ORIGINAL ARTICLE

Unraveling the sequence dynamics of the formation of genus-specific satellite DNAs in the family solanaceae

S-H Jo[1,4,5], H-M Park[2,3,5], S-M Kim[1,4], HH Kim[3], C-G Hur[1] and D Choi[2]

[1]*Plant Systems Engineering Research Center, KRIBB, Daejeon, Korea;* [2]*Department of Plant Science and Plant Genomics and Breeding Institute, Seoul National University, Seoul, Korea;* [3]*Department of Life Science, Sahmyook University, Seoul, Korea and* [4]*Seeders Inc., Daejeon, Korea*

Tandemly repeated DNAs, referred to as satellite DNAs, often occur in a genome in a genus-specific manner. However, the mechanisms for generation and evolution for these sequences are largely unknown because of the uncertain origins of the satellite DNAs. We found highly divergent genus-specific satellite DNAs that showed sequence similarity with genus-specific intergenic spacers (IGSs) in the family Solanaceae, which includes the genera *Nicotiana, Solanum* and *Capsicum*. The conserved position of the IGS between 25S and 18S rDNA facilitates comparison of IGS sequences across genera, even in the presence of very low sequence similarity. Sequence comparison of IGS may elucidate the procedure of the genesis of complex monomer units of the satellite DNAs. Within the IGS of *Capsicum* species, base substitutions and copy number variation of subrepeat monomers were causes of monomer divergence in IGS sequences. At the level of inter-generic IGS sequences of the family Solanaceae, however, genus-specific motif selection, motif shuffling between subrepeats and differential amplification among motifs were involved in formation of genus-specific IGS. Therefore, the genus-specific satellite DNAs in Solanaceae plants can be generated from differentially organized repeat monomers of the IGS rather than by accumulation of mutations from pre-existent satellite DNAs.
Heredity (2011) **106,** 876–885; doi:10.1038/hdy.2010.131; published online 10 November 2010

Introduction

Although some plants with large genomes are polyploid (Clarkson *et al.*, 2005), most differences in size and composition arise from differences in the repetitive DNA complement of the species (Olmstead *et al.*, 1999). Repetitive sequences exhibit diverse compositions in different genomes owing to a much faster evolutionary rate than that of coding sequences, and, therefore, changes in repetitive sequences may be particularly useful in understanding genome evolution. One abrupt change that occurs in genomic sequences during evolution is the sudden appearance or disappearance of tandem repetitive sequences. These sequences are amplified rapidly and produce repeats specific to a genus, species or even an individual chromosome (Kishii and Tsujimoto, 2002). Several mechanisms have been proposed to explain this rapid expansion of satellite DNA, including a rolling circle mechanism, unequal crossing over, segmental duplication and replication slippage (Hourcade *et al.*, 1973; Smith, 1976; Tautz and Renz, 1984; Ma and Jackson, 2006). Copy number changes are accompanied by rapid evolution of nucleotide sequences and can explain the species specificity of satellite profiles according to the 'library model' (Salser *et al.*, 1976; Ugarkovic and Plohl, 2002). To alter specific profiles of satellite repeats, the interplay of stochastic events and selective pressure also represents a factor for inter-satellite variability in a set of related repeats that have been differentially amplified in a group of taxa (Mestrovic *et al.*, 2006). These mechanisms, however, are not entirely sufficient to explain the phenomenon of genus-specific satellite profiles.

Satellite DNA sequence divergence has been studied in related taxa that share the same satellite DNA. These studies revealed that satellite DNA sequence divergence proceeds in a gradual manner mostly because of the accumulation of nucleotide substitutions (Hemleben *et al.*, 2007; Plohl *et al.*, 2008), as deletions and insertions represent rare events in these sequences (Pons *et al.*, 2004; Mravinac *et al.*, 2005a, b). However, this comparison is limited to satellite DNAs that show sequence similarity in closely related species, as the origins of these sequences cannot be determined. In other words, whether these sequences result from a common sequence that then becomes highly divergent via accumulation of mutations or whether these sequences result from distinct satellite DNAs remains indistinguishable. Thus, identification of the origins of the satellite DNA allows a better understanding of species-specific satellite repeat development.

The IGS-derived satellite DNA may offer good material to study the mechanism of formation of

species-specific satellite DNA among related species. First, highly amplified satellite DNAs with sequence homology to the IGS subrepeats of rDNA have been reported to occur in dispersed patterns over several chromosomes in several plants, including legumes (Flaquet et al., 1997), potatoes (Stupar et al., 2002), tomatoes (Jo et al., 2009) and tobacco (Lim et al., 2004). In our previous work, development of satellite DNA from the IGS of rDNA was demonstrated using intermediate sequences in the tomato genome (Jo et al., 2009). Second, the position of IGS in 45S rDNA is well conserved between the 25S rRNA and 18S rRNA genes. Knowledge of the origin of satellite DNAs may provide a better understanding of the transition of highly divergent satellite DNAs among taxa. Third, comparative analysis of IGS divergence has been intensively conducted in most plant species, including solanaceous species (Volkov et al., 2003; Lim et al., 2004), as well as polyploidy species (Wendel et al., 1995). Thus, we speculated that whether IGS-derived satellite DNA commonly occurs in Solanaceae members, these molecules could serve as a model for investigating the molecular changes involved in the constitution of highly divergent satellite DNAs among related species.

Therefore, the presence of satellite DNA was examined in the Capsicum genus, and genus specificity was confirmed using hybridization in three Solanaceae genera: Capsicum (pepper), Solanum (tomato and potato) and Nicotiana (tobacco). Analysis of the IGS sequences and the satellite repeats revealed that the genus-specific satellite DNAs was generated by explosive amplification of distinct IGS sequences rather than gradual accumulation of base substitutions from common satellite DNA. The major mechanism involved in construction of the novel monomer of the IGS in Solanaceae plants was then analyzed by comparing the current IGS variants across genera.

Materials and methods

Cloning and sequencing of IGS regions

To clone the IGS regions of the five Capsicum species (C. baccatum cv. Pl260549, C. annuum cv. Bukang, C. frutescens cv. 3CA129, C. chinense cv. habanero and C. chacoense cv. 3CA87), PCR amplification was performed with 40–110 ng of genomic DNA as the template. The reaction contained Taq polymerase buffer, MgCl$_2$ at a final concentration of 2.5 mM, dinucleotide triphosphate mix at 2.5 mM, primers at 10 pmol each and one unit of LA Taq DNA polymerase (TaKaRa, Shiga, Japan) in a reaction volume of 50 μl. The primer sequences were reported previously (Pr1, 5′-CATAGCGGCCGCAGACGACTTAAATACGCG AC-3′; Pr2, 5′-CATAGCGGCCGCATGGCTTAATCTTTG AGACAA-3′; Volkov et al., 2003; Komarova et al., 2004). PCR was performed with the following parameters: initial DNA denaturation at 95 °C for 2 min; followed by 10 cycles of 95 °C for 30 s, 65 °C for 1 min and 75 °C for 3.5 min; followed by 25 cycles of 95 °C for 30 s, 65 °C for 1 min and 72 °C for 3.5 min + 20 s per cycle; and a final extension at 72 °C for 10 min. Probes were amplified from genomic DNA with the following primer pairs: C. annuum tandem repeat (TR) (forward primer 5′-GGTTT TTGTCACCGTCGAGT-3′, reverse primer 5′-ATGGATG AACGTCGGAAAAA-3′), N. tabacum A1/A2 TR (forward primer 5′-AGTGGTTGCGGGCAAAAT-3′, reverse pri-

mer 5′-CGTTGCCCAAAAGCCTAT-3′) and S. lycopersicum subrepeat I (forward primer 5′-CGACGTACCATTT GTGCTT-3′, reverse primer 5′-TTACCTATGGGCAGCA CACATGGTC-3′). The PCR products were subcloned into the pGEM-T Easy Vector (Promega, Madison, WI, USA), and the sequences were determined at NICEM (Seoul National University, Korea).

