Progress in Forensic Genetics: New Markers Validation Studies and Population Data

Anna Barbaro

PhD Academic Thesis

2012

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Progress in Forensic Genetics:
New Markers Validation Studies
and Population Data

Dedicated to

My beloved Aysha

Anna Barbaro

Academic thesis submitted for the

"Programa de Doutoramento in Ciencias Forenses e Patoloxía"

Academic Year 2011-2012

UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

FACULTADE DE MEDICINA

DEPARTAMENTO DE ANATOMÍA PATOLÓXICA

La Doctora María Victoria Lareu Huidobro y el Doctor Ángel Carracedo Álvarez, Catedráticos de Medicina Legal de la Universidad de Santiago de Compostela

CERTIFICAN

Que la presente memoria que lleva por título *Progress in Forensic Genetics: New Markers Validation Studies and Population Data* realizada pro la licenciada Anna Barbaro, ha sido realziada bajo nuestra dirección, considerándola en condiciones para optar al Grado de Doctor y autorizándola para su presentación y defensa ante el Tribunal correspondiente.

Y para que así conste firmamos la presente certificación en Santiago de Compostela a 20 de julio de 2012

Fdo. Dra.María Victoria Lareu Huidobro **Fdo. Dr. Ángel Carracedo Álvarez**

Fdo. Dª Anna Barbaro

INDEX

Chapter III: DNA Markers:Single Nucleotide Polymorphisms (SNPS)

Chapter IV : DNA Procedure Standardization

Aims and outline of the thesis

More than 20 years passed from the first application of DNA fingerprints in forensics and DNA analysis has played a crucial role in the investigation and resolution of thousands of violent crimes.

In the last years DNA analysis is rapidly developed in particular after the introduction of forensic DNA databases useful in the fight against crime. DNA has become a powerful forensic tool for solving cases such as linking a suspect to a crime scene, resolving biological relationship issues and identifying disaster victims.

Three different types of DNA marker, Short Tandem Repeats (STRs), Single Nucleotide Polymorphisms (SNPs) and DNA sequence data, represent the absolute majority of polymorphisms used in forensic genetic applications.

They all have characteristics, making them especially useful for solving criminal cases and for relationship testing.

Short tandem repeats (STRs) are the most widely used markers for forensic DNA testing, because of their high differentiating power, good resolution of alleles and the ability to process samples rapidly using multiplexed polymerase chain reaction (PCR).

13 STRs have been chosen as the core loci upon which the FBI's Combined DNA Index System (CODIS) database has been built.

But other genetic polymorphisms, such as those found in the mitochondrial DNA (mtDNA) genome and the X or Y chromosome, have been shown to provide effective results that can improve traditional STR data.

The demand for tools and technologies in forensic DNA testing, is continuous: common problems in forensics are concerning the genetic identification of degraded biological samples such as the ones collected from crime scenes or mass disaster that may have been exposed to harsh environmental conditions (sunlight, humidity, etc.) that damage DNA structure, or the presence of inhibitors interfering with the ability to obtain a full DNA profile from a biological evidence.

To overcome these problems, new markers has been selected in the last years, in order to recover as more information as possible from smaller regions of DNA, which are more likely to be intact following DNA damage.

These include mini-STRs and single nucleotide polymorphisms (SNPs).

In this perspective, not only the range of genetic markers used is widely increased but also new sophisticated analytical methods (automation, miniaturization, high-throughput performance) have been adopted in order to give to investigators as more informations as possible about a perpetrator solely on the biological evidence left at the crime scene.

Moreover, more recently, the analysis of genes useful for physical characteristics determination (such as hair, eye or skin colour) have been introduced and this application may have in the near future a fundamental role in forensics. The ability to perform genetic typing of biological traces collected at the crime scene, in order to obtain information about a donor's physical characteristics, is a very attractive prospect for forensic analysis and it could potentially offer a powerful new tool for crime scene investigations.

Obviously before the introduction in routine casework analysis, it's relevant for the forensic community to establish which markers may be useful for catching up the procedure to a level acceptable for forensic application and than to validate protocols with sufficient analysis repeat rates. Moreover in order to calculate the correct representative weight of DNA evidence, prior knowledge about the DNA markers for a relevant population sample is required. Important properties such as how frequently certain DNA-variants (i.e. alleles) occur in the population, the differences in such frequencies between populations and the forensic efficiency of the DNA markers in casework should be studied to determine the probability that a particular genotype might occur at random in a population.

<u>The aims of this thesis are:</u>

- to validate a next generation pentaplex, previously we developed, including the new five loci recommended by the European Union Council for the expansion of the

European Standard Set (ESS) evaluating the STR data informativeness and success rate on a wide range of forensic samples and to compare its performance with the one of other commercially available kits .

- to create a useful population database that includes an estimate of the frequency of each possible allele and genotype (or haplotype), by studying the variability in Mediterranean Area of the well established 15 autosomal STRs together with the 5 new ESS, the highly discriminative SE33 and some sex linked STRs commonly used in forensics.Moreover to evaluate the variability of the 52 SNPplex recently introduced for forensic applications.

Justificación y objetivos de la tesis

Han pasado más de 20 años desde el primer uso de las huellas de ADN (DNA fingerprints) en un caso forense y el análisis de ADN ha jugado un papel crucial en la investigación y la resolución de crímenes violentos.

El análisis de ADN se ha desarrollado rápidamente en los ultimos años en particular después de la introducción de bases de datos de ADN forenses, de gran utilidad en la lucha contra el crimen.

Las repeticiones cortas en tándem (STRs) son los marcadores más habitualmente usados para pruebas de ADN forenses, debido a su alto poder de diferenciación, la buena resolución de los alelos y la capacidad de analizarlos rápidamente mediante electroforesis capilar después de amplificarlos mediante la reacción en cadena de polimerasa (PCR).

Trece han sido los STRs escogidos como los marcadores principales sobre los cuales se ha construído la base de datos del CODIS y otros más se han añadido en Europa como parte del ESS (European Standard Set).

Pero otros polimorfismos genéticos, como aquellos encontrados en el ADN mitochondrial (mtDNA) y en los cromosomas X o Y, han mostrado también proporcionar resultados eficaces que pueden complementar los datos proporcionados por los STRs tradicionales.

La introducción de nuevos instrumentos y tecnologías en pruebas de ADN forense es continua: un problema común es la identificación genética de muestras biológicas degradadas que aparecen a menudo en la escena del crimen o en desastre de masas, donde las muestras pueden haber sido expuestas a condiciones ambientales agresivas (la luz del sol, la humedad, etc.) que dañan la estructura de ADN, o implican la presencia de inhibidores que interfieren con la capacidad de obtener un perfil de ADN adecuado.

Para solucionar estos problemas, han sido seleccionados nuevos marcadores en los años pasados, con la idea común de utilizar las más pequeñas regiones de ADN que con mayor probabilidad están preservadas de daños. Estos nuevos marcadores incluyen miniSTRs y SNPs.

Desde esta perspectiva, no sólo la gama de marcadores genéticos usados extensamente se ha visto aumentada, sino también nuevos métodos analíticos sofisticados han sido adoptados para dar a los investigadores más información sobre el autor de las pruebas biológicas dejadas en la escena de crimen.

Además, más recientemente, el análisis de genes útiles para la determinación de características física (como el pelo, el ojo o el color de la piel) ha sido investigada y este uso puede tener en un futuro próximo un papel fundamental en la la investigación criminal. La capacidad de analizar muestras biológicas para obtener la información sobre las características físicas de un donante, es una perspectiva muy atractiva para el análisis forense y potencialmente podría ofrecer un nuevo instrumento poderoso para investigaciones a partir de vestigios obtenidos en la escena de crimen.

Obviamente antes de la introducción en el análisis rutinario, es relevante para la comunidad forense establecer cuales marcadores pueden ser útiles para mantener el procedimiento a un nivel aceptable para el uso forense y validar protocolos con un numero de repeticiónes de análisis suficientes.

Además antes de la introducción de nuevos marcadores de ADN en forense, para determinar la probabilidad que un genotipo particular ocurre en una población, es necesario realizar estudios demográficos para hacer una estimación de la frecuencia de cada allelo posible y genotipo (o haplotipo).

Los objetivos de esta tesis son:

- validar una pentaplex de nueva generación que incluya los cinco nuevos marcadores recomendados por el Consejo de la Unión Europea para la extensión del lo European Standard Set (ESS) mediante evaluación del nivel de información de los datos y del éxito del análisis sobre una amplia gama de muestras forenses y también comparar su funcionamiento con el que de otros equipos disponibles comercialmente.

- crear una base de datos demográfica útil en forense que incluye una estimación de la frecuencia de cada allelo posible y genotipo (o haplotipo), estudiando la variabilidad

en el Área mediterránea de 15 STRs autosómicosl bien establecidos juntos con los 5 nuevos marcadores EES, el sumamente discriminante SE33 y algún STRs de los chromosomes sexuales comúnemente usados en la investigación forense.

Finalmente pretendemos evaluar la variabilidad de la 52 SNPplex recientemente introducida para usos forenses.

Chapter I

GENERAL INTRODUCTION ABOUT FORENSIC DNA TYPING

1. A brief History

Ever since the early 1900s when Karl Landsteiner discovered the A,B, and O blood types, it has been the dream of forensic serologists (scientists who study body fluids) to be able to positively identify the individual source of a small blood stains. Advances were made during the 1930s in discovering new blood types such as the Rh factor and again in the 1970s with the invention of electrophoresis, a new method for separating and identifying some of the variable proteins found in blood that are the keys to classic genetic typing. Proteins comprise the majority of the structural and functional substances that make up our bodies. Application of more and more of these typing systems allowed an increasing percentage of the wrongly suspected individuals to be excluded as the source of an evidence stain. A major problem with the classical systems that test proteins found in blood, is that very few of these proteins are also found in semen and other body fluids. Because the majority of cases that require genetic typing are sexual assaults, the lack of a definitive set of useful genetic markers in semen has long been a great handicap to the scientific analysis of rape evidence

In 1953 James Watson and Francis Crick working at Cambridge University first described the double helix structure of DNA and at that time they could probably not imagine the future usefulness of their finding.DNA is found in almost every cell in the body. There are about 100 trillion cells in the adult human body. Most of them have a nucleus that contains thread-like bundles of chromosomes. Each parent contributes one chromosome to each of the 23 pairs found in all normal people. Within the chromosomes, are up to 100,000 paired genes, the fundamental units of heredity. Each gene can have different versions (as many as 100 or more in rare cases) called alleles, but most are the same from person to person. Genes determine all inherited traits including those that give the individual specific characteristics (blue eyes rather than brown eyes) as well as common characteristics (two eyes, two arms, etc.).

Since there a wide range of technologies have been developed: by analyzing DNA, information about diseases, evolution of biological life and population history can be retrieved. Nowadays, DNA is used in everyday practice for applications within different areas, such as medical genetics, the food processing industry and in forensic situations when solving crimes as well as in disputes about biological relationships.[1]

1.1 RFLPs (Restriction Fragments Polymorphisms) analysis

DNA fingerprinting was first described in 1985 by Dr. Alec Jeffreys, an English scientist at Leicester University, who found that certain region of human genome contained DNA sequences that were repeated over and over again next to each other. He also discovered that the number of repeated sections present in a sample could differ from individual to individual.[2] By developing a technique to examine length variation of these DNA repeat sequences, Dr. Jeffreys created the ability to perform human identity tests. These DNA repeat regions became known as VNTRs, which stands for variable number of tandem repeats.

The technique used by Dr. Jeffreys to examine the VNTRs was called restriction fragment length polymorphism (RFLP) because it involved the use of a restriction enzyme to cut the regions of DNA surrounding the VNTRs.[3]

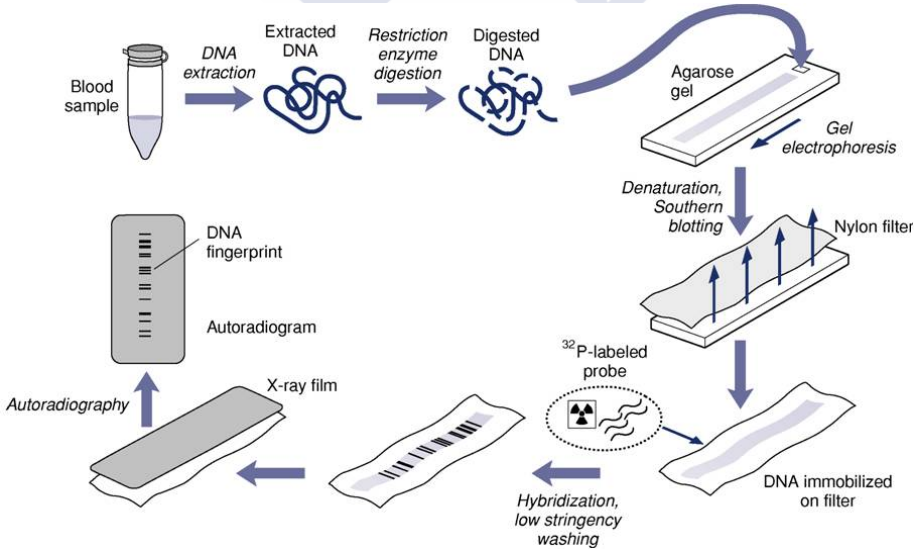


Fig.1 *RFLPs analysis by Southern Blotting*

DNA fragments obtained by restriction enzymes were separated in size by electropho resis in agar gel and than transferred to a filter membrane for subsequent detection by radioactively labelled probes using a procedure called Southern blotting. [4]
Jeffreys proved that, even if VNTRs loci are very similar between closely related humans, however the small cut fragments of DNA molecules were so variable that unrelated individuals are extremely unlikely to have the same VNTRs, so they were virtually unique to individuals.

With appropriate dramatic flair, he called the process he invented "DNA fingerprinting," a term most forensic scientists dislike because it is confusing and can be misleading. With his co-workers, he also demonstrated that forensic samples, dried stains several years old, contained sufficient DNA to yield conclusive results.
Like the fingerprints that came into use by detectives and police labs during the 1930s, each person has a unique DNA fingerprint. Unlike a conventional fingerprint that occurs only on the fingertips and can be altered by surgery, a DNA fingerprint is the same for every cell, tissue and organs of a person and it cannot be altered by any known treatment. Consequently, DNA fingerprinting rapidly became the primary method for identifying and distinguishing among individual human beings.

DNA fingerprinting was first used as a police forensic test to identify the rapist and killer of two teenagers, Lynda Mann and Dawn Ashworth, who were both murdered in Narborough Leicestershire, in 1983 and 1986 respectively. A young man Colin Pitchfork, was identified and convicted of murder after samples taken from him matched semen samples taken from the two dead girls. [5]
This turned out to be a specifically important identification for without it, British Authorities believe that Richard Buckland, the main suspect, would have inevitably been convicted. Therefore, not only did Jeffrey's work in this case prove who the real killer was, but exonerate someone who likely would have spent his life in prison otherwise. This procedure was also used to help in some English immigration cases.[6]
Unfortunately this method, while powerful in its ability to differentiate individuals, was limited by the quantity and quality of DNA required for an unambiguous result because it required a large amounts of un-degraded sample DNA and in addition it was laborious for the amount of time it took to obtain a result.

In summary RFLP analysis of VNTRs has several drawbacks, including:

- The process is extremely laborious and time-consuming
- Radioactive probes pose health and disposal risks (although chemiluminescent technology eliminated this risk)
- A relatively large amount of sample is required to perform the tests
- The method requires high molecular weight, un-degraded DNA
- The use of yield gels is an essential, but time consuming, step in the analysis not only to estimate the amount of DNA recovered but also to determine the suitability of the sample for analysis

1.2 Polymerase Chain Reaction (PCR)

The field of molecular biology was revolutionized by the invention of the polymerase chain reaction (PCR), technology that is ideally suited for the analysis of forensic DNA samples because it's sensitive and rapid and not has limited by the quality/quantity of DNA as the RFLPs method. This revolutionary method was developed in April 1983 by Kary Mullis and some members of the Human Genetics group at the Cetus Corporation (now Roche Molecular Systems) Mullis received in 1993 the Nobel prize for it. [7]

The method relies on thermal cycling consisting of cycles of repeated heating and cooling of the reaction for DNA melting and enzymatic replication of the DNA. Primers containing sequences complementary to the target region along with a DNA polymerase are key components to enable selective and repeated amplification. As PCR progresses, the DNA generated is itself used as a template for replication, setting in motion a chain reaction in which the DNA template is exponentially amplified.

Each cycle has three steps:

- The two DNA strands are denatured by heat.
- The sample is then cooled to allow the primers to anneal to the DNA segments.
- The temperature is raised to allow the DNA polymerase to add nucleotides to extend the primers to produce a copy of each DNA template strand

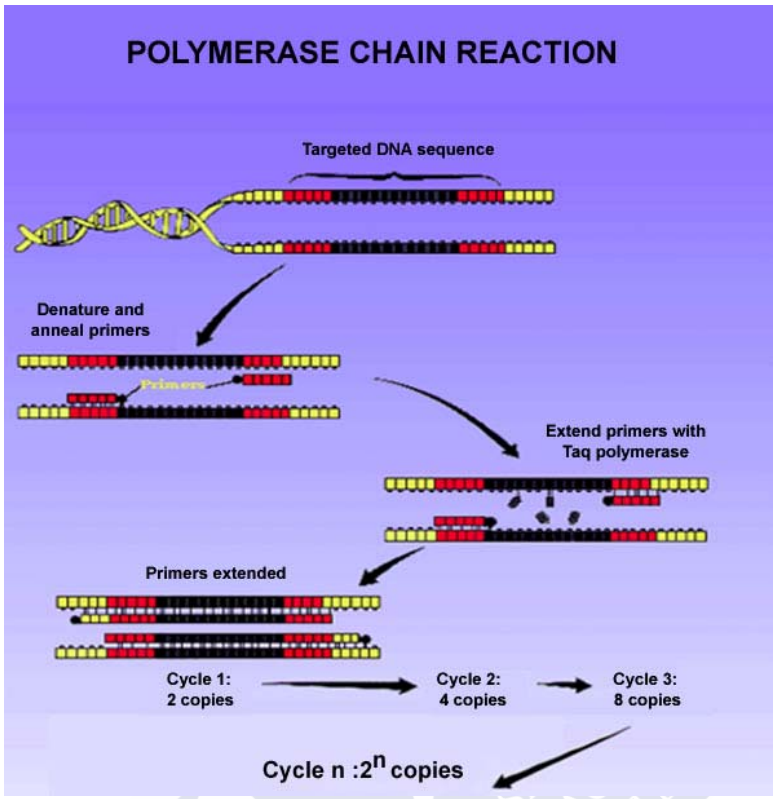


Fig.2 *Polymerase Chain Reaction steps*

The PCR product is sometimes referred to as an amplicon. Each cycle results in the doubling of amplicons. The result is an exponential accumulation of the specific target fragment, approximately 2^n, where "n" is the number of cycles of amplification performed. However, the process loses efficiency at higher cycle numbers. After 30 cycles, approximately a billion copies of the target DNA template are generated.

Polymerase chain reaction (PCR) is used to make millions of exact copies of DNA from a biological sample. DNA amplification with PCR allows DNA analysis on biological samples as small as a few skin cells. The ability of PCR to amplify such tiny quantities of DNA enables even highly degraded samples to be analyzed. Great care, however, must be taken to prevent contamination with other biological materials during the identifying, collecting, and preserving of a sample.

The PCR process was originally performed manually. The thermolabile Klenow DNA polymerase was used and had to be replenished at the beginning of each cycle.

The subsequent introduction of *Thermus aquaticus* (Taq) polymerase, a thermostable DNA polymerase, represented a considerable advance. Taq DNA polymerase is the most widely used polymerase in forensic DNA analysis and is available from multiple vendors activity. Ampli *Taq* Gold® DNA Polymerase (supplied by Applied Biosystems) is a chemically modified form of Taq DNA Polymerase, which is delivered in an inactive state and requires a pre-PCR heating step to be activated. [8,9]

PCR is now a common and often indispensable technique used in medical and biological research labs for a variety of applications. More than one region can be copied simultaneously by adding more than one primer set to the reaction this is known as multiplexing. Primer design and the optimization of thermal cycling parameters are more complex with multiplex reactions than for a single-locus reaction.

With the introduction of PCR it became possible to analyze another type of markers called Short Tandem Repeat (STRs).[10]

1.3 Applications of DNA typing

Since 1987, FBI and police labs around the U.S. have begun to use DNA fingerprints to link suspects to biological evidence - blood or semen stains, hair, or items of clothing - found at the scene of a crime, so since that time, human identity testing using DNA typing methods has been widespread. The past 15 years have seen tremendous growth in the use of DNA evidence in crime scene investigations as well as paternity testing. DNA typing has become the most important tool for the identification. Today public forensic laboratories and private paternity testing laboratories conduct hundreds of thousands of DNA test annually and a lot of cases have been decided with the assistance of DNA fingerprint evidence. DNA typing has greatly expanded the sources of evidence that can be tested, while simultaneously reducing the amount of evidence necessary to perform a conclusive test. DNA profiles can be obtained from any source of biological material, provided, it contains nucleated cells. Furthermore, DNA is resistant to many conditions that would destroy other compounds of forensic interest such as polymorphic proteins and blood group substances.[11,12]

Practical applications of forensic DNA typing include :

a) Criminal Identification and Forensics

DNA isolated from blood, hair, skin cells, or other genetic evidence left at the scene of a crime can be compared, with the DNA of a criminal suspect to determine guilt or innocence. DNA profiles are also useful in establishing the identity of a homicide victim, either from DNA found as evidence or from the body itself.

b) Personal Identification

Like the fingerprints that came into use by detectives and police labs during the 1930s, each person has a unique DNA profile. Unlike a conventional fingerprint can be altered by surgery, a DNA fingerprint is the same for every cell, tissue, and organ of a person. It cannot be altered by any known treatment. Consequently, DNA fingerprinting is rapidly became the primary method for identifying and distinguishing among individual human beings in particular in Disaster Victim Identification cases.

c) Paternity and Maternity

Because a person inherits his/her DNA from parents, STRs patterns are so specific and can be used to establish paternity and maternity as well as more complicated cases of confirming legal nationality and, in instances of adoption, biological parenthood.

2. How DNA typing works : statistical evaluations

2.1 Criminal Caseworks

DNA forensic scientists are presented with the situation were they are given two samples related to a crime scene, about which they know nothing in advance, and are asked whether or not they are identical. Only one-tenth of a single percent of DNA (about 3 million bases) differs from one person to the next. These variable regions are used to generate a DNA profile of an individual, using samples from blood, bone, hair, and other body tissues and products. In criminal cases,this generally involves obtaining

samples from crime-scene evidence and a suspect, extracting the DNA, and analyzing it for the presence of a set of specific DNA regions (markers). [13,14]

DNA profiles are compared to determine whether the suspect's sample matches the evidence sample found at crime scene. A marker by itself usually is not unique to an individual; if, however, two DNA samples are alike at some regions, odds are great that the samples are from the same person. If the sample profiles don't match, the person did not contribute the DNA at the crime scene. If the patterns match, the suspect may have contributed the evidence sample. The possibility of a close relative (typically a brother) of the accused being in the pool of potential contributors of crime scene evidence should be considered in case-specific context. It is not appropriate to proffer that a close relative is a potential contributor of the evidence when there are no facts in evidence to suggest this instance is relevant. However, if a relative had access to a crime scene and there is reason to believe he/she could have been a contributor of the evidence, then the best action to take is to obtain a reference sample from the relative. After all, this scenario should be sufficient probable cause for obtaining a reference sample. Typing with the same battery of short tandem repeat (STR) loci will resolve the question of whether or not the relative carries the same DNA profile as the accused. A typical DNA case involves the comparison of two samples – an unknown or *evidence* sample and a known or *reference* sample, such as a blood/saliva sample from a suspect. If DNA profiles obtained from the two samples are indistinguishable (they "match"), that of course is evidence for the court that the samples have a common source.

- If the DNA profile obtained from the two samples are distinguishable (they " NOT match"), that of course is evidence for the court that the samples have a different source

- If the DNA profile obtained from the two samples are indistinguishable (they "match"), that of course is evidence for the court that the samples have a common source

Once an individual's STR profile is identified, it is statistically improbable that anyone else in the world will have the same profile, unless that person has an identical twin. Identical twins (twins derived from a single fertilized egg) have identical STR DNA profiles. For evidence yielding full single source DNA profiles, it's possible to calculate the *random match probabilities and likelihood ratio.*

a) Random match probability

It's the chance of a random DNA profile match within a given population and is the reciprocal of the DNA profile frequency.

PI = the probability that a match would occur by chance.

A DNA profile frequency is estimated by determining the genotype frequency for each locus and then multiplying the frequency across all loci. Rare genotypes provide stronger evidence, and population databases sorted by race will yield somewhat different results, but it is important to understand that this is a representation of how rare a DNA profile is in a representative population.

Population data with allele frequencies for used markers are collected for different populations and contain only DNA profiles from anonymous donors of specific populations tested.

b) Likelihood Ratio

A likelihood ratio (LR) is a ratio of two probabilities of the same evidence under two mutually exclusive hypotheses, specifically the position of the prosecution and the position of the defence. It conveys the relative support for the weight of DNA evidence under the hypothesis that the defendant is the source of the DNA profile, versus an unrelated individual from the population at large.

While interpretation of the strength of the statistical value can be variable, and should ultimately be considered in context with all case circumstances

c) Combined Probability of Exclusion

The combined probability of exclusion can be used to conservatively interpret complex DNA mixtures. This calculation provides an estimate of the portion of the population that has a genotype of at least one allele not observed in the DNA mixture. This is a conservative approach since all other alleles not observed are considered and an individual can be excluded if he has any allele at any locus that is not detected in the mixture.

2.2 Paternity Test

In paternity tests to determine if the alleged father is the true biological father, the DNA profiles of the child, mother, and alleged father are compared. A child inherits two different alleles at each genetic locus—one from the mother and one from the father. If a child has an allele that the mother does not have, this obligate allele has to come from the biological father.

If the tested man (alleged father) does not have the genetic characteristics necessary to be the biological father of the child, the result is an exclusion (the alleged father is not the biological father).

If the tested man's DNA has the same allele as the obligate allele does, the result is an inclusion. In this last case the alleged father has the same allele as the obligate allele and a Paternity Index (PI) can be calculated. This is the relative probability that the alleged father and not an unrelated, randomly selected male of the same ethnic background transmitted the obligate allele to the child.

This is a likelihood ratio and is presented in the formula X/Y, where X is the chance that the alleged father could transmit the obligate allele and Y is the chance that an unrelated man of the same race could have the allele.

X is assigned the value of 1 if the alleged father is homozygous for the allele of interest and 0.5 if the alleged father is heterozygous.

The probability of an unrelated, randomly selected man possessing the obligate allele is determined by using a database that lists the frequency distribution of individual alleles. If there is more than one obligate allele, the individual paternity

indexes can be multiplied and the total across all loci is called the Combined Paternity Index (CPI). This is a measure of the strength of the genetic evidence and is an odds ratio, not a probability.

CPI can range from 0 to infinity, an interpretation of the CPI is as follows:

- If CPI is between 0-1, the genetic evidence is more consistent with non-paternity than paternity.

- If CPI>1, the genetic evidence is more consistent with paternity than non-paternity

It is normal practice to establish a threshold value for CPI, above which it is accepted that the tested man is the true biological father. This threshold is 1000 in Europe, but can be as low as 100 in the USA.

Sometimes a likelihood ratio is converted into a probability. This probability is known as the probability of paternity. This formula tests the hypothesis that the alleged father is indeed the biological father of the child. For example, a value of 99% reflects a 99% probability that the hypothesis is correct and a 1% probability that it is not.

PI sometimes is called L and the probability of paternity is W (from the German word *Wahrscheinlichkeit*, "probability").

W and L are related as :

$W = L / (1+L)$, or $L = W / (1-W)$.

Probability of paternity is not widely used in the United States. A more common approach, similar to the probability of exclusion, is the Random Man Not Excluded (RMNE) statistic. This is the proportion of the population that could contribute all of the obligate alleles and therefore could not be excluded, or would be falsely included.

A single locus RMNE is calculated by $1-(1-p)^2$.

Combining the RMNE statistics over all loci gives the combined RMNE (CRMNE) which is equivalent to the CPI.

The value of the CRMNE is typically small (less than one), and is analogous to 1-CRMNE or exclusionary power (PE). PE represents the probability of excluding a falsely accused man.

3. DNA databases

DNA analysis is a powerful crime-fighting tool for prosecuting criminals and exonerating the innocent. A national DNA database is a government database of DNA profiles which can be used by law enforcement agencies to identify suspects of crimes. The first government database was set up by the United Kingdom in April 1995, The second one in New Zealand, and than in France in 1998. [15,16]

The growing public approval of DNA databases has seen the creation and expansion of many states' own DNA databases. California currently maintains the third largest DNA database in the world. The size of DNA database, and its rate of growth, is giving concern to civil liberties and political groups in the UK, where police have wide-ranging powers to take samples and retain them even in the event of acquittal.

Originally intended for sex offenders, they have since been extended to include almost any criminal offender.DNA databases are effective because a majority of crimes are committed by repeat offenders. In fact has been evaluated that sixty percent of those individuals released from prison for violent offenses and subsequently released were re-arrested for a similar offense in less than 3 years.

The value of the DNA database is in its ability to apprehend criminals that are not direct suspects in a case and to prevent further victims from crimes committed by those individuals.

In particular :

- Link an unknown sample to a convicted offender. This gives the investigator the name of a previously unidentified suspect.

- Link an unknown sample to a solved case. This would also identify a suspect for the investigator.

- Link two or more unsolved cases. Linking unsolved cases can help an investigator look for similarities in the crimes, define geographical areas, compare victim

statements, etc. If the crimes occurred in different jurisdictions, linked cases would enable the investigators to compare notes and possibly develop a suspect.

- Exclude suspects. This can often be as important as identifying a suspect. Exclusion of a particular individual can allow the investigator to change the focus of the investigation

The administration of DNA databanks are handled differently from state to state. Regulations for their operation are found in statute and administrative code, and they are overseen by various state departments.

In USA each state has established a central repository to store and compare DNA profiles to make the best use of this tool. Profiles loaded into these databanks are searchable through the FBI CODIS. The CODIS software allows forensic laboratories at the national, state and local levels to compare DNA samples and use the database in accordance with state and local laws. CODIS is split into three parts – LDIS (Local DNA Index System), SDIS (State DNA Index System) and NDIS (National DNA Index System). To participate in NDIS and make their profiles nationally searchable, states must take and analyze samples in accordance with the FBI's Quality Assurance Standards for Forensic DNA Testing Laboratories**.** Each state has a laboratory that participates in NDIS. All DNA profiles are entered into the database and then evidences profiles are compared to other samples or to suspects profiles or to DNA found at other scenes to close open cases.

CODIS contains over 10,194,686 offender profiles and 395,105 forensic profiles as of September 2011. Ultimately, the success of the CODIS program will be measured by the crimes it helps to solve. CODIS's primary metric, the "Investigation Aided," tracks the number of criminal investigations where CODIS has added value to the investigative process. As of September 2011, CODIS has produced over 161,100 hits assisting in more than 155,100 investigations. (data from U. S. Department of Justice - Federal Bureau of Investigation CODIS Unit**).**

For a criminal DNA database to be successful, both convicted offender DNA samples must be entered and crime scene material from case where there's no suspect must be tested. All convicted offender samples are blood or saliva which improves the

capability for automating the DNA typing process. Collecting the actual samples can be a challenge considering the fact that the convicted offenders know that their blood or saliva could be used to catch them committing other crimes in the future or match them to previous unsolved crimes they committed.
On the other hand, forensic cases can involve the examination of a dozen or more pieces of biological evidence from a variety of formats (semen stain,bloodstains, hair, so on) which makes them much more complex. In spite of the time and effort required to obtain results on crime scene samples, it's working these cases that make DNA databases effective.

Several categories of persons may be included in a DNA-database.

- Convicted persons, persons who have been found guilty of a crime by a court of law and may (or not) be (conditionally) convicted to imprisonment, a penalty, labor, hospitalization or combinations of those. A conviction can be overturned by a successful appeal to a higher court. In some countries it is possible to include persons in the National DNA-database who have been convicted in the past and who have already completed their imprisonment. This is called retrospective sampling.

- Suspects, persons who have not yet been found guilty but are officially the subject of investigation and/or prosecution.

- Arrestees, persons who have been taken into custody by the police but are not (yet) a suspect.

- Volunteers, persons outside the abovementioned categories who have agreed to give a DNA-sample for investigative purpose.

Most countries develop and implement their DNA testing programs, including databases, through coordinating bodies or other consensus working groups.
For instance, the FBI and SWGDAM are responsible for setting standards, for training, and for development for the United States. Similarly, in Europe the ENFSI and

EDNAP represent government institutions and coordinate efforts to develop European DNA databases, in part by recommending markers.

Subsequently, the ENFSI DNA working group has established recommendations for DNA database management, including criteria for including and deleting DNA profiles In the USA 13 loci are required for inclusion of a reference profile in the National DNA-database of the USA (CODIS).

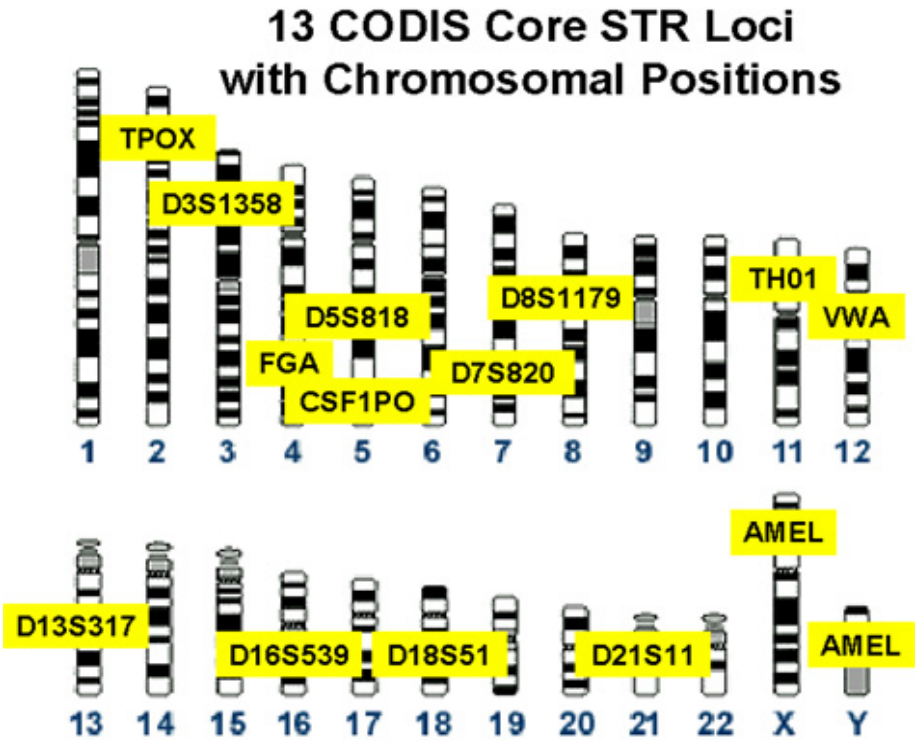


Fig.3 : *select Codis core STR loci*

The Council of the European Union has already invited Member States in 1997 to consider establishing DNA Databases. In 2001 a European Standard Set (ESS) of loci was established to enable comparison of DNA-profiles from different countries.

The EU-Council resolution 2009/C 296/01 calls upon European countries to use the European Standard Set (ESS) as a minimum to enable international comparison of DNA-profiles. The Interpol Standard Set of Loci (ISSOL) is equal to the European Standard Set plus the Amelogenin locus. The European Standard Set of Loci until

recently contained only 7 loci. This was enough for occasional exchanges of DNA-profiles between countries. In 2005 some European countries firmed the Treaty of Prum with the purpose of cross-border cooperation, in the fight against terrorism, cross-border crime and illegal migration. As a consequence all 27 EU countries became able to exchange DNA profiles in the near future. In June 2008 the Council of the European Union has converted the Treaty of Prüm into EU legislation (The EU-Prüm-Decision). However when massive exchanges of DNA-profiles are undertaken (as has been made possible by the Interpol DNA-database and the EUPrüm- Decision) 7 loci are not enough because the chance of adventitious matches becomes unacceptable. In addition each DNA-database contains a significant portion of partial profiles with much higher probability to match randomly.

That is why ENFSI has recommended that the European Standard Set of Loci should be extended by 5 additional loci and the Council of the has taken over this recommendation in 2009. [17]

National DNA databases have become one of the most efficient tools to provide intelligence about unknown perpetrators in criminal investigations.

At present, almost six million DNA profiles from both suspects and convicted offenders are stored in European databases, and more than one million person-to-stain and stain-to-stain hits have been obtained.

One of the major challenges for maintaining a DNA database is the issue of privacy and security of the information stored in the database.

Blood or saliva samples contain genetic information that could be used against an individual if not handled properly. The issue of privacy is approached in two ways. First DNA markers analysed for forensic purpose are in non coding regions of the DNA and are not known to have any association with a genetic disease or any other genetic predisposition. Thus the information in the database is only useful for human identity testing. Second, no names of individuals or other characterizing data is stored with the DNA profiles. Specific case data are secured and controlled by local law enforcement agencies. Only the crime laboratory that submitted the DNA profile has the capability to link DNA results with a known individual.

Another important facet to the privacy and security of the information in DNA databases is the fact that access to it it's solely for law enforcement purposes and there're strict penalties for anyone using the information or samples for any purpose other than law enforcement. No personal information, criminal history information or case-related information are contained within database.

After the offenders DNA specimen is tested, the resulting profile is expressed as a series of numbers much like a long social security number. These numerical profiles are entered into a computer database in an offender file.

When DNA evidence is recovered from crime scene evidence (for example, bloodstains, saliva, hairs, vaginal swabs,etc), the evidence DNA profile is entered into the database computer. A comparison is then made by the computer to determine if the evidence profile matches a known offender.

There are three possible results:

1) there is a match between a suspect and the DNA in the evidence, then that suspect is included in the group of individuals who could be the source of that evidence;

2) there is not a match between the samples. The suspect could not be the source of the evidence. They are excluded from the group of individuals who could have contributed the specimen;

3) Data is inconclusive, meaning that it is not possible to make a determination.

This can be caused by a variety of factors, and usually occurs when the DNA is old and heavily contaminated.

A match in a national DNA database that links a crime scene to an offender who has provided a DNA sample to a database that link is often referred to as a *cold hit*.

A cold hit is of value in referring the police agency to a specific suspect

If a hit is made, the investigator is promptly advised of the match. The match is used as a probable cause to seek a new DNA sample from the suspect which will be used to

perform confirmation testing for subsequent use in a judicial proceeding. When the software identifies a potential match, the laboratories responsible for the matching profiles are notified. After the match has been confirmed by qualified DNA analysts, which often involves retesting of matching convicted offender DNA sample, laboratories may exchange additional information, such as names and phone numbers of criminal investigators and case details. If a match is obtained the identity and location of the convicted offender is determined and an arrest warrant procured.

Otherwise if a match between the evidence specimen and an offender is not found in the database, the evidence DNA profile is compared with profiles from evidence recovered from other unsolved crimes. In this way, it can be determined if the same perpetrator may be linked to other assaults even though his identity is not yet known.

.

4. References

[1]Watson J.D. and Crick F.H.C. (1953),A Structure for Deoxyribose Nucleic Acid, Nature 171: 737-738

[2]Jeffreys A.J., Wilson V., Thein S.W.(1984), Hypervariable 'minisatellite' regions in human DNA, Nature 314: 67–73.

[3]Jeffreys AJ, Wilson V, Thein SL.(1985), Individual-specific 'fingerprints' of human DNA, Nature, 316: 76

[4]Southern E.M.(1975),Detection of specific sequences among DNA fragments separated by gel electrophoresis, J Mol Biol., 98:503-517.

[5]Sanders J.(2000),Forensic Casebook of Crime, London: True Crime Library, Forum Press. pp. 229.

[6]Jeffreys A.J., Brookfield J.F., Semeonoff R., (1985) Positive identification of an immigration test-case using human DNA fingerprints, Nature 6: 317

[7]Mullis K. (1990), The unusual origin of the polymerase chain reaction, Scientific American 262 (4): 56–61, 64–5.

[8]Sambrook J. and Russel D.W. (2001), Molecular Cloning: A Laboratory Manual (3rd ed.). Cold Spring Harbor, N.Y.: chapter 8: In vitro Amplification of DNA by the Polymerase Chain Reaction

[9]Saiki, RK; Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA (1988), Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase,Science 239: 487–91.

[10]Butler J, (2005),Forensic DNA Typing – Biology, Technology, and Genetics of STR Markers, Academic Press, ISBN 0-12-147952-8

[11]Houck M.M., Siegel J.A.(2006),Fundamentals of forensic science. Burlington, MA : Elsevier Academic Press, 2 - Chapter 1

[12]Kiely T.F., (2006),Forensic evidence: science and the criminal law, CRC Press, Taylor & Francis.

[13]Evett IW, Weir BS (1998),Interpreting DNA Evidence, Sinauer Associates. ISBN 0-87983-155-4

[14]Buckleton J, Triggs CM, Walsh SJ (2005),Forensic DNA Evidence Interpretation CRC Press, 534

[15] Schneider PM, Martin PD (2001),Criminal DNA database: the Europe situation, in Forensic Sci Int.119(2):232-8.

[16]Linacre A.,(2003),The UK national DNA database, The Lancet, 361:1842

[17]DNA-Database Management Review and Recommendation ENFSI DNA Working Group - April 2010

Chapter II

DNA MARKERS : SHORT TANDEM REPEATS LOCI (STRs)

1. AUTOSOMAL STRs MARKERS

1.1 Introduction

The human genome is full of repeated DNA sequences that are widespread throughout almost every chromosome in the genome. surrounding the chromosomal centromere. These repeated sequences come in various sizes and are classified according to the length of the core repeat units, the number of contiguous repeat units, and/or the overall length of the repeat region.Minisatellites (variable number of tandem repeats, VNTRs) have core repeats with 9-80 bp, while microsatellites (short tandem repeats, STRs) contain 2-5 bp repeats and are typically in the non-coding intron region. An individual inherits one copy of an STR from each parent, which may or may not have similar repeat sizes.

The number of repeats in STR markers can be highly variable among individuals, the variety of alleles (generally more than 10 alleles for the commonly used STRs) present in a population is such that a high degree of discrimination among individuals in the population may be obtained when multiple STR loci are examined. That means a multi locus STR DNA profile is unique.

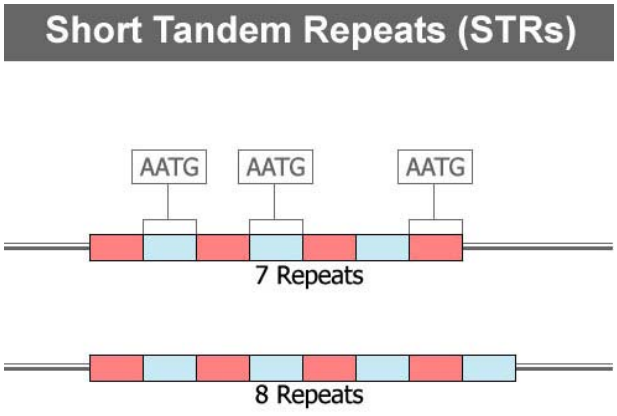


Fig.1: *example of STRs structure*

There are hundreds of STR systems which have been mapped throughout the human genome**.** Several dozen have been investigated for application to human identity testing**.** This make these STRs have become important in several fields including genetic mapping, linkage analysis, and human identity testing. It is often challenging to obtain PCR amplification products from forensic samples because either the DNA in those samples is degraded, or mixed, such as in a sexual assault case.

STRs have become popular DNA markers for the forensic community because they are easily amplified by polymerase chain reaction (PCR) and show several benefits that make them especially suitable for human identification, such as:

- high heterozygosity
- regular repeat unit
- distinguishable alleles
- robust amplification
- low mutation rate

The smaller size of STR alleles make STR markers better candidates for use in forensic applications, in which degraded DNA is common. PCR amplification of degraded DNA samples can be better accomplished with smaller target product sizes. Moreover because of their smaller size, STR alleles can also be separated from other chromosomal locations more easily to ensure closely linked loci are not chosen. Closely linked loci do not follow the predictable pattern of random distribution in the population, making statistical analysis difficult. [1]

Because of these characteristics, STRs with higher power of discrimination are chosen for human identification in forensic cases on a regular basis. It is used to identify victim, perpetrator, missing persons, and others. It makes them effective for human identification application since for this purpose, it is important to have DNA markers that exhibit the highest possible variation in order to discriminate between samples.

In October 1993, the DNA Commission of the International Society of Forensic Genetics (ISFG) recommended the nomenclature for STR systems which is commonly used today. Alleles are generally named by the number of repeats which they contain.

When an allele does not conform to the standard repeat motif of the system in question, it should be designated by the number of complete repeat units and the number of base pairs of the partial repeat. [2-4]

1.2 A brief History

Beginning in 1996, the FBI Laboratory launched a nationwide forensic science effort to establish core STR loci for inclusion within the national database known as CODIS (Combined DNA Index System). The 13 CODIS loci are CSF1PO, FGA, TH01, TPOX, VWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51 and D21S11. These loci are nationally and internationally recognized as the standard for human identification.

In 1999 the DNA working group of the European Network of Forensic Science Institutes (ENFSI) decided on a European Standard Set (ESS), which includes seven loci: TH01, vWA, FGA, D21S11, D3S1358, D8S1179 and D18S51. These loci have been confirmed by a resolution of the European Council in 2001 and now form the core of all national DNA databases in Europe.

Due to the overwhelming success of DNA databases, a political process was initiated by a number of European countries to establish a legal basis for exchanging DNA database profiles between countries in criminal investigations. This led to the Treaty of Prüm, which was signed in 2005 with the purpose of stepping up cross-border cooperation, particularly in combatting terrorism, cross-border crime and illegal migration. Subsequently, the ENFSI DNA working group has established recom mendations for DNA database management, including criteria for including and deleting DNA profiles, matching rules, and handling of partial profiles. Furthermore, the occurrence of adventitious matches between DNA profiles that have no case-related connection has been addressed in detail. When massive exchanges of DNA profiles are undertaken following the implementation of the Treaty of Prüm, the seven ESS loci will not be sufficient because the chance of adventitious matches will no longer be negligible. In addition, each DNA database contains a significant portion of partial profiles with an even higher probability to match randomly.

