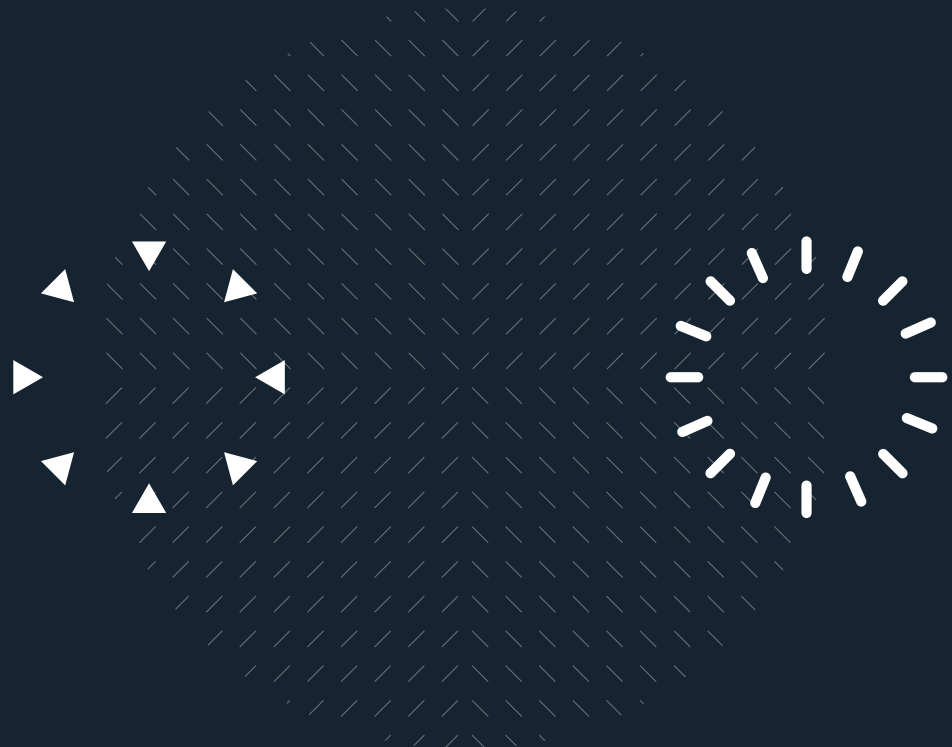




GeneMarker HTS

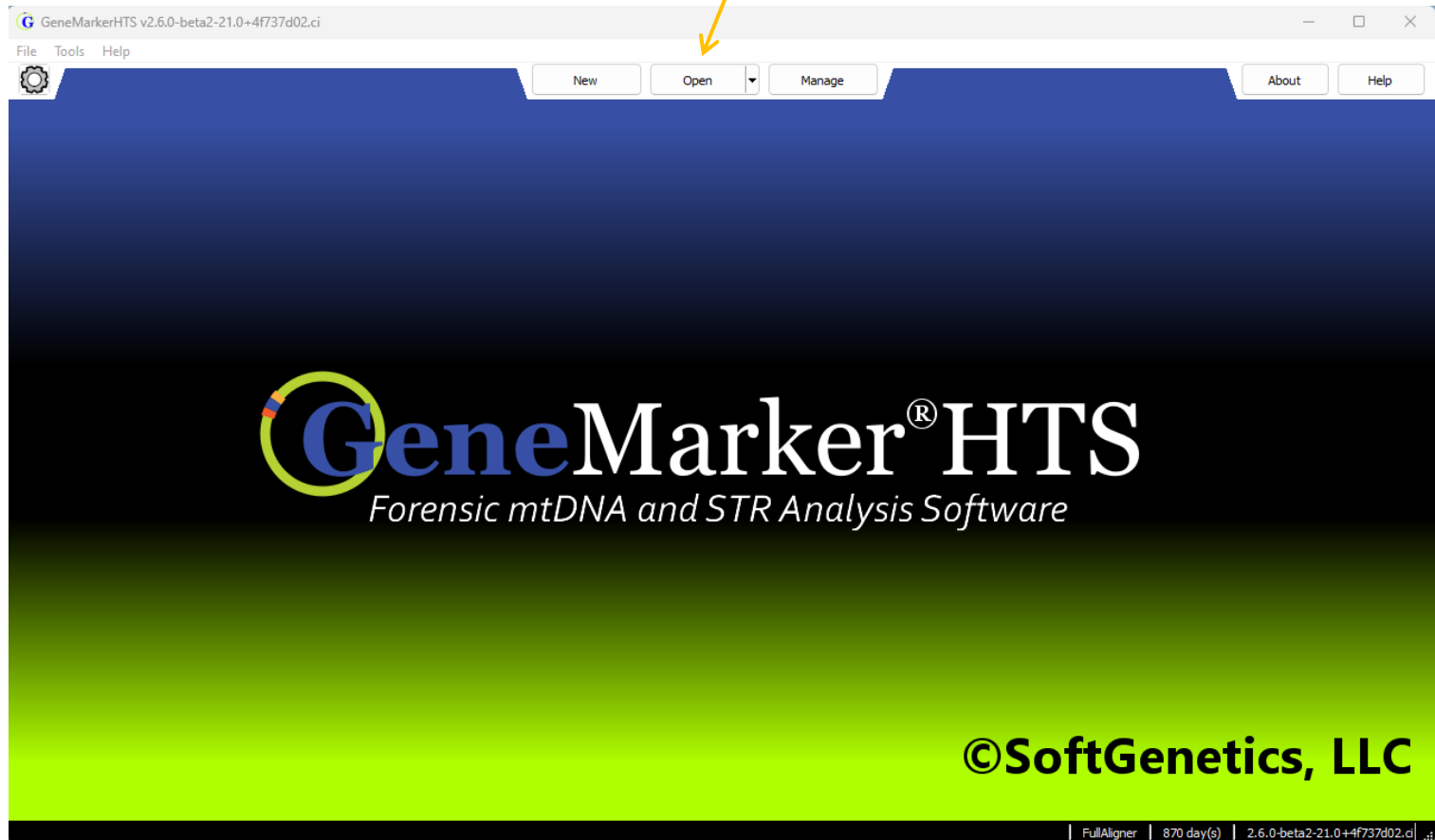
Quick Start Guide - mtDNA



August 2023

Launching GeneMarkerHTS

Upon launching the software, the user will have the option to start a *New* project or *Open* a previously saved project.



Creating a new project

The screenshot shows the 'New Project' dialog box with the following fields and options:

- Project Folder:** C:/Users/sarah/Desktop/HTS_QCTestData/PowerSeq_56GMY/projects/2023-06-20T11-13_v2.6.0-beta2-21.0+4f737d02.d
- Reference Path:** C:/Users/sarah/Downloads/mtDNA_NC_012920.gb
- Motif Path:** (Empty field)
- Panel:** Promega_PowerSeq_CRM_v1
- File 2:** L001_R2_001.fastq
- Annotations:**
 - Use Default Reference:** A checkbox that, when selected, makes the software use the revised Cambridge reference sequence (rCRS) for alignment. It is highlighted with a purple box and an arrow pointing to the 'Reference Path' field.
 - Use Default Motif:** A checkbox that, when selected, allows the user to use built-in motif data. It is highlighted with a purple box and an arrow pointing to the 'Motif Path' field.
 - Motifs:** A checkbox in the 'Panel Options' section that, when selected, allows the user to align data with or without using a Motif file. It is highlighted with a purple box and an arrow pointing to the 'Motif Path' field.
- Panel Options:**
 - ☐ Allow Primer Mismatches
 - Alignment Options:**
 - ☒ Consensus
 - ☐ Remove PCR Duplicates
 - ☐ Keep Only Proper Pairs
 - ☐ Merge Pairs
 - ☐ Pre-alignment merging
 - ☐ Post-alignment merging
 - ☒ **Motifs**
 - ☐ Match Proportion:
 - Percent \geq 90
 - Identity:
 - ☒ Percent \geq 90
 - ☐ Number \leq 0
