Theor Appl Genet (2001) 103:1273–1282 

**R.A.Volkov · C. Zanke · I.I. Panchuk · V. Hemleben**

# Molecular evolution of 5S rDNA of *Solanum* species (sect. *Petota*): application for molecular phylogeny and breeding

Received: 11 January 2001 / Accepted: 18 April 2001

**Abstract** Nucleotide sequences of 5S rRNA genes (5S rDNA) of 26 wild species of the genus *Solanum* (sect. *Petota*) originating from Middle or South America, four *Solanum tuberosum* breeding lines and one European species, *Solanum dulcamara* (sect. *Dulcamara*) were compared with each other and with the 5S rDNA of *Lycopersicon esculentum*. The length of the repeat ranges from 285 bp to 349 bp. The complete 5S repeat unit consists of the 120-bp long conserved coding region and of a intergenic spacer with a high variability in the central portion as result of deletions/duplications of short motifs demonstrating sequence similarity to box C in the 5S rRNA coding region. Numerous structural rearrangements found in the spacer region can be applied to design species-specific molecular markers for *Solanum* species involved in breeding programs. Characteristic insertions/deletions (indels) were used to reconstruct phylogenetic relationships among the species studied. *S. dulcamara* forms a separate clade; *L. esculentum* is more related to *Solanum* species of sect. *Petota*. Conservation of ancestral 5S spacer organization was demonstrated for the representatives of several series of sect. *Petota*, both *Stellata* and *Rotata*. Further rearrangements of the spacer organization occurred in at least four independent lineages: (1) *L. esculentum,* (2) ser. *Polyadenia*, (3) other *Stellata* species from Middle America (ser. *Pinnatisecta* and *Bulbocastana*), (4) superser. *Rotata*. In this last group, series *Megistacroloba* and *Conocibaccata* show a common origin, and separation from ser. *Tuberosa. Solanum chacoense* and *Solanum maglia* demonstrate a close relatedness to species of ser. *Tuberosa* and should be included into this group, whereas *Solanum bukasovii* should be excluded due to conservation of ancestral spacer organization. Three major subgroups may be distinguished for species from ser. *Tuberosa*, although a high sequence similarity was found here. Several wild species (diploids *Solanum phureja* and *Solanum spegazzinii*) probably participated in the natural origin of tetraploid *S. tuberosum*; others were later used for crossing in breeding programs (e.g. *Solanum demissum*). Clear separation of Middle-American *Stellata* species from South-American *Stellata* and from Middle-American *Rotata* polyploids is shown.

**Keywords** *Lycopersicon* · Potato · Speciation · Species-specificity · *Solanum dulcamara* · Solanaceae

Communicated by G. Wenzel

R.A. Volkov · C. Zanke · I.I. Panchuk · V. Hemleben (✉)
Department of General Genetics, ZMBP, University of Tübingen, Auf der Morgenstelle 28, 72076 Tübingen, Germany
e-mail: vera.hemleben@uni-tuebingen.de
Fax: +49-7071-295042

## Introduction

Species of the genus *Solanum* that comprises more than 1000 members (D'Arcy 1991) are distributed worldwide. Especially, *Solanum* species from sect. *Petota* are widely spread in Middle and South America. Two subsections are distinguished (Hawkes 1990): subsect. *Estolonifera* (54 non-tuber-bearing species) and *Potatoe* (180 tuber-bearing species some of which are cultivated) with superseries *Stellata* and *Rotata*. Close genetic relationship, high morphological polymorphism and interspecific hybridization of *Solanum* species of sect. *Petota* make it extremely difficult to understand the evolution of the group. In order to clarify the situation, chloroplast DNA variability was intensively studied by several authors (Spooner et al. 1993; Hosaka 1995; Spooner and Castillo 1997). Major groups (such as genus *Lycopersicon*, sect. *Etuberosum*, sect. *Petota*, etc.) can be successfully separated (Spooner et al. 1993), and even four clades in sect. *Petota* may be distinguished, but the resolving power of cpDNA restriction-site analysis was too low to describe the evolution within the terminal clades (Spooner and Castillo 1997). Investigations of RFLP and RAPD were also performed for nuclear genes of *Solanum* (Debener et al. 1990; Van den Berg et al. 1996; Miller and Spooner 1999). Very close relatedness, hybridization and introgression of *Rotata* species of the *Solanum brevicaule*

complex was confirmed, but no concordance of topology of taxa within the complex was found using RFLP and RAPD data sets. Kardolus et al. (1998) showed suitability of the AFLP technique for the biosystematics of 17 species of sect. *Petota*, but the results obtained are partially in conflict with the data of other authors and additional studies applying other sensitive markers are still necessary. Especially, questions about presumptive wild diploid progenitors of the cultivated tetraploid potato (*Solanum tuberosum*) and the origin of other polyploid species remained a subject of debate.

Previously, we searched for species-specific molecular probes to distinguish the species of sect. *Petota* selected for use in breeding programs (e.g. Menke et al. 1996). In eukaryotes, the 5S rRNA genes are tandemly arranged; each repeated unit is composed of conserved, approximately 120-bp long, coding sequences and of a more-variable intergenic spacer region. In the *Solanum* species investigated so far the length of the spacer ranges between 190 to 220 bp (Zanke et al. 1995), and the question was whether this region contains enough informational sites for a molecular phylogenetic search among closely related species. For *Solanum pinnatisectum* a 28-bp oligonucleotide of the 5S rDNA spacer can be used as specific hybridization probe to prove the hybrid nature of regenerates obtained after protoplast fusion with *S. tuberosum* (Menke et al. 1996; Hemleben et al. 1998). Therefore, we further investigated the molecular evolution of 5S rDNA repeats in order to clarify the relationships between the *Solanum* species from sect. *Petota* as well as to design species-specific markers for those species.

## Materials and methods

### Plant material

The plants used (Table 1) were cultivated in vitro on Murashige and Skoog medium with the exception of *Solanum dulcamara* where leaves were collected in the field.

### Isolation, cloning and sequencing 5S rDNA

Total DNA was isolated from leaves according to the method described (Zanke et al. 1995). Standard procedures were carried out according to Sambrook et al. (1989). For cloning of the 5S rDNA of 17 species the strategy described by Zanke et al. (1995) was applied and clones containing complete 5S rDNA repeats were obtained.

For direct sequencing primers, pr5S14 (5′-GGC GAG AGT AGT ACT AGG ATC CGT GAC-3′) and pr5S15 (5′-GCT TAA CTT CGG AGT TCT GAT GGG A-3′) deduced from the 5S rRNA coding sequence of *Lycopersicon esculentum* (Venkateswarlu et al. 1991) were used. PCR products were electrophoretically separated in agarose gels, purified with a Gel Band Purification Kit (APBiotech), and sequenced with the BigDye Terminator Cycle Sequencing Kit (PE Applied Biosystems) using pr5S14 or pr5S15. The sequences obtained appeared in the EMBL data bank under the accession nos. listed in Table 1.

