

MPKin-YSTR User Manuel

Version 1.1

Contents

Introduction	2
Software Installation	2
Program execution.....	2
Input files	2
Y-STR haplotype profiles	2
Common Message Format (CMF) file	3
Excel file	3
Y-STR mutation rates	3
Functions.....	4
Profile Comparisons	5
Pedigree Likelihood Ratios.....	8
Mismatch distributions	14
Contacts	18

Introduction

MPKin-YSTR is a software program designed for kinship analysis with Y-STR haplotypes. This software was developed with the support from the National Institute of Justice (NIJ) 2020-DQ-BX-0018. There are three major functions implemented in this tool:

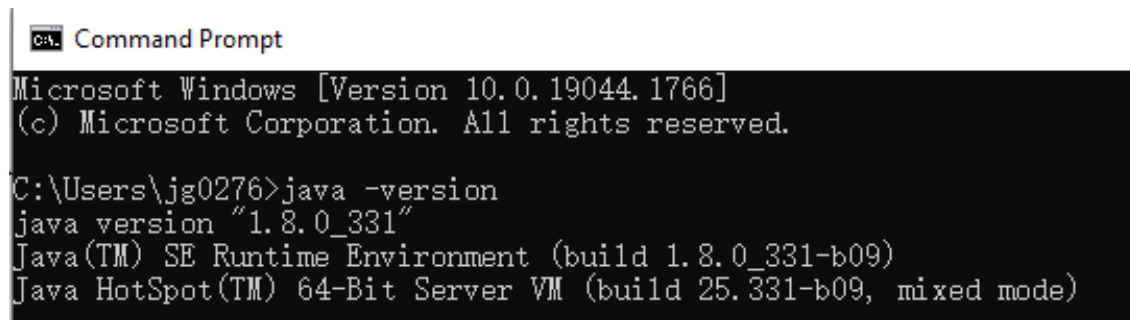
- Y-STR haplotype comparisons (i.e., 1:1, 1:N, and N:N comparisons)
- Pedigree likelihood ratio for kinship analysis with Y-STR profiles
- Draw mismatch distributions to decide a threshold for determining if two profiles are from the same lineage or not

Software Installation

Please make sure your computer (either Windows, Linux, or Mac) has the latest version of Java installed. Please go to <https://www.java.com/> to download and install Java. To find out the version of Java installed on your computer or to find out if your computer has Java installed, please type the following command in the command window:

java -version

The following is an example for the Windows system.



```
Command Prompt
Microsoft Windows [Version 10.0.19044.1766]
(c) Microsoft Corporation. All rights reserved.

C:\Users\jg0276>java -version
java version "1.8.0_331"
Java(TM) SE Runtime Environment (build 1.8.0_331-b09)
Java HotSpot(TM) 64-Bit Server VM (build 25.331-b09, mixed mode)
```

Program execution

The program can be executed by double-clicking the “MPKin-YSTR.jar” file.

The program may also be executed by typing “java -jar MPKin-YSTR.jar” in the command windows. The approach is recommended if a bug is found, so that the error trace may be seen in the command windows.

Input files

There are two types of data files that may be needed in this software: Y-STR haplotype profiles and Y-STR mutation rates

Y-STR haplotype profiles

MPKin-YSTR accepts two formats for the Y-STR haplotype profiles: CMF format defined by FBI (version 3.2 or later) or Excel file (.xlsx).

Common Message Format (CMF) file

The CMF file is a specific XML file, which is defined by the FBI CODIS system. More details may be found by <https://www.fbi.gov/file-repository/codis-rapid-import-cmf-interface-specification.pdf>. An example CMF file is provided together with the software.

Excel file

The format of the Excel file that contains the Y-STR haplotypes is as follows:

1. The first row contains the "SampleName" and the marker names; one cell for one marker name
2. The followed rows contain the samples; one row for one sample;
3. The first column has to be the sample name
4. The followed columns contain the genotypes of the Y-STR markers; one cell in each column for one genotype of one specific sample
5. An empty cell means missing data

The following table is an example. An example Excel file is also provided with the software.

SampleName	DYS389 I	DYS389 II	DYS390
AA-1	13	28	22
AA-2	14	31	
AA-3	14	32	21

Y-STR mutation rates

The mutation rates file is a simple text file. The file contains two columns, and these two columns are separated by tab.

The first column contains the marker name. The marker name may contain space. The marker names with space and without space are different. For example, "DYS389 I" is different from "DYS389I". Please make sure the marker names in the mutation file and the haplotype file are consistent. Different software may export different Y-STR marker names, which may lead to errors in the subsequent analysis. Specifically, please check the space within the marker names, and check that "YGATAH4" is different from "GATAH4".

