Construction and characterization of two BAC libraries from *Brachypodium distachyon*, a new model for grass genomics

Naxin Huo, Yong Q. Gu, Gerard R. Lazo, John P. Vogel, Devin Coleman-Derr, Ming-Cheng Luo, Roger Thilmony, David F. Garvin, and Olin D. Anderson

Abstract: *Brachypodium* is well suited as a model system for temperate grasses because of its compact genome and a range of biological features. In an effort to develop resources for genome research in this emerging model species, we constructed 2 bacterial artificial chromosome (BAC) libraries from an inbred diploid *Brachypodium distachyon* line, Bd21, using restriction enzymes *Hin*dIII and *Bam*HI. A total of 73 728 clones (36 864 per BAC library) were picked and arrayed in 192 384-well plates. The average insert size for the *Bam*HI and *Hin*dIII libraries is estimated to be 100 and 105 kb, respectively, and inserts of chloroplast origin account for 4.4% and 2.4%, respectively. The libraries individually represent 9.4- and 9.9-fold haploid genome equivalents with combined 19.3-fold genome coverage, based on a genome size of 355 Mb reported for the diploid *Brachypodium*, implying a 99.99% probability that any given specific sequence will be present in each library. Hybridization of the libraries with 8 starch biosynthesis genes was used to empirically evaluate this theoretical genome coverage; the frequency at which these genes were present in the library clones gave an estimated coverage of 11.6- and 19.6-fold genome equivalents. To obtain a first view of the sequence composition of the *Brachypodium* genome, 2185 BAC end sequences (BES) representing 1.3 Mb of random genomic sequence were compared with the NCBI GenBank database and the GIRI repeat database. Using a cutoff expectation value of $E < 10^{-10}$, only 3.3% of the BESs showed similarity to repetitive sequences in the existing database, whereas 40.0% had matches to the sequences in the EST database, suggesting that a considerable portion of the *Brachypodium* genome is likely transcribed. When the BESs were compared with individual EST databases, more matches hit wheat than maize, although their EST collections are of a similar size, further supporting the close relationship between *Brachypodium* and the Triticeae. Moreover, 122 BESs have significant matches to wheat ESTs mapped to individual chromosome bin positions. These BACs represent colinear regions containing the mapped wheat ESTs and would be useful in identifying additional markers for specific wheat chromosome regions.

Key words: Brachypodium distachyon, BAC library, BAC end sequencing, comparative genomics, genomic mapping.

Résumé : Le *Brachypodium* constitue un bon système modèle pour les graminées de climat tempéré en raison de son génome compact et de plusieurs caractéristiques biologiques. En vue de développer des ressources génomiques chez ce modèle émergent, les auteurs ont produit deux banques BAC à partir d'une lignée fixée diploïde du *B. distachyon*, Bd21, à l'aide des enzymes *Hin*dIII et *Bam*HI. Au total, 73 728 clones (36 864 par banque) ont été disposés dans 192 plaques à 384 puits. La taille moyenne des inserts dans les banques *Bam*HI et *Hin*dIII est estimée à 100 et 105 kb, respectivement et les inserts d'origine chloroplastique représentent 4,4 % et 2,4 %, respectivement, des clones. Les banques offrent une redondance de 9,4 et 9,9 fois la taille du génome haploïde. Ensemble, ces banques présentent une couverture de 19,3 équivalents de génome sur la base d'un génome estimé à 355 Mb chez le *Brachypodium* diploïde, ce qui confère une probabilité de 99,99% de trouver une quelconque séquence dans chaque banque. Les banques ont été criblées avec huit gènes de la voie de biosynthèse de l'amidon pour évaluer de manière empirique la couverture du génome. Ces gènes se trouvaient au sein des banques à une fréquence permettant d'estimer une couverture génomique entre 11,6 et 19,6 génomes. Pour obtenir un premier aperçu de la composition du génome du *Brachypodium*, 2 185 séquences terminales de BAC (BES), totalisant 1,3 Mb d'ADN génomique aléatoire, ont été comparées à la banque GenBank du NCBI ainsi qu'à la banque GIRI de séquences répétées. En utilisant un seuil de $E < 10^{-10}$, seuls 3,3 % des BES ont montré de la similarité avec des séquences

Received 5 April 2006. Accepted 1 July 2006. Published on the NRC Research Press Web site at http://genome.nrc.ca on 19 October 2006.

Corresponding Editor: F. Belzile.

N. Huo and D. Coleman-Derr. United States Department of Agriculture – Agricultural Research Service, Western Regional Research Center, 800 Buchanan St., Albany, CA 94710, USA; Department of Plant Sciences, University of California, Davis, CA 95616, USA.
Y.Q. Gu,[1] G.R. Lazo, J.P. Vogel, R. Thilmony, and O.D. Anderson. United States Department of Agriculture – Agricultural Research Service, Western Regional Research Center, 800 Buchanan St., Albany, CA 94710, USA.
M.-C. Luo. Department of Plant Sciences, University of California, Davis, CA 95616, USA.
D.F. Garvin. USDA–ARS, Plant Science Research Unit and Department of Agronomy and Plant Genetics, 411 Borlaug Hall, University of Minnesota, 1991 Upper Buford Circle, University of Minnesota, St. Paul, MN 55108, USA.

[1]Corresponding author (e-mail: ygu@pw.usda.gov).

 doi:10.1139/G06-087

répétées au sein de la banque de données, tandis que 40,0 % étaient semblables à des EST, ce qui suggère qu'une portion considérable du génome du *Brachypodium* est vraisemblablement transcrit. Lorsque les BES ont été comparés à des banques individuelles d'EST, plus de positifs ont été trouvés au sein des EST du blé que du maïs bien que le nombre d'EST disponibles chez ces deux espèces soit assez semblable. Cela vient confirmer la proximité entre le *Brachypodium* et les hordées. De plus, 122 BES étaient semblables à des EST préalablement assignés à des régions spécifiques sur des chromosomes du blé. Ces BAC correspondent à des régions colinéaires contenant des EST de blé de position connue et ils pourraient s'avérer utiles pour identifier d'autres marqueurs pour des régions chromosomiques spécifiques chez le blé.

Mots clés : Brachypodium distachyon, banque de BAC, séquençage des extrémités de BAC, génomique comparée, cartographie génétique.

[Traduit par la Rédaction]

Introduction

The grass family (Poaceae), which includes major crops such as rice, maize, wheat, and barley, is the most economically important family in the plant kingdom. The wealth of genomics information obtained from model plants (i.e., *Arabidopsis* and rice) has greatly increased our knowledge of related plant species and allows us to design effective strategies for map-based cloning of genes from crop species with large and complex genomes. This is partially due to conservation of gene content and order (colinearity) in related grass genomes (Devos and Gale 1997). Despite this general trend of conservation, exceptions to colinearity are often observed among grass genomes (Guyot et al. 2004; Li and Gill 2002; Ramakrishna et al. 2002). Grass genomes are evolutionarily labile for many characteristics, including chromosome number and genome size (Gaut 2002). In addition, dynamic sequence changes such as insertions, deletions, duplications, and translocations can further complicate the use of many regions of the rice genome for cross-species comparison with other grasses, such as wheat and barley (Sorrells et al. 2003). Therefore, for genome studies in wheat, barley, and other temperate grasses, a "model grass genome" that is more closely related to these species than rice will serve as a better intermediate for comparative analysis (Draper et al. 2001; Foote et al. 2004).

