Case report

Validation of a combined autosomal/Y-chromosomal STR approach for analyzing typical biological stains in sexual-assault cases

Josephine Purps, Maria Geppert, Marion Nagy, Lutz Roewer*

Department Forensic Genetics, Institute of Legal Medicine and Forensic Sciences, Charité-Universitätsmedizin Berlin, Germany

ABSTRACT

DNA testing is an established part of the investigation and prosecution of sexual assault. The primary purpose of DNA evidence is to identify a suspect and/or to demonstrate sexual contact. However, due to highly uneven proportions of female and male DNA in typical stains, routine autosomal analysis often fails to detect the DNA of the assailant. To evaluate the forensic efficiency of the combined application of autosomal and Y-chromosomal short tandem repeat (STR) markers, we present a large retrospective casework study of probative evidence collected in sexual-assault cases. We investigated up to 39 STR markers by testing combinations of the 16-locus NGMSelect kit with both the 23-locus PowerPlex Y23 and the 17-locus Yfiler kit. Using this dual approach we analyzed DNA extracts from 2077 biological stains collected in 287 cases over 30 months. To assess the outcome of the combined approach in comparison to stand-alone autosomal analysis we evaluated informative DNA profiles. Our investigation revealed that Y-STR analysis added up to 21% additional, highly informative (complete, single-source) profiles to the set of reportable autosomal STR profiles for typical stains collected in sexual-assault cases. Detection of multiple male contributors was approximately three times more likely with Y-chromosomal profiling than with autosomal STR profiling. In summary, 1/10 cases would have remained inconclusive (and could have been dismissed) if Y-STR analysis had been omitted from DNA profiling in sexual-assault cases. © 2015 The Authors. Published by Elsevier Ireland Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Autosomal short tandem repeat (STR) analysis is a central method in forensic laboratories due to its individualizing potential and the existence of national DNA databases of standardized STR profiles from autosome alone. Other methods (e.g. mitochondrial or Y-chromosomal DNA analysis) have restricted applications; for example Y-STR analyses are highly valuable for investigations of sexually motivated crimes in which the minor male proportion in DNA mixtures frequently remains undetected in standard analyses [1,2]. These markers trace male cells even when long periods of time have passed since the assault [3–6]. Moreover, Y-STR typing is useful for the determination of the number of donors in mixtures from multiple male contributors [7,8].

Nevertheless, most high-throughput forensic laboratories are often reluctant to establish workflows for sexually motivated crimes that may require time-consuming and cost-intensive additional or alternative typing steps and may demand special training for correct interpretation of Y-STR results. The lower power of exclusion and the lack of Y profiles in national police databases are serious obstacles for the application of Y-STRs. However, several essential prerequisites for widespread application of Y-STR analysis are currently available: commercial kits with high resolution and high sensitivity, acceptance of the technique and the resulting biostatistics in the scientific community, hundreds of publications and large population databases. New, highly discriminative Y-STR markers have been evaluated [9–11]; some of these have been implemented in current-generation Y-STR kits like the PowerPlex® Y23 System (Promega Corp., Mannheim, Germany) [12]. A recent multicenter study confirmed the significantly increased discriminatory power of this system compared to the AmpFLSTR® Yfiler® PCR Amplification Kit (Life Technologies GmbH, Darmstadt, Germany) [13]. Nevertheless, systematic casework studies of the value of current Y-STR kits are lacking.

Here we present a retrospective study of 287 sexual-assault cases in which 2077 stains were analyzed with up to 39 autosomal and Y-STR markers via the application of two commercial kits. Approximately 10% of these cases were analyzed with the AmpFLSTR® NGM Select™ PCR Amplification Kit (Life Technologies GmbH, Darmstadt, Germany) and the Yfiler kit and 90% were analyzed with the NGM Select kit and the PowerPlex Y23. Based on the number of male profiles with negative autosomal but positive Y chromosomal results

* Corresponding author. Fax: +49 30 450 525912.
E-mail address: lutz.roewer@charite.de (L. Roewer).

http://dx.doi.org/10.1016/j.fsgen.2015.08.002
1872–4073 © 2015 The Authors. Published by Elsevier Ireland Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
we estimated the proportion of cases for which information on male donor(s) and subsequent investigative leads would have been lost if Y-STRs had been excluded from analysis.