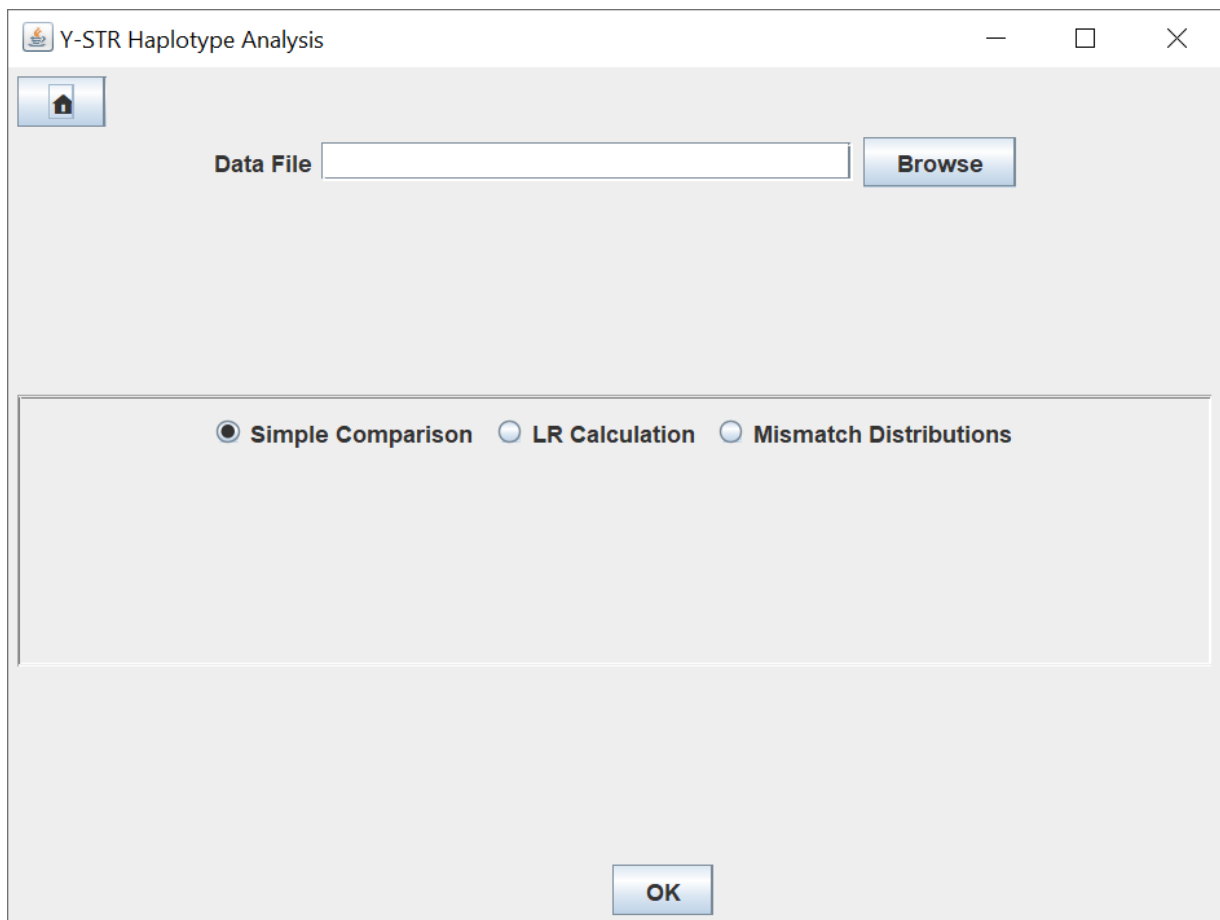
The second column contains the mutation rates of the Y-STR markers.

MPKin-YSTR has provided a mutation rate file "AllCommonYSTRMutationRates – nameWithSpace.txt". The mutation rates were obtained from YHRD (https://yhrd.org/pages/resources/mutation_rates) in March 2020.

Functions

MPKin-YSTR has three major functions: Simple Comparison, LR Calculation, and Mismatch Distributions.

First, users need to browse and upload a data file (i.e., Y-STR haplotype file). Second, select a function. Third, click “OK” button to move forward.



The screenshot shows a software window titled "Y-STR Haplotype Analysis". The window has a standard Windows-style title bar with minimize, maximize, and close buttons. Inside the window, there is a home icon in the top-left corner. Below it, the text "Data File" is followed by a text input field and a "Browse" button. In the center of the window, there are three radio buttons for selecting a function: "Simple Comparison" (which is selected), "LR Calculation", and "Mismatch Distributions". At the bottom center of the window is an "OK" button.

Profile Comparisons

MPKin-YSTR allows comparing Y-STR profiles in 1:1 (i.e., one profile vs. another profile), 1:N (i.e., one profile vs. multiple profiles), and N:N (i.e., pairwise comparisons with a profile set).

1:1 → Somebody vs. Somebody

1:N → Everybody vs. Somebody, or vice verse

N:N → Everybody vs. Everybody

Two thresholds were defined to screen the profile comparison results: (1) the maximum number of mismatched loci (e.g., 3) and (2) the maximum number of mismatched steps (e.g., 5). The software only displays the profile pairs passing the screening thresholds.

The screenshot shows a Windows-style dialog box titled "Who do you want to compare?". It has a standard window frame with minimize, maximize, and close buttons. Inside the dialog, there is a home icon in the top-left corner. The main content area is divided into two sections: "Who?" and "Whom?". Each section contains three radio button options: "Everybody" (which is selected), "Somebody", and "External". Below these sections are two input fields for thresholds: "Maximum Mismatched Loci" with a value of 9, and "Maximum Total Mismatched Steps" with a value of 16. At the bottom of the dialog are two buttons: "Cancel" on the left and "OK" on the right.

Users may also upload an additional data profile by the “External” option. One profile in the additional uploaded file should be decided in comparison.

The screenshot shows a dialog box titled "Who do you want to compare?". It contains two sections: "Who?" and "Whom?".

Who? Section:

- Radio buttons: ☐ Everybody, ☐ Somebody, ☒ External.
- A "Browse" button is located to the right of the radio buttons.
- Below the radio buttons, the file path "C:\Software\Test population.xlsx" is displayed.
- A dropdown menu below the file path shows "AA-10".

Whom? Section:

- Radio buttons: ☒ Everybody, ☐ Somebody, ☐ External.

Comparison Parameters:

- "Maximum Mismatched Loci" is set to 3.
- "Maximum Total Mismatched Steps" is set to 5.

