ELSEVIER

Journal of Chromatography A, 894 (2000) 191–201

JOURNAL OF CHROMATOGRAPHY A

www.elsevier.com/locate/chroma

Sub-microliter DNA sequencing for capillary array electrophoresis

Andrew G. Hadd, Michael P. Goard, David R. Rank, Stevan B. Jovanovich*

Molecular Dynamics/Amersham Pharmacia Biotech, 928 East Arques Avenue, Sunnyvale, CA 94086, USA

Received 22 February 2000; received in revised form 28 March 2000; accepted 4 April 2000

Abstract

DNA sequencing from sub-microliter samples was demonstrated for capillary array electrophoresis by optimizing the analysis of 500 nl reaction aliquots of full-volume reactions and by preparing 500 nl reactions within fused-silica capillaries. Sub-microliter aliquots were removed from the pooled reaction products of 10 μl dye–primer cycle-sequencing reactions and analyzed without modifying either the reagent concentrations or instrument workflow. The impact of precipitation methods, resuspension buffers, and injection times on electrokinetic injection efficiency for 500 nl aliquots were determined by peak heights, signal-to-noise ratios, and changes in base-called readlengths. For 500 nl aliquots diluted to 5 μl in 60% formamide–1 m*M* EDTA and directly injected, a five-fold increase in signal-to-noise ratios was obtained by increasing injection times from 10 to 80 s without a corresponding increase in peak widths or reduction in readlengths. For 500 nl aliquots precipitated in alcohol, 80±5% template recovery and a two-fold decrease in conductivity was obtained, resulting in a two-fold increase in peak heights and 50 to 100 bases increase in readlengths. In a comparison of aliquot volumes and precipitation methods, equivalent readlengths were obtained for 500 nl, 4 μl, and 8 μl aliquots by simply adjusting the electrokinetic injection conditions. To ascertain the robustness of this methodology for genomic sequencing, 96 *Arabidopsis thaliana* subclones were sequenced, with a yield of 38 624 bases obtained from 500 nl aliquots versus 30 764 bases from standard scale reactions. To demonstrate 500 nl sample preparation, reactions were performed in fused-silica capillary reaction chambers using air-based thermal cycling. A readlength of 690 bases was obtained for the polymerase chain reaction product of an *Arabidopsis* subclone without modifying the reagent concentrations, post-reaction processing or electrokinetic injection workflow. These results demonstrated the fundamental feasibility of small-volume DNA sequencing for high-throughput capillary electrophoresis.

Keywords: Electrokinetic injection; Capillary array electrophoresis; DNA

1. Introduction

The anticipated completion of the human genome sequence by 2003, the continued need to sequence the complete genomes of model organisms, the detection of single nucleotide polymorphisms (SNPs), and the growth of environmental genomics will require significant expansion of sequencing throughput and reductions in the associated costs of DNA sequencing. The preferred technology to meet this growth in DNA sequencing is capillary array electrophoresis (CAE) [1]. CAE uses an array of narrow-bore, gel-filled capillaries for high-speed, high-resolution separations of multiple sequencing samples per run [2–6]. CAE has the advantages of automated gel loading, sample injection, and data-

*Corresponding author. Tel.: +1-408-737-3159; fax: +1-408-737-4808.

E-mail address: sjovanovich@mdyn.com (S.B. Jovanovich).

0021-9673/00/$ – see front matter
PII: S0021-9673(00)00459-3

processing of discrete sample lanes [6]. Because only a fraction of the prepared sample is consumed during electrokinetic injection, low-volume sample preparation could significantly reduce sequencing costs while maintaining the high-throughput requirements of genomic sequencing. Towards this goal, we investigated the preparation, purification, injection, and analysis of sub-microliter dye–primer DNA sequencing reactions for capillary array electrophoresis.

The analysis of sub-microliter DNA sequencing samples by capillary array electrophoresis is possible because of the efficiency of electrokinetic injection. Currently, most sequencing protocols generate pmol of fluorescently labeled fragments, in large excess of the amol detection limits in capillary electrophoresis [6]. In electrokinetic injection, the amount of sample loaded onto the capillary depends on the injection time and buffer conductivity [7]. By exploiting the difference in the ionic strength of the capillary running buffer and the sample buffer, the sample can be effectively concentrated during the injection into a narrow ''field-amplified'' injection plug [8,9]. Furthermore, longer injection times can be used without compromising the separation efficiency. However, electrokinetic injection causes DNA sequencing by CAE to be sensitive to contaminants in the sample such as template DNA, buffer anions and other impurities.

The sensitivity of CAE to sample constituents requires reproducible and robust purification methods that are compatible with high-throughput, automated DNA sequencers. Typically, sequencing samples are precipitated in ethanol and reconstituted in a formamide–EDTA mixture prior to injection [10]. However, any variability of residual template DNA and contaminating ions can reduce sample-to-sample reproducibility [11]. Improvements in injection efficiency have been achieved by the elimination of EDTA in the resuspension buffer [9] and by employing alternative cleanup protocols such as magnetic-bead capture [12]. Recently, a method for the removal of both template DNA and contaminating salts was developed that increased overall injection efficiency and minimized capillary failures caused by template overloading [13]. This cleanup protocol, along with changes in the sieving matrix and capillary operating conditions, resulted in the routine sequencing of 1000 bases in 60 min run times [14].

In addition to improving the reliability of CAE for DNA sequencing, several groups have investigated strategies for integrating sample preparation and for reducing the reaction volume. A flow-through device using air-thermal cycling and in-line size-exclusion chromatography for sample cleanup prior to capillary electrophoresis has been developed [15]. An integrated, multi-capillary flow-through system using a novel freeze–thaw valving technique was used for direct genotyping of blood samples [16,17], and for multiplexed DNA sequencing from single bacterial colonies [18]. An automated sample-processing system using sub-microliter piezo-dispensing into glass capillaries followed by air-thermal cycling has been developed [19]. For small-scale reactions, a solid-phase nanoreactor has been described for preparing 64 nl sequencing reactions from DNA bound to a capillary surface using biotin–streptavidin affinity [20]. These methods have demonstrated preparation and analysis of small reaction volumes by capillary electrophoresis, but require further improvement for compatibility with high-throughput CAE.

