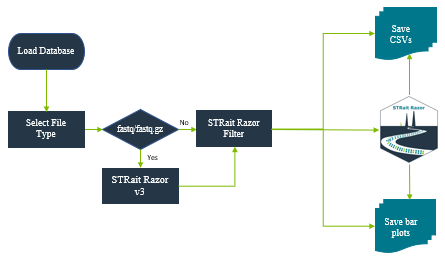
STRait Razor Online Manual



**Contents**

[Purpose 2](#_Toc55294386)

[Installation (Beta Test) 3](#_Toc55294387)

[Landing Page 4](#_Toc55294388)

[Data Analysis 5](#_Toc55294389)

[All Loci Bar Plot 6](#_Toc55294390)

[Final Profile Bar Plot 6](#_Toc55294391)

[Manual 7](#_Toc55294392)

[Contact 7](#_Toc55294393)

[Settings 7](#_Toc55294394)

[Troubleshooting 8](#_Toc55294395)

[Windows Electron Application 8](#_Toc55294396)

[Appendix A: DatabasePath.csv 9](#_Toc55294397)

[Appendix B: STRaitRazorConfig.csv 9](#_Toc55294398)

[Appendix C: HaplotypeDatabase.csv 9](#_Toc55294399)

[Appendix D: STRaitRazorAnalysisConfig.csv 10](#_Toc55294400)

[Changelog 12](#_Toc55294401)

# Purpose

*STRait Razor Online (SRO) serves as the user-interface (UI) for analyzing sequencing data with STRait Razor. Using the SRO, converts STRait Razor\* sequence-based allele calls into genotype tables for and/or bar plots for downstream analysis. STRait Razor Online also allows the user to control the genotype calls using several thresholds (e.g., heterozygous balance, strand balance, etc.). Read strands are merged and reported as single strand (Default = Forward) in Genotype Tables.*

*\*STRait Razor use is detailed in separate manual. Use of this workbook assumes a degree of STRait Razor knowledge prior to use.*

A close up of a logo

Description automatically generated

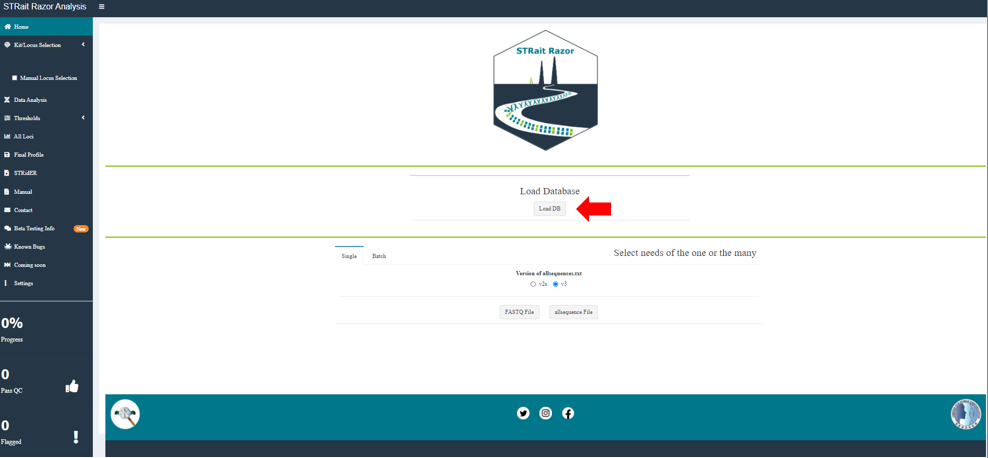
# Installation (Beta Test)

*Congrats on joining us on this beta journey. It will be painful [in comparison to typing install.packages(“STRaitRazor”)]. But I sincerely thank you for your bravery. The final version will be much easier to bring online.*

