



**UNIVERSIDAD DE GRANADA
FACULTAD DE MEDICINA
DEPARTAMENTO DE MEDICINA LEGAL**

TESIS DOCTORAL

***ANÁLISIS Y COMPARACIÓN DE 16 LOCI
STR EN CROMOSOMA Y DE VARIOS
GRUPOS POBLACIONALES
(APELLIDADOS COLÓN Y NO
APELLIDADOS COLÓN) EN LA CUENCA
NOROESTE MEDITERRÁNEA.***

Esther Martínez Espín

Editor: Editorial de la Universidad de Granada
Autor: Esther Martínez Espín
D.L.: En trámite
ISBN: En trámite



UNIVERSIDAD DE GRANADA
DEPARTAMENTO DE MEDICINA LEGAL

***ANÁLISIS Y COMPARACIÓN DE 16 LOCI
STR EN CROMOSOMA Y DE VARIOS
GRUPOS POBLACIONALES
(APELLIDADOS COLÓN Y NO
APELLIDADOS COLÓN) EN LA CUENCA
NOROESTE MEDITERRÁNEA.***

Memoria que presenta para optar al grado de doctor en Ciencias
Biológicas,
Esther Martínez Espín

Granada, Mayo 2008

D. JOSÉ ANTONIO LORENTE ACOSTA, PROFESOR TITULAR DE
MEDICINA LEGAL Y TOXICOLOGÍA DE LA UNIVERSIDAD DE GRANADA

CERTIFICA: Que el trabajo de investigación que se expone en la presente tesis "**ANÁLISIS Y COMPARACIÓN DE 16 LOCI STR EN CROMOSOMA Y DE VARIOS GRUPOS POBLACIONALES (APELLIDADOS COLÓN Y NO APELLIDADOS COLÓN) EN LA CUENCA NOROESTE MEDITERRÁNEA.**", ha sido realizado bajo mi dirección en el Departamento de Medicina Legal y Toxicología por la Licenciada en Biología Doña Esther Martínez Espín y corresponde fielmente a los resultados obtenidos.

Una vez redactada la presente Tesis, ha sido revisada por mí y la encuentro conforme para ser presentada y aspirar al Grado de Doctor en Biología ante el Tribunal que en su día se designe.

Y para que así conste, en cumplimiento de las disposiciones vigentes, extiendo el presente en Granada, a 03 de Mayo de dos mil ocho.

D. JUAN CARLOS ÁLVAREZ MERINO, TÉCNICO SUPERIOR EN DOCENCIA E INVESTIGACIÓN EN MEDICINA LEGAL Y TOXICOLOGÍA DE LA UNIVERSIDAD DE GRANADA.

CERTIFICA: Que el trabajo de investigación que se expone en la presente tesis "**ANÁLISIS Y COMPARACIÓN DE 16 LOCI STR EN CROMOSOMA Y DE VARIOS GRUPOS POBLACIONALES (APELLIDADOS COLÓN Y NO APELLIDADOS COLÓN) EN LA CUENCA NOROESTE MEDITERRÁNEA.**", ha sido realizado bajo mi dirección en el Departamento de Medicina Legal y Toxicología por la Licenciada en Biología Doña Esther Martínez Espín y corresponde fielmente a los resultados obtenidos.

Una vez redactada la presente Tesis, ha sido revisada por mí y la encuentro conforme para ser presentada y aspirar al Grado de Doctor en Biología ante el Tribunal que en su día se designe.

Y para que así conste, en cumplimiento de las disposiciones vigentes, extiendo el presente en Granada, a 03 de Mayo de dos mil ocho.

D. CARMEN ENTRALA BERNAL, DIRECTORA TÉCNICA DE LORGEN
G.P. PARQUE TECNOLÓGICO DE CIENCIAS DE LA SALUD DE
GRANADA

CERTIFICA: Que el trabajo de investigación que se expone en la presente tesis "**ANÁLISIS Y COMPARACIÓN DE 16 LOCI STR EN CROMOSOMA Y DE VARIOS GRUPOS POBLACIONALES (APELLIDADOS COLÓN Y NO APELLIDADOS COLÓN) EN LA CUENCA NOROESTE MEDITERRÁNEA.**", ha sido realizado bajo mi dirección en el Departamento de Medicina Legal y Toxicología por la Licenciada en Biología Doña Esther Martínez Espín y corresponde fielmente a los resultados obtenidos.

Una vez redactada la presente Tesis, ha sido revisada por mí y la encuentro conforme para ser presentada y aspirar al Grado de Doctor en Biología ante el Tribunal que en su día se designe.

Y para que así conste, en cumplimiento de las disposiciones vigentes, extiendo el presente en Granada, a 03 de Mayo de dos mil ocho.

***A mis Padres
a mi Abuelo***

**El que lleva su farol
a la espalda,
no echa delante
más que su sombra.**

Rabindranath Tagore

Agradecimientos

Mi gratitud a todas las personas que han colaborado en esta tesis, sin ellas no hubiera podido realizar este trabajo.

A Doña María Castellano Arroyo y a Don Enrique Villanueva Cañadas, catedráticos del departamento de Medicina Legal de la Universidad de Granada, por concederme la oportunidad de formarme y realizar este trabajo.

A José Antonio Lorente, Juan Carlos Álvarez y Carmen Entrala por la confianza depositada en mí para formar parte de su equipo de investigación y por haber sido mis directores de esta Tesis.

A Olga Rickard por abrimé las puertas de su departamento en Roma y Cristina Martínez por que es una gran persona y mejor profesional, por su ayuda incondicional, y, por que sin ella hubiera sido imposible terminar este trabajo.

A todos mis compañeros de laboratorio, que más que compañeros son amigos, a Encarni, Olga, Espe, Marga, Esther, Alicia, Antonio, M^a Jesús, y, de forma muy especial a Luís Javier quién se ha implicado profundamente con este trabajo y me ha ayudado mucho. A Javier "compañero de fatigas". A Marga porque a lo largo de estos años me ha demostrado que siempre está ahí, para lo bueno y para lo no tan bueno. A todos aquellos que han pasado por el laboratorio periodos más o menos largos, de forma especial a Carla y a Rhonda.

A mis padres, por estar siempre a mi lado, por su ayuda, apoyo y cariño, porque fueron los primeros que comenzaron mi formación y porque con su ayuda hicieron posible que siguiera adelante. A mis hermanos, Juan y Lourdes porque gracias a su insistencia en que hablara con José Antonio hoy estoy aquí.

A mi abuelo, por transmitirme un poquito de su sabiduría y su filosofía de vida.

A Lourdes, M^a Jesús, Pilar y Joaquín y muy especialmente a M^a Carmen por todas aquellas tardes de domingo que nos pasamos estudiando, por su paciencia y dedicación.

A Paco por estar dispuesto siempre a escucharme, enseñarme, aconsejarme, tranquilizarme, disfrutar y alegrarse conmigo, animarme en situaciones difíciles y estar siempre a mi lado.



Ayuntamiento engalanado para la
visita de Alfonso XIII (1904)
AUTOR: José Martínez Rioboó

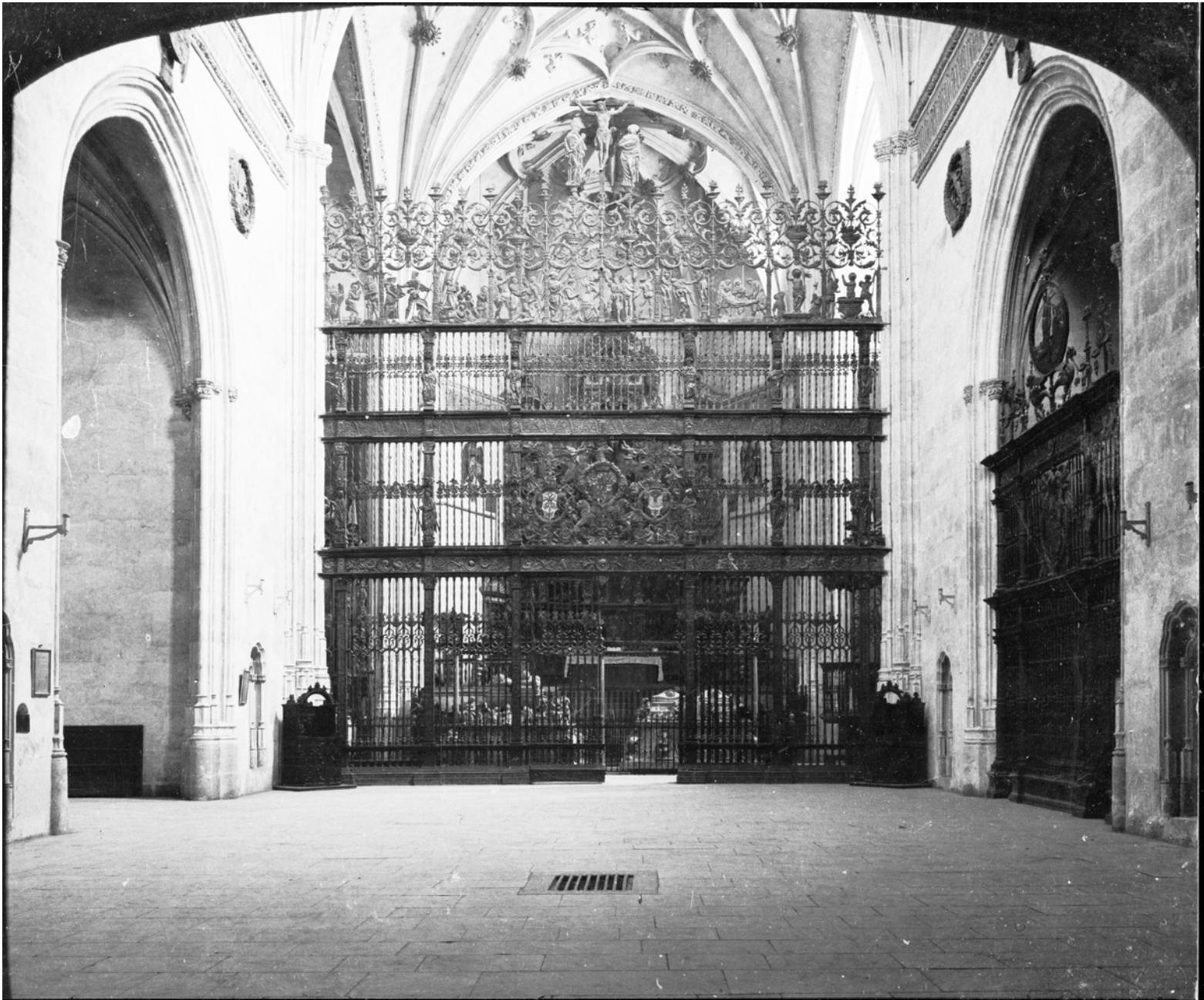
Fundación Rodríguez-Acosta
Donación Martínez Sóla.

ÍNDICE

ÍNDICE

1.- OBJETIVOS	1
2.- INTRODUCCIÓN	3
2.1.- Identificación genética en medicina legal.....	3
2.2.- El ADN	4
2.2.1.- El ADN codificante	4
2.2.2.- El ADN no codificante	5
2.2.3.- Polimorfismos del ADN	7
2.2.3.1.- ADN minisatélite.....	7
2.2.3.2.- ADN microsatélite	9
2.2.3.3.- Los SNPs	14
2.3.- Aplicación a la identificación genética	15
2.4.- Métodos de estudio de los polimorfismos del ADN... 17	
2.4.1.- Reacción en cadena de la polimerasa.....	19
2.4.1.1.- Procedimiento analítico.....	21
2.4.1.2.- Ventajas y limitaciones prácticas	25
2.4.2.- Sistemas de análisis de los polimorfismos	27
2.4.3.- Marcaje fluorescente	28
2.5.- Introducción al cromosoma Y	33
2.5.1.- Su naturaleza	34
2.5.2.- Las regiones pseudoautosómicas	35
2.5.3.- La región específica del cromosoma Y.....	37
2.5.4.- Algunas peculiaridades del cromosoma Y; analogía con el ADN mitocondrial.....	38
2.6.- Haplogrupos	41
2.7.- Introducción a la antropología	43
2.8.- Bases de datos genéticos.....	48
3.- MATERIAL Y MÉTODOS	53
3.1.- Muestras	53
3.1.1.- Población analizada.....	53
3.2.- Extracción del ADN	54
3.2.1.- Extracción orgánica mediante fenol-cloroformo- Isoamilalcohol.....	55
3.2.2.- Extracción del ADN con chelex.....	55
3.3.- Cuantificación del ADN	56
3.4.- Amplificación.....	57
3.4.1.- Kit Y-Filer	58
3.4.2.- Comprobación de la amplificación	59
3.5.- Electroforesis capilar	59
3.6.- Análisis estadístico	64

3.6.1.-	Indices de diversidad	65
3.6.1.1.-	Métodos intrapoblacionales	66
3.6.1.2.-	Métodos intrapoblacionales	67
3.6.1.3.-	Otros parámetros estadísticos	73
4.-	RESULTADOS.....	75
4.1.-	Resultados del estudio poblacional	75
4.1.1.-	Visualización de los resultados obtenidos	75
4.1.2.-	Resultados obtenidos con el kit Y-Filer	77
4.1.3.-	Análisis haplotipos de individuos Colón	148
4.1.4.-	Análisis de los haplogrupos Colón	158
5.-	DISCUSIÓN	170
5.1.-	Consideraciones metodológicas	170
5.1.1.-	Sobre el método del ADN.....	170
5.1.2.-	Sobre el método de cuantificación.....	171
5.1.3.-	Sistema electroforético.....	171
5.2.-	Estudio genético-poblacional	172
5.2.1.-	Caracterización de frecuencias alélicas.....	172
5.2.2.-	Caracterización de los haplotipos-haplogrupos	183
5.2.3.-	Caracterización de nuestras poblaciones con otras	203
5.3.-	Eficacia de los sistemas y su utilidad médico-legal ..	204
6.-	CONCLUSIONES.....	206
7.-	BIBLIOGRAFÍA	208



Reja de la capilla Real
AUTOR: José Martínez Rioboó

Fundación Rodríguez-Acosta
Donación Martínez Sóla.

1.-OBJETIVOS Y JUSTIFICACIÓN

1.-JUSTIFICACIÓN DEL TRABAJO:

1.1 ANTECEDENTES:

El almirante Cristóbal Colón, marinero, explorador, geógrafo y descubridor de América, ha sido uno de los personajes más estudiados de la historia no solo española, sino universal.

El nacimiento de Cristóbal Colón así como el lugar de descanso de sus restos tras fallecer han sido un misterio hasta nuestros días. Son muchas las teorías y los documentos de la época estudiados por miles de historiadores y biógrafos para descubrir estos enigmas.

La teoría más aceptada hasta el momento en cuanto a la fecha y lugar de nacimiento es la siguiente: fue en 1451, en Génova, Italia, hijo de padres italianos, llamados Domenico Colombo, un modesto comerciante de lanas y Susana Fontanarossa, pero desde hace más de 20 años se está investigando el auténtico lugar de nacimiento. Según la tesis mallorquina, Colón habría nacido en una cueva de Felanitx, hijo bastardo del Príncipe de Viana, navarro, y de una judía mallorquina, Margarita Colom.

Otras tesis afirman que era gallego o portugués. O que intentara ocultar su origen al ser judío converso.

En la década de 1920 el historiador peruano Luís de Ulloa, afirmaba que Colón era realmente Joan Colom, un noble catalán. Otros estudios más actuales también afirman que pertenecía a la nobleza barcelonesa.

La tesis alcarreña, castellana, nos presenta a la madre como una noble manchega, Aldonza Mendoza, con un hijo que habría sido escondido desde su nacimiento para evitar que fuera asesinado (cuestiones de herencia) y que habría sido educado por unos criados genoveses llamados Colombo.

Incluso algunas teorías delirantes: Nacido en América, descendiente de los templarios expulsados del continente europeo a principios del siglo XIV, que habrían recalado en el Caribe para desde allí realizar un viaje de ida y vuelta.

Tras la muerte de Colón, 20 de mayo de 1506. Su cuerpo fue enterrado en Valladolid, posteriormente, tras un tiempo, sus restos mortales fueron enterrados en el monasterio de la Cartuja de Sevilla. En 1795 los Franceses entraron a Sevilla, cuando Napoleón tomó España, los restos del almirante fueron trasladados a América y se notifica que estos restos fueron enterrados en Cuba. En 1898 y a causa de la guerra con Cuba, los restos fueron trasladados de nuevo a Sevilla y custodiados dentro de la catedral de Sevilla.

En 1877 periodo en el que se supone que los huesos de Colón estaban en Cuba, aparece en Santo Domingo una urna con una inscripción que asegura que su contenido son los restos mortales del Almirante.

Tras estas dos posibilidades son muchos los historiadores que se han preocupado en desvelar cuál de las urnas es la que realmente encierra los restos del descubridor.



Interior de la capilla Real
AUTOR: José Martínez Rioboó

Fundación Rodríguez-Acosta
Donación Martínez Sóla.

2.-INTRODUCCIÓN

2.- INTRODUCCIÓN.

2.1.- IDENTIFICACIÓN GENÉTICA EN MEDICINA LEGAL.

A lo largo de la historia, siempre ha existido un gran interés por conocer la identidad de los individuos. Identificar a una persona, establecer su individualidad, es determinar aquellos rasgos o conjunto de cualidades que la distinguen de todos los demás y que hacen que sea ella misma (Gisbert Calabuig, 1983).

La identificación humana en Medicina Legal siempre es un proceso de comparación, ya se trate de un dato odontoestomatológico, antropométrico o genético. En todas las circunstancias se compara el resultado obtenido en el estudio o análisis con otros cuyo origen es conocido. En caso de que sean iguales servirán para confirmar que se trata de la misma persona, y en caso de que no coincidan permitirá descartar que se trate del mismo individuo. El proceso no es directo, sino que se llega a la identificación por un camino indirecto a través de la individualización, es decir, conforme se van consiguiendo características de la persona, el número de individuos que los comparten es menor, y va reduciéndose de forma progresiva con la ampliación del estudio.

Las pruebas para llegar al diagnóstico de identificación genética se realizan mediante el estudio de caracteres hereditarios. Para que estos caracteres sean útiles en los diagnósticos de identificación deben cumplir ciertos requisitos:

- ✚ Ser altamente variables, esto es necesario ya que si todos los individuos presentaran los mismos alelos sería imposible diferenciarlos entre sí.
- ✚ El modelo de herencia debe ser conocido y estar firmemente determinado. Esta característica es importante en los diagnósticos biológicos de la paternidad, en los que se analizan los caracteres que el hijo ha heredado de su padre biológico y se establece si son compatibles con las del presunto padre.
- ✚ Inmutables a lo largo de la vida del individuo, con modelos idénticos en los distintos tejidos del organismo.

El estudio de los polimorfismos de ADN, revolucionó las técnicas de identificación genética a principios de la década de los 80. Analizando un número suficiente de regiones de ADN es posible demostrar que la coincidencia por azar entre dos personas es extraordinariamente pequeña. Analizando un número suficiente de regiones de ADN que muestren gran variabilidad entre personas, se puede reducir la probabilidad de coincidencia a niveles extremadamente bajos. De esta manera, la probabilidad puede, en principio, llegar a ser tan baja que el tipado de ADN se convierte, no sólo en un método de exclusión o inclusión, sino un método de absoluta identificación.

2.2.- EL ADN.

La unidad fundamental de la vida es la célula, y cada una de las células que conforman cualquier ser vivo tiene núcleo, que va a tener en su interior el ADN necesario para desarrollarse y cumplir la misión encomendada por la naturaleza. Existen otras células, las menos, que no tienen núcleo, y por tanto, no poseen ADN, siendo las más conocidas y abundantes los glóbulos rojos sanguíneos.

El ADN fue aislado por Frederich Miescher en Alemania, en el pus de las vendas en 1869, pero no fue hasta 1941 que Aver y McLead demostraron que era la molécula portadora de la información genética. Posteriormente, en 1953, Watson y Crick, en Inglaterra, descubrieron, en base a la información de otros científicos la estructura de la molécula de ADN.

Según su localización y estructura nos encontramos con dos tipos de ADN en el organismo, el ADN nuclear, situado en el interior del núcleo de las células, y el ADN mitocondrial, situado, como su propio nombre indica, en el interior de las mitocondrias. Ambos tienen interés medico-forense y poseen características que los hacen complementarios a la hora de la identificación e investigación de relaciones familiares.

2.2.1.- ADN CODIFICANTE.

Los genes que contienen este ADN tienen la propiedad de codificar la fabricación de proteínas que actúan a nivel celular y que se expresan en la persona como un carácter individual y puede ser normal o patológico. Por ADN codificante o expresivo se entiende aquellos fragmentos de ácido

nucleico que determinan a los diferentes genes que definirán las características de las personas a través de la síntesis proteica, determinando la secuencia de los aminoácidos de las proteínas que codifican, y el grado de expresión del gen en cada tejido y en cada tiempo.

No todo el ADN que conforma el genoma es codificante. Hay que considerar, por otra parte, que los conocimientos actuales de la ciencia no permiten asignar a cada fragmento de ADN una función.