Detection of IGS homologous satellite DNA

To determine the presence of genus-specific satellite DNAs in three genera of Solanaceae (Capsicum, Nicotiana and Solanum), DNA was analyzed by Southern hybridization. Genomic DNAs (20 μg) from seven species (C. baccatum cv. Pl260549, C. frutescens cv. 3CA129, N. glutinosa, N. tabacum, S. tuberosum, S. lycopersicum cv. Microtom and S. pimpinellifolium) were digested with the HaeIII restriction enzyme and transferred from agarose gels to three Nytran membranes (Amersham Pharmacia, Sunnyvale, CA, USA). Three radioactive probes were prepared from the IGS subrepeats from C. annum, N. tabacum and S. lycopersicum by nick translation. Cross-hybridization was confirmed by hybridizing each probe to a filter independently. Hybridization was performed in 40 ml of Church's buffer (Church and Gilber, 1984) at 65 °C overnight in a hybridization oven. After hybridization, membranes were washed in solutions with increasing stringency, starting with washing solution I, which contained 2 × saline-sodium citrate (SSC) and 0.1% SDS, followed with washing solution II, which contained 1 × SSC and 0.1% SDS, and finally with washing solution III, which contained 0.5 × SSC and 0.1% SDS at 60 °C. Phosphorimaging was done on a BAS-1800II (Fijifilm, Tokyo, Japan) instrument, according to the manufacturer's instructions.

Fluorescence in situ hybridization (FISH)

To determine the presence of IGS variants independent of the coding sequence of rDNA in the Capsicum genus, FISH was performed with 25S rDNA and IGS probes from C. annuum. The FISH procedure that was applied to mitotic chromosomes was the same as previously reported (Koo et al., 2004). Probes were labeled with avidin-fluorescein isothiocyanate and anti-digoxigenin-Cy3 (Roche, Basel, Switzerland). Chromosomes were counterstained with 1 μg μl^{-1} 4,6 diamidino-2-phenylindole (Sigma-Aldrich Corporation, St Louis, MO, USA). The signals were detected with a Cooled CCD Camera (CoolSNAP, Photometrics, Pleasanton, CA, USA). Images were processed with the Meta Imaging Series 4.6 software using a Leica epi-fluorescence microscope equipped with fluorescein isothiocyanate-4,6 diamidino-2-phenylindole two-way or fluorescein isothiocyanate-rhodamine-4,6 diamidino-2-phenylindole three-way filter sets (Leica, Tokyo, Japan). The final printed images were prepared with Adobe Photoshop 7.0.

Sequence analyses

The IGS sequences of S. lycopersicum (AY366528), S. tuberosome (AF464863), N. tabacum (Y08422), N. sylvestris (X76056) and N. tomentosiformis (Y08427) were obtained from GenBank (http://www.ncbi.nlm.nih.gov/blast/Blast.cgi). To identify conserved motifs, sequence alignment was performed using DNAMAN software (Lynnon Corporation, Quebec, Canada) followed by manual adjustments. The dot-matrix analyses were conducted

878

Table 1 List of IGS sequences of Solanaceae species used in this study

Genus	Species	GenBank accession	Size (bp)
Solanum	S. lycopersicum	AY366528	3363
	S. tuberosum	AF464863	3385
Nicotiana	N. tabacum	Y08422	5140
	N. sylvestris	X76056	4349
	N. tomentosiformis	Y08427	6512
Capsicum	C. annuum	HM446335	798
		HM352910	1268
		HM352915	3107
	C. baccatum	HM352911,	855
		HM352916	848
	C. frutescens	HM352914,	662
		HM352909	2188
	C. chacoense	HM352912,	814
		HM352917	2403
	C. chinense	HM352913,	818
		HM352918	2621

using Java Dot Plot alignment (http://athena.bioc.uvic.ca/) and DNAMAN software.

Results

Cloning of IGS from five *Capsicum* species
To characterize IGS homologous repeats, IGSs from five *Capsicum* species were cloned and sequenced (Table 1). The IGS from peppers was amplified using primers designed from the 3′ end of the 25S rRNA gene and the 5′ end of the 18S rRNA gene. All the five pepper species produced multiple bands, indicating that IGS of variable sizes are present in each genome. Two IGS variants (818 and 2621 bp) from *C. chinense*, three variants (1268, 1455 and 3107 bp) from *C. annuum*, two variants (855 and 1024 bp) from *C. baccatum*, three variants (662, 851 and 2188 bp) from *C. frutescens* and two variants (814 and 2587 bp) from *C. chacoense* were cloned and sequenced. Comparison between the differently sized IGS variants using dot plot analysis revealed that these variants occurred mainly as a result of sequence loss (Supplementary Figure 1). For example, comparisons between the three different sizes of IGS of *C. annuum* revealed that the length of the IGS depended on TR size and the length of the sequence downstream of the TR.

The transcription initiation site (TIS) sequence (5′-TATA(G)TAGGGG-3′), which is necessary for transcription by RNA polymerase I, was examined in IGS variants. The TIS was found in the longest IGS of *C. annuum* (3107 bp), *C. chacoense* (2587 bp) and *C. chinense* (2621 bp), but was not found in the *C. frutescens* (2188 bp) and *C. baccatum* (1024 bp) variants or in the short variant IGS. Thus, we concluded that both transcribing and non-transcribing rDNA paralogs commonly occur in the pepper genomes.

Distribution of rDNA variants in the *Capsicum* genome
The rDNA loci and IGS-derived (CaIGSD) satellite DNA positions were analyzed using FISH analysis. The rDNA positions are highly polymorphic and well known for their potential intragenomic mobility (Schubert and Wobus, 1985). The occurrence of independent satellite repeats that are homologous to the IGS subrepeats was

reported for some solanaceous species, such as tobacco (Lim *et al.*, 2004), potato (Stupar *et al.*, 2002) and tomato (Jo *et al.*, 2009). Thus, somatic metaphase chromosomes from *Capsicum* species were examined using FISH analysis. These analyses showed that *Capsicum* species contained IGS-homologous sequences not only at the rDNA loci but also at the IGS homologous loci independent of 25S rDNA (Figure 1). The number of rDNA and rDNA-independent IGS homologous signals was highly polymorphic among pepper species. In *C. frutescens*, the 25S rDNA probe yielded four FISH signals; however, the signals for the IGS probe were preferentially localized to five major sites and several minor sites on the chromosome ends. The four rDNA probe signals co-localized with the IGS probe signals, indicating that four loci contain rRNA genes and IGS as a normal 45S rDNA locus. The signals, however, did not co-localize for the 25S rDNA and IGS, which indicates the occurrence of independent repeats with homology to IGS in *Capsicum* species. Similar results were obtained for *C. baccatum* and *C. annuum*, but the number of signals varied depending on the species. The *Capsicum* species were also highly polymorphic with regard to distribution of rDNA loci and the occurrence of non-NOR signals derived from one of the subrepeated regions of IGS similar to the *Nicotiana* and *Solanum* genera.

