Methylation Sequence Data Analysis with Mutation Surveyor Software

One of the methods of deducing the methylation state of a DNA molecule is the use of bisulfite treatment. Bisulfite treatment of DNA samples converts C to T in a DNA sequence, without converting methylated C bases in CpG sites. DNA sequencing is an excellent way to measure DNA methylation state. Mutation Surveyor software is a powerful software tool using unique comparison technology to detect nucleotide changes between two sequence traces.

In this application of our software, a GenBank sequence is used as a "ruler" to report nucleotide changes, including methylations and mutations. The reference trace is physically compared to the sample traces to find nucleotide differences. Mutation Surveyor automatically creates a synthetic reference trace from the GenBank sequence text or derivatives of the sequence text.

As previously mentioned, we use the GenBank sequence as a ruler for methylation studies. Therefore, the appropriate GenBank file **must** be input into the Reference Dialog Box. A reference trace is then synthesised from the derivative of the GenBank file(s).

Do NOT input a reference file in the box labeled Reference Files.

Input the sample trace files into the box labeled Sample Files. The user can choose the method to modify the GenBank sequence based on your interest. The modifications are categorised as CG to TG, or CA to TA, or CC to TC, or CT to TT, or any combination above. If the user is not sure which one to choose, please use the Auto Methylation setting. In the Auto Methylation mode, Mutation Surveyor chooses the best conditions so that the number of mutations is minimised. The Auto Methylation mode will simplify data analysis.

Procedures to Analyze DNA Methylation Sequence Data:

1. Set Mutation Surveyor software to Methylation application in the Process > Options > Others: Methylation.

   If methylation is NOT checked, the software will process the data as normal for mutation detection.

   The GenBank file **must** be the forward strand so that CpG is well defined. If only the **reverse complement** sequence is available, it must first be converted to forward strand using Tools: Seq File Editor and saved as reverse complement.

   Modification of GenBank files should be carefully considered. The modifications are categorized as CG to TG, or CA to TA, or CC to TC, or CT to TT, or any combination above. The reference synthetic trace is created from the sequence after modification parameters are set.

   a) Set by User means that user can choose the modification methods.
   b) Auto Methylation allows computer to modify the sequence text, minimizing the number of nucleotide changes (mutations) detection.

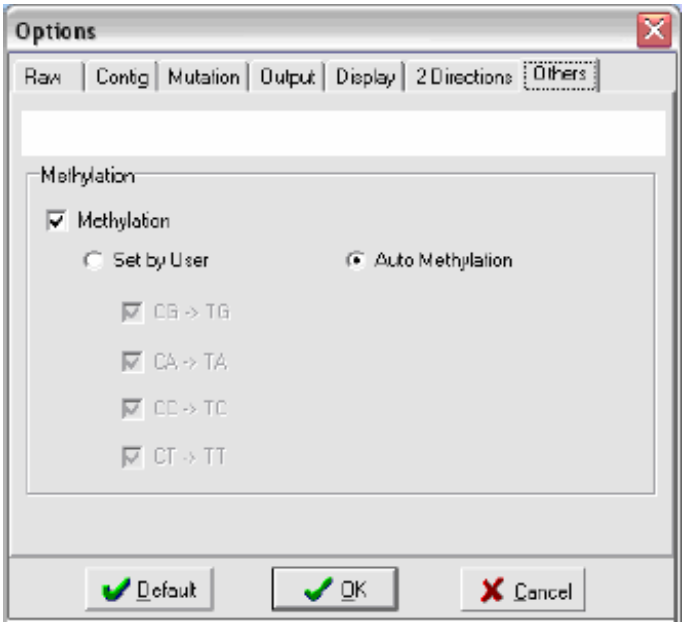


Figure 1. Methylation Options. Auto Methylation modifies the GenBank sequence to create the synthetic reference trace with minimum number of nucleotide changes. The user may also set the sequence modification.