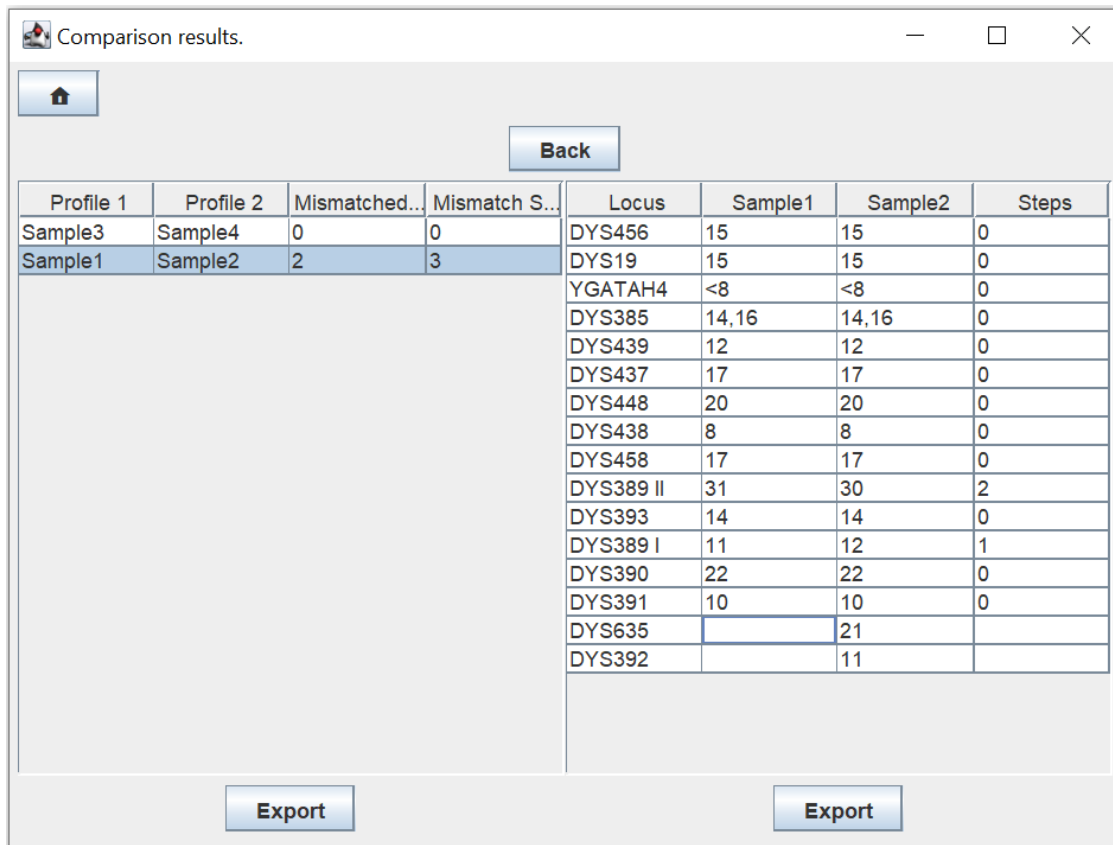
At the bottom, there are "Cancel" and "OK" buttons.

The comparison results will show all profile pairs passing the screening thresholds.

The left table shows the profile pairs with the numbers of mismatched loci and mismatched steps.

The right table shows the detailed comparison of a selected pair in the left table.

Users may export the tables as TSV (Tab-Separated Values) files.



Comparison results.

Back

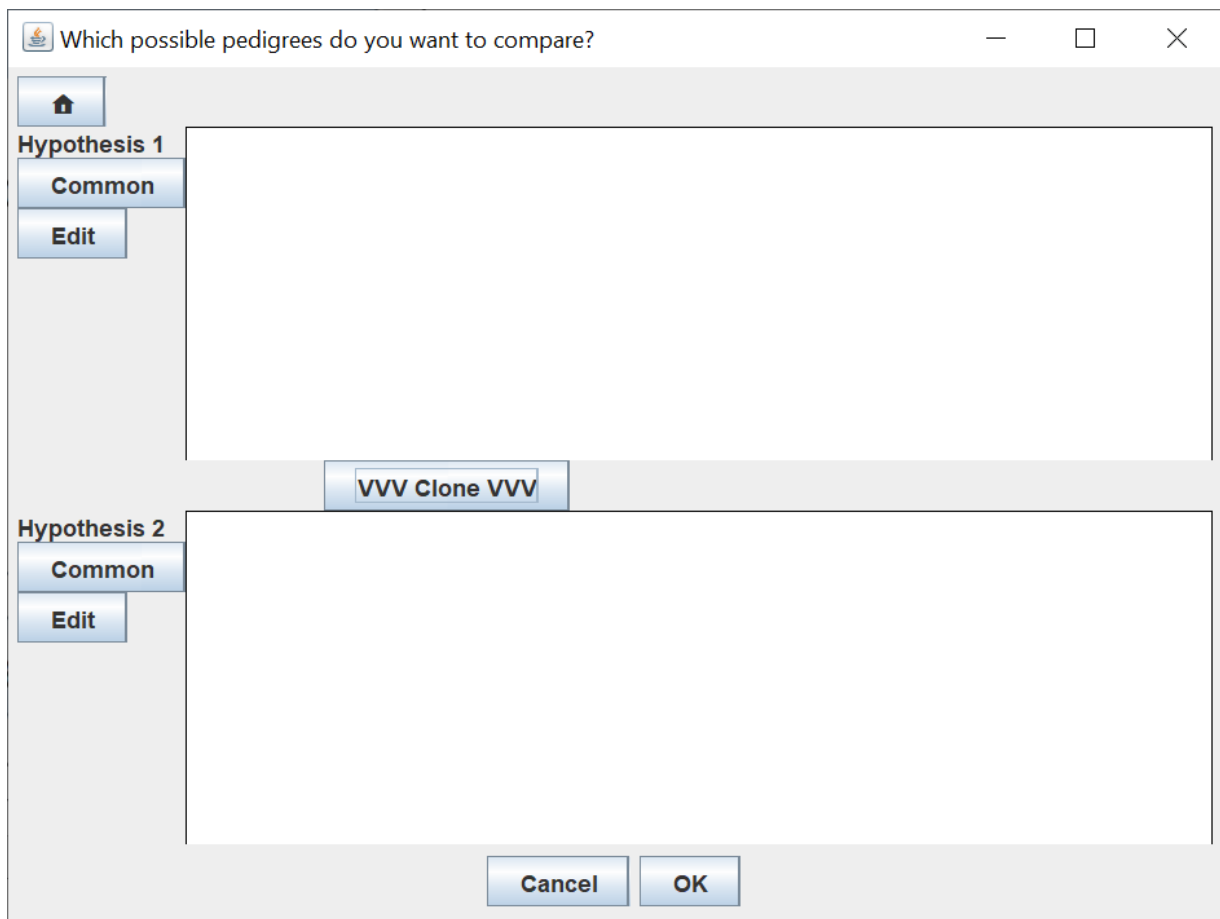
Profile 1	Profile 2	Mismatched...	Mismatch S...
Sample3	Sample4	0	0
Sample1	Sample2	2	3

