Original article

Toward a *Pinus pinaster* bacterial artificial chromosome library

Rocío Bautista, David P. Villalobos, Sara Díaz-Moreno, Francisco R. Cantón,
Francisco M. Cánovas, M. Gonzalo Claros*

Departamento de Biología Molecular y Bioquímica, Facultad de Ciencias & Instituto Andaluz de Biotecnología,
Universidad de Málaga, 29071 Málaga, Spain

Abstract – Conifers are of great economic and ecological importance, but little is known concerning their genomic organization. This study is an attempt to obtain high-quality high-molecular-weight DNA from *Pinus pinaster* cotyledons and the construction of a pine BAC library. The preparation incorporates modifications like low centrifugation speeds, increase of EDTA concentration for plug maintenance, use of DNase inhibitors to reduce DNA degradation, use of polyvinylpyrrolidone and ascorbate to avoid secondary metabolites, and a brief electrophoresis of the plugs prior to their use. A total of 72 192 clones with an average insert size of 107 kb, which represents an equivalent of 11X pine haploid genomes, were obtained. The proportions of clones lacking inserts or containing chloroplast DNA are both approximately 1.6%. The library was screened with cDNA probes for seven genes, and two clones containing Fd-GOGAT sequences were found, one of them seemingly functional. Ongoing projects aimed at constructing a pine bacterial artificial chromosome library may benefit from the methods described here.

BAC / library / glutamate synthase / high molecular weight DNA / pine

Résumé – Construction d'une banque BAC pour le pin maritime (*Pinus pinaster*). Les conifères présentent un intérêt économique et écologique de premier plan mais restent très mal connus du point de vue de l'organisation de leur génome. Cette étude présente une tentative réussie de construction d'une banque BAC de séquences d'ADN de haute qualité et de poids moléculaire élevé à partir de cotylédons de *Pinus pinaster*. Le protocole de préparation se base sur des ajustements comme une baisse de la vitesse de centrifugation, une augmentation des concentrations d'EDTA dans les culots, l'utilisation d'inhibiteurs des ADNases pour limiter la dégradation de l'ADN, l'utilisation de polyvinylpyrrolidone et d'ascorbate pour éliminer les métabolites secondaires, et de brèves électrophorèses des culots. Un total de 72 192 clones a été obtenu, d'une dimension moyenne d'inserts de 107 kb et représentant l'équivalent de 11X du génome haploïde de pin. La proportion de clones dépourvus d'inserts ou contenant de l'ADN chloroplastique était de 1.6%. La banque a été testée avec des ADN complémentaires de 7 gènes, et deux clones contenant la séquence de la Fd-GOGAT ont été détectés. Des projets visant à construire une banque bactérienne artificielle (BAC) de chromosome de pin tireront bénéfice de l'utilisation de cette méthode.

banque BAC / glutamate synthase / brins d'ADN à fort poids moléculaire / pin

1. INTRODUCTION

Conifers are of great economic and ecological importance as they are widely used for reforestation, to obtain paper pulp, and as a primary source of wood for furniture. In particular, pines, which cover vast areas of the globe, are one of the most important genera of forest trees, dominate the ecology of many temperate and subtropical forest ecosystems, and provide a major fraction of the world's wood. They are also valuable as model organisms because they are the best-characterized gymnosperms and widely used for genomic mapping, genetic analysis and functional genomics, among other studies [6, 24]. Isolation, sequencing, and characterization of pine genes are necessary for genomic studies and are becoming available, although mapped based cloning is not still well developed in conifers, among other things, because large insert collections are not available due to their genome size and complexity.

Over the last few years, libraries with large genomic DNA inserts have become more and more important in the analysis of whole plant, animal, and fungal genomes. Bacterial artificial chromosomes (BACs) [37] are the most widely used cloning system in the construction of a genomic DNA library since their use presents several advantages relative to other genomic cloning systems [29, 38]. The construction of BAC libraries have provided desirable substrates for large-scale sequencing of large, complex genomes, they are essential resources for development of sequence-ready physical map of the target genome, to span long-range repetitive sequence regions, to confirm the entire sequence structure of the genome, to study the horizontal gene transfer, and to analyze clustered genes [36, 45]. However, BAC libraries are underdeveloped in pine since its genome, like other conifer genomes, is composed primarily of non-genic repetitive elements and is exceedingly large, making the characterization of gene-containing genome regions difficult. For example, the estimated size of the pine genome is 2.3×10^{10} nt distributed on 12 large chromosomes, that is, 48 pg/2C [23, 31, 35], which is even greater than the common wheat [20].

Studies of conifer genomics usually experience additional difficulties in obtaining high quality and large size genomic DNA: pine genome is about 40× larger than poplar and 160× larger than *Arabidopsis*. In grasses, legumes, vegetables and fruit trees like sweet orange, peach, walnut and willow, it is enough to break plant cell walls physically in order to isolate and embed nuclei in agarose plugs to easily manipulate DNA as in aqueous solutions without significant shearing or contamination [22, 28, 44]. By contrast, there have been contradictory results reported concerning high quality genomic DNA from conifers [19, 22]. Hence, in conifers, isolation of nuclei suitable for BAC library construction is a critical step because: (i) megabase DNA tends to shear more during tissue freezing and homogenization, (ii) plant cell walls make it much more difficult to obtain intact high-molecular-weight DNA (HMW-DNA), (iii) chloroplasts must be separated from the nuclei and selectively eliminated since the buoyant densities of nuclei and chloroplasts are similar, (iv) HMW-DNA is more sensitive to endogenous nucleases, and (v) HMW-DNA deteriorates rapidly when stored at 4 °C [27] due to waxes and polyphenols, which can mediate DNA damage by binding to DNA after cell lysis.

BAC libraries are laborious and costly, especially from organisms with large and complex genomes, and require additional efforts in organization [1]. HMW-DNA preparation from conifers should focus on avoiding the above-mentioned critical steps concerning the preparation of intact nuclei, the protection of DNA from breakage by DNase activity as well as shearing, and the avoidance of contamination with polysaccharides and phenolic compounds. It would be of interest to construct a *Pinus pinaster* BAC library since it is the most widely used conifer species for reforestation in Southwestern European countries, and it is expected that the breeding of improved varieties of *P. pinaster* will be accelerated by a better knowledge of the pine genome using such a BAC library. Moreover, a large collection of ESTs from *Pinus pinaster* is publicly available at NCBI. Therefore, here we report on how to circumvent these obstacles in order to obtain high-quality HMW-DNA, how to construct a pine BAC library, and how to conduct its screening in order to identify clones that contain genes.

2. MATERIALS AND METHODS

2.1. BAC vector preparation

The modified, high-copy BAC vector pCUGIBAC1 containing the pIndigoBAC536 was kindly provided by David Frisch (Clemson University Genomics Institute, USA) and used for library construction. Supercoiled pCUGIBAC1 DNA was isolated according to previously described methods [27, 39], digested with *Bam*HI, and dephosphorylated with 0.4 U of calf intestinal alkaline phosphatase (Roche) per μg of vector DNA. It was then self-ligated at room temperature for 2 h and fractionated on 0.8% low melting agarose gel in 1X TAE buffer (40 mM Tris-acetate, 1 mM EDTA). The linear dephosphorylated pIndigoBAC536 DNA (7.5 kb) was sliced from the gel and purified with Qiaquick kit (Qiagen) following the manufacturer's instructions. The final concentration of vector was adjusted to 10 ng/μL, aliquoted, and stored in 50% glycerol at –20 °C (for frequent use) or –80 °C (for long-term storage).

