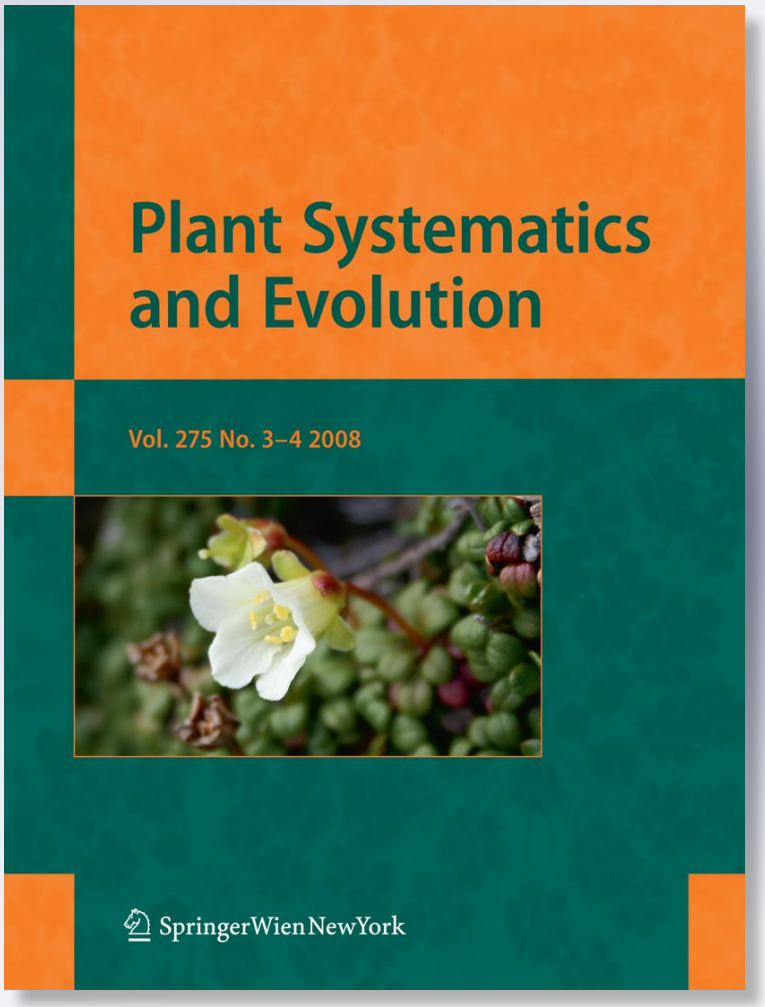
25S–18S rDNA IGS of Capsicum: *molecular structure and comparison*

Mauro Grabiele, Humberto J. Debat, Eduardo A. Moscone & Daniel A. Ducasse

Plant Systematics and Evolution

ISSN 0378-2697
Volume 298
Number 2

Plant Syst Evol (2012) 298:313-321
DOI 10.1007/s00606-011-0546-8



Springer

Plant Syst Evol (2012) 298:313–321
DOI 10.1007/s00606-011-0546-8

ORIGINAL ARTICLE

25S–18S rDNA IGS of *Capsicum*: molecular structure and comparison

Mauro Grabiele · Humberto J. Debat · Eduardo A. Moscone · Daniel A. Ducasse

Received: 9 July 2010 / Accepted: 20 December 2010 / Published online: 28 October 2011

Abstract The primary and secondary structures of the intergenic spacer (IGS) between the 3′-end of 25S ribosomal RNA (rRNA) gene and the 5′-end of 18S rRNA gene are described for the cultivated chili pepper *Capsicum pubescens*. The recognized functional IGS is 2,078 bp in length. According to nucleotide base composition, regulatory elements, and conserved and repeated sequences the IGS can be divided into seven structural regions (SRI–VII). SRI comprises three copies of GAGGTTTTT-like motif, a probable transcription termination site in Solanaceae. At 3′-end, there are 21 bp matching the 18S rDNA. SRII is formed by 47 repeats of CACCATGG-like motif, the shortest repetitive region found in plant rDNA to date. SRIII is highly AT-rich, preceding SRIV, a highly conserved region in Solanaceae containing the transcription initiation site (TIS) TATA-TAAGGGGGG. The external transcribed spacer (ETS) is 966 bp in length. SRV-VII, downstream of the TIS, possesses eight inverted repeats, and three predicted stem-loops show pre-micro RNA (miRNA)-like structural features. Intragenomic variation is presented, and data are compared with characterized Solanaceae 25S–18S rDNA IGS.

M. Grabiele (✉)
Facultad de Ciencas Exactas, Químicas y Naturales,
Universidad Nacional de Misiones (FCEQyN-UNaM),
Félix de Azara 1552, 3300 Posadas, Argentina
e-mail: maurograbiele@conicet.gov.ar

M. Grabiele · E. A. Moscone
Instituto Multidisciplinario de Biología Vegetal (IMBIV),
Universidad Nacional de Córdoba-CONICET, C.C. 495,
5000 Córdoba, Argentina

H. J. Debat · D. A. Ducasse
Instituto de Fitopatología y Fisiología Vegetal (IFFIVE-INTA),
Camino a 60 Cuadras Km 5½, 5119 Córdoba, Argentina

Keywords Solanaceae · *Capsicum* · rDNA IGS

Introduction

In eukaryotes, the ribosomal DNA (rDNA) located at the nucleolar organizer regions (NORs) is composed of the 18S, 5.8S, and 25–28S rRNA genes, the internal transcribed spacer 1 and 2 (ITS), and the intergenic spacer (IGS), forming units arranged in tandem arrays comprising several hundreds to thousands copies that exceed what is physiologically essential (Gruendler et al. 1991; Rogers and Bendich 1987); i.e., there is a variable population of active rRNA according to the physiological needs of the cell, which appears to rely on dosage control, and a second level of control involved in fine-tuning regulation of the transcription of the active rRNA genes in which the IGS might be implicated (Tucker et al. 2010). The rRNA genes, which are transcribed by the RNA polymerase I, are highly conserved throughout plants, animals, and fungi (Gruendler et al. 1991; Zentgraf and Hemleben 1992). In contrast, the IGS, which comprises the sequence between the 3′-end of 25S rRNA and 5′-end of 18S rRNA, is more variable in both length and organization even between closely related taxa. However, its functional role is conserved, containing a transcription initiation site (TIS), a transcription termination site (TTS), and different regulatory elements (Gruendler et al. 1991; Volkov et al. 1999a).

In Solanaceae, a large family that comprises 2,300 species (Martins and Barkman 2005) including important cultivated plants (tobacco, tomato, potato, eggplant, petunia, chili peppers), the rDNA IGS has been characterized in depth in only some taxa of *Solanum* and *Nicotiana* (Schmidt-Puchta et al. 1989; Perry and Palukaitis 1990; Borisjuk and Hemleben 1993; Borisjuk et al. 1994, 1997;

Volkov et al. 1996, 1999a, 2003, 2007; Komarova et al. 2008).

