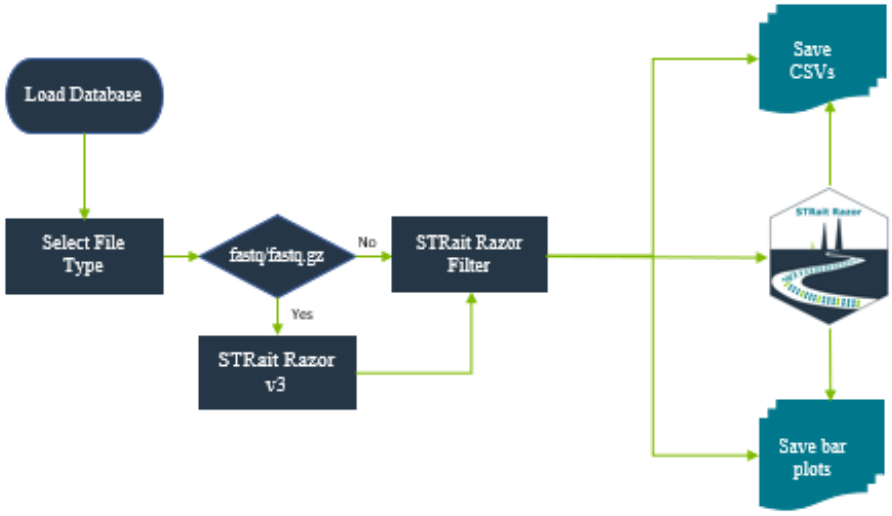
**Center for Human Identification**
**Research and Development Unit**

# STRait Razor Shiny Manual




## Contents

**Center for Human Identification**
**Research and Development Unit**

## Purpose

*STRait Razor Shiny (SRS) serves as the user-interface (UI) for analyzing sequencing data with STRait Razor. Using the SRS, converts STRait Razor* sequence-based allele calls into genotype tables for and/or bar plots for downstream analysis. STRait Razor Shiny 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.*

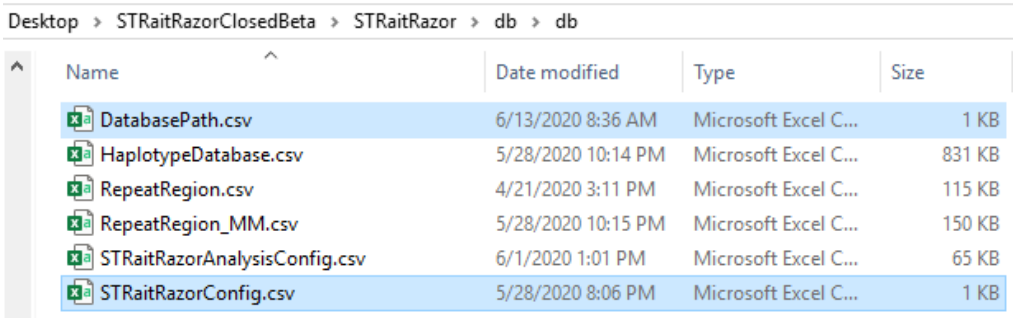
**Center for Human Identification**
**Research and Development Unit**

