

Hypervariable Mutation Detection with Mutation Surveyor®

Kevin LeVan, ChangSheng Jonathan, Liu

Introduction

Organisms contain areas in the genome with hypervariability. A bacteriophage that infects *Bordetella bronchiseptica*, a relative of the bacterium that causes whooping cough, contains genes to specifically change its proteins thus permitting binding to different cell receptors (1). These genes allow the phage to rapidly evolve new variants that can recognize and attack bacteria that may have become resistant to the previous phage. Similarly, infection by the influenza virus requires receptor binding through cellspecific tropism (2). Several strains of *Helicobacter pylori*, the organism responsible for many cases of peptic ulcers and stomach cancer, show antibiotic resistance possibly due to efficiency of DNA repair or DNA polymerase (3). Additional causes for high mutation frequency come from antigenic variation found in sexually transmitted HIV (4) and the development of drug resistance mutations as in the case of the blood borne pathogen Hepatitis C virus (HCV).

Projects involving a high level of variability demand fully automated calls of SNPs and homozygous and heterozygous indels to identify variance between strains. Mutation Surveyor software is the only software package that detects heterozygous indels by deconvoluting DNA sequence traces. Mutation Surveyor software provides an accuracy >99% in detecting homozygous and heterozygous SNPs (5), when both forward and reverse traces are of Phred 20 quality. The software features ease of use, a low-learning curve, full automation, and an exclusive migration time comparison to detect indels.

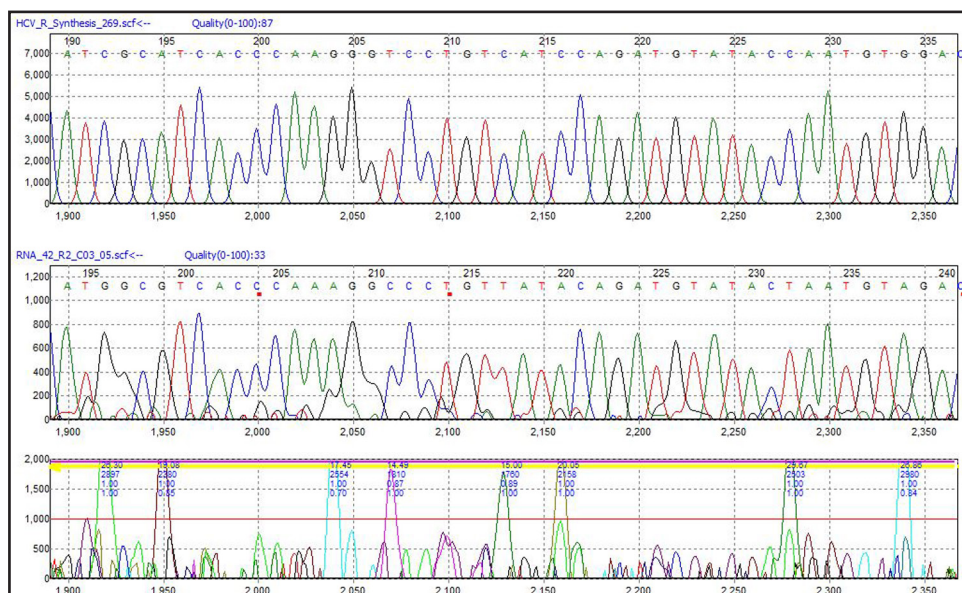


Figure 1: HCV sample (middle) shows high frequency of mutations when compared to the reference (top). The bottom trace displays the level of correlation between the sample and reference and mutations are indicated with high intensity peaks.

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. Adjust Process Settings Mutation Detection to High Sensitivity.
3. Click Process Run.

Data Analysis

1. After processing the data, a variation report is displayed.
2. Click Reports-Customer reports to select Vertex 2 report and choose settings appropriate for analysis.
3. Double-clicking on any one of the cells will activate the graphic analysis display where the variations can be reviewed and edited.

