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

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-

Short Technical Reports

ing method can be applied in an AFLP context by using adapters and unlabeled primers containing a restriction enzyme recognition sequence that can be cut and fluorescently labeled after the final, selective, AFLP amplification. Fluorescent AFLP (fAFLP) products can be multiplexed by simultaneously using all of the fluor channels detected by the sequencer.

We have been working with the CEQ™ 2000-XL Capillary-Based Sequencer and a CEQ Dye-Terminator Cycle Sequencing (DTCS) Kit (both from Beckman Coulter, Fullerton, CA, USA), but our fRFLP and fAFLP labeling methods should be feasible on any sequencing platform in which the reagents are packaged individually (i.e., not in a terminator dNTP-buffer-enzyme premixture). Table 1 illustrates that the post-digestion labeling chemistry costs approximately $0.45/reaction, and the electrophoretic separation costs approximately $1.35. These costs, per site, decrease with multiplexing.

MATERIALS AND METHODS

fRFLP

Here we illustrate fRFLP by using the method to distinguish *Attacin A* alleles generated by gene conversion in the *Attacin AB* cluster of *Drosophila melanogaster* (7). DNA from *D. melanogaster* lines 2CPA 1, 2CPA 14, 2CPA 105, and 2CPA 129 was amplified using primers Att98-1 and Att98-2 (7) under standard PCR conditions. PCR was carried out for 35 cycles with 45-s annealing and extension steps at 58°C and 72°C, respectively. Following amplification, 15 μL each PCR product were incubated with 7.5 U exonuclease I and 3 U shrimp alkaline phosphatase (USB, Cleveland, OH, USA) for 1 h at 37°C to inactivate residual primer and unincorporated nucleotides. The enzymes were then inactivated for 15 min at 80°C.

We digested the PCR products with the restriction enzymes *Msp*I, *Mse*I, and *Dpn*II (New England Biolabs, Beverly, MA, USA) because diagnostic polymorphisms for the *Attacin AB* gene conversion events disrupt cut sites recognized by these enzymes and because all three enzymes cut efficiently in PCR

Table 1. Post-Digestion Cost of fRFLP and fAFLP

Reagent	Stage	Per-Site Cost for Multiplexed Samples			
		Simplex	3-fold[a]	6-fold[a]	9-fold[a]
Exonuclease I	sample preparation	0.060	0.060	0.030	0.020
Shrimp alkaline phosphatase	sample preparation	0.111	0.111	0.056	0.037
DTCS label	labeling	0.163	0.163	0.082	0.054
10× Seq. Buffer	labeling	0.116	0.116	0.058	0.039
Separation Buffer	electrophoresis	0.490	0.163	0.082	0.054
Separation Gel	electrophoresis	0.550	0.183	0.092	0.061
Size Standard-600	electrophoresis	0.309	0.103	0.052	0.034
TOTAL (US dollars)		**1.799**	**0.900**	**0.450**	**0.300**

[a]Multiplex assumes all three fluor channels are used with either one, two, or three amplification products in each channel. It is assumed that digested reactions are pooled for a single labeling reaction when more than one product is represented in each channel and that each product queries only a single polymorphic site. Estimated per-site costs vary slightly with other multiplex structures.

buffer, which obviates the need for a buffer change at this step. Five microliters of each PCR product were incubated overnight at 37°C with 5 U each enzyme in single digestions. For simplicity, only one amplification product was digested in each of these reactions, but we have had excellent success multiplexing by pooling distinct PCRs at the digestion stage (unpublished results).

The restriction digestion reactions were labeled using components of the DCTS kit, with extra 10× sequencing buffer ordered directly from the manufacturer. The molar concentrations of the materials in this kit are proprietary, so we can only provide volume measures of reagents, but we have found this labeling reaction to be robust over a wide range of template, terminator, and DNA polymerase concentrations. For the present demonstration, each labeling reaction included 1 μL digested DNA, 1.9 μL 10× Sequencing Buffer, and 33 nL sequencing DNA polymerase. Only the fluorescent terminator nucleotide specific to each digestion was included in the labeling reactions (ddGTP for *Dpn*II, ddCTP for *Msp*I, and ddUTP for *Mse*I). Thirty-three nanoliters of fluorescent terminator nucleotide were used in the ddCTP and ddUTP labeling reactions, but 100 nL

were required for adequate signal with ddGTP. We added 17 μL distilled water to make 20 μL total volume, and the reactions were incubated for 1 h at 60°C.

