Genetics: Early Online, published on September 13, 2013 as 10.1534/genetics.113.157115

Genomic resources for gene discovery, functional genome annotation, and evolutionary studies of maize and its close relatives

Chao Wang[*,#], Xue Shi[*,#], Lin Liu[*], Haiyan Li[*], Jetty S.S. Ammiraju[§], David A. Kudrna[§], Wentao Xiong[*], Hao Wang[*], Zhaozhao Dai[*], Yonglian Zheng[*], Jinsheng Lai[†], Weiwei Jin[†], Joachim Messing[‡], Jeffrey L Bennetzen[**], Rod A. Wing[§], and Meizhong Luo[*,1]

[*] National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan 430070, China

[§] Arizona Genomics Institute, University of Arizona, Tucson, AZ 85721, USA

[†] State Key Lab of Agrobiotechnology, China Agricultural University, Beijing, China

[‡] Waksman Institute, Rutgers University, Piscataway, NJ 08854, USA

[**] Department of Genetics, University of Georgia, Athens, Georgia, USA

[#] These authors contributed equally to this work.

Maize BAC/BIBAC resources

Keywords: BAC; BIBAC; maize; parviglumis; Tripsacum

[1]Correspondence author: Meizhong Luo, National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan 430070, China. E-mail: mzluo@mail.hzau.edu.cn, Tel/Fax +86 27 87284213

ABSTRACT Maize is one of the most important food crops and a key model for genetics and developmental biology. A genetically anchored and high-quality draft genome sequence of maize inbred B73 has been obtained to serve as a reference sequence. To facilitate evolutionary studies in maize and its close relatives, much like the OMAP (www.OMAP.org) BAC resource did for the rice community, we constructed BAC libraries for maize inbred lines Zheng58, Chang7-2 and Mo17 and maize wild relatives *Zea mays* ssp. *parviglumis* and *Tripsacum dactyloides*. Furthermore, to extend functional genomic studies to maize and sorghum, we also constructed BIBAC libraries for the maize inbred B73 and the sorghum land race Nengsi-1. The BAC/BIBAC vectors facilitate transfer of large intact DNA inserts from BAC clones to the BIBAC vector and functional complementation of large DNA

fragments. These seven ZMAP BAC/BIBAC libraries have average insert sizes ranging from 92kb to 148kb, organellar DNA from 0.17% to 2.3%, empty vector rates between 0.35% and 5.56%, and genome equivalents of 4.7- to 8.4-fold. The usefulness of the *Parviglumis* and *Tripsacum* BAC libraries was demonstrated by mapping clones to the reference genome. Novel genes and alleles present in these ZMAP libraries can now be used for functional complementation studies and positional or homology-based cloning of genes for translational genomics.

Introduction

Maize is one of the most important crops worldwide, already producing more tons of grain than any other plant species, and with continuing expansion of planted acreage in the developed and developing world. It is used as a staple food crop, but also for animal feed, biofuel and various industrial byproducts, such as starches, oil, sugars, sweeteners, beverages, and adhesives. It is also a key model species for research in genetics, epigenetics, genomics, development, physiology and evolution (BENNETZEN and HAKE 2009; SCHNABLE *et al.* 2009; WALBOT 2009). Maize was domesticated about 10,000 years ago from an annual wild teosinte, *Zea mays ssp. parviglumis,* that originated in the balsas river lowland in Mexico (DOEBLEY *et al.* 2006; VAN HEERWAARDEN *et al.* 2011). One feature of contemporary maize is that it is probably the most genetically diverse of all crop species. Analyses of haplotypes, sequences around the *bz, z1C1,* and *adh1* loci and sequences of large genomic regions

all showed striking structural diversity in maize inbred lines (BRUNNER *et al.* 2005; CHIA *et al.* 2012; FU and DOONER 2002; GORE *et al.* 2009; JUNG *et al.* 2004; SONG *et al.* 2002; SONG and MESSING 2003; WANG and DOONER 2006). It also has been shown that the two ancestral progenitors of maize and the progenitor of sorghum split about 11.9 mya and the progenitors of maize hybridized about 4.8 mya to form allotetraploid maize (SWIGONOVA *et al.* 2004). Hybridization resulted in broken and fused chromosomes with the reduced set of today's ten chromosomes (WEI *et al.* 2007). Divergence of the two homologous chromosomal regions continued through loss of duplicated genes and uneven expansion through retrotransposition (BRUGGMANN *et al.* 2006; MESSING *et al.* 2004; WOODHOUSE *et al.* 2010). A very recent study has indicated the nature of the major rearrangements that reduced ancestral *Zea* chromosome numbers from 20 to 10, including which ten centromeres were retained and which ten were lost (WANG and BENNETZEN 2012).

BAC (bacterial artificial chromosome) libraries are important tools for structural, functional and comparative genomics studies of eukaryotic genomes. Hundreds of BAC libraries for crop plants have been constructed to date, such as those for rice and wild rice (AMMIRAJU *et al.* 2006; WANG *et al.* 1995), wheat (CENCI *et al.* 2003; JANDA *et al.* 2006; NILMALGODA *et al.* 2003), maize (O'SULLIVAN *et al.* 2001; TOMKINS *et al.* 2002; YIM *et al.* 2002), soybean (WU *et al.* 2004), sorghum (WOO *et al.* 1994), and pearl millet (ALLOUIS *et al.* 2001). These BAC libraries have been used for developing sequence-ready physical maps (CHEN *et al.* 2002; WEI *et al.* 2009), genome sequencing (INTERNATIONAL RICE GENOME SEQUENCING PROJECT 2005;

SCHNABLE *et al.* 2009), positional cloning of genes of agronomic importance (LI *et al.* 2011; MAO *et al.* 2000; SONG *et al.* 1995), and comparative analysis of genome structures (AMMIRAJU *et al.* 2008; CHEN *et al.* 1997; LU *et al.* 2009; MESSING and LLACA 1998; WANG and DOONER 2012). BAC end sequencing has also been used to survey genome structure (MESSING *et al.* 2004), to assist in (and/or confirm) genome sequence assemblies (BENNETZEN *et al.* 2012), and to develop SSR and SNP markers (QI *et al.* 2001; QI *et al.* 2004). BAC end sequences (BESs) embedded in physical maps have been used to detect structural differences between the target genomes and reference sequence (DENG *et al.* 2013; HURWITZ *et al.* 2010; KIM *et al.* 2007; LIN *et al.* 2012). We have flagged many putative assembly inversion errors of the Nipponbare reference sequence by aligning physical maps of rice ZH11 and 93-11 to the rice Nipponbare reference sequence through BESs (LIN *et al.* 2012) (Pan; Deng et al., manuscript in submission) and confirmed 5 representatives by further detailed analysis (DENG *et al.* 2013).

A binary BAC (BIBAC) vector was developed from BAC vector, and BIBAC libraries or the like (transformation-competent artificial chromosome, TAC) can be used not only for general purposes as BAC libraries but also for *Agrobacterium*-mediated plant transformation of large genomic DNA fragments (HAMILTON 1997; HAMILTON *et al.* 1996; LIU *et al.* 1999). Many BIBAC libraries have been constructed to date, such as those for rice (LIU *et al.* 2002; TAO *et al.* 2002), wheat (LIU *et al.* 2000), soybean (MEKSEM *et al.* 2000), sunflower (FENG *et al.* 2006), cotton (LEE *et al.* 2013), and ginseng (ZHAI *et al.* 2013). Recently, we modified a pair

of BAC and BIBAC vectors coordinately in order to make the construction of BAC/BIBAC libraries easier and to allow simple large intact DNA insert exchange between BAC and BIBAC vectors (SHI *et al.* 2011).

Three BAC libraries containing genomic DNA from maize inbred B73 have been constructed, each containing DNA digested with a different restriction enzyme (YIM *et al.* 2002), and they have been fingerprinted by two complementary methods (NELSON *et al.* 2005). A large number of these clones have also been end-sequenced (MESSING *et al.* 2004). A BAC physical map for B73 was constructed (WEI *et al.* 2009) and genetically anchored, allowing the generation of a BAC-based draft sequence of the maize B73 genome (SCHNABLE *et al.* 2009). However, to functionally annotate the maize genome and make better use of the maize reference sequence in future studies, additional related genomic resources of a broader phylogenetic diversity are needed. As an example, BAC libraries for 12 species of genus *Oryza* (one cultivar and 11 wild species) achieved a better understanding of the diversity of the genus and serves as a resource for new traits in cultivars (AMMIRAJU *et al.* 2006; WING *et al.* 2005). This project is also known as the *Oryza* Map Alignment Project (OMAP; http://www.omap.org). Here, we report the construction and quality analysis of BAC libraries for three elite inbred lines (Zheng58, Chang7-2 and Mo17) and two wild relatives of maize (*Zea mays* ssp. *parviglumis* and *Tripsacum dactyloides*), and BIBAC libraries for maize B73 and sorghum Nengsi-1. To ascertain that the BAC/BIBAC resources are optimal, the inbreds and accessions were selected after input from a group of maize geneticists and evolutionary biologists. Zheng58 and

Chang7-2 are the respective male and female parents of Zhengdan958, the most widely planted hybrid maize in China. They are predicted to be rich in beneficial alleles and thus are appropriate subjects for gene cloning, heterosis and genomic studies (DING *et al.* 2011; JIAO *et al.* 2012; LAI *et al.* 2010). Maize inbred Mo17 represents the genetic background of many current elite inbreds in the U.S. and world seed industry. Mapping populations (B73xMo17) have been used to construct an integrated physical and genetic map of the maize genome and to clone maize QTLs (BARBAZUK and SCHNABLE 2011; EICHTEN *et al.* 2011; FU *et al.* 2006). Although a Mo17 BAC library has been constructed in the industry, it is not publicly available and does not have deep coverage (BRUNNER *et al.* 2005). *Zea mays* ssp. *parviglumis* is the most likely wild progenitor of maize and is a main subject for maize domestication and evolution studies (DOEBLEY 2004; SWANSON-WAGNER *et al.* 2012). *Tripsacum dactyloides* with a large genome size of 3,730 Mb (ARUMUGANATHAN and EARLE 1991; BENNETT and LEITCH 2011) is a member of the Tribe *Maydeae*, closely related to maize, and is phylogenetically well-defined and genetically tractable. It represents an untapped genetic resource for crop improvement and contains many useful agronomic traits like disease and pest resistance and harsh-environment adaptation (CHIA *et al.* 2012; HOISINGTON *et al.* 1999). It is also a natural model system to study parthenogenesis, an important trait for crop improvement (BANTIN *et al.* 2001; HOISINGTON *et al.* 1999). *Tripsacum* can hybridize with maize and the hybrids can grow well (HARLAN and DE WET 1977). B73 has served as a maize genetics and genomics platform worldwide, and is rich in genetics and genomics

resources such as genetic maps, BAC libraries, a physical map, and a high-quality draft reference sequence (FU *et al.* 2006; SCHNABLE *et al.* 2009; WEI *et al.* 2009; YIM *et al.* 2002). However, it lacked a BIBAC library to facilitate functional genomics studies. Sorghum is also the staple food for people in semi-arid tropic areas especially in Sub- Saharan Africa (DICKO *et al.* 2006). It also is superior as a biofuel crop over maize because of lower input cost (CALVINO and MESSING 2012). Both maize and sorghum are C4 plants with high photosynthetic efficiency. Partly because of its smaller genome size (730 Mb), sorghum became a model for research related to maize (BENNETZEN and FREELING 1993), including C4 and drought tolerance research (PATERSON *et al.* 2009; SASAKI and ANTONIO 2009). The high output of grain, biomass and sugars from sorghum can provide new resources for biofuels (BYRT *et al.* 2011; VERMERRIS 2011) in particular from sweet sorghum (CALVINO and MESSING 2012). BAC libraries for sorghum are available (LIN *et al.* 1999; WOO *et al.* 1994). A high quality draft genome sequence of sorghum bicolor BTx623 has been obtained by a whole genome shotgun sequencing approach as discussed above (PATERSON *et al.* 2009). Therefore, a BIBAC library for sorghum has great potential. The sorghum land race Nengsi-1 used here is an important variety in China and has many elite features for both energy and fodder use. The wide selection of maize and related germplasms represented in these libraries will be critical for researchers interested in comparative and functional genomics and for cloning of agriculturally important genes because of the tremendous structural and allelic variation available in these BAC/BIBAC libraries. They should also be of interest to breeders for use in engineering

quantitative traits. Furthermore, variation in both gene and transposable element content will provide a wealth of information to evolutionary biologists. This selection took into consideration the fact that BAC libraries were already available for outgroup species at intermediate evolutionary distances between maize and rice: *Coix lacryma-jobi*, also a member of the Tribe *Maydeae* (MENG *et al.* 2010), *Sorghum*, a member of the closely related Tribe *Andropogonea* (LIN *et al.* 1999; WOO *et al.* 1994) and two other members of the Tribe *Paniceae*, *Pennisetum* (pearl millet) (ALLOUIS *et al.* 2001) and *Setaria italica* (foxtail millet) (BENNETZEN *et al.* 2012) .

