Theor Appl Genet (2000) 101:7–14

A. Schmidt · R.L. Doudrick · J.S. Heslop-Harrison
T. Schmidt

The contribution of short repeats of low sequence complexity to large conifer genomes

Received: 1 October 1999 / Accepted: 3 November 1999

Abstract The abundance and genomic organization of six simple sequence repeats, consisting of di-, tri-, and tetranucleotide sequence motifs, and a minisatellite repeat have been analyzed in different gymnosperms by Southern hybridization. Within the gymnosperm genomes investigated, the abundance and genomic organization of micro- and minisatellite repeats largely follows taxonomic groupings. We found that only particular simple sequence repeat motifs are amplified in gymnosperm genomes, while others such as (CAC), and (GACA), are present in only low copy numbers. The variation in abundance of simple sequence motifs reflects a similar situation to that found in angiosperms. Species of the two- and three-needle pine section *Pinus* are relatively conserved and can be distinguished from *Pinus strobus* which belongs to the five-needle pine section *Strobus.* The hybridization pattern of *Picea* species, bald cypress and gingko were different from the patterns detected in the *Pinu*s species. Furthermore, sequences with homology to the plant telomeric repeat (TTTAGGG), have been analyzed in the same set of gymnosperms. Telomere-like repeats are highly amplified within two- and three-needle pine genomes, such as slash pine (*Pinus elliottii* Engelm. *var. elliottii),* compared to *F! strobus, Picea* species, bald cypress and gingko. *P. elliottii var. elliottii* was used as a representative species to investigate the chromosomal organization of telomere-like sequences by fluorescence in situ hybridization (FISH). The telomere-like sequences are not restricted to the ends of chromosomes; they form large intercalary and pericentric blocks showing that they are a repeated component of the slash pine genome. Conifers have genomes larger than 20000 Mbp, and our results clearly demonstrate that repeats of low sequence complexity, such to (CA),, (GA),, (GGAT), and (GATA),, and minisatellite- and telomere-like sequences represent a large fraction of the repetitive DNA of these species. The striking differences in abundance and genome organization of the various repeat motifs suggest that these repetitive sequences evolved differently in the gymnosperm genomes investigated.

Key words *Pinus* . Gymnosperms . Simple sequence repeats . Microsatellites . Minisatellites . Telomere

Communicated by P.M.A. Tigerstedt

A. Schmidt ([XI) . T. Schmidt
Institute of Crop Science and Plant Breeding, Olshausenstraße 40, D-241 18 Kiel, Germany
e-mail: kamm@plantbreeding.uni-kiel.de
Tel.: +49/431/8803288, Fax: +49/431/8802566

R.L. Doudrick
Southern Institute of Forest Genetics, Southern Research Station, USDA Forest Service, 23332 Highway 67, Saucier, MS 39574-9344, USA

J.S. Heslop-Harrison
Karyobiology Group, Department of Cell Biology, John Innes Centre, Colney Lane, Norwich NR4 7UH, U.K.

Introduction

Plant genomes contain several classes of repetitive DNA organized in tandem arrays. Satellite DNA repeats and rRNA genes have been extensively investigated at the molecular and chromosomal level, respectively, showing that they range from 150-180 bp or 300-360 bp up to several kilobases in length.

However, a large proportion of the plant repetitive DNA also consists of tandemly arranged repeats which are much shorter and of lower complexity compared to satellite DNA and rRNA genes. Monotonous repetitions of sequence motifs (1-5 bp) are known as simple sequence repeats or microsatellites. They have been found in many plant species (Weising et al. 1992), and although they are ubiquitous elements of eukaryotic genomes, there are remarkable differences in the abundance of microsatellite sequence motifs between plants and vertebrates (Lagercrantz et al. 1993).

Because of their widespread occurence, hypervariability in repeat number and their interspersed multilocus distribution in eukaryotic genomes, microsatellites are highly informative markers for plant and mammalian ge-

nome analysis and mapping. Although some of these loci may consist of complex or compound microsatellites, the overall simplicity of the basic repeating motifs is maintained. In plants, microsatellites have been used to map the genomes of rice (Wu and Tanksley 1993), soybean (Akkaya et al. 1992) and cereal species (Becker and Heun 1995). In rapeseed (*Brassica napus*), chickpea (*Cicer arietinum)* and sugar beet *(Beta vulgaris),* DNA fingerprinting has shown that di-, tri- and tetranucleotide repeats are amplified sequences of the genome and are highly informative markers for the differentiation of accessions, cultivars and breeding lines (Weising et al. 1992; Poulsen et al. 1993; Schmidt et al. 1993).

Another type of short DNA repeats are minisatellites, which are characterized by longer repeating units (10–50 bp). First isolated from mammalian genomes, minisatellites have been detected in many plant species (Broun and Tanksley 1993; Winberg et al. 1993; Tourmente et al. 1994). Minisatellite repeats share a GC-rich core motif and are highly polymorphic enabling their application as markers for genome mapping. Minisatellite repeats can be amplified to high copy numbers and make-up a considerable fraction of the repetitive DNA sequences of plant genomes (Tourmente et al. 1994; Metais et al. 1998).

Functional repeats of low sequence complexity are the telomeres. Occurring at the outermost ends of plant chromosomes, these repeats are highly conserved, forming stabilizing (TTTAGGG), stretches at the physical ends of chromosomes that protect the linear chromosomes from exonucleolytic degradation. Proximal telomere repeat copies are less conserved and sequence heterogeneity increases with distance from the chromosome ends, although the basic structure of these telomere-like repeats can still be recognized. In *Arabidopsis thaliana,* it has been reported that telomere-like sequence blocks occur also in intercalary positions on chromosome I (Richards et al. 1993). By fluorescence in situ hybridization Fuchs et al. (1995) have shown that intercalary blocks of telomere-like sequences are found in many plant genomes.

