

GeneMarker[®]HTS

Quick Start Guide - mtDNA

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

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

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

Motif Path:
Fullpath of the Custom Motif File

Panel:
Promega_PowerSeq_CRM_v1

Use Default Reference

Use Default Motif

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

Motifs

Match Proportion:

Percent ≥ 90

Identity:

Percent ≥ 90

Number ≤ 0

Soft Clipping at 3' Q ≤ 25

Clip mismatched ends

Sequencer:

Ion Torrent

Illumina

Other

Amplicon Settings

PHI Settings

Add Remove Remove All Filter Settings Clear Settings OK Cancel

Selecting **Use Default Reference** will make the software use the revised Cambridge reference sequence (rCRS) for the alignment.

Alternatively, a *Reference file* may be loaded using the ellipsis icon to the right of the Reference field. The software supports input of FASTA or GBK files for the reference sequence.

If the user chooses to align using the *Motif* option, they may use the built-in motif data by selecting **Use Default Motif**.

If the user would like to use a customized motif file, it may be imported using the ellipsis icon to the right of the **Motif Path** field.

The user has the option to align their data with or without using a Motif file. **Motifs** provide phylogenetic alignment.

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.ci

Reference Path: Use Default Reference
C:/Users/sarah/Downloads/mtDNA_NC_012920.gb

Motif Path: Use Default Motif
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 [?] [X]

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

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

Motif Path: Use Default Motif
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

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

Buttons: Add, Remove, Remove All, Filter Settings, Clear Settings, Amplicon Settings, PHI 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.ci

Reference Path: Use Default Reference
C:/Users/sarah/Downloads/mtDNA_NC_012920.gb

Motif Path: Use Default Motif
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 \geq 90
Identity:
 Percent \geq 90
 Number \leq 0
Soft Clipping at 3' Q \leq 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 following settings and annotations:

- Mito Variant** (selected tab)
- Mito Alignment** (sub-tab)
- STR Filter Rules** (sub-tab)
- STR Stutter Rules** (sub-tab)
- Variant Percentage**: 10.0% (Annotation: Percent of reads at a position with the variant)
- Variant Allele Coverage**: 200 (Annotation: Number of reads with the variant)
- Total Coverage**: 200 (Annotation: Total number of reads)
- Allele Score Difference**: 10 (Annotation: 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.)
- Allele Balance Ratio**
- SNP**: 2.5
- Indel**: 5.0

The **Allele Balance Ratio** section includes the following text and list:

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 settings can be returned to their original state by using the *Default* button. A project settings file can be imported using the *Load* button.

Buttons at the bottom: Save, Load, Default, OK, Cancel.

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)

