Cot-based cloning and sequencing of the short arm of wheat chromosome 1B

H. Šimková, J. Janda, E. Hřibová, J. Šafář, J. Doležel

Institute of Experimental Botany, Academy of Sciences of the Czech Republic, Olomouc, Czech Republic

ABSTRACT

Sequencing of cereal genomes is not a feasible task due to their large size and high content of repetitive DNA sequences. There are two basic approaches to simplify analysis of such genomes: reduced representation approaches, such as EST sequencing, methyl filtration and Cot-based cloning and sequencing; on the other side there is analysis of genomes in a step-wise manner, e.g. through creation of chromosome-specific genomic resources. Combination of both approaches – i.e. Cot-based cloning and sequencing of DNA obtained from a chromosome-arm-specific BAC library – was tested in this work.

Keywords: Cot; CBCS; Cot filtration; BAC library; 1 BS; wheat; genome analysis

Cot-based cloning and sequencing (CBCS), a method combining Cot analysis, DNA cloning and high-throughput sequencing, was described as a new and efficient approach to capture unique sequences from eukaryotic genomes (Peterson et al. 2002b; available at: http://www.mgel.msstate.edu/cf.htm). In particular, CBCS should (1) permit efficient gene discovery in species with substantial quantities of repetitive DNA, (2) allow the sequence complexity of large genomes to be elucidated at a fraction of the cost of whole-genome shotgun sequencing, and (3) facilitate capture of low-copy sequences not secured by EST sequencing.

CBCS includes Cot analysis, a biochemical method set on principles of DNA reassociation kinetics. This method is based on the observation that in a solution of heat-denatured, sheared genomic DNA, a specific sequence reassociates at a rate proportional to the number of times it occurs in the genome (Britten and Kohne 1968). Whereas highly repetitive sequences reassociate at a low Cot value (Cot value = product of a sample's nucleotide concentration in moles per liter, its reassociation time in seconds and a factor based upon the cation concentration of the buffer, see Britten et al. 1974), low-copy and unique sequences reassociate at high Cot values. Based on these observations, Cot analysis can be used to fractionate the genome into subpopulations of DNA fragments differing in their iteration frequency. Hydroxyapatite (HAP) chromatography is then employed to separate single-stranded DNA (ssDNA) from double-stranded DNA (dsDNA) (see Peterson et al. 1998, Peterson et al. 2002a for details). Until now, CBCS was used to analyze highly repetitive part of the genome of chicken (Wicker et al. 2005), banana (Hřibová et al. 2004) or *Panax ginseng* (Ho and Leung 2002), to characterize the genome of *Sorghum* by separating highly repetitive, moderately repetitive, and single/low-copy components of the genome (Peterson et al. 2002a) and to analyze high-Cot fraction of the maize genome (Yuan et al. 2003). Recently, the efficacy of this technique for gene enrichment was also tested in wheat (Lamoureux et al. 2005), in which the high-Cot libraries exhibited 13.7 fold enrichment in genes compared with sequences from a shotgun-library. In maize, representation of genic sequences in the high-Cot fraction was fourfold higher compared with the non-fractionated genomic DNA, increasing from

Supported by the Czech Science Foundation, Project No. 521/05/P257.

This contribution was presented at the 4th Conference on Methods in Experimental Biology held in 2006 in the Horizont Hotel in Šumava, Czech Republic.

5% for a random library to more than 20% for a high-Cot library (Yuan et al. 2003).

A completely different approach to simplify analysis of large genomes is based on their analysis in a step-wise manner. The pre-requisite of this approach is the availability of subgenomic large-insert libraries, usually cloned in bacterial artificial chromosome (BAC). In bread wheat, one of the most important crops ranking with its composed genome of about 17 Gbp among the largest, were constructed several subgenomic BAC libraries from flow-sorted chromosomes: BAC library specific for chromosomes 1D, 4D and 6D (Janda et al. 2004), for chromosome 3B (Šafář et al. 2004) and for the short arm of chromosome 1B (Janda et al. 2006). These libraries markedly simplify genome analyses, construction of physical maps and isolation of genes (Kota et al. 2006).

Combination of both strategies, i.e. CBCS and creation of chromosome-specific genomic resources, could provide an even more focused and efficient tool to capture unique sequences from extremely large genomes.