Genus-specific satellite DNAs in Solanaceae
The presence of genus-specific satellite DNAs in the three genera (*Capsicum*, *Nicotiana* and *Solanum*) of Solanaceae was analyzed by Southern hybridization (Figure 2). Genomic DNAs of *C. baccatum*, *C. frutescens*, *N. glutinosa*, *N. tabacum*, *S. tuberosum*, *S. lycopersicum* and *S. pimpinellifolium* were digested with the *Hae*III restriction enzyme and hybridized with IGS subrepeats of *C. annum*, *N. tabacum* and *S. lycopersicum*. At the interspecies level, the signals were detected by cross-hybridization, although the band patterns were not identical. Across genera, however, cross-hybridization did not occur. Also, a probe from the IGS subrepeat from *S. lycopersicum* (former *Lycopersicum* genus) failed to cross-hybridize with the *S. tuberosum* genome, which is from the same genus. A sequence comparison showed 67% identity between the total length of the *S. lycopersicum* and *S. tuberosum* IGSs and a 58% identity between the type I subrepeats. The distinct satellite DNAs from the three genera exhibited a high sequence similarity with the IGS of rDNA of the species.

Comparison of the IGS between five *Capsicum* species
Comparison of the IGS sequences from five species of *Capsicum* demonstrated the mechanism of sequence divergence at the intra-genus level. The dot-matrix plot with the longest IGS from each *Capsicum* species revealed the organization of *Capsicum* IGS. The IGS consisted of a central repeated region and two unique regions that flank the repeats (Figure 3). The central TR regions were variable, and this variability depended on the copy number of the monomer. The monomer was repeated 39 times in *C. chinense*, 89 times in *C. annuum*, 17 times in *C. chacoense* and 85 times in *C. frutescens*. The length of the representative subrepeat monomer of pepper IGS was conserved with a length of 8 bp and a sequence of 5′-A(G)T(A)GGCACC-3′. Variants of the subrepeat contained one or more substitutions in the eight bases of
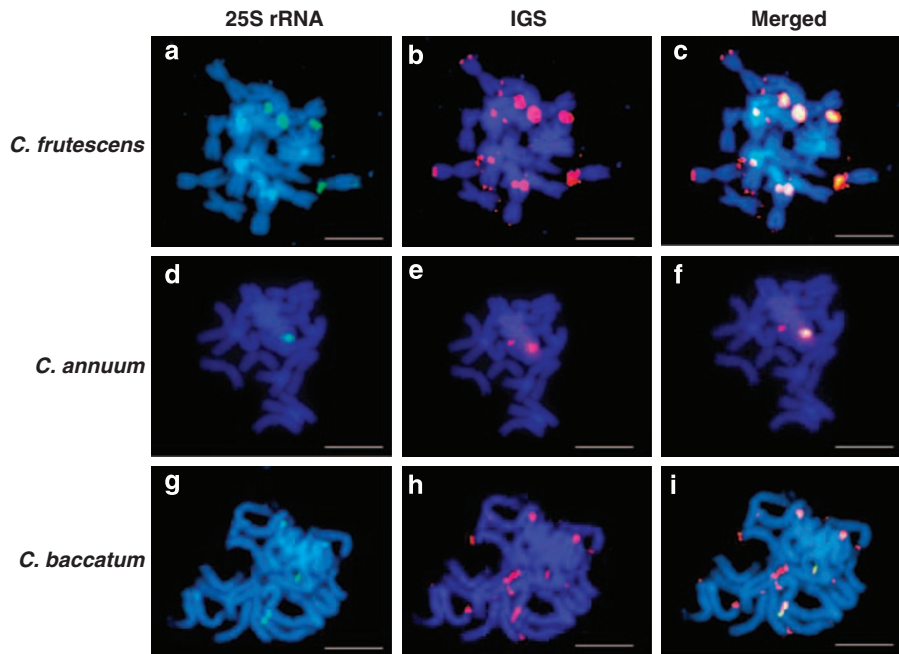
Heredity

Figure 1 Distribution of ribosomal DNA and IGS-derived satellite DNA in *Capsicum* species. The 45S ribosomal DNA loci have 25S, 5.8S and 18S ribosomal genes and an IGS between the 18S and 25S rRNA genes. Co-localized signals with the 25S rRNA probe (green) and the IGS probe (red) indicate the 45S rDNA loci. Note that red signals (IGS probe) that did not co-localize with the green signals (25S rRNA probe) are also found in all species. 4,6 diamidino-2-phenylindole was used to counterstain the DNA (blue). (**a**) 25S rDNA probe on *C. frutescens*. (**b**) IGS probe on *C. frutescens*. (**c**) Merged image of (**a**) and (**b**). (**d**) 25S rDNA probe (green) on *C. annuum*. (**e**) IGS probe (red) on *C. annuum*. (**f**) Merged image of (**d**) and (**e**). (**g**) 25S rDNA probe (green) on *C. baccatum*. (**h**) IGS probe (red) on *C. baccatum*. (**i**) Merged image of (**g**) and (**h**). Bar, 10 μm.
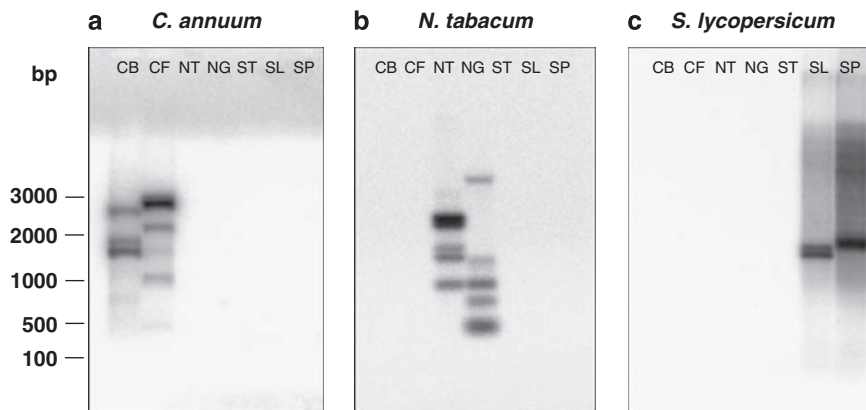

Figure 2 Genus-specific satellite DNA in Solanaceae. Genus specificity was confirmed by hybridization with IGS subrepeat probes prepared from *C. annuum* (**a**), *N. tabacum* (**b**) and *S. lycopersicum* (**c**). Cross-hybridization was limited to inter-genus species, although the band pattern was distinct even in the same genus. *S. tuberosum* (potato) failed to cross-hybridize with the IGS of *S. lycopersicum*, a former member of the lycopersicum genus. A relatively smaller amount of the satellite repeat was present in *Solanum* species as evidenced in the lack of a ladder of bands. CB, *C. baccatum*; CF, *C. frutescens*; NT, *N. tabacum*; NG, *N. glutinosa*; ST, *S. tuberosum*; SL, *S. lycopersicum*; SP, *S. pimpinellifolium*.

representative subrepeat. Six subrepeat variants were commonly found in the four *Capsicum* species; however, the total number of variants was different among the *Capsicum* species (Supplementary Table 1). For example, *C. frutescens* and *C. annuum* contained a similar monomer copy number (85 and 89), but the variant number was different (33 and 13, respectively).

