Theor Appl Genet (2003) 107:931–939
DOI 10.1007/s00122-003-1331-z

A. Cenci · N. Chantret · X. Kong · Y. Gu · O. D. Anderson · T. Fahima · A. Distelfeld · J. Dubcovsky

Construction and characterization of a half million clone BAC library of durum wheat (*Triticum turgidum* ssp. *durum*)

Received: 10 January 2002 / Accepted: 14 March 2003 / Published online: 27 June 2003

Abstract Durum wheat (*Triticum turgidum* ssp. *durum*, $2n = 4x = 28$, genomes AB) is an economically important cereal used as the raw material to make pasta and semolina. In this paper we present the construction and characterization of a bacterial artificial chromosome (BAC) library of tetraploid durum wheat cv. Langdon. This variety was selected because of the availability of substitution lines that facilitate the assignment of BACs to the A and B genome. The selected Langdon line has a 30-cM segment of chromosome 6BS from *T. turgidum* ssp. *dicoccoides* carrying a gene for high grain protein content, the target of a positional cloning effort in our laboratory. A total of 516,096 clones were organized in 1,344 384-well plates and blotted on 28 high-density filters. Ninety-eight percent of these clones had wheat DNA inserts (0.3% chloroplast DNA, 1.4% empty clones and 0.3% empty wells). The average insert size of 500 randomly selected BAC clones was 131 kb, resulting in a coverage of 5.1-fold genome equivalents for each of the two genomes, and a 99.4% probability of recovering any gene from each of the two genomes of durum wheat. Six known copy-number probes were used to validate this theoretical coverage and gave an estimated coverage of 5.8-fold genome equivalents. Screening of the library with 11 probes related to grain storage proteins and starch biosynthesis showed that the library contains several clones for each of these genes, confirming the value of the library in characterizing the organization of these important gene families. In addition, characterization of fingerprints from colinear BACs from the A and B genomes showed a large differentiation between the A and B genomes. This library will be a useful tool for evolutionary studies in one of the best characterized polyploid systems and a source of valuable genes for wheat. Clones and high-density filters can be requested at http://agronomy.ucdavis.edu/Dubcovsky/BAC-library/BAC_Langdon.htm

Keywords BAC · Genomic library · Wheat · Polyploid · *Triticum durum*

Communicated by P. Langridge

The first two authors contributed equally to the investigation

A. Cenci · N. Chantret · T. Fahima · A. Distelfeld · J. Dubcovsky (✉)
Department of Agronomy and Range Science, University of California, One Shields Avenue, Davis, CA 95616-8515, USA
e-mail: jdubcovsky@ucdavis.edu
Fax: +1-530-7524361

X. Kong · Y. Gu · O. D. Anderson
USDA Western Reg. Research Center, 800 Buchanan St, Albany, CA 94710, USA

Present address:
A. Cenci, Dip. Biologia e Chimica Agroforestale e Ambientale, Università degli Studi di Bari, Via G. Amendola 165/a, 70126 Bari, Italy

Present address:
N. Chantret, CIRAD-AMIS-Biotrop, av. Agropolis, TA 40/30, 34398 Montpellier Cedex 5, France

Present address:
T. Fahima · A. Distelfeld, The Institute of Evolution, University of Haifa, Mt. Carmel, Haifa. Israel

Introduction

Yeast artificial chromosomes (YACs) and bacterial artificial chromosomes (BACs) are valuable tools that allow the cloning of large fragments of DNA and are used to construct representative libraries of entire genomes (Burke et al. 1987; Shizuya et al. 1992). BAC libraries are generally preferred over YAC libraries because of their relatively simple production process, greater clone stability and low frequency of chimerism (Woo et al. 1994). BAC libraries are also easy to manipulate, and the isolation of exogenous DNA is relatively similar to normal plasmid preparations (Woo et al. 1994).

Large-insert libraries are useful for gene positional cloning strategies (Tanksley et al. 1995), for genome structure analysis (Chen and Foolad 1997; Panstruga et al.

1998; SanMiguel et al. 2002), for genome organization comparisons between related species (Chen et al. 1997; Tikhonov et al. 1999; Dubcovsky et al. 2001) and for map saturation with molecular markers and microsatellite discovery in targeted regions (Cregan et al. 1999).

The required number of BAC clones for a library mainly depends on the genome size of the species considered, the required genome coverage and the average insert size. Considering average inserts of 100 kb, a five fold genome coverage of the small genomes of *Arabidopsis thaliana* (125 Mb) and rice (450 Mb) would require 7,000 and 23,000 BAC clones, respectively (Choi et al. 1995; Zhang et al. 1996). However, species with large genomes, such as barley (5,000 Mb) or diploid wheat (5,600 Mb), require 250,000 and 280,000 clones for a similar coverage (Lijavetzky et al. 1999; Yu et al. 2000).

Tetraploid wheat *Triticum turgidum* L. (genomes AABB) originated from the hybridization of *T. urartu* Thum. and a species closely related to *T. speltoides* (Tausch) Gren. (Dvorak et al. 1988; Dvorak and Zhang 1990). The modern hexaploid bread wheats (genomes AABBDD) originated from a more recent hybridization of *T. turgidum* with diploid *T. tauschii* (Coss.) Schmalh. (genome DD; Kihara 1944). The combination of large diploid genomes in the polyploid species resulted in even larger genomes, varying from 13,000 Mb in tetraploid wheat to 16,000 Mb in hexaploid wheat *T. aestivum* L. (Arumuganathan and Earle 1991).

