Advanced resources for plant genomics: a BAC library specific for the short arm of wheat chromosome 1B

Jaroslav Janda1, Jan Šafář1, Marie Kubaláková1,2, Jan Bartoš1, Pavlína Kovářová1, Pavla Suchánková1, Stephanie Pateyron3, Jarmila Číhalíková1,2, Pierre Sourdille4, Hana Simková1, Patricia Faivre-Rampant3, Eva Hribová1, Michel Bernard4, Adam Lukaszewski5, Jaroslav Doležel1,2,* and Boulos Chalhoub3

1Laboratory of Molecular Cytogenetics and Cytometry, Institute of Experimental Botany, Sokolovská 6, CZ-77200 Olomouc, Czech Republic,
2Department of Cell Biology and Genetics, Palacký University, Šlechtitelů 11, CZ-78371 Olomouc, Czech Republic,
3Organization and Evolution of Plant Genomes, Unité de Recherches en Génomique Végétale (INRA-URGV), 2 rue Gaston Crémeux, CP 5708, F-91057 Évry Cedex, France,
4Génétique Moléculaire des Céréales, UMR INRA-UBP, Domaine de Crouelle, 234 Avenue du Brézet, F-63039 Clermont-Ferrand Cedex 2, France, and
5Department of Botany and Plant Sciences, University of California, Riverside, CA 92521, USA

Received 1 November 2005; revised 2 May 2006; accepted 30 May 2006.
*For correspondence (fax + 420 585 205 853; e-mail dolezel@ueb.cas.cz).

Summary

Common wheat (Triticum aestivum L., 2n = 6x = 42) is a polyploid species possessing one of the largest genomes among the cultivated crops (1C is approximately 17 000 Mb). The presence of three homoeologous genomes (A, B and D), and the prevalence of repetitive DNA make sequencing the wheat genome a daunting task. We have developed a novel ‘chromosome arm-based’ strategy for wheat genome sequencing to simplify this task; this relies on sub-genomic libraries of large DNA inserts. In this paper, we used a di-telosomic line of wheat to isolate six million copies of the short arm of chromosome 1B (1BS) by flow sorting. Chromosomal DNA was partially digested with HindIII and used to construct an arm-specific BAC library. The library consists of 65 280 clones with an average insert size of 82 kb. Almost half of the library (45%) has inserts larger than 100 kb, while 18% of the inserts range in size between 75 and 100 kb, and 37% are shorter than 75 kb. We estimated the chromosome arm coverage to be 14.5-fold, giving a 99.9% probability of identifying a clone corresponding to any sequence on the short arm of 1B. Each chromosome arm in wheat can be flow sorted from an appropriate cytogenetic stock, and we envisage that the availability of chromosome arm-specific BAC resources in wheat will greatly facilitate the development of ready-to-sequence physical maps and map-based gene cloning.

Keywords: wheat, genomics, physical mapping, BAC resources, chromosome sorting, flow cytometry.

Introduction

Common wheat (Triticum aestivum L.) is grown on a larger acreage than any other crop plant. Because of its socio-economic importance, a sustained increase in production must be ensured to keep pace with the growing human population. It is expected that the breeding of improved varieties will be accelerated by a better knowledge of the wheat genome. However, genomics in wheat has been hampered by the size and the complexity of its genome. Common wheat is an allohexaploid species consisting of three closely related genomes (AABBDD, 2n = 6x = 42), resulting in sequence redundancy (Gill et al., 2004). The estimated size of the hexaploid wheat genome is 17 billion bases (Bennett and Smith, 1991), mainly consisting of repetitive DNA (approximately 80%, Smith and Flavell, 1975), and is >100 times larger than the genome of the model plant Arabidopsis thaliana. Each individual wheat chromosome is larger than the recently sequenced rice genome (Goff et al., 2002). These unique features pose great challenges for gene discovery and genome sequencing.
Because of its large size and complexity, sequencing of the wheat genome requires careful consideration and development of appropriate strategies and technologies. In addition to clone-by-clone sequencing and selective sequencing of random bacterial artificial chromosome (BAC) clones from gene-rich regions, shotgun sequencing of ‘gene-enriched’ DNA obtained after genome filtration was suggested (Gill et al., 2004). Regardless of which strategy is chosen, a robust physical map will be required. This holds true not only for the sequencing approaches based on large-insert DNA clones but also for short sequences obtained from gene-enriched DNA that will have to be ordered on the genomic scaffold (Meyers et al., 2004). Beside its utility for genome sequencing, the availability of a global physical map would greatly accelerate positional gene cloning.

