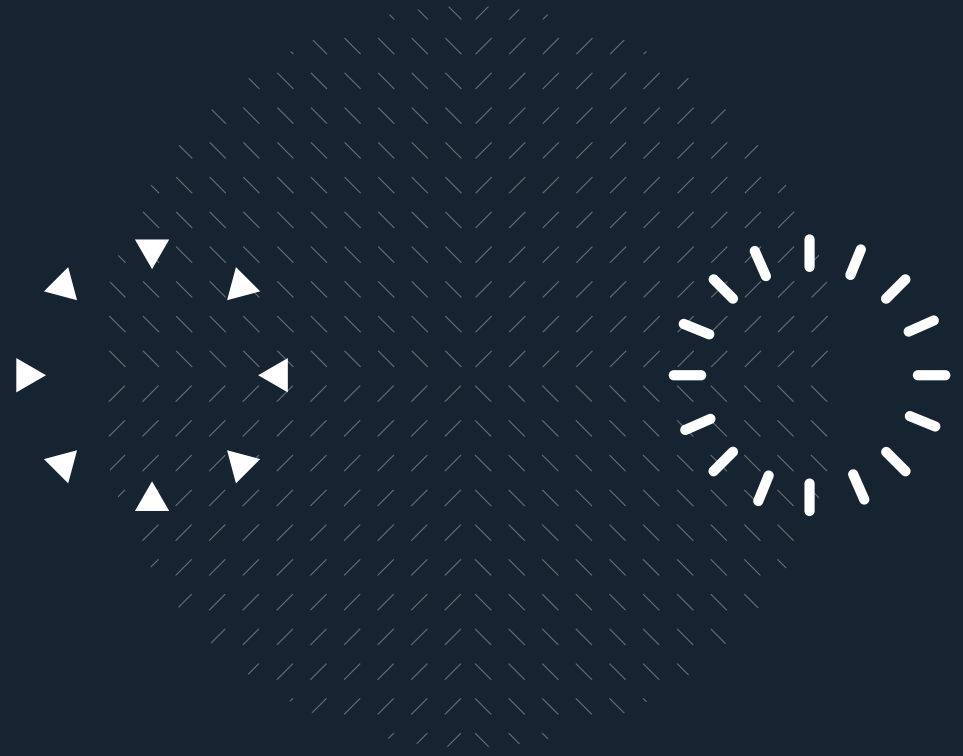
Dotmatics
SoftGenetics

# GeneMarker HTS

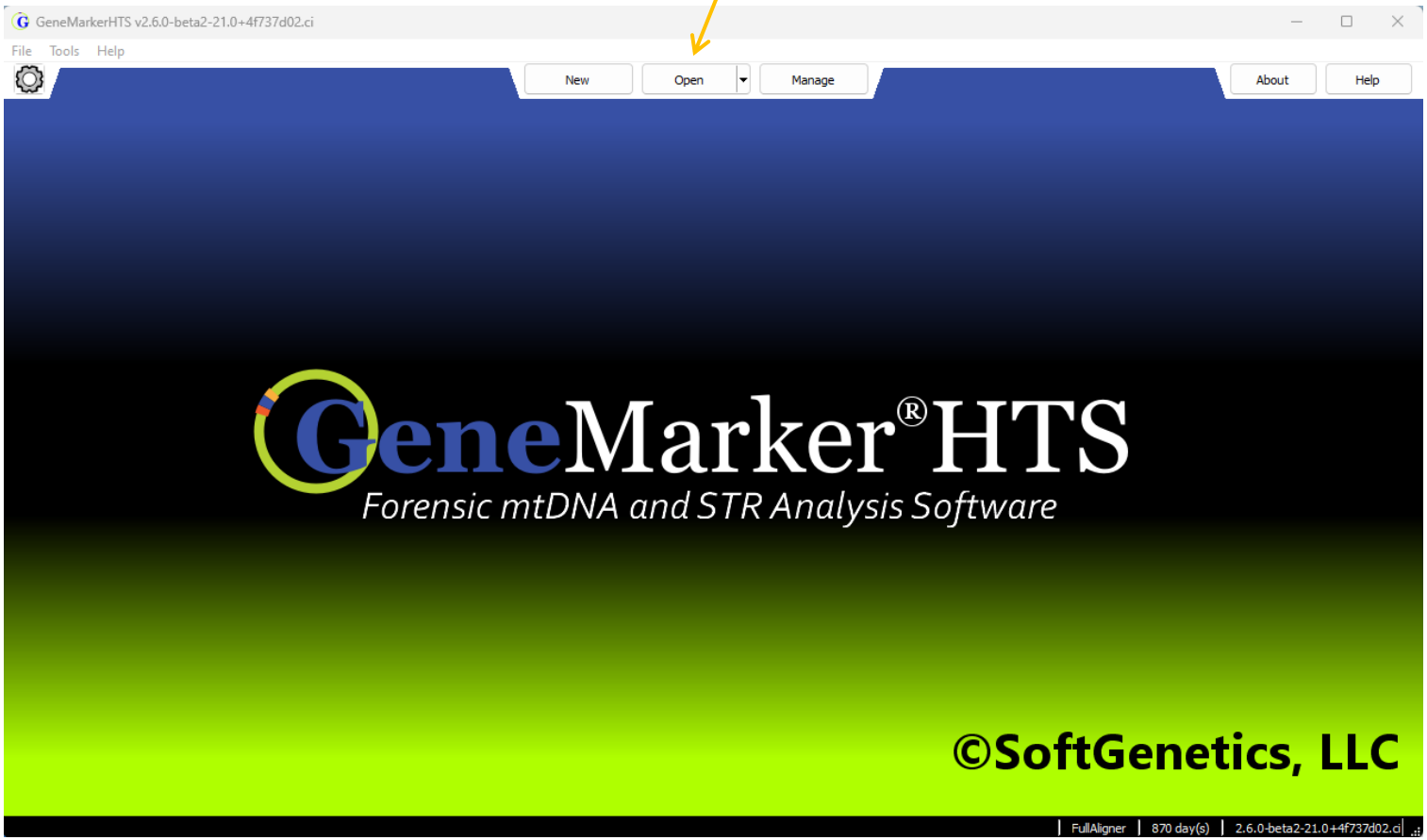
## Quick Start Guide - mtDNA



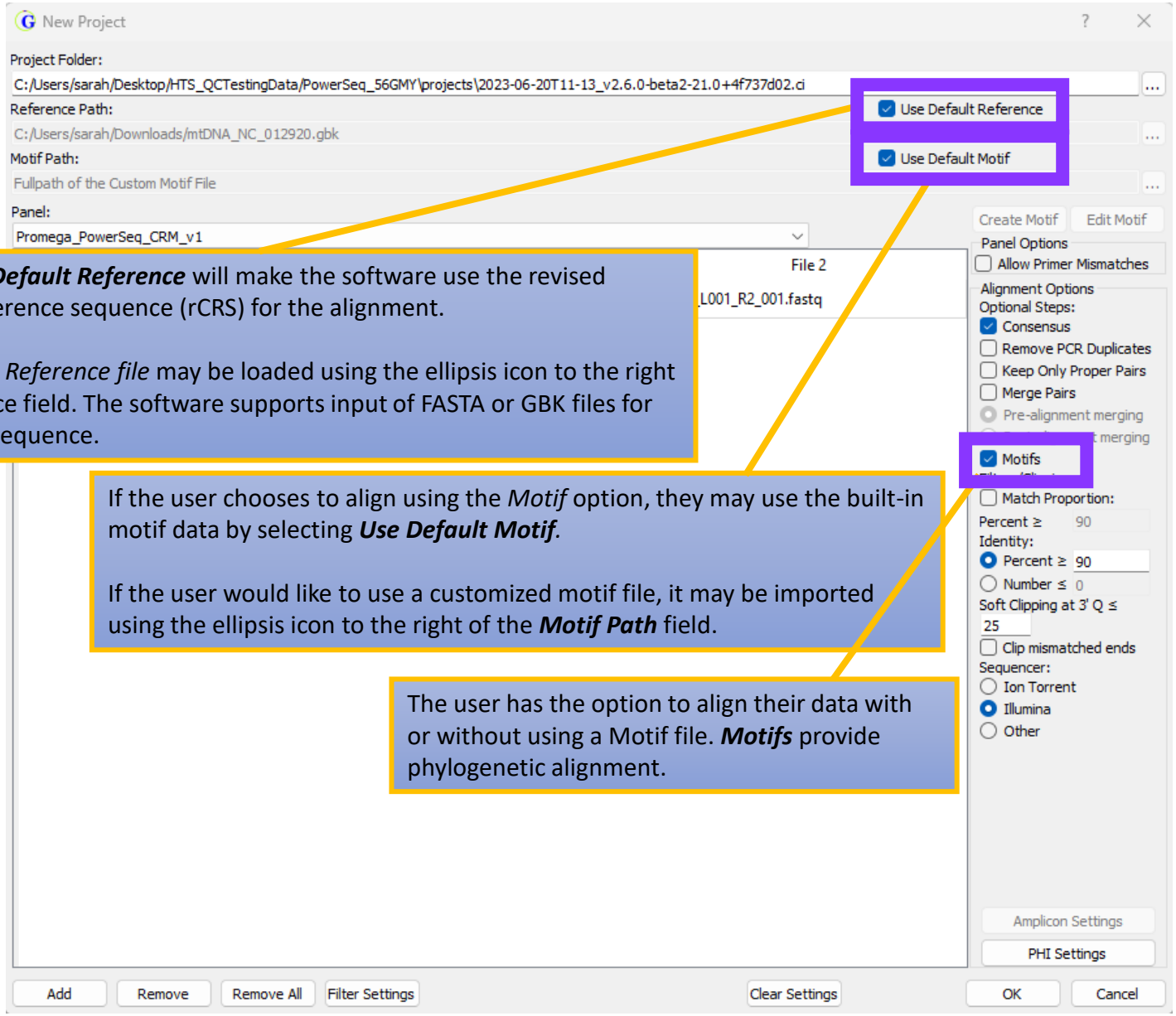
August 2023

# 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




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

