SoftGenetics Application Note

Analysis of MegaBACE™ Data with GeneMarker®Software

Wan Ning, Tamela Serensits, ChangSheng (Jonathan) Liu

Introduction

Capillary array electrophoresis uses narrow-bore capillaries filled with a separation matrix to resolve DNA sequencing fragments. An applied electric field causes the DNA fragments to migrate through the capillaries. The DNA fragments are separated by size, with the shorter fragments moving faster (peaks earlier in the electropherogram) than the longer fragments (peaks later in the electropherogram). This basic concept of capillary electrophoresis has been applied in different ways by several companies. One company in particular, GE Healthcare (formerly AmershamBiosciences), created a capillary array electrophoresis (CAE) instrument and chemistry line called MegaBACE.

MegaBACE systems use capillaries filled with a replaceable, non-cross-linked linear polyacrylamide (LPA) matrix. The inside walls of the capillary are coated with hydrophilic polymers. The MegaBACE instrument applies a variable time/variable voltage pulse to electrokinetically inject all DNA samples from a microtitre plate simultaneously. Each capillary has a clear detection window located at a fixed distance from the sample loading point through which the samples are scanned. During electrophoresis the instrument uses two laser excitation and a confocal optical detection system to both excite and detect the dye-labelled DNA fragments as they migrate past the detection window. MegaBACE instruments have been validated for use with such genotyping applications as microsatellites (1), STRs (2), AFLP® (3), and SNP detection (SNuPe™) (4).

In addition to the MegaBACE instruments, GE Healthcare also provides the DYEnamic™ ET Dye Terminator and Primer kits. Energy transfer (ET) dye terminators and primers have two fluorescent dyes attached to the DNA fragment. One dye, called the donor dye, was chosen because it efficiently absorbs light of the wavelength of the argon-ion laser (488nm). This dye transfers absorbed light energy to the second dye attached to the same DNA molecule. This acceptor dye emits the absorbed energy as fluorescence at its normal emission wavelength for detection by fluorescent sequencing instruments.

To reduce reagent costs, many labs use MegaBACE's ET dye kits with their Applied Biosystems Inc. (ABI) capillary electrophoresis instruments. The average cost of the ET dye kits is only a fraction of ABI's BigDye® kits. In contrast some labs have instruments from several manufacturers and may use a single dye system; for example, using the BigDye kits with a MegaBACE instrument. As a result, data analysis becomes tricky when crossing different instrument platforms with other third-party reagents.

SoftGenetics GeneMarker software is designed specifically for analyzing DNA fragment data and takes into account the uniqueness of MegaBACE instruments and chemistries by recognizing the specific color spectrum order of the raw data files for accurate genotyping and correcting for the DNA migration time pattern associated with MegaBACE's unique LPA matrix. In addition, GeneMarker makes analysis of MegaBACE raw data simple by including the MegaBACE pre-defined size standards: ET400-R, ET550-R, and ET900-R.

Procedure

To analyze MegaBACE data efficiently with GeneMarker, the user must apply a matrix and change the default color channels. The following are steps a user can take to optimize GeneMarker for analyzing MegaBACE data.
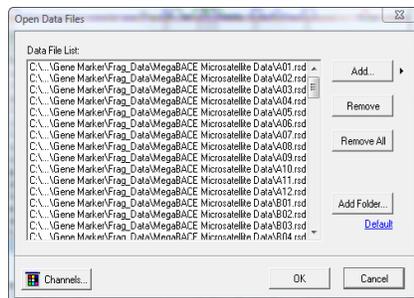
1. **Load RSD Files**
   -Click the **Open Data** icon in the main toolbar

   -Click **Add...** button in the *Open Data Files* box

   -Choose *RSD* files from the directory tree

   -Select all files for analysis and click **Open**
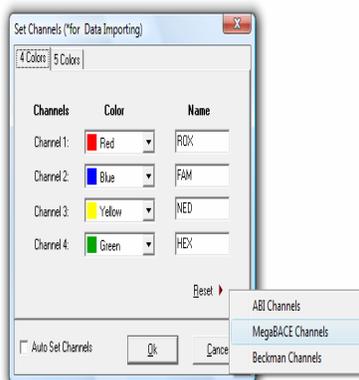

2. **Adjust Color Channels**
   -Click the **Channels** button in the *Open Data Files* box

   -Deselect **AutoSet Channels**

   -Click the **Reset** button and choose **MegaBACE Channels** from the fly out

   -Click **OK** in the *Set Channels* box and the *Open Data Files* box

3. **Apply Matrix**
   -Once the data is uploaded, select the **Manual Pull-up Correction** icon from the main toolbar

   -Click the **Load** button to upload a TXT matrix file or manually enter the matrix values
   NOTE: The order color of the matrix may be different than that of the Pull-up Correction module. Take care to enter the numbers in the correct order.