Physical maps are produced by ordering DNA clones from large-insert (typically BAC) libraries on the basis of a clone fingerprint pattern (Luo et al., 2003). Several genomic BAC libraries have been constructed for bread wheat (Allouis et al., 2003; Liu et al., 2000; Ma et al., 2000; Nilmaligoda et al., 2003). To achieve the required genome coverage, one to two million BAC clones have to be fingerprinted. Although it is possible to fingerprint such large numbers of clones with the automated fingerprinting technique of Luo et al. (2003), it is doubtful that biological constraints due to polyploidy and current contig assembly technology would allow contigs and physical maps to be built from such a large number of fingerprints (Jan Dvořák, University of California, Davis, CA, USA, personal communication).

One approach to simplify this task is to employ diploid relatives of wheat whose genomes are smaller. Several BAC libraries representing genomes of individual wheat progenitors have been constructed. Moulet et al. (1999) created a BAC library from the D genome donor of wheat Aegilops tauschii, and Lijavetzky et al. (1999) developed a BAC library from T. monococcum, which is related to T. urartu, the progenitor of the A genome. BAC libraries from T. urartu and Ae. speltoides, the presumed progenitor of the B genome, were developed recently (Akhunov et al., 2005). However, the hexaploid wheat has diverged compared with its wild diploid progenitor species. Polyploidization events cause important DNA loss and rearrangements (Akhunov et al., 2003; Anderson et al., 2003; Chantret et al., 2005; Dvovar et al., 2004; Kong et al., 2004; Wicker et al., 2003) suggesting that the diploid species constitute poor surrogates for genomic studies in bread wheat.

We believe that a fundamental solution to the problem of sequencing large genomes is to create BAC libraries specific for small and defined parts thereof. In line with this, we have developed procedures for purification of individual chromosomes and chromosome arms by flow cytometry (Kubaláková et al., 2002, 2005; Vrána et al., 2000) and a protocol for preparation of intact DNA from sorted chromosomes that is suitable for cloning (Šimková et al., 2003). These advances, and a highly efficient protocol for BAC cloning (Chalhoub et al., 2004), have allowed the creation of two sub-genomic BAC libraries from hexaploid wheat: a composite library specific for chromosomes 1D, 4D and 6D (Janda et al., 2004) and a library specific for chromosome 3B (Šafář et al., 2004). Our ability to sort single chromosome arms from polyploid wheat (Kubaláková et al., 2002, 2005) offers the possibility of constructing BAC libraries from even smaller parts of the genome. The size of individual wheat chromosome arms ranges from 224 to 582 Mb, representing only 1.3–3.4% of the whole genome. This is within the size range of plant genomes that have already been sequenced and/or will be sequenced in the near future.

Even with a reliable chromosome arm purification protocol, creation of a BAC library still represents a challenge because of the minute amounts of DNA obtained from flow-sorted chromosome arms. To the best of our knowledge, no chromosome arm-specific BAC library has been constructed for any organism. In this paper, we describe the construction and characterization of a BAC library specific for the short arm of chromosome 1B (1BS) of wheat. This chromosome arm represents only 1.9% of the hexaploid wheat genome and carries a large number of important genes, including disease and pest resistance genes and genes for storage proteins that directly affect grain quality (Erayman et al., 2004; Sandhu et al., 2001). The Wheat Gene Catalog (http://wheat.pw.usda.gov/ggpages/wgc/98/) is a good reference to the gene content of chromosome arm 1BS. The ability to create a BAC library from a particular chromosome arm opens avenues for the development of a physical map of the complex hexaploid wheat genome and its analysis in a targeted and step-wise manner, working with one particular chromosome arm at a time. In addition to simplifying the physical mapping efforts by several orders of magnitude, this chromosome-based approach has the potential for division of labor during sequencing of the wheat genome (Gill et al., 2004).

Results

The short arm of chromosome 1B (1BS) was purified from a di-telosomic line of bread wheat cultivar ‘Pavon 76’. The work involved preparation of liquid suspensions of intact chromosomes from synchronized root tip meristems, staining with a DNA-specific fluorescent dye 4’,6-diamidino-2-phenylindole (DAPI), and analysis of the fluorescence intensity of stained chromosomes using flow cytometry. We obtained a reproducible distribution of relative fluorescence intensity (‘flow karyotype’), which was characterized by five peaks (Figure 1). We confirmed the chromosome content of each peak by sorting particles from each peak onto microscope slides and performing fluorescence in situ hybridization (FISH) with probes for GAA microsatellites and the Afa repeat. The two probes

© 2006 The Authors
facilitated identification of any chromosome arm in hexaploid wheat (Pedersen and Langridge, 1997). We established that peaks I–III were composites of various chromosomes and that the small peak to the right of peak III represented chromosome 3B (Kubala et al., 2002; Vrána et al., 2000). The peak on the far left was found to represent the short arm of chromosome 1B.