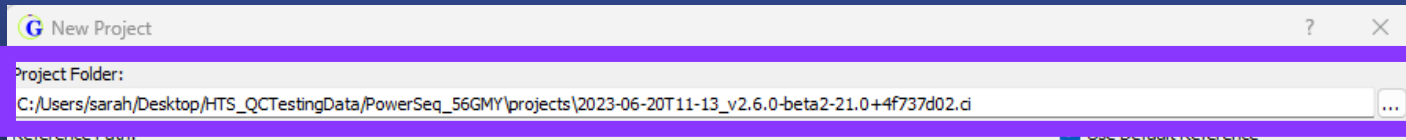
Add Remove

Load BED File Remove All

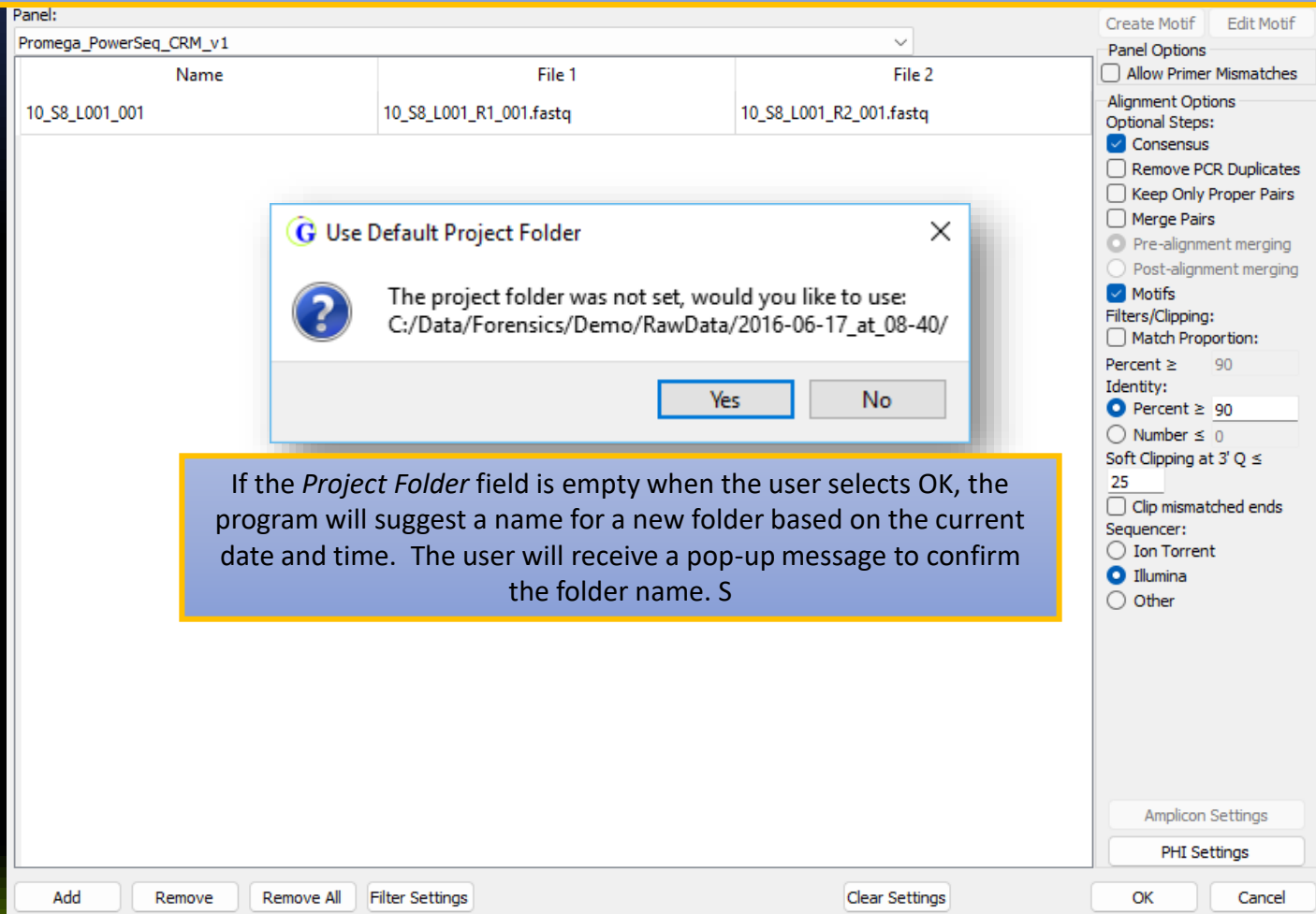
Save Load Default

OK Cancel

Creating a new project



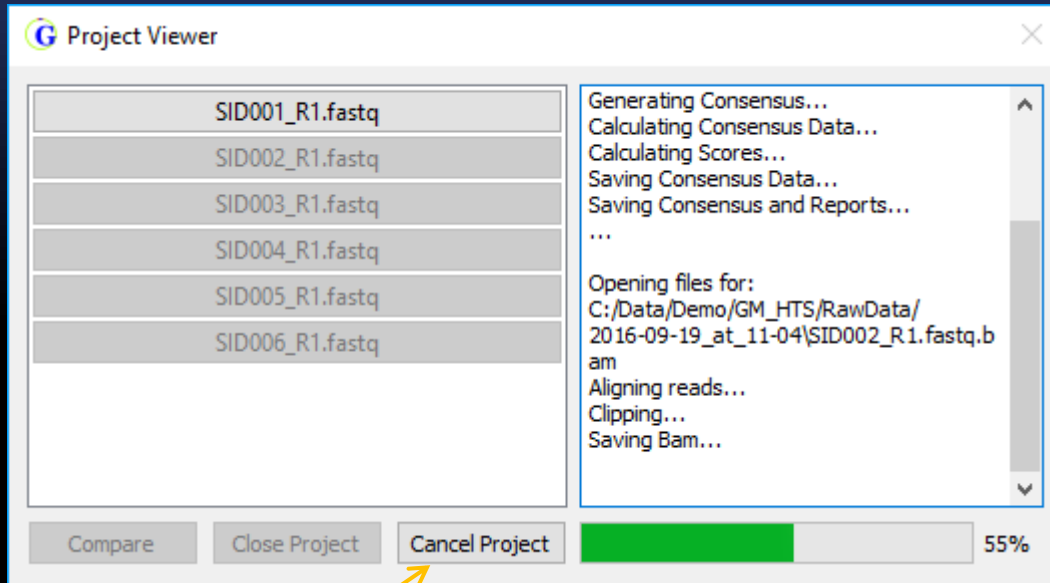
In the *Project Folder* field, a location can be selected for the data output by the program. A location can be set using the ellipsis button to the right of the field, or it can be typed manually. The folder will be created if it does not exist.



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

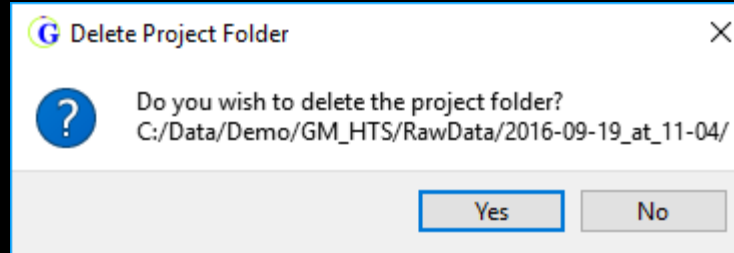
Sample Processing

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



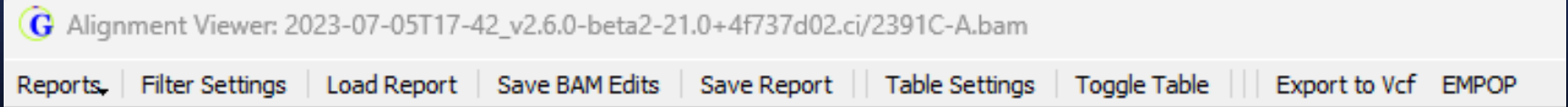
When a sample is finished it is possible to click on the button in the *Project Viewer* to open it - even before all samples finish processing

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