Here, we present a molecular characterization at the primary and secondary structure levels of the rDNA IGS of a cultivated member of chili peppers (*Capsicum pubescens*) and a comparison of our data with those available in the family, in order to delineate the major conserved elements within Solanaceae.

Materials and methods

The plant material, *Capsicum pubescens* Ruiz and Pav. cultivar "locoto rojo" EAM 256, was from Salta, Salta Province, Argentina.

Total DNA was isolated and purified from fresh leaves according to Rogers and Bendich (1994). The rDNA IGS was amplified by polymerase chain reaction (PCR) using two primers designed from consensus sequences of Solanaceae, specific to the 3′-end of the 25S rRNA gene, 5′TAAA TACGCGACGGGGTATTGTAA3′ IGS-1 (24-mer), and the 5′-end of the 18S rRNA gene, 5′GACTACTGGC AGGATCAACCAGGT3′ IGS-2 (24-mer). In the PCR reaction, Taq DNA polymerase, sequencing grade from Promega (USA) was used. Thirty-six amplification cycles were performed, each involving denaturation at 94°C for 1 min, annealing at 57°C for 1 min, and extension at 72°C for 2 min. PCR products were gel-isolated, purified by the GFX kit (Amersham Pharmacia, USA), cloned in pCR2.1-TOPO vector, and transformed into TOP10 One Shot *E. coli* (Invitrogen, USA) according to manufacturer instructions. Restriction enzyme analyses with *Eco*RI (NEB, USA) and *Nco*I (NEB, USA) of the obtained clones were also performed according to the manufacturer′s instructions.

DNA sequencing was performed by Macrogen (Korea). The nucleotide sequence data are available in the GenBank database under accession numbers FJ460246 and FJ460247.

Nucleotide sequences of 25S–18S rDNA of different taxa of Solanaceae and other angiosperms used for comparison were obtained from the NCBI database: *Solanum tuberosum* L. (AF464863, AF464865, AY366530, AY366531, X65489; X67238); *Solanum lycopersicum* L. (AY366528, AY366529, X146639, X52215; X51576); *Nicotiana tabacum* L. (D76443, Y08422; AJ236016); *Nicotiana sylvestris* Speg. & S. Comes, (X76056); *Nicotiana tomentosiformis* Goodsp. (X76055; Y08427); *Solanum etuberosum* Lindl. (AF46413, AF464145); *Solanum bulbocastanum* Dunal (AF464857, AF464859); *Solanum raphanifolium* Cárdenas & Hawkes (AF464853, AF464855); *Solanum maglia* Schltdl. (AF464147); *Solanum circaeifolium* Bitter (AF447414).

Editing and multiple alignments of the nucleotide sequences were performed using MEGA 4.0 (Tamura et al. 2007), DNAStar 7.0 (Lasergene Inc.), BioEdit 7.0.9 (Ibis Biosciences inc.), and Geneious Pro 4.8.3 (Biomatters Ltd.).

Secondary structure analyses from different IGS regions were performed using UNAFold (Markham and Zuker 2008), DINAMelt (Markham and Zuker 2005), MiRAlign (Wang et al. 2005), Structure-Based miRNA analysis tool (Ritchie et al. 2007), VMIR (Grundhoff et al. 2006), and MiPred (Jiang et al. 2007).

Results and discussion

General features of the IGS organization

PCR amplification of the IGS of *C. pubescens* using designed primers from 25S rDNA 3′-end and 18S rDNA 5′-end showed in all cases two distinct products of ca. 1,800 and 2,100 bp, recognized as A- and B-types of IGS, respectively. Cloned fragments from each type of IGS, CpIGS-A3 and CpIGS-B4, were sequenced, showing 1,767 and 2,078 bp, respectively. Dot matrix self-comparison analyses showed that both types of IGS have a block of repetitive DNA (SRII/*Nco*I box, see below), being 183 bp larger in A-type (Fig. 1a, b). In addition, dot matrix alignment of both types of IGS showed high sequence homology, but a deletion of 529 bp in A-type, ca. 150 bp downstream of the 3′-end of the repetitive block, was observed (Fig. 1c). No differences in GC content were found, being 51.84% in A-type and 51.44% in B-type of IGS. Variation other than described is mainly caused by moderate transitions and transversions.

According to nucleotide base composition, regulatory elements, and conserved and repeated sequences, B-type of IGS can be recognized as a functional sequence and is divided into seven structural regions (SRI–VII) (Table 1; Fig. 2). PCR amplification in several clones using designed primers from SRIII 5′-end to SRVI 3′-end showed that the lack of 529 bp is a constant feature in the A-type of IGS including SRIII–VI (Fig. 2). Restriction map of CpIGS-A3 and CpIGS-B4 shows that *Nco*I sites are abundant but only present within the block of repetitive DNA (*Nco*I box SRII). Restriction enzyme analyses in 48 clones with *Eco*RI, to release the cloned fragment from the vector, along with *Nco*I demonstrated that the whole length of A- and B-types of IGS is invariable while the block of repetitive DNA is conserved in base composition and also in length for each type of IGS. Based on the above, CpIGS-A3 and CpIGS-B4 are fairly representatives of A- and B-types of IGS of *C. pubescens*, respectively. Differences in length in SRII, the block of repetitive DNA, between the two types of IGS may be explained by unequal crossover events. In contrast, the lack of 529 bp, a region with major conserved motifs and internal regulatory elements (see