Subsequently, we sorted six million copies of chromosome arm 1BS representing approximately 4 lg DNA in batches of 10^5 and embedded them in 60 agarose plugs. The sorting exercise took 27 working days, and approximately 7000 seeds were required to prepare 267 samples of chromosome suspensions. We repeatedly checked the purity of the flow-sorted fractions by microscopic observation of chromosomes sorted on a glass slide and subjected to FISH with probes for GAA microsatellites and Afa repeats. On average, the sorted fractions consisted of 89% chromosome 1BS. Contaminating particles were arms and chromatids of other chromosomes without an apparent prevalence of specific types. Pulsed-field gel electrophoresis showed that the DNA of sorted chromosome arms was intact, indicating its usefulness for BAC library construction (data not shown).

The DNA of sorted chromosomes was partially digested with HindIII and used to prepare a 1BS-specific BAC library. The complete 1BS-specific BAC library (TA-1BS) consists of five sub-libraries created from five independent ligation reactions using different size classes of DNA: H01, 50–75 kb; H02, 75–100 kb; H1, 100–150 kb; H2, 150–200 kb; H3, 200–250 kb. We estimated the average insert size of the library after the analysis of 96 BAC clones selected randomly from the five sub-libraries (Figure 2). The H01 sub-library has 9600 clones and an average insert size of 62 kb, and is similar to the H02 sub-library with 23 808 clones and average insert size 66 kb. Sub-library H1 is the largest, with 27 264 clones and an insert size of 96 kb, while sub-library H2 is the smallest sub-library, with only 384 clones and an average insert size of 110 kb. Finally, the H3 sub-library has 4224 clones and its average insert size is 118 kb. The complete 1BS-specific BAC library comprises 65 280 clones ordered in 170 384-well plates with an average insert size of 82 kb. Almost a half of the TA-1BS library (45%) has inserts larger than 100 kb, while 18% of inserts are 75–100 kb, and 37% of inserts are shorter than 75 kb (data not shown).

Taking into account the wheat genome size of 16 974 Mb 1C⁻¹ (Bennett and Smith, 1991) and the relative

© 2006 The Authors
chromosome lengths given by Gill et al. (1991), we estimate the molecular size of the short arm of chromosome 1BS to be 315 Mb. Thus, this chromosome arm represents only 1.9% of the hexaploid wheat genome. Given the number of clones, the average insert size and the frequency of contaminating particles, this library represents 14.5 size equivalents of 1BS. Using the formula of Clarke and Carbon (1976), we estimated the probability of recovering any DNA sequence present on chromosome 1BS in the library to be 99.9%. On the other hand, the probability of finding any DNA sequence from other chromosomes is as low as 4.5%.

To confirm the chromosome arm specificity and verify the 14.5-fold genome coverage, we screened the TA-1BS library by PCR using a set of microsatellite (SSR) markers. To do this, we divided the library into 170 pools; each pool representing one 384-well plate. Then we screened the pools with nine SSR markers specific for the short arm of chromosome 1B. The number of pools that gave PCR products of the expected lengths with each SSR marker ranged from 10 to 21 (Table 1). Based on these results, we estimated the coverage of the TA-1BS BAC library to be at least 15.2-fold.

Cytogenetic mapping is needed to support the development of physical contig maps by anchoring BAC contigs onto chromosomes. In order to test the usefulness of our new BAC library for cytogenetic mapping, we hybridized 75 randomly selected ‘low-copy’ BAC clones to wheat chromosomes by FISH. Only four of them (5.3%) localized exclusively on 1BS. Another BAC clone (4B9) gave a distinct signal on the satellite of 1BS and on chromosome arms 5AS, 5BS, 1DS and 5DS. Five BAC clones gave FISH patterns similar to that of repetitive DNA family pSc119 (Bedbrook et al., 1980), while FISH with four BAC clones resulted in a distinct signal on 1BS and dispersed signals on all wheat chromosomes. Fifty-two BAC clones hybridized to all wheat chromosomes along their length, indicating the presence of interspersed repeated sequences. The remaining nine BAC clones did not yield any hybridization signals.

Table 1 Results of TA-1BS library screening on pools of 384-well plates with 1BS-specific SSR markers

<table>
<thead>
<tr>
<th>SSR marker</th>
<th>Reference</th>
<th>Number of positive pools</th>
</tr>
</thead>
<tbody>
<tr>
<td>gwm 18</td>
<td>a</td>
<td>21</td>
</tr>
<tr>
<td>gwm 264</td>
<td>a</td>
<td>14</td>
</tr>
<tr>
<td>gwm 413</td>
<td>a</td>
<td>17</td>
</tr>
<tr>
<td>gwm 550</td>
<td>a</td>
<td>12</td>
</tr>
<tr>
<td>gpw 2067</td>
<td>b</td>
<td>10</td>
</tr>
<tr>
<td>gpw 3122</td>
<td>b</td>
<td>13</td>
</tr>
<tr>
<td>gpw 4069</td>
<td>b</td>
<td>14</td>
</tr>
<tr>
<td>gpw 7059</td>
<td>b</td>
<td>21</td>
</tr>
<tr>
<td>gpw 7070</td>
<td>b</td>
<td>15</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td></td>
<td><strong>15.2</strong></td>
</tr>
</tbody>
</table>

a Röder et al. (1998).
b Sourdille and Bernard (unpublished data).