BAC libraries are available for the D genome of diploid *T. tauschii* (Moullet et al. 1999) and for the A^m genome of diploid *T. monococcum* L. (genome A^mA^m) (Lijavetzky et al. 1999). *Triticum monococcum* is a species closely related to *T. urartu*, the A-genome donor of cultivated polyploid wheats (Dvorak et al. 1993). None of the diploid species has a genome that matches perfectly the B genome from polyploid wheats (Dvorak and Zhang 1990; Huang et al. 2002). Therefore, the B genome needs to be cloned from a polyploid species. A transformation-competent artificial chromosome (TAC) BAC library of hexaploid wheat variety Chinese Spring has been produced, but its clones were stored as bulked cultures and no high-density filters are available (Liu et al. 2000; Ma et al. 2000).

To fill this gap, we decided to construct a gridded BAC library of the tetraploid species *T. turgidum* var. *durum*. Among the tetraploid species, durum wheat (*T. turgidum* ssp. *durum*) is the most cultivated and is an economically important cereal, especially in the Mediterranean area, since it is the raw material for making pasta and semolina. Therefore, a BAC library from tetraploid wheat will provide access to genes directly involved with pasta quality. Our laboratory is involved in the positional cloning of a gene from *T. dicoccoides* that increases protein content in the grain (Joppa et al. 1997; Khan et al. 2000). This BAC library will be an essential tool for positional cloning of this and other genes and will also be a valuable resource for comparative genomic studies of the wheat genomes.

In this paper we present the construction and organization of a BAC library of the tetraploid species *T. turgidum* ssp. *durum* and a preliminary characterization of its potential to study genes involved in quality traits. We also discuss the differentiation of colinear regions in the A and B genomes.

Materials and methods

Plant material

Tetraploid cultivar Langdon was selected for the construction of this library because the availability of cytogenetic stocks will facilitate the assignment of BAC clones to individual genomes. A complete set of substitution lines with D-genome chromosomes replacing each of the 14 chromosome pairs of cv. Langdon (Joppa and Williams 1988) was used to assign selected BAC clones to the A or B genomes.

The line selected for the construction of this library is not a pure Langdon line but a homozygous recombinant substitution line (Langdon no. 65; Joppa et al. 1997). Langdon no. 65 has 13 pairs of chromosomes identical to Langdon and a recombinant chromosome 6B of Langdon carrying a 30-cM homozygous introgression of the *T. turgidum* ssp. *dicoccoides* 6B chromosome. This 30-cM segment encompasses a gene for high grain protein content (Joppa et al. 1997), the target of a positional cloning project at the UC Davis (USA) and Haifa University (Israel) laboratories.

Preparation of the BAC vector

The plasmid pCUGIBAC1 (BAC vector pINDIGO536 cloned into the high-copy vector pGEM-4Z; Luo et al. 2001) was kindly provided by Dr. R. Wing (Clemson University Genomics Institute, S.C.). Plasmid DNA was extracted from the *Escherichia coli* host by an AutoGen 850α robot (AutoGen, Framingham, Mass.). Ten micrograms of plasmid were digested with 120 U *Hin*dIII for 3 h dephosphorylated by CIAP (Promega, Madison, Wis.), purified by phenol/chloroform extraction, and self-ligated with 10 U T4 ligase (to eliminate phosphorylated plasmids). The dephosphorylated 7.5-kb fragment of the pINDIGO536 was digested with restriction enzyme *Hin*dIII, and recovered by electrophoresis and electroelution from the agarose gel. Aliquots of the vector were stored in 50% glycerol at –80 °C.

BAC library construction

BAC library construction was performed as described previously (Lijavetzky et al. 1999). The main difference from the previous protocol was the doubling of the amount of the high-molecular-weight (HMW) DNA that was separated by pulse field gel electrophoresis (PFGE).

Approximately 100 clones were tested for each ligation. Plasmid DNAs were isolated using an AutoGen 850α robot (AutoGen) and digested with *Not*I to release the insert. PFGE was performed using a 1% agarose gel at 200 V, with a 5- to 15-s pulse ramp for 14 h at 14 °C in 0.5× TBE, to check for the presence of inserts and to determine their size. Only those ligations that produced clones with an average size greater than 110 kb and with an empty clone percentage lower than 2% were included in the library. Clones were individually isolated by hand or using a Q-bot (Genetix, Dorset, UK) and were used to inoculate 384-well plates containing freezing medium (Woo et al. 1994) with 12.5 μg/ml of chloramphenicol. High-density filters, each containing 18,432 double-spotted clones, were produced using a Q-bot (Genetix).

Table 1 Genome coverage estimation based on hybridization of the durum wheat BAC library with known copy-number probes. Total copy numbers in the A plus B genomes are indicated in the second column

Probe	Copy number	Number of BACs	Genome coverage	Reference or GeneBank accession number for probe
Triplet	2	14	7	Singh et al. 1993
HMW-glutenin	4	23	5.8	Anderson et al. 1989
GBSS1	2	11	5.5	AF250137
PSR164	2	11	5.5	Gale et al. 1995
Wheat EST	2	14	7	BE604883
Wheat EST	2	8	4.5	BF484238
Average	14	81	5.8	

Table 2 Inferred copy number of quality related genes based on the number of selected BAC clones

Probe	Number of BACs	Inferred total copy number	Reference or GeneBank accession number for probe
LMW-glutenin	92	18	Cassidy and Dvorak 1991
α-gliadin	223	44	XO1130
γ-gliadin	56	11	AF234646
SBEI	8	2	Y12320
SBEII	19	4	Nair et al. 1997
SUI (Isoamylase)	17	3	AF438328
SSI	40	8	Peng et al. 2001
SSIIa	40	8	Gao and Chibbar 2000

Hybridization procedures

Plant nuclear DNAs were isolated from leaves of single plants following a procedure described previously (Dvorak et al. 1988). Hybridizations for both the high-density filters and the genomic Southern blots were performed as described by Dubcovsky et al. (1994). Probes used to determine the chloroplast (cp) DNA contamination of the BAC library were obtained by polymerase chain reaction (PCR) amplification of two chloroplast genes, *rbcL* and *petD*, from total wheat DNA. For *rbcL*, the primers were 5′-GGACTTATGTCACCACAAAC-3′ and 5′-CTACTTAT-CAATAGTATCTACC-3′; for *petD*, 5′-CTAAAGGGATGGGA-CATA-3′ and 5′-ATCAATGGTAATGTTGC-3′ (Ogihara and Tsunewaki 2000). The probes used to screen the BAC library are listed in Tables 1 and 2.

