The Use of Y-Chromosome Profiles in Forensic Genetics

Fa Kul

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Public Health

University of Washington

2022

Committee:

Bruce Weir

Anna C. Mastroianni

Program Authorized to Offer Degree:

Public Health Genetics

Fa Kul

University of Washington

Abstract

The Use of Y-Chromosome Profiles in Forensic Genetics

Fa Kul

Chair of the Supervisory Committee:
Bruce Weir
Department of Biostatistics

In forensic science, Y-STR testing is one of the most common tools used to interpret DNA typing results, specifically for human identification purposes (SWGDAM, 2014). With the complexity in DNA samples, not only can they be different from another individual, but each unique haplotype contains several markers with different combinations of alleles. While unique haplotypes do not arbitrate identity of individuals, methods in calculating numerical strength to a match between forensic evidence and a suspect remains an area of debate.

This study used Y-STR profiles from worldwide databases to evaluate and explore the alleles shared within and between each population. It also raises some of the ethical considerations associated with the use of Y-STR profiles for forensic genetics purposes.

If Y-STR typing or genetic profiling is the future of forensic genetics, more research needs to be done to ensure that the methods and technologies are effectively delivered and to assure public trust.

Introduction

Introduction to Forensic Genetics

Forensic DNA markers are used worldwide for current forensic casework. The use of these markers includes studying populations, identifying forensic identity, ancestry, and phenotype determination (Alonso et al., 2018). The markers are known as short tandem repeats (STRs), which were introduced as polymorphic DNA loci. In forensic DNA analysis, alleles within each STR are identified by PCR-based amplicon sizing method, as well as electrophoretic systems (Walter et al., 2016). Y-chromosome short tandem repeats or Y-STR typing is used in the forensic field due to its ability to analyze previously problematic DNA samples (Budowle et al., 2007). Since Y-STR loci are male-specific, they allow forensic scientists to analyze mixtures that predominantly comprise female DNA when low quantities of male DNA is present. This, therefore, makes Y-STR typing a suitable method for analyzing DNA in criminal cases, specifically, sexual assault cases (Kayser, 2017).

Autosomal STR typing is a technique used to identify individual profiles for forensic casework (Wyner et al., 2020). Similar to prior existing analyses for autosomal STRs, Y-STR analysis also provides evidence and reference samples using specific loci for forensic applications. By using these profiles, scientists can analyze whether the reference samples and evidence are matched, or whether there are shared components in the obtained mixtures (SWGDAM, 2014). The analysis leads to exploring whether the sample originated from the same source as the reference. One unique feature of Y-STR that distinguishes it from autosomal STR is that Y-STR profiles are not affected by recombination, which simplifies the analysis in the complexity of DNA markers but prevents the multiplication of match probabilities over loci (Kayser, 2017).

While Y-STR analysis can be a successful tool for analyzing genetic profiles in forensic settings, there are some ethical challenges.[1] Ethics concerns revolve around considerations about privacy invasions (LaRue, 2019) and equity (Murphy & Tong, 2020). Despite this knowledge, we still know very little about how accessing a DNA database affects individuals and families.

This study aims to evaluate the forensic use of Y-STR profiles and to explore recent compilations of current existing publications of Y-STR profiles. By examining the available forensic databases, we will be able to identify barriers to the use of Y-STR profiles in forensic genetics settings, which will improve understanding in the use of DNA profiling, as well as protect suspects from inaccurate analysis and wrongful conviction.

The specific aims are:

1. Describe the forensic use of Y-STR profiles and the publicly available datasets for Y-STR profiles
2. Describe the current methods of attaching numerical strength to matching Y-STR profiles and produce a report of numerical strength of that profile found to match between some evidence and a suspect
3. Describe some of the ethical considerations related to the use of Y-STR profiles in forensic genetics.

Before discussing the use of Y-STR typing in the context of forensic genetics and its unique features, I will be introducing relevant terms and summarizing the existing DNA typing used in the field, as well as how genetic profiles are stored.

---

[1] Potential violation of the US Constitution's protections against unreasonable search and seizure under the 4th Amendment is a significant legal concern in the US criminal justice system (LaRue, 2019) and is not discussed in this thesis.

Defining STRs

Short tandem repeats (STRs) are short repeated sequences of DNA, which can be 2 to 6 base pairs in length. STRs account for approximately 3% of the human genome with different varieties of the STRs (Wyner et al., 2020). STRs are noncoding regions and not implicated in gene expression. The number of repeat units in the STR offers the power of discrimination for identifying individuals, as they are highly heterozygosity in each person.

Autosomal STRs

Autosomal STR loci are both used and their profiles are stored in CODIS. They provide a higher power of discrimination. Although autosomal STRs are the preferred method in STR typing, the limitation includes reduced capabilities when using female-male mixtures, and differential extraction such as fingernail scraping (Hebda et al., 2014).

mtDNA

While both autosomal and Y-STRs are linear DNA located in the cell nucleus, mitochondrial DNA (mtDNA) are circular DNA located in the cell cytoplasm. mtDNA copies are not only highly abundant but they are not as degradable in highly degradable specimens or low quantity of samples. Similarly to Y-STR, mtDNA is treated as a single haplotype due to its lack of recombination, conveying a uniqueness or rarity of individuals (Budowle et al., 1999).

Y-STR Typing and its Forensic Applications

The application of Y-STR typing itself allows investigators to find more information on biological parts of Y-STR, or specifically focusing on male fraction. Y-STR typing is capable of both extracting and identifying DNA genotypes on the Y-chromosome, especially in a male-female sample (MacLean, 2013) or when evidence from autosomal typing is inconclusive or compromised (Hebda et al., 2014). That being said, Y-STR typing is especially useful for female rape cases.

In addition, Y-STR typing is especially useful when more than one offender is involved in the crime scene. This means that there are multiple source samples, as well as when the mixtures of DNA compromised the use of autosomal STR profiling (Hebda et al., 2014). However, due to the inheritance feature of the Y chromosome, the profiles are indistinguishable when they are compared between male relatives derived from the same paternal lineage, which can be father to son, siblings, and cousins (Fig 1).

NRY regions in the Y-STR loci

Non-recombining region (NRY) on the Y chromosome plays an important role in forensic analysis. Since Y-STR loci reside in this region, the assumption of independence between loci is not valid. Thus, the product rule is not an appropriate method to estimate the rarity of a Y-STR haplotype. Instead, the counting method is applied with the assumption that the haplotypes of Y-STR are treated as alleles at a single locus (Budowle et al., 2007 & Roewer et al., 2020). Since Y-STR lack independence of the markers on the NRY, the Y-STR haplotypes have reduced inclusion probabilities when compared to autosomal STRs, which can have independence of locus (Roewer et al., 2020). In addition, the Y-STRs are affected more by genetic drift due to the lack of recombination feature. The lack of recombination reduces the number of Y-chromosomes relative to autosomes, and the reduction in the effective number of the Y-chromosomes in a population is the result of gender-specific behaviors.

Current Methods for Numerical Strength

In order to calculate matching probabilities between two profiles used in identifying persons, it is essential to understand what methods are available and how they are used to calculate the matching probabilities. Below I summarize two current methods used within forensic genetics.

Counting Method

The counting method uses the evidentiary profile to query against a database profile. It estimates the profile probability Pr(A) = p_A. By searching a given haplotype against a size N reference database, it determines the number of times n the haplotype is observed in the database. The counting method is sometimes modified from n/N to n+1/N+1, where the haplotype of profile or an evidentiary profile is added to both the observations and the database (Roewer et al., 2020). Although, counting method is shown to be an effective and acceptable approach through estimating the rarity of evidentiary mtDNA haplotypes for more than a decade, the counting method can appear to reduce the evidentiary power (Budowle et al., 1999).

The counting method is used to report the proportion of haplotypes in a database of interest. The lack of recombination imposes dependencies at each locus among profiles. There is also dependence arising from relatedness and population membership, which results in match probability exceeding profile probability, thus, reduces the evidential strength of a match from the implied profile probability (Weir, 2007).

Theta Method

The theta method is used in the calculation of haplotype match probabilities (Syndercombe Court, 2021). This method is different from the counting method because its match probability is given A as a conditional probability.

$\Pr(A|\,A) = \theta + (1 - \theta)\, p_A$

The parameter, θ, is defined as a measure of normalized variance of allele frequencies over populations, which is usually interpreted as a measure of similarity of alleles within populations compared to the similarities of alleles between populations (Buckleton et al., 2011).

Theta is also known as FST, in which the SWGDAM lists theta values that are incorporated into the formula based on the number of loci typed and the amplification kit used (SWGDAM, 2014).

During the Y-STR typing, it is important to recognize the differences between Y-STR profile probability and match probability. The Y-STR probability is also known as profile probability. It is estimated from the observed haplotype frequency along with a confidence interval to the haplotype frequency estimate. Its goal is to capture the effect of database size. The match probability refers to the probability of observing a profile given that it has already been observed and it depends on the history of the population. Its goal is to address a match between the evidentiary reference samples given that the reference donor is the source of the evidentiary sample (Balding and Nichols, 1994).