After the incubation, we pooled the three labeled digestions from each PCR product and added 15 μL stop solution (1.2 M sodium acetate, 40 mM EDTA, 4 μg/μL glycogen). The fRFLP reactions were precipitated by the addition of 200 μL ice-cold 95% ethanol to each reaction and were then centrifuged for 20 min at 1800× *g*. The supernatant was aspirated off, and the pellets were washed three times with 200 μL ice-cold 70% ethanol and then allowed to air-dry. The dry pellets were resuspended in 40 μL deionized formamide with 0.25 μL CEQ DNA Size Standard-600 (Beckman Coulter) and overlayed with one drop of mineral oil. The fragments were then separated on the CEQ 2000-XL capillary-based sequencer using the default "Frag-4" method and analyzed with the fragment analysis component of the CEQ 2000-XL software, version 4.3.9 (Beckman Coulter).

fAFLP

We performed AFLP on DNA extracted from hydrothermal vent tubeworms, *Ridgeia piscesae* (Polychaeta:

Short Technical Reports

Siboglinidae), collected at the Endeavor Segment of the Juan de Fuca Ridge in the northeast Pacific Ocean. Genomic DNA was extracted using a standard phenol:chloroform technique (2), and AFLP was carried out using AFLP® Analysis System I and AFLP Starter Primer kits (Invitrogen, Carlsbad, CA, USA), which follow the protocol of Vos et al. (8). DNA (250 ng) from each individual served as the starting template. For fAFLP, the final amplification was performed with a selective extension of -AGG on the *Eco*RI adapter primer and -CTC on the *Mse*I adapter primer. In this example, we make use of the fact that the *Mse*I adapter and primer contain a recognition site for the four-cutting restriction enzyme *Dde*I.

After the selective amplification, 5 µL each AFLP amplification were incubated with 2.5 U exonuclease I and 1 U shrimp alkaline phosphatase for 1 h at 37°C, followed by 15 min at 80°C. The reactions were digested overnight at 37°C with 2.5 U *Dde*I (New England Biolabs), and 1 µL of the digestion was labeled with ddUTP as described earlier. Following labeling, 5 µL stop solution and 60 µL ice-cold 95% ethanol were added to each of the fAFLP reactions, which were then centrifuged for 20 min at 1800× *g*. The pellets were washed twice with 200 µL ice-cold 70% ethanol and air-dried. As in fRFLP, the dry pellets were resuspended in 40 µL deionized formamide containing 0.25 µL Size Standard-600 and overlayed with a drop of mineral oil. The fAFLP fragments were then resolved on the CEQ 2000-XL automated sequencer using the Frag-4 separation method.

