KEYWORDS: bacterial artificial chromosome, vector, molecular biology

Special section on techniques:

So you want to Work with Giants: The BAC Vector

Kevin She

Pathology, University of British Columbia
Submitted April 2003

Introduction

Bacterial Artificial Chromosome (BAC) libraries have been prominently used for the construction of physical genetic maps of many model species including *Mus musculus*, *Arabidopsis thaliana,* and of course, *Homo sapiens*. Also being studied are maps of economically important species such as *Zea mais* (corn), *Oryza sativa* (rice) as well as dangerous pathogens including herpes simplex virus and Epstein-Barr virus. Although BACs were created primarily to facilitate complex genome analyses and has been an extremely valuable tool for this purpose, there is a wide range of other uses, some of which will be discussed later in this manuscript.

Development of BAC vectors

The BAC vector was constructed [1] based on *Escherichia coli* Factor F' (F') since methods of DNA manipulation in bacteria are well established. F' is an incompatibility group involved in *E. coli* chromosomal transfer and conjugative ability, which can exist as an extra-chromosomal element. The original BAC vector, pBAC108L, is based on a mini-F plasmid, pMBO131 (Figure 1) which encodes genes essential for self-replication and regulates its copy number inside a cell. The unidirectional self-replicating genes are *oriS* and *repE* while *parA* and *parB* maintain copy number to one or two for each *E. coli* genome. Added to the vector were multiple cloning sites flanked by "universal promoters" T7 and SP6, all flanked by GC-rich restriction enzyme sites for insert excision. *cosN* and *loxP* sites were cloned in (by bacteriophage l terminase and P1 Cre recombinase, respectively) to permit linearization of the plasmid for convenient restriction mapping. Additionally, there is a chloramphenicol resistance gene for negative selection of non-transformed bacteria. This vector is capable of maintaining insert DNA in excess of 300 kilobases (kb). Because there is no positive selection of clones with successful DNA fragment insertion, libraries created with pBAC108L had to be hybridization screened with whole DNA. Since the inception of BAC vectors, there have been many modifications intended to increase the ease-of-use as well as for use in specific systems and situations [2-5]. pBeloBAC11 [2] and pBACe3.6 [6] (Figure 1) are modified BAC vectors based on pBAC108L and are commonly used as a basis for further modification.

pBeloBAC11

The primary characteristic of pBeloBAC11 is the addition of a *lacZ* gene into the multiple cloning site [2] of pBAC108L. On plates supplemented with X-gal/IPTG, an intact lacZ gene encodes b-galactosidase which catalyses the supplemented substrate into a blue substance. Successful ligation of insert DNA into the vector inactivates *lacZ*; hence, white colonies indicate the presence of a successful vector-insert ligation. However, pBeloBAC11 is still a low-copy number plasmid from the presence of *parA* and *parB*.

pBACe3.6

pBACe3.6 is also based on pBAC108L but is more highly modified than pBeloBAC11 [6]. In order to address the issue of low plasmid copy numbers, the P1 replicon in F' was deleted and a removable high copy number replicon originating from an inserted pUC19 was introduced. Additionally, a different positive selection mechanism from blue/white selection used in pBeloBAC11 is employed. pBACe3.6 includes a 2.7 kb pUClink stuffer fragment flanked by two sets of six restriction sites within a *sacB* region. The *sacB* gene product is levansucrase, which converts sucrose, supplemented in the media, to levan, which is toxic to *E. coli* host cells. Hence, if the vector is re-ligated without

an insert, the functional *sacB* produces levansucrase and the cells die before forming colonies. Successful ligation of an insert into the vector increases the distance from the promoter to the coding region of *sacB*, disrupting toxic gene expression in the presence of sucrose.

Other BAC vectors

In addition to the commonly used pBeloBAC11 and pBACe3.6 vectors, there are many specialized BAC vectors carrying a variety of different combinations of drug resistance genes. Also, many different selection mechanisms and markers are available. Modification of cloning sites (unique restriction endonuclease sites) are also common as are the addition of genes and promoters specific to different strains of bacteria.

Why BACs?