With the aim of overcoming the difficulties observed with BAC-FISH, we explored a possibility of using ‘low-copy’ subclones for FISH. Out of the 75 BAC clones tested, eight BAC clones with different FISH patterns were chosen for subcloning (Table 2). This strategy provided shotgun subclones that hybridized exclusively to unique loci on 1BS. FISH with subclones selected from BAC clones 77E3, 79P7 and 81P18, which hybridized to a single locus on 1BS and multiple loci on all chromosomes, resulted in discrete signals on 1BS only. To some measure, the subcloning strategy was also successful with BAC clones that hybridized evenly over the whole chromosome complement. We succeeded in selecting ‘low-copy’ subclones hybridizing exclusively to specific loci on 1BS from two such BAC clones (78M16 and 81L2). However, subclones obtained from two other BAC clones (80E19, 80M16) gave only dispersed hybridization signals. Finally, FISH with two subclones (4B9-A9 and 4B9-F6) resulted in the same hybridization pattern as FISH with the complete BAC 4B9. Representative examples of FISH with BAC clones and subclones are shown in Figure 3.

Selected ‘low-copy’ subclones that gave a range of FISH patterns were completely sequenced (Table 3). DNA sequences were assembled and edited using BioEdit software (http://www.mbio.ncsu.edu/BioEdit/bioedit.html), and further analyzed using Tandem Repeats Finder (TRF; Benson, 1999) and JDotter (Brodie et al., 2003). All sequences were successfully assembled, except subclones 4B9-A9 and 78M16-G18, for which TRF and JDotter showed the presence of duplicated and tandem organized regions. In case of 4B9-A9, this result could provide an explanation for the FISH pattern obtained (Figure 3). A homology search was performed against the sequences deposited in GenBank, and against the Triticeae Repeat Sequence Database (TREP; http://wheat.pw.usda.gov/ITMI/Repeats/index.shtml). Most of the subclones were found to carry various parts of repetitive elements, the most frequent being various parts of retrotransposons (Table 3). No homology was found for subclones 81P18-B18 and 77E3-E15.

The cytogenetically mapped subclones were then used to test the feasibility of developing chromosome sequence tags. Based on the nucleotide sequences of BAC subclones, we designed primers for PCR and tested their specificity on flow-sorted chromosome arm 1BS, genomic DNA of ‘Pavon 76’, and di-telosomic line 1BL of ‘Pavon 76’. The primers designed for 81P18-B18 and 77E3-F15 subclones were 1BS-specific and gave one PCR product of about 1000 bp (Figure 4).

Discussion

In this paper, we report an important step forward in developing a new generation of genomic resources for bread wheat. We demonstrate that it is possible to create a
BAC library from a specific chromosome arm. The TA-1BS library is specific for the short arm of chromosome 1B, and is the third sub-genomic BAC library produced from individual plant chromosomes, after the composite BAC library specific for wheat chromosomes 1D, 4D and 6D (Janda et al., 2004) and a BAC library from wheat chromosome 3B (Šafář et al., 2004). In addition to creating a unique resource for wheat genomics, the current work confirms that our procedure for

Table 2 Patterns of hybridization to wheat chromosomes observed after FISH with probes for putative ‘low-copy’ BAC clones selected from the 1BS-specific BAC library, and for their ‘low-copy’ shotgun subclones

<table>
<thead>
<tr>
<th>BAC clones</th>
<th>BAC subclones</th>
</tr>
</thead>
<tbody>
<tr>
<td>BAC address</td>
<td>Insert size (kb)</td>
</tr>
<tr>
<td>4B9</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>77E3</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>78M16</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>79P7</td>
<td>45</td>
</tr>
<tr>
<td>80E19</td>
<td>50</td>
</tr>
<tr>
<td>80M16</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>81L2</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>81P18</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 3. Fluorescence in situ hybridization (FISH) on the short arm of wheat chromosome 1B (1BS) using three BAC clones that give three different types of hybridization pattern, and three subclones derived from them. BAC 81L2 (left) and its subclone 81L2-A1 hybridize to the entire length of 1BS except the NOR region. BAC 4B9 (centre) and its subclone 4B9-A9 hybridize to a single locus on 1BS. In contrast, BAC 81P18 hybridizes strongly to a single locus and weakly along the entire length of 1BS, while its subclone 81P18-B18 maps to a single locus on 1BS. The probes were labeled by digoxigenin and the sites of probe hybridization were detected using anti-digoxigenin–FITC (yellow). The chromosomes were counterstained by DAPI (shown in red pseudocolor). Three representative examples of hybridizations on 1BS are given.
the construction of large-insert DNA libraries from DNA of flow-sorted chromosomes is reproducible, reliable and efficient. This advance, together with the availability of a complete series of telosomic lines in wheat (Sears, 1954) and the possibility of purifying wheat telosomes by flow cytometry (Kubalaková et al., 2002), make it possible to create a BAC library from every chromosome arm of wheat.