 - Soft Clipping at 3' Q \leq 25
 - ☐ Clip mismatched ends
 - Sequencer:**
 - ☐ Ion Torrent
 - ☒ Illumina
 - ☐ Other
- Buttons:** Add, Remove, Remove All, Filter Settings, Clear Settings, OK, Cancel

Creating a new project

New Project

Project Folder:
C:/Users/sarah/Desktop/HTS_QCTestData/PowerSeq_56GMY/projects/2023-06-20T11-13_v2.6.0-beta2-21.0+4f737d02.d

Reference Path:
C:/Users/sarah/Downloads/mtDNA_NC_012920.gbk

Motif Path:
Fullpath of the Custom Motif File

Panel:

Pro

10 38 1001 001

10 38 1001 R1 001.fastq

10 38 1001 R2 001.fastq

Selecting **Allow Primer Mismatches** will allow a 1 bp difference in the primer binding region.

Consensus refers to performing local indel realignment based on the consensus sequence.

The user may enable the optional **Remove PCR Duplicates** step. This is used to remove reads identified as PCR duplicates based on their start positions. When using paired data, each pair of reads is considered together.

Keep Only Proper Pairs: Ignore reads not marked as part of a proper pair – both reads must be aligned and on opposite strands, 5' end of reverse strand must be to the right of the 5' end of the forward read.

Identity: Reads that are less similar to the reference than this percentage or number are not aligned.

Soft Clipping: The 3' end of reads are trimmed when basecall quality is low.
Note: Soft-clipping is also performed on the 3' ends of reads when mismatched bases are found near the end of the alignment.

The **Identity** and **Soft Clipping** settings will change depending on the **Sequencer** selected. If the user manually changes either setting, the program will change the selected **Sequencer** to **Other**.