An AT-rich region (67.5% AT in *C. annuum*, 71.3% AT in *C. chinense*, 68.8% AT in *C. frutescens* and 74.3% AT in *C. chacoense*) was found between the central TR region and the TIS. The length of the AT-rich region was quite

similar among species (543 bp in *C. chinense*, 515 bp in *C. annuum*, 478 bp in *C. frutescens* and 534 bp in *C. chacoense*). Sequence and dot-matrix comparisons of the IGS of the *Capsicum* species revealed two variable regions (VRs), the central TR (VR1) and the region downstream of the TIS (VR2). The length of the IGS is dependent on the degree of nucleotide loss in the VRs. VR2, which is downstream of the TIS, is in the same position as the A1/A2 TR in *Nicotiana* (Borisjuk *et al.*, 1997) and the type II subrepeat TR in *Solanum* species (Borisjuk and Hemleben, 1993). Overall, two VRs have been identified in pepper IGS, and
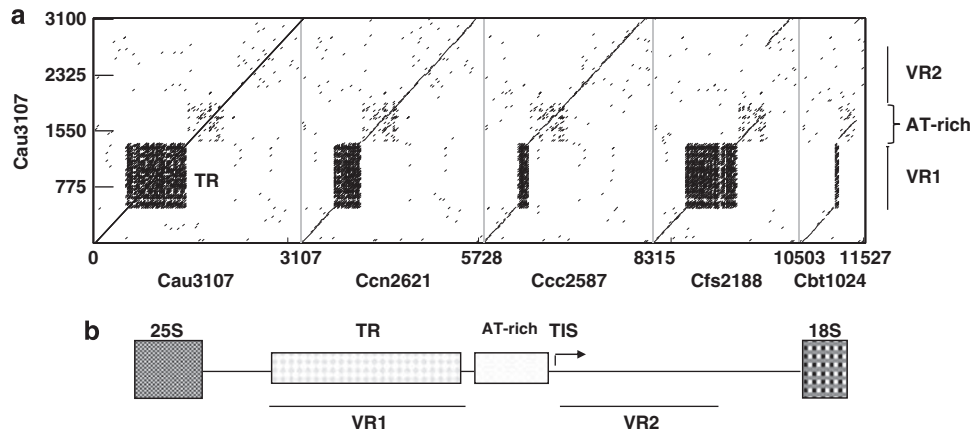
Figure 3 Comparison of IGS sequences of five *Capsicum* species. (**a**) Dot plot of IGS sequences from five *Capsicum* species. Dark box represents IGS-TR. The length of the IGS-TR monomer is conserved as 8 bp, but the copy number is highly variable across species. The IGS sequences of *C. baccatum* and *C. frutescens* may be variants because they do not contain a TIS sequence. (**b**) Schematic diagram of the *Capsicum* IGS. *Capsicum* species have an IGS-TR and two variable regions (VR). VR1 is a tandem repeat upstream of the TIS, and VR2 is the region downstream of the TIS. VR2 is the same position as the A1/A2 TR in *Nicotiana* species and the type II TR in *Solanum* species. TR was not observed in the VR2 region of *Capsicum* species. Cau, *C. annuum*; Ccc, *C. chacoense*; Ccn, *C. chinense*; Cbt, *C. baccatum*; Cfs, *C. frutescens*. The number after each name represents the length of the sequence.

the central TR (VR1) causes major differences among the IGS of the *Capsicum* species via accumulation of base substitutions and variation in copy number of the subrepeat monomer at the intra-genus level.

Comparison of IGS sequences at the inter-generic level of Solanaceae

Comparisons of the three IGS sequences from the three genera revealed divergence not only in the number of repeat blocks but also in the repeat monomer organization. The IGS sequences of tomato (Schmidt-Puchta *et al.*, 1989), potato (Borisjuk and Hemleben, 1993) and tobacco (Borisjuk *et al.*, 1997) contain two TR blocks and an AT-rich region between the two TR blocks. In other words, a TR block occurs upstream of the TIS, and the other large TR block is located downstream of the TIS. The IGS of *Capsicum* species, however, contains only a TR block upstream of the TIS.

On the other hand, the repeats were highly divergent among the genera of Solanaceae. To find conserved motifs among interspecies TRs in IGS, a multi-dot-matrix plot with *Nicotiana*, *Solanum* and *Capsicum* species was constructed (Figure 4). Comparisons of the three IGS sequences from the Solanaceae genera revealed two conserved regions, the TIS and the unique sequence upstream of the 18S rRNA coding region. Close comparisons between the TRs of the three different Solanaceae species revealed short common motifs between *Nicotiana* and *Capsicum*, as well as between *Nicotiana* and *Solanum*. Common motifs were not noted between *Capsicum* and *Solanum* species. The motifs conserved between the IGS of *Capsicum* species and the A1/A2 repeat block in the IGS of *N. tabacum* were examined (Figure 5). The common motifs were found 9–10 bp long (5′-TCGATTGTTG-3′ (10 bp) in *C. chinense*, 5′-CAACATCGG-3′ (9 bp) in *C. baccatum*, 5′-GTAGCGCCG-3′ (9 bp) in *C. chacoense* and 5′-CGACATCGC-3′ (9 bp) in *C. frutescens* and *C. annuum*); however, these motifs were in different positions of the A1/A2 TR of *N. tabacum*.

Common motifs were also found in the IGS of *Solanum* and the A1/A2 TR of *N. tabacum* (5′-CTCGTGCAAG-3′ (10 bp) in *S. tuberosum* and 5′-GGCAACTGAA-3′ (10 bp) in *S. lycopersicum*). These 9- to 10-bp motifs were a part of the A1/A2 subrepeat monomer in *N. tabacum*, but were unique and not used as a TR monomer in *Capsicum* and *Solanum* species.

