





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 gives you the ability to compare traces within each contig to the reference trace in order to easily depict the location of variances. **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 variances 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	667	668	669	670	671	672	673	674	675	676	677	678	Total #
Consensus	A	I	I	P	D	R	E	V	L	Y	Q	E	mutations
RNA_42_A03_01.sb1	-	-	-	-	-	-	-	-	-	-	R	-	5
RNA_42_D03_07.sb1	-	-	-	P/P	-	-	-	-	-	-	R	-	12
RNA_42_E03_09.sb1	A/A	-	-	-	-	-	-	-	-	-	R	-	3
3 sequences	1			1							2		

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, SoftGenetics' Mutation detection 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 and Mutation Explorer are 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 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_4	RNA_4	RNA_4	RNA_4	RNA_4	RNA_4	RNA_4	RNA_4	RNA_4
33	V	I	I	I						
40	T	A	A	A						
64	I	-	-	M						
122	S	G		G						
153	L	I		I						
200	A	A/V				A/V				
248	V	I				I				
332	S	P				P				
358	V	T				T	T			
557	F	L						L	F/L	
609	V	I							I	
677	Q	R							R	R

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 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 Mutation Surveyor 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 is the only software analysis package that can separate frame shift Indels and also identify mutations in samples containing hypervariability.

## Acknowledgements

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. Ibrcvic, 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.