The seven BAC/BIBAC libraries are now called the ZMAP (*Zea* Map Alignment Project) libraries, following the designation of *Oryza* BAC libraries as the OMAP collection (AMMIRAJU *et al.* 2006; WING *et al.* 2005). These libraries all have high qualities and comprise a total of 940,032 clones that have been stored in 384-well plates in -80 ℃ freezers in the Genome Resource Laboratory of Huazhong Agricultural University (http://GResource.hzau.edu.cn). Copies of the *Parviglumis* and *Tripsacum* BAC libraries have also been stored at AGI (http://www.genome.arizona.edu). Whole libraries, filters and clones are publicly available through above websites.

Materials and Methods

Plant materials

Seed of *Zea mays* ssp. *parviglumis* (PI 384062, lot number: 83ncso01) and *Tripsacum dactyloides* (PI 595898, lot number: 96ncfo01) were obtained from the USDA germplasm bank at Ames, Iowa (http://www.ars-grin.gov/), where they are reproduced from the originally released line, not modified by introgression of other segments. Seed of maize inbreds Zheng58 and Chang7-2 are available and maintained at Beijing Agricultural University, currently by Professor Jinsheng Lai. Seed of maize Mo17 with an introgressed *Rf3* gene were maintained by Professor Yonglian Zheng. Seed of maize inbred B73 were from the same source as that used for the generation of the B73 reference sequence (TOMKINS *et al.* 2002; WEI *et al.* 2007) and were obtained from the Maize Genetics Cooperation Stock Center (Urbana, IL). Seed of sorghum landrace Nengsi-1 were provided by Dr. Ruiheng Du of Hebei Academy of Agriculture and Forestry Sciences, China. Young leaves of *Zea mays* ssp. *parviglumis*, *Tripsacum*，Zheng58, Chang7-2, and Nengsi-1, and young ears of Mo17 and B73 were used as DNA sources for BAC/BIBAC library construction.

BAC/BIBAC vector preparation

The BAC and BIBAC vectors pIndigoBAC536-S and BIBAC-S were prepared with *Hin*dIII (Fermentas) and *Bam*HI (Fermentas) from the high-copy composite vectors pHZAUBAC1 and pHZAUBIBAC1 (SHI *et al.* 2011), respectively. Vector plasmid DNA was isolated using the plasmid midkit (QIAGEN). The digested products were dephosphorylated with CIP (NEB), self-ligated overnight and separated

on a CHEF-gel at 6 V/cm and 14 ℃ in 0.5 × TBE buffer with a linear ramp time from 1 to 40 s for 16.5 h. The linearized and dephosphorylated pIndigoBAC536-S and BIBAC-S vectors were excised and electroeluted from the gel as previously described (LUO and WING 2003).

BAC/BIBAC library construction

BAC libraries were constructed with *Hin*dIII, as previously described (AMMIRAJU *et al.* 2006; LUO *et al.* 2006; LUO and WING 2003; SONG *et al.* 2011). BIBAC libraries were constructed with *Bam*HI using the same procedure as for BAC library construction except that 20μg/ml kanamycin was used for selection. DH10B T1-resistant cells (Invitrogen) were used for both BAC and BIBAC library construction. Individual BAC/BIBAC colonies were picked and the entire collection was stored in 384-well microtiter plates at -80 ℃. For the B73 BIBAC library, clone pools were made for each 384-well plate to facilitate PCR screening of targeted clones.

Insert size analysis

BAC/BIBAC clones were randomly selected from each library for insert sizing. BAC/BIBAC plasmids were isolated as described by Luo and Wing (LUO and WING

2003), digested with I-*Sce*I and separated on 1% agarose CHEF (CHEF-DRIII apparatus, Bio-Rad) gels at 6 V/cm and 14℃ in 0.5 × TBE buffer with a linear ramp time from 5 to 15 s for 16 h.

Stability analysis of sorghum BIBAC clones in *E. coli* and *Agrobacterium*

Stability tests of sorghum BIBAC clones were conducted following the previously described method for that of maize B73 BIBAC clones (SHI *et al.* 2011). To test the stability of sorghum BIBAC clones in *E. coli*, BIBAC clones maintained in the DH10B T1-resistant host cells were cultured for different times and subjected to insert size analysis, as described above. To test the stability of sorghum BIBAC clones in *Agrobacterium*, we used an indirect method because preparing enough large plasmid DNA directly from *Agrobacterium* is difficult. The BIBAC plasmid DNA was transferred into *Agrobacterium* strain EHA105, extracted after cultured for different times and transferred back into *E.coli* DH10B T1-resistant cells. Then the plasmid DNA was re-extracted from the *E.coli* culture after overnight growth and subjected to insert size analysis.

Probe preparation

Twelve single-copy gene probes (Table S1) were selected according to the maize

B73 reference sequence via MaizeGDB (http://www.maizegdb.org/). They represented all the 10 chromosomes of B73: one from each chromosome except two from chr1 and two from chr3. PCR primers were designed using Primer5. Probes *adh1*, *adh2,* and *ps1* were generated by RT-PCR amplification from B73 mRNA. The other gene probes were amplified by PCR using B73 genomic DNA as the template. All the PCR products were purified from an agarose gel with the MinElute Gel Extraction Kit (Qiagen). Three barley chloroplast probes (*ndhA, rbcL,* and *psbA;* (obtained from J. Mullet, Texas A&M University) were used to estimate the content of chloroplast DNA and four rice mitochondrial probes (*atpA, cob, atp9,* and *coxE;* obtained from T. Sasaki, MAFF, Japan) were used to estimate the content of mitochondrial DNA, as previously described (AMMIRAJU *et al.* 2006; LUO *et al.* 2001).

BAC/BIBAC library screening

The 12 maize gene probes were used to screen the *Parviglumis* and *Tripsacum* BAC libraries, whereas the chloroplast and mitochondrial probes were used to screen a subset of each of all seven BAC/BIBAC libraries. High-density colony filters for the *Parviglumis* and *Tripsacum* BAC libraries were prepared with Nybond-N+ nylon membranes (GE healthcare) using a Q-bot (Genetix) following previously described methods (AMMIRAJU *et al.* 2006; LUO *et al.* 2006; LUO *et al.* 2001). Each 22.5×22.5cm filter holds 18,432 independent clones from 48x384-well plates arrayed

in a 4×4 double spotted pattern in 6 fields. Colony filters with half of the above densities for the other 5 BAC/BIBAC libraries were prepared with Nybond-N+ nylon membranes using a semi-automatic machine (Beckman Coulter Biomek 2000). The filters were prehybridized in pre-warmed hybridization solution (2xSSC) at 55℃ for 3 hours in a hybridization oven at 30 rpm. The probes were labeled with [$^{-32}$P]dCTP (Amersham Biosciences, USA) using a DecaprimeII random prime labeling kit (Ambion), denatured at 95 ℃ for 5 min and added to the hybridization solution for hybridization at 55 ℃ overnight. For the two maize probes *adh1* (probe11) and *adh2* (probe12), single probes were used for each hybridization experiment. For the other ten maize gene probes (probes1-10), pools of 4 probes were used for each hybridization experiment (Table S2). For the three chloroplast probes and the four mitochondrial probes, probe pools were also used. After hybridization, the filters were processed by washing twice with 1.5×SSC, 0.1% SDS at 55 ℃ for 10 minutes each in the hybridization oven at 30 rpm, then once with 1.5×SSC, 0.1% SDS and once with 1×SSC, 0.1% SDS at 55 ℃ for 10 minutes each in a shaker at 100 rpm. The filters were tested for signal strength with a Geiger counter to determine if more washes were needed. After washing, the filters were dried, marked by dotting 8 positions with aliquots of hybridization solution containing labeled probes and exposed to phosphorimager screens at room temperature overnight. The screens were scanned with the Typhoon scanner (GE Healthcare). All putative positive clones of the maize gene probes were located, re-arrayed onto colony filters, and re-hybridized with individual probes labeled with digoxin for confirmation. Probe labeling with digoxin

and colony hybridization with the digoxin-labeled probes followed the commercial protocol in the DIG High Prime DNA labeling and Detection Smart kit I (Roche).

BAC/BIBAC end sequencing, fingerprinting and contig assembly

Selected BAC and BIBAC clones were sequenced at both ends with an ABI 3730 DNA Analyzer using BigDye v3.1 (Applied Biosystems, Foster City, California) as previously described (WANG *et al.* 2013). The two primers BACf (5'aacgacggccagtgaattg3') and BACr (5'gataacaatttcacacagg3'), common to all BAC/BIBAC libraries constructed with our BAC/BIBAC vectors (SHI *et al.* 2011), were used as forward and reverse sequencing primers, respectively. Sequences were base-called using Phred software (EWING and GREEN 1998; EWING *et al.* 1998). The vector and low-quality sequences (Phred value <16) were trimmed using the Applied Biosystems Sequence Scanner v1.0 Software. Reads less than 50 bp in length were removed. BAC/BIBAC clones were fingerprinted using the SNaPshot method, as previously described (KIM *et al.* 2007; LUO *et al.* 2003; WANG *et al.* 2013), and assembled into contigs with FPC v9.4 at a fixed tolerance value 7 and a cutoff score $1e^{-10}$. Those containing two or more overlapping clones were counted as contigs.

Anchoring BAC/BIBAC clones to the reference sequences through BESs

Repeat Masker version 3.3.0 (http://www.repeatmasker.org) was used to mask end sequences of the maize B73 and sorghum Nengsi-1 BIBAC clones with the maize and sorghum databases, respectively. The end sequences of the *Parviglumis* and *Tripsacum* BAC clones (unmasked) and the maize B73 BIBAC clones (masked and unmasked) were aligned to the maize B73 genome sequence, and the end sequences of the sorghum Nengsi-1 (masked and unmasked) were aligned to the sorghum genome sequence with MegaBLAST using the ncbi-BLAST 2.27+ (CAMACHO *et al.* 2009). The parameters of the program were set to –evalue 1e-100 –outfmt 6 for *Parviglumis, Tripsacum* and sorghum Nengsi-1, and –evalue 1e-150 –outfmt 6 for maize B73. The putative mapped sites of the BAC/BIBAC clones on the reference genome sequences were determined using the Dalrymple's and Zhao's method with the criteria specified (DALRYMPLE *et al.* 2007; ZHAO *et al.* 2004).

