Linear *E. coli* Vector for Cloning of Difficult DNAs

Ronald Godiska[1], Nikolai Ravin[2], Karen Usdin[3], Attila Karsi[4], David A. Mead[1]

[1] Lucigen Corporation, Middleton, WI 53562 USA
[2] Centre Bioengineering, Russian Academy of Science, Moscow, Russia
[3] NIH, Bethesda, MD 20892 USA
[4] Mississippi State University, MS, USA

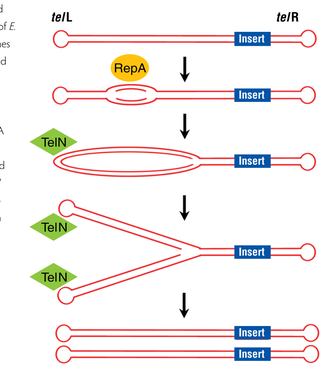
Abstract

We have developed a novel linear vector for unbiased cloning of 0-30 kb inserts in *E. coli*. This vector, termed "pJAZZ®", shows an unprecedented ability to maintain large inserts from very AT-rich genomes. For example, the otherwise difficult-to-clone genome from *Flavobacterium columnare* (70% AT, 3.2 Mb) was sequenced to seven fold coverage using the pJAZZ vector, with only 10 sequencing gaps. The linear vector was able to maintain 20-30 kb fragments from *Lactobacillus helveticus* (65% AT) and 2-4 kb inserts from *Piromyces* (up to 96% AT), which were unclonable in conventional plasmids. The transcription-free, linear pJAZZ vector also minimizes "sequence gaps" caused by secondary structures, as shown by its stable cloning of inverted repeats and di- and tri-nucleotide repeats. Unlike fosmid cloning, construction of large-insert libraries (10-20 kb) in pJAZZ is simple and robust, using standard methods of transformation and plasmid purification. We are evaluating use of a single pJAZZ shotgun library to eliminate the need for multiple libraries, making finishing easier and more cost effective.

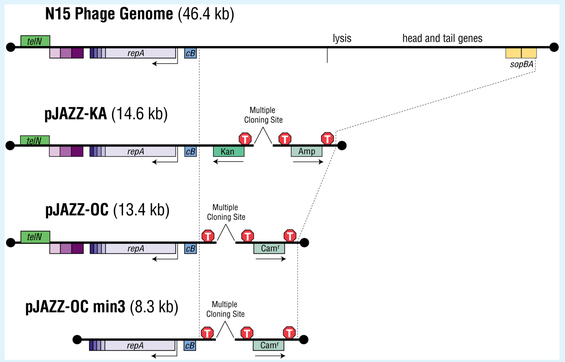
Enhanced stability of inserts in the pJAZZ vector is attributed to both the lack of supercoiling and the lack of transcriptional interference. Torsional strain inherent to supercoiled plasmids can induce localized melting and generate secondary structures, which are substrates for deletion or rearrangement by resolvases and replication enzymes. For example, the instability of tandem repeats and palindromic sequences is presumably due to cleavage of hairpin structures or to replication slippage across the secondary structures. Most conventional plasmid vectors also induce strong transcription and translation of inserted fragments, and they allow transcription from cloned promoters to interfere with plasmid stability. As a result, certain DNA sequences are deleterious or highly unstable, leading to sequence "stacking", clone gaps, or a complete inability to construct libraries, especially from AT-rich genomes or toxic cDNAs.

Replication of the pJAZZ Linear Vector

The pJAZZ vector is derived from the linear phage N15 of *E. coli*. The *rep*A and *tel*N genes of N15 encode replicase and protelomerase. Bi-directional replication is initiated by RepA and is carried out by the host DNA polymerase. TelN cleaves the replicated telomeres and covalently joins the 5' and 3' strands of each free end, re-creating the terminal hairpin loops (*tel*L and *tel*R).



The pJAZZ® Linear Vector



The linear pJAZZ vectors lack the structural and lysis genes of N15.