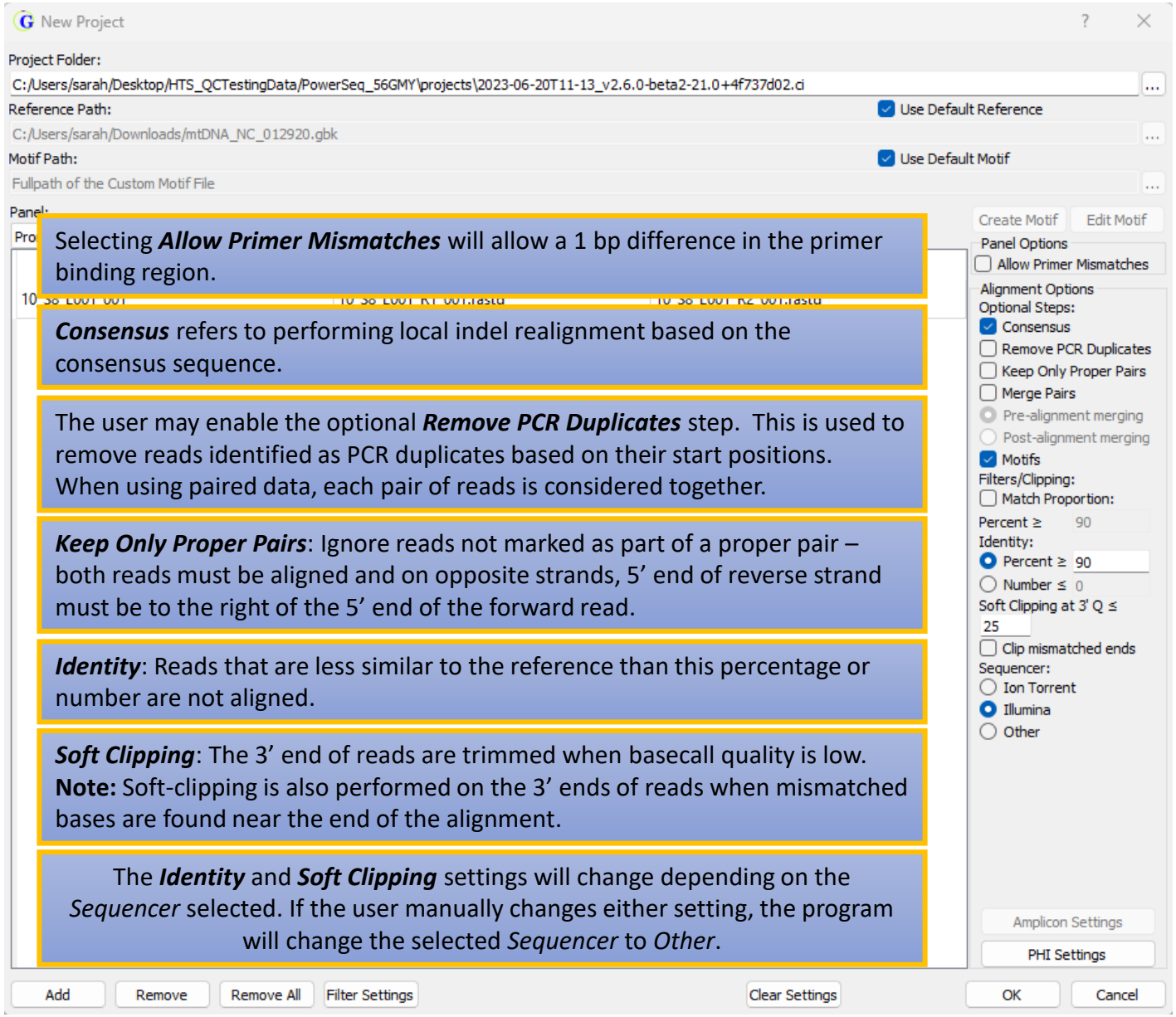
Alternatively, a *Reference file* may be loaded using the ellipsis icon to the right of the Reference field. The software supports input of FASTA or GBK files for the reference sequence.

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




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.