The C subrepeat, the other TR in the *Nicotiana* IGS, was divided into seven subgroups according to a previous report (Volkov *et al.*, 1999) for comparison with *Nicotiana*, *Solanum* and *Capsicum* species (Table 2). The number of matching C subrepeat variants was different between genera. *Nicotiana* species have the most C subrepeat variants. Nevertheless, the composition was different between *Nicotiana* species. For instance, the C2 variant group is specific for *N. sylvestris*, whereas the C5 variant group occurs in *N. tabacum* and *N. tomentosiformis* but not in *N. sylvestris*. This result suggested that the IGS of *N. sylvestris* and *N. tomentosiformis* have a different composition of subrepeats. In addition, the IGS of allotetraploid *N. tabacum* mirrored the composition of the subrepeats of the paternal diploid *N. tomentosiformis*, although the copy numbers were different (Volkov *et al.*, 1999). The *Solanum* genus possessed six C subrepeat variants (C1b, C1d, C1h, C1i, C5a and C7). C1b was found only in *S. tuberosum*, whereas C7 was found only in *S. lycopersicum*. The *Capsicum* genus possessed two C subrepeat variants (C6b and C6c) with a difference of only 1 bp between them. Similar to the A1/A2 repeat sequences, no common motifs between the subrepeats of *Solanum* and *Capsicum* genera were observed. Interestingly, the *Solanum* species contain a longer monomer in their IGS subrepeats than the *Capsicum* species. The *Solanum* species shared more C motifs with the *Nicotiana* species than the *Capsicum* species. The copy numbers of the C subrepeat variants in each species were also different, even though each species contained the common variants. Thus, the varied composition of the monomer motif has an important role in the formation of genus-specific subrepeat monomers in the IGS of Solanaceae.
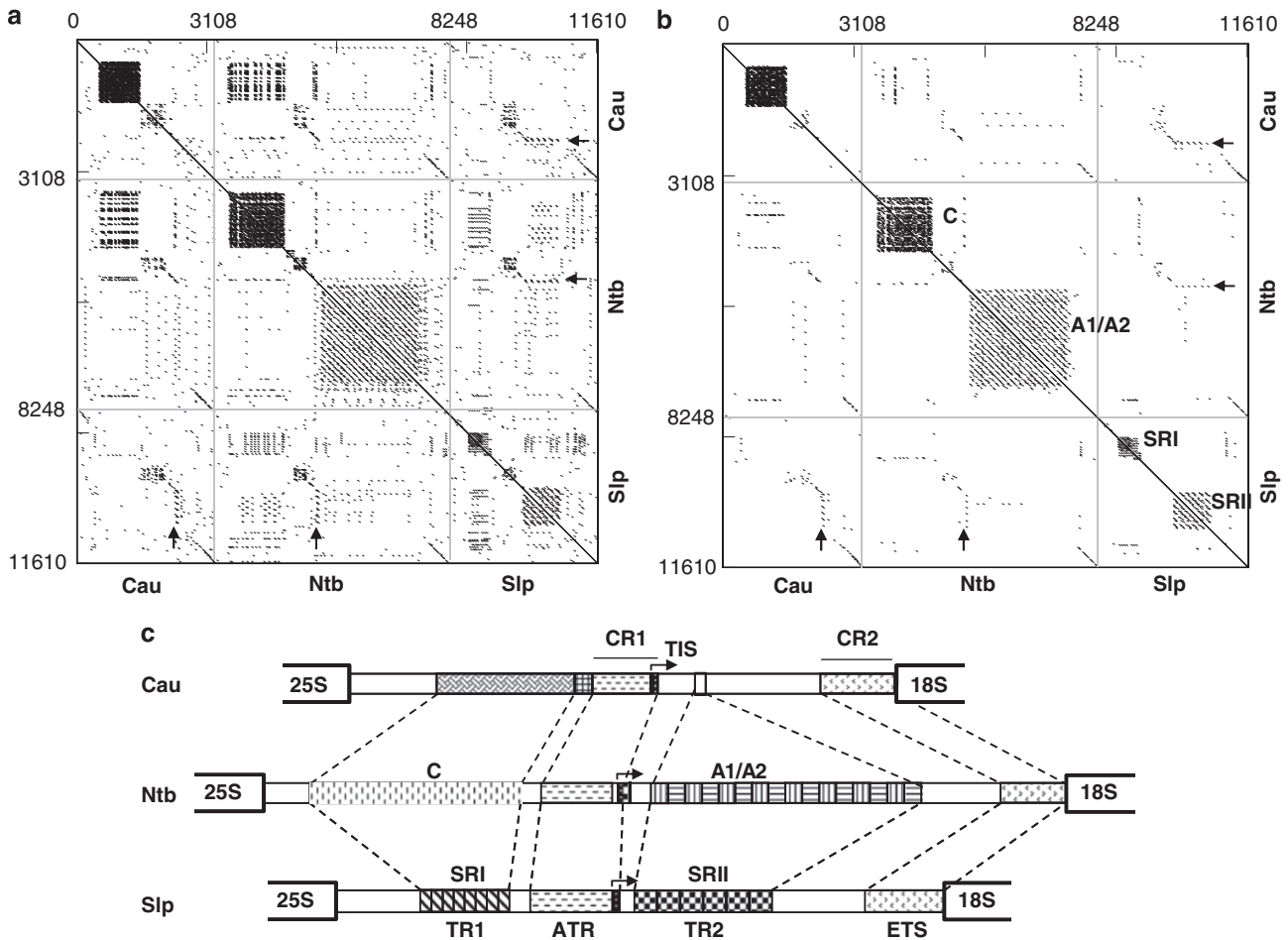
Figure 4 Inter-generic comparison of the IGS organization of *Nicotiana*, *Capsicum* and *Solanum*. (**a**) Dot plot analysis of the IGS of *C. annuum*, *N. tabacum* and *S. lycopersicum*. Two tandem repeat (TR) blocks were found in the IGS of *Nicotiana* (C and A1/A2) and *Solanum* (SRI and SRII), but only one TR block was found in the IGS of *Capsicum*. *Capsicum* species may have one copy of the monomer in TR2 region. Arrows indicate common motifs (with different copy numbers) across the genera. The six copies of the motif were present in *S. lycopersicum*, but only one copy occurred in *C. annuum* and *N. tabacum*. The motif was used in the TR monomer of *S. lycopersicum* but not in the TR of *C. annuum* and *N. tabacum*. Window size, 50. Zoom factor, 16 base per pixel. Pixel factor, 28. Scoring matrix, DNA $+5/-4$. GreyMap tool, top 10/bottom 30. (**b**) Adjusted GreyMap tool of A to top 25/bottom 70. Sequence identity between the common motifs (arrows) of *C. annuum* and *S. lycopersicum* is greater than that of *N. tabacum* and *S. lycopersicum*. (**c**) Comparison of the IGS sequences from *C. annuum*, *N. tabacum* and *S. lycopersicum*. Most of the regions were highly divergent except for the external transcribed spacers (ETSs) of the 18S rDNA and the AT-rich region (ATR), including the TIS, which may be related with the functional domain. CR indicates the conserved region. Cau, *C. annuum*; Ntb, *N. tabacum*; Slp, *S. lycopersicum*.

Rearrangement of subrepeat monomers by recombination between subrepeats in the *Nicotiana* genus

Dot-matrix comparisons revealed that reorganization of the repeat monomer occurred in the *Nicotiana* genus (Supplementary Figure 2). The subrepeats of *C. annuum* and *S. lycopersicum* were divided and aligned to the C subrepeats and A1/A2 subrepeats in the IGS of *N. tomentosiformis*. In the 676–798 bp of the subrepeat in *C. annuum*, the sequence 5′-ACCATGGC-3′, which is located at 676–682 bp, was found 55 times in the C subrepeat of *N. tomentosiformis*, and the adjacent subrepeat sequence, 5′-TCGTGGC-3′, located at 684–690 bp, was found nine times in the A1/A2 subrepeat and one time in the C subrepeat of *N. tomentosiformis*. The remaining sequence 5′-CCATGGC-3′, located at 692–698 bp, was found 25 times in the C subrepeat. These results showed that the region from 676 to 798 bp of the subrepeat of *C. annuum* was a linear strand but that the

sequence was likely shuffled into the C and A1/A2 subrepeats of *N. tomentosiformis*.