# Installation (Beta Test)

*Congrats on joining on us 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

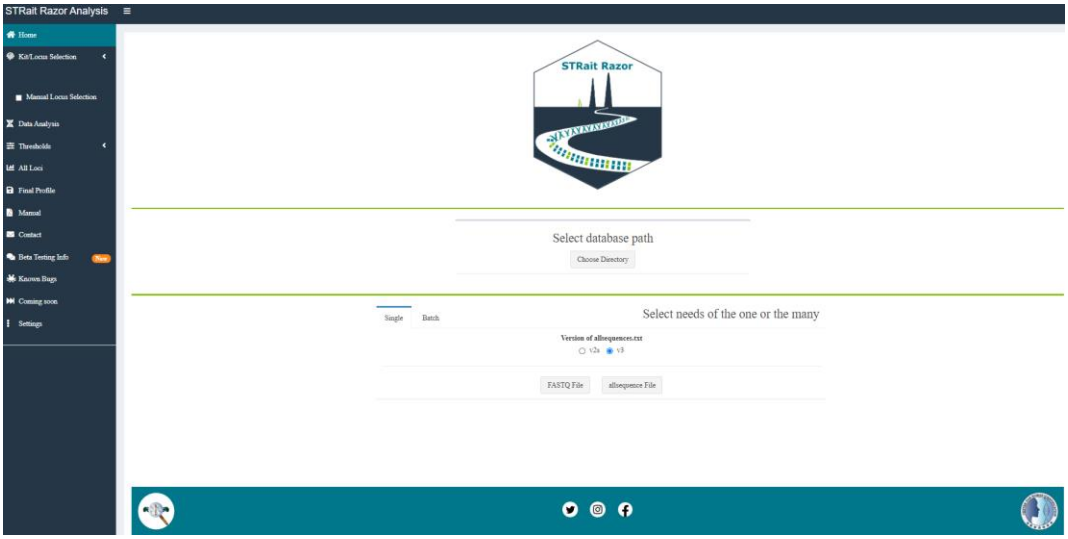
1) Download the package at the attached link.
2) Unzip and place folder on Desktop. For this beta and all the discussion below, this will be your home and ~\\STRaitRazorClosedBeta\\STRaitRazor\\ will contain all your toys.
3) Update path listings in the following file
    a. Change jlk0260 → euid*
        i. DatabasePath.csv (Appendix A)
        ii. STRaitRazorConfig.csv (Appendix B)




* Note: If you did not place your file on the desktop or if you are using a personal computer, please update path accordingly.

4) If you don't have R installed, install R from your mirror of choice https://www.r-project.org/
5) A number of "Imports" will be installed automatically when the package releases on GitHub; however, at this point, you can install these individually using the included STRait_RazoR_Installation.Rmd found in the ~\\STRaitRazorClosedBeta\\STRaitRazor\\ directory. These packages are Imports not Suggests. So, make sure you get them all. ☺
6) Once all the packages are installed, it is time to roll. Open 'app.R' script from the same directory as the installation.Rmd.

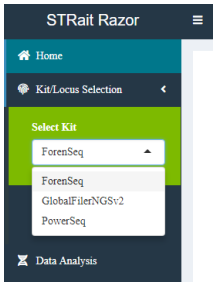
# Landing Page




1. Select the path ~\\STRaitRazorClosedBeta\\STRaitRazor\\db\\db\\ using the 'Choose Directory' Button.

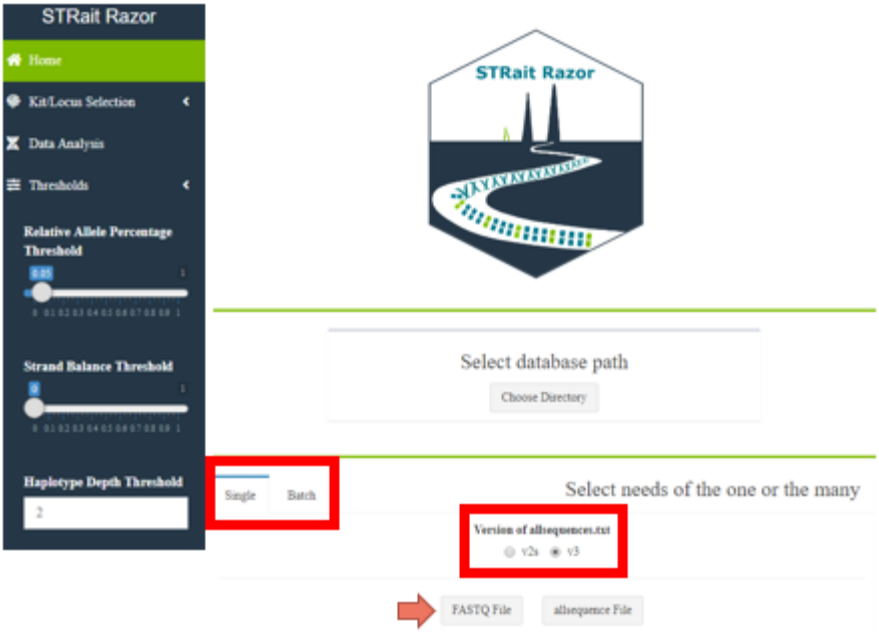
   *Note: This dialog box sometimes launches behind the app…*

