UPTEC X07 033
Examensarbete 20 p
Juni 2007

Validation of AmpFlSTR® Yfiler™ and evaluation of the use of Y-STR analysis in forensic casework in Europe

Carl-Olof Joachimsson

Validation of AmpFℓSTR® Yfiler™ and evaluation of the use of Y-STR analysis in forensic casework in Europe

Carl-Olof Joachimsson

Sammanfattning

Den största delen av grova våldsbrott, såsom mord och våldtäkter, begås av män. Genom att analysera Y-kromosomalt DNA från våldtäktsfall med manliga förövare och kvinnliga offer kan man erhålla manliga DNA-profiler även om andelen kvinnligt DNA är högt. Till hjälp finns kommersiella kit som specifikt binder till variabla delar på Y-kromosomen (Y-STRs). Ett sådant kit är AmpFℓSTR® Yfiler™, som i en reaktion känner igen 17 olika Y-STRs. De analyserade regionerna varierar mellan olika individer och resultatet kan, som ett slags molekylärt fingeravtryck, jämföras mot en misstänkt gärningsmans DNA-profil.

Yfiler™ validerades för användning på Statens Kriminaltekniska Laboratorium (SKL) och visades tillmötesgå uppställda krav med avseende på hur låga koncentrationer av manligt DNA som kan analyseras och hur specifik metoden är för manligt DNA. I en jämförande studie med det Y-STR kit som för närvarande används på SKL, PowerPlex®Y, visades Yfiler™ ha en något högre kapacitet än PowerPlex®Y. En enkätundersökning som skickades ut till forensiska laboratorier i Europa visade att tillförlitligheten för Y-STR generellt är låg, främst beroende på otillräckliga referensdatabaser och svårtydda resultat. Internationella riktlinjer och standardförfaranden skulle i framtiden kunna förbättra tillförlitligheten, men det är fortfarande oklart hur stor genomslagskraft Y-STR kommer få inom forensisk DNA-analys.

Examensarbete 20 p i Molekylär bioteknikprogrammet

Uppsala universitet Juni 2007

UPPSALA UNIVERSITET

Molecular Biotechnology Programme

Uppsala University School of Engineering

UPTEC X 07 033	**Date of issue 2007-06**

Author

Carl-Olof Joachimsson

Title (English)

Validation of AmpFℓSTR® Yfiler™ and evaluation of the use of Y-STR analysis in forensic casework in Europe

Title (Swedish)

Abstract

Y-chromosomal short tandem repeat (Y-STR) analysis is a new and expanding field in forensic science. AmpFℓSTR® Yfiler™ amplifies 17 polymorphic Y-STR loci in one single multiplex reaction. The kit was shown to be able to detect male DNA with a high success rate even at low amounts of male DNA (63 pg). The amplification was shown to be specific to male DNA, and male DNA was successfully amplified even in the presence of excess female DNA (ratio 1:1000). In a comparison study with PowerPlex®Y, another Y-STR kit on the market, Yfiler™ was shown to have a lower drop-out ratio and displayed a higher discriminatory power than PowerPlex®Y. Evaluation on the use of Y-STR analysis in Europe showed that the need for international guidelines and standards for database searches and haplotype frequency calculations is apparent.

Keywords

Y-STR, Yfiler™ , PowerPlex®Y SWGDAM core set, forensic science, male DNA typing

Supervisors

Christina Valgren — **Senior Molecular Biologist, SKL**

Marianne Mattsson — **Senior Forensic Expert, SKL**

Scientific reviewer

Ricky Ansell, Ph.D.
Forensic Adviser, SKL

Project name	Sponsors
Language **English**	Security
ISSN 1401-2138	Classification
Supplementary bibliographical information	Pages **50**

Biology Education Centre Biomedical Center Husargatan 3 Uppsala
Box 592 S-75124 Uppsala Tel +46 (0)18 4710000 Fax +46 (0)18 555217

Table of Contents

1. INTRODUCTION

DNA analysis has become an invaluable part of forensic casework during the last decades. The analysis methods are improving in sensitivity and accuracy, which increases the power of evidence of DNA profiling in court.

Males commit the vast majority of violent crimes, such as murder and sexual assault. Many of the DNA samples collected in rape cases contain mixtures of DNA from both the assailant and the victim, usually with a male perpetrator and a female victim. The female DNA part of the mixed samples is usually in excess, and the male DNA could go undetected in the routine autosomal DNA analysis. In these cases, there is a need to single out and profile the male portion of the DNA. The Y chromosome is the only genetic difference between men and women since men exclusively inherit the Y chromosome from the paternal lineage and an X chromosome from the maternal lineage. Women inherit two X chromosomes, one from each parent, and thus lack the Y chromosome. By targeting the Y chromosome, the male specific parts of male/female DNA mixtures can be analysed.

The study was conducted in order to evaluate and validate the AmpFℓSTR Yfiler™ Y-STR kit. With the use of this kit, 17 variable regions of the male-specific Y chromosome can be amplified and detected. The purpose of this is to acquire a DNA profile that either strengthens the evidence against a male suspect or clears him from the investigation. As a supplement to the validation study, a comparison study between Yfiler™ and the current kit used in Y-STR analysis at SKL (Swedish National Laboratory of Forensic Science), PowerPlex®Y, will be set up. The use of Y-STR analysis methods and the thoughts on Y-STR around Europe will be analysed.

The report is intended for readers with a level of education corresponding to or higher than that of the author.

2. BACKGROUND

2.1. History of DNA profiling

The first pioneering analysis method using DNA in forensic science occurred in 1984, where minisatellites (repetitive sequences with repeat units of more than 10 nucleotides) were separated using restriction enzyme digestion and Southern hybridisation blotting. The differences in sequence lengths appeared as multi-band patterns. The use of single locus probes (SLPs) followed where separate minisatellites were cloned and blotted (Jobling and Gill 2004).

With the introduction of PCR (polymerase chain reaction), the amounts of DNA needed for DNA profiling dramatically decreased. Combined analysis of several bi-allelic SNPs (single nucleotide polymorphisms) was introduced as a complement to the SLP method. However, a more powerful method in reference to discriminatory power is the use of STRs (short tandem repeats). STRs are short repeat sequences (by definition shorter than the minisatellites) frequently occurring in the genome. STR regions are usually highly multi-allelic and different alleles are easily separated due to the allelic size differences (Jobling and Gill 2004). Today, STR analysis is by far the most important method used in forensic DNA analysis, and has been in use in Sweden since 1993 (Ansell 2004).

2.2. From crime scene to DNA profile

When a crime scene investigator arrives at a crime scene, DNA sampling may present a challenge due to the fact that biological traces sometimes are hard to detect. The risks of contamination by the sampler, or the collection of irrelevant samples, are other factors that could slow down the criminal investigation (Butler 2005). Although the exact chain of handling may vary between cases depending on the nature of the crime and the type of sample acquired, the general chain between crime scene sampling and the acquisition of a DNA profile will be covered in this section.

Collecting the evidence

When at a crime scene, care must be taken to acquire appropriate samples and doing this in an appropriate manner. A crime scene sample could be anything that has been in contact with, or that originates from, the perpetrator or the victim, such as clothes, cigarette butts, or chewing gum. In sexual assault cases, underwear and genital samples from the victim and the suspect are obtained if possible.

The unwanted effect of contamination is primarily avoided by good sample handling. Good sample handling includes the use of protective gloves in every handling step of the sample and the isolation of secured samples to avoid contact and mixing with other samples from the crime scene. Gloves should be changed as often as considered necessary, preferably between the collections of each sample. Other sample-conserving strategies include packaging the sample in paper bags, air-drying wet samples to avoid degradation of the DNA, and quick transportation from the crime scene to the forensic laboratory (Butler 2005). When the sample reaches the forensic laboratory, a sample of the stain is cut out from the collected material and is prepared for DNA extraction.

Extracting the DNA

There are mainly three strategies for DNA extraction used at SKL, and the choice of method depends largely on the nature of the sample. These three methods are organic extraction, chelex extraction, and extraction using an FTA matrix[1].

Organic extraction

The organic extraction method involves the use of organic substances, such as phenol and chloroform, which separates proteins from the DNA in organic- and water-soluble phases.

Chelex extraction

The chelex extraction uses the chelating properties of the styrene divinylbenzene copolymer known as Chelex® 100 (Bio-Rad Laboratories). This polymer attracts polyvalent metal ions and thereby excludes them from the solution. The cell's nucleases use magnesium ions as a co-enzyme, and the lack of magnesium ions therefore inhibits the degradation of the DNA. In sexual assault case samples with mixtures of sperm and epithelial cells, the chelex method is complemented with a separation method called differential lysis which isolates the sperm cell fraction of the sample (Butler 2005, Jobling and Gill 2004).

FTA paper

The FTA method is exclusively used for reference samples. The method uses the DNA-binding qualities of a cellulose-based paper called FTA™ paper. A swab sample collected from a suspect or a victim is smeared over the paper. The paper binds the DNA and at the same time protects it from degradation. The FTA™ paper can then be used in amplification processes without further extraction by just transferring a small sample to a PCR reaction tube.

Analysis of the extracted samples

Analysis of the extracted DNA involves four highly automated steps: quantification, amplification of STR regions, fragment length separation and detection, and *in silico* interpretation.

Quantification

Real-time PCR is used to determine the amount of DNA in the sample. The quantification results are used to determine if dilution is needed to obtain optimal STR analysis DNA concentrations. Real-time PCR uses a thermal cycler, the quencher-fluorescent properties of the TaqMan® probe, the 5' nuclease activity of the *Taq* polymerase, an argon laser, and a fluorescent detection device to quantify the amplified fragments while still running the amplification (Wilson and Walker 2005). The TaqMan® probe consists of an oligo-nucleotide sequence, a fluorescent indicator at one end of the oligo-nucleotide, and a proximity quencher at the other end. As long as the fluorophore and the quencher stay in close proximity to each other, no fluorescence is emitted. The oligo-nucleotide probe is designed to bind complementary to the amplified sequences. When the polymerase reaches the bound probe in the extension phase of the thermal cycle, the 5' nuclease activity degrades the oligo-nucleotide and the fluorescent indicator is separated from the quencher. This results in the excitation of the fluorophore and the detection of the emitted electromagnetic radiation. As more cycles passes, more and more fluorophores will be released from their quenchers and the

[1] Personal communication: Christina Valgren, Senior Molecular Biologist. SKL, Linköping, Sweden (February 2007).

detected signal will increase. By using samples of known concentration as a standard comparison set, the detected signal can be compared to this set and the original sample concentration can be estimated.

Amplification of STR regions

When amplifying STR loci multiplex PCR techniques are used. In a multiplex PCR reaction, several primer pairs are used in one reaction tube (Butler 2005). Each fragment amplified in the PCR reaction usually requires a separate primer pair. Optimising a PCR reaction with multiple primer pairs is a laborious task. Primer sets need to have similar annealing temperatures and similar affinity to their respective DNA binding sites and yet they have to maintain sufficient specificity to avoid unspecific binding. The annealing temperature (T_m) of the primers is largely dependent on the GC-content in the primer-binding region. A high GC-content in the primer raises the annealing temperature compared to an equally sized primer with a lower GC-content. This is mainly due to increased stability in the primer-template DNA binding because of the greater number of hydrogen bonds between G≡C than between A=T. This leaves little freedom to the primer designer since the primers have to bind close to the STRs in order to amplify as small fragments as possible. Consequently, this renders multiplex PCR more vulnerable to degraded DNA and DNA samples of poor quality. Even so, the gains of using multiplex PCR overshadow the difficulties in the form of time gains and less sample volume needed. Low sample volume usage leads to more effective sample handling, where full profiles of samples with low DNA amounts can be obtained.

Fragment length separation and detection

Capillary electrophoresis is a separation method used to separate charged particles. This method uses small sample volumes and the separation efficiency is high with a fragment resolution of less than one nucleotide (Wilson and Walker 2005). An argon laser is set at the end of the capillary to excite fluorescent fragments passing. Fluorescent primers are used in the amplification step, and these primers together with their amplified fragments are detected by a detection camera as they pass the argon laser. Small fragments of DNA move faster through the gel than larger fragments, and fragments of increasing sizes will be detected. The use of multiple dyes (see TABLE 2.6.1) allows separation of fragments of similar sizes. Thus the separation is two-dimensional. Primarily, the direct capillary electrophoresis method separates the fragments by size, and secondarily, the fluorescent detector interprets the emitted wavelengths to indirectly separate fragments with similar sizes.

In silico analysis

The automated laboratory steps are followed by computational and manual analysis *in silico*. The detector measures the fluorescence in relative fluorescence units (RFU). The raw data from the detector is modified and interpreted, and the results are modified by an algorithmic comparison, which uses the internal lane standard (ILS), allelic ladders, and fluorescent overlap matrices to separate the signals and defining the fluorescent peaks as alleles. The resulting peaks are presented in an electropherogram (*Figure 2.2.1*), which is analysed and interpreted manually. The peaks correspond to specific STR alleles and a DNA profile is created based on these peaks. The profiles are handled differently depending on the nature and type of sample. Some profiles are entered in national databases, while others are compared to reference samples of suspects[2].

[2] Personal communication: Marianne Mattsson, Senior Forensic Expert. SKL, Linköping, Sweden (March 2007).

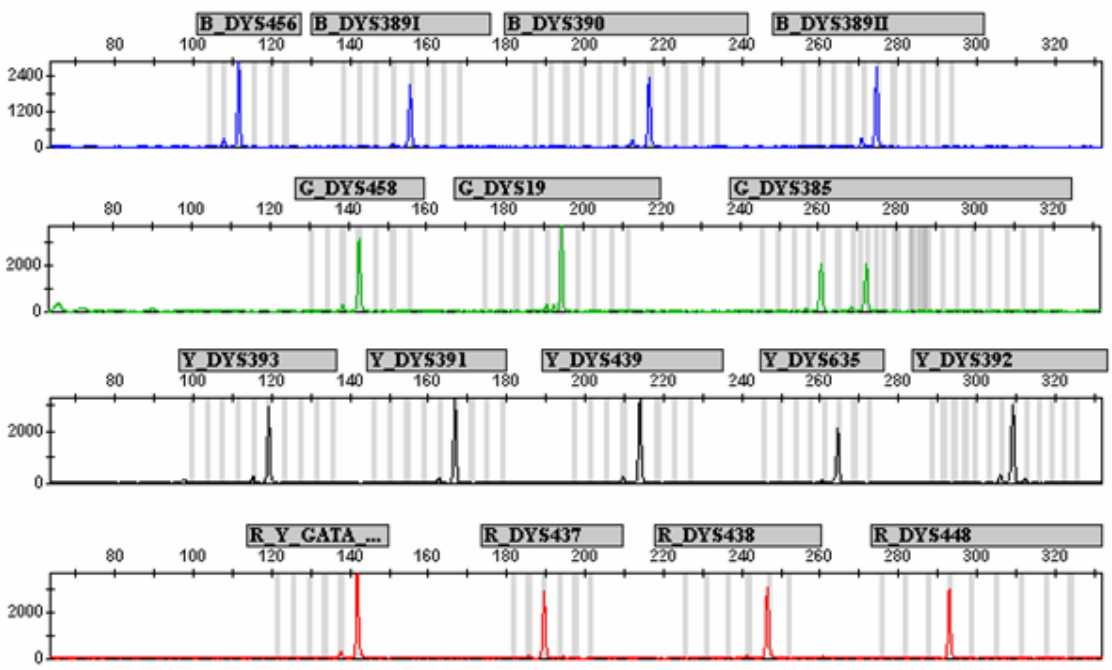


Figure 2.2.1. A typical electropherogram of a Yfiler™ Y-STR DNA profile. The four lanes separate the markers according to their labelled dyes (from the top: Blue, Green, Yellow, and Red).

2.3. STR loci

The occurrence of STR regions and their mechanisms of molecular evolution provide a versatile tool to differentiate between different sets of genomic DNA. The technical advantages in genotyping using STR loci will be discussed, as well as some of the problems frequently encountered during the analysis. To conclude this section, Y-STR nomenclature and repeat sequence characterisation will be covered.