Anchoring B73 BIBAC clones to the B73 reference sequence through fingerprints

The putative mapped regions of some representative B73 BIBAC clones on the B73 reference sequence were cut *in silico* with the restriction enzymes *Bam*HI, *Eco*RI, *Xba*I, *Xho*I, *Hae*III that were used in the SNaPshot method for BAC/BIBAC fingerprinting. The real fingerprints of the B73 BIBAC clones and the *in silico* fingerprints of the putative mapped regions on the B73 reference sequence were compared by FPCv9.4 (SODERLUND *et al.* 1997). We calculated Background (GC) and

Significant Quality (LARMANDE *et al.* 2008) to measure their relationships using the algorithms described below.

Background Coincidence Score (GC): GC reflects the coincidence score of a clone's real fingerprint with any random *in silico* fingerprint.

$$GC = 10^{\frac{\sum_{i=1}^{N} \lg e_i}{N}}$$

In the formula, N is the number of clones involved, so we have a cohort of $I = \{i_1, \ldots, i_N\}$ clones. e_i is the minimum coincidence score of the real fingerprints of the clone *i* with the *in silico* fingerprints corresponding to other clones.

Significant Quality (LARMANDE *et al.* 2008): SQ is a quality reflecting the difference between BC_i and Background Coincidence Score.

$$SQ = -\frac{\lg BC_i - \lg GC_i}{\left|\lg GC - \lg SC_i\right|}$$

In this formula, BC_i (Best Coincidence Score) is the minimum coincidence score of the real fingerprints of the clone *i* with the *in silico* fingerprints corresponding to this clone. SC_i (Secondary Coincidence Score) is the secondary minimum coincidence score of the real fingerprints of the clone *i* with the *in silico*

fingerprints corresponding to this clone. If a clone was anchored to a single site on the reference sequence (only *in silico* fingerprints of a single region on the reference is available), then one takes e_i instead of SC_i.

Results

BAC library construction and characterization

We constructed BAC libraries for three elite inbreds of maize (Zheng58, Chang7-2 and Mo17), one maize progenitor, *Zea mays* ssp. *parviglumis* and one species from the closest genus to *Zea* within the tribe maydeae, *Tripsacum dactyloides*. All 5 BAC libraries were constructed with genomic DNA fragments prepared by *Hin*dIII partial digestion and two size selections and in the pIndigoBAC536-S vector prepared from pHZAUBAC1 (SHI *et al.* 2011). BAC clones were arrayed into 384-well microtiter plates and stored in -80 ℃ freezers. The numbers of BAC clones for *Zea mays* ssp. *parviglumis, Tripsacum*, Zheng58, Chang7-2 and Mo17 are 100,608, 128,256, 148,992, 145,920 and 172,416, respectively (Table 1). Each BAC library was constructed from more than one ligation having different insert sizes and transformation efficiencies.

The average insert sizes and empty vector rates were estimated by digesting

randomly selected 238-412 clones from each library with I-*Sce*I and separating the digested clones on 1% agarose CHEF gels. Figure 1 shows the DNA analysis results of 40 randomly picked clones from the *Parviglumis* BAC library. Because the vector was engineered to contain two 18-bp I-*Sce*I recognition sites flanking the cloning site and I-*Sce*I sites are rare in genomic DNA, all BAC clones released a single genomic DNA insert. Insert sizes ranged from 20 to 257kb, with the majority larger than 115kb. The average insert sizes of these BAC libraries ranged between 122kb and 148kb. The rates of empty vector clones ranged from 0.72% to 5.56% (Table 1). The insert size distributions of each BAC library are shown in Figure 2. Some libraries do not show typical Poisson patterns, probably due to the use of different ligations. To determine the organellar DNA content, BAC libraries were screened with three chloroplast probes and four mitochondrial probes. Chloroplast DNA content of the BAC libraries ranged between 0.16% and 0.50% and mitochondrial DNA content ranged from 0 to 0.10%. However, it has been reported that maize contains numerous dynamic nuclear-mitochondrial sequences (NUMTs) and nuclear-plastid sequences (NUPTs) (LOUGH *et al.* 2008; ROARK *et al.* 2010). So screening out the mitochondrial and plastid sequences will also eliminate the NUMTs and NUPTs. The calculated genome coverage of the BAC libraries for *Zea mays* ssp. *parviglumis*, *Tripsacum*, Zheng58, Chang7-2 and Mo17 are 5.5, 4.7, 8.4, 8.0 and 7.9, respectively (Table 1).

Table 1 Summary of the seven ZMAP BAC/BIBAC libraries

Species or varieties	Library type	Genome size (Mb)	Total clone numbers	% Empty vectors	% Chl DNA content	% Mito DNA content	Average insert size (kb)	Fold genome coverage
Parviglumis	BAC	2700[1]	100,608	0.77	0.27	0.04	148	5.5
Tripsacum	BAC	3730[2, 3]	128,256	0.72	0.26	0	139	4.7
Zheng58	BAC	2500[2, 3]	148,992	1.14	0.50	0.05	144	8.4
Chang7-2	BAC	2500[2, 3]	145,920	0.85	0.48	0.10	139	8.0
Mo17	BAC	2500[2, 3]	172,416	5.56	0.16	0.01	122	7.9
B73	BIBAC	2300[2- 4]	192,000	0.35	0.15	0.09	92	7.6
Sorghum Nengsi-1	BIBAC	750[2,3,5]	51,840	2.29	2.20	0.10	111	7.3

1. Dermastia, M., et al., *A cellular study of teosinte Zea mays subsp. parviglumis (Poaceae) caryopsis development showing several processes conserved in maize.* Am J Bot, 2009. **96**(10): p. 1798-807.
2. Bennett, M.D. and I.J. Leitch, *Nuclear DNA amounts in angiosperms: targets, trends and tomorrow.* Ann Bot, 2011. **107**(3): p. 467-590.
3. Arumuganathan, K. and E.D. Earle, *Nuclear DNA content of some important plant species.* Plant Mol Biol Rep, 1991. **9**(3): p. 208-218.
4. Schnable, P.S., et al., *The B73 maize genome: complexity, diversity, and dynamics.* Science, 2009. **326**(5956): p. 1112-5.
5. Paterson, A.H., et al., *The Sorghum bicolor genome and the diversification of grasses.* Nature, 2009. **457**(7229): p. 551-6.

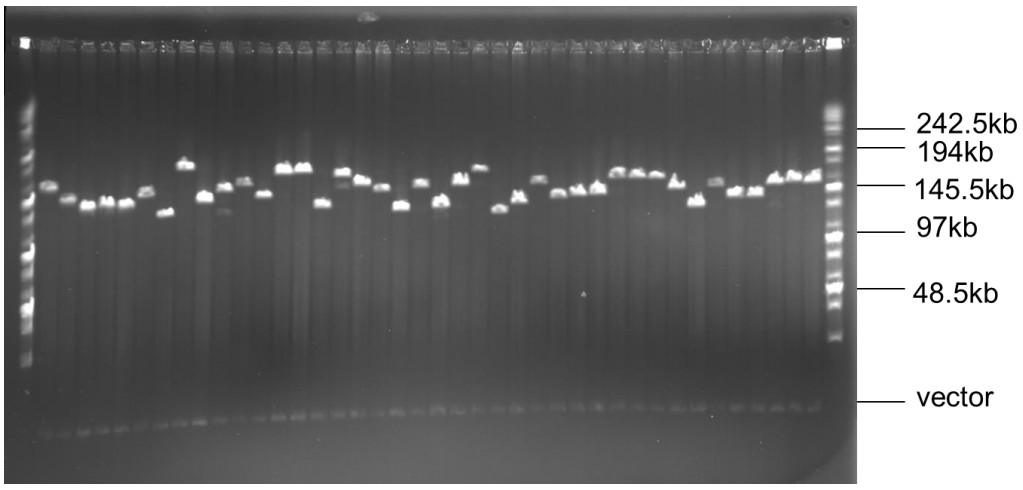

Fig. 1 DNA analysis of random BAC clones from the *Parviglumis* BAC library by pulse-field gel

electrophoresis. DNA samples were digested with I-*Sce*I and separated on a 1% agarose CHEF gel. The 6.38-kb common band is the vector pIndigoBac536-S I-*Sce*I fragment. Markers used are MidRange PFG marker I (New England Biolabs).

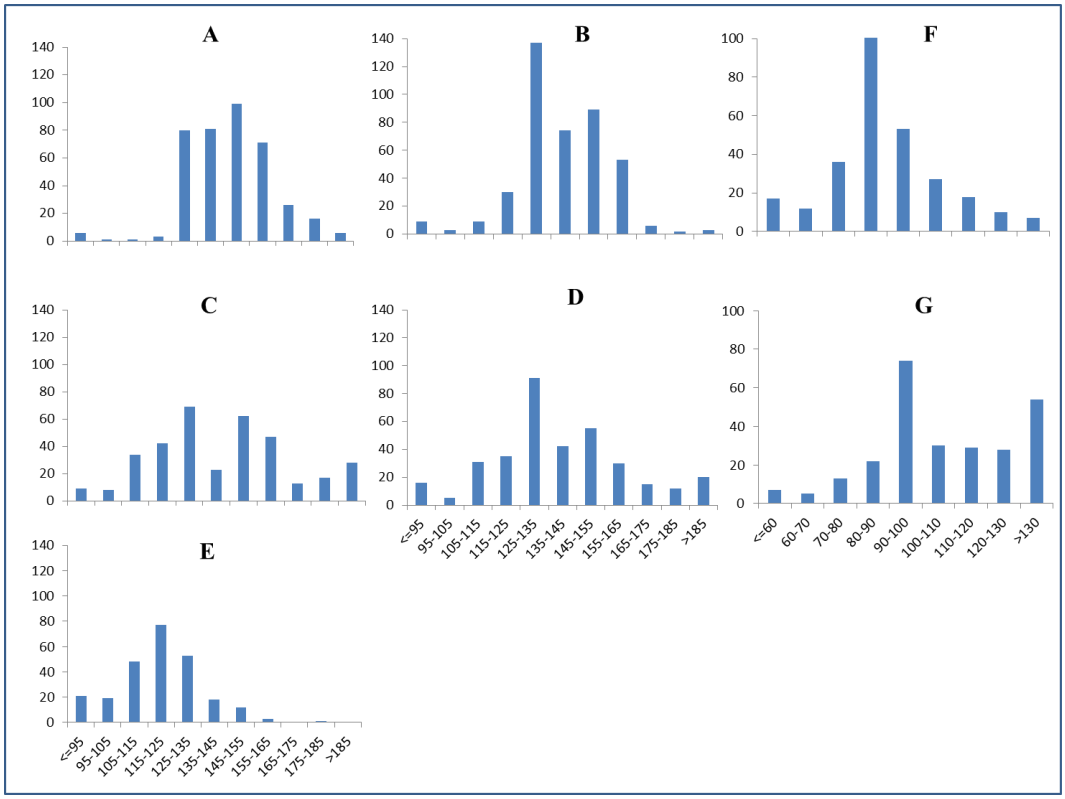

Fig. 2 Insert size distributions of the seven ZMAP BAC and BIBAC libraries. In each histogram, the x-axis represents insert size range (kb) and the y-axis represents the number of clones within a particular insert size range. (A) *Parviglumis* BAC library, the average insert size is 148kb based on 387 clones. (B) *Tripsacum* BAC library, the average insert size is 139kb based on 412 clones. (C) Zheng58 BAC library, the average insert size is144kb based on 348 clones. (D) Chang7-2 BAC library, the average insert size is 139kb based on 349 clones. (E) Mo17 BAC library, the average insert size is 122kb based on 238 clones. (F) B73 BIBAC library, the average insert size is 92kb based on 282 clones. (G) Sorghum BIBAC library, the average insert size is 111kb based on 256 clones.

