Federal Agency:	Office of Justice Programs, National Institute of Justice, Department of Justice
Federal Grant:	Award Number: 2020-DQ-BX-0018
Project Title:	Interpretation of Y chromosome STRs for missing persons cases
PI:	Jianye Ge, Associate Professor & Associate Director, Center for Human Identification, University of North Texas Health Science Center at Fort Worth, Center for Human Identification, Fort Worth, TX 76107 Tel: 817-735-2477; email: Jianye.Ge@unthsc.edu
Submission Date:	June 30, 2022
DUNS Number:	
EIN:	
Recipient Organization:	University of North Texas Health Science Center 3500 Camp Bowie Blvd., Fort Worth, TX 76107
Institutional Profile Number:	
Project/Grant Period:	01-01-2020 – 06-30-2021
Report Frequency:	Final Technical Report
Signature:	Jianye Ge, Ph.D.

Introduction

The human Y chromosome is inherited through the paternal line. It contains a good number of Short Tandem Repeat (STR) loci throughout the non-recombining region of the chromosome, and a set of these STRs was adopted to characterize the lineage of male individuals. Y-STR haplotypes have been used to assist in various forensic investigations, such as rape cases, other violent crimes, kinship in missing persons cases, and identifications in mass disasters [1-21]. Y-STR haplotypes, typed with various commercial testing kits, consist of specific sets of Y-STR loci and have been recorded in databases, such as U.S. Y-STR (out of service in 2019) [18] and Y Chromosome Haplotype Reference Database (YHRD) [19] for use in casework. There are also Y-STR haplotype data that are publicly accessible, such as those in the scientific litrature [22, 23]. These data can be compiled, analyzed, and used to support forensic applications. Based on these databases, Y-STR haplotype frequencies can be estimated for identification, kinship analysis, and mixture analysis purposes [12, 17, 24].

The primary use of Y-STR haplotypes is to determine the relative high probability of observing two haplotypes from individuals of the same lineage as opposed to from individuals of different lineages [12], particularly for samples from the sexual assault cases, as well as the Y-STR screening in familial searching [21, 25]. A comparison is typically based on a direct comparison of an evidence profile with a reference profile, and if a “match” is obtained, the analysis is then followed up by searching a relevant population database (e.g., YHRD) containing a large number of reference profiles to estimate a haplotype frequency. The SWGDAM Lineage Marker Committee has recommended methods of estimating Y-STR haplotype frequencies [26] using the counting method followed by applying a 95% upper confidence limit of the counting estimate and/or calculating a Y-STR match probability using a population substructure correction. The YHRD provides capabilities to perform the SWGDAM recommended haplotype frequency estimations.

However, the SWGDAM interpretation guidelines [26] do not address scenarios in which two (or more) mismatched profiles are compared. For example, an alleged father and a male child have a one locus mismatch with a two-step repeat difference and paternity is questioned. For many missing person cases, multiple male family members may be available as references, and these

individuals may have different Y-STR haplotypes due to mutation or false claimed relationships. In such complex cases, more sophisticated interpretation methods and guidelines are needed. The SWGDAM Guidelines for Missing Persons Casework [27] recommended using a likelihood ratio approach for interpreting Y-STR haplotypes, but no detailed method was described.

Studies [20, 28] have developed computational methods and interpretation guidelines specifically addressing this issue. In Ge et al. [20], a more general likelihood ratio (LR) based approach for Y-STR haplotype relationship estimation was developed. Similar to the pedigree likelihood ratio (PLR) approach applied to autosomal loci [29, 30], the PLR method [20] for lineage markers (both Y-STR and mitochondrial DNA haplotypes) was developed to evaluate the relationship among multiple individuals in the same pedigree with a likelihood ratio approach. This method requires alternate hypotheses with fixed pedigree relationships for the evidence profiles. A pedigree likelihood given each hypothesis can be calculated with family structure and mutation incorporated, and a LR can be obtained by comparing the likelihoods from the alternate hypotheses.

In practice, the pedigree structure in the hypotheses may not always be available (i.e., two individuals are hypothesized to be related, but the exact relationship may not be known), and the haplotype frequencies may not be precisely estimated (e.g., due to limited number of profiles of the relevant population in the database or uncertain ethnicity). For such scenarios, a simple method [28], based on the distributions of the mismatched loci or steps between unrelated and related haplotype pairs, was developed to quickly determine if there is strong support that two profiles are from the same lineage or not. Thresholds can be determined to minimize false interpretations using the mismatch distributions for concluding Y-STR haplotype pairs as related or unrelated (e.g., two Yfiler Plus profiles with ≤5 mismatched steps are more likely from the same lineage). This method is very simple and also is accurate (e.g., >99.975% accuracy for two unrelated pairs determined as unrelated; >99.9999% of the father-son pairs have ≤ 5 mismatched steps) [28, 31]. This study [28] was done with a Chinese population for Yfiler and Yfiler Plus kits. A different population or kit may have a different threshold to minimize false interpretations due to different Y-STRs in a kit and/or haplotype frequencies in the reference populations.

These methods and interpretation guidelines have not been adopted by forensic laboratories, likely because the guidelines and tools necessary to perform the analysis are not readily available. In this study, a software program, MPKIN-YSTR, was developed with a user friendly graphic interface to facilitate the Y-STR haplotype interpretations. This tool includes three major functions, Y-STR profile comparisons (i.e., 1:1, 1:N, and N:N comparisons), PLR for kinship analysis with Y-STR profiles, and drawing mismatch distributions to select a threshold for determining if two profiles are from the same lineage or not. The methods described in [20, 28] were implemented in this software.

Methods

Y-STR population data

The haplotype frequencies and mutation rates of the Y-STR loci are the foundations for kinship analysis of Y-STR haplotypes. YHRD [19], as the largest publicly accessible Y-STR database, allows users to upload a Y-STR haplotype, search it against the database, and return haplotype frequency estimations based on configurable population groupings. The YHRD also summarizes the mutation rates of the Y-STR loci in the major commercial kits (https://yhrd.org/pages/resources/mutation_rates). Population specific mutation rates also are available through publications, such as Ge et al. [32].

Y-STR haplotypes from various populations and commercial kits are also available from scientific literature, including Budowle et al. [22], Gopinath et al. [33], and Purps et al. [34]. Table 1 shows the number of profiles that were collected. All these Y-STR profiles were used to test the software and develop Y-STR interpretation guidelines for missing persons cases. The profiles with null alleles were removed from the data, because the number of steps between null alleles and any other alleles can not be calculated when developing the Y-STR interpretation guidelines following the methods described in Liu et al. (2016). In addition, estimating the distributions of the number of mismatched loci or steps for unrelated pairs requires unrelated profiles. Thus, duplicated profiles within a study were removed to minimize the impact of related pairs in the sampled profiles.

Table 1. The collection of Y-STR profiles for developing interpretation guidelines and testing the interpretation software program.

Study	Kit	Population	Subpopulation	Sample size*	Sample size without duplicates
Budowle et al. [22]	Yfiler	Texas	African	931	887
			Caucasian	946	914
			Hispanic	982	889
Gopinath et al. [33]	Yfiler Plus	USA	African	256	256
			Asian	255	240
			Caucasian	237	237
			Hispanic	185	182
Purps et al. [34]	Y23	USA	African	672	666
			Asian	154	154
			Caucasian	718	716
			Hispanic	513	510
			Native American	493	402

* Samples with null alleles were not included in this table and subsequent analyses.

Y-STR 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) manners. The number of mismatched loci and steps can be calculated between a pair of Y-STR profiles, where the number of mismatched steps is defined as the minimum number of mutation steps between two sets of alleles. 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 established screening thresholds.

Pedigree likelihood ratio

The PLR method implemented in this software has been described in Ge et al. [20]. The general principle of the LR calculations on lineage markers is the same as that for autosomal markers, which compares the likelihoods of Y-STR haplotypes given two competing hypotheses. For missing persons cases, the hypotheses could be the hypothesis of kinship H_k: the questioned person (Q) is a specific member of the family (F), or the hypothesis of non-kinship or unrelated H_{nk}: Q is unrelated to F. The LR is represented by the following expression:

$$LR = \frac{\Pr(Q, F \mid H_k)}{\Pr(Q, F \mid H_{nk})} = \frac{\Pr(Q, F \mid H_k)}{\Pr(Q) * \Pr(F)} \quad (1)$$

where $\Pr(Q)$ is the haplotype frequency of Q, $\Pr(F)$ is the pedigree likelihood of F, and $\Pr(Q, F \mid H_k)$ is the pedigree likelihood of both Q and F given H_k. Note that H_k posits a specific relationship: it is not "Q belongs to the lineage of F", where the position of Q in F is left unspecified.

Given there are no missing data (i.e., the Y-STR profile of every individual in the pedigree is available), the pedigree likelihood is the product of the haplotype frequency of the founder (i.e., the person without antecedent relatives in the pedigree) and the transmission probability from the founder(s) to all descendants, which is dependent on the mutation steps between each father-son transmission event. The transmission probability is the product of the transmission probabilities of all father-son pairs in the pedigree. The transmission probability of each father-son pair may be calculated by a mutation model. In this software, the Two-Phase mutation model [29, 35, 36] was used.