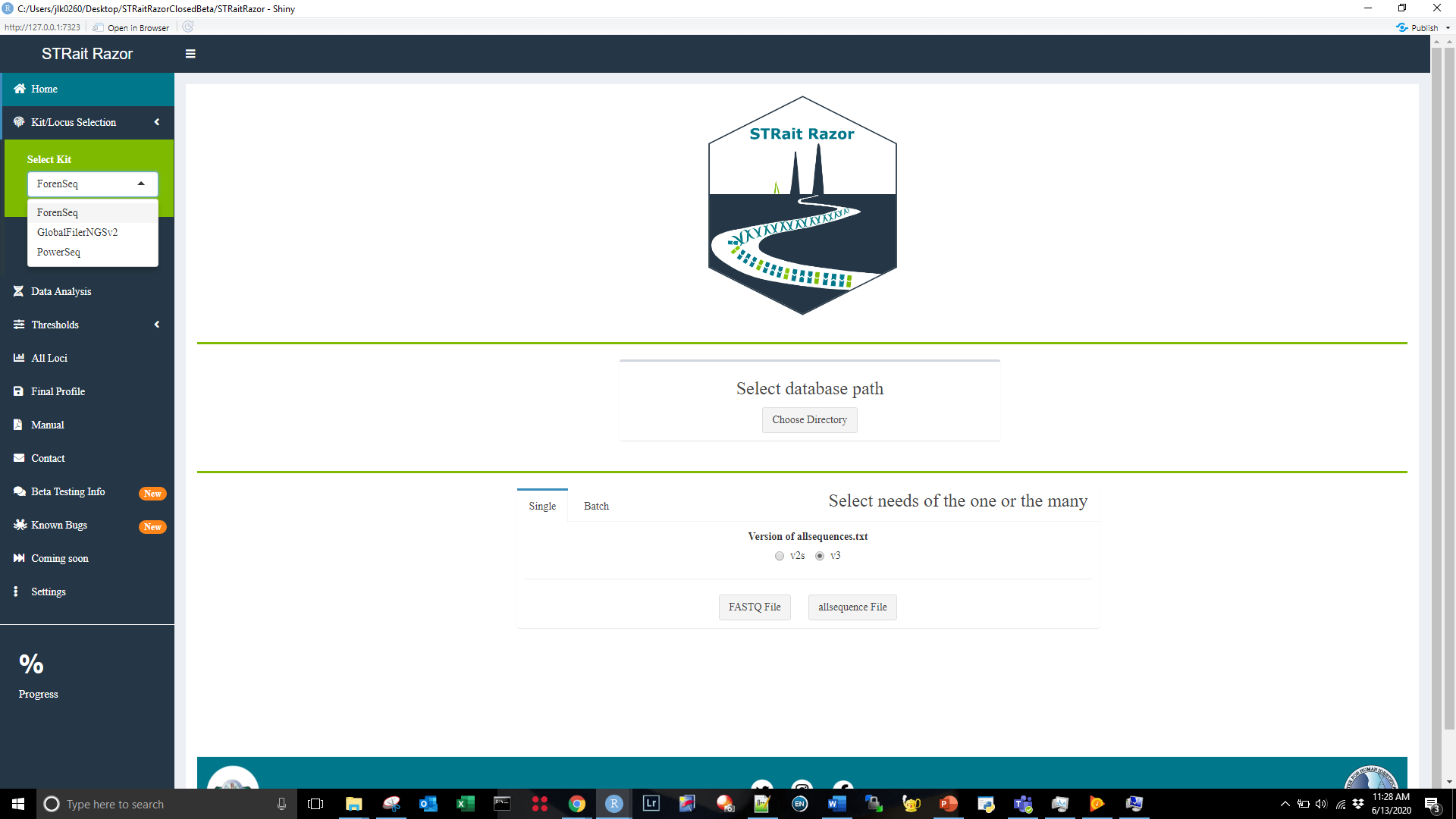
Note to **macOS** users: Please install Xquartz from <https://www.xquartz.org/> and then logout and log back in.

1. Download the package at the attached link.
2. Unzip the STRaitRazorOnline.zip file.
   1. **Linux** users: rename *str8rzr\_linux* to *str8rzr*
   2. **macOS** users: rename *str8rzr\_osX* to *str8rzr*
3. If you don’t have R installed, install R from your mirror of choice <https://www.r-project.org/>
4. If you don’t have RStudio installed, install RStudio Desktop from <https://rstudio.com/products/rstudio/download/#download>
5. Once you have R and RStudio installed, a few packages need to be installed. You can install these packages individually using the included STRait\_Razor\_Online\_Installation.Rmd found in the ~\\STRaitRazorOnline\\ directory.
   1. **Note**: These packages are Imports not Suggests. So, make sure you get them all. 😊
6. Once all the packages are installed, open ‘app\_standalone.R’ script from the same directory as the SROInstallation.Rmd.

# Landing Page



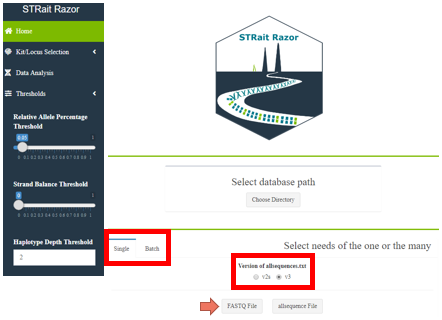
1. Load the database files into the environment using the ‘Load DB’ Button.
2. In sidebar, use the dropdown under ‘Select Kit’ based on the amplification primers\* used.



\***Note**: ForenSeq PMB is the default state for analysis.

