U.PORTO
FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

MSc

2.º CICLO

FCUP
LPC-PJ
2016

U.PORTO

Evaluation and validation of retrotransposons-based kits for DNA analysis of degraded biological evidence and rootless hair

Cátia Liane Teixeira Martins

FC

U.PORTO
FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

Evaluation and validation of retrotransposons-based kits for DNA analysis of degraded biological evidence and rootless hair

Cátia Liane Teixeira Martins

Dissertação de Mestrado apresentada à
Faculdade de Ciências da Universidade do Porto, Laboratório de
Polícia Cientifica da Polícia Judiciária
Genética Forense

2016

Evaluation and validation of retrotransposons-based kits for DNA analysis of degraded biological evidence and rootless hair

Cátia Liane Teixeira Martins

Mestrado em Genética Forense

Departamento de Biologia

2016

Orientador

Maria Manuela Martins Oliveira, Professora Auxiliar Convidada

i3S-Instituto de Investigação e Inovação em Saúde, Universidade do Porto

FCUP-Faculdade de Ciências da Universidade do Porto

Coorientador

Luísa Azevedo, Professora Auxiliar Convidada

i3S-Instituto de Investigação e Inovação em Saúde, Universidade do Porto

FCUP-Faculdade de Ciências da Universidade do Porto

Orientador Externo

Paulo Miguel Ferreira, Especialista Superior

Laboratório de Polícia Cientifica da Polícia Judiciária

Todas as correções determinadas pelo júri, e só essas, foram efetuadas.

O Presidente do Júri,

Porto, _____/_____/_____

Agradecimentos

Com o culminar de mais uma etapa da minha vida que me proporcionou mais conhecimento e que me enriqueceu, não só do ponto de vista académico mas também pessoal, é imprescindível agradecer às pessoas e instituições que, de diferentes formas me ajudaram na realização deste trabalho.

Começo por agradecer à minha irmã, Andreia, pela força de vontade que me impingiu desde nova e por me proporcionar as condições necessárias para a concretização deste meu objetivo.

Gostaria também de agradecer aos meus pais, em especial à minha mãe Cândida por toda a sua dedicação e sacrifício. Sei que nem sempre foi fácil, e por tudo, o meu muito obrigada mãe.

Quero agradecer à minha orientadora Manuela Oliveira, pela orientação científica deste trabalho e pela disponibilidade e dedicação que me dispensou na realização da redação da presente dissertação. Agradeço a transmissão e partilha de mais e novos conhecimentos, que em muito contribuíram para a minha formação e que, acredito continuarão a ser valiosos no futuro, por ter tido sempre uma palavra amiga e otimista para me dar, e por toda a confiança que depositou em mim e neste trabalho.

Agradeço ao meu orientador externo, Especialista Superior Paulo Miguel Ferreira do Laboratório de Polícia Cientifica da Polícia Judiciária, pela sugestão do tema, pela orientação prática deste trabalho, e pelos seus ensinamentos na área da Genética Forense que em muito contribuíram para a minha formação. Obrigada por me ter apresentado a “Genética Forense no mundo real”.

À minha co-orientadora Luísa Azevedo, pela ajuda e disponibilidade demonstrada durante a redação deste trabalho, expresso os meus mais sinceros agradecimentos.

Não posso deixar de apresentar o meu mais profundo agradecimento à chefe de área de Biotoxicologia do LPC-PJ, Sandra Cristina Costa, por todo o apoio que me prestou, pelos conhecimentos que me transmitiu, mostrando-se sempre disponível, e por todas as novas experiências que me proporcionou e que ficarão para sempre na minha memória.

Agradeço também à Raquel Carvalho, por todo o apoio prestado no decorrer do meu trabalho laboratorial, pela sua disponibilidade e paciência e pelos conhecimentos transmitidos.

Agradeço a todos os elementos da Especialidade Forense de Biologia do LPC-PJ, em especial à Ana Paula Cardoso, António Goulding, João Paulo Júlio, Isabel Ribeiro, Ana Paula Carvalho, Ana Marta Silva, Cláudia Broeiro, Rita Almeida, Maria de Lurdes Prisal, Sónia Cardoso e Maria do Carmo Lopes, por toda a simpatia e disponibilidade demonstrada durante a realização da parte prática do presente trabalho e por me terem feito sentir em casa mesmo estando distante.

Gostaria ainda de agradecer a algumas instituições e empresas envolvidas diretamente na realização deste trabalho. Ao Laboratório de Polícia Cientifica da Polícia Judiciária, na pessoa do Doutor Carlos Farinha, agradeço a confiança e a possibilidade da realização deste estágio. À empresa *InnoGenomics Technologies* nomeadamente ao Jonathan Tabak, Gina Pineda, Gabriela Hirschfeld Campolongo e Andrew Loftus, agradeço toda a disponibilidade e ajuda prestada no desenvolvimento prático deste trabalho. Agradeço ainda à Faculdade de Ciências da Universidade do Porto (FCUP), nomeadamente ao Coordenador do Mestrado em Genética Forense Professor António Amorim, ao Instituto de Patologia e Imunologia Molecular da Universidade do Porto (Ipatimup) e ao Instituto de Investigação e Inovação em Saúde da Universidade do Porto (i3S) que tão bem me receberam e acolheram.

Um especial agradecimento a todos os meus amigos, sobretudo à Letícia Fortes, Sara Abreu e Ana Almeida com quem tive a oportunidade de partilhar sete anos de momentos inesquecíveis e loucuras divertidas e a quem devo muito daquilo que sou hoje. Obrigada, meninas, por estarem sempre presentes em todos os momentos. Agradeço também a todos os outros que fizeram parte desta minha caminhada académica.

Por fim, e não menos importante, ao meu namorado Fábio Costa, a quem privei da minha atenção e companhia, e que mesmo assim sempre me transmitiu a força e a motivação necessária para continuar, principalmente nos momentos mais difíceis e em que tudo parecia estar errado, agradeço de coração pelo apoio, amizade, companheirismo, sinceridade e auxílio em todos os momentos deste trabalho e deste mestrado.

Acknowledgments

With the conclusion of another step of my life that has provided me with more knowledge and enriched both as a student and as a person, it is imperative to thank some people and institutions that somehow helped me to get this far.

I begin by thanking my sister, Andreia, for teaching me the virtues of willpower and endurance during my youth and for providing me all the necessary conditions that allowed me to achieve this aim.

I would also like to thank both my parents, but especially my mother Cândida, for all their dedication and sacrifice. I know it was not always easy... For everything, my deepest and sincerest gratitude, Mom.

I want to thank my academic supervisor, Manuela Oliveira, for the orientation of this work, the availability, and dedication during the preparation of this thesis. I appreciate all the knowledge transmitted and shared which significantly contributed to my training, and that will continue to be valuable in my future. Also, I would like to express my gratitude for always having a friendly and optimistic word to give, and for all the confidence invested both in me and my work.

I thank my external supervisor, Paulo Miguel Ferreira, from Laboratório de Polícia Cientifica da Polícia Judiciária, for suggesting the theme and the orientation of the practical component of this work, and his teaching in the field of Forensic Genetics which significantly contributed to my training. Thank you for introducing me to the "Forensic Genetics in the real world".

To my co-supervisor, Luísa Azevedo, for the help and availability provided during the preparation of this work, I express my most sincere thanks.

I have to present my deepest gratitude to the Coordinator of Biotoxicology area from LPC-PJ, Sandra Cristina Costa, for all her support and knowledge transmitted, for always being unconditionally available, and for all the new experiences that will be forever in my memory.

I also thank Raquel Carvalho, for all her support during my laboratory work, for her availability, patience, and for the knowledge transmitted.

I thank all the members from Forensic Specialty of Biology, especially Ana Paula Cardoso, Antonio Goulding, João Paulo Júlio, Isabel Ribeiro, Ana Paula Carvalho, Ana Marta Silva, Claudia Broeiro, Rita Almeida, Maria de Lurdes Prisal, Sónia Cardoso, and Maria do Carmo Lopes for all the friendship and helpfulness demonstrated in the course of practical component of this work and for making me feel at home.

I would also like to thank some institutions and companies directly involved in this work. To the Laboratório de Polícia Cientifica da Polícia Judiciária, personified by Carlos Farinha, thanks for the trust and for the possibility of realizing this internship. To the *InnoGenomics Technologies* Company, in the persons of Jonathan Tabak, Gina Pineda, Gabriela Hirschfeld Campolongo, and Andrew Loftus, thanks for all the availability and assistance provided in the practical development of this work. I am also grateful to the Faculdade de Ciências da Universidade do Porto (FCUP), personified by António Amorim – the coordinator of the Master Degree in Forensic Sciences, to the Instituto de Patologia e Imunologia Molecular da Universidade do Porto (Ipatimup), and to the Instituto de Investigação e Inovação em Saúde da Universidade do Porto (i3S).

Huge thanks to all my friends, especially Leticia Fortes, Sara Abreu, and Ana Almeida with who I had the opportunity to share seven years of unforgettable moments and fun, and to whom I owe much of what I am today. Thank you, girls, for always being present at all times. Thanks also to all others who were part of this my academic journey.

Finally, but not least, to my boyfriend Fábio Costa, who was deprived of my attention and company, and to nonetheless always gave me the strength and motivation to continue, especially in difficult times and when everything seemed to be wrong. From the bottom of my heart thanks for the support, friendship, companionship, sincerity, and assistance at all times from this work and this master's degree.

Resumo

Os investigadores de cenas de crime enviam, frequentemente, para os Laboratórios forenses amostras biológicas limitadas em quantidade e qualidade do ADN, devido a condições ambientais, tempo e condições de preservação. Para facilitar a análise destas amostras, a *InnoGenomics Technologies* desenvolveu novos kits para quantificação e amplificação de ADN (InnoQuant® HY and InnoTyper® 21) baseados em retrotransposões. O objetivo do presente trabalho foi validar internamente os kits InnoQuant® HY e InnoTyper® 21, sob as condições de trabalho do Laboratório de Polícia Científica da Polícia Judiciária (LPC-PJ) e avaliar a sua capacidade para produzir resultados em amostras recolhidas em cenas de crime reais.

Para a validação interna foram realizados testes de sensibilidade, repetibilidade e reprodutibilidade. Adicionalmente, foi também realizado um estudo de concordância entre os kits atualmente em uso no LPC-PJ e estes novos kits. Para os testes anteriormente descritos foram selecionadas 132 amostras de extratos do LPC-PJ (principalmente cabelo com folículo, mas também sangue, vestígios de contacto e dentes). Finalmente, foi analisado um conjunto de 40 amostras de cabelo sem raiz para observar a capacidade dos kits InnoQuant® HY e InnoTyper® 21 para produzir resultados nestas amostras consideradas difíceis do ponto de vista forense.

Os resultados obtidos para os testes de sensibilidade, repetibilidade e reprodutibilidade permitiram validar internamente os kits InnoQuant® HY and InnoTyper® 21. Das 132 amostras testadas, os resultados demonstraram que o kit InnoQuant® HY foi capaz de produzir resultados de quantificação em cerca de 28,8 % mais das amostras analisadas do que o kit the Quantifiler® Duo, indicando o novo kit é mais efetivo no tratamento de amostras difíceis. No que respeita ao kit InnoTyper® 21, os resultados demonstraram que no conjunto de amostras selecionadas e sob as condições de trabalho utilizadas neste estudo, as diferenças observadas entre ambos os kits (InnoTyper® 21 e GlobalFiler™) não foram significativas. Contudo, foi comprovada a utilidade do kit InnoTyper® 21, especialmente devido a amplificação de um grande número de perfis genéticos completos. Os resultados obtidos para o conjunto de 40 amostras de cabelos sem raiz permitiram confirmar que ambos os kits InnoQuant® HY e InnoTyper® 21 representam boas alternativas para análise de amostras difíceis.

Keywords: *Retrotransposões, InnoQuant® HY, Indice de Degradatção, InnoTyper® 21, INNULs*

Abstract

Forensic laboratories frequently receive, from crime scene investigators, biologic samples that are limited regarding DNA quantity and quality, due to environmental stresses, storage time and preservation conditions. In an attempt to facilitate the analysis of this type of biological samples, *InnoGenomics Technologies* developed new commercial DNA quantification and amplification kits (InnoQuant® HY and InnoTyper® 21) which use retrotransposons as molecular markers. The aim of this work was to perform the internal validation of InnoQuant® HY and InnoTyper® 21 kits, under the working conditions of the Laboratório de Polícia Científica da Polícia Judiciária (LPC-PJ) and to evaluate the ability of these kits for obtaining results with samples taken from real crime scenes.

For the internal validation, sensitivity, repeatability, and reproducibility tests were performed. A concordance study between the kits currently in use at LPC-PJ (Quantifiler® Duo and GlobalFiler™) and these new retrotransposon-based kits was also performed. Therefore, 132 samples were selected from extracts of LPC-PJ caseworks (mainly hairs with follicle but also blood, contact traces, and teeth). Additionally, a set of 40 rootless hair samples was analyzed to observe the ability of InnoQuant® HY and InnoTyper® 21 kits to produce results with these challenging samples.

The results obtained for sensitivity, repeatability, and reproducibility tests allowed internally validated both InnoQuant® HY and InnoTyper® 21 kits. For the 132 samples tested, results demonstrate that InnoQuant® HY kit was able to produce quantification results in nearly 28.8 % more of the analyzed samples than the Quantifiler® Duo kit, indicating that the new kit is more effective in treating challenging samples. Regarding the InnoTyper® 21 kit, results demonstrate that for the samples selected and under the conditions used in this study, the differences observed between the kits (InnoTyper® 21 and GlobalFiler™) were not significant. Therefore, the utility of InnoTyper® 21 has been proven, especially by the successful amplification of a greater number of complete genetic profiles. The results obtained for the set of 40 rootless hair samples allowed to confirm that both InnoQuant® HY and InnoTyper® 21 kits represent very good alternatives for the analysis of challenging samples.

Keywords: *Retrotransposons, InnoQuant® HY, Degradation Index, InnoTyper® 21, INNULs*

Contents

Figures

Tables

Abbreviations

Av – Average

bp - Base Pairs

CC - Coiled-Coiled Domain

cDNA – Complementary Deoxyribonucleic Acid

CE - Capillary Electrophoresis

C_T – Threshold Cycle

CTD - C-Terminal Domain

DI - Degradation Index

DNA – Deoxyribonucleic Acid

DR - Direct Repeats

DTT – Dithiothreitol

EDTA – Ethylenediamine Tetraacetic Acid

EN – Endonuclease

ENFSI - European Network Of Forensic Sciences Institutes

HERV - Human Endogenous Retrovírus

IN – Integrasse

INNUL - Insertion And Null Alleles

IPAC - Instituto Português de Acreditação

IPC - Internal Positive Control

IS - Insertion Sequences

kb – Kilobase Pair

kDa – Kilodalton

LINE - Long Interspersed Elements

LPC-PJ - Laboratório de Polícia Científica da Polícia Judiciária

LTR - Long Terminal Repeat

M – Molar

mM – Milimolar

ng – Nanogram

NTC - No Template Control

ºC – Celsius Degrees

ORF - Open-Reading Frames

PCR – Polymerase Chain Reaction

QD – Quantifiler® Duo

qPCR – Quantitative Polymerase Chain Reaction

R^2 – Correlation coefficient

RFLP - Restriction Fragment Length Polymorphism

RFU – Relative Fluorescence Units

RH - Ribonuclease H

RNA – Ribonucleic Acid

RNP – Ribonucleoprotein

rpm – Rotation Per Minute

RRM - RNA Recognition Motif

RS – Reference Sample

RT - Reverse Transcriptase

SINE - Short Interspersed Elements

SNP – Single-Nucleotide Polymorphism

SSR - Simple Sequence Repeat

STR - Short Tandem Repeats

SVA - SINE-R/VNTR/Alu

SWGDAM - Scientific Working Group on DNA Analysis Methods

TE - Transposable Elements

TIR - Terminal Inverted Repeats

TPRT - Target-Primed Reverse Transcription

TSD - Target Site Duplications

TWGDAM - Technical Working Group on DNA Analysis Methods

UTR - Untranslated Region

UV - Ultraviolet Light

VNTR - Variable Number of Tandem Repeats

µL – Microliter

µM – Micromolar

1. Introduction

Forensic Genetics deals with the application of genetic knowledge to resolve legal cases.

In a crime scene, two major types of evidence can be obtained. Testimonial evidence is obtained by a statement from people who may have witnessed the crime or part of it. Physical evidence is something left at the crime scene that can take several forms and dimensions, being able to prove that the crime was committed or establish a key element of a crime. Amongst physical evidence, the direct evidence supports a fact without inference; while the circumstantial evidence compels an inference to associate the evidence to the conclusion of a fact. Thus, the interpretation and use of physical evidence becomes critical in forensic research [1, 2].

In the recent years, the use of deoxyribonucleic acid (DNA) as forensic evidence gained importance in forensic investigations. The use of DNA evidence has rapidly increased and considerably benefited justice due to the recent developments in the field of DNA testing [3]. Evidence collected in casework and processed at forensic laboratories enables to establish a probability-based connection between a crime scene, particular individuals (either suspects or victims), and objects involved in a crime. Such link, determined by the comparison of the genetic profiles, complemented by other means of proof, is often critical to assist in acquitting or convicting the suspect [3, 4].

Aims

The main aims of the present work were to internally validate both InnoQuant® HY and InnoTyper® 21 kits (InnoGenomics Technologies) for later implementation in Laboratório de Polícia Científica da Polícia Judiciária (LPC-PJ) routine laboratory, and to evaluate the ability of these new retrotransposons-based kits to produce results in challenging samples. Thus, the genetic profiles from challenging samples (hairs, contact traces, blood, and teeth) obtained from crimes scenes and suspects using conventional approaches (Quantifiler® Duo and GlobalFiler™, Thermo Fisher Scientific) were compared with the results obtained using these new retrotransposons-based kits.

1.1. Historical contextualization of Forensic Genetics

Over the years, several techniques have been developed to resolve legal conflicts or criminal cases. Early in the 20th century, Karl Landsteiner described the ABO blood system group and found that it was possible to allocate people into four groups according to their blood type. This system, based on four blood groups (A, B, AB, and O), has been proven useful to determine if a particular donor is the contributor of a sample left at a crime scene [4, 5].

Fifteen years later, Leone Lattes reported the use of ABO typing system in resolving a paternity test. The first step towards the use of biological markers to distinguish between individuals was given [4, 5].

In the 1960s and 1970s, the developments of Molecular Biology techniques, such as the use of restriction enzymes, Sanger sequencing, and Southern blotting, allowed the examination of DNA sequences, having been reported in the 1980s the analysis of the first polymorphic locus [4].

In 1985, Alec Jeffreys described DNA fingerprinting (also known as DNA typing or DNA profile), reporting that certain DNA regions contain tandemly repeated sequences (variable number of tandem repeats, VNTR) and that the number of repeats differs amongst individuals [6]. Thus, the restriction fragment length polymorphism (RFLP) technique, that allows examining the variation in the size of these repetitive DNA sequences, has been developed for application in human identification testing. The RFLP consists in the DNA cleavage by restriction enzymes to produce fragments of variable lengths. Then, fragments are separated by size in an agarose gel electrophoresis being polymorphic loci detected by Southern blotting using hybridization probe [5, 6]. Each locus has a different number of repeats being created a distinctive pattern for each individual; therefore, RFLP analysis has proven to be a highly discriminative method. However, this technique requires large amounts of DNA (50 to 500 ng) [4]. This limitation associated with DNA degradation in samples left at the crime scene which prevents the amplification of long fragments as are the VNTR and the difficulty of interpreting the results obtained with this method acted as a starter for the introduction of polymerase chain reaction (PCR).

PCR is an automated process to be used in the laboratory and consists in the enzymatic amplification of a specific region of genomic DNA sequence, using specific primers [7]. Developed in 1984, by Kary Mullis, the PCR technique is a powerful molecular biology

tool that has revolutionized all areas of medical and biological science [8]. After its discovery, the PCR technique was quickly incorporated into the forensic analysis, being used for the first time in 1988 to analyze skeletal remains. As VNTRs, the mostly used in casework, required large amounts of DNA, PCR technology was applied to the analysis of VNTR loci amplifying alleles between 5-10 kilobases (kb). However, the existence of forensic samples which were highly degraded and consequently had low amounts of DNA continued to be a problem. This limitation was overcome, at the beginning of the 1990s, with the development, and characterization of short tandem repeats (STR), that showed to be simpler and shorter than VNTR and, therefore, more suitable for the analysis of biological samples recovered at crime scenes [5].

