Optimization of the use of biological traces in forensic routine: DNA extraction from fluid identification tests

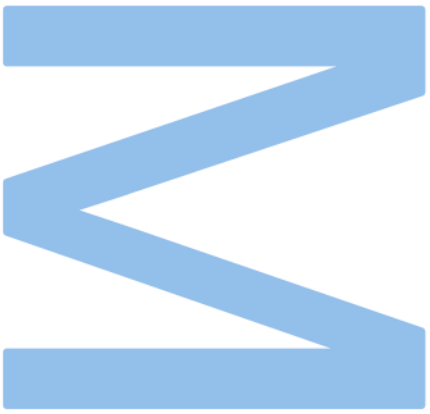
Mafalda Palhares Falcão Ferreira da Silva
Master in Forensic Genetics
Department of Biology
2022/2023

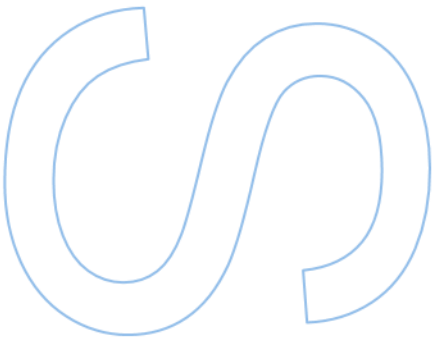
Supervisor
Ana Mosquera Miguel, Researcher, USC/INCIFOR

Co-supervisor
Iva Gomes, Associate Researcher, i3S/IPATIMUP

Maria de la Puente Vila, Researcher, USC/INCIFOR


U.PORTO
FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO





Sworn Statement

I, Mafalda Palhares Falcão Ferreira da Silva, enrolled in the Master Degree in Forensic Genetics at the Faculty of Sciences of the University of Porto hereby declare, in accordance with the provisions of paragraph a) of Article 14 of the Code of Ethical Conduct of the University of Porto, that the content of this dissertation reflects perspectives, research work and my own interpretations at the time of its submission.

By submitting this dissertation, I also declare that it contains the results of my own research work and contributions that have not been previously submitted to this or any other institution.

I further declare that all references to other authors fully comply with the rules of attribution and are referenced in the text by citation and identified in the bibliographic references section. This dissertation does not include any content whose reproduction is protected by copyright laws.

I am aware that the practice of plagiarism and self-plagiarism constitute a form of academic offense.

Mafalda Palhares Falcão Ferreira da Silva

29 de junho, 2023

Acknowledgements

Nearing the culmination of one of the most laborious phases in my academic training, I would like to thank the people without whom this work would not have been possible.

To the director of the Institute of Forensic Science Luís Concheiro, Professor Maria Victoria Lareu Huidobro, for accepting me in her laboratory.

To Professor António Amorim for being the intermediary for the realization of this dissertation at INCIFOR.

To Doctor Ana Mosquera, and Doctor Maria de La Puente from INCIFOR, my master thesis supervisor and co-supervisor respectively, and Doctor Iva Gomes from i3S/IPATIMUP, also my co-supervisor, for being as enthusiastic about this work as I was, for helping me in everything I needed, for accompanying me during the whole project, and for reading and re-reading this work until everything was well done.

To the entire team working in the forensic genetics' laboratory at INCIFOR for helping me and sharing knowledge that I will carry with me for life.

To my friends and family for supporting me when everything seemed to go wrong, for hugging me when the motivation faded, and for always believing in me and what I am capable of.

Resumo

Na rotina de um laboratório de Genética Forense, são efetuadas análises genéticas a todos os tecidos e/ou fluidos biológicos, tais como sangue, saliva, cabelo, sémen, etc. que foram, por exemplo, recuperados do local do crime, da vítima ou do suspeito. Quando chegam ao laboratório para análise, as provas que poderão conter fluidos corporais, são testadas para determinar a presença ou ausência de um determinado fluido. Estes testes são chamados testes presuntivos e são rápidos, fáceis de utilizar e interpretar e não devem consumir grandes quantidades de ADN. Após a determinação da presença ou ausência do fluido, as amostras seguem para a extração de ADN (quando confirmada a presença de material biológico). A extração de ADN é o processo através do qual as amostras são submetidas para separar quaisquer componentes do ADN, como, por exemplo, proteínas. Uma vez extraído, é enviado para a quantificação para determinar a concentração de ADN presente na amostra e garantir que é de origem humana. Segue-se a amplificação do ADN por Reação em Cadeia da Polimerase (PCR, do inglês, *Polymerase Chain Reaction*). O número de amostras recolhidas em locais de crime de abuso sexual é frequentemente limitado e a quantidade de ADN presente poderá ser escassa. Por este motivo, a PCR é utilizada para produzir milhões de cópias de ADN uma vez que se trata de uma técnica muito sensível e que, com apenas minutas quantidades de ADN, é possível obter perfis genéticos.

A agressão sexual é um crime prevalente em todo o mundo que pode afetar qualquer pessoa, independentemente da sua etnia, riqueza, idade e/ou sexo. É definida como qualquer tipo de contacto ou tentativa de contacto físico sexual sem consentimento. De acordo com a OMS, 30% das mulheres reportaram violência física e/ou sexual por parte do parceiro e 7% das mulheres em todo o mundo reportaram violência sexual por um indivíduo que não o parceiro. Em Portugal, 96,2% das vítimas adultas são mulheres. A recolha e análise de fluidos corporais neste crime é extremamente importante, uma vez que, muitas vezes, é a única forma de provar que o crime ocorreu e de identificar o seu autor. Sémen é um fluido corporal muito comum encontrado em casos de agressão sexual e é uma importante fonte de ADN para encontrar o agressor.

Neste trabalho analisámos os testes presuntivos utilizados em casos encerrados e recebidos pelo laboratório entre 2019 e 2022, e comparámos a sensibilidade entre o Seratec PSA Semiquant (Seratec, Germany) e o RSID-Sémen (Independent Forensics, USA). Tentámos, através da comparação de metodologias, estabelecer um novo

método (Método 2) para o processamento de zaragatoas vaginais que consiste na utilização de toda a zaragatoa. Na rotina forense deste laboratório, um corte da zaragatoa é utilizado para o RSID-Sémen e o restante é utilizado para a extração de ADN (Método 1). Incubámos a zaragatoa inteira com tampão RSID e extraímos ADN da mesma zaragatoa. Foram utilizadas amostras com vários intervalos de pós-coito (ICP) e diluições de sémen fresco e congelado. Todas as amostras utilizadas na comparação dos dois métodos foram processadas de acordo com ambos para permitir a comparação de resultados entre diferentes metodologias. As restantes amostras foram analisadas de acordo com o Método 1. As amostras foram extraídas utilizando a extração padrão e a extração diferencial com o sistema de extração de ADN forense *AutoMate Express* e o kit comercial de extração de ADN forense *PrepFiler Express* (*Applied Biosystems, Thermofisher Scientific*, USA). O ADN foi quantificado utilizando o kit de quantificação de ADN Quantifiler Trio e amplificado com o kit comercial GlobalFiler e o kit comercial *YFiler Plus* (*Applied Biosystems, Thermofisher Scientific*, USA).

De um total de 44 casos de agressão sexual recebidos pelo laboratório, em 30 casos tanto o Seratec PSA Semiquant como o RSID-Sémen foram positivos ou negativos, mostrando resultados concordantes. Em 11 casos, o Seratec PSA Semiquant foi positivo e o RSID-Sémen foi negativo, e em 3, observou-se o contrário. A comparação da sensibilidade do Seratec PSA Semiquant e do RSID-Sémen mostrou que o Seratec PSA Semiquant tem uma sensibilidade mais elevada. Os perfis genéticos autossómicos obtidos pelo Método 1 apresentaram mais alelos amplificados do que os obtidos pelo Método 2. Das três amostras amplificadas para cada método, duas apresentaram um perfil mais completo dos STRs autossómicos da componente masculina para o Método 1. A análise dos Y-STRs mostrou que, utilizando o Método 2, conseguimos amplificar mais alelos às 12h IPC e às 24h IPC. Os resultados deste estudo não são conclusivos, mas sugerem que o Método 2 pode ser mais vantajoso em relação ao Método 1 quando se tratam de amostras com baixa quantidade de material biológico e, consequentemente, com pouco conteúdo genético. No entanto, são necessários mais estudos para suportar os resultados preliminares do presente trabalho e melhorar o método de extração de ADN a partir do resto das amostras utilizadas nos testes presuntivos.

Palavras-chave: Genética forense, abuso sexual, testes presuntivos, identificação de sémen, lise diferencial

Abstract

In routine case-work of a forensic genetics' laboratory, genetic analysis is performed on every biological tissue and/or fluid such as blood, saliva, hair, semen, etc. that has been retrieved from, for example, a crime scene, a victim or a suspect. When evidence first arrives at the laboratory, it is tested to determine the presence or absence of a determined body fluid. These tests are called presumptive tests and are fast, easy to use and to interpret and do not consume large quantities of DNA. After determining the presence or absence of the fluid, samples proceed to DNA extraction. DNA extraction is the process under which samples are subjected to separate any components, such as proteins, from the DNA. Once it is extracted it is sent to DNA quantification to determine the concentration of DNA present in the sample and ensure that it is of human origin. DNA amplification by Polymerase Chain Reaction (PCR) is followed. The number of samples collected from crime scenes may be in some cases limited and the amount of DNA present scarce. For this reason, PCR is used to produce millions of copies of the representative DNA.

Sexual assault is a prevalent crime worldwide that can affect anyone regardless of their ethnicity, wealth, age and/or gender. It is defined as any type of physical sexual contact or attempt without consent. According to World Health Organization (WHO) 30% of women reported partner physical and/or sexual violence and 7% of women worldwide reported non-partner sexual violence. In Portugal, 96.2% of the adult victims are women. The collection and analysis of body fluids in this crime is extremely important since it is often the only way to prove that it happened and to find or convict the perpetrator. Semen is a very common body fluid found in sexual assault cases and is a great source of DNA to identify the assailant.

In this work we analysed the presumptive tests used in closed cases received by the laboratory, compared the sensitivity of Seratec PSA Semiquant (Seratec, Germany) and RSID-Semen (Independent Forensics, USA) and tried to establish a new method (Method 2) for processing vaginal swabs consisting in the use of the entire swab. In the forensic routine a cut from the swab is used for RSID-Semen and the rest is used for DNA extraction (Method 1). We incubated the entire swab with RSID Universal buffer and from the same swab extracted DNA. Post-coital interval (PCI) samples were used and semen dilutions were obtained from fresh and frozen semen samples. All samples used to test Method 2 were processed according to both methods to allow comparison of results. Samples were then extracted using standard extraction and differential

extraction with the AutoMate Express Forensic DNA Extraction System with PrepFiler *Express* Forensic DNA Extraction Comercial Kit (Applied Biosystems, Thermofisher Scientific, USA). DNA was quantified using Quantifiler Trio DNA Quantification kit and amplified with GlobalFiler Commercial kit and YFiler Plus commercial kit (Applied Biosystems, Thermofisher Scientific, USA).

In sexual assault cases received by the laboratory, out of 44 cases tested, Seratec PSA Semiquant and RSID-Semen were either positive or negative in 30 of those cases, showing concordant results. In 11 cases, Seratec PSA Semiquant was positive and RSID-Semen was negative, and in three the opposite was observed. Comparing the sensitivity of Seratec PSA Semiquant and RSID-Semen results showed that Seratec PSA Semiquant has higher sensitivity. Autosomal genetic profiles obtained for Method 1 had more amplified alleles than for Method 2. Out of three samples amplified for each method, two showed a more complete profile for Method 1 when analysing autosomal STRs. Analysis of Y-STRs showed that using Method 2 we were able to amplify more alleles at 12h PCI and at 24h PCI. The preliminary results of this study are inconclusive but suggest that Method 2 may be advantageous over Method 1 when dealing with very small samples with little genetic content. However, more data is needed to prove and improve the method of extracting DNA from the remaining material of the presumptive tests.

Keywords: Forensic genetics, sexual assault, presumptive testing, semen identification, differential lysis

Table of Contents

List of Tables

List of Charts

List of Figures

List of Abbreviations

ALS	Alternate Light Source
APAV	Associação Portuguesa de Apoio à Vítima
bp	Base pairs
DI	Degradation Index
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
LA	Large Autosomal Target
m-RNA	Messenger RNA
mi-RNA	Micro RNA
NIR	Near Infra-red
NFW	Nuclease Free Water
PCI	Post-Coital Interval
PCR	Polymerase Chain Reaction
PSA	Prostate Specific Antigen
qPCR	Quantitative Polymerase Chain Reaction
RFU	Relative Fluorescence Unit
RNA	Ribonucleic Acid
rpm	Rotations per minute
RFLP	Restriction Fragment Length Polymorphism
RSID	Rapid Stain Identification Test
SA	Small Autosomal Target
SF	Sperm Fraction
Sg	Semelogenin
STR	Short Tandem Repeat
VNTRs	Variable Number of Tandem Repeats
WHO	World Health Organization
Y-STR	Y-Short Tandem Repeat

1.Introduction

1.1. Forensic Genetics

Over the years forensic genetics has revolutionized forensic science. In 1985, Alec Jeffreys, an English geneticist, found repeated sequences in the genome where the number of repeats might vary from person to person (Jeffreys et al., 1985). Analysing the number of repeats allowed the identification and discrimination of individuals (Butler, 2005). These DNA repeats are now known as variable number of tandem repeats (VNTRs) and range in size from 6 to 100 bp. Jeffreys and colleagues used a technique called restriction fragment length polymorphism (RFLP) to analyse these regions which involved the use of a restriction enzyme that cut the DNA surrounding VNTRs. RFLP, however, required large quantities of DNA and would not work in degraded samples (Butler, 2005). Short tandem repeats (STRs) are short DNA markers that were introduced in the mid-1990s and are currently the most used in the forensic routine due to their small size (1 to 6 bp), robustness and high discrimination power (Goodwin et al., 2007). In 1985 polymerase chain reaction (PCR) was developed by Kary Mullis to amplify specific regions of the DNA (Mullis et al., 1986). With this technique we can generate DNA profiles from degraded samples with just a few cells (Goodwin et al., 2007). Nowadays, PCR is a technique used worldwide allowing the analysis of very small samples with sparse genetic content. DNA can be collected from any biologically forensic relevant material such as blood, semen, hair, bones, skin and saliva and allow “DNA fingerprinting” – the unique identification of an individual through its genome.

1.2. Workflow in a Forensic Genetic laboratory

Nowadays the analysis of body fluids left at crime scenes is highly important because they can incriminate or exonerate an individual (Virkler & Lednev, 2009a). Fluid identification (i.e., the identification of the DNA source) might be helpful in determining where the crime occurred and give useful information about what happened and the relevance of the evidence collected (de Beijer et al., 2018). Even though different types of body fluids might be present, the most common are blood, saliva and semen (Virkler & Lednev, 2009a). All nucleated cells have DNA so this molecule is present in all biological fluids (Butler, 2005). DNA is the blueprint that every individual inherits from the mother and the father. Apart from identical twins, every individual has a different DNA profile due to differences in the DNA sequence and/or length polymorphisms.

The evidence collected is analysed and the DNA obtained from biological material is then directly compared to a potential suspect or run on a database (Butler, 2010). In practice, at least 15 cells are needed to generate a DNA profile. In forensic samples, it is usually required more to produce a profile due to sample degradation (Goodwin et al., 2007). From sample collection to DNA profiling, identification of the fluid type, extraction, quantification, amplification and capillary electrophoresis are performed as the common workflow in a routine case-work forensic laboratory, as represented in *Figure 1*.

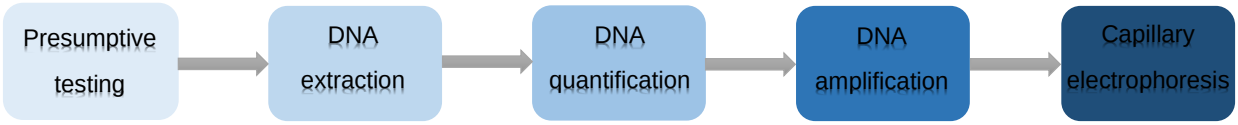


Figure 1 - Workflow in a Forensic Genetic laboratory.

1.2.1. Presumptive and confirmatory tests

Once the samples arrive at the laboratory, they are usually tested to evaluate the presence of any biological material. This is done using presumptive and confirmatory tests (Butler, 2012). Presumptive tests should be easy to use, safe and simple to read so the results are easy to interpret (Butler, 2012, 2010; Li, 2021). They should not damage the DNA present in the sample nor consume large quantities (Butler, 2012, 2010). These tests only indicate the possibility of a certain fluid being present and, therefore, do not rule out the need for a confirmatory test. A negative result indicates the absence of the fluid and a positive result indicates the possibility of the fluid being present in the sample (Li, 2021). Confirmatory assays are, on the other hand, more specific to the body fluid and can identify it with higher certainty (Li, 2021).