Structure, occurrence and evolutionary history

STRs (short tandem repeats) are repetitive DNA sequences spread throughout the human genome. The initial human genome sequencing project suggests that STRs, or microsatellites, make up for as much as 1.5% of the human genome (International Human Genome Sequencing Consortium 2001). STRs consist of short nucleotide sequences (1-6 nucleotides) linked together to form repetitive regions of variable lengths. The distinguishing factor between different alleles of one STR locus is the number of repeats, i.e. the length of the STR repeat. The most common motives of the repetitive sequences are di-nucleotide repeat sequences, followed by tetra-nucleotides (Ellegren 2004). As will be discussed later on in this section, the use of longer repeat units, such as tetra- and penta-nucleotides, are preferred in forensic science due to greater replication stability when amplifying the fragments with the Taq-polymerase.

Most STR regions are located in introns and other non-transcribing areas of the genome and are therefore believed to be under neutral selection (Ellegren 2004). The mutation rate of STR regions is high compared to the overall mutation rate of the human genome. This has also been shown in Y-STR regions, where *Kayser et al*. (Kayser 2000) estimates the average Y-STR mutation rate to be 0.28%. The exact rate is difficult to estimate, but a study by *Heyer et al*. (Heyer 1997) presents similar results. The high mutation rate and the influence of neutral selection suggest a high multi-allelic diversity of the STR loci.

The lengths of the regions are determined by complex and not yet fully determined evolutionary dynamic processes, which expand or shorten the regions. One theory on the evolution of STR regions (Ellegren 2004) suggests that shorter regions tend to increase in length due to replication slippage, while longer regions, which generally will have existed for a longer period of time, are more prone to acquire point mutations. This breaks the repetitiveness and splits the STR region into two smaller regions. These mutational processes could reach equilibrium, where the STR regions are kept from expanding into vast areas of repetitive arrays in the genome.

Typical problems involved in STR analysis

Stutter peaks

The main problem with the polymerase is the occurrence of *stutter peaks*. A stutter is a peak in the electropherogram positioned close to the actual allele peak, usually one whole repeat sequence away (see *Figure 2.3.1*). The occurrence of stutter peaks impedes the interpretation of DNA profiles, since allele peaks of a minor DNA contributor could be masked by the stutter peak of the allele peak of the major contributor.

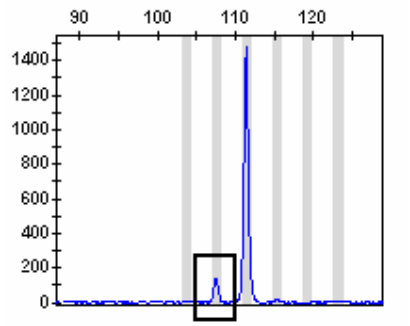


Figure 2.3.1. An example of a stutter peak. The vertical grey bars represent allelic positions of the locus. The stutter is positioned 1 repeat sequence away from the allele peak.

The stutter peak is the result of a replication slippage made by the polymerase early on in the amplification process (Shinde 2003). The mechanism creating the replication slippage is thought to be the partial dissociation and re-association of the newly synthesised DNA from the template strand. If the two strands partially dissociates from a replication site within the STR region, the strands could re-associate one repeat further away from the replication site, thus creating a non-replicated loop on the template strand. This would then create a strand one repeat shorter than the original one (Ellegren 2004). This motif will in this report be referred to as *stutter –1*.

The opposite event is also known to occur, but much more seldom (Shinde 2003). The newly synthesised strand dissociates and re-associates to the template strand one repeat upstream of the original replication site (Ellegren 2004). Thus that particular repeat sequence will be replicated twice, creating a strand one repeat longer than the original template strand. This motif will be referred to as *stutter +1*. The energetic argument explaining the infrequent occurrence of stutter +1 peaks is that at least one extra nucleotide has to dissociate from the template strand in order to form a loop in the newly synthesised strand (Shinde 2003). This is not energetically favourable compared to the slippage creating a stutter –1.

Stutter peaks are more common if the repeat sequence is short (Shinde 2003); e.g. the stutter peaks from a tri-nucleotide repeat sequence are larger relative to the allele peak in the electropherogram compared to the stutter peak from a tetra- or penta-nucleotide repeat sequence. The total number of repeats is also known to be an important factor determining the relative stutter height. A longer repeat sequence is more likely to acquire replication slippage during the amplification than a shorter sequence mainly due to the increased possibility for the polymerase to slip (Shinde 2003).

Degraded DNA and amplification inhibitors

Another problem frequently encountered in the forensic analysis of STRs is the occurrence of degraded DNA. DNA degradation takes place in dead cells, where endonucleases cut the DNA into smaller pieces, making it hard to amplify long sequences of DNA. A typical electropherogram pattern of a degraded sample shows that the shorter fragments are amplified with a higher success rate than the longer fragments (Butler 2005).

There are a number of known substances which either degrades or binds to the DNA, or inhibits the DNA polymerase. The result of inhibition ranges from low intra-colour peak balance to allelic drop-outs and total PCR failure.
Known inhibitors described in the literature common in forensic case samples include calcium ions, cellulose, haematin, denim dyes, soil components and high levels of non-target DNA (Butler 2005, Wilson 1997).

Mutations

The *Taq* polymerase lacks the 3'-5' exonuclease proof-reading capacity of the eukaryotic DNA polymerases (Diaz and Sabino 1998). This could affect the amplified PCR products in a number of ways. A point mutation could occur early on in the amplification process, resulting in most of the fragments having a slightly different sequence than the original sequence. This, however, is not a problem in STR analysis since the exact sequence is irrelevant because the fragments are separated by size and not by sequences.

Y-STR nomenclature

STR regions are present throughout the human genomic DNA. The evolutionary mechanisms determining the mutation rate and STR formation are the same regardless of the location of the repeat. However, in this report, the main focus will be set to STR regions located on the Y chromosome (Y-STR), and only nomenclature and repeat sequence characterisation of Y-STRs will be covered.

Y-STR sequences ranges in structure and length, making nomenclature guidelines necessary for inter-laboratory comparison. The International Society of Forensic Genetics (ISFG) has set up guidelines concerning the nomenclature of Y-STRs (Gusmão 2006). These guidelines include nomenclature of Y-STR loci as well as the definition of a Y-STR locus and the different alleles it comprises. In this section, the guidelines regarding typical Y-STR alleles will be covered. For further information regarding nomenclature of Y-STRs, see the ISFGs recommendations (Gusmão 2006).

The number of repeats of a Y-STR allele is referred to as a numerical value, e.g. 11. Intermediate-sized alleles could arise from indels (insertions or deletions) in the repeat sequence. In these cases, the alleles are referred to as integers depicting the number of whole repeat sequences followed by a suffix describing the number of remaining bases. A hypothetical tetra-nucleotide STR allele with 11 whole repeats and a 3 nucleotide insertion

would be defined as 11.3. Note that the evolutionary origin of the incomplete repeat is irrelevant to the nomenclature. A deletion of one nucleotide in the repeat sequence region of a 12 allele would be denoted in the same way as a 3 nucleotide insertion into the repeat sequence of an 11 allele, i.e. 11.3.

Alleles not included in the allelic range of the STR locus are defined as less than or larger than the boundary alleles (see TABLE 2.6.1 for boundary alleles of the allelic ladder supplied by Applied Biosystems) of the locus (Gusmão 2006). An allele with 6 repeats or less of the locus DYS391, with a defined allelic range of 7-13 repeat units, would be denoted as a <7 allele.

In order to simplify the statistical calculations, off-ladder alleles will be denoted as the corresponding allele in this report, i.e. a 6 repeat in the DYS391 locus will be called as a 6 allele and not as a <7 allele. For similar reasons, intermediate-sized alleles will be denoted as the closest known whole-repeat allele, i.e. an 11.2 allele will be called as an 11 allele. These denotations are used throughout this study.

2.4. The Y chromosome

In this section the evolutionary history of the sex chromosomes will be summarised. The details on the molecular evolutionary processes leading up to the differentiation of the sex chromosomes are out of the scope for this analysis. It is only meant to serve as a brief introduction to the structure and repetitive nature of the modern Y chromosome described later on in this section.

Evolutionary history of the sex chromosomes

Müller proposed the ruling theory on the evolution of sex chromosomes as early as 1914 (Graves 2006). He suggested that the sex chromosomes originated from an ordinary pair of autosomal chromosomes. This evolutionary process is believed to have started some 300 million years ago (Bachtrog 2006, Graves 2006, and Ross *et al*. 2006). The introduction of a sex-specific gene on one of the chromosomes together with a sexually antagonistic allele induced the creation of a non-recombining area between the chromosomes (Bachtrog 2006, Graves 2006). The creation of this non-recombining region is believed to be one of the most important factors for further clustering of male specific genes on the Y chromosome (Bachtrog 2006).

The lack of recombination between the sex chromosomes induced other types of evolutionary events, including the selective sweep. Selective sweeps occur when a favourable mutation, derived either from a point mutation that improves gene function or an insertion of a transposable element, arises on the male chromosome (Bachtrog 2006). The favoured mutation increases the overall fitness of the bearer who will thus have a selective advantage and the mutation will tend to go to fixation in the population. Due to the lack of recombination, this will also lead to the fixation of the whole non-recombining area of the chromosome. The drawback of selective sweeps is the reduced variability and the possibility of fixation of neighbouring deleterious mutations, known as hitchhiking sequences (Bachtrog 2006). The introduction of deleterious mutations is another factor degrading the Y chromosome. In autosomes, these deleterious mutations are excluded from the population due to recombination of the chromosomes and a negative selection. These mutations will tend to accumulate in the parts of the genome that lack recombination (Bachtrog 2006). Thus, the loss of recombination allows genetic sex determination, but may eventually degrade the non-recombining sex chromosome until it ultimately withers away (Graves 2006).

Y chromosome structure

The 60 Mb (megabases) long Y chromosome can be divided into two sections defined by the amount of interaction with the X chromosome during meiosis: the PARs (pseudoautosomal regions) and the MSY (male specific regions of the Y chromosome). The PARs are located at both ends of the Y chromosome and make up approximately 3 Mb of the total Y chromosome (Rappold 1993). These parts interact and recombine with the equivalent parts on the X chromosome during meiosis. The MSY does not recombine and can in turn be divided into two parts: one euchromatic part and one highly repetitive heterochromatic part. The euchromatic content of the MSY is made up of about 23 Mb and contains 156 transcription units coding for 27 unique proteins (Skaletsky 2003). There are several long repetitive sections in the euchromatic regions, including palindromic sequences referred to as amplicons containing the testis genes (Skaletsky 2003). The highly repetitive nature of the heterochromatic regions and the amplicons in the euchromatic region has made it difficult to create detailed maps of the Y chromosome (Ross *et al*. 2006, Skaletsky 2003). However, the Yfiler™ STR loci are all defined and located within the euchromatic parts of the chromosome.

2.5. Interpreting Y-STR profiles

Interpreting Y-STR profiles is an important part of the forensic process. Even if a profile is available, the power of evidence has to be assessed in order to use the results in court. The problems of estimating haplotype frequencies and the tools that helps solve the problems are introduced in this section.

Y-STR databases

SWGDAM core set

The SWGDAM core set consists of 11 Y-STR markers. These markers are chosen by the Scientific Working Group of DNA Analysis Methods (SWGDAM) because of their high gene diversity, and they are included in most Y-STR databases (Gusmão 2006). The Y-STR markers included in the SWGDAM core set are DYS19, DYS385a/b, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS438 and DYS439.

YHRD haplotype database

The Y chromosome haplotype reference database (YHRD) allows searches in different meta-populations as well as conducting haplotype frequency calculations using the information of the entire database. The STR markers used are those defined in the SWGDAM core set (see above). The information in the database is substantial with 51,253 recorded haplotypes (Institute of Legal Medicine, Charité - Universitätsmedizin Berlin 2007). This makes it the largest Y-STR database currently available. This database uses the counting method and neighbouring haplotypes to estimate haplotype frequencies. The large database and the statistical methods allow accurate estimates of rare haplotypes (Roewer 2000).

Yfiler database

The Yfiler Y-STR haplotype database contains haplotype frequencies of all the Yfiler Y-STR markers (see TABLE 2.6.1). More markers allow better separation and differentiation between different haplotypes. However, as of today, the database contains substantially fewer entries than the YHRD database and the population diversity is lower with mainly three subgroups of the U.S. population being covered: Caucasians, Hispanics and African Americans (Applied Biosystems 2007). This database use the counting method combined with statistical Bayesian models to estimate haplotype frequencies. The overall allelic gene diversity in the database

varies from 0.381 to 0.950 depending on the subgroup and locus (Mulero 2006). The gene diversity, D, is calculated using the allelic frequencies of the different loci according to EQUATION 2.6.1 below.

EQUATION 2.6.1. $$D = \left(\frac{n}{n-1}\right)(1 - \sum p_i^2)$$

n: sample size
D: gene diversity of the individual loci
p_i: allele frequencies of the individual alleles of each locus.

The gene diversity calculations are useful in estimating haplotype frequencies through *a priori* methods, since the diversity allows a comparison calculation between closely related haplotypes.

The Yfiler database is expected to grow through potential cooperation and outsourcing. More subpopulations will be included in the future, which will allow detailed subpopulation database searches hopefully increasing the power of evidence of the estimated haplotype frequencies. As of now, the database is fairly good for Caucasians and the use of 17 markers significantly improves differentiation compared to the European minimal haplotype and the SWGDAM extended haplotype alone[3].

Power of evidence

Determining the power of evidence of a Y-STR profile is not a trivial task. The reference population databases available present a selection of a specific population. But how should the information in these databases be used? The easiest and most hands-on approach is to use the counting method to determine the haplotype frequencies (Krawczak 2001, Roewer 2000). This method simply counts the absolute haplotype frequencies in the database without compensating for stochastic differences. This method could be complemented with a statistical calculation using Bayesian *a priori* distribution estimates (Krawczak 2001, Roewer 2000). The STR loci on the Y chromosome are not to be considered as statistically independent due to the lack of recombination, and thus a cumulative approach multiplying allele frequencies will be erroneous (Roewer 2000).

When estimating haplotype frequencies, the occurrence of the exact same haplotype is calculated. This frequency is only applicable to men unrelated to the suspect. However, when interpreting Y-STR profiles, one has to consider the fact that all males in the paternal lineage share exactly the same Y-STR profile barring mutationally induced changes. So even if the haplotype frequency is low enough to present strong arguments in court, special concerns have to be taken to assure that a male from the suspect's paternal lineage did not deposit the sample.

Y-STR analysis today

The field of application for Y-STR analysis ranges from crime cases with male perpetrators and mixed male/female DNA samples, to paternal lineage investigations tracing genealogical history. Some of the most applicable areas are listed below:

[3] Personal communication: Nicola J Oldroyd. Senior Forensic Specialist. Applied Biosystems (March 2007).

Mixed profiles

Y-STR analysis could be used as a complement to autosomal STR analysis to strengthen the power of evidence, especially when male/female DNA mixtures are encountered (Butler 2005). When analysing DNA samples from a rape case, the sperm cells are usually vastly outnumbered by the female epithelial cells. Differential lysis could be used to fractionate these types of cells, but sometimes this is not enough to get a clean autosomal male profile. In these cases, Y-STR analysis is used to isolate the male fraction of the DNA.

Identification

Y-STR analysis could be used to identify a member of a paternal lineage with the use of any other member of the same paternal lineage as reference. This is useful in paternity testing as well as murder victim identification processes where a close relative could be hard to find. In recent years, the analysis method has proven to be a useful tool in the identification processes of mass disasters (Butler 2005).

Genealogical studies

Y-STR analysis is a powerful tool in tracing the kinship of distant relatives and ancestral roots. The use of Y-STR analysis in the study of European demographic history has proven unparalleled in comparison to previous methods (Roewer *et al.* 2005). Obviously, only the paternal inheritance is traceable using this analysis method. The corresponding method of tracing maternal inheritance is by mitochondrial DNA-analysis, due to the maternal inheritance of the mitochondrion.

2.6. Introduction to Yfiler™ and PowerPlex®Y

Yfiler™

The Yfiler™ kit is supplied by Applied Biosystems. It is a 17-plex Y-STR system amplifying the SWGDAM core set and 6 additional Y-STR markers. The STR markers and their respective repeat sequences, the allelic range of the allelic ladder, and the dyes with which the primers are labelled are listed in TABLE 2.6.1. The repeat sequences are published in the ISFGs recommendations on Y-STRs (Gusmão 2006).