Short tandem repeats (STR), also known as microsatellites or simple sequences repeats (SSRs), firstly identified in the early 1980s, are tandemly repeated DNA sequences. Consist of repeated regions with 1-6 base pairs (bp) long (mono-, di, tri-, tetra-, penta- and hexanucleotide repeats) and in alleles that are shorter than 350 bp. STRs are distributed throughout the entire human genome being the majority located in non-coding regions, either in introns or intergenic sequences [4, 5, 9, 10], STRs are highly polymorphic in nature being important genetic markers for mapping studies, disease diagnosis, and human identity tests [11]. The simplicity of PCR amplification without the problems of differential amplification (due the fact that both alleles from a heterozygous individual are similar in size since the repeat size is small) associated with the possibility of successfully typing low amounts of DNA even in a degraded form, made STRs widespread for current forensic applications [11-13]. However, and even though thousands of polymorphic STRs have been characterized in the human genome, only a small set of loci were chosen for use in forensic DNA and human identity tests [14]. STRs can be divided into three different groups depending on their repeat structure: i) simple or perfect repeats have the same length and sequence in every repeat unit (e.g. (GATA)(GATA)); ii) compound or imperfect repeats contain stretches of two or more different repeat types (e.g. (GATA)(GATA)(GACA)); iii) complex or interrupted repeats have several blocks of repeats with different unit lengths but also contain intervening sequences (e.g.:(GATA)(GACA)(CA)(CATA)) [4]. Concerning human identification, the most important feature is to have DNA markers that display the highest possible variation or a number of less polymorphic markers that can be combined to discriminate between samples [12]. Thus, among the various STR systems, forensic laboratories prefer tetra- or pentanucleotide locis, since the amount of stutters (additional peaks produced by amplicons that are typically one or more repeat units less in size than true allele and that

can complicate the interpretation of DNA profiles) produced during PCR is lower [9, 12]. In general, the analysis of autosomal STR is highly discriminant and sensitive. However, in certain cases, this analysis fails, or becomes difficult to interpret [4]. These cases usually occur during the analysis of challenging samples, such as severely degraded samples, hairs, DNA mixtures, and presence of PCR inhibitors.

1.2. DNA analysis of challenging samples

Forensic laboratories frequently receive, from crime scene investigators, biologic samples (mostly bone fragments, teeth, blood, and hairs) that are very limited in terms of DNA quantity and quality, mainly due to environmental stresses (which degrades DNA molecules by randomly breaking them into smaller pieces), storage time, and preservation conditions [6, 12]. In most cases, forensic geneticists are confronted with incomplete DNA profiles or lack of results from the analysis of the samples. This is because STRs, the genetic markers of choice for DNA analysis, produce long amplicon sizes. As DNA degradation takes place, this molecule becomes progressively more fragmented and difficult to amplify. Additionally, samples collected from crime scenes are often contaminated with PCR inhibitors, such as hematin, which interfere with PCR process and produce profiles with alleles and/or locus drop-out [15-17].

Hair shafts are one of the most common types of evidence found during crime scene investigation, and as such becomes an ideal sample for DNA testing. In cases where hair samples sent to the laboratory still containing the root, the probability of obtaining a complete DNA profile using the standard PCR technology is around 60% - 70%. However, most of the hairs found at the crime scenes are rootless. Unfortunately, the highest amount of DNA is located in the root. Therefore, it is expected that the analysis of rootless hair samples does not produce results making them a problem for forensic laboratories despite their great evidence value in court [18, 19]. When receiving rootless hair samples, genetic experts are faced with the need for special approaches to achieve better results. Thus, new approaches, such as mini-STRs and also the mitochondrial DNA have been developed [16]. However, the difficulties in obtaining profiles remain, and the mitochondrial DNA analysis has the disadvantage of providing a result of maternal lineage and not individual identification. In the recent years, an alternative methodology using mobile elements has been proven useful in such cases.

1.3. Mobile elements in forensic DNA analysis

Mobile DNA elements, also known as transposable elements (TEs) or "jumping genes", are genomic sequences that have the ability to move independently into different parts

of the genome (within and sometimes between genomes) [20, 21]. Although the primary source of TEs is not entirely known, it has been hypothesized that its origin occurred millions of years ago from viruses that have invaded the eukaryotic genome. Reasons that support this hypothesis are the structural and functional similarities found between viruses and TEs, the fact that the drugs that inhibit virus activity can also inhibit TEs, and that proteins involved in the inhibition of TEs within the cells also have inhibitory effects on viruses [22].

Firstly discovered in the1940s, by Barbara McClintock, in maize, the function of these elements has been reported in several organisms [21, 22]. After the conclusion of the Human Genome Project, it has been revealed that approximately half of our genome (45 - 47%) resulting from TEs activity. However, only a small portion of these elements remains active. Also, it has been estimated that nearly 25% of human promoter regions, as well as 4% of the human exons, have TEs derived sequences [23]. In some plants, TEs constitute about 90% of the genome [20, 22]. Moreover, TEs have a significant biological importance, especially in genome evolution, gene expression, genetic instability, and genetic diseases [20]. These elements induce modifications into the genome either by his capability to "jump" to new locations, or by facilitating chromosomal rearrangements [23].

According to the method used to duplicate into the genome, TEs can be grouped into two distinct classes: transposons (class 2 elements) and retrotransposons (class 1 elements), as illustrated in Figure 1.1 [24, 65].

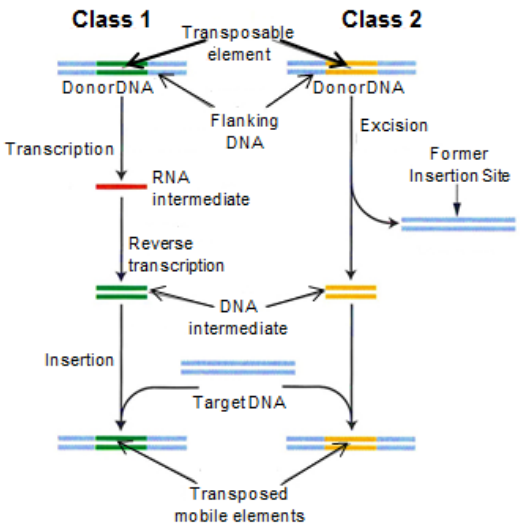


Figure 1.1 - Schematic representation of the two classes of transposable elements and their transposition method (Adapted from Lodish *et al*., 2102 [24])

1.3.1. DNA transposons

Constituting nearly 3% of the human genome, DNA transposons have the ability to move in the genome [21, 25]. These elements act through a cut and paste mechanism in which the transposon excises itself from a particular location, moves as DNA, and reintegrates into new genomic sites [20, 22, 23]. Despite not being currently active in the human genome, these elements were active during the early primate evolution [20, 21, 26]. Currently, transposons remain mainly active in the genome of simpler organisms, such as bacteria, plants, and flies, conferring the benefit of spreading antibiotic-resistance genes and reproduction through sexual mode [22, 25].

The simplest DNA transposons, known as insertion sequences (IS), are independent units that encode one or two enzymes necessary for transposition, and consist of a gene for transposase flanked by two terminal inverted repeats (TIRs) sequences [23, 27, 28] (Figure 1.2). These inverted sequences, apparently identical, are oriented in opposite directions which are characteristic of a particular IS element and define the end of the transposon. As such, the transposase recognizes these TIR, excises the DNA transposon from its original genomic location and then generates a break in another genomic location where subsequently reinserts the transposon (Figure 1.3 a). Upon the insertion of the IS element, the sequence of the host DNA is duplicated creating target site duplications (TSDs), which correspond to a particular characteristic for each DNA transposons. At the insertion site, the IS sequence is always flanked by very short direct repeats (DRs) [23, 27] (Figure 1.3 b). However, non-autonomous elements that do not encode proteins required for mobilization have been reported. These non-autonomous elements depend on the autonomous transposons for their mobilization [23] (Figure 1.4).

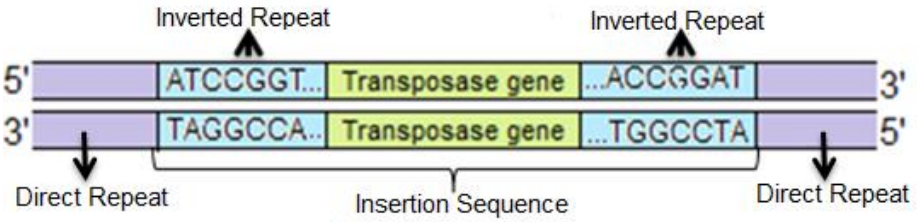


Figure 1.2 - Simplest DNA transposon containing a transposase gene flanked by two inverted repeats (Adapted from Russell, 2002 [29])

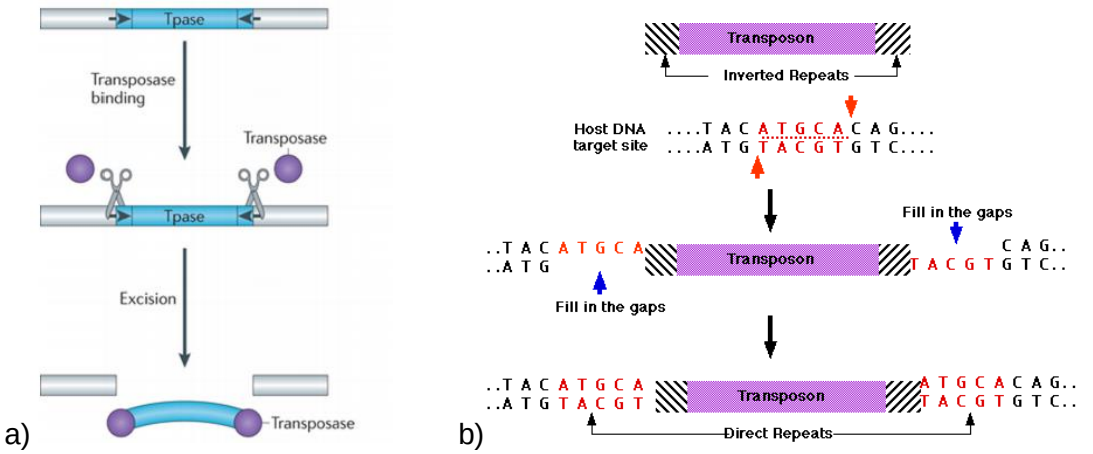


Figure 1.3 - Mobilization mechanism of Transposons. a) Transposase recognizes the terminal inverted repeats, excises the transposon and then, incorporated it into a new genomic location; b) insertion of the IS element in the host DNA and duplication of it sequence, creating target site duplications (TSDs) (Adapted fom Levin and Moran, 2011 [30] and Russell, 2002 [29]).

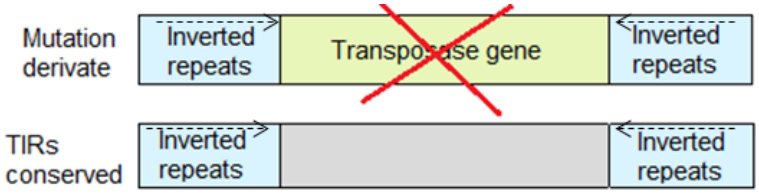


Figure 1.4 - Schematic representation of a non-autonomous transposon element. TIR, terminal inverted repeats (Adapted from Slotkin and Martienssen, 2007 [31])

In addition to these "classic" types of transposons, the DNA transposons can still be divided into two more main subclasses. The *Helitrons* whose mechanism used is possibly related to the rolling-circle replication (Figure 1.5 a), and the *Mavericks* whose transposition mechanism is not well understood, but it is thought that possibly replicates using a self-encoded DNA polymerase [28] (Figure 1.5 b).

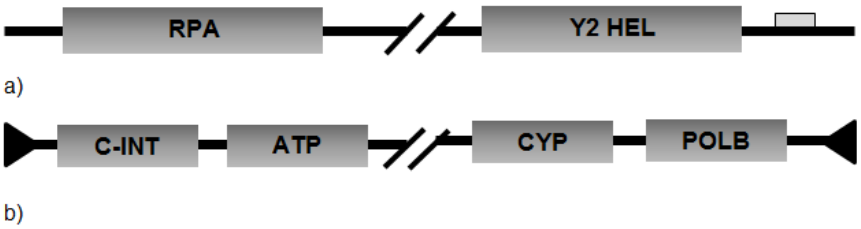


Figure 1.5 - Schematic representation from subclass 2 of DNA transposons. a) *Helitron*: gray rectangles represent protein coding regions; RPA, replication protein A; Y2 HEL, YR with YY motif helicase; bond breaking represents regions that can contain one or more additional ORFs; b) *Marverick-Polinton*: gray rectangles represent protein coding regions; C-INT, C-integrase; ATP, packaging ATPase; CYP, cysteine protease; POLB, DNA polymerase B; bond breaking represents regions that can contain one or more additional ORFs; black triangles represents terminal inverted repeats (Adapted from Piégu *et al*., 2015 [32])

1.3.2. Retrotransposons

Constituting more than 40% of the human genome, retrotransposons are another class of TEs (class 1) that also resort to a copy-and-paste mechanism for mobilization [33-35]. This mobilization mechanism resorts to an RNA intermediate which is then reverse transcribed into a complementary DNA (cDNA) copy by a mechanism called target-primed reverse transcription (TPRT), and then inserted into new genomics locations [25, 26].

Based on the presence of a structure named long terminal repeat (LTR), retrotransposons can be classified into two distinct groups: LTR and Non-LTR retrotransposons [20, 22] (Figure 1.6).

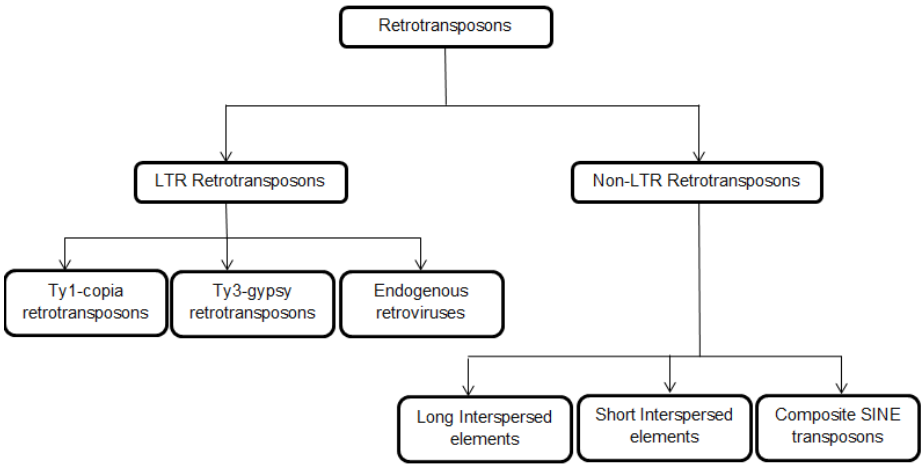


Figure 1.6 - Retrotransposons subdivision based on their structural characteristics and mobilization (Adapted from Ayarpadikannan and Kim, 2014 [20])

LTR retrotransposons, also known as human endogenous retrovirus (HERVs), constitute approximately 8% of the human genome [21, 36], and possess many characteristics similar to retroviruses, such as reverse transcriptase (RT), ribonuclease H (RH) and integrase (IN) genes [37]. However, normally these elements do not have a gene envelope or have a nonfunctional one [36]. Encoding Pro, Gag, Pol and sometimes Env-like proteins, its insertion in the human genome dates back over 25 million years ago, evolving from ancient viral infections of the germline [20, 21, 38]. As mentioned above, the HERVs mobilization occurs through transcription by Ribonucleic acid (RNA) polymerase II, followed by reverse transcription and re-integration into the host genome [39]. Currently, the activity of HERVs is limited in humans and the active human-specific LTRs identified belonging to the HERV-K family [20, 21]. Evidence suggests that the long terminal repeats of HERVs aid in the regulation of the expression of nearby genes due to the capacity of regulatory sequences found into retroviral LTR can alter or inactivate

the expression of adjacent genes. Moreover, HERVs insertions can potentially benefit the host, since its insertion in a particular location of the genome can reverse a harmful mutation [20].

Non-LTR elements are retrotransposons that do not contain long terminal repeats, and possess a terminal poly-A tail [36]. These old genetic elements were persevered in the eukaryotic genomes for hundreds of millions of years [40]. Within the non-LTR category, retrotransposons can be further divided into three subgroups according with their capacity to mobilize [22, 25]: long interspersed elements (LINEs), short interspersed elements (SINEs) and a composite element named SVA (SINE-R/VNTR/*Alu*). Non-LTR elements, represent the major class of mobile elements in the human genome (nearly 21% of human DNA), due its activity and are comprised of active elements such as L1, *Alu*, and SVA - which is believed that are the only three families of retrotransposons currently active in the human genome - and by inactive elements as such L2 and mammalian-wide interspersed repeats [20, 21, 34-36, 41].

1.3.2.1. Line-1 or L1 elements

Long interspersed elements-1 or L1 are autonomous retrotransposons (the only active autonomous mobile DNA element in the human genome) that can encode the enzymatic machinery necessary for transposition [22, 25]. Due to the high number of copies of these elements, more than 500.000 copies in the human genome resulting from the continued mobilization activity in the last 150 million years, L1 elements are the most abundant retrotransposons constituting approximately 17% of the human genome [21, 25, 36]. However, despite the high number of copies of these elements, only 100 of them remain intact [20, 21].

A full-length L1 element (Figure 1.7) has approximately 6 kb and is composed of a 5' untranslated region (UTR), containing an internal RNA polymerase II promoter, two open-reading frames (ORF1 and ORF2) separated by a short inter-ORF spacer, and a 3' UTR that contains a polyadenylation signal ending with an oligo dA rich tail of variable length [20, 21, 25, 34, 36]. Usually, L1s are flanked by target site duplications created by the retrotransposition process and vary from 7 to 30 bp in length, approximately [36]. Both ORFs are necessary for the occurrence of L1 retrotransposition since they code the necessary proteins for this to occur. ORF1 encodes an RNA-binding protein (a 40 kDa protein known as ORF1p) that contains a coiled-coiled domain (CC), a non-canonical RNA recognition motif (RRM) domain, and a basic C-terminal domain (CTD). ORF2 encodes a protein (a 150 kDa protein known as ORF2p) with endonuclease (EN)

and reverse-transcriptase (RT) activities. ORF2 also contains two other conserved domains, the Z domain (which is immediately proximal to the RT domain and mutations in this area can nullify reverse transcriptase activity in vitro) and cysteine-rich motif (C-domain) of unknown function [21, 34, 36]. This is the necessary molecular machinery that enables the retrotransposition process to occur and is known as target-primed reverse transcription (TPRT) [21, 37].

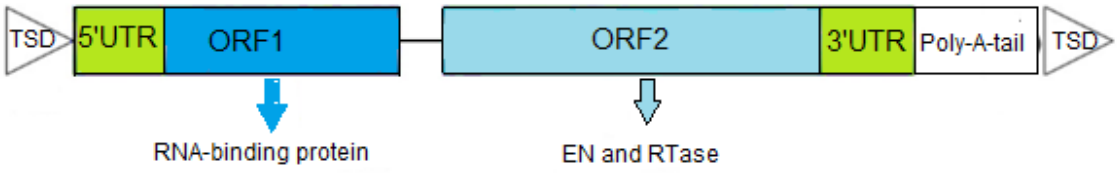


Figure 1.7 - Full-length active L1 element and its constitution: two open-reading frames (ORF1 and ORF2) for autonomous mobilization; TSD, target site duplication; UTR, untranslated region; EN, endonuclease domain; RT, reverse transcriptase domain; VNTR, variable number of tandem repeats; POLY (A), oligo dA rich tail (Adapted from Mills *et al*., 2007 [42]).

The retrotransposition of the human L1 begins with transcription from the internal RNA polymerase II promoter. After transcription, the L1 RNA is transported to the cytoplasm where ORF1 and ORF2 proteins undergo translation. Several genetic, biochemical, and phylogenetic studies demonstrate that both proteins present a strong cis-preference and, consequently, they preferentially associate with the L1 RNA transcript that encodes them to form a ribonucleoprotein (RNP) particle which is a proposed retrotransposition intermediate. After this, the RNP formed is transported back into the nucleus by a mechanism that is still understood [21, 36]. Inside the nucleus, the L1 retrotransposition process probably occurs by a mechanism mentioned above named target-primed reverse transcription during which it is thought that the L1 endonuclease breaks the first strand of target DNA releasing a 3' hydroxyl, which is then used as a primer for reverse transcription of L1 RNA by L1 reverse transcriptase, leading to the first strand of L1 cDNA. The second strand of the target DNA is cleaved and used as a primer for the synthesis of the second strand by poorly understood mechanisms. The integration process completes retrotransposition and results in L1 structural hallmarks, including frequent 5' truncations, the presence of an oligo dA-rich tail at the 3' end, and variable-length target site duplications [21, 36] (Figure 1.8).