1.2.2. DNA extraction

DNA extraction is the process through which these molecules and others, as well as, PCR inhibitors are separated (Bogas, Vanessa, 2013; Butler, 2012, 2010). This is a crucial step because there are proteins that coat the DNA and protect them, preventing its analysis. The goal is to extract enough quantity, quality and purity to obtain a good profile for further analysis (Goodwin et al., 2007). In Forensic Genetic laboratories this is the process most susceptible to contamination, which is defined by the introduction of exogenous DNA in a sample (Li, 2021). To avoid contamination, the evidence and the reference samples should be processed in different rooms/areas and, if not possible, the evidence should be analysed prior to the reference sample (Li, 2021). Automated

extraction processes allow hands free extraction decreasing the risk of contamination (Liu et al., 2012).

The first step in the extraction process is the cellular lysis followed by the removal of the components that interfere with the extraction and DNA purification (Bogas, Vanessa, 2013; Butler, 2012; Liu et al., 2012). There are different methods and the choice of which one to perform depends on several factors such as the quantity and type of material present (Goodwin et al., 2007). For a long time, extraction with Phenol-Chloroform, also called organic extraction, was the most used (Butler, 2012). It uses sodium dodecylsufate and proteinase k to lyse the cell membranes and break down the proteins and a mixture of phenol and chloroform to separate the proteins from the DNA (Butler, 2012). Even though it extracts DNA successfully, this method uses toxic chemicals and multiple tubes which increases the risk of contamination and is time consuming (Liu et al., 2012). Another commonly used method is the Chelex method (Liu et al., 2012). This does not involve high costs, it is quick to execute and only uses one tube, decreasing the risk of contamination (Liu et al., 2012). A 5% Chelex suspension is usually added to the sample which is then exposed to high temperatures to denature DNA and break down the cell's membranes. This is followed by a centrifuge step, the supernatant is collected and is ready to be used for PCR amplification (Butler, 2012).

In this work, we used the Automate Express Forensic DNA Extraction System with the PrepFiler Forensic DNA Extraction kit (Applied Biosystems, Thermo Fisher Scientific) as illustrated in *Figure 2*.

Figure 2 - AutoMate Express Forensic DNA Extraction System. https://www.thermofisher.com.

The kit enables the automated extraction of up to 13 samples even with very low amounts of biological material such as semen or blood. It is very efficient removing PCR inhibitors and the final DNA concentration is high enough for further analysis requiring little extract volume (Liu et al., 2012). This technique allows the extraction to be done in a single tube by placing and removing different solutions from it (Butler, 2012). It uses silica-coated magnetic particles that reversibly adhere to DNA by lowering the pH of the medium (<7.5) (Cunha, Mariana, 2019). In the presence of a magnetic support the magnetic particle-DNA complex is isolated from the remaining cellular components and inhibitors (Bogas, Vanessa, 2013; Butler, 2012; Cunha, Mariana, 2019). For DNA purification successive washes are performed to eliminate all inhibitors (Cunha, Mariana, 2019). Extraction using magnetic particles allows the process to be fast, simple and automated, decreasing the risk of sample contamination (Butler, 2012).

1.2.2.1. Differential DNA extraction

First described in 1985 by Gill and colleagues (Gill et al., 1985), differential DNA extraction is a useful method in the analysis of sexual assault cases. In these cases, there is often an excess amount of female DNA resulting in an unfavourable male/female DNA ratio. As a result, the complete loss or drop-out of some male alleles is likely to happen when analysing autosomal STRs (Jankova et al., 2019). Despite being considered a golden method for the analysis of sexual assault samples, it is not applicable in samples from azoospermic or vasectomized men since they do not have spermatozoa (Jankova et al., 2019).

Differential extraction allows the separation of epithelial and sperm cells making it easier to interpret mixed DNA profiles and increasing the chances of obtaining the perpetrator's profile since in these situations the male profile is of greatest forensic value (Li, 2021; McKiernan & Danielson, 2017). The separation happens based on differential resistance of spermatozoa and epithelial cells to chemicals (Gill et al., 1985; Li, 2021). Firstly, the epithelial cells are lysed with proteolytic degradation using proteinase K and SDS. The sperm cells contain proteins cross-linked by disulphide bonds which are resistant to these chemicals (Goodwin et al., 2007). Once the epithelial cells have been lysed, the sample is centrifuged to pellet intact sperm cells. In the supernatant, what we call the first fraction, are the male and female epithelial cells (McKiernan & Danielson, 2017). To successfully lyse sperm cells it is necessary to add dithiothreitol (DTT), a reducing agent, to cleave the disulphide bridges, proteinase K and SDS (Butler, 2012; Li, 2021; McKiernan & Danielson, 2017). Despite being effective it is labour intensive

and time consuming. In addition, in highly degraded samples, the sperm fraction might be already lysed and therefore not separated from the epithelial cells (Li, 2021).

1.2.3. DNA quantification

Following DNA extraction is the DNA quantification procedure which is performed to determine the amount of DNA present in the extracted sample (Butler, 2010). The goals are to determine the amount of amplifiable DNA and ensure that its origin is human (Butler, 2012). The quantification of DNA is of utmost importance since if it is present in too high or too low quantities the amplification results will not be interpretable (Butler, 2012, 2010). In most of the available STR kits the optimal range of input human DNA is 0.5 to 2.0 ng (Butler, 2010). The amplification of too much DNA will likely result in "overblown" electropherograms and too little will result in loss of alleles. A sample with excessive amounts of DNA should be diluted and a sample with not enough DNA should be concentrated, re-extracted or sent directly to PCR amplification using low copy number approaches (Butler, 2012). The closer the DNA concentration is to the ideal range, the more likely a complete and clear DNA profile will be obtained.

Real-time quantitative PCR (qPCR) is the most common quantification approach in forensic genetics routine case-work. This method was first described by Higuchi and his co-workers at the Cetus Corporation in the early 1990s and allows to monitor the generation of PCR products as they are generated (Butler, 2012). Most commonly it is done using the TaqMan assay (Applied Biosystems, Thermo Fisher Scientific). This method uses two primers and a probe located within the region of interest defined by the forward and reverse primers (Goodwin et al., 2007; McKiernan & Danielson, 2017). On the 5' end of the probe there is a reporter and on the 3' end there is a fluorescence quencher. As the DNA polymerase forms new copies of the DNA it cleaves the TaqMan probe separating the reporter and the quencher. Once separated, the quencher is no longer in close proximity to the reporter and it fluoresces (Goodwin et al., 2007). The fluorescence signal is proportional to the amount of DNA copies generated during each PCR signal (McKiernan & Danielson, 2017).

In this work we used Quantifiler Trio DNA Quantification Kit (Applied Biosystems, Thermo Fisher Scientific) which uses the TaqMan based-assay real-time quantitative PCR (Holt et al., 2016). This kit can detect very low amounts of DNA, can distinguish between total human DNA and male human DNA and can determine the level of degradation in a sample. It detects three different targets, two in the autosomes and one in the Y chromosome (Gouveia, Nair, 2015). The autosomal targets are small autosomal

(SA) and large autosomal (LA) with amplicon sizes of 80 bp and 214 bp, respectively (Gouveia, Nair, 2015; Holt et al., 2016). The Y-chromosome target has an amplicon size of 75 bp (Holt et al., 2016). The quantification of different sized targets allows conclusions to be drawn about sample degradation (Holt et al., 2016). LA target is more prone do degradation due to its bigger size and is, therefore, used to determine the degradation level. Additionally, the degradation index (DI) can be calculated using the SA/LA ratio. The higher the degradation level, the lower the LA concentration and consequently the higher the DI (Gouveia, Nair, 2015; Holt et al., 2016). The kit has also an internal PCR control (IPC) which confirms the correct functioning of the reaction and also helps assessing problems with PCR inhibition (Holt et al., 2016).

1.2.4. DNA amplification by PCR

PCR was first described in 1985 by Kary Mullis and members of the Human Genetics group at the Cetus Corporation (now Roche Molecular Systems) (Mullis et al., 1986). In 1993, Kary Mullis received the Nobel Prize in Chemistry for developing PCR, highlighting the importance of his discovery (Butler, 2012). The goal is to make hundreds of millions of copies of a specific DNA sequence to allow the analysis of forensic samples that could otherwise be impossible to analyse (Butler, 2012). From one DNA molecule two are produced, and then four, then eight and so forth (Joshi & Deshpande, 2011). This process allows the amplification of challenging casework samples, uses human-specific primers preventing the amplification of non-human DNA and is a very sensitive and fast process (Butler, 2012). On the other hand, if PCR inhibitors are present the target DNA might not be amplified, is very susceptible to contamination and changes in the primer-binding region can prevent amplification and produce misleading results (Butler, 2012).

The process of DNA amplification by PCR involves heating and cooling of the samples in a determined number of cycles that are divided into three steps as represented in *Figure 3*. The first is the denaturation process in which the hydrogen bonds connecting both DNA strands break and the strands are separated This process occurs at high temperatures at around 94ºC - 95ºC (Li, 2021). The next step is the annealing of the primers to the DNA template promoted by the decrease in the temperature usually to 50ºC - 60ºC (Joshi & Deshpande, 2011). The last step is the extension or DNA replication which happens at 72ºC with the help of a DNA polymerase (Butler, 2012). The enzyme chosen has to be thermally stable and the most often used is *Taq* polymerase. *Taq* polymerase was isolated from a bacterium called *Thermus*

aquaticus that inhabits hot springs (Butler, 2012). To successfully generate millions or even billions of DNA copies around 30 cycles are needed (Joshi & Deshpande, 2011).

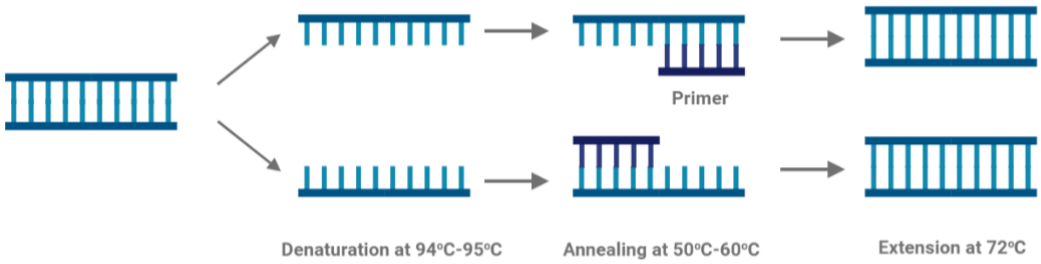


Figure 3 - Thermal cycling temperature parameters for DNA amplification with PCR. First is the denaturation step in which the DNA strands are broken. At 50ºC-60ºC the primers anneal to the single stranded DNA and at 72ºC Taq polymerase extends both primers. The process is repeated around 30 times. Created with BioRender.com.

The regions of interest to be amplified in a forensic laboratory are usually STRs – the gold standards of genetic markers for human identification. These are DNA regions with two to seven nucleotides that are easily amplified by PCR (Butler, 2012). STRs are the preferred polymorphisms due to their high variability among individuals allowing individual discrimination (except for identical twins) (Butler, 2012). In sexual assault cases there is usually an excess amount of female DNA that will supress the amplification of the male component. It is, therefore, advisable to complement the analysis with Y-STR haplotyping (Purps et al., 2015). It is also beneficial to detect the number of male contributors, for example, when a women is raped by more than one male and in situations where no spermatozoa is detected (Purps et al., 2015). The last might be due to infertility problems such as azoospermia, lack of ejaculation or the male has been vasectomised (Purps et al., 2015).

In this work we used GlobalFiler and YFiler Plus kits (Applied Biosystems, Thermo Fisher Scientific). GlobalFiler kit is a single multiplex assay that amplifies a set of 24 markers (Ludeman et al., 2018). Out of the 24 markers, 13 STRs are the original CODIS loci with 7 non-overlapping loci from the expanded European Standard Set (ESS), amelogenin which is the sex determining marker, two Y-based loci and the SE33 locus which has a higher power of discrimination than any other STR studied for human identification (Ludeman et al., 2018). As well as being a highly effective kit for amplifying challenging samples, it also has the advantage of completing amplification in approximately 80 minutes. Using this kit it is possible to generate DNA profiles even in

the presence of PCR inhibitors (Ludeman et al., 2018). YFiler Plus kit amplifies a set of 27 markers which are 17 loci previously used in the YFiler kit and 10 additional loci. These new loci allow the discrimination of related males due to their high mutation rate (Gopinath et al., 2016). Out of the ten new loci, three are highly polymorphic loci and seven are rapidly mutating loci (Gopinath et al., 2016).

1.2.5. Capillary electrophoresis

After DNA amplification, PCR products are separated according to their size using capillary electrophoresis. DNA is a negatively charged molecule due to the phosphate groups on the nucleotides so the molecule will migrate in an electric field. As it migrates from the cathode to the anode it passes the polymer sieving medium allowing the separation of different sized STRs (McKiernan & Danielson, 2017). The smaller the DNA molecules are the faster they pass through the capillary and the quicker they are detected (Butler, 2005). During the amplification process, fluorescent dyes were incorporated to the 5' end of one primer for each primer set per locus. As the STR products pass through the capillary, a laser excites the primer dyes which emits a fluorescence signal that is detected, measured and an electropherogram is produced, all done by an appropriate software (McKiernan & Danielson, 2017). An internal size standard is added to each sample containing DNA fragments of known lengths labelled with a different fluorochrome to allow allele designation (Goodwin et al., 2007). An allelic ladder is also added, an artificial mixture of the common alleles present in the human population for a particular STR marker, used for accurate genotype determination (Butler, 2012).

1.3. Sexual assault

Sexual assault is a worldwide problem that can affect anyone regardless of wealth, age, ethnicity, sexual orientation, etc., but is more prominent among women (Du Mont et al., 2007). It refers to any type of physical sexual contact or attempt without consent (DeVore & Sachs, 2011; Dworkin et al., 2021). It includes anything with or without vaginal and/or anal penetration and physical force or psychological coercion (Mollen et al., 2012). There are two forms of sexual violence which are sexual violence committed by an intimate partner or by a non-partner (World Health Organization, 2013). The first is defined as being physically forced to have intercourse when the person does not want to or is afraid of saying no to the partner and/or being forced to do something sexually that the person finds humiliating or degrading (World Health Organization, 2013). Non-partner sexual violence is any kind of sexual or attempted sexual violence

committed by someone other than a partner (World Health Organization, 2013). According to the World Health Organization (WHO) (World Health Organization, 2013), 35% of women worldwide have experienced one or both types of sexual violence (World Health Organization, 2013). The same study estimated that 7% of women worldwide reported non-partner sexual violence and 30% reported partner physical and/or sexual violence (World Health Organization, 2013). Unfortunately, there are groups of people that are more vulnerable such as adolescents, disabled people, sex workers, survivors of childhood sexual or physical abuse, etc (Luce et al., 2010). Young women from the age of 16 to 24 are four times more likely to be sexually assaulted than women of any other age (Luce et al., 2010). Sexual assault also victimizes children and consists of any activity with a child before the age of consent for the sexual gratification of older people (Sakelliadis et al., 2009). It includes the use of a child in prostitution or other unlawful sexual practices, use of children in pornographic performances, etc (Sakelliadis et al., 2009). In 2006, the WHO estimated that 1950 million girls and 73 million boys were victims of child sexual abuse worldwide (Murray et al., 2014).

In Portugal, according to APAV (Associação Portuguesa de Apoio à Vítima) (APAV), between 2013 and 2018 the number of victims of sexual assault per year increased. As expected, the majority of victims are female, 77.4% girls and 96.2% female adults.

There are several consequences that a sexual assault survivor might suffer short- and long-term. These consequences include sexually transmitted infections, unwanted pregnancy, sexual dysfunction, pelvic pain, depression, anxiety, post-traumatic stress disorder, suicidal ideation and attempts, among others (Du Mont et al., 2007; Luce et al., 2010).

1.4. Semen fluid identification

To successfully prove the sexual assault and aid in the identification of a suspect, evidence must be collected. An example of evidence is the presence of body fluids deposited on the body and/or clothing of the victim or at the crime scene itself. A very common fluid deposited in these circumstances is semen. Semen is the liquid ejaculated after sexual stimulation composed of secretions from various accessory glands and spermatozoa (Bell, 2019). A healthy man releases about 2-6 mL of semen per ejaculation containing on average 100 to 150 million sperm cells/mL (Bell, 2019, 2021). Exceptions to this definition are oligospermic men, men with abnormal low sperm count,

azoospermic men, no production of spermatozoa at all or men subjected to a vasectomy (surgical removal of a bilateral segment of the ductus deferens) (2021; Li, 2021).