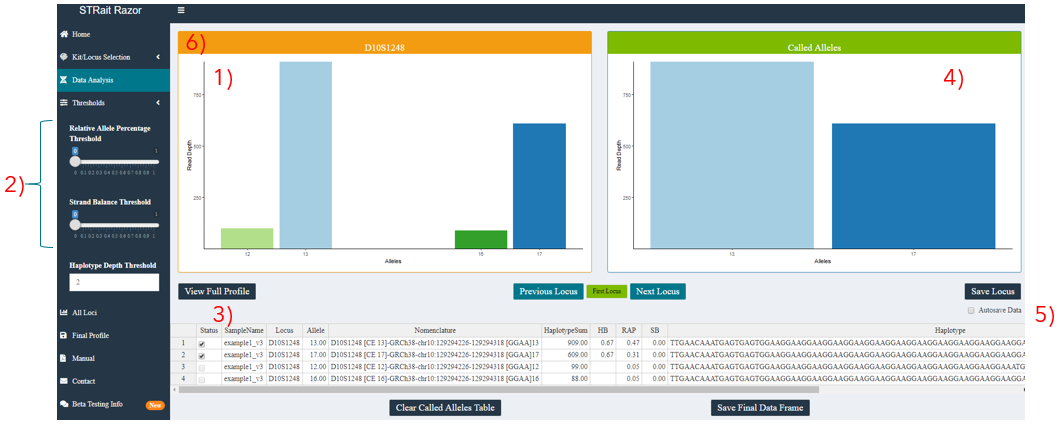
1. Select ‘Single’ or ‘Batch’ Tab for analysis of one “sample” or >= 2 “samples”.

*Note: the process of compiling the list of fastq files can occur recursively or not (default = TRUE). If you have subdirectories with the same sample name, results may be overwritten* ☹*. See ‘Settings’ tab on sidebar to change.*



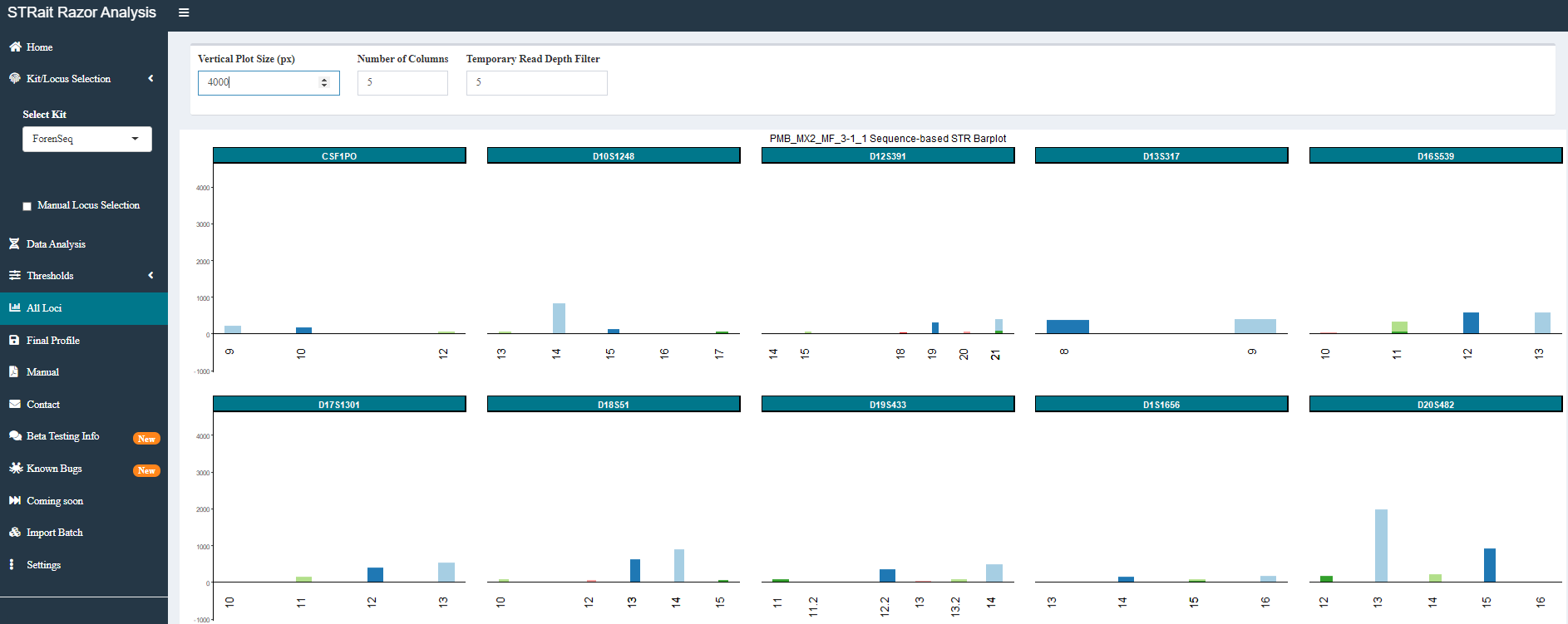
1. STRait Razor Online accepts both fastq/fastq.gz and allsequences.txt files from previous analyses to reduce analysis time of “reprocessing” fastq files. If processing allsequences.txt, please select ‘v2s’ or ‘v3’ format.

