

UNIVERSIDADE DE SANTIAGO DE COMPOSTELA  
FACULTADE DE MEDICINA  
DEPARTAMENTO DE ANATOMÍA PATOLÓXICA E CIENCIAS  
FORENSES



**“Estudo de polimorfismos autosómicos e de cromosoma Y humanos con aplicacións en xenética forense e poboacional”**

Memoria presentada para a obtención do grao de doutor por

**Eva Ramos Luis**



O Doutor Ángel Carracedo Álvarez, Catedrático de Medicina Legal da Facultade de Medicina da Universidade de Santiago de Compostela, a Doutora María Brión Martínez, investigadora do Complexo Hospitalario Universitario de Santiago (CHUS), e a Doutora Paula Sánchez Diz,

CERTIFICAN

Que a presente memoria, que leva por título **“ESTUDO DE POLIMORFISMOS AUTOSÓMICOS E DE CROMOSOMA Y HUMANOS CON APLICACIONES EN XENÉTICA FORENSE E POBOACIONAL”**, da licenciada en Bioloxía pola Universidade de Santiago de Compostela Eva Ramos Luis, foi realizada baixo a nosa dirección, considerándoa en condicións para optar ao Grao de Doutor e autorizándoa para a súa presentación ante o Tribunal correspondente.

E para que así conste, firmamos a presente en Santiago de Compostela, a 25 de xuño de 2014

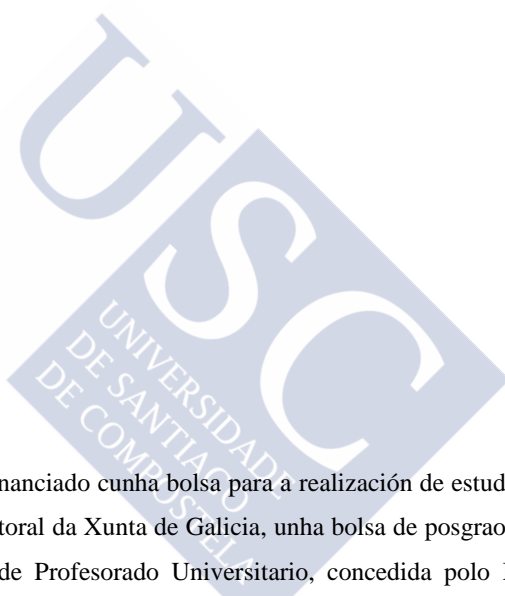
Asdo.: Dr. Ángel Carracedo Álvarez.

Asdo.: Dra. María Brión Martínez.

Asdo.: Dra. Paula Sánchez Diz.

Asdo.: Eva Ramos Luis.





Este traballo foi financiado cunha bolsa para a realización de estudos de terceiro ciclo e unha bolsa predoutoral da Xunta de Galicia, unha bolsa de posgrao do Programa Nacional de Formación de Profesorado Universitario, concedida polo Ministerio de Educación y Ciencia, e cos proxectos BMC2003-09822 e FFI2008-03252.



## AGRADECEMENTOS

Estas páxinas só pretenden achegar doutro xeito o meu agradecemento a todas as persoas que me axudaron na elaboración desta tese doutoral, fora coa súa axuda profesional como persoal. A máis importante veu de man dos meus directores de tese que ademais de transmitirme os seus coñecementos, sempre me deron unha oportunidade detrás doutra. A Mari, por todo o tempo que me dedicou e a infinita paciencia que tivo comigo, non sei como poderei agradecer que me fixera ver a luz ao final do bucle. A Ángel, porque é o noso fareiro, porque sempre nos transmitiu a ilusión pola investigación e pola nosa cultura, e porque non deixa de loitar por este grupo, pese aos tempos que corren. A Paula, porque me axudou a dar os primeiros pasos e me transmitiu as súas gañas de loitar sempre, moitas grazas.

A Bea, que sempre cun sorriso me axudou en todas as dificultades que me fun atopando neste camiño e non deixou que me desanimara nunca.

A Maviki, porque sempre nos animou a todos a non tirar a toalla e a aproveitar a gran oportunidade que tivemos ao estar nun grupo coma este. A Chris e Toño, pola ilusión no seu traballo e todo o que nos ensinaron.

A Álex porque é o mellor compañeiro que calquera desexaría ter, e eu tiven a sorte de telo sempre ao meu lado. A Carolina, Juan, María Torres, Xabi e Xulio, por todo o apoio que me deron sempre. Ao resto de compañeiros do grupo de cardio: Ángela, Marina, Montse, Rocío, Sandra e especialmente Mariña, por todo o que padeceron coa recta final desta tese. E ao resto de compañeiros da Fundación e de CeGen: Cata (e ao seu sorriso), Inés, Jens, Joja, Jorge Amigo, Julio, Marián, Montse Camiña, Noa, Peñas e Xabi Bello, entre outros, por terme axudado e animado tantas veces.

A todos “os do depar”, cos que compartín tantas experiencias e que me axudaron e apoiaron de tantas formas neste camiño: Alberto, Ana F. Aradas, Ana Pastoriza, Ángela, Anita, Carla, Danel, Esther, Fonde, Francesca, Jens, José Manuel, Leonor Gusmão, Luís, Manu, María Cerezo, María Dosil, María Mayo, Miguel, Nuria Naverán, Olalla, Raquel, Rui, Sandra Filippini, Ulises, Vane e Yari, entre outros cos que tiven a sorte de coincidir. Grazas por facer que me custara tanto marcharme dalí.

Grazas a xentes doutros sitios ou disciplinas coas que tiveron a sorte de colaborar, como Francisco Villar, por todo o que aprendín ao seu carón e o que confiou en nós, e a xente do laboratorio de París: Amandine, Camilo, Périnne e Veronique, polo ben que me fixeron sentir entre eles. Tamén quero agradecer o apoio de Ana Canda, Rosi e ao meu profesor de ciencias, Javier.

A Gloria e Meli, compañeiras de traballo nun principio, grandes amigas despois e das persoas que máis me animaron e me aturaron ata chegar ata aquí, moitas grazas.

Ás amigas coas que empecei esta aventura da bioloxía e do doutorado: Amparo, Ana, Irene, Leti e Rebe (Ampe, Txokyto, Foz, Pinki e Bbk), ben deitei de todas elas!

A todas as miñas amizades de Verín (en especial a Elena, Eva, Irene, Iria, Lidia, Marta e Uxía), de Lamardeite (*á moxedá*) e de Santiago. Moitas grazas por non deixar que me desanimara nunca. Tamén aos amigos de Cantigas e Agarimos, Agustín Delicias, Charan, Fumaces e Pili, Leo de Olvera, Marina, Pedro e Petru entre tantos outros.

A toda a miña familia, en especial á miña avoa, a meus tíos e ao *Calros*, o seu apoio foi sempre imprescindible. Tamén aos que xa non están. A Vicente, porque día a día axudou cun sorriso a que seguira avante. E por último, a meus pais, por todo, non hai palabras para agradecer o que fixeron para que eu puidera chegar ata aquí.

Moitísimas grazas a todos vós.

# ÍNDICE

ABREVIATURAS.....	xi
<b>1 INTRODUCCIÓN.....</b>	<b>1</b>
1.1 <i>O xenoma humano</i> .....	3
1.2 <i>A variabilidade xenética humana</i> .....	6
1.3 <i>Os polimorfismos de ADN</i> .....	8
1.3.1 Haplotipos e desequilibrio de ligamento.....	10
1.3.2 Os microsátélites.....	10
1.3.2.1 Definición e características xerais.....	10
1.3.2.2 Mecanismo e taxa de mutación.....	11
1.3.2.3 Métodos de detección e artefactos.....	13
1.3.2.4 Estandarización e bases de datos.....	15
1.3.2.5 Os marcadores mini-STR.....	17
1.3.3 Os marcadores SNP.....	18
1.3.3.1 Definición e características xerais.....	18
1.3.3.2 Mecanismo e taxa de mutación.....	20
1.3.3.3 Métodos de detección e artefactos.....	21
1.3.3.4 Estandarización e bases de datos.....	23
1.3.3.5 <i>Multiplexes</i> de marcadores SNP: o 52-plex.....	24
1.4 <i>O cromosoma Y</i> .....	26
1.4.1 Definición e características xerais.....	26
1.4.2 Estrutura do cromosoma Y.....	30
1.5 <i>Os polimorfismos de cromosoma Y</i> .....	31
1.5.1 Marcadores STR-Y.....	31
1.5.1.1 Taxa de mutación dos STR-Y.....	34
1.5.2 Marcadores SNP-Y.....	36
1.5.2.1 A filoxenia do cromosoma Y.....	38
1.6 <i>Aplicacións do estudo da variabilidade xenética</i> .....	47
1.6.1 Marcadores STR.....	47
1.6.1.1 Aplicacións de interese forense.....	47
1.6.1.2 Aplicacións de interese poboacional.....	48
1.6.2 Marcadores SNP.....	49
1.6.2.1 Aplicacións de interese forense.....	49
1.6.2.2 Aplicacións de interese poboacional.....	51

<b>2</b>	<b>XUSTIFICACIÓN E OBXECTIVOS .....</b>	<b>53</b>
<b>3</b>	<b>MATERIAL E MÉTODOS .....</b>	<b>59</b>
<b>4</b>	<b>RESULTADOS .....</b>	<b>65</b>
4.1	<i>Polimorfismos autosómicos .....</i>	67
4.1.1	Análise de ADN degradado utilizando marcadores STR e SNP –resultados dun exercicio europeo de colaboración (EDNAP)- .....	67
4.1.2	Validación forense do ensaio 52-plex de SNPforID .....	81
4.1.3	Avaliación do sistema de xenotipado de SNP <i>GenPlex</i> e o panel de marcadores forenses 49-plex ..	89
4.2	<i>Polimorfismos de cromosoma Y .....</i>	97
4.2.1	Taxas de mutación de microsatélites específicos de cromosoma Y .....	97
4.2.2	Análise do ADN de cromosoma Y de liñaxes masculinas francesas.....	109
4.2.3	O poboamento de Europa e o cauteloso relato da liñaxe R-M269 de cromosoma Y.....	119
4.2.4	Evidencia de estrutura xenética tribal prerromana en vascos a través de marcadores herdados uniparentalmente .....	131
<b>5</b>	<b>DISCUSIÓN.....</b>	<b>145</b>
<b>6</b>	<b>CONCLUSIÓNS .....</b>	<b>157</b>
<b>7</b>	<b>BIBLIOGRAFÍA.....</b>	<b>161</b>
<b>8</b>	<b>ANEXOS .....</b>	<b>179</b>

## ABREVIATURAS

A: Adenina.

ADN: Ácido desoxirribonucleico.

ADNmt: ADN mitocondrial.

AIM: *Ancestry Informative Marker* (marcador informativo de ascendencia).

ACP: Análise de componentes principais.

ASD: *Average Squared Distance* (media do cadrado das distancias).

C: Citosina.

CEPH: *Centre d'Etude du Polymorphisme Humain* (centro de estudos dos polimorfismos humanos).

CODIS: *COmbined DNA Index System* (sistema indexado combinado de ADN).

DHPLC: *Denaturing High-Performance Liquid Chromatography* (cromatografía líquida desnaturalizante de alta resolución).

DYS: *DNA Y-chromosome Segment* (segmentos de ADN de cromosoma Y).

EDNAP: *European DNA Profiling Group*.

ENCODE: *Encyclopedia Of DNA Elements*.

ENFSI: *European Network of Forensic Science Institutes*.

ESS: *European Standard Set* (set estándar europeo).

FBI: *Federal Bureau of Investigation*.

FRET: *Fluorescence Resonance Energy Transfer* (transferencia de enerxía por resonancia de fluorescencia).

FSS: *Forensic Science Service*.

G: Guanina.

GHEP-ISFG: Grupo de fala española e portuguesa da ISFG (tamén GEP-ISFG).

HVSI: Segmento hipervariable I.

HVSII: Segmento hipervariable II.

IC: Intervalo de confianza.

Indel: Inserción/ deleción.

ISFG: *International Society for Forensic Genetics*.

kb: Quilobase (mil bases/ pares de bases).

LCN: *Low Copy Number* (ADN en número de copias baixo).

LINE: *Long Interspersed Nuclear Element* (elementos nucleares dispersos longos).

LT: *Life Technologies*.

MAF: *Minor Allele Frequency* (alelo de menor frecuencia).

Mb: Megabase (millón de bases/ de pares de bases).

MSY: *Male-Specific region* (rexión específica masculina do cromosoma Y).

NDNAD: *The United Kingdom National DNA Database* (base de datos nacional de ADN do Reino Unido).

NIST: *National Institute of Standards and Technology*.

NRY: *Non-Recombining Region* (rexión non recombinante do cromosoma Y).

PAR: Rexión pseudoautosómica.

pb: pares de bases.

PCR: *Polymerase Chain Reaction* (reacción en cadea da polimerasa).

RFLP: *Restriction Fragment Length Polymorphism* (polimorfismos na lonxitude dos fragmentos de restrición).

SINE: *Short Interspersed Nuclear Element* (elementos nucleares dispersos curtos).

SNP: *Single Nucleotide Polymorphism* (polimorfismo de nucleótido único).

SNP-Y: Marcadores SNP específicos de cromosoma Y.

SMM: *Stepwise Mutation Model* (modelo de mutación “paso a paso”).

STR: *Short Tandem Repeat* (repetición curta en tándem).

STR-Y: Marcadores STR específicos de cromosoma Y.

SWGAM: *Scientific Working Group on DNA Analysis and Methods* (grupo científico de traballo en métodos de análise de ADN).

T: Timina.

TMRCA: *Time to the Most Recent Common Ancestor* (tempo ata o antepasado común máis recente).

USC: Universidade de Santiago de Compostela.

YCC: *Y Chromosome Consortium* (consorcio de cromosoma Y).

YHRD: *Y-STR Haplotype Reference Database* (bases de datos de referencia de haplotipos de marcadores STR de cromosoma Y).



## **1 INTRODUCCIÓN**



## 1.1 O xenoma humano

O xenoma humano é o conxunto de información xenética que se encontra nas células dun individuo, e está formado polo xenoma nuclear e o xenoma mitocondrial ou citoplasmático.

O xenoma nuclear está formado por 22 pares de cromosomas autosómicos e un par de cromosomas sexuais (XX no caso das mulleres e XY no caso dos varóns) para as células con dotación diploide, é dicir, as células somáticas. No caso das células con dotación haploide, as da liña xerminal, o xenoma nuclear está composto por 22 cromosomas e un cromosoma sexual (X nos óvulos e X ou Y nos espermatozoides).

O xenoma mitocondrial atópase no interior das mitocondrias e existen miles de copias en cada célula somática. Nos óvulos a cantidade é cen veces superior á das células somáticas e nos espermatozoides existen moi poucas copias. Actualmente acéptase a idea de que as mitocondrias dos gametos masculinos non se transmiten á descendencia, polo tanto, considérase que ten unha herdanza exclusivamente materna.

En ambos casos, xenoma nuclear e mitocondrial, o material xenético está constituído pola molécula do ácido desoxirribonucleico (ADN). Todos os seres vivos coñecidos compartimos esta mesma molécula, isto dá idea da versatilidade case ilimitada do ADN para orixinar formas de vida diferentes.

O ADN contén toda a información necesaria para xerar un organismo nos chamados xenes, segmentos de ADN codificante que transmiten instrucións para a síntese de proteínas. Porén, máis do 98% do noso xenoma está formado por secuencias non codificantes. Este ADN inicialmente denominábase “ADN lixo” porque non se coñecía ningunha función para o mesmo. Na actualidade, identificáronse varias funcións para este xenoma, como é a regulación transcricional de secuencias codificadoras, con todo, a maioría aínda ten unha función descoñecida. O alto grado de homoloxía e conservación destas secuencias parece indicar que xogan un papel esencial nas células.

O xenoma non codificante está formado por secuencias de copia única e secuencias con múltiples copias. Estas últimas constitúen o chamado ADN repetitivo, que clasificamos de acordo coa lonxitude da unidade de repetición e coa disposición ao longo do xenoma, en dous tipos:

- ADN repetitivo disperso: as unidades de repetición están espalladas por todo o xenoma, falamos así de elementos LINE (*Long Interspersed Nuclear Elements*), como o elemento L1, e de elementos SINE (*Short Interspersed Nuclear Elements*), como é a secuencia Alu, entre outros.
- ADN repetitivo en tándem: as unidades de repetición están dispostas no xenoma de forma consecutiva, hai tres tipos principalmente:
  - ADN satélite: abrangue desde centos de quilobases (kb) a megabases (Mb), as súas unidades de repetición son dun extenso rango de tamaños, co que poden formar estruturas de orde superior e ser compoñentes funcionais importantes dos cromosomas, como o que está situado nos centrómeros (Jobling 2004).
  - ADN minisatélite: a unidade de repetición mide entre 10 e 100 pares de bases (pb) aproximadamente. A maior parte deste tipo de ADN atópase nas rexións teloméricas dos cromosomas.
  - ADN microsátélite/ marcadores STR (*Short Tandem Repeats*): a unidade de repetición mide entre 1 e 8 pb aproximadamente. Adoitan ser moi polimórficos polo que son usados frecuentemente como marcadores moleculares en estudos de xenética poboacional e forense (ver apartado 1.3.2).

O feito de que o funcionamento e a regulación do xenoma humano son moito máis complexos do que se viña pensando, foi confirmado á luz dos resultados do proxecto ENCODE, un proxecto de análise exhaustiva do xenoma que tiña como obxectivo deliñar todos os elementos funcionais codificados no xenoma humano (Birney et al. 2007; Dunham et al. 2012), e veu mesmo modificar a definición que se tiña ata o momento de xene, propoñéndose agora como a unión das secuencias xenómicas que codifican un conxunto coherente de produtos funcionais e que potencialmente poden solapar (Gerstein et al. 2007).

### **Composición química do ADN**

O ADN é un polímero formado polos nucleótidos, monómeros que están formados por tres partes: unha molécula de fosfato, un monosacárido (a desoxirribosa) e unha base nitroxenada. Esta última pode ser adenina (A), timina (T), guanina (G) ou citosina (C), e normalmente emprégase o nome da base nitroxenada para referirse ao

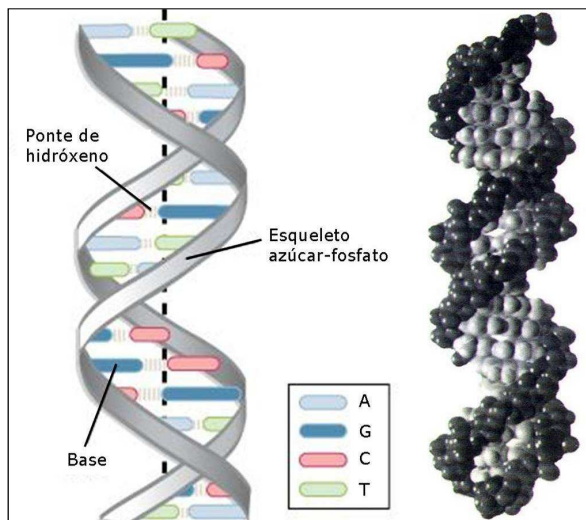
nucleótido. Sobre a base de criterios estruturais a A e a G son bases púricas ou purinas e a T e a C, bases pirimidínicas ou pirimidinas. Así, a información xenética dunha célula está contida na orde lineal dos nucleótidos do seu ADN, e a multitude de combinacións destes catro tipos de nucleótidos, é o que posibilita a variabilidade que existe no xenoma de cada individuo.

### **Estrutura do ADN**

O descubrimento da estrutura secundaria do ADN é un dos eventos máis importantes na historia da bioloxía, sendo para moitos, a orixe da bioloxía molecular moderna. Dilucidar a estrutura do material hereditario tería unha gran importancia á hora de coñecer o funcionamento de dita molécula. En 1953 James Watson e Francis Crick describiron a estrutura secundaria do ADN como a dunha hélice de dobre cadea (Watson et al. 1953). Este descubrimento foi posible en gran medida grazas ás fotografías realizadas por Rosalind Franklin mediante a técnica de difracción de raios X, que informaban sobre a posición dos átomos na molécula de ADN, e tamén se apoiou nos datos obtidos por Erwin Chargaff sobre a composición nucleotídica do ADN, a chamada "Regra da Equivalencia de Bases" (Chargaff 1950).

Deste xeito, ambos investigadores concluíron que a estrutura secundaria do ADN é unha dobre hélice formada por cadeas orientadas en direccións opostas. A estrutura mantense grazas aos enlaces de hidróxeno entre as bases nitroxenadas, que se encontran orientadas cara o interior das cadeas, e os enlaces fosfodiéster, que unen os nucleótidos da mesma cadea (figura 1).

Nas células, o xenoma eucariótico atópase asociado a proteínas que lle permiten o seu progresivo dobramento, a este complexo núcleo proteico denomínaselle cromatina. Durante a mitose celular o material hereditario forma unidades discretas, morfoloxicamente diferentes, que constitúen o estado de máxima compactación da cromatina, os cromosomas. Non obstante, durante a interfase celular, a cromatina atópase distribuída de forma irregular polo núcleo da célula. Existen dúas clases de cromatina, a eucromatina, que é a parte do xenoma que contén o ADN transcricionalmente activo e posúe unha conformación relativamente dispersa, e a heterocromatina que, pola contra, é transcricionalmente inerte e moito máis compacta que a eucromatina.



**Figura 1:** Estrutura secundaria do ADN, unha hélice de dobre cadea.

## 1.2 A variabilidade xenética humana

A maior parte da variabilidade xenética da nosa especie débese a diferenzas entre os individuos de cada poboación, máis que a diferenzas entre poboacións. Este repartimento comunmente aceptado crese que é debido a que as poboacións humanas non evolucionaron como entidades independentes se non que interactuaron entre si ao longo da historia e, por outra banda, tamén se debería a que somos unha especie demograficamente nova, a maioría da variabilidade que portamos compartímola cos nosos devanceiros africanos. Unha poboación pódese definir como un conxunto de individuos que comparten determinadas características sociais ou físicas, aínda que frecuentemente se empregue o termo en sentido teórico, para referirse a un grupo de individuos cuxo apareamento é ao azar.

A diversidade do xenoma humano está regulada por procesos evolutivos diferentes dependendo da función de dito xenoma. Desta maneira, a variabilidade do ADN codificante está modulada principalmente pola selección natural, e a do ADN non codificante está fortemente influenciada por forzas de carácter estocástico como son a deriva xenética, a migración ou fluxo xénico, e a mutación. A continuación descríbense brevemente estas catro forzas evolutivas, factores que modulan ou orixinan a variabilidade xenética humana.

### **A selección natural**

Concibida por Charles Darwin e Alfred Wallace a mediados do século XIX, a selección natural é un mecanismo evolutivo que se define como a contribución diferencial de individuos á xeración seguinte, segundo a súa capacidade de sobrevivir e reproducirse, é dicir, segundo a súa eficacia biolóxica, a denominada “*fitness*” en inglés. Porén, desde a teoría xerárquica da evolución, non só os xenes, se non tamén as células, os organismos, as especies e os clados, poden actuar como unidades evolutivas e estar sometidas á selección natural.

### **A migración**

O intercambio de xenes entre poboacións debido á migración dos individuos entre as mesmas, é outro factor importante de cambio xenético. A migración é o movemento de individuos dunha área ocupada a outra; cando esta implica desprazamento de xenes, denomínase fluxo xenético. Isto unicamente se produce cando se alcanza o éxito reprodutivo, é dicir, cando existe contribución xenética á xeración seguinte. Os cambios nas frecuencias alélicas serán proporcionais ás diferenzas de frecuencia entre a poboación doadora e a receptora e á taxa de migración (figura 2).

### **A deriva xenética**

Non toda a variabilidade xenética dunha xeración se transmite á xeración seguinte. Posto que as poboacións naturais teñen un tamaño finito, cada xeración implica unha transmisión aleatoria de xenes (durante a transferencia de gametos de pais a fillos) que fai que as frecuencias dos alelos flutúen de xeración en xeración. A deriva xenética, ou xénica, é o efecto acumulativo desta flutuación durante moitas xeracións. A consecuencia esencial deste axente evolutivo é a perda da variación xenética (figura 2).

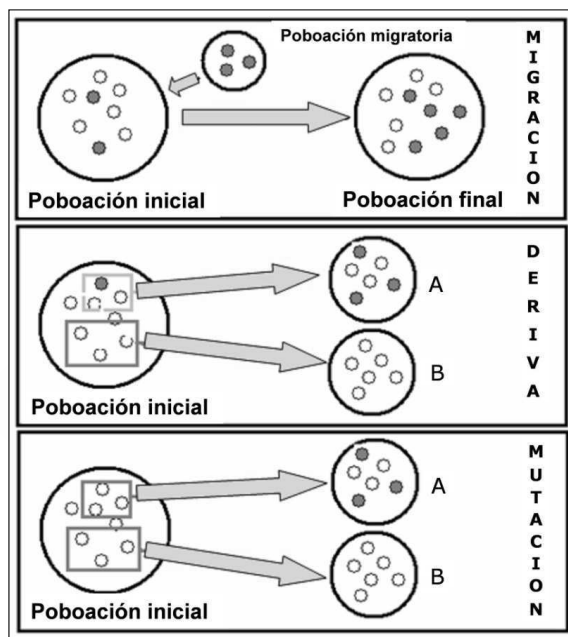
Dous casos extremos da deriva xenética son o efecto fundador e o colo de botella xenético. O primeiro prodúcese cando unha poboación se forma a partir de moi poucos individuos e, o segundo, cando se reduce drasticamente o tamaño efectivo dunha poboación. Ambos casos implican unha gran redución da diversidade xenética.

### **A mutación**

As mutacións son cambios estables que se producen na secuencia do ADN e constitúen a fonte última de variación xenética. A mutación é un proceso aleatorio, e cando afecta ao xenoma das células da liña xerminar dun individuo, proporciona a materia prima sobre a cal actuarán o resto de forzas evolutivas. As taxas espontáneas de

mutación son moi baixas, debido á fidelidade das ADN polimerasas e aos mecanismos de reparación do ADN, por iso non poden producir cambios de frecuencias rápidos nas poboacións (figura 2).

Efectos similares aos da mutación posúe a **recombinación**, outro proceso que aumenta a variabilidade xenética, e que se refire ao intercambio de ADN entre os compoñentes dun par cromosómico, normalmente na meiose.



**Figura 2:** Exemplo do posible efecto de tres forzas evolutivas na frecuencia das variantes dun carácter, partindo dunha poboación inicial hipotética ata chegar a unha posible poboación final tras varias xeracións. Imaxe tomada de (Fondevila 2009).

### 1.3 Os polimorfismos de ADN

A variabilidade do ADN preséntase en forma de diferentes alelos ou posibilidades nun *locus* particular. A pesar de que tradicionalmente se denomina alelo a cada unha das formas alternativas dun xene, que ocupa un *locus* concreto, tamén se pode usar, de maneira xeral, referido a calquera secuencia de ADN e, por tanto, ás variantes dunha rexión polimórfica. Un polimorfismo implica a existencia de dúas ou máis variantes (de secuencias de ADN, de fenotipos, de proteínas, etc.) cunha

frecuencia significativa na poboación. No caso do ADN, segundo a definición histórica, o alelo con menor frecuencia debe atoparse na poboación cunha frecuencia maior ou igual ao 1%, en caso contrario, estaríamos falando de mutación e non de polimorfismo. Acéptase así a definición de polimorfismo de E. B. Ford de 1940, que consideraba polimorfismo como "a presenza nunha poboación de dúas ou máis formas alélicas discretas, das cales, a de menor frecuencia, non pode manterse só por mutación recorrente" (Ford 1940). Actualmente considérase unha clasificación das variantes xenéticas máis detallada, na que podemos citar como variantes moi raras aquelas nas que a frecuencia do alelo menor (MAF, do inglés *Minor Allele Frequency*) é inferior ao 0,1%, variantes raras nas que  $0,1\% < \text{MAF} < 0,5\%$ , variantes de baixa frecuencia nas que  $0,5\% < \text{MAF} < 5\%$ , e variantes comúns ás que dita frecuencia é maior ao 5% (Manolio et al. 2009).

Os primeiros marcadores de variabilidade xenética que se utilizaron foron os polimorfismos proteicos, tamén chamados "polimorfismos clásicos". Descuberto por Landsteiner en 1900, o sistema sanguíneo ABO foi o primeiro polimorfismo empregado no estudo da diversidade xenética humana. Co tempo, empezáronse a empregar polimorfismos xenéticos como os RFLP (*Restriction Fragment Length Polymorphism*), seguidos polos STR autosómicos, os polimorfismos presentes no ADN mitocondrial (ADNmt) e no cromosoma Y e os SNP (*Single Nucleotide Polymorphisms*). Actualmente os polimorfismos clásicos non se utilizan debido a súa baixa variabilidade, xa que están sometidos a unha forte presión selectiva.

Pódese considerar que existen, *grosso modo*, dúas formas posibles de variación: polimorfismos de secuencia e polimorfismos de lonxitude, segundo nos refiramos a cambios nos nucleótidos dunha secuencia de ADN ou a cambios na lonxitude de dita secuencia, respectivamente. Os polimorfismos de lonxitude son os máis frecuentes no ADN microsátelite, onde as variacións entre os alelos débense a diferenzas no número de copias da unidade repetitiva. Con todo, os polimorfismos máis frecuentes no xenoma son os SNP, que son polimorfismos de secuencia. Ambos tipos de polimorfismos, os SNP e os STR, son os máis utilizados á hora de estudar a variabilidade xenética humana, sexa no campo da xenética forense como no da xenética poboacional.

### **1.3.1 Haplotipos e desequilibrio de ligamento**

Un haplotipo fai referencia á combinación dos estados alélicos de marcadores polimórficos ao longo dunha molécula de ADN, é dicir, no mesmo cromosoma, ou ben, no ADN mitocondrial (Jobling 2004). Pode referirse a calquera tipo de polimorfismo, desde unha substitución de base a un polimorfismo de lonxitude.

O xenotipado de marcadores de cromosoma Y ou de ADNmt proporciona directamente un haplotipo, xa que son moléculas haploides. Ocorre o mesmo co xenotipado do cromosoma X en varóns, debido a que estes só posúen un único cromosoma X. No caso de mulleres para marcadores de cromosoma X, ou no caso de calquera marcador de cromosomas autosómicos en ambos sexos, non se obteñen os haplotipos directamente (excepto se os dous cromosomas teñen o mesmo, é dicir, en homocigosidade), se non que se obtén un xenotipo diploide, formado pola combinación dos dous haplotipos (Jobling 2004).

Debido a que o ADNmt e a maior parte do cromosoma Y non son recombinantes, a diversidade haplotípica nestas moléculas débese principalmente a mutacións. No resto do xenoma, a variabilidade haplotípica vese influenciada tamén pola recombinación.

O desequilibrio de ligamento é a propiedade pola cal alelos específicos de *loci* separados tenden a herdarse conxuntamente, isto produce asociacións alélicas na poboación. Esta característica fai que poida considerarse que o xenoma humano posúe unha estrutura en bloques haplotípicos; estes encontraríanse separados por rexións cunha taxa de recombinación significativamente maior á media do xenoma, os chamados puntos quentes de recombinación ou “*hotspots*”.

### **1.3.2 Os microsatélites**

#### **1.3.2.1 Definición e características xerais**

Os microsatélites, tamén coñecidos como marcadores STR (*Short Tandem Repeats*), son secuencias de ADN que conteñen un número de repeticións curtas en tándem normalmente inferior a 50 e que, frecuentemente, son marcadores polimórficos. As unidades de repetición dos microsatélites poden variar entre 1 e 8 pb, sendo os STR máis usados os formados por unidades de repetición de 3 ou 4 pb. Existen tamén

microsatélites compostos que están formados por unidades de repetición de distinta lonxitude (Jobling 2004).

A maioría dos STR atópanse distribuídos ao longo de todo o xenoma e considéranse selectivamente neutros, aínda que se describiron algúns con efectos fenotípicos, agrupados en zonas determinadas. Así, hai varios STR asociados ao desenvolvemento de doenzas, como a enfermidade de Huntington, varios tipos de ataxias, e o síndrome do cromosoma X fráxil, entre outras (Jobling 2004). Actualmente considérase que os microsatélites non están distribuídos de forma aleatoria por todo o xenoma eucariota, e mostran diferentes propiedades en rexións xenómicas con diferente funcionalidade (Katti et al. 2001).