Locus	Sample1	Sample2	Steps
DYS456	15	15	0
DYS19	15	15	0
YGATAH4	<8	<8	0
DYS385	14,16	14,16	0
DYS439	12	12	0
DYS437	17	17	0
DYS448	20	20	0
DYS438	8	8	0
DYS458	17	17	0
DYS389 II	31	30	2
DYS393	14	14	0
DYS389 I	11	12	1
DYS390	22	22	0
DYS391	10	10	0
DYS635		21	
DYS392		11	

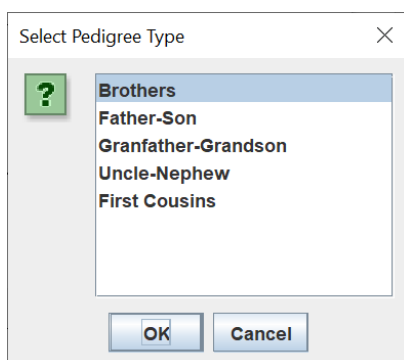
Export Export

Pedigree Likelihood Ratios

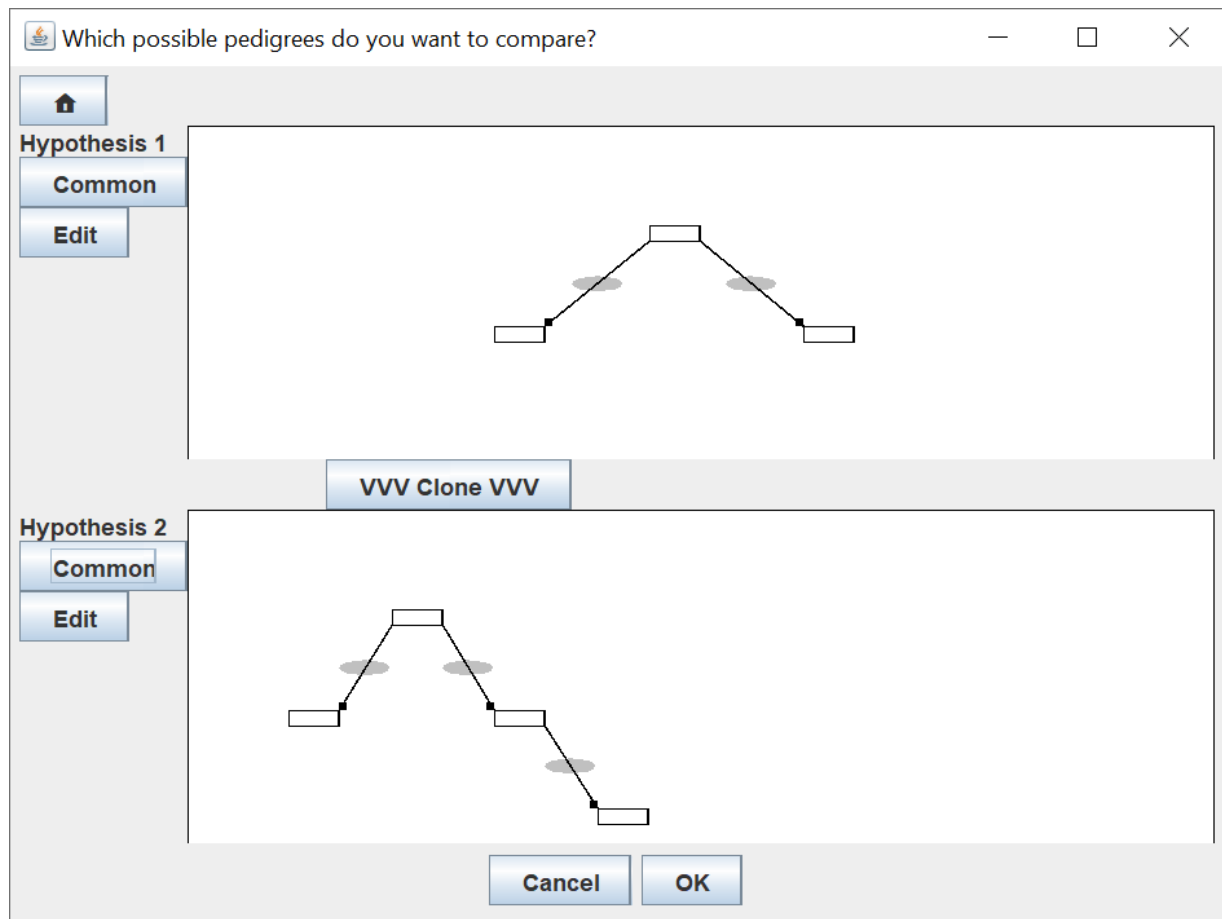
MPKin-YSTR allows the users to draw two competing pedigree hypotheses, assign Y-STR haplotypes to the individuals in the pedigrees, and calculate the likelihood of each hypothesis, as well as the likelihood ratio (LR).



Some common pedigrees have been provided to facilitate the pedigree drawing.