The TA-1BS library consists of five sub-libraries differing in the number of clones and average insert sizes. These features offer considerable flexibility in the use of the library. Sub-libraries H1, H2 and H3, with an average insert size of about 100 kb and tenfold 1BS coverage, could be used preferentially for contig assembly and chromosome walking. Sub-libraries H01 and H02, which contain clones with smaller inserts of around 65 kb, represent useful resources for FISH mapping. We have confirmed the representative coverage of the short arm of chromosome 1B by library screening with SSR markers that map to 1BS. The 15.2-fold 1BS coverage, as determined by screening with SSR markers, suggests that our estimate of clone representation using average insert size (14.5-fold) is accurate.

Compared with other genomic and sub-genomic BAC resources for hexaploid wheat, the TA-1BS library has a comparable average insert size and the highest genome coverage. Liu et al. (2000) reported on the construction of a TAC (transformation-competent artificial chromosome) library with an average insert size of 54 kb and 3.07-fold genome coverage. A genomic BAC library prepared by Nilmalgoda et al. (2003) comprises $6.5 \times 10^6$ clones, an average insert size of 79 kb, and 3.1-fold genome coverage.

![Figure 4. Agarose gel electrophoresis of PCR products obtained with primers (5'-TCTCCTCCTGTAAGCTCGT-3' and 5'-CTAGAAGGATGCCGAGGATG-3') derived from BAC subclone 81P18-B18. The reaction was performed on flow-sorted short arm of chromosome 1B (1BS), BAC clone 81P18 (BAC), genomic DNA prepared from a line of wheat cultivar 'Pavon 76' di-telosomic for 1BL (Dt1BL) and genomic DNA of 'Pavon 76'. The absence of a product on Dt1BL and its presence on 1BS indicate specificity of the primers for the short arm of 1B. Lane M, 100 bp DNA Ladder Plus marker (Fermentas, Vilnius, Lithuania).](image)

**Table 3** DNA sequences obtained from ‘low-copy’ BAC subclones

<table>
<thead>
<tr>
<th>Subclone address</th>
<th>Sequence length (bp)</th>
<th>GenBank accession number</th>
<th>DNA sequence</th>
<th>GenBank accession number</th>
<th>Smallest sum probability</th>
<th>TREP accession number</th>
<th>Smallest sum probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>4B9-A9</td>
<td>1334*</td>
<td>DX572372</td>
<td>Gypsy-like retrotransposon</td>
<td>AY494981</td>
<td>8e-113</td>
<td>TREP1238</td>
<td>9e-31</td>
</tr>
<tr>
<td>77E3-E15</td>
<td>4079</td>
<td>DX572368</td>
<td>Putative non-LTR retroelement reverse transcriptase</td>
<td>NP922330</td>
<td>3e-28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>77E3-F15</td>
<td>2787</td>
<td>DX572367</td>
<td>Transposon, CACTA</td>
<td>ABA98219</td>
<td>3e-10</td>
<td>TREP746</td>
<td>2e-22</td>
</tr>
<tr>
<td>77E3-O23</td>
<td>2549</td>
<td>DX572366</td>
<td>Transposon protein, CACTA</td>
<td>AAT67049</td>
<td>4e-14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>78M16-G18m13</td>
<td>590*</td>
<td>DX572369</td>
<td>Germin-like protein 4</td>
<td>AAT67049</td>
<td>4e-14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>78M16-G18r</td>
<td>718*</td>
<td>DX572370</td>
<td>Germin-like protein 4</td>
<td>AAT67049</td>
<td>4e-14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>78M16-G18m13f</td>
<td>642*</td>
<td>DX572371</td>
<td>Transposon protein, CACTA</td>
<td>ABA98219</td>
<td>3e-10</td>
<td>TREP821</td>
<td>e-157</td>
</tr>
<tr>
<td>81L2-A1</td>
<td>2376</td>
<td>DX572365</td>
<td>Transposon, LTR, gypsy</td>
<td>AAT67049</td>
<td>4e-14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>81P18-B18</td>
<td>2321</td>
<td>DX572364</td>
<td>Unclassified</td>
<td>TREP1218</td>
<td>6e-20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>81P18-C3</td>
<td>2513</td>
<td>DX572363</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Subclone could not be completely sequenced and/or assembled.
Another published BAC library from hexaploid wheat was created by Allouis et al. (2003). The library consists of $1.2 \times 10^6$ clones, with an average insert size of 130 kb, and represents 9.3 haploid genome equivalents.