If any individual in the pedigree is untyped, the Y-STR profile of the individual needs to be inferred before computing the PLR. In theory, any allele in a defined population could be the allele of the untyped individual at a specific locus. In practice, not all allele combinations could be a real haplotype in the population, and very distant alleles from the typed allele of the related individuals are unlikely (e.g., allele 20 of a son may be unlikely if the father has an allele 10 at a locus). Thus, only the haplotypes from the other typed individuals in the same pedigree would be assigned to be the possible haplotypes of the untyped founders, but the other untyped individuals were allowed to have alleles other than the existing ones with a predefined maximum step (e.g., 1). Thus, an untyped individual may have one of many possible haplotypes. A recursive algorithm was used to assign possible haplotypes to the untyped individuals and generate complete sub-pedigrees with no missing data. For example, if two individuals in the pedigree are untyped and each has 3 possible haplotypes, 9 complete sub-pedigrees would be generated. The PLR would be calculated for each complete sub-pedigree, and the final PLR would be the average of the PLRs for each complete sub-pedigree.

Mismatch distribution and threshold determination

Following the same methods described in Liu et al. [28], the distributions of the numbers of mismatched loci and steps given various relationships were calculated to determine the thresholds to decide if two profiles are from the same lineage or not. The distributions of mismatched steps for unrelated pairs were generated by pairwise comparison of all unrelated haplotypes given the population haplotype data. The distributions of mismatched steps for related pairs, including 1 meiosis (parent-child), 2 meioses (full-sibling), 3 meioses (uncle-nephew), and 4 meioses (first cousin), were generated by simulations with the Two-Phase mutation model and mutation rates from the YHRD. By combining the distributions from related and unrelated pairs, interpretation guidelines to determine lineage based on mismatched loci or steps were generated for given relationships, populations, and Y-STR haplotypes, and the accompanying error rates (false positive and false negatives) also were estimated.

Software development and running environment

MPKin-YSTR was developed with Java JDK 1.8.0. JFreeChart Java library (version 1.5.3) (www.jfree.org) was used to plot the mismatch distributions. MPKin-YSTR can be run on any platform installed with the Java JRE 1.8.0 or higher versions.

Results

MPKin-YSTR includes three major functions, Y-STR profile comparisons (i.e., 1:1, 1:N, and N:N comparisons), the calculation of the PLR for kinship analysis with Y-STR profiles, and drawing mismatch distributions to decide a threshold for determining if two profiles are from the same lineage or not. The methods described in [20, 28] were implemented in this software.

Y-STR profile comparisons

MPKin-YSTR provides detailed information for all profile pairs passing the screening thresholds, including genotypes of the profiles and the number of mismatched steps between the pairs. This detailed information can be exported into tab-separated values (i.e., TSV) files as reports. Figure S1 in the supplementary material shows one example of the profile comparison results.

Pedigree likelihood ratio

MPKin-YSTR allows users to draw pedigrees for two competing hypotheses. With haplotypes and haplotype frequencies assigned to the pedigrees, the likelihood of each pedigree, as well as the likelihood ratio that compares two competing likelihoods of the pedigrees, can be calculated. All details can be exported into Excel files as reports. To facilitate the YHRD database search, this software provides the function to export the Y-STR profiles in the pedigrees into a CSV file, which can be uploaded to YHRD as an input file to obtain the haplotype frequencies.

For scenarios where the relationships among the profiles are uncertain, multiple pedigrees can be drawn, and the corresponding PLRs can be calculated. Table 2 gives an example of determining the relationship between two Yfiler profiles, in which there is a mismatch at the DYS458 locus. By comparing the PLRs of the common relationships, the grandfather-grandson was the most likely relationship with the highest PLR (Table 3).

Table 2. An example used to determine the relationship between two Y-STR profiles. There is a mismatch between these two profiles at the DYS458 locus.

Marker	Sample1	Sample2
DYS389 I	13	13
DYS389 II	31	31
DYS390	21	21
DYS456	15	15
DYS19	15	15
DYS385	16,17	16,17
DYS458	17	16
DYS437	14	14
DYS438	11	11
DYS448	21	21
YGATAH4	12	12
DYS391	10	10
DYS392	11	11
DYS393	13	13
DYS439	11	11
DYS635	21	21
Haplotype Frequency	0.0004	0.0008

Table 3. The likelihood ratios between two Y-STR profiles (see details in Table 2) given various relationships.

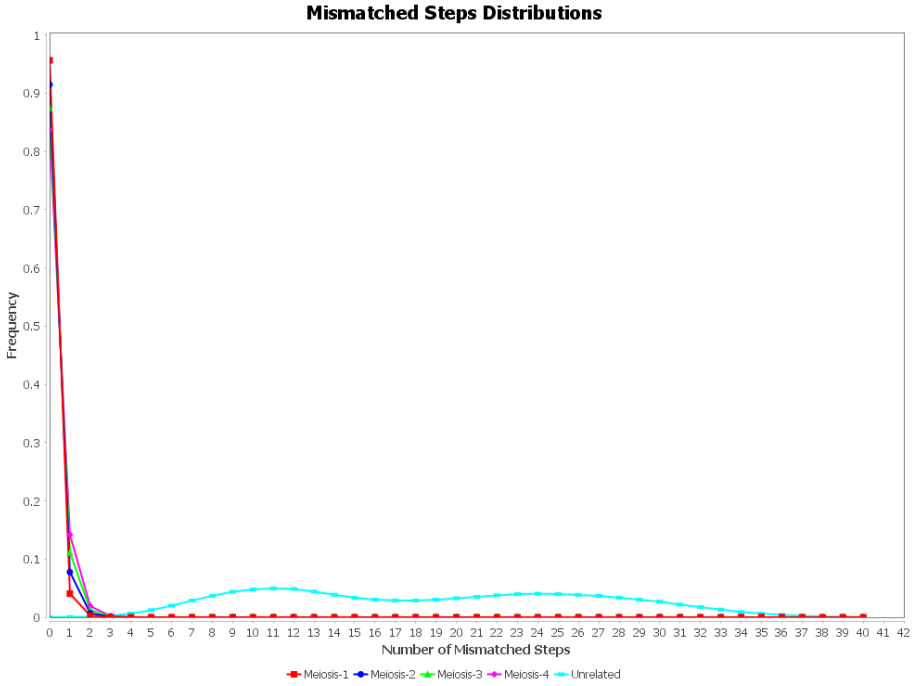
Relationship	PLR
Father	3.63E+00
Brother	5.19E+00
Uncle	8.26E+00
Grandfather	1.38E+01
Cousin	9.45E+00

Thresholds to determine lineage

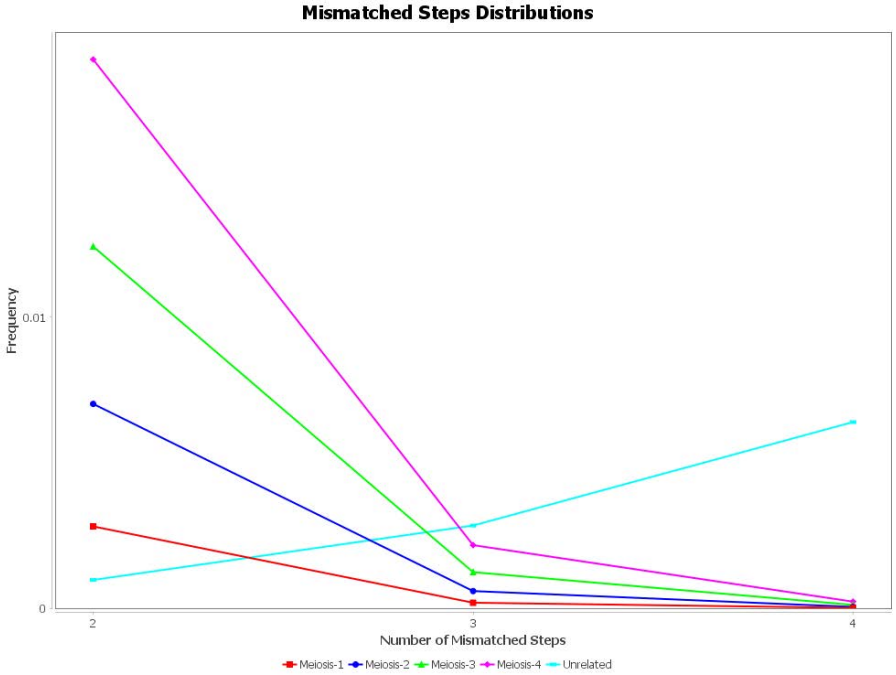
In general, the distributions of mismatched steps, compared with the distributions of mismatched loci, may be more appropriate to determine the lineage threshold, as the mismatched steps include more detailed information. The mismatch distributions can vary for different relationships, populations, and Y-STR marker sets, and the accompanying error rates. Following the same method as in Liu et al. [28], the distributions of mismatched steps for unrelated pairs were generated by pairwise comparisons of all unrelated haplotypes collected for the US major populations (e.g., Caucasian, African American, Hispanics) and the total population. The distributions of mismatched steps for related pairs up to 4 meioses also were generated by a dynamic programming approach [28] with the Two-Phase mutation model using the YHRD mutation rates. By combining the distributions from related and unrelated pairs, the thresholds to determine lineage based on mismatched steps were generated for given relationships, populations, and Y-STR haplotypes.