2. Material and methods

2.1. Forensic specimen and characterization of DNA samples

Over 30 months (2012–2014), we collected DNA profiling results from 287 cases and 2077 stains that were analyzed via a dual workflow with both autosomal and Y-chromosomal STRs. The cases, which were generally classified as sexual assaults, were legally categorized according to the German Criminal Code (Chapter 13; §§174–184) (Fig. 1). The majority of offenses in this study belonged to the category “Sexual coercion and rape” (74%; 211/287), followed by “Sexual abuse of children” (13%; 37/287). The definition of rape differs between countries and within federal countries even between states but generally penetration of the victim’s body is defined as the salient element of rape. Sexual coercion is the act of using subtle pressure, drugs, alcohol or force to have sexual contact with someone against their will. Approximately 58% of all cases in this study (165/287) were sexually motivated attacks by individual or group offenders without penile penetration. Typical evidence items from coerced women included touch samples, e.g. epithelial cells from clothing. Of these cases, 58/287 (20%) were characterized as attempted rape (§§174,176,177); in most of these cases, penetration was digital or stopped before penile penetration. In many cases the victim was reported to be unconscious or disoriented due to the application/consumption of drugs or alcohol or due to the violent attack itself, and the victim was not able to confirm penile penetration. The proportion of cases with manifest penile penetration and ejaculation was therefore small in our incoming casework (~2%; 5/287).

For all cases of suspected rape, swabs collected from the interior and exterior of the body of the victim were mandatorily provided by medical examiners (20% of our incoming evidence, i.e. 423/2077 stains). A further 14% of the stains (296/2077) were collected by police examiners at the crime scene using swabs or adhesive tapes, including hair samples. Approximately 66% (1358/2077) of the analyzed stains were collected at the Department of Forensic Genetics from the victim’s clothing (88%; 1197/1358) or from evidentiary items like cigarette butts, bottles, or drinking glasses (12%; 163/1358). Approximately 69% (937/1358) of the stains collected in-house were touch DNA (epithelial cells from clothing and items) and 31% (421/1358) were suspected bodily fluids (e.g. semen, blood, saliva). Fig. 2 depicts the entire set of samples collected externally or in-house. On average 6–8 stains per case are collected; in the rape cases evaluated here, the average number was 15. We used cutting and swabbing for bodily fluids and mostly swabbing for touch stains. Where necessary, presumptive tests for semen, blood, or saliva were conducted using RSID® assays (Galantos Genetics, Mainz, Germany), the rapid benzidine test for blood (FERAK, Berlin, Germany), and screening with the SUPERLITE M 05 multispectral light source (Lumatec, Deisenhofen, Germany). In 53% of our cases (152/287), a reference sample from the victim was available; male reference samples were provided in 28% of the cases (80/287), which means that no-suspect cases prevailed in our dataset.

2.2. DNA extraction

DNA was extracted via a magnetic-bead method on the QiaSymphony robot station (QIAGEN, Hilden, Germany) using a 96 well PCR plate. Microscopy was not used to determine the cell types present in swabs, and thus swabs were treated uniformly without differential lysis or other cell-specific extraction protocols. Before extraction case samples were lysed in 500 μL ATL lysis buffer (QIAGEN) with proteinase K (final concentration 1:25) and incubated for at least 15 min at 56°C. For other samples such as hair, nail, and sperm, 20 μL dithiothreitol was added and samples were incubated for one day at 56°C. Highly concentrated DNA samples from blood and saliva were pre-diluted. The extraction volume was 200 μL and the elution volume was 100 μL. When necessary, DNA was concentrated in a Thermo Scientific™ Savant™ Universal SpeedVac™ Vacuum System. No quantification step was included in the workflow.

