Genomic Resources of Broomcorn Millet: Demonstration and Application of a High-throughput BAC Mapping Pipeline

Wei Xu[1], Mengjie Liang[1], Xue Yang[1], Hao Wang[1], Meizhong Luo[1*]

[1]College of Life Science and Technology, Huazhong Agricultural University, Wuhan 430070, China

*Correspondence: mzluo@mail.hzau.edu.cn

Abstract

Background: With high-efficient water-use and drought tolerance, broomcorn millet has emerged as a candidate for food security. To promote its research process for molecular breeding and functional research, a comprehensive genome resource is of great importance.

Results: Herein, we constructed a BAC library for broomcorn millet, generated BAC end sequences based on the clone-array pooled shotgun sequencing strategy and Illumina sequencing technology, and integrated BAC clones into genome by a novel pipeline for BAC end profiling. The BAC library is consisted of 76,023 clones with an average insert length of 123.48 Kb, covering about 9.9-fold of the 850 Mb genome. Of 9,216 clones tested using our pipeline, 8,262 clones were mapped on the broomcorn millet cultivar longmi4 genome. These mapped clones covered 308 of the 829 gaps left by the genome. To our knowledge, this is the only BAC resource for broomcorn millet.

Conclusions: We constructed a high-quality BAC libraray for broomcorn millet and designed a novel pipeline for BAC end profiling. BAC clones can be browsed and obtained from our

website (http://eightstarsbio.com/gresource/JBrowse-1.16.5/index.html). The high-quality BAC clones mapped on genome in this study will provide a powerful genomic resource for genome gap filling, complex segment sequencing, FISH, functional research, and genetic engineering of broomcorn millet.

Keywords: Broomcorn millet, BAC, BES, Genomics resources, Gap filling, CAPSS, Jbrowse

1 Introduction

With the increased global water scarcity caused by climate change and population growth, it is of great importance to exploit the high-efficient water-use crop for the food security of the human in the future. Broomcorn millet (*Panicum miliaceum*), also known as proso millet, panic millet, and wild millet, is one of the traditional five-grain crops in the north of China [1]. It is a typical C4 plant with high photosynthetic efficiency. It also has a high-efficient utilization ratio of water resource and a capacity for drought resistance, exquisitely adapting to semi-drought or drought conditions [2]. Furthermore, its growing cycle, 60–90 days from sowing to maturity, is shorter than other cereals [3]. Broomcorn millet contains more protein than most grains, and a relatively balance array of trace elements and vitamins. More than 8700 accessions of broomcorn millet including landraces and cultivars have been conserved in the National Gene Bank of the Institute of Crop Science, Chinese Academy of Agricultural Sciences, thus providing an abundance of resources for genetic improvement of broomcorn millet.

At present, high-quality chromosome-scale genome assemblies of two allotetraploid (2n = 4x = 36) broomcorn millet varieties decoded by Chinese researchers are available [4, 5]. These genome assemblies provide the foundation for the molecular breeding of broomcorn

47 millet. However, other gemonic resources are still required to complete and make the full use
48 of the genome assemblies.

49 The tranditional bacterial artificial chromosome (BAC) libraries with genomic DNA
50 inserts of 50 kbp – 300 kbp [6–8] are also important resources for genomic research. They
51 provide natural DNA materials for a variety of experiments, such as intact gene cluster
52 cloning, map-based cloning, whole genome sequencing, comparative genomics analysis and
53 fluorescence *in situ* hybridization that aim to understand functional elements in the genome
54 [9, 10]. The utility of the BAC libraries can be greatly enhanced by mapping the BAC clones
55 on the genome assemblies.

56 Hence, we constructed a high-quality BAC library for broomcorn millet, developed a
57 new pipeline for cost-effectively decoding BAC end seqeunces generated by clone-array
58 pooled shotgun sequencing strategy (abbr. CAPSS) and Illumina sequencing technology, and
59 mapped the BAC clones on the genome assemblies of the broomcorn millet [11].

60

61 2 Results

62 2.1 BAC library construction

63 A BAC library of broomcorn millet was constructed with the restriction enzyme *Hin*dIII
64 using high-molecular-weight genomic DNA prepared from etiolated seedlings. In total, the
65 library consists of 76,032 BAC clones, that were arrayed into 198 independent 384-well
66 microtiter plates. Insert sizing of randomly picked clones showed that the majority of
67 genomic BAC inserts fell into the length range of 97–145.5 kb with an average insert size of
68 132 kb (Fig. 1).