Sequence analysis was performed using several softwares: “DNA-Star”, “sequence alignment program” Align Plus (Ver. 2.0), “Clone/sequence clone program” (Myers and Miller 1988), and others. Phylogenetic trees were bootstrapped (200 to 500 replicates; Felsenstein 1985).

## Results

### Length of 5S rDNA

5S repeat lengths in *Solanum* wild species and *S. tuberosum* breeding lines were determined after Southern hybridization of *Bam*HI-digested genomic DNA; a 5S rDNA repeat from *S. pinnatisectum* (Zanke et al. 1995) was used as hybridization probe. One intensive band corresponding to monomer size and a weak ladder-pattern characteristic for tandemly arranged repeats were observed for the majority of species (data not shown), indicating that mainly one repeat size class is present in the genome. Nevertheless, for seven species intragenomic polymorphisms of 5S rDNA repeat length were found (Table 1). The size of 5S rDNA repeats was also checked after PCR amplification. PCR products were digested with the restriction enzyme *Bam*HI and electrophoretically separated on agarose gels. The size of the fragments obtained was in agreement with the data from Southern hybridization (Table 1).

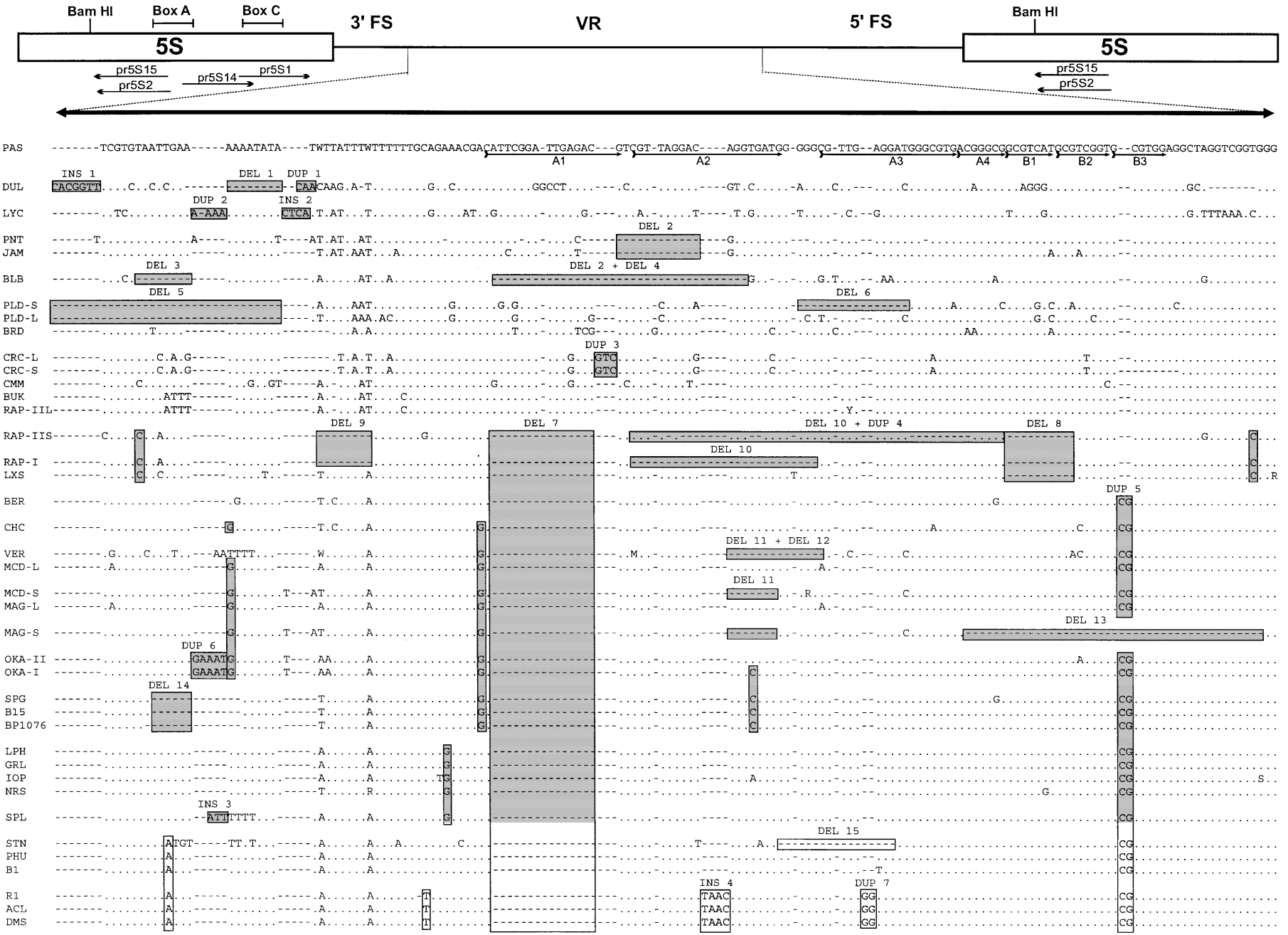
### 5S-repeat intraspecific heterogeneity

*Bam*HI-digested PCR products for 14 wild species and four *S. tuberosum* breeding lines were cloned and sequenced. For each accession, between 3 to 9 individual clones were compared showing mostly similar lengths with only some exceptions: Two distinct repeat-length variants were found in the genomes of seven species (Fig. 1). For *Solanum brevidens*, six 5S rDNA clones were characterized: two clones contained a complete repeat with a length of 343 bp each (Table 1); additionally, three clones (Y16660–62) with an insert length of 230 bp each, and one clone (Y16663) with 461 bp, were sequenced. Sequence analysis indicated that the 230 bp-long clones represent pseudogenes with a 113-bp deletion of the 3′ part of the coding and adjacent spacer regions. The 461 bp-long clone contains a pseudogene-dimer with a mutation in the *Bam*HI restriction site. These data are in agreement with the results of Southern hybridization where two overlapping ladder patterns were observed for *S. brevidens* DNA after digestion with *Bam*HI (data not shown). The stronger signal with a monomer size of approximately 340 bp corresponds to the complete 5S rRNA gene of *S. brevidens*, whereas the weaker signal with a monomer size of 230 bp was obviously caused by amplification of tandemly arranged 5S rDNA pseudogenes.