The ENFSI and EDNAP groups met in Glasgow in 2005 and discussed extension of the ESS and recommendations for additional European STR systems. Since the ESS loci are typically part of larger multiplexes with 10–15 loci, which are already used in forensic laboratories throughout Europe, it would have been straightforward to choose among these loci. However, at the same meeting, the results of a collaborative exercise carried out by the EDNAP group to examine typing of heavily degraded DNA samples were presented This exercise addresses the fact that many casework samples include only minimal amounts of DNA or DNA that is degraded due to environmental A decision was adopted by the ENFSI and EDNAP groups to increase the number of ESS loci and a recommendation was published to include more robust loci with short amplicons, rather than already established STRs which frequently fail to give results, and/or have a poor power of discrimination.

In particular Europe adopted 5 new loci D2S441 D10S1248, D22S1045, D1S1656, and D12S391.[5]

Short tandem repeats (STRs) located on autosomes are the genetic markers of choice in paternity investigation and they are also the most widely used in other cases of kinship analysis. Nevertheless, in some complex cases, independent of the number of polymorphisms being typed, autosomal markers convey very little information. Depending on the parentage constellation available for the analysis, as well as the gender of the subjects, this problem can sometimes be solved by using markers with different modes of transmission.

Therefore, most forensic laboratories are nowadays prepared to analyze lineage markers (Y-chromosome and mitochondrial DNA) and many have recently introduced the analysis of X-STR markers in their routine.

2. ALTERNATIVE STRS MARKERS : Y- STRs

2.1 Introduction

Y-STRs are Short Tandem Repeats found on the male-specific Y Chromosome. The human Y-chromosome has often been considered an evolutionary relic of the X chromosome. The Y-chromosome has retained the ability to dictate gender but has little other functional significance. Recent studies have demonstrated that it possesses numerous functional genes, including some that appear to be critical for normal male development. Approximately 300 million years ago, the X and Y-chromosomes were true homologues, comparable in size and genetic content. Through the passage of time, the Y-chromosome underwent a series of deletion mutations reducing it to its present size of approximately 50 megabases (Mg). This notwithstanding, significant X chromosome sequence homology still persists.

The chromatin of the Y-chromosome exists in at least three functionally different forms including:

- Pseudoautosomal regions (PARs)
- Euchromatin
- Heterochromatin

The PARs, located in the telomeric regions of the chromosome, pair and recombine with the X-chromosome during male meiosis. The euchromatin (containing the functional genes) and the transcriptionally inert heterochromatin form the non-recombining region (NRY) of the Y-chromosome. Sequencing of the euchromatic region has revealed a patchwork of three distinct sequence classes.

The coding genes, mostly found on the short arm of the Y Chromosome, are vital to male sex determination, spermatogenesis and other male related functions.

The NRY region of the Y-chromosome is inherited in a patrilineal manner in which a haplotype of physically linked genetic markers is transmitted unchanged, barring the occasional rare mutation, from father to son. Reduced genetic variability results from:

- Non-independent segregation of genetic markers on the Y-chromosome
- Enhanced genetic drift potential (due to the smaller effective population size of the Y-chromosome – one-fourth that of autosomes)

Thus, significantly more Y-chromosome markers would be required to provide the same ability to discriminate individuals (the discriminating power) as that obtained by autosomal STR markers

The Y-STRs are polymorphic among unrelated males and are inherited through the paternal line and remains virtually unchanged through many generations.

By examining specific locations on the Y chromosome, we can generate a Y-STR profile for each male tested. Males who are related through their fathers will tend to have the same or similar Y-STR profiles, and males who are not related will likely have different Y-STR profiles.

In humans, the Y chromosome spans about 58 million base pairs and represents approximately 2% of the total DNA in a male cells. The human Y chromosome contains 86 genes, which code for only 23 distinct proteins. Traits that are inherited via the Y chromosome are called holandric traits.

The human Y chromosome is unable to recombine with the X chromosome, except for small pieces of pseudoautosomal regions at the telomeres (which comprise about 5% of the chromosome's length). These regions are relics of ancient homology between the X and Y chromosomes. The bulk of the Y chromosome which does not recombine is called the "NRY" or non-recombining region of the Y chromosome [6].

2.2Y-STRs applications

The analysis of Y-chromosome short tandem repeats (YSTRs) has become a very useful tool, both in evolutionary studies and forensic casework. [7-9]

Although more than three hundred STR loci have been described on the Y-chromosome a much more limited number have been appropriately evaluated for forensic casework use and some of these have presented a particular challenge for assay design. The Y-STR loci comprise di-, tri-, tetra-, and penta-nucleotide repeats

with the di-nucleotides exhibiting the most polymorphism but an excessively high level of stutter artifacts.

The ability to identify male-specific DNA renders polymorphic Y-chromoso mal sequences an invaluable addition to the standard panel of autosomal loci used in forensic genetics. Y-STR haplotyping is particularly important for sensitive typing of male DNA in mixed stains as well as for rapid assortment of biological crime scene evidence. Males commit the majority of violent crimes. For example, the U.S. Bureau of Justice Statistics reports that males commit about 80% of all violent crimes and 95% of sexual offenses in the United States. Many times autosomal STR (Short Tandem Repeats) markers are able to fully discriminate between unrelated individuals, but there are several circumstances in which Y-chromosome polymorphisms are useful in forensic analysis.

In a sexual assault case, evidence, such as vaginal swabs, contain both female and male DNA. Differential extraction is often used to separate the male component from the female component, but sometimes, the two components cannot be separated completely. As a result, the female component could exist prominently even in the male component after separation. When the sample undergoes the PCR amplification process, the female DNA component is amplified as well, sometimes masking the male DNA, which makes analysis difficult. This masking effect obviously does not occur when Y-STRs are examined.

Since there is no Y-STR in the female evidence, the only contribution of Y-STR can only come from the assailant(s) in a sexual assault case. So the male component is easily detected, since only this part of DNA will be amplified. Thus the ability to specifically detect a male profile could obviate the need for the time-consuming and frequently inefficient differential extraction procedure for the separation of sperm and non-sperm fractions

The Y-STRs analysis is especially helpful when there are more than one assailant since the mixed pattern in the evidence can help to identify them.

Y chromosome specific systems may prove invaluable for the identification of the genetic profile of the male component in mixed male/female specimens from non-sexual assault cases, in which the female portion is present in overwhelming

quantities, (not balanced mixture) where there's a masking effect due to very small quantity of male DNA in the sample. Performing Y-STR testing can help to identify all males who have contributed to the evidence.

Male specific systems may also aid in the investigation of cases involving mixtures or degraded DNA specimens (displaying partial autosomal STR profiles) by providing additional statistical discriminating power, but in some circumstances, Y-STR data might be the only data that can be obtained. It is important to note that a Y-STR haplotype is shared by males from the same paternal lineage. This fact must be taken into account when drawing conclusions. In fact two individuals that share the same Y-STR haplotype are very likely related through the same paternal line.

STRs loci are located on the non-recombining part of the Y-chromosome and, therefore, should be considered linked as a single locus because they are inherited as a block of linked haplotypes, so estimates of the multi-locus frequency cannot proceed by the product rule.[10,11] Although more than three hundred STR loci have been described on the Y-chromosome a much more limited number have been appropriately evaluated for forensic casework use and some of these have presented a particular challenge for assay design. The Y-STR loci comprise di-, tri-, tetra-, and penta-nucleotide repeats with the di-nucleotides exhibiting the most polymorphism but an excessively high level of stutter artifacts.

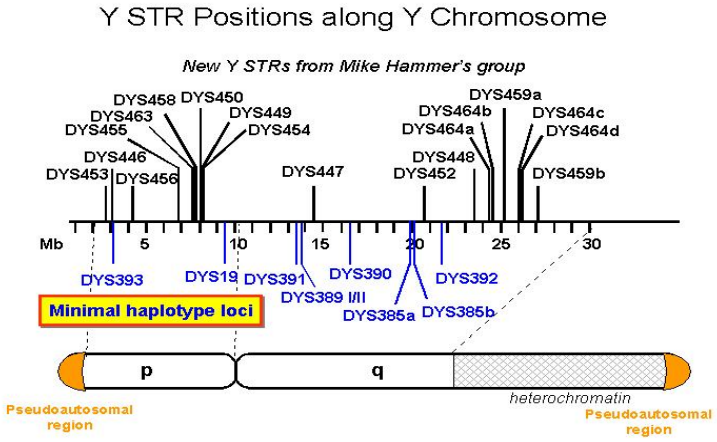


Fig.2 : *example of Y-STRs markers*

Resuming Y-STRs may be useful in :

a) Sexual assault cases

- when a mixture is present where either more than one male may be present, it can be used to determine the number of males present
- a male and female mixture is present, the male sample can be masked due to the high concentration of female sample; the Y-STR method will ignore the female portion of the mixture, and only focus on the male sample.the male sample.
- Deposition of semen by an azoospermic or oligospermic males.
- Cases of oral sodomy where only trace amounts of male buccal epithelial cells may be present.
- Normal post-coital degradative and semen sample loss processes that occur with the passage of time.
- Determination of the presence of the number of semen donors in cases of multiple perpetrator rape.

b) Criminal paternity analysis

- When an alleged father is not available for testing, male relatives of the alleged father can be used as a reference; only male children can be tested via this method.
- When a mother is not available for testing, the Y-STR method can be used in conjunction with nuclear DNA STR testing to prove paternity again, only for male children.

c) Disaster victim identification and/or missing person

- male individual can be identified by typing a male relative (such as a son, brother, father, nephew, or uncle) who can be used as a reference to provide a DNA match.

d) Ancestry

- since Y-STRs are polymorphic among unrelated males and are inherited through the paternal line with little change through generations chromosomal profiling can trace back paternal lineages into the past in genealogical testing.

e) Mixed samples

- Y-STR is also used for non-sexual assault cases where mixed samples are collected from evidence. Sometimes, regular STR will cause the masking effect if there is a very small quantity of male DNA in the mixed sample. Performing Y-STR testing can help to identify all males who have contributed to the evidence.

2.3 Population Study

Since the Y STR loci are located on the NRY part of the Y-chromosome and are inherited as a block of linked haplotypes, estimates of the multi-locus frequency cannot proceed by the product rule. Instead an estimate of the frequency of occurrence of a particular haplotype requires the counting method which is based upon how many times a particular multi-locus haplotype is observed in a particular database.

Thus a Y-STR database must consist of haplotype frequencies rather than allele frequencies.[13] The ability to obtain a reliable estimate thus depends upon the nature of the databases that are used. Specifically, the ethno-geographic composition of the database, the number of individuals deposited therein and the number of searchable Y-STR loci are factors that should be maximized to increase confidence in the accuracy of a frequency estimate.

In the year 2000 the **Y-STR haplotype reference database (YHRD)** was established in order to pursue two important objectives:

- the generation of reliable Y-STR haplotype frequency estimates for Y-STR haplotypes to be used in the quantitative assessment of matches in forensic and genealogical casework.

- the assessment of male population stratification among world-wide populations as far as reflected by Y-STR haplotype frequency distributions.

A large and continuously growing number of diagnostic and research laboratories have joined in a collaborative effort to collect population data and to create a sufficiently large reference database. The database "YHRD - Y chromosome haplotype reference database" is interactive and can be searched at http://www.yhrd.org/index.html .

The individuality of the male-specific part of the Y chromosome can be optimally explored by the Y-STR haplotype analysis using a set of highly variable short tandem repeat markers approved by the forensic and scientific community.

A major international multi-center study of 13 candidate Y-STR markers in 1997 resulted in recommendations for the use of the following nine core loci for standard forensic haplotyping (designated the minimal haplotype loci, MHL or minHt): DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS385a/b. [14]

Subsequent to the development of the MHL loci, additional microsatellite loci were described that proved to have utility in forensic genetics.

In the United States, the development of Y-STR core loci has followed a pattern similar to the development of the CODIS core STR loci. The Scientific Working Group for DNA Analysis Methods (SWGDAM) created a subcommittee to investigate Y-STRs. The subcommittee reviewed published data and ran several hundred population database samples to evaluate various loci. In January 2003 the Scientific Working Group on DNA Analysis Methods (SWGDAM) recommended for US forensic casework use a set of eleven core loci that included the MHL (minimal haplotype loci) loci plus DYS438 and DYS 439.[15]

With the establishment of the core loci, manufacturers have the specifications needed for the development of Y-STR kits. European laboratories have also established core Y-STR loci for forensic testing. ENFSI decided to use the same loci as the United States, minus DYS438 and DYS439.

3. ALTERNATIVE STRS MARKERS : X STRs

3.1 Introduction

Some Short Tandem Repeats are located on the X chromosome (ChrX).
The X chromosome in humans spans more than 153 million base pairs (the building material of DNA). It represents about 2000 out of 20,000 - 25,0000 genes.
Normal human females have 2 X-chromosomes (XX), for a total of nearly 4000 "sex-tied" genes (many of which have nothing to do with sex, other than being attached to the chromosome that is believed to cause sexual bimorphism. Men have, depending on the study, 1400-1900 fewer genes, as the Y chromosome is thought to have only 45 - 78 remaining genes down from an estimated 1438 -~2000.

Approximately 300 million years ago, the X and Y-chromosomes were true homologues, comparable in size and genetic content. Through the passage of time, the Y-chromosome underwent a series of deletion mutations reducing it to its present size of approximately 50megabases (Mg). This notwithstanding, significant X chromosome sequence homology still persists.

The X degenerate sequences are surviving relics from common autosomes from which the X and Y both arose and comprise 39% of contemporary euchromatin. Sequence similarity with the X-chromosome ranges from 60-96%. The X-transposed sequence comprises approximately 15 % of the euchromatin sequence and consists of two blocks that are 99% identical to Xq21 sequences. This represents an ancient, massive single X to Y transposition event that occurred 3-4 million years ago after the divergence of the human/chimpanzee lineages. Approximately 46% of the euchromatin contains ampliconic sequences which occur in seven segments that have as much as 99.9% sequence similarity to other sequences in the euchromatin. Amplicons are characterized by eight palindromic sequences and the extreme sequence conservation has been attributed to intra-chromosomal gene conversion events

The sexual chromosomes X and Y are unique and differ in several aspects from the autosomes. In the cells of human males gonosomes do not normally occur in pairs but comprise one X and one Y chromosome. Males carry one X-chromosome, so

the ChrX markers appear in hemizygous state and their haplotype is transferred to their biological daughters. Females carry two X chromosomes but only one ChrX is active per cell. In females, ChrX is present as a homologous pair and resembles autosomes in this respect. The two female X-chromosomes are prone to recombination during meiosis, which necessitate a consideration of two features that can have an impact on the interpretation and calculation of probabilities in relationship testing. [16-20] In contrast to ChrY markers, which do not recombine during meiosis, the genetic localisation is an important issue for ChrX markers when used in kinship testing. Since all ChrX markers are located on the same chromosome within an area of 240cM, working with these markers requires an exact knowledge about their genetic localisation.

The fundamental idea of making wide use of X-chromosomal markers in forensic practice developed from the experiences made during the second half of the last century in the field of the clinical genetics.

Gender identification (sex-typing) is commonly performed in conjunction with mutiplex STRs typing kit using PCR products generated from the Amelogenin gene that occurs on both the X- and Y-chromosome. A commonly used PCR primer set was first published by Sullivan et al. (1993) . It targets a 6 bp deletion that occurs on the X-chromosome, which enables amplicons generated from the X- and Y-chromosome to be distinguished from one another when by electrophoresis. Since females are X,X, only a single peak is observed when testing female DNA whereas males, which possess both X and Y chromosomes, exhibit two peaks. The ratio of amelogenin X and Y PCR products can be helpful in deciphering mixtures involving male and female DNA such as sexual assault evidence.[21]

3.2 X-STRs applications

Paternity cases involving the common trio constellation of mother, offspring and alleged father can usually be solved with autosomal STRs alone, and do not seem to require any additional or alternative markers.

X-chromosome (ChrX) genotyping can complete the analysis of autosomal (AS) and Y chromosomal (ChrY) markers very efficiently, especially in complex kinship testing cases. These insights, which arose in the late nineties and in the first years of the current decade, increasingly induced investigations on ChrX markers for forensic use.
Many forensic laboratories have introduced the analysis of X-STR markers in their routine, due to the advantages in the use of these markers in cases of complex kinship analysis markers were selected taking into account the gene diversity values reported in different populations; and the potential for multiplexing. Preference was given to simple rather than complex STRs, following the ISFG recommendations concerning locus selection for forensic application.D etailed information regarding more than 50 X-STRs has been collected (www.chrx-str.org) and used in different PCR multiplexes (Becker et al., 2008; Hundertmark et al., 2008; Gomez et al., 2007; Diegoli et al., 2010). Typing of X chromosomal (ChrX) markers increasingly becomes an issue in kinship testing, especially some special deficiency cases can be solved this way rather than by using autosomal and Y chromosomal (ChrY) markers [22].
Paternity trio cases can most easily be solved with autosomal short tandem repeat (STR) markers alone, while test of paternity duos involving a daughter or more complex family relations could gain from X-chromosomal testing.

The main application of X-STR markers is in deficient paternity cases, where the alleged father is absent and where only his close relatives are available for testing. especially the DNA-analysis of multiple females under the hypothesis that they share the same father. For testing mother– daughter relationships, ChrX markers are similar to AS markers and do not provide any specific advantage. Testing mother–son kinship, however, is more efficiently performed using ChrX markers and in addition if a father/ daughter parentage is in question, it may be worthwhile using also ChrX markers for testing.
A positive proof of paternity is also possible with lacking maternal genotype information : this is due to the fact that sisters usually inherit partially by matching haplotypes from their mother (even if the co-inheritance of two identical maternal ChrX without recombination is possible, but rare).

Consider, for example, a case where two sisters are tested to establish whether or not they have the same father, and where DNA profiles are only available for the sisters. In such instances, autosomal DNA markers cannot exclude paternity, since two sisters can inherit different alleles despite being full siblings. The use of X-chromosome markers can, however, exclude paternity, since two sisters would share the same paternal allele if they have the same father.

DNA markers on the X-chromosome have been shown to be powerful tools for assigning pedigree members over long distances with respect to X-chromosomal tracks. However, they fail if X-chromosomal lines are interrupted by a father–son relationship. If female individuals have the same father, they always share the same paternal ChrX. An investigation of ChrX markers of two sisters or stepsisters can thus exclude paternity, even if DNA of the parents is not available. So typing of ChrX STR clusters provides a powerful tool.

In paternity cases involving close blood-relatives as putative fathers, the exclusion power of STRs is considerably reduced and ChrX STRs may be superior to AS markers. For example, if two alleged fathers are father and son, they would not share any X-chromosomal alleles identical by descent , and hence ChrX markers would be more efficient than AS markers.

Brothers, in contrast, share a given maternal ChrX allele with a probability of 0.5, which corresponds to the probability of exactly one allele shared ibd at an AS locus.

A specific request for kinship tests in which only remote relatives are available for testing can be expected to arise, particularly from the need to rejoin families in the context of the identification of wars and mass disasters victims or also of world-wide migration.

3.3 Population Study

The X-chromosome has features that make it a good source of information for population genetic studies. The X-chromosome is present in a single copy in males, which makes it possible to determine the X-chromosome haplotypes in men. Compared with autosomes, the X-chromosome has lower recombination rate, lower

mutation rate and smaller effective population size resulting in a faster genetic drift. In consequence, both linkage disequilibrium (LD) and population structure in the X chromosome are expected to be stronger than those in autosomes. Two thirds of the X-chromosome history has been spent in females. Thus, X chromosome polymorphisms mainly reflect the history of females. Due to recombination, X chromosome markers in females provide a multilocus system, while the mtDNA and Y-chromosome are linked haplotypes.

Thus, X-chromosome markers are valuable for population genetic studies.[23-26]

Furthermore, if other scientific disciplines such as evolutionary anthropology will focus their attention also at ChrX markers, they would need reliable data for ChrX markers found in different ethnics all over the world. Due to the quite different inheritance mode, the ChrX typing can never achieve the same significance in this field as ChrY marker research has obtained.

In males, the ChrX marker appears in hemizygous state. Hence, ChrX typing of marker clusters automatically provides haplotypes. Since very closely linked markers regularly exhibit a linkage disequilibrium, hence, frequencies of haplotypes cannot be calculated by multiplicating the frequencies single alleles of the haplotypes involved but they must be estimated by the analysis of population samples. If two or more STR loci are used, the count of haplotypes may extend several hundreds or even more than thousand haplotypes.

Thus a X-STR database must consist of haplotype frequencies rather than only allele frequencies. A website **(http://www.chrx-str.org/)** accessible for the forensic community was established in order to provide a reference database for ChrX STRs and ChrX STR haplotypes comprising published population data for populations from several countries.

.

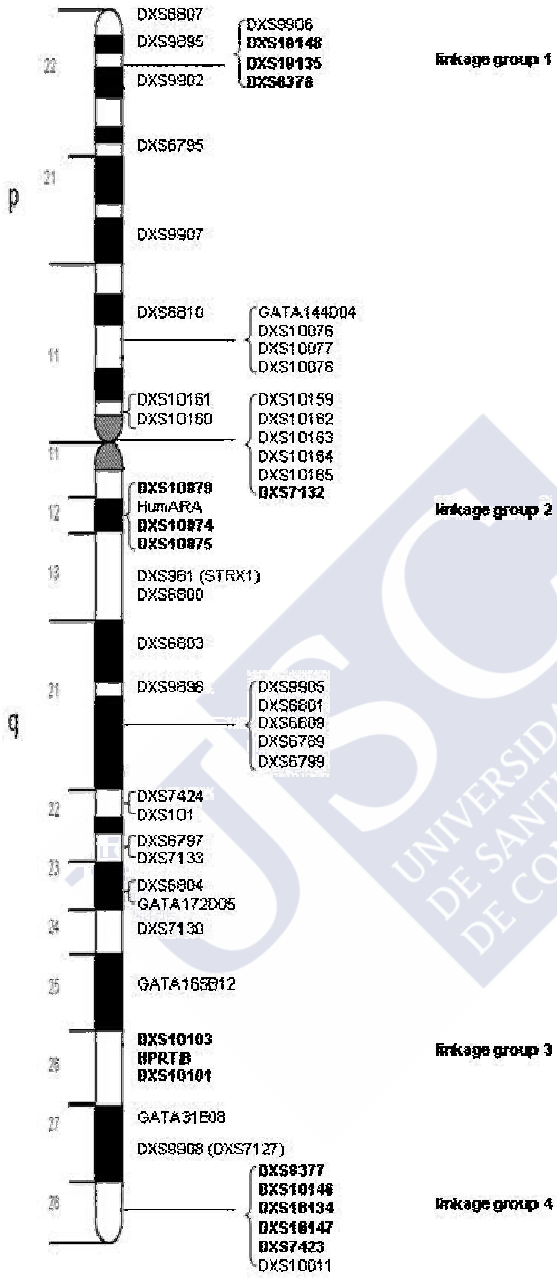


Fig.4 : *X- chromosome ideogram*

4. ALTERNATIVE STRS MARKERS : MINI STRs

4.1 Introduction

Degraded DNA samples are commonly observed in forensic investigations involving biological evidence: forensic DNA laboratory often has to deal with DNA samples that are less than ideal. The biological material serving as evidence of a crime may been left exposed to a harsh environment for days, months or even years such as in the case of the investigations of missing person. The victims of homicides are typically taken to out of the way places where they remain until their bodies are discovered. Instead of being preserved in a freezer away from caustic chemicals that can break it down, the DNA molecule may have been left in direct sunlight or in damp woods. Regardless of the situation, the DNA molecules from a crime scene come from a less than pristine environment that is normally found in molecular biology laboratories. Just as important is the fact that the retrieved biological sample may be limited in quantity. Thus accurate sample analysis is critical since a forensic scientist may only obtain enough evidence for one attempt at analysis. [27]

The versatility and the unequalled sensitivity of the DNA test has established its use in many forensic case scenario's. The implementation of the DNA technology has a great impact on how a scene or a victim of crime must be investigated. Investigators have to look for biological traces so tiny that they cannot be detected.

In cases where DNA evidence is limited, either in quantity or quality, such as highly degraded samples that are exposed to environmental insults or inhibitors, standard STR testing is often inadequate. Analysis of these compromised DNA samples often result in dropout of the larger STR loci from the sample, and only a partial DNA profile can be obtained.. The problem is further exacerbated when large multiplex PCR reactions are used due to the wide size range of PCR products generated.

Partial DNA profiles generally do not provide the power of discrimination to include or exclude a potential contributor to the sample. Recovery of information from these degraded samples is often enhanced by analyzing smaller PCR products called **Mini-STRs.** This innovative approach exploits the ability of specially designed

primers that preferentially target the larger STR loci. Reduced-size STR amplicons can be created by moving the forward and reverse PCR primers in close to the STR repeat region. In fact while standard STR primers target longer sequences that include the STR loci, mini-STR primers "zoom in" on the STR locus so that the resulting DNA product is smaller, thereby increasing the chances of successful amplification of the larger loci. This technology dramatically increases the sensitivity of DNA detection and optimizes the opportunity to obtain a DNA profile from compromised samples helping to recover information from degraded DNA samples that typically produce partial profiles and a total loss of information from larger STR amplicons.

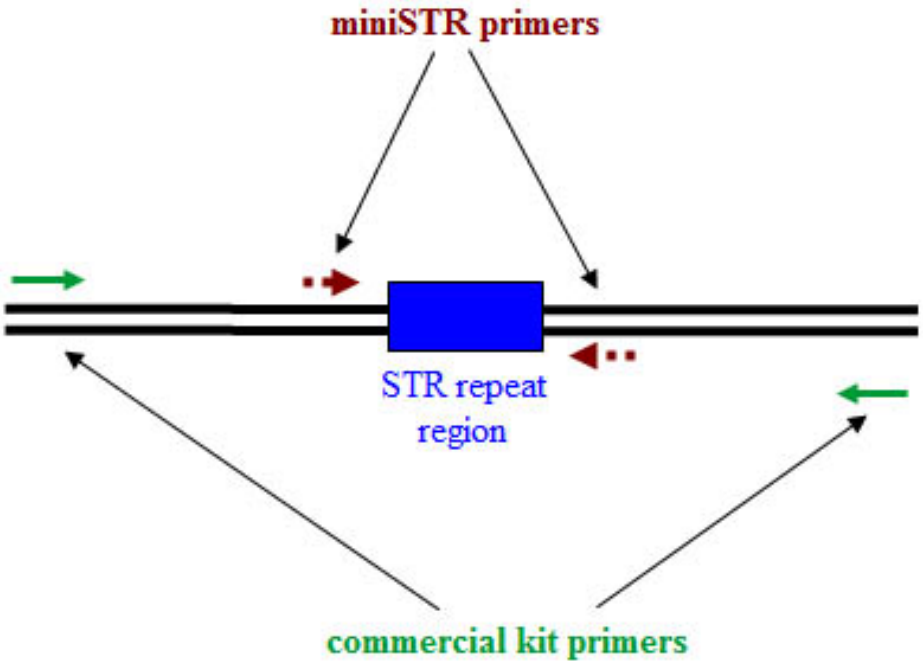


Fig.5 : *Mini Strs primers*

Because of the ability to type very degraded samples, mini-STR technology provides a useful tool for obtaining data from samples with extremely low DNA quality and

quantity, that typically produce partial profiles and a total loss of information from larger STR amplicons.[28-30]

Thus, many previously unsolvable human identity cases may be resolved with mini-STR technology. MiniSTRs had a fairly macabre introduction into the world of forensics, as they were first used extensively in a forensic lab to help identify the victims of the World Trade Center disaster. They were needed because many of the remains collected from the debris left where the Twin Towers once stood were horribly degraded from fire, heat, and sometimes bacterial degradation.

The majority of the bone samples were able to be analyzed and typed via conventional STR analysis. But there was a portion of the samples that were in such poor condition that little to no DNA results could be obtained. Miniaturized STRs primers were at that time in developing by John Butler and Bruce McCord of the National Institute of Science and Technology (NIST).

These "MiniSTRs", as they came to be known, were shown to be very successful at amplifying DNA from highly degraded samples—and became really helpful with the most degraded remains from the WTC site in late 2002. [31]

4.2 Mini-STRs applications

DNA fragments created through Mini STRs amplification are much smaller than traditional STR analysis, the types of samples that can benefit from this technology are those that are degraded or inhibited. Degraded samples typically can include bones, teeth, burnt items, items exposed to heat and humidity, etc and consist of DNA that is highly fragmented or broken down. Traditional STR typing can work with slightly degraded samples, but as the fragmentation increases even traditional STR markers (large amplicons) may yield a negative result. In addition, forensic samples often contain some substance that slows or inactivates the PCR reaction (inhibitors). Many inhibitors are known to forensic scientists.

These include: certain dyes (such as the indigo dye in denim fabrics), humic acid present in some soils, heme from blood samples, melanin from skin and hair samples, and tannins from leather. While traditional STR testing works best with about

1-2 nanograms of DNA, Mini-STRs are more sensitive.(for example Minifiler kit works between 0.25-0.5 nanograms of DNA and has been shown to yield usable results with even less than this amount of DNA). This sensitivity allows analysts to obtain results from samples with very little DNA quantity. Trace biological evidence arising from casual handling of objects ('touch DNA') is increasingly being recovered from crime scenes. Many of these 'touch DNA' samples contain low amount of DNA. Recovery of genetic profiles from LCN samples is difficult using standard STR methods and such attempts often result in total failure or recovery of partial profiles. The technology of Mini-STRs, using reduced-size STR amplicons, can help to recover information from these samples, increasing the success rate in difficult sample typing. One prime area for MiniSTR testing is unidentified remains in missing persons cases (DVI). Bones that have been located in sub-optimal conditions, such as buried, underwater, or in locations with high heat and humidity are perfect for MiniSTR analysis. Typically these samples would be analyzed in combination with the traditional STR testing. Traditional testing could allow the analyst to obtain results from the majority of the loci tested, and MiniSTR typing could provide many of the loci that may be missing from the traditional results. Cold cases can also benefit from MiniSTR testing. Samples that even a year or two ago would not be considered good candidates for DNA testing may now yield results with this extremely sensitive systems. However, one must take into account the condition and handling of the sample prior to its arrival in the lab. MiniSTR typing is so sensitive it is possible to pick up DNA from officers or others who may have handled the item years ago. MiniSTR typing opens the doors for a whole new genre of samples in the forensic DNA laboratory. In fact a collaborative study with the European DNA Profiling Group evaluated several methods of analysis to assess how effective each was for genotyping degraded DNA. STR systems (miniSTR assays and standard STR kits) and single nucleotide polymorphisms were compared and in general, miniSTR systems were observed to be the most effective in the analysis of degraded.

One major advantage of these smaller STRs, or "miniSTRs," is that database compatibility can be maintained with convicted offender samples processed using commercial STR multiplexes. In fact 3 new miniSTR loci (D10S1248, D2S441, and

D22S1045) have been recommended for adoption by the European DNA community as new core loci The addition of new loci into the ESS decreases the chance of obtaining false positive matches with cross-border DNA data exchanges – especially when there are partial (incomplete) profiles, whilst the small amplicon sizes of the new loci increases the chance of amplification in degraded sample, where DNA may be fragmented and/or in low quantity. In addition, these miniSTR markers have the potential to provide additional discrimination in complex paternity cases or missing persons cases. [32,33]

5. References

[1]John M. Butler, Advanced Topics in Forensic DNA Typing Methodology (2011) CRC Accademic Press.

[2]Bär W., Brinkmann B., Lincoln P., Mayr W., Rossi U., Budowle B., Eisenberg A., Fourney R., Gill P., Rand S. (1992), Editorial: Recommendations of the DNA Commission of the International Society for Forensic Haemogenetics relating to the use of PCR-based polymorphisms, Forensic Sci. Int. 55, 1-3

[3]Bär W., Brinkmann B., Budowle B., Carracedo A., Gill P., Lincoln P., Mayr W., Olaisen B. (1997), DNA recommendations. Further report of the DNA Commission of the ISFG regarding the use of short tandem repeat systems, Forensic Sci Int. 87(3), 179-4

[4]Bär W., Brinkmann B., Lincoln P., Mayr W., Rossi U., Budowle B., Fourney R., Gill P., Rand S. (1993), Editorial: Statement by DNA Commission of the International Society for Forensic Haemogenetics concerning the National Academy of Sciences report on DNA Technology in Forensic Science in the USA, Forensic Sci.Int.59(1),1-2

[5]Graves, J.A.M. (2006),Sex chromosome specialization and degeneration in mammals, Cell 124 (5): 901–914.

[6]de Knijff P., Kayser M., Caglia A., Corach D., Fretwell N., Gehrig C., Graziosi G., Heidorn F., Herrmann S., Herzog B., Hidding M., Honda K., Jobling M., Krawczak, M., Leim K., Meuser S., Meyer E., Oesterreich W., Pandya A., Parson W., Penacino, G., Perez-Lezaun, A., Piccinini, A., Prinz M., Schmitt, C., Schneider,P.M., Szibor R., Teifel-Greding J., Weichhold G. M., and Roewer, L. (1997) Chromosome Y microsatellites: population genetic and evolutionary aspects, Int.J.Legal Med. 110(3): 134-140

[7]Kayser M., de Knijff P., Dieltjes P., Krawczak M., Nagy M., Zerjal T., Pandya A., Tyler-SmithC.,and Roewer,L.(1997),Applications of microsatellite-based Y chromosome haplotyping, Electrophoresis. 18: 1602-1607.

[8]Butler,J.M.,Kline,M.C.,Decker,A.E.(2008),AddressingY-chromosome short tandem repeat (Y-STR) allele nomenclature.Journal of Genetic Genealogy 4(2):125-148

[9]Gill P., Brenner C., Brinkmann B., Budowle B., Carracedo A., Jobling MA., De Knijff P., Kayser M., Krawczak M., Mayr WR., Morling N., Olaisen B., Pascali V., Prinz M., Roewer L., Schneider PM., Sajantila A., Tyler Smith C. (2001),DNA Commission of the International Society of Forensic Genetics: recommendations on forensic analysis using Y-chromosome STRs., Forensic Sci Int 124(1), 5-10

[10]Gusmao L., Butler JM., Carracedo A., Gill P., Kayser M., Mayr WR., Morling N., Prinz M., Roewer L., Tyler Smith C., Schneider PM. (2006), DNA Commission of the International Society of Forensic Genetics (ISFG): an update of the recommendations on the use of Y-STRs in forensic analysis, Forensic Sci Int 157: 187-97

[11]Kayser, M., Kruger, C., Nagy, M., Geserick, G., de Knijff, P., and Roewer, L. (1998),Y-chromosomal DNA-analysis in paternity testing: experiences and recommendations, Advances in Forensic Genetics 7: 494-496.

[12]Roewer L., Krawczak M., Willuweit S.,Nagy M., Alves C., Amorim A., Anslinger K., Augustin C., Betz A., Bosch E., Caglia A., Carracedo Kayser M. et al. (2001), Online reference database of European Y-chromosomal short tandem repeat (STR) haplotypes, Forensic Sci Int 118: 106-13

[13]Willuweit S., Roewer L.(2007), Y chromosome haplotype reference database (YHRD): Update, Forensic Science International: Genetics 1(2), 83-7

[14]M. Kayser,A. Caglia, D. Corach, N. Fretwell, C. Gehrig, G. Graziosi, F. Heidorn, et al. (1997), Evaluation of Y-chromosomal STRs: A multicenter study. Int J Legal Med 110 (3): 125–33, 141–49.

[15]Y-STR Subcommittee (2004),Report on the current activities of the Scientific Working Group on DNA Analysis Methods, Forensic Science Communications *6* (3).

[16] Szibor R.(2007) X-chromosomal markers: Past, present and future, For. Sci. Int. Genetics 1 : 93–99

[17] Szibor R, Edelmann J, Hering S, Plate I, Wittig H, Roewer L, Wiegand P, Cali F, Romano V, Michael M (2003), Cell line DNA typing in forensic genetics--the necessity of reliable standards, Forensic Sci Int 138: 37-43.

[18]Szibor R, Krawczak M, Hering S, Edelmann J, Kuhlisch E, Krause D (2003), Use of X-linked markers for forensic purposes,Int J Legal Med 117: 67-74

[19]Desmarais D, Zhong Y., Chakraborty R., Perreault C, Busque L.(1998), Development of a highly polymorphic STR marker for identity testing purposes at the human androgen receptor gene (HUMARA), J. Forensic Sci. 43 (5):1046–1049.

[20]Sullivan, K. M., Mannucci, A., Kimpton, C. P., and Gill, P. (1993), A rapid and quantitative DNA sex test: fluorescence-based PCR analysis of X-Y homologous gene amelogenin, BioTechniques. 15(4): 637-641.

[21]Hering S, Augustin C, Edelmann J, Heidel M, Dressler J, Rodig H, Kuhlisch E, Szibor R, (2005), DXS10079 DXS10074 and DXS10075 are STRs located within a 280 kb region of Xq 12 and provide stable haplotypes useful for complex kinship cases, Int. J. Legal Med.116:144–149.

[22]J.Edelmann,S.Hering, M. Michael, L. Rudiger, D. Deichsel, G. Meier Sundhausen, L. Roewer, I. Plate, R. Szibor (2001),16 X-chromosome STR loci frequency data from a German population, Forensic Sci. Int. 124 215–218.

[23]T. Kishida,W.Wang, M. Fukuda, Y. Tamaki, Duplex PCR of the Y-27H39 and HPRT loci with reference to Japanese population data on the HPRT locus, (1997) Nippon Hoigaku Zasshi 51 (2):67–69.

[24] D.Becker, H. Rodig, C. Augustin, J. Edelmann, F. Gotz, S. Hering, R. Szibor, W.Brabetz (2008),Population genetic evaluation of eight X-chromosomal short tandem repeat loci using Mentype Argus X-8 PCR amplification kit, Forensic Sci. Int. Genet.2 :69–74.

[25]M.Gelabert-Besada, C. Alves, S. Ferreira, M. García-Magariños, L. Gusmão, P. Sánchez-Diz, (2012) Genetic characterization of Western Iberia using Mentype Argus X-8 kit, Forensic Sci. Int. Genet: 6:39–41.

[26]P.Wiegand, M. Kleiber (2001) Less is more—length reduction of STR amplicons using redesigned primers, Int. J. Legal Med. 114:285–287.

[27]H. Ohtaki, T. Yamamoto, T. Yoshimoto, R. Uchihi, C. Ooshima, Y.Katsumata, K. Tokunaga (2002), A powerful, novel, multiplex typing system for six short tandem repeat loci and the allele frequency distributions in two Japanese regional populations, Electrophoresis 23: 3332– 3340.

[28]P.Grubwieser,R.Muhmann,W.Parson(2003),New sensitive amplification primers for the STR locus D2S1338 for degraded casework DNA,Int. J.Leg Med. 117:185–188.

[29] D.T. Chung, J. Drabek, K.L. Opel, J.M. Butler, B.R. McCord, (2004),A study on the effects of degradation and template concentration on the amplification efficiency of the STR miniplex primer sets, J. Forensic Sci. 49:733– 740.

[30]M. Schneider, K. Bender, W. Mayr, W. Parson, B. Hoste, et al. (2004), STR analysis of artificially degraded DNA—results of a collaborative European exercise, Forensic Sci. Int. 139:123–134.

[31]Schumm, J.W., Wingrove R.S, Douglas E.K, (2004), Robust STR multiplexes for challenging casework samples, Progress in Forensic Genetics 10: 547-549.

[32]C. R. Hill, M.C. Kline, M. D. Coble, J. M. Butler, (2008),Characterization of 26 MiniSTR Loci for Improved Analysis of Degraded DNA Samples, J Forensic Sci 53:1

[33]M. D. Coble, J. M. Butler, (2005) Characterization of New MiniSTR Loci to Aid Analysis of Degraded DNA, J. Forensic. Sci. 50:43–53

Chapter III

DNA MARKERS: SINGLE NUCLEOTIDE POLYMORPHISM (SNPs)

1. Introduction

Single Nucleotide Polymorphisms (SNPs) represent the most common form of natural genetic variation in the human genome (approximately 90%) and are considered the major genetic source to phenotypic variability that differentiate individuals. Because SNPs occur frequently throughout the genome and tend to be relatively stable genetically, they serve as excellent biological markers for identification of genes in parts of the genome that may have some relation to a specific disease and even have influence on response to drug regimens.

A single-nucleotide polymorphism (SNP) is a single base change in a DNA sequence. It occurs when a single nucleotide (A, T, C, or G) in the genome is replaced by any of the other three bases and the DNA sequence differs between individuals of a species or between paired chromosomes in an individual. (An example: in the DNA sequence AGCT, a SNP occurs when the G base changes to a C, and the sequence becomes ACCT).A variation is considered to be a SNP, if it occurs in at least 1% of the population.

SNPs occur with a very high frequency, with estimates ranging from about 1 in 1000 bases to 1 in 100 to 300 bases along the 3-billion-base human genome: about 7-10 million SNPs exist in human populations that represent about 90% of all human genetic variation. has been estimated that approximately seven million SNPs, exist, with a minimum allele frequency (MAF) of 5% across the genome, and an additional approximately four million occurring with a MAF of 1%.

The rate of SNPs varies some along human chromosomes: the Y chromosome, such as the X chromosome, has less genetic variation than autosomes because the number of chromosomes (effective population size) is fewer than for autosomes.

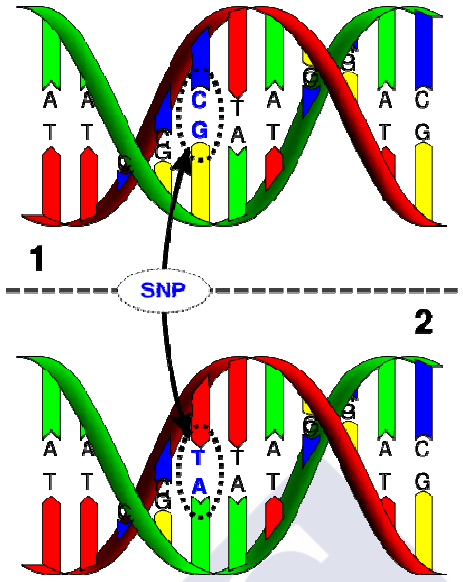


Fig.1 : *SNPs structure*

The amount of variation present in the human genome is somewhat reduced due to substantial linkage disequilibrium between closely linked SNP markers, in effect creating haplotype blocks separated from one another by recombination hotspot. SNPs initially arise via rare spontaneous mutations (approximately 10^{-8} per base pair per generation) and obtain an appreciable population frequency by genetic drift and other evolutionary forces, such as selection. The spontaneous mutation rate is so low that the chances of a further mutation at the same site on the same individual chromosome (either reversal to the original base or conversion to any of the other two) are negligible.

The most common type of SNPs has alleles A and G in a strand while in the opposite has alleles T and C. So an A/G SNP can also be described as a T/C SNP, depending upon strand orientation. It has been estimated that the distribution of the types of SNPs in human genome could be the following: 63 % A/G (and T/C), 17 % A/C (and T/G), 8 % CG, 4 % AT and 8% insertion/deletions [1]. Even if a SNP could theorically have three or four alleles, almost all common SNPs have only two alleles : the minor allele is the one showed the lowest frequency at a locus analyzed in a particular population.

There are variations between human populations, so an SNP allele that is common in one geographical or ethnic group may be much rarer in another.

A set of associated SNP alleles in a particular region of a chromosome is defined “haplotype". Almost most regions show only a few common haplotypes, with a frequency of at least 5%, which account for most of the variation from person to person in a population. Even if a chromosome region may contain many SNPs, only a few tag SNPs can provide most of the useful information about the pattern of genetic variation in that region.

SNPs are also evolutionarily stable, not changing much from a generation to another : it makes them easier to follow in population studies

SNPs may be found within coding or non-coding regions of genes, or in the intergenic regions between genes and in both nuclear and mitochondrial DNA. But only about 3 - 5 % of a person's DNA sequence codes for proteins, so most SNPs are found outside of the coding sequences.

SNPs within a coding sequence not always change the amino acid sequence of the protein, due to degeneracy of the genetic code. A SNP in which both forms lead to the same polypeptide sequence is called *synonymous* (a silent mutation) while if a different polypeptide sequence is produced, it’s termed *nonsynonymous*.

A *nonsynonymous* change may either be missense or nonsense: a missense change produces a different amino acid while a nonsense change results in a premature stop codon. All that could alter the protein, which in turn could influence a person's health. Single nucleotide polymorphisms that are not in protein-coding regions may have consequences for gene splicing, transcription factor binding, or the sequence of non-coding RNA.

Although more than 99% of human DNA sequences are the same, variations in DNA sequence can predispose people to disease or influence their response to disease, environmental factors (such as bacteria, viruses, toxins, and chemicals) drugs and therapies. [2,3]. The abundance of SNPs and the ease with which they can be measured make these genetic variations useful for biomedical research and for developing pharmaceutical products or medical diagnostics.

Single nucleotide polymorphisms within a gene region have often been studied to evaluate their effect on phenotype. Although a single base pair change can produce a phenotypic change, however a phenotype is often influenced by the presence of multiple polymorphisms and their relative positions within a given region. This means that it is essential to study the haplotype, or the combination of multiple SNPs alleles on each chromosome in order to associate genomic changes with a particular phenotype.

SNP markers are preferred over microsatellite markers for association studies because of their abundance along the human genome (SNPs with minor allele frequency > 10% occur in 1 of every 600 bp), the low mutation rate and the potential to high-throughput genotyping.

Different genotyping applications require screening of different numbers of SNPs.

The determination of a single SNP can be sufficient to screen for the presence of a Mendelian disease even if to accurately evaluate whether mutations within a class of genes contribute to a disease, hundreds to thousands of SNPs must be studied in association studies. In fact the number of SNPs required for genomewide association studies depends on the LD pattern. Recent studies have shown that the human genome can be partitioned into discrete blocks of high LD and relatively limited haplotype diversity, separated by shorter regions of low LD. One of the practical implications of this observation is that only a small fraction of all the single-nucleotide polymorphisms (SNPs) (referred as "tag SNPs") is sufficient to capture most of haplotype structure of the human genome in each block. So it can be extremely useful for association studies in which it is not necessary to genotype all SNPs since it permits significantly to reduce genotyping effort. [4-6]

Many efforts in both the public (Human Genome Project) as well as the private (The SNP Consortium) sectors have been made underway to generate high-density SNPs maps that could provide the framework for research studies designed to identify genes involved in the physiology of multigenic diseases, as well as diagnostic markers or responsible of different individual response to drug or pharmaceuticals.