DNA is the basic code that determines how a terrestrial organism functions and the manipulation of DNA is a well-established field with many different techniques and applications. The underlying premise in working with DNA is the ability to modify and amplify the sequence of interest. There are many methods of managing and amplifying DNA and many take advantage of the rapid and controllable biosynthetic ability of bacteria. Phages, viruses that infect bacteria, were a commonly used system that could produce very pure DNA of interest. However, it could be difficult to work with and the maximum size of each non-virus DNA fragment is usually limited [7]. Plasmids are double stranded DNA vectors that are maintained and replicated in bacteria, which are easy to manipulate and maintain. However, the drawback of plasmids is that their non-vector insert size limit is around 10 kilobases (kb) [8]. Cosmids are essentially plasmids with at least one cohesive end site (cos) from a bacteriophage and require viral packaging prior to transfection into bacteria. Cosmids can accommodate non-vector inserts ranging from 5 kb to 23 kb [9]. Inserts up to 35 kb can be achieved, albeit at a sacrifice of packaging and transformation efficiency (personal observation) which are critical for library construction. However, even the increased maximum insert size of a cosmid system is insufficient or cumbersome for the study of a cluster of bacterial genes, large double-stranded DNA viral genomes, genes encoding non-ribosomal polypeptide synthetic proteins, or constructing physical maps of whole genomes.

BACs vs YACs

Yeast artificial chromosomes (YAC) can accommodate insert sizes in excess of 2 megabases (Mb) which vastly overcome the size limitation of previous vectors. However, yeast spheroplast transformation is relatively inefficient, and large amounts of DNA are required for library construction [10]. YAC DNA, in addition, is linear and can be difficult to isolate intact due to its susceptibility to shear. Most importantly, YAC clones are often unstable and chimeric [11] in nature and sequences with repetitious elements are prone to rearrangement [12] or are un-clonable.

Bacterial artificial chromosomes overcome many of the problems involved with YACs [1]. BACs can be transfected into *E. coli* by electroporation at efficiencies up to 100 times greater than yeast transformation. BAC DNA exists in supercoiled circular form that permits easy isolation and manipulation with minimal breaking. In addition, clones can be effortlessly isolated via miniprep alkaline lysis and directly re-introduced into bacterial cells. Importantly, bacterial recombination systems are well characterized and recombination deficient strains of *E. coli* are readily available. It is not surprising, then, that BAC DNA is very stable, a trait that is aided by the low copy numbers maintained in each cell. However, there are BAC vectors that can attain very high copy numbers while maintaining DNA stability [6]. One drawback of BAC vectors compared to YAC vectors is that the maximum insert size that BACs can accommodate merely exceeds about 300kb although clones in the mid-300 kb range are obtainable. Additionally, the number of successfully generated clones decreases when attempting to achieve higher insert sizes (personal observation) and there has been suggestion that there are species-specific library insert-size limitations based on base-pair content and sequence dissimilarities [13].

Making Your Own

BAC vectors are primarily used in the construction of libraries, however, BAC vectors have also been used in other applications such as to study and modify complete double-stranded DNA viruses [3, 5] which are too large to fit into a plasmid or cosmid vector. In such cases, many of the techniques involved in manipulating plasmids can also be applied to BAC vectors. Nonetheless, owing to the large size of BAC plasmids there are special considerations and techniques, some of which will be briefly described.

There are a number of protocols for BAC library construction on the internet, available both at academic institutions as well as commercial services. One protocol from an academic institution can be found here: http://www.tree.caltech.edu/protocols/BAC_lib_construction.html. A BAC library is a collection of clones that theoretically should have several-fold coverage of the genome of interest, thus the library should represent every nucleotide in that genome.