A single but serious disadvantage of genomic libraries from hexaploid wheat is their large size, which makes their screening laborious and expensive. Maintenance and handling of genomic libraries with approximately $10^8$ BAC clones remains a non-trivial task (Allouis et al., 2003). However, sub-genomic BAC libraries offer a solution to this problem. Our composite BAC library specific for chromosomes 1D, 4D and 6D, which consists of only 87 168 clones with an average insert size of 85 kb, represents at least 3.4-fold clone coverage (Janda et al., 2004), while the BAC library specific for chromosome 3B, which comprises only 67 968 clones with an insert size of 103 kb, represents at least 6.2 equivalents of the chromosome (Šafař et al., 2004). The TA-1BS library described in this paper comprises 65 280 clones with an average insert size of 82 kb, and represents 14.5 size equivalents of the short arm of chromosome 1B.

In addition to chromosome coverage, specificity is an important quality parameter for chromosome- and chromosome arm-specific BAC libraries. To ensure that our library is specific for chromosome 1BS, we have periodically checked the flow-sorted fractions for contamination. The average content of other chromosome arms and chromatids was 11% as determined by FISH using probes for the GAA microsatellite and the Afa repeat. This FISH approach offers unambiguous identification of all chromosome arms in hexaploid wheat (Kubaláková et al., 2002). Our previous studies (Janda et al., 2004; Šafař et al., 2004) demonstrated that FISH analysis of sorted chromosome fractions provided the most detailed and reliable data on the purity in sorted fractions and the extent of contamination by other chromosomes. Hence, we did not consider it useful to check the TA-1BS library for contamination using PCR with markers specific for other wheat chromosomes.

DNA libraries cloned in a BAC vector are a favorable genomic resource for use in map-based cloning. The availability of a BAC library specific for the short arm of 1B, which represents only 1.9% of the bread wheat genome, should accelerate cloning of important genes in wheat. Preliminary results obtained with our BAC library specific for chromosome 3B (Šafař et al., 2004) confirm these expectations. The small size and the specificity of the library greatly facilitated map-based cloning of major QTL for Fusarium head blight resistance (Liu et al., 2005) and a durable rust-resistance gene (Kota et al., 2006).

Whole-genome physical maps provide a basis for genome sequencing and will be needed whatever the sequencing method employed (Meyers et al., 2004). However, the genome of hexaploid wheat is so complex that the development of a whole-genome physical map seems out of reach of the current technology (Jan Dvořák, University of California, Davis, USA, personal communication). Chromosome- and chromosome arm-specific BAC libraries provide a solution to this problem. As they represent only a small fraction of the whole genome, sufficient coverage of a particular chromosome can be achieved with about 50 000 BAC clones. Using a suitable high-throughput method, a library of this size can be fingerprinted in a few months (Luo et al., 2003). Furthermore, the reduced complexity should greatly simplify the assembly of BAC contigs. Recent results on the production of a physical contig map of wheat chromosome 3B (Paux et al., 2006) fully support these assumptions. The ability to create a BAC library from a particular chromosome arm opens avenues for the development of a physical map of the complex hexaploid wheat genome and its analysis in a targeted and step-wise manner, working with a particular chromosome arm at a time. In addition to simplifying the physical mapping efforts by at least two orders of magnitude, the chromosome-based approach has the potential for a division of labor (Gill et al., 2004).

The development of physical contig maps requires support from cytogenetic mapping to confirm the precise physical positions of centromeres, to determine the distances from the actual chromosome termini of markers near the termini of linkage maps, to resolve the order of contigs and clones, and to estimate the numbers and sizes of gaps (Harper and Cande, 2000). Our results demonstrate that mapping complete BAC clones to specific loci in wheat may be difficult due to the presence of repetitive DNA sequences that hybridize throughout the genome. This conclusion is supported by the observations of Zhang et al. (2004). In this paper, we show that shotgun subcloning and the use of ‘low-copy’ subclones may overcome this problem. Our success in the localization of clones shorter than 2 kb was made possible by performing FISH on flow-sorted chromosomes that are free of cell wall and cytoplasmic remnants. Because thousands of chromosomes can be sorted on one slide, FISH on sorted chromosomes offers higher throughput, higher sensitivity and higher resolution when compared to mitotic metaphase spreads (Valárik et al., 2004). Hence, flow-based cytogenetic mapping could play an important role in the development of a sequence-ready physical map of bread wheat.