Sperm cells have three morphological different structures. The head is where the nucleus is and is connected to the tail by the middle piece as represented in *Figure 4* (Bell, 2019). In the middle piece are the mitochondria responsible for providing energy for the tail motility (Li, 2021). In a healthy man about 60% of the spermatozoa have a normal morphology (Li, 2021).

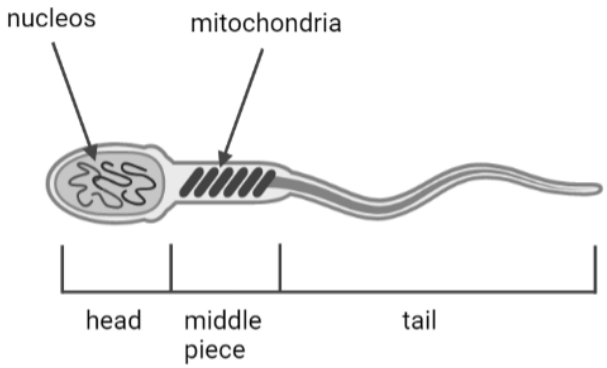


Figure 4 - Human sperm cell structure. Created with BioRender.com.

The seminal vesicles produce approximately 60% of the seminal fluid and 30% correspond to secretions of the prostatic fluid (Li, 2021; Lilja et al., 1989). The latter contains high concentrations of prostate-specific antigen (PSA) and acid phosphatase (AP) which are biological markers used to assess the presence of semen (Li, 2021). First ejaculated semen has a gel consistency provided by proteins semenogelin I (SgI), semenogelin II (SgII) and fibronectin together (2021; Laffan et al., 2011; Old et al., 2012). This is produced in the seminal vesicles and secreted into de seminal fluid. PSA is responsible for the degradation of Sg within 15 minutes of ejaculation (Laffan et al., 2011; Li, 2021; Lilja et al., 1989). Detecting the presence of this fluid is crucial in determining the existence of sexual activity. To detect it, several methods of orientation and confirmation are described (Virkler & Lednev, 2009a).

1.4.1. Semen detection methods

1.4.1.1. Alternate Light Source

The simplest way to detect body fluids invisible to the naked eye is to use an alternate light source (ALS). This is a presumptive test that uses light of different wavelengths, such as ultraviolet (Sijen & Harbison, 2021). Different body fluids fluoresce

at different wavelengths (Sijen & Harbison, 2021). Semen is best detected using wavelengths between 420 to 450 nm allowing the visualization of fluorescence with orange goggles (Nelson, 2002). Urine and saliva samples might fluoresce with lower intensity meaning the test sensibility is not high (Li, 2021). A device used for semen detection is Wood's Lamp which emits wavelengths between 320 to 400 nm (J.-H. An et al., 2012). Another example is Bluemaxx BM500 (Sirchie, USA) whose wavelengths range from 390 to 500 nm. The latter has higher sensibility but semen can still be misidentified as urine in the eyes of an untrained person (Nelson, 2002).

1.4.1.2. Acid Phosphatase

The detection of AP is a presumptive test and one of the most commonly used methods for semen detection. The enzyme is secreted by the prostatic gland and is present in concentrations up to 400 times higher in semen than in other body fluids (Butler, 2010). This is a catalytic test based on the enzymatic activity. Briefly, AP catalysis the hydrolysis of phosphatases. The product reacts with the reagent causing colour change (Li, 2021). For example, the Brentamine test promotes a colour change from orange to purple, *Figure 5*, using the combination of α-naphthyl phosphate and Brentamine Fast Blue and substrate/reagent, respectively (Virkler & Lednev, 2009a). In the reaction α-naphthyl phosphate is hydrolysed to phosphate and α-naphthol by AP. The last product reacts with Brentamine Fast Blue promoting the colour change within 30 seconds to 2 minutes, depending on the protocol used (2021; Li, 2021, p. 263). Even though it is an easy and non-laborious test, it is destructive to the sample, prone to false positives and the enzyme can be degraded when exposed to heat, mould, putrefaction or chemicals (Virkler & Lednev, 2009a, 2009b).

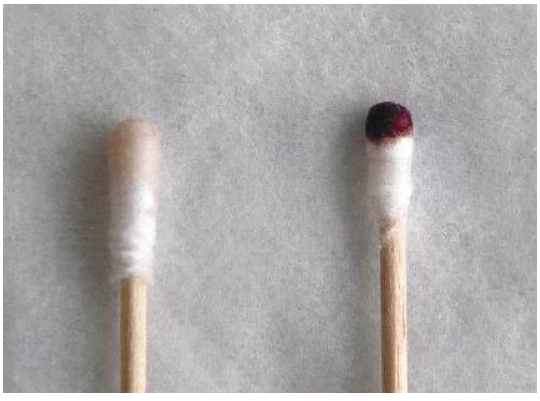


Figure 5 - Result of a positive test for semen using Bretamine Fast Blue (Li, 2021) The colour change from light orange to dark purple indicates the presence of AP and consequently semen.

1.4.1.3. Microscopic examination of spermatozoa

A confirmatory method for semen detection is the microscopically determination of the presence of sperm cells, considering that semen is the only body fluid that has these cells (Laffan et al., 2011; Virkler & Lednev, 2009a). To better observe the cells, if present, the sample should be stained. The most common staining technique is the *Christmas tree staining* (Li, 2021; Virkler & Lednev, 2009a). This method uses Nuclear Fast Red to stain the head red and picroindigocarmine to stain the tail green as represented in *Figure 6* (Li, 2021). However, this method is not useful when analysing semen from oligospermic, aspermic, or vasectomized men (Li, 2021). In addition, in cases where semen adheres to clothing it may be difficult to observe intact sperm cells, and in situations where the sample is washed or comes into contact with other objects the semen may disintegrate (Laffan et al., 2011).

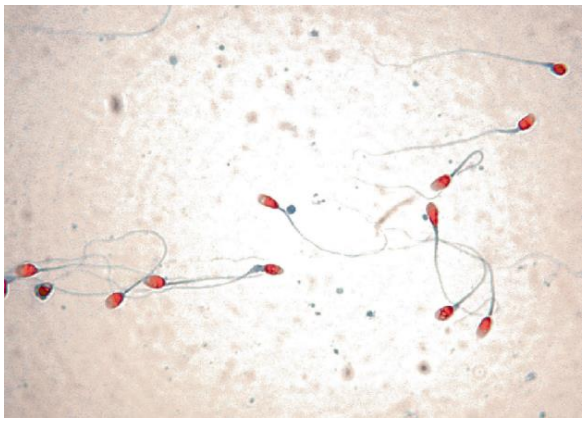


Figure 6 - Example of human sperm cells by Christmas Tree stain procedure (Bell, 2019).

1.4.1.4. Raman spectroscopy

Raman spectroscopy is a confirmatory and non-destructive method that requires sample quantities as small as picograms or phentoliters (Li, 2021; Virkler & Lednev, 2009a, 2009b). It is based on the inelastic scattering of light from the analysed sample. The light source used is a near-infrared excitation light (NIR). To perform this test no chemical reagents are used and there is no consumption or destruction of the sample, allowing subsequent DNA analysis (Virkler & Lednev, 2009b). It does not react significantly with water and is therefore ideal for the analysis of body fluids (Virkler & Lednev, 2009b). The Raman spectrum shows several peaks, and the combination of those are characteristic for each type of body fluid. Thus, it is possible to relate a fluid type to a single spectrum, as shown in *Figure 7*. The semen spectrum contains three

main components, tyrosine, choline and spermine phosphate hexahydrate (Virkler & Lednev, 2009b).

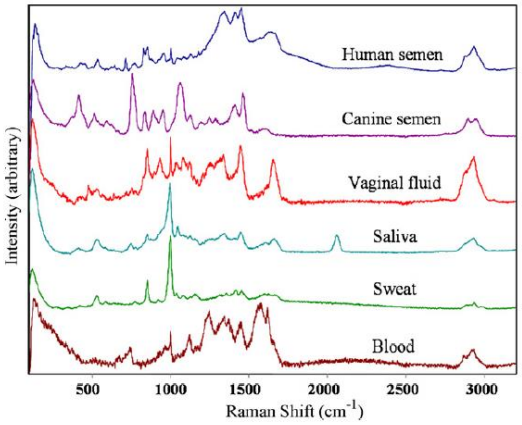


Figure 7 - Raman spectrum for human semen, canine semen, vaginal fluid, saliva, sweat and blood (Virkler & Lednev, 2009a).

1.4.1.5. Messenger RNA

Conventional body fluid identification methods have limitations such as lack of specificity and sensitivity, can be destructive to samples and use large quantities, and do not allow the identification of multiple fluids at the same time. The application of RNA for body fluid identification allows to overcome these limitations (Lynch & Fleming, 2021). The use of messenger RNA (mRNA) is a confirmatory technique that allows the identification of body fluids based on the detection of specific genes expressed only in certain cells (Li, 2021). When compared to the conventional methods, mRNA offers higher sensitivity and specificity, allows analysis of multiple body fluids at the same time and can be co-extracted with DNA (Lynch & Fleming, 2021).

These specific genes are candidate genes for the identification of body fluids and tissues in forensic context (Li, 2021). For internal control, housekeeping genes are used, i.e., constitutively expressed genes (expressed in all cell types) (Li, 2021). *GAPDH*, *G6PD* and *18S* RNA are examples of housekeeping genes commonly used in forensics (Harbison & Fleming, 2016).

The most commonly used process for determining body fluid/tissue using mRNA involves sample collection, RNA extraction, complementary DNA synthesis, touch-down PCR and capillary electrophoresis to separate the specific amplicons based on size, and finally analysis of results (Lynch & Fleming, 2021). For semen identification five

candidate genes were used in the study of Tian H. *et al*, KLK3, TGM4, SEMG1, PRM1 and PRM2, that aimed to explore the effect of semen from infertile men on body fluid/tissue identification (Li, 2021; Tian et al., 2021). This work also aimed to identify semen-specific mRNAs that can efficiently and accurately distinguish normal and infertile semen samples from other body fluids (Tian et al., 2021).

The marker KLK3 belongs to the kallikrein-associated peptidase family. It is specific for the prostate and produced by the prostate epithelium (Tian et al., 2021). TGM4 is a member of the transglutaminases family and is known as the human prostate-specific transglutaminase gene (Tian et al., 2021). SEMG1 encodes the major component of seminal fluid. This marker binds to the spermatozoa to form the gelatinous clot after ejaculation. This binding depends on the amount of spermatozoa present in the semen, so in infertile men SEMG1 can be detected in low amounts (Tian et al., 2021). Finally, PRM1 and PRM2 are protamines whose function lies in the condensation and compaction of the sperm nuclear genome in the sperm head. Abnormal expression of this marker may be indicative of infertility (Tian et al., 2021). In the study of Tian H. et al (Tian et al., 2021), KLK3 was the most optimal mRNA biomarker in semen identification.

This is just an example of some of the most robust and useful mRNA markers for semen identification. However, over the past years, many studies have proposed mRNA markers for other body fluid and tissue identification that are relevant in a sexual assault context (such as vaginal secretions and saliva) and therefore a range of these types of markers are available for selection together with mRNA semen specific markers (Albani & Fleming, 2018; Hanson et al., 2018; Jakubowska et al., 2013)

A disadvantage when working with RNA is the more unstable nature when compared to DNA due to ribonuclease degradation (Lynch & Fleming, 2021) and the lack of a standard, straightforward applicable method for the interpretation of results in forensic body fluid/tissue identification cases (Lindenbergh et al., 2013).

1.4.1.6. MicroRNA

MicroRNA (miRNA) are small non-coding RNA molecules that regulate gene expression on a post-transcriptional level silencing mRNA (J.-H. An et al., 2012; Bexon KJ, 2017; Li, 2021). Their small size makes them less susceptible to/prone to degradation (compared to mRNA), they exist in high amounts in the human cells and can be co-extracted with DNA (J.-H. An et al., 2012). These characteristics are highly advantageous when the amount of sample is very low or it is very degraded (Bexon KJ, 2017; Lynch & Fleming, 2021). A single miRNA may have multiple RNA targets which

means there expression is not tissue-specific but the patterns in which they are expressed are specific to certain body fluids (Harbison & Fleming, 2016; Li, 2021). For semen identification there are a few miRNAs described in the literature. These include miR-10a, 10b, 135a, 135b, 507, 888, 891a e 943 (Bexon KJ, 2017; Li, 2021; Zubakov et al., 2010). The first five and the last miRNA marker, miR-943, were detected in other body fluids but are over-expressed in semen (Bexon KJ, 2017; Zubakov et al., 2010). According to the doctoral study of Bexon KJ (Bexon KJ, 2017), miR-10a, 135a and 888 are highly expressed in all semen samples supporting its use in semen detection, with and without sperm cells.

1.4.1.7. DNA methylation

DNA methylation is a process initiated by methyltransferases which affect gene expression regulation without altering the DNA sequence (Choi et al., 2014). This process occurs at 5′-position of the pyrimidine ring of cytosine in CpG dinucleotides in both DNA strands (J. H. An et al., 2013; Lee et al., 2012). Different cell types or tissues have different methylation patterns and those are called tissue-specific differentially methylated regions. DNA methylation profiling can be achieved using different methods such as methylation sensitive restriction enzyme digestion polymerase chain reaction, bisulfite sequencing and methyl-DNA immunoprecipitation (Li, 2021)**.** Considering that different cell types have different methylation patterns, DNA methylation may be a useful tool to differentiate between them (J. H. An et al., 2013; Lee et al., 2012). Semen may be potentially identified using hypomethylation markers such as cg22407458, PRMT, ZC3H12D, cg04382920, cg11768416, ZNF282, DACT1, USP49 e PRMT2 (J. H. An et al., 2013; Ghai et al., 2020; Lee et al., 2012).

1.4.2. Immunochromatographic tests

Seratec PSA Semiquant (Seratec, Germany) and Rapid Stain Identification (RSID)-Semen (Independent Forensics, USA) are immunochromatographic tests used to detect the presence of semen. Both detect the presence of a specific antigen (Harbison & Fleming, 2016). In the sample well of immunochromatographic tests there are dye-labelled antibodies that will bind to the target antigen, if present in the sample. As represented in *Figure 8*, in the test area there is an immobilized antibody and in the control zone an antiglobulin that recognizes the antibody (Li, 2021). In the test zone a pink line will appear if the target antigen is detected by the dye-labelled antibody in the conjugate pad. In the control area unbound antibody will bind to the antiglobulin regardless of the presence of the target antigen and a pink line will appear (Li, 2021).

The control line ensures that the test works properly and it is only valid if a pink line appears (Li, 2021). Before testing the sample, it is first mixed with the respective supplied buffer and the results are shown within 10-20 minutes after pipetting the sample into the sample well (Dawnay & Sheppard, 2023).

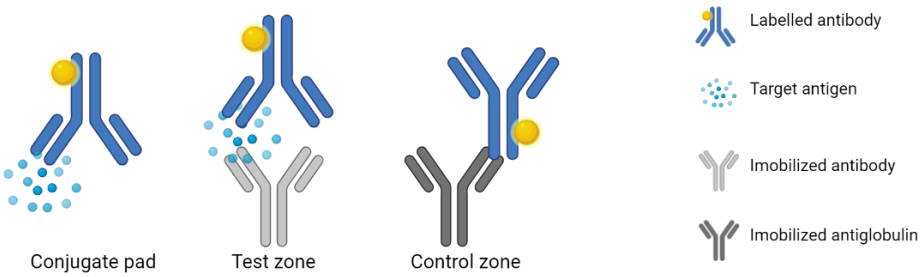


Figure 8 - Immunochromatographic test. Created with BioRender.com.

These tests are, however, affected by the high-dose hook effect (*Figure 9*). High-dose hook effect happens when the sample to be tested presents high amounts of the target antigen, contributing to a large part of it not binding to the dye-labelled antibody in the conjugate pad (Old et al., 2012). The free antigen moves freely, exceeds the dye-labelled antigen and binds to the antibody present in the test area. The binding of the free antigen to the antibody in the test area prevents the binding of the labelled antigen, reducing the intensity of the positive signal or even causing a false negative (Old et al., 2012).