The genus *Brachypodium* and wheat belong to the same subfamily, Pooideae, which diverged from the subfamily Ehrhartoideae, to which rice belongs, nearly 50 million years ago (mya) (Gaut 2002). Within the genus *Brachypodium*, *Brachypodium distachyon* is an annual species with 1C nuclear genome size of about 0.36–0.4 pg in diploid accessions ($2n = 2x = 10$), equivalent to 300 Mb (Bennett and Leitch 2005; Vogel et al. 2006*a*). Therefore, the size of the diploid *Brachypodium* genome falls between *Arabidopsis* (0.16 pg) and rice (0.495 pg) (Bennett and Leitch 2005). Recently, a phylogenetic analysis with nuclear DNA sequence data from a subset of *Brachypodium* ESTs has reinforced the close evolutionary relationships between *Brachypodium*, wheat, and barley (Vogel et al. 2006*b*). Because of its compact genome size, along with a range of other biological features, such as small physical size, short life cycle, and undemanding growth requirements, *Brachypodium* has been proposed as a model system for temperate grasses (Draper et al. 2001).

There have been recent reports of progress in developing *Brachypodium* as a model system. For instance, more than 2 dozen reference inbred lines, including diploids derived from single-seed descent, have been developed for the research community and rapid-cycling conditions were identified (Vogel et al. 2006*a*). High-efficiency genetic transformation of *Brachypodium distachyon* via microprojectile bombardment and *Agrobacterium* has been reported (Christiansen et al. 2005; Vogel et al. 2006*a*). A total of 20 440 expressed sequence tags (ESTs) were generated from 5 *Brachypodium* cDNA libraries constructed from various tissues (Vogel et al. 2006*b*). A large-insert library is one important remaining requirement necessary for structural and functional genomics studies on this emerging grass model species. The availability of large DNA fragments cloned into BACs and organized in a library with redundant genome coverage has made it possible to construct physical maps for entire genomes. Such overlapping BAC-based physical maps are valuable to the process of map-based or positional cloning of genes of interest. In addition, BAC libraries are useful in elucidating genome structure and organization and in comparisons with other related genomes. Previously, a BAC library was constructed for *Brachypodium sylvaticum* and successfully used for comparative analysis between this species, rice, and wheat (Foote et al. 2004). Further, the isolation of the major chromosome-pairing locus *Ph1* in polyploid wheat demonstrated the usefulness of the *B. sylvaticum* sequence in comparative analysis and marker development in targeted wheat regions (Griffiths et al. 2006). However, *B. sylvaticum* is a perennial species with an estimated genome size of 470 Mb, which is considerably larger than the diploid *B. distachyon* accessions (Bennett and Leitch 2005; Foote et al. 2004; Vogel et al. 2006*a*). A BAC library was constructed from the diploid *B. distachyon* (Hasterok et al. 2006). However, the genome coverage of this library is only 2.22-fold, which is insufficient for either developing a BAC-based whole-genome physical map through BAC fingerprinting or positional cloning of genes. Also, this library was not constructed using inbred lines. Thus, polymorphisms in the population of plants used to construct the library would lead to difficulties in assembling a map based on BAC fingerprints.

In this study, we report the construction and characterization of 2 deep-coverage BAC libraries for *B. distachyon*, which are to be used as resources for genomics studies on this emerging model species. Pilot sequencing of BAC ends was performed to provide a first view of the sequence composition of *Brachypodium*'s compact genome. Our results indicate that BAC end sequencing is an efficient and informative approach to anchoring BAC clones to the well-

Table 1. Estimated genome coverage (*x*) based on hybridization of the *Brachypodium* BAC libraries with starch biosynthesis gene probes.

EST probes	Copy number	No. of positive clones		Estimated genome coverage (*x*)		Match to wheat or barley starch biosynthesis genes	*E* value	References or GenBank accessions
		*Bam*HI	*Hin*dIII	*Bam*HI	*Hin*dIII			
DV482119	2	19	22	19.0	22.0	GBSSI (waxy)	1.0×10^{-107}	Clark et al. 1991
DV471309	2	16	22	16.0	22.0	SBEII	2.0×10^{-40}	Rahman et al. 2001
DV484026	1	6	9	12.0	18.0	SBEI	5.0×10^{-50}	Rahman et al. 1999
DV480922	2	6	13	6.0	13.0	GBSSII	1.0×10^{-105}	Vrinten et al. 2000
DV478009	3	19	43	12.6	24.6	AGPase	6.0×10^{-71}	Ali et al. 2000
DV482739	1	3	9	6.0	18.0	Isoamylase	0.0	Rahman et al. 2003
DV474839	1	3	10	6.0	20.0	β-amylase	0.0	X98504
DV475397	2	16	23	16.0	23.0	α-amylase	1.0×10^{-158}	Whittier et al. 1987

annotated rice genome and identifying orthologous regions of wheat for comparative analysis.

Materials and methods

Plant material

An inbred diploid line of *B. distachyon*, Bd21 derived from a single-seed descent (Vogel et al. 2006*a*), was used for BAC library construction. Seeds were germinated on filter paper and cold treated at 4 °C for 7 d to synchronize germination, then transferred to a soilless mixture in a greenhouse. The greenhouse was maintained at ~24 °C during the day and ~18 °C at night. Before harvesting, 3-week-old seedlings were placed in the dark for an additional 5 d. Approximately 60 g of leaf material was harvested from each tray. The seedlings were frozen in liquid nitrogen and then stored at –80 °C.

BAC library construction

Nuclei were isolated from ~60 g of seedling tissue according to the protocol previously reported (Zhang et al. 1995). The nuclei containing the high-molecular-weight DNA was then embedded in agarose plugs. Partial digestions of DNA with *Bam*HI (New England BioLabs, Beverly, Mass.) and *Hin*dIII (Roche, Indianapolis, Ind.) restriction endonucleases were performed using different enzyme concentrations ranging from 0.1 to 10 U. DNA fragments were separated by pulsed-field gel electrophoresis (PFGE; CHEF-DRII system, BioRad, Hercules, Calif.), and fragments in the size range of 150–350 kb were excised from the gel. DNA was electro-eluted from the excised gel slices in 0.5× Tris–borate–EDTA (TBE) buffer using Spectra/Pro 7 dialysis membranes (SpectrumLabs, Rancho Dominguez, Calif.). Purified high-molecular-weight DNA was ligated to CopyControl™ pCC1BAC™ (either *Hin*dIII or *Bam*HI Cloning-Ready) vector (Epicentre, Madison, Wis.). The ligation mixture was then used to transform TransforMax™ EPI300™ Electrocompetent *Escherichia coli* (Epicentre). Transformants were plated onto Q-trays (Genetix, Hampshire, UK) with Luria–Bertani agar medium containing 90 μg/mL isopropyl-β-D-thiogalactopyranoside (IPTG), 90 μg/mL X-gal, and 12.5 μg/mL of chloramphenicol to provide blue–white and antibiotic selection. The amount of plated cells was adjusted to produce no more than 1500 colonies/Q-Tray. Plates were incubated at 37 °C overnight. White colonies were picked into 384-well plates using a Q-Bot (Genetix). Each well in the 384-well plates contained 80 μL freezing media with 12.5 μg/mL of chloramphenicol (Woo et al. 1994). Plates were incubated overnight at 37 °C and stored at –80 °C. Picked clones from each library were also gridded on 22.5 cm × 22.5 cm Biodyne Plus membrane filters (Pall-Gelman, Ann Arbor, Mich.) using a Q-Bot. Each clone was double-spotted in 4 × 4 arrays, allowing representation of 18 432 different clones per filter; each library fit onto 2 filters.