To develop high-throughput methods for the analysis of sub-microliter DNA sequencing reactions, the post-reaction processing and analysis of 500 nl aliquots from dye–primer reactions were investigated. Sub-microliter aliquots from 10 μl reactions were shown to have adequate peak heights and comparable readlengths relative to full-volume reactions. These results were validated by showing the injection and analysis of genomic subclones prepared from 10 μl reactions and from 500 nl aliquots were similar. In addition, cycle-sequencing reactions performed in 500 nl fused-silica capillaries demonstrated the ability to prepare and analyze small volume reactions. The preparation of 500 nl volumes was amenable to high-throughput analysis requiring no changes to the normal reagent concentrations or the instrument workflow.

2. Experimental

2.1. Chemicals

Fluorescently labeled energy-transfer primers, ThermoSequenase, 7-deaza-GTP sequencing kits, shrimp alkaline phosphatase, exonuclease I, and formamide were obtained from Amersham Life

Sciences (Piscataway, NJ, USA). Isopropanol and magnesium chloride were purchased from Sigma (St. Louis, MO, USA); EDTA and electrophoresis grade agarose from Gibco Life Technologies (Frederick, MD, USA); and Sybr Green II from Molecular Probes (Eugene, OR, USA). All solutions were prepared with deionized water from a Milli-Q water purification system (Millipore, Worchester, MA, USA). Single stranded M13mp18 DNA was prepared and purified using a standard protocol [21], and stored as a stock solution of 200 ng/μl in 0.1×TE buffer. Glycerol stock solutions of previously sequenced *Arabidopsis thaliana* BAC subclones (Genbank accession number AF077407, NID g3319339) were obtained from the Washington University Genome Sequencing Center (St. Louis, MO, USA). All other reagents were electrophoresis or analytical grade and used without further purification.

2.2. DNA sequencing chemistry

Dye–primer sequencing reactions were performed by combining template DNA with a dideoxynucleotide mixture containing an emission-specific primer. The color-specific primers were based on the M13–40 FWD primer (5′-FAM-GTTTTCCCAGT*CACGACG-3′) [22], with 5-carboxyfluorescein, FAM, as the donor dye, and a termination specific acceptor dye attached to the indicated thymine (T*). The thymine was labeled with FAM for ddC-terminated reactions (C-FAM), 6-carboxyrhodamine for ddA reactions (A-REG), *N*,*N*,*N*′,*N*′-tetramethyl-5-carboxyrhodamine for ddG reactions (G-TMR), and 5-carboxy-X-rhodamine for ddT reactions (T-ROX). A master mix for 100 dye–primer sequencing reactions was prepared by combining 65 μl reaction buffer [Amersham Pharmacia Biotech (APB), part 79607], 100 μl dye–primer solution (either 1 μ*M* T-ROX, 1 μ*M* G-TMR, 0.5 μ*M* A-REG, or 0.5 μ*M* C-FAM), 100 μl of the corresponding deoxy- and dideoxynucleotide mix (APB part 79591), 10 μl of enzyme (32 U/μl ThermoSequenase with pyrophosphatase, APB part E79000Y), and 225 μl filtered deionized water. This solution was stable at −20°C for up to 3 weeks. The sequencing cocktail was mixed with an equal volume of 20 ng/μl M13mp18 for most of the cycle-sequencing reactions. Thermal cycling reactions were performed in 0.20 ml reaction tubes (MJ Research, Watertown, MA, USA) using a PTC-200 DNA engine containing a 48- and 96-well adapter block. Cycle sequencing of M13mp18 was performed with 30 cycles of 95°C for 10 s, 50°C for 15 s, and 72°C for 60 s. Reaction products were pooled into the A-REG reaction tube and aliquoted for further processing.

2.3. Polymerase chain reaction (PCR) amplification and DNA sequencing of Arabidopsis thaliana subclones

Random subclones generated from a bacterial artificial chromosome (BAC) were inserted into M13 and PCR amplified followed by enzymatic removal of nucleotides and primers [23,24]. A 1 μl aliquot of glycerol stock template was added to 24 μl of a PCR mix which contained 5 pmol each of M13–100 universal (5′-GCTATTACGCCAGCTGGCGA-3′) and M13–100 reverse (5′-ATGCAGCTGGCACGACAGGT-3′) primers (Gibco Life Technologies); 200 μ*M* each dATP, dCTP, dGTP and dTTP; 1.25 U of AmpliTaq (Sigma, 5 U/μl); and 1 × GeneAmp buffer (Sigma). Reactions were cycled 35 times with denaturing at 95°C for 5 s, annealing at 55°C for 10 s, and extension at 72°C for 60 s. After PCR amplification, 9.2 μl of an Exonuclease I/shrimp alkaline phosphatase reaction mix was added to the PCR products and incubated at 37°C for 30 min followed by enzyme deactivation at 85°C for 15 min. The treated PCR products were diluted to 80 μl with water.

Full-volume dye–primer cycle sequencing reactions were performed in a CyclePlate-384 (Robbins Scientific, Sunnyvale, CA, USA) which had 96 reaction wells divided into four quadrants per well with molded dividers. Five microliters of the corresponding dye–primer reagents were added by pipetting the appropriate reagents into one of four-quadrants per well. Template DNA, either purified M13 or purified PCR products, were added using a Hydra-96 microdispenser and split-well plate-positioner (Robbins Scientific) to a final volume of 10 μl. The split-well plate was sealed with Microfilm Plate Sealer A (MJ Research) prior to cycle sequencing at 95°C for 5 s, 55°C for 15 s, and 72°C for 60 s for 25 cycles. The reaction products were aspirated from the well quadrants using the Hydra-96 microdispenser and dispensed to a clean 96-well plate (Robbins Scientific).

