

RESEARCH ARTICLE

A bacterial artificial chromosome-based physical map of *Manihot esculenta* ssp. *flabellifolia*

Yuhua FU^{1,2}, Zhiqiang XIA¹, Shujuan WANG¹, Xin CHEN¹, Cheng LU¹, Mingcheng LUO³, Hongbin ZHANG⁴,
Wenquan WANG (✉)¹

¹ Institute of tropical Biosciences & Biotechnology, Chinese Academy of Tropical Agricultural Sciences, Haikou 571101, China

² Guizhou Institute of Subtropical Crops, Guizhou Academy of Agricultural Sciences, Xinyi 562400, China

³ Department of Plant Sciences, UC Davis, Davis, CA 95616, USA

⁴ Department of Soil and Crop Sciences, Texas A&M University, College Station, TX 77843, USA

Abstract Cassava (*Manihot esculenta*) is known as the third most important food crop in the tropics and also used for industrial feedstock for biofuels. Two new bacterial artificial chromosome (BAC) libraries were constructed for W14 (*M. esculenta* ssp. *flabellifolia*), a wild ancestor of domesticated cassava. The libraries were constructed with *EcoRI* and *HindIII* insertion vectors, respectively. The *EcoRI* library has 29952 clones with an average insert size of 115 kb, while the *HindIII* library consists of 29952 clones with an average insert of 129 kb. The combined libraries contain a total of 59904 clones with an average insert size of 125 kb, representing approximately 10 × haploid genome equivalents. A total of 29952 clones were fingerprinted and resulted in a cassava physical map composed of 2485 contigs with an average physical length of 336 kb and 2909 singletons, representing approximately 762 Mb of the cassava genome. 5000 clones located at the ends of BAC contigs were selected and sequenced. A total of 6077 SNPs and 231 indels were identified, that covered 459 gene sequences, of which 6 genes were associated with starch and sucrose metabolism. This BAC-based physical map provides valuable tools to understand the genetics and evolution of cassava.

Keywords cassava, BAC library, physical map

($2n = 36$) with an estimated genome size of approximately 770 Mb^[1], is an important tropical crop belonging to the Euphorbiaceae family. It is grown throughout Africa, Asia and South America, and its starchy roots and edible leaves are consumed by more than 800 million people^[2]. Besides being a staple food, it is also an industrial feedstock for biofuels, animal feed and financial source in some countries^[3–5]. Cassava is considered the third most important source of calories in the tropics^[6].

Despite the importance of cassava, progress has been slow in the areas of both conventional breeding and molecular breeding compared to other major crops, in part this is due to its heterozygous genome, long breeding cycle and low seed production^[7]. Great efforts worldwide have been made to improve cassava properties and many important QTLs have been identified underlying cassava mosaic disease^[8], bacterial blight^[9,10], plant architecture and productivity^[11,12], yield-related traits^[13], cyanogenic potential^[14] and dry matter content^[15]. Current genomic resources for cassava research include several genetic linkage maps constructed with multiple molecular markers^[16–22]; a large EST database^[23–27]; DNA microarrays^[28,29]; a non-redundant set of SSRs and validated SNPs^[26,30]; bacterial artificial chromosome (BAC) libraries^[31,32] and annotated draft cassava genome sequence in Phytozome (<http://www.phytozome.net/cassava>). To date, four cassava BAC libraries have been developed. The first two libraries were constructed to facilitate the cloning of resistance genes by using a cassava mosaic disease (CMD) resistant and cassava bacterial blight (CBB) resistant clone TMS 30001 and a cassava white fly resistant clone MECU72^[31]. The TMS 30001 library has a 5 × haploid genome equivalent with an average insert size of 80 kb and was screened by southern hybridization using a cassava analog (CBB1) of the *Xa21*

1 Introduction

Cassava (*Manihot esculenta*), an allopolyploid species

Received August 5, 2016; accepted December 5, 2016

Correspondence: wangwenquan@itbb.org.cn

from rice, while the MECU72 library has a genome coverage of $11.3 \times$ with an average insert size of 93 kb and was analyzed by end sequencing of 2301 clones in order to scan for genes and repeat DNA contents. The third BAC library with $11 \times$ genome coverage was constructed from TME3, a CMD-resistant genotype, for contig mapping. Two RAPD markers which flanked a resistance gene (CMD2) were used to screen the TME3 library by PCR amplification of BAC pools. A total of 16 positive clones were identified. A recent cassava BAC library was developed from a partially inbred genotype AM560-2 for physical mapping^[32].

A bacterial artificial chromosome physical map is an important genomic resource for basic and applied research in plants. Physical maps of either the whole genome, a specific chromosome or region of interest have proven to be useful tools for comparative genomics, gene cloning and understanding the size and complexity of a genome^[33–37]. Physical maps which are integrated with genetic map are more valuable in map-based cloning, genomic analysis of closely related species and enhance the understanding of unsequenced genomes^[38–40]. For any genome sequencing project, a physical map is a fundamental component of the clone-by-clone whole-genome sequencing method that was the predominant approach used to access crop genomes until 2010 or an important intermediate layer for providing a good template for completing gaps and correcting misassemblies in a hybrid sequencing strategy, especially for large and complex genomes^[35–44]. For the species with a sequenced genome, physical maps are being constructed for wild relatives for comparative purposes^[35,45]. Studies in other plant species have demonstrated an evolution of BAC physical map construction and the high-information-content fingerprinting method (HICF) and five-enzyme digestion plus SNaPshot labeling have proven to be the most effective^[35,46–48]. One BAC-based physical map in cassava was constructed from AM560-2 by Pablo Rabinowicz and Mingcheng Luo (unpublished data), which has been used in improving the genome sequence assembly (Cassava Genome Database, <http://cassava.igs.umaryland.edu/cgi-bin/index.cgi>).

Phylogeography shows cultivated cassava originated from wild progenitors from the southern rim of the Amazon basin. Chromosome variation studies show *M. esculenta* ssp. *flabellifolia* (W14) is probably the only ancestor of cultivated cassava. Scientists carried out cassava genetic breeding using W14 and cultivars such as SC 205, SC 8 and Arg 7 as materials to explore the mechanism of the high photosynthesis efficiency of the cultivated cassava through physiology and biochemistry, transcriptomics and proteomics. In this study, two additional cassava BAC libraries were constructed from W14 using different restriction enzymes. In addition, a physical map was built using the HICF method and single

nucleotide polymorphisms (SNPs) were developed from the BAC-end sequence.