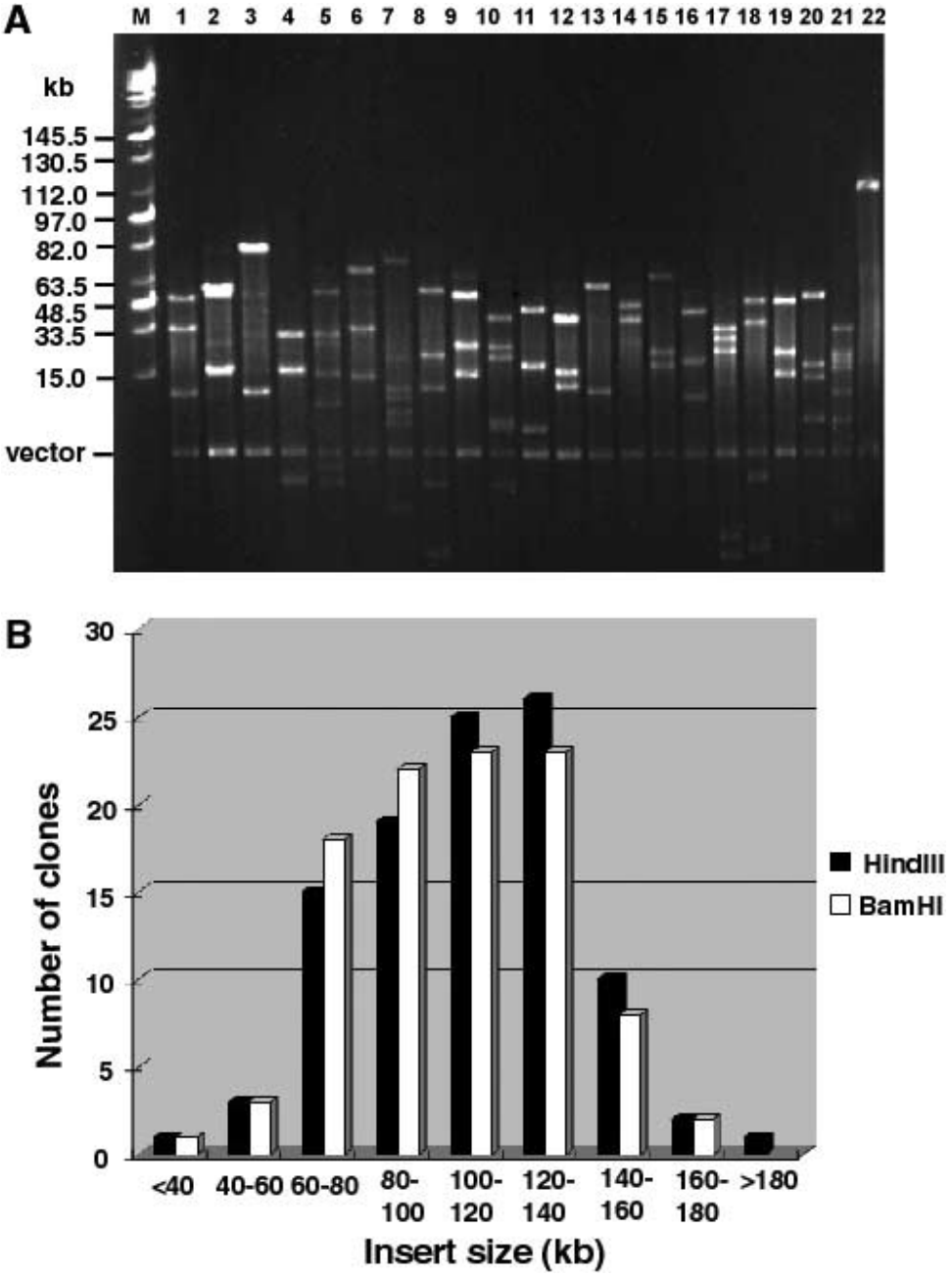
BAC insert analysis

One hundred random BACs from each library were selected for estimating the average insert size of the library. BAC DNA was isolated using a REAL Prep 96 Plasmid Kit (Qiagen, Valencia, Calif.). Purified BAC DNA was digested with *Not*I to release the BAC insert (New England BioLabs, Beverly, Mass.) and then separated by PFGE in 1% *w*/*v* agarose gel in 0.5× TBE buffer. The sizes of BAC inserts were determined based on comparison with MidRange I PFG markers (New England BioLabs).

BAC library screening

Sequences of genes involved in starch biosynthesis in wheat and barley were used to BLAST against *Brachypodium* ESTs derived from leaf, callus, stem, and seed cDNA libraries (Vogel et al. 2006*a*). Eight *Brachypodium* ESTs (Table 1) were selected as probes to screen the *Brachypodium* BAC library. Inserted DNA fragments were prepared free of plasmid sequence by excising the insert with a restriction enzyme and subsequent gel purification using a QIAquick Gel Extraction Kit (Qiagen). The level of chloroplast contamination for the libraries was determined by hybridization of a single high-density BAC filter from each library with fragments of the *rbcL* and *petD* chloroplast genes as described previously (Cenci et al. 2003) . DNA was labeled with ^{32}P dCTP using a DECAprime™ II Random Priming DNA Labeling Kit (Ambion, Austin, Tex.) following the manufacturer's instructions. Hybridization was carried out at 65 °C overnight in a buffer containing 0.36 mol/L Na_2HPO_4, 0.14 mol/L NaH_2PO_4, 7% sodium dodecyl sulfate (SDS), 1% bovine serum albumi (BSA), 3 mmol/L EDTA, and 0.1% sheared salmon sperm DNA. Filters were washed at 65 °C in a solution containing 0.5× SSC and 0.1% SDS.

Fig. 1. Characterization of randomly selected *Brachypodium distachyon* BAC clones. (A) Sizing of BAC inserts. DNA was extracted from randomly selected clones and digested with *Not*I, then resolved by PFGE. Lanes 1– 22 show some typical *Not*I digestion patterns of BAC clones. Lane M contains MidRange I PFG markers (New England Biolabs). A 7.4 kb cloning vector band is indicated. (B) Insert size distribution of the *Hin*dIII and *Bam*HI BAC libraries. The results are based on the sizing of 100 random clones by PFGE for each library.



Southern hybridization analysis

Genomic DNA (~5 μg) of *Brachypodium* line Bd21 was digested with *Hin*dIII, *Eco*RI, *Bam*HI, *Sac*II (New England BioLabs), and *Sst*I (Invitrogen, Carlsbad, Calif.) at concentrations of 2 U/μg DNA, separated in a 0.8% *w/v* agarose gel, and transferred to Hybond N+ membranes (Amersham, Piscataway, N.J.). The filter blot was individually probed with each ^{32}P-labeled EST insert fragment. Hybridization was performed as described for BAC library screening.