Alignment Viewer: 2023-07-05T17-42_v2.6.0-beta2-21.0+4f737d02.ci/2391C-A.bam

Reports | Filter Settings | Load Report

Global View - Forward read coverage in blue and reverse read coverage in red

Blocks
 Variants
 Zoom Global

Reference: G G A C A T A A T A A T A A C A A T T G A A T G T C T

Consensus: G G A C A T A A T A A T A A C A A T T G A A T G T C T

Motifs (green): [Green bar]

Amplicon (blue): [Blue bar]

Pileup

Toggle between different reports

2391C-A.bam ← Sample Name

Major Minor Both

Index	Position	Ref	Alt	IUPAC	Variants	Variants %	Coverage	A(#F;#R)	C(#F;#R)	G(#F;#R)	T(#F;#R)	Del(#F;#R)	Ins(#F;#R)	A%	C%
1	263	A	G	263G	A263G	99.95	13399	2;3	0;1	4387;9006	0;0	0;0	0;0	0.03	0.00
2	315	C	insC	315.1C	C315insC	93.32	5529	5;5529	0;0	0;0	0;1	9;17	3;5187	0.00	99.51
3	523	A	del	523del	A523del	99.84	13166	0;21	0;0	0;0	0;0	192;12953	0;0	0.15	0.00
4	524	C	del	524del	C524del	99.85	13166	0;0	0;19	0;0	0;0	192;12955	0;0	0.00	0.14