MATERIAL AND METHODS

Plant materials and DNA preparation

BAC library specific for the short arm of chromosome 1B was constructed from flow-sorted chromosome arms 1BS present as telosomes in a ditelosomic line of hexaploid wheat *Triticum aestivum* L. cv. Pavon 76 ($2n = 40 + 2t1BS$) (Janda et al. 2006). The library consists of 65 280 BAC clones with average insert size of 82 kb. The library is ordered in 170 plates. BAC clones from each of the 384-well plates were grown overnight on a solid medium and then resuspended in 4 ml of TE buffer (10mM Tris, 1mM EDTA). Bacterial suspensions from several plates were combined. The suspensions were used to isolate DNA by standard alkaline lysis method. Remnants of bacterial chromosomal DNA were removed by Microcon YM-100 column (Millipore). BAC vector (7.5 kb) was not removed as it was expected to become a part of the highly repetitive fraction, which was not the target in this experiment. Moreover, it served as an internal control to prove that highly repetitive sequences were not included in the Cot fraction obtained.

Cot fractionation and cloning of the DNA was performed following the protocol of Yuan et al. (2003) with some modifications.

Shearing, denaturation and reassociation

The DNA was filtered using Ultrafree-MC filter unit (Millipore) to remove tiny particles. The filtered DNA was sheared mechanically using Hydroshear DNA Shearing Device (GeneMachines) to obtain fragments of 0.8–2.5 kb (1.5 kb average). 40 µg of DNA were diluted with 20 × SSC and water to reach final concentration of SSC 1× and that of DNA 400 ng/µl. The samples were denatured in boiling water for 5 min and the DNA was let reassociate at 65°C for 23 h and 32 min to reach Cot value 100. The reassociation was stopped by freezing the samples at –80°C.

Considering the relatively small size of 1BS (315 Mbp), the Cot value of 100 represents a relatively high Cot value. This was supported by estimation of ssDNA proportion at various Cot values (50, 100, and 200). For comparison, an attempt to clone even higher Cot fraction (Cot200) was made as well.

HAP chromatography

HAP chromatography was performed in columns with 125 mg hydroxyapatite (Bio-Rad Laboratories) at 60°C. Samples were loaded in 40mM sodium phosphate buffer (40mM SPB, 40mM sodium phosphate, 10mM EDTA), the ssDNA was eluted with about 3 ml 110mM SPB (110mM sodium phosphate, 10mM EDTA) and afterwards the dsDNA was eluted with 3 ml 300mM SPB (300mM sodium phosphate, 10mM EDTA). The samples were purified and concentrated using Microcon YM-30 column (Millipore). The ssDNA:dsDNA ratio was checked by comparing intensity of fluorescence of smears obtained for eluted ssDNA and dsDNA fractions, respectively, after electrophoresis and subsequent staining in ethidium bromide (0.5 µg/ml).

Concentration of ssDNA was estimated by Fluorimeter TD700 (Turner Designs) using Oligreen (Molecular Probes).

Synthesis of the second strand and cloning of DNA fragments

The ssDNA was converted into dsDNA using hexamer random priming system. The reaction was carried out in a 20-µl volume containing about 400 ng ssDNA, 12µM random hexanucleotides, 250µM dNTPs and 0.1U Klenow exo (Fermentas) and performed at room temperature for 4–5 hours.

The reaction was stopped by adding 1 µl of 0.5M EDTA and concentration of dsDNA was estimated. The reaction buffer was removed using Microcon YM-30 column (Millipore) and the DNA was washed out with a suitable amount of water to reach final concentration of DNA 100 ng/µl. The dsDNA was treated with mung bean nuclease (Roche) in concentration 2U/µg DNA for 30 min at 30°C to generate blunt-ended molecules. The reaction was stopped by adding 1/10 volume of 0.1% SDS. The buffer was exchanged and the DNA was concentrated using Microcon YM-30 column (Millipore).

To clone fragments of desired size, size selection was performed in 1% low-melting agarose gel in 1 × TAE. Fragments of 800–1500 bp were isolated from the gel using Gel Extraction Kit (Qiagen) and cloned into the pCR – BluntII-TOPO vector (Invitrogen), which was used to transform One Shot TOP10 Electrocomp *Escherichia coli* cells (Invitrogen).

DNA sequencing and sequence analysis

The obtained clones were checked for the insert size and 85 clones with inserts above 500 bp were sequenced at the Department of Plant Sciences, University of Arizona, Tucson, USA. DNA sequences were assembled and edited using the BioEdit software. The homology search was done against sequences deposited in the GenBank and the GrainGenes databases for *Triticum* and *Avena* (available at: http://wheat.pw.usda.gov/cgi-bin/westsql/map_locus.cgi) using BLASTn and BLASTx (Altschul et al. 1997). The presence of duplicated and tandem-organized regions was detected using Tandem Repeats Finder (TRF, Benson 1999).