2.3. DNA typing and evaluation of profiling data

We applied a combination of the NGM SElect™ and PowerPlex® Y23 kits (257 cases with 1801 stains) as well as a

Fig. 1. Legal classification of offences against sexual self-determination according to the German Criminal Code, Chapter 13, Sections 174–185.
combination of the NGM SElect™ and Yfiler® kits (30 cases with 276 stains). The NGM SElect PCR reaction mix had a final volume of 17.5 μL which includes 7.5 μL DNA extract. PCR was performed with 29 cycles [14]. For the PowerPlex® Y23 [15], the final volume was reduced to 10 μL and the maximum DNA input was 7 μL. For the Yfiler® kit [16], 7.5 μL DNA were added to each reaction in a final amplification volume of 17.5 μL. Thermal cycling was performed according to the manufacturer’s recommendations using 30 cycles for both Y-STR multiplexes. Amplified products were separated and detected on an ABI PRISM® 3130xl Genetic Analyzer (Life Technologies GmbH) following the manufacturer’s protocols. Autosomal PCR samples (1.5 μL) or ladder were mixed with 20 μL HiDi formamide (Life Technologies, Darmstadt, Germany) and a 0.8 μL size standard (LIZ 600; Life Technologies, GmbH). One microliter of the Y-STR PCR products or ladder was mixed with 11 μL HiDi formamide and 1 μL size standard (ILS 500 of Promega Corp. or LIZ 600).

2.4. Casework statistics and analysis

GeneMapper® ID-X1.1.1 (Life Technologies) was used for data analysis. The Amelogenin system, which is included in the NGM SElect kit, was used to assess the male DNA proportion in female/male mixtures. The minimal requirement for an informative autosomal profile was set to five fully detectable STR systems. The average exclusion chance for this minimal genotype was between 1 in 10^8 and 1 in 10^10 based on the database created by the European Network of Forensic Science Institutes (ENFSI) and the product rule (http://strbase.org). Each Y-STR profile was assessed in terms of the number of sources (single source or male/male mixture) and the number of alleles called. A profile was considered to be highly informative when it was single-source and all Y-STRs were detected; at least one allelic peak needed to be detected at the multi-copy locus DYS385. The exclusion chance for full single-source profiles was between 10^8 and 10^9 based on the size of the Y Chromosome Haplotype Reference Database (YHRD) and the counting method (91,231 Yfiler and 25,499 Powerplex Y23 profiles in release 49 from February 2015; https://yhrd.org).

3. Results

An overview of the results of this study is given in Fig. 3. Each of the 2077 cases was analyzed twice in order to resolve the expected female/male DNA mixture with both autosomal and Y-chromosomal markers. Only 8 out of 287 cases (~3%) do not show a result with both profiling methods. In the autosomal analysis, ~24% of cases (68/287) do not exhibit any sign of male admixture in the Amelogenin system, but an additional male component was detected in 211/287 cases (74%). In 93/287 cases (32%), a total of 110 informative autosomal profiles were reported to the German Criminal DNA Database (DAD).

Parallel Y-STR analysis (with Yfiler for 276 stains and PowerPlex Y23 for 1801 stains) did not detect a male profile in ~13% (4/30 for Yfiler) and 13% (33/257 for PowerPlex Y23) of our cases. However, in ~84% of PowerPlex Y23–examined cases (216/257) and 87% of Yfiler-examined cases (26/30), at least one male profile was detectable. In contrast, only 211/287 (74%) cases exhibit any autosomal profile. This corresponds to an increased yield of ~11% for Y-chromosomal analysis versus autosomal analysis. In ~40% (102/257 PowerPlex Y23) and 47% (14/30 Yfiler) of the cases, at least one complete, single-source Y-STR profile was generated. In total, 133 full profiles were detected with the PowerPlex Y23 or Yfiler kits. Thus, the combined autosomal/Y-chromosomal STR approach increased the number of informative profiles by 21% compared to standard stand-alone autosomal analysis (Fig. 3).