Conifers are commercially important and inherently interesting because they dominate many terrestrial ecosystems of the world. Despite the large genome size of gymnosperms (> 20 000 Mbp for pine species; Wakamiya et al. 1993) little is known about the genomic structure and composition of the nuclear genome of conifer species. With the exception of chromosome numbers, which are very conserved (2n = 24) (Khoshoo 1961), there are few investigations of the undoubtedly high proportion of noncoding DNA in the genomes of gymnosperms. Reassociation kinetics data revealed 75% of the genome to be repetitive DNA (Dhillon 1987). For slash pine (*Pinus elliottii* Engelm. var. *elZiottii;* 2n = 2x = 24) only 18S-5.8S-25S rRNA genes and the *Ty*1-*copia*-retrotransposon *TPE1* have been characterized as major classes of repetitive DNA (Doudrick et al. 1995; Kamm et al. 1996) and a Ty3-*gypsy* retrotransposon element was characterized as highly amplified in *Pinus radiata* (Kossack et al. 1990). A satellite DNA family has been cloned from *Picea* species and localized along the chromosomes (Brown et al. 1998).

Here we demonstrate that short repeats of simple sequence structure and low complexity are amplified in gymnosperm species, in particular conifers. We aimed to investigate the abundance and genomic organization of simple sequence repeats consisting of di-, tri-, and tetranucleotide motifs, and a minisatellite repeat in different gymnosperms by Southern hybridization using synthetic oligonucleotides as probes. By fluorescence in situ hybridization on slash pine chromosomes we show that sequences with homology to the plant telomeric repeat (TTTAGGG), are amplified at many intercalary regions and are a major component of the *P. elliottii.* var. *elliottii* genome.

Material and methods

Plant material, DNA extraction, and in-gel hybridization

Total genomic DNA was extracted from needle tissue (10 g fresh weight) following the protocol of Wagner et al. (1987). The plant species used are listed in Table 1.

For in-gel hybridization, genomic DNA was digested using *Rsa*I and separated on 1% agarose gels. Gels were dried in a slab dryer, pre-treated and hybridized with [γ-^{32}P]dATP end-labelled oligonucleotides (Ali et al. 1986). The following oligonucleotides were employed: $(CA)_5$; $(GA)_5$; $(CAC)_5$; $(GGAT)_4$; $(GACA)_4$; $(GATA)_4$ and the A. *thaliana* minisatellite sequence GGAGGTGGAGGACATGGCGG (Tourmente et al. 1994). The hybridization and stringent washes were performed at the appropriate T_m - 5°C according to Thein and Wallace (1986). Exposure times were dependent on the amount of bound radio-labelled probe and varied, for example, between 5 h for (GA), 26 h for the minisatellite motif and 40 h at -80°C using intensifying screens for (GACA).

DNA labelling and Southern hybridization

The nonradioactive chemiluminescence method ECL (Amersham) was used for DNA labelling of pAtT4, a clone containing several copies of the *A. thaliana* telomeric repeat TTTAGGG (Richards and Ausubel 1988), hybridization and detection. For preparation of Southern blots, 10 µg of genomic DNA were digested using *Rsa*I, size-fractionated in a 1% agarose gel and transferred onto a nylon membrane (Sambrook et al. 1989). The hybridization, with a membrane DNA concentration of 10 ng/cm^2, was carried out overnight at a stringency of 90%.

Fluorescence in situ hybridization

Seedling root tips of *P. elliottii* var. *elliottii* were used for chromosome preparations. After treatment using colchicine, root tips were fixed in methanol/acetic acid (3: 1). Chromosome preparations and in situ hybridization were done according to Doudrick et al. (1995). pAtT4, carrying the *A. thaliana* telomeric sequence, was labelled with biotin-11-dUTP (Sigma) by PCR. The in situ hybridization was performed with the most stringent wash being 40% formamid in 2 x SSC at 42°C. Sites of hybridization were detected using a streptavidin-Cy3 conjugate (Sigma). Slides were counterstained with DAPI (4',6-diamidino-2-phenylindole).

Results

Short repeats of low sequence complexity in gymnosperm genomes

The existence and distribution of sequences complementary to $(CA)_8$, $(GA)_8$, $(CAC)_5$, $(GACA)_4$, $(GGAT)_4$, (GATA)$_4$, as well as a plant minisatellite consensus sequence, were analyzed in various two- and three-needle pines of the section *Pinus*, the five-needle pine *Pinus strobus* of the section *Strobus*, and in two *Picea* species and *Taxodium distichum* (Table *1). Gingko biloba* was used as an out-group, non-coniferous gymnosperm species. Most of these gymnosperm species are characterized by extremely large genomes. In order to optimize hybridization experiments, several enzymes commonly used for fingerprint studies such as *Rsa*I, *Hin*fI, *Hae*III, *Taq*I were employed for the digestion of genomic DNA (data not shown). All restriction patterns resulted in a hybridization smear, and *Rsa*I was chosen for further investigations because it cuts frequently and is not inhibited by methylation.

Hybridization of the dinucleotide repeat (CA), to *Rsa*I-digested genomic DNA revealed that this sequence motif is equally amplified and present in high copy numbers in all gymnosperms investigated here. Strong hybridization signals visible for all species as a smear over the whole lane up to 5 kb could be detected, and most CA repeat copies are present in genomic fragments of approximately 0.5–3.0 kb (Fig. 1A).

