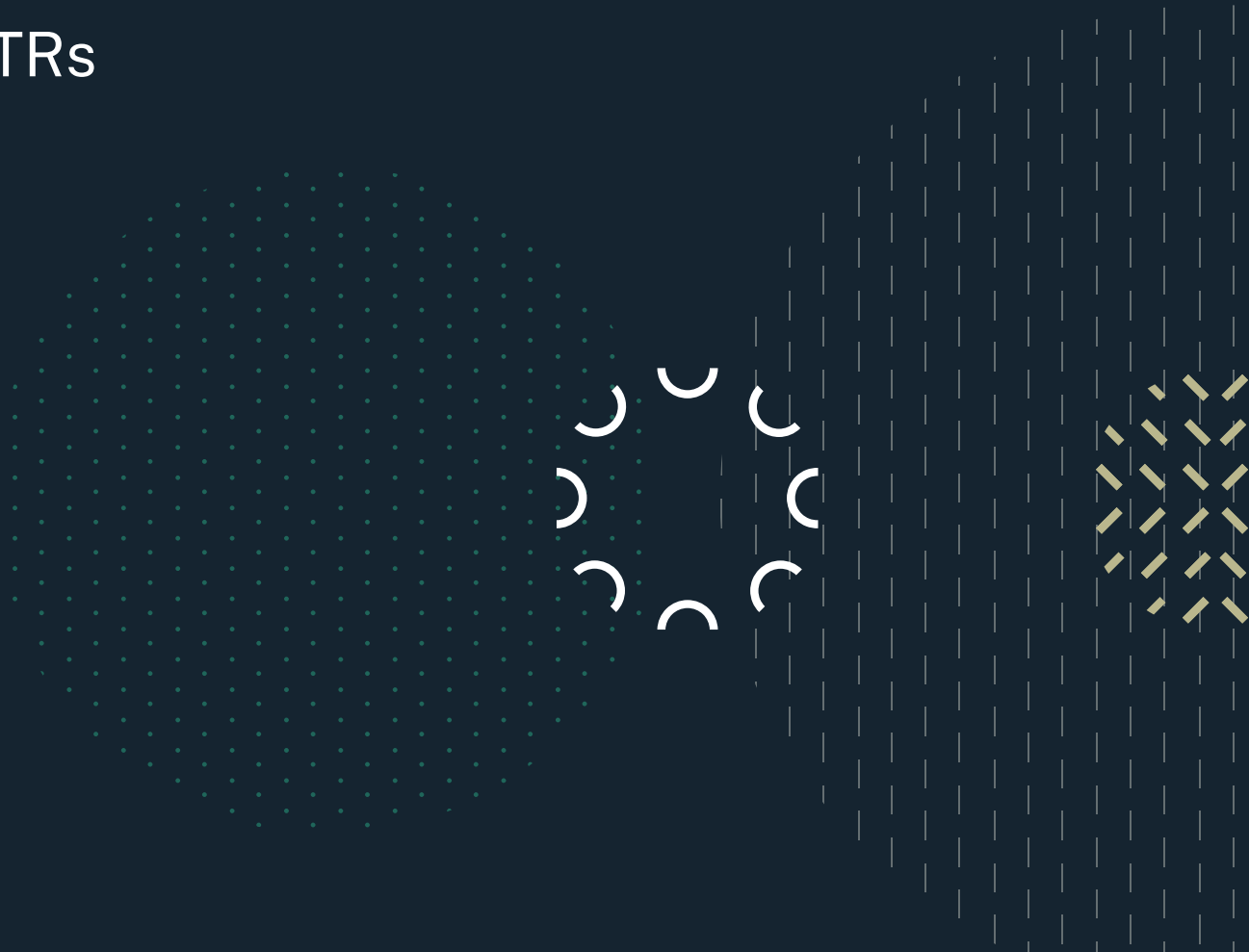
Dotmatics
SoftGenetics

GeneMarker HTS

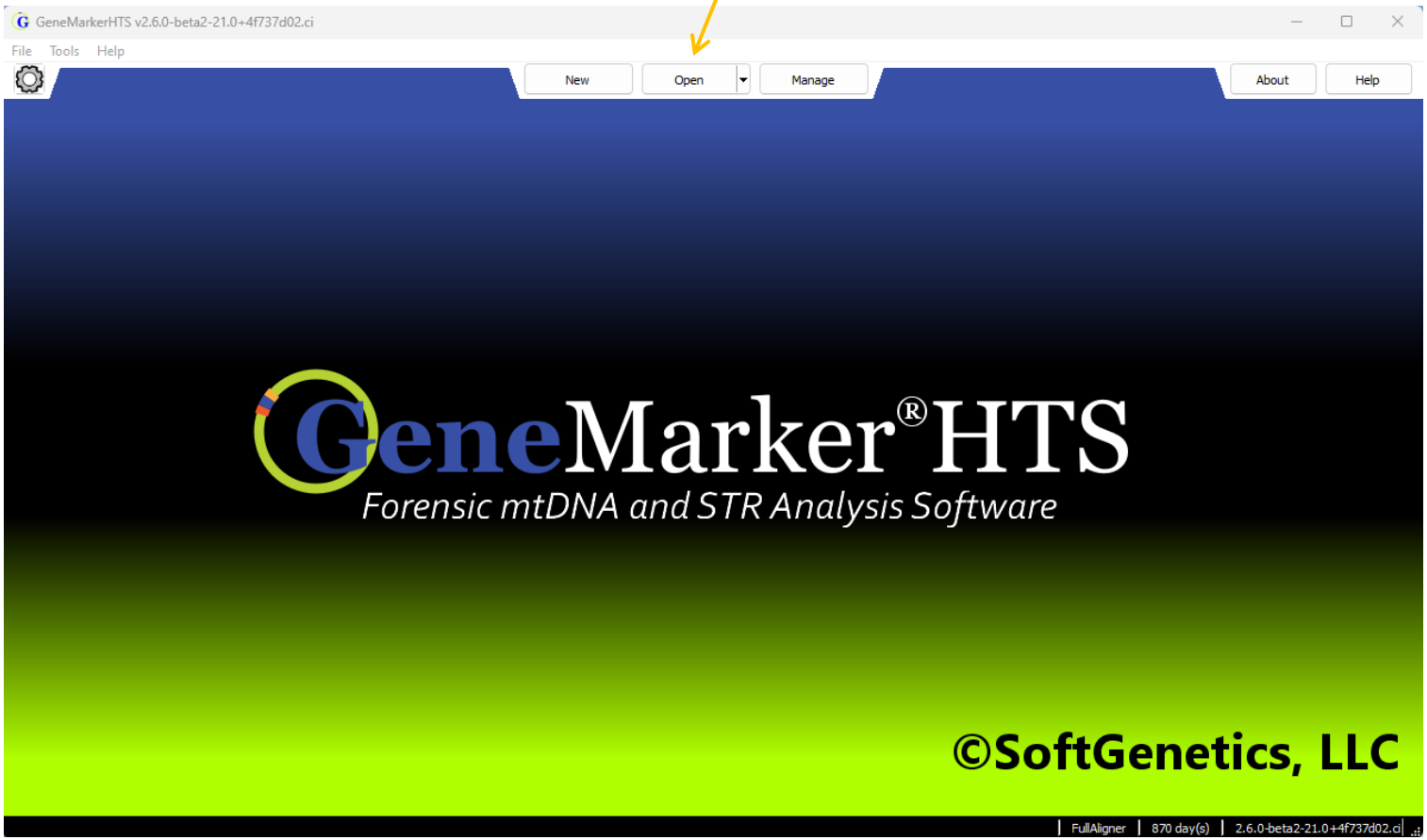
Quick Start Guide - STRs



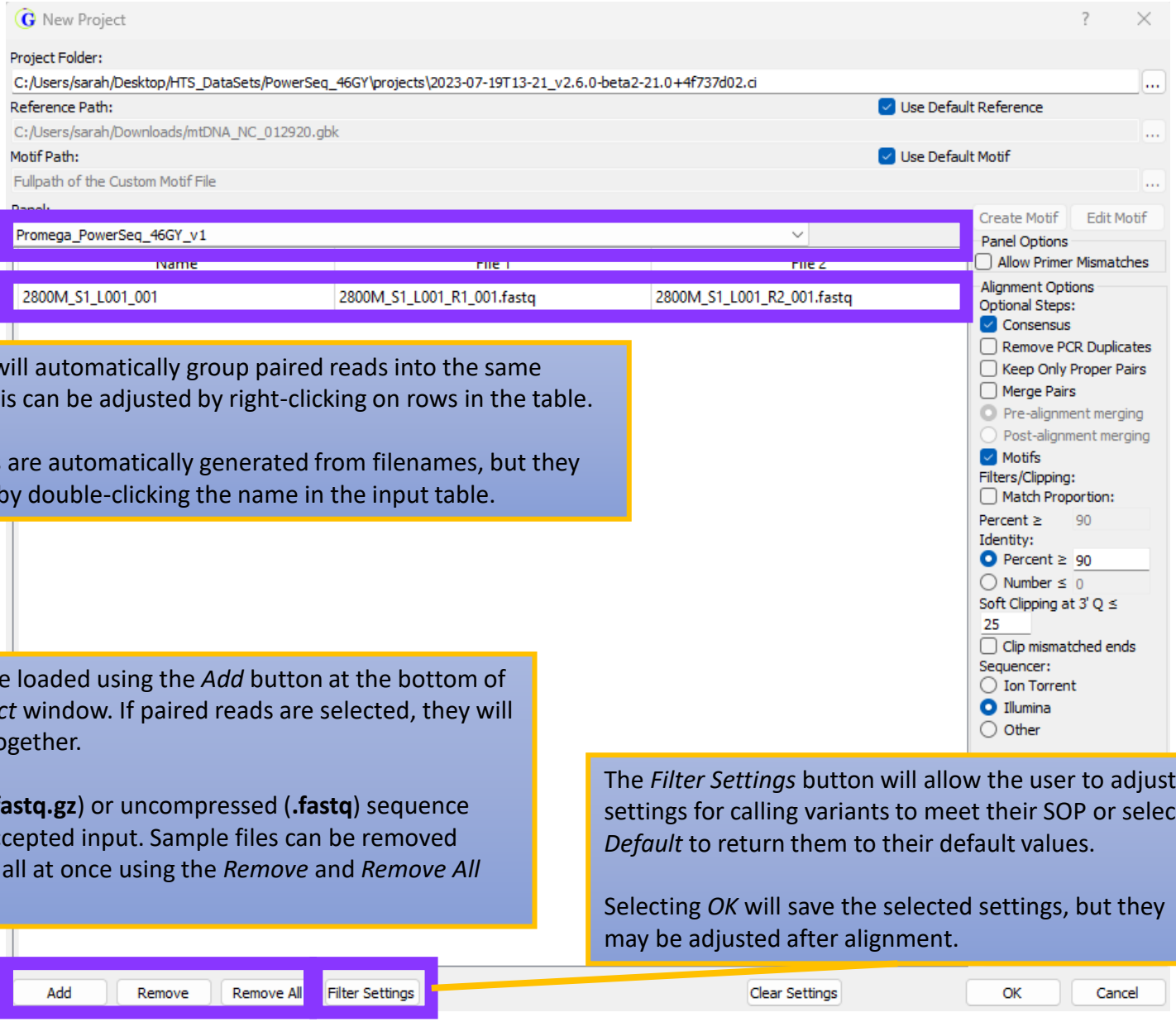
August 2023