BIBAC library construction and characterization

We also constructed two BIBAC libraries, one for maize B73 and one for sorghum Nengsi-1. Both BIBAC libraries were constructed with genomic DNA fragments prepared by *Bam*HI partial digestion and two size selections into the BIBAC-S vector prepared from pHZAUBIBAC1 (SHI *et al.* 2011). Multiple ligations were also used for the BIBAC libraries. As for BAC clones, BIBAC clones were arrayed into 384-well microtiter plates and stored in -80℃ freezers. Clone pools per each 384-well microtiter plate of the maize B73 BIBAC library were made for PCR screening. The numbers of BIBAC clones for maize B73 and sorghum Nengsi-1 are 192,000 and 51,840, respectively (Table 1). DNA samples of 282 and 256 clones, randomly selected from the maize B73 and sorghum Nengsi-1 BIBAC libraries, respectively were analyzed with I-*Sce*I. All BIBAC clones released an intact genomic DNA insert band (Figure 3). Insert sizes ranged from 30 to 190kb. The average insert size, empty vector rate, chloroplast and mitochondrial DNA content were 92kb, 0.35%, 0.15% and 0.09%, respectively, for the maize B73 BIBAC library, and 111kb, 2.29%, 2.20% and 0.10%, respectively, for the sorghum Nengsi-1 BIBAC library (Table 1). The insert size distribution of each BIBAC library is shown in Figure 2. We did not use very large genomic DNA fragment fractions for BIBAC library construction on purpose because previously it was reported that large BIBACs were not stable in *Agrobacterium* (SONG *et al.* 2003). The calculated genome coverage of

the maize B73 and sorghum Nengsi-1 BIBAC libraries are 7.6 and 7.3, respectively (Table 1).

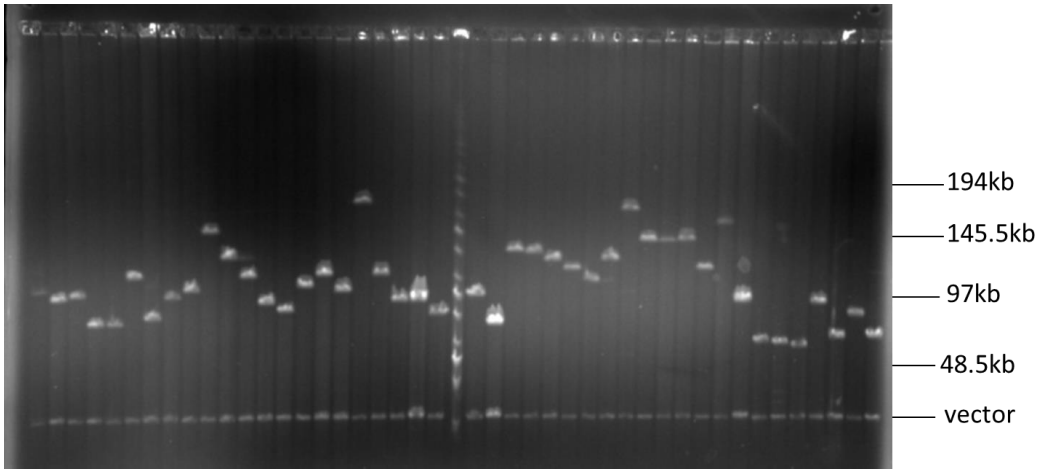

Fig. 3 DNA analysis of random BIBAC clones from the sorghum Nengsi-1 BIBAC library by pulse-field gel electrophoresis. DNA samples were digested with I-*Sce*I and separated on a 1% agarose CHEF gel. The 23.2-kb common band is the vector BIBAC-S I-*Sce*I fragment. Markers used are MidRange PFG marker I (New England Biolabs).

Utility demonstration for the *Parviglumis* and *Tripsacum* BAC libraries.

To demonstrate the potential usefulness of the *Parviglumis* and *Tripsacum* BAC libraries, we chose 12 single copy genes as probes from the 10 maize chromosomes according to the B73 reference sequence and screened the *Parviglumis* and *Tripsacum* BAC libraries. First, we screened the two libraries with P^{32}- labeled probes and then re-hybridized the clones collected from the first round of screening with digoxin-labeled probes. A total of 259 clones hybridized in the two libraries with the 12

P^{32}–labeled maize probes and the identity of 205 (79%) clones were confirmed with the 12 digoxin-labeled probes (Table 2). All 12 P^{32}–labeled maize gene probes hybridized to clones in both *Parviglumis* and *Tripsacum* BAC libraries, and 96 out of the 111 *Parviglumis* clones (86%) and 109 out of the 148 *Tripsacum* clones (74%) were confirmed with digoxin-labeled probes. The average confirmed positive clones per probe were 8.0 and 9.1 for *Parviglumis* and *Tripsacum* BAC libraries, respectively, even higher than predicted by their respective 5.5 and 4.7 genome equivalencies. The 54 clones that hybridized with P^{32}–labeled probes but not with digoxin-labeled probes could be false positive clones or, more likely, missed positive clones due to the weak hybridization signals of the digoxin-labeled heterologous probes and inefficient digoxin labeling of the probes *ba1* and *ps1* in *Tripsacum* colony hybridization (see below).

Table 2 Numbers of positive clones of the *Parviglumis* and *Tripsacum* BAC libraries for the 12 maize gene probes.

Maize gene probe	Maize Chr	Parviglumis clones*	Tripsacum clones*
vp1	Chr3	2 (2)	4 (2)
ba1	Chr3	7 (7)	6 (0)
ps1	Chr5	19 (10)	21 (1)
why1	Chr6	11 (11)	11 (9)
o2	Chr7	9 (7)	6 (5)
gst1	Chr8	7 (7)	10 (10)
waxy1	Chr9	5 (5)	12 (12)
du1	Chr10	16 (16)	6 (6)
tb1	Chr1	8 (7)	39 (34)
hrg1	Chr2	12 (11)	12 (11)
adh1	Chr1	9 (7)	13 (13)

adh2	Chr4	6 (6)	8 (6)
total		**111 (96)**	**148 (109)**
Avg clones per probe		**9.2 (8.0)**	**12.3 (9.1)**

* Number of clones collected in the first round screening (Number of clones confirmed with the second hybridization).

We SNaPshot fingerprinted and sequenced both ends of all 259 clones from the first round screening. A total of 238 (92%) of the BAC clones resulted in qualified fingerprints and 149 BAC clones were assembled into contigs. Of the 149 BAC clones, 130 were from the 205 confirmed positive BAC clones and 19 were from the 54 BAC clones that were not confirmed in the second hybridization. These 19 BAC clones were considered as missed positive clones by the second hybridization probably due to the reasons mentioned above. Together, 64 *Parviglumis* BAC clones and 85 *Tripsacum* BAC clones were assembled into 20 *Parviglumis* BAC contigs and 27 *Tripsacum* BAC contigs, respectively (Table 3). The average clones per contig were 3.1 and 3.2 for *Parviglumis* and *Tripsacum* BAC libraries, respectively. Positive clones for some maize probes remained singletons (e.g. those of both BAC libraries for *vp1* and those of the *Parviglumis* BAC library for *adh2*) or formed more than one contig (e.g. those of both BAC libraries for *ps1*, *why1*, *gst1*and *hrg1*). This is probably due to the technical limitation of contig assembly, heterozygosity, and/or, most likely according to the much higher positive rates than predicted, duplication of these genes in these genomes relative to maize (see discussion below).

We used the fingerprints of all the 238 *Parviglumis* and *Tripsacum* BAC clones together for contig assembly and found that the clones that were assembled into each

contig were from the same library except that one *gst1* clone from *Parviglumis* BAC library was misassembled into the *Tripsacum adh1* contig, which contains 5 *Tripsacum* clones. This result indicated that a good level of diversity exists between *Zea mays* ssp. *parviglumis* and *Tripsacum* at all of the homologous regions to the 12 maize genes.

Table 3 Contigs assembled with the positive *Parviglumis* and *Tripsacum* BAC clones for the 12 maize gene probes.

Maize gene probes	*Parviglumis*		*Tripsacum*	
	contig numbers*	clones in contigs**	contig numbers*	clones in contigs**
vp1	0	0	0	0
ba1	1	2	1 (1)	2 (2)
ps1	4 (1)	11 (3)	2 (2)	4 (4)
why1	2	7	3	8
o2	1 (1)	2 (2)	1	2
gst1	2	5	3	8
waxy1	1	4	3	9
du1	4	16	1	5
tb1	1	4 (1)	6	22 (5)
hrg1	3	9	3	11 (1)
adh1	1	4 (1)	3	11
adh2	0	0	1	3
Total	20 (2)	64 (7)	27 (3)	85 (12)
Avg clones Per contig	**3.1**		**3.2**	

* Number of contigs assembled with the clones collected in the first round screening (Number of contigs assembled with the clones that were not confirmed with the second hybridization);
** Number of clones collected in the first round screening (Number of clones that were not confirmed with the second hybridization.

We anchored the *Parviglumis* and *Tripsacum* positive clones and contigs to the maize B73 high-quality draft reference sequence through BESs (Table 4 and Table S3). From the 224 positive BAC clones (205+19), we obtained a total of 431 qualified

BESs (98% success). The 103 *Parviglumis* positive BAC clones generated 202 BESs with an average read length of 654bp and the 121 *Tripsacum* positive BAC clones generated 229 BESs with an average read length of 599bp. For *Parviglumis* BAC clones, 102 out of the 103 positive BAC clones possess BESs with hits to the B73 reference sequence, 94 with paired BES hits and 8 with single-end BES hits. Of the 202 BESs, 196 hit to the B73 reference sequence: 36 hit single sites and 160 hit multiple sites. Using criteria that the aligned paired BESs should locate vis-à-vis to the same chromosome with a distance between 35kb and 500kb, we could anchor 70 clones to the B73 reference sequence, 60 to multiple sites and 10 to single sites. Thirty-four clones were anchored to the expected sites for the maize gene probes. Of the 34 clones, 18 clones were from 9 contigs and 16 clones were singletons (Table 4, and Table S3). For *Tripsacum* BAC clones, 104 out of 121 positive BAC clones possessed BESs with hits to the B73 reference sequence: 46 with paired BES hits and 58 with single-end BES hits. Of the 229 BESs, 150 hit the B73 reference sequence, 54 hit single sites and 96 multiple sites. Twenty-two clones could be anchored to the B73 reference sequence, 15 to multiple and 7 to single sites. Seven clones, 5 from 3 contigs and 2 singletons, were anchored to the expected sites for the maize gene probes (Table 4, and Table S3). The placement of *Parviglumis* and *Tripsacum* BAC clones and contigs to the expected maize B73 sites indicated that large regions surrounding the corresponding sites were generally co-linear between *Zea mays* ssp. *parviglumis* and maize and between *Tripsacum* and maize. Fewer BESs, clones, and contigs of *Tripsacum* were aligned to the maize B73 reference sequence than those of

Zea mays ssp. *parviglumis*, consistent with the phylogenetic distance of *Tripsacum* to maize.