Result Table

Viewing the Results

Alignment Viewer: 2023-07-05T17-42_v2.6.0-beta2-21.0+4f737d02.ci/2391C-A.bam

Reports | Filter Settings | Load Report | Save BAM Edits | Save Report | Table Settings | Toggle Table | Export to Vcf | EMPOP

Blocks 12,000
 Variants 8,000
 Zoom Global 4,000

256 The green lines in the *Global View* represent the location of variants. 311

Reference
 Consensus

Motifs (green)
 Amplicon (blue)

Zoom in: draw a box from left to right
Zoom out: draw a box from right to left
Horizontal scroll: Right mouse click and drag
Vertical scroll: Scroll bar or mouse wheel

White area/space: in between reads
 Deletion: **Burnt Orange**
 Insertion: **Purple**
 A: **Green**
 C: **Blue**
 G: **Black**
 T: **Red**
 N: Gray text
 PHI region: Gray box

Greater-than and less-than symbols point to the 5' end of the read:
 ">" for forward reads
 "<" for reverse reads

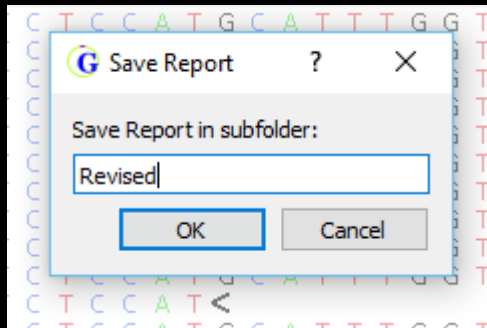
2391C-A.bam Major Minor Both

Index	Position	Ref	Alt	Del(#F;#R)	Ins(#F;#R)	A%	C%
1	263	A	G	0;0	0;0	0.03	0.00
2	315	C	insC	315.1C	C315insC	93.32	5561
3	523	A	del	523del	A523del	99.84	13166
4	524	C	del	524del	C524del	99.85	13166

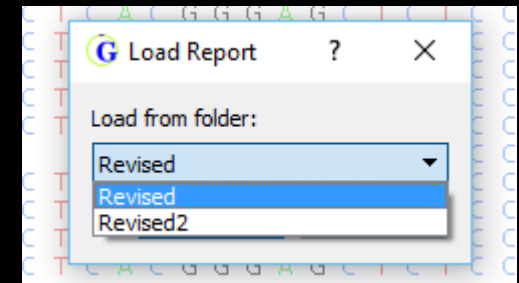
Adding and Removing Variants

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



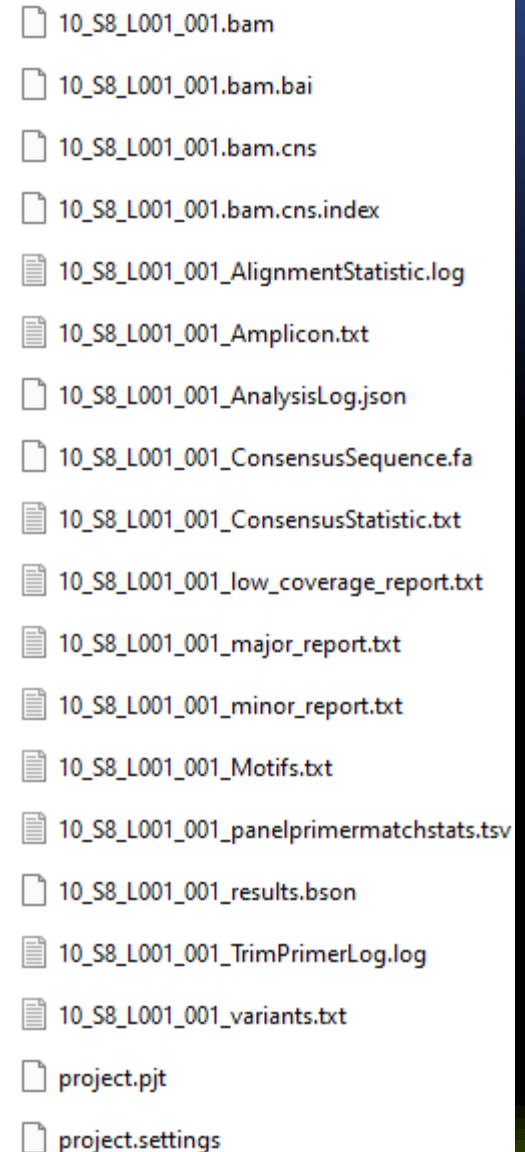
After changes to settings or manual edits, the report can be saved as a new report. The original project is opened by default, but all previously saved reports are maintained in subfolders with changelogs.