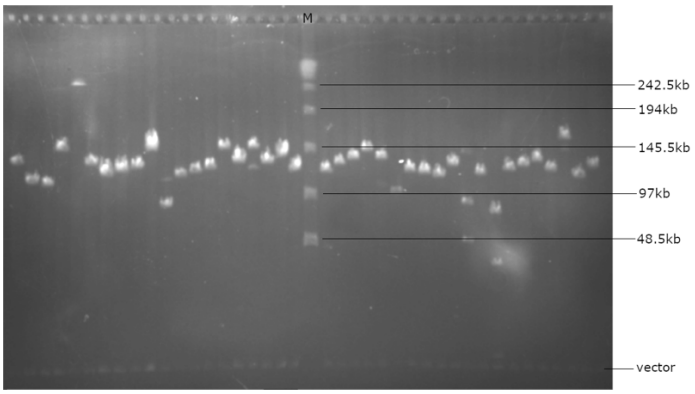

Fig. 1 The insert sizes of randomly selected BAC clones determined by PFGE. The maker in the middle is λ DNA ladder.

2.2 Construction and sequencing of DNA pools

In order to obtain BAC end sequences with high efficiency and low cost, we designed and performed a pipeline based on the clone array pooled shotgun sequencing strategy (Fig. 2). Randomly selected 24 384-plates (each plate consists of 16 x 24 clones) from the broomcorn millet BAC library were arranged to a square superpool with 6 row plates and 4 column plates. Therefore, the superpool consisted of 96 row pools (6 plates x 16) and 96 column pools (4 plates x 24). These row pools and column pools were called as the secondary pools, and each secondary pool also consisted of 96 BAC clones. In total, one superpool consisted of 192 secondary pools and 9,216 individual BAC clones.

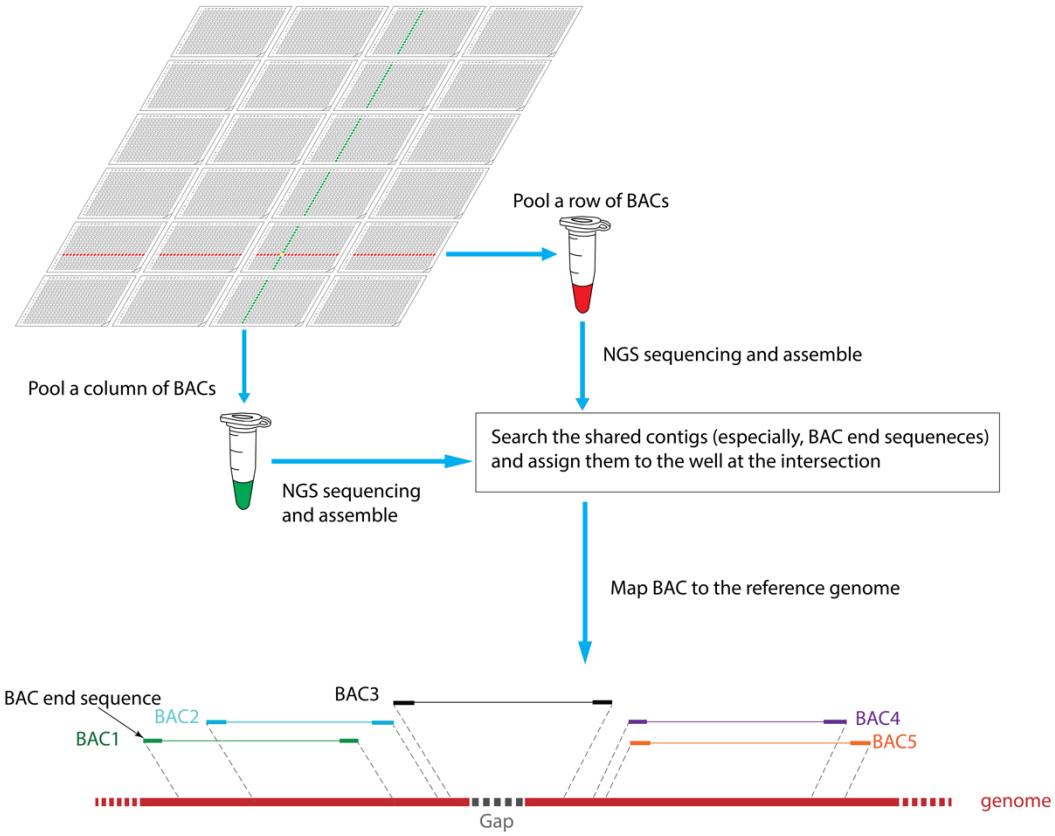

Fig. 2 The strategy of BAC end sequence localization based on CAPSS. Twenty-four 384-plates are arranged to a square superpool, and then 96 row pools and 96 column pools are prepared and sequenced by NGS platform. The sequences of each pool are assembled into contigs. If a contig (especially, BAC end sequences, abbr. BES) is shared in a row pool and a column pool, it will be assigned to the well at the intersection of the row pool and the column pool. BAC clones will be further mapped to the reference genome according to assigned contigs.