From *Solanum circaeifolium*, 2 of 8 clones sequenced show a 7-bp deletion in the spacer (short repeat variant; Table 1, Fig. 1), and one clone (AJ226017) contains a 14-bp deletion in the 5S RNA coding region. From six clones sequenced for two accessions of *Solanum okadae*, one clone (AJ226066) contains a 55 bp-long deletion at the beginning of the spacer region and probably represents a pseudogene. Two repeat classes found for *Solanum microdontum*, *Solanum maglia* and *Solanum poly-*

**Table 1** List of *Solanum* species and 5S rDNA sequences analyzed. Note: Taxonomy, species names and abbreviations are shown according to Hawkes (1990). Bio, Fa. Bioplant, Ebstorf, Germany; IPK, the Institut für Pflanzengenetik und Kulturpflanzenforschung, Gatersleben, Germany; GDC, the German-Dutch Curatorium for Plant Genetic Resources, Braunschweig, Germany; MPI, Max-Planck-Institute für Züchtungforschung, Köln; GFP, Gesellschaft zur Förderung der Pflanzenzüchtung, Bonn, Germany; CPBR, Center for Plant Breeding and Reproduction Research CPRO, Wageningen, The Netherlands; CIP, Centre Internacionale de la Palma, Lima Peru; BLBP, Bayrische Landesanstalt für Bodenkultur und Pflanzenbau, Freising, Germany; RAGIS, RAGIS-Züchtstation Heidehof, Heidehof, Germany; TUB, Herbarium Tubingense, University of Tübingen, Germany. CS, cloning and sequencing; DS, direct sequencing of PCR products. Long (L) and short (S) repeat variants are shown separately; presumptive pseudogenes are not included

| Classification | Species | Abbreviation | Accession no. | Source | Sequencing | EMBL Accession nos. | Intergenic spacer region | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | Length (bp) | GC (%) |
| Section *Petota* | | | | | | | | |
| Subsect. *Estolonifera* | | | | | | | | |
| Ser. *Etuberosa* | *S. brevidens* | BRD | BGRC N 17441 | GDC | CS | AJ226035–36 | 223 | 48.0 |
| Subsect. *Potatoe* | | | | | | | | |
| Superser. *Stellata* | | | | | | | | |
| Ser. *Bulbocastana* | *S. bulbocastanum* | BLB | BGRC N 08006 | GDC | CS | AJ226012–14 | 189–190 | 50.7 |
| Ser. *Pinnatisecta* | *S. pinnatisectum* | PNT | BGRC N 08168 | GDC | CS | X82779, AJ22608–11 | 210 | 49.1 |
| | *S. jamesii* | JAM | BGRC N 10054 | GDC | DS | AF331058 | 213 | 50.2 |
| Ser. *Polyadenia* | *S. polyadenium* | PLD | BGRC N 08176 | GDC | CS | AF331045 (L) | 206 | 51.5 |
| | | | | | | AF331044, 46 (S) | 193 | 49.2 |
| Ser. *Commersoniana* | *S. commersonii* | CMM | BGRC N 17654 | GDC | DS | AF331056 | 224 | 51.3 |
| Ser. *Circaeifolia* | *S. circaeifolium* ssp. *quimense* | CRC | BGRC N 27036 | GDC | CS | AJ226015–16, 18–20 (L) | 228–229 | 49.8 |
| | | | | | | AJ226021–22 (S) | 220, 222 | 49.6 |
| Ser. *Yungasensa* | *S. chacoense* | CHC | B2 | MPI | DS | AF331055 | 213 | 50.7 |
| Superser. *Rotata* | | | | | | | | |
| Ser. *Megistacroloba* | *S. raphanifolium*-I | RAP | BGRC N 07207 | GDC | DS | AF332131 | 172 | 50.0 |
| | *S. raphanifolium*-II | | BGRC N 08189 | | DS | AF332133 (L) | 223 | 49.8 |
| | | | | | | AF332132 (S) | 178 | 51.1 |
| Ser. *Conicibaccata* | *S. laxissimum* | LXS | GLKS 154.3 | IPK | CS | AJ226046–50 | 202 | 49.1 |
| Ser. *Maglia* | *S. maglia* | MAG | BGRC N032571 | GDC | CS | AF331051–54 (L) | 213 | 49.1 |
| | | | | | | AF331047–50 (S) | 165 | 45.6 |
| Ser. *Tuberosa* | *S. berthaultii* | BER | BGRC 18548 | GDC | CS | AJ226037–41 | 213 | 51.0 |
| – wild species | *S. bukasovii* | BUK | BGRC N 15424 | GDC | DS | AF332130 | 222 | 48.2 |
| | *S. gourlayi* | GRL | 5.6 | GFP | DS | AF331057 | 213 | 49.8 |
| | *S. leptophyes* | LPH | 8.27 | GFP | DS | AF331059 | 213 | 49.8 |
| | *S. microdontum* | MCD | BGRC 27351 | GDC | CS | AJ226051–54, 59 (L) | 212–213 | 49.3 |
| | | | | | | AJ226055–58, 60–62 (S) | 208 | 51.6 |
| | *S. neorossii* | NRS | 11.42 | GFP | DS | AF331060 | 213 | 50.2 |
| | *S. okadae*-I | OKA | BGRC 17550 | GDC | CS | AJ226060–62 | 220 | 50.3 |
| | *S. okadae*-II | | BGRC 24719 | GDC | CS | AJ226063–64, 66 | 220 | 49.1 |
| | *S. sparsipilum* | SPL | 14.9 | GFP | DS | AF331062 | 216 | 49.1 |
| | *S. spegazzinii* | SPG | 17.45 | GFP | DS | AF331063 | 205 | 52.2 |
| | *S. vernei* | VRN | – | GDC | DS | AF332129 | 202 | 49.5 |
| – cultivated species | *S. phureja* | PHU | IVP 101 | CPBR | DS | AF331061 | 212 | 50.0 |
| and breeding lines | *S. stenotomum* | STN | – | CIP | DS | AF331064 | 200 | 47.5 |
| | *S. tuberosum* | TBR | B1 | BLBP | CS | Y16656–59 | 212–213 | 49.5 |
| | dihaploid breeding lines | | R1 | RAGIS | CS | X82781 | 216 | 49.1 |
| | | | B15 | BLBP | CS | X82780, Y16650–51 | 204–205 | 52.2 |
| | | | BP1076 | Bio | CS | Y16652–55 | 205 | 52.3 |
| Ser. *Acaulia* | *S. acaule* | ACL | – | CIP | CS | AJ226031–34 | 219–220 | 49.3 |
| Ser. *Demissa* | *S. demissum* | DMS | – | CIP | CS | AJ226023–25 | 219 | 49.0 |
| | *S. iopetalum* | IOP | GLSK 161 | IPK | CS | AJ226042–45 | 211–213 | 49.0 |
| Section *Dulcamara* | *S. dulcamara* | DUL | 96065 | TUB | CS | AJ226026–30 | 219–224 | 57.7 |


Bam HI
Box A
Box C
3' FS
VR
5' FS
Bam HI
5S
5S
pr5S15
pr5S2
pr5S14
pr5S1
pr5S15
pr5S2
PAS
A1
A2
A3
A4
B1
B2
B3
DUL
LYC
PNT
JAM
BLB
PLD-S
PLD-L
BRD
CRC-L
CRC-S
CMM
BUK
RAP-IIL
RAP-IIS
RAP-I
LXS
BER
CHC
VER
MCD-L
MCD-S
MAG-L
MAG-S
OKA-II
OKA-I
SPG
B15
BP1076
LPH
GRL
IOP
NRS
SPL
STN
PHU
B1
R1
ACL
DMS
INS 1
DEL 1
DUP 1
DUP 2
INS 2
DEL 2
DEL 3
DEL 2 + DEL 4
DEL 5
DEL 6
DUP 3
DEL 9
DEL 7
DEL 10 + DUP 4
DEL 8
DEL 10
DUP 5
DEL 11 + DEL 12
DEL 11
DEL 13
DUP 6
DEL 14
INS 3
DEL 15
INS 4
DUP 7