Using the data in Budowle et al. [22] as an example, Figure 1 shows the distributions of the number of mismatched steps between related and unrelated Yfiler profile pairs. The full distributions for African American profiles are shown in Figure 1.a, and Figure 1.b-f only shows the part of the distributions where the distributions for related and unrelated pairs intersect. For all three tested populations, as well as the total population, the related and unrelated distributions intersect between 2 and 3 mismatched steps. Based on these data, two Yfiler profiles with $\leq$2 mismatched steps are more likely from the same lineage (if a male lineage is limited to a maximum of 4 meioses); otherwise, they are more likely from two different lineages. This threshold is consistent with the results in Liu et al. [28].

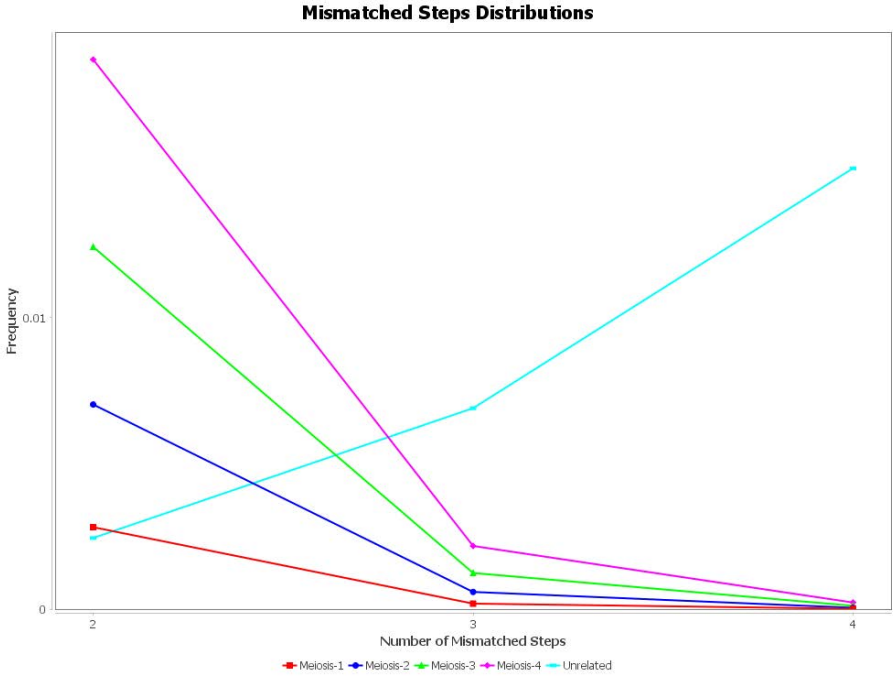
(a) African Americans (number of mismatched steps from 0 to 40)



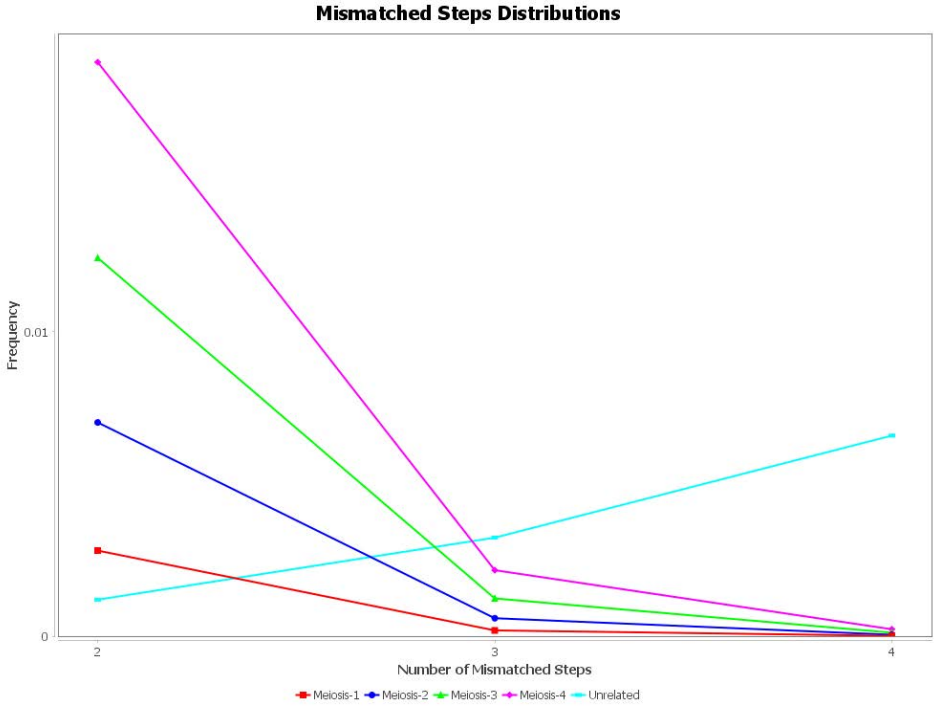
(b) African Americans (number of mismatched steps from 2 to 4)



(c) Caucasians (number of mismatched steps from 2 to 4)



(d) Hispanics (number of mismatched steps from 2 to 4)



(e) Total population (number of mismatched steps from 2 to 4)

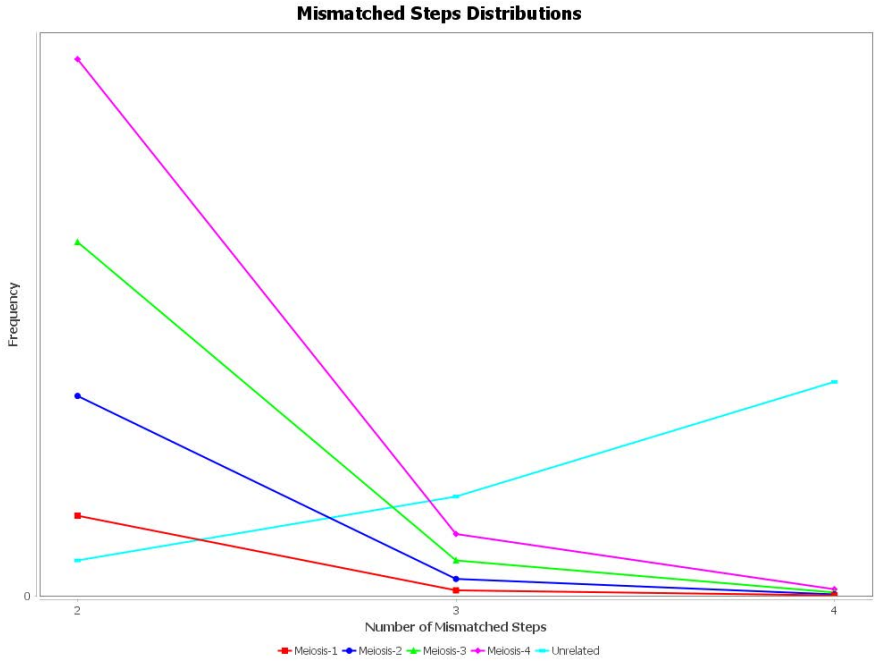


Figure 1. The distributions of the number of mismatched steps between related and unrelated Yfiler profile pairs for African, Caucasian, Hispanic, and total populations in Budowle et al. [22]. Meiosis-1 = father-son; Meiosis-2 = full-sibling; Meiosis-3 = uncle-nephew; Meiosis-4 = first-cousin.

The distributions of the mismatched steps for Y23 and Yfiler Plus marker sets were also plotted (Figures S3-S11 in the supplementary materials). The distributions of the unrelated were below the distributions of the related for the number of steps ≤4 and ≤5 across all populations with Y23 and Yfiler Plus, respectively. Thus, it may be appropriate to set the thresholds at ≤4 and ≤5 for Y23 and Yfiler Plus, respectively, to minimize false negative errors (i.e., to minimize the chance of missing potential true relatives). The thresholds could be lowered if the false positive errors (i.e., include unrelated as related) need to be balanced. Apparently, with more markers included, the chance of mutation between relatives increases, and thus the corresponding thresholds increases. Due to the limited number of samples, these distributions were not as smooth as those in Liu et al. [28], which could be improved by adding more samples.

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Summary

The Missing Persons Unit under the Center of Human Identification at the University of North Texas Health Science Center (UNTCHI) specializes in the DNA analysis and identification of missing persons cases and has processed a large proportion of the unidentified human remains of the missing persons cases in the US . It is common within UNTCHI to encounter some complex cases (i.e., those that are focused on identification of an unknown male) that would be better served with enhanced Y-STR interpretation procedures. In this study, a software program, MPKin-YSTR, was developed to facilitate the Y-STR haplotype analysis, particularly for complex missing persons cases, by implementing more sophisticated methods. This tool should be able to improve the interpretation of complex cases with Y-STR haplotype evidence, which are relatively common among forensic DNA laboratories employing Y-STR typing. Thus, more biological evidence can be interpreted, which in turn can result in more investigation leads to help solve crimes.

Data achieving

The software MPKin-YSTR is available on GitHub at https://github.com/Ge-Lab/MPKin-YSTR, which is freely accessible to the public. The data uploaded includes

1. README.MD to introduce the software
2. Tutorial videos: https://www.youtube.com/playlist?list=PLZ_kVTJ53UPB2LFr7o0HYA-3WXfL_F3Jr
3. A user manual
4. The runnable file MPKin-YSTR.jar, which also includes
5. The source codes of MPKin-YSTR
6. All population data used in testing. A CMF format input file used in testing

This study only used publicly accessible data, data obtained from the publications, or simulated data. All samples were de-identified. No informed consent form was used when collecting these data.