2.2. Pine nuclei isolation

Since pine needles seem to be difficult to extract their DNA with HMW quality [19, 22], *Pinus pinaster* seeds (supplied by INFOCA, Málaga, Spain) were used and germinated as described previously [7] in spite of their genotype heterogeneity. Nuclei were isolated from seedling cotyledons according to commonly used plant methods [26, 29, 44] modified as follows: 20 g of *P. pinaster* cotyledons were ground into fine powder in liquid nitrogen with mortar and pestle, and then immediately transferred into ice-cold homogenization buffer HB (10 mM Tris-HCl pH 9.4, 80 mM KCl, 10 mM EDTA, 4 mM spermidine, 1 mM spermine, 0.5 M sucrose) supplemented with 0.5% Triton X-100, 0.15% 2-mercaptoethanol, 2% PVP-40 and 0.1% ascorbic acid. The contents were mixed thoroughly on ice by magnetic stirring and filtered through two layers of cheesecloth and two layers of Miracloth (Calbiochem-Novabiochem, USA) by squeezing with gloved hands, and centrifuged at $1200 \times g$ for 20 min at 4 °C with a fixed-angle rotor. The supernatant was poured off and the nuclear pellet (pale green) was re-suspended in 30 mL of ice-cold HB with the assistance of a small paintbrush soaked in ice-cold HB buffer. The suspension was filtered through one layer of Miracloth and centrifuged again at $1000 \times g$ for 15 min at 4 °C in a swinging bucket rotor. The supernatant was decanted and the pellet was washed at least twice in ice-cold HB buffer. Finally, the nuclear pellet was suspended in 500 μL SCE buffer (1 M sorbitol, 0.1 M sodium citrate, 60 mM EDTA, pH 7.0).

2.3. Nuclei embedding in agarose plugs and DNA digestion

Nuclei were pre-warmed at 45 °C for nearly 5 min and then mixed with an equal volume of 1.6% low melting point agarose (Seaplaque, FMC Bioproducts, USA) in SCE, kept at 45 °C, using a micropipette. The mixture was then aliquoted into ice-cold plug moulds with the same micropipette. Solidified plugs were stored in TE at 4 °C until use. The plugs were transferred to 10 vol. of lysis buffer (0.5 M EDTA pH 9.0-9.3, 1% sodium laurylsarcosine, 2% PVP-40, 0.1% ascorbic acid, and 0.1 mg/mL proteinase K) and incubated for 36 h at 50 °C without shaking, changing buffer and enzyme every 12 h. Plugs were washed with 10 vol. of 0.5 M EDTA pH 9 for 1 h at 50 °C and then washed with ice-cold 0.5 M EDTA pH 8 for 1 h until the plugs became transparent. The plugs could be stored on 0.5 M EDTA pH 8 for more than 3 months.

Just prior to use, the plugs were electrophoresed to remove broken, chloroplast and mitochondrial DNA in a 0.7% agarose gel using a FIGE (Field Inversion Gel Electrophoresis system; FIGE Mapper, BioRad, USA). The migration was performed in 0.33 X TBE at 4 V/cm and 11 °C for 12 h with an initial pulse time of 2.5 s and a final pulse time of 5.5 s. After electrophoresis, plugs were removed from wells and washed twice for 30 min in 10 vol. of ice-cold TE. DNA concentration was determined by melting a plug and measuring the absorbance at 260 nm.

One half-plug (a volume of nearly 40 μL) was transferred into a sterile microcentrifuge tube and equilibrated with 1 mL of incubation buffer (restriction enzyme buffer plus 4 mM spermidine, 1 mM DTT,

2.5 mM Lys-HCl and 60 μM EGTA) for 1 h at 4 °C. Then, the mixture was replaced with 170 μL of digestion buffer (restriction enzyme buffer plus 0.75 U *Bam*HI, 4 mM spermidine, 1 mM DTT, 0.5 mg/mL BSA, 2.5 mM Lys-HCl and 60 μM EGTA), and incubated for 1 h in ice and then for 3 min at 37 °C. The reaction was stopped with 20 μL of 0.5 M EDTA.

2.4. HMW-DNA electro-elution

Ten half-plugs containing the digested HMW-DNA were electrophoresed in a FIGE at 11 °C with a 0.7% agarose gel in 0.33 X TBE. The FIGE was performed to resolve DNA fragments between 25 and 200 kb using 4 V/cm for 20 h, setting the initial pulse time of 2.5 s and final pulse of 5.5 s. Two molecular weight markers were used: one was the lambda DNA concatemer ladder (BioRad, USA) and the other was the *E. coli* chromosomal DNA digested with *Xba*I prepared following the FIGE Mapper procedure. The region between 100 and 240 kb was excised from the gel and electroeluted in a thin dialysis bag (Sigma, USA) filled with approximately 300 μL of 0.33 X TBE buffer. Electro-elution was performed for 2.5 h in the FIGE system in the same conditions as above. Buffer containing the DNA fragments was taken from the bag with a cut tip. The electro-elution yields a low concentration DNA that was drop-dialysed against 0.5 X TE with 10% PEG for 90 min at room temperature on 0.025 μm VSWP type MF- Millipore Membranes.

2.5. BAC library construction

About 40 ng of the dephosphorylated pIndigoBAC536 DNA were ligated overnight to 400 ng of pine HMW-DNA in a total volume of 150 μL with 400 U of T4 DNA ligase (New England Biolabs) in 66 mM Tris-HCl pH 7.6, 6.6 mM MgCl$_2$, 10 mM DTT, 0.1 mM ATP at 16 °C. The ligation mixture was then desalted and concentrated against TE/PEG. Two microliters of the concentrated ligation reaction were used to electroporate 30 μL of *E. coli* ElectroMax DH10B cells (Gibco-BRL, USA) using the Eppendorf cell-porator system and the following settings for a 1 mm cuvette (BioRad, USA): 1800 V, 10 μF capacitance, and a maximum resistance of 600 W. After transformation, cells were suspended in 1 mL SOC medium (2% Bacto tryptone, 0.5% Bacto yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl$_2$, 10 mM MgSO$_4$, 20 mM glucose, pH 7.0) and incubated for 1 h at 37 °C with shaking at 100 rpm. Cells were spun down, suspended in 200 μL of SOC medium and plated in LB plates (1% Bacto tryptone, 0.5% Bacto yeast extract, 1% NaCl, 1.5% agar, pH 7.5) containing 12.5 mg/mL chloramphenicol, 40 μg/mL X-gal and 100 μg/mL IPTG. After incubating for 36–48 h at 37 °C, single white colonies were manually picked and stored in 384-well microtitre plates (Nunc, Denmark) as follows: individual wells containing 75 μL of LB freezing buffer (LB broth supplemented with 36 mM K$_2$HPO$_4$, 13.2 mN KH$_2$PO$_4$, 1.7 mM sodium citrate, 0.4 mM MgSO$_4$, 6.8 mM (NH$_4$)SO$_4$, 4.4% v/v glycerol) with 12.5 mg/mL chloramphenicol were inoculated with a single colony. The plates were incubated overnight at 37 °C, and stored at –80 °C.

The average insert size of the BAC library was assessed using 128 randomly selected clones and isolating their DNA with a modified alkaline lysis protocol [40]. Half a microgram of BAC DNA were digested overnight with 3 U of *Not*I to release the insert from the BAC vector. The insert size was estimated via 1% agarose gel by using FIGE conditions of 4 V/cm for 16 h at 11 °C with an initial pulse time of 0.1 s and a final pulse time of 3.5 s.