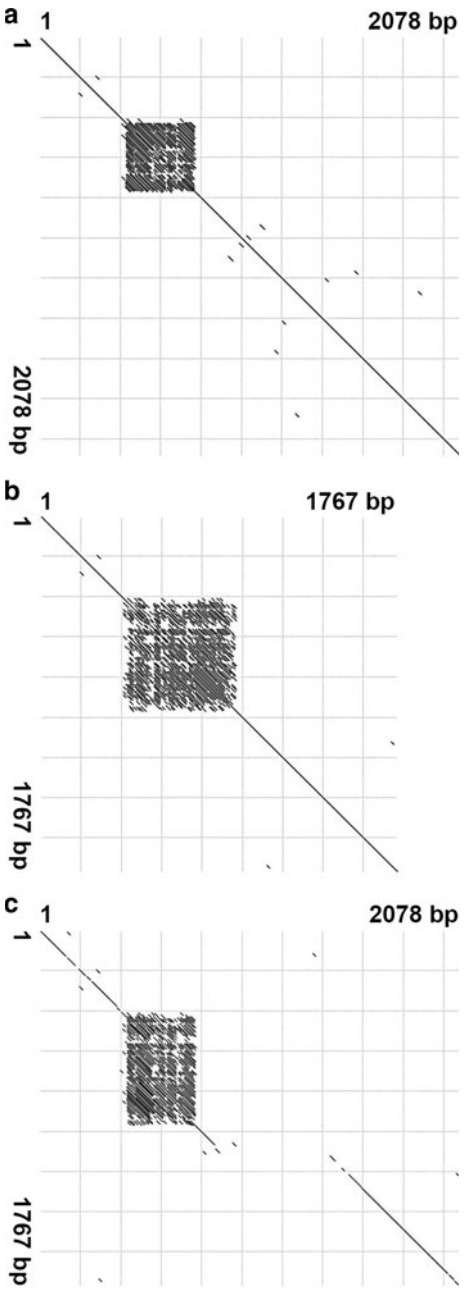


Fig. 1 Dot matrix analysis in the rDNA IGS of *Capsicum pubescens*. **a**, **b** Self-comparison of the CpIGS-B4 and CpIGS-A3 clones, respectively. **c** Alignment of both clones. Analyses were performed using MegAlign software (DNASTAR 7.0, Lasergene Inc.) with threshold of 80% and window size of 16 nucleotides

below), in the A-type of IGS, may be the result of selective pressure on the copy number of functional rDNA sequences.

Structure analysis of B-type IGS

SRI is 393 bp in length and 45.04% GC-rich (Table 1). Analogous regions in the rDNA of other plants also present low GC content (Volkov et al. 1999a). At the 5′-end of the SRI there is a typical pyrimidine-rich block of 17 bp (CCCTCCCCCTTCAATCC) characteristic of higher plants (Schmidt-Puchta et al. 1989; Gruendler et al. 1991; Borisjuk and Hemleben 1993; Borisjuk et al. 1997; Volkov et al. 1999a). The GAGGTTTTT motif has been proposed as a putative transcription termination site (TTS) in Solanaceae (Volkov et al. 1999a), occurring in the SRI of several taxa, i.e., *N. sylvestris*, *N. tabacum*, *N. tomentosiformis*, *S. lycopersicum*, *S. tuberosum*, *S. etuberosum*, *S. maglia*, *S. bulbocastanum*, and *S. raphanifolium*. SRI of *C. pubescens* presents three copies (GAGGTTT, GAGGTTTT, and GAGGTTTTT) analogous to the TTS, being part of oligomer repetitions of 18 bp (Fig. 2). In addition, CTTTT motif adjacent to the 3′-end of the 25S rRNA gene has been proposed as a presumptive 25S rRNA 3′ processing signal (Volkov et al. 1999a). This motif is found in several Solanaceae (those described above), but the corresponding alignment of these regions with those homologous to *C. pubescens* shows that in the latter this motif is absent where it is expected, being replaced by CTCCATCT in both sequenced clones. At 3′-end of SRI (353-374), a 21-bp sequence identical to the 18S rRNA gene (near position 330) of several higher plants was found (Fig. 3). In this regard, numerous small RNAs (sRNAs) have already been described for this 18S rRNA 21-bp sequence in *Arabidopsis* (http://asrp.cgrb.oregonstate.edu) in addition to thousands of sRNAs described for the 18S, 5.8S, and 25S rRNA genes, as targets for transcriptional regulation of rRNA loci (Mayer et al. 2006; Preuss et al. 2008; Backman et al. 2008; Daxinger et al. 2009). Nevertheless, it is worth mentioning that these 21 bp identical to 18S rRNA were found upstream of the putative transcription initiation site, which may imply that it could only presumably be a *cis*-target element of RNA-mediated DNA silencing (Matzke et al. 2009). However recent results of deep sequencing reveal that the spacer region

Table 1 GC content and length of the structural regions of the B-type IGS of *Capsicum pubescens*

IGS SR	I	II	III	IV	V	VI	VII	Total
GC content (%)	45.04	60.27	28.80	35.62	55.63	62.00	59.45	51.44
Length (bp)	393	375	309	73	151	200	577	2,078