Rehybridization of the same gel with (GA), resulted in a pattern which is more pronounced in regions of larger restriction fragments compared to CA. Strong hybridization with high-molecular weight fragments up to 20 kb were detected in all *Pinus* species. The hybridization signals of GA are more complex and, despite the smear, many individual bands were visible. The motif GA was particularly abundant in the genome of *Pinus resinosa* (Fig. 1B, lane 8). The strength of hybridization with CA and GA showed that these dinucleotide repeats are major components of the gymnosperm genomes investigated but show a different genome organization (cf. Fig. 1A and 1B).

Only a very weak signal could be observed after hybridization of (CAC), and (GACA), with all species, suggesting that these sequence motifs are infrequent in the genomes investigated (Fig. 1C, E). We note that (GACA) shows 75% homology to the abundant (GA) and (CA) repeats, demonstrating that hybridization conditions clearly discriminated the sequences.

Hybridization of (GGAT), and (GATA), showed medium to strong hybridization in the gymnosperm genomes investigated (Fig. 1D, F). (GGAT), revealed a strong hybridization with a fragment of 1.3 kb which is conserved in most two- and three-needle pine species such as *Pinus echinata, P. elliottii* var. *elliottii* and *Pinus palustris*. However, this fragment was not detected in the two- and three-needle *Pinus* species *l? resinosa* and *Pinus massoniana,* nor in *P. strobus, Picea* species, *T. distichum* and *G. biloba.* Among the investigated gymnosperm species, the highest copy number of both (GGAT), and (GATA), was detected in *T. distichum* and *G. biloba* demonstrating that these simple sequence repeats are very abundant sequences of these genomes. In *G. biloba,* (GATA), hybridized to multiple fragments between 3.5 kb and approximately 8.0 kb.

In addition to simple sequence repeats, many plant genomes also contain minisatellite repeats, which are characterized by a longer repeating unit and a high CC-rich base constitution. Therefore, we used an oligonucleotide representing an *A. thaliana* minisatellite consensus motif to investigate the occurrence of this sequence in the gymnosperm genomes included in this study. The minisatellite-like motif is abundant in two- and three-needle pines, while it is only moderately amplified in the five-needle pine, *P. strobus,* in *Picea* species, *T. distichum* and *G. biloba* (Fig. 2A). The results show clearly that these sequences form a considerable fraction of *Pinus* genomes.

Table 1 Gymnosperm species used in this study. All samples were from the collection of the Southern Institute of Forest Genetics, U.S. Department of Agriculture, Saucier (USA)

Lane	Genus	Section	Subsection	Species	Common name	Source
	Pinus L.	*Pinus*	***Australes*** Loud.	*P.* ***echinata*** Mill.	Shortleaf pine	Harrison County, MS
2				*P.* ***elliottii*** Engelm. var. ***elliottii***	Slash pine	Harrison County, MS
				P. ***palustris*** Mill.	Longleaf pine	Harrison County, MS
4				*P. caribaea* Morelet	Caribbean pine	Puerto Rico
5			***Oocarpae*** Little & Critchfield	*P.* ***oocarpa*** Schiede		Puerto Rico
6			*Contortae* Little & Critchfield	*P. banksiana* Lamb.	Jack pine	Oneida County, WI
7			***Sylvestres*** Loud.	*P.* ***massoniana*** Lamb.	Masson pine	Harrison County, MS
8				*P.* ***resinosa*** Ait.	Red pine	Oneida County, WI
9		***Strobus*** Lemm.	***Strobi*** Loud.	*F!* ***strobus*** L.	White pine	Oneida County, WI
10	***Picea*** Diet.			*P.* ***glauca*** (Moench.) Voss.	White spruce	Oneida County, WI
11				*P.* ***abies*** (L.) Karst.	Norway spruce	Oneida County, WI
12	*Taxodium* Rich.			*T.* ***distichum*** (L.) Rich.	Bald cypress	Harrison County, MS
13	***Gingko*** L.			***G. biloba*** L.	Gingko	Harrison County, MS