Most of the repeat sequences are tetra-nucleotides. Note that the repeat sequences of the markers differ substantially between different loci. In some cases, e.g. DYS390, the repeat sequences are broken up into slightly different repeat sequences, while in other loci, e.g. DYS389II, the repeat is broken up by an intermediate oligo-nucleotide spacer.

The Yfiler™ kit uses a five-dye chemistry to label the forward primers of the STR regions. These dyes are referred to as 6-FAM™ (blue), VIC® (green), NED™ (yellow), PET® (red) and LIZ® (orange) (Mulero 2006). The first four dyes serve as STR primer labels. The fifth dye, LIZ®, is used to label the internal lane standard (ILS), which is run with every sample. To evenly space the markers, some of the primers have been equipped with a linker sequence between the fluorophore and the primer sequence. This lengthens the amplified fragments and allows tailored separation of the STR fragments.

TABLE 2.6.1. The 17 Y-STR markers of Yfiler™. The allelic ranges are those present in the Yfiler Kit Allelic Ladder (Applied Biosystems 2006).

Y-STR marker	Repeat length	Repeat sequence	Allelic range	Labelled dye
DYS19	4	$(TAGA)_3tagg(TAGA)_n$	10-19	VIC®
DYS385a/b	4	$(aagg)_{6-7}(GAAA)_n$	7-25	VIC®
DYS389I	4	$(TCTG)_3(TCTA)_n$	10-15	6-FAM™
DYS389II	4	$(TCTG)_n(TCTA)_nN_{28}(TCTG)_3(TCTA)_n$	24-34	6-FAM™
DYS390	4	$(tcta)_2(TCTG)_n(TCTA)_n(TCTG)_n(TCTA)_ntca(tcta)_2$	18-21	6-FAM™
DYS391	4	$(tctg)_3(TCTA)_n$	7-13	NED™
DYS392	3	$(TAT)_n$	7-18	NED™
DYS393	4	$(AGAT)_n$	8-16	NED™
DYS438	5	$(TTTTC)_1(TTTTA)_{0-1}(TTTTC)_n$	8-13	PET®
DYS439	4	$(GATA)_n$	8-15	NED™
DYS437	4	$(TCTA)_n(TCTG)_{1-3}(TCTA)_4$	13-17	PET®
DYS448	6	$(AGAGAT)_nN_{42}(AGAGAT)_n$	17-24	PET®
DYS456	4	$(AGAT)_n$	13-18	6-FAM™
DYS458	4	$(GAAA)_n$	14-20	VIC®
DYS635	4	$(TCTA)_4(TGTA)_2(TCTA)_2(TGTA)_2(TCTA)_2(TGTA)_{0,2}(TCTA)_n$	20-26	NED™
Y GATA H4	4	$(AGAT)_4CTAT(AGAT)_2(AGGT)_3(AGAT)_nN_{24}(ATAG)_4(ATAC)_1(ATAG)_2$	8-13	PET®

PowerPlex®Y

PowerPlex®Y is supplied by the Promega Corporation. It uses a single multiplex PCR reaction to amplify 12 markers. These include the markers recommended by SWGDAM and an additional marker, DYS437 (Krenke 2005). It uses a 4-dye system with one dye marking the internal lane standard (ILS). Concordance studies (Gross *et al*. 2006) display 100% concordance between PowerPlex®Y and Yfiler™. This allows comparison studies on the efficiency and detection capacity of the two systems.

2.7. Aim of the study

Validation of Yfiler™

The overall aim is to validate the Yfiler™ kit and to decide whether or not it meets the demands to be used in forensic routine analysis. This kit would then, if proven significantly better, replace the Y-STR analysis kit currently used, namely PowerPlex®Y. The validation of the Yfiler™ kit will include analysis of the sensitivity, specificity and robustness in detecting small amounts of male DNA under different conditions. These tests are recommended by the Scientific Working Group on DNA Analysis Methods (SWGDAM 2004). The sensitivity will be tested through dilution series analysis of male DNA. The specificity and robustness will be analysed by determining the ability to detect small amounts of male DNA along with excess female DNA. These tests are generally referred to as mixture studies, but to avoid confusion, the terms specificity and robustness will be used. Mixed samples will consistently refer to male/male DNA mixtures throughout this study unless anything else is specified. The main focus will be to analyse authentic case samples to observe stutter height and occurrence, amplification artefacts and intra-colour peak balance. The sensitivity and specificity will be analysed in a comparison study between two commercial Y-STR kits: PowerPlex®Y and Yfiler™.

The use of Y-STR in Europe

A questionnaire regarding the use of Y-STR analysis in forensic science will be sent out to other forensic laboratories in Europe. This will be done in order to get an overview on the possibilities and drawbacks of Y-STR analysis in general.

3. MATERIALS AND METHODS

3.1. The ABI chain of amplification and detection

Applied Biosystems (ABI) supplies the whole chain of amplification and detection of the Y-STR analysis used at SKL. The settings and specifications of the components are listed in this section.

AmpFℓSTR Yfiler™ kit

The contents of the Yfiler™ kit are listed in TABLE 3.1.1 below. The volumes in the table correspond to the contents of one kit of approximately 100 reactions.

TABLE 3.1.1. The contents of the Yfiler™ kit. Total volume refers to one kit of approximately 100 reactions. For the concentrations of the contents of the Primer Set and the PCR Reaction Mix, see text.

	Component	Volume (µl)	Concentration
Mastermix	PCR Reaction Mix	1100	-
	Primer Set	550	-
	AmpliTaq Gold®	100 (2 tubes)	5 U/µl
	Allelic Ladder	50	-
Control DNA	Male DNA 007	300	0.1 ng/µl
	Female DNA 9948	50	10 ng/µl

The concentration of the primers in the Primer Set varies between 0.12-1.5 µM. The components and their concentrations in the PCR Reaction Mix are listed below (Mulero 2006):

$MgCl_2$: 1.60 mM
KCl: 50 mM
BSA: 160 µg/ml
Sodium azide: 0.020%
dNTP mixture (equimolar): 800 µM

The concentrations above are the final diluted PCR setting concentrations. The total volumes needed of each of the components for one PCR reaction are listed below (Applied Biosystems 2006):

PCR Reaction Mix: 9.2 µl
Primer Set: 5 µl
AmpliTaq Gold®: 0.8 µl
Sample: 10 µl (recommended amount of DNA: 0.5 – 1.0 ng)

Total PCR reaction volume: 25 µl

GeneAmp® PCR System 9700

The PCR amplification program was set up on a GeneAmp® PCR System 9700 thermal cycler according to the recommendations of Applied Biosystems (Applied Biosystems 2006). The cycling parameters of the program most often used will hereafter be referred to as *Yfiler30* and this amplification program is described in TABLE 3.1.2. The total volume of each sample was 25 µl and the ramp speed on the thermal cycler was set to 9600.

TABLE 3.1.2. PCR cycling parameters. These cycling parameters were used throughout this analysis, if nothing else is specified. This program will be referred to as *Yfiler30*.

	1 cycle	---------- 30 cycles ----------			1 cycle	
Phase	Initiation	Denaturation	Annealing	Extension	Final extension	Storage
Temperature (°C)	95	94	61	72	60	4
Time (min)	11:00	01:00	01:00	01:00	80:00:00	∞

Prism® 3130xl Genetic Analyser

Fragment separation and fluorescent detection was done on the Prism® 3130xl using the settings recommended by Applied Biosystems (Applied Biosystems 2006).
1 µl of every sample was added to a mixture of 8.7 µl Hi-Di™ formamide and 0.3 µl GeneScan-500 LIZ Size Standard (in this study referred to as ILS). The prepared samples were heated in a heat block at 95 °C for 3 minutes and then cooled on ice for 3 minutes. The samples were injected for 10 seconds at 3 kV and were run at 15 kV for 1500 seconds.

These separation settings were used throughout this report.

GeneMapper® v3.2

The *in silico* analysis of the samples was made using GeneMapper® v3.2 software. The baseline cut-off value was set to 10 RFU. All peaks below this value were automatically interpreted as baseline fluctuations. Peaks and stutter peaks below 20 RFU resulted in inconsistent results due to stochastic effects of baseline fluctuations and were consequently discarded manually. Allele calls with values of 20 RFU or higher were either approved or discarded manually.
The internal lane standard (ILS) cut-off value was set to 50 RFU. The peak calls of the ILS were handled and interpreted automatically by the software.

These analysis parameters were used throughout the report.

3.2. Sensitivity, specificity and robustness

Male DNA sensitivity analysis and PCR optimisation

A male DNA dilution series was created using the male control DNA (male DNA 007). The following amounts of male DNA were amplified at 29, 30 and 31 PCR cycles respectively: 1000 pg, 500 pg, 250 pg, 125 pg and 63 pg. The amount of cycles was the only change made in the PCR programming settings of *Yfiler30*. The results were interpreted to find the optimal PCR cycle parameter. The parameter of choice was 30 cycles (also recommended by Applied Biosystems) and a duplicate using the same amounts of male DNA was done using *Yfiler30*.

An additional sensitivity assay with a triplicate of runs on *Yfiler30* was conducted using the same dilution series.

Specificity analysis using female DNA

A female dilution series was created using a female DNA reference sample with a DNA concentration of approximately 40 ng/µl. The following amounts of female DNA were amplified using *Yfiler30*: 400 ng, 130 ng, 67 ng, 33 ng, 17 ng, 8.3 ng and 4.2 ng. The amounts were chosen to cover the range of female DNA normally encountered in routine analysis.

Robustness analysis using male/female DNA mixtures

An analysis of the robustness of the method was done, where small amounts of male DNA were amplified in the presence of excess female DNA. Male DNA of three different amounts (0.25 ng, 0.125 ng and 0.0625 ng) was amplified in the presence of female DNA of ratios 1:50, 1:200, 1:400 and 1:1000 as schematically shown in TABLE 3.2.1.
The ratios were chosen to cover the possible ranges encountered in routine analysis.

TABLE 3.2.1. A schematic view of the male/female DNA robustness analysis. Amounts indicated in bold represent the amount of female DNA. The analysed ratios were 1:50, 1:200, 1:400 and 1:1000 as indicated in the header of the table.

	Male:**Female** DNA (ng:ng)			
	1:50	1:200	1:400	1:1000
0.25 ng male DNA	0.25:**12.5**	0.25:**50**	0.25:**100**	0.25:**250**
0.125 ng male DNA	0.125:**6.3**	0.125:**25**	0.125:**50**	0.125:**125**
0.063 ng male DNA	0.063:**3.1**	0.063:**12.5**	0.063:**25**	0.063:**63**

3.3. Authentic case samples

139 crime case samples were selected for analysis. The criteria when selecting the samples was that a clean and full male DNA profile or an obvious one male/one female DNA mixture was detected in routine autosomal STR analysis. Samples extracted with different extraction methods were chosen with the following ratio:

- 16 samples extracted using the organic extraction method.
- 64 samples extracted using the chelex extraction method.
- 19 samples extracted using the chelex extraction method and further purified using Centricon®-100.
- 40 samples extracted using the chelex-based differential lysis method to separate sperm and epithelial fractions. Both sperm and epithelial fractions were chosen if a male profile was detected in the autosomal STR analysis.

All samples had previously been quantified using Quantifiler™ Human kit and real-time PCR and the samples were diluted to a male DNA concentration of 0.05–0.10 ng/µl (recommended by Applied Biosystems).

23 additional crime case samples were run on routine analysis using PowerPlex®Y and analysed using Yfiler™ the same day in a comparison study.

3.4. Questionnaire

A multiple-choice questionnaire regarding the use of Y-STRs in forensic casework was created and was sent out to different forensic laboratories in Europe. The questions and the answers are shown in section 4.4.

3.5. Statistical calculations

The statistical presentation of the results will sometimes be illustrated by expectation values (arithmetic mean $\bar{x}$) of normally distributed stochastic variables with error bars representing a confidence interval (C) of significance*** ($\alpha = 0.001$). The confidence interval calculations are shown in EQUATION 3.5.1 below. These calculations are made under the assumption that the stochastic variables follow the Gaussian distribution, and thus $\alpha = 0.001$ corresponds to $\lambda_\alpha = 3.2902$ (tabulated value).

EQUATION 3.5.1.

$$C = \pm 3.2902\left(\frac{\sigma}{\sqrt{n}}\right)$$

The confidence interval calculations include the standard deviation (σ), and the sample size (n). The standard deviation calculations are shown in EQUATION 3.5.2 below.

EQUATION 3.5.2.

$$\sigma = \sqrt{\sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1}}$$

where $\bar{x}$ represents the expectation value of the stochastic variable (the arithmetic mean of the values x_i).

The level of significance and the statistical calculations above are used throughout this report when the results are presented as average values with error bars, unless anything else is specified.

4. RESULTS

4.1. Sensitivity, specificity and robustness

PCR optimisation and male DNA sensitivity analysis

PCR optimisation

The results from the PCR optimisation are shown in TABLE 4.1.1 below. Average peak heights and average peak balance are compared between the runs at 29, 30, and 31 PCR cycles. The peak balance has been calculated by dividing the lowest intra-colour peak height to the highest. The resulting four colour balance ratios have been averaged and are presented in the table below. Average values of the duplicate runs at 30 cycles are shown.

TABLE 4.1.1. Average peak heights and peak balance of the PCR optimisation assay.

	29 cycles	**30 cycles**	31 cycles
Average peak height (RFU)	305	**885**	1359
Lowest peak height at 1000 pg (RFU)	228	**1495**	2729
Peak balance (%)	41	**38**	33

The difference in peak balance between the runs is not statistically significant (confidence intervals of 95%; data not shown).

Sensitivity analysis

The results of the triplicate runs at 30 cycles are shown as average peak heights in *Figure 4.1.1* below. Note the similar patterns in peak height ratios at all the amplified DNA concentrations.

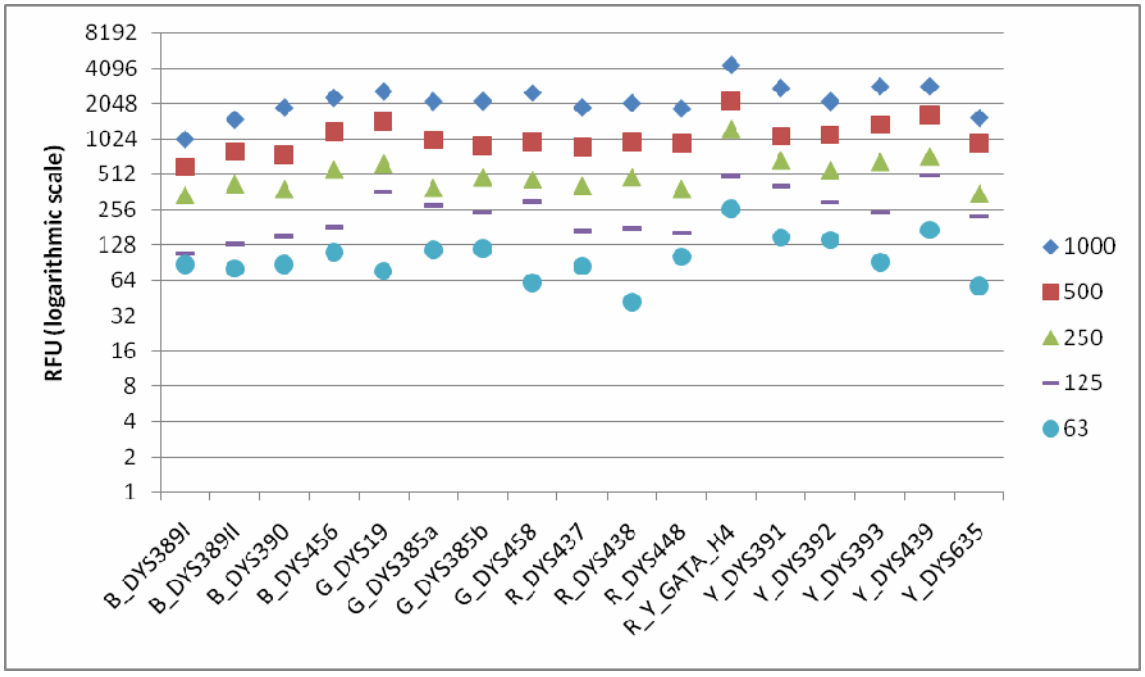


Figure 4.1.1. Average peak heights of the Y-STR markers. The vertical axis displays RFU-values on a logarithmic scale. The data series represent average peak heights of the different DNA amounts (1000 pg, 500 pg, 250 pg, 125 pg, and 63 pg). The indexes of the markers correspond to their labelled dyes (B: blue; G: green; R: red; Y: yellow).