Create Motif Edit Motif

Panel Options

☐ Allow Primer Mismatches

Alignment Options

Optional Steps:

☒ Consensus

☐ Remove PCR Duplicates

☐ Keep Only Proper Pairs

☐ Merge Pairs

☐ Pre-alignment merging

☐ Post-alignment merging

☒ Motifs

Filters/Clipping:

☐ Match Proportion:

Percent \geq 90

Identity:

☒ Percent \geq 90

☐ Number \leq 0

Soft Clipping at 3' Q \leq 25

☐ Clip mismatched ends

Sequencer:

☐ Ion Torrent

☒ Illumina

☐ Other

Amplicon Settings

PHI Settings

Add Remove Remove All Filter Settings Clear Settings OK Cancel

Creating a new project

New Project

Project Folder:
C:/Users/sarah/Desktop/HTS_QCTestData/PowerSeq_56GMY/projects/2023-06-20T11-13_v2.6.0-beta2-21.0+4f737d02.d

Reference Path:
C:/Users/sarah/Downloads/mtDNA_NC_012920.gbk

Motif Path:
Fullpath of the Custom Motif File

Panel:
Promega_PowerSeq_CRM_v1

Name	File 1	File 2
10_S8_L001_001	10_S8_L001_R1_001.fastq	10_S8_L001_R2_001.fastq

Amplicon Settings: Set amplicon regions when not using a panel

PHI Settings: Load a BED file or enter in regions you would like masked due to potential health concerns

AddRemoveRemove AllFilter Settings

Clear Settings

OKCancel

Create MotifEdit Motif

Panel Options
☐ Allow Primer Mismatches

Alignment Options
Optional Steps:
☒ Consensus
☐ Remove PCR Duplicates
☐ Keep Only Proper Pairs
☐ Merge Pairs
☐ Pre-alignment merging
☐ Post-alignment merging
☒ Motifs

Filters/Clipping:
☐ Match Proportion:
Percent \geq 90
Identity:
☒ Percent \geq 90
☐ Number \leq 0
Soft Clipping at 3' Q \leq 25
☐ Clip mismatched ends

Sequencer:
☐ Ion Torrent
☒ Illumina
☐ Other

Amplicon Settings
PHI Settings

Creating a new project

New Project

Project Folder:
C:/Users/sarah/Desktop/HTS_QCTestData/PowerSeq_56GMY/projects/2023-06-20T11-13_v2.6.0-beta2-21.0+4f737d02.d

Reference Path:
C:/Users/sarah/Downloads/mtDNA_NC_012920.gbk

Motif Path:
Fullpath of the Custom Motif File

Panel:
Promega_PowerSeq_CRM_v1

Name	File 1	File 2
10_S8_L001_001	10_S8_L001_R1_001.fastq	10_S8_L001_R2_001.fastq

Create Motif Edit Motif

Panel Options
☐ Allow Primer Mismatches

Alignment Options
Optional Steps:
☒ Consensus
☐ Remove PCR Duplicates
☐ Keep Only Proper Pairs
☐ Merge Pairs
☐ Pre-alignment merging
☐ Post-alignment merging
☒ Motifs

Filters/Clipping:
☐ Match Proportion:
Percent ≥ 90
Identity:
☒ Percent ≥ 90
☐ Number ≤ 0
Soft Clipping at 3' Q ≤ 25
☐ Clip mismatched ends

Sequencer:
☐ Ion Torrent
☒ Illumina
☐ Other

Add Remove Remove All Filter Settings

Clear Settings OK Cancel

The software will automatically group paired reads into the same sample, but this can be adjusted by right-clicking on rows in the table.

Sample names are automatically generated from filenames, but they can be edited by double-clicking the name in the input table.

Samples can be loaded using the *Add* button at the bottom of the *New Project* window. If paired reads are selected, they will be displayed together.

Compressed (**fastq.gz**) or uncompressed (**.fastq**) sequence files are the accepted input. Sample files can be removed individually or all at once using the *Remove* and *Remove All* buttons.

The *Filter Settings* button will allow the user to adjust settings for calling variants to meet their SOP or select *Default* to return them to their default values.

Selecting *OK* will save the selected settings, but they may be adjusted after alignment.

Mito Variant Filter Settings

The *Filter Settings* dialog allows for variant calling settings to be adjusted.

The screenshot shows the 'Filter Settings' dialog box with the 'Mito Variant' tab selected. The dialog has four tabs: 'Mito Variant', 'Mito Alignment', 'STR Filter Rules', and 'STR Stutter Rules'. The 'Mito Variant' tab contains the following settings:

- Variant Percentage**: A text input field with the value '10.0%'. A callout box explains: 'Percent of reads at a position with the variant'.
- Variant Allele Coverage**: A text input field with the value '200'. A callout box explains: 'Number of reads with the variant'.
- Total Coverage**: A text input field with the value '200'. A callout box explains: 'Total number of reads'.
- Allele Score Difference**: A checked checkbox. A callout box explains: 'Each minor allele has its score subtracted from the major allele score and the difference must be less than or equal to the set value.' The corresponding text input field has the value '10'.
- Allele Balance Ratio**: A checked checkbox. A callout box explains: 'The balance ratio is the maximum value of:
 - Allele forward % / Total forward %
 - Total forward % / Allele forward %
 - Allele reverse % / Total reverse %
 - Total reverse % / Allele reverse %This is calculated for each allele. Indels and SNPs have different cutoffs for filtering.' The corresponding text input field has the value '2.5'.
- SNP**: A text input field with the value '2.5'.
- Indel**: A text input field with the value '5.0'.

At the bottom of the dialog, there are four buttons: 'Save', 'Load', 'Default', and 'Cancel'. A callout box explains: 'The settings can be returned to their original state by using the *Default* button. A project settings file can be imported using the *Load* button.'

Mito Alignment Filter Settings

Filter Settings

?