   -Click **OK** when finished - the matrix will be applied.
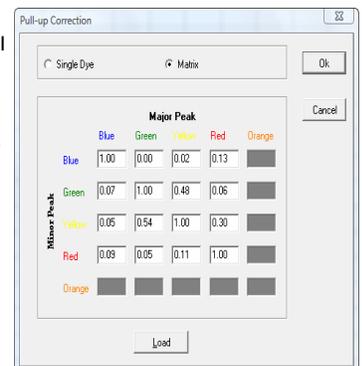
4. **Choose ET Size Standard**
   -Click the **Run Project** icon in the main toolbar

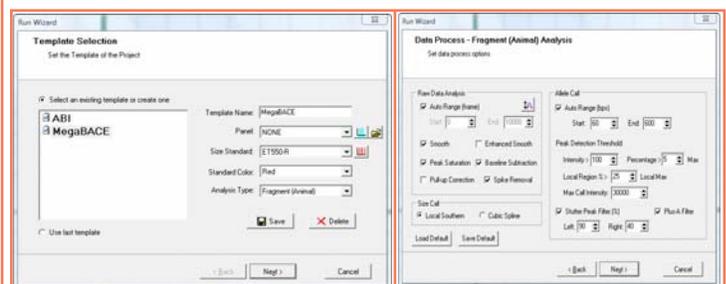   -The *Run Wizard Template Selection* box appears

   -Select the appropriate ET size standard from the *Size Standard* drop-down menu

   -Click **Next**

   -De-select **Pull-up Correction** in the *Run Wizard Data Process* box

   -Click **OK** in the *Run Wizard Additional Settings* box

   -GeneMarker sizes the data with the ET size standard

5. **Verify Size Call was Successful**
-Click the **Calibration Charts** icon in the main toolbar

-Samples with low sizing scores can be edited and the score improved

-Green triangles at the top of the peaks indicate the peak was used for sizing

-Right-click in the electropherogram to **Add/Delete Peaks** for sizing

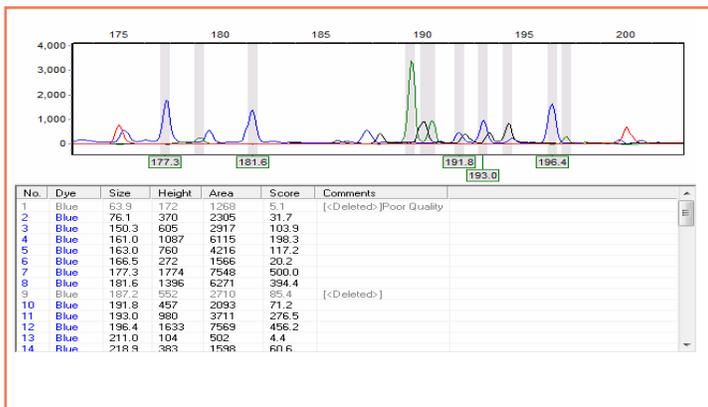-Right-click and select **Update Calibration** to improve size scoring.
Close *Calibration Charts* when finished editing, changes will be applied to the samples
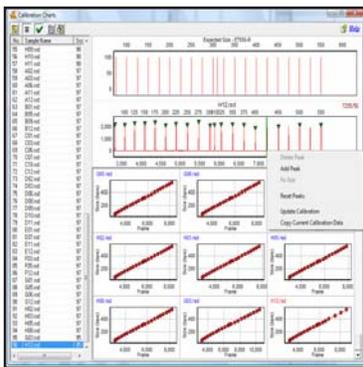
6. **Analyze and Edit the Samples**
-Right-click peaks in the electropherogram or report table to add/delete and edit comments

-Click the **Peak Table** icon to display the peak table below the electropherogram

No.	Dye	Size	Height	Area	Score	Comments
1	Blue	63.9	172	1268	5.1	[<Deleted>]Poor Quality
2	Blue	76.1	370	2305	31.7	
3	Blue	150.3	605	2917	103.9	
4	Blue	161.0	1087	6115	198.3	
5	Blue	163.0	760	4216	117.2	
6	Blue	166.5	272	1566	20.2	
7	Blue	177.3	1774	7548	500.0	
8	Blue	181.6	1396	6271	394.4	
9	Blue	187.2	552	2710	85.4	[<Deleted>]
10	Blue	191.8	457	2093	71.2	
11	Blue	193.0	980	3711	276.5	
12	Blue	196.4	1633	7569	456.2	
13	Blue	211.0	104	502	4.4	
14	Blue	218.9	383	1598	60.6	

Results

Once MegaBACE instrument RSD files are uploaded to GeneMarker, the software automatically corrects for the instrument's unique mobility. The MegaBACE CAE instrument uses a linear polyacrylamide (LPA) gel as a separation matrix which has a high molecular weight (~7 million Dalton) and high viscosity (~50 000 cP). The DNA migration time in the LPA gel is non-linear such that, smaller fragments are tightly bunched and the larger fragments are more spread out. To correct for this bias, a calculation has been incorporated into GeneMarker which measures the curve of DNA size versus the migration time (**Figure 1**).

