GeneMarker®
by SoftGenetics
March 2009

New Technology Improves Size Calling of Kilobase DNA Fragments

Teresa Snyder-Leiby, Haiguo He, David Hulce and ChangSheng Jonathan Liu

Introduction

Capillary 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). Separation of ssDNA fragments by capillary gel electrophoresis has reached 1,000 base pairs of DNA sequencing and fragment analysis with 1 bp resolution using POP7™ gels. Estimation of unknown DNA fragment sizes is routinely accomplished by using synthetic molecular weight size standards, such as, ABI GeneScan™ 1200 LIZ® , ROX™1000, or MapMarker® 1000. However, the migration of ssDNA fragments ranging from 20-1200 base pairs is sigmoidal and not linear over the entire fragment size range (Figure 1). Obtaining accurate size calls for data over 800 bps is challenging due to the different conformations of small and large DNA fragments during migration. Due to differences in size and shape there is a limited linear range when separating fragments ranging in size from 20-80, 80-900 and over 1000 bps. Traditionally, different sizing algorithms are used for size calling; local Southern for large fragments, cubic spline and third order least squares for middle fragments. All of these methods are local and do not project beyond the last fragment. Analysis of large fragment data is limited by current sizing algorithms.

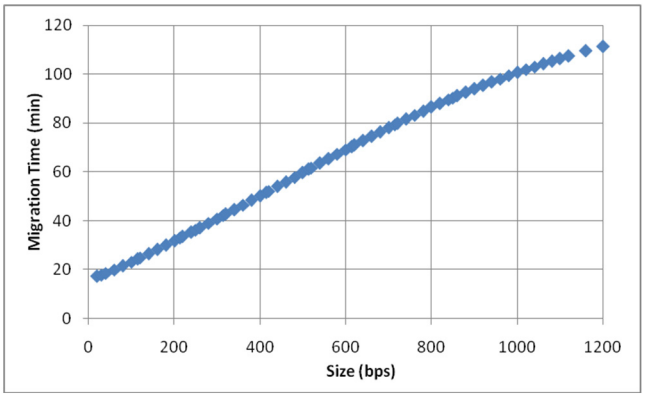
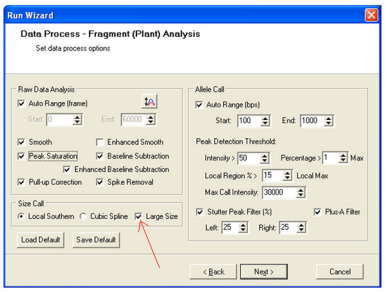


Figure 1: Illustrates the non-linear DNA migration of GeneScan®1200 fragments in a POP7™gel. The migration time is linear for 100-800 bp fragments. The larger and smaller fragments deviate from the linear function.

GeneMarker's new algorithm provides accurate, linear sizing of the data using a DNA derivative migration time correction to large DNA fragments. Current sizing separation technology has its optimum efficiency at 470 bp. GeneMarker's Large Size algorithm enables accurate sizing from small to large (30-1400) base pair fragments. The peak area in base space accurately determines the copy numbers because the peak area normalization is much less variable than that of the height normalization. The peak width is less than 1 bps for all fragments less than 1k.



Procedure

1. Load data files
2. Run Wizard → select size standard
3. Run Wizard → select Size Call method 'Large Size'
4. Next → Okay to make size calls