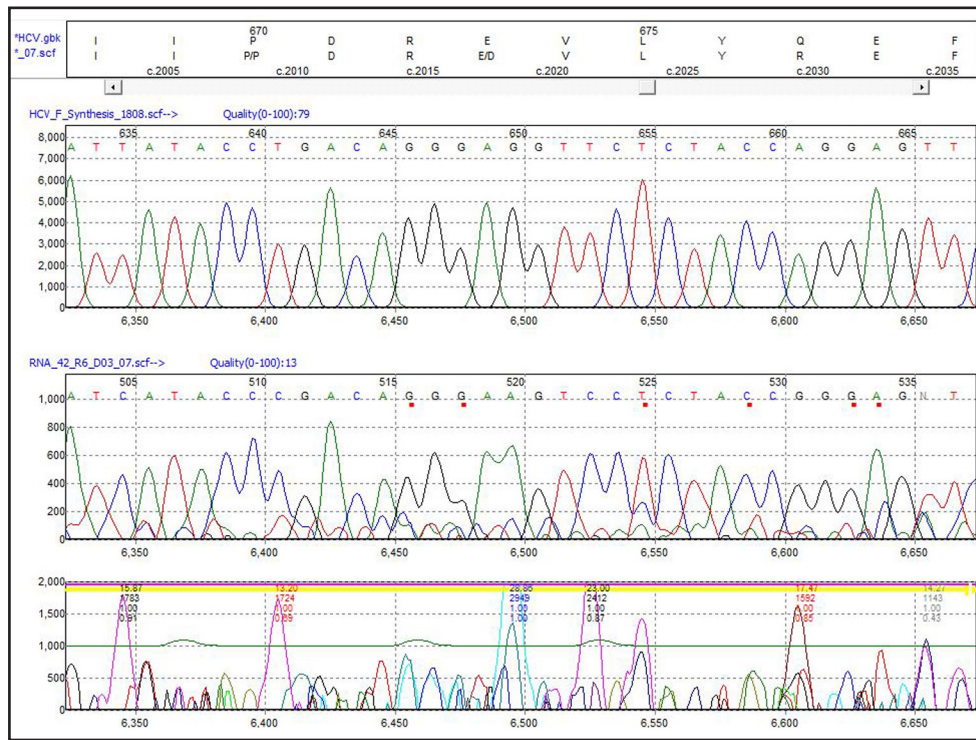


Figure 2: The HCV sample contains homozygous and heterozygous mutations; some altering the amino acid sequence.

Results

Mutation Surveyor software makes it easy to visualize variations between samples. When analyzing samples that have a high mutation frequency (see Figure 1) the High Sensitivity setting is best to allow for more mutation calls to be displayed. While Medium Sensitivity is more stringent for 2 directional data and will filter out calls that are more likely false positives, High Sensitivity is less selective in filtering mutation calls. As displayed in Figure 2, a high level of variability is evident when comparing the HCV samples to the reference trace. In this example, the mutation at residue 670 results in a silent mutation where the reference nucleotide T is substituted with an C. The mutation at residue 673 contains a homozygous substitution G>A changing the amino acid from glutamic acid to aspartic acid.

Contig	30	35	40	45															
HCV.gbk	N	Q	V	E	G	E	V	Q	I	V	S	T	A	T	Q	T	F	L	A
RNA_42_B03_03.scf	-	Q/Q-	-	-	-	-	W	-	W	-	-	-	-	-	A	-	-	-	-
RNA_42_C03_05.scf	-	-	-	-	-	-	I	-	-	-	-	-	-	A/GT/A-	-	-	L	L	-
RNA_42_R1_B03_03.scf	-	Q/Q-	-	-	-	-	W	-	W	-	-	-	-	A	-	-	-	-	-
RNA_42_R2_C03_05.scf	-	-	-	-	-	-	I	-	-	-	-	-	-	A/GT/A-	-	-	L	L	-

Figure 3: The amino acid sequence for each sample in Contig 1 is aligned in the Vertex report to the reference sequence, all variations are displayed with the amino acid lettering. Dashes indicate no variance within nucleotides for respective amino acid residue.

Mutation Surveyor software gives you the ability to compare traces within each contig to the reference trace in order to easily depict the location of variants. Figure 3 shows the Vertex report with the reference amino acid sequence at the top. All of the samples within contig 1 are aligned to the reference sequence; the dashes refer to amino acids that are the same as the reference and the amino acid letter is displayed in cases that differ from the reference sequence. This report is also able to display the variants associated with the nucleotide sequence and variations associated with insertions and deletions. When saved as an *.xls file the report includes color-coding of the mutations as well as the number of mutations in each sample (see Figure 4).

The software package is capable of assembling all variations from multiple sequences or clones within one or multiple projects. By utilizing the Vertex 2 report (see Figure 5), a matrix is created displaying each sample and all relevant positions with variations. Additional options include the ability to show either amino acid or nucleotide information, show only the variations and show only confirmed mutations. Columns showing the consensus of all the samples and comments created during the analysis are also available within this report.

Sample	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46						
Consensus	A	P	I	T	A	Y	A	Q	Q	T	R	G	L	L	G	C	I	I	T	S	L	T	G	R	D	K	N	Q	V	E	G	E	V	Q	I	V	S	T	A	T	Q	T	F	L	A	T						
RNA_42_B03_03.scf																		I/V										Q/C				I/V		I/V						A							T/T					
RNA_42_C03_05.scf										Q/C/T																																							L/L		T/T	
RNA_42_R1_B03_03.scf																													Q/C				I/V		I/V							A								T/T		
RNA_42_R2_C03_05.scf										Q/C/T																																								L/L		T/T
4 sequences									2	2								2										2					4		2						2	4			2			4				

Figure 4: The Vertex report can be saved in *.xls format, providing additional color-coding, total number of mutations in each sample and total number of samples with a mutation at each residue.