2. In sidebar, use the dropdown under 'Select Kit' based on the amplification primers used.



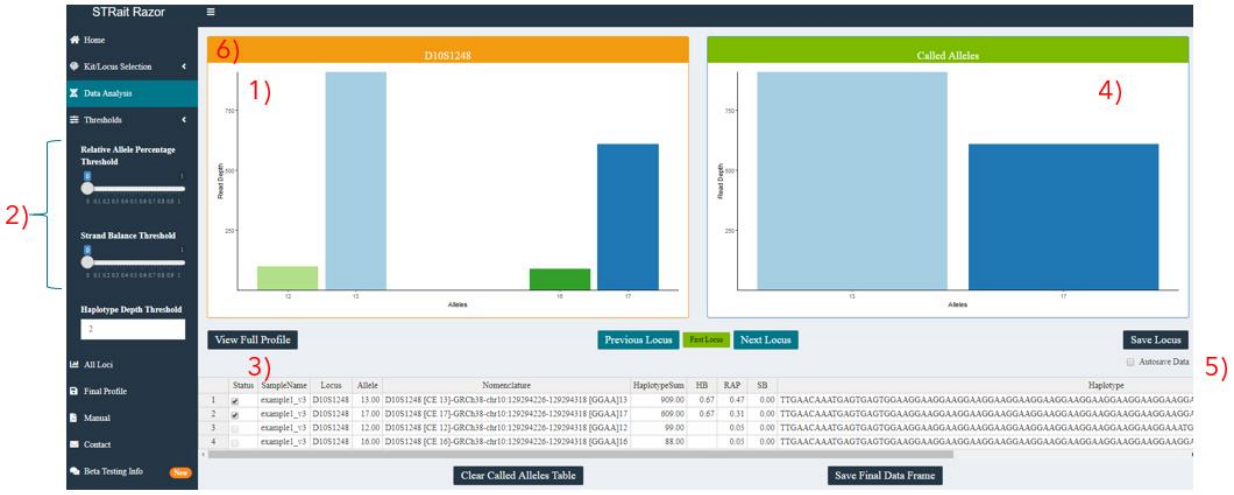

3. Select 'Single' or 'Batch' Tab for analysis of one "sample" or >= 2 "samples".

   *Note: the processing of compiling the list of 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.*



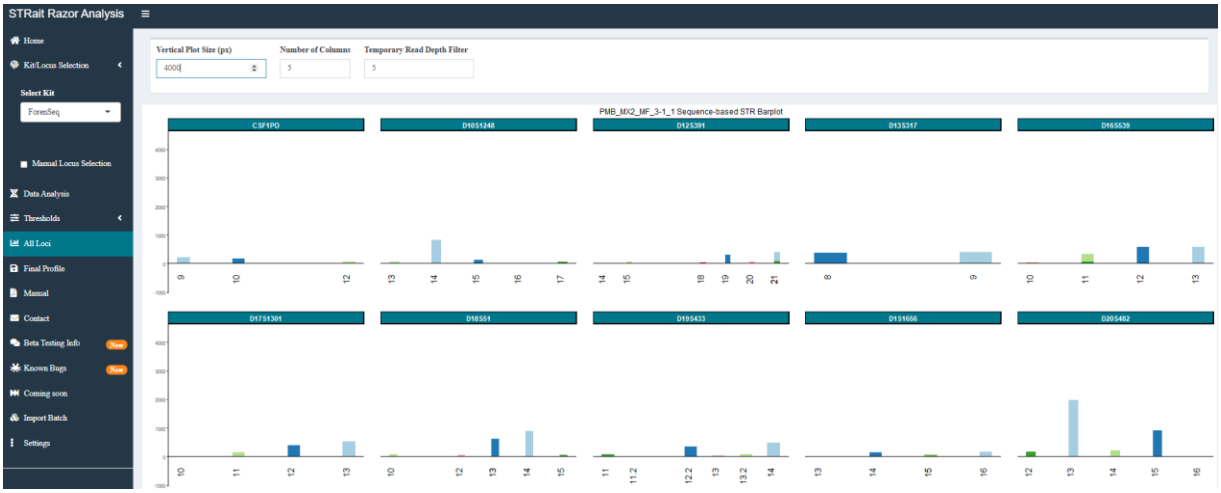

4. STRait Razor Shiny accepts both fastq/fastq.gz and allsequnces.txt files from previous analyses to reduce analysis time of “reprocessing” fastq files. If processing allsequences.txt, please select ‘v2s’ or ‘v3’ format.