×

Mito Variant

Mito Alignment

STR Filter Rules

STR Stutter Rules

☐ Template Length \geq 100

Reads shorter than this length will be filtered out

Mito Region

☒ Entire Reference

☐ Input Region

Chromosome	Start	End
chrM	16013	16569
chrM	1	592

Positions are inclusive (start through end)

Add

Remove

Load BED File

Remove All

Save

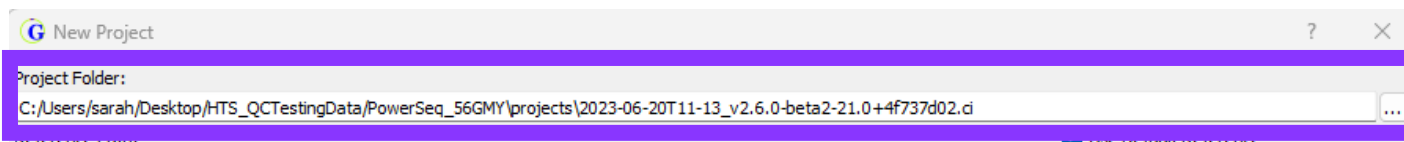
Load

Default

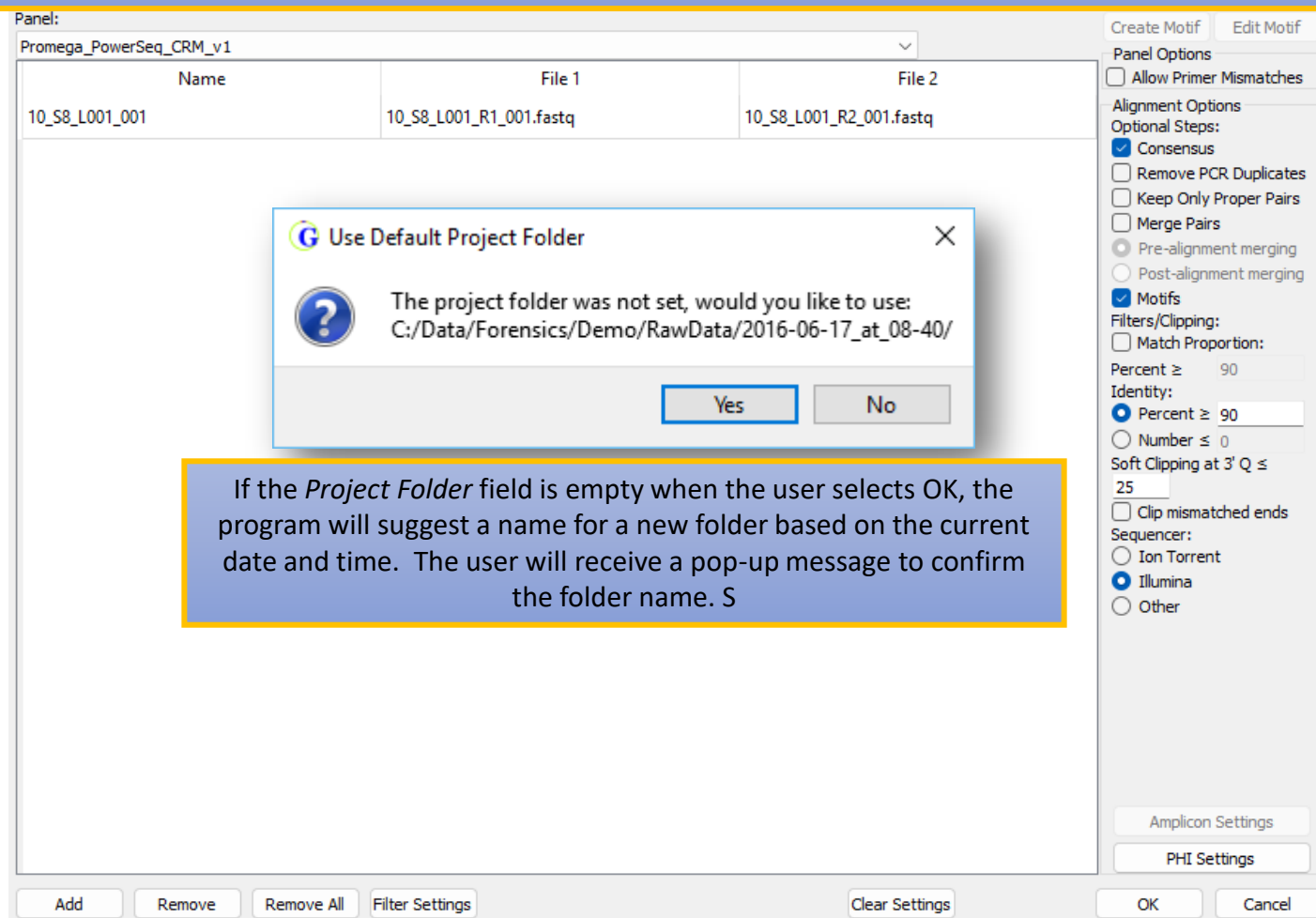
OK

Cancel

Creating a new project



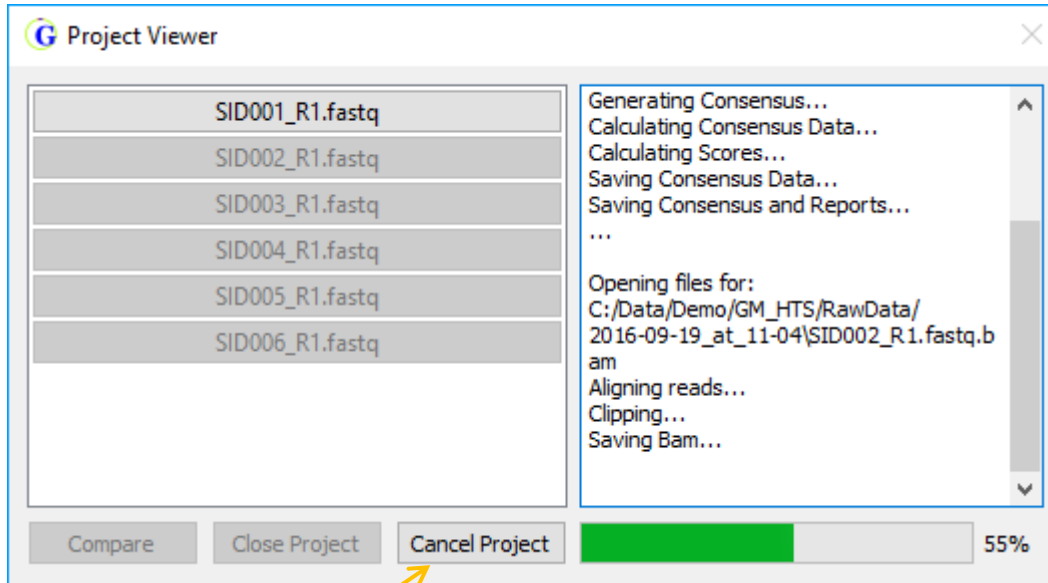
In the *Project Folder* field, a location can be selected for the data output by the program. A location can be set using the ellipsis button to the right of the field, or it can be typed manually. The folder will be created if it does not exist.



If the *Project Folder* field is empty when the user selects OK, the program will suggest a name for a new folder based on the current date and time. The user will receive a pop-up message to confirm the folder name. S

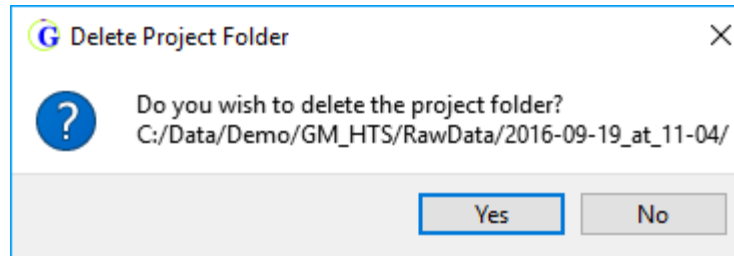
Sample Processing