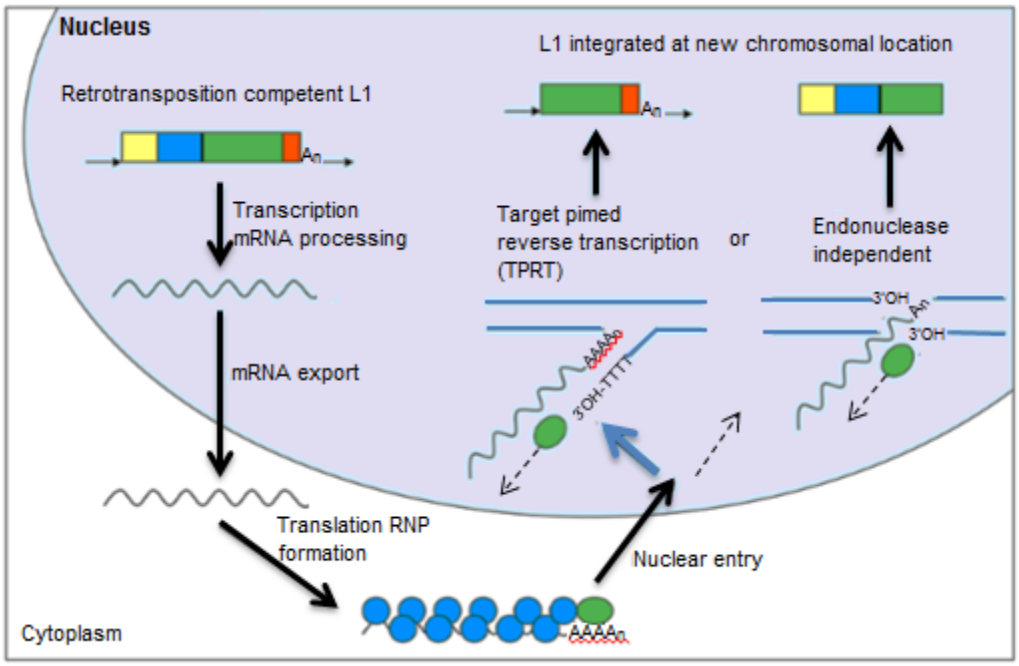


Figure 1.8 - Model of L1 retrotransposition cycle. Blue circles represents ORF1, green oval represents ORF2, black arrows represents variable length target site duplications, blue arrow represent the integration into the genome and dashed arrow represents endonuclease independent retrotransposition (Adapted from Hulme *et al*., 2006 [36]).

1.3.2.2. Alu elements

Alu elements are another type of retrotransposon, representing the family of SINEs, frequently called “a parasite’s parasite” due to their inability to encode for a polymerase. For this reason, *Alu* are non-autonomous elements once they need and depend on the enzymatic machinery of L1 elements for its retrotransposition [20].

Although its transposition is dependent on the L1 autonomous elements, there are more than one million *Alu* copies in the human genome, resulting from their continued mobilization activity over the last 65 million years, a period that coincides with the radiation of primates [21, 25, 43, 44]. Therefore, *Alu* elements are the most successful mobile elements in the human genome regarding copy number (comprise approximately 10% of the human genome sequence) [20, 21, 43, 45, 46].

Alu elements are divided into several subfamilies that share specific positions and that have been activated at different time points during primate evolution. The subfamilies Y, Ya and Yb of *Alu* elements remain active in the genome, being able to produce new *Alu* insertions that are polymorphic in the human population. However, these subfamilies only became active at the time of divergence between the human lineage and their last common ancestor, the nonhuman apes. Thus, the inclusion of *Alu* elements is restricted to humans and the ancestral allele is considered to be the absence of the insertion. The

human genome contains about 30 distinct categories of Alu subfamilies, being the AluYa5 subfamilies, AluYb8 and AluYc1 the most recently inserted and the most active in the human lineage [20, 46-48].

The constant mobilization of *Alu* repeats in the human genome may have numerous consequences, such as insertional mutations, gene conversion, recombination, changes in gene expression, pseudogenization, structural variation, and formation of segmental duplications, leading to diversity and genomic instability [20, 25, 43, 46].

A full-length *Alu* element (Figure 1.9) has approximately 300 bp in length and is composed of a dimeric structure formed by the union of two monomers derived from 7SL RNA gene (first originated from 7SL RNA as a monomeric element of about 150 bp in an ancestor of primates and rodents which later, in primate evolution, was fused into two monomers and became the nearly 300 bp dimer that characterizes most active *Alu* elements nowadays), separated by an A-rich linker region (with a consensus sequence: A_5TACA_6). The 5'-region comprises the A and B boxes that are promoters of transcription by internal RNA polymerase III and the element ends with an 3' oligo dA-rich tail of variable length [21, 25, 37].

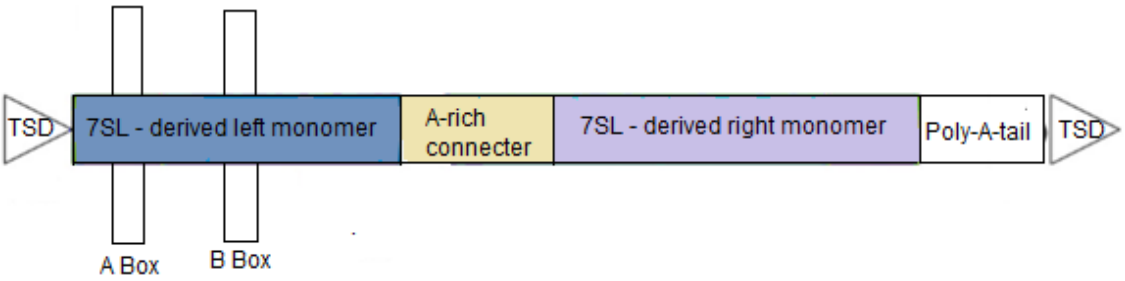


Figure 1.9 - Full-length active *Alu* element and its constitution: two monomers (derived from 7SL RNA gene) separated by an middle A-rich linker region; A and B boxes are promoters of transcription by internal RNA polymerase III; TSD, target site duplication; POLY (A), oligo dA rich tail (Adapted from Mills *et al*., 2007 [42]).

As mentioned above, *Alu* elements do not encode proteins required for its retrotransposition, requiring the machinery of L1 elements. Accordingly, it has been assumed that retrotransposition probably also occurs via TPRT. However the trans-mobilization mechanism by L1 proteins remains undefined. Because *Alu* elements do not have RNA polymerase III termination signals, *Alu* transcripts extend into the downstream flanking sequence until the terminator (typically a run of four or more consecutive Ts) is reached. Thus, these RNA polymerase III-mediated *Alu* transcripts are exported to the cytoplasm and linked to SRP9/14 proteins to form RNPs stable. It has been postulated that these RNPs interact with ribosomes, allowing positioning *Alu* transcripts nearby of nascent L1 ORF2 protein. Nevertheless, it is unclear whether *Alu*

RNPs have access to the L1 retrotransposition machinery in the cytoplasm or in the nucleus, since *Alu* RNPs should recruit L1 ORF2 proteins in the nucleus and immediately perform with TPRT [21].

*Alu*s constitute an important source of human variation. Several studies suggested that *Alu* elements insertions originate present/absence variants that potentially can be used not only as DNA markers in human population studies but also in forensic and paternity analyses [20, 49].

1.3.2.3. SVA elements

Firstly reported in 1994, human SVA elements are composite retrotransposons with approximately 3,000 copies in the human genome, being 27 - 38% of these copies structural polymorphic and actively mobile, resulting from its continued activity over around 25 million years of hominid evolution [21, 25, 50].

A full-length SVA element (Figure 1.10) has approximately 2 kilobases (kb) in length and presents an hexamer repeat region, an *Alu*-like region, a variable number of tandem repeats region, an HERV-K10-like region (SINE-R element was first reported as derived from human endogenous retrovirus HERV-K10 [50]), and a polyadenylation signal ending with an oligo dA-rich tail of variable length [20, 21, 25] and is flanked by TSDs [51, 52] [47, 48].

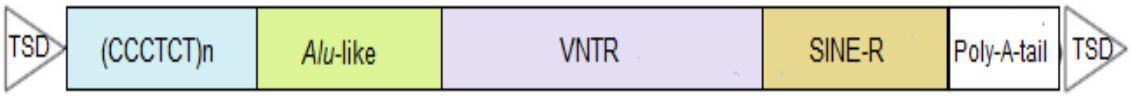


Figure 1.10 - Full-length active SVA element and its constitution: an hexamer repeat region; an *Alu*-like region; VNTR, a variable number of tandem repeats region: a SINE-R region; TSD, target site duplication; POLY (A), oligo dA rich tail (Adapted from Mills *et al*., 2007 [42]).

Although it has been proposed that SVA elements are transcribed by RNA polymerase II, these elements do not have internal promoters and, as such it is thought that can focus, at least in part, on the promoter activity in the flanking regions. Thus, it can be assumed that SVAs are non-autonomous elements and that, as the *Alu* elements, dependent on the retrotransposition machinery of L1 elements to proliferate into the human genome [21, 51].

The ancestral SVA has its origins about 14 million years ago and then diverged into six subgroups (A-F) based on their sequence divergence and evolutionary time estimates. According to subfamily age estimates, based on nucleotide divergence, the expansion of the four SVA subfamilies (A, B, C and D) began before the divergence of human,

chimpanzee, and gorilla, while the subfamilies E and F are characteristic to the human lineage (Figure 1.11). Therefore, it is thought that nearly 80% of the SVA members are probably human-specific in origin or, in other words, their integration in the human genome occurred after the human - great apes' radiation [51, 52].

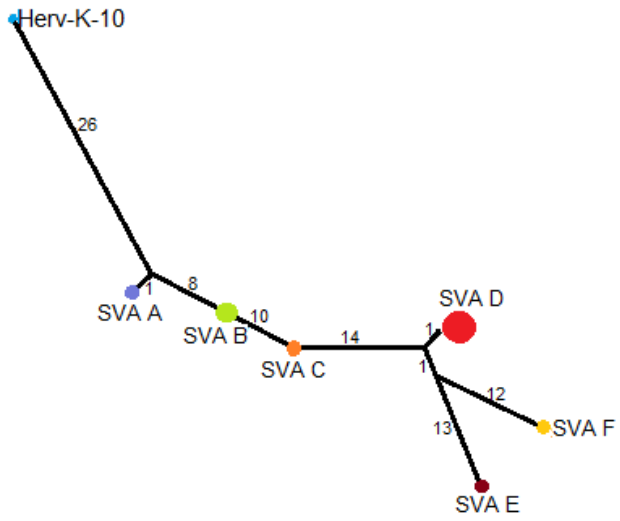


Figure 1.11 - Network of SVA subfamilies (Adapted from Wang *et al*., 2005 [52])

1.3.3. Retrotransposons insertion polymorphisms as genetic markers for human identification

Transposable elements have prominent influence regarding evolutionary change since they can impact gene expression by introducing alternative regulatory elements, exons, and splice junctions. Their presence, in a wide variety of genomes, is associated with deletions, inversions, duplications, translocations, and chromosome breaks. Together with recombination, independent assortment, and sexual reproduction, mutations caused by TEs plays a major role in creating genetic diversity [53].

The use of retrotransposons, particularly SINEs, as genetic markers demonstrated that these present some advantages over other widely used systems. SINES have proven to be almost ideal markers for phylogenetic and population genetic analysis, as well as for the application in human identity. The bi-allelic nature of retrotransposons conveys the technical simplicity of typing only two alleles per genetic marker instead of typing an array of nucleotides (which sometimes have hundreds of nucleotides in length). The existence of an element in several different individuals at a given locus, when there is a significant number of potential insertions sites for any element, demonstrates that retrotransposons are only identical by descent. This makes them more accurately and stable than other systems regardless inheritance (they have been shown to be essentially homoplasy-

free). The ancestral state of a SINE insertion locus is known to be the absence of the element which makes assumptions about this aspect of analysis unnecessary. Additionally, retrotransposons do not produce technical artifacts due to slippage during PCR, and thus should reduce some of the problems associated with the interpretation of STR currently used [53, 54]. Lastly, a crucial advantage is the fact that retrotransposons have low mutation rates, making them interesting for population studies and missing person kindship analysis [55]. The homoplasy-free nature of retrotransposon insertions allowed performing several studies that applied variation in *Alu* insertion frequencies to verify the human demography and their extension, forensic identification of specific individuals or groups. The use of *Alu* and/or L1 elements as genetic markers allowed not only explore ancient human origins, their migrations, and group continental human populations but also has proven to be preponderant for forensic applications enabling genotyping of unknown individuals and identifying their genetic ancestry with high probability [53].

1.3.3.1. InnoQuant® HY

For forensic laboratory analyses, it is important to determine the quality and quantity of the human DNA extracted from the sample, since these factors allow to estimate the optimal range of input DNA amounts for amplification, as well as, to facilitate the choice of the best amplification procedure to be applied. Therefore, it is crucial that the quantification method to be also extremely sensitive. Real-time PCR has several advantages when compared with other quantification methods, such as nonnucleic acid-based quantification methods, total genomic methods, or DNA hybridization-based human and higher primate-specific DNA methods [56].

InnoQuant® HY kit is a real-time PCR system (qPCR) that allows evaluating both the quantity and quality of human DNA existing on biological samples from forensic caseworks. This system was developed to detect total human and male DNA, and uses two independent genomic targets, a multi-copy sequence of short length and a separate multi-copy sequence of long length, to qualitatively measure the degree of degradation of a sample. The degradation level of a certain sample, designated as Degradation Index (DI), is given by the ratio between the amount of a short strand of amplified DNA with the amount of a long strand of amplified DNA target, allowing to know the quality of DNA and thus select the best typing method [57-59].

For the development of this multiplex, two independent retrotransposons targets (short and long targets), a 79 bp male specific target and a 172 bp amplicon from a synthetic

template as Internal Positive Control (IPC) to assess PCR inhibition were used to design the primers and the TaqMan probes. The “Short” target is an 80 bp sequence from an *Alu* element (more precisely an Yb8 *Alu*, which lineage is the second largest young group of *Alu* subfamilies and contains nearly 1800 copies per genome) while the “Long” target is a 207 bp sequence from an SVA element (which contains an average of nearly 1700-1800 copies per genome) (Figure 1.12) [57-60].

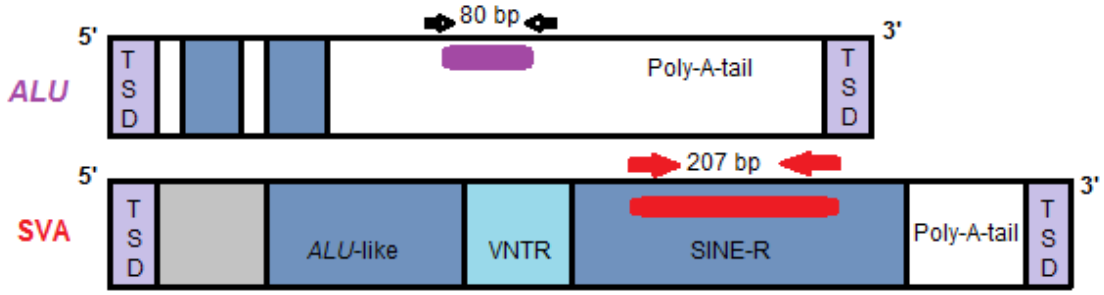


Figure 1.12 - Representation of two independent retrotransposons (*Alu* and SVA) genomic targets; TSD, target site duplication (Adapted from Pineda *et al*., 2014 [57, 59]).

When compared with commonly other quantification methods, the *Alu* systems present the advantage of using a large number of fixed insertions. Since they have a high copy number into the human genome, these tests have a high sensitivity dynamic range and can be successful used to quantify the human DNA in biological samples in forensic caseworks. The same is verified with the SVA elements which, while there are smaller families, contains nearly 1800 copies per genome. Thus, this large copy numbers will permit to minimize the effect of variation among individuals producing highly reproducible quantification values [57].

Previous studies using Ya5 *Alu* elements to assess the quantity of degraded DNA in forensic samples demonstrated a lack of reproducibility and/or sensitivity and did not present high PCR efficiencies [57]. These low PCR efficiencies occurred due to the competition between the two amplified fragments, mainly because the two Ya5 *Alu* element fragment sequences targeted were not independent [57]. To overcome this difficulty, this multiplex system uses two multi-copy independent targets. The inclusion of a sensitive Y-chromosome specific male target allows to determine accurately the amount of male DNA in challenging male-female mixtures (i.e. sexual assault samples) and easily identify when the sample processing should be terminated, modified or if the sample should be preserved for alternative amplification methods, such as InnoTyper® 21 [57, 58, 60]. The inclusion of an IPC target in the multiplex allows measuring the

inhibition and, in some cases, indicates potential mechanisms of inhibition in forensic samples [56].

1.3.3.2. InnoTyper® 21

STR amplification systems are preferably used for human identification tests because they are highly polymorphic and provide a high degree of sensitivity [55, 61]. Thus, these systems allow samples containing as little as 250 pg of DNA can be typed. However, as already mentioned the vast majority of samples sent for analysis in forensic laboratories are degraded and have significantly reduced amounts of DNA. To outwit this problem, several authors have proven the potential of dimorphic *Alu* elements in cell lines identification, paternity testing, and forensic analysis [55]. These studies were based on the size difference between the insertion and absence (null allele) of the *Alu* element by amplifying the whole region with the same primers [61]. However, with this inherent size difference greater than 300 bp between the two allelic states (insertion and null alleles, INNULs) was quite difficult to use retrotransposons for human identification [54]. This limitation is due to the wide difference between amplicons of INNULs caused in amplification efficiency during PCR, once there is a preferential amplification of the smaller allele and possibly the drop-out of the insertion element [55, 61].

To overcome the limitations that prevented the use of *Alu* elements for human identification, a new primer design methodology (Mini-Primer Strategy) has been developed. This methodology removes the intra-specific locus competitiveness in heterozygous by reducing the overall amplification size as well as the difference in amplification sizes amongst the two allelic states of INNULs. Thus, the resulting INNUL allelic amplicons can differ by as little as one base pair instead of the previous nearly 300 bp and can be reduced to smaller sizes than those presented by the STRs markers [54, 55, 61].

InnoTyper® 21 kit is a multiplex system based on *Alu* elements to determine small amplicon fragments (60 - 125 bp). It is compatible with existing PCR and capillary electrophoresis platforms and may be an optimum system to be used for DNA typing of highly degraded and/or low quality samples [62]. This multiplex system consists of 21 genetic markers, including 20 retrotransposons and Amelogenin [54]. For each marker, amplification of the two allelic states of INNULs is preformed using a common unlabeled forward primer and two fluorescently labeled reverse primers (Figure 1.13). The forward primer (CF) is used either for inserting or the absence of the retrotransposon. The labeled reverse primer for the null allele (RN) overlaps the insertion site of the RE and anneals

in its absence. The labeled reverse primer for the insertion allele (RI) has an overlap region with the junction and the retrotransposon itself, or just inside the retrotransposon, and when retrotransposon is present, the annealing site of the reverse primer is disrupted allowing the anneal of the insertion specific reverse primer at the site that overlaps with the insertion site and the adjacent portion of retrotransposon [55, 61].

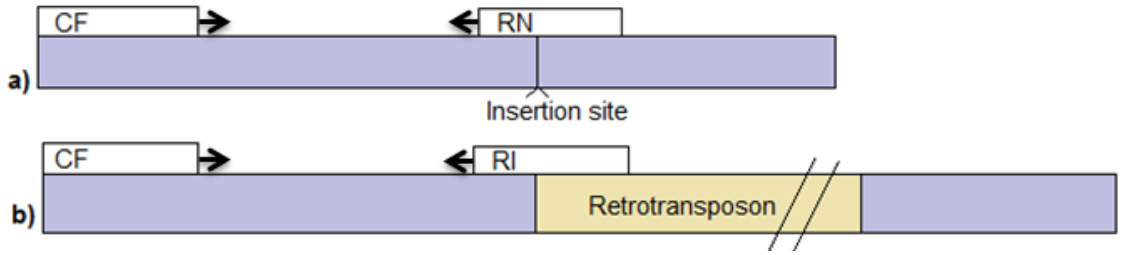


Figure 1.13 -Amplification strategy of InnoTyper® 21 kit (Adapted from LaRue *et al*., 2012 [55])

The selection of genetic markers to include in multiplexes must be careful and include markers that are highly polymorphic in all major populations (i.e., reaching 50% heterozygosity) and therefore desirable for testing human identification, as well as markers with high inbreeding coefficients (such as SNPs) which allow bio-ancestry analysis. Thus, the markers used for the development of InnoTyper® 21 kit are based on *Alu* elements and were selected based on molecular characteristics and existing population data [61] (Appendix I).