# Data Analysis



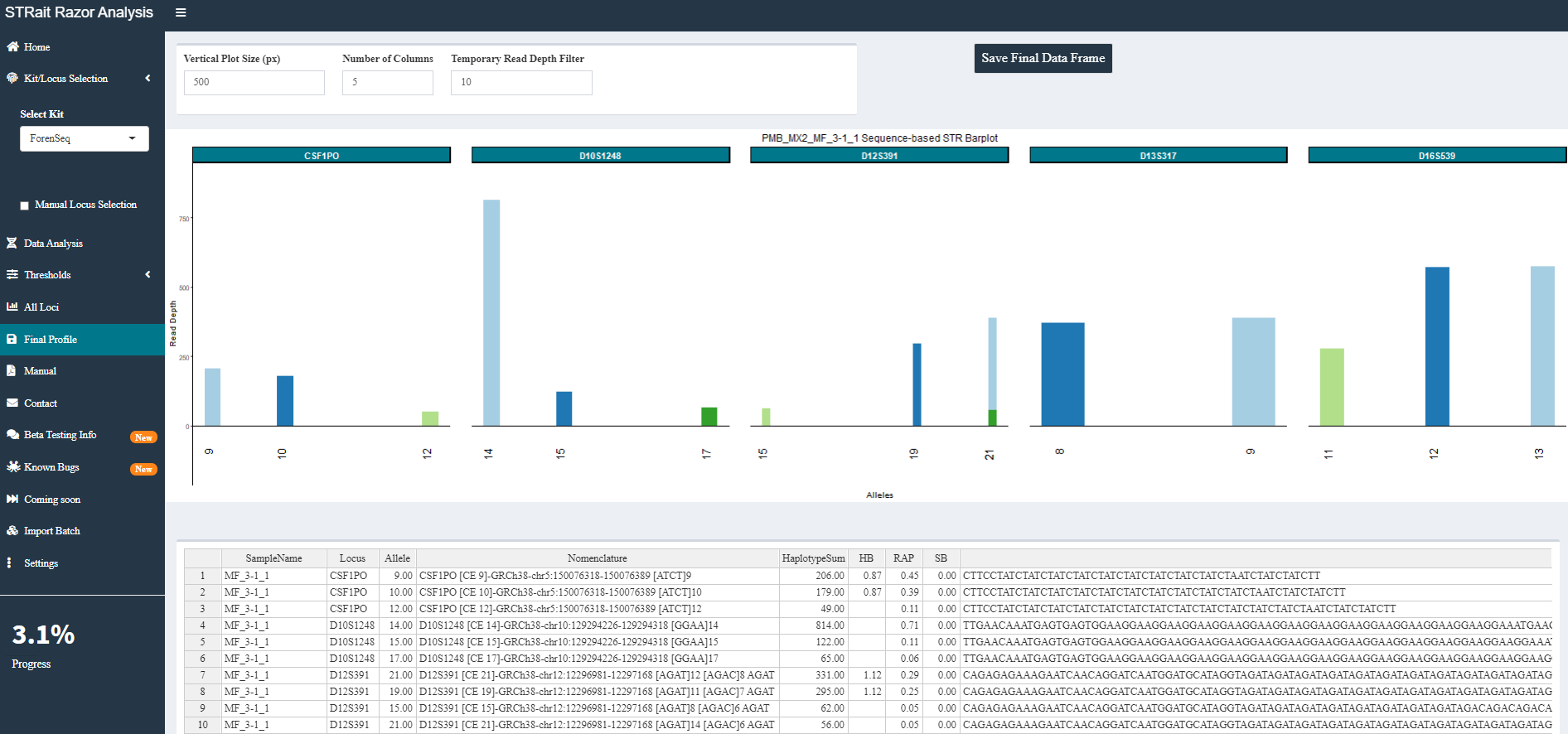
1. Raw plot of reads above threshold
2. Thresholds
   1. per locus
   2. editable via input in sidebar (dynamically for UI only) or STRaitRazorAnalysis.config (static within an analysis; haplotypes passing conditionals create data frame for analysis; Appendix A)
3. Table of alleles > threshold
   1. “Checking” the ‘Status’ of an allele will pass the result to the “Called Alleles” plot
4. Called Alleles bar plot
   1. Final set of alleles
5. Autosave Data
   1. “Checking” this will push called alleles to final data frame when you press ‘Next Locus’
6. Status Bar
   1. In current version, status bar reflects loci with high Relative Allele Proportion (RAP) (e.g., Orange < AutoRAPT). Release version will include more optimized elements. (See settings page for more info)
7. Other buttons
   1. Most other buttons are self-explanatory

# All Loci Bar Plot



1. Bar plots for all STR loci (plus Amelogenin) are shown for profile-level view of haplotype counts.
2. This plot may be scaled using the controls at the top.

# Final Profile Bar Plot



1. Bar plots for STR loci (plus Amelogenin) after analyst’s interpretation are shown for profile-level view of haplotype counts read for export.
2. This plot may be scaled using the controls at the top.