El ADN codificante, a pesar de ser el más interesante desde el punto de vista médico, posee una escasa variabilidad entre personas, con la excepción de ciertas regiones, como la que informa para el sistema HLA.

2.2.2.- ADN NO CODIFICANTE.

En cada célula eucariota parece haber un gran exceso de ADN, o cuando menos de ADN cuyas funciones son desconocidas. Se estima que en las células alrededor del 2% de todo el ADN codifica proteínas, mientras que el resto no tiene una función claramente codificante y desempeñan funciones reguladoras, estructurales y, en gran medida, su función es desconocida.

De entre los 100.000 y 150.000 genes que los expertos calculaban que poseíamos, hemos quedado reducidos a una humilde cifra de apenas 30.000. Por tanto, el 95% del genoma no es codificante.

En efecto, del 10-25% del total del ADN humano y de otros eucariotas superiores está formado por secuencias cortas de cinco a diez pares de bases, que se repiten en tándem miles de veces. En los cromosomas, estas repeticiones cortas se localizan en los centrómeros, lugar donde los dos cromosomas hermanos continúan unidos tras la replicación, hasta que se separen al dividirse la célula. Aparentemente, esta área del ADN tiene un papel estructural, más que genético, en la célula y, de algún modo, opera durante la replicación y separación. También se encuentran repeticiones cortas en los extremos de los cromosomas, que son conocidos como telómeros. Otra gran parte del genoma está compuesta por repeticiones de secuencias más largas, repartidas por todo el genoma, cuya función permanece aún desconocida.

Por último, otros componentes de este ADN no codificante son los "pseudogene", que son secuencias que fueron funcionales como genes en estadios evolutivos anteriores, pero que han perdido su capacidad codificadora debido a diversos tipos de mutaciones. Estos pseudogenes pueden ser importantísimos a la hora de llevar a cabo una investigación de la historia evolutiva del hombre.

En general, el ADN no codificante puede ser de dos tipos:

- ✚ ADN de copia sencilla o no repetitivo que actúa como ADN espaciador entre regiones codificantes del genoma, entre las que se encuentran los intrones, que no contienen información útil para la síntesis de la cadena polipeptídica y son eliminados durante el proceso de "splicing".
- ✚ ADN de copia múltiple por lo que se conoce como ADN repetitivo (Singer, 1982 a; Kao, 1985).

Las repeticiones en tándem de una secuencia sencilla se denomina "ADN estable" debido a que las primeras secuencias de este tipo que fueron descubiertas tenían una proporción poco habitual de nucleótidos que hacía posible su separación del grueso de ADN de la célula como componente minoritario o satélite, coincidiendo, por otra parte, con los niveles que permanecían separados del grueso del ADN cuando se centrifugaban en gradientes de CsCl, como si de un satélite se tratara.

El hecho de que estas secuencias de ADN carezcan de cualquier función conocida y la comprobación de que son capaces de propagarse y extenderse por ellas mismas a través del genoma sin objetivo ni beneficio aparente para él, hace que este tipo de ADN sea conocido como "ADN egoista". Además la capacidad para moverse ha hecho que en el transcurso de la evolución las secuencias satélite se haya modificado extraordinariamente, cambiando, incluso, sus posiciones en los cromosomas.

2.2.3.- POLIMORFISMO DEL ADN.

Aproximadamente la mitad del ADN no codificante es ADN repetitivo y, aunque, gran parte del mismo es extremadamente polimórfico, por diversos motivos, el ADN utilizado con fines forenses es el ADN repetido en tándem y, dentro de él, el ADN minisatélite y microsatélite.

El ADN minisatélite y microsatélite consiste en repeticiones de fragmentos de ADN de número variable, por lo que genéricamente se denomina VNTR (variable number of tandem repeats). Las repeticiones en el ADN microsatélite son de tamaño pequeño (de 1 a 6 pares de bases), por lo que se suelen denominar STRs (short tandem repeats). Las repeticiones de un locus minisatélite tienen un tamaño medio de unos 28 pares de bases.

Los individuos nos diferenciamos por el número de repeticiones de esa secuencia. Un individuo 8-12 para un STR significa que tiene 8 veces una unidad de repetición en un lugar específico de un cromosoma (locus génico) y 12 veces en el mismo lugar de su otro cromosoma.

Los minisatélites y microsatélites además de ser extraordinariamente polimórficos (variables entre los individuos), poseen una herencia mendeliana simple. Esto significa que el individuo 8-12 ha heredado uno de los alelos de su madre y el otro de su padre biológico.

Además de los STRs en cromosomas autosómicos, son de gran importancia los STRs del cromosoma Y, particularmente para el caso de las agresiones sexuales, especialmente si son cometidas por individuos azoospermicos o que tengan muy pequeña cantidad de células masculinas en una gran cantidad de células femeninas.

2.2.3.1.- ADN MINISATÉLITE.

El término ADN minisatélite hace referencia a una clase particular de secuencia de ADN de 10-60 pares de bases repetidas en tándem, con un núcleo o unidad "core" central, denominada también VNTRs. El término minisatélite proviene por extensión del de ADN satélite, una parte considerable del genoma eucariota que se puede separar del resto por configuración en gradiente de densidad.

El primer locus altamente polimórfico de esta clase, encontrado en humanos, fue detectado por azar utilizando una sonda de ADN arbitraria (Wyman y White, 1980). Pero la base molecular de estos polimorfismos sólo fue establecida después de la secuenciación de varios alelos de la región hipervariable 5' del gen de la insulina del cromosoma 11 (Bel *et al.*, 1982; Ulrich *et al.*, 1982). Posteriormente, el análisis de otros loci hipervariables ha confirmado la existencia y estructura de este tipo de secuencias de ADN.

Después del descubrimiento de los primeros VNTRs, se vio que éstos podían ser una fuente importante de marcadores en medicina forense y que podían sustituir a los marcadores convencionales (Jeffreys *et al.*, 1985b). Posteriormente se demostró que se podía obtener un patrón de VNTRs o "fingerprint" estable y reproducible en manchas secas de sangre de más de 4 años de antigüedad, de manchas de semen, de raíces de cabello y de manchas de sangre expuestas a diferentes condiciones ambientales (Gill *et al.*, 1985; Kanter *et al.*, 1986; Giusti *et al.*, 1986; McNally *et al.*, 1989a; Adams *et al.*, 1991). Un avance adicional en la diferenciación genética de individuos sospechosos de delitos de agresión sexual, fue desarrollado de los métodos de lisis diferencial para separar los núcleos espermáticos de las células vaginales de muestras obtenidas mediante hisopos vaginales (Gill *et al.*, 1985; Giusti *et al.*, 1986).

Estas secuencias de ADN minisatélite pueden analizarse mediante el uso de enzimas de restricción cuya diana no se encuentra dentro de las unidades de repetición. El tamaño de los fragmentos de restricción es función del número de copias presentes en los loci estudiados, por lo tanto, un marcador basado en este tipo de polimorfismos, podría ser altamente informativo en contraste con los polimorfismos RFLPs que al producirse por un cambio en un par de bases, son bialélicos.

Dependiendo de las condiciones de la hibridación y de lavado, se pueden obtener diferentes resultados. Bajo condiciones de hibridación de baja astringencia, las sondas desarrolladas por Jeffreys *et al.* (1985 a) detectan los loci que contienen repeticiones en tándem de secuencias similares. Estas sondas fueron denominadas "multi-locus" o MLPs y el patrón de múltiples bandas obtenidas se denominó huella genética multilocus o "DNA-fingerprint". Sin embargo, bajo condiciones de hibridación de alta astringencia y mediante sondas complementarias a un solo locus se pueden detectar regiones minisatélites en un locus único (Nakamura *et al.*, 1987; Wong *et al.*, 1987). Estas sondas fueron

denominadas "single-locus" o SLPs y patrón de bandas obtenidas se denominó perfil unilocus de ADN.

Sin embargo, desde el descubrimiento de la PCR, el análisis de los VNTRs se ha simplificado enormemente al ofrecer una serie de ventajas que no ofrecía el análisis por RFLP (Lorente y Lorente, 1995), ya que consume menos tiempo, como se analizan fragmentos pequeños permite amplificar material antiguo y parcialmente degradado, la cantidad de ADN que se requiere es menor, con lo que se posibilita el análisis de indicios biológicos mínimos (un pelo, una mancha pequeña de sangre o semen, partículas de caspa, etc,...) y finalmente posibilita amplificar conjuntamente varios marcadores (reacciones tipo multiplex).

2.2.3.2.- ADN MICROSATELITE.

Los loci microsatélite se pueden definir como pequeñas secuencias de ADN (menos de 300 pb) dispersas por el genoma, compuestas de monómeros repetidos en tándem de 2-5 pb de longitud, por lo que también reciben el nombre de STRs.

La presencia de STRs particulares en el genoma es tan alta que el screening, por hibridación con un conjunto limitado de oligonucleótidos STR, permitiría identificar un locus STR en la mayoría de los segmentos genómicos, es decir, cada 6-10 kb según la estima de Richards y Sutherland (1992). Si se restringen los loci STRs a los triméricos y tetraméricos, puede estimarse que se distribuyen cada 15 kb, existiendo por tanto aproximadamente 200.000 loci de estas características en el genoma humano (Edwards *et al.*, 1991; Beckmann y Weber, 1992).

A diferencia de los loci VNTRs, que se sitúan preferentemente en las regiones teloméricas, los STRs parecen ser abundantes por todo el genoma y han sido descritos tanto en regiones génicas como extragénicas. Los STRs situados en regiones génicas están presentes no sólo en intrones y secuencias flanqueantes, sino también dentro de regiones codificantes. Los loci STR están asociados frecuentemente con el extremo 3' de las inserciones Alu, resultando, probablemente, de la degeneración de la secuencia de la cola poli-A (Economou *et al.*, 1990; Beckmann y Weber, 1992).

Desde su descubrimiento han sido los responsables de un gran avance en el campo de la identificación humana, en el desarrollo de los

mapas de ligamiento humano y en nuevos procedimientos de diagnóstico clínico. Así, numerosos loci microsátélites han sido usados para el diagnóstico predictivo de muchas enfermedades genéticas, incluyendo distrofia miotónica, fibrosis quística, distrofia muscular de Duchenne/Becker y enfermedad de Huntington (Mully *et al.*, 1991; Oudet *et al.*, 1991; Morral y Estivill, 1992; Weber *et al.*, 1992).

Al igual que los VNTRs minisatélite, los loci STRs presentan un alto grado de polimorfismo de longitud debido a la variación en el número de unidades de repetición (Litt y Luty, 1989; Weber y May, 1989). Además de la variación en el número de repeticiones, las secuencias STRs varían también en la longitud de la unidad de repetición y el rigor con el cual llegan a conformar un modelo de repeticiones. Las unidades de repetición "simples" contienen unidades de idéntica longitud y secuencia; las repeticiones "complejas" pueden contener varios bloques de repetición de unidades de longitud variable, junto con mayor o menor variación de las secuencias. En consecuencia, los loci STR parecen distribuirse entre dos grupos principales: aquéllos que muestran un pequeño número de alelos bien diferenciados (< 12) y los que poseen un elevado número (>35) pero cuya diferenciación entre alelos es compleja.

Los STRs del segundo grupo son significativamente más polimórficos, sin embargo, estos loci parecen ser mucho más complejos y poseen múltiples alelos que difieren en un número incompleto de repeticiones, además de tener repeticiones incompletas intercaladas entre repeticiones normales, presentando microheterogeneidad extrema. Estas características no excluyen a los loci complejos de su inclusión como marcadores para la identificación genética, pero las diferencias entre alelos, en ocasiones de una sola base, requieren un análisis muy preciso.

Los STRs del primer grupo presentan alelos cuya diferencia en tamaño es una unidad de repetición, pero existen algunas excepciones. Así, en el locus HUMF13A01 (unidad de repetición pb) el alelo más pequeño es sólo 2 pb menor que el alelo de mayor tamaño inmediato (Urquhart *et al.*, 1994). En el locus HUMTH01 recientemente se han detectado deleciones en la región de repetición de los alelos 8, 9 y 10. Estos alelos son denominados 8.3, 9.3 y 10.3 y consisten en 8, 9 y 10 repeticiones AATG respectivamente más un solo trinucleótido ATG. Considerando las altas frecuencias del alelo 9.3, las otras dos variantes alélicas parecen haber sido generadas a partir de este alelo, por la deleción de una unidad de repetición (8.3) o por la inserción de una

unidad de repetición (10.3). Mientras que el alelo 9.3 ha sido observado en todas las poblaciones estudiadas excepto en los aborígenes australianos, los otros alelos 8.3 y 10.3 únicamente han sido descritos en una población alemana (Brinkmann *et al.*, 1996).

Dentro del grupo de STRs simples existe un interés preferencial por los sistemas tri-tetra-pentanucleotídicos, frente a los STRs dinucleotídicos debido a que estos presentan bandas sombra o tartamudas, como artefactos de amplificación (Litt y Luty, 1989). Así, destacan HUMTH01 (Polymeropoulos *et al.*, 1991a), HUMVWFA31 (Kimpton *et al.*, 1992), HUMFESFPS (Polymeropoulos *et al.*, 1991b), HUMF13A01 (Polymeropoulos *et al.*, 1991c), HUMFXIIIIB (Nishimura y Murray, 1992), HUMTPOX (Anker *et al.*, 1992), HUMCSF1PO (Hammond *et al.*, 1994) y HUMLIPOL (Zulani y Hobbs, 1990), entre otros.

Los loci STRs pueden ser amplificados usando la reacción en cadena de la polimerasa (PCR) y primers complementarios a las regiones de secuencia única que flanquean a las unidades de repetición. La electroforesis en gel de poliacrilamida (Sambrook *et al.*, 1989) empleada en combinación con tinción plata (Bassam *et al.*, 1991) permite el análisis preciso de los fragmentos STRs una vez que han sido amplificados, aunque actualmente se usa la electroforesis capilar.

La asignación de los tamaños de los alelos, se realiza mediante ladders alélicos que consisten en la mezcla de secuencias amplificadas de varios alelos que se emplean como marcadores para cada loci STRs. La comparación de las muestras con el ladder alélico permite una determinación visual rápida de los tamaños de los alelos sin necesidad de realizar cálculos o análisis adicionales. Por otro lado la utilización de ladders alélicos hace posible la identificación incluso de variantes alélicas que difieren en una sola base, bien sea en la región de repetición o en la secuencia flanqueante.

A la hora de seleccionar un STR para uso forense se requiere que éste cumpla una serie de características (Budowle, 1997) como son:

- ✚ Alta heterocigosidad.
- ✚ Una tasa mutación baja.
- ✚ Que no originen artefactos durante la amplificación como bandas sombra.
- ✚ Fácilmente amplificable.

- ✚ De pequeño tamaño.
- ✚ Fácilmente coamplificable junto a otros STRs.

La utilización de ADN microsatélite presenta grandes ventajas frente a los sistemas anteriormente mencionados. La amplificación de la técnica de PCR, permite obtener varios millones de copias de un fragmento determinado de ADN en un tiempo de 2 o 3 horas permitiendo analizar muestras cuya cantidad de ADN es escasa. Por otra parte el tamaño de los productos amplificados es pequeño (100-500 pb) de forma que es posible obtener resultados a partir de ADN sumamente degradado cuyos fragmentos son aproximadamente 1000 pb.

Otra ventaja de los sistemas STRs es el corto tiempo necesario para completar los análisis debido a la técnica de PCR empleada, tiempo que se ha reducido aún más con la utilización de tecnología automatizada basada en fluorescencia (Schwartz *et al.*, 1992; Ziegle *et al.*, 1992; Suilivan *et al.*, 1992; Kimpton *et al.*, 1993). Este sistema incorpora la medición de tamaños de los productos de PCR y elimina las diferencias en la movilidad electroforética entre las calles del gel ya que incorpora marcadores de tamaño internos con cada muestra. Además el pequeño tamaño de las unidades amplificadas permite el estudio simultáneo de varios loci mediante amplificación PCR multiplex y análisis con identificación precisa de los alelos en geles de secuenciación de ADN. Y si combinamos la técnica de multiplex STR-PCR y la detección fluorescente automatizada, el resultado es una técnica sumamente rápida y poderosa para la obtención de perfiles de ADN (Kimpton *et al.*, 1995). Actualmente se consiguen analizar un gran número de STRs a partir de la misma muestra biológica, utilizando los sistemas automáticos y PCR multiplex, con un poder de discriminación enorme.

Los STRs analizados en este trabajo mediante PCR, con el kit Y-Filer , han sido DYS 456, DYS 389I, DYS 390, DYS 389II, DYS 458, DYS 19, DYS 385, DYS 393, DYS 391, DYS 439, DYS 635, DYS 392, Y-GATA H4, DYS 437, DYS 438 y DYS 448. Estos 16 loci se encuentran situados en el cromosoma Y, por lo que la muestra de población escogida es exclusivamente masculina.

STR	Fluorocromo	Color	Secuencia de repetición	Nº. unidades de repetición
DYS 456	6-FAM	Azul	AGAT	13-18
DYS 389I	6-FAM	Azul	[TCTG] _q [TCTA] _r	10-15
DYS 390	6-FAM	Azul	[TCTG] _n [TCTA] _m [TCTG] _p [TCTA] _q	18-27
DYS 389II	6-FAM	Azul	[TCTG] _n [TCTA] _p [TCTG] _q [TCTA] _r	24-34
DYS 458	VIC	Verde	CTTT	14-20
DYS 19	VIC	Verde	[TAGA] ₃ tagg[TAGA] _n	10-19
DYS 385	VIC	Verde	[GAAA] _n	7-25
DYS 393	NED	Amarillo	[AGAT] _n	8-16
DYS 391	NED	Amarillo	[TCTA] _n	7-13
DYS 439	NED	Amarillo	[GATA] _n	8-15
DYS 635	NED	Amarillo	TCTA/TGTA	20-26
DYS 392	NED	Amarillo	[TAT] _n	7-18
Y-GATA H4	PET	Rojo	[TAGA] _n ATGGATAGATTA[GATG] _p AA[TAGA] _q	8-13
DYS 437	PET	Rojo	[TCTA] _n [TCTG] ₂ [TCTA] ₄	13-17
DYS 438	PET	Rojo	[TTTTTC] _n	8-13
DYS 448	PET	Rojo	[AGAGAT] _n AGAGAGATAG[AGAGAT] _p N ₁₄ [AGAGAT] _q	17-24

Una vez los STRs son amplificados se procede al tipaje o asignación de los alelos por electroforesis capilar y asignación de alelos automatizada a través de un software.

Las principales ventajas que ofrece el análisis de los STRs frente a los VNTRs son:

- ✚ Al ser fragmentos de pequeño tamaño (menores de 400 pb) son menos susceptibles a la degradación, de manera que es posible obtener resultados a partir de ADN sumamente degradado.
- ✚ Poseen alelos discretos, lo cual permite simplificar el análisis estadístico de las frecuencias alélicas en las poblaciones por un lado, y por otro facilitar la interpretación de los resultados al migrar cada alelo a una longitud determinada.

- ✚ El tiempo consumido se reduce considerablemente si comparamos con los VNTRs y más aun si consideramos los últimos avances en electroforesis capilar.

2.2.3.3.- LOS SNPs.

Los SNPs (single nucleotide polymorphisms) representan la clase más abundante de polimorfismos humanos. Actualmente, resulta muy difícil dar una estimación del número de SNPs existentes en el genoma humano, pero en algunas publicaciones se ha hablado de unos cinco millones de SNPs y han sido validados unos cuatro millones.

El interés de este tipo de polimorfismo en el campo forense aumenta continuamente, debido a que tienen un serie de características que los hacen muy apropiados para estos estudios:

- ✚ Tiene una tasa de mutación baja, la cual, se estima en 10^{-8} (Reich *et al.*, 2002), mientras que la tasa de mutación de los STRs es de 10^{-3} a 10^{-5} (Huang *et al.*, 2002; Dupuy *et al.*, 2004), lo cual es importante para pruebas de paternidad.
- ✚ Son susceptibles de análisis por medio de alta tecnología y el tamaño del producto amplificado es pequeño (algunos pueden ser de 45-55 pb) lo que implica que puede ser el único medio útil para el análisis de muestras muy degradadas. Por ello, también tienen un papel muy importante en identificación humana, sobretodo cuando los restos no se encuentran en muy buenas condiciones (Sánchez *et al.*, 2003; Gill *et al.*, 2004 y Amorim *et al.*, 2005).

Pero si bien, poseen grandes ventajas, también cuentan con algunas limitaciones:

- ✚ El número de SNPs requerido para alcanzar un poder de discriminación similar al de los STRs es de cuatro veces el número de éstos (Chakraborty *et al.*, 1999 y Gill *et al.*, 2001).

