Single-Stranded Conformational Polymorphism Analysis Using Automated Capillary Array Electrophoresis Apparatuses

BioTechniques 34:746-750 (April 2003)

ABSTRACT

We describe a new environment of a single-stranded conformational polymorphism (SSCP) analysis using automated capillary array sequencers (e.g., ABI PRISM® 3100 and 3700). In this environment, electrophoretic conditions, settings for instrument management, and software for data analysis are adjusted for SSCP analysis. Highly reproducible results are obtained with this new system, and fragments with mutations and/or polymorphisms in different capillaries or different runs can be reliably detected. The relative peak heights between alleles are quantitative and reproducible between runs, and so allele frequencies of single nucleotide polymorphisms can be accurately estimated by a pooled DNA strategy. The method allows unattended, low-cost, and quantitative SSCP analysis using instruments that are widely accessible.

INTRODUCTION

There has been an increasing need for an efficient and easily accessible method for the characterization of polymorphisms [e.g., single nucleotide polymorphisms (SNPs)] in many genomic regions by examining a large number of samples (1).

PCR-single-stranded conformational polymorphism (SSCP) has been widely used to detect polymorphisms/mutations because it is simple, highly sensitive, and cost effective (2,3). We previously developed a method, PLACE-SSCP, in which PCR products are post-labeled with fluorescent dyes, and labeled products are analyzed using an automated capillary electrophoresis system under SSCP conditions (4,5). However, the employed apparatus required manual loading of sample plates and was not suitable for unattended operation, which is essential in large-scale analyses.

Thus, we sought a SSCP environment where sample loadings are automated without human intervention and instrument operation/data processing is customized for SSCP.

We focused on the ABI PRISM® 3100 and 3700 (Applied Biosystems, Foster City, CA, USA), which are widely accessible as capillary array sequencers. Both apparatuses are potentially fit for a high-throughput SSCP analysis because of their automated sample handling, high-resolution electrophoretic separation, precise temperature control, and multicolor fluorescence detection. However, these capillary array apparatuses have been designed primarily for sequencing or microsatellite genotyping and are not readily usable for SSCP analysis. We found that electrophoretic run conditions had to be optimized for SSCP analysis, parameters for running the instruments needed modification, and using a dedicated software was required for efficient processing and interpretation of SSCP data.

Here we show a detailed procedure to redefine parameters of the run module of the associated software and to newly construct the color matrix. We also present the results of evaluation of reproducibility of positions and heights of the peaks of SSCP analysis in the new environments, and of the suitability of the method for allele frequency estimation by pooled DNA strategy, using fragments of various sequence contexts analyzed by both capillary array machines.

MATERIALS AND METHODS

DNA, PCR, and Fluorescence Labeling

Genomic DNA from anonymous Japanese individuals was used as templates for PCR. Sequence information of the primers for the analysis of some of the SNPs was obtained from the Human SNP Database of Whitehead Institute (6) and dbSTS (http://www.ncbi.nlm.nih.gov/dbSTS/). Primers for other SNPs were designed using Primer3 software (7). All primers (purchased from Amersham Biosciences, Tokyo, Japan) were made to carry either 5′-ATT or 5′-GTT for the post-labeling purposes (8). Primer sequences, product sizes, and GenBank® accession numbers (if applicable) of all sequence tagged sites (STSs) used in this work are available on request to authors.

PCR was performed in a 10-µL reaction volume following conditions as described previously (4). The thermal cycling profile was 1 min at 94°C for initial heating, followed by 40 cycles of 30 s at 94°C, 30 s at 60°C, and 1 min at 72°C. Post-PCR labeling was performed according to the previously described procedures (5).

Electrophoresis

Linear poly-N,N-dimethylacrylamide (PDMA) served as a sieving matrix. The polymer was synthesized as described elsewhere (4,9). The polymer was dissolved in 60 mM Tris, 70 mM 2-morpholinoethanesulfonic acid (MES), and 2 mM Na2EDTA (2× TME), filtered through Millipore®-G, 0.22 µm pore size filter (Millipore, Bedford, MA, USA) and degassed under vacuum for 10 min immediately before use. The concentration of PDMA in the capillary was 10% or 12% (w/v), as indicated.

Loading solution for PLACE-SSCP consisted of 1 µL post-labeled sample fragment, 0.5 µL TAMRA-labeled reference DNA, 0.5 µL purified GeneScan®-500 TAMRA Size Marker (Applied Biosystems), and 28 µL (ABI 3100) or 13 µL (ABI 3700) of 0.5 mM EDTA. The size marker was purified by ultra-filtration using Microcon® YM-50 (Millipore) to remove small fragments and salts, beforehand. The mixture of TAMRA-labeled DNAs served as an internal standard to calibrate variation of mobility among capillaries.