Comparing the subrepeats of *S. lycopersicum* and the subrepeats of *N. tomentosiformis* also yielded similar results. For instance, the sequence 5′-GGGCGTGGC-3′, located at 831–839 bp in the IGS of *S. lycopersicum*, was found 18 times in the C subrepeat of *N. tomentosiformis* and the sequence 5′-TGCCATC-3′, located at 841–847 bp, was found 10 times in the A1/A2 subrepeat of *N. tomentosiformis*. In contrast, a comparison of the C subrepeat or the A1/A2 subrepeat of *N. tomentosiformis* with the subrepeats of *S. lycopersicum* did not support a connected sequence of shuffling between the two subrepeats of *S. lycopersicum*. No common sequence motifs were identified between the *Capsicum* species and the *Solanum* species. These results suggest that the C subrepeat and the A1/A2 subrepeat of *Nicotiana* underwent rearrangement via shuffling of their ancestral
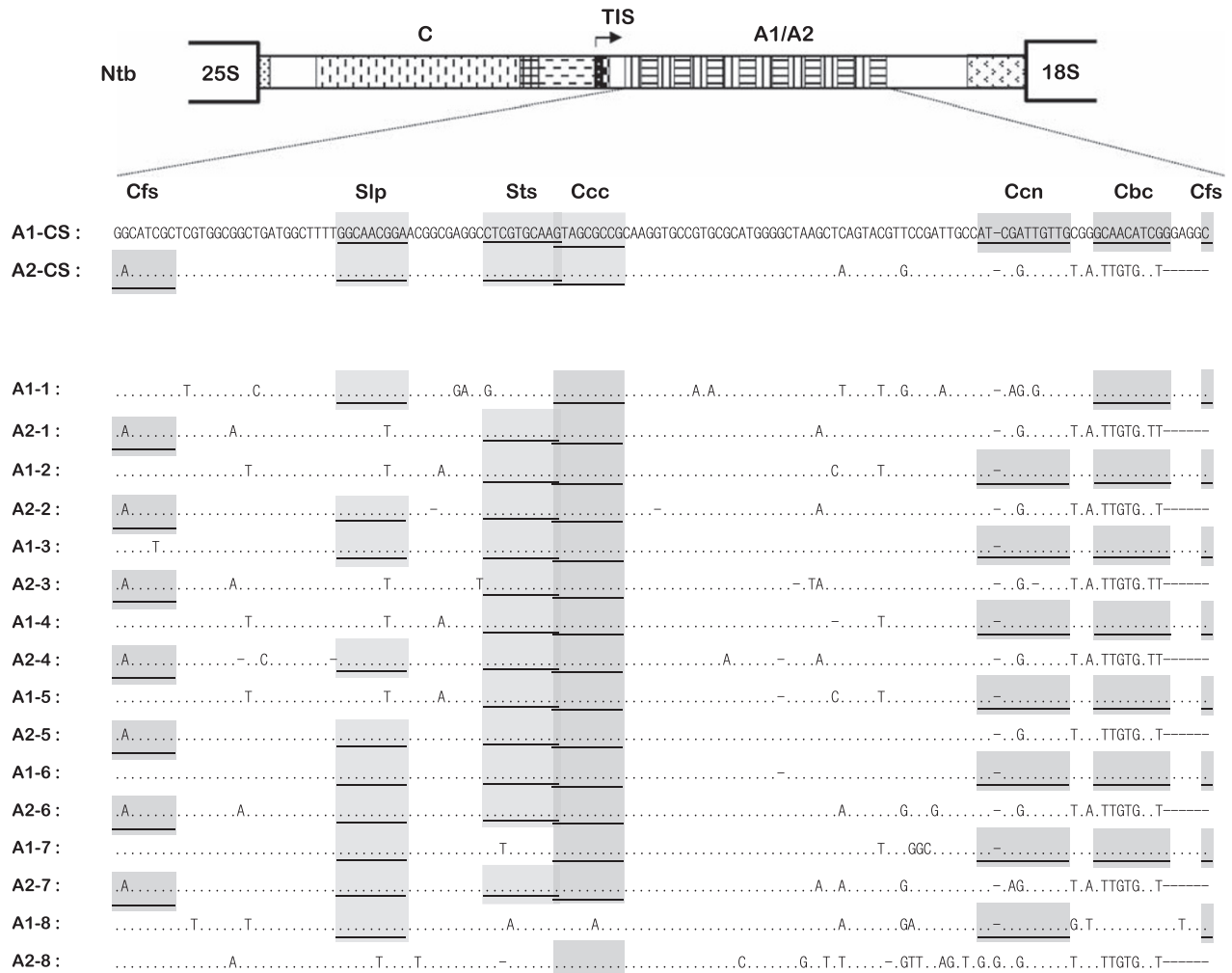
Figure 5 Differential multiplication of subrepeat motifs among Solanaceae species. Conserved sequence motifs in the A1/A2 subrepeat in the IGS of *N. tabacum* were found by comparing the IGS of the *Solanum* and *Capsicum* species. Underlined sequences (9–10-bp motifs) indicate a motif conserved between the sequence and the named species. The copy numbers of the motifs were different between the species. The motifs in the A1/A2 subrepeat of *N. tabacum* were reiterated by composing the repeat monomer. The motifs in the *Solanum* and *Capsicum* species were not contained in the monomer of the IGS-TR, and the motifs were present once in the unique region of the IGS. A dot (.) indicates an identical base, a dash (−) indicates a missing base and variations from the top row are also shown. A1-CS, A1-consensus sequence; A2-CS, A2-consensus sequence; Ccc, *C. chacoense*; Ccn, *C. chinense*; Cbc, *C. baccatum*; Cfs, *C. frutescens*; Slp, *S. lycopersicum*; Sts, *S. tuberosum*.

sequences between two TR blocks, but not in the IGS subrepeats in *Capsicum* and *Solanum* species.

Discussion

Our results suggest that the genus-specific satellite DNAs in the family Solanaceae were derived from the distinct IGS subrepeats of each species following formation of the current IGS rather than from a gradual accumulation of mutations or rearrangements from a common satellite DNA. We offer this conclusion on the basis of our finding that the satellite repeat of each genus was similar to its own IGS subrepeats but not to the satellite repeats across genera. Major differences of the satellite repeats across genera were driven by the unique monomer sequence of IGS-TR and the copy number of the monomer.

The coding regions of rDNA are highly conserved among eukaryotic organisms, whereas the sequence of the IGS region varies broadly between even closely related species. Dot plot analysis of IGS across genera showed that different regions evolved at different rates. The region neighboring the TIS and the external transcribed spacer of 18S rRNA gene are relatively well conserved across genera; however, other regions are quite variable. The IGS-TR regions are hypervariable in copy number even within a plant. The IGS-TR may become satellite repeats when IGS variant copies are dispersed to other genomic regions. The region of preferential amplification between subrepeats in the IGS is highly variable across genera. For example, the satellite repeats found in the *Solanum* genus showed sequence similarity with the upstream subrepeat of the TIS in the IGS of *Solanum* (Stupar *et al.*, 2002). On the other hand, the A1/A2 satellite repeat of the *Nicotiana* genus is similar to the downstream subrepeat of the TIS in the IGS of *Nicotiana* (Lim *et al.*, 2004). A satellite DNA family could arise in a phylogenetically short period of time via explosive amplification. Tomato and potato belong to the genus *Solanum*, and molecular dating

