Rapid identification of the three homoeologues of the wheat dwarfing gene *Rht* using a novel PCR-based screen of three-dimensional BAC pools

Melanie Febrer, Ed Wilhelm, Nadia Al-Kaff, Jonathan Wright, Wayne Powell, Michael W. Bevan, and Margaret I. Boulton

Abstract: A high-throughput two-step PCR strategy for the identification of selected genes from a BAC library derived from hexaploid wheat (16 974 Mbp) is described. The screen is based on the pooling of DNA from BAC clones into 675 "superpools" arrayed in a three-dimensional configuration. Each BAC clone is represented in three superpools to allow the identification of candidate 384-well plates of clones after the first round of PCR; identification is facilitated by an associated Perl script. A second round of PCR detects the specific BAC clone within the candidate plate that corresponds to the gene of interest. Thus, a single copy of the target gene can be identified from the library of over 700 000 clones (approximately 5 genome equivalents) by assaying only three 384-well plates. The pooling strategy was validated by screening the library with primers specific for the reduced height (*Rht-1a*) gene. Using relatively stringent selection criteria, 13 *Rht*-containing clones were identified from 17 candidate plates, and sequence analysis of the amplified products showed that all three *Rht* homoeologues were represented. Furthermore, the method confirmed the estimated coverage of the BAC library. Thus, this methodology allows the rapid and cost-effective identification of genes, and their homoeologues, from large-insert libraries of complex genomes such as hexaploid wheat.

Key words: hexaploid wheat, BAC library, *Rht* gene, library screening.

Résumé : Une stratégie de criblage PCR à haut débit en deux étapes pour l'identification de gènes ciblés au sein d'une banque de clones BAC, issue du blé hexaploïde (16 974 Mb), est décrite. Le crible repose sur des ADN composites provenant de 675 « superpools » de clones BAC disposés dans une configuration en trois dimensions. Chaque clone BAC est représenté dans trois « superpools », ce qui permet d'identifier des plaques à 384 puits candidates après une première amplification PCR. Cette identification est facilitée par l'emploi d'un script Perl. Une seconde ronde d'amplification PCR permet de détecter un clone BAC spécifique qui contient le gène d'intérêt au sein de la plaque candidate. Ainsi, on peut identifier une copie unique du gène ciblé au sein d'une banque de plus de 700 000 clones (l'équivalent d'environ 5 génomes) en interrogeant seulement trois plaques à 384 puits. Cette stratégie a été validée en criblant la banque à l'aide d'amorces spécifiques pour le gène reduced height (*Rht-1a*). À l'aide de critères relativement stricts, 13 clones contenant *Rht* ont été identifiés parmi 17 plaques candidates et le séquençage des produits amplifiés a montré que les trois homéologues *Rht* étaient représentés. De plus, la méthode a permis de confirmer la couverture estimée de la banque de clones BAC. Ainsi, cette méthodologie permet d'identifier rapidement et à peu de frais les gènes et leurs homéologues au sein de banques d'inserts de grande taille pour des génomes complexes comme celui du blé hexaploïde.

Mots-clés : blé hexaploïde, banque de clones BAC, gène *Rht*, criblage de banque.

[Traduit par la Rédaction]

Introduction

Producing sufficient healthy food and feed during a period of increasing population and the threat of global climate change is of paramount importance. Cereals such as rice, wheat, sorghum, and maize are staple foods world-wide, and the improvement of these crops is consequently a high priority. Genome sequencing facilitates crop improvement by identifying markers for breeding and candidate genes for transgenic approaches. Although bread wheat (*Triticum aestivum* L.) is the most important food crop world-wide after rice, its complex genome is currently a barrier to genomic

Received 12 August 2009. Accepted 9 September 2009. Published on the NRC Research Press Web site at genome.nrc.ca on 17 November 2009.

Corresponding Editor: P. Donini.

M. Febrer, N. Al-Kaff, J. Wright, M.W. Bevan,[1] and M.I. Boulton. John Innes Centre, Norwich Research Park, Colney, Norwich, NR4 7UH, UK.
E. Wilhelm. John Innes Centre, Norwich Research Park, Colney, Norwich, NR4 7UH, UK; National Institute of Agricultural Botany, Huntingdon Road, Cambridge, CB3 0LE, UK.
W. Powell. National Institute of Agricultural Botany, Huntingdon Road, Cambridge, CB3 0LE, UK.

[1]Corresponding author (e-mail: michael.bevan@bbsrc.ac.uk).

analysis. Bread wheat has one of the largest and most complex genomes of all crop species; it is an allohexaploid species containing 3 genomes (AA, BB, and DD), each of which has 7 pairs of chromosomes. Its ≥16 000 Mbp genome (Arumuganathan and Earle 1991), consisting of approximately 80% repetitive (noncoding) DNA sequences interspersed amongst low-copy-number or singleton genes (Francki and Appels 2002), requires novel resources and techniques to facilitate gene identification and polymorphism analysis.

Large-insert libraries are essential tools for physical mapping, positional cloning, and genome sequencing. The bacterial artificial chromosome (BAC) cloning system (Shizuya et al. 1992) has become the favoured method to clone large genomic DNA inserts (Xia et al. 2005) because of the high transformation efficiency of *Escherichia coli* and the stability of inserts, coupled with the ease of BAC DNA purification and its compatibility with sequence determination protocols. Several BAC libraries have been constructed with genomic DNA from bread wheat and its close relatives, including *Aegilops tauschii* (diploid; D genome (Moullet et al. 1999)) and *Triticum monococcum* (diploid; A genome (Lijavetzky et al. 1999)), and the tetraploid *Triticum turgidum* (AB genome (Cenci et al. 2003)). For *T. aestivum*, a bulked BAC library was reported for genotype Hartog (Ma et al. 2000) and gridded BAC libraries were reported for genotypes Renan (Chalhoub et al. 2002), Glenlea (Nilmalgoda et al. 2003), and Chinese Spring (Allouis et al. 2003).

