AFLP analysis using GeneMapper® software and an Excel® macro that aligns and converts output to binary

Timothy A. Rinehart
United States Department of Agriculture, Poplarville, MS, USA

BioTechniques 37:__-__ (August 2004)

Amplified fragment-length polymorphism (AFLP) is a popular technique that uses linkers to combine elements from restriction fragment-length polymorphism (RFLP) and random-amplified polymorphic DNA (RAPD) methods to PCR amplify restriction-based molecular markers from random locations in the genome. AFLP markers are widely utilized because of reproducibility, high band yield per primer pair, and a priori sequence knowledge is not required (1). Automation using capillary array systems and fluorescently labeled primers makes AFLP particularly appealing to researchers with large numbers of samples. Among the diverse applications for AFLP-based genetic fingerprints, it is most useful in breeding programs involving germplasm characterization and linkage mapping (2,3).

Applied Biosystems (Foster City, CA, USA), LI-COR Biosciences (Lincoln, NE, USA), and Amersham Biosciences' (Piscataway, NJ, USA) MegaBACE™ are the most common capillary array systems for sequencing and fragment analysis. Both LI-COR and MegaBace systems have current software packages supporting AFLP. Previous Applied Biosystems software, GeneScan™ and Genotyper®, allowed AFLP analysis, but these programs are no longer supported and are rapidly becoming incompatible with current Data Collection software. For example, files generated with Data Collection version 2.0 on instruments other than ABI PRISM® 3100 and 3100-Avant Genetic Analyzers are not recognized by GeneScan. Laboratories are motivated to upgrade Data Collection versions to maximize versatility, allowing researchers to perform DNA sequencing, single nucleotide polymorhisms (SNP) detection, sequence variant discovery, and simple sequence repeat (SSR) detection via multiple analysis programs on a single instrument. As software is upgraded, researchers performing AFLP must resolve compatibility issues with the discontinued GeneScan program or manually evaluate their data.

GeneMapper®, the latest fragment analysis software from Applied Biosystems, is automatically installed when upgrading to Data Collection version 2.0 and was designed for SSR and SNP analyses. Unfortunately, AFLP analysis is not supported, and the peak detection algorithms are optimized for detecting stutter in SSR samples. AFLP customers are on their own to adjust parameters for fragment sizing, allele calling, and format the output to be compatible with other phylogenetic analysis software. The methods described here enable semi-automated AFLP fragment characterization using GeneMapper, and the available Excel® macro quickly converts the output to a format compatible with phylogenetic software. This procedure works equally well for both GeneMapper version 3.0 and version 3.5.

GeneMapper's fragment sizing and automated allele calling parameters for microsatellite data can be modified to accurately size AFLP fragments. First, create a panel, marker, and binset for the sample using the Panel Manager. Marker sizes should range from 50 to 500 base pairs. For high-density AFLP samples, multiple markers can be created with sizes as small as 50 bases (bases 50 to 100, 100 to 150, etc.). A maximum of 50 alleles, or bins, can be detected for each marker, so multiple markers with smaller sizes allow the identification of more AFLP peaks per primer pair. Set the marker repeat type to 3. Once markers have been created, Use the Analysis Editor to modify the analysis methods. Select the appropriate binset and check the box labeled *use marker-specific stutter ratio if available*. Reduce all values under Trinucleotide to 0 except for PlusA distance, which should be changed to 0.1. Change the Peak Detector to *advanced* algorithm. Select *heavy smoothing*, and change the baseline window size to 151

```
#NEXUS
BEGIN DATA;
DIMENSIONS NTAX=10 NCHAR=10;
FORMAT DATATYPE=BINARY INTERLEAVE MISSING=-;
MATRIX
Sample #1  10011111110
Sample #2  01101101111
Sample #3  10101010101
Sample #4  00111100111
Sample #5  10001001101
Sample #6  01001010110
Sample #7  01001101111
Sample #8  01001100100
Sample #9  00011101111
Sample #10 10011011111
```

Figure 1. Worksheet created by the macro condenses binary data into a single column with the appropriate formatting for further analysis.

Table 1. AFLP Results for Ten Samples Using GeneMapper Software

Sample Name	Allele 1	Allele 2	Allele 3	Allele 4	Allele 5	Allele 6	Allele 7	Allele 8	Allele 9	Allele 10
Sample #1	50	78	148	?	220	261	?	267	328	403
Sample #2	55	62	148	220	267	328	?	403	?	446
Sample #3	50	?	?	62	148	261	328	?	446	
Sample #4	62	?	78	148	220	328	403	?	?	446
Sample #5	?	50	148	267	328	?	?	?	?	446
Sample #6	55	148	261	328	?	?	403			
Sample #7	55	148	?	?	220	267	328	?	403	446
Sample #8	?	55	148	?	220	328	?	?	?	
Sample #9	78	148	220	267	328	?	403	?	?	446
Sample #10	50	78	148	261	267	328	403	?	446	

Optimized amplified fragment-length polymorphism (AFLP) detection classifies peaks with a bin number representing the closest base pair or a question mark (?) for peaks detected but not recognized as alleles. Alleles with the same bin number can be found in multiple samples, but data are not organized in vertical columns for easy polymorphism detection.