O uso dos microsatélites como marcadores polimórficos de ADN foi estendéndose ao longo do tempo, usándose en cada vez máis especies e en moitos tipos de estudos, como os de cartografía xenética, inestabilidade xenómica en cancro, bioloxía da conservación, antropoloxía molecular, xenética de poboacións, xenética forense e estudos de historia evolutiva humana (Bhargava et al. 2010). O custo inicial asociado ao estudo destas rexións xenómicas é elevado xa que se require información da secuencia de ADN, así e todo, unha vez desenvolvidas as técnicas para a súa análise, estas son estables e poden ser reproducidas en distintos laboratorios dun xeito sinxelo (Maughan et al. 1995). A facilidade de manexo, a alta reproducibilidade, o baixo custo e a abundancia dos STR no xenoma de todos os organismos, fai dos microsatélites uns marcadores ideais para a análise xenética (Bhargava et al. 2010).

### **1.3.2.2 Mecanismo e taxa de mutación**

A dinámica de mutación dos microsatélites aínda non está totalmente comprendida, é unha cuestión sometida a debate e que aínda suscita gran controversia (Bhargava et al. 2010). Desenvolvéronse así, varios modelos que intentaron explicar a evolución dos STR, o máis aceptado é o modelo de mutación “paso a paso” ou “*stepwise*” (SMM: *Stepwise Mutation Model*) (Ohta et al. 1973), modelo segundo o cal a lonxitude dos STR varía por unidades, a unha taxa constante e independentemente da lonxitude do alelo, e coa mesma probabilidade de expansión que de contracción. Tendo en conta características destes marcadores, como é a correlación positiva que existe

entre a lonxitude alélica e a mutabilidade, este modelo actualmente considérase demasiado simplista (Jobling 2004).

Hai moitos modelos desenvolvidos ao respecto, como o modelo do alelo infinito (Kimura et al. 1964), o modelo en dúas fases mutacionais (Di Rienzo et al. 1994) ou o modelo do alelo K (Crow et al. 1970). Foise contrastando así, en primeiro lugar, se a mutación se produce nunha ou en dúas fases, en segundo lugar se a taxa de mutación é ou non constante e, por último, se existe ou non un nesgo mutacional polo que a probabilidade de mutación dependería, dunha u doutra maneira, da lonxitude do motivo repetitivo (Bhargava et al. 2010).

A pesar de que o mecanismo de mutación dos microsátélites aínda se descoñece, está amplamente aceptado que as mutacións ocorrerían, principalmente, como consecuencia dunha replicación do xenoma onde se produce unha escorregada, coñecida en inglés como "*slippage*". Este fenómeno é un erro que ocorre durante a replicación do ADN e prodúcese cando a ADN polimerasa se despraza sobre unha rexión da cadea molde e se reensambla noutra posición, omitindo, desta maneira, algunha unidade de repetición ou engadindo algunha de máis. O resultado é que a nova cadea non ten o mesmo número de copias que a cadea molde. Este mecanismo podería explicar tamén como se orixinan os *loci* microsátélites *de novo*.

Ante a observación de que o número e a variabilidade dos *loci* STR da rexión non recombinante do cromosoma Y (microsátélites haploides) non son significativamente menores que no resto dos cromosomas, considérase case inexistente o papel da recombinación *intercromosómica*, e acéptase como mecanismo mutacional maioritario en dita rexión, o "*slippage*" (Kayser et al. 2000).

Sobre a variabilidade dos *loci* microsátélites, ademais dos procesos xa descritos, tamén inflúe a mutación de secuencia, constitúe unha proba diso a presenza de alelos intermedios. Considérase que estes alelos son consecuencia dun evento mutacional, que pode xerar un novo alelo cun tamaño intermedio entre dous preexistentes.

De acordo co observado ata agora, os microsátélites non posúen unha taxa de mutación constante, esta tende a diferir entre *loci*, entre alelos e mesmo entre especies (Ellegren 2000). A taxa de mutación depende en parte das características intrínsecas dos STR, como o número de unidades de repetición, a lonxitude e o motivo repetitivo; tamén inflúen características xenómicas que varían localmente, como as taxas de recombinación e de substitución locais, e mesmo inflúen outros procesos como son a

transcrición, a heterocigosidade, a posición cromosómica, a división celular, o sexo e o contido en GC das rexións flanqueantes do STR (Bhargava et al. 2010).

En xeral, os STR teñen elevadas taxas de mutación se se comparan coas mutacións que se producen nos xenes codificantes. O rango de variación de dita taxa vai desde  $10^{-6}$  a  $10^{-2}$  eventos por *locus* por xeración (Bhargava et al. 2010). A taxa de mutación dos STR presentes no cromosoma Y exporase no apartado 1.5.1.1.

### 1.3.2.3 Métodos de detección e artefactos

O primeiro paso para detectar os STR consiste na localización destes no xenoma; actualmente isto realízase mediante buscadores como o “*Tandem Repeats Finder*” (Benson 1999), que se despraza ao longo dunha secuencia de ADN e detecta unidades de repetición adxacentes. Posteriormente, análises experimentais sobre a variación de lonxitude de moitas mostras de ADN xenómico, demostrarán se os *loci* son ou non polimórficos e, por tanto, se posúen ou non utilidade para estudos de variabilidade xenética.

No xenotipado dos STR utilízanse dúas técnicas amplamente estendidas: a PCR (*Polymerase Chain Reaction*) e a electroforese. A PCR ou reacción en cadea da polimerasa, descrita en 1986 por Kary Mullis, considérase un dos descubrimentos máis revolucionarios da ciencia. A PCR é unha técnica que permite a amplificación exponencial dunha secuencia específica de ADN *in vitro*, mediante o uso de cebadores ou “*primers*” oligonucleotídicos que flanquean a rexión a amplificar, unha ADN polimerasa termoestable e un protocolo con temperatura cíclica, fundamentalmente. Os métodos de xenotipado baseados na PCR varreron rapidamente aos que se baseaban nos polimorfismos de lonxitude de fragmentos de restrición (RFLP), xa que os primeiros permiten obter bos resultados con menor cantidade de mostra e de peor calidade, en menos tempo e, ademais, permiten o procesado e automatización dun maior número de mostras á vez (Butler 2001).

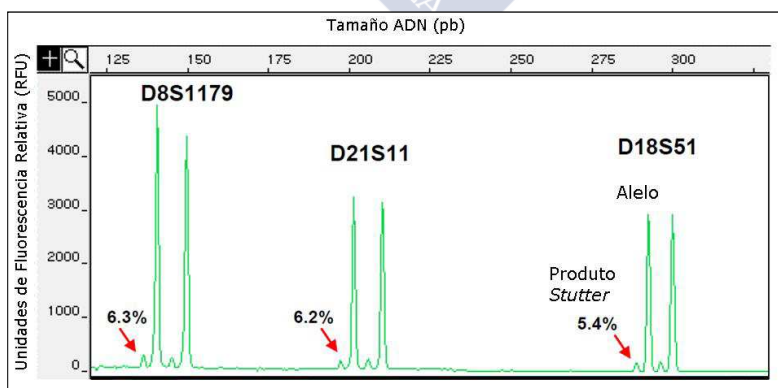
Nun principio, a detección dos *loci* microsátélites unha vez amplificados levábase a cabo, normalmente, mediante a autorradiografía dun xel de poliacrilamida no que tiña lugar unha electroforese dos produtos de PCR marcados radiactivamente ou mediante unha tinguidura de prata de dito xel. Tamén se utilizaban xeles de agarosa, con

produtos de PCR marcados con bromuro de etidio que eran detectables baixo luz ultravioleta.

A dispoñibilidade de marcaxe do extremo 5' dos cebadores usados na reacción de PCR con diferentes marcadores fluorescentes permitiu que os produtos de PCR puideran ser detectados en plataformas de secuenciación capilar, utilizando unha tecnoloxía láser. Isto posibilitou, ademais, a amplificación de varios *loci* microsatélites nunha única reacción, e a súa posterior separación e detección simultánea, o que se denomina reacción de PCR *multiplex*. O uso de *multiplexes* de microsatélites autosómicos proporciona alta sensibilidade, alta especificidade e alto poder de discriminación, permitindo xenotipar varios STR nunha única reacción, con cantidades limitadas de ADN e mostras degradadas.

Durante o proceso de amplificación dos microsatélites pódense producir varios tipos de artefactos que probablemente interfiran na interpretación e no xenotipado dos alelos presentes no ADN molde (Butler 2001). Os máis frecuentes son os que seguen:

- **Produtos “stutter” ou tartamudos:** obsérvanse no electroferograma como picos cun tamaño uns pares de bases menor que o alelo xenotipado, xeralmente unha unidade de repetición menor (figura 3). Débense a un erro de “*slippage*” producido cando a ADN polimerasa está replicando o STR en cuestión. No xenotipado dos microsatélites usados con fins forenses, estes artefactos poden supoñer un gran problema cando se traballa cunha mestura de mostras de varios individuos, xa que o produto tartamudo podería ser indistinguible dun alelo pertencente ao compoñente minoritario da mestura.



**Figura 3:** Alelos de STR con produtos tartamudos ou “stutter” indicados con frechas. Imaxe modificada a partir de (Butler 2001).

- **Alelos nulos:** artefactos producidos por existir un cambio de base, unha inserción ou unha deleción no sitio de unión do cebador empregado na amplificación dun STR. Son pouco frecuentes xa que as secuencias que flanquean aos *loci* STR adoitan ser estables e constantes entre mostras (Butler 2001). A este fenómeno denomínaselle en inglés “*allele drop-out*” e un dos maiores problemas que pode causar é que, para un alelo determinado, un heterocigoto aparente ser un homocigoto.
- **Adición nucleotídica “extra”:** frecuentemente as enzimas ADN polimerasas, sobre todo a Taq polimerasa (usada na reacción de PCR), engaden un nucleótido no extremo 3’ do produto de PCR que non está na cadea molde. O máis frecuente é que se engada unha adenina, por iso a este fenómeno adóitaselle chamar “**adenilación**”. Isto fai que o produto de PCR teña unha base máis que a secuencia diana, e no electroferograma se poderán ver picos alélicos cunha pequena fenda no centro (“*split peaks*”) ou ben dous picos seguidos (-A e +A), separados por unha base, de maior ou menor altura dependendo da eficiencia da reacción de adenilación (Butler 2001).
- **Bandas inespecíficas:** aparecen debido a que durante a reacción de PCR producíronse hibridacións inespecíficas dos cebadores.

#### 1.3.2.4 Estandarización e bases de datos

A estandarización do tipo de marcadores xenéticos, das tecnoloxías e dos métodos de análise empregados polos laboratorios, é unha característica esencial para que sexa posible realizar contra pericias. Con este propósito formáronse varios comités, entre os que destacan o EDNAP (*European DNA Profiling Group*) en Europa e o SWGDAM (*Scientific Working Group on DNA Analysis and Methods*) en América, constituídos por representantes de varios países e laboratorios.

Por outra banda, é necesario que os laboratorios que realizan pericia xenética cumpran uns requisitos de calidade. Deste modo, grupos como o *Grupo de Habla Española y Portuguesa* da *International Society for Forensic Genetics* (GHEP-ISFG), que aglutina a laboratorios de 19 países a día de hoxe, coordina exercicios de validación entre laboratorios periodicamente, acreditando a calidade dos laboratorios que superan ditos exercicios. A ISFG, ademais, nomeou unha comisión de ADN (“*DNA*

*Commission*”) para publicar regularmente guías e recomendacións sobre os polimorfismos de ADN e a súa aplicación aos problemas na identificación humana.

Paralelamente, estendeuse o uso dos mesmos marcadores na maioría de estudos, para facilitar a reproducibilidade e a comparación de resultados entre os laboratorios de casuística forense. Desde 1997 o FBI utiliza, para as probas de identidade humana, un conxunto de 13 *loci* de marcadores STR autosómicos chamado *COmbined DNA Index System* (CODIS). Na información xenética que se foi reunindo ao longo destes anos predomina a relativa a este conxunto de marcadores STR, debido ao seu uso na base de datos nacional de ADN do Reino Unido (NDNAD) e tamén noutras bases de datos de ADN de xustiza criminal de todo o mundo (Butler 2006). Na actualidade, a maioría dos laboratorios utilizan *kits* comerciais de *multiplexes* de STR en casuística forense, xa que permiten obter un rendemento máis homoxéneo e un maior control e uniformidade dos resultados. Os *kits* usados maioritariamente ata hai uns anos para STR autosómicos eran o *AmpFISTR® Identifiler® PCR Amplification Kit* (Life Technologies) e o *PowerPlex® 16 System* (Promega), porén, actualmente estendeuse o uso doutros como o *AmpFISTR® Identifiler® Plus PCR Amplification Kit* e *AmpFLSTR® SGM Plus® PCR Amplification Kit* (Life Technologies) e o *PowerPlex® ESX e ESI Systems* (Promega).

Non obstante, máis recentemente, a rede europea de institutos de ciencias forenses chamada ENFSI (*European Network of Forensic Science Institutes*), publicou novas recomendacións nas que incorporaban cinco novos STR ao denominado “*European Standard Set*” (ESS), un conxunto de sete *loci* microsátélites utilizado desde o ano 2001 de maneira xeneralizada en toda Europa (The-Council-of-the-European-Union 2001) e que inclúe os *loci* TH01, vWA, FGA, D21S11, D3S1358, D8S1179 e D18S51. Isto provocou que varias empresas foran comercializado novos *kits* que inclúen estes novos marcadores, e pouco a pouco parece que se están introducindo na rutina dos laboratorios forenses (Phillips et al. 2010). Os cinco novos marcadores do ESS son os chamados D1S1656, D2S441, D10S1248, D12S391 e D22S1045, considéranse *loci* mini-STR e serán tratados con maior profundidade no seguinte apartado.

O uso dos STR no cartografado xenético e nas probas de identidade humana foise expandindo enormemente nos últimos anos, sen que houbera un lugar concreto onde se puidese atopar información sobre os *loci* STR utilizados en xenética forense. Fíxose necesaria así, unha revisión da abundante literatura descrita ata o momento, para

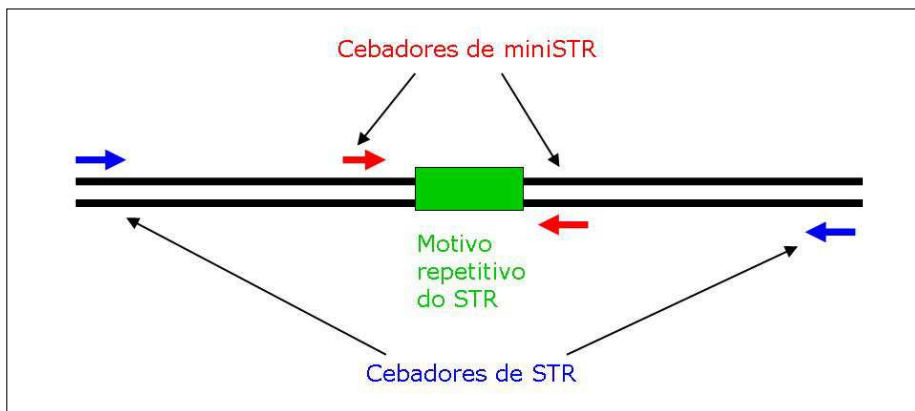
facilitar o futuro traballo neste campo. Por iso, a partir de 1997, o *National Institute of Standards and Technology* (NIST) foi reunindo na chamada “*Short Tandem Repeat DNA Internet Database*”, comunmente denominada *STRBase* (Butler 1997), aos STR máis utilizados como marcadores forenses (Ruitberg et al. 2001). Neste sitio *web* ([www.cstl.nist.gov/strbase/](http://www.cstl.nist.gov/strbase/)) pódese atopar a información relativa á secuencia dos STR, aos cebadores e ás condicións utilizadas nas reaccións de PCR, ás tecnoloxías empregadas, ver cales son os *multiplexes* máis utilizados, e mesmo hai datos de frecuencias en diferentes poboacións. Existe outro tipo de bases de datos de marcadores STR, as relativas aos existentes no cromosoma Y, que serán comentadas no apartado 1.5.1.

### 1.3.2.5 Os marcadores mini-STR

Nas investigacións que se levan a cabo en xenética forense é moi común dispoñer de mostras con ADN altamente degradado como evidencia biolóxica. Por este motivo, nos últimos anos aumentáronse os esforzos destinados a incrementar a posibilidade de amplificar ADN altamente degradado, máis que a aumentar o poder de discriminación das técnicas existentes (Dixon et al. 2006). Neste contexto é onde apareceron os mini-STR, microsátélites cun amplicón menor que os STR convencionais (figura 4). A utilización destes marcadores, por tanto, aumenta a probabilidade de obter produto amplificado en mostras nas que o ADN está degradado. A maioría dos mini-STR deseñáronse simplemente desprazando os sitios de unión do cebador da PCR a unha posición máis próxima ao motivo de repetición do STR, así se podían manter os *loci* usados nas bases de datos internacionais.

Dende que os mini-STR foron descritos no ano 2003 (Butler 2003) ata a actualidade, leváronse a cabo multitude de estudos xenéticos utilizando estes marcadores (Abrahams et al. 2010; Asamura et al. 2007; Butler 2003; Coble and Butler 2005; Dixon et al. 2005; Grubwieser et al. 2006; Hill et al. 2008; Martin et al. 2007; Santos et al. 2010; Vullo et al. 2010; Welch 2011; Senge et al. 2011; Raimann et al. 2012; Jin et al. 2013; Previdere et al. 2013). Deste xeito, os *multiplexes* de mini-STR axudaron a recuperar información de mostras con ADN degradado que cos STR convencionais non se poderían xenotipar ou ben só se poderían obter perfís xenéticos parciais. A inclusión deste tipo de *locus* microsátélite no ESS permite que, mostras de casos forenses complicados nos que só se dispón de pequenas cantidades de ADN en

estado degradado, poidan ser xenotipadas completamente, evitando taxas inaceptables de coincidencias espontáneas debido ao incremento de perfís parciais nas bases de datos (Schneider 2009). Así e todo, nas mostras máis degradadas o problema persiste, por iso, buscáronse outras técnicas que usen amplicóns aínda máis reducidos, como é o xenotipado dos SNP.



**Figura 4:** Cebadores de STR e de mini-STR para un mesmo motivo de repetición. Imaxe modificada a partir de <http://www.cstl.nist.gov/strbase/miniSTR.htm>.

### 1.3.3 Os marcadores SNP

#### 1.3.3.1 Definición e características xerais

Os polimorfismos de nucleótido único (*Single Nucleotide Polymorphisms*) ou marcadores SNP, son substitucións dunha base nucleotídica da molécula de ADN. Tamén se inclúen na categoría de SNP ás insercións ou delecións (*indels*) dunha única base, con todo, os mecanismos que subxacen baixo estes polimorfismos difiren dos das substitucións de base (Jobling 2004).

Os SNP denomínanse tamén polimorfismos binarios ou bialélicos xa que, na súa maioría, só teñen dúas variantes, a variante ancestral e a variante derivada ou mutada. Dada a coñecida filoxenia entre os humanos e os grandes simios, pódese coñecer cal é o alelo ancestral para un SNP en humanos, determinando o alelo que posúen os grandes simios na secuencia homóloga á humana (Jobling 2004).

Son os polimorfismos máis frecuentes do xenoma humano, aparecendo un SNP cada 1.000 bases aproximadamente, segundo os resultados obtidos polo Proxecto

Xenoma Humano e Celera Genomics (Sachidanandam et al. 2001; Venter et al. 2001). Dependendo de onde apareza, terá diferentes consecuencias a nivel fenotípico, aínda que a maioría dos SNP atópanse en rexións non codificantes do xenoma, por tanto non teñen un impacto directo coñecido no fenotipo do individuo, e son os que se utilizan como marcadores xenéticos en estudos poboacionais e evolutivos (Sylvänen 2001).

Principalmente podemos diferenciar catro tipos de SNP con interese forense e poboacional (Budowle et al. 2008):

- **Marcadores SNP usados en probas de identidade:** posúen elevada heterocigosidade e baixo coeficiente de endogamia. Para lograr inferir a identidade cos SNP requírese un maior número de marcadores que no caso de utilizar os STR.
- **Marcadores SNP informativos de liñaxe:** encóntranse no ADNmt e no cromosoma Y. Son moi informativos en estudos evolutivos e nas análises de parentesco, sobre todo nos casos nos que a mostra de referencia e a evidencia están separadas por varias xeracións.
- **Marcadores SNP informativos de ascendencia:** coñecidos polas siglas en inglés “AIMs” (*Ancestry Informative Markers*), establecen cunha elevada probabilidade a orixe bioxeográfica do antepasado do individuo. Son marcadores SNP con baixa heterocigosidade e elevado coeficiente de endogamia.
- **Marcadores SNP informativos de características fenotípicas:** informan sobre determinados trazos fenotípicos dun individuo, como pode ser a pigmentación (cor de pel, de ollos e de cabelo).

Nos últimos anos, o estudo dos SNP no xenoma humano, espertou un gran interese debido, por unha parte, á información que proporcionan sobre a historia evolutiva dun *locus* determinado (incluíndo o impacto da selección natural) e, por outra, debido a que permiten encontrar mutacións que subxacen a doenzas humanas monoxénicas e complexas (Jobling 2004). Estes estudos víronse favorecidos cos avances de proxectos internacionais como o Proxecto HapMap (The-International-HapMap-Consortium 2003), que consistiu na elaboración dun mapa haplotípico do xenoma humano, o proxecto 1000 xenomas (1000-Genomes-Project-Consortium et al.

2012) no que se describen as variantes xenéticas obtidas no secuenciamento do xenoma de 1.092 individuos, ou tamén o proxecto Exome Sequencing Project (<http://evs.gs.washington.edu/EVS/>) pioneiro na aplicación da secuenciación de nova xeración nas rexións do xenoma que codifican proteínas, coa finalidade de descubrir novos xenes e mecanismos implicados en diversas doenzas.

### 1.3.3.2 Mecanismo e taxa de mutación

Fundamentalmente existen dous procesos que dan lugar aos cambios nunha soa base (SNP): a non incorporación de nucleótidos durante a replicación e a mutaxénese, que pode producirse por unha modificación química das bases ou por un dano físico debido a, por exemplo, radiación ultravioleta ou ionizante (Jobling 2004). As mutacións que afecten ás células da liña xerminar e non sexan resoltas polos mecanismos celulares de reparación do dano no ADN, serán as que se poidan transmitir á descendencia.

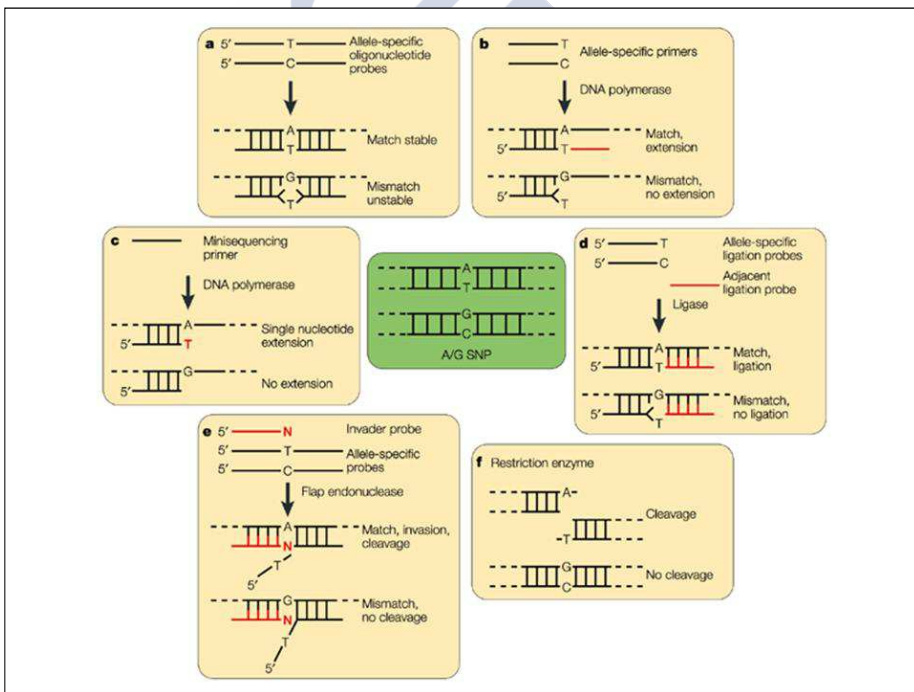
As taxas de mutación dos SNP son da orde de  $10^{-8}$  mutacións por *locus* e por xeración, polo tanto, son polimorfismos moito máis estables que os microsátélites, cuxa taxa de mutación é da orde de  $10^{-3}$  mutacións por *locus* e por xeración. Porén, encontráronse diferenzas significativas nesta taxa dependendo do tipo de SNP do que se trate, así por exemplo, os *indels* son menos frecuentes que as substitucións de base, as transversións menos frecuentes que as transicións, e os SNP “CpG” posúen unha taxa de mutación maior que o resto (Nachman et al. 2000).

A baixa taxa de mutación das substitucións de base implica que sexa case inexistente a aparición dunha mutación recorrente, dada a escala temporal da evolución dos humanos modernos (Jobling 2004). Por este motivo, a coincidencia do mesmo alelo para un SNP determinado en dúas persoas diferentes implicaría, na maioría dos casos, unha coincidencia por descendencia e non por estado, é dicir, significaría que terían herdado dito alelo dun devanceiro común, e non se debería a un fenómeno de converxencia evolutiva. Non obstante, documentáronse xa varias mutacións recorrentes, como é o caso do primeiro SNP descuberto no cromosoma Y, SRY10831/SRY1532, que mutou dúas veces ao longo da historia da nosa especie (Hammer et al. 1998).

### 1.3.3.3 Métodos de detección e artefactos

A día de hoxe, existen multitude de métodos de xenotipado dos SNP, e na maioría deles subxace un dos seguintes principios de reacción bioquímicos (Syvänen 2001) (figura 5):

- Hibridación con oligonucleótidos alelo específicos (*ASOs, allele-specific oligonucleotides*).
- Extensión de cebadores alelo específicos (*Allele-specific primer extension*).
- Extensión de cebadores dun único nucleótido ou minisequenciación (*Single nucleotide primer extension*).
- Ligadura oligonucleotídica (*Oligonucleotide ligation*).
- Fragmentación invasiva (*Invasive cleavage*).
- Fragmentación en lugares de restrición (*Restriction site cleavage*).



**Figura 5:** Detección do alelo A dun SNP producido por unha transición dunha adenina por unha guanina, segundo os seis principios bioquímicos mencionados no texto. A detección do alelo G faríase analogamente nunha reacción paralela. Imaxe obtida de (Syvänen 2001).

Nos últimos anos desenvolveuse un gran número de tecnoloxías diferentes para o xenotipado dos SNP, baseadas nos distintos métodos de discriminación alélica existente, anteriormente mencionados. Considérase que hai dúas categorías dependendo do tipo de análise necesaria: reaccións homoxéneas, cando ocorren en disolución, e reaccións nun soporte sólido (Sobrinó et al. 2005). As reaccións homoxéneas son xeralmente máis axeitadas para a automatización, xa que non hai pasos de separación ou de purificación despois da reacción da discriminación alélica, pero presentan unha limitación á hora de facer reaccións en *multiplex*, ao contrario das reaccións nun soporte sólido (Sobrinó et al. 2005).

Desta maneira, encontramos métodos de detección dos marcadores SNP por fluorescencia, por luminescencia, por espectrometría de masas, por fluorescencia polarizada e por transferencia de enerxía por resonancia de fluorescencia (FRET), principalmente. Entre as tecnoloxías máis usadas encóntrase unha das utilizadas nos traballos desta tese, o *SNaPSHOT*<sup>®</sup> (*Life Technologies*).

É moi común que, durante o xenotipado dos SNP, aparezan artefactos que poidan provocar unha falsa asignación alélica ou un falso positivo ou negativo. Os artefactos máis comúns que aparecen son o **desequilibrio alélico** (gran diferenza de tamaño entre os distintos alelos), os **alelos nulos** ou “*drop-out*” (ausencia dun alelo no electroferograma), os **alelos “drop-in”** (aparición dun alelo non real no electroferograma) e a aparición de **bandas inespecíficas** (produtos de reacción inespecíficos, que non corresponden cun alelo). Por outra banda, é frecuente que apareza un nesgo de mostraxe no xenotipado, é dicir, unha distorsión sistemática no conxunto de datos obtidos, debido a como foron seleccionados os marcadores, sobre todo se os *loci* a estudar presentan unha gran diferenciación entre poboacións (Jobling 2004). Desta maneira, por exemplo, moitos estudos temperáns utilizando marcadores SNP subestimaban a diversidade existente en poboacións non europeas, debido á errónea escolla dos marcadores (Jobling 2004). Os resultados, pois, débense interpretar con precaución e tendo en conta este tipo de erros, así como os posibles nesgos no xenotipado.

### 1.3.3.4 Estandarización e bases de datos

Actualmente hai máis de 62 millóns de marcadores SNP caracterizados no xenoma humano, e 44 millóns deles están validados, segundo a información da base de datos dbSNP. Previamente ao xenotipado destes marcadores, é habitual que se realice unha busca dos SNP de interese nalgunha das bases de datos dispoñibles hoxe en día, entre as que destacan as seguintes:

- dbSNP (*database of Single Nucleotide Polymorphisms*):  
<http://www.ncbi.nlm.nih.gov/SNP/index.html>.
- Ensembl: <http://www.ensembl.org/index.html>.
- HapMap (levado a cabo por *The SNP Consortium* (Thorisson et al. 2003)):  
<http://hapmap.ncbi.nlm.nih.gov/>.
- ALFRED (*Allele frequency database*):  
<http://alfred.med.yale.edu/alfred/index.asp>.
- HGVbase (*Human Genome Variation database*) actualmente incluída na base de datos *GWAS Central*: <https://www.gwascentral.org/>.
- 1000 Genomes: <http://www.1000genomes.org/>.
- *Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP)*, Seattle, WA (ESP): <http://evs.gs.washington.edu/EVS/>.

As mesmas sociedades de estandarización citadas no apartado 1.3.2.4 que establecen as normas para o emprego dos marcadores STR na xenética forense, nos últimos anos, tamén traballan nas recomendacións a seguir na utilización de marcadores SNP. Un exemplo de estandarización realizado pola EDNAP é un dos resultados que se presenta nesta memoria de tese (Dixon et al. 2006).

No grupo da GHEP-ISFG anteriormente mencionado, mesmo existe unha comisión de traballo centrada nos marcadores SNP en xenética forense. Actualmente, este grupo de traballo pretende prestar apoio teórico e técnico en materia de marcadores SNP a todos os laboratorios membros do GHEP-ISFG que o soliciten, así como promover un foro aberto de discusión e debate sobre aspectos relacionados coa análise dos SNP na casuística forense. Ademais, o grupo desenvolve exercicios entre laboratorios adicionais aos propios controis de calidade do GHEP-ISFG.

### 1.3.3.5 **Multiplexes de marcadores SNP: o 52-plex**

A incorporación do xenotipado dos SNP á resolución de casos forenses non foi proposta como unha substitución das técnicas existentes (os marcadores STR e o ADN mitocondrial principalmente), se non como un complemento das mesmas. Na actualidade está amplamente aceptado que, a curto ou medio prazo, os SNP non substituirán aos STR como método de análise de mostras forenses (Gill et al. 2004). Isto é debido, principalmente, ás vantaxes que presentan estes últimos marcadores fronte aos primeiros, entre as que destacan a existencia de extensas bases de datos de mostras xenotipadas cos STR, o elevado custo que aínda supón realizar xenotipados masivos de marcadores SNP no campo forense, o requirimento do xenotipado dun menor número de marcadores (no caso dos STR) para a maioría das aplicacións forenses e a maior facilidade coa que se poden interpretar mesturas, debido a que cada *locus* dun STR contén múltiples alelos, ao contrario dos SNP. Non obstante, os SNP posúen outras vantaxes como son a ausencia de determinados artefactos no seu xenotipado (os produtos tartamudos), o cal é beneficioso na interpretación de determinados escenarios, a capacidade de análise en mostras altamente degradadas, e as aplicacións concretas de determinados marcadores SNP, como son os que se atopan no ADN mitocondrial (utilizados principalmente para xenotipar ósos, dentes ou pelos sen raíz), no cromosoma Y (empregados sobre todo para coñecer o compoñente masculino dunha mestura ou dilucidar a posible orixe poboacional da mesma) e os que se asocian a características físicas (como a cor do pelo, da pel e dos ollos) (Gill et al. 2004).