2.4. Preparation of 500 nl reaction aliquots

The reaction aliquots corresponding to 500 nl per dye–primer reaction were obtained by transferring 2 μl from the pooled reaction products to a new 96-well plate. The 2 μl aliquots were diluted to 5 μl by the addition of an injection buffer consisting of 95% formamide–1.6 m*M* EDTA; diluted samples were injected without further purification (direct injection). In another method, samples were precipitated by the addition of 8 μl ethanol to the pooled 500 nl aliquots. Precipitated samples were centrifuged for 30 min at 1900 *g*, followed by an inversion of the plate and centrifugation for 30 s at 50 *g* to remove the supernatant. The samples were resuspended in 5 μl of an injection buffer of 80% formamide–1 m*M* EDTA, followed by gentle vortexing for 20 s, and storing on ice prior to injection.

The recovery of template DNA was determined by agarose gel electrophoresis using a 0.7% agarose gel with 1 × Tris–acetate–EDTA buffer, pH 8.0, followed by staining with Sybr Green II (Molecular Probes). Samples were separated for 40 min at 15 V/cm and imaged using a two-dimensional fluorescence scanner (FluorImager, Molecular Dynamics, Sunnyvale, CA, USA). Conductivity measurements were performed with a Model 125 Orion conductivity meter (Orion Research, Boston, MA, USA) and a MI-905 conductivity electrode. Duplicate measurements were made for a minimum of three samples.

2.5. Cycle-sequencing of 500 nl reactions in capillaries

Dye–primer sequencing reactions were performed within uncoated 4 cm×150 μm I.D.×360 μm O.D. fused-silica capillaries (Polymicro Technologies, Phoenix, AZ, USA). Four capillaries per sample were filled by capillary action after dipping the capillary end into the appropriate dye–primer reaction mix. The capillary ends were sealed with syringe septa (Model 75810, Hamilton, Reno, NV, USA) and placed in the foam capillary holder of an air-based thermal cycler (Rapid Cycler, Idaho Technology, Salt Lake City, UT, USA). The capillaries were arranged in 9 mm spacing and enough length of the capillary placed below the foam holder so that 3 cm (~500 nl) was exposed to the air chamber of the cycler. A cycle sequencing protocol was adapted from Idaho Technology literature using shorter denature and annealing times, such that amplification was achieved using 30 cycles of 95°C for 2 s, 55°C for 2 s, and 60°C for 60 s. After the reaction, the capillaries were removed from the holder and the reaction products were dispensed into 40 μl of 80% aqueous isopropanol using a low-pressure flow of nitrogen gas through a pipette tip. The samples were centrifuged, resuspended, and injected into MegaBACE as previously described. The capillaries were not pretreated with any wash solutions prior to use and disposed of after a single reaction.

2.6. MegaBACE capillary array electrophoresis system

DNA sequencing runs were performed with MegaBACE, a 96-capillary array electrophoresis instrument (Molecular Dynamics) using scanning confocal laser-induced fluorescence [2,5] and motorized beamsplitters and bandpass filters (Omega Optical, Battleboro, CT, USA) for four-color detection with two photomultiplier tubes (PMTs). The confocal objective scanned the 96 capillaries at 1.7 Hz and peaks were detected by capillary position across the array and binned according to scan direction. In the forward scan, the fluorescence emission was split with a 540DRLP beamsplitter and filtered with a 555DF15 filter for detection of A-REG on PMT-1 and a 525DF15 bandpass filter for detection of C-FAM on PMT-2. In the reverse scan, a dichroic 600DRLP beamsplitter was used to transmit fluorescence emission to a longpass 610LP filter for detection of T-ROX on PMT-1 and to reflect emission to a 580DF20 filter for detection of G-TMR on PMT-2.

Separations were performed in 62 cm×75 μm I.D.×200 μm O.D. fused-silica capillaries with a working separation distance of 40 cm. Electroosmotic flow was reduced by Grignard coupling of a vinyl group to the capillary surface and acrylamide polymerization [25]. The capillaries were filled with a fresh solution of 3% linear polyacrylamide (MegaBACE Long Read Matrix, Amersham Life Sciences), which was pumped through the capillaries under high-pressure from the anode chamber to

individual wells of a 96-well buffer plate contained in the cathode chamber. Each well was filled with 100 μl of running buffer which was prepared with 30 mM Tris, 100 mM TAPS, and 1 mM EDTA at pH 8.0. The LPA matrix was equilibrated to relax strand-stretching in the capillaries for 20 min followed by pre-electrophoresis for 5 min at 180 V/cm. The capillary ends on the cathode side were rinsed with double-distilled water to remove residual LPA from the capillary and electrode ends prior to sample injection. DNA sequencing samples were electrokinetically injected at constant voltage from a 96-well microtiter plate according to the specified conditions. After injection, the capillary ends were rinsed with water, the buffer plate was placed in the cathode chamber, and the electrophoresis run was commenced.

2.7. Base-calling software and data processing

Four-color electropherograms were processed using the Sequence Analyzer base-calling software package (Molecular Dynamics). The base-calling algorithm used data-processing steps analogous to those described by Yin et al. with baseline subtraction, spectral separation of cross-talk from other channels, signal intensity normalization, mobility shift correction, and assignment of a letter code corresponding to the fluorescent-dye of the terminated reaction [26].

The Phred algorithm was used to assess base-calling accuracy [27,28]. Error discrimination for correct base-calls was determined by measuring peak spacing and peak resolution, determining the ratio of called to uncalled peaks in a seven-peak window, and determining the same ratio in a three-peak window. From this algorithm, each base was assigned a quality value, q, according to Eq. (1):

$$q = -10\log_{10}(P) \tag{1}$$

where P represents the estimated error probability for that base-call. Thus, a base-call having a probability of 1/1000 being incorrect was assigned a quality value of 30. The error probabilities were fit to MegaBACE sequencing data by comparing the base identity error against known sequences [29]. Sample readlengths were reported as the range of Phred20 scores (1/100 error rate) between the start and end of the trace.

Injection efficiency was evaluated by measuring the peak height of baseline-subtracted T-ROX peaks across the trace. Counting from the end of the FWD −40 priming site (base 6306 in M13mp18), average peak heights were determined from the relative fluorescence of the early (bases 60, 68, and 74), middle (bases 295, 297, and 298), and late (bases 589, 596, and 605) parts of the sequence trace. The signal-to-noise ratios and peak widths were determined using IgorPro 3.14 (Wavemetrics, Lake Oswego, OR, USA).