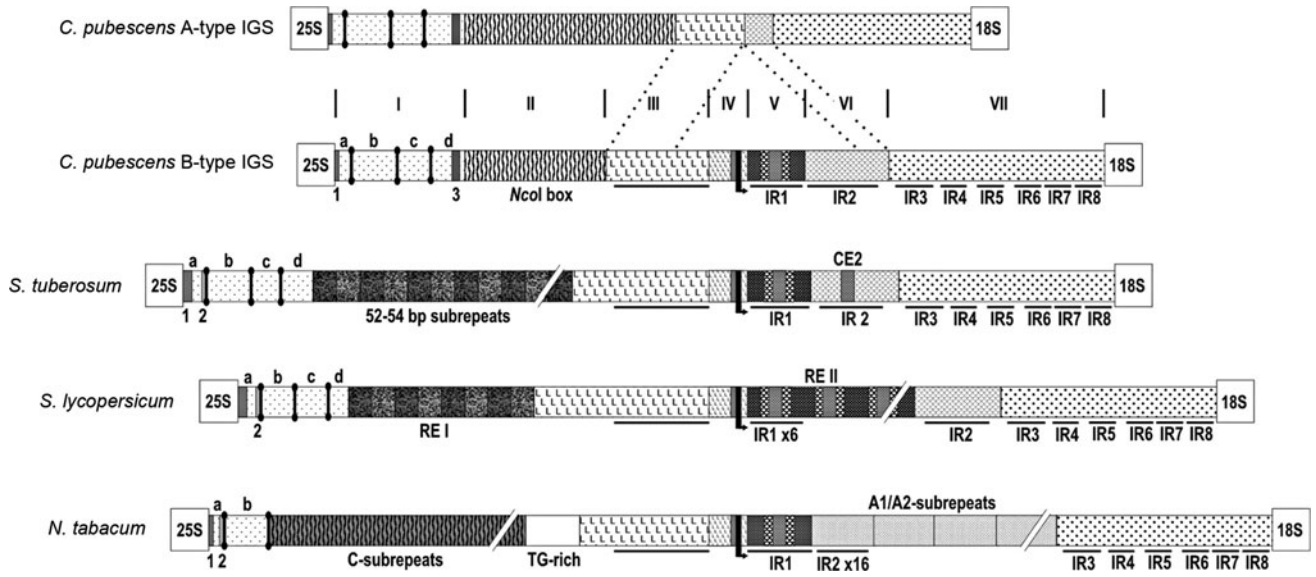


Fig. 2 Organization of the rDNA IGS of *Capsicum pubescens* and comparison with *Solanum tuberosum*, *Solanum lycopersicum*, and *Nicotiana tabacum*. Roman numerals (*I–VII*) correspond to the different structural regions (SRs); 52–54-bp subrepeats defined by Borisjuk and Hemleben (1993). CE2: Conservative Element-2 (40 bp) of B/C-types of ETS of *Solanum* sect. Petota defined by Volkov et al. (2003). RE I and RE II: repetitive elements I and II, respectively, defined by Schmidt-Puchta et al. (1989). C-subrepeats, TG-rich, and A1/A2 subrepeats defined by Borisjuk et al. (1997). *1* pyrimidine residues. *2* CTTTT motif. *3* 21-bp region, 100% identical to 18S rRNA. SRI subregions; whole similarity in a: 63.2%, b: 46.3%, c: 59.6%, d: 60.2%. GAGGTTTTT and GAGGTTTTT-like motif. AT-rich region; whole similarity in the 300-bp stretch (*underlined*): 57.1%. Highly similar region (84.5%), containing a poly-T stretch (*grey*) and the putative TIS (*black with arrow*). IR: inverted repeats. ca. 80-bp conserved region within the IR1; whole similarity: 78.6%; at its middle part, the Conservative Element-1 (CE1, 40 bp) defined by Volkov et al. (2003) for the ETS of *Solanum* sect. Petota is shown (*grey*)

```
IGS SRI C. pubescens       ATAGTGGCCTACCATGGTGGT
18S C. annuum EF564281     .....................
18S Petunia AJ236020       .....................
18S N. tabacum AJ236016    .....................
18S S. lycopersicum X51576 .....................
18S S. tuberosum X67238    .....................
18S S. melongena X63311    .....................
18S A. thaliana X16077     .....................
18S T. aestivum X07841     ....G................
```

Fig. 3 18S rRNA gene region of different angiosperms and identical site in the IGS SRI of *Capsicum pubescens*

could also be a generator of small RNAs, although the source of the spacer transcripts remain unclear (Preuss et al. 2008).

IGS SRII is 375 bp in length and 60.21% GC-rich (Table 1). It is formed by 47 repeats of 8-bp sequences similar to the most abundant motifs, CACCATGG and CACCAAGG (Table 2), the shortest ones found to date in plant rDNA. C1-subrepeats of SRII of *Nicotiana* are 10 bp in length, and the abundant CAGGACATGG sequence (Borisjuk et al. 1997; Volkov et al. 1999a, b) resembles the CACCATGG motif of *Capsicum*. SRII of *C. pubescens* is also called the *Nco*I box, because the restriction enzyme recognition site CCATGG is present 15 times. This motif is present 22 times in *N. tabacum* SRII (1,287 bp, Borisjuk et al. 1997), and the CCnnGG recognition site for *Bsa*JI is present 26 times in 52-54-bp subrepeats of SRII of

Table 2 Repeat sequences and frequencies in the IGS SRII of *Capsicum pubescens*

Oligomer	No.	Oligomer	No.
CACCATGG	12	CATCAAGG	1
CACCAAGG	9	CAACATGG	1
CACCGTGG	3	CCCCATGG	1
CACGAAGG	2	C-CCAAAC	1
CACTGTGG	2	CCTTCGCC	1
CACCAGGG	1	CGCCATGG	1
CACCATGA	1	GCACTGTG	1
CACCAAGT	1	AACCAAGG	1
CACCAAGA	1	AACCATGG	1
CACCAATA	1	AACCAAAG	1
CACCGTGC	1	AATCATGG	1

S. tuberosum (1,296 bp, Borisjuk and Hemleben 1993), compared with 27 in *N. tabacum* and 29 in *C. pubescens*. The 52–54-bp subrepeats of RE I of *S. lycopersicum* (SRII, 420 bp, Schmidt-Puchta et al. 1989) present few *Bsa*JI sites, but some similar oligomer repeated sequences are shared with SRII of *C. pubescens* (e.g., CATGTCATCA, CATGCCATCA, CATGCCACCA). SRII of *C. pubescens*,

S. lycopersicum, *S. tuberosum*, and *Nicotiana* species present similar high GC content and some oligomer sequence similarity. These repeats may have arisen by cycles of amplification, deletions, and base substitutions from common ancestral sequences. Repetitive regions upstream of the TIS are found in the rDNA IGS of several higher plants (Barker et al. 1988; Schmidt-Puchta et al. 1989; Gruendler et al. 1991; Borisjuk and Hemleben 1993; Borisjuk et al. 1997), acting as enhancers of transcription by rRNA PolI (Hemleben and Zentgraf 1994).

IGS SRIII is 309 bp in length, and its major feature is a high AT content (28.80% GC-rich; Table 1), just before the putative TIS. An AT-rich region upstream of the TIS is found also in several plants (Kelly and Siegel 1989; Gruendler et al. 1991; Borisjuk et al. 1997; Volkov et al. 1999b), probably affecting the initiation of transcription by reducing the local double-strand stability. Some sequence similarity (57.1%) is found in a stretch of ca. 300 bp between *C. pubescens* and AT-rich regions upstream of the TIS of *N. tabacum* (386 bp, 83% AT; Borisjuk et al. 1997), *S. tuberosum* (459 bp, 73% AT; Borisjuk and Hemleben 1993), and *S. lycopersicum* (599 bp, 73% AT; Schmidt-Puchta et al. 1989) (Fig. 2).

IGS SRIV, which is 73 bp in length and AT-rich (35.62% GC content; Table 1), presents the putative TIS for RNA PolI conserved in Solanaceae, the TATA-TAAGGGGGG motif. In the flanking regions of the TIS of *C. pubescens*, *N. tabacum*, *S. tuberosum*, *S. lycopersicum*, and several Solanaceae, high sequence similarity (ca. 85%) is found. Just 5–6 bp upstream of the putative TIS, a poly-T stretch 6–11 bp in length appears in those taxa, probably being an additional TTS (Delcasso-Treymousaygue et al. 1988). The external transcribed spacer (ETS) of *Capsicum*, of 966 bp in length, is shorter than any other described in Solanaceae [973–1,041 bp in *Solanum* sect. Petota (Volkov et al. 2003), 1,691–1,695 bp in *S. lycopersicum* (Schmidt-Puchta et al. 1989; Perry and Palukaitis 1990; Komarova et al. 2004), ca. 1,400–3,005 bp in *Nicotiana* (Borisjuk et al. 1997; Volkov et al. 1996, 1999a, b)]. As in other IGS of Solanaceae previously described, and unlike *A. thaliana* (Gruendler et al. 1991), no spacer promoters have been found in *Capsicum*.