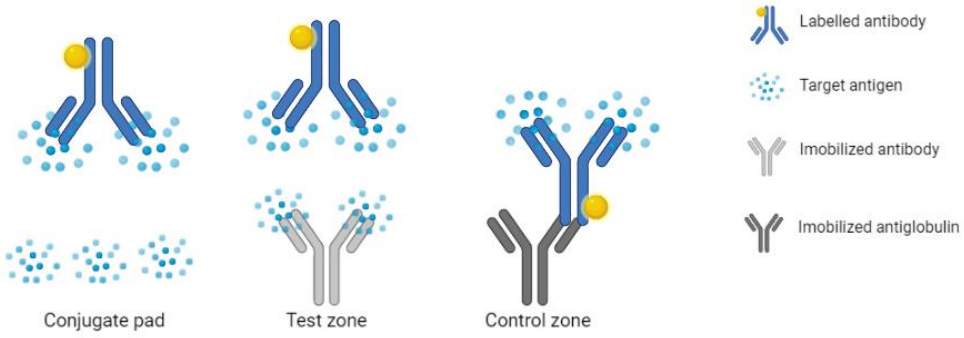


Figure 9 - High-dose hook effect. Created with BioRender.com.

1.4.2.1. Seratec PSA Semiquant

Seratec PSA Semiquant is a presumptive immunochromatographic test which detects PSA. PSA is a glycoprotein produced by prostatic epithelium cells and is found in semen in high quantities (0.2-5 mg/mL) (Li, 2021). It is also present in vasectomized, men and those exhibiting oligospermia and azoospermia allowing the test to be used in samples from these men (Berrena, Taylor, 2015; Virkler & Lednev, 2009a). This test was initially used to detect prostate cancer in patients and is, therefore, a semiquantitative test (DiFrancesco & Sutton, 2015; Hochmeister et al., 1999). An elevated concentration of PSA in the plasma is present in these cancer patients and also in cases of benign prostatic hyperplasia and prostatitis (Li, 2021).

Even though it was initially thought to be prostate specific, its presence was found in different body fluids which suggests it is not prostate nor male specific (Mannello et al., 1998; Schmidt et al., 2001). In addition to being present in semen it is also present in female urine (Mannello et al., 1998; Schmidt et al., 2001), amniotic fluid (Wolff et al., 1999) and breast milk (Yu & Diamandis, 1995). Alternatively, PSA can be produced as a steroid hormone-mediated response (Mannello et al., 1998)**.** In breast milk PSA existence might be explained by the high amounts of steroids hormones produced during pregnancy (Yu & Diamandis, 1995). The hormones stimulate PSA production in the pregnant women breasts and the glycoprotein is then secreted into breast milk (Yu & Diamandis, 1995) . In this body fluid, PSA is present in quantities ranging from <0.01 mg/L to 350 mg/L (*Table 1*), and these numbers decrease with time after birth (Yu & Diamandis, 1995). In the amniotic fluid, PSA concentrations varies from 0.003 ng/mL to 1.22 ng/mL (*Table 1*). As the gestation period progresses, the concentrations increase. Between the 16th and 18th week the mean PSA value is 0.19 ng/mL increasing to 0.39 ng/mL at 36th week (Wolff et al., 1999). These differences in PSA concentration may be indicative of a role in fetal growth and development (Wolff et al., 1999). In female urine, the presence of PSA can be explained by the same embryological origin of the female periurethral glands as the male prostate (Schmidt et al., 2001). According to a study that compares PSA levels in urine samples from women who take oral contraceptives and women who do not, the numbers are higher in women who take oral contraceptives. In the urine of females that take oral contraceptives, PSA values range from 0.09 mg/L to 1.24 mg/L in contrast to 0.02 mg/L to 0.15 mg/L in the urine of women that do not take oral contraceptives (*Table 1*) (Mannello et al., 1998)**.** In these fluids PSA is present in much lower concentrations compared to semen and it is therefore possible to continue to use it as a biomarker for the presence of semen (Laffan et al., 2011). However,

because it is present in all of these different body fluids, it cannot be considered a confirmatory test for semen detection.

Table 1 - Different PSA concentrations in breast milk, amniotic fluid, and urine of female individuals on oral contraceptive and not.

Body fluid	**Breast milk (mg/L)**	**Amniotic fluid (ng/µL)**	**Urine (µg/L)**	
			Oral contraceptive	No oral contraceptive
[PSA]	<0.01 - 350	0.003 – 1.22	0.09 – 1.24	0.02 – 0.15

In the present test the glycoprotein is detected by two monoclonal anti-PSA antibodies. After incubating the sample in running buffer, it is placed into the test well where it will bind to colloidal gold-labelled anti-PSA antibodies (Kishbaugh et al., 2019). If PSA is present in the sample the antigen will bind to this antibody forming an antibody-PSA complex (DiFrancesco & Sutton, 2015; Kishbaugh et al., 2019). In the test-line “T”, the complex formed binds to another anti-PSA monoclonal antibody and a pink line appears. The immunochromatographic test has two control lines, the control-line “C” for possible procedural errors and for the integrity of test components and the internal standard. The latter is responsible for the semi quantification of PSA with the colour intensity being related to a PSA concentration of approximately 4 ng/mL (DiFrancesco & Sutton, 2015). Both control regions bind freely to the gold-labelled PSA antibodies regardless of the presence of the glycoprotein. For the test to be valid both control lines must appear pink (DiFrancesco & Sutton, 2015). The results must be read within 10 minutes after the sample was put in the sample well. In *Figure 10* is a representation of a negative and a positive Seratec PSA Semiquant results.

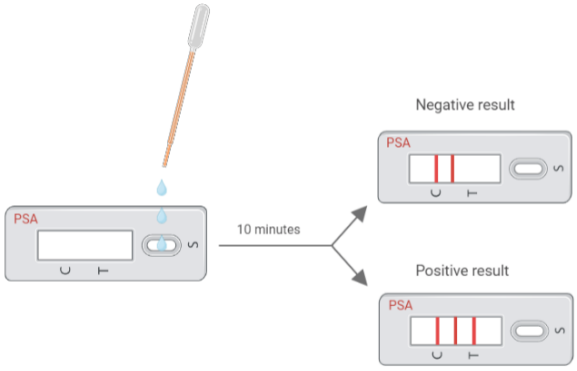


Figure 10 – Seratec PSA Semiquant representation of a negative (above) and a positive (below) test result. Created with BioRender.com.

Seratec PSA Semiquant gives positive results with PSA concentrations as low as 0.5 ng/mL (Laux & Barnhart, 2011). Samples with concentrations higher than 50.000 ng/mL are susceptible to high dose hook effect (Hochmeister et al., 1999). To avoid this it is recommended to dilute the samples to at least 1:100 (Hochmeister et al., 1999). For this reason, a negative result may be indicative of absence of semen in the sample, an amount below the detection limit, or an excess of antigen in the sample (Hochmeister et al., 1999).

1.4.2.2. RSID-Semen

RSID (Independent Forensics) is a confirmatory immunochromatographic strip test to assess the presence of human semen. This test detects the presence of Sg I and II, proteins which usage has advantages over the detection of PSA (Laffan et al., 2011; Old et al., 2012). The average concentration of Sg in the seminal fluid is 19 mg/mL, 10 times higher than the concentration of PSA (Laffan et al., 2011). Unlike PSA, Sg is not present in other body fluids. Its presence was detected in other tissues such as kidneys, retina and trachea but as these are not usually submitted in sexual assault cases, it is not a concern (2021; Laffan et al., 2011).

The RSID-Semen test uses two antihuman semenogelin monoclonal antibodies to detect the presence of semen (Old et al., 2012). One antibody is conjugated to colloidal gold and is on a conjugate pad beneath the sample window and the other is on the test-line “T”. In the control zone there is an anti-mouse antibody (Old et al., 2012). In cases in which semen is present in the sample, the antigen binds to the antibody conjugated to colloidal gold forming an antigen-antibody complex. The complex binds to the antibody in the test region forming the red line (Old et al., 2012). The colloidal gold conjugated

antibody that did not bind to the sample will keep diffusing through the conjugate pad and will bind to the anti-mouse at the control region (Old et al., 2012). The test is only valid if a line appears in the control region and must be read 10 minutes after loading the sample (*Figure 11)* (Old et al., 2012).

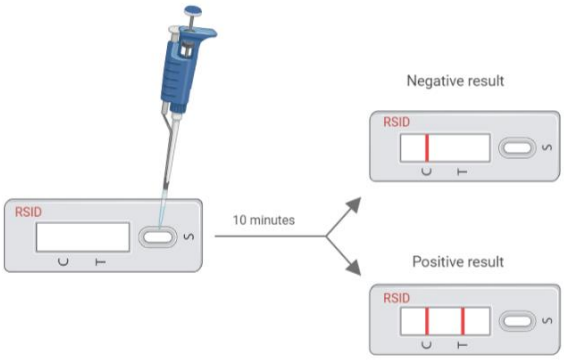


Figure 11 – RSID-Semen representation of a negative (above) and a positive (below) test result. Created with BioRender.com.

RSID-Semen test can detect as little as 1 μL of semen in a sample (Old et al., 2012). Samples containing large amounts of semen, from 3 μL, might give a weak result due to high dose hook-effect (Old et al., 2012). In these situations samples should be diluted to 1:20 to decrease the probability of high dose hook-effect (Old et al., 2012). According to Old J and co-authors (Old et al., 2012), RSID-Semen does not react with other important body fluids such as blood, urine, saliva and breast milk.

2. Aims

Sexual assault samples are very often scarce and, in these cases, may contain little DNA. This can pose a threat to a successful investigation due to lack of sufficient DNA proof. The main goal of this work is to analyse comparatively the presumptive testing currently in-use in forensic routine case-work in our laboratory. This will allow to optimize the workflow in order to obtain improved results for both semen detection tests and subsequent genetic profiling. With this main aim traced, specific aims were established in this work:

1) To analyse results of the presumptive tests applied in common routine case-work in closed cases available and previous tested at the laboratory;
2) To compare the sensitivity of Seratec PSA Semiquant (Seratec, Germany) and RSID-Semen (Independent Forensics, USA);
3) To establish a new method for processing vaginal swabs consisting in using the entire swab;
4) To validate a new method, by comparing the sensitivity and capacity for obtaining full DNA profiles with the previous method* used in-house.

* The previous method consists in using one part of the swab for presumptive testing and the rest for DNA extraction. In the present work, the entire swab was used for both presumptive testing and DNA extraction to avoid cutting samples when they are already scarce. The routinely used method will be labelled as “Method 1” and the method to be tested as “Method 2”. A brief description of both methods is represented in *Figure 12* in the section Materials and Methods.

3. Materials and methods

3.1. Analysis of sexual assault cases from 2019 to 2022

We studied the sexual assault cases received at the Institute of Forensic Science Luís Concheiro (INCIFOR) between the years of 2019 and 2022. Within the cases received, those that were closed investigations were selected for this study. For each case and for every sample, we analysed: 1) the presumptive testing performed; 2) the DNA quantification results and; 3) the autosomal and Y chromosomal DNA STR profiles.

3.2. Sample and test preparation

The samples used in this study were semen and post-coital interval (PCI) samples and were provided by an anonymous donor under informed consent.

Samples were divided in five series: Serie A, Serie B, Serie C, Serie D and Serie E. A brief description of each series is presented in *Table 2*. Both Seratec PSA Semiquant and RSID-Semen were performed in Serie A according to Method 1, i.e., following the manufacturer's instructions (Seratec PSA Semiquant, Art.-Nr. PSM400F, Seractec, 2011 and RSID-Semen Technical Information Sheet, Independent Forensics, 2011). RSID-Semen was done in the remaining series following two alternative methods that will be explained below. All Series are independent from each other. Even though some dilutions and PCI are the same, each series was prepared separately.

Table 2 - Description of the series used in this work. "Yes" means the test was performed and "No" means the test was not performed.

Serie	Post-coital interval (PCI)	Dilutions	PSA	RSID
A	2h, 4h, 8h, 12h, 24h, 36h	-	Yes	Yes
B	-	1:10, 1:25, 1:50, 1:75, 1:100, 1:250, 1:500, 1:750, 1:1000	No	Yes
C	4h, 12h, 24h, 36h, 48h, 60h, 72h	-	No	Yes
D	8h, 12h, 24h, 36h, 48h	-	No	Yes
E	-	1:50, 1:250, 1:500, 1:1000, 1:1250, 1:1500, 1:2000	No	Yes

Semen dilutions were prepared using frozen semen for Serie B (stored at approximately -20ºC) and fresh semen for Serie E. The respective volumes of semen and nuclease-free water (NFW) are in *Table 3* and *Table 4*. "Input DNA" corresponds to the dilution stocks that were used in the "Sample" volumes to obtain the "Final dilution" factor. Semen dilutions swabs were prepared by pipetting 100 μL of diluted semen into cotton swabs and left to air dry overnight. Vaginal post coital samples were collected with a swab by the donor and were also left in the tube to air dry overnight. All swabs were prepared in duplicate sets, one set for Method 1 and the other for Method 2. The swabs were kept at room temperature until examination.

Table 3 - Components and respective volumes of the dilutions done on Serie B. The volume in the "Sample" column corresponds to the volumes used from the "Input DNA" stocks to achieve the "Final dilution" factor shown in the 1st column.

Final dilution	Input DNA	Sample (μL)	NFW (μL)
1:10	-	200	1800
1:25	1:10	600	900
1:50	1:25	300	300
1:75	1:25	200	400
1:100	1:10	100	900
1:250	1:25	100	900
1:500	1:50	100	900
1:750	1:75	100	900
1:1000	1:100	100	900

Table 4 - Components and respective volumes of the dilutions done on Serie E. The volume in the "Sample" column corresponds to the volumes used from the "Input DNA" stocks to achieve the "Final Dilution" factor shown in the 1st column.

Final dilution	Input DNA	Sample (μL)	NFW (μL)
1:50	-	10	490
1:250	1:50	200	800
1:500	1:250	500	500
1:1000	1:500	250	250
1:1250	1:500	120	240
1:1500	1:500	100	200
1:2000	1:1000	150	150

To perform the Seratec PSA Semiquant test, a piece of swab was cut and incubated in 160 μL of the provided buffer. The samples were incubated for two hours at room temperature with agitation. After the incubation, using a Pasteur pipette, 3 drops of the sample and buffer mixture were placed on the sample well of the test. The results were recorded after 10 minutes.

To perform the RSID-Semen test, two different methods were followed. Method 1 was considered the standard method used in the laboratory. In this Method, a piece that was cut from the cotton swab was placed in a 1.5 mL eppendorf tube and incubated with 100 μL of RSID Universal buffer for two hours with agitation. After incubation the sample and buffer mixture were transferred to the sample well of the test strip (*Figure 12*).

According to Method 2, instead of cutting the sample, we used the entire swab to do RSID-Semen. In a 1.5 mL eppendorf tube the entire swab was placed and incubated with 250 μL of RSID Universal buffer for two hours with agitation. After the incubation approximately 100 μL of sample and buffer mixture was transferred to the sample well of the test strip. The remaining was kept for DNA extraction. Series D and E were subjected to an extra centrifugation step for 10 minutes at 14000 g after incubation with RSID Universal buffer as illustrated in *Figure 12*. 100 μL of the supernatant was retained for RSID-Semen and pelleted cells were kept for DNA extraction. The results were recorded after 10 minutes.

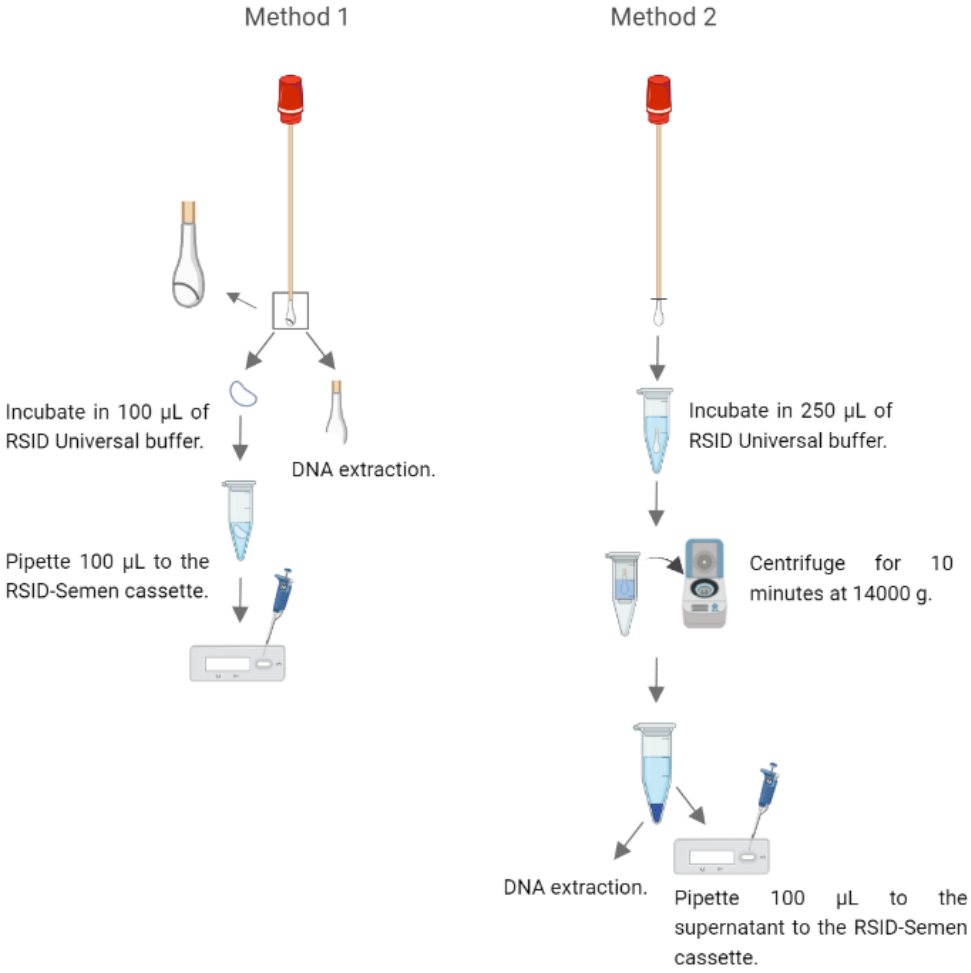


Figure 12 - Description of Method 1 and Method 2. In Method 1 a cut of the sample is incubated in 100 µL of RSID-Semen Universal buffer and the rest is used for DNA extraction. In Method 2, the entire sample is used for both. Firstly, it is incubated in 250 µL of RSID-Semen Universal buffer. Secondly, it is centrifuged for 10 minutes. The supernatant is used to do RSID-Semen test and the pellet kept for DNA extraction. Created with BioRender.com.