3. Results and discussion

Analysis of sub-microliter DNA sequencing reactions by capillary electrophoresis depends on the efficiency of electrokinetic injection. In electrokinetic injection, the quantity, Q_i, of analyte introduced into the capillary under an applied electric field can be described by:

$$Q_i = [(\mu_i + \mu_{eo})\pi r_c^2 V/L]C_i t_i \tag{2}$$

where μ_i and μ_{eo} are the electrophoretic and electroosmotic mobilities of species i; r, the capillary radius; V, applied voltage applied; L, capillary length; C, concentration of analyte; and t, injection time. With electrokinetic injection of DNA into a gel-filled capillary, a biased injection towards low molecular weight DNA is observed because of the differences in the free solution and gel matrix mobility. This bias and the injection parameters that affect injection into gel-filled capillaries can be described by [30]:

$$m_{DNA} = C_{DNA} \cdot \frac{\mu_{buffer}}{\kappa_{buffer}} \cdot I\Delta t \cdot \left(1 - \frac{\mu_{eo}}{\mu_{gel}}\right) \tag{3}$$

where m_{DNA} is the mass of DNA transferred into the capillary; C_{DNA}, the concentration of DNA in solution; μ_{buffer}, the mobility of DNA in the analyte buffer; κ_{buffer}, the buffer conductivity; I, the current; and Δt, the injection time. Preferential loading of smaller fragments occurs because of the counter-flow component of electroosmotic mobility, μ_{eo}, and the higher mobility of small DNA molecules in the sieving matrix, μ_{gel}. According to Eq. (3), the injection of sub-microliter aliquots depends simply

on the injection time and sample conductivity; controlling these two parameters should allow for efficient injection of reduced sample volumes.

Since sequencing reactions are typically prepared in high salt solutions, improving the injection efficiency requires dilution of the sample, alcohol precipitation, or other purification methods to remove ions and reduce the conductivity. Furthermore, the optimal buffer solution for diluting sub-microliter aliquots or for resuspending precipitated reaction products had not been determined. As a way of characterizing the impact of excess ions on injection efficiency, standard reaction samples were resuspended in formamide containing a range of EDTA concentration. These experiments were used to determine an optimal buffer for studying the analysis of sub-microliter reactions by capillary electrophoresis.

3.1. Effect of EDTA on injection efficiency

The effect of EDTA, a common additive in formamide resuspension buffers [21], on injection efficiency was determined by comparing signal-to-noise (S/N) ratios and readlengths for different EDTA concentrations in formamide. As shown in Fig. 1, increasing the concentration of EDTA in the injection buffer reduced the S/N ratio and, when the concentration of EDTA exceeded 20 m*M*, the number of bases per sample. The readlengths, which averaged 540±100 bases from 0.2 to 20 m*M* EDTA, were nearly constant despite the decrease in S/N ratio. Above 20 m*M* EDTA, however, the decrease in the signal-to-noise ratio reduced the base-calling accuracy. Small fragment bias in the sample injection was evident from the four-fold decrease in the signal from the early bases when the EDTA was increased from 0.5 to 5 m*M*. The signal from the middle and late bases remained relatively constant. Above 5 m*M* EDTA, the total amount of DNA loaded into the capillary decreased, reflecting a decrease in injection efficiency under these conditions.

Although an increase in EDTA concentration correlated with a decrease in injected sample, reducing the concentration of EDTA below 1 m*M* resulted in sample failure. Sample failure was characterized by a decrease (ca. 40%) in capillary current during electrophoresis or by an undetected sample. Current reduction and loss of resolution in capillary electrophoresis for DNA sequencing has been associated with the injection of template DNA [10,31]. The presence of small ions, such as EDTA, in the loading buffer reduces the deleterious effect of template DNA [11]. Given the strong dependence of injection reproducibility and peak height on the EDTA concentration, 1 m*M* EDTA in 80% formamide was used for subsequent analysis of sub-microliter aliquots. This concentration of formamide was determined in a related experiment where a two-fold increase in peak height was obtained by adjusting the formamide concentration from 95 to 80%, with the signal intensity remaining constant between 80 and 20% formamide (data not shown). The use of 1 m*M* EDTA in 80% formamide also had the benefit of reducing the differences between the signal intensities in the early, middle, and late parts of the electropherogram.

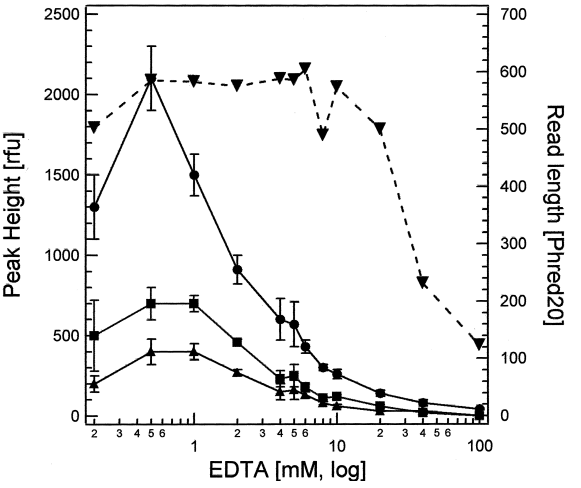

Fig. 1. Effect of EDTA concentration on signal-to-noise ratio and Phred20 readlength for the analysis of four-color M13mp18 sequencing standards. The signal-to-noise ratio is indicated by solid lines corresponding to (●) early peaks, (■) middle peaks, and (▲) late peaks with values indicated on the left axis. The readlength (▼) is indicated by the dotted line with values on the right-hand axis. Four-color sequencing standards were aliquoted to a loading plate, mixed to yield the indicated concentrations of EDTA, and injected at 10 kV for 20 s. Samples were separated at 10 kV for 90 min.