Y-STR Markers and Their Profile Databases

During Y-STR typing, Y-STR markers carry identifiable features. Here I will summarize tools that amplify Y-STR markers for profiling purposes, as well as publicly available databases for Y-STR typing.

There are commercially available multiplex kits capable of amplifying as many as 23 STR loci on the Y Chromosome. One of the well-known multiplex kits is PowerPlex Y23.

PowerPlex Y23 (PPY 23) system is a 5-dye multiplex genotyping kit. This kit can analyze up to 23 Y-STR loci (Jain & Shrivastava, 2016). Usually, the PPY23 can analyze 17 Y-STR loci, which are common in other commercially available kits, these include DYS19, DYS385a/b, DYS389I/II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS635, and Y-GATA-H4. The 6 new Y-STR loci are DYS481, DYS533, DYS549, DYS570, DYS576, and DYS643 (Jain & Shrivastava, 2016), these loci add up to 23 loci in total. More loci available in testing kits allow Y-STR analysis to be more discriminating.

The PPY 23 kit is known to provide a robust and sensitive multiplex system. Therefore, it is useful for analyzing forensic casework (Jain & Shrivastava, 2016).

Other multiplex kits are Yfiler and Yfiler Plus. Yfiler Plus and PPY23 are considered more sensitive than Yfiler, as they reduce the occurrence of off ladder alleles. Thus, both kits can produce better performance for amplifying Y-STR and analyzing evidence for sexual assault casework (Ferreira-Silva et al., 2018).

Profile Databases

Y-Chromosome STR Haplotype Reference Database

The Y-Chromosome STR Haplotype Reference Database (YHRD) is a worldwide scientific public reference database specifically designed for searching Y-STR haplotypes through different formats (Roewer et al., 2020). The database can generate reliable Y-STR haplotype frequency estimates.

Individuals working in the field of forensic, human and population genetics submit their population data onto YHRD site, where they will be given accession number to add population sample once the YHRD examines the sample validity and originality. The population samples are categorized into subgroups defined by their nationality and by ancestry (Roewer et al., 2020).

STRBase

Another database is STRBase, which is an internet accessible DNA database. STRBase consists of more than 3000 reference STR profiles. STRBase is compiled by the National Institute of Standards and Technology (NIST), where it provides an information resource for the forensic DNA typing community. The NIST compiles and organizes information on observed alleles for sequences found on STR loci as well as commercially available STR multiplex kits and STR analysis technologies (Ruitberg et al., 2001).

Y-STR Profiles Study

The goal of this study is to explore the current forensic use of Y-STR profiles and current methods in Y-STR profile matching.

Methods

Using previously existing and current research, I reviewed forensic use of Y-STR profiles. Once I reviewed relevant literature, I described current methods of attaching numerical strength to match Y-STR profiles. These methods include the counting method, and the theta method. The publicly available datasets for Y-STR profiles were also described earlier in the paper.

To calculate and produce a report for numerical strength for population haplotype, I used 10 Y-STR profiles from Purps (Purps et al. 2014) to explore the shared Y-STR sequence variants within and between populations in YHRD databases.

Data Collection

10 Y-STR profiles from Purps were used for the numerical strength analysis. All the profiles included in the study were males, as Y-STR loci are male-specific markers. Exclusion criteria were profiles with fewer than 17 Y-STR loci, and profiles from country-specific databases raising concerns about potential identifiability of samples (Mainland China, Hong Kong, and Taiwan). I did not include all the populations from Purps. However, 10 profiles from different populations were selected and analyzed in the following order: Belgium [Belgian], Spain [Spanish], Bolivia [Mestizo, Native American], the US [Native American - Athapaskan], Japan [Japanese], Vietnam [Vietnamese], and the US [African, European, and Hispanic].

Data Analysis

Data from Purps was analyzed in YHRD. Each profile was checked to ensure that they consisted of at least 17 loci. When any profile consists of less than 17 loci, I removed them from the analysis.

To run the numerical analysis, I manually input each of the individual profiles into the YHRD website. Before running the analysis, I created a sheet consisting of all the profiles that I would be analyzing (Appendix 1). I then selected PPY23 for Y23 and Yfiler Plus for Y17, which represented the kind of commercial kit used and the number of loci included in the profile.

For each profile that I ran with Y17, I selected Yfiler Plus as its tool kit. I then used it to run for another round of analysis. This second round of analysis was done with the PPY 23 but with the Y23 option. This was done to compare Y17 and Y23 numerical strength results using the same profile for both kits.

To select database, I selected a population from the metapopulation database based on the 10 data profiles, which are categorized as follows: Belgian - Eurasian - European, Spanish (Eurasian - European), Mestizo [Bolivia] - Admixed, Native American [Bolivia] - Admixed Native American, Athapaskan - Admixed Native American, Japanese - East Asian - Japanese, Vietnamese - Admixed Asian American, African American - Admixed African American, European American - Admixed Caucasian American, and Hispanic American - Admixed Hispanic American. The categorization of each metapopulation group was based on each population's geographic location, with the exception of Vietnam, which was categorized under Asian American, as there was no subgroup under "Southeast Asian."

While the worldwide database is always set as default for each analysis, I proceeded to select the US national database and analyze the 10 haplotypes. If there was no match found within the database, I would not select the US subpopulation. However, if there was a match found within the US national database, I would look at the subpopulation to see the match as well as the theta values.

Report

Data from Purps was analyzed in YHRD using available markers to produce a numerical strength report for each haplotype profile.

For further analysis, I also selected "add more features to the report," where I would select different databases I want to compare the haplotype profile against. The databases I used were the worldwide, which was a default database; metapopulation, US national database, and US national database with subpopulations. Metapopulation is defined as an admixed population, where in this analysis, is specific to diverse populations in the US.

For all the reports produced, I created tables summarizing the match probabilities for each database (Tables 1-3). Note that theta value calculations shown in the reports are only available under the US subpopulation database.

Results

Using Search on the YHRD website to calculate numerical strength and match probabilities for profile haplotype, the results were shown based on its selected database features: worldwide, metapopulation, US national database, and US national database with subpopulations. Table 1 showed match probabilities using counting, and theta methods for all populations included in the analysis, these include both profiles with 17 and 23 markers. Each row showed each population had different haplotypes, as well as upper confidence intervals. The same haplotype profile provided different match probabilities when changing its number of loci: Y17 and Y23.

An example of the report was produced from YHRD using Y17's Yfiler Plus kit (Fig 2). This is done using a Belgian profile obtained from Purps.

For each of the following 10 profiles obtained from Purps, at least one match is expected from the worldwide population database. This is because the Purps data are within YHRD.

Belgium

There was one match found in both Y17 and Y23 in the worldwide and Eurasian-Europe databases, where there was zero match found within the US national database, as the Purps data were not from US national database populations. Although with Y17, the profile was used to compare against a larger haplotype pool, it did not show more matching other than the one found in the worldwide and subpopulation of Eurasian-Europe databases.

Spain

For the Spanish population, there were two matches found in Y17's and one match in Y23's worldwide database. Both Y17 and Y23 did not find any matches present within the US national database. Under the metapopulation of Eurasian-European, there were two matches found in Y17, and one match found in Y23 due to Purps being stored in YHRD.

Bolivia

For the Mestizo population, there were 18 matches found in Y17's worldwide, nine matches in the metapopulation - Admixed, and five matches in the US national pool. The subpopulations are as follows: one match in African American, three matches for Caucasian American, and one match for Hispanic American. While Y23's worldwide and metapopulation databases found one match as Purps is stored in YHRD, and no match for other databases.

For the Native American population, there were two matches found in the worldwide database with Y17, and only one match found in Y23's worldwide database. No matches found under other databases.

USA Part I

For the Athapaska, which is a large group of Native American population; there was one match found for each Y17 and Y23, both of them had one match found in the US national database. The match found was under Native American as the subpopulation within the US.

Japan

There were 7 matches found under the worldwide database using Y17, in which there were no matches found within the US national database. Under the metapopulation database, there were three matches found in East Asian-Japanese.

There was only one match found in Y23's worldwide database and East Asian-Japanese. No matches found within the US national database.

Vietnam

There were four matches found in Y17's worldwide database. However, there was only one match found within both the metapopulation and the US national database, where the subpopulation was Asian American.

For Y23, there were two matches found in the worldwide database, with one match in both metapopulation and the US national, specifically in Asian American.

USA Part II

For the African American population, there were three matches found in Y17's worldwide database, with two matches in the US national and metapopulation databases. The subpopulation for the two matches were African American. While for Y23's worldwide, there were two matches found. One match was found under the US national and the metapopulation databases, which were under the subpopulation of African American.

For the European American population, there was one match found in each Y17's worldwide, metapopulation, and US national. The subpopulation was Caucasian American.

One match was also found in the Y23's worldwide, metapopulation and the US national databases. The subpopulation was Caucasian American. However, Y23 profile was compared against a smaller number of haplotypes due to higher discriminatory power in loci.

For the Hispanic American population, there were 14 matches found in Y17's worldwide database. However, only one match was found within the US national and metapopulation databases, in which the subpopulation was Hispanic American.