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Reference

[1] R. Chakraborty, J.M. Opitz, J.F. Reynolds, Paternity testing with genetic markers: Are Y-linked genes more efficient than autosomal ones?, American journal of medical genetics 21(2) (1985) 297-305.
[2] M.A. Jobling, A. Pandya, C. Tyler-Smith, The Y chromosome in forensic analysis and paternity testing, International journal of legal medicine 110(3) (1997) 118-124.

[3] D. Corach, L.F. Risso, M. Marino, G. Penacino, A. Sala, Routine Y-STR typing in forensic casework, Forensic science international 118(2-3) (2001) 131-135.

[4] L. Roewer, M. Kayser, P. de Knijff, K. Anslinger, A. Betz, A. Caglia, D. Corach, S. Füredi, L. Henke, M. Hidding, A new method for the evaluation of matches in non-recombining genomes: application to Y-chromosomal short tandem repeat (STR) haplotypes in European males, Forensic Science International 114(1) (2000) 31-43.

[5] M. Kayser, A. Sajantila, Mutations at Y-STR loci: implications for paternity testing and forensic analysis, Forensic science international 118(2-3) (2001) 116-121.

[6] B. Rolf, W. Keil, B. Brinkmann, L. Roewer, R. Fimmers, Paternity testing using Y-STR haplotypes: assigning a probability for paternity in cases of mutations, International journal of legal medicine 115(1) (2001) 12-15.

[7] P. Gill, C. Brenner, B. Brinkmann, B. Budowle, A. Carracedo, M.A. Jobling, P. de Knijff, M. Kayser, M. Krawczak, W. Mayr, DNA Commission of the International Society of Forensic Genetics: recommendations on forensic analysis using Y-chromosome STRs, International journal of legal medicine 114(6) (2001) 305-309.

[8] E. Avila, A.B. Felkl, P. Graebin, C.P. Nunes, C.S. Alho, Forensic characterization of Brazilian regional populations through massive parallel sequencing of 124 SNPs included in HID ion Ampliseq Identity Panel, Forensic Sci Int Genet 40 (2019) 74-84.

[9] L. Gusmão, J.M. Butler, A. Carracedo, P. Gill, M. Kayser, W.R. Mayr, N. Morling, M. Prinz, L. Roewer, C. Tyler-Smith, P.M. Schneider, DNA Commission of the International Society of Forensic Genetics (ISFG): An update of the recommendations on the use of Y-STRs in forensic analysis, Forensic Science International 157(2–3) (2006) 187-197.

[10] L. Roewer, Y chromosome STR typing in crime casework, Forensic science, medicine, and pathology 5(2) (2009) 77-84.

[11] A. Ambers, J. Votrubova, D. Vanek, A. Sajantila, B. Budowle, Improved Y-STR typing for disaster victim identification, missing persons investigations, and historical human skeletal remains, Int J Legal Med 132(6) (2018) 1545-1553.

[12] M. Kayser, Forensic use of Y-chromosome DNA: a general overview, Human genetics 136(5) (2017) 621-635.

[13] J.S. Buckleton, M. Krawczak, B.S. Weir, The interpretation of lineage markers in forensic DNA testing, Forensic Science International: Genetics 5(2) (2011) 78-83.

[14] C.H. Brenner, Understanding Y haplotype matching probability, Forensic Science International: Genetics 8(1) (2014) 233-243.

[15] A. Caliebe, M. Krawczak, Match probabilities for Y-chromosomal profiles: A paradigm shift, Forensic Science International: Genetics 37 (2018) 200-203.

[16] M.M. Andersen, D.J. Balding, Y-profile evidence: Close paternal relatives and mixtures, Forensic Sci. Int. Genet. 38 (2019) 48-53.

[17] L. Roewer, Y-chromosome short tandem repeats in forensics—Sexing, profiling, and matching male DNA, Wiley Interdisciplinary Reviews: Forensic Science 0(0) (2019) e1336.

[18] J. Ge, B. Budowle, J.V. Planz, A.J. Eisenberg, J. Ballantyne, R. Chakraborty, US forensic Y-chromosome short tandem repeats database, Legal medicine 12(6) (2010) 289-295.

[19] S. Willuweit, L. Roewer, The new Y chromosome haplotype reference database, Forensic Science International: Genetics 15 (2015) 43-48.

[20] J. Ge, A. Eisenberg, J. Yan, R. Chakraborty, B. Budowle, Pedigree likelihood ratio for lineage markers, International journal of legal medicine 125(4) (2011) 519-525.

[21] J. Ge, R. Chakraborty, A. Eisenberg, B. Budowle, Comparisons of familial DNA database searching strategies, J Forensic Sci 56(6) (2011) 1448-56.

[22] B. Budowle, J. Ge, X.G. Aranda, J.V. Planz, A.J. Eisenberg, R. Chakraborty, Texas population substructure and its impact on estimating the rarity of Y STR haplotypes from DNA evidence, Journal of forensic sciences 54(5) (2009) 1016-1021.

[23] B. Budowle, J. Ge, J. Low, C. Lai, W.H. Yee, G. Law, W.F. Tan, Y.M. Chang, R. Perumal, P.Y. Keat, N. Mizuno, K. Kasai, K. Sekiguchi, R. Chakraborty, The effects of Asian population substructure on Y STR forensic analyses, Leg Med (Tokyo) 11(2) (2009) 64-9.

[24] J. Ge, B. Budowle, R. Chakraborty, Interpreting Y chromosome STR haplotype mixture, Legal Medicine 12(3) (2010) 137-143.

[25] S.P. Myers, M.D. Timken, M.L. Piucci, G.A. Sims, M.A. Greenwald, J.J. Weigand, K.C. Konzak, M.R. Buoncristiani, Searching for first-degree familial relationships in California's offender DNA database: Validation of a likelihood ratio-based approach, Forensic Science International: Genetics 5(5) (2011) 493-500.

[26] SWGDAM, Interpretation Guidelines for Y-Chromosome STR Typing, (2014).

[27] SWGDAM, Guidelines for Missing Persons Casework, (2014).

[28] H. Liu, X. Li, J. Mulero, A. Carbonaro, M. Short, J. Ge, A convenient guideline to determine if two Y-STR profiles are from the same lineage, Electrophoresis 37(12) (2016) 1659-1668.

[29] J. Ge, B. Budowle, R. Chakraborty, DNA identification by pedigree likelihood ratio accommodating population substructure and mutations, Investig Genet 1(1) (2010) 8.

[30] C.H. Brenner, Symbolic kinship program, Genetics 145(2) (1997) 535-42.

[31] A. Lackey, https://www.thermofisher.com/blog/behindthebench/how-to-determine-if-two-y-str-haplotypes-are-from-the-same-male-lineage/, (2018).

[32] J. Ge, B. Budowle, X.G. Aranda, J.V. Planz, A.J. Eisenberg, R. Chakraborty, Mutation rates at Y chromosome short tandem repeats in Texas populations, Forensic Science International: Genetics 3(3) (2009) 179-184.

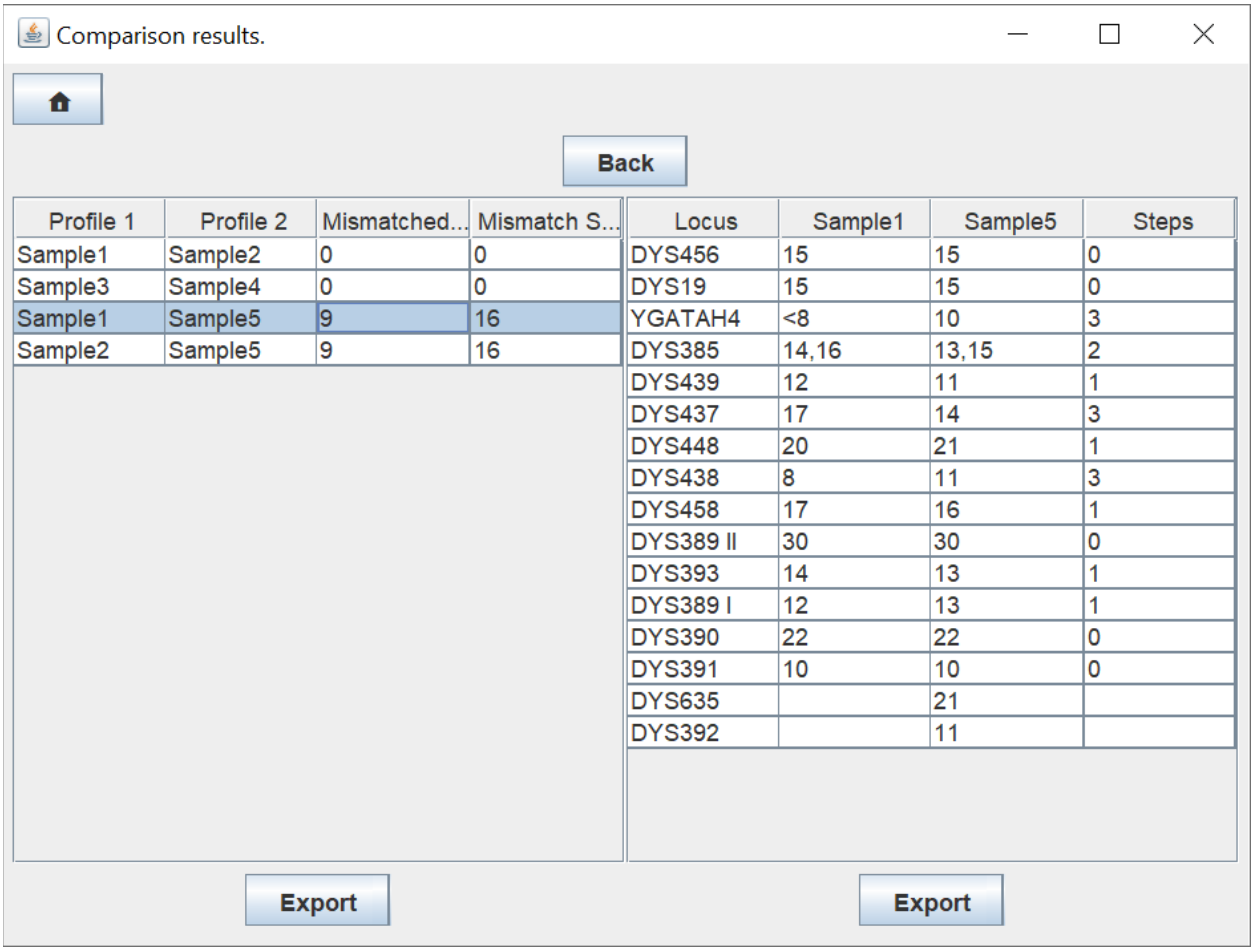
[33] S. Gopinath, C. Zhong, V. Nguyen, J. Ge, R.E. Lagacé, M.L. Short, J.J. Mulero, Developmental validation of the Yfiler® Plus PCR Amplification Kit: An enhanced Y-STR multiplex for casework and database applications, Forensic Science International: Genetics 24 (2016) 164-175.