This amplification system is useful for the DNA analysis of forensic samples, presenting high discrimination power for samples previously analyzed with conventional STR methods that rendered no results, such as rootless hair shafts and degraded DNA samples. Their bi-allelic nature offers the advantage of type only two alleles for genetic marker, and thus simplifying the analysis and interpretation of results. However, InnoTyper® 21 kit presents limitations in the interpretation of samples containing DNA mixture when used in conventional capillary electrophoresis (CE) platforms. The same is not expected to happen whenever the analyses of DNA mixtures are performed in NGS platforms [54]. It is important to mention that it is possible to perform population statistical analysis with InnoTyper® 21 kit. However, the existing allelic frequencies were calculated from the American population database samples. Therefore, the lack of a Portuguese population database for retrotransposons-based genetic markers represents another limitation for the use of this kit.

1.4. Internal validation for forensics laboratories

The implementation of new procedures, methods, kits, software or equipment in forensic laboratories to obtain results leading to the conviction or acquittal of a suspect should be carefully evaluated and validated. When discussing the importance of maintaining good laboratory practice to obtain accurate scientific results have to take into consideration the quality assurance and quality control. Quality assurance refers to planned or systematic actions that are necessary to provide adequate confidence that a product or service will satisfy certain quality requirements. Quality control refers to daily operating techniques and the activities used to meet the quality requirements. As such, an organization plans quality assurance measures and performs quality control activities in the laboratory [6].

The constant development of new methodologies for DNA testing leads to a constant process of validation in forensic laboratories [59]. Validate a particular laboratory process is to demonstrate that it is robust (successful results are achieved a high percentage of the time and few, if any, samples need to be repeated), reproducible (equal or very similar results are obtained every time a sample is tested) and reliable (the results obtained are accurate and correctly reflect the sample being tested) through a defined range of conditions [6, 63]. Two types of validation exist: i) developmental validation is performed by the manufacturer of a DNA test or a group of laboratories with the intuit to test new kits, new primers sets and new technologies for detecting alleles; ii) internal validation, more specific to the needs of a particular forensic laboratory, consists of verifying that the established procedures examined previously by developmental validation given by another laboratory will work effectively in the laboratory [6, 63].

Validation guidelines in DNA analysis have been previously published by Technical Working Group on DNA Analysis Methods (TWGDAM) in 1989, 1991 and 1995, but in July 2004, the FBI's Scientific Working Group on DNA Analysis Methods (SWGDAM) published revised validation guidelines providing more detailed validation information [63, 64]. In December 2012, SWGDAM made minor revisions to the 2004 validation guidelines, and the internal validation process was changed to included five studies (known and nonprobative evidence samples, sensitivity and stochastic studies, precision and accuracy, mixture studies, and contamination assessment) [65].

Known and nonprobative evidence samples refers to methods proposed for casework samples that must be evaluated and tested using known samples, nonprobative evidence samples or mock case samples and, when possible, authentic case samples.

Results from these studies must be compared to the previous results of known samples and/or nonprobative evidence or mock case samples to guarantee concordance [65].

Sensitivity and stochastic studies are used to demonstrate sensitivity levels of the test. As such, by testing a range of DNA concentrations, these studies estimate the dynamic range, ideal target range, limit of detection, limit of quantitation, heterozygote balance (e.g., peak height ratio), and the signal to noise ratio associated with the assay. Sensitivity studies may also be used to assess stochastic effects (stochastic threshold) usually resulting from low quantity and/or low quality samples [65, 66].

Precision and accuracy are demonstrated by repeatability and reproducibility tests. Reproducibility tests are used to evaluate the variation in average obtained by different operators using the same measurement equipment to measure repeatedly the same sample. Repeatability tests are used to evaluate the variation of the measures obtained by a single operator, using the same measurement equipment and method, to measure repeatedly the same sample [65, 67].

Mixture studies are conducted to help forensic laboratories to establish guidelines for the interpretation of mixed DNA samples which are those usually found at crime scenes. These guidelines include determination of the number of contributors to a biological mixture, determination of the major and minor contributor profiles, and contributor ratios or proportions. Results from mixture tests can also be used to evaluate laboratories contamination [65, 66].

Finally, contamination assessment is performed using negative controls as well as known samples, to detect exogenous DNA (including allele drop-in and heteroplasmy) which may be originated from reagents, consumables, operator and/or laboratory environment [65, 66].

According to the European Network of Forensic Sciences Institutes (ENFSI) for a certain forensic laboratory implement new DNA quantification system and/or new multiplex kit (as is the case of the present work), at these minimum parameters to be validated internally must be added concordance studies.

2. Material and Methods

The practical component of this work was developed at the Forensic Specialty of Biology, in the Biotoxicology area of the Laboratório de Polícia Científica da Polícia Judiciária (LPC-PJ). The LPC is a unit of the PJ that supports criminal investigation, with scientific and technical autonomy to develop its activities with independence, impartiality and integrity. This activity is applied to multiple fields of forensic science and, in scientific technical terms, is organized into three main areas: Biotoxicology, Criminology and Physical Document.

The Forensic Specialty of Biology is inserted in the Biotoxicology area and performs forensic tests within the biology and genetics criminalistics, such as search and characterizing of biological traces; determining genetic profiles (autosomal STRs and Y chromosome markers); comparing genetic profiles; evaluating results by statistical analysis; bioresearch kinship in criminal situations; and sending genetic profiles to the "Base de Dados de Perfis de ADN", as determined by Law 5/2008 of 12 February.

This laboratory analysis of forensic samples requires the use of a set of preventive measures to avoid any undesirable contamination that may adversely alter the results. Thus, the LPC has implemented various procedures to prevent contamination, exchanges or any other error that may arise in the treatment of cases, lying currently in the process of obtaining accreditation of NP EN ISO/IEC 17025, having submitted the case to the Instituto Português de Acreditação (IPAC).

In the Forensic Specialty of Biology from LPC-PJ are analyzed two types of forensic samples: reference samples corresponding to samples that are properly identified and which have a reasonable DNA quantity and quality, and casework samples of unknown origin and compromised DNA quantity and quality. All the samples received in the LPC for analysis are properly identified with the examination number and a preparation number assigned by the preparation models developed by the service.

One of the first methods to prevent cross-contamination in LPC-PJ is the assignment of two independent and separate rooms for processing both types of samples, to certify the veracity of the results. The LPC-PJ also has a pre-PCR room for the plates preparation of quantification and amplification of each of these types of sample, as well as a post-PCR room. To circulate within these rooms is required the use of exclusive work clothes (gown, disposable gloves and masks), taking care to not circulate between different rooms with work clothes that are using in a particular room. The same applies for

transportation of materials and/or equipment. Before and after any laboratory procedure that refer to the use of a room, workbench, workstation and/or utensils, the same must be sterilized using ultraviolet light (UV) and/or washing with sodium hypochlorite and subsequent passage by ethyl alcohol. Another parameter for prevention of contamination is the addition of a negative control in all the different stages of analysis in order to enable monitoring of these contaminations.

2.1. Sample selection

Challenging samples such as hairs, blood, contact trace, bones fragments and teeth were chosen from extracts of LPC-PJ casework, based on its profiles (no results, inconclusive results, and possible degradation) and quantification values. Samples that matched the guidelines of the ENFSI guidelines [67] were also selected as control. Additionally to the casework samples whose extraction is performed from hairs with follicle was, a new set of samples of rootless hair was also included.

2.2. Extraction Methods

DNA extractions from samples selected from existing LPC extracts were performed using the commercial kit PrepFiler™ BTA Forensic DNA Extraction in the equipment AutoMate™ Express Instrument, both used according to the protocol, within a laminar flow camera to prevent contamination (Appendix II, Figure 1). Thus, each sample was cut to a PrepFiler column to which was added 260 µL of a mixture containing 248 µL of PrepFiler™ BTA Lysis Buffer; 3.4 µL of 1M Dithiothreitol (DTT) solution and 7.9 µL of proteinase K. The set "*PrepFiler*" column/tube was placed into the TERMOMIXER at 56 ºC for 40 minutes at 900 rpm. The set column/tube was centrifuged for two minutes at 10000 rpm. Finally, after checking whether the end of the tube was a volume of at least 150 µL, the column was discarded and the tube preserved for AutoMate™ Express Instrument use. For the set of 40 samples whose extraction was performed from rootless hair samples, the extraction method used was the same as that used by *InnoGenomics Technologies* [62]. For this, the hair selected for this set of samples was observed under a microscope to identify the presence of the root, which was subsequently removed with a sterile scissors. A hair fragment with approximately 2 cm was cut and placed inside a tube to be cleaned by sonication in 5 % *Terg-a-zyme*, for 20 minutes, followed by a rinse in 100% ethanol and molecular biology grade water. DNA extraction was performed using the commercial kit PrepFiler™ BTA Forensic DNA Extraction (according to protocol with minor modifications) in the equipment AutoMate™ Express Instrument. Each hair fragment was placed inside a PrepFiler column and added 260 µL of a mixture containing

248 µL of PrepFiler™ BTA Lysis Buffer; 3.4 µL of 1M Dithiothreitol (DTT) solution and 7.9 µL of proteinase K. The set column/tube was placed in the TERMOMIXER at 56 ºC for 60 minutes at 900 rpm. Then 200 µL of *Buffer AL* were added and samples were incubated for 10 minutes. After, all procedures were performed as described by user protocol and in accordance with what has been above reported for samples of the LPC extracts.

2.3. Quantification Method - InnoQuant® HY kit

The DNA quantification of selected samples was performed with the InnoQuant® HY kit (Table 2.1) according to the manufacturer's instructions, described in the *InnoQuant® HY Human and Male DNA Quantification & Degradation Assessment Kit* user guide v1.2.

Table 2.1 - InnoQuant® HY Kit contents.

InnoQuant® HY Kit Reagents	InnoQuant® HY Primer Mix
	InnoQuant® HY DNA Standard at 100 ng/µL
	InnoQuant® HY Dilution Buffer A
	Agilent Technologies Brilliant Multiplex QPCR Master Mix
	ROX reference dye at 1mM

Before starting any quantification/amplification/capillary electrophoresis run, a planning sheet (or plate layout) was created to serve as guidance during the plate preparation. For this purpose, the Forensic Specialty of Biology from LPC has created files templates (models) originated from the results analysis, not only allowing automate procedures, but also serving as quality control. All the procedures to prepare the plate for the quantification of DNA were performed in a pre-PCR room of sample problem, inside of a laminar flow chamber previously irradiated with ultraviolet radiation (UV) to prevent any contamination. Since InnoQuant® HY kit (Appendix III, Figure 1) was designed to be highly sensitive, it is extremely important have in attention their storage conditions (as fluorescent dyes attached to the probes are light-sensitive, the InnoQuant® HY Primer Mix and the Agilent reference dye should be protect from light when not in use) and aseptic working conditions.

2.3.1. Preparation of DNA Standards

To initiate a quantification reaction must be previously prepared serial dilutions of InnoQuant® DNA HY Standard, whose obtained values will be used to draw a calibration curve. These serial dilutions were made in duplicate to enable the increased precision of the calibration curve. According to the manufacturer's recommendations, five dilutions standards with known concentrations were prepared from the original solution of 100 ng/µL varying their concentrations from 20 ng/µL until 0.005 ng/µL. These five solutions

were prepared with the InnoQuant® HY Dilution Buffer A, such as is described in the Table 2.2.

Table 2.2 - Preparing serial dilutions from the initial solution of 100 ng/µL.

Standard	Concentration (ng/µl)	Recommended dilutions amounts	Dilution factor
Std. 1	20	10 µL [Stock 100 ng/µL] + 40µL Dilution Buffer A	5X
Std. 2	2.5	10 µL [Std. 1] + 70 µL Dilution Buffer A	8X
Std. 3	0.3125	10 µL [Std. 2] + 70 µL Dilution Buffer A	8X
Std. 4	0.0391	10 µL [Std. 3] + 70 µL Dilution Buffer A	8X
Std. 5	0.005	10 µL [Std. 4] + 70 µL Dilution Buffer A	8X

To prepare the dilutions was necessary equilibrate InnoQuant® DNA HY Standard to room temperature, for 15 minutes and when defrosted, vortexed for five seconds and centrifuged for 10-15 seconds at 3000 rpm (protected from light). After, fresh serial dilutions were performed by the mixture of InnoQuant® DNA HY Standard with InnoQuant® HY Dilution Buffer A in low adhesion tubes.

2.3.2. Reaction and Sample Setup

The preparation of the diluted standards for the calibration curve was followed by the preparation of reaction mix composed by InnoQuant® HY Primer Mix, Agilent Technologies Brilliant Multiplex QPCR Master Mix, and Agilent reference dye reagents, as shown in Table 2.3.

Table 2.3 - Components required for master mix preparation and required volume for each reaction.

PCR Components	Volume per reaction
Agilent Technologies Brilliant Multiplex QPCR Master Mix	10 µL
Agilent reference dye (2 µM)	0.3 µL
InnoQuant® HY Primer Mix	7.7 µL
Total Volume	18 µL

These reagents must be equilibrated at room temperature for 15 minutes in the dark before use. Then, 1 µL of 1mM Agilent reference dye was added to 499 µL nuclease-free water for a final concentration of 30 nM in the PCR reaction. Then, the reaction mix was prepared by combining the master mix, primer mix and reference dye, with the calculated volume per reaction as indicated in table 2.3 and vortexed. In a 96-well reaction plate were dispensed 18 µL of the reaction mix into each appropriate well and 2 µL of DNA standard dilutions, followed by the unknown samples and 2 µL of InnoQuant® HY Dilution Buffer A as negative control (or NTC from “no template control”). The plate was sealed with an optical adhesive cover using the plate cover applicator, and centrifuged for one minute at 1500 rpm.

After, the plate is then introduced in the 7500 HID Real-Time PCR equipment existing in the LPC-PJ to run (Appendix II, Figure 2).

2.3.3. 7500 Instrument Setup

Before running, the equipment was calibrated for FAM, Cy5, TAMRA, and HEX dyes. The calibration plates for these dyes were provided by *InnoGenomics Technologies* in the *InnoQuant® Spectral calibration kit* (Appendix III, Figure 2).

Quantification run comprises an initial heating time followed by 40 cycles of PCR as shown in figure 2.1.

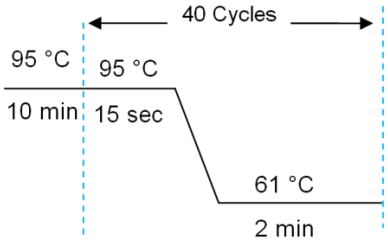


Figure 2.1 - Cycling parameters of Quantification run.

The analysis software generated calibration curves based on standards and calculated the DNA concentration in each sample. The reaction efficiency values must be greater than 90 % for the Short, Long, and Y targets; acceptable slope values must be between -3.6 and -3.1; and R^2 values must be greater than 0.98. Regarding to IPC target it is expected to have C_T values no more than 2 units above the mean IPC C_T for all quantification standards on the plate.

2.4. Amplification Method - InnoTyper® 21 kit

The DNA amplification of selected samples was performed with the InnoTyper® 21 kit (Table 2.4) according to the manufacturer's recommendations which are described in the *InnoTyper® 21 Human DNA Analysis Kit* user guide v3.7.

Table 2.4 – InnoTyper® 21 Kit contents.

InnoTyper® 21 Kit Reagents	InnoTyper® 21 Primer Mix
	InnoTyper® 21 Master Mix
	IGT DNA Polymerase
	DNA Control at 0.12 ng/µL
	InnoTyper® 21 Allelic Ladder
	ILS-155 Internal Lane Standard

All procedures were performed in the pre-PCR room of sample problem, inside a laminar flow chamber previously irradiated with ultraviolet radiation. InnoTyper® 21 kit (Appendix III, Figure 3) was designed to be highly sensitive and contains fluorescent dyes, being important to have precautions concerning storage conditions and light exposure.

2.4.1. PCR reaction Setup

The PCR reaction mix and the volume calculation of each component were prepared according to Table 2.5.

Table 2.5 - Components required for master mix preparation and required volume for each reaction.

PCR Components for Multiplex	Volume per reaction 25 µL Total reaction Mix
InnoTyper® 21 Primer Mix	3.5 µL
InnoTyper® 21 Master Mix	5.0 µL
IGT DNA Polymerase	0.5 µL
DNA template/ Positive Control/ TE^{-4} buffer	Up to 16.0 µL

The InnoTyper® 21 Primer Mix, InnoTyper® 21 Master Mix, and IGT DNA polymerase were thaw and equilibrated to room temperature. After, these reagents were vortexed for 3-5 seconds and centrifuged for two seconds. To prepare the reaction mix, the required volumes of master mix, primer mix, and DNA polymerase multiplex components was pipetted into a sterile 1.5 mL polypropylene tube which was vortexed for 5-10 seconds and centrifuged for two seconds. In a 96-well reaction plate, 9 µL of the reaction mix were dispensed. Samples were prepared according to the Table 2.6 and added to the reaction mix for a final reaction volume of 25 µl. For selected samples, the results obtained by quantification allowed to determine the DNA extract volume to be added to the amplification reaction, to obtain a maximum input of 0.5 ng/µL. The final reaction mix and samples were mixed and the plate was sealed with an optical adhesive film using the plate cover applicator.

Table 2.6 - Components and volume required in preparing samples for the PCR reaction.

Sample	Components and Volume
Negative Control	Add 16.0 µL of TE^{-4} buffer (10 mM Tris; 0,1 mM EDTA, pH 8.0)
Positive Control	Add 4.0 µL of provided DNA Control plus 12.0 µL of TE^{-4} buffer (DNA Positive Control should be equilibrated at room temperature, vortexed, and centrifuged briefly prior to adding into the reaction)
Test Sample	Add up to 16.0 µL of the DNA sample to the reaction mix (if the required sample volume was less than 16.0 µL, it was added the required volume of TE to totalize the 16.0 µL needed)

2.4.2. PCR Amplification

Once the plate prepared, was loaded to the thermal cycler GeneAmp® PCR System 9700 with gold-plated silver 96-Well Block equipment existing in post-PCR room of LPC-PJ, using the 9600 Emulation Mode and the heated cover was closed (Appendix II, Figure 3). An adhesive clear film was used to seal the plate wells and a MicroAmp® compression pad was positioned on top of the plate to prevent evaporation during thermal cycling. The thermocycler programming conditions for the PCR amplification reaction are described in Figure 2.2.

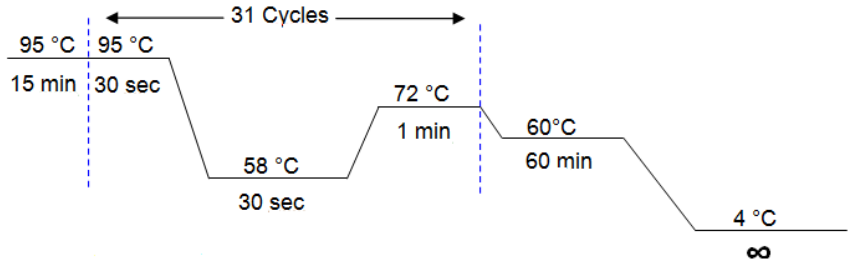


Figure 2.2 - Amplification program from InnoTyper® 21 kit.

At the end of amplification, samples were inserted in the automatic sequencer 3130XL Genetic Analyzer (*Applied Biosystems*), where the amplified products are separated and detected by a capillary electrophoresis process (Appendix II, Figure 4).

2.4.3. Genetic Analysis

To perform the genetic analysis, it is necessary that the InnoGenomics IGT 5 Dye Matrix Standard matrix file exists in the 3130XL Genetic Analyzer. The IGT 5 Dye Matrix Standard (Appendix III, Figure 4) consists of DNA fragments labeled with five different fluorescent dyes (FAM, JOE, TMR, ROX, and TGI-ORANGE) used to perform a spectral calibration on a specific dye set (in this case, G5). Once generated, this file was applied during sample detection to calculate the spectral overlap between the five different dyes and separate the raw fluorescent signals into individual dye signals. The matrix standard was prepared by combining the components as described in Table 2.7. To this, it was necessary first to thaw and mixe the matrix standard by vortexing for 10 seconds. Then, 10 µL of this mixture was dispensed in 16 reaction well of the plate, sealed with septa strip, and placed in the thermal cycler at 95 ºC for three minutes to denature the DNA. Finally, the plate was covered with plate retainer, placed into autosampler tray, and run.