2 Materials and methods

2.1 Plant materials

A wild ancestral cassava species, *M. esculenta* ssp. *flabellifolia* accession W14, was used to construct the two BAC libraries. W14 has low storage root yield with an average of 0.8 kg per plant, low starch content and high disease resistance. Seeds were planted in a greenhouse and young fully-developed leaves were collected three weeks after planting and stored at -80°C before use.

2.2 BAC libraries construction

Megabase-sized nuclear DNA was prepared according to Zhang's protocol^[49]. Approximately 60.0 g of frozen young leaves was used for the extraction of nuclear DNA; nuclei were washed and centrifuged several times in order to minimize plastid DNA contamination. Nuclei were resuspended in 2 mL of $1 \times$ homogenization buffer (nuclei isolation buffer), mixed with an equal volume of 1% low-melting-point agarose and dispensed into ice-cold 100- μL plug molds. The solidified nuclei plugs were incubated in lysis buffer (0.5 mol·L⁻¹ EDTA, pH 9.0–9.3, 1% sodium lauryl sarcosine, 0.3 mg·mL⁻¹ proteinase K) with gentle shaking at 50°C overnight. The plugs were washed three times in ice-cold $1 \times$ TE plus 0.1 mmol·L⁻¹ phenylmethylsulfonyl fluoride (PMSF) on ice, 1 h each wash, followed by three washes on ice in $1 \times$ TE without PMSF, 1 h each wash. The plugs were stored in $1 \times$ TE at 4°C before use.

Size selection after partial digestion and BAC cloning were performed according to the protocol developed by Tao et al.^[50] with minor modifications. Two plugs were first used to determine the optimal partial digestion condition. They were cut into equal small pieces (nine slices for each plug), equilibrated on ice in the enzyme reaction buffer plus 2 mmol·L⁻¹ spermidine and 1 mmol·L⁻¹ DTT twice for 30 min each time, and then another 60 min in the enzyme reaction buffer plus 2 mmol·L⁻¹ spermidine, 1 mmol·L⁻¹ DTT and 0.5 mg·mL⁻¹ BSA with a range of *EcoRI* or *HindIII*. The reactions were incubated at 37°C for 7 min and stopped by transferring onto ice and adding 1/10 volume of 0.5 mol·L⁻¹ EDTA. Partially digested DNA was fractionated on a 1% agarose gel in $0.5 \times$ TBE using the CHEF DRIII System (Bio-Rad Laboratories, Hercules, CA, USA) and the enzyme conditions producing the largest number of fragments ranging from 100 to 350 kb were selected. Six DNA plugs were used to construct the library and partially digested using *EcoRI* or *HindIII* as described as above. The

fragments were size-selected twice by pulsed-field electrophoresis (PFGE) and ligated into either the CopyControl pCC1BAC (*Hind*III-cloning ready) vector or the CopyControl pCC1BAC (*Eco*RI-cloning ready) vector (Epicenter Technology, Madison, WI, USA). The ligation mixture was transformed into *Escherichia coli* strain DH10B electrocompetent cells by electroporation. The cells were cultured in 0.5 mL SOC medium (Invitrogen, Carlsbad, CA, USA) and plated on LB agar medium containing IPTG, X-GAL and chloramphenicol. White colonies were arrayed in 384-well microplates containing freezing medium with appropriate chloramphenicol, incubated at 37°C overnight and stored at – 80°C.

2.3 Analysis of BAC clones

Random clones from the two BAC libraries were selected to estimate the insert sizes. BAC plasmids were isolated and digested with *Not*I to release the inserts from the vector. The insert size was analyzed by PFGE and the genome coverage of the libraries was estimated.

2.4 BAC fingerprinting

In total, 29952 clones were randomly selected from the *Hind*III and *Eco*RI BAC libraries and then fingerprinted as described by Luo et al.^[47] with minor modifications. From each 384-well plate from the two libraries, samples were inoculated into 96-deep well plates containing 1 mL TB (Terrific Broth) medium with 12.5 g·mL⁻¹ chloramphenicol using a 96-pin replicator (Boeckel Scientific, Feasterville, PA, USA), such that the clones from one 384-well plate were inoculated into four 96-deep well plates. Each of the 96-deepwell plates were covered with air-permeable seals (Excel Scientific, Wrightwood, CA, USA) and incubated in an orbital shaker agitated at 350 r·min⁻¹ at 37°C for 22 h. Cells were harvested by centrifugation at 2500 r·min⁻¹ for 10 min and BAC DNA was isolated using a modified alkaline lysis method^[53]. Approximately 0.5–1.2 µg BAC DNA was simultaneously digested using *Bam*HI, *Eco*RI, *Nde*I, *Xba*I and *Hae*III enzymes (New England Biolabs, Ipswich, MA, USA) for 3 h at 37°C and then the restricted products were labeled using the SNaPshot Multiplex Kit (Applied Biosystems, Foster City, CA, United States). Fragmented and labeled DNA was precipitated by adding chilled 95% ethanol, washed, air-dried and dissolved in a mixture of 9.9 µL of Hi-Di formamide and 0.3 µL of internal size marker LIZ 1200 (Applied Biosystems). The DNA was denatured at 95°C for 5 min and loaded on the ABI3730xl DNA Analyzer (Applied Biosystems) using the default module (50 cm array and POP7).

2.5 Fingerprints editing and contig assembly

The fragment size from the fingerprint profile of each BAC

clone was collected by the GeneMapper v3.7 (Applied Biosystems) and output data of size-calling files was edited with the FPMIner program (<http://www.bioinformsoft.com>). This software package was further used to distinguish peaks corresponding to true fragments from those generated by background noise and remove vector bands and substandard restriction fragments. The edited profiles were exported as sizes files. Fragments in the size range 75 to 1000 bp were used to perform contig assembly using the program FPC V 8.9 (<http://www.agcol.arizona.edu/software/fpc>) with adjusted parameters for the HICF method^[47,48,54].

2.6 BAC-end sequencing

The BAC DNAs were isolated using a modified alkaline lysis method. BAC-end sequencing was carried out using an ABI Big Dye Terminator v3.1 and analyzed on the ABI 3730xl Analyzer. The sequences were processed and trimmed by the programs Phred and Lucy^[55,56].