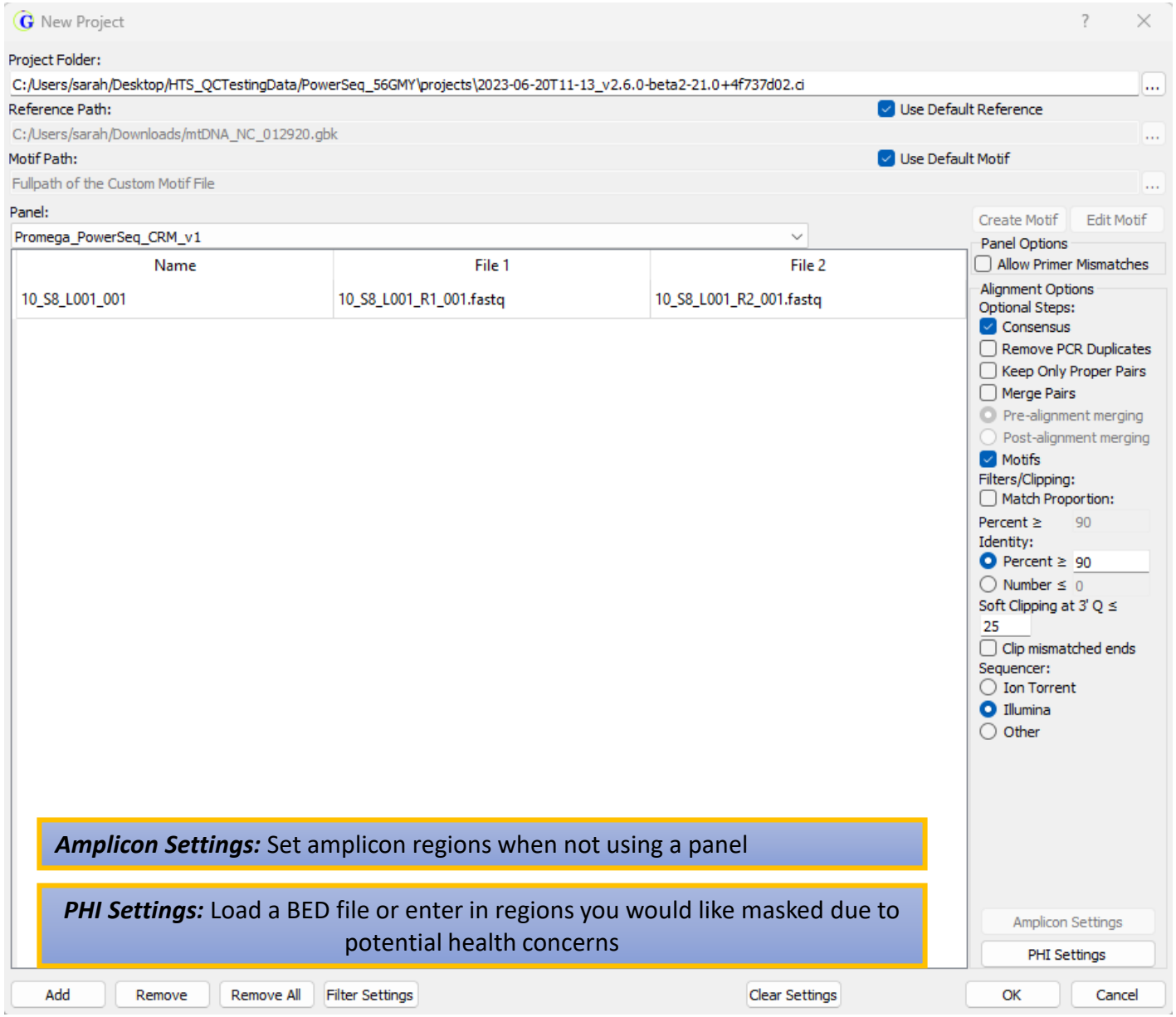
***Keep Only Proper Pairs***: Ignore reads not marked as part of a proper pair – both reads must be aligned and on opposite strands, 5' end of reverse strand must be to the right of the 5' end of the forward read.

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

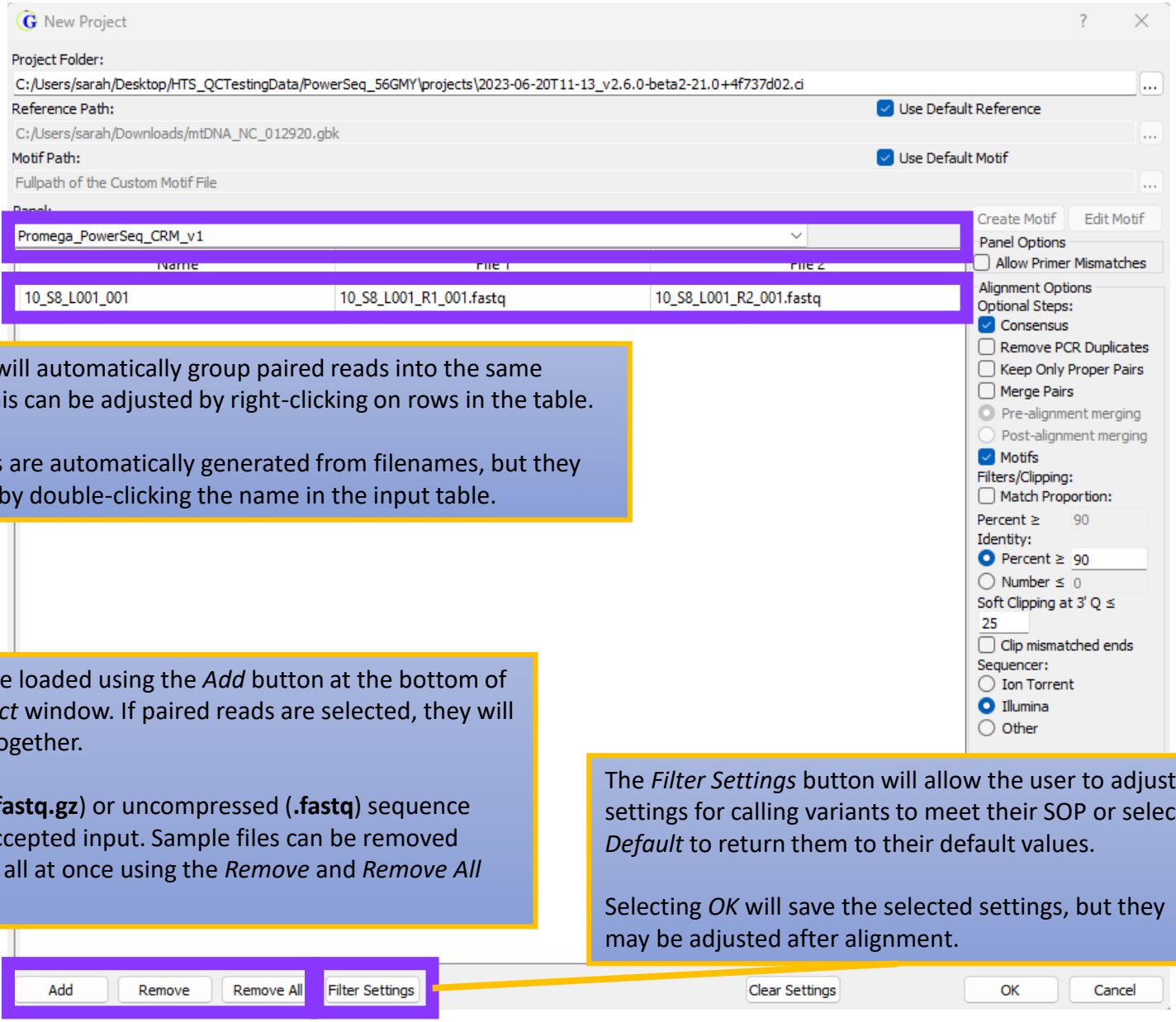
***Soft Clipping***: The 3' end of reads are trimmed when basecall quality is low. **Note:** Soft-clipping is also performed on the 3' ends of reads when mismatched bases are found near the end of the alignment.