Table 2.7 – Components and volumes required for the Matrix Standard preparation**.**

Components	Volume to 3130XL
IGT 5 Dye Matrix Standard	18 µL
HI-DI™ Formamide	162 µL

The spectral calibration was performed as described in the ABI 3130XL Genetic Analyzer user manual with only few modifications described in the *Multi-Capillary IGT 5 Dye Matrix Standard ABI 3130 & 3500 Series Genetic Analyzers*.

2.4.3.1. Sample preparation and data analysis using *GeneMapper® ID-X*

Samples were prepared based on the components and volumes shown in the Table 2.8.

Table 2.8 - Components and volumes required for the Sample preparation.

Components	Volume to 3130XL
Hi-Di™ Formamide	10.8 µL
ILS-155 Internal Lane Standard	0.2 µL

After prepared, the formamide/size standard mixtures were vortexed and centrifuged. On the 96-well sample plate were dispensed 11 µL of the mixture and 1 µL of PCR product or allelic ladder (Appendix III, Figure 5) which was them mixed (is critical to use an allelic ladder run under the same conditions as the samples because size values obtained for the same sample can differ between instrument platforms due to different polymer matrices and electrophoretic conditions). The reaction plate was sealed with appropriated septa, and placed in a thermal cycler for three minutes at 95 ºC.

The results produced by the capillary electrophoresis instrument were provided as electropherograms, posteriorly analyzed using the GenneMapper® ID-X v1.4 software.

2.5. The use of *Qiagen® QIAgility™* robot

In this study, the Qiagen® QIAgility™ robot was also used to evaluate the possibility of automatization of these kits. The QIAgility™ robot was designed to perform rapid, high-precision automated PCR setup and liquid handling actions in molecular biology applications. The use of this automated instrument enables to decrease the number of human faults that can occur during repetitive steps during the DNA analysis process.

The QIAgility™ robot was used to prepare the DNA Standards required for DNA quantification as well as pipetting 18 µL of the master mix into each well of the 96-well quantitation plate and 2 µL from each sample or DNA standard. This instrument was also used in the PCR plate preparation pipetting 9 µL of the reaction mix into each well of the 96-well followed by the required volume of each sample (test sample, negative control and positive control) necessary to a final reaction volume of 25 µL (Appendix II, Figure 5). It was also used in the post-PCR room for preparing the plate for capillary electrophoresis, to mixture 11 µL of formamide/size standard and 1 µL of PCR product or allelic ladder (Appendix II, Figure 6). All procedures performed by this automated

instrument were according to the requirements of the kits manufacturers described in the user guides. The use of two different techniques (manual and automated) for the preparation of quantification and amplification plates also will serve as an internal validation method.

2.6. Internal validation procedures

The parameters analyzed for the internal validation of InnoQuant® HY and InnoTyper® 21 kits were the minimum required by the ENFSI and include sensitivity, repeatability tests, reproducibility tests, and mixture studies. Additionally, to these parameters, a concordance study between the results obtained with the STR kits currently used by LPC-PJ (Quantifiler® Duo and GlobalFiler™) and the results obtained in the quantification and amplification kits were compared.

For the InnoQuant® HY kit, the evaluation of these parameters were considered the results for the Short target, since this target represent the total DNA present in the sample. In the mixture studies, were also considered the results obtained for the Y chromosome target, concerning the quantification of male DNA. For InnoTyper® 21 kit, the evaluation of these parameters was performed based on the percentage of amplified fragments and it subsequently detection by capillary electrophoresis (profile quality), with different injection times.

2.6.1. Sensitivity

The sensitivity test for the InnoQuant® HY kit was performed using a series of six dilutions (from 1 – 0.03125 ng/µL) from InnoQuant® HY DNA standard, tested in duplicate. As InnoQuant® HY DNA standard has an initial concentration of 100 ng/µL was proceeded to an initial 1:10 dilution yielding a new dilution with concentration of 10 ng/µL and then it proceeded further dilution 1:10 creating a dilution with 1 ng/µL from which was carried out the serial dilution required to this study. For the InnoTyper® 21 kit, this test was performed by the amplification of a series of five positive controls with different input volumes (5 µL, 4 µL, 3 µL, 2 µL, 1 µL) with concentrations ranging from 6.0 ng/µL to 0.12 ng/µL, prepared from InnoTyper® 21 DNA Control and tested in duplicates.

2.6.2. Precision and accuracy

In the study of the precision and accuracy of these kits, repeatability and reproducibility tests were performed. Thus, for the InnoQuant® HY kit these tests were based on the analysis of the C_T values obtained for the five standards used to create the calibration curve. These five standards were performed in duplicate and analyzed in four different

quantification runs, thus obtaining eight C_T values for each standard in all the four quantification runs. For the InnoTyper® 21 kit, the performance of these tests was based on the amplification of the same series of five positive controls used in sensitivity tests, with different injection times tested in duplicates. During the validation of the kits two different methodologies (manual and automatic) were used and the results obtained by both enable the verification of the reproducibility of the kits.

2.6.3. Mixture studies

This parameter was only evaluated for the validation of InnoQuant® HY kit since the InnoTyper® 21 kit presents limitations in the interpretation of samples containing DNA mixture when used in conventional capillary electrophoresis. The validity of this study was performed by preparing mixtures from a male sample and a female sample, with known concentrations. For this, both samples were normalized to a concentration of 0.5 ng/µL and the preparation of mixture set was performed as described in Table 2.9 in order to test the ability of InnoQuant® HY kit to determine the two contributors in these proportions.

Table 2.9 – Volume of each diluted DNA sample required to prepare the mixture set.

	Mixture Ratio						
	19:1	**9:1**	**3:1**	**1:1**	**1:3**	**1:9**	**1:19**
Female Sample	95µL	90 µL	75 µL	50 µL	25 µL	10 µL	5 µL
Male Sample	5 µL	10 µL	25 µL	50 µL	75 µL	90 µL	95 µL

2.6.4. Concordance study

The concordance study was conducted by comparing the results obtained with these new quantification and amplification kits using retrotransposons as genetic markers with the kits currently in use in the LPC, Quantifiler® Duo and GlobalFiler™, which use conventional STRs as genetic markers. Thus, from the samples selected from existing extracts of the LPC-PJ casework previously quantified with Quantifiler® Duo kit and amplified with GlobalFiler™ kit, and whose quantification results were not concordant with the genetic profiles produced, 132 samples were tested with InnoQuant® HY kit and the results obtained were compared with those previously obtained by Quantifiler® Duo. The same procedure was performed to the concordance study of amplification kit, in which the 132 samples quantified with InnoQuant® HY kit were amplified with InnoTyper® 21 kit and the genetic profiles produced were compared to previously obtained with the GlobalFiler™ kit.

2.7. Evaluation of InnoQuant® HY and InnoTyper® 21 kits in rootless hair analysis

To evaluate the performance of both kits in rootless hair samples, 40 samples were analyzed in which the existing roots were removed by cutting. The DNA extraction methodology was similar to the one currently used by LPC-PJ, with some modifications, mainly the samples incubation time (which is higher) and the addition of buffer AL. These samples, obtained from crime scenes, were quantified with InnoQuant® HY kit and amplified with InnoTyper® 21kit. Thus, this evaluation besides allowing to know the behavior of both kits for this type of samples, allowing to determine the level of degradation to the samples sent to the LPC-PJ (through the degradation index calculated by the InnoQuant® HY kit) and obtaining amplified fragments from rootless hair samples.

2.8. Statistical analysis

Data were statistically analyzed using Microsoft Excel 2010. T-student and Chi-square tests were performed to establish the significance of the results obtained. P values of 0.05 or lower were considered statistically significant. Results were expressed as the mean ± S.E.M. of the indicated number of experiments.

3. Results and Discussion

3.1. Sample selection

Sample selection, made from existing extracts of LPC-PJ, was accomplished through extensive analysis of all the exams admitted in the forensic specialty of Biology, since 2015. From each exam analyzed, several samples of blood, contact trace, teeth and hairs were chosen. This choice was based on the results of the genetic profile (no results, inconclusive, complete, and/or possible degradation/inhibition) produced by the amplification kit currently at use in the LPC-PJ as well as in the quantification results presented. After, samples were divided into six separate groups according to the genetic profile presented, as represented in Table 3.1 (Appendix IV).

Table 3.1 – Type of selected samples and their division according to the genetic profile presented.

Group		Sample	Ext	Trace	Quantification value (QD)	Type of profile
I		1	430-11	Hair	----------	No Results
		2	442-02	Hair	----------	
		3	442-07	Hair	----------	
II		40	442-01	Hair	----------	Inconclusive
		42	358-03	Hair	0.003	
		43	442-04	Hair	----------	
III		56	430-04	Hair	0.001	Complete
		57	447-08	Hair	0.024	
		58	442-09	Hair	0.087	
IV	IPC +	64	452-12	Biologic	----------	No Results
		65	454-06	Biologic	----------	No Results
		66	454-06	Biologic	----------	No Results
V	Degraded DNA	84	465-02	RS	0.006	Inconclusive
		85	457-17	RS	0.006	
		86	457-03	RS	0.010	
		87	447-05	Hair	0.011	
		88	375-24	RS	0.012	
VI	Degraded or inhibited DNA	113	101-01	Teeth	----------	No Results
		114	113-02	Teeth	----------	No Results
		115	113-01	Teeth	0.001	No Results
		116	113-03	Teeth	0.260	Complete
		118	126-04	Teeth	0.008	Inconclusive
		119	122-03	Teeth	----------	No Results

Additionally, another set of 40 rootless hair samples extracted using manufacturer's instructions (*InnoGenomics Technologies*) was included to determine the behavior of the InnoQuant® HY and InnoTyper® 21 kits in this type of samples, often sent to the LPC-PJ and that mostly do not produce results.

To internally validate the InnoQuant® HY and InnoTyper® 21 kits, the InnoQuant® HY DNA Standard and the InnoTyper® 21 DNA Control, respectively, were used as samples.

3.2. Internal Validation of InnoQuant® HY kit

The DNA quantification of a given sample is an extremely important step in forensic laboratories, since it allows to determine not only the approximated concentration of DNA present in a given sample but also its quality, and thus select the most suitable amplification method to be used. Thus, knowing the best amplification kit to be used and the optimal volume of DNA required for the run, allows saving time and financial resources.

One the mainly aims of the present study was the internal validation of InnoQuant® HY kit, a human and male quantification and degradation assessment kit, for later implementation in the LPC-PJ laboratory routine. To this, sensitivity, repeatability, and reproducibility parameters, as well as mixtures studies were evaluated.

3.2.1. Sensitivity

To study the sensitivity of the quantification kit was prepared a serial dilution of the InnoQuant® HY DNA Standard (1.0 ng/µL; 0.5 ng/µL; 0.250 ng/µL; 0.125 ng/µL; 0.063 ng/µL and 0.031 ng/µL) and tested in duplicate. An intermediate dilution of 10 ng/µL (performed due to the fact that the standard DNA having an initial concentration of 100 ng/µL) it was also analyzed. All dilutions were quantified according to manufacturer's instructions, and using two different methodologies - manual and automated (QIAgility™ robot). The results for the short target are described in the Table 3.2. The quantification results for all the targets analyzed by InnoQuant® HY, for both methodologies can be observed in Appendix V.

The aim of this study was to evaluate the kit ability to measure very low DNA concentrations, as well as comparing expected results with the respective quantification values obtained.

Table 3.2 – Quantification results obtained for the intermediate and serial dilution with the InnoQuant® HY kit using manual and QIAgility™ robot procedures.

Dilution	Expected values (ng/µL)	Observed values (ng/µL) - QIAgility™ robot		Observed values (ng/µL) - Manual
		1st repetition	2nd repetition	1st repetition
Intermediate	10	13.067	11.288	Not analyzed
Dil.1	1.0	1.097	1.133	0.900
Dil.2	0.500	0.676	0.601	0.463
Dil.3	0.250	0.284	0.299	0.215
Dil.4	0.125	0.143	0.141	0.124
Dil.5	0.063	0.064	0.071	0.056
Dil.6	0.031	0.034	0.038	0.033

Using both methods, the concentrations obtained for the serial dilutions were very similar to the expected theoretical concentrations. In intermediate dilution, the observed values were slightly different, although not significant [$p > 0.05$ (Appendix V)], from the expected theoretical values, a fact that may be attributed to pipetting errors made by the QIAgility™ robot or to the dilution factor, since the variability encountered with pipetting larger dilutions resulted in larger deviations from the expected values. Also, the quantification method did not provide the exact amount of DNA present in a given sample, but approximate results. Thus, it was verified that this quantification kit was able to quantify samples containing both large and reduced DNA concentrations, presenting, therefore, good sensitivity.

3.2.2. Repeatability and Reproducibility

To study repeatability and reproducibility was performed the analysis of C_T values obtained for the five standards used to create the calibration curve (four independent quantification runs). Three of these quantification plates were performed by QIAgility™ robot, at different times, and the fourth run was manually prepared and applied in the following day. Thus, as the same standards were prepared and applied at different times and on different days, being possible to verify the repeatability of the kit. The preparation and application of standards using two different operators (manual and automated) allowed evaluating its reproducibility. All quantification plates were run on ABI 7500 HID Real-Time PCR System.

For each of the three quantification runs performed for the evaluation of these parameters, the QIAgility™ robot prepared dilutions from InnoQuant® DNA HY Standard solution, as described in the kit user manual. Thus, five standards were obtained, with concentrations of 20 ng/µL; 2.5 ng/µL; 0.3125 ng/µL; 0.0391 ng/µL and 0.005 ng/µL. Each standard was applied in duplicate in each run, producing six C_T values (C_T 1 to C_T 6) in the set of three automated runs. To perform the fourth run, the same dilutions were manually prepared and applied, in duplicate, producing more two C_T values: C_T7 and C_T8.

The C_T values for each standard of the short target in the four quantification runs prepared are illustrated in Figure 3.1 (Appendix VI).

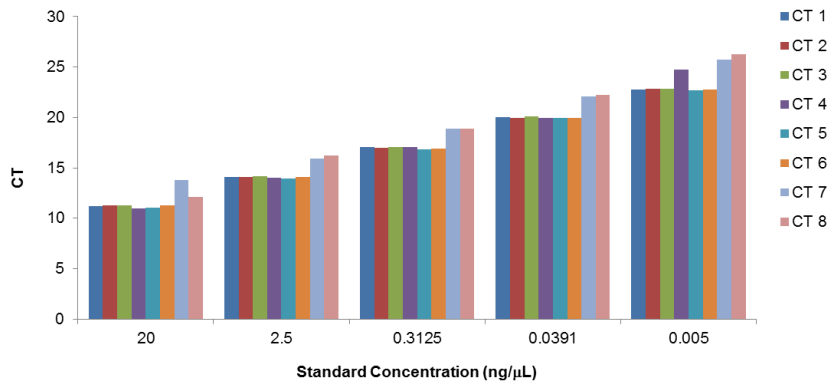


Figure 3.1– C_T values obtained for the short target of five Standards in the three quantification runs.

Considering the same run, no significant statistically differences of C_T values between duplicates were observed, as well as among the different standards ($p > 0.05$). The same was observed for the fourth run, except for the standards 1 and 5 (20 ng/µL and 0.005 ng/µL, respectively) whose values between duplicates were different, although without compromising the results (Appendix VI). This difference may be related to the fact that these standards presented higher and lower concentrations, respectively.

Also, as the concentration of DNA standards decreases, an increase of the C_T values was detected. This was observed in all quantification runs and presented the inverse proportionality between the concentration values of the DNA standards and the C_T values. These data demonstrated the high consistency of the kit.

Comparing the C_T values obtained for the three quantification runs prepared by the robot (C_T1 to C_T6) with the C_T values in the manual run (C_T7 and C_T8), an increase in the C_T values for this last run was observed. This variability may be associated with a variation in the machine conditions since the runs were performed on different days by different operators, as well as the fact that the preparation of the standards for the manual run was performed from a new InnoQuant® DNA HY Standard solution. However, it is important to note that these changes observed for the C_T values between runs were not significant ($p > 0.05$) (Appendix VI) and, therefore, did not affect the kit reproducibility, since all values were within the optimal range of 30 C_T stipulated by manufacturers.

For each of the three quantification runs prepared by the robot, the respective calibration curve for DNA concentration was drawn, using the C_T average values from each standard. It was also drawn a calibration curve for the set of the runs (Av. run), considering the average values of total C_T obtained for each of the five standards. These calibrations curves are illustrated in Figure 3.2. The calibration curve in function of the

DNA concentration logarithm was also designed to the fourth quantification run and is illustrated in Figure 3.3.

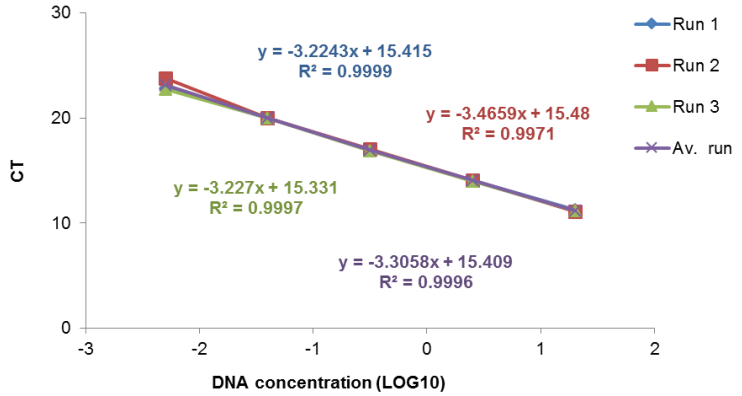


Figure 3.2 - Calibration curves designed for each quantification run performed by the robot as well as for the average run (Av. run).

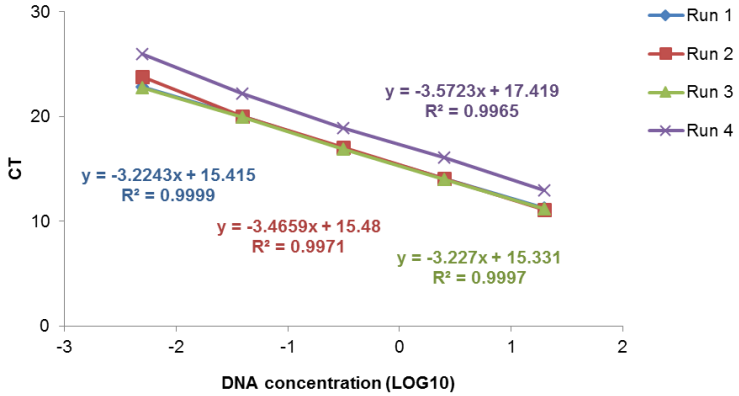


Figure 3.3 - Calibration curve designed for all four quantification runs.

In all calibration curves, was observed a linear regression with negative slope showing the existence of inverse proportionality between the C_T values and the DNA concentration. All calibration curves, including the calibration curve for the set of the three runs (Av. run), were similar, substantially overlapping with each other with the exception of run 4 that although similar, was parallel to the others. All curves presented correlation coefficients (R^2) higher than 0.99, demonstrating primer efficiency and showing the proximity between the regression line and the C_T values for each standard. Thus, the consistency of results for the C_T values of each of the analyzed standards, demonstrated the kit repeatability and reproducibility.

3.2.3. Mixture studies

The ability to quantify different concentrations of male and female DNA in mixtures was evaluated by the analysis of quantification results for each Male : Female (M:F) ratio. For

that, two samples - a female and a male sample - with known concentrations (5.68 ng/µL and 4.17 ng/µL, respectively) were selected. As the samples selected for preparing the mixtures contained different concentrations, were normalized to 0.5 ng/µL by diluting the initial extract with TE. Then, was performed the mixture of both samples to prepare seven solutions with different M:F ratios (Table 3.3). These seven solutions were manually prepared and distributed in quantification plate using the QIAgility™ robot.

Table 3.3- Quantification results obtained for the different ratios M:F from prepared mixtures.

M:F ratio obtained (ng/µL)	Total DNA (ng/µL)	Male DNA (ng/µL)	Female DNA (ng/µL)	M:F ratio obtained (ng/µL)
19:1	0.182	0.175	0.006	25:1
9:1	0.168	0.148	0.020	7.40:1
3:1	0.188	0.128	0.060	2.13:1
1:1	0.180	0.083	0.097	1:1.17
1:3	0.215	0.051	0.163	1:3.13
1:9	0.266	0.030	0.235	1:7.83
1:19	0.227	0.014	0.213	1:15.21

The concentration values of total DNA and male DNA resulted from the quantification results obtained for the short and Y targets, respectively. The concentration of female DNA is an approximate value, resulting from the subtraction of the Male DNA concentration to the total DNA concentration.

