

Evolution and structure of 5S rDNA loci in allotetraploid *Nicotiana tabacum* and its putative parental species

J Fulneček¹, KY Lim², AR Leitch², A Kovařík¹ and R Matyášek¹

¹Institute of Biophysics, Academy of Sciences of the Czech Republic, CZ-61265 Brno, Czech Republic; ²School of Biological Sciences, Queen Mary, University of London, London E1 4NS, UK

Nicotiana tabacum (tobacco) is an allotetraploid derived from ancestors of the modern diploids, *N. sylvestris* and *N. tomentosiformis*. We identified and characterized two distinct families of 5S ribosomal DNA (rDNA) in *N. tabacum*; one family had an average 431 bp unit length and the other a 646 bp unit length. In the diploid species, *N. sylvestris* and *N. tomentosiformis*, the 5S rDNA unit lengths are 431 bp and 644 bp respectively. The non-coding spacer sequence of the short unit in tobacco had high sequence homology to the spacer of *N. sylvestris* 5S rDNA, while the longer spacer of tobacco had high homology with the 5S spacer of *N. tomentosiformis*. This suggests that the two 5S families in tobacco have their origin in the diploid ancestors. The longer spacer sequence had a GC rich sub-region (called the T-genome sub-region) that was absent in the short spacer. Pulsed field gel analysis and fluorescent *in situ* hybridization to tobacco metaphase chromosomes showed that the two families of

5S rDNA units are spatially separate at two chromosomal loci, on chromosomes S8 (short family) and T8 (long family). The repeat copy number at each chromosomal locus showed heterogeneity between different tobacco cultivars, with a tendency for a decrease in the copy number of one family to be compensated by an increase in the copy number of the second family. Sequence analysis reveals there is as much diversity in 5S family units within the diploid species as there is within the T and S-genome 5S family units respectively, suggesting 5S diversification within each family had occurred before tobacco speciation. There is no evidence of interlocus homogenization of the two 5S families in tobacco. This is therefore substantially different to 18–26S rDNA where interlocus gene conversion has substantially influenced most sequences of S and T genome origin; possible reasons are discussed.

Heredity (2002) 88, 19–25. DOI: 10.1038/sj/hdy/6800001

Keywords: allopolyploids; 5S rRNA; tobacco; gene conversion; evolution

Introduction

Nicotiana tabacum (tobacco) is a 5–6 million year old natural amphidiploid derived from ancestors of *N. sylvestris*, the maternal S genome donor, and *N. tomentosiformis* the paternal T genome donor (Goodspeed, 1954; Okamuro and Goldberg, 1985; Parokony and Kenton, 1995; Lim *et al.*, 2000b). Comparisons of the tobacco genome with the genomes of modern diploid relatives have revealed several contrasting evolutionary patterns. Evolution of telomere repeat lengths in tobacco does not show an additive pattern to those found in *N. sylvestris* and *N. tomentosiformis*, but instead a tobacco-specific pattern (Kovařík *et al.*, 1996). Numerous studies based on genomic *in situ* hybridization showed clear discrimination of S and T genome chromosomes, with several macroscopic S/T or T/S translocations involving, for example, the subtelomeric HRS60 repetitive sequence (Kenton *et al.*, 1993; Moscone *et al.*, 1996; Lim *et al.*, 2000b). Recombination between S and T genomes has also been identified in members of

the 1,3 β -glucanase gene family (Sperisen *et al.*, 1991). But perhaps the best characterized S and T genome interaction is for the 18–26S rDNA loci, where gene conversion has homogenized the *N. tomentosiformis*-type units and eliminated most units from the *N. sylvestris* parent (Kovařík *et al.*, 1996; Volkov *et al.*, 1999). This 18–26S rDNA gene conversion has not gone to completion, and it is possible that heavily methylated 18–26S rDNA units may not be as vulnerable to conversion as undermethylated units (Lim *et al.*, 2000a).

Previously, two 5S rDNA families in *N. tabacum* were identified that differed in the length of the non-coding spacer sequence (Fulneček *et al.*, 1998). *Nicotiana rustica*, another natural *Nicotiana* allotetraploid, also contains two 5S rDNA families that also differ in the length of the intergenic spacers (Venkateswarlu *et al.*, 1991). It is possible that the two 5S rDNA families in *N. tabacum* and *N. rustica* have their origins in the diploid progenitor species. Certainly in the Triticeae (Scoles *et al.*, 1988) and in the *Gossypium* allopolyploids, *G. hirsutum* and *G. mustelinum* (Cronn *et al.*, 1996), two 5S rDNA families have been found that remained closely similar to the diploid progenitor species. In this paper we extend our analysis of 5S rDNA families in tobacco to determine: (i) the organization, location and copy numbers of the two 5S families;

Correspondence: R Matyášek, Institute of Biophysics, Academy of Sciences of the Czech Republic, Královopolská 135, CZ-612 65 Brno, Czech Republic. E-mail: matyasek@ibp.cz

Received 7 September 2001; accepted 18 September 2001

(ii) whether the two families do indeed have their origin in the diploid progenitor species; and (iii) if gene conversion has influenced 5S evolution, as has occurred for 18–26S rDNA. We employed sequencing of multiple clones from each species, and a distance tree was constructed. Using *in situ* hybridization and pulsed field gel electrophoresis (PFGE), the chromosomal distribution and large-scale structure of the 5S gene families were studied, and the numbers and distributions of sites compared with tobacco's diploid relatives.