Por outra banda, na análise de relación xenética en casos de ausencia do presunto pai, os marcadores STR teñen desvantaxes debidas á, ás veces escasa, eficiencia bioestatística, así como á posibilidade de observar unha ou máis inconsistencias xenéticas que tamén puideran ser explicadas por eventos de mutación. Nestas situacións, poderían empregarse os marcadores SNP e *indels* como marcadores adicionais, xa que a insignificante taxa de mutación que posúen permitiría aportar a información xenética adicional decisiva a favor ou en contra da relación asumida (Schneider 2012). Tamén debido á menor taxa de mutación dos SNP, de xeito xeral, permiten descubrir máis facilmente se existe subestruturación nunha poboación determinada, xa que cos STR esta podería pasar desapercibida (Moreno et al. 2014).

Recentemente conseguíuse, ademais, mellorar a capacidade de análise de ADN de mostras extremadamente degradadas, onde o xenotipado con STR pode non ser

suficiente para acadar un adecuado poder de discriminación, mediante a escolla de marcadores SNP localizados en rexións nucleosómicas, xa que os complexos de ADN-histonas parece que dotan ao ADN de maior protección fronte aos procesos de degradación (bacterianos, ambientais e apoptóticos) (Freire-Aradas et al. 2012).

En casuística forense, un dos principais factores limitantes é a cantidade de mostra biolóxica, debido a iso é de suma importancia obter moita información xenética co mínimo gasto de mostra de ADN. Por este motivo, a capacidade de facer un ensaio *multiplex* co maior número posible de marcadores, ten tanta transcendencia. Deste modo, nos últimos anos, deseñáronse reaccións *multiplex* dos SNP autosómicos que permiten definir a identidade dun vestixio biolóxico coa mesma probabilidade de asignación que se obtén cos *multiplexes* dos STR de *kits* de identificación utilizados comunmente en laboratorios forenses. É o caso, por exemplo, do SNPforID 52-plex (Sanchez et al. 2006) que, en mostras de determinadas poboacións, proporciona un poder de discriminación maior que moitos dos *kits* de xenotipado dos STR autosómicos (Phillips et al. 2007a).

Este ensaio está formado por 52 SNP autosómicos polimórficos en poboacións europeas, asiáticas e africanas, que se amplifican nunha mesma reacción de PCR e se detectan mediante dúas reaccións de extensión de base única (“*single base extension*”) utilizando electroforese capilar. O 52-plex superou, ademais, unha validación forense levada a cabo por varios laboratorios do consorcio europeo SNPforID (The-European-SNPforID-Consortium). Mediante esta validación se demostrou que, en determinadas circunstancias relevantes en casuística forense, como é o xenotipado de mostras degradadas, o 52-plex proporciona mellores resultados que un *multiplex* de STR (Musgrave-Brown et al. 2007).

Por outra parte, avalíouse tamén a reproducibilidade do 52-plex en diferentes sistemas de xenotipado dos SNP, obténdose mellores resultados ao utilizar a tecnoloxía *GenPlex*<sup>TM</sup> (*Life Technologies*) cando se trata de mostras compostas por mesturas de ADN de diferentes individuos (Phillips et al. 2007a), a pesar de que tiveron que escollerse unicamente 49 dos 52 *loci* de SNP de dito *multiplex*, por limitacións de dita tecnoloxía.

Paralelamente, elaboráronse numerosos *multiplexes* de SNP, non só utilizando marcadores SNP de identificación individual, como no caso do 52-plex, se non tamén,

por exemplo, con marcadores informativos de orixe poboacional, como é o caso do 34-plex (Phillips et al. 2007b), entre outros.

Na presente memoria de tese exporase, ademais, o desenvolvemento de *multiplexes* de SNP de cromosoma Y, cuxa finalidade principalmente é a predición da orixe xeográfica das liñaxes masculinas (Brion et al. 2005a).

## 1.4 O cromosoma Y

### 1.4.1 Definición e características xerais

O cromosoma Y é un dos dous cromosomas sexuais que teñen os mamíferos de xénero masculino e é o que determina dito sexo, mediante a acción do xene *SRY*. Este xene dominante masculino provoca a diferenciación das gónadas a testículos en lugar de a ovarios, nunha etapa temperá do desenvolvemento embrionario (Jobling 2004). É, por tanto, un cromosoma haploide, específico dos homes, e que se transmite exclusivamente de pais a fillos varóns.

É o único cromosoma que non posúe un homólogo, polo que se esperaríase que non houbera recombinación posible, porén, posúe dúas pequenas rexións nos seus extremos, chamadas rexións pseudoautosómicas, coas cales é posible a recombinación co cromosoma X, debido á súa elevada homoloxía. O comportamento de recombinación dos cromosomas sexuais, por tanto, non é usual: en varóns, o cromosoma X e o Y recombinan unicamente nas rexións teloméricas pseudoautosómicas e, nas mulleres, a recombinación pode ocorrer ao longo de todo o cromosoma X (Jobling 2004).

Así pois, na maior parte do cromosoma Y non existe recombinación, por iso a esa rexión se lle denomina NRY (*Non-recombining region*), rexión non recombinante do cromosoma Y ou ben MSY (*Male-specific region*), rexión específica masculina do cromosoma Y. Actualmente, emprégase máis o nome de MSY xa que algúns estudos (Skaletsky et al. 2003) cuestionaron a ausencia total de recombinación, afirmando que se podía producir un fenómeno de recombinación interna en dita rexión, a conversión xénica. Así e todo, ata o momento non se acepta á conversión xénica como unha forma de recombinación *sensu stricto*.

O proceso máis frecuente no seo do cromosoma Y é, por tanto, a conversión xénica (Rozen et al. 2003), mecanismo que se considera que protexe a integridade

xenética de dito cromosoma e que consiste no intercambio non recíproco dunha secuencia (Jobling et al. 2003), dentro da propia estrutura do cromosoma Y, neste caso. Crese que, ante a ausencia de recombinación, este mecanismo de intercambio xenético foi o que permitiu a adquisición e posterior conservación das funcións espermatoxénicas (Skaletsky et al. 2003; Rosser et al. 2009). Ademais, a diferenza da recombinación, a conversión xénica ocorre múltiples veces en cada xeración e ten lugar, sobre todo, nas secuencias amplicónicas da rexión MSY. Considérase que a conversión xénica actuou durante a evolución humana retardando o cambio de secuencia nas rexións palindrómicas onde ten lugar, xa que se comprobou que dito proceso foi significativamente nesgado cara o estado ancestral ademais de cara a fixación de nucleótidos G ou C (Hallast et al. 2013).

Por outra banda, existen tamén numerosos eventos de conversión xénica entre secuencias da rexión MSY do cromosoma Y e o cromosoma X (Cruciani et al. 2010; Trombetta et al. 2010). A taxa de conversión entre X e Y é, ademais, moi elevada, considérase que de catro a cinco ordes de magnitude máis rápida que a taxa de mutacións de substitucións de base do cromosoma Y (Rosser et al. 2009).

Nunha orixe, os cromosomas sexuais eran un par de cromosomas autosómicos, ata que nun momento temperán da evolución dos mamíferos, un deles adquiriu a función de determinación do sexo seguido da supresión da recombinación, polo menos na maior parte de dito cromosoma, foi así como apareceu o cromosoma Y (Jobling 2004). A supresión da recombinación foi debida a unha serie de inversións que sufriu o cromosoma Y ao longo da súa historia evolutiva, en concreto, acéptase que foron polo menos catro os eventos de inversión cromosómica, o primeiro deles ocorrería hai entre 240 e 320 millóns de anos, despois da diverxencia da liñaxe dos mamíferos da das aves (Lahn et al. 1999).

Crese que o cromosoma Y segue sufrindo un proceso de dexeneración, tendo perdido xa a maioría de xenes que tería, nun principio, en mesmo número que o cromosoma X, hai uns 170 millóns de anos (Graves 2001). A perda de xenes débese á incapacidade do cromosoma Y de eliminar alelos mutantes con alelos homólogos non mutados mediante recombinación, provocándose, por tanto, a dexeneración de moitos xenes. Este proceso, que permite a acumulación de mutacións deleterias nos organismos, denomínase “*Muller’s ratchet*” e dase en organismos con reprodución

asexual, e nas partes do xenoma de organismos que se reproducen sexualmente pero que non están suxeitas a recombinación, como é o caso do MSY.

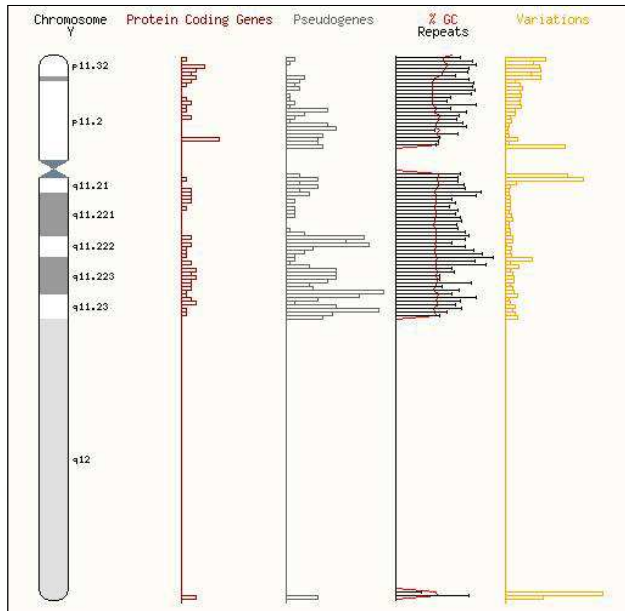
Por outra parte, estudos recentes nos que compararon as secuencias das rexións MSY do cromosoma Y de humanos e chimpancés, demostran que este foi un cromosoma que evolucionou rapidamente nos últimos 6 millóns de anos, xa que ambos cromosomas difiren radicalmente na estrutura de secuencia e no contido xénico (Hughes et al. 2010).

Ao proceso de dexeneración do cromosoma Y acompañouno outro de adquisición de novos xenes, que adoptaron as funcións específicas de varóns, como é o xene *SRY*, que se cre é unha versión degradada do xene *SOX3*, que se encontra no cromosoma X (Graves 2001). O cromosoma Y (figura 6), polo tanto, non pode conter ningún xene esencial para a vida que non comparta co cromosoma X, xa que as mulleres non o posúen, de feito, este cromosoma é moi pobre en xenes. Así e todo, un estudo recente parece demostrar que algúns xenes do cromosoma Y si serían esenciais para a viabilidade masculina e, ademais, poderían contribuír ás diferenzas na susceptibilidade a doenzas que existe entre mulleres e homes, debido a que estarían activos en células de todo o corpo (Bellott et al. 2014).

Por outra banda, a recombinación homóloga non alélica que se produce no cromosoma Y fai que este teña uns inusuais niveis elevados de polimorfismos estruturais, e varios tipos diferentes de repeticións (Jobling 2004). Os fenómenos de recombinación homóloga non alélica descritos no cromosoma Y teñen lugar entre longos amplicóns da rexión MSY, tanto entre pares de amplicóns do brazo curto do cromosoma (Yp), como entre pares de amplicóns do brazo longo (Yq) e mesmo entre pares de amplicóns, con repeticións invertidas, localizados en brazos diferentes de dito cromosoma (Lange et al. 2013).

A ausencia de recombinación na rexión MSY provoca que os polimorfismos que se encontran nela se transmitan como un único bloque de ligamento, formando todos eles un haplotipo. Considérase así, que as rexións non recombinantes dos cromosomas Y actuais posúen un antepasado común. Ademais, tendo en conta o reducido número de mutacións acumuladas en dita rexión, e aceptando a teoría da evolución dos cromosomas sexuais, considérase que o devanceiro común para as liñaxes de cromosoma Y é recente (Underhill et al. 2001; Jobling et al. 2003). Parece existir consenso cando se afirma que non se encontrou ningunha liñaxe de cromosoma Y

anterior a 200.000 anos, e que a filoxenia deste cromosoma arranca en África (Jobling et al. 2003). A estrutura xeográfica da rexión MSY unida á súa recente idade molecular, permiten utilizar dita rexión para trazar patróns de microevolución da diversidade dos humanos modernos (Underhill et al. 2001).



**Figura 6:** Idiograma do cromosoma Y, representando tamén a densidade de xenes e pseudoxenes coñecidos, a porcentaxe de GC e as variantes ao longo do cromosoma (Ensembl *release 74*, decembro de 2013; [www.ensembl.org](http://www.ensembl.org)).

Debido á súa forma de herdanza, o tamaño efectivo de poboación do cromosoma Y é tres veces menor que o do cromosoma X e catro veces menor que o de calquera cromosoma autosómico (Jobling et al. 1995). Esta característica permite un maior impacto do efecto da deriva xenética que, unido á ausencia de recombinación na maior parte do cromosoma, fai que a presión selectiva sexa elevada. Deste modo, a diversidade poboacional do cromosoma Y é moi baixa, calquera tipo de selección afectaría ao cromosoma enteiro, debido á súa ausencia de recombinación, polo que aumentaría a frecuencia dunha liñaxe máis rapidamente do esperado por deriva (Jobling et al. 2003).

Outra singularidade evolutiva que diferencia ao cromosoma Y do resto de *loci* nucleares é a alta taxa de mutación. Esta é debida á exclusiva herdanza do cromosoma

Y a través da liña xerminar masculina, a cal é moito máis mutaxénica que a liña xerminar feminina, por ter maior número de replicacións xenómicas (Haldane 1947). A comparación das taxas evolutivas de pseudoxenes ligados ao cromosoma X e a autosómicos suxiren que a taxa de mutación masculina é catro veces maior que a taxa de mutación feminina (Nachman et al. 2000).

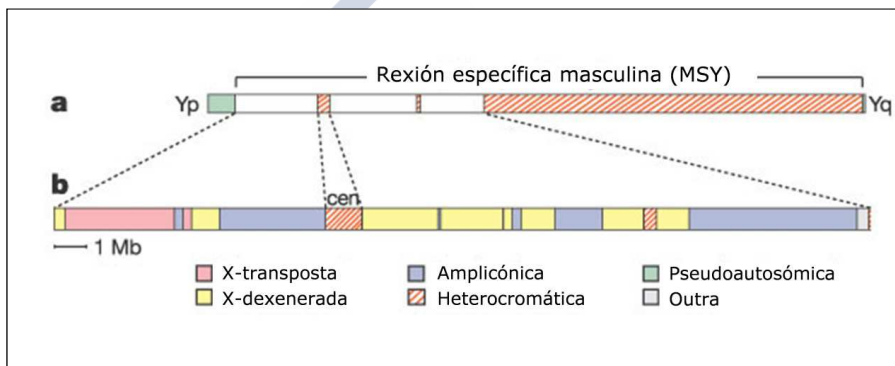
### **1.4.2 Estrutura do cromosoma Y**

O cromosoma Y é un dos cromosomas de menor tamaño en humanos, de tan só unhas 60 Mb. Como todos os cromosomas hemicigotos (tamén ocorre no cromosoma W dos sistemas ZW) tende, pois, a ser pequeno, pobre en xenes e rico en secuencias repetitivas, ao contrario que o cromosoma X que tende a semellarse máis aos autosomas, tanto en forma como en contido (Lahn et al. 2001).

Aproximadamente o 5% do cromosoma Y está formado polas rexións pseudoautosómicas mencionadas anteriormente, chamadas PAR1 e PAR2. A rexión PAR1 é a máis antiga, reflicte a orixe primitiva dos cromosomas sexuais dos mamíferos. Encóntrase no extremo do brazo curto do cromosoma Y (Yp) e posúe unha elevada taxa de recombinación co cromosoma X, o que permite a correcta segregación dos cromosomas durante a meiose (Jobling 2004). A rexión PAR2, encóntrase no extremo do brazo longo do cromosoma Y (Yq), é de moito menor tamaño que a rexión PAR1 e considérase que é unha adquisición evolutiva recente dos cromosomas sexuais humanos. A rexión PAR2 ten escasa importancia na segregación cromosómica, xa que a recombinación que sofre co cromosoma X non é obrigada, ao contrario que a rexión PAR1 (Jobling 2004). As rexións PAR semellan ser vestixios de adicións diferenciais, perdas, reorganizacións e procesos de degradación dos cromosomas X e Y, en distintas liñaxes de mamíferos (Helena Mangs et al. 2007).

A rexión específica masculina do cromosoma Y (MSY) (figura 7) constitúe o 95% da lonxitude de dito cromosoma (Skaletsky et al. 2003). Considérase un mosaico heteroxéneo de secuencias de heterocromatina (secuencias transcricionalmente inactivas, son a maior parte, unhas 40Mb), e tres tipos de secuencias eucromáticas (de menor tamaño, unhas 23Mb). As tres clases de secuencias de eucromatina presentes no cromosoma Y son (Skaletsky et al. 2003):

- “X-transpostas”: secuencias propias do cromosoma X, están no Y debido a unha transposición cromosómica. Só conteñen dous xenes, homólogos no cromosoma X.
- “X-dexeneradas”: son secuencias dexeneradas do cromosoma X, que se encontraban nos autosomas orixinais dos que proceden os cromosomas sexuais. Entre elas encóntranse varios xenes, pero só un expresado nos testículos, o xene *SRY*.
- “Amplicónicas”: secuencias propias do cromosoma Y, debidas principalmente a recombinación intracromosómica. Reúnen a maior parte das secuencias codificantes do cromosoma, e son xenes que se expresan maioritariamente nos testículos.



**Figura 7:** Rexión específica masculina do cromosoma Y (MSY). **a:** Representación esquemática do conxunto do cromosoma, incluíndo as rexións pseudoautosómicas e as heterocromáticas. **b:** Visión aumentada da rexión MSY, mostrando as tres clases de secuencias eucromáticas e heterocromáticas. Imaxe modificada a partir de (Skaletsky et al. 2003).

## 1.5 Os polimorfismos de cromosoma Y

### 1.5.1 Marcadores STR-Y

Os marcadores STR específicos de cromosoma Y (STR-Y) teñen, en xeral, as mesmas características que os que se encontran no resto de cromosomas, coa diferenza de que, como na maior parte do cromosoma Y non se produce recombinación, os STR-Y transmítense ligados como bloques haplotípicos. Esta peculiaridade fai destes

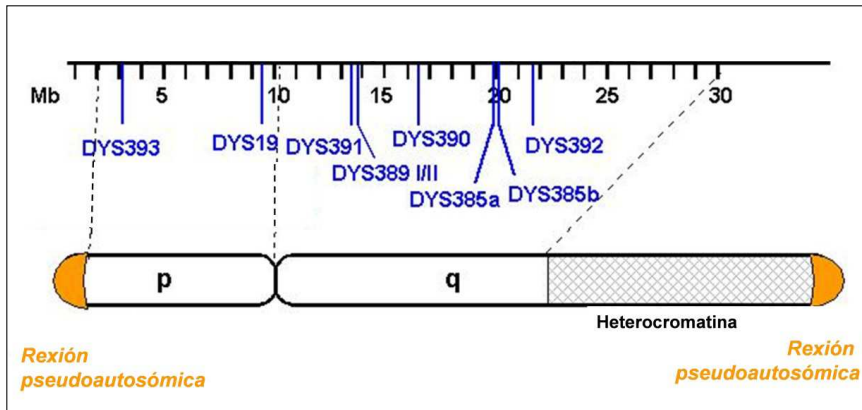
marcadores unhas ferramentas moi útiles en xenética forense e poboacional, como veremos ao describir as súas aplicacións no apartado 1.6.1 da presente tese.

Dende que se describiu o primeiro *locus* STR no cromosoma Y (Roewer et al. 1992), segundo Butler e colaboradores caracterizáronse máis de 400 STR-Y, porén, non todos son o suficientemente polimórficos como para ter interese forense ou aplicacións xenealóxicas (Butler 2008).

A comisión de ADN da ISFG tratou en profundidade os problemas de nomenclatura dos STR-Y e a definición dos *loci* e dos alelos destes marcadores (Gusmão et al. 2006). Isto permitiu diminuír, en gran medida, os problemas que existían á hora de compartir datos entre laboratorios, xa que ata o momento se nomeaban de maneiras diferentes. Actualmente, a nomenclatura recomendada pola ISFG, é a máis utilizada. Ademais, por outra parte, a estandarización da nomenclatura facilitou a elaboración de bases de datos de haplotipos de STR-Y, imprescindibles para as aplicacións forenses de ditos polimorfismos.

Na actualidade, unha das bases de datos dos STR de cromosoma Y máis utilizada é a chamada YHRD (*Y STR haplotype reference database*, [www.yhrd.org](http://www.yhrd.org)); esta foi substituíndo ás anteriores versións que recollían datos de poboacións europeas, asiáticas e americanas por separado (Willuweit et al. 2007). Xurdiu con dous obxectivos, por unha banda, xerar estimacións fiables das frecuencias de haplotipos dos STR-Y, para poder utilizalas en casos forenses e, por outra, calcular a estratificación poboacional existente entre todas as poboacións mundiais, que se reflicte na distribución das frecuencias haplotípicas dos STR-Y.

A base de datos YHRD orixinalmente incluía o chamado “haplotipo mínimo europeo” (Kayser et al. 1997; Pascali et al. 1999), un conxunto de marcadores STR-Y aceptados pola comunidade científica forense extremadamente informativos. Conforman este “haplotipo mínimo” os seguintes *loci*: DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393 e DYS385ab (figura 8). Posteriormente engadíronse outros *loci* hipervariables de STR-Y (DYS438, DYS439, DYS437, DYS448, DYS456, DYS458, DYS635, YGATAH4, DYS481, DYS533, DYS549, DYS570, DYS576 e DYS643) coa finalidade de aumentar o poder de discriminación obtido só co “haplotipo mínimo” (Ayub et al. 2000; Redd et al. 2002; Mulero et al. 2006; Vermeulen et al. 2009).



**Figura 8:** Loci STR-Y que forman o haplotipo mínimo (imaxe modificada a partir de <http://www.cstl.nist.gov/strbase/ystrpos1.htm>).

Outra das bases de datos empregada actualmente é a *US Consolidated Y-STR Database*, accesible na rede no enderezo: <http://www.usystrdatabase.org/>.

Coma no caso dos microsatélites de cromosomas autosómicos, o crecente interese da aplicabilidade destes marcadores, fixo que se desenvolveran varios *multiplexes*, entre os que destacan:

- *AmpFISTR® Yfiler® PCR Amplification Kit (Life Technologies)* permite *coamplificar* 17 STR-Y nunha única reacción de PCR. Este *kit* comercial inclúe os *loci* definidos como “haplotipo mínimo europeo”, os recomendados polo “*Scientific Working Group on DNA Analysis Methods*” (SWGAM), e seis *loci* altamente polimórficos, para incrementar a capacidade de discriminación na análise haplotípica.
- *PowerPlex® Y System Amplification (Promega)* *coamplifica* 12 *loci* STR-Y nunha única reacción, e tamén inclúe o “haplotipo mínimo europeo” e os recomendados polo SWGAM.
- *PowerPlex® Y23 System (Promega)* é o *multiplex* máis novidoso e contén 23 *loci* STR-Y, en concreto os 12 do *PowerPlex® Y System*, cinco *loci* do *AmpFISTR® Yfiler®* e seis novos *loci*.

### 1.5.1.1 Taxa de mutación dos STR-Y

Na análise estatística utilizada na comparación de perfís xenéticos, é importante ter en conta os valores das taxas de mutación para os diferentes *loci* e mesmo para cada alelo de cada *loci* microsatélite do cromosoma Y. Isto sería necesario no caso no que, por exemplo, o haplotipo do presunto pai difira do haplotipo do fillo por un único STR, só sería posible calcular o índice de paternidade coñecendo a taxa de mutación do *locus* en cuestión. Existe, por tanto, unha taxa de mutación por microsatélite e por xeración, pero hai diferenzas alelo específicas (Jobling et al. 2003). Por outra parte, unha boa estimación da taxa de mutación dos STR-Y é esencial para realizar unha datación precisa das liñaxes de cromosoma Y definidas por marcadores SNP, entre outras aplicacións evolutivas importantes.

Ao longo destes últimos anos, numerosos grupos de investigación levaron a cabo estudos sobre as taxas de mutación dos STR-Y e fóronse reunindo na YHRD (Willuweit et al. 2007). Nela podemos comprobar como a taxa de mutación destes *loci* depende da estrutura do *locus* e da lonxitude do alelo (Brinkmann et al. 1998b; Di Rienzo et al. 1998).

A estimación da taxa de mutación media dos *loci* de STR-Y varía segundo o estudo. As máis aceptadas, por obterse en estudos que tiñan en conta un elevado número de transferencias alélicas (meioses), son de  $1,998 \times 10^{-3}$  (95% IC:  $1,501 \times 10^{-3}$ - $2,606 \times 10^{-3}$ ) mutacións/ xeración/ *locus* (Gusmao et al. 2005), de  $2,1 \times 10^{-3}$  mut./xen./*locus* (95% IC:  $1,7 \times 10^{-3}$ - $2,5 \times 10^{-3}$ ) (Ge et al. 2009), de  $2,8 \times 10^{-3}$  (95% IC:  $2,2 \times 10^{-3}$ - $3,5 \times 10^{-3}$ ) (Goedbloed et al. 2009), de  $2,8 \times 10^{-3}$  mut./xen./*locus* (95% IC:  $1,72 \times 10^{-3}$ - $4,27 \times 10^{-3}$ ) (Kayser et al. 2000) e de  $3,35 \times 10^{-3}$  (95% IC:  $1,79 \times 10^{-3}$ - $6,38 \times 10^{-3}$ ) mut./xen./*locus* (Ballantyne et al. 2010), entre moitas outras. Ao ser da mesma orde que as taxas de mutación obtidas para os microsatélites de cromosomas autosómicos, considérase que o mecanismo mutacional xeral dos STR é independente da recombinación (Sajantila et al. 1999; Kayser et al. 2000).

Segundo os datos recollidos na base de datos YHRD, na actualidade (maio de 2014) o *locus* DYS438 presenta a frecuencia de mutación específica máis baixa, cun valor de  $3,748 \times 10^{-4}$  (95% IC:  $1,021 \times 10^{-4}$  -  $9,593 \times 10^{-4}$ ). Outros *loci* como o DYS533 presentan un valor de 0 (95% IC: 0 -  $8,219 \times 10^{-3}$ ), probablemente debido ao reducido número de meioses estudadas ata o momento.

Alguns estudos parecen demostrar que existe unha tendencia a unha maior mutabilidade cando os alelos dos STR-Y son maiores (Gusmao et al. 2005), pero para determinar dunha forma máis precisa esta relación, sería necesario un tamaño de mostraxe maior do que se utilizou ata o momento, debido a que hai outros estudos que observan o contrario (Järve et al. 2009). Ademais, os resultados obtidos mostran que, xeralmente, a maioría de mutacións son dun só paso mutacional (Kayser et al. 2000; Decker et al. 2008; Sánchez-Diz et al. 2008; Goedbloed et al. 2009; Ballantyne et al. 2010), feito que apoia o modelo de mecanismo mutacional máis aceptado para os STR, o modelo de mutación “paso a paso”, en inglés coñecido como “*stepwise*”. Por outra banda, na maioría de estudos obsérvase que o número de ganancias de unidades de repetición supera significativamente ao número de perdas (Kayser et al. 2000; Dupuy et al. 2004; Budowle et al. 2005; Gusmao et al. 2005; Decker et al. 2008). Outros traballos, pola contra, observan un maior número de perdas de unidades de repetición (Hohoff et al. 2007; Ge et al. 2009), e outros ven unha relación da tendencia á perda ou á ganancia de unidades de repetición co tamaño do alelo do que se trate (Dupuy et al. 2004). Isto demostra, unha vez máis, que o estudo da dinámica mutacional dos microsátélites aínda non concluíu e son necesarias aproximacións máis exhaustivas.

Mención especial merecen as duplicacións e incluso triplicacións que nos encontramos nos STR-Y e que, en moitas ocasións, dificultan a interpretación do xenotipado dos mesmos, sobre todo nos casos nos que se parte dunha mestura de mostras. A detección de máis dun alelo a nivel dun *locus* débese á ocorrencia dunha duplicación e unha posterior mutación que altera o número de repeticións nun dos dous *loci*. As mutacións máis frecuentes son as dun paso único de mutación, polo tanto, espérase unha diferenza dunha soa repetición nos alelos duplicados (Butler et al. 2005). Esta característica, unida á observación de que a altura dos picos que corresponden a cada alelo dunha duplicación dun *locus* é similar, podería permitir diferenciar entre unha mostra formada por varios individuos masculinos e unha mostra dun individuo con varias duplicacións (Butler et al. 2005). Entre os *loci* STR-Y máis utilizados nos que se detectaron duplicacións atópanse os *loci* DYS19, DYS389I, DYS389II, DYS437 e DYS439, entre outros. Documentáronse tamén, casos de eventos independentes de duplicación nun mesmo *locus* (Capelli et al. 2007) e casos de triplicacións de *loci* (Santos 1996; Schoske et al. 2004).

### 1.5.2 Marcadores SNP-Y

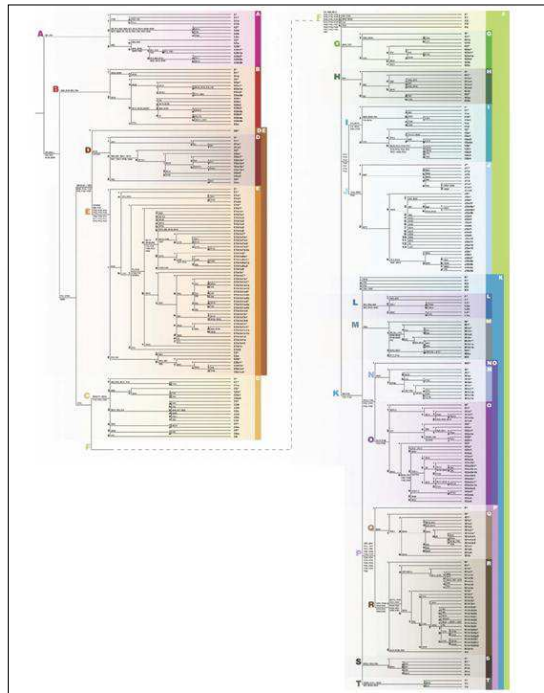
As características xerais dos SNP presentes no cromosoma Y (SNP-Y) son as mesmas que as do resto dos SNP, exceptuando as que confire a ausencia de recombinación na maior parte deste cromosoma sexual.

Os SNP-Y combínanse en haplogrupos, é dicir, haplotipos definidos por marcadores binarios, considerados máis estables que os haplotipos definidos por microsátélites, a pesar de seren tamén menos informativos. No caso do cromosoma Y, esta combinación normalmente pasa intacta de xeración en xeración, permitindo conservar así un rexistro da súa historia, xa que no cromosoma Y, só a mutación xeraría diversidade haplotípica (Jobling et al. 2003). En moitas ocasións atopamos na bibliografía existente unha confusión entre os termos “haplotipo” e “haplogrupo”, porén, de maneira xeral considérase que os cromosomas Y definidos polo estado dos seus marcadores binarios denomínanse “haplogrupos”, os identificados por microsátélites denomínanse “haplotipos”, e os cromosomas Y definidos pola combinación de ambos, marcadores SNP e STR, denomínanse “liñaxes” (de Knijff 2000).

Os factores que inflúen en gran medida na diversidade dos haplogrupos son, principalmente, o tamaño efectivo da poboación e a mobilidade da mesma. Deste xeito, se as poboacións son grandes e están asentadas nun lugar determinado, os cambios na frecuencia dos haplogrupos debidos á deriva serán lentos e o fluxo xénico local provoca a formación de patróns clinais (Jobling et al. 2003). Existen multitude de ocasións na historia das poboacións humanas nas que se produciron estas circunstancias, dando lugar así á heteroxeneidade de patróns clinais de haplogrupos de cromosoma Y que atopamos na actualidade. No apartado 4.2.3 da presente memoria de tese expónse un exemplo do estudo do patrón clinal do haplogrupo máis frecuente do continente europeo, o chamado R1b1b2 ou R-M269.