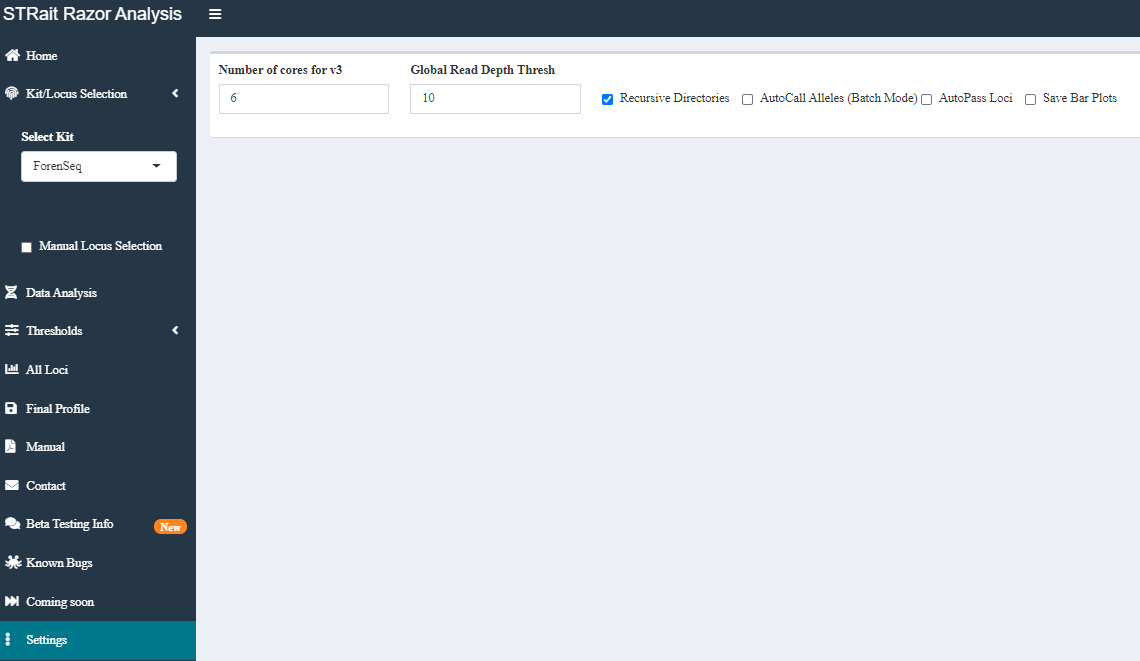
# Manual

Have you ever seen the movie Inception? It is not quite the same. But it does launch a pdf of this doc.

# Contact

Bat signal, but for data analysis help.

# Settings



While this page will look significantly different with the final batch of settings added, this implementation controls a few different functions.

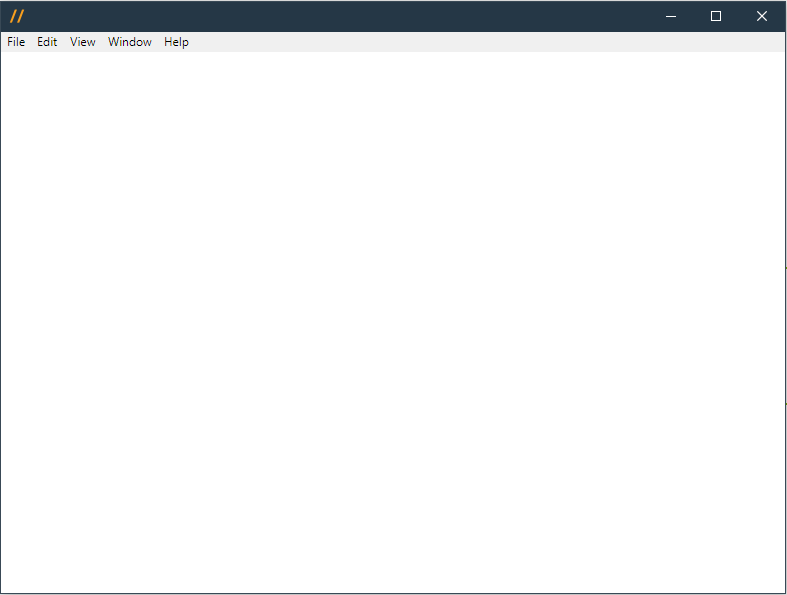
* 1. **Number of cores for v3**: Allocate system resources for fastq processing using str8rzr program
  2. **Global Read Depth Thresh**: Filter out haplotypes with fewer than X reads
  3. **Recursive Directories**: When processing a batch of files (either fastq or allsequences), process chosen directory or directory plus subdirectories
  4. **AutoPass Loci**: When processing single files, “passing” green loci are automatically moved to final data frame and “warning” orange loci are passed to ‘Data Analysis’ tab for interpretation (with some more optimization, will likely move this from Default = False 🡪 True)
  5. **Save Bar Plots**: To save a .png of all loci bar plot to sample folder additionally a conditional setting for separating STR and SNP loci into separate image files

# Troubleshooting

## Windows Electron Application

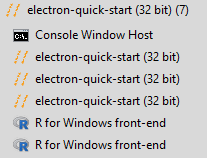
* **Issue**: When launching electron app A close up of a sign

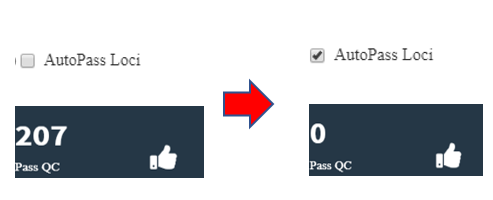
  Description automatically generated, users are met with blank screen.