In conclusion, this paper demonstrates that high-quality sub-genomic BAC resources representing only few per cent of the whole genome can be prepared for large-genome species. While our two previous studies utilized flow-sorted chromosomes to construct sub-genomic BAC libraries (Janda et al., 2004; Šafař et al., 2004), this study marks a fundamental advance and reports on the creation of a large-insert library specific for a particular chromosome arm. We envisage that the use of chromosome arm-specific BAC resources will accelerate the development of sequence-ready physical contig maps and gene cloning, as well as comparative analyses aiming at revealing the genome changes accompanying the evolution of polyploid wheat.
Experimental procedures

Plant material

A di-telosomic line of hexaploid wheat Triticum aestivum L. cv. ‘Pavon 76’ (2n = 40 + 21BS) carrying the short arm of chromosome 1B (1BS) in the form of a telosome was used. Seeds were germinated in the dark at 25 ± 0.5°C on moistened filter paper for 3 days to obtain a root length around 2–3 cm. Approximately 7000 seeds were germinated in batches of 25–30 for the preparation of 267 chromosome suspensions which were further used for flow sorting.

Preparation of chromosome suspensions and flow-cytometric sorting

Cell-cycle synchronization and the preparation of suspensions of intact chromosomes were performed according to the method described by Vrana et al. (2000). Briefly, each of the 267 samples of chromosome suspension was prepared by mechanical homogenization of 25 formaldehyde-fixed meristem root tips in 1 ml ice-cold isolation buffer (Šimková et al., 2003). Chromosomes in suspension were stained with 2 μg ml⁻¹ DAPI (4',6-diamidino-2-phenylindole) and analyzed using a FACSVantage flow cytometer equipped with a UV laser set to multi-line UV and running at 300 mW output power. In order to sort the chromosome 1BS, the sort window was set on a dot plot of fluorescence area versus fluorescence pulse width. The chromosome arm was sorted in aliquots of 1×10⁵ into 160 μl of 1.5x isolation buffer. The purity of sorted fractions was checked microtitre plates filled with freezing medium (Woo et al., 1994). The purity of sorted fractions was checked using fluorescence microscope slide. After air-drying, sorted chromosomes were identified using fluorescence in situ hybridization (FISH) with probes for the GAA microsatellite and Afa repeat, which give chromosome-specific fluorescent labeling patterns (Kubaláková et al., 2002).

BAC library construction

Preparation of high-molecular-weight DNA and BAC library construction were performed as described previously (Chalhoub et al., 2004; Janda et al., 2004; Šafář et al., 2004). Briefly, each batch of 10⁵ flow-sorted 1BS chromosomes was pelleted (200 g), and the chromosomes were embedded in 15 μl of low-melting-point agarose (0.8% w/v). In total, 60 agarose mini-plugs representing 6×10⁶ sorted chromosomes were prepared and used for DNA cloning. Chromosomal DNA was partially digested with HindIII and size-selected by pulsed-field gel electrophoresis (PFGE) in 0.25x TBE buffer at 6 V cm⁻¹, with a 1–40 sec switch time ramp, angle 120°, for 12 h, and with a 2.5–5 sec switch time ramp, angle 120°, for 6 h at 14°C. Five regions of the gel, corresponding to approximately 50–75 kb (H01 fraction), 75–100 kb (H02 fraction), 100–150 kb (H1 fraction), 150–200 kb (H2 fraction) and 200–250 kb (H3 fraction) were excised, the high-molecular-weight DNA was electro-eluted and ligated into a dephosphorylated vector pIndigoBAC5 (Caltech, Pasadena, CA, USA) to reveal clones potentially containing a low amount of repetitive DNA sequences. Then the filters were screened with a probe for BAC vector to avoid the selection of non-growing clones. Seventy-five ‘low-copy’ BAC clones showing no or weak hybridization signals were tested as BAC-FISH, and eight of them were selected for subcloning. In addition, they were localized on flow-sorted wheat chromosomes using FISH as described previously (Šafář et al., 2004). No blocking DNA was used in the hybridization mix for FISH.

Estimation of the insert size of BAC clones

Ninety-six BAC clones were randomly selected from all five ligations (H01–H5), and incubated overnight at 37°C in 1.5 ml of 2YT medium (Sambrook and Russell, 2001) containing 12.5 μg ml⁻¹ chloramphenicol. BAC DNA was extracted by a standard alkaline lysis method and digested with NotI restriction endonuclease (0.25 U/20 μl). PFGE was used to separate DNA fragments in 1% agarose gel with 0.5x TBE buffer at 6 V cm⁻¹, with a 1–40 sec switch time ramp, angle 120°, for 14 h at 14°C. The insert sizes were estimated after comparison with a lambda size standard run in the same gel.

Screening the library with 1BS-specific probes

After growing overnight in 2YT medium supplemented with chloramphenicol (12.5 μg ml⁻¹) in 384-well plates, BAC clones from each of the 384-well plates were pooled, pelleted and resuspended in 4 ml TE buffer (10 mM Tris, 1 mM EDTA). Bacterial suspensions were lysed at 95°C for 30 min, pelleted at 3000 g for 60 min, and the supernatant was diluted 25-fold using deionized water for PCR reactions. The reaction mix (10 μl) consisted of 2 μl template DNA, 1x PCR buffer, 1.5 mM MgCl₂, 0.2 mM dNTPs, 1.25 μM primers and 0.5 U AmpliTag DNA polymerase (Roche, Mannheim, Germany). The PCR reaction was performed as follows: nine cycles consisting of 5 min at 94°C, 12 sec at 65°C and 18 sec at 72°C, and 35 cycles of 12 sec at 94°C, 12 sec at 55°C and 18 sec at 72°C, and a final extension at 72°C for 7 min. The presence of PCR products was checked by electrophoresis on 4% MetaPhor agarose gel (FMC Bioproducts, Philadelphia, PA, USA).