2. SNPs RESEARCH PROJECT

In the past years, several research groups worked to create SNP maps of the human genome. Among these were the U.S. Human Genome Project (HGP) and a group of companies called the SNP Consortium.

The U.S. Human Genome Project was a 13-year effort coordinated by the U.S. Department of Energy (DOE) and the National Institutes of Health (NIH) with the aim to discover all the estimated 20,000-25,000 human genes and make them accessible for further biological study. [7]

The project begun in October 1990 and originally was planned to last 15 years, but rapid technological advances accelerated the completion date to 2003.

In 1998, as a part of the last five years research plans, the DOE and NIH established the following goals about Human Genome Sequence Variation:

- Develop technologies for rapid, large-scale identification and/or scoring of single nucleotide polymorphisms and other DNA sequence variants.
- Identify common variants in the coding regions of the majority of identified genes during this five-year period.
- Create a SNP map of at least 100,000 markers.
- Develop the intellectual foundations for studies of sequence variation.
- Create public resources of DNA samples and cell lines.

The initial aim was briefly reached and widely exceeded since in February 2003 were mapped 3.7 million human SNPs. All data informations are stored in a public database accessible as a common resource for scientits. The SNP Consortium (TSC) was established in april 1999 under the leading of Arthur L. Holden as a collaboration of ten large pharmaceutical companies and the U.K. Wellcome Trust philanthropy.The goal was to discover in two years 300,000 SNPs and to produce a public widely accepted, high-quality SNPs map resource. [8]

The international member companies APBiotech, AstraZeneca Group PLC, Aventis, Bayer Group AG, Bristol-Myers Squibb Co., F. Hoffmann-La Roche, Glaxo Wellcome PLC, IBM, Motorola, Novartis AG, Pfizer Inc., Searle, and SmithKline Beecham PLC

contribuited at least $30 million to the consortium while the Wellcome Trust gave around $14 million. Laboratories funded by these companies to identify SNPs are located at the Whitehead Institute, Sanger Centre, Washington University (St. Louis), and Stanford University. Data management and analysis take place at Cold Spring Harbor Laboratory.

The final results largely exceeded the initial purpose and a high-density map with 1.8 million SNPs was created. Now that the first phase of the TSC project is essentially complete, the current goal is to determine the of the allele frequency/genotype frequency of certain SNPs in the major world populations.

A public website (http://snp.cshl.org), maintained at Cold Spring Harbor Laboratory, was established to make all TSC project data available to the research community, to provide information about the project itself and also to improve existing data browsing and searching facilities.

The mapping of the human genome has made possible to develop a haplotype map in order to better define human SNP variability. The haplotypes map or ‘‘HapMap’’ (www. hapmap.org) is a powerful tool that allow researchers to find genes and genetic variations that affect health and disease. The International HapMap Project is a multi-country effort started on October 2002 as a collaboration among scientists from public and private organizations in six countries (Canada, China, Japan, Nigeria, United States, United Kingdom). The goal of the project is to compare the genetic sequences of different individuals to identify the common patterns of genetic variation in humans. This includes the chromosome regions with sets of strongly associated SNPs, the haplotypes in those regions, the SNPs that tag them, the identification of regions where associations among SNPs are weak.

All of the information generated by the Project are released into the public domain, in order to help researchers in finding genes that affect health, disease, and individual responses to therapeutic drugs and environmental factors. By October 2007 more than 3 millions SNPs were found and discovery still continues. [9,10]

Genotyping quality was assessed by using duplicate samples, since all centers genotyped a standard set of SNPs and checked some of the genotypes produced by other centers.

3. RELEVANT SNPs CLASSES

2.1 Autosomal SNPs

Highly degraded DNA presents a major challenge to the standard identification markers available for forensic analyses; though shortening the amplified fragments generated in PCR markedly improves genotyping success. The rate of DNA degradation is accelerated by the effect of environmental factors including temperature, humidity, ultraviolet radiation, pH, presence of microorganisms and the localized geochemical properties of the soil. All these factors have a greater bearing on the condition of DNA than the time since deposition or death. Chemical reactions affecting DNA stability and consequently PCR efficiency, can be categorized into three groups: hydrolysis leading to base loss, oxidation leading to base modification and single/ double strand breakage. Post mortem, a corpse is subject to the action of a range of bacterial enzymes originating from the gastro-intestinal tract and from the immediate environment. The principal catalytic activity of bacterial enzymes is to cleave DNA to generate a pool of small oligonucleotides where average fragment sizes and their range of 80–200 base pairs (bp) fall within most forensic markers' inter-primer lengths and therefore compromise PCR amplification efficiency. Short Tandem Repeats (STRs) represent the first-choice markers for forensic identification due in large part to their high discrimination power. However STR analysis of highly degraded samples is often inadequate in terms of profile completeness and this compromises the discrimination power that can be expected from genotyping of these markers alone. The need to decrease amplicon sizes to the smallest possible amplifiable fragments has led to the development of several alternative marker sets specifically aimed at analyzing highly degraded DNA.

These include: mini-STRs, and single nucleotide polymorphism (SNPs) [11].

SNPs offer ideal candidate loci for typing degraded DNA due to their simplified binary polymorphisms that allow large-scale multiplexing as well as their obvious potential for designing PCR amplicon sizes in a feasible range of 50–120 bp.

Unfortunately in the case of identification,the disadvantage, however, is that since the number of alleles per locus is limited, the information content is low. The amount of information from one STR marker is the same as from approximately four SNPs (Sobrino et al., 2005).

A SNP with high heterozygosity and essentially identical allele frequencies in all populations would be ideal because the match probability would be nearly constant irrespective of population. High heterozygosity maximizes the information at each SNP and low Fst minimizes the chance effects between populations.

Thus, it should be possible to select SNPs that are useful for human identification purposes in the majority of populations, and to supplement these with SNPs showing highly contrasting allele frequency distributions in particular populations. These latter SNPs can provide valuable information for population admixture detection, in addition to the estimation of biogeographical ancestry.

In addition autosomal single nucleotide polymorphisms (SNPs) are widely investigated as markers of biogeographical ancestry due to their low mutation rate, high abundance in the genome and wide range of allele frequencies amongst populations.

The development of autosomal SNP-based forensic assays which can infer ancestral origin from biological evidence samples has considerable potential in forensic intelligence but is relatively limited, particularly with respect to populations studied. Inferences of ancestry could be utilized to narrow, or create, a pool of suspects particularly when STR profiling has been unsuccessful and when eyewitnesses are unavailable. Such techniques could also assist in the identification of victims in mass disasters and enable more efficient use of police and forensic resources in the early stages of an investigation.

2.2 SNPs on Chromosome Y

Males have one X chromosome and one Y chromosome, that contains a gene which triggers the embryonic development as a male.

Since some years, Y chromosome analysis has became a common method for tracing human evolution through male lineages as well as application to male identification in

forensic situations. In fact the ability to separate and identify the male component from evidences containing mixtures of male and female DNA is strongly useful in many forensic situations.

In fact, for example in case of sexual assault, the use of Y chromosome specific primers can improve the chances of detecting low levels of the perpretator's DNA in a high background of the female victim's DNA without any procedure for differential DNA extraction between male and female cells. Y chromosome analysis can also benefit paternity testing when a male offspring is in question: in fact since fathers pass their Y chromosome onto their sons unchanged (except for an occasional mutation), all males in a paternal lineage will possess a common Y chromosome haplotype.

The lack of recombination along most of the Y-chromosome makes it a useful tool in difficult paternity analysis for reconstruction of male linage or application in kinship analysis, in human evolutionary studies and for assessing male migration patterns .

To assess the reliability of a database as representation of actual population haplotype frequencies, however the extent of structure among populations also needs to be considered in particular because Y chromosome haploid and paternal mode of inheritance makes it more sensitive to genetic drift than the autosomes.

Extensive studies are still performed to identify numerous single nucleotide polymorphisms (SNPs) on the Y chromosome. A variety of polymorphic genetic markers have been identified in the euchromatin portion of the Y-chromosome, including a number of STR and SNPs loci. These SNPs are single base changes or insertion/deletions, which are slowly evolving in comparison with the short tandem repeat markers, which evolve more rapidly.

The analysis of single nucleotide polymorphisms located within the male-specific region of the Y-chromosome (MSY) is widely used as a powerful tool for evolutionary studies and for measuring the variability between populations. Every man can trace his Y- chromosome back to an ancestor who lived in East Africa around 140.000 years ago. DNA has changed slightly during years: if one brother had a SNP mutation, and another didn't, the brothers go separate ways. Because each of their respective sons had these different mutations and all of their descendants, at the end two large branches of the Y-chromosome tree were created.

Roughly 80 thousand years ago men decided to move, first just within Africa, but then to every part of the globe and they took this Y-chromosome mutation, (and thus the identifying branch) with them. When men adapted to new surroundings many new mutations in the DNA strand have occurred. Series of mutations form molecular lineages and each SNP mutation may define a set of specific Y chromosomes called haplogroups. Because of the special feature of Y-DNA (no recombination) mutations remain fixed in place on both types of DNA and the historical sequence of these mutations can be inferred.

In fact due to the specific distribution of Y-haplogroups among populations, Y-SNP permit to infer the origin, evolution, and history of humans by tracing back male initiated patterns of migration from modern human populations. The non-random distribution of the Y chromosome lineages worldwide permits an accurate characterisation of haplogroups associated with specific geographic areas. [12,13].

At the present day, there are many of these large branches 'haplogroups' (called A trough to R) in different regions around the globe.

Even if the validation of the Y chromosome SNPs multiplexes described for forensic application is still in progress, however SNP typing could in a near future significantly contribute to forensic investigation by providing information on the ethnic origin of a male DNA sample and combined with STR markers, could be a powerful tool for mass disasters or terrorist attacks being able to identify people from various geographical areas involved. [14,15]

2.3 SNPs on Chromosome X

The X-chromosome is present in a single copy in males, who inherit their one X-Chr from their mother, while female individuals receive one X from the mother and the other one from the father. So, female individuals fathered by the same man share their paternal Chromosome X. Female individuals fathered by the same man share their paternal Chromosome X.

X chromosome analysis have been proven to be useful in case of deficiency paternity testing and in effective mother-son kinship and father-daughter testing.
Hence in case of deficiency paternity in which the mother is available for typing, the possible X alleles of the putative father can be determined and the paternal profile can be reconstructed. [16]

The X-chromosome has features that make it a good source of information for population genetic studies. It has a lower recombination and mutation rate than autosomes and also a small population size that results in a faster genetic drift. As consequence the linkage disequilibrium (LD) and population structure in the X chromosome are stronger than in autosomes.
X chromosome polymorphisms reflect the history of females: following to recombination, X-chromosome markers in females provide a multilocus system, while the mtDNA and Y-chromosome are linked haplotypes.
The transmission pattern of the human X chromosome reduces its population size relative to the autosomes, subjects it to disproportionate influence by female demography, and leaves X-linked mutations exposed to selection in males. As a result, the analysis of X-linked genomic variation can provide insights into the influence of demography and selection on the human genome.

X chromosomes tend to be more differentiated between human populations than autosomes with several notable exceptions. Comparisons between genetically distant populations also showed an excess of X-linked SNPs with large allele frequency differences. The relationship between male and female demographic histories is likely to be complex as evidence supporting different conclusions can be found in the same dataset. Although demography may have contributed to the excess of SNPs with large allele frequency differences observed on the X chromosome, however the selection is at least partially responsible.

X-chromosome SNPs markers can be used to complement the results obtained from STR markers since they show some advantages compared to STRs such as the low mutation rate, the high number in the human genome and the ability to be typed also in partly degraded samples: all features that makes them particularly useful in forensic caseworks, complex kinship analysis or immigration case. [17,18].

2.4 Mitochondrial SNPs

Mitochondrial genome is highly polymorphic, making it useful for human identification. The vast majority of the human genome is located within the nucleus of each cell, however also mitochondria which are placed in the cytoplasm, contain a small circular genome.

Human mt-DNA was first sequenced in 1981 in the laboratory of Frederick Sanger in Cambridge, England. The original sequence is the reference sequence to which new sequences are compared and is commonly known as the Anderson sequence or the Cambridge reference sequence.

Mt-DNA is useul to the forensic DNA community because it can be efficiently amplified from limited or severaly degraded biological material.

The likelihood of recovering mtDNA in small or degraded biological samples is greater than for nuclear DNA because mtDNA molecules are present in high copy number (hundreds to thousands) in each cell compared to the nuclear complement of two copies per cell. Therefore samples that lack sufficient nuclear DNA as shed hairs, old bones and in general scarce human remains, even if degraded by environmental insult or time, may provide enough material for typing the mtDNA locus.

Unlike nuclear DNA, which is passed from both mother and father to the offspring, mtDNA is only maternally inherited so that in situations where an individual is not available for a direct comparison with a biological sample, any maternally related individual may provide a reference sample. Moreover it has a relatively infrequent mutation rate and it remains the same through many generations. Thus, mt-DNA analysis will not differentiate women that are in the same maternal lineage or children with the same mother.[19,20]

Since considerable effort and expense are required to obtain a full HVI (positions 16024–16365) and HVII (positions 73–340) mtDNA sequence so several mtDNA screening methods have been developed that permit rapid resolution of non-matching samples.Moreover the discrimination power of an mtDNA analysis is limited because common haplotypes exist in HVI/HVII mtDNA sequences that can reduce the ability to differentiate two unrelated samples.

In all this cases it can be useful the analysis of some coding region variations in addition to the non-coding polimorphysms. The sequence analysis of the coding region require more material than the one generally present in forensic samples and for that an alternative SNP analysis approach is possible in order to analyze SNPs polymorphisms within the hypervariable region as well as in the coding region. Even though the number of markers in the current system is limited, it can easily be extended to yield a greater power of discrimination. When fully developed, microarray analysis provides a promising system for efficient sensitive SNP analysis of forensic samples in the future.

The typing of mithocondrial SNPs allows the differentiation between individuals possessing an identical HV1/HV2 sequence.[21]

Multiplex SNPs panels are in development to resolve mitotypes in some populations such as Caucasian, Hispanic, and African American.

For example a set of 11 SNPs has been selected by NIST reserachers for distinguishing individuals of the most common Caucasian HV1/HV2 mitotype.

Resolution and detection of products were achieved by electrophoresis on a capillary sequencer The development of the mtSNP 11-plex assay is ian accurate method for typing sequence variant mtSNPs on a platform common to almost all forensic laboratories. Currently are in developing additional multiplex SNP panels to resolve other common mitotypes such as Caucasian, Hispanic, and African American. [22,23].

Therefore, the forensic genetics fields have been increasingly interested in studying these polymorphisms, assembling information on genetic variation of human populations and their history and also using SNPs for individual identification purposes. Coding region SNPs can fulfil a useful role for separating common HV1/HV2 mitochondrial DNA types and assays have been developed to reliably examine mtDNA coding region SNP variation

3. Forensic Applications

Since many years forensic laboratories commonly use short tandem repeats (STRs) as the standard DNA identification method, because they have been widely

validated and some multiplexes are commercially available. Even if STRs represent the ideal approach for personal identification and paternity tests, however also SNPs could be a valuable tool for this application.[24,25]

The primary advantage of using SNPs is that a higher recovery of information from degraded DNA samples is possible since a smaller target region is needed to be analysed. In fact when working with degraded or scarce DNA samples, the analysis of autosomal single nucleotide polymorphisms (SNPs) might be more successful because only a single nucleotide needs to be measured instead of hundreds of nucleotides in length as with STRs.

Common STRs markers generally have a range in length between 150 and 450 bp, so they often fail to amplify in degraded DNA samples.

SNPs allow designing small PCR target sequences in order to obtain amplicon shorter in length than other markers and that enable efficient amplification of evidences from crime scenes or formalin-fixed paraffin-embedded tissues that contain often degraded DNA or Low Copy Number DNA (less than 100 pg DNA per sample).

Another advantage of SNPs is that they possess mutation rates approximately 100 thousand times lower than STR. Thus, since they're more stable in terms of inheritance, SNPs could aid parentage testing in complex cases or kinship analysis for example for identifying mass disaster victims.

Unfortunately several significant disadvantages exist with SNPs markers when considered as a possible replacement for currently used STR loci.

First of all SNPs are not as polymorphic as STRs, so more SNPs are required to reach equivalent powers of discrimination or random match probabilities. To obtain a similar discrimination power than 13–15 STR loci commonly in use today in forensic identification, around 40–60 well balanced SNPs are necessary.

Moreover 15 STRs can be routinely amplified simultaneously in a single multiplex amplification reaction from minimal amounts of DNA sample using commercially available kits. Multiplex PCR amplification of such a large number of SNPs has been only recently obtained and the commercialisation of robust assays with their use in routine practice will require a long time being more difficult than initially supposed.

In multiplex PCR, the number of undesired interactions between the PCR primers increases exponentially as the number of primers included in the reaction mixture increases. This interaction usually produces preferential amplification of unwanted 'primer-dimer' artefacts instead of the tested DNA templates. Moreover another problem that often occurs in a multiplex reaction are variations in PCR efficiency between the amplicons, due to sequences differences in the template.

The problem of multiplexing can be reduced by using PCR primers that are as similar as possible to one another, even if the multiplexing level that can be achieved not yet is able to reach the same capacity offered by present technologies for producing STRs multiplex. Also the cost of examining more loci appear to be higher than with traditional STRs even if it seems that for high-throughput the cost of SNPs could be lower than the cost per STR using commercial kits.

Another problem is concerning data interpretation: in fact when attempting to analyse a greater number of loci, there is an increased complexity of data to be examined. In fact more loci mean also more peaks and the possibility to observe more artefacts. With so many loci being typed, an accurate data analysis could become very difficult and necessary requires the support of validated computer expert systems.

Assays with a larger number of loci are more sensitive to the quantity and integrity of the input DNA template in particular when working with low amount DNA so many loci may be lost during analysis. Unfortunately with limited amounts of starting material there may not be opportunities to do further attempts in order to recover missing loci. Moreover failed loci on reference material samples may be different from those on the evidentiary sample leaving even less of an overlap of successfully typed loci for comparison purposes. So if a few dozen or even hundreds of loci fail to produce a result on a sample, these data are excluded from the final analysis. In fact this type of data loss when attempting to perform a direct comparison between a suspect and evidence is unacceptable under the current paradigm of traditional sample matching performed with STR typing.

In addition great experience on STRs has been accumulated during the last 10 years. For instance mutations or polymorphisms in flanking regions are increasingly

more known for STRs and can be a problem for SNPs since an extensive validation in population groups is required for an increased number of markers.

Another significant disadvantage of SNPs analysis is that the limited number of alleles per SNP locus (typically only 2) that can strongly affect the interpretation of mixture. On the contrary since STRs are highly polymorphic, a STR profile on the basis of the number of alleles observed at multiple loci can clearly suggest if there's one or more donor of a sample: this means that multiple contributors to a mixture can be identified because they have non-overlapping alleles.

Since the past years, a wide number of different SNP typing technologies have been developed based on various method, different chemistries and detection platforms. Products of the allelic discrimination reactions can be detected with more than one method, and the same detection method can analyze products obtained with different reactions or assay formats.

Before SNPs markers can be introduced in routine forensic application a decision must be taken by the forensic community on the SNP markers to be employed and the detection system to be utilized, even if since technologies and validations are still in evolution it becomes quite difficult to decide on the best options available.

Most SNPs genotyping assays can be attributed to one of the following groups based on molecular mechanism: allele specific hybridisation, primer extension, invasive cleavage, oligonucleotide ligation and several detection methods (luminescen ce, fluorescence, mass measurement, etc.) are available for analysing the products of that type of reactions.

Reactions that occur in solution are more suitable for the automation because they do not require any separation or purification step after the allele discrimination reaction. Unfortunately they have a limited multiplex capability.

On the contrary reactions taking place on a solid support such as a bead, a glass slide, a chip, etc.. have greater multiplex capability but they're less flexible because further manipulations are required before the automation.

These methods offer the capability of accurate genotyping, even if they all rely on standard PCR amplification of target sequences as the initial front-end step in generating material. Because of the nature of standard solution-based multiplexed

PCR, which often can require extensive optimisation of primers and reaction conditions the inherent amplification requirement effectively limits the extent to which these varied platforms can be modified for highly multiplexed genotyping.

The ability to convert hundreds of PCR primer-pairs into a single-tube, multiplexed reaction producing specific, robust products from a complex genomic DNA template would greatly reduce the requirements for large-scale, population-based SNP genotyping.

Table 2 : *SNPs main features in forensic applications*

Main Advantages
• Abundance in the genome • Low mutation rates • Reduced amplicon sizes (ability to analyze degraded DNA) • Simple multiplex assays • High-throughput genotyping • Potential for phenotypic trait prediction
Main Disadvantages
• Low polymorphism and discrimination power • Requirement for large numbers of individual SNPs to be analyzed • Difficulties with body fluid mixture detection and interpretation • High analysis cost • Uncertainty in reliability of multiplex-capable platforms

The SNP*for*ID Consortium (www.SNPforID.org) is a research group having as partners:

- Institute of Legal Medicine,Johannes Gutenberg University Mainz, Germany
- Institute of Legal Medicine,University of Santiago de Compostela, Spain
- Institute of Forensic Medicine, University of Copenhagen, Denmark
- Institute of Cell and Molecular Sciences, Queen Mary, University of London,UK
- Institute of Legal Medicine, Medical University Innsbruck, Austria

The consortium main objectives were :

- Selection of 50 SNPs suitable for the identification of persons of unknown ethnic origin, and determination of population genetic frequencies in major ethnic groups.
- Development of a highly efficient DNA amplification strategy for the simultaneous analysis of up to 50 independent SNPs in a single assay.
- Assessment of high-throughput DNA typing platforms for reliable and accurate multiplex SNP typing.
- Investigation of the efficiency of multiplex high-throughput SNP typing using different microarray technologies in forensic casework.

All data generated by the SNPforID consortium are available online on the SNPforID browser that is a public accessible database for searching and review SNPs allele frequencies of the studied markers from all the available populations used by SNPforID. The web tool has been designed to favourite the combination of populations in groupings, the comparison between populations individually or amongst groupings or with equivalent HapMap data. [26]

SNPs may play a useful role in several forensic applications such as [27]:

a) Individual Identity test - SNPs for individualization

b) Lineage test, family reconstruction - sets of linked SNPs useful as haplotype markers to identify missing persons through kinship analyses (Y-SNPs, mt-SNPs)

c) Ancestry study - informative SNPs for establishing high probability of a person geographical ancestry (AIMs)

d) Phenotypic identification - selected SNPs for establishing if an individual has a particular characteristic, such as skin colour, hair colour or eye colour for investigative purpose (identikit).

Tab.3 : *schematic review of SNPs main forensic applications*

	Autosomal SNPS	Y-SNPs – X-SNPs	Mt-SNPs
Identity test	X	X	
Lineage test		X	X
Ancestry study	X	X	X
Phenotypic identification	X		

a) **Identity Informative SNPs**

SNPs selected for Identity-tests have the same function as the forensic STR loci: in fact they provide genetic information to differentiate people in order to exclude/ include a suspect as a source of an evidentiary sample or to attribute/exclude the origin as a putative family member.

The best SNPs for identity testing are those that have the highest heterozygosity and low coefficient of inbreeding, because these characteristics permit to reach high levels of power of discrimination even using fewer SNPs and also a fewer reference population data is required for statistical evaluations.

Regarding SNP multiplexes, there is no commercial forensic kit available, although some work has taken place and efforts made to develop such multiplexes for use in criminal cases and for relationship testing (Borsting et al. (2009), Philips et al. 2008). A 52 SNP-multiplex that amplifies 52 DNA fragments with 52 autosomal SNP loci in only one multiplex PCR was developed by the SNPforID consortium for human identification application using as main criteria for selection of the SNP loci the high levels of heterogeneity in the three major population groups and a minimum distance of 100 kb between the SNPs and neighboring genes.

The 52 SNPs are detected in two separate single base extension (SBE) multiplex reactions with 29 and 23 SNPs, respectively, using SNaPshot kit and capillary electrophoresis. The 52 multiplex assay developed has been validated by testing its performance on a wide variety of forensic samples. 52plex amplified fragments are all less than 120 bp offering greater success than standard STRs with highly degraded DNA. [28] Results obtained were really good and demonstrate that SNP typing with

SBE, capillary electrophoresis and multicolor detection methods can be applied to forensic caseworks. [29-31].

b) **Lineage Informative SNPs**

Lineage SNPs are placed on the Y chromosome or in mitochondrial DNA genome. They show a lack of recombination and a low mutation rate, so they are informative for evolutionary studies and kinship analyses, in particular in complex cases when the evidence and the reference sample are separated by several generations.

In fact the most useful forensic application of lineage SNPs is for missing person or mass disaster identifications, even if the success of analysis in kinship test is limited by the amount of DNA in samples, the number of family members available for comparison, and the characteristics of the used genetic markers.

In fact the lineage markers, currently available have a limited power of discrimination. Coble et al. selected for lineage forensic applications 59 SNPs that have been sub-divided into 8 different multiplex panels targeting 18 specific common Caucasian HVI/HVII types. [32,33]

However other studies are in progress to select more SNPs either on Y chromosome and mt-DNA than on the autosomes that all together may serve as lineage-based markers. [34,35]

c) **Ancestry Informative SNPs**

In all cases where no suspects are available for a comparison with an evidentiary sample or not match is found against a DNA database , it may be useful, for investigative purpose, to define the genetic bio geographical ancestry of a perpetrator.

Forensic STR loci are powerful identity markers, but they are poor informative as ancestry markers because of the high degree of allele-sharing among different populations. Y chromosome and mt-DNA markers used for evolutionary purposes may give some informations also about the genetic ancestry even they're not good candidates for ancestry studies because of their uniparental inheritance (haplotypes) and limited representation of the human genome.

Ancestry informative markers (AIMs) are SNPs that reveal ancestral origin of a sample donor but not identify directly physical characteristics.
They are distributed throughout all the human genome and show different frequencies in different populations .

Tests that infer the ancestral origin of a DNA sample may have a considerable potential in the development of forensic tools that can assist crime investigation.
Since this method is based on the correlation of phenotypic expression with certain elements of population ancestry structure, thus it strongly requires the assessment of the genetic variation that correlates with specific populations and the development of specific databases to quantify AIMs.
Moreover a complex statistical classification algorithm based on maximum likelihood, is required to predict ancestral origin from the profiles obtained.
A reliable forensic test for assigning the most likely ancestry can be achieved from multiplexed assays by choosing SNPs that exhibit significative allele frequency differences between population so to characterize sequences of DNA that are more prevalent in people from one continent than another.

The investigation of a series of five unsolved serial murders in southern Louisiana between September 2001 and March 2003 was aided by the use of AIM-SNPs. Prior to their use, psychological profiling had indicated the likelihood that a Caucasian male was the culprit. However, AIM-SNP analysis revealed that the killer was likely to be of African-American ancestry. Acting upon this lead, investigators eventually arrested an African-American suspect, Derek Todd Lee and tried him for the murder of Charlotte Murray Pace. Lee was subsequently linked by DNA evidence to seven other homicides from 1998 to 2003.

d) **Phenotype Informative SNPs**

The association between genetic variation and phenotypic features has been explored in several studies. The ability to perform genetic typing of biological traces collected at the crime scene, in order to obtain information about a donor's physical characteristics, is a very attractive prospect for forensic analysis and it could potentially offer a powerful new tool for crime scene investigations.

SNPs can be taken into consideration as DNA markers for phenotypic traits (eye colour, hair, skin, etc) that enable a genetic prediction of appearance for investigative purpose to identify the perpetrator of a crime. They also may have value in anthropology studies for the reconstruction of unknown human remains. AIMs provide useful information regarding the likely appearance of a suspect connected only with biogeographic ancestry, so they can be indirect measures of the phenotype of an individual. [36]

Studies are performed to determine the genetic polymorphisms, simple and complex, responsible for these different phenotypic traits, SNPs in a number of pigmentation genes have been associated with various human hair, skin, and eye colour phenotypes.This requires an assessment of a set of SNPs that strongly affects a specific phenotype as well as development of databases to relate these variants to the specific traits. To date most work on phenotype SNPs has concentrated on pigmentation, since the genetic basis of hair, skin and eye colour is well understood from animal model studies.[37,38]

Thus, they have very limited value for describing the physical appearance of an individual and the informative value must be taken into consideration on a case-by case basis. DNA markers that describe phenotypic traits would enable a more precise genetic prediction of appearance for investigative leads to identify the perpetrator of a crime. They also may be of value in anthropology studies for the facial reconstruction of unknown human remains (i.e., the skull).

DNA evidence left by a perpretor at a crime scene or on a victim's body can be analyzed to obtain physical informations about the donor in order to construct a physical portrait of the person , giving an high improvement to the investigation.

The most obvious descriptors of an individual's appearance are colouring, height, and facial features, which are all highly heritable It should therefore be possible to determine responsible for different phenotypic traits variation. [39,40]

5. REFERENCES

[1]Miller, R.D., P. Taillon-Miller, and P.Y. Kwok. (2001), Regions of Low Single-Nucleotide Polymorphism Incidence in Human and Orangutan Xq: Deserts and Recent Coalescences,Genomics 71: 78-88.

[2]Third International Meeting on Single Nucleotide Polymorphism and Complex Genome Analysis (2000) Eur. J. Hum. Genet. 9, 316-18.

[3]Weiner MP, Hudson TJ (2002),Introduction to SNPs: Discovery of Markers for Disease. BioTechniques Suppl:4-7, 12-3

[4]Wang N, Akey JM, Zhang K, Chakraborty R, Jin L(2002), Distribution of recombination crossovers and the origin of haplotype blocks: the interplay of population history, recombination and mutation,Am J Hum Genet 71:1227–1234

[5]Daly MJ, Rioux JD, Schaffner SF, Hudson TJ, Lander ES (2001),High-resolution haplotype structure in the human genome,Nat Genet 29:229–232

[6]Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, De Felice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D (2002), The structure of haplotype blocks in the human genome,Science 296:2225–2229

[7]International_Human_Genome_Sequencing_Consortium (2001), Initial sequencing and analysis of the human genome. Nature 409: 860-921.

[8]The SNP Consortium Website: Past, Present, and Future (2003),Nucleic Acids Research 31(1), 124-27.

[9]TheInternational HapMap Consortium (2007), A second generation human haplotype - map of over 3.1 million SNPs ,Nature 449: 851-861

[10]The_International_SNP_Map_Working_Group.(2001),A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms,Nature 409: 928-933.

[11]Gill P.(2001),An assessment of the utility of single nucleotide polymorphisms (SNPs) for forensic purposes. Int J Legal Med 114(4–5):204–10.

[12]Sobrino B. and Carracedo A. (2005), SNP Typing in Forensic Genetics, Forensic DNA Typing Protocols, 297: 1064-3745

[13]Blanco-Verea A, Brion, M.. Ramos-Luis E,. Lareu, M.V Carracedo A. (2008), Forensic validation and implementation of Y-chromosome SNP multiplexes, For.Sci. Int. Genetics Suppl. Series 1 181–183

[14]Brion M., Sanchez J.J.,. Balogh K, Thacker C., Blanco-Verea A., Børsting C.,. Stradmann-Bellinghausen B,. Bogus M, Syndercombe-Court D., Schneider P.M., A. Carracedo, N. Morling (2006), Analysis of 29 Y-chromosome SNPs in a single multiplex useful to predict the geographic origin of male lineages. ICS 1288, 13– 15

[15]Bouakaze C., Keyser C., Amory S., Crubézy E. and Ludes B.(2007), First successful assay of Y-SNP typing by SNaPshot minisequencing on ancient DNA
Int J Legal Med, 121 (6):493-9

[16]Tomas C.,Sanchez J.J.,Castro J.A., Børsting C.,Morling N.,Utility of X-chromoso me SNPs in relationship testing, (2008) For. Sci. Int. Genetics Supplement Series 1: 528–530

[17] M.T. Zarrabeitia, V.Mijares and J.A.Riancho (2007),Forensic efficiency of microsatellites and single nucleotide polymorphisms on the X chromosome, Int J Legal Med, 121 (6), 433-437

[18]Tomas C.,Sanchez J.J.,Castro J.A.,Børsting C.,Morling N.(2010), Forensic usefulness of a 25 X-chromosome single-nucleotide polymorphism marker set, Transfusion 50: 2258–2265

[19]Kimberly A. S., Coble M. D., Barritt. S. M., Parsons T. J., Just R. S. (2008),The application of mtDNA SNPs to a forensic case, For. Sci. Int. Genetics Supplement Series 1 :295–297

[20]Just, R.S., Irwin, J.A., O'Callaghan, J.E., Saunier, J.L., Coble, M.D., Vallone, P.M., Butler, J.M., Barritt, S.M., and Parsons, T.J. (2004), Toward increased utility of mtDNA in forensic identifications. Forensic Sci. Int. 146S: S147-S149

[21]Kline, M.C., Vallone, P.M., Redman, J.W., Duewer, D.L., Calloway, C.D., and Butler, J.M. (2005),Mitochondrial DNA typing screens with control region and coding region SNPs, J. Forensic Sci. 50(2): 377-385

[22]Vallone, P.M., Just, R.S., Coble, M.D., Butler, J.M., and Parsons, T.J. (2004),
A multiplex allele-specific primer extension assay for forensically informative SNPs distributed throughout the mitochondrial genome, Int. J. Legal Med. 118: 147-157

[23]Coble, M.D., Just, R.S., O'Callaghan, J.E., Letmanyi, I.H., Peterson, C.T., Irwin, J.A., Parsons, T.J. (2004),Single nucleotide polymorphisms over the entire mtDNA genome that increase the power of forensic testing in Caucasians,Int. J. Legal Med. 118: 137-146.

[24]Butler J.M., Coble M.D., Vallone P.M. (2007), STRs vs. SNPs: thoughts on the future of forensic DNA testing. Forensic Sci Med Pathol. 3:200–205 201

[25]Gill, P., D.J. Werrett, B. Budowle, and R. Guerrieri **(**2004),An assessment of whether SNPs will replace STRs in national DNA database: joint considerations of the DNA working group of the European Network of Forensic Science Institutes (ENFSI) and the Scientific Working Group on DNA Analysis Methods (SWGDAM), Sci. Justice 44:51-53.

[26]Amigo J, Phillips C, Lareu M, Carracedo A.(2008),The SNPforID browser: an online tool for query and display of frequency data from the SNPforID project. Int J Legal Med 2008, 122(5):435-440

[27]Budowle B, van Daal A.(2008),Forensically relevant SNP classes, BioTechniques 44:603-610,2008 pp. 603–610

[28]Sanchez, J.J., Phillips C., Børsting C., Balogh K., Bogus M., Fondevila M., Harrison C.D, Musgrave-Brown E., Salas A., Syndercombe-Court D., Schneider P., Carracedo A., Morling N. (2006), A multiplex assay with 52 singlenucleotide polymorphisms for human identification, Electrophoresis 27:1713-1724.

[29]Kidd, K.K., A.J. Pakstis, W.C. Speed, E.L. Grigorenko, S.L. Kajuna, N.J. Karoma, S.Kungulilo, J.J. Kim, et al. (2006), Developing a SNP panel for forensic identification of individuals, Forensic Sci. Int. 164:20-32.

[30]J.J. Sanchez,C. Børsting,K. Balogh,B. Berger,M. Bogus,J.M. Butler,A.Carracedo D. Syndercombe-Court L.A. Dixon, B. Filipovi , M. Fondevila, P. Gill, C.D. Harrison, C. Hohoff, R. Huell, B. Ludes, W. Parson, T.J. Parsons,E. Petkovski,C. Phillips, H. Schmitter, P.M. Schneider, P.M. Vallone, N. Morling (2008),Forensic typing of autosomal SNPs with a 29 SNP-multiplex-Results of a collaborative EDNAP exercise, For.Sci.Int. Genet. 2:176-183

[31]Costa G.,Dario P., Lucas I. Ribeiro T., Espinheira R., Geada H.(2008),Autosomal SNPs in paternity investigation. For.Sci.Int. Genet. Suppl. Series 1 :507–509

[32]Coble, M.D., R.S. Just, J.E. O'Callaghan, I.H. Letmanyi, C.T. Peterson, J.A. Irwin, and T.J. Parsons (2004),Single nucleotide polymorphisms over the entire mtDNA genome that increase the power of forensic testing in Caucasians,Int. J. Legal Med.118:137-146.

[33] Allan F. McRae, Enda M. Byrne, Zhen Zhen Zhao, Grant W. Montgomery, and Peter M. Visscher (2008),Power and SNP tagging in whole mitochondrial genome association studies,Genome Res. 18(6): 911–917.

[34]Frudakis, T., K. Venkateswarlu, M.J. Thomas, Z.Gaskin, S. Ginjupalli, S. Gunturi, V. Ponnuswamy, S. Natarajan, and P.K. Nachimuthu. (2003),A classifier for the SNP-based inference of ancestry,J. Forensic Sci. 48:771-782.

[35]Phillips C, Salas A, Sánchez JJ, Fondevila M, Gómez-Tato A, Alvarez-Dios J, Calaza M, de Cal MC, Ballard D, Lareu MV, Carracedo A , SNPforID Consortium. (2007),Inferring ancestral origin using a single multiplex assay of ancestry informative marker SNPs. Forensic Sci Int Genet.1(3-4):273-80.

[36]Frudakis, T. N.(2007),Molecular Photofitting: Predicting Ancestry and Phenotype from DNA,Academic Press Publishers (Elsevier), Amsterdam, Netherlands. Edition - 2007-09-21

[37]Grimes, E.A., P.J. Noake, L. Dixon, and A. Urquhart (2001), Sequence polymorphism in the human melanocortin 1 receptor gene as an indicator of the red hair phenotype,Forensic Sci. Int. 122:124-129.

[38]Sulem, P., D.F. Gudbjartsson, S.N. Stacey, A. Helgason, T. Rafnar, K.P. Magnusson, A. Manolescu, A. Karason, et al. (2007),Genetic determinants of hair, eye and skin pigmentation,Nat. Genet. 39:1443-1452.

[39]Sang Hong Lee, Julius H. J. van der Werf, Ben J. Hayes,Michael E. Goddard, and Peter M. Visscher (2008,'Predicting Unobserved Phenotypes for Complex Traits from Whole-Genome SNP, Data PLoS Genet. 4(10).

[40]Frudakis, T., M. Thomas, Z. Gaskin, K. Venkateswarlu, K.S. Chandra, S. Ginjupalli, S. Gunturi, S. Natrajan, et al. (2003). Sequences associated with human iris pigmentation,Genetics 165:2071-2083.

Chapter IV

DNA PROCEDURE STANDARDIZATION

1. Introduction

Although DNA analysis in most courts is generally admissible in principle, the analysis of the evidence presented by forensics labs is the area that is brought under careful inspection. In many cases, judges have ruled that DNA evidence was not admissible because the analysis did not follow generally accepted principles of forensic analysis. As a new technology, DNA fingerprinting had to be found in each of the courts to satisfy well-established standards for the admissibility of novel scientific evidence. Also, in dozens of cases prosecutors have chosen to withdraw DNA evidence when defence attorneys have hired their own forensic experts who raised questions about the validity of the evidence.

As example the DNA results can be ruled inadmissible by the judge and therefore were never presented to the jury. This was not because of a problem with the scientific validity of the test, but the statistics of the result were complicated for example when no sample are available from the reference person.

The reliability of the results is maintained by stringent quality management program, which includes proficiency testing, validation studies and quality control procedures. Any scientific test which results in information that may lead to the loss of liberty for an individual accused of a crime needs to be performed with the most care. DNA typing is no exception. It's a multi step technical process that needs to be performed by qualified and effectively trained personnel to ensure that accurate results are obtained and interpreted correctly.

In this perspective laboratory personnel must have the education, training and experience commensurate with the examination and testimony provided.

The technical manager or leader and examiner or analyst(s) must stay abreast of developments within the field of DNA typing and also to have a minimum of three years of forensic DNA laboratory experience.

It is important to remember that in many cases DNA is not the only evidence, but some of the consequences of the DNA evidence range from charges against suspects being withdrawn, to defendants pleading guilty.

Courts have accepted DNA fingerprinting because in theory the procedure is faultless. If enough sites of genetic variation are examined, it is certainly possible to determine whether two samples come from the same source. All forensic methods for individualization—fingerprints, dental impressions, striations on bullets, hair and fiber comparisons, voice spectrograms, neutron-activation analysis, blood-grouping and serum-protein and enzyme typing, as well as DNA profiling—demand an ability to match samples with reasonable accuracy with respect to characteristics that can help to differentiate one source from another. If such evidence is to be useful in court, scientifically acceptable procedures must permit the reliable measurement and comparison of physical features. Likewise, a scientific basis must exist for concluding that properly performed comparisons can distinguish possible sources.

Courts have deemed it necessary for experts not only to demonstrate that DNA profiles usually vary from one person to another, but also to produce uncontroversial, quantitative estimates of how rare the identifying characteristics are within particular groups and subgroups. Whether many other forms of identification-evidence could survive comparable demands is doubtful.

In practice, however, DNA fingerprinting presents problems.

In DNA forensics the crime lab is constrained by whatever samples happen to be found at the scene of a crime. Samples may have been degraded and may be mixtures of samples from different individuals, as happens in a multiple murder.

Also, the forensic scientist often has only a small amount of DNA, only enough to do one test, and as a result the test cannot be repeated because the sample will have been used up. DNA forensic scientists are presented with the situation were they are given two samples related to a crime scene, about which they know nothing in advance, and are asked whether or not they are identical. They first need to determine if the profiles match, a decision which requires them to make fine judgments about whether small differences between patterns are meaningful. If the scientists decide that they match at for a sufficient number of loci, then they must assess the mathematical probability that

the match might have occurred by chance. For this purpose, the distribution of loci variability in the general population must be known by the forensic scientist.

Whether scientific evidence is admissible in criminal cases depends on whether the evidence tends to prove or disprove a fact that, under the applicable law, might matter to the outcome of the case; whether the expert presenting the evidence is qualified; whether the information is derived from scientifically acceptable procedures; and whether the potential for unfair prejudice or time-consumption substantially outweighs the probative value of the information.

We recognize that some risk of error is inevitable, as in any human endeavor, whatever efforts a laboratory takes to eliminate mistakes. Nonetheless, safeguards can be built into the system to prevent both types of errors and to identify and correct them. It is important that forensic laboratories use strict quality-control standards to minimize the risk of errors.

1.1 Quality Assurance

The maintenance of high laboratory standards rests on a foundation of sound quality control (QC) and quality assurance (QA). *Quality control* and *quality assurance* refer to related but distinct components of a laboratory's effort to deliver a quality product. *Quality control* refers to measures that are taken to ensure that the product, in this case a DNA-typing result and its interpretation, meets a specified standard of quality. *Quality assurance* refers to measures that are taken by a laboratory to monitor, verify, and document its performance. Regular proficiency testing and regular auditing of laboratory operations are both essential components of QA programs. QA thus serves as a functional check on QC in a laboratory. Demonstration that a laboratory is meeting its QC objectives provides confidence in the quality of its product. [1]

The DNA Advisory Board (DAB) was established by the Director of the FBI under the DNA Identification Act of 1994. The Objectives of the DAB were essentially to develop quality assurance standards for forensic DNA testing. The first meeting of the Board was held on May 12, 1995, under the chairmanship of Dr. Joshua Lederberg, Nobel Laureate.

The work of the DAB culminated with the promulgation of the Quality Assurance Standards for DNA Testing, which became effective nationally on October 1, 1998. Subsequent deliberations resulted in the publication of the Quality Assurance Standards for Convicted Offender Databasing Laboratories, which became effective on April 1, 1999. [2,3]

The legislation establishing the DAB provided for its dissolution on March 9, 2000 .

No extension was implemented and responsibility for maintenance of the Quality Assurance Standards was transferred to the Director of the FBI with provision for recommendations by the TWGDAM.The Technical Working Group on DNA Analysis Methods (TWGDAM) is a group composed of forensic DNA analysts from government and private laboratories around the United States and Canada. TWGDAM meets several times a year to discuss problems, report on cooperative studies, and share procedures and experiences. It has published guidelines and reports that address various aspects of forensic DNA analysis and laboratory procedures. [4,9]

Proficiency-testing and audits are key assessment mechanisms in any program for critical self-evaluation of laboratory performance. Proficiency-testing entails the testing of specimens submitted to the laboratory in the same form as evidence samples. Audits are independent reviews of laboratory operations conducted to determine whether the laboratory is performing according to a defined standard. Both forms of assessment can be conducted internally or externally, that is, by people inside or outside the laboratory. Good QA programs have a mixture of regular internal and external assessment. A benefit of open proficiency-testing conducted by external entities is that many laboratories can test the same set of samples, thus allowing interlaboratory comparison of performance and statistical evaluation of collective results. In this way lab it may identify systematic problems due to equipment, materials, the laboratory environment (such as contamination) and analyst misjudgment.

Regular audits of laboratory operations complement proficiency-testing in the monitoring of general laboratory performance. The objective of the audit is to compare a laboratory's performance with its professed quality policies and objectives. Audits cover all phases of laboratory operations related to performance and accordingly touch

on matters not covered by proficiency-testing, such as equipment-calibration schedules and case-management records. Open proficiency-testing is required under TWGDAM guidelines and is a requirement for laboratory accreditation. [10]

Federal legislation authorizing the Federal Bureau of Investigation (FBI) to establish a National DNA Index System also authorized the creation of the Federal DNA Advisory Board. The Federal DNA Advisory Board was responsible for recommending quality assurance standards, and revisions as necessary, to the FBI Director and when their statutory time period expired, they charged SWGDAM (Scientific Working Group on DNA Analysis Methods) that replaced TWGDAM with this responsibility. [11]

Accreditation for laboratories is crucial for quality DNA testing services.

The International Organization for Standardization, or ISO, is an organization that establishes standards and requirements for international businesses, including various testing laboratories. UNI EN ISO/IEC 17025 specifically standardizes laboratory practices of testing and calibration. In particular for a DNA testing lab, this means that ISO/IEC 17025 is a set of standards that has been internationally established to ensure high-quality and technically acceptable laboratory practices. Specifically, a lab that is in compliance with ISO 17025 is a lab that produces reliable reports and is overseen by qualified and competent staff, meeting all international standards.

Council Framework Decision 2009/905/JHA of 30 November 2009 on Accreditation of forensic service providers carrying out laboratory activities" requires ISO 17025 accreditation for laboratory activities, for DNA and fingerprints, when "locating and recovering traces on items, as well as developing, analyzing and interpreting forensic evidence, with a view to providing expert opinions or exchanging forensic evidence". The accreditation program requires extensive documentation of all aspects of laboratory operations (including the education, training, and experience of personnel; the specification and calibration of equipment and reagents; the validation and description of analytic methods, the definition of appropriate standards and controls, the procedures for handling samples, and the guidelines for interpreting and reporting data), proficiency testing, internal and external audits of laboratory operations, and a

plan to address deficiencies with corrective action and weigh their importance for laboratory competence.

