Molecular authentication of three Italian melon accessions by ARMS-PCR and ITS1 (internal transcribed spacer 1) secondary structure prediction

Stefania Fantaccione[1], Pasqualina Woodrow[1] and Giovanni Pontecorvo[1, *]

[1]Department of Life Science, II University of Naples, via Vivaldi 43, 81100 Caserta, Italy;
Giovanni Pontecorvo*- E-mail: giovanni.pontecorvo@unina2.it; * Corresponding author

Abstract:
Genetic assessment was carried out on three Italian melon accessions by sequence and structural analysis of the internal transcribed spacer 1 (ITS1) from three populations belonging to two *Cucumis melo* L. varieties (*madras* and *tendral*). Alignment of the 18S-5.8S-26S sequences from three melon accessions showed that there were three single-nucleotide polymorphisms (SNPs) and one short insertion-deletion (indel) at the 5'end ITS1. An amplification refractory mutation system (ARMS)-PCR-based analysis was successfully applied to the SNP markers of the ITS1 sequences for the fingerprinting analysis of three melon populations. Secondary structure models for each ITS1 were derived. The prediction of ITS1 RNA secondary structure from each accession was improved by detecting key functional elements shared by all sequences in the alignments. Our results demonstrated that the ITS1secondary structure models can be used to improve the preliminary genetic assessment of the three melon accessions, suggesting a new tool in plant fingerprinting analysis.

Keywords: secondary structure prediction; ARMS-PCR; *Cucumis melo* L.; ITS1; SNP

Background:

Melon (*Cucumis melo* L.) is an outcrossing horticultural crop of worldwide economic importance belonging to the subsp. *melo* in the Cucurbitaceae family. Within the genus *Cucumis*, melon is considered to be the most diverse species for morphological features, such as fruit size, shape, colour and texture. The polymorphism in melon is relatively high as reported by simple-sequence repeat (SSR or microsatellite) marker exploration in a wide collection of melon germplasm [1]. Genetic diversity in melon accessions was evaluated using different types of molecular markers. To date, the microsatellite [2] and single-nucleotide polymorphism (SNP) markers [3] have shown to be excellent tools for melon cultivar identification.

Internal transcribed spacers (ITS1 and ITS2) are part of the 18S-5.8S-26S nuclear ribosomal cistrons. The ITS functionality is related to specific 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 (rRNAs). The structural integrity of ITSs is an essential prerequisite for the correct processing of mature rRNA and for the biogenesis of active ribosomal subunits [4]. ITS sequences evolve rapidly, representing a popular and versatile tool for reconstructing the plant molecular phylogeny at the specific, generic or family levels [5]. To assess the phylogenetic relationship among species of the Cucurbitaceae family [6] and the *Cucumis* genus [7] ITS sequence data were used. ITS1, located between the 18S and 5.8S genes, may be especially valuable as a versatile genetic marker at the species level and below [8].

The understanding genetic diversity is necessary for efficient conservation and management of germplasm collection in agronomically important plants. In this preliminary study, we first analyzed and compared the nucleotide sequence of the ITS1-5.8S-ITS2 region from three melon populations belonging to two *C. melo* varieties (*madras* and *tendral*). This study revealed four SNPs at 5'end ITS1, used in amplification refractory mutation system (ARMS) analysis for melon discrimination. Moreover, to examine the effects of the SNPs on the overall structure of ITS1 folds we explored the ITS1 secondary structure for each accession and differences in structural elements were observed. Prediction of ITS1 RNA structure, in conjunction with the ARMS-PCR strategy, could be an attractive tool for plant fingerprinting studies.

Methodology:
Plant materials and DNA extraction
For each investigated population (Table 1 in supplementary material), genomic DNA was isolated from young leaves of three individual plants following the protocol described in [9].