A high redundancy of genome coverage and a large insert size are important if a library is to be used for physical mapping, but the resulting large size of the library, concomitant with the large genome, make screening a wheat BAC library a laborious and time-consuming procedure. Until recently, hybridization was used to screen high-density filters but this procedure is cumbersome, requires the use of radioactive probes, and may lack specificity. Recently, PCR-based screening has provided a substitute for hybridization-based screens because it is simple, fast, and sensitive (Klein et al. 2000; Yim et al. 2007). The increasing efficiency of PCR-based techniques relies upon improved primer design and thermal cyclers and, most importantly, the design of multidimensional clone pooling strategies for the efficient, sensitive, and specific detection of individual BAC clones (Barillot et al. 1991; Bruno et al. 1995). Once pooled, candidate clones may be located by identifying the subset of pools containing the corresponding markers (Klein et al. 2000; Yim et al. 2007). Preparation of BAC pools in three or more dimensions is a requirement for high-throughput PCR-based screening of large libraries. Recently, Farrar and Donnison (2007) described the technique for screening a BAC library for *Brachypodium distachyon* using 343 96-well microtitre plates and a three-dimensional (3-D) 7 × 7 × 7 microtitre plate grid.

Here we describe the first 3-D pooling strategy for a wheat BAC library that contains over 700 000 clones. The suitability of the pooling system for gene discovery was validated by PCR screening of the library for clones containing the single-copy reduced height (*Rht-1a*) gene that contributes to the increased yields of the green revolution (Peng et al. 1999). Sequence is available for two of the three *Rht-1a* homoeologues (Peng et al. 1999; N. Harberd, University of Oxford, unpublished data), and sequence analysis of the amplification products confirmed the predicted genome coverage of the library and the validity of the technique for the isolation of homoeologous genes.

Material and methods

Wheat BAC library

Part of the wheat BAC library constructed by Allouis et al. (2003) for cultivar Chinese Spring was used in this study. The library was constructed in a collaboration between BBSRC (UK) and INRA (France) and comprises 1 200 000 clones (3125 384-well plates) with an average insert size of 130 kb. For this study, we used the UK component of the library, which contains 715 776 clones (1864 384-well plates), representing 5 haploid genome equivalents.

BAC library screening

Pooling strategy

To facilitate screening of the wheat library, a pooling strategy based on a cube design similar to that described by Farrar and Donnison (2007) was used. However, the large size of the wheat library required 384-well plates and four 3-D microtitre plate grids to be used. Thus, two 8 × 8 × 8 cubes (each containing 512 384-well plates), one 9 × 9 × 9 cube (729 plates), and one 4 × 4 × 4 (64 plates) cube were created. A diagram of the pooling procedure (4 × 4 × 4 cube) is shown in Fig. 1.

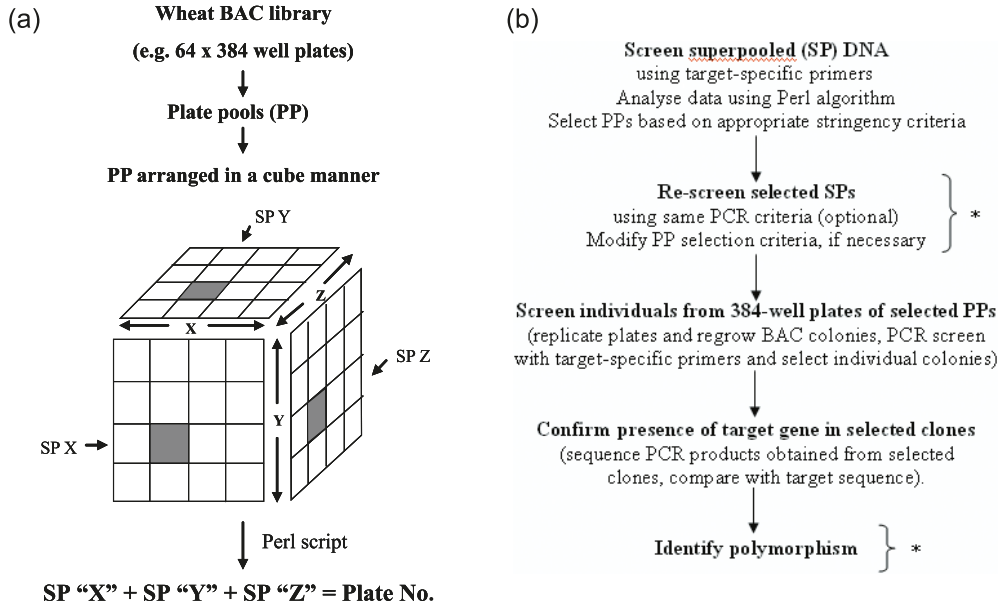
For each grid, each plate was replicated onto a 140 mm triple-vent Petri dish containing 100 mL of LB agar and 12.5 $\mu g \cdot mL^{-1}$ chloramphenicol. After incubation overnight at 37 °C, the colonies were collected by the addition of 40 mL of LB broth and the resulting culture was transferred to a 50 mL tube to give a plate pool (PP). The PPs for each grid were displayed in a virtual cube (*X*, *Y*, and *Z* axes). Superpools (SPs) were produced by combining 5 mL of culture from each of the PPs so that each plate was represented in three different pools. The 675 SPs were produced essentially as described by Farrar and Donnison (2007), and the correlation of PP and SP positions is described therein.