[34] J. Purps, S. Siegert, S. Willuweit, M. Nagy, C. Alves, R. Salazar, S.M. Angustia, L.H. Santos, K. Anslinger, B. Bayer, A global analysis of Y-chromosomal haplotype diversity for 23 STR loci, Forensic Science International: Genetics 12 (2014) 12-23.

[35] A. Di Rienzo, A. Peterson, J. Garza, A. Valdes, M. Slatkin, N. Freimer, Mutational processes of simple-sequence repeat loci in human populations, Proceedings of the National Academy of Sciences 91(8) (1994) 3166-3170.

[36] A. Estoup, P. Jarne, J.M. Cornuet, Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis, Mol Ecol 11(9) (2002) 1591-1604.

Appendix



Profile 1	Profile 2	Mismatched...	Mismatch S...
Sample1	Sample2	0	0
Sample3	Sample4	0	0
Sample1	Sample5	9	16
Sample2	Sample5	9	16

Locus	Sample1	Sample5	Steps
DYS456	15	15	0
DYS19	15	15	0
YGATAH4	<8	10	3
DYS385	14,16	13,15	2
DYS439	12	11	1
DYS437	17	14	3
DYS448	20	21	1
DYS438	8	11	3
DYS458	17	16	1
DYS389 II	30	30	0
DYS393	14	13	1
DYS389 I	12	13	1
DYS390	22	22	0
DYS391	10	10	0
DYS635		21	
DYS392		11	

Figure S1. An example of the profile comparison results.

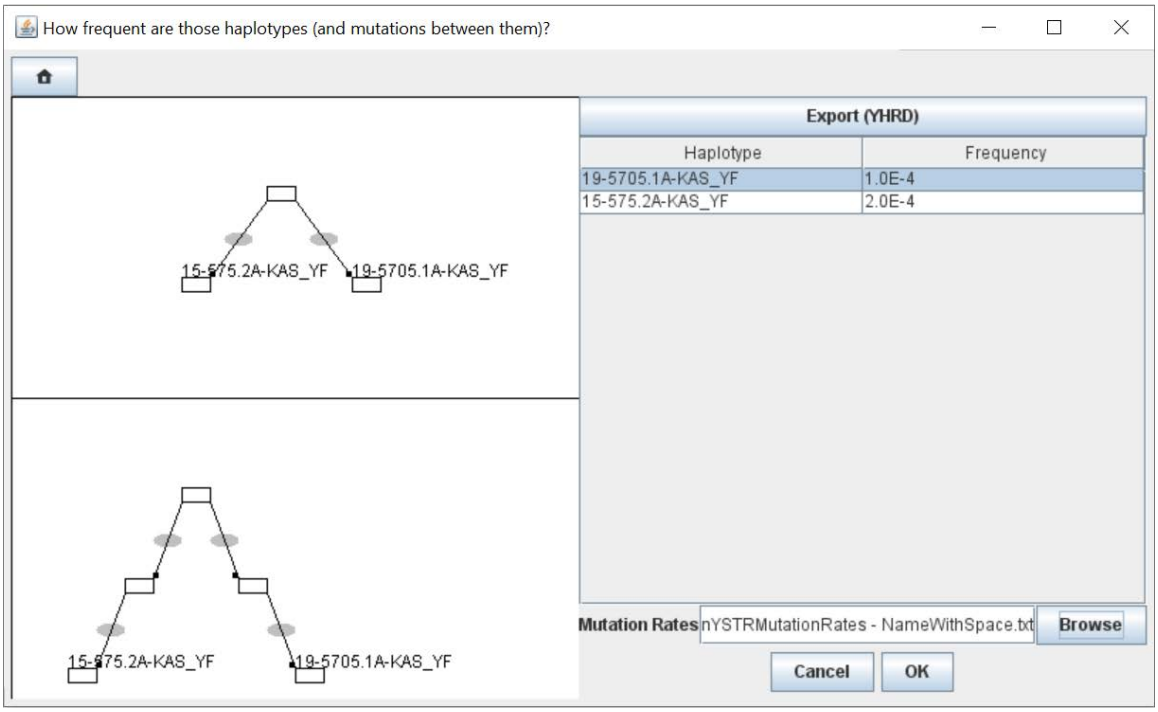


Figure S2. User interface of the defined pedigrees, the assignments of the haplotype frequencies and mutation rates.

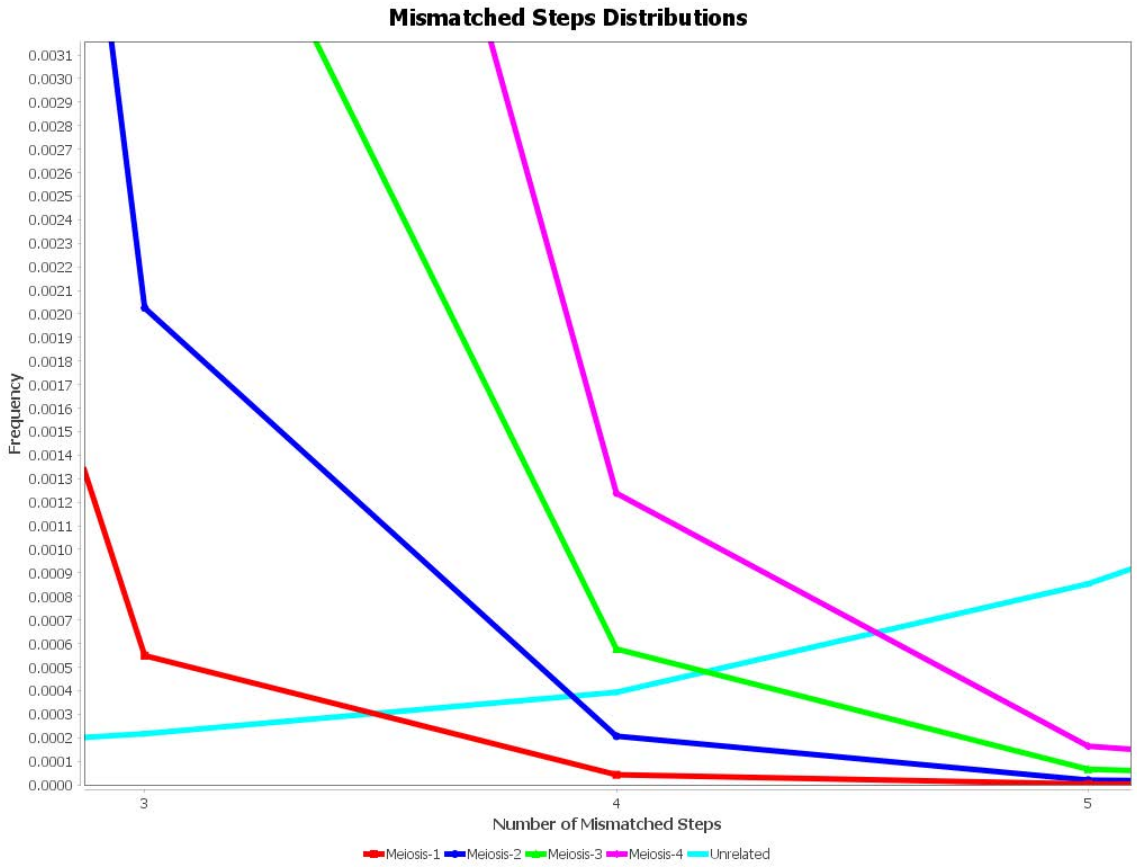


Figure S3. The distributions of the number of mismatched steps between related and unrelated Y23 profile pairs for **African Americans** in Purps et al. [35]. Meiosis-1 = father-son; Meiosis-2 = full-sibling; Meiosis-3 = uncle-nephew; Meiosis-4 = first-cousin.