Isolation of ITS1-5.8S-ITS2 sequences
The forward and reverse primers, located in the 3'end of 18S rDNA and 5'end of 25S rDNA, respectively, were designed according to the known sequence of *C. melo* (GenBank accession no. Z48805). PCR amplifications were performed in a 9700 PerkinElmer thermal cycler

(Applied Biosystem) with the following cycling profile: denaturation at 94°C for 2 min and 30 cycles of 1 min denaturation at 94°C, 1 min annealing at 53.8°C and 1 min extension at 75°C. PCR products were gel-purified and cloned into pGEM-T Easy vector System II (Promega). Ten clones for each plant (thirty clones per population) were sequenced in both directions using an automated DNA sequencer (Applied Biosystems). Sequence data have been deposited in the GenBank Data Library under accession nos. EU12157-EU12159.

ARMS-PCR strategy
Forward primers harbouring the SNP at their 3' ends and a reverse common primer (located at the 3' end of ITS1) for each population were designed. Primer sequences for the internal control, a 90 bp fragment from the 5.8S rDNA, were also designed. ARMS-PCRs were performed as described in [10]. Control primers were included in all reactions. PCR cycling program: 35 cycles of 30 s at 94°C, 45 s at 57 or 59 °C, and 30 s at 72°C. ARMS primer sequences: 1, 5'-GATCATTGTCGATGCCTAAACATCTAACC**t**-3'; 2, 5'-AGGATCATTGTCGATGCC**c**TAAACATCTA**t**-3'; 3, 5'-AAGGATCATTGTCGATGCCTAAACATCa-3' (the mismatches introduced in the ARMS primers are shown in lowercase letters); common primer sequence: c, 5'-TTAATAATACGACTAGAATGCTCCATCCC-3'; control primer sequences: C1, 5'-CGACTCTGGCAACGGATATCTC-3'; C2, 5'-GTTCGCGGATCCTGCAATTCA-3'.

DNA sequence analysis and alignments
Nucleotide sequence data were compared to the GenBank-NCBI database using the BLAST network service [11]. Multiple sequence alignments were performed using Clustal W program [12].

ITS1 secondary structure prediction
ITS1 RNA secondary structures were predicted with mfold using the free energy optimization methodology according to the Zuker dynamic programming algorithm [13]. ITS1 RNA common secondary structure prediction was performed using the Pfold server, which 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 [14].

RNA fold
ITS1 probable target accessibility sites (loops) for trans-cleaving ribozymes were predicted using the Sribo program [15]. The probability profiling approach reveals target sites that are commonly accessible for a large number of statistically representative structures in the target RNA. This 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:
**ARMS-PCR for three melon differentiation**
We first compared the nucleotide sequences of the ITS1-5.8S-ITS2 region obtained from multiple sequence alignment

of the ninety clones of two *C. melo* varieties from populations of three restricted geographical locations of South Italy (Figure 1A). At the 5' end of ITS1 there were four SNPs, which included three sites of nucleotide substitutions and 1 bp insertion-deletion (indel). Length compensation between ITS1 and ITS2 was that we expected, as well as the GC content, which is related to the stability maintenance of the RNA secondary structures [16], is in the mean range previously reported [5, 6]. Transitions from C to T play an important role in the ITS evolution of Cucurbitaceae plants, by achieving part of the GC balance recognized in the spacers [16]; we found that the C to T transitions predominated, representing 46.6% of all observed base pair exchanges.

Melon SNPs are an excellent way to obtain high-quality markers [3] for the development of an ARMS-PCR assay [10] that could discriminate between the different *C. melo* populations. To this end, we designed three primers (1, 2 and 3) that contain the specific SNP at their 3' ends, used in conjunction with a single common primer. ARMS-PCR strategy was highly specific to differentiate the three melon accessions, as only the targeted PCR product was amplified by the corresponding ARMS primers (Figure 2). These results demonstrated the feasibility and usefulness of the ARMS-PCR for the three *C. melo* ecotype differentiation, suggesting that the SNPs found in the ITS1 sequences could be used as molecular marker in cultivar identification and discrimination.

