

## GeneMarker® Software for Single Base Extension Analysis

David Hulce, Wan Ning, Jonathan Liu

### Introduction

Single nucleotide polymorphisms (SNPs) occur every 100 to 300 bases along the human genome and make up to 90% of human genetic variation **(1)**. Functional SNPs or non-synonymous SNPs (nsSNPs) occur in the coding region of a gene. Regulatory SNPs (rSNPs) occur in the promoter of a gene and may alter protein function or gene expression **(2, 3)**. Intronic or intergenic SNPs may not alter gene or protein function, but can be used to address questions in evolutionary biology **(4)** or in association studies with complex diseases, drug response, environmental insults **(1)** or quantitative trait loci (QTL) **(5)**.

One method to determine SNP genotypes is single base primer extension or SBE. An unlabeled primer with its 3' end directly flanking the SNP is extended one nucleotide by Taq polymerase and fluorescently-labeled ddNTPs complementary to the polymorphic base are added. The resulting fragment is one nucleotide longer, but the observed fragment size on a gel will be greater than expected due to the influence of the fluorescent dye on the electrophoretic mobility of these small fragments. SNPs can be identified by the one- or two-color peaks associated with the incorporated labeled ddNTP and the length of the primer.

The SBE technique can quickly interrogate a small number of SNPs. SNP genotyping using the MegaBACE™ SNUpe™ Genotyping Kit (Amersham Bioscience) demonstrated colocalization of Korrigan (beta 1-4 endo-glucanase) with a wood quality QTL of *Pinus pinaster* **(5)**. SNaPshot™ Genotyping System (Applied Biosystems) was used to demonstrate the hybrid nature of the Baltic sturgeon by detecting both *Acipenser sturio* and *A. oxyrinchus* alleles in museum specimens of the extinct Baltic sturgeon **(4)**. To fully utilize the investigative potential of SBE, a robust genotyping and data analysis system should be employed. GeneMarker genotyping software is designed for fast, accurate and efficient analysis and reporting of primer extension data.

GeneMarker is user-friendly software containing a robust size calling algorithm that resolves fragment lengths to less than one base pair with high efficiency allele calling, adjusts for mobility shift of small fragments and displays two-color SNPs on one trace, essential for data generated by SNUPE and SNaPshot. The software is capable of analyzing data files from all major capillary electrophoresis systems including ABI, MegaBACE and Beckman. GeneMarker, in combination with our JelMarker™ can also be used for analysis of slab gel output.

### Procedure

The GeneMarker software settings shown below are recommended to produce the best results for SBE analysis. The process from raw data to review of the analysis report occurs in three steps: size calling of the data, creating the SBE panel, running the sized data against the panel.

Import the data files (up to 1000 lanes) into GeneMarker using the Open Files icon.

#### Step 1: (Suggested Analysis Parameters for Size Calling)

Click the **Run Project** icon to activate the Run Wizard for size calling of the data.

##### Template Selection

1. Panel: *None*
2. Size Standard: *GeneScan™-120*
3. Standard Color: *Orange (LIZ™)*
4. Analysis Type: *SNaPshot*

##### Data Process - SNaPshot Analysis

1. Accept defaults

#### Step 2: Create Panel

**Tools** → **Panel Editor** → **Create New Panel.**

1. The *Panel Editor* will create a panel based on the called peaks for each dye color of the selected samples.
2. Name the peaks associated with each SNP using X\_N format, where X is the SNP descriptor and N is the nucleotide. For example, SNP2\_A and SNP2\_G would identify alleles A and G of locus SNP2. The details for editing the panel can be located in the manual.
3. Reprocess the data with the new panel (see Step 3).

#### Step 3: Compare Sized Fragments to Panel

1. Under *Panel* in the *Template Selection* window of the *Run Wizard* use the drop-down menu to select the appropriate SNUPE/SNaPshot panel for the analysis.
2. Continue through *Run Wizard* using all previously selected options.
3. To activate the SNUPE/SNaPshot analysis report, click **Applications** in the menu bar and select **SNPlex/SNaPshot**.

### Results

The SNP analysis report window displays a synthetic gel image, sample electropherogram, cluster plot to analyze peak information and assigns SNP genotypes. A SNP genotyping report **(Figure 1)** is also available. The information contained in the report is interlinked. Double-clicking on the cell in the report containing SNP 24-G/24-A for locus SNP1\_G/SNP1\_A of sample XX019 highlights the cell, corresponding sample ID in the sample list, data point in the cluster plot and locus in the electropherogram.