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 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.

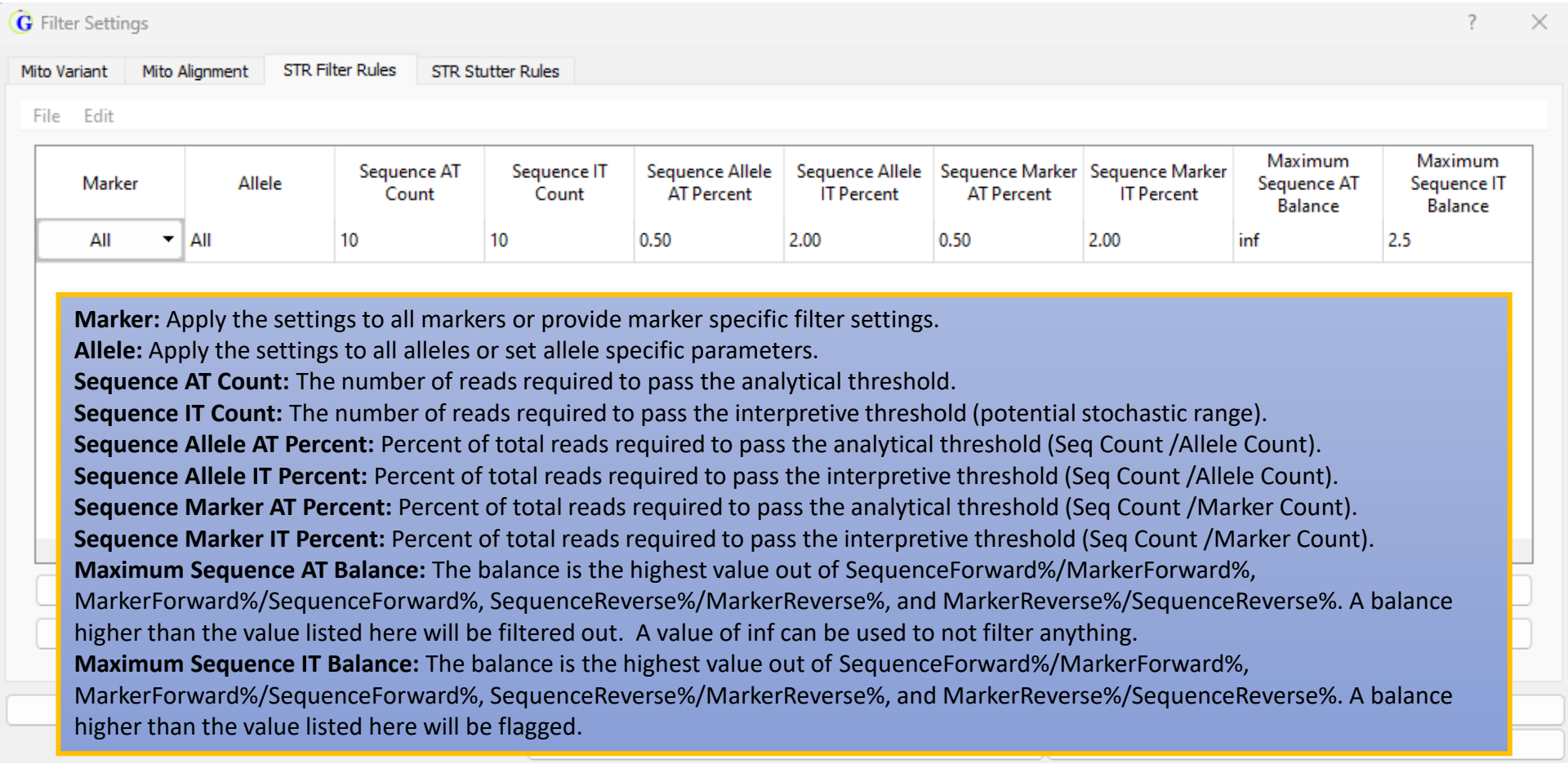
Compressed (**fastq.gz**) or uncompressed (**.fastq**) sequence files are the accepted input. Sample files can be removed individually or all at once using the *Remove* and *Remove All* buttons.

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.