3.2. Direct injection of sub-microliter aliquots.

To ascertain the possibility of eliminating post-

reaction processing for the analysis of sub-microliter volumes, 500 nl aliquots of each dye–primer were diluted and directly injected without additional post-reaction cleanup (direct injection). The impact of injection time on peak heights and readlengths was determined for 500 nl reaction aliquots, obtained by removing 2 μl (500 nl per dye–primer reaction) from the pooled reaction products of 10 μl cycle sequencing reactions. The pooled 500 nl aliquots were diluted to 5 μl with the addition of 3 μl injection buffer (96% formamide containing 1.6 m*M* EDTA). The samples were electrokinetically injected under constant voltage conditions, 30 V/cm, for injection times ranging from 10 to 80 s. A plot of S/N ratio and readlength versus injection time is shown in Fig. 2. The S/N ratio increased linearly from 8 ± 3 with a 10 s injection to 70 ± 25 with an 80 s injection consistent with the relationship of mass transfer to injection time (Eq. (3)).

The increase in injection time improved peak heights without compromising the peak parameters that affect readlength. For an injection time of 20 s, the width of the peak corresponding to 298 bases was 0.033 cm at 20 s, and remained constant between 40 and 80 s, with values of 0.070 cm at 40 s, 0.061 cm at 60 s, and 0.062 cm at 80 s. The S/N ratio was insufficient for peak identification from a 10 s injection, however, a 20 s injection afforded a modest increase in the S/N ratio resulting in a readlength of 525 ± 50 bases (99% accuracy). At an 80 s injection, the readlength improved slightly to 580 ± 20 bases indicating constant peak spacing, peak width, and adequate S/N ratio for these injection conditions.

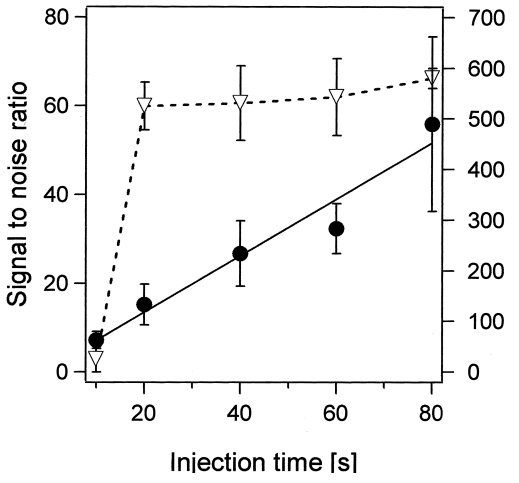

Fig. 2. Effect of injection time on the signal-to-noise ratio and readlength for the analysis of 500 nl aliquots diluted in formamide and injected without further purification. The signal-to-noise ratio (●) is indicated by a solid line with values on the left axis, and the readlength (▽) with a dotted line and values on the right axis. Sequencing samples were injected at 2 kV for indicated time, with electrophoresis at 8 kV for 120 min. Data points represent the average and standard deviation of nine samples at each injection condition.

These results confirmed the efficiency of electrokinetic injection for the analysis of sub-microliter aliquots without additional post-reaction processing. Dilution and direct injection has been demonstrated as an effective technique for capillary electrophoresis of dilute analytes [7]. By lowering the conductivity of the sample relative to the running buffer, stacking of analytes into the capillary was obtained; the amount of stacking increased with increased injection times. Furthermore, increasing the injection time improved the S/N ratio without loss of peak resolution [8]. Dilution of the sample and direct injection had the advantage of reducing the post-reaction processing time compared to alcohol precipitation. Since the mass transfer efficiency depends also on sample conductivity (Eq. (3)), an alcohol precipitation step should improve the signal-to-noise ratio by removing excess competing anions.

3.3. Recovery and conductivity of precipitated 500 nl aliquots

The most widely used method for concentrating nucleic acids is ethanol precipitation [32]. The precipitate is usually recovered by centrifugation, followed by decanting, and resuspension in an appropriate buffer at the desired concentration. Quantitative recovery of DNA down to 20 pg precipitated in ethanol at 0°C has been demonstrated [33], which is well below the ca. 20 to 80 ng of DNA in a 500 nl dye–primer sequencing aliquot.

The recovery of template DNA from an alcohol precipitation and the resulting conductivity was determined to evaluate the application of this technique to the analysis of 500 nl aliquots. Four 500 nl aliquots from a 15 ng/μl M13mp18 reaction were pooled, precipitated with the addition of 40 μl of 80% isopropanol, banded by agarose gel electro-

phoresis, and quantified by fluorescent staining. The amount of DNA estimated from the 500 nl precipitated aliquots was 25±2 ng template DNA compared to 28±1 ng template DNA from an unprecipitated sample. The conductivity of a precipitated 500 nl aliquot was determined by reconstituting the products in 5 μl of 80% deionized formamide–1 m*M* EDTA, and diluting with the addition of 45 μl ddH_2O. These samples had a two-fold decrease in conductivity, 50±4 μS for ethanol precipitated samples compared to 96±5 μS for unprecipitated samples.

The high sample recovery, as well as the two-fold reduction in sample conductivity resulted in more efficient sample injections (Eq. (3)). For a 10 s injection time, precipitated samples had a signal-to-noise ratio of 16±3 and readlengths of 450±80 bases compared to a *S*/*N* ratio of 7±2 and a non-determined readlength for unprecipitated samples (the *S*/*N* was too low for adequate peak identification). When the injection time was doubled to 20 s, the *S*/*N* ratio and readlength increased to 32±10 and 610±33 bases, respectively. For unprecipitated samples injected at 20 s, the *S*/*N* ratio was only 15±2 and the readlength was only 480±150 bases. Removing ions from the solution through alcohol precipitation, and thus lowering the conductivity (Eq. (3)) improved the *S*/*N* ratio and readlength. This improvement favored the implementation of alcohol precipitation for the cleanup of microaliquots at the expense of a somewhat lengthy precipitation step compared to direct injection.