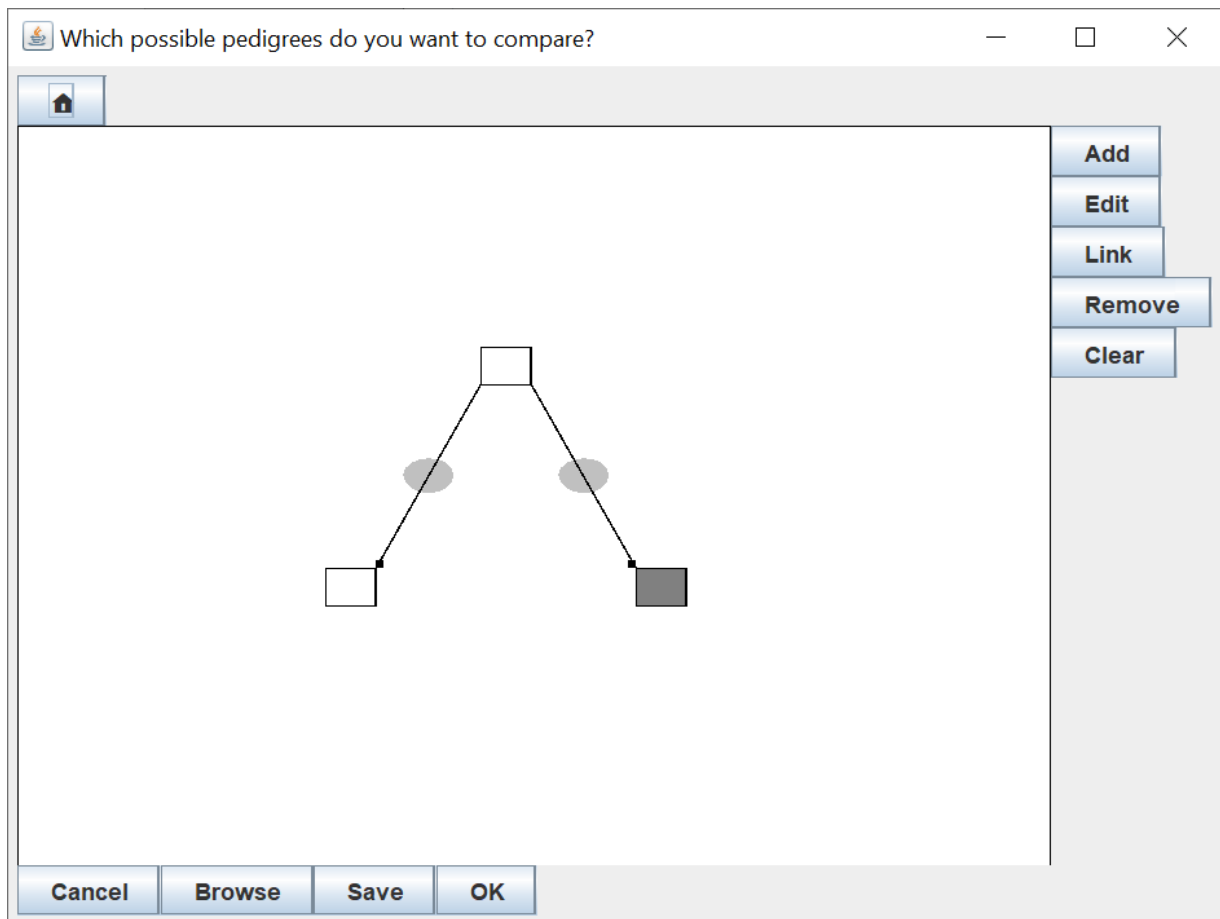


An example of two competing pedigree hypotheses.

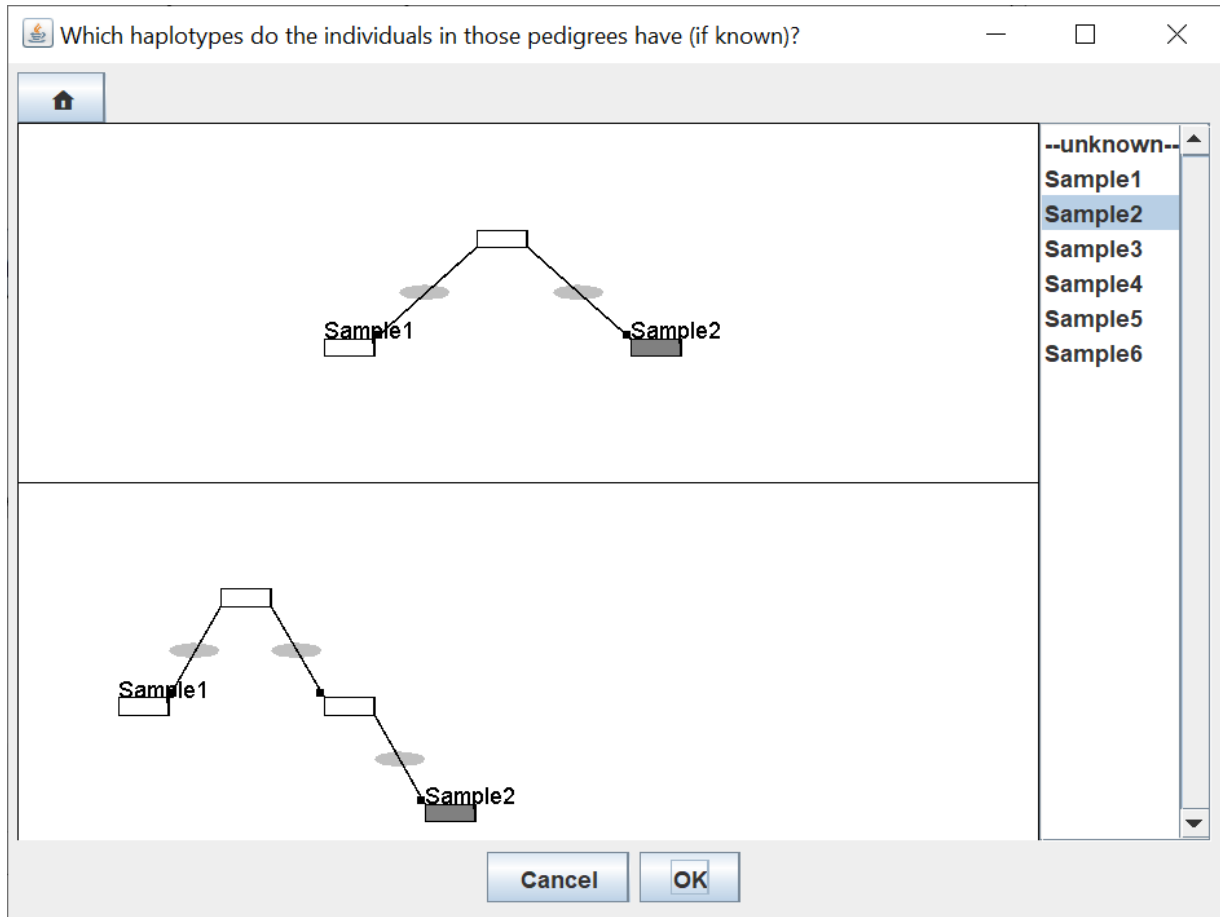


Users may define a pedigree by adding, editing, linking, or removing individual(s).