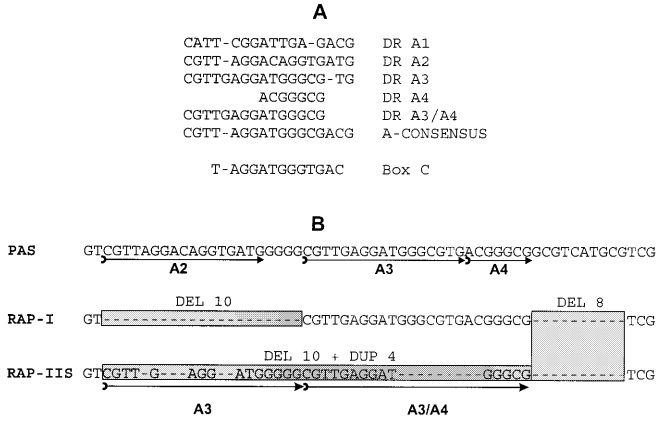



**Fig. 2A, B** Direct repeats A (*DR A*) in the 5S rDNA spacer of *Solanum*. **A** Sequence similarity between individual copies of DR A and Box C in the 5S rRNA coding region. **B** Rearrangement of DR A: DR A2 was deleted (*DEL 10*) in the spacer region of *S. raphanifolium*-I. In the short repeat variant of *S. raphanifolium*-II this deletion was compensated by reconstruction of DR A3 and A4 (*DEL 10+DUP 4*) that led to the appearance of two perfect copies of DR A. PAS presumptive ancestral sequence

*adenium* differ by deletions (7–43 bp) in the central portion of the spacer (Fig. 1). Different repeat variants are also present in *Solanum raphanifolium*-II, but this accession has an obviously hybrid origin (see Discussion).

In order to evaluate not only length but also sequence heterogeneity of the 5S rDNA repeats, comparison of all clones of the same size for each species studied was performed, and corresponding consensus sequences were produced. Sequence similarity between the consensus sequence and individual clones was 99–100% for the coding region and 95–100% for the spacer. The difference observed is mainly due to occasionally distributed base substitutions.

Taking into consideration such a high level of intragenomic sequence similarity of 5S repeats, direct sequencing of PCR products was further applied for 13 other diploid species (Table 1) that showed only one band after electrophoretic separation of PCR products. Here the complete spacer region and adjacent 3′ and 5′ ends of the coding region were amplified, sequenced, and the sequences obtained were used for alignment with the consensus sequences previously produced for other species.

## Interspecific variability

Different parts of the 5S repeated units demonstrate different rates of molecular evolution. As expected, the most-conserved sequences represent the 5S rRNA coding region; only a few base substitutions were found here. In the non-transcribed spacer, however, there are regions with different degree of conservation. Homology reaches its maximum 3′ downstream (18–19 bp) from and 5′ upstream (56–59 bp) of the coding region (3′ and 5′ **f**lanking **s**equences, FS), whereas the lowest level of sequence similarity was observed in the middle of the spacer (**v**ariable **r**egion, VR) where short imperfect **d**irect **r**epeated motives (DR A and B; Figs. 1 and 2) can be distinguished. Generally, there are three and a half copies of DR A and three copies of DR B, but some species show indels here, whereas the level of base substitutions remains low (Fig. 1). Thus, the 5S rDNA spacer region evolved preferentially through fast indel events rather than by point mutations.

◀ **Fig. 1** Organization of 5S rDNA repeats in the genus *Solanum* and sequence comparison of the variable region (*VR*) in the spacer: deletions (*DEL*), duplications (*DUP*), incertions (*INS*) and point mutations common for several species are shown; *FS*, rather conserved flanking sequences; *PAS*, presumptive ancestral sequence. Positions of PCR primer and direct repeats A and B are indicated. For species abbreviations see Table 1; *LYC=L. esculentum* (Venkateswarlu et al. 1991; Acc. No X55697)

## G+C content

For the evolutionary variable spacer region the G+C content was determined. Compared with the South American *Solanum* species, the G+C value of *Solanum dulcamara* 5S rDNA (57.7%) is significantly higher than the G+C values of potato wild species (47.5–52.2%) and breeding lines (49.1–52.3%; see Table 1). Remarkably, the upper strand of the spacer region of all *Solanum* species analyzed is characterized by a high guanidine content of approximately 35% and only 18% cytosines, although there appears an AT-rich region in the VR (see Fig. 1).

## Mode of 5S rDNA molecular evolution and phylogeny of *Solanum*

Sequence data were used for phylogeny reconstruction applying different computer algorithms (see Materials and methods), and corresponding dendrograms were produced. The topology of these dendrograms differed essentially, depending on the algorithm used. It is well known that there is no widely accepted theory for selecting “gap costs” (Pearson 1995). Taking into account the mode of molecular evolution of 5S rRNA genes (high abundance of indels) we have for calculations applied different values for gap costs and found that the dendrogram topology is extremely unstable. Especially sensitive to the changes of gap-parameters were positions of species containing long deletions such as *S. pinnatisectum*, *Solanum laxissimum*, *S. raphanifolium*, etc. Some of the dendrograms obtained were obviously in conflict with the morpholological/biogeographical data. Taking together, this analysis shows that, due to the high abundance of indels in comparison to base substitutions in the spacer, these region cannot be applied for correct estimation of genetic distances and phylogeny reconstruction using standard algorhithms. On the other hand, indels cannot be excluded from calculations because an essential part of taxonomically significant information will be lost.