DNA of the secondary pools were extracted and Illumina sequencing libraries with individual index sequences for each secondary pool were prepared. Finally, 96 row pool libraries were mixed in equal amounts as the library X, and 96 column pool libraries were also mixed in equal amounts as the library Y. The average insert sizes of the library X and Y analyzed by the Agilent 2100 Bioanalyzer (Additional file 1: Fig. S1) were 491 bp and 443 bp, respectively. Sequencing was accomplished using the Illumina HiSeq 2000 sequencing

platform with paired-end protocol (PE150). The Library X and Y generated 130.82 Gb and 133.50 Gb raw data, respectively. Trimmomatic was employed for trimming adaptors and filtering low-quality or shorter reads. FastQC was employed for evaluating the quality of the preprocessing reads. Then, the valid reads were obtained by filtering the reads traced to *E. coli* DH10B reference genome using Bowtie2. Finally, we obtained valid yields of 60.40 Gb and 102.37 Gb for libraries X and Y, respectively (Table 1). Demultiplexer was employed for demultiplexing and generating each secondary pool reads according to the 7-bp index (Additional file 2). The average valid sequencing depths of row pools and column pools were 45x and 78x, respectively (Additional file 1: Fig. S2).

Table 1 The summary of NGS data

Library	Total size (Gb)	Size after QC (Gb)	Valid size (Gb)
X	130.82	103.78	60.40
Y	133.50	107.92	102.37

2.3 Parsing BAC end sequences

We focused our interest on the two special contigs in a BAC: the forward and reverse BAC end seqeunces. We designed two pathways to find them. On one hand, the paired-end reads overlapping with the border sequences of the vector harboring the *Hin*dIII restriction enzyme site were extracted from the valid data of each secondary pool, and then assembled using Cap3 based on the overlap-layout-consensus method. The vector sequence parts in the consensus were truncated to obtain short BAC end sequences (abbr. BES) starting with AAGCTT (*Hin*dIII site). These short BAC end sequences were assigned to the corresponding wells according to the clone array pooled shotgun sequencing strategy. Of the 9,216 wells in 384-plates tested in the pipeline, 8,183 (88.79%) wells were assigned one forward BES and

113 7,897 (85.69 %) wells were assigned one reverse BES (Additional file 3: Table S1). All
114 assigned BAC end sequences have an average length of 358 bp (Additional file 1: Fig. S3),
115 which is consistent with the insert size of the Illumina libraries (Additional file 1: Fig. S1).

116 On the other hand, the valid NGS data of each secondary pool were firstly assembled
117 using SPAdes, and the contig N50 sizes of all row pools and column pools were counted
118 (Additional file 1: Fig. S4). The average N50 of row pools and column pools were 13.97 kb
119 and 11.45 kb, respectively. Likewise, the vector sequence parts at the ends of the contigs
120 were removed to retain the long BESs starting with AAGCTT. Finally, these long BESs were
121 assigned to the corresponding wells in 384-plates. Of the 9,216 wells, 5,454 (59.18%) wells
122 obtained one forward BES and 5,108 (55.43%) wells obtained one reverse BES (Additional
123 file 3: Table S1). The N50 of the long BESs is 60.87 kb (Additional file 1: Fig. S5).

124 2.4 Determination of BAC locations on the genome

125 In order to determine the locations of the broomcorn millet BAC clones on the genome,
126 we aligned the short and long BAC end sequences onto the cultivar longmi4 genome with
127 Blastn. Table 2 listed the alignment results. With BAC end sequences with repeats, 14,862
128 (83.44%) short BAC end sequences including 12,971 single-hit and 1,891 multi-hit sequences
129 were mapped to the genome, and 12,120 (95.28%) long BAC end sequences (all single-hit)
130 were mapped to the genome. With the BAC end sequences without repeats, 7,760 (43.57%)
131 short BAC end sequences including 7,295 single-hit and 465 multi-hit sequences were
132 mapped to the genome, and 10,626 (83.53%) long BAC end sequences (all single-hit) were
133 mapped to the genome. In order to map as many as possible BACs to the genome, the
134 alignment results of the end sequences with repeats were adopted. We wrote a python script
135 to extract the Blast results from short and long BESs. As a result, 5,795 BACs were mapped
136 to the genome using short BAC end sequences, and 6,973 BACs were mapped to the genome

137 using long BAC end sequences. Finally, the Blast results generated by short and long BAC
138 end sequences were integrated, and in total 8,262 BACs (89.65%) were mapped to the
139 genome.

