

GeneMarker[®] Software for SNPWave[™] Analysis

David Hulce, Wan Ning, ChangSheng 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—classified as non-synonymous SNPs (nsSNPs) that occur in the coding region of a gene or as regulatory SNPs (rSNPs) that occur in the promoter region of a gene—are often associated with altered 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)** quantitative trait loci (QTL) **(5)** or genotyping plants and animals **(6)**.

One high-throughput method to determine SNP genotypes is SNPWave (Keygene N.V.). SNPWave uses multiplex oligonucleotide ligation amplification of allele-specific probes coupled with AFLP[®]-primer selective amplification. SNPWave can detect up to 100 SNPs **(6)**.

Circularizing padlock ligation probes are constructed that are specific to the SNP and flanking sequences. Locus-specific probes will hybridize to complementary denatured genomic DNA. Allele-specificity is determined by the SNP at the 5' end of the padlock probe. Probes that contain a 5' nucleotide complementary to the SNP will be ligated and amplified by PCR in subsequent reactions. Probes that do not contain a 5' nucleotide complementary to the SNP will not be ligated and will not be amplified **(6)**.

The padlock probes contain stuffer regions and primer binding sites for AFLP-selective amplification. The ligated padlock probe is amplified using fluorescently-labeled +2-selective and unlabeled non-selective AFLP primers. The stuffer region provides length discrimination between alleles and among loci. SNPs are separated by 2 base pairs and loci are separated by 3 bp. The fragments are separated by size using capillary electrophoresis. Fragment dye color and length indicate SNP locus and allele.

To fully utilize the investigative potential of SNPWave, a robust genotyping and data analysis system should be employed. GeneMarker genotyping software is designed for fast, accurate and efficient analysis and reporting of SNP data generated by capillary electrophoresis.

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, essential for data generated by SNPWave. The software is capable of analyzing data files from major capillary electrophoresis systems including ABI, MegaBACE[™] and Beckman. GeneMarker, in combination with our JelMarker[™], can also be used for analysis of slab gel outputs from Li-Cor and FMBIO.

Procedure

The software settings shown below are recommended to produce the best results for SNPWave analysis. The process from raw data to review of the analysis report occurs in three steps: size calling of the data, creating the SNPWave 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: ET400-R
3. Standard Color: red
4. Analysis Type: *SNPlex*

Data Process-*SNPlex* Analysis

1. Accept defaults

Step 2: Create Panel

After the data is processed, create a 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. Rename any of the SNPs as necessary to improve readability of final SNP report. The details for editing the panel can be located in the manual.
3. Click OK and the panel is automatically uploaded to the **Panel Editor**.
4. 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 panel for the analysis (use all previously selected options).
2. To activate the SNP analysis report, click **Applications** in the menu bar and select **SNPlex/SNaPshot**.

Results

The SNP analysis report window displays a synthetic gel image, list of samples, cluster plot to analyze peak information and assign SNP genotypes, and sample electropherogram (**Figure 1**). A SNP genotyping report is available (on right, **Figure 1**) and can be saved as Excel (*.xls) or Text (*.txt) files. The information contained in the report is interlinked. Double-clicking on the cell in the report containing SNP 107/109 for locus SNP_06 of sample E01 highlights the cell, corresponding sample ID in the sample list, data point in the cluster plot and locus in the electropherogram.