Laboratories that seek accreditation must submit all their documentation to an accreditation review team and must undergo a week-long site inspection by that team. The site inspection includes a critical evaluation of randomly selected case files to verify that the QC standards as documented are being met. Accredited laboratories must annually certify to the accreditation organism that they continue to meet defined standards [12]

1.2 Laboratory Accreditation:

- Provides formal recognition to laboratories that demonstrate technical competency
- Maintains this recognition through periodic evaluations to ensure continued compliance with requirements;
- Provides laboratories with the opportunity to determine whether work is performed correctly and to appropriate standards;
- Identifies areas for improvement through discussion and detailed reporting;
- Monitors areas for improvement through follow-up action

When a lab achieves accreditation, it certifies that its practices are reviewed by top professionals in the forensic science community. This does not mean that accredited labs are always error free or use best practices on every case. It does mean that labs are held accountable for the quality of their operations and that mechanisms are in place to quickly identify and correct. Through accreditation programs, proficiency testing, professional certification requirements and ethics codes can be instilled in crime laboratories.

For evidence to be eligible to be uploaded into the national DNA database, ISO 17025 Standards for Forensic DNA Testing Laboratories are required.

In fact according to **ENFSI-recommendation 7 : "** Labs producing DNA-profiles for a DNA-database should, as a minimum, be ISO-17025 (and/or nationally equivalent) accredited and should participate in challenging proficiency tests". [13]

2. TROUBLESHOOTING : DNA CONTAMINATION

Current used methods are sensitive and capable of amplifying and detecting low quantities of DNA. The ability to detect low DNA quantities of DNA, increases chance of detecting contaminant DNA in both samples and controls. Contamination can happen during any step of the process and can result from a variety of substances. It is not always possible to prevent contamination; however, laboratories should ensure procedures are able to :

- Minimize the risk of contamination
- Detect contamination
- Document and implement corrective measures for incidents of contamination

The most common sources of extraneous DNA are:

1. Investigators and laboratory staff
2. Reagents and consumables
3. Sample cross contamination during analysis

1) Investigators and laboratory staff can deposit their own DNA during collection, handling, and analysis of samples. The most probable means of depositing DNA from investigators and laboratory staff onto a sample is from skin cells and saliva spray. Individuals handling evidence should avoid talking or sneezing over evidence, and gloves should be worn so that sloughing of cells onto the evidence is prevented.

2) Reagents and consumable supplies can introduce exogenous DNA into the analysis process. Negative controls and reagent blanks will assist in detecting contamination associated with reagents. Since contamination introduced by consumables (e.g., plastic ware, pipette tips) are commonly single tube events they may or may not be detected through the use of controls. Reagent contamination is generally more easily detected than that from consumable products. Negative controls and reagent blanks provide a good way to monitor reagent contamination. This is not true for contamination events from consumable products. These are usually single tube events and the level of contamination is low. This type of event is primarily detected in negative controls,

reagent blanks, and evidentiary samples with low levels of DNA. Consumables may be contaminated during the manufacturing and/or packaging process. Contamination events have shown that these sterilized products can carry DNA from individuals working in the manufacturing and/or packaging process.

The Forensic Science Service (FSS) has reported incidents of casework-related STR contamination from staff of plastic ware manufacturers. Investigations carried out by the FSS prompted the novel establishment of a vendor database consisting of DNA profiles from individuals employed by various vendors of consumable products. The database has subsequently sourced unknown profiles developed in the laboratory to the manufacturing process.

The first incident in the United States was reported after DNA profiles were uploaded into the Combined DNA Index System (CODIS) and subsequently linked multiple crimes across multiple states. After it was determined that the FSS had also observed this same profile on more than one occasion, it was understood that the profile must have originated from a consumable used in the analysis process.

3) One risk of batch analysis is the inadvertent cross contamination of DNA from one sample to another sample that was processed concurrently. Generally contamination will be from samples with higher concentrations of DNA to those with lower concentration.

There are numerous processes that laboratories can establish to minimize the risk of contamination. It is important for each laboratory to assess their specific needs both technically and administratively prior to establishing a process.

Laboratories must demonstrate that they have a facility that is designed to minimize contamination. This mainly includes restricting the movement of staff, equipment, and consumables between pre- and post-amplification areas and also:

- Staff training
- Quality control testing of reagents and consumables
- Storage and treatment of consumables
- Implementation of clean techniques

Laboratory managers should ensure that all laboratory personnel are appropriately trained in the handling and processing of evidence and offender samples. The most effective way of protecting evidence from contamination from investigators and laboratory staff is to use personal protective equipment (PPE), such as gloves, gowns, and masks. In general, universal precaution methods not only protect the investigator and analyst but also ensure that the evidence is protected from contamination by handlers. Negative controls and reagent blanks are critical quality control steps to detect contamination from reagents. Laboratories should run quality control checks on reagents prior to use in casework. These checks assist in determining if a reagent is free of contamination at that time. Negative controls can then be assessed on an ongoing basis to demonstrate that they remain contaminant free. Because many contamination events are sporadic, negative results in these controls do not necessarily mean that samples from the same batch are contaminant free. Additionally, the detection of contamination in these controls does not mean that all batch samples have been affected.

Some consumables can be treated with ultraviolet (UV) light and/or autoclaved. These preventive measures may be useful in limiting contamination events even if sometimes may not be entirely effective since they may not penetrate all surfaces of the consumable. Some laboratories have established procedures whereby a percentage of consumables from each lot number is evaluated prior to use in casework. This may be especially useful for laboratories that have observed contamination suspected to be from consumable products. While this approach will not prevent contamination, it can provide data from any profile(s) developed during these checks, which it is recommended that laboratories store their consumables in such a way as to limit exposure to the environment and consider effective pretreatment.

One problem with contamination is that an individual may be falsely linked to a crime. Reference samples are generally good quality DNA samples and result in high quantities of extracted DNA. Many laboratories process samples in a way that isolates evidentiary samples from reference samples during the screening, extraction, and PCR stages. Therefore, the possibility of contaminating an evidentiary sample

with reference DNA is avoided. Most contamination events involve small quantities of DNA and therefore will be detected at lower threshold values.

Laboratories establish reporting thresholds based on their validation studies. Because most contamination is below that threshold, it will not be reported; analysts should assess any allelic activity under the reporting threshold to determine if it could be from contamination.

As stated above, negative controls and reagent blanks can greatly assist in the detection of contamination. Positive controls and samples from known sources may also aid in the detection of contamination. This because they are single-source samples of a known type so the detection of additional alleles may indicate contamination.

It's known the most likely cause of contamination of evidence is from the staff involved in handling of samples. So it is highly desirable that the laboratory maintains a staff DNA database including everyone involved from collection to completion of analysis.

And shall also be expanded to the following:

- DNA profiles from contractors who work in the laboratory area
- DNA profiles from visitors to the laboratory
- DNA profiles from employees of subcontract vendor laboratories

The comparison with this database can ensure that no contamination from a staff member is mistakenly reported.

It is important to compare the contaminant profiles to:

- Other samples from the same batch
- Samples from other batches processed in the same time frame
- Staff profiles
- Previously detected contaminant profiles
- Other persons involved in the collection and handling of the evidence

If the profile contains too few alleles to effectively screen against the above, consideration can be given to boosting the signal strength by using one or more of the following:

- Amplifying additional extract
- Extending the injection time
- Concentrating the extract or amplicon
- Increasing the number of PCR cycles

The point at which the contamination has occurred may be determined by reworking the samples in reverse in a step-by-step manner .
Although alleles under the threshold are not reported in casework, these should be considered when performing investigations/corrective actions to assist in the determination of the source.

In conclusion there are four steps to taking corrective actions:

- Identify the problem
- Determine the root cause
- Implement preventive measures
- Document the event

Each event should be documented and included in the lab documentation:

- Description of deficiency
- Description of root cause of deficiency
- Description of the impact of deficiency on past work and remedial action taken
- Description of resolution/completion

3. REFERENCES

[1]Balazic, J. and I. Zupanic (1999),Quality control and quality assurance in DNA laboratories: Legal, civil and ethical aspects, Forensic Sci Int Suppl. no.103:S1–5.

[2]DNA Advisory Board (1998)Quality assurance standards for forensic DNA testing laboratories, *Forensic Science Communications* 2 (3).

[3]DNA Advisory Board (1999),Quality assurance standards for convicted offender DNA databasing laboratories,*Forensic Science Communications* 2 (3).

[4]TWGDAM (1989),Guidelines for a quality assurance program for DNA restriction fragment length polymorphism analysis,Crime Lab Dig 16: 40-59.

[5]TWGDAM (1991),Guidelines for a quality assurance program for DNA analysis, Crime Lab Dig 18: 44-75.

[6]TWGDAM (1993),A guide for conducting a DNA quality assurance audit,Crime Lab Dig 20: 8-18.

[7]TWGDAM (1994a),Notes from the Technical Working Group on DNA Analysis Methods,Crime Lab Dig 21: 9-13.

[8]TWGDAM (1994b),Notes from the Technical Working Group on DNA Analysis Methods,Crime Lab Dig 21: 69-74.

[9]TWGDAM (1995),Guidelines for a quality assurance program for DNA analysis, Crime Lab Dig 22: 21-50.

[10]Guidelines for a Proficiency Testing Program for DNA Restriction Fragment Length Polymorphism Analysis, Crime Laboratory Digest, 1990 Vol. 17: 59-64

[11]Scientific Working Group on DNA Analysis Methods (SWGDAM)(2001), Training guidelines, Forensic Science Communications 3 (4).

[12]Understanding DNA Evidence: A Guide for Victim Serice Providers, May 2001, Brochure, National Institute of Justice and Office for Victims of Crime

[13] DNA-Database Management Review And Recommendation ENFSI DNA Working Group - April 2010

[14]National Research Council,The Evaluation of Forensic DNA Evidence, Washington, DC: The National Academies Press, 1996

Chapter V : RESULTS

In order to treat adequately the aims of the thesis, results of the investigation work have been divided in 2 different groups, each including published (or in process) papers.

a) Validation of New STRs Multiplex

Investigation was performed in order to validate a previously developed next generation pentaplex, including the new five ESS loci, evaluating the STR data informativeness and success rate on a wide range of forensic samples and to compare its performance with the one of other commercially available kits .

1.**Development and validation of a next generation-STR pentaplex,** Forensic Sci. Int. Genet. Suppl. 2 (2009) 25-26

2**.Casework application of a standalone pentaplex assay of extended-ESS STRs**, Legal Medicine (2012), *in process*.

3. **Validation Study of AmpF*l*STR *NGM SElect*™ PCR Amplification Kit,** Journal of Forensic and Legal Medicine (2012), *in process*

Forensic Science International: Genetics Supplement Series 2 (2009) 25–26

Contents lists available at ScienceDirect

Forensic Science International: Genetics Supplement Series

journal homepage: www.elsevier.com/locate/FSIGSS

Research article

Development and validation of a next generation STR ESS-pentaplex

Christopher Phillips [a,*], Anna Barbaro [b], Luís Fernandez Formoso [a], David Ballard [c], Denise Syndercombe Court [c], Ángel Carracedo [a], Maviky Lareu [a]

[a] Forensic Genetics Unit, Institute of Legal Medicine, University of Santiago de Compostela, Santiago de Compostela, Spain
[b] Department of Forensic Genetics, SIMEF, Reggio Calabria, Italy
[c] Haematology, ICMS, Barts and The London, UK

ARTICLE INFO

Article history:
Received 26 August 2009
Accepted 27 August 2009

Keywords:
Short tandem repeat
STR
Human identification
Multiplex PCR
Degraded DNA

ABSTRACT

We constructed a simple STR pentaplex of new loci recommended as next generation markers for the European Standard Set (ESS) comprising normal-amplicon STRs: D12S391 and D1S1656, plus mini-amplicon STRs: D2S441, D10S1248 and D22S1045. Validation of the pentaplex included evaluation of its ability to amplify DNA from a variety of degraded forensic casework samples. Although the ESS-pentaplex was designed in the first instance to generate allele frequency data to supplement existing databases of established STRs, the multiplex proved to be a valuable tool for the analysis of challenging DNA when certain markers of Identifiler or MiniFiler occasionally failed.

1. Introduction

Two problems regularly confront forensic DNA analyses with the routine use of standard STRs: insufficient discrimination power and presence of highly degraded DNA where locus and allele drop-out can lead to complex interpretative problems. Success with highly degraded DNA is improved using short amplicon mini-STRs. We decided to develop a bolt-on STR pentaplex of five new loci, recommended as next generation markers for the European Standard Set (ESS) in order to generate allele frequency data ahead of the release of ESS kits. The ESS-pentaplex comprises two tried and tested STRs: D12S391 and D1S1656, that are highly informative but with conventional amplicon lengths, plus three mini-STRs: D2S441, D10S1248 and D22S1045 typed with amplicon size ranges 74–135 bp. Space exists in this multiplex amongst the fragment sizes and green/yellow dye labels to allow additional STRs to be included in future. As part of the validation of the ESS-pentaplex we assessed its ability to amplify DNA from a range of degraded casework samples including hairs, bones, nails and washed bloodstains. In routine forensic use the ESS-pentaplex provided a valuable additional approach for the analysis of challenging DNA, even when some standard STRs in commercial kits failed or were too weak.

2. Materials and methods

As commercial primer designs for the five new ESS STRs are not published we used our original primers for D1 and D12 [1,2] together with those detailed in STRbase from the original developers for D2, D10 and D22 [3]. Amplicon sizes, primer sequences and dye labels are outlined in Table 1. These show that sufficient space exists for inclusion of additional informative STRs such as SE33 or D9S1120 [4] labeled with NED or VIC. For each STR reference ladders were constructed from sequenced alleles using standard procedures as previously described [4].

The quality of results obtained from challenging forensic material was evaluated by assessing the relative performance and locus drop-out of STRs in partial profiles measured as percentage genotyping success. Detectable peaks below a prescribed minimum signal of 100 RFU were also recorded.

3. Results and discussion

The percentage genotyping success rates observed in 49 challenging casework samples for Identifiler, MiniFiler and the ESS-pentaplex are summarized in Fig. 1. Although this study examined a wide range of degraded forensic material, the three multiplexes showed a consistent pattern of relative success. The ESS-pentaplex showed an average 97.6% success (94.7% when peaks below 100 RFU were excluded); MiniFiler an average 89.8% (88.0%) and; Identifiler 81.5% (80.8%). Clearly calculating success for the small-scale pentaplex is not completely comparable to larger multiplexes, but the limited number of PCR components

* Corresponding author. Tel.: +34 981 582 327; fax: +34 981 580 336.
E-mail address: c.phillips@mac.com (C. Phillips).

1875-1768/$ – see front matter © 2009 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.fsigss.2009.08.190

Table 1
PCR primer designs, dye labels and amplicon sizes of the pentaplex STRs. Observed (obs.) allele sizes obtained from an AB 3730xl and POP7.

STR		Dye	PCR primer sequence	Obs. repeat numbers	Obs. sizes	Actual sizes
D10S1248	F	6-FAM	TTAATGAATTGAACAAATGAGTGAG	8	79	82
	R		gCAACTCTGGTTGTATTGTCTTCAT	19	123	126
D1S1656	F		GTGTTGCTCAAGGGTCAACT	8	131	135
	R	6-FAM	ctctctctctctctccttGAGAAATAGAATCACTAGGGA	19.3	181	182
D12S391	F		AACAGGATCAATGGATGCAT	12	194	197
	R	6-FAM	TGGCTTTTAGACCTGGACTG	27.2	261	259
D2S441	F	VIC	CTGTGGCTCATCTATGAAAACTT	8	74	77
	R		gAAGTGGCTGTGGTGTTATGAT	17	112	113
D22S1045	F	NED	ATTTTCCCCGATGATAGTAGTCT	9	105	106
	R		CGGCACAGTGTGAGTGATCAC	19	135	136

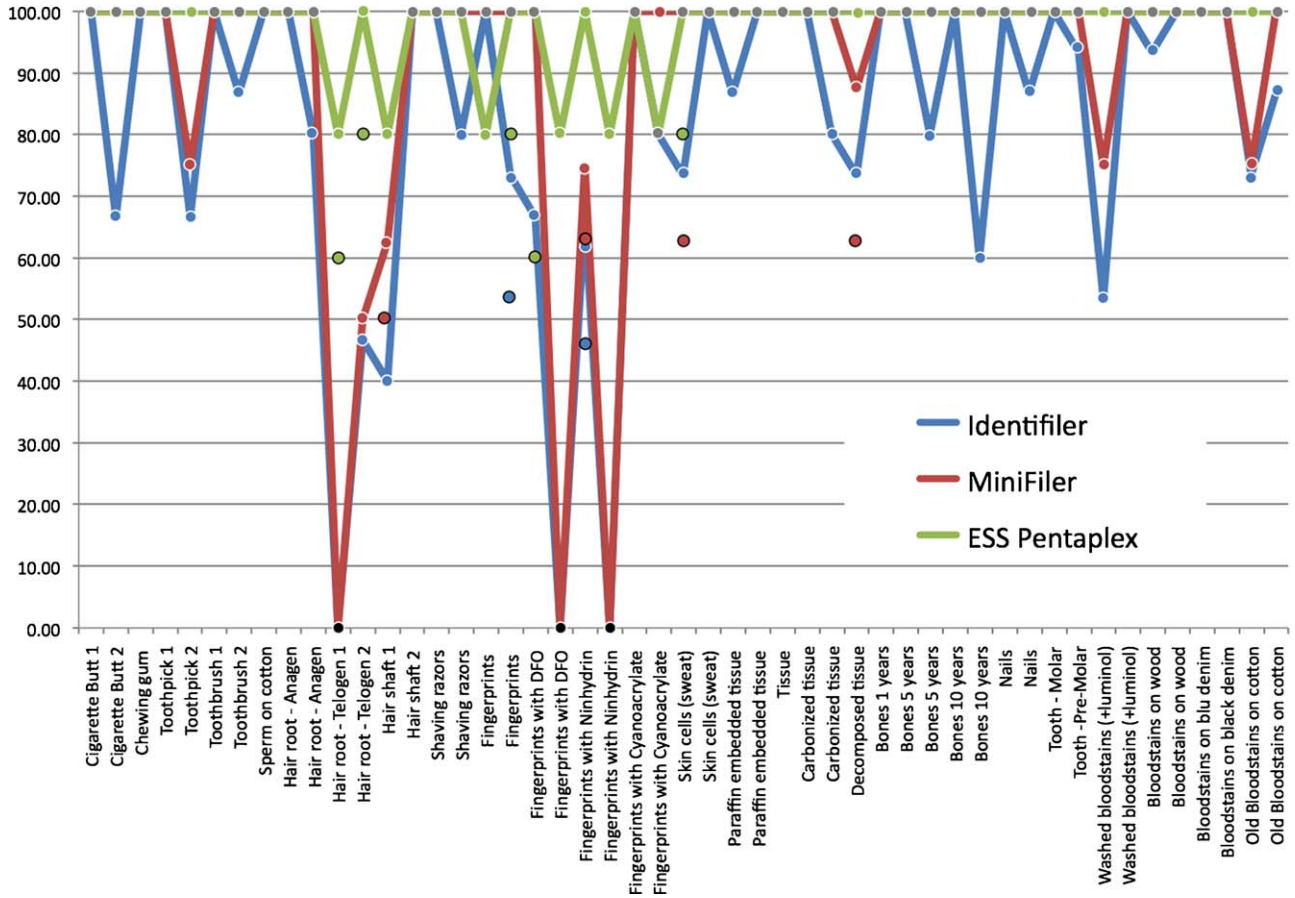


Fig. 1. Percent genotyping success for three forensic multiplexes (Identifiler, 15 STRs; MiniFiler, 8; ESS-pentaplex, 5) in 49 challenging casework samples. Multiple overlaying points shown as grey (black if 0%). Points with dark outlines show success when excluding genotype peaks below a prescribed minimum 100 RFU.

benefits performance and the pentaplex is an informative supplement to either Identifiler or MiniFiler with better overall chance of success. It is interesting to note that the only ESS-pentaplex STR showing locus drop-out (8%) was D10S1248, while both normal-amplicon STRs worked almost as well as the other two mini-STRs that showed complete success with all material genotyped.

4. Conflict of interest statement

None.

References

[1] M.V. Lareu, C. Pestoni, M. Schürenkamp, S. Rand, B. Brinkmann, Á. Carracedo, A highly variable STR at the D12S391 locus, Int. J. Legal Med. 109 (1996) 134–138.
[2] M.V. Lareu, S. Barral, A. Salas, C. Pestoni, Á. Carracedo, Sequence variation of a hypervariable short tandem repeat at the D1S1656 locus, Int. J. Legal Med. 111 (1998) 244–247.
[3] J.M. Butler, Y. Shen, B.R. McCord, The development of reduced size STR amplicons as tools for analysis of degraded DNA, J. Forensic Sci. 48 (2003) 1054–1064.
[4] C. Phillips, A. Rodriguez, A. Mosquera-Miguel, M. Fondevila, L. Porras-Hurtado, F. Rondon, A. Salas, Á. Carracedo, M.V. Lareu, D9S1120, a simple STR with a common Native American-specific allele: forensic optimization, locus characterization and allele frequency studies, Forensic Sci. Int. Genet. 3 (2008) 7–13.

Elsevier Editorial System(tm) for Legal Medicine
Manuscript Draft

Manuscript Number:

Title: Casework application of a stand-alone pentaplex assay of extended-ESS STRs

Article Type: Brief Communication

Keywords: Extended ESS-STRs; challenging DNA; mini-STRs; D12S391; D1S1656; D2S441, D10S1248; D22S1045

Corresponding Author: Mr. Christopher Phillips,

Corresponding Author's Institution: University of Santiago de Compostela

First Author: Anna Barbaro

Order of Authors: Anna Barbaro; Luis Fernandez-Formoso; Christopher Phillips; Ángel Carracedo; Maria V Lareu

Abstract: Using a stand-alone pentaplex comprising two standard-length short tandem repeats (STRs): D12S391 and D1S1656 plus three mini-STRs: D2S441, D10S1248 and D22S1045, all recently adopted to extend the European Standard Set (ESS) STRs, we have examined the genotyping performance of the new markers in 111 challenging casework samples. Although commercial kits now combine the five new STRs with existing core loci, we found the ESS-pentaplex we developed in-house performed better than both MiniFiler (comprising eight miniaturised STRs) and the NGM kit that includes the new STRs in a 15-marker multiplext. Our findings suggest at least part of the improved sensitivity of recently available ESS STRs can be attributed to the loci themselves as well as applying long-standing, robust primer designs that were first designed for the extended ESS markers by the laboratories that originally developed them. Therefore the ESS-pentaplex provides an ideal adjunct to Identifiler or MiniFiler to allow laboratories to assess the new STRs alongside existing standard loci, measure performance with challenging material and generate population frequency data ahead of a final decision on which additional STRs will extend the reconfigured CODIS core set.

Short communication

Casework application of a stand-alone pentaplex assay of extended-ESS STRs

A. Barbaro[1,2], L. Fernandez-Formoso [1,3], C. Phillips[3], Á. Carracedo[3], M.V. Lareu[3]

[1] Contributed equally to this study
[2] SIMEF, via Nicolo' da Reggio 4, 89128 Reggio Calabria, Italy.
[3] Forensic Genetics Unit, Institute of Legal Medicine, University of Santiago de Compostela, Santiago de Compostela, Galicia, Spain.

1. Introduction

Forensic DNA analysis commonly encounters two problems in the routine use of short tandem repeats (STRs): insufficient discrimination power from partial profiles and reduced success typing highly degraded DNA samples - when locus or allele dropout can lead to complex profile interpretation problems. Despite these problems, in the last six years STR typing success has been markedly improved by a shift towards shortened amplicon mini-STRs [1,2]. Furthermore the above issues have been partly addressed in Europe [3,4] by replacing five existing STRs with new markers that provide improved informativeness in two: D12S391 and D1S1656 [5,6] and miniaturised amplicon sizes in three: D2S441, D10S1248 and D22S1045 [7]. The five STRs have been approved by the European Union for expansion of the European Standard Set (ESS) markers up to 15 standard loci and a similar approach will be implemented for the CODIS loci in the near future, to potentially include four of the five STRs, with D22S1045 recommended not required [8,9].

With the initial aim to generate population frequencies efficiently, we developed a stand-alone pentaplex comprising the five new ESS STRs detailed above. We retained the conventional amplicon length primers for D12S391 and D1S1656 that we had previously developed [5,6] and had used successfully for many years in challenging casework applications. For the D2S441, D10S1248 and D22S1045 mini-STRs we used published primers [7], giving amplicon length ranges of 74 to 135 bp. Once we established the ESS-pentaplex for population frequency studies, we assessed the sensitivity of the multiplex when applied to a wide range of degraded casework samples, including hairs, bones, nails and washed bloodstains, that had previously given partial or negative profiles with established forensic STR kits. We present results of the application of the ESS-pentaplex to routine DNA profiling use and show the multiplex provided a valuable additional approach for the analysis of challenging DNA, at least as good as Applied Biosystem's (AB) MiniFiler and NGM kits , while providing better results than AB Identifiler for the most challenging material. Although several commercial kits now include the five new ESS STRs, our results suggest that the loci themselves as much as their primer designs/amplicon lengths or the kit chemistries contribute to the success that can be observed with modified combinations of STRs. Furthermore, adding smaller multiplexes up to 5-6 STRs to kit-based DNA profiling may represent a suitable strategy for maximizing the total number of genotypes when typing degraded DNA.

2. Materials and methods

2.1. Development of the pentaplex assay and allelic reference ladders

PCR primer sequences were as previously described by Lareu et al. for D1S1656 and D12S391 [5,6], and as previously described by Coble and Butler [7] for the three mini-STRs. Primers and dye-label combinations are outlined in Table 1. FAM, VIC and NED dyes were used for end-labeling the primers, therefore considerable space exists in the ESS-pentaplex amongst the green/yellow dye-labelled loci to allow additional novel STRs to be included such as D18S535 [10], and D9S1120 [11].

The PCR reactants and cycling conditions were briefly: 1 µl DNA (0.5 ng/µl) 1 µl primer mix (ratios in Table 2), 3 µl H_2O, 5 µl Qiagen Multiplex PCR Master Mix. PCR conditions were:, 95°C for 15 mins pre-denaturation, then 34 cycles of: 94°C for 30 secs, 58°C for 90 secs, 72°C for 90 secs, then a final elongation of 72°C for 10 mins.

We tested and further optimized the resulting multiplex by genotyping the 944 samples of the HGDP-CEPH diversity panel [12,13] and this population data is available at the *pop.STR* online frequency browser (http://spsmart.cesga.es/popstr.php).

Rare and intermediate alleles identified from the above population studies were further characterized by sequence analysis, as previously described [13], using the sequencing primers listed in Table 1. This enabled the ladders we had previously constructed for D1S1656 and D12S391 to be enhanced with additional alleles while allelic ladders for the mini-STRs were completely built de-novo. Allele designations for D10S1248 were determined following the changes noted by Coble and Butler [14].

2.2. Analysis of challenging casework material with ESS-pentaplex in parallel to AB Identifiler, MiniFiler and NGM STR kits

We assessed a total of 111 casework DNA extracts that were considered to constitute more challenging material than is routinely profiled, including: extracts containing strong PCR inhibitors (e.g. blue denim); low level DNA sources (sweat stains, fingerprints, washed bloodstains); hairs; fingerprints with commonly used enhancers (ninhydrin, Luminol, Cyano, DSO); and decomposed tissue and bones or teeth samples with 5 to 20 years internment. The aim was to thoroughly assess the performance of the ESS-pentaplex assay as it was available to use as a compliment to Identifiler or MiniFiler and we were able to explore its potential eight months before commercial kits had been released. This allowed an assessment of the relative sensitivity of the mini-STRs compared to the longer amplicons and allele ranges of D1S1656 and D12S391. It was also useful to assess rates of allele/locus dropout and peak balance as well as the ability to differentiate single nucleotide repeat length differences in routine profiling. This latter characteristic was particularly important for the reliable genotyping of D1S1656 and D12S391 where intermediate alleles with single nucleotide differences comprised a significant proportion of genotypes (D1S1656: 14.3-19.3 alleles = 0.254 combined frequency and 15-19 = 0.362; D12S391: 17.1/.3-20.1/.3 = 0.042 and 17-21 = 0.658). A total of sixty challenging DNA samples were compared to AB Identifiler and miniaturised amplicon MiniFiler kits, of which eleven were also analyzed with AB NGM that includes all five STRs of the ESS-pentaplex. An additional 51 challenging samples were analyzed with NGM and Identifiler in parallel with the ESS-pentaplex. Finally, we recorded individual locus dropout for the components of the the ESS-pentaplex but in cases where reference genotypes were available for comparison (~80% of cases) no allele dropout was observed. To graphically compare performance of different multiplexes we charted casework profile completeness in heatmaps ordered best to worst, left to right and with percentage loci present, in order to adjust for different multiplex sizes ranging from 5 to 15.

3. Results and discussion

The genotyping success recorded in 97 challenging casework samples for the ESS-pentaplex and three commercial STR kits is shown in Table 2 and summarized in the heat-map charts of Figs. 1A and 1C that simplify the percentage profile completeness in each case (going from hot to cold colours) as: 100%; 80-99%; 60-79%; 40-59%; 20-39%; up to 20% and no profile. By taking the average percentage success from three multiplexes the DNA samples can be ordered left to right to indicate extracted DNA in the best to the worst condition. This system of representing profile quality with a heatmap scale allows a more direct comparison of multiplexes comprising different marker numbers of 5, 8 (MiniFiler) and 15. The charts in Figs. 1A-C and underlying data in Table 2 indicate the ESS-

pentaplex to be directly comparable in performance to MiniFiler with 93% vs. 92% average profile completeness respectively with both multiplexes giving fifty full profiles from the challenging casework samples analyzed. This compares to ~81% average profile completeness observed using Identifiler on the same extracts and 28 full profiles achieved.

When the individual performance of component STRs in the ESS-pentaplex are examined no discernible pattern is evident from the minimal locus drop-out observed. No allele dropout was observed, but was controlled by comparison to reference profiles in each case. Fig. 1B lists the five STRs in ascending order of smallest amplified fragment size and only D10S1248, the second shortest STR in ESS-pentaplex, gives slightly less success than the others with six dropouts (90% success) compared to 3 or 4 in the others (93 or 95%). Therefore we did not detect a strong effect of amplicon size amongst the five ESS-pentaplex components and the so-called midi systems of D1S1656 and D12S391 perform as well as the three mini-STRs in the analysis of very challenging casework material. The same detailed breakdown of the five component STRs was not made comparing NGM vs. the ESS-pentaplex but we observed very similar performance in both multiplexes, i.e. the five ESS STRs worked equally well in both assays for any one casework sample.

The comparison of NGM and ESS-pentaplex performance indicated similar results to the parallel typing with MiniFiler but the ESS-pentaplex genotyping was noticeably more sensitive than NGM with six more complete profiles (52 vs. 44 full profiles = 97.3% success vs. 95.3%) albeit with a much smaller number of STRs in the multiplex. With this second range of casework extracts Identifiler performed better overall (average 86.7% success) but had slightly fewer complete profiles (26). Excluding cases with full profiles in all multiplexes as likely in better condition than the exhibits initially suggest, i.e. when samples giving partial profiles for some or all multiplexes are considered, the average profile completeness shows marginally more contrast between the ESS-pentaplex and MiniFiler or NGM with 88.6% vs. 85.7% for MiniFiler and 95.3 vs. 91.9% for NGM. Again, indicating a marginal but consistent improvement in performance using the ESS-pentaplex compared to both MiniFiler and NGM.

4. Discussion

The examination of a full range of one hundred and eleven different casework DNA extracts provides a reasonably comprehensive survey of the expected performance of different STR multiplexes applied to the analysis of the most challenging forensic material. It can be expected that well-balanced and commercially produced STR multiplex kits will provide the optimum performance but our results suggest it is also likely that the properties of the STRs themselves are directly responsible for the performance of a multiplex when analyzing scant or highly degraded material. Although we originally intended to build population frequency data only, ahead of the release of kits containing the five new STRs, when we used the ESS-pentaplex alongside MiniFiler and Identifiler to expand the profile completeness obtained from challenging material we found the success rate and therefore the informativeness of the STR data was enhanced considerably. Running the ESS-pentaplex in parallel to NGM containing the same STRs in a 15-plex suggests slightly better performance from the ESS-pentaplex most likely as a result of a much smaller multiplex allowing improved chances of successful amplification in a less competitive PCR. Therefore we can recommend use of small-scale STR multiplexes based on well founded primer designs as an informative and robust adjunct to MiniFiler or Identifiler, thereby improving performance and informativeness while keeping a laboratory's existing STR

protocols and data intact while transitioning to new multiplex combinations. This has relevance for those laboratories currently using CODIS STRs and seeking to assess new loci under consideration for an expanded CODIS combination in the near future [8,9].

References

[1] J.J. Mulero, C.W. Chang, R.E. Lagacé, D.Y. Wang, J.L. Bas, T.P. McMahon, L.K. Hennessy, Development and validation of the AmpFlSTR MiniFiler PCR Amplification Kit: a MiniSTR multiplex for the analysis of degraded and/or PCR inhibited DNA, J. Forensic Sci. 53 (2008) 838-852.

[2] J.M. Butler, Y. Shen, B.R. McCord, The development of reduced size STR amplicons as tools for analysis of degraded DNA, J. Forensic Sci. 48 (2003) 1054–1064.

[3] P. Gill, L. Fereday, N. Morling, P.M. Schneider, The evolution of DNA databases recommendations for new European STR loci, Forensic Sci. Int. 156 (2006) 242–244.

[4] P. Gill, L. Fereday, N. Morling, P.M. Schneider, New multiplexes for Europe. Amendments and clarification of strategic development, Forensic Sci. Int. 163 (2006) 155–157.

[5] M.V. Lareu, C. Pestoni, M. Schurenkamp, S. Rand, B. Brinkmann, Á. Carracedo, A highly variable STR at the D12S391 locus, Int. J. Legal Med. 109 (1996) 134–138.

[6] M.V. Lareu, S. Barral, A. Salas, C. Pestoni, Á. Carracedo, Sequence variation of a hypervariable short tandem repeat at the D1S1656 locus, Int. J. Legal Med. 111 (1998) 244–247.

[7] M.D. Coble, J.M. Butler, Characterization of new miniSTR loci to aid analysis of degraded DNA, J. Forensic Sci. 50 (2005) 43–53.

[8] D.R. Hares, Expanding the CODIS core loci in the United States, Forensic Sci. Int. Genet. 6 (2012): e52-4.

[9] J. Ge, A.B. Eisenberg, B. Budowle, Developing criteria and data to determine best options for expanding the core CODIS loci, Investig. Genet. 3 (2012) 1.

[10] M. V. Lareu, S. Barral, A. Salas, Á. Carracedo, Sequence variation of a variable short tandem repeat at the D18S535 locus, Int. J. Legal Med. 111 (1998) 337–339.

[11] C. Phillips, A. Rodriguez, A. Mosquera-Miguel, M. Fondevila, L. Porras-Hurtado, F. Rondon, A. Salas, Á. Carracedo, M.V. Lareu, D9S1120, a simple STR with a common Native American-specific allele: forensic optimization locus characterization and allele frequency studies, Forensic Sci. Int. Genet. 3 (2008) 7–13.

. Carracedo, M. Lareu, Development and validation of a next generation STR ESS-pentaplex, Forensic Sci. Int. Genet. Suppl. 2 (2009) 25–26.

[13] - - , L. Porras, T. Tvedebrink, J. Amigo, M. Fondevila, A. Gomez-Tato, J. Alvarez-Dios, A. Freire-Aradas, A. Gomez-Carballa, A. Mosquera-Miguel, Á. Carracedo, M.V. Lareu, Analysis of global variability in 15 established and 5 new European Standard Set (ESS) STRs using the CEPH human genome diversity panel, Forensic Sci. Int. Genet. 5 (2011) 155–169.

Figure 1. Genotyping performance summarized as a graded heatmap aligning casework DNA condition from best to worst, left to right. 1A. Profile completeness of ESS-pentaplex (5-plex), MiniFiler and Identifiler compared in 60 casework samples considered to represent challenging material due to degradation or inhibition. 1B. ESS-pentaplex component performance for the same samples as arranged in 1A. STRs arranged in ascending order of amplicon size and showing no advantage from the smaller amplicons of the three mini-STR components. 1C. Genotyping performance of ESS-pentaplex, NGM and Identifiler in forty additional challenging casework samples. A further eleven samples were typed in common with the MiniFiler as shown of 1A, with gray backgrounds to the

case sample descriptions used as a visual aid to their linkage. Underlying data to these charts is given in Table 2.

Table 1. Primer and repeat number details for the five extended ESS STRs typed in the ESS-pentaplex.

STR	Dye	Genotyping Primers	Ratio in PCR primer mix	Observed repeats	Amplicon size range (including tails)	Sequencing primers
D10S1248	6-FAM	TTAATGAATTGAACAAATGAGTGAG	0.75 µl	7	78 (79)	CTCTGTATCCCACCCCTG
		gCAACTCTGGTTGTATTGTCTTCAT		19	126 (127)	AAAGCAAACCTGAGCATTAGCC
D1S1656	6-FAM	GTGTTGCTCAAGGGTCAACT	0.75 µl	8	117 (135)	CCATATAAGTTCAAGCCTGTGTT
		ctctctctctctctccttGAGAAATAGAATCACTAGGGA		19.3	164 (182)	GAGAAATAGAATCACTAGGGA
D12S391	6-FAM	AACAGGATCAATGGATGCAT	0.75 µl	12	197	AGAGACTGTATTAGTAAGGCTTC
		TGGCTTTTAGACCTGGACTG		27.2	259	TGGCTTTTAGACCTGGACTG
D2S441	VIC	CTGTGGCTCATCTATGAAAACTT	0.72 µl	8	76 (77)	CTGAGCCCTAATGCACCCA
		gAAGTGGCTGTGGTGTTATGAT		17	112 (113)	gAAGTGGCTGTGGTGTTATGAT
D22S1045	NED	ATTTTCCCCGATGATAGTAGTCT	0.72 µl	9	104	AGCTGCTATGGGGGCTAGATT
		CGGCACAGTGTGAGTGATCAC		19	134	CGGCACAGTGTGAGTGATCAC

Table 2A. Genotyping success of challenging casework samples analyzed with ESS-pentaplex (5-plex), MiniFiler and Identifiler. The case average profile completeness values are listed at the base of the table and multiplex averages (for all profiles and for those with partial data in at least one multiplex).
Table 2B. Genotyping success for a different set of challenging casework samples (two partial profile sets of results in common with 2A) analyzed with ESS-pentaplex, NGM and Identifiler. This data underlies the heatmaps of Fig. 1A and 1C respectively.

2A

		5 plex	MiniFiler	Identifiler	average success of ESSMiniFiler
Full profiles	1-25	100	100	100	**100**
Bloodstains on wood 2	26	100	100	93.3	**100**
Tooth - pre-molar	27	100	100	93.3	**100**
Nail	28	100	100	86.7	**100**
Bloodstains on cotton 1	29	100	100	86.7	**100**
Saliva on adesive tape	30	100	100	86.7	**100**
Bone - recently deceased femur	31	100	100	80	**100**
Bone 5 years internment 1	32	100	100	80	**100**
Carbonized tissue - 2	33	100	100	80	**100**
Hair root - anagenic 2	34	100	100	80	**100**
Disposable razor 1	35	100	100	80	**100**
Tooth with caries	36	100	100	80	**100**
Degraded bone 1	37	100	100	73.3	**100**
Fingerprints 1	38	100	100	73.3	**100**
Sweat stains 1	39	100	100	73.3	**100**
Cigarette butt 2	40	100	100	66.7	**100**
Fingerprints with DFO 1	41	100	100	66.7	**100**
Bone 10 years internment 2	42	100	100	60	**100**
Degraded bone 2	43	100	100	20	**100**
Tooth	44	100	100	0	**100**
Decomposed tissue	45	100	87.5	73.3	**93.75**
Degraded bone 2	46	80	100	100	**90**
Fingerprints 2	47	80	100	100	**90**
Fingerprints with Cyano 1	48	80	100	80	**90**
Paraffin embedded tissue 2	49	100	75	86.7	**90**
Bloodstains on cotton 2	50	100	75	73.3	**87.5**
Fingerprints with Ninhydrin 1	51	100	75	66.7	**87.5**
Toothpick 2	52	100	75	66.7	**87.5**
Hair root - telogenic 2	53	100	50	46.6	**87.5**
Hair root - telogenic 1	54	80	62.5	40	**71.25**
Bone - humid environment	55	40	100	100	**70**
Bone with putrified tissue	56	0	100	73.3	**50**
Degraded bone 1	57	0	100	66.7	**50**
Fingerprints with DFO 2	58	80	0	0	**40**
Fingerprints with Ninhydrin 2	59	80	0	0	**40**
Hair shaft	60	80	0	0	**40**
Average success across all profiles		**93.3**	**91.7**	**80.6**	
Average success when partial profiles observed		**88.6**	**85.7**	**66.7**	
No. of full profiles		**50**	**50**	**28**	

2B

		5plex	NGM	Identifiler	average success of 5-plex/ESSMiniFiler
Full profiles	61-77	100	100	100	**100**
Epithelial cells on leather belt	78	100	100	95	**98.3**
Carbonized tissue 3	79	100	100	95	**98.3**
Saliva stain	80	100	100	93.3	**97.8**
Molar 5 years internment	81	100	100	93.3	**97.8**
Bone 5 years internment 2	82	100	100	93.3	**97.8**
Washed bloodstain on blue denim 1	83	100	100	93.3	**97.8**
Bloodstains with luminol	84	100	100	93.3	**97.8**
Nasal mucus	85	100	100	93.3	**97.8**
Handprint on a gun	86	100	100	93.3	**97.8**
Skull 5 years internment	87	100	100	90	**96.7**
Disposable razor 2	88	100	100	87.5	**95.8**
Nail	28	100	100	86.7	**95.8**
Clavicle 5 years internment	89	100	100	86.7	**95.8**
Vertebra 5 years internment	90	100	100	86.7	**95.8**
Jaw 5 years internment	91	100	100	86.7	**95.8**
Fingerprints with Cyano 3	92	100	93.3	93.3	**95.5**
Bloodstain on blue denim	93	100	93.3	93.3	**95.6**
Hair root anagenic 3	94	100	100	80	**93.3**
Sweath on a balaclava	95	100	93.3	86.7	**93.3**
Fingerprints with Cyano 4	96	100	100	75	**91.7**
Fingerprint on a bullet	97	93.3	87.5	90	**90.3**
Hairs from cadaver	98	100	95	75	**90.0**
Washed bloodstains (+luminol)	99	100	93.3	75	**89.4**
Hair root - telogenic 3	100	100	93.3	73.3	**88.9**
Toothpick 3	101	100	87	73	**86.6**
Bloodstain on black denim	102	93.3	87.5	75	**85.3**
Medulla 5 years	103	100	86.7	66.7	**84.4**
Fingerprint on paper	104	90	86.7	73.3	**83.3**
Decomposed tissue	105	80	86.7	73.3	**80.0**
Blood on leather shoes	106	83.3	80	75	**79.4**
Washed bloodstains (+luminol)	107	80	100	50	**76.7**
Fingerprints with Ninhydrin 3	108	100	80	30	**70.0**
Hair root - telogenic 1	54	90	65	53.3	**69.4**
Tooth 20 years internment	109	80	73.3	45	**66.1**
Fingerprints with DFO 3	110	80	80	30	**63.3**
Washed bloodstains on blue denim 2	111	60	45	20	**41.7**
Average success across all profiles		**97.3**	**95.3**	**86.7**	
Average success when partial profiles observed		**95.3**	**91.9**	**77.0**	
No. of full profiles		**52**	**44**	**26**	

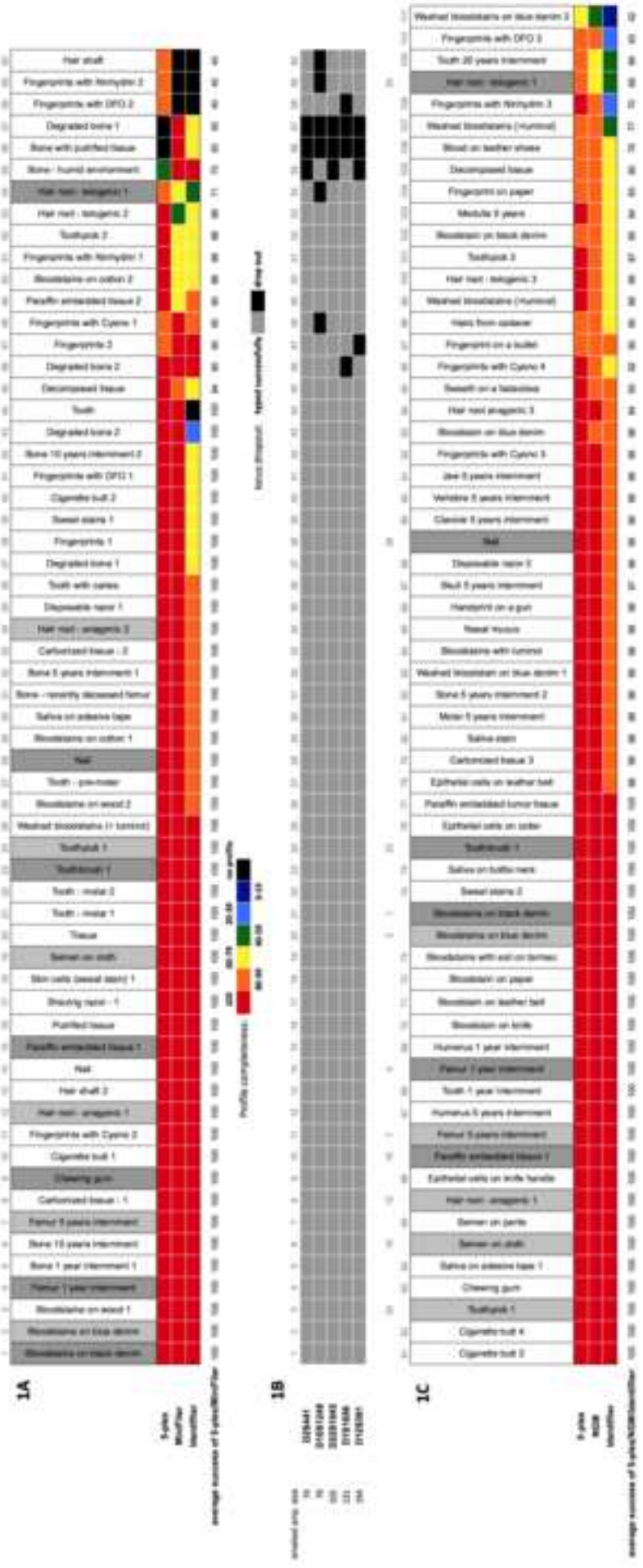


Figure 1
Click here to download high resolution image

Elsevier Editorial System(tm) for Journal of Forensic and Legal Medicine
Manuscript Draft

Manuscript Number: JCFM-E-1447

Title: Validation Study of AmpFlSTR NGM SElect™ PCR Amplification Kit

Article Type: Short Report

Keywords: AmpFLSTR NGM SElect™; ESS; internal validation; challenging samples

Corresponding Author: Dr. Anna Barbaro,

Corresponding Author's Institution:

First Author: Anna Barbaro

Order of Authors: Anna Barbaro; Patrizia Cormaci; Stefano Votano; Giacomo Falcone

Abstract: The AmpFlSTRs NGM SElect™ is a next generation kit (Applied Biosystems) containing the 5 new loci specified in the recently expanded European Standard Set (ESS) together with the remaining markers from the SGM Plus® kit, plus the highly discriminating SE33.
An internal validation study has been performed with evaluation of some critical parameters such as: species specificity, sensitivity, degradation/inhibition, mixture sample analysis.
NGM Select genotyping performance has been evaluated on a wide variety of forensic samples in comparison with the one of AmpFlSTRs Identifiler.
Our feedback confirmed the multiplex shows a robust PCR chemistry and improved performances, especially in regards to its sensitivity and greater tolerance to high levels of PCR inhibitors, allowing maximum recovery of information from difficult samples,producing useful data even when working with few DNA.