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*.

# Creating a new project



# Creating a new project




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.

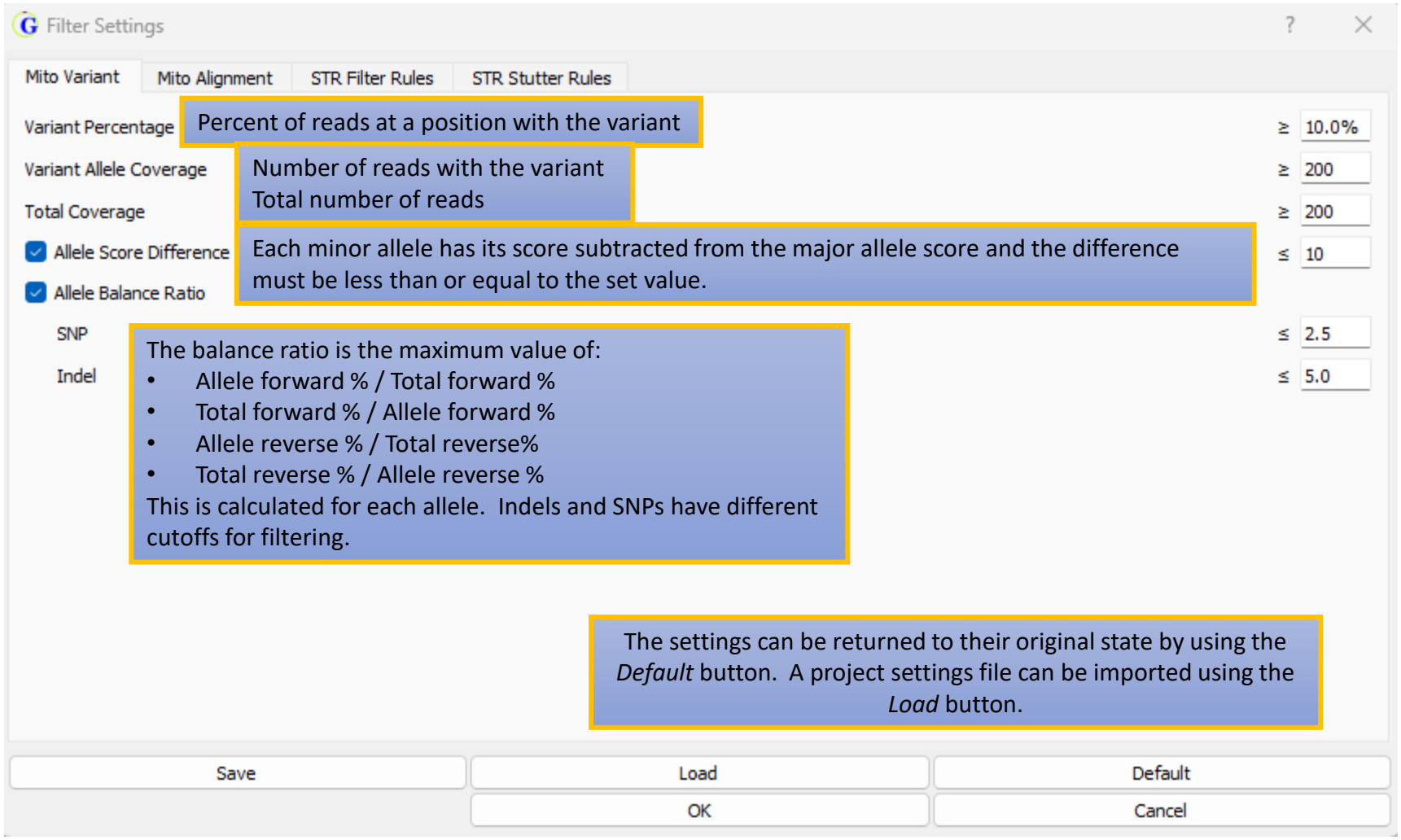
Compressed (**fastq.gz**) or uncompressed (**.fastq**) sequence files are the accepted input. Sample files can be removed individually or all at once using the *Remove* and *Remove All* buttons.

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.

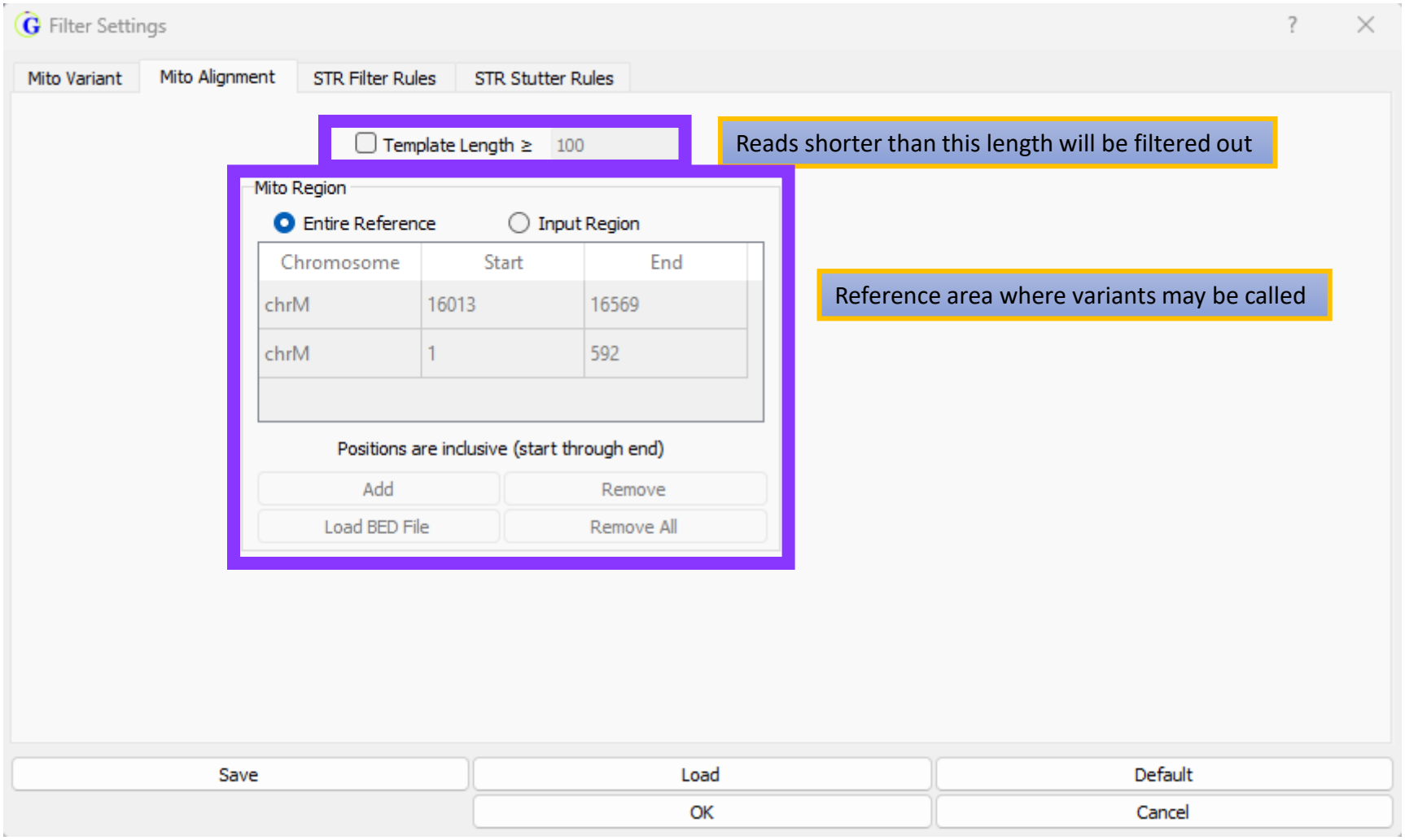
Selecting *OK* will save the selected settings, but they may be adjusted after alignment.