Post-AFLP digestion with *Dde*I cut both in the adapter/primer sequences and at internal sites in the amplified products. All of these cut sites were subsequently labeled. In addition to supplementing genetic information about the samples, the presence of internal cut sites in larger products can bring the peaks down into the effective resolution range of the capillary sequencer (60–600 bp on default settings). However, in some cases, it may be desirable to engineer adapters and primers carrying recognition sites for rare-cutting enzymes to minimize the digestion of AFLP amplification products.
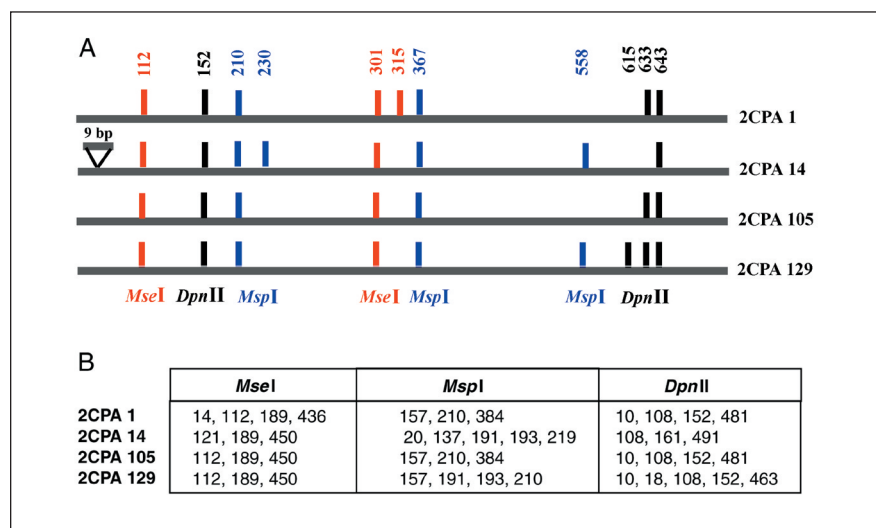

Figure 1. Digestion of the *Attacin A* gene of *D. melanogaster* with only three restriction enzymes distinguishes six polymorphisms among four alleles. (A) Schematic representation of the pattern of restriction sites in the four alleles. The numbers along the top row represent the positions of the cut sites along the 751-bp product. (B) Expected fragment sizes in the four lines, following digestion with each of the three enzymes. Fragments smaller than 60 bp are not detected, but all of the other peaks were accurately detected and sized within 2 bp.
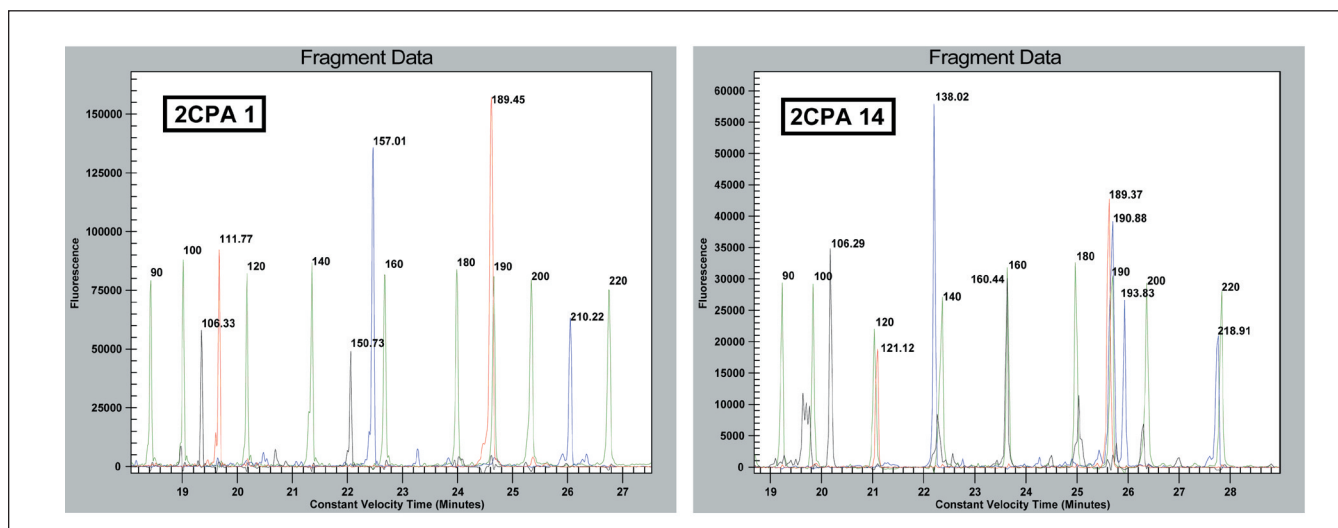

Figure 2. A zoomed-in view of chromatograms showing fRFLP peaks between 90 and 220 bp in lines 2CPA 1 and 2CPA 14. The high resolution of the automated sequencer allows easy identification of a 9-bp insertion/deletion polymorphism [112 vs. 121 bp in the red (T-labeled) channel; 210 vs. 219 in the blue (C-labeled) channel, and 152 vs. 161 in the black (G-labeled) channel]. Note the clear distinction between 191- and 193-base fragments in the blue channel of 2CPA 14. Green peaks are the CEQ DNA Size Standard-600 internal reference size standard.