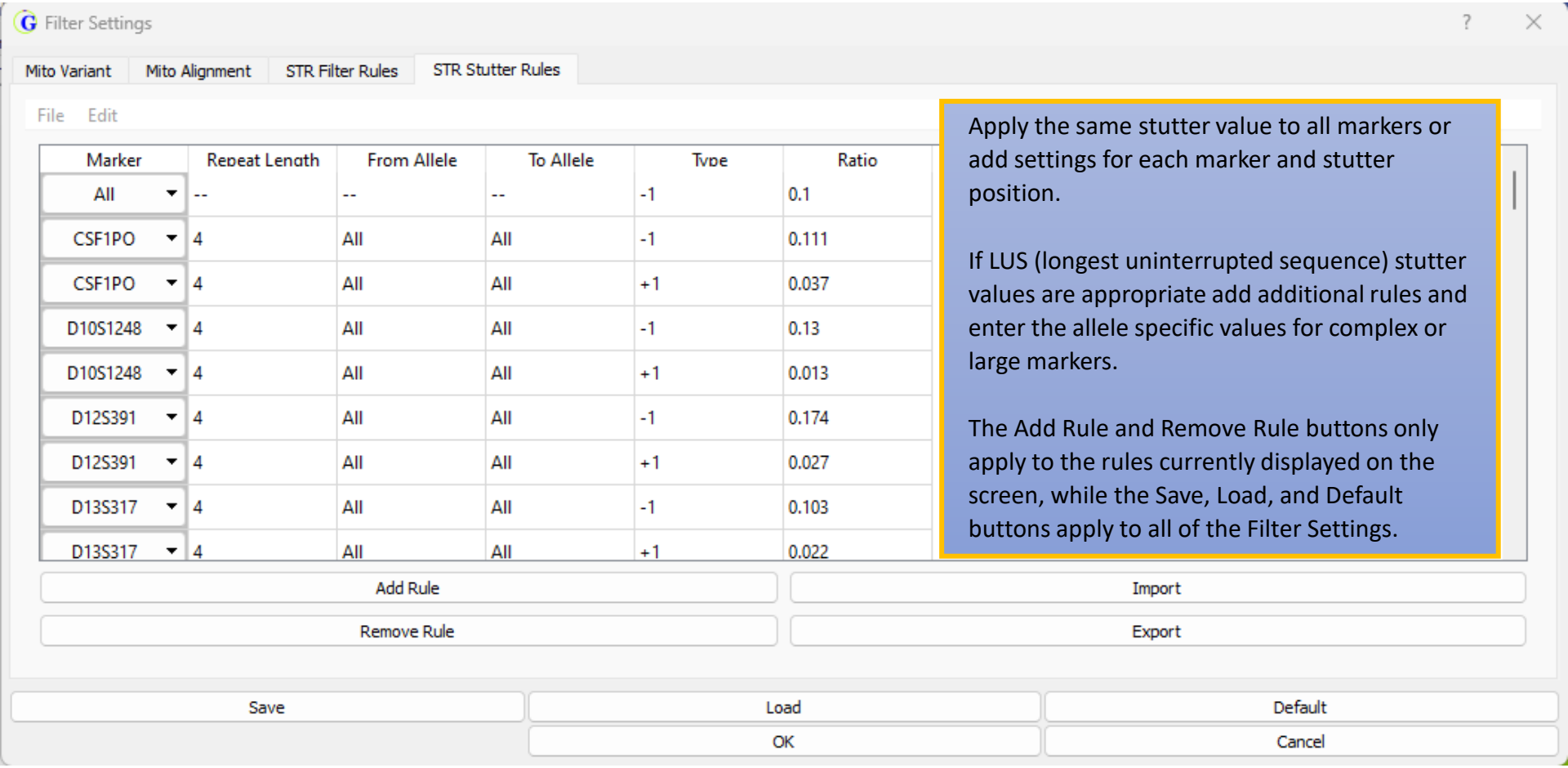
STR Filter Settings

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



Marker: Apply the settings to all markers or provide marker specific filter settings.
Allele: Apply the settings to all alleles or set allele specific parameters.
Sequence AT Count: The number of reads required to pass the analytical threshold.
Sequence IT Count: The number of reads required to pass the interpretive threshold (potential stochastic range).
Sequence Allele AT Percent: Percent of total reads required to pass the analytical threshold (Seq Count /Allele Count).
Sequence Allele IT Percent: Percent of total reads required to pass the interpretive threshold (Seq Count /Allele Count).
Sequence Marker AT Percent: Percent of total reads required to pass the analytical threshold (Seq Count /Marker Count).
Sequence Marker IT Percent: Percent of total reads required to pass the interpretive threshold (Seq Count /Marker Count).
Maximum Sequence AT Balance: The balance is the highest value out of SequenceForward%/MarkerForward%, MarkerForward%/SequenceForward%, SequenceReverse%/MarkerReverse%, and MarkerReverse%/SequenceReverse%. A balance higher than the value listed here will be filtered out. A value of inf can be used to not filter anything.
Maximum Sequence IT Balance: The balance is the highest value out of SequenceForward%/MarkerForward%, MarkerForward%/SequenceForward%, SequenceReverse%/MarkerReverse%, and MarkerReverse%/SequenceReverse%. A balance higher than the value listed here will be flagged.