En referencia ás taxas de mutación dos SNP de cromosoma Y, hai que mencionar que son tan baixas que poden ser considerados eventos únicos na evolución humana. Estudos recentes encontran unha taxa de mutación para os SNP-Y da mesma orde que a dos SNP autosómicos, en concreto, de  $3 \times 10^{-8}$  mutacións/ nucleótido/xeración (95% IC:  $8,9 \times 10^{-9}$ – $7 \times 10^{-8}$ ) (Xue et al. 2009). Así e todo, como se indicou no apartado 1.3.3.2 da presente memoria de tese, varias mutacións recorrentes xa foron documentadas. Estas son mutacións que xeran independentemente un estado derivado

que xa fora observado previamente na poboación (Jobling 2004). A estrutura filoxenética estable e a haploidía do cromosoma Y permiten identificar sen ambigüidades as mutacións recorrentes (Jobling et al. 2003). De feito, na árbore filoxenética deseñada por Karafet e colaboradores no ano 2008 (figura 9) hai aproximadamente 600 polimorfismos binarios descritos, dos que oito son mutacións recorrentes (Karafet et al. 2008). É o caso, por exemplo, do primeiro SNP de cromosoma Y descuberto, chamado SRY10831/SRY1532, que mutou dúas veces ao longo da historia da nosa especie (Hammer et al. 1998).



**Figura 9:** Árbore filoxenética do cromosoma Y. Imaxe tomada de (Karafet et al. 2008).

Nos últimos anos asistimos a unha explosión de novos polimorfismos descritos na rexión MSY do cromosoma Y humano. Deste xeito, ata 1997 só se coñecían 11 polimorfismos binarios nesta rexión, que eran xenotipados por técnicas baseadas na PCR. En 1997 describíronse 19 polimorfismos binarios máis, grazas á incorporación dunha nova técnica moi eficiente na detección de mutacións, a tecnoloxía DHPLC, cromatografía líquida desnaturizante de alta resolución (Underhill et al. 1997). Esta técnica permitiría que no ano 2001 xa se tiveran descrito máis de 200 polimorfismos

binarios (Hammer et al. 2001; Underhill et al. 2001), no 2008 máis de 600 (Karafet et al. 2008) e, actualmente, debido aos recentes avances nas tecnoloxías de secuenciación de ADN, xa se describiron varios miles (van Oven et al. 2014).

Dispoñemos de varias bases de datos que tratan de reunir os SNP-Y descritos ata o momento, entre as que destaca a base de datos da YHRD (*Y-STR haplotype reference database*; <http://www.yhrd.org/Research/YSNPs>) e a de *PhyloTree-Y* (<http://www.phylotree.org/Y>).

Os polimorfismos binarios da rexión MSY permiten inferir a evolución humana, a afinidade entre poboacións e a súa historia demográfica (Underhill et al. 2000). Nos últimos anos deseñáronse numerosos *multiplexes* de SNP-Y, con fins forenses e poboacionais (Brion et al. 2004; Vallone et al. 2004; Brion 2005; Brion et al. 2005b; Sanchez et al. 2005; Onofri et al. 2006; He et al. 2010; van Oven et al. 2012), como os que permiten unha predición da orixe xeográfica das liñaxes masculinas (Brion et al. 2005a), ou aqueles utilizados na busca de relacións filoxenéticas entre poboacións (Geppert et al. 2010), entre outros. As aplicacións específicas destes polimorfismos serán expostas en maior profundidade no apartado 1.6.2 desta tese.

### **1.5.2.1 A filoxenia do cromosoma Y**

As características diferenciadoras do cromosoma Y como son a súa herdanza paterna, a ausencia de recombinación na maior parte do cromosoma e a conservación de mutacións de xeración en xeración, fan deste cromosoma unha útil ferramenta filoxenética. Estas peculiaridades permiten a reconstrución de liñaxes masculinas nas poboacións actuais, mediante o estudo dos polimorfismos deste cromosoma (Jobling et al. 2003). A ausencia de recombinación en concreto, implica que os haplogrupos monofiléticos poidan ser relacionados cunha única filoxenia utilizando o principio de máxima parsimonia. Este principio define a mellor árbore filoxenética como aquela que, para explicar os datos existentes, require o menor número de cambios evolutivos posibles.

A filoxeografía asume que hai unha correspondencia entre a distribución de haplotipos e haplogrupos e os movementos poboacionais acontecidos no pasado. Demostrouse, deste modo, que o estudo da distribución da diversidade do cromosoma Y permite describir os movementos humanos prehistóricos máis importantes e trazar a

composición das poboacións humanas modernas (Underhill et al. 2001). A distribución moderna dunha liñaxe pode proporcionar información sobre a distribución pasada da mesma, debido a que se asume que os movementos xeográficos das poboacións humanas foron limitados (Jobling 2004).

Nos estudos xenéticos de cromosoma Y hai que ser cautos á hora de analizar os datos e non obviar que se trata dun *locus* específico masculino, polo que non se deben facer interpretacións simplistas, como pode ser igualar unha liñaxe cunha poboación. Outra das consideracións a ter en conta ao estudar a distribución xeográfica das liñaxes de cromosoma Y é o feito de que proporcionan información sobre eventos antigos, sempre e cando se poidan excluír aquelas distribucións que, se sabe, foron consecuencia de eventos recentes. Por ilustrar cun exemplo coñecido, á luz dos datos de cromosoma Y obtidos na illa de Rapa (Polinesia), poderíase interpretar que a colonización de dita illa foi amerindia, cando por evidencias históricas se sabe que ese compoñente amerindio atopado procede dun fenómeno de tráfico de escravos nativo americanos do século XIX, sendo a colonización da illa orixinariamente asiática (Hurles et al. 2003).

Moitas destas cuestións acláranse ao comparar os resultados obtidos cos estudos baseados no ADN mitocondrial. Cando se contrastan ambos tipos de estudos podemos coñecer o fluxo xénico específico de cada sexo, como é o caso da expansión dos europeos en América e Oceanía nos últimos 500 anos. Nesta expansión, por exemplo, observouse que en América do Sur ou Polinesia hai un gran compoñente europeo nos cromosomas Y, con todo, mantéñense as liñaxes indíxenas de ADNmt, o cal reflicte unha política sexual na actividade colonial (Jobling et al. 2003). Existen, pois, diferentes forzas demográficas moduladas polo sexo, isto fai que, ás veces, os relatos ofrecidos polos dous sistemas xenéticos uniparentalmente herdados, sexan diferentes, de aí a importancia de comparalos (Underhill et al. 2001).

Paralelamente, na interpretación dos datos filoxenéticos ten un papel esencial o denominado tempo de coalescencia ou **TMRC**A (do inglés, *Time to the Most Recent Common Ancestor*), é dicir, o tempo ata o antepasado común máis recente. O TMRC A dun *locus* é proporcional ao tamaño efectivo da poboación, supoñendo a neutralidade de dito *locus* e outros supostos comunmente aceptados polos xenetistas, como son un tamaño de poboación constante e un apareamento aleatorio. Unha maneira apropiada de estimar o tempo de coalescencia que existe nun conxunto próximo de cromosomas Y, tendo en conta ditas asuncións e estimacións precisas de taxas de mutación, consiste en

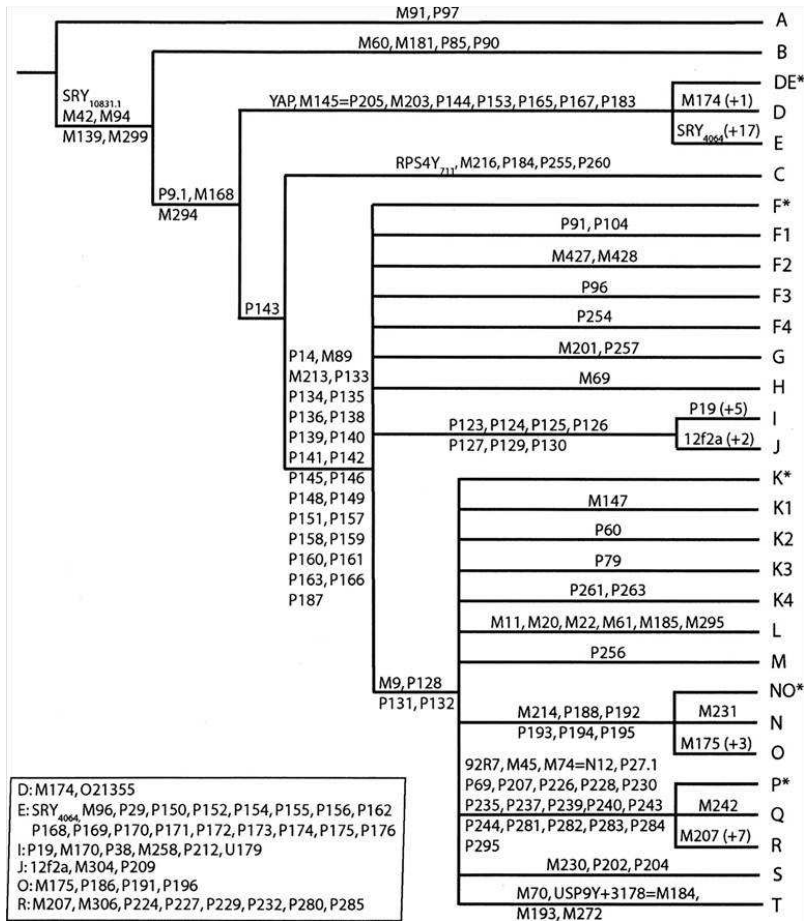
analizar a variación que existe nos microsátélites de ditos cromosomas (Jobling et al. 2003). Utilizando estas ferramentas, realizáronse numerosas estimacións para coñecer a antigüidade de cada haplogrupo, e poder reconstruír así a historia evolutiva das liñaxes do cromosoma Y. Un haplogrupo será máis antigo cando maior diversidade haplotípica acumulase co paso do tempo, é dicir, canto maior sexa a variabilidade dos seus microsátélites.

Así e todo, existe moita controversia ao respecto, sobre todo referida ás asuncións que hai que ter en conta para estimar as idades de cada haplogrupo. Unha delas é a referida ao tempo interxeracional, onde adoita producirse unha subestimación no caso dos varóns. Outro dos factores que complican o cálculo do TMRCA consiste no feito de que as taxas de mutación coñecidas dos STR non son precisas, e varían entre cada *locus* e incluso entre os distintos alelos dun mesmo *locus*, como se mostrou no apartado 1.5.1.1 desta tese. Ademais, habería tamén que ter en conta que, realmente, a variación non se acumula a unha taxa constante.

No ano 2002, o chamado “*Y Chromosome Consortium*” (YCC), grupo implicado no estudo cooperativo sobre a variación xenética do cromosoma Y humano, publicou a primeira filoxenia consenso deste cromosoma, describindo 153 haplogrupos (Y-Chromosome-Consortium 2002). Nesta publicación propúñase tamén un novo sistema de nomenclatura para designar aos haplogrupos do cromosoma Y que, ata agora, é o máis utilizado. Trátase dun sistema de nomenclatura alfanumérico alternante que, ademais, é xerárquico, propiedade que facilita enormemente a incorporación á árbore filoxenética dos novos SNP-Y que se van descubriendo. Deste xeito, é un sistema o suficientemente flexible como para permitir os cambios inevitables que aparecen nas liñaxes da MSY, como consecuencia da aparición de novas mutacións. Con anterioridade a esta publicación, existían sete sistemas diferentes de nomenclatura en uso, o que facía case imposible a comparación de resultados entre as distintas publicacións ao respecto. Por este motivo intentábase tamén establecer unha equivalencia entre as nomenclaturas anteriores ((Jobling et al. 2000) e (Kalaydjieva et al. 2001), (Underhill et al. 2000), (Hammer et al. 2001), (Karafet et al. 2001), (Semino et al. 2000), (Su et al. 1999) e (Capelli et al. 2001)) e a proposta nesta publicación (Y-Chromosome-Consortium 2002).

Para designar aos haplogrupos de cromosoma Y hoxe en día, ademais do sistema de nomenclatura xerárquico proposto polo YCC, está vixente outra nomenclatura que

consiste en nomear ao haplogrupo utilizando o nome do paragrupo ao que pertence seguido da mutación que define esa liñaxe. Considérase paragrupo a unha liñaxe que non está definida pola presenza do estado derivado dun marcador. Pódese definir tamén paragrupo como un conxunto de haplotipos que conteñen algúns dos descendentes, pero non todos, dunha liñaxe ancestral (Jobling et al. 2003). Para diferenciarlos visualmente dos haplogrupos engádesse un asterisco (\*) á letra que representa un paragrupo.

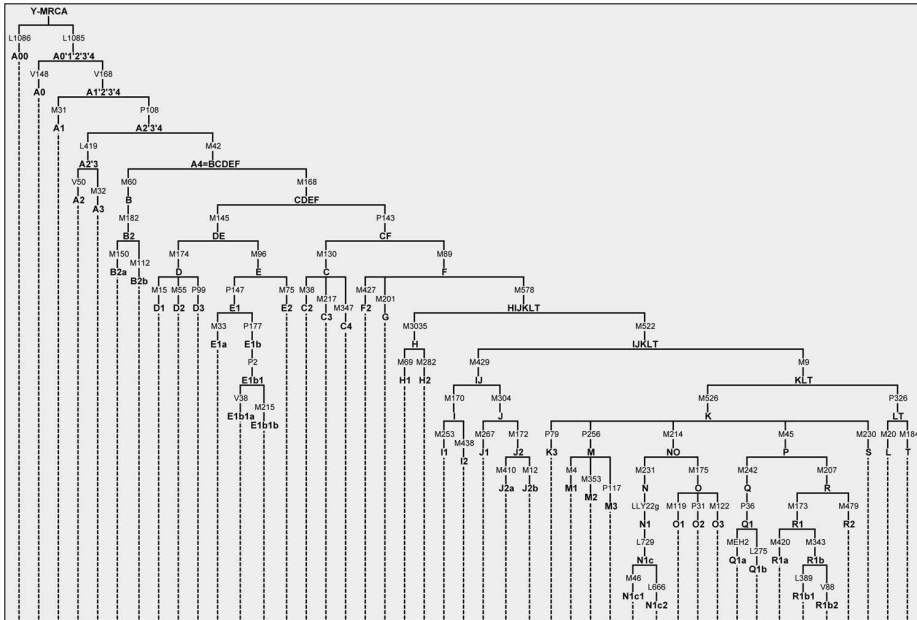


**Figura 10:** Esquema da árbore de parsimonia do cromosoma Y. Nas ramas indícanse as mutacións que definen cada clado. Imaxe obtida de Karafet et al. 2008.

A filoxenia descrita por Karafet e colaboradores no ano 2008, inclúe, aproximadamente, 600 polimorfismos binarios e 311 haplogrupos (Karafet et al. 2008). Un dos cambios máis substanciais que introduciu este estudo, en comparación coa

filoxenia anterior (Y-Chromosome-Consortium 2002), foi a aparición dun novo polimorfismo binario, o chamado P143, que unía os haplogrupos C e FT anteriormente descritos, formando un “*supercluster*” que contiña as liñaxes que non se encontraban na África subsahariana. Tamén se uniron os clados I e J, e N e O, ao aparecer novas mutacións que os relacionaban. Por outra banda, dúas liñaxes do haplogrupo K consideráronse dous haplogrupos diferentes e máis recentes, denominados S e T. Pódese apreciar unha visión xeral de dita filoxenia na figura 10.

Coa recente chegada das tecnoloxías de secuenciación de nova xeración, produciuse unha explosión na cantidade de SNP descubertos no cromosoma Y, polo que a resolución da árbore filoxenética é cada vez maior, e parece necesaria unha nova actualización desta árbore proximamente. Existen varios recursos na rede que xa amosan unha visión da enorme topoloxía da árbore filoxenética completa, como o que mantén a *International Society of Genetic Genealogy* (ISOGG; <http://www.isogg.org/tree>) ou Thomas Krahn (<http://ytree.ftdna.com>). Porén, a elevada cantidade de información incluída nesta filoxenia completa, con xa miles de SNP-Y identificados, complica a súa comprensión e a escollo dos marcadores axeitados para o xenotipado con finalidades evolutivas, demográficas, antropolóxicas, xenealóxicas, médicas e forenses. Por iso, recentemente apareceu unha nova proposta de filoxenia minimizada (figura 11), baseada nunha árbore de referencia concisa, moi estable e que unicamente inclúe os SNP-Y máis informativos, é dicir, só aqueles que ocupan posicións filoxenéticas clave e definen liñaxes con frecuencias significativas nun nivel de poboación rexional, pretendendo acadar así unha capacidade de discriminación global óptima. A versión actualizada desta árbore está dispoñible na rede no enderezo seguinte: <http://www.phylotree.org/Y>, e non pretende substituír á filoxenia completa do cromosoma Y, moi importante para certas aplicacións, se non proporcionar unha árbore de referencia práctica e sinxela de empregar para os xenetistas humanos (van Oven et al. 2014).



**Figura 11:** Esqueleto da filoxenia do cromosoma Y humano coa nomenclatura do haplogrupo basal (en negra) e a definición dos marcadores SNP-Y indicados nas ramas, segundo van Oven et al. 2014.

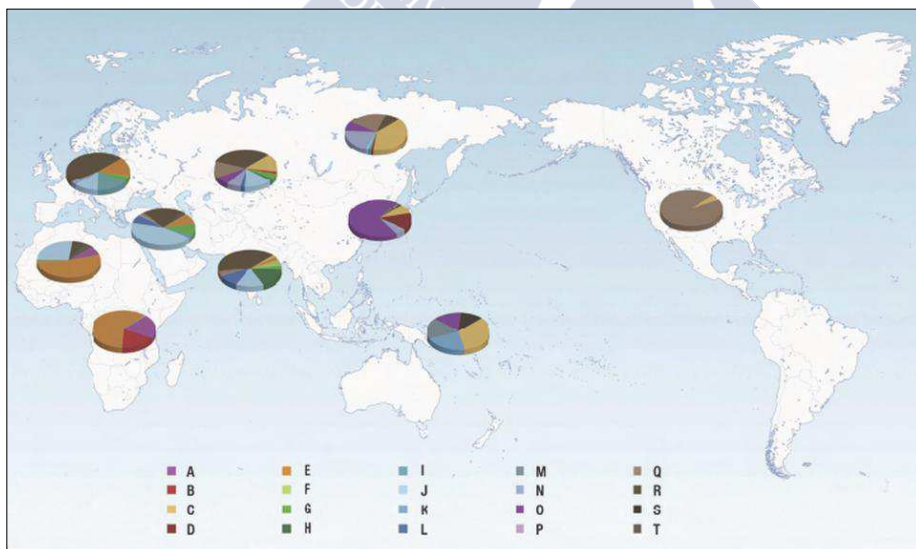
Dende a publicación da filoxenia de Karafet e colaboradores foron aparecendo varias actualizacións topolóxicas que tiveron importantes consecuencias na nomenclatura dos clados basais de haplogrupos, tal e como recolle a filoxenia recentemente publicada por van Oven e colaboradores (figura 11). Deste xeito, os subclados A coñecidos como A1a, A1b e A3, deixan de considerarse un clado monofilético para seren catalogados como unha estrutura parafilética. Ademais, o haplogrupo A1b considérase o A0, o A1a o A1, e o A4 sería o equivalente ao BCDEF. Estes autores, ademais, son partidarios de empregar unha nomenclatura máis intuitiva, onde se nomeen os nomes de todos os clados que compoñen unha determinada rama da filoxenia, como poden ser CDEF e BCDEF en lugar de CF ou BT, como se adoita a nomear noutras publicacións. Segundo esta proposta, nomear un rango de clados só coas letras do primeiro e do último podería levar a confusión xa que, nalgunha ocasión, ao nomear dúas letras non se está a referir a un rango, se non á raíz que une dous clados determinados.

Outro dos cambios topolóxicos informados nesta filoxenia é a bifurcación máis profunda entre os haplogrupos M, NO, P (incluíndo Q e R) e S nun subclado, definido polo marcador M526, e os haplogrupos L e T noutro subclado diferente, definido polo marcador P326. Deste xeito, os haplogrupos L e T non están incluídos na rama do haplogrupo K-M526.

Por último, outro dos cambios topolóxicos desta filoxenia implicou nomear ao haplogrupo H-M69 como H1 e ao F3-M282 como H2, tendo ambos unha raíz común definida polo marcador M3035.

### Breve descrición da distribución filoxeográfica das liñaxes do cromosoma Y

Segundo a filoxenia descrita por Karafet e colaboradores (Karafet et al. 2008) a árbore filoxenética proposta para o cromosoma Y (figura 9) está formada por 20 clados principais, para os que describen a súa distribución xeográfica a nivel mundial, e para 11 deles, ademais, estiman a súa antigüidade, permitindo inferir así cales foron os grandes movementos poboacionais da nosa especie. De maneira esquematizada, pódese observar a distribución actual destes clados principais na figura 12 e dunha maneira máis descriptiva na táboa 1.



**Figura 12:** Frecuencias dos 20 haplogrupos maioritarios actualmente aceptados da rexión MSY, para 10 rexións xeográficas diferentes. Imaxe tomada de (Karafet et al. 2008).

A pesar da incerteza sobre o TMRCA do cromosoma Y, como xa foi mencionado, non hai vacilación á hora de afirmar que non se encontrou, en ningunha parte do mundo, ningunha liñaxe de cromosoma Y anterior a 200.000 anos, e que a filoxenia deste cromosoma arranca en África (Jobling et al. 2003).

A hipótese da orixe africana recente, coñecida como “*Out of Africa*” ou hipótese da “Eva mitocondrial”, foi postulada a partir dun estudo realizado utilizando o ADN mitocondrial de 147 individuos con cinco orixes xeográficas diferentes. Dita hipótese propón que todos os tipos de ADNmt actuais remóntanse a un único antepasado que viviu en África hai aproximadamente 200.000 anos, postulando así que África é a fonte de todo o acervo xenético mitocondrial humano actual (Cann et al. 1987). Esta hipótese está baseada no principio de coalescencia, segundo o cal, asumindo que houbo unha orixe única para todos os organismos vivos, dedúcese que toda variación en calquera segmento do ADN nas xeracións presentes debe proceder, en última instancia, dun único devanceiro que existiu en xeracións previas. Esta é a hipótese maioritariamente aceptada na actualidade, en contraposición á hipótese multirrexional (Templeton 1997).

Considérase que tiveron lugar dúas saídas masivas do *Homo sapiens* a partir do continente africano e por dúas rutas diferentes: unha primeira hai uns 50.000 anos, polo sur, seguindo probablemente unha ruta costeira, que poboaría o sueste de Asia e Australia, e unha posterior, polo norte, que se dirixiría cara o noroeste de Asia e Europa. Isto podería explicar, de forma xeral, o feito de que os haplogrupos que se encontran no sueste asiático e en Australia sexan tan diferentes aos que aparecen no resto de Asia e Europa (Jobling et al. 2003). O último continente colonizado sería o de América, e a hipótese maioritariamente aceptada considera que esta colonización se produciu desde o noreste de Asia hai uns 14.500 anos, a pesar de que existe aínda certo debate sobre o número de colonizacións que tiveron lugar.

O modo de vida baseado na agricultura unido ao clima temperán e estable que acompañou ao Holoceno, permitiu varias expansións demográficas desde hai aproximadamente uns 10.000 anos. Ao aumentar a poboación mundial humana, os patróns de diversidade que se estableceran durante o período Paleolítico, en gran parte por migración e deriva, quedaron como suspendidos e os cambios a gran escala foron menos frecuentes. Porén, eventos recentes poderían ocasionalmente ter unha influencia significativa na diversidade do cromosoma Y (Jobling et al. 2003).

Na seguinte táboa amósase a distribución global dos clados principais da filoxenia do cromosoma Y.

**Táboa 1:** Exemplo de marcadores bialélicos do cromosoma Y que determinan algúns dos clados principais da árbore filoxenética, así como a súa respectiva distribución poboacional, indicando as rexións e/ou as poboacións onde a súa frecuencia é máis elevada (Adaptada de (Gusmão 2010)).

HG	SNP-Y	Mutación	Distribución poboacional
A	M91	9T→8T	África (poboacións khoisánidas, Etiopía e Sudán)
B	M60	Ins 1pb	África subsahariana (elevada frecuencia en pigmeos, África central)
C	M216	C→T	Asia, Australia, Oceanía e poboacións <i>Na-Dené</i> do norte de Norteamérica
D	M174	T→C	Asia Central (Himalaia, Tibet) e Xapón
E	SR <sub>Y</sub> <sub>4064</sub>	G→A	África (elevadas frecuencias en todo o continente), Oriente Medio e Mediterráneo
G	M201	G→T	Oriente Medio, Mediterráneo e Cáucaso
H	M69	T→C	Subcontinente hindú e poboacións roma (xitanas) europeas
I	M170	A→C	Europa (frecuencias elevadas no norte de Europa, Escandinavia)
J	M304	A→C	Oriente Medio, Norte de África, Europa, Asia Central, Paquistán e India
L	M22	A→G	Oriente Medio, Norte de África, Europa, Asia Central e Mediterráneo
M	P256	G→A	Oceanía e Indonesia Oriental (Papúa Nova Guinea e Melanesia)
N	M231	G→A	Norte de Eurasia
O	M175	Del 5pb	Asia Oriental, Asia Central e Oceanía
Q	M242	C→T	Norte da Eurasia (Siberia), con elevada frecuencia en poboacións nativas de América
R	M207	A→G	Europa (frecuencias elevadas no oeste europeo), Asia central e occidental
S	P202	T→A	Oceanía e Indonesia
T	M70	A→C	Oriente Medio, África, Europa, Asia Central e Mediterráneo

Os estudos de filoxeografía e filoxenia permitiron reconstruír a historia das poboacións humanas actuais, existindo, incluso, grandes consistencias entre as filoxenias do ADNmt e a do cromosoma Y, ademais da raíz africana de ambas. Entre estas coincidencias destaca o feito de que só un pequeno subconxunto das dúas árbores se dispersou fora de África. Deste xeito, existen tres haplogrupos fundadores en ambas árbores que compoñen toda a variación xenética non africana, e constitúen só unha

pequena parte da diversidade xenética africana. En concreto, son os haplogrupos M, N e R na filoxenia do ADNmt e os haplogrupos C, D e F na filoxenia do cromosoma Y (Underhill et al. 2007).

## 1.6 Aplicacións do estudo da variabilidade xenética

### 1.6.1 Marcadores STR

#### 1.6.1.1 Aplicacións de interese forense

Non cabe dúbida de que unha das aplicacións do estudo dos polimorfismos xenéticos humanos que acadou maior relevancia nos nosos días, foi a do uso dos STR como ferramenta de identificación de persoas. Deste modo, os STR autosómicos con alto grado de polimorfismo utilízanse de forma rutineira na obtención dun perfil xenético que é específico de cada individuo, excepto no caso de xemelgos univitelinos. Os microsátélites son, por tanto, de gran utilidade na investigación criminalística, na identificación de persoas e de restos humanos, e tamén na investigación biolóxica de paternidade. En cada un destes casos preténdese obter o perfil de interese para o seu posterior cotexo con outros perfís xenéticos implicados. Así, por exemplo, pódese obter o perfil xenético procedente dos vestixios biolóxicos atopados na escena dun crime, ou ben dos individuos dos que se pretende establecer unha relación de parentesco, e mesmo de restos cadavéricos. Estes perfís cotexaríanse con outros que fosen de interese en cada caso, como poderían ser os perfís dos sospeitosos de dito crime, os dos familiares dunha persoa desaparecida, ou os perfís dunha base de datos xenética.

Por conseguinte, nunha pericia xenético biolóxica é necesario diferenciar dous tipos de mostras, as mostras de **referencia** (“*indubitadas*”), de procedencia coñecida, e as **evidencias** (“*dubitadas*”), de procedencia descoñecida. Cando se comparan os perfís xenéticos de ambos tipos de mostras e non coinciden, dise que son perfís incompatibles, e non é necesario realizar ningunha valoración estatística, se non que se informa da exclusión con seguridade. Pola contra, cando o perfil xenético evidenciado en ambos tipos de mostras é **coincidente** (en estudos de criminalística) ou **compatible** (en casos de paternidade), é necesario valorar estatisticamente se a coincidencia é debida ao azar ou a que realmente a evidencia biolóxica pertence ao individuo de referencia (Prieto Solla 2004). Esta valoración estatística realízase en termos de probabilidade.

No referente aos microsátélites presentes no cromosoma Y hai que sinalar que permiten diferenciar o compoñente masculino nunha mestura formada polos restos biolóxicos de individuos de distinto sexo. Esta propiedade é de gran utilidade en casos de agresións sexuais, nos que normalmente existe unha mestura formada por un compoñente maioritario feminino que enmascara ao compoñente minoritario masculino, de forma que, se non se contase cos STR do cromosoma Y, sería moi difícil chegar a conseguir o perfil xenético masculino, sobre todo nos casos nos que non se logra realizar unha extracción diferencial. Son especialmente útiles en casos problemáticos de agresións sexuais, como son aqueles nos que existen varios compoñentes masculinos na mestura. Ademais, o perfil xenético obtido cos microsátélites do cromosoma Y tamén é de gran interese en probas de paternidade nas que, por exemplo, non contamos co perfil do suposto pai biolóxico, podéndose utilizar os perfís dos familiares da liñaxe paterna como mostras de referencia, ou ben cando non se dispón do perfil materno e o fillo é varón.

A haploidía e a herdanza patrilineal complican a interpretación das coincidencias dos haplotipos de STR-Y, porque os familiares emparentados por vía paterna comparten durante varias xeracións un perfil idéntico dos STR-Y (Roewer 2009), é dicir, un haplotipo de STR-Y é unha liñaxe non recombinante que pode ser compartido por moitos individuos (Gusmão et al. 2006). É por este motivo polo que, polo momento, utilizando os microsátélites do cromosoma Y só se pode excluír a un individuo, non se pode chegar a identificar, como ocorre cos STR autosómicos. Nun futuro, a expansión da secuenciación completa do cromosoma Y podería permitir chegar a discriminar entre individuos da mesma liña paterna.

### **1.6.1.2 Aplicacións de interese poboacional**

Nos últimos anos, o estudo da variación existente nos microsátélites do cromosoma Y representou un achegamento apropiado para estimar a antigüidade do antepasado común máis recente (TMRCA) e poder datar así liñaxes de cromosoma Y (Jobling et al. 2003), como xa se expuxo no apartado 1.5.2.1 da presente tese.

Estimouse que o 70% das sociedades modernas actuais son patrilocais (Burton 1996), isto implica que os familiares emparentados por vía paterna tenden a vivir no mesmo territorio cultural e xeográfico que os seus antepasados, por iso, outra das

aplicacións da análise do cromosoma Y é a de realizar inferencias sobre a poboación orixinaria dun perfil de ADN dado (Roewer 2009).

Ao longo da historia pódese considerar que houbo dous tipos de movementos poboacionais: os de longa distancia, nos que maioritariamente estaban implicados homes, e os de curta distancia, que consistiron, principalmente, en migracións locais femininas, debido á elevada porcentaxe de sociedades patrilocais. No caso do cromosoma Y, ao existir un maior grado de diferenciación xenética entre poboacións que no ADN mitocondrial, compróbase que foron os movementos locais os que tiveron un maior efecto nos patróns xenéticos (Jobling 2004).

Desta forma, nalgunhas poboacións o estudo dos haplotipos de cromosoma Y unido a estudos de apelidos de herdanza patrilineal proporcionaron moita información sobre xenealoxías, indicando, por exemplo, os niveis de *coascendencia* ou os efectos da deriva xenética nunha poboación determinada. Algúns autores defenden incluso que a combinación da xenética molecular e a análise dos apelidos podería esclarecer a estrutura poboacional e a historia, sendo unha área dun crecente interese público (King et al. 2009).

Por todo iso, os microsátélites do cromosoma Y son considerados como marcadores xenéticos adecuados para estudos de estrutura poboacional local e estudos de historia demográfica recente (Roewer et al. 2005).