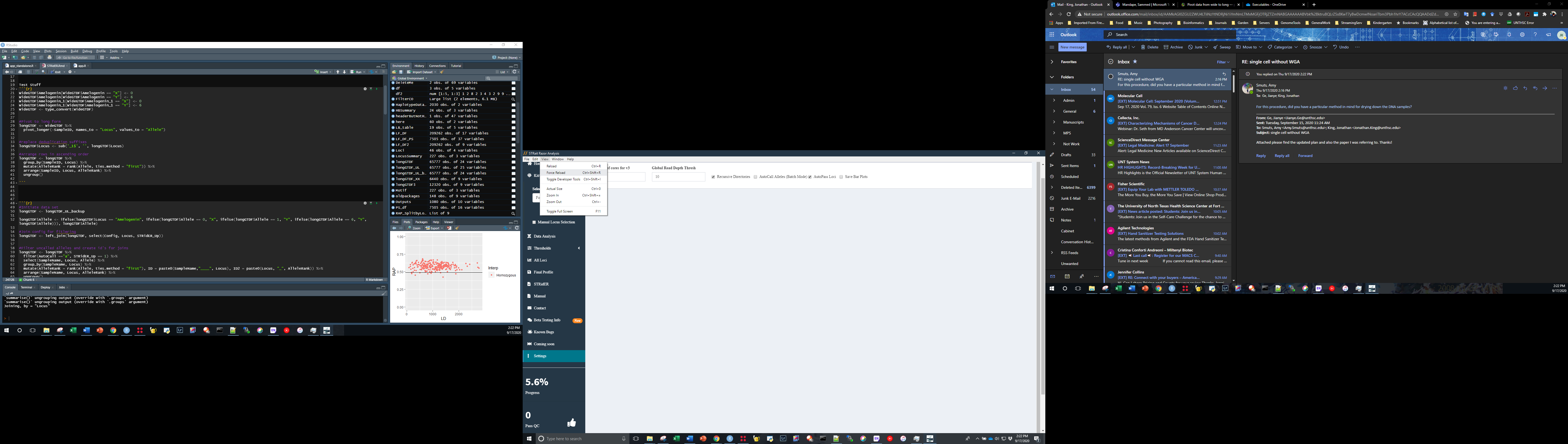


* **Solution**: View 🡪 Force Reload
* **Issue**: When launching electron app A close up of a sign

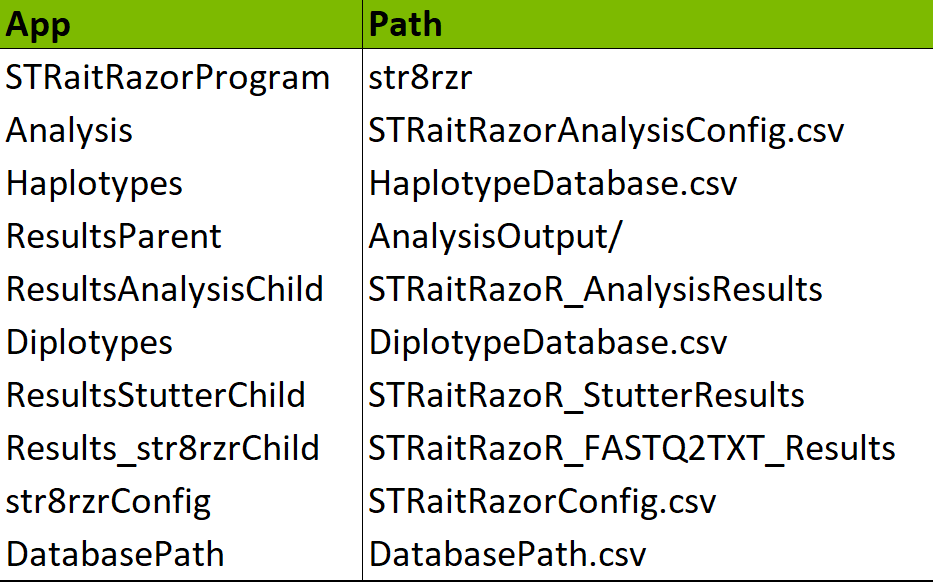
  Description automatically generated, users are met with error message “An error has occurred”.
* **Solution**: Open Task Manager and close any instances of *electron-quick-start* or *R for Windows front-end*. After these processes have been closed, relaunch the application.