Validation Study of AmpF*l*STR NGM SElect™ PCR Amplification Kit

Anna Barbaro[a*],
Patrizia Cormaci[a] ,
Stefano Votano[a],
Giacomo Falcone[a]
[a]Dept. Forensic Genetics
Studio Indagini Mediche E Forensi (SIMEF)- Reggio Calabria - Italy

*Corresponding Author
Email address: simef_dna@tiscali.it (Anna Barbaro)
Tel. +39 0965891184 - Fax +39 0965891125

Validation Study of AmpF*l*STR NGM SElect™ PCR Amplification Kit

1.Introduction

The AmpF*L*STR NGM SElect™ is a next generation kit developed by Applied Biosystems that contains the 5 new loci specified in the recently expanded European Standard Set of Loci(ESS)together with the remaining markers from the SGM Plus® kit, plus the highly discriminating SE33. This permits exchange of data with several central European countries using SE33 routinely. [1-9] Reformulated reaction reagents and a greater number of loci concentrated in the low molecular weight region deliver greater sensitivity.
We performed an internal validation study of the NGM SElect™ Kit in order to evaluate some critical parameters as species specificity, sensitivity, degradation/inhibition study, mixture sample analysis, performance on a wide variety of forensic samples, according to SWGDAM reccomandations [10-12]

2. Materials and Methods

DNA samples were extracted by Prepfiler/BTA™ system and quantified using the Quantifiler® Human DNA Quantification kit. PCR amplification was performed in the GeneAmp® PCR Systems 2720 (Applied Biosystems) according NGM SElect™ PCR Amplification kit protocol.
PCR products were separated and detected on the AB 3130 Genetic Analyzer using recommended conditions and data analysis performed by GeneMapper® *IDX* v1.0 software (Applied Biosystems).

2.1 *Sensitivity Study*

DNA quantity affects typing results: too much DNA can result in off scale data and incomplete A nucleotide addition while extremely low quantity can produce unbalanced amplification .
Serial two-fold dilutions of 007 human control DNA were made to give final concentrations from 0.5 ng to 0.01 ng per reaction. DNA dilutions were tested in replicates and assessed for the number of alleles detected, intra-colour balance and heterozygote balance. Full profiles were obtained reproducibly with 0.016 ng of input DNA.

2.2 Inhibition Study

Inhibitors are often co-extracted and co-purified with the DNA and subsequently interferes with PCR by inhibiting polymerase activity.
Two series of test samples were formulated, containing 1ng of 007 DNA Control together with increasing concentrations of haematin as PCR inhibitor (10,50,100,150,200,250 uM).
Results were reliable and full profiles were obtained till to the highest concentrations of inhibitor tested.

Fig.1 Results from Sensitivity Study

Fig.2 Results from Inhibition Study

2.3 Degradation Study

As the average size of degraded DNA approaches the size of the target sequence, the amount of PCR product generated is reduced because of the reduced number of intact templates in the size range necessary for amplification

Control DNA 007 was treated with increasing concentration of DNase I (2,4,6,8U) to simulate DNA degradation. The longer loci gradually disappear as the amount of DNase I increases but the 3 new miniSTR (D10S1248, D22S1045 and D2S441) amplify successfully even at 6U DNase.

Fig.3 Results from Degradation Study

2.4 Species Specificity Study

Nonhuman DNA may be present in forensic casework samples. 27 species (Gorilla, Chimpanzee, Amadriade, Macaque,Fox, Gazelle, Puma, Ox, Sheep, Horse, Goat, Horse, Rabbit, Jaguar, Turkey, Dog, Raccoon, Chicken, Cat, Pig, Rape, Fish, Ram, Hare, Hippo, Panther, Snake) were tested. Chimpanzee and Gorilla DNA samples produced partial profiles, while Macaque DNA produced a strong Amelogenin-X peak and two small out-of-marker-range peaks in PET. Among non-primates, only Horse DNA produced a 96-bp fragment near the Amelogenin locus in the VIC® dye. The other animals did not yield detectable products. Results are outlined in Table 1.

Table 1 Results from Species Specificity Study

2.5 Mixture studies

Forensic casework samples may contain DNA from more than one individual. Therefore, it is essential to ensure that the DNA typing system is able to detect DNA mixtures.

Mixtures of two known DNA samples (from saliva) were examined at various ratios (1:1, 1:3, 1:7, 1:10 1:15). The total amount of genomic input DNA mixed at each ratio was 1 ng.

Detection of full profiles for the minor contributor was possible till to ratio 1:10 , while 1:15 ratios resulted in partial profiles for the minor component.

Fig4 Results from Mixture studies

2.6 Casework samples Study

The ability to obtain results from DNA recovered from biological samples deposited on various substrates and subjected to various environmental and chemical insults has been documented analyzing a wide variety of casework samples (blood, saliva sperm stains, washed bloodstains, cadaveric tissues, bones, teeth, prints, sweat). DNA was than amplified in duplicate using AmpF*l*STR Identifiler™ (28cycles) and NGM Select™ (29cycles). The quality of STRs profiles obtained has been evaluated considering peaks balance, preferential amplification, allelic drop-out,etc Genotyping performance of NGM Select™ and Identifiler™ has been compared in 20 casework challenging samples and results are summarized in a heatmap.(Fig.4).

NGM SElect genotyping on challenging samples was more sensitive than Identifiler with 7 more complete profiles (81,76 % success vs. 42,5%)

Samples with low DNA (<100pg) produced no profiles or very little genotyping information with Identifiler™ kit, while they gave successful amplification for some loci by NGM™ . Therefore, even this partial NGM™ kit profile were informative because include the 5 ESS new loci.

Different kind of samples at almost the same DNA concentration showed different typing success.

This means the nature of the evidence and its storing condition (i.e. environmental factors) have a big impact on final results.

Fig.5 Heatmap showing profile completeness ordered, left to right, best to worst

The individual performance of each STR in both kits are examined and the average rate of success for each locus is reported in Fig.6: CSF1PO, according to its size, showed the lowest success (20%), while the 2 mini D22S1045 and D2s441 were the most successful loci (97,5%).

Fig.6 Average rate of success for each locus

3.Conclusions

Results of our validation study demonstrate that NGM SElect™ kit is a reliable multiplex well suited for typing a wide variety of forensic samples. It shows improved performances, especially in regards to its sensitivity and greater tolerance to high levels of PCR inhibitors, allowing maximum recovery of information from difficult samples,producing useful data even when working with very few DNA.
STRs profiles by NGM™ were generally better balanced than Identifiler™ one showing clear baseline, less noise and PCR artefacts. This confirms NGM™ multiplex shows a robust PCR chemistry an the improved performance requested by the forensic community for challenging casework samples as well as paternity testing . [13]

4. Ethical standards

The study described in the present paper have been carried out using samples taken from people where informed consent had been previously obtained for research studies in accordance with Italian Law D.Lgs. 196/2003 and to approved SIMEF UNI EN ISO 17025 procedure.

5. Conflict of interest

None

6. References

[1] Lareu M.V., Pestoni C, Schürenkamp M., Rand S., Brinkmann B., Carracedo A., A highly variable STR at the D12S391 locus, Int J Legal Med. 109(3) (1996) 134-138.
[2] Lareu M.V., Barral S.,. Salas A, Pestoni C., Carracedo A., Sequence variation of a hypervariable short tandem repeat at the D1S1656 locus. Int J Legal Med. 111(5) (1998) 244-247.
[3] Wenda S., Dauber E. M.,. Schwartz M, Jungbauer C.,. Weirich V, Wegener R. and. Mayr W. R, ACTBP2 (alias ACTBP8) is localized on chromosome 6 (band 6q14), Forensic Sci. Int. Volume 148, Issues 2-3, 10 March 2005, Pages 207-209
[4] Coble M.D., Butler J.M., Characterization of new miniSTR loci to aid analysis of degraded DNA, J. Forensic Sci. 50 (2005) 43–53.
[5] Olaisen B., Bär W., Brinkmann B., Budowle B.,. Carracedo A, Gill P, Lincoln P., Mayr WR, DNA recommendations 1997 of the International Society for Forensic Genetics.Vox Sang. 1998;74(1):61-3.
[6] Gill P., Fereday L., Morling N., Schneider P.M., The evolution of DNA databases recommendations for new European STR loci, Forensic Sci. Int. 156 (2006) 242–244.
[7] Gill P., Fereday L., Morling N., Schneider P.M, New multiplexes for Europe. Amendments and clarification of strategic development, Forensic Sci. Int. 163 (2006) 155–157.
[8] Butler J. M., Coble M.D.,Regarding nomenclature for new miniSTR locus D10S1248, J. Forensic Sci. 52 (2007) 494.

[9] Butler J.M and Coble M.D, Authors' Response to Letter to Editor regarding nomenclature for new miniSTR locus D10S1248*]*. J. Forensic Sci. 52 (2007) 494
[10]Scientific Working Group on DNA Analysis Methods (SWGDAM),Revised Validation Guidelines, Forensic Sci. Communications (2004) 6(3);
[11] Sparkes R., Kimpton C.,. Watson S, Oldroyd N., Clayton T., Barnett L. ,Arnold J., Thompson C., Hale R., Chapman J., Urquhart A. and Gill P., The validation of a 7-locus multiplex STR test for use in forensic casework. (I). Mixtures, ageing, degradation and species studies. Int. J. Legal Med.109 (1996) 186–194.
[12]Sparkes R., Kimpton C., Gilbard S., Carne P., Andersen J., Oldroyd N., Thomas D.,Urquhart A., and Gill P., The validation of a 7-locus multiplex STR test for use in forensic casework. (II), Artifacts, casework studies and success rates. Int. J. Legal Med.109 (1996)195–204
[13] Sprecher C.J, McLaren R.S., Rabbach D., Krenke B., Ensenberger M.G., Fulmer P.M. , Downey L., McCombs E., Storts D.G., PowerPlex1 ESX and ESI Systems: A suite of new STR systems designed to meet the changing needs of the DNA-typing community, Forensic Sci. Int. Genet. Supplement Ser 2 (2009) 2–4

Table 1: *Results from Species Specificity Study*

Animal DNATested	**PCR products: size reffered to the closer peak in the human range**
Amadriade	**178 bp (Fam)**= allele 17 locus vWA **246 bp (Fam)** = allele 19 locus D16 **101 bp (Joe)** = allele X locus Amelogenin **219 bp (Joe)=** allele 32.2 locus D21
Gorilla, chimpanzee	**165 bp (Fam)** = allele 15 locus vWA **101 bp (Fam)** = allele 14 locus D10 **156 bp (Pet)** = allele 17 locus D3 **101 bp (Joe)** = allele X locus Amelogenin **149 bp (Joe)** = allele 14 locus D8 **204 bp (Ned)** = allele 9.3 locus TH01
Macaque	**97 bp (Fam)** = out range locus D10 **101 bp (Joe)** = allele X locus Amelogenin **170 bp (Joe)** = out range locus D3
Horse	**96 bp (Joe)** = out range locus Amelogenin
Fox, Gazelle, Puma, Ox, Sheep, Horse, Goat, Rabbit, Jaguar, Turkey, Dog, Raccoon, Chicken, Cat, Pig, Rape, Fish, Ram, Hare. Hippo, Panther, Snake.	**no PCR products**

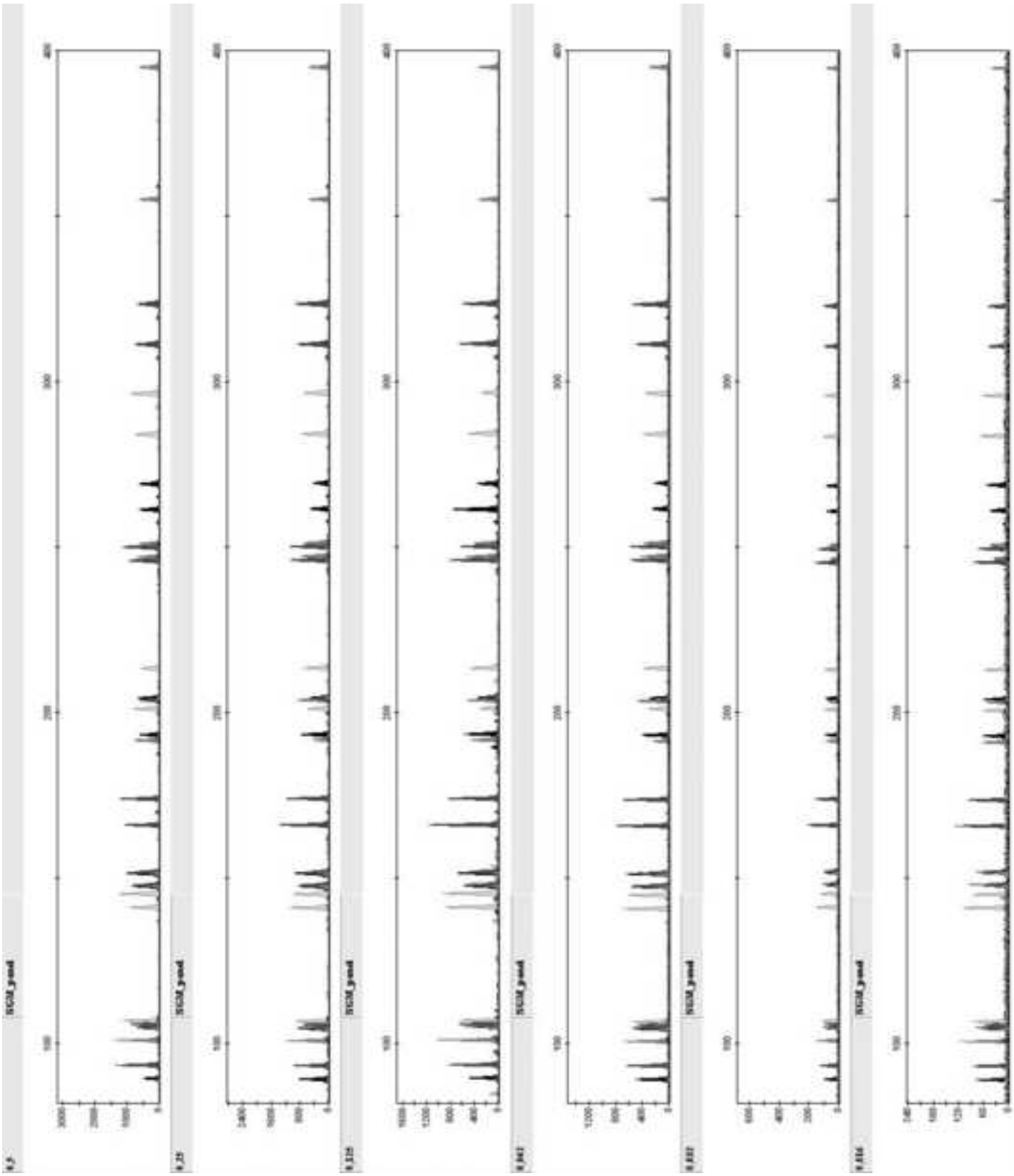


Figure1
Click here to download high resolution image

Figure2

Click here to download high resolution image

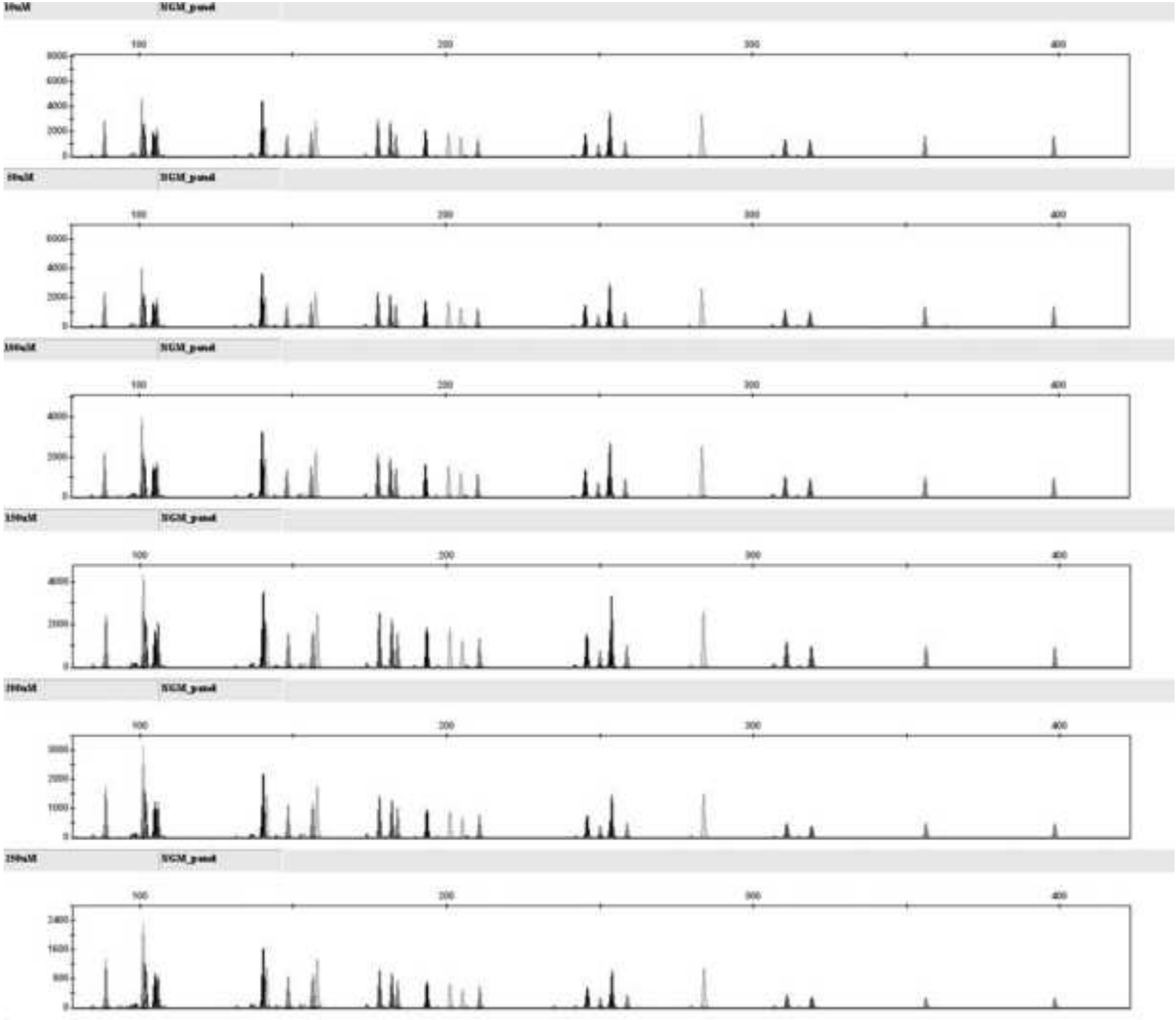


Figure3

Click here to download high resolution image

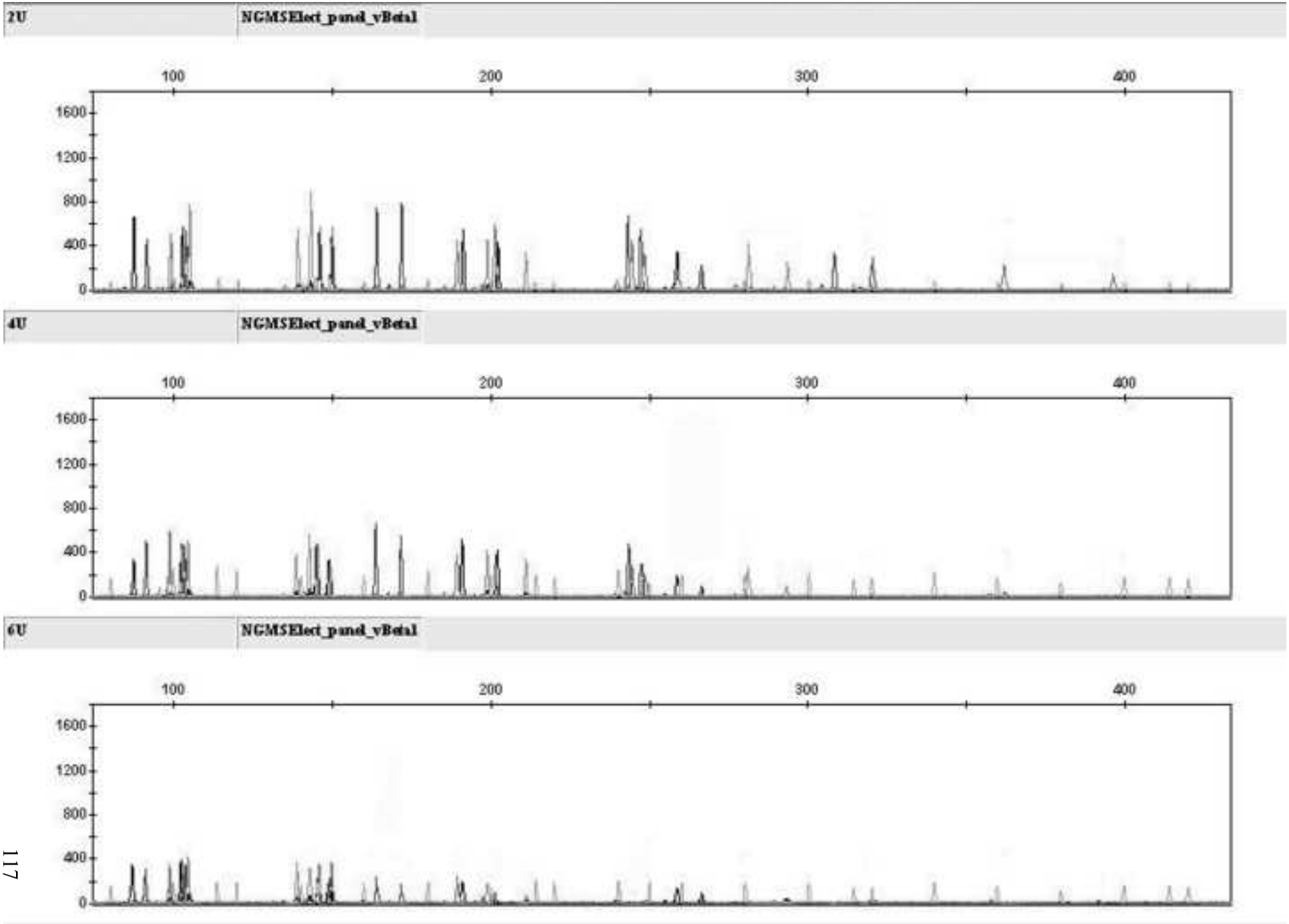


Figure4

Click here to download high resolution image

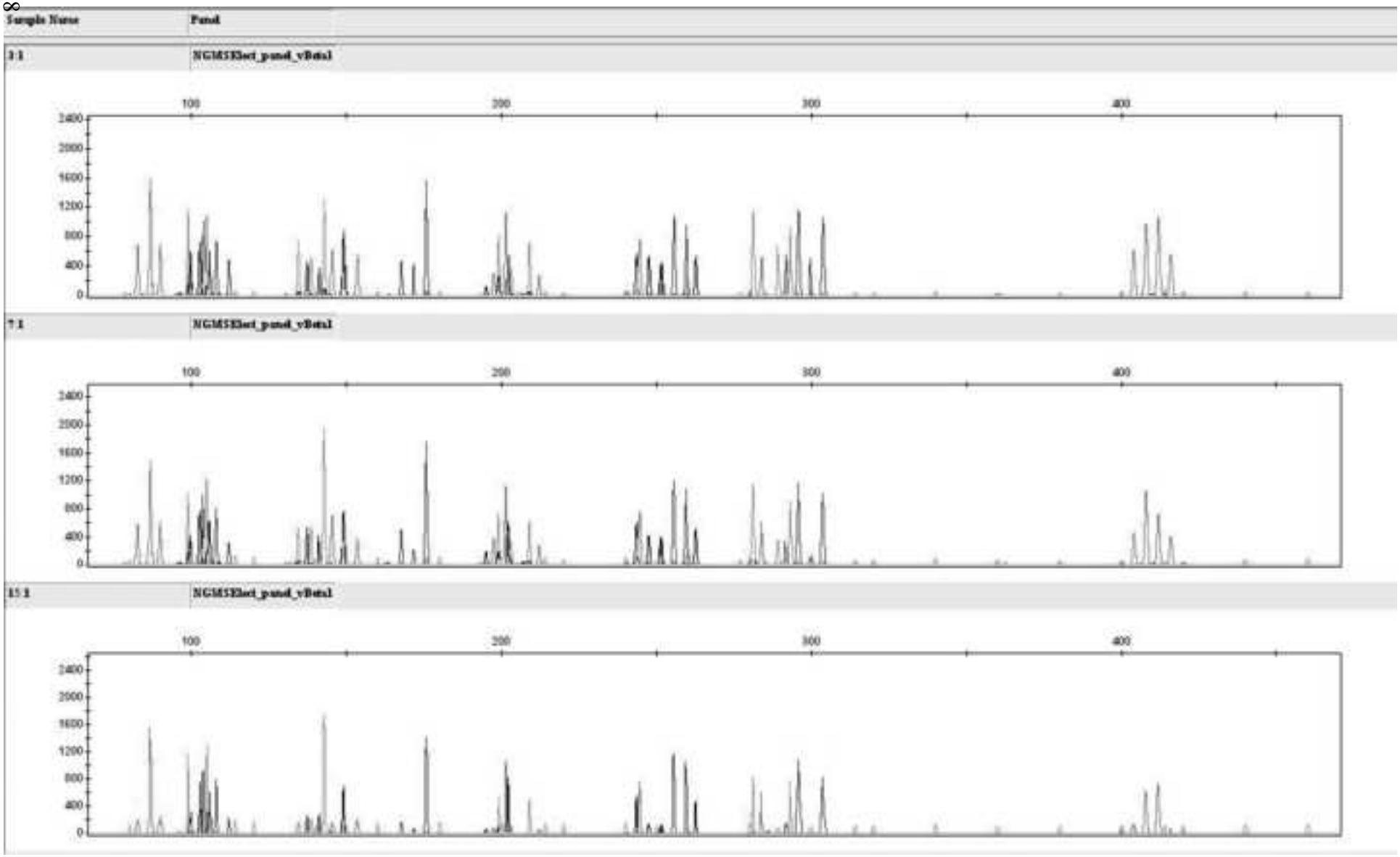


Figure5
Click here to download high resolution image

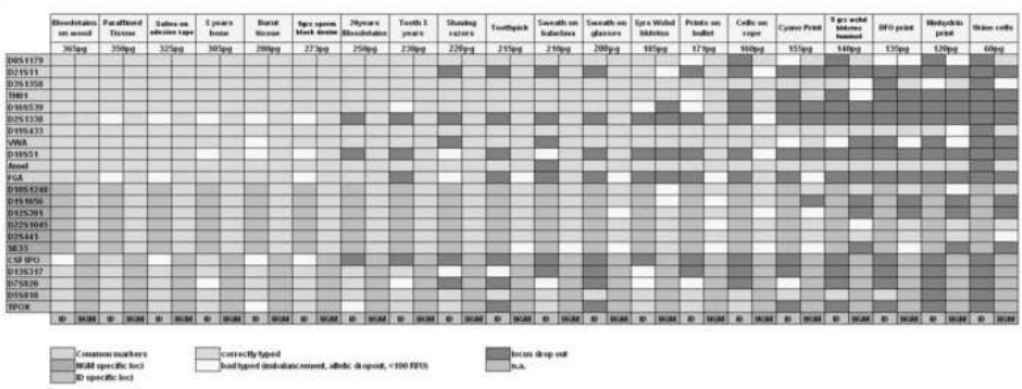
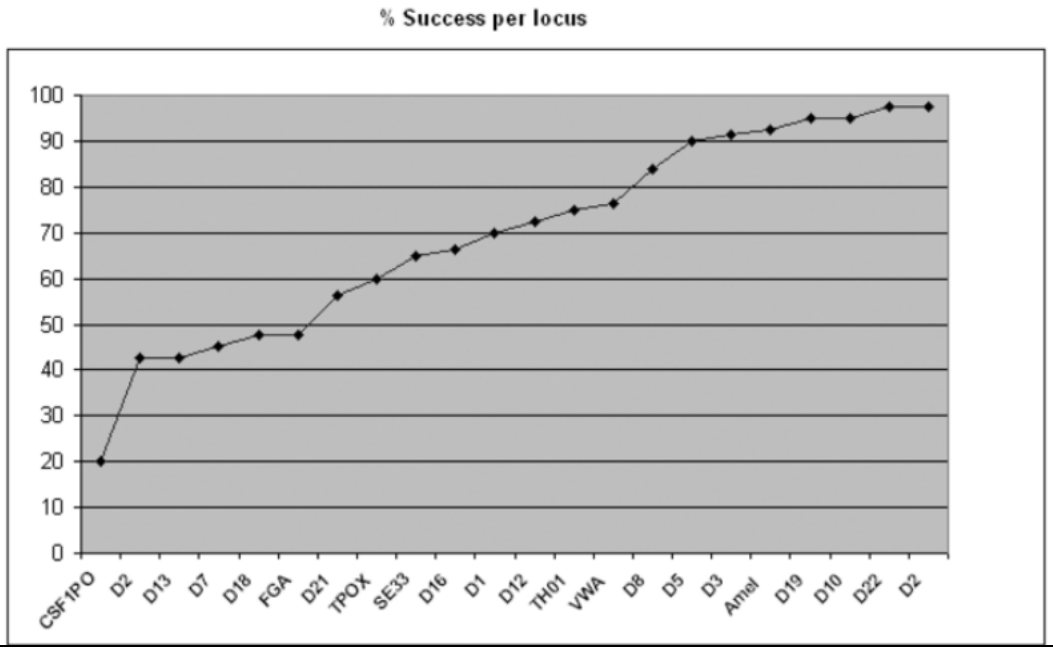


Figure6
Click here to download high resolution image



USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

b) Population Study for forensic statistical evaluations

We studied the variability in Mediterranean Area in order to create a useful population database, including the well established 15 autosomal STRs together with the 5 new ESS, the SE33 and some sex linked STRs routinely used in forensics. In addition it has been evaluated the variability of the 52 SNPplex recently introduced for forensic applications.

1. **Validation of a large Italian Database of 15 STR loci,** Forensic Sci Int. 156 (2006):266-268.

2. **Allele frequencies of 20 STRs from Northwest Spain (Galicia),** Forensic Sci. Int. Genet. 6 (2012) 149–150.

3. **Distribution of allele frequencies of 20 STRs loci in a population sample from Calabria, Southern Italy,** Forensic Sci. Int. Genet. 6 (2012) 137–138.

4.,**Variability of SE33 Locus in 2 Mediterranean Populations,** Journal of Forensic and Legal Medicine, (2012) *in process*

5. **Distribution of 8 X chromosomal STR loci in an Italian population sample (Calabria)** Forensic Sci. Int. Genet. (2012), doi:10.1016/j.fsigen.2012.05.011

6. **Genetic sub-structure in western Mediterranean populations revealed by 12 Y-chromosome STR loci.** Int J Legal Med. 123 (2009) 137-41.

7. **Microgeographic variation of Y-chromosome haplotypes in Italy,** Forensic Sci. Int. Genet. Suppl. Series 1(2008)239–241

8. **Study about the genetic variability of the SNP*for*ID 52-plex panel in Italian population samples,** Forensic Sci. Int. Genet. (2012),DOI: 10.1016/j.fsigen.2012.07.002

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Forensic Science International 156 (2006) 266–268
Forensic Science International
www.elsevier.com/locate/forsciint

Announcement of population data

Validation of a large Italian Database of 15 STR loci

Silvano Presciuttini [a,1,*], Nicoletta Cerri [b,2], Stefania Turrina [c,2], Benedetto Pennato [a], Milena Alù [d], Alessio Asmundo [e], Anna Barbaro [f], Ilaria Boschi [g], Loredana Buscemi [h], Luciana Caenazzo [i], Eugenia Carnevali [j], Domenico De Leo [c], Cosimo Di Nunno [l], Ranieri Domenici [m], Michela Maniscalco [n], Gabriella Peloso [o], Susi Pelotti [p], Andrea Piccinini [q], Daniele Podini [r], Ugo Ricci [s], Carlo Robino [t], Luigi Saravo [u], Andrea Verzeletti [b], Marina Venturi [v], Adriano Tagliabracci [h,3]

[a] *Center of Statistical Genetics, SS Abetone e Brennero 2, 56127 Pisa, Italy*
[b] *Department of Surgery, Radiology and Forensic Medicine, University of Brescia, Italy*
[c] *Department of Medicine and Public Health, University of Verona, Italy*
[d] *Department of Morphological and Forensic Sciences, University of Modena, Italy*
[e] *Institute of Legal Medicine, University of Messina, Italy*
[f] *Department of Forensic Genetics SIMEF, Reggio Calabria, Italy*
[g] *Institute of Legal Medicine, Catholic University, Rome, Italy*
[h] *Chair of Legal Medicine, University of Ancona, Italy*
[i] *Department of Environmental Medicine and Public Health, University of Padova, Italy*
[j] *Department of Surgery and Forensic Sciences, University of Perugia and Section of Legal Medicine, Hospital of Terni, Italy*
[l] *Department of Internal Medicine, University of Bari, Italy*
[m] *Department of Neurosciences, University of Pisa, Italy*
[n] *Andros Center srl, Palermo, Italy*
[o] *Department of Enviromental Medicine and Public Health., University of Pavia, Italy*
[p] *Department of Medicine and Public Health, Section of Legal Medicine, University of Bologna, Italy*
[q] *Institute of Legal Medicine, University of Milan, Italy*
[r] *Genoma srl, Roma, Italy*
[s] *Center of Medical and Molecular Genetics, Hospital "A. Meyer", Florence, Italy*
[t] *Department of Anatomy, Pharmacology and Legal Medicine, University of Turin, Italy*
[u] *Ra.C.I.S., Section of Biology, Messina, Italy*
[v] *Department of Biomedical Sciences, Section of Legal Medicine, University of Ferrara, Italy*

Received 14 December 2004; received in revised form 2 March 2005; accepted 2 March 2005
Available online 22 April 2005

Abstract

Results from a collaborative exercise with proficiency testing conducted by 20 Italian laboratories on the 15 loci included in the Identifiler® kit were analyzed by allele sharing methods and by standard population genetics tests. The validated database,

* Corresponding author. Tel.: +39 050 2213797; fax: +39 050 2213524.
E-mail address: sprex@biomed.unipi.it (S. Presciuttini).
[1] Responsible for data analysis.
[2] Exercise coordinator.
[3] President of the GeFI.

0379-0738/$ – see front matter
doi:10.1016/j.forsciint.2005.03.001

including about 1500 subjects, was merged with that of a previous exercise conducted on nine loci, and the resulting allele frequencies, subdivided by Italian region, were published on-line.

Keywords: Identifiler STR; Database validation; Population data

Population: Twenty laboratories scattered around Italy (16 from hospitals/universities, three from private companies, one from a national criminal justice service) typed 41–197 unrelated subjects of both sexes born in their region, totaling 1541 individuals.

Extraction: All labs but one extracted at least part of DNA samples from blood; saliva was also used as a source by 11 labs and three labs indicated other additional sources. Extraction methods varied by laboratory; most labs used Chelex-100, others indicated Qiagen, and three used phenol–chloroform; other commercial kits were also indicated (Amersham, Promega, Epicentre, Mac/Nag).

PCR: The Identifiler® kit (Applied Biosystems) was used by 12 labs; the combination of ProfilerPlus® + SGM Plus® + Green™ I (Applied Biosystems) was used by two labs, and one lab used a custom multiplex combination; four labs used the combination ProfilerPlus® + Cofiler® (Applied Biosystems, 13 loci).

Typing: Electrophoresis was carried out using five-color capillary separation by 12 labs, whereas six used a four-color separation apparatus. Two labs used vertical gels. Allele call was carried out by the Genotyper® software (Applied Biosystem) by 16 labs, whereas four used visual comparison with ladder.

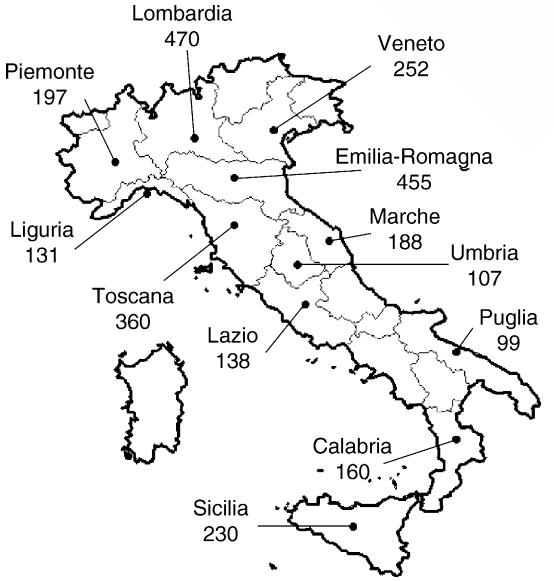


Fig. 1. Map of Italy showing regional boundaries. Numbers are sample sizes of the published database.

Analysis of data: Allele sharing between pairs of individuals within local datasets was analyzed with the Excel workbook AlleleSharingSheet.xls, and allele sharing among the entire database with the program AlleleSharingMacro.xls (both are available at http://statgen.dps.unipi.it/downloads/). Exact tests for Hardy–Weiberg (HW) equilibrium, F_{ST} analysis and tests of population differentiation were performed by Arlequin 2000 [3]. Homozygosity test was performed by Chi-square.

Results: Data of laboratories from the same region were merged. Allele frequencies were available for 12 (out of 20) Italian regions from north, center and south, together representing 77% of the entire Italian population (2001 census [1], Fig. 1). Data from a previous GeFI collaborative exercise [2], which included nine of the 15 loci examined here, were crosschecked against the new database; the repeated samples were discarded, and the two databases were merged. Thus, the final published tables (http://www.gefi-forensicdna.it) include nine loci typed in about 2800 individuals and six loci typed in more than 1500 individuals.

Quality control: Blind typing of two stains provided by the organizing committee.

Other remarks: Allele sharing analysis allowed correcting local databases for duplicate records and presence of possibly related individuals; global allele sharing analysis highlighted two pairs of individuals typed independently by different laboratories. Analysis of outlier genotypes (those with very low HW or contingency-table expectations) allowed correcting for typos. One lab sample that remained out of HW equilibrium for a locus even after applying the Bonferroni correction was discarded. Allele frequency distributions of 9 of the 15 loci have already been compared across different Italian studies [2]. The other six loci showed frequencies consistent with those published in the following reports (from groups not participating in the present exercise): D2S1338 [4], D16S539 [5], D19S43 [6], CFS1PO–TH01–TPOX [7]. The level of genetic differentiation among regions was low at all loci, so that the overall allele frequencies can be used in general forensic analyses in Italy.

This paper follows the guidelines for publication of population data requested by the journal [8].

References

[1] ISTAT - 14° censimento generale della popolazione e delle abitazioni 2001. Popolazione legale (ISBN: 88-458-1069-0) Roma, 2003.

[2] S. Presciuttini, F. Ciampini, M. Alù, N. Cerri, M. Dobosz, R. Domenici, G. Peloso, S. Pelotti, A. Piccinini, E. Ponzano, U. Ricci, Adriano Tagliabracci, J.E. Baley-Wilson, Francesco De Stefano and Vincenzo Pascali. Allele sharing in first-degree and unrelated pairs of individuals in the GeFI AmpFlSTR® Profiler Plus™ database, Forensic Sci. Int. 131 (2003) 85–89.

[3] S. Schneider, D. Roessli, L. Excoffier, Arlequin (Version 2000): A Software for Population Genetics Data Analysis, Genetics and Biometry Laboratory, University of Geneva, Switzerland, 2000.

[4] L. Garofano, M. Pizzamiglio, F. Donato, F. Biondi, M. Rossetti, B. Budowle, Italian population data on two new short tandem repeat loci: D2S1338 and Penta E, Forensic Sci. Int. 105 (1999) 131–136.

[5] L. Garofano, M. Pizzamiglio, C. Vecchio, G. Lago, T. Floris, G. D'Errico, G. Brembilla, A. Romano, B. Budowle, Italian population data on thirteen short tandem repeat loci: HUMTH01, D21S11, D18S51, HUMVWFA31, HUMFIBRA, D8S1179, HUMTPOX, HUMCSF1PO, D16S539, D7S820, D13S317, D5S818, D3S1358, Forensic Sci. Int. 97 (1998) 53–60.

[6] L. Garofano, M. Pizzamiglio, G.P. Bizzaro, F. Donato, M. Rossetti, B. Budowle, Italian population data on two new short tandem repeat loci: D6S477 and D19S433, Forensic Sci. Int. 101 (1999) 203–208.

[7] R. Biondo, A. Spinella, P. Montagna, P.S. Walsh, C. Holt, B. Budowle, Regional Italian Allele frequencies at nine short tandem repeat loci, Forensic Sci. Int. 115 (2001) 95–98.

[8] P. Lincoln, A. Carracedo, Publication of population data of human polymorphisms, Forensic Sci. Int. 110 (2000) 3–5.

Forensic Science International: Genetics 6 (2012) e149–e150

Contents lists available at SciVerse ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig

Forensic Population Genetics—Letter to the Editor

Allele frequencies of 20 STRs from Northwest Spain (Galicia)

Dear Sir,

Allele frequencies and forensic informativeness parameters for 15 established autosomal STRs and 5 new ESS autosomal STRs were obtained from 204 unrelated individuals of Northwest Spain (Galicia) with Identifiler® Plus kit of Applied Biosystems (typing D2S1338, D3S1358, D8S1179, D16S539, D18S51, D19S433, D21S11, FGA, TH01, vWA and Amelogenin) and an in-house designed pentaplex typing the five STRs (D1S1656, D2S441, D10S1248, D12S391 and D22S1045) adopted for the European Standard Set (ESS) in 2009 [1–5]. This study followed the guidelines for publication of forensic population data [6] as well as the recommendations of the ISFG with particular reference to the characterization of new forensic STR markers [7,8].

Samples were taken from paternity trios where informed consent had been previously obtained for extended population studies and this procedure was approved by the ethics committee of the University of Santiago de Compostela. DNA was extracted using QIAamp® DNA Micro kit and QIAamp® DNA Blood Mini kit following the manufacturer's protocol. DNA quantification were

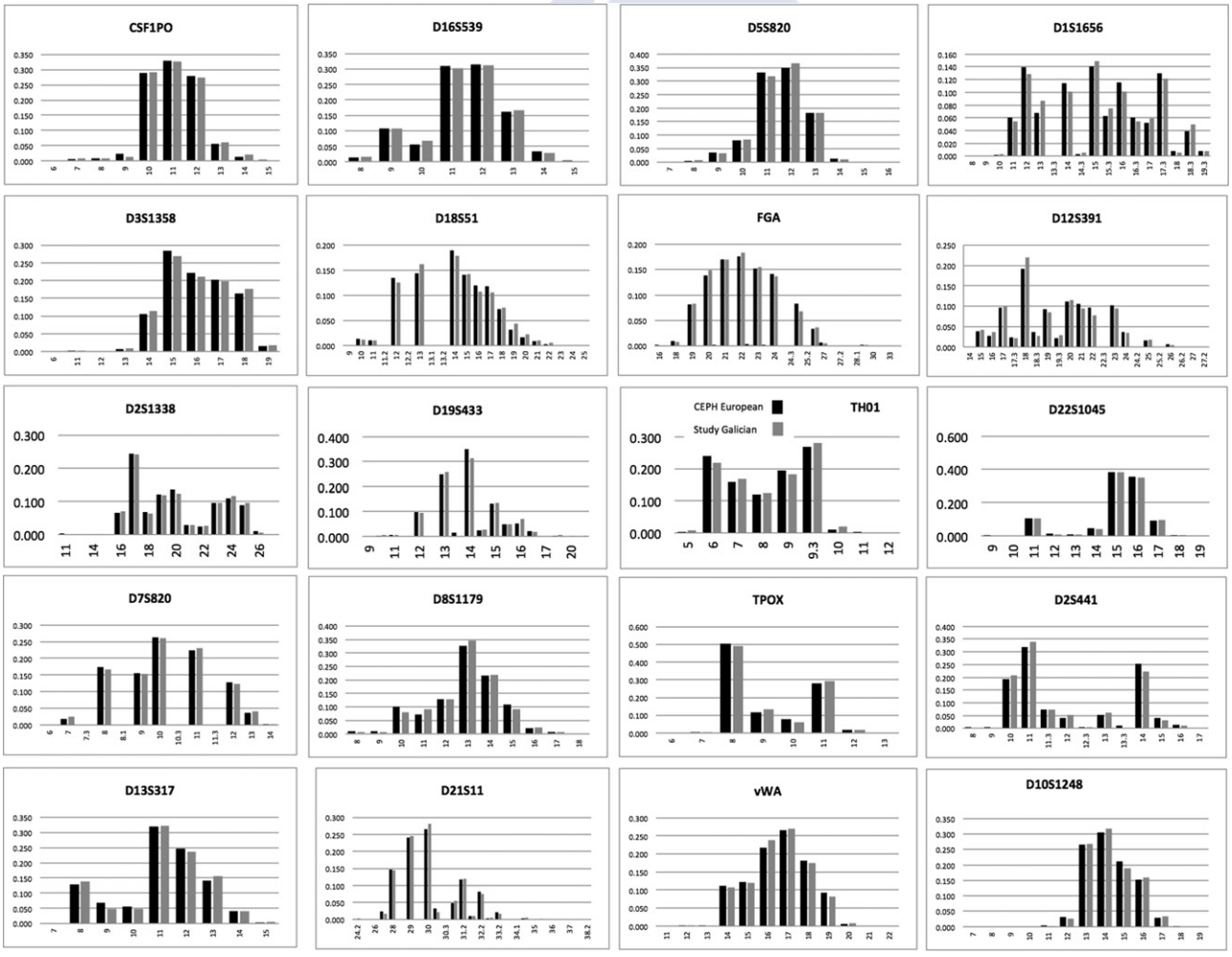


Fig. 1. Allele frequency distributions observed for 20 STRs in the study population of Galicians from NW Spain (light grey bars) compared with a combined European population group from the HGDP-CEPH genome diversity panel comprising: French from France; French Basque; Adygei from Caucasus; Russian; Orcadian from UK; Sardinian; Tuscan; and North Italian population samples.