Materials and methods

Plant material

The following plant materials were used: *Nicotiana tabacum* L. cv. Vielblättriger, cv. Samsun, cv. Petite Havana SR-1, cv. 95–55, cv. Wisconsin 38, a callus of cv. Vielblättriger and suspension cultures, TBV-2 (Nagata *et al*, 1992) and VBI-0 (Opatrný and Opatrná, 1976), *N. tomentosiformis* Goodsp., *N. otophora* Griseb. and *N. sylvestris* Speng.

Cloning and sequencing

The intergenic spacer sequences of 5S rDNA from *N. tabacum*, *N. otophora*, *N. sylvestris* and *N. tomentosiformis* were amplified using primers derived from the *N. tabacum* 5S rRNA genic sequence (Fulneček *et al*, 1998): 5SLF 5'-CCT GGG AAT TCC TCG TGT T, and 5SLR 5'-TGC GTT AAA GCT TGT ATG ATC GCA T. *EcoRI* and *HindIII* restriction sites were included at the 5' ends for cloning purposes. PCR conditions were as follows: 1 µg total genomic DNA, 0.5 µM each primer, 0.2 mM each dNTP, 1× cloned *Pfu* DNA polymerase buffer, 2.5 U *Pfu* DNA polymerase (Stratagene) in a total volume of 100 µl. After denaturation at 94°C for 5 min, the polymerase was added and 20 cycles (94°C for 30 s, 50°C for 30 s, 72°C for 120 s) were performed (MJ Research thermocycler). A final extension was done at 72°C for 5 min. PCR products were cleaved by *EcoRI* and *HindIII*, separated by polyacrylamide gel electrophoresis. The fragments of expected size were isolated from the gel and ligated into the *EcoRI/HindIII*-cleaved vector pBluescriptII SK(-). The recombinant constructs were transformed into the DH5α strain of *E. coli*. Plasmid DNAs were isolated and both strands were sequenced with T7 Sequenase Version 2.0 DNA Sequencing Kit (USB) using T3 and T7 primers. Sequences were submitted to GenEMBL database under the following accession numbers: *N. tabacum* short family, AJ131161–AJ131165; *N. tabacum* long family, AJ131166–AJ131167; *N. sylvestris*, AJ131168–AJ131171; *N. otophora*, AJ131172–AJ131174; *N. tomentosiformis*, AJ131175–AJ131179.

DNA preparation

For most experiments, DNA was prepared from fresh young leaf tissues using standard procedures (Saghai-Marooif *et al*, 1984). The quantity and quality of DNA preparations were checked by absorbance at 260/280 nm. DNA of high molecular weight in agarose plugs for analysis by PFGE (Gene Navigator System, Pharmacia Biotech) was prepared using the modified protocol of Guidet *et al* (1990). Frozen leaf tissue was ground in liquid nitrogen to a fine powder using a mortar and pestle and transferred to 2.5 volumes of TEM (1 mM Tris-HCl pH 8.0; 50 mM EDTA pH 8.0; 0.4 M mannitol) at

50°C, and the mixture quickly stirred. The same volume of 1.5% agarose (gelling point 28.5°C) in TEM preheated to 50°C was added and the mixture stirred. The paste was transferred to a horizontally positioned plug mold (Bio-Rad) and chilled at 8°C for 20 min. The plugs were transferred to tubes containing a six-fold excess of TES (10 mM Tris-HCl pH 8.0; 500 mM EDTA pH 8.0; 1.0% N-lauroylsarcosine sodium salt) and incubated at 55°C for 60 min with one change of solution, and stored at 5°C. Some plugs were transferred to TES with 0.5 mg/ml proteinase K and incubated at 55°C for 3 days, with one change of solution and washed five times, for 1 h each, in ET buffer (50 mM EDTA pH 8.0; 20 mM Tris-HCl pH 8.0) at 55°C (first wash) and room temperature (2–4 washes) with gentle agitation. Phenylmethylsulfonyl fluoride (PMSF) was included at a concentration of 1 mM in the second and third wash. The plugs were stored in ET buffer at 5°C. Size fractionation of high molecular weight DNA was carried out in 1% (w/v) agarose in 0.5× TBE buffer for 24 h using PFGE with a time ramp of 50 to 120 s at 6 V/cm.

Restriction endonuclease digestion

Standard DNA preparations were digested with restriction enzymes using standard methods (Sambrook *et al*, 1988). For digestion of DNA for PFGE, agarose plugs were digested with 70–100 units of enzyme per 50 µl plug. The DNA was digested at 37°C for 16 h and then 30 units of fresh enzyme was added for another 4 h.

Distance tree analysis

Multiple alignment of all sequenced 5S rDNA spacers was done using the PileUp program (gap creation penalty 6, gap extension penalty 0, Wisconsin Package ver.10.1, Genetics Computer Group (GCG), Madison, WI, USA) and edited by hand. A set of 100 distance matrices was calculated using the maximum likelihood method (ts/tv ratio = 1.535, WebPhylip, version 1.3., <http://sdmc.krdl.org:8080/~lxzhang/phylyip>). The neighbour-joining method and 'consense' program were used to generate a consensus unrooted distance tree.

DNA probes

The following 5S rDNA probes were used for FISH and Southern hybridization experiments:

- Genic 5S rDNA probe – the transcribed part of 5S rDNA monomeric unit from *N. tabacum* (Fulneček *et al*, 1998; accession number AJ222659).
- T-genome sub-region probe – a GC-rich part of the long non-coding spacer of 5S rDNA from *N. tabacum* (Figure 1).