140 **Table 2 A summary of the Broomcorn millet BESs and the anchoring results of the Broomcorn**
141 **millet BAC clones to the longmi 4 genome using the BESs**

Categories	Short BESs	Long BESs	Integrity
BAC end sequences			
Clones in superpool	9216	9216	
Clones with successful BESs	9126	7725	
with paired BESs	7790	3890	
with single-end BESs	1336	3835	
only forward	921	2189	
only reverse	415	1646	
Total successful BESs	17811 (96.63%)	12721 (66.17%)	
Alignment			
Aligned BESs with repeats unmasked	14862 (83.44%)	12120 (95.28%)	
Single-hit BESs	12971	12120	
Multi-hit BESs	1891	0	
Aligned BESs with repeats masked	7760 (43.57%)	10626 (83.53%)	
Single-hit BESs	7295	10626	
Multi-hit BESs	465	0	
Anchoring to reference sequences			
Clones anchored to single sites	5795	6973	8262
Clones anchored with single BES	2507	3907	2871
Clones anchored with paired BESs	3288	3066	5391

142

143 To verify the accuracy of BAC locations on the genome, 55 BAC clones were randomly
144 picked from 384-plates for BAC end sequencing using Sanger method. After quality control,
145 the 55 paired Sanger end sequences were blasted with the above short and long BESs, and the
146 genome. All BAC clones but one (32G16) were consistent. By checking the 32G16 BAC end
147 sequences, we found that this clone (or rather 32G16 well) was assigned two long forward
148 BESs, one short forward BES, and no reverse BES. Only one long forward BES and the short
149 forward BES were perfectly identical to the Sanger end sequence. Another long forward BES
150 confused the mapping. However, it can be solved by weighting. In summary, the accuracy of
151 the BAC mapping approach in this study was extremely high.

152 The distribution of the 5,391 BACs mapped to the genome by paired BAC end
153 sequences were counted (Table 3). These clones covered a total of 432.47 Mb of
154 chromosomes with a total coverage of 50.97%. Among the 18 chromosome sequences of the
155 cultivar longmi4 genome, there left 829 gaps. Our BACs covered 308 of them. The insert
156 sizes of these BACs presented the Gaussian distribution, with an average insert size of 123.48
157 kb (Fig. 3), which is lower than that predicted by pulse field electrophoresis. These BACs are
158 valuable resources for further improvement of the genome.

159 **Table 3 The location result of BACs on broomcorn millet chromosomes**

Chr.	Length (Mb)	No. gaps	No. BACs	BACs Covered	
				length (Mb)	No. gaps
chr 1	69.18	85	439	35.25	34
chr 2	61.15	59	379	31.04	19
chr 3	57.97	50	394	31.03	20
chr 4	56.29	34	359	27.90	14
chr 5	54.13	52	361	29.86	19
chr 6	52.84	46	358	28.24	16

chr 7	51.23	67	286	23.67	27
chr 8	48.26	29	339	26.49	8
chr 9	45.11	70	240	20.96	21
chr10	44.65	53	308	25.44	28
chr11	43.18	30	267	21.90	14
chr12	42.47	30	254	20.79	11
chr13	40.72	50	261	21.11	17
chr14	38.49	32	269	20.23	10
chr15	34.36	34	213	18.14	8
chr16	33.61	45	212	16.72	15
chr17	32.99	25	226	18.26	10
chr18	32.24	38	199	15.32	17
unplaced	9.48	0	27	-	0
Total	848.47	829	5391	432.47	308