Implementation of the melon fingerprinting: ITS1 secondary structure prediction
To analyze the effects of the SNPs on the ITS1 folds, we compared the ITS1 secondary structure from each investigated accession. A general topology with the multibranch loops and several paired regions can be generalized for the three ITS1 models. The stem regions stabilize the RNA secondary structures [17] and the stem number of each ITS1 is given (Table 1 in supplementary material). Several differences in the secondary structure models are caused by the SNPs found at the 5'end of the ITS1 sequences (Figure 1B). One hairpin loop, two stems (stem I and stem II) separated by a junction region can be identified at the 5'end of the three models. Hairpin loop is 6 bp long in Pop1 and Pop 3, whereas it is elongated up to 7 bp in Pop2 on account of the C insertion at position 20 (Figure 1A). The length of the junction region is variant (1 base in Pop1, 2 bases in Pop3 and 3 bases in Pop2) and the stem II no has a common position and sequence in the three secondary structures. In addition, to improve the prediction of RNA secondary structures [18] we inspected Pfold output using an alignment of the ITS1 RNA sequences as input. Comparative analysis showed clear evidence of conservation of several secondary structural elements shared by the three ITS1 folds (Figure 3). On the one hand, three ITS1 models fall into a common topology, on the other hand, at the 5' end of ITS1 secondary structure there are several local differences between the three models. This suggests that the secondary structure analysis of the molecular markers can provide useful information on the plant uniqueness in the plant fingerprinting methods.
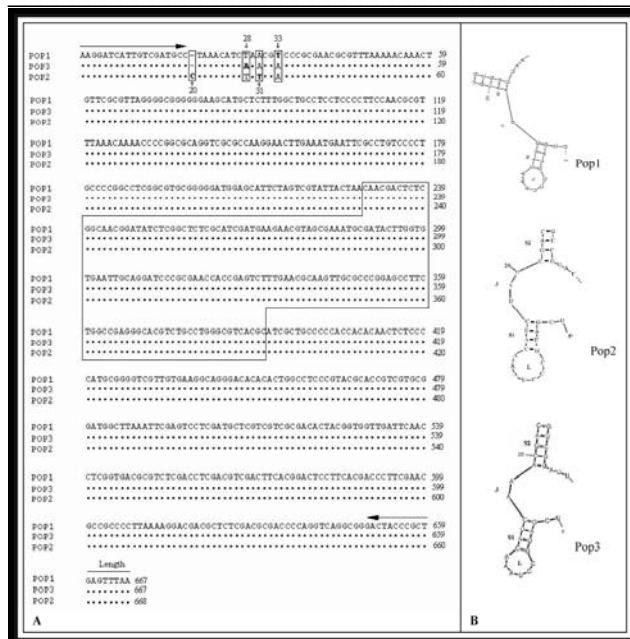
Figure 1: (A) Multiple sequence alignment of ITS1-5.8S-ITS2 rDNA from 3 Italian melon accessions. Terminal arrows indicate the 3'end of 18S rDNA and 5'end of 26S rDNA, respectively. The centrally positioned 163 bp 5.8S gene is enclosed in the box. SNPs found at the ITS1 5'end are enclosed in the boxes and their respective nucleotide positions are reported; (B) Secondary structural models for the 5'end of ITS1 sequences (S-stem; L-hairpin loop; J-junction).
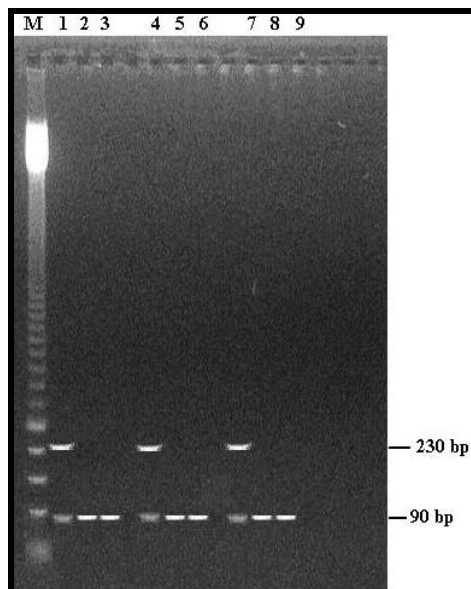
Figure 2: ARMS-PCR strategy based on the SNPs. Lanes: M, molecular mass markers (Sigma); 1, 2, 3, amplification PCR products obtained from the Pop1 genome with the 1-C, 2-C and 3-C primer combinations, respectively; 4, 5, 6, amplification PCR products obtained from the Pop2 genome with the 2-C, 1-C and 3-C primer combinations, respectively; 7, 8, 9, amplification PCR products obtained from the Pop3 genome with the 3-C, 2-C and 1-C primer combinations, respectively (230bp-ARMS PCR products; 90bp-internal control).
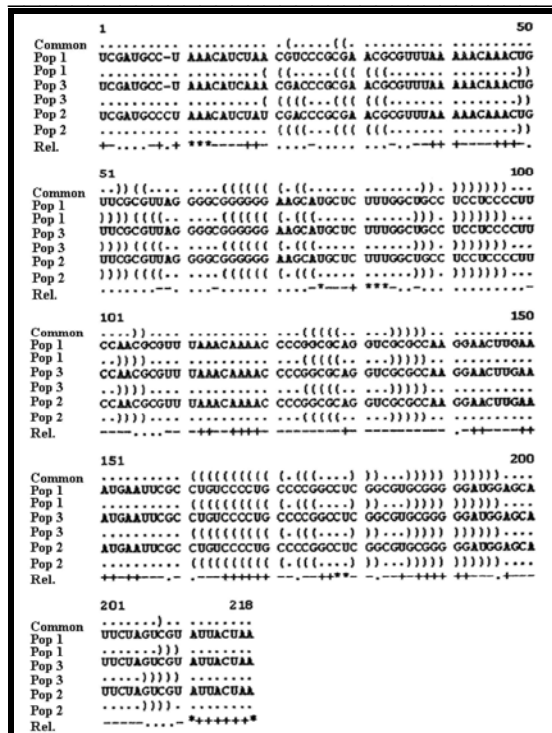
Figure 3: Pfold output ("*" > 90%; "+" > 80%; "-" > 50%; "." <= 50% indicate the reliability of the common structural motifs).