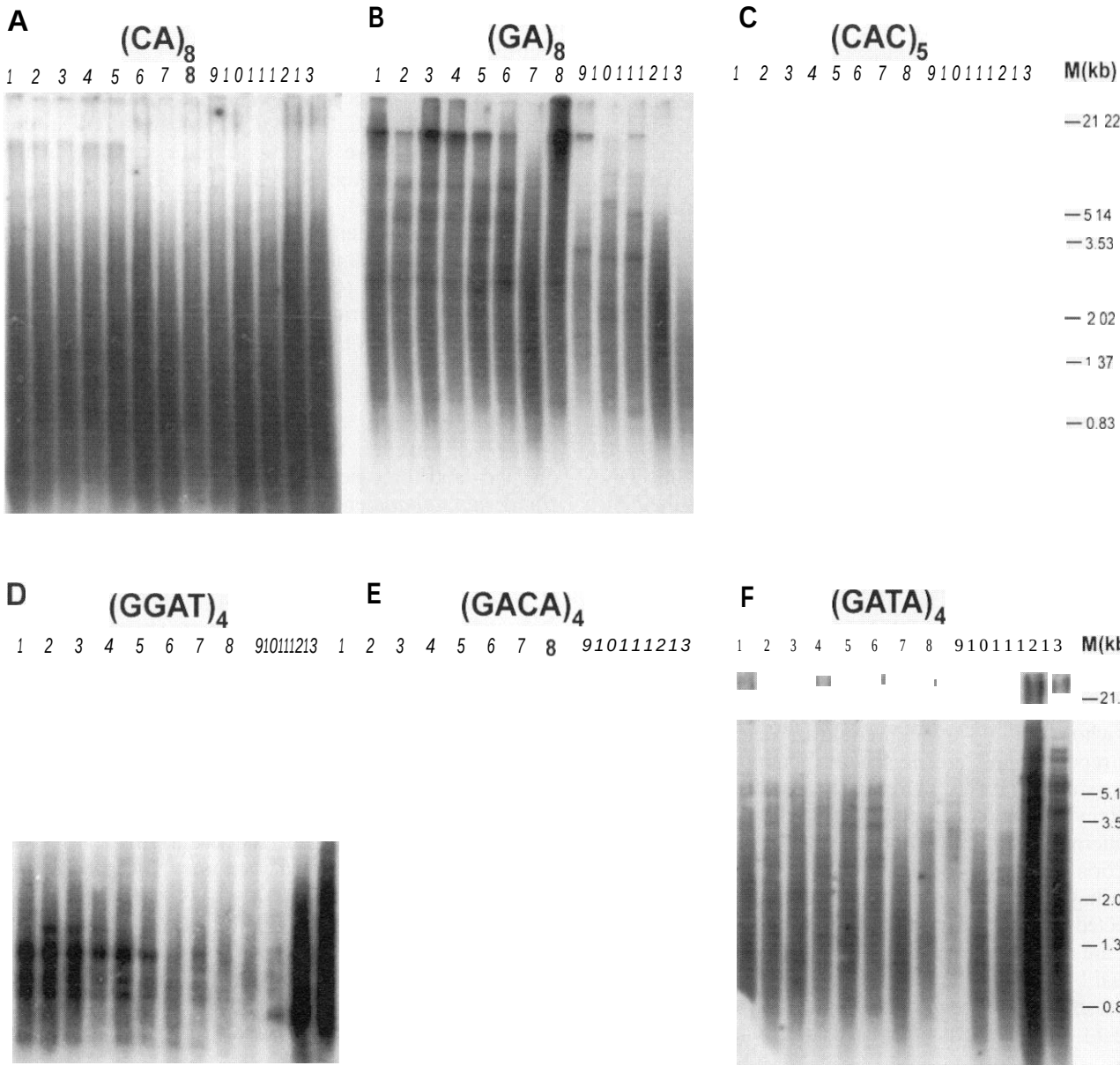

Fig. 1 A-F Genomic distribution of simple sequence repeats in several gymnosperm genomes. Different simple sequence repeats (motif given above each panel) were hybridized to *Rsa*I-digested genomic DNA of *Pinus echinata* (I). *Pinus elliottii* var. *elliottii* (2), *Pinus palustris* (3), *Pinus caribaea* (4), *Pinus oocarpa* (5), *Pinus banksiana* (6), *Pinus massoniana* (7), *Pinus resinosa* (R), *Pinus strobus* (9), *Picea glauca* (10), *Picea abies* (II), *Taxodium distichum* (12) and *Gingko biloba* (13). Lambda *Hin*dIII/*Eco*RI-digested DNA was used as the DNA size marker (lane *M*)

Amplification of telomere-like repeats in pine genomes

Plant genomes, with their large amount of nuclear DNA might contain blocks of the telomeric repeat (TTTAGGG), at intercalary positions (Richards et al. 1991).

Probe pAtT4, carrying the TTTAGGG tandem repeats of the A. *thaliana* telomere (Richards and Ausubel 1988), was used for Southern hybridization to digested DNAs from various *Pinus* and *Picea* species, *T. distichum* and G. *biloba*. Information for *Pinus* species was of most interest, and hence short exposures were needed, otherwise the lanes containing *Pinus* DNA were overexposed. The Southern filters showed that the telomeric repeat is highly amplified within two- and three-needle pines of the section *Pinus* (Fig. 2B, lanes l-8). The hybridization pattern revealed a smear over the whole range up to high-molecular weights with some amplified fragments, suggesting the presence of the telomere repeats at different genomic loci with a conserved organiration. Hybridization signals larger than 20 kb indicate