IGS SRV is 151 bp in length and 55.63% GC-rich (Table 1). Its major feature is an inverted repeat region of 143 bp which includes 80 bp also conserved in other Solanaceae at its core region (Figs. 4, 5). This conserved ca. 80 bp region has been previously described as a short 53-bp element shared by *Nicotiana* species, *S. tuberosum*, and different repeats of RE II of *S. lycopersicum* (Borisjuk et al. 1997). Moreover, the 40-bp Conservative Element-1 (CE1) of *Solanum* sect. Petota (Volkov et al. 2003; Komarova et al. 2008) is found within this region also. Volkov et al. (1999a) described a possible stem-loop structure of 113 bp that includes this region in *N. sylvestris*. In fact, folding analysis showed that the inverted repeat conserved region of ca. 80 bp in conjunction with flanking sequences can form similar stem-loop structures in *C. pubescens* (143 bp), *N. sylvestris* (113 bp), *N. tabacum* and *N. tomentosiformis* (162 bp), *S. lycopersicum* (139 bp), and ETS variants A, B, and C of *Solanum* sect. Petota (Volkov et al. 2003), with 186 bp for *S. tuberosum* (Figs. 4, 5). In *Solanum*, the repeated sequences flanking this ca. 80-bp conserved region which are similar to the 5′-end and 3′-end of this latter do not affect the folding of the secondary structure.

In addition, this ca. 80-bp stem-loop structure (Fig. 5) has distinctive features related to micro-RNAs precursors (pre-miRNAs) described by Yousef et al. (2006), Ritchie et al. (2007), and Jiang et al. (2007). Even though the putative resultant miRNA (https://bioinfo.wistar.upenn.edu/miRNA/miRNA/index.htm) is not present in the miRBase (http://microrna.sanger.ac.uk/sequences/), which includes all the described miRNAs, nor in the Tomato Functional Genomics Database for putative miRNAs (http://ted.bti.cornell.edu/cgi-bin/TFGD/sRNA/mi-RNA.cgi), there is emerging evidence that, additionally to the rRNA loci, the ETS is per se a generator of several small interference RNAs that in turn would trigger de novo cytosine methylation of homologous regions resulting in large-scale silencing of rRNA gene loci (Preuss et al. 2008). It is worth mentioning that Schmidt-Puchta et al. (1989) described a region of the complementary strand of RE II of *S. lycopersicum* IGS, with 19/25 nucleotide identity with the potato spindle tuber viroid (PSTV). In fact, the base pairing described by those authors would take place at the complementary strand of the potential pre-miRNA-like element described here, from 5′-end position 1–25. Further DNA sequence comparison of the entire IGS of *C. pubescens* with viruses of several genera that affect *Capsicum*, i.e., *Alfamovirus*, *Begomovirus*, *Comovirus*, *Cucumvirus*, *Curtovirus*, *Potyvirus*, *Tobamovirus*, and *Tospovirus*, did not show significant similarity.

On the other hand, this ca. 80-bp inverted repeat conserved region in Solanaceae could be an internal rRNA regulatory element, as for the inverted repeats A0 and A1 of *Saccharomyces cerevisiae*, which act as processing signals for RNase III during the liberation of mature rRNA (Lafontaine and Tollervey 2001). Additionally, the CE within this ca. 80-bp repeat conserved region (Figs. 4, 5) appears independently amplified, alone or within subrepeats, in different groups of *Solanum* species (Komarova et al. 2008). In conjunction, the 45S rDNA of *Solanum* allopolyploids containing IGS with more CE downstream of TIS, which is an evolutionary novelty, appears transcriptionally dominant over the 45S rDNA with ancestral organization, indicating feasible selective advantage and reinforcing the functional role for

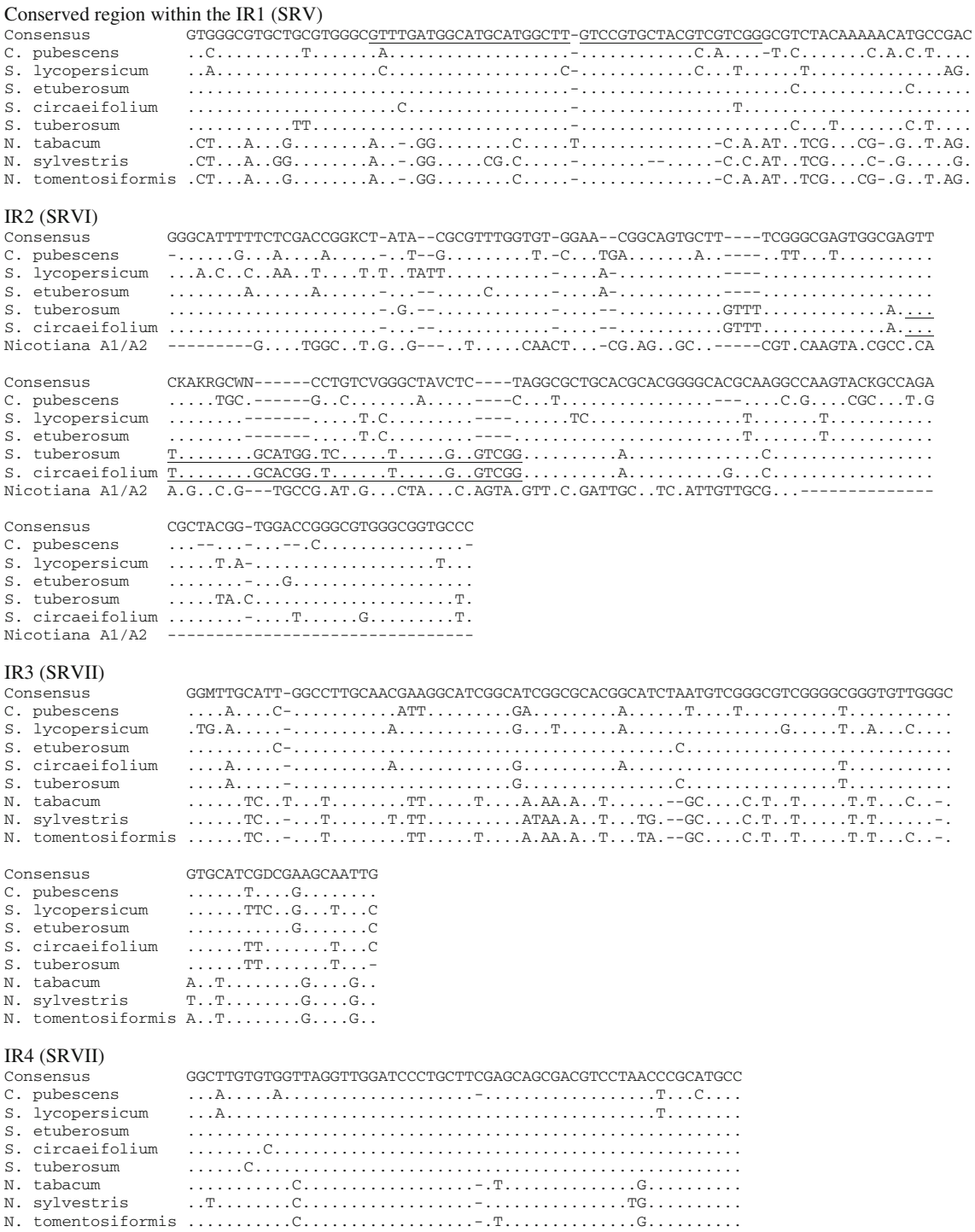


Fig. 4 Inverted repeats (IR) found in the ETS of *Capsicum pubescens* and comparison with consensus homologous regions in different Solanaceae. *Solanum etuberosum*, *S. circaeifolium*, and *S. tuberosum* represent ETS variants A, B and C of *Solanum* sect. Petota (Volkov et al. 2003), respectively. *Underlined regions* in the first two alignments correspond to the CE1 and CE2 of Volkov et al. (2003), respectively. For *S. lycopersicum* the consensus between different repeats of RE II (Schmidt-Puchta et al. 1989) is also shown

this sequence in Solanaceae (Komarova et al. 2004, 2008; Volkov et al. 2007).