Users may also save a defined pedigree or load a saved pedigree.



To assign the Y-STR haplotypes in the uploaded Y-STR haplotype file to the individuals in the pedigrees, the users need to click an individual first, and then click a profile on the right panel.



Users may manually enter the haplotype frequencies of all the haplotypes in the pedigrees.

To facilitate the haplotype frequency entering, MPKin-YSTR allows the users to export all haplotypes in the pedigrees to a CSV file, which can be further uploaded to YHRD to obtain the frequencies. Since YHRD does not provide web API and multiple haplotype frequencies may be obtained in a YHRD search (for different populations, frequency calculation methods, etc.), the users have to manually enter the frequencies after obtaining the frequencies from YHRD.

Users are required to upload a Y-STR mutation rate file, which has been explained above.

How frequent are those haplotypes (and mutations between them)?

Export (YHRD)

Haplotype	Frequency
Sample2	1.0E-4
Sample1	2.0E-4

Mutation Rates: ates - NameWithSpace.txt **Browse**

Cancel **OK**

MPKin-YSTR calculates the likelihoods, the number of mismatched loci, and the number of the mismatched steps for both pedigrees. In addition, the LR will be calculated.

The users may export the results of the calculation, which will include all detail information, such as the pedigrees, the Y-STR haplotypes in the pedigrees, haplotype frequencies, mutation rates, and the calculated mismatched steps of all markers, etc. In addition, the users may add notes in the exported file.

Relative likelihood between the hypothesized pedigrees.

```
graph TD; A[ ] --- B[Sample1]; A --- C[Sample2];
```

```
graph TD; A[ ] --- B[Sample1]; B --- C[ ]; C --- D[Sample2];
```

H1 Likelihood: 1.8384E-4
H1 Mismatched Loci: 0
H1 Mismatch Steps: 0
H2 Likelihood: 8.7277E-5
H2 Mismatched Loci: 0
H2 Mismatch Steps: 0
LR: 2.1064E0
Notes
An example

Export

Back

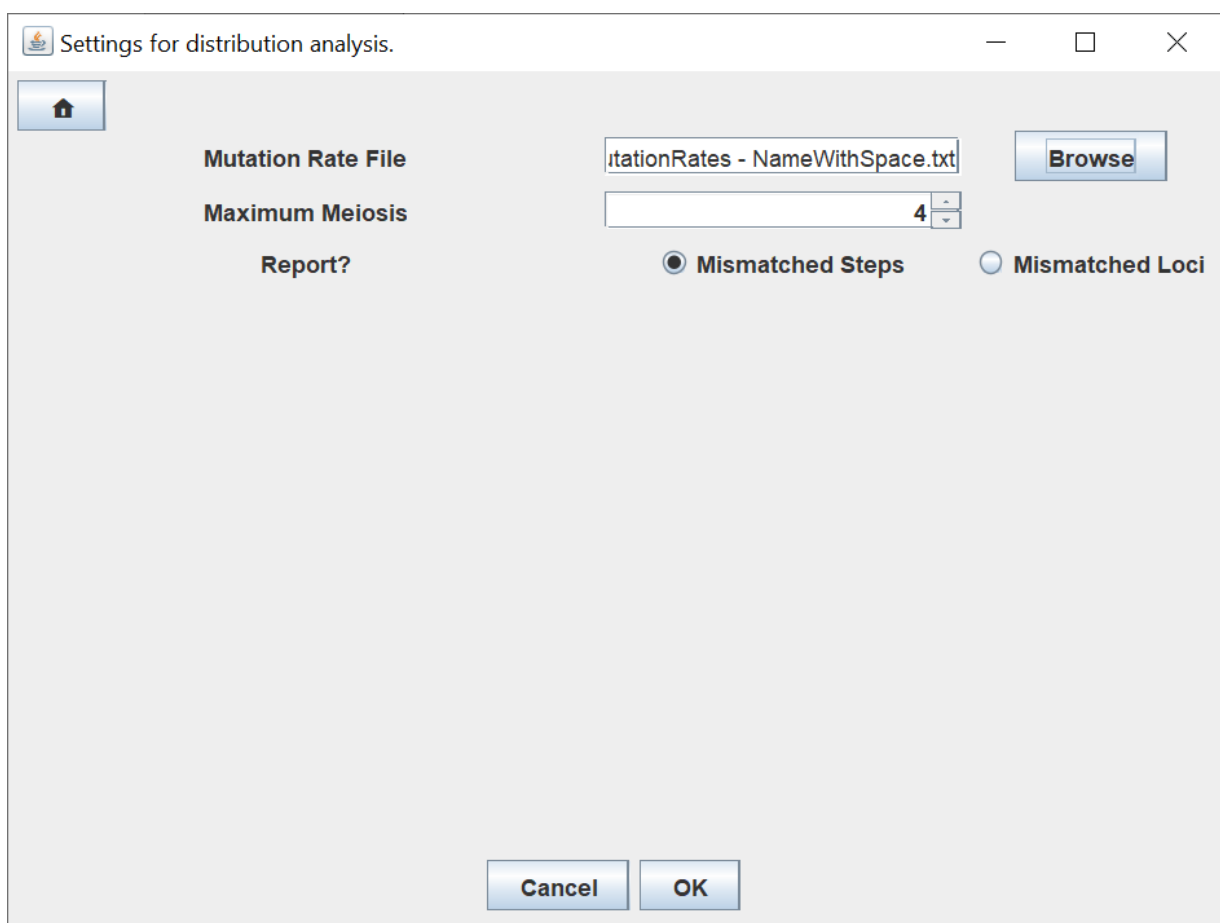
13

Mismatch distributions

This function draws the mismatch distributions, either loci or steps, for unrelated and related pairs. Ultimately, it is to assist the DNA analysts to quickly determining if there is strong support that two profiles are from the same lineage or not.