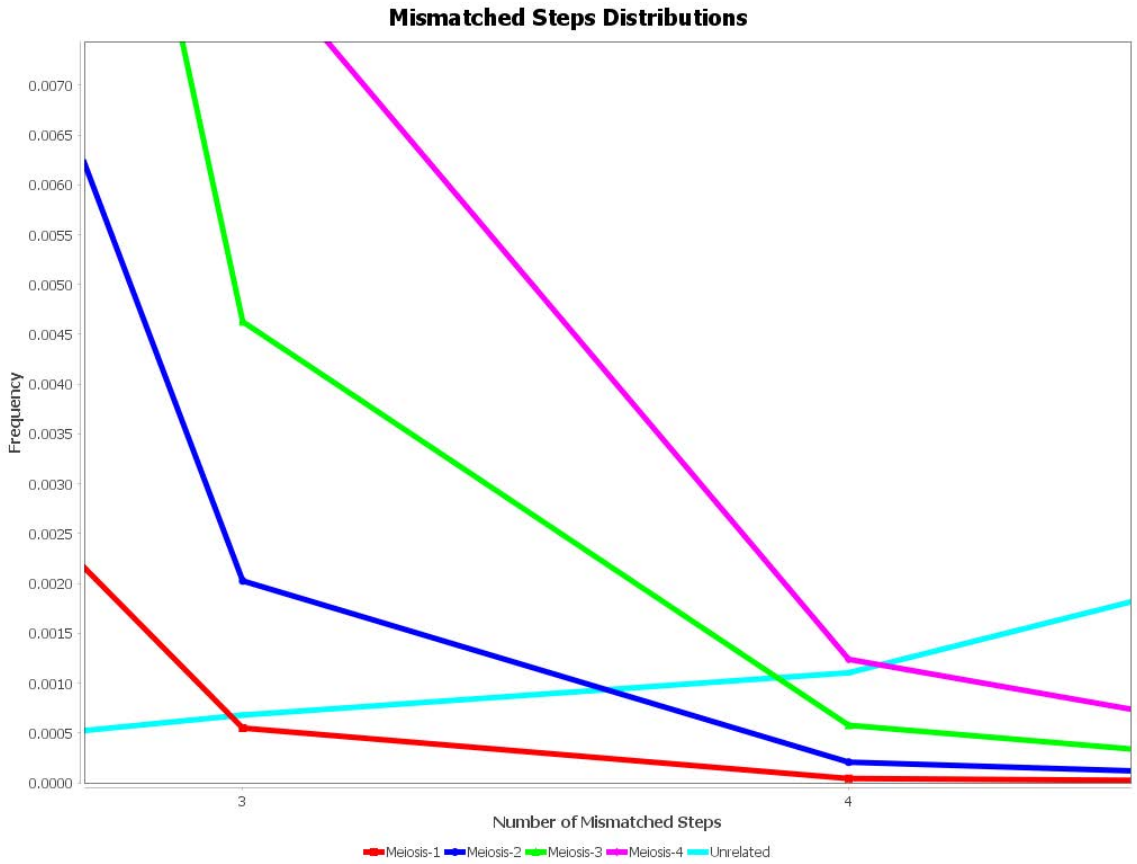


Figure S4. The distributions of the number of mismatched steps between related and unrelated Y23 profile pairs for **Asian Americans** in Purps et al. [35]. Meiosis-1 = father-son; Meiosis-2 = full-sibling; Meiosis-3 = uncle-nephew; Meiosis-4 = first-cousin.

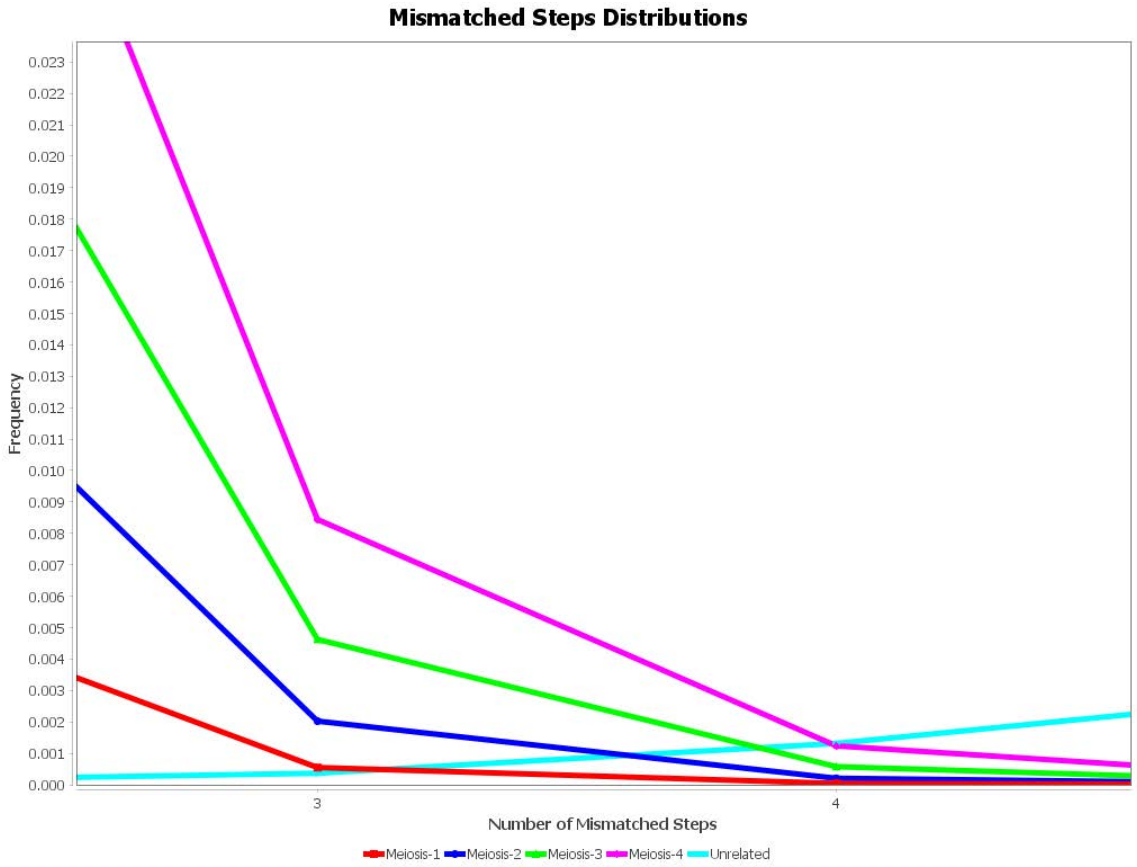


Figure S5. The distributions of the number of mismatched steps between related and unrelated Y23 profile pairs for **Caucasian Americans** in Purps et al. [35]. Meiosis-1 = father-son; Meiosis-2 = full-sibling; Meiosis-3 = uncle-nephew; Meiosis-4 = first-cousin.

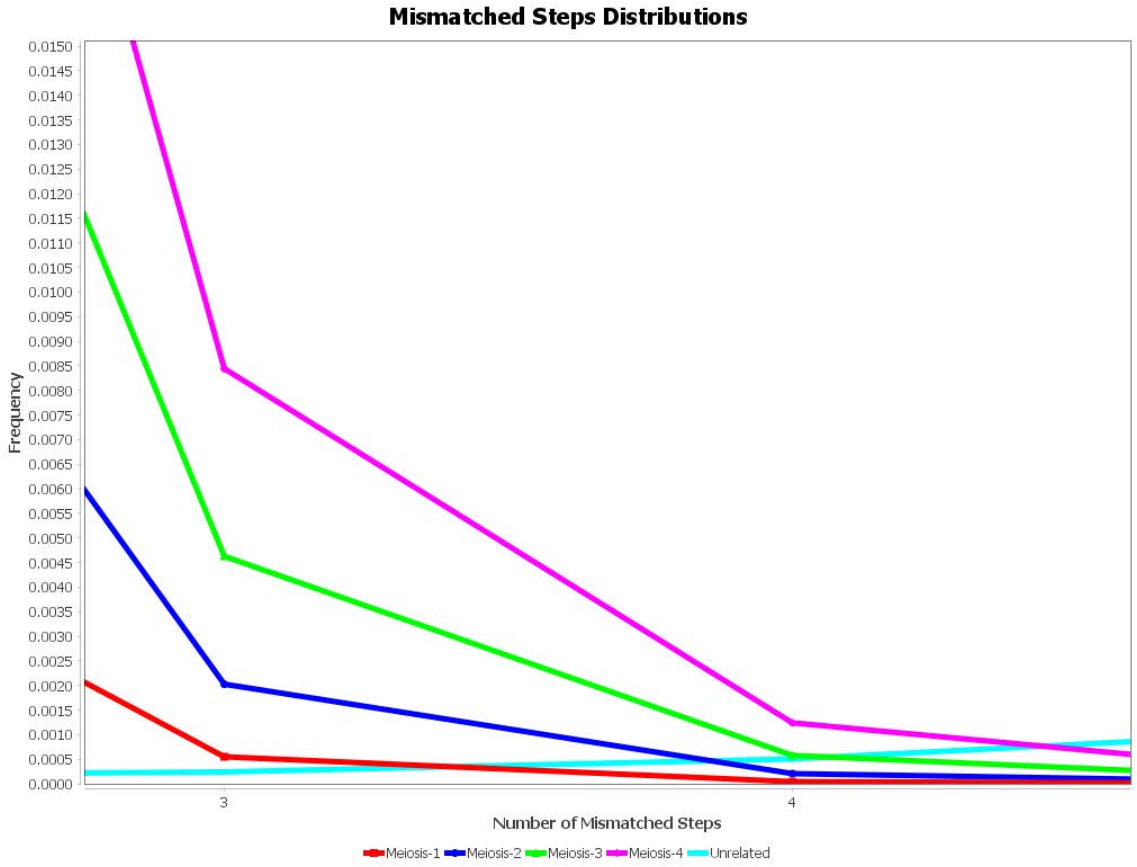


Figure S6. The distributions of the number of mismatched steps between related and unrelated Y23 profile pairs for **Hispanic Americans** in Purps et al. [35]. Meiosis-1 = father-son; Meiosis-2 = full-sibling; Meiosis-3 = uncle-nephew; Meiosis-4 = first-cousin.