The biological function of an RNA molecule is determined by its secondary structure [18] and the sequence variations that contribute to differences between species are those that preserve the RNA functions [17]. Key functional elements of ITS secondary structures are the different unpaired regions, which are the sites of nucleation for overall three-dimensional folding and protein recognition [17, 19, 20]. Distribution among different melon accessions of different types of loops having score ≥50 with respect to the target accessibility sites for trans-cleaving ribozymes is reported (Figure 4). The probability profiling approach revealed that the interior loops represent the targeting preferential sites, whereas the exterior loops are target sites of difficult accessibility. These results suggest that the ITS1 folds assessed by the SNPs do not affect the target accessibility, an essential pre-requisite for the correct RNA processing [20]. 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 [20].
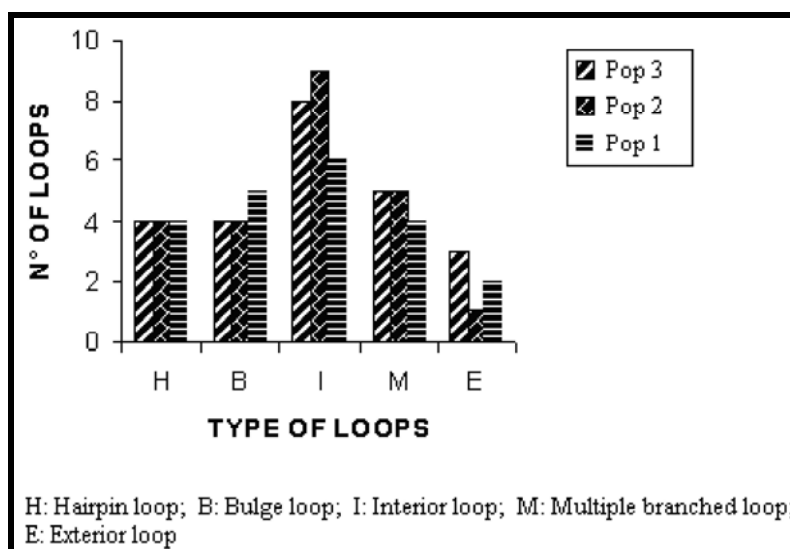

H: Hairpin loop; B: Bulge loop; I: Interior loop; M: Multiple branched loop; E: Exterior loop

Figure 4: Distribution of the different types of loops (hairpin, bulge, multi branched, interior and exterior) among different melon accessions.