# Mito Variant Filter Settings

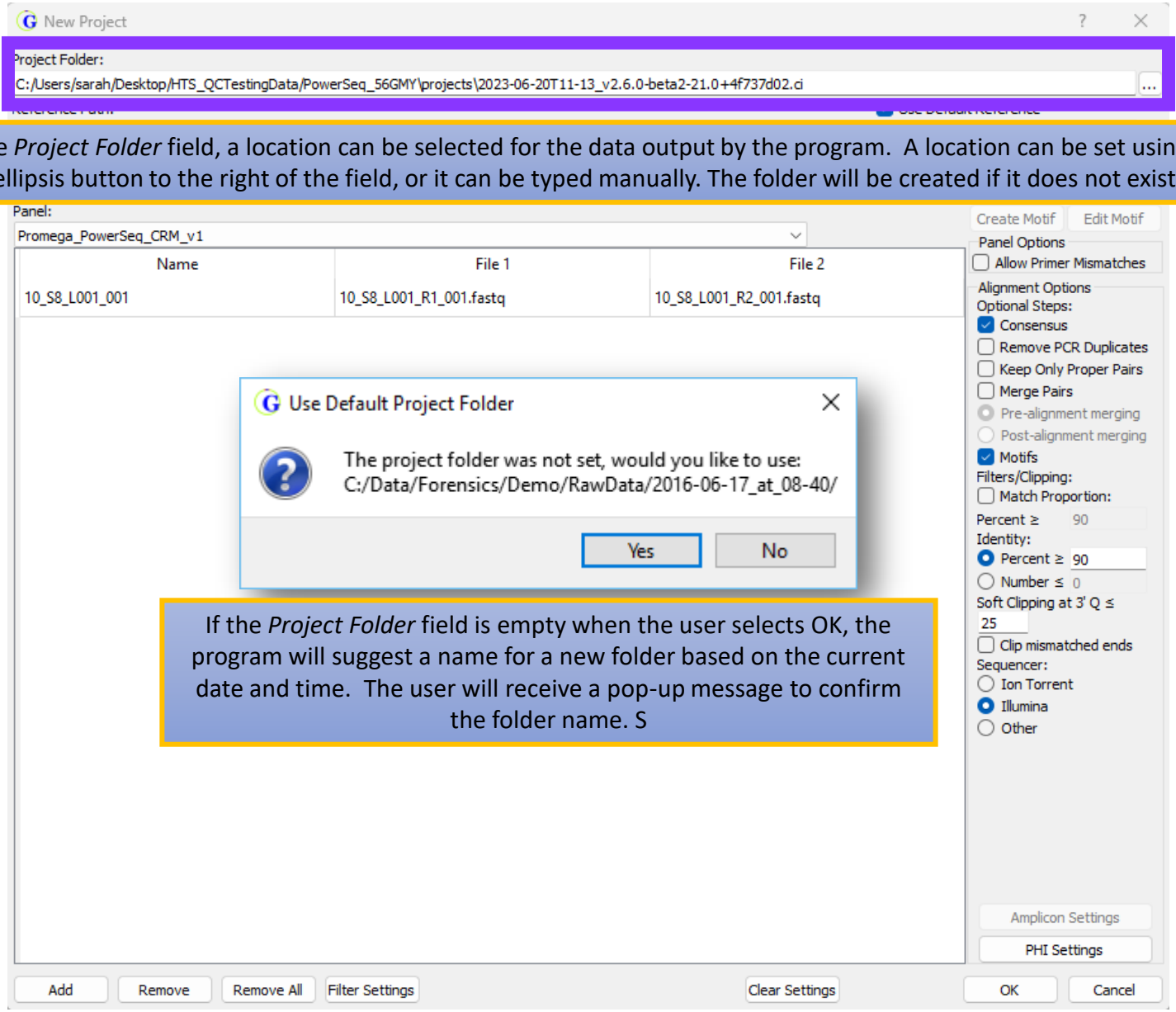
The *Filter Settings* dialog allows for variant calling settings to be adjusted.



# Mito Alignment Filter Settings



# Creating a new project

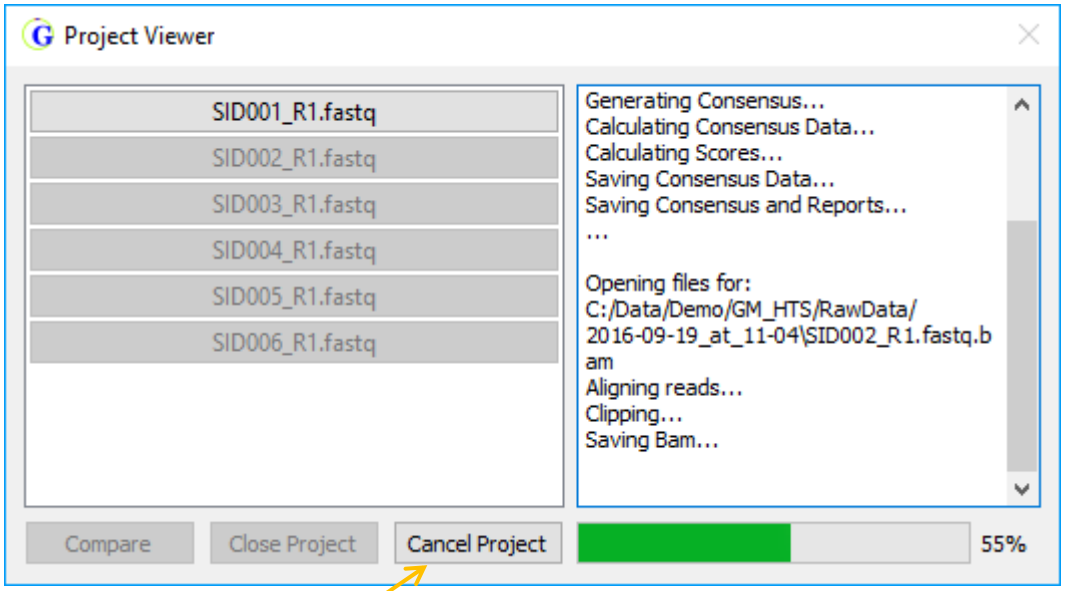



In the *Project Folder* field, a location can be selected for the data output by the program. A location can be set using the ellipsis button to the right of the field, or it can be typed manually. The folder will be created if it does not exist.