1872-4973/$ – see front matter
doi:10.1016/j.fsigen.2012.02.009

made using Quantifiler™ Human DNA Quantification Kit (AB) using a 7500 Real-Time PCR System (AB).

Electrophoresis was performed using a 3130xl genetic analyzer (AB) with 36 cm capillary filled with POP-4TM Polymer (AB). Allele designations were made following manufacturer's protocol in the 15 established STRs of Identifiler® Plus kit except using 10 μl final PCR volume and for the 5 new ESS as previously described by Phillips et al. [1].

Allele frequency data and basic forensic statistics were obtained with Promega Powerstats software [9] for Galician population data and are outlined in supplementary Table S1. Hardy–Weinberg analysis was made using Arlequin ver. 3.5 [10] and is summarized in supplementary Table S2. No significant deviations from Hardy–Weinberg equilibrium were found.

As a point of reference, allele frequency data for combined European populations from the HGDP-CEPH human diversity panel were obtained using the pop.STR database [11] comprising: French from France; French Basque; Adygei from Caucasus; Russian; Orcadian from UK; Sardinian; Tuscan; and North Italian population samples. Summary allele frequencies for this population grouping are shown in supplementary Table S3. A graphic comparison of allele frequency distributions between CEPH Europeans and Galicians for 20 STRs is shown in Fig. 1, indicating very similar frequencies in both populations. We observed that certain alleles are present at low frequency in CEPH European populations but not found in the Galician population studied, these are: CSF1PO Allele: 15; D10S1248 18; D16S539 15; D19S433 13.2; D21S11 35.2; D22S1045 9; D2S1338 11; D2S441 8, 9 and 13.3; FGA 16, 20.2 and 23.2; TH01 11, and; vWA 13. In contrast, three alleles were observed uniquely in the Galician population in STR D19S433: repeats 13.2, 20 and 23.

Observed heterozygosity is above 0.650 in all STRs in both the European group and the Galician study population except for TPOX that has a value of 0.647 in the Europe population group. The most informative STR in the Galician population is D12S391 with a discrimination index of 0.900, near identical to the most informative STR in the European population group: D1S1656 that gives a discrimination index of 0.898.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.fsigen.2012.02.009.

References

[1] C. Phillips, L. Fernandez-Formoso, M. Garcia-Magariños, L. Porras, T. Tvedebrink, J. Amigo, M. Fondevila, A. Gomez-Tato, J. Alvarez-Dios, A. Freire-Arada, A. Gomez-Carballa, A. Mosquera-Miguel, Á. Carracedo, M.V. Lareu, Analysis of global variability in 15 established and 5 new European Standard Set (ESS) STRs using the CEPH human genome diversity panel, Forensic Sci. Int. Genet. 5 (2011) 155–169.

[2] M.V. Lareu, C. Pestoni, M. Schurenkamp, S. Rand, B. Brinkmann, Á. Carracedo, A highly variable STR at the D12S391 locus, Int. J. Legal Med. 109 (1996) 134–138.

[3] M.V. Lareu, S. Barral, A. Salas, C. Pestoni, Á. Carracedo, Sequence variation of a hypervariable short tandem repeat at the D1S1656 locus, Int. J. Legal Med. 111 (1998) 244–247.

[4] M.D. Coble, J.M. Butler, Characterization of new miniSTR loci to aid analysis of degraded DNA, J. Forensic Sci. 50 (2005) 43–53.

[5] T. Lederer, G. Braunschweiger, Commentary on: Coble MD, Butler JM. Characterization of new miniSTR loci to aid analysis of degraded DNA, J. Forensic Sci. 50 (2005) 43–53, J. Forensic Sci. 52 (2007) 493 and 494.

[6] P. Lincoln, Á. Carracedo, Publication of population data of human polymorphisms, Forensic Sci. Int. 110 (2000) 3–5.

[7] W. Bär, B. Brinkmann, B. Budowle, Á. Carracedo, P. Gill, P. Lincoln, W.R. Mayr, B. Olaisen, Further report of the DNA Commission of the ISFH regarding the use of short tandem repeat systems. International Society for Forensic Haemogenetics, Int. J. Legal Med. 110 (1997) 175–176.

[8] B. Olaisen, W. Bär, B. Brinkmann, B. Budowle, Á. Carracedo, P. Gill, P. Lincoln, W.R. Mayr, S. Rand, DNA recommendations 1997 of the International Society for Forensic Genetics, Vox Sang. 74 (1998) 61–63.

[9] Promega Powerstats Download Page: http://www.promega.com/geneticidtools/powerstats/.

[10] L. Excoffier, H.E. Lischer, Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows, Mol. Ecol. Resour. 10 (2010) 564–567.

[11] J. Amigo, C. Phillips, A. Salas, L. Fernandez-Formoso, Á. Carracedo, M.V. Lareu, pop.STR—an online population frequency browser for established and new forensic STRs, Forensic Sci. Int. Genet. Suppl. Series 2 (2009) 361–362.

L. Fernandez-Formoso
C. Phillips*
A. Rodriguez
R. Calvo
Forensic Genetics Unit, Institute of Legal Medicine, University of Santiago de Compostela, Santiago de Compostela, Galicia, Spain

A. Barbaro[a,b]
[a]*Forensic Genetics Unit, Institute of Legal Medicine, University of Santiago de Compostela, Santiago de Compostela, Galicia, Spain*
[b]*Studio Indagini Mediche E Forensi (SIMEF), Reggio Calabria, Italy*

M.V. Lareu
Forensic Genetics Unit, Institute of Legal Medicine, University of Santiago de Compostela, Santiago de Compostela, Galicia, Spain

Á. Carracedo[a,b]
[a]*Forensic Genetics Unit, Institute of Legal Medicine, University of Santiago de Compostela, Santiago de Compostela, Galicia, Spain*
[b]*Genomics Medicine Group, CIBERER, University of Santiago de Compostela, Galicia, Spain*

*Corresponding author. Tel.: +34 981 582 327; fax: +34 981 580 336
E-mail address: c.phillips@mac.com (C. Phillips)

8 November 2011

Forensic Science International: Genetics 6 (2012) e137–e138

Contents lists available at SciVerse ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig

Letter to the Editor

Distribution of allele frequencies of 20 STRs loci in a population sample from Calabria, Southern Italy

Dear Editor,

Allele frequencies of 20 STRs including the 5 new loci (D10S1248, D2S441, D1S1656, D12S391, D22S1045) approved by the European Union Council for the expansion of the European Standard Set (ESS) were calculated from a population sample from Calabria in southern Italy using the Applied Biosystems (AB) *AmpflSTR* Identifiler™ kit plus a next-generation 5-plex we previously developed as a supplementary assay to Identifiler™ [1–6].

Blood or saliva samples were collected from unrelated healthy donors belonging to the Calabrian population for at least 3 generations. Samples were taken from donors with previously obtained informed consent for population studies in accordance with Italian Law D.Lgs. 196/2003 and approved by SIMEF ISO-17025 procedures.

DNA was extracted by rapid resin (IstaGene Matrix System-Biorad) and then quantified with the Quantifiler™ Human DNA Quantification Kit using a 7300 Real Time System kit [7]. PCR amplification was performed using the *AmpFlSTR* Identifiler™ kit that amplifies the well-established loci: D2S1338, D3S1358, D8S1179, D16S539, D18S51, D19S433, D21S11, FGA, TH01, vWA, and amelogenin. We supplemented this analysis with a pentaplex we designed for the amplification of the five new ESS loci: D10S1248, D22S1045, D2S441, D1S1656 & D12S391, as previously described [8,3]. Positive and negative controls were used during all amplification steps.

PCR products were analyzed by capillary electrophoresis with an AB 3130 genetic analyzer and allele assignments made by comparison with Identifiler™ ladder or in the case of the pentaplex typing with reference to sequenced allelic ladders assembled in-house. For the five new STRs allele designations were determined following the repeat structure changes noted by Coble and Butler [9,10].

Statistical parameters of forensic interest (Dp: power of discrimination, PE: power of exclusion, RMP: random matching probability, etc.) were calculated using PowerStats v.1.2 software [11]. Hardy–Weinberg equilibrium and other population parameters were calculated using Arlequin software v.3.1. [12].

Allelic frequencies for all twenty STRs were compared to previously published population data. No significant differences were found in comparison with other European population data [13–18]. No significant deviations from Hardy–Weinberg expectations were found ($p > 0.05$). In all STRs except TPOX the observed heterozygosity was greater than 0.7, with the highest value in D1S1656. With the exception of D12S391 individual STRs showed a low exclusion power (PE) but the combined PE reached 0.99999999. Combined RMP using 20 loci was calculated to be 4.47×10^{-24}, therefore used together these twenty loci can distinguish samples with a probability of 99.99999%.

Allele frequencies and the resulting statistical parameters are given in Tables 1–2 available as e-components. A population comparison was made between the Italian samples described here and previously available data of Galicia (NW Spain) and the analysis is outlined in Table 3. Allele frequencies from these two southern European populations were very similar for each of the STRs studied.

The study laboratory has ISO17025 accreditation and participates in the quality control/proficiency testing of the GEP-ISFG WG (www.gep-isfg.org). This paper follows the guidelines for publication of population data requested by the journal. [19].

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.fsigen.2012.02.006.

References

[1] M.V. Lareu, C. Pestoni, M. Schürenkamp, S. Rand, B. Brinkmann, A. Carracedo, A highly variable STR at the D12S391 locus, Int. J. Legal Med. 109 (1996) 134–138.
[2] M.V. Lareu, S. Barral, A. Salas, C. Pestoni, A. Carracedo, Sequence variation of a hypervariable short tandem repeat at the D1S1656 locus, Int. J. Legal Med. 111 (1998) 244–247.
[3] M.D. Coble, J.M. Butler, Characterization of new miniSTR loci to aid analysis of degraded DNA, J. Forensic Sci. 50 (2005) 43–53.
[4] B. Olaisen, W. Bär, B. Brinkmann, B. Budowle, Á. Carracedo, P. Gill, P. Lincoln, W.R. Mayr, S. Rand, Vox Sang. 74 (1998) 61–63.
[5] P. Gill, L. Fereday, N. Morling, P.M. Schneider, The evolution of DNA databases recommendations for new European STR loci, Forensic Sci. Int. 156 (2006) 242–244.
[6] P. Gill, L. Fereday, N. Morling, P.M. Schneider, New multiplexes for Europe. Amendments and clarification of strategic development, Forensic Sci. Int. 163 (2006) 155–157.
[7] P.S. Walsh, D.A. Metzger, R. Higuchi, Chelex 100 as a medium for the simple extraction of DNA for PCR-based typing from forensic materials, Biotechniques 10 (1991) 506–513.
[8] C. Phillips, A. Barbaro, L. Fernandez-Formoso, A. Carracedo, M.V. Lareu, Development and validation of a next generation-STR pentaplex, Forensic Sci. Int. Genet. (Suppl. 2) (2009) 25–26.
[9] J.M. Butler, M.D. Coble, Regarding nomenclature for new miniSTR locus D10S1248, J. Forensic Sci. 52 (2007) 494.
[10] J.M. Butler, M.D. Coble, Author's response to letter to editor regarding nomenclature for new miniSTR locus D10S1248, J. Forensic Sci. 52 (2007) 494.
[11] A. Tereba, Tools for analysis of population statistics, Profiles in DNA 9 (1999) 14–16 (free software distributed at http://www.promega.com/geneticidtools).
[12] L. Excoffier, G. Laval, S. Schneider, Arlequin ver. 3.0: an integrated software package for population genetics data analysis, Evol. Bioinform. (Online) 1 (2005) 47–50.
[13] P. Hatzer-Grubwieser, B. Berger, D. Niederwieser, M. Steinlechner, Allele frequencies and concordance study of 16 STR loci—including the new European Standard Set (ESS) loci-in an Austrian population sample, Forensic Sci. Int. Genet. 6 (2012) 50–51.
[14] M. Arlindo, T. Lagoa, V. Martins, L.M. Cainé, M. Fátima Pinheiro, Allele frequencies of six miniSTR loci in the population of Northern Portugal, Forensic Sci. Int. Genet. 2 (2008) 379–381.

1872-4973/$ – see front matter
doi:10.1016/j.fsigen.2012.02.006

[15] V. Lopes, A. Serra, J. Gamero, L. Sampaio, F. Balsa, C. Oliveira, L. Batista, F. Corte-Real, D.N. Vieira, M.C. Vide, M.J. Anjos, M. Carvalho, Allelic frequency distribution of 17 STRs from Identifiler and PowerPlex-16 in Central Portugal area and the Azores archipelago, Forensic Sci. Int. Genet. 4 (2009) e1–e7.
[16] F. Brisighelli, C. Capelli, I. Boschi, P. Garagnani, M.V. Lareu, V.L. Pascali, A. Carracedo, Allele frequencies of fifteen STRs in a representative sample of the Italian population, Forensic Sci. Int. Genet. 3 (2009) 29–30.
[17] A. Berti, F. Brisighelli, A. Bosetti, E. Pilli, C. Trapani, V. Tullio, C. Franchi, G. Lago, C. Capelli, Allele frequencies of the new European Standard Set (ESS) loci in the Italian population, Forensic Sci. Int. Genet. 5 (2011) 548–549.
[18] L. Fernandez-Formoso, C. Phillips, A. Rodriguez, R. Calvo, A. Barbaro, M.V. Lareu, A. Carracedo, Allele frequencies of 20 STRs from Northwest Spain (Galicia), Forensic Sci. Int. Genet. 6 (2012) e149–e150.
[19] A. Carracedo, J.M. Butler, L. Gusmao, W. Parson, L. Roewer, P.M. Schneider, Publication of population data for forensic purposes, Forensic Sci. Int. Genet. 4 (2010) 145–147.

Anna Barbaro[a,b,*]
[a]*Studio Indagini Mediche E Forensi (SIMEF), Reggio Calabria, Italy*
[b]*Institute of Legal Medicine, University of Santiago de Compostela, Santiago de Compostela, Spain*

Chris Phillips
Luis Fernandez Formoso
Maria Victoria Lareu
Ángel Carracedo
Institute of Legal Medicine, University of Santiago de Compostela, Santiago de Compostela, Spain

*Corresponding author at: Studio Indagini Mediche E Forensi (SIMEF), Reggio Calabria, Italy
E-mail address: simef_dna@tiscali.it (A. Barbaro)

7 December 2011

Elsevier Editorial System(tm) for Journal of Forensic and Legal Medicine
Manuscript Draft

Manuscript Number:

Title: VARIABILITY OF SE33 LOCUS IN 2 MEDITERRANEAN POPULATIONS

Article Type: Short Report

Keywords: SE33; Variability; Calabria; Malta

Corresponding Author: Dr. Anna Barbaro,

Corresponding Author's Institution:

First Author: Anna Barbaro

Order of Authors: Anna Barbaro; Marisa Cassar; Patrizia Cormaci; Janessa C Grech

Abstract: The SE33 (ACTP2—human actin beta-actin-related pseudogene H-beta-Ac-psi-2) is one of the most informative STR systems for biological identification .
Variability of SE33 was studied in 2 Mediterranean populations (Calabria and Malta) using the AmpFlSTR NGM SElect™ PCR Amplification Kit (Applied Biosystems) and the PowerPlex ESI 17 (Promega).
A total of 41 different alleles were observed in the 2 examined populations with no allele being more frequent than 10,5%. Allelic frequencies, Hardy-Weinberg and statistical parameters were calculated using. No significant deviations from Hardy-Weinberg equilibrium were found.
Allelic frequencies were compared to previously published population data and no significant differences were found. Moreover comparison with non-European population showed no big distances between Germany and Morocco and between Hungary and Turkey.
Results confirmed the usefulness of SE33 for forensic identification, which should be added to the set of STRs loci routinely studied in caseworks and in paternity cases.

*Title Page (WITH Author Details)

VARIABILITY OF SE33 LOCUS IN 2 MEDITERRANEAN POPULATIONS

A.Barbaro[a] *
M. Cassar[b]
P.Cormaci[a]
J. C. Grech[b]

[a]Studio Indagini Mediche E Forensi (SIMEF), Reggio Calabria , Italy
[b] MLS BioDNA , Paola, Malta.

*Corresponding Author
Studio Indagini Mediche E Forensi (SIMEF)
Via Nicolò da Reggio 4, Reggio Calabria , Italy
Fax +390965891125
Email address: simef_dna@tiscali.it

VARIABILITY OF SE33 LOCUS IN 2 MEDITERRANEAN POPULATIONS

1.Introduction

The SE33 (ACTP2—human actin beta-actin-related pseudogene H-beta-Ac-psi-2) is one of the most informative STR systems for biological identification [1,2].Allele frequencies of SE33 were calculated from a population sample from Calabria in Southern Italy and another population sample from the Maltese Islands using PowerPlex ESI 17 (Promega).
Both laboratories performing this study are accredited UNI CEI/EN ISO 17025 and participate in the quality control/proficiency testing of the GEP-ISFG WG (www.gep-isfg.org), College of American Pathologists (CAP) and Collaborative Testing Services (CTS).

2. Materials and Methods

Blood or saliva samples were collected from 200 unrelated healthy donors belonging to the Calabrian (Southern Italy) population for at least 3 generations. Samples were taken from donors with previously obtained informed consent for population studies in accordance with Italian Law D.Lgs. 196/2003 and approved by SIMEF ISO-17025 procedures. Unrelated healthy donors having both grandparents of Maltese origin were recruited for the study following their informed consent.
DNA was extracted by rapid resin (IstaGene Matrix System-Biorad) and silica columns (AccuPrep Genomic DNA Extraction Kit- Bioneer) and then quantified with the Quantifiler™ Human DNA Quantification Kit using a 7300 Real Time System kit [3].
PCR amplification was performed using the next generation kits NGM SElect™ (Applied Biosystems) and PowerPlex® ESI 17 (Promega) which include the new 5 ESS markers and the highly discriminating SE33. This allows exchange of data with several central European countries using SE33 routinely. [4,5] Positive and negative controls were used during all amplification steps.
PCR products were analyzed by capillary electrophoresis with an AB 3130 genetic analyzer and allele assignment was carried out by comparison with the reference allelic ladders available in the kits using GeneMapper 3.2 software.

3. Results and Discussion

Allelic frequencies and statistical parameters of forensic interest (Dp: power of discrimination, PE: power of exclusion, RMP: random matching probability, etc.) were calculated using PowerStats v.1.2 software [6]. Hardy-Weinberg equilibrium and other population parameters were calculated using Arlequin software v.3.1 [7] and TFPGA v1.3.[8]
A total of 41 different alleles were observed in the 2 examined populations with no allele being more frequent than 10,5%. In the Maltese population more intermediate alleles than in Calabria were found; moreover 6 out of ladder alleles (17.3, 18.3, 20.3,32,33,34) were present.
The high number of alleles observed at SE33 locus confirmed the high degree of polymorphism.
No significant deviations from Hardy–Weinberg equilibrium were found.
Based on heterozygosity (greater than 0.7) and polymorphic information content PIC (greater than 0.9), SE33 could be considered as an informative locus in both populations.
Even if it shows a low exclusion power (PE) degree when used individually, however combined PE with the other 5 ESS new STRs is increased to 0.999 in both populations and the combined power of discrimination (PD) was 0.999999.
This means when used together these loci can distinguish samples from different individuals from Calabria with a probability of 99,9999%.
Allelic frequencies of SE33 were compared to previously published population data and no significant differences were found.[9-13]
When comparing with Sicily no overall significant genetic distances were found, while comparison to other populations showed significant ones. Moreover comparison with non-European population showed no big distances between Germany and Morocco and between Hungary and Turkey.

Allele frequencies and the resulting statistical parameters are given in Table1 available as e-component. Data of populations comparisons analysis are outlined in Tables 2,3, also available as e-components. This paper follows general guidelines for publication of population data. [14]

4. Conclusions

SE33 showed a similar variability in the 2 examined populations (Calabria and Malta) in comparison with other European populations: a total of 41 different alleles were observed with no allele being more frequent than 10,5%.
This confirmed the locus is effectively highly polymorphic and useful for forensic identification: it should be routinely added to the set of STRs loci commonly studied in casework and in paternity cases using as reference the database we generated.

5. Ethical standards

The study described in the present paper have been carried out using samples taken from people where informed consent had been previously obtained for research studies in accordance with Italian and Maltese Laws and to both labs approved UNI EN ISO 17025 procedures.

6. Conflict of interest

None

7. REFERENCES

[1] Rolf B, Schürenkamp M, Junge A, Brinkmann B,Sequence polymorphism at the tetranucleotide repeat of the human beta-actin related pseudogene H-beta-Ac-psi-2 (ACTBP2) locus, Int. J. Leg. Med. 110 (1997) 69–72.

[2] Wenda S, Dauber EM, Schwartz M, Jungbauer C,Weirich V, Wegener R , Mayr WR, ACTBP2 (alias ACTBP8) is localized on chromosome 6 (band 6q14), Forensic Sci. Int. Volume 148, Issues 2-3, 10 March 2005, Pages 207-209

[3] Walsh PS, Metzger DA, Higuchi R, Chelex 100 as a medium for the simple extraction of DNA for PCR-based typing from forensic materials, Biotechniques 10 (1991) 506–513.

[4]Gill P, Fereday L, Morling N, Schneider PM, The evolution of DNA databases recommendations for new European STR loci, Forensic Sci. Int. 156 (2006) 242–244.

[5] Gill P, Fereday L, Morling N, Schneider PM, New multiplexes for Europe. Amendments and clarification of strategic development, Forensic Sci. Int. 163 (2006) 155–157.

[6] Tereba A, Tools for analysis of population statistics, Profile in DNA (1999) 14–16
(free software distributed at http://www.promega.com/geneticidtools/).

[7]Excoffier L, Laval G, Schneider S, Arlequin ver. 3.0: an integrated software package for population genetics data analysis, Evol. Bioinform. (Online) 1 (2005) 47–50

[8] Miller MP,TFPGA (genetic data analysis software) http://herb.bio.nau.edu/miller/tfpga.htm.

[9] Barbaro A, Phillips C, Formoso LF, Lareu MV, Carracedo Á, Distribution of allele frequencies of 20 STRs loci in a population sample from Calabria, Southern Italy, Forensic Sci. Int. Genet. 6 (2012) 149–150.

[10] Berti A,Brisighelli F, Bosetti A, Pilli E, Trapani C, Tullio V, Franchi C,Lago G,Capelli C, Allele frequencies of the new European Standard Set (ESS) loci in the Italian population, Forensic Sci Int Genet. (2011) 548-9.

[11] Cruz C, Vieira-Silva C, Ribeiro T, Espinheira R, Genetic data for the locus SE33 in a south Portuguese population with Powerplex ES System, ICS 1288 (2006) 427– 429.

[12] Lászik A, Sótonyi P, Rand S, Hohof C, Frequency data for the STR locus ACTBP2 (SE33) in eight populations, Int J Legal Med 115 (2001) 94–96.

[13] Krause M, Heide KG, Krawczak M,SE33 allele and genotype frequencies in the population of Schleswig-Holstein (Northern Germany),ICS 1288 (2006) 400–402.

[14] Carracedo A, Butler JM, Gusmao L, Parson W, Roewer L,Schneider PM, Publication of population data for forensic purposes, Forensic Sci. Int. Genet. 4 (2010) 145–147.

20 July 2012

Forensic Science International: Genetics xxx (2012) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig

Letter to the Editor

Distribution of 8 X-chromosomal STR loci in an Italian population sample (Calabria)

Dear Editor,

DNA markers on the X-chromosome have been shown to be a powerful tool for solving complex relationship cases such as deficiency paternity testing, mother–son, father–daughter, grandmother–granddaughter, putative sisters kinship testing. Male individuals inherit their one X-Chr from their mother, while female individuals receive one X from the mother and the other one from the father. So, females fathered by the same man share their paternal chromosome X [1–4].

The distribution of 8 X-chromosomal STR loci in a Southern Italy population sample (Calabria) was investigated using the Mentype® Argus X-8 kit (Biotype AG, Dresden, Germany) that includes four X-STR duos in the linkage groups 1–4 (DXS10135–DXS8378, DXS7132–DXS10074, HPRTB–DXS10101 and DXS10134–DXS7423) [5].

Each of the four STR clusters spans less than 0.5 cM and the genetic distance between linkage groups is more than 50 cM so the clusters represent stable haplotypes that can be treated as unlinked, thereby providing highly informative tools for kinship testing.

Blood or saliva samples were collected from 200 unrelated healthy donors (110 women, 90 men) belonging to Calabria for at least 3 generations. Samples were taken with previously obtained informed consent for population studies, in accordance with Italian Law D.Lgs. 196/2003 and approved by SIMEF ISO-17025 procedures.

DNA was extracted by rapid resin (IstaGene Matrix System-Biorad) and then quantified with the Quantifiler™ Human DNA Quantification Kit using a 7300 Real Time System [6].

PCR amplification was carried out using the Mentype® Argus X-8 PCR Amplification Kit.

Positive and negative controls were included during all amplification steps.

PCR products were analyzed by capillary electrophoresis with an AB 3130 genetic analyzer and allele assignments made by comparison with the allelic ladder provided in the Argus X-8 kit.

Allele frequencies, Hardy–Weinberg equilibrium and genetic distances between different populations (Fst) were calculated using Arlequin software v.3.1 [7].

No significant deviations from Hardy–Weinberg expectations were found ($p > 0.05$).

Different statistical parameters (polymorphism information content (PIC), mean exclusion chance (MEC), average power of discrimination in females (PDF) and in males (PDM)) were calculated using GenoProof 2.0 Software (Qualitype AG, Dresden, Germany).

In all STRs except DXS8378 the observed heterozygosity was greater than 0.7, with the highest degree (0.9128) in DXS10135.

DXS10135 had the highest PIC value at 0.9062 while DXS8378 had the lowest at 0.6138.

In all loci, the PDF value was higher than the PDM value. The DXS10135 locus had the highest PDF value at 0.9858. The locus with the lowest PDF value was DXS8378.

In female combined RMP using 8 loci was calculated to be 4.06×10^{-8}, therefore used together these 20 loci can distinguish samples with a probability of 99.999996%.

In male combined RMP using 8 loci was calculated to be 4.91×10^{-5}, therefore used together these 20 loci can distinguish samples with a probability of 99.995090%.

Allelic frequencies for all 8 STRs were compared to previously published population data and no significant differences were found [8–12].

When comparing to other European samples, no overall significant genetic distances were found.

The only significant *p*-value was obtained between Calabria and North Italy at DXS10074 locus and DXS8378 while in Germany at DXS10074 locus.

Comparison to other populations showed significant genetic distances in Morocco at DXS7132 and Korea in all the studied markers with the exception of DXS10135.

Allele frequencies and the resulting statistical parameters for all 8 X-STRs are given in Table 1; results of population comparison analysis are outlined in Table 2.

Both tables are available as e-components.

The laboratory where the study was carried out is accredited ISO17025 and participates in the quality control/proficiency testing of the GEP-ISFG WG (www.gep-isfg.org).

This paper follows the guidelines for publication of population data requested by the journal [13].

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.fsigen.2012.05.011.

References

[1] R. Szibor, M. Krawczak, S. Hering, J. Edelmann, E. Kuhlisch, D. Krause, Use of X-linked markers for forensic purposes, Int. J. Legal Med. 117 (2003) 67–74.
[2] R. Szibor, X-chromosomal markers: past, present and future, Forensic Sci. Int. Genet. 1 (2007) 93–99.
[3] D. Desmarais, Y. Zhong, R. Chakraborty, C. Perreault, L. Busque, Development of a highly polymorphic STR marker for identity testing purposes at the human androgen receptor gene (HUMARA), J. Forensic Sci. 43 (1998) 1046–1049.
[4] R. Szibor, I. Plate, J. Edelmann, S. Hering, E. Kuhlisch, M. Michael, D. Krause, Chromosome X haplotyping in deficiency paternity testing principles and case report, Int. Congr. Ser. 1239 (2003) 815–820.
[5] W. Branicki, P. Wolanska-Nowak, A. Parys-Proszek, T. Kupie, Application of the Mentype Argus X-8 kit to forensic casework, Prob. Forensic Sci. 73 (2008) 53–64.
[6] P.S. Walsh, D.A. Metzger, R. Higuchi, Chelex 100 as a medium for the simple extraction of DNA for PCR-based typing from forensic materials, Biotechniques 10 (1991) 506–513.

1872-4973/$ – see front matter © 2012 Published by Elsevier Ireland Ltd.
http://dx.doi.org/10.1016/j.fsigen.2012.05.011

[7] L. Excoffier, G. Laval, S. Schneider, Arlequin (version 3.0): an integrated software package for population genetics data analysis, Evol. Bioinform. 1 (2005) 47–50.
[8] N. Cerri, A. Verzeletti, F. Gasparini, A. Poglio, E. Mazzeo, F. De Ferrari, Population data for 8 X-chromosome STR loci in a population, sample from Northern Italy and from the Sardinia Island, Forensic Sci. Int. Genet.: Suppl. Series 1 (2008) 173–175.
[9] D. Becker, H. Rodig, C. Augustin, J. Edelmann, F. Götz, S. Hering, R. Szibor, W. Brabetz, Population genetic evaluation of eight X-chromosomal short tandem repeat loci using Mentype Argus X-8 PCR amplification kit, Forensic Sci. Int. Genet. 2 (2008) 69–74.
[10] M. Gelabert-Besada, C. Alves, S. Ferreira, M. García-Magariños, L. Gusmão, P. Sánchez-Diz, Genetic characterization of Western Iberia using Mentype1 Argus X-8 kit, Forensic Sci. Int. Genet. 6 (2012) 39–41.
[11] K. Bentayebi, A. Picornell, M. Bouabdeallah, J.A. Castro, R. Aboukhalid, D. Squalli, M. Misericòrdia, S. Amzazi, Genetic diversity of 12 X-chromosomal short tandem repeats in the Moroccan population, Forensic Sci. Int. Genet. 6 (2012) 48–49.
[12] E. Lim, J.E. Sim, H.Y. Lee, M.J. Park, N.Y. Kim, H.J. Ahn, W.I. Yang, K.J. Shin, Genetic polymorphism and haplotype analysis of 4 tightly linked X-STR duos in Koreans, Croat. Med. J. 50 (2009) 305–312.
[13] A. Carracedo, J.M. Butler, L. Gusmao, W. Parson, L. Roewer, P.M. Schneider, Publication of population data for forensic purposes, Forensic Sci. Int. Genet. 4 (2010) 145–147.

A. Barbaro*
P. Cormaci
G. Falcone
S. Votano
A. La Marca
Studio Indagini Mediche E Forensi (SIMEF), Reggio Calabria, Italy

*Corresponding author
E-mail address: simef_dna@tiscali.it (A. Barbaro)

29 February 2012

Please cite this article in press as: A. Barbaro, et al., Distribution of 8 X-chromosomal STR loci in an Italian population sample (Calabria), Forensic Sci. Int. Genet. (2012), http://dx.doi.org/10.1016/j.fsigen.2012.05.011

Int J Legal Med (2009) 123:137–141
DOI 10.1007/s00414-008-0302-y

ORIGINAL ARTICLE

Genetic sub-structure in western Mediterranean populations revealed by 12 Y-chromosome STR loci

V. Rodríguez · C. Tomàs · J. J. Sánchez · J. A. Castro · M. M. Ramon · A. Barbaro · N. Morling · A. Picornell

Received: 10 June 2008 / Accepted: 18 November 2008 / Published online: 6 December 2008

Abstract Haplotype and allele frequencies of 12 Y-chromosome short tandem repeat (Y-STR) loci (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS385 a/b, DYS437, DYS438 and DYS439) included in the Powerplex® Y System were determined in seven western Mediterranean populations from Valencia, Majorca, Ibiza (eastern Spain), Sicily and three Calabrian provinces (southern Italy). Amongst the 554 males included in the study, 443 different haplotypes were observed of which 372 were only observed once. The other haplotypes were shared by two to seven men. The overall haplotype diversity was 0.9988±0.0002. These Y-STRs markers showed a low capacity of discrimination (56.3%) in the Ibiza population probably due to genetic drift. Comparisons between the populations studied and other neighbouring populations showed a clear genetic sub-structure in the western Mediterranean area.

Keywords Y-STRs · Eastern Spain · Ibiza · Sicily · Calabria

Electronic supplementary material The online version of this article (doi:10.1007/s00414-008-0302-y) contains supplementary material, which is available to authorized users.

V. Rodríguez · J. A. Castro · M. M. Ramon · A. Picornell (✉)
Institut Universitari d'Investigacions en Ciències de la Salut (IUNICS) i Laboratori de Genètica, Departament de Biologia, Universitat de les Illes Balears,
Carretera de Valldemossa, km 7.5,
07122 Palma de Mallorca, Illes Balears, Spain
e-mail: apicornell@uib.es

C. Tomàs · J. J. Sánchez · N. Morling
Section of Forensic Genetics, Department of Forensic Medicine, Faculty of Health Sciences, University of Copenhagen,
11 Frederik V's Vej,
2100 Copenhagen, Denmark

J. J. Sánchez
Delegación de Canarias,
Instituto Nacional de Toxicología y Ciencias Forenses,
Campus de Ciencias de la Salud, La Cuesta,
38320 Tenerife, Spain

A. Barbaro
Sezione di Genetica Forense,
Studio Indagini Mediche e Forensi (SIMEF),
89128 Reggio Calabria, Italy

Introduction

The analysis of Y-chromosome short tandem repeats (Y-STRs) has become a very useful tool, both in evolutionary studies and forensic casework. Two sets of Y-STR systems called minimal and extended haplotypes have been widely accepted by the forensic community and are currently used in the worldwide database: "Y-STR haplotype reference database" (YHRD; http://www.ystr.org) [38]. Here, we present a population study carried out using the Powerplex® Y System (Promega, Madison, WI, USA) containing 12 Y-STR loci (the nine Y-loci of the minimal haplotype, as well as the DYS437, DYS438 and DYS439 loci).

The International Society of Forensic Genetics (ISFG) recommended the use of regional Y-STR haplotype databases and the verification that no population sub-structure exists before pooling data from different regions [16]. It is especially important to evaluate genetic sub-structure in the western Mediterranean area, due to the existence of some geographical and/or cultural isolates such as Balearic populations [24, 37] and Calabrian populations [36]. We studied the genetic constitution of seven western Mediterranean populations (Valencia and two Balearic Islands, from eastern Spain, Sicily and three provinces in the Calabria region, southern Italy) with the aim of establishing a Y-STR

haplotype regional database and to evaluate sub-structuring of geographical sub-sets in this Mediterranean region.

Materials and methods

Population samples

Blood samples were obtained from 554 unrelated males belonging to seven populations in the western Mediterranean area (Figure shown in ESM 1)—three Spanish populations: Valencia (eastern coast of the Iberian Peninsula; n=59), Majorca (n=91) and Ibiza (n=96) (Balearic Islands) and four Italian populations: Sicily (n=115) and three populations from the region of Calabria (southern Italy), Reggio Calabria (n=97), Cosenza (n=37) and Catanzaro (n=59). Protocols were approved by the Danish local ethical committee (KF-01-037/03). DNA was extracted by using QIAamp spin columns (Qiagen, Hilden, Germany) following the manufacturer's recommendations.

Y-STR typing

Amplification of the 12 Y-chromosome STRs loci DYS19, DYS389 I/II, DYS390, DYS391, DYS392, DYS393, DYS385 a/b, DYS437, DYS438 and DYS439 was carried out according to the kit Powerplex® Y System protocol, from 2 ng DNA template in a 10-μl final reaction volume, using a GeneAmp PCR system 2400 Thermal Cycler (Perkin-Elmer, Waltham, MA, USA).

For genetic typing, an ABI PRISM® 3100 Genetic Analyser along with GeneScan® 3.7 and Genotyper® v. 3.7 software (Applied Biosystems, Foster City, CA, USA) were used. All samples were tested twice. Allelic designation was based on comparison to the Powerplex® Y System allelic ladder. Allele nomenclature was according to the ISFG guidelines [16].

Sequence analysis

A new variant allele was sequenced on both strands. Briefly, the samples were amplified using unlabelled primers [3], amplicons were purified with a QIAquick PCR purification kit (Qiagen) and the sequence was determined using the Big Dye® Terminator Cycle Sequencing kit v. 3.1 (Applied Biosystems) and an ABI PRISM® 3130 Genetic Analyser (Applied Biosystems). Sequences were aligned using the Bioedit program v. 7.0.5.3 [17].

Quality control

Proficiency testing of the Spanish and Portuguese Working Group of the International Society for Forensic Genetics (GEP-ISFG, http://www.gep-isfg.org/) was carried out as quality control.

Statistical analysis

Allele and haplotype frequencies were estimated by gene counting. Haplotype and gene diversities, population differentiation parameters (F_{ST} and R_{ST}) and analysis of molecular variance (AMOVA) were calculated using ARLEQUIN v. 3.01 [11]. Discrimination capacity was calculated as the percentage of different haplotypes and haplotype match probability as 1-haplotype diversity. All statistical parameters were calculated for both minimal and extended haplotypes.

In order to examine the relationship of the populations studied with other neighbouring populations, Reynolds' genetic distances [30], calculated using PHYLIP v. 3.67 [12], were used to generate the multi-dimensional scaling (MDS) plot performed using the SPSS v. 15.0 (SPSS, Inc., Chicago, IL, USA).

Results and discussion

Allele frequencies and gene diversities of each Y-STR of the populations under study are shown in the table in ESM 2. DYS392 and DYS438 showed bimodal distribution of allele frequencies. In DYS392, modality was shared by DYS392-11 and DYS392-13 alleles, with DYS392-13 the most common allele amongst the Spanish populations and DYS392-11 the most frequent in southern Italy. These results are consistent with previous studies showing a longitudinal decrease of frequencies from the west to the east of the European landscape for DYS392-13 and, conversely, a decrease in the opposite direction for DYS392-11. The Neolithic demic diffusion could explain these two opposite patterns, with the DYS392-13 allele present in the proto-European gene pool [27, 28]. The clinal frequency pattern observed in the DYS438 system, with DYS438-12 as the most frequent in the Spanish and Sicilian populations and DYS438-10 the most frequent in Calabria, could also be due to the same Neolithic effect.

Gene diversities ranged from around 0.85 (in DYS385) to approximately 0.50 (in DYS392). Generally, Italian populations had higher gene diversities than the Spanish populations, following the same pattern found in bi-allelic Y-chromosome markers, with an increasing diversity trend from Spain to Greece, maybe due to the impact of the arrival of haplotypes in Europe from the Middle East [13]. Ibiza showed especially low gene diversities for DYS389II, DYS390, DYS391, DYS385, DYS438 and DYS439 loci.

Extra peaks were reproducibly obtained at DYS19 and DYS385, representing the presence of duplicated regions in

Table 1 Forensic parameters for the seven western Mediterranean populations studied using the minimal and the Powerplex® Y haplotypes

	Ibiza (n=96)	Majorca (n=91)	Valencia (n=59)	Sicily (n=115)	Catanzaro (n=59)	Cosenza (n=37)	R. Calabria (n=97)	Total (n=554)
Minimal 9 Y-STR haplotype								
Number of haplotypes	43	74	49	98	52	33	83	379
Unique haplotypes	19	53	37	72	37	23	58	299
Haplotype diversity ± SD	0.9695± 0.0061	0.9927± 0.0035	0.9930± 0.0047	0.9968± 0.0018	0.9959± 0.0039	0.9925± 0.0088	0.9968± 0.0020	0.9968± 0.0006
Discrimination capacity (%)	44.79	81.32	83.05	85.22	88.14	89.19	85.57	68.41
Match probability (%)	3.05	0.73	0.70	0.32	0.41	0.75	0.32	0.32
Powerplex 12 Y-STR haplotype								
Number of haplotypes	54	82	56	103	54	33	85	443
Unique haplotypes	30	70	49	86	43	26	68	372
Haplotype diversity ± SD	0.9807± 0.0050	0.9968± 0.0025	0.9982± 0.0035	0.9976± 0.0017	0.9971± 0.0037	0.9925± 0.0088	0.9972± 0.0020	0.9988± 0.0002
Discrimination capacity (%)	56.25	90.11	94.92	89.57	91.53	89.19	87.63	79.96
Match probability (%)	1.93	0.32	0.18	0.24	0.29	0.75	0.28	0.12

the Y-chromosome (e.g. [2, 6, 19, 20, 32]). Duplications were observed in six individuals: at locus DYS19, alleles 13 and 14 (once) and at loci DYS385, alleles 13–14–15 (once), 13–17–18 (twice) and 13–18–19 (twice).

Three alleles not included in the Powerplex® Y allelic ladder were observed. DYS438-7 and DYS438-13 alleles have been reported in other populations (e.g. [8, 23]), but to our knowledge, DYS19-9 has not been reported before. Sequence analysis (GenBank: FJ196286) confirmed the number of repeats attributed: $(TAGA)_3tagg(TAGA)_6$.

Amongst the 554 western Mediterranean males analysed, 443 different haplotypes were observed (Table in ESM 3), of which 372 were only observed once. The other haplotypes were shared by two to seven men. The most frequent haplotypes were h314 (15–12–29–22–10–11–14–14,14–16–10–12) and h356 (15–13–29–24–11–13–13–11,14–14–12–13), both found in seven men from the Ibiza population but absent from the other studied populations.

The haplotypes (without locus DYS437, not included in the YHR database) were searched against the haplotypes in the YHRD (release 18), and 149 haplotypes were matched to at least one YHRD sample. The most frequent haplotypes h314 and h356 matched with three and six samples of European origin, respectively, in a worldwide database of 38,761 haplotypes. It is noteworthy that 46 of the other haplotypes (almost all from Valencia, the Balearic Islands and Sicily) matched with north African or African samples. This result is concordant with other studies showing African influences in these Mediterranean populations (e.g. [14, 15, 24, 35]).

Table 1 shows the forensic parameters for the 12-loci Powerplex® Y System haplotypes compared with the diversity values of haplotypes based on the nine-loci minimal haplotype. The overall haplotype diversity only increased by 0.20% (ranging from 0% for Cosenza to 1.15% for Ibiza) by using the 12-loci Powerplex® Y System instead of the minimal haplotype.

The discrimination capacity ranged from 87.63% (Reggio Calabria) to 94.92% (Valencia) except in the Ibizan population (56.25%). Ibiza also showed a reduced genetic diversity in previous genetic studies [24, 25, 37]. These results are in accordance with the historical and demographic data of the island population (an isolated, consanguineous population with a reduced effective population size) [1, 22]. Therefore, the high haplotype match probability in Ibiza (1.93%) must be taken into account in forensic practice.

AMOVA analysis of the seven Mediterranean populations showed a significant value (F_{ST}=0.0499, P<0.0001). Pairwise analyses (Table in ESM 4) evidenced two

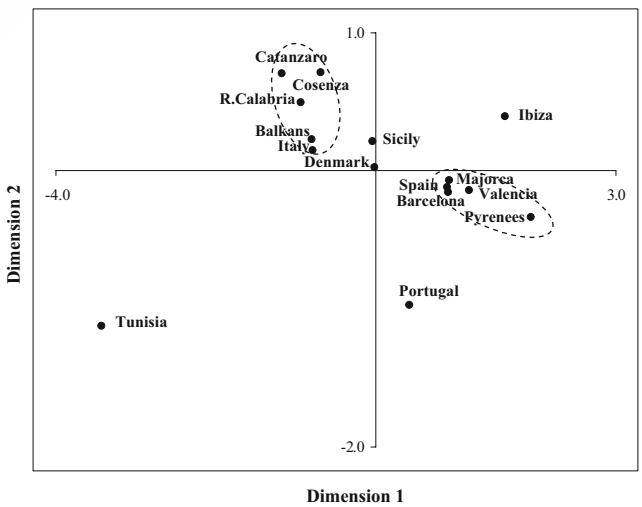


Fig. 1 MDS plot based on Reynolds' distances (Spain [23], Pyrenees [21], Barcelona [33], Portugal [26], Italy [29], the Balkans [5], Tunisia [9], Denmark [18], Majorca, Valencia, Ibiza, Catanzaro, Cosenza, Reggio Calabria and Sicily (this study))

significantly different sub-sets: one with the Spanish populations (Valencia, Ibiza and Majorca) and another with the Calabrian samples (Catanzaro, Cosenza and Reggio Calabria). Sicily was closer to the Italian than to the Spanish populations, although presented significant differences with Reggio Calabria ($P<0.005$). Significant differences between groups ($F_{CT}=0.0603$, $P<0.0001$) were found with a three-hierarchical AMOVA performed grouping the populations according to pairwise analysis results (Spanish populations, Sicily and Calabrian populations).