STR Stutter Rules

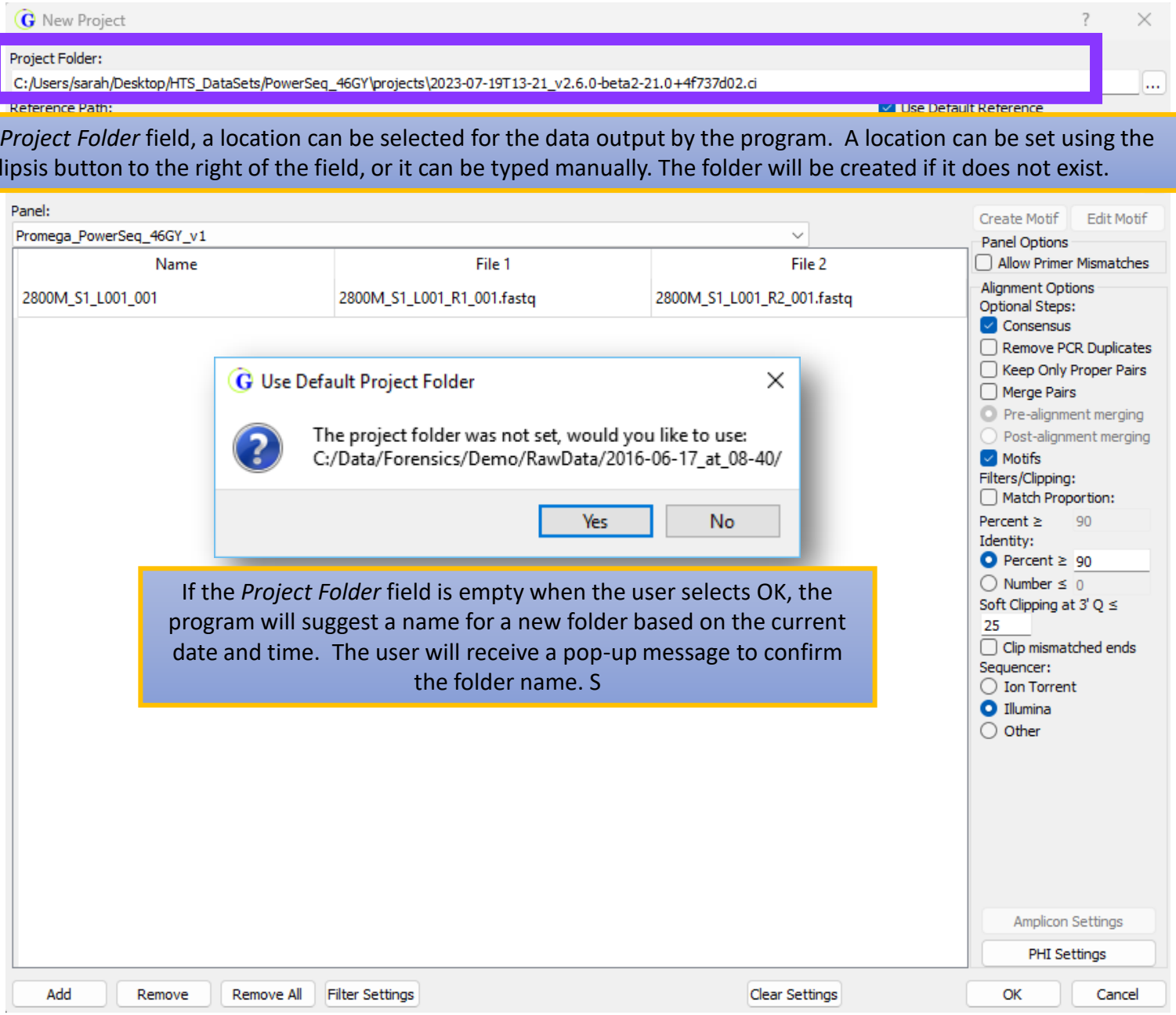


Apply the same stutter value to all markers or add settings for each marker and stutter position.

If LUS (longest uninterrupted sequence) stutter values are appropriate add additional rules and enter the allele specific values for complex or large markers.

The Add Rule and Remove Rule buttons only apply to the rules currently displayed on the screen, while the Save, Load, and Default buttons apply to all of the Filter Settings.