The analysis of the quantification values present in Table 3.3 demonstrated that the values obtained for the quantification of total DNA were different from the expected theoretical concentrations of 0.5 ng/µL. However, it was possible to observe that the M:F ratios obtained were similar to the expected theoretical M:F ratios, suggesting variations during pipetting that can possibly had occurred during mixtures preparation. Analyzing the quantification values obtained for both male and female DNA, was possible to determinate the major and minor contributors for each sample, attesting the kit's ability to discriminate between two different contributors with different proportions. Thus, the results obtained for this parameter demonstrated the ability of the InnoQuant® HY kit to quantify different concentrations of male and female DNA, in mixtures with different M:F ratios.

3.2.4. Concordance study

A concordance study was performed to evaluate the performance of InnoQuant® HY kit in challenging samples in forensic analysis, when compared with the Quantifiler® Duo kit that is currently in use at the LPC-PJ. A total of 132 samples (including three control samples) was selected from existing extracts of the LPC-PJ casework, previously quantified with Quantifiler® Duo kit and amplified with GlobalFiler™ kit. This set of

samples presented quantification results non-concordant with the genetic profiles produced. The quantification results obtained with InnoQuant® HY kit and those previously obtained by Quantifiler® Duo are illustrated in Figure 3.4.

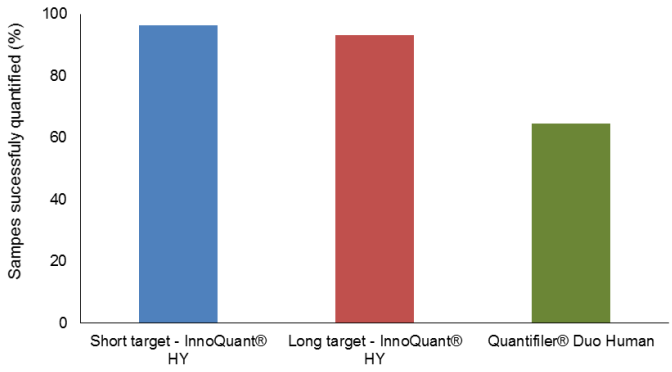


Figure 3.4 - Percentage of quantified samples for each target of the InnoQuant® HY and Quantifiler® Duo kits (n=132).

Among the 132 samples analyzed, three presented no quantification values for any of the kits tested. This indicates that these samples contained no DNA, possibly due to a less accomplished extraction. Also, approximately 96.2% and 93.2% (Short and Long targets, respectively) of the casework samples analyzed were successfully quantified with InnoQuant® HY kit. Regarding Quantifiler® Duo kit, approximately 64.4% of the samples presented quantification values.

To make a comparison between the two kits, was necessary to note that the Quantifiler® Duo kit quantifies the total amount existing in the sample (Human target), which induces that the comparison should be made with the short target of InnoQuant® HY. However, the amplicon length quantified by Human and Short target is different (140 bp and 80bp, respectively), which could lead to doubts due to the quality of DNA present in the sample. Therefore, the comparison was performed between the Long target of InnoQuant® HY kit and the Human target of Quantifiler® Duo kit. Thus, the comparison between these two targets demonstrated that the InnoQuant® HY was able to produce quantification results in nearly 28.8% more of the analyzed samples than with the Quantifiler® Duo kit, indicating that the InnoQuant® HY was more effective in treating challenging samples.

In addition to this ability to produce better results, InnoQuant® HY also allowed calculating the degradation index (DI) of samples, was absent in the kit used currently in LPC-PJ. This DI was determined by the ratio between the concentrations of long and short targets as presented in the following equation:

$$\text{Equation 3.1: } \mathbf{DI} = \frac{[\mathbf{Short}]}{[\mathbf{Long}]}$$

As mentioned above, the short target results represent the concentration of total DNA present in the sample, while the long target contribute to determining the degradation index. The first target to express changes when the sample is degraded is the long target due to the susceptibility conferred by its size. Therefore, using the calculated DI was possible observe that the samples analyzed possessed some level of degradation (Table 3.4).

Table 3.4– Degradation index in the set of analyzed samples.

DI	Number of Samples	Percentage	Number of Samples with QDuo values > 0.1 ng/uL	Full Profile	Percentage
< 3	28	22.8	8	8	100
3-5	33	26.8	1	0	0
6-10	24	19.5	2	1	50
11-15	6	4.9	0	0	0
16-20	6	4.9	0	0	0
21-30	9	7.3	0	0	0
31-60	6	4.9	0	0	0
> 60	11	8.9	0	0	0

The results for this parameter indicated that 46.3% of the analyzed samples had a DI between 3 and 10, while 30.9% had DI values superior to 10. These values indicated that a significant percentage of the samples analyzed (77.2%) had moderate to high degradation. These high sample degradation values can cause problems in assessing the amount of DNA input for amplification which may lead to subsequent rework. Of the 28 samples that had an index of degradation lower than three, only 13 had a degradation index of 1 (not degraded). These samples were distributed among sub groups II, III, IV and VI (inconclusive, complete, IPC^{+}, and inhibited DNA, respectively).

Thus, this new tool provided by the InnoQuant® HY allowed a prior knowledge about the DNA quality present in the samples and, therefore, to determine the best strategy to be adopted for the production of results, minimizing unnecessary re-amplifications.

3.3. Internal Validation of InnoTyper® 21 kit

Determining the genetic profile of an individual in a given sample is the purpose of forensic genetics, whether in a criminal investigation or missing person identification. For this, amplification methods of the DNA extracted from these samples is a crucial step and as such, the most important. As the DNA from forensic samples typically has low quality and quantity, there arises the need to implement an amplification method which overcomes these problems and permits obtaining genetic profiles in this type of samples. Thus, the development of a kit with the ability to amplify fragments as small as 125 bp becomes an advantage for forensic laboratories.

The internal validation of the InnoTyper® 21 kit in LPC-PJ to enable its subsequent implementation in the laboratory routine for the analysis of this type of challenging samples was the main objective of this work, and for that, sensitivity, repeatability and reproducibility tests were performed.

3.3.1. Sensitivity

The optimal DNA template target for the InnoTyper® 21 kit is between 0.2 – 0.5 ng/µL. Thus, to study the sensitivity of this kit were prepared five Positive Controls (PC1 to PC5) from InnoTyper® 21 DNA Control with different concentrations ranging from 0.6 ng/µL to 0.12 ng/µL. All solutions were prepared and amplified according to manufacturer's instructions and applied in duplicate. The results of amplification obtained through capillary electrophoresis are described in Table 3.5. The aim of this study was to evaluate the kit capacity to produce good quality genetic profiles outside their optimal range of operation.

Table 3.5 – Number of amplified alleles and respective percentage for each of the Positive Controls analyzed.

Samples	Concentration (ng/µL)	number of alleles obtained		Percentage (%)	
		1st repetition	2nd repetition	1st repetition	2nd repetition
PC1	0.12	34	28	80.95	66.67
PC2	0.24	40	40	95.24	95.24
PC3	0.36	40	41	95.24	97.62
PC4	0.48	42	41	100	97.62
PC5	0.60	42	42	100	100
NTC	0	0	0	0	0

The kit was able to amplify DNA quantities above and below the optimal DNA template target range. However, when the DNA concentration was below the minimum limit of optimal DNA input, the percentage of amplified alleles was considerably reduced. Also, it was possible to observe a small difference between the first and the second repetition performed ($p > 0.05$). However, this difference did not compromise the results (Appendix VII). Concerning the concentrations within optimal amplification range, the results were not in accordance with the expected since complete profiles were not obtained. However, the number of amplified alleles was similar to the entire profile, and an amplification failure of the same marker in the Positive Controls 2, 3 and 4 was observed. These small discrepancies observed may be related to amplification problems.

In general, the results obtained from the analysis of this parameter allowed to confirm the sensitivity of the InnoTyper® 21 kit, either to amplify samples with low concentrations or with concentrations above the optimal amplification range.

3.3.2. Repeatability and Reproducibility

For the study of repeatability and reproducibility was proceeded to the analysis of genetic profiles obtained for the five Positive Controls (PC1 to PC5) used for the sensitivity test, performed in duplicate and run with three different injection times: 16 seconds, 18 seconds (default), and 20 seconds. The percentage of amplified alleles in each positive control is illustrated in Figure 3.5 (Appendix VIII).

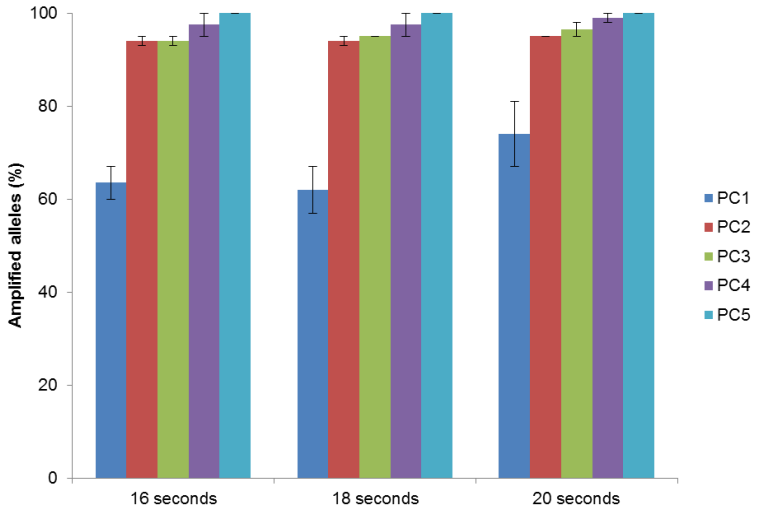


Figure 3.5 - Percentage of amplified alleles for each Positive Control run in replicates with three different injection times. Data are mean ± stdev of two experiments.

Through individual assessment of each capillary electrophoresis run with the different injection times, was possible to observe that the percentage of amplified alleles for each control was similar among duplicates. Also, it was possible to observe a large percentage of amplified alleles for all the controls (superior to 94%) except for the PC1 whose percentage of amplified fragments was expressively lower (between 64% and 74%) when compared with the others, however not significant ($p > 0.05$) (Appendix VIII). This discrepancy may be related to the low concentration of DNA existing in this sample. The analysis of Figure 3.5 also allowed observing that no significant differences were detected between the different runs performed, except in the second repetition, where there was a decrease, although not significant ($p < 0.05$), in the number of amplified fragments that was constant in all the controls and runs (Appendix VIII). The quality of the amplification results can also be related to the fact that the reagents were not properly mixed, yielding lower peak height when compared with in-house manufacturer's results (Figure 3.6).

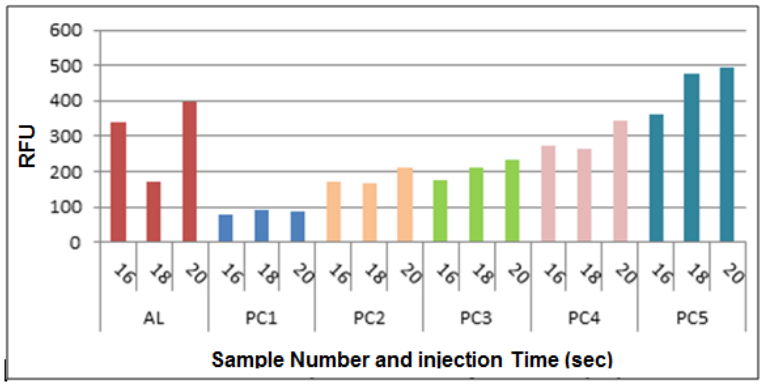


Figure 3.6 - Peak heights average, in RFU, for each of the controls analyzed and Allelic ladder.

Overall, in the three runs performed, the percentage of amplified alleles for each of the five controls was similar. Thus, from the genetic profiles obtained for these two parameters, was possible prove the repeatability of InnoTyper® 21 kit, as well as its reproducibility, allowing to internally validate this new amplification kit.

3.3.3. Concordance Study

As described for quantification kit, this study was performed to evaluate the performance of InnoTyper® 21 kit in challenging samples for forensic analysis, comparing this new kit with GlobalFiler™, the kit currently in use at the LPC-PJ. Therefore, from the 132 samples previously quantified with the InnoQuant® HY kit, 127 were selected for subsequent amplification with the InnoTyper® 21 kit, using a DNA input of 0.5 ng/µL. The results obtained by InnoTyper® 21 kit, as well as the results obtained by GlobalFiler™ kit are illustrated Figure3.7.

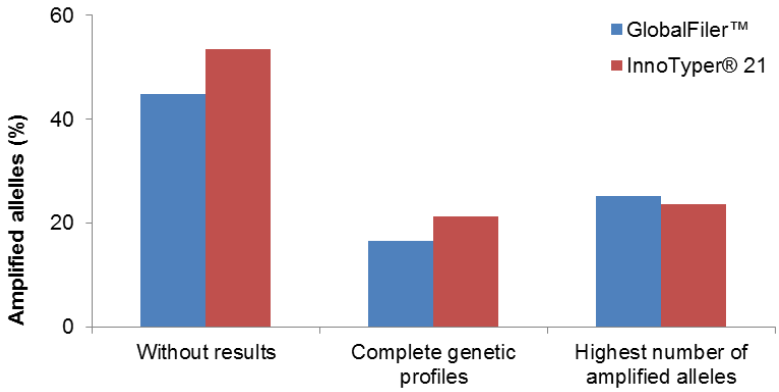


Figure 3.7 - Amplified alleles from the samples analyzed with InnoTyper® 21 and GlobalFiler™ kits (n=127).

The comparison between the results obtained by the two kits demonstrated that both presented difficulties in the amplification of the samples selected, and with this new

amplification kit nearly 9% more samples without results were obtained. Although InnoTyper® 21 kit produced more complete genetic profiles (27 vs. 21 obtained by GlobalFiler™), overall, the GlobalFiler™ kit originated approximately 2% more genetic profiles with the highest number of amplified alleles. However, it should be noted that the amplification with the GlobalFiler™ kit was performed with fresh extracts, whereas when it proceeded to amplification with InnoTyper® 21 kit, the selected extracts were subjected to thawing and, in some cases, the samples had already undergone at least one thawing and refreezing process before being used by InnoTyper® 21 kit. Therefore, this defrosting and refreezing of samples and the use of samples extracted by another method than the recommended by the manufacturer can be the determining factor in explaining the difference between the two kits compared in this study.

Overall, the results demonstrated that for challenging samples and under the conditions used in this study, the differences observed between the two kits were not significant. Therefore, the utility of InnoTyper® 21 has been proven, especially by the successful amplification of a greater number of complete genetic profiles.

3.4. Evaluation of InnoQuant® HY and InnoTyper® 21 kits for DNA analysis of rootless hair

One of the main goals of the development of these two retrotransposons-based kits was, somehow, to facilitate the forensic analysis of challenging samples with a limited number of complete genetic profiles. Therefore, in this study, challenging samples collected from real crime scenes were used to evaluate the performance of InnoQuant® HY and InnoTyper® 21 kits. A set of 40 strands of hair received from different crime scenes, were analyzed under the microscope for the detection and removal of the root. The processed hair fragments were subjected to several washes and subsequently to DNA extraction using the extraction method mentioned above. The obtained DNA extracts were quantified with InnoQuant® HY kit (only was possible to quantify 17 from the 40 samples, due to kit lack), and sample degradation index was calculated. Finally, the amplification of theses DNA extracts (DNA input of 0.5 ng/µL) with InnoTyper® 21 kit was performed. This set of samples was amplified with the GlobalFiler™ kit, to compare the two kits accurately. The quantification results, as well as the degradation index for the 17 samples analyzed with InnoQuant® HY kit are presented in Table 3.6.

Table 3.6– Quantification results obtained for Short, Long and Y targets and respective degradation index (n=17).

Sample	Short (ng/µL)	Long (ng/µL)	Y (ng/µL)	Degradation Index
14	0.000041	0.000004	-	9.44
15	0.000026	0.000001	-	22.65
16	0.000030	0.000001	-	42.35
17	0.000100	0.000017	0.000009	5.81
18	0.000020	0.000000	-	81.32
19	0.000021	0.000001	-	17.73
20	0.000020	0.000002	0.000011	9.13
21	0.000013	0.000001	-	15.04
22	0.000014	0.000001	-	13.95
23	0.000017	0.000001	-	12.20
24	0.000015	0.000000	-	29.77
25	0.000014	-	-	Undetermined
26	0.000021	0.000000	-	83.01
27	0.000024	0.000001	-	46.25
28	0.000149	0.000012	-	12.27
29	0.000043	0.000003	-	16.81
30	0.000189	0.000013	-	14.70

All the tested samples presented degradation. Among these samples, 17.65% had moderate degradation, while 76.47% presented a very high degradation rate. For one of the analyzed samples (sample 25), the InnoQuant® HY kit was unable to quantify the long target, indicating that the existing DNA was fully degraded to fragments smaller than 125 bp. These results allowed to determine the best amplification method to used, since the probability of obtaining results with this type of samples will be quite difficult due the highly fragmentation of the DNA.

Based on the quantitation results, these samples were amplified with InnoTyper® 21 and GlobalFiler™ kits, taking into account the DNA input required. The percentages of the amplified alleles for both kits are illustrated in Figure 3.8.

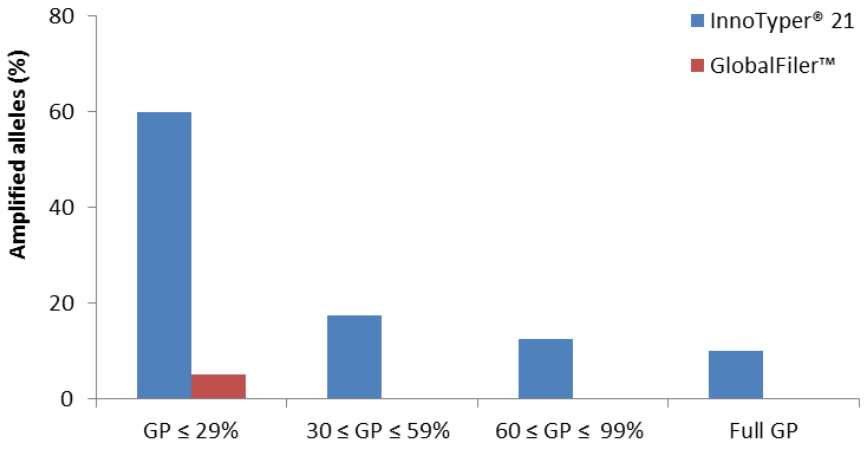


Figure 3.8 - Percentages of amplified alleles for InnoTyper® 21 and GlobalFiler™ kits (n= 40).

The analysis of the genetic profiles obtained for each amplification kit allowed to verify that the InnoTyper® 21 kit was able to amplify fragments in the 40 samples while the GlobalFiler™ kit only produced results in two samples. Also, it was possible to observe that in nearly 23% of the samples, the retrotransposon-based kit presented genetic profiles with quality to be reported (Appendix IX), of which 10% presented complete genetic profiles (Appendix XI). It was also possible to observe that 18% of the samples presented between 30 to 59% of amplified alleles in their genetic profile. Thus, the comparison between the results produced by both kits demonstrated that the InnoTyper® 21 kit produced better results than GlobalFiler™ kit when operated in fresh extracts with a modified extraction method. Additionally, it was possible to find that many of the samples exhibited amplification peaks in genetic profile (nearly 60%). However, these samples did not have quality to be reported since due to the reduced RFU peak size (Appendix XI). As was already verified during the internal validation of this new kit, the peak heights were lower indicating that the reagents necessary for amplification were not be properly mixed by the robot used for sample preparation.

In general, the results obtained for rootless hair samples allowed to confirm that both InnoQuant® HY and InnoTyper® 21 kits represent very good alternatives for the analysis of challenging samples frequently received in forensic laboratories. These results also demonstrated the importance of having knowledge on the quality of DNA present in a given sample, because a high degradation rate leaded to a decreased quality of the genetic profiles. Thus, the InnoQuant® HY can be used to predict a particular genetic profile (via DI) and, thus, to determine the quantity of DNA input to be added to the amplification reaction to overcome the degradation effect and obtain better results.

4. Conclusion

Obtaining a genetic profile from pieces of evidence collected at a crime scene is the primary objective of forensic laboratories. However, most samples sent for forensic laboratories do not present good DNA quality and/or quantity to be amplified. In this sense, the quantification of the samples is crucial in the forensic analysis since it allows knowing the amount of DNA available in a given sample and thus, determining the better amplification method.

The present study aimed to evaluate and internally validate two new retrotransposons-based kits, InnoQuant® HY and InnoTyper® 21, for its later implementation in the routine laboratory of the Laboratório de Polícia Científica da Polícia Judiciária to analyze challenging samples.

To internally validate both InnoQuant® HY and InnoTyper® 21 kits the following parameters were studied: sensitivity, repeatability and reproducibility, mixtures (only for the InnoQuant® HY kit), and concordance. Additionally, the functionality of both kits for results production in rootless hair samples was also evaluated.