Southern blot hybridization

Following electrophoresis, gels were blotted onto Hybond N+ membranes and hybridized to ³²P-labelled DNA probes (>10⁸ dpm/µg DNA, Fermentas, Dekapriime kit). Southern hybridization was carried out in 0.25 M Na-phosphate buffer, pH 7.0, supplemented with 7% (w/v) sodium dodecyl sulfate (SDS) at 65°C for 16 h followed by washing with 2× SSC (1× SSC = 150 mM NaCl, 15 mM Na₃-citrate, pH 7.0), 0.1% SDS (twice 5 min), 0.2× SSC, 0.1% SDS (twice 15 min). Membranes were exposed to X-ray film for 4 to 48 h (Sambrook *et al*, 1988) or visualized using a PhosphorImager (Storm,

Molecular Dynamics) and quantified using ImageQuant software.

In situ hybridization

In situ hybridization was carried out as described in Leitch *et al* (1994) with modifications as described in Lim *et al* (2000b). Probes were labelled with digoxigenin-11 dUTP and detected with antidigoxigenin-FITC or with biotin-11-dUTP and detected with avidin-Cy3. Conditions of hybridization gave an estimated hybridization stringency of 80–85%. Chromosomes and interphase nuclei from root tip meristematic cells were counterstained with 2 µg/ml DAPI (4',6-diamidino-2-phenylindole).

Results

5S intergenic spacers in tobacco and progenitor species

Using 5S rDNA sequence data (Fulneček *et al*, 1998), a pair of primers was designed to amplify the non-coding spacer sequences. PCR amplification of *N. tabacum* genomic DNA generated two prominent products of about 580 and 370 bp suggesting two 5S rRNA gene families. In contrast there was one prominent band of about 580 bp and a minor band of about 450 bp from *N. tomentosiformis* and one band of about 370 bp in *N. sylvestris*. PCR products were eluted from the gel and cloned. The following sequences were submitted to the EMBL database (accession numbers AJ131161–AJ131179): five short and two long 5S spacer sequences from *N. tabacum*; four short 5S spacers from *N. sylvestris*; three long 5S spacers from *N. otophora*; and three long 5S spacer sequences from *N. tomentosiformis*. The range in length of short spacers was 306–319 bp and the long spacers 509–527 bp. Three further 5S rDNA spacers were cloned from the low abundance variant in *N. tomentosiformis*, they ranged in size from 456–459 bp and had a number of small deletions compared to the longer variant (Figure 1).

The average divergence between pairs of similar clones was expressed as percentage of nucleotide substitutions. The short spacer from *N. tabacum* was very similar to that of *N. sylvestris* (4% divergence). The long spacer from *N. tabacum* was very similar to the spacer of *N. tomentosiformis* and *N. otophora* (4% divergence). The sequence heterogeneity between individual clones of each family in tobacco also showed approximately 4% divergence. The sequences of the long and short spacer variants of *N. tabacum* 5S units showed 80% sequence identity. The differences are largely attributed to a number of missing regions in the spacer of the short family (Figure 1). A phylogenetic scheme was constructed (Figure 2) using all the 5S non-coding spacer sequences cloned. These resolve into two well-supported groups; one contains the spacers found in *N. sylvestris* and the short spacers of *N. tabacum* (the S-genome family), and the other contains the spacers of *N. tomentosiformis*, *N. otophora*, and the long spacers of *N. tabacum* (T-genome family). The tree does not cluster the spacers into species groups as might be expected.

The G+C content of the spacer was 38% for the T-genome family and 34% for the S-genome family. A 130-bp sequence, found only in the T-genome family clones (position 267–397 in Figure 1), had a higher 54% G+C content and contained three GGCC motifs. This sequence is called here the T-genome sub-region, and specific pri-

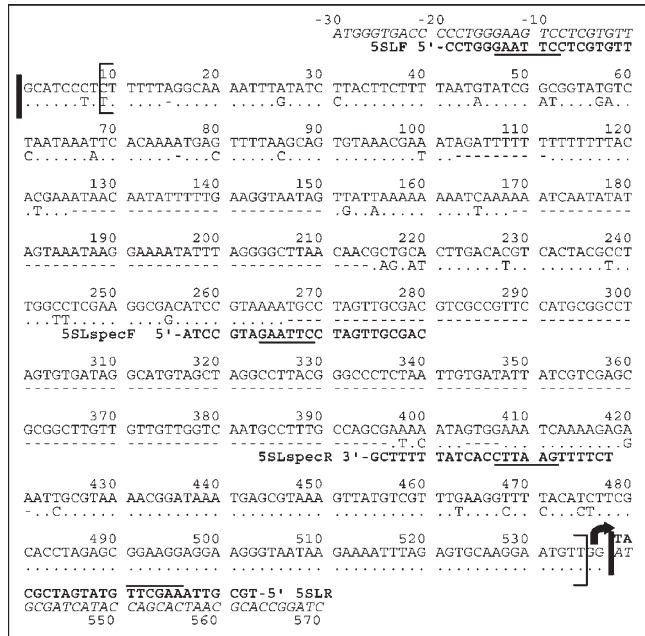


Figure 1 Sequence alignment of long (upper) and short (lower) *Nicotiana tabacum* 5S rDNA intergenic spacers (accession numbers AJ131166 and AJ131164; database names NTA131166 and NTA131164, respectively). The regions that correspond to these accession numbers are bordered by thick vertical lines. Sequences were obtained by PCR amplification of *N. tabacum* genomic DNA with 5SLF and 5SLR primers. The non-coding spacer sequences are in square brackets. The T genome sub-region is amplified with 5SLspecF and 5SLspecR occurring from 265–407 bp (accession number AJ131166). The curved arrow represents the start of transcription. Horizontal lines mark restriction enzyme sites.

mers were designed (Figure 1) for its amplification to generate a probe for further experiments (Figure 3). This probe hybridized only to the T-genome 5S family of *N. tabacum* in Southern hybridisation experiments against *Bst*NI digested genomic DNA (not shown).