Results

Local Southern Algorithm has a limited linear range

New GeneMarker Algorithm provides linear range from 20–1200 bps

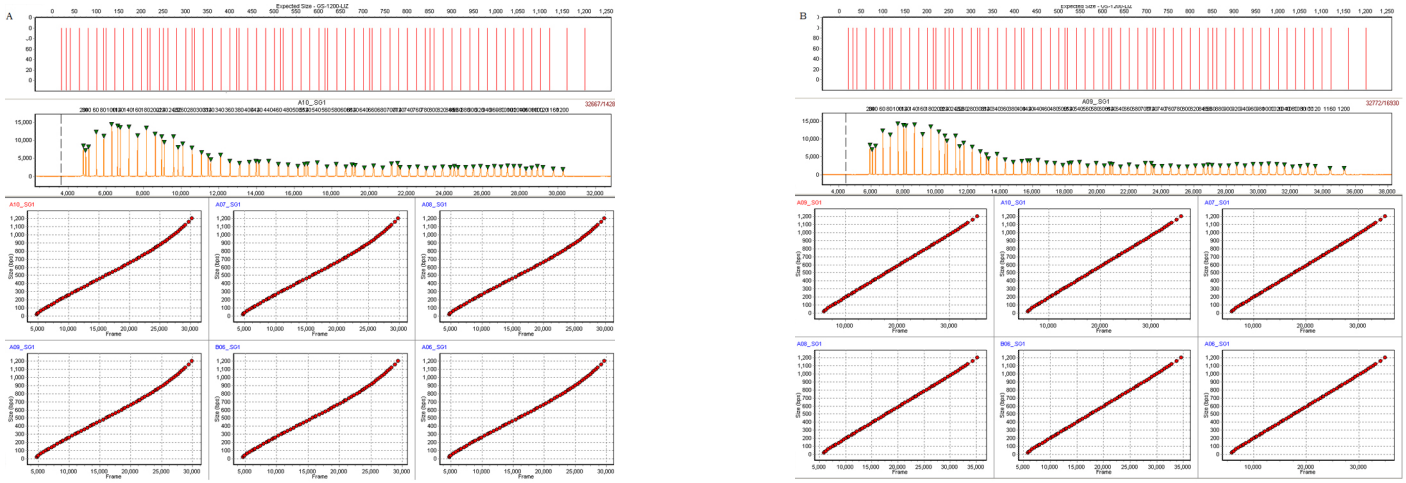


Figure 2: Comparison of the linear range for DNA fragments before (A) with Local Southern (80-800 bps) and after (B) with GeneMarker's Large Size (20-1200 bps) algorithms. The migration times are almost a perfect linear function to the DNA fragment size after applying the Large Size call in GeneMarker.

Results

Applications

Multiplex ligation-dependent probe amplification (MLPA®) analysis molecular diagnostic techniques are used in detecting copy number variations associated with cancers and genetically inherited diseases. One disease test is separated into multiple kits covering 80 probes, such as, P034 and P035 for diagnosis of Duchenne Muscular Dystrophy (DMD). Current technologies require running the same sample multiple times using different PCR kits to cover all the probes of interest; followed by manual combination and normalization of data by the analyst. These steps are time consuming and error prone. MLPA will benefit from the ability to use large (up to 1200) base pair fragments – quadrupling the number of detectable peaks from 45 to 200 probes. Allele specific amplification (ARMs™, ASA, OLA for diagnosis/pre-gestational diagnosis of diseases such as Tay Sachs and cystic fibrosis) is another molecular diagnostic technique where extending the size range of fragments will enable screening of four times as many mutations in one panel – making detection of less common mutations more cost effective and removing the error potentially caused by multiple kits and multiple steps to complete a single test in multiple runs.

Additionally, several research techniques are limited by the current size calling technology (TILLING®, STR / microsatellite profiles, AFLP, T-RFLP, VNTR, and BAC fingerprinting). Targeted Induced Local Lesions in Genomes (TILLING) and Endonucleolytic Mutation Analysis by Internal Labeling (EMAIL) analyses require the ability to accurately call the sizes of large DNA fragments and are effective tools for SNP detection.

Forensic applications (STR analysis) will benefit from increasing the size range of fragments by increasing the probe density. Tripling the range of markers in the linear range of 100-1200 bps sets the stage for using 14 STR markers for individual genetic profiles in one dye. This would quadruple the throughput and greatly reduce the cost of profile generation. Profiles for four individuals could be generated with a 5-dye system, instead of requiring four dyes to produce one individual profile.

Experimental Results: Migration Time Correction

The measured derivative of migration time as a function of DNA fragment size in figure 3 can be fitted with a polynomial function which is transformed to the base space. The derivative migration time is represented as a function of dt/dSize=ymax{1-0.333[1+0.25(Size-Size0)/Size0][(Size-Size0)^2/Size0^2)]} where ymax is the derivative of DNA migration time to the fragment size (Size) at its optimized size, Size0 = 470 bp. ymax is a parameter to be determined in the experiment, and the unit of ymax is in min/bps. The other parameters such as 0.333 and 0.25 are independent of the running experimental conditions.

We have improved the curve from our former algorithm used in processing MegaBACE™ and Beckman traces within the size range from 20 to 500 bps. dt/dSize=ymax{1-0.5[(Size-Size0)^2/Size0^2)]} After integrating the equation, GeneMarker converts the migration time to the pseudo base space. Figure 2B shows high linearity from 30-1200 bps for all DNA fragments in the calibration curves of multiple sample files. In contrast, Figure 2A shows a sigmoidal curve.

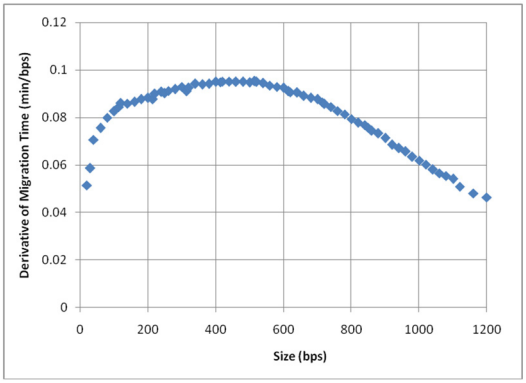


Figure 3: Measured derivative of migration time as a function of DNA fragment size.