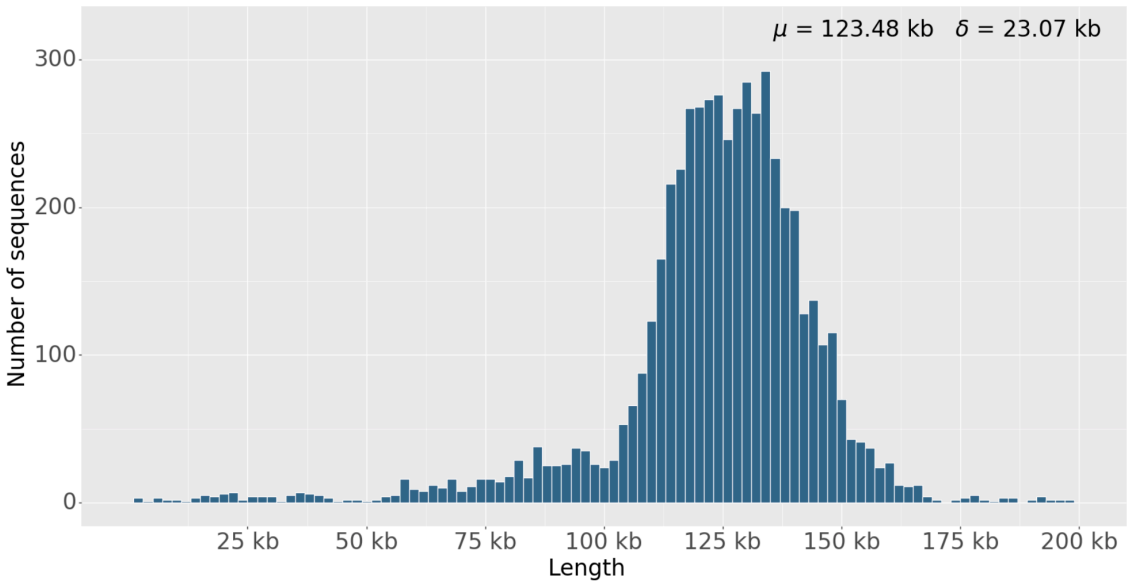

Fig. 3. the statistics of the BAC inserts mapped by paired BESs

2.5 Presentation of BAC locations by JBrowse

In order to effectively use the BAC resource, quickly and easily retrieve BAC clones and view other annotaion information, we established a resource website employing a lightweight JBrowse to display the BAC resource information on the broomcorn millet genome (Fig. 4).

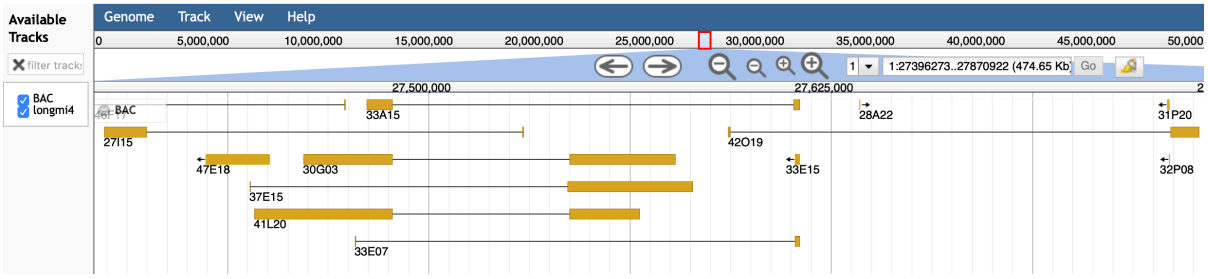

Fig. 4. Presentation of broomcorn millet BAC locations by JBrowse. A barbell icon idicates a BAC mapped by forward and reverse BESs (yellow rectangle). An icon with arrow indicates a BAC mapped by only single BES.

2.6 Organelle genomes of the broomcorn millet

We aligned all BESs in this study to the chloroplast (CM009689) genome of broomcorn millet by local blast. A total of 65 BAC clones were mapped to the chloroplast with more than 99% of similarity, of which 16 BAC clones were determined by paired BESs (25C07, 25N09, 27F16, 27O19, 29B02, 29K11, 32J09, 32N16, 33D23, 33K24, 36A05, 39C22, 42C10, 46C16, 48B15, 48C03). In the absence of mitochrondrial sequence in broomcorn millet, we aligned BESs to all mitochondrial genomes in *Gramineae*. Of 16 BAC clones with homologous BES sequences, 13 BACs were simultaneously aligned to the chloroplast and 3 BACs were simultaneously aligned to the nuclear genome.

3 Discussion

BAC library is still a powerful resource for genome assembly, functional genomics research, and long-term genetic resoure storage for the endangered species. BAC seqeunces,

182 espcially paired BESs, are usally used to detect assembly errors or assist assembly [12]. The
183 most conventional and convenient approach to decode BESs is Sanger sequencing method.
184 However, its one-by-one style is laborious, time-consuming and expensive. Fortunately, a
185 few of efficient approaches had developped, based on next generation sequencing
186 technologies, such as pBACcode and BAC-anchor[13, 14]. pBACode determines paired
187 BESs by a pair of random barcodes flanking the cloning site. BAC-anchor determines general
188 paired BESs by using specific restriction enzyme sites and searching for utlra-long paired-
189 end subreads containing large internal gaps. We previously also developed a high-throughput
190 approach for long accurate BES profiling with PacBio sequencing technology [15]. However,
191 these approaches focused on paired BES profiles, and they are hard to trace BESs of specific
192 clones in 384-plate wells. In this study, we generated BESs and assigned them to physical
193 wells in 384-plates by applying the characteristics of row-cross-column in two-dimension
194 arrays and cost-effective Illumina platform. These BAC end sequences in average are longer
195 and accurater than that generated by Sanger method, for assembling by SPAdes
196 assembler[16]. After alignment, 89.65% of clones were successfully associated with the
197 broomcorn millet longmi 4 genome. These clones covered 308 of the 829 gaps left by the
198 genome and can be uesd to close the genome gaps.