3.4. Comparison of precipitation methods for 500 nl volumes

Since several precipitation protocols exist for sample cleanup of large volume reactions [32], five methods were compared for the analysis of 500 nl, 4 μl, and 8 μl aliquots. Aliquots were precipitated with either ethanol (EtOH) or isopropanol (IPA) with and without a 70% alcohol wash. From a 15 μl dye–primer reaction, the reaction aliquots were distributed to individual tubes and precipitated by adding four volumes of the indicated alcohol. The wash step, if used, was performed by the addition of the same volume of a 70% solution of the indicated alcohol. After precipitation, the samples were resuspended in 5 μl of 80% formamide–1 m*M* EDTA. The injection conditions were adjusted to compensate for the differing concentrations of fluorescent fragments. This was necessary because using identical injection conditions either lowered the yield for the smallest aliquot or caused sample overloading and capillary failure for the larger volume samples.

The percent success and average readlength was equivalent for 500 nl, 4 μl, and 8 μl aliquots and five precipitation methods. The readlengths, plotted in Fig. 3 by volume and precipitation method, were 630±25 bases for 500 nl aliquots, 610±35 bases for 4 μl aliquots, and 610±35 bases for 8 μl aliquots. The readlengths were essentially independent of either reaction volume or precipitation method. By simply adjusting the injection conditions, differences in starting reaction volume were normalized. This comparison demonstrated the compatibility of conventional protocols for the reproducible analysis of sub-microliter aliquots.

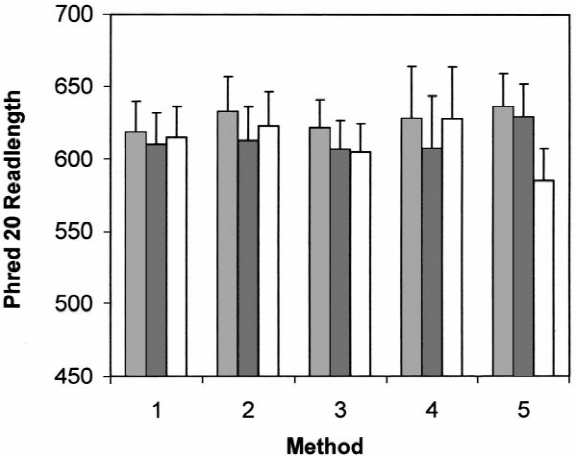

Fig. 3. The effect of reaction volume and alcohol precipitation method for 500 nl (left bar), 4 μl (center bar), and 8 μl (right bar) reaction volumes. The precipitation methods were (1) 80% ethanol, (2) 80% isopropanol, (3) 80% ethanol with a 70% ethanol wash, (4) 80% IPA with a 70% IPA wash, (5) 80% IPA with a 70% ethanol wash. Inverted centrifugation was used to remove the supernatant after each step. After decanting, the samples were resuspended in 5 μl of 80% formamide, 1 m*M* EDTA and injected for 20 s at 10 kV for 500 nl aliquots, 10 s at 8 kV for 4 μl aliquots, and 10 s at 4 kV for 8 μl aliquots; electrophoretic separation was performed for 90 min at 9 kV.

3.5. Sequencing 500 nl aliquots of Arabidopsis thaliana subclones

The use of inexpensive and dirty template preparation techniques for genomic sequencing requires the demonstration of any new sequencing technologies with ''real-world'' templates [34,35]. To test post-reaction processing and analysis of 500 nl aliquots from genomic samples, 96 PCR products of *Arabidopsis thaliana* subclones were sequenced. The subclones were inserted into M13 from a randomly sheared bacterial artificial chromosome (BAC) and PCR amplified from glycerol stock solutions. The samples were then enzymatically purified and sequenced with 10 μl dye–primer reactions. Aliquots from the pooled reaction products were precipitated, injected and compared to the samples prepared from 10 μl reactions by percent success, readlength, and yield (the total number of Phred20 bases obtained per sample set). The yield of bases per plate is an important indicator of sequencing coverage, with greater readlengths and yields reducing the number of samples required for shotgun sequencing [36,37].

Both standard reaction volumes and 500 nl aliquots yielded a large number of high quality bases indicating the reliability of small-volume analysis for genomic sequencing. The PCR yield and histogram of percent success versus readlength window is shown in Fig. 4. For 10 μl reactions, the percentage of samples with a readlength above 100 bases was 74%, the average readlength was 430±100 bases, and the sum of Phred20 bases for all samples was 30 754. For 500 nl aliquots from the same reactions the percentage of samples with a readlength above 100 bases was 82%, the average readlength was 480±100 bases, and the Phred20 sum was 38 264 bases. The failed and short readlength samples in both traces were due to short-insert subclones (nine samples) and failed PCR reactions (six samples). The higher average readlength for the sub-microliter aliquots was due to lack of overloading observed for several of the 10 μl samples. Reducing the injection voltage corrected the sample overloading and increased the yield from 10 μl analysis to 35 000 bases and the percent success to 78%. These results clearly demonstrated the ability of CAE to analyze sub-microliter aliquots without compromising the data quality obtained with larger volume reactions.

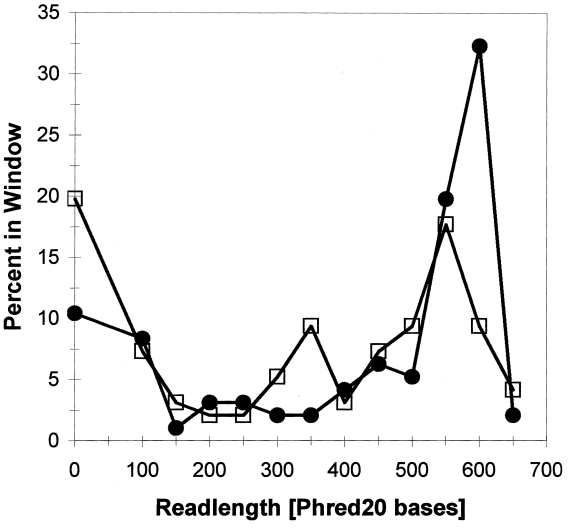

Fig. 4. Histogram of percent versus readlength window for 500 nL (●) and 10 μL (□) reactions. The histogram was derived from the sequencing of 96 samples prepared from PCR products of random subclones from an *Arabidopsis thaliana* bacterial artificial chromosome. Injection conditions: 4 kV for 10 s for 10 μL reactions, 8 kV for 20 s for 500 nL aliquots.