After all the desired settings are chosen, selecting *OK* will begin alignment.



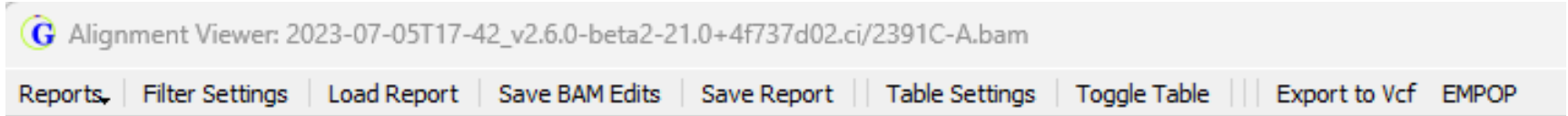
When a sample is finished it is possible to click on the button in the *Project Viewer* to open it - even before all samples finish processing

Projects can be canceled using the *Cancel Project* button. The *Project Viewer* will be closed after the next alignment finishes.



If the project is cancelled, the program will ask the user if they would like to delete the project folder that was created.

Viewing the Results



The **Reports** gives options for coverage Report, Template Length Report, and Read Length Report

The **Filter Settings** button will open the *Table Filter Settings* window that was available in the *New Project* window.

The **Load Report** button will allow the user to change to a different saved report.

The **Save BAM Edits** will save a new BAM file that includes the edits users have made.

The **Save Report** button will allow the user to save the report using the current filter settings and edited variants.