* **Issue**: When selecting Settings 🡪 AutoPass Loci, Pass Loci Count changes to zero.
* 
* **Solution**: View 🡪 Force Reload

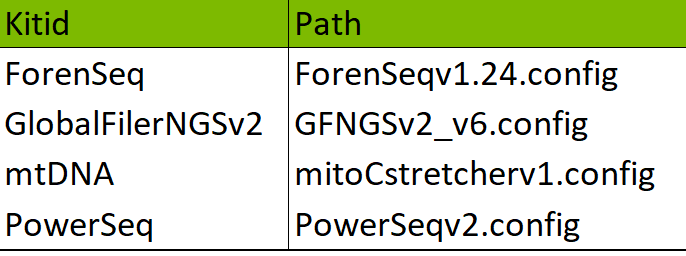


# Appendix A: DatabasePath.csv



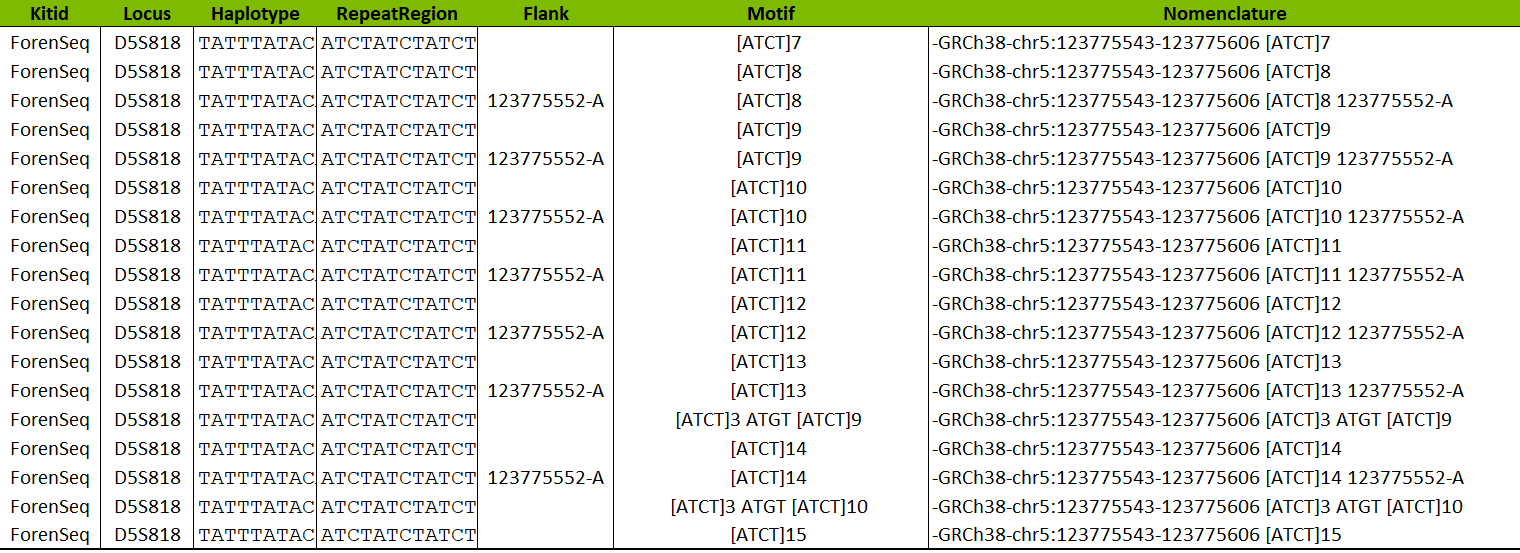
1. Database files and directory for fastq processing and analysis functions

# Appendix B: STRaitRazorConfig.csv



1. Configuration files for str8rzr application to analyze fastq/fast.gz files

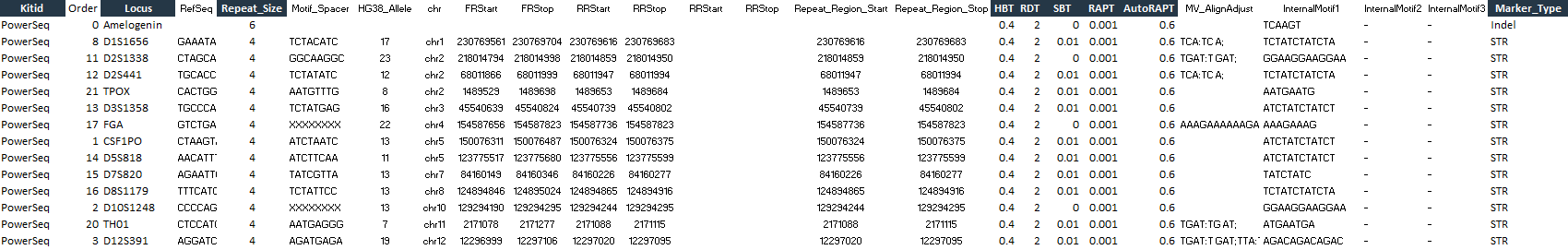
# Appendix C: HaplotypeDatabase.csv



This database file contains haplotypes with associated metadata (i.e., flanking region SNPs, repeat region motif, and region bounds).