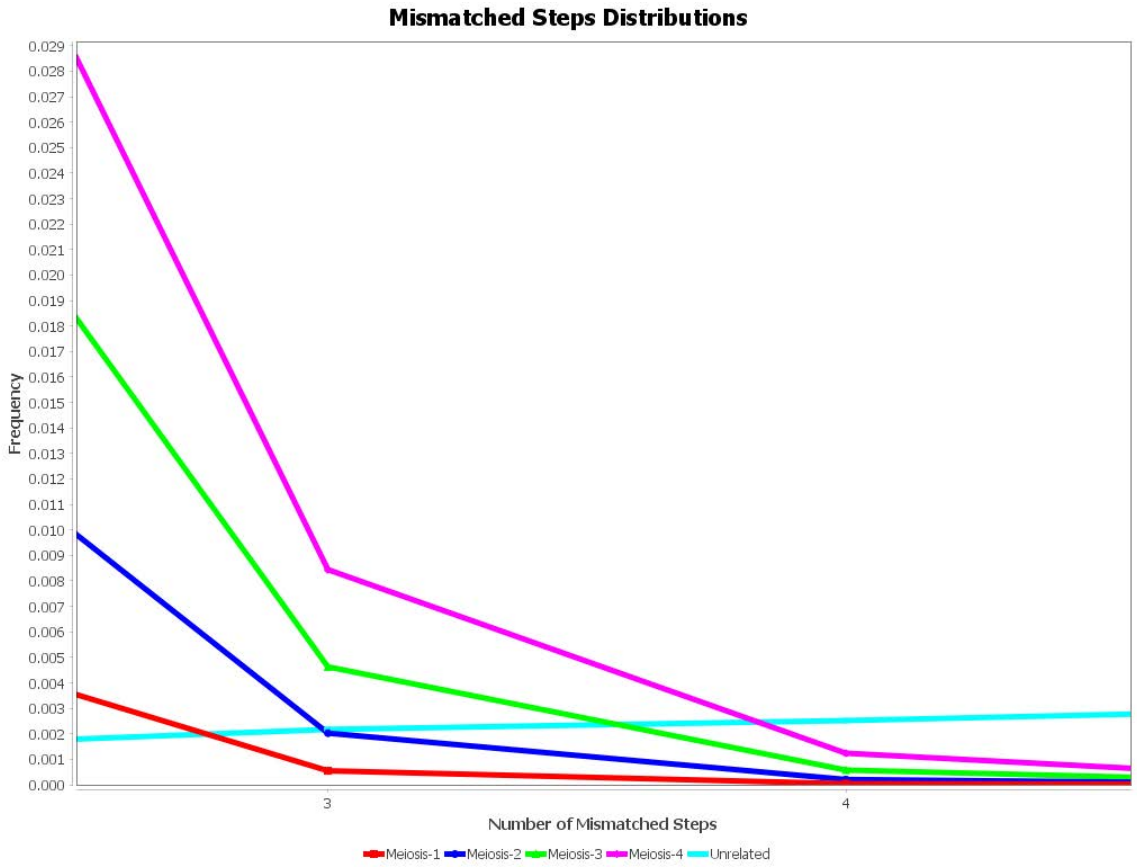


Figure S7. The distributions of the number of mismatched steps between related and unrelated Y23 profile pairs for **Native Americans** in Purps et al. [35]. Meiosis-1 = father-son; Meiosis-2 = full-sibling; Meiosis-3 = uncle-nephew; Meiosis-4 = first-cousin.

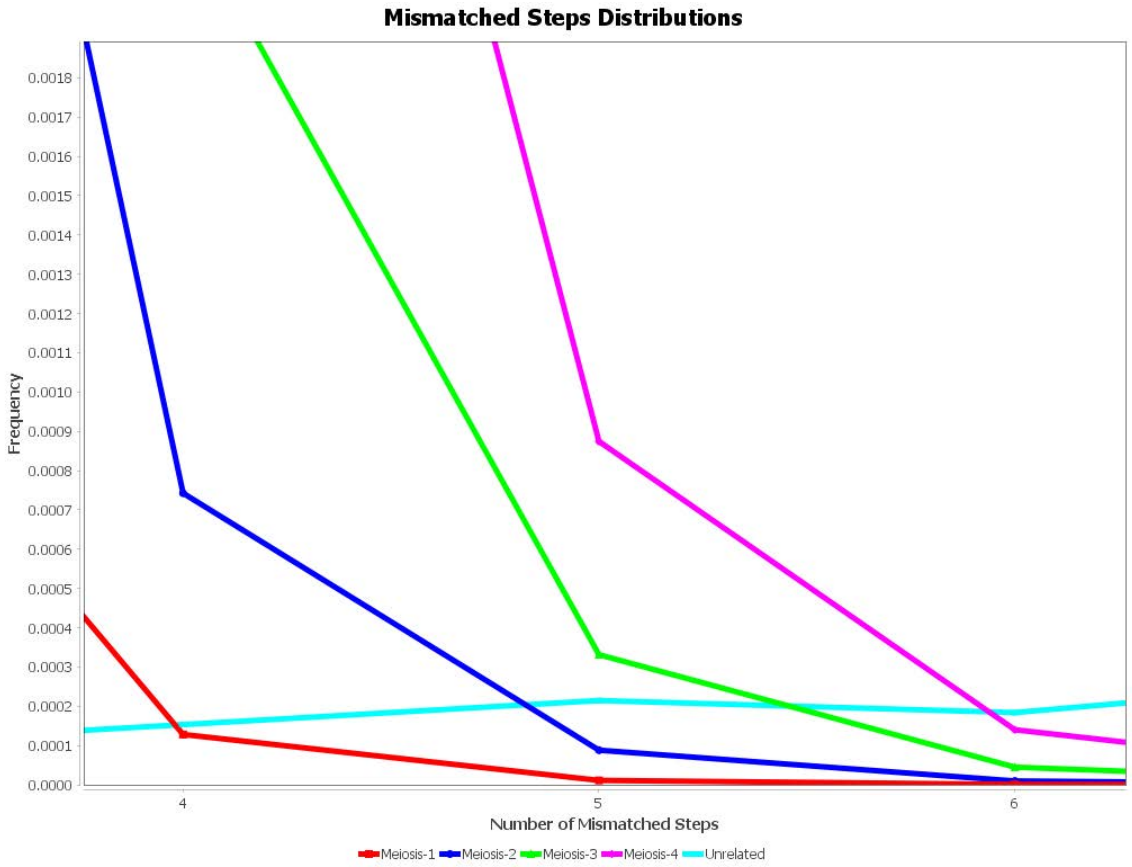


Figure S8. The distributions of the number of mismatched steps between related and unrelated Yfiler Plus profile pairs for **African Americans** in Gopinath et al. [34]. Meiosis-1 = father-son; Meiosis-2 = full-sibling; Meiosis-3 = uncle-nephew; Meiosis-4 = first-cousin.