3.3. DNA extraction

After the presumptive testing phase, samples were taken for the extraction process. Not all samples submitted for presumptive testing were extracted. The samples extracted are listed in *Table 5*. The samples were chosen in order to catch the loss of intensity from the positive signal to the negative result of the RSID-Semen test.

Table 5 - Samples submitted to normal and differential extraction with AutoMate Express Forensic DNA Extraction System with PrepFiler Express Forensic DNA Extraction Comercial Kit (Applied Biosystems, Thermo Fisher Scientific).

DNA extraction	**Standard extraction**	**Differential extraction**
Serie A	-	-
Serie B	-	-
Serie C	24h, 36h, 48h, 60h, 72h (Method 1 and Method 2)	24h, 36h, 48h, 60h, 72h (Method 2 RSID-Semen cassetes)
Serie D	-	8h, 12h, 24h, 36h (Method 1 and Method 2)
Serie E	-	1:50, 1:250, 1:500, 1:1000, 1:1500, 1:2000 (Method 1 and Method 2)

3.3.1. Standard DNA extraction using AutoMate *Express* Forensic DNA Extraction System with PrepFiler *Express* Forensic DNA Extraction commercial Kit

The extraction process was done using the PrepFiler *Express* Forensic DNA Extraction Kit (Applied Biosystems, Thermo Fisher Scientific). The LySep Columns were placed into the PrepFiler sample tubes and properly labelled. Samples were carefully transferred to the columns. For the extraction of the samples according to Method 1, the remaining of the swab was placed inside the columns. For Method 2, the entire swab that was used for the presumptive testing was transferred to the columns. To each column/tube assembly 500 μL of the lysis solution plus 5 μL 1M DTT were added. The column/tube assembly was then placed in the thermal shaker for 40 minutes at 70^{o}C and 750 rpm. After the lysis, samples were centrifuged for 2 minutes at 10000 g to transfer the lysate to the sample tube. The columns were then disposed of and the tubes placed in the AutoMate *Express* Forensic DNA Extraction System (Applied Biosystems, Thermo Fisher Scientific). A scheme of the normal extraction process for the two methods is shown in *Figure 13*. Samples were extracted for 30 minutes. After extraction all samples were stored in the fridge prior to quantification.

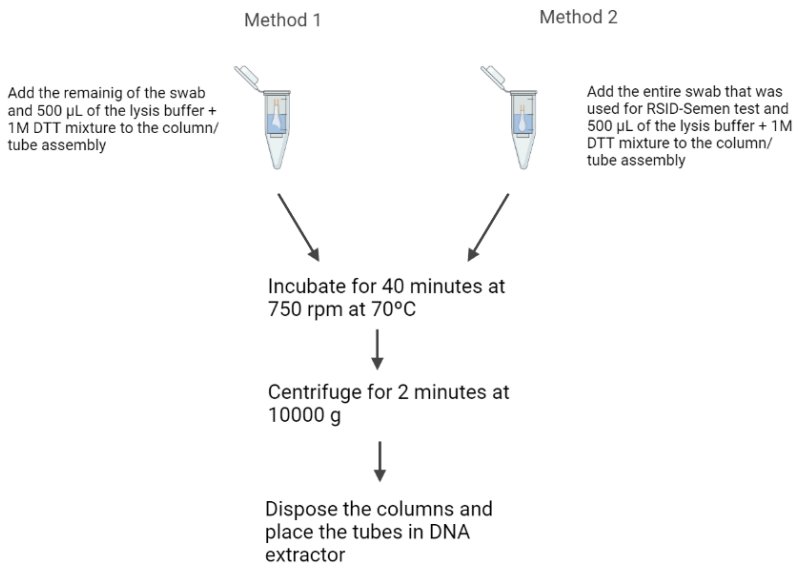


Figure 13 - Schematic representation of standard extraction for Method 1 and Method 2. Created with BioRender.com.

3.3.2. Differential extraction

A representation of differential extraction process is shown in *Figure 14*. Differential extraction of samples analysed according to Method 1 started with the washing of the remaining of the swab in 3 mL of NFW for 2h at room temperature (RT) with agitation. After incubation, 1 mL of the liquid was pipetted into a new eppendorf tube and centrifuged for 5 minutes at 14000 g. The supernatant was discarded, leaving 50 µL of precipitate. The process was repeated until the liquid in the first tube was finished. After, a column was placed in the new tube, the swab transferred to the column and centrifuged for 10 minutes at 14000 g. The column and supernatant were discarded and around 50 µL of the pellet left in the tube. After the washing steps, the first lysis started. All the components and respective volumes needed for the 1st and 2nd lysis are listed in *Table 6*.

Table 6 - Components and respective volumes needed for the 1st and 2nd lysis per sample.

Components	**Volumes (µL)**	
	1st lysis	**2nd lysis**
Extraction buffer	400	600
Proteinase k	20	-
DTT 1M	-	5
Prepfiler extraction buffer	450	500

To each sample tube 400 µL of extraction buffer and 20 µL of Proteinase K was pipetted. Samples with RSID-Semen positive result were incubated for 2h at 56ºC at 1000 rpm and samples with a negative result were incubated for 1h under the same conditions. The supernatant was transferred to a new tube and labelled as female fraction. Now the epithelial cells are separated from the sperm cells. To the pellet left in each tube was added 200 µL of extraction buffer. A 10-minute centrifugation step at 13000 g followed. The supernatant was discarded. The process was repeated three times. To the pellet 500 µL of PrepFiler extaction buffer and DTT 1M mixture was added. A PrepFiler tube/column was assembled and the mixture was transferred to the column. Samples were incubated at 70ºC for 40 minutes at 750 rpm. A centrifugation step of 2 minutes at 140000 g followed. Discard the column. The second fraction is now ready to go into the extraction robot. 50 µL of the female fraction was taken and transferred to a new tube to which were added 450 µL of PrepFiler extraction buffer. The first fraction is now ready to go to the extraction robot.

Samples analysed according to Method 2 without the additional centrifugation step before RSID-Semen presumptive testing were extracted following the same procedure of Method 1 samples. The extraction process of Method 2 samples with the additional centrifugation step started in the 1st lysis, skipping the washing steps.

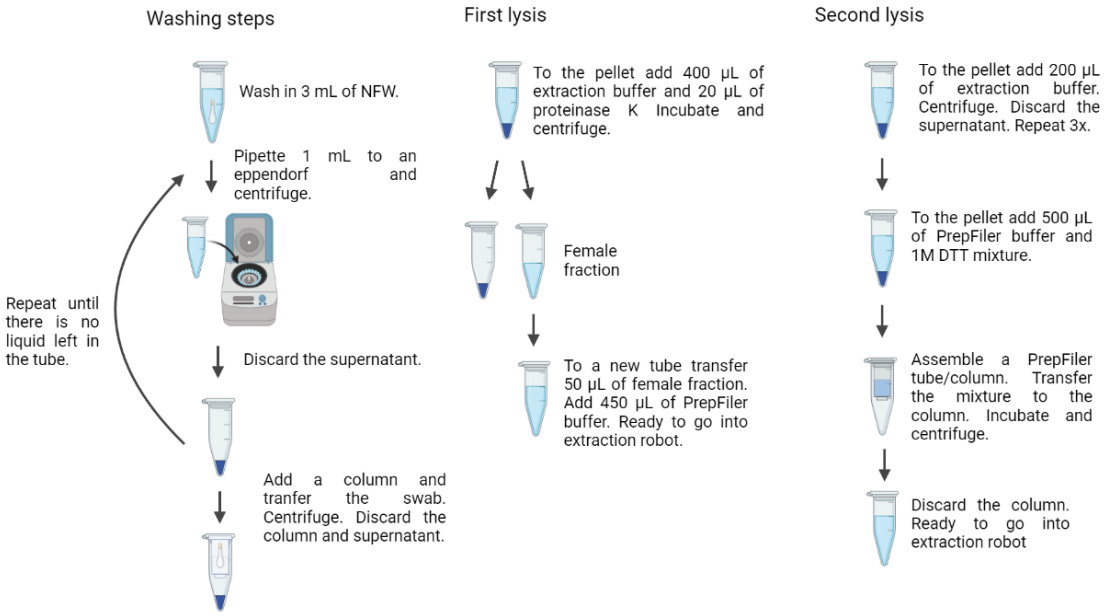


Figure 14 - Schematic representation of the differential extraction process. Created with BioRender.com.

The sample pad of Serie C RSID-Semen cassettes was also extracted with differential lysis. The sample pad was cut in half, one to be extracted and the other to keep (*Figure 15*). The extraction process followed the same procedure described for Method 1 samples.

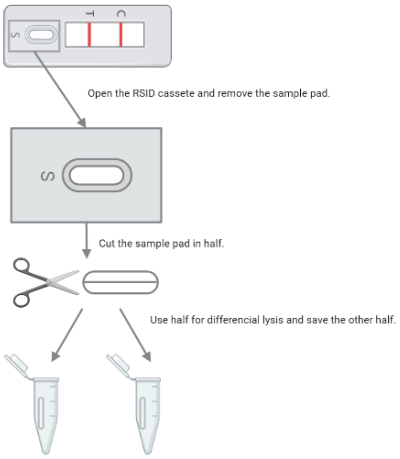


Figure 15 - A description of the preparation of the RSID-Semen cassette extraction. Created with BioRender.com.

3.4. Quantification of extracted DNA

3.4.1. DNA quantification using Quantifiler Trio DNA Quantification Kit

The quantification of the extracted DNA samples was performed using Quantifiler Trio DNA Quantification Kit (Applied Biosystems, Thermo Fisher Scientific). The procedure was done following the manufacturer's instructions (Applied Biosystems, 2018). Standards were prepared with concentrations ranging from 50 ng/mL to 0.005 ng/mL according to *Table 7*. All standards were prepared in duplicate. Comparing the standards calibration curve with the samples signal intensity produced in each cycle allows quantification of the samples.

Table 7 - Standards dilutions preparation. Adapted from Quantifiler Trio User Guide (Applied Biosystems, 2018).

Standard	**Concentration (ng/μL)**	**Example Volumes**	**Dilution Factor**
1	50.000	10μL (100 ηg/μL stock) + 10μL Quantifiler® THP dilution buffer	2x
2	5.000	10 μL [Std. 1] + 90 μL Quantifiler™ THP DNA dilution buffer	10x
3	0.500	10 μL [Std. 2] + 90 μL Quantifiler™ THP DNA dilution buffer	10x
4	0.050	10 μL [Std. 3] + 90 μL Quantifiler™ THP DNA dilution buffer	10x
5	0.005	10 μL [Std. 4] + 90 μL Quantifiler™ THP DNA dilution buffer	10x

In each well of a MicroAmp Optical 96-Well Reaction Plate (Applied Biosystems, Thermo Fisher Scientific) was pipetted 18 μL of Master Mix containing 10 μL of Quantifiler Trio Reaction Mix and 8 μL of Quantifiler Trio Primer Mix. In the standard's wells, 2 μL of each standard was pipetted, and in the sample wells, 2 μL of each sample was pipetted. Positive and negative controls were included. The plate was sealed with a MicroAmp Optical Adhesive Film (Applied Biosystems, Thermo Fisher Scientific) and centrifuged for 1 minute at 1000 g in a Centrifuge 5424 (Eppendorf, Germany). Quantitative PCR was performed on 7500 Real Time PCR System (Thermo Fisher Scientific) for 40 cycles. Results were analysed using Applied Biosystems 7500 Real-Time PCR Software.

For post-coital male-female samples, an estimate of the female DNA concentration was calculated subtracting the male DNA concentration (Y-chromosome DNA) to total DNA. Male:female DNA ratio was calculated dividing the concentration of male DNA by the concentration of the female DNA.

3.5. DNA amplification by PCR

PCR procedures were performed on a PCR set-up room following the laboratory standard operating procedure. Both procedures were conducted a bench laminar flow clean air cabinet (Mini V/PCR, Telstar).

3.5.1. DNA amplification by PCR using GlobalFiler commercial kit

GlobalFiler kit (Applied Biosystems, Thermo Fisher Scientific) amplifies 21 autossomal loci (D3S1358, vWA, D16S539, CSF1PO, TPOX, D8S1179, D21S11, D18S51, D2S441, D19S433, TH01, FGA, D22S1045, D5S818, D13S317, D7S820, SE33, D10S1248, D1S1656, D12S391, D2S1338), 1 Y-STR (DYS391), 1 Y InDel and amelogenin (Ludeman et al., 2018). The procedure was done following the manufacturer's instructions (Applied Biosystems, 2019a).

A mixture of 3 μL of Master Mix ("hot-start" DNA polymerase, buffer, salts, dNTPs, detergent, carrier protein, sodium azide and stabilizing components), 1 μL of Primer Set and 5 μL of water per sample was prepared. To each tube was added 9 μL of Master Mix and 1 μL of each DNA sample. In the negative control was added 1 μL of water instead of DNA and in the positive control a commercial DNA with known genotype. Lastly, samples were amplified in a Veriti 96-Well Thermal Cycler (Applied Biosystems, Thermo Fisher Scientific) for 30 cycles following standard thermal cycler parameters described in *Table 8*.

Table 8 - PCR cycles used during DNA amplification with GlobalFiler Amplification kit.

STEPS	**Initial denaturation**	**30 cycles**		**Final extension**	**Final hold**
		Denaturation	**Annealing/ Extension**		
Temperature	95ºC	94ºC	59ºC	60ºC	10ºC
Time	1 minute	10 seconds	90 seconds	10 minutes	∞

DNA amplification of these samples was repeated for a total volume of 25 μL with the exclusion of the samples 8h PC. DNA and water volumes used to repeat PCR are in *Table 9*.

Table 9 - DNA and water volumes used for the second PCR with GlobalFiler Amplification kit. M1 and M2 stand for Method 1 and Method 2, respectively. SF stands for Sperm Fraction.

Samples	DNA (μL)	Water (μL)
M1 12h SF	15	-
M1 24h SF	9.20	5.80
M1 36h SF	3.50	11.50
M2 12h SF	4.10	10.90
M2 24h SF	5.90	9.10
M2 36h SF	6.50	8.50

3.5.2. DNA amplification by PCR using YFiler Plus commercial kit

YFiler Plus kit (Applied Biosystems, Thermo Fisher Scientific) amplifies 27 chromosome Y markers (DYS576, DYS389I, DYS635, DYS389II, DYS627, DYS460, DYS458, DYS19, YGATAH4, DYS448, DYS391, DYS456, DYS390, DYS438, DYS392, DYS518, DYS570, DYS437, DYS385, DYS449, DYS393, DYS439, DYS481, DYF387S1, DYS533) (Gopinath et al., 2016). The procedure was done following the manufacturer's instructions (Applied Biosystems, 2019b).

A mixture of 4 μL of Master Mix (hot-start DNA polymerase, buffer, salts dNTPs, carrier proteins and sodium azide), 2 μL of Primer Set, 3 μL of water per sample was prepared. To each tube was pipetted 9 μL of Master Mix and 1 μL of each DNA sample. The positive control was prepared diluting 1 μL of the control in 9 μL of water. In the negative control was added 1 μL of water instead of DNA. Lastly, samples were amplified in a Veriti 96-Well Thermal Cycler (Applied Biosystems, Thermo Fisher Scientific) for 30 cycles following standard thermal cycler parameters described in *Table 10*.