RESULTS AND DISCUSSION

For preparation of the Cot100 fraction, 40 µg of DNA were isolated from the pooled 1BS BAC library. After HAP chromatography, 730 ng of ssDNA were eluted. For Cot50 and Cot200, 1 240 ng and 230 ng of ssDNA, respectively, were obtained from the same amount of DNA. 660 ng of the Cot100 ssDNA were used for synthesis of the second strand of DNA. The yield of the random primed synthesis was 1 511 ng of dsDNA. After size selection, 480 ng of DNA were isolated from the gel and 78 ng were used for the TOPO cloning reaction, providing thousands of clones. 1 536 clones were picked and ordered in four 384-well plates. Despite the size selection, average insert size was only 435 bp. 85 clones with inserts above 500 bp were sequenced and analysed and the data were sent to the GenBank database. The sequenced clones were screened for homology with sequences deposited in the GenBank and the GrainGenes databases for *Triticum* and *Avena*. Five clones showed homology with 1BS-specific ESTs, nine clones were homologous to different parts of retrotransposons (five clones were homologous to genes for different retrotransposon proteins, three clones showed homology with reverse transcriptase and one clone was homologous to transposase) and further 14 clones were homologous to other known DNA sequences (Figure 1). Tandem Repeat Finder analysis revealed that 31 clones contained tandem-organized repetitive units. Most of them (23) consisted of duplicated sequence units, two clones consisted of microsatellite sequences and six clones contained more than two tandem-organized repetitive units (with minimal unit size 20 bp). None of the sequences showed homology with the BAC vector, which indicates that we succeeded in filtering out the highest repetitive fraction.

For comparison, Cot200 fraction was prepared from 1BS using the same protocol, however the yield of ssDNA after HAP chromatography and number of obtained clones were significantly lower. Moreover, prolonged reassociation led to shortening of fragments obtained after HAP chromatography. Thus cloning of higher Cot fractions of 1BS does not seem feasible. On the other hand, Lamoureux et al. (2005) succeeded in cloning wheat DNA collected at Cot values of 1 188 and 1 639. This can be due to the fact that the size of the whole wheat genome is 53 fold larger than that of 1BS, thus its complexity must be higher and the reassociation proceeds much slower. Moreover, 1BS belongs to chromosome arms with the highest

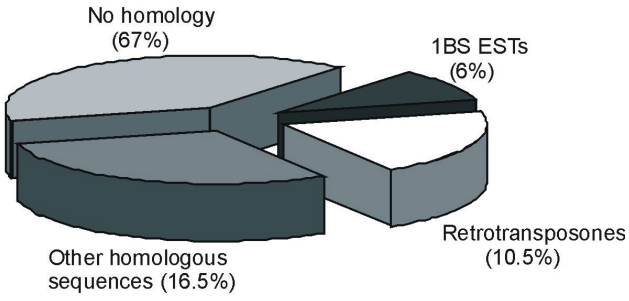

Figure 1. Homology of sequences from the Cot100 library with sequences deposited in the GenBank and the GrainGenes databases

content of heterochromatic regions (Kubaláková, personal communication), which accelerates the reassociation. However, even at Cot value of 1 639 for wheat Lamoureux et al. (2005) obtained 31.4% repetitive sequences (22.6% retrotransposons) and Yuan et al. (2003) gained 31.8% repetitive sequences (15% retroelements) at Cot 466 for maize. This shows limited potential of CBCS for filtering out the repetitive sequences of large genomes.

There are some critical steps of this procedure, which can result in lower yield or production of artifacts. Prolonged denaturation should be avoided as it showed to damage the DNA. The most critical step is the synthesis of the second strand by random priming reaction. The composition of the reaction must be carefully considered, as lower concentration of random hexanucleotides leads to low yield but higher concentration results in significant production of reaction artifacts. Relatively high content of sequences with no homology to known sequences present in our library (67%) but also in Cot466 library of maize (44%) (Yuan et al. 2003) could be caused by such artifacts although it can be also interpreted as a potential of this method to discover previously unknown functional sequences. Moreover, Britten and Kohne (1968) and Britten et al. (1974) found that the resolution of distinct kinetic components can only be achieved if the DNA fragments in the Cot analysis are relatively short (200–600 bp). Such a length of fragments is not efficient for sequencing by classical Sanger technology but could be potentially exploited in projects based on 454 sequencing technology.

