Review Article

Bacterial Artificial Chromosome Libraries of Pulse Crops: Characteristics and Applications

Kangfu Yu

Greenhouse and Processing Crops Research Centre, Agriculture and Agri-Food Canada, 2585 County Road 20, Harrow, ON, Canada N0R 1G0

Correspondence should be addressed to Kangfu Yu, kangfu.yu@agr.gc.ca

Received 27 April 2011; Revised 29 May 2011; Accepted 30 May 2011

Academic Editor: J. Birchler

Copyright © 2012 Kangfu Yu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Pulse crops are considered minor on a global scale despite their nutritional value for human consumption. Therefore, they are relatively less extensively studied in comparison with the major crops. The need to improve pulse crop production and quality will increase with the increasing global demand for food security and people's awareness of nutritious food. The improvement of pulse crops will require fully utilizing all their genetic resources. Bacterial artificial chromosome (BAC) libraries of pulse crops are essential genomic resources that have the potential to accelerate gene discovery and enhance molecular breeding in these crops. Here, we review the availability, characteristics, applications, and potential applications of the BAC libraries of pulse crops.

1. Introduction

“Pulse crops” refers to a group of more than sixty different grain legumes grown around the world. Pulse crop seeds, which are important for human nutrition, typically have 20–25% protein and 40–50% starch, are rich in dietary fibre, and usually have only small amounts of oil. Pulse protein is high in the amino acids lysine and methionine, making pulses nutritionally complementary to cereals, which are deficient in these two essential amino acids. Despite the importance of pulse crops for nutrition and food security in developing countries, they are considered to be minor on a global scale, and pulse genomes have been less extensively studied than those of major crops.

Large insert genomic DNA libraries are essential genomic resources for physical mapping, positional cloning, and genome sequencing of higher eukaryotes [1–4]. The bacterial artificial chromosome (BAC) cloning system has become an invaluable tool in genomic studies because of its ability to stably maintain large DNA fragments and its ease of manipulation [2, 5, 6]. BAC libraries are an important resource for the development of molecular markers that can be used for marker-assisted selection (MAS) for desirable agronomic traits. Development of simple sequence repeat (SSR) markers from BAC-end sequences is very cost effective [7, 8] and offers genome-wide coverage as all repeat types are systematically sampled in the randomly selected BACs [9].

Since the development of the BAC vector in 1989 [10], many BAC libraries have been developed for the major crop species, such as wheat, rice, corn, and soybean. In recent years, however, BAC libraries have also been developed for several pulse crops including mungbean (Vigna radiata L.), cowpea (V. unguiculata L.), lupin (Lupinus angustifolius L.), chickpea (Cicer arietinum L.), pigeonpea (Cajanus cajan L.), field pea (Pisum sativum L.), Lima bean (Phaseolus lunatus L.), and common bean (P. vulgaris L.) (Table 1). Understanding the characteristics of these libraries would facilitate the utilization of these resources in genomic studies of pulse crops. In this paper, we will review the characteristics and utilizations of the BAC libraries in pulse crops and will discuss the potential applications of these libraries in pulse crop genomic research.

2. Mungbean (Vigna radiata)

Mungbean (V. radiata) is an important grain legume in Asia, particularly India. It has 25–28% of protein; 4.5% of fibre, and 1.5% of fat [11]. It is eaten mainly as a dry seed but is also used for forage and as a vegetable in the form
of sprouted seeds or green pods. Mungbean is a relatively low-yielding crop that suffers from abiotic stresses such as heat and drought as well as many biotic stresses including foliar diseases and insect damage, particularly from bruchid (Callosobruchus chinensis), a weevil-like insect that can account for up to 40% loss of stored seed.

Miyagi et al. [12] constructed two mungbean BAC libraries that together gave a 3.5 × coverage of the 587 Mb genome (Table 1). The libraries were constructed from both radiata ssp. (green gram) and its wild progenitor sublobata ssp. (golden gram). By screening these libraries using RFLP probes, including Mgm213 that is very closely linked (1.3 cM) to the major locus conditioning bruchid resistance, two PCR-based markers were developed closely linked to this major locus. This information should facilitate the introgression of this resistance locus into agriculturally valuable cultivars. These libraries also could be used in the development of other PCR-based markers linked to other desirable traits.