# Data Analysis



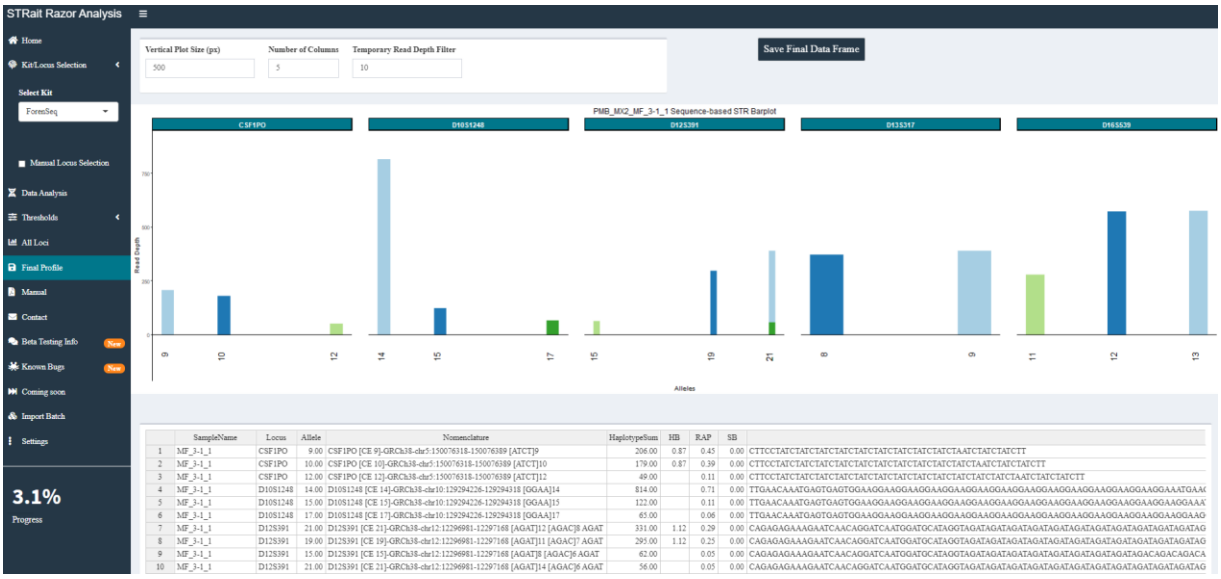

1. Raw plot above threshold
2. Thresholds
   a. per locus
   b. editable via input in sidebar (dynamically for UI only) or STRaitRazorAnalysis.config (static within an analysis; haplotypes passing conditionals create data frame for analysis; Appendix A)
3. Table of alleles > threshold
   a. “Checking”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

## All Loci Bar Plot




1. Bar plots for all STR loci are shown for profile-level view of haplotype counts.
2. This plot maybe scaled using the controls at the top.

## Final Profile Bar Plot




1. Bar plots for STR loci after analyst's interpretation are shown for profile-level view of haplotype counts read for export.
2. This plot maybe scaled using the controls at the top.