In Y23's worldwide database, only one match was found. There was also one match found in the US national and metapopulation databases. The subpopulation was Hispanic American.

Discussion

This project aimed to look at the forensic application in the Y-STR typing. This was accomplished by analyzing data in individuals from collected profiles from Purps through the use of YHRD, which showed how numerical strength methods were used to calculate the match probability between profiles. Additionally, I aimed to tie the numerical strength approaches to real world forensic application that exists for individuals who are interested in calculating profile probabilities and finding match probabilities within different databases available in YHRD, which are worldwide, metapopulation, US national database, and US national database with subpopulations.

I hypothesized that there would be at least one match found in profiles from countries outside of the US when filtering within the US database because Purps is in YHRD. My numerical strength results showed that for every profile from other countries outside of the US domain, there was at least one match profile found from a different total number of haplotypes. Given the results of this analysis, it may be possible that individuals from different countries share the same profile as ones in the US. This rarely happened but it was still possible to occur as the US consists of a diverse population from different countries. However, it is possible that when there was one match found within the worldwide or the US database, the match was due to the existing profile within the databases, the two profiles were close relatives - shared the same paternal lineage, or the manually searched profile belonged to the same individual. The data I obtained from Purps is one of the examples, as the profiles used were stored within the YHRD databases.

While producing reports for numerical strength calculations for each unique haplotype, I observed that the matching of profiles are mostly within the same populations. This was also observed in the analysis of matching pairs in all populations in Purps using 17 and 23 loci (Tables 4-7). These tables did not use YHRD, but only used Purps' all population data. Table 5 and 7 showed that it was more probable to observe more matching pairs within the same population rather than between pairs of populations in both using Y-STR's 17 and 23 markers. These results supported the assumption that population membership can affect the probability of getting matched profiles.

Looking back to the reports produced across 10 profiles from different populations, I also observed that a profile that had more than one match usually shared profiles from the same subpopulation, indicating that profiles are more likely to match when they are within the same subpopulation.

In addition, population databases with different numbers of loci can lead to different match probabilities. Haplotypes with more loci generally give less matching and more discrimination when analyzing the profiles through commercial kits. Compared to other available DNA typing, Y-STR typing is still new in the forensic field.

Some Ethical Considerations Related to Y-STR Typing in Forensic Genetics

While Y-STR typing is helpful in identifying suspects and missing persons through paternity and lineage analyses, it is also capable of exonerating innocent people (LaRue, 2019). To ensure that the benefits of Y-STR typing are fully realized, it will be necessary to ensure that public trust is maintained. That requires attention to ethics. Some of the ethical concerns raised by Y-STR typing include privacy and equity, which are illuminated below through discussions of the CODIS database and familial searching using the example of the Usry case.

Forensic DNA Databases: Issues of Equity and Privacy

One example of a forensic DNA Database is the Combined DNA Index System (CODIS), a DNA database and the software that searches for matches with targeted DNA (Aherne, 2018). It consists of DNA profiles belonging to criminals, arrestees, as well as missing persons. The National DNA Index System (NDIS) is a part of CODIS, and contains DNA profiles contributed from a variety of sources, which include federal, state, and local forensic laboratories (FBI; Murphy & Tong, 2020).

Although the CODIS database keeps individuals' profiles confidential from public access and the stored genetic data do not reveal genetic traits of those individuals, the system includes identifiable data associating individuals with their criminal records, which could reveal familial data. This raises privacy concerns since CODIS has the potential to identify relatives of individuals found in the database (Ram, 2015).

In addition to a potential violation of individuals' and relatives' privacy from searches that do not lead to criminal convictions, the lack of racial and ethnic diversity in forensic DNA databases raises equity considerations. According to legal scholars Erin Murphy and Jun Tong (2020), of the total US population, 62% identify as white and 13% identify as Black. However, the populations represented in forensic genetic databases reviewed by the authors are disproportionate in comparison: 49% white and 34% Black. In addition, they estimate that the burden of collection and storage is disproportionately within people of color, and is concentrated on the Black population. Specifically, the scholars reveal annually collected statistics in forensic DNA databases: 2.26% of Black samples were subjected to be collected, followed by 0.69% Hispanic, and 0.12% Asian (Murphy & Tong, 2020). CODIS also disproportionately represents racial and minority populations (Aherne, 2018; Murphy & Tong, 2020). One concern about the overrepresentation of Blacks in forensic DNA databases is that it increases the potential for confirmation bias. Confirmation bias is defined as the tendency to interpret, recall and search for information that support one's beliefs (Casad, 2016). When an evidentiary profile is searched against existing profiles within a database with disproportionate racial and ethnic diversity, in which this overrepresentation of Blacks can exacerbate or reinforce beliefs and biases of law enforcement of criminality in the Black population.

At least one scholar has suggested that equity concerns can be addressed by ensuring that everyone's genetic information be stored in a national database and equally accessible (Aherne, 2018). While it seems fair that everyone's genetic profile is equally accessible, that raises other potential ethical, social and legal questions that would need to be further investigated in depth.

Although the rate of wrongful convictions are not clear, there have been occurrences of relatives of individuals whose profiles are stored in the CODIS being targeted for crimes they did not commit. LaRue (2019) describes examples of individuals who were interrogated and pressured to confess for uncommitted crimes. Such errors not only could lead to wrongful accusations, but in the worst case, wrongful conviction (Lyttle, 2014). This is another example of how families of offenders and accusers may be impacted by confirmation bias, in which they were stigmatized and discriminated against by law enforcement. Also, when families are targeted, it raises additional issues of privacy, as noted above.

Familial Searching and the Usry Case: Individual and Familial Privacy

Familial searching looks for close relatives within a DNA database instead of searching for the perpetrator of an offense (Murphy & Tong, 2020). Y-STR typing is one approach of several that can be used in familial searching. Ethical concerns more broadly raised by this investigative tool include violations of a suspect's or wrongfully convicted individual's self and relatives' privacy.

The ethical issues raised in other types of DNA profiling analysis led to the use of Y-STR typing. Since Y-STR typing does not retrieve or compare family member's genotypes, which is beneficial for the individual's families, who have been accused or have their profiles stored in CODIS and other forensic databases (Ram, 2015). Its approach of providing evidence is similar to unknown fingerprints, in which Y-STR provides identifiable genetic information of an individual (MacLean, 2013; Murphy & Tong, 2020), making the technique appealing for investigative forensic applications.

However, there are states that use Y-STR via familial searching, in which this approach has the potential to invade an individual's privacy (Aherne, 2018; Ram, 2015). Below I discuss the Usry case as an example of how Y-STR typing can misidentify individuals with serious ramifications for an individual and family members.

Usry Case

In 1996, Angie Dodge was sexually assaulted and murder in the Idaho Falls. The case had been unsolved until 2014, when the investigation team found a match of 34 of 35 in searched Y-STR alleles. This result belonged to Michael Usry's father, who decided to do genetic testing for personal reasons (Akpan, 2019). Law enforcement compared DNA evidence from the crime scene to DNA profiles in Ancestry.com (Swenson, 2019). Paternal lineage analysis strongly indicated that a close relative of Michael Usry's father might be the suspect. Usry's son, Michael, became a suspect for the cold case (Akpan, 2019). Thus, Usry was accused as a criminal suspect through his father's genetic genealogy, specifically using Y-STR analysis. One month after he was accused, Usry's DNA result from the standard autosomal DNA typing was found to not match the DNA at Angie Dodge's crime scene. The criminal allegation was dropped by law enforcement, leaving Angie Dodge's suspect unknown again.

Michael Usry Jr.'s case is an example of familial searching, where it not only invaded individual's and their families' privacy, but also almost led to wrongful conviction. Although Usry was not imprisoned or considered guilty, his one month of being considered a murder suspect reportedly, and not surprisingly, left him unsettled and stressed. The stigma against Usry being accused as a criminal suspect reportedly remained even when his allegation was dropped (Scudder and McNevin, 2018). In addition, the use of the DNA profile belonging to Usry's father for criminal search purposes rather than for its original intended purposes raises questions about whether such actions by law enforcement's actions justifiably compromised privacy.

Conclusion

Based on the Y-STR Profiles Study, I observed that matching profiles within the same populations are more likely to occur than finding the matches between random populations. Although the matching pairs of Y-STR haplotypes between different populations are rare, there is still potential for its occurrence. When using YHRD data search to produce a numerical strength report to calculate matching of the profiles, it is critical to be aware that a total of one match between profiles could mean the existence of the same profile within the YHRD database.

It is important to understand the use, and limitations of Y-STR typing, as well as to acknowledge existing ethical concerns. Some ethical considerations include potential for the violation of privacy resulting from the sample's inherent genetic identifiability and equity issues resulting from disproportion of racial and ethnic diversity, which specifically negatively impact the people of color.

While it is undeniable that Y-STR typing is a powerful tool, which is predominantly used in identifying suspects in the justice system, as it can identify and exclude suspects based on their Y-STR profiles, as well as exonerate innocent convicts. The downside of Y-STR typing is that it also has a capacity to facilitate wrongful conviction as well as overturn the wrongfully condemned and exonerate the innocent. Misinterpretation of statistical results and profile matches due to human and laboratory errors can exacerbate into coincidental matches and result into miscarriage of justice (LaRue, 2019).