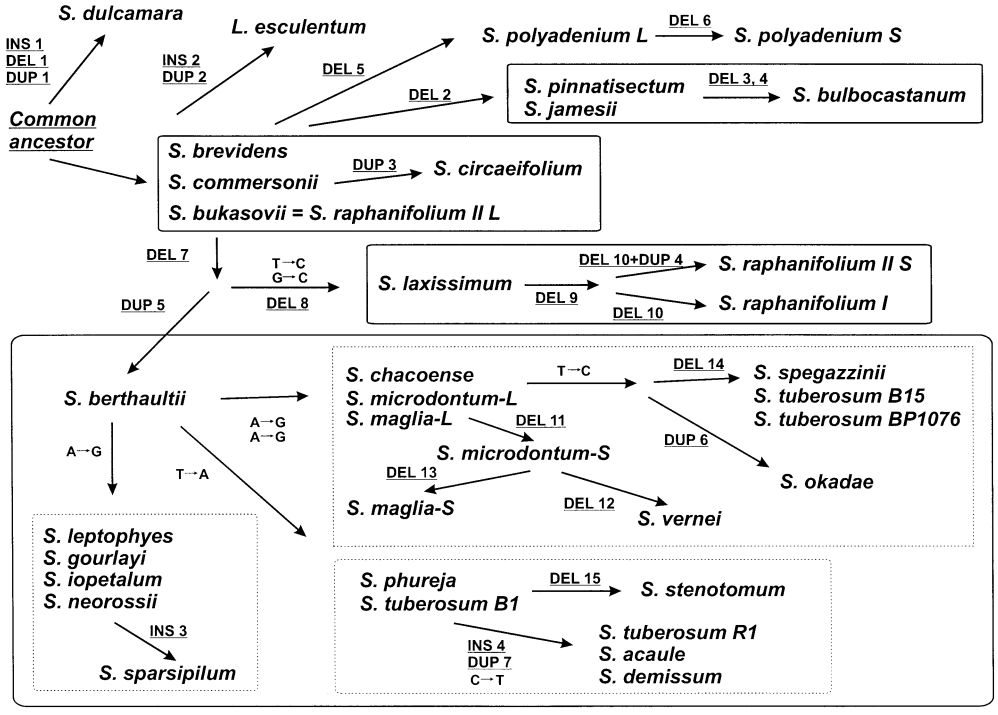



**Fig. 3** Schematic representation of molecular evolution and phylogenetic lineages deduced from the variable spacer region in 5S rDNA repeats of the genus *Solanum*/*Lycopersicon*: deletions (*DEL*), duplications (*DUP*), insertions (*INS*) and point mutations common for several species are shown

Therefore, 5S rDNA sequences can be extremely valuable for phylogenetic reconstructions of *Solanum*, because it is possible to group species and to distinguish between groups by manually evaluating the indels (see Fig. 1). So, INS 1 is only present in *S. dulcamara*; the majority of species in superser. *Rotata* bear an identical deletion (DEL 7), *Solanum acaule* and *Solanum demissum* contain an identical insertion (INS 4), etc. Comparing localizations and the sequence-specificity of indels, and also taking into consideration identical base substitutions present at least in two species, it was possible to define positions for all species studied. All results obtained by this "manual" approach are summarized as a phylogenetic scheme in Fig. 3 and further evaluated in the Discussion.

## Discussion

General organization of 5S rDNA

5S rRNA genes of the *Solanum* species analyzed exhibit a tandem organization typical for higher plants (Ellis et al. 1988; Hemleben and Werts, 1988; Scoles et al. 1988; Lapitan et al. 1991). The length of the complete repeat unit of *Solanum* ranges between 285 bp and 349 bp. Analysis of sequences available in the NCBI-database shows that *Solanum* species, especially from superser. *Rotata*, have relatively short 5S repeats in comparison to other Solanaceae and representatives of many other families. Hence, reduction of the repeat length is a general direction of molecular evolution of 5S rDNA in the genus *Solanum*. Each unit consists of the 120 bp-long highly conserved coding region which is separated by relatively variable spacers. The high conservation of the 5S genes, especially at the beginning of the coding region, demonstrates some family specificity. Thus, 5S rRNA genes within the Solanaceae (Frasch et al. 1989; Barciszewska et al. 1994) and Poaceae (Van Campenhout et al. 1998; Grebenstein et al. 2001) begin with GGA, whereas the typical beginning of the coding region is AGG in the Fabaceae and GGG in the Brassicaceae (Hemleben and Werts 1988; Barciszewska et al. 1994)

In comparison to the middle VR both termini of the spacer region (3′ and 5′ FS) are more-conservative, suggesting its presumptive functional significance. A pyrimidine-rich stretch that probably functions as a termination site is present in 3′ FS**.** At the position –28 to –24 bp in 5′ FS a TAATA-motif was found that is similar to the conserved "TATA"-box found at the same position in *Vigna radiata* and *Matthiola incana* (Hemleben and Werts 1988). The comparative conservation of FS was also reported for the Fabaceae (Crisp et al. 1999), Poaceae (Van Campenhout et al. 1998), and even for the Gymnospermae (Trontin et al. 1999). In the VR two further portions can be distinguished: (1) the AT-rich and (2) the subrepeated regions. A rather AT-rich region with similarity to the amplification-promoting sequences (Borisjuk et al. 2000) which may be involved in amplification of the 5S rDNA repeats was also observed in the 5S spacer region of the Fabaceae (Hemleben and Werts, 1988) and Poaceae (Grebenstein et al. 2001). A function for the subrepeated region is not yet clear, however, DR A show obvious homology to box C (Fig. 2), an internal control region of the 5S rRNA genes (Pieler et al. 1987; Hemleben and

Werts 1988). Box C is also duplicated in presumptive 5S-pseudogenes in wheat (Van Campenhout et al. 1998).

Rearrangements in VR are preferentially associated with the repeated motifs. DEL 7 and DEL 10 correspondingly removed DR A1 and DR A2. In the short repeat variant of *S. raphanifolium*-II DEL 10 was compensated by reconstruction of DR A3 and A4 that led to the appearance of two perfect copies of DR A (Fig. 2: DEL 10+DUP 4). Several homologous bases surround the borders of DEL 7 (GAC), DEL 8 [GCG$^{G}_{T}$CG(G)T], DEL 10 [CGTT(G)AGGA], DEL 13 (GTG) and DEL 15 (GATGGG) (Fig. 1), as well as the border of a 55-bp long deletion found in clone pOII5S3–1 (AJ226065) of *S. okadae* (GAAA). Deletion in pseudogenes of *S. brevidens* also stretches from the GGATGG motif in box C to GGATTG in DR A1. A similar observation was made for the spacer region of 5S rDNA in *Larix* (Trontin et al. 1999). In *S. brevidens* a new class of repeats evolved after a 113-bp deletion comprising part of the 3′ coding and 5′ spacer region and subsequent amplification of this variant forming an independent satellite DNA-like component.

The relatively low level of intraspecific length and sequence polymorphism of 5S repeats indicates a concerted evolution of this tandem repeat family. Among 28 *Solanum* species studied an essential amount of different size classes was found only for *S. brevidens, S. polyadenium, S. microdontum* and *S. maglia,* whereas in other species different 5S repeat classes were not found or are present only as minor components. Low intragenomic length variability was observed also for 5S repeats of *Petunia hybrida* (Frasch et al. 1989).

Intralocus concerted evolution is much more effective than interlocus homogenization: repeats from the same locus are normally more similar as those from different loci. So, different 5S rDNA classes originating from diploid progenitors can still be distinguished in allopolyploid *Nicotiana tabacum* (Fulnecek et al. 1998) and in hexaploid *Triticum aestivum* (Van Campenhout et al. 1998). Hence, the high level of intragenomic similarity of the 5S repeat units of *Solanum* agree with the observation that this repeated family are present in *Solanum*/*Lycopersicon* only as a single locus at chromosome 1 (Gebhardt et al. 1991; Lapitan et al. 1991)**.** Nevertheless, interspersion of repeats of different length was observed in pea (Ellis et al. 1988) and in *Larix* (Trontin et al. 1999), suggesting a different efficiency of concerted evolution in plants.