3.6. DNA sequencing in 500 nl capillaries

To demonstrate sub-microliter sample preparation, a subset of samples were prepared in 500 nl fused-silica capillary reaction chambers using air-based thermal cycling for amplification. The use of thin-walled small diameter capillaries allowed for rapid temperature equilibrium and improved thermal cycling rates [38]. In addition, the capillaries are inexpensive reaction chambers and was easily sealed to prevent evaporative loss. In these experiments, only 2 mm of capillary, ca. 7% of the volume, was lost to evaporation. To perform the reactions, the capillaries were filled by dipping the ends of the capillaries into pre-mixed dye–primer reactions containing reagents and template DNA. The capillaries were sealed with septa and placed in an air-based thermal cycler. After thermal cycling, the samples were dispensed into 0.20 ml reagent tubes containing 40 μl of isopropanol; the larger volume of alcohol prevented spattering of the sample on the walls of the reagent tube and retention of the sample on the end of the capillary. After post-reaction precipitation,

the samples were resuspended in 80% formamide with 1 m*M* EDTA, and analyzed by MegaBACE. As controls, the 500 nl aliquots of the full-volume reactions were analyzed under similar conditions.

Preparation of the reaction in a 500 nl capillary produced comparable results to a microaliquot sample. The readlength was 690 Phred20 bases for the capillary-prepared sample compared to 710 Phred20 bases from the 500 nl aliquots of the full-volume reactions. The base-called sequencing trace from the capillary reaction is shown in Fig. 5. Although the readlength was equivalent, the peak height in the capillary reaction was reduced four-fold relative to the microaliquot sample. The reduced peak height may be caused by sample handling differences (four capillaries versus one pooled aliquot), differences between air- and peltier-based thermal cycling, and differences in the surface area-to-volume ratio between the capillaries and polypropylene tubes. These effects are being further investigated.

Genomic sequencing and analysis of 500 nl dye–primer reactions was demonstrated for capillary array electrophoresis. The excellent readlengths coupled with the approximately 20-fold reduction in sample

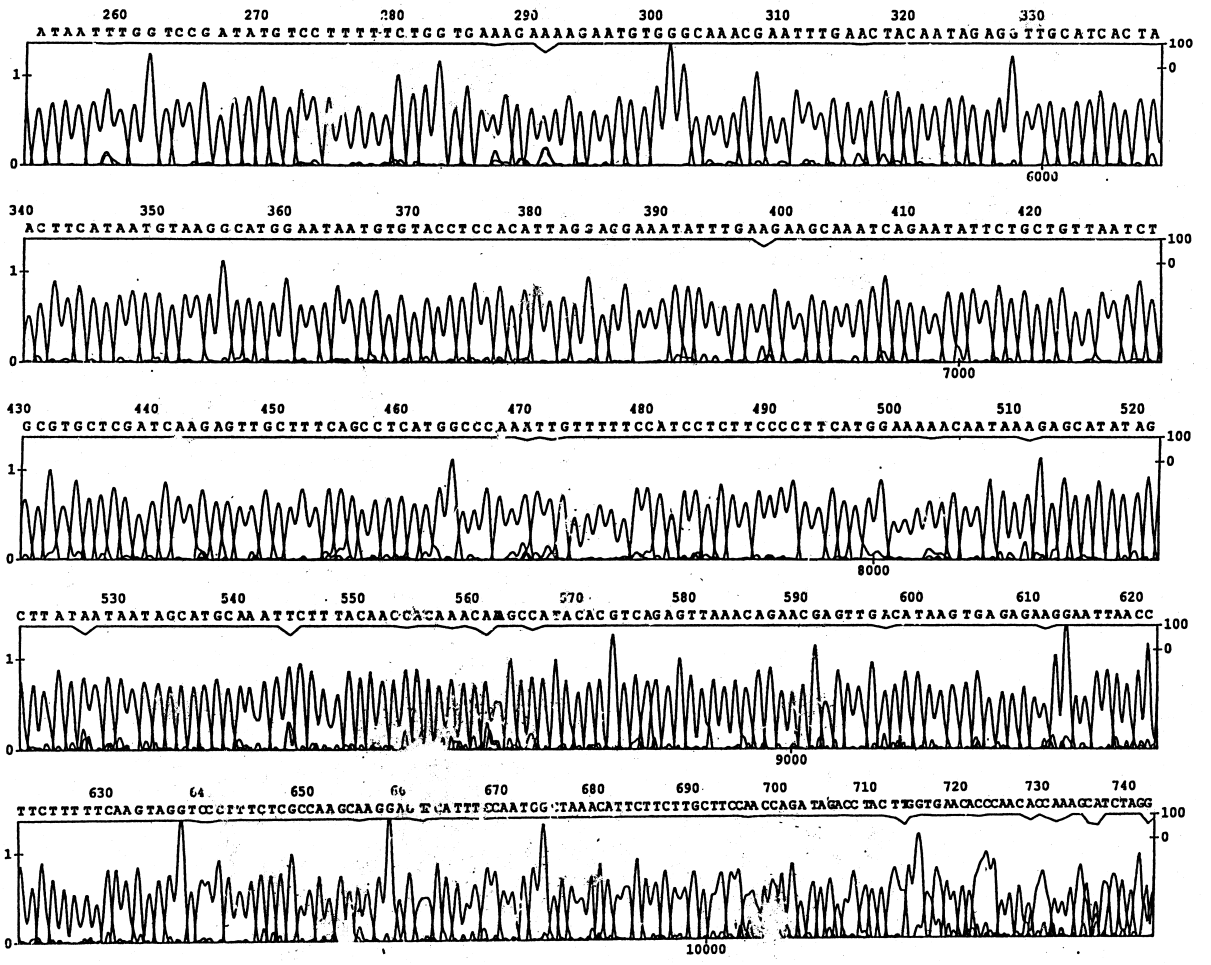

Fig. 5. Base-called electropherogram of an *Arabidopsis* BAC subclone prepared in a 500 nl cycle-sequencing reaction. Individual dye-primer reactions were prepared in 3 cm sections of 150 μm i.d. fused-silica capillaries using air-thermal cycling, precipitated in 80% IPA and resuspended in 5 μl of 80% formamide–1 m*M* EDTA prior to injection for 40 s at 2 kV and separation for 120 min at 8 kV. A Phred20 readlength of 690 bases was obtained.

volumes, and therefore reagent consumption, suggests that the preparation of sub-microliter samples may be an enabling approach for high throughput sequencing. Current work focuses on implementing these approaches in a format compatible with high throughput sample preparation.