199 In conventional whole genome sequencing project, sequencing coverage is usallly at
200 least 30x, and not more than 200x. Higher coverage will result in more sequences that are
201 generated by PCR mutations or sequencing errors and lead to contigs with shorter N50. We
202 added index into each secondary pool and mixed 96 secondary pools as a sequencing library
203 to run in a lane of Illumina flow cell for greatly reducing the cost of sequencing. As a result,
204 after removing the contaminated *E. coli* genomic DNA reads the average valid sequencing
205 depths of row pools and column pools were 45x and 78x, respectively.

206 In the process of BESs extraction, we designed two pathways: short BES pathway and
207 long BES pathway. Short BES pathway searched all reads overhanging with verctor end
208 sequences before assembling by Cap3; consequently, it generated all potenital BESs. Long
209 BES pathway identified all contigs overhanging with vector end seqeunces after assembling
210 by SPAdes; consequently, it generated longer but less BESs than the first pathway. The
211 assignment of BAC end sequences at the intersection sites is affected by many factors, such
212 as the overlapping rate of the BAC clones and the correct rate of the sequences. The
213 overlapping rate of BAC clones in the superpool is the most important factor. If overlapping
214 BAC clones appear in the same row or column pool, we can assign them to wells easily and
215 correctly. Also If two overlapping BAC clones appear in a different row or column pool, we
216 can rectify them by our previous method [17]. However, if more than two overlapping BAC
217 clones appear in a different row or column pool, our method will filter out potential BESs, so
218 that the BESs of the intersection well will be absent. In the process of BES assignment, the
219 flow of the forward and reverse BESs are completely independent. When more than one
220 forward and/or reverse BESs were assigned to a well, we cannot determine which BESs are a
221 pair of BESs. If a high-qualty genome is available, it is easy to assess which pair of BESs are
222 derived from the same BAC by mapping. However, if the variety used for BAC library
223 construction is not the same as that the reference genome stands for, the alignment results
224 that do not satisfy the location requirement of BES pairs will be discarded. The clones that
225 may contain a large structural variation can be picked out for further analysis from the 384-
226 plate wells.

227 Plant cells contain an abundance of chloroplasts and mitochondria. The chloroplast
228 genome is generally around 150 kb, while the mitochondrial genome size varies widely,
229 typically between 200 kb and 750 kb [18]. Although nuclei are extracted for BAC library
230 construction, a trace of organelle DNA cotamination is inevitable. In this study, a small

231 number of clones of chloroplast genome was found by BESs, while clones of mitochondrial
232 genome were almost absent.

233 By high-throughput sequencing and mapping clones in the secondary pool of BAC
234 libraries, we can find the coordinates of genome sequences in the broomcorn millet BAC
235 libraries. Therefore, if we find genes that play an important role in biology, we can quickly
236 locate the BAC clones containing this gene, and obtain the experimental materials for further
237 research and analysis. At the same time, because BAC contains a long DNA fragment (about
238 120 kb), it is also convenient for us to quickly analyze the upstream and/or downstream DNA
239 elements of interested genes or adjacent genes.

240 4 Conclusions

241 We constructed a high-quality BAC library for broomcorn millet, developed a high-
242 efficient and low-cost pipeline which can generate and parse ten thousand of paired BAC end
243 sequences at each time, and mapped a total of 8,262 broomcorn millet BACs to the
244 chromosomes. These clones covered 308 of the 829 gaps left by the genome. The high-
245 quality BAC clones mapped on genome in this study will provide a powerful genomic
246 resource for gap filling, complex segment sequencing, FISH, functional research, and genetic
247 engineering of broomcorn millet.

248 5 Methods

249 5.1 Plant materials, growth conditions and BAC library construction

250 The seeds of broomcorn millet were provided by professor Mingsheng Chen of the
251 Institute of Genetics and Developmental Biology, Chinese Academy of Sciences and grown
252 at dark conditions under 25°C. The young leaves of seedlings were harvested and mixed,

253 frozen immediately in liquid nitrogen for the extraction of nuclear DNA. High molecular
254 weight nuclear DNA was extracted and BAC library was constructed following our previous
255 protocol [6]. Partial digestions of DNA plugs with dilution *Hin*dIII were performed. DNA
256 fragments ranging from 100 kb to 200 kb were recovered from pulse field gel and ligated
257 with pIndigoBAC536-S vector [7]. The ligation product was used to transform DH10B cells
258 by electroporation. White colonies were picked up and stored in 384-plates at -80 ºC.