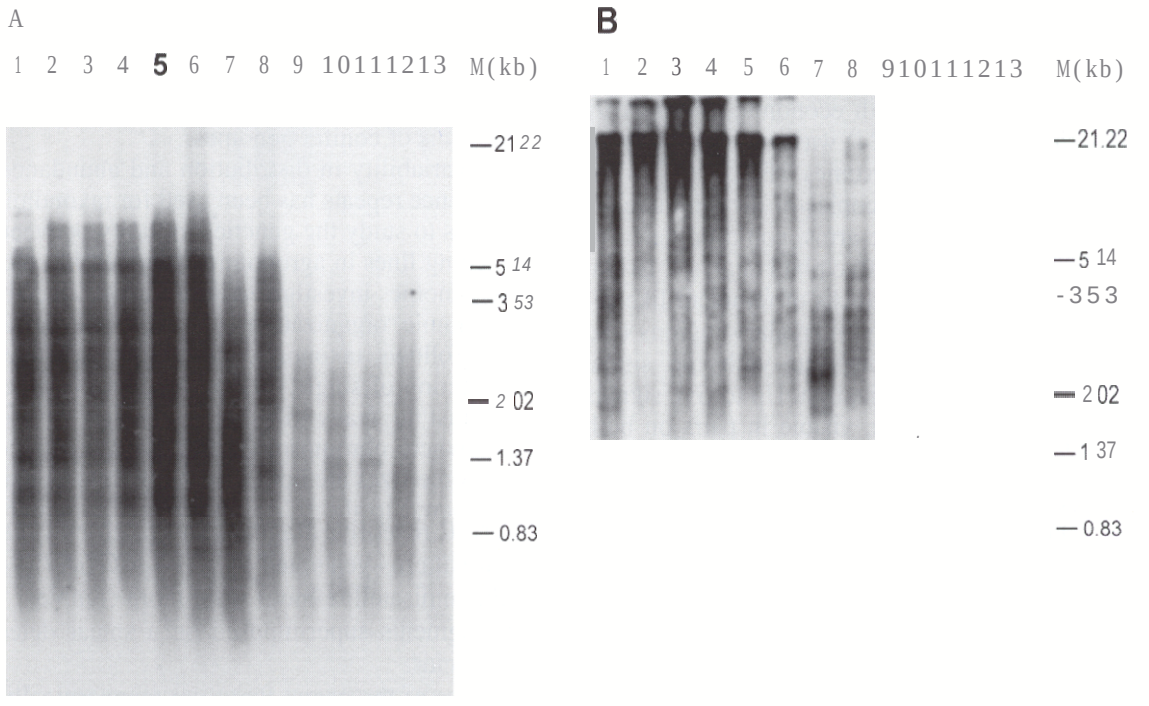

Fig. 2 A, B Genomic and chromosomal organization of a minisatellite sequence and the telomeric sequence 5'(TTTAGGG)$_n$-3' in various gymnosperm genomes. **A** A minisatellite-like repeat was hybridized to *Rsa*I-digested genomic DNA of *Pinus echinata* (*1*), *Pinus elliottii* var. *elliottii* (*2*), *Pinus palustris* (*3*), *Pinus caribaea* (*4*), *Pinus oocarpa* (*5*), *Pinus banksiana* (*6*), *Pinus massoniana* (*7*), *Pinus resinosa* (*8*), *Pinus strobus* (*9*), *Picea glauca* (*10*), *Picea abies* (*11*), *Taxodium distichum* (*12*) and *Gingko biloba* (*13*). Lambda *Hin*dIII/*Eco*RI-digested DNA (*M*) was used as the DNA size marker. **B** Hybridization of pAt4 containing the telomeric sequence 5'(TTTAGGG3)$_n$-3' to the same set of species

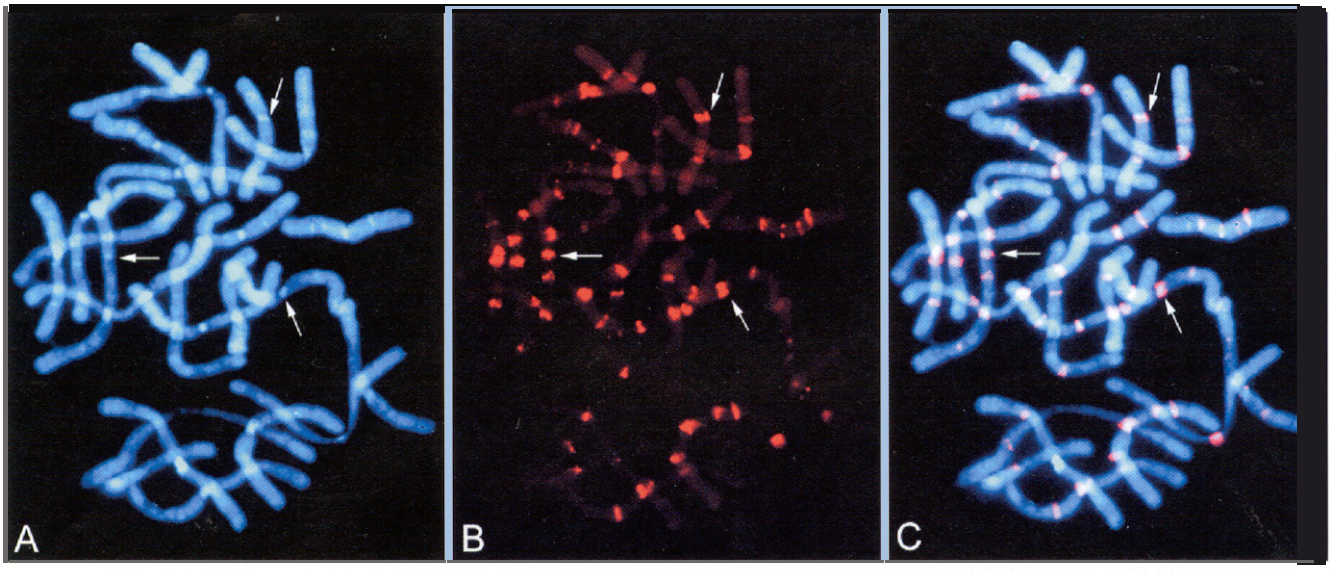

Fig. 3 Chromosomal localization of telomere-like sequences along chromosomes of *P. elliottii* var. *elliottii* by fluorescent in situ hybridization. Arrows point to examples of intercalary and pericentromeric regions which stain brightly with DAPI and consist of telomere-like DNA sequences. **A** DAPI staining (blue fluorescence) of a metaphase plate shows the chromosome morphology of *P. elliottii* var. *elliottii*. **B** Sites of hybridization of the biotinylated probe pAt4 were detected with streptavidin-Cy3 (bright red fluorescence) and show regions consisting of telomere-like sequences. **C** Computer-generated overlay of both images shows the co-localization of telomere-like sequence blocks with many DAPI-positive regions

that some fragments lack recognition sites of the frequently cutting enzyme *Rsa*I. Weaker signals were observed in *P. massoniana* and *P. resinosa* of the subsection *Sylvestris* as compared to other species of the section.

Much less signal was observed in *P. strobus* (five-needle pine section *Strobus,* Fig. 2B, lane 9), in *Picea* species, *T. distichum* and *G. biloba* (lanes 10–13), demonstrating much lower copy numbers in these species.