Table 10 - PCR cycles used during DNA amplification with YFiler Plus Amplification kit.

STEPS	**Initial denaturation**	**30 cycles**		**Final extension**	**Final hold**
		Denaturation	**Annealing/ Extension**		
Temperature	95ºC	94ºC	61.5ºC	60ºC	8ºC
Time	1 minute	4 seconds	32 seconds	22 minutes	∞

DNA amplification of these samples was repeated for a total volume of 25 μL with the exclusion of the samples 8h PCI. All DNA volumes used were 10 μL and no water was added.

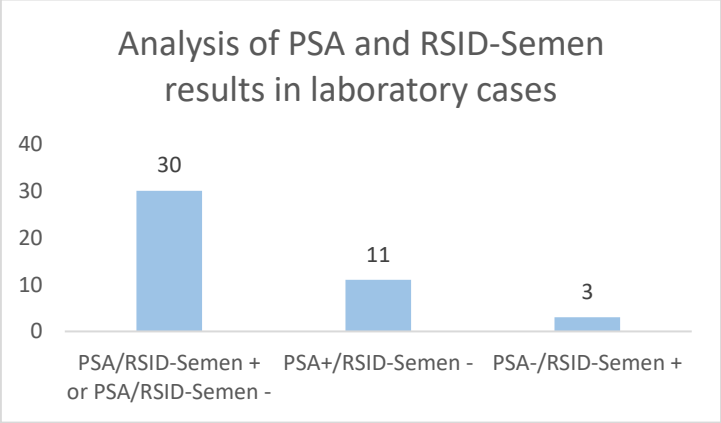
3.6. Capillary electrophoresis

In each MicroAmp® Optical 96-Well Reaction Plate (Applied Biosystems, Thermo Fisher Scientific) was pipetted 9,6 μL of formamide and 0,4 μL of LIZ 600. 1 μL of each PCR product was added and 1 μL of ladder (GlobalFiler Allelic Ladder or YFiler Plus, respectively). To the samples amplified with a total volume of 25 μL the process was repeated but 2 μL of PCR product was used. A 3-minute denaturing process at 85ºC followed. Finally, the plate was placed in a Genetic Analyzer 3500 (Applied Biosystems, Thermo Fisher Scientific) and capillary electrophoresis was performed. The obtained electropherograms were analysed using GeneMapper® ID-X v1.4 software.

4. Results and discussion

4.1. Analysis of Seratec PSA Semiquant and RSID-Semen results for cases received by the laboratory between 2019 and 2022

Of the total number of sexual assault cases received by the laboratory in this time period, 44 cases were closed investigations at the date of this analysis. As shown in *Graph 1*, in 30 cases, both Seratec PSA Semiquant and RSID-Semen were either positive or negative, showing concordant results. In 11 cases Seratec PSA Semiquant was positive and RSID-Semen was negative. In the remaining 3 cases the opposite was observed.



Graph 1 - Analysis of Seratec PSA Semiquant and RSID-Semen results of sexual assault closed cases received by the laboratory between 2019 and 2022. On the Y axis are represented the number of cases analysed.

In situations in which Seratec PSA Semiquant is positive and RSID-Semen is negative results are considered to be expected having into account that Seratec PSA Semiquant has higher sensitivity than RSID-Semen (Suttipasit & Wongwittayapanich, 2018). On the other hand, situations in which Seratec PSA Semiquant is negative and RSID-Semen is positive are not as usual. As previously explained, Seratec PSA Semiquant is a presumptive test where a negative result should mean “absence of semen” and a positive result indicative of the “possibility of its presence" (DiFrancesco & Sutton, 2015). Furthermore, a positive RSID-Semen test result is considered confirmatory for the presence of semen (Old et al., 2012). A situation where the PSA result is negative and the RSID-Semen is positive can be explained by the persistence

of the respective proteins in the semen. According to Laffan and co-authors (Laffan et al., 2011), in postcoital high vaginal swabs Sg can be detected for longer periods of time than PSA. Most often, the time interval between sexual assault and sample collection is not transmitted to the laboratory, and this may be an explanation for a positive RSID-Semen and a negative PSA result.

4.2. RSID-Semen *vs* Seratec PSA Semiquant detection interval assessment

After the inconsistencies found in some cases between Seratec PSA Semiquant and RSID-Semen results, both tests were compared using anonymously postcoital interval (PCI) vaginal samples. These samples comprised 2h, 4h, 8h, 12h, 24h, 36h and 48h PCI. RSID-Semen and Seractec PSA Semiquant results are shown in *Table 11.* All samples are positive for Sg with the intensity of the signal decreasing after 36h. Seratec PSA Semiquant was positive in samples taken until after 12h, the latter being a very low intensity positive signal. At 24h, 36h and 48h PCI, Seratec PSA Semiquant results were negative.

Table 11 - Comparison of RSID-Semen and Seratec PSA Semiquant results with post-coital interval vaginal samples. Arrows represent the intensity of the positive signals observed: low ↓, very low ↓↓ and very very low ↓↓↓.

Serie A	RSID-Semen	Seratec PSA Semiquant
2h	+	+
4h	+	+
8h	+	+
12h	+	+ ↓↓
24h	+	-
36h	+ ↓	-
48h	+ ↓↓↓	-

The differences of detection interval when comparing both systems is, of course, related to the different detection target (Sg vs PSA). According to the study of Suttipasit & Wongwittayapanich (Suttipasit & Wongwittayapanich, 2018), the detection rate of Sg is higher than the one for PSA. Their results showed that after 24h the detection rate of PSA declined significantly. On the other hand, the decline observed with Sg testing was not as big, favouring the use of this until approximately 72 hours after the sexual assault (Suttipasit & Wongwittayapanich, 2018). Other studies concluded that 48h is the end of

the window for PSA positivity, with a positive result being rarely obtained after this time (Graves et al., 1985; Jamshidi et al., 2013).

In sexual assault cases, the probability of more than 24 hours having elapsed between the assault and sample collection cannot be ruled out, which affects the reliability of a Seractec PSA Semiquant result. In addition to Sg persisting longer in semen, this protein exists in much higher concentrations when compared to PSA (Laffan et al., 2011). This factor contributes to the ability of RSID-Semen detecting Sg body fluid for a longer period of time (Laffan et al., 2011; Old et al., 2012).

4.3. Comparison of methods for vaginal swab processing

4.3.1. Sensitivity

The sensitivity of RSID-Semen was tested on various dilutions of previously frozen (Serie B) and freshly obtained semen (Serie E). For Serie B the results are shown in *Table 12*. For Method 1, semen was detected down to 1:750 dilution. Nonetheless, the intensity of the positive signal decreases at 1:500. In the 1:1000 dilution the positive signal at the test line was no longer observed. In Method 2, Sg was detected in every dilution. However, the positive line in 1:1000 dilution was less intense than the others.

Table 12 - Sensitivity of RSID-Semen on different frozen semen dilutions (Serie B) for both Method 1 and Method 2. The symbol ↓ means that the intensity of the positive line is low.

Dilutions	**RSID-Semen**	
	Method 1	**Method 2**
1:10	+	+ ↓
1:25	+	+
1:50	+	+
1:75	+	+
1:100	+	+
1:250	+	+
1:500	+ ↓	+
1:750	+ ↓	+
1:1000	-	+ ↓

In Method 2, the positive line in the first dilution was less intense than the following ones revealing a potential high-dose hook effect. As explained previously, high-dose hook effect happens when the sample has a very high concentration of the target

antigen. A large portion of the target antigen will not bind to the gold-labelled antibody in the conjugate pad. In the test zone the unlabelled antigen will compete with the labelled antigen and, as the first is in excess, it will bind to the antibody. Subsequently, the result will be negative or a very weak positive (Li, 2021). To confirm or rule out the possibility of high-dose hook effect, a new 1:10 dilution was made from the previous 1:10 dilution and RSID-Semen was repeated. As can be seen in *Figure 16*, the positive line for the new 1:10 dilution was stronger than the first 1:10 dilution. This means that in the first dilution too much of the target antigen was present, which contributed to the fact that it did not fully bind to the labelled antibody on the conjugate pad, confirming the high-dose hook effect phenomenon.

Figure 16 – High-dose hook effect. MS1:10(2) is the first dilution tested. MS1:10(2)2º is the new 1:10 dilution made from the first 1:10 dilution.

The results of Serie E samples are represented in *Table 13*. For Method 1, Sg is detected until 1:2000 dilution and the positive signal weakens as of 1:1500 dilution. For Method 2, Sg is only undetectable until 1:6000 dilution. Yet, the positive signal intensity starts to decrease at 1:3000.

Table 13 - Sensitivity of RSID-Semen on different fresh semen dilutions (Serie E) for Method 1 and Method 2. Arrows represent the intensity of the positive signals observed: low ↓ and very, very low ↓↓↓.

DILUTIONS	RSID-Semen	
	Method 1	Method 2
1:500	+	+
1:1000	+	+
1:1250	-	+
1:1500	+ ↓	+
1:2000	+ ↓	+
1:3000	-	+ ↓↓↓
1:4000	-	+ ↓↓↓
1:5000	-	+ ↓↓↓
1:6000	-	-
1:7000	-	-

In the literature we found conflicting values for the sensitivity of RSID-semen test. Laffan and collaborators state that RSID-Semen test is able to detect Sg in semen diluted up to 1:500 which contrasts with a dilution of 1:2000 in the study done by Idris & Goodwin (Idris & Goodwin, 2022; Laffan et al., 2011). Also, in the present work, the RSID-Semen detected semen up to 1:750 for Serie B and 1:2000 for Serie E. These differences can be explained by the difference in the methods used to determine the sensitivity, whether the semen used was fresh or frozen (Boward & Wilson, 2013) and the different materials which are tested (Idris & Goodwin, 2022). For example, in the study conducted by Laffan and collaborators (Laffan et al., 2011), a cut from the swab with semen diluted with sterile distilled water was incubated with tris-buffered saline running buffer instead of RSID-Semen Universal buffer. The act of freezing and thawing semen is also known to decrease the detection of the Sg response, which explains why in Serie E semen was detected in lower concentrations (Boward & Wilson, 2013).

Comparing the results obtained within Serie B (*Table 12*) and Serie E (*Table 13*) for both methods, the differences in sensitivity are noticeable. In both Series, using Method 2 contributes to the detection of semen in less concentrations than in Method 1. Using the entire swab instead of a small cut enhances the sensitivity of the presumptive test. This is especially beneficial in situations where the sample is too scarce and the concentration of the body fluid to be tested is very low. In terms of presumptive testing, the use of Method 2 for processing vaginal swabs has proved to be more advantageous.

4.3.2. Postcoital interval (PCI)

The detection interval of RSID-Semen was also tested on PCI samples. As represented in *Table 14* and *Table 15*, Sg was detected until 24h and 48h using Method 1 in Serie C and Serie D, respectively. According to Method 2 the result was positive up until two days after unprotected vaginal intercourse for both Series. Our expectation was that in both Series Sg would be detected for a longer period of time after intercourse in Method 2, which matches the results obtained. By using the entire swab to do RSID-Semen instead of a cut, a greater amount of semen is tested, allowing detection for a longer period of time.

A study conducted by Old and co-authors (Old et al., 2012) showed that RSID-Semen test can detect semen up to two days after unprotected intercourse and Suttipasit & Wongwittayapanich (Suttipasit & Wongwittayapanich, 2018) states that Sg can be detected up to 72h. According to the results in the study by Laffan et al.(Laffan et al., 2011), at the 48h mark it is possible to get a negative result. There are discrepancies between our results and the ones found in the literature. These might be because the samples are: 1) from different donors; 2) the concentration of semen is not the same even though PCI is; and 3) the protocols followed might have slight differences.

Table 14 - Results of the comparison of Method 1 and Method 2 for Serie C samples. The symbol ↓ means that the intensity of the positive signal is low.

Post-coital interval	**Method 1**	**Method 2**
4h	+	+
12h	+	+
24h	+ ↓	+
36h	-	+ ↓
48h	-	+ ↓
60h	-	-
72h	-	-

Table 15 - Results of the comparison of Method 1 and Method 2 for Serie D samples. Arrows represent the intensity of the positive signals observed: low ↓ and very, very low ↓↓↓.

Serie D	Method 1	Method 2
8h	+	+
12h	+	+
24h	+	+
36h	+↓↓↓	+↓
48h	-	+↓↓↓

4.3.3. DNA extraction and quantification

To determine if extracting DNA from the entire swab (Method 2) is beneficial comparing with the extraction of only the remaining part of the swab (Method 1), DNA must be quantified to determine its concentration. Also, DNA quantification is key to determine the correct volume of DNA extract to be used in PCR (Dawnay & Sheppard, 2023).

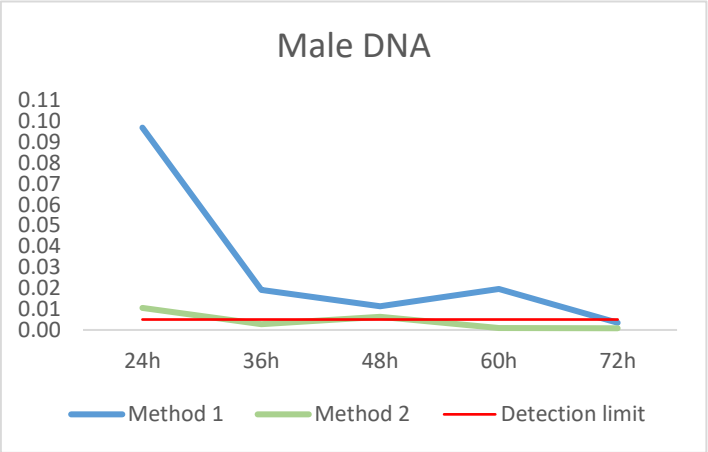
Table 16 shows the results of DNA quantification of Serie C samples extracted using standard extraction, as well as the male:female DNA ratio. Male DNA (Y-chromosomal DNA) concentration is quantified to confirm the presence of male DNA and analyse the amounts in which it is present. In Method 1 male DNA ranges from 0.0035 ng/μL to 0.0970 ng/μL and in Method 2 from 0.0008 ng/μL to 0.0062 ng/μL. Having into account the ideal DNA input (0.2 ng/μL), the concentration of male DNA was very low. Quantification values of Y chromosome DNA is below the limit of detection (0.005 ng/μL) in three samples in Method 2 in contrast to just one in Method 1 (*Graph 1*). As expected, female DNA is in higher concentrations in both methods.

Table 16 - Quantification of male and female DNA (ng/μL) and ratio between both genders obtained from Serie C samples and extracted with standard extraction procedure.

Samples	Method 1			Method 2		
	Female DNA	Male DNA	Male:Female ratio	Female DNA	Male DNA	Male:Female ratio
24h	86.2071	0.0970	1:889	4.6188	0.0106	1:437
36h	123.0522	0.0191	1:6439	5.1292	0.0028	1:1865
48h	114.7947	0.0114	1:10060	14.4478	0.0062	1:2324
60h	151.0755	0.0196	1:7699	9.8479	0.0010	1:10344
72h	61.4266	0.0035	1:17799	5.1630	0.0008	1:6429

In male:female mixtures, male autosomal DNA cannot be detected when analysing autosomal STRs if present below 1:20 male:female ratio, which is much higher than what is usually found in PCI vaginal swabs of rape victims (Vuichard et al., 2011). This happens due to competition for the primers during STR amplification which leads to preferential amplification of female DNA (Vuichard et al., 2011). As shown in *Table 16*, every sample has male:female DNA ratio disproportionately low. In Method 1, the ratio ranged from 1:889 to 1:17799 which corresponds to 24h PCI and 72h PCI, respectively. In Method 2, it ranged from 1:437 to 1:10344, corresponding to 24h PCI and 60h PCI, respectively. Although in both methods the amount of male DNA is too low, in Method 2, the difference between the concentrations of male and female DNA is not as high.

Even though Method 2 showed better results for male:female DNA ratio, the concentration of male DNA extracted was very low. As shown in *Graph 2*, the overall male DNA quantification results for Method 1 samples were better than in Method 2. Since using Method 2 the entire sample is depleted, we tried to extract DNA from the RSID-Semen cassettes, however this approach did not show satisfactory results so they will not be analysed in this work. Research shows that RSID buffer can undergo DNA extraction, so poor results cannot be explained by the incubation and subsequent extraction of the samples with the buffer (Dawnay & Sheppard, 2023). The aliquot used for RSID-Semen after incubating the entire swab with the buffer may have carried with it too much of the DNA present in the sample, leaving low to no DNA for extraction.