The sensitivity of the InnoQuant® HY kit was assessed using a serial dilution with six different concentrations, as well as an intermediate dilution. The results obtained confirmed the sensitivity of the kit in the quantification of high DNA concentrations, as well as in the quantification of very low concentrations of DNA, even revealing that the concentrations obtained for the serial dilutions were similar to the expected theoretical concentrations. The sensitivity of the InnoTyper® 21 kit was evaluated using five Positive Controls with DNA concentration values higher and lower the optimal DNA template target range. The results obtained proven the sensitivity of the InnoTyper® 21 kit to amplify samples with concentrations higher and lower their optimal amplification range.

In the repeatability and reproducibility studies were evaluated the five standards used to obtain a calibration curve to InnoQuant® HY kit and the five Positive Controls for InnoTyper® 21 kit. The consistency of results for the C_T values of each of the analyzed standards in the four quantification runs performed, proved the InnoQuant® HY kit repeatability and reproducibility. For InnoTyper® 21 kit, the percentage of amplified alleles for each control conducted in both repetitions was similar to each other as well as along the three different runs performed, proving the kit repeatability and reproducibility.

Regarding the mixtures study, performed only for quantification kit, was observed that the InnoQuant® HY kit was able to determine different concentrations of male DNA *versus* female DNA in mixtures with different ratios M:F, and so determinate the major and minor contributors in each sample. These results attested the kit ability to discriminate between two different contributors with different proportions.

Even within the internal validation, were performed concordance studies between InnoQuant® HY and InnoTyper® 21 kits and kits currently in use at the LPC-PJ - Quantifiler® Duo and GlobalFiler™, respectively. The results obtained allowed to determine that InnoQuant® HY had the ability to produce quantification results in 29% more of the analyzed samples than the Quantifiler® Duo kit, making the InnoQuant® HY kit more efficient in the analysis of challenging samples. Additionally, the InnoQuant® HY kit also allowed to calculate sample degradation index (DI), the missing tool in the kit used currently in LPC-PJ, allowing a better approach during the amplification step. Comparison the results obtained with both amplification kits, it was demonstrate that in casework samples with freezing and thawing cycles, the differences observed between them were not significant. However, it can be hypothesized that the fact that the InnoTyper® 21 kit was operated with extracts subjected to these conditions may have impaired the results produced by this kit since the results produced by GlobalFiler™ were obtained from fresh extracts.

Thus, the results obtained internally validated both InnoQuant® HY and InnoTyper® 21 kits, allowing its implementation in the routine laboratory at the LPC-PJ for the treatment of challenging samples.

Finally, the evaluation of the functionality of both InnoQuant® HY and InnoTyper® 21 kits was performed in rootless hair samples, collected from real crime scenes. The results demonstrated that InnoQuant® HY kit was able to determine the concentration of DNA present in each sample as well as its degradation index. Among these samples, 17.65% presented moderate degradation while 76.47% exhibited a very high degradation rate. Through these results, it will be possible for a forensic expert to select the best amplification method to apply and the DNA input to be added to the amplification reaction. Thus, in samples with higher DNA concentration was possible to obtain a more complete genetic profile. The analysis of the genetic profiles obtained allowed verifying that the InnoTyper® 21 kit was able to amplify fragments in all the analyzed samples, with nearly 23% of these samples producing genetic profiles with quality to be reported, of which 10% presented complete genetic profiles (Appendix IX). Also, it was observed

that the amplification of these samples with the GlobalFiler™ kit failed, allowing to demonstrate that the InnoTyper® 21 kit was most efficient when operated with freshly obtained extracts.

In general, these results proved that both InnoQuant® HY and InnoTyper® 21 kits represent very good alternatives for the analysis of challenging samples, which are the most commonly received in forensic laboratories.

References

1. Osterburg, J.W. and R.H. Ward, *Criminal investigation: A method for reconstructing the past*. 2010: Routledge.

2. Fisher, B.A. and D.R. Fisher, *Techniques of crime scene investigation*. 2012: crc Press.

3. Butler, J.M., et al., *Forensic DNA typing by capillary electrophoresis using the ABI Prism 310 and 3100 genetic analyzers for STR analysis.* Electrophoresis, 2004. 25(10-11): p. 1397-1412.

4. Edlund, H., *Sensitive Identification Tools in Forensic DNA Analysis.* 2010.

5. Goodwin, W., A. Linacre, and S. Hadi, *An introduction to forensic genetics*. Vol. 2. 2011: John Wiley & Sons.

6. Butler, J.M., *Forensic DNA typing: biology, technology, and genetics of STR markers*. 2005: Academic Press.

7. Chen, B.-Y. and H.W. Janes, *PCR cloning protocols*. Vol. 192. 2002: Springer Science & Business Media.

8. Innis, M.A., et al., *PCR protocols: a guide to methods and applications*. 2012: Academic press.

9. Butler, J.M. and C.H. Becker, *Improved analysis of DNA short tandem repeats with time-of-flight mass spectrometry*. 2001: US Department of Justice, Office of Justice Programs, National Institute of Justice.

10. Hameed, I.H., et al., *Genetic variation and DNA markers in forensic analysis.* African Journal of Biotechnology, 2014. 13(31).

11. Ruitberg, C.M., D.J. Reeder, and J.M. Butler, *STRBase: a short tandem repeat DNA database for the human identity testing community.* Nucleic Acids Research, 2001. 29(1): p. 320-322.

12. Butler, J.M., *Advanced topics in forensic DNA typing: methodology*. 2011: Academic Press.

13. Zhong, Y., B. Budowle, and F. Center, *The utility of short tandem repeat loci beyond human identification: implications for development of new DNA typing systems.* Electrophoresis, 1999. 20: p. 1682-1696.

14. Butler, J.M., *Short tandem repeat typing technologies used in human identity testing.* Biotechniques, 2007. 43(4): p. 2-5.

15. Fondevila, M., et al., *Challenging DNA: assessment of a range of genotyping approaches for highly degraded forensic samples.* Forensic Science International: Genetics Supplement Series, 2008. 1(1): p. 26-28.

16. Constantinescu, C.M., et al., *Challenging DNA samples solved with MiniSTR analysis.* Rom J Leg Med, 2012. 20: p. 51-56.

17. Barbaro, A., P. Cormaci, and G. Falcone, *Validation of BTA™ lysis buffer for DNA extraction from challenged forensic samples.* Forensic Science International: Genetics Supplement Series, 2011. 3(1): p. e61-e62.

18. Almeida, M., et al., *Efficient DNA extraction from hair shafts.* Forensic Science International: Genetics Supplement Series, 2011. 3(1): p. e319-e320.

19. Hughes, C., *Challenges in DNA Testing and Forensic Analysis of Hair Samples.* Forensic Magazine, 2013: p. 1-4.

20. Ayarpadikannan, S. and H.-S. Kim, *The impact of transposable elements in genome evolution and genetic instability and their implications in various diseases.* Genomics & informatics, 2014. 12(3): p. 98-104.

21. Cordaux, R. and M.A. Batzer, *The impact of retrotransposons on human genome evolution.* Nature Reviews Genetics, 2009. 10(10): p. 691-703.

22. Habibi, L., et al., *Mobile DNA Elements: The Seeds of Organic Complexity on Earth.* DNA and cell biology, 2015. 34(10): p. 597-609.

23. Muñoz-López, M. and J.L. García-Pérez, *DNA transposons: nature and applications in genomics.* Current genomics, 2010. 11(2): p. 115-128.

24. Lodish, H., et al., *Molecular Cell Biology*

Seventh Edition. 7 ed. 2012: Freeman & Company, W. H. 1247.

25. Callinan, P. and M. Batzer, *Retrotransposable elements and human disease*, in *Genome and disease*. 2006, Karger Publishers. p. 104-115.

26. Burns, K.H. and J.D. Boeke, *Human transposon tectonics.* Cell, 2012. 149(4): p. 740-752.

27. Krebs, J.E., et al., *Lewin's genes XI*. 2014: Jones & Bartlett Publishers.

28. Feschotte, C. and E.J. Pritham, *DNA transposons and the evolution of eukaryotic genomes.* Annual review of genetics, 2007. 41: p. 331.

29. Russell, P.J., *IGenetics*. 2002: Benjamin Cummings San Francisco.

30. Levin, H.L. and J.V. Moran, *Dynamic interactions between transposable elements and their hosts.* Nature Reviews Genetics, 2011. 12(9): p. 615-627.

31. Slotkin, R.K. and R. Martienssen, *Transposable elements and the epigenetic regulation of the genome.* Nature Reviews Genetics, 2007. 8(4): p. 272-285.

32. Piégu, B., et al., *A survey of transposable element classification systems–a call for a fundamental update to meet the challenge of their diversity and complexity.* Molecular phylogenetics and evolution, 2015. 86: p. 90-109.

33. Wang, J., et al., *dbRIP: A highly integrated database of retrotransposon insertion polymorphisms in humans.* Human Mutation, 2006. 27(4): p. 323-329.

34. Hancks, D.C. and H.H. Kazazian, *Active human retrotransposons: variation and disease.* Current opinion in genetics & development, 2012. 22(3): p. 191-203.

35. Faulkner, G.J., *Retrotransposons: Mobile and mutagenic from conception to death.* FEBS Letters, 2011. 585(11): p. 1589-1594.

36. Hulme, A.E., et al., *The Impact of LINE-1 retro transposition on the human genome*, in *Genomic Disorders*. 2006, Springer. p. 35-55.

37. Ray, D.A., *SINEs of progress: Mobile element applications to molecular ecology.* Molecular Ecology, 2007. 16(1): p. 19-33.

38. Mita, P. and J.D. Boeke, *How retrotransposons shape genome regulation.* Current opinion in genetics & development, 2016. 37: p. 90-100.

39. Thompson, P.J., T.S. Macfarlan, and M.C. Lorincz, *Long Terminal Repeats: From Parasitic Elements to Building Blocks of the Transcriptional Regulatory Repertoire.* Molecular cell, 2016. 62(5): p. 766-776.

40. Han, J.S., *Non-long terminal repeat (non-LTR) retrotransposons: mechanisms, recent developments, and unanswered questions.* Mobile DNA, 2010. 1(1): p. 1.

41. Muotri, A.R., et al., *The necessary junk: new functions for transposable elements.* Human molecular genetics, 2007. 16(R2): p. R159-R167.

42. Mills, R.E., et al., *Which transposable elements are active in the human genome?* TRENDS in Genetics, 2007. 23(4): p. 183-191.

43. Cordaux, R., et al., *In search of polymorphic Alu insertions with restricted geographic distributions.* Genomics, 2007. 90(1): p. 154-158.

44. Bennett, E.A., et al., *Active Alu retrotransposons in the human genome.* Genome research, 2008. 18(12): p. 1875-1883.

45. Hedges, D.J., et al., *Mobile element-based assay for human gender determination.* Analytical biochemistry, 2003. 312(1): p. 77-79.

46. Hormozdiari, F., et al., *Alu repeat discovery and characterization within human genomes.* Genome research, 2011. 21(6): p. 840-849.

47. Teixeira-Silva, A., et al., *The role of recombination in the origin and evolution of alu subfamilies.* PloS one, 2013. 8(6): p. e64884.

48. Watkins, W., et al., *Patterns of ancestral human diversity: an analysis of Alu-insertion and restriction-site polymorphisms.* The American Journal of Human Genetics, 2001. 68(3): p. 738-752.

49. Kass, D.H., et al., *Identification of a unique Alu-based polymorphism and its use in human population studies.* Gene, 2007. 390(1): p. 146-152.

50. Ostertag, E.M., et al., *SVA elements are nonautonomous retrotransposons that cause disease in humans.* The American Journal of Human Genetics, 2003. 73(6): p. 1444-1451.

51. Kulski, J.K., A. Shigenari, and H. Inoko, *Polymorphic SVA retrotransposons at four loci and their association with classical HLA class I alleles in Japanese, Caucasians and African Americans.* Immunogenetics, 2010. 62(4): p. 211-230.

52. Wang, H., et al., *SVA elements: a hominid-specific retroposon family.* Journal of molecular biology, 2005. 354(4): p. 994-1007.

53. Ray, D.A. and M.A. Batzer, *Reading TE leaves: new approaches to the identification of transposable element insertions.* Genome research, 2011. 21(6): p. 813-820.

54. Sinha, S., et al., *Retrotransposable elements: Novel and sensitive DNA markers and their application in human identity.* Forensic Science International: Genetics Supplement Series, 2015. 5: p. e627-e629.

55. LaRue, B.L., et al., *INNULs: A Novel Design Amplification Strategy for Retrotransposable Elements for Studying Population Variation.* Human heredity, 2012. 74(1): p. 27-35.

56. Lee, S.B., B. McCord, and E. Buel, *Advances in forensic DNA quantification: a review.* Electrophoresis, 2014. 35(21-22): p. 3044-3052.

57. Pineda, G.M., et al., *Development and validation of InnoQuant™, a sensitive human DNA quantitation and degradation assessment method for forensic samples using high copy number mobile elements Alu and SVA.* Forensic Science International: Genetics, 2014. 13: p. 224-235.

58. Goecker, Z.C., et al., *Comparison of Quantifiler® Trio and InnoQuant™ human DNA quantification kits for detection of DNA degradation in developed and aged fingerprints.* Forensic science international, 2016. 263: p. 132-138.

59. Pineda, G.M., et al. *Development of a highly sensitive human and male quantification system for assessing DNA quality in forensic samples*. in *Poster A67 from 66th Amer. Acad. Forensic Sci. Conf.* 2014.

60. Aaron LeFebvre, S.A., Laura Gahn, Gina Pineda Murphy, Sid Sinha, and Sudhir K. Sinha, *InnoQuant® as a tool to determine profile suitability and improve profile success rates for high throughput property crime specimens.* Poster, 2014.

61. Sinha, S., *Method for genetic detection using interspersed genetic elements: a multiplexed dna analysis system*. 2013, Google Patents.

62. Murphy, G.P., et al., *Novel Methods to Obtain Nuclear DNA Profiles from Rootless Hair Shafts.*

63. Butler, J.M., C.S. Tomsey, and M.C. Kline. *Can the validation process in forensic DNA typing be standardized*. in *Proceedings of the 15th International Symposium on Human Identification*. 2004. Citeseer.

64. (SWGDAM), S.W.G.o.D.A.M., *Revised Validation Guidelines*. 2004: Forensic Science Communications. p. 1-6.

65. Methods, S.W.G.o.D.A., *Validation Guidelines for DNA Analysis Methods*. 2012. p. 1-13.

66. Corporation, P., *Internal Validation Guide of Autossomal STR Systems for Forensic Laboratories*, P. Corporation, Editor. 2013. p. 1-4.

67. Group, E.D.W., *Recommended minimum criteria for the validation of various aspects of the DNA profiling process*. 2010.

Appendix

Appendix I – InnoTyper® 21 marker Information: loci amplified and the corresponding fluorescent marker dyes, observed allele sizes (in base pairs) and the genotype of the DNA Positive Control

	Marker	Dye	Chromo-some	Band	Amplicon Size* (I = Insertion)	Amplicon Size* (N = No Insertion)	Positive Control Profile
1	AC4027	FAM	7	7q21.11	66.1	68.4	I,I
2	MLS26	FAM	3	3p22.1	79.8	82.3	I,N
3	ALU79712	FAM	20	20p12.2	91.6	95.4	N,N
4	NBC216	FAM	7	7p14.1	99.8	109.2	I,I
5	NBC106	FAM	21	21q22.2	115.9	120.0	I,N
6	RG148	JOE	2	2q23.3	73.3	80.5	I,N
7	NBC13	JOE	16	16p12.1	85.4	89.4	N,N
8	AC2265	JOE	13	13q33.1	96.4	100.5	I,N
9	MLS09	JOE	1	1q25.3	111.6	116.7	I,N
10	AC1141	TMR	3	3q11.2	63.4	66.2	I,I
11	TARBP	TMR	1	1q42.2	70.3	73.7	I,N
12	AMEL	TMR	X, Y	Xp22.1-22.3 Yp11.2	X=78.1	Y=80.9	X,Y
13	AC2305	TMR	13	13q13.3	92.0	97.2	I,N
14	HS4.69	TMR	5	5q34	109.2	113.9	I,N
15	NBC51	TMR	3	3q28	123.5	119.7	N,N
16	ACA1766	ROX	8	8q12.1	69.5	75.5	I,N
17	NBC120	ROX	22	22q11.21	79.2	83.6	I,N
18	NBC10	ROX	4	4q31.21	86.9	93.1	N,N
19	NBC102	ROX	17	17q23.3	101.0	97.1	N,N
20	SB19.12	ROX	19	19q13.43	108.7	113.6	I,N
21	NBC148	ROX	14	14q31.1	117.5	120.2	I,I

Appendix II – Equipment from LPC-PJ used in this study

Figure 1 - AutoMate™ Express Instrument and laminar flow camera with pipes support on inside

Figure 2 – ABI 7500 HID Real-Time PCR analysis

Figure 3 – Thermal cycler GeneAmp® PCR System 9700 with gold-plated silver 96-Well Block equipment

Figure 4 – ABI 3130XL Genetic Analyzer equipment

Figure 5 – QIAgility™ robot present in pre-PCR room used to prepare the plates for DNA quantification

Figure 6 – QIAgility™ robot present in post-PCR room used to prepare the plates for the capillary electrophoresis

Appendix III – Retrotransposons-based kits used in the present study

Figure 1 - InnoQuant® HY kit

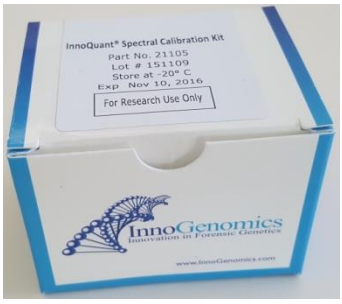


Figure 2 - InnoQuant® HY Spectral Calibration kit

Figure 3 - InnoTyper® 21 kit

Figure 4 – IGT 5 Dye Matrix Standard kit

Figure 5 – ILS- Internal Line Standard and Allelic ladder

Appendix IV - Selected samples and their division according to the genetic profile presented

Group	Sample	Ext	Trace	Quantification value (QD)	Type of profile
I	1	430-11	Hair	----------	No Results
	2	442-02	Hair	----------	
	3	442-07	Hair	----------	
	4	430-02	Hair	----------	
	5	447-02	Hair	----------	
	6	447-12	Hair	----------	
	7	447-04	Hair	----------	
	8	450-01	Hair	----------	
	9	450-04	Hair	----------	
	10	450-05	Hair	----------	
	11	450-06	Hair	----------	
	12	450-07	Hair	----------	
	13	450-11	Hair	----------	
	14	450-12	Hair	----------	
	15	442-06	Hair	0.001	
	16	430-07	Hair	----------	
	17	442-10	Hair	0.001	
	18	442-05	Hair	0.002	
	19	103-10	Hair	----------	
	20	154-12	Hair	----------	
	21	257-05	Hair	----------	
	22	310-13	Hair	----------	
	23	386-06	Hair	----------	
	24	033-10	Hair	----------	
	25	245-06	Hair	----------	
	26	266-02	Hair	0.000	
	27	266-07	Hair	0.000	
	28	266-04	Hair	0.000	
	29	033-04	Hair	0.001	
	30	266-01	Hair	0.001	
	31	295-11	Hair	0.001	
	32	317-07	Hair	0.001	
	33	033-05	Hair	0.002	
	34	033-12	Hair	0.002	
	35	245-05	Hair	0.002	
	36	033-06	Hair	0.003	
	37	245-04	Hair	0.003	
	38	266-06	Hair	0.003	
	39	081-09	Hair	0.006	

Group		Sample	Ext	Trace	Quantification value (QD)	Type of profile
II		40	442-01	Hair	-------	Inconclusive
		41	450-02	Hair	-------	
		42	358-03	Hair	0.003	
		43	442-04	Hair	--------	
		44	430-12	Hair	0.002	
		45	447-03	Hair	0.003	
		46	447-09	Hair	0.023	
		47	317-01	Hair	-------	
		48	317-04	Hair	-------	
		49	015-11	Hair	-------	
		50	046-06	Hair	--------	
		51	155-08	Hair	0.003	
		52	358-04	Hair	---------	
		53	348-08	Hair	---------	
		54	348-07	Hair	0.003	
		55	310-06	Hair	0.009	
III		56	430-04	Hair	0.001	Complete
		57	447-08	Hair	0.024	
		58	442-09	Hair	0.087	
		59	430-03	Hair	0.435	
		60	295-09	Hair	0.021	
		61	295-10	Hair	0.531	
		62	151-01	Hair	1.316	
		63	033-09	Hair	1.587	
IV	IPC +	64	452-12	Biologic	---------	No Results
		65	454-06	Biologic	---------	No Results
		66	454-06	Biologic	---------	No Results
		67	452-11	Biologic	---------	Mixture (only quantification)
		68	452-09	Biologic	0.012	Mixture (only quantification)
		69	109-05	Hair	0.002	No Results
		70	109-09	Hair	1.201	Complete
		71	109-10	Hair	0.610	Complete
		72	103-13	Hair	---------	No Results
		73	263-03	Biologic	0.002	Inconclusive
		74	134-12	Biologic	0.049	Inconclusive
		75	129-01	Semen	0.016	Inconclusive
		76	097-10	Semen	0.003	Inconclusive
		77	096-10	Semen	0.004	No Results
		78	097-11	Semen	0.015	Complete
		79	297-04	Biologic	---------	No Results
		80	297-05	Biologic	0.001	No Results