Electrophoresis was performed using 36-cm (ABI 3100) or 50-cm (ABI 3700) capillaries. In the case of ABI 3100, samples were injected at 1.5–2 kV for 10 s and separated at 15 kV for 50 min. As for the ABI 3700, sample injection was at 2–5 kV for 10 s, and separation was at 5 kV. Run temperature was set at 27°C for both instruments.

Parameters in the Software for Instrument Operation

The ABI 3100 has both heating and cooling functions that can strictly maintain the temperature of the cham-
Temperature for the electrophoresis was changed by altering run temperature specified in the run module, “GeneScan36 POP4DefaultModule”, by using the “Module Editor” included in the “Data Collection” software.

ABI 3700 is not equipped with a positive cooling device; thus, the running temperature is restricted to at least several degrees above ambient. We found that isothermal electrophoresis can be performed at 27°C using an ABI 3700, by controlling the room temperature to, for example, 22°C. Two modules control the temperature for the electrophoresis using this machine. “PreBatchRunInternalUseOnly” module, which defines pre-run heating, is in plain text format and can be edited with a text editor. In this module, the values that specified temperature for the capillary and cuvette (“45” and “35” in lines 33 and 37, respectively) were replaced by “27”. In addition, the run temperature in “GeneScan1_1DefaultModule” was changed using the Module Editor as described for the ABI 3100.

Construction of Color Matrix

The fluorophores we used in the present PLACE-SSCP analysis were R110, R6G, and TAMRA. As none of the available color matrices matched the dye set, it was necessary to construct a new matrix optimized for proper color correction. To do this, we used the module called “Spectral Calibration Run” in the “Data Collection” software of the ABI 3100 (version 1.0.1) or ABI 3700 (version 1.0). “Spectral Calibration Run” uses the “MatrixStandard” algorithm, which requires a mixture of four kinds of DNA fragments having preset size ranges that are labeled by appropriate colors.

Three hemi-stranded fragments of sizes 93, 104, and 117 bp, each carrying R110, TAMRA and R6G, respectively, were prepared by PCR using one of the primers phosphorylated by polynucleotide kinase (10), post-labeling, and lambda exonuclease treatment as described previously (4,11). These fragments were chosen so that their sizes were close to those of the fragments supplied by the manufacturer of the instruments for the color matrix construction. ROX-labeled dNTP was not commercially available, and so we used a ROX-labeled fragment from the “Matrix Standard Set for Dye Set D” supplied by the manufacturer.

The four fragments were mixed and subjected to a spectral calibration run in SSCP conditions, using parameter set “MtxStd{AnyDyeSet}.par”. The quality of matrix created for each capillary was checked by the software using default parameters and by visual inspection.

Data Processing

Output data from each capillary machine were converted to ASCII format using ABI Toolkit and were imported to QUISCA (12). The mobility calibration

<table>
<thead>
<tr>
<th>STS Name</th>
<th>ABI 3100 Peak Position</th>
<th>ABI 3100 Peak Height Ratio</th>
<th>ABI 3700 Peak Position</th>
<th>ABI 3700 Peak Height Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Peak 1</td>
<td>Peak 2</td>
<td>(Peak 1/Peak 2)</td>
<td>Peak 1</td>
</tr>
<tr>
<td>STS1</td>
<td>0.014</td>
<td>0.020</td>
<td>0.35</td>
<td>0.063</td>
</tr>
<tr>
<td>STS2</td>
<td>0.077</td>
<td>0.012</td>
<td>0.93</td>
<td>0.026</td>
</tr>
<tr>
<td>STS3</td>
<td>0.020</td>
<td>0.004</td>
<td>1.92</td>
<td>0.028</td>
</tr>
<tr>
<td>STS4</td>
<td>0.018</td>
<td>0.010</td>
<td>2.02</td>
<td>0.050</td>
</tr>
<tr>
<td>STS5</td>
<td>0.107</td>
<td>0.017</td>
<td>1.41</td>
<td>0.012</td>
</tr>
</tbody>
</table>

CVs are expressed as percentages. See text for details. Peaks 1 and 2 represent fast and slow peaks of alleles in heterozygotes, respectively. In ABI 3100, the number of capillary tested in this study was 16 in all STSs. In ABI 3700, the number of capillary was 89 in STS1, 93 in STS2, 95 in STS3, 92 in STS4, and 93 in STS5, respectively. The product size of STS1 through 5 is 155, 155, 130, 130, and 160 bp, respectively.
was done by referring to the peaks of the internal standard following local Southern or cubic spline method. Sequencing data were interpreted by Phred/Phrap/Polyphred/Consed with visual inspection of trace data (13–16). The default parameter of PolyPhred was chosen for the SNP detection (i.e., PolyPhredRank 3 and quality threshold 30).