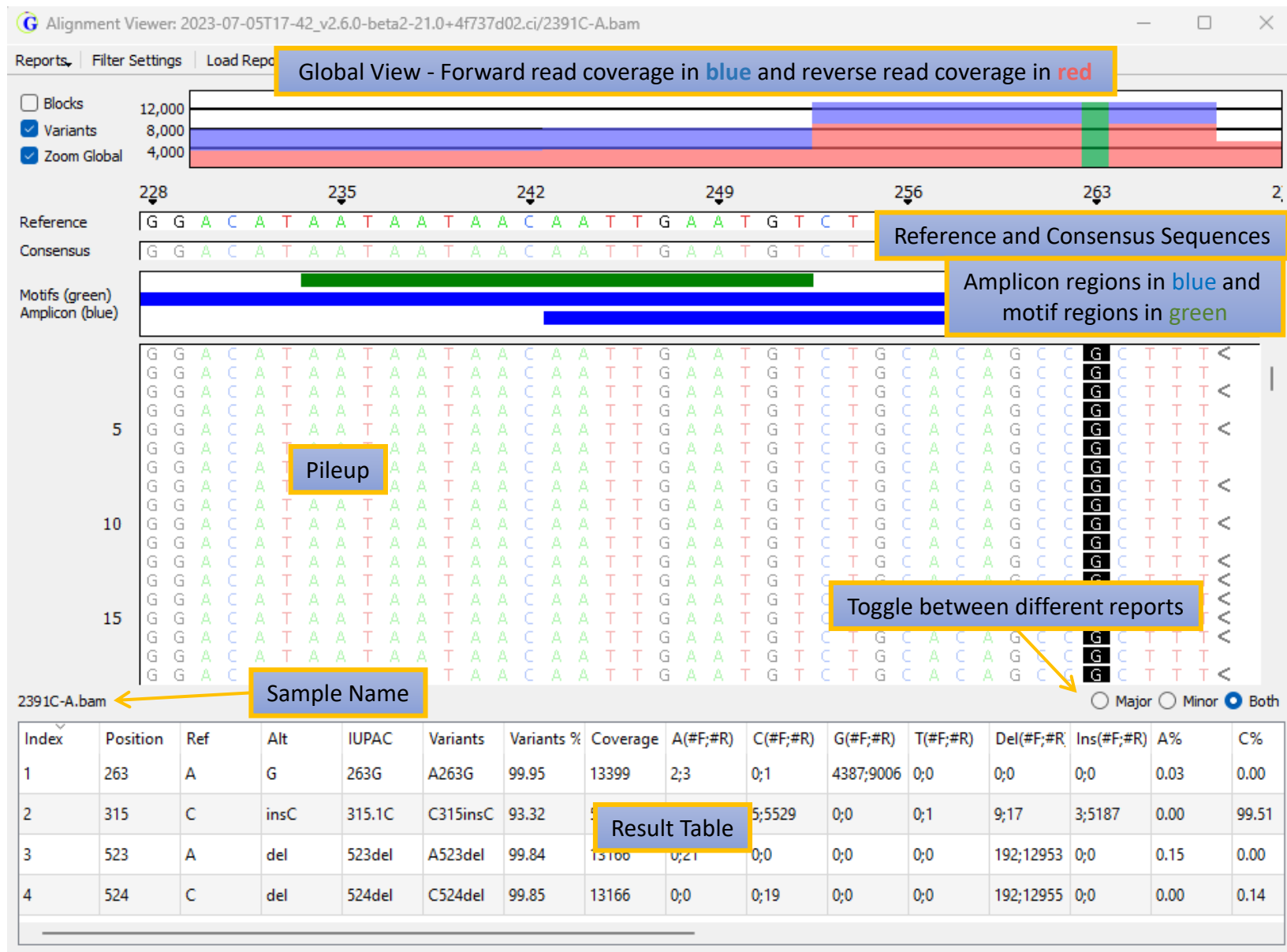
The **Table Settings** button will open the *Table Settings* window. Here the user can choose what information is displayed in the table.

The **Toggle Table** button will allow the user to display or hide the table below the pile-up.