Peak Height and Peak Area

Accurate copy number determination requires normalization of variation. The intensity varies about 5 times, with height of large DNA fragments having lower intensities in counts and optimal intensity around 100 bps. However, the peak areas in counts x basepairs are relatively constant (Figure 4), which is significant for determining the accurate copy numbers of the DNA fragments.

Diagnostic and forensic applications require the ability to resolve fragments that differ by only one base pair. Figure 5 illustrates the resolving power of this approach. The peak width profile indicates that fragments in the 1000 bp range have a 1bp resolution.

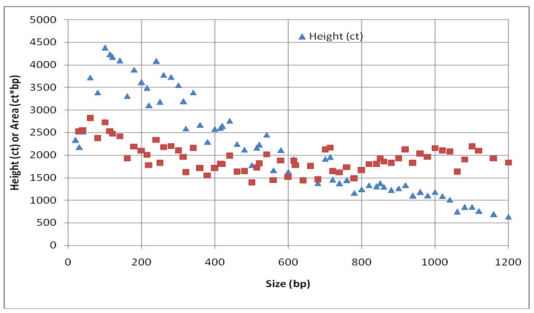


Figure 4: Intensity Profile comparing height (blue) or area (red) versus size for GeneScan 1200 LIZ size standard in POP7 gel through Applied Biosystems™ 3730 capillary electrophoresis systems.

Discussion

The previous GeneMarker algorithm was optimized for large DNA fragments generated in POP6 gels and in linear polyacrylamide (LPA) used in MegaBACE and Beckman CEQ™ systems.[4] The new sizing algorithm is developed with a denaturing POP7 gel of poly-acrylamide-co-dimethylacrylamide[2] using both small and large DNA fragments. We have found that the slope (derivative) is almost a constant after correction from 80 bp to 950 bps. This is the first experimental result measuring DNA phase changing in POP7 gel, from spherical to linear movement at 80bps and from linear to highly stretched linear movements at 950 bps. Size repeatability is approximately 0.15 bps using the new size standard.

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™, MegaBACE™ from GE Healthcare, CEQ™ and PACE™ systems from Beckman Coulter, and even experimental instruments. It provides the solutions for 1-color to 5-color systems. It supports all of the internal lane size standards, such as GeneScan® 500 LIZ and ROX, GeneScan® 600, GeneScan® 1200, from Applied Biosystems; CEQ™ 400 and 600 from Beckman; MapMarker™ 1000 from BioVenture; ILS™ 600 from Promega, etc. GeneMarker is an excellent replacement for GeneScan®, Genotyper®, GeneMapper®, GenoSpectrum™, Genographer©, Genoprofiler© and STRand©.

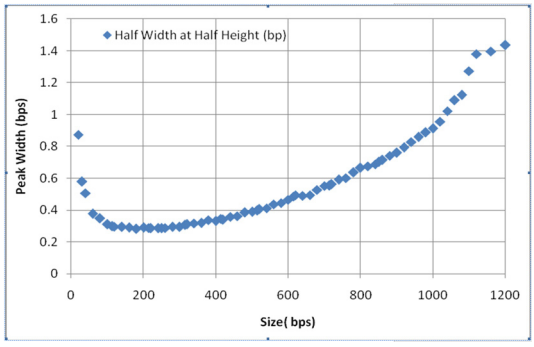


Figure 5: The measured peak width (HWHH, half width at the half height) of DNA fragments. The peak width profile indicates that the resolution of 1bp is reached at 1000bps.

References

1. M M Fisher and EW Triplett. 1999. Automated Approach for Ribosomal Intergenic Spacer Analysis of Microbial Diversity and Its Application to Freshwater Bacterial Communities Appl. Env. Micro. 65:4630-4636.
2. CS Liu and Q Li, Copolymers for capillary gel electrophoresis, US Patent: 6926815, 2005.
3. M J Cross, D L E Waters, L S Lee, R J Henry. 2008. Endonucleolytic mutation analysis by internal labeling (EMAIL). Electrophoresis 29:1291-1301
4. W Ning, T Serensits, CS Liu. 2007. Analysis of MegaBACE™data with GeneMarker® software. SoftGenetics LLC App Note www.softgenetics.com

SoftGenetics LLC 200 Innovation Blvd. Suite 235 State College, PA 16803 USA
Phone: 814/237/9340 Fax 814/237/9343
www.softgenetics.com email: info@softgenetics.com