If Y-STR typing is the future of forensic genetics, more ethical research needs to be done. In order to ensure that public trust is maintained to reap the benefits of Y-STR typing, future research on how misuse of Y-STR typing results impact accusers, their relatives, as well as families of criminal offenders are needed.

References

Aherne, A. "Support of Familial DNA Testing in Illinois Criminal Investigation." *Northern Illinois University Law Review*, vol. 38, no. 3, Summer 2018, p. 553-584. HeinOnline, https://heinonline-org.offcampus.lib.washington.edu/HOL/P?h=hein.journals/niulr38&i=579.

Akpan, N. "Genetic genealogy can help solve cold cases. It can also accuse the wrong person." *Pbs*, Nov 7, 2019.

Alonso, A., et al. “Current State‐of‐Art of STR Sequencing in Forensic Genetics.” *Electrophoresis*, vol. 39, no. 21, 2018, pp. 2655–2668.

Balding, D. J., and Nichols, R. A. “DNA Profile Match Probability Calculation: How to Allow for Population Stratification, Relatedness, Database Selection and Single Bands.” *Forensic Science International*, vol. 64, no. 2, 1994, pp. 125–140.

Buckleton, J. S., et al. “The Interpretation of Lineage Markers in Forensic DNA Testing.” *Forensic Science International: Genetics*, vol. 5, no. 2, 2011, pp. 78–83.

Budowle, B., et al. “Basic principles for estimating the rarity of Y-STR haplotypes derived from forensic evidence.” In: Proceedings of the Eighteenth International Symposium on Human Identification, 2007.

Budowle, B., et al. “Mitochondrial DNA Regions HVI and HVII Population Data.” *Forensic Science International*, vol. 103, no. 1, 1999, pp. 23–35.

Casad, B. J. “Confirmation Bias.” Britannica. 2016. Britannica. https://www.britannica.com/science/confirmation-bias. Accessed, Jun 8, 2022.

FBI. Frequently Asked Questions on CODIS and NDIS. https://www.fbi.gov/services/laboratory/biometric-analysis/codis/codis-and-ndis-fact-sheet. Accessed, Jun 7, 2022.

Ferreira‐Silva, B., et al. “A Comparison Among Three Multiplex Y‐STR Profiling Kits for Sexual Assault Cases.” *Journal of Forensic Sciences*, vol. 63, no. 6, 2018, pp. 1836–1840.

Hebda, L. M., et al. “Collecting and Analyzing DNA Evidence from Fingernails: A Comparative Study.” *Journal of Forensic Sciences*, vol. 59, no. 5, 2014, pp. 1343–1350.

Jain, T., and Shrivastava, P. “PowerPlex Y23 System: A Fast, Sensitive and Reliable Y-STR Multiplex System for Forensic and Population Genetic Purpose.” *Journal of Molecular Biomarkers & Diagnosis*, 2016, Vol.7 (3).

Kayser, M. “Forensic Use of Y-Chromosome DNA: a General Overview.” *Human Genetics*, vol. 136, no. 5, 2017, pp. 621–635.

LaRue, R. A. "The Science of Change: Familial Searches and Y-STR DNA." *Ohio State Journal of Criminal Law*, vol. 17, no. 1, Fall 2019, pp. 241-[iv]. HeinOnline, https://heinonline-org.offcampus.lib.washington.edu/HOL/P?h=hein.journals/osjcl17&i=251.

Lyttle, A. J. "Return of the Repressed: Coping with Post-Conviction Innocence Claims in Wyoming." *Wyoming Law Review*, vol. 14, no. 2, 2014, pp. 555-632. HeinOnline, https://heinonline-org.offcampus.lib.washington.edu/HOL/P?h=hein.journals/wylr14&i=580.

MacLean, C. E. "Creating a Wanted Poster from a Drop of Blood: Using DNA Phenotyping to Generate an Artist's Rendering of an Offender Based Only on DNA Shed at the Crime Scene." *Hamline Law Review*, vol. 36, no. 3, 2013, p. 357-386. HeinOnline, https://heinonline-org.offcampus.lib.washington.edu/HOL/P?h=hein.journals/hamlrv36&i=403.

Murphy, E., and Tong, J. H. “The Racial Composition of Forensic DNA Databases.” *California Law Review*, vol. 108, no. 6, 2020, p. 1847.

Purps, J., et al. “A Global Analysis of Y-Chromosomal Haplotype Diversity for 23 STR Loci.” *Forensic Science International : Genetics,* vol. 12, no. 100, 2014, pp. 12–23.

Ram, N. "DNA by the Entirety." *Columbia Law Review*, vol. 115, no. 4, May 2015, p. 873-940. HeinOnline, https://heinonline-org.offcampus.lib.washington.edu/HOL/P?h=hein.journals/clr115&i=953.

Roewer, L., et al. “DNA Commission of the International Society of Forensic Genetics (ISFG): Recommendations on the Interpretation of Y-STR Results in Forensic Analysis.” *Forensic Science International: Genetics*, vol. 48, 2020, p. 102308.

Ruitberg, C. M., et al. “STRBase: a Short Tandem Repeat DNA Database for the Human Identity Testing Community.” *Nucleic Acids Research*, vol. 29, no. 1, 2001, pp. 320–322.

Scudder, N. and McNevin, D. “Is your DNA really your own? The public and forensic value of DNA.” *The conversation*, May 1, 2018. https://theconversation.com/is-your-genome-really-your-own-the-public-and-forensic-value-of-dna-95786 . Accessed, Jun 4, 2022.

Syndercombe Court, D. “The Y Chromosome and Its Use in Forensic DNA Analysis.” *Emerging Topics in Life Sciences*, vol. 5, no. 3, 2021, pp. 427–441.

Swenson, K.. “Police twice targeted the wrong men for a brutal 1996 killing. A cigarette butt changed everything.” *Washington Post*. May, 17, 2019.

SWGDAM. “Interpretation Guidelines for Y-Chromosome STR Typing.” SWGDAM Documents, 2014. https://www.swgdam.org/publications. Accessed, Jun 5, 2022.

Weir, B. S. “The Rarity of DNA Profiles.” *The Annals of Applied Statistics*, vol. 1, no. 2, 2007, pp. 358–370.

Wyner, N., et al. “Forensic Autosomal Short Tandem Repeats and Their Potential Association with Phenotype.” *Frontiers in Genetics*, vol. 11, 2020, p. 884.

Figures

Fig 1. Y-STR Mode of Inheritance

Pedigrees showing all Y-STR in males represented by purple squares. Each square represents a biological male, while each circle represents a biological female. The Y-STRs are passed on through patrilineal inheritance, which means they are inherited from father to son.