Validation of the pooling procedure: screening BAC pools for target sequences

Plasmid DNA was extracted from bacterial pellets of each of the SP cultures using the alkaline-lysis method (Sambrook et al. 1989). Buffers P1, P2, and P3 (QIAGEN Ltd.) were used to extract the BAC DNA, with volumes being determined by the volume of the overnight culture. The DNA was precipitated by addition of ice-cold isopropanol and resuspended in 200 μL sterile water. After quantification, a 90 $ng \cdot \mu L^{-1}$ working stock was used for further analyses.

To screen the BAC library for the presence of *Rht-1a*-containing clones, a region known to contain sequence polymorphism between the *Rht-B1a* and *Rht-D1a* alleles of Chinese Spring (D genome, GenBank accession No. AJ242531 (Peng et al. 1999); B genome, N. Harberd, unpublished data) was amplified using PCR. The primers were chosen to produce a 527 bp product (based on the *Rht-D1a* sequence) that encompasses the region conserved in DELLA protein-encoding genes. For the first round of PCR, SPs were screened using a 10 μL PCR mix consisting of 90 ng of BAC SP DNA, 1× Green GoTaq Reaction Buffer

Fig. 1. Representation of the 3-D pooling strategy for PCR screening of the BAC library (adapted from Farrar and Donnison 2007). (*a*) A virtual 4 × 4 × 4 cube which is sufficient for assessing the contents of 64 plates of the library. Each plate is replicated and all 384 clones of the replicated plate are combined to form a plate pool (PP). Each PP is pooled in each of three dimensions (*X*, *Y*, *Z*) such that, for a 4 × 4 × 4 cube, four PPs are present in each superpool (SP). For the 8 × 8 × 8 and 9 × 9 × 9 cubes, eight and nine PPs are present in each SP, respectively. The grey squares depict the plate that is represented in three SPs (*X*, *Y*, and *Z*). (*b*) A resume of the steps involved in identifying target genes from the SP DNA. *, optional steps.



(Promega), 3% glycerol, 0.2 mmol·L^{-1} of each dNTP, 2 mmol·L^{-1} $MgCl_2$, 1 μmol·L^{-1} of each of the gene-specific primer pairs (Rht F1, 5′-GGACACCGTGCACTACAACC-3′; Rht R1, 5′-CCAGCAAGGGTATCTGCTTC-3′), and 0.25 units of *Taq* polymerase. The glycerol was added to facilitate amplification of the GC-rich *Rht-1a* sequence. The PCR consisted of initial denaturation for 5 min at 95 °C, followed by 40 cycles of 95 °C for 30 s, 58 °C for 30 s, and 72 °C for 45 s. Amplification products were separated on a 1.5% agarose gel in 1× Tris – acetic acid – EDTA buffer and visualized using ethidium bromide. To monitor the reliability of the amplification of the BAC DNA, several controls were used: NC1, in which water replaced DNA in the reaction; NC2 and PC1, miniprep DNA of BAC clones known not to contain *Rht-1a* or known to contain *Rht-1a*, respectively; PC2, genomic DNA isolated from leaves of Chinese Spring seedlings; and PC3, a 1:3072 (*v*/*v*) dilution of PC1:NC2. PC3 was established to mimic a SP from an 8 × 8 × 8 cube that included a single BAC clone containing *Rht-1a*. The BAC DNA concentration was 90 ng per reaction with the exception of PC1, where 20 ng was used. The SP DNA samples were scored for the presence and intensity (clear, strong (++) or weak (+)) of an approximately 530 bp amplification product. As DNA from each PP is represented in each of three SPs, identification of the candidate plates likely to harbour an *Rht-1a*-containing clone was facilitated by the use of a custom Perl script. Once the candidate plates were selected, the corresponding SP DNAs were subjected to a second round of amplification to validate the robustness of the PCR screening.

To detect the individual BAC clones containing the *Rht-1a* gene, the selected 384-well plates were screened by colony PCR using the same PCR conditions except that 1 μL of cells was collected directly from the plates to serve as the template and initial denaturation was for 10 min. To allow further characterization, DNA was isolated, as described above, from BAC clones that produced a *Rht-1a*-specific product. Cells were grown by shaking overnight at 37 °C in LB broth containing chloramphenicol (12.5 μg·mL^{-1}).

Characterization of the putative *Rht-1a*-containing BAC clones

Fingerprint analysis

To determine whether multiple copies of a single clone had been selected, DNA from each of the identified BAC clones was subjected to fingerprint analysis by pulse field gel electrophoresis as described by Allouis et al. (2003), except that samples were digested overnight with *Hin*dIII, electrophoresis was carried out for 24 h at 4 °C, and the DNA was visualized using CyberGold (Molecular Probes). Insert size was estimated by comparison with molecular weight markers (HyperLadder I, Bioline; kb ladder, BRL).