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

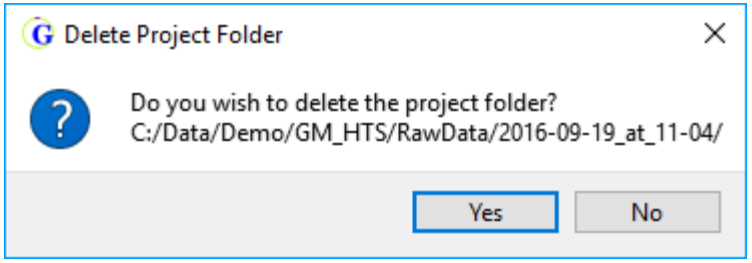
# Sample Processing

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



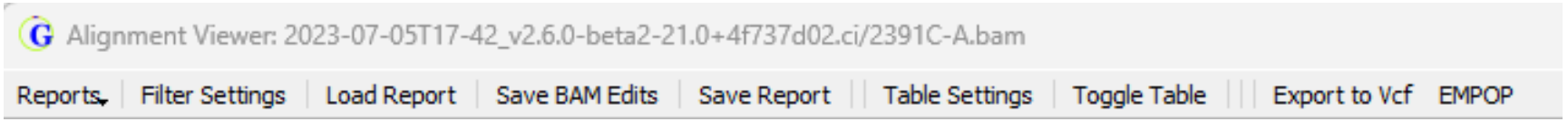

When a sample is finished it is possible to click on the button in the *Project Viewer* to open it - even before all samples finish processing

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




***Reports*** gives options for coverage Report, Template Lenth 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.

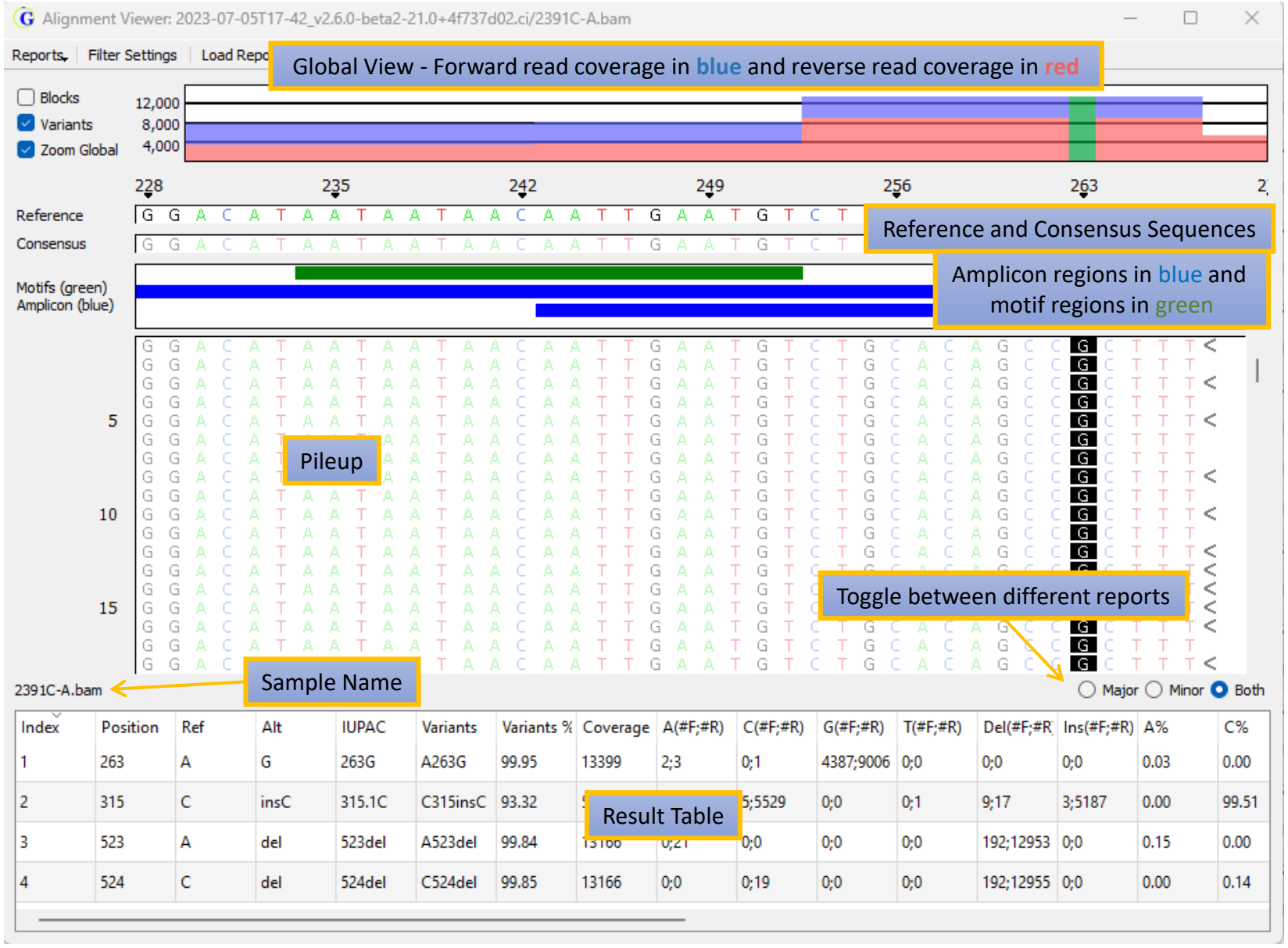
The ***Table Settings*** button will open the *Table Settings* window. Here the user can choose what information is displayed in the table.