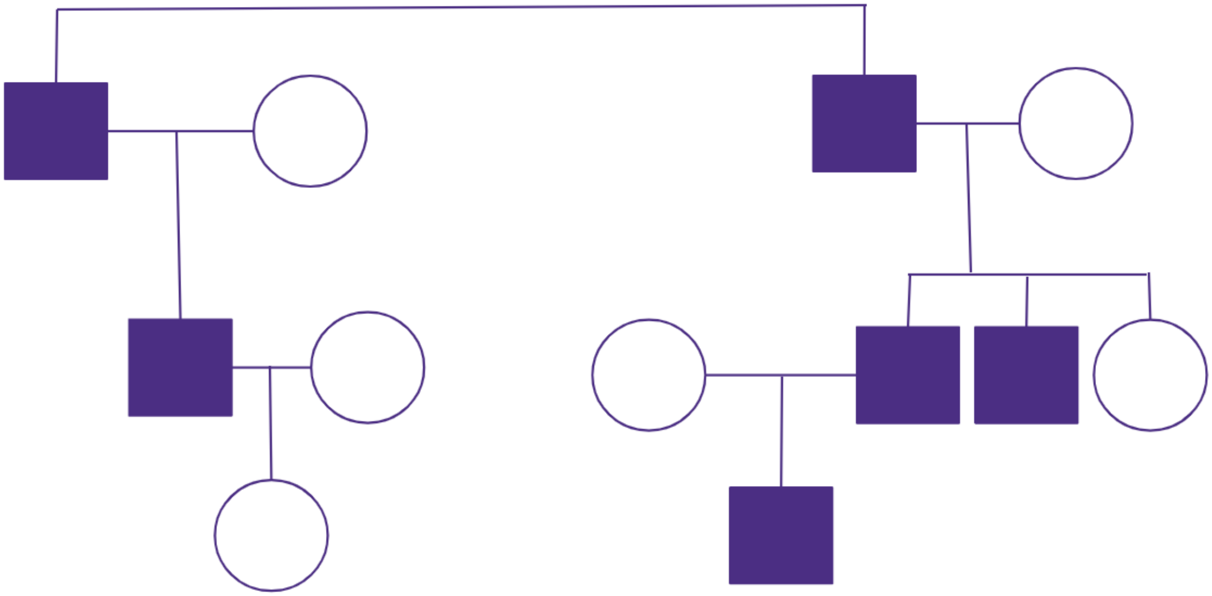
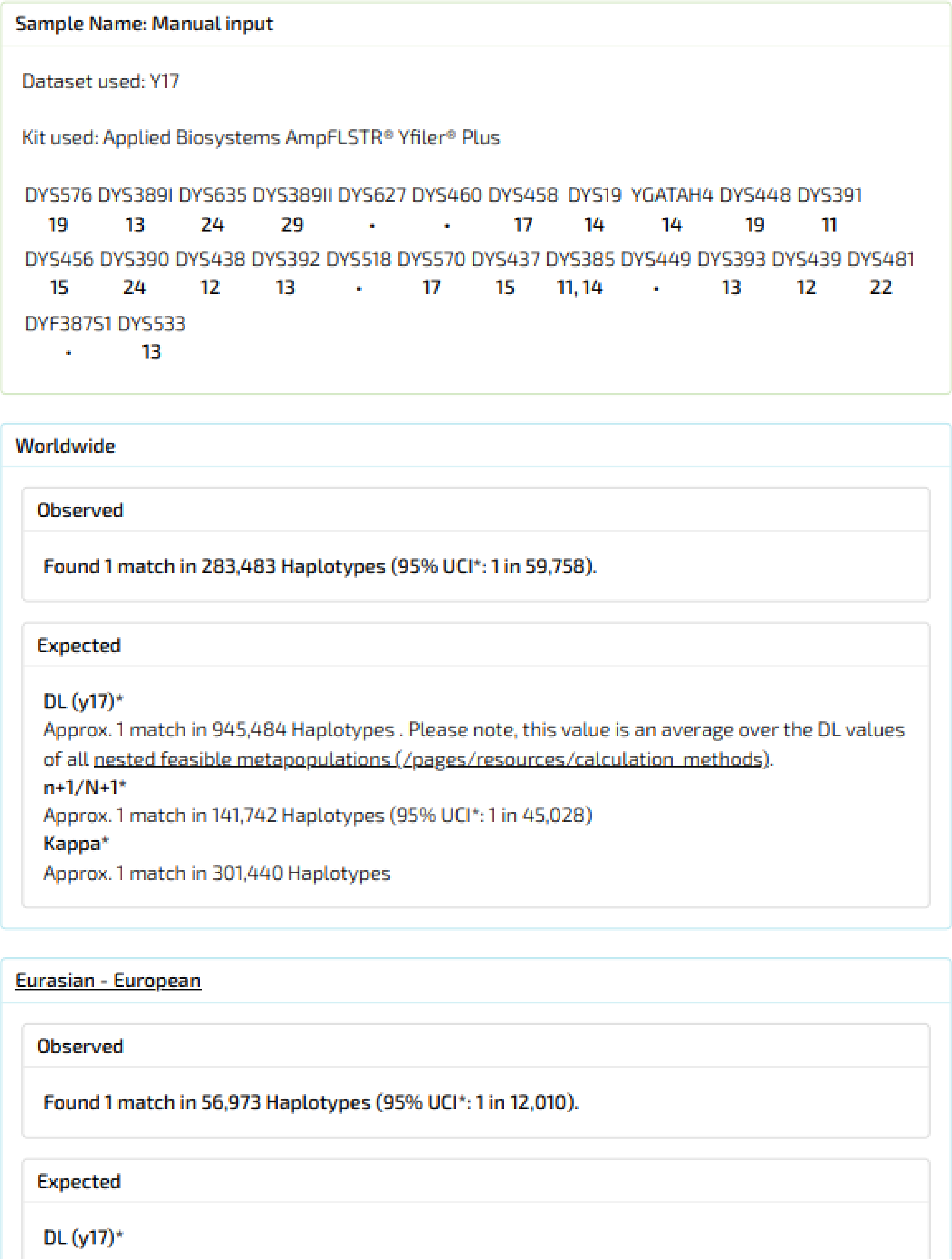


Fig 2. YHRD Search Report for a Belgian Profile using Y17's YFiler Plus



Sample Name: Manual input

Dataset used: Y17

Kit used: Applied Biosystems AmpFLSTR® Yfiler® Plus

DYS576	DYS389I	DYS635	DYS389II	DYS627	DYS460	DYS458	DYS19	YGATAH4	DYS448	DYS391
19	**13**	**24**	**29**	**·**	**·**	**17**	**14**	**14**	**19**	**11**

DYS456	DYS390	DYS438	DYS392	DYS518	DYS570	DYS437	DYS385	DYS449	DYS393	DYS439	DYS481
15	**24**	**12**	**13**	**·**	**17**	**15**	**11, 14**	**·**	**13**	**12**	**22**

DYF387S1	DYS533
·	**13**

Worldwide

Observed

Found 1 match in 283,483 Haplotypes (95% UCI*: 1 in 59,758).

Expected

DL (y17)*
Approx. 1 match in 945,484 Haplotypes . Please note, this value is an average over the DL values of all nested feasible metapopulations (/pages/resources/calculation_methods).
n+1/N+1*
Approx. 1 match in 141,742 Haplotypes (95% UCI*: 1 in 45,028)
Kappa*
Approx. 1 match in 301,440 Haplotypes

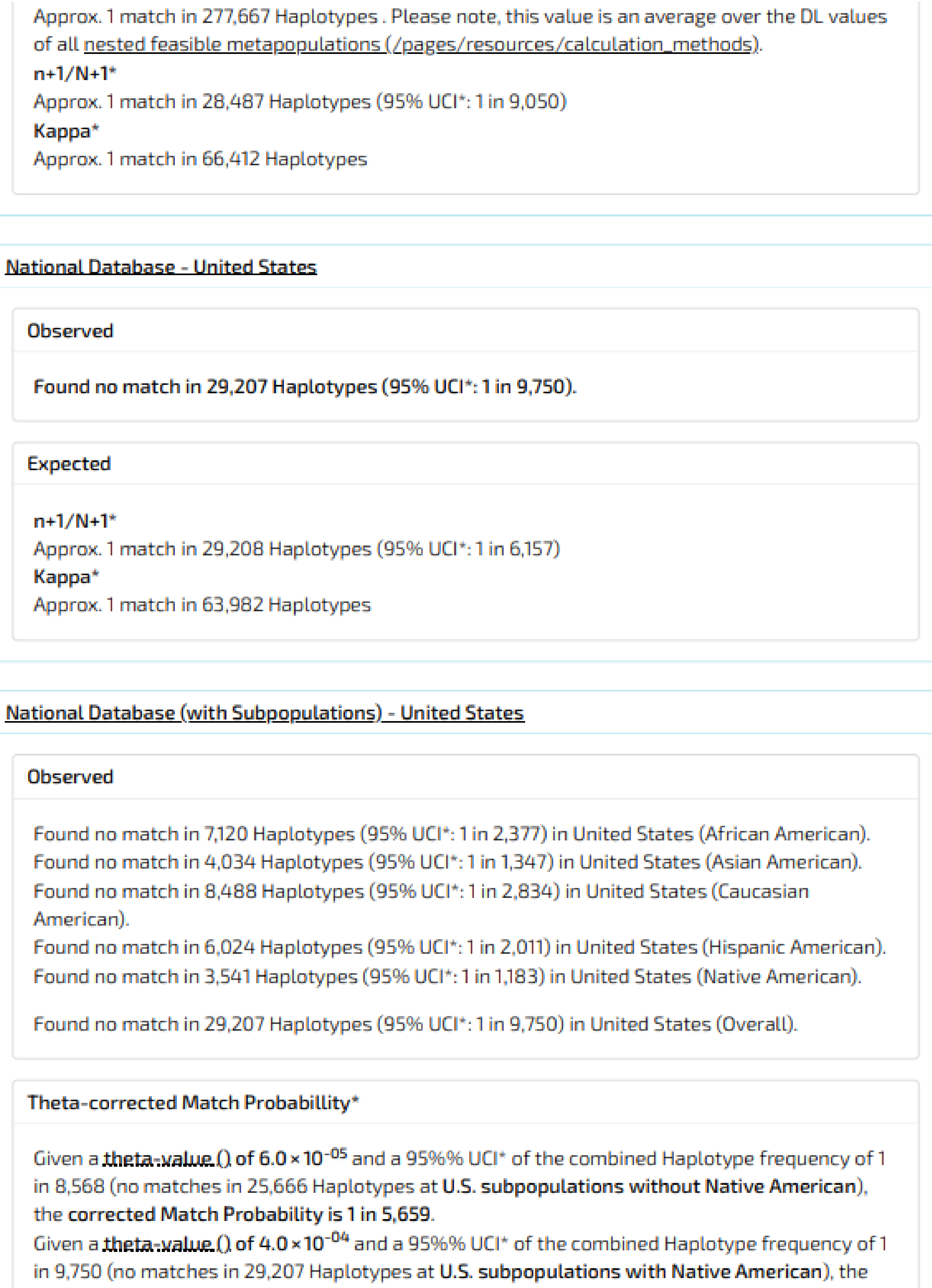
Eurasian - European

Observed

Found 1 match in 56,973 Haplotypes (95% UCI*: 1 in 12,010).