## **1.6.2 Marcadores SNP**

### **1.6.2.1 Aplicacións de interese forense**

Existe un crecente interese no xenotipado dos polimorfismos nucleotídicos simples no campo forense, non só pola utilidade dos SNP para definir a orixe dos haplogrupos de cromosoma Y ou ADNmt, ou para analizar a orixe xeográfica das mostras, se non tamén polas aplicacións potenciais dos SNP autosómicos. O interese débese ás potenciais vantaxes que posúen os SNP autosómicos para as probas forenses, vantaxes debidas principalmente á súa baixa taxa de mutación e a que permiten o uso de amplicóns moi pequenos, sendo especialmente útiles na análise de mostras degradadas (Sobrinho et al. 2005).

Atendendo á clasificación dos tipos de SNP de Budowle e colaboradores (Budowle et al. 2008) son, por tanto, marcadores SNP dun elevado interese forense

aqueles que se utilizan nas probas de identidade (de forma xeralmente complementaria ao uso de microsátélites) e os que son informativos de características fenotípicas, é dicir, aqueles que permiten realizar predicións dos trazos físicos dun individuo como son a cor de ollos, de cabelo ou de pel. Como tamén se mencionou no apartado 1.3.3.1 desta tese, os SNP informativos de ascendencia (“AIMs”) e os informativos de liñaxe, teñen aplicacións forenses importantes ademais de poboacionais, xa que, en casuística forense, pode ser de gran importancia inferir a orixe bioxeográfica do antepasado dun individuo ou ben a liñaxe á que pertencen a mostra de referencia e a evidencia, para chegar a resolver un caso (Phillips et al. 2009). No eido da predición bioxeográfica veñen de obterse resultados moi precisos co algoritmo de estrutura poboacional xeográfica (*Geographic Population Structure*, GPS) e o uso dun chip duns 100.000 marcadores informativos de ascendencia, chegándose a predicir o lugar orixinario de individuos de todo o mundo ata unhas centenas de quilómetros (Elhaik et al. 2014).

A incorporación do uso de *multiplexes* ao xenotipado dos SNP, tanto en casuística forense, como é o caso do SNPforID 52-plex, como noutro tipo de campos, supuxo un gran avance, xa que a cantidade de mostra requirida é cada vez menor e a información que se obtén en cada reacción é cuantiosa. En multitude de estudos avaliouuse o uso dos SNP para as análises forenses e chegouse a comprobar que, cunha escolma relativamente pequena duns 50 *loci*, obtéñense resultados equiparables aos obtidos cos *multiplexes* de STR existentes (Gill 2001; Musgrave-Brown et al. 2007). Coa fin de deseñar paneis de SNP universais, os *loci* elixidos deben reunir unha serie de características, entre as que destacan o feito de posuír unhas frecuencias xénicas próximas a 0,5 no maior número de poboacións posible (Pakstis et al. 2007) e a segregación independente dos marcadores elixidos.

No tocante aos SNP presentes no cromosoma Y, ademais de seren informativos da liñaxe, cabe mencionar a súa utilidade na resolución de casos complexos de agresións sexuais, complementando a información que outorgan os STR-Y. Os SNP autosómicos, pola contra, adoitan ter gran importancia en casos complexos de paternidade nos que, por exemplo, os posibles pais teñan unha estreita relación de parentesco. Neste caso, unha análise cun *multiplex* de SNP autosómicos, ademais dos marcadores convencionais, aumentaría o poder de discriminación, incrementando só unha única reacción de amplificación.

### 1.6.2.2 Aplicacións de interese poboacional

O campo onde os SNP tiveron maior utilidade ata o momento, foi o da xenética de poboacións. Por unha parte encóntranse os chamados “AIMs”, marcadores SNP que permiten facer unha inferencia sobre a poboación de orixe máis probable dunha mostra determinada, e, por outra, os coñecidos como marcadores SNP informativos de liñaxe, polimorfismos que se atopan no cromosoma Y e no ADNmt e informan sobre a orixe da liñaxe da mostra, non sobre a orixe da propia mostra.

A achega dos SNP presentes no cromosoma Y no estudo das poboacións humanas ten sido especialmente significativa xa que, como se detallou no apartado 1.5.2 da presente memoria de tese, constituíron unha ferramenta filoxenética esencial. Deste xeito, considérase ao cromosoma Y como o *locus* que proporciona maior información xeográfica de todo o xenoma (Jobling 2004). Así o demostran os innumerables estudos filoxeográficos realizados ata o momento, posto que posibilitaron a progresiva reconstrución da historia evolutiva das liñaxes masculinas humanas, definindo, dunha maneira cada vez máis detallada, os movementos poboacionais que tiveron lugar no pasado, así como as relacións de afinidade entre as distintas poboacións humanas e mesmo as súas respectivas historias demográficas.

Así mesmo, a xenética molecular constituíu unha ferramenta complementaria de multitude de disciplinas, algunhas *a priori* tan afastadas como a lingüística. Neste ámbito, queremos mencionar os estudos que comparan os patróns distribucionais de conxuntos toponímicos antigos (anteriores ao s.VII d.C.) e os patróns distribucionais de haplogrupos de ADNmt e de cromosoma Y, reveladores sobre a indoeuropeidade do noso continente (Villar 2005; Villar 2006; Villar 2012). Segundo estes estudos, determinados haplogrupos de cromosoma Y e de ADNmt xunto cunhas series toponímicas concretas (de etimoloxía indoeuropea) serían introducidos conxuntamente no continente europeo desde Anatolia e Oriente Próximo, durante a difusión démica Neolítica. Os patróns distribucionais doutro conxunto de haplogrupos e coleccións toponímicas tamén indoeuropeas, indicarían unha expansión demográfica desde varios puntos meridionais do continente (península ibérica, itálica e balcánica, principalmente), coincidindo probablemente cunha repoboación do continente posterior ao Dryas recente, durante o Mesolítico. Estes estudos tamén pretenderon dilucidar a orixe de poboacións portadoras de linguas non indoeuropeas como é o éuscaro. Neste termo, ao non encontrarse na península ibérica toponimia de etimoloxía éuscara anterior á época de

Pompeio, trasládase a busca da orixe destas poboacións a Aquitania, de onde parece que proceden (Villar 2005).

Por outra parte, coñecer a variabilidade das poboacións humanas e determinar a existencia de subestruturación poboacional é transcendental para o coñecemento de bases de datos xenéticas empregadas no campo forense, así como en estudos de asociación (como son os estudos caso-control) e, por tanto, de gran relevancia tamén no campo da xenética clínica. En xeral, podemos concluír afirmando que todos os marcadores situados na porción MSY do cromosoma Y posúen aplicacións en moitos campos, incluídos a bioloxía evolutiva, as ciencias forenses, a xenética médica e a xenética poboacional (Karafet et al. 2008).





## **2 XUSTIFICACIÓN E OBXECTIVOS**



A xenética forense está contribuíndo de forma moi extensa a una declaración máis precisa do concepto de identidade persoal, entendida non como un concepto filosófico, se non como un valor biolóxico. O impacto en aplicacións da vida diaria é extraordinario, no ámbito xurídico ofrece a posibilidade de caracterizar as mostras biolóxicas na escena do crime ou, de forma retrospectiva, a revisión dos artigos incautados permite a reconstrución da dinámica delituosa, permitindo por tanto, identificar aos culpables, exonerar os inocentes ou recoñecer ás persoas desaparecidas.

No ámbito civil, a determinación de relacións de parentesco son numerosas, con probas de paternidade e incluso de maternidade. A capacidade para realizar exames a persoas falecidas, mesmo hai moitos anos, ofrece a posibilidade dunha plena aplicación da lei a aqueles herdeiros que tiveron que loitar coa lentitude do noso sistema xudicial. Os acontecementos dos últimos atentados en diversos países, amosaron tamén a posibilidade de aplicar a xenética forense na análise de desastres masivos.

A xenética de poboacións encárgase de dilucidar a orixe da variabilidade xenética, a cal é responsable das diferenzas fenotípicas entre seres humanos e da patoloxía con compoñente xenético. O seu enfoque constitúe unha visión matemática da distribución, así como da dinámica dos alelos e xenotipos dentro das poboacións e entre elas. A variabilidade xenética ten a súa orixe primaria nas mutacións, pero as frecuencias alélicas nun *locus* varían dunha xeración a outra debido a diversas forzas naturais tales como a selección natural, o efecto fundador, a deriva xénica e as mesturas poboacionais. Esta disciplina permitiu descubrir as orixes de moitas variantes xenéticas de crucial importancia para a supervivencia do ser humano e posibilitou coñecer as rutas migratorias dos nosos devanceiros.

Baseándonos nestas premisas, e dispoñendo de multitude de marcadores xenéticos xa descritos e caracterizados, de aplicación común en ambas disciplinas, no traballo desenvolvido na presente tese doutoral propuxémonos os seguintes obxectivos principais:

1. Coñecer cales son os mellores marcadores xenéticos para o estudo do ADN degradado, tanto en xenética forense como en xenética de poboacións.
2. Pór en práctica o uso de marcadores SNP en xenética forense.
3. Aumentar o coñecemento de marcadores de cromosoma Y.

4. Aplicar o uso de marcadores de cromosoma Y para o estudo de poboacións humanas e da súa evolución.

De forma máis detallada, as tarefas propostas para acadar os citados obxectivos foron as seguintes:

1. Coñecer cales son os mellores marcadores xenéticos para o estudo do ADN degradado, tanto en xenética forense como en xenética de poboacións.

- 1.1. Atopar os marcadores xenéticos máis axeitados para incrementar a posibilidade de amplificación do ADN moi degradado (marcadores STR/ marcadores mini-STR/ marcadores SNP) e incluílos nun novo núcleo de *loci* que constitúa as bases de datos internacionais de ADN.

2. Pór en práctica o uso de marcadores SNP en xenética forense.

- 2.1. Comprobar que o xenotipado de marcadores SNP é un complemento real nos ensaios dispoñibles na investigación forense, para o cal se propón un estudo entre varios laboratorios que permita validar un *multiplex* de 52 SNP.

- 2.2. Avaliar a tecnoloxía *GenPlex* no xenotipado dun *multiplex* de 49 SNP, así como valorar o carácter informativo de dito *multiplex* para uso forense.

3. Aumentar o coñecemento de marcadores de cromosoma Y.

- 3.1. Realizar unha estimación fiable das taxas de mutación dos marcadores STR de cromosoma Y, atendendo á variación específica de *locus* e de alelo. Preténdese así poder mellorar a interpretación dos datos nos estudos de parentesco, así como conseguir datar dun xeito máis fiable a antigüidade das liñaxes definidas polos SNP (haplogrupos), polo tanto, con relevantes aplicacións na xenética forense e poboacional.

- 3.2. Caracterización a nivel xenético das liñaxes masculinas de Francia. Preténdese coñecer o grado de subestruturación poboacional da poboación francesa actual, de especial interese na utilización das bases de datos de cromosoma Y en casuística forense.

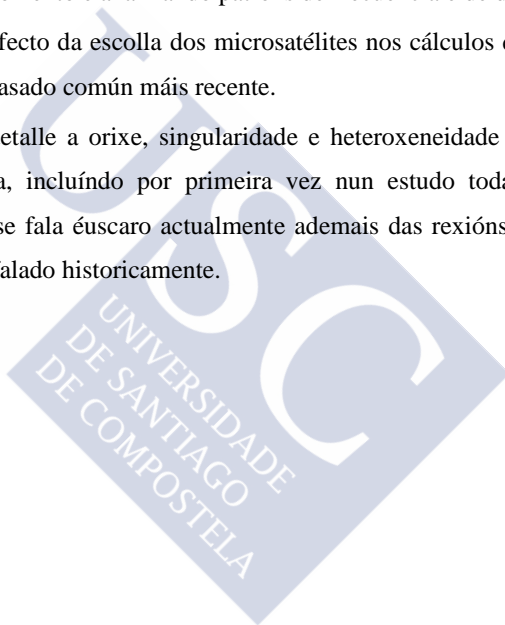
4. Aplicar o uso de marcadores de cromosoma Y para o estudo de poboacións humanas e da súa evolución.

4.1. Caracterización a nivel xenético das liñaxes masculinas de Francia, unha poboación ata ese momento escasamente caracterizada a nivel de cromosoma Y, e cunha localización xeográfica crucial nos movementos de repoboación de Europa ao longo do tempo, polo que tamén se pretende contribuír á reconstrución da historia demográfica das poboacións europeas.

4.2. Esclarecer a orixe e dispersión da liñaxe R-M269 do cromosoma Y, o principal haplogrupo europeo, reunindo a maior colección de cromosomas de dita liñaxe ata o momento e analizando patróns de frecuencia e de diversidade.

4.3. Investigar o efecto da escolla dos microsátélites nos cálculos da estimación do tempo ao antepasado común máis recente.

4.4. Coñecer en detalle a orixe, singularidade e heteroxeneidade dos vascos a través da xenética, incluíndo por primeira vez nun estudo todas as rexións xeográficas onde se fala éuscaro actualmente ademais das rexións circundantes onde se puido ter falado historicamente.





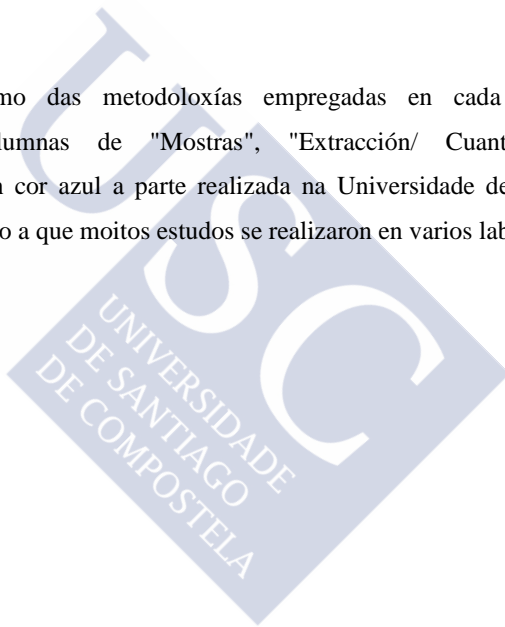


### **3 MATERIAL E MÉTODOS**



A presente memoria de tese reúne un conxunto de traballos de investigación xa publicados que se presentan no apartado de resultados. En cada un destes artigos descríbese o apartado de material e métodos correspondente, explicando detalladamente o estudo que se desenvolveu en cada un deles. Coa finalidade de ter unha visión conxunta das metodoloxías empregadas, resúmense na seguinte táboa os aspectos máis relevantes dos traballos presentados, atendendo ás mostras empregadas, aos métodos de extracción, cuantificación e xenotipado utilizados e ás análises que se realizaron.

**Táboa 2:** Táboa resumo das metodoloxías empregadas en cada traballo de investigación. Nas columnas de "Mostras", "Extracción/ Cuantificación" e "Xenotipado" márcase en cor azul a parte realizada na Universidade de Santiago de Compostela (USC), debido a que moitos estudos se realizaron en varios laboratorios.



ARTIGO	MOSTRAS	EXTRACCIÓN/ CUANTIFICACIÓN	XENOTIPADO	ANÁLISE
Dixon et al. (2006)	<p>-Mostras de sangue e saliva degradadas artificialmente.</p> <p>-Mostras de ADN en baixo número de copias (Low Copy Number).</p>	<p><b>Extracción:</b></p> <p>-QIAamp ou QIAshredder (Qiagen™).</p> <p>-Fenol-cloroformo.</p> <p>-Chelex.</p> <p><b>Cuantificación:</b></p> <p>-Quant-IT™ PicoGreen® dsDNA Assay Kit (Life Technologies). Quantifiler™ Human DNA Quantification kit (LT), metodoloxías de Slot-blot e PCR cuantitativa en tempo real con sonda TaqMan® en secuencia Alu.</p>	<p><b>STR:</b></p> <p>-AMPF/STR® SGM Plus™ ou AMPF/STR® Identifier (LT).</p> <p>-PowerPlex® 16 system (Promega).</p> <p>-mini-SGM e miniNCO1 (National Institute of Standards and Technology (NIST)).</p> <p><b>SNP:</b></p> <p>-21 loci no kit multiplex Foren-SNP™ (The Forensic Science Service®).</p>	<p>-Análise estatística empregando ANOVA, a mediana "pulida" ("median polish method") e diagramas de caixas.</p>
Musgrave-Brown et al. (2007)	<p>-Mostras de ADN en baixo número de copias (ADN control das liñas celulares de Promega 9948 e 9947a).</p> <p>-Mostras de ADN degradado.</p> <p>-Mostras de ADN degradado e en baixo número de copias (extractos de saliva e sangue baixo condicións cálidas e húmidas durante varios meses).</p>	<p><b>Extracción:</b></p> <p>-Cada laboratorio participante empregou os seus propios procedementos de rutina.</p> <p><b>Cuantificación:</b></p> <p>-Quantifiler™ Human DNA Quantification kit (LT).</p>	<p><b>STR:</b></p> <p>-PowerPlex® 16 system (Promega).</p> <p><b>SNP:</b></p> <p>-SNPforID 52-plex: SNaPshot® Multiplex kit (LT).</p>	<p>-Análise estatística calculando a porcentaxe de xenotipado correcto e a porcentaxe de alelos "drop-in".</p>
Phillips et al. (2007)	<p>-Placa control de Applied Biosystems: 44 mostras de ADN duplicadas da European Collection of Cell Cultures.</p> <p>-24 tríos alemáns.</p> <p>-3 familias extensas Utah do CEPH.</p>		<p><b>STR:</b></p> <p>-SEfiler™ (LT) (tríos alemáns).</p> <p>-PowerPlex® 16 system (Promega) (tríos alemáns).</p> <p>-AMPF/STR® Identifier (LT) (familias CEPH).</p> <p><b>SNP:</b></p> <p>-48 SNP: GenPlex™ (LT).</p> <p>-39 SNP: iPLEX® MassARRAY (Sequenom).</p> <p>-SNPforID 52-plex: SNaPshot® Multiplex kit (LT).</p> <p>-43 SNP: TaqMan® real-time PCR (LT).</p>	<p>-Valor informativo dos 48 SNP: probabilidade de coincidencia.</p> <p>-DNAVIEW v.28.01: análise estatística dos marcadores STR e SNP dos tríos alemáns.</p> <p>-CERVUS v2.0: análise de distintos parámetros de paternidade das familias CEPH.</p>

ARTIGO	MOSTRAS	EXTRACCIÓN/ CUANTIFICACIÓN	XENOTIPADO	ANÁLISE
Gusmão et al. (2005)	-3.026 pares de mostras pai/filho (100 deles na USC).	-Cada laboratorio participante empregou os seus propios procedementos de rutina.	<b>STR-Y:</b> -17 loci (cada laboratorio participante empregou os seus propios procedementos de rutina).	-Estimación dos intervalos de confianza da desviación binomial estándar para as taxas de mutación.  -Cálculo da desviación á media das taxas de mutación locus e alelo específicas mediante a proba de $\chi^2$ .
Ramos-Luis et al. (2014)	-558 mostras de sangue de homes non emparentados de 7 poboacións francesas.	<b>Extracción:</b> -QIAamp® DNA Blood Mini Kit (Qiagen). <b>Cuantificación:</b> -Quant-iT™ PicoGreen® dsDNA Assay Kit (LT).	<b>STR-Y:</b> -AmpFSTR® Yfiler® PCR Amplification Kit (LT). <b>SNP-Y:</b> -27 loci: SNaPshot® Multiplex kit (LT).	-Arlequin 3.5.1.3: --cálculo de frecuencias e diversidade de haplogrupos e haplotipos; --análise de varianza molecular (AMOVA); --cálculo de distancias xenéticas por pares de poboacións, con valores de $F_{ST}$ para SNP e $R_{ST}$ para STR.  -SPSS v. 17.0: gráficas de escalada multidimensional (MDS, <i>Multidimensional scaling plot</i> ) das distancias xenéticas entre poboacións.
Busby et al. (2012)	-2.486 cromosomas Y do haplogrupo R-M269 de Europa, Oriente Próximo e Oeste asiático (novos e publicados) (deles 327 eran de 7 poboacións francesas e xenotipáronse na USC).	-Cada laboratorio participante empregou os seus propios procedementos de rutina. <b>-USC:</b> --Extracción: QIAamp® DNA Blood Mini Kit (Qiagen). --Cuantificación: Quant-iT™ PicoGreen® dsDNA Assay Kit (LT).	<b>STR-Y:</b> -10 loci: AmpFSTR® Yfiler® PCR Amplification Kit (LT) ou PowerPlex® Y System (Promega). <b>SNP-Y:</b> -6 loci: SNaPshot® Multiplex kit (LT).	-A <sub>IC</sub> M <sub>AP</sub> GIS (v. 9.2; ESRI): mapas de frecuencias de SNP.  -Paquete estatístico R: media da varianza dos STR, e comparación coa lonxitude e latitude por unha análise de regresión.  -Estimación do tempo de coalescencia utilizando ASD ( <i>Average squared distance</i> , a media do cadrado das distancias) e BATWING.

ARTIGO	MOSTRAS	EXTRACCIÓN/ CUANTIFICACIÓN	XENOTIPADO	ANÁLISE
Martínez-Cruz et al. (2012)	<p>-886 mostras de individuos non relacionados de 18 áreas xeográficas do País Vasco e arredores.</p> <p>-Na contextualización das poboacións do estudo:</p> <p>--mostras de cromosoma Y de 14 poboacións ibéricas e de 7 poboacións francesas (USC).</p> <p>--mostras de ADNmt de 15 poboacións ibéricas e de 9 francesas.</p>	<p>-Cada laboratorio participante empregou os seus propios procedementos de rutina.</p> <p>--USC:</p> <p>--Extracción: QIAamp® DNA Blood Mini Kit (Qiagen).</p> <p>--Cuantificación: Quant-iT™ PicoGreen® dsDNA Assay Kit (LT).</p>	<p><b>STR-Y:</b></p> <p>-19 loci: 17 loci (AmpFISTR® Yfiler® PCR Amplification Kit (LT)) e 2 loci en multiplex (DYS426 e DYS388).</p> <p>-10 loci: AmpFISTR® Yfiler® PCR Amplification Kit (LT).</p> <p><b>SNP-Y:</b></p> <p>-54 loci: TaqMan® real-time PCR (LT).</p> <p>-6 Indels (multiplex).</p> <p>-27 loci: SNaPshot® Multiplex kit (LT).</p> <p><b>ADNmt:</b></p> <p>-Secuenciación de HVS-I e HVS-II da rexión control.</p> <p>-22 SNP: GenoCoRe22 (SNaPshot® Multiplex kit (LT)).</p> <p>-17 SNP (multiplex "HPLEX17" que diferencia subhaplogrupos do haplogrupo H).</p>	<p>-Arlequin 3.4: cálculo de parámetros de diversidade</p> <p>--diversidade de haplogrupos;</p> <p>--número de haplotipos de STR observado;</p> <p>--distancias xenéticas por pares, con valores de <math>F_{ST}</math> para SNP e <math>R_{ST}</math> para STR, corrixidos por Bonferroni;</p> <p>--valores de diversidade de secuencia;</p> <p>--media das diferenzas por pares.</p> <p>-"Test-T".</p> <p>-STATISTICA 7: Análise de compoñentes principais (ACP).</p> <p>-Análise de varianza molecular (AMOVA)</p> <p>-Genepop: "Test Mantel".</p> <p>-Estimación do TMRCA: BATWING, empregando valores da media da varianza dos STR-Y (haplogrupos de cromosoma Y) e de Network 4.5.0.0 (haplogrupos de ADNmt).</p>



## **4 RESULTADOS**



## 4.1 Polimorfismos autosómicos

### 4.1.1 Análise de ADN degradado utilizando marcadores STR e SNP –resultados dun exercicio europeo de colaboración (EDNAP)-

Nos últimos anos existiu un gran debate sobre que tipo de marcadores xenéticos deberían empregarse nun novo conxunto de *loci* para que constituían as bases de datos nacionais de ADN. As opcións existentes radican no uso de marcadores STR (cuxo tamaño de amplición varía entre 100 e 450 pb), marcadores mini-STR (de amplicións menores de 200 pb) e marcadores SNP. Os grupos EDNAP (*European DNA Profiling Group*) e ENFSI (*European Network of Forensic Science Institutes*) consideraron que a razón para buscar novos marcadores é incrementar a posibilidade de amplificar ADN moi degradado, máis que para incrementar o poder de discriminación das técnicas actuais.

Con esta fin realizouse un estudo de colaboración entre nove laboratorios europeos e de EEUU, baixo a dirección da EDNAP. Cada laboratorio, coas súas propias técnicas, debía xenotipar un set de ADN degradado (formado por manchas de saliva e de sangue) utilizando un *kit Foren-SNP™* (un *multiplex* de SNP do FSS - *Forensic Science Service*-) e dous *kits* mini-STR (dous *multiplexes* de marcadores STR do NIST - *National Institute of Standards and Technology*-).

En xeral, os marcadores mini-STR resultaron ser os máis efectivos, por este motivo, os grupos de traballo EDNAP e ENFSI recomendan redeseñar os *loci* STR existentes para dispoñer de amplicións menores e, por outra parte, incluír tres novos *loci* aos estándar europeos.



As páxinas correspondentes ao artigo foron eliminadas da versión dixital da memoria de tese e pode atoparse nos seguintes enlaces:

<http://www.ncbi.nlm.nih.gov/pubmed/16343834>

<http://www.sciencedirect.com/science/article/pii/S0379073805006146>





#### 4.1.2 Validación forense do ensaio 52-plex de SNPforID

As vantaxes da análise dos marcadores SNP en xenética forense son ben coñecidas e inclúen a posibilidade de empregar un amplo rango de tecnoloxías de xenotipado, baixas taxas de mutación, e melloras na análise de mostras degradadas. Non obstante, se os marcadores SNP queren converterse nunha alternativa real aos actuais métodos de xenotipado de marcadores STR, deben demostrar a súa fiabilidade e éxito na análise de mostras problemáticas, moi frecuentes en casuística forense.

O consorcio europeo SNPforID, pertencente ao programa “EU GROWTH”, desenvolveu un *multiplex* de 52 SNP para análise forense, amplificando os 52 *loci* nunha única reacción seguida de dúas reaccións de extensión de base única (“*single base extension*”, SBE) que son detectadas mediante electroforese capilar.

Coa fin de validar este ensaio, preparáronse varias mostras tratando de reflectir unha problemática frecuente en casuística forense, como é o ADN degradado e en número de copias baixo. Preparáronse un total de 40 mostras e cada unha delas foi enviada a dous dos cinco laboratorios participantes, para ser xenotipada por duplicado ou triplicado. Aos laboratorios indicóuselle que tiñan que analizar as mostras como se estiveran tratando con mostras de casos reais. Os resultados foron enviados ao laboratorio coordinador e comparados cos resultados da análise das mesmas mostras co *kit* de marcadores STR *PowerPlex® 16 System (Promega)*. Os resultados amosan as vantaxes do 52-plex fronte aos STR na análise de mostras degradadas e nos casos de mostras degradadas con baixa cantidade de ADN, o que suxire que a análise de marcadores SNP ofrecería vantaxes fronte aos STR en circunstancias forenses relevantes. Porén, ao longo do estudo presentáronse problemas relacionados coa contaminación e coa calidade dos cebadores, que son tamén discutidos.



As páxinas correspondentes ao artigo foron eliminadas da versión dixital da memoria de tese e pode atoparse nos seguintes enlaces:

<http://www.ncbi.nlm.nih.gov/pubmed/19083753>

<http://www.sciencedirect.com/science/article/pii/S1872497307000415>





### 4.1.3 Avaliación do sistema de xenotipado de SNP *GenPlex* e o panel de marcadores forenses 49-plex

Empregando o ensaio 52-plex previamente desenvolvido con fins forenses, preséntase un novo ensaio, o 49-plex, coa tecnoloxía *GenPlex*<sup>TM</sup>. Esta consiste nunha modificación do sistema *SNPlex*<sup>TM</sup> (ambos de *Life Technologies*) baseado na ligazón de oligonucleótidos do ADN previamente amplificado e no uso de sondas con marcaxe fluorescente e con modificadores da mobilidade, para a súa detección mediante electroforese capilar. Os *loci* escollidos son os 48 *loci* máis informativos dos SNP autosómicos de *SNPforID* e engadíuselle a ameloxenina como marcador de sexo. Estes marcadores SNP están distribuídos ao longo dos 22 cromosomas autosómicos, mostran un grado de polimorfismo equilibrado entre os tres grupos poboacionais maioritarios e demostraron previamente ser marcadores eficaces para análises forenses. Analizouse a precisión e reproducibilidade do sistema *GenPlex* en tres laboratorios do consorcio *SNPforID*, cada un cun secuenciador de *Life Technologies* diferente. A concordancia do xenotipado mediuse utilizando réplicas de 44 controis de ADN estandarizados e comparando cos xenotipos das mesmas mostras xerados polos sistemas de xenotipado de SNP: *TaqMan*<sup>®</sup> *real-time* PCR (LT), *SNaPshot*<sup>®</sup> *Multiplex kit* (LT) e *iPLEX*<sup>®</sup> *MassARRAY* (Sequenom). O grado de información para a análise forense dos 48 SNP foi medido utilizando as frecuencias alélicas previamente estimadas, para obter a probabilidade de coincidencia acumulada e, na análise de paternidade, empregando 24 tríos previamente xenotipados con 18 STR xunto con tres familias da colección do CEPH con gran descendencia xenotipada cos 15 STR do *kit* *AMPFISTR*<sup>®</sup> *Identifiler* (LT).



As páxinas correspondentes ao artigo foron eliminadas da versión dixital da memoria de tese e pode atoparse nos seguintes enlaces:

<http://www.ncbi.nlm.nih.gov/pubmed/19083752>

<http://www.sciencedirect.com/science/article/pii/S1872497307000610>





## 4.2 Polimorfismos de cromosoma Y

### 4.2.1 Taxas de mutación de microsatélites específicos de cromosoma Y

Levou-se a cabo un estudo de colaboración dirixido polo grupo de traballo da ISFG do grupo español e portugués (GEP-ISFG) para estimar taxas de mutación de marcadores STR de cromosoma Y. Nunha mostra de 3.026 pares pai/fillo analizáronse 17 *loci* de STR de cromosoma Y (DYS19, DYS385, DYS389I e II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS460, DYS461, DYS635 [GATA C4], GATA H4, e GATA A10). Observáronse 54 mutacións entre 27.029 transferencias alélicas, cunha taxa de mutación global dos 17 *loci* de  $1,998 \times 10^{-3}$  (95% IC:  $1,501 \times 10^{-3}$ - $2,606 \times 10^{-3}$ ). Todas as mutacións foron dun único evento de mutación e foron observadas só unha vez por gametoxénese. As ganancias de repeticións foron máis frecuentes que as perdas, os alelos longos resultaron ser máis mutables e a taxa de mutación parecía aumentar coa idade do pai.





As páxinas correspondentes ao artigo foron eliminadas da versión dixital da memoria de tese e pode atoparse nos seguintes enlaces:

<http://www.ncbi.nlm.nih.gov/pubmed/16220553>

<http://onlinelibrary.wiley.com/doi/10.1002/humu.20254/abstract;jsessionid=32D063A1FC6A862F1363FD43A3B1973F.f03t02>





#### 4.2.2 Análise do ADN de cromosoma Y de liñaxes masculinas francesas

A poboación francesa, a pesar da súa crucial localización xeográfica para os movementos de repoboación de Europa ao longo do tempo, ten sido insuficientemente caracterizada a nivel xenético, especialmente para a variación do ADN de cromosoma Y. Coa fin de realizar unha caracterización da estrutura xenética, analizamos a diversidade de cromosoma Y de 558 individuos, espallados ao longo de sete rexións francesas: *Alsace (Strasbourg)*, *Auvergne (Clermont-Ferrand)*, Bretaña ou *Bretagne (Rennes)*, *Île-de-France (Paris)*, *Midi-Pyrénées (Toulouse)*, *Nord-Pas-de-Calais (Lille)* e *Provence-Alpes-Côte d'Azur (Marseille)*. Para cada individuo se xenotiparon un total de 17 *loci* STR e 27 *loci* SNP, todos no cromosoma Y, e atopouse que a maioría das poboacións francesas non se diferenciaban unhas das outras. Porén, a poboación de Bretaña mostrou subestruturación poboacional, un feito importante a ser considerado cando se establecen bases de datos de poboación xeral.