The 5S spacer sequences revealed similarities with other plant spacers that also have a GC dinucleotide at –12 (with respect to the start site of transcription, Figure 1), a TATA-like motif at about –31, and a long oligo-dT tract downstream from the transcribed region. At –1 from the start of transcription, there is frequently a C-nucleotide in other species (Wolters and Erdmann, 1988; Venkateswarlu *et al*, 1991), but this is substituted by a T nucleotide in most sequences cloned from *N. sylvestris* and the S-genome family of *N. tabacum*.

Large-scale structure of 5S rDNA arrays in *N. tabacum* PFGE mapping and *in situ* hybridization were used to determine the number, size and locations of 5S rDNA arrays in the *Nicotiana* species. For PFGE, high molecular weight leaf DNA of *N. tabacum* cv. Vielblättriger was digested with several restriction endonucleases and hybridized sequentially with the probe against the T-genome sub-region (absent in spacers of S-type units) and with the 5S rDNA genic probe (Figure 3). With each of *Bcl*I, *Hpa*I and *Bgl*II DNA digestions, the T-genome sub-region probe revealed one dominant band of approximately 500 kb in length suggesting that this is the length of the T-genome 5S locus in the *N. tabacum* genome. No hybridization signal appeared after *Hae*III digestion,

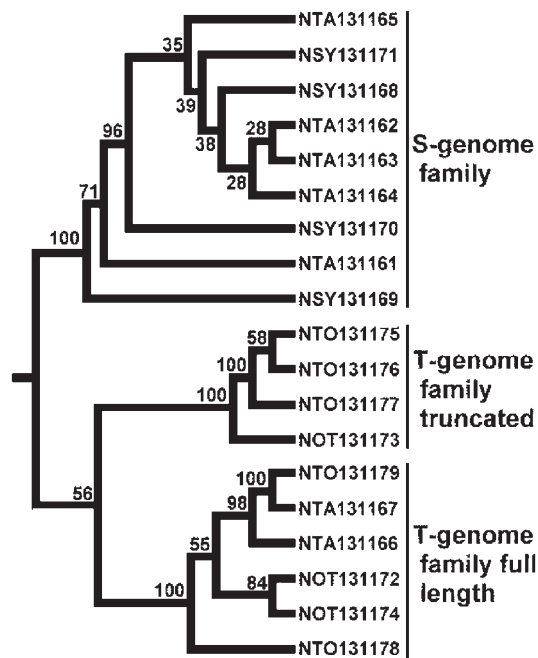


Figure 2 Consensus unrooted distance tree of 5S rDNA spacers from *Nicotiana tabacum* (NTA131161–67), *N. tomentosiformis* (NTO131175–79), *N. otophora* (NOT131172–74), and *N. sylvestris* (NSY131168–71). The names of spacers are as published in GenBank. The tree was generated from 100 bootstrapped multiple analyses (see methods). The numbers at the forks indicate the number of times the branch point occurred among the trees (out of 100). The tree is divided into two main groups corresponding to the S-genome and T-genome 5S rDNA families. The T-genome family is divided into two subgroups that cluster sequences according to spacer length.

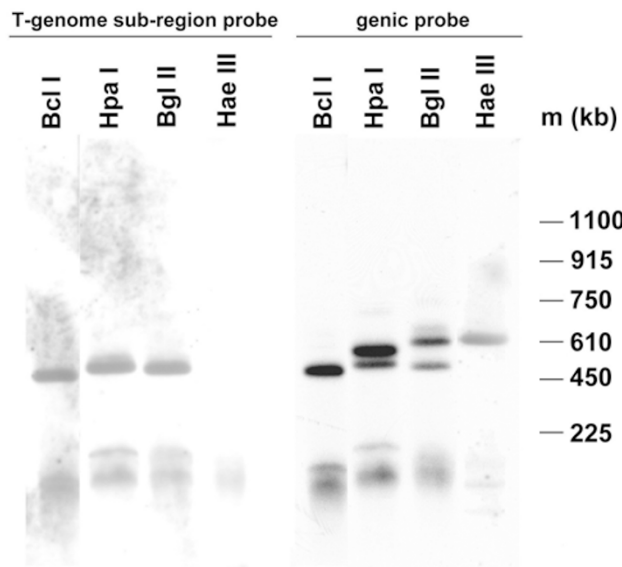


Figure 3 PFGE determination of the large scale organization of 5S rDNA clusters in *Nicotiana tabacum* cv. Vielblättriger. The restriction endonucleases used are indicated. The Southern blots were probed first with the T-genome sub-region probe, and then with the genic 5S rDNA probe. m, *Saccharomyces cerevisiae* size markers.

probably because the T-genome 5S units contain frequent *Hae*III restriction sites and were digested to small fragments that migrated off the gel; these sites are absent in the non-coding spacer of the S-genome family (Figure 1). After rehybridization of the same blot with the genic probe, a 600 kb band appeared in *Hpa*I, *Bgl*III and *Hae*III digests. Since this band is absent with the T-genome sub-region probe, the band must represent blocks of S-genome derived units. The 600 kb band was not detected in the lane loaded with *Bcl*II-digested DNA probably because some, but not all, of the S-genome 5S units contained a *Bcl*II target site.