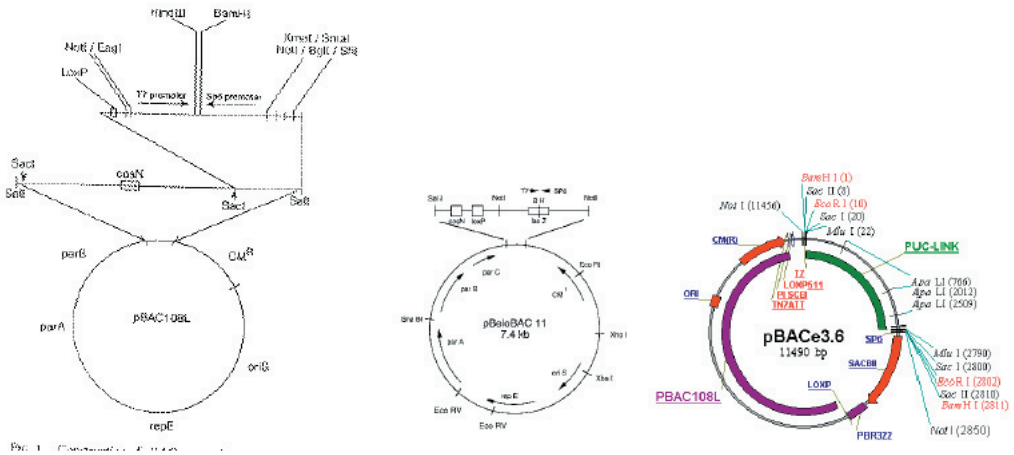

Figure 1 Comparison of the original BAC vector and a more recent widely used vectors

Step 1: pick your vector

The first step is to choose an appropriate vector from the wide variety of F' based plasmids. pBAC108L, pBeloBAC11, and pBACe3.6 are available and commonly used; hence, these are well characterized for their stability within different hosts, their cloning efficiency with DNA from different species, and other practical characteristics. However, available vectors may not suitable for your needs and may require modification. An important consideration is in choosing a plasmid that is maintained at a low copy-number or a high copy-number. Low copy-number plasmids tend to be more stable than high copy-number plasmids in that there are lower percentages of recombined clones in each ligation reaction. Long term stability of clones is not significantly different between the two types of plasmids. If your application requires a large amount of DNA or is throughput dependent, the increased number of recombined clones may be reasonable trade-off. If however, your application is intolerant of false positives or false negatives, using a low copy-number plasmid may be advisable.

Positive selection mechanisms is another point of deliberation, although many of the high-end automated colony pickers and sorters are able to use colour to determine which colonies to pick, if you are using an automated system, a toxic positive selection system such as *sacB* may be more appropriate. Antibiotic selection should also be considered based on the bacterial host that you are using. DH10B is a well-characterized recombination deficient *E. coli* that is often used with pBAC108L, pBeloBAC11, pBACe3.6, and their derivatives. However, if you are investigating or screening protein expression, DH10B may be an inappropriate production host or may only be an intermediate host. Chloramphenicol is a fairly standard resistance marker, as is ampicillin or apramycine. However, different antibiotics are effective to varying degrees, vary in mammalian toxicity, or may not be compatible with your host bacteria. Chloramphenicol and hygromycin are relatively more hazardous to humans, and ampicillin allows satellite colony growth, while apramycine - especially in conjunction with *sacB*/sucrose selection - is relatively more effective. The relative cost between different antibiotics may also be a factor in your decision.

Step 2: select your DNA

Once you have chosen your vector, you will have to size-select your DNA. The primary difficulty in this step is to prevent shearing or other degradation of the DNA. Shear forces will not only result in damaged DNA ends, affecting its ability to ligate into a restriction site on the vector, but also decrease the maximum insert size possible for the library. The most common method is to embed the cells containing the genome of interest in low-melting point agarose followed by digestion with lysozyme and proteinase K. The addition of spermine/spermidine is recommended by many protocols, but their benefit in reducing DNA degradation is debatable. Treatment method and incubation times will depend greatly on the characteristics of the embedded cells as well as their concentration. Cells from different strains of bacteria and cells from different species may have vastly different characteristics. After lysis treatment, the genomic DNA of interest should be free from potentially degradative enzymes and in a shear-resistant environment (in the agarose "plugs"). There are also methods of isolating the DNA in liquid, but damage from shear forces will greatly decrease both the quality and the maximum size of the DNA. However, some applications may require such a method, such as

extraction of DNA from soil [14, personal observation].