In contrast to high intragenomic homogeneity of the spacer region in 5S rDNA, a considerable interspecific variability between the *Solanum* species studied has been observed. Numerous species-specific indels of various length as well as base substitutions are present here, which allowed phylogenetic reconstructions among this group.

## Taxonomic status of *Lycopersicon*

The taxonomic position of *Lycopersicon* still remains a subject of debate (Child 1990; D'Arcy 1991). Comparison of the 5S rDNA spacers shows that *Solanum* species from sect. *Petota* are more-closely related to *L. esculentum* as to *S. dulcamara,* which belongs to sect. *Dulcamara* of the same subgenus *Potatoe*. Hence, *Lycopersicon* should be replaced into *Solanum* as a sect. *Lycopersicum* (as proposed by Child 1990) or even as a subsection of sect. *Petota*. This interpretation agrees well with other molecular data on chloroplast DNA (Spooner et al. 1993) or 18–25S ribosomal RNA genes (Borisjuk et al. 1994).

## Conservation of ancestral 5S repeat structure

*S. brevidens* from subsect. *Estolonifera* (non-tuber-bearing species) is regarded as an "evolutionary link" between tomato and tuber-producing *Solanum* species (subsect. *Potatoe*; Hawkes 1990). This position is confirmed by sequence analysis of 5S rDNA (Figs. 1 and 3). The shape of flowers – primitive "*stellata*" or advanced "*rotata*" – is a significant taxonomic criterion for tuber-bearing species of subsect. *Potatoe*. Our data show that 5S rDNA remains very similar in several *Solanum* species with a different flower morphology and taxonomical position, such as *S. brevidens*, tuber-bearing *Solanum commersonii, S. circaeifolium* (both superser. *Stellata*) and *Solanum bukasovii* (superser. *Rotata*) indicating a relative conservation of 5S rDNA organization at early stages of divergence in sect. *Petota*.

Other tuber-bearing *Solanum* species evolved further as two or even three independent lineages. The first of these lineages contains Middle-American wild species from superser. *Stellata*: series *Bulbocastana* (*Solanum bulbocastanum*) and *Pinnatisecta* (*S. pinnatisectum* and *Solanum jamesii*) group together. Unusually, species of ser. *Polyadenia* bear a specific deletion at the beginning of VR (DEL 5) and should be accepted as a parallel second lineage. The third group characterized by a common DEL 7 includes representatives of ser. *Yungasensa*, *Megistacroloba* and superser. *Rotata* (with the exception of *S. bukasovii*; Fig. 3).

The close relationship between *S. pinnatisectum, S. jamesii* and *S. bulbocastanum* corresponds to the taxonomy established by Hawkes (1990) where these *Stellata* species are characterized as the most primitive among tuber-bearing potatoes. At the second step of evolution more advanced species appeared, like *S. circaeifolium, S. commersonii* and finally *Solanum chacoense* that developed further into *Rotata*. Our data do not support the first part of such a scenario, because the Middle American primitive *Stellata* and the South America *Rotata* probably evolved independently from a common ancestor. Our findings agree with the distribution of repetitive satellite elements among *Solanum*/*Lycopersicon* species (Stadler et al. 1995) and with the phylogenetic RFLP analyses of chloroplast DNA that shows early separation of *S. bulbocastanum* and *S. pinnatisectum* from *S. commersonii* and *Solanum capsicibaccatum* (ser. *Circaeifolia*) together with *Rotata* species (Spooner and

Castillo 1997). The crossing experiments for *Solanum* species support such a point of view: Louwes et al. (1992) have obtained fertile offspring by the sexual crossing of *S. circaeifolium* with *S. tuberosum* whereas no direct crossings were successfully carried out between wild species *S. bulbocastanum* or *S. pinnatisectum* and *S. tuberosum* (e.g. Matsubayashi 1991; Menke et al. 1996).

Combining this information with our new data it is possible to claim that different repetitive elements and/or morphological features evolve independently, i.e. they endure reconstruction at a different time. So, "primitive" *S. pinnatisectum, S. jamesii, S. bulbocastanum* and *S. polyadenium* demonstrate rearrangements within the 5S rDNA spacer, whereas the more "developed" *S. commersonii, S. circaeifolium* and even *S. bukasovii* show an ancestral organization. Taking together it is possible to propose that a species related to *S. commersonii* or *S. circaeifolium* with *Stellata*-morphology was an evolutionary link between non-tuber-bearing species and the first *Rotata*-species similar to *S. bukasovii*.

## Evolution of superser. *Rotata*

According to Hawkes (1990) *Solanum* species with a "*stellata*" or "*rotata*" morphology of the corolla should be combined together into corresponding superseries. Our data in general conform to the monophyletic origin of superser. *Rotata*, because all *rotata* species – with the exception of *S. bukasovii* – contain the same DEL 7 (Fig. 3). Apparently, *rotata* flowers appeared in the evolution before the separation of *S. bukasovii* and other *rotata* species, whereas DEL 7 took place just after that. *Rotata* species are traditionally divided into several series which now can be also distinguished by a comparison of 5S rDNA sequences (Fig. 1). *Solanum laxissimum* (ser. *Conicibacata*) and *S. raphalifolium* (ser. *Megistacroloba*) diverged from other *rotata* species at early stage of radiation (DEL 8; Fig. 3) and representatives of ser. *Tuberosa*, *Acaulia* and *Demissa*, both wild and cultivated species, are grouped together (DUP 5).

Remarkably, some *stellata* or primitive *rotata* species demonstrate close relationship with advanced *rotata* species. Thus: (1) VR of *S. chacoense* (ser. *Yungasensa*) is nearly identical to that of the long variant of *S. microdontum*, and (2) *S. raphanifolium* (ser. *Megistacroloba*) shows point mutations and DEL 8 in common with *S. laxissimum* (Fig. 1), indicating a secondary return to *stellata* morphology in at least two independent lineages (Fig. 3). This means that the morphology of flowers is a significant but not absolute taxonomic criterion for subsect. *Potatoe*. Superser. *Stellata* obviously has a polyphyletic nature, and the position of species from ser. *Yungasensa* should be re-examined, e.g. the position of *S. chacoense* should be discussed. Regarding s*tellata* morphology, it was originally placed together with *S. commersonii* into ser. *Commersoniana* by Hawkes (1978), Correll (1962) and Bukasov (1978), and only later replaced into ser. *Yungasensa* (Hawkes 1990) which agrees with the observation that *S. commersonii* does not cross with *S. chacoense* due to a difference in embryo balance number (EBN). Molecular data (Debener et al. 1990; Spooner and Castillo 1997; Miller and Spooner 1999) also placed *S. chacoense* among species of ser. *Tuberosa*.