Acknowledgements

The authors gratefully acknowledge support by NIH grant R01HG01775. Dr. John Bashkin, Dr. Nels Olson, and Dr. Oscar Solas-Solano provided helpful comments and suggestions. The *Arabidopsis* BAC subclones, prepared at the Washington University Genome Sequencing Center, were kindly provided by Dr. Elaine Mardis.

References

[1] F.S. Collins, A. Patrinos, E. Jordan, A. Chakravarti, R. Gesteland, L. Walters, Science 282 (1998) 682.
[2] X.C. Huang, M.A. Quesada, R.A. Mathies, Anal. Chem. 64 (1992) 2149.
[3] K. Ueno, E.S. Yeung, Anal. Chem. 66 (1994) 1424.
[4] S. Takahashi, K. Murakami, T. Anazawa, H. Kambara, Anal. Chem. 66 (1994) 1021.
[5] M. Marsh, O. Tu, V. Dolnik, D. Roach, N. Solomon, K. Bechtol, P. Smietana, L. Wang, X. Li, P. Cartwright, A. Marks, D. Barker, D. Harris, J. Bashkin, J. Cap. Elec. 4 (1997) 83.
[6] I. Kheterpal, R.A. Mathies, Anal. Chem. 80 (1999) 31A.
[7] Z.K. Shihabi, in: J.P. Landers (Ed.), Handbook of Capillary Electrophoresis, CRC Press, Boca Raton, FL, 1997, p. 457.
[8] R.L. Chien, D.S. Burgi, J. Chromatogr. 559 (1991) 141.
[9] D. Figeys, H. Ahmadzedeh, E. Arriaga, N.J. Dovichi, J. Chromatogr. 744 (1996) 324.
[10] J.A. Luckey, T.B. Norris, L.M. Smith, J. Phys. Chem. 97 (1993) 3067.
[11] O. Salas-Solano, M.C. Ruiz-Martinez, E. Carrilho, L. Kotler, B.L. Karger, Anal. Chem. 70 (1998) 1528.
[12] T. Hultman, S. Stahl, E. Hornes, M. Uhlen, Nucleic Acids Res. 17 (1989) 4937.
[13] M.C. Ruiz-Martinez, O. Salas-Solano, E. Carrilho, L. Kotler, B.L. Karger, Anal. Chem. 70 (1998) 1516.
[14] O. Salas-Salano, E. Carrilho, L. Kotler, A.W. Miller, W. Goetzinger, Z. Sosic, B.L. Karger, Anal. Chem. 70 (1998) 3996.
[15] H. Swerdlow, B.J. Jones, C.T. Wittwer, Anal. Chem. 69 (1997) 848.
[16] H. Tan, E.S. Yeung, Anal. Chem. 70 (1998) 4044.
[17] N. Zhang, H. Tan, E.S. Yeung, Anal. Chem. 71 (1999) 1138.
[18] Y. Zhang, H. Tan, E.S. Yeung, Anal. Chem. 71 (1999) 5018.
[19] N.A. Friedman, D.R. Meldrum, Anal. Chem. 70 (1998) 2997.
[20] S.A. Soper, D.C. Williams, Y. Xu, S.J. Lassiter, Y. Zhang, S.M. Ford, R.C. Bruch, Anal. Chem. 70 (1998) 4036.
[21] J. Sambrook, E.F. Fritsch, T. Maniatis, Molecular Cloning, 2nd ed, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1989.
[22] J. Ju, C. Ruan, C.W. Fuller, A.N. Glazer, R.A. Mathies, Proc. Natl. Acad. Sci. USA 92 (1995) 4347.
[23] X. Meltzer, Biotechniques 8 (1990) 142.
[24] E. Werle, C. Schneider, M. Renner, M. Volker, W. Fiehn, Nucleic Acids Res. 22 (1994) 4354.
[25] V. Dolnik, D. Xu, A. Yadav, J. Bashkin, M. Marsh, O. Tu, E. Mansfield, M. Vainer, R. Madabhushi, D. Barker, D.J. Harris, Microcol. Sep. 10 (1998) 175.
[26] Z. Yin, J. Severin, M.C. Giddings, W.-A. Huang, M.S. Westphall, L.M. Smith, Electrophoresis 17 (1996) 1143.
[27] B. Ewing, P. Green, Genome Res. 8 (1998) 186.
[28] B. Ewing, L. Hillier, M.C. Wendl, P. Green, Genome Res. 8 (1998) 175.
[29] P. Richterich, Genome Res. 8 (1998) 251.
[30] K. Kleparnik, Z. Mala, P. Bocek, J. Chromatogr. A 772 (1997) 243.
[31] H. Swerdlow, K.E. Dew-Jager, K. Brady, R. Grey, N.J. Dovichi, R. Gesteland, Electrophoresis 13 (1992) 475.
[32] J. Sambrook, E.F. Fritsch, T. Maniatis (Eds.), Molecular Cloning, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1989, p. E.10.
[33] D.J. Shapiro, Anal. Biochem. 110 (1981) 229.
[34] E.R. Mardis, Nucleic Acids Res. 22 (1994) 2173.
[35] M. Marra, T.A. Kucaba, R.H. Waterston, Nucleic Acids Res. 27 (1999) e37.
[36] E.S. Lander, M.S. Waterman, Genomics 2 (1988) 231.
[37] J.C. Roach, Genome Res. 5 (1995) 464.
[38] C.T. Wittwer, G.C. Fillmore, D.R. Hillyard, Nucleic Acids Res. 17 (1989) 4353.