- ✚ El estudio de los SNPs ha sido desarrollado durante los últimos diez años, por lo tanto la experiencia y los datos acumulados son mucho más amplios. Por ejemplo, las mutaciones y los polimorfismos de las regiones flanqueantes de los STRs son más conocidos, lo que puede suponer un problema para la validación de los SNPs en grupos poblacionales (Amorim *et al.* 2005).
- ✚ En el campo forense, además existen dos problemas añadidos, por los cuales, actualmente, los SNPs no pueden reemplazar a los STRs. Por una parte, cuando la muestra está compuesta por una mezcla de fluidos de dos o más individuos; y, por otra la inexistencia de bases de datos de SNPs como puede ser el CODIS para los STRs.

2.3.- APLICACIÓN A LA IDENTIFICACIÓN GENÉTICA.

Las aplicaciones de la huella genética, analizando tanto el ADN mini como microsatélite, son muy numerosas y han experimentado un importante desarrollo en los últimos cinco años. Las secuencias repetitivas en tándem del ADN representan una rica fuente de marcadores altamente polimórficos para el ligamiento genético, el mapeo y la identificación genética de individuos.

Numerosos loci STRs han sido utilizados para el diagnóstico predictivo de muchas enfermedades genéticas, incluyendo distrofia miotónica, fibrosis quística, distrofia muscular de Duchenne-Becker y enfermedad de Huntington (Mully *et al.*, 1991; Oudet *et al.*, 1991 y Morral y Estivill, 1992), entre otras. Actualmente, también se están utilizando mucho los SNPs con este mismo fin (Gabriel *et al.*, 2002)

Así mismo, el tipaje del ADN es una técnica poderosa para determinar la relación entre dos muestras de ADN genómico. Las aplicaciones del tipaje del ADN incluyen la identificación genética personal, para determinar relaciones de parentesco (diagnóstico de paternidad), identificación neonatal y postnatal y con especial relevancia identificación de vestigios biológicos en el campo forense.

La introducción de las técnicas de ADN ha posibilitado la investigación biológica de la paternidad en un gran número de nuevas situaciones que hasta hace poco era impensable. Por ejemplo, en la actualidad y gracias a la tecnología del ADN, es posible la investigación de la paternidad en ausencia del presunto padre, o a partir de restos cadavéricos si éste estuviera fallecido. Por otro lado, el ADN ha hecho posible la investigación de la paternidad a partir de restos fetales, lo cual es de trascendental importancia en la resolución de delitos de aborto y/o infanticidio, así como en la resolución de delitos de violación en los que se produce embarazo.

Otra de las aplicaciones de la huella genética es la identificación de recién nacidos (en casos de que haya sospecha de intercambio) a partir de una gota de sangre del cordón umbilical tomada en la sala de partos, esta identificación puede ser llevada a cabo en la clínica maternal, pero en ocasiones es necesario la identificación tiempo después del parto. En este caso la identificación genética también es posible a partir de la mancha de sangre guardada desde el momento del nacimiento (Rodríguez-Alarcón *et al.*, 1996).

La utilización de la técnica de PCR permite también estudios de determinación del sexo del feto. En la sangre periférica de una mujer embarazada se encuentran aproximadamente 3 ml. de sangre del feto, por tanto, es posible la amplificación de ADN a partir de los escasos linfocitos del feto circulantes en la sangre materna (Lo *et al.*, 1989; Suzumori *et al.*, 1992). Esta aproximación al diagnóstico genético en el feto, sin recurrir a procedimientos invasivos, está siendo objeto de investigación para intentar aplicarlo al diagnóstico de enfermedades hereditarias.

El reciente descubrimiento de la existencia de regiones STR tetraméricas en el cromosoma Y, con una variabilidad comparable a la de los STRs autosómicos, ha permitido disponer de una nueva gama de marcadores genéticos de gran interés, no sólo en el campo de la identificación genética humana, sino también en el campo de la genética de poblaciones humanas, y en estudios evolutivos. El cromosoma Y cuenta con dos minisatélites y unos doscientos marcadores tipo STRs (Butler, 2003 y Kayser *et al.*, 2004). Ambos (minisatélites y microsatélites) son muy diferentes en cuanto a la tasa de mutación. Así, los minisatélites tienen una tasa de mutación de 6-11% por generación (Jobling *et al.*,

1999), mientras que la tasa de mutación de los loci STRs es de 0-2% por generación (Kayser *et al.*, 2000 y Dupuy *et al.*, 2004). La herencia exclusivamente paterna de la región no recombinante del cromosoma Y se traduce en el mantenimiento, a través de las generaciones, de polimorfismos ligados que pueden ser utilizados para trazar la evolución de los linajes paternos. Por otro lado, los STRs localizados en el cromosoma Y se están convirtiendo en una herramienta de gran valor en una gran variedad de situaciones forenses. Los STRs del cromosoma Y, por ejemplo, pueden ser de gran interés para detectar específicamente los restos celulares de un varón en casos en los que sea necesario realizar una identificación genética de una mezcla de fluidos biológicos de varón y de mujer, en los que la fracción de células procedentes del varón sea minoritaria con respecto a la fracción de células de la mujer. Tal situación ocurre en una proporción de casos de violación en los que, en la toma vaginal, por las circunstancias de los hechos o la toma de la muestra, se detecta una muy pequeña cantidad de espermatozoides del agresor que se encuentran mezclados con una gran cantidad de células de descamación del epitelio vaginal de la víctima. Por otra parte, los STRs del cromosoma Y son de gran interés en la investigación biológica de la paternidad de hijos varones, cuando el presunto padre haya fallecido y no se pueda disponer de una muestra biológica del mismo, ya que permite investigar a cualquier pariente por vía paterna y comprobar si su haplotipo de marcadores STR-Y es coincidente o no con el haplotipo de STR-Y del supuesto hijo.

2.4.- MÉTODOS DE ESTUDIO DE LOS POLIMORFISMOS DEL ADN.

La hemogenética forense nace a principios de siglo, cuando Karl Landsteiner describe el sistema ABO de los hematíes y Von Durgan y Hirschfeld descubren su transmisión hereditaria. Esta ciencia surgió como una rama de la Criminalística cuyo objetivo era la identificación genética tanto en casos de investigación criminal como en estudios biológicos de la paternidad. Inicialmente las investigaciones se centraban en el estudio de antígenos eritrocitarios (sistema ABO, Rh, MN), proteínas séricas, enzimas eritrocitarias y sistema HLA. Con el estudio de dichos marcadores podía incluirse o excluirse una persona como posible sospechoso por poseer una combinación genética igual o diferente a la del vestigio biológico hallado en el lugar de los hechos.

Pero fue a mediados de siglo cuando gracias al descubrimiento del ADN y de su estructura y al posterior avance en las técnicas de análisis de

dicha molécula la Hemogenética Forense evolucionó considerablemente hasta el punto de que hoy en día puede hablarse de una nueva subespecialidad dentro de la Medicina Forense: la Genética Forense. Dicha ciencia estudia básicamente unas regiones del ADN que presentan variabilidad entre los distintos individuos, es decir, estudia regiones polimórficas del ADN. Así, analizando un determinado número de regiones polimórficas, la probabilidad de que dos individuos sean genéticamente iguales es prácticamente nula (excepto en el caso de gemelos univitelinos).

A pesar de que la Ciencia poseía las herramientas necesarias para el estudio del ADN su aplicación en la resolución de casos judiciales no se produjo hasta 1.985 cuando el Ministerio del Interior Británico solicitó la ayuda de Alec J. Jeffreys, profesor de Genética de la Universidad de Leicester. Los primeros casos de Criminalística fueron resueltos gracias a la técnica de los RFLPs (fragmentos de restricción de longitud polimórfica). Jeffreys descubrió la existencia de unas regiones minisatélites hipervariables dispersas por el genoma humano que al ser tratadas con enzimas de restricción generaban unos fragmentos de longitud variable. Estudios posteriores realizados por el mismo Jeffreys demostraron que las diferencias en el tamaño de estos fragmentos se debían a que estas regiones consistían en un determinado número de repeticiones en tándem de una secuencia central, el cual variaba de unos individuos a otros. El origen de dichas variaciones puede deberse a dos hechos:

- ✚ En primer lugar, la mutación de un nucleótido por otro en una región que actúe como secuencia diana para la enzima de restricción originará un fragmento de mayor tamaño al no reconocer la enzima dicha secuencia. O bien si corta en otro lugar podrá originar un fragmento de menor tamaño. Variaciones debidas a dichos cambios se denominan polimorfismos mutacionales.
- ✚ O puede que las variaciones se deban a la inserción o delección de una secuencia dentro del fragmento que es cortado por la enzima de restricción en cuyo caso se denomina polimorfismo de longitud o de inserción-delección.

El primer locus de ADN polimórfico fue descubierto usando una sonda de ADN arbitraria (Wyman y White, 1980). De esta manera observaron fragmentos de más de 15 longitudes diferentes en una pequeña muestra de individuos. Posteriormente se encontraron otros loci hipervariables como la secuencia del gen de la insulina humana (Bell *et al.*, 1982), en el oncogen "ras" (Capon *et al.*, 1983), en el pseudogen de la zeta-globina (Proudfoot *et al.*, 1982) y en el gen de la mioglobina (Jeffreys *et al.*, 1985b). Estos loci hipervariables constaban de repeticiones en tándem de una secuencia de oligonucleótidos (11 a 60 pb), de manera que las diferentes longitudes de los fragmentos originados dependían del número de dichas repeticiones y se les denominó VNTR ("Variable Number of Tandem Repeat").

Tras el descubrimiento de los primeros VNTRs se vio que éstos podían ser aplicados a la medicina forense y sustituir a los marcadores clásicos (Jeffreys *et al.*, 1985b).

La PCR o reacción en cadena de la polimerasa, supuso una revolución en muchos campos de la Biología, Medicina y, como no, especialmente, en la Genética Forense.

2.4.1.- REACCIÓN EN CADENA DE LA POLIMERASA.

La reacción en cadena de la polimerasa (conocida como PCR por sus siglas en inglés, Polymerase Chain Reaction) permite amplificar más de un millón de veces un ADN obtenido a partir de la región seleccionada del genoma, siempre y cuando se conozca una parte de su secuencia de nucleótidos. Esta técnica fue ideada en 1983 por Kary B. Mullis que obtuvo el premio Nobel de Química en 1993 por dicho invento.

Para la PCR se utilizan dos oligonucleótidos sintéticos de unos 15-20 nucleótidos que son complementarios a las zonas flanqueantes de la región que se quiere amplificar. Estos oligonucleótidos (habitualmente conocidos por su nombre en inglés, "primers") actúan como cebadores para la síntesis *in vitro* de ADN, la cual, está habitualmente catalizada por una enzima llamada Taq polimerasa. Dicha enzima se aísla de una bacteria termófila, denominada *Thermus aquaticus*, que es capaz de crecer a temperaturas elevadas. A esta temperatura dicha enzima es capaz de mantener una media de extensión de más de 60 nucleótidos por segundo en regiones ricas en uniones G-C. La temperatura óptima a la

que actúa la Taq polimerasa permite el uso de elevadas temperaturas para la unión de los primers y para la extensión, de esta manera se aumenta el nivel de exigencia de la reacción y se reduce la extensión de los primers unidos inespecíficamente al ADN.

La reacción se lleva a cabo en una serie de ciclos cada uno de los cuales incluye tres fases o pasos (Vosberg HP *et al*, 1989; Mullis K,1990):

- ✚ DESNATURALIZACIÓN: Para que comience la reacción es necesario que el ADN molde se encuentre en forma de cadena sencilla. Esto se consigue aplicando temperaturas de 90 a 95°C que producen la rotura de los puentes de hidrógeno intercatenarios y por lo tanto la separación de ambas cadenas. Para conseguir la completa separación de las hebras de toda la muestra esta temperatura debe mantenerse unos minutos. Si el ADN solo se desnaturaliza parcialmente éste tenderá a renaturalizarse rápidamente, evitando así una eficiente hibridación de los primers y una posterior extensión.
- ✚ HIBRIDACIÓN: Esta fase se denomina también fase de "annealing" o de emparejamiento. Una vez que el ADN está desnaturalizado se disminuye la temperatura hasta un rango comprendido entre los 40 y los 60°C para que se pueda producir la unión de los primers a las secuencias flanqueantes del fragmento que se va a amplificar. La temperatura de fusión o annealing (T_m , "melting temperature") depende de varios factores y es relativamente específica para cada primer. La longitud de los primers y la secuencia son críticas en la designación de los parámetros de una amplificación, una fórmula simple para calcular la T_m es la siguiente:

$$T_m = 4(G+C) + 2(A+T)$$

No obstante, cada primer exige una serie de estudios experimentales para determinar su temperatura de annealing específica ya que si la temperatura es muy baja la unión se hará de forma inespecífica y si es muy alta no se producirá una unión completa.

✚ **EXTENSIÓN:** Durante este paso la *Taq polimerasa* incorpora nucleótidos en el extremo 3' del primer utilizando como molde la cadena de ADN previamente desnaturalizada. La temperatura a la que se lleva a cabo este paso suele ser de 72°C ya que es la temperatura a la que la *Taq polimerasa* alcanza su máxima actividad. Normalmente una extensión de 20 segundos es suficiente para fragmentos menores de 500 pb, y 40 segundos para fragmentos por encima de 1.2Kb.

Un factor importante en el transcurso de las diferentes fases es el **tiempo de rampa**. Este se define como el tiempo invertido en pasar de una temperatura a otra y depende del diseño y de las características del aparato donde se realiza automáticamente este proceso, el termociclador. En las nuevas generaciones de termocicladores este factor se ha ido optimizando para hacerlo mínimo.

2.4.1.1.- PROCEDIMIENTO ANALÍTICO.

El mantenimiento de la vida y su transmisión depende, en cualquier ser vivo, de la copia adecuada del ADN que poseen unas células y que ha de ser transmitido, de modo íntegro y exacto, a las células hijas. El ADN de cualquier célula necesita, pues, ser copiado continuamente por los organismos, por lo que éstos poseen unas enzimas capaces de, a partir de un molde dado, crear una copia de ADN exactamente igual, hecho que se basa en la complementariedad de las cadenas de ADN.

Recientemente, la ciencia ha conseguido aislar todos los productos que intervienen en esta reacción, de tal modo que se pueden copiar artificialmente fragmentos determinados de ADN hasta que haya cantidad suficiente para poder realizar los estudios deseados. Los componentes básicos para poder efectuar PCR son:

❖ BUFFER DE AMPLIFICACIÓN

Los buffer de PCR que se utilizan normalmente contienen KCl, Tris y MgCl₂. El buffer empleado por nosotros es un buffer comercial concentrado diez veces (10X) y que contiene:

- Gelatina ..0.01% (w/v), en un volumen de 1,4 ml

- KCl500 mM
- Tris-ClH...100 mM, pH 8,3
- MgCl₂.....15 mM

El **MgCl₂** es el componente que más influye en la especificidad y rendimiento de la reacción ya que los iones Mg²⁺ son necesarios para la actividad de la *Taq polimerasa*, es decir, actúan como cofactores de la polimerasa.

La concentración óptima de MgCl₂ está en torno a 1.5 mM si se emplean concentraciones de 200 mM de cada uno de los dNTPs. No obstante, a veces es necesario probar con diferentes cantidades de Mg ya que un exceso del mismo origina una acumulación de productos inespecíficos y una cantidad insuficiente hace que disminuya el rendimiento de la amplificación.

❖ PRIMERS

A la hora de elegir unos primers para amplificar un determinado fragmento de ADN hay una serie de reglas a seguir (Coyne *et al*, 1996):

- La *longitud de cada uno de los primers* debe estar comprendida entre 18 y 24 bases ya que se ha comprobado que primers de mayor longitud (30-35 bases) no aumentan el rendimiento y los primers cortos carecen de suficiente especificidad.
- Ambos primers deben tener una *T_m* similar (como mucho la diferencia entre ambas temperatura debe ser de 5°C).
- La *relación bases púricas: bases pirimidínicas* debe ser 1:1 (o como mucho 40-60%).
- La *secuencia de los primers* debe comenzar y terminar con 1-2 bases púricas.
- Para evitar la formación de *dímeros de primers* es necesario comprobar que los primers no contengan secuencias complementarias entre sí.

Los dímeros de primers consisten en fragmentos de doble cadena cuya longitud es muy próxima a la de la suma de los primers y se

producen cuando un primer es extendido a continuación del otro. El mecanismo exacto por el que se forman estos dímeros no está completamente determinado. La observación de que primers con los extremos 3' complementarios favorecen su formación sugiere que el paso inicial se debe a interacciones transitorias que aproximan los extremos complementarios. Algunas polimerasas, incluida la *Taq*, han mostrado una débil actividad polimerizadora no dirigida por un ADN patrón, la cual puede unir nucleótidos adicionales al doble extremo apareado. Si esta actividad puede producirse sobre una hebra sencilla de oligonucleótidos, resultaría una buena oportunidad para que la extensión formara un corto solapamiento en el extremo 3' con el otro primer, suficiente para promover la formación del dímero.

❖ DESOXINUCLEÓTIDOS TRIFOSFATOS

Las concentraciones de dNTPs que suelen usarse están en torno a 200 μM para cada uno de ellos. En un volumen de reacción de 25 μl con esta concentración de dNTPs se sintetizarían entre 6-6.5 μg de ADN. La concentración de dNTPs y de MgCl_2 va relacionadas ya que el Mg se une a los dNTPs con lo que concentraciones elevadas de dNTPs inhibirían la reacción al no tener la *Taq polimerasa* suficiente Mg como para incorporar dNTPs. Para una concentración de 200 μM de cada dNTP se suele añadir MgCl_2 a una concentración de 1.5 mM.

❖ TAQ-POLIMERASA

Las cantidades óptimas de *Taq polimerasa* necesarias para la síntesis de ADN están alrededor de 2 unidades en 25 μl de volumen final de reacción. La actividad de este enzima se ve influenciada por la concentración de dNTPs, de Mg^{2+} y de algunos iones monovalentes de manera que concentraciones elevadas de los mismos inhiben dicha actividad.

Por otro lado, pequeñas concentraciones de KCl estimulan la actividad sintética de la *Taq* en un 50-60% con un máximo aparente cuando su concentración es de 50 mM. Existen algunos datos relacionados con la influencia de ciertos reactivos que se emplean antes de la amplificación y que alteran la actividad de la *Taq*. Por ejemplo concentraciones 1M de urea estimulan la actividad, el SDS a bajas concentraciones que la inhibe al igual que concentraciones mayores del 10% de etanol.

Actualmente, nosotros usamos *Taq polimerasa gold* (Applied Biosystems), que tiene la ventaja de que para activarse necesita un aumento de temperatura, concretamente 90°C durante 11 minutos, con lo cual su actividad no comienza hasta que no se introduce en el termociclador.

❖ ADN MOLDE O "TEMPLATE"

Es el ADN del cual queremos copiar un determinado fragmento, es, por tanto, el ADN que la *Taq* polimerasa utiliza como molde para la síntesis de nuevas cadenas polinucleotídicas. La cantidad de ADN necesaria para la PCR depende de varios factores:

- Del marcador que se va a amplificar: hay marcadores cuyos primers son más específicos o bien cuyas condiciones de amplificación están mejor optimizadas que las de otros. Por esta razón puede darse el caso de que cierta cantidad de ADN (sobre todo cuando jugamos con cantidades mínimas) amplifique para unos marcadores pero no para otros. Por ello, cuando en un laboratorio se va a utilizar un nuevo marcador es necesario hacer un estudio de validación en él que se incluye un estudio de sensibilidad. De dicho estudio de sensibilidad puede sacarse como conclusión cuál es la mínima cantidad de ADN que amplifica en condiciones estándar. De manera general, para casi todos los STR utilizados en Genética Forense la cantidad óptima de ADN que asegura un rendimiento adecuado está en torno a los 1 ng.
- Calidad del ADN: Cuando se trabaja con ADN cuya calidad es óptima no suele haber problemas en la amplificación y cantidades del mismo por encima e incluso por debajo de los 1 ng rinden buenos resultados. El problema aparece cuando la calidad del ADN obtenido no es la idónea, bien porque esté degradado o bien porque dicho ADN vaya ligado a una serie de contaminantes que pueden inhibir la actividad de la polimerasa. Si el ADN está degradado por la acción de enzimas de restricción el que obtengamos o no resultado en la amplificación va a depender de que el fragmento a amplificar haya sido dañado o no. En el caso en el que tengamos ADN sin degradar pero unido a una serie de contaminantes habría que intentar diluir al

máximo la muestra para disminuir dichos contaminantes, pero siempre dentro de un rango de ADN que no esté por debajo del límite de sensibilidad de la PCR. El problema de las cantidades mínimas de ADN y la presencia de contaminantes o inhibidores de la Taq es un hecho habitual en Criminalística y requiere un estudio pormenorizado de la muestra antes de la amplificación.

❖ ADYUVANTES DE LA PCR

Son elementos que mejoran el rendimiento y la especificidad de la PCR. Aunque algunos autores han recomendado el uso del **DMSO** y del **glicerol**, el adyuvante más extendido y utilizado es el **BSA**. A concentraciones por encima de 0.8 µg/µl el BSA incrementa la eficiencia de la PCR ya actúa como una proteína captadora de iones que pueden ser inhibidores de la Taq polimerasa.

2.4.1.2.- VENTAJAS Y LIMITACIONES PRÁCTICAS.