Selection of ‘low-copy’ BAC clones

A total of 6144 BAC clones from the H01 sub-library (sixteen 384-well plates) were spotted onto two 8 × 12 cm Hybond N+ filters (AP Biotech, Piscataway, NJ, USA). The filters were hybridized with wheat genomic DNA labeled with alkaline phosphatase using Alk-Phos Direct (Amersham Biosciences, Little Chalfont, Buckinghamshire, UK) to reveal clones potentially containing a low amount of repetitive DNA sequences. The filters were screened with a probe for BAC vector to avoid the selection of non-growing clones. Seventy-five ‘low-copy’ BAC clones showing no or weak hybridization signals were tested as BAC-FISH, and eight of them were selected for subcloning. In addition, they were localized on flow-sorted wheat chromosomes using FISH as described previously (Šafář et al., 2004). No blocking DNA was used in the hybridization mix for FISH.

Subcloning of ‘low-copy’ BAC clones

DNA of ‘low-copy’ BAC clones was isolated using the Large-Construct Kit (Qiagen, Valencia, CA, USA), and physically fragmented using a HydroShear DNA shearing device (GeneMachines, San Carlos, CA, USA). Fragments of 3–5 kb were ligated into pCR-XL-TOPO vector (Invitrogen Life Technologies, Carlsbad, CA, USA). Ligation mixtures were transformed into One-Shot TOP10 E. coli (Invitrogen). A total of 384 subclones from each of the eight BAC clones were ordered in 384-well plates filled with freezing medium (Woo et al., 1994), incubated at 37°C overnight, and stored at –80°C. All 3072 BAC subclones were spotted onto one 8 × 12 cm Hybond N+ filter and screened with labeled genomic DNA as described above. One to three subclones from each BAC showing a
weak signal after hybridization with labeled genomic DNA were randomly chosen for mapping using FISH.

**Fluorescence in situ hybridization**

Preparation of digoxigenin- or biotin-labeled probes and FISH on flow-sorted chromosomes were performed as described by Šafař et al. (2004). The sites of probe hybridization were detected using anti-digoxigenin-fluorescein isothiocyanate and streptavidin-Cy3, and the preparations were counterstained with DAPI or propidium iodide. The preparations were evaluated using a Olympus BX60 microscope (Olympus, Tokyo, Japan) with a CCD camera interfaced to a PC running ISIS software (Metasystems, Altlussheim, Germany).

**Sequence analysis of ‘low-copy’ BAC subclones**

Six BAC subclones that were successfully localized by FISH to single loci on the short arm of chromosome 1B and two that gave hybridization patterns dispersed on all chromosomes were randomly chosen for sequencing. The sequencing was performed by Agowa (Berlin, Germany). The sequence parts of the subclones were assembled in the Chromas software (Griffith University, Southport, Qld, Australia), edited using BioEdit (http://www.mbio.ncsu.edu/BioEdit/bioedit.html), and the assembled sequences were deposited in GenBank (GSS database). They were then searched for similarities to known sequences in GenBank using BLASTN (Altschul et al., 1997), and tested for the presence of repetitive units using JDotter (Brodie et al., 2003) and Tandem Repeat Finder (Benson, 1999). Finally, the sequences of subclones were used to design primers for internal sequence amplification to verify their specificity for 1BS. PCR was performed as described by Vrána et al. (2000) on flow-sorted chromosome arm 1BS. BAC DNA, genomic DNA of a di-telosomic line of cv. ‘Pavon 76’ carrying the long arm of 1B (Dt1BL), and genomic DNA of cv. ‘Pavon 76’.

**Acknowledgements**

We are grateful to our colleagues Arnaud Bellec, Michaela Libosvárová, Radka Tušková and Jitka Weiserová for excellent technical assistance. This work was supported by research grants from the Czech Science Foundation (grant awards 521/04/0607, 521/05/P257 and 521/05/H013), the Ministry of Education, Youth and Sports of the Czech Republic (grant award LC06004), research grants from the French Ministry of Research and Agriculture, and by a collaborative project ‘Barrande’ (2003-035-2) between the Ministry of Education, Youth and Sports of the Czech Republic and the French Ministry of Foreign Affairs and Education. J.J. acknowledges the receipt of a Marie Curie PhD student Fellowship at URGV-INRA.

**References**


Genbank GSS database accession numbers DX572363-DX572372 (low-copy BAC subclones).