Table 4 Anchoring of the positive *Parviglumis* and *Tripsacum* BAC clones and contigs for the 12 maize gene probes to the maize B73 reference sequences through BESs

	*Parviglumis**	*Tripsacum**
Total clones	103(7)	121(12)
Clones with BESs	102(7)	120(12)
without BES hits on B73	0	16(4)
with BES hits on B73	102(7)	104(8)
with paired BES hits	94(6)	46(4)
with single-end BES hits	8(1)	58(4)
Total BESs	202(14)	229(23)
BES without hits on B73	6(1)	79(11)
BES with hits on B73	196(13)	150(12)
with single hit	36(4)	54(4)
with multi-hits	160(9)	96(8)
Clones anchored to B73	70(3)	22(1)
Clones anchored to multi-sites	60(3)	15(1)
Clones anchored to single-sites	10	7
Clones anchored to the expected sites	34(1)	7
Clones from contigs	18(1)	5
Clones not from contigs	16	2
Contigs anchored to B73	9	3

* The data from all the positive clones (the data from only the 19 clones that were not confirmed with the second hybridization but were assembled into contigs)

Stability of sorghum Nengsi-1 BIBAC clones in *E. coli* and *Agrobacterium*

We previously demonstrated that maize B73 BIBAC clones of different sizes from small to 160kb were all stable in both *E. coli* and *Agrobacterium* (SHI *et al.* 2011). In this study, we tested the stability of sorghum Nengsi-1 BIBAC clones and found that the BIBAC clones of different sizes were all stable in *E. coli* (data not shown). When 10 sorghum BIBAC clones with different insert sizes from 58kb to 160kb were tested for their stability in *Agrobacterium* EHA105 via an indirect method

(SHI *et al.* 2011), nine were found to be stable after 48h and 96h of culture (sub-cultured after 48h). Figure 4 shows the results for 8 clones. The 48h and 96h culture samples of the *Agrobacterium* EHA105 colonies showed the same I-*Sce*I digestion patterns as the original *E. coli* clones. Only one clone with an insert size of 131kb was not stable. Three types of smaller insert fragments were found from both the 48h and 96h culture samples of this clone (data not shown).

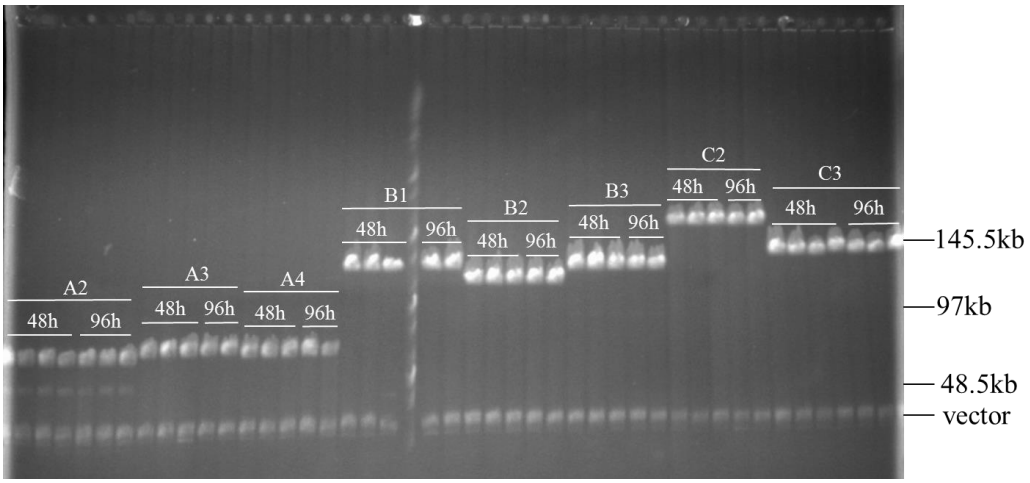

Fig.4 Stability of sorghum Nengsi-1 BIBAC clones in *Agrobacterium*. BIBAC clones with different insert sizes from 58kb to 160kb were selected randomly from the sorghum Nengsi-1 BIBAC library. Plasmid DNA of each BIBAC clone was transferred into *Agrobacterium* EHA105 cells. Single *Agrobacterium* EHA105 transformed colonies were cultured at 28℃ for 48h and 96h (sub-cultured after 48h). BIBAC DNA of each culture was prepared via an indirect method (SHI *et al.* 2011) and analyzed with I-*Sce*I. The molecular weight marker was the MidRange PFG marker I (NEB).

Mapping maize B73 and sorghum Nengsi-1 BIBAC clones to reference sequences

High-quality draft reference sequences for both maize and sorghum are available. Especially, our maize B73 BIBAC library was constructed using the same source of seeds as that used for B73 sequencing. To assign sequence information to our BIBAC libraries, we tried to map random maize B73 and sorghum Nengsi-1 BIBACs to the maize B73 and sorghum reference sequences, through BESs (Table 5), respectively.

We sequenced 1,152 random BIBACs at both ends for the maize B73 BIBAC library and obtained a total of 2,235 qualified sequences (97.0% success) with an average read length of 679bp. A total of 1,087 clones generated paired BESs and 61 clones generated single-end BESs. Of the 2,235 BESs, 2,064 BESs (92.3%) were aligned to the maize B73 reference sequence, 371 (18%) hit single and 1,693 (82%) multiple sites. After repeat sequences were masked, 491 BESs were aligned to the maize B73 reference sequence, 293 hit single and 198 multiple sites. Using criteria that the aligned paired BESs should locate vis-à-vis to the same chromosome with a distance between 35kb and 180kb, we could anchor 852 clones (74.2% of the clones with successful BESs) to the B73 reference sequence, 625 (73.3%) to multiple and 227 (26.6%) to single sites.

We also sequenced 144 random BIBACs at both ends for the sorghum Nengsi-1 BIBAC library and obtained a total of 251 qualified sequences (87.2% success) with an average read length of 562bp, 118 clones generated paired BESs and 15 clones single-end BESs. Of the 251 BESs, 195 BESs (77.4%) were aligned to the sorghum reference sequence, 62 (31.8%) hit single and 133 (68.2%) multiple sites. After repeat

sequences were masked, 167 BESs were aligned to the sorghum reference sequence, 52 hit single and 115 multiple sites. Using criteria that the aligned paired BESs should locate vis-à-vis to the same chromosome with a distance between 35kb and 500kb, the same as those used for anchoring of *Parviglumis* and *Tripsacum* BACs, we anchored 66 sorghum Nengsi-1 BIBACs (49.2% of the clones with successful BESs) to the sorghum reference sequence, 44 (66.7%) to multiple and 22 (33.3%) to single sites. The percentage of sorghum Nengsi-1 BIBAC clones anchored to the sorghum reference sequence (49.2%) was significantly lower than that of maize B73 BIBAC clones to the B73 reference sequence (74.2%) even though lower stringency was used for anchoring of the sorghum BIBACs. Because the sorghum Btx623 has been generated with a higher genome coverage than the maize B73 genome, paired reads of different insert libraries, and an average high quality long read of 735 bp, it is unlikely that the difference is due to gaps in the genome assembly. Therefore, it appears that the sequence of sorghum Nengsi-1 has more diverged from sorghum Btx623 reference sequence than expected.

Table 5 Anchoring of maize B73 and sorghum Nengsi-1 BIBAC clones to their respective reference sequences through BESs

	B73	Nengsi-1
BAC end sequences		
Clones for BAC end sequencing	1152	144
Clones with successful BESs	1148	133
with paired BESs	1087	118
with single-end BESs	61	15
Total successful BESs	2235 (97.0%)	251(87.2%)
Alignment		
Aligned BESs with repeats unmasked	2064 (92.3%)	195 (77.4%)
Single hit BESs	371 (18%)	62 (31.8%)
Multi-hit BESs	1693 (82%)	133 (68.2%)

Aligned BESs with repeats masked	491	167
Single hit BESs	293	52
Multi-hit BESs	198	115
Anchoring to reference sequences		
Clones anchored	852 (74.2%)	66 (49.2%)
Clones anchored to multi-sites	625 (73.3%)	44 (66.7%)
Clones anchored to single sites	227 (26.6%)	22 (33.3%)

To test if the multi-sites anchored clones could be further refined and to evaluate single-site anchored clones, we fingerprinted 48 B73 BIBAC clones and obtained qualified fingerprints for 36 clones, 9 single-site anchored clones, 27 multi-sites anchored clones. Because the maize B73 physical map was constructed with BAC fingerprints using earlier methods and different restriction enzymes (WEI *et al.* 2007; WEI *et al.* 2009) than the SNaPshot method (LUO *et al.* 2003), we could not merged our fingerprints into a common maize B73 BAC physical map. To circumvent this problem, we digested the target regions of the B73 reference sequence with the same 5 enzymes *in silico* that were used in SNaPshot method and compared our B73 BIBAC fingerprints with the *in silico* fingerprints using the software FPCv9.4 at the cutoff value of 1e-12 and tolerance value of 3bp. Significant Quality (LARMANDE *et al.* 2008) is defined as a value which reflects the confidence level of the match between the BIBAC fingerprints and *in silico* fingerprints under the setting cutoff value of 1e-12; the algorithm of SQ is described in the Methods section. If SQ is greater than 2, it indicates that the result is significant. If SQ is lower than 2, the result would not reach a significant threshold. When SQ is a negative number, the results are not usable.

Indeed, 8 out of the 9 fingerprints from single-site anchored clones did match *in silico* fingerprints of the target regions of the B73 reference sequence under the FPC cutoff value. The average SQ of these 8 clones is 5.084, and the FPC coincidence score ranges from 4.00e-22 to 6.47e-17. This result indicates that the single-site anchored clones determined by paired BESs have a high confidence level.

Of the 27 multi-sites anchored clones, fingerprints of each of ten clones (37%) matched to *in silico* fingerprints of one site with SQ value greater than 2. Hence, the 10 clones that were anchored to multiple sites of reference sequence with BESs could be further resolved into single sites of reference sequence with fingerprints at high confidence. For the 17 remaining multi-sites anchored clones, fingerprints from 10 clones (37% of the multi-sites anchored clones analyzed) matched to *in silico* fingerprints with SQ values between 0 and 2, whereas 2 clones (7%) had negative SQs and 5 clones (19%) did not match under the FPC cutoff value at all.

Discussion

Construction of the ZMAP BAC/BIBAC libraries

Large genome comparative projects, such as for *Drosophila* (SONG *et al.* 2011) and for *Oryza* (AMMIRAJU *et al.* 2006; WING *et al.* 2005)(www.OMAP.org), have served as foundations for evolutionary and functional genomics studies. For a similar purpose, we constructed BAC libraries for the three inbred lines Zheng58, Chang7-2,

and Mo17 and the wild species *Zea mays* ssp. *parviglumis* and *Tripsacum dactyloides*, in addition to BIBAC libraries for maize inbred B73 and sorghum land race Nengsi-1. The seven BAC/BIBAC libraries were called ZMAP BAC/BIBAC libraries following the OMAP nomenclature. The BAC libraries for *Zea mays* ssp. *parviglumis*, *Tripsacum*, Zheng58, Chang7-2, and Mo17, and the BIBAC libraries for B73 and Nengsi-1 have average insert sizes of 148kb, 139kb, 144kb, 139kb, 122kb, 92kb and 111kb, average organellar DNA of 0.31%, 0.26%, 0.55%, 0.58%, 0.17%, 0.24% and 2.3%, empty vector rates of 0.77%, 0.72%, 1.14%, 0.85%, 5.56%, 0.35% and 2.29%, and genome equivalents of 5.5, 4.7, 8.4, 8.0, 7.9, 7.6 and 7.3, respectively (Table 1). The BAC libraries for Zheng58, Chang7-2, and Mo17 are the first public BAC resources for these three elite maize inbreds, the *Parviglumis* and *Tripsacum* BAC libraries are the first BAC libraries for close wild relatives of maize, and the two BIBAC libraries are the first BIBAC resources for maize and sorghum. These new BAC/BIBAC libraries, together with the already available BAC libraries and genome map resources of other maize inbred lines, sorghum, pearl millet, *Setaria* and rice, will provide a wide-range of trait diversity and evolutionary distances. These powerful new resources will complement use of the maize reference sequence.

