Achievement, Challenge and Future Perspective of BAC Library Technology with a Focus on its Application to Soybean Genomics Researches

Hongyan Wu1, Zhengjun Xia2* and Kyuya Harada2
1Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Harbin, China
2Graduate School of Engineering, Osaka University, Suita, Japan

Abstract
Large genomic DNA insert-containing libraries e.g. bacterial artificial chromosome (BAC) libraries have made great contributions to advanced genomic studies, such as genome-wide sequencing, synteny analysis between genomes, physical mapping, and positional cloning of important QTL genes in soybean or other legume. As recent advances in next-generation sequencing (NGS) technology, application of BAC or other large insert-containing library has become less demanding. We briefly review the history and applications of soybean BAC library technology. The challenge and future prospective of soybean BAC library technology in the post genomic-era are also discussed.

Keywords: Bacterial artificial chromosome; Genome; Library; Gene cloning; Soybean

Introduction
In the late 1980s, the discovery of yeast artificial chromosomes (YACs) made cloning of megabase-sized DNA fragments possible [1]. However, several drawbacks for YAC, i.e. almost 50% of YAC clones are chimeric or resulted from insert rearrangements [2-4], lead such clones unsuitable for advanced genomic application, e.g. sequencing and mapping research [3,5]. Also it is rather time consuming to manipulate the YAC clones and deal with chimerase or insert rearrangement [6,7]. Bacterial artificial chromosomes (BACs) library was developed as alternative to YACs [8]. BACs rather are modified bacterial F factors and are not really artificial chromosomes in themselves, which is contrary to their name [9]. BAC had been widely employed as best of choice for advanced genomics research in many plant species including soybean. The advantages of BAC libraries can be summarized as follow:

- BAC clones are almost free of chimerism and insert rearrangements [7,10].
- BAC clone is also very stable since F factor genes (parA and parB) that prevent more than one BAC from simultaneously inhabiting a bacterium [4,8,11].
- The insert sizes of BAC library are large enough for downstream advanced genomics research e.g. developing physical contig of the region that QTL of interest are located as well as physical map of whole genome. Typically, average insert sizes are with a ranges of 80 and 200 kb for a good quality BAC library, although some clones can reach 500 kb or more [8,12].

It is also easy to manipulate in terms of preparation, screening and application of BAC libraries [6,13,14]. Most BAC vectors carry an antibiotic resistance gene in order to ensure all acquired colony carrying plasmid. Also polycloning sites within a reporter gene are very suitable for many restriction enzymes as well as colony selection, e.g. the blue/ white selection which is based on the principle of α-complementation of the β-galactosidase gene. The most common BAC vector is pBeloBAC11 [9,15].

Application of BAC libraries
Many DNA probes and markers can be used to screen the BAC libraries to localize specific BACs, therefore, we can superimpose genetic maps directly onto BAC-based physical maps [16,17]. Thus map-based cloning of genes responsible for specific phenotypes is greatly facilitated [18,19]. Many important genes were cloned and functionally confirmed [20-22].

High-throughput physical mapping (BAC-based mapping in conjunction with efficient multiplex screening methods) can be applied to the construction of BAC contigs encompassing entire chromosomes and/or complete chromosome sets of a genome [4,17]. Genome sequence information can be achieved through strategy of BAC by BAC i.e. soybean genome [23-25]. BAC-based physical mapping does require DNA polymorphism [26]. Therefore it can replace radiation hybrid mapping in which chromosomes are broken by radiation and propagated in cell cultures [27].

In the post genomic-era, a gap filling is very important and hard task for a perfect genome assembly. BAC libraries, contig assembly, STS-based mapping can facilitate the gap filling in genome assembly using whole-genome shotgun (WGS) sequencing. Fluorescence in situ hybridization (FISH) using BACs as probe [28-29] allows molecular and physical maps to be directly superimposed onto the framework of chromosomes, and subsequently the relationship between chromosome structure, DNA sequence, and recombination can be revealed [30].

Application of BAC library technology in soybean genomics research
In 1997, Marek and Shoemaker constructed a soybean BAC 40000 clones (4-5 genome equivalents) in 384-well plate microtiter dishes for studying disease resistance gene [31]. Danshet al., made approximately 30000 clones with an average insert size of 120 kilobase pairs for their research in cyst nematode resistance [18]. In 1999 Salimath and Bhattacharyya constructed – 45 000 clones for cultivar Williams 82, which carries the Rps1-k gene to Phytophthora sojae.
The usefulness of the BAC library has been constrained since construction, storage and screening are very difficult and laborious. Xia developed a simple and novel hand-made collection device for pooling without the needs for robotic manipulation [40]. In 2014, Xia et al. also developed a non-gridded BAC library for saving storage space as well as procedure for retrieving the target BAC clones from the positive pools [41]. Currently, here we list the websites regarding protocol, resource and application of BAC library of soybean or other leguminous species in Table 1.

**Challenge and future perspectives of BAC library**

Since the emergence of the next generations of sequencing technology, BAC library technology has faced strong challenge since construction, screening and application of BAC libraries are time and labor intensive [42-49]. Recent third-generation sequencing technology, which can yield longer multi kilo base reads and is getting popular in various genomic researches, further despoils the applications of BAC library technology. However, the special uniqueness of BAC library still makes a BAC library technology of first choice in modern genomic researches, e.g. refinement of physical map of target region of interest or at the whole genome level, comparative analyses between different species, and preservation of the genetic resources.

**Funding**

This work was supported by the Programs (31471518, 31271742, and 31301338) from National Natural Science of China; by Program (2011BAD35B06-2) from National Science and Technology Ministry of China.

**References**


**Table 1**: Online resources of BAC library of soybean and related leguminous species.

<table>
<thead>
<tr>
<th>Website</th>
<th>URL</th>
<th>Information or resource available</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soybase</td>
<td><a href="http://soybase.org">http://soybase.org</a></td>
<td>BAC based physical map of soybean; a comprehensive repository for professionally curated genetics, genomics and related data resources for soybean</td>
<td>[42]</td>
</tr>
<tr>
<td>LIS - Legume Information System</td>
<td><a href="http://legumeinfo.org/">http://legumeinfo.org/</a></td>
<td>A collaborative, community resource to facilitate crop improvement by integrating genetic, genomic, and trait data across legume species</td>
<td>[43]</td>
</tr>
<tr>
<td>The Soybean Genome Database (SoyGD)</td>
<td><a href="http://soybeangenome.siu.edu">http://soybeangenome.siu.edu</a></td>
<td>(BAC) fingerprint database, genetic map associated genomic data, soybean physical map</td>
<td>[44]</td>
</tr>
<tr>
<td>Clemson University</td>
<td><a href="https://www.genome.clemson.edu/services/bacrc/BAC_library">https://www.genome.clemson.edu/services/bacrc/BAC_library</a></td>
<td>BAC library construction</td>
<td>[45]</td>
</tr>
<tr>
<td>Daizubase</td>
<td><a href="http://daizu.dna.affrc.go.jp/">http://daizu.dna.affrc.go.jp/</a></td>
<td>BAC-end sequence of BAC library of a Japanese cultivar Enrei</td>
<td>[46]</td>
</tr>
<tr>
<td>The Arizona Genomics Institute (AGI)</td>
<td><a href="http://www.genome.arizona.edu/">http://www.genome.arizona.edu/</a></td>
<td>Targeted sequencing using BACs and BAC pools; - Whole Genome Profile BAC Physical Mapping</td>
<td>[47]</td>
</tr>
<tr>
<td>LegumeBase</td>
<td><a href="https://www.legumebase.brc.miyazaki-u.ac.jp/">https://www.legumebase.brc.miyazaki-u.ac.jp/</a></td>
<td>BAC clones of Lotus japonicus</td>
<td>[48]</td>
</tr>
<tr>
<td>Zhang Hongbin’s Lab (Texas A&amp;M University)</td>
<td>doi:10.1038/nprot.2011.455</td>
<td>Preparation of megabase-sized DNA from a variety of organisms using the nuclei method</td>
<td>[49]</td>
</tr>
</tbody>
</table>

[32]. Since William 82 was used for genome sequencing, this kind library is very useful for gap-filling and verification of sequencing data. In 1999 Tomkins et al., made a soybean BAC library using the plant introduction (PI) 437654 with 9 haploid genome equivalents for studying soybean cyst nematode (SCN) resistance genes [12]. In 2001 Lewerst et al., built a physical mapping of resistant and susceptible soybean genomes near the soybean cyst nematode resistance gene Rhg1, which has contribute the successful cloning of this gene [33,34]. Xia et al., constructed 53,760 clones with an average insert size of 116 kb for a Japanese cultivar, Misuzudaizu. This library represents 5.2 genome equivalents [35]. Also three-dimensional pools of the BAC library facilitate rapid and efficient PCR-based screening. This library had facilitated successful cloning of E1 [22], E2 [21] and E3 [20]. In 2004, Wu et al., constructed (RIBAC)-based physical map of the soybean genome based on plant-transformation-competent binary large-insert plasmid clone, which will benefit research in positional cloning of genes and QTLs. The BIBAC library contained 78,001 clones representing 9.6 haploid genomes [36,37]. Together with previous build BIBAC library for soybean cv. Forrest, two (BAC and BIBAC) libraries encompass 13.2 haploid genomes, providing the comprehensive clone resource for a single soybean genotype.

Integrated genetic and physical maps are extremely valuable for genomic studies as well as for guiding assembling whole genome shotgun sequences. Comparative physical mapping can reveal features of microsynteny between Glycine max, Medicago truncatula, and Arabidopsis thaliana. In 2013, Yan et al., used BAC-end cross-hybridization among G. max and M. truncatula contigs and uncovered microsynteny [38]. The result showed 85% of coding and 75% of noncoding sequences between G. max homoeologous (within genome duplicate) contigs were conserved in cross-hybridization, which confirmed one recent genome-wide duplication in G. max and suggested that M. truncatula also experienced ancient large-scale genome duplications.

Wu tried to use genetic marker anchoring by six-dimensional pools for development of a soybean physical map [39]. Six-dimensional soybean BAC pools can be efficiently used to anchor markers to soybean BACs despite the complexity of the soybean genome. In addition to anchoring markers, the 6-D pooling method was also effective for targeting BAC clones for investigating gene families and duplicated regions in the genome, as well as for extending physical map contigs.

Since the emergence of the next generations of sequencing technology, BAC library technology has faced strong challenge since construction, screening and application of BAC libraries are time and labor intensive [42-49]. Recent third-generation sequencing technology, which can yield longer multi kilo base reads and is getting popular in various genomic researches, further despoils the applications of BAC library technology. However, the special uniqueness of BAC library still makes a BAC library technology of first choice in modern genomic researches, e.g. refinement of physical map of target region of interest or at the whole genome level, comparative analyses between different species, and preservation of the genetic resources.

**Funding**

This work was supported by the Programs (31471518, 31271742, and 31301338) from National Natural Science of China; by Program (2011BAD35B06-2) from National Science and Technology Ministry of China.

**References**