3. Lupin (Lupinus angustifolius)

Narrow-leafed lupin (L. angustifolius) is a pulse crop native to the Mediterranean. It is grown as green manure as far north as 60 degrees N latitude. Used as a feed crop, the seeds contain 32–34% of protein. The plant is also used as a medicinal source of vitamins and of insecticides [26]. Some varieties can grow on marginal soil, and they have been used for the removal of heavy metals from contaminated soil [29]. Crop improvement, however, has been slow, partly due to lack of knowledge of its genomic structure.

In 2006, Kasprzak et al. [26] reported the construction of a BAC library (Table 1) from diploid L. angustifolius L. cv. Sonet. This genotype has valuable agronomic traits such as low alkaloid content (iuc), nonshattering pods (ta and le), and soft seed coat (moll). The library has been used for cytogenetic mapping of mitotic metaphase chromosomes to karyotype L. angustifolius. This first lupin ideogram was produced using BAC-fluorescence in situ hybridization (FISH), combined with primed in situ DNA labelling (PRINS) and computer measurement of chromosomes [30]. It was also used to produce a BAC-end sequence tag (BEST) marker, which facilitated aligning a new reference genetic map of lupin with the Lotus japonicus genome sequence [31].

4. Cowpea (Vigna unguiculata)

Cowpea (V. unguiculata) is the major source of dietary protein for millions of people in Africa and other regions of the developing world. Humans consume the seeds predominantly in the dry form, but fresh pods and green peas are eaten in some areas. Cowpea hay is an important fodder for animals in parts of West Africa [32]. Cowpea is well adapted to drought and heat and can grow where few other food crops can [33].

To date, a moderate number of SSR and single nucleotide polymorphism (SNP) markers have been developed to provide a reasonably densely covered map; however, the physical map needs work, and there has been little trait mapping or molecular breeding [34]. Two independent cowpea BAC libraries have been developed (Table 1), and high-information-content fingerprinting analysis [35] of 60,000 of these clones has allowed their assembly into 790 contigs with an average of 52 BACs per contig. This assembly represents 10 × coverage of the genome. End sequencing yielded 30,611 high-quality BESs (BAC end sequences) with an average length of 674 bp for a total of 20.6 Mb. In a BLAST search against a nonredundant protein database, 891 of the BESs had significant hits [20].

5. Chickpea (Cicer arietinum L.)

Chickpea (C. arietinum L.) is the third most important pulse crop in the world [FAO 2003, http://www.fao.org/]. Chickpea is primarily produced for human consumption both as whole seeds and as flour (dahl). The larger-seeded type, also known as garbanzo beans, is used in salads and vegetable mixes. Chickpea was one of the first grain legumes to be domesticated in the old world although there is considerable debate as to the exact area of origin [36]. The crop is highly susceptible to fungal diseases, particularly fusarium wilt (Fusarium oxysporum) and ascochyta blight (Ascochyta rabiei, Didymella rabiei), so attempts to improve yield have not been very successful to date. Any genetic resources that could assist in developing resistant cultivars would be of significant benefit to people who depend on plant sources for their protein.

Studies of resistance to ascochyta blight indicate that it is inherited in a complex and quantitative manner [37] and, although moderate resistance has been detected in existing accessions within germplasm collections [38, 39], the resistance is likely recessive, which hinders the development of superior cultivars. In addition, there is a paucity of genetic diversity among cultivars and wild accessions of C. arietinum, making it difficult to find material to use in crop improvement programs [40, 41].

The first genetic map of chickpea was developed in 1997 [42], but it had a low marker density. There are several molecular tools now available for chickpea improvement, including molecular genetic maps [37, 43–45] and identified quantitative trait loci (QTLs) [9, 46, 47], but new resources must be developed to obtain markers close enough to the loci of interest to be of use in MAS.