Specificity analysis using female DNA

The female dilution series showed little more than a few off-ladder baseline fluctuations. The electropherogram of the sample with the highest female DNA concentration (400 ng/25 µl) is shown in *Figure 4.1.2* below. The possible artefacts are boxed and marked 1-7.
As seen in the figure, the baseline fluctuates of up to and occasionally above 10 RFU. These fluctuations decrease slightly when the concentration of female DNA is lowered (data not shown). Also, the artefact in the DYS456 allelic area (marked 1) was only distinguishable at the highest of the female DNA concentrations.
A positive control sample with male DNA and a negative control sample without any presence of DNA were run together with these samples. The results of these runs were consistent with previous results using the same control samples (data not shown).

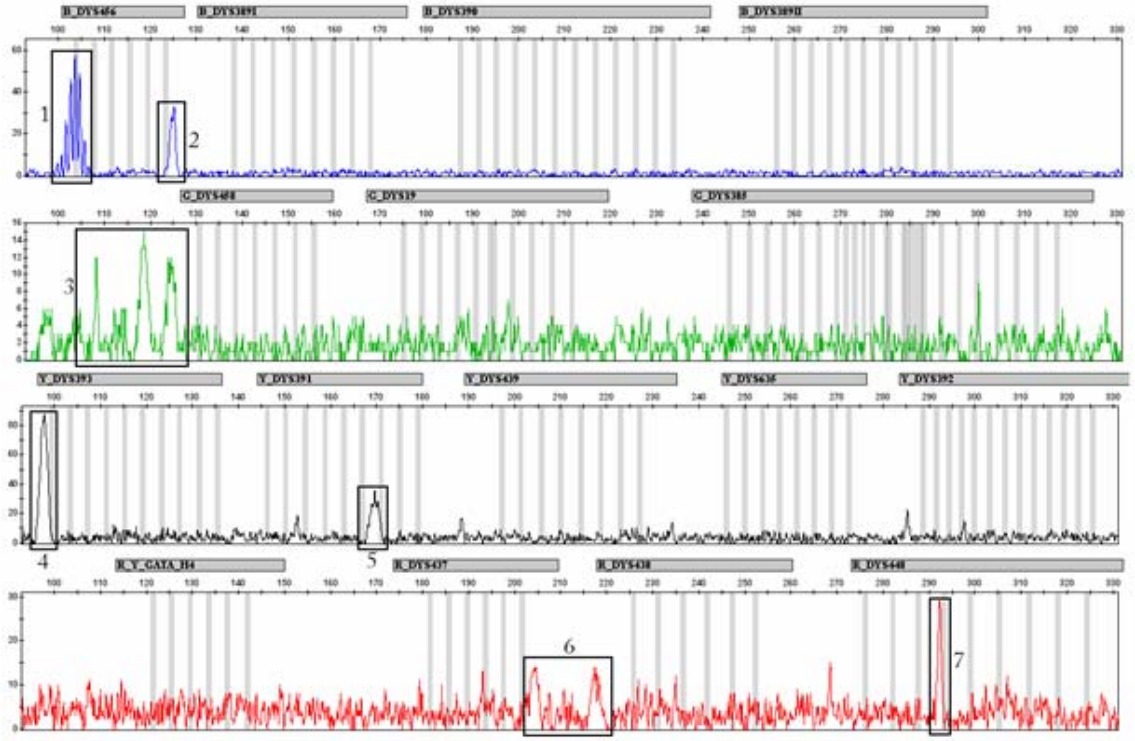


Figure 4.1.2. Specificity analysis using female DNA. The total amount of added female DNA was 400 ng. The vertical grey stripes represent known alleles and the horizontal grey bars represent the allelic area for the different markers. Possible artefacts are boxed and numbered 1-7.

Robustness analysis using male/female DNA mixtures

The results from the robustness analysis are presented in *Figure 4.1.3*. The columns in the figure represent the relative peak height ratios between the average peak heights in the robustness analysis and the average peak heights in the authentic sample analysis (see *Figure 4.2.4*). The peak height ratios have been normalised using the highest peak height ratio (Y GATA H4) as the normalising value. The figure shows the relative amplification success of the markers in the presence of excess non-target DNA. The authentic sample analysis is used as a reference data set since most of these samples were shown to consist of clean male profiles.

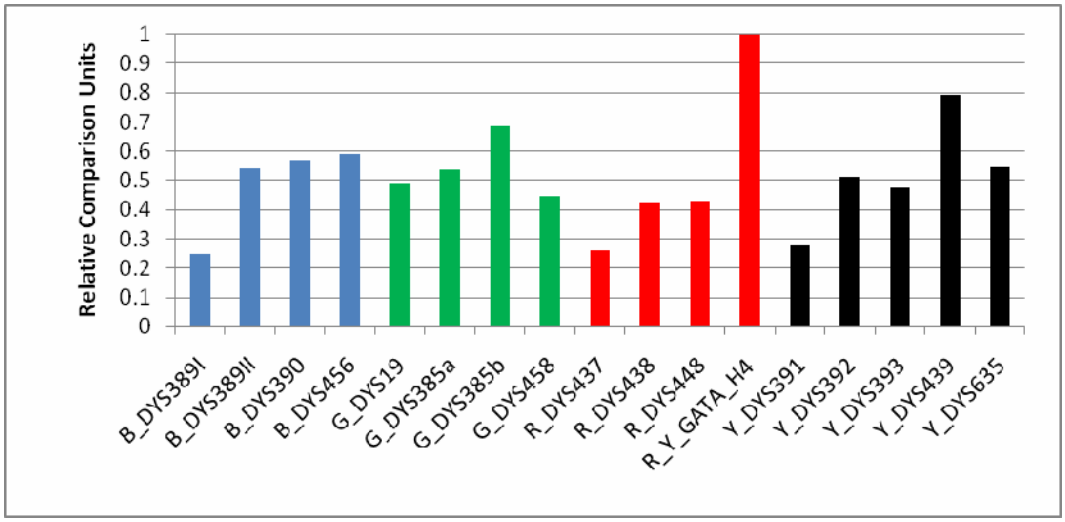


Figure 4.1.3. The relative amplification success of the markers in the presence of female DNA. The relative comparison units should only be used as internal reference values (see text for more information). The indexes of the markers correspond to their labelled dyes (B: blue; G: green; R: red; Y: yellow).

As seen in the figure, the primers labelled with the red dye (indexed with R in the figure) become unstable in the presence of non-target DNA. Y GATA H4 shows the strongest relative amplification success and DYS437 shows the weakest amplification success. The primers labelled with the green dye (indexed with G in the figure) are equally affected by the non-target DNA and maintain a relatively high intra-colour peak balance.

4.2. Authentic samples

Of the 139 samples analysed, 3 samples contained allelic drop-outs. These samples have been omitted from the peak balance and stutter calculations. Thus, the calculations below are based on a total of 136 samples.

Stutter-1 ratios

Average stutter ratios

The stutter calculations are presented as average stutter ratios (see *Figure 4.2.1*). The error bars represent a confidence interval of significance***.

Stutter sum

The sum of the total stutter ratios was also calculated as a weighted measure of occurrence and magnitude of the stutters (see *Figure 4.2.2*). The stutter ratios of each locus have been added together, and thus loci with frequent stutters will have a relatively high column height compared to the average values (*Figure 4.2.1*). Note the low allelic sums of the penta- and hexa-nucleotide repeat loci (DYS438 and DYS448).

Average allelic stutters

Stutter calculations of different alleles of the individual loci are presented as average values with confidence intervals of significance***. Two of the loci are presented in *Figure 4.2.3* and all the loci are presented in Appendix A. Note the general trend of increased stutter ratios with increased allelic repeats.

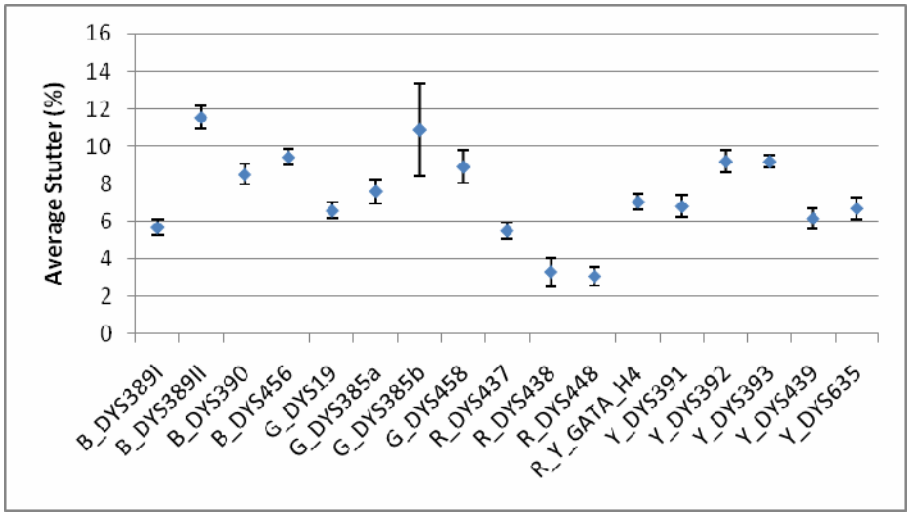


*Figure 4.2.1. Average stutter percentage of the Yfiler loci. Error bars represent a confidence interval of significance***.*

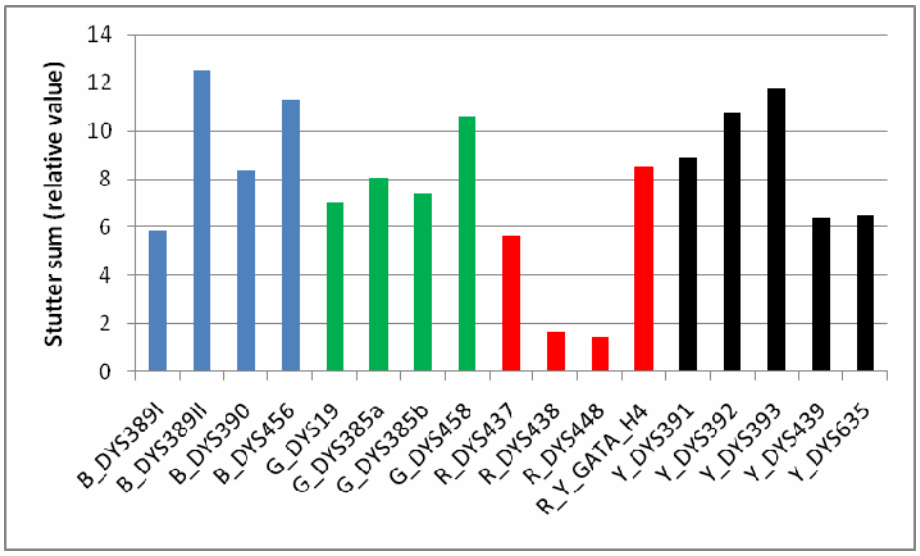


Figure 4.2.2. Stutter sum of the Yfiler loci. The column height is only to be used as a relative comparison among the loci.

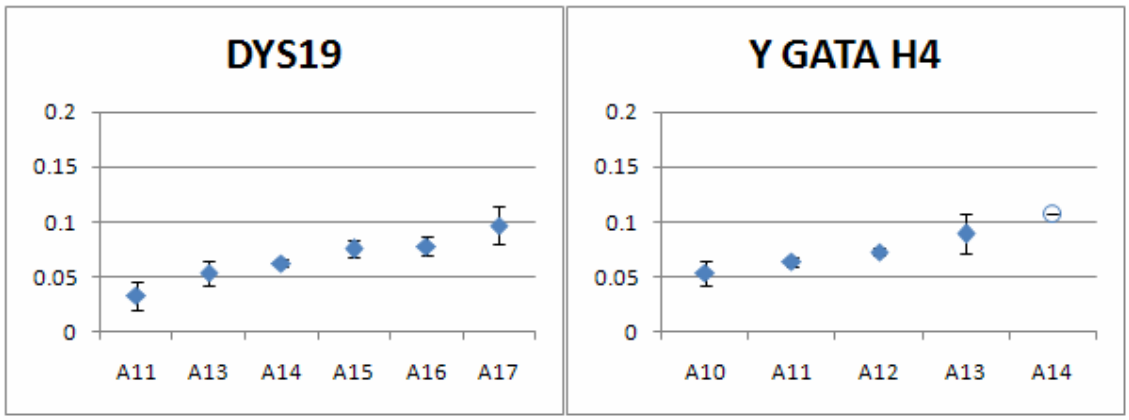


*Figure 4.2.3. Average stutter ratios of different alleles in two of the Yfiler loci (DYS19 and Y GATA H4). Single values are shown as non-filled circles. Error bars represent a confidence interval of significance***.*

Peak balance

The average peak heights of the 136 samples are presented in *Figure 4.2.4*. As seen in the figure, the amplification success of DYS391 is, on average, almost twice as high as the other intra-colour markers (DYS392, DYS393, DYS439, and DYS635). This results in a low intra-colour peak balance (as seen in *Figure 4.2.5*). The peak balance calculations are presented as average values of the quotient of the lowest and the highest intra-colour peak heights. The confidence intervals (error bars) are calculated with significance***. The DYS385 peaks were doubled in the calculations when multiple alleles were present (DYS385a and DYS385b).

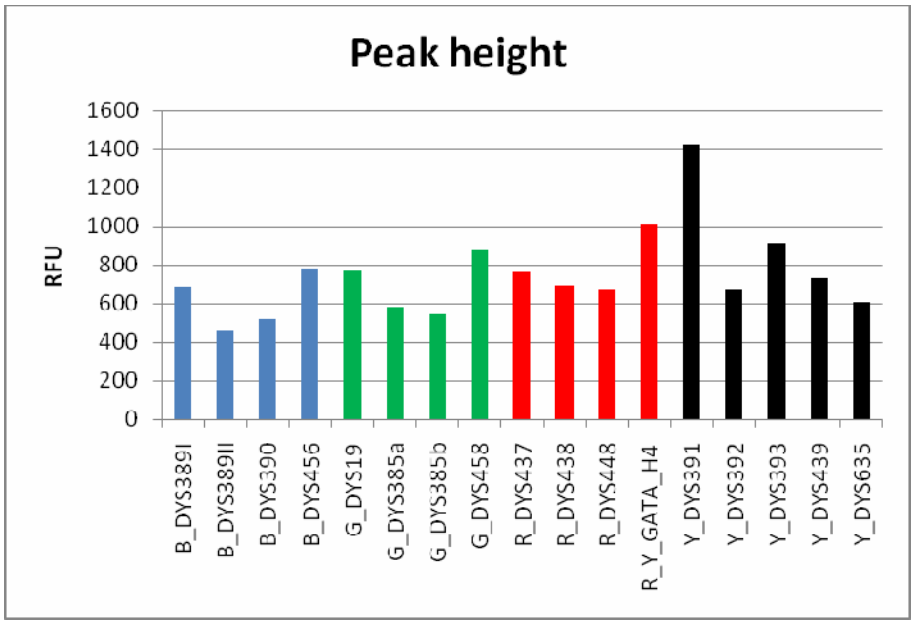


Figure 4.2.4. Average peak heights of the Yfiler loci.