Based on the above-mentioned data, the inverted repeat conserved region of ca. 80 bp in conjunction with the CE of Solanaceae merit more detailed analysis.

IGS SRVI is 200 bp in length and has 62% GC content (Table 1). As in the preceding structural region, its major feature is an inverted repeat of 164 bp (Fig. 5). SRVI of *C. pubescens* presents some similarity (57.5%) to the corresponding regions of *S. lycopersicum* (215 bp; Schmidt-Puchta

IR5 (SRVII)

```
Consensus          GGTWTCCTGTGCTGCATACCTAATGCCTAGGCATTATNCACGTGCAATCGGTCGCC
C. pubescens       A.........................C.........-..A..TT.....A.....
S. lycopersicum    AAC...................C...........G...........A........
S. etuberosum      ...C................G.....G.........C..G.................
S. circaeifolium   ....................................C..G.................
S. tuberosum       ....................................C...................
N. tabacum         .....T.....T................C.........-..A.CA........T...
N. sylvestris      .....T.....T................A.........-..A.C.........T...
N. tomentosiformis .....T.....T................C.........-..A.CA........T...
```

IR6 (SRVII)

```
Consensus          AGCYGCTCTYGCGTCCCAC-GCCTTCC-TCGCTTCGTCGTGCGATGGCGTGGTCCGTGAGCGGCG-GCT
C. pubescens       ...T.......T.......-......A-...................T...................-A..
S. lycopersicum    ..A................-.......C.....................................-C..
S. etuberosum      ......C..........G.-.......-.......................C......C........-...
S. circaeifolium   ...................-.......-.....................................C........-...
S. tuberosum       .........T.........-.......-.....................................C........-...
N. tabacum         ...T..AG.T.T.......-...A...-...................T.TA.....A...CTA.TAT...
N. sylvestris      ...T.TAG.T.T......GA.T.G...-.T....T.C.T.........T......A...CTA.TAT...
N. tomentosiformis ...T..AG.T.T.......-...A...-...................T.TA.....A...CTA.TAT...
```

IR7 (SRVII)

```
Consensus          CGGATTCGGTAGACGCAGTGGGCATGGGGYCTTCACCGGCTCCTATCTGCCCAAAACGAATGC
C. pubescens       .....G..............T.......T.............................-...
S. lycopersicum    .....A.....................................T.....................
S. etuberosum      ..........T.............................................G.........
S. circaeifolium   ..................................G.....................A.....
S. tuberosum       ....................................................................
N. tabacum         ..........A.....A...........T.....TT..................--......
N. sylvestris      ..........A.....A...........T.....TT.........-....--..--......
N. tomentosiformis ..........A.....A...........T.....TT..................--......
```

IR8 (SRVII)

```
Consensus          GACGGCCGCGCTCGCCTTGGACCCGGCCGTGCCCTYACGGKCGCGCCGGGCTCATGCGGYGCGCGGCGTC
C. pubescens       ....................T.T.T.............T..................T..-.......
S. lycopersicum    ...T.....................A........GA.A.....................................
S. etuberosum      ..........................C....GA........T.....G.....................
S. circaeifolium   ........................A.....................................................
S. tuberosum       .......................T...................G.....................
N. tabacum         .....T..T.....T.......TT.......G...TTG..T..-...AT..........T..-..A....
N. sylvestris      .....T..T.....T.......T-...G...-...TTG..T..-...AT.........AT..-.......
N. tomentosiformis .....T..T.....T.......TT.......G...TTG..T..-...AT..........T..-..A....
```

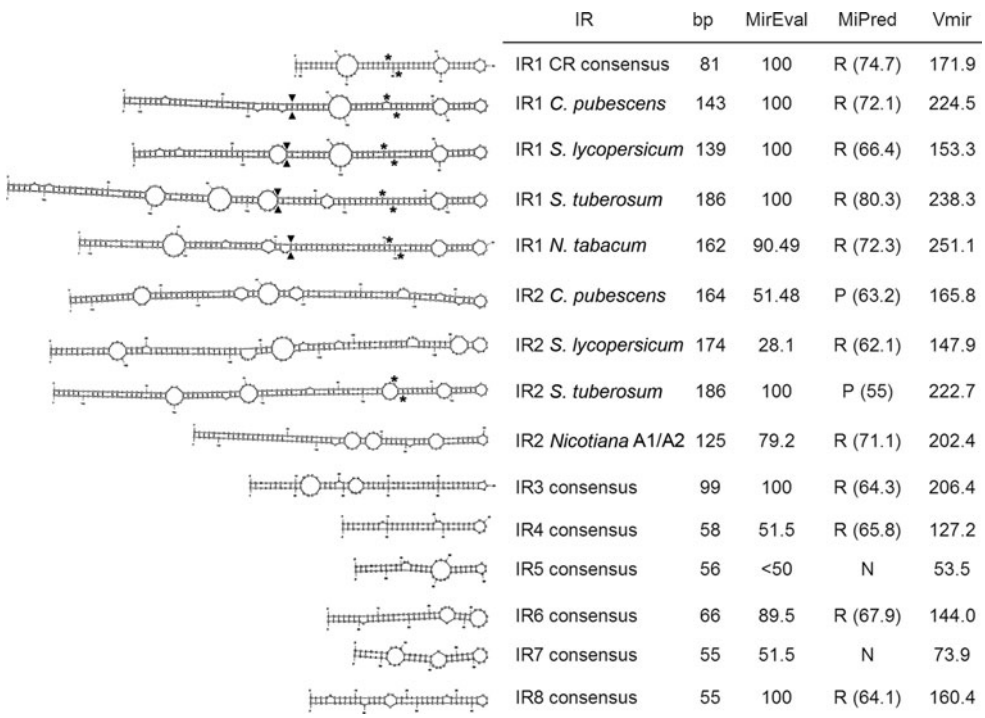
Fig. 4 continued

et al. 1989) and *S. tuberosum* (242 bp; Borisjuk and Hemleben 1993) (Fig. 2), but no significant similarity is found between them and A1/A2 repeats of *Nicotiana* (Borisjuk et al. 1997; Volkov et al. 1996, 1999a, b) (Fig. 4), probably denoting early divergence from common ancestral sequences. In those regions, inverted repeats are also found in *S. lycopersicum* (174 bp), A1/A2 of *Nicotiana* (125 bp), and ETS variants A (174 bp) and B/C (186 bp) of *Solanum* sect. Petota, all of which are potentially able to form stem-loop structures as in *Capsicum* (Figs. 4, 5). Conservative Element-2 (CE2) of ETS variants B and C of *Solanum* sect. Petota is a 40-bp sequence identical to CE1 (Volkov et al. 2003), and both of them occur in the middle region of the predicted stem-loop (Figs. 4, 5). The analysis performed to evaluate the ability of the above-described possible stem-loops to be pre-miRNAs shows contradictory results (Fig. 5).