Expected

DL (y17)*



Approx. 1 match in 277,667 Haplotypes . Please note, this value is an average over the DL values of all nested feasible metapopulations (/pages/resources/calculation_methods).

n+1/N+1*

Approx. 1 match in 28,487 Haplotypes (95% UCI*: 1 in 9,050)

Kappa*

Approx. 1 match in 66,412 Haplotypes

National Database - United States

Observed

Found no match in 29,207 Haplotypes (95% UCI*: 1 in 9,750).

Expected

n+1/N+1*

Approx. 1 match in 29,208 Haplotypes (95% UCI*: 1 in 6,157)

Kappa*

Approx. 1 match in 63,982 Haplotypes

National Database (with Subpopulations) - United States

Observed

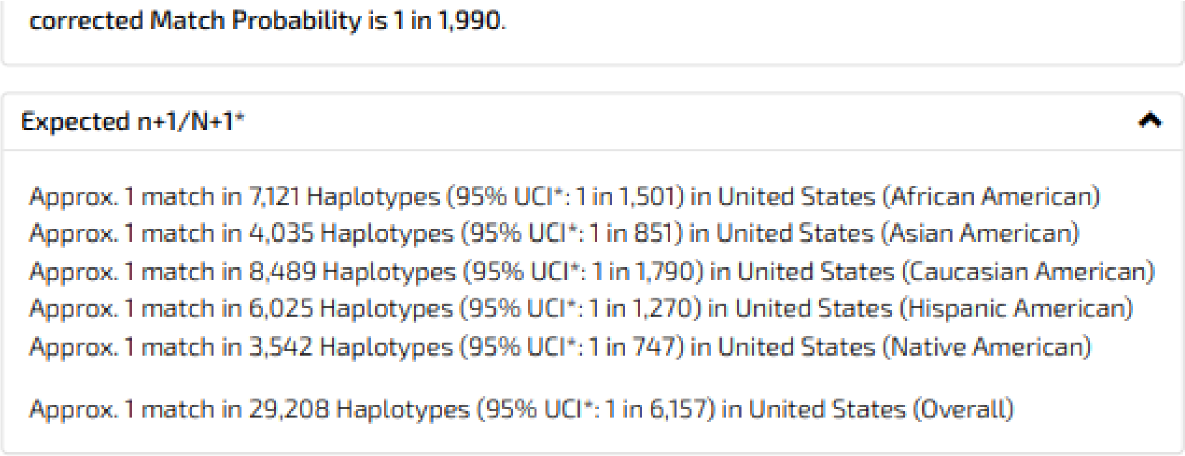
Found no match in 7,120 Haplotypes (95% UCI*: 1 in 2,377) in United States (African American).
Found no match in 4,034 Haplotypes (95% UCI*: 1 in 1,347) in United States (Asian American).
Found no match in 8,488 Haplotypes (95% UCI*: 1 in 2,834) in United States (Caucasian American).
Found no match in 6,024 Haplotypes (95% UCI*: 1 in 2,011) in United States (Hispanic American).
Found no match in 3,541 Haplotypes (95% UCI*: 1 in 1,183) in United States (Native American).

Found no match in 29,207 Haplotypes (95% UCI*: 1 in 9,750) in United States (Overall).

Theta-corrected Match Probabillity*

Given a theta-value () **of 6.0×10^{-05}** and a 95%% UCI* of the combined Haplotype frequency of 1 in 8,568 (no matches in 25,666 Haplotypes at **U.S. subpopulations without Native American**), the **corrected Match Probability is 1 in 5,659**.

Given a theta-value () **of 4.0×10^{-04}** and a 95%% UCI* of the combined Haplotype frequency of 1 in 9,750 (no matches in 29,207 Haplotypes at **U.S. subpopulations with Native American**), the



corrected Match Probability is 1 in 1,990.

Expected n+1/N+1*

Approx. 1 match in 7,121 Haplotypes (95% UCI*: 1 in 1,501) in United States (African American)
Approx. 1 match in 4,035 Haplotypes (95% UCI*: 1 in 851) in United States (Asian American)
Approx. 1 match in 8,489 Haplotypes (95% UCI*: 1 in 1,790) in United States (Caucasian American)
Approx. 1 match in 6,025 Haplotypes (95% UCI*: 1 in 1,270) in United States (Hispanic American)
Approx. 1 match in 3,542 Haplotypes (95% UCI*: 1 in 747) in United States (Native American)

Approx. 1 match in 29,208 Haplotypes (95% UCI*: 1 in 6,157) in United States (Overall)

Tables

Table 1: YHRD Database Search Results for 10 Profiles Obtained from Purps using Y17's YFiler Plus and Y23's PPY 23

		Worldwide	Metapopulation	US National Database	US National Database with subpopulations	
Country and Marker	**Population**	**Observed**[2]	**Observed**	**Observed**	**Observed Subpopulation(s)**	**Theta Correction**[3]
Belgium; Y17	Belgian	n(N): 1(284,483) u(U): 1(59,758)	n(N): 1(56,973) u(U): 1(9,050) (Eurasian - European)	NA	NA	Theta: 6.0 × 10-05 n(N): & 1(8,568) MPn: 1(5,659) Theta: 4.0 × 10-04 m(M): 1(9,750) MPm: 1(1,990)
Belgium; Y23	Belgian	n(N): 1(99,962) u(U): 1(21,072)	n(N): 1(25,350) u(U): 1(5,344) (Eurasian - European)	NA	NA	Theta: 2.0 × 10-05 n&(N): 1(4,413) MPn: 1(4,055) Theta:3.0 × 10-04 m(M): 1(5,471) MPm: 1(2,072)
Spain; Y17	Spanish	n(N): 2(283,483)	n(N): 2(56,973)	NA	NA	Theta: 6.0 × 10-05

---

[2] For each observed column in worldwide, metapopulation, and US national databases, n is the observed match number and (N) is the number of profiles searched; u(U) is the observed match number in the 95% upper confidence interval.

[3] n(N) is the 95% upper confidence interval and MPn is the corrected match probability for the population without Native American; m(M) is the 95% upper confidence interval and MPm is the corrected match probability for the population with Native American.

		u(U): 1(45,028)	u(U): 1(9,050) (Eurasian - Europe)			n(N): 1(8,568) MPn: 1(5,659) Theta: 4.0 × 10-04 m(M): 1(9,750) MPm: 1(1,990)
Spain; Y23	Spanish	n(N): 1(99,962) u(U): 1(21,072)	n(N): 1(12,676) u(U): 1(4,027) (Eurasian - European)	NA	NA	Theta: 2.0 × 10-05 n(N): 1(4,413) MPn: 1(4,055) Theta: 3.0 × 10-04 m(M): 1(5,471) MPm: 1(2,072)
Bolivia; Y17	Mestizo	n(N): 18(283,483) u(U): 1(10,621)	n(N): 9(48,337) u(U): 1(3,078) (Admixed)	n(N): 5(29,207) u(U): 1(2,778)	n(N): 1(7,120) u(U): 1(1,501) (African American) n(N): 3(8,488) u(U): 1(1,095) (Caucasian American) n(N): 1(6,024) u(U): 1(1,270) (HispanicAmerican)	Theta: 6.0 × 10-05 n(N): 1(2,442) MPn: 1(2,130) Theta: 4.0 × 10-04 m(M): 1(2,778) MPm: 1(1,316)
Bolivia; Y23	Mestizo	n(N): 1(99,962) u(U): 1(21,072)	n(N): 1(20,878) u(U): 1(4,401) (Admixed)	NA	NA	Theta: 2.0 × 10-05 n(N): 1(4,413) MPn: 1(4,055) Theta: 3.0 × 10-04 m(M): 1(5,471) MPm: 1(2,072)
Bolivia; Y17	Native	n(N): 2(283,	NA	NA	NA	Theta: 6.0 × 10-05