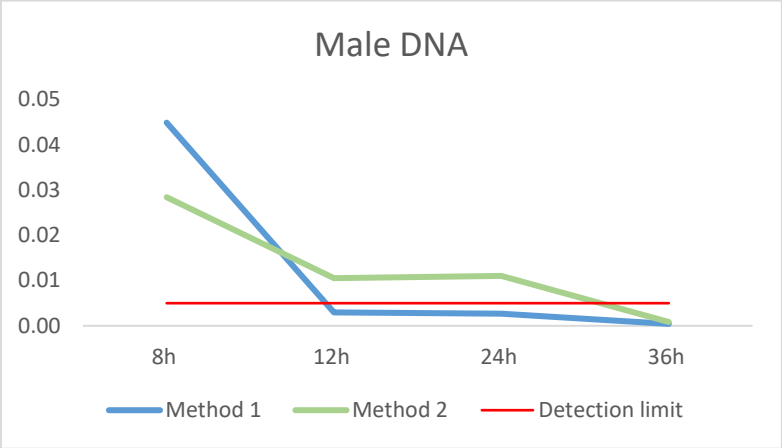


Graph 2 - Comparison of male DNA concentration of Serie C samples for Method 1 and Method 2. The red line is representative of the detection limit.

To try to improve DNA extraction, we performed differential lysis with an extra centrifugation step before the RSID-Semen testing. Centrifuging before RSID-Semen enables the separation of the proteins and DNA. Protein supernatant was retained for RSID-Semen testing and DNA was pelleted and used for DNA extraction (Ng et al., 2017). In *Table 17* are displayed the results obtained for the DNA quantification of Serie D samples and male:female ratio estimates for both methods. In Method 1, the concentration of male DNA ranges from 0.0004 ng/μL at 36h PCI to 0.0449 ng/μL at 8h PCI. In Method 2, the concentration varies from 0.0008 ng/μL and 0.0284 ng/μL at 36h PCI and 8h PCI, respectively. In Method 1, male DNA is present at a value above the detection limit (0.005 ng/μL) in only one sample (8h PCI), contrasting with Method 2 where values are higher in 3 out of 4 samples *Graph 3*.

Table 17 - DNA quantification results of Serie D differential extraction and male:female DNA ratios for Method 1 and Method 2. SF= Sperm Fraction.

Samples	**Method 1**			**Method 2**		
	Male DNA	**Female DNA**	**Male:Female ratio**	**Male DNA**	**Female DNA**	**Male:Female ratio**
8h SF	0.0449	0.0319	1:0.7111	0.0284	0.0235	1:0.8261
12h SF	0.0030	0.0100	1:3.3659	0.0106	0.0381	1:3.6096
24h SF	0.0027	0.0190	1:7.0157	0.0110	0.0227	1:2.0670
36h SF	0.0004	0.0567	1:134.3182	0.0008	0.0301	1:35.5625



Graph 3 - Comparison of male DNA concentration of SF of Serie D samples for Method 1 and Method 2. The red line is representative of the limit of detection.

Ideally, in a fully efficient differential extraction process, the concentration of female DNA in the sperm fraction (SF) should be null. However, as one can see in *Table 17*, not only is it not zero, it is higher than the concentration of male DNA at 12h PCI, 24 PCI and 36h PCI for both methods. The concentration of female DNA in the SF remains more constant over time in contrast to the concentration of male DNA that decreases. However, with the exception of 36h PCI, all samples have a male-to-female DNA ratio within the limits suggested by Vuichard et al. to detect the male component of the mixture (Vuichard et al., 2011). Comparing both methods, the male:female DNA ratio is higher in Method 2. Having into account that female DNA is usually in concentrations high enough to mask male DNA, the higher male:female DNA ratio, the higher the probability of obtaining a full male autosomal DNA profile.

With these results one can conclude that performing differential lysis with the extra centrifugation step and using the entire swab from the beginning increases the concentration male DNA recovered from the sample and, consequently, reduces male:female DNA ratio. As shown in *Graph 3*, the overall DNA quantification results are better for Method 2 than for Method 1.

4.3.4. Genetic profiles

Obtaining a male autosomal profile in cases of sexual assault between a female victim and a male perpetrator is of utmost importance since autosomal STRs allow for individual identification, enabling identification of the assailant (Kowalczyk et al., 2018). In most cases, as female DNA is in overly high concentrations it will mask male DNA, thereby preventing the amplification of male DNA. In some cases, a few male alleles are amplified, allowing a partial male autosomal genetic profile to be obtained. Partial profiles

are the ones in which alleles are missing in any locus. Full genetic profiles are the ones which all alleles amplify at a locus included in the amplification. To categorize the profiles obtained, they were compared with the reference male and female DNA genetic profiles. Serie C samples were not amplified due to the extremely low male:female DNA ratio.

Table 18 shows the number of specific male markers amplified of SF. The markers counted were those where at least one allele is present solely in the reference profile of the male individual. Markers with both alleles shared with the female individual were not taken into account. Of the samples processed with Method 1, only from one (12h PCI) was obtained a full male autosomal genetic profile. In Method 2, the same was observed for sample 24h PCI (*Table 18*). In Method 1, the number of male markers amplified decreases as PCI interval increases. In Method 2, there is a smooth increase from 12h PCI to 24h PCI, in the latter all specific markers were amplified (*Table 18*).

Table 18 - Description of the autosomal-STRs obtained from de SF using both methods of Serie D.

Method	**Post-Coital interval**	**Profile**	**Number of male specific markers amplified**
1	12h	Mixture with full male profile	21
	24h	Mixture with partial male profile	18
	36h	Mixture with partial male profile	5
2	12h	Mixture with partial male profile	18
	24h	Mixture with full male profile	21
	36h	Mixture with partial male profile	1

To analyse the completeness (in this study refers to a “complete genetic profile”) of each autosomal profile, we calculated the presence of specific alleles for each contributor. Once again, markers with both alleles shared between both individuals were not included. According to *Table 19*, in Method 1, at 12h PCI was obtained a mixture with a full male and female autosomal profile. In Method 2, the same result was obtained at 24h PCI. In Method 1, at 24h PCI and in Method 2 at 12h PCI was obtained a mixture with full female and partial male autosomal genetic profile. At 36h PCI a mixture with male and female partial profiles was obtained. In Method 1 20% of male specific alleles were amplified and only 4% in Method 2. Considering that, at 36h PCI, the male:female DNA ratio is below 1:20 in both methods, as one can see in *Table 17*, the male profile obtained was expected to be very incomplete. In Method 1, as expected, the number of samples with complete DNA profiles decrease as PCI increases. For Method 2, there is

a slight increase in the profile completeness between 12h and 24h PCI interval, which matches with the increase of the male:female DNA ratio (*Table 17*). However, this increase could be a one-time observation, thus needing further information.

Table 19 – DNA profile completeness of autosomal- STRs of the samples submitted for capillary electrophoresis.

Method	Samples	Total (%)	Male (%)	Female (%)
1	12h	100	100	100
	24h	92	80	100
	36h	68	20	95
2	12h	94	84	100
	24h	100	100	100
	36h	62	4	95

Usually, a DNA profile mixture is detected when two or more peaks are present at a locus in the electropherogram. However, caution must be taken because multiple alleles at a locus might also be due to artefacts (Torres et al., 2003). An example of a two-donor profile DNA mixture is shown in *Figure 17*.

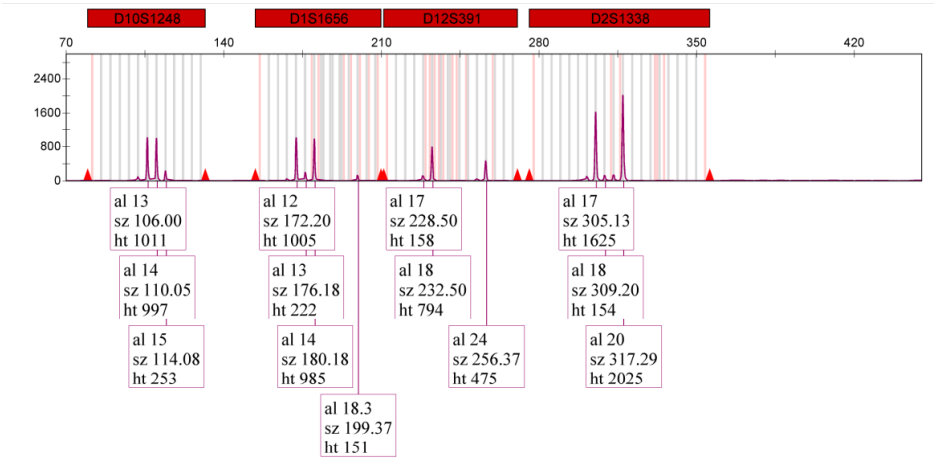


Figure 17 - A mixed autosomal genetic profile using GlobalFiler amplification kit.

Most often, sexual assault scenarios take place between a male individual (abuser) and a female individual (victim). In these situations, the major component belongs to the victim which can and will most likely mask the perpetrator's genetic profile. In such cases, it is recommended to do Y-STR analysis. Y-STRs are proven to successfully generate a profile in male/female DNA mixture in a ratio of 1:2.000, in azoospermic and/or vasectomised men and helps determine the number of male contributors in multiple rape cases (Gusmão et al., 2008). Even though Y-STR markers are very helpful and recommended in a mixture of male and female DNA, they do not determine the identity of the perpetuator. This is due to the fact that all males of the same paternal lineage share the same Y-STR genetic profile, however, Y-STR profiling can strongly support the investigation (Garvin et al., 2012).

According to *Table 20*, in Method 1 approximately 69% of the alleles were amplified at 12h PCI opposing to 96% with Method 2. At 24h PCI the difference is even bigger with 46% in Method 1 and 100% in Method 2. 36h PCI is an exception because 27% of the alleles were amplified with Method 1 and with Method 2 no allele was amplified. Overall, more alleles were amplified in samples processed with Method 2 than in Method 1, and it was even possible to obtain a complete Y-STR profile. At 36h PCI DNA quantification of the Y-chromosome was below the limit of detection for both methods which is in accordance with the amplification of few or almost no genetic markers (*Table 20*).

Table 20 - Profile completeness of Y-STRs of the SF of Serie D.

Method	Samples	Profile completeness (%)
1	12h	69
	24h	46
	36h	27
2	12h	96
	24h	100
	36h	0

Peak height is proportional to the amount of amplified DNA. The more DNA there is in the sample, the higher the peaks are (Kumar et al., 2018).

Table 21 - Mean peak heights of Y-Chromosomal DNA of SF. Mean peak heights are in RFU's (Relative Fluorescence Units).

Method	Post-coital interval	Mean peak heights (RFU)
1	12h	192
	24h	176
	36h	140
2	12h	485
	24h	408
	36h	0

The obtained profiles showed that as the concentration of male DNA decreases, peaks also decrease, as shown in *Table 21*. However, the difference in the average peak values (Relative Fluorescence Unit – RFU) between the two methods is evident. In Method 1, at 12h PCI the average peak height is 192 RFU comparing with 485 RFU in Method 2. The same tendency is observed for the 24h PCI with average heights of 176 RFU and 408 RFU for Method 1 and Method 2, respectively. At 36h PCI no alleles were amplified in Method 2 so no peaks were obtained. Given that the tendency is to get more peaks with method 2 samples, the sample at 36h PCI must have been a one-time failure. In Method 1, the average height is the lowest corresponding to 140 RFU. Once again, excluding 36h PCI, overall results are better for Method 2 than for Method 1.

Overall, the results for the new method to process vaginal swabs showed inconclusive results. On the one hand, this method loses the ability to amplify autosomal STRs comparing with Method 1 since the completeness of the genetic profile is better in the samples processed with Method 1. On the other hand, it gains in the amplification of STRs of the Y chromosome. With Method 2 it was possible to amplify all male alleles in one of the samples, which did not happen in any of the samples of Method 1. Although Method 2 shows potential, further analysis are required to draw strong conclusions.

5. Conclusions

The results of the present study demonstrate that:

1. Presumptive tests used in closed cases received by the laboratory between the year of 2019 and 2022 showed that out of 44 cases, 30 showed concordant results by Seratec PSA Semiquant and RSID-Semen. However, 11 cases showed positive results for Seratec PSA Semiquant and negative for RSID-Semen and 3 showed the opposite. The observed differences between methodologies demonstrates that in some cases this may lead to false negatives regarding “semen detection” in forensic case work depending on the method chosen.

2. Comparing the detection interval of Seratec PSA Semiquant and RSID-Semen, using post-coital interval samples, showed that RSID-Semen detects the correspondent protein for a longer period of time.

3. A new assay for processing vaginal swabs was tested (Method 2) and compared with the method used in the laboratory (Method 1). For Method 1 a piece of swab was cut and incubated with RSID Universal buffer and the rest of the swab was kept for DNA extraction. For Method 2 the entire swab was incubated and RSID Universal buffer and then used for DNA extraction. Results support that:

3.1. RSID-semen sensitivity tested on frozen and fresh semen and post-coital samples was higher using Method 2 for all samples. However, with Method 2 caution must be taken when samples are very concentrated due to high-dose hook effect.

3.2. Extracting DNA with standard extraction showed better quantification results for Method 1 than Method 2. Extracting DNA with differential extraction showed the opposite results. Differential extraction showed overall more male DNA concentration and consequently higher male-to-female DNA ratio.

3.3. Genetic profiles of the sperm fraction showed that samples processed with Method 1 showed more amplified alleles on two out of three samples for autosomal STRs, contrasting with just one sample in Method 2. For Y-STRs,

using Method 2 we were able to amplify more alleles and obtain a more complete profile in two out of three samples.

3.4. The data generated in this study are preliminary results and therefore further studies are needed so that the method of processing vaginal swabs is as efficient as possible for cases of sexual assault where, in particular, samples obtained may contain little DNA.

6. References

Albani, P. P., & Fleming, R. (2018). Novel messenger RNAs for body fluid identification. *Science & Justice*, *58*(2), 145–152. https://doi.org/10.1016/j.scijus.2017.09.002

An, J. H., Choi, A., Shin, K.-J., Yang, W. I., & Lee, H. Y. (2013). DNA methylation-specific multiplex assays for body fluid identification. *International Journal of Legal Medicine*, *127*(1), 35–43. https://doi.org/10.1007/s00414-012-0719-1

An, J.-H., Shin, K.-J., Yang, W.-I., & Lee, H.-Y. (2012). Body fluid identification in forensics. *BMB Reports*, *45*(10), 545–553. https://doi.org/10.5483/BMBRep.2012.45.10.206

APAV. *APAV, Estatísticas APAV Crimes Sexuais [2013—2018]*. https://apav.pt/apav_v3/index.php/pt/estatisticas-apav

Applied Biosystems. (2018). *Quantifiler™ HP and Trio DNA Quantification kits: User Guide*. *www.thermofisher.com*

Applied Biosystems. (2019a). *GlobalFiler™ and GlobalFiler™ IQC PCR Amplification Kits: User Guide*. *www.thermofisher.com*

Applied Biosystems. (2019b). *Yfiler™ Plus PCR Amplification Kit: User Guide*. *www.thermofisher.com*

Bell, S. (Ed.). (2019). *Forensic science: An introduction to scientific and investigative techniques* (5th edition). CRC Press, Taylor & Franics Group.

Berrena, Taylor, T. L. B. (2015). *Persistence of seminal constituents in vaginal drainage samples*. Boston University.

Bexon KJ. (2017). *Forensic MicroRNA Analysis of Body Fluids Relating to Sexual Assaults* [Doctoral Thesis]. University of Huddersfield.

Bogas, Vanessa. (2013). *Métodos de extração de ADN de amostras biológicas expostas a diferentes condições—Estudo comparativo*. Faculdade de Medicina da Universidade do Porto.

Boward, E. S., & Wilson, S. L. (2013). A comparison of ABAcard® p30 and RSID™-Semen test kits for forensic semen identification. *Journal of Forensic and Legal Medicine*, *20*(8), 1126–1130. https://doi.org/10.1016/j.jflm.2013.09.007

Butler. (2012). *Advanced Topics in Forensic DNA Typing: Methodology*. Elsevier. https://doi.org/10.1016/C2011-0-04189-3

Butler, J. M. (2005). *Forensic DNA typing: Biology, technology, and genetics of STR markers* (2nd ed). Elsevier Academic Press.

Butler, J. M. (2010). *Fundamentals of forensic DNA typing*. Academic Press/Elsevier.