As páxinas correspondentes ao artigo foron eliminadas da versión dixital da memoria de tese e pode atoparse nos seguintes enlaces:

<http://www.ncbi.nlm.nih.gov/pubmed/24528594>

<http://www.sciencedirect.com/science/article/pii/S187249731300255X>





### **4.2.3 O poboamento de Europa e o cauteloso relato da liñaxe R-M269 de cromosoma Y.**

Recentemente, reavivouse o debate sobre as orixes do principal haplogrupo europeo de cromosoma Y, o chamado R1b1b2 ou R-M269, e a opinión afastouse das orixes paleolíticas do mesmo cara a idea dunha difusión neolítica máis temperá destes cromosomas desde Oriente Próximo. Nós tratamos este debate estudando patróns de frecuencia e de diversidade na maior colección de cromosomas R-M269 reunida ata ese momento. A nosa análise revela que non hai tendencias xeográficas na diversidade, en contraposición coas expectativas baixo a hipótese neolítica, e suxire unha explicación alternativa para a aparente clina de diversidade recentemente descrita. Ademais, examinamos as estimacións do tempo ao antepasado común máis recente baseadas en marcadores STR propostas ata o momento para as liñaxes relacionadas con R-M269, e encontramos evidencias dun considerable efecto da escolla dos microsátélites nas estimacións de idade. Como consecuencia, os datos e ferramentas existentes son insuficientes para facer estimacións fiables da idade deste haplogrupo, e as conclusións sobre a súa orixe e dispersión débense considerar cun elevado nivel de precaución.

*Nota: o material suplementario deste traballo de investigación inclúese no apartado de “Anexos” da presente memoria de tese.*



As páxinas correspondentes ao artigo foron eliminadas da versión dixital da memoria de tese e pode atoparse nos seguintes enlaces:

<http://www.ncbi.nlm.nih.gov/pubmed/21865258>

<http://rspb.royalsocietypublishing.org/content/279/1730/884.long>





#### **4.2.4 Evidencia de estrutura xenética tribal prerromana en vascos a través de marcadores herdados uniparentalmente.**

O pobo vasco recibiu unha atención considerable de antropólogos, xenetistas e lingüistas durante o último século debido á singularidade da súa lingua e a outras características culturais e biolóxicas. A pesar dos esforzos multidisciplinares realizados para abordar as cuestións sobre a orixe, a singularidade e a heteroxeneidade dos vascos, os estudos xenéticos realizados ata o momento sufrían dun deseño de estudo feble, no cal as poboacións non eran analizadas nun contexto xeográfico e poboacional axeitado. Para abordar as cuestións anteriores e superar estas limitacións de deseño, analizamos os marcadores de herdanza uniparental (cromosoma Y e ADN mitocondrial) duns 900 individuos de 18 poboacións, incluíndo aquelas onde o vasco se fala actualmente e poboacións de rexións adxacentes onde o vasco podería terse falado en tempos históricos. Os nosos resultados indican que as poboacións de fala vasca caen dentro do acervo xenético de Europa occidental e son similares ás poboacións non vascas xeográficamente circundantes, e tamén que a súa singularidade xenética está baseada nunha menor cantidade de influencias externas comparado con outras poboacións ibéricas e francesas. Os nosos datos suxiren que a heteroxeneidade e estrutura xenética observada no territorio vasco resulta dunha estrutura tribal prerromana relacionada coa xeografía e podería estar asociada ao incremento de complexidade das sociedades emerxentes durante a Idade de Bronce. A irregular superposición do emprazamento das tribos prerromanas e os actuais límites dialectais apoia a idea de que a diversidade ambiental na rexión xogou un papel recorrente na diferenciación cultural e na etnoxénese nos diferentes períodos de tempo.

*Nota: o material suplementario deste traballo de investigación inclúese no apartado de “Anexos” da presente memoria de tese.*



As páxinas correspondentes ao artigo foron eliminadas da versión dixital da memoria de tese e pode atoparse nos seguintes enlaces:

<http://www.ncbi.nlm.nih.gov/pubmed/22411853>

<http://mbe.oxfordjournals.org/content/29/9/2211.full.pdf+html>







## **5 DISCUSIÓN**



### Os SNP autosómicos na xenética forense

Os avances científicos en xenética forense viñeron supeditados aos avances tecnolóxicos, pondo de manifesto a dispoñibilidade dun gran número de tecnoloxías de aplicación na disciplina e dun gran número de marcadores, diversos en canto ás súas condicións e características. Con todo, tanto os marcadores como as tecnoloxías requiren extensas e precisas validacións que permitan que as estritas sociedades de estandarización, reguladoras no campo, os acepten. Neste sentido a presente tese doutoral intentou colaborar en ditas validacións, en primeiro lugar, valorando cales serían os mellores marcadores xenéticos para o estudo de mostras problemáticas en xenética forense, como poden ser aquelas con ADN degradado ou ADN limitante e, en segundo lugar, tratando de estender o uso dos marcadores SNP no campo, a través da validación de diversas tecnoloxías de análise, como o *SNaPshot<sup>®</sup> Multiplex kit (LT)* ou o *GenPlex<sup>™</sup> (LT)*.

Deste xeito, un dos primeiros estudos que se levou a cabo pretendía valorar a eficacia de distintos tipos de marcadores no xenotipado de ADN artificialmente degradado procedente de saliva e de sangue, lográndose simular a secuencia temporal de degradación das manchas tal e como sería no seu estado natural. Cando se compararon os resultados de xenotipado obtidos cos marcadores STR (empregando o *kit SGM+*), co *kit de Foren-SNP<sup>™</sup>* e con dous *kits* de marcadores mini-STR, foron con estes últimos cos que se obtiveron, en xeral, os mellores resultados. Con todo, o método de amplificación empregado nos experimentos de ADN en número de copias baixo ou LCN (*Low Copy Number*), foi o que tivo maior éxito de xenotipado, debido a que nese caso se aumenta o número de ciclos de amplificación por PCR: dos 28 ciclos que se empregan nas PCR estándar, pásase a 34 cando se amplifica o ADN en número de copias baixo (Gill et al. 2000). Cando se utilizan as condicións de LCN, aumenta a posibilidade de amplificación das moléculas, pero tamén se dificulta a interpretación do perfil, ao verse empeorado o equilibrio alélico e ao aumentaren os alelos nulos.

Os marcadores mini-STR foron amplificados nunhas condicións intermedias en canto ao número de ciclos de amplificación, utilizándose 32 ciclos. Probablemente a combinación do aumento do número de ciclos da amplificación destes marcadores unido ao menor tamaño de amplicón, foi o que provocou que foran os ensaios con maior taxa de éxito no xenotipado. Ademais, o *kit* que obtivo maior porcentaxe de perfís

completos foi o *kit* mini-NC01, debido tamén a que era o *multiplex* con menor número de *loci*, só tres, e polo tanto, adoita ser moito máis doado de optimizar que os *multiplex* con maior número de *loci*.

Aínda que neste primeiro estudo o *kit Foren-SNP*<sup>TM</sup> obtivo malos resultados con ADN degradado artificialmente, no seguinte estudo de validación doutro *multiplex* de SNP, o *SNPforID 52-plex*, os resultados cos SNP foron mellores que os obtidos co *kit* de STR *PowerPlex*<sup>®</sup> *16 system* (Promega), cando se analizaron mostras degradadas.

Os mellores marcadores para ADN degradado parecen ser os de amplicón de menor tamaño, con independencia de que sexan marcadores SNP ou mini-STR, porén, os SNP son os que permiten obter menor tamaño de amplicón. Unha posible explicación para os malos resultados obtidos cos *Foren-SNP*<sup>TM</sup> podería ser que a maioría dos cebadores deste ensaio eran moi longos, maiores de 35 pb, e o transporte entre distintos países puido aumentar os ciclos de conxelación-desconxelación, provocando a degradación dos cebadores longos e, polo tanto, a falta de uniformidade nos resultados presentados polos diferentes grupos. De feito, cando se escolleron os 10 mellores SNP de dito ensaio, os resultados xa se igualaron aos obtidos cos mini-STR.

Por outra banda, debido á natureza bialélica dos SNP, a interpretación de mesturas de mostras é máis complicada se non se consegue un ensaio ben balanceado. Por estes motivos, para os casos forenses onde son moi frecuentes as mesturas de mostras, préfirense os STR. En casos de ADN moi degradado, a redución do tamaño do amplicón dos STR é recomendada, e pódense incrementar as posibilidades de éxito ao usar as condicións de LCN ademais das convencionais.

Na validación forense do 52-plex de SNP, tal e como era de esperar, obtivéronse mellores resultados que cos STR, tanto no caso da análise de mostras degradadas coma no de mostras degradadas con ADN limitante; non obstante, os resultados coas mostras con ADN limitante non degradado foron peores que os dos STR.

Nestes tres tipos de mostras sempre se observou unha maior asignación de alelos que non estaban presentes na mostra de referencia, os chamados alelos “*drop-in*”. Existen varias posibles explicacións para este feito, como son as posibles diferenzas nos equipamentos que puidera haber entre os distintos laboratorios. Non obstante, a principal explicación é que a aparición de alelos “*drop-in*” é un feito aleatorio, e como neste estudo só se asignaba o alelo cando este aparecía como mínimo en dous dos dous ou tres xenotipados que se facían por mostra, a posibilidade de que aparecese por azar

dúas veces o mesmo alelo para calquera *locus* dado é maior no caso do 52-plex, porque só existen dous alelos posibles na poboación para cada *locus*, mentres que, para os marcadores STR do *Powerplex*<sup>®</sup>, o número de alelos posibles para cada *locus* é maior, polo que a posibilidade de observar o mesmo alelo “*drop-in*” máis dunha vez, redúcese moito. Así e todo, no noso estudo o “*drop-in*” foi sempre moito maior nun dos dous laboratorios que xenotipou co 52-plex, feito que fai pensar que tamén se debe a cuestións particulares do propio laboratorio, non só a cuestións do propio *multiplex*. Ademais, tamén se valorou a variación entre os resultados obtidos polos laboratorios participantes en dita validación, permitindo comprobar como os laboratorios que tiñan menor éxito no xenotipado eran os que detectaban maior grado de “*drop-in*”, e viceversa, poñendo de manifesto que a calidade dos resultados acadados co 52-plex depende en gran medida do laboratorio onde se obteñan.

No tocante aos resultados obtidos nos extractos con ADN degradado, independentemente da cantidade de ADN dispoñible, son claramente mellores os conseguidos co 52-plex, xa que co *PowerPlex*<sup>®</sup> o rendemento do xenotipado diminúe rapidamente coa media do tamaño do fragmento de ADN. Isto débese a que o rango do tamaño dos amplicóns do 52-plex é de 59 a 115 pb, mentres que no caso do *PowerPlex*<sup>®</sup> o rango vai de 107 a 473 pb. Polo tanto, tal e como xa mencionamos anteriormente, aquí radica a maior potencial vantaxe da análise dos SNP en casuística forense.

En relación á avaliación da tecnoloxía *GenPlex*<sup>™</sup> (LT) para a análise dun *multiplex* de 49 SNP, comprobouse a existencia dunha concordancia total entre os resultados obtidos con esta tecnoloxía e os acadados coas outras tres alternativas coas que se realizou o xenotipado do 49-plex: *TaqMan*<sup>®</sup> *real-time* PCR (LT), *SNaPshot*<sup>®</sup> *Multiplex kit* (LT) e *iPLEX*<sup>®</sup> *MassARRAY* (Sequenom). Ademais, o *GenPlex* demostrou ser unha tecnoloxía fiable e consistente ao ser reproducible nos tres laboratorios onde se probou, un deles sen experiencia previa en “oligo-ligación”.

A análise dos resultados obtidos co *GenPlex* levouse a cabo co software *GeneMapper*<sup>™</sup>, que asigna os alelos de cada SNP comparando as posicións do conxunto de datos dispostos en tres grupos ou *clusters* (dous de homocigotos e un de heterocigotos), nunha representación de dúas dimensións. Esta característica é moi vantaxosa na asignación dos heterocigotos xa que estes raramente teñen un equilibrio perfecto entre os dous alelos, pero ao ser este un desequilibrio constante en todas as mostras para un SNP dado, os heterocigotos pódense asignar de forma inequívoca. Por

outra parte, nalgúns SNP observábase un sinal non específico na posición do alelo ausente no caso dos homocigotos, pero do mesmo xeito que acontece co desequilibrio de heterocigotos, o sinal inespecífico tamén é constante para cada SNP, independentemente das mostras e do ensaio onde se analizaran estas.

En estudos posteriores (Musgrave-Brown et al. 2008) comprobaron que esta característica facía tamén perfectamente diferenciables ás mostras formadas pola mestura de ADN de dous individuos diferentes, feito de crucial importancia en casuística forense.

Coa intención de valorar o carácter informativo deste ensaio tamén se calculou a probabilidade de coincidencia para o 49-plex e viuse que para poboación europea era a mesma que a obtida co *kit* AMPFISTR® *Identifiler* (LT), un *multiplex* de 15 STR autosómicos totalmente validado para casuística forense, o que indicaría que o 49-plex tamén é axeitado para este uso. No caso de poboación africana, a probabilidade de coincidencia diminuía lixeiramente, debido á menor variabilidade dalgúns dos SNP do *multiplex* en poboación africana. Así e todo, a probabilidade de coincidencia obtida co *multiplex* SGM+ en africanos chega a equipararse coa obtida cuns 42 SNP do *GenPlex*.

Por outra banda, valorouse a capacidade do 49-plex para realizar análises de paternidade comparada coa dos STR, confirmando que os 49 SNP eran menos informativos que os STR. Non obstante, o feito de apareceren tres exclusións en segunda orde en varios marcadores STR, nun número tan reducido de mostras analizadas, demostra a súa maior inestabilidade en comparación cos SNP.

Con todo, pese ao menor valor informativo dos 49 SNP incorporados con éxito no ensaio de *GenPlex* nos casos de paternidades normais, comprobamos que estes marcadores son máis eficaces para discriminar individuos emparentados.

A pesar de quedar demostrada a utilidade forense da tecnoloxía *GenPlex*, soamente estivo dispoñible un breve período de tempo, debido a que *Applied Biosystems* retirou do mercado a tecnoloxía *SNPlex*<sup>TM</sup> e con ela tamén *GenPlex*, por compartiren os mesmos reactivos.

## O estudo do cromosoma Y

Coa intención de valorar adecuadamente as posibles exclusións acontecidas por causa dalgunha mutación nos STR, é de vital importancia coñecer exactamente cales son as taxas específicas de mutación de cada un dos STR e cales son as taxas específicas de cada un dos alelos de cada STR. Neste sentido, a pesar de que a taxa de mutación na liña xerminal masculina é maior que a da liña xerminal feminina (Weber et al. 1993; Brinkmann et al. 1998a), non se atopan diferenzas significativas entre a taxa de mutación media dos STR do cromosoma Y e a dos autosómicos.

O valor da taxa de mutación media estimada para os 16 STR-Y estudados nesta tese foi de  $1,998 \times 10^{-3}$  (95% I. C.:  $1,501 \times 10^{-3}$ - $2,606 \times 10^{-3}$ ) e non mostra diferenzas significativas á obtida en estudos anteriormente publicados (Kayser et al. 2000; Dupuy et al. 2004). As diferenzas só son significativas para a taxa de mutación específica dalgún dos *loci* STR, en concreto, a taxa de mutación do DYS392 é significativamente menor á taxa de mutación media, e a dos *loci* DYS391 e DYS439 é maior á media. Confírmase tamén unha tendencia no aumento de mutabilidade nos alelos máis longos, aínda que se necesitaría un maior número de mostras para precisar máis esta relación.

En referencia ao modelo mutacional, o noso estudo apoia ao modelo de mutación “paso a paso” xeralmente aceptado, o cal considera que os alelos mutan principalmente por ganancias e perdas de unidades simples de repetición (Weber et al. 1993; Di Rienzo et al. 1994; Zhivotovsky et al. 2004), xa que así foron todas as mutacións atopadas, coa excepción dunha na que se perderon catro unidades de repetición. Ademais, as ganancias de unidades de repetición resultaron ser dúas veces máis frecuentes que as perdas, de acordo tamén con estudos previos realizados neste eido (Cooper et al. 1999; Kayser et al. 2000; Dupuy et al. 2004; Kurihara et al. 2004) e co esperado para microsatélites novos (Xu et al. 2000).

Os datos dispoñibles neste estudo non permitiron evidenciar a hipótese de Dupuy e colaboradores (Dupuy et al. 2004) que considera que existe un exceso de perdas de repeticións nos alelos máis longos fronte as ganancias, feito que impediría a expansión indefinida da lonxitude dos alelos dos STR, debido a que neste estudo o exceso de perdas non se observou unicamente nas clases de maior tamaño. Por este motivo, crese que o que debe acontecer é que os microsatélites teñen tendencia á expansión ata que, cando o STR chega a unha determinada lonxitude, a taxa de mutación de expansión igualase á taxa de mutación de contracción.

Ao contrario do estudo de Dupuy e colaboradores de 2004, atopamos unha asociación entre a presenza de mutacións e a idade do pai. Así e todo, analizando por clases de idade, aínda que parece que a taxa de mutación aumenta ao incrementarse o grupo de idade, as diferenzas só son significativas nunha das clases de idade.

Outro dos puntos cruciais da aplicación de marcadores xenéticos na práctica forense para poder dar un valor *probabilístico* á proba, é a dispoñibilidade de bases de datos que permitan determinar como de frecuente é a presenza dun alelo nun determinado marcador autosómico, ou como de frecuente é un haplotipo concreto de marcadores uniparentais. Tal e como xa citamos no apartado 1.5 da introdución, no caso do cromosoma Y existe unha gran base de datos de haplotipos de STR, a YHRD, que actualmente inclúe (con data 10/06/2014):

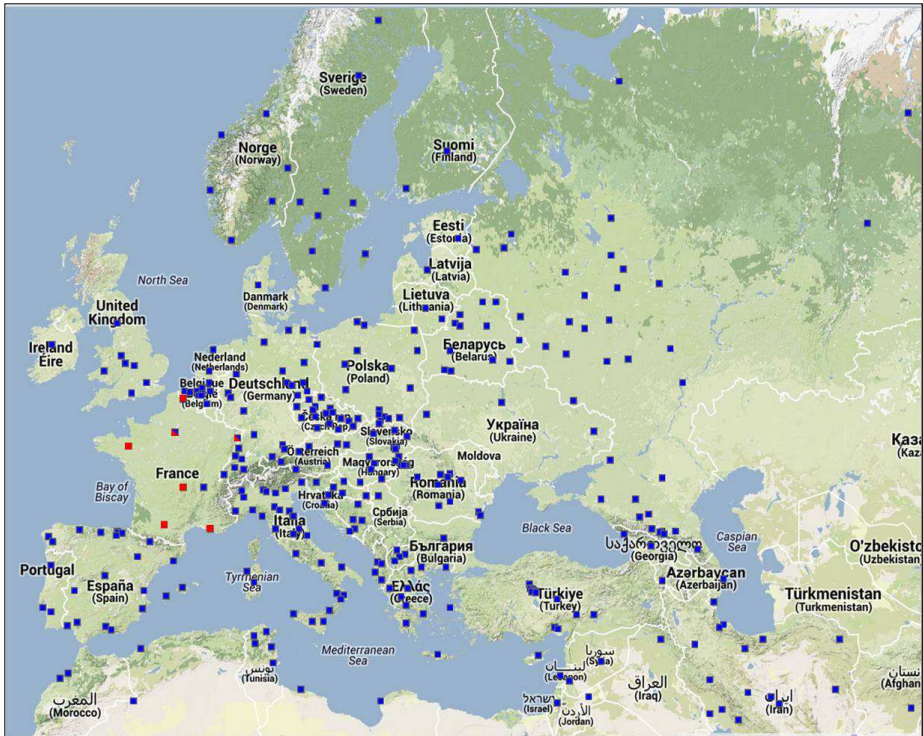
-Haplotipo mínimo (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS385a/b): 125.148 haplotipos.

-*Promega PowerPlex Y* (+ DYS437, DYS438, DYS439): 83.280 haplotipos.

-*Life Technologies AmpFISTR Yfiler* (+ DYS448, DYS456, DYS458, DYS635, YGATAH4): 71.235 haplotipos.

-*Promega PowerPlex Y23* (+ DYS576, DYS481, DYS549, DYS533, DYS570, DYS643): 19.592 haplotipos.

A pesar da elevada cantidade de haplotipos, na base de datos, tal e como podemos ver na figura 13, existen diversas rexións xeográficas moi mal cubertas, como é o caso de Francia, onde a gran maioría dos datos incluídos na base de datos son os obtidos na presente tese doutoral. Analizáronse 558 individuos de sete poboacións francesas diferentes, atopando que só a poboación da Bretaña se diferencia do resto de poboacións, tanto no xenotipado con marcadores STR como con SNP. A característica máis salientable é a baixa diversidade de haplogrupos desta rexión, o cal se explica porque máis do 68% da poboación ten o haplogrupo predominante europeo, o R-M269, e tamén porque o haplogrupo I-M170 ten unha frecuencia lixeiramente elevada en comparación co resto de poboacións. Polo tanto, a incorporación destas poboacións ás bases de datos, permite unha mellor interpretación dos estudos forenses levados a cabo con polimorfismos de cromosoma Y, permitindo valorar a frecuencia en determinadas poboacións que presentan subestruturación poboacional, é dicir, diferenzas significativas en canto á frecuencia de haplotipos con respecto ás poboacións veciñas.



**Figura 13:** Imaxe tomada da páxina da YHRD en xuño de 2014, na cal se indican as poboacións euroasiáticas das que se dispoñían datos de xenotipado nese momento. En vermello sinálase a contribución do noso grupo de investigación (Ramos-Luis et al. 2014) a dito mapa (as poboacións centradas en París e Estrasburgo foron tamén caracterizadas posteriormente por outros grupos de investigación –marcadas en vermello e azul-).

O estudo en profundidade do haplogrupo maioritario europeo, o R-M269, a partir da análise de datos máis ampla realizada ata ese momento, incluíu as mostras francesas caracterizadas nesta tese e permitiu confirmar a distribución espacial deste haplogrupo en dúas subliñaxes definidas polo SNP S127, unha subliñaxe europea e outra do oeste eurasiático. Corroboramos, polo menos en parte, a dicotomía observada por Morelli baseada na combinación de dous alelos de dous STR (Morelli et al. 2010).

Datar as liñaxes de cromosoma Y é unha cuestión claramente controvertida, principalmente porque a escolla dunha determinada taxa de mutación dos STR pode levar a estimacións moi diferentes. Así, na procura dun conxunto de STR-Y axeitado

para estimar a media do tempo de coalescencia (T) do subhaplogrupo R-S127, vimos que non todos os STR funcionan igual neste contexto. Comprobamos que tanto as estimacións baseadas no cálculo da ASD (*Average Squared Distance*, a media do cadrado das distancias) como as obtidas mediante o método Batwing, dependían en gran medida do conxunto de STR utilizados; si seleccionabamos os STR en función da duración da linearidade, D, a cal fai referencia á duración de tempo no pasado sobre a cal a ASD e o T manteñen unha relación lineal para un STR determinado, as estimacións feitas con ambos métodos variaban amplamente. A pesar de que a linearidade depende en gran medida da taxa de mutación ( $\mu$ ) e do rango de posibles alelos que un STR pode ter (R), a maior parte dos investigadores só teñen en conta a taxa de mutación dos STR cando estiman o tempo de diverxencia, e os STR maioritariamente usados non teñen os atributos específicos que permitan asumir unha linearidade máis alá no pasado. Por este motivo, a maioría das datacións de haplogrupos baseadas en ditos conxuntos de STR poden estar subestimadas.

No noso traballo non puidemos ofrecer ningunha estimación da idade dos haplogrupos R-M269 ou R-S127, pero permitiunos suxerir, por unha banda, que as estimacións recentes de R-M269 de Balaesque e colaboradores (Balaesque et al. 2010) e de R-S116 de Myres e colaboradores (Myres et al. 2011) son máis novas que os verdadeiros valores e, por outra banda, que a homoxeneidade da varianza dos STR e a distribución de subtipos a longo de todo o continente europeo, parecen inconsistentes coa hipótese da difusión neolítica da liñaxe R-M269 de cromosoma Y.

O estudo da poboación francesa permitiunos tamén colaborar na comprobación da hipótese da singularidade xenética dos vascos dentro da Europa occidental, a cal fora baseada en resultados contraditorios de marcadores clásicos, liñaxes uniparentais e SNP autosómicos. Os nosos datos amosan que o *pool* xenético das presentes liñaxes uniparentais vascas aseméllase ao resto de poboacións occidentais europeas, a pesar de que aparezan lixeiras diferenzas nas súas frecuencias. Cando a diversidade dos haplogrupos vascos se enmarca na das poboacións circundantes, a gráfica da análise de compoñentes principais suxire que todas as poboacións comparten un *pool* xenético ancestral común, aínda que algunhas poboacións se viron afectadas por influencias externas en distinta medida. As poboacións vascas, ademais, agrupan con outras ibéricas e francesas circundantes non vascas, o que suxire a existencia dunha singularidade

xenética nas poboacións que habitan esa rexión xeográfica, pero que non é exclusiva de falantes de vasco.

Ademais, a poboación da Bretaña francesa, distante xeograficamente, que non mostra haplogrupos norte africanos e que tivo escasas influencias neolíticas, cae no centro das poboacións vascas, sen que outras poboacións francesas máis próximas xeograficamente así o fagan. Os bretóns falan unha lingua céltica con raíces nas Illas Británicas e que non ten relación co éuscaro. Polo tanto, poboacións europeas occidentais xeográfica e etnicamente afastadas, poden presentar unha composición xenética similar aos vascos e a algunhas poboacións circundantes a estes, de novo sen que isto estea vinculado ao feito de ter unha cultura vasca.

As poboacións de fala éuscara, así como algunhas poboacións ao longo dos Pirineos (Lopez-Parra et al. 2009) amosan baixos niveis de diversidade nos seus cromosomas Y en comparación coas poboacións circundantes. Isto podería ser explicado polos elevados niveis de consanguinidade, especialmente en Gipuzkoa e nas zonas rurais, unido ao reducido fluxo xenético externo e concordando co illamento demográfico proposto para estas e outras poboacións vascas, sexa por motivos xeográficos ou culturais (Lopez-Parra et al. 2009). Este illamento demográfico suxerido podería producir algunha heteroxeneidade xenética entre as mostras vascas respecto ás non vascas, como se ve nas liñaxes de cromosoma Y pero non nas de ADNmt. Estes resultados opostos para as liñaxes maternas e paternas poderían explicar, polo menos parcialmente, a contradición entre os resultados que apoian e rexeitan a heteroxeneidade de vascos.

Os nosos resultados amosan que non existe unha microestrutura xenética significativa relacionada cos subdialectos que se falan actualmente en Euskal Herria. Con todo, algún nivel de estrutura xenética intrínseco aparece nos vascos, o que pode ser consecuencia de diferentes factores culturais, xeográficos e históricos, como pode ser a distribución das tribos prerromanas no territorio.

A reducida influencia externa que recibiron historicamente, tanto o País Vasco como a Bretaña francesa, fai pensar que estas dúas poboacións poden ser boas candidatas para representar o perfil xenético das poboacións europeas antigas.





## **6 CONCLUSIÓNS**



1. Os mellores resultados na análise de ADN degradado obtéñense con amplicóns pequenos, independentemente do marcador utilizado. Porén, os marcadores SNP son os que permiten o deseño dun amplicón máis reducido.
2. Os marcadores mini-STR resultaron ser máis efectivos que os *Foren-SNP*<sup>TM</sup> para a análise de ADN moi degradado. Recoméndase a redución do tamaño dos amplicóns dos *loci* STR comunmente empregados sempre que sexa posible e, ademais, os grupos da EDNAP e da ENFSI recomendan a inclusión de tres novos *loci* mini-STR nas bases de datos de ADN: os *loci* do *multiplex* de mini-STR NC01 (D10S1248, D14S1434 e D22S1045).
3. O 52-plex de SNP amosou mellores resultados que os STR, tanto no caso da análise de mostras degradadas coma no de mostras degradadas con ADN limitante; pola contra, os STR foron máis efectivos no estudo de mostras con ADN limitante non degradado.
4. A elevada presenza de alelos “*drop-in*” no 52-plex resalta a importancia de posuír uns equipamentos adecuados nos laboratorios, unha calidade de reactivos consistente así como unha maior estandarización na asignación de alelos.
5. O xenotipado con SNP, sen substituír ao dos STR, é unha boa alternativa cando os microsátélites non logran obter bos resultados con mostras problemáticas, como pode ser na análise de ADN degradado ou no caso de existir a necesidade de excluír individuos emparentados.
6. A reproducibilidade e fiabilidade da tecnoloxía *GenPlex*<sup>TM</sup> (*Life Technologies*) para a análise de SNP quedaron demostradas ao xenotipar con éxito paralelamente mostras estandarizadas en tres laboratorios e ao existir, ademais, unha total concordancia coas outras tres plataformas de xenotipado de SNP utilizadas: *TaqMan*<sup>®</sup> *real-time* PCR (LT), *SNaPshot*<sup>®</sup> *Multiplex kit* (LT) e *iPLEX*<sup>®</sup> *MassARRAY* (Sequenom).
7. A taxa de mutación global para marcadores STR específicos de cromosoma Y é de  $1,998 \times 10^{-3}$  (95% I.C.:  $1,501 \times 10^{-3}$ - $2,606 \times 10^{-3}$ ) aínda que varía moito entre os distintos *loci* estudados. Os nosos resultados apoian o modelo mutacional “paso a paso”,

comprobándose que nos STR estudados eran máis frecuentes as ganancias que as perdas de repeticións, e que os alelos longos eran máis susceptibles a mutación.

8. O estudo da variabilidade xenética das liñaxes masculinas da poboación francesa amosa que é unha poboación con baixo nivel de diferenciación xenética con respecto ao cromosoma Y, sendo unicamente a rexión da Bretaña a que amosa subestruturación poboacional, feito que debe ser considerado cando se establecen as bases de datos xenéticas de poboación xeral.

9. A caracterización xenética da poboación francesa xunto a doutras poboacións europeas permitiu definir a distribución espacial do haplogrupo R-M269 en función do SNP S127 en liñaxes europeas e liñaxes do oeste eurasiático.

10. Actualmente non podemos ofrecer ningunha estimación da idade dos haplogrupos R-M269 ou R-S127, pero a nosa análise de STR suxire que as estimacións previamente publicadas son máis novas que os verdadeiros valores e que a homoxeneidade da varianza dos STR e a distribución de subtipos a longo de todo o continente europeo, son inconsistentes coa hipótese da difusión neolítica da liñaxe R-M269 de cromosoma Y.

11. As estimacións do tempo de coalescencia dunha liñaxe dependen explicitamente do conxunto de STR que se utilice para inferilas. A maior parte das estimacións de TMRCA das liñaxes masculinas están feitas con STR cos que non se pode asumir linearidade no tempo, polo que poden estar subestimadas.

12. En contra da chamada singularidade xenética dos vascos, os seus xenomas uniparentais aseméllanse aos dalgunhas poboacións circundantes e aos dalgunhas poboacións non vascas afastadas xeograficamente como a da Bretaña francesa.

13. A poboación vasca amosa una reducida subestruturación, que non se correlaciona coa localización xeográfica dos diferentes dialectos falados, pero si coa distribución tribal prerromana desta rexión.