RESULTS AND DISCUSSION

PLACE-SSCP

Figure 1A shows typical examples of SSCP analysis of DNA samples carrying SNPs. Alleles of the strands labeled with R6G (red) and R110 (blue) were usually well separated, and samples can be easily genotyped. Occasionally, however, a single allele was observed as two or more peaks or was broadened as depicted in Figure 1B. This is interpreted to be because of the presence of multiple iso-conformers (3). In these cases, heterozygotes gave a composite electropherogram of the two homozygotes. Thus, the presence or absence of mutations/polymorphisms can still be unambiguously determined by comparing the peak patterns of two or more samples.

Reproducibility of Peak Positions and Peak Height Ratios

The reproducibility of mobility among the capillaries after calibration was evaluated using five STSs each carrying single SNPs, which were chosen from a public database (6). DNA samples were from heterozygous individuals for these SNPs. SSCP electrophoresis was performed using ABI 3100 and ABI 3700 under the same conditions (i.e., 10% PDMA in 2× TME buffer at 27°C).

All STSs revealed two well-separated peaks, each representing alleles at least in one strand. Table 1 summarizes CVs of positions and relative heights of peaks of all alleles. In all cases, the CV of peak position was less than 0.11%. Assuming that the value follows normal distribution, a change in mobility greater than 0.33% of CV is indicative of the presence of sequence difference at a confidence limit higher than 99.7%. The CV of peak height ratios was less than 2% for ABI 3100, but those of ABI 3700 are distinctively larger, reaching as high as 15.3%. We observed that the peak height ratio was gradually changed across the array of the capillary of ABI 3700, possibly because of uneven temperature across the array. Run-to-run variation in ABI 3100 was estimated using one of the STSs used in Table 1 (STS 2). The CVs of peak positions and peak height ratios of each of 16 capillaries among five successive runs were calculated using the same heterozygous sample. The CVs of peak positions and peak height ratios were less than 0.06% and less than 1.7%, respectively, for all capillaries (data not shown). Thus, the reproducibility between capillaries and between runs was similar.

Table 2. Quantitative Detection of Alleles

<table>
<thead>
<tr>
<th>Mixing Ratio (Allele 1:Allele 2)a</th>
<th>Corrected Allele Frequency of Allele 1b</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>STS6</td>
</tr>
<tr>
<td>0:1</td>
<td>0.0</td>
</tr>
<tr>
<td>1:9</td>
<td>11.8</td>
</tr>
<tr>
<td>2:8</td>
<td>19.4</td>
</tr>
<tr>
<td>3:7</td>
<td>29.5</td>
</tr>
<tr>
<td>4:6</td>
<td>39.5</td>
</tr>
<tr>
<td>5:5</td>
<td>50.0</td>
</tr>
</tbody>
</table>

[a]DNAs from two homozygous individuals of opposite alleles were mixed at the indicated ratio. The alleles that were observed as fast and slow peaks were named allele 1 and allele 2, respectively.

[b]Allele frequencies are in percentage. Reference values are italicized.

The product size of STS6 through 9 is 304, 306, 399, and 351 bp, respectively.

Among the 164 STSs that gave readable signals in both analyses, 112 turned out to be monomorphic by sequencing. None of the monomorphic STSs were judged to be polymorphic by SSCP. Thus, the specificity was 100% (112/112). All but four of the 52 polymorphic STSs detected by sequencing were also judged to be polymorphic by SSCP. Thus, the detection rate of SSCP in this case was 92.3% (48/52). The rate can be lower than this value, if the detection rate of less than 100% for sequencing is assumed. In the dye-terminator sequencing, the accuracy of Polyphred is significantly lower than that in dye-primer sequencing, as has been reported (16).

Allele Frequency Estimation

Precise allele frequency can be estimated by the ratio of peak heights of two alleles, after correction of amplification bias using the ratio of the heterozygotes (17). Quantitative detection of alleles in the present PLACE-SSCP system was tested by mixing experiments using four polymorphic STSs (Table 2). For each STS, the genomic DNAs of two individuals who were homozygous for the opposite alleles were combined at various ratios (from 0:5 to 5:5) and analyzed by PLACE-SSCP using ABI3100. Peak heights were determined by QUISCA (12), and their ratios were calculated. Then, the ratio at each mixing was corrected by the ratio at the 1:1 mixing (17). As shown in Table 2, the allele frequencies estimated from the corrected peak-height ratios were in good agreement with those obtained from the mixing ratios for all
four STSs (difference less than 2.2%), demonstrating that this method is suitable as the technique of allele frequency estimation in the pooling strategy, even in the recently proposed strict criteria (18).

**Throughput**

The turnaround time of the analysis using ABI 3100 is approximately 40 and 60 min for the fragments of around 200 and 400 bp, respectively. Thus, the throughput is 384–576 samples per day, if 24 h continuous run is assumed. These values are a remarkable improvement when compared with the previous method using a single-capillary machine such as the ABI PRISM 310, which can process only 24 samples per day for 400-bp fragments (8). The turnaround time using ABI 3700 is approximately 90 min per run for 400-bp fragments, but the number of samples that can be analyzed per run is six times that of ABI 3100. Consequently, 768 samples can be processed per day with the ABI 3700. SSCP analysis of comparable throughput can be achieved with MegaBace™ 1000 (Amersham Biosciences) (4), although this machine is not equipped with automatic sample loading capability. The throughput values of the present system are not particularly impressive when compared with the ultra-high-throughput technologies specialized in SNP typing such as hybridization using DNA chips or parallel primer extension (19). Nevertheless, the advantage of the present system is that the major instrumentation is widely accessible capillary array sequencers, and the throughput is sufficiently high for routine analysis in many situations (e.g., association studies of particular genetic traits in candidate gene approach), especially when considering the remarkable feature of

---

**Figure 1. Examples of SSCP analysis using ABI 3100.** Two loci, STS1 (A) and STS10 (B), in three individuals were subjected to SSCP analysis using an ABI 3100, as described in the Materials and Methods section, and interpreted with QUISCA. Red, blue, and black lines indicate fluorescence of R6G, R110, and TAMRA, respectively. In panel A, each allele gave a narrow, well-separated single peak in both strands (red and blue), and the genotype of each individual was obvious (1-1, 1-2, and 2-2 for individuals from top to bottom, assuming alleles with fast- and slow-moving peaks as 1 and 2, respectively). Panel B shows an example of peak splitting and broadening. In this case, both alleles gave well-separated split peaks in one strand (red). However, in blue, one of the alleles (allele 2) gave a single narrow peak, while the other allele (allele 1) was seen as a broad band. Genotypes of individuals were still unambiguously assigned (1-1, 1-2, and 2-2 for individuals from top to bottom, respectively, assuming the allele definition as described above).
the present method; that is, allele frequency can be precisely estimated by analyzing pooled DNA samples.

REFERENCES


This work was supported by a Grant-in-Aid for Genome Science from the Ministry of Education, Culture, Sports, Science and Technology, Japan. The authors thank other members of the Division of Genome Analysis, Research Center for Genetic Information, Medical Institute of Bioregulation, Kyushu University, for their constant cooperation and encouragement. Address correspondence to Dr. Kenshi Hayashi, Division of Genome Analysis, Research Center for Genetic Information, Medical Institute of Bioregulation, Kyushu University, 3-1-1 Maidashi, Fukuoka, Japan. e-mail: khayashi@gen.kyushu-u.ac.jp

Received 3 September 2002; accepted 21 January 2003.

S. Baba1, Y. Kukita1,2, K. Higasa1, T. Tahira1, and K. Hayashi1
1Kyushu University Fukuoka
2University of Tokyo
Tokyo, Japan

3′ RACE Walking along a Large cDNA Employing Tiered Suppression PCR


ABSTRACT

Large genes present particular cloning difficulties, especially when expressed at relatively low levels. We describe a novel method, termed 3′ rapid amplification of cDNA ends (RACE) walking, for the rapid determination of unknown 3′ flanking sequence of a large cDNA. The technique is a derivative of the anchored PCR 5′ RACE procedure but includes a specific and limited second-strand cDNA synthesis and a tiered “panhandle” suppression of nonspecific products. The method generated 900 bp of new sequence for the large tammar wallaby ATRY gene in two easy steps, in which standard 3′ RACE and PCR-based cDNA library walking proved unsuccessful. This robust approach represents a new tool for isolating unknown sequence under challenging cloning scenarios such as poor library representation, long coding regions, long 3′ untranslated regions, and difficult template regions.

INTRODUCTION

Several rapid PCR-based methods have been described for determining unknown sequence flanking known regions, with no need of a library. These include various forms of one-sided PCR such as anchored PCR, 5′ or 3′ rapid amplification of cDNA ends (5′/3′ RACE) (1), cassette ligation anchored PCR (2), vectorette or bubble PCR (3), splinkerette PCR (4), panhandle PCR (5,6), Alu PCR (7), restriction site PCR (8), thermal asymmetric interlaced PCR (9), extension primer tag selection/ligation-mediated PCR (EPTS/LM PCR) (10), and the recently described universal fast walking method (11).

Many of these methods have limitations such as low specificity, restriction enzyme site dependence, or the requirement for an efficient ligation step. Universal fast walking and anchored PCR 5′/3′ RACE address these issues. How-