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

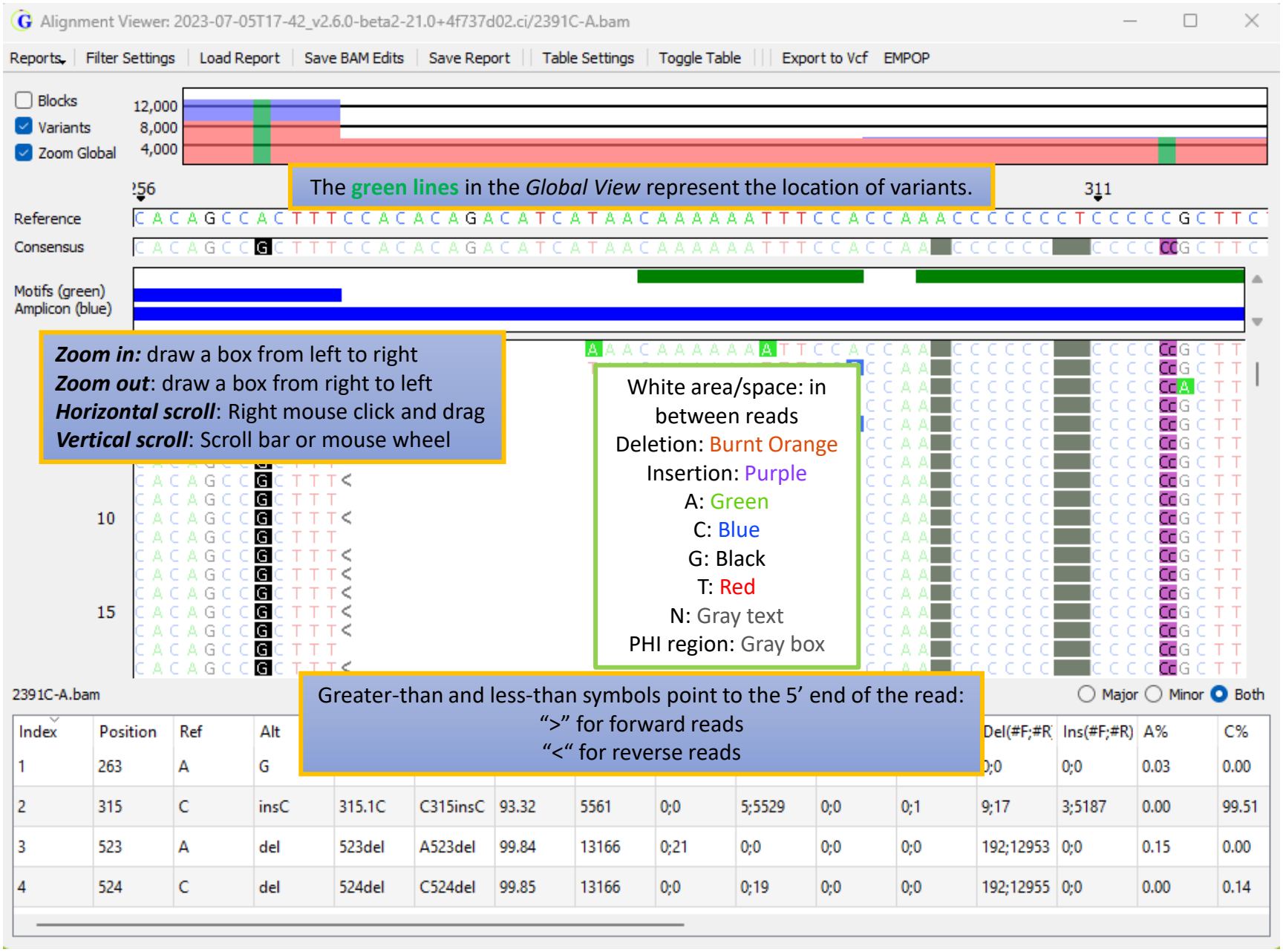
The ***EMPOP*** button streamlines the process of searching for profiles in the EMPOP database

# Viewing the Results




| 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 | | | 5;5529 | 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 |

# Viewing the Results

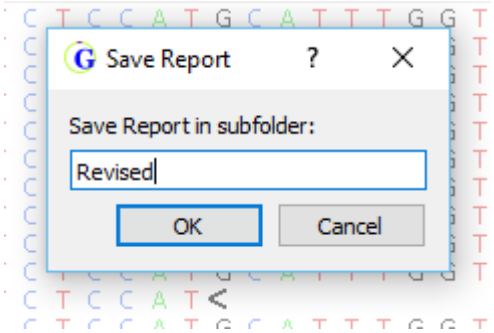



| Index | Position | Ref | Alt | | | | | | | | | Del(#F;#R | Ins(#F;#R) | A% | C% |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 263 | A | G | | | | | | | | | 0;0 | 0;0 | 0.03 | 0.00 |
| 2 | 315 | C | insC | 315.1C | C315insC | 93.32 | 5561 | 0;0 | 5;5529 | 0;0 | 0;1 | 9;17 | 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 |

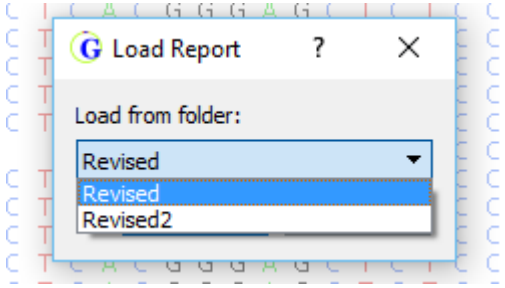
# Adding and Removing Variants

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 |



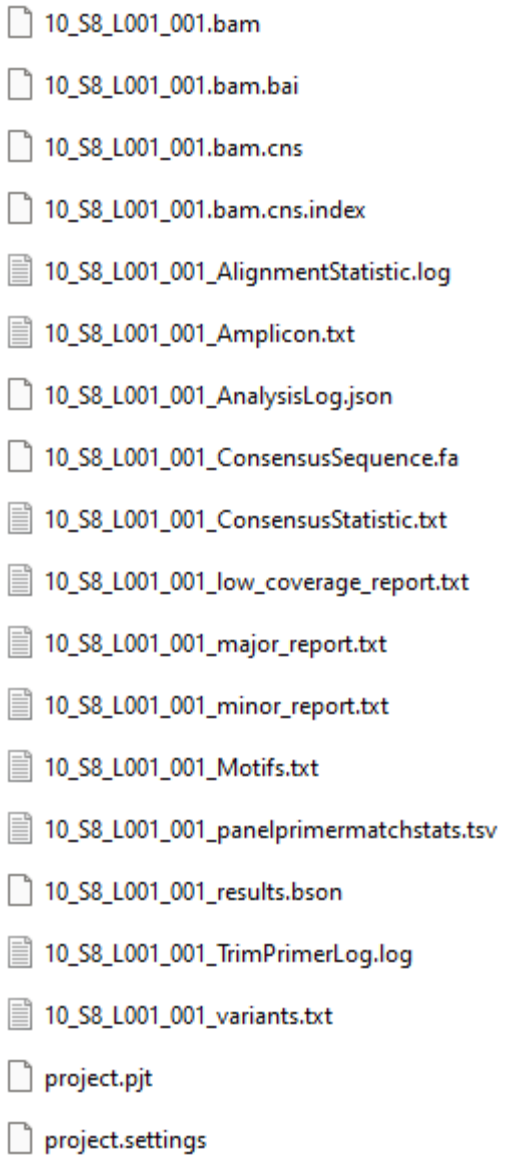

After changes to settings or manual edits, the report can be saved as a new report. The original project is opened by default, but all previously saved reports are maintained in subfolders with changelogs.