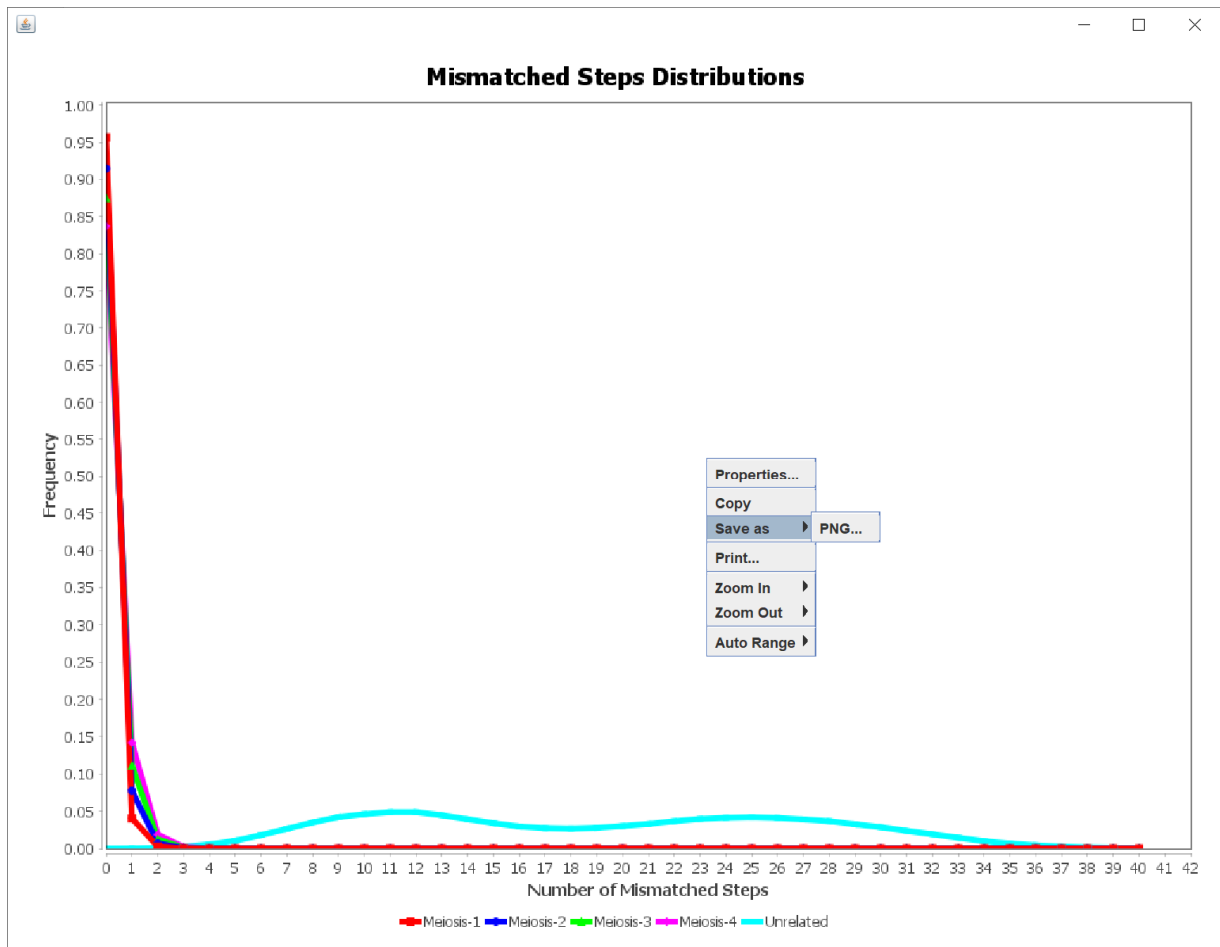
A mutation rate file is required. The maximum meiosis also needs to be decided, which is to define the scope of a male lineage. If the Maximum Meiosis = 4, a male lineage is defined as relatives up to the first-cousin relationship.

Meiosis-1 = father-son; Meiosis-2 = full-sibling; Meiosis-3 = uncle-nephew; Meiosis-4 = first-cousin.