Creating a new project

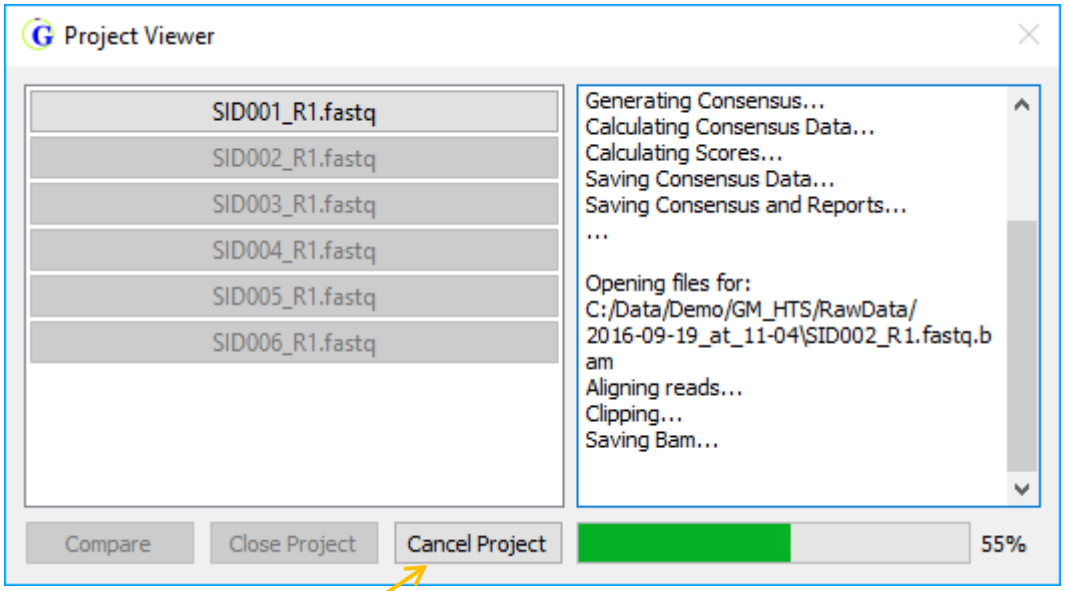


In the *Project Folder* field, a location can be selected for the data output by the program. A location can be set using the ellipsis button to the right of the field, or it can be typed manually. The folder will be created if it does not exist.

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

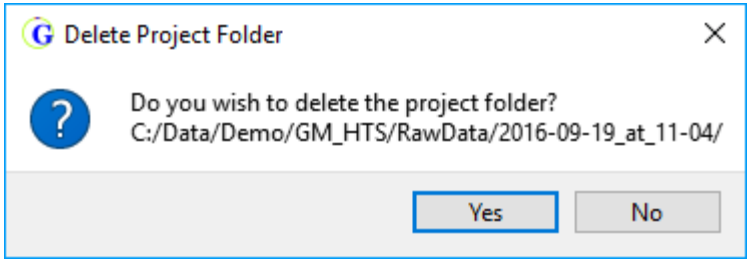
Sample Processing

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



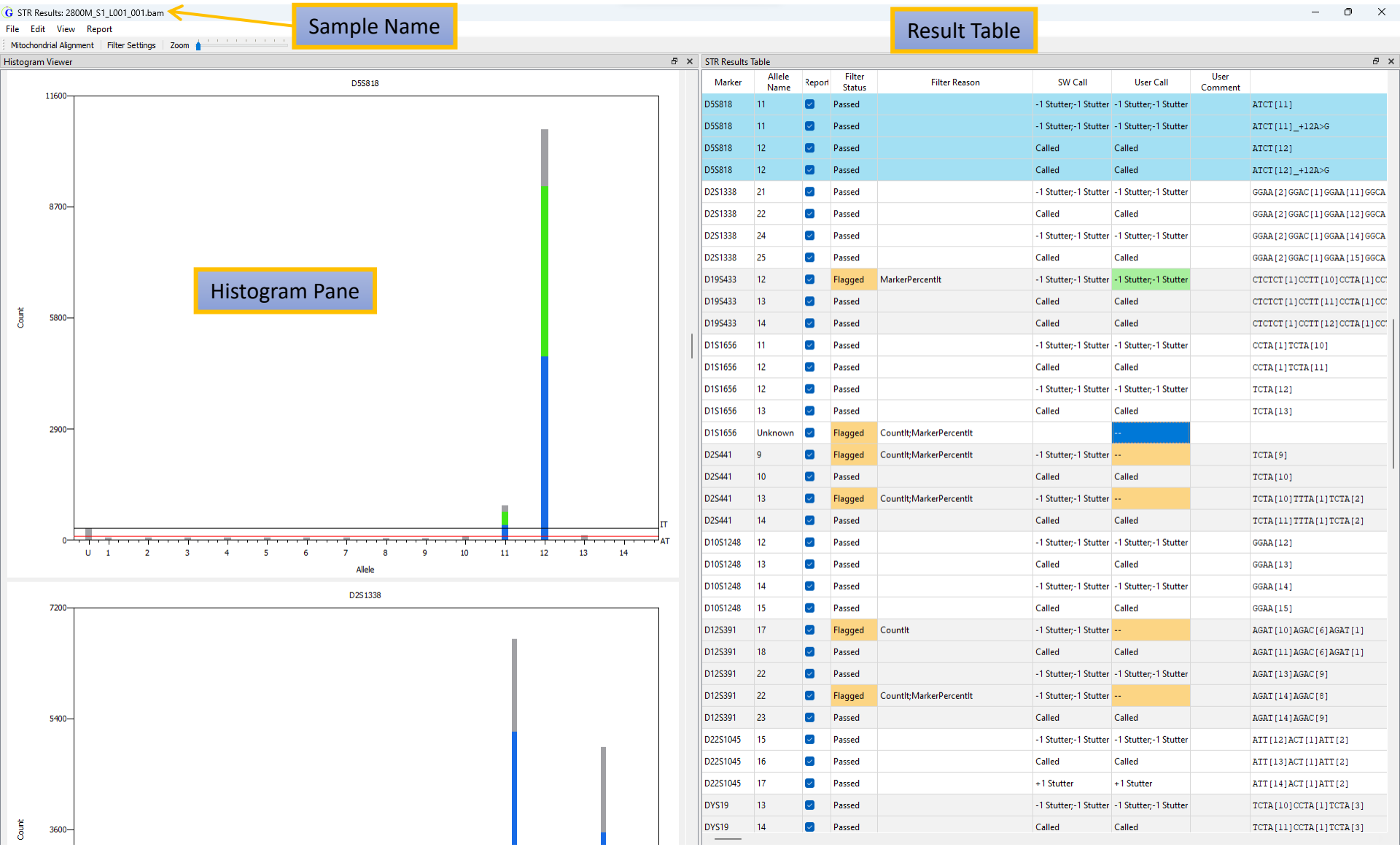
When a sample is finished it is possible to click on the button in the *Project Viewer* to open it - even before all samples finish processing

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

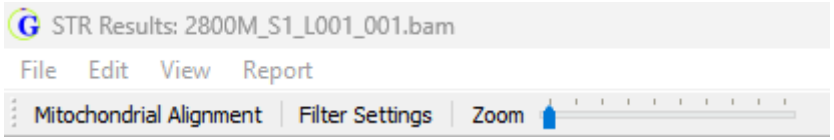


If the project is cancelled, the program will ask the user if they would like to delete the project folder that was created.