Conclusion:
With increasing interest in the plant systematic community for resource improvement and conservation, several studies have been undertaken to genetically assess and molecularly authenticate agronomically important cultivars and species from a defined geographical location. The variety/species identification is an important link between the conservation and utilization of a commercially important plant such as *C. melo*. Thus, new and fast molecular authentication methods for the accession discrimination are needed. ITS variation sequence at intraspecies level seems an efficient way for accession identification and discrimination in studies based on "ITS fingerprinting" methods. ARMS-PCR strategy applied to the ITS1 SNPs, in conjunction with secondary structure prediction, may allow ITS1 to be utilized in its fullest potential for reconstructing the variety and cultivar identity in the same plant species. To expand the comparative genetic studies of agronomically important cultivars, secondary structure models can be used as the sensitive tools for plant fingerprinting analysis.

References:
[01] P. S. Ritschel *et al., BMC Plant Biol.,* 4: 9 (2004) [PMID: 15149552]
[02] M. J. Gonzalo *et al., Theor. Appl. Genet.,* 110: 802 (2005) [PMID: 15700148]
[03] M. Morales *et al., Genome,* 47: 352 (2004) [PMID: 15060588]
[04] W. Musters *et al., EMBO J.,* 9: 3989 (1990) [PMID: 2249660]
[05] B. G. Baldwin, *Mol. Phylogenet. Evol.,* 1: 3 (1992) [PMID: 1342921]
[06] J. Jobst *et al., Mol. Phylogenet. Evol.,* 9: 204 (1998) [PMID: 9562980]
[07] J. Garcia-Mas *et al., Plant Syst. Evol.,* 248: 191 (2004)
[08] P. S. Soltis & R. K. Kuzoff, *Mol. Phylogenet. Evol.,* 2: 166 (1993) [PMID: 8025723]
[09] G. C. Allen *et al., Nat Protoc.,* 1: 2320 (2006) [PMID: 17406474]
[10] C. R. Newton *et al., Nucleic Acids Res.,* 17: 2503 (1989) [PMID: 2785681]
[11] http://www.ncbi.nlm.nih.gov/blast/Blast.cgi
[12] http://clustalw.ddbj.nig.ac.jp/top-e.html
[13] http://www.bioinfo.rpi.edu/applications/mfold
[14] http://www.daimi.au.dk/~compbio/pfold
[15] Y. Ding *et al., Nucleic Acids Res.,* 32: W135 (2004) [PMID: 15215366]
[16] R. A. Torres *et al., J. Mol. Evol.,* 30: 170 (1990) [PMID: 2107331]
[17] S. Smit *et al., Nucleic Acids Res.,* 35: 3339 (2007) [PMID: 17468501]
[18] X. Fang *et al., Bioinformation,* 2: 222 (2007) [PMID: 18305832]
[19] C. S. Campbell *et al., Mol. Phylogenet. Evol.,* 35: 165 (2005) [PMID: 15737589]
[20] R. Bhargavi *et al., Bioinformation,* 1: 52 (2005) [PMID: 17597853]

Supplementary material

Cultivar	ITS1		ITS2		N° of Stems	Energy (Kcal/mol)
	Length (bp)	G+C (%)	Length (bp)	G+C (%)		
tendral var. Capua (Pop1)	217	56	238	61	15	-83.5
madras var. Maddaloni (Pop2)	218	56	238	61	14	-83.6
madras var. Piana M.V. (Pop3)	217	56	238	61	14	-86.6

Table 1: Comparison in length and GC content of the ITSs from the three melon cultivars and energy of the ITS1 secondary structures.