Rajesh et al. [21] reported the construction of a BAC library from germplasm of the line FLIP 84-92C (Table 1) in order to investigate resistance to fusarium wilt. He screened the library with a Sequence Tagged Microsatellite Site (STMS) marker, Ta96, which is tightly linked (1 cm) to a gene for fusarium wilt resistance (Foc3) and isolated two clones with a combined insert size of 200 kb. This marker was mapped to linkage group 2 where other wilt R genes against races 1, 3, 4, and 5 of the same pathogen were located [33]. These clones overlapped at the interior sequences and, upon homology search, showed resemblance over 165 bp to a ribosomal protein of Medicago truncatula (mtgs005e03), a close relative of chickpea, and to a zinc finer-like protein of Arabidopsis (AT1G14580.1). Further screening of the library with markers positioned near the R genes on linkage group 2
Table 1: Bacterial artificial chromosome (BAC) libraries of pulse crops.

<table>
<thead>
<tr>
<th>Species (common name)</th>
<th>Genotype</th>
<th>G-size(^a) Mb</th>
<th>Vector</th>
<th>R-site(^b)</th>
<th>No. of clones</th>
<th>A.I-size(^c) kb</th>
<th>Cov(^d)</th>
<th>Use(^e)</th>
<th>Ref.(^f)</th>
</tr>
</thead>
</table>
| **Phaseolus vulgaris L.**  
(Common bean) | Sprite    | 636            | pECSBC4 | RcoRI      | 33,792        | 100             | 5.3    | Ph.M.    | [13] |
| Bat 93                | pIndigoBac5 | HindIII      | 110,592 | 125         | 21.7          | Evolution       | [14] |
| G21245                | pIndigoBac5 | HindIII      | 55296   | 105         | 9.2           | Evolution       | [14] |
| G02771                | pIndigoBac5 | HindIII      | 55296   | 139         | 12.1          | Evolution       | [14] |
| G12949                | pIndigoBac5 | HindIII      | 30720   | 135         | 6.5           | Cloning         | [15] |
| G19833                | pIndigoBac5 | HindIII      | 55296   | 145         | 12.6          | Ph.M BES     | [16] |
| HR45                  | pIndigoBac5 | HindIII      | 33,024  | 107         | 5.5           | Ph.M. Cloning | [17] |
| HR67                  | BiBAC2     | BamHI        | 22,560  | 125         | 10.6          | Cloning         | [18] |
| OAC-Rex               | BiBAC2     | BamHI        | 31,776  | 150         | 7.5           | Cloning         | [18] |
| **P. lunatus L.**  
(Lima bean) | Henderson  | 636            | pIndigoBac5 | HindIII      | 55,296        | 110             | 9.6    | Evolution       | [14] |
| **Vigna radiate**  
(Mungbean) | ACC41      | 587            | pBeloBacII | HindIII      | 6,912         | 107             | 1.3    | Med-dev. | [12] |
| ATT3640               | pBeloBacII | BamHI        | 11,904  | 113         | 2.3           | Med-dev.       | [12] |
| **V. unguiculata**  
(Cowpea) | IT97K-499-35 | 611          | pCC1    | HindIII      | 36,864        | 150             | 9.1    | Cloning       | [20] |
| IT97K-499-35          | pCC1      | MboI       | 36,864  | 130         | 7.8           | Cloning       | [20] |
| FLIP84-92C            | pCLD04541     | HindIII      | 23,780  | 100         | 3.1           | Cloning       | [21] |
| **Cicer arietinum**  
(Chickpea) | Hadas      | 758            | pIndigoBac5 | HindIII      | 14,976        | 121             | 2.4    | Med-dev. | [22] |
| Hadas                 | pCLD04541   | BamHI        | 23,040  | 145         | 4.4           | Med-dev.       | [22] |
| Hadas                 | pECBAC1     | MboI       | 22,272  | 130         | 3.8           | Ph.M.       | [23] |
| Hadas                 | pCLD04541   | HindIII      | 23,800  | 142         | 7.2           | Ph.M.       | [23] |
| **Pisum sativum**  
(Bean) | PI269818    | 4397          | pCLD04541 | HindIII      | 55,680        | 105             | 1.3    | [24] |
| PI269818              | pCLD04541   | HindIII      | 55,280  | 105         | 2.2           | [24] |
| **Cajanus cajan**  
(Pigeonpea) | Asha       | 831            | pCC1BAC | HindIII      | 34,560        | 120             | 5.0    | Med-dev. | [25] |
| Asha                  | pCC1BAC     | BamHI        | 34,560  | 115         | 4.8           | Med-dev.       | [25] |
| **Lupinus Angustifolius**  
(Lupin) | Sonet      | 905            | pIndigoBac5 | HindIII      | 55,296        | 100             | 6.1    | Karyotype | [26] |