Series *Maglia* is not confirmed. This series was established by Hawkes (1978, 1990) in the latest version of his taxonomy and was not distinguished by other authors (Correll 1962; Bukasov 1978). Our data show a close relationship between *S. maglia* and *S. microdontum*: long variants of 5S repeats in these two species are nearly identical (even specific point mutations are the same), and the only difference between short repeat variants is DEL 13 in *S. maglia*.

The molecular data presented agree well with the sporadic natural hybridization between *S. chacoense* and *S. microdontum*, *S. chacoense* and *Solanum spegazzinii*, *S. raphalifolium* and *Solanum sparsipilum* reported by Hawkes (1990). Remarkably, in accession II of *S. raphanifolium* two distinct 5S repeat variants were found. The long repeat is identical with *S. bukasovii* whereas the short one is very similar to the 5S rDNA of *S. raphanifolium*-I (Fig. 1). The simplest explanation for this observation is that *S. raphanifolium*-II actually represents a hybrid between *S. bukasovii* and *S. raphanifolium*. This example clearly demonstrates that: (1) morphological data may be not enough for the correct identification of closely related *rotata*-species and/or hybrids, and (2) comparison of 5S rDNA can be successfully used for this purpose.

## Polyploid species

The series *Acaulia* and *Demissa* containing polyploids are not confirmed by our data: the 5S rDNA of hexaploid *Solanum iopetalum* (ser. *Demissa*) is nearly identical to diploid *Solanum leptophyes* or *Solanum gourlayi* (ser. *Tuberosa*) whereas hexaploid *Solanum demissum* (ser. *Demissa*) is very similar to tetraploid *Solanum acaule* (ser. *Acaulia*). These findings agree with data obtained by RFLP (Debener et al. 1990) or AFLP (Kardolus et al. 1998) analyses: *S. demissum* and *S. acaule* exhibit a close relationship and may be clearly distinguished from the *Tuberosa* wild species as well as from the breeding lines of potato**.** Independent origin of Mexican hexaploid *S. demissum* and *Solanum iopetalum* demonstrated here explain the lack of hybridization of these species (Hawkes 1990).

According to Hawkes (1990) Middle-American *stellata* diploids participated in the origin of *rotata* hexaploids *S. demissum* and *S. iopetalum*. In spite of their presumptive allopolyploid nature (Matsubayashi 1991), all 5S clones isolated from these species are nearly identical. Two possibilities can explain a high intragenomic homogeneity of 5S rDNA repeats in *S. demissum* and *S. iopetalum*: (1) Parental diploids were closely related, or (2) high sequence similarity appeared as a result of

the differential elimination of 5S repeats of one of the diploid progenitors and/or interlocus conversion after alloploidization, as was demonstrated for 18–25S rDNA in alloploid *Gossypium* (Wendel et al. 1995) and *Nicotiana* (Volkov et al. 1999). Therefore, the rate of 5S rDNA rearrangement in alloploids seems to be much lower than that for 18–25S rDNA because 5S rDNA repeats originating from different progenitors are still present in such "old" polyploids as *Nicotiana tabacum* (Fulnecek et al. 1998). Hence, the first possibility seems to be more reasonable and, therefore, the participation of Middle-American *stellata* diploids in the origin of *S. demissum* and *S. iopetalum* should be rejected.

### Origin of *S. tuberosum*

A comparison of the VR demonstrates that breeding lines of *S. tuberosum* may be distinguished and grouped with different diploid species e.g. B15, BP1076 and *S. spegazzinii* bear a specific DEL 14, B1 differs only by one base substitution from *Solanum phureja*, and R1 contains two insertions identical to those of *S. demissum* and *S. acaule*. The question of the origin of tetraploid *S. tuberosum* has been actively discussed and several diploid species – *Solanum bukasovii, S. gourlayi, S. leptophyes, S. phureja, S. sparsipilum, S. spegazzinii, Solanum stenotomum,* etc. – were proposed as presumptive candidates (Bukasov 1978; Hawkes 1990; Hosaka 1995). Our data confirm such a point of view for *S. phureja* and *S. spegazzinii* whereas *S. bukasovii* should probably be excluded from the list of candidates. Therefore, more accessions of wild species and breeding lines should be investigated in order to finally clarify the situation. Nevertheless, our new data clearly confirm the suggestion that several species participated in the natural origin of *S. tuberosum* or were later used for crossing in order to introduce agronomically useful traits into the cultivated potato. For instance, *S. demissum* had often been used as a source of resistance to *Phytophthora infestans* or to potato virus X (Hawkes 1990).

### 5S rDNA as a molecular marker for breeding purposes

The spacer region in the 5S rDNA of *Solanum*, especially of superser. *Rotata*, demonstrates a high rate of molecular evolution that produced numerous indels. Specificity of these indels allowed the discrimination of species and even breeding lines that can be used for reconstruction of phylogeny and for the design of species-specific molecular markers for breeding programs. Previously, a synthetic oligonucleotide corresponding to the specific sequence within the spacer was used as a hybridization probe (Zanke et al. 1995) in order to confirm the presence of the parental *S. pinnatisectum* genome in somatic hybrids obtained by protoplast fusion of this species with the breeding line *S. tuberosum* B15 (Menke et al. 1996).

Therefore, the considerable interspecific variability of 5S rDNA between the *Solanum* species studied, especially the species-specific indels of various lengths, allows detailed phylogenetic reconstruction among this group of Solanaceae and can be used for interspecific differentiation.

**Acknowledgements** This work was supported by Bundesministerium für Forschung und Technologie (BMFT), Gesellschaft zur Förderung der Pflanzenzüchtung (GFP) and Alexander von Humboldt Foundation. Authors are grateful to Dr. L. Schilde and Prof. Dr. H. Ninnemann for providing in vitro plant material and to N. Komarova for technical assistance.

## References

Barciszewska M, Erdman VA, Barciszewski J (1994) A new model for the tertiary structure of 5S ribonucleic acid in plants. Plant Mol Biol Rep 12:116–131

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 N, Borisjuk L, Komarnytsky S, Timeva S, Hemleben V, Gleba Y, Raskin I (2000) The tobacco ribosomal DNA spacer element stimulates amplification and expression of heterologous genes. Nature Biotechnol 18:1303–1306

Bukasov SM (1978) Systematics of the potato. Bull Appl Bot Genet Breed 62:1–42

Child A (1990) A synopsis of *Solanum* subgenus *Potatoe* (G. Don) (D'Arcy) [Tuberarium (Dun.) Bitter (s.l.)]. Feddes Rep 95:283–298

Correll DS (1962) The potatoe and its wild relatives. Contrib Texas Res Found, Bot Stud 4:1–606

Crisp MD, Appels R, Smith FM, Keys WMS (1999) Phylogenetic evaluation of 5S ribosomal RNA gene and spacer in the *Callistachys* group (*Fabaceae*: *Mirbelieae*). Plant Syst Evol 218:33–42