BAC fingerprinting

The objective of this experiment was to test if BACs from colinear regions of the A and B genomes would be assembled either into two separate contigs or be combined into one contig containing a combination of BACs from the A and B genomes. BAC clones from colinear regions of the A and B genomes were selected by hybridization with five single-copy probes (Table 1) that showed one restriction fragment per genome in hybridizations with Langdon genomic DNA. BAC fingerprints were produced using the restriction enzyme *Hin*dIII according to the protocols described by Marra et al. (1997). Southern blots from the fingerprints were hybridized with their respective probes to differentiate the BACs from the two different genomes. Contigs were assembled using the FPC program (Marra et al. 1997) using a tolerance of 3 and cutoff of $1e^{-22}$.

Results

Construction of durum wheat BAC library

Five ligations were used to construct this BAC library of durum wheat. The average insert size and the proportion of empty clones were determined independently for each ligation (Figs. 1 and 2). The library has 516,096 clones, arrayed in 1344 384-well plates. Approximately 150,000 clones were handpicked, and the rest were picked using a Q-bot (Genetix). The complete library was blotted on 28 high-density filters using a Q-bot (18,432 double-spotted clones per filter).

Characterization of the durum wheat BAC library

To estimate the average insert size, we estimated PFGE band sizes of 500 clones after *Not*I digestion (Fig. 1). The average size obtained was 131 kb with a standard deviation of 30 kb. The insert sizes ranged from 30 kb to 290 kb and showed a normal distribution (Fig. 2). Of the clones, 90% showed inserts larger than 100 kb, and only 1.4% of the clones had no inserts. The percentage of empty wells resulting from Q-bot errors was 0.3%.

More than 20% of the BAC inserts have internal *Not*I sites (2–6 bands in addition to the vector), resulting in an average of 1.71 restriction sites per insert. This result agreed with previous observations for *T. tauschii* (Moullet et al. 1999) and other monocot species (Woo et al. 1994; Zhang et al. 1996; Tomkins et al. 1999) that suggested a higher frequency of *Not*I restriction sites relative to dicot species. This result was also expected based on the higher GC content of the wheat genome (45.7%; SanMiguel et al. 2002) and the GCGGCCGC recognition sequence of *Not*I.

To evaluate the level of chloroplast contamination in the library, six high-density filters (representing a total of 110,592 clones) were hybridized with a mixture of two probes corresponding to the chloroplast genes *rbcL* and *petD* (Ogihara and Tsunewaki 2000). A total of 376

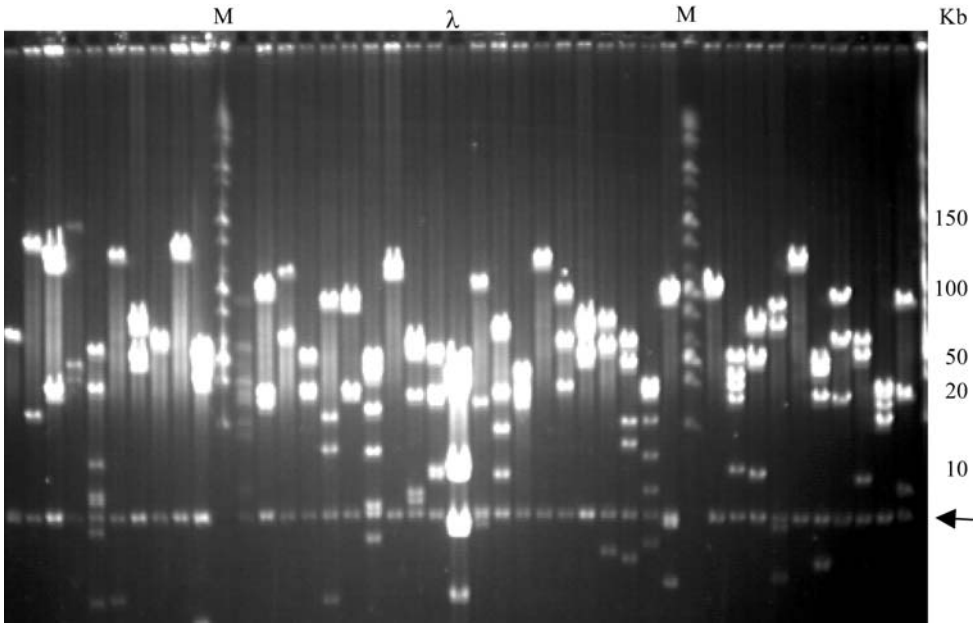

Fig. 1 Analysis of randomly selected *Triticum durum* BAC clones: ethidium bromide-stained PFGE gel of Langdon BAC clones digested with *Not*I. Molecular weight standards are: *M* MidRange PFG Marker I (New England Biolabs), λ *Hin*dIII-digested lambda DNA. *Arrow* indicates the vector

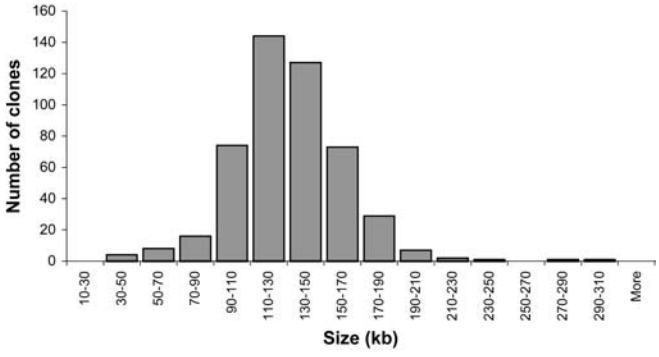

Fig. 2 Distribution of the insert size of 500 randomly selected *T. durum* BAC clones

clones hybridized with the chloroplast genes, corresponding to 0.34% contamination by cpDNA.