2. Interface and data Interpretation:

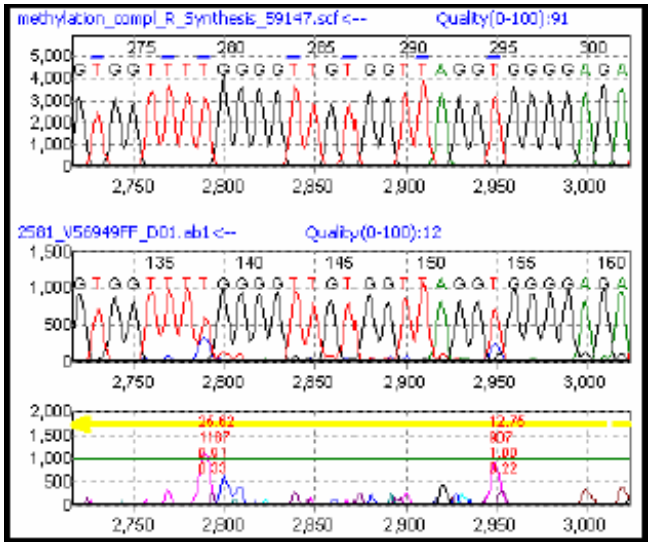
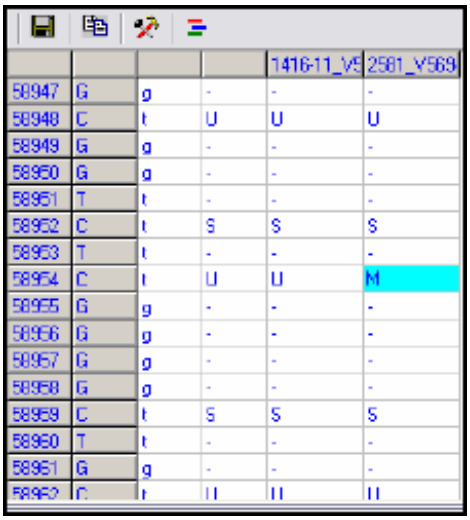


Figure 2. The top panel is the modified reference trace synthesized from the derivative GenBank file. The blue lines above the text indicate that the GenBank text was C (blue colour) prior to modification. The second panel is the sample trace. The third panel is the mutation detection panel.

3. Analysis Output:

Click the icon for the customer output  and select the template: Methylation.



				1416-11_V5	2581_V569
58947	G	g	-	-	-
58948	C	t	U	U	U
58949	G	g	-	-	-
58950	G	g	-	-	-
58951	T	t	-	-	-
58952	C	t	S	S	S
58953	T	t	-	-	-
58954	C	t	U	U	M
58955	G	g	-	-	-
58956	G	g	-	-	-
58957	G	g	-	-	-
58958	G	g	-	-	-
58959	C	t	S	S	S
58960	T	t	-	-	-
58961	G	g	-	-	-
58962	C	t	U	U	U

Figure 3: Report Table: M means methylated and U means unmethylated. S stands for success of the conversion from C to T, and I represents incomplete conversion.

Saving: Click on disk icon to save report in a text format (.txt) which can be exported to Excel or other program for printing.

4. Graphical output of methylation results:

To review global graphical results in text format, click on  icon in the Analysis output report.

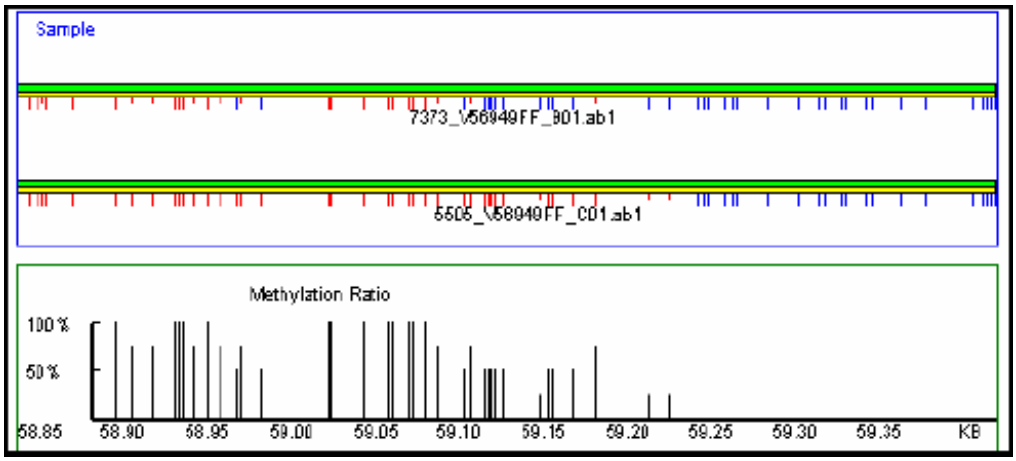


Figure 4. Graphic output of methylation results. Red color lines represent methylated sites and blue color lines represent unmethylated locations.