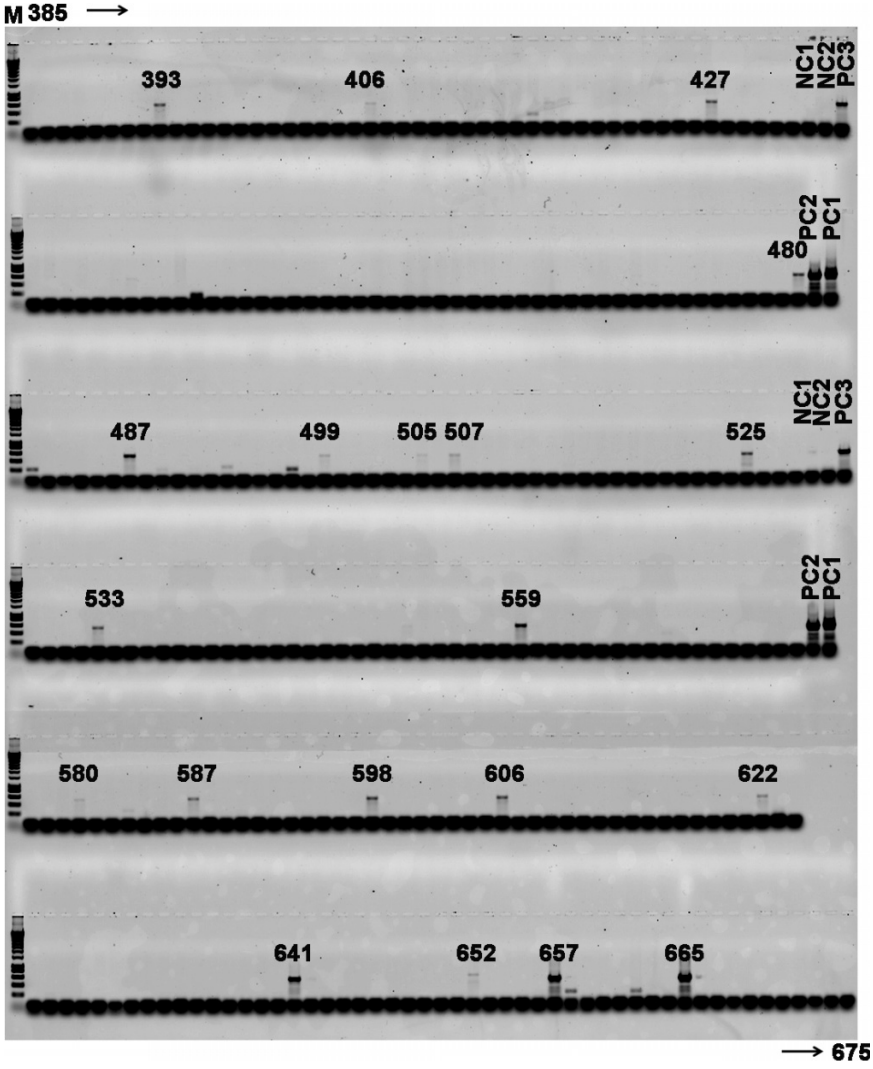
Sequence analysis of the amplified products

DNA of BAC minipreps or BAC colonies was PCR-amplified as described above, and 1 μL of the reaction was used in a 10 μL BigDye (Applied Biosystems) cycle sequencing reaction that also contained 1 μmol·L^{-1} of primer Rht F1 or Rht R1. The sequencing reaction consisted of 96 °C for 1 min followed by 25 cycles of 96 °C for 10 s, 50 °C for 5 s, and 60 °C for 4 min. A forward and reverse read of each amplified product was performed to identify potential PCR-induced errors and to confirm nucleotide polymorphisms. Sequencing was performed by the John Innes Centre Genome Sequencing Facility. Sequences

Table 1. Organisation of the wheat (Chinese Spring) BAC library in the different microtitre plate grids.

	Microtitre grid				
	1	2	3	4	Total
Grid type	8×8×8	8×8×8	9×9×9	4×4×4	
Number of plates	512	512	729	64	1817
Number of superpools	192	192	243	48	675
DNA range (ng/μL)	91–3747	436–5844	91–4235	2071–5047	91–5844
DNA mean (ng/μL)	1013	4164	3521	1484	2545

Fig. 2. PCR screening of BAC superpools (SPs) 385–675 using *Rht-1a*-specific primers. The numbers of the SPs from which an approximately 530 bp PCR product was obtained are shown. Controls used to confirm the specificity of the PCR are shown on the right-hand side of the gel: NC1, water replaces template DNA; NC2, DNA from a BAC clone known not to contain *Rht-1a*; PC1, DNA from a BAC clone known to contain *Rht-1a*; PC2, Chinese Spring genomic DNA; PC3, a 1:3072 dilution of PC1:NC2, designed to mimic a SP that contained one *Rht-1a*-containing BAC clone. M, molecular weight marker (kb ladder, BRL).



were aligned using the ContigExpress software of VectorNTI (Invitrogen).

Results and discussion

Pooling strategy and library screening

To allow efficient PCR screening of the >700 000 clones that comprise the 5 genome equivalent coverage of the UK-produced part of the wheat (Chinese Spring) BAC library (Allouis et al. 2003), a pooling strategy was designed. The cube design allowed 1817 of the 1864 384-well plates of the library to be screened; the organisation of the microtitre grids is shown in Table 1. Each microtitre grid was configured in three dimensions such that each plate of the library

Table 2. *Rht-1a* superpool (SP) screen summary.

	Rht-1a superpool screen[a]						
	Original score			Retest score			
Plate No.[b]	SP1	SP2	SP3	SP1	SP2	SP3	Results[c]
141	++	–	++	++	–	+	√
143	++	++	–	++	+	–	X
167	++	++	–	++	+	+	√
224	++	+	–	+	++	–	√
305	++	++	+	++	++	–	√
679	ND	+	++	ND	++	++	√
967	++	+	+	++	++	+	√
1003	++	+	+	++	–	–	X
1006	++	++	++	+	++	++	√
1297	–	++	++	+	++	–	√
1315	+	++	–	++	++	–	X
1318	+	–	–	++	+	++	√
1399	–	–	+	–	++	++	X
1417	–	+	++	–	++	++	√
1506	++	+	++	–	++	++	√
1668	–	++	++	–	+	++	X
1730	–	++	++	+	++	++	√
1905	++	++	++	++	++	++	√

[a]The intensity of the approximately 530 bp PCR product obtained from SPs was assessed visually. ++, strong band; +, weak band; –, no band; ND, no DNA available.

[b]The corresponding 384-well plates (each plate is represented by 3 SPs) were selected for further *Rht-1a* screening if at least two SPs scored ++ or 3 SPs scored + or ++ in either the original or retest PCR.