## **7 BIBLIOGRAFÍA**



- 1000-Genomes-Project-Consortium; Abecasis, G. R.; Auton, A.; Brooks, L. D.; DePristo, M. A.; Durbin, R. M.; Handsaker, R. E.; Kang, H. M.; Marth, G. T. and McVean, G. A. (2012). "An integrated map of genetic variation from 1,092 human genomes." *Nature* **491**(7422): 56-65.
- Ayub, Q.; Mohyuddin, A.; Qamar, R.; Mazhar, K.; Zerjal, T.; Mehdi, S. Q. and Tyler-Smith, C. (2000). "Identification and characterisation of novel human Y-chromosomal microsatellites from sequence database information." *Nucleic Acids Res* **28**(2): e8.
- Balaresque, P.; Bowden, G. R.; Adams, S. M.; Leung, H. Y.; King, T. E.; Rosser, Z. H.; Goodwin, J.; Moisan, J. P.; Richard, C.; Millward, A.; Demaine, A. G.; Barbujani, G.; Previdere, C.; Wilson, I. J.; Tyler-Smith, C. and Jobling, M. A. (2010). "A predominantly neolithic origin for European paternal lineages." *PLoS Biol* **8**(1): e1000285.
- Ballantyne, K. N.; Goedbloed, M.; Fang, R.; Schaap, O.; Lao, O.; Wollstein, A.; Choi, Y.; van Duijn, K.; Vermeulen, M.; Brauer, S.; Decorte, R.; Poetsch, M.; von Wurmb-Schwark, N.; de Knijff, P.; Labuda, D.; Vezina, H.; Knoblauch, H.; Lessig, R.; Roewer, L.; Ploski, R.; Dobosz, T.; Henke, L.; Henke, J.; Furtado, M. R. and Kayser, M. (2010). "Mutability of Y-chromosomal microsatellites: rates, characteristics, molecular bases, and forensic implications." *Am J Hum Genet* **87**(3): 341-53.
- Bellott, D. W.; Hughes, J. F.; Skaletsky, H.; Brown, L. G.; Pyntikova, T.; Cho, T. J.; Koutseva, N.; Zaghul, S.; Graves, T.; Rock, S.; Kremitzki, C.; Fulton, R. S.; Dugan, S.; Ding, Y.; Morton, D.; Khan, Z.; Lewis, L.; Buhay, C.; Wang, Q.; Watt, J.; Holder, M.; Lee, S.; Nazareth, L.; Rozen, S.; Muzny, D. M.; Warren, W. C.; Gibbs, R. A.; Wilson, R. K. and Page, D. C. (2014). "Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators." *Nature* **508**(7497): 494-9.
- Bhargava, A. and Fuentes, F. F. (2010). "Mutational dynamics of microsatellites." *Mol Biotechnol* **44**(3): 250-66.
- Birney, E.; Stamatoyanopoulos, J. A.; Dutta, A.; Guigo, R.; Gingeras, T. R.; Margulies, E. H.; Weng, Z.; Snyder, M.; Dermitzakis, E. T.; Thurman, R. E.; Kuehn, M. S.; Taylor, C. M.; Neph, S.; Koch, C. M.; Asthana, S.; Malhotra, A.; Adzhubei, I.; Greenbaum, J. A.; Andrews, R. M.; Flicek, P.; Boyle, P. J.; Cao, H.; Carter, N. P.; Clelland, G. K.; Davis, S.; Day, N.; Dhami, P.; Dillon, S. C.; Dorschner, M. O.; Fiegler, H.; Giresi, P. G.; Goldy, J.; Hawrylycz, M.; Haydock, A.; Humbert, R.; James, K. D.; Johnson, B. E.; Johnson, E. M.; Frum, T. T.; Rosenzweig, E. R.; Karnani, N.; Lee, K.; Lefebvre, G. C.; Navas, P. A.; Neri, F.; Parker, S. C.; Sabo, P. J.; Sandstrom, R.; Shafer, A.; Vetrie, D.; Weaver, M.; Wilcox, S.; Yu, M.; Collins, F. S.; Dekker, J.; Lieb, J. D.; Tullius, T. D.; Crawford, G. E.; Sunyaev, S.; Noble, W. S.; Dunham, I.; Denoeud, F.; Reymond, A.; Kapranov, P.; Rozowsky, J.; Zheng, D.; Castelo, R.; Frankish, A.; Harrow, J.; Ghosh, S.; Sandelin, A.; Hofacker, I. L.; Baertsch, R.; Keefe, D.; Dike, S.; Cheng, J.; Hirsch, H. A.; Sekinger, E. A.; Lagarde, J.; Abril, J. F.; Shahab, A.; Flamm, C.; Fried, C.; Hackermuller, J.; Hertel, J.; Lindemeyer, M.; Missal, K.; Tanzer, A.; Washietl, S.; Korb, J.; Emanuelsson, O.; Pedersen, J. S.; Holroyd, N.; Taylor, R.; Swarbreck, D.; Matthews, N.; Dickson, M. C.; Thomas, D. J.; Weirauch, M. T.; Gilbert, J.; Drenkow, J.; Bell, I.; Zhao, X.; Srinivasan, K. G.; Sung, W. K.; Ooi, H. S.; Chiu, K. P.; Foissac, S.; Alioto, T.;

- Brent, M.; Pachter, L.; Tress, M. L.; Valencia, A.; Choo, S. W.; Choo, C. Y.; Ucla, C.; Manzano, C.; Wyss, C.; Cheung, E.; Clark, T. G.; Brown, J. B.; Ganesh, M.; Patel, S.; Tammana, H.; Chrast, J.; Henrichsen, C. N.; Kai, C.; Kawai, J.; Nagalakshmi, U.; Wu, J.; Lian, Z.; Lian, J.; Newburger, P.; Zhang, X.; Bickel, P.; Mattick, J. S.; Carninci, P.; Hayashizaki, Y.; Weissman, S.; Hubbard, T.; Myers, R. M.; Rogers, J.; Stadler, P. F.; Lowe, T. M.; Wei, C. L.; Ruan, Y.; Struhl, K.; Gerstein, M.; Antonarakis, S. E.; Fu, Y.; Green, E. D.; Karaoz, U.; Siepel, A.; Taylor, J.; Liefer, L. A.; Wetterstrand, K. A.; Good, P. J.; Feingold, E. A.; Guyer, M. S.; Cooper, G. M.; Asimenos, G.; Dewey, C. N.; Hou, M.; Nikolaev, S.; Montoya-Burgos, J. I.; Loytynoja, A.; Whelan, S.; Pardi, F.; Massingham, T.; Huang, H.; Zhang, N. R.; Holmes, I.; Mullikin, J. C.; Ureta-Vidal, A.; Paten, B.; Seringhaus, M.; Church, D.; Rosenbloom, K.; Kent, W. J.; Stone, E. A.; Batzoglu, S.; Goldman, N.; Hardison, R. C.; Haussler, D.; Miller, W.; Sidow, A.; Trinklein, N. D.; Zhang, Z. D.; Barrera, L.; Stuart, R.; King, D. C.; Ameur, A.; Enroth, S.; Bieda, M. C.; Kim, J.; Bhinge, A. A.; Jiang, N.; Liu, J.; Yao, F.; Vega, V. B.; Lee, C. W.; Ng, P.; Yang, A.; Moqtaderi, Z.; Zhu, Z.; Xu, X.; Squazzo, S.; Oberley, M. J.; Inman, D.; Singer, M. A.; Richmond, T. A.; Munn, K. J.; Rada-Iglesias, A.; Wallerman, O.; Komorowski, J.; Fowler, J. C.; Couttet, P.; Bruce, A. W.; Dovey, O. M.; Ellis, P. D.; Langford, C. F.; Nix, D. A.; Euskirchen, G.; Hartman, S.; Urban, A. E.; Kraus, P.; Van Calcar, S.; Heintzman, N.; Kim, T. H.; Wang, K.; Qu, C.; Hon, G.; Luna, R.; Glass, C. K.; Rosenfeld, M. G.; Aldred, S. F.; Cooper, S. J.; Halees, A.; Lin, J. M.; Shulha, H. P.; Xu, M.; Haidar, J. N.; Yu, Y.; Iyer, V. R.; Green, R. D.; Wadelius, C.; Farnham, P. J.; Ren, B.; Harte, R. A.; Hinrichs, A. S.; Trumbower, H.; Clawson, H.; Hillman-Jackson, J.; Zweig, A. S.; Smith, K.; Thakkapallayil, A.; Barber, G.; Kuhn, R. M.; Karolchik, D.; Armengol, L.; Bird, C. P.; de Bakker, P. I.; Kern, A. D.; Lopez-Bigas, N.; Martin, J. D.; Stranger, B. E.; Woodroffe, A.; Davydov, E.; Dimas, A.; Eyas, E.; Hallgrimsdottir, I. B.; Huppert, J.; Zody, M. C.; Abecasis, G. R.; Estivill, X.; Bouffard, G. G.; Guan, X.; Hansen, N. F.; Idol, J. R.; Maduro, V. V.; Maskeri, B.; McDowell, J. C.; Park, M.; Thomas, P. J.; Young, A. C.; Blakesley, R. W.; Muzny, D. M.; Sodergren, E.; Wheeler, D. A.; Worley, K. C.; Jiang, H.; Weinstock, G. M.; Gibbs, R. A.; Graves, T.; Fulton, R.; Mardis, E. R.; Wilson, R. K.; Clamp, M.; Cuff, J.; Gnerre, S.; Jaffe, D. B.; Chang, J. L.; Lindblad-Toh, K.; Lander, E. S.; Koriabine, M.; Nefedov, M.; Osoegawa, K.; Yoshinaga, Y.; Zhu, B. and de Jong, P. J. (2007). "Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project." *Nature* **447**(7146): 799-816.
- Brinkmann, B.; Junge, A.; Meyer, E. and Wiegand, P. (1998a). "Population genetic diversity in relation to microsatellite heterogeneity." *Hum Mutat* **11**(2): 135-44.
- Brinkmann, B.; Klintschar, M.; Neuhuber, F.; Huhne, J. and Rolf, B. (1998b). "Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat." *Am J Hum Genet* **62**(6): 1408-15.
- Brion, M. (2005). "Y chromosome SNP analysis using the single-base extension: a hierarchical multiplex design." *Methods Mol Biol* **297**: 229-42.
- Brion, M.; Quintans, B.; Zarrabeitia, M.; Gonzalez-Neira, A.; Salas, A.; Lareu, V.; Tyler-Smith, C. and Carracedo, A. (2004). "Micro-geographical differentiation in Northern Iberia revealed by Y-chromosomal DNA analysis." *Gene* **329**: 17-25.

- Brion, M.; Sánchez, J. J.; Balogh, K.; Thacker, C.; Blanco-Verea, A.; Borsting, C.; Stradmann-Bellinghausen, B.; Bogus, M.; Syndercombe-Court, D.; Schneider, P. M.; Carracedo, A. and Morling, N. (2005a). "Introduction of a single nucleotide polymorphism-based "Major Y-chromosome haplogroup typing kit" suitable for predicting the geographical origin of male lineages." Electrophoresis **26**(23): 4411-20.
- Brion, M.; Sobrino, B.; Blanco-Verea, A.; Lareu, M. V. and Carracedo, A. (2005b). "Hierarchical analysis of 30 Y-chromosome SNPs in European populations." Int J Legal Med **119**(1): 10-5.
- Budowle, B.; Adamowicz, M.; Aranda, X. G.; Barna, C.; Chakraborty, R.; Cheswick, D.; Dafoe, B.; Eisenberg, A.; Frappier, R.; Gross, A. M.; Ladd, C.; Lee, H. S.; Milne, S. C.; Meyers, C.; Prinz, M.; Richard, M. L.; Saldanha, G.; Tierney, A. A.; Viculis, L. and Krenke, B. E. (2005). "Twelve short tandem repeat loci Y chromosome haplotypes: genetic analysis on populations residing in North America." Forensic Sci Int **150**(1): 1-15.
- Budowle, B. and van Daal, A. (2008). "Forensically relevant SNP classes." Biotechniques **44**(5): 603-8, 610.
- Burton, M. L., Moore, C.C., Whiting, J.W.M., Romney, A.K. (1996). "Regions based on social structure." Curr Anthropol **37**: 87-123.
- Butler, J. M. (1997). from <http://www.cstl.nist.gov/strbase/>.
- Butler, J. M. (2001). Forensic DNA typing, Academic Press.
- Butler, J. M. (2006). "Genetics and genomics of core short tandem repeat loci used in human identity testing." J Forensic Sci **51**(2): 253-65.
- Butler, J. M.; Decker, A. E.; Kline, M. C. and Vallone, P. M. (2005). "Chromosomal duplications along the Y-chromosome and their potential impact on Y-STR interpretation." J Forensic Sci **50**(4): 853-9.
- Butler, J. M., Kline, M.C., Decker, A.E. (2008). "Addressing Y-Chromosome Short Tandem Repeat Allele Nomenclature." Journal of Genetic Genealogy **4**(2): 125-148.
- Butler, J. M., Shen, Y., McCord, B.R. (2003). "The development of reduced size STR amplicons as tools for analysis of degraded DNA." J Forensic Sci **48**(5): 1054-64.
- Cann, R. L.; Stoneking, M. and Wilson, A. C. (1987). "Mitochondrial DNA and human evolution." Nature **325**(6099): 31-6.
- Capelli, C.; Brisighelli, F.; Scarnicci, F.; Blanco-Verea, A.; Brion, M. and Pascali, V. L. (2007). "Phylogenetic evidence for multiple independent duplication events at the DYS19 locus." Forensic Sci Int Genet **1**(3-4): 287-90.
- Capelli, C.; Wilson, J. F.; Richards, M.; Stumpf, M. P.; Gratrix, F.; Oppenheimer, S.; Underhill, P.; Pascali, V. L.; Ko, T. M. and Goldstein, D. B. (2001). "A predominantly indigenous paternal heritage for the Austronesian-speaking peoples of insular Southeast Asia and Oceania." Am J Hum Genet **68**(2): 432-43.
- Cooper, G.; Burroughs, N. J.; Rand, D. A.; Rubinsztein, D. C. and Amos, W. (1999). "Markov chain Monte Carlo analysis of human Y-chromosome microsatellites provides evidence of biased mutation." Proc Natl Acad Sci U S A **96**(21): 11916-21.

- Crow, J. F. and Kimura, M. (1970). An introduction to population genetics theory. New York, Harper and Row.
- Cruciani, F.; Trombetta, B.; Macaulay, V. and Scozzari, R. (2010). "About the X-to-Y gene conversion rate." Am J Hum Genet **86**(3): 495-7; author reply 497-8.
- Chargaff, E. (1950). "Chemical specificity of nucleic acids and mechanism of their enzymatic degradation." Experientia **6**(6): 201-9.
- de Knijff, P. (2000). "Messages through bottlenecks: on the combined use of slow and fast evolving polymorphic markers on the human Y chromosome." Am J Hum Genet **67**(5): 1055-61.
- Decker, A. E.; Kline, M. C.; Redman, J. W.; Reid, T. M. and Butler, J. M. (2008). "Analysis of mutations in father-son pairs with 17 Y-STR loci." Forensic Sci Int Genet **2**(3): e31-5.
- Di Rienzo, A.; Donnelly, P.; Toomajian, C.; Sisk, B.; Hill, A.; Petzl-Erler, M. L.; Haines, G. K. and Barch, D. H. (1998). "Heterogeneity of microsatellite mutations within and between loci, and implications for human demographic histories." Genetics **148**(3): 1269-84.
- Di Rienzo, A.; Peterson, A. C.; Garza, J. C.; Valdes, A. M.; Slatkin, M. and Freimer, N. B. (1994). "Mutational processes of simple-sequence repeat loci in human populations." Proc Natl Acad Sci U S A **91**(8): 3166-70.
- Dixon, L. A.; Dobbins, A. E.; Pulker, H. K.; Butler, J. M.; Vallone, P. M.; Coble, M. D.; Parson, W.; Berger, B.; Grubwieser, P.; Mogensen, H. S.; Morling, N.; Nielsen, K.; Sanchez, J. J.; Petkovski, E.; Carracedo, A.; Sanchez-Diz, P.; Ramos-Luis, E.; Brion, M.; Irwin, J. A.; Just, R. S.; Loreille, O.; Parsons, T. J.; Syndercombe-Court, D.; Schmitter, H.; Stradmann-Bellinghausen, B.; Bender, K. and Gill, P. (2006). "Analysis of artificially degraded DNA using STRs and SNPs--results of a collaborative European (EDNAP) exercise." Forensic Sci Int **164**(1): 33-44.
- Dunham, I.; Kundaje, A.; Aldred, S. F.; Collins, P. J.; Davis, C. A.; Doyle, F.; Epstein, C. B.; Frietze, S.; Harrow, J.; Kaul, R.; Khatun, J.; Lajoie, B. R.; Landt, S. G.; Lee, B. K.; Pauli, F.; Rosenbloom, K. R.; Sabo, P.; Safi, A.; Sanyal, A.; Shores, N.; Simon, J. M.; Song, L.; Trinklein, N. D.; Altshuler, R. C.; Birney, E.; Brown, J. B.; Cheng, C.; Djebali, S.; Dong, X.; Ernst, J.; Furey, T. S.; Gerstein, M.; Giardine, B.; Greven, M.; Hardison, R. C.; Harris, R. S.; Herrero, J.; Hoffman, M. M.; Iyer, S.; Kellis, M.; Kheradpour, P.; Lassmann, T.; Li, Q.; Lin, X.; Marinov, G. K.; Merkel, A.; Mortazavi, A.; Parker, S. C.; Reddy, T. E.; Rozowsky, J.; Schlesinger, F.; Thurman, R. E.; Wang, J.; Ward, L. D.; Whitfield, T. W.; Wilder, S. P.; Wu, W.; Xi, H. S.; Yip, K. Y.; Zhuang, J.; Bernstein, B. E.; Green, E. D.; Gunter, C.; Snyder, M.; Pazin, M. J.; Lowdon, R. F.; Dillon, L. A.; Adams, L. B.; Kelly, C. J.; Zhang, J.; Wexler, J. R.; Good, P. J.; Feingold, E. A.; Crawford, G. E.; Dekker, J.; Elinitzki, L.; Farnham, P. J.; Giddings, M. C.; Gingeras, T. R.; Guigo, R.; Hubbard, T. J.; Kellis, M.; Kent, W. J.; Lieb, J. D.; Margulies, E. H.; Myers, R. M.; Stamatoyannopoulos, J. A.; Tennebaum, S. A.; Weng, Z.; White, K. P.; Wold, B.; Yu, Y.; Wrobel, J.; Risk, B. A.; Gunawardena, H. P.; Kuiper, H. C.; Maier, C. W.; Xie, L.; Chen, X.; Mikkelsen, T. S.; Gillespie, S.; Goren, A.; Ram, O.; Zhang, X.; Wang, L.; Issner, R.; Coyne, M. J.; Durham, T.; Ku, M.; Truong, T.; Eaton, M. L.; Dobin, A.; Tanzer, A.; Lagarde, J.; Lin, W.; Xue, C.; Williams, B. A.; Zaleski, C.; Roder, M.; Kokocinski, F.; Abdelhamid, R. F.; Alioto, T.; Antoshechkin, I.; Baer, M. T.; Batut, P.; Bell, I.; Bell, K.; Chakraborty, S.; Chrast, J.; Curado, J.;

Derrien, T.; Drenkow, J.; Dumais, E.; Dumais, J.; Duttagupta, R.; Fastuca, M.; Fejes-Toth, K.; Ferreira, P.; Foissac, S.; Fullwood, M. J.; Gao, H.; Gonzalez, D.; Gordon, A.; Howald, C.; Jha, S.; Johnson, R.; Kapranov, P.; King, B.; Kingswood, C.; Li, G.; Luo, O. J.; Park, E.; Preall, J. B.; Presaud, K.; Ribeca, P.; Robyr, D.; Ruan, X.; Sammeth, M.; Sandu, K. S.; Schaeffer, L.; See, L. H.; Shahab, A.; Skancke, J.; Suzuki, A. M.; Takahashi, H.; Tilgner, H.; Trout, D.; Walters, N.; Wang, H.; Hayashizaki, Y.; Reymond, A.; Antonarakis, S. E.; Hannon, G. J.; Ruan, Y.; Carninci, P.; Sloan, C. A.; Learned, K.; Malladi, V. S.; Wong, M. C.; Barber, G. P.; Cline, M. S.; Dreszer, T. R.; Heitner, S. G.; Karolchik, D.; Kirkup, V. M.; Meyer, L. R.; Long, J. C.; Maddren, M.; Raney, B. J.; Grasfeder, L. L.; Giresi, P. G.; Battenhouse, A.; Sheffield, N. C.; Showers, K. A.; London, D.; Bhinge, A. A.; Shestak, C.; Schaner, M. R.; Kim, S. K.; Zhang, Z. Z.; Mieczkowski, P. A.; Mieczkowska, J. O.; Liu, Z.; McDaniell, R. M.; Ni, Y.; Rashid, N. U.; Kim, M. J.; Adar, S.; Zhang, Z.; Wang, T.; Winter, D.; Keefe, D.; Iyer, V. R.; Sandhu, K. S.; Zheng, M.; Wang, P.; Gertz, J.; Vielmetter, J.; Partridge, E. C.; Varley, K. E.; Gasper, C.; Bansal, A.; Pepke, S.; Jain, P.; Amrhein, H.; Bowling, K. M.; Anaya, M.; Cross, M. K.; Muratet, M. A.; Newberry, K. M.; McCue, K.; Nesmith, A. S.; Fisher-Aylor, K. I.; Pusey, B.; DeSalvo, G.; Parker, S. L.; Balasubramanian, S.; Davis, N. S.; Meadows, S. K.; Eggleston, T.; Newberry, J. S.; Levy, S. E.; Absher, D. M.; Wong, W. H.; Blow, M. J.; Visel, A.; Pennachio, L. A.; Elnitski, L.; Petrykowska, H. M.; Abyzov, A.; Aken, B.; Barrell, D.; Barson, G.; Berry, A.; Bignell, A.; Boychenko, V.; Bussotti, G.; Davidson, C.; Despacio-Reyes, G.; Diekhans, M.; Ezkurdia, I.; Frankish, A.; Gilbert, J.; Gonzalez, J. M.; Griffiths, E.; Harte, R.; Hendrix, D. A.; Hunt, T.; Jungreis, I.; Kay, M.; Khurana, E.; Leng, J.; Lin, M. F.; Loveland, J.; Lu, Z.; Manthavadi, D.; Mariotti, M.; Mudge, J.; Mukherjee, G.; Notredame, C.; Pei, B.; Rodriguez, J. M.; Saunders, G.; Sboner, A.; Searle, S.; Sisu, C.; Snow, C.; Steward, C.; Tapanari, E.; Tress, M. L.; van Baren, M. J.; Washietl, S.; Wilming, L.; Zadissa, A.; Zhengdong, Z.; Brent, M.; Haussler, D.; Valencia, A.; Raymond, A.; Addleman, N.; Alexander, R. P.; Auerbach, R. K.; Bettinger, K.; Bhardwaj, N.; Boyle, A. P.; Cao, A. R.; Cayting, P.; Charos, A.; Cheng, Y.; Eastman, C.; Euskirchen, G.; Fleming, J. D.; Grubert, F.; Habegger, L.; Hariharan, M.; Harmanci, A.; Iyenger, S.; Jin, V. X.; Karczewski, K. J.; Kasowski, M.; Lacroute, P.; Lam, H.; Larnar-Vincent, N.; Lian, J.; Lindahl-Allen, M.; Min, R.; Miotto, B.; Monahan, H.; Moqtaderi, Z.; Mu, X. J.; O'Geen, H.; Ouyang, Z.; Patasil, D.; Raha, D.; Ramirez, L.; Reed, B.; Shi, M.; Slifer, T.; Witt, H.; Wu, L.; Xu, X.; Yan, K. K.; Yang, X.; Struhl, K.; Weissman, S. M.; Tenebaum, S. A.; Penalva, L. O.; Karmakar, S.; Bhanvadia, R. R.; Choudhury, A.; Domanus, M.; Ma, L.; Moran, J.; Victorsen, A.; Auer, T.; Centarin, L.; Eichenlaub, M.; Gruhl, F.; Heerman, S.; Hoeckendorf, B.; Inoue, D.; Kellner, T.; Kirchmaier, S.; Mueller, C.; Reinhardt, R.; Schertel, L.; Schneider, S.; Sinn, R.; Wittbrodt, B.; Wittbrodt, J.; Jain, G.; Balasundaram, G.; Bates, D. L.; Byron, R.; Canfield, T. K.; Diegel, M. J.; Dunn, D.; Ebersol, A. K.; Frum, T.; Garg, K.; Gist, E.; Hansen, R. S.; Boatman, L.; Haugen, E.; Humbert, R.; Johnson, A. K.; Johnson, E. M.; Kutayavin, T. M.; Lee, K.; Lotakis, D.; Maurano, M. T.; Neph, S. J.; Neri, F. V.; Nguyen, E. D.; Qu, H.; Reynolds, A. P.; Roach, V.; Rynes, E.; Sanchez, M. E.; Sandstrom, R. S.; Shafer, A. O.; Stergachis, A. B.; Thomas, S.; Vernot, B.; Vierstra, J.; Vong, S.; Weaver, M. A.; Yan, Y.; Zhang, M.; Akey, J. A.; Bender, M.; Dorschner, M. O.; Groudine, M.; MacCoss, M. J.; Navas, P.;

- Stamatoyannopoulos, G.; Stamatoyannopoulos, J. A.; Beal, K.; Brazma, A.; Flicek, P.; Johnson, N.; Lusk, M.; Luscombe, N. M.; Sobral, D.; Vaquerizas, J. M.; Batzoglu, S.; Sidow, A.; Hussami, N.; Kyriazopoulou-Panagiotopoulou, S.; Libbrecht, M. W.; Schaub, M. A.; Miller, W.; Bickel, P. J.; Banfai, B.; Boley, N. P.; Huang, H.; Li, J. J.; Noble, W. S.; Bilmes, J. A.; Buske, O. J.; Sahu, A. O.; Kharchenko, P. V.; Park, P. J.; Baker, D.; Taylor, J. and Lochovsky, L. (2012). "An integrated encyclopedia of DNA elements in the human genome." *Nature* **489**(7414): 57-74.
- Dupuy, B. M.; Stenersen, M.; Egeland, T. and Olaisen, B. (2004). "Y-chromosomal microsatellite mutation rates: differences in mutation rate between and within loci." *Hum Mutat* **23**(2): 117-24.
- Elhaik, E.; Tatarinova, T.; Chebotarev, D.; Piras, I. S.; Maria Calo, C.; De Montis, A.; Atzori, M.; Marini, M.; Tofanelli, S.; Francalacci, P.; Pagani, L.; Tyler-Smith, C.; Xue, Y.; Cucca, F.; Schurr, T. G.; Gaieski, J. B.; Melendez, C.; Vilar, M. G.; Owings, A. C.; Gomez, R.; Fujita, R.; Santos, F. R.; Comas, D.; Balanovsky, O.; Balanovska, E.; Zalloua, P.; Soodyall, H.; Pitchappan, R.; Ganeshprasad, A.; Hammer, M.; Matisoo-Smith, L. and Wells, R. S. (2014). "Geographic population structure analysis of worldwide human populations infers their biogeographical origins." *Nat Commun* **5**: 3513.
- Ellegren, H. (2000). "Microsatellite mutations in the germline: implications for evolutionary inference." *Trends Genet* **16**(12): 551-8.
- Fondevila, M. (2009). Desarrollo de paneles de SNPs autosómicos y estudio de su aplicación con fines forenses. *Departamento de anatomía patológica y ciencias forenses*, Universidade de Compostela.
- Ford, E. B. (1940). Polymorphism and taxonomy. *The New Systematics*. Oxford, Clarendon Press.
- Freire-Aradas, A.; Fondevila, M.; Kriegel, A. K.; Phillips, C.; Gill, P.; Prieto, L.; Schneider, P. M.; Carracedo, A. and Lareu, M. V. (2012). "A new SNP assay for identification of highly degraded human DNA." *Forensic Sci Int Genet* **6**(3): 341-9.
- Ge, J.; Budowle, B.; Aranda, X. G.; Planz, J. V.; Eisenberg, A. J. and Chakraborty, R. (2009). "Mutation rates at Y chromosome short tandem repeats in Texas populations." *Forensic Sci Int Genet* **3**(3): 179-84.
- Geppert, M.; Baeta, M.; Nunez, C.; Martinez-Jarreta, B.; Zweynert, S.; Cruz, O. W.; Gonzalez-Andrade, F.; Gonzalez-Solorzano, J.; Nagy, M. and Roewer, L. (2010). "Hierarchical Y-SNP assay to study the hidden diversity and phylogenetic relationship of native populations in South America." *Forensic Sci Int Genet*.
- Gerstein, M. B.; Bruce, C.; Rozowsky, J. S.; Zheng, D.; Du, J.; Korbel, J. O.; Emanuelsson, O.; Zhang, Z. D.; Weissman, S. and Snyder, M. (2007). "What is a gene, post-ENCODE? History and updated definition." *Genome Res* **17**(6): 669-81.
- 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.
- Gill, P.; Werrett, D. J.; Budowle, B. and Guerrieri, R. (2004). "An assessment of whether SNPs will replace STRs in national DNA databases--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 (SWGDM)." *Sci Justice* **44**(1): 51-3.
- Gill, P.; Whitaker, J.; Flaxman, C.; Brown, N. and Buckleton, J. (2000). "An investigation of the rigor of interpretation rules for STRs derived from less than 100 pg of DNA." *Forensic Sci Int* **112**(1): 17-40.
- Goedbloed, M.; Vermeulen, M.; Fang, R. N.; Lembring, M.; Wollstein, A.; Ballantyne, K.; Lao, O.; Brauer, S.; Kruger, C.; Roewer, L.; Lessig, R.; Ploski, R.; Dobosz, T.; Henke, L.; Henke, J.; Furtado, M. R. and Kayser, M. (2009). "Comprehensive mutation analysis of 17 Y-chromosomal short tandem repeat polymorphisms included in the AmpFISTR Yfiler PCR amplification kit." *Int J Legal Med* **123**(6): 471-82.
- Graves, J. A. (2001). "From brain determination to testis determination: evolution of the mammalian sex-determining gene." *Reprod Fertil Dev* **13**(7-8): 665-72.
- Gusmão, L., Alves, C. (2010). Marcadores dos cromossomas sexuais. Relevância na identificação genética. *Genética forense: perspectivas da identificação genética*. Pessoa, E. U. F.
- Gusmão, L.; Butler, J. M.; Carracedo, A.; Gill, P.; Kayser, M.; Mayr, W. R.; Morling, N.; Prinz, M.; Roewer, L.; Tyler-Smith, C. and Schneider, P. M. (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**(2-3): 187-97.
- Gusmao, L.; Sanchez-Diz, P.; Calafell, F.; Martin, P.; Alonso, C. A.; Alvarez-Fernandez, F.; Alves, C.; Borjas-Fajardo, L.; Bozzo, W. R.; Bravo, M. L.; Builes, J. J.; Capilla, J.; Carvalho, M.; Castillo, C.; Catanesi, C. I.; Corach, D.; Di Lonardo, A. M.; Espinheira, R.; Fagundes de Carvalho, E.; Farfan, M. J.; Figueiredo, H. P.; Gomes, I.; Lojo, M. M.; Marino, M.; Pinheiro, M. F.; Pontes, M. L.; Prieto, V.; Ramos-Luis, E.; Riancho, J. A.; Souza Goes, A. C.; Santapa, O. A.; Sumita, D. R.; Vallejo, G.; Vidal Rioja, L.; Vide, M. C.; Vieira da Silva, C. I.; Whittle, M. R.; Zabala, W.; Zarrabeitia, M. T.; Alonso, A.; Carracedo, A. and Amorim, A. (2005). "Mutation rates at Y chromosome specific microsatellites." *Hum Mutat* **26**(6): 520-8.
- Haldane, J. B. (1947). "The mutation rate of the gene for haemophilia, and its segregation ratios in males and females." *Ann Eugen* **13**(4): 262-71.
- Hallast, P.; Balaesque, P.; Bowden, G. R.; Ballereau, S. and Jobling, M. A. (2013). "Recombination dynamics of a human Y-chromosomal palindrome: rapid GC-biased gene conversion, multi-kilobase conversion tracts, and rare inversions." *PLoS Genet* **9**(7): e1003666.
- Hammer, M. F.; Karafet, T.; Rasanayagam, A.; Wood, E. T.; Altheide, T. K.; Jenkins, T.; Griffiths, R. C.; Templeton, A. R. and Zegura, S. L. (1998). "Out of Africa and back again: nested cladistic analysis of human Y chromosome variation." *Mol Biol Evol* **15**(4): 427-41.
- Hammer, M. F.; Karafet, T. M.; Redd, A. J.; Jarjanazi, H.; Santachiara-Benerecetti, S.; Soodiyall, H. and Zegura, S. L. (2001). "Hierarchical patterns of global human Y-chromosome diversity." *Mol Biol Evol* **18**(7): 1189-203.
- He, B. F.; Wu, Y. B.; Li, J.; Chen, H.; Yang, Q. E. and Huang, D. X. (2010). "Development of SNP multiplexes utilizing universal reporter primers for forensic purposes." *Fa Yi Xue Za Zhi* **26**(1): 26-9.