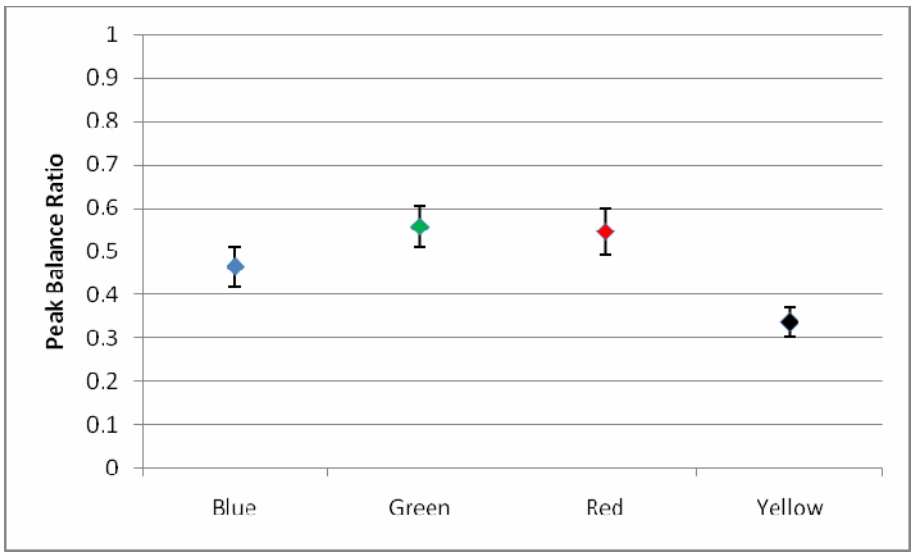


*Figure 4.2.5. Intra-colour peak balance of the different dyes. Error bars represent a confidence interval of significance***.*

4.3. Comparison study

The comparison study of 23 samples amplified with both Yfiler™ and PowerPlex®Y were analysed in respect to samples with low male DNA concentration and the ability to detect potentially mixed male profiles.

Allelic drop-outs

The total amount of allelic drop-outs were counted for each sample and analysis method. Drop-outs were detected in 5 of the 23 samples and the results of the drop-out ratio (number of drop-out loci/total number of loci) are presented in *Figure 4.3.1*.

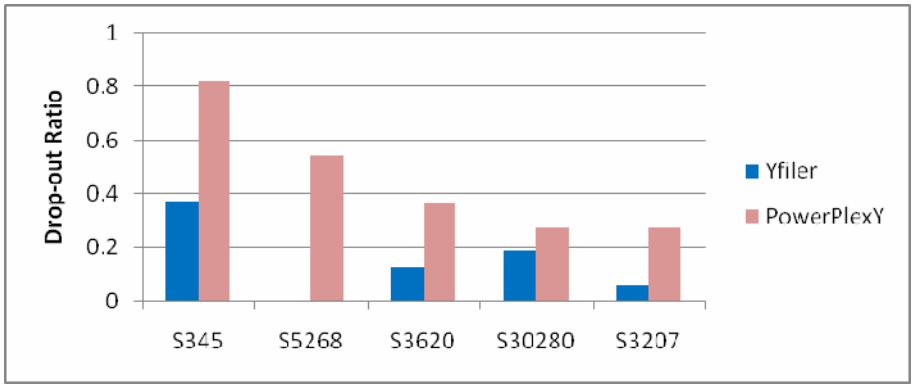


Figure 4.3.1. Drop-out ratio comparison between Yfiler™ and PowerPlex®Y.

Mixed samples

The ability to detect mixed profiles was defined as when two or more peaks were present at a locus. Minor profile peaks in stutter-1 position were interpreted as allele peaks if the peak height was higher than 15% of the major profile allele peak. 12 potentially mixed profiles were detected using PowerPlex®Y, compared to 16 mixed profiles detected using Yfiler™. The results are presented as a mixed loci ratio (number of mixed loci/total number of loci) in *Figure 4.3.2*.

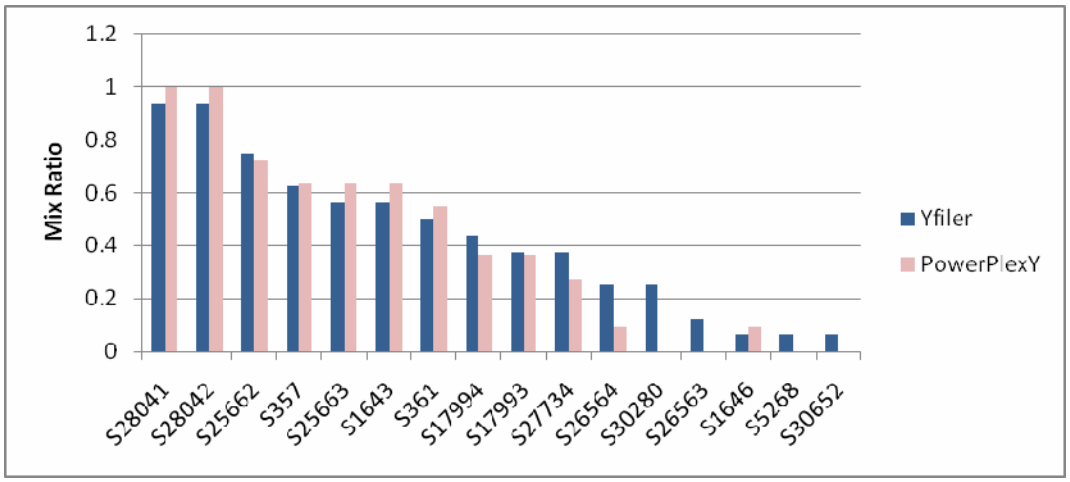


Figure 4.3.2. Ratio of potentially mixed loci. Comparison between Yfiler™ and PowerPlex®Y.

4.4. Questionnaire

The questionnaire was sent out to 15 forensic laboratories around Europe. 12 of the laboratories replied, but only nine of these had any experience of Y-STR analysis (see TABLE 4.4.1). The three-letter abbreviation in the right column will be used to indicate the answers of the laboratories in this section. Three of the forensic laboratories, located in Denmark, Ireland, and Scotland, are in the planning stages of Y-STR analysis, but as of today they have no routine protocols and chose not to answer the questionnaire.

Y-STR analysis today

Two questions concerning the commercial kits used and the current use of Y-STR analysis are presented in TABLE 4.4.2 and TABLE 4.4.3. The question and the options are presented in the middle column of the table. The three-letter abbreviations in the right column show how the laboratories answered as shown in TABLE 4.4.1.

Evaluation of partial and mixed profiles

The answers concerning the interpretation and evaluation of partial and mixed profiles are presented in TABLE 4.4.4 and TABLE 4.4.5.

The forensic report

Answers concerning the value of evidence of a Y-STR haplotype, and the statistical interpretation and presentation of the results in the forensic report are presented in TABLE 4.4.6, TABLE 4.4.7, and TABLE 4.4.8 below.

Problems with and future of Y-STR analysis

The recipients of the questionnaire were asked to grade the importance of some important developmental areas in Y-STR analysis. The results are presented in TABLE 4.4.9. Possible answers were 1, 2, 3, 4, and 5. A high number corresponds to the addressed issue being a considerate problem, and a low number corresponds to the addressed issue being an irrelevant problem. The results are presented as separate answers and as a numerical average. TABLE 4.4.10 shows the general thoughts on the future perspectives on the use of Y-STR markers.

TABLE 4.4.1. The nine laboratories that answered the questionnaire and their locations. The last column in the table represents a three-letter abbreviation used to indicate the answers throughout this section.

Forensic laboratory	Location	Abbreviation
Bundeskriminalamt	Wiesbaden, Germany	**GEW**
Forensic Science Institut (KTI) of the Central Investigation Office (LKA) of North Rhine Westfalia (NRW)	Düsseldorf, Germany	**GED**
National Institute for Forensic Sciences	Belgium	**BEL**
Crime Laboratiry/National Bureau of Investigation	Finland	**FIN**
Institute of Legal Medicine, Innsbruck Medical University	Austria	**AUS**
DNA Laboratory	Greece	**GRE**
Forensic Service Center	Estonia	**EST**
Institute of Forensic Medicine	Norway	**NOR**
Institute of Legal Medicine Zürich	Switzerland	**SWI**

TABLE 4.4.2. Some general areas of use of Y-STR. Multiple answers possible.

	To what extent do you use Y-chromosomal STR profiling?	Answers
1	Y-STR analysis is never done	
2	Y-STR analysis is used in sexual assault cases	GEW, BEL, FIN, GED, GRE, EST, NOR, SWI
3	Y-STR analysis is used in crime cases other than sexual assault cases	GEW, BEL, FIN, GED, GRE, EST, NOR, SWI
4	Y-STR analysis is used as a complement to autosomal STR profiling when mixed profiles are encountered	BEL, FIN, AUS, GRE, EST, NOR, SWI
5	Y-STR analysis is used as a complement to autosomal STR profiling when a male profile is expected	GEW, AUS, GRE, EST, NOR, SWI
6	Y-STR analysis is used as a routine analysis when mixed profiles are encountered	
7	Y-STR analysis is used as a routine analysis when a male profile is expected	
8	Y-STR analysis is used in kinship analysis to determine paternal origin	GEW, BEL, GED, GRE, EST, SWI
9	Y-STR analysis is used in identification of unknown males	GEW, GRE, EST
10	Other*	GED, NOR

*: See section 5.4 for specifications.

TABLE 4.4.3. The current use of commercial kits. Some laboratories use several Y-STR kits.

	What kit do you use for Y-chromosomal STR analysis?	Answers
1	DYSplex-1 (Serac, Germany)	
2	DYSplex-2 (Serac, Germany)	AUS
3	Mentype® Argus Y-MHQS (Biotype, Germany)	GED, AUS
4	PowerPlex®Y (Promega Corporation)	GEW, BEL, AUS, GRE, EST, NOR, SWI
5	Yfiler™ (Applied Biosystems)	GEW, FIN, GED, AUS, EST, NOR, SWI
6	Other*	AUS

*: See section 5.4 for specifications.

TABLE 4.4.4. Question regarding the interpretation of partial profiles. Multiple answers possible.

	How do you regard partial profiles?	Answers
1	A partial profile could actually be a full profile from one male with unusual chromosomal aberrations (indels)	
2	A partial profile carries no significant value of evidence	GED, SWI
3	The haplotype frequency of a partial profile should be calculated before any further assessment is done	GEW, BEL, EST, NOR
4	Other*	FIN, AUS, GRE

*: See section 5.4 for specifications.

TABLE 4.4.5. Question regarding the interpretation of mixed profiles. Multiple answers possible.

	How do you regard potentially mixed profiles?	Answers
1	A seemingly mixed profile could actually be a full profile from one male with unusual chromosomal aberrations	BEL
2	A mixed profile carries no significant value of evidence	GED, EST, SWI
3	The discriminative power of mixed profiles should be evaluated before any further assessment is done	GEW, BEL, GRE, NOR
4	Other*	FIN

*: See section 5.4 for specifications.

TABLE 4.4.6. Question regarding the value of evidence of a Y-STR profile match. One answer possible.

	How high do you think the value of evidence is of a match between a crime scene sample and a reference sample using Y-chromosomal STR analysis?	Answers
1	Very low	
2	Low	BEL, GED, NOR
3	High	GEW, EST, SWI
4	Very high	
5	Other*	FIN, AUS, GRE

*: See section 5.4 for specifications.

TABLE 4.4.7. Question regarding statistical estimation of the value of evidence of a Y-STR profile. One answer possible.

	What kind of statistical method do you use to determine the value of evidence from a Y-chromosomal STR analysis?	Answers
1	Database frequency searches are transformed to likelihood ratios. Comment below, please!	
2	Haplotype frequencies are estimated using the counting method from database searches. Comment below please!	GEW, BEL, AUS, EST, NOR, SWI
3	No statistical calculations are done to estimate the value of evidence.	FIN, GED, GRE
4	Other	

TABLE 4.4.8. Question regarding the presentation of a Y-STR analysis in the forensic report. One answer possible.

	How do you present the result from a Y-chromosomal STR analysis in a forensic report?	Answers
1	The result is presented as a numerical value or interval, describing some sort of statistically significant probability.	GEW, BEL, EST
2	The result is presented as a juridical statement with predefined intervals based on statistical calculations.	NOR
3	The result is presented as a juridical statement with predefined intervals based on estimates other than statistical calculations.	GED, SWI
4	Other*	FIN, GRE

*: See section 5.4 for specifications.

TABLE 4.4.9. The importance of some of the main issues of Y-STR analysis. The answers are presented separate and as an average. Possible answers were 1, 2, 3, 4, and 5.

	Grade the importance of some of the main problems with Y-STR analysis.	Answers	Average
1	Not enough specificity (too few STR markers to achieve satisfactory value of evidence)	2, 2, 2, 2, 2, 3, 3, 5, 5	2.9
2	Paternal inheritance reduces the discriminatory power.	1, 2, 2, 3, 3, 3, 5, 5	3.0
3	Insufficient support in reference databases makes it difficult to accurately calculate haplotype frequencies.	2, 3, 3, 3, 3, 4, 5	3.3
4	Database support does not include all Y-STR markers used in the analysis.	1, 2, 2, 3, 3, 4, 5	2.9

TABLE 4.4.10. The future of Y-STR analysis. Multiple answers possible.

	What are your thoughts about the future (in the next 5-10 years) of Y-chromosomal STR analysis?	Answers
1	The commercially available kits will continue to improve in robustness and multiplexing capacity.	GRE, EST, NOR, SWI
2	SNP analysis will probably render Y-STR analysis obsolete.	GED
3	Y-STR marker analysis doesn't have the discrepancy power to yield results that are reliable and will probably not have a significant influence on forensic science.	
4	Expanding databases will lead to more reliable results thus increasing the value of evidence.	GEW, BEL, GRE, EST, NOR, SWI
5	Other*	FIN, AUS

*: See section 5.4 for specifications.

5. DISCUSSION

5.1. Sensitivity, specificity and robustness

Male DNA sensitivity analysis and PCR optimisation

PCR optimisation

The initial optimisation part of the analysis was made in order to determine the detection limits of male DNA using different cycling conditions during the amplification step. A two-fold increasing dilution series of male DNA was used ranging from 6.3 pg/µl to 100 pg/µl. The amounts of cycles tested were 29, 30, and 31. Even if the allele peaks increase in amplitude when increasing the number of PCR cycles, an assessment of the necessity of adjusting the PCR cycles must be done. No significant differences were seen when comparing stutter occurrence and stutter ratios when the numbers of cycles were increased.

When analysing the electropherograms from the PCR optimisation a general conclusion was made to use the recommended 30 cycles. No significant improvement was seen in intra-colour peak balance at 31 cycles compared to 30 cycles and the allelic amplitude was considered to meet the requirements at 250 pg male DNA at 30 cycles. The results from the 29-cycle amplification proved to be slightly better than the 30-cycle amplification in regard to peak balance. However, the general peak height at 250 pg male DNA was considered to be too low with several peaks below 150 RFU.

The results from the duplicate run at 30 cycles showed consistent results to the previous run. The overall conclusion on the sensitivity of the kit at 30 cycles is that an average of 8 allele peaks (9 and 7 in the duplicate runs) reached the 150 RFU when amplifying 125 pg of male DNA. When analysing 250 pg of male DNA only one of the markers from the two duplicate runs had an amplitude of less than 150 RFU (DYS456).

Sensitivity analysis

The triplicate runs at 30 cycles meant to serve as a basis for determination of detection limits and cut-off values for the markers when used in routine analysis. As seen in *Figure 4.1.1*, there are some apparent amplification patterns of the different markers. Y GATA H4 usually shows the greatest amplification success rate, while markers such as DYS389I and DYS389II show weaker amplification success. Other markers, such as DYS19 (dropped out in two of the three runs at 63 pg of male DNA), DYS458, DYS438, and DYS635, became unstable when the DNA concentration was lowered. This could either be the effect of low primer concentrations in the primer mix, or due to differences in primer affinities. For a more thorough discussion on detection limits and cut-off values, see section 5.5.

Specificity analysis using female DNA

The specificity analysis was made using a female DNA dilution series to analyse if unspecific binding to regions of autosomal DNA was observed in the presence of high female DNA concentrations. The dilution series consisted of 7 samples ranging from 4.2 ng to 400 ng of female DNA. Some baseline fluctuations were seen, but no detectable allele peaks were identified at any of the concentrations.

Only the sample with the highest female DNA concentration will be discussed here, since the artefacts and baseline fluctuations were most prominent in this sample. Some of the artefacts and re-occurring baseline fluctuations seen in the electropherogram of the highest

concentrated female DNA sample (*Figure 4.1.4*) occurred at lower concentrations and even in the negative control samples.

As seen in *Figure 4.1.2*, there is a jagged artefact in the allelic area of DYS456 (marked 1 in the figure). The pattern of this artefact is similar to mono-nucleotide peaks presented in previous mono-nucleotide stutter studies (Shinde 2003), and the artefact could be a mono-nucleotide repeat sequence amplified due to unspecific binding. The jagged peak to the right of the allelic bins in the DYS456 allelic area (marked 2) is present in all female DNA samples. This could indicate unspecific binding of one of the primers used, but even at the highest female DNA concentrations, the peak heights of the artefacts rarely reached the values of 30 RFU.