Y-STR analysis was especially valuable for complex mixtures for which the contributor profiles could not be deconvoluted and for mixtures for which an expected male component was not detected via autosomal analysis. In 34/287 cases the composition of the autosomal mixture became clear upon Y-STR analysis, which provided complete male profile(s). Y-STR profiles were detected for 53% of Amelogenin Y negative cases (36 of 68 cases). In six of these cases complete single source Y profiles were generated. Closer inspection revealed that stains from these cases typically consisted of visible bodily-fluids on victim’s clothing (these fluids were supposed to be female because they tested negative for semen), including hidden touch DNA from a male contributor. Here, Y-STR analysis selectively identified a male component in a high female background to yield a single-source Y-STR profile, whereas autosomal analysis failed due to preferential amplification [17–19]. Comparable results were obtained for vaginal, anal, or oral swabs from the interior of the body of the victim. Our case collection included 58 attempted–rape cases that almost exclusively lacked sperm-positive stains; 911 evidentiary samples from these cases were collected in the hospital. Of these samples, 423 (46%) were swabs taken from the body of the victim (vaginal, anal, oral, and skin). Although an autosomal male component was identified in 12% of samples, only four stains (1%) resulted in profiles that were sufficiently high quality for database reporting. Stains from clothing were much more informative because the negative effect of unbalanced mixture proportions was less pronounced. Overall, we obtained informative autosomal profiles in 29/58 of attempted rape cases (50%).

![Fig. 2. Probative sexual-assault casework samples (n=2077) submitted to the Department of Forensic Genetics, Berlin.](image-url)
Difficulties in analyzing body swabs via autosomal markers were partially mitigated by the use of Y-STRs. Here, 64,423 body swabs collected at the hospital (~15%) (Fig. 2) yielded complete, single-source Y-STR profiles. Of these 64 swabs, ~11% did not generate an Amelogenin Y signal with NGM SELECT analysis. Autosomal STR analysis detected multiple male donors in 9% of these cases compared to 26% with Y-STR analysis.

In 12/287 offences (4%) two men contributed to the mixture due to the victim and the offender being male. In four cases of such male–male mixture both profiles were deconvoluted using Y-STRs but not with autosomal markers. However, in three other cases, Y-STR testing with 17 and 23 markers did not differentiate paternal relatives that contributed to the stain. In one of these cases, the two paternally related contributors were only discriminated by autosomal STRs; profiling revealed that one man matched the stain. In a second case, a suspect’s autosomal profile did not match the stain, but a Y-chromosomal match confirmed the same patrilineal descent. In a third case, autosomal STRs identified three male contributors in different stains and Y-STRs identified two of them as patrilineal relatives. Only the profile of the third man matched the trace profile.

To assess the information value of the 133 complete Y-STR profiles (116 PowerPlex Y23, 17 Yfiler profiles), all haplotypes were searched against the YHHRD database [20], release 49, February 2015. Of these haplotypes, 93% (124/133) did not match a reference sample. The augmented counting results for the unmatched haplotypes were 1 in 25,500 for the 23-locus haplotype and 1 in 91,232 for the 17-locus haplotype based on the size of the actual database. The 23-locus haplotype matched substantially less often (2 out of 116 PowerPlex Y23 profiles) than the 17-locus haplotype (7 out of 17 Yfiler profiles). Two Yfiler haplotypes matched 31 and 45 times.

4. Discussion

Sexual assault is a forensic casework category for which numerous biological traces are expected to contain sufficient numbers of cells for DNA analysis due to physical contact between individuals. Most forensic laboratories concentrate only on swabs from the female genital tract in order to obtain an autosomal profile of the sperm donor. Sophisticated methods such as cytological detection in combination with laser-microdissection of sperm cells [21] as well as differential lysis [22] in combination with Sperm Elution® [23] have been proposed for the separation of evidentiary male samples from female background. However, many sexual offense cases lack detectable spermatozoa for reasons including azoospermic or vasectomized perpetrators, no ejaculation, extended time intervals between the incident and sampling [24], and false allegation. However, male DNA can still be detected in cervicovaginal samples, skin, clothing, bedding, and other items that were touched due to the presence of epithelial cells and other cell types, e.g. leukocytes. Christian et al. reported that clothing and linens yield the majority of evidence and should be pursued vigorously for analysis [25]. Thus, the potential for collecting DNA evidence from the victim, the suspect, and the crime scene is nearly unlimited if methods to deconvolute the female–male mixture are implemented.