Table I. Primers used to amplify three pine chloroplast genes, and the length of each amplified fragment.

Gene	Primers	Amplicon size (kb)
RpoC2	5'-GCACTATCCAAGGTATTTTCGTAA-3' 5'-GTCACCTGATCTATGTTCCACTTG-3	0.92
PsaA	5'-CCGGAACCAGAAGTAAAGAAAGTA-3' 5'-TATGGCCCTCCCCCGTAAAT-3'	1.00
ChiN	5'-GATTTTTGCAGAACCCCGCTATG-3' 5'-CGTATCTGATTCGAATTGTCTGGT-3'	0.98

2.6. BAC library screening

BAC clones were spotted on nylon membranes using a 384-pin replicator system (Nunc). Each 7.5 × 11.5 cm membrane was prepared to contain clones from four different microtitre plates (1536 clones per membrane). DNA was fixed to membranes by UV cross-linking.

Screening for chloroplast DNA in the library was carried out using three pine chloroplast PCR-generated gene-specific probes for genes *RpoC2*, *PsaA* and *ChiN* (Tab. I), which are spaced almost equally around the 119.7 kb pine chloroplast genome [41]. Each gene fragment was amplified by PCR using 50 ng *Pinus pinaster* total DNA isolated as previously reported [7]. The samples for PCR amplification were denatured for 1 min at 94 °C and subjected to 30 cycles of 1 min denaturation at 94 °C, 1 min annealing at 50–65 °C, and 2 min extension at 72 °C, followed by a final extension of 10 min at 72 °C. *EcoTaq* 0.025 U/μL (EcoGen, Spain) was used. The three amplicons were cloned into pBluescript and their presence was confirmed by sequencing.

Eight probes derived from 7 nuclear genes (Tab. II) were used to study the suitability of the partial BAC library. Hybridization was carried out at 65 °C overnight using standard techniques [34].

3. RESULTS

3.1. Protocol modifications to preserve the integrity of pine nuclei and their DNA

Commonly used protocols to obtain suitable nuclei [26, 29, 44] have to be optimized to make them suitable for pine. Better yields were obtained using brief centrifugation times (15 min) at speeds of 1200 or 1000 × *g*. The number or layers of Miracloth was increased from 1 to 2 to compensate for the decrease in centrifugation speed. DNA does not become oxidized (brown pellet) when PVP-40 and ascorbic acid were added to the homogenization buffer. Nuclei were finally resuspended in SCE buffer instead of HB since the former contains much more EDTA (60 mM vs. 10 mM) in order to inhibit pine Mg^{2+}-dependent DNases that were not removed neither using previous methods (lanes Z and P in Fig. 1) nor using our optimized one (lanes B in Fig. 1). Nuclei were embedded in agarose and the resulting plugs were migrated briefly to remove sheared DNA (lanes 'TE' in Fig. 1). The final protocol yields 4.8 μg of unoxidized unbroken DNA per gram of fresh weight.

For enzymatic manipulation, EDTA was replaced with other inhibitors that do not affect restriction endonucleases.

Table II. cDNA probes used to screen the pine BAC library.

| Gene	Size (bp)	Species	References
GS 1a	1400	*Pinus sylvestris*	[7, 8]
GS 1b	800	*Pinus sylvestris*	[3, 4]
Fd-GOGAT, 3'-end	1300	*Pinus sylvestris*	[13, 14]
Fd-GOGAT, 5'-end	2000	*Hordeum vulgare*	[2]
PII	1100	*Pinus pinaster*	C. Avila, unpublished
Nitrite reductase	800	*Pinus sylvestris*	C. Avila, unpublished
Chloroplast dicarboxylic translocator DIT1	2000	*Spinacea oleacea*	[42]
Chloroplast dicarboxylic translocator DIT2	2000	*Spinacea oleacea*	[42]
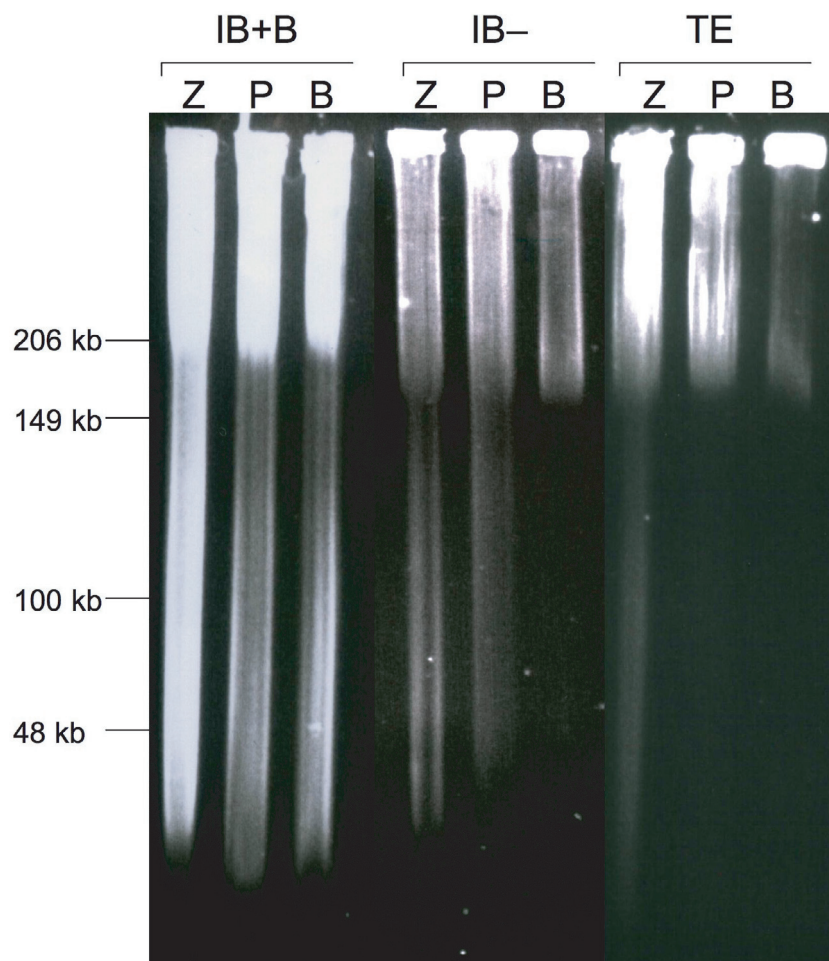
Figure 1. Comparison of pine nuclei DNA obtained with three different protocols: Z according to Zhang et al. (1995), P according to Peterson et al. (2000), and B according to this work. Agarose plugs were incubated for 2 h at 37 °C with TE (TE), incubation buffer without restriction enzyme (IB–), or incubation buffer with *Bam*HI (IB+B) and then migrated in a FIGE. The sizes of the molecular weight markers are shown on the left.