The artefacts marked 3-6 were present in all female DNA samples as well as negative control samples. These peaks could be the result of primer-dimer formations or the detection of other substances in the analysed mixtures. These peaks were not occurring in the allelic areas, and rarely reached amplitudes of more than 20 RFU.

The artefact marked 7 in Figure 4.1.2 is only occurring in the sample with the highest female DNA concentration. This peak could either be the result of unspecific primer association, or spectral bleed-through from the internal lane standard (ILS).

To summarise, no allele peaks apart from some re-occurring artefacts with relatively low peak heights were detected as seen in *Figure 4.1.2*. These results show that the method is highly specific to male DNA, e.g. Y chromosomal DNA.

Robustness analysis using male/female DNA mixtures

Male/female DNA mixtures were analysed to observe the differences in amplification success of the different markers in the presence of excess autosomal DNA. The female DNA acts as a bulk of inert DNA, and the primers might be inhibited by sterical hindrance or unspecific binding. Applied Biosystems has not publically released the primer sequences, and this makes it impossible to make any *a priori* suggestions of which primers that are most prone to be effected by excess DNA.

The degree of primer inhibition due to sterical presence of excess autosomal DNA was small, even when the highest ratios of female DNA (1:1000) were used. However, the peak balance of the mixed samples was on average lower than the peak balance of the male DNA sensitivity analysis as well as the average peak balance of the authentic sample analysis (data not shown). The lower peak balance could be the result of lower male DNA concentrations due to stochastic differences in amplification success and is therefore not to be seen as a statistically significant correlation. To verify that excess non-binding DNA lowers peak balance, a larger comparison study using male/female DNA mixtures and male DNA samples should be set up using the same male DNA concentrations in the mixtures and the clean samples.

Since the average DNA concentrations and experimental set-ups differ between the authentic sample analysis and the robustness analysis, the only comparison possible is a relative amplification success rate analysis (*Figure 4.1.3*). The figure presents the amplification rates as normalised values referred to as relative comparison units (RCU). Y GATA H4 in particular has the highest relative amplification success rate compared to the results of the authentic

sample analysis, and the markers have been normalised using the value of Y GATA H4 as normalising factor.

The markers labelled with the red dye show the lowest robustness in the presence of excess DNA with both the highest and the lowest amplification success rate (Y GATA H4 and DYS437, respectively), and this reflects a poor relative peak balance (data not shown). The green dye presents the most stable results, with a relative amplification success rate of 0.45 RCU to 0.69 RCU (*Figure 4.1.3*). No significant differences in amplification success and peak balance could be seen between the different ratios of excess DNA (1:50, 1:200, 1:400, and 1:1000). This indicates that excess autosomal DNA has a low inhibitory effect on the amplification.

The peak heights of the mixture analysis were compared to the average peak heights of the male sensitivity analysis as well (data not shown). The results indicated that some of the differences in amplification success could be due to low DNA concentrations in the samples.

Increased baseline fluctuations discussed in the specificity analysis above could be an issue in allele peak interpretation. However, the detectable allele peaks, even at the lowest male DNA concentration used in this analysis (63 pg), proved to be significantly stronger than the baseline.

To conclude the robustness analysis, excess non-binding autosomal DNA could lower the peak balance and general amplification success. However, these effects are small, and even ratios of male DNA as low as 1:1000 could be amplified with a high success rate.

5.2. Authentic samples

139 authentic case samples were analysed as a part of the validation study. The samples were chosen to cover all the extraction methods used at SKL as described earlier in the report. Amplification failure with resulting allelic drop-outs occurred in three of the samples. The probable explanation for these drop-outs is dilution errors, since quantification and routine autosomal STR analysis suggested the presence of a male profile. These samples were omitted from further calculations, and thus 136 samples were included in the analysis. Stutter occurrence, differences in peak balance, and re-occurring artefacts based on these samples will be discussed in this section.

Stutters

Mixed male profiles can be disguised in the allele peaks or the stutter peaks of the primary male profile. Since stutter peaks have been shown to occur -2, -1, and +1 repeat sequences away from the actual allele, there is a high degree of freedom for a secondary male profile. This could explain some of the unusually high stutter peaks and the occurrence of -2 stutters. However, the samples were selected using the results from autosomal STR analysis, and mixed male profiles detected at that stage were not considered.

The detection limit for allele peaks as well as stutter peaks was set to 20 RFU. Peaks lower than these were not used in any calculations since baseline fluctuations could influence the peak size as much as 10 RFU. Indeed, even peaks with the magnitude of 20-100 RFU are influenced by baseline fluctuations, and any interpretations and calculated values using these

peaks have to be used with caution. A t-test was made using the stepwise Sidak statistical model[4] to compare stutter peaks of amplitudes of between 20 RFU – 50 RFU to stutter peaks of amplitudes higher than 50 RFU. The test showed statistically significant differences for one out of 55 comparable alleles (data not shown). This discrepancy should not influence the results presented in this report, but a high degree of significance (***) was used when interpreting the results of the stutter analysis to compensate for this.

Stutters in -1 position

Stutters were by far most common in -1 position. All loci produced stutter peaks in -1 position, although the occurrence and amplitude of the stutter peaks differed significantly between different loci (see *Figure 4.2.1*). As seen in the figure, the highest average stutter peaks occurred in DYS389II, DYS392, DYS456, and DYS393 (for repeat structures, see TABLE 2.6.1). These loci present common patterns of stutter-rich repeat structures as discussed below:

DYS389II: DYS389I and DYS389II partly amplify the same STR repeat. However, DYS389II amplifies two closely spaced repeat regions, while DYS389I only amplifies one of the repeats (for structure comparison, see TABLE 2.6.1). The differences in stutter occurrence are apparent as seen in *Figure 4.2.1* and *Figure 4.2.2*. According to these results, two repeat sequences amplified together seem to double the stutter occurrence and the stutter ratio. The individual STR repeat lengths are not unusually long compared to other STR loci, which imply that the stutter formation occurs independently in the two amplified STR regions.

DYS392: This is a tri-nucleotide STR locus (see TABLE 2.6.1). Earlier reports on stutter occurrence and repeat length (Shinde 2003) show that increased repeat lengths reduces the average stutter height. The repeat sequence of DYS392 is uniform $(TAT)_n$, which further increases the chance of the polymerase to slip. The total lengths of the alleles (number of repeats) are not as high as in many of the other loci (Appendix A), and a stutter height to allelic length analysis would undoubtedly result in DYS392 having the highest stutter height to allelic length ratio (data not shown).

DYS456 and DYS393: These loci both have a repeat sequence of $(AGAT)_n$. Uniform repeat sequences have been shown to increase stutter ratios compared to broken repeat sequences (Klintschar and Wiegand 2003). This holds true in this analysis as well, since many of the loci with broken repeat sequences (e.g. DYS437 and DYS635) have significantly lower stutter ratios.

The lowest stutter ratios were seen in the loci with repeat sequences of more than four nucleotides (DYS438 and DYS448). These results are consistent with previous results (Shinde 2003), and both the occurrence of stutters and average stutter ratios (see *Figure 4.2.1* and *Figure 4.2.2*) were significantly lower at these loci than in all the other loci. The similar results of the penta-nucleotide repeat (DYS438) and the hexa-nucleotide repeat (DYS448) could be explained by the longer allele sizes (total repeat number) in the DYS448 alleles than in the DYS438 alleles (see Appendix A). The longer allelic lengths compensates for the differences in repeat structure.

Trends of increasing stutters in position -1 were observed with increasing allele sizes. This trend was significant in almost all loci barring a few exceptions (see Appendix A). The results

[4] Personal communication: Joakim Englund, Statistician, Ashford, United Kingdom (June 2007).

of the DYS635 locus with the decrease of stutter ratios in the A23 allele could be explained by a point mutation in the repeat sequence, which would decrease the probability for the polymerase to slip. The stutter -1 patterns of all the loci, including DYS635, were highly consistent with previous validation results obtained by Applied Biosystems (Applied Biosystems 2006)(data not shown).

DYS385a/b: The differences in average stutter percentage between DYS385a and DYS385b seen in Figure 4.2.1 is the result of the definition of the two alleles. The longer of the two alleles is always referred to as DYS385b, which would imply that the average stutter ratio is higher. The larger spread in the confidence interval is a natural effect of this, since the spread in allele lengths is larger in DYS385b than in any of the other loci, where alleles were observed ranging from 13 to 22 repeats (data not shown).

Stutters in -2 position

Stutter-2 peaks rarely reached the cut-off value of 20 RFU. In fact, many of the loci showed no signs of stutter-2 peaks at all. Because of this, any quantitative analysis of the -2 stutters would be highly unsecure, and only general conclusions on the occurrence of the stutter peaks will be discussed here.

The most remarkable observation in the stutter-2 analysis was the unsuspected pattern of DYS448. The stutter-2 peaks, when observed, were almost as high as the stutter-1 peaks (data not shown). This could be due to the formation of a hairpin structure, as illustrated in *Figure 5.2.1*. When the polymerase dissociates, the formation of the hairpin structure could prevent the polymerase to re-associate to the same site, and thereby causing the stutter-2 slippage. The formation of this hairpin structure is highly speculative, and is not supported by any reference literature or thermodynamic calculations.

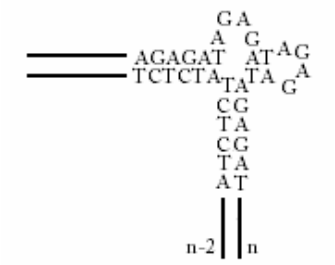


Figure 5.2.1. Possible replication slippage model of DYS448 $(AGAGAT)_n$. A 2-repeat hairpin formation results in the creation of a stutter-2 strand (n-2).

Other loci where stutter-2 peaks were relatively common were DYS385, DYS458, DYS391, DYS392, and DYS393. These loci all consist of uniform repeat sequences, as seen in TABLE 2.6.1, which implies that these stutter-2 and stutter-1 formations follow the same biochemical mechanisms.

Stutters in +1 position

As described earlier in the report, the formation of a stutter+1 strand is energetically less favourable than the formation of a stutter-1 strand. This is concordant with the results obtained, where stutter+1 peaks were rare. DYS392, however, showed amplification patterns where stutter+1 peaks were almost as common as stutter-1 (data not shown). DYS392 consists of uniform tri-nucleotide repeat sequences, and uniformity and repeat length are probably the most important factors in the stutter+1 formation.

Intermediate-sized stutters

DYS19 usually showed an amplification pattern of a stutter-1 peak, and an intermediate-sized stutter peak 2 nucleotides shorter than the allele peak (referred to as allele-2). The occurrence of this peak has been omitted from the stutter analysis, but the height and occurrence was usually of the same amplitude as the stutter-1 peak.

Peak balance

The peak balance analysis was made to evaluate the robustness of the multiplex reaction. The balance has been calculated through intra-colour ratios between the loci with the lowest and the highest amplification success. The results of the different dyes will be discussed, as well as differences in peak balance at low DNA concentrations.

Intra-colour peak balance

The multiplex reaction is tuned to maximise allelic peak balance. This, however, may be a hard task, due to differences in annealing temperatures of the primers and due to unspecific binding to autosomal DNA. Primer concentrations have been set to compensate for these problems, and concentrations vary between 0.12-1.5 μM (Mulero 2006). However, the primer sequences and the individual primer concentrations are not released by Applied Biosystems, and this renders affinity analysis impossible.

There are three main reasons of low peak balance. The first is skew results due to degraded DNA samples, where the shorter amplified fragments will have a higher amplification success than the longer fragments. The second reason is the possibility that the primer concentrations are biased, leading to high amplification success of the fragments with primers of high relative concentrations. Finally, the third reason would be differences in amplification success due to inhibition of primer sites or general inhibition of the polymerase, which would result in low amplification success, especially for the primers with low binding affinities.

The average peak heights of the loci in *Figure 4.2.4* show that the amplification success of the marker DYS391 is relatively high. In fact, it is almost twice as high as the rest of the intra-colour markers. DYS391 has an allelic size window of roughly 145-180 nucleotides. Five markers have, on average, shorter amplified STR fragments, which would rule out degraded samples as the sole factor for this biased peak height. This leaves inhibition and, more likely, skew primer concentrations the probable reasons for this high amplification success. As seen in *Figure 4.1.3*, the relative amplification success rate of the markers in the presence of excess non-binding DNA shows that DYS391 amplifies poorly. It remains unclear if the low amplification success rate is due to unspecific binding of the primers to autosomal regions of the DNA, or if the binding affinities of the primers to their binding sites are low.

There are significant differences in peak balance when comparing the different dyes (see *Figure 4.2.5*). The reasons for these differences could partly be explained by the number of amplified markers of each dye. More markers should result in lower peak balance. The labelled primers of the yellow dye (NED™) amplify five markers, while the blue and the red dyes (6-FAM™ and PET®) amplify 4 markers each. Three primers are labelled with the green dye (VIC®), and two of the markers (DYS385a/b) are amplified by the same primers. As seen in the figure, the yellow dye shows the significantly lowest peak balance of the dyes.

Low peak heights

Another possible effect on the balance calculations is the interpretation of allele peaks close to the baseline. The baseline usually fluctuates up to and sometimes over 10 RFU, and interpretation of peaks in the range of 10-100 RFU could be influenced by this.
When discriminating the selection to allele peaks where the average peak height is higher than the cut-off value of 100 RFU, the intra-colour peak balance stabilises (data not shown). This indicates that a cut-off value is vital for the interpretation of mixed male profiles, since random variations in peak balance impedes the possibility to identify major and minor male allele peaks.

Artefacts

The artefact patterns discussed previously in this report were consistent with the results in the authentic sample assay. Since the artefacts were all low in peak height and not occurring in any allelic positions, the impact on interpretation of the results was small.

However, spectral bleed-through between the dyes proved to be high in some samples, where peaks heights comparable to the stutters appeared in the lanes with close spectral relation (e.g. between NED™ and PET® or between 6-FAM™ and VIC®). Bleed-through was also observed from the internal lane standard (ILS), mainly resulting in extra peaks in the red and yellow lanes (PET® and NED™). The bleed-through rarely resulted in peaks in the allelic positions, but could potentially amplify stutter peaks if the peaks overlap. Spectral bleed-through ratios were increasingly observed with increasing allelic peak heights (usually peaks >1000 RFU), and could be avoided by diluting samples with high male DNA concentrations.

5.3. Comparison study

23 DNA crime case samples were chosen for a comparison study between Yfiler™ and PowerPlex®Y. The samples were amplified simultaneously to obtain comparable results. Many of the samples had earlier presented questionable results, where some samples indicated the presence of sperm cells, but no trace of male DNA could be detected in routine autosomal STR analysis.

Allelic drop-outs

The total amount of allelic drop-outs was counted. A ratio of allelic drop-outs/total number of loci was calculated and presented earlier (see *Figure 4.3.1*). As seen in the figure, drop-out ratios were higher when analysing the samples using PowerPlex®Y, which could be the effect of non-optimal electrophoresis settings. The samples included in the drop-out analysis generally presented low peak heights, probably due to low concentrations of male DNA.

Mixed samples

Mixed male profiles could result in interpretation difficulties of haplotypes and haplotype frequencies of the major and minor DNA contributor if they are hard to distinguish. These problems include low intra-colour peak balance, stutter occurrence, and other related issues involved in the interpretation of mixed profiles (Ansell 2005).

The average intra-colour peak balance for the yellow dye (NED™) was below 50% (*Figure 4.2.5*), and this implies that DNA mixtures where the ratio between the minor and the major contributor is higher than, or just below, 1:2 would be difficult to interpret.

The stutter percentage of some of the markers (DYS389II in particular) often exceeded 10% for the longer alleles (see Appendix A). The stutter ratio sets the lower boundary for mixed male samples. This implies that an allele peak of a minor contributor with a DNA concentration ratio of less than, or just above, 10% of the major DNA concentration could go undetected in some loci if the minor allele is in stutter-1 position or in allele position of the major allele.