259 5.2 Pool construction

260 Twenty-four 384-plates were chosen form the BAC library of broomcorn millet. These
261 plates were arranged in a 2-dimension superpool. The superpool contains 96 row pools and
262 96 column pools. A total of 192 pools were processed independently for high quality DNA
263 extraction with the AxyGen AxyPrep Easy-96 plasmid kit. Then, these BAC DNAs were
264 completely digested with ATP-Dependent DNase (Epicentre) to remove the host *E. coli* DNA.
265 After digestion, these BAC DNAs were sheared in the Bioruptor to an average of 500 bp.
266 During blunt-end repair, overhanging 5' and 3' ends were filled in or removed by T4 DNA
267 polymerase. 5'-phosphates were attached using T4 polynucleotide kinase. Tail-A were added
268 using Taq DNA polymerase. Then, adapters were ligated to both ends of the molecules using
269 T4 DNA ligase. The ligation products were cleaned using MagBead DNA Purification Kit
270 (Sangon Biotech, shanghai, CN). Sequencing adaptors were added using PCR. Finally, the
271 products of KOD PCR were cleaned again.

272 5.3 Illumina sequencing and BAC end sequence analysis

273 NGS sequencing of 2 mixed DNA libraries was performed via the Illumina HiSeq 2000
274 with 150-bp paired-end protocol (Genewiz, suzhou, CN). Trimmotatic was employed to filter
275 and trim raw reads. FastQC was employed to assess data quality. After adapter filtering and

276 quality assessment, BBMAP/demuxbyname script was employed for deconvolution
277 depending on unique index sequence in each pool. A python script was employed to extract
278 target paired-end reads that cover BES-VES site. Cap3 was employed to assemble those
279 reads to consensuses, and consensuses were then trimmed to remove the part of vector to
280 generate BESs called "short BES". SPAdes was employed to directly assemble pool reads to
281 contigs. Then, the contigs containing BAC vector sequences were extracted and trimmed
282 using python script to generate the trimmed contigs called "long BES". Blastn was used to
283 align BESs from raw and column pools, and then the shared BESs were assigned to the wells
284 at the intersection.

285 5.4 Validation of BAC end sequences

286 The analysis results of the BAC end sequences were validated by Sanger sequencing.
287 Fifty-five BAC clones were randomly selected and their DNAs were extracted using an
288 improved alkaline lysis protocol. Sanger sequencing was accomplished using BAC-F (5'-
289 AACGACGGCCAGTGAATTG-3') and BAC-R (5'-GATAACAATTTCACACAGG-3')
290 primers from pIndigoBAC536-S vector backbone. The BAC end sequences from Sanger
291 sequencing were aligned to BES from the analysis results of illumina data using Blastn.

292 5.5 BAC Mapping on broomcorn millet longmi4 genome

293 The genome sequence of broomcorn millet longmi4 was downloaded from NCBI
294 genome database under GCA_002895445 accession, which was submitted by researchers
295 from China Agricultural University. The local blastn was employed to map all BESs to the
296 genome with the following options: qcov_hsp_perc=99, perc_identity=99, outfmt=6,
297 culling_limit=1. The results were further converted to a GFF3 format file using a Python
298 script. In this script, following conditions were set: if both forward and reverse BESs in each

299 clone were mapped to the same chromosome, and their orientations were opposite and their
300 interval lengths were less than 250 kb, such clones were recorded in GFF3 format file with
301 three lines; if either forward or reverse BES was uniquely mapped to chromosome, such
302 clones also were recorded in GFF3 format file with two lines; other conditions would be
303 discarded. The GFF3 file was sorted using GFF3sort and presented with JBrowse in our
304 website (http://eightstarsbio.com/gresource/JBrowse-1.16.5/index.html).

305 ## 5.6 Organelle genome analysis

306 The chloroplast genome (CM009689) of brromcorn millet and all mitochondrial
307 genomes (NC_008331, NC_007982, NC_036024, NC_031164, NC_029816, NC_022714,
308 NC_022666, NC_013816, NC_007886, NC_011033, NC_008362, NC_008360, NC_008332,
309 NC_008333) in *Gramineae* were downloaded form NCBI nucleotide database. All potential
310 BACs of organelle genome were identified by the local blastn with following options:
311 qcov_hsp_perc=99, perc_identity=99, outfmt=6.

312 # 6 Abbreviations:

313 BAC: bacterial artificial chromosome;
314 BES: BAC end sequence;
315 CAPSS: clone array pooled shotgun sequencing;
316 CM: chloramphenicol;
317 NGS: next generation sequencing;
318 PCR: polymerase chain reaction;
319 VES: vector end sequence.
320

7 Declarations

7.1 Ethics approval and consent to participate

Not applicable

7.2 Consent for publication

Not applicable

7.3 Availability of data and materials

The source codes are openly available in a GitHub repository (https://github.com/xuweixw/broomcorn-millet-BAC-library). Illumina sequencing data are available at Sequence Read Archive (SRA) under the accession PRJNA576359. BAC clones can be browsed and obtained through our website (http://eightstarsbio.com/gresource/JBrowse-1.16.5/index.html).