The combination of 160 mM L-lysine and 4 mM EGTA has been shown to inhibit unwanted DNases while preserving the activity of common restriction enzymes [26]. However, in our hands, this combination inhibited all nuclease activity, including restriction endonucleases. Hence, incubation buffers were then tested with decreasing concentrations of L-lysine/EGTA of 80 mM/2 mM, 40 mM/1 mM, 20 mM/0.5 mM, 10 mM/0.25 mM, 5 mM/0.12 mM,

2.5 mM/60 μM and 1.25 mM/30 μM. Each agarose plug was then separated into two halves: one half-plug was not treated, and the other half-plug was digested with *Bam*HI for 2 h. Afterwards, they were subjected to a FIGE. 2.5 mM L-lysine and 60 μM EGTA were sufficient to reduce unspecific HMW-DNA degradation while permitting digestion with restriction endonucleases. Plugs can be stored in 0.5 M EDTA up to 5 years. Additionally, sheared DNA is nearly absent and most
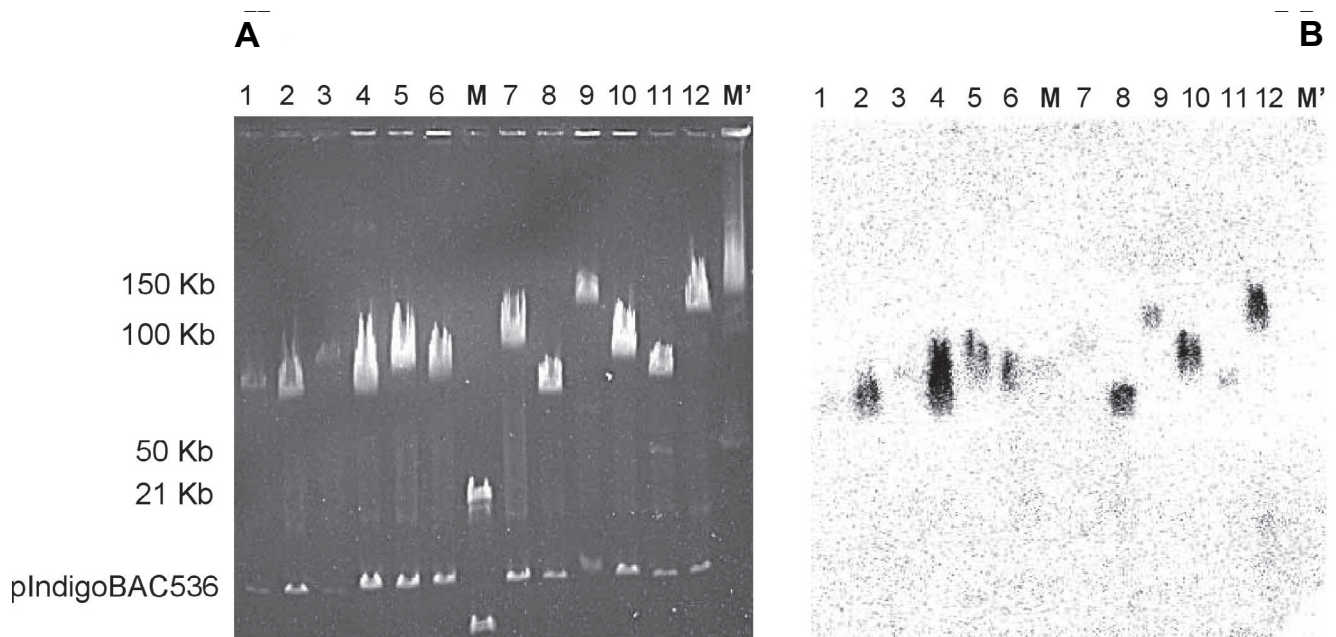
Figure 2. Analysis of 12 out of the 128 randomly selected *P. pinaster* BAC clones. **A**. Ethidium bromide-stained FIGE showing inserted DNA larger than the common 7.5 kb pIndigoBAC536 vector band. **B**. Autoradiograph of the gel in A after transfer and hybridization with total pine genomic DNA. Lane numbers correspond to different BAC clones. The sizes of the different molecular weight markers (M and M') are shown on the left.

nucleases were removed (compare lane B with respect to lanes Z and P in blocks 'TE' and 'IB–' at Fig. 1).

3.2. Partial digestion of HMW-DNA and FIGE

Special care was devoted to setting up the HMW-DNA partial digestion because restriction reactions were carried out in the presence of L-lysine and EGTA, whose presence alters the activity of restriction endonucleases. The half-plugs were serially digested in 170 μL of reaction buffer for 30, 15, 5 and 3 min with 2, 0.75 and 0.1 U of restriction enzyme. The three single-cut restriction sites in the BAC vector (*Hin*dIII, *Eco*RI and *Bam*HI) were tested and 0.75 U *Bam*HI for 3 min results in a maximal amount of DNA digested in the range of 100 to 200 kb. The optimal conditions for *Hin*dIII and *Eco*RI were very similar but not identical to *Bam*HI, perhaps due to the presence of L-lysine and EGTA in the digestion mixture [26].

Partial digestion of 64 μg of DNA in 10 agarose plugs yielded 4 μg of DNA ranging from 100 to 200 kb. Eluted DNA was not subjected to a second round of size selection [25] since plugs were previously electrophoresed to remove any other DNA. Our results suggest that HMW-DNA can be efficiently digested using this protocol.

3.3. BAC library construction and characterization

Optimal dephosphorylation and ligation conditions were determined with lambda DNA and produced 90% of white colonies. Partially digested pine DNA was ligated into

pIndigoBAC536 vector and transformed by electroporation into *E. coli* DH10B (efficiency, 3.75×10^6 colonies per μg of DNA). Twelve BAC clones were digested with *Not*I and hybridized with radiolabeled total genomic DNA. Since *Not*I is a GC-8-base cutter, the digestion typically generates two fragments, one of 7.5 kb corresponding to the BAC, and a larger one corresponding to the insert (Fig. 2A). These fragments were hybridized with labeled genomic DNA in a Southern blot (Fig. 2B), indicating that the insert source was pine. The strongly hybridizing lanes should correspond to highly repetitive genomic DNA while the weak ones should probably correspond to low-copy DNA, since clones with the same DNA amount on the gel (lanes 4, 5 and 7, for example) provide different signal intensity. The same filter was also probed with total genomic DNA from *E. coli* and only the pIndigoBAC vector hybridized to this probe, confirming that no bacterial DNA was cloned.

The partial BAC library that we produced consisted of 72 192 clones, whose average insert size was determined by digesting 128 randomly selected clones with *Not*I. The clones were grouped by insert size, and the insert size of each clone was plotted against the frequency of each group of clones represented in the library (Fig. 3). Based on this analysis, 1.6% of the clones did not contain inserts. Most of the library was between 76 and 125 kb: 10% of the clones had an insert larger than 150 kb, while 25% were smaller than 80 kb. Therefore, the average insert size of the library was 107 kb. This average size together with the 72 192 clones makes the representation to be 0.11X. The proportion of clones in the library containing organelle DNA was estimated as follows: 4 high-density filters containing 6 144 (corresponding to 8.5% of the current
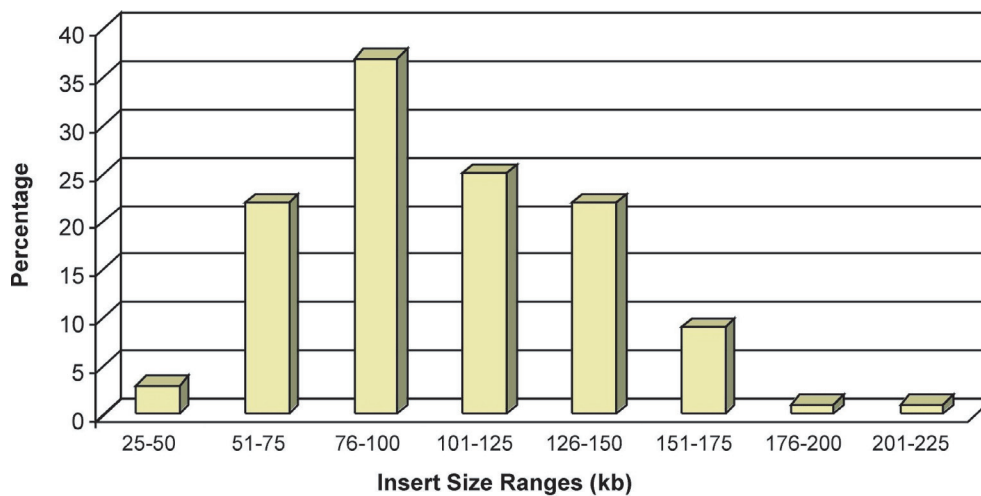
Figure 3. Insert size distribution of the BAC library. The insert size distribution of 128 BAC clones was analyzed as shown in Figure 2A. The average insert size is 107 kb.

library) were screened with the chloroplast-specific probes *rpo*C2, *Chl*N and *psa*A (see Materials and methods). Ninety-eight clones hybridized with at least one probe, corresponding to 1.59% of the BAC clones containing inserts of chloroplast DNA.

