Nucleic Acids Research, 1999, Vol. 27, No. 6 **1539–1546**

High throughput direct end sequencing of BAC clones

Jenny M. Kelley[+], Casey E. Field[§], M. Brook Craven, Diana Bocskai[1], Ung-Jin Kim[1], Steve D. Rounsley[§] and Mark D. Adams*

The Institute for Genomic Research, Rockville, MD 20850, USA and [1]Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA

Received September 30, 1998; Revised and Accepted December 22, 1998 DDBJ/EMBL/GenBank accession nos[¶]

ABSTRACT

Libraries constructed in bacterial artificial chromosome (BAC) vectors have become the choice for clone sets in high throughput genomic sequencing projects primarily because of their high stability. BAC libraries have been proposed as a source for minimally overlapping clones for sequencing large genomic regions, and the use of BAC end sequences (i.e. sequences adjoining the insert sites) has been proposed as a primary means for selecting minimally overlapping clones for sequencing large genomic regions. For this strategy to be effective, high throughput methods for BAC end sequencing of all the clones in deep coverage BAC libraries needed to be developed. Here we describe a low cost, efficient, 96 well procedure for BAC end sequencing. These methods allow us to generate BAC end sequences from human and *Arabidopsis* libraries with an average read length of >450 bases and with a single pass sequencing average accuracy of >98%. Application of BAC end sequences in genomic sequencing is discussed.

INTRODUCTION

A cloning system based on the *Escherichia coli* F factor, referred to as bacterial artificial chromosomes (BACs), was initially described by Shizuya and colleagues (1). Libraries constructed in BAC vectors have become the choice for high throughput genomic sequencing projects because of their higher stability as compared with their YAC or cosmid counterparts (1–5). The use of BAC end sequences (i.e. sequences adjacent to the insert sites) has been proposed as a primary means for selecting minimally overlapping clones for sequencing large genomic regions (6–8). A necessary prerequisite of this is a collection of end sequences from all the clones in deep coverage BAC libraries. This is now being pursued for both the human and *Arabidopsis* genomes (1,2,9).

The *E.coli* F-factor replicon offers strict copy number control which limits the number of BACs to one to two copies per cell. This minimizes the opportunity for DNA rearrangements or deletions and also minimizes any toxic effects of the cloning in *E.coli* cells. However, the low abundance of BAC DNA poses a challenge for high throughput direct sequencing of the BAC clone ends due to the difficulty in obtaining sufficient quantities of high quality template from standard minipreps. There has been some success with obtaining BAC clone end sequences by inverse PCR or end rescue PCR methods (10,11), but these involved multiple complex steps. Previous successful attempts at direct BAC end sequencing have relied either on large volume preps (3,4,7,12–14) or on very expensive automation (Autogen 740, Integrated Separation Systems) (5,7,8,12). Our aim was to develop a 96 well template purification procedure that yielded enough good quality BAC DNA for one or two sequencing reactions and a DNA fingerprint. Using a basic alkaline lysis method (15), combined with a protein-binding reagent, Procipitate™ (LigoChem) (16,17), we were able to develop an alkaline lysis filter plate prep (TIGR Procipitate™ Filter Method or TPF) that is processed completely in a 96 well format.

The sequencing reactions are also processed in a 96 well format, including the removal of excess dyes before loading onto ABD 377XL automated sequencers, so that the entire direct end sequencing of BAC clones can be performed in an efficient high throughput manner.

MATERIALS AND METHODS

BAC clones were inoculated into triplicate 2 ml 96 well blocks containing 1.3 ml Luria Broth (LB) (15) plus the appropriate antibiotic (either 12.5 μg/ml chloramphenicol or 25 μg/ml kanamycin). The blocks were covered with breathable AirPore tape Strips (Qiagen) or plastic lids and incubated for ~18–20 h at 37°C at 325 r.p.m. in a shaking incubator. The BAC DNA purification method is described in Figure 1. Qiagen reagents R1, R2 and R3 are routinely used with the RNase concentrations listed below for Solution I. We have also used an alternate set of alkaline lysis solutions as follows (15). Solution I: 50 mM glucose, 25 mM Tris–HCl (pH 8.0), 10 mM EDTA (pH 8.0). Solution I may be prepared in batches, autoclaved and stored at 4°C for up to 6 months. Before use, RNase A was added to a final concentration

*To whom correspondence should be addressed at present address: Celera Genomics Corp., 45 West Gude Drive, Rockville, MD 20852, USA. Tel: +1 240 453 3700; Fax: +1 240 453 4000; Email: adamsmd@celera.com

Present addresses: [+]Laboratory of Population Genetics, DCEG, NCI, NIH, Bethesda, MD 20892, USA and [§]Cereon Genomics LLC, 270 Albany Street, Cambridge, MA 02139, USA