BAC-end sequencing and sequence analysis

For BAC-end sequencing, 5 μL of BAC DNA (~0.2–0.5 μg), isolated using a REAL Pre 96 Plasmid Kit, was used in a sequencing reaction with BigDye version 3.1 (Applied Biosystems, Foster City, Calif.). Template DNA was sequenced from both directions with pCC1BAC–pIndigoBAC-5 forward and reverse end-sequencing primer (Epicentre). The trace files generated by an ABI 3730xl genetic analyzer were quality-analyzed with the PHRED software package (http://www.phrap.com). Vector sequence was removed using the

Fig. 2. Genomic Southern hybridizations with starch biosynthesis genes. *Brachypodium* genomic DNA was extracted from leaf tissues and digested with *Hin*dIII (lane 1), *Eco*RI (lane 2), *Bam*HI (lane 3), *Sac*II (lane 4), and *Sst*I (lane 5). The digested DNA was resolved by 0.8% *w*/*v* agarose gel and transferred to a Hybond N+ membrane. Duplicate blots were individually probed with the starch biosynthesis genes described in the text. Only the hybridization results with SBEI (A) and GBSSII (B) probes are shown. λ DNA digested with *Hin*dIII was used as size standards as indicated.

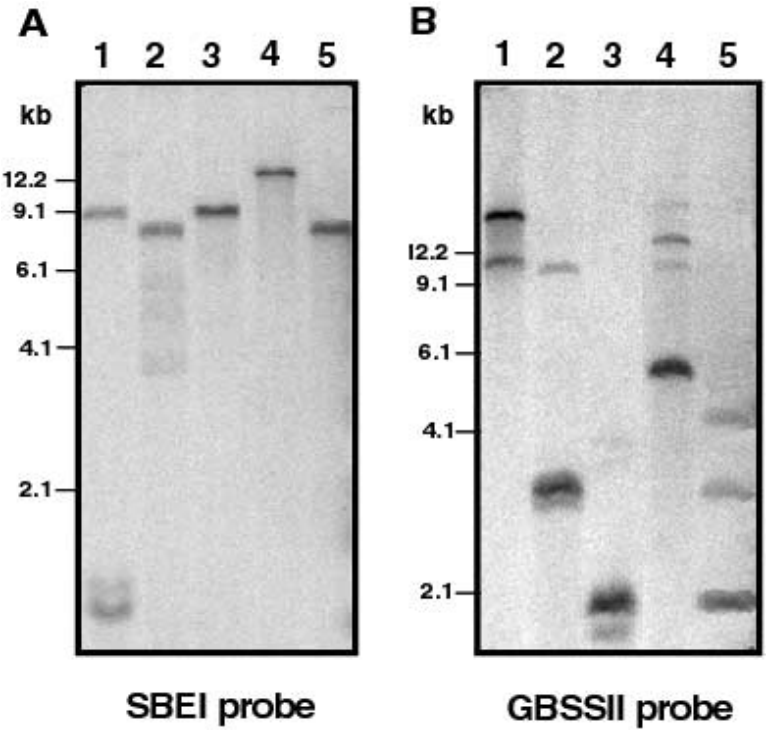


Table 2. Number and percentage (in parentheses) of BESs that have matches in BLAST searches at different *E*-value cutoffs.

Databases	$\leq 10^{-10}$	$\leq 10^{-30}$	$\leq 10^{-50}$
dbEST	866(40.0)	533(25.3)	360(16.5)
Brachypodium EST	295(13.5)	203(9.3)	137(6.3)
Wheat EST	502(23.0)	328(15.0)	203(9.3)
Maize EST	324(14.8)	179(8.2)	92(4.2)
Rice EST	393(18.0)	209(9.6)	112(5.1)
Rice GAAS	658(30.1)	401(18.4)	255(11.7)
GIRI repeat DNA	72(3.3)	31(1.4)	14 (0.6)

CROSS_MATCH software with minmatch 10 and minscore 20 settings (Ewing et al. 1998; Lazo et al. 2004).

General steps of database comparison for the quality BAC-end sequences (BES) were first compared with the GenBank nonredundant database to identify matches to known sequences, and then to individual EST databases using the BLASTn algorithm. Both the expect value (*E* value) and the bit score of the best match for each BES were extracted. At the time the BLAST searches were conducted, the entire EST database contained 19 872 141 sequences, whereas the wheat, maize, rice, and *Brachypodium* EST databases contained 600 205, 686 395, 407 545, and 20 440 sequences, respectively. The matches in each BLAST search were divided into 3 categories based on *E* value thresholds of $\leq 10^{-10}$, $\leq 10^{-30}$, and $\leq 10^{-50}$. The number of BESs that fell into each category was counted. When different bit scores of 75, 125, and 200 were used to categorize sequence matches, similar results were obtained as compared with those using different *E* value cutoffs.

Table 3. Distribution of *Brachypodium* BAC-end sequences anchored to individual rice chromosomes.

Rice chromosome	Length (Mb) of chromosome	Predicted rice genes	No. of *Brachypodium* BES matches
1	43.26	4856	81
2	35.95	3964	76
3	36.19	4159	72
4	36.49	3400	67
5	29.73	2956	56
6	30.73	3079	56
7	29.64	3044	43
8	28.43	2708	46
9	22.69	2175	43
10	22.68	2185	34
11	28.38	2650	37
12	27.56	2368	47

Brachypodium BESs were also used in a homology search against the RiceGAAS (http://RiceGAAS.dna.affrc.go.jp/) and mapped wheat ESTs to determine matches to mapped sequences in rice and wheat. The RiceGAAS dataset is the culmination of genome annotation of the BACs for the rice genome, each with annotated gene predictions and its location against the rice genome (Sakata et al. 2002). When a collection of 7772 ESTs from wheat were used as probes, over 16 000 loci were mapped to the wheat genome (Qi et

Fig. 3. In silico mapping of *Brachypodium* BAC clones to wheat chromosome bins. BESs of *Brachypodium* clones were used to conduct BLAST searches against a set of mapped wheat ESTs (Qi et al. 2004). BESs that found matches with $E < 10^{-10}$ were assigned to individual wheat chromosome bins based on the corresponding map positions of matched wheat ESTs. The color in the heat map represents the relative density of mapped wheat ESTs in a bin. The number to the right of the bin represents the number of matched BESs. The number in the bracket represents matches to a region larger than a single bin. The number below the chromosome represents BESs that were matched to the chromosome only and not placed in a bin.

al. 2004); a general comparison of BESs to the mapped ESTs was performed in a BLAST search using *E* value threshold of 10^{-10}. In addition, the proportion of BESs represented by known repetitive elements was estimated by a search against the GIRI repeat database (http://www.girinst.org) that contains a collection of repetitive elements from various organisms, including rice, maize, wheat, and barley (Jurka et al. 1996).