Output Files

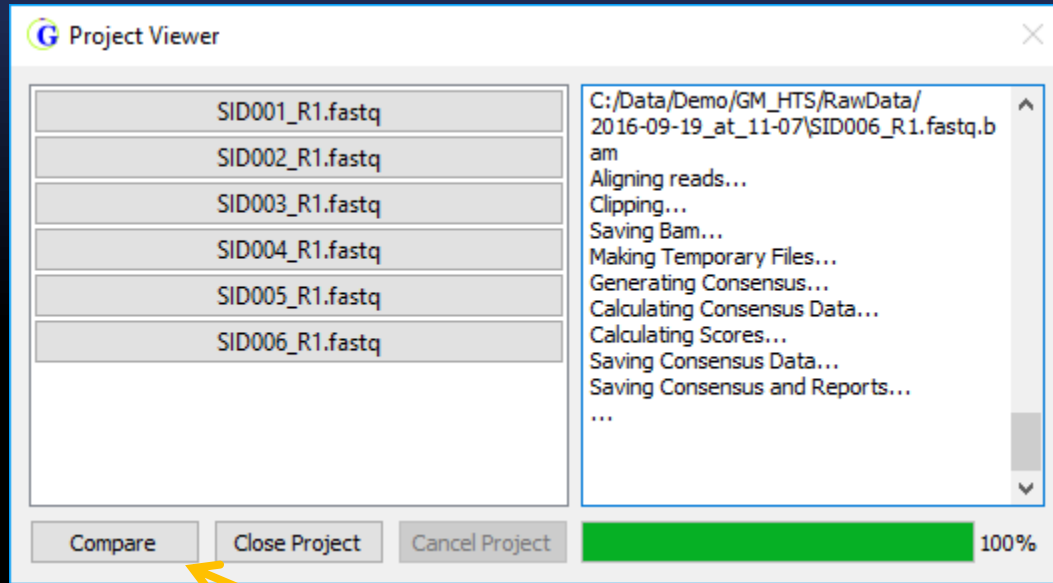
The program will output the following pieces of information for each sample in the project:

- **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.fasta
- 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.

The screenshot shows the 'Comparison Viewer' window. At the top, there are filter settings: 'Filter Settings', 'Major to Major', 'Major to Minor', and 'Minor to Minor'. Below this is the 'Sample to Sample Comparison' section, which includes a slider for 'Proportion of Shared Variants' set to 'None' and a button for 'All'. The main part of the window is a table with 7 rows and 7 columns, each representing a sample: SID001_R1.fa, SID002_R1.fa, SID003_R1.fa, SID004_R1.fa, SID005_R1.fa, and SID006_R1.fa. The diagonal cells (where row and column labels are the same) are black and contain '100%'. Other cells contain percentage values representing the similarity between samples. Below the table, there are controls for 'Row' and 'Column' selection, both set to 'Major'. There are also checkboxes for 'Show Percentage' and 'Use Jaccard Index', both of which are checked. Two yellow arrows point from text boxes below to the 'Major' radio buttons and the 'Show Percentage' checkbox.

	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!