*Note: The SNPs are largely annotated by position. Ultimately, these will be migrating to rs#.*

# Appendix D: STRaitRazorAnalysisConfig.csv



*For this implementation, we will be focusing on the columns currently implemented [graphite shaded].*

**Kit**: Amplification kit(s) used for target enrichment\*

**\*Note:** ForenSeq PMB is the default state for analysis.

**Locus**: Full list of markers in each kit

**Repeat\_Size**: Period of the repeat (e.g., CSF1PO: ATCT; 4 base repeat)

**HBT**: Heterozygote Balance Threshold on a per marker & per kit basis. This is used for assignment of second allele prior to data frame passing to UI

*Danger\_math\_ahead*: The threshold is calculated by dividing the second largest, in terms of coverage or read depth, by the largest allele (e.g., the allele 12 has 932 reads associated with it and allele 17.3 has 950 reads. Thus, the heterozygote balance is 0.98). However, the heterozygote balance output in the genotype tables is calculated by dividing the largest (lexicographically) allele by the second largest allele (e.g., . or ).

**RDT**: Read Depth Threshold on a per marker & per kit basis filter prior to data frame passing to UI

**SBT**: Strand Balance Threshold on a per marker & per kit basis filter prior to data frame passing to UI

**RAPT**: Relative Allele Proportion Threshold on a per marker & per kit basis filter prior to data frame passing to UI

**AutoRAPT**: Locus flagging variable. In this beta, loci with a proportion of called alleles above this value will be flagged “green” (e.g., CSF1PO; 1000 reads aligned to the locus, 12: 450, 14: 450, AutoRAPT = 0.857; 900/1000 == 0.90; 0.9 > 0.857, CSF1PO = “green”)

**Marker\_Type**: Categories of marker type (e.g., SNP, microhaplotype, STR) used for parsing markers

**Appendix E: DiplotypeDatabase.csv**

\*Under Construction\*

\*Check Back Soon\*

**Appendix F: RepeatRegion.csv**

\*Under Construction\*

\*Check Back Soon\*

**Appendix G: RepeatRegion\_MM.csv**

\*Under Construction\*

\*Check Back Soon\*

# Changelog

v 0.1.2: **07/20/2020**

-Launch Day!!!!!!!!!!!

v 0.1.3: **08/19/2020**

-minor bug fixes

v 0.1.4: **09/01/2020**

-added STRidER tab

-STRidER input fields

-added CV to RAP batch output

-fixed bug affecting SNP loci with secondary SNP variant in matching sequence

-added progress bar for batch mode or file processing

v 0.1.5: **09/17/2020**

-added data.table package to address STRidER appending issue

-finalized STRidER functions for multi-sample processing

-cleaned up in script comments regarding page titles

-removed readxl package (not used in current implementation)

-added code to clean up UI on start-up

-Launched Windows Electron application

-added function subThresh for saving data frame of reads >= GlobalThreshold, but < locus threshold

-added Indels (e.g., Amelogenin) to All Loci and Final Profile ggplots

v 0.1.51: **09/18/2020**

-corrected bug for fastq processing related to relative path of configs in unzipped fastq pipeline (app\_online.R only)

v 0.1.6: **11/10/2020**

-adjusted calculation of heterozygote balance for isoalleles and SNP loci

-added toggle switch for expected vs. observed

-adjusted calculation of progress box

-added spacer box for missing loci when cycling through under data analysis

-modified allele reporting to SNP/Amel to character rather than numeric in table output

-removed closed beta tabs

v 0.1.7 (#NoCodeUpdate): **01/05/2021**

-Updated recommended PowerSeq config v2-->v2.1

v 0.1.8: **05/10/2021**

-Changed Kit IDs

ForenSeq --> ForenSeq DNA Signature

PowerSeq --> PowerSeq 46GY

-Added Kit ForenSeq MainstAY

-Updated config files

ForenSeq DNA Signature: ForenSeqv1.25 --> ForenSeqv1.26

GlobalFilerNGSv2: GFNGSv2\_v7 --> GFNGSv2\_v7.1

PowerSeq 46GY: PowerSeqv2.1 --> PowerSeqv3.1

-Updated Sample Name bug in All Loci Tab to display sample name rather than locus

v 0.1.9 (#NoCodeUpdate): **07/23/2021**

-Added Kit IDseek SNP85

v 0.2: **08/19/2021**

-Added Locus to haplotype for revComp filter to account for repeat region similarities

v 0.2.1: **10/04/2021**

-Added hotfix for processing multi-format data alongside single-format (i.e., STRs with STR & SNP data in batch processing) credit: S.P. from VCU

v 0.2.2 (#NoCodeUpdate): **11/05/2021**

-Updated MainstAY config file to include flanking region information for DYS393

-Added 7z.dll to bin to account for dll module failure on fastq unzipping

v 0.2.3: **06/13/2022**

-Updated 'Find New Alleles' section to address no new data bug (Standalone only)

v 0.2.4: **07/21/2022**

-Added reactive element to df2() to address occasional bug where Data Analysis tab active locus would not refresh when loading new sample

V 0.2.5: 08/12/2022

-Added IDseek Auto 30

-Added IDseek Y-STR 27

-Modified note (online only) regarding Sample Name input to address bug where sample names without file extension would not parse correctly and result in crash