Viewing the Results

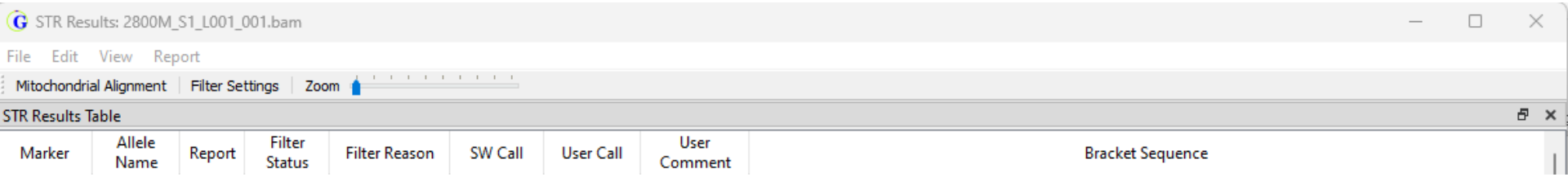


Viewing the Results



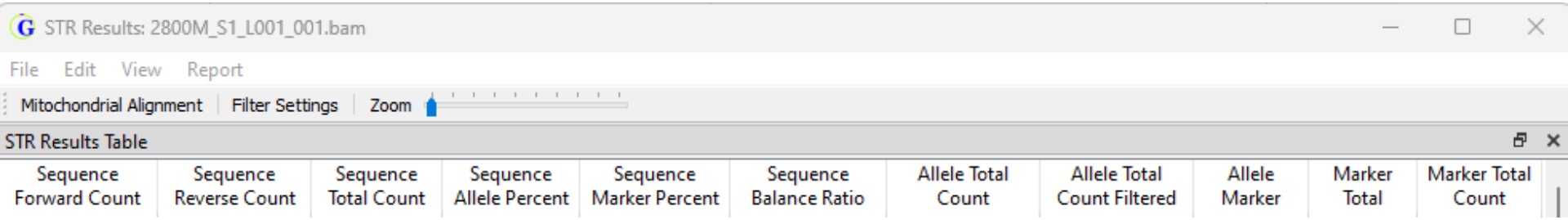
File: Options to export histograms and results table
Edit: Opens the STR Filter Rules and Stutter Rules Settings
View: Contains options and hot key shortcuts for viewing the table and histograms
Report: Contains options for the NGS STR Allele Report and the CE STR Allele Report
Mitochondrial Alignment: A direct link to open the Mitochondrial Alignment Viewer for projects that include mtDNA data
Filter Settings: A direct link to the STR Filter Settings
Zoom: A slide bar for displaying histograms

Viewing the Results



Field	Description
Marker	The autosomal STR or Y-STR locus (marker) name.
Allele Name	The corresponding capillary electrophoresis allele name.
Report	Check box giving the option to include the sequence in the CE or NGS reports
Filter Status	Passed status indicates the read passed STR and Stutter Filters. Flagged status indicates the read fired one or more of the STR and Stutter rules.
Filter Reason	The rule(s) fired if a read was flagged. If the read passed, was not flagged, the Filter Reason will be >IT, counts are above the interpretive threshold.
SW Call	The software call based on the analysis parameters. Called = met all parameters, -1 Stutter (and any other stutter positions) = Sequence Total Count of potential stutter peak/Allele Total Count of true peak = peak height ratio of potential stutter peak to true allele. If value is below stutter filter settings, then it is called stutter, if it is above filter settings then it is a true peak.
User Call	Flagged calls will have orange background in this cell (for example, the counts are AT< x <IT) Double click on the cell to enter the analyst's decision to call or not call the allele.
User Comment	Double click to comment on the User Call. Type the comment in the field onthe right and press Save Current to save the comment in the list on the left, press OK to apply the selected comment to the User Comment field.
Bracket Sequence	The STR repeat sequence is displayed in brackets

Viewing the Results



Sequence Forward Counts	**The number of forward reads with this sequence.**
Sequence Reverse Counts	**The number of reverse reads with this sequence.**
Sequence Total Counts	**Total number of reads with this sequence.**
Sequence Allele Percent	**The percent of the reads of that sequence for that allele (Sequence Total Count/Allele Total Count)**
Sequence Marker Percent	**The percent of the reads of that sequence for that marker (Sequence Total Count/Marker Total Count)**
Sequence Balance Ratio	**The highest value between: SequenceForward%/MarkerForward%, MarkerForward%/SequenceForward%, SequenceReverse%/MarkerReverse%, and MarkerReverse%/SequenceReverse%**
Allele Total Counts	**All reads for this allele (including filtered reads and sequence variants having the same CE allele name).**
Allele Total Counts Filtered	**The number of reads filtered for that allele.**
Allele Marker Percent	**The ratio of this allele to all alleles for the marker.**
Marker Total Count	**All read counts for the marker.**
Marker Total Count Filtered	**The number of reads for the marker that were filtered.**