The total percentage of clones without genomic DNA inserts was 2% (empty wells plus empty clones plus chloroplast clones). Considering the above percentage, a 131-kb average insert size and assuming that the genome size of durum wheat is 13,000 Mb, we estimated the coverage of the library to be 5.1-fold genome equivalents. This coverage provides a 99.4% probability of recovering any gene present in each of the two genomes of durum wheat (Clarke and Carbon 1976).

Hybridization of the library with known copy-number clones and clones of genes affecting grain quality

The 28 high-density filters containing the complete BAC library were hybridized with five single-copy probes and one two-copy probe (Table 1). A total of 81 positive clones were found for these 14 genes, resulting in an observed average coverage of 5.8-fold genome equivalents, similar to the estimated coverage of 5.1-fold genome equivalents. Among these probes, HMW glutenin gene is present in two copies per genome (designated x and y) that are tightly linked in the three genomes of bread wheat (Payne et al. 1981). None of the BACs with the HMW-glutenin probe included more than one subunit, suggesting that the x-type and y-type HMW glutenin genes are not adjacent to each other in both genomes.

The 28 high-density filters were also screened with three probes for endosperm storage proteins and five probes from genes related to starch biosynthesis. The 5.1-fold estimated coverage of the BAC library was used to provide an indirect estimate of the total number of copies of these genes in the tetraploid genome (Table 2). These calculations suggested the presence of approximately 79 endosperm storage protein genes in both genomes, including two Triplet genes, four HMW glutenins, 18 low-molecular-weight (LMW) glutenins, 44 α-gliadins and 11 γ-gliadins. All except one of the BAC clones selected with the LMW glutenin probe showed a single LMW glutenin gene per BAC. This indicates that most of the LMW-glutenin genes are not adjacent in both genomes. Hybridization of the fingerprinting membranes from the LMW-glutenin BACs with a gliadin probe (Hor1 & 2; Forde et al. 1985) showed the presence of gliadin genes in three of the BACs, suggesting some interspersion of these two classes of storage protein genes.

The number of detected BACs containing genes related to starch biosynthesis suggested the presence of two copies of the starch branching enzyme (SBE) gene SBEI, four of SBEII, three of the isoamylase type debranching enzyme gene SU1 and eight for each of the starch synthase (SS) SSI and SSIIa genes (for a review on wheat starch biosynthesis genes, see Morell et al. 2001). The estimates in Table 2 are based on a 5.1-fold genome

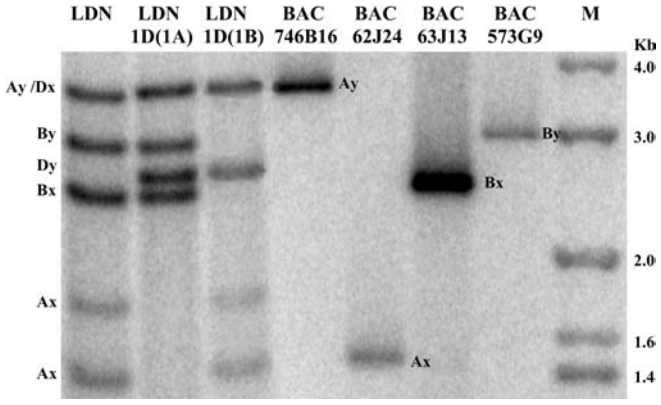

Fig. 3 Assignment of Langdon (*LDN*) BAC clones selected with the HMW-glutenin probe to the A or B genome using chromosome substitution lines Langdon 1D(1A) and Langdon 1D(1B). Genomic DNA and BAC clones were digested with the restriction enzyme *Taq*I

coverage and should be considered only as an approximation because of the random variation in the representation of different regions of the genome within the library.

Determination of the genome source of selected BACs using chromosome substitution lines

Since durum wheat is a tetraploid species, most of the genes are present in at least two copies, located once in each genome. In many cases it is important to determine the genome source of the selected BACs, a task that is facilitated by the availability of substitution lines of each of the 14 Langdon chromosomes by the homeologous chromosome from the D genome (Joppa and Williams 1988). The high level of restriction fragment length polymorphisms (RFLPs) between the different genomes can be used to assign BACs to the A or B genomes. Figure 3 shows an example of the assignment of four positive BAC clones obtained with the HMW-glutenin probe to the two different genomes and some possible complications in this strategy. In this example, the RFLP fragments that disappear in the LDN1D(1A) and LDN1D(1B) substitution lines were assigned to the A and B genome, respectively. Comparison of the RFLP pattern of these lines with that of the selected BAC clones digested with the same restriction enzyme was used to assign BAC 62J24 to the A genome and BACs 63J13 and 573G9 to the B genome.

Some complications can be found in this strategy. The 3.2-kb fragment corresponding to the Ay gene is not missing in LDN1D(1A) because there is a fragment of similar mobility in the D genome (Fig. 3). Assignment of this BAC to the A genome was confirmed by the connection of BACs 746B16 with 62J24 by fingerprinting. BAC 62J24 has only part of the Ax HMW glutenin gene at one end of the BAC and, therefore, the 1.7-kb fragment present in the A genome is missing in this BAC. The 1.7-kb and 1.3-kb *Taq*I fragments were present together in all other BACs including the Ax gene.