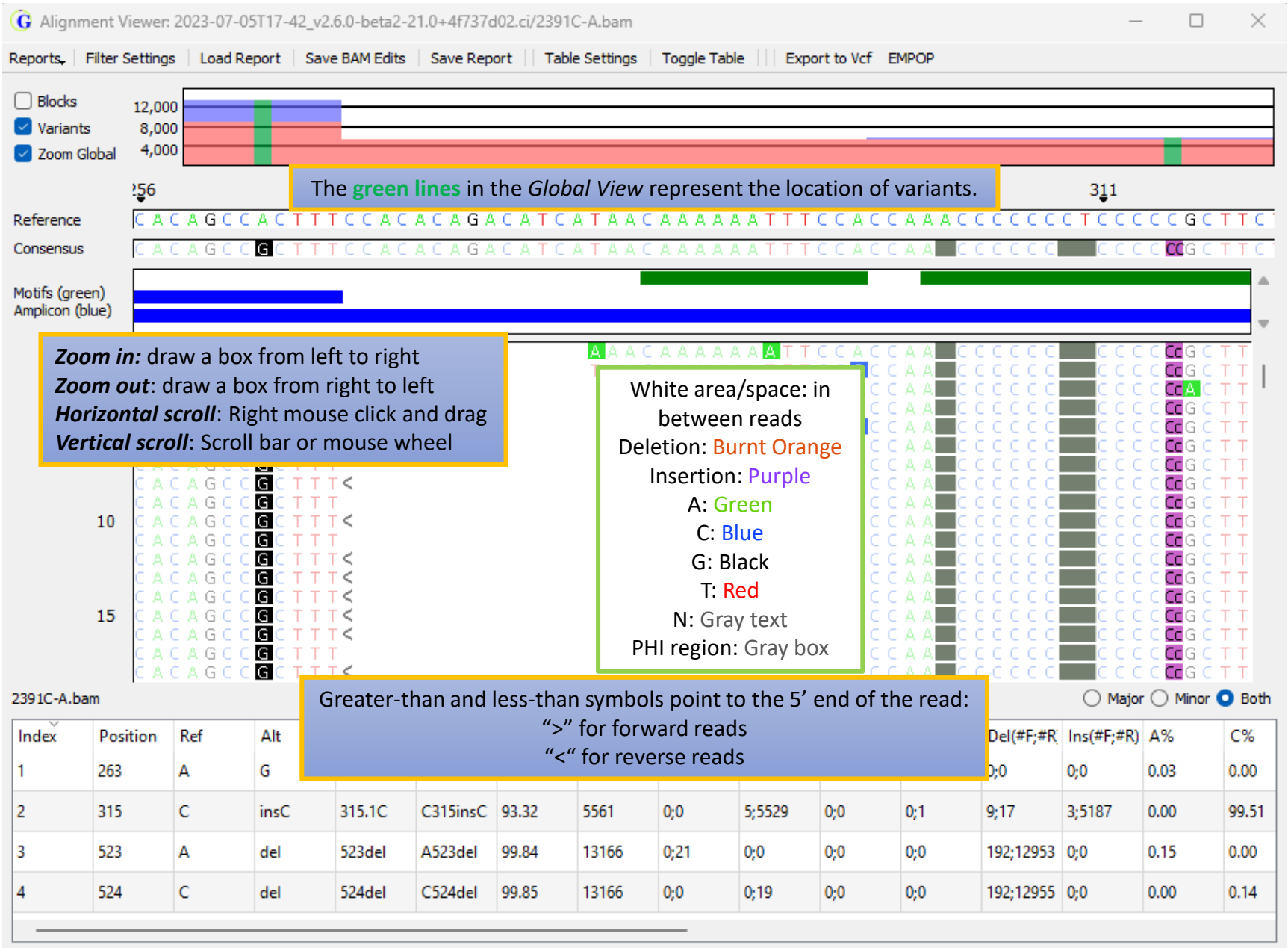
Export to VCF will export variants in VCF format

The **EMPOP** button streamlines the process of searching for profiles in the EMPOP database

Viewing the Results



Viewing the Results



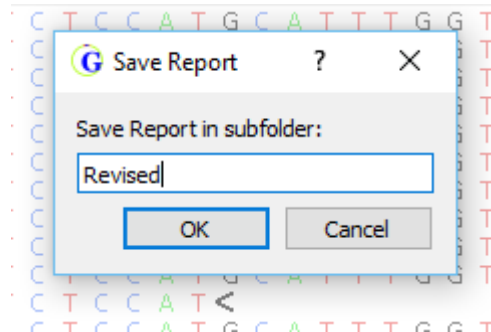
Adding and Removing Variants

Variants may be manually added (right-click in the pileup) or removed (right-click in the table).

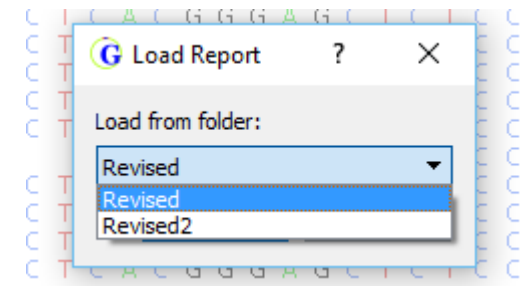
Added variants are **shaded green** and removed variants are **shaded red**.

Either kind of edit can include comments

Index	Position	Ref	Alt	Variants	Variants %	Coverage	A%	C%	G%	T%	Del%	Ins%
1	73	A	G	A73G	99.83	1805	0.11	0.00	99.83	0.05	0.00	0.00
2	114	C	A	C114A	12.19	2001	12.19	87.75	0.00	0.04	0.00	0.04
3	143	G	C	G143C	0.04	2240	0.00	0.04	99.95	0.00	0.00	0.00
4	146	T	C	T146C	12.01	2206	0.00	12.01	0.00	87.98	0.00	0.00
5	152	T	C	T152C	16.98	2220	0.00	16.98	0.00	83.01	0.00	0.00
6	195	T	C	T195C	16.75	2172	0.00	16.75	0.00	83.24	0.00	0.00



After changes to settings or manual edits, the report can be saved as a new report. The original project is opened by default, but all previously saved reports are maintained in subfolders with changelogs.