The ends of the vectors are free to rotate during replication, so cloned inserts are not subject to torsional stress caused by supercoiling. Transcriptional terminators at the cloning site minimize transcriptional interference between the insert and the vector, increasing insert stability. The pJAZZ vectors are low-copy (~5/cell) to further promote stable propagation of inserts, and their copy number can be induced 5-20X for DNA preparation.

Dual selection ensures that recombinant clones contain both arms of the vector. The origin of replication is used to select for the left arm of the pJAZZ vectors, and antibiotic resistance selects for the right arm.

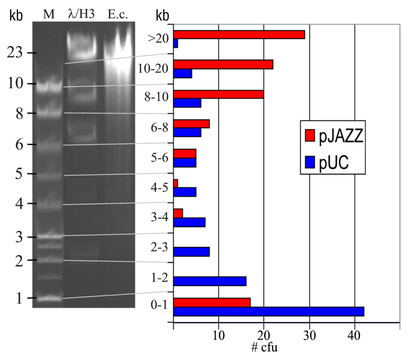
BigEasy®-TSA cells are the *E. coli* host strain for the pJAZZ vectors. This strain contains several required genetic elements of N15: telN encodes protelomerase for replication; sopAB is needed for stable maintenance, and antA regulates copy number.

Inserts are ligated to the vector arms in a standard ligation reaction, which is then electroporated into BigEasy-TSA *E. coli* cells. Linear plasmid DNA is efficiently isolated from recombinants using standard methods, such as alkaline lysis and binding to a silica matrix.

TelN, telomerase; repA, replication protein; cB, replication regulator; Cmr, chloramphenicol resistance; terminators; black circles, closed hairpin telomeres.

Minimal Size Bias

Sheared *E. coli* genomic DNA was cloned into the pJAZZ vector or pUC19 without size selection. As expected, the size distibution of the pUC19 inserts was clearly skewed toward smaller inserts. In contrast, the distribution of the pJAZZ inserts closely matched the input DNA.

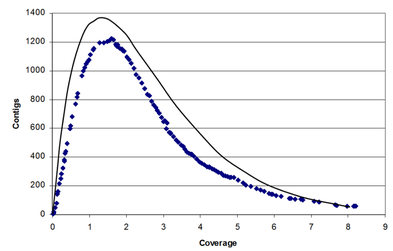


E. coli genomic DNA was sheared to ~8-20 kb. One aliquot was visualized on a gel (E.c.), along with a 1-kb size marker (M) and a HindIII digest of lambda DNA (λ/H3). Another aliquot was cloned without size selection into the pJAZZ vector or pUC19. Clones from each library were randomly isolated for analysis of insert size. The bar graph depicts the number of inserts of each size.

Efficient Library Construction and Assembly

The pJAZZ vector was used to construct genomic libraries of *Flavobacterium columnare* (75% AT, 3.2 Mb). Clones of 2-10 kb were readily obtained, and gave sequence reads of ~800 bp.

Automated assembly closely followed the theoretical Lander Waterman curve, yielding 21 major contigs. Further analysis revealed just 10 sequence gaps. Therefore, this AT-rich genome was completely cloned and nearly completely sequenced without the use of fosmid or BAC libraries.



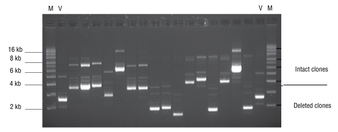
Number of contigs vs. genomic sequence coverage of *Flavobacterium columnare.*

Cloning "Unclonable" DNA

AT-rich genomic DNA

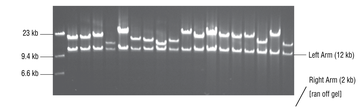
Using conventional circular vectors, AT-rich DNA is often difficult to clone in *E. coli*, producing very few stable, intact clones (Figure A).

A) *Lactobacillus helveticus* (>67% AT) inserts of 1-2 kb are unstable in pUC19. V, vector; M, marker.