Comparison of fingerprints from A and B genome BACs from colinear regions

One critical question for the future construction of physical maps of tetraploid wheat is if the degree of divergence among the genomes will be sufficient to generate separate contigs for colinear regions of the A and B genomes. Southern blots of the *Hin*dIII fingerprints from the BACs clones selected by single-copy probes (Table 1) were hybridized with their respective probes. In all cases, only two different hybridization patterns were observed, confirming that these were single-copy genes with one copy per genome. The fingerprints were analyzed together by FPC and assembled into contigs. In all cases, the A and B genome BACs were assembled into separate groups. BACs from the same genome were assembled together with the exception of two BACs from the Triplet gene that were not connected under the parameters used in this experiment (tolerance = 3, cutoff = $1e^{-22}$).

A similar result was obtained with the fingerprinting of 23 BACs from the HMW glutenins. The BACs were assembled into two separate contigs, each including the x and y genes from the same genome. The connection of the BACs including the x and y subunits from each genome in a contig indicates that the distance between the x and y subunits is less than twice the average BAC insert size.

Discussion

The construction of a fivefold coverage BAC library of tetraploid wheat was a significant effort due to the large size of its genome (13,000 Mb). The number of clones present in this library is equal to the number of clones present in 15 rice BAC libraries with a tenfold genomic coverage each. An alternative strategy was recently used in hexaploid wheat to reduce the effort required for the construction of large BAC libraries. BAC clones were stored as bulked cultures without picking and gridding (Liu et al. 2000). However, libraries stored as single clones and organized in 384-well plates as the one presented in this paper are necessary to make high-density filters which facilitate screening by hybridization in a single step and distribution of the library for international collaborations. At the time of this publication complete copies of this library have been distributed to Australia, France, Israel, Italy, Switzerland and the UK. Clones and high-density filters can be requested at http://agronomy.ucdavis.edu/Dubcovsky/BAC-library/BAC_Langdon.htm.

To reduce the number of clones required for a fivefold coverage, we made an effort to increase the average size of the inserts and to minimize the proportion of clones containing cpDNA or no inserts. Large inserts are also a

requirement for chromosome walking in wheat because large regions of repetitive elements are present, even in gene-rich regions (SanMiguel et al. 2002). The average insert size of this library (131 kb) was larger than the average insert size of previous published Triticeae BAC libraries (*T. tauschii*, 119 kb; *T. monococcum*, 115 kb; *H. vulgare*, 106 kb) and the proportion of clones with cpDNA or no inserts was reduced to very low levels (<2%).

The Langdon BAC library as a tool to characterize polyploid evolution

Durum wheat is a relatively young polyploid with extensive colinearity between the two genomes (Blanco et al. 1998; Nachit et al. 2001). This library will provide a valuable tool to characterize the genomic changes that have occurred during these early stages of polyploidization. Polyploidization is a dynamic process playing an important role in the evolution of plants (Soltis and Soltis 1995; Wendel 2000). Rapid genomic changes, including elimination of DNA sequences, activation of transposons, gene silencing and methylation changes, have been observed after the formation of new polyploid species (Comai et al. 2000; Wendel 2000; Ozkan et al. 2001; Shaked et al. 2001). Comparison of BACs from the A genome of the Langdon BAC library with colinear regions from the A genome of diploid wheat species will provide additional insights into the nature and extent of these processes.

Fingerprints of orthologous BACs from the A and B genomes show a very limited number of common restriction fragments, suggesting extensive differentiation between the A and B genome. This differentiation was expected based on the high level of RFLP polymorphism observed between the A and B genomes and the limited sequence similarity observed between the wheat genomes outside gene regions (Anderson et al. 1998). The A and B genomes diverged approximately 2.5–4.5 million years ago (Huang et al. 2002), providing a long period of time for divergence. Dating of 11 retrotransposon insertions in a 215-kb region of *T. monococcum* (SanMiguel et al. 2002) showed that all of them were inserted during the last 4.5 million years and that more than half of them were inserted in the last 2.5 million years, after the divergence of the A and B genome diploid species. In addition, an active process of deletion of repetitive elements was detected in the intergenic regions, leading to an almost complete differentiation of the colinear intergenic regions from wheat and barley after 10–15 million years of divergence (SanMiguel et al. 2002). The divergence of the A-genome species, *T. monococcum* (A^m) and *T. urartu* (A), occurred more recently (approximately 0.5–1 million years; Huang et al. 2002). A comparison between 300 kb of homologous sequence from chromosomes $5A^m$ and 5A (from tetraploid wheat) has shown that 75% of the intergenic regions present in *T. monococcum* have no similar sequences in the A genome of tetraploid wheat (J. Dubcovsky and P. San Miguel, unpublished). Based on the previous data, the extensive differentiation of the fingerprints from colinear BACs from the A and B genomes is not surprising. This result suggests that global fingerprinting strategies can be used in the construction of separate physical maps of the A and B genomes in tetraploid wheat.

We confirmed this result by repeating the fingerprinting experiment for 16 BACs selected with three single-copy probes using a more sophisticated fingerprinting technique based on capillary sequencers and precise estimates of fragment sizes. Drs. J. Dvorak and M.C. Luo (University of California Davis) kindly provided access to the fingerprinting technique currently being used in their laboratory to assemble a complete physical map of the D genome of *T. tauschii* (http://wheat.pw.usda.gov/PhysicalMapping/). The FPC analysis of these 16 fingerprints resulted in six contigs, one per probe/genome combination, demonstrating that the A and B genome BACs from colinear regions can be easily separated by different fingerprinting techniques.

The Langdon BAC library as a tool to characterize genes for important agronomic traits

Durum wheat is second in importance to bread wheat among the cultivated wheats. Between 20 and 30 million tons per year have been produced worldwide during the last 10 years (USDA Foreign Agricultural Service online: http://www.fas.usda.gov/). This library will provide a direct resource to clone agronomically important genes in wheat, including disease resistance genes, genes related to abiotic stress and genes related to quality. Although the model species rice has been almost completely sequenced and is a powerful tool for wheat genetics, many traits can be studied only in wheat. Genes affecting bread-making and pasta-making characteristics are good examples of these types of traits.