Results and discussion

BAC library construction and characterization

In an effort to develop critical genomics resources for *Brachypodium*, we constructed 2 BAC libraries from the inbred diploid line Bd21, previously used for the construction of cDNA libraries and EST sequencing (Vogel et al. 2006*b*). The libraries were constructed using different restriction enzymes, *Hin*dIII and *Bam*HI, to avoid the bias and cloning artifacts associated with BAC libraries constructed using single restriction enzymes. To estimate the average insert size, DNA purified from randomly selected BAC clones was digested with *Not*I and DNA fragments were resolved by PFGE (Fig. 1A). The average insert sizes for the *Hin*dIII and *Bam*HI libraries were 105 and 100 kb, respectively, with insert sizes ranging from 30 to 200 kb. Based on this analysis, over 80% of the BAC clones were shown to carry a DNA insert larger than 80 kb, and only 3% had inserts smaller than 50 kb (Fig. 1B). The total percentage of empty clones for each library was estimated to be 4.6% and 5.1%, respectively. Because *Not*I recognizes an 8 bp GC sequence, digestion typically generated between 1 and 6 fragments per BAC clone insert (Fig. 1A), suggesting the *B. distachyon* genome might be moderately GC rich like other monocot genomes (Wang et al. 1995; Woo et al. 1994).

To evaluate the level of chloroplast contamination in the libraries, one BAC filter from each *Hin*dIII and *Bam*HI library was hybridized with probes corresponding to the wheat chloroplast genes *rbcL* and *petD*. Hybridization of these filters showed that 2.4% of the *Hin*dIII library clones and 4.4% of the *Bam*HI library clones were contaminated by cpDNA. Individually, the 2 libraries theoretically represent 9.9- and 9.4-fold coverage of the *Brachypodium* genome based on a genome size of 355 Mb (Bennett and Leitch 2005; Vogel et al. 2006*a*). The coverage provides a 99.99% probability of recovering any gene sequence present in the genome from either of the *Brachypodium* BAC libraries (Clarke and Carbon 1976).

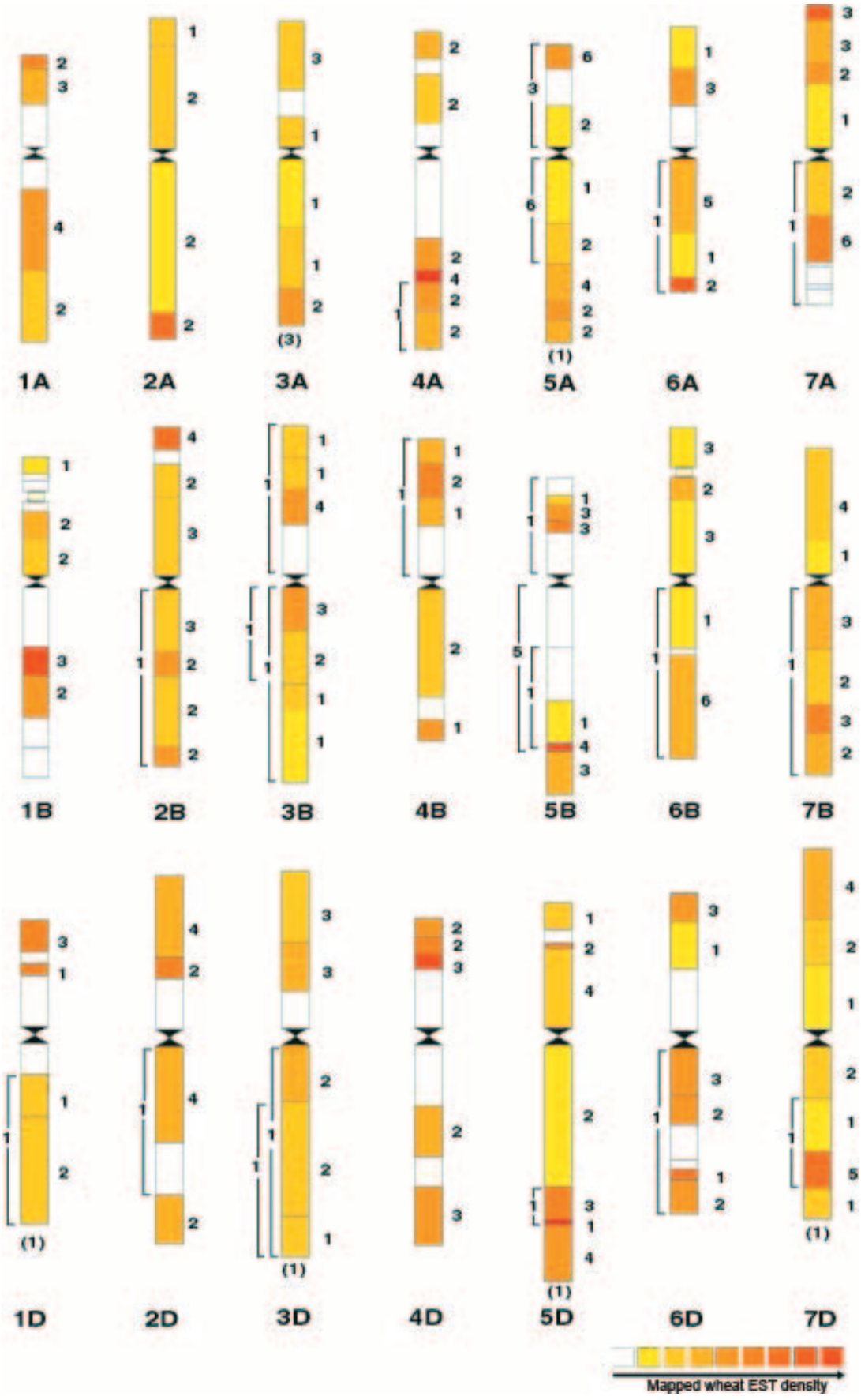
BAC library screening and genomic hybridization with starch biosynthesis genes

To further validate the theoretical genome coverage of each BAC library, we hybridized the BAC libraries with probes for genes with well-characterized functions. Genes encoding enzymes involved in starch biosynthesis have been extensively studied (Tetlow et al. 2004). Because of the close evolutionary relationship between *Brachypodium* and Triticeae species, we used 10 wheat and barley starch biosynthesis gene sequences for a BLAST search against a collection of 20 440 *Brachypodium* ESTs (Vogel et al. 2006*b*). Eight wheat and barley starch biosynthesis genes had strong matches to the *Brachypodium* EST sequences (Table 1). These 8 *Brachypodium* ESTs were selected and used as probes in genomic Southern hybridization analysis to determine the copy number of the corresponding genes in the *Brachypodium* genome (Fig. 2). Among them, homologues of the wheat *SBE1* (Rahman et al. 1999), *isoamylase* (Rahman et al. 2003), and β-*amylase* (GenBank accession No. X98504) are single-copy genes, whereas the other genes, *GBSSI* (Clark et al. 1991), *GBSSII* (Vrinten and Nakamura 2000), *SBEII* (Rahman et al. 2001), and α-*amylase* (Whittier et al. 1987), each have 2 copies in the *Brachypodium* genome (Table 1). Only *AGPase* (Ali et al. 2000) has 3 copies (Table 1). The remaining 2 genes, *Ss2a-1* and *Ss2a-2*, which encode the same class of wheat starch synthase isoforms (Gao and Chibbar 2000), did not show any significant matches in the current *Brachypodium* EST collection and were not used in the analysis.