Discussion

Due to the increasing emergence of organisms with drug resistance mutations, pharmaceutical companies need ways to easily find these high frequency mutations. By adjusting the sensitivity setting, Mutation Surveyor software has the ability to analyze data of high variability aiding in the identification of these mutations. In the case where all of the samples show a high level of similar variants (see Figure 1) a homolog consensus sequence can be constructed and used as the reference. In addition to minimizing the number of mutation calls, an improvement in accuracy and a reduction in labor and analysis time will be gained. This is particularly useful for the determination of a patient’s treatment progress by analyzing mutations at different stages.

Mutation Surveyor software is superior by not relying solely on the trace’s base call. Through the use of anti-correlation technology and a unique physical comparison of the migration time for reference and sample traces, higher accuracy of calling is attained. This process also allows for the detection of heterozygous insertions and deletions while eliminating false calls caused by text-based comparison and alignment. This is especially helpful with samples that have a high mutation rate.

Samples exhibiting a high mutation frequency may contain genotypic mixtures. Mutation Surveyor software offers tools for the quantification of the viral strains within these sample mixtures by evaluating the relative intensity drop or gain with respect to its reference. The software will automatically evaluate all of the sequence traces, determine the best samples to set as controls, and identify the percentage of contribution each sample contains with respect to these controls. Manual assignment of the controls is available for additional user flexibility.

Index	Ref	RNA_42	RNA_42	RNA_42	RNA_42	RNA_42	RNA_42	RNA_42
33	V	I	I	I	I	I		
39	A	-	-	A/G	-	A/G		
40	T	A	A	A	A	A		
64	I	I/M	I/M	M	I/M	M		
69	G	D	D	-	D	-		
70	P	T	T	-	T	-		
91	S	S/T		S/T		S/T		
122	S	G		G		G		
139	S	>		>		>		
142	P	Q/H		Q/H		Q/H		
143	L	W		W		W		
153	L	I		I		I		
248	V	I				I	I	
296	D	D/Y				D/Y	D/Y	
316	R	R/T				R/T	R/T	

Figure 5: The Vertex 2 report assembles all contigs into one report. The first column contains the amino acid residues with detected variation. The second column displays the reference amino acid and the third column shows a consensus of amino acid variations within all of the samples. Subsequent columns describe the individual variations within each sample.

Color-coding is a useful feature integrated throughout Mutation Surveyor software to assist the user. The mutation depicted by the black font shown in Figure 2 has been marked as “confirmed” by the user. The mutations with blue font meet all mutation-calling criteria while the mutation with red font falls in a borderline range. In this case the drop factor, the measured decrease in intensity of the normal allele in the sample trace as compared to the reference trace, is within a range set for inspection by the user. Drop factor is one of the many criteria evaluated by the software to automate mutation calling. Background color-coding in reports assist in identifying missense mutations, reported variations and more.

Notes

Some software packages that are capable of detecting substitution changes include Sequencher™ from Gene Code, Ann Arbor, Michigan; SeqScape® from Applied Biosystems Inc., Foster City, CA; PolyPhred University of Washington, Seattle; inSNP, novoSNP, seeSNP, spotSNP, Codon Code Aligner, PolyBayer, and Paracel Agent. Mutation Surveyor software is the only software analysis package that can separate frame shift Indels and also identify mutations in samples containing hypervariability.

Acknowledgment

We would like to thank Tara Kieffer and Doug Bartels of Vertex Pharmaceuticals Incorporated for their collaboration and for providing data.

References

1. J F Miller, et al. 2004. Tropism switching in Bordetella bacteriophage defines a family of diversity-generating retroelements. *Nature*, 431: 476-81.
2. Ibrcevic, et al. 2006. Influenza virus receptor specificity and cell tropism in mouse and human airway epithelial cells. *Journal of Virology*, 80: 7469-7480.
3. Bjorkholm, et al. 2001. Mutation frequency and biological cost of antibiotic resistance in *Helicobacter pylori*. *Proceedings of the National Academy of Sciences USA*, 98: 14607-14612.
4. Wei-Kung Wang, et al. 1999. Hypervariable region 3 residues of HIV type 1 gp120 involved in CCR5 coreceptor utilization: therapeutic and prophylactic implications. *Proceedings of the National Academy of Sciences USA*, 96: 4558-4562.

Trademarks are property of their respective owners.