	American	483) u(U): 1(45,028)	(Admixed - Native American)			n(N): 1(8,568) MPn: 1(5,659) Theta: 4.0 × 10-04 m(M): 1(9,750) MPm: 1(1,990)
Bolivia; Y23	Native American	n(N): 1(99,962) u(U): 1(21,072)	NA (Admixed - Native American)	NA	NA	Theta: 2.0 × 10-05 n(N): 1(4,413) MPn: 1(4,055) Theta: 3.0 × 10-04 m(M): 1(5,471) MPm: 1(2,072)
USA; Y17	Athapaskan	n(N): 1(283,483) u(U): 1(59,758)	n(N): 1(3,541) u(U): 1(747) (Admixed - Native American)	n(N): 1(29,207) u(U): 1(6,157)	Same as metapopulation and US national databases	Theta: 6.0 × 10-05 n(N): 1(8,568) MPn: 1(5,659) Theta: 4.0 × 10-04 m(M): 1(6,157) MPm: 1(1,778)
USA; Y23	Athapaskan	n(N): 1(99,962) u(U): 1(21,072)	n(N): 1(3,168) u(U): 1(668) (Admixed - Native American)	n(N) 1(16,388) u(U): 1(3,455)	Same as metapopulation and US national databases	Theta: n(N): 2.0 × 10-05 & 1(4,413) MPn: 1(4,055) Theta: 3.0 × 10-04 m(M): 1(3,455) MPm: 1(1,697)
Japan; Y17	Japanese	n(N): 7(283,483) u(U): 1(21,561)	n(N): 3(1,909) u(U): 1(247) (East Asian - Japanese)	NA	NA	Theta: n(N): 6.0 × 10-05 & 1(8,568) MPn: 1(5,659) Theta: 4.0 × 10-04

						m(M): 1(9,750) MPm: 1(1,990)
Japan; Y23	Japanese	n(N): 1 (99,962) u(U): 1(21,072)	n(N): 1(32) u(U): 1(68) (East Asian - Japanese)	NA	NA	Theta: 2.0 × 10-05 n(N): 1(4,413) MPn: 1(4,055) Theta: 3.0 × 10-04 m(M): 1(5,471) MPm: 1(2,072)
Vietnam; Y17	Vietnamese	n(N): 4(283,483) u(U): 1(30,970)	n(N): 1(4,034) u(U): 1(851) (Admixed - Asian American)	n(N): 1(29,207) u(U): 1(6,157)	Same as metapopulation and US national databases	Theta: 6.0 × 10-05 n(N): 1(5,411) MPn: 1(4,085) Theta: 4.0 × 10-04 m(M): 1(6,157) MPm: 1(1,778)
Vietnam; Y23	Vietnamese	n(N): 2(99,962) u(U): 1(15,878)	n(N): 1(3,149) u(U): 1(664) (Admixed - Asian American)	n(N): 1(16,388) u(U): 1(3,455)	Same as metapopulation and US national databases	Theta: 2.0 × 10-05 n(N): 1(2,787) MPn: 1(2,640) Theta: 3.0 × 10-04 m(M): 1(3,455) MPm: 1(1,697)
USA; Y17	African	n(N): 3(283,483) u(U): 1(36,562)	n(N): 2(7,120) u(U): 1(1,131) (Admixed - African American)	n(N): 2(29,207) u(U): 1(4,639)	Same as metapopulation and US national databases	Theta: 6.0 × 10-05 n(N): 1(4,077) MPn: 1(3,276) Theta: 4.0 × 10-04 m(M): 1(4,639) MPm: 1(1,625)
USA; Y23	African	n(N): 2(99,962)	n(N): 1(3,289)	n(N): 1(3,289)	Same as	Theta: 2.0 × 10-05

		u(U): 1(15,878)	u(U): 1(694) (Admixed - African American)	u(U): 1(694)	metapopulation and US national databases	n(N): 1(2,787) MPn: 1(2,640) Theta: 3.0 × 10-04 m(M): 1(3,455) MPm: 1(1,697)
USA; Y17	European	n(N): 1(283,483) u(U): 1(59,758)	n(N): 1(8,488) u(U): 1(1,790) (Admixed - Caucasian American)	n(N): 1(29,207) u(U): 1(6,157)	Same as metapopulation and US national databases	Theta: 6.0 × 10-05 n(N): 1(5,411) MPn: 1(4,085) Theta: 4.0 × 10-04 m(M): 1(6,157) MPm: 1(1,778)
USA; Y23	European	n(N): 1(99,962) u(U): 1(21,072)	n(N): 1(3,625) u(U): 1(765) (Admixed - Caucasian American)	n(N): 1(16,388) u(U): 1(3,455)	Same as metapopulation and US national databases	Theta: 2.0 × 10-05 n(N): 1(2,787) MPn: 1(2,640) Theta: 3.0 × 10-04 m(M): 1(3,455) MPm: 1(1,697)
USA; Y17	Hispanic	n(N): 14(283,483) u(U): 1(12,953)	n(N): 1(6,024) u(U): 1(1,270) (Admixed - Hispanic American)	n(N): 1(29,207) u(U): 6,157	Same as metapopulation and US national databases	Theta: 6.0 × 10-05 n(N): 1(5,411) MPn: 1(4,085) Theta: 4.0 × 10-04 m(M): 1(6,157) MPm: 1(1,778)
USA; Y23	Hispanic	n(N): 1(99,962) u(U): 1(21,072)	n(N): 1(3,157) u(U): 1(666) (Admixed - Hispanic	n(N): 1(16,388) u(U): 1(3,455)	Same as metapopulation and US national databases	Theta: 2.0 × 10-05 n(N): 1(2,787) MPn: 1(2,640)

			American)			Theta: 3.0 × 10-04 m(M):1(3,455) MPm: 1(1,697)

Table 2 YHRD Database Search Results for 10 Profiles Obtained from Purps using Y23's PPY23

		Worldwide	Metapopulation	US National Database	US National Database with subpopulations	
Country	**Population**	**Observed**[4]	**Observed**	**Observed**	**Observed Subpopulation(s)**	**Theta Correction**[5]
Belgium	Belgian	n(N): 1(99,962) u(U): 1(21,072)	n(N): 1(25,350) u(U): 1(5,344) (Eurasian - European)	NA	NA	Theta: 2.0 × 10-05 n&(N): 1(4,413) MPn: 1(4,055) Theta:3.0 × 10-04 m(M): 1(5,471) MPm: 1(2,072)
Spain	Spanish	n(N): 1(99,962) u(U): 1(21,072)	1n(N): 1(12,676) u(U): 1(4,027) (Eurasian - European)	NA	NA	Theta: 2.0 × 10-05 n(N): 1(4,413) MPn: 1(4,055) Theta: 3.0 × 10-04 m(M): 1(5,471) MPm: 1(2,072)
Bolivia	Mestizo	n(N): 1(99,962) u(U): 1(21,072)	n(N): 1(20,878) u(U): 1(4,401) (Admixed)	NA	NA	Theta: 2.0 × 10-05 n(N): 1(4,413) MPn: 1(4,055) Theta: 3.0 × 10-04 m(M): 1(5,471)

---

[4] For each observed column in worldwide, metapopulation, and US national databases, n is the observed match number and (N) is the number of profiles searched; u(U) is the observed match number in the 95% upper confidence interval.

[5] n(N) is the 95% upper confidence interval and MPn is the corrected match probability for the population without Native American; m(M) is the 95% upper confidence interval and MPm is the corrected match probability for the population with Native American.

						MPm: 1(2,072)
Bolivia	Native American	n(N): 1(99,962) u(U): 1(21,072)	NA (Admixed - Native American)	NA	NA	Theta: 2.0 × 10-05 n(N): 1(4,413) MPn: 1(4,055) Theta: 3.0 × 10-04 m(M): 1(5,471) MPm: 1(2,072)
USA	Athapaskan	n(N): 1(99,962) u(U): 1(21,072)	n(N): 1(3,168) u(U): 1(668) (Admixed - Native American)	n(N) 1(16,388) u(U): 1(3,455)	Same as metapopulation and US national databases	Theta: n(N): 2.0 × 10-05 & 1(4,413) MPn: 1(4,055) Theta: 3.0 × 10-04 m(M): 1(3,455) MPm: 1(1,697)
Japan	Japanese	n(N): 1 (99,962) u(U): 1(21,072)	n(N): 1(32) u(U): 1(68) (East Asian - Japanese)	NA	NA	Theta: 2.0 × 10-05 n(N): 1(4,413) MPn: 1(4,055) Theta: 3.0 × 10-04 m(M): 1(5,471) MPm: 1(2,072)
Vietnam	Vietnamese	n(N): 2(99,962) u(U): 1(15,878)	n(N): 1(3,149) u(U): 1(664) (Admixed - Asian American)	n(N): 1(16,388) u(U): 1(3,455)	Same as metapopulation and US national databases	Theta: 2.0 × 10-05 n(N): 1(2,787) MPn: 1(2,640) Theta: 3.0 × 10-04 m(M): 1(3,455) MPm: 1(1,697)
USA	African	n(N): 2(99,962) u(U): 1(15,878)	n(N): 1(3,289) u(U): 1(694)	n(N): 1(3,289) u(U): 1(694)	Same as metapopulation and US national databases	Theta: 2.0 × 10-05 n(N): 1(2,787)

			(Admixed - African American)			MPn: 1(2,640) Theta: 3.0 × 10-04 m(M): 1(3,455) MPm: 1(1,697)
USA	European	n(N): 1(99,962) u(U): 1(21,072)	n(N): 1(3,625) u(U): 1(765) (Admixed - Caucasian American)	n(N): 1(16,388) u(U): 1(3,455)	Same as metapopulation and US national databases	Theta: 2.0 × 10-05 n(N): 1(2,787) MPn: 1(2,640) Theta: 3.0 × 10-04 m(M): 1(3,455) MPm: 1(1,697)
USA	Hispanic	n(N): 1(99,962) u(U): 1(21,072)	n(N): 1(3,157) u(U): 1(666) (Admixed - Hispanic American)	n(N): 1(16,388) u(U): 1(3,455)	Same as metapopulation and US national databases	Theta: 2.0 × 10-05 n(N): 1(2,787) MPn: 1(2,640) Theta: 3.0 × 10-04 m(M):1(3,455) MPm: 1(1,697)