**Center for Human Identification**
**Research and Development Unit**

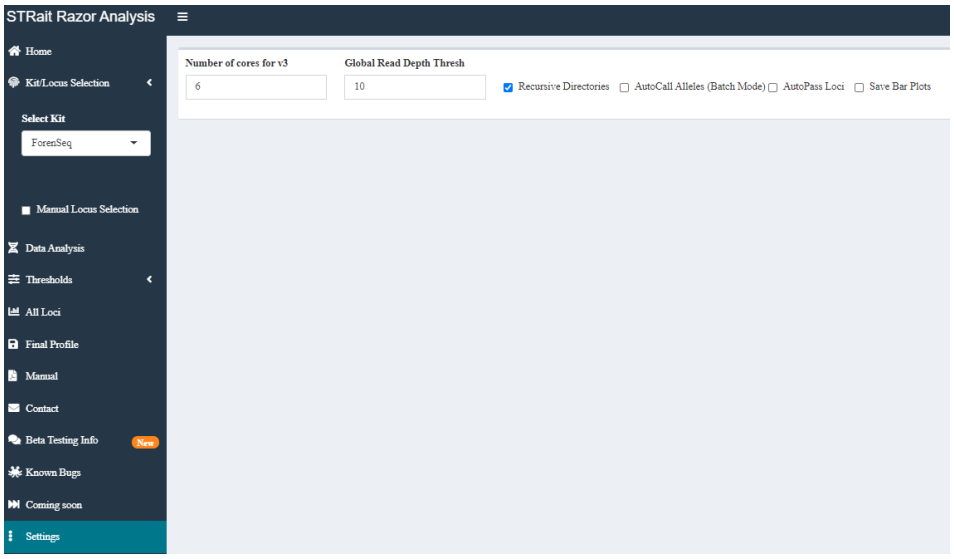
# Manual

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

# Contact

Bat signal, but for data analysis.

# Settings




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

a. **Number of cores for v3**: Allocate system resources for fastq processing using str8rzr program
b. **Global Read Depth Thresh**: Filter out haplotypes with fewer than X reads
c. **Recursive Directories**: When processing a batch of files (either fastq or allsequences), process chosen directory or directory plus subdirectories
d. **AutoCall Alleles (Batch Mode)**: Filter out unchecked alleles when processing multiple files. Feels ugly (might delete later)
e. **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)
f. **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

## Appendix A: DatabasePath.csv

| DBFile/Directory | Path |
|---|---|
| STRaitRazorProgram | C:/Users/jlk0260/Desktop/STRaitRazorClosedBeta/STRaitRazor/bin/str8rzr |
| Analysis | C:/Users/jlk0260/Desktop/STRaitRazorClosedBeta/STRaitRazor/db/db/STRaitRazorAnalysisConfig.csv |
| Haplotypes | C:/Users/jlk0260/Desktop/STRaitRazorClosedBeta/STRaitRazor/db/db/HaplotypeDatabase.csv |
| ResultsParent | C:/Users/jlk0260/Desktop/STRaitRazorClosedBeta/STRaitRazor/data/AnalysisOutput/ |
| ResultsAnalysisChild | STRaitRazoR_AnalysisResults |
| Diplotypes | C:/Users/jlk0260/Desktop/STRaitRazorClosedBeta/STRaitRazor/db/db/DiplotypeDatabase.csv |
| ResultsStutterChild | STRaitRazoR_StutterResults |
| Results_str8rzrChild | STRaitRazoR_FASTQ2TXT_Results |
| str8rzrConfig | C:/Users/jlk0260/Desktop/STRaitRazorClosedBeta/STRaitRazor/db/db/STRaitRazorConfig.csv |
| DatabasePath | C:/Users/jlk0260/Desktop/STRaitRazorClosedBeta/STRaitRazor/db/db/DatabasePath.csv |

Paths to database files, directory for fastq processing program and analysis

## Appendix B: STRaitRazorConfig.csv

| Kitid | Path |
|---|---|
| ForenSeq | C:\Users\jlk0260\Desktop\STRaitRazorClosedBeta\STRaitRazor\db\configs\ForenSeqv1.2.config |
| GlobalFilerNGSv2 | C:\Users\jlk0260\Desktop\STRaitRazorClosedBeta\STRaitRazor\db\configs\GFNGSv2_v5.config |
| mtDNA | C:\Users\jlk0260\Desktop\STRaitRazorClosedBeta\STRaitRazor\db\configs\mitoCstretcherv1.config |
| PowerSeq | C:\Users\jlk0260\Desktop\STRaitRazorClosedBeta\STRaitRazor\db\configs\PowerSeqv2.config |

Paths to configuration files for str8rzr application to analyze fastq/fast.gz files