The quality of different wheat products is primarily determined by the variation in the gluten and starch properties. The main genes affecting these two characteristics are those encoding seed storage proteins and starch biosynthesis enzymes. Therefore, knowledge of the structure and organization of these genes is important for genetic improvement of grain quality. The large number of genes present in some of these families and the polyploid nature of wheat has complicated the characterization of these gene families using whole genome approaches. Results from the present study show that individual members of these families can be separated in individual BACs, thereby facilitating the sequencing of individual genes, the characterization of their promoters and the design of specific PCR primers for mapping and expression studies.

The gliadins represent more than 50% of the total endosperm proteins. The α-gliadin genes, located at the *Gli-2* loci on the short arm of the chromosomes of homoeologous group 6, are estimated to have between

25–150 copies depending on the exclusion or inclusion of pseudogenes in the gene number determinations (Anderson and Greene 1997). A lower number of copies, varying between 15 and 40 different genes, has been estimated for the γ-gliadins located at the *Gli-1* and *Gli-3* loci on the short arm of the chromosomes from homoeologous group 1 (Sabelli and Shewry 1991). These estimates are similar to those predicted in this study based on the number of BAC clones detected with these genes (Table 2). Fingerprint analysis of the 223 α-gliadin BAC clones and subsequent contig assembly using the FPC program generated two large contigs and several small contigs, suggesting that the two large contigs are the major α-gliadin-rich regions (Y. Gu and O.D. Anderson, unpublished data). Some BAC clones showed the simultaneous presence of gliadin and LMW-glutenin genes, suggesting some level of interspersion of these two gene families. Further characterization and sequencing of contigs from the main gliadin and LMW-glutenin loci will provide a better understanding of the organization and evolution of this gene family in the wheat genome.

Multiple BAC clones were detected with each of the genes involved in the starch biosynthesis (Table 2). Under the hybridization and washing conditions used in our study, the number of BAC clones selected with the complete endosperm-expressed SBEI cDNA (Accession number Y12320) suggests the presence of one SBEI gene copy per genome. The number of restriction fragments in wheat genomic DNA hybridized with SBEI varied depending on the region of the gene used as a probe (Rahman et al. 1997, 1999). Analysis of the genes identified from *Triticum tauschii* demonstrated that there are multiple copies of the SBEI present in the genome but that only one of them corresponds to the endosperm-expressed protein (Rahman et al. 1997, 1999). There is strong evidence in cereals that the SBEII class contains two genes (Rahman et al. 2001), in agreement with the observed number of positive BACs selected with the SBEII probe. This is also the case for the isoamylase debranching enzyme SU1, where the 17 positive BAC clones detected with this probe are within the expected number based on the reported single-copy nature of this gene (Peng et al. 2001). A larger number of BACs than expected was found for both starch synthase genes SSI and SSII. Although a single-copy gene per genome was reported for each of these genes (Li et al. 1999a, b), 40 positive BAC clones were detected for each gene. Only one BAC clone was detected simultaneously by SSI and SSII, suggesting that this difference did not originate in cross hybridization between the two genes. One possible explanation is the presence of a repetitive element (e.g. MITE) within the clones used to screen the BAC library. Detailed analyses of these clones are being performed to explain the differences between the expected and observed numbers (R.N. Chibbar, personal communication).

This library is currently being used in the University of California, Davis and the University of Haifa laboratories for the positional cloning of a gene for high grain protein content located on chromosome 6BS (Khan et al. 2000). In France, BACs from the A and B genome are being compared with the A^m genome to characterize the evolution of the hardness loci. BAC clones including genes related to pasta quality are being studied in Italy and those including seed storage proteins are being sequenced in USDA-Albany. In addition, this library has been already distributed to more than eight laboratories worldwide. Clones and high-density filters can be requested at http://agronomy.ucdavis.edu/Dubcovsky/BAC-library/BAC_Langdon.htm.

Acknowledgements This research was supported by Research Grant No. US-3224-01R from BARD (The United States -Israel Binational Agricultural Research and Development Fund), by INRA-France and by USDA-Albany. The authors thank Dr R. N. Chibbar for the screening of the Langdon BAC library with the genes involved in starch biosynthesis.

References

Anderson OD, Greene FC (1997) The alpha-gliadin gene family.II. DNA and protein sequence variation, subfamily structure, and origins of pseudogenes. Theor Appl Genet 95:59–65

Anderson OD, Greene FC, Yip RE, Halford NG, Shewry PR, Malpica-Romero J-M (1989) Nucleotide sequences of the two high-molecular-weight glutenin genes from the D genome of hexaploid wheat, *Triticum aestivum* L. cv. Cheyenne. Nucleic Acids Res 17:461–462

Anderson OD, AbrahamPierce FA, Tam A (1998) Conservation in wheat high-molecular-weight glutenin gene promoter sequences: comparisons among loci and among alleles of the *Glu-B1-1* locus. Theor Appl Genet 96:568–576

Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. Plant Mol Biol Rep 9:208–218

Blanco A, Bellomo MP, Cenci A, DeGiovanni C, Dovidio R, Iacono E, Laddomada B, Pagnotta MA, Porceddu E, Sciancalepore A, Simeone R, Tanzarella OA (1998) A genetic linkage map of durum wheat. Theor Appl Genet 97:721–728

Burke DT, Carle GF, Olson MV (1987) Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. Science 236:806–812

Cassidy BG, Dvorak J (1991) Molecular characterization of a low-molecular-weight glutenin cDNA clone from *Triticum durum*. Theor Appl Genet 81:653–660