Fluorescence in situ hybridization was used to test whether telomere-like sequences are located in other regions than the physical ends of the *Pinus* chromsomes. As a representative species with economic relevance *P. elliotti var. elliottii* (2n = 2x = 24) was used for the preparation of metaphase chromosomes (Fig. 3A). Weak sites, indicating relatively few copies of the telomere repeat at the end of chromosome arms, could be detected by computer-enhanced quantitative pixel analysis. Bright signals representing extended blocks of telomere-like sequences were observed at many intercalary regions (red fluorescence in Fig. 3B), consistent with the signal strength of the Southern hybridization. Up to five signals were detected per chromosome, visible in pairs on both chromatids. The strongest amplification at non-telomeric sites occurred in the pericentric regions of many chromosomes (Fig. 3C). Most of the telomere-like DNA sites correspond to DAPI-positive regions.

Discussion

We have investigated the occurrence and genomic organization of simple sequence repeats, minisatellite and telomere-like sequences and their contribution to the extremely large genomes of various gymnosperms species.

The conifer species analyzed in this study have nuclear DNA amounts greater than 20 000 Mbp, and 75% of the genomes consist of repetitive sequences (Dhillon 1987; Wakamiya et al. 1993; Murray 1998). The hybridization patterns, often detected as a smear including many fragments, are presumably a consequence of the large genome size, since a very large number of fragments contribute to the Southern pattern if the investigated simple repeat motif is present in the genome. The wide range of molecular weights demonstrates that these sequences are present at many different loci and are presumably dispersed among other sequences. This pattern is in contrast to the clear-cut polymorphic fingerprints often detected in many angiosperm plant species (Weising et al. 1991; Poulsen et al. 1993; Schmidt et al. 1993). However, repetitive fragments were detected in several taxa, e.g. with (GA) in *Pinus* species (Fig. 1B) and with $(GATA)_4$ in *G. biloba* (Fig. 1F).

In general, simple sequence repeats are an abundant class of repeats in plant genomes, but it is not yet clear why some sequence motifs are more abundant or more polymorphic than others (Beyermann et al. 1992; Poulsen et al. 1993; Depeiges et al. 1995). The low copy number of $(CAC)_5$ and $(GACA)_4$ in conifers and G. *biloba* is in agreement with the findings in angiosperms and indicates that the structure of large conifer genomes and the organization of simple sequence repeats is to some extent similar to other higher plants. Studies of Smith and Devey (1994) showed that the dinucleotides GA and CA are highly amplified in three other pine genomes, namely *P. radiata, Pinus taeda* and *Pinus sylvestris.* Based on genomic library screening Echt and May-Marquardt (1997) calculated that GA and CA are the most abundant simple sequence motifs in *P. strobus* and *P. tueda.* Thus, our hybridization data (Fig. 1 A and B) are consistent with previous studies and show that, in particular, dinucleotides such as GA and CA are major components of conifer genomes.

The variability in distribution and abundance of simple sequence repeats has been used for taxonomic studies as well as to study the genetic similarity between species or breeding lines in agronomic crops. Within the gymnosperm genomes investigated in this report, the abundance and genomic organization of simple sequence repeats largely follows taxonomic groupings, but unequivocal identification of single species was not possible. This is in contrast to many fingerprint studies in angiosperm groups where chromosome numbers are less conserved and genome sizes are both smaller and more variable.

Compared to the (CA), hybridization pattern, where conserved repetitive fragments were absent, the hybridization patterns generated by (GA),, (GGAT),, (GATA), and the minisatellite-motif were more complex, with distinct fragments superimposed on a smear detectable in most species.

Species of the genus *Pinus* are taxonomically divided into sections and subsections based on morphology (e.g. two- and three-needle, and five-needle species). It is noteworthy that the two- and three-needle pines of the subsections *Australes, Oocarpae* and *Contortae* (all belonging to the section *Pinus*) showed a conservation of fragments and similar hybridization patterns (Figs. 1 B, D and Fig. 2A). *P. massoniana* and *P. resinosa* of the subsection *Sylvestris* differed in hybridization with micro- and m&satellites as well as telomere-like sequences, from the remaining two- and three-needle pines. Furthermore, *P. strobus,* a five-needle pine of the section *Strobus,* can be distinguished from two- and three-needle pines by a different minisatellite and telomere hybridization pattern, although it is not greatly different for the six tested simple sequence motifs. Since the two sections of the genus *Pinus, Pinus* and *Strobus,* had separated relatively early in conifer phylogeny, the differences of the simple sequence repeats between the two sections are not surprising. The diverged hybridization detected in *G. biloba* and the more distantly related conifers, *Pinus glauca, Pinus alba* and *T. distichum,* reflects the taxonomic position of these species. Nevertheless, both *Pinus abies* and *P. glauca* share some conserved minisatellite fragments.

Despite the ecological and economic importance of conifers, little is known 'about the structure and composition of their nuclear genomes. In *Z? elliottii var. elliottii,* genes encoding the 18S-5.8S-25S rRNA and the 5S rRNA are abundant sequences and occur on 10 of the 12 chromosome pairs (Brown et al. 1993, Doudrick et al. 1995). Other highly repetitive DNA sequences of pine genomes are retrotransposable elements. The Ty1-*copia*-like retrotransposon *TPE*1 is distributed over all chromosomes, but excluded from major rDNA sites, and form a large proportion of the *P. elliottii var. elliottii* genome (Kamm et al. 1996). *ZFG7* is an abundant sequence of

the Ty3-*gypsy* family found in *P. radiata* (Kossack et al. 1990).

There are only a few reports about tandem repeats in conifers, such as satellites with monomers of 150-180 bp or 300-360 bp (Brown et al. 1998), in contrast to angiosperm genomes, where many satellite repeats have been cloned and localized along chromosomes by FISH (Kamm et al. 1994; Schmidt and Heslop-Harrison 1998). Molecular cytological studies of *P. elliottii var. elliottii* revealed the presence of brightly fluorochrome-staining regions along the chromosomes. While some of them could be assigned to rRNA gene blocks (Doudrick et al. 1995), there are indications that others are unrelated repetitive DNA sequences.