Previously, Ali *et al* (2000) detected two rDNA loci in *N. tabacum* and Lim *et al* (2000b) suggested that they occur on chromosomes S8 and T8. It was reasoned that using the T-genome sub-region probe it would be possible to localize each family to a particular chromosome and to determine the extent (if any) of mutual mixing. In *N. tabacum* two signals were obtained with the T-genome sub-region probe (yellow/green fluorescence) and four with the genic probe (red fluorescence) (Figure 4c–e). These results demonstrate that the S- and T-genome 5S units occur in separate clusters at different chromosomal loci. In *N. tomentosiformis* and *N. sylvestris*, the genic probe labelled one locus only on chromosome 8 for each species respectively (Figure 4a,b). In *N. tabacum* it is already known that the two 5S loci occur on S- and T-genome chromosomes (Lim *et al*, 2000b), and it is reasonable to assume that the short family of *N. sylvestris* origin occur on chromosome S8 and the long 5S family on chromosome T8 of *N. tomentosiformis* origin. An examination of several hundred root-tip meristematic interphase nuclei revealed no clustering of 5S rDNA loci. Both S- and T- genome 5S loci showed signs of DNA decondensation, suggesting activity of units at both loci (Figure 4f).

5S rDNA polymorphisms in tobacco cultivars

Three widely used tobacco cultivars (Vielblättriger, Petite Havana SR-1, and Samsun), two undifferentiated tobacco cell lines (TBY-2 and VBI-0), and the progenitor species of tobacco were analyzed for polymorphism in the lengths of the basic 5S units. Southern hybridization of *Bst*NI restricted DNAs showed the presence of both T- and S-genome 5S families in all *N. tabacum* DNAs tested. However, differences were observed in the copy number of both families (Figure 5). The S-genome family was significantly reduced in both cell culture lines. This explains the strong *in situ* hybridization signal to only one locus in TBY-2 interphase nuclei (Figure 4g,h). The densitometric evaluation of radioactivity in individual bands was used to determine the copy number of monomers in individual families (Table 1).

Discussion

Two classes of 5S rDNA in the genome of *N. tabacum* *Nicotiana tabacum* is an allotetraploid with ancestors close to modern *N. sylvestris* (maternal donor of S genome) and *N. tomentosiformis* (most probable paternal donor of T genome; cf. Goodspeed, 1954; Lim *et al*, 2000b). Two distinct families of 5S rDNA have previously been localized in *N. tabacum*, a short 450 bp family and a longer 630 bp family (Fulneček *et al*, 1998). In this paper the two 5S families have been localized to separate loci. Large-scale

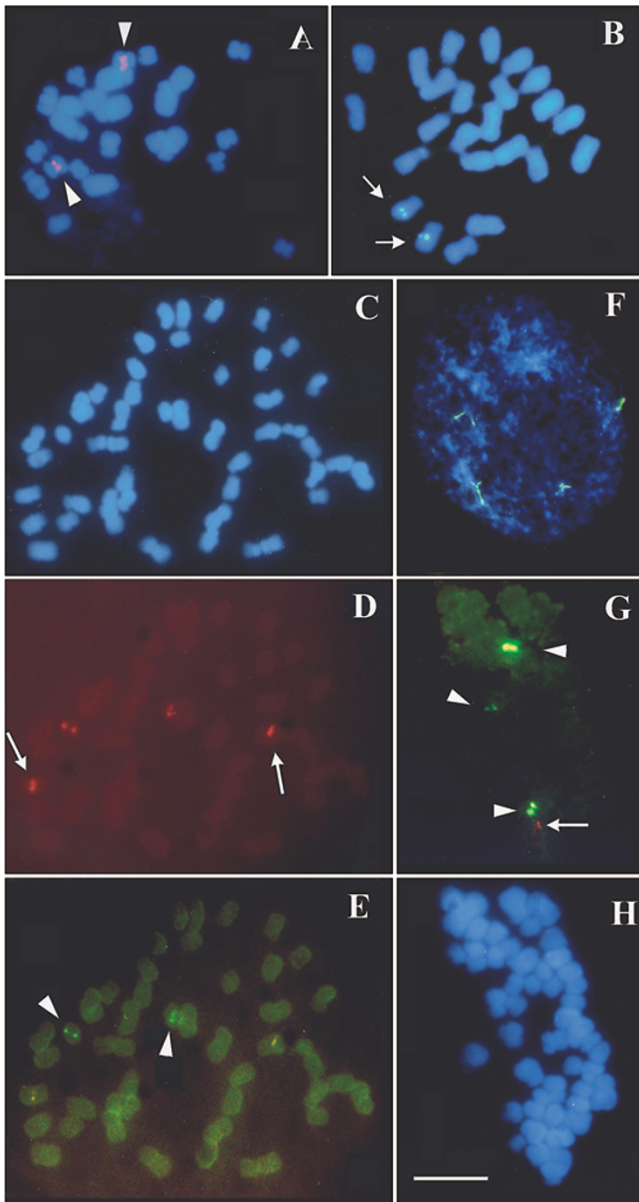


Figure 4 FISH using (a, b, f) 5S gene probe detected with either FITC (yellow-green) or cy3 (red). (c–e, g, h) 5S gene probe (cy3) double-labelled with T-genome sub-region probe (FITC). Metaphases (a–e, g, h) and interphase nucleus (f) have been counterstained blue with DAPI. Sites of probe hybridization are arrowed, homologous sites carry the same type of arrow. (a) *Nicotiana tomentosiformis*; (b) *N. sylvestris*, (c–f) *N. tabacum*; (g, h) TBY-2, showing two large and one minor site with the T-genome sub-region probe (arrow-head) and a minor site that is labelled only with the 5S genic probe (arrow). Scale bar = 10 μ m.