Coverage estimations

Genome coverage for a BAC/BIBAC library is usually estimated mathematically from the average insert size, total number of clones, and size of the corresponding

genome (AMMIRAJU *et al.* 2006; LEE *et al.* 2013; LUO *et al.* 2001; SONG *et al.* 2011; WU *et al.* 2004). Hybridization and contig analysis were also used as complementary approaches to estimate genome coverage (AMMIRAJU *et al.* 2006; SONG *et al.* 2011). When the probe sequences were from the same genome for which the BAC/BIBAC library was constructed or from a close genome, the average hybridization results were usually close to the mathematic results (LEE *et al.* 2013; LUO *et al.* 2001). However, compared to the mathematic estimation, hybridization and contig-based estimation can be affected by many uncounted factors. In hybridization-based estimation, technically, inefficient probe labeling, improper hybridization conditions, filter qualities, etc. could all result in underestimation (AMMIRAJU *et al.* 2006), and at the sequence level, sequence diversity of the heterologous probes could also result in underestimation. In contig-based estimation, in addition to factors affecting the hybridization-based estimation, an unknown multi-copy number and/or heterozygosity of a locus could also affect the outcomes. Furthermore, low quality fingerprints, too many or too few fingerprint bands due to nested retrotransposons and uneven distribution of restriction sites of the enzymes for fingerprinting, or not enough overlap in BAC clones also affect contig assemblies. We screened the *Parviglumis* and *Tripsacum* BAC libraries with 12 maize "single copy" gene probes and obtained an average of 8.0 and 9.1 confirmed positive clones per probe, respectively. Still, some clones may have remained undetected as reported in our previous work (AMMIRAJU *et al.* 2006). The average hybridization-based coverage of 12 maize probes of the BAC libraries for *Zea mays* ssp. *parviglumis* and *Tripsacum*

(8.0 and 9.1) is higher than the mathematically estimated coverage (5.5 and 4.7). Copy number variation in gene families corresponding to some maize probes could exist in *Zea mays* ssp. *parviglumis* and *Tripsacum*. For instance, *ps1*- and *du1*-homologous clones were over-represented 2-3 fold in the *Parviglumis* library, whereas *tb1* was 5-fold over-represented in *Tripsacum* BACs. This suggested that these genes contain lineage or species-specific amplifications, a fact that was confirmed by the observation of 6 homologous contigs possibly containing 5-6 different *tb1* alleles in *Tripsacum*.

We fingerprinted and end-sequenced the 259 BAC clones collected from the first round of screening and could only assemble fingerprints of 149 clones into a total of 47 contigs: 64 *Parviglumis* BAC clones and 85 *Tripsacum* BAC clones into 20 and 27 contigs, with average clones per contig 3.1 and 3.2, respectively. The average clones per contig, if used, could underestimate the coverage due to the missed clones in hybridization screening and failed assembly of some homologous contigs or, especially, singletons. For example, of the 34 *Parviglumis* BAC clones that were anchored to the expected sites of the B73 reference sequence through BESs, 16 were singletons; of the 7 *Tripsacum* BAC clones that were anchored to the expected sites of the B73 reference sequence through BESs, 2 were singletons. Some homologous contigs and singletons were anchored to the same sites of B73 reference sequence through BESs (Table 4). These contigs and singletons potentially could be manually merged, but more sequence information is required.

Utilization of the ZMAP BAC/BIBAC resources

The ZMAP BAC/BIBAC libraries provide a platform for maize research in several areas. They are particularly useful for positional or homologous cloning of genes or alleles desired for maize or even cereal crop improvement, not only because they were derived from the elite lines and untapped maize relatives that stacked or maintained many beneficial genes or alleles, but also because they were all cloned in our new BAC/BIBAC vectors (SHI *et al.* 2011). Our BIBAC vector is stable in *Agrobacterium*. Interesting BIBAC clones can be directly used for complementary experiments. For interesting BAC clones, the intact large inserts can be easily exchanged to the BIBAC vector. Our BAC and BIBAC vectors were both engineered to have two I-*Sce*I sites flanking the cloning site in an inverted orientation and therefore add two important features. First, the I-*Sce*I site is an 18bp sequence, and its presence in any organism will be rare (e.g., only two sites were found in the maize B73 reference sequence). Therefore, intact large inserts can be released from the BAC and BIBAC clones. Second, the I-*Sce*I site is non-palindromic. The same non-complementary 3' protruding ATAA ends of the BAC and BIBAC vectors produced by I-*Sce*I digestion suppress the self- and inter-ligations. Because the BAC and BIBAC vectors contain chloramphenicol and kanamycin resistance genes for selection, respectively, intact large BAC inserts can be directly re-cloned into the BIBAC vector by digesting BAC clones with I-*Sce*I, adding the BIBAC vector

prepared with I-*Sce*I to the I-*Sce*I-digested BAC products without the purification of large insert DNA, ligating, transforming and selecting the transformants on kanamycin-containing medium (SHI *et al.* 2011).

Comparative genome analysis among closely related species is a powerful approach to explore genome structure and function (LIN *et al.* 2012). However, to do this, it is necessary to have high-quality whole genome or target region sequences of the investigated species.

BAC sequencing has been extensively used in comparative studies (AMMIRAJU *et al.* 2008; BENNETZEN and MA 2003; CHEN *et al.* 1997; CLARK *et al.* 2004; ILIC *et al.* 2003; LAI *et al.* 2004; LU *et al.* 2009; MESSING and LLACA 1998; SWIGONOVA *et al.* 2004; TIKHONOV *et al.* 1999; WANG and DOONER 2012). The ZMAP BAC/BIBAC libraries provide an important new set of sequencing templates for such studies in the *Zea* lineage. To see if the orthologous loci of maize can be identified in *Zea mays* ssp. *parviglumis* and *Tripsacum*, we screened the *Parviglumis* and *Tripsacum* BAC libraries with 12 maize B73 gene probes. Positive clones were obtained for every probe. Many clones can be anchored to the expected sites for the probes on the B73 reference sequence by BESs. These results demonstrated the high utility of the two maize relative libraries.

The ZMAP BAC/BIBAC libraries provide essential tools for construction of physical maps. BAC physical map-based genome sequencing and whole genome shotgun (WGS) sequencing are the two main approaches to obtain genome sequences.

Lewin et al (LEWIN *et al.* 2009) reported that a good physical map is needed to obtain a high-quality genome sequence. We evaluated the two rice reference sequences, the Nipponbare reference sequence obtained with the physical map-based approach (IRGSP, 2005) and the 93-11 reference sequence obtained with the WGS approach (YU *et al.* 2005), and found that the Nipponbare reference sequence has a high quality whereas the 93-11 reference sequence, though also used long Sanger reads as the Nipponbare reference sequence, has a much lower quality (Pan; Deng et al., manuscript in submission). Many large genomes were sequenced with the short next-generation sequencing (NGS) reads. However, without physical maps as frames for assembly, these genome sequences will be kept fragmented. We recently developed a new method for simultaneously physical mapping and sequencing of complex genomes using the same set of NGS reads of BAC/BIBAC pools that can largely reduce the cost (Pan, Wang et al., manuscript in submission), making the physical map-based genome sequencing approach and BAC/BIBAC resources more attractive.

Genetic transformation of plants with large insert BIBAC/TAC clones is a desirable approach in gene functional analysis and molecular breeding because large insert BIBAC/TAC clones can contain complete large genes with regulatory sequences, gene clusters or multiple genes (LEE *et al.* 2013; SHI *et al.* 2011; ZHAI *et al.* 2013). Although the biolistic bombardment method has been successfully used to deliver 45 kb of a sorghum BAC comprising a cluster of kafirin gene into the maize genome that were tissue specifically expressed (SONG *et al.* 2004) and 83 kb of an

Arabidopsis BIBAC into tobacco (CHANG *et al.* 2011), the *Agrobacterium*-mediated transformation method is still considered the best method to deliver BIBACs/TACs to different plants (HAMILTON *et al.* 1996; LEE *et al.* 2013; LIU *et al.* 2002; LIU *et al.* 1999; QU *et al.* 2003; SHI *et al.* 2011), because of fewer instances of multiple (e.g., concatamerized) vector uptake and fewer cases of vector/insert rearrangement. A large DNA fragment including a herbicide resistance marker gene, a 30 kb yeast genomic fragment and a 35S-lox-cre recombination cassette was successfully transferred from a BIBAC clone into the maize Hi II line through *Agrobacterium* (VEGA *et al.* 2008). However, for *Agrobacterium*-mediated transformation, the stable maintenance of BIBAC clones in *Agrobacterium* is critical. We showed that the maize B73 BIBAC clones were stable in *Agrobacterium* (SHI *et al.* 2011). In fact, we have successfully transferred many large maize DNA fragments from maize B73 BIBAC clones into rice through *Agrobacterium*-mediated transformation (our unpublished data). In this study, we also demonstrated that the sorghum Nengsi-1 BIBAC clones were essentially stable in *Agrobacterium*. For the ten sorghum Nengsi-1 BIBAC clones tested, 9 were stable in *Agrobacterium*.

The maize B73 seeds that we used in the BIBAC library construction were from the same source as those used in B73 genome sequencing. The current version of the B73 reference sequence still contains many physical and sequence gaps and assembly errors (WEI *et al.* 2009). The B73 BIBAC library could potentially be used to improve the B73 reference sequence.

Accession Numbers

The BES sequences reported in this paper have been deposited in the GenBank database [Accession Nos. KG619880 - KG622865].

Acknowledgments

We thank the USDA germplasm bank at Ames, Iowa for the seeds of *Zea mays* ssp. *parviglumis* and *Tripsacum dactyloides*; Dr. Ruiheng Du of Hebei Academy of Agriculture and Forestry Sciences, China and Dr. Xianmin Diao of the Institute of Crop Science, Chinese Academy of Agricultural Sciences for the sorghum seeds and leaves; the Maize Genetics Cooperation Stock Center (Urbana, IL) for the B73 seeds. We also thank Fusheng Wei, Xiang Song, Jose ´L. Goicoechea, Yeisoo Yu, Jianwei Zhang, and other members of the Arizona Genomics Institute, and Yonglong Pan and Xiaoming Wang of Huazhong Agricultural University for technical assistance. This work was supported by National Basic Research Program (the “973” project, 2009CB118404), National Programs for High Technology Research and Development (863 Project: 2012AA10A305) and the Chinese 111 Project B07041.

Literature Cited

ALLOUIS, S., X. QI, S. LINDUP, M. D. GALE and K. M. DEVOS, 2001 Construction of a BAC library of pearl millet, Pennisetum glaucum. Theoretical and Applied Genetics 102: 1200-1205.

AMMIRAJU, J. S., F. LU, A. SANYAL, Y. YU, X. SONG *et al.*, 2008 Dynamic evolution of oryza genomes is revealed by comparative genomic analysis of a genus-wide vertical data set. Plant Cell 20: 3191-3209.

AMMIRAJU, J. S., M. LUO, J. L. GOICOECHEA, W. WANG, D. KUDRNA *et al.*, 2006 The Oryza bacterial artificial chromosome library resource: construction and analysis of 12 deep-coverage large-insert BAC libraries that represent the 10 genome types of the genus Oryza. Genome Res 16: 140-147.

ARUMUGANATHAN, K., and E. D. EARLE, 1991 Nuclear DNA content of some important plant species. Plant Mol Biol Rep 9: 208-218.