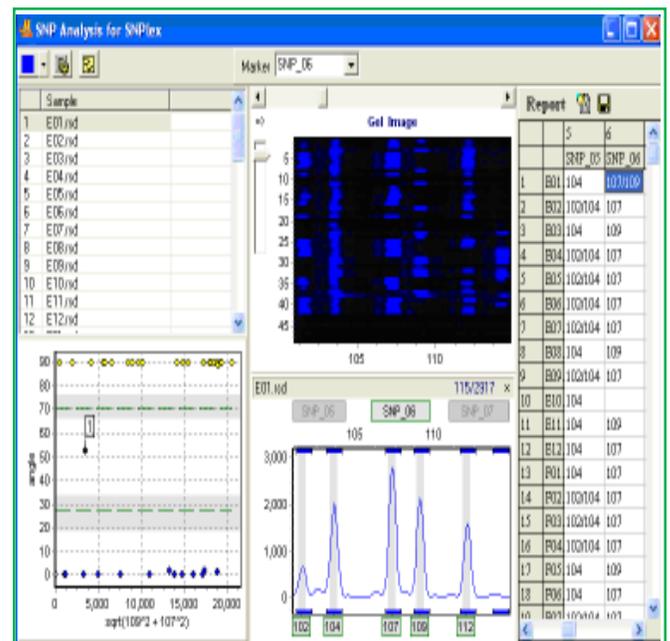


Figure 1: SNP Analysis for SNPWave report

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 derivative to remove 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. SNP_06 for sample E02 is heterozygous for alleles 107 and 109 (**Figure 1**).

The default cluster plot layout is 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 cluster 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.

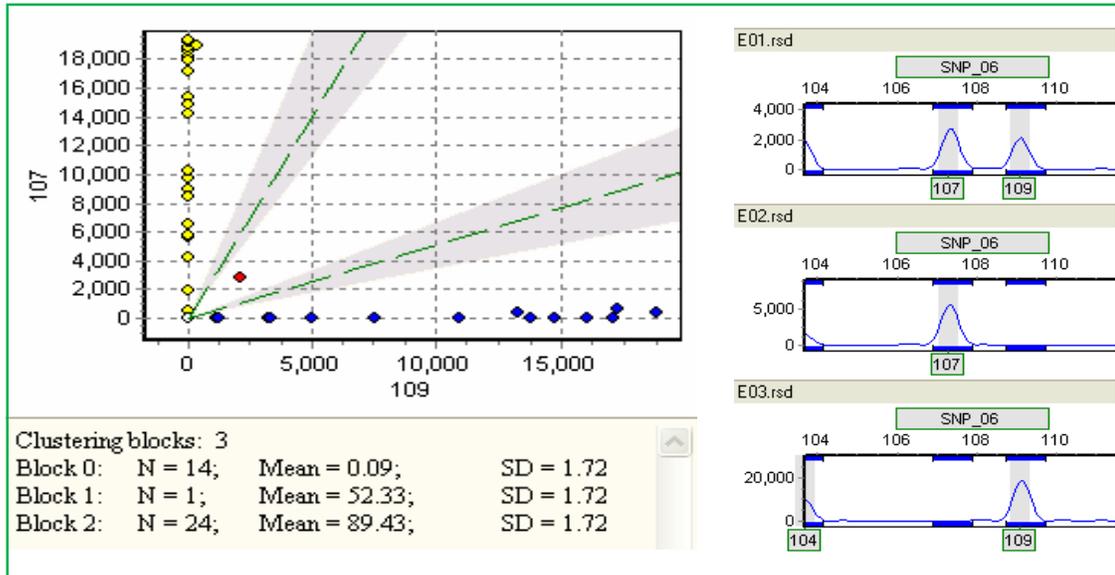


Figure 2: Cartesian view of cluster plot displaying statistics (left) and examples of heterozygous 107/109, homozygous 107 and homozygous 109 tomato lines (right).

Discussion

The data presented in Figure 1 was obtained from a collection of tomato lines. This data set shows efficient allele calls for *.rsd files from MegaBACE DNA Analysis System. One can quickly confirm allele calls using the cluster plot and SNP report. For SNP_06 the majority of the samples are homozygous for either allele 107 or allele 109. One sample is heterozygous for both alleles (**Figure 2**, right).

Various techniques have been developed to interrogate SNPs, including single nucleotide primer extension (SNUPE), SNPlex™, SNaPshot™, SNUpe™, SNPWave®, SNP chips and DNA sequence analysis. SNPWave has the advantage of accurate high-throughput genotyping of up to 100 SNPs.

GeneMarker is highly accurate and easy to use. The robust sizing and efficient allele calling along with advanced clustering algorithms can quickly and efficiently determine SNP genotypes while maintaining the advantages of SNPWave 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.
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.
6. van Eijk, MJT, JLN Broekhof, HJA van der Poel, RCJ Hogers, H Schneiders, J Kamerbeek, E Verstege, JW van Aart, H Geerlings, JB Buntjer, AJ van Oeveren and P Vos. 2004. SNPWave™: a flexible multiplexed SNP genotyping technology. *Nucleic Acids Research* 32(4): e47.