mapping by pulsed field gel electrophoresis showed that both families form large homogenous blocks of tandem repeats of around 500 kb (T-genome family) and 600 kb (S-genome family) most probably located to chromosome T8 and chromosome S8, respectively (cf. also Lim *et al*, 2000b). The data revealed: (i) the length of basic unit of the S-genome family of tobacco is the same as the length of basic unit in *N. sylvestris*, likewise the length of basic unit of the T-genome family of tobacco is the same as the full length unit in *N. tomentosiformis*; (ii) the T-genome family of tobacco and *N. tomentosiformis* 5S units have the

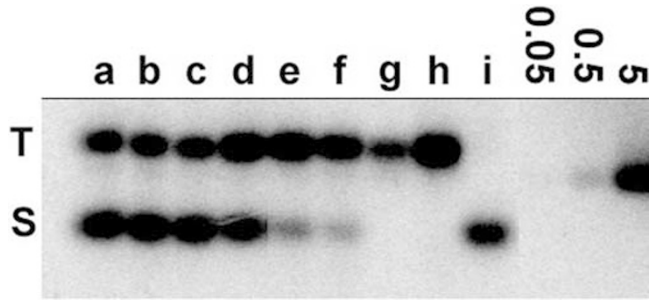


Figure 5 Determination of copy number of S- and T-genome 5S rDNA units in different cultivars of *Nicotiana tabacum*: (a) cv. Vielblättriger; (b) cv. Vielblättriger-callus; (c) cv. Samsun; (d) Petit Havana SR-1; (e) TBY-2 suspension culture; (f) VBI-0; and in the progenitor species (g) *N. tomentosiformis*; (h) *N. otophora*; and (i) *N. sylvestris*. 0.5 μ g *Bst*NI digested DNA was loaded in each lane. Lanes labelled 0.05, 0.5 and 5 contain respectively 0.05 ng, 0.5 ng and 5 ng of recombinant pBluescript II SK(-) plasmid DNA carrying the 154 bp fragment of 5S rDNA gene from *N. tabacum* used as standards for quantitative evaluation of copy numbers.

Table 1 Estimated 5S rDNA unit copy numbers per haploid genome in S- and T-genome 5S families of *Nicotiana* species. Densitometric evaluation of Southern hybridization signal (Figure 5) was used for copy number determination. Error estimate is about 10%

<i>Nicotiana</i> species	Cultivar	Copy number	
		S-family	T-family
<i>N. tabacum</i>	Vielblättriger	1100	900
	Samsun	1200	900
	Petit Havana SR-1	950	1700
	TBY-2 cell culture	400	1600
	VBI-0 cell culture	400	1600
<i>N. tomentosiformis</i>		0	400
<i>N. otophora</i>		0	1400
<i>N. sylvestris</i>		900	0

same GC-rich sub-region that is absent in the S-genome family and *N. sylvestris*; and (iii) the non-coding spacer of S-genome family has high sequence similarity to the spacer of *N. sylvestris* while the non-coding spacer of T-genome family has high sequence similarity to *N. tomentosiformis*. This organization of 5S units in tobacco in which parental 5S units can be distinguished by length polymorphisms resembles other well-studied plant allopolyploids in the genera *Triticum* (Scoles *et al*, 1988) and *Gossypium* (Cronn *et al*, 1996).

Evolution of 5S repeats at the unit level

It is informative that there is a similar diversity (4%) of 5S rDNA sequence structure: (i) among the T-genome units of tobacco; (ii) among the units of *N. tomentosiformis* and *N. otophora*; (iii) between these units and the T-family units of tobacco; (iv) among the S-genome units of tobacco; (v) among the units of *N. sylvestris*; and (vi) between these units and the S-genome units of tobacco. For this reason the phylogenetic reconstruction of 5S units does not reflect a species phylogeny (Figure 2) and it suggests high sequence diversity in the progenitors of tobacco that are maintained in modern-day diploids and tobacco. Although only a subset of repeats was sequenced, it was not possible to detect any conserved base substitutions in the non-coding spacers of the S- and T-genome subunits

compared with the diploid progenitor species. This is supported by extensive restriction analysis of numerous sites in both the non-coding (*Hae*III and *Taq*I) and genic sequence (*Bst*NI and *Rsa*I). These data suggest a slow nucleotide substitution and fixation rate of 5S units in *Nicotiana*. It is reasonable to conclude that the time required for the fixation of a nucleotide change is longer than the 5–6 million years (Okamuro and Goldberg, 1985) since the formation of tobacco. This appears true not only for the genic region but also for the relatively rapidly evolving spacer sequence.