Considering relatively high costs, time consumption and difficulty of CBCS together with difficulties of assembly of short DNA fragments into contigs, we do not consider it as a progressive method for analyzing large genomes in combination with chromosome-specific and arm-specific libraries. These libraries can be more reasonably used for contig assembly and clone-by-clone sequencing.

REFERENCES

Altschul S.F., Madden T.L., Schäffer A.A., Zhang J., Zhang Z., Miller W., Lipman D.J. (1997): Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res., *25*: 3389–3402.

Benson G. (1999): Tandem repeat finder – a program to analyze DNA sequences. Nucleic Acids Res., *27*: 573–580.

Britten R.J., Graham D.E., Neufeld B.R. (1974): Analysis of repeating DNA sequences by reassociation. Method. Enzymol., *29*: 363–405.

Britten R.J., Kohne D.E. (1968): Repeated sequences in DNA. Science, *161*: 529–540.

Ho I.S., Leung F.C. (2002): Isolation and characterization of repetitive DNA sequences from *Panax ginseng*. Mol. Genet. Genomics, *266*: 951–961.

Hřibová E., Macas J., Neumann P., Doležel J. (2004): Isolation of highly repetitive DNA sequences in *Musa acuminata* using reassociation kinetics. In: Abstr. 1st Int. Congr. Musa, Penang, Malaysia.

Janda J., Bartoš J., Šafář J., Kubaláková M., Valárik M., Číhalíková J., Šimková H., Caboche M., Sourdille P., Bernard M., Chalhoub B., Doležel J. (2004): Construction of a subgenomic BAC library specific for chromosomes 1D, 4D and 6D of hexaploid wheat. Theor. Appl. Genet., *109*: 1337–1345.

Janda J., Šafář J., Kubaláková M., Bartoš J., Kovářová P., Suchánková P., Hřibová E., Pateyron S., Číhalíková J., Sourdille P., Šimková H., Faivre-Rampant P., Bernard M., Lukaszewski A., Doležel J., Chalhoub B. (2006): Novel resources for plant genomics: BAC library specific for short arm of wheat chromosome 1B. Plant J., *47*: 977–986.

Kota R., Spielmeyer W., Doležel J., Šafář J., Paux E., Mcintosh R.A. Lagudah E.S. (2006): Towards isolating the durable stem rust resistance gene *Sr2* from hexaploid wheat (*Triticum aestivum* L.). In: Abstr. Plant and Animal Genome XIV Conf., San Diego, California.

Lamoureux D., Peterson D.G., Li W.L., Fellers J.P., Gill B.S. (2005): The efficacy of Cot-based gene enrichment in wheat (*Triticum aestivum* L.). Genome, *48*: 1120–1126.

Peterson D.G., Pearson W.R., Stack S.M. (1998): Characterization of the tomato (*Lycopersicon esculentum*) genome using *in vitro* and *in situ* DNA reassociation. Genome, 41: 346–356.

Peterson D.G., Schulze S.R., Sciara E.B., Lee S.A., Bowers J.E., Nagel A., Jiang N., Tibbitts D.C., Wessler S.R., Paterson A.H. (2002a): Integration of Cot analysis, DNA cloning, and high-throughput sequencing facilitates genome characterization and gene discovery. Genome Res., *12*: 795–807.

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

Šafář J., Bartoš J., Janda J., Bellec A., Kubaláková M., Valárik M., Pateyron S., Weiserová J., Tušková R., Číhalíková J., Vrána J., Šimková H., Faivre-Rampant P., Sourdille P., Caboche M., Bernard M., Doležel J., Chalhoub B. (2004): Dissecting large and complex genomes: flow sorting and BAC cloning of individ-

ual chromosomes from bread wheat. Plant J., *39*: 960–968.

Wicker T., Robertson J.S., Schulze S.R., Feltus F.A., Magrini V., Morrison J.A., Mardis E.R., Wilson R.K., Peterson D.G., Paterson A.H., Ivarie R. (2005): The repetitive landscape of the chicken genome. Genome Res., *15*:126–136.

Yuan Y., Sanmiguel P.J., Bennetzen J.L. (2003): High-Cot sequence analysis of the maize genome. Plant J., *34*: 249–255.

Received on February 27, 2007

Corresponding author:

M.Sc. Hana Šimková, Ph.D., Akademie věd České republiky, Ústav experimentální botaniky, v. v. i., Sokolovská 6, 772 00 Olomouc, Česká republika
e-mail: simkovah@ueb.cas.cz