🚦 **Ventajas:** La amplificación de ADN que se encuentra en indicios muy pequeños (un solo pelo o cabello, la saliva depositada en la colilla de un cigarrillo, una mancha de sangre,...) es, sin duda alguna, la principal ventaja que para la ciencia criminalística ofrece la PCR. Y no sólo porque se va a poder trabajar con un material de partida muy pequeño, sino porque se va a poder guardar parte del mismo sin usar para posteriormente realizar contraperitajes u otras pruebas que se estimen necesarias para confirmar o desmentir un resultado.

Cantidades tan pequeñas 0.1 ng pueden ser suficientes (en condiciones ideales) para conseguir resultados de amplificación de un locus.

Los indicios biológicos obtenidos en casos criminales suelen verse afectados por procesos degradativos y contaminantes, propios de una conservación en medios hostiles (temperaturas medioambientales extremas, coexistencia con otros restos orgánicos que se contaminan biológicamente con bacterias y hongos,...).

El tamaño de los fragmentos de amplificación (alelos) oscila, aproximadamente, entre los 200 y los 1000 pares de bases, por lo que son muy pequeños en comparación con los 15000 a 30000 pares de bases

que pueden llegar a tener los fragmentos estudiados con la hibridación de sondas monolocus.

Esto les confiere una gran estabilidad y resistencia a la degradación de cualquier origen (físico, químico o biológico). En consecuencia, llegado el momento de analizar un indicio, será mucho más probable obtener resultados positivos si estudiamos un trozo (entre los 6000 millones de pb que tiene todo el ADN) que posee 300 pb, que si estudiamos un trozo que tiene 30000: por pura lógica, es más difícil que se rompa el trozo más pequeño que el más grande.

Esta circunstancia tiene una importancia enorme en criminalística e identificación humana, ya que la estabilidad de alelos pequeños, como los denominados STRs estudiados por PCR garantizan a los laboratorios la conservación del ADN incluso después de muchos años, o a pesar de que existan condiciones ambientales extremas, lo que adquiere especial importancia en la identificación de restos humanos que, por ejemplo, han estado desaparecidos mucho tiempo o sometidos a temperaturas extremas (grandes incendios, accidentes de aviación...), etc.

En los grandes laboratorios de criminalística y biología forense, así como en casos especialmente complicados o donde existen muchos indicios diferentes, no deja de tener importancia la característica de la rapidez con que pueden obtenerse resultados, que en los casos más favorables permite a los laboratorios ofrecer datos iniciales a las 24-48 horas de recibir muestras, cuando el análisis con sondas requiere, en los mejores casos, un mínimo de 72-96 horas, si no mucho más.

Por el tipo de técnicas utilizadas, la disminución de la probabilidad de error es una realidad evidente. El error disminuye no sólo a la hora de interpretar los resultados, sino también al disminuir la manipulación de las muestras, es decir, el número de veces que una muestra es cambiada de tubo o aparatos para su estudio.

 **Inconvenientes:** La presencia de contaminación biológica es el gran problema de la PCR. Esta técnica tiene su talón de Aquiles en la que es su gran ventaja: la posibilidad de amplificar cualquier resto de ADN, por pequeño que sea.

Las muestras de origen biológico recogidas como indicios en las escenas donde han ocurrido hechos violentos tienen muchas posibilidades

de estar mezcladas con muestras procedentes de otras personas, hecho este conocido como contaminación biológica, y que se descubre cuando tras analizar un resto o indicio biológico con PCR aparece ADN de más de una persona.

De modo general, sí que ha de mencionarse que la contaminación biológica puede ser de tres tipos, según acontezca antes de la recogida, durante la inspección, recogida y traslado de las muestras y durante su estudio en el laboratorio. De este modo, la contaminación criminal o accidental es la que ocurre antes de que los indicios sean encontrados y manipulados por los investigadores. La contaminación manipulativa es totalmente evitable, es aquella que ocurre al llegar a la escena del crimen las fuerzas de orden público, autoridades, miembros de comisiones judiciales y todo tipo de auxiliares y colaboradores, ya que penetran sin tomar todas las medidas de prevención de la contaminación biológica. Por último, la contaminación analítica ocurre al manipular las muestras en el laboratorio y, al igual que la manipulativa, es perfectamente evitable.

Existen circunstancias en las que hay una falta de amplificación del ADN remitido al laboratorio, bien porque el ADN se encuentre muy degradado, de tal modo que los fragmentos (alelos) que intentamos amplificar, pese a que sean muy pequeños, están rotos y no pueden actuar de molde para que se copie una nueva cadena. También puede existir falta de amplificación por contaminación química, ya que existen miles de productos (colorantes, tóxicos, metales, etc...) capaces de inhibir la acción de la enzima polimerasa. La gran mayoría de los contaminantes pueden ser obviados y desechados con una buena extracción del ADN de los indicios, pero otras veces es totalmente imposible, sobre todo en indicios muy pequeños y muy contaminados.

2.4.2.- SISTEMAS DE ANÁLISIS DE LOS POLIMORFISMOS.

El análisis de los loci STR en el campo de la identificación genética humana se ha facilitado enormemente gracias a la posibilidad de amplificar de manera simultánea dos o más loci STR en una única PCR mediante el procedimiento denominado multiplex-PCR (Edwards, 1991). Las ventajas de los sistemas de amplificación múltiple con respecto al análisis individual de cada loci son obvias: la posibilidad de obtener información de varios loci en una sola reacción aumenta el poder de discriminación del sistema, disminuyendo el tiempo de análisis, así como la cantidad de ADN templado y la cantidad de reactivos necesarios.

Los productos de amplificación de una reacción multiplex-PCR pueden ser separados de acuerdo a su tamaño mediante electroforesis en gel de poliacrilamida bajo condiciones desnaturalizantes y posteriormente detectados mediante dos tipos de métodos: detección manual mediante tinción con nitrato de plata y detección fluorescente semiautomática, en nuestro caso, hemos utilizado el segundo método, por lo tanto, nos vamos a ceñir a él.

2.4.3.- MARCAJE FLUORESCENTE Y DETECCIÓN SEMIAUTOMÁTICA.

La electroforesis capilar es un método alternativo para la separación de fragmentos y obtención de ADN que supe, en parte, los inconvenientes de los sistemas convencionales. En este caso el soporte o medio de separación es un polímero incluido en un capilar de silica de unos 50µm y de longitud variable lo cual hace que la cantidad de calor generado sea menor y que puedan aplicarse voltajes mayores. La gran ventaja en cuanto a rapidez de la técnica se debe a que la preparación del gel y la carga de muestras se hacen de manera automática. Por otro lado se consigue una mayor sensibilidad, al poderse obtener resultados interpretables en los casos en los que el número de copias obtenidas por PCR es bastante bajo, y aumenta la capacidad para diferenciar alelos que se distinguen en un par de bases. Además, los resultados obtenidos son analizados por un software evitándose así problemas de interpretación y permitiendo que éstos queden almacenados para posibles futuros análisis.

Para que pueda llevarse a cabo el análisis por electroforesis capilar es necesario que el ADN sea amplificado utilizando un par de primers o más (en el caso de los multiplex) marcados en el extremo 5' con unas moléculas llamadas **fluorocromos** los cuales emiten fluorescencia a diferentes longitudes de onda cuando son excitados por láser. En los kits de amplificación los fluorocromos van ligados a uno de los primers (con lo cual se evitan los problemas derivados de la diferente movilidad de las dos cadenas de ADN), mientras que en kits de secuenciación pueden ir marcados o bien los primers o bien los dideoxinucleótidos. En el kit utilizado en este trabajo (Y-Filer Applied Biosystems) los primers vienen marcados con los fluorocromos: *6-FAM*, *VIC*, *NED* y *PET*.

El software de recogida de datos define ciertas áreas en una **cámara CCD** para recoger exclusivamente las emisiones procedentes del grupo de

fluorocromos que estemos utilizando. Estas áreas se conocen con el nombre de **filtros virtuales** y su funcionamiento es similar al de los filtros físicos que se utilizan para separar las diferentes longitudes de onda. A pesar de que existe cierto solapamiento entre los diferentes rangos, este problema es minimizado gracias a la creación en el ordenador de una **matriz** matemática. Dicha matriz consta de cinco filas y cinco columnas y los números indican el grado de solapamiento entre las longitudes de onda emitidas por los cinco fluorocromos. Los valores de la matriz varían según el analizador, el filtro virtual y las condiciones de electroforesis por lo que ésta debe ser creada para cada analizador y para unas determinadas condiciones electroforéticas.

El analizador ABI Prism 310 utiliza varios tipos de filtros virtuales denominados *A, B, C, D, E, F* y *G*. Los filtros *A, C, D, F* y *G* son usados en aplicaciones de GeneScan que es un software para análisis de fragmentos, mientras que el *A, B,* y *E* se usan para aplicaciones de secuenciación. En nuestro caso, para el análisis de fragmentos amplificados con el kit Y-Filer el filtro utilizado fue el *G5*. Cada filtro combina cuatro o cinco fluorocromos (incluyendo el reservado para el estándar de tamaño) asignando un color a cada uno de ellos.

El **capilar** en el que se lleva a cabo la electroforesis se encuentra cubierto por poliamida opaca excepto por una pequeña zona denominada "ventana del capilar" la cual es atravesada por el láser y permite que las muestras sean excitadas a su paso por la misma.

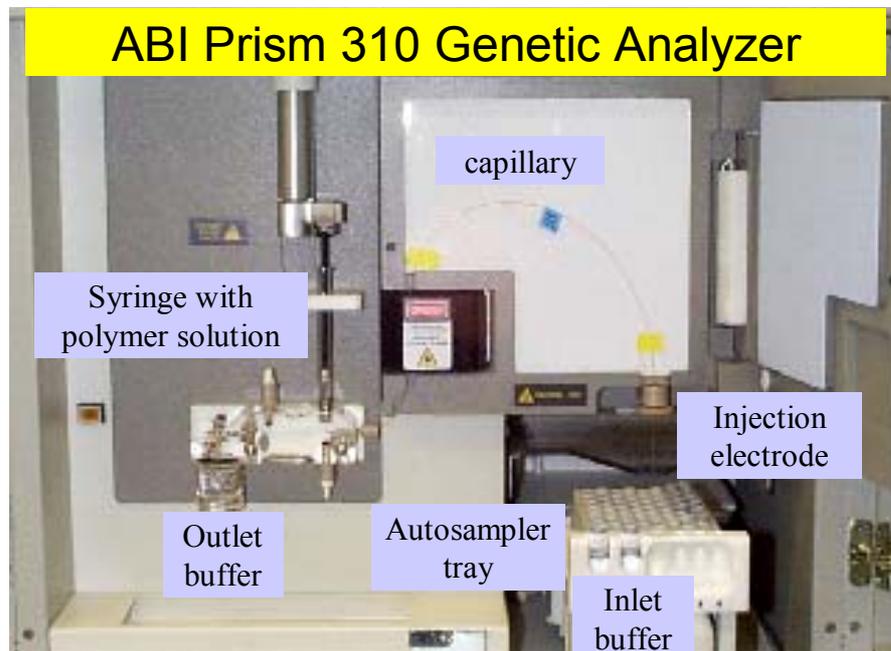
Finalmente los datos analizados son enviados al ordenador el cual los transforma en secuencias de ADN o en fragmentos analizados con sus correspondientes alelos asignados.

DESCRIPCIÓN DEL ANALIZADOR

El instrumento de electroforesis capilar utilizado en este trabajo ha sido el analizador genético **ABI Prism 310** de Perkin Elmer Applied Biosystems. Fundamentalmente el analizador genético consta de dos partes: una en la que se lleva a cabo todo el proceso de inyección de la muestra, electroforesis y detección de la fluorescencia y otra que es el sistema informático acoplado al analizador.



Las distintas partes donde se lleva a cabo la electroforesis, que posteriormente describiremos, se pueden observar en la siguiente imagen:

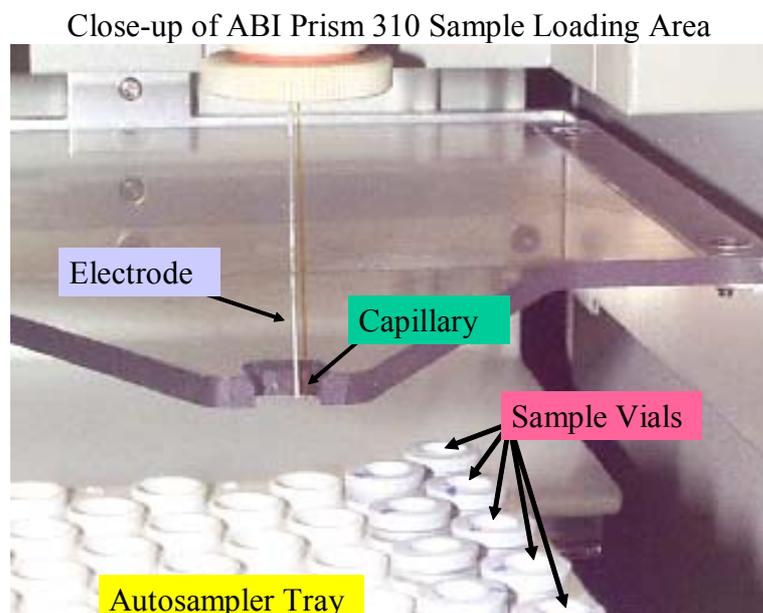


Una vez que las muestras amplificadas han sido preparadas y desnaturalizadas se colocan en un soporte con capacidad para 48 o 96 tubos y que encaja en la parte del analizador destinada a la carga de

muestras (conocida como distribuidor de muestras o **autosampler**, que es su denominación inglesa). El autosampler se encarga de mover el soporte y, por lo tanto, los tubos de manera que el capilar pueda ser introducido en los mismos. Además dispone de:

- 2 viales de 4 ml: uno en el que se añade el buffer y otro con agua destilada para lavar el electrodo negativo y el capilar entre unos análisis y otros.
- Un tubo eppendorf de 1.5 ml que es el vial de deshecho, al cual irá la carga de polímero utilizada para cada muestra una vez terminado el análisis.

Encima del autosampler y junto al capilar a una distancia de unos 2 mm se encuentra el **electrodo negativo** que está conectado a una fuente de alto voltaje (15000 V) y proporciona la corriente necesaria para la electroforesis.



Existe otra zona en el aparato que es la encargada de bombear el polímero hacia el capilar. Esta zona consta de una jeringa que se rellena de polímero y un bloque de metacrilato con conductos internos y donde se encuentra además el electrodo positivo o ánodo al cual va acoplado el reservorio de buffer. A través de los conductos internos, los electrodos

inmersos en el buffer quedan conectados durante la electroforesis, generándose así la diferencia de potencial necesaria para la separación de los fragmentos de ADN.

Entre las válvulas del bloque de metacrilato se encuentran:

- Válvula del buffer del ánodo: se mantiene cerrada mientras se realiza el relleno del capilar y se abre para contactar el polímero con el buffer durante la electroforesis.
- Válvula de deshecho: permite la salida de exceso de polímero.

La última zona importante en el analizador es la dedicada a la electroforesis y detección de fluorescencia la cual consta de:

- Una **placa** o plato que proporciona al capilar la temperatura adecuada para la electroforesis (60°C en análisis de fragmentos y 50°C en secuenciación)
- El **capilar de sílica** relleno de polímero. Es el soporte de la electroforesis.
- Un **láser de argón**.
- Una **cámara CCD**.
- Una **ventana detectora** en la cual se hace coincidir la ventana del capilar.

Una vez que las muestras han sido analizadas los datos generados son recogidos por el **ordenador**. El ordenador acoplado al secuenciador ABI Prism 310 es un modelo de Macintosh que permite la instalación de diversos programas informáticos con diferentes funciones. El ABI Prism 310 Data Collection Software es el programa que recoge lo que se denomina el dato bruto o raw data y que es la fluorescencia detectada en la cámara CCD cuando los fragmentos marcados, separados por electroforesis y excitados por el láser pasan por la ventana de detección. Existen dos programas que analizan el dato bruto y lo convierten en secuencia de ADN o en fragmentos de un determinado tamaño:

- **ABI Prism®DNA Sequencing analysis:** para secuenciación
- **GeneScan® Analysis:** para análisis de fragmentos.

En este trabajo, puesto que se han analizado fragmentos de ADN, los programas utilizados han sido el ABI Prism® 310 Data Collection y el GeneScan® Analysis. Una vez analizados los fragmentos y con sus correspondientes tamaños (en pb) se hizo una asignación automática de los alelos mediante otro programa informático llamado **Genotyper®2.0**.

VENTAJAS DEL USO DE LA ELECTROFORESIS CAPILAR FRENTE A SISTEMA DE TIPIFICACION MANUAL

1. El hecho de utilizar diferentes fluorocromos para cada locus permite el análisis simultáneo de varios loci aunque éstos posean alelos con tamaños solapantes.
2. La utilización de un estándar de tamaño interno permite medir de manera automática el tamaño de los alelos eliminando las diferencias de movilidad electroforética que a veces existen en las calles de un gel.
3. Al marcarse sólo uno de los primers se elimina el problema del desdoblamiento de bandas que tan frecuentemente se observa con la tipificación manual como consecuencia de la diferente movilidad de las cadenas complementarias de algunos STR.
4. Los resultados se obtienen de manera informatizada lo cual facilita su análisis a través de programas informáticos.

2.5.- INTRODUCCIÓN AL CROMOSOMA Y.

Dentro de los interrogantes que se plantean los seres humanos desde lejanas épocas, la curiosidad de sus orígenes ocupó siempre un lugar muy relevante. Tradicionalmente, el pasado ha sido investigado por historiadores, arqueólogos y paleontólogos, en tanto que hoy en día se tiende a buscar la evidencia indirecta de poblaciones humanas modernas a través de la lingüística y, cada vez en mayor medida, la biología molecular. Desde que ha sido factible el análisis de nuestro material genético con nuevas técnicas, el estudio del cromosoma Y, por sus especiales características, está despertando grandes expectativas al

revelarse como una útil fuente de polimorfismos de ADN, especialmente en el ámbito de la evolución humana y las ciencias forenses.

Tras la fecundación, la mitad del ADN del cigoto procede de la madre y la otra mitad del padre. La mujer cede a su descendencia obligatoriamente un cromosoma X, el varón, en cambio, puede contribuir con un cromosoma X o Y determinando, por tanto, el sexo de su descendencia. Este cromosoma en los mamíferos es el responsable del desarrollo testicular.

2.5.1.- SU NATURALEZA.

Es uno de los cromosomas más pequeños del genoma humano, con un tamaño aproximado de 60Mb (Morton 1991). Desde el punto de vista citológico, el cromosoma Y está formado por regiones de heterocromatina y eucromatina.

La región heterocromatínica se sitúa sobre el brazo largo (Yq) en posición distal. Se compone de secuencias altamente repetitivas, como por ejemplo DYZ1 y DYZ2, presentando polimorfismos de longitud.

La región de eucromatina se localiza en el brazo corto (Yp), centrómero y en la zona proximal del brazo largo. Su tamaño es bastante constante en varones normales (alrededor de 30 Mb). Es la región de mayor interés genético. Aquí se encuentran algunos genes, así como secuencias que muestran homología con el cromosoma X y los autosomas, y también secuencias repetitivas específicas del cromosoma Y (DYZ3, DYZ4, DYZ5).

Aunque el grueso de su volumen se considera genéticamente inerte, en el cromosoma Y se pueden diferenciar tres regiones interesantes desde el punto de vista genético: dos pseudogenes (PAR1 y PAR2) y una específica del cromosoma Y. (Figura 1).

2.5.2.- LAS REGIONES PSEUDOAUTOSÓMICAS.

A pesar de su distinta naturaleza morfológica, los cromosomas X e Y pueden aparearse durante la meiosis e intercambiar información. Este fenómeno ocurre en pequeñas regiones que contienen secuencias homólogas presentes en ambas cromosomas, denominadas regiones pseudoautosómicas porque en ellas las secuencias de ADN no muestran una herencia ligada al sexo en sentido estricto. En los cromosomas sexuales humanos hay dos de estas regiones, denominadas PAR1 y PAR2, situadas respectivamente en los extremos terminales del brazo corto y largo de dichos cromosomas.

La región pseudoautosómica mayor (PAR1) del cromosoma Y tiene un tamaño aproximado de 2.6 Mb. La recombinación a nivel de PAR1 es necesaria para una segregación normal de los cromosomas X e Y en la meiosis (Ellis y Goodfellow 1989a) y, por tanto, recombina siempre. Pueden existir alteraciones, como en el síndrome de Klinefelter, cuyo cariotipo 47, XXY es producto de la no-disyunción, no produciéndose recombinación a nivel del PAR1 en la gran mayoría de casos (Hassold *et al.*, 1991).

La región pseudoautosómica menor (PAR2) del cromosoma Y mide sobre 320 Kb y no siempre participa en procesos de recombinación. Además, la actividad de recombinación de PAR2 no puede sustituir a la de PAR1 (Freije *et al.*, 1992), no siendo, por tanto, ni necesaria ni suficiente para el éxito de la meiosis masculina.