Sequence comparisons of S- and T-genome spacers revealed several interesting features. The region of high sequence identity is interrupted at three sites with insertions/deletions found in the diploid progenitors (Figure 1). The 130-bp T-genome sub-region has a distinct nucleotide composition which contrasts with the flanking sequences in being relatively GC-rich and forming a small 'GC-rich island'. The sequence could be an ancient transposition of non-5S DNA as happened in the 25S rRNA gene of *Drosophila* (Fedoroff, 1979), although it is too short to demonstrate any characteristic features of either transposons or retroelements. Alternatively, its origin could be a duplication of pre-existing short elements in the spacer region, and indeed, three GGCCT sites are observed. Sub-repeats are also found in the non-coding spacer of 5S rDNA of other plants (Scoles *et al*, 1988; Venkateswarlu *et al*, 1991; Singh *et al*, 1994) and perhaps amplification/reductions in sub-repeat numbers play a similar role in the evolution of 5S rDNA as they do for 18–26S rDNA (Flavell, 1989; Volkov *et al*, 1999).

Evolution of 5S repeats at the chromosome level

Since the formation of *N. tabacum* there have been chromosomal interchanges (Kenton *et al*, 1993), amplification and diversification of GRD sequences (Ashby *et al*, 1997) and homogenization of 18–26S rDNA sequences via gene conversion (Volkov *et al*, 1999). Here data is presented suggesting that there is little or no change of parental 5S rDNA units in allopolyploid tobacco. Thus interlocus gene conversion has not occurred for 5S rDNA as it did for 18–26S rDNA. Similarly, interlocus homogenization at 18–26S rDNA, but not 5S rDNA, has been observed in allopolyploid *Gossypium* species (Cronn *et al*, 1996), suggesting similar evolutionary patterns.

There are several, although not necessarily exclusive, hypotheses to explain why 18–26S rDNA undergoes interlocus gene conversion while 5S rDNA seems less vulnerable to this type of genetic modification.

(1) The length of mutual sequence homology and the presence of repetitive elements could facilitate genetic recombination. In 18–26S rDNA of both *N. tomentosiformis* and *N. sylvestris*, a significant part of the intergenic spacer is formed by short 10–15 bp sub-repeats some of which are conserved (Volkov *et al*, 1999). In contrast, the sub-repeats of 5S non-coding region are less abundant, highly degenerated and specific for either S- or T- genome units. Perhaps the length of mutual sequence homology, and the abundance of sub-repeats facilitates interlocus genetic recombination at meiosis or interphase in 18–26S rDNA, but not in 5S rDNA.

(2) Cronn *et al* (1996) predicted that genetic recombination is influenced by chromosomal location. Perhaps sequences positioned near the telomere have a greater

likelihood to interact with non-homologous chromosomes (eg, during end-to-end fusions) than interstitial or centromeric sequences. In both *Gossypium* and *N. tabacum*, the 18–26S rDNA loci which show interlocus homogenization are sub-telomeric, while 5S repeats, which do not show this modification, are interstitial. Perhaps chromosomal position influences interlocus gene conversion rates, a hypothesis that could be tested in species where 5S rDNA is located sub-telomerically, eg, in pea or wheat (Ellis *et al*, 1988).

(3) The interphase nucleus of plants is a highly structured entity showing domains of condensed and decondensed chromatin (Leitch, 2000). We argued in Lim *et al* (2000a) that gene conversion between homoeologous 18–26S rDNA loci is most likely to occur during interphase of the cell cycle. It is at interphase that the different loci are decondensed, active and in close proximity within a nucleolus. Unlike 18–26S rDNA, the 5S sequences showed no such interphase associations, and were regularly well separated across the nucleus (Figure 4). This lack of interphase association may contribute to a lack of interlocus homogenization of the two 5S families in *N. tabacum* and elsewhere (Scoles *et al*, 1988; Cronn *et al*, 1996)

The copy number of 5S repeats seems to be a characteristic feature of a given cultivar and is not influenced by environmental or developmental factors (plants and derived calluses had essentially the same number of repeats; data not shown). Interestingly, a decrease in the copy number in one 5S rDNA family in tobacco is often associated with an increase in copy number in the second family suggesting that a mechanism maintaining a relatively fixed number of 5S repeats may be active. It would be interesting to analyze other stocks of *N. tomentosiformis* from different wild populations to determine if they have variable 5S rDNA copy numbers which could explain the polymorphisms in *N. tabacum*. Differences in copy number between cultivars has already been reported for 5S rDNA in wheat (Roder *et al*, 1992) and soybean (Danna *et al*, 1996). Contractions or expansions of an array seem to be frequent genetic events in the evolution of 5S repeats.

Acknowledgements

We thank the Grant Agency of the Czech Republic (grant numbers 204/98/0191, 204/01/0313, 204/99/D001 and 521/01/0037), the Natural Environment Research Council, and The Royal Society for support. The hardware and the GCG software has been acquired with the help of the grant LB98210 of the Ministry of Education of the Czech Republic Program of research and development 'Promotion of information infrastructure for research and development'. B Koukalová is thanked for valuable assistance. We thank Ms L Murad, Professor CP Lichtenstein and Dr Nichols for their help.

