

# Genetic Diversity Resequencing with Mutation Surveyor<sup>®</sup>

Kevin LeVan, Ni Shouyong, Jonathan Liu

## Introduction

The tomato (*Solanum lycopersicum*, formerly *Lycopersicon lycopersicum*) is a member of the Solanaceae or nightshade family, native to the western coast and Andes mountains of South America and to the Galapagos Islands (1). Patterns of single-nucleotide polymorphisms (SNPs) and insertions and deletions (indels) in DNA sequence traces can be used to analyze the distribution of genetic diversity among cultivars and within genes (2). *Mutation Surveyor has also been used to successfully analyze sequencing data from other crops such as sugar cane, barley and potato.*

Large resequencing projects demand fully automated calls of SNPs and homozygous and heterozygous indels to identify genetic relationships of wild-relative and domesticated strains. Mutation Surveyor is the only software package that detects heterozygous indels by deconvoluting DNA sequence traces. Mutation Surveyor also provides an accuracy of >99% in detecting homozygous and heterozygous SNPs (3), when both forward and reverse traces are of Phred 20 quality. Mutation Surveyor software features ease of use, low-learning curve, full automation and an exclusive migration time comparison to detect both homozygous and heterozygous indels.

## Procedure

### Data Entry (open files window)

1. Add 1- or 2-directional sample traces into the Sample Files field.
2. Add GenBank or annotated control sequence into the GenBank Sequence File field.

### Variation Detection

1. Click OK.
2. Click Run.

### Data Analysis

1. After processing the data, a variation report is displayed.
2. Double-clicking on any one of the cells will activate the graphic analysis display where the variation can be reviewed and edited.

## Results

A fragment of the gene ATGPAT6/GPAT6 was sequenced in three tomato samples: LR (Landrace) 015 is from Colombia, LR018 is from Peru and LR029 is from Ecuador. LR015 was used as a control sequence since Columbia is closest to the area of domestication (Central America). The LR015.\*.ab1 DNA sequence was opened in the Advanced GBK File Editor (the user can easily revise control sequences as new information becomes available), SNP 88C/G was added to the variation field (for illustrative purposes) and the sequence was saved as a \*.gbk file that was entered into the GenBank File field for variation analysis.

The report table (Fig 1) summarizes detected variations: three SNPs, one homozygous deletion (125\_130delGAGAAG) and one heterozygous deletion (125\_130het\_delGAGAAG). The purple highlighting indicates the SNP of interest that was recorded in the \*.gbk file. Blue font indicates variations of high confidence; green font indicates a negative SNP (displaying negative SNPs can be turned off). The variation report table can be saved in \*.txt, \*.xml, \*.htm or \*.xls formats.

No.	Sample File	Reference File	Gene	Mut#	Mutation1	Mutation2	Mutation3
1	LR015-LeSNP30R-061305-G10-14.ab1	LR15-Columbia-ATG	ATGPAT6/GPA	1		88C>G\$100	
2	LR018-LeSNP30R-061305-B01-03.ab1	LR15-Columbia-ATGIATGPAT6/GPA	3	61T>C\$104	88C>G\$100	125_130delGAGAAG	
3	LR029-LeSNP30R-061305-E02-10.ab1	LR15-Columbia-ATGIATGPAT6/GPA	3	61T>TC\$110	88C>CG\$33	125_130het_delGAGAAG	

Figure 1: Variation report table

The Graphic Analysis Display is arranged with the reference file (created from the \*.gbk file) at the top, the sample trace in the center and the mutation electropherogram at the bottom (Figures 2 and 3). The mutation electropherogram reports the differences between the reference and sample traces. The homozygous deletion in LR018 can be identified by the red bar denoting the nucleotides deleted from the sample trace; the sample trace has been gapped to align the sample trace to the reference trace (Fig 2). The heterozygous indel in LR029 can be identified by the brown bar above the nucleotides deleted in the sample trace (Fig 2).

Sample LR029 contains the variation 88C>CG (Fig 3). Purple highlighting indicates that this variation matches the SNP entered in the \*.gbk file—88C/G. The negative SNP (88C>G) for sample LR018 is displayed in green and the short vertical bar in the mutation electropherogram indicates the position of the reported SNP. The nucleotides displayed next to the short vertical bar are associated with that SNP.

Figure 2: Graphic analysis display comparing the heterozygous deletion (125-130GAGAAG) of LR029 and the homozygous deletion (125-130GAGAAG) of LR018.