BANTIN, J., F. MATZK and T. DRESSELHAUS, 2001 Tripsacum dactyloides (Poaceae): a natural model system to study parthenogenesis. Sexual Plant Reproduction 14: 219-226.

BARBAZUK, W. B., and P. S. SCHNABLE, 2011 SNP discovery by transcriptome pyrosequencing. Methods Mol Biol 729: 225-246.

BENNETT, M. D., and I. J. LEITCH, 2011 Nuclear DNA amounts in angiosperms: targets, trends and tomorrow. Ann Bot 107: 467-590.

BENNETZEN, J. L., and M. FREELING, 1993 Grasses as a single genetic system: genome composition, collinearity and compatibility. Trends Genet 9: 259-261.

BENNETZEN, J. L., and S. HAKE, 2009 *Handbook of Maize Genetics and Genomics*. Springer.

BENNETZEN, J. L., and J. MA, 2003 The genetic colinearity of rice and other cereals on the basis of genomic sequence analysis. Curr Opin Plant Biol 6: 128-133.

BENNETZEN, J. L., J. SCHMUTZ, H. WANG, R. PERCIFIELD, J. HAWKINS *et al.*, 2012 Reference genome sequence of the model plant Setaria. Nat Biotechnol 30: 555-561.

BRUGGMANN, R., A. K. BHARTI, H. GUNDLACH, J. LAI, S. YOUNG *et al.*, 2006 Uneven chromosome contraction and expansion in the maize genome. Genome Res 16: 1241-1251.

BRUNNER, S., K. FENGLER, M. MORGANTE, S. TINGEY and A. RAFALSKI, 2005 Evolution of DNA sequence nonhomologies among maize inbreds. Plant Cell 17: 343-360.

BYRT, C. S., C. P. GROF and R. T. FURBANK, 2011 C4 plants as biofuel feedstocks: optimising biomass production and feedstock quality from a lignocellulosic perspective. J Integr Plant Biol 53: 120-135.

CALVINO, M., and J. MESSING, 2012 Sweet sorghum as a model system for bioenergy crops. Curr Opin Biotechnol.

CAMACHO, C., G. COULOURIS, V. AVAGYAN, N. MA, J. PAPADOPOULOS *et al.*, 2009 BLAST+: architecture and applications. Bmc Bioinformatics 10: 421.

CENCI, A., N. CHANTRET, X. KONG, Y. GU, O. D. ANDERSON *et al.*, 2003 Construction and characterization of a half million clone BAC library of durum wheat (Triticum turgidum ssp. durum). Theor Appl Genet 107: 931-939.

CHANG, Y. L., H. W. CHUANG, K. MEKSEM, F. C. WU, C. Y. CHANG *et al.*, 2011 Characterization of a plant-transformation-ready large-insert BIBAC library of Arabidopsis and bombardment transformation of a large-insert BIBAC of the library into tobacco. Genome 54: 437-447.

CHEN, M., G. PRESTING, W. B. BARBAZUK, J. L. GOICOECHEA, B. BLACKMON *et al.*, 2002 An integrated physical and genetic map of the rice genome. Plant Cell 14: 537-545.

CHEN, M., P. SANMIGUEL, A. C. DE OLIVEIRA, S. S. WOO, H. ZHANG *et al.*, 1997 Microcolinearity in sh2-homologous regions of the maize, rice, and sorghum genomes. Proc Natl Acad Sci U S A 94: 3431-3435.

CHIA, J. M., C. SONG, P. J. BRADBURY, D. COSTICH, N. DE LEON *et al.*, 2012 Maize HapMap2 identifies extant variation from a genome in flux. Nat Genet 44: 803-807.

CLARK, R. M., E. LINTON, J. MESSING and J. F. DOEBLEY, 2004 Pattern of diversity in the genomic region near the maize domestication gene tb1. Proc Natl Acad Sci U S A 101: 700-707.

DALRYMPLE, B. P., E. F. KIRKNESS, M. NEFEDOV, S. MCWILLIAM, A. RATNAKUMAR *et al.*, 2007 Using comparative genomics to reorder the human genome sequence into a virtual sheep genome. Genome Biol 8: R152.

DENG, Y., Y. PAN and M. LUO, 2013 Detection and correction of assembly errors of rice Nipponbare reference sequence. Plant Biology Accepted.

DICKO, M. H., H. GRUPPEN, A. S. TRAORE, A. G. J. VORAGEN and W. J. H. VAN BERKEL, 2006 Sorghum Grain as Human Food in Africa: Relevance of Starch Content and Amylase Activities. African Journal of Biotechnology 5: 384-395.

DING, J. Q., J. L. MA, C. R. ZHANG, H. F. DONG, Z. Y. XI *et al.*, 2011 QTL mapping for test weight by using F(2:3) population in maize. J Genet 90: 75-80.

DOEBLEY, J., 2004 The genetics of maize evolution. Annu Rev Genet 38: 37-59.

DOEBLEY, J. F., B. S. GAUT and B. D. SMITH, 2006 The molecular genetics of crop domestication. Cell 127: 1309-1321.

EICHTEN, S. R., J. M. FOERSTER, N. DE LEON, Y. KAI, C. T. YEH *et al.*, 2011 B73-Mo17 near-isogenic lines demonstrate dispersed structural variation in maize. Plant Physiol 156: 1679-1690.

EWING, B., and P. GREEN, 1998 Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res 8: 186-194.

EWING, B., L. HILLIER, M. C. WENDL and P. GREEN, 1998 Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res 8: 175-185.

FENG, J., B. A. VICK, M. K. LEE, H. B. ZHANG and C. C. JAN, 2006 Construction of BAC and BIBAC libraries from sunflower and identification of linkage group-specific clones by overgo hybridization. Theor Appl Genet 113: 23-32.

FU, H., and H. K. DOONER, 2002 Intraspecific violation of genetic colinearity and its implications in maize. Proc Natl Acad Sci USA 99: 9573-9578.

FU, Y., T. J. WEN, Y. I. RONIN, H. D. CHEN, L. GUO *et al.*, 2006 Genetic dissection of intermated recombinant inbred lines using a new genetic map of maize. Genetics 174: 1671-1683.

GORE, M. A., J.-M. CHIA, R. J. ELSHIRE, Q. SUN, E. S. ERSOZ *et al.*, 2009 A First-Generation Haplotype Map of Maize. Science 326: 1115-1117.

HAMILTON, C. M., 1997 A binary-BAC system for plant transformation with high-molecular-weight DNA. Gene 200: 107-116.

HAMILTON, C. M., A. FRARY, C. LEWIS and S. D. TANKSLEY, 1996 Stable transfer of intact high molecular weight DNA into plant chromosomes. Proc Natl Acad Sci U S A. 93: 9975-9979.

HARLAN, J. R., and J. M. DE WET, 1977 Pathways of genetic transfer from Tripsacum to Zea mays. Proc Natl Acad Sci U S A 74: 3494-3497.

HOISINGTON, D., M. KHAIRALLAH, T. REEVES, J. M. RIBAUT, B. SKOVMAND *et al.*, 1999 Plant genetic resources: what can they contribute toward increased crop productivity? Proc Natl Acad Sci U S A. 96: 5937-5943.

HURWITZ, B. L., D. KUDRNA, Y. YU, A. SEBASTIAN, A. ZUCCOLO *et al.*, 2010 Rice structural variation: a comparative analysis of structural variation between rice and three of its closest relatives in the genus Oryza. Plant J 63: 990-1003.

ILIC, K., P. J. SANMIGUEL and J. L. BENNETZEN, 2003 A complex history of rearrangement in an orthologous region of the maize, sorghum, and rice genomes. Proc Natl Acad Sci U S A 100: 12265-12270.

INTERNATIONAL RICE GENOME SEQUENCING PROJECT, 2005 The Map-based Sequence of the Rice Genome. Nature 436: 793-800.

JANDA, J., J. SAFAR, M. KUBALAKOVA, J. BARTOS, P. KOVAROVA *et al.*, 2006 Advanced resources for plant genomics: a BAC library specific for the short arm of wheat chromosome 1B. Plant J 47: 977-986.

JIAO, Y., H. ZHAO, L. REN, W. SONG, B. ZENG *et al.*, 2012 Genome-wide genetic changes during modern breeding of maize. Nat Genet 44: 812-815.

JUNG, M., A. CHING, D. BHATTRAMAKKI, M. DOLAN, S. TINGEY *et al.*, 2004 Linkage disequilibrium and sequence diversity in a 500-kbp region around the adh1 locus in elite maize germplasm. Theoretical and Applied Genetics 109: 681-689.

KIM, H., P. SAN MIGUEL, W. NELSON, K. COLLURA, M. WISSOTSKI *et al.*, 2007 Comparative physical mapping between Oryza sativa (AA genome type) and O. punctata (BB genome type). Genetics 176: 379-390.

LAI, J., J. MA, Z. SWIGONOVA, W. RAMAKRISHNA, E. LINTON *et al.*, 2004 Gene loss and movement in the maize genome. Genome Res 14: 1924-1931.

LAI, J. S., R. Q. LI, X. XU, W. W. JIN, M. L. XU *et al.*, 2010 Genome-wide patterns of genetic variation among elite maize inbred lines. Nat Genet 42: 1027-U1158.

LARMANDE, P., C. GAY, M. LORIEUX, C. PERIN, M. BOUNIOL *et al.*, 2008 Oryza Tag Line, a phenotypic mutant database for the Genoplante rice insertion line library. Nucleic Acids Research 36: D1022-D1027.

LEE, M. K., Y. ZHANG, M. ZHANG, M. GOEBEL, H. J. KIM *et al.*, 2013 Construction of a plant-transformation-competent BIBAC library and genome sequence analysis of polyploid Upland cotton (Gossypium hirsutum L.). BMC Genomics 14: 208.

LEWIN, H. A., D. M. LARKIN, J. PONTIUS and S. J. O'BRIEN, 2009 Every genome sequence needs a good map. Genome Res 19: 1925-1928.

LI, Y., C. FAN, Y. XING, Y. JIANG, L. LUO *et al.*, 2011 Natural variation in GS5 plays an important role in regulating grain size and yield in rice. Nat Genet 43: 1266-1269.

LIN, H., P. XIA, A. W. R, Q. ZHANG and M. LUO, 2012 Dynamic intra-japonica subspecies variation and resource application. Mol Plant 5: 218-230.

LIN, Y.-R., L. ZHU, S. REN, J. YANG, K. SCHERTZ *et al.*, 1999 A Sorghum propinquum BAC library, suitable for cloning genes associated with loss-of-function mutations during crop domestication. Molecular Breeding 5: 511-520.

LIU, Y. G., H. LIU, L. CHEN, W. QIU, Q. ZHANG *et al.*, 2002 Development of new transformation-competent artificial chromosome vectors and rice genomic libraries for efficient gene cloning. Gene 282: 247-255.

LIU, Y. G., K. NAGAKI, M. FUJITA, K. KAWAURA, M. UOZUMI *et al.*, 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.

LIU, Y. G., Y. SHIRANO, H. FUKAKI, Y. YANAI, M. TASAKA *et al.*, 1999 Complementation of plant mutants with large genomic DNA fragments by a transformation-competent artificial chromosome vector accelerates positional cloning. Proc Natl Acad Sci U S A 96: 6535-6540.

LOUGH, A. N., L. M. ROARK, A. KATO, T. S. REAM, J. C. LAMB *et al.*, 2008 Mitochondrial DNA transfer to the nucleus generates extensive insertion site variation in maize. Genetics 178: 47-55.

LU, F., J. S. AMMIRAJU, A. SANYAL, S. ZHANG, R. SONG *et al.*, 2009 Comparative sequence analysis of MONOCULM1-orthologous regions in 14 Oryza genomes. Proc Natl Acad Sci U S A 106: 2071-2076.