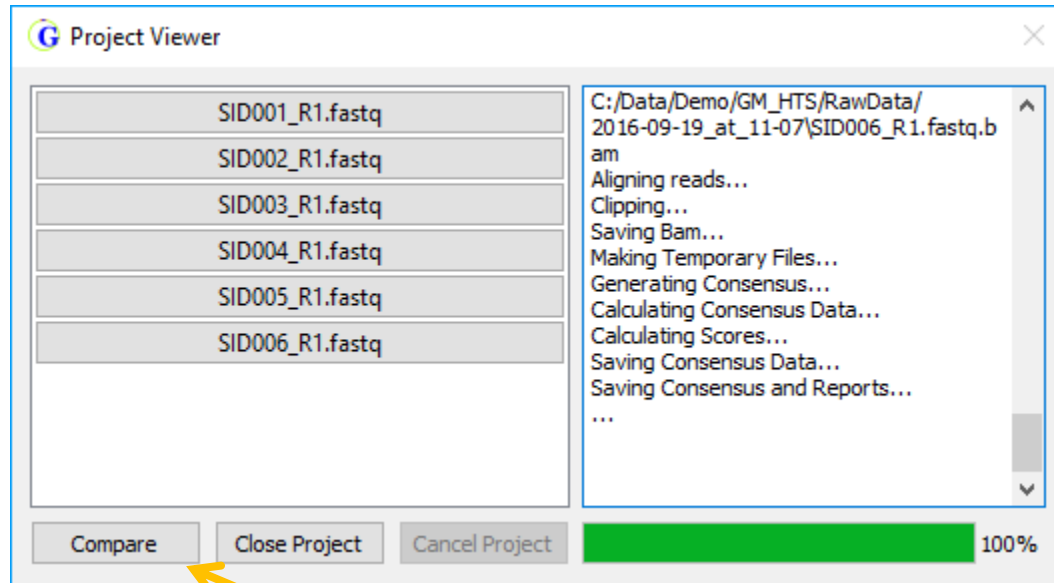
Output Files

The program will output the following pieces of information for each sample in the project:

- **BAM files:** Alignment results
- **Alignment Statistic Log:** Summary of alignment results
- **Amplicon:** Summary of reads in each amplicon.
- **AnalysisLog.json:** stats about mito alignment in an easy to parse (for computers) json format
- **Consensus Sequence:** Consensus sequence with primary alleles in FASTA format
- **Consensus Statistics:** Report listing the number of copies of each allele at each position in the sample.
- **Low Coverage Report:** Information about regions with low coverage
- **Major Report:** Variant report for primary alleles
- **Minor Report:** Variant report for minor alleles
- **Motifs:** List of all motifs used in project
- **Panel Primer Match Stats:** Information about amplicon sorting results
- **Results.bson:** analysis results in a compressed binary format
- **Trim Primer Log:** Information about amplicon sorting results
- **Variants:** List of all variants in sample
- **Project and Project Settings:** Used by software to track settings and data

10_S8_L001_001.bam
10_S8_L001_001.bam.bai
10_S8_L001_001.bam.cns
10_S8_L001_001.bam.cns.index
10_S8_L001_001_AlignmentStatistic.log
10_S8_L001_001_Amplicon.txt
10_S8_L001_001_AnalysisLog.json
10_S8_L001_001_ConsensusSequence.fa
10_S8_L001_001_ConsensusStatistic.txt
10_S8_L001_001_low_coverage_report.txt
10_S8_L001_001_major_report.txt
10_S8_L001_001_minor_report.txt
10_S8_L001_001_Motifs.txt
10_S8_L001_001_panelprimermatchstats.tsv
10_S8_L001_001_results.bson
10_S8_L001_001_TrimPrimerLog.log
10_S8_L001_001_variants.txt
project.pjt
project.settings

Comparison Tool



Any of the samples within a project can be opened in a *Comparison Viewer* using the "Compare" button

Comparison Tool

The *Sample to Sample Comparison* table (top half) shows a similarity table for the selected samples.

Comparison Viewer

Filter Settings | Major to Major | Major to Minor | Minor to Minor

Sample to Sample Comparison Proportion of Shared Variants: None All

	SID001_R1.fa	SID002_R1.fa	SID003_R1.fa	SID004_R1.fa	SID005_R1.fa	SID006_R1.fa
SID001_R1.fa	100%	100%	41%	41%	20%	17%
SID002_R1.fa	100%	100%	41%	41%	20%	17%
SID003_R1.fa	41%	41%	100%	100%	10%	18%
SID004_R1.fa	41%	41%	100%	100%	10%	18%
SID005_R1.fa	20%	20%	10%	10%	100%	7%
SID006_R1.fa	17%	17%	18%	18%	7%	100%

Row: ☐ Both ☒ Major ☐ Minor
Showing (Row's Major ∩ Column's Major) / (Row's Major ∪ Column's Major)

Column: ☐ Both ☒ Major ☐ Minor
☒ Show Percentage ☒ Use Jaccard Index

Comparisons can be made between major alleles, minor alleles, or both on each axis (row and column)

It is possible to show the absolute number of variants in common, or the percentage.

Comparison Tool

The *Variant Comparison* table (bottom half) simply shows the allele frequency of all variants called in at least one sample. Cells in the table are colored according to whether or not the variant was a major allele or minor allele in that sample, or if the total coverage was below the set threshold.

Variant Comparison							Major Allele	Minor Allele	Low Coverage
	SID001_R1.fa	SID002_R1.fa	SID003_R1.fa	SID004_R1.fa	SID005_R1.fa	SID006_R1.fa			
A73A	0%	0%	0%	6%	99%	0%			
A73G	99%	99%	99%	93%	0%	99%			
A93G	0%	4%	0%	0%	0%	0%			
T146C	0%	4%	0%	0%	0%	0%			
C150T	0%	15%	0%	0%	0%	0%			
T152C	0%	15%	0%	0%	0%	0%			

**Please contact tech_support@softgenetics.com
if further assistance is needed.**

**Visit our website for more information:
softgenetics.com**

Thank you for using GeneMarker HTS!