2.7 SNP mining

All the BAC-end sequences were mapped to a reference genome using software Bowtie 2^[57]. The SAM tools^[58] pileup command was used to determine the variable positions of single-nucleotide polymorphisms (SNPs). The annotation was performed by BLAST, with blastp parameter as 1×10^{-5} to search the best match of peptide sequence of genes from KEGG (<http://www.genome.jp/kegg>) and NCBI nrviridiplantae. The functional annotation was based on BLASTX results with the NCBI nrviridiplantae database. The program BLAST2GO^[59] was used to assign biological functions, cellular components and cellular processes to the transcripts.

3 Results

3.1 BAC library construction and characterization

Two BAC libraries were successfully constructed from the nuclear DNA of a wild ancestral cassava species, W14, in the CopyControl pCC1BAC vectors. The HMW DNA was partially digested either by *Hind*III or by *Eco*RI and restricted fragments generated from these two digestions were subjected to two-size selections in order to select big restricted fragments for the library construction. Each of the two libraries, *Eco*RI library and *Hind*III library, contained 29952 clones arrayed in 78384-well microplates, respectively. To estimate the insert size of the *Eco*RI library, a random sample of 96 clones was analyzed by *Not*I digestion and pulsed-field gel electrophoresis (Fig. 1). The results showed *Eco*RI library has an average insert size of 115 kb in a range of 9 to 183 kb. 75% of the clones contained inserts of greater than 100 kb with a majority



Fig. 1 A random sample of bacterial artificial chromosomes (BAC) selected from the *EcoRI* library and analyzed on a pulsed-field gel. BAC DNA was digested with *NorI* to release the entire insert from the BAC vector. M indicates lambda ladder pulsed-field gel electrophoresis marker.

having a range from 101 to 150 kb. 5 out of 96 clones (approximately 5%) were found to be insert-empty clones. From *HindIII* library, a sample of 96 clones was also selected randomly and analyzed by PFGE in order to evaluate the insert size. Approximately 97% of the clones contained DNA inserts with an average size of 129 kb, ranging from 8 to 286 kb. More than 71% of the clones had inserts larger than 125 kb with a majority having a range from 126 to 150 kb (Fig. 2).

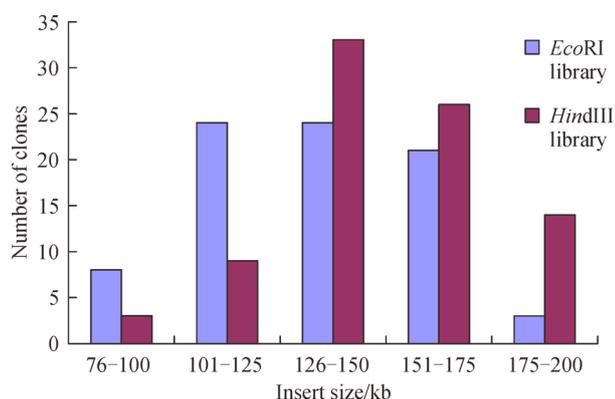


Fig. 2 Insert size distribution of the bacterial artificial chromosome libraries with 80 clones from the *EcoRI* library and 85 clones from *HindIII* library

Therefore, these two libraries consist of a total of 59904 clones with an average insert size of 125 kb. Based on the estimated size of the haploid cassava genome of 770 Mb, the combined two libraries have a genome coverage equivalent to approximately $10 \times$ haploid genomes.

3.2 BAC fingerprinting and contig assembly

A total of 29952 clones from the two libraries representing approximately $5.0 \times$ genome equivalents were finger-

printed using the HICF method of Luo et al.^[47]. Of these clones, 5168 clones were removed during fingerprint editing because they were small insert clones, contaminated clones, empty clones or clones with poor quality fingerprints. Therefore, a total of 24784 clones were successfully fingerprinted and subsequently subjected to contig assembly using the FPC software.

To determine the optimal cutoff and minimize the contig number, a tolerance of 0.5 bp and an initial cutoff of $1E-45$ were set, followed by DQer and several rounds of end-to-end and single-to-end merging at progressively lower stringencies. In the end, a total of 2485 contigs were assembled from the validated 24784 BAC fingerprints whereas 2909 clones remained as singletons (Table 1). Each contig contained an average of nine BACs and spanned an average length of 336 kb. The contiguous bands across all contigs were equivalent to 762 Mb (98.96% of the cassava genome).

Table 1 Statistics of the bacterial artificial chromosome (BAC)-based physical map of the wild cassava genome

Description	Index
Number of BAC clones fingerprinted	29952
Number of high quality fingerprints used for assembly	24784
Number of contigs	2485
Number of singletons	2909
N50 contig length/kb	336
Average number of clones per contig	9
Longest contig/kb	1867
Total physical length of the contig/Mb	762

3.3 BAC-end sequencing and integrated into draft assembly sequences

A total of 5000 BAC clones located at the ends of BAC contigs were randomly selected for end sequencing and 8361 BAC-end sequences (BESs) were generated. The average read length was 699 bp with minimum length of 19 bp, representing 5.8 Mb of genomic DNA in total (Table 2). The GC content was estimated to be 38.54%.

The BESs were matched to the whole-genome sequence of W14 (sequenced using a next generation sequencing platform Illumina HiSeq2000 and Roche/454 GS FLX) in order to improve the sequence assembly and create superscaffolds. This analysis allowed the original 33166 contigs from the primary sequence assembly to be merged into 7393 scaffolds and a total of 357 mega scaffolds in the physical map were generated with the average scaffold length of 183 kb (Table 3). The N50 was 202 kb in the physical map assembly and 33 kb in the scaffolds of W14. This alignment made a total of 65 Mb of the physical map successfully integrated with the sequence assembly of W14.