[c]The 384-well plates were marked as "positive" (√) if one or more clones gave a *Rht-1a*-specific PCR product; plates in which no clones gave a PCR product were designated X.

was present in 3 superpools (SPs). Thus, 675 SPs, represented in 4 grids, cover the 1817 plates of the BAC library and PCR screening of the SP DNA can be carried out with only two 384-well plates. The concentration of the DNA extracted from each SP varied (minimum 91 ng·μL^{-1}, maximum 5844 ng·μL^{-1}, mean 2545 ng·μL^{-1}; Table 1) and no attempt was made to equalize the concentrations.

After PCR-based identification of SPs containing DNA from the clones of interest, it is necessary to deconvolute the data to identify which of the 1817 384-well plates contain clones specifying the gene of interest. Previous studies have approached this either by reference to a tabulated matrix (Farrar and Donnison 2007) or by use of a Perl script (Klein et al. 2000). As the former method is arduous and time-consuming, a Perl script was written. The script loads the coordinates of all 384-well plates into a two-dimensional array and iterates through the array until it finds a plate with a coordinate matching one in a comma-separated list of coordinates supplied by the user. The other two addresses of the plate are then checked for matches. At this point the user may elect to screen only those candidate plates containing three matching coordinates, which will decrease the number of false positives, or to select plates containing two out of three matching coordinates. The latter option will result in fewer "missed" BACs but potentially more false positives, and an increased number of samples in the second round of PCR screening. In this BAC library, which contains approximately 5 haploid genome equivalent coverage of the hexaploid genome, a single-copy gene is predicted to occur in 15 plates. Thus, rigorous screening for a candidate single-copy gene should require an average of 17 × 384 PCRs (2 PCRs for the SP screen followed by 15 PCRs for the plate screen).

Validation of the screening method

PCR screening

Validation of the method was carried out by PCR screening of the BAC library with a set of primers that were designed to perfectly match the *Rht-B1a* and *Rht-D1a* alleles of Chinese Spring. Prior to screening the BAC DNA, PCR was carried out to confirm the specificity of the primers and to determine whether a single BAC clone containing *Rht-1a* could be detected when present in a SP from an 8 × 8 × 8 cube (1:3072 (*v*/*v*) dilution of BAC DNA from a previously identified *Rht-D1a*-containing clone : 3071 parts DNA from a *Rht-1a*-negative BAC clone). The expected 527 bp amplification product was obtained, and no amplification was obtained with DNA from the *Rht-1a*-negative clone (data not shown, but see control lanes PC3 and NC2 in Fig. 2), thereby confirming the specificity and the sensitivity of the PCR screening procedure.

The first round of PCR (SP screening) identified SPs from which a clear, approximately 530 bp amplification product was obtained. Deconvolution of these data showed that the gene of interest, *Rht-1a*, was potentially present in 36 384-well plates. An example of the amplification from SPs 385–675 is shown in Fig. 2. To further test the robustness of the PCR screen, the SPs containing the "candidate" plates were subjected to a second PCR. For all PCRs, the intensity of

Fig. 3. Alignment of the nucleotide sequences obtained from *Rht-1a*-specific products amplified from 13 BAC clones. Designation of the homoeologues is based on the known sequences for *Rht-D1a* (GenBank accession No. AJ242531) and *Rht-B1a*; the designation of *Rht-A1a* was inferred. The primer sequences used for amplification are shown in bold text; nucleotide polymorphisms compared with the *Rht-D1a* accession sequence are highlighted in grey. The coordinates shown at the beginning and end of the sequence are those for *Rht-D1a*.

```
            264
AJ242531    GGACACCGTG CACTACAACC CCACCGACCT GTCGTCTTGG GTCGAGAGCA TGCTGTCGGA GCTCAACGCG CCGCCGCCGC CCCTCCCGCC
Rht-B1a     GGACACCGTG CACTACAACC CCACCGACCT CTCCTCCTGG GTCGAGAGCA TGCTGTCGGA GCTCAACGCG CCGCCGCCGC CCCTCCCGCC
Rht-A1a     GGACACCGTG CACTACAACC CCACCGACCT CTCCTCCTGG GTCGAGAGCA TGCTGTCGGA GCTCAACGCG CCGCCGCCGC CCCTCCCGCC

AJ242531    CGCCCCGCAG C...TCAACG CCTCCACCTC CTCCACCGTC ACGGGCAGCG GCGGCTACTT CGATCTCCCG CCCTCCGTCG ACTCCTCCAG
Rht-B1a     CGCCCCGCAG C...TCAACG CCTCCACCTC CTCCACCGTC ACCGGCCGCG GG...TACTT CGATCTCCCG CCCTCCGTCG ACTCCTCCTG
Rht-A1a     CGCCCCGCAG CAGCTCAACG CCTCCACCTC CTCCACCGTC ACGGGCGGTG GG...TACTT CGATCTCCCG CCCTCGGTTG ACTCCTCCTG

AJ242531    CAGCATCTAC GCGCTGCGGC CGATCCCCTC CCCGGCCGGC GCGACGGCGC CGGCCGACCT GTCCGCCGAC TCCGT...GC GGGATCCCAA
Rht-B1a     CAGCACCTAC GCGCTGCGGC CGATCCCGTC CCCGGCCGTC GCGCCGG... ...CCGACCT CTCCGCCGAC TCCGTCGTGC GGGATCCCAA
Rht-A1a     CAGCAGCTAC GCCCTGCGGC CGATCCCCTC CCCGGCCGGC GCCGTCGGGC CGGCCGACCT GTCCGCCGAC TCCGT...GC GGGACCCCAA

AJ242531    GCGGATGCGC ACTGGCGGGA GCAGCACCTC GTCGTCATCC TCCTCCTCGT CGTCTCTCGG .TGGGGGCGC CAGGAGCTCT GTGGTGGAGG
Rht-B1a     GCGGATGCGC ACTGGCGGGA GCAGCACCTC GTCGTCATCC TCATCGTCCT C.TCTCGGCG GTGGCGGCGC CAGGAGCTCT GTGGTGGAGG
Rht-A1a     GCGGATGCGC ACTGGCGGGA GCAGCACCTC GTCGTCGTCA TCCTCCTCGT CCTCTCTCGG .TGGGGGCGC CAGGAGCTCT GTGGTGGAGG

AJ242531    CTGCCCCGCC GGTCGCGGCC GCGGCCAACG CGACGCCCGC GCTGCCGGTC GTCGTGGTCG ACACGCAGGA GGCCGGGATT CGGCTGGTGC
Rht-B1a     CTGCCCCGCC GGTGGCCGCC GCGGCC...G GTGCGCCCGC GCTGCCGGTC GTCGTGGTCG ACACGCAGGA GGCCGGGATT CGGCTGGTGC
Rht-A1a     CTGCTCCGCC GGTCGCGGCC GGGGCCAACG ...CGCCCGC GCTGCCGGTC GTCGTGGTCG ACACGCAGGA GGCCGGGATT CGGCTGGTGC

AJ242531    ACGCGCTGCT GGCGTGCGCG GAGGCCGTGC AGCAGGAGAA CCTCTCCGCC GCGGAGGCGC TGGTGAAGCA GATACCCTTG CTGG 791
Rht-B1a     ACGCGCTGCT GGCGTGCGCA GAGGCCGTGC AGCAGGAGAA CTTCTCTGCC GCGGAGGCGC TGGTGAAGCA GATACCCTTG CTGG
Rht-A1a     ACGCGCTGCT GGCCTGCGCG GAGGCCGTGC AGCAGGAGAA CTTCTCTGCC GCGGAGGCGC TGGTGAAGCA GATACCCTTG CTGG
```