When comparing the samples analysed using Yfiler™ and PowerPlex®Y, the 17 loci amplified by Yfiler™ present increased possibilities to detect mixed profiles (see *Figure 4.3.2*). This is mainly due to increased haplotype diversity and the added gene diversity of more loci. This comparison is highly qualitative and is only meant to serve as a general overview of the gains of more included loci. Haplotype diversity comparisons and linked allele analysis are out of scope for this validation study, but could present a helpful insight if a quantitative comparison of the two Y-STR kits is done in the future.

Since the comparison assay was made, a new and improved primer mix for PowerPlex®Y has been released from Promega Corporation. The primer sequences are unchanged, but the improved primers are synthesised using one specific dye isomer, compared to previous multi isomeric dyes. This should, according to Promega, improve sensitivity and spectral separation of the different dyes. Internal unpublished comparison analyses performed at SKL have shown that the fluorescent signals of some loci were twice as high when using the improved primers. The improvements in the primer mix renders peak height and sensitivity comparisons of the sub-optimal analysis redundant, since those variables most probably have been affected.

5.4. Questionnaire

The answers to the questions of the questionnaire will be discussed separately in the following section using the same classification as in section 4.4.

Y-STR analysis today

To what extent do you use Y-chromosomal STR profiling?

The question regarding the use of Y-STR analysis was designed to overview the possibilities of Y-STR markers and the general reliance towards the analysis method. As seen in TABLE 4.4.2, Y-STR analysis is used in crime cases as well as kinship analysis and to identify unknown males. The answers in the 10th category (Other) specified the use of Y-STR to cases when autosomal analysis fails, or to support autosomal STR results. However, the reliance of the results is low, since none of the forensic laboratories use Y-STR analysis as a routine method when a male profile is expected. This is probably due to the fact that the discriminatory capacity of autosomal STR analysis is high enough on its own, and Y-STR analysis is seen as a complement when normal autosomal analysis fails.

What kit do you use for Y-chromosomal STR analysis?

There are a few commercial kits on the market today (see TABLE 4.4.3). Many of the laboratories use several kits, and the most commonly used kits were Yfiler™ and PowerPlex®. Two of the laboratories used the Mentype® Argus Y-MH^{QS} kit. Due to patent rules, this kit is only available to laboratories in German speaking countries. One of the laboratories (located in Austria) also uses a homemade kit specified in the 6th category (Other).

Evaluation of partial and mixed profiles

How do you regard partial profiles?

All but two of the laboratories use the information of partial profiles, but mainly on a qualitative inclusion/exclusion basis (see TABLE 4.4.4). The value of evidence of a partial profile is low, and three of the answers pointed out this in the fourth option (Other).

How do you regard potentially mixed profiles?

Mixed profiles are generally handled with caution, if handled at all (see TABLE 4.4.5). A mixed profile could still be used for inclusions/exclusions, but the value of evidence is low. This scepticism is well founded, since the number of possible haplotypes grows exponentially for every locus with two or more allelic variants.

The forensic report

How high do you think the value of evidence is of a match between a crime scene sample and a reference sample using Y-chromosomal STR analysis?

As seen in TABLE 4.4.6, one third of the laboratories regard the value of evidence as high. However, two thirds of the laboratories regard the value of evidence as too low to give any reliable results, unless used as a method for inclusion/exclusion of suspects. This, and the use of the method as a complement to autosomal analysis, is specified by the laboratories that checked the fifth option (Other). Thus, results from Y-STR analysis alone is not considered to present strong enough evidentiary support on its own, but is to be used for inclusion/exclusion only or as a complement to autosomal analysis.

What kind of statistical method do you use to determine the value of evidence from a Y-chromosomal STR analysis?

There is a noticeable divergence in the evaluation of the results of this question (see TABLE 4.4.7). Six of the laboratories calculate haplotype frequencies of some sort and use database references to determine the value of evidence of the haplotypes. The remaining three laboratories interpret the results without the use of database references and haplotype frequency estimations. This discrepancy among the laboratories is probably due to the novelty of the analysis method and the lack of local databases with reliable haplotype frequency estimations. If Y-STR analysis is to become a part of routine analysis, international guidelines concerning the use of reference databases and haplotype frequency estimations should be established.

How do you present the result from a Y-chromosomal STR analysis in a forensic report?

Three of the laboratories present the results from the Y-STR analysis as a numerical value or interval of some sort, describing the probability that the suspect deposited the sample (see TABLE 4.4.8). The two answers in the fourth category (Other) comment on the lack of local databases and the use of Y-STR analysis as a qualitative method for exclusion. In general, the scepticism towards Y-STR analysis is directed towards the lack of reliable local databases and the linear paternal inheritance, which includes all males sharing the same paternal lineage.

Problems with and future of Y-STR analysis

Grade the importance of some of the main problems with Y-STR analysis.

Four known bottlenecks were addressed in this question (see TABLE 4.4.9) and the recipient laboratories were asked to grade these according to their importance and impact on the future of Y-STR analysis. Paternal inheritance and the lack of sufficiently large and accurate databases were seemingly seen as slightly more troublesome than the specificity of the Y-STR markers. However, as seen in the table, the results varied substantially between different

recipients of the questionnaire, and these diverging results present further indications that there is still a lot of ground to cover before Y-STR analysis will become an accepted routine analysis method throughout Europe.

What are your thoughts about the future (in the next 5-10 years) of Y-chromosomal STR analysis?

Most of the recipients seem to believe in a bright future of Y-STR analysis (see TABLE 4.4.10). Improvements in the commercial kits and the available reference databases will likely improve the discriminatory power of the method. Two of the laboratories, however, checked the fifth option (Other), and stated that the paternal inheritance and complications due to haploid markers would render Y-STR analysis a tool only used in individual cases and as a method used to exclude suspects. It remains to be seen whether or not Y-STR analysis will continue to expand and become a part of the routine analysis in forensic laboratories, or if it will stabilise at the present level as a complement to currently used analysis methods. In fact, one of the laboratories even suggested that Y-SNP analysis would render Y-STR analysis obsolete in the future. This is along the lines of the author's thoughts on the future of forensic DNA analysis in general, since the potential of SNP analysis and microarray technology seem to be vastly greater than the multiplex amplification of a handful of STR markers. Only time will tell the direction of the development in the forensic community.

5.5. Cut-off values and detection limits

DNA concentrations

The results of the male DNA sensitivity assay showed that many of the loci became unstable when the DNA concentration was lowered. This raises the question of whether to use detection limits and cut-off values or not. The introduction of cut-off values is an important tool to standardise the interpretation of profiles in routine analysis, and an essential step to guarantee scientific reliance.

The sensitivity analysis indicated that 250 pg of male DNA gave reliable results with low probability of allelic drop-outs. In fact, even as low amounts as 125 pg proved to be enough to amplify all loci, but with unsecure and diverging peak heights. A fair suggestion for a detection limit based on these results would be 150 pg of male DNA. Given that the sample is not inhibited or degraded, this amount would be enough to obtain reliable results with most peak heights well over 150 RFU. An expanded sensitivity assay with more intermediate amounts of DNA would give a more accurate estimation of the detection limit.

Mixed profiles should not be evaluated at all below the detection limit, since stochastic differences in peak height and naturally occurring baseline fluctuations could result in misinterpretation of the results.

Individual loci

A blunt estimation of cut-off categories based on the results of the sensitivity assay in this report (see *Figure 4.1.1*) is presented in TABLE 5.5.1. The loci are clustered into three classes: high (H), medium (M), and low (L) amplified loci. DYS385a/b should be considered as two separate loci in respect to peak height cut-off interpretation. If the two loci are of the same lengths, the cut-off value should be upgraded one level, from medium to high. These relative cut-off categories could be translated into absolute cut-off values, but would require the discussion of appropriate detection limits to avoid the risks of poor peak balance and allelic drop-outs.

TABLE 5.5.1. Suggested cut-off categories based on the male DNA sensitivity analysis.

	DYS389I	DYS389II	DYS390	DYS456	DYS19	DYS385a	DYS385b	DYS458	DYS437	DYS438	DYS448	Y GATA H4	DYS391	DYS392	DYS393	DYS439	DYS635
Amplification category	L	L	L	M	M	M	M	M	L	L	L	H	M	M	M	M	L

Baseline fluctuations

The baseline occasionally results in noise spikes (narrow peaks) of more than 10 RFU. The risk of erroneously interpreting these spikes as allele peaks is low, but to avoid excess information provided by the interpretation software, an absolute lower limit of detection should be set to 20 RFU. Values above 20 RFU should include stutters as well as allele peaks, since stutters with peak heights lower than 20 RFU could be severely influenced by the baseline. This could lead to the erroneous interpretation of a mixed profile if the resulting stutter ratio is unusually high. To further avoid the risk of misinterpretation, a more conservative detection limit could be set to e.g. 50 RFU, which is the detection limit used in autosomal STR analysis.

5.6. Haplotype frequency estimations

The haplotype databases estimate haplotype frequencies using the counting method and statistical *a priori* methods (Institute of Legal Medicine, Charité - Universitätsmedizin Berlin 2007). These Bayesian estimations are based on haplotype frequencies from closely related haplotypes. The statistical calculations provide the possibility to increase the power of evidence based on the contents of the database, since the counting method is believed to be too conservative when calculating rare haplotype frequencies. The lowest haplotype frequency (f) measurable, using only the counting method, corresponds to the incident where no entries in the database match the haplotype as seen in EQUATION 5.6.1 below.

EQUATION 5.6.1. $$f = \frac{1}{N+1}$$

The frequency simply corresponds to the inverse of the sum of the total number of entries in the database (N) and the analysed sample. This frequency estimating method could result in skew frequencies if the database is not large or accurate enough and if the calculations are not complemented with a Bayesian *a priori* method. The Bayesian methods in the YHRD database use frequencies of closely related haplotypes to estimate frequencies of uncommon haplotypes. The assumption that the haplotypes have evolved from the same ancestral root is the basis of this estimation (Krawczak 2001). Thus the frequencies of closely related haplotypes could be used as an indication of the frequency of the uncommon entry. The mathematical and statistical details of Bayesian methods are out of scope for this analysis. However, the use of these methods helps to increase the resolution of the databases, especially for rare haplotypes.

Local variations of haplotype diversity are influenced by the gene flow between meta-populations and the mating habits within the population. A genealogical study on haplotype diversity in different Swedish geographical populations (Karlsson 2006), using a reference

population in Österbotten (Finland), showed that the haplotype diversity in the Österbotten population was unusually low with skew haplotype frequencies. *Karlsson et al* suggests that this could be the result of a potential founder effect resulting from a small, isolated population.

Regardless of the local variations in haplotype frequencies, the mutation rates of Y-STR regions are high compared to the average mutation rate of the human genome (Kayser 2000, Heyer 1997). This implies that mutational forces will counteract the fixation of one specific haplotype in the worldwide human population. However, in small, relatively isolated meta-populations that lack random mating, the haplotype frequencies can differ significantly from the global population. These differences make haplotype frequency estimations based on blunt database references almost impossible to interpret. The estimated frequency depicts the probability of encountering the specific haplotype in an unrelated individual with the general, and often erroneous, assumption that the population frequencies coincide with the database frequencies. If the database is large enough, this assumption could be useful. However, when estimating local frequencies, which is usually the case in forensic casework, the global frequency is of lesser interest[5].

One substantial issue when creating Y-STR databases is the need for vast amounts of input data. When estimating autosomal genotype frequencies, the stochastic variables of the allele frequencies are statistically independent due to the fact that the markers are located on different chromosomes. This enables the use of a cumulative allele frequency calculation (Butler 2005). Because of this, there is no need for large databases. But with lineage markers, such as mitochondrial and Y chromosomal markers, individual allele frequencies are not statistically independent. This forces the use of complete haplotype frequency calculations, and, as a consequence, increases the need for large databases.

The need for large local databases in order to make reliable estimations of haplotype frequencies is perhaps the main bottleneck when determining the discriminative power of Y-STR profiles. But, even if reliable local databases are to be created, the possibility that the perpetrator originates from a different meta-population remains. This would then instigate the use of dual frequency estimations, where one estimate depicts the frequency under the assumption that the offender is a part of the local population, and where the other estimation depicts the global haplotype frequency.

5.7. Conclusions

The future of Yfiler™

The results obtained in the validation study are consistent with previous validation studies performed by Applied Biosystems in respect to sensitivity, specificity, robustness, and stutter ratios. These results should prove that the method is accurate and sensitive enough to meet the demands as an analysis method at SKL.

The comparison study between Yfiler™ and PowerPlex®Y showed results slightly favouring Yfiler™ as the more sensitive and accurate method of the two. This, together with the obvious discriminatory advantages of more included loci, leads to the recommendation to implement Yfiler™ as Y-STR analysis method at SKL.

---

[5] Personal communication: Weine Drotz, Forensic Adviser. SKL, Linköping, Sweden (March 2007).

The future of Y-STR analysis

Y-STR analysis is still under development, and there is a lot of ground to cover before any results attained will be accepted as strong evidentiary support. The results of the questionnaire shows that this seems to be the general view on Y-STR analysis in the forensic community, where the analysis method is used as a complement to autosomal STR analysis and as a qualitative inclusion/exclusion method.

Haplotype reference databases will undoubtedly continue to grow, resulting in more reliable local haplotype frequencies. The imminent question still to answer will be if the information in the databases can be used to accurately estimate a frequency with enough evidentiary impact to sustain a criminal investigation on its own.

Acknowledgements

Christina Valgren and Marianne Mattsson, SKL
For helping me throughout this project, for patiently answering all my stupid questions, and for proof-reading my report on your spare-time

Dr Ricky Ansell, SKL
Scientific reviewer

Nicola J Oldroyd, Applied Biosystems
For answering questions regarding the haplotype database

Linda Albinsson and Katarina Bjuhr, SKL
For DNA typing and helping me with the in silico *analysis*

Ulla Sjunnesson, SKL
For showing me how to secure evidence

Weine Drotz, SKL
For helping me to understand frequency estimations and pointing out the need for local databases

Ian Jamieson, SKL
For proof-reading the questionnaire

Tomas Johansson, SKL
For introducing me to the rules and general guidelines at SKL

Joakim Englund
For helping me with the statistical calculations

All members of the extraction group at SKL
For helping me when I got lost in the laboratory, and for sharing your coffee breaks with me

All employees at the Biology Department at SKL
For making me feel welcome

References

Ansell, Ricky. *Tio år med DNA.* SKL report, Linköping: SKL, 2005:02.

Applied Biosystems. "AmpFℓSTR® Yfiler™ PCR Amplification Kit User's Manual." 2006.

Applied Biosystems. *Yfiler Haplotype Database.* 2007. http://www.appliedbiosystems.com/yfilerdatabase/ (accessed March 29, 2007).

Bachtrog, D. "A dynamic view on sex chromosome evolution." *Current Opinions in Genetics & Development*, 2006: 16:578-585.

Butler, John M. *Forensic DNA Typing*. Second Edition. London: Elsevier, 2005.

Diaz, RS, and EC Sabino. "Accuracy of replication in the polymerase chain reaction. Comparison between Thermotoga maritima DNA polymerase and Thermus aquaticus DNA polymerase." *Brazilian Journal of Medical and Biological Research*, 1998: 31:1239-1242.

Ellegren, Hans. "Microsatellites: simple sequences with complex evolution." *Nature*, 2004: 5:435-445.

Graves, J. "Sex Chromosome Specialization and Degeneration in Mammals." *Cell*, 2006: 124:901-914.

Gross, AM *et al*. "Y-STR Concordance Study Between Y-Plex™5, Y-Plex™6, Y-Plex™12, Powerplex®Y, Yfiler™, MPI, and MPII." *J Forensic Sci*, 2006: 51:6:1423-1428.

Gusmão, L *et al*. "DNA Commission of the International Society of Forensic Genetics (ISFG): an update of the recommendations on the use of Y-STRs in forensic analysis." *Int J Legal Med*, 2006: 120:191-200.

Heyer, Evelyn *et al*. "Estimating Y chromosome specific microsatellite mutation frequencies using deep rooting pedigrees." *Human Molecular Genetics*, 1997: 6:5:799-803.

Institute of Legal Medicine, Charité - Universitätsmedizin Berlin. *YHRD STR Database.* 17 January 2007. http://www.yhrd.org/index.html (accessed March 29, 2007).

International Human Genome Sequencing Consortium. "Initial sequencing and analysis of the human genome." *Nature*, 2001: 409:860-921.