## Appendix C: HaplotypeDatabase.csv

| Kitid | Locus | Haplotype | RepeatRegion | Flank | Motif | Nomenclature |
|---|---|---|---|---|---|---|
| ForenSeq | D5S818 | TATTTATACCTCTA | ATCTATCTATCTATC | | [ATCT]7 | -GRCh38-chr5:123775543-123775606 [ATCT]7 |
| ForenSeq | D5S818 | TATTTATACATCTA | ATCTATCTATCTATC | 123775552-A | [ATCT]8 | -GRCh38-chr5:123775543-123775606 [ATCT]8 123775552-A |
| ForenSeq | D5S818 | TATTTATACCTCTA | ATCTATCTATCTATC | | [ATCT]8 | -GRCh38-chr5:123775543-123775606 [ATCT]8 |
| ForenSeq | D5S818 | TATTTATACATCTA | ATCTATCTATCTATC | 123775552-A | [ATCT]9 | -GRCh38-chr5:123775543-123775606 [ATCT]9 123775552-A |
| ForenSeq | D5S818 | TATTTATACCTCTA | ATCTATCTATCTATC | | [ATCT]9 | -GRCh38-chr5:123775543-123775606 [ATCT]9 |
| ForenSeq | D5S818 | TATTTATACATCTA | ATCTATCTATCTATC | 123775552-A | [ATCT]10 | -GRCh38-chr5:123775543-123775606 [ATCT]10 123775552-A |
| ForenSeq | D5S818 | TATTTATACCTCTA | ATCTATCTATCTATC | | [ATCT]10 | -GRCh38-chr5:123775543-123775606 [ATCT]10 |
| ForenSeq | D5S818 | TATTTATACATCTA | ATCTATCTATCTATC | 123775552-A | [ATCT]11 | -GRCh38-chr5:123775543-123775606 [ATCT]11 123775552-A |
| ForenSeq | D5S818 | TATTTATACCTCTA | ATCTATCTATCTATC | | [ATCT]11 | -GRCh38-chr5:123775543-123775606 [ATCT]11 |
| ForenSeq | D5S818 | TATTTATACATCTA | ATCTATCTATCTATC | 123775552-A | [ATCT]12 | -GRCh38-chr5:123775543-123775606 [ATCT]12 123775552-A |
| ForenSeq | D5S818 | TATTTATACCTCTA | ATCTATCTATCTATC | | [ATCT]12 | -GRCh38-chr5:123775543-123775606 [ATCT]12 |
| ForenSeq | D5S818 | TATTTATACCTCTA | ATCTATCTATCTATC | | [ATCT]3 ATGT [ATCT]9 | -GRCh38-chr5:123775543-123775606 [ATCT]3 ATGT [ATCT]9 |
| ForenSeq | D5S818 | TATTTATACATCTA | ATCTATCTATCTATC | 123775552-A | [ATCT]13 | -GRCh38-chr5:123775543-123775606 [ATCT]13 123775552-A |
| ForenSeq | D5S818 | TATTTATACCTCTA | ATCTATCTATCTATC | | [ATCT]13 | -GRCh38-chr5:123775543-123775606 [ATCT]13 |
| ForenSeq | D5S818 | TATTTATACCTCTA | ATCTATCTATCTATC | | [ATCT]3 ATGT [ATCT]10 | -GRCh38-chr5:123775543-123775606 [ATCT]3 ATGT [ATCT]10 |
| ForenSeq | D5S818 | TATTTATACATCTA | ATCTATCTATCTATC | 123775552-A | [ATCT]14 | -GRCh38-chr5:123775543-123775606 [ATCT]14 123775552-A |
| ForenSeq | D5S818 | TATTTATACCTCTA | ATCTATCTATCTATC | | [ATCT]14 | -GRCh38-chr5:123775543-123775606 [ATCT]14 |
| ForenSeq | D5S818 | TATTTATACCTCTA | ATCTATCTATCTATC | | [ATCT]15 | -GRCh38-chr5:123775543-123775606 [ATCT]15 |

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. Prior to release, I plan on migrating all these to rs#.*