the amplification product band was scored. The 384-well plates selected for screening of individual BAC clones were those for which the diagnostic PCR product was identified in all three matching SPs or a clear, strong band was identified in at least two of the SPs in either the original or repeated assay. Thus, 17 "best candidate" 384-well plates were selected. An additional plate (224, Table 2) was also chosen to determine the likely extent of false negatives ("missed" clones) if only one round of screening was performed or if plates with a single SP producing a strong PCR product were not selected. For plate 224, in each of the first and repeated assays, only one SP produced a strong band, whereas a second SP produced a low-intensity band and the third SP did not produce a band (Table 2).

The 18 candidate plates were replicated for the second round of PCR to identify the address corresponding to the *Rht-1a*-containing clone. Following colony PCR, amplification products were obtained from 13 clones (Table 2), including a clone present in plate 224, thereby indicating that some *Rht-1a*-containing clones may have been missed using our relatively stringent selection criteria.

To further assess our selection criteria, the PCR data obtained from SPs containing the 18 plates (Table 2) selected for further testing were analysed. This revealed that if only the first SP screen had been carried out, and only those plates for which the diagnostic PCR product was obtained in all three matching SPs or a strong, clear band was obtained from two of the three SPs had been selected, 12 plates would have been regarded as potential candidates for further screening. Of these, 9 would have yielded a positive clone in the final round of PCR. In the repeat screening, 12 plates would also have been selected. Although some of the plates selected would have differed, 10 would have yielded a positive clone in the final round of PCR (Table 2). The discrepancy between the data obtained in the two screens probably results from the difficulty of amplifying the GC-rich *Rht-1a* sequence and because the amplification conditions were not fully optimized, but despite this the number of positive BACs identified would have been very similar (9 or 10, first or second screen, respectively). For more easily amplified genes, or if automated procedures are used, such a discrepancy may not be seen. If candidate plates had been selected only when strong amplification for three matching coordinates was obtained, only two plates would have been selected in each of the duplicated screens, but both would have yielded a clone from which a *Rht-1a* amplification product was produced. Thus, if the aim is to identify only one clone of a single-copy gene, the stringency of the selection parameters may be increased, thereby decreasing the total number of PCRs. In the current study, where the aim was to identify most (but not necessarily all) of the *Rht-1a*-containing clones (to verify the theoretical genome coverage of the library and to validate the screening procedure), the stringency of selection was relaxed, although this was compensated by a confirmatory duplicate testing of the SP DNA, since this is the critical point at which to limit the final number of PCR tests. It is possible that we have underestimated the number of clones that contain *Rht-1a* (that is, we have designated them falsely as "negative" clones), given that there were additional candidate plates that were not selected because the re-screen did not produce clear specific PCR products from two of the SPs. However, because we had identified 13 positive clones (compared with a predicted 15, based on the genome redundancy of the library), we did not attempt an additional second round of PCR screening of the additional unselected plates or the other 4 plates that were identified as potentially having *Rht-1a*-containing clones.

Characterization of the selected clones

Sequence analysis of the amplification products from each of the 13 selected clones showed that 7 were identical to the published sequence for *Rht-D1a* and 4 were identical to the sequence of *Rht-B1a*. The remaining 2 clones showed high