The *Brachypodium* EST probes were then used to screen a BAC filter containing 18 432 BAC clones from each library. A total of 88 and 151 clones were recovered from the *Bam*HI and *Hin*dIII libraries, respectively (Table 1). As expected, the number of positive BAC clones for each gene varied from 3 for both the *isoamylase* and β-*amylase* genes to 43 for *AGPase*. Using the genome copy number for each of the 8 genes determined by Southern hybridization analysis and the number of positive clones recovered in the BAC screening, we empirically calculated the genome coverage for the *Bam*HI and *Hin*dIII libraries to be 11.6 and 19.6, respectively. The empirical data is higher than that estimated based on average insert size of BAC clones, particularly for the *Hin*dIII BAC library. This may simply be due to a sampling variation. However, since we sampled 8 genes, an alternative explanation is that the *Brachypodium* genome is smaller than the estimated value (355 Mb) we used to predict genome coverage. It has been noted that different C values have been reported for the diploid *B. distachyon* genome, with estimated genome sizes ranging from 294 to 355 Mb (Bennett and Leitch 2005; Vogel et al. 2006*a*). A closer estimate of the true genome size will depend on more experimental information from a whole-genome physical map or the sequence of the entire genome.

BAC-end sequencing and data analysis

It has been reported that genomes of *Brachypodium* spe-



cies contain less than 15% highly repetitive DNA (Catalan et al. 1995). To obtain a first view of the sequence composition of this compact genome, we performed a pilot BAC-end sequencing of the library clones. A total of 2185 high-quality BESs were generated from the 2 BAC libraries. The high-quality sequences were defined as those having Phred scores ≥ 20 and read lengths of over 100 bp after removing the vector and organelle-contaminated sequences. The total length of BESs, derived from the *Brachypodium* nuclear genome, was 1.3 Mb, with an average length of 563 bp/BES. These BESs represent a random sample of the genomic sequences of *Brachypodium* species. The GC content of the BESs is 45.2%, which is very similar to those obtained previously for other monocot genomes (SanMiguel et al. 2002; Yu et al. 2002).

The BESs were used in BLAST searches against several databases including the GIRI repeat sequence database, various EST databases, and the rice genome database (RiceGAAS) (http://ricegaas.dna.affrc.go.jp/). When the sequences were used to perform BLAST searches against the NCBI EST database (excluding human and mouse ESTs), 40.0% of the sequences had matches with *E* values $<10^{-10}$ (Table 2). Even at $E < 10^{-50}$, 16.5% of BESs still found a match, suggesting that a considerable portion of the *Brachypodium* genome is similar to known transcribed sequences. To evaluate the amount of repetitive sequences in the genome, a search was performed against the GIRI repeat database, which contains a collection of repeat sequences from various organisms including *Arabidopsis*, rice, maize, wheat, and barley (http://www.girinst.org/). Seventy-two BESs had matches to the repeat sequences with *E* values $< 10^{-10}$, representing only 3.3% of the BESs. Although this result suggests a much lower content of repetitive DNA in the *Brachypodium* genome than in other cereal genomes (Akhunov et al. 2005; Haberer et al. 2005; Mao et al. 2000; Peterson et al. 2002), it is also likely that this number is underestimated since repeat sequences unique to the *Brachypodium* genome are not represented in the GIRI repeat database and may have been missed in the BLAST search. Large-scale sequencing of genomic regions in *Brachypodium* will help us further understand the composition and evolution of repeat DNA in this compact genome.

The BESs were also used to search against individual *Brachypodium*, wheat, maize, and rice EST databases. At the time the BLAST search was performed, the number of ESTs present in each collection was 20 440 for *Brachypodium*, 600 205 for wheat (*Triticum aestivum*), 686 395 for maize (*Zea mays*), and 407 545 for rice (*Oryza sativa*). Despite the fact that *Brachypodium* has by far the smallest number of ESTs in the database, over 10% of the BESs have matches to the *Brachypodium* EST sequences at $E < 10^{-10}$. At $E < 10^{-50}$, 6.3% still find a match to the *Brachypodium* EST sequences. Maize has the highest number of ESTs in the database; however, nearly twice as many BESs had matches to wheat than to maize. Although it is possible that the ESTs sampled from different cDNA libraries for wheat and maize may have influenced the BLAST result, given the considerable difference in the number of matches at every different *E*-value cutoff, our data support that wheat and *Brachypodium* are more closely related. Despite the fact that the BESs have more matches to wheat than to rice ESTs, it is difficult to predict which species has a closer relationship to *Brachypodium* based on our BLAST results, since rice has a smaller EST collection in the database. However, our previous studies using a set of EST sequences in phylogenetic analysis has revealed that *Brachypodium* is more closely related to wheat and barley than it is to rice (Vogel et al. 2006*b*). This result supports previous phylogenic studies (Kellogg 2001).

The *Brachypodium* BESs were also used to perform BLAST search against the completed rice genome sequence. We found that almost 12% more sequences have matches to the rice genome sequence at $E < 10^{-10}$ compared with the BLAST search against the rice EST database, suggesting that these sequences are not transcribed or their expression is low and thus they are not represented in the rice dbEST collection. These sequences were further examined by searching the annotation of the corresponding regions in the rice genome. In most cases, these regions were annotated to contain putative protein-encoding sequences.

The availability of the complete rice genome sequence allows us to in silico map *Brachypodium* BESs to 12 individual rice chromosomes on the basis of $E < 10^{-10}$ (Table 3). Rice chromosomes with larger sizes and greater numbers of genes tended to have more anchored *Brachypodium* BESs, suggesting that matched sequences are likely related to gene sequences. Taken together, our data indicate that BAC end sequencing is an efficient and informative approach to anchor markers to BAC clones. When more *Brachypodium* BESs are generated, an alignment of these sequences to the rice genome will not only provide a direct comparison of genome structure and organization between *Brachypodium* and rice, but also will be useful in developing a BAC-based physical map for the *Brachypodium* genome (Chen et al. 2004; Vollrath and Jaramillo-Babb 1999).

Brachypodium has been proposed as a model system for temperate grasses including important crops with large genomes such as wheat and barley. In wheat, 7104 wheat EST unigenes have been mapped into a chromosome bin map using a set of wheat aneuploids and deletion stocks (Qi et al. 2004). We searched the BESs against this set of mapped wheat ESTs. A significant match might suggest that the BAC insert represents a *Brachypodium* region that is colinear to the wheat region containing the mapped EST. There are 122 significant matches to different mapped wheat probe sequences. The positions with respect to the wheat A, B, and D genomes could be placed on 316 different loci (112 loci in the A genome, 110 in the B genome, and 98 in the D genome). Figure 3 shows the tentative assignment of 98 *Brachypodium* BAC clones into individual chromosome bin positions from 21 wheat chromosomes. It appears that the *Brachypodium* BACs are relatively evenly distributed across the wheat genome, supporting the randomness of our pilot BAC-end sequencing. These BAC clones will be useful in identifying additional markers near the mapped EST regions by BAC sequencing if detailed sequence information is required for a specific region (Ramakrishna and Bennetzen 2003).