IGS SRVII of *C. pubescens* is 577 bp in length and 59.45% GC-rich (Table 1). It is the most conserved region of the ETS in Solanaceae, with 75.30% sequence similarity between *Capsicum* and the corresponding regions of *S. tuberosum* (590 bp; Borisjuk and Hemleben 1993), *S. lycopersicum* (594 bp; Schmidt-Puchta et al. 1989), *N. tabacum* (567 bp; Borisjuk et al. 1997), and several related taxa (Volkov et al. 1996, 1999a, b, 2003). This highly conserved region presents six different inverted repeats covering 67% of its extension, and two of them (IR3 and IR8) show the features of pre-miRNAs (Figs. 4, 5).

In this regard, previous whole-genome analyses of *Homo sapiens* and *Arabidopsis* as model organisms found that hundreds of thousands of stem-loops actually resemble pre-miRNAs, but just a few of them were confirmed as biologically active (Ritchie et al. 2007). Nevertheless, in order to test the pre-miRNA-like elements proposed in this report, Northern blotting of both precursor and expected mature microRNAs is needed.

Furthermore, close interaction between 24-bp small interference RNAs (siRNAs) derived from IGS transcripts



IR	bp	MirEval	MiPred	Vmir
IR1 CR consensus	81	100	R (74.7)	171.9
IR1 *C. pubescens*	143	100	R (72.1)	224.5
IR1 *S. lycopersicum*	139	100	R (66.4)	153.3
IR1 *S. tuberosum*	186	100	R (80.3)	238.3
IR1 *N. tabacum*	162	90.49	R (72.3)	251.1
IR2 *C. pubescens*	164	51.48	P (63.2)	165.8
IR2 *S. lycopersicum*	174	28.1	R (62.1)	147.9
IR2 *S. tuberosum*	186	100	P (55)	222.7
IR2 *Nicotiana* A1/A2	125	79.2	R (71.1)	202.4
IR3 consensus	99	100	R (64.3)	206.4
IR4 consensus	58	51.5	R (65.8)	127.2
IR5 consensus	56	<50	N	53.5
IR6 consensus	66	89.5	R (67.9)	144.0
IR7 consensus	55	51.5	N	73.9
IR8 consensus	55	100	R (64.1)	160.4

Fig. 5 Predicted foldings of the inverted repeats (IR) described in Fig. 4 and evaluation by different pre-miRNA predictor softwares. MirEval (Ritchie et al. 2007): scale 1–100, >90 indicates good candidates to be pre-miRNA. MiPred (Jiang et al. 2007): *R* real microRNA precursor, *P* pseudo-pre-miRNA, *N* not pre-miRNA; between brackets, the prediction confidence (%). Vmir (Grundhoff et al. 2006): scores above 100 indicate good candidates. *Arrowheads* in IR1 denote the limits from right to left of the conserved region (CR). *Asterisks* in IR1 and IR2 denote the limits from right to left of the CE1 and CE2 of Volkov et al. (2003), respectively. Stem-loops are shown from left to right: 5′ to 3′. For other details, see Fig. 4. The predicted value corresponds to the core region of the stem-loop with the best hairpin score

and silencing at the rRNA loci was described by Preuss et al. (2008). According to McStay and Grummt (2008), this would represent a new level of regulation of rRNA genes which may unveil the complex concert of ribosomal genes switching on and off associated with development and variable physiological states that have been observed. We postulate that the population of these intergenic non-coding transcripts might be directly related to the complex structures found in the ETS, which may represent the core triggers of the RNAi silencing machinery, by recruiting Dicer-like (DCL) activity. Recognition of key features associated with the primary and secondary structure of the IGS might reveal in the near future the requirements that in turn mediate the exquisite basis of selective silencing and activation.

Additional evidence of an intimate association between high-complexity structures and functionality and coevolution of regulatory proteins and structural regions in the IGS of 45S rDNA of a number of eukaryotes (Baldridge et al. 1992; Zentgraf and Hemleben 1992; Ricci et al. 2008) supports the probable role of all the inverted repeats able to form stem-loop structures in the ETS of *Capsicum* (IR1-IR8; SRV-VII) and related Solanaceae.

Acknowledgments This study was supported by grant no. PICT 20196 of the Agencia Nacional de Promoción Científica y Tecnológica de Argentina (ANPCyT) and Consejo Nacional de Investigaciones Científicas y Técnicas de Argentina (CONICET) as part of the doctoral thesis of M.G. Additional support came from doctoral and postdoctoral research fellowships of CONICET to H.J.D. and M.G., respectively, who contributed equally to this work. H.J.D. carried out the DNA methods with the aid of M.G. and both performed the bioinformatic and secondary structures analyses. M.G. generated the figures and wrote the manuscript, and H.J.D., D.A.D., and E.A.M. revised it. The authors thank Prof. Elba Villanueva for critical reading of the manuscript.

References

Backman TW, Sullivan CM, Cumbie JS, Miller ZA, Chapman EJ, Fahlgren N, Givan SA, Carrington JC, Kasschau KD (2008) Update of ASRP: the *Arabidopsis* Small RNA Project database. Nucleic Acids Res 36:D982–D985

Baldridge GD, Dalton MW, Fallon AM (1992) Is higher-order structure conserved in eukaryotic ribosomal DNA intergenic spacers? J Molec Evol 35:514–523

Barker RF, Harberd NP, Jarvis MG, Flavell RB (1988) Structure and evolution of the intergenic region in a rDNA repeat union of wheat. J Molec Biol 201:1–17

Borisjuk N, Hemleben V (1993) Nucleotide sequence of *S. tuberosum* rDNA intergenic spacer. Pl Molec Biol 21:381–384

Borisjuk N, Borisjuk L, Petjuch G, Hemleben V (1994) Comparison of nuclear ribosomal RNA genes among *Solanum* species and other *Solanaceae*. Genome 37:271–279

Borisjuk NV, Davidjuk YM, Kostishin SS, Miroshnichenco GP, Velasco R, Hemleben V, Volkov RA (1997) Structural analysis of rDNA in the genus *Nicotiana*. Plant Mol Biol 35:655–660