Table 2 Composition of subrepeat motifs in the three genera in Solanaceae

C subrepeat variant	Sequence	Number of copies							
		Ntb	Nsv	Ntm	Slp	Stb	Ccn	Cau	Ccc
C1a	GGACATGG	26	30	59	0	0	0	0	0
C1b	GGACGTGG	11	4	32	0	4	0	0	0
C1c	GGACGTGT	1	24	2	0	0	0	0	0
C1d	GGACGTCG	2	7	4	1	1	0	0	0
C1e	GGATGTGT	0	3	0	0	0	0	0	0
C1f	GGACGTGC	0	2	1	0	0	0	0	0
C1g	GGACATGA	7	0	17	0	0	0	0	0
C1h	GGGCGTGG	8	1	19	9	8	0	0	0
C1i	GGGCATGG	3	1	4	1	8	0	0	0
C1j	GGATATGG	1	0	4	0	0	0	0	0
C2a	CCATGGCACGTGT	0	10	0	0	0	0	0	0
C2b	CCATGGCACGAGT	0	10	0	0	0	0	0	0
C2c	GATGGCACGAGT	0	3	0	0	0	0	0	0
C3	GCCATGGCAGGCGT	0	0	0	0	0	0	0	0
C4	CGCACCACGT	1	1	1	0	0	0	0	0
C5a	TCGGACGTG	9	0	28	1	4	0	0	0
C5b	TCGGACGTCT	3	0	5	0	0	0	0	0
C5c	TCGGACGTCG	2	0	4	0	0	0	0	0
C5d	TCGTACGTGG	2	0	3	0	0	0	0	0
C5e	TCGGAGATGG	0	0	0	0	0	0	0	0
C6a	CCATGGCTC	7	0	16	0	0	0	0	0
C6b	CCATGGCA	13	22	37	0	0	11	29	6
C6c	CCATGGCAC	12	22	31	0	0	11	27	5
C6d	CCATGGCCC	0	0	2	0	0	0	0	0
C6h	CCACGGCAG	3	0	10	0	0	0	0	0
C7	CAGGACG	13	36	28	5	0	0	0	0

Abbreviations: Cau, *C. annuum*; Ccc, *C. chacoense*; Ccn, *C. chinense*; Slp, *S. lycopersicum*; Stb, *S. tuberosum*; Nsy, *N. sylvestris*; Ntb, *N. tabacum*; Ntm, *N. tomentosiformis*.

suggests that these species split 7.3 million years ago (MYA) (Wu and Tanksley, 2010). Alignment of the satellite repeat and IGS suggest that the satellite repeats of tomato and potato arose after speciation (Stupar *et al.*, 2002). Then, after formation, the monomer sequence of the repeats likely followed a gradual mode of sequence evolution during a long evolutionary period (Bachmann and Sperlich, 1993). The IGS homologous satellite repeat, however, may have followed rapid amplification of the repeat or transposition into other genomic regions (Jo *et al.*, 2009). The satellite repeats show a distinct distribution pattern in closely related species.

Our study indicated that several mechanisms, including base substitution, genus-specific motif composition, species-specific motif amplification and motif sequence shuffling, were involved in the IGS-TR sequence divergence. Within an individual genome of the *Capsicum* species, the IGS variants mainly resulted from deletion and base substitutions. The degree of sequence similarity in a genome, except for the deletions, was higher than in intra-species genomes, implying that the variants in a genome were generated and spread after the formation of the existing IGSs.

The copy numbers of monomers and base substitutions were major factors in IGS-TR divergence among the five *Capsicum* species. Different copy numbers of monomers among closely related species were consistent with previous results for crucifers (Delseny *et al.*, 1990), rice (Sano and Sano, 1990), Cucurbitaceae (Torres *et al.*, 1989; King *et al.*, 1993) and barley (Saghai-Maroof *et al.*, 1984). Although the length of the monomer was conserved as an 8-mer in the *Capsicum* species, the level of accumulation of base substitutions was different between species.

For instance, the IGS-TR of *C. frutescens* contained more monomers and base substitutions than the IGS sequences of other pepper species. Thus, this result indicates that copy number and base substitution is highly dependent on the species.

Comparison of IGS sequences among three intergeneric level demonstrated that different selective pressures on monomer composition might be involved in genus-specific motif formation. The *Capsicum* species have the simplest monomer in the IGS-TR. Although the rate of nucleotide substitution is variable depending on the species, the *Capsicum* species has an 8-bp single monomer. The IGS of the *Capsicum* species contains one type of common variant (C6) of the C subrepeat of *Nicotiana*. The IGS-TR of *S. lycopersicum* contains longer monomers than that of *Capsicum*. That is, a 52-bp monomer block is located upstream of the TIS, and a 142-bp monomer block is located downstream of the TIS (Borisjuk and Hemleben, 1993). The IGS sequence of the *Solanum* species shares three types of variants, C1, C5 and C7, in type I subrepeats, the 52-bp monomer block. The genus *Nicotiana* contains multifarious motifs in its C subrepeat compared with the IGS-TRs of *Solanum* and *Capsicum* species. Although the major part of the C subrepeat is composed of the 10-bp C1 variant with the consensus sequence 5′-CAGGAC(A/G)TG(G/A)-3′, seven types of C1 variants that also show length variation were found (Volkov *et al.*, 1999). Thus, we concluded that the genus *Nicotiana* may have more of the IGS-TR progenitor motifs than the genus *Solanum* or the genus *Capsicum*, as the *Nicotiana* genus contains both types of motifs found in the *Solanum* and *Capsicum* subrepeats.

The genus-specific motif formation was consistent with comparative analysis of inter-specific hybrids from several plant families. Species-specific IGS organization has been described in detail for polyploid and diploid species of the genus *Nicotiana* to explain molecular evolution during speciation (Volkov *et al.*, 1999). Comparison of *N. tabacum* and the progenitor diploid species revealed elimination/replacement of maternal sylvestris-originated rDNA and rearrangement of paternal tomentosiformis-originated rDNA in the allopolyploid genome of *N. tabacum*. In synthetic *Brassica* allotetraploids, significant changes in the paternally derived genome were also observed, whereas the maternally donated genome demonstrated no modification (Song *et al.*, 1995). Results of comparing the C motifs of *Nicotiana* with the IGS of *Solanum* and *Capsicum* revealed that each genus contained different C motifs. The IGS of the *Nicotiana* genus has motifs in common with the IGS of *Solanum* and *Capsicum*, but no conserved motifs between the IGS of *Solanum* and *Capsicum* were found.

The specific motifs in the IGS sequences of the common ancestor were used as the IGS-TR depends on the genus. Comparison of the A1/A2 subrepeats of *N. tabacum* with the IGS of *Capsicum* or *Solanum* species indicated that the motifs that shared sequences with *Capsicum* or *Solanum* were used as a component of the monomer of the A1/A2 subrepeats and that these motifs polymerized in the *Nicotiana* species. However, the motifs were not used as a part of the repeat monomer in the IGS-TR of *Capsicum* or *Solanum* species. Instead, the motifs remained as a unique sequence in the IGS of *Capsicum* or *Solanum*.

Furthermore, sequence shuffling between two subrepeat blocks in *Nicotiana* species may be one of the important mechanisms that increased IGS sequence diversity across genera. Modeling of the AT-rich region topology predicts intrinsically bent DNA with two elements of bending upstream of the TIS in the IGS of *Cruciferae* (Da Rocha and Bertrand, 1995). Thus, the two TR blocks may cross over. Interestingly, sequence shuffling between two TR blocks is only found in the *Nicotiana* genus and not in the *Solanum* or *Capsicum* genera. We found vestiges of the C subrepeat and the A1/A2 subrepeat of the *Nicotiana* genus, and thus these sequences may have experienced recombination between two subrepeats during evolution. No evidence of sequence exchanges between the two TR blocks of *C. annuum* and *S. lycopersicum* was found.

In summary, we demonstrated that the genus-specific satellite repeat in Solanaceae is spread from the highly divergent IGS of rDNA rather than the accumulation of mutations in satellite DNA. The conserved location of the IGS allowed comparing highly divergent IGS sequences. In addition to accumulation of mutations, a different number of motifs and different motif arrangements underlie the creation of distinct IGS sequences between related species.