Application of the *Brachypodium* BAC libraries

The 2 deep-coverage *Brachypodium* BAC libraries con-

structed with *Hin*dIII and *Bam*HI restriction enzymes will provide a useful resource for genomics studies on this emerging model species. In wheat, map-based cloning of genes controlling important traits still represents a great challenge because of its extremely large genome size, high content of repetitive sequences, and polyploid nature (3 related subgenomes). Thus far, rice has served as a model for monocots. However, colinearity is often disturbed owing to rapid sequence rearrangements during genome evolution (Guyot et al. 2004; Li and Gill 2002). Recently, the isolation and characterization of the *Ph1* locus for chromosome pairing in polyploid wheat proved the usefulness of a second model species for comparative genomics and map-based cloning of genes in large and complex genomes (Griffiths et al. 2006). In the *Ph1* region, more orthologous genes were identified between the related species *B. sylvaticum* and wheat as compared with rice and wheat. This work also demonstrated that the genus *Brachypodium* will be more useful for additional marker development in the *Ph1* locus region, since sequences of *B. sylvaticum* can directly be used as probes for mapping through genomic Southern hybridization analyses. Several studies have been directed toward the development of genomics tools for *Brachypodium* (Chen et al. 2004; Christiansen et al. 2005; Draper et al. 2001; Foote et al. 2004; Hasterok et al. 2004; Hasterok et al. 2006; Vogel et al. 2006*a*; Vogel et al. 2006*b*). The 2 BAC libraries constructed in this study will provide additional genomics resources for promoting *Brachypodium* as a model species for temperate grasses.

To increase the utility of genomic resources for this model species, an integrated physical and genetic map for *Brachypodium* needs to be developed. BAC-based physical maps can be readily generated by fingerprinting deep-coverage BAC libraries and subsequently assembling overlapping BACs to form contigs spanning large chromosome regions (Zhang and Wu 2001). Such research is underway in our laboratories. In addition, the analysis of BESs shows that a high percentage of BESs has significant matches to gene-related sequences. The BAC end sequences can serve as sequence-tag-connectors (STC) to anchor the BAC-based physical map. A BAC-based physical map with STC will be useful not only for accessing the *Brachypodium* genome for comparative analysis, but also will facilitate the development of an STC framework for sequencing the completed genome (Vollrath and Jaramillo-Babb 1999). Clones and BAC filters from each library are available upon request from the authors.

Acknowledgements

We thank C.C. Crossman for technical assistance in BAC-end sequencing and Xiaohua He for careful reading of the manuscript. This work was supported in part by USDA–ARS CRIS 532502100-010 and 532502100-011.

References

Akhunov, E.D., Akhunova, A.R., and Dvorak, J. 2005. BAC libraries of *Triticum urartu, Aegilops speltoides* and *Ae. tauschii*, the diploid ancestors of polyploid wheat. Theor. Appl. Genet. **111**: 1617–1622. doi:10.1007/s00122-005-0093-1. PMID:16177898.

Ali, S., Holloway, B., and Taylor, W.C. 2000. Normalisation of cereal endosperm EST libaries for structural and functional genomic analysis. Plant Mol. Biol. Rep. **18**: 123–132.

Bennett, M.D., and Leitch, I.J. 2005. Nuclear DNA Amounts in Angiosperms: Progress, Problems and Prospects. Ann. Bot. (London), **95**: 45–90. PMID:15596457.

Catalan, P., Shi, Y., Armstrong, L., Draper, J., and Stace, C.A. 1995. Molecular phylogeny of the grass genus Brachypodium P-Beauv based on RFLP and RAPD analysis. Bot. J. Linn. Soc. **117**: 263–280. doi:10.1006/bojl.1995.0018.

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

Chen, R., Sodergren, E., Weinstock, G.M., and Gibbs, R.A. 2004. Dynamic building of a BAC clone tiling path for the rat genome sequencing project. Genome Res. **14**: 679–684. doi:10.1101/gr.2171704. PMID:15060010.

Christiansen, P., Andersen, C.H., Didion, T., Folling, M., and Nielsen, K.K. 2005. A rapid and efficient transformation protocol for the grass *Brachypodium distachyon.* Plant Cell Rep. **23**: 751–758. doi:10.1007/s00299-004-0889-5. PMID:15503032.

Clark, J.R., Robertson, M., and Ainsworth, C.C. 1991. Nucleotide sequence of a wheat (*Triticum aestivum* L.) cDNA clone encoding the waxy protein. Plant Mol. Biol. **16**: 1099–1101. doi:10.1007/BF00016086. PMID:1863765.

Clarke, L., and Carbon, J. 1976. A colony bank containing synthetic CoI EI hybrid plasmids representative of the entire *E. coli* genome. Cell, **9**: 91–99. doi:10.1016/0092-8674(76)90055-6. PMID:788919.

Devos, K.M., and Gale, M.D. 1997. Comparative genetics in the grasses. Plant Mol. Biol. **35**: 3–15. doi:10.1023/A:1005820229043. PMID:9291955.

Draper, J., Mur, L.A., Jenkins, G., Ghosh-Biswas, G.C., Bablak, P., Hasterok, R., and Routledge, A.P. 2001. *Brachypodium distachyon*. A new model system for functional genomics in grasses. Plant Physiol. **127**: 1539–1555. doi:10.1104/pp.127.4.1539. PMID:11743099.

Ewing, B., Hillier, L., Wendl, M.C., and Green, P. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. **8**: 175–185. PMID:9521921.

Foote, T.N., Griffiths, S., Allouis, S., and Moore, G. 2004. Construction and analysis of a BAC library in the grass *Brachypodium sylvaticum*: its use as a tool to bridge the gap between rice and wheat in elucidating gene content. Funct. Integr. Genomics, **4**: 26–33. PMID:14727148.

Gao, M., and Chibbar, R.N. 2000. Isolation, characterization, and expression analysis of starch synthase IIa cDNA from wheat (*Triticum aestivum* L.). Genome, **43**: 768–775. doi:10.1139/gen-43-5-768. PMID:11081966.

Gaut, B.S. 2002. Evolutionary dynamics of grass genomes. New Phytol. **154**: 15–28. doi:10.1046/j.1469-8137.2002.00352.x.

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 (London), **439**: 749–752. doi:10.1038/nature04434. PMID:16467840.

Guyot, R., Yahiaoui, N., Feuillet, C., and Keller, B. 2004. In silico comparative analysis reveals a mosaic conservation of genes within a novel colinear region in wheat chromosome 1AS and rice chromosome 5S. Funct. Integr. Genomics, **4**: 47–58. doi:10.1007/s10142-004-0103-4. PMID:14767678.

Haberer, G., Young, S., Bharti, A.K., Gundlach, H., Raymond, C., Fuks, G., et al. 2005. Structure and architecture of the maize

genome. Plant Physiol. **139**: 1612–1624. doi:10.1104/pp.105.068718. PMID:16339807.

Hasterok, R., Draper, J., and Jenkins, G. 2004. Laying the cytotaxonomic foundations of a new model grass, *Brachypodium distachyon* (L.) Beauv. Chromosome Res. **12**: 397–403. doi:10.1023/B:CHRO.0000034130.35983.99. PMID:15241018.

Hasterok, R., Marasek, A., Donnison, I.S., Armstead, I., Thomas, A., King, I.P., et al. 2006. Alignment of the genomes of *Brachypodium distachyon* and temperate cereals and grasses using BAC landing with fluorescent in situ hybridization. Genetics, **173**: 349–362. doi:10.1534/genetics.105.049726. PMID:16489232.