Fluorescence in situ hybridization confirmed the substantial amplification of intercalary and pericentric telomere-like repeats in the *P. elliottii* var. *elliottii* pine genome, which was already evident from Southern-hybridization, and also confirmed their existence as a major genomic component (Figs. 2B and 3B–C). Taking into account that the restriction enzyme *RsaI* cuts frequently in plant DNA, it is very likely that the large genomic fragments on the Southern experiments (Fig. 2B) have an unusual base composition or are characterized by a low sequence complexity.

Fluorescence in situ hybridization resulted in signals similar to a chromosome-banding pattern that might serve for chromosome identification, perhaps in combination with rRNA fluorescence in situ hybridization or fluorochrome staining. The telomere-like motif is not restricted to the end of chromosomes in *P. elliottii var. elliottii,* and similar to many eukaryote genomes ranging from mammalian species (Moyzis et *al.* 1988) to *A. thaliana* (Richards et al. 1991) shows the intercalary and pericentric telomeric repeat as a repetitive genomic component. Compiled data about the chromosomal localization of the telomeric sequence repeats in 44 plant species (Fuchs et al. 1995) revealed intercalary telomeric sequences in addition to the terminal ones in nine of them, and in three species also pericentric positions (e.g. in *P. sylvestris).*

According to Biessmann and Mason (1992), interstitial telomeric repeats have developed either as random short sequence arrays which may have become extended by slippage replication, or by attachment via telomerase or recombination to extrachromosomal linear DNA fragments which may then integrate into the genome, or by chromosome rearrangements such as fusions or inversions. The morphology and number of chromosomes ($2n = 2x = 24$) are extremely conserved in the more than 100 extant *Pinus* species; hence, if fusions of chromosomes occurred in the genus *Pinus*, these events must be considered as ancient in *Pinus* evolution and unusual in comparison to other plants. If amplification of telomere-like sequences is not the result of chromosome fusions, the increase in copy number probably still dates back long ago in the phylogeny of the genus *Pinus*. Since the two sections of the genus, namely section *Pinus* and section *Strobus*, had become distinct taxa in the early Cretaceous period (136 million years ago), the telomeric sequence has been amplified only in the section *Pinus* after separation of the two sections.

Fuchs et al. (1995) have demonstrated that terminal telomeric signals are difficult to detect by fluorescence in situ hybridization in *Picea* species. Similarly, in the experiments presented here, terminal sites are only detectable by computer-enchanced image analyses. Although these data indicate that conifer telomeres are relative short, improved hybridization conditions at appropriate stringencies recently enabled the reliable detection of telomeres at the termini of slash pine chromosomes (M. Oard, personal communication).

Gene duplication and the formation of complex gene families have been widely cited as a cause of plant genomes being large, and Kinlaw and Neale (1997) suggested that levels of multiplication were greater in conifers than in other plant species. However, considering the relatively low proportion which coding sequences make up of a total plant genome, it is unlikely that the evolution of multigene families alone can explain the enormous size of pine nuclear genomes. It is clear that repeats of low complexity such as $(CA)_n$, $(GA)_n$, $(GGAT)_n$, $(GATA)_n$, as well as minisatellite- and telomere-like sequences, represent a large fraction of the conifer genomes used in this report.

The fingerprint approach reported here gives a multilocus survey about the abundance of microsatellite motifs within the genome. Polymorphisms of simple sequence repeat loci are widely used for plant genome mapping. However, the isolation and molecular characterization of DNA fragments containing simple sequence repeat stretches is an expensive and laborious process, in particular for large genomes. The results reported here provide valuable information for the genome mapping of *Pinus* species and show which simple sequence repeat motifs are abundant, and hence suitable for the preparation of enriched libraries, and which are present at low abundance and more likely to give single-copy markers.

Acknowledgements T. Schmidt acknowledges financial support by a grant from the USDA Forest Service (SRS 33-G-97-082). J.S. Heslop-Harrison acknowledges EU grant BIO4CT962125.

References

Akkaya MS, Bhagwat AA, Cregan PB (1992) Length polymorphisms of simple sequence repeat DNA in soybean. Genetics 132:1131–1139

Ali S, Müller CR, Epplen JT (1986) DNA fingerprinting by oligonucleotide probes specific for simple repeats. Hum Genet 74: 239-243

Becker J, Heun M (1995) Barley microsatellites: allele variation and mapping. Plant Mol Biol 27: 835-845

Beyermann B, Nürnberg P, Weihe A, Meixner M, Epplen JT, Börner T (1992) Fingerprinting plant genomes with' oligonucleotide probes specific for simple repetitive DNA sequences. Theor Appl Genet 83:691–694

Biessmann H, Mason JM (1992) Genetics and molecular biology of telomeres. Adv Genet 30: 185-249

Broun P, Tanksley SD (1993) Characterization of tomato DNA clones with sequence similarity to human minisatellites 33.6 and 33.15. Plant Mol Bio123: 231-242

Brown GR, Amarasinghe V, Kiss G, Carlson JE (1993) Preliminary karyotype and chromosomal localization of ribosomal DNA sites in white spruce using fluorescence in situ hybridization. Genome 36:310–316

Brown GR, Newton CH, Carlson JE (1998) Organization and distribution of a *Sau*3A tandem repeated DNA sequence in *Picea* (Pinaceae) species. Genome 41:560—565

Depeiges A, Goubely C, Lenoir A, Cocherel S, Picard G, Raynal M, Grellet F, Dlseny M (1995) Identification of the most-represented repeated motifs in *Arabidopsis thaliana* microsatellite loci. Theor Appl Genet 91: 160-168