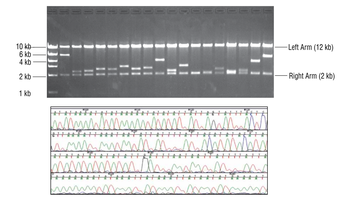


In contrast, large clones of extremely AT-rich DNA are stable in the linear, transcription-free pJAZZ vectors (Figures B, C, D)

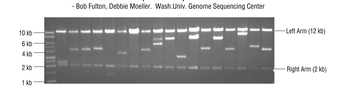
B) *L. helveticus* inserts of 10-20 kb (67% AT) in the pJAZZ vector



C) *Piromyces* inserts of 2-6 kb (85-96% AT), with a sequence trace from one of the clones.



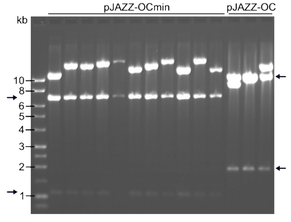
D) *Clostridium* inserts of 6-12 kb (65% AT)
- Bob Fulton, Debbie Moeller. Wash.Univ. Genome Sequencing Center



DNAs were sheared to 2-6 kb (Figures A & C) or 6-20 kb (Figures B & D), end-repaired, size-selected, and cloned into the vector. Plasmid DNA from pJAZZ transformants was cut with NotI to excise the insert and analyzed by gel electrophoresis. Vector bands are 12 kb and 2 kb. AT content of inserts was as high as 96% (Figure D).

pJAZZ-OCmin vector

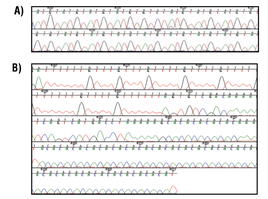
One of the most difficult templates is 10-20 kb fragments of *Tetrahymena* genomic DNA (>70% AT). Libraries were constructed in the pJAZZ-OCmin and pJAZZ-OC vectors. Unexpectedly, the minimal vector yielded approximately 10-fold more colonies, and inserts were all within the expected size range. The DNA yield from the smaller vector was approximately 5-fold lower.



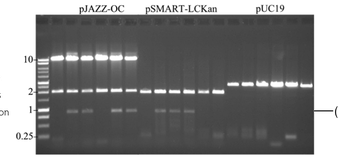
Miniprep DNA from randomly selected colonies was digested with NotI to excise the inserts. Bands corresponding to the vector arms are denoted by arrows. Inserts are 10-20 kb.

Repetitive DNA

Repetitive DNA sequences can be extremely difficult to clone in circular plasmids. For example, mollusk cDNA of 0.3-2 kb was unclonable in all circular plasmids tested. However, the pJAZZ linear vector produced a library of full length clones. Repetitive DNA from two of the clones is shown (Figures A & B).

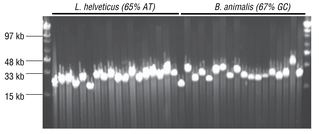


In *E. coli*, clones of 50-100 CGG repeats from the Fragile X locus are highly unstable in circular vectors, and the frequency of deletion is increased by transcription and supercoiling. Corroborating these results, the pJAZZ vector was able to maintain fragments containing 220 copies of the CGG repeat, which has not been achieved with circular vectors.



GC-rich genomic DNA

The GC-rich genome of *B. animalis* (67% GC) was sheared to 6-20 kb, end-repaired, and cloned into the pJAZZ vector. Clones from this library were similar in size and number to those from the AT-rich *L. helveticus* library (see below). Uncut DNA from transformants was 24-40 kb, corresponding to inserts of 10-26 kb.



Summary

- Linear, transcription-free pJAZZ vectors allow cloning of repetitive DNA and large AT-rich or GC-rich DNAs.
  Libraries of 10-20 kb clones of AT-rich genomic DNA were routinely created with the linear transcription-free vector. Likewise, repetitive DNAs were maintained without rearrangement.
- Efficient Genomic Cloning and Sequencing.
  Powerful, unbiased cloning greatly reduced the need for manual finishing of genomic libraries. A 3-Mb microbial genome was assembled without the use of fosmid clones.
- Rapid and simple protocol.
  No vector preparation or special techniques are needed to generate high quality, large-insert libraries of otherwise "unclonable" DNAs.

Lucigen® Advanced Products for Molecular Biology
Lucigen Corporation 2120 W. Greenview Dr., Ste. 9 Middleton, WI 53562 www.lucigen.com