MARKER SELECTION IN FORENSIC GENETICS

Martin Ramos, Edgar. Genetics Degree. Facultat de Biociències. Universitat Autònoma de Barcelona. Cerdanyola del Valles, Barcelona, Spain.

UAB Universitat Autònoma de Barcelona

INTRODUCTION

Forensic genetics bases its applications in human genome variation. This variation can present itself, for example, as a difference in a single nucleotide at the same position in the same species, thus it is considered a marker of the difference between individuals. When analyzed, these markers can help identify every single human being, so they have been used for judicial investigation and parental testing.

An extensive bibliographic research has been done to know which markers are used actually and their current applications. This review aims to explain all their characteristics to deduce why using one marker in certain circumstances may be more suitable than using another.

SHORT TANDEM REPEATS (STR)

2-5 pb
6 or more repeats → $(TAGC)_n$

↓

Every repeat is a different allele

↓
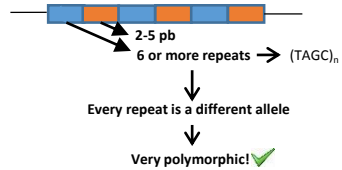
Very polymorphic! ✓

Individual identification using multiplex PCR to analyze more than 20 high polymorphic STRs ✓

GlobalFiler® analizes 23 loci. You can find above two of them: locus D16S539 is heterozygous (alleles 11 and 12) and locus DYS291 is homozygous for allele 11[1]

→ **High discrimination power** ✓
   └ Random match probability is very low

→ **The cheapest** ✓

→ **Do not work in degraded samples** ✗
   └ Allele drop out or Allele imbalance

→ **Can't identify tissue of origin** ✗

SINGLE NUCLEOTIDE POLYMORPHISMS (SNP)

Individual A ...GCA**A**CGTTAGA...
Individual B ...GCA**G**CGTTAGA...

↓

Biallelic

↓

Low heterozygosity ✗

Need 50 SNP to achieve same statistic power as STR ✗

SNP array by Affymetrix **Expensive!** ✗

→ **Explain variation in phenotypical characters**

Hair color   Skin color   Eye color

→ **Combined with STR, increase statistical power** ✓
→ **Smaller amplicons** → **Can be used for degraded samples** ✓

Hair	Dark	Light		
	0.839	0.161		
Black	Brown	Red	Blond	
0.631	0.309	0.001	0.059	

Eye	Blue	Int.	Brown
	0.058	0.183	**0.759**

Prediction of Externally Visible Characteristics by Hirisplex®[2] ✓

MITOCHONDRIAL DNA (mtDNA)

Cell
Nucleus
DNA
Mitochondrion

A cell has an elevated number of mitochondria, each with multiple copies of mtDNA.
Image from Muscular Distrophy Assotiation

Every cell has a high copy number

Low effective population size

Its mutation rate is 10-fold higher than nuclear's.

Maternally inherited

→ **Higher chance to survive degradation** ✓

The most polymorphic regions are **hypervariable regions** (HVR).
Image from National Forensic Science Technology Center

SNPs analysis

Missing person's identification by comparing with a matrilineal relative ✓

Identification of tsar Nicholas II (sequence 1) thanks to HVR sequences from his living relatives[3]

→ **Heteroplasmy** ✗

↓

Two mtDNA subpopulations (in this case C and T) in the same individual.

Haplogroup identification → **Predicts geographic origin**[4] ✓

Image from Wikipedia

mtDNA macro-haplogroup L is more frequent in Sub-Saharan Africa.

↓

An individual with macro-haplogroup L will probably have the same phenotype as a Sub-Saharan individual

Y CHROMOSOME

The analyzed Y chromosome STR are located in the **non-recombining region.** Image from National Geographic

Only present in men

— Pseudoautosomal region (homologous to X)

— Non-recombining region (NRY)

→ **Paternally inherited as a single locus**

SNPs analysis → **Missing person's identification by comparing with a patrilineal relative** ✓

STRs analysis

Useful for sexual crimes

→ **Low discrimination power** ✗

Población reclusa por nacionalidad, sexo y periodo.
Unidades: número de reclusos

	2008	2009	2010	2011	2012
TOTAL					
Ambos sexos	73.558	76.079	73.929	70.472	68.597
Varones	67.608	70.003	68.141	65.184	63.372
Mujeres	5.950	6.076	5.788	5.288	5.225

Men commit more violent crimes than women
Data from Instituto Nacional de Estadística.

Male component identification[5] ✓

• For mixed stains where proportion of female component is higher than the male component.
• To determine the number of semen donors in gang rape.
• In cases where evidence is expected to include cells of a male perpetrator even when there is no ejaculation.

EXPRESSION ANALYSIS

Based on expression differences between tissues:
• Messenger RNAs have a shortest half-life
• Micro RNAs (miRNAs) are shorter → less susceptible to degradation

↓

Screening of body fluid RNA using quantitative PCR

Not useful for identification ✗

Can identify tissue of origin ✓
relative expression ratio

miR214's expression is higher in menstrual blood (MB) than venous blood (VB), semen (SE), vaginal secretion (VS) or saliva (SA)[6]

DATABASES ✓

All information obtained from analysis of STR and SNP from nucleus and mitochondria is stored in databases to:

FBI   INTERPOL

• Increase the possibility of solving crimes.
• Increase the possibility to link unsolved crimes.
• Increase the possibility to identify false identities.
• Increase the possibility to identify disaster victims.
• Allow cooperation between countries.

CODIS' 13 core loci plus sex determining amelogenin[7].

CSF1PO	FGA	THO1	TPOX
VWA	D3S1358	D5S818	D7S820
D8S1179	D13S317	D16S539	D18S51
D21S11	AMEL (sex)		

CONCLUSIONS

• STRs are the most used markers for individual identification because of their high liability and their low cost. However, they are incapable of predicting the phenotype and to identify the tissue of origin. Also, their analysis presents problems when the samples are degraded.
• SNPs are not recommended for identification, but to increase STRs discrimination power. They can be used to predict phenotypes like hair color and can be analyzed when the sample is degraded.
• mtDNA's use is more recommended when nuclear DNA is too degraded for analysis, thanks to its elevated copy number by cell ⎫
• Y chromosome analysis is useful in sexual crimes, where male part needs to be analyzed ⎬ Haplogroup identification can predict the geographic origin
• Databases store information and are useful tools to improve the efficiency in crime investigation. ⎭
• Expression analysis can be useful to identify the sample's tissue of origin.

BIBLIOGRAPHY

1. Flores et al. 2014. *Forensic Science International: Genetics*, 10, 33-39
2. Walsh et al. 2013. *Forensic Science International: Genetics*, 7, 98-115
3. Gill et al. 1994. *Nature Genetics*. 6(2), 130-135
4. Kayser & de Kniff, 2011. *Nature Reviews Genetics*, 12(3), 179-192
5. Roewer et al. 2009. *Forensic science, medicine and pathology*, 5(2), 77-84
6. Wang et al. 2012. *Forensic Science International: Genetics*. 7(1), 116-123
7. http://www.fbi.gov/about-us/lab/biometric-analysis/codis/codis-and-ndis-fact-sheet