[¶]All sequence data have been deposited in GenBank

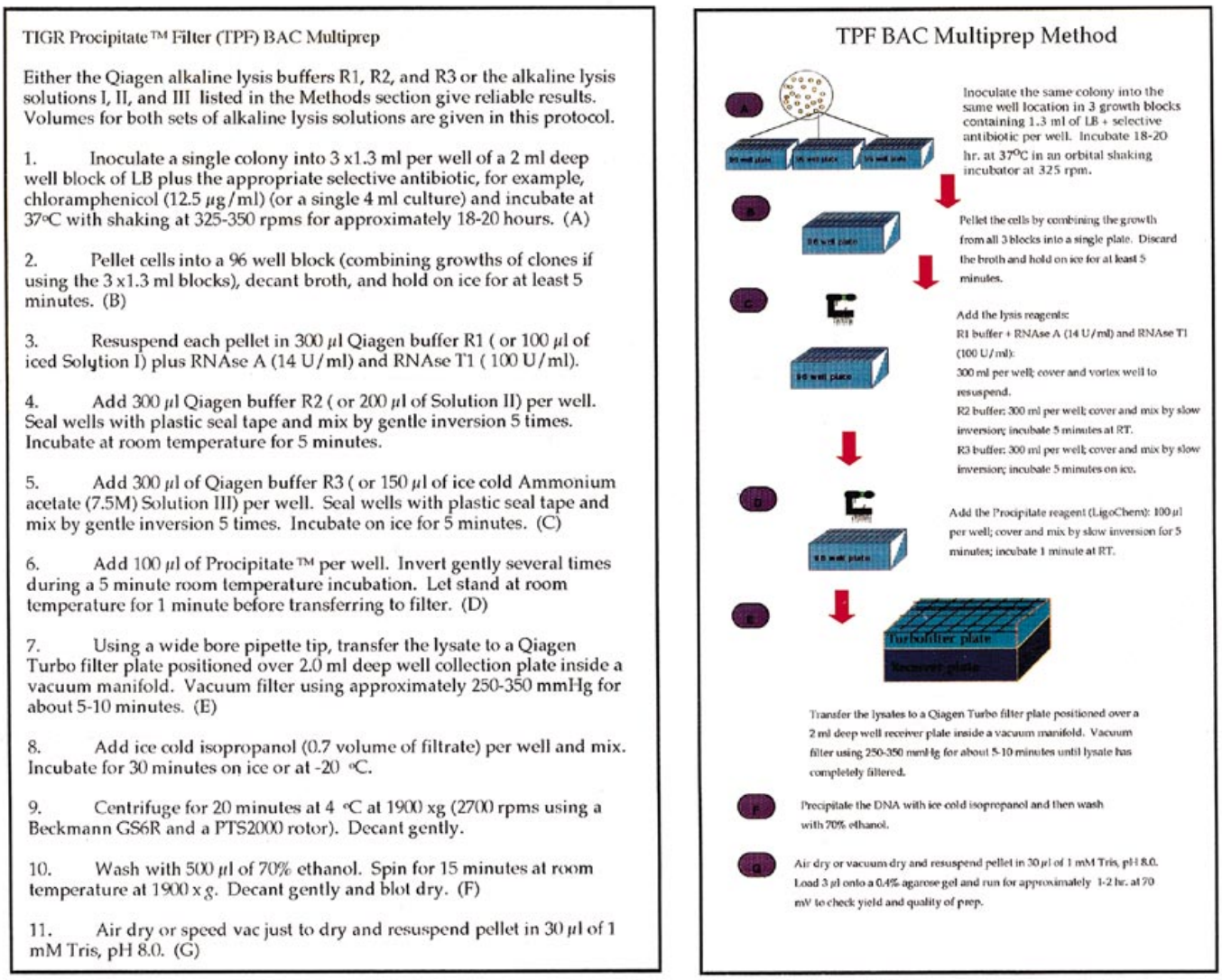

Figure 1. Procedure for the TIGR Procipitate™ Filterplate prep.

of 14 U/ml and RNase T_1 to a final concentration of 100 U/ml. Solution II: 0.2 N NaOH (freshly diluted from a 10 N stock), 1% SDS (diluted from 10% stock). It is important to prepare this solution fresh daily. Solution III: 7.5 M ammonium acetate (may be stored at room temperature for up to 1 month).

After the DNA had been resuspended and checked by agarose gel (Fig. 2), dye terminator sequencing reactions were run as described in Figure 3, and post-reaction removal of excess dyes was performed using a Sephadex G50 mini-column filter plate method as described by the Stanford DNA Sequence & Technology Center (http://sequence-www.stanford.edu/protocols/dye-terms.html). The reactions were then analyzed with an ABD 377XL automated sequencer using 48 lanes and either 5.0% premixed acrylamide gel mix (FMC) and 34 cm well-to-read plates or 4.5% premixed acrylamide gel mix (FMC) and 48 cm well-to-read plates. One pBluescript control reaction per 48 samples was included to aid in quality control and sample tracking. In addition, 0.5 µl of a 1:10 dilution of the Genescan Rox500 Lane Standard (ABD/Perkin-Elmer) was loaded in lanes 0 and 49 as lane position markers.

BAC clones from the human libraries are available from Research Genetics, Inc. (Huntsville, AL) (http://www.resgen.com), and BAC clones from the *Arabidopsis* libraries are available from the Arabidopsis Biological Resource Center, ABRC (Ohio State University, OH) (http://www.aims.cps.msu.edu/aims/).