## Appendix D: STRaitRazorAnalysisConfig.csv

| Kitid | Order | Locus | RefSeq | Repeat_Size | Motif_Spacer | HG38_Allele | chr | FRStart | FRStop | RRStart | RRStop | RRStart | RRStop | Repeat_Region_Start | Repeat_Region_Stop | HBT | RDT | SBT | RAPT | AutoRAPT | MV_AlignAdjust | InternalMotif1 | InternalMotif2 | InternalMotif3 | Marker_Type |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PowerSeq | 0 | Amelogenin | | 6 | | | | | | | | | | | | 0.4 | 2 | 0 | 0.001 | 0.6 | | TCAAGT | - | - | Indel |
| PowerSeq | 8 | D1S1656 | GAAATA | 4 | TCTACATC | 17 | chr1 | 230769561 | 230769704 | 230769616 | 230769683 | | | 230769616 | 230769683 | 0.4 | 2 | 0.01 | 0.001 | 0.6 | TCA:TC A; | TCTATCTATCTA | - | - | STR |
| PowerSeq | 11 | D2S1338 | CTAGCA | 4 | GGCAAGGC | 23 | chr2 | 218014794 | 218014938 | 218014859 | 218014950 | | | 218014859 | 218014950 | 0.4 | 2 | 0 | 0.001 | 0.6 | TGAT:T GAT; | GGAAGGAAGGAA | - | - | STR |
| PowerSeq | 12 | D2S441 | TGCACC | 4 | TCTATATC | 12 | chr2 | 68011866 | 68011999 | 68011947 | 68011994 | | | 68011947 | 68011994 | 0.4 | 2 | 0.01 | 0.001 | 0.6 | TCA:TC A; | TCTATCTATCTA | - | - | STR |
| PowerSeq | 21 | TPOX | CACTGG | 4 | AATGTTTG | 8 | chr2 | 1489529 | 1489638 | 1489653 | 1489684 | | | 1489653 | 1489684 | 0.4 | 2 | 0.01 | 0.001 | 0.6 | | AATGAATG | - | - | STR |
| PowerSeq | 13 | D3S1358 | TGCCCA | 4 | TCTATGAG | 16 | chr3 | 45540639 | 45540824 | 45540739 | 45540802 | | | 45540739 | 45540802 | 0.4 | 2 | 0.01 | 0.001 | 0.6 | | ATCTATCTATCT | - | - | STR |
| PowerSeq | 17 | FGA | GTCTGA | 4 | XXXXXXXX | 22 | chr4 | 154587656 | 154587823 | 154587736 | 154587823 | | | 154587736 | 154587823 | 0.4 | 2 | 0 | 0.001 | 0.6 | AAAGAAAAAAGA | AAAGAAAG | - | - | STR |
| PowerSeq | 1 | CSF1PO | CTAAGT, | 4 | ATCTAATC | 13 | chr5 | 150076311 | 150076487 | 150076324 | 150076375 | | | 150076324 | 150076375 | 0.4 | 2 | 0.01 | 0.001 | 0.6 | | ATCTATCTATCT | - | - | STR |
| PowerSeq | 14 | D5S818 | AACATT | 4 | ATCTTCAA | 11 | chr5 | 123775517 | 123775680 | 123775556 | 123775599 | | | 123775556 | 123775599 | 0.4 | 2 | 0.01 | 0.001 | 0.6 | | ATCTATCTATCT | - | - | STR |
| PowerSeq | 15 | D7S820 | AGAATT | 4 | TATCGTTA | 13 | chr7 | 84160149 | 84160346 | 84160226 | 84160277 | | | 84160226 | 84160277 | 0.4 | 2 | 0.01 | 0.001 | 0.6 | | TATCTATC | - | - | STR |
| PowerSeq | 16 | D8S1179 | TTTCATC | 4 | TCTATTCC | 13 | chr8 | 124894846 | 124895024 | 124894865 | 124894916 | | | 124894865 | 124894916 | 0.4 | 2 | 0.01 | 0.001 | 0.6 | | TCTATCTATCTA | - | - | STR |
| PowerSeq | 2 | D10S1248 | CCCCAG | 4 | XXXXXXXX | 13 | chr10 | 129294190 | 129294295 | 129294244 | 129294295 | | | 129294244 | 129294295 | 0.4 | 2 | 0 | 0.001 | 0.6 | | GGAAGGAAGGAA | - | - | STR |
| PowerSeq | 20 | TH01 | CTCCAT( | 4 | AATGAGGG | 7 | chr11 | 2171078 | 2171277 | 2171088 | 2171115 | | | 2171088 | 2171115 | 0.4 | 2 | 0.01 | 0.001 | 0.6 | TGAT:TG AT; | ATGAATGA | - | - | STR |

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

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

**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 (Appendix XXXXX) is calculated by dividing the largest (by length) allele by the second largest allele

**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!!