Table 2 The index statistics of bacterial artificial chromosome-end sequence and assembly

Category	Span/bp	Numbers
Total	5831315	10000
Mean	699.0308	6703
Median	701	5978
Max	1231	1
Min	19	8335
N50	701	3700
GC/%	38.54	

Table 3 Integrated draft genome of W14 based on the physic map and 5000 bacterial artificial chromosome-end sequences

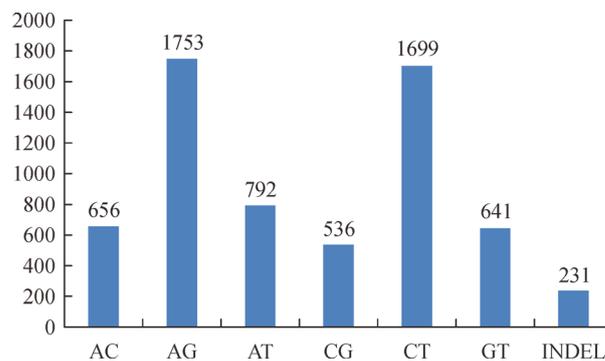
	W14 all contigs /scaffolds	Physic map assembly /mega scaffolds
Fold genome coverage	136	8361 BES
Total number of contigs/scaffolds	33166	359
Total span/Mb	426	65
N50/kb	33	202
Number of scaffolds	7393	357
Largest scaffold/kb	277	431
Average scaffold length/kb	35	183
Scaffolds N50/kb	51	203
GC/%	35.98	34.30

3.4 Mining SNP involved in functional genes

A total of 6077 SNPs and 231 indels were discovered through alignment of BESs and whole-genome sequence assembly of AM560. Among the 6077 SNPs, the most abundant SNPs were A/G and C/T, accounting for 28.85% and 27.96% of the total SNPs, respectively (Fig. 3). In addition, 5172 single-nucleotide variations (SNVs) fell into 459 gene sequences, of which 2317 SNVs positioned in the coding sequence of 399 genes and 167 SNVs were in the untranslated region of 50 genes. Gene ontology analysis of the 459 genes showed these SNPs and indels to be related to pathways in cellular component, molecular function and biological process (Fig. 4). Further KEGG analysis showed a total of six genes were associated with starch and sucrose metabolism including sucrose synthase, trehalose 6-phosphate synthase, pectinesterase, trehalose 6-phosphate phosphatase, α -glucosidase and α -1,4-galacturonosyltransferase.

4 Discussion

Two new large-insert libraries were constructed for cassava with different restriction endonucleases. The combined

**Fig. 3** Statistics of single-nucleotide polymorphisms and indels developed from bacterial artificial chromosome-end sequences

libraries contain a total of 59904 clones with an average insert size of 125 kb and approximate $10 \times$ genome coverage. In comparison to the previously described cassava BAC libraries, these two libraries were constructed from a wild ancestral cassava subspecies and have relatively large insert size with two different cloning sites. The strategy of constructing libraries with multiple enzymes with a different recognition sequence is to limit the risk of under-representation of genome regions of interest due to restriction enzyme bias, and this strategy has been applied in several physical mapping projects^[60–62]. Therefore, the two cassava libraries reported here are valuable complements for cassava genomic resources, especially in physical map construction and studying genes that were lost during domestication.

A genome-wide physical map of cassava, consisting of 2485 contigs and 2909 singletons with an estimated physical length of 762 Mb, was built based on two BAC libraries using the HICF method. The quality of a BAC-based map was determined by several key factors such as source BAC library coverage, the density of bands per fingerprint and the critical parameters (tolerance and cutoff value). It has been reported that some singletons can be produced not only because of low quality fingerprints but also because of low coverage in the source BAC library^[61], and that the contig length increases as the genomic coverage of source clones increases^[46]. In the present study, although the source of clones for the cassava physical map was equivalent to $5 \times$ haploid genomes, the fact that they were from two BAC libraries of different origin ensured higher genomic representation in the physical map and compensated for the low coverage of the source clones. In addition, a high stringency setting was used to build the contig map in order to minimize the false positive and obtain high accuracy of contig assemblies. However, the high stringency parameters employed in physical mapping are likely lead to over-estimation of genome coverage due to failures in detecting overlaps between adjacent contigs, as reported for other plant physical maps such as poplar^[63], soybean^[60] and

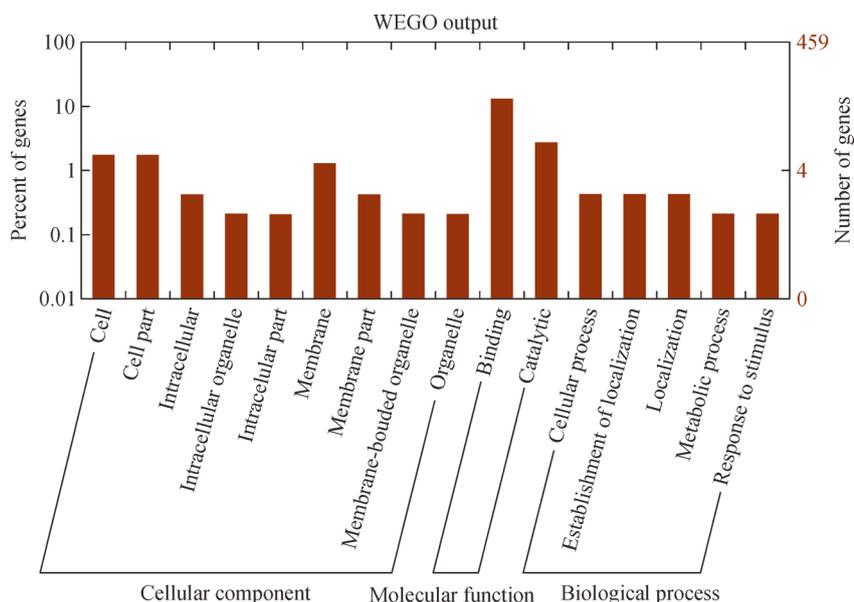


Fig. 4 Gene ontology analysis of 459 genes involving single-nucleotide polymorphisms and indels

apple^[64]. Therefore, the estimate of the genome coverage of the cassava physical map reported here may be greater than the actual percentage. Furthermore, the reliability of this cassava map was assessed by aligning the BAC-end sequences of the BAC clones with the whole-genome sequence assembly of W14. This BAC-based physical map will be of great help in improving the quality of the W14 genome sequence, cloning potentially valuable genes and comparative evolutionary studies.