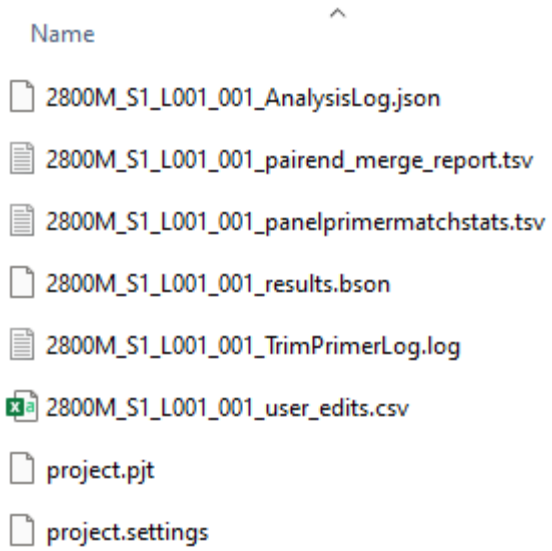
Viewing the Results

Sequence	**The sequence for the reads with flanking sequence in lower case and repeat sequence in upper case.**
Left Flank	**The sequence for the left flank of the reads.**
Repeat	**The sequence for the repeat portion of the reads (not in bracket format).**
Right Flank	**The sequence for the right flank of the reads.**

Output Files

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

- **AnalysisLog.json**: stats about alignment in an easy to parse (for computers) json format
- **Pairend Merge Report:** information about merged and unmerged reads
- **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
- **User Edits:** List of user edits
- **Project and Project Settings:** Used by software to track settings and data



NGS STR Allele Report

	A	B	C	D	E	F	G
1	#Report: NGS STR Allele Report						
2	#Format: 1						
3	#Version: 2.6.0-beta2-21.0+4f737d02.ci						
4	#Datetime: 2023-08-02T14:39:24						
5	#User: sarah						
6	Sample	Marker	CE Allele	NGS Allele	Count	User Call	User Comment
7	2800M_S1_L	Amelogen	chrX		8837		
8	2800M_S1_L	Amelogen	chrY		7671		
9	2800M_S1_L	PentaE	7	TCTTT[7]	6713	Called	
10	2800M_S1_L	PentaE	13	TCTTT[13]	231	--	
11	2800M_S1_L	PentaE	14	TCTTT[14]	4457	Called	
12	2800M_S1_L	PentaE	14	TCTTT[14]	291	--	
13	2800M_S1_L	D18S51	15	AGAA[15]AAA[1]	251	--	
14	2800M_S1_L	D18S51	16	AGAA[16]AAA[1]	2971	Called	
15	2800M_S1_L	D18S51	17	AGAA[17]AAA[1]	252	--	
16	2800M_S1_L	D18S51	18	AGAA[18]AAA[1]	2595	Called	

The NGS STR Allele Report includes information from the Result table in a .tsv or .fasta format

CE STR Allele Report

	A	B	C	D	E
1	#Report: CE STR Allele Report				
2	#Format: 1				
3	#Version: 2.6.0-beta2-21.0+4f737d02.ci				
4	#Datetime: 2023-08-02T14:40:09				
5	#User: sarah				
6	Sample	Marker	Allele	Height	
7	2800M_S1	Amelogen	X	8837	
8	2800M_S1	Amelogen	Y	7671	
9	2800M_S1	D8S1179	13	462	
10	2800M_S1	D8S1179	14	6224	
11	2800M_S1	D8S1179	15	5550	
12	2800M_S1	D12S391	17	242	
13	2800M_S1	D12S391	18	3826	
14	2800M_S1	D12S391	22	485	
15	2800M_S1	D12S391	23	3318	
16	2800M_S1	TPOX	10	472	

	A	B	C	D	E	F	G	H	I	J
1	#Report: CE STR GM Allele Report									
2	#Format: 1									
3	#Version: 2.6.0-beta2-21.0+4f737d02.ci									
4	#Datetime: 2023-08-02T14:40:29									
5	#User: sarah									
6	Sample	Marker	Allele#1	Allele#2	Allele#3	Allele#4	Height#1	Height#2	Height#3	Height#4
7	2800M_S1	Amelogen	X	Y			8837	7671		
8	2800M_S1	D8S1179	13	14	15		462	6224	5550	
9	2800M_S1	D12S391	17	18	22	23	242	3826	485	3318
10	2800M_S1	PentaE	7	13	14		6713	231	4748	
11	2800M_S1	TH01	6	9.3			4094	3586		
12	2800M_S1	TPOX	10	11			472	10180		
13	2800M_S1	DYS19	13	14	15		475	8228	194	
14	2800M_S1	D21S11	28	29	30.2	31.2	421	5119	378	5392
15	2800M_S1	FGA	19	20	22	23	254	5257	310	3749
16	2800M_S1	DYS389II	30	31	32		791	4939	363	

The CE STR Allele Report includes allele and height (number of reads) information. The GM version (right) mimics the reporting style found in GeneMarkerHID and other CE analysis software.

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!