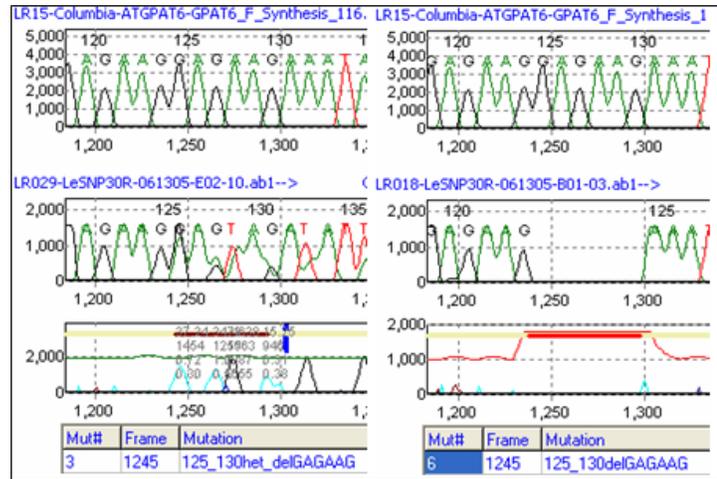
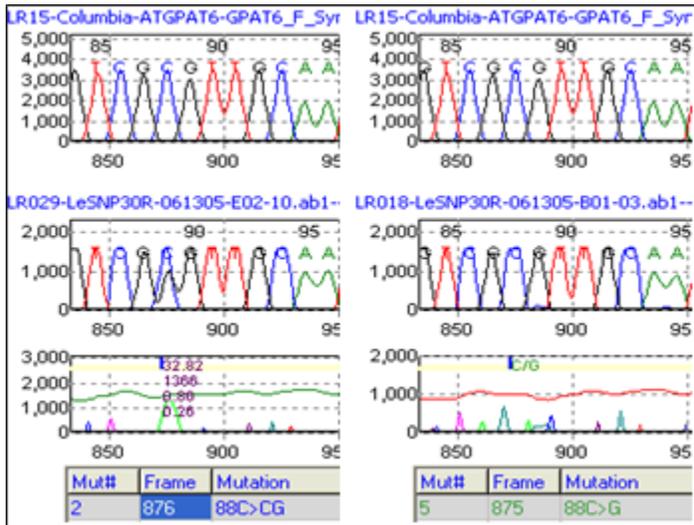


Figure 3: Graphic analysis display comparing a variation that matches a SNP recorded in the \*.gbk file (LR029) and a negative SNP—no matching variation at that position (LR018).



## Discussion

Variations displayed by LR029 are SNPs 61T>TC, 88C>CG and a heterozygous deletion 125\_130, while LR018 contains one SNP 61T>C and a homozygous deletion 125\_130, relative to LR015. Mutation Surveyor can quickly identify variations associated with diversity (SNPs, homozygous and heterozygous indels). Variation information may be useful for strain and germplasm verification, phylogeny studies, or possibly, identifying SNPs associated with traits of importance.

SoftGenetics' mutation detection software utilizes anti-correlation technology and a unique physical comparison of the migration time of the sample and the reference traces, rather than relying on the text base call. By monitoring the migration time of the sample and the reference, Mutation Surveyor can accurately detect het-indels while eliminating false calls caused by text-based comparison and alignment.

## Acknowledgements

We would like to thank the following for providing the sequencing traces from tomato: **Angela M Baldo**<sup>1,2</sup>, **Yizhen Wan**<sup>2,3</sup>, **Warren F Lamboy**<sup>1</sup>, **Charles J Simon**<sup>1,2</sup>, **Joanne A Labate**<sup>1,2</sup>, **Susan M Sheffer**<sup>1</sup>

<sup>1</sup> USDA-ARS Plant Genetic Resources Unit 630 W. North St. Geneva, NY 14456 USA

<sup>2</sup> Cornell University Department of Horticultural Sciences 630 W. North St. Geneva, NY 14456 USA

<sup>3</sup> Northwest Sci-Tech University of Agriculture and Forestry Shaanxi Province, Yangling, China

## References

- Peralta, I.E. and D. M. Spooner. 2001. Granule-bound starch synthase (GBSSI) gene phylogeny of wild tomatoes (*Solanum* L. section *Lycopersicon* [Mill.] Wettst. subsection *Lycopersicon*). American Journal of Botany 88(10): 1888-1902.
- Angela M Baldo, Yizhen Wan, Warren F Lamboy, Charles J Simon, Joanne A Labate, Susan M Sheffer. P171. NP Validation and Genetic Diversity in Cultivated Tomatoes and Grapes. Plant & Animal Genomes XV Conference, San Diego, CA. January 13-17, 2007.
- Haines, et al. 2005. Complement Factor H Variant Increases the Risk of Age Related Macular Degeneration. Science 308: 419-21.