Taking into account that the pine genome contains 2.3×10^{10} bp, and that the average insert size of our library is 107 kb, a pine BAC library should contain at least 643 941 clones in order to represent 1 X copy of the pine genome with a probability of 95%. The 1.6% of clones without insert and the 1.59% of clones with chloroplast DNA would increase this number to 664 647 clones.

3.4. Suitability of the partial BAC library

Since we only have a 0.11X library of a genome that is mostly composed of non-coding DNA [35], the library coverage was tested by screening 47 filters that contain the 72 192 clones using 8 probes for 7 genes, which are among the best characterized genes in gymnosperms (Tab. II). Two GOGAT-positive clones, 16A9 and 176P12, were obtained (Fig. 4). They were digested with *Bam*HI and *Hin*dIII and hybridized with the pine GOGAT probe that recognizes the 3'-end of the gene (Fig. 4, right), and the barley GOGAT probe that recognizes the 5'-end of the gene (Fig. 4, left). As expected, hybridization with the homologous pine 3'-end probe resulted in more intense signals than with the heterologous barley 5'-end probe. The hybridizing fragments were cloned into pBluescript and sequenced; the primary BAC clones 16A9 and 176P12 were also end-sequenced. The resulting sequences contained highly homologous sequences to Fd-GOGAT cD-NAs from other plant species (Fig. 5).

Clone 16A9 aligned with parts of the Fd-GOGAT cDNA of *Arabidopsis thaliana* (AC# AL391716), maize (AC# M59190.1) and rice (AC# AP003833) indicating the occurrence of internal deletions in the coding sequence of 16A9

while no putative intron DNA was found (Fig. 5A, left); the three regions in the dotplot that display similarity between 16A9 and AL391716 are 78% identical in average. In relation with *P. sylvestris* Fd-GOGAT cDNA [13] and clone 16A6 (Fig. 5A, right), when we compared the theoretical overlap between both sequences (604 nt), only a stretch of 380 nt displayed a high percentage of identity (95%) while the rest (224 nt) is strikingly divergent. Moreover, no open reading frame can be inferred from the 16A9 sequence. Taken together, these data suggest that clone 16A9 is a non-functional Fd-GOGAT sequence, perhaps a pseudogene.

With regard to clone 176P12, an intron-exon series can be predicted from its sequence analysis and alignments (data not shown). Furthermore, the exon sequences seem to be nearly identical to *P. sylvestris* Fd-GOGAT cDNA [13]. The deduced amino acid sequence from the exons is very similar to the protein sequences of rice and *Arabidopsis* Fd-GOGAT, particularly concerning the conserved locations for their splicing sites (Fig. 5B). In conclusion, unlike 16A9, 176P12 seems to contain a functional Fd-GOGAT gene.

4. DISCUSSION

The first attempt to construct a BAC DNA library from *P. pinaster* is described. This protocol will considerably advance genomic studies in pine since, although incomplete, the partial BAC library has led to the isolation of a clone containing a gene of interest. In consequence, it could be used to assist pine genome mapping and sequencing projects in the near future.

Several aspects of previously described protocols to prepare HMW-DNA were optimized to (i) preserve the nuclei integrity, (ii) avoid DNA shearing, (iii) manage the presence of secondary metabolites (mainly polyphenols), and (iv) inhibit non specific DNAse activities. This changes can be summarized as follows: as pine organs are waxy and rich in polyphenolic substances, and pine cell walls are difficult, expensive
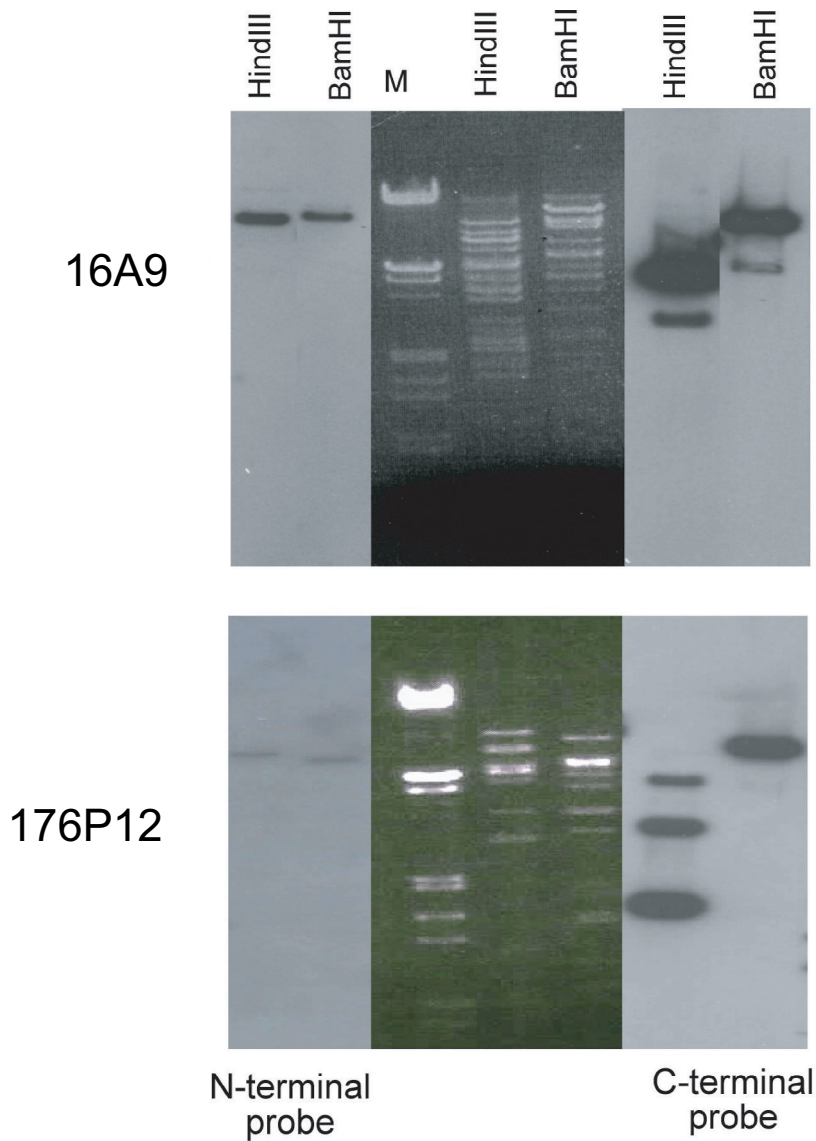
Figure 4. Hybridization of clones 16A9 and 176P12 with GOGAT probes. Left: Autoradiograph obtained hybridizing with the heterologous 5'-end probe. Centre: Ethidium bromide-stained FIGE gel of each clone digested with *Bam*HI and *Hin*dIII. Right: Autoradiograph obtained hybridizing with the pine 3'-end probe.