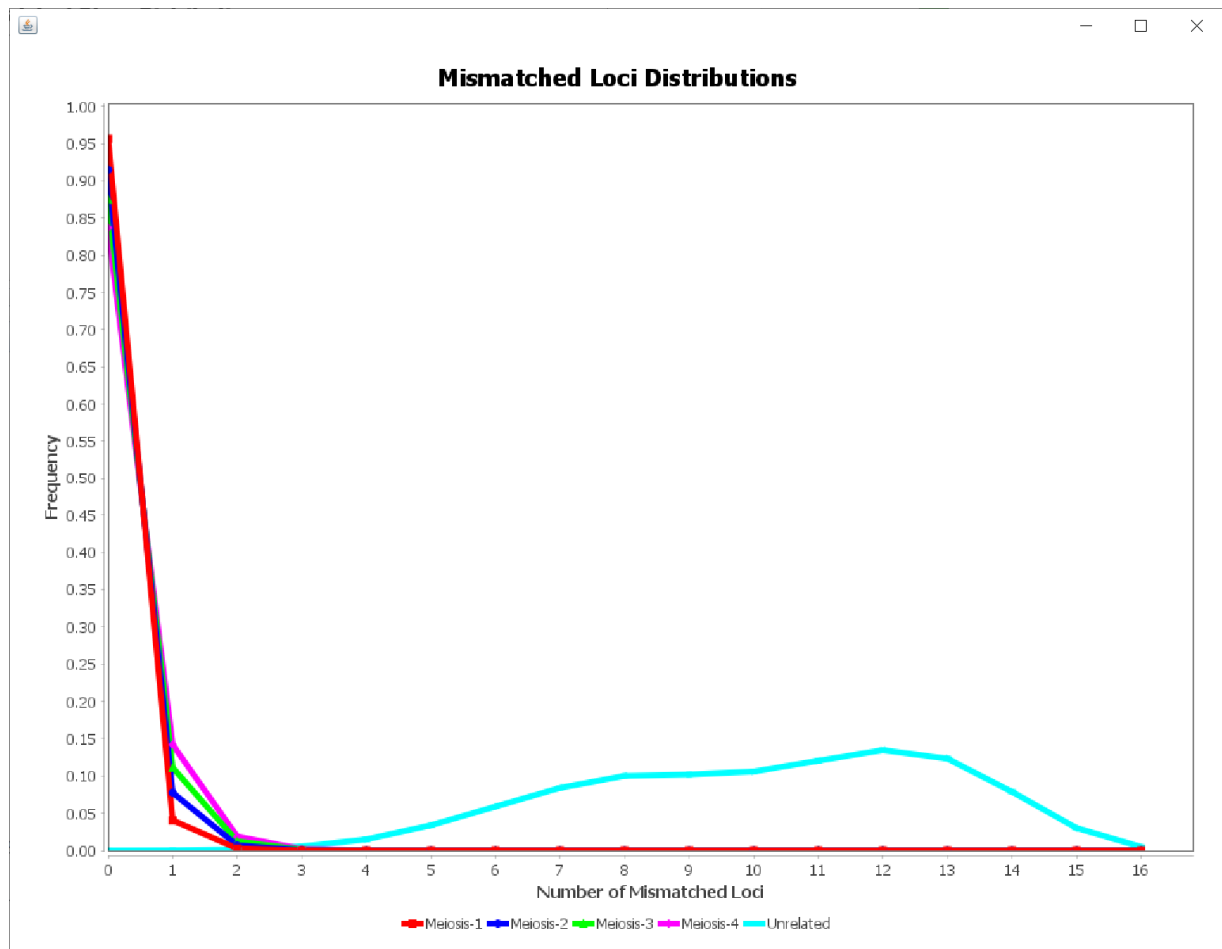


The screenshot shows a software window titled "Settings for distribution analysis." with standard Windows window controls (minimize, maximize, close). Inside the window, there is a home icon in the top-left corner. The main area contains three settings: "Mutation Rate File" with a text box containing "utationRates - NameWithSpace.txt" and a "Browse" button; "Maximum Meiosis" with a spinner box set to "4"; and "Report?" with two radio buttons, "Mismatched Steps" (which is selected) and "Mismatched Loci". At the bottom of the window are "Cancel" and "OK" buttons.

A figure example of the mismatched steps distributions. The figure may be zoomed in or out and saved as a PNG file by clicking the right key of the mouse.

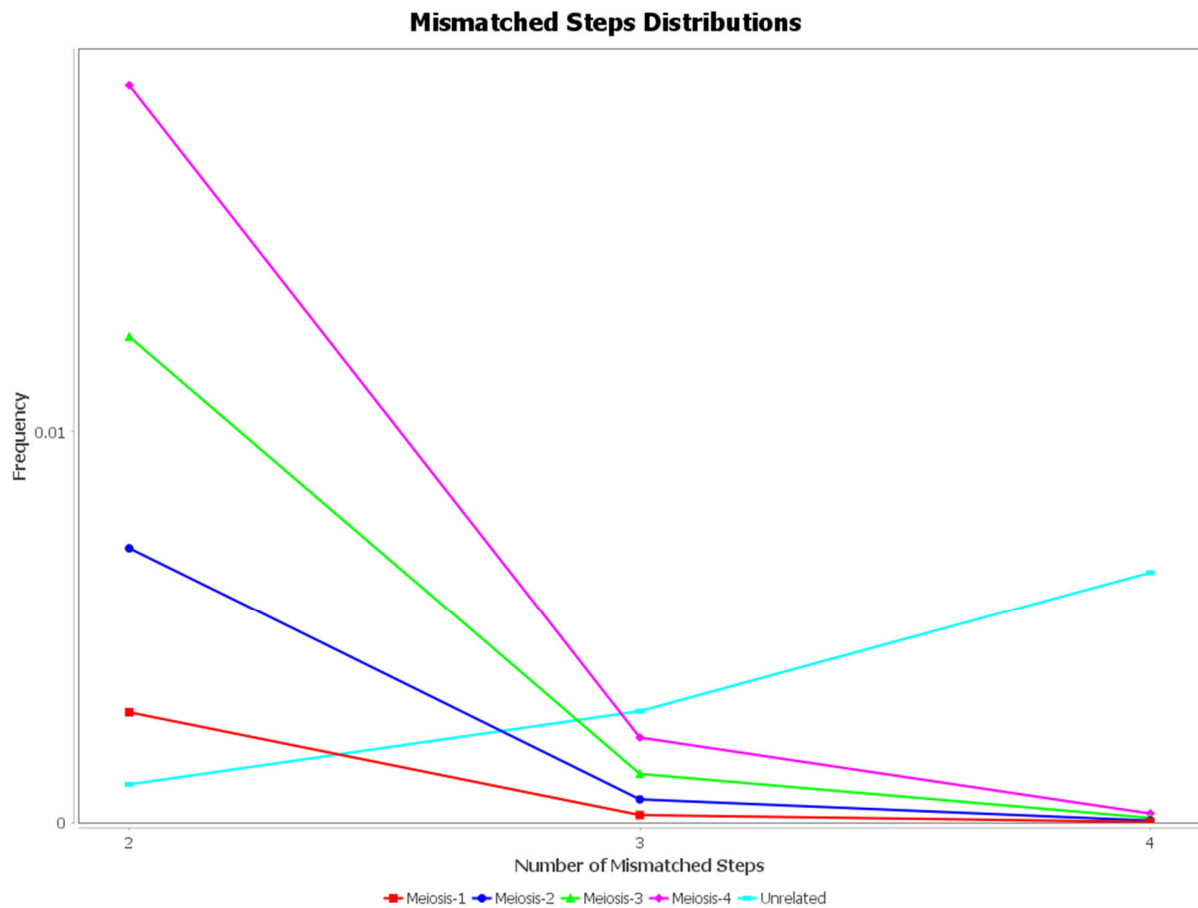


A figure example of the mismatched loci distributions.



Users may zoom in to the regions where the distribution of the unrelated pairs comes across the distributions of the related pairs, and then decide a threshold to decide paternal lineage.

In the example below, a threshold ≤ 2 mismatched steps would be good to conclude a male lineage.



Contacts

Questions may be sent to gejianye@gmail.com or jianye.ge@unthsc.edu.

In case a bug or error is found, please share all information (except the identifications of the test profiles) to the MPKin-YSTR authors. Thus, the authors can quickly replicate the bug or error.