Chen FQ, Foolad MR (1997) Molecular organization of a gene in barley which encodes a protein similar to aspartic protease and its specific expression in nucellar cells during degeneration. Plant Mol Biol 35:821–831

Chen M, SanMiguel P, Oliveira ACd, Woo S-S, Zhang H, Wing RA, Bennetzen JL (1997) Microcolinearity in *sh2*-homologous regions of the maize, rice, and sorghum genomes. Proc Natl Acad Sci USA 94:3431–3435

Choi S, Creelman RA, Mullet JE, Wing RA (1995) Construction and characterization of a bacterial artificial chromosome library of *Arabidopsis thaliana*. Plant Mol Biol Rep 13:124–128

Clarke L, Carbon J (1976) A colony bank containing synthetic ColE1 hybrid plasmids representative of the entire *E. coli* genome. Cell 9:91–100

Comai L, Tyagi AP, Winter K, Holmes-Davis R, Reynolds SH, Stevens Y, Byers B (2000) Phenotypic instability and rapid gene silencing in newly formed *Arabidopsis* allotetraploids. Plant Cell 12:1551–1567

Cregan PB, Mudge J, Fickus EW, Marek LF, Danesh D, Denny R, Shoemaker RC, Matthews BF, Jarvik T, Young ND (1999) Targeted isolation of simple sequence repeat markers through the use of bacterial artificial chromosomes. Theor Appl Genet 98:919–928

Dubcovsky J, Galvez AF, Dvorak J (1994) Comparison of the genetic organization of the early salt stress response gene system in salt-tolerant *Lophopyrum elongatum* and salt-sensitive wheat. Theor Appl Genet 87:957–964

Dubcovsky J, Ramakrishna W, SanMiguel P, Busso C, Yan L, Shiloff B, Bennetzen J (2001) Comparative sequence analysis of colinear barley and rice BACs. Plant Physiol 125:1342–1353

Dvorak J, Zhang HB (1990) Variation in repeated nucleotide sequences sheds light on the phylogeny of the wheat B and G genomes. Proc Natl Acad Sci USA 87:9640–9644

Dvorak J, McGuire PE, Cassidy B (1988) Apparent sources of the A genomes of wheats inferred from the polymorphism in abundance and restriction fragment length of repeated nucleotide sequences. Genome 30:680–689

Dvorak J, di Terlizzi P, Zhang HB, Resta P (1993) The evolution of polyploid wheats: identification of the A genome donor species. Genome 36:21–31

Forde BG, Heyworth A, Pywell J, Kreis M (1985) Nucleotide sequence of a B1 hordein gene and the identification of possible upstream regulatory elements in endosperm storage protein genes from barley, wheat and maize. Nucleic Acids Res 13:7327–7339

Gale MD, Atkinson MD, Chinoy CN, Harcourt RL, Jia J, Li QY, Devos KM (1995) Genetic maps of hexaploid wheat. In: Li ZS, Xin ZY (eds) Proc 8th Int Wheat Genetic Symp. China Agricultural Scientech Press, Beijing, pp 29–40

Gao M, Chibbar RN (2000) Isolation, characterization, and expression analysis of starch synthase IIa cDNA from wheat (*Triticum aestivum* L.). Genome 43:768–775

Huang S, Sirikhachornkit A, Su XJ, Faris J, Gill B, Haselkorn R, Gornicki P (2002) Genes encoding plastid acetyl-CoA carboxylase and 3-phosphoglycerate kinase of the *Triticum/Aegilops* complex and the evolutionary history of polyploid wheat. Proc Natl Acad Sci USA 99:8133–8138

Joppa LR, Williams ND (1988) Langdon durum disomic substitution lines and aneuploid analysis in tetraploid wheat. Genome 30:222–228

Joppa LR, Du C, Hart GE, Hareland GA (1997) Mapping a QTL for grain protein in tetraploid wheat (*Triticum turgidum* L.) using a population of recombinant inbred chromosome lines. Crop Sci 37:1586–1589

Khan IA, Procunier JD, Humphreys DG, Tranquilli G, Schlatter AR, Marcucci-Poltri S, Frohberg R, Dubcovsky J (2000) Development of PCR based markers for a high grain protein content gene from *Triticum turgidum* ssp. *dicoccoides* transferred to bread wheat. Crop Sci 40:518–524

Kihara H (1944) Discovery of the DD-analyser, one of the ancestors of *Triticum vulgare*. Agric Hortic 19:13–14

Li Z, Rahman S, Kosar-Hashemi B, Mouille G, Appels R, Morell MK (1999a) Cloning and characterization of a gene encoding wheat starch synthase I. Theor Appl Genet 98:1208–1216

Li ZY, Chu XS, Mouille G, Yan LL, Kosar-Hashemi B, Hey S, Napier J, Shewry P, Clarke B, Appels R, Morell MK, Rahman S (1999b) The localization and expression of the class II starch synthases of wheat. Plant Physiol 120:1147–1155

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

Liu YG, Nagaki K, Fujita M, Kawaura K, Uozumi M, Ogihara Y (2000) Development of an efficient maintenance and screening system for large-insert genomic DNA libraries of hexaploid wheat in a transformation-competent artificial chromosome (TAC) vector. Plant J 23:687–695

Luo MZ, Wang YH, Frisch D, Joobeur T, Wing RA, Dean RA (2001) Melon bacterial artificial chromosome (BAC) library construction using improved methods and identification of clones linked to the locus conferring resistance to melon Fusarium wilt (*Fom-2*). Genome 44:154–162

Ma Z, Weining S, Sharp PJ, Liu C (2000) Non-gridded library: a new approach for BAC (bacterial artificial chromosome) exploitation in hexaploid wheat (*Triticum aestivum*). Nucleic Acids Res 28:e106