Daxinger L, Kanno T, Bucher E, van der Winden J, Naumann U, Matzke AJ, Matzke MA (2009) Stepwise pathway for biogenesis of 24-nt secondary siRNAs and spreading of DNA methylation. EMBO J 28:48–57

Delcasso-Treymousaygue D, Grellet F, Panabieres F, Ananiev E, Delseny M (1988) Structural and transcripcional characterization of the external spacer of a ribosomal RNA nuclear gene from a higher plant. Eur J Biochem 172:767–776

Gruendler P, Unfried K, Pascher K, Schweizer D (1991) rDNA intergenic region from *Arabidopsis thaliana*: structural analysis, intraspecific variation and functional implications. J Molec Biol 221:1209–1222

Grundhoff A, Sullivan CS, Ganem D (2006) A combined computational and microarray-based approach identifies novel microRNAs encoded by human gamma-herpesviruses. RNA 12:733–750

Hemleben V, Zentgraf U (1994) Structural organization and regulation of transcription by RNA polymerase I of plant nuclear ribosomal RNA genes. In: Nover L (ed) Results and problems in cell differentiation 20: plants promoters and transcription factors. Springer, Berlin/Heidelberg, pp 3–24

Jiang P, Wu H, Wang W, Ma W, Sun X, Lu Z (2007) MiPred: calssification of real and pseudo microRNA precursors using random forest prediciton model with combined features. Nucleic Acid Res 35:W339–W344

Kelly R, Siegel A (1989) The *cucurbita maxima* ribosomal DNA intergenic spacer has a complex structure. Gene 80:239–248

Komarova NY, Grabe T, Huigen DJ, Hemleben V, Volkov RA (2004) Organization, differential expression and methylation of rDNA in artificial *Solanum* allopolyploids. Plant Mol Biol 56:439–463

Komarova NY, Grimm GW, Hemleben V, Volkov RA (2008) Molecular evolution of 35S rDNA and taxonomic status of *Lycopersicon* within *Solanum* sect. *Petota*. Plant Syst Evol 276:59–71

Lafontaine D, Tollervey D (2001) Ribosomal RNA Encyclopedia of life sciences. Wiley, Chichester

McStay B, Grummt I (2008) The epigenetics of rRNA genes: from molecular to chromosome biology. Annu Rev Cell Dev Biol 24:131–157

Markham NR, Zuker M (2005) DINAMelt web server for nucleic acid melting prediction. Nucleic Acids Res 33:W577–W581

Markham NR, Zuker M (2008) UNAFold: software for nucleic acid folding and hybriziation. In: Keith JM (ed) Bioinformatics, volume II structure, functions and applications, number 453 in methods in molecular biology, chapter 1. Humana, Totowa, pp 3–31

Martins TR, Barkman TJ (2005) Reconstruction of *Solanaceae* phylogeny using the nuclear gene SAMT. Syst Bot 30:435–447

Matzke M, Kanno T, Daxinger L, Huettel B, Matzke AJ (2009) RNA-mediated chromatin-based silencing in plants. Curr Opin Cell Biol 21:367–376

Mayer C, Schmitz K-M, Li J, Grummt I, Santoro R (2006) Intergenic transcripts regulate the epigenetic state of rRNA genes. Mol Cell 22:351–361

Perry KL, Palukaitis P (1990) Transcription of *S. lycopersicum* ribosomal DNA and the organization of the intergenic spacer. Molec Gen Genet 221:102–112

Preuss SB, Costa-Nunes P, Tucker S, Pontes O, Lawrence RJ, Mosher R, Kasschau KD, Carrington JC, Baulcombe DC, Viegas W, Pikaard CS (2008) Multimegabase silencing in nucleolar dominance involves siRNA-directed DNA methylation and specific methylcytosine-binding proteins. Mol Cell 32:673–684

Ricci A, Scali V, Passamonti M (2008) The IGS-ETS in *Bacillus* (Insecta Phasmida): molecular characterization and the relevance of sex in ribosomal DNA evolution. BMC Evol Biol 8:278

Ritchie W, Legendre M, Gautheret D (2007) RNA stem-loops: to be or not to be cleaved by RNAse III. RNA 13:457–462

Rogers SO, Bendich AJ (1987) Ribosomal RNA genes in plants: variability in copy number and in intergenic spacer. Plant Mol Biol 9:509–520

Rogers SO, Bendich AJ (1994) Extraction of total cellular DNA from plant, algae and fungi In: Stanton BG, Schilperoort RA (eds) A plant molecular biololgy manual. Kluwer Academic Publ, Dordrecht, pp D1/1-8

Schmidt-Puchta W, Guenther I, Saenger HL (1989) Nucleotide sequence of the intergenic spacer (IGS) of the *S. lycopersicum* ribosomal DNA. Plant Mol Biol 13:251–253

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596–1599

Tucker S, Vitins A, Pikaard C (2010) Nucleolar dominance and ribosomal RNA gene silencing. Curr Opin Cell Biol 22:351–356

Volkov R, Kostishin S, Ehrendorfer F, Schweizer D (1996) Molecular organization and evolution of the external transcribed rDNA spacer region in two diploid relatives of *Nicotiana tabacum* (Solanaceae). Plant Syst Evol 201:117–129

Volkov RA, Bachmair A, Panchuk II, Kostyshyn SS, Schweizer D (1999a) 25S–18S rDNA intergenic spacer of Nicotiana sylvestris (Solanaceae): primary and secondary structure analysis. Plant Syst Evol 218:89–97

Volkov RA, Borisjuk NV, Panchuk II, Schweizer D, Hemleben V (1999b) Elimination and rearrangement of parental rDNA in the allotetraploid *Nicotiana tabacum*. Mol Biol Evol 16:311–320

Volkov RA, Komarova NY, Panchuk II, Hemleben V (2003) Molecular evolution of rDNA external transcribed spacer and phylogeny of sect. *Petota* (genus *Solanum*). Molec Phyl Evol 29:187–202

Volkov RA, Komarova NY, Hemleben (2007) Ribosomal DNA in plant hybrids: inheritance, rearrangement, expression. Syst Biod 5:261–276

Wang X, Zhang J, Li F, Gu J, He T, Zhang X, Li Y (2005) MicroRNA identification based on sequence and structure alignment. Bioinformatics 21:3610–3614

Yousef M, Nebozhyn M, Shatkay H, Kanterakis S, Showe LC, Showe MK (2006) Combining multi-species genomic data for microRNA identification using a Naïve Bayes classifier. Bioinformatics 22:1325–1334

Zentgraf U, Hemleben V (1992) Complex formation of nuclear proteins with the RNA polymerase I promoter and repeated elements in the external transcribed spacer of *Cucumis sativus* ribosomal DNA. Nucl Acids Res 20:3685–3691