Group		Sample	Ext	Trace	Quantification value (QD)	Type of profile
IV	IPC +	81	297-03	Biologic	--------	No Results
		82	134-05	Biologic	0.088	Complete
		83	134-06	Biologic	0.007	Inconclusive
		84	297-10	Biologic	0.005	No Results
		85	134-04	Biologic	0.009	Inconclusive
		86	134-11	Biologic	0.012	Inconclusive
V	Degraded DNA	87	465-02	AR	0.006	Inconclusive
		88	457-17	AR	0.006	
		89	457-03	AR	0.010	
		90	447-05	Hair	0.011	
		91	375-24	AR	0.012	
		92	457-02	AR	0.016	
		93	457-11	AR	0.021	
		94	457-01	AR	0.041	
		95	457-16	AR	0.053	
		96	457-07	AR	0.060	
		97	401-15	AR	0.078	
		98	457-21	AR	0.099	
		99	447-10	Hair	0.111	
		100	465-01	AR	0.265	
		101	455-04	Biologic	0.052	
		102	423-01	Biologic	0.002	
		103	462-11	Biologic	0.004	
		104	033-08	Hair	0.012	
		105	151-04	Hair	0.001	
		106	310-07	Hair	0.054	
		107	435-11	Hair	0.011	
		108	368-01	Hair	0.000	
		109	369-02	Biologic	0.004	
		110	369-04	Biologic	0.004	
		111	380-04	RBC	---------	
		112	381-03	Biologic	0.003	
		113	381-07	Biologic	0.007	
		114	418-02	Biologic	0.007	
		115	423-09	Biologic	0.024	
		116	370-05	---------	0.012	
		117	151-07	Hair	0.006	

Group		Sample	Ext	Trace	Quantification value (QD)	Type of profile
VI	Degraded or inhibited DNA	118	101-01	Teeth	-------	No Results
		119	113-02	Teeth	--------	No Results
		120	113-01	Teeth	0.001	No Results
		121	113-03	Teeth	0.26	Complete
		122	452-04	RBC	1.092	Inconclusive
		123	126-04	Teeth	0.008	Inconclusive
		124	122-03	Teeth	--------	No Results
		125	054-05	Teeth	1.275	Inconclusive
		126	106-02	Teeth	-------	Inconclusive
		127	122-06	Teeth	--------	No Results
		128	432-02	Teeth	0.001	Inconclusive
		129	432-03	Teeth	--------	Inconclusive

Appendix V - Quantification results, respective averages and standard deviation for the targets Short, Long and Y analyzed by InnoQuant® HY kit in the sensitivity test

Target	Expected Values (ng/µL)		Observed values (ng/µL) - QIAgility™ robot				Observed values (ng/µL) - Manual
			1st Rep	2nd Rep	Average	Desviation	
Short	Intermediate Dilution	10	13.07	11.29	12.18	1.26	-
	Dilution 1	1	1.10	1.13	1.12	0.03	0.9
	Dilution 2	0.5	0.68	0.60	0.64	0.05	0.46
	Dilution 3	0.250	0.284	0.299	0.292	0.01	0.215
	Dilution 4	0.125	0.143	0.141	0.142	0.001	0.124
	Dilution 5	0.063	0.064	0.071	0.068	0.005	0.056
	Dilution 6	0.031	0.034	0.038	0.036	0.003	0.033
	Significance						NS
	Pvalue						0.064
Long	Intermediate Dilution	10	10.25	5.67	7.96	3.24	-
	Dilution 1	1	1.07	1,03	1.05	0.03	0.83
	Dilution 2	0.5	0.64	0.57	0.60	0.05	0.47
	Dilution 3	0.250	0.270	0.298	0.284	0.02	0.183
	Dilution 4	0.125	0.159	0.156	0.158	0.002	0.109
	Dilution 5	0.063	0.069	0.088	0.079	0.013	0.070
	Dilution 6	0.031	0.032	0.034	0.033	0.001	0.026
	Significance						NS
	Pvalue						0.054
Y	Intermediate Dilution	10	9.19	11.85	10.52	1.88	-
	Dilution 1	1	1.08	1.19	1.13	0.08	0.88
	Dilution 2	0.5	0.69	0.60	0.64	0.06	0.294
	Dilution 3	0.250	0.299	0.297	0.298	0.001	0.161
	Dilution 4	0.125	0.138	0.152	0.145	0.01	0.103
	Dilution 5	0.063	0.067	0.065	0.066	0.001	0.071
	Dilution 6	0.031	0.032	0.031	0.0315	0.0007	0.018
	Significance						NS
	Pvalue						0.064

Annex VI - C_T values, respective averages and standard deviation for all the four targets analyzed by InnoQuant® HY kit in the four runs performed to repeatability and reproducibility tests

Target	Standard	Concentration (ng/µL)	Run 1				Run 2				Run 3				Run 4				Global Average	Standard error
			CT1	CT2	Average	Desviation	CT3	CT4	Average	Desviation	CT5	CT6	Average	Desviation	CT7	CT8	Average	Desviation		
Short	Std 1	20	11.23	11.27	11.25	0.03	11.27	10.99	11.13	0.20	11.08	11.24	11.16	0.11	13.81	12.1	12.96	1.21	11.62	0.47
	Std 2	2.5	14.10	14.12	14.11	0.01	14.20	14.03	14.12	0.12	13.96	14.1	14.03	0.10	15.93	16.23	16.08	0.21	14.58	0.46
	Std 3	0.3125	17.03	16.96	17.00	0.05	17.04	17.07	17.06	0.02	16.82	16.92	16.87	0.07	18.91	18.85	18.88	0.04	17.45	0.44
	Std 4	0.0391	20.03	19.96	20.00	0.05	20.11	19.94	20.03	0.12	19.98	19.98	19.98	0.00	22.11	22.24	22.18	0.09	20.54	0.50
	Std 5	0.005	22.79	22.88	22.84	0.06	22.84	24.74	23.79	1.34	22.68	22.77	22.73	0.06	25.7	26.3	26.00	0.42	23.84	0.75
	Significance					NS				NS				NS				NS		NS
	Pvalue					0.074				0.150				0 .363				0 .146		0.057
Long	Std 1	20	14.86	15.04	14.95	0.13	15.31	14.95	15.13	0.25	14.95	15.07	15.01	0.09	11.65	11.16	11.40	0.34	14.12	0.84
	Std 2	2.5	17.71	17.61	17.66	0.07	17.64	17.48	17.56	0.11	17.57	17.70	17.64	0.10	14.13	15.02	14.57	0.63	16.86	0.72
	Std 3	0.3125	20.67	20.79	20.73	0.09	20.74	20.74	20.74	0.00	20.55	20.51	20.53	0.03	17.39	17.61	17.50	0.15	19.88	0.74
	Std 4	0.0391	23.70	23.49	23.60	0.15	23.74	23.60	23.67	0.10	23.56	23.45	23.50	0.08	20.65	20.93	20.79	0.19	22.89	0.65
	Std 5	0.005	26.56	26.36	26.46	0.14	26.43	28.64	27.53	1.57	26.43	26.34	26.39	0.06	23.99	24.03	24.01	0.03	26.10	0.75
	Significance					NS				NS				NS				NS		NS
	Pvalue					0.153				0.317				0.173				0.219		0.178
Y	Std 1	20	20.07	20.10	20.08	0.02	20.03	20.10	20.06	0.05	19.83	20.02	19.92	0.13	18.69	18.83	18.76	0.10	19.71	0.30
	Std 2	2.5	22.92	22.86	22.89	0.05	22.88	22.78	22.83	0.08	22.79	22.70	22.74	0.06	21.50	20.37	20.93	0.80	22.35	0.46
	Std 3	0.3125	25.70	25.60	25.65	0.07	25.50	25.62	25.56	0.08	25.48	25.53	25.51	0.04	23.89	25.18	24.54	0.91	25.31	0.30
	Std 4	0.0391	28.51	28.82	28.66	0.23	28.59	28.42	28.51	0.12	28.47	28.58	28.53	0.07	27.62	28.09	27.85	0.33	28.39	0.19
	Std 5	0.005	31.68	31.65	31.66	0.02	31.51	32.66	32.08	0.81	31.17	31.67	31.42	0.35	30.64	31.38	31.01	0.53	31.54	0.29
	Significance					NS				NS				NS				NS		NS
	Pvalue					0.452				0.470				0.415				0.084		0.066
IPC	Std 1	20	21.11	20.92	21.02	0.14	21.01	20.98	21.00	0.02	21.07	20.99	21.03	0.06	17.47	17.41	14.95	0.13	19.50	0.83
	Std 2	2.5	20.19	19.75	19.97	0.31	20.15	19.69	19.92	0.32	20.16	19.79	19.97	0.26	17.86	17.75	17.66	0,07	19.38	0.51
	Std 3	0.3125	19.98	19.93	19.95	0.03	19.87	19.89	19.88	0.01	19.86	19.82	19.84	0.03	18.43	18.42	20.73	0,09	20.10	0.34
	Std 4	0.0391	20.53	20.55	20.54	0.02	20.53	20.57	20.55	0.03	20.44	20.50	20.47	0.04	19.18	19.72	23.60	0,15	21.29	0.26
	Std 5	0.005	21.06	21.12	21.09	0.04	20.88	21.07	20.97	0.13	21.04	20.92	20.98	0.08	19.54	19.71	26.46	0,14	22.38	0.33
	Significance					NS				NS				*				NS		NS
	Pvalue					0.160				0.304				0.042				0.315		0.997

Note: Global Average represents the average from the four runs. Standard error corresponds to the standard deviation amongst the four runs. For the comparison within runs, a Chi-square test was performed.

Annex VII- Number of amplified alleles, respective averages and standard deviation obtained by InnoTyper® 21 kit in the sensitivity test

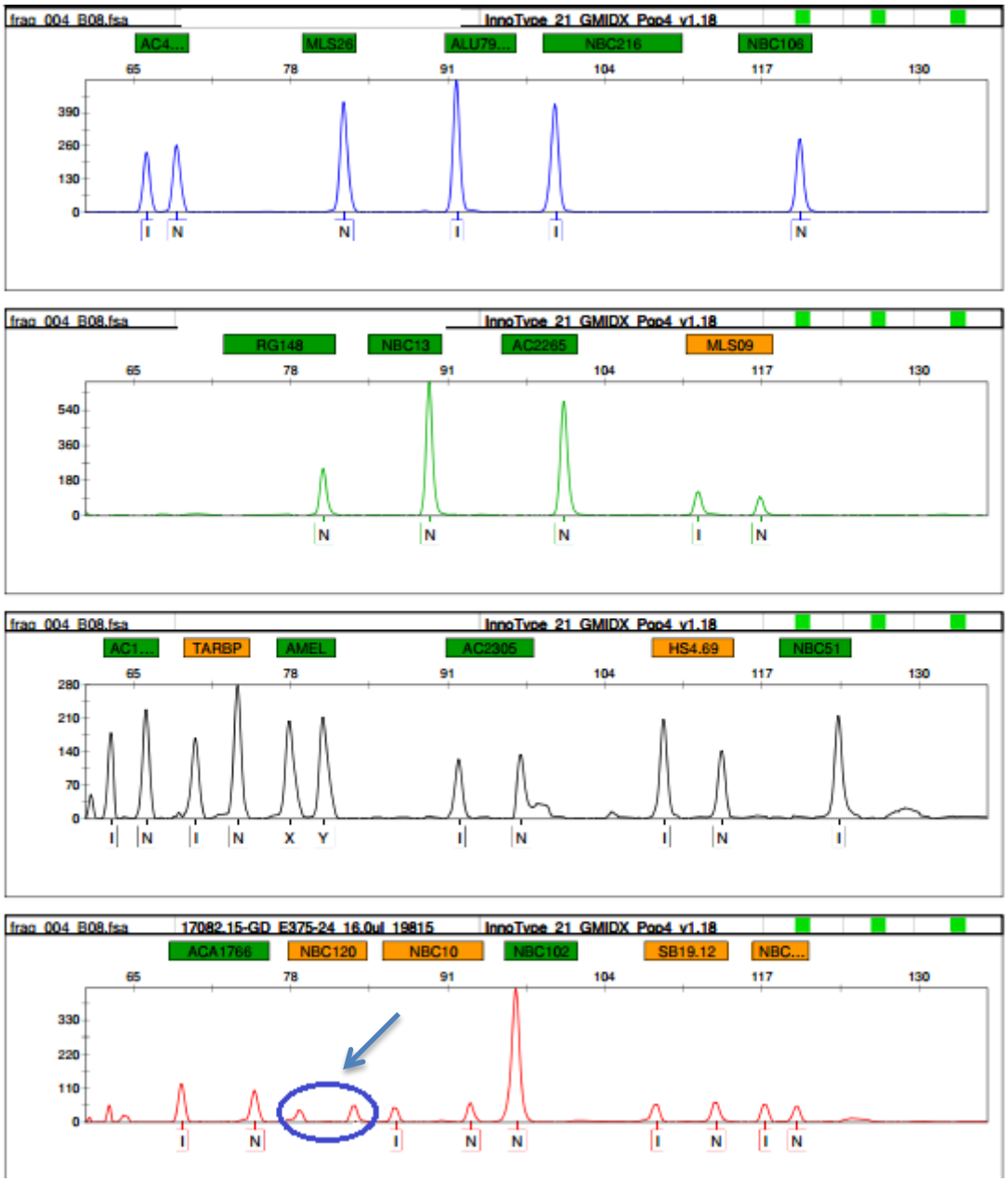
Samples	Concentration (ng/µL)	Number of alleles obtained				Percentage (%)			
		1st Rep	2nd Rep	Average	Stdev	1st Rep	2nd Rep	Average	Stdev
PC1	0.12	34	28	31	4.24	80.95	66.67	73.81	10.10
PC2	0.24	40	40	40	0	95.24	95.24	95.24	0
PC3	0.36	40	41	40.5	0.71	95.24	97.62	96.43	1.68
PC4	0.48	42	41	41.5	0.71	100	97.62	98.81	1.68
PC5	0.6	42	42	42	0	100	100	100	0
Significance				NS				NS	
Pvalue				0.459				0.474	

Annex VIII - Number of amplified alleles, respective averages and standard deviation obtained for each positive control tested by InnoTyper® 21 kit with different injection times to test repeatability and reproducibility

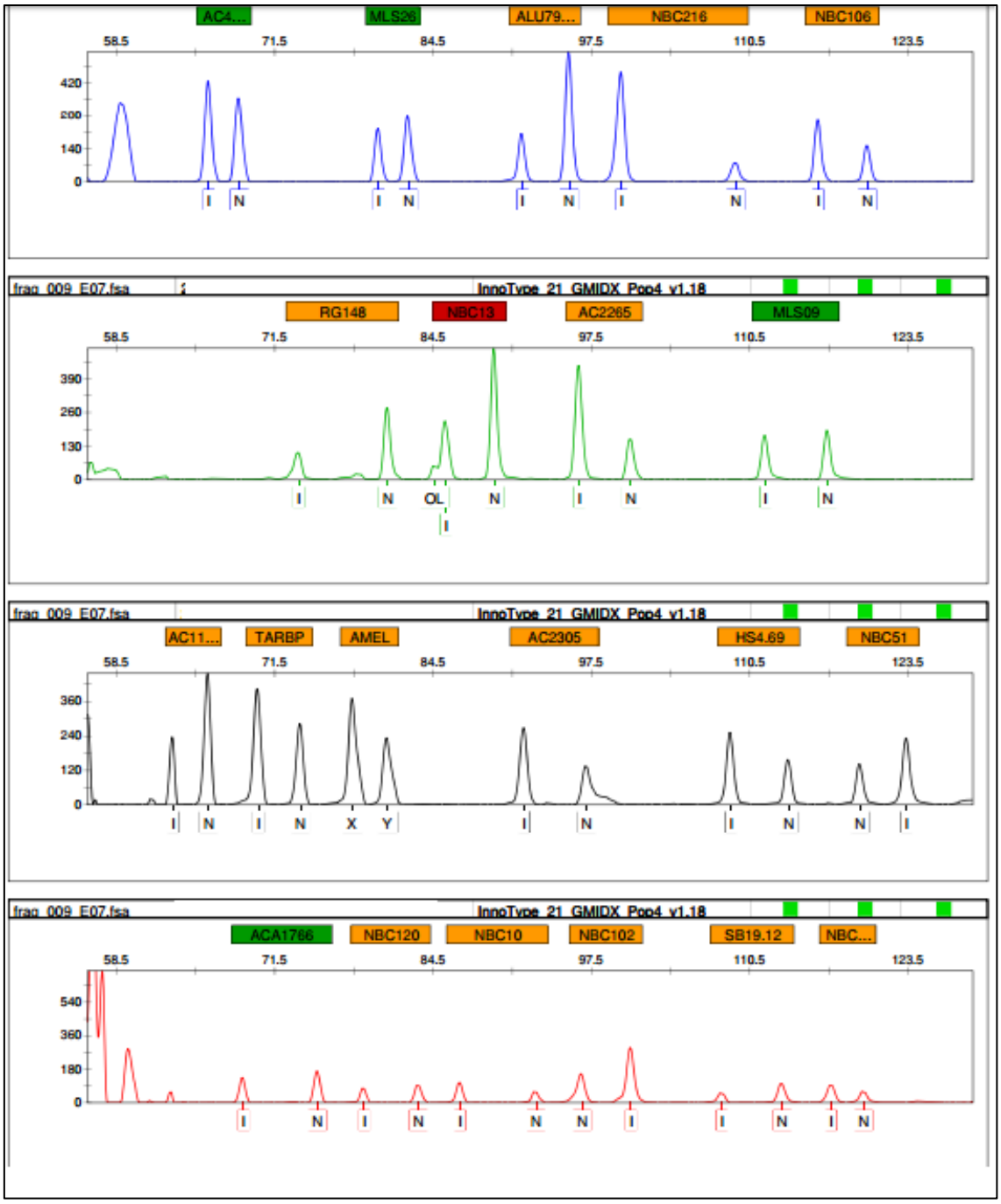
Sample				PC1	PC2	PC3	PC4	PC5		
Concentration (ng/µL)				0.12	0.24	0.36	0.48	0.6	Significance	Pvalue
Time Injection	16 seconds	Number of Alleles	1st rep	28	39	39	42	42	NS	0.443
			2nd rep	25	40	40	40	42		
			Average	26,5	39,5	39,5	41	42		
			Stdev	2.12	0.71	0.71	1.41	0		
		%	1st rep	67	93	93	100	100	NS	0.436
			2nd rep	60	95	95	95	100		
			Average	63,5	94	94	97,5	100		
			Stdev	2.45	0.71	0.71	1.77	0		
	18 seconds	Number of Alleles	1st rep	28	39	40	42	42	NS	0.409
			2nd rep	24	40	40	40	42		
			Average	26	39,5	40	41	42		
			Stdev	1.41	0.35	0	0.71	0		
		%	1st rep	67	93	95	100	100	NS	0.401
			2nd rep	57	95	95	95	100		
			Average	62	94	95	97.5	100		
			Stdev	3.54	0.71	0	1.77	0		
	20 seconds	Number of Alleles	1st rep	34	40	40	42	42	NS	0.35
			2nd rep	28	40	41	41	42		
			Average	31	40	40,5	41,5	42		
			Stdev	2.12	0	0.35	0.35	0		
		%	1st rep	81	95	95	100	100	NS	0.362
			2nd rep	67	95	98	98	100		
			Average	74	95	96,5	99	100		
			Stdev	4.95	0	1.06	0.71	0		

Note: For the comparison within runs, a Chi-square test was performed.

Annex IX – Example of a genetic profile with quality to be reported



Annex X – Example of a complete genetic profile



Annex XI – Example of a genetic profile exhibiting amplification peaks which could not be reported since due to the reduced RFU peak size