Step 3: clone into a vector

The most efficient method of cloning into a vector is to make use of the unique restriction sites in the multiple cloning site. Many of these restriction endonuclease sites result in overhangs which increases cloning efficiency, although blunt-end ligations are also possible. In order to generate mixed DNA of the desired size, the DNA plugs have to undergo partial digestion with a restriction enzyme corresponding to the restriction site on the vector that you wish to clone into. Since different genomic DNA have different distributions of restriction sites the amount of restriction enzyme and the reaction time required to generate DNA of the appropriate size will differ. Indeed, genomic DNA in different batches of plugs will respond differently to restriction enzyme digestion. Commonly, to determine conditions to generate DNA of the desired size, a small amount of DNA plugs will be chopped up, and aliquots subjected to either varying amounts of restriction enzyme or to a set amount of enzyme but treated for different lengths of time. A wide range of conditions may have to be tried in order to generate a successful ligation reaction to generate the library. It is not surprising that it is easier to generate libraries with a smaller mean insert size than a library with a larger mean insert size. Additionally, libraries with smaller mean insert sizes tend to have a higher number of clones. Minimum mean insert size will depend on the intended application of the library or the estimated size of the genes/pathways of interest.

Since the maximum size that a standard agarose gel can resolve is about 40 kb, an alternative method will be required to separate and visualize DNA in the desired 100 to 300 kb range. Contour-clamped Homogenous Electric Field – Pulsed Field Gel Electrophoresis (CHEF-PFGE) is a method capable of separating DNA by size to well over 10 Mb (10,000 kb) and is commonly used to visualize DNA used in BAC creation as well as the BAC clones themselves. Although the theory of PFGE is a matter of debate, it is understood that DNA above 30 to 50 kb migrate with the same mobility under continuous field electrophoresis. However, if the DNA is forced to change direction during electrophoresis (i.e., if the electric field moves), smaller sized DNA will begin moving in the new direction more quickly than larger DNA, thus the larger DNA lags behind and the smaller DNA is separated out [15]. There are many commercial set-ups available and information on program settings is available [16] although empirically determining the correct parameters such as separation area, field strength, pulse time, re-orientation angle, agarose concentration, and temperature will likely be required.

Commonly, the PFGE-run with the partially digest plugs will be set so that 150-250 kb range and 250-350 kb range (if the intended size of the BAC clones is from 100 to 200 kb) is spread over about 1 cm each. Preferably, with smaller and larger DNA separated over a much greater distance. The two ranges (150-250 kb, 250-350 kb) are excised from the gel. The two main reasons for this are that (1) by "focusing" on the desired size range, a higher concentration of DNA can be eluted from the gel and (2) by having DNA of undesired sizes, especially smaller fragments, more separated there is less contamination of small fragments in the DNA of interest. From a practical standpoint, the 150-250 kb region on the gel will contain a range of DNA sizes mostly in the 100 kb to 300 kb range. Since smaller DNA fragments have an increased likelihood of inserting into a vector upon library ligation, in order to maintain a high mean insert size in the library, care should be taken to avoid the presence of small DNA fragments.

There are two main approaches to recovering the size-selected DNA. Agarase is an enzyme that degrades polymerized agarose. Although this magnificently prevents shearing, the enzyme can be expensive and the buffers needed may interfere with the function of the ligase enzyme. Electroelution of DNA from the gel excision is another method of recovering the DNA fragments. Although there is theoretically increased shear damage, the practical results are insignificant, especially when factoring in the cost of agarase. The importance of recovering the DNA in high concentration is that precipitation and resuspension in a smaller volume will greatly degrade the DNA, again, through shearing forces. It is, however, acceptable to concentrate the DNA through dialysis.

The vector should be prepared by restriction enzyme digestion at the intended restriction site (the same or compatible enzyme that the genomic DNA was partially digested with) followed by phosphatase treatment. Calf intestinal alkaline phosphatase and shrimp alkaline phosphatases are commonly used. Depending on the restriction enzyme used, the size of the insert DNA, and the ligation reaction volume, the ligation setup should have between 3:1 to 10:1 molar ratio of vector to insert. A wide range of vector to insert ratios may have to be tried in order to generate a successful ligation. Ligation reactions are usually carried out at 16^{o}C overnight.