RESULTS AND DISCUSSION

Several steps in this BAC DNA prep method have been identified as critical to the success of the process. First, the growth blocks must be shaken at a minimum of 325 r.p.m. and the incubation times must be adjusted for each cell line and vector combination to insure that late log or early stationary phase has been reached without cell lysis and loss of the BAC plasmid into the media. Although rich media (2× YT broth and MacConnell) substituted for LB increased cell numbers, the increased amount of cellular

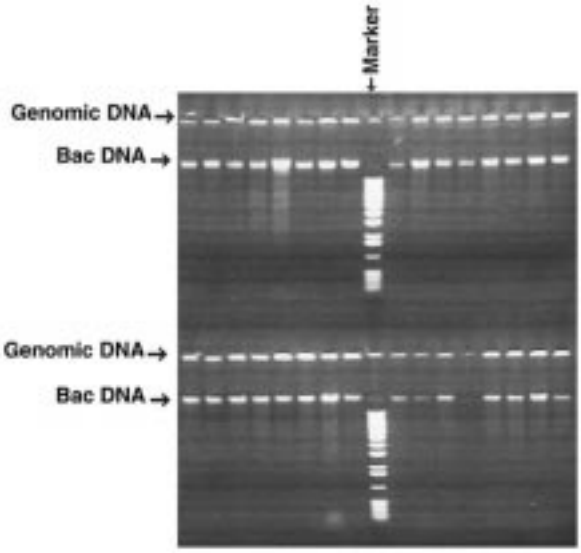

Figure 2. Agarose gel of TPF prep. Typical agarose gel electrophoresis results of BAC purification using the TPF prep as described in the text and showing the consistent well-to-well reproducibility and sufficient DNA present in ~95% of the wells. BAC clones are from a human BAC library. Lanes marked M are 1 kb molecular weight marker ranging from 500 bp to 12 kb (Life Technologies). The sample lanes contain 3 µl DNA (10% of the volume of the prep) and 1 µl electrophoresis loading buffer (OWL Separation Systems) per well. The gel was 0.4% agarose (Gibco BRL) with 1× TAE buffer run at 70 mV for 1 h.

TIGR Direct Sequencing of BAC Clones
Using ABD FS+ or BigDye Terminator Chemistry

Using approximately 0.5-1.0 µg DNA per reaction or approximately 10-20 µl* of the template prepared by our TIGR BAC Clone Multiprep (TPF): FS+ Dye Terminator reactions in 40 µl total volume or Big Dye DyeTerminator sequencing reactions in 30 µl total volume with 4 mM extra $MgCl_2$ and 10-20 pmoles of the T7 or Sp6 primers or of the M13-21 or M13R primers, depending on the BAC vector used.

Per sequencing reaction:	BigDye Term	FS+ Term
terminator mix (+ 4 mM extra $MgCl_2$)	12 µl	16 µl
primer @ 10-20 total pmoles/rxn	x µl	x µl
BAC DNA (approx. 0.5-1 µg/rxn)	10 µl*	20 µl*
sterile H_2O to bring up volume	qs	qs
total reaction volume	30 µl	40 µl

The following cycling protocol is used on a Perking Elmer 9600 thermocycler or a MJResearch PT-100 thermocycler.

96 °C	2 min	initial denaturation
96 °C	10 sec	
50 °C	5 sec	x30 cycles
60 °C	4 min	
4 °C	hold	

Following the DyeTerminator sequencing reactions: excess dyes are removed with a Sephadex G50 (fine) Slurry 96 well protocol. The samples are dried in a speed vacuum and stored covered at 4 °C until loading onto a ABD 377 Automated DNA Sequencer using a 5.0% LongRanger (FMC) gel according to manufacturer's recommendation except that the entire reaction sample is loaded by resuspending the sample in 2.0 µl loading buffer and loading 1.8-2.0 µl per well for 48 lanes.

Figure 3. Procedure for ABD FS+ DyeTerminator or ABD BigDye Terminator reactions.

proteins and RNA in the prep yielded less satisfactory sequencing results and we recommend using LB for the growth media. Several growth times were tested for yield of DNA and purity for sequencing for each host/vector combination. The next critical step is the thorough resuspension of the cell pellet in buffer R1 (or Solution I). Complete resuspension is necessary for effective cell lysis and release of maximum amount of BAC DNA at the next step. Another important factor was increasing the concentration of RNase relative to plasmid minipreps due to the larger starting culture volume. The fourth critical step is the addition of the Procipitate™ reagent to bind the additional protein, cell debris and membrane-associated chromosomal DNA before filtration. Because of the low copy number and the triple volume of cells in the starting culture, there is a much higher ratio of proteins and cellular debris that must be removed to provide the clean DNA required for fluorescent dye terminator reactions on large inserts. The Procipitate™ aided in the removal of proteins and the proper flow through the filter plates. The last critical factor was the type and quality of the filter plates. Two alternative methods for filtering the lysates, vacuum and centrifugation, were tested. Although centrifugation works, our method of choice is vacuum filtration because it is faster, more amenable to automation and produces more consistent results. We tested several filtration plates from a number of manufacturers and found the Qiagen TurboFilter™ plate to be the most reliable in terms of flow through and consistency from well to well. Experiments on the stability of the BAC DNA isolated by TPF have shown templates to be stable for at least 12 weeks when stored at 4°C (Fig. 4A), and good sequencing results are obtained even on more difficult templates as shown in Figures 4 and 5.

The most reliable method for determining the success of a BAC multiprep has been by running ~10% of the final sample volume on a 0.4% agarose gel at 80 mV for 2 h, staining with ethidium bromide and viewing using a Biorad GelDoc system. A typical gel image (Fig. 2) shows that the prep success rate is >90%.