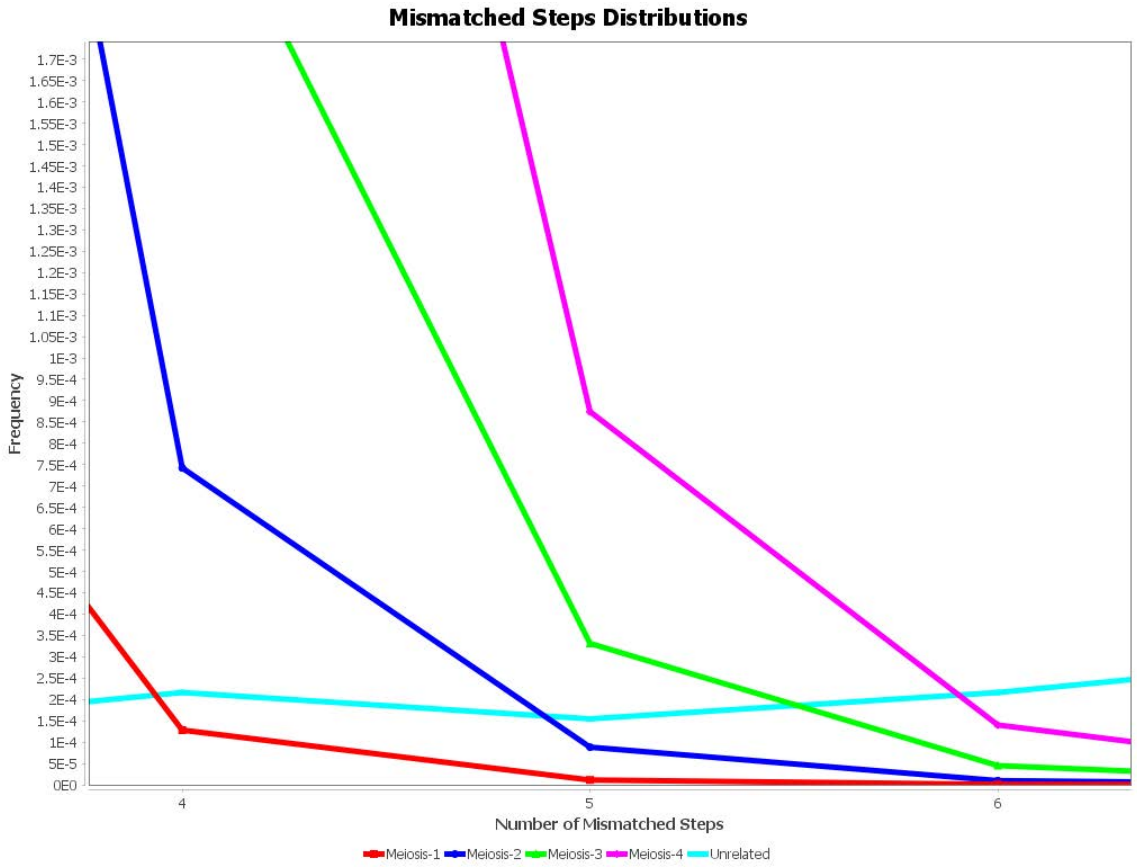


Figure S9. The distributions of the number of mismatched steps between related and unrelated Yfiler Plus profile pairs for **Asian Americans** in Gopinath et al. [34]. Meiosis-1 = father-son; Meiosis-2 = full-sibling; Meiosis-3 = uncle-nephew; Meiosis-4 = first-cousin.

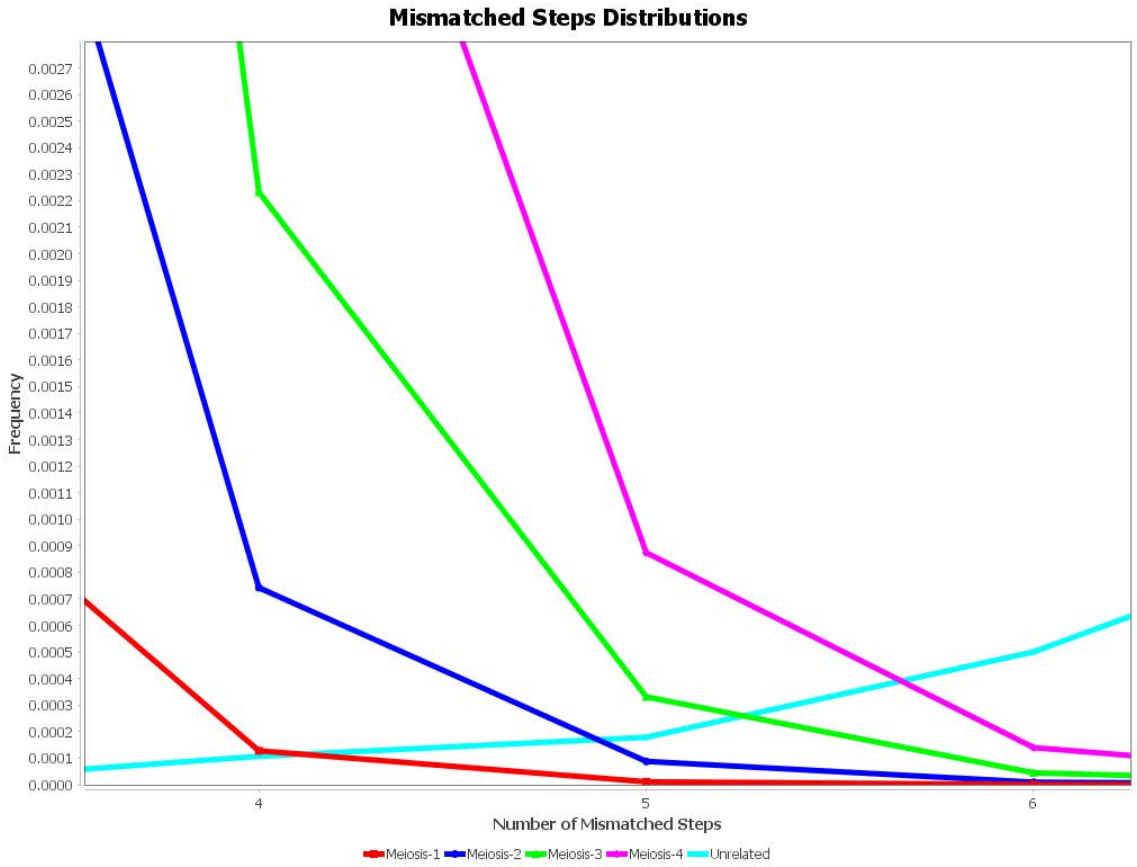


Figure S10. The distributions of the number of mismatched steps between related and unrelated Yfiler Plus profile pairs for **Caucasian Americans** in Gopinath et al. [34]. Meiosis-1 = father-son; Meiosis-2 = full-sibling; Meiosis-3 = uncle-nephew; Meiosis-4 = first-cousin.

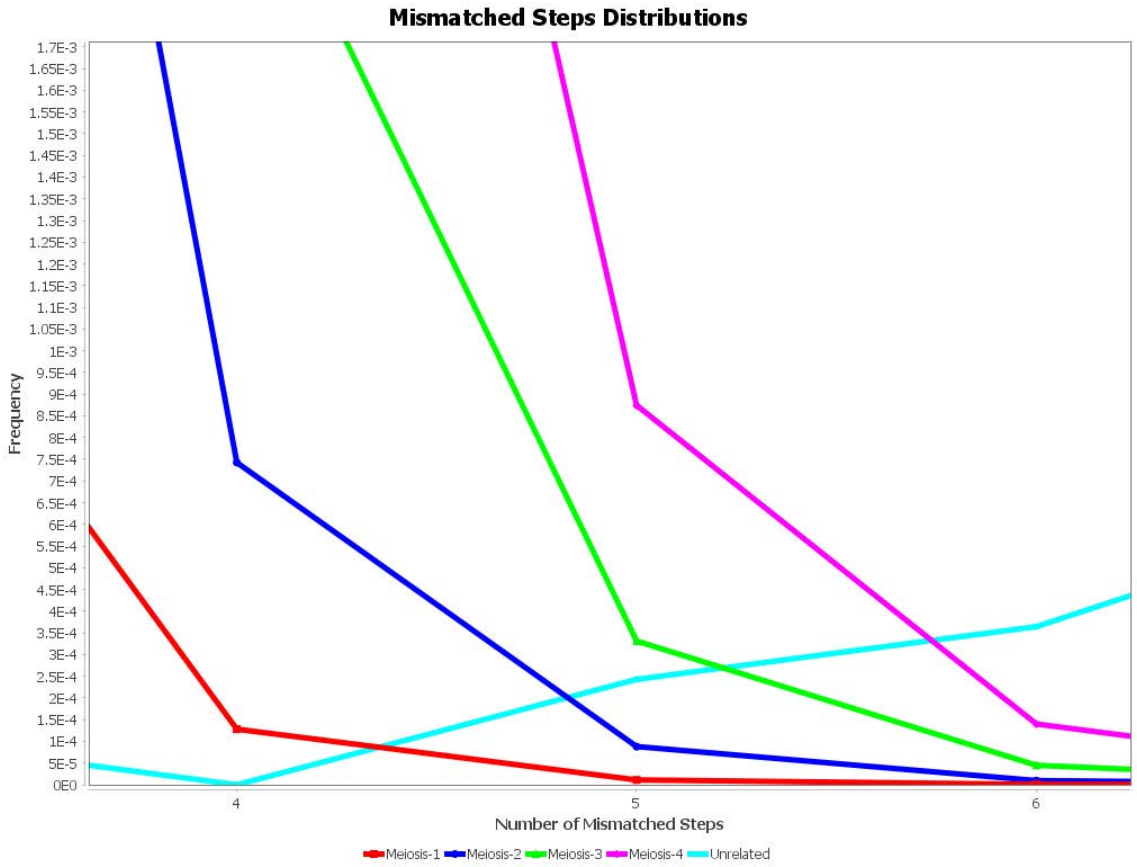


Figure S11. The distributions of the number of mismatched steps between related and unrelated Yfiler Plus profile pairs for **Hispanic Americans** in Gopinath et al. [34]. Meiosis-1 = father-son; Meiosis-2 = full-sibling; Meiosis-3 = uncle-nephew; Meiosis-4 = first-cousin.