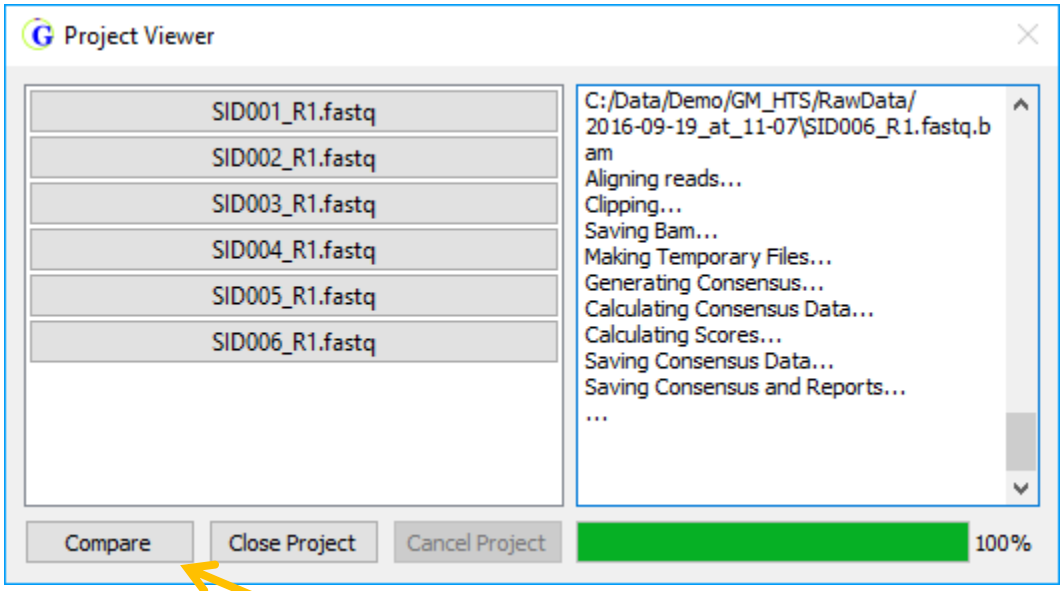
# Output Files

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



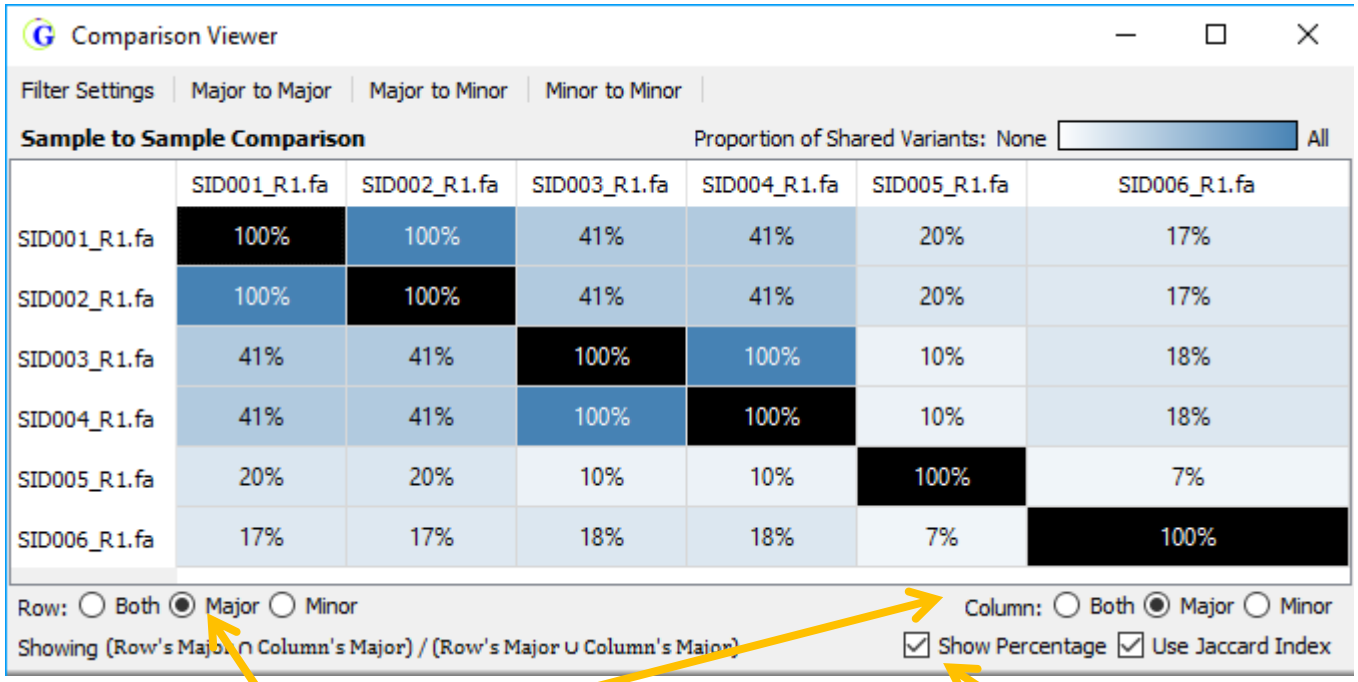
# Comparison Tool




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.

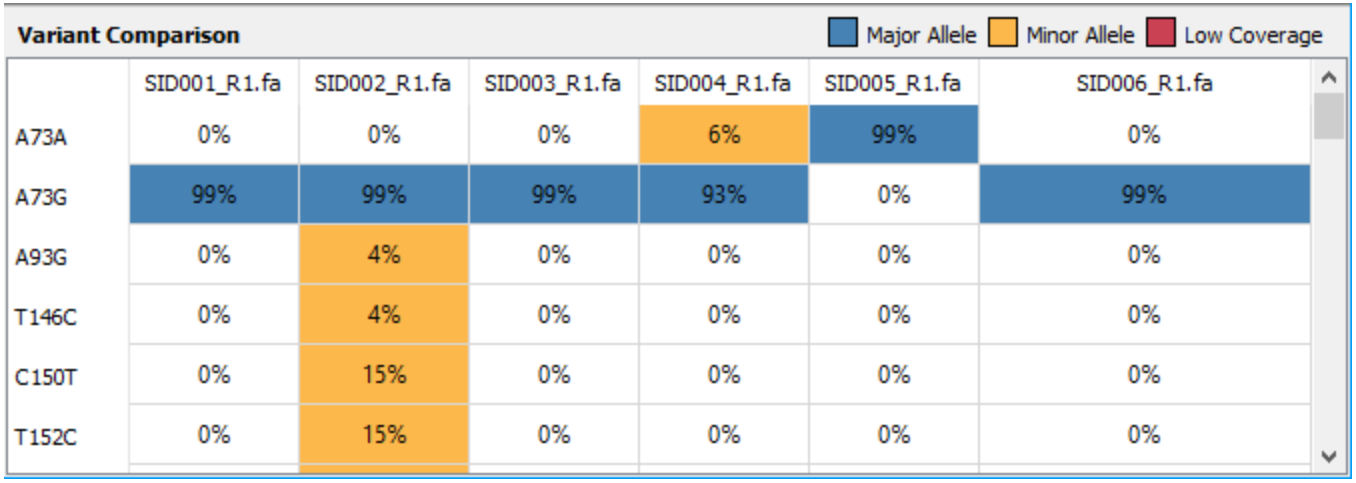



Comparisons can be made between major alleles, minor alleles, or both on each axis (row and column)

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 lease 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.




| | 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!**