For the dye terminator sequencing reactions, ABD dichlororhodamine terminator or ABD BigDye terminator kits (FS+) are used with the modifications indicated in Figure 3. We found that with the dichlororhodamine chemistry, a double volume reaction mix is necessary for consistent high quality sequencing of BACs, possibly due to the fewer number of available priming sites or non-specific polymerase binding. The total reaction volume of 40 µl includes 16 µl of FS+ rhodamine dye terminator mix and 20 µl of BAC template containing ~500 ng DNA. Recently, we began sequencing both ends from a single 3.9 ml well prep using ABD BigDye Terminator chemistry. This new sequencing chemistry enables us to use 30 µl total reaction volume which contains only 15 µl of BAC DNA and 12 µl of BigDye dye terminator mix. We found that both of these dye terminator chemistries do an excellent job of sequencing difficult regions in the template DNA (Figs 4 and 5).

Magnesium chloride is essential for optimum activity of the *Taq* polymerase. We found improved signal to noise ratio and consistency of reactions when the overall concentration of $MgCl_2$

in the reaction premix is increased. Because a small amount of Procipitate™ leaks through the filter and sequesters some of the Mg^{2+} in the reaction mix, $MgCl_2$ is routinely added to the reaction premix to a final nominal concentration of 4 mM.

Dye terminator reactions for plasmid vectors usually require ~3.2 pmol of primer per reaction. For an equivalent amount of BAC DNA, there are 50–100 times fewer priming sites available. In addition, the BAC target contains an increased number of imperfect binding sites which may titrate away some primer molecules. We found that using 12–20 pmol of primer in a reaction prevents primer concentration from being a limiting factor. Also, by increasing the initial denaturing temperature and denaturing time to denature the BAC, the reaction yield is better.

Due to the larger amount of dyes in the reaction mixtures, the excess dyes cannot be removed completely using a simple ethanol precipitation commonly used in plasmid dye primer and terminator reactions. It is necessary to use a 96 well Sephadex G50 (Sigma) mini-column method that efficiently removes residual dye labeled dideoxynucleotides from the reaction before analysis on ABD 377XL sequencers.

There are several advantages of the TPF prep. The entire process is performed in a 96 well format, from clone storage and growth through DNA purification, isopropanol precipitation and final resuspension. Sequencing reactions are also processed in a 96 well format, including the removal of excess dyes before loading onto the sequencers. Advantages of using the dye terminator chemistry rather than the dye primer chemistry for direct end sequencing of BAC clones include higher throughput setup and thermocycler use for a single tube reaction (as opposed to a four tube reaction) and the ability to use custom primers for more flexibility with primer selection for various BAC vectors and to permit primer walking directly from BAC clones. Other advantages are better resolution of GC rich areas with the elimination of hardstops commonly seen in direct BAC sequencing (Fig. 4B) and the flexibility with primer concentration that can be adjusted to account for the small number of priming sites relative to the total DNA in BACs. Both the prep and reaction high throughput methods are amenable to future automation. In addition, there is no requirement for equipment other than common laboratory centrifuges and vacuum manifolds, and the relatively low cost per prep is <$1.40. To date, we have generated 106 916 human BAC end sequences and 17 263 *Arabidopsis* BAC end sequences. Our end sequencing success rate for both projects is ~80%. This percentage reflects both the template preparation and sequencing reactions.

As may be expected from the lower yield of DNA and the larger template size, BAC end sequencing reactions are not as robust as those on plasmid DNA templates. We have examined several parameters of quality of the BAC end sequences. BAC end sequence reads are processed by the same vector- and quality-trimming routines as our plasmid sequence reads. A key feature of the processing is base-calling using phred (18,19), which reports quality values for each base-call. A phred quality score of ≥20 indicates that there is less than 1:100 chance that the base-call is incorrect. These are called high quality bases or HQ bases. The trimming software was calibrated to result in sequence data that is accurate to ≥97% in each 50 bp sliding window. BAC end sequences have, on average, 100 fewer HQ bases than plasmid sequences trimmed by the same algorithm. The average quality-trimmed read lengths for the *Arabidopsis* and human BAC end-sequences are 533 and 515 bp, respectively.

Use of BAC end sequences for selecting clones to sequence

We have employed BAC end sequencing as a primary means of constructing sequence-ready maps for human chromosome 16 and for *Arabidopsis thaliana* chromosome 2. For the *Arabidopsis* project, BACs were selected by screening the TAMU *Arabidopsis* BAC library (9) with YAC clones from the map constructed by Howard Goodman (20; http://weeds.mgh.harvard.edu/goodman/bac_bin/index.html). For the human chromosome project, we began by screening the CalTech BAC library (1) with 110 STSs from the LANL physical map (21) located on the p arm between D16S2782 and the centromere. To date, over 2800 BAC clones have been identified and end-sequenced from this ~20 Mb region. The average resolution of the LANL STS map is ~180 kb between markers, so it was expected that few, if any, contigs of BACs would be built from the primary screen based on STS-content and fingerprinting (22). We therefore selected 36 initial BACs for sequencing spread throughout the region. End-probes have been developed from each of the seed BACs and used to rescreen the library to identify additional clones. From the 36 seed BACs, 33 new clones have been selected for sequencing based on BAC end sequence indications of overlap. In each case, the clone selected based on the BAC end sequence match was a true overlap clone. Table 1 summarizes the BACs obtained for the *Arabidopsis* and human projects.