RESULTS AND DISCUSSION

Using only three restriction enzymes and a single electrophoretic run of fRFLP, we typed six polymorphisms, including a 9-bp insertion/deletion polymorphism, in the *Attacin A* gene from four lines of *D. melanogaster* (Figures 1 and 2) at a post-digestion cost of about $0.45/site (Table 1). We also used fAFLP to generate a genomic fingerprint for the hydrothermal vent tubeworm *R. piscesae* (Figure 3) at a labeling cost of $0.45, without needing to buy a fluorescently labeled primer. The total per-site cost of fRFLP varies depending on the price of the diagnostic restriction enzymes and can be reduced by higher degrees of multiplexing. We have had success in multiplexing by pooling independent reactions before restriction digestion, pooling multiple digests before fluorescent labeling (data not shown), and simultaneously
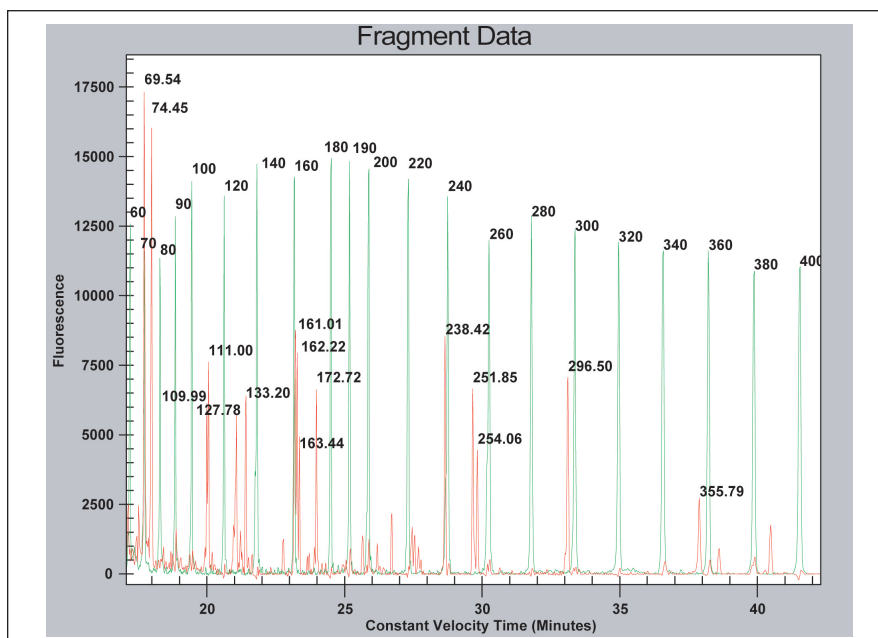

Figure 3. Chromatogram showing fAFLP peaks obtained by using *Dde*I to digest an unlabeled *R. piscesae* AFLP reaction and then incorporating a fluorescent ddUTP terminator nucleotide into the cut site (red peaks). Green peaks are the CEQ DNA Size Standard-600 internal reference size standard.

electrophoresing multiple independently labeled products.

