End sequence determination from large insert clones using energy transfer fluorescent primers.

M Marra, L A Weinstock and E R Mardis

Genome Res. 1996 6: 1118-1122
Access the most recent version at doi:10.1101/gr.6.11.1118

GENOME METHODS

End Sequence Determination from Large Insert Clones Using Energy Transfer Fluorescent Primers

Marco Marra, Lori A. Weinstock, and Elaine R. Mardis[1]

Genome Sequencing Center, St. Louis, Missouri 63108

Genome mapping strategies depend heavily on confirmatory data of several types to establish overlaps between contiguous stretches of cloned DNA derived from genomic regions. One type of ancillary data that can contribute to establishing these overlaps is DNA sequence data derived from the ends of large (>30 kb) inserts in genomic clones. This type of data can be difficult to obtain routinely, because large clones are often unstable and microgram quantities of highly purified DNA are required in each sequencing reaction to obtain sufficient signal for accurate base calling and maximum read length. Recently, we have been experimenting with methods to consistently obtain up to 800 bases of high-quality sequence data from the ends of large insert clones using ThermoSequenase DNA polymerase and Energy Transfer fluorescent primers. Our experimental approach and results, described in this paper, indicate that routinely obtaining high-quality sequence data from the ends of large insert genomic clones is feasible. Such data can contribute to the assessment of common regions between large insert clones, to the establishment of conservation of synteny between closely related species, and to the detection of additional contiguous clones.

Conventional bottom-up approaches to genomic physical mapping typically employ fragment size similarities and other gross sequence similarity assays [such as hybridization and sequence-tagged site (STS) content] to determine overlaps between genomic subclones. While such information is sufficient to produce high-resolution sequence-ready maps (Coulson et al. 1986), it has been recognized that the ability to obtain DNA sequence information from the ends of large insert clones could facilitate map generation by absolute confirmation of overlapping sequences (Chen et al. 1993). End sequence data also can provide oligonucleotide sequences to directly probe genomic libraries or to use as PCR primers in efforts to identify new clones within an area of interest. Until recently, however, obtaining end sequence data has proven largely irreproducible due primarily to the difficulty of achieving analyzable fluorescent signal from large clones without requiring large quantities of highly purified template DNA. The development of energy transfer (ET) primers (Ju et al. 1995) recently has provided a means to generate enhanced fluorescent signal from labeled DNA sequencing reaction products using reduced amounts of input DNA in large clone sequencing reactions. Furthermore, the use of a mutant thermostable polymerase, ThermoSequenase DNA polymerase (Amersham), has enhanced the overall data quality and base-calling accuracy of DNA sequencing reactions by providing increased uniformity of peak heights (Tabor and Richardson 1995). We have investigated the utility of large clone end sequencing with several bacterial artificial chromosome (BAC) clones from the CCM-1 region of human chromosome 7, and here demonstrate that high-quality sequence data can be obtained using both universal and reverse ET primers in reactions catalyzed by ThermoSequenase DNA polymerase.

[1]**Corresponding author.**
E-MAIL emardis@watson.wustl.edu; FAX 314-286-1810.

RESULTS AND DISCUSSION

One drawback of conventional fluorescent reaction chemistries in large clone sequencing has been the need for large amounts of highly purified DNA template to produce sufficient amounts of labeled sequencing fragments and therefore detectable fluorescent signal and peak resolution (Voss et al. 1990). The introduction of energy transfer primers (Ju et al. 1995) recently has provided an alternative labeling strategy that enables sequencing fragments to be detected more

6:1118–1122 ©1996 by Cold Spring Harbor Laboratory Press ISSN 1054-9803/96 $5.00

readily by the fluorescent sequencing instrument, as as result of their enhanced quantum efficiency over conventional, singly labeled dye primers. To illustrate this enhancement, Figure 1 shows a comparison of sequencing results obtained using the same BAC clone preparation (O39A08) with either ET (panel A) or standard, single fluor dye primers (panel B). A comparison of these data indicates that the overall signal strength for the ET primed reaction products is higher and that the trace data quality for the ET primed reaction is much higher in the region shown. Base calling using the phred algorithm of P. Green (pers. comm.) and its output quality measures were generated for the entire trace data from the two Figure 1 sequences, as shown in Table 1. The base-calling program phred computes a probability (p) of an error for each base call at each position. The probability is then converted to a quality value using the transformation: $q = -10 \log_{10}(p)$. Thus, a quality of 30 corresponds to an error probability of 1/1000, a quality of 40 to an error probability of 1/10000, and so on. Finally, these quality values are averaged for specific regions of the data, given in Table 1 for 100 base increments. The average quality values associated with the conventional, single fluor dye primer data decline more rapidly with increasing base number than those generated from the ET primer data. Data quality measures for both sequences converge near the end of the data, which is probably a reflection of gel resolution limits. Because both reactions were catalyzed by ThermoSequenase DNA polymerase and because electrophoresis and detection occurred on the same gel, the signal strengths and confidence measures obtained are attributable largely to the differences in quantum efficiency between ET and conventional fluorescent primers. We have performed similar comparative experiments on four other BAC clones in this study and have found the aforementioned trend in quality measures for phred base calls of the data to be reproducible (data not shown).

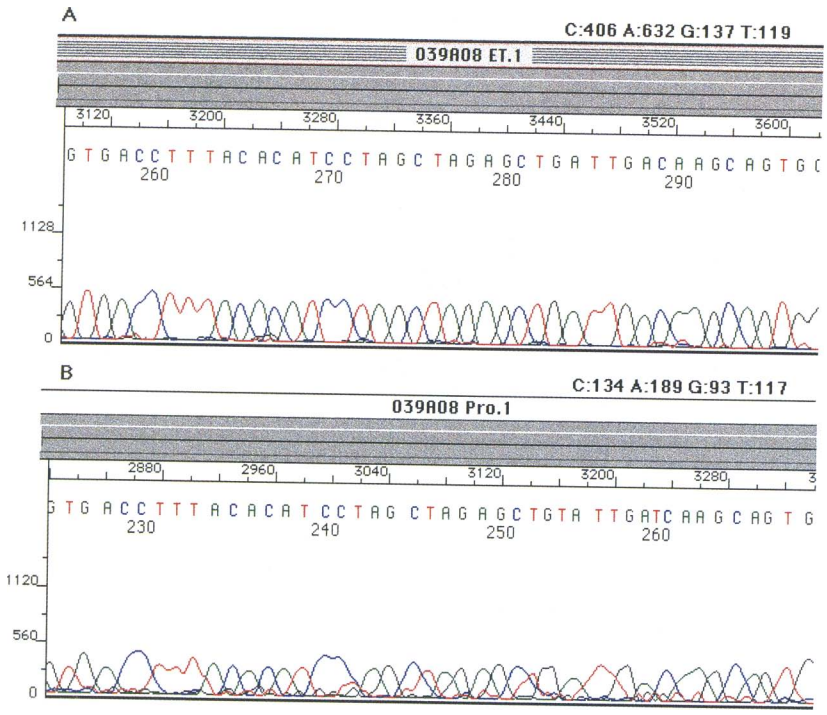

Figure 1 A comparison of sequence data quality for ET fluorescent primers and conventional, single dye-labeled fluorescent primers. Sections of DNA sequence trace data obtained from sequencing of BAC clone 039A08 are shown. (*A*) Data from reactions primed with universal (−40) Energy Transfer dye primers. (*B*) Data from reactions primed with singly labeled universal (−21) dye primers from Promega Biotech (Madison, WI), where all reagents for the G and T reactions were doubled, according to standard procedure. Electrophoresis and detection for these two sequences were performed on the ABI 377 DNA Sequencer during the same gel run, as described in the text.

Experiments conducted to determine whether increased primer concentration would enhance the read length of the sequencing reactions by increasing the amount of product are illustrated in Figure 2. Here a twofold increase in primer concentration showed no change in fragment distribution and no increase in peak height or effective read length, in spite of the higher signal strength indicated by the ABI software. We also experimented with increased cycle number in anticipation of increasing the amount of fluorescent labeled sequencing products and therefore the peak heights and read length. However, as shown in Figure 3, performing more than 20 cycles for these reactions contributed to a less even, more top-heavy distribution of products, even with increased primer amounts.

To substantiate the robustness of our protocol, we have attempted

Table 1. Phred Average Quality Values for ET Primed and Single Fluor Primed BAC End Sequence Data from Figure 1

Range of called bases	1–100	101–200	201–300	301–400	401–500	501–600	601–700	701–800
039A08 (ET)	31.08	43.00	38.04	28.69	24.32	18.83	11.77	8.93
039A08 (Promega)	23.94	32.75	15.22	9.71	8.82	8.38	7.86	8.49

to generate sequence from the ends of 13 different BAC clones from the CCM-1 region of human chromosome 7. Our overall success rate in obtaining sequence data using forward (−40) and reverse (−28) ET primers for this data set was 79.2%, based upon whether sequence data was obtained. The resulting sequence data were converted to phred basecalls using the algorithm of P. Green (pers. comm.). The phred output quality measures of base-calling confidence for these sequences, described earlier, are represented graphically in Figure 4. This analysis indicates that the mean phred quality measures out to 800 bases of data are greater than 10^9 (overall error probability of one in nine for the 100-base region). The size range of BAC clones tested further indicates the robustness of this protocol, as successful reactions were obtained over a >3.5-fold clone size variation from 41 to 144 kb. The high quality of sequencing data obtained from ET primed reactions thus routinely provides extended read length capability over conventional labeling chemistry and, with the optimized reaction conditions described here, provides more end sequence data per clone. This, in turn, may reduce the number of large clones sequenced to confirm clonal contiguity in a regional mapping effort. One variation of this concept is to use end sequence data to substantiate clonal map order in closely related organisms, one of which already has been entirely sequenced, in an effort to examine conservation of synteny. In addition, the high quality of this data can be used to provide reliable primer sequences for new STS generation that, when used to probe libraries, can identify more clones and thereby increase map coverage. We currently are investigating these possibilities using the end sequencing protocol described herein and hope to report our results in a subsequent publication.

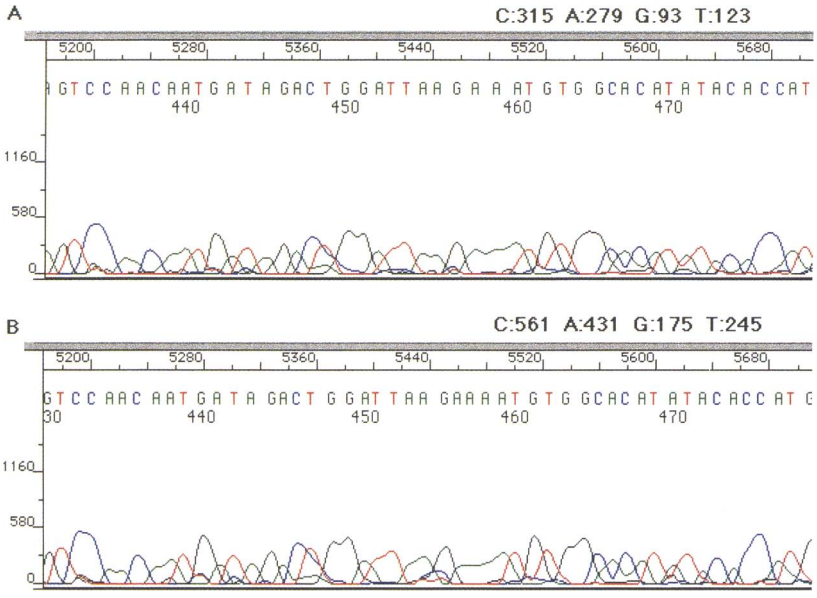

Figure 2 A comparison of peak resolution and signal strength for BAC clone sequencing reactions using different amounts of input ET primer. (*A*) BAC clone 013L03 sequenced using 0.5 pmole of the A and C, and 1 pmole of the G and T universal (−40) ET primers. (*B*) The same clone sequenced with 1 pmole of the A and C, and 2 pmole of the G and T universal ET primers. Both samples were sequenced and analyzed on the ABI 377 DNA Sequencer during the same gel run. ABI software-generated signal strengths for each reaction are indicated in the *upper right* of each panel.

METHODS

Preparation of DNA

The following growth and preparation conditions have produced DNA suitable for end sequence analysis using the reaction conditions described in the next section. Glycerol stocks of individual

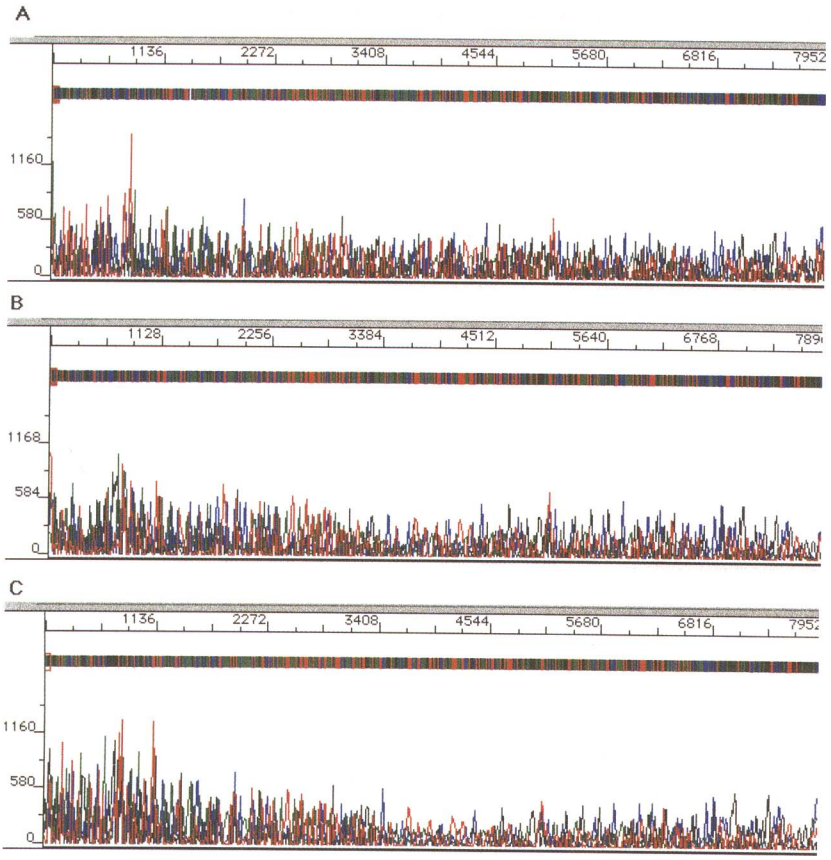

Figure 3 The effect of cycle number on fragment-length distribution for ET primer reactions using BAC clone 013L03. Universal (−40) Energy Transfer primers and ThermoSequenase reactions were performed as described in the text, with variations in the number of cycles following the initial denaturation step at 95°C for 2 min. Shown is the compressed view of data obtained over the entire electrophoresis and data collection time of 10 hr for this sample, cycled as follows: (*A*) 20 cycles; (*B*) 30 cycles; (*C*) 50 cycles. The observable trend for these and other samples tested was a more top-heavy fragment distribution with increasing cycle number.

BAC clones were used to streak 2 × YT agar plates containing 12.5 µg/ml chloramphenicol, with overnight growth at 37°C. Individual colonies were picked to inoculate 50 ml of 2 × YT media (supplemented with 12.5 µg/ml chloramphenicol) held in 500 ml Erlenmeyer flasks. The resulting cultures were grown for 20–24 hr at 37°C with ~300 rpm agitation. Cells were pelleted after culture transfer to 50 ml conical tubes (Falcon) by centrifugation at 4500 rpm for 30 min in a Jouan GR422 centrifuge. The supernatant was aspirated and the tubes inverted for 5 min to drain. Each pellet was resuspended in 5 ml GET buffer (50 mM Tris-HCl at pH 8.0, 50 mM glucose, 10 mM EDTA) and 60 µl of 20 mg/ml RNase A by stirring with a sealed Pasteur pipette followed by gentle repeated pipetting. The resulting suspension then was transferred to an Oak Ridge tube, and cell lysis was achieved by the addition of 5 ml of 0.2 N NaOH, 1% SDS solution (made fresh), followed by gentle inversion two times. The lysate was neutralized with the addition of 5 ml of 3 M potassium acetate at pH 5.5 and gently inverted five times prior to placing on ice for 5 min. An additional 10 inversions were performed, followed by 10 min on ice. Another five inversions preceded pelleting of the cell debris by centrifugation at 15,000 rpm for 30 min in a Beckman JA-17 rotor. The resulting supernatant was transferred by pipet into a fresh Oak Ridge tube, avoiding transfer of floating debris. Adding 0.67 volumes isopropanol to the transferred supernatant, inversion and incubation at room temperature for 15 min precipitated the DNA, which was pelleted by another centrifugation at 15,000 rpm for 30 min in the JA-17 rotor. A 70% ethanol rinse followed, after which the DNA pellet was dried under vacuum. Pellets were resuspended in 500 µl TE buffer (10:1) and transferred to a microcentrifuge tube. This solution was extracted once with an equal volume of phenol and once with an equal volume of phenol–chloroform, followed by precipitation of the resulting upper phase with an equal volume of isopropanol. After a 15-min incubation at room temperature, a 10-min centrifugation at 13,000 rpm in a benchtop microfuge and a 70% ethanol wash, the DNA pellet was dried under vacuum. The pellet was resuspended in 40 µl of TE (10:0.1) and analyzed by electrophoresis on 0.4% agarose with ethidium bromide stain, then used for sequencing, as described below.

DNA Sequencing Reactions

Each sequencing reaction contained 0.4 µl of ThermoSequenase Reaction Buffer (260 mM Tris-HCl at pH 9.5 and 65 mM $MgCl_2$), 1 µl of Dye-namic Energy Transfer Primer (−40 universal or −28 reverse, Amersham Life Science) at 0.5 µM (A and C primers) or 1 µM (G and T primers), 1.6 µl of ThermoSequenase nucleotide mix (750 µM each dNTP, 2.5 µM ddNTP), 1 µl of ThermoSequenase polymerase (Amersham) at 1.5 U/µl and 1 µg of DNA template (prepared as described above), in a total volume of 5–6 µl. The reactions were assembled in 0.2 ml thin-walled polypropylene tubes (Robbins Scientific) and cycled in an MJ Research PTC-200 DNA Engine using the following parameters: 95°C for 2 min, followed by 20 cycles of 95°C for 5 sec, 55°C for 10 sec (50°C for reverse primer), 72°C for 60 sec, then cooled to 4°C. Reaction products were pooled and precipitated in 1.5 ml tubes with the addition of 1 µl glycogen (Boehringer Mannheim, 20 mg/ml) and 100 µl of 100% ethanol. Products were pelleted by centrifugation at 13,000 rpm for 15 min in a benchtop microfuge. Ethanol was decanted and the pellets dried under vacuum prior to resuspension in 3 µl of formamide–dH_2O (5:1) with blue dextran. Products were elec-

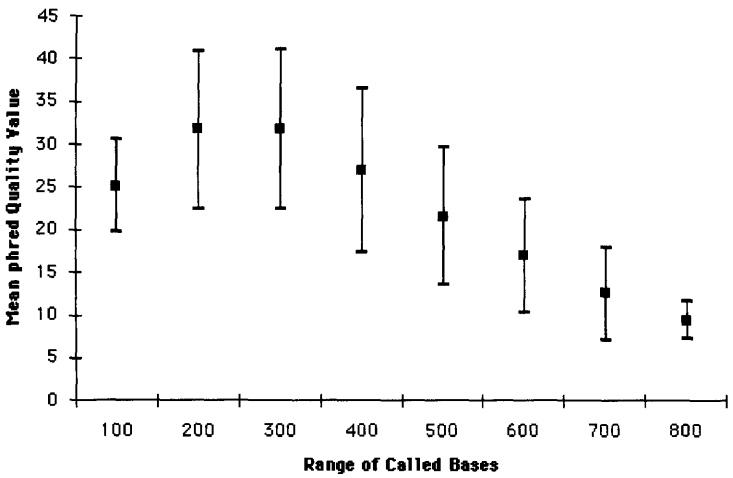

Figure 4 Mean phred quality measures as a function of called base position for the BAC end sequences generated in this study (n = 34). Error bars indicate 95% confidence limits. Sequences of obvious overall poor quality (phred quality values <8 for the entire range) were excluded from these calculations. The highest quality sequence occurs, on average, over the range of bases 101–200 (overall error probability of 1/1000), and steadily declines with increasing distance from the priming site. Decreases in gel resolution capability contribute to decreased phred confidence values near the end of the sequence data.

trophoresed through a 5% Page Plus (Amresco), 6 M urea, 1× TBE gel on the Applied Biosystems 377 DNA Sequencer using the 2 X A run module and an extended run time of 10 hr.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Eric Green of the National Institutes of Health for providing the BAC clones from the chromosome 7 CCM-1 region used in this study, and Drs. Robert Waterston and Richard Wilson, in whose lab the study was conducted. We gratefully acknowledge the help of LaDeana Hillier with phred analysis of the BAC clone sequences. We would also like to acknowledge the generous contribution of ET primers from Dr. Peter B. Vander Horn and Amersham Life Sciences.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

Chen, E.Y., D. Schlessinger, and J. Kere. 1993. Ordered shotgun sequencing, a strategy for integrated mapping and sequencing of YAC clones. *Genomics* **17:** 651–656.

Coulson, A., J. Sulston, S. Brenner, and J. Karn. 1986. Toward a physical map of the genome of the nematode *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci.* **83:** 7821–7825.

Ju, J., C. Ruan, C.W. Fuller, A.N. Glaser, and R. Mathies. 1995. Fluorescence energy transfer dye-labeled primers for DNA sequencing and analysis. *Proc. Natl. Acad. Sci.* **92:** 4347–4351.

Tabor, S. and C.C. Richardson. 1995. A single residue in DNA polymerases of the *Escherichia coli* DNA polymerase I family is critical for distinguishing between deoxy- and dideoxyribonucleotides. *Proc. Natl. Acad. Sci.* **92:** 6339–6343.

Voss, H., J. Zimmermann, C. Schwager, H. Erfle, J. Stegemann, K. Stucky, and W. Ansorge. 1990. Automated fluorescent sequencing of cosmid DNA. *Nucleic Acids Res.* **18:** 1066.

Received May 3, 1966; accepted in revised form September 3, 1996.