References

- Ali HBM, Fransz P, Shubert I (2000). Localization of 5S RNA genes on tobacco chromosomes. *Chromosome Res* 8: 85–87.
- Ashby MK, Warry A, Bejarano ER, Khashoggi A, Burrell M, Lichtenstein CP (1997). Analysis of multiple copies of geminiviral DNA in the genome of four closely related *Nicotiana*

- species suggest a unique integration event. *Plant Mol Biol* **35**: 313–321.
- Cronn RC, Zhao X, Paterson AH, Wendel JF (1996). Polymorphism and concerted evolution in a tandemly repeated gene family: 5S ribosomal DNA in diploid and allopolyploid cottons. *J Mol Evol* **42**: 685–705.
- Danna KJ, Workman R, Coryell V, Keim P (1996). 5S rRNA genes in tribe Phaseolae: array size, number, and dynamics. *Genome* **39**: 445–455.
- Ellis TH, Lee D, Thomas CM, Simpson PR, Cleary WG, Newman MA *et al* (1988). 5S rRNA genes in *Pisum*: sequence, long-range and chromosomal organization. *Mol Gen Genet* **214**: 333–342.
- Fedoroff NV (1979). On spacers. *Cell* **16**: 697–710.
- Flavell RB (1989). Variation in structure and expression of ribosomal DNA loci in wheat. *Genome* **31**: 963–968.
- Fulneček J, Matyášek R, Kovařík A, Bezděk M (1998). Mapping of 5-methylcytosine residues in *Nicotiana tabacum* 5S rRNA genes by genomic sequencing. *Mol Gen Genet* **259**: 133–141.
- Goodspeed TH (1954). *The genus Nicotiana*. *Chronica Botanica*, Waltham, Massachusetts.
- Guidet F, Rogowsky P, Langridge P (1990). A rapid method of preparing megabase plant DNA. *Nucleic Acids Res* **18**: 4955.
- Kenton A, Parokony AS, Gleba YY, Bennett MD (1993). Characterization of the *Nicotiana tabacum* L. genome by molecular cytogenetics. *Mol Gen Genet* **240**: 159–169.
- Kovařík A, Fajkus J, Koukalová B, Bezděk M (1996). Species-specific evolution of telomeric and rDNA repeats in the tobacco composite genome. *Theor Appl Genet* **92**: 1108–1111.
- Leitch AR, Schwarzscher T, Leitch IJ, Jackson D (1994). *In Situ Hybridisation: A Practical Approach*. Bios Scientific Publishers: Oxford.
- Leitch AR (2000). Higher levels of organization in the interphase nucleus of cycling and differentiated cells. *Microbiol Mol Biol Rev* **64**: 138–152.
- Lim KY, Kovařík A, Matyášek R, Bezděk M, Lichtenstein CP, Leitch AR (2000a). Gene conversion of ribosomal DNA in *Nicotiana tabacum* is associated with undermethylated, decondensed and probably active gene units. *Chromosoma* **109**: 161–172.
- Lim KY, Matyášek R, Lichtenstein CP, Leitch AR (2000b). Molecular cytogenetic analyses and phylogenetic studies in the *Nicotiana* section *Tomentosae*. *Chromosoma* **109**: 245–258.
- Moscone EA, Matzke MA, Matzke AJ (1996). The use of combined FISH/GISH in conjunction with DAPI counterstaining to identify chromosomes containing transgene inserts in amphidiploid tobacco. *Chromosoma* **105**: 231–236.
- Nagata T, Nemoto Y, Hasezawa S (1992). Tobacco BY-2 cell line as a “HeLa” cell in the cell biology of higher plants. *Int Rev Cytol* **132**: 1–30.
- Okamuro JK, Goldberg RB (1985). Tobacco single-copy DNA is highly homologous to sequences present in the genomes of its diploid progenitors. *Mol Gen Genet* **198**: 290–298.
- Opatrný Y, Opatrná J (1976). The specificity of the effect of 2,4-D and NAA of the growth, micromorphology, and occurrence of starch in long-term *Nicotiana tabacum* L. cell strains. *Biologia Plantarum* **18**: 359–365.
- Parokony AS, Kenton AY (1995). Comparative physical mapping and evolution of the *Nicotiana tabacum* L. karyotype, In: Brandham PE, Bennett MD (eds) *Kew Chromosome Conference IV*, Royal Botanic Gardens: Kew.
- Roder MS, Sorrells ME, Tanksley SD (1992). 5S ribosomal gene clusters in wheat: pulsed field gel electrophoresis reveals a high degree of polymorphism. *Mol Gen Genet* **232**: 215–220.
- Saghai-Maroo 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.
- Sambrook J, Fritsch EF, Maniatis T (1988). *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press: New York.
- Scoles CJ, Gill BSZ, Xin Y, Clarke BC, McIntyre CL, Chapman C *et al* (1988). Frequent duplication and deletion events in the 5S rRNA genes and the associated spacer regions of the *Triticeae*. *Plant Syst Evol* **160**: 105–122.
- Singh K, Bhatia S, Lakshmikumaran M (1994). Novel variants of the 5S rRNA genes in *Eruca sativa*. *Genome* **37**: 121–128.
- Sperisen C, Ruals J, Meins F (1991). Comparison of cloned genes provides evidence for intergenomic exchange of DNA in the evolution of a tobacco glucan endo-1,3-beta-glucosidase gene family. *Proc Natl Acad Sci USA* **88**: 1820–1824.
- Venkateswarlu K, Lee SW, Nazar RN (1991). Conserved upstream sequence elements in plant 5S ribosomal RNA-encoding genes. *Gene* **105**: 249–254.
- 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.
- Wolters J, Erdmann VA (1988). Cladistic analysis of ribosomal RNAs—the phylogeny of eukaryotes with respect to the endosymbiotic theory. *Biosystems* **21**: 209–214.