Table 1. Summary of BAC sequencing

	Arabidopsis	Human
Size of region	14 Mb[a]	20 Mb
No. of BACs identified	N/A	3008
Total BAC end sequences	41 496[b]	157 542
Chr.-specific BAC end sequences	N/A	4328
No. of seed BACs	21	36
No. of overlapping BACs	71[c]	33
Average size of overlaps	9.2 kb	12 kb

[a]Size of Chr2.
[b]Combined total from TIGR (16 392), University of Pennsylvania (9230) and Genoscope (15 874).
[c]Total number of overlaps for the *Arabidopsis* Chr2 project identified using entire BAC end database including data from TIGR, UPenn and Genoscope.

The BAC end sequence data are stored both in a relational database structure and as searchable databases. In order to minimize the impact of repetitive elements in the human genome, the human BAC end sequence database is searched in two steps. The first step is a BLAST search of a database of BAC end sequences which have been masked for repetitive elements by the repeatmasker program (23). Liberal criteria are used to select sequences to be searched in the second phase. In the second phase search, a mini-database is constructed from unmasked copies of all BAC end sequences from the BLAST search which may potentially match the query sequence. FASTA is used to compare the query to the mini-database of unmasked sequences. *Arabidopsis* BAC end sequences are searched using the same two-step method, except unmasked sequences are used for both the BLAST and FASTA stages. A WWW interface has been developed which allows complete BAC sequences (or other sequences) to be used as queries against the BAC end sequence database with the results returned in both graphical and tabular formats (for human

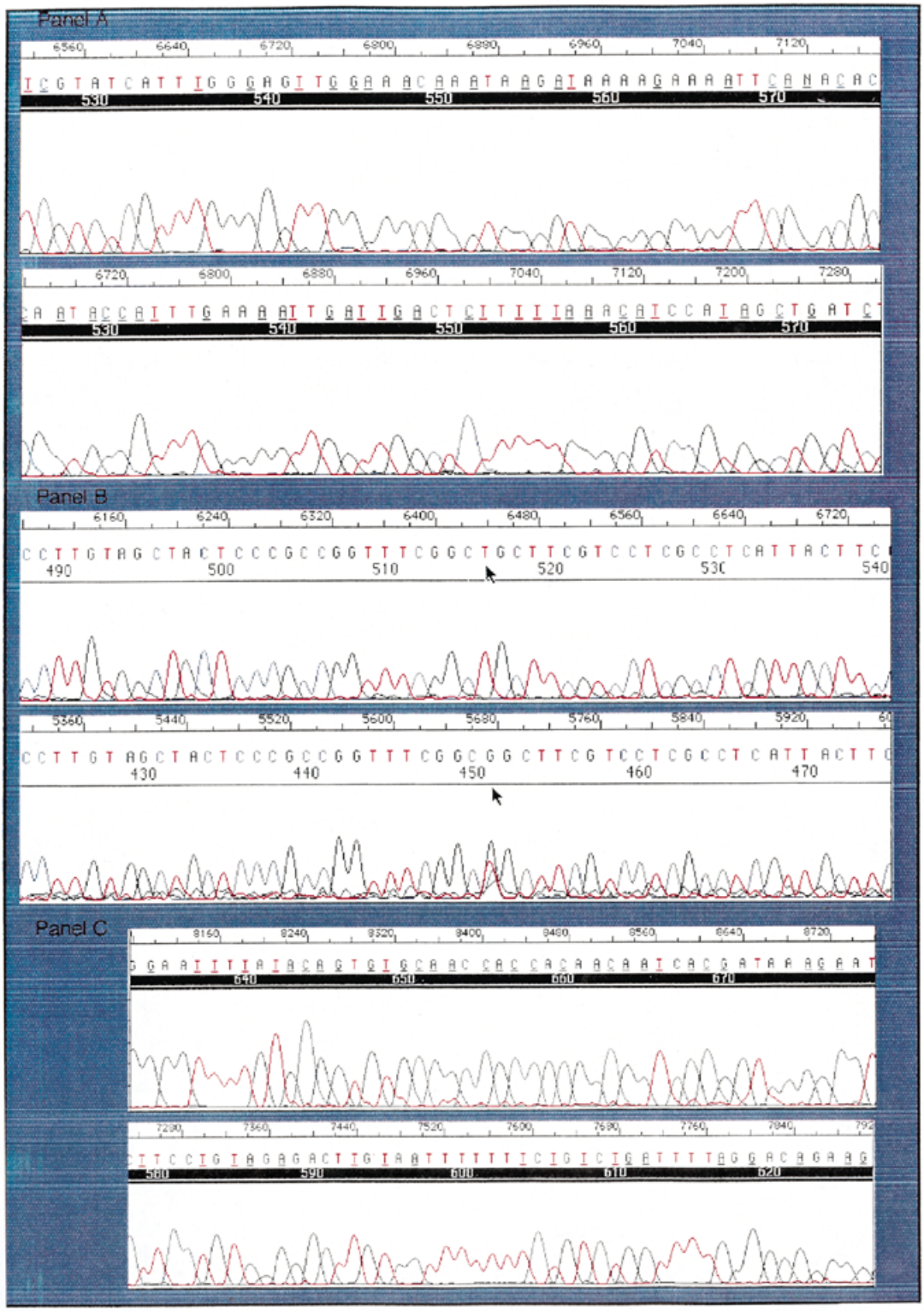

Figure 4. Direct BAC end sequencing electropherograms. (**A**) shows data from a sequencing reaction using a template after 3 months storage at 4°C. The template contained *Arabidopsis* DNA reacted with FS+ DyeTerminator chemistry using the M13 -21 universal forward primer as described in Figure 3. There is good signal to noise ratio and clean data beyond 500 bases. (**B**) shows a comparison of the same template containing human DNA reacted with the BigDye DyeTerminator sequencing chemistry (top) and the BigDye DyePrimer sequencing chemistry (bottom) as described in Figure 3. There is low background and good resolution in the top panel compared with the increased background and base miscall in the lower panel (indicated by arrow). (**C**) shows two examples of long read data. The top sequence reads >650 bases and was generated from a template containing human DNA reacted with the BigDye DyeTerminator sequencing chemistry. The bottom sequence reads >600 bases and was generated from another template containing human DNA reacted with the FS+ DyeTerminator sequencing chemistry.

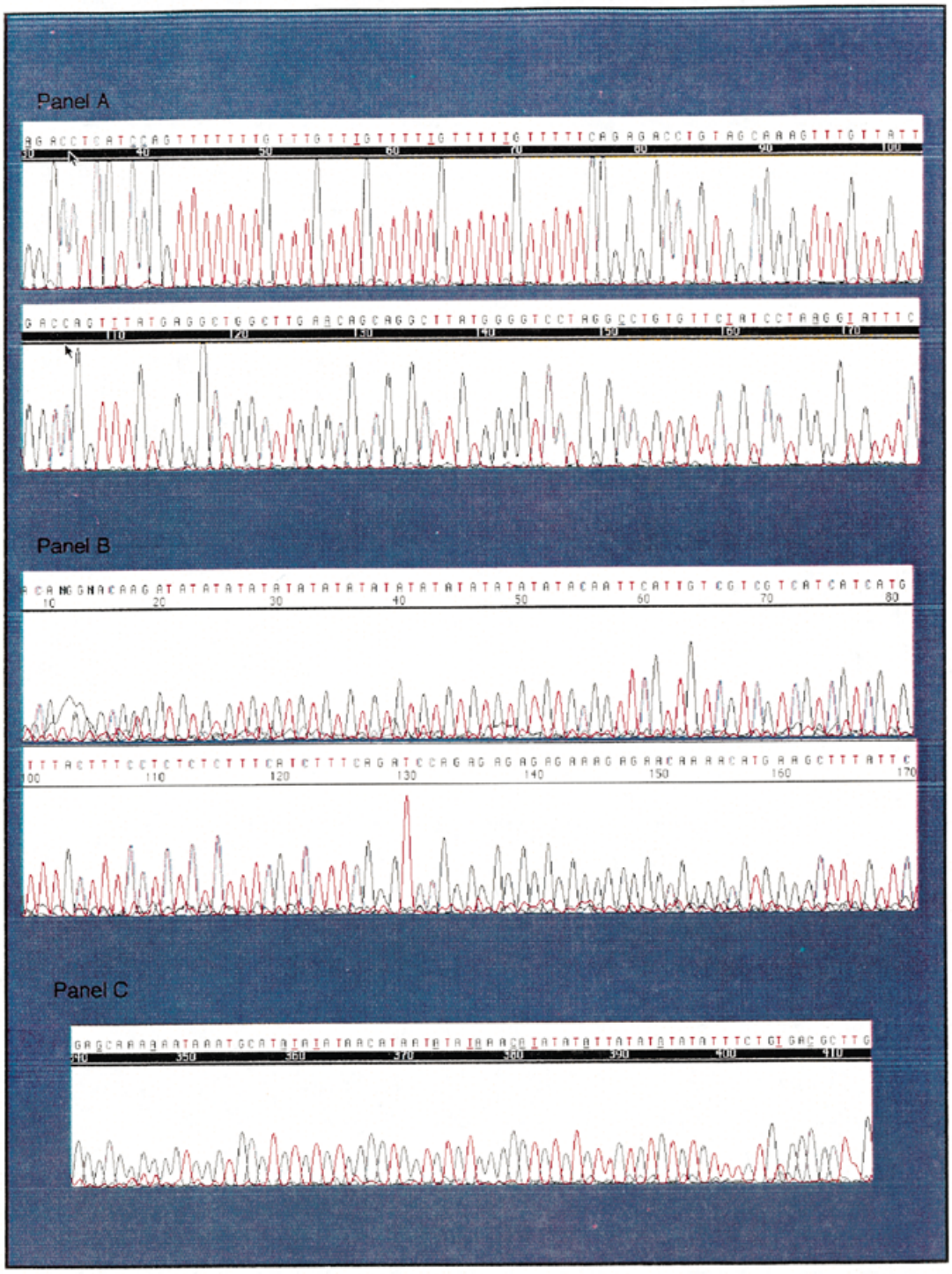

Figure 5. Sequencing data from templates with difficult repeat areas. (**A**) Sequencing data from a template containing human DNA reacted with FS+ DyeTerminator chemistry using the Sp6 primer as described in Figure 3. This clone has repeat areas at the beginning of the insert site. (**B**) Sequencing data from a template containing *Arabidopsis* DNA reacted with FS+ DyeTerminator chemistry using the M13 -21 forward primer as described in Figure 3. This clone has an AT repeat area at the beginning of the insert. (**C**) Sequencing data from a template containing human DNA reacted with BigDye DyeTerminator chemistry using the T7 primer as described in Figure 3. This clone has initial repeat areas and then more repeats at bases 300–400.

BACs: http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html ; for *Arabidopsis* BACs: http://www.tigr.org/tdb/at/atgenome /bac_end_search/bac_end_search.html).

An example of the alignments of a contig and several possible BACs to choose for contig extension is shown in Figure 6.

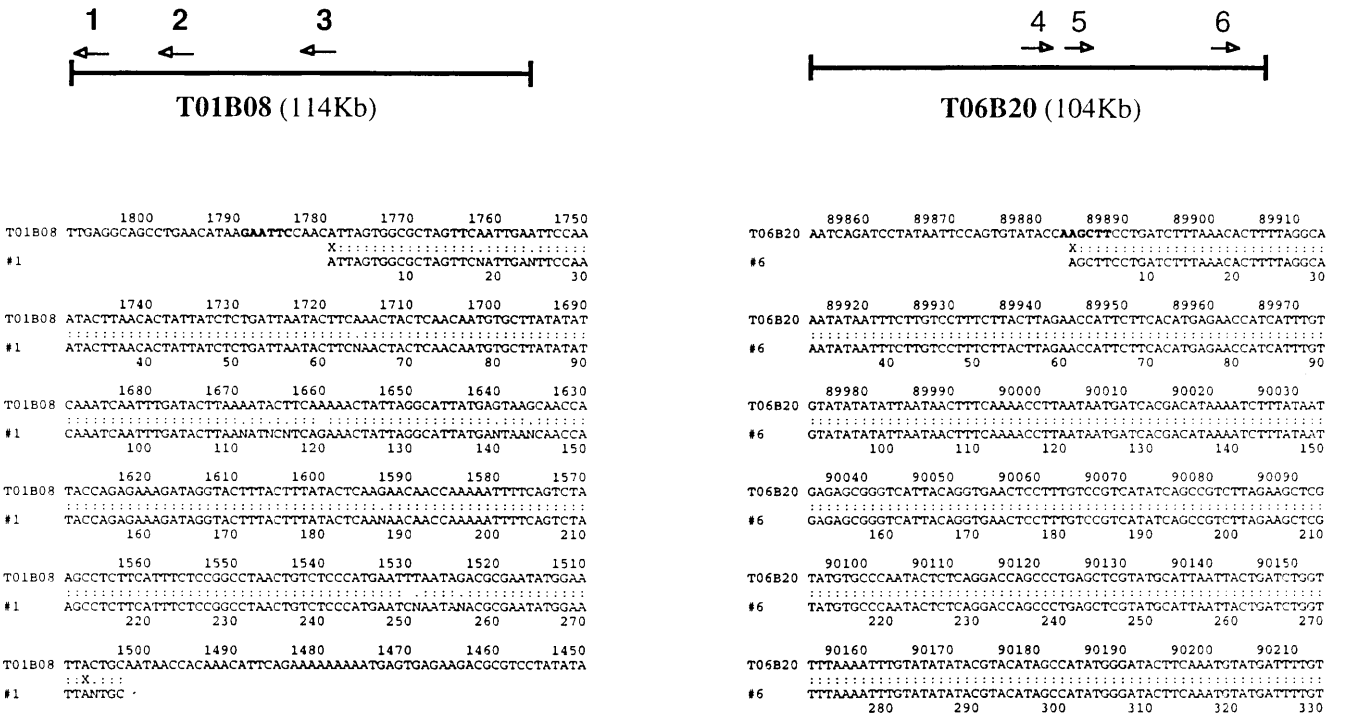

Figure 6. An example of BAC end sequence matches used to select clones for continued genome sequencing. Gapped alignments between BAC end sequences and genomic sequence identify possible BAC clones to extend sequencing contigs with a minimum amount of sequencing redundancy. The schematic above shows the relative position of BAC end sequence matches to completed BAC sequences. The alignments of the *Arabidopsis* BAC clones T01B08 and T06B20 against the end sequences numbered 1 and 6, respectively, are shown to illustrate the quality of the sequence data obtained from direct BAC end sequencing using the described methods. The sequence indicated in bold represents the restriction enzyme cleavage sites in the genome that were used when BAC clones 1 and 6 were constructed.

As noted above, each human BAC end sequence has been analyzed by the repeatmasker program. Results of the repeat content analysis are presented in Table 2. The overall fraction of DNA in each repeat class is as expected (23), providing some indication that the BAC library and the end-sequence dataset are representative of the genome. Interestingly, >40% of the sequences are completely free of repetitive DNA and are thus likely to be useful as a source of new STS markers. Furthermore, >90% of the BAC end sequences contain at least 20 bp of unique sequence.

Table 2. Repetitive DNA content of human BAC end sequences

Repeat type	% sequences	% bases
Simple sequence	12	1
Alu	16	6
L1	22	14
LTR	1	1
MER/MIR	14	5
MST/THE	3	1
Other	9	4
Total	58[a]	32

[a]Percentages do not add up to 100% because some sequences have more than one repeat type.

In summary, we are able to generate a large number of end sequences from random human and *Arabidopsis* BAC libraries, and these data are very useful in providing information that identifies new BAC clones to add to the genomic sequencing efforts. BAC end sequences also can be made into STSs. The Stanford Genome Center is currently using BAC end sequences as STS markers on the whole genome radiation hybrid map, thus providing a link between chromosome location and a physical DNA clone (E.Beasley, personal communication).

ACKNOWLEDGEMENTS

The authors would like to thank Cheryl Heiner, Cecelia Boysen, Greg Mahairas, Lee Hood, Jean-Francois Tomb, Robert Fleischmann, Robert Blakesley, Jeannine Gocayne and Andreas Duesterhoeft for their helpful technical suggestions and comments. We would also like to thank Rachel Tripi, Samir Kaul and Steven Bass for their technical help and Lily Fu, Michael Heaney and John Scott for their database support. Funding support for this project was received through grants to J. Craig Venter from NSF (DBI-9632085) and DOE (DE-FG02-96ER20249) for the Arabidopsis BAC end sequencing and to M.D.A. from DOE (DE-FC02-97ER62500) for the human BAC end sequencing.

REFERENCES

1 Shizuya,H., Birren,B., Kim,U., Mancino,V., Slepak,T., Tachiiri,Y. and Simon,M.I. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 8794–8797.
2 Goodman,H.M., Ecker,J.R. and Dean,C. (1995) *Proc. Natl Acad. Sci. USA*, **92**, 10831–10835.
3 Kim,U., Shizuya,H., Birren,B., Slepak,T., de Jong,P. and Simon,M.I. (1994) *Genomics*, **22**, 336–339.

4 Kim,U., Birren,B., Mancino,V., Slepak,T., Boysen,C., Kang,H., Simon,M.I. and Shizuya,H. (1996) *Genomics*, **34**, 213–218.
5 Ross,M., Dewar,K., Kim,U.-J. and Dunham,I. (1998) *Genome Analysis: A Laboratory Manual Series, Vol. 3: Cloning Systems*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, in press.
6 Venter,J.C., Smith,H.O. and Hood,L. (1996) *Nature*, **381**, 364–366.
7 Boysen,C., Simon,M.I. and Hood,L. (1997) *Genome Res*., **7**, 330–338.
8 Boysen,C., Simon,M.I. and Hood,L. (1997) *BioTechniques*, **23**, 978–980.
9 Choi,S., Creelman,R.A., Mullet,J.E. and Wing,R.A. (1996) *Weeds World*, **2**, 17–20. (http://genome-www.stanford.edu/Arabidopsis/ww/Vol2/choi.html)
10 Cai,L., Taylor,J.F., Wing,R.A., Gallagher,D.S., Woo,S.-S. and Davis,S.K. (1995) *Genomics*, **29**, 413–425.
11 Woo,S.-S., Jiang,J., Gill,B.S., Paterson,A.H. and Wing,R.A. (1994) *Nucleic Acids Res*., **22**, 4922–4931.
12 Hubert,R.S., Mitchell,S., Chen,X.-N., Ekmekji,K., Gadomski,C., Sun,Z., Noya,D., Kim,U.-J., Chen,C., Shizuya,H., Simon,M., de Jong,P. and Korenberg,J.R. (1997) *Genomics*, **41**, 218–226.
13 Marra,M., Weinstock,L.A. and Mardis,E.R. (1996) *Genomic Methods*, **6**, 1118–1122.
14 Zimmer,R. and Verinder Bibbins,A.M. (1997) *Genomics*, **42**, 217–226.
15 Maniatis,T., Fritsch,E.F. and Sambrook,J. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
16 Huang,G.M., Wang,K., Kuo,C., Paeper,B. and Hood,L. (1994) *Biochemistry*, **223**, 35–38.
17 Kuruc,M. and Krupey,J. (1992) *Am. Biotechnol. Lab*., **10**, 12–14.
18 Green,P., Ewing,B., Hillier,L. and Wendl,MC. (1998) *Genome Res*., **8**, 175–185.
19 Green,P. and Ewing,B. (1998) *Genome Res*., **8**, 186–194.
20 Zachgo,E.A., Wang,M.L., Dewdney,J., Bouchez,D., Camilleri,C., Belmonte,S., Huang,L., Dolan,M. and Goodman,H.M. (1996) *Genome Res*., **6**, 19–25.
21 Doggett,N.A., Goodwin,L.A., Tesmer,J.G., Meincke,L.J., Bruce,D.C., Clark,L.M., Altherr,M.R., Ford,A.A., Chi,H.C., Marrone,B.L. *et al.* (1995) *Nature*, **377**, 335–365.
22 Marra,M.A., Kucaba,T.A., Dietrich,N.L., Green,E.D., Brownstein,B., Wilson,R.K., McDonald,K.M., Hillier,L.W., McPherson,J.D. and Waterston,R.H. (1997) *Genome Res*., **7**, 1072–1084.
23 Smit,A.F. (1996) *Curr. Opin. Genet. Dev*., **6**, 743–748.