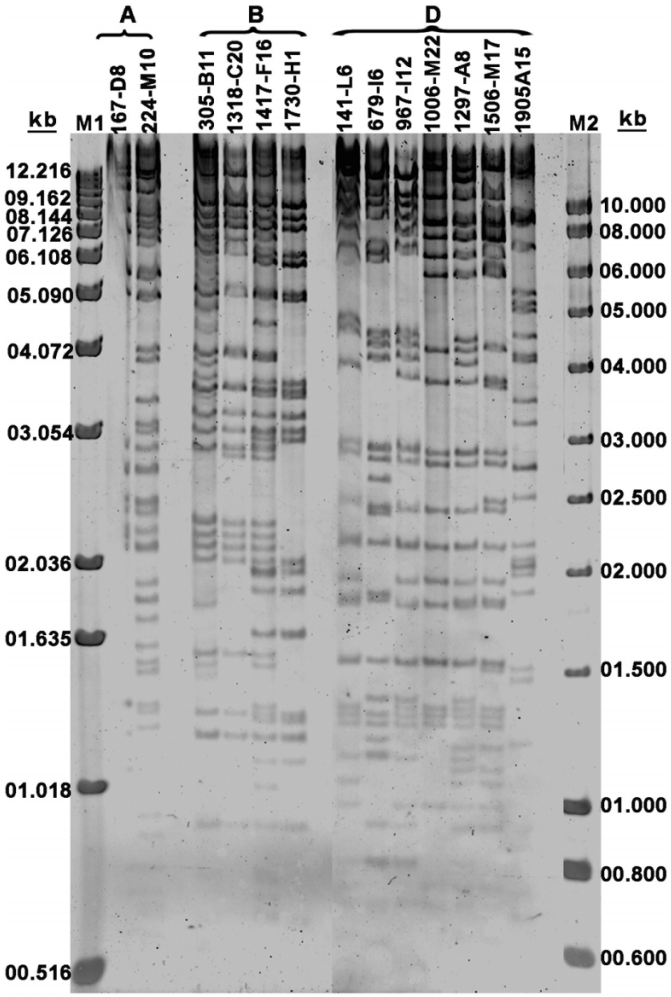


Fig. 4. Gel electrophoretic analysis of DNA extracted from BAC clones from which *Rht-1a* sequence was obtained. DNA from these clones was digested with *Hin*dIII and the gel was stained with CyberGold. Molecular weight markers (M1 and M2) were kb ladder (BRL) and HyperLadder 1 (Bioline), respectively. The analysis was done to determine whether all clones were unique, rather than to size the clones, but estimation of their size by comparison with the size markers revealed that insert sizes are between 70 and 200 kb. Clones were designated according to their plate number and their position in the 384-well plate (e.g., clone 224-M10 is present in plate 224, at position M10). The genome derivation (A, B, or D) of the clones is also shown above their designations.

sequence homology to these two *Rht-1a* homoeologues, but contained several polymorphisms and are therefore likely to be the *Rht-A1a* homoeologue (Fig. 3). Further sequence analysis of clone 224-M10 (not shown) identified the presence of the conserved DELLA domain (Peng et al. 1999), confirming the *Rht-A1a* identity of these clones. Thus, all 13 clones from which an amplification product was obtained contain *Rht-1a*, thereby validating the pooling and screening strategy, even for genes that are recalcitrant to PCR amplification. Furthermore, our data show that if only the first round of SP screening had been carried out, all three *Rht-1a* homoeologues would have been represented in the 9 *Rht-1a*-containing clones (1 A, 2 B, and 6 D genome-derived clones were identified). A similar result would have been obtained in the repeated screen, where 1 A, 3 B, and 6 D genome-derived clones would have been identified.

The identification of homoeologous genes is not possible using hybridization-based screening of BAC libraries, but even with the PCR screening technique it depends upon the design of appropriate primer pairs. In this study the primers were designed to a region that was conserved between the two known homoeologues. Primer design for identification of multiple-copy genes or members of gene families will be more demanding and this step will be integral to the success of the screen. The Rht protein is a member of the DELLA clade of the GRAS domain family of transcription factors (Pysh et al. 1999; Bolle 2004), and for this study primers were designed to nucleotides that encode the N-terminal domain of the protein, which is specific to members of the DELLA clade.

The quality of a BAC library depends not only on the number of clones and their insert size but also on the presence of a high percentage of unique clones. To determine whether all of the selected BAC clones were unique, they were subjected to fingerprint analysis. Although the clones containing a specific homoeologue had bands in common, each of the 13 clones produced a distinctive fingerprint (Fig. 4), and the size of the BAC clones was in agreement with the insert sizes reported for this library (Allouis et al. 2003), thereby further confirming the quality of the library. Estimation of the size of the BAC clones and the position of the candidate gene within these clones provides valuable information to inform further sequence analysis. Our studies show that this library will provide a key resource for systematic physical mapping of the wheat genome (Moolhuijzen et al. 2007) and for gene discovery (Griffiths et al. 2006).

Conclusions

We have identified all three homoeologous copies of a single gene (here *Rht-1a*) in the large polyploid wheat genome using a rapid PCR-based screen of pooled BAC clones. The entire PCR screening procedure and sequence analysis was completed, without automation, by 2 people in approximately 2 weeks. Thus we have demonstrated the coverage and quality of the BAC library and the utility of the 3-D pooling strategy for rapidly isolating target genes and obtaining the single nucleotide polymorphism information. Our research will facilitate the large-scale production of molecular markers and analysis of homoeologue-specific gene expression in wheat, one of the world's most important food crops.

Acknowledgements

We would like to thank Prof. Nicholas Harberd, University of Oxford, UK, for providing unpublished sequence of *Rht-B1a*, and Tracie Foote, Simon Griffiths, and Graham Moore for providing the BAC library. This work was supported by a grant in aid from the Biotechnology and Biological Sciences Research Council (BBSRC) (MIB, BB/E004725/1 to M.B.) and by the National Institute of Agricultural Botany (NIAB) Trust (E.W., W.P.).

References

Allouis, S., Moore, G., Bellec, A., Sharp, R., Faivre Rampant, P., Mortimer, K., et al. 2003. Construction and characterisation of a hexaploid wheat (*Triticum aestivum* L.) BAC library from the reference germplasm 'Chinese Spring'. Cereal Res. Commun. **31**(3–4): 331–338.

Arumuganathan, K., and Earle, E.D. 1991. Nuclear DNA content of some important plant species. Plant Mol. Biol. Rep. **9**(3): 208–218. doi:10.1007/BF02672069.