Step 4: transformation into a host

After the ligation reaction is finished, transformation into an *E. coli* host is trivial, although the amount of DNA to be transformed and the amount of bacteria to transform into may have to be empirically determined for each ligation reaction in order to maxi-

mize the titre of the library. Regeneration and plating protocols are readily available. If you are lucky (or just plain good), you now have a BAC library!

A consideration when working with BAC clones is that they tend to be much larger than plasmid or cosmid clones. PCR based manipulations, such as when screening or sequencing, may require more clonal DNA and may require agents such as DMSO or betaine to disrupt secondary structure.

Applications

BACs are useful for the construction of genomic libraries but their range of use is vast. Application of BACs as tools span basic science, economically rewarding industrial research, and fields as prosaic as animal husbandry. Related to genomic analyses, the ease with which phylogenetic lineage determination between species has been vastly aided by easily manipulated and easily sequenced [17] BAC libraries [14]. Also related and aided by BAC libraries is the study of horizontal gene transfer [18] and since bacterial genes are usually clustered, the ability of BAC vectors to accommodate large inserts has allowed the study of entire bacterial pathways. Of academic interest is the definition of what constitutes a minimal genome - BAC vectors because of their accommodation of large insert sizes and, more importantly, their stability are marvellous tools for this area of investigation. The biosphere is dominated by micro-organisms [19] however, only a minuscule fraction has been studied because the vast majority of micro-organisms or either uncultureable or are termed viable- but uncultureable. By isolating DNA directly from soil or from marine environments, the "metagenomes" of these organisms can be cloned into BAC vectors and indirectly studied [14].

Indeed, like the unidentified therapeutic compounds in jungle flora, tiny micro-organisms are a potentially gargantuan source of antibiotic molecules. It stands to reason that micro-organisms will produce compounds to inhibit the survival of competitors and millions of years of co-evolution would surely provide a bounty of these molecules. Undeniably, *Actinomycetes*, a common class of soil bacteria – many of which are difficult or impossible to culture, has been a major source of novel antibiotics in the previous decades. There are several labs working on isolating biosynthetic pathways from environmental samples such as soil, marine-environments, and heavily contaminated or polluted areas. Other industrial research fields where BAC vectors are invaluable tools in cataloguing novel genomes is in the discovery of novel enzymes. Work has been done on identifying enzymes that are involved in biopolymer hydrolysis or even radioactive waste management. In addition to antibiotic and industrially important enzymes, BAC vectors have been instrumental in studying large double stranded DNA viruses [3, 5, 20] both from an academic viewpoint and as a tool to develop improved vaccines.

Bacterial Artificial Chromosome libraries are relatively easy to generate and BAC vector-based clones are readily manipulated using a wide field of well-established techniques. BACs are useful in academic research, in industrial research, and even in fields as prosaic as animal husbandry [21, 22]. Current and future work such as the developing the ability to use BAC transgenic animals [23] continues to expand the versatility and the ease of use of bacterial artificial chromosomes.

References

1. Shizuya, H., Birren, B., Kim, UJ., Valeria, M., Slepak, T., Tachiiri, Y., & Simon, M. 1992 *Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia coli using an F-factor-based vector.* Proc. Natl. Acad. Sci. **89; 879-8797**

2. Kim, UJ., Birren, BW., Slepak, T., Mancino, V., Boysen, C., Kang, HL., Simon, MI., & Shizuya, H. 1996 *Construction and characterization of a human bacterial artificial chromosome library.* Genomics **34;213-218**

3. Smith, GA. & Enquist, LW. 1999 *A self-recombining bacterial artificial chromosome and its application for analysis of herpesvirus pathogenesis.* Proc. Natl. Acad. Sci. **97; 4873-4878**

4. Fu, H., & Dooner, HK. 2000 *A gene-enriched BAC library for cloning large allele-specific fragments from Maize: Isolation of a 240-kb contig of the bronze region.* Genome Res. **10; 866-873**

5. Yu, D., Smith, GA., Enquist, L., & Shenk, T. 2001 *Construction of a self-excisable bacterial artificial chromosome containing the human cytomegalovirus genome and mutagenesis of the diploid TRL/IRL13 gene.* J. Vir. **76; 2316-2328**