7.4 Competing interests

The authors declare that they have no competing interests.

7.5 Funding

This work was supported by a grant from the National Natural Science Foundation of China (Grant no. 31671268).

7.6 Authors' contributions

WX and MLuo conceived and designed the research framework; WX, MLiang, XY and HW performed the experiments; WX analyzed the data and wrote the manuscript; MLuo supervised the work and finalized this manuscript. All authors read and approved the manuscript.

7.7 Acknowledgements

We are grateful to Dr. Mingsheng Chen (the Institute of Genetics and Developmental Biology, Chinese Academy of Sciences) for providing the seeds of broomcorn millet.

7.8 Authors' information (optional)

[1]College of Life Science and Technology, Huazhong Agricultural University, Wuhan 430070, China

References

1. Kalinova J, Moudry J. Content and Quality of Protein in Proso Millet (Panicum miliaceum L .) Varieties. Plant Foods Hum Nutr. 2006;61:45–9.

2. Washburn JD, Schnable JC, Davidse G, Pires JC. Phylogeny and photosynthesis of the grass tribe Paniceae. Am J Bot. 2015;102:1493–505.

3. Baltensperger DD. Progress with proso, pearl and other millets. 2002.

355 4. Shi J, Ma X, Zhang J, Zhou Y, Liu M, Huang L, et al. Chromosome conformation
356 capture resolved near complete genome assembly of broomcorn millet. Nat Commun.
357 2019;10:1–9. doi:10.1038/s41467-018-07876-6.
358 5. Zou C, Li L, Miki D, Li D, Tang Q, Xiao L, et al. The genome of broomcorn millet.
359 Nat Commun. 2019;10:1–12. doi:10.1038/s41467-019-08409-5.
360 6. Luo M, Wing RA. An improved method for plant BAC library construction. In: Plant
361 Functional Genomics. 2003. p. 3–19.
362 7. Shi X, Zeng H, Xue Y, Luo M. A pair of new BAC and BIBAC vectors that facilitate
363 BAC/BIBAC library construction and intact large genomic DNA insert exchange. Plant
364 Methods. 2011;7:33. doi:10.1186/1746-4811-7-33.
365 8. Shizuya H, Birren B, Kim UJ, Mancino V, Slepak T, Tachiiri Y, et al. Cloning and
366 stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia coli using
367 an F-factor-based vector. Proc Natl Acad Sci U S A. 1992;89:8794–7.
368 9. Pan Y, Deng Y, Lin H, Kudrna DA, Wing RA, Li L, et al. Comparative BAC-based
369 physical mapping of Oryza sativa ssp. indica var. 93-11 and evaluation of the two rice
370 reference sequence assemblies. Plant J. 2014;77:795–805.
371 10. Dong G, Shen J, Zhang Q, Wang J, Yu Q, Ming R, et al. Development and
372 Applications of Chromosome-Specific Cytogenetic BAC-FISH Probes in S. spontaneum.
373 Front Plant Sci. 2018;9:218. doi:10.3389/fpls.2018.00218.
374 11. Cai WW, Chen R, Gibbs RA, Bradley A. A clone-array pooled shotgun strategy for
375 sequencing large genomes. Genome Res. 2001;11:1619–23.
376 12. Deng Y, Pan Y, Luo M. Detection and correction of assembly errors of rice
377 Nipponbare reference sequence. Plant Biol. 2014;16:643–50.

378 13. Wei X, Xu Z, Wang G, Hou J, Ma X, Liu H, et al. PBACode: A random-barcode-
379 based high-throughput approach for BAC paired-end sequencing and physical clone mapping.
380 Nucleic Acids Res. 2017;45(7):e52.
381 14. Yang X, Yang Y, Ling J, Guan J, Guo X, Dong D, et al. A high-throughput BAC end
382 analysis protocol (BAC -anchor) for profiling genome assembly and physical mapping. Plant
383 Biotechnol J. 2019;18(2):364–72.
384 15. Zhaozhao D, Tong L, Jiadong L, Zhifei H. High-throughput long paired-end
385 sequencing of a Fosmid library by Pacbio. Plant Methods. 2019;15:142.
386 16. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al.
387 SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell
388 Sequencing. J Comput Biol. 2012;19:455–77.
389 17. Pan Y, Wang X, Liu L, Wang H, Luo M. Whole Genome Mapping with Feature Sets
390 from High-throughput Sequencing Data. PLoS One. 2016;11:1–17.
391 18. Gualberto JM, Newton KJ. Plant Mitochondrial Genomes: Dynamics and
392 Mechanisms of Mutation. Annu Rev Plant Biol. 2017;68:225–52. doi:10.1146/annurev-
393 arplant-043015-112232.
394