Sequencing paired ends of BAC clones to discover molecular markers, mainly SSRs and SNPs, has been successfully applied in several plant species, and BES-based markers have proven to be useful tools for integration of physical and genetic maps^[65–68]. Alignment BES to a reference genome, not only allow researchers to develop SNPs in a cost-effective way but also to make use of the genome annotation to predict SNPs, both in coding and non-coding regions. A total of 5172 SNVs discovered in this research are within 459 gene sequences of which six genes are associated with starch and sucrose metabolism. Given that starch content is one of the most important traits of cassava, the SNPs discovered in these six genes are valuable marker resources for better understanding of starch and sucrose metabolism in cassava and potentially for improving the yield and quality. However, further study of the association between these SNPs and heritable phenotypes is needed before they can be applied for marker-assisted selection.

5 Conclusions

The present study reported the construction of a genome-wide BAC contig map of W14, an accession of wild

M. esculenta spp. *flabellifolia*. Two BAC libraries were constructed with a total of 59904 clones with an average insert size of 125 kb. These libraries are valuable complementary resources for cassava genomic studies providing a useful platform for multiple genetic researches. A physical map was constructed using 24784 BAC clones with 8361 BAC-end sequences. This map is suitable for comparative genomics study among species of *M. esculenta* and positional cloning of disease resistance genes from wild species. We hope this study could accelerate the genetic and genomic researches on this important tropical crop. The BAC library and physical map are shared through our database at <http://www.cassava-genome.cn>.

Acknowledgements This project was funded by the National Basic Research and Development Program (2010CB126600), China Agriculture Research System (CARS-12), and National International Science and Technology Cooperation Plan (2011DFB31690). Authors are grateful to bioinformatics researchers from the Cassava Genome Consortium for their help in integration of the BAC-based physical map with the assembled genome.

Compliance with ethics guidelines Yuhua Fu, Zhiqiang Xia, Shujuan Wang, Xin Chen, Cheng Lu, Mingcheng Luo, Hongbin Zhang, and Wenquan Wang declare that they have no conflict of interest or financial conflicts to disclose.

This article does not contain any studies with human or animal subjects performed by any of the authors.

References

1. Awolaye F, van Duren M, Dolezel J, Novak F J. Nuclear DNA content and *in vitro* induced somatic polyploidization cassava (*Manihot esculenta* Crantz) breeding. *Euphytica*, 1994, **76**(3): 195–

202

2. Mohammed B B, Isah A B, Ibrahim M A. Influence of compaction pressures on modified cassava starch as a binder in paracetamol tablet formulations. *Nigerian Journal of Pharmaceutical Sciences*, 2009, **8**(1): 80–88
3. Dixon A G O, Bandyopadhyay R, Coyne D, Ferguson M, Ferris R S B, Hanna R, Hughes J, Ugent I I, Legg J, Mahungu N, Manyong V, Mowbray D, Neuenschwander P, Whyte J, Hartmann P, Ortiz R. Cassava: from poor farmer's crop to pacesetter of African rural development. *Chronica Horticulturae*, 2003, **43**(4): 8–15
4. Tonukari N J. Cassava and the future of starch. *Electronic Journal of Biotechnology*, 2004, **7**(1): 5–8
5. Rettinassababady C, Ramadoss N, Thirumeni S. Present situation and future potential of cassava in Thailand. *Ifad Stakeholder Consultation Meeting*, 2007, **19**(1): 392–415
6. Fauquet C M, Taylor N. The potential for biotechnology to improve the nutritional value of cassava. *Food and Nutrition Bulletin*, 2002, **23**(4): 364–366
7. Ceballos H, Iglesias C A, Perez J C, Dixon A G O. Cassava breeding: opportunities and challenges. *Plant Molecular Biology*, 2004, **56**(4): 503–516
8. Akano A O, Dixon A G O, Mba C, Barrera E, Fregene M. Genetic mapping of a dominant gene conferring resistance to cassava mosaic disease. *Theoretical and Applied Genetics*, 2002, **105**(4): 521–525
9. López C E, Quesada-Ocampo L M, Bohórquez A, Duque M C, Vargas J, Tohme J, Verdier V. Mapping EST-derived SSRs and ESTs involved in resistance to bacterial blight in *Manihot esculenta*. *Genome*, 2007, **50**(12): 1078–1088
10. Wydra K, Zinsou V, Jorge V, Verdier V. Identification of pathotypes of *Xanthomonas axonopodis* pv. *manihotis* in Africa and detection of quantitative trait loci and markers for resistance to bacterial blight of cassava. *Phytopathology*, 2004, **94**(10): 1084–1093