Marra MA, Kucaba TA, Dietrich NL, Green ED, Brownstein B, Wilson RK, McDonald KM, LaHillier LW, McPherson JD, Waterston RH (1997) High throughput fingerprint analysis of large-insert clones. Genome Res 7:1072–1084

Morell MK, Rahman S, Regina A, Appels R, Li Z (2001) Wheat starch biosynthesis. Euphytica 119:55–58

Moullet O, Zhang H-B, Lagudah ES (1999) Construction and characterization of a large DNA insert library from the D genome of wheat. Theor Appl Genet 99:305–313

Nachit MM, Elouafi I, Pagnotta MA, El Saleh A, Iacono E, Labhilili M, Asbati A, Azrak M, Hazzam H, Benscher D, Khairallah M, Ribaut JM, Tanzarella OA, Porceddu E, Sorrells ME (2001) Molecular linkage map for an intraspecific recombinant inbred population of durum wheat (*Triticum turgidum* L. var. *durum*). Theor Appl Genet 102:177–186

Nair RB, Baga M, Scoles GJ, Kartha KK, Chibbar RN (1997) Isolation, characterization and expression analysis of a starch branching enzyme II cDNA from wheat. Plant Sci 122:153–163

Ogihara Y, Tsunewaki K (2000) Chinese Spring wheat (*Triticum aestivum* L.) chloroplast genome: complete sequence and contig clones. Yokohama City University and Fukui Prefetural University, Japan

Ozkan H, Levy AA, Feldman M (2001) Allopolyploidy-induced rapid genome evolution in the wheat (*Aegilops-Triticum*) group. Plant Cell 13:1735–1747

Panstruga R, Buschges R, Piffanelli P, Schulze-Lefert P (1998) A contiguous 60 kb genomic stretch from barley reveals molecular evidence for gene islands in a monocot genome. Nucleic Acids Res 26:1056–1062

Payne PI, Corfield KG, Holt LM, Blackman JA (1981) Correlations between the inheritance of certain high molecular weight subunits of glutenin and bread-making quality in progenies of six crosses of bread wheat. J Agric Sci 32:51–60

Peng MS, Hucl P, Chibbar RN (2001) Isolation, characterization and expression analysis of starch synthase I from wheat (*Triticum aestivum* L.). Plant Sci 161:1055–1062

Rahman S, Abrahams S, Abbott D, Mukai Y, Samuel M, Morell M, Appels R (1997) A complex arrangement of genes at a starch branching enzyme I locus in the D-genome donor of wheat. Genome 40:465–474

Rahman S, Li Z, Abrahams S, Abbott D, Appels R, Morell MK (1999) Characterisation of a gene encoding wheat endosperm starch branching enzyme-I. Theor Appl Genet 98:156–163

Rahman S, Regina A, Li ZY, Mukai Y, Yamamoto M, Kosar-Hashemi B, Abrahams S, Morell MK (2001) Comparison of starch-branching enzyme genes reveals evolutionary relationships among isoforms. Characterization of a gene for starch-branching enzyme IIa from the wheat D genome donor *Aegilops tauschii*. Plant Physiol 125:1314–1324

Sabelli PA, Shewry PR (1991) Characterization and organization of gene families at the *Gli-1* loci of bread and durum wheats by restriction fragment analysis. Theor Appl Genet 83:209–216

SanMiguel P, Ramakrishna W, Bennetzen JL, Busso CS, Dubcovsky J (2002) Transposable elements, genes and recombination in a 215-kb contig from wheat chromosome 5A. Functional Integrative Genomics 2:70–80

Shaked H, Kashkush K, Ozkan H, Feldman M, Levy AA (2001) Sequence elimination and cytosine methylation are rapid and reproducible responses of the genome to wide hybridization and allopolyploidy in wheat. Plant Cell 13:1749–1759

Shizuya H, Birren B, Kim U-J, Mancino V, Slepak T, Tachiiri Y, 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 USA 89:8794–8797

Singh NK, Donovan GR, Carpenter HC, Skerritt JH, Langridge P (1993) Isolation and characterization of wheat triticin cDNA revealing a lysine-rich repetitive domain. Plant Mol Biol 22:227–237

Soltis DE, Soltis PS (1995) The dynamic nature of polyploid genomes. Proc Natl Acad Sci USA 92:8089–8091

Tanksley SD, Ganal MW, Martin GB (1995) Chromosome landing: a paradigm for map-based gene cloning in plants with large genomes. Trends Genet 11:63–68

Tikhonov AP, SanMiguel PJ, Nakajima Y, Gorenstein NM, Bennetzen JL, Avramova Z (1999) Colinearity and its exceptions in orthologous *Adh* regions of maize and sorghum. Proc Natl Acad Sci USA 96:7409–7414

Tomkins JP, Yu Y, Miller-Smith H, Frisch DA, Woo SS, Wing RA (1999) A bacterial artificial chromosome library for sugarcane. Theor Appl Genet 99:419–424

Wendel JF (2000) Genome evolution in polyploids. Plant Mol Biol 42:225–249

Woo SS, Jiang J, Gill BS, Paterson AH, Wing RA (1994) Construction and characterization of a bacterial artificial chromosome library of *Sorghum bicolor*. Nucleic Acids Res 22:4922–4931

Yu Y, Tomkins JP, Waugh R, Frisch DA, Kudrna D, Kleinhofs A, Brueggeman RS, Muehlbauer GJ, Wise RP, Wing RA (2000) A bacterial artificial chromosome library for barley (*Hordeum vulgare* L.) and the identification of clones containing putative resistance genes. Theor Appl Genet 101:1093–1099

Zhang H-B, Choi S, Woo SS, Li Z, Wing RA (1996) Construction and characterization of two rice bacterial artificial chromosome libraries from the parents of a permanent recombinant inbred mapping population. Mol Breed 2:11–24