Table 3 YHRD Database Search Results for 10 Profiles Obtained from Purps using Y17's YFiler Plus

		Worldwide	**Metapopulation**	**US National Database with subpopulations**		
Country	**Population**	**Observed[6]**	**Observed**	**Observed**	**Observed Subpopulation(s)**	**Theta Correction[7]**
Belgium	Belgian	n(N): 1(284,483) u(U): 1(59,758)	n(N): 1(56,973) u(U): 1(9,050) (Eurasian - European)	NA	NA	Theta: 6.0 × 10-05 n(N): & 1(8,568) MPn: 1(5,659) Theta: 4.0 × 10-04 m(M): 1(9,750) MPm: 1(1,990)
Spain	Spanish	n(N): 2(283,483) u(U): 1(45,028)	n(N): 2(56,973) u(U): 1(9,050) (Eurasian - European)	NA	NA	Theta: 6.0 × 10-05 n(N): 1(8,568) MPn: 1(5,659) Theta: 4.0 × 10-04 m(M): 1(9,750) MPm: 1(1,990)
Bolivia	Mestizo	n(N): 18(283,483) u(U): 1(10,621)	n(N): 9(48,337) u(U): 1(3,078) (Admixed)	n(N): 5(29,207) u(U): 1(2,778)	n(N): 1(7,120) u(U): 1(1,501) (African American) n(N): 3(8,488) u(U): 1(1,095) (Caucasian American)	Theta: 6.0 × 10-05 n(N): 1(2,442) MPn: 1(2,130) Theta: 4.0 × 10-04 m(M): 1(2,778) MPm: 1(1,316)

---

[6] For each observed column in worldwide, metapopulation, and US national databases, n is the observed match number and (N) is the number of profiles searched; u(U) is the observed match number in the 95% upper confidence interval.

[7] n(N) is the 95% upper confidence interval and MPn is the corrected match probability for the population without Native American; m(M) is the 95% upper confidence interval and MPm is the corrected match probability for the population with Native American.

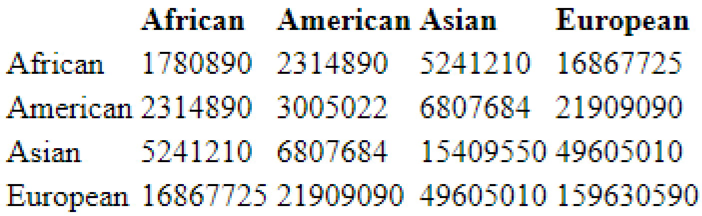
					n(N): 1(6,024) u(U): 1(1,270) (HispanicAmerican)	
Bolivia	Native American	n(N): 2(283, 483) u(U): 1(45,028)	NA (Admixed - Native American)	NA	NA	Theta: 6.0 × 10-05 n(N): 1(8,568) MPn: 1(5,659) Theta: 4.0 × 10-04 m(M): 1(9,750) MPm: 1(1,990)
USA	Athapaskan	n(N): 1(283,483) u(U): 1(59,758)	n(N): 1(3,541) u(U): 1(747) (Admixed - Native American)	n(N): 1(29,207) u(U): 1(6,157)	Same as metapopulation and US national databases	Theta: 6.0 × 10-05 n(N): 1(8,568) MPn: 1(5,659) Theta: 4.0 × 10-04 m(M): 1(6,157) MPm: 1(1,778)
Japan	Japanese	n(N): 7(283,483) u(U): 1(21,561)	n(N): 3(1,909) u(U): 1(247) (East Asian - Japanese)	NA	NA	Theta: n(N): 6.0 × 10-05 & 1(8,568) MPn: 1(5,659) Theta: 4.0 × 10-04 m(M): 1(9,750) MPm: 1(1,990)
Vietnam	Vietnamese	n(N): 4(283,483) u(U): 1(30,970)	n(N): 1(4,034) u(U): 1(851) (Admixed - Asian American)	n(N): 1(29,207) u(U): 1(6,157)	Same as metapopulation and US national databases	Theta: 6.0 × 10-05 n(N): 1(5,411) MPn: 1(4,085) Theta: 4.0 × 10-04 m(M): 1(6,157) MPm: 1(1,778)

USA	African	n(N): 3(283,483) u(U): 1(36,562)	n(N): 2(7,120) u(U): 1(1,131) (Admixed - African American)	n(N): 2(29,207) u(U): 1(4,639)	Same as metapopulation and US national databases	Theta: 6.0 × 10-05 n(N): 1(4,077) MPn: 1(3,276) Theta: 4.0 × 10-04 m(M): 1(4,639) MPm: 1(1,625)
USA	European US	n(N): 1(283,483) u(U): 1(59,758)	n(N): 1(8,488) u(U): 1(1,790) (Admixed - Caucasian American)	n(N): 1(29,207) u(U): 1(6,157)	Same as metapopulation and US national databases	Theta: 6.0 × 10-05 n(N): 1(5,411) MPn: 1(4,085) Theta: 4.0 × 10-04 m(M): 1(6,157) MPm: 1(1,778)
USA	Hispanic US	n(N): 14(283,483) u(U): 1(12,953)	n(N): 1(6,024) u(U): 1(1,270) (Admixed - Hispanic American)	n(N): 1(29,207) u(U): 6,157	Same as metapopulation and US national databases	Theta: 6.0 × 10-05 n(N): 1(5,411) MPn: 1(4,085) Theta: 4.0 × 10-04 m(M): 1(6,157) MPm: 1(1,778)

The following tables (4-7) do not use YHRD, but only Purps data.

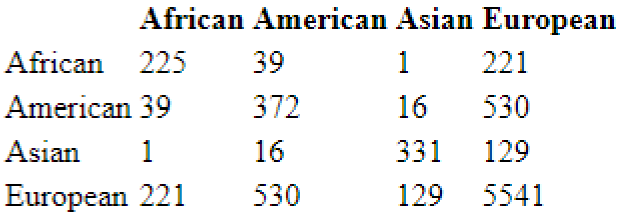
Acknowledge help by Michael Yung.

Table 4 Numbers of Pairs Compared Between and Within Populations in the Yfiler kit Using 17 Y-STR Markers



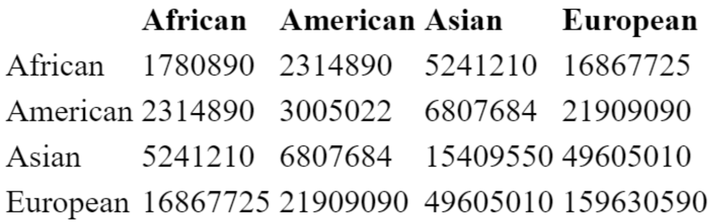
	African	American	Asian	European
African	1780890	2314890	5241210	16867725
American	2314890	3005022	6807684	21909090
Asian	5241210	6807684	15409550	49605010
European	16867725	21909090	49605010	159630590

Table 5 Numbers of Matching Pairs Between and Within Populations in the Yfiler Kit Using 17 Y-STR Markers



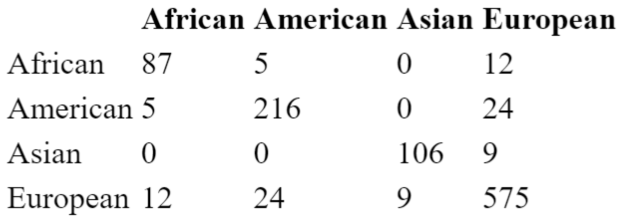
	African	American	Asian	European
African	225	39	1	221
American	39	372	16	530
Asian	1	16	331	129
European	221	530	129	5541

Table 6 Numbers of Pairs Compared Between and Within Populations in the Yfiler Kit Using 23 Y-STR Markers



	African	American	Asian	European
African	1780890	2314890	5241210	16867725
American	2314890	3005022	6807684	21909090
Asian	5241210	6807684	15409550	49605010
European	16867725	21909090	49605010	159630590

Table 7 Numbers of Matching Pairs Between and Within Populations in the Yfiler Kit Using 23 Y-STR Markers



	African	**American**	**Asian**	**European**
African	87	5	0	12
American	5	216	0	24
Asian	0	0	106	9
European	12	24	9	575

Appendices

Appendix 1: 10 Y-STR Profile Identifications Obtained from Purps

Population	**ID**
Antwerpen, Belgium	ID007
Aragon, Spain [Spanish]	ARA001
Bolivia [Mestizo]	BO14
Bolivia [Native American]	BO05
Central Alaska, USA [Athapaskan]	C.10113.S
Gunma, Japan [Japanese]	11-001
Ho Chi Minh City, Vietnam [Kinh]	HG01596
Illinois, USA [African American]	1000
Illinois, USA [European American]	1002
Illinois, USA [Hispanic American]	100

Appendix 2: Y-STR Commercial Tool Kit

Kit	**No. of Y-STR loci**
Yfiler Plus	17
PowerPlex 23 (PPY 23)	23