11. Okogbenin E, Marin J, Fregene M. QTL analysis for early yield in a pseudo F₂ population of cassava. *African Journal of Biotechnology*, 2008, **7**(2): 131–138
12. Okogbenin E, Fregene M. Genetic mapping of QTLs affecting productivity and plant architecture in a full-sib cross from non-inbred parents in Cassava (*Manihot esculenta* Crantz). *Theoretical and Applied Genetics*, 2003, **107**(8): 1452–1462
13. Chen X, Fu Y, Xia Z, Jie L, Wang H, Lu C, Wang W. Analysis of QTL for yield-related traits in cassava using an F₁ population from non-inbred parents. *Euphytica*, 2012, **187**(2): 227–234
14. Whankaew S, Poopear S, Kanjanawattanawong S, Tangphatsornruang S, Boonseng O, Lightfoot D A, Triwitayakorn K. A genome scan for quantitative trait loci affecting cyanogenic potential of cassava root in an outbred population. *BMC Genomics*, 2011, **12**(1): 266
15. Kizito E B, RÖnnberg-Wästljung A C, Egwang T G, Gullberg U, Fregene M, Westerbergh A. Quantitative trait loci controlling cyanogenic glucoside and dry matter content in cassava (*Manihot esculenta* Crantz) roots. *Hereditas*, 2007, **144**(4): 129–136
16. Chavarriaga-Aguirre P, Maya M M, Bonierbale M W, Kresovich S, Fregene M A, Tohme J, Kochert G. Microsatellites in cassava (*Manihot esculenta* Crantz): discovery, inheritance and variability. *Theoretical and Applied Genetics*, 1998, **97**(3): 493–501
17. Fregene M, Angel F, Gomez R, Rodriguez F, Chavarriaga P, Roca W M, Tohme J, Bonierbale M. A molecular genetic map of cassava (*Manihot esculenta* Crantz). *Theoretical and Applied Genetics*, 1997, **95**(3): 431–441
18. Okogbenin E, Marin J, Fregene M A. An SSR-based molecular genetic map of cassava. *Euphytica*, 2006, **147**(3): 433–440
19. Chen X, Xia Z, Fu Y, Lu C, Wang W. Constructing a genetic linkage map using an F₁ population of non-inbred parents in cassava (*Manihot esculenta* Crantz). *Plant Molecular Biology Reporter*, 2010, **28**(4): 676–683
20. Kunkeaw S, Tangphatsornruang S, Smith D R, Triwitayakorn K. Genetic linkage map of cassava (*Manihot esculenta* Crantz) based on AFLP and SSR markers. *Plant Breeding*, 2010, **129**(1): 112–115
21. Mba R E C, Stephensen P, Edwards K, Melzer S, Nkumbira J, Gullberg U, Apel K, Gale M, Tohme J, Fregene M. Simple sequence repeat (SSR) markers survey of the cassava (*Manihot esculenta* Crantz) genome: towards and SSR-based molecular genetic map of cassava. *Theoretical and Applied Genetics*, 2001, **102**(1): 21–31
22. Sraphet S, Boonchanawiwat A, Thanyasiriwat T, Boonseng O, Tabata S, Sasamoto S, Shirasawa K, Isobe S, Lightfoot D A, Tangphatsornruang S, Triwitayakorn K. SSR and EST-SSR-based genetic linkage map of cassava (*Manihot esculenta* Crantz). *Theoretical and Applied Genetics*, 2011, **122**(6): 1161–1170
23. Lopez C, Jorge V, Piégu B, Mba C, Cortes D, Restrepo S, Soto M, Laudie M, Berger C, Cooke R, Delseny M, Tohme J, Verdier V. A unigene catalogue of 5700 expressed genes in cassava. *Plant Molecular Biology*, 2004, **56**(4): 541–554
24. Lokko Y, Anderson J V, Rudd S, Raji A, Horvath D P, Mikel M A, Kim R, Liu L, Hernandez A G, Dixon A G O, Ingelbrecht I. Characterization of an 18,166 EST dataset for cassava (*Manihot esculenta* Crantz) enriched for drought-responsive genes. *Plant Cell Reports*, 2007, **26**(9): 1605–1618
25. Sakurai T, Plata G, Rodrigue-Zapata F, Seki M, Salcedo A, Toyoda A, Ishiwata A, Tohme J, Sakaki Y, Shinozaki K, Ishitani M. Sequencing analysis of 20,000 full-length cDNA clones from cassava reveals lineage specific expansions in gene families related to stress response. *BMC Plant Biology*, 2007, **7**(1): 66–72
26. Ferguson M, Hearne S J, Close T J, Wanamaker C S, Moskal W A, Town C D, Young J D, Marri P R, Rabbi I Y, Villiers E P D. Identification, validation and high-throughput genotyping of transcribed gene SNPs in cassava. *Theoretical and Applied Genetics*, 2012a, **124**(4): 685–695
27. Utsumi Y, Sakurai T, Umemura Y, Ayling S, Ishitani M, Narangajavana J, Sojikul P, Triwitayakorn K, Matsui M, Manabe R, Shinozaki K, Seki M. RIKEN cassava initiative: establishment of a cassava functional genomics platform. *Tropical Plant Biology*, 2012, **5**(1): 110–116
28. Lopez C, Soto M, Restrepo S, Piegu B, Cooke R, Delseny M, Tohme J, Verdier V. Gene expression profile in response to *Xanthomonas axonopodis* pv. *Manihotis* infection in cassava using a cDNA microarray. *Plant Molecular Biology*, 2005, **57**(3): 393–410
29. Yang J, An D, Zhang P. Expression profiling of cassava storage roots reveals an active process of glycolysis/gluconeogenesis. *Journal of Integrative Plant Biology*, 2011, **53**(3): 193–211
30. Ferguson M, Rabbi I Y, Kim D, Gedil M, Lopez-Lavalle L A B, Okogbenin E. Molecular markers and their application to cassava

- breeding: past, present and future. *Tropical Plant Biology*, 2012b, **5** (1): 95–109
31. Tomkins J P, Fregene M, Main D, Kim H, Wing R A, Tohme J. Bacterial artificial chromosome (BAC) library resource for positional cloning of pest and disease resistance genes in cassava (*Manihot esculenta* Crantz). *Plant Molecular Biology*, 2004, **56**(4): 555–561
 32. Okogbenin E, Setter T L, Ferguson M, Mutegi R, Ceballos H, Olanami B, Fregene M. Phenotypic approaches to drought in cassava. *Frontiers in Physiology*, 2013, **4**: 93
 33. Adam-Blondon A F, Bernole A, Faes G, Lamoureux D, Pateyron S, Grando M S, Caboche M, Velasco R, Chalhoub B. Construction and characterization of BAC libraries from major grapevine cultivars. *Theoretical and Applied Genetics*, 2005, **110**(8): 1363–1371
 34. Yim Y, Davis G L, Duru N A, Musket T A, Linton E W, Messing J W, McMullen M D, Soderlund C A, Polacco M L, Gardiner J M, Coe E H. 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 Physiology*, 2002, **130**(4): 1686–1696
 35. Meyers B C, Scalabrini S, Morgante M. Mapping and sequencing complex genomes: let's get physical! *Nature Reviews Genetics*, 2004, **5**(8): 578–589
 36. Lin L, Pierce G J, Bowers J E, Estill J C, Compton R O, Rainville L K, Kim C, Lemke C, Rong J, Tang H, Wang X, Braidotti M, Chen A H, Chicola K, Collura K, Epps E, Golser W, Grover C E, Ingles J, Karunakaran S, Kudrna D, Olive J, Tabassum N, Um E, Wissotski M, Yu Y, Zuccolo A, Rahman M, Peterson D G, Wing R A, Wendel J F, Paterson A H. A draft physical map of a D-genome cotton species (*Gossypium raimondii*). *BMC Genomics*, 2010, **11**(1): 395
 37. Schulte D, Ariyadasa R, Shi B, Fleury D, Saski C, Atkins M, Dejong P, Wu C, Graner A, Langridge P, Stein N. BAC library resources for map-based cloning and physical map construction in barley (*Hordeum vulgare* L.). *BMC Genomics*, 2011, **12**(1): 247
 38. Shinozuka H, Cogan N O I, Smith K F, Spangenberg G C, Forster J W. Fine-scale comparative genetic and physical mapping supports map-based cloning strategies for the self-incompatibility loci of perennial ryegrass (*Lolium perenne* L.). *Plant Molecular Biology*, 2010, **72**(3): 343–355
 39. Kucuktas H, Wang S, Li P, He C, Xu P, Sha Z, Liu H, Jiang Y, Baoprasertkul P, Somridhivej B, Wang Y, Abernathy J, Guo X, Liu L, Muir W M, Liu Z. Construction of genetic linkage maps and comparative genome analysis of catfish using gene-associated markers. *Genetics*, 2009, **181**(4): 1649–16660
 40. Philippe R, Paux E, Bertin I, Sourdille P, Choulet F, Laugier C, Simkova H, Safar J, Bellec A, Vautrin S, Frenkel Z, Cattonaro F, Magni F, Scalabrini S, Martis M M, Mayer K F X, Korol A B, Berges H, Doležel J, Feuillet C. A high density physical map of chromosome 1BL supports evolutionary studies, map-based cloning and sequencing in wheat. *Genome Biology*, 2013, **14**(6): 1–22
 41. Tuskan G A, Difazio S P, Jansson S, Bohlmann J, Grigoriev I V, Hellsten U, Putnam N H, Ralph S, Rombauts S, Salsmov A, Schein J E, Sterck L, Aerts A, Bhalariao R R, Bhalariao R P, Blaudez D, Boerjan W, Brun A, Brunner A M, Busov V, Campbell M M, Carlson J E, Chalot M, Chapman J, Chen G, Cooper D M L, Coutinho P M, Couturier J, Covert S F, Cronk Q C B, Cunningham R, Davis J M, Degroove S, Dejardin A, Depamphilis C W, Detter J C, Dirks B, Dubchak I, Duplessis S, Ehling J, Ellis B E, Gendler K, Goodsterin D, Gribskov M, Grimwood J, Groover A, Gunter L E, Hamberger B, Heinze B, Helariutta Y, Henrissat B, Holligan D, Holt R A, Huang W, Islamfaridi N, Jones S, Jonesrhoades J K, Jorgensen R, Joshi C P, Kangasjarvi J, Karlsson J, Kelleher C T, Kirkpatrick R, Kirst M, Kohler A, Kalluri U C, Larimer F W, Leebensmack J, Leple J C, Locascio P, Lou Y, Lucas S, Martin F, Montanini B, Napoli C, Nelson D R, Nelson C D, Nieminen K, Nilsson O, Pereda V, Peter G, Philippe R, Pilate G, Poliakov A, Razumovskaya J, Richardson P M, Rinaldi C, Ritland K, Rouze P, Ryaboy D, Schmutz J, Schrader J, Segerman B, Shin H, Siddiqui A, Sterky F, Terry A, Tsai C, Uberbacher E, Unneberg P, Vahala J, Wall K, Wessler S R, Yang G, Yin T, Douglas C J, Marra M A, Sandberg G, Peer Y V D, Rokhsar D S. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science*, 2006, **313**(5793): 1596–1604
 42. Schnable P S, Ware D, Fulton R S, Stein J C, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves T, Minx P J, Reily A D, Courtney L, Kruchowski S S, Tomlinson C, Strong C, Delehaunty K D, Fronick C, Courtney B, Rock S M, Belter E A, Du F, Kim K, Abbott R, Cotton M, Levy A, Marchetto P, Ochoa K, Jackson S M, Gillam B, Chen W, Yan L, Higginbotham J, Cardenas M, Waligorski J, Applebaum E, Phelps L, Falcone J, Kanchi K L, Thane T, Scimone A, Thane N, Henke J, Wang T, Ruppert J, Shah N, Rotter K, Hodges J, Ingenthron E, Cordes M, Kohlberg S, Sgro J, Delgado B, Mead K, Chinwalla A T, Leonard S, Crouse K, Collura K, Kudrna D, Currie J, He R, Angelova A, Rajasekar S, Mueller T, Lomeli R, Scara G, Ko A, Delaney K, Wissotski M, Lopez G, Campos D, Braidotti M, Ashley E, Golser W, Kim H, Lee S, Lin J, Dujmic Z, Kim W, Talag J, Zuccolo A, Fan C, Sebastian A, Kramer M, Spiegel L, Nascimento L, Zutavern T, Miller B, Ambrose C, Muller S M, Spooner W, Narechania A, Ren L, Wei S, Kumari S, Faga B, Levy M J, McMahan L, Buren P V, Vaughn M W, Kai Y, Yeh C, Emrich S J, Jia Y, Kalyanaraman A, Hsia A, Barbazuk W B, Baucom R S, Brutnell T P, Carpita N C, Chaparro C, Chia J, Deragon J, Estill J C, Fu Y, Jeddelloh J A, Han Y, Lee H, Li P, Lisch D, Liu S, Liu Z, Nagel D H, Mccann C M, Sanmiguel P, Myers A M, Nettleton D, Nguyen J, Penning B W, Ponnala L, Schneider K L, Schwartz D C, Sharma A, Soderlund C, Springer N M, Sun Q, Wang H, Waterman M S, Westerman R, Wolfgruber T K, Yang L, Yu Y, Zhang L, Zhou S, Zhu Q, Bennetzen J L, Dawe R K, Jiang J, Jiang N, Presting G G, Wessler S R, Aluru S, Martienssen R A, Clifton S W, Mccombie W R, Wing R A, Wilson R K. The B73 maize genome: complexity, diversity, and dynamics. *Science*, 2009, **326**(5956): 1112–1115
 43. Bevan M W, Uauy C. Genomics reveals new landscapes for crop improvement. *Genome Biology*, 2013, **14**(6): 206
 44. Ariyadasa R, Stein N. Advances in BAC-based physical mapping and map integration strategies in plants. *Journal of Biomedicine & Biotechnology*, 2012, **2012**: 1–11
 45. Ha J, Abernathy B, Nelson W, Grant D, Wu X, Nguyen H T, Stacey G, Yu Y, Wing R A, Shoemaker R C, Jackson S A. Integration of the draft sequence and physical map as a framework for genomic research in soybean (*Glycine max* (L.) Merr.) and wild soybean (*Glycine soja* Sieb. and Zucc.). *G3: Genes, Genomes, Genetics*,

- 2012, 2(3): 321–329
46. Xu Z, Sun S, Covaleda L, Ding K, Zhang A, Wu C, Scheuring C F, Zhang H. Genome physical mapping with large-insert bacterial clones by fingerprint analysis: methodologies, source clone genome coverage, and contig map quality. *Genomics*, 2004, **84**(6): 941–951
47. Luo M, Thomas C, You F M, Hsiao J, Ouyang S, Buell C R, Malandro M E, Mcguire P E, Anderson O D, Dvorak J. High-throughput fingerprinting of bacterial artificial chromosomes using the SNaPshot labeling kit and sizing of restrictin fragments by capillary electrophoresis. *Genomics*, 2003, **82**(3): 378–389
48. Nelson W M, Bharti A K, Butler E, Wei F, Fuks G, Kim H, Wing R A, Messing J, Soderlund C. Whole-genome validation of high-information-content fingerprinting. *Plant Physiology*, 2005, **139**(1): 27–38
49. Zhang H, Zhao X, Ding X, Paterson A H, Wing R A. Preparation of megabase-size DNA from plant nuclei. *Plant Journal*, 1995, **7**(1): 175–184
50. Wu C, Xu Z, Zhang H. DNA libraries. In: Encyclopedia of Molecular Cell Biology and Molecular Medicine. Meyers RA (ed.). Vol. 3 (2nd Edition). Weinheim: *Wiley-VCH Verlag GmbH*, 2004, 385–425
51. Ren C, Xu Z, Sun S, Lee M, Wu C, Scheuring C F, Zhang H. Genomic DNA Libraries and Physical Mapping. In: The Handbook of Plant Genome Mapping: Genetic and Physical Mapping. Meksem K and Kahl G (eds.). Weinheim: *Wiley-VCH Verlag GmbH*, 2005, 173–213
52. Zhang M, Li Y, Zhang H. Isolation of megabase-sized DNA fragments from plants. In: Handbook of Nucleic Acid Purification. D. Liu (ed.). Florida: *Taylor & Francis Group, LLC*, 2008, 513–524
53. Sambrook J, Russell D W. Molecular cloning: a laboratory manual (3rd ed.). New York: *Cold Spring Harbor Laboratory Press*, 2001, 1–68
54. Quiniou S M, Waldbieser G C, Duke M V. A first generation BAC-based physical map of the channel catfish genome. *BMC Genomics*, 2007, **8**(1): 40
55. Ewing B, Hillier L, Wendl M C, Green P. Base-calling of automated sequencer traces using Phred.I. Accuracy assessment. *Genome Research*, 1998, **8**(3): 175–185
56. Chou H, Holmes M H. DNA sequence quality trimming and vector removal. *Bioinformatics*, 2001, **17**(12): 1093–1104
57. Langmead B, Salzberg S L. Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 2012, **9**(4): 357–359
58. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G T, Abecasis G R, Durbin R. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 2009, **25**(16): 2078–2079
59. Conesa A, Götz S. Blast2GO: a comprehensive suite for functional analysis in plant genomics. *International Journal of Plant Genomics*, 2008, **2008**: 1–12
60. Wu C, Sun S, Nimmakayala P, Santos F A, Meksem K, Springman R, Ding K, Lightfoot D A, Zhang H A. BAC and BIBAC-based physical map of the soybean genome. *Genome Research*, 2004, **14**(2): 319–326
61. Mun J, Kwon S, Yang T, Kim H, Choi B, Baek S, Kim J S, Jin M, Kim J A, Lim M, Lee S I, Kim H, Kim H, Lim Y P, Park B. The first generation of a BAC-based physical map of *Brassica rapa*. *BMC Genomics*, 2008, **9**(1): 280
62. Coe E, Cone K, McMullen M, Chen S S, Davis G, Gardiner J, Liscum E, Polacco M, Paterson A, Sanchez-Villeda H, Soderlund C, Wing R. Access to the maize genome: an integrated physical and genetic map. *Plant Physiology*, 2002, **128**(1): 9–12
63. Kelleher C T, Chiu R, Shin H, Bosdet I E, Krzywinski M I, Fjell C D, Wilkin J, Yin T, DiFazio S P, Ali J, Asano J K, Chan S, Cloutier A, Girn N, Leach S, Lee D, Mathewson C A, Olson T, O’connor K, Prabhu A L, Smailus D E, Stott J M, Tsai M, Wye N H, Yang G S, Zhuang J, Holt R A, Putnam N H, Vrebalov J, Giovannoni J J, Grimwood J, Schmutz J, Rokhsar D, Jones S J, Marra M A, Tuskan G A, Bohlmann J, Ellis B E, Ritland K, Douglas C J, Schein J E. A physical map of the highly heterozygous *Populus* genome: integration with genome sequence and genetic map and analysis of haplotype variation. *Plant Journal*, 2007, **50**(6): 1063–1078
64. Han Y, Gasic K, Marron B M, Beever J E, Korban S S. A BAC-based physical map of the apple genome. *Genomics*, 2007, **89**(5): 630–637
65. Córdoba J M, Chavarro C, Rojas F, Muñoz C, Blair M W. Identification and mapping of simple sequence repeat markers from Common Bean (*Phaseolus vulgaris* L.) bacterial artificial chromosome end sequences for genome characterization and genetic-physical map integration. *Plant Genome*, 2010, **3**(3): 154–165
66. Han Y, Chagné D, Gasic K, Rikkerink E H A, Beever J E, Gardiner S E, Korban S S. BAC-end sequence-based SNPs and Bin mapping for rapid integration of physical and genetic maps in apple. *Genomics*, 2009, **93**(3): 282–288
67. Ollitrault P, Terol J, Garcia-Lor A, Bérard A, Chauveau A, Froelicher Y, Belzile C, Morillon R, Navarro L, Brunel D, Talson M. SNP mining in *C. clementina* BAC end sequences; transferability in the *Citrus* genus (Rutaceae), phylogenetic inferences and perspectives for genetic mapping. *BMC Genomics*, 2012, **13**(1): 13
68. Shultz J L, Kazi S, Bashir R, Afzal J A, Lightfoot D A. The development of BAC-end sequenced-based microsatellite markers and placement in the physical and genetic maps of soybean. *Theoretical and Applied Genetics*, 2007, **114**(6): 1081–1090