and time-consuming to remove [15], the starting material were pine cotyledons and the preferred homogenization method was grinding in liquid nitrogen. Peterson et al (2000) described that lower centrifugation speeds ($600 \times g$) increased nuclei integrity due to the high DNA content (48 pg) and large size of pine nuclei [23, 35]. But in our hands, yield is so compromised that we have to make use of brief centrifugation times (15 min) at speeds of 1200 or $1000 \times g$. The method of Zhang et al. (2000) provides brownish plugs due to oxidation; therefore, PVP-40 and ascorbic acid avoid the oxidation of DNA since they adsorb polyphenols [10]. The EDTA concentration in SCE or HB buffers was not described to be important [44] but, in our procedure, 60 mM EDTA has been reported here to be critical in order to stabilize HMW-DNA and provide DNA in agarose plugs stable for up to 5 years. Nuclease inhibitors such as L-lysine and EGTA [26] have to be used, although in a lower concentration (2.5 mM L-Lys and 60 μM EGTA). Therefore, the method described here (Fig. 1) provides nuclei

that can be stored for a long time and provides long DNA fragments (average size of 107 kb) in contrast to previous studies that reported the extraction of extremely short fragments from conifers [22].

Addition of nuclease inhibitors at the beginning of the protocol to isolate DNA bands from 300 to 500 kb [29] has been reported; however, this is not applicable when using EGTA because it removes the Ca^{2+} that is necessary for the activity of proteinase K, resulting in the presence of active nucleases after the treatment [5]. Consequently L-Lys and EGTA were added only after proteinase K treatment (see Materials and Methods).

Although DNA fragments used for ligation were selected in the range of 100 to 240 kb, the 107 kb final average insert size was lower (Fig. 3). Similar findings were observed in the construction of other BAC libraries [43] and it was hypothesized that smaller fragments could comigrate in agarose gels, which results in decreasing the average insert size [11].

A
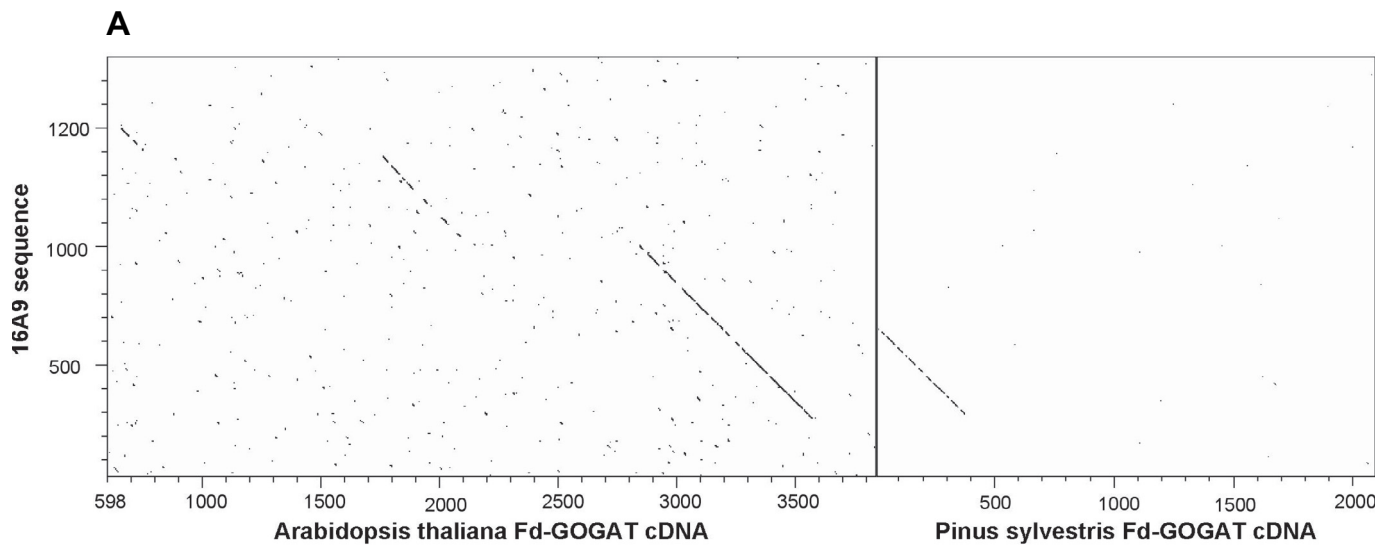
B

```
              1                                                         5
AP003833     GEDPIRWSPL ADVEDGYSPT LPHLKGLQNG DTATSAIKQV ASGRFGVTPT
AL391716     GEDPIRWKPL TDVVDGYSPT LPHLKGLQNG DIATSAIKQV ASGRFGVTPT
176P12        DPVRWRPL SDVVDGYSST FPHLKGLRNG DTATSAIKQV ASGRFGVTPT

              51                                                        100
AP003833     FLVNAEQIEI KIAQGAKPGE GGQLPGKKVS AYIARLRNSK PGVPLISPPP
AL391716     FLVNADQLEI KVAQGAKPGE GGQLPGKKVS AYIARLRSSK PGVPLISPPP
176P12       FLVNADQLEI KIAQGAKPGE GGQLPGKKVS TYIARLRNSK PGVPLISPPP

              101                                                       150
AP003833     HHDIYSIEDL AQLIYDLHQI NPKAKVSVKL VAEAGIGTVA SGVSKGNADI
AL391716     HHDIYSIEDL AQLIFDLHQI NPNAKVSVKL VAEAGIGTVA SGVAKGNADI
176P12       HHDIYSIEDL AQLIFDLHQV NPNAKVSVKL VAEAGIGTVA SGVAKGNADI

              151                                                       200
AP003833     IQISGHDGGT GASPISSIKH AGGPWELGLS ETHQTLIQNG LRERVVLRVD
AL391716     IQISGHDGGT GASPISSIKH AGGPWELGLT ETHQTLIANG LRERVILRVD
176P12       IQISGHDGGT GASPISSIKH AGGPWELGLT ETHQTLIENG LRERVVLRVD

              201                                                       250
AP003833     GGFRSGLDVL MAAAMGADEY GFGSVAMIAT GCVMARICHT NNCPVGVASQ
AL391716     GGLKSGVDVL MAAAMGADEY GFGSLAMIAT GCVMARICHT NNCPVGVASQ
176P12       GGFKSGVDVM LAAAMGADEY GFGSVAMIAT GCIMARICHT NNCPVGVASQ

              251                                                       300
AP003833     REELRARFPG VPGDLVNYFL FVAEEVRATL AQLGFEKLDD IIGRTDILKA
AL391716     REELRARFPG VPGDLVNYFL YVAEEVRGIL AQLGYNSLDD IIGRTELLRP
176P12       REELRARFPG VPGDLVNFFL YVAEEVRGIL AQLGYAKLDD IIGRTDLLRP

              301                                                       351
AP003833     KHVSLAKTQH IDLKYLLSSA GLPKWSSSQI RSQDVHSNGP VLDETILADPD
AL391716     RDISLVKTQH LDLSYLLSSV GTPSLSSTEI RKQEVHTNGP VLDDDILADPL
176P12       RHISLMKTQQ LDLSYLLSNV GRPKWSSTKI RQQNVHTNGP VLDDILLAD
```

Figure 5. Sequence analysis of Fd-GOGAT sequences isolated from the BAC library. **A**. 16A9 sequence is compared in a dotplot with Fd-GOGAT cDNAs of *Arabidopsis* (left) using a window of 15 amino acids and a stringency of 11, and with *Pinus sylvestris* (right) using a window and stringency of 9 amino acids. **B**. Alignment of the amino acid sequence deduced from 8 exons of *Arabidopsis* (AC# AL391716), rice (AC# AP003833) and *Pinus pinaster* (176P12). The first amino acid coded by an exon is marked in bold-italic, while the last amino acid coded by an exon is marked in bold.