D'Arcy WG (1991) The Solanaceae since 1976, with a review of its biogeography. In: Hawkes JG, Lester RN, Nee M, Estrada N (eds) Solanaceae III: taxonomy, chemistry, evolution. Roy Bot Garden, Kew, pp 75–137

Debener T, Salamini F, Gebhardt C (1990) Phylogeny of wild and cultivated *Solanum* species based on nuclear restriction fragment length polymorphism (RFLPs). Theor Appl Genet 79: 360–368

Ellis THN, Lee D, Thomas CM, Simpson PR, Cleary WG, Newmann M-A, Burcham KWG (1988) 5S rRNA genes in *Pisum*: sequence, long range and chromosomal organization. Mol Gen Genet 214:333–342

Felsenstein J (1985) Confidence limits on phylogenetic trees. Science 155:279–284

Frasch M, Wenzel W, Hess D (1989) The nucleotide sequences of nuclear 5S rRNA genes and spacer regions of *Petunia hybrida*. Nucleic Acids Res 17:2857

Fulnecek J, Matyasek R, Kovarik A, Bezdek M (1998) Mapping of 5-methylcytosine residues in *Nicotiana tabacum* 5S rRNA genes by genomic sequencing. Mol Gen Genet 259:133–141

Gebhardt C, Ritter E, Barone A, Debener T, Walkemeier B, Schachtschabel U, Kaufmann H, Thomson RD, Bonierbale MW, Ganal MW, Tanksley SD, Salamini F (1991) RFLP maps of potato and their alignment with the homologous tomato genome. Theor Appl Genet 83:49–57

Grebenstein B, Winterfeld G, Röser M, Hemleben V (2001) Molecular diversity and physical mapping of 5S rDNA in wild and cultivated oat grasses (Poaceae: Agrostideae). Mol Phylog Evol (in press)

Hawkes JG (1978) Biosystematics of the potato In: Harris PM (ed) The potato crop – the scientific basis for improvement. Chapman and Hall, London, pp 15–69

Hawkes JG (1990) The potato evolution, biodiversity and genetic resources. Smithsonian Institution Press, Washington DC

Hemleben V, Werts D (1988) Sequence organization and putative regulatory elements in the 5S rRNA genes of two higher plants (*Vigna radiata* and *Matthiola incana*). Gene 62:165–169

Hemleben V, Zanke C, Panchuk I, Volkov R (1998) Repetitive elements as molecular markers in potato breeding. Beitr Züchtungsforsch 4:61–67

Hosaka K (1995) Successive domestication and evolution of the Andean potatoes as revealed by chloroplast DNA restriction endonuclease analysis. Theor Appl Genet 90**:**356–363

Kardolus JP, Van Eck HJ, van den Berg RG (1998) The potential of AFLPs in biosystematics: a first application in *Solanum* taxonomy (Solanaceae). Plant Syst Evol 210**:**87–103

Lapitan NLV, Ganal MW, Tanksley SD (1991) Organization of the 5S ribosomal RNA genes in the genome of tomato. Genome 34:509–514

Louwes KM, Hoekstra R, Mattheij WM (1992) Interspecific hybridisation between the cultivated potato *Solanum tuberosum* subspecies *tuberosum* L and the wild species *S. circaeifolium* subsp. *Circaeifolium* Bitter exhibiting resistance to *Phytophytora infestans* (Mont.) de Bary and *Globodera pallida* (Stone) Behrens. Theor Appl Genet 84:362–370

Matsubayashi M (1991) Phylogenetic relationships in the potato and its related species In: Tsuchiya T, Gupta PK (eds) Chromosome engineering in plants: genetics, breeding, evolution, Part B. Elsevier, Amsterdam, pp 93–118

Menke U, Schilde-Rentschler L, Ruoss B, Zanke C, Hemleben V, Ninnemann H (1996) Somatic hybrids between the cultivated potato *Solanum tuberosum* L and the 1EBN wild species *Solanum pinnatisectum* Dun: morphological and molecular characterization. Theor Appl Genet 92:617–626

Miller JT, Spooner DM (1999) Collapse of species boundaries in the wild potato *Solanum brevicaule* complex (Solanaceae, sect. Petota): molecular data. Plant Syst Evol 214:103–130

Myers EW, Miller W (1988) Optimal alignment in linear space. CABIOS 4:11–17

Pearson WR (1995) Comparison of methods for searching protein sequence databases. Protein Sci 4:1145–1160

Pieler T, Hamm J, Roeder RG (1987) The 5S gene internal control region is composed of three distinct sequence elements, organised as two functional domains with variable spacing. Cell 48:91–100

Sambrook J, Fritsch E, Maniatis T (1989) Molecular cloning. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

Scoles CJ, Gill GS, Xin Z-Y, Clarke BC, McIntyre CL, Chapman C, Appels R (1988) Frequent duplication and deletion events in the 5S RNA genes and the associated spacer region of the Triticeae. Plant Syst Evol 160:105–122

Spooner DM, Castillo RT (1997) Reexamination of series relationsphips of South American wild potatoes (Solanaceae: *Solanum* Sect. *Petota*): evidence from chloroplast DNA restriction site variation. Amer J Bot 84:671–685

Spooner DM, Anderson GJ, Jansen RK (1993) Chloroplast DNA evidence for the interrelationships of tomatoes, potatoes, and pepinos (Solanaceae). Am J Bot 80:676–688

Stadler M, Stelzer T, Borisjuk N, Zanke C, Schilde-Rentschler L, Hemleben V (1995) Distribution of novel and known repeated elements of *Solanum* and application for the identification of somatic hybrids among *Solanum* species. Theor Appl Genet 91:1271–1278

Trontin J-F, Grandemange C, Favre J-M (1999) Two highly divergent 5S rDNA unit size classes occur in composite tandem array in European larch (*Larix deciduas* Mill.) and Japanese larch [*Larix kaempferi* (Lamb.) Carr.]. Genome 42:837–848

Van Campenhout S, Aert R, Volckaert G (1998) Orthologous DNA sequence variation among 5S ribosomal RNA gene pacer sequences on homoeologous chromosomes 1B, 1D and 1R of wheat and rye. Genome 41:244–255

Van den Berg RG, Groenduk-Wilders N, Kardolus JP (1996) The wild ancestors of cultivated potato: the *brevicaule* complex. Acta Bot Neerl 45:157–171

Venkateswarlu K, Lee S-W, Nazar RN (1991) Conserved upstream sequence elements in plant 5S ribosomal RNA-encoding genes. Gene 105:249–253

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

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

Zanke C, Borisjuk N, Ruoss B, Schilde-Rentschler L, Ninnemann H, Hemleben V (1995) A specific oligonucleotide of the 5S rDNA spacer and species-specific elements identify symmetric somatic hybrids between *Solanum tuberosum* and *S. pinnatisectum*. Theor Appl Genet 90**:**720–726