\(^a\)G-size: Genome size; genome size estimates based on Arumuganathan, and Earle [27] for all species but pigeonpea, which was based on Greilhuber and Obermayer, [28].

\(^b\)R-site: Restriction site.

\(^c\)A.I-size: Average insert size.

\(^d\)Cov: Genome coverage; for consistency of comparisons, genome coverage was calculated using the genome estimates in Table 1, which may be different from the cited literature.

\(^e\)Ph.M, physical mapping; BES: BAC end sequences; Mar-dev: marker development.

\(^f\)Most of the libraries, if not all, may be accessed through collaborative research by contacting the authors who developed the library.

will further elucidate the organization of the r gene complex and generate more tightly linked markers.

Lichtenzveig et al. [22] constructed both a BAC and a BIBAC (plant transformation-competent binary BAC) library of chickpea cv. Hadas (Table 1). Winter et al. noted that the abundant SSRs in chickpea have a high level of polymorphism and therefore would be useful for mapping and gene tagging [43]. Lichtenzveig screened the libraries with eight different synthetic oligos; sequenced the SSR regions with their flanking sequences of positive clones; designed primers complementary to the flanking sequences. A total of 233 new SSR markers for chickpea were developed and characterized, which will be of significant value in further genomic studies.

A genome-wide BAC/BIBAC-based physical map of chickpea was developed by Zhang et al. using fingerprint analysis of four libraries, two of which were constructed for their study (Table 1) [23]. The map consists of 55,029 clones assembled into 1,945 contigs. Each contig contains 2 to >199 clones with an average of 28.3 clones per contig. The average contig size is 559 kb, and the contigs collectively span 1,088 MB that is 47% larger than the estimated size of the chickpea genome. This can be explained by overlap of contigs, underestimation of genome size, or overestimation of the average insert size of the source clones, the first explanation being most likely. The 10 × coverage of the chickpea genome proved adequate to construct a high-quality physical map sufficient for use in various aspects of
chickpea genomics. The accuracy of the map was verified using independent contig building methods, different fingerprinting methods, and SSR marker hybridization [23].

Using their map, Zhang et al. identified three contigs that likely contain, or are closely linked to, the RDR QTL4.1, which accounted for 14.4% of the variability in Ascochyta blight resistance. They also identified a contig associated with DTF QTL8, which results in approximately four days earlier flowering, with early flowering being an advantage in escaping Ascochyta blight that tends to strike later in the season. In addition, since the physical map was constructed from both BAC and Agrobacterium-mediated plant transformation ready BIBACs, the inclusion of the BIBACs should facilitate cloning; promote functional analysis of the chickpea genome by genetic transformation at the whole genome level; improve the identification and utilization of QTL for other genes of agronomic interest [23,48].

6. Pigeonpea (Cajanus cajan L.)

Pigeonpea (C. cajan L.) is an important food legume predominantly cultivated in tropical and subtropical regions. It is drought tolerant with a large range of maturities. Its seeds have 20–22% protein and are generally consumed as green peas, whole grain, or split peas. Pod husks are used as fodder and branches, and stems are used as domestic fuel [25].