Because automated sequencers can distinguish fragments that differ in size by a single nucleotide (Figure 2) and can detect hundreds of fragments in a single electrophoretic run, the only practical limit to the degree of fRFLP multiplexing that can be employed seems to be in the informatics of analyzing the output. The CEQ Fragment Analysis software does not have a pattern-recognition component that would allow for the translation of highly multiplexed fRFLP peak patterns into component multi-locus genotypes, but the software does export a tab-delimited text file that lists, among other attributes, the estimated size in base pairs, and under-peak area for every peak in each fluor channel. It should be straightforward to program an independent script that analyzes the CEQ text output for expected fRFLP genotype patterns. Without the use of automated genotype calling, we have found 6-fold multiplexing, with two independent PCR products digested and labeled in each the C, G, and T channels (3–10 peaks in each channel), to provide

an adequate compromise among cost per assay, throughput, and simplicity of analysis. For fAFLP analysis, a macro that generates a presence/absence table for all peaks in each fluor channel in a given sample set is available from Beckman Coulter technical support. This macro, along with other improvements that ease data analysis, is incorporated into the next generation CEQ 8000 software for more seamless fAFLP analysis.

The ability of automated sequencers to resolve extremely small size differences accurately can also be exploited in the examination of microsatellite repeats. Although we have not tested the possibility, we believe that the post-labeling procedure used in fRFLP and fAFLP could also be used in microsatellite analysis if rare-cutting restriction enzyme recognition sites are incorporated into the amplification primers. Non-overlapping repeat size ranges and the use of multiple fluors could allow a high degree of multiplexing in microsatellite analysis. Our labeling method could conceivably even be applied to the inexpensive generation of the internal size standard required in each well for accurate sample peak sizing.

Overall, both fRFLP and fAFLP provide a simple, rapid, and inexpensive means of genotyping moderate numbers of individuals at a large number of loci with no need to order allele-specific or fluorescently labeled oligonucleotides. It is not even necessary to have prior knowledge of the polymorphisms that will be typed because both methods can be used to query anonymous or uncharacterized sequences. The cost, ease, and flexibility of these methods make them appropriate for many medium-throughput applications on a scale likely to be employed by individual research laboratories.

REFERENCES

1. **Arnold, C., L. Metherell, J.P. Clewley, and J. Stanley.** 1999. Predictive modeling of fluorescent AFLP: a new approach to the molecular epidemiology of *E. coli*. Res. Microbiol. *150*:33-44.
2. **Ausubel, F.M., R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith, and K. Struhl.** 1989. Current Protocols in Molecular Biology, Vol. 1. Greene Publishing Associates and Wiley-Interscience, New York.
3. **Bruce, K.D. and M.R. Hughes.** 2000. Terminal restriction fragment length polymorphism monitoring of genes amplified directly from bacterial communities in soils and sediments. Mol. Biotechnol. *16*:261-269.
4. **Desai, M., A. Tanna, R. Wall, A. Efstratiou, R. George, and J. Stanley.** 1998. Fluorescent amplified-fragment length polymorphism analysis of an outbreak of group A streptococcal invasive disease. J. Clin. Microbiol. *36*:3133-3137.
5. **Fishman, L., A.J. Kelly, E. Morgan, and J.H. Willis.** 2001. A genetic map in the *Mimulus guttatus* species complex reveals transmission ratio distortion due to heterospecific interactions. Genetics *159*:1701-1716.
6. **Koeleman, J.G., J. Stoof, D.J. Biesmans, P.H. Savelkoul, and C.M. Vandenbroucke-Grauls.** 1998. Comparison of amplified ribosomal DNA restriction analysis, random amplified polymorphic DNA analysis, and amplified fragment length polymorphism fingerprinting for identification of *Acinetobacter* genomic species and typing of *Acinetobacter baumannii*. J. Clin. Microbiol. *36*:2522-2529.
7. **Lazzaro, B.P. and A.G. Clark.** 2001. Evidence for recurrent paralogous gene conversion and exceptional allelic divergence in the *Attacin* genes of *Drosophila melanogaster*. Genetics *159*:659-671.
8. **Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. van de Lee, M. Hornes, A. Frijters, J. Pot, et al.** 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res. *23*:4407-4414.
9. **Voss, S.R., J.J. Smith, D.M. Gardiner, and D.M. Parichy.** 2001. Conserved vertebrate chromosome segments in the large salamander genome. Genetics *158*:735-746.

Brian P. Lazzaro, Bonnielin K. Sceurman, Susan L. Carney, and Andrew G. Clark
Pennsylvania State University University Park, PA, USA