Choi, A., Shin, K.-J., Yang, W. I., & Lee, H. Y. (2014). Body fluid identification by integrated analysis of DNA methylation and body fluid-specific microbial DNA. *International Journal of Legal Medicine*, *128*(1), 33–41. https://doi.org/10.1007/s00414-013-0918-4

Cunha, Mariana. (2019). *Análise de misturas de DNA no âmbito de perícias criminais: Estudo comparativo da lise diferencial em Agressões Sexuais* [Faculdade de Ciências da Universidade do Porto]. https://hdl.handle.net/10216/125688

Dawnay, N., & Sheppard, K. (2023). From crime scene to courtroom: A review of the current bioanalytical evidence workflows used in rape and sexual assault investigations in the United Kingdom. *Science & Justice*, *63*(2), 206–228. https://doi.org/10.1016/j.scijus.2022.12.006

de Beijer, R. P., de Graaf, C., van Weert, A., van Leeuwen, T. G., Aalders, M. C. G., & van Dam, A. (2018). Identification and detection of protein markers to differentiate between forensically relevant body fluids. *Forensic Science International*, *290*, 196–206. https://doi.org/10.1016/j.forsciint.2018.07.013

DeVore, H. K., & Sachs, C. J. (2011). Sexual Assault. *Emergency Medicine Clinics of North America*, *29*(3), 605–620. https://doi.org/10.1016/j.emc.2011.04.012

DiFrancesco, J., & Sutton, J. (2015). Re-evaluation of the Seratec® PSA Semiquant test: Comparison of kit provided buffer with phosphate buffer saline. *Canadian Society of Forensic Science Journal*, *48*(3), 137–151. https://doi.org/10.1080/00085030.2015.1051315

Du Mont, J., World Health Organization, White, D., & Sexual Violence Research Initiative. (2007). *The uses and impacts of medico-legal evidence in sexual assault cases: A global review / Janice Du Mont, Deborah White*. 80.

Dworkin, E. R., Krahé, B., & Zinzow, H. (2021). The global prevalence of sexual assault: A systematic review of international research since 2010. *Psychology of Violence*, *11*(5), 497–508. https://doi.org/10.1037/vio0000374

Garvin, A. M., Fischer, A., Schnee-Griese, J., Jelinski, A., Bottinelli, M., Soldati, G., Tubio, M., Castella, V., Monney, N., Malik, N., & Madrid, M. (2012). Isolating DNA from sexual assault cases: A comparison of standard methods with a nuclease-based approach. *Investigative Genetics*, *3*(1), 25. https://doi.org/10.1186/2041-2223-3-25

Ghai, M., Naidoo, N., Evans, D. L., & Kader, F. (2020). Identification of novel semen and saliva specific methylation markers and its potential application in forensic analysis. *Forensic Science International: Genetics*, *49*, 102392. https://doi.org/10.1016/j.fsigen.2020.102392

Gill, P., Jeffreys, A. J., & Werrett, D. J. (1985). Forensic application of DNA 'fingerprints'. *Nature*, *318*(6046), 577–579. https://doi.org/10.1038/318577a0

Goodwin, W., Linacre, A., & Hadi, S. (2007). *An introduction to forensic genetics*. John Wiley & Sons.

Gopinath, S., Zhong, C., Nguyen, V., Ge, J., Lagacé, R. E., Short, M. L., & Mulero, J. J. (2016). Developmental validation of the Yfiler® Plus PCR Amplification Kit: An enhanced Y-STR multiplex for casework and database applications. *Forensic*

Science International: Genetics, *24*, 164–175. https://doi.org/10.1016/j.fsigen.2016.07.006

Gouveia, Nair. (2015). *Validação e implementação do kit Quantifiler Trio DNA Quantification em amostras forenses*. Universidade de Lisboa.

Graves, H. C. B., Sensabaugh, G. F., & Blake, E. T. (1985). Postcoital Detection of a Male-Specific Semen Protein: Application to the Investigation of Rape. *New England Journal of Medicine*, *312*(6), 338–343. https://doi.org/10.1056/NEJM198502073120603

Gusmão, L., Brión, M., & Gomes, I. (2008). Chapter 30 The human Y chromosome male-specific polymorphisms and forensic genetics. Em *Handbook of Analytical Separations* (Vol. 6, pp. 969–1000). Elsevier. https://doi.org/10.1016/S1567-7192(06)06034-7

Hanson, E., Ingold, S., Haas, C., & Ballantyne, J. (2018). *Messenger RNA biomarker signatures for forensic body fluid identification revealed by targeted RNA sequencing* [Preprint]. Genomics. https://doi.org/10.1101/247312

Harbison, S., & Fleming, R. (2016). Forensic body fluid identification: State of the art. *Research and Reports in Forensic Medical Science*, 11. https://doi.org/10.2147/RRFMS.S57994

Hochmeister, M. N., Budowle, B., Rudin, O., Gehrig, C., Borer, U., Thali, M., & Dirnhofer, R. (1999). Evaluation of prostate-specific antigen (PSA) membrane test assays for the forensic identification of seminal fluid. *Journal of Forensic Sciences*, *44*(5), 1057–1060.

Holt, A., Wootton, S. C., Mulero, J. J., Brzoska, P. M., Langit, E., & Green, R. L. (2016). Developmental validation of the Quantifiler® HP and Trio Kits for human DNA quantification in forensic samples. *Forensic Science International: Genetics*, *21*, 145–157. https://doi.org/10.1016/j.fsigen.2015.12.007

Idris, B., & Goodwin, W. H. (2022). Evaluating the sensitivity of presumptive and confirmatory tests for body fluids. *Forensic Science International: Genetics Supplement Series*, *8*, 276–278. https://doi.org/10.1016/j.fsigss.2022.10.061

Jakubowska, J., Maciejewska, A., Pawłowski, R., & Bielawski, K. P. (2013). MRNA profiling for vaginal fluid and menstrual blood identification. *Forensic Science International: Genetics*, *7*(2), 272–278. https://doi.org/10.1016/j.fsigen.2012.11.005

Jamshidi, R., Penman-Aguilar, A., Wiener, J., Gallo, M. F., Zenilman, J. M., Melendez, J. H., Snead, M., Black, C. M., Jamieson, D. J., & Macaluso, M. (2013). Detection of two biological markers of intercourse: Prostate-specific antigen and Y-chromosomal DNA. *Contraception*, *88*(6), 749–757. https://doi.org/10.1016/j.contraception.2013.08.003

Jankova, R., Jakovski, Z., & Janevski, R. (2019). Differential extraction method as a golden standard in analyzing semen stains in sexual-assault cases. *Forensic Science International: Genetics Supplement Series*, *7*(1), 838–840. https://doi.org/10.1016/j.fsigss.2019.10.196

Jeffreys, A. J., Wilson, V., & Thein, S. L. (1985). Hypervariable 'minisatellite' regions in human DNA. *Nature*, *314*(6006), 67–73. https://doi.org/10.1038/314067a0

Joshi, M., & Deshpande, J. D. (2011). POLYMERASE CHAIN REACTION: METHODS, PRINCIPLES AND APPLICATION. *International Journal of Biomedical Research*, *2*(1), 81–97. https://doi.org/10.7439/ijbr.v2i1.83

Kimberley Ellen Wing. (2021). *The evaluation of semenogelin proteins I and II in semen and other body fluids of males and females for forensic detection.* [Murdoch University]. http://researchrepository.murdoch.edu.au/id/eprint/63413

Kishbaugh, J. M., Cielski, S., Fotusky, A., Lighthart, S., Maguire, K., Quarino, L., & Conte, J. (2019). Detection of prostate specific antigen and salivary amylase in vaginal swabs using SERATEC® immunochromatographic assays. *Forensic Science International*, *304*, 109899. https://doi.org/10.1016/j.forsciint.2019.109899

Kowalczyk, M., Zawadzka, E., Szewczuk, D., Gryzińska, M., & Jakubczak, A. (2018). Molecular markers used in forensic genetics. *Medicine, Science and the Law*, *58*(4), 201–209. https://doi.org/10.1177/0025802418803852

Kumar, N., Maitray, A., Gupta, R., Sharma, D., & Sk, S. (2018). Importance of Y- STR profiling in sexual assault cases with mixed DNA profile. *International Journal of Molecular Biology*, *3*(1). https://doi.org/10.15406/ijmboa.2018.03.00048

Laffan, Á., Sawyer, I., Quinones, I., & Daniel, B. (2011). Evaluation of semen presumptive tests for use at crime scenes. *Medicine, Science and the Law*, *51*(1), 11–17. https://doi.org/10.1258/msl.2010.010040

Laux, D. L., & Barnhart, J. P. (2011). Validation of the Seratec ® SeraQuant™ for the Quantitation of Prostate-Specific Antigen Levels on Immunochromatographic Membranes. *Journal of Forensic Sciences*, *56*(6), 1574–1579. https://doi.org/10.1111/j.1556-4029.2011.01893.x

Lee, H. Y., Park, M. J., Choi, A., An, J. H., Yang, W. I., & Shin, K.-J. (2012). Potential forensic application of DNA methylation profiling to body fluid identification. *International Journal of Legal Medicine*, *126*(1), 55–62. https://doi.org/10.1007/s00414-011-0569-2

Li, R. (2021). *Forensic biology* (Second edition). CRC Press, Taylor & Francis Group.

Lilja, H., Abrahamsson, P. A., & Lundwall, A. (1989). Semenogelin, the predominant protein in human semen. Primary structure and identification of closely related proteins in the male accessory sex glands and on the spermatozoa. *The Journal of Biological Chemistry*, *264*(3), 1894–1900.

Lindenbergh, A., Maaskant, P., & Sijen, T. (2013). Implementation of RNA profiling in forensic casework. *Forensic Science International: Genetics*, *7*(1), 159–166. https://doi.org/10.1016/j.fsigen.2012.09.003

Liu, J. Y., Zhong, C., Holt, A., Lagace, R., Harrold, M., Dixon, A. B., Brevnov, M. G., Shewale, J. G., & Hennessy, L. K. (2012). AutoMate Express™ Forensic DNA Extraction System for the Extraction of Genomic DNA from Biological Samples*:

AUTOMATE EXPRESS™ FORENSIC DNA EXTRACTION SYSTEM. *Journal of Forensic Sciences*, *57*(4), 1022–1030. https://doi.org/10.1111/j.1556-4029.2012.02084.x

Luce, H., Schrager, S., & Gilchrist, V. (2010). Sexual assault of women. *American Family Physician*, *81*(4), 489–495.

Ludeman, M. J., Zhong, C., Mulero, J. J., Lagacé, R. E., Hennessy, L. K., Short, M. L., & Wang, D. Y. (2018). Developmental validation of GlobalFiler™ PCR amplification kit: A 6-dye multiplex assay designed for amplification of casework samples. *International Journal of Legal Medicine*, *132*(6), 1555–1573. https://doi.org/10.1007/s00414-018-1817-5

Lynch, C., & Fleming, R. (2021). RNA -based approaches for body fluid identification in forensic science. *WIREs Forensic Science*, *3*(4). https://doi.org/10.1002/wfs2.1407

Mannello, F., Condemi, L., Cardinali, A., Bianchi, G., & Gazzanelli, G. (1998). High Concentrations of Prostate-Specific Antigen in Urine of Women Receiving Oral Contraceptives. *Clinical Chemistry*, *44*(1), 181–183. https://doi.org/10.1093/clinchem/44.1.181

McKiernan, H. E., & Danielson, P. B. (2017). Molecular Diagnostic Applications in Forensic Science. Em *Molecular Diagnostics* (pp. 371–394). Elsevier. https://doi.org/10.1016/B978-0-12-802971-8.00021-3

Mollen, C. J., Goyal, M. K., & Frioux, S. M. (2012). Acute Sexual Assault: A Review. *Pediatric Emergency Care*, *28*(6), 584–590. https://doi.org/10.1097/PEC.0b013e318258bfea

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., & Erlich, H. (1986). Specific Enzymatic Amplification of DNA In Vitro: The Polymerase Chain Reaction. *Cold Spring Harbor Symposia on Quantitative Biology*, *51*(0), 263–273. https://doi.org/10.1101/SQB.1986.051.01.032

Murray, L. K., Nguyen, A., & Cohen, J. A. (2014). Child Sexual Abuse. *Child and Adolescent Psychiatric Clinics of North America*, *23*(2), 321–337. https://doi.org/10.1016/j.chc.2014.01.003

Nelson, D. G. (2002). An Alternate Light Source to Detect Semen. *Academic Emergency Medicine*, *9*(10), 1045–1048. https://doi.org/10.1197/aemj.9.10.1045

Ng, H. H., Lim, M.-L., Hoe, S. Y., Yong, Z. D., Ping, Y. S., Ang, H. C., & Syn, C. K. C. (2017). Modified differential DNA extraction to reduce processing time of sexual assault exhibits. *Forensic Science International: Genetics Supplement Series*, *6*, e252–e254. https://doi.org/10.1016/j.fsigss.2017.09.094

Old, J., Schweers, B. A., Boonlayangoor, P. W., Fischer, B., Miller, K. W. P., & Reich, K. (2012). Developmental Validation of RSID^{TM}-Semen: A Lateral Flow Immunochromatographic Strip Test for the Forensic Detection of Human Semen*: IMMUNOCHROMATOGRAPHIC STRIP TEST FOR THE FORENSIC DETECTION OF HUMAN SEMEN. *Journal of Forensic Sciences*, *57*(2), 489–499. https://doi.org/10.1111/j.1556-4029.2011.01968.x

Purps, J., Geppert, M., Nagy, M., & Roewer, L. (2015). Validation of a combined autosomal/Y-chromosomal STR approach for analyzing typical biological stains in sexual-assault cases. *Forensic Science International: Genetics*, *19*, 238–242. https://doi.org/10.1016/j.fsigen.2015.08.002

Sakelliadis, E. I., Spiliopoulou, C. A., & Papadodima, S. A. (2009). Forensic investigation of child victim with sexual abuse. *Indian Pediatrics*, *46*(2), 144–151.

Schmidt, S., Franke, M., Lehmann, J., Loch, T., Stöckle, M., & Weichert-Jacobsen, K. (2001). Prostate-specific antigen in female urine: A prospective study involving 217 women. *Urology*, *57*(4), 717–720. https://doi.org/10.1016/s0090-4295(00)01093-1

Sijen, T., & Harbison, S. (2021). On the Identification of Body Fluids and Tissues: A Crucial Link in the Investigation and Solution of Crime. *Genes*, *12*(11), 1728. https://doi.org/10.3390/genes12111728

Suttipasit, P., & Wongwittayapanich, S. (2018). Detection of prostate specific antigen and semenogelin in specimens from female rape victims. *Journal of Forensic and Legal Medicine*, *54*, 102–108. https://doi.org/10.1016/j.jflm.2017.12.017

Tian, H., Huang, S., Bai, P., Xiao, X., Peng, D., Zhao, H., Liu, Y., Feng, Q., Liao, M., Li, F., & Liang, W. (2021). The effect of infertile semen on the mRNA-based body fluid identification. *ELECTROPHORESIS*, *42*(16), 1614–1622. https://doi.org/10.1002/elps.202000238

Torres, Y., Flores, I., Prieto, V., López-Soto, M., Farfán, M. J., Carracedo, A., & Sanz, P. (2003). DNA mixtures in forensic casework: A 4-year retrospective study. *Forensic Science International*, *134*(2–3), 180–186. https://doi.org/10.1016/S0379-0738(03)00161-0

Virkler, K., & Lednev, I. K. (2009a). Analysis of body fluids for forensic purposes: From laboratory testing to non-destructive rapid confirmatory identification at a crime scene. *Forensic Science International*, *188*(1–3), 1–17. https://doi.org/10.1016/j.forsciint.2009.02.013

Virkler, K., & Lednev, I. K. (2009b). Raman spectroscopic signature of semen and its potential application to forensic body fluid identification. *Forensic Science International*, *193*(1–3), 56–62. https://doi.org/10.1016/j.forsciint.2009.09.005

Vuichard, S., Borer, U., Bottinelli, M., Cossu, C., Malik, N., Meier, V., Gehrig, C., Sulzer, A., Morerod, M.-L., & Castella, V. (2011). Differential DNA extraction of challenging simulated sexual-assault samples: A Swiss collaborative study. *Investigative Genetics*, *2*(1), 11. https://doi.org/10.1186/2041-2223-2-11

Wolff, J. M., Heyl, W., Borchers, H., Brehmer Jr., B., Rath, W., & Jakse, G. (1999). Detection of Prostate-Specific Antigen Immunoreactivity in Amniotic Fluid. *Urologia Internationalis*, *62*(3), 139–142. https://doi.org/10.1159/000030377

World Health Organization. (2013). *Global and regional estimates of violence against women: Prevalence and health effects of intimate partner violence and non-partner sexual violence*. World Health Organization. https://apps.who.int/iris/handle/10665/85239

Yu, H., & Diamandis, E. P. (1995). Prostate-specific antigen in milk of lactating women. *Clinical Chemistry*, *41*(1), 54–58.

Zubakov, D., Boersma, A. W. M., Choi, Y., van Kuijk, P. F., Wiemer, E. A. C., & Kayser, M. (2010). MicroRNA markers for forensic body fluid identification obtained from microarray screening and quantitative RT-PCR confirmation. *International Journal of Legal Medicine*, *124*(3), 217–226. https://doi.org/10.1007/s00414-009-0402-3