Pigeonpea has a relatively low level of genetic diversity, which has made conventional breeding and development of genomic tools relatively ineffective [25]. In 2006, the Pigeonpea Genomics Initiative, with funding from the Indian Council of Agriculture and the US National Science Foundation, began an in-depth analysis of the pigeonpea genome. One result of this initiative was the development of a BAC library, made at UC Davis, from the reference genotype “Asha” that provides an 11 × coverage of the 808 Mb genome (Table 1).

This BAC library is an important resource for the development of molecular markers that can be used for MAS for desirable agronomic traits. Development of SSR markers from BAC-end sequences is very cost effective [7, 8] and offers genome-wide coverage as all repeat types are systematically sampled in the randomly selected BACs [9]. Varshney et al. end-sequenced 50,000 randomly selected clones from this BAC library generating a total of 87,590 BAC end sequences (BESs) [25]. These were screened with a microsatellite search module resulting in the identification of 18,149 SSRs representing 6,590 BAC clones. Amplified products were obtained from 2,565 of the designed primer pairs that will be used to identify polymorphism in a set of 24 pigeonpea genotypes [25].

Using NBS-LRR (nucleotide binding site—leucine-rich repeat disease resistance) homologues based on Medicago truncatula, researchers at UC Davis found 756 BAC clones that could form the basis for an SSR molecular resource linked to 90 BAC contigs [25]. This information will be extremely useful in molecular breeding and disease-resistance dissection. Along with high-density molecular maps, transcription sequences, and so forth, the availability of a BAC library and all that it leads to could revolutionize pigeonpea crop improvement.

7. Pea (Pisum sativum L.)

Pea (P. sativum L.) is a relatively inexpensive and highly nutritious crop, and pea proteins are of great nutritional importance, being one of the major food legumes grown in different parts of the world [49]. Processed pea can be utilised in specific food formulations for preschool children to improve their protein intake. They are rich in lysine and complement cereals, producing an amino acid profile complying with the FAO reference pattern [50]. The pea genome is estimated to be 3947–4397 Mbp/1C or 10–30 times the size of the genomes of Arabidopsis thaliana, Lotus japonicus, and Medicago truncatula [24]. Two BAC libraries of pea were constructed for the genotype PI 269818. One BAC library was constructed in the binary vector pCLD04541 for direct transformation of candidate pea gene BACs into plants [51]. The pCLD04541 library is composed of 55,680 clones with an average insert size of 105 kb. The other library used the single-copy oriS-based vector (pIndigoBac5) and contains 65,280 clones with a mean insert size of 105 kb. Partially HindIII-digested DNA fragments were cloned in both BAC vectors and the libraries encompassed about 3.2 × of the large haploid pea genome with about 1% of the clones from chloroplast and 0.1% of empty vectors [24]. Successful amplifications of low copy pea-specific resistance gene analogs (RGA) indicated that the libraries should be useful for many applications in genetic studies of pea.