Jobling, M, and P Gill. "Encoded evidence: DNA in forensic analysis." *Nature reviews Genetics*, 2004: 5:739-752.

Karlsson, Andreas O *et al*. "Y-chromosome diversity in Sweden - A long-time perspective." *European Journal of Human Genetics*, 2006: 14:963-970.

Kayser, M *et al*. "Characteristics and Frequency of Germline Mutations at Microsatellite Loci from the Human Y Chromosome, as Revealed by Direct Observation in Father/Son Pairs." *Am J Hum Genet*, 2000: 66:1580-1588.

Klintschar, Michael, and Peter Wiegand. "Polymerase slippage in relation to the uniformity of tetrameric repeat stretches." *Forensic Science International*, 2003: 135:163-166.

Krawczak, M. "Forensic evaluation of Y-STR haplotype matches: a comment." *Forensic Science International*, 2001: 118:114-115.

Krenke, Benjamin E. *et al*. "Validation of a male-specific, 12-locus flourescent short tandem repeat (STR) multiplex." *Forensic Science International*, 2005: 151: 111-124.

Mulero, J *et al*. "Development and Validation of the AmpFℓSTR® Yfiler™ PCR amplification kit: A Male Specific, Single Amplification 17 Y-STR Multiplex System." *J Forensic Sci*, 2006: 51:1:64-75.

Rappold, G. "The pseudoautosomal regions of the human sex chromosomes." *Human Genetics*, 1993: 92:315-324.

Roewer, L *et al*. "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*, 2000: 114:31-43.

Roewer, L, and PJP Croucher *et al*. "Signature of recent historical events in the European Y-chromosomal STR haplotype distribution." *Hum Genet*, 2005: 116:279-291.

Ross, M *et al*. "The sequences of the human sex chromosomes." *Current Opinins in Genetics & Development*, 2006: 16:213-218.

Shinde, D *et al*. "Taq DNA polymerase slippage mutation rates measured by PCR and quasi-likelihood analysis: (CA/GT)n and (A/T)n microsatellites." *Nucleic Acids Research*, 2003: 31:3:974-980.

Skaletsky, H *et al*. "The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes." *Nature*, 2003: 423:825-837.

SWGDAM. "Revised Validation Guidelines." *Forensic Science Communications.* 10 07 2004. http://www.fbi.gov/hq/lab/fsc/backissu/july2004/standards/2004_03_standards02.htm (accessed 04 11, 2007).

Wilson, Ian G. "Inhibition and Facilitation of Nucleic Acid Amplification." *Applied and Environmental Microbiology*, 1997: 3741-3751.

Wilson, Keith, and John Walker. *Principles and Techniques of Biochemistry and Molecular Biology.* Sixth Edition. Hong Kong: Cambridge University Press, 2005.

Additional literature

Ansell, Ricky. *Tio år med DNA.* SKL report, Linköping: SKL, 2005:02.

Blom, Gunnar, and Björn Holmquist. *Statistikteori med tillämpningar.* Lund: Studentlitteratur, 1998.

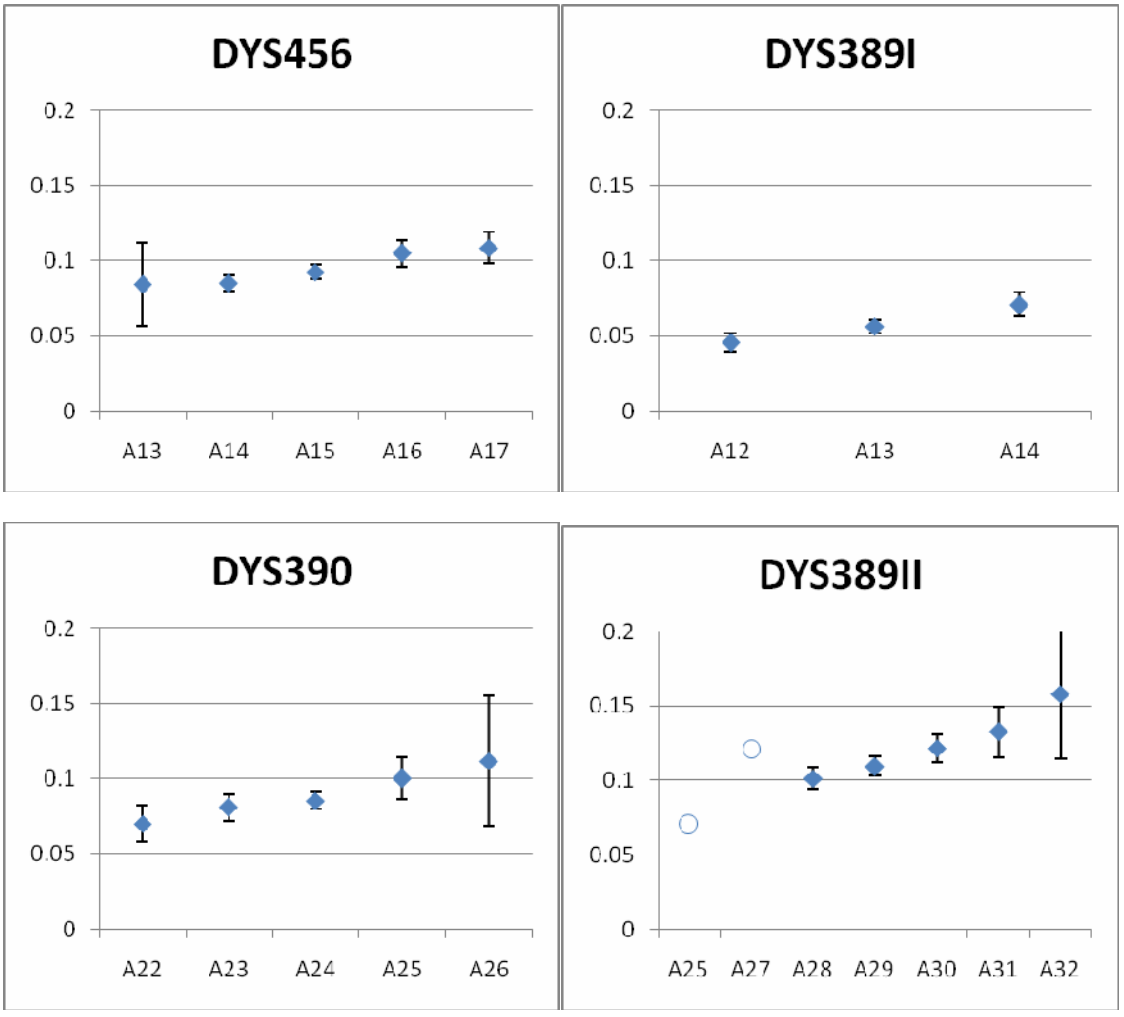
Butler, John M. *Forensic DNA Typing.* London: Elsevier Press, 2005.

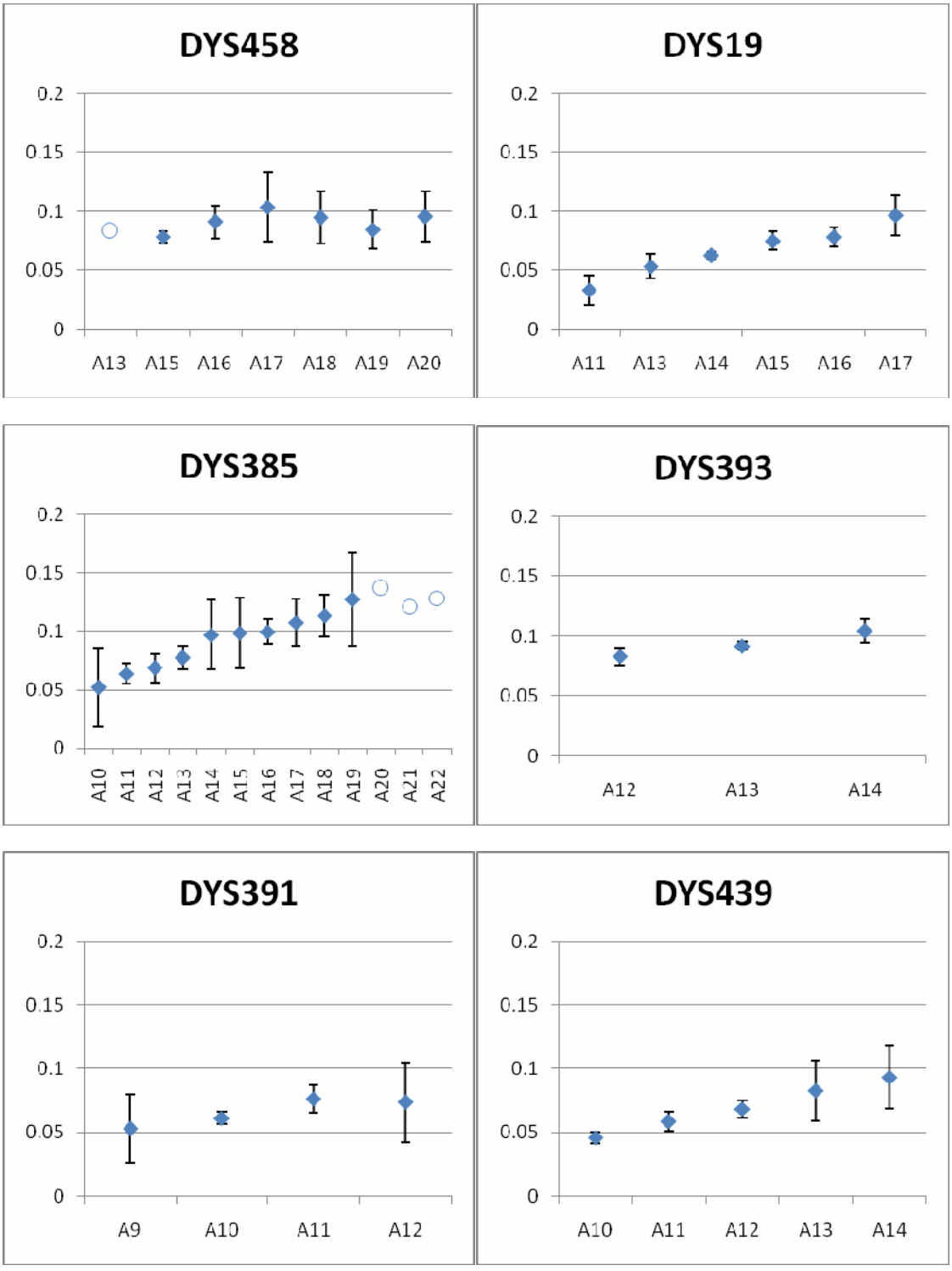
Jägsell, Malin. *Evaluation and adjustment of the Y-STR kit PowerPlex®Y to forensic casework - limitations and possibilities.* Masters thesis, Linköping: Linköping University, 2005.

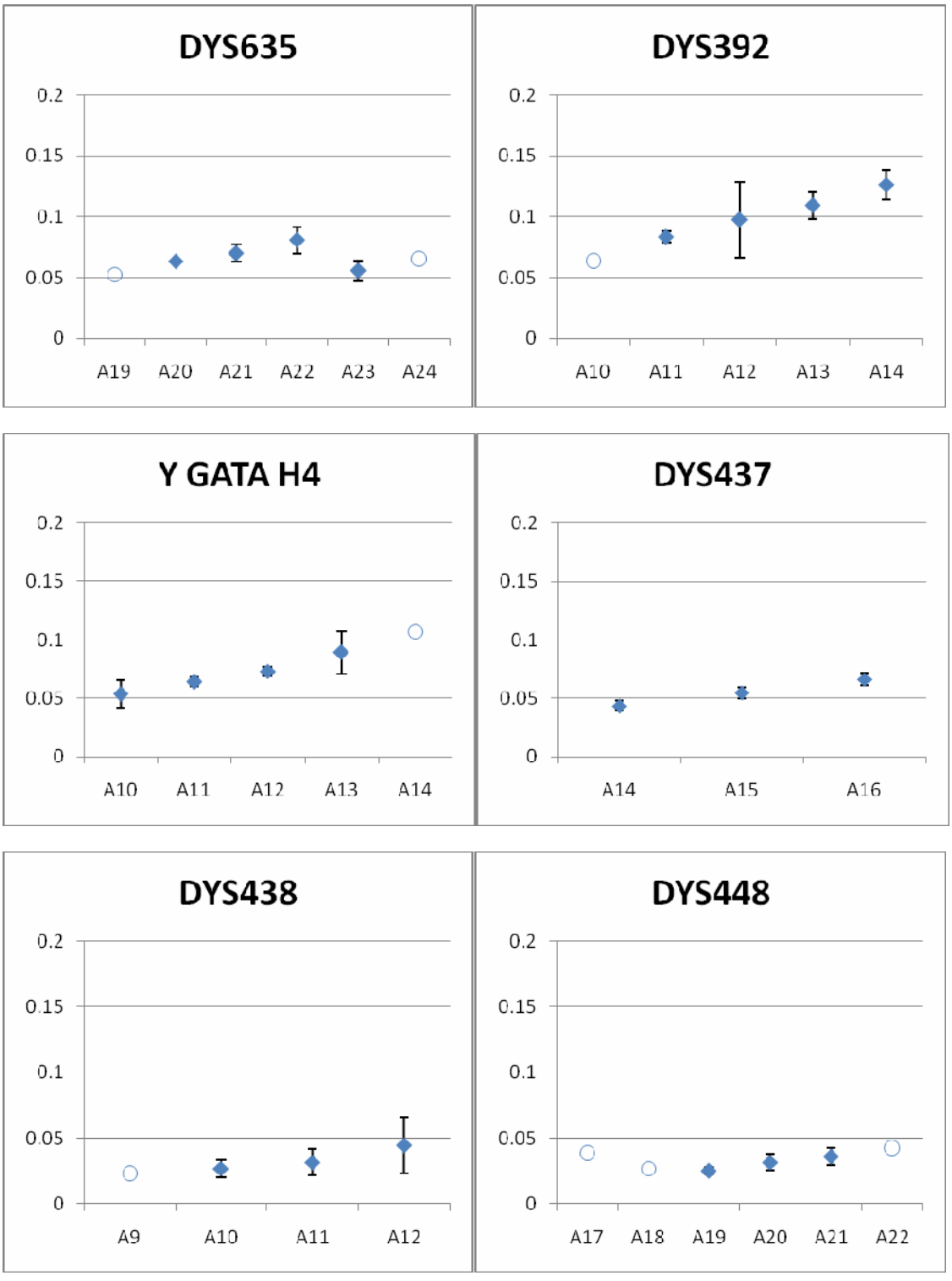
Wilson, Keith, and John Walker. *Principles and Techniques of Biochemistry and Molecular Biology.* Hong Kong: Cambridge University Press, 2005.

Appendix A

Stutter-1 ratios of the Yfiler STR loci in relation to repeat length. Error bars represent a confidential interval of significance*** ($\alpha = 0.001$). Single values are denoted by non-filled circles. The vertical axes of the figures represent stutter ratios (stutter peak height/allele peak height). The horizontal axes represent the length (n) of the allele peak (A). The allelic variants are thus denoted as An.




DYS458
0.2
0.15
0.1
0.05
0
A13 A15 A16 A17 A18 A19 A20
DYS19
0.2
0.15
0.1
0.05
0
A11 A13 A14 A15 A16 A17
DYS385
0.2
0.15
0.1
0.05
0
A10 A11 A12 A13 A14 A15 A16 A17 A18 A19 A20 A21 A22
DYS393
0.2
0.15
0.1
0.05
0
A12 A13 A14
DYS391
0.2
0.15
0.1
0.05
0
A9 A10 A11 A12
DYS439
0.2
0.15
0.1
0.05
0
A10 A11 A12 A13 A14


DYS635
0.2
0.15
0.1
0.05
0
A19
A20
A21
A22
A23
A24
DYS392
0.2
0.15
0.1
0.05
0
A10
A11
A12
A13
A14
Y GATA H4
0.2
0.15
0.1
0.05
0
A10
A11
A12
A13
A14
DYS437
0.2
0.15
0.1
0.05
0
A14
A15
A16
DYS438
0.2
0.15
0.1
0.05
0
A9
A10
A11
A12
DYS448
0.2
0.15
0.1
0.05
0
A17
A18
A19
A20
A21
A22