6. Frengen, E., Weichenhan, D., Zhao, B., Osoegawa, K., van Geel, M., & Jong, PJ. 1999 *A modular, positive selection bacterial artificial chromosome vector with multiple cloning sites.* Genomics **58; 250-253**

7. Flotte, TR. 2000 *Size does matter: overcoming the adeno-associated virus packaging limit.* Respir. Res. **1; 16-18**

8. Hershfield, V., Boywer, HW., Yanofsky, C., Lovette, MA., & Helinksi, DR. 1974 *Plasmid*

ColE1 as a molecular vehicle for cloning and amplification of DNA. Proc. Natl. Acad. Sci. **71; 3455-3462**

9. Koukolikova-Nicola, HB., Lindenmaier, W., & Collins, J. 1988 *Cosmids.* Biotechnology **10; 113-127**

10. Monaco, AP. & Larin, Z. 1994 *YACs, BACs, PACs and MACs: Artificial chromosomes as research tools.* Trends. Biotech. **67; 4566-4576**

11. Green, ED., Riethman, HC., Dutchik, JE., & Olson, MV. 1991 *Detection and characterization of chimeric yeast artificial chromosome clones.* Genomics **11; 658-669**

12. Ioannou, PA., & de Jong, PJ. 1996 *Construction of bacterial artificial chromosome libraries using the modified P1 (PAC) system.* "Current Protocols in Human Genetics" (Dracopoli, NC., Haines, JL., Korf, BR., Moir, DT., Morton, CC., Seidman, CE., Seidman, JG., & Smith, DR. eds) **5.15; 1-24**

13. Rondon, MR., Raffel, SJ., Goodman, RM., & Handelsman, J. 1998 *Towards functional genomics in bacteria: Analysis of gene expression in Escherichia coli from a bacterial artificial chromosome library of Bacillus cereus.* Proc. Natl. Acad. Sci. **96; 6451-6455**

14. Rondon, MR., August, PR., Bettermann, AD., Brady, SF., Grossman, TH., Liles, MR., Loiacono, KA., Lynch, BA., MacNeil, IA., Minor, C., Tiong, CL., Gilman, M., & Goodman, RM. 2000 *Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms.* Appl. Env. Microbiol. **66; 2541-2547**

15. Chu, G., Vollrath, D., & Davis, RW. 1986 *Separation of large DNA molecules by contour-clamped homogenous electric fields.* Science **234; 1582-1585**

16. Anand, R. & Southern, EM. 1990 *Pulsed field gel electrophoresis*. "Gel Electrophoresis of Nucleic Acids: A Practical Approach" (Rickwood, D., & Hames, BD., eds) **101-123**

17. Chen, CX. & Gmitter, FG. Jr. 1999 *Direct cloning and sequencing of bacterial artificial chromosome (BAC) insert ends based on double digestion.* Plant. Mol. Biol. Rep. **17; 231-238**

18. Hugenholtz, P. & Pace, NR. 1996 *Identifying microbial diversity in the natural environment: a molecular phylogenetic approach.* Trends. Biotech. **14; 190-197**

19. Whitman, WB., Coleman, DC., & Wiebe, WJ. 1998 *Prokaryotes: the unseen majority.* Proc. Natl. Acad. Sci. **95; 6578-6583**

20. Adler, H., Messerle, M., & Koszinowski, UH. 2001 *Virus reconstituted from infectious bacterial artificial chromosome (BAC)-cloned murine gammaerpesvirus 68 acquires wild-type properties in vivo only after excision of BAC vector sequences.* J. Vir. **75; 5692-5696**

21. Anderson, SI., Lopez-Corrales, NL., Gorick, B., & Archibald, AL. 2000 *A large-fragment porcine genomic library resource as a BAC vector.* Mamm. Genome **11; 811-814**

22. Buitkamp, J., Kollers, S., Durstewitz, G., Fries, R., Welzel, K., Schäfer, K., Kellermann, A., & Lehrach, H. 2000 *Construction and characterizatio of a gridded cattle BAC library.* Anim. Genet. **31; 347-353**

23. Heintz, N. 2001 *BAC to the future: The use of BAC transgenic mice for neuroscience research.* Nature Rev. Neur. **2; 861-870**