Figure 1 shows a multi-dimensional scaling plot where 15 populations have been included. The Tunisia population showed a displaced position, in accordance with other studies that have suggested that the Mediterranean Sea may have acted as a relative north-to-south geographic barrier to gene flow [4, 10, 28]. With the exception of Sicily and Ibiza, the Spanish and Italian populations under study grouped together with other Spanish and Italian samples, respectively. On the one hand, Sicily presented an intermediate position between Italian and Spanish populations. No consensus on the genetic landscape of the Sicilian population has been established to date. Whilst some authors claim the differentiation of the Sicilian population from Italy and from the western Mediterranean basin [7, 31], other studies indicate Sicily is closely related to other Italian populations [13, 15, 34]. On the other hand, Ibiza showed a large distance from the Spanish group, in accordance with the fact that Ibiza has important historical and genetic differences from other insular and continental populations in the western Mediterranean area [24, 25, 37].

In conclusion, the results of the present study provide a useful Y-STR haplotype dataset, for the western Mediterranean region, where some geographical and/or cultural isolates exist based on demographic, historical and genetic data. A clear genetic sub-structure between population groups (Spanish, Sicilian and Calabrian populations) was observed. Therefore, local databases must be used in the forensic field to correctly weigh the value of the evidence of a Y profile match. Special care should be taken in male identification in the Ibizan population, due to the very low discrimination capacity found for the 12 Y-STR loci included in the Powerplex® Y System.

Acknowledgements This work was partially supported by grants PRDT-2003-12099 and PRDIB-2006-687872 from the Direcció General de R+D+I (Comunitat Autònoma de les Illes Balears) and by Ellen and Aage Andersen's Foundation.

References

1. Alarco Von Perfall C (1981) Cultura y personalidad en Ibiza. Editora Nacional, Madrid
2. Balaresque P, Parkin EJ, Roewer L et al (2008) Genomic complexity of the Y-STR DYS19: inversions, deletions and founder lineages carrying duplications. Int J Legal Med (in press)
3. Beleza S, Alves C, González-Neira A, Lareu M, Amorim A, Carracedo A, Gusmão L (2003) Extending STR markers in Y chromosome haplotypes. Int J Legal Med 117:27–33
4. Bosch E, Calafell F, Comas D, Oefner PJ, Underhill PA, Bertranpetit J (2001) High-resolution analysis of human Y-chromosome variation shows a sharp discontinuity and limited gene flow between northwestern Africa and the Iberian Peninsula. Am J Hum Genet 68:1019–1029
5. Bosch E, Calafell F, González-Neira A et al (2006) Paternal and maternal lineages in the Balkans show a homogeneous landscape over linguistic barriers, except for the isolated Aromuns. Ann Hum Genet 70:459–487
6. Butler JM, Decker AE, Kline MC, Vallone PM (2005) Chromosomal duplications along the Y-chromosome and their potential impact on Y-STR interpretation. J Forensic Sci 50:853–859
7. Calò CM, Garofano L, Mameli A, Pizzamiglio M, Vona G (2003) Genetic analysis of a Sicilian population using 15 short tandem repeats. Hum Biol 75:163–178
8. Chang YM, Perumal R, Keat PY, Kuehn DL (2007) Haplotype diversity of 16 Y-chromosomal STRs in three main ethnic populations (Malays, Chinese and Indians) in Malaysia. Forensic Sci Int 167:70–76
9. Cherni L, Pereira L, Goios A et al (2005) Y-chromosomal STR haplotypes in three ethnic groups and one cosmopolitan population from Tunisia. Forensic Sci Int 152:95–99
10. Comas D, Calafell F, Benchemsi N, Helal A, Lefranc G, Stoneking M, Batzer MA, Bertranpetit J, Sajantilla A (2000) Alu insertion polymorphisms in NW Africa and the Iberian peninsula: evidence for a strong genetic boundary through the Gibraltar straits. Hum Genet 107:312–329
11. Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0: an integrated software package for population genetics data analysis. Evol Bioinf Online 1:47–50
12. Felsenstein J (2007) PHYLIP (Phylogeny Inference Package) version 3.67. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle
13. Francalacci P, Morelli L, Underhill PA et al (2003) Peopling of three Mediterranean islands (Corsica, Sardinia, and Sicily) inferred by Y-chromosome biallelic variability. Am J Phys Anthropol 121:270–279
14. Gérard N, Berriche S, Aouizerate A, Dieterlen F, Lucotte G (2006) North African Berber and Arab influences in the western Mediterranean revealed by Y-chromosome DNA haplotypes. Hum Biol 78:307–316
15. Ghiani ME, Piras IS, Mitchell RJ, Vona G (2004) Y-chromosome 10 locus short tandem repeat haplotypes in a population sample from Sicily Italy. Legal Med 6:89–96
16. Gusmão L, Butler JM, Carracedo A et al (2006) DNA Commission of the International Society of Forensic Genetics (ISFG): an update of the recommendation on the use of Y-STRs in forensic analysis. Int J Legal Med 120:191–200
17. Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 41:95–98
18. Hallenberg C, Nielsen K, Simonsen B, Sánchez J, Morling N (2005) Y-chromosome STR haplotypes in Danes. Forensic Sci Int 155:205–210
19. Hohoff C, Dewa K, Sibbing U, Hoppe K, Forster P, Brinkmann B (2007) Y-chromosomal microsatellite mutation rates in a population sample from northwestern Germany. Int J Legal Med 121:359–363

20. Lim SK, Xue Y, Parkin EJ, Tyler-Smith C (2007) Variation of 52 new Y-STR loci in the Y Chromosome Consortium worldwide panel of 76 diverse individuals. Int J Legal Med 121:124–127
21. López AM, Alvarez S, Gusmão L et al (2004) Population data for 16 Y-chromosome STRs in four populations from Pyrenees (Spain). Forensic Sci Int 140:125–129
22. Macabich I (1966) Historia de Ibiza. Volumen I. Daedalus, Palma de Mallorca
23. Martín P, García-Hirschfeld J, García O et al (2004) A Spanish population study of 17 Y-chromosome STR loci. Forensic Sci Int 139(1):231–235
24. Picornell A, Gómez-Barbeito L, Tomàs C, Castro JA, Ramon MM (2005) Mitochondrial DNA HVRI variation in Balearic populations. Am J Phys Anthropol 128:119–130
25. Picornell A, Miguel A, Castro JA, Ramon MM, Arya R, Crawford MH (1996) Genetic variation in the population of Ibiza (Spain): genetic structure, geography and language. Hum Biol 68:899–913
26. Pontes ML, Cainé L, Abrantes D, Lima G, Pinheiro MF (2007) Allele frequencies and population data for 17 Y-STR loci (AmpFlSTR Y-filer) in a northern Portuguese population sample. Forensic Sci Int 170:62–67
27. Quintana-Murci L, Semino O, Minch E, Passarino G, Brega A, Santachiara-Benerecetti AS (1999) Further characteristics of proto-European Y chromosomes. Eur J Hum Genet 7:603–608
28. Quintana-Murci L, Veitia R, Fellous M, Semino O, Poloni ES (2003) Genetic structure of Mediterranean populations revealed by Y-chromosome haplotype analysis. Am J Phys Anthropol 121:157–171
29. Rapone C, Geraci A, Capelli C et al (2007) Y chromosome haplotypes in central-south Italy: implication for reference database. Forensic Sci Int 172:67–71
30. Reynolds J, Weir BS, Cockerham CC (1983) Estimation of the coancestry coefficient: basis for a short-term genetic distance. Genetics 105:767–779
31. Robino C, Inturri S, Gino S et al (2006) Y-chromosomal STR haplotypes in Sicily. Forensic Sci Int 159:235–240
32. Roewer L, Willuweit S, Krüger C et al (2008) Analysis of Y chromosome STR haplotypes in the European part of Russia reveals high diversities but non-significant genetic distances between populations. Int J Legal Med 122:219–223
33. Sánchez C, Barrot C, Xifró A et al (2007) Haplotype frequencies of 16 Y-chromosome STR loci in the Barcelona metropolitan area population using Y-Filer™ kit. Forensic Sci Int 172:211–217
34. Scozzari R, Cruciani F, Pangrazio A et al (2001) Human Y-chromosome variation in the western Mediterranean area: implications for the peopling of the region. Hum Immunol 62:871–884
35. Semino O, Torroni A, Scozzari R, Brega A, De Benedictis G, Santachiara Benerecetti AS (1989) Mitochondrial DNA polymorphisms in Italy. III. Population data from Sicily: a possible quantitation of maternal African ancestry. Ann Hum Genet 53:193–202
36. Tagarelli A, Piro A, Tagarelli G, Zinno F (2000) Color-blindness in Calabria (southern Italy): a north–south decreasing trend. Am J Hum Biol 12:17–24
37. Tomàs C, Jiménez G, Picornell A, Castro JA, Ramon MM (2006) Differential maternal and paternal contributions to the genetic pool of Ibiza Island, Balearic Archipelago. Am J Phys Anthropol 129:268–278
38. Willuweit S, Roewer L, on behalf of the International Forensic Y Chromosome User Group (2007) Y chromosome haplotype reference database (YHRD): update. Forensic Sci Int: Genetics 1:83–87

Available online at www.sciencedirect.com
ScienceDirect
Forensic Science International: Genetics Supplement Series 1 (2008) 239–241
www.elsevier.com/locate/FSIGSS

Research article

Microgeographic variation of Y-chromosome haplotypes in Italy

S. Pelotti [a,*], C. Bini [a], A. Barbaro [b], L. Caenazzo [c], E. Carnevali [d], N. Cerri [e], R. Domenici [f], G. Ferri [g], M. Maniscalco [h], V. Onofri [i], A. Piccinini [j], C. Previderè [k], U. Ricci [l], C. Robino [m], F. Scarnicci [n], F. Torricelli [o], M. Venturi [p], S. Presciuttini [q]

GeFI's group of Y-chromosome characterization

[a] *Department of Medicine and Public Health, Section of Legal Medicine, University of Bologna, Italy*
[b] *Department of Forensic Genetics, Studio Indagini Mediche e Forensi (SIMEF), Italy*
[c] *Department of Environmental Medicine and Public Health, University of Padova, Italy*
[d] *Department of Surgery and Forensic Sciences, University of Perugia and Section of Legal Medicine, Hospital of Terni, Italy*
[e] *Department of Surgery, Radiology and Forensic Medicine, University of Brescia, Italy*
[f] *University of Pisa, Italy*
[g] *Department of Diagnostic and Laboratory Service and Legal Medicine, Section of Legal Medicine, University of Modena and Reggio Emilia, Italy*
[h] *Andros Day Surgery, Reproduction Medicine Center, Palermo, Italy*
[i] *Institute of Legal Medicine, University of Ancona, Italy*
[j] *Institute of Legal Medicine, University of Milan, Italy*
[k] *Department of Environmental Medicine and Public Health, University of Pavia, Italy*
[l] *Medical Genetic Unit, Azienda Ospedaliero Universitaria "A. Meyer", Florence, Italy*
[m] *Department of Anatomy, Pharmacology and Legal Medicine, University of Turin, Italy*
[n] *Institute of Legal Medicine, Università Cattolica Sacro Cuore, Rome, Italy*
[o] *Genetic Diagnostic Unit, Azienda Ospedaliero Universitaria Careggi, Florence, Italy*
[p] *Department of Biomedical Sciences, Section of Legal Medicine, University of Ferrara, Italy*
[q] *Center of Statistical Genetics, University of Pisa, Italy*

Received 17 August 2007; accepted 8 October 2007

Abstract

Within an Italian collaborative exercise on the extended haplotype of the Y-chromosome, 1288 subjects were typed by the AmpFlSTR YFiler Amplification Kit (AB Applied Biosystems) and other 526 were typed by the PowerPlex Y® System (Promega). The sampling scheme included either a ''regional'' or a ''local'' recruitment, the first referring to individuals born in the region of the participating lab, the second referring to individuals coming from small villages. Total sample sizes were $N = 954$ and 860, respectively. A significant decrease of haplotype diversity was found in the local samples. The results may be of interest in forensic applications of the Y-chromosome.

Keywords: Y-chromosome; Population data; Short tandem repeats; Haplotypes

1. Introduction

Up to now the haplotype diversity using nine Y-STRs comprising the so-called minimal haplotype loci was studied among worldwide population samples showing that there are significant portions of haplotypes in several populations which cannot be resolved. Evaluation of haplotype discrimination capacity of 35 Y-STRs was recently evaluated and complete resolution of the pooled population was achieved by additional genotyping of further loci [1]. Y-STRs generating haplotypes were studied in 2001 by GeFI collaborative exercise on 1176 Italian individuals from different regions [2]. The typed loci were DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS385 and a low degree of variations was shown among regions. In forensic genetics laboratories Y-STRs multiplex kit, based on 17 markers validated for forensic applications, have become widely used in the last years for the high power of discrimination at minimal samples consumption.

* Corresponding author. Tel.: +39 0512088343; fax: +39 0512088358.
E-mail address: susi.pelotti@unibo.it (S. Pelotti).

1875-1768/$ – see front matter © 2008 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.fsigss.2007.10.083

	Northern Italy							Central Italy									
Nine-locus haplotype	Bergamo	Brescia	Cuneo	Rimini	Valmarecchia	Urbino	AscoliPiceno	Fabriano	Matelica	Buti	Col	Msb	Nor	Sezze	Total counts	Expected counts	Freq.	Rank in Italy
14-13-29-24-11-13-13-11-14		6	4	3		3	1			2		1	4	1	25	15.6	0.040	1
14-13-29-24-10-13-13-11-14		4	4	2			2			1					13	6.2	0.021	2
14-13-29-24-11-13-12-11-14				3						1			3	1	8	5.2	0.013	16
14-13-29-23-11-13-13-11-14		3		1	1		1						1		7	4.5	0.011	3
14-13-30-24-11-13-13-11-14		3	1	2						1					7	3.8	0.011	71
14-13-29-24-10-13-13-11-15		1		1							4				6	3.8	0.010	6
15-12-29-22-10-11-14-14-14		1	2	1	1								1		6	3.6	0.010	41
13-13-30-24-10-11-13-16-18				3	2										5	3.4	0.008	5
14-13-29-24-11-13-13-11-15	1		1	1	1					1					5	3.2	0.008	4
14-13-29-24-11-14-13-11-14			2		1					2					5	3.1	0.008	19
15-13-29-24-11-13-13-11-14		1		1		2					1				5	3.1	0.008	22
13-13-30-23-10-11-13-16-18				1	2									1	4	2.8	0.006	48
14-13-29-24-11-13-14-11-14		1	2					1							4	2.8	0.006	65
14-13-29-25-11-13-13-11-14	1			2						1					4	2.8	0.006	7
15-12-28-23-10-11-12-13-17				1									3		4	2.8	0.006	#N/A
13-12-29-24-10-11-13-17-18				2				1							3	2.7	0.005	128
13-13-29-24-11-13-13-11-14			2									1			3	2.4	0.005	168
13-13-30-23-10-11-13-16-17				1	1	1									3	2.2	0.005	#N/A
13-13-30-24-09-11-15-15-17								3							3	2.0	0.005	#N/A
13-13-30-24-10-12-13-16-18		1			1			1							3	2.0	0.005	#N/A
14-12-28-23-11-11-12-13-17													3		3	2.0	0.005	#N/A
14-12-29-24-10-11-13-17-18		2	1												3	2.0	0.005	#N/A
14-13-28-24-10-13-13-11-14								2		1					3	2.0	0.005	27
14-13-28-24-11-13-13-11-14		1		1									1		3	2.0	0.005	344
14-13-29-23-10-13-13-11-14					1	2									3	2.0	0.005	9
14-13-29-23-11-13-13-11-15				2	1										3	2.0	0.005	13
14-13-30-25-11-13-13-11-14					3										3	2.0	0.005	#N/A
14-14-31-24-11-13-13-11-15						3									3	2.0	0.005	#N/A
15-12-28-24-10-11-12-14-17			1	1								1			3	2.0	0.005	83
Others	**20**	**63**	**66**	**69**	**50**	**29**	**34**	**36**	**9**	**28**	**16**	**26**	**19**	**16**	**481**	**448**	**0.762**	
Total	**22**	**87**	**86**	**98**	**65**	**40**	**38**	**44**	**9**	**38**	**21**	**29**	**35**	**19**	**631**		**1**	

Fig. 1. The most frequent ''minimal haplotypes'' in 14 local samples from central and northern Italy. #N/A: not present in the Italian database.

This further GEFI collaborative project was designed for studying the diversity of 17-locus Y-STR profiles, usually used in casework, on different Italian population groups, sampling by regional or local ways for a total of 1288 typed samples to determine individual loci gene diversity, multiplex discriminatory capacity and to increase data for reference database. In addition a total of 526 samples were typed for 12 loci by a few number of laboratories and the results were collected to increase the minimal haplotype Italian database.

2. Materials and methods

Participating laboratories were asked to type at least 100 unrelated individuals born in their region for 17 loci by Y filer kit. Blind control samples were prepared for each laboratory. Laboratories were left free to use their preferred DNA extraction methods. PCR and analysis of amplified products were performed according to the manufacturer's recommendations. In addition other 526 Italian samples were typed by the PowerPlex Y® System (Promega).

3. Results and discussion

Within an Italian collaborative exercise on the extended haplotype of the Y-chromosome, 1288 subjects were typed by the AmpFlSTR YFiler Amplification Kit (AB Applied Biosystems) and other 526 were typed by the PowerPlex Y® System (Promega). The sampling scheme included either a ''regional'' or a ''local'' recruitment, the first referring to individuals born in the region of the participating lab, the second referring to individuals coming from small villages. In the second case, only non-isonymous subjects were sampled. Fig. 1 shows the 9-locus haplotype counts in 631 individuals from 14 local samples compared with the expected counts in a sample of the same size if

Seventeen-locus haplotype	Bergamo	Brescia	Cuneo	Urbino	AscoliPiceno	Fabriano	Matelica	Buti	Col	Msb	Nor	Sezze	Total
14-13-29-24-10-13-13-11-15-14-13-12-18-15-18-23-12									4				4
14-14-31-24-11-13-13-11-15-14-12-12-19-16-18-27-12				3									3
13-13-29-24-11-13-13-11-14-15-12-11-19-15-16-23-12			2										2
13-13-30-25-10-11-13-17-17-14-10-11-20-15-15-24-12										2			2
13-13-30-25-11-11-13-17-17-14-10-13-20-15-17-22-12						1			1				2
13-13-32-22-11-11-13-17-17-14-10-12-20-15-14-21-13						2							2
14-12-28-23-10-11-13-13-13-16-10-10-20-14-15-23-10			2										2
14-12-28-23-10-11-13-13-14-16-10-12-20-14-15-21-11			2										2
14-12-28-23-11-11-12-13-17-14-09-12-21-14-19-21-11											2		2
14-13-28-24-10-13-13-11-14-15-11-11-19-17-16-23-12						2							2
14-13-29-24-11-13-12-11-14-15-12-13-19-15-15-23-12											2		2
14-13-29-24-11-13-13-10-13-15-12-12-19-15-18-23-12											2		2
14-13-29-24-11-13-13-11-14-15-12-14-20-16-17-23-12											2		2
14-13-29-24-11-13-13-11-14-15-13-13-19-16-18-23-11											2		2
14-13-30-23-10-11-12-13-15-15-09-12-20-17-17-21-11						2							2
15-12-28-23-10-11-12-13-17-16-09-12-19-13-17-23-11											2		2
15-12-29-22-10-11-14-13-17-15-11-10-21-15-16-20-12				2									2
15-14-30-23-10-11-12-12-16-15-09-11-20-15-16-23-11											2		2
Unique haplotypes	22	78	79	35	38	37	9	38	16	27	21	19	419
Total	**22**	**78**	**85**	**40**	**38**	**44**	**9**	**38**	**21**	**29**	**35**	**19**	**458**

Fig. 2. Non-unique 17-locus haplotypes in 12 local samples from northern and central Italy.

it were randomly sampled from the general Italian population. The expected counts were obtained by numerical resampling of the Italian database. For example, the first two most frequent haplotypes among the local samples display the highest rank in Italy also; however, their relative frequency is higher in the local samples (observed counts 38 vs. expected counts 21.8). In general, the haplotype frequency distribution is biased in the local samples towards a lower number of haplotypes with higher frequency. Fig. 2 shows the non-unique 17-locus haplotype counts in 12 local samples (two local samples were not typed with the 17-locus kit). It is remarkable that all non-unique haplotypes but one are present in a single local sample.

Conflict of interest

None.

References

[1] H. Roding, L. Roewer, A. Gross, T. Richter, P. de Knijff, M. Kaiser, W. Brabetz, Evaluation of haplotype discrimination capacity of 35 Y-chromosomal short tandem repeat loci, Forensic Sci. Int. 174 (2008) 182–188.

[2] S. Presciuttini, A. Caglià, M. Alù, A. Asmundo, L. Buscemi, L. Caenazzo, E. Carnevali, E. Carra, Z. De Battisti, F. De Stefano, R. Domenici, A. Piccinini, N. Resta, U. Ricci, V.L. Pascali, Y-chromosome haplotypes in Italy: the GEFI collaborative database, Forensic Sci. Int. 122 (2001) 184–188.

Forensic Science International: Genetics xxx (2012) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig

Letter to the Editor

Genetic variability of the SNP*for*ID 52-plex identification SNP panel in Italian population samples

Dear Editor,

The potential application of SNPs in place of supplementary STRs in paternity testing and forensic casework has been the subject of debate in recent years [1–4]. In fact SNPs show a range of characteristics that make them well suited to forensic analysis, such as low mutation rate, much reduced amplicons sizes and relatively simple multiplex assays [5–8].

Previously we characterized variation within Italy, studying two geographically separated populations from the north of Italy (Veneto) and the south (Calabria). In this study we update existing data analyzing more samples (200) from the same populations.

DNA was extracted from blood samples of healthy, unrelated volunteers that gave informed consent for population studies in accordance with Italian Law D. Lgs. 196/2003 and approved by SIMEF ISO-17025 procedures. DNA extractions were made with the Promega Wizard® DNA purification kit and quantification with the Applied Biosystems (AB) Quantifiler Human DNA Quantification Kit using an AB 7300 real-time PCR system.

The 52plex SNaPshot assay was applied, as previously described by Sanchez et al. validated for forensic applications [9]. Amplifications were made in a single PCR followed by two parallel 23- and 29-plex single base extension reactions (SBEs) using primers and reaction conditions described by Sanchez et al. [9]. Detection of the SBE products was performed by capillary electrophoresis on an AB Prism 3130 using GeneScan LIZ 120 for internal calibration.

Allele frequencies, forensic and statistical parameters are given in Supplementary Tables 1 and 2. No significant differences were found in comparison with our previous data already published in the SPSmart open-access online frequency browser (http://spsmart.cesga.es/snpforid.php?dataSet=snpforid52). Moreover, when comparing our data to other European populations (specifically, Spanish, Portuguese and Danish data), no overall significant differences were found for the same markers [12]. Comparison analysis is outlined in Table 3. The main differences for allelic distributions were found with Denmark in rs1335873, rs2046361, with Portugal in rs1357617 and with both populations in rs826472.

Statistical parameters of forensic interest were calculated, comprising: Dp: power of discrimination, PE: power of exclusion, RMP: random matching probability, using PowerStats v.1.2 software [10]. The SNPs showed low discrimination power (PD) and exclusion power (PE) when used individually, but in combination PD and PE were raised to 0.9999 in both populations, representing values comparable to those obtained analyzing a standard 15 STRs set. Hardy–Weinberg equilibrium and other population parameters were calculated using Arlequin software v.3.1 [11]. The highest average heterozygosity for both populations was found in rs2831700. No significant deviation from Hardy–Weinberg expectations was observed ($P > 0.05$).

The typical paternity index that can be expected applying these SNPs was assessed by calculating the average PI values obtained from three trio cases and, separately, from three deficient family studies (lacking the mother in each case) and the genotypes obtained are listed in full in Supplementary Tables 4a and 4b. Individual paternity indices were calculated in the standard way and a combined paternity index (CPI) determined as the product of these individual values [13]. From three combined PIs the average trio PI was 7.23E+10 and the average deficient family PI was 1.22E+8. With both values in accordance with previously published data [6].

In conclusion, all 52 SNPs were informative in the population samples analysed, this means they can provide valuable information not only for population studies but also for forensic applications (e.g. identification cases with highly degraded samples) as well as the analysis of complex pedigrees (e.g. distant relationships or incomplete pedigrees) as a complement to standard STRs typing.

The laboratory performing this study participates in the quality control/proficiency testing of the GEP-ISFG WG (www.gep-isfg.org). This paper follows the guidelines for publication of population data requested by the journal [14].

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.fsigen.2012.07.002.

References

[1] C. Phillips, M.V. Lareu, J. Sanchez, M. Brion, B. Sobrino, N. Morling, P. Schneider, D. Syndercombe Court, Á. Carracedo, Selecting single nucleotide polymorphisms for forensic applications, Progress Forensic Genet. 10 (2004) 18–20.
[2] J.M. Butler, M.D. Coble, P.M. Vallone, STRs vs. SNPs: thoughts on the future of forensic DNA testing, Forensic Sci. Med. Pathol. 3 (2007) 200–205.
[3] B. Budowle B, A. van Daal, Forensically relevant SNP classes, BioTechniques 44 (2008) 603–610.
[4] P. Gill, An assessment of the utility of single nucleotide polymorphisms (SNPs) for forensic purposes, Int. J. Legal Med. 114 (2001) 204–210.
[5] C. Phillips, M. Fondevila, M. García-Magariños, A. Rodriguez, A. Salas, Á. Carracedo, M.V. Lareu, Resolving relationship tests that show ambiguous STR results using autosomal SNPs as supplementary markers, Forensic Sci. Int. Genet. 2 (2008) 198–204.
[6] C. Børsting, J.J. Sanchez, H.E. Hansen, A.J. Hansen, H.Q. Bruun, N. Morling, Performance of the SNP*for*ID 52 SNP-plex assay in paternity testing, Forensic Sci. Int. Genet. 2 (2008) 292–300.
[7] M. Fondevila, C. Phillips, N. Naverán, M. Cerezo, A. Rodríguez, A. Salas, Á. Carracedo, M.V. Lareu, Identification of skeletal remains using short-amplicon marker analysis of severely degraded DNA extracted from a decomposed and charred femur, Forensic Sci. Int. Genet. 2 (2008) 212–218.
[8] L.A. Dixon, A.E. Dobbins, H.K. Pulker, J.M. Butler, P.M. Vallone, M.D. Coble, W. Parson, B. Berger, P. Grubwieser, H.S. Mogensen, N. Morling, K. Nielsen, J.J. Sanchez, E. Petkovski, Á. Carracedo, P. Sanchez-Diz, E. Ramos-Luisf, M. Brion,

1872-4973/$ – see front matter
http://dx.doi.org/10.1016/j.fsigen.2012.07.002

J.A. Irwin, R.S. Just, O. Loreille, T.J. Parsons, D. Syndercombe Court, H. Schmitter, B. Stradmann-Bellinghausen, K. Bender, P. Gill, Analysis of artificially degraded DNA using STRs and SNPs – results of a collaborative European (EDNAP) exercise, Forensic Sci. Int. 164 (2006) 33–44.

[9] J.J. Sanchez, C. Phillips, C. Borsting, K. Balogh, M. Bogus, M. Fondevila, C.D. Harrison, E. Musgrave-Brown, A. Salas, D. Syndercombe-Court, P.M. Schneider, Á. Carracedo, N. Morling, A multiplex assay with 52 single nucleotide polymorphisms for human identification, Electrophoresis 27 (2006) 1713–1724.

[10] A. Tereba, Tools for Analysis of Population Statistics Profiles in DNA, Promega Corp., 1999.

[11] L. Excoffier, G. Laval, S. Schneider, Arlequin ver. 3.0: an integrated software package for population genetics data analysis, Evol. Bioinform. (2005) 47–50.

[12] J. Amigo, C. Phillips, A. Salas, L. Fernandez Formoso, Á. Carracedo, M.V. Lareu, *pop.STR*—an online population frequency browser for established and new forensic STRs, Forensic Sci. Int. Genet. Suppl. Series 2 (2009) 361–362.

[13] D.W. Gjertson, C.H. Brenner, M.P. Baur, Á. Carracedo, F. Guidet, J.A. Luque, R. Lessig, W.R. Mayr, V.L. Pascali, M. Prinz, P.M. Schneider, N. Morling, ISFG: recommendations on biostatistics in paternity testing, Forensic Sci. Int. Genet. 1 (2007) 223–231.

[14] Á. Carracedo, J.M. Butler, L. Gusmao, W. Parson, L. Roewer, P.M. Schneider, Publication of population data for forensic purposes, Forensic Sci. Int. Genet. 4 (2010) 145–147.

Anna Barbaro[a,b,*]
[a]*Studio Indagini Mediche E Forensi (SIMEF), Reggio Calabria, Italy*
[b]*Institute of Legal Medicine, University of Santiago de Compostela, Spain*

Chris Phillips
Manuel Fondevila
Maviky Lareu
Ángel Carracedo
Institute of Legal Medicine, University of Santiago de Compostela, Spain

*Corresponding author at: Studio Indagini Mediche E Forensi (SIMEF), Reggio Calabria, Italy
E-mail address: simef_dna@tiscali.it (A. Barbaro)

8 March 2012

Please cite this article in press as: A. Barbaro, Genetic variability of the SNP*for*ID 52-plex identification SNP panel in Italian population samples, Forensic Sci. Int. Genet. (2012), http://dx.doi.org/10.1016/j.fsigen.2012.07.002

Chapter VI : GENERAL DISCUSSION

1. Introduction

Forensic labs have often to deal with the analysis of highly degraded DNA samples that can result in locus or allele dropout leading to complex interpretative problems. In cases where DNA evidence is limited, either in quantity or quality, such as highly degraded samples that are exposed to environmental insults or inhibitors, standard **STR testing** is often inadequate. Analysis of these compromised DNA samples often result in dropout of the larger STR loci from the samples and only a partial DNA profile can be obtained. Partial DNA profiles generally do not provide the power of discrimination to include or exclude a potential contributor to the sample.

Success with highly degraded DNA is improved using short amplicon mini-STRs. While standard STR primers target longer sequences that include the STR loci, mini-STR primers are redesigned so resulting DNA product is smaller, thereby increasing the chances of successful amplification of the larger loci.

The use of more robust loci, rather than already established STRs which frequently fail to give results, and/or have a poor power of discrimination increases the sensitivity of DNA detection and optimizes the opportunity to obtain a DNA profile from compromised samples, providing forensic scientists with a tool that captures genetic data from DNA samples of marginal and extremely low quality and quantity.

Thus, many previously unsolvable human identity cases may be resolved with mini-STR technology. Obviously before the introduction in routine casework analysis, it's relevant for the forensic community to establish which markers may be useful for catching up the procedure to a level acceptable for forensic application and to validate protocols with sufficient analysis repeat rates. Moreover in order to calculate the correct representative weight of DNA evidence, prior knowledge about the DNA markers for a relevant population sample is required. Important properties such as how frequently certain DNA-variants (i.e. alleles) occur in the population, the differences in such frequencies between populations and the forensic efficiency of the DNA markers

in casework should be studied to determine the probability that a particular genotype might occur at random in a population,

As previously discussed the aims of this thesis are:

- to validate a next generation pentaplex we developed including the new five loci recommended by the European Union Council for the expansion of the European Standard Set (ESS) evaluating the STR data informativeness and success rate on a wide range of forensic samples and to compare its performance with commercially available kits .

- to create a useful population database that includes an estimate of the frequency of each possible allele and genotype(or haplotype), by studying the variability in Mediterranean region of the well established 15 autosomal STRs together with the 5 new ESS, the highly discriminant SE33 and some sex linked STRs commonly used in forensics .

1.1 Validation of New STRs Multiplex

The EU-Council resolution 2009/C 296/01 calls upon European countries to use the European Standard Set (ESS) as a minimum to enable international comparison of DNA-profiles.The European Standard Set until recently contained only 7 loci. This was enough for occasional exchanges of DNA-profiles between countries.
However when massive exchanges of DNA-profiles are undertaken as has been made possible by the Interpol DNA-database and the Prüm Treaty, 7 loci are not enough because the chance of adventitious matches becomes unacceptable. In addition each DNA-database contains a significant portion of partial profiles with much higher probability to match randomly.

DISCUSSION

A decision was adopted by the ENFSI and EDNAP groups to increase the number of ESS loci and a recommendation was published to include more robust loci, rather than already established STRs which frequently fail to give results, and/or have a poor power of discrimination. In particular Europe adopted 5 new loci D2S441 D10S1248, D22S1045, D1S1656, and D12S391.

In the meantime, commercial companies have already produced kits which contain these new loci to enable the implementation of the new ESS loci. Also the FBI has recently published an intended expansion of the CODIS core loci set. (D.R. Hares (2012) *Expanding the CODIS Core Loci in the United States,* Forensic Sci. Int. Genet. 6: e52-e54,*Addendum to expanding the CODIS core loci in the United States,* Forensic Sci. Int. Genet. (2012) doi:10.1016/j.fsigen.2012.01.003)

We developed a new STR pentaplex including the five new loci recommended as next generation markers for the European Standard Set (ESS) : D12S391 and D1S1656, that are highly informative but with conventional amplicon lengths, plus three mini-STRs: D2S441, D10S1248 and D22S1045.

Space exists in this multiplex amongst the fragment sizes and green/yellow dye labels to allow additional STRs to be included in future. Furthermore we tested and optimized the resulting multiplex by genotyping 944 samples of the HGDP-CEPH .

Rare and intermediate alleles identified from the above population studies were further characterized by sequence analysis. This enabled the ladders we had previously constructed for D1S1656 and D12S391 to be enhanced with additional alleles while allelic ladders for the mini-STRs were completely built de-novo. Allele designations for D10S1248 were determined following the changes noted by Coble and Butler [14].

As part of the ESS pentaplex validation we assessed its ability to amplify DNA from a total of 111 casework DNA extracts that were considered to constitute more challenging material than is routinely profiled, including: extracts containing strong PCR inhibitors (e.g. blue denim); low level DNA sources (sweat stains, fingerprints, washed bloodstains); hairs; fingerprints with commonly used enhancers (ninhydrin, Luminol, Cyano, DSO); and decomposed tissue and bones or teeth samples with 5 to 20 years internment.

The aim was to thoroughly assess the performance of the ESS pentaplex assay as it was available to use as a complement to Identifiler or MiniFiler and we were able to explore its potential eight months before commercial kits had been released.

This allowed an assessment of the relative sensitivity of the mini-STRs compared to the longer amplicons and allele ranges of D1S1656 and D12S391. It was also useful to evaluate rates of allele/locus dropout and peak balance as well as the ability to differentiate single nucleotide repeat length differences in routine profiling.

This latter characteristic was particularly important for the reliable genotyping of D1S1656 and D12S391 where intermediate alleles with single nucleotide differences comprised a significant proportion of genotypes (D1S1656: 14.3-19.3 alleles = 0.254 combined frequency and 15-19 = 0.362; D12S391: 17.1/.3-20.1/.3 = 0.042 and 17-21 = 0.658).

A total of sixty challenging DNA samples were compared to AB Identifiler and short amplicon MiniFiler kits. Eleven of them were also analyzed with AB NGM that also includes the five ESS markers. Additional 51 challenging samples were analyzed with NGM and Identifiler in parallel with the ESS-pentaplex.

We recorded also individual locus dropout for the components of the ESS-pentaplex and in cases where reference genotypes were available for comparison (~80% of cases) no allele dropout was observed.

A genotyping success for the ESS pentaplex was observed in 97 challenging casework samples in comparison with other three commercial STR kits.

To better compare performance of different multiplexes we charted casework profile completeness in heatmaps ordered best to worst, left to right and with percentage loci present (going from hot to cold colours) as: 100%; 80-99%; 60-79%; 40-59%; 20-39%; up to 20% and no profile.

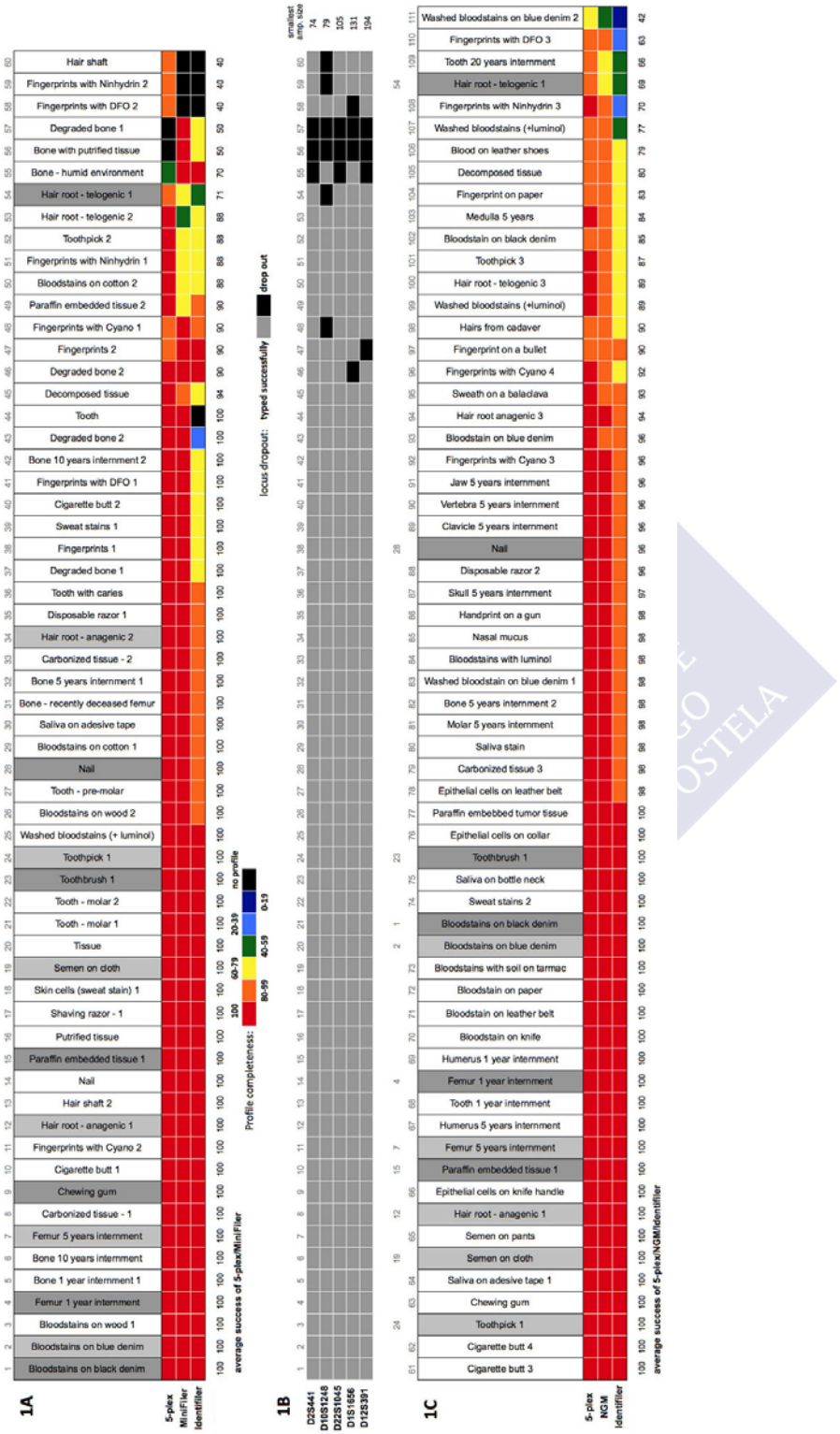


Fig.1: *Genotyping performance summarized as a graded heatmap aligning casework DNA condition from best to worst, left to right.*

The ESS pentaplex is directly comparable in performance to MiniFiler with 93% vs. 92% average profile completeness respectively with both multiplexes giving fifty full profiles from the challenging casework samples analyzed. This compares to ~81% average profile completeness observed using Identifiler on the same extracts and 28 full profiles achieved.

When the individual performance of component STRs in the ESS-pentaplex was examined no allele dropout was observed when controlled by comparison to reference profiles in each case. Only D10S1248, the second shortest STR in ESS-pentaplex, gives slightly less success than the others with six dropouts (90% success) compared to 3 or 4 in the others (93 or 95%). Therefore we did not detect a strong effect of amplicon size amongst the five ESS-pentaplex components and the so-called midi systems of D1S1656 and D12S391 perform as well as the three mini-STRs in the analysis of very challenging casework material. We observed very similar performance in both multiplexes, so the five ESS STRs worked equally well in both assays for any one casework sample. The comparison of NGM and ESS-pentaplex performance indicated similar results to the parallel typing with MiniFiler but the ESS-pentaplex genotyping was noticeably more sensitive than NGM with six more complete profiles (52 vs. 44 full profiles = 97.3% success vs. 95.3%) even if with a much smaller number of STRs in the multiplex. With this second range of casework extracts Identifiler performed better overall (average 86.7% success) but had slightly fewer complete profiles. Excluding cases with full profiles in all multiplexes as likely in better condition than the exhibits initially suggest (i.e. when samples giving partial profiles for some or all multiplexes are considered) the average profile completeness shows marginally more contrast between the ESS-pentaplex and MiniFiler or NGM with 88.6% vs. 85.7% for MiniFiler and 95.3 vs. 91.9% for NGM.

Again, indicating a marginal but consistent improvement in performance using the ESS-pentaplex compared to both MiniFiler and NGM.

So in routine forensic use the ESS-pentaplex provided a valuable additional approach for the analysis of challenging DNA, even when some standard STRs in commercial kits failed or were too weak. The examination of a full range of more than 100 different casework DNA extracts provides a reasonably comprehensive survey of the

expected performance of different STR multiplexes applied to the analysis of the most challenging forensic material.

It can be expected that well-balanced and commercially produced STR multiplex kits will provide the optimum performance but our results suggest it is also likely that the properties of the STRs themselves are directly responsible for the performance of a multiplex when analyzing scant or highly degraded material.

Although we originally intended to build population frequency data only, ahead of the release of kits containing the five new STRs, when we used the ESS-pentaplex alongside MiniFiler and Identifiler to expand the profile completeness obtained from challenging material we found the success rate and therefore the informativeness of the STR data was enhanced considerably. Running the ESS pentaplex in parallel to NGM containing the same STRs in a 15-plex suggests slightly better performance from the ESS-pentaplex most likely as a result of a much smaller multiplex allowing improved chances of successful amplification in a less competitive PCR.

In conclusion we recommend use of small-scale STR multiplexes based on well founded primer designs as an informative and robust adjunct to MiniFiler or Identifiler, thereby improving performance and informativeness while keeping a laboratory's existing STR protocols and data intact while transitioning to new multiplex combinations. This has relevance for those laboratories currently using CODIS STRs and seeking to assess new loci under consideration for an expanded CODIS combination in the near future.

Since during our validation study a next generation kit NGMSElect containing the above 5 ESS plus the highly discriminating SE33, was released in commerce by Applied Biosystems, an internal validation study was carried out in order to verify the kit real performance because it has been expected that commercial STR multiplex show well-balance and better performance especially on "difficult" samples.

Some critical parameters such as species specificity, sensitivity, degradation/inhibition, mixture sample analysis, performance on a wide variety of real forensic samples were evaluated.Very often DNA labs have to deal with samples that are degraded or in small traces. DNA quantity affects typing results: too much DNA can result in off scale data

and incomplete A nucleotide addition while extremely low quantity can produce unbalanced amplification.

Sensitivity study was carried out testing serial DNA dilutions in replicates and evaluating the number of alleles detected, intra-colour balance and heterozygote balance. Observed full profiles were reliable till to 0.016 ng of input DNA.

Inhibition study performed with variable concentration of inhibitors showed reliable results full profiles obtained till to the highest concentrations of inhibitor tested. (250 uM hematin). It's known the average size of degraded DNA approaches the size of the target sequence, the amount of PCR product generated is reduced because of the reduced number of intact templates in the size range necessary for amplification.

Degradation Study showed that the longer loci gradually disappear as the amount of DNase I increases but the 3 new miniSTR (D10S1248, D22S1045 and D2S441) amplify successfully even at 6U DNase.

Species-specificity study performed on DNA from 27 animal species showed Chimpanzee and Gorilla DNA samples produced partial profiles, while Macaque DNA produced a strong Amelogenin-X peak and two small out-of-marker-range peaks in PET. Among non-primates, only Horse DNA produced a 96-bp fragment near the Amelogenin locus in the VIC® dye. The other animals did not yield detectable products. This is relevant because often nonhuman DNA may be present in forensic casework samples and it confirms the specie specificity of primers used in the kit .

Since forensic casework samples may contain DNA from more than one individual, therefore, it is essential to ensure that the DNA typing system is able to detect DNA mixtures. With NGM SElect detection of full profiles for the minor contributor was possible till to ratio 1:10, while 1:15 ratios resulted in partial profiles for the minor component.The ability to obtain results from DNA recovered from biological samples deposited on various substrates and subjected to various environmental and chemical insults has been documented analyzing a wide variety of casework samples (such as blood,saliva,sperm stains, washed bloodstains, cadaveric tissues, bones, teeth, prints, sweat).Genotyping performance of NGM Select™ and Identifiler™ has been compared in 20 casework challenging samples. For a better evaluation Results are summarized in a heatmap showing profile completeness ordered, left to right, best to worst (Fig.2).