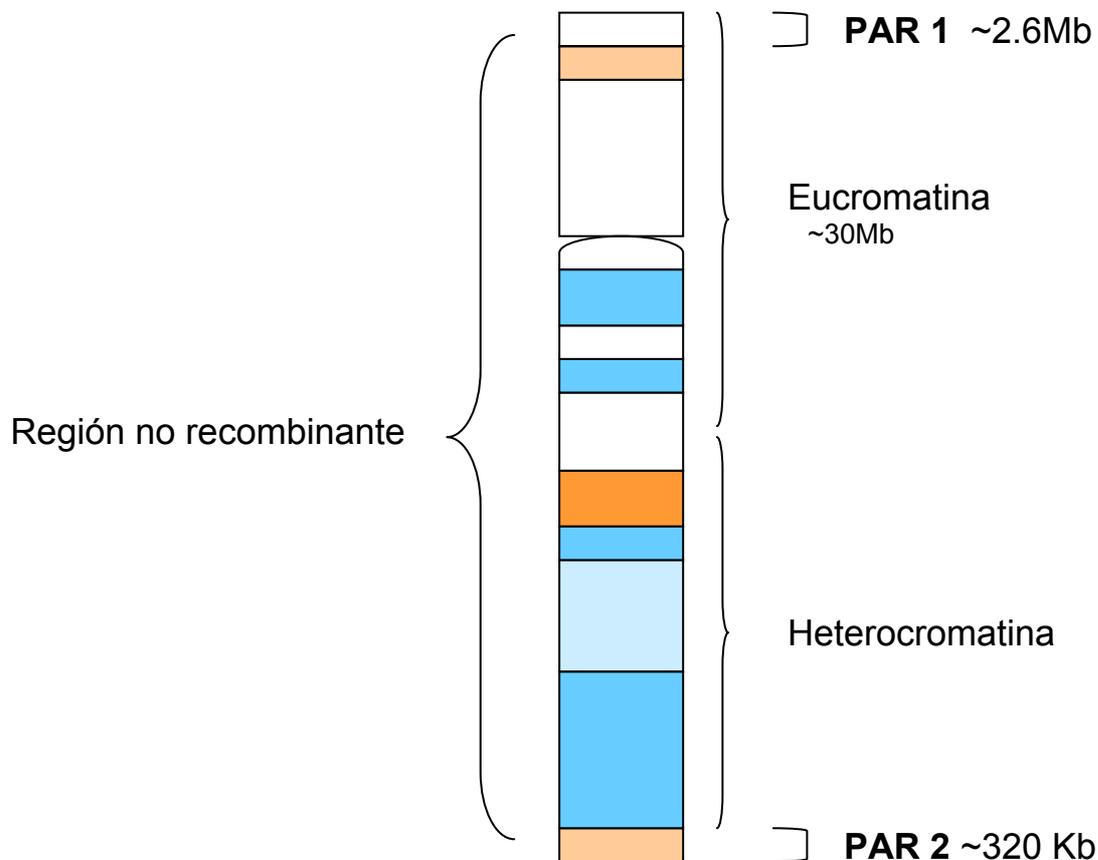


Figura 1: Esquema del cromosoma Y

Se espera que los genes localizados en estas regiones escapen al mecanismo de inactivación, para asegurar de este modo un reparto equivalente de genes entre los sexos. De los genes que se citan a continuación, todos tienen homólogo en X.

- ✚ El gen **MIC2Y** (PAR1) ha sido el primer gen pseudoautosómico descrito en el hombre (Goodfellow *et al.*, 1983). Codifica una proteína estructural del antígeno de superficie celular CD99. Su homólogo en X es **MIC2**.
- ✚ El gen **ASMTY** (PAR1) codifica una enzima que cataliza la última etapa de la síntesis de la hormona melatonina (Yi *et al.*, 1993). Su homólogo en X es **ASMT**.
- ✚ El gen **XE7Y** se localiza proximalmente respecto al anterior. Fue aislado en 1992 (Ellison *et al.*, 1992) y codifica el antígeno

XE7, de función desconocida, pero que presenta homología con el CD99. Su homólogo en X es **XE7**.

- ✚ **IL3RAY** (PAR1) es el gen del receptor de la interleucina-3. Se localiza proximal a CSFRAY (Kremer *et al.*, 1993). Su homólogo en X es **IL3RA**.
- ✚ **CSF2RAY** (PAR1) es el gen más distal de los descritos en PAR1 hasta el momento. Codifica el receptor del factor estimulante de las colonias de granulocitos-macrófagos (GSM-CSF) (Goug *et al.*, 1990). Su homólogo en X es **CSF2RA**.

También han sido descritos algunos genes en la región PAR2, como el gen **SYBL1Y** que está inactivado en el cromosoma Y (D'Esposito *et al.*, 1996) y el gen del receptor de la interleucina-9, **IL9RY** (Kermouni *et al.*, 1995).

2.5.3.- LA REGIÓN ESPECÍFICA DEL CROMOSOMA Y.

Esta región no recombina con ninguno de los cromosomas, por lo cual sus loci se transmiten inalteradamente por línea paterna de generación en generación (en ausencia de mutaciones). Presenta algunos genes funcionales, de gran importancia para el desarrollo sexual del hombre.

- ✚ El **SMCY** es un gen que codifica el antígeno menor de histocompatibilidad H-Y, fue el primer gen identificado en la región específica del cromosoma Y (Wachtel *et al.*, 1975). Su homólogo en X es **SMCX**.
- ✚ El gen **ZFY** codifica una proteína que presenta un motivo en dedo de zinc (Page *et al.*, 1987) cuya función no es conocida por el momento. Su homólogo en X es **ZFX**.
- ✚ El gen **SRY** ha sido aislado en el brazo corto del cromosoma Y, a 5 Kb en dirección proximal a PAR1 (Sinclair *et al.*, 1990). Presenta un solo exón y carece de homólogo en X.

- ✚ La familia de genes **YRRM** se cree que podría intervenir en la regulación de la espermatogénesis. Como el gen anterior, este grupo carece de homólogo en X.
- ✚ El **RPS4Y**, localizado entre SRY y ZFY, es un gen que codifica la proteína ribosómica S4. Posee un homólogo en X que escapa a la inactivación (Fisher *et al.*, 1990).
- ✚ El **AMGY** es un gen que codifica un producto amelogenine-like, también tiene un homólogo en X (Nakahori *et al.*, 1991), que es responsable de la formación normal del esmalte dentario.

2.5.4.- ALGUNAS PECULIARIDADES DEL CROMOSOMA Y. ANALOGÍA CON EL ADN MITOCONDRIAL.

El cromosoma Y presenta una compleja secuencia, con grandes fragmentos de ADN no codificante y numerosas familias de secuencias repetitivas dispersas o en tándem. El hecho de que el cromosoma Y sea extraordinariamente pobre en genes (Lahn *et al.*, 1997) hace pensar en una baja presión selectiva en las zonas no codificantes del cromosoma, por lo que podría acumular más mutaciones que los otros cromosomas. Además, el que permanezcan más tiempo en la línea germinal, y que la formación de esperma conlleve un alto número de divisiones celulares, también podría favorecer la aparición de mutaciones (Weber y Wong 1993, Brinkmann *et al.*, 1998).

Paradójicamente, la frecuencia de polimorfismos en regiones no codificantes del cromosoma Y es mucho menor que en el resto del genoma nuclear. Hay varias posibles explicaciones que intentan justificar este bajo nivel de polimorfismos observado en el cromosoma Y respecto a otros cromosomas (Jobling y Tyler-Smith 1995).

Una de las proposiciones se basa en una razón aritmética, ya que la población efectiva de cromosomas Y (número de cromosomas que pasan a la descendencia) en un momento determinado es claramente inferior a la de otros cromosomas (por cada cromosoma Y están presentes 3

cromosomas X y 4 de cada autosoma), lo que refleja el hecho de que existe correlación entre diversidad y tamaño de la población cromosómica (Kimura 1983). Todo esto hace que procesos como la deriva genética hayan tenido un efecto mucho más patente sobre el cromosoma Y que sobre cualquier otro, lo que acrecienta su interés para estudios de evolución humana (Roewer *et al.*, 1996a, Underhill *et al.*, 1996).

Otro de los argumentos hace referencia a la falta de recombinación del cromosoma Y. Debido a este hecho, si el proceso de selección actuase sobre alguno de los genes del cromosoma, se produciría una selección global de todo el haplotipo que portase dicho cromosoma. El haplotipo en concreto aumentaría mucho de frecuencia y podría llegar a fijarse, lo que provocaría una bajada de los niveles de polimorfismo global del cromosoma Y en la población. De todas maneras, para que un barrido selectivo de este tipo se pueda detectar en el presente, tendría que haber sucedido bastante recientemente, ya que si nos remontamos a los comienzos de la historia evolutiva de la especie, la mutación ya habría tenido tiempo de regenerar los niveles de polimorfismo.

El análisis de las primeras secuencias de ADN mitocondrial (ADNmt) (Anderson *et al.*, 1981) tuvo una gran repercusión en el estudio de las relaciones filogenéticas entre los diferentes grupos humanos actuales. El hecho de que una molécula se herede exclusivamente por vía materna y sin sufrir recombinación, permite preservar la información de los sucesos mutacionales que han surgido en un linaje mitocondrial determinado en las generaciones pasadas. Pero el ADNmt no es la única molécula que tiene un gran potencial filogenético. El cromosoma Y (Figura 2), portador de los genes determinantes del sexo masculino, también se hereda uniparentalmente (en este caso por vía paterna) y no padece fenómenos de recombinación en prácticamente toda su extensión (Wolf *et al.*, 1992).

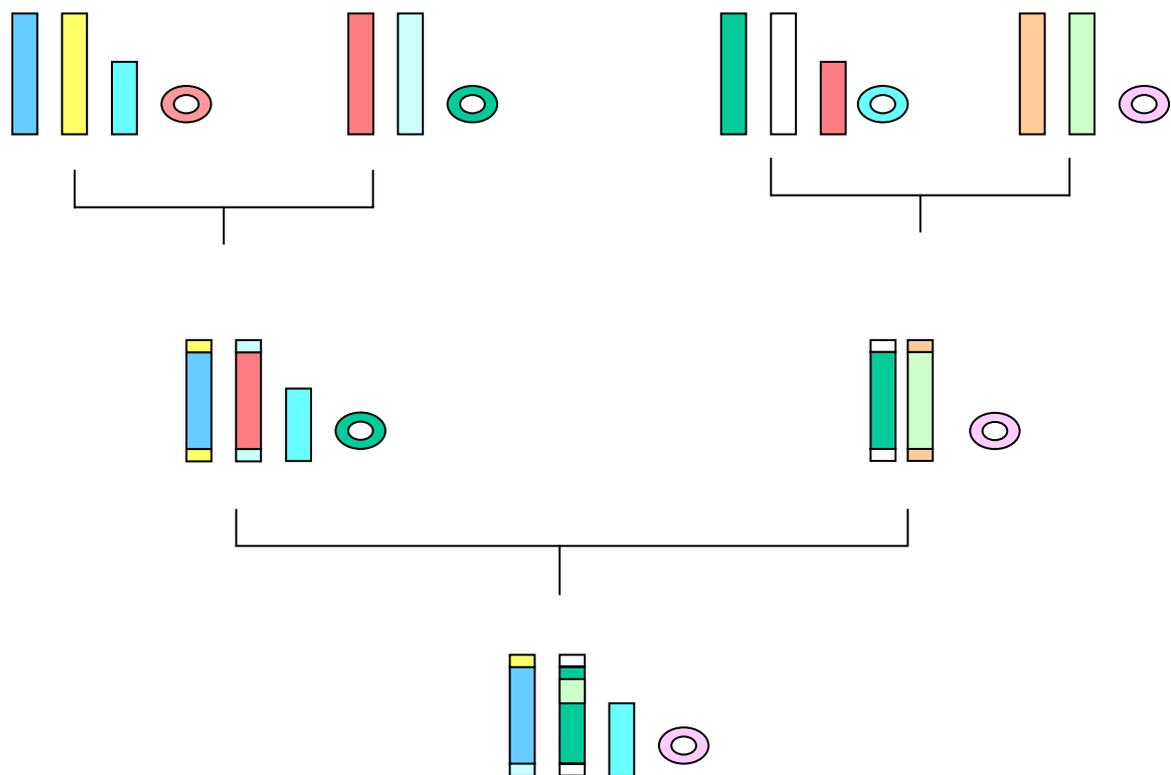


Figura 2: Esquema de la transmisión diferencial de un cromosoma Y, una molécula de ADNmt y un par de autosomas.

Esta falta de recombinación hace que el cromosoma se comporte como un gran fragmento único de ADN, portador de un haplotipo que se transmite intacto de padres a hijos, a no ser que actúen fenómenos de mutación.

El ADNmt presenta una tasa de sustitución de bases diez veces superior a la del ADN nuclear, lo que significa que existe un gran número de secuencias diferentes de ADNmt en la población, por tanto, una gran diversidad. Al mismo tiempo, esta fuerte tasa de mutación también es causa de inconvenientes: mutaciones puntuales se pueden producir varias veces en el mismo punto o aparecer independientemente en diferentes linajes, haciendo imposible determinar el estado ancestral. Por el contrario, algunas mutaciones puntuales en el cromosoma Y se pueden considerar como acontecimientos únicos que son específicos de un linaje,

ya que cuando un grupo de cromosomas Y portan una mutación puntual dada es muy probable que tengan el mismo origen.

2.6.- HAPLOGRUPOS.

Según los distintos alelos presentes en cada uno de los marcadores (haplotipo), pertenece a un haplogrupo determinado. Los principales haplogrupos de la zona que hemos estudiado, tal como se pueden observar en la figura 3, son:

- ✚ Haplogrupo E: Este haplogrupo se divide a su vez en tres subgrupos denominados E1, E2 y E3. E1 y E2 se encuentran exclusivamente en África. E3 se divide a su vez en E3a y E3b, pero sólo E3b tiene frecuencias significativas en Europa, en la mitad este y oeste de Asia, además de grupos minoritarios de blancos dentro de África. E3b es uno de los más comunes en Etiopía, Somalia, norte de África y Arabia, y también es la tercera más común en los haplogrupos de cromosoma Y de Europa. Esto se atribuye a las antiguas migraciones del este de África a través del norte de África en el periodo Neolítico.
- ✚ Haplogrupo G: Este haplogrupo tiene una baja frecuencia, pero su distribución es en el continente Europeo, Oeste de Asia, noroeste de África, Asia Central y sur de Asia. Se divide en dos subgrupos G1 y G2.
- ✚ Haplogrupo I: Este haplogrupo es un haplogrupo Europeo, representa cerca de 1 de cada 5 individuos de la población. Este haplogrupo prácticamente no existe fuera de Europa. Se piensa que se originó en los Balcanes durante la época glacial y posteriormente con la recolonización se extendió por Europa. Los subgrupos en los que se divide este haplogrupo son I1, I2a, I2a1 y I2b (I2b1, I2b2, I2b3 y I2b4). El I1 es típico de la región Escandinava y noroeste de Europa. El I1b es típico de la zona Este de Europa.
- ✚ Haplogrupo J: Este haplogrupo es el más cercano al haplogrupo I. Se subdivide en haplogrupo J1 y J2. El haplogrupo J1 aparece en una alta frecuencia en la mitad este y norte de África y Etiopía. El J2 es común en Turquía, Irak, Siria, Armenia, Italia, zona Mediterránea, Iran, Asia Central y

sur de Europa. Tanto el J1 como el J2 se subdividen a su vez en varios grupos.

- ✚ Haplogrupo K: Se piensa que el haplogrupo K es el ancestro común para la gente que vive en el hemisferio norte, incluyendo la mayor parte de europeos, asiáticos y nativos americanos. Este haplogrupo desciende del haplogrupo F. Este haplogrupo se subdivide en cuatro grupos a su vez, que son, K1, K2, K3 y K4, que se encuentran en baja frecuencia en India, Oceanía y Australia. La forma de los subgrupos K2 y K5 fueron renombrados como los haplogrupos T y S, los subgrupos K1 y K7 fueron reasignados a nuevos subgrupos M2 y M3 (dentro del haplogrupo M), y las formas de los subgrupos K3, K4 y K6 fueron renombrados como K1, K2 y K3. (Karafet *et.al.* 2008).
- ✚ Haplogrupo L: Este haplogrupo está asociado con el sur de Asia. Se encuentra en baja frecuencia en la población de Asia Central, Suroeste de Asia, Norte de África y Sur de Europa a lo largo de la costa del Mar Mediterráneo. Desciende del haplogrupo K. La presencia del haplogrupo L es bastante raro en grupos nativos (Cordaux *et.al.* 2004, Sengupta *et.al.* 2006, Thamseem *et.al.* 2006), lo cuál indica que el haplogrupo L no era original de la población paleolítica de India.
- ✚ Haplogrupo Q: Se cree que este haplogrupo tiene su origen en Siberia. Este haplogrupo contiene el ancestro patrilineal de muchos Siberianos, centro asiáticos e indígenas americanos. Una migración de Asia a Alaska por el Estrecho de Bering llevó este haplogrupo a tierras americanas, donde ahora se encuentra en una proporción muy alta. Los subgrupos en los que se divide son Q1, Q2, Q3 (típico de los indígenas americanos), Q4, Q5, y Q6.
- ✚ Haplogrupo R: Se cree que se originó en algún lugar del norte de Asia. Este haplogrupo se subdivide en R1 y R2, a su vez R1 está formado por R1a y R1b. La mayoría de los miembros del haplogrupo R, pertenecen al haplogrupo R1, y, es muy común en Europa y oeste de Eurasia. El R1 se concentra en el oeste de Europa con una frecuencia del 70%, se encuentra especialmente en Irlanda, España y Países Bajos.

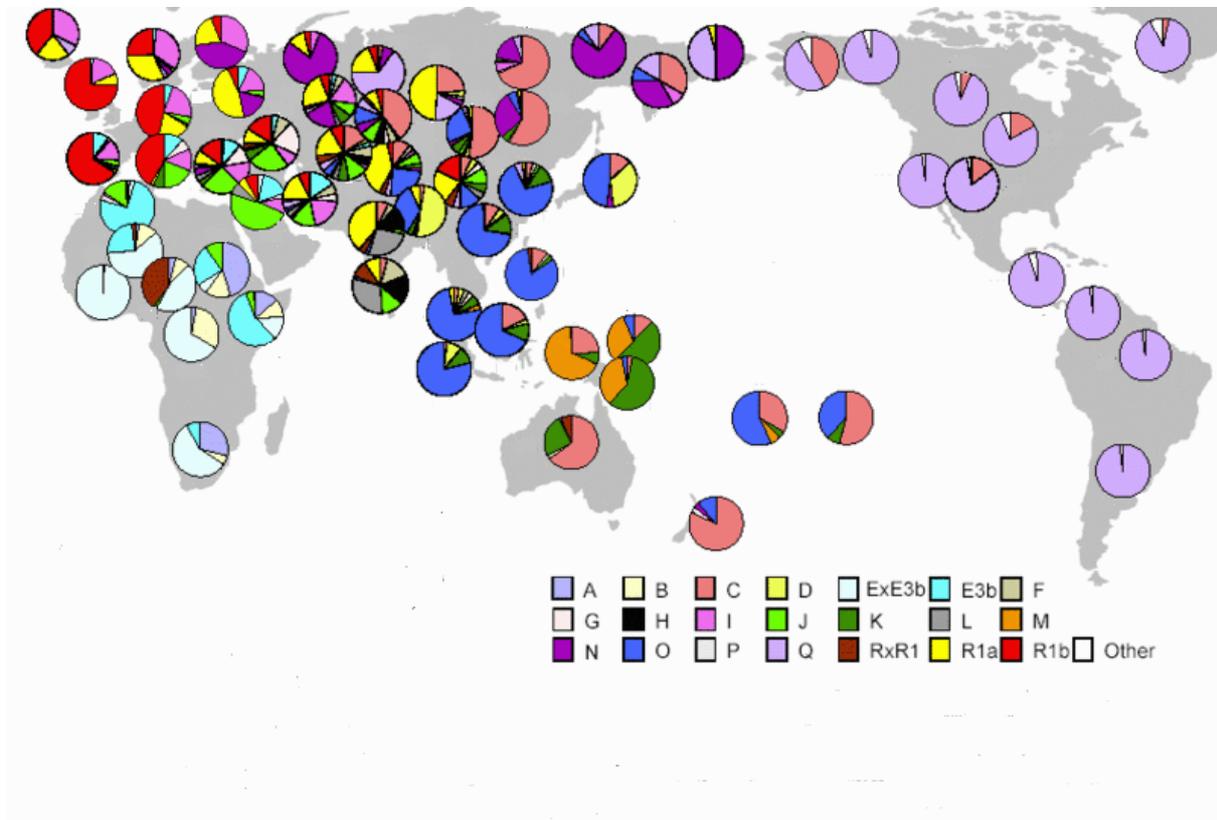


Figura 3: Distribución mundial de los haplogrupos.

2.7.- INTRODUCCIÓN A LA ANTROPOLOGÍA.

Los polimorfismos genéticos (PG) son características genéticas de las poblaciones que la antropología utiliza como instrumento para describir cuantitativamente las relaciones entre individuos de una misma población y entre poblaciones diversas. Por lo tanto, los genes que forman parte de esta categoría son aquellos para los cuales en las poblaciones están presentes variantes individuales debidas a dos o más formas alternativas (alelos): la existencia de estas diferencias es a lo que se denomina polimorfismo.

La utilidad de un PG como marcador antropológico está altamente correlacionada con el grado de diversidad de los individuos respecto a ese

fragmento de genoma y a su distinta distribución en los diversos grupos humanos. En particular, ese PG debe satisfacer algunos requisitos elementales: el primero de los cuales es que se haya originado una sola vez, es decir, que sea monofilético. Esto es posible sólo en el caso de que la frecuencia de los eventos moleculares que hayan generado los distintos alelos sea tan pequeña que haga irracional suponer que se haya instaurado más de una vez. Estos polimorfismos, llamados únicos, son los instrumentos ideales para las reconstrucciones filogenéticas. En efecto, encontrar el mismo polimorfismo en dos o más poblaciones distintas permite inferir que éste sea idéntico por descendencia, es decir, que existía ya en una población ancestral común, y no que se haya originado por mutaciones independientes.

Los PG recurrentes deben, sin embargo, ser utilizados con una cierta cautela debido a su posible polifiletismo: éstos pueden haber aparecido de la misma forma más veces y en más lugares independientemente. En este caso, la interpretación de las relaciones de descendencia entre las formas alélicas iguales o similares en estado es necesariamente más compleja. Los ejemplos más típicos de polimorfismos que pueden entrar en esta categoría son los maláricos, como la HbS y la α -talasemia.

Otro requisito fundamental para el estudio de perfil antropológico es que el PG sea selectivamente neutro. Sólo en este caso las informaciones que aportan podrán ser utilizadas para comprender la historia evolutiva (relaciones de antecesor-descendiente), y no los factores ecológicos (adaptación ambiental) de las poblaciones.

La última característica a considerar en la elección de un polimorfismo para llevar a cabo estudios destinados a analizar la variabilidad dentro de nuestra especie es su capacidad informativa o su grado de variabilidad entre poblaciones. En otras palabras, el marcador será informativo cuando su variabilidad interpoblacional (politipia) sea superior a la intrapoblacional (polimorfismo) Rickards *et.al.* 1998.

Los PG del ADN se pueden subdividir respecto a la sede topográfica y funcional del sitio y a su grado de variabilidad Rickards *et.al.* 1998. Respecto a la sede topográfica, los PG se pueden dividir en: mitocondrial y nuclear, que a su vez se divide en autosómico y heterocromosómico (del cromosoma X y del cromosoma Y). Respecto a la sede funcional pueden ser: PG de regiones intergénicas y alejados de los genes estructurales

entre los cuales están interpuestos; como es el caso de muchos STRs, y de las secuencias adyacentes implicadas en su expresión.

En cuanto al grado de variabilidad, existen PG de elevada variabilidad intrapoblacional (caracterizados por un grado de heterocigosidad H muy elevada) y aquellos de alta variabilidad interpoblacional.

Entre los PG de alta variabilidad intrapoblacional se pueden identificar dos clases:

- ✚ Secuencias con alta variabilidad haplotípica (presencia de muchos loci variables estrechamente asociados), entre los cuales está el cluster de los genes estructurales de los antígenos de histocompatibilidad (HLA).
- ✚ Sitios polimórficos puntuales constituidos por una secuencia repetida en tándem n veces. En este caso, se trata de variabilidad alélicas en sentido estricto, y los alelos difieren entre sí por el número de repeticiones (Ej. STRs).

Los polimorfismos de los microsatélites son marcadores particularmente ilustrativos. En efecto, son muy abundantes en el genoma humano y altamente informativos porque presentan numerosos alelos, algunos de los cuales presentan una distribución geográfica específica; son además fáciles de tipificar mediante PCR y de interpretar. Se trata generalmente de loci con una secuencia de 2-5 nucleótidos repetidos en tándem hasta una docena de veces y más. La elevada tasa de mutación, que en aquellos constituidos por repeticiones de 2-3 bases puede llegar a ser de 1.2×10^{-3} , hace más rápida su evolución respecto a las secuencias de ADN no repetitivas. Los distintos alelos de los microsatélites son debidos esencialmente a mutaciones que provocan variaciones del número de unidades repetidas.

Los microsatélites, cuyos alelos son definidos por el distinto número de repeticiones, es muy alta la probabilidad de que los alelos de igual longitud vengán generados repetidamente a través de modificaciones sucesivas y sean, por lo tanto, polifiléticos. Como consecuencia, la información obtenida de estos marcadores no resulta siempre simple y unívoca a causa de posibles fenómenos de homoplasia, que pueden ser verificados en el curso de las generaciones. Es decir, mientras es razonable sostener que poblaciones con repertorios micro- y minisatélites

muy diversos sean distintas, no se puede concluir que cuadros moleculares similares impliquen estrechas relaciones evolutivas, puesto que la semejanza genética podría ser el resultado de fenómenos de convergencia evolutiva.

Para el estudio de las relaciones entre poblaciones es particularmente útil construir haplotipos combinando diferentes formas alélicas de polimorfismos diversos (bien sean polimorfismos únicos, bien sean recurrentes) y analizar las frecuencias relativas de estos haplotipos en las diversas poblaciones. En efecto, si bien existe un gran riesgo de homoplasia para un simple locus, la probabilidad de que tal fenómeno se repita en más loci disminuye cuanto mayor sea el número de éstos considerados.

El ADN del cromosoma Y (ADN-Y), que constituye, aproximadamente, el 2% del genoma haploide, representa la contrapartida masculina al ADNmt. Se trata de una larga molécula lineal de aproximadamente 60Mb, cuya secuencia es casi totalmente conocida, salvo las partes relativamente largas de ADN no codificante y las secuencias repetitivas dispersas a lo largo del cromosoma, y colocadas en tándem. Aunque dos pequeñas porciones del cromosoma Y, situadas en sus extremos y denominadas regiones pseudoautosómicas, son homólogas a otras del cromosoma X (y pueden recombinarse con ellas), el resto del cromosoma es haploide y se transmite por vía paterna. Estos rasgos específicos del Y están en su mayor parte, situados hacia la región heterocromática del centromero y en la zona distal del brazo.

Los primeros estudios sobre la variabilidad del cromosoma Y, destinados a identificar polimorfismos por lugares de restricción, no permitieron evidenciar variaciones apreciables: el número de polimorfismos encontrados era extremadamente bajo respecto al de otras partes del genoma (Ngo *et.al.*, 1986; Jakubiczka *et.al.*, 1989). Sin embargo, el cromosoma Y está constituido principalmente por secuencias repetidas, probablemente no está sujeto a vínculos funcionales y por lo tanto es más libre de acumular mutaciones (Makova y Li 2002).

En los años siguientes, el descubrimiento de nuevas tecnologías de análisis de ADN ha aumentado de forma progresiva el espectro de polimorfismos individuales, de modo análogo a lo que ha sucedido para los polimorfismos autosómicos y mitocondriales. De todos estos estudios se ha concluido que la tasa de mutación del cromosoma Y es

considerablemente más baja respecto a la de otros cromosomas nucleares y del ADNmt. Esto es debido a distintas causas: El diferente número de cromosomas autosómicos y de cromosomas Y en una población, la estructura matrimonial de las poblaciones humanas y los probables efectos selectivos debidos a hitch-hiking.

Este bajo nivel de variabilidad hace que los eventos mutacionales sean tan raros que pueden ser considerados como únicos, y, por lo tanto, específicos de una línea, resultando así óptimos marcadores monofiléticos.

La invención de la PCR ha ampliado enormemente la posibilidad de investigar polimorfismos debidos a pequeñas inserciones/delecciones, minisatélites y microsatélites. Actualmente, la automatización de la secuenciación del ADN ha hecho extremadamente más simple y rápido este proceso, consiguiendo poner en evidencia muchos polimorfismos representados por simples sustituciones nucleotídicas (Underhill *et.al.*, 1997).

Hoy sabemos que el cromosoma Y presenta diversos tipos de polimorfismos, cada uno de los cuales está caracterizado por una tasa de mutación propia y por ello adaptados a estudios evolutivos de diversa escala temporal: reordenaciones cromosómicas (delecciones, inserciones, inversiones y duplicaciones), polimorfismos asociados a la presencia de repeticiones en tándem de breves secuencias (mini- y microsatélites) y simples sustituciones nucleotídicas.

Hasta hace poco tiempo, las diferencias genéticas que se utilizaban como marcadores antropológicos se limitaban a aquellas variaciones que se ponían de manifiesto a través de las reacciones antígeno-anticuerpo o por electroforesis de las proteínas de la sangre. Hay que tener en cuenta que, éstas representan una pequeña parte de la variabilidad presente y no permiten una reconstrucción detallada de las relaciones genéticas entre las poblaciones. Con este tipo de estudios no se analizan directamente los genes, sino su producto, perdiéndose así información, bien porque se concentra sólo en una parte del genoma, la de los genes estructurales (en algunos casos sólo de sus partes codificantes), bien porque no todas las mutaciones a cargo de los exones se reflejan en variaciones aminoacídicas.

Los notables progresos que se han producido en la últimas décadas en las técnicas moleculares, como el descubrimiento de las tecnologías del

ADN recombinante, la clonación genética, el descubrimiento de las enzimas de restricción y, sobretodo, la invención de la reacción en cadena de la polimerasa, con la consecuente posibilidad de secuenciar directamente los ácidos nucleicos, han permitido resolver estos inconvenientes. A partir de los años 80, ha sido posible analizar directamente el genotipo de los individuos y el ADN, poniendo en evidencia un elevado grado de variabilidad. Cuando se comparan individuos distintos para una secuencia nucleotídica de un gen o para cualquier segmento del genoma, se observan variaciones polimórficas con una frecuencia que oscila de uno a pocos nucleótidos mutados por cada 100-1000 pares de bases.

2.8.- BASES DE DATOS GENÉTICAS.

Los avances científicos y técnicos han generado y generan una cantidad de información extraordinaria, la cual ha de almacenarse de modo racional y ordenado para su posterior uso. Surge así la necesidad de las denominadas "bases de datos", lugares donde se almacena (usualmente en ordenadores o informatizadas) de modo ordenado y coherente cualquier tipo de información, que luego es rescatada de modo automático de acuerdo a parámetros previamente establecidos. El acceso a las mismas es más o menos restringido, dependiendo de la trascendencia de los datos almacenados, siendo los más protegidos aquéllos que contienen información sobre la "intimidad" de las personas.

En el ámbito científico nadie duda que una de las más importantes vías de conocimiento es el que deriva del manejo apropiado de los datos que se van generando en diversas investigaciones.

Por ello, nadie puede dudar que el acúmulo de datos es necesario, y nadie puede acotar la libre generación de los mismos, siempre que sean útiles a la sociedad en general y a las personas en particular.

CLASIFICACIÓN DE LAS BASES DE DATOS.

Básicamente son dos los criterios que pueden usarse para clasificar las bases de datos:

- ✚ Clasificación según el contenido: las bases de datos pueden contener sólo datos alfanuméricos, ADN extraído o material biológico. Como tal, bases de datos son sólo las primeras, es decir, aquéllas que en un soporte de papel o informático, contienen letras y números que identifican a una persona de entre todas las demás. Agrupaciones de muestras biológicas y de ADN ya extraídos son realmente “archivos biológicos”.
 - Bases de datos de identificación genética: corresponden al primer grupo y contienen números y letras asociados al código de identificación de una persona.
 - Archivos de ADN: ya no son bases de datos como tales, en tanto en cuanto no acumulan “datos” a los que se puede acceder de modo automatizado, sino que contienen muestras de ADN ya extraído del núcleo celular y de las mitocondrias, dispuesto para ser analizado (en sus regiones codificantes y no codificantes) con la técnica o técnicas adecuadas.
 - Archivos de muestras biológicas: tampoco son “bases de datos” en si, ya que representan una especie de “escalafón inferior” en comparación con los archivos de ADN. Cualquier laboratorio forense o de análisis clínicos, hospital, sanatorio, etc., almacena muestras biológicas (sangre, orina, biopsias) de modo más o menos controlado. A las mismas se puede o podría acceder y proceder a obtener ADN para realizar posteriores análisis de uno u otro tipo.
- ✚ Clasificación según su finalidad: pueden ser generales, profesionales y judiciales o forenses.
 - Generales: comprenderían a toda una población o país, y en este momento no hay ninguna en marcha en el mundo.
 - Profesionales: incluyen a profesionales de riesgo, siendo el principal y mayor de los ejemplos el repositorio de material biológico de las fuerzas armadas de los Estados Unidos. Realmente es un archivo biológico cuyas muestras han sido obtenidas con consentimiento y sólo se analizan en caso de necesidad.

- Judiciales o Forenses: son las que realmente nos ocupan, y a su vez pueden ser de tipo criminal o civil.
 - Bases de datos forenses criminales: almacenan datos procedentes de personas que han sido procesadas o condenadas, así como de indicios biológicos encontrados en la escena del crimen. Su característica principal es que algunas de las muestras y datos considerados se obtienen sin el consentimiento de las personas implicadas.
 - Bases de datos forenses civiles: su único fin es la identificación de personas desaparecidos, lo cual se hace comparando el ADN de las personas no identificadas (normalmente restos óseos) con el de los familiares. Por sus características humanitarias, es requerimiento ético que los familiares colaboren voluntariamente y tras firmar un consentimiento informado, hecho éste que las diferencia de las bases de datos criminales.

El ADN, componente químico del núcleo celular, se ha convertido en un instrumento esencial de las técnicas que la moderna medicina forense utiliza para la investigación de delitos por parte de las autoridades judiciales y policiales.

Desde que en 1988, en el Reino Unido y por primera vez, la información obtenida del ADN fuese utilizada para identificar y condenar al culpable de un delito, tanto en España como en el resto de los países de nuestro entorno se ha tomado conciencia de la trascendencia de los marcadores genéticos en las investigaciones criminales, como la identificación de cadáveres o la determinación de relaciones de parentesco.

Sin embargo, y a pesar de esa importancia, el uso de los datos relacionados con el ADN, en el ámbito de la persecución de delitos, cuenta hoy con numerosas dificultades, especialmente en lo relativo a su obtención y registro de cara a su empleo en el curso de ulteriores investigaciones. Ello viene dado tanto por el carácter sensible que dichos datos tienen y el importante grado de protección con que, naturalmente,

deben contar, como por la inexistencia de un marco jurídico que regule adecuadamente su empleo.

En el año 2003, y mediante lo dispuesto en la Disposición Final Primera de la Ley Orgánica 15/2003, de 25 de noviembre, de modificación del Código Penal, se reformó la Ley de Enjuiciamiento Criminal a fin de proporcionar cobertura jurídica, de la que carecían hasta entonces, a determinadas prácticas de investigación.

La nueva redacción dada a los artículos 326 y 363 de la Ley de Enjuiciamiento Criminal consistió, esencialmente, en regular la posibilidad de obtener el ADN a partir de muestras biológicas provenientes de pruebas halladas en el lugar del delito o extraídas de sospechosos, de manera que dichos perfiles de ADN puedan ser incorporados a una base de datos para su empleo en esa concreta investigación.

Sin embargo, la reforma no contempló otros aspectos importantes, como la posibilidad de crear una base de datos en la que, de manera centralizada e integral, se almacenase el conjunto de los perfiles de ADN obtenidos, a fin de que pudiesen ser utilizados, posteriormente, en investigaciones distintas o futuras, incluso sin el consentimiento expreso del titular de los datos.

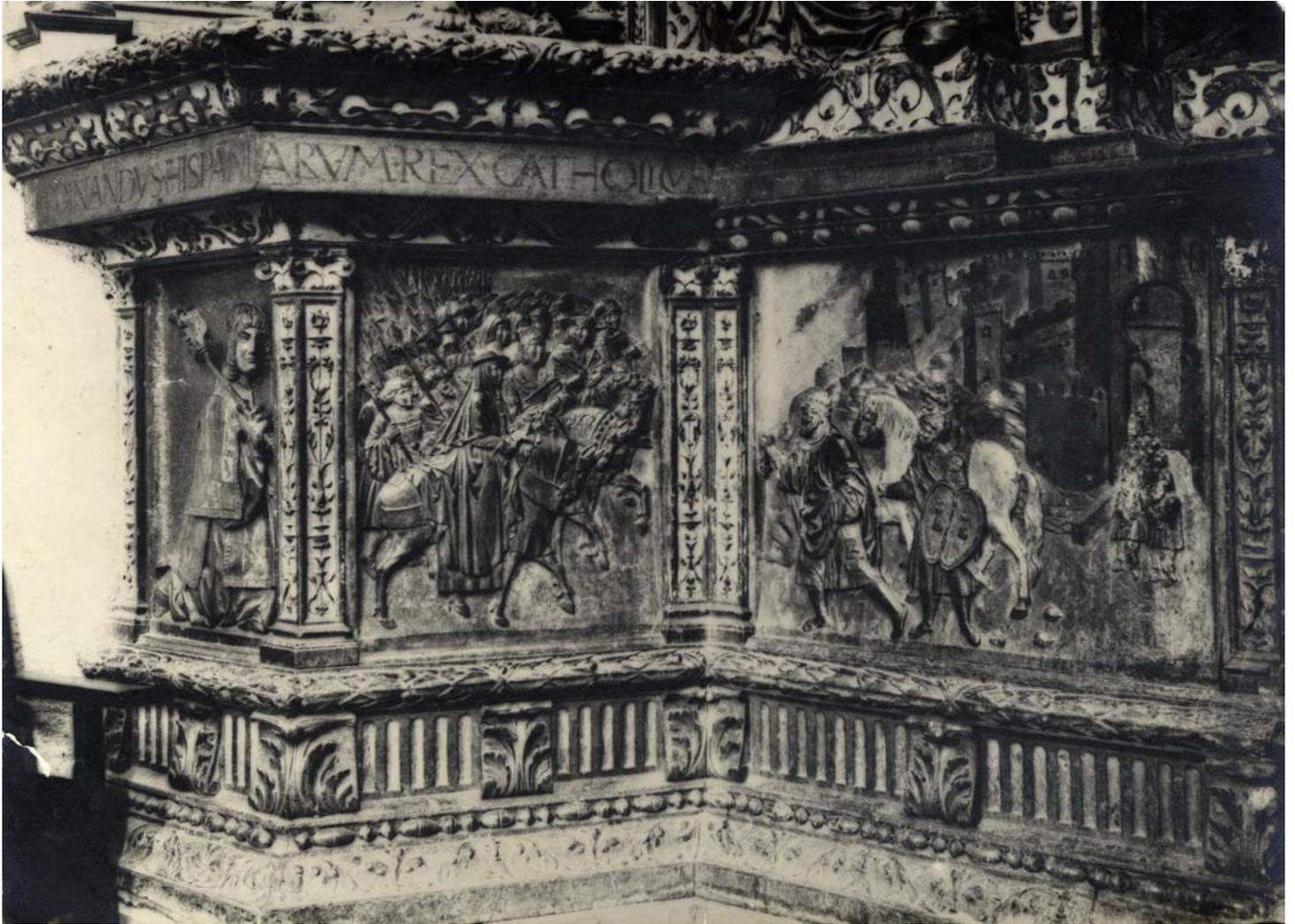
La creciente globalización de los delitos y la paralela asunción por parte de España de una serie de obligaciones recíprocas con otros países para compartir la información disponible en los respectivos ficheros y bases de datos, exigen la adopción de las medidas materiales y jurídicas adecuadas. La adopción de esas medidas jurídicas, así como la creación de bases de datos que permitan intercambiar la información entre los Estados miembros, ha sido reiteradamente expuesta desde las Instituciones comunitarias a través de sendas Resoluciones del Consejo relativas al intercambio de resultados de análisis de ADN, de 9 de junio de 1997 y 25 de julio de 2001, respectivamente. En el mismo sentido se ha venido pronunciando el Consejo de Europa a partir de la Recomendación (92) 1, de 10 de febrero de 1992, de su Comité de Ministros, sobre la utilización de los resultados de análisis de ADN en el marco del sistema de justicia penal.

Así, el 8 de Octubre de 2007, se aprueba esta ley orgánica, en la que como no podría ser de otra manera, se han tenido en cuenta los criterios que, sobre la protección de los derechos fundamentales en la

obtención de pruebas a partir de los perfiles de ADN, ha venido conformando el Tribunal Constitucional en diversas Sentencias como la 207/1996, de 16 de diciembre.

El artículo de la presente Ley comienza determinando lo que constituye su objetivo fundamental, que no es otro que la creación de una base de datos en la que, de manera única, se integren los ficheros de las Fuerzas y Cuerpos de Seguridad del Estado en los que se almacenan los datos identificativos obtenidos a partir de los análisis de ADN que se hayan realizado en el marco de una investigación criminal, o en los procedimientos de identificación de cadáveres o de averiguación de personas desaparecidas. Además posibilita que para determinados delitos de especial gravedad y repercusión social, así como, en el caso de los patrones identificativos obtenidos en los procedimientos de identificación de restos cadavéricos o de personas desaparecidas, o cuando el titular de los datos haya prestado su consentimiento para la inscripción-, los resultados obtenidos a partir del análisis de las muestras biológicas del sospechoso, detenido o imputado, sean inscritos y conservados en la base de datos policial, a fin de que puedan ser utilizados en esa concreta investigación, o en otras que se sigan por la comisión de alguno de los delitos para los que la propia Ley habilita la inscripción de los perfiles de ADN en la base de datos.

En relación con el período de la conservación de los perfiles identificativos en la base de datos, la Ley fija unos períodos de cancelación cuya duración dependerá del tipo de delito y de la resolución judicial con que finalice el procedimiento penal.



Predela retablo Capilla Real
AUTOR: José Martínez Rioboó

Fundación Rodríguez-Acosta
Donación Martínez Sóla.

3.-MATERIAL Y MÉTODOS

3.1.- MUESTRAS.

3.1.1.- POBLACIÓN ANALIZADA.

En este trabajo se analizan individuos, no emparentados, apellidos Colón, Colóm y Colombo procedentes de Cataluña, Valencia, Islas Baleares, Sur de Francia y tres regiones Italianas (Liguria, Lombardía y piamonte). Esto se realizó gracias a la colaboración de Francesc Albardaner i Llorens miembro del Centre d'EstudisColombins de Òmnium Cultural de Barcelona, Alfonso Ballesteros Presidente de la Reial Acadèmia de Medicina de las Illes Balears, y, Olga Rickards (Dipartimento di Biologia, Università di Roma "Tor Vergata"). Para ello, se hizo un estudio sobre todos los apellidos Colon, de los que constara que tuvieran, al menos, tatarabuelos (tres generaciones) nacidos en la zona seleccionada. Entre ellos, se eligió, de forma aleatoria, una muestra de voluntarios de diferente localización geográfica que será sometida a una prueba para recoger su ADN. Así, se pusieron en contacto con los seleccionados, enviándoles una carta con los hisopos (para la recogida de muestras) y un consentimiento informado, que nos sería posteriormente remitido. De estas mismas poblaciones también se escoge una muestra de referencia, es decir, de individuos no apellidos Colón y no emparentados entre sí.



En la siguiente tabla se muestra un resumen de las poblaciones y los individuos analizados.

Poblaciones	Individuos apellidados Colón	Individuos no apellidados Colón
Cataluña	125	59
Valencia	45	
Islas Baleares	50	
Sur de Francia	35	50
Liguria	47	
Lombardía	52	
Piamonte	14	

En primer lugar, voy a analizar todas las muestras de estas poblaciones, aplicarles una serie de parámetros estadísticos a cada una de ellas, para ver su diversidad y otros datos de interés, y finalmente comparar cada una de estas poblaciones con las demás. Posteriormente, compararé estas poblaciones con los datos, obtenidos de la bibliografía, de otras poblaciones muy diversas. Así, las compararé con Austria (Berger *et al.* 2005), Bereberes (Frigi *et al.* 2006), Brasil (Pereira *et al.* 2007), Bubi (guinea ecuatorial) (Barrot *et al.* 2007), China (Han Yan *et al.* 2007), fangs (guinea ecuatorial) (Barrot *et al.* 2007), Japón (Hara *et al.* 2007), Polonia (Soltyszewski *et al.* 2007), Tibet (provincias de Qinghai, Gansu, Sichuan y Yunnan) (Zhu *et al.* 2007), Tibet-China (Zhang *et al.* 2006), Tibet Lassa (Kang Long Li *et al.* 2007), China, Malasia e India (Yuet Meng Chang *et al.* 2007) Portugal (Alves *et al.* 2007), Norte de Portugal (Pontes *et al.* 2006), Norte de Italia (Turrina *et al.* 2006) y Barcelona (Sanchez *et al.* 2007).

Las muestras analizadas pueden ser de dos tipos:

- ✚ Hisopo de saliva: A la mayor parte de los individuos se les toma una muestra de saliva mediante dos hisopos que son pasados por la cara interna de la boca. Una vez se han secado ambos hisopos, a uno de ellos se le practica una extracción con fenol/cloroformo y el otro se conserva en el congelador.
- ✚ Sangre: a parte mínima de los individuos, en este caso, se les practica una extracción de sangre periférica, depositándola en un trocito de tela estéril. Una vez seca se le va a practicar una extracción con chélex.

3.2.- EXTRACCIÓN DE ADN.

3.2.1.- EXTRACCIÓN ORGÁNICA MEDIANTE FENOL/ CLOROFORMO-ISOAMILALCOHOL.

Este tipo de extracción se realiza a las muestras de saliva depositadas en hisopos de algodón. De este modo, se corta con un escalpelo estéril la cabeza completa de algodón que se introduce en un eppendorf. Posteriormente, se añaden 300 µl de buffer de extracción (10 mM Tris, 100 mM NaCl, 39 mM DTT, 10 mM EDTA, 2% SDS), 7.5 µl de proteinada K (10 mg/ml) y 12 µl de DTT. Se agita muy bien a baja velocidad y se centrifuga en una nanofuga durante dos segundos. A continuación se incuba a 56°C toda la noche (18-24 horas).

Una vez terminada la incubación, se elimina el algodón del hisopo con una punta de pipeta, presionando contra las paredes del eppendorf para no arrastrar nada de muestra. Este paso se realiza en una campana extractora, donde también se le añaden 300 µl de fenol/cloroformo/isoamilalcohol 24:25:1 (saturado con 10 mM Tris, pH 8, 1 mM EDTA, Sigma). Se agita hasta conseguir una emulsión lechosa y se centrifuga en una microcentrífuga 3 minutos a máximas revoluciones (13.000 rpm).

Una vez separadas la fase fenólica y clorofórmica se recupera la fase acuosa o fenólica que es la contiene al ADN y se transfiere a un tubo Microcon-100 (Microcon YM-100, Millipore) al cual previamente se le ha añadido 100 µl de agua estéril.

Se centrifuga a 2.500 rpm durante 20 minutos. Posteriormente, se elimina el filtrado y se añaden 200 µl de agua estéril para limpiar la membrana, volviendo a centrifugar a 2.500 rpm durante 20 minutos. Finalmente, se añade el volumen de recuperación deseado (100 µl de agua estéril), resuspendiendo varias veces con la pipeta, se invierte el filtro en un tubo limpio y se centrifuga a 3.500 rpm durante 5 minutos. Descartar el filtro y cerrar el tubo.

3.2.2.- EXTRACCIÓN DE ADN CON CHELEX.

El "chelex 100" es una resina quelante captadora de iones polivalentes que permite extraer el ADN a partir de diferentes substratos de manera rápida y sencilla (Walsh *et al.*, 1991). Además presenta la ventaja de poder extraer ADN a partir de muestras pequeñas puesto que al ser el número de pasos reducido la probabilidad de pérdida de ADN por manipulación es extremadamente baja. Sin embargo, presenta el

inconveniente de que el ADN obtenido está menos purificado que el obtenido en la extracción orgánica con fenol/cloroformo. No obstante, es un método muy efectivo para extraer ADN a partir de muestras de sangre total que se encuentra en buen estado, por lo tanto ha sido utilizado para la extracción de ADN del grupo de muestras compuestas por sangre líquida.

El primer paso consiste en un choque hipotónico que hace que las membranas citoplasmáticas de los eritrocitos y linfocitos se lisen, para ello se toman dos cuadraditos de tela de medio centímetro, aproximadamente, de tela con la sangre y se le añade 1 ml de agua estéril, se agita durante 2 ó 3 segundos y se deja en incubación media hora a temperatura ambiente.

Posteriormente, se agita de nuevo y se centrifuga a máximas revoluciones (13.000 rpm) durante 3 minutos, tras los cuales se formará un sedimento constituido mayoritariamente por los núcleos de los linfocitos.

Se añaden 200 µl de chelex al 5% (chelex-100 Resin, BioRad) asegurándonos de tomar bolitas de resina. Se agita muy bien y se incuba a 56°C durante media hora.

Una vez transcurrida la media hora, se sacan las muestras del baño, se agitan e incuban a 100°C durante 8 minutos. Finalmente, se vuelven a agitar y se centrifugan a máximas revoluciones durante 3 minutos. Se guardan las muestras en el frigorífico a 4°C.

3.3.- CUANTIFICACIÓN DEL ADN.

La cuantificación del ADN se ha realizado mediante geles de agarosa al 0.8%. Para la preparación del gel se depositan 0.8g de agarosa en 100 ml del buffer TBE 1X (0.89 M Tris, 0.89 M ácido bórico, 20 mM EDTA), se deja que la agarosa se hidrate un poco y posteriormente se introduce en el microondas. Cuando la agarosa comienza a hervir se saca y se agita brevemente, volviendo a repetir este paso tres veces para que la agarosa se disuelva bien. Una vez la agarosa está lista, se puede hacer dos tipos de geles según tamaño:

- ✚ Gel pequeño: se miden 30 ml de agarosa y 50µl de bromuro de etidio (0.3mg/ml) mezclándolos muy bien.
- ✚ Gel grande: se miden 140 ml de agarosa y 240µl de bromuro de etidio.

El tamaño del gel depende del número de muestras que tengamos para cuantificar, ya que en un gel pequeño sólo le entran 18 muestras (incluyendo los controles que se utilizarán para saber la cantidad de ADN que tienen nuestras muestras), si el número de muestras es superior a 18 se prepara un gel grande.

Una vez tenemos ya preparado el gel, mientras este gelee, preparamos las muestras. Para ello se pone en un tubo de amplificación 2 µl de buffer de carga (50% glicerol, 0.1% azul de bromofenol, 0.1M EDTA en TE⁻⁴) junto con 5 µl de la muestra. Al principio y al final del gel se pone 100 pb para comprobar exactamente que el tamaño del ADN de las muestras es el correcto. Además, intercaladas entre las muestras se colocan controles de los que conocemos a ciencia cierta la concentración del ADN (K2 que contiene 2ng/5µl, K5 que contiene 5ng/5µl, K10 que contiene 10ng/5µl, etc...). Tanto los controles como los 100 pb se preparan igual que las muestras. Todas aquellas muestras en las que obtenemos una concentración superior al control que contiene mayor cantidad de ADN (20ng/5µl), se vuelven a cuantificar utilizando diluciones de las mismas.

Una vez tenemos ya preparado todo, comprobamos que el gel ha gelificado, lo colocamos en la cubeta donde queda totalmente cubierto por TBE 1X. Posteriormente, vamos colocando las muestras en los distintos pocillos cuya capacidad es de 7 µl (2 µl del buffer de carga y 5 µl de la muestra, control o 100 pb). Una vez hemos cargado todos los pocillos, lo dejamos correr a 100 v, 100 mA y 10 W durante 10 ó 15 minutos.

Finalmente, sólo hay que comparar las muestras con los distintos controles según la intensidad de la fluorescencia que emiten al someterlos a luz ultravioleta.

3.4.- AMPLIFICACIÓN.

Hemos analizado un total de 17 STRs, pertenecientes todos ellos al cromosoma Y masculino, mediante el Kit Y-Filer de Applied Biosystem, preparado para ser detectado por electroforesis capilar con emisión de fluorescencia. La amplificación de todas las muestras se realizó en el termociclador Biometra. A continuación se muestra cada uno de los loci amplificados con algunas características:

Color	Locus	Alelos en ladder	Rango de tamaño en pb	Control positivo	Otros alelos
Azul	DYS 456	13-18	100-125	15	<13
	DYS 389I	10-15	135-175	13	
	DYS 390	18-27	180-235	24	
	DYS 389II	24-34	245-300	29	
Verde	DYS 458	14-20	125-160	17	<14, 16.2, 17.2, 18.2, 19.2, >19 y >20 13.2
	DYS 19	10-19	165-220	15	
	DYS 385	7-25	235-320	11, 14	
Amarillo	DYS 393	8-16	95-135	13	15.1
	DYS 391	7-13	145-180	11	
	DYS 439	8-15	190-230	12	
	DYS 635	20-26	240-275	24	
	DYS 392	7-18	280-330	13	
	Y-GATA H4	8-13	115-150	13	
	DYS 437	13-17	175-210	15	
Rojo	DYS 438	8-13	215-255	12	>13
	DYS 448	17	270-330	19	

3.4.1.- KIT Y-FILER:

La amplificación se llevó a cabo siguiendo las recomendaciones del manual.

🧪 Mezcla de reacción:

STR 10X Buffer	9.20 µl
Primer Pair Mix	5 µl
AmpliTaq Gold™ (5 u/µl)	0.80 µl
H ₂ O (estéril)	X µl
DNA (0.2-1ng)	X µl
Volumen final	25.00 µl

✚ Condiciones de amplificación:

95° C	11 min.	} 30 ciclos
94° C	1 min.	
61° C	1 min.	
72° C	1 min	
60° C	1h20min.	
4° C	∞	

3.4.2.- COMPROBACIÓN DE LA AMPLIFICACIÓN.

Una vez finalizada la reacción de amplificación y antes de comenzar con el tipado de la muestra es recomendable, para ahorrar tiempo y reactivos, comprobar los productos de la amplificación mediante la migración electroforética de los mismos en geles de agarosa al 2% en solución TBE 1X y con 50 µl de bromuro de etidio (0.3 mg/ml). En cada pocillo se depositan 5 µl de producto amplificado más 2 µl de buffer de carga. Las condiciones electroforéticas que se aplican son: 10 watios, 100 mA y 100 voltios. Tras unos 30 minutos las muestras son visualizadas iluminando el gel con luz UV. De esta manera es posible saber si la amplificación ha sido positiva y, en caso afirmativo puede observarse la intensidad de amplificación, indicándonos si debemos aumentar la cantidad de producto amplificado que se va a utilizar para el tipado de las muestras.

3.5.- ELECTROFORESIS CAPILAR.

Las muestras amplificadas con el kit Y-Filer son tipadas por electroforesis capilar en un analizador genético ABI Prism 310 de Perkin Elmer. Antes de cargar las muestras en el 310 son necesarios una serie de pasos para prepararlo adecuadamente y que las muestras puedan ser analizadas sin problema.

- ✚ Establecer comunicación entre el ordenador y el 310.
- ✚ Abrir el software de recogida de datos "ABI PRISM□ 310 Collection" y permitir a través de dicho programa que el 310 reconozca la posición del distribuidor de muestras ("autosampler") y de la bomba

("syringe").

- ✚ Colocar el bloque de metacrilato que ha de estar limpio y seco.
- ✚ Colocar el capilar. Antes de colocarlo se limpia la ventana del capilar con un papel embebido en etanol absoluto, a continuación con otro embebido en agua destilada y seguidamente se seca. Tras limpiar la ventana se introduce el capilar en la válvula que le conecta con el bloque de metacrilato hasta llegar a la primera intersección y se aprieta bien la válvula; es importante no rebasar esta primera intersección ya que si no se sitúa el extremo del capilar correctamente puede impedirse que el polímero penetre en él. A continuación se sitúa la zona de lectura del capilar en la ventana del láser y se cierra la puerta del láser, se introduce el otro extremo en el bloque que contiene el electrodo negativo o cátodo y se desliza hasta que queden a la misma altura y separados 2 mm. Por último, se sujeta el capilar con una cinta adhesiva contra la placa termoestabilizadora.
- ✚ Preparar el buffer 1X. Hacer una dilución 1:10 con agua bidestilada del buffer 10X (310 Genetic Analyzer buffer) con un volumen final de 15 ml (11 ml para el vial del ánodo y 4 ml para el vial del buffer).
- ✚ Relleno del capilar y del bloque de metacrilato con el polímero. El polímero utilizado para el análisis de fragmentos es el POP-4 (Performance Optimized Polymer 4, Perkin Elmer) y debe sacarse media hora antes de su uso del frigorífico para que se atempere. Con la jeringa bien limpia se toma el volumen de polímero necesario en función del número de muestras, se enrosca bien para evitar pérdidas de polímero y se llenan los conductos del bloque de metacrilato con mucho cuidado de que no se formen burbujas de aire. Una vez colocado el capilar y relleno el bloque de metacrilato con el polímero, lo cual equivaldría a la preparación del gel en los métodos convencionales de electroforesis, el siguiente paso es la preparación de las muestras.

Para preparar las muestras se añade en un tubo de 0.5 ml los siguientes volúmenes por muestra:

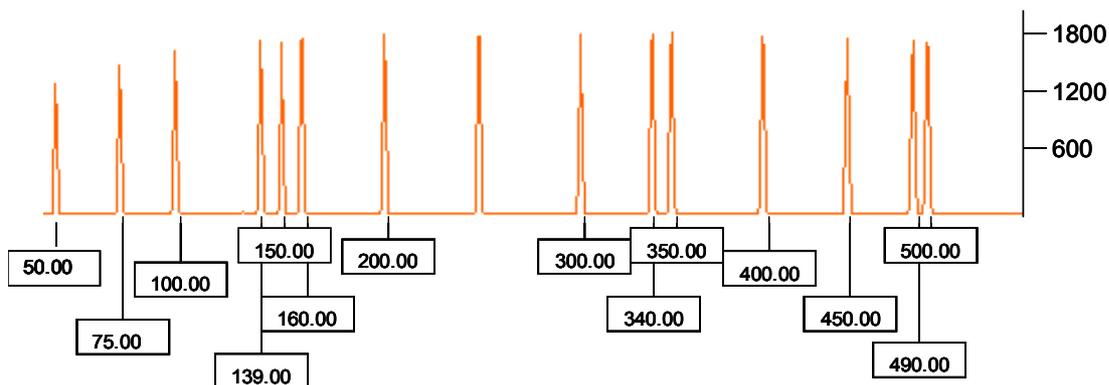
- ✚ 24 μ l de formamida desionizada (Ultra Pure Grade, Amresco).
- ✚ 0.5 μ l de estándar de tamaño (GeneScan-500 [LIZ] Size Standard, Perkin Elmer).
- ✚ 1 μ l de producto amplificado (o de ladder Y-Filer).

Al llevar cada muestra un estándar de tamaño se evitan errores en la asignación del tamaño de los fragmentos analizados debidos a diferencias en la movilidad electroforética entre las distintas inyecciones (bien por variaciones de temperatura, de concentración de polímero o en las condiciones de electroforesis). Con el estándar utilizado es posible analizar fragmentos de entre 35-500 pares de bases. En total consta de 16 fragmentos de las siguientes longitudes:

	139	250	400
50	150	300	450
75	160	340	490
100	200	350	500

El pico de 250 se deja sin asignar y su único papel es indicar las posibles variaciones entre los diferentes procesos electroforéticos. Los distintos standard de tamaño se pueden ver en la siguiente figura:

GS500 LIZ (APPLIED)



Una vez preparadas las muestras éstas se desnaturalizan a 96°C durante 3 minutos, se ponen en hielo y se coloca cada tubo en su posición adecuada dentro del soporte que va acoplado al distribuidor de muestras.

✚ Antes de iniciar la electroforesis una vez que las muestras han sido preparadas es necesario establecer una serie de parámetros de análisis en el programa de GeneScan. Entre los parámetros de mayor interés y que pueden afectar más a la hora de interpretar los resultados se encuentran, por un lado, el rango de análisis, y por otro la intensidad de los picos. En el rango es importante especificar el rango de tamaños (en pares de bases o en puntos de datos) de los fragmentos que se van a analizar, por lo tanto este parámetro deberá de ser programado en función de los marcadores que se vayan a analizar.

Para la intensidad de picos, normalmente se recomienda 150 RFU para evitar picos de menor intensidad que podrían corresponder a un ADN extraño. No obstante, cuando los datos van a ser genotipados con el Genotyper se recomienda bajar la intensidad hasta 50 RFU para que este programa pueda identificar señales bajas de manera automática.

Para preparar la hoja de muestras es necesario abrir el programa ABIPrism 310 Collection. Una vez abierto, se le indica que queremos analizar cinco colores y se procede a rellenar una nueva hoja con el nombre de las muestras en función de su colocación en el tray. Para cada muestra se requiere especificar el color del estándar, en la columna "Std" de la hoja aparecerá un rombo junto al color elegido para el estándar (en nuestro caso en el que las muestras fueron amplificadas con el kit Y-Filer el color reservado para el estándar es el naranja). A continuación se completa la columna "Sample Info" en la que se anotan los marcadores analizados y se guarda la hoja de muestras en la carpeta "Sample Sheets".

En esta lista se incorpora la hoja de muestras creada y grabada anteriormente y en ella es posible modificar el orden de análisis de las muestras e incluso repetir algunas de las muestras cambiando alguno de los parámetros de análisis como el tiempo de inyección. En la segunda columna se especifica el módulo de análisis el cual contiene las funciones específicas ejecutadas para el proceso de análisis de las muestras. En el caso de análisis de fragmentos el módulo seleccionado es el GS STR POP-4 G5v2(1 ml). Una vez seleccionado dicho módulo se establecen automáticamente unos valores para los siguientes parámetros:

Tiempo de inyección 3 seg.
Voltaje de inyección 15 Kv.
Voltaje de electroforesis 15 Kv.
Temperatura de electroforesis ... 60°C.
Tiempo de electroforesis 28 min.

A pesar de preestablecer unas condiciones de electroforesis es posible modificarlas mientras el instrumento trabaja y siempre que las muestras no hayan sido procesadas.

✚ **Voltaje y tiempo de inyección** (para inyección electrocinética): ambos parámetros pueden ser variados para regular la cantidad de ADN que entra en el capilar. Así, en el caso de obtener una señal débil en una muestra es posible aumentar el tiempo de inyección para permitir que se inyecte más cantidad de producto amplificado al capilar y que aumente la señal de fluorescencia. No obstante, un tiempo de inyección excesivo puede originar una pérdida de resolución.

✚ **Voltaje de electroforesis:** para fragmentos analizados utilizando el polímero POP-4 el voltaje es de 319 V/cm. Para un capilar de 47 cm esto supone un voltaje total de 15000 V y la corriente a este voltaje es de 8-9 μ A. Bajo estas condiciones electroforéticas el tiempo necesario para que un fragmento de 400 pb llegue a la ventana detectora es de unos 25 min. Si se aumenta el voltaje, el tiempo necesario es menor pero se pierde resolución.

✚ **Temperatura de electroforesis:** para secuenciación esta temperatura debe ser de 50°C y para fragmentos de 60°C.

✚ **Tiempo de electroforesis:** puede acortarse cuando sólo se necesita información de productos de amplificación de pequeño tamaño. Normalmente, el tiempo de electroforesis programado es un 10% superior a la media del tiempo necesario para que el fragmento de mayor tamaño migre lo suficiente.

Una vez completada la lista de inyección y habiéndose alcanzado la temperatura de inyección, la electroforesis se inicia pulsando el comando "run".

ANÁLISIS DE LOS RESULTADOS:

✚ **Análisis de los resultados con el software GeneScan:** Una vez recogido el dato bruto o "raw data", el software GeneScan procesa la información para proporcionar una cuantificación en pares de bases de todos los picos. Como cada muestra lleva un estándar interno éste migra a la vez que el ADN amplificado y bajo las mismas condiciones, con lo que se evita el problema de las variaciones entre muestras.

El estándar de tamaño se define una vez que la electroforesis ha terminado y antes de que el programa GeneScan comience el análisis de

las muestras. Para dicha definición existen dos posibilidades: definir un estándar nuevo o bien utilizar un estándar utilizado previamente en otros análisis. Un vez definido el estándar se comprueba que en todas las muestras coincidan los picos del mismo, lo cual se hace solapando todos los estándares. Si en alguna muestra los picos no coinciden se define uno nuevo. En nuestro caso utilizamos un estándar que había sido usado en análisis anteriores y definiéndolo sólo en caso necesario.

Tras esto, es necesario definir los parámetros de análisis, que fueron los que el programa tiene asignados automáticamente para el Y-Filer y se procedió al análisis de las muestras. El método por el que el programa determina los tamaños de los fragmentos se denomina Southern Local ("Local Southern Method"). Dicho método utiliza tres puntos de la curva generada con los puntos del estándar: los dos que están por debajo en tamaño del fragmento a determinar y uno por encima, con dichos valores realiza una interpolación y asigna el valor del tamaño.

Los datos analizados pueden ser visualizados de forma gráfica (electroferograma), de forma numérica o bien de ambas formas. Estos datos fueron exportados al programa Genotyper.

 **Asignación de los alelos:** El programa Genotyper es el encargado de asignar los alelos a los fragmentos analizados en el GeneScan por comparación con los alelos del ladder. Cuando se importa un archivo desde GeneScan a Genotyper éste último genera una línea (dye/lane list) para cada fluorocromo en la que están incluidos el tamaño, cantidad y la información de todos los fragmentos marcados con un mismo fluorocromo.

Según la aplicación Genotyper contiene una serie de "macros" que son secuencias de comandos predefinidos necesarios para un determinado tipo de análisis. Para fragmentos amplificados con el kit Y-Filer el macro seleccionado fue el "Kazam". Una vez ejecutado dicho macro aparecieron las muestras con los correspondientes alelos asignados de manera automática. Aquellos alelos que se encontraron por encima o por debajo de los alelos del ladder en un rango de 0.5 pares de bases no fueron designados por el programa.

3.6.-ANÁLISIS ESTADÍSTICO.

El análisis de la variación genética de individuos contemporáneos ha sido muy útil en la actualidad para conocer aspectos importantes del pasado de los humanos. Los estudios de diversidad genética humana están, normalmente limitados por un único periodo de tiempo, tal como el

análisis de diversidad de poblaciones modernas, de las cuales debemos reconstruir el pasado. Así, la diversidad humana actual debe estar influenciada por procesos históricos y prehistóricos. Antropológicamente, existe un gran interés en las poblaciones prehistóricas, sus orígenes, movimientos y demografías.

La comparación de la diversidad genética entre poblaciones humanas continentales ha sido muy útil para la construcción de modelos alternativos sobre el origen de las poblaciones humanas modernas. Un aspecto muy interesante, es que los humanos parecen tener una diversidad nucleotídica inferior a los simios, a pesar de que su tamaño poblacional es muy superior.

Unos bajos niveles de diversidad genética en determinadas poblaciones puede ser fruto de una inmigración limitada, un tamaño poblacional pequeño o una presión selectiva contra un determinado grupo de alelos. Mientras que, un alto nivel de diversidad genética, puede ser el resultado de una gran inmigración, un tamaño poblacional considerable y unos procesos selectivos que favorezcan un aumento de la diversidad genética.

3.6.1.- ÍNDICES DE DIVERSIDAD.

Se han calculado un conjunto de índices de diversidad, todos ellos computados con el programa Arlequín Ver. 2.000 (Schneider *et al.* 2000).

El objetivo de Arlequín es proporcionar el medio, en genética poblacional, con un amplio conjunto de métodos y test estadísticos para poder obtener toda la información sobre aspectos genéticos y demográficos, en un conjunto de muestras poblacionales. Así pues, con Arlequín podemos tratar un amplio abanico de datos, como pueden ser:

- ✚ Secuencias de ADN.
- ✚ RFLP
- ✚ Microsatélites
- ✚ Frecuencias alélicas
- ✚ Datos Standard: datos para los cuales la base molecular de los polimorfismos no están particularmente definidos o cuando las diferencias alélicas son consideradas como mutaciones.

Los análisis de los datos con Arlequín pueden encontrarse dentro de dos categorías: métodos intra-poblacionales y métodos Inter.-poblacionales. Como nuestro trabajo consta de datos obtenidos de siete poblaciones (más dos de referencia) y posteriormente la hemos

comparado con 18 poblaciones más obtenidas de la bibliografía, podemos realizar los análisis de ambas categorías.

3.6.1.1.- MÉTODOS INTRA-POBLACIONALES.

Los métodos intra-poblacionales a los que he sometido mis datos son los siguientes:

✚ **Diversidad genética:** La frecuencia de un alelo en una población se denomina frecuencia génica o alélica y se puede definir como el cociente resultante de dividir el número de alelos iguales en una población entre el número total de alelos, cuyo error se calcularía:

$$S = \sqrt{p_i (1 - p_i)} / n$$

Siendo p_i la frecuencia alélica y n la población.

Mientras que la diversidad alélica o genética se define como la probabilidad de que dos alelos escogidos al azar sean diferentes, y, se calcula:

$$1 - \sum p_i^2$$

Siendo p_i la frecuencia alélica.

Una alta frecuencia alélica es de esperar en una población antigua, ya que se requiere tiempo para incrementar la frecuencia de dicho alelo. Sin embargo, una baja frecuencia alélica puede ser por ser antigua o joven y puede, por tanto, tener una alta o baja diversidad. Una selección positiva puede incrementar, rápidamente, la frecuencia de un alelo, lo que puede llegar a ser común sin llegar a conseguir un alto nivel de diversidad alélica.

✚ **Mismatch distribution:** Es la distribución del número de diferencias observadas entre parejas de haplotipos. Comúnmente usado como medio de representar la diversidad de diferentes datos moleculares, conociendo, también, la pairwise differences. La mismatch distribution es apropiada para datos donde las diferencias discretas de los alelos pueden ser contadas. La distribución del número de estas diferencias entre los

alelos nos indica la diversidad genética. La diversidad que aparece de cada una de las muestras, y la forma de la distribución de la misma (representación gráfica en un histograma) nos indica la historia poblacional y concretamente los episodios de expansión popular. Así pues, la forma de la distribución es muy informativa. Una mismatch distribution con forma de campana indica un crecimiento rápido de una población, mientras que una distribución multimodal indica que el tamaño de la población ha sido constante durante un periodo largo de tiempo (Rogers y Harpending, 1992). Así pues, los episodios de crecimiento o declive de una población dejan huellas discernibles en la distribución de las pairwise differences.

✚ **Frecuencias haplotípicas o diversidad haplotípica:** Es la probabilidad de que de que dos individuos cogidos al azar, pertenecientes a la misma población, posean diferentes haplotipos. Se calcularía como

$$(n/n-1)(1 - \sum p_i^2)$$

Debido a que, en este trabajo, lo que se trata es cromosoma Y, que se encuentra de forma haplotípica en la población masculina, no se puede estudiar el equilibrio Hardy-Weinberg.

3.6.1.2.- MÉTODOS INTER-POBLACIONALES.

✚ **AMOVA:** Existe un método alternativo para tratar las frecuencias alélicas que es Analysis of Molecular Variance (AMOVA), que es un método que considera, por ejemplo, la varianza en el número de unidades de repetición de microsatélites dadas en un locus. La AMOVA puede ser aplicada en datos donde las distancias genéticas entre los alelos se calculan. De hecho, nosotros concretamente realizamos distintas agrupaciones de las distintas poblaciones, para ver si las diferencias entre nuestras poblaciones son más importante a nivel cultural, geográfico, etc...

✚ **Distancias genéticas:** Las distancias genéticas F_{st} se calcularon con el software Arlequin 2000 ver (Schneider *et. al.*, 2000). Las distancias genéticas de la poblaciones se calcularon a partir de dos datos. Por un lado, se calcularon a partir de las frecuencias alélicas y, por otro, a partir de las frecuencias haplotípicas.

- A partir de las frecuencias alélicas presentes en las distintas poblaciones se calculó la distancia genética de Reynolds (Reynolds *et al.* 1983). Esta distancia asume que no existe mutación y que la deriva genética es el único factor responsable del cambio en las

frecuencias de los alelos. A diferencia de otras distancias, esta medida no asume un tamaño efectivo de la población constante e igual para todas las poblaciones. De hecho, las fluctuaciones en el tamaño poblacional juegan un importante papel en la magnitud de la distancia puesto que pueden modificar el efecto de la deriva genética.

El cálculo de la matriz de distancias de Reynolds fue calculada con el programa Arlequín 2.000ver (Schneider *et. al.* 2000). A partir de la matriz de distancias de Reynolds, entre todas las poblaciones estudiadas generada con el programa Phylip se construyeron árboles sin raíz, utilizando los métodos de reconstrucción filogenético de Neighbor-Joining (Saitou y Nei 1987).

El método de Neighbor-Joining se basa en el agrupamiento sucesivo de las diferentes líneas para la construcción del árbol, asignando la longitud de las ramas a medida que éstas se van uniendo. Este método de reconstrucción no asume reloj molecular, por lo que los árboles generados con él no presentan raíz. Para la construcción de un árbol de este tipo a partir de la matriz de distancias de Reynolds entre las diferentes poblaciones, se empleó el programa Neighbor del paquete Phylip 3.5c (Felsenstein 1993).

Al igual que se hizo con la matriz de Reynolds, se realizó también un árbol filogenético a partir de la matriz de distancia de Cavalli-Sforza.

- Otra forma sencilla de comparar la información obtenida de las secuencias, es el estudio de la frecuencia y distribución de los haplotipos en las diferentes poblaciones.

La búsqueda de haplotipos comunes entre las poblaciones estudiadas, y posteriormente con las obtenidas a partir de la bibliografía, se realizó con el programa Arlequín Ver. 2.000.

A partir de la frecuencia haplotípica de cada una de las poblaciones se volvió a calcular una matriz de distancias de Reynolds y se construyó un nuevo árbol filogenético, para ver las distancias entre las poblaciones. Posteriormente, se realiza también un árbol filogenético con la matriz de distancia de Slatkin.

La pairwise F_{st} se puede usar para poblaciones que tienen una distancia poblacional corta, aplicando una ligera transformación para alinear las distancias con un tiempo de divergencia poblacional (Reynolds *et. al.* 1983; Slatkin 1995). Los valores de pairwise se obtienen en forma de matriz. La distribución nula de los valores de pairwise bajo la hipótesis de que no hay diferencias entre poblaciones se obtiene por haplotipos permutados entre las poblaciones. El valor de P de este trabajo es

proporcional a la principal permutación para un valor de F_{st} mayor o igual al observado. En este trabajo se utilizó la alineación F_{st} de Slatkin.

Slatkin consideró un modelo demográfico simple donde dos poblaciones haplotípicas de un tamaño N habían divergido (se habían separado) hace ζ generaciones de una población de tamaño idéntico. Estas dos poblaciones habían permanecido aisladas desde entonces, sin ningún intercambio de migración. Bajo estas condiciones, F_{st} puede ser expresado en términos de tiempo de fusión t_1 , que es el tiempo de fusión de dos genes obtenidos de dos poblaciones diferentes y t_0 , que es el tiempo de fusión de dos genes de una misma población. Usando el análisis de varianza, F_{st} se expresa como:

$$F_{st} = \frac{t_1 - t_0}{t_1}$$

Como t_0 es igual para N generaciones (Hudson 1990) y t_1 es igual para $\zeta + N$ generaciones, la expresión de forma reducida sería:

$$F_{st} = \frac{\zeta}{\zeta + N}$$

Por lo tanto, la proporción $D = F_{st}/(1 - F_{st})$ es igual a ζ/N y es por tanto proporcional al tiempo de divergencia entre dos poblaciones.

✚ Multidimensional de Scaling: Multidimensional de Scaling es una técnica de reducción dimensional, en la cual (δ_{ij}) la distancia observada entre n ítems es representada en un sistema de coordenadas de pocas dimensiones (P.ej. dos o tres dimensiones) y las distancias d_{ij} relaciona, estrechamente, las distancias originales δ_{ij} (Rencher 2002), que es:

$$d_{ij} \approx \delta_{ij} \text{ para todos los } ij$$

El programa de estadística "Statistica 6.0ver" fue usado para el análisis MDS. El MDS descubre las relaciones entre las poblaciones y puede representarlas de una manera gráfica. El MDS, sin embargo, no asume las relaciones lineales entre estos grupos, por esta razón, es aplicable a muchos más tipos de datos. El gráfico producido por el MDS es de dos o tres dimensiones, aunque el número apropiado de dimensiones es necesario visualizarlo porque a veces es difícil de interpretar. EL MDS desarrolla un gráfico dimensional de las poblaciones que trata de minimizar las diferencias entre las distancias entre los puntos en el gráfico y la matriz de datos de entrada, usando una función de "stress" (Kruskal y Desea, 1978). En muchos casos, algunos patrones interesantes (entre los

cuales se pueden observar determinados agrupamientos que parecen reflejar ciertas variables como la distancia geográfica) pueden identificarse por la inspección visual de estos gráficos. En términos más técnicos, el programa usa un algoritmo de minimización de función que evalúa configuraciones diferentes con el objetivo de obtener la máxima calidad. La medida de "stress" nos sirve para evaluar la matriz de distancias observada en una configuración particular. El valor de "stress" de una configuración está definido por:

$$\text{Stress} = \sum [d_{ij} - f(\delta_{ij})]^2$$

Donde la $f(\delta_{ij})$ indica un valor no métrico, la transformación de monotonía de la entrada observada se distancia de δ_{ij} . Así el programa intentará reproducir el ordenamiento de fila general de distancias entre los grupos en el análisis. El valor más pequeño de stress es el reproducido por una matriz de distancias para una matriz de distancias observada. Kruskal sugiere las evaluaciones siguientes para varios niveles de stress:

Stress	Calidad
0.40	Pobre
0.20	Bajo
0.10	Bueno
0.05	Excelente
0.00	Perfecto

En este estudio hemos utilizado las matrices de Slatkin para la representación bidimensional de las distancias entre los distintos grupos.

✚ Análisis de correspondencias o componentes principales: Además de expresar las relaciones entre poblaciones usando árboles o redes, se pueden visualizar similitudes (también diferencias) genéticas entre poblaciones mediante gráficos de dos o más dimensiones. Estas representaciones gráficas de las relaciones entre los grupos tienen la ventaja de no admitir ningún orden de ramificación, simplemente distribuirán los grupos de modo que estén asociados los más estrechamente relacionados entre sí. Es una técnica gráfica para representar la información en una tabla de contingencia de doble dirección, que contiene las frecuencias de los ítems para una clasificación a través de las dos variables categóricas. Con CA se puede reconstruir un argumento que muestra la interacción de dos variables categóricas en filas y columnas (Rencher, 2002). Cuando están implicadas más de dos variables se trata de una MCA (CA múltiple).

Para ello, se utilizó el programa Statistica 6.0ver. Se realizó con las frecuencias haplotípicas y alélicas obtenidas en nuestras poblaciones,

así como con los haplogrupos, realizando posteriormente una representación gráfica en dos y tres dimensiones. Esto mismo, se realizó también con nuestras poblaciones y las obtenidas de la bibliografía, para observar las posibles distancias entre los distintos continentes y las distintas etnias de cada uno.

✚ Análisis Filogenético: La taxonomía es una disciplina clásica que tiene entre sus objetivos clasificar a los seres vivos. Una de las técnicas más eficientes usadas para esto son los métodos filogenéticos. Estos métodos utilizan información relacionada con rasgos de las especies para clasificarlas.

Existen diferentes métodos filogenéticos basados en distintos criterios para reconstruir el árbol de parentesco que relaciona entre sí a los haplotipos estudiados. El objetivo de la inferencia filogenética es construir un árbol evolutivo a partir de la información disponible de los diferentes caracteres que se han medido en los taxones objetos de estudio.

Los árboles constan de nodos conectados por ramas. Los nodos terminales suelen denominarse unidades taxonómicas operacionales (OTUs) o taxones terminales y representan aquellas entidades de las que disponemos de información en forma de caracteres. Tales unidades pueden existir o haber desaparecido. Los nodos internos representan ancestros hipotéticos. Al ancestro de todas las unidades taxonómicas estudiadas se le denomina raíz. Dependiendo de los métodos utilizados para la obtención de los árboles evolutivos, podemos obtener información variada sobre los nodos y las ramas, por ejemplo la naturaleza hipotética de los caracteres que componen los nodos internos, o las longitudes de los brazos.

Los árboles evolutivos pueden ser representados con o sin raíz. Un árbol con raíz tiene un nodo al que se identifica como raíz y del que derivan el resto de los nodos descendientes, siguiendo el tiempo evolutivo. Por ello, cuanto más próximo se encuentre un nodo determinado a la raíz, mayor es su antigüedad. Por otro lado, los árboles con raíz permiten identificar las relaciones ancestro-descendientes entre los nodos. De tal modo, que si tenemos dos nodos conectados por una rama podemos decir que el nodo más próximo a la raíz es el nodo ancestral del nodo más alejado del mismo, el nodo descendiente.

El nodo sin raíz difiere del árbol con raíz en que tiene un nodo menos y una rama menos (la correspondiente a la raíz). De este árbol no se puede deducir las relaciones entre ancestros y descendencia entre los nodos que aparecen.

Es muy difícil evaluar la exactitud de los árboles filogenéticos, aunque se han empleado algunos métodos que pueden dar fiabilidad a estos árboles. El Bootstrapping (Felsenstein, 1985) es el más popular de éstos métodos. Es un método adoptado de la estadística y consiste en producir un número elegido de replicaciones de los datos y posteriormente se seleccionan las columnas de datos al azar para obtener un conjunto de datos variado. Las replicaciones se analizan para generar un árbol filogenético para cada uno de ellos y posteriormente se calcula un árbol consenso.

Los árboles de Bootstrap tienen números asociados a cada nudo, representando la proporción de replicaciones que la generaron.

El árbol consenso trata de resolver el problema de la comparación entre árboles que tienen una misma longitud o igualmente parsimoniosos, tratando de optimizar las diferentes topologías y no de optimizar la matriz de datos. La búsqueda de este árbol, se condensa en una única representación en la que se diferencian claramente las partes concordantes entre los distintos árboles, de las de las partes discordantes.

 **Networks (redes):** Otra manera de aproximarse a las relaciones evolutivas es mediante el uso de una red. En muchos casos, ésta puede ser una representación más apropiada de la realidad que un árbol. Es muy adecuada para la investigación en los casos donde existe una alta tasa de mutación y donde pueden ocurrir mutaciones recurrentes que puede causar ambigüedad en el modelo evolutivo. Una red nos permite dibujar estas ambigüedades en las conexiones entre los nudos en una topología, conectando nudos a través de vías múltiples.

Los algoritmos de esta red, generalmente utilizan los datos de la secuