Resuspension of DNA Sequencing Reaction Products in Agarose Increases Sequence Quality on an Automated Sequencer

BioTechniques 33:532-538 (September 2002)

ABSTRACT

We are investigating approaches to increase DNA sequencing quality. Since a major factor in sequence generation is the cost of reagents and sample preparations, we have developed and optimized methods to sequence directly plasmid DNA isolated from alkaline lysis preparations. These methods remove the costly PCR and post-sequencing purification steps but can result in low sequence quality when using standard resuspension protocols on some sequencing platforms. This work outlines a simple, robust, and inexpensive resuspension protocol for DNA sequencing to correct this shortcoming. Resuspending the sequenced products in agarose before electrophoresis results in a substantial and reproducible increase in sequence quality and read length over resuspension in deionized water and has allowed us to use the aforementioned sample preparation methods to cut considerably the overall sequencing costs without sacrificing sequence quality. We demonstrate that resuspension of unpurified sequence products generated from template DNA isolated by a modified alkaline lysis technique in low concentrations of agarose yields a 384% improvement in sequence quality compared to resuspension in deionized water. Utilizing this protocol, we have produced more than 74 000 high-quality, long-read-length sequences from plasmid DNA template on the MegaBACE™ 1000 platform.

INTRODUCTION

The introduction and subsequent commercial availability of high-throughput capillary array electrophoresis DNA sequencing instruments in the late 1990s increased enormously DNA sequencing capacity. The utilization of large numbers of these machines in public and private ventures around the globe made possible the announcement of the essentially complete human genome years ahead of schedule (8,18). These instruments employ an array of narrow-bore, gel-filled capillaries for high-speed, high-resolution separations of multiple sequencing samples and feature automated sample injection, data collection, and processing (4–6,9). However, despite the many hardware improvements in sequencing technologies, a major hurdle of modern genome-scale sequencing undertakings remains the high cost of reagents and sample preparation.

Our Genome Sciences Centre is participating in a variety of large-scale, high-throughput sequencing activities that utilize two types of capillary array DNA sequencers, the MegaBACE™ 1000 (Molecular Dynamics, Sunnyvale, CA, USA) and the ABI PRISM® 3700 (Applied Biosystems, Foster City, CA, USA). In addition to various sequencing projects, we are engaged in ongoing technology development, researching methods to utilize the existing hardware more efficiently and cost-effectively.

This paper details our work optimizing DNA sequence quality with the MegaBACE 1000 sequence analysis platform. This instrument utilizes electrokinetic injection to introduce the samples into the capillaries. During electrokinetic injection, the capillary tips and a negatively charged electrode are immersed in the sample, while the positive electrode resides at the other end of the capillary. Upon application of an electric potential, a fraction of the negatively charged molecules in the sample, including DNA, is drawn into and through the capillary. The amount of material loaded into the capillary during electrokinetic injection depends on the injection time, voltage, and buffer conductivity (4). Sequencing products can be effectively concentrated during the injection by exploiting the differences in the ionic strengths of the electrophoresis buffer and the sample resuspension buffer leading to a phenomenon known as “sample stacking” (4,16,17). When a sample is injected from a low-conductivity environment, such as water, into a buffer possessing a much higher conductivity, such as the MegaBACE running buffer, the sample ions, DNA molecules, and proteins are stacked in the front of the sample zone upon application of the electric field (7,17).

A major disadvantage of the electrokinetic injection procedure is that the amount of sample loaded can be strongly affected by variable experimental conditions such as salt or template concentration (4,7). The performance of the MegaBACE 1000 sequencing platform is known to be sensitive to the amount and purity of template that is loaded, as overloaded capillaries have been shown to give a reduced signal or no signal at all (10).

Sequencing template generated by PCR amplification is recommended by the manufacturers of the both the MegaBACE 1000 and the ABI PRISM 3700 sequencers, as it yields a DNA starting material that is essentially free of contaminating bacterial genomic and plasmid DNA. However, PCR-based sequencing template preparation is expensive, and amplification of certain sequences (e.g., homopolymeric runs) can be problematic. Hence, we have investigated direct sequencing of plasmid DNA purified by an alkaline lysis procedure (J. Schein, T. Kucaba, M. Sekhon, D. Smailus, R. Waterston, and M. Marra, submitted). DNA sequencing template isolated by our alkaline lysis technique may contain more contaminants than PCR-generated template, but the procedure can be performed efficiently and inexpensively in 96-well format.

When cycle sequencing reactions from our alkaline lysis DNA templates were resuspended in water and loaded on the MegaBACE 1000, we found that sequence read length and quality were poor, presumably because of contaminating template DNA, genomic DNA, or some other impurity present in the samples. To address this, we experimented with resuspension of the cycle sequencing products in various concentrations of agarose, with the rationale that the agarose should limit the amount of residual bacterial genomic or plasmid DNA loaded onto the capillaries during the electrokinetic injection.

Optimum read length and sequence quality were obtained when the sequenced products were resuspended in 0.06% agarose before electrokinetic in-

jection. Utilizing this resuspension protocol, we have produced more than 74 000 high-quality sequences from crudely purified template on the MegaBACE 1000 platform without having to perform post-reaction cleanup steps.

MATERIALS AND METHODS

Template Preparation

Bacteria containing a single cDNA clone from the Mammalian Gene Collection (MGC-10790) were inoculated into each well of a 96-well culture block (Beckman Coulter, Fullerton, CA, USA) in 1.2 mL 2× YT media (BD Biosciences, San Jose, CA, USA) supplemented with chloramphenicol (Sigma, St. Louis, MO, USA) at 12.5 µg/mL. The block was sealed with a sheet of AirPore™ tape (Qiagen, Valencia, CA, USA) and incubated for 16 h at 37°C with agitation at 290 rpm in a C25 Incubator/Shaker (New Brunswick Scientific, Edison, NY, USA) fitted with custom holders. Five microliters of this culture were then used to seed each well of eight similar blocks, which were incubated under identical conditions. Following growth, cell pellets were collected by centrifugation for 20 min at 1400× *g*, and the media were decanted. Draining of residual media was achieved by inverting the blocks over paper towelling. The plasmid DNA was harvested using a modified alkaline lysis protocol (J. Schein, T. Kucaba, M. Sekhon, D. Smailus, R. Waterston, and M. Marra, submitted). This protocol has undergone continuous refinement over several years. The improved sequence quality obtained by resuspending in agarose products sequenced from this template has been consistent. Therefore, we believe that resuspension in agarose would improve the sequence quality from other cell-mediated DNA amplification methods. The pelleted DNA in each well was resuspended in 140 µL water, and all samples were pooled and re-aliquoted into eight fresh 96-well plates (Nalge Nunc International, Rochester, NY, USA). The template DNA was pooled to provide a homogeneous starting material for the development and optimization of the resuspension protocol. This DNA solution was used directly, without further purification, for DNA sequencing and restriction analysis. The integrity of the DNA was checked by excising the cDNA insert from the pOTB7 vector by digestion with *Eco*RI and *Xho*I (Invitrogen, Carlsbad, CA, USA) and resolving the fragments on a 0.7% agarose gel (data not shown).

Sequencing Reactions

DNA was sequenced (13) using DYEnamic™ energy transfer dye terminators (Amersham Biosciences, Piscataway, NJ, USA). To increase the speed and efficiency of transfers and decrease well-to-well variability all fluid transfer steps were performed with HYDRA® 96-channel microdispensers (Robbins Scientific, Sunnyvale, CA, USA). All solutions and reactions were prepared with deionized 18 MΩ water from a Milli-Q® water purification system (Millipore, Bedford, MA, USA). As part of our sequencing pipeline, "master mixtures" of sequencing reaction components are assembled and dispensed into multiple CyclePlates®-96 ET PCR plates (Robbins Scientific) using a HYDRA microdispenser, which is also used subsequently to add the DNA template. In this manner, 6 µL of the homogenized template DNA (164 ng/µL) were sequenced in each well of 28 96-well PCR plates. Each 20 µL reaction contained 8 µL DYEnamic energy transfer brew mixture (Amersham Biosciences) and 5 pmol -21M13 forward primer (5′-TGTAAAACGACGGCCA-GT-3′). A MicroAmp® Full Plate Cover (Perkin Elmer Life Science, Gaithersburg, MD, USA) was placed onto the cycleplates, and the reaction mixtures were collected in the bottom of the wells by brief centrifugation at 700× *g*. Cycle sequencing was performed in a PTC-225™ DNA engine tetrad (MJ Research, Waltham, MA, USA) with a

ramp speed of 3°C/s using 30 cycles of 95°C for 20 s, 48°C for 15 s, 60°C for 1 min, followed by incubation at 4°C. The sequencing reactions from all plates were collected by centrifugation at 700× *g* after thermal cycling, pooled, and re-aliquoted into 28 fresh 96-well cycle-plates. To concentrate the sequenced products and remove extra salts and dye-terminators, the reactions were ethanol-precipitated. Sixty microliters of 95% ethanol and 183 mM ammonium acetate (pH 7.5) were added to each well, and, after mixing by repeated pipetting with a HYDRA microdispenser, DNA precipitates were collected by centrifuging the cycleplate for 30 min at 2750× *g* at 4°C. The ethanol/salt mixture was decanted immediately following centrifugation by inverting and vigorously shaking the plates to remove the liquid from the wells. Following a wash with 150 µL 70% ethanol, the plates were spun inverted at 700× *g* over paper towelling for 1 min to remove any residual ethanol. The reaction pellets were dried in a SpeedVac® (ThermoSavant, Holbrook, NY, USA) with the rotor removed, on high heat for 2 min. The plates were sealed with foil tape, and the precipitated sequencing reactions were stored at -20°C in plastic bags.

Sequencing Parameters

Immediately before electrophoresis, reactions were resuspended in 20 µL water or 20 µL molten (64°C) 0.02–0.6% SeaKem® Gold agarose (BioWhittaker Molecular Applications, Rockland, ME, USA) in water by gentle vortex mixing for 20 s. For each water and agarose resuspension experiment, one 96-well plate was resolved on one MegaBACE 1000 96-capillary array sequencer, while a second plate was resolved on a second machine. The run parameters (electrokinetic injection was for 10 s at 3 kV and electrophoresis was continued at 6 kV for 240 min) were identical for all plates. These settings represent the optimized conditions on our equipment when alkaline lysis plasmid DNAs are used as template.

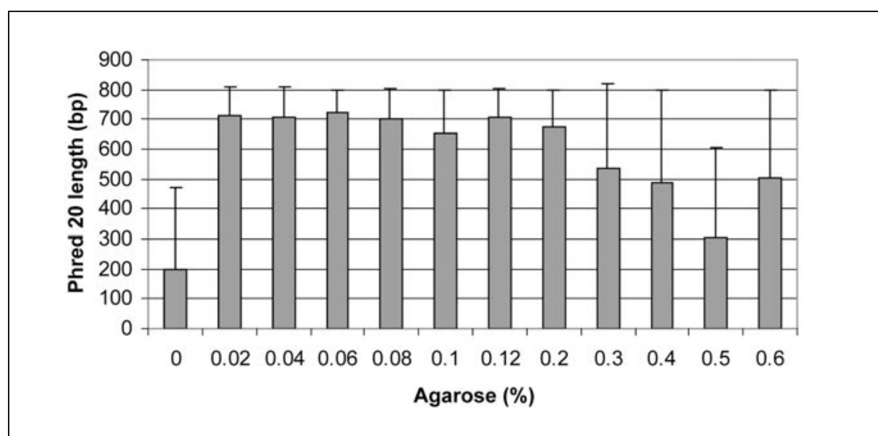

Figure 1. Summary of average Phred 20 lengths when reactions are resuspended in various concentrations of agarose. For each agarose concentration (%) listed on the horizontal axis, the average Phred 20 lengths (bp) are depicted by the height of the bars. Data for each column are from 192 wells (two 96-well plates); error bars depict the standard deviation of the Phred 20 lengths.

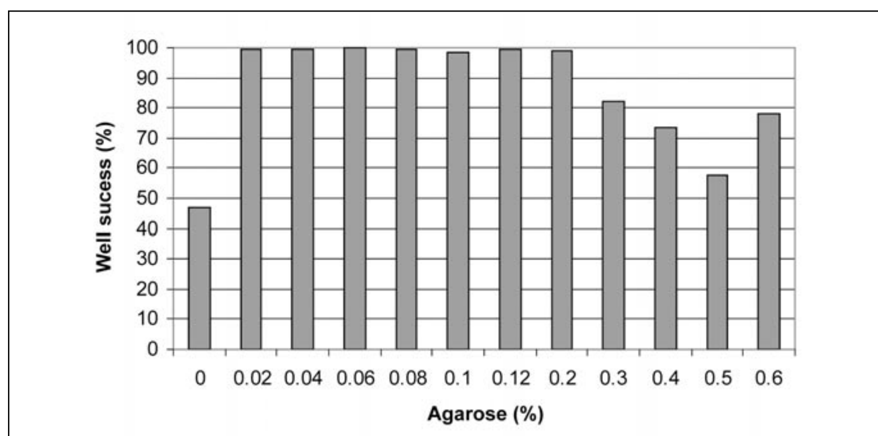

Figure 2. Well successes when reactions are resuspended in water and various concentrations of agarose. The height of the bars for each agarose concentration (%) lists the percentage of wells (from a total of 192 wells) that give useful sequence. For a given well, success is defined as at least 50 Phred 20 bases.

Data Analysis

To reduce the contribution of experimental artifacts and concentrate on the effects that the resuspension media had on sequence quality, the analytes were standardized where possible by the multiple poolings outlined above. Hence, well-to-well variability due to differences in DNA template concentration and inconsistency in cycling reactions were minimized. To further reduce variables, all plates contained the same cDNA clone, were sequenced with the same primer, and electrophoresed under identical conditions.

After electrophoresis, scripts automatically process the DNA sequence chromatograms using the PHRED software (available free of charge at http://www.phrap.org) (1,2) and move the processed data into a MySQL™ database (available free of charge at http://www.mysql.com). Sequence quality was assessed using a combination of Consed (3) and a suite of in-house Web tools (unpublished data). During the course of these experiments, 2304 sequencing reactions (representing 24 96-well plates) were analyzed. For the water resuspension and for each of the 11 agarose concentrations tested, two 96-well plates were electrophoresed. For each of the 12 different resuspension media experiments, two parameters were assayed: (*i*) average Phred 20 length (Figure 1), which is defined by the average Phred 20 lengths of each of 195 sequences (two 96-well plates); and (*ii*) sequence success (Figure 2)—a given read was called successful if the Phred 20 length was greater than 50 bp.

When the sequencing reaction products were resuspended in water and resolved on the MegaBACE 1000, the results were disappointing. Many of the resulting reads failed, while those that did work were short and of poor quality. When water was used as a resuspension media, the average Phred 20 length per 96-well plate was only 198 bp (Figure 1). The low averages for the water plates were due in part to the number of failed wells. Using the definition of success outlined above (a Phred 20 length of greater than 50 bp), only half of the wells succeeded (46.9%; Figure 2). If only successful wells are included, then the average Phred 20 for water resuspension is 416 bp.

Believing a possible contributing factor to the poor quality of the sequence to be the electrokinetic introduction of charged contaminants into the capillaries, we resuspended the sequencing products in 0.1% agarose before injection and noticed a substantial improvement. The average Phred 20 read length increased to 653 bp (Figure 1). The addition of agarose also considerably increased sequencing success, with 98.4% of the wells having more than 50 Phred 20 bases (Figure 2). We subsequently tested a range of agarose concentrations from 0.02% to 0.6% to find the optimum value for resuspension to achieve the highest-quality sequence.

As summarized in Figures 1 and 2, the maximum Phred 20 length with the fewest failed wells was achieved when the sequencing reactions were resuspended in 0.06% agarose. When resuspended in this concentration of agarose, the resulting average Phred 20 read length was 721 bp with a 100% success rate. This represents a 374% increase in average Phred 20 sequence length when compared to plates resuspended in water. It has become standard practice in our laboratory to resuspend in 0.06% agarose sequencing reactions that are to be resolved on the MegaBACE 1000 platform.

When this modified resuspension protocol was applied to production sequencing projects, where every well contained a different plasmid subclone template, similar improvements in quality were obtained (data not shown).

DISCUSSION

Our continuing goal is to generate, at high throughput, high-quality, long-read-length DNA sequence while minimizing the cost per base. Toward this end, we developed a protocol for sequencing inexpensive DNA templates on the MegaBACE 1000 platform.

Electrokinetic injection, which is used by the MegaBACE 1000, causes DNA sequenced by capillary array electrophoresis to be sensitive to conta-

minants in the sample such as template DNA, genomic DNA, buffer anions, and other impurities (4,7,14). Reasoning that the plasmid DNA template generated by alkaline lysis probably contained many impurities, we resuspended the sequenced products in low concentrations of agarose, with the rationale that the agarose would act as a molecular sieving matrix, excluding the larger contaminants. In effect, resuspending the sequencing products in agarose before loading acts to "gel-purify" the cycle sequencing products as they are electrokinetically injected.

The manner by which sequence quality is degraded as a result of charged contaminants in the sample is not yet well understood. Capillary currents often exhibit either dramatic variability, perhaps due to bubble formation, or rapid drop off due to low steady-state values. These effects will at best interfere with the signal peak spacing, and at worst effectively stop the migration of analytes. Although poorly understood, these phenomena were documented during early development of capillary electrophoresis (15) and have caused problems ever since (11). Degradation of sequence quality because of excessive injection of template DNA from samples with very low salt concentration has also been shown (12). Our own experiments with a custom-built capillary electrophoresis test-bed indicate that bubble formation leads to unstable electric current within the capillary, and experiments with the MegaBACE 1000 confirm that poor-quality reads are well correlated with a decrease in current through the capillary. Further work will be performed to uncover exactly how the injection of large DNA fragments or other contaminants such as bacterial proteins may lead to bubble formation and current instabilities.

As the pore size of agarose is too large to exclude salts, we believe that the increased sequence quality with agarose resuspension is due to the ability of agarose to exclude selectively large DNA and protein contaminants from the capillary. Contaminating salts are removed or reduced by concentrating the sequencing products with ethanol precipitation before electrokinetic injection.

The purpose of our experiment was to optimize the quality of sequence obtained from alkaline lysis-purified template DNA. We determined that agarose at a range of concentrations from 0.02% to 0.2% could be used as an effective resuspension medium, possibly impeding the movement of contaminating DNA or other negatively charged macromolecules during electrokinetic injection. At higher concentrations of agarose (0.3%–0.6%), the sequence quality starts to decline (Figure 1) and the number of failed wells increases (Figure 2). It is likely that as the agarose concentration increases the movement of smaller DNA fragments, including sequencing products, is affected as well. We routinely use 0.06% agarose to resuspend sequencing reactions before injection, as this concentration yielded the highest average sequence length (721 bp). In summary, we have shown that agarose resuspension dramatically improves both sequence quality and read length with a negligible cost increase. Eliminating the need to PCR-amplify the sequencing template and subsequently purify the sequenced products substantially reduces the overall cost of sequencing.

REFERENCES

1. **Ewing, B. and P. Green.** 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res. *8*:186-194.
2. **Ewing, B., L. Hillier, M.C. Wendl, and P. Green.** 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. *8*:175-185.
3. **Gordon, D., C. Abajian, and P. Green.** 1998. Consed: a graphical tool for sequence finishing. Genome Res. *8*:195-202.
4. **Hadd, A.G., M.P. Goard, D.R. Rank, and S.B. Jovanovich.** 2000. Sub-microliter DNA sequencing for capillary array electrophoresis. J. Chromatogr. A *894*:191-201.
5. **Huang, X.C., M.A. Quesada, and R.A. Mathies.** 1992. DNA sequencing using capillary array electrophoresis. Anal. Chem. *64*:2149-2154.
6. **Kheterpal, I. and R.A. Mathies.** 1999. Capillary array electrophoresis DNA sequencing. Anal. Chem. *71*:31A-37A.
7. **Krivacsy, Z., A. Gelencser, J. Hlavay, G. Kiss, and Z. Sarvari.** 1999. Electrokinetic injection in capillary electrophoresis and its application to the analysis of inorganic compounds. J. Chromatogr. A *834*:21-44.
8. **Lander, E.S., L.M. Linton, B. Birren, C. Nusbaum, M.C. Zody, J. Baldwin, K. Devon, K. Dewar, et al.** 2001. Initial sequencing and analysis of the human genome. Nature *409*:860-921.
9. **Marsh, M., O. Tu, V. Dolnik, D. Roach, N. Solomon, K. Bechtol, P. Smietana, L. Wang, et al.** 1997. High-throughput DNA sequencing on a capillary array electrophoresis system. J. Capillary Electrophor. *4*:83-89.
10. **Marziali, A. and M. Akeson.** 2001. New DNA sequencing methods. Annu. Rev. Biomed. Eng. *3*:195-223.
11. **Ruiz-Martinez, M.C., O. Salas-Solano, E. Carrilho, L. Kotler, and B.L. Karger.** 1998. A sample purification method for rugged and high-performance DNA sequencing by capillary electrophoresis using replaceable polymer solutions. A. Development of the cleanup protocol. Anal. Chem. *70*:1516-1527.
12. **Salas-Solano, O., M.C. Ruiz-Martinez, E. Carrilho, L. Kotler, and B.L. Karger.** 1998. A sample purification method for rugged and high-performance DNA sequencing by capillary electrophoresis using replaceable polymer solutions. B. Quantitative determination of the role of sample matrix components on sequencing analysis. Anal. Chem. *70*:1528-1535.
13. **Sanger, F., S. Nicklen, and A.R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA *74*:5463-5467.
14. **Shihabi, Z.K.** 1999. Capillary electrophoresis of double-stranded DNA in an untreated capillary. J. Chromatogr. A *853*:349-354.
15. **Swerdlow, H., K.E. Dew-Jager, K. Brady, R. Grey, N.J. Dovichi, and R. Gesteland.** 1992. Stability of capillary gels for automated sequencing of DNA. Electrophoresis *13*:475-483.
16. **Tagliaro, F., G. Manetto, F. Crivellente, and F.P. Smith.** 1998. A brief introduction to capillary electrophoresis. Forensic Sci. Int. *92*:75-88.
17. **Tseng, W.L., M.M. Hsieh, S.J. Wang, C.C. Huang, Y.C. Lin, P.L. Chang, and H.T. Chang.** 2001. Analysis of large-volume DNA markers and polymerase chain reaction products by capillary electrophoresis in the presence of electroosmotic flow. J. Chromatogr. A *927*:179-190.
18. **Venter, J.C., M.D. Adams, E.W. Myers, P.W. Li, R.J. Mural, G.G. Sutton, H.O. Smith, M. Yandell, et al.** 2001. The sequence of the human genome. Science *291*:1304-1351.

G.P.V. is supported by a Natural Sciences and Engineering Research Council of Canada (NSERC) fellowship and a Michael Smith Foundation for Health Research Fellowship. M.A.M. is a Michael Smith Foundation for Health Research Scholar. This work was supported by the BC Cancer Foundation. R.C. and A.M. are supported through the National Human Genome Research Institute (NHGRI) grant no. R01-HG02072. *Address correspondence to Dr. Marco Marra, Genome Sciences Centre, BC Cancer Agency, 600 West 10th Avenue, Vancouver, BC, Canada, V5Z 4E6. e-mail: mmarra@bcgsc.bc.ca*

Received 22 March 2002; accepted 11 June 2002.

G. Vatcher, D. Smailus, M. Krzywinski, R. Guin, J. Stott, M. Tsai, S. Chan, P. Pandoh, G. Yang, J. Asano, T. Olson, A.-L. Prabhu, R. Coope[1], A. Marziali[1], J. Schein, S. Jones, and M. Marra
Genome Sciences Centre
BC Cancer Agency
[1]University of British Columbia
Vancouver, BC, Canada

fRFLP and fAFLP: Medium-Throughput Genotyping by Fluorescently Post-Labeling Restriction Digestion

BioTechniques 33:539-546 (September 2002)

ABSTRACT

Genome-scale studies of population structure and high-resolution mapping of genetically complex traits both require techniques for accurately and efficiently genotyping large numbers of polymorphic sites in multiple individuals. Many high-throughput genotyping technologies require the purchase of expensive equipment or consumables and are therefore out of reach of some individual research laboratories. Conversely, less expensive technologies are often labor intensive so that the effort involved in typing large numbers of samples or polymorphic sites is prohibitive. Here we present a method of fluorescently post-labeling restriction digestion using standard dye-terminator sequencing chemistry so that RFLP and AFLP products can be visualized on an automated sequencer. This labeling method is efficient, inexpensive, easily multiplexed, and requires no unusual equipment or reagents, thus striking a balance between cost and throughput that should be appropriate for many research groups and core facilities.

INTRODUCTION

Restriction endonucleases have long been used in population studies of allelic variation, where the presence or absence of a cut site is indicative of a mutation in the enzyme target sequence. In genetically well-characterized organisms, diagnostic RFLPs are often used as markers for previously characterized alleles. In organisms with less well-studied genetics, the selective amplification of anonymous restriction digestion products (e.g., AFLPs) (8) has also been used with considerable success (5,9). Although they are extremely powerful in the appropriate contexts, many existing RFLP and AFLP methods have limitations in throughput, sensitivity, or consumable cost that lessen their utility in many individual research laboratories. Traditionally, RFLP and AFLP products have been electrophoretically resolved on either agarose or acrylamide slab gels, with the digested fragments visualized by ethidium bromide staining or radiolabeling. Existing fluorescence-based RFLP and AFLP methods use labeled amplification primers (1,3,4,6) that could cost over $100 apiece (Research Genetics, Huntsville, AL, USA). This quickly increases the cost of an experiment using multiple primers and, in the case of RFLP, only allows the visualization of a single terminal fragment, thus losing all sequence information 3′ of the first restriction cut site. To increase RFLP information content and allow higher RFLP and AFLP throughput at lower costs, we have developed a simple and inexpensive method for post-labeling restriction digestion fragments using "off-the-shelf" sequencing chemistry before separation on an automated sequencer. We call this labeling method fluorescent RFLP (fRFLP).

Briefly, we amplify the target locus in a standard PCR and then digest the product with a restriction enzyme that leaves a 5′ overhang. This overhang acts as a template for the single base incorporation of a fluorescent dye-terminator nucleotide from a standard cycle sequencing kit. Thus labeled, the fragments are resolved on a capillary-based sequencer. Multiplexing is facilitated by the high resolution of the sequencer and its ability to read simultaneously multiple fluor wavelengths. This label-