Previously, Y-STR testing effectively and selectively amplified the male fraction and clearly identified male profiles in evidentiary samples from the interior and exterior of the victim’s body [26]. However, to realize the full potential of a combined autosomal/Y chromosomal STR analysis for solving sexual-assault cases, it is essential to collect all appropriate DNA evidence from the victim’s body and clothing as well as from the crime scene, and to define a case-sensitive workflow that is high-throughput and generates the maximum number of informative DNA profiles. The lack of widespread implementation of Y-STR methods in routine casework may be influenced by budget concerns but could also be caused due to the lack of available success-rate data. In the current retrospective study (2012–2014) we evaluated a large number of real cases that included typical stains from sexual-assault cases. Particular attention was paid to touch stains, which were collected and prepared in our department. These stains are often the sole evidentiary material in cases of attempted sexual assault.

Here we implemented and evaluated a workflow for all types of stains submitted in several categories of sexual assault. We applied one autosomal kit and one Y-chromosomal kit and amplified up to 39 STR markers for each stain. Several studies support our decision to omit additional steps from our workflow, such as sperm identification and scoring, differential lysis, and quantification. For
example, a recent study showed that samples with low sperm counts (observed in the majority of swabs relating to sexual coercion and attempted rape) were misclassified as spermatozoa-free in 73% of tested samples [27]. Another study indicated that more than 90% of the male DNA initially present in simulated sexual-assault samples was lost after differential DNA extraction [28]. Current quantification methods based on q-PCR cannot reliably determine which samples will yield a DNA profile [2]. STR kits were previously shown to be more sensitive than real-time PCR approaches in terms of DNA quantitation [29].

In the current study, DNA profiles were generated in 279 of all 287 cases (97%) via autosomal or Y-chromosomal analysis or both. In ~8% of all cases (23/287), informative profiles were only detected with Y-STR kits. The advantages of our dual workflow are particularly strong for complex mixtures that are inconclusive on autosomal analysis as well as for mixtures that contain a clear excess of female DNA, which results in preferential amplification of the female component and suppression of the male component. Another advantage of Y-STR typing is the increased chance of detecting multiple male donors. Compared with autosomal analysis, Y-STR analysis was more successful at detecting and individualizing males in the mixture. Only in three cases did Y-STR analysis lack sufficient discriminatory power to distinguish paternally related males that contributed to the stain. Most of the generated Y-STR profiles were associated with a low frequency of occurrence and therefore displayed substantial power of exclusion. In cases with inconclusive autosomal results, Y-STR profiling can be used to exclude or match a suspect if reference samples are available for direct comparison.

Based on the current investigation, we propose the following evidence-based recommendations. First, touch stains are an important source of male DNA in sexual-assault cases and should always be collected and analyzed. Second, mandatory analysis of both autosomal and Y-chromosomal STRs ensures the retrieval of the maximum amount of information about different DNA sources in a stain. Third, although Y-STR profiles currently cannot be included in national DNA databases, reference samples should be submitted to forensic laboratories and typed for both marker categories.

In summary: combined autosomal and Y-chromosomal analysis provides additional leads for the investigation and prosecution of sexual assault. The two-kit analysis implemented here is currently the method of choice. However, future kit development should pursue a one-kit solution using next-generation sequencing or classical fragment-length analysis.

Disclosure statement

The authors declare no conflict of interest.

Acknowledgement

The authors thank Mechthild Prinz for critical comments on a previous version of this manuscript.

References