LUO, M., H. KIM, D. KUDRNA, N. B. SISNEROS, S. J. LEE *et al.*, 2006 Construction of a nurse shark (Ginglymostoma cirratum) bacterial artificial chromosome (BAC) library and a preliminary genome survey. Bmc Genomics 7: 106.

LUO, M., Y. H. WANG, D. FRISCH, T. JOOBEUR, R. A. WING *et al.*, 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.

LUO, M., and R. A. WING, 2003 An improved method for plant BAC library construction. Methods Mol Biol 236: 3-20.

LUO, M. C., C. THOMAS, F. M. YOU, J. HSIAO, S. OUYANG *et al.*, 2003 High-throughput fingerprinting of bacterial artificial chromosomes using the snapshot labeling kit and sizing of restriction fragments by capillary electrophoresis. Genomics 82: 378-389.

MAO, L., D. BEGUM, H. W. CHUANG, M. A. BUDIMAN, E. J. SZYMKOWIAK *et al.*, 2000 JOINTLESS is a MADS-box gene controlling tomato flower abscission zone development. Nature 406: 910-913.

MEKSEM, K., K. ZOBRIST, E. RUBEN, D. HYTEN, T. QUANZHOU *et al.*, 2000 Two large-insert soybean genomic libraries constructed in a binary vector: applications in chromosome walking and genome wide physical mapping. Theoretical and Applied Genetics 101: 747-755.

MENG, X., B. HUANG, L. ZHOU, Y. HE, Q. CHEN *et al.*, 2010 Construction of a Coix BAC library and isolation of the 22 kDa α-coixin gene cluster. Genome 53: 667-674.

MESSING, J., A. K. BHARTI, W. M. KARLOWSKI, H. GUNDLACH, H. R. KIM *et al.*, 2004 Sequence composition and genome organization of maize. Proc Natl Acad Sci U S A 101: 14349-14354.

MESSING, J., and V. LLACA, 1998 Importance of anchor genomes for any plant genome project. Proc Natl Acad Sci U S A 95: 2017-2020.

NELSON, W. M., A. K. BHARTI, E. BUTLER, F. WEI, G. FUKS *et al.*, 2005 Whole-genome validation of high-information-content fingerprinting. Plant Physiol 139: 27-38.

NILMALGODA, S. D., S. CLOUTIER and A. Z. WALICHNOWSKI, 2003 Construction and characterization of a bacterial artificial chromosome (BAC) library of hexaploid wheat (Triticum aestivum L.) and validation of genome coverage using locus-specific primers. Genome 46: 870-878.

O'SULLIVAN, D. M., P. J. RIPOLL, M. RODGERS and K. J. EDWARDS, 2001 A maize bacterial artificial chromosome (BAC) library from the European flint inbred line F2. Theor Appl Genet 103: 425-432.

PATERSON, A. H., J. E. BOWERS, R. BRUGGMANN, I. DUBCHAK, J. GRIMWOOD *et al.*, 2009 The Sorghum bicolor genome and the diversification of grasses. Nature 457: 551-556.

QI, X., S. LINDUP, T. S. PITTAWAY, S. ALLOUIS, M. D. GALE *et al.*, 2001 Development of simple sequence repeat markers from bacterial artificial chromosomes without subcloning. Biotechniques 31: 355, 358-362.

QI, X., T. S. PITTAWAY, S. LINDUP, H. LIU, E. WATERMAN *et al.*, 2004 An integrated genetic map and a new set of simple sequence repeat markers for pearl millet, Pennisetum glaucum. Theor Appl Genet 109: 1485-1493.

QU, S., G. COAKER, D. FRANCIS, B. ZHOU and G. L. WANG, 2003 Development of a new transformation-competent artificial chromosome (TAC) vector and construction of tomato and rice TAC libraries. Molecular Breeding 12: 297-308.

ROARK, L. M., A. Y. HUI, L. DONNELLY, J. A. BIRCHLER and K. J. NEWTON, 2010 Recent and frequent insertions of chloroplast DNA into maize nuclear chromosomes. Cytogenet Genome Res 129: 17-23.

SASAKI, T., and B. A. ANTONIO, 2009 Plant genomics: Sorghum in sequence. Nature 457: 547-548.

SCHNABLE, P. S., D. WARE, R. S. FULTON, J. C. STEIN, F. WEI *et al.*, 2009 The B73 maize genome: complexity, diversity, and dynamics. Science 326: 1112-1115.

SHI, X., H. ZENG, Y. XUE and M. LUO, 2011 A pair of new BAC and BIBAC vectors that facilitate BAC/BIBAC library construction and intact large genomic DNA insert exchange. Plant Methods 7.

SODERLUND, C., I. LONGDEN and R. MOTT, 1997 FPC: a system for building contigs from restriction fingerprinted clones. Comput Appl Biosci 13: 523-535.

SONG, J., J. M. BRADEEN, S. K. NAESS, J. P. HELGESON and J. JIANG, 2003 BIBAC and TAC clones containing potato genomic DNA fragments larger than 100 kb are not stable in Agrobacterium. Theor Appl Genet 107: 958-964.

SONG, R., V. LLACA and J. MESSING, 2002 Mosaic organization of orthologous sequences in grass genomes. Genome Res 12: 1549-1555.

SONG, R., and J. MESSING, 2003 Gene expression of a gene family in maize based on noncollinear haplotypes. Proc Natl Acad Sci U S A 100: 9055-9060.

SONG, R., G. SEGAL and J. MESSING, 2004 Expression of the sorghum 10-member kafirin gene cluster in maize endosperm. Nucleic Acids Res 32: e189.

SONG, W. Y., G. L. WANG, L. L. CHEN, H. S. KIM, L. Y. PI *et al.*, 1995 A receptor kinase-like protein encoded by the rice disease resistance gene, Xa21. Science 270: 1804-1806.

SONG, X., J. L. GOICOECHEA, J. S. AMMIRAJU, M. LUO, R. HE *et al.*, 2011 The 19 genomes of Drosophila: a BAC library resource for genus-wide and genome-scale comparative evolutionary research. Genetics 187: 1023-1030.

SWANSON-WAGNER, R., R. BRISKINE, R. SCHAEFER, M. B. HUFFORD, J. ROSS-IBARRA *et al.*, 2012 Reshaping of the maize transcriptome by domestication. Proc Natl Acad Sci U S A 109: 11878-11883.

SWIGONOVA, Z., J. LAI, J. MA, W. RAMAKRISHNA, V. LLACA *et al.*, 2004 Close split of sorghum and maize genome progenitors. Genome Res 14: 1916-1923.

TAO, Q., A. WANG and H. B. ZHANG, 2002 One large-insert plant-transformation-competent BIBAC library and three BAC libraries of Japonica rice for genome research in rice and other grasses. Theor Appl Genet 105: 1058-1066.

TIKHONOV, A. P., P. J. SANMIGUEL, Y. NAKAJIMA, N. M. GORENSTEIN, J. L. BENNETZEN *et al.*, 1999 Colinearity and its exceptions in orthologous adh regions of maize and sorghum. Proc Natl Acad Sci U S A 96: 7409-7414.

TOMKINS, J. P., G. DAVIS, D. MAIN, Y. YIM, N. DURU *et al.*, 2002 Construction and Characterization of a Deep-Coverage Bacterial Artificial Chromosome Library for Maize. Crop Sci. 42: 928-933.

VAN HEERWAARDEN, J., J. DOEBLEY, W. H. BRIGGS, J. C. GLAUBITZ, M. M. GOODMAN *et al.*, 2011 Genetic signals of origin, spread, and introgression in a large sample of maize landraces. Proc Natl Acad Sci U S A 108: 1088-1092.

VEGA, J. M., W. YU, F. HAN, A. KATO, E. M. PETERS *et al.*, 2008 Agrobacterium-mediated transformation of maize (Zea mays) with Cre-lox site specific recombination cassettes in BIBAC vectors. Plant Mol Biol 66: 587-598.

VERMERRIS, W., 2011 Survey of genomics approaches to improve bioenergy traits in maize, sorghum and sugarcane. J Integr Plant Biol 53: 105-119.

WALBOT, V., 2009 10 reasons to be tantalized by the B73 maize genome. PLoS Genet 5: e1000723.

WANG, G. L., T. E. HOLSTEN, W. Y. SONG, H. P. WANG and P. C. RONALD, 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.

WANG, H., and J. L. BENNETZEN, 2012 Centromere retention and loss during the descent of maize from a tetraploid ancestor. Proc Natl Acad Sci U S A 109: 21004-21009.

WANG, Q., and H. K. DOONER, 2006 Remarkable variation in maize genome structure inferred from haplotype diversity at the bz locus. Proc Natl Acad Sci U S A 103: 17644-17649.

WANG, Q., and H. K. DOONER, 2012 Dynamic evolution of bz orthologous regions in the Andropogoneae and other grasses. Plant J 72: 212-221.

WANG, X., Q. LIU, H. WANG, C.-X. LUO, G. WANG *et al.*, 2013 A BAC based physical map and genome survey of the rice false smut fungus Villosiclava virens. BMC Genomics.

WEI, F., E. COE, W. NELSON, A. K. BHARTI, F. ENGLER *et al.*, 2007 Physical and genetic structure of the maize genome reflects its complex evolutionary history. PLoS Genet 3: e123.

WEI, F., J. ZHANG, S. ZHOU, R. HE, M. SCHAEFFER *et al.*, 2009 The physical and genetic framework of the maize B73 genome. PLoS Genet 5: e1000715.

WING, R. A., J. S. AMMIRAJU, M. LUO, H. KIM, Y. YU *et al.*, 2005 The oryza map alignment project: the golden path to unlocking the genetic potential of wild rice species. Plant Mol Biol 59: 53-62.

WOO, S. S., J. JIANG, B. S. GILL, A. H. PATERSON and R. A. WING, 1994 Construction and characterization of a bacterial artificial chromosome library of Sorghum bicolor. Nucleic Acids Res 22: 4922-4931.

WOODHOUSE, M. R., J. C. SCHNABLE, B. S. PEDERSEN, E. LYONS, D. LISCH *et al.*, 2010 Following tetraploidy in maize, a short deletion mechanism removed genes preferentially from one of the two homologs. PLoS Biol 8: e1000409.

WU, C. C., P. NIMMAKAYALA, F. A. SANTOS, R. SPRINGMAN, C. SCHEURING *et al.*, 2004 Construction and characterization of a soybean bacterial artificial chromosome library and use of multiple complementary libraries for genome physical mapping. Theor Appl Genet 109: 1041-1050.

YIM, Y. S., G. L. DAVIS, N. A. DURU, T. A. MUSKET, E. W. LINTON *et al.*, 2002 Characterization of three maize bacterial artificial chromosome libraries toward anchoring of the physical map to the

genetic map using high-density bacterial artificial chromosome filter hybridization. Plant Physiol 130: 1686-1696.

YU, J., J. WANG, W. LIN, S. LI, H. LI *et al.*, 2005 The Genomes of Oryza sativa: a history of duplications. PLoS Biol 3: e38.

ZHAI, J., Y. WANG, C. SUN, S. JIANG, K. WANG *et al.*, 2013 A plant-transformation-competent BIBAC library of ginseng (Panax ginseng C. A. Meyer) for functional genomics research and characterization of genes involved in ginsenoside biosynthesis. Molecular Breeding 31: 685-692.

ZHAO, S., J. SHETTY, L. HOU, A. DELCHER, B. ZHU *et al.*, 2004 Human, mouse, and rat genome large-scale rearrangements: stability versus speciation. Genome Res 14: 1851-1860.