Jurka, J., Klonowski, P., Dagman, V., and Pelton, P. 1996. CENSOR–a program for identification and elimination of repetitive elements from DNA sequences. Comput. Chem. **20**: 119–121. doi:10.1016/S0097-8485(96)80013-1. PMID:8867843.

Kellogg, E.A. 2001. Evolutionary History of the Grasses. Plant Physiol. **125**: 1198–1205. doi:10.1104/pp.125.3.1198. PMID:11244101.

Lazo, G.R., Chao, S., Hummel, D.D., Edwards, H., Crossman, C.C., Lui, N., et al. 2004. Development of an expressed sequence tag (EST) resource for wheat (*Triticum aestivum* L.): EST generation, unigene analysis, probe selection and bioinformatics for 16,000-locus bin-delineated map. Genetics, **168**: 585–593. doi:10.1534/genetics.104.034777. PMID:15514037.

Li, W., and Gill, B.S. 2002. The colinearity of the *Sh2/A1* orthologous region in rice, sorghum and maize is interrupted and accompanied by genome expansion in the triticeae. Genetics, **160**: 1153–1162. PMID:11901130.

Mao, L., Wood, T.C., Yu, Y., Budiman, M.A., Tomkins, J., Woo, S.-S., et al. 2000. Rice transposable elements: a survey of 73,000 sequence-tagged-connectors. Genome Res. **10**: 982–990. doi:10.1101/gr.10.7.982. PMID:10899147.

Peterson, D.G., Schulze, S.R., Sciara, E.B., Lee, S.A., Bowers, J.E., Nagel, A., et al. 2002. Integration of Cot analysis, DNA cloning, and high-throughput sequencing facilitates genome characterization and gene discovery. Genome Res. **12**: 795–807. doi:10.1101/gr.226102. PMID:11997346.

Qi, L.L., Echalier, B., Chao, S., Lazo, G.R., Butler, G.E., Anderson, O.D., et al. 2004. A chromosome bin map of 16,000 expressed sequence tag loci and distribution of genes among the three genomes of polyploid wheat. Genetics, **168**: 701–712. doi:10.1534/genetics.104.034868. PMID:15514046.

Rahman, S., Li, Z., Abrahams, S., Abbott, D., Appels, R., and Morell, M.K. 1999. Characterization of a gene encoding wheat endosperm starch branching enzyme-I. Theor. Appl. Genet. **98**: 156–163. doi:10.1007/s001220051052.

Rahman, S., Nakamura, Y., Li, Z., Clarke, B., Fujita, N., Mukai, Y., et al. 2003. The sugary-type isoamylase gene from rice and *Aegilops tauschii*: characterization and comparison with maize and *Arabidopsis*. Genome, **46**: 496–506. doi:10.1139/g02-130. PMID:12834068.

Rahman, S., Regina, A., Li, Z., Mukai, Y., Yamamoto, M., Kosar-Hashemi, B., Abrahams, S., and Morell, M.K. 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. doi:10.1104/pp.125.3.1314. PMID:11244112.

Ramakrishna, W., and Bennetzen, J.L. 2003. Genomic colinearity as a tool for plant gene isolation. Methods Mol. Biol. **236**: 109–122. PMID:14501061.

Ramakrishna, W., Dubcovsky, J., Park, Y.J., Busso, C., Emberton, J., SanMiguel, P., and Bennetzen, J.L. 2002. Different types and rates of genome evolution detected by comparative sequence analysis of orthologous segments from four cereal genomes. Genetics, **162**: 1389–1400. PMID:12454082.

Sakata, K., Nagamura, Y., Numa, H., Antonio, B.A., Nagasaki, H., Idonuma, A., et al. 2002. RiceGAAS: an automated annotation system and database for rice genome sequence. Nucleic Acids Res. **30**: 98–102. doi:10.1093/nar/30.1.98. PMID:11752265.

SanMiguel, P.J., Ramakrishna, W., Bennetzen, J.L., Busso, C.S., and Dubcovsky, J. 2002. Transposable elements, genes and recombination in a 215-kb contig from wheat chromosome 5A(*m*). Funct. Integr. Genomics, **2**: 70–80. doi:10.1007/s10142-002-0056-4. PMID:12021852.

Sorrells, M.E., La Rota, M., Bermudez-Kandianis, C.E., Greene, R.A., Kantety, R., Munkvold, J.D., et al. 2003. Comparative DNA sequence analysis of wheat and rice genomes. Genome Res. **13**: 1818–1827. PMID:12902377.

Tetlow, I.J., Morell, M.K., and Emes, M.J. 2004. Recent developments in understanding the regulation of starch metabolism in higher plants. J. Exp. Bot. **55**: 2131–2145. doi:10.1093/jxb/erh248. PMID:15361536.

Vogel, J., Garvin, D., Leong, O., and Hayden, D. 2006*a*. *Agrobacterium*-mediated transformation and inbred line development in the model grass *Brachypodium distachyon.* Plant Cell Tissue Organ Cult. **84**: 100179–100191. doi:10.1007/s11240-005-9023-9.

Vogel, J.P., Gu, Y.Q., Twigg, P.,s Lazo, G.R., Laudencia-Chingcuanco, D., Hayden, D.M., et al. 2006*b*. EST sequencing and phylogenetic analysis of the model grass *Brachypodium distachyon*. Theor. Appl. Genet. **113**: 186–195.

Vollrath, D., and Jaramillo-Babb, V.L. 1999. A sequence-ready BAC clone contig of a 2.2-Mb segment of human chromosome 1q24. Genome Res. **9**: 150–157. PMID:10022979.

Vrinten, P.L., and Nakamura, T. 2000. Wheat granule-bound starch synthase I and II are encoded by separate genes that are expressed in different tissues. Plant Physiol. **122**: 255–264. doi:10.1104/pp.122.1.255. PMID:10631269.

Wang, G.L., Holsten, T.E., Song, W.Y., Wang, H.P., and Ronald, P.C. 1995. Construction of a rice bacterial artificial chromosome library and identification of clones linked to the Xa-21 disease resistance locus. Plant J. **7**: 525–533. doi:10.1046/j.1365-313X.1995.7030525.x. PMID:7757120.

Whittier, R.F., Dean, D.A., and Rogers, J.C. 1987. Nucleotide sequence analysis of alpha-amylase and thiol protease genes that are hormonally regulated in barley aleurone cells. Nucleic Acids Res. **15**: 2515–2535. PMID:3031602.

Woo, S.S., Jiang, J., Gill, B.S., Paterson, A.H., and Wing, R.A. 1994. Construction and characterization of a bacterial artificial chromosome library of Sorghum bicolor. Nucleic Acids Res. **22**: 4922–4931. PMID:7800481.

Yu, J., Hu, S., Wang, J., Wong, G.K., Li, S., Liu, B., et al. 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). Science (Washington, D.C.), **296**: 79–92. doi:10.1126/science.1068037. PMID:11935017.

Zhang, H.-B., and Wu, C. 2001. BAC as tools for genome sequencing. Plant Physiol. Biochem. **39**: 195–209. doi:10.1016/S0981-9428(00)01236-5.

Zhang, H.-B., Zhao, X., Ding, X., Paterson, A.H., and Wing, R.A. 1995. Preparation of megabase-size DNA from plant nuclei. Plant J. **7**: 175–184. doi:10.1046/j.1365-313X.1995.07010175.x.