- Helena Mangs, A. and Morris, B. J. (2007). "The Human Pseudoautosomal Region (PAR): Origin, Function and Future." Curr Genomics **8**(2): 129-36.
- Hohoff, C.; Dewa, K.; Sibbing, U.; Hoppe, K.; Forster, P. and Brinkmann, B. (2007). "Y-chromosomal microsatellite mutation rates in a population sample from northwestern Germany." Int J Legal Med **121**(5): 359-63.
- Hughes, J. F.; Skaletsky, H.; Pyntikova, T.; Graves, T. A.; van Daalen, S. K.; Minx, P. J.; Fulton, R. S.; McGrath, S. D.; Locke, D. P.; Friedman, C.; Trask, B. J.; Mardis, E. R.; Warren, W. C.; Repping, S.; Rozen, S.; Wilson, R. K. and Page, D. C. (2010). "Chimpanzee and human Y chromosomes are remarkably divergent in structure and gene content." Nature **463**(7280): 536-9.
- Hurles, M. E.; Maund, E.; Nicholson, J.; Bosch, E.; Renfrew, C.; Sykes, B. C. and Jobling, M. A. (2003). "Native American Y chromosomes in Polynesia: the genetic impact of the Polynesian slave trade." Am J Hum Genet **72**(5): 1282-7.
- Järve, M.; Zhivotovsky, L. A.; Rootsi, S.; Help, H.; Rogaev, E. I.; Khusnutdinova, E. K.; Kivisild, T. and Sánchez, J. J. (2009). "Decreased rate of evolution in Y chromosome STR loci of increased size of the repeat unit." PLoS One **4**(9): e7276.
- Jobling, M. A., Hurles, M., Tyler-Smith, C. (2004). Human Evolutionary Genetics. New Delhi, India, Garland Publishing.
- Jobling, M. A. and Tyler-Smith, C. (1995). "Fathers and sons: the Y chromosome and human evolution." Trends Genet **11**(11): 449-56.
- Jobling, M. A. and Tyler-Smith, C. (2000). "New uses for new haplotypes the human Y chromosome, disease and selection." Trends Genet **16**(8): 356-62.
- Jobling, M. A. and Tyler-Smith, C. (2003). "The human Y chromosome: an evolutionary marker comes of age." Nat Rev Genet **4**(8): 598-612.
- Kalaydjieva, L.; Calafell, F.; Jobling, M. A.; Angelicheva, D.; de Knijff, P.; Rosser, Z. H.; Hurles, M. E.; Underhill, P.; Tournev, I.; Marushiakova, E. and Popov, V. (2001). "Patterns of inter- and intra-group genetic diversity in the Vlax Roma as revealed by Y chromosome and mitochondrial DNA lineages." Eur J Hum Genet **9**(2): 97-104.
- Karafet, T.; Xu, L.; Du, R.; Wang, W.; Feng, S.; Wells, R. S.; Redd, A. J.; Zegura, S. L. and Hammer, M. F. (2001). "Paternal population history of East Asia: sources, patterns, and microevolutionary processes." Am J Hum Genet **69**(3): 615-28.
- Karafet, T. M.; Mendez, F. L.; Meilerman, M. B.; Underhill, P. A.; Zegura, S. L. and Hammer, M. F. (2008). "New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree." Genome Res **18**(5): 830-8.
- Katti, M. V.; Ranjekar, P. K. and Gupta, V. S. (2001). "Differential distribution of simple sequence repeats in eukaryotic genome sequences." Mol Biol Evol **18**(7): 1161-7.
- 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.; Roewer, L. and et al. (1997). "Evaluation of Y-chromosomal STRs: a multicenter study." Int J Legal Med **110**(3): 125-33, 141-9.

- Kayser, M.; Roewer, L.; Hedman, M.; Henke, L.; Henke, J.; Brauer, S.; Kruger, C.; Krawczak, M.; Nagy, M.; Dobosz, T.; Szibor, R.; de Knijff, P.; Stoneking, M. and Sajantila, A. (2000). "Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son pairs." *Am J Hum Genet* **66**(5): 1580-8.
- Kimura, M. and Crow, J. F. (1964). "The Number of Alleles That Can Be Maintained in a Finite Population." *Genetics* **49**: 725-38.
- King, T. E. and Jobling, M. A. (2009). "What's in a name? Y chromosomes, surnames and the genetic genealogy revolution." *Trends Genet* **25**(8): 351-60.
- Kurihara, R.; Yamamoto, T.; Uchihi, R.; Li, S. L.; Yoshimoto, T.; Ohtaki, H.; Kamiyama, K. and Katsumata, Y. (2004). "Mutations in 14 Y-STR loci among Japanese father-son haplotypes." *Int J Legal Med* **118**(3): 125-31.
- Lahn, B. T. and Page, D. C. (1999). "Four evolutionary strata on the human X chromosome." *Science* **286**(5441): 964-7.
- Lahn, B. T.; Pearson, N. M. and Jegalian, K. (2001). "The human Y chromosome, in the light of evolution." *Nat Rev Genet* **2**(3): 207-16.
- Lange, J.; Noordam, M. J.; van Daalen, S. K.; Skaletsky, H.; Clark, B. A.; Macville, M. V.; Page, D. C. and Repping, S. (2013). "Intrachromosomal homologous recombination between inverted amplicons on opposing Y-chromosome arms." *Genomics* **102**(4): 257-64.
- Lopez-Parra, A. M.; Gusmao, L.; Tavares, L.; Baeza, C.; Amorim, A.; Mesa, M. S.; Prata, M. J. and Arroyo-Pardo, E. (2009). "In search of the pre- and post-neolithic genetic substrates in Iberia: evidence from Y-chromosome in Pyrenean populations." *Ann Hum Genet* **73**(1): 42-53.
- Manolio, T. A.; Collins, F. S.; Cox, N. J.; Goldstein, D. B.; Hindorf, L. A.; Hunter, D. J.; McCarthy, M. I.; Ramos, E. M.; Cardon, L. R.; Chakravarti, A.; Cho, J. H.; Guttacher, A. E.; Kong, A.; Kruglyak, L.; Mardis, E.; Rotimi, C. N.; Slatkin, M.; Valle, D.; Whittemore, A. S.; Boehnke, M.; Clark, A. G.; Eichler, E. E.; Gibson, G.; Haines, J. L.; Mackay, T. F.; McCarroll, S. A. and Visscher, P. M. (2009). "Finding the missing heritability of complex diseases." *Nature* **461**(7265): 747-53.
- Maughan, P. J.; Saghi Maroof, M. A. and Buss, G. R. (1995). "Microsatellite and amplified sequence length polymorphisms in cultivated and wild soybean." *Genome* **38**(4): 715-23.
- Morelli, L.; Contu, D.; Santoni, F.; Whalen, M. B.; Francalacci, P. and Cucca, F. (2010). "A comparison of Y-chromosome variation in Sardinia and Anatolia is more consistent with cultural rather than demic diffusion of agriculture." *PLoS One* **5**(4): e10419.
- Moreno, F.; Freire-Aradas, A.; Phillips, C.; Fondevila, M.; Carracedo, A. and Lareu, M. V. (2014). "SNP variation with latitude: Analysis of the SNPforID 52-plex markers in north, mid-region and south Chilean populations." *Forensic Sci Int Genet* **10**: 12-6.
- Mulero, J. J.; Chang, C. W.; Calandro, L. M.; Green, R. L.; Li, Y.; Johnson, C. L. and Hennessy, L. K. (2006). "Development and validation of the AmpFISTR Yfiler PCR amplification kit: a male specific, single amplification 17 Y-STR multiplex system." *J Forensic Sci* **51**(1): 64-75.

- Musgrave-Brown, E.; Ballard, D.; Álvarez, M. F.; Fang, R.; Harrison, C.; Phillips, C.; Prasad, Y.; Rey, B. S.; Thacker, C.; Wiluhn, J.; Carracedo, A.; Schneider, P. M. and Court, D. S. (2008). "Forensic validation of the Genplex SNP typing system- Results of an inter-laboratory study." Forensic Science International: Genetics Supplement Series **1**(1): 389-393.
- Musgrave-Brown, E.; Ballard, D.; Balogh, K.; Bender, K.; Berger, B.; Bogus, M.; Borsting, C.; Brion, M.; Fondevila, M.; Harrison, C.; Oguzturun, C.; Parson, W.; Phillips, C.; Proff, C.; Ramos-Luis, E.; Sanchez, J. J.; Sanchez Diz, P.; Sobrino Rey, B.; Stradmann-Bellinghausen, B.; Thacker, C.; Carracedo, A.; Morling, N.; Scheithauer, R.; Schneider, P. M. and Syndercombe Court, D. (2007). "Forensic validation of the SNPforID 52-plex assay." Forensic Sci Int Genet **1**(2): 186-90.
- Myres, N. M.; Rootsi, S.; Lin, A. A.; Jarve, M.; King, R. J.; Kutuev, I.; Cabrera, V. M.; Khusnutdinova, E. K.; Pshenichnov, A.; Yunusbayev, B.; Balanovsky, O.; Balanovska, E.; Rudan, P.; Baldovic, M.; Herrera, R. J.; Chiaroni, J.; Di Cristofaro, J.; Villems, R.; Kivisild, T. and Underhill, P. A. (2011). "A major Y-chromosome haplogroup R1b Holocene era founder effect in Central and Western Europe." Eur J Hum Genet **19**(1): 95-101.
- Nachman, M. W. and Crowell, S. L. (2000). "Estimate of the mutation rate per nucleotide in humans." Genetics **156**(1): 297-304.
- Ohta, T. and Kimura, M. (1973). "A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population." Genet Res **22**(2): 201-4.
- Onofri, V.; Alessandrini, F.; Turchi, C.; Pesaresi, M.; Buscemi, L. and Tagliabracci, A. (2006). "Development of multiplex PCRs for evolutionary and forensic applications of 37 human Y chromosome SNPs." Forensic Sci Int **157**(1): 23-35.
- Pakstis, A. J.; Speed, W. C.; Kidd, J. R. and Kidd, K. K. (2007). "Candidate SNPs for a universal individual identification panel." Hum Genet **121**(3-4): 305-17.
- Pascali, V. L.; Dobosz, M. and Brinkmann, B. (1999). "Coordinating Y-chromosomal STR research for the Courts." Int J Legal Med **112**(1): 1.
- Phillips, C.; Fang, R.; Ballard, D.; Fondevila, M.; Harrison, C.; Hyland, F.; Musgrave-Brown, E.; Proff, C.; Ramos-Luis, E.; Sobrino, B.; Carracedo, A.; Furtado, M. R.; Syndercombe Court, D. and Schneider, P. M. (2007a). "Evaluation of the Genplex SNP typing system and a 49plex forensic marker panel." Forensic Sci Int Genet **1**(2): 180-5.
- Phillips, C.; Fernandez-Formoso, L.; Garcia-Magarinos, M.; Porras, L.; Tvedebrink, T.; Amigo, J.; Fondevila, M.; Gomez-Tato, A.; Alvarez-Dios, J.; Freire-Aradas, A.; Gomez-Carballa, A.; Mosquera-Miguel, A.; Carracedo, A. and Lareu, M. V. (2010). "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.
- Phillips, C.; Prieto, L.; Fondevila, M.; Salas, A.; Gomez-Tato, A.; Alvarez-Dios, J.; Alonso, A.; Blanco-Verea, A.; Brion, M.; Montesino, M.; Carracedo, A. and Lareu, M. V. (2009). "Ancestry analysis in the 11-M Madrid bomb attack investigation." PLoS One **4**(8): e6583.
- Phillips, C.; Salas, A.; Sanchez, J. J.; Fondevila, M.; Gomez-Tato, A.; Alvarez-Dios, J.; Calaza, M.; de Cal, M. C.; Ballard, D.; Lareu, M. V. and Carracedo, A. (2007b). "Inferring ancestral origin using a single multiplex assay of ancestry-informative marker SNPs." Forensic Sci Int Genet **1**(3-4): 273-80.

- Prieto Solla, L. (2004). Aplicaciones forenses del ADN Publicaciones fiscales. Centro de estudios jurídicos, Ministerio de Justicia.
- Ramos-Luis, E.; Blanco-Verea, A.; Brion, M.; Van Huffel, V.; Sanchez-Diz, P. and Carracedo, A. (2014). "Y-chromosomal DNA analysis in French male lineages." Forensic Sci Int Genet **9**: 162-8.
- Redd, A. J.; Agellon, A. B.; Kearney, V. A.; Contreras, V. A.; Karafet, T.; Park, H.; de Knijff, P.; Butler, J. M. and Hammer, M. F. (2002). "Forensic value of 14 novel STRs on the human Y chromosome." Forensic Sci Int **130**(2-3): 97-111.
- Roewer, L. (2009). "Y chromosome STR typing in crime casework." Forensic Sci Med Pathol **5**(2): 77-84.
- Roewer, L.; Arnemann, J.; Spurr, N. K.; Grzeschik, K. H. and Epplen, J. T. (1992). "Simple repeat sequences on the human Y chromosome are equally polymorphic as their autosomal counterparts." Hum Genet **89**(4): 389-94.
- Roewer, L.; Croucher, P. J.; Willuweit, S.; Lu, T. T.; Kayser, M.; Lessig, R.; de Knijff, P.; Jobling, M. A.; Tyler-Smith, C. and Krawczak, M. (2005). "Signature of recent historical events in the European Y-chromosomal STR haplotype distribution." Hum Genet **116**(4): 279-91.
- Rosser, Z. H.; Balaesque, P. and Jobling, M. A. (2009). "Gene conversion between the X chromosome and the male-specific region of the Y chromosome at a translocation hotspot." Am J Hum Genet **85**(1): 130-4.
- Rozen, S.; Skaletsky, H.; Marszalek, J. D.; Minx, P. J.; Cordum, H. S.; Waterston, R. H.; Wilson, R. K. and Page, D. C. (2003). "Abundant gene conversion between arms of palindromes in human and ape Y chromosomes." Nature **423**(6942): 873-6.
- Ruitberg, C. M.; Reeder, D. J. and Butler, J. M. (2001). "STRBase: a short tandem repeat DNA database for the human identity testing community." Nucleic Acids Res **29**(1): 320-2.
- Sachidanandam, R.; Weissman, D.; Schmidt, S. C.; Kakol, J. M.; Stein, L. D.; Marth, G.; Sherry, S.; Mullikin, J. C.; Mortimore, B. J.; Willey, D. L.; Hunt, S. E.; Cole, C. G.; Coggill, P. C.; Rice, C. M.; Ning, Z.; Rogers, J.; Bentley, D. R.; Kwok, P. Y.; Mardis, E. R.; Yeh, R. T.; Schultz, B.; Cook, L.; Davenport, R.; Dante, M.; Fulton, L.; Hillier, L.; Waterston, R. H.; McPherson, J. D.; Gilman, B.; Schaffner, S.; Van Etten, W. J.; Reich, D.; Higgins, J.; Daly, M. J.; Blumenstiel, B.; Baldwin, J.; Stange-Thomann, N.; Zody, M. C.; Linton, L.; Lander, E. S. and Altshuler, D. (2001). "A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms." Nature **409**(6822): 928-33.
- Sajantila, A.; Lukka, M. and Syvanen, A. C. (1999). "Experimentally observed germline mutations at human micro- and minisatellite loci." Eur J Hum Genet **7**(2): 263-6.
- Sánchez-Diz, P.; Alves, C.; Carvalho, E.; Carvalho, M.; Espinheira, R.; Garcia, O.; Pinheiro, M. F.; Pontes, L.; Porto, M. J.; Santapa, O.; Silva, C.; Sumita, D.; Valente, S.; Whittle, M.; Yurrebaso, I.; Carracedo, A.; Amorim, A. and Gusmao, L. (2008). "Population and segregation data on 17 Y-STRs: results of a GEP-ISFG collaborative study." Int J Legal Med **122**(6): 529-33.
- Sanchez, J. J.; Borsting, C. and Morling, N. (2005). "Typing of Y chromosome SNPs with multiplex PCR methods." Methods Mol Biol **297**: 209-28.
- Sanchez, J. J.; Phillips, C.; Borsting, C.; Balogh, K.; Bogus, M.; Fondevila, M.; Harrison, C. D.; Musgrave-Brown, E.; Salas, A.; Syndercombe-Court, D.;

- Schneider, P. M.; Carracedo, A. and Morling, N. (2006). "A multiplex assay with 52 single nucleotide polymorphisms for human identification." Electrophoresis **27**(9): 1713-24.
- Santos, F. R., Gerelsaikhan, T., Munkhtuja, B., Oyunsuren, T., Epplen, J.T., Pena, S.D.J. (1996). "Geographic differences in the allele frequencies of the human Y-linked tetranucleotide polymorphism DYS19." Hum Genet **97**: 309–313.
- Schneider, P. M. (2009). "Expansion of the European Standard Set of DNA database loci-the current situation." Profiles in DNA **12**.
- Schneider, P. M. (2012). "Beyond STRs: The Role of Diallelic Markers in Forensic Genetics." Transfus Med Hemother **39**(3): 176-180.
- Schoske, R.; Vallone, P. M.; Kline, M. C.; Redman, J. W. and Butler, J. M. (2004). "High-throughput Y-STR typing of U.S. populations with 27 regions of the Y chromosome using two multiplex PCR assays." Forensic Sci Int **139**(2-3): 107-21.
- Semino, O.; Passarino, G.; Oefner, P. J.; Lin, A. A.; Arbuzova, S.; Beckman, L. E.; De Benedictis, G.; Francalacci, P.; Kouvatsi, A.; Limborska, S.; Marcikiae, M.; Mika, A.; Mika, B.; Primorac, D.; Santachiara-Benerecetti, A. S.; Cavalli-Sforza, L. L. and Underhill, P. A. (2000). "The genetic legacy of Paleolithic Homo sapiens sapiens in extant Europeans: a Y chromosome perspective." Science **290**(5494): 1155-9.
- Skaletsky, H.; Kuroda-Kawaguchi, T.; Minx, P. J.; Cordum, H. S.; Hillier, L.; Brown, L. G.; Repping, S.; Pyntikova, T.; Ali, J.; Bieri, T.; Chinwalla, A.; Delehaanty, A.; Delehaanty, K.; Du, H.; Fewell, G.; Fulton, L.; Fulton, R.; Graves, T.; Hou, S. F.; Latrielle, P.; Leonard, S.; Mardis, E.; Maupin, R.; McPherson, J.; Miner, T.; Nash, W.; Nguyen, C.; Ozersky, P.; Pepin, K.; Rock, S.; Rohlfing, T.; Scott, K.; Schultz, B.; Strong, C.; Tin-Wollam, A.; Yang, S. P.; Waterston, R. H.; Wilson, R. K.; Rozen, S. and Page, D. C. (2003). "The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes." Nature **423**(6942): 825-37.
- Sobrinho, B.; Brión, M. and Carracedo, A. (2005). "SNPs in forensic genetics: a review on SNP typing methodologies." Forensic Sci Int **154**(2-3): 181-94.
- Su, B.; Xiao, J.; Underhill, P.; Deka, R.; Zhang, W.; Akey, J.; Huang, W.; Shen, D.; Lu, D.; Luo, J.; Chu, J.; Tan, J.; Shen, P.; Davis, R.; Cavalli-Sforza, L.; Chakraborty, R.; Xiong, M.; Du, R.; Oefner, P.; Chen, Z. and Jin, L. (1999). "Y-Chromosome evidence for a northward migration of modern humans into Eastern Asia during the last Ice Age." Am J Hum Genet **65**(6): 1718-24.
- Syvänen, A. C. (2001). "Accessing genetic variation: genotyping single nucleotide polymorphisms." Nat Rev Genet **2**(12): 930-42.
- Templeton, A. R. (1997). "Out of Africa? What do genes tell us?" Curr Opin Genet Dev **7**(6): 841-7.
- The-Council-of-the-European-Union (2001). European Council: Resolution of 25 June 2001 on the exchange of DNA analysis results. UNION, T. C. O. T. E.
- The-European-SNPforID-Consortium. "High throughput analysis of single nucleotide polymorphisms for the forensic Identification of persons." from <http://www.snpforid.org/>.
- The-International-HapMap-Consortium (2003). "The International HapMap Project." Nature **426**(6968): 789-96.

- Thorisson, G. A. and Stein, L. D. (2003). "The SNP Consortium website: past, present and future." *Nucleic Acids Res* **31**(1): 124-7.
- Trombetta, B.; Cruciani, F.; Underhill, P. A.; Sellitto, D. and Scozzari, R. (2010). "Footprints of X-to-Y gene conversion in recent human evolution." *Mol Biol Evol* **27**(3): 714-25.
- Underhill, P. A.; Jin, L.; Lin, A. A.; Mehdi, S. Q.; Jenkins, T.; Vollrath, D.; Davis, R. W.; Cavalli-Sforza, L. L. and Oefner, P. J. (1997). "Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography." *Genome Res* **7**(10): 996-1005.
- Underhill, P. A. and Kivisild, T. (2007). "Use of y chromosome and mitochondrial DNA population structure in tracing human migrations." *Annu Rev Genet* **41**: 539-64.
- Underhill, P. A.; Passarino, G.; Lin, A. A.; Shen, P.; Mirazon Lahr, M.; Foley, R. A.; Oefner, P. J. and Cavalli-Sforza, L. L. (2001). "The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations." *Ann Hum Genet* **65**(Pt 1): 43-62.
- Underhill, P. A.; Shen, P.; Lin, A. A.; Jin, L.; Passarino, G.; Yang, W. H.; Kauffman, E.; Bonne-Tamir, B.; Bertranpetit, J.; Francalacci, P.; Ibrahim, M.; Jenkins, T.; Kidd, J. R.; Mehdi, S. Q.; Seielstad, M. T.; Wells, R. S.; Piazza, A.; Davis, R. W.; Feldman, M. W.; Cavalli-Sforza, L. L. and Oefner, P. J. (2000). "Y chromosome sequence variation and the history of human populations." *Nat Genet* **26**(3): 358-61.
- Vallone, P. M. and Butler, J. M. (2004). "Y-SNP typing of U.S. African American and Caucasian samples using allele-specific hybridization and primer extension." *J Forensic Sci* **49**(4): 723-32.
- van Oven, M.; van den Tempel, N. and Kayser, M. (2012). "A multiplex SNP assay for the dissection of human Y-chromosome haplogroup O representing the major paternal lineage in East and Southeast Asia." *J Hum Genet* **57**(1): 65-9.
- van Oven, M.; Van Geystelen, A.; Kayser, M.; Decorte, R. and Larmuseau, M. H. (2014). "Seeing the wood for the trees: a minimal reference phylogeny for the human Y chromosome." *Hum Mutat* **35**(2): 187-91.
- Venter, J. C.; Adams, M. D.; Myers, E. W.; Li, P. W.; Mural, R. J.; Sutton, G. G.; Smith, H. O.; Yandell, M.; Evans, C. A.; Holt, R. A.; Gocayne, J. D.; Amanatides, P.; Ballew, R. M.; Huson, D. H.; Wortman, J. R.; Zhang, Q.; Kodira, C. D.; Zheng, X. H.; Chen, L.; Skupski, M.; Subramanian, G.; Thomas, P. D.; Zhang, J.; Gabor Miklos, G. L.; Nelson, C.; Broder, S.; Clark, A. G.; Nadeau, J.; McKusick, V. A.; Zinder, N.; Levine, A. J.; Roberts, R. J.; Simon, M.; Slayman, C.; Hunkapiller, M.; Bolanos, R.; Delcher, A.; Dew, I.; Fasulo, D.; Flanigan, M.; Florea, L.; Halpern, A.; Hannenhalli, S.; Kravitz, S.; Levy, S.; Mobarry, C.; Reinert, K.; Remington, K.; Abu-Threideh, J.; Beasley, E.; Biddick, K.; Bonazzi, V.; Brandon, R.; Cargill, M.; Chandramouliswaran, I.; Charlab, R.; Chaturvedi, K.; Deng, Z.; Di Francesco, V.; Dunn, P.; Eilbeck, K.; Evangelista, C.; Gabrielian, A. E.; Gan, W.; Ge, W.; Gong, F.; Gu, Z.; Guan, P.; Heiman, T. J.; Higgins, M. E.; Ji, R. R.; Ke, Z.; Ketchum, K. A.; Lai, Z.; Lei, Y.; Li, Z.; Li, J.; Liang, Y.; Lin, X.; Lu, F.; Merkulov, G. V.; Milshina, N.; Moore, H. M.; Naik, A. K.; Narayan, V. A.; Neelam, B.; Nusskern, D.; Rusch, D. B.; Salzberg, S.; Shao, W.; Shue, B.; Sun, J.; Wang, Z.; Wang, A.; Wang, X.; Wang, J.; Wei, M.; Wides, R.; Xiao, C.; Yan, C.; Yao, A.; Ye, J.; Zhan, M.; Zhang, W.;

- Zhang, H.; Zhao, Q.; Zheng, L.; Zhong, F.; Zhong, W.; Zhu, S.; Zhao, S.; Gilbert, D.; Baumhueter, S.; Spier, G.; Carter, C.; Cravchik, A.; Woodage, T.; Ali, F.; An, H.; Awe, A.; Baldwin, D.; Baden, H.; Barnstead, M.; Barrow, I.; Beeson, K.; Busam, D.; Carver, A.; Center, A.; Cheng, M. L.; Curry, L.; Danaher, S.; Davenport, L.; Desilets, R.; Dietz, S.; Dodson, K.; Doup, L.; Ferreira, S.; Garg, N.; Gluecksmann, A.; Hart, B.; Haynes, J.; Haynes, C.; Heiner, C.; Hladun, S.; Hostin, D.; Houck, J.; Howland, T.; Ibegwam, C.; Johnson, J.; Kalush, F.; Kline, L.; Koduru, S.; Love, A.; Mann, F.; May, D.; McCawley, S.; McIntosh, T.; McMullen, I.; Moy, M.; Moy, L.; Murphy, B.; Nelson, K.; Pfannkoch, C.; Pratts, E.; Puri, V.; Qureshi, H.; Reardon, M.; Rodriguez, R.; Rogers, Y. H.; Romblad, D.; Ruhfel, B.; Scott, R.; Sitter, C.; Smallwood, M.; Stewart, E.; Strong, R.; Suh, E.; Thomas, R.; Tint, N. N.; Tse, S.; Vech, C.; Wang, G.; Wetter, J.; Williams, S.; Williams, M.; Windsor, S.; Winn-Deen, E.; Wolfe, K.; Zaveri, J.; Zaveri, K.; Abril, J. F.; Guigo, R.; Campbell, M. J.; Sjolander, K. V.; Karlak, B.; Kejariwal, A.; Mi, H.; Lazareva, B.; Hatton, T.; Narechania, A.; Diemer, K.; Muruganujan, A.; Guo, N.; Sato, S.; Bafna, V.; Istrail, S.; Lippert, R.; Schwartz, R.; Walenz, B.; Yooseph, S.; Allen, D.; Basu, A.; Baxendale, J.; Blick, L.; Caminha, M.; Carnes-Stine, J.; Caulk, P.; Chiang, Y. H.; Coyne, M.; Dahlke, C.; Mays, A.; Dombroski, M.; Donnelly, M.; Ely, D.; Esparham, S.; Fosler, C.; Gire, H.; Glanowski, S.; Glasser, K.; Glodek, A.; Gorokhov, M.; Graham, K.; Gropman, B.; Harris, M.; Heil, J.; Henderson, S.; Hoover, J.; Jennings, D.; Jordan, C.; Jordan, J.; Kasha, J.; Kagan, L.; Kraft, C.; Levitsky, A.; Lewis, M.; Liu, X.; Lopez, J.; Ma, D.; Majoros, W.; McDaniel, J.; Murphy, S.; Newman, M.; Nguyen, T.; Nguyen, N.; Nodell, M.; Pan, S.; Peck, J.; Peterson, M.; Rowe, W.; Sanders, R.; Scott, J.; Simpson, M.; Smith, T.; Sprague, A.; Stockwell, T.; Turner, R.; Venter, E.; Wang, M.; Wen, M.; Wu, D.; Wu, M.; Xia, A.; Zandieh, A. and Zhu, X. (2001). "The sequence of the human genome." *Science* **291**(5507): 1304-51.
- Vermeulen, M.; Wollstein, A.; van der Gaag, K.; Lao, O.; Xue, Y.; Wang, Q.; Roewer, L.; Knoblauch, H.; Tyler-Smith, C.; de Knijff, P. and Kayser, M. (2009). "Improving global and regional resolution of male lineage differentiation by simple single-copy Y-chromosomal short tandem repeat polymorphisms." *Forensic Sci Int Genet* **3**(4): 205-13.
- Villar, F. (2006). "Genes and Languages in Europe and South-Western Asia during the Mesolithic and Neolithic Periods." *Quaderni di semantica: rivista internazionale di semantica teorica e applicata* **27**.
- Villar, F., Prósper, B.M. (2005). *Vascos, celtas e indoeuropeos: genes y lenguas*. Salamanca, Ediciones Universidad de Salamanca.
- Villar, F., Prósper, B.M., Jordán, C., Fernández M.P. (2012). *Lenguas, genes y culturas en la prehistoria de Europa y Asia suroccidental*. Salamanca, Ediciones Universidad de Salamanca.
- Watson, J. D. and Crick, F. H. (1953). "Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid." *Nature* **171**(4356): 737-8.
- Weber, J. L. and Wong, C. (1993). "Mutation of human short tandem repeats." *Hum Mol Genet* **2**(8): 1123-8.
- Willuweit, S. and Roewer, L. (2007). "Y chromosome haplotype reference database (YHRD): update." *Forensic Sci Int Genet* **1**(2): 83-7.

- Xu, X.; Peng, M. and Fang, Z. (2000). "The direction of microsatellite mutations is dependent upon allele length." Nat Genet **24**(4): 396-9.
- Xue, Y.; Wang, Q.; Long, Q.; Ng, B. L.; Swerdlow, H.; Burton, J.; Skuce, C.; Taylor, R.; Abdellah, Z.; Zhao, Y.; MacArthur, D. G.; Quail, M. A.; Carter, N. P.; Yang, H. and Tyler-Smith, C. (2009). "Human Y chromosome base-substitution mutation rate measured by direct sequencing in a deep-rooting pedigree." Curr Biol **19**(17): 1453-7.
- Y-Chromosome-Consortium (2002). "A nomenclature system for the tree of human Y-chromosomal binary haplogroups." Genome Res **12**(2): 339-48.
- Zhivotovsky, L. A.; Underhill, P. A.; Cinnioglu, C.; Kayser, M.; Morar, B.; Kivisild, T.; Scozzari, R.; Cruciani, F.; Destro-Bisol, G.; Spedini, G.; Chambers, G. K.; Herrera, R. J.; Yong, K. K.; Gresham, D.; Tournev, I.; Feldman, M. W. and Kalaydjieva, L. (2004). "The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time." Am J Hum Genet **74**(1): 50-61.







## **8 ANEXOS**



As páxinas correspondentes ao "b cvgtkn'lw rgo gpvctkq'f q"  
cr ctvcf q"6045 foron"eliminadas da versión dixital da memoria  
de tese e poden atoparse nos seguintes enlaces:

[http://rspb.royalsocietypublishing.org/content/279/1730/  
884/suppl/DC1](http://rspb.royalsocietypublishing.org/content/279/1730/884/suppl/DC1)

[http://rspb.royalsocietypublishing.org/content/suppl/2011  
/08/18/rspb.2011.1044.DC1/rspb20111044suppl.pdf](http://rspb.royalsocietypublishing.org/content/suppl/2011/08/18/rspb.2011.1044.DC1/rspb20111044suppl.pdf)





As páxinas correspondentes ao material suplementario do apartado 4.2.4 foron eliminadas da versión dixital da memoria de tese e poden atoparse nos seguintes enlaces:

<http://mbe.oxfordjournals.org/content/29/9/2211/suppl/D>

C1

[http://mbe.oxfordjournals.org/content/suppl/2012/03/12/mss091.DC1/MSS091\\_SupFigure1.pdf](http://mbe.oxfordjournals.org/content/suppl/2012/03/12/mss091.DC1/MSS091_SupFigure1.pdf)