Dhillon SS (1987) DNA in tree species. In: Bonga JM, Durzan DJ (eds.): Cell and tissue culture in forestry. Vol. 1. Marinus Nijhoff Publ. Dordrecht, Netherlands, pp 298-313

Doudrick RL, Heslop-Harrison JS, Nelson CD, Schmidt T, Nance WL, Schwarzacher T (1995) Karyotype of slash pine (*Pinus elliottii var. elliottii)* using patterns of fluorescence in situ hybridization and fluorochrome banding. J Hered 86: 289-296

Echt CS, May-Marquardt P (1997) Survey of microsatellite DNA in pine. Genome 40: 9-17

Fuchs J, Brandes A, Schubert I (1995) Telomere sequence localization and karyotype evolution in higher -plants. Plant Syst Evol 196:227–241

Kamm A, Schmidt T, Heslop-Harrison JS (1994) Molecular and physical organization of highly repetitive, undermethylated DNA from *Pennisetum glaucum*. Mol Gen Genet 244:420–425

Kamm A, Doudrick RL, Heslop-Harrison JS, Schmidt T (1996) The genomic and physical organization of *Ty*1-*copia*-like sequences as a component of large genomes in *Pinus elliottii var. elliottii* and other gymnosperma Proc Natl Acad Sci USA 93: 2708-2713

Khoshoo TN (1961) Chromosome numbers in gymnosperms. Silvae Genet 10: l-9

Kinlaw CS, Neale DB (1997) Complex gene families in pine genomes. Trends Plant Sci 2:356–359

Kossack DS, Barrios M, Kinlaw CS (1990) An abundant family of retrotransposons in pines. International Union of Forestry Research Organizations (IUFRG) Molecular Genetics Meeting, September 30-October 4, 1990, Lake Tahoe, Nevada (abstract)

Lagercrantz U, Ellegren H, Andersson L (1993) The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. Nucleic Acids Res 21: 111 l-l 115

Metais I, Aubry C, Hamon B, Peltier D, Jalouzot R (1998) Cloning, quantification and characterization of a minisatellite DNA sequence from common bean *Phaseolus vulgaris* L. Theor Appl Genet 97:232–237

Moyzis RK, Buckingham JM, Cram LS, Dani M, Deaven LL, Jones MD, Meyne J, Ratliff RL, Wu JR (1988) A highly conserved repetitive DNA sequence, $(TTAGGG)_n$, at the telomeres of human chromosomes. Proc Natl Acad Sci USA 85: 6622-6626

Murray BG (1998) Nuclear DNA amounts in gymnosperms. Ann of Bot (Suppl) 82:3–15

Poulsen GB, Kahl G, Weising K (1993) Abundance and polymorphism of simple repetitive DNA sequences in *Brassica napus* L. Theor Appl Genet 85: 994-1000

Richards ER, Ausubel FM (1988) Isolation of a higher eukaryotic telomere from *Arabidopsis thaliana*. Cell 53: 127-136

Richards ER, Goodman HM, Ausubel FM (1991) The centromere region of *Arabidopsis thaliana* chromosome 1 contains telomere-similar sequences. Nucleic Acids Res 19. 3351-3357

Richards ER, Vongs A, Walsh M, Yang J, Chao S (1993) Substructure of telomere repeat arrays. In: Heslop-Harrison JS, Flavell RD (eds). The chromosome. BIOS Scientific. Oxford, pp 103–114

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor. New York

Schmidt T, Heslop-Harrison JS (1998) Genomes, genes and junk: the large-scale organization of plant chromosomes. Trends Plant Sci 3:195–199

Schmidt T, Boblenz K, Metzlaff M, Kaemmer D, Weising K, Kahl G (1993) DNA-fingerprinting in sugar beet (*Beta vulgaris)* - identification of double-haploid breeding lines. Theor Appl Genet 85: 653-657

Smith DN, Devey ME (1994) Occurrence and inheritance of microsatellites in *Pinus radiata*. Genome 37: 977-983

Thein SL, Wallace RB (1986) The use of synthetic oligonucleotides as specific hybridization probes in the diagnosis of genetic disorders. In: Davies KE (ed) Human genetic diseases - a practical approach, IRL Press, Oxford, pp 33–50

Tourmente S, Deragon JM, Lafleuriel J, Tutois S, Pelissier T, Cuvillier C, Espagnol MC, Picard G (1994) Characterization of minisatellites in *Arabidopsis thaliana* with sequence similarity to the human minisatellite core sequence. Nucleic Acids Res 22: 3317-3321

Wagner DB, Furnier GR, Shaghai-Maroof MA, Williams SM, Dancik BP, Allard RW (1987) Chloroplast DNA polymorphism in lodgepole and jack pines and their hybrids. Proc Natl Acad Sci USA 84: 2097-2100

Wakamiya I, Newton RJ, Johnston JS, Price HJ (1993) Genome size and environmental factors in the genus *Pinus*. Am J Bot 80: 12361241

Weising K, Beyermann B, Ramser, J, Kahl G (1991) Plant DNA fingerprinting with radioactive and digoxigenated oligonucleotide probes complementary to simple sequence repetitive sequences. Electrophoresis 12:159–169

Weising K, Kaemmer D, Weigand F, Epplen JT, Kahl G (1992) Oligonucleotide fingerprinting reveals various probe-dependent levels of informativeness in chickpea (*Cicer arietinum* L.). Genome 35: 436-442

Winberg BC, Zhou Z, Dallas JF, McIntyre CL, Gustafson JP (1993) Characterization of minisatellite sequences from *Oryza sativa*. Genome 36: 978-983

Wu K-S, Tanksley SD (1993) Abundance, polymorphism and genetic mapping of microsatellites in rice. Mol Gen Genet 241: 225-235