Conflict of interest

The authors declare no conflict of interest.

References

Bachmann L, Sperlich D (1993). Gradual evolution of a specific satellite DNA family in Drosophila ambigua, D. tristis and D. obscura. *Mol Biol Evol* **10**: 647–659.

Borisjuk N, Hemleben V (1993). Nucleotide sequence of the potato rDNA intergenic spacer. *Plant Mol Biol* **21**: 381–384.

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

Church GM, Gilbert W (1984). Genomic sequencing. *Proc Natl Acad Sci USA* **81**: 1991–1995.

Clarkson JJ, Lim KY, Kovarik A, Chase MW, Knapp S, Leitch AR (2005). Long-term genome diploidization in allopolyploid Nicotiana section Repandae (Solanaceae). *New Phytol* **168**: 241–252.

Da Rocha PS, Bertrand H (1995). Structure and comparative analysis of the rDNA intergenic aspacer of Brassica rapa. Implications for the function and evolution of the Cruciferae spacer. *Eur J Biochem* **229**: 550–557.

Delseny M, McGrath JM, This P, Chevre AM, Quiros CF (1990). Ribosomal RNA genes in diploid and amphidiploid Brassica and related species-organisation, polymorphism and evolution. *Genome* **33**: 733–744.

Flaquet J, Creusot F, Drone M (1997). Molecular analysis of Phaseolus vulgaris rDNA unit and characterization of a satellite DNA homologous to IGS subrepeats. *Plant Physiol and Biochem* **35**: 611–622.

Hemleben V, Kovarik A, Torres-Ruiz RA, Volkov RA, Beridze T (2007). Plant highly repeated satellite DNA: molecular evolution, distribution and use for identification of hybrids. *Systematics and Biodiversity* **5**: 277–289.

Hourcade D, Dressler D, Wolfson J (1973). The amplification of ribosomal RNA genes involves a rolling circle intermediate. *Proc Natl Acad Sci USA* **70**: 2926–2930.

Jo SH, Koo DH, Kim JF, Hur CG, Lee S, Yang TJ et al. (2009). Evolution of ribosomal DNA-derived satellite repeat in tomato genome. *BMC Plant Biol* **9**: 42.

King K, Torres RA, Zentgraf U, Hemleben V (1993). Molecular evolution of the intergenic spacer in the nuclear ribosomal RNA genes of cucurbitaceae. *J Mol Evol* **36**: 144–152.

Kishii M, Tsujimoto H (2002). Genus-specific localization of the TaiI family of tandem-repetitive sequences in either the centromeric or subtelomeric regions in Triticeae species (Poaceae) and its evolution in wheat. *Genome* **45**: 946–955.

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.

Koo DH, Plaha P, Lim YP, Hur Y, Bang JW (2004). A high-resolution karyotype of Brassica rapa ssp. pekinensis revealed by pachytene analysis and multicolor fluorescence in situ hybridization. *Theor Appl Genet* **109**: 1346–1352.

Lim KY, Skalicka K, Koukalova B, Volkov RA, Matyasek R, Hemleben V et al. (2004). Dynamic changes in the distribution of a satellite homologous to intergenic 26–18S rDNA spacer in the evolution of Nicotiana. *Genetics* **166**: 1935–1946.

Ma J, Jackson SA (2006). Retrotransposon accumulation and satellite amplification mediated by segmental duplication facilitate centromere expansion in rice. *Genome Res* **16**: 251–259.

Mestrovic N, Castagnone-Sereno P, Plohl M (2006). Interplay of selective pressure and stochastic events directs evolution of the MEL172 satellite DNA library in root-knot nematodes. *Mol Biol Evol* **23**: 2316–2325.

Mravinac B, Plohl M, Ugarkovic D (2005a). Preservation and high sequence conservation of satellite DNAs suggest functional constraints. *J Mol Evol* **61**: 542–550.

Mravinac B, Ugarkovic E, Franjevic D, Plohl M (2005b). Long inversely oriented subunits form a complex monomer of Tribolium brevicornis satellite DNA. *J Mol Evol* **60**: 513–525.

Olmstead RG, Sweere JA, Spangler RE, Bohs L, Palmer JD (1999). Phylogeny and provisional classification of the Solanaceae based on chloroplast DNA. In: Nee M, Symon DE, Lester RN, Jessop JP (eds). *Solanaceae IV*. Royal Botanic Gardens, Kew, UK. pp 111–137.

Plohl M, Luchetti A, Mestrovic N, Mantovani B (2008). Satellite DNAs between selfishness and functionality: structure, genomics and evolution of tandem repeats in centromeric (hetero)chromatin. *Gene* **409**: 72–82.

Pons J, Bruvo B, Petitpierre E, Plohl M, Ugarkovic D, Juan C (2004). Complex structural features of satellite DNA sequences in the genus Pimelia (Coleoptera: Tenebrionidae): random differential amplification from a common 'satellite DNA library'. *Heredity* **92**: 418–427.

Saghai-Maroof MA, Soliman KM, Jorgensen RA, Allard RW (1984). Ribosomal DNA spacer-length polymorphisms in barley: mendelian inheritance, chromosomal location, and population dynamics. *Proc Natl Acad Sci USA* **81**: 8014–8018.

Salser W, Bowen S, Browne D, el-Adli F, Fedoroff N, Fry K *et al.* (1976). Investigation of the organization of mammalian chromosomes at the DNA sequence level. *Fed Proc* **35**: 23–35.

Sano Y, Sano R (1990). Variation of the intergenic spacer region of ribosomal DNA in cultivated and wild rice species. *Genome* **33**: 209–218.

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

Schubert I, Wobus U (1985). *In situ* hybridization confirms jumping nucleolus organizing regions in *Allium. Chromosoma* **92**: 143–148.

Smith GP (1976). Evolution of repeated DNA sequences by unequal crossover. *Science* **191** (4227): 528–535.

Song K, Lu P, Tang K, Osborn TC (1995). Rapid genome change in synthetic polyploids of Brassica and its implications for polyploid evolution. *Proc Natl Acad Sci USA* **92**: 7719–7723.

Stupar RM, Song J, Tek AL, Cheng Z, Dong F, Jiang J (2002). Highly condensed potato pericentromeric heterochromatin contains rDNA-related tandem repeats. *Genetics* **162**: 1435–1444.

Tautz D, Renz M (1984). Simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acids Res* **12**: 4127–4138.

Torres RA, Zentgraf U, Hemleben V (1989). Species and genus specificity of the intergenic spacer (IGS) in the ribosomal RNA genes of Cucurbitaceae. *Z Naturforsch [C]* **44**: 1029–1034.

Ugarkovic D, Plohl M (2002). Variation in satellite DNA profiles–causes and effects. *EMBO J* **21**: 5955–5959.

Volkov RA, Borisjuk NV, Panchuk II, Schweizer D, Hemleben V (1999). 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). *Mol Phylogenet Evol* **29**: 187–202.

Wendel JF, Schnabel A, Seelanan T (1995). Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (Gossypium). *Proc Natl Acad Sci USA* **92**: 280–284.

Wu F, Tanksley S (2010). Chromosomal evolution in the plant family Solanaceae. *BMC Genomics* **11**: 182.

Supplementary Information accompanies the paper on Heredity website (http://www.nature.com/hdy)