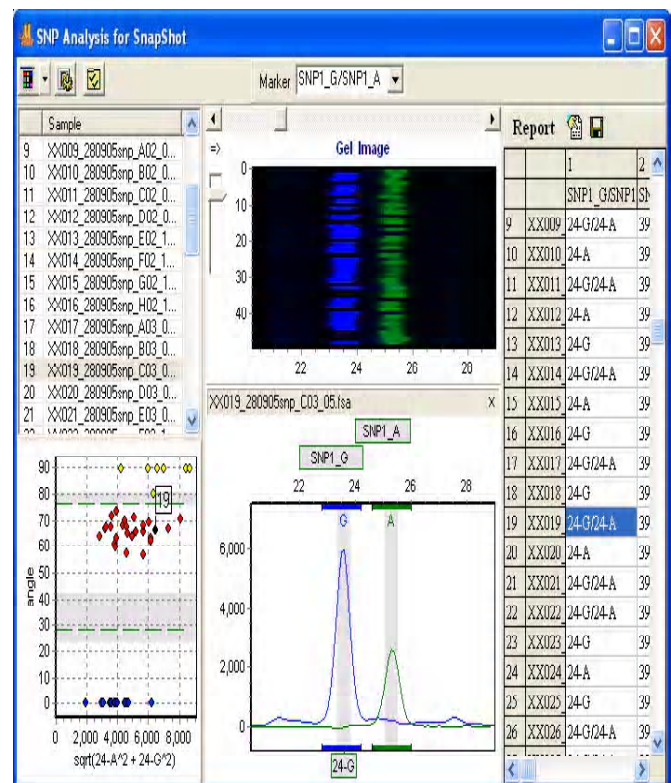
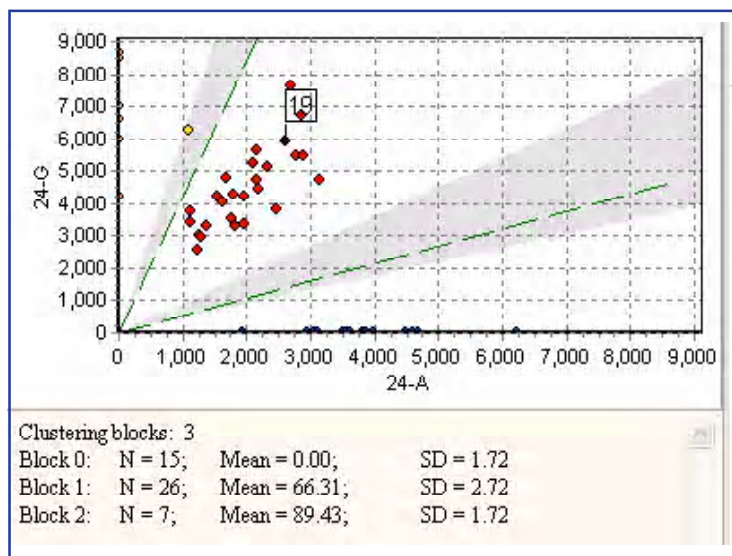


Figure 1: SNP Analysis for SNaPshot report.

The electropherogram displays peak intensity, fragment size (in box along bottom of trace) and SNP name (along top of trace). Low intensity peaks (<200 rfu) are detected with high sensitivity using a derivative to remove the influence of the baseline. The intensity threshold may be set to filter out low intensity peaks that contribute to incorrect SNP calling. A SNP is represented by the fragment size, height and color of the peaks. SNP1 for sample XX019 is heterozygous for G and A (**Figure 1**).

The default cluster plot layout is a polar view. Polar view plots the angle of the vector connecting the peak intensities of the SNP and the origin. This vector is plotted against the square root of the sum of the square of the longer fragment and the square of the shorter fragment. A Cartesian view (**Figure 2**) may be selected by clicking the **Layout Settings** icon on the toolbar. The number of samples, the mean and the standard deviation for each group can be displayed under the clustering plot.

Data points within the gray region of the cluster plot indicate SNP calls of low confidence. The clustering algorithm calculates a t-value for each data point in each cluster. Samples that do not meet the 95% confidence level for the cluster will appear in the gray region. Editing SNPs can be done by right-clicking on a data point in the cluster plot or on the cell of the report table.



**Figure 2: Cartesian view of cluster plot displaying statistics.**

The SNP analysis report can be saved as Excel (\*.xls) or Text (\*.txt) files.

## Discussion

The size of a short DNA fragment following capillary electrophoresis will vary from the expected size. This mobility shift is due to the nucleotide sequence, dye used to label the fragment and run conditions. It is recommended that genotyping projects should be started and completed on the same instrument.

Various techniques have been developed to interrogate SNPs, including single nucleotide primer extension, SNPlex™, SNaPshot™, SNuPe™, SNPwave™ (KeyGene), SNP chips and DNA sequence analysis. Primer extension method has the advantage of accurate genotyping using a low number of unlabeled user-defined primers.

GeneMarker is highly accurate and easy to use. The robust sizing, efficient allele calling and mobility shift correction along with advanced clustering algorithms can quickly and efficiently determine SNP genotypes while maintaining the advantages of primer extension technology.

## References

1. SNP Fact Sheet. Human Genome Project Information. U.S. Department of Energy-Office of Science, Office of Biological and Environmental Research. [http://www.ornl.gov/sci/techresources/Human\\_Genome/faq/snps.shtml](http://www.ornl.gov/sci/techresources/Human_Genome/faq/snps.shtml).
2. Salim Mottagui-Tabar, MA Faghihi, Y Mizuno, PG Engström, B Lenhard, WW Wasserman and C Wahlestedt. 2005. Identification of functional SNPs in the 5-prime flanking sequences of human genes. *BMC Genomics*. 6:18. <http://www.biomedcentral.com/1471-2164/6/18>.
3. Yan H, W Yuan, VE Velculescu, B Vogelstein, KW Kinzler. 2002. Allelic variation in human gene expression. *Science*. 297(5584):1143.
4. Tiedemann R, K Moll, KB Paulus, M Scheer, P Williot, R Bartel, J Gessner and F Kirschbaum. 2007. Atlantic sturgeons (*Acipenser sturio*, *Acipenser oxyrinchus*): American females successful in Europe. *Naturwissenschaften*. 94(3): 213–217.
5. Pot D, J-C Rodrigues, P Rozenberg, G Chantre, J Tibbits, C Cahalan, F Pichavant, C Plomion. 2006. QTLs and candidate genes for wood properties in maritime pine (*Pinus pinaster Ait.*). *Tree Genetics & Genomes*. 2: 10–24.