Barillot, E., Lacroix, B., and Cohen, D. 1991. Theoretical analysis of library screening using a N-dimensional pooling strategy. Nucleic Acids Res. **19**(22): 6241–6247. doi:10.1093/nar/19.22.6241. PMID:1956784.

Bolle, C. 2004. The role of GRAS proteins in plant signal transduction and development. Planta, **218**(5): 683–692. doi:10.1007/s00425-004-1203-z. PMID:14760535.

Bruno, W.J., Knill, E., Balding, D.J., Bruce, D.C., Doggett, N.A., Sawhill, W.W., et al. 1995. Efficient pooling designs for library screening. Genomics, **26**(1): 21–30. doi:10.1016/0888-7543(95)80078-Z. PMID:7782082.

Cenci, A., Chantret, N., Kong, X., Gu, Y., Anderson, O.D., Fahima, T., et al. 2003. Construction and characterization of a half million clone BAC library of durum wheat (*Triticum turgidum* ssp. *durum*). Theor. Appl. Genet. **107**(5): 931–939. doi:10.1007/s00122-003-1331-z. PMID:12830387.

Chalhoub, B., Lefevre, A., Rouault, P., Pateyron, S., Bellec, A., Laffont, P., et al. 2002. Construction of an hexaploid wheat BAC library for studying genome organization and cloning of interesting genes and QTLs. *Presented at* Plant, Animal & Microbe Genomes X Conference, San Diego, Calif., 12–16 January 2002. Poster 350.

Farrar, K., and Donnison, I.S. 2007. Construction and screening of BAC libraries made from *Brachypodium* genomic DNA. Nat. Protoc. **2**(7): 1661–1674. doi:10.1038/nprot.2007.204. PMID:17641631.

Francki, M., and Appels, R. 2002. Wheat functional genomics and engineering crop improvement. Genome Biol. **3**(5): reviews1013.1–1013.5. doi:10.1186/gb-2002-3-5-reviews1013. PMID:12049670.

Griffiths, S., Sharp, R., Foote, T.N., Bertin, I., Wanous, M., Reader, S., et al. 2006. Molecular characterization of *Ph1* as a major chromosome pairing locus in polyploid wheat. Nature, **439**(7077): 749–752. doi:10.1038/nature04434. PMID:16467840.

Klein, P.E., Klein, R.R., Cartinhour, S.W., Ulanch, P.E., Dong, J., Obert, J.A., et al. 2000. A high-throughput AFLP-based method for constructing integrated genetic and physical maps: progress toward a sorghum genome map. Genome Res. **10**(6): 789–807. doi:10.1101/gr.10.6.789. PMID:10854411.

Lijavetzky, D., Muzzi, G., Wicker, T., Keller, B., Wing, R., and Dubcovsky, J. 1999. Construction and characterization of a bacterial artificial chromosome (BAC) library for the A genome of wheat. Genome, **42**(6): 1176–1182. doi:10.1139/gen-42-6-1176. PMID:10659785.

Ma, Z., Weining, S., Sharp, P.J., and Liu, C. 2000. Non-gridded library: a new approach for BAC (bacterial artificial chromosome) exploitation in hexaploid wheat (*Triticum aestivum*). Nucleic Acids Res. **28**(24): e106. doi:10.1093/nar/28.24.e106. PMID:11121493.

Moolhuijzen, P., Dunn, D.S., Bellgard, M., Carter, M., Jia, J., Kong, X., et al. 2007. Wheat genome structure and function: genome sequence data and the International Wheat Genome Sequencing Consortium. Aust. J. Agric. Res. **58**(6): 470–475. doi:10.1071/AR06155.

Moullet, O., Zhang, H.-B., and Lagudah, E.S. 1999. Construction and characterisation of a large DNA insert library from the D genome of wheat. Theor. Appl. Genet. **99**(1–2): 305–313. doi:10.1007/s001220051237.

Nilmalgoda, S.D., Cloutier, S., and Walichnowski, A.Z. 2003. Construction and characterization of a bacterial artificial chromosome (BAC) library of hexaploid wheat (*Triticum aestivum* L.) and validation of genome coverage using locus-specific primers. Genome, **46**(5): 870–878. doi:10.1139/g03-067. PMID:14608404.

Peng, J., Richards, D.E., Hartley, N.M., Murphy, G.P., Devos, K.M., Flintham, J.E., et al. 1999. 'Green revolution' genes encode mutant gibberellin response modulators. Nature, **400**(6741): 256–261. doi:10.1038/22307. PMID:10421366.

Pysh, L.D., Wysocka-Diller, J.W., Camilleri, C., Bouchez, D., and Benfey, P.N. 1999. The GRAS gene family in Arabidopsis: sequence characterization and basic expression analysis of the *SCARECROW-LIKE* genes. Plant J. **18**(1): 111–119. doi:10.1046/j.1365-313X.1999.00431.x. PMID:10341448.

Sambrook, J., Fitsch, E.F., and Maniatis, T. 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Shizuya, H., Birren, B., Kim, U.J., Mancino, V., Slepak, T., Tachiiri, Y., and Simon, M. 1992. Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. Proc. Natl. Acad. Sci. U.S.A. **89**(18): 8794–8797. doi:10.1073/pnas.89.18.8794. PMID:1528894.

Xia, Z.J., Sato, H., Watanabe, S., Kawasaki, S., and Harada, K. 2005. Construction and characterization of a BAC library of soybean. Euphytica, **141**(1–2): 129–137. doi:10.1007/s10681-005-6380-8.

Yim, Y.-S., Moak, P., Sanchez-Villeda, H., Musket, T.A., Close, P., Klein, P.E., et al. 2007. A BAC pooling strategy combined with PCR-based screenings in a large, highly repetitive genome enables integration of the maize genetic and physical maps. BMC Genomics, **8**(1): 47. doi:10.1186/1471-2164-8-47. PMID:17291341.