Table 2. GeneMapper Data Converted to Binary and Aligned Using an Excel Macro

Sample Name	50	55	62	78	148	220	261	267	328	403	446
Sample #1	1	0	0	1	1	1	1	1	1	1	0
Sample #2	0	1	1	0	1	1	0	1	1	1	1
Sample #3	1	0	1	0	1	0	1	0	1	0	1
Sample #4	0	0	1	1	1	1	0	0	1	1	1
Sample #5	1	0	0	0	1	0	0	1	1	0	1
Sample #6	0	1	0	0	1	0	1	0	1	1	0
Sample #7	0	1	0	0	1	1	0	1	1	1	1
Sample #8	0	1	0	0	1	1	0	0	1	0	0
Sample #9	0	0	0	1	1	1	0	1	1	1	1
Sample #10	1	0	0	1	1	0	1	1	1	1	1

Bin numbers from all samples are used to create the header row. Samples scoring positive for a particular bin are assigned 1 under the corresponding header, while samples that do not contain a bin are assigned 0.

and the relative fluorescent unit (rfu) value to 100, for the appropriate dye label color you are using. Under the Peak Quality tab, change the signal levels to 50, heterozygote balance to 0.1, peak morphology to 1.5, Pull-up peak to 0.1, and allele number to 50.

After the AFLP samples are analyzed, return to the Panel Manager to automatically create a binset. Designate all the samples using a particular marker as reference data and select the *autobin* function. Change the default settings for minimum quality value to 0 and the allele naming scheme to *rounded base pair*. Once autobinning is complete, the panel will contain all possible bins, and your AFLP samples can be reanalyzed. The results shown under the genotype tab should display up to 50 allele columns and bins can now be compared across samples (Table 1). For example, samples #1, #3, #5, and #10 in Table 1 have alleles labeled 50, indicating that a peak was detected at approximately base pair 50 in all four samples. This peak was automatically characterized, or binned, by the software according to the analysis settings. Peaks that are detected but not binned are shown in Table 1 as a question mark (?). Depending on the density of markers and differences in peak heights between samples, the Peak Quality parameters may need to be adjusted. Bins can be added, deleted, widened, or narrowed, depending on the researchers needs. Samples should be reanalyzed after any changes are made. Automated peak detection is easily overwhelmed by the large number of peaks found in AFLP samples. In some cases, AFLP protocols must be modified to reduce sample complexity.

A custom genotype table must be created using the table editor in order to use the Excel macro. Table content should be reduced to *sample names* and *alleles*. Export the custom genotype data and import the saved file into Excel. Data should look identical to the display in GeneMapper, with alleles organized in rows for each sample starting with the first peak detected (Table 1). Bins are not organized in vertical columns, which is a serious problem for phylogenetic software that requires data to be aligned and converted to binary (presence = 1 and absence = 0).

The newly available Excel macro converts and aligns allele calls in a new worksheet and then creates a NEXUS format interleaved alignment in a second worksheet. With macro file open in the background, select *macro* under Tools while viewing the genotype data. Running the binary macro will create a new worksheet with sample names and allele calls converted to binary for presence and absence (Table 2). Invoking the *align* macro will create a new worksheet with the binary data condensed into a single column with appropriate NEXUS formatting for further analysis (Figure 1). This worksheet can be saved as a text file for use with PAUP or PHYLIP (4,5). The *binary_alignment* macro combines both procedures into one macro routine.

This Excel macro is written in visual basic and easily distributed or modified for particular researcher needs. Simple modifications can be programmed by a novice and are legal under the General Public License agreement for freely available software. Copies may be distributed and modified in accordance with guidelines that include legal provisions against profiting from freely distributed software. The macro can be downloaded free of charge from

the BioTechniques web site at http//www.BioTechniques.com/August04/RinehartSoftware.html or contact the author for an e-mail version. For those researchers using Applied Biosystems capillary array systems for AFLP analysis, the ultimate solution is the inclusion of AFLP-specific tools in the next version of GeneMapper. Until this happens, the methods and conversion macro described in this paper make AFLP analyses semi-automated using the current software.

ACKNOWLEDGMENTS

Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture. Lisa Davis and Adonis Reece from Applied Biosystems provided critical technical expertise for optimizing GeneMapper software for AFLP analysis. Jeff Gilbert provided assistance with visual basic programming. Special thanks to Thom Yorke.

COMPETING INTERESTS STATEMENT

The author declares that he has no competing interests.

REFERENCES

1. **Mueller, U.G. and L.L. Wolfenbarger.** 1999. AFLP genotyping and fingerprinting. Trends Ecol. Evol. *14*:389-394.
2. **Nyborn, H.** 2003. Navigating the "smorgasbord" of DNA marker methods for improvement of horticultural crops. Acta Hort. *625*:15-28.
3. **Rajapakse, S.** 2003. Progress in application of molecular markers to genetic improvement of horticultural crops. Acta Hort. *625*:29-36.
4. **Jin, L. and M. Nei.** 1991. Relative efficiencies of the maximum-parsimony and distance-matrix methods of phylogeny construction for restriction data. Mol. Biol. Evol. *8*:356-365.
5. **Swofford, D.L.** 2000. PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods). Sinauer Associates, Sunderland, MA.

Received 12 April 2004; accepted 29 April 2004.

Address correspondence to Timothy A. Rinehart, United States Department of Agriculture, Agricultural Research Service, Small Fruit Research Station, P.O. Box 287, 306 South High Street, Poplarville, MS 39470, USA. e-mail: trinehart@ars.usda.gov