The migration times of smaller fragments are significantly deviated from the linear curve in the original plot (blue line). To correct for this curvature, the following correction formula is applied and the corrected plot is almost a straight line (red line):

$$t_{corrected} = [t_0+(t-t_0)]*\{1+(0.5)(t-t_0)^2/[3(slope * s_0)^2]\}$$

Migration Time at Max of the Derivative = t_0
Instrument Constant = **0.5**
Average Migration Time Derivative = **slope**

Optimized Size = s_0
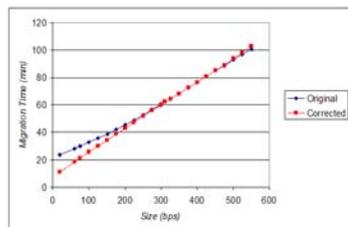Integration Constant = **3**

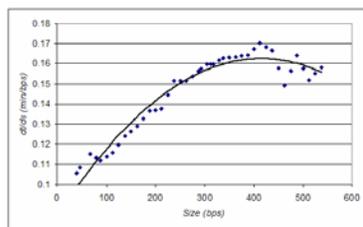**Figure 1**. The curve of DNA size vs. DNA migration time from a MegaBACE ET-550 size standard.

Figure 2. The local derivative of DNA migration time vs. DNA size as measured on a MegaBACE system.

The migration time is corrected for with an experimental slope calculation. To determine this slope, a local derivative (dt/ds) calculation is used and is as follows:

$$y = y_{max}[1-0.5(s-s_0)^2]$$

After observing several sets of experimental data, the curvature and max position were found to be independent of the experimental conditions. S_0 is about 400 bps and the fitting parameter 0.5 is fixed. As you can see in **Figure 2**, the DNA migration pattern shows the smaller fragments to be highly compacted due to the high viscosity of the LPA matrix used in MegaBACE instruments. Without GeneMarker's correction factor, these smaller fragments may be incorrectly identified due to the presence of extra peaks at the beginning of a trace.
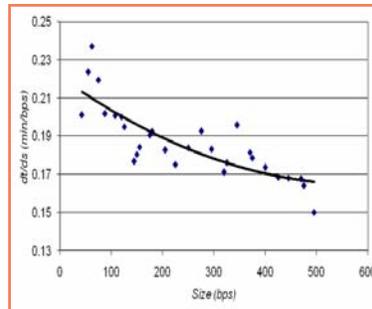

In contrast, the POP™ series gels in ABI instruments are made with polydimethylacrylamide which has a very low viscosity (~70cP), therefore the fragments are influenced more by the hydrophobicity of the gel. The smaller fragment peaks are expanded and therefore a correction factor is not necessary. In this case, the size calls are accurate most of the time for smaller DNA fragments (**Figure 3**).

Figure 3. The local derivative plot of migration time vs. DNA size measured on an ABI system.

Discussion

GeneMarker is a unique genotyping software program in that it has been developed independently of any one sequencing instrument. Because of this, GeneMarker can accept and analyze the raw data files from many instrument manufacturers including Applied Biosystems, GE Healthcare, Beckman Coulter, and even experimental instruments. The advantage that GeneMarker gives to MegaBACE instrument users is the many application modules specifically designed to report results for several experimental techniques including a pedigree module for parentage analysis, a data clustering module for phylogeny, and a trace comparison feature when the wild-type reference sample is known. GeneMarker takes advantage of the MegaBACE instrument's high resolution output (0.5bp resolution up to 550 nucleotides), decreased signal-to-noise ratio due to the confocal optics system, and accurate sizing precision with the ET dye terminator/primer kits.

References

1. Sensitivity, reproducibility and accuracy in short tandem repeat genotyping using capillary array electrophoresis. E Mansfield, et al. *Genome Research*. 2006. 6 (893-903).
2. Evaluation and validation of the ABI 3700, ABI 3100, and the MegaBACE 1000 capillary array electrophoresis instruments for use with short tandem repeat microsatellite typing in forensic environment. P Koumi, et al. Electrophoresis. 2004. 25 (2227-2241).
3. Highly automated AFLP fingerprint analysis on the MegaBACE capillary sequencer. M van der Meulen, et al. *International Plant and Animal Genome Conference*. 2002.
4. *CYP3A4* and *CYP3A5* genotypes, haplotypes, and risk of prostate cancer. S Plummer, et al. *Cancer Epidemiology Biomarkers & Prevention*. 2003. 12 (928-932).

*Trademarks are property of their respective owners.