Our method overcomes the drawbacks of ligation inhibition by sheared DNA and electroporation efficiency by (i) the treatment with laurylsarcosine which breaks nuclear and chloroplast membranes, and (ii) the brief electrophoresis of the plugs prior to their use [26] removed sheared DNA, chloroplast DNA, and any mitochondrial DNA (around 120 kb [41]). The ligation efficiency seems to be improved with respect to previous literature, where the proportion of false positive ranges from 3% to 17% [43], while the proportion was 1.6% in our case.

Vector preparation also plays a major role in the success of BAC library construction [9]. Hence, mild dephosphorylation was performed to avoid vector damage. After ligation, the remaining linearized form was purified from an agarose gel. The transformation efficiency after ligation of the partially digested HMW-DNA suggests that negligible contamination of sheared DNA is present, indicating that the protocol yields DNA suitable for BAC cloning. The high transformation efficiency could have the following, mutually non-exclusive explanations: (i) the use of 0.33 X TBE results in less contamination with borate, which increases the ligation efficiency [9], (ii) the mild dephosphorylation, (iii) the purification of the digested unreligated pIndigoBAC vector prior to use, and (iv) the removal of sheared DNA from the plugs with one round of FIGE, which also decreased the possible contamination with organellar DNA. Hence, the method described here seems to provide an affordable approach to megabase DNA preparation from conifers and facilitates long-range physical mapping and large DNA cloning in conifer genome research programs. Moreover, the reproducibility of our protocol makes it suitable for other woody plants like the olive tree (V. Sánchez-Vera, M.G. Claros, unpublished results). Construction of BAC libraries with 'gene-enriched' DNA has been suggested using differential metilation [32] or reassociation curves [30]. These approaches were discarded since they will produce BAC libraries only useful for gene cloning, as the construction of a sublibrary enriched in a gene of interest [12]. Isolation of a particular chromosome to construct a BAC library [20] seems not very useful in pine where the 12 chromosomes are very big and are close in size [17]. In fact, when we attempted to separate them by flow cytometry, they were resolved only in two main peaks (R. Bautista and M.G. Claros, unpublished results). However, it can be of future interest to maintain a pooled, nongridded BAC library that would be easier to organize and faster to screen [18].

The hybridizations performed with genomic DNA suggested that the BAC library contains a significant proportion of nonrepetitive DNA (Fig. 2), enough to isolate genomic clones containing genes (Figs. 4 and 5), or that pine gene sequences present in multiple copies can facilitate their cloning [21]. The preliminary characterization of the 176P12 clone indicates that it seems to contain a Fd-GOGAT gene (Fig. 5) while 16A9 is possibly a non-functional GOGAT sequence inserted in the genome that can also be found in both *P. pinaster* and *P. sylvestris* genomes [40]. The 16A9 clone may be an usual event [16, 17, 21] since, in pine, the number of gene-like sequences is extremely large in comparison to other plants [33]. The preliminary analysis of the 179P12 clone indicates that it contains a functional copy of the Fd-GOGAT gene with intron-exon structure similar to other orthologous genes. In conclusion, the 72 192 clones obtained with this method were of sufficient quality to identify gene sequences. The library and the method represent a first step toward a future physical mapping of the pine genome. Moreover, both the library and the method can be used in functional genomics with the ultimate goal of identifying and cloning specific genes but not for sequencing strategies as far as nuclei were isolated from non-clonal material.

REFERENCES

[1] Allouis S., Moore G., Bellec A., Sharp R., Faivre Rampant P., Mortimer K., Pateyron S., Foote T.N., Griffiths S., Caboche M., Chalhoub B., Construction and characterisation of a hexaploid wheat (*Triticum aestivum* L.) BAC library from the reference germplasm 'Chinese Spring', Cereal Res. Commun. 31 (2003) 331–338.

[2] Avila C., Marquez A.J., Pajuelo P., Cannell M.E., Wallsgrove R.M., Forde B.G., Cloning and sequence analysis of a cDNA for barley ferredoxin-dependent glutamate synthase and molecular analysis of photorespiratory mutants deficient in the enzyme, Planta 189 (1993) 475–483.

[3] Avila C., Muñoz-Chápuli R., Plomion C., Frigerio J.M., Cánovas F.M., Two genes encoding distinct cytosolic glutamine synthetases are closely linked in the pine genome, FEBS Lett. 477 (2000) 237–243.

[4] Avila C., Suarez M.F., Gomez-Maldonado J., Canovas F.M., Spatial and temporal expression of two cytosolic glutamine synthetase genes in Scots pine: functional implications on nitrogen metabolism during early stages of conifer development, Plant J. 25 (2001) 93–102.

[5] Bajorath J., Raghunathan S., Hinrichs W., Saenger W., Long-range structural changes in proteinase K triggered by calcium ion removal, Nature 337 (1989) 481–484.

[6] Boerjan W., Biotechnology and the domestication of forest trees, Curr. Opin. Biotechnol. 16 (2005) 159–166.

[7] Cantón F.R., García-Gutierrez A., Gallardo F., Vicente A. de, Cánovas F.M., Molecular characterization of a cDNA clone encoding glutamine synthetase from a gymnosperm, *Pinus sylvestris*, Plant Mol. Biol. 22 (1993) 819–828.

[8] Canton F.R., Suarez M.F., Jose-Estanyol M., Canovas F.M., Expression analysis of a cytosolic glutamine synthetase gene in cotyledons of Scots pine seedlings: developmental, light regulation and spatial distribution of specific transcripts, Plant Mol. Biol. 40 (1999) 623–634.

[9] Chalhoub B., Caboche M., Efficient cloning of plant genomes into bacterial artificial chromosome (BAC) libraries with larger and more uniform insert size. Plant Biotechnol. J. 2 (2004) 181–188.

[10] Claros M.G., Cánovas F.M., Rapid high quality RNA preparation from pine seedlings, Plant Mol. Biol. Rep. 16 (1998) 9–18.

[11] Frijters A.C.J., Zhang Z., vanDamme M., Wang G.L., Ronald P.C., Michelmore R.W., Construction of a bacterial artificial chromosome library containing large EcoRI and HindIII genomic fragments of lettuce, Theor. Appl. Genet. 94 (1997) 390–399.

[12] Fu H.H., Dooner H.K., A gene-enriched BAC library for cloning large allele-specific fragments from maize: Isolation of a 240-kb contig of the bronze region, Genome Res. 10 (2000) 866–873.

[13] García-Gutiérrez Á., Cantón F.R., Gallardo F., Sánchez-Jiménez F., Cánovas F.M., Expression of ferredoxin-dependent glutamate synthase in dark-grown pine seedlings, Plant Mol. Biol. 27 (1995) 115–128.

[14] García-Gutiérrez A., Dubois F., Canton F.R., Gallardo F., Sangwan R.S., Canovas F.M., Two different modes of early development and nitrogen assimilation in gymnosperm seedlings, Plant J. 13 (1998) 187–199.

[15] Gómez-Maldonado J., Crespillo R., Avila C., Cánovas F.M., Efficient preparation of maritime pine (*Pinus pinaster*) protoplasts suitable for transgene expression analysis, Plant Mol. Biol. Rep. 19 (2001) 361–366.

[16] Hipkins V.D., Marshall K.A., Neale D.B., Rottmann W.H., Strauss S.H., A mutation hotspot in the chloroplast genome of a conifer (Douglas-fir: *Pseudotsuga*) is caused by variability in the number of direct repeats derived from a partially duplicated tRNA gene, Curr. Genet. 27 (1995) 572–579.

[17] Hizume M., Shibata F., Matsusaki Y., Garajova Z., Chromosome identification and comparative karyotypic analyses of four *Pinus* species, Theor. Appl. Genet. 105 (2002) 491–497.

[18] Isidore E., Scherrer B., Bellec A., Budin K., Faivre-Rampant P., Waugh R., Keller B., Caboche M., Feuillet C., Chalhoub B., Direct targeting and rapid isolation of BAC clones spanning a defined chromosome region, Funct. Integr. Genomics 5 (2005) 97–103.

[19] Islam-Faridi N., Chang Y.-L., Zhang H., Kinlaw C., Doudrick R.L., Neale D.B., Echt C., Price H.J., Stelly D.M., in: Plant & Animal Genome VI Conference, San Diego, USA, 1998.

[20] Janda J., Safar J., Kubalakova M., Bartos J., Kovarova P., Suchankova P., Pateyron S., Cihalikova J., Sourdille P., Simkova H., Faivre-Rampant P., Hribova E., Bernard M., Lukaszewski A., Dolezel J., Chalhoub B., Advanced resources for plant genomics: a BAC library specific for the short arm of wheat chromosome 1B, Plant J, 47 (2006) 977–986.

[21] Katari M.S., Balija V., Wilson R.K., Martienssen R.A., McCombie W.R., Comparing low coverage random shotgun sequence data from *Brassica oleracea* and *Oryza sativa* genome sequence for their ability to add to the annotation of *Arabidopsis thaliana*, Genome Res. 15 (2005) 496–504.

[22] Kim C.S., Lee C.H., Shin J.S., Chung Y.S., Hyung N.I., A simple and rapid method for isolation of high quality genomic DNA from fruit trees and conifers using PVP, Nucl. Acids Res. 25 (1997) 1085–1086.

[23] Kinlaw C.S., Neale D.B., Complex gene families in pine genomes, Trends Plant Sci. 2 (1997) 356–359.

[24] Lev-Yadun S., Sederoff R., Pines as model gymnosperms to study evolution, wood formation, and perennial growth, J. Plant Growth Regul. 19 (2000) 290–305.

[25] Lijavetzky D., Muzzi G., Wicker T., Keller B., Wing R., Dubcovsky J., Construction and characterization of a bacterial artificial chromosome (BAC) library for the A genome of wheat, Genome 42 (1999) 1176–1182.

[26] Liu D., Wu R., Protection of Megabase-Sized Chromosomal DNA from Breakage by DNase Activity in Plant Nuclei, BioTechniques 26 (1999) 258–261.

[27] Luo M., Wang Y.-H., Frisch D., Joobeur T., Wing R.A., Dean R.A., Melon bacterial artificial chromosome (BAC) library construction using improved methods and identification of clones linked to the locus conferring resistance to melon *Fusarium* wilt (*Fom-2*), Genome 44 (2001) 154–162.

[28] Luro F., Laigret F., Preparation of high molecular weight genomic dna from nuclei of woody plants, BioTechniques 19 (1995) 388–391.

[29] Peterson D.G., Tomkins J.P., Fritsch D.A., Wing R.A., Paterson A.H., Construction of plant bacterial artificial chromosome (BAC) libraries: an illustrated guide, J. Agric. Genomics 5 (2000) http://www.ncgr.org/jag/index.html.

[30] Peterson D.G., Wessler S.R., Paterson A.H., Efficient capture of unique sequences from eukaryotic genomes, Trends Genet. 18 (2002) 547–550.

[31] Plomion C., Hurme P., Frigerio J.M., Ridolfi M., Pot D., Pionneau C., Avila C., Gallardo F., David H., Neutlings G., Campbell M., Cánovas F.M., Savolainen O., Bodénès C., Kremer A., Developing SSCP markers in two *Pinus* species, Mol. Breed. 5 (1999) 21–31.

[32] Rabinowicz P.D., Constructing gene-enriched plant genomic libraries using methylation filtration technology, Methods Mol. Biol. 236 (2003) 21–36.

[33] Rabinowicz P.D., Citek R., Budiman M.A., Nunberg A., Bedell J.A., Lakey N., O'Shaughnessy A. L., Nascimento L.U., McCombie W.R., Martienssen R.A., Differential methylation of genes and repeats in land plants, Genome Res. 15 (2005) 1431–1440.

[34] Sambrook J., Russell D.W., Molecular Cloning, CSHL Press, Cold Spring Harbour, 2001.

[35] Schmidt T., Heslop-Harrison J.S., Genomes, genes and junk: the large-scale organization of plant chromosomes, Trends Plant Sci. 3 (1998) 195–198.

[36] She K., So you want to work with giants: the BAC vector, BioTech. J. 1 (2003) 69–74.

[37] Shizuya H., Birren B., Kim U.J., Mancino V., Slepak T., Tachiiri Y., Simon M., Cloning and stable maintenance of 300-kb pair fragments of human DNA in *Escherichia coli* using a F-factor-based vector, Proc. Natl. Acad. Sci. USA 89 (1992) 8794–8797.

[38] Sinnett D., Richer C., Baccichet A., Isolation of Stable Bacterial Artificial Chromosome DNA Using a Modified Alkaline Lysis Method, BioTechniques 24 (1998) 752–754.

[39] Vanhouten W., MacKenzie S., Construction and characterization of a common bean bacterial artificial chromosome library, Plant Mol. Biol. 40 (1999) 977–983.

[40] Villalobos D.P., Bautista R., Cánovas F.M., Claros M.G., Isolation of Bacteria Artificial Chromosome DNA by means of improved alkaline lysis and double potassium acetate precipitation, Plant Mol. Biol. Rep. 22 (2004) 419–425.

[41] Wakasugi T., Tsudzuki J., Ito S., Nakashima K., Tsudzuki T., Sugiura M., Loss of all *ndh* genes as determined by sequencing the entire chloroplast genome of the black pine *Pinus thunbergii*, Proc. Natl. Acad. Sci. USA 91 (1994) 9794–9798.

[42] Weber A., Flügge U.I., Interaction of cytosolic and plastidic nitrogen metabolism in plants, J. Exp. Bot. 53 (2002) 865–874.

[43] Yüksel B., Paterson A.H., Construction and characterization of a peanut HindIII BAC library, Theor. Appl. Genet. 111 (2005) 630–639.

[44] Zhang H.-B., Zhao X., Ding X., Paterson A.H., Wing R.A., Preparation of megabase-size DNA from plant nuclei, Plant J. 7 (1995) 175–184.

[45] Zhang H.B., Wu C.C., BAC as tools for genome sequencing, Plant Physiol. Biochem. 39 (2001) 195–209.