (2010) Development of ITS sequence based molecular marker to distinguish *Tribulus terrestris* L. (Zygophyllaceae) from its adulterants. *Fitoterapia*. doi:10.1016/J.fitote.210.01.002
32. Small RL, Wendel JF (2000) Copy number lability and evolutionary dynamics of the Adh gene family in diploid and tetraploid cotton (*Gossypium*). *Genetics* 155(4):1913–1926
33. Coleman AW (2003) ITS2 is a double-edged tool for eukaryote evolutionary comparisons. *Trends Genet* 19:370–375
34. Coleman AW (2007) Pan-eukaryote ITS2 homologies revealed by RNA secondary structure. *Nucleic Acids Res* 35:3322–3329
35. Wolf M, Achtziger M, Schultz J, Dandekar T, Müller T (2005) Homology modeling revealed more than 20,000 rRNA internal transcribed spacer 2 (ITS2) secondary structures. *RNA* 11:1616–1623
36. Müller T, Philippi N, Dandekar T, Schultz J, Wolf M (2007) Distinguishing species. *RNA* 13:1469–1472
37. Xia X (2000) Data analysis in molecular biology and evolution. Kluwer Academic Publishers, Boston
38. Wuyts J, Van de Peer Y, De Wachter R (2001) Distribution of substitution rates and location of insertion sites in the tertiary structure of ribosomal RNA. *Nucleic Acids Res* 29:5017–5028
39. Liu JS, Schardl CL (1994) A conserved sequence in internal transcribed spacer 1 of plant nuclear rRNA genes. *Plant Mol Biol* 26:775–778
40. Goertzen LR, Cannone JJ, Gutell RR, Jansen RK (2003) ITS secondary structure derived from comparative analysis: implications for sequence alignment and phylogeny of the Asteraceae. *Mol Phylogenet Evol* 29:216–234
41. Ritz CM, Schmutz H, Wissemann V (2005) Evolution by reticulation: European dogroses originated by multiple hybridization across the genus *Rosa*. *J Hered* 96:4–14
42. Hershkovitz MA, Zimmer EA (1996) Conservation patterns in angiosperm rDNA ITS2 sequences. *Nucleic Acids Res* 24:2857–2867
43. Fox GE, Woese CR (1975) 5S RNA secondary structure. *Nature* 256:505–507
44. Bhargavi R, Vishwakarma S, Murty US (2005) A secondary structural common core in the ribosomal ITS2 (internal transcribed spacer) of *Culex* species from diverse geographical locations. *Bioinformatics* 1:52–55
45. Dover GA (1986) Molecular drive in multigene families: how biological novelties arise, spread and are assimilated. *Trends Genet* 6:159–165
46. Asp T, Frei UK, Didion T, Nielsen KK, Lübberstedt T (2007) Frequency, type, and distribution of EST-SSRs from three genotypes of *Lolium perenne*, and their conservation across orthologous sequences of *Festuca arundinacea*, *Brachypodium distachyon*, and *Oryza sativa*. *BMC Plant Biol* 12:36–48
47. Fantaccione S, Woodrow P, Pontecorvo G (2008) Molecular authentication of three Italian melon accessions by ARMS-PCR and ITS1 (internal transcribed spacer 1) secondary structure prediction. *Bioinformatics Biomed Inform Publ Group* 2(7):311–315
48. Xue HG, Zhou SD, He XJ, Yu Y (2007) Molecular authentication of the traditional medicinal plant *Croton caudatus*. *Plant Med* 73:611–613
49. Li X, Ding X, Chu B, Ding G, Gu S, Qian L, Wang Y, Zhou Q (2007) Molecular authentication of *Alisma orientale* by PCR-RFLP and ARMS. *Plant Med* 73:67–70
50. Sukrong S, Zhu S, Ruangrunsi N, Phadungcharoen T, Palanuvej C, Komatsu K (2007) Molecular analysis of the genus *Myrtagyna* existing in Thailand based on rDNA ITS sequences and its application to identify a narcotic species: *Myrtagyna speciosa*. *Biol Pharm Bull* 30:1284–1288
51. Wang H, Sun H, Kwon WS, Jin H, Yang DC (2010) A PCR-based SNP marker for specific authentication of Korean ginseng (*panax ginseng*) cultivar “Chunpoong”. *Mol Biol Rep* 37(2): 1053–1057
52. Ishikawa N, Yokoyama J, Ikeda H, Takabe E, Tsukaya H (2006) Evaluation of morphological and molecular variation in *Plantago asiatica* var. *densiuscula*, with special reference to the systematic treatment of *Plantago asiatica* var. *yakusimensis*. *J Plant Res* 119:358–395
53. Joy N, Abraham Z, Soniya EV (2007) A preliminary assessment of genetic relationships among agronomically important cultivars of black pepper. *BMC Genet* 8:42–49
54. Old JM, Petrou M, Modell B, Weatherall DJ (1984) Feasibility of antenatal diagnosis of beta thalassaemia by DNA polymorphisms in Asian Indian and *Cypriot* populations. *Br J Haematol* 57:255–263
55. Newton CR, Heptinstall LE, Summers C, Super M, Schwarz M, Anwar R, Graham A, Smith JC, Markham AF (1989) Amplification refractory mutation system for prenatal diagnosis and carrier assessment in cystic fibrosis. *Lancet* 2:1481–1483