8. Common Bean (Phaseolus vulgaris L.)

Common bean (P. vulgaris) is grown and consumed principally in developing countries in Latin America, Africa, and Asia. It is largely a subsistence crop eaten by its producers and, hence, is underestimated in production and commerce statistics. Common bean is a major source of dietary protein, which complements carbohydrate-rich sources such as rice, maize, and cassava. It is also a rich source of dietary fibres, minerals, such as iron and zinc, and certain vitamins [18]. Among the eleven BAC libraries constructed in the genus Phaseolus today, ten of them were common bean (P. vulgaris) (Table 1). The first common bean BAC library was developed by Vanhouten and Mackenzie in 1999 with the Sprite snap bean-derived genotype for physical mapping of the nuclear fertility restorer Fr. locus [13]. In 2006, four BAC libraries, three for common bean genotypes BAT93, G21245, and G02771, and one for lima bean, cv. Henderson (P. lunatus), were developed to study the evolution of the arcelin-phytohemagglutinin-a-amylase inhibitor (APA) multigene family [14]. The four BAC libraries have a range of 9–20-fold genome coverage that should make them useful genetic resources for studying common bean and lima bean. The BAT 93 BAC library has been used successfully for cytogenetic studies of bean chromosomes [52,53]. BAC libraries were also developed for common bean genotypes G19833 [16], G12949 [15], HR45 [17], G02333 [19], HR67 [18], and OAC-Rex [18].
of the restriction sites for any enzyme in genomic DNA [63] or by nonclonable or unstable DNA segments in the E. coli host [64]. As an alternative, the fosmid cloning system is rapidly emerging as a method of choice to rapidly create high-titer “mini-BAC libraries” with an average insert size of 40 kb [65]. Because the genomic DNA used in fosmid library construction is usually mechanically shared, fosmid libraries are quite useful for closing gaps in physical mapping [66]. Today, the improved fosmid vector from Lucigen can accommodate DNA fragment as large as 90 kb [http://lucigen.com/store/Custom-Fosmid-Libraries.html].

Although the problems above mentioned were not reported for the BAC libraries developed in pulse crops, instability associated with insert size over 100 kb was observed in potato [67]. The mechanism of the instability of BAC clones is unknown. It is likely that the current BAC vectors and a host strain like DH10B are not able to stably maintain DNA sequences with certain unique features, including tandemly repeated sequences. Therefore, it may be possible to partially or completely overcome the BAC instability problem by selecting appropriate E. coli strains [67].

10. Summary and Future Prospects

In summary, twenty-five BAC libraries of pulse crops have been reported in the literature. The BAC libraries are important genomic resources that have been used for (1) physical mapping of pulse genomes, (2) molecular marker development of microsatellite markers, (3) map-based cloning of genes or QTL for important agronomic traits, (4) evolutionary study of multigene families, (5) karyotyping genomes through BAC-FISH, and (6) whole genomic sequencing.

Looking into the future, the BACs of pulse crops should have potential applications in pulse comparative genomics and functional genomics in addition to those above-mentioned. It is well known that macro- and microsynteny are widespread within legumes. Based on 1000 anchored BAC ends, more than half of all soybean BAC contig groups exhibit microsynteny with Medicago truncatula [68]. By comparing BAC end sequences, microsynteny was found among M. truncatula, G. max, and Lotus japonicus [69]. Significant macro- and microsynteny were observed among G. max, P. vulgaris and Vigna radiata [70]. Large-scale macrosyntentic blocks were also observed among P. vulgaris, M. truncatula, and L. japonicus [71]. Because extensive genomic information is available for soybean (http://soybase.org/), medicago (http://gbrowse.jcvi.org/cgi-bin/gbrowse/medicago/#search), and Lotus (http://www.plantgdb.org/LjGDB/), the genetic synteny between pulse and the model legume species will help pulse researchers to speed up the understanding of pulse genomes by comparative genomics (http://www.comparative-legumes.org/).

Most of the BAC applications in pulse crops to date are of structural genomics nature; however, the application of BACs in functional genomics analysis of pulses also has great potential. Since large insert clones in BAC vectors are more likely to contain the necessary promoter, enhancer, and silencer combination, mimicking the natural expression of
the gene of interest, the advantages of the BAC transgenic approach are significant compared to the conventional transgenic approach [72]. Since the presence of appropriate regulatory elements could cause a gene in BACs to be expressed with spatial and temporal accuracy at similar levels to the endogenous loci, the integration site effects experienced by conventional gene transformation methods might be eliminated or reduced [73]. In addition, the ability of modification techniques to insert or delete sequences or alter sequences as discrete as a single-point mutation could make the BAC transgenic system a powerful tool for addressing both mechanistic and functional questions.

Acknowledgments

The author acknowledges the financial supports from Ontario White Bean and Colored Bean Producers and Agriculture and Agri-Food Canada. Critical review of the paper by Dr. V. Poysa is gratefully acknowledged as well.

References