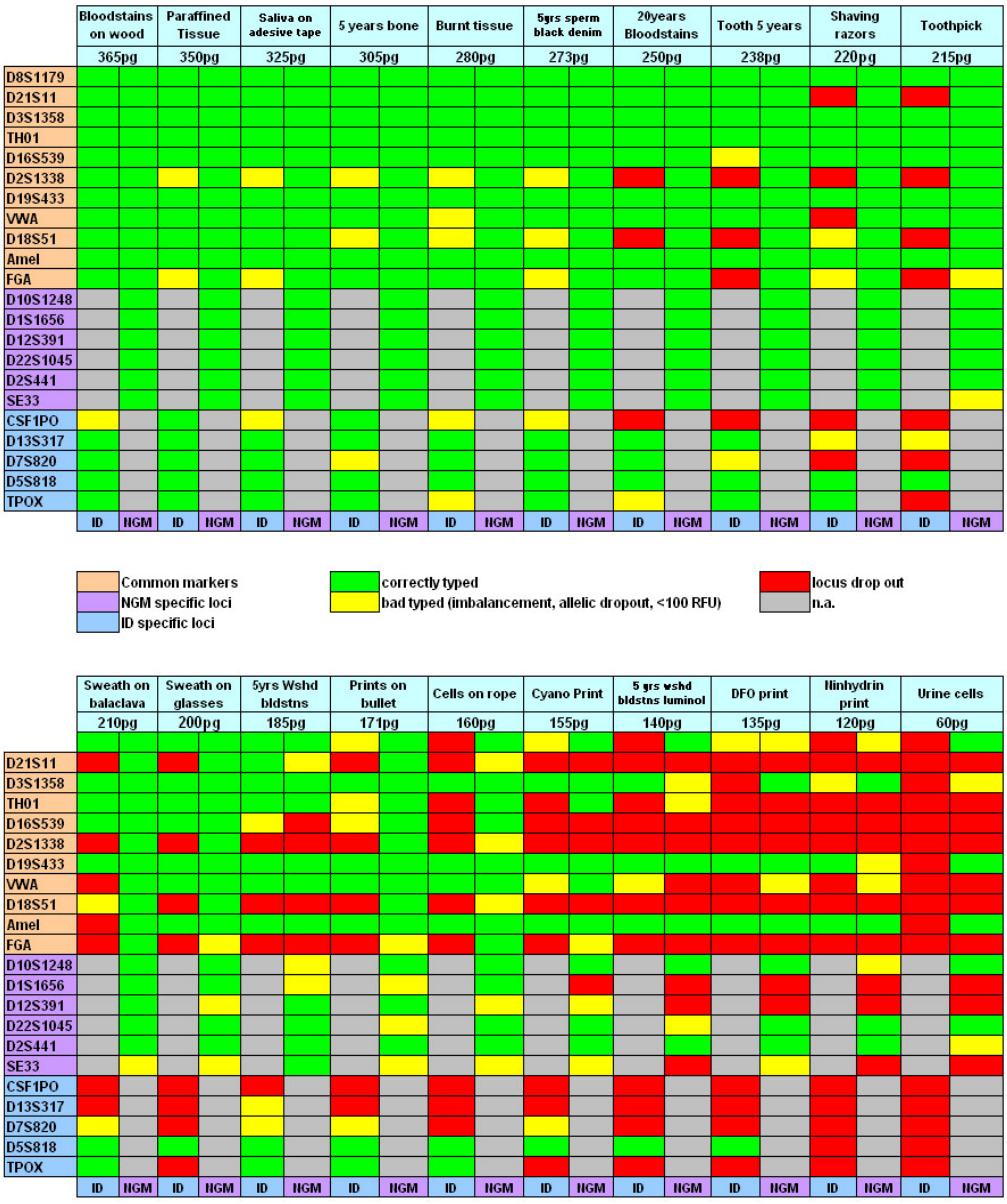


Fig.2 : *Heatmap showing profiles completeness*

NGM SElect genotyping on challenging samples was more sensitive than Identifiler with 7 more complete profiles (81,76 % success vs. 42,5%).

Samples with low DNA (<100pg) produced no profiles or very little genotyping information with Identifiler™ kit, while they gave successful amplification for some loci by NGM . Therefore, even this partial NGM™ kit profile were informative because include the 5 ESS new loci. Different kind of samples at almost the same DNA concentration showed different typing success.This means the nature of the evidence and its storing condition (i.e. environmental factors) have a big impact on final results.

The individual performance of each STR in both kits are examined and the average rate of success for each locus is reported in Fig.2: CSF1PO, according to its size, showed the lowest success (20%), while the 2 mini D22S1045 and D2S441 were the most successful loci (97,5%).

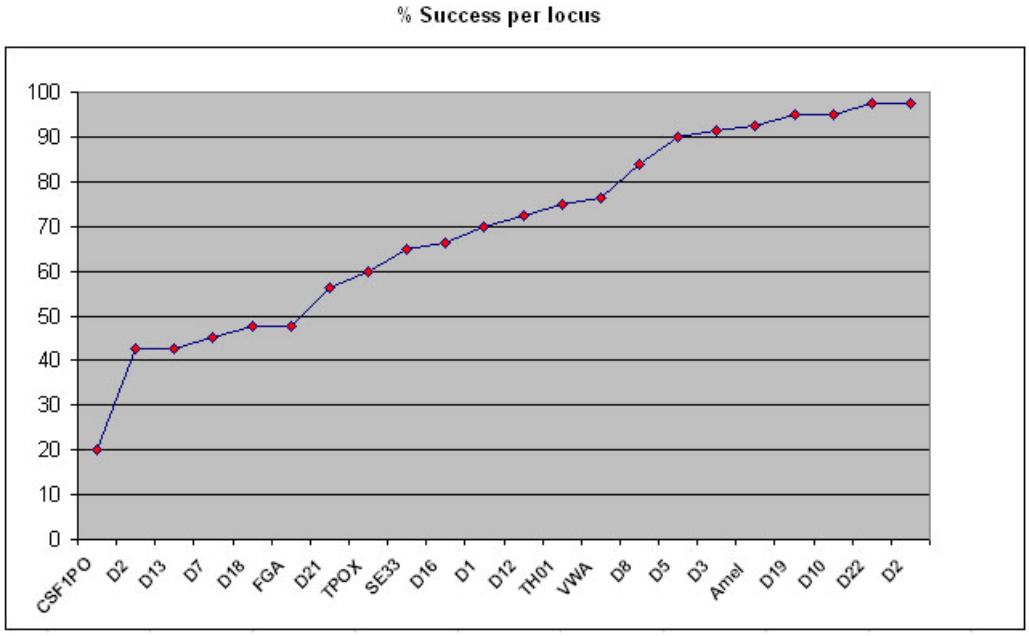


Fig.3: *average rate of success for each locus*

In conclusion results of our validation study demonstrate that NGM SElect™ kit is a reliable multiplex well suited for typing a wide variety of forensic samples.

It shows improved performances, especially in regards to its sensitivity and greater tolerance to high levels of PCR inhibitors, allowing maximum recovery of information from difficult samples,producing useful data even when working with very few DNA. STR profiles by NGM™ were generally better balanced than the Identifiler™ showing clear baseline, less noise and no PCR artefacts. This confirms NGM multiplex shows a robust PCR chemistry an the improved performance requested by the forensic community for challenging casework samples as well as paternity testing and the utility of the new 5 ESS loci analysis.

1.1 Population studies for forensic statistical evaluations

DNA forensic scientists are presented with the situation where they are given two samples related to a crime scene, about which they know nothing in advance, and are asked whether or not they are identical. Only one-tenth of a single percent of DNA (about 3 million bases) differs from one person to the next. These variable regions are used to generate a DNA profile of an individual, using samples from blood, bone, hair, and other body tissues and products. In criminal cases, this generally involves obtaining samples from crime-scene evidence and a suspect, extracting the DNA, and analyzing it for the presence of a set of specific DNA regions (markers).

DNA profiles are compared to determine whether the suspect's sample matches the evidence sample found at crime scene. If two DNA samples are alike at some regions, odds are great that the samples are from the same person. If the sample profiles don't match, the person did not contribute the DNA at the crime scene. If the patterns match, the suspect may have contributed the evidence sample.

When a comparison of DNA profiles derived from evidence and reference samples fails to exclude an individual(s) as a contributor(s) of the evidence sample, statistical assessment and/or probabilistic reasoning are used to evaluate the significance of the association. Traditionally, DNA investigations mainly involve two steps; the establishment of DNA profiles from biological samples and the interpretation of the evidential weight given by theses DNA profiles.

Prior to the introduction of new DNA markers into real forensic caseworks, in order to calculate the correct representative weight of DNA evidence, a knowledge about the DNA markers for a relevant population sample is required.

Important properties that should be studied are, for example, how frequently certain DNA-variants (i.e. alleles) occur in the population, the differences in such frequencies between subpopulations, in order to determine the probability that a particular genotype might occur at random in a population. Another feature to consider is the forensic efficiency of using the DNA markers in casework involving criminal cases and relationship testing. Such estimates describe the theoretical value of using the specific markers for different forensic genetic situations and differ from case specific values. The estimation of such parameters is most often based on the number of distinctive alleles found in the population and their corresponding frequencies.

This means before a new locus can be introduced in the forensic current practice a database for the relevant population must be established to evaluate its effectiveness (Brinkmann B. (1996). Usually a sample size of greater than 100 samples is sufficient to make reliable projections about a genotype's frequency in a larger population (see Chakraborty, R. (1992)..

In this perspective with the aim to establish a useful database we studied the variability in Mediterranean area of the well established 15 autosomal STRs together with the 5 new ESS and some sex linked STRs commonly used in forensics. Moreover it was evaluated the variability of the 52 SNP plex recently introduced for forensic applications.

a) For the 15 autosomal STRs as a point of reference, allele frequency data for combined European populations from the HGDP-CEPH human diversity panel were obtained using the pop.STR database.

We observed that certain alleles are present at low frequency in CEPH European populations but not found in the Galician (NW Spain) population studied.

These are: CSF1PO Allele: 15; D10S1248 18; D16S539 15; D19S433 13.2; D21S11 35.2; D22S1045 9; D2S1338 11; D2S441 8, 9 and 13.3; FGA 16, 20.2 and 23.2; TH01 11, and vWA 13. In contrast, three alleles were observed uniquely in the Galician

population in STR D19S433: repeats 13.2, 20 and 23. Observed heterozygosity is above 0.650 in all STRs in both the European group and the Galician study population except for TPOX that has a value of 0.647 in the Europe population group.

The most informative new STR in the Galician population is D12S391 with a discrimination index of 0.900, near identical to the most informative one in the European population group, D1S1656, that gives a discrimination index of 0.898 and that is also the most informative in Calabria (Sothern Italy) with a value of 0.974.

A population comparison made between Galicia and the Italian (Sothern Italy) samples showed allele frequencies from these two southern European populations were very similar for each of the STRs studied. Moreover no significant differences were found between Calabria and other European population data. No significant deviations from Hardy–Weinberg expectations were found ($p > 0.05$). In all STRs except TPOX the observed heterozygosity was greater than 0.7, with the highest value in D1S1656. With the exception of D12S391 individual STRs showed a low exclusion power (PE) but the combined PE reached 0.99999999. Combined RMP using 20 loci was calculated to be 4.47×10^{-24}, therefore used together these twenty loci can distinguish samples with a probability of 99.99999%.

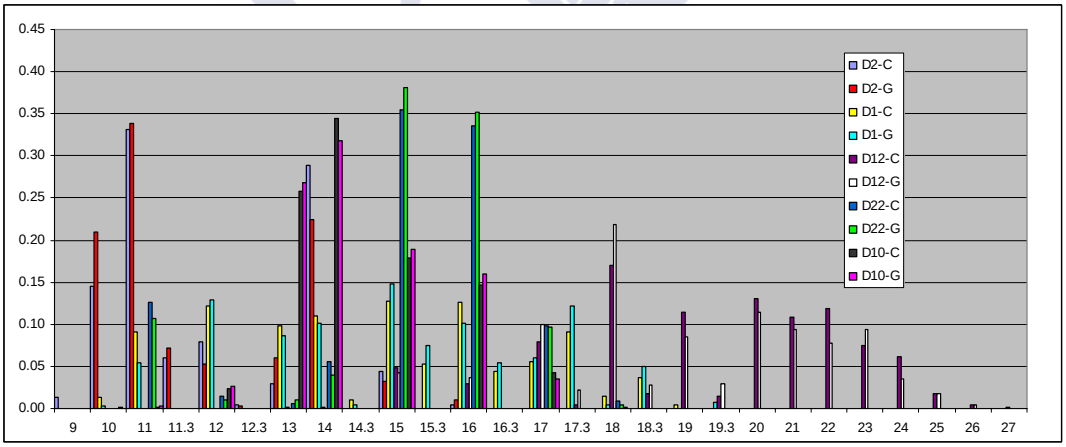


Fig.1 5ESS loci *variability in Calabria (South Italy) and Galicia (NW Spain)*

b) The SE33 (ACTP2—human actin beta-actin-related pseudogene H-beta-Ac-psi-2) is one of the most informative STR systems for biological identification that is routinely used in several central European countries and in particular for data exchange between databases.

Since the practical application of a polymorphic marker requires the availability of a suitable profile database from the reference population, SE33 variability was studied in 2 Mediterranean populations (Calabria and Malta) .

A total of 41 different alleles were observed in the 2 examined populations with no allele being more frequent than 10,5%. In the Maltese population more intermediate alleles than in Calabria were found; moreover 3 out of ladder alleles (17.3, 18.3, 20.3) were present. The high number of alleles observed at SE33 locus confirmed the high degree of polymorphism. No significant deviations from Hardy–Weinberg equilibrium were found.Based on heterozygosity (greater than 0.7) and polymorphic information content PIC (greater than 0.9), SE33 could be considered as an informative locus in both populations.

Even if it shows a low exclusion power (PE) degree when used individually, however combined PE with the other 5 ESS new STRs is increased to 0.999 in both populations and the combined power of discrimination (PD) was 0.999999. This means when used together these loci can distinguish samples from different individuals with a probability of 99,9999%.

Allelic frequencies of SE33 were compared to other previously published population data and no significant differences were found.

When comparing with Sicily no overall significant genetic distances were found, while comparison to other populations showed significant ones. Moreover comparison with non-European population showed no big distances between Germany and Morocco and between Hungary and Turkey.

Results confirmed the locus is effectively highly polymorphic and should be routinely added to the set of STRs loci commonly studied in caseworks and in paternity cases using as reference the database we generated.

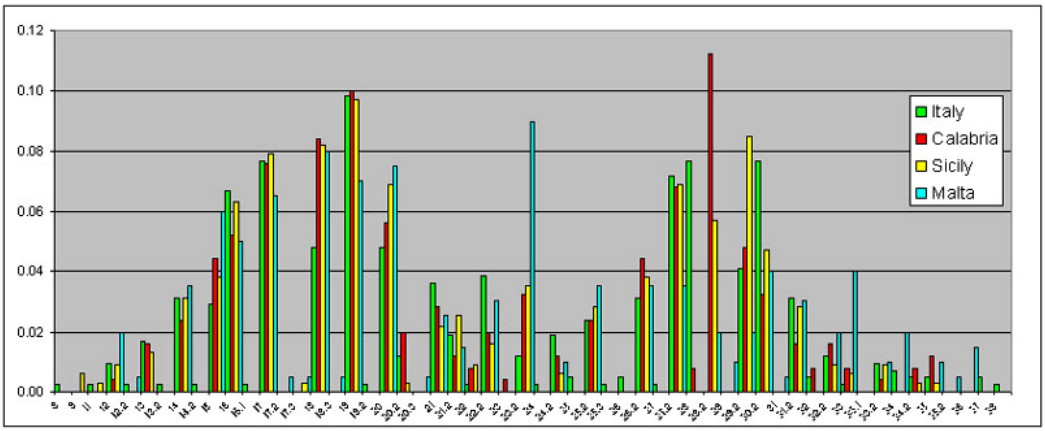


Fig.2 *SE variability in 4 close populations*

c) X-chromosome markers may complement the results obtained from other genetic markers in particular in complex relationship cases.

Consider, for example, a case where two sisters are tested to establish whether or not they have the same father, and where DNA profiles are only available for the sisters. In such instances, autosomal DNA markers cannot exclude paternity, since two sisters can inherit different alleles despite being full siblings. The use of X-chromosome markers can, however, exclude paternity, since two sisters would share the same paternal allele if they have the same father.

In this perspective we studied eight X-chromosomal markers in terms of their informativeness and usefulness in relationship testing. In particular four X-STR duos in the linkage groups 1-4 (DXS10135-DXS8378,DXS7132-DXS10074,HPRTB-DXS10101 and DXS10134-DXS7423) were tested in a population sample from Southern Italy (Calabria).

Each of the four STR clusters spans less than 0.5 cM and the genetic distance between linkage groups is more than 50 cM so the clusters represent stable haplotypes that can be treated as unlinked, thereby providing highly informative tools for kinship testing.

In all STRs except DXS8378 the observed heterozygosity was greater than 0.7, with the highest degree (0.9128) in DXS10135.

DXS10135 had the highest PIC value at 0.9062 while DXS8378 had the lowest at 0.6138. together these loci can distinguish samples with a probability of 99,995090 %.

Allelic frequencies for all 8 STRs were compared to previously published population data and no significant differences were found; in addition when comparing to other European samples, no overall significant genetic distances were found.
The only significant p-value was obtained between Calabria and North Italy at DXS10074 locus and DXS8378 while in Germany at DXS10074 locus.
Comparison to other populations showed significant genetic distances in Morocco at DXS7132 and Korea in all the studied markers with the exception of DXS10135.

The power of discrimination (PD) value of ChrX markers varies depending on the sex. In all loci, the PDF value was higher than the PDM value. The DXS10135 locus had the highest PDF value at 0.9858 The locus with the lowest PDF value was DXS8378. Consequently in female combined RMP using 8 loci was calculated to be $4.06x10^{-8}$, therefore used together these twenty loci can distinguish samples with a probability of 99,999996 %. In male combined RMP using 8 loci was calculated to be $4.91x10^{-5}$, therefore used.

If female traces are to be assigned to female individuals, ChrX markers yield the same results as AS. For matching male traces to male suspects, the PD value of ChrX markers is generally lower than that of AS markers. This is due to the fact that for male ChrX analysis only one allele per STR is used.
In a mixed female/male stain, the chance of having all X male alleles included in the female component is higher forChrX than for AS markers. Therefore, it is not advisable to use ChrX markers for testing male traces in a female background.
In order to identify female traces in male contamination, however, ChrX markers are more efficient than AS marker since only if the female component coincidentally happens to be homozygous at all loci then female component is not distinguishable from the male one and this is really unlikely.

d) The Y-chromosome normally exists in one copy in males and is absent in females.
Y-STR haplotyping is particularly important for sensitive typing of male DNA in mixed stains as well as for rapid assortment of biological crime scene evidence.

Since it is inherited from father to son, thus all men in a paternal lineage share an identical Y-chromosome, so Y chromosomal profiling can trace back paternal lineages into the past and has thus been proven a useful tool in genealogical and kinship testing. The individuality of the male-specific part of the Y chromosome can be optimally explored by the Y-STR haplotype analysis using a set of highly variable STRs markers approved by the forensic and scientific community. The minimal haplotype (MHL), consists of nine STR loci well characterised and recommended for court use.

This core haplotype can be extended by other hypervariable Y-STR loci, combining a number of the most polymorphic markers in a YSTR multiplex to increase the power of discrimination.

Apart from the recombination region (~5%), mutation is the only force that leads to new variation on the Y chromosome. Due to this and the fact that the Y-chromosome has one-fourth of the relative population size compared with autosomal loci, the Y chromosomal variation has been found to be fairly population specific (Hammer et al., 2003; Jobling et al.,2004). As a result, regional Y-STR haplotype database population databases must be collected.

With this aim, a total of 554 males from seven Western Mediterranean populations were genotyped for 12 Y-chromosome STR loci: three from East Spain (Ibiza, Majorca, Valencia) and four from South Italy (Catanzaro, Cosenza, Reggio Calabria, Sicily). Alleles not included in the allelic ladder were found for two loci: DYS19 and DYS438. The allele 7 and 13 for DYS438 have been previously described, but not the allele 9 for the locus DYS19.

In DYS392, allele 13 was the most common amongst the Spanish populations and allele 11 the most frequent in Southern Italy. These results are consistent with previous studies showing a longitudinal decrease of frequencies from the west to the east of the European landscape for 13 and conversely, a decrease in the opposite direction for 11. The Neolithic demic diffusion could explain these two opposite patterns, with the DYS392-13 allele present in the proto-European gene pool.

The frequency pattern observed in the DYS438 system, with DYS438-12 was the most frequent in the Spanish and Sicilian populations and DYS438-10 the most frequent in Calabria, could also be due to the same Neolithic effect.

Gene diversities ranged from around 0.85 (in DYS385) to approximately 0.50 (in DYS392). Generally, Italian populations had higher gene diversities than the Spanish populations, following the same pattern found in bi-allelic Y-chromosome markers, with an increasing diversity trend from Spain to Greece, maybe due to the impact of the arrival of haplotypes in Europe from the Middle East.

The use of Y-chromosome markers does not follow the basic rules of Mendelian genetics. The statistical assessment can still be applied when comparing the DNA profiles obtained from forensic evidence to that from a known individual.

The common approach is simply to state the number of occurrences of a Y-STR type present in the database. This is commonly referred to as the counting method.

Among the 554 males analyzed, 443 different haplotypes were obtained, of which 372 were unique. The other haplotypes were shared by two to seven men; 51 were present in 1 population, 17 in 2, 2 in 3 and 1 was found in 4 populations.

The most frequent haplotypes for markers DYS19,DYS389I, DYS389II, DYS390, DYS391, DYS392,DYS393, DYS385 a/b, DYS437, DYS438 and DYS439 were 15–12–29–22–10–11–14–14,14–16–10–12 and 15–13–29–24–11–13–13–11,14–14–12–13, both found in seven men from the Ibiza population but absent from the other studied populations.

These haplotypes (without locus DYS437) matched with three and six samples of European origin, respectively, in the YHR worldwide database of 38,761 haplotypes.

It is noteworthy that 46 haplotypes (almost all from Valencia, the Balearic Islands and Sicily) matched with north African or African samples. This result is concordant with other studies showing African influences in the Mediterranean populations.

The discrimination capacity ranged from 87.63% (Reggio Calabria) to 94.92% (Valencia) except in the Ibizan population (56.25%). These results are in accordance with the historical and demographic data of the island population (an isolated, consanguineous population with a reduced effective population size)

In conclusion, the results of the present study provide a useful Y-STR haplotype dataset, for the western Mediterranean region, that can be used in the forensic field to correctly weigh the value of the evidence of a Y profile match. Special

care should be taken in male identification in the Ibiza population, due to the very low discrimination capacity found for the 12 Y-STR loci

Further study on different Italian population groups for a total of 1814 (1288 typed for 17 loci and 526 samples were typed for 12 loci) was performed to determine individual loci gene diversity and to increase data for Italian reference database.

It has been observed that the 2 most frequencies haplotypes in the Ibiza population were also diffused in North and Central Italy.

	Northern Italy					Central Italy											
Nine-locus haplotype	Bergamo	Brescia	Cuneo	Rimini	Valmarecchia	Urbino	AscoliPiceno	Fabriano	Matelica	Buti	Col	Msb	Nor	Sezze	Total counts	Expected counts	Freq.	Rank in Italy
14-13-29-24-11-13-13-11-14		6	4	3		3	1			2		1	4	1	25	15.6	0.040	1
14-13-29-24-10-13-13-11-14		4	4	2			2			1					13	6.2	0.021	2
14-13-29-24-11-13-12-11-14				3						1			3	1	8	5.2	0.013	16
14-13-29-23-11-13-13-11-14		3		1	1		1						1		7	4.5	0.011	3
14-13-30-24-11-13-13-11-14		3	1	2						1					7	3.8	0.011	71
14-13-29-24-10-13-13-11-15		1		1							4				6	3.8	0.010	6
15-12-29-22-10-11-14-14-14		1	2	1	1								1		6	3.6	0.010	41
13-13-30-24-10-11-13-16-18				3	2										5	3.4	0.008	5
14-13-29-24-11-13-13-11-15	1		1	1	1					1					5	3.2	0.008	4
14-13-29-24-11-14-13-11-14			2		1					2					5	3.1	0.008	19
15-13-29-24-11-13-13-11-14		1		1		2					1				5	3.1	0.008	22
13-13-30-23-10-11-13-16-18				1	2									1	4	2.8	0.006	48
14-13-29-24-11-13-14-11-14		1	2					1							4	2.8	0.006	65
14-13-29-25-11-13-13-11-14	1			2						1					4	2.8	0.006	7
15-12-28-23-10-11-12-13-17				1									3		4	2.8	0.006	#N/A
13-12-29-24-10-11-13-17-18				2				1							3	2.7	0.005	128
13-13-29-24-11-13-13-11-14			2									1			3	2.4	0.005	168
13-13-30-23-10-11-13-16-17				1	1	1									3	2.2	0.005	#N/A
13-13-30-24-09-11-15-15-17								3							3	2.0	0.005	#N/A
13-13-30-24-10-12-13-16-18		1			1			1							3	2.0	0.005	#N/A
14-12-28-23-11-11-12-13-17													3		3	2.0	0.005	#N/A
14-12-29-24-10-11-13-17-18		2	1												3	2.0	0.005	#N/A
14-13-28-24-10-13-13-11-14								2		1					3	2.0	0.005	27
14-13-28-24-11-13-13-11-14		1		1									1		3	2.0	0.005	344
14-13-29-23-10-13-13-11-14					1	2									3	2.0	0.005	9
14-13-29-23-11-13-13-11-15				2	1										3	2.0	0.005	13
14-13-30-25-11-13-13-11-14					3										3	2.0	0.005	#N/A
14-14-31-24-11-13-13-11-15						3									3	2.0	0.005	#N/A
15-12-28-24-10-11-12-14-17			1	1								1			3	2.0	0.005	83
Others	**20**	**63**	**66**	**69**	**50**	**29**	**34**	**36**	**9**	**28**	**16**	**26**	**19**	**16**	**481**	**448**	**0.762**	
Total	**22**	**87**	**86**	**98**	**65**	**40**	**38**	**44**	**9**	**38**	**21**	**29**	**35**	**19**	**631**		**1**	

Fig.3:*Most frequent 'minimal haplotypes' in 14 local samples from central and northern Italy*

e) The use of single nucleotide polymorphisms (SNPs) in forensic genetic caseworks has been widely discussed in recent years, mainly because SNPs have three important advantages compared to short tandem repeats (STRs): (1) low mutation rates, (2) short amplicons sizes (<100 bp) and (3) simple multiplex assays. In addition, SNPs may

become an important investigative tool for the police when SNP packages for physical characteristics or ethnic origin are fully developed and validated.

The potential application of SNPs in place of supplementary STRs in paternity testing and forensic casework has been the subject of debate in recent years.

In paternity testing the genetic profiles of the individuals are used to compare the relative likelihoods of the alleged father and the child being related as father/offspring against, usually, being unrelated. In the great majority of the cases, analyses with the widely used sets of short tandem repeat markers (STRs) provide powerful statistical evidence favoring one of the alternative hypotheses. Nevertheless, there are situations where the final statistical result is ambiguous. In these cases, the possibility that the alleged father is actually a close relative of the real one (son, father or brother) can reasonably be raised. In such cases, when the statistical evidence obtained is considered as insufficient, the common practice is to extend the set of analyzed markers. In this context, many authors have suggested that biallelic markers, such as single nucleotide (SNP) polymorphisms, are markers of choice, as they are less prone to mutation than STRs.

The low mutation rate of SNPs (typically: 2.5×10^{-8} compared with 10^{-3} to 10^{-4} in STRs) is a major advantage in relationship testing. A mutation will often lead to a genotype that is incompatible with Mendel's law of inheritance, and the results may indicate that more than one type of relationship between the tested individuals is possible. This is especially true for immigration cases, where the tested individuals are often related and a mutation may result in comparable likelihoods for different family scenarios. However, a locus with a low mutation rate is also a locus with few alleles, (in multi allele systems) and SNPs are much less polymorphic than STRs. This means more SNPs are required to reach the same level of discrimination obtained with 13-15 STR loci.

Some SNPs that are polymorphic in one population may be almost or completely monomorphic in another populations, while others are known to be polymorphic in all major population groups.Because allele frequencies can vary greatly among populations, the population genetics of match probabilities is a critical issue.

So before a new locus can be introduced in the forensic current practice a database for the relevant population must be established to evaluate its effectiveness.

With this aim we characterized variation within Italy of the 52 SNP-plex assay developed by the SNPforID consortium by studying two geographically separated populations from the north of Italy (Veneto) and the south (Calabria).

The highest average heterozygosity for both populations was found in rs2831700 .

Moreover when comparing to other European populations (Spanish, Portuguese, Danish) data, no overall significant differences were found for the same markers according to previous accumulated data showing that allele frequencies tend to be similar in geographically close populations. However main differences for allelic distributions have been found with Denmark in rs1335873, rs2046361, with Portugal in rs1357617 and with both populations in rs826472.

The performance of the 52plex assay in paternity cases was tested in 3 trios and 3 motherless cases simulation. Individual PIs are likelihood ratios which compare the alleged father's likelihood of contributing the obligate paternal allele for any SNP locus to that of a composite randomly chosen man of his race. The combined paternity index (CPI) was determined as the product of all the individual paternity indices and values were ranged between 10^6-10^{11} in accordance to previously published data.

All 52 SNPs were informative in the population samples analysed, this means they can provide valuable information not only for population studies but also for forensic application (e.g high degraded samples) as well as the implementation of criminal DNA databases or to solve complex pedigree (e.g. distant relationships or incomplete pedigrees) as a complement to standard STRs methodology.

Match probability and the typical paternity indices calculated on basis of the 52 SNP-plex would be satisfactory in both crime and paternity cases.

The application of SNPs in relationship testing has not been widespread to date because nearly all paternity cases are adequately resolved with existing well validated STR sets but since a consistent proportion of relationship tests can involve analysis of human remains, SNPs small size provides an important way to avoid a further source ambiguous results from incomplete STRs profiles commonly obtained from degraded DNA.

Chapter VII

CONCLUSIONS

1. Validation of New STRs Multiplex

1.1 The examination of a full range of more than 100 different casework DNA extracts provides a reasonably comprehensive survey of the expected performance of different STR multiplexes applied to the analysis of the most challenging forensic material.
We found the properties of the STRs themselves are directly responsible for the performance of a multiplex when analyzing scant or highly degraded material.
Because of this, we recommend the use of small-scale STR multiplexes based on well founded primer designs as an informative and robust adjunct for improving performance and informativeness when traditional STRs typing of challenging samples fails or give partial results.
1.2 The use of commercial next generation multiplexes containing ESS combinations can assure a drastic increase of the information providing a valuable additional approach for the analysis of challenging DNA and has relevance for those laboratories currently using CODIS STRs and seeking to assess new loci in high sensitive multiplex .
1.3 In summary we can conclude that the new 5ESS loci have been validated successfully so that they provide a valuable solution for the forensic analysis and especially for those complex cases where the high degradation of the sample sensitively reduce the information degree obtained with the traditional STRs multiplexes currently in use.

2. Population studies for forensic statistical evaluations

2.1 A database for some Mediterranean populations was set up for further reliable statistical analyses.

CONCLUSIONS

In particular studies were performed on the relevant populations in order to evaluate the effectiveness of markers routinely used and especially of the new ones recently introduced in the forensic current practice.

2.2 A population comparison made between Galicia (NW Spain) and Calabria (Southern Italy) samples for 15 autosomal STRs and the new 5 ESS markers showed similar allele frequencies for the STRs studied. With respect to the well established 15STRs, locus TPOX confirmed to be the less informative one, while D12S391 and D1S1656, are the most discriminative new STRs.

While many studies are available about classic STRs, up to now, data for the new 5ESS markers solely were available for few population samples, so our study provides additional information on the genetic variation within European populations, confirming these 20 loci when combined all together are useful genetic markers for forensic identification and paternity testing.

2.3 SE33 showed a similar variability in the 2 examined populations (Calabria and Malta) in comparison with other European populations: a total of 41 different alleles were observed with no allele being more frequent than 10,5%.

This confirmed that this locus is effectively highly polymorphic and should be routinely added to the set of STRs loci commonly studied in casework and in paternity cases using as reference the database we generated.

2.4 Concerning the X_STRs, all the 8 X-STRs analyzed showed an important degree of variability and DXS10135 was the most informative.

The power of discrimination (PD) value of ChrX markers varies depending on the sex: in all loci, the PDF value was higher than the PDM value.

In a mixed female/male stain, the chance of having all male alleles included in the female component is higher for ChrX than for AS markers. Therefore, it is not advisable to use ChrX markers for testing male traces in a female background.

On the contrary we recommend the use of ChrX markers, in order to identify female traces in male background.

In fact only if the female component coincidentally happens to be homozygous at all loci, hence female alleles are coincident with male ones and female component is not distinguishable from the male one , this is really unlikely.

2.5 Seven Western Mediterranean populations were genotyped for 12 Y-chromosome STR loci. Results were consistent with previous studies showing a longitudinal cline of frequencies from the west to the east of the European landscape: for example in DYS392, allele 13 was the most common amongst the Spanish populations and allele 11 the most frequent in southern Italy.

Moreover it was noteworthy that 46 haplotypes (almost all from Valencia, the Balearic Islands and Sicily) matched with African samples. This result is concordant with other studies showing African influences in the Mediterranean populations.

It has been observed that the 2 most frequencies haplotypes in the Ibiza population were also diffused in Northern and Central Italy.

Results of the present study provide a useful Y-STR haplotypes dataset, for Mediterranean populations that can be used in the forensic field to correctly weigh the value of the evidence of a Y profile match.

2.6 52 SNPs were analyzed and have shown to be informative in the population samples studied. This means they can provide valuable information not only for population studies but also for forensic applications due to the short amplicon size (e.g highly degraded samples) as well as the implementation of criminal DNA databases or to solve complex pedigree (e.g. distant relationships or incomplete pedigrees) as a complement to standard STRs methodology.

The application of SNPs in relationship testing has not been widespread to date because nearly all paternity cases are adequately resolved with existing well validated STR sets.

However since a consistent proportion of relationship tests can involve analysis of human remains, SNPs can be useful to solve ambiguous results from incomplete STRs profiles commonly obtained from degraded DNA.

CONCLUSIONES

1. Validación de nuevos multiplex de STRs

1.1 El examen de una serie de más de 100 extractos de ADN de muestras forenses diferentes ha permitido una valoración exhaustiva del funcionamiento de diferentes STRs. Así, pudimos verificar que las propiedades individuales de los STRs son directamente responsables del funcionamiento de una múltiplex cuando se analizan muestras degradadas o con una mínima cantidad de ADN.
Por ello, recomendamos el empleo de STR de pequeño tamaño basado en diseños de primers bien establecidos, cuando los STRs tradicionales no funcionan o dan resultados parciales con muestras complejas.

1.2 El empleo de kit comerciales de nueva generación gracias a la incorporación de los nuevos marcadores del ESS posibilitan un drástico aumento de la informatividad en una única reacción de PCR lo que representa una ayuda adicional valiosa para casos reales y además tienen importancia para los laboratorios que actualmente usan STRs del sistema CODIS y quieran incorporar nuevos marcadores y aumentar la eficacia de sus multiplexes..

1.3 Se puede concluir a modo de resumen que se han validado con éxito los nuevos 5 sistemas del ESS y probado que constituyen una solución valiosa para el analisis forense y en particular para aquellos casos complejos en los que la elevada degradación de la muestra reduce sensiblemente la informatividad de los multiplexes clásicos de STRs de uso habitual.

2. Estudio Demográfico para evaluaciones forenses estadísticas

2.1 Se creó una base de datos para algunas poblaciones mediterráneas para análisis los análisis estadísticos de las pruebas forenses. Los estudios poblacionales fueron realizados sobre poblaciones de interés y se evaluó la eficacia de marcadores clásicos y sobre todo de los recientemente introducidos en la práctica forense.

2.2. Se estudiaron 15 STRs clásicos y 5 nuevos STRs del ESS en muestras de Galicia (NW España) y Calabria (Sur de Italia) y se encontró que las frecuencias de los alelos de estas dos poblaciones eran muy similares para cada uno de los STRs estudiados.
Además por lo que concierne los 15 STRs clásicos, el marcador TPOX confirmó ser el menos informativo y los nuevos D12S391 y D1S1656 los más discriminatorios.
Mientras muchos estudios están disponibles para los STRs clásicos, se han obtenido aun pocos datos para los marcadores 5 nuevos marcadores del ESS. Por ello nuestro estudio representa una información adicional sobre la variación genética dentro de las poblaciones europeas, confirmando que el total de los 20 marcadores es muy útil para la identificación forense y pruebas de paternidad.

2.3 El STR SE33 mostró una importante variabilidad en las 2 poblaciones examinadas (Calabria y Malta) en comparación con otras poblaciones europeas ya que fueron observados un total de 38 alelos diferentes y ninguno de ellos con una frecuencia superior al 10,5 %.Esto confirma que este marcador es sumamente polimórfico y debería ser añadido al grupo de STRs habitualmente estudiados en trabajos forenses de identificación y en casos de paternidad, en los que se podrá utilizar como referencia la base de datos que hemos generado.

2.4 Se analizaron 8 X-STRs que mostraron un importante grado de variabilidad siendo el DXS10135 el más informativo, de acuerdo con lo que ocurre en otros grupos europeos.
El valor del poder de discriminación (PD) de los X-STRs varía según el sexo: en todos los marcadores, el valor de PDF demostró ser más alto que el valor de PDM.
En una mancha mixta femenina/masculina, la posibilidad de tener todos los alelos masculinos incluidos en el componente femenino es más alto para los X-STRs que para los marcadores autosómicos. Por lo tanto, no es aconsejable usar marcadores del cromosoma X para buscar el componente masculino en un trasfondo femenino. Al contrario recomendamos el empleo de marcadores de cromosoma X para identificar huellas femeninas en un trasfondo masculino.

2.5 Se analizaron 7 poblaciones mediterráneas occidentales para 12 Y-STRs. Los resultados fueron compatibles con estudios anteriores y demostraron una variación clinal de frecuencias del oeste al este de Europa: por ejemplo en el sistema DYS392, el alelo 13 es el más común entre las poblaciones españolas y el 11 el más frecuente en el sur de Italia

Además es significativo que 46 haplotipos (principalmente de Valencia, Islas Baleares y Sicilia) están emparejados con haplotipos norteafricanos. Este resultado es concordante con otros estudios que mostraron la influencia africana en las poblaciónes mediterráneas.

Ha sido observado que los 2 haplotipos mas frecuentes en la población de Ibiza también están difundidos en el Norte y en el Centro de Italia.

Los resultados del estudio presente proporciona una base de datos de Y-STR para poblaciones mediterráneas, que pueden ser usadas en el campo forense para valorar estadísticamente la prueba forense en caso de coincidencia de haplotipos Y.

2.6 Se analizaron 52 SNPs del grupo SNPforID que demostraron ser informativos en las poblaciones estudiadas. Esto significa que no solo pueden contribuir a estudios poblacionales sino sobre todo a trabajos forenses debido al tamaño más corto de los productos amplificados (por ejemplo en muestras degradadas) y serían de utilidad para la ampliación de bases de datos de ADN criminales o para solucionar pedigríes complejos (p.ej.relaciones distantes o pedigríes incompletos) como un complemento a la metodología STRs estándar.

El uso de SNPS en pruebas de parentesco no está popularizado hasta el momento porque casi todos los casos de paternidad son resueltos adecuadamente con marcadores STR, pero ya que una proporción cada vez más importante de pruebas de parentesco implica el análisis de restos humanos, el pequeño tamaño de los SNPs puede ser útil para evitar resultados ambiguos como los perfiles incompletos de STRs que son obtenidos frecuentemente en muestras de ADN degradado.

Chapter VIII

FUTURE PERSPECTIVES

The advent of modern DNA technology has resulted in the increased ability to perform human identity testing: so now the way to perform crime scene analysis has changed and also the traditional figure of the forensic haematologist. Advances have increased the need for law enforcement officers, including seasoned investigators, to regularly attend education and training programs in the forensic sciences.

The ability to recover and analyze the DNA found in a small or degraded crime scene samples can make forensic DNA analysis an extremely powerful investigative tool. Small drops of blood (as small as the size of a pin), perspiration, saliva, dandruff and hair shafts can now be analyzed to provide a DNA profile of the source of that biological sample.

If we look back at the developments in this field, we can see enormous progress over the past 25 years: DNA testing is now an essential part of the crime laboratory's armamentarium in the investigation of crimes. STRs have become popular DNA markers because they are easily amplified by polymerase chain reaction (PCR) and have small size and low mutation rates, which makes the data more stable and predictable. Furthermore the number of repeats in STR markers is highly variable among individuals, which make these STRs effective for human identification purposes where it is important to have DNA markers that exhibit the highest possible variation in order to discriminate between samples.

Because of these characteristics, STRs with higher power of discrimination are nowadays used in forensic cases on a regular basis to identify victim, perpetrator, missing persons, and others. [Butler,2011]

STRs placed on chromosome Y are also used for sexual and non-sexual assault cases where mixed samples are collected from evidence. Y-STR testing can help to identify all males who have contributed to the evidence. X-chromosome markers may complement the results obtained from other genetic markers in particular in complex relationship cases, where the alleged father is absent and where only his close relatives

are available for testing. especially the DNA-analysis of multiple females under the hypothesis that they share the same father. [Kayser,1997 ; Szibor,2007]

In cases where DNA evidence is limited, either in quantity or quality, such as highly degraded samples that are exposed to environmental insults or inhibitors, standard STR testing is often inadequate. Analysis of these compromised DNA samples often result in dropout of the larger STR loci from the samples and only a partial DNA profile can be obtained. Partial DNA profiles generally do not provide the power of discrimination to include or exclude a potential contributor to the sample.

To overcome these problems new markers has been selected in the last years, in order to recover as more information as possible from smaller regions of DNA, which are more likely to be intact following DNA damage. These include miniSTRs and single nucleotide polymorphisms (SNPs). [Butler,2003]

Because of the ability to type very degraded samples, mini-STR technology can be expected to provide forensic scientists with another tool that captures genetic data from DNA samples of marginal and extremely low quality and quantity.

Thus, many previously unsolvable human identity cases may be resolved with mini-STR technology

Moreover in particular the interest in SNPs is rising because they show a range of characteristics that make them well suited even to forensic analysis, including: abundance in the genome, Low mutation rates, reduced amplicon sizes (ability to analyze degraded DNA), relatively simple multiplex assays, potential for automation.

SNPs can be detected in short amplicons (less than 150bp) that is really useful for the analysis of degraded samples. In forensic paternity casework and anthropological investigations, where biological samples are often poor or degraded, the particular advantage of SNPs is that the studied DNA sequences are much shorter than those used for “classic” DNA analysis. Moreover they show a low mutation rate, that is relevant for the paternity investigations and finally they can be adapted to automation with high throughput.

In addition they could be useful for population studies in order to estimate the allele frequencies of the SNPs selected, for increasing the power of kinship test or family reconstruction analysis and also for creating criminal DNA databases.

Single nucleotide polymorphisms may have in the near future a fundamental role in forensics, not only in specialized applications such as phylogeographical or ancestry studies (mtDNA,Y-SNPs) but also for the potential applications of autosomal SNPs either in the prediction of phenotypic traits than in real forensic caseworks applications.

Unfortunately SNPs are less informative for identity testing than STR, because most of them have only two alleles. This means more SNPs are required to reach the same level of discrimination obtained with 13-15 STR loci. It has been demonstrated that the information from 50 SNP arrays would be comparable to that obtained by existing STR multiplexes.[Butler,2007] This involves the necessity to produce multiplexes with a large panel of SNPs that obviously require for typing far more DNA than the one needed for several multiplex STR systems. In the last years biotechnology companies are making big efforts in developing new strategies for SNP typing: this rapid technological progress makes difficult to choose the appropriate method for specific applications.

Before the introduction in routine casework analysis, it will be relevant for the forensic community, to establish which SNP markers/platforms may be useful for catching up the procedure to a level acceptable for forensic application and then to validate selected markers panels with sufficient analysis repeat rates.

Probably, different technologies are to be used for forensic casework since it is often difficult to collect adequate amounts of DNA from scarce samples, and in paternity testing or for DNA databases where the DNA quality/quantity is much less critical. In general, all technologies that have limited multiplexing capability should be excluded as candidates for routine forensic analysis either because of the limited quantity of DNA available from evidences, than because the number of markers must be as large as possible to yield a significative power of discrimination.

For forensic applications a medium throughput is enough for criminal casework and paternity test while a high throughput is necessary for DNA databases.

Actually, for routine application minisequencing by SNaPshot analysis seems to be the best choice method in forensic laboratories in terms of cost and efficiency, because of the high multiplex capacity, the sensitivity of the system, the ability to perform

detection on an automatic capillary electrophoresis instrument, without the requirement of any special platform.

However MALDI-TOF MS analysis might be a useful future tool in routine analysis of SNPs in forensic, genealogical and other applications, because it's time-saving and cost-efficient. However in comparison with other sequencing methods, the MALDI-TOF MS technology is not only always similar reliable and the multiplex capability is lower. Moreover it requires high-purity samples for genotyping that increase time consume for analysis and sample-processing costs. [Petkovski,2005]

As a consequence, this technology can be useful for validating candidate SNPs or for setting DNA databases,but not for routine application in forensic casework.[Gill,2004]

Personal identification seems likely to be the first application of SNPs in routine caseworks, followed by the lineage based family reconstructions.

Moreover the ability to perform genetic typing of biological traces collected at the crime scene, in order to obtain information about a donor's physical characteristics, is a very attractive prospect for forensic analysis and it could potentially offer a powerful new tool for crime scene investigations. SNP typing could be useful to help investigators with info obtained from the biological evidence left at the crime scene.

In fact predicting ancestry or phenotypic characteristics, such as hair, skin or eye colour, is another role that SNPs may play: obviously this more specialized applications will require considerably more research and development efforts to provide reliable predictions about a persons. [Grimes,2001; Sulem2007]

The application of SNPs in relationship testing has not been widespread to date because nearly all paternity cases are adequately resolved with existing well validated STR sets but since a consistent proportion of tests can involve analysis of human remains for relationship or DVI, SNPs small size provides an important way to avoid a further source ambiguous results from incomplete STRs profiles commonly obtained from degraded DNA [Phillips,2008].

However, since the expertise in STR typing is more than decennial and all forensic databases are well-established on STR loci, thus it is unlikely that SNPs will replace STR loci as the primary forensic markers for human forensic identification.

Even if it is not likely that SNPs typing will totally replace STRs as the principal method for human identification, however new SNPs genotyping methods, chemistries and platforms are continuously being developed and considerable researches are still undergone to establish adequate scientific foundations for these applications, even if probably the community will not continue to invest the same level of time and money in research as in the past.

It is favourable that once validated for forensic applications, SNPs may be used as a supplementary battery of forensic markers, while retaining STRs as the primary set of markers for forensic DNA analysis. It's evident that future development will become more diversified, which means less money and time spent on each topic. [Gill,2001]

ACKNOWLEDGEMENTS

I would like to thank firstly Prof. Angel Carracedo who kindly offered me the great opportunity to fulfill my PhD at the Institute of Legal Medicine in Santiago de Compostela and Prof. Maria Victoria Lareu for her support, guidance, and encourage ments through these years.

A special thank to Prof. Chris Phillips for sharing with me his knowledge within the field: without whose contribute the thesis could not be prepared. It was a great privilege for me to met all of them and spend this time all together .

I thank also all my colleagues at the Institute, in particular Luis Fernandez-Formoso,for their help and support and for making the time at the lab productive and happy.

Finally, a thank to all co-authors of my papers for the work they have done and to all others who I not mentioned but has been helpful in different ways

UNIVERSIDAD DE SANTIAGO DE COMPOSTELA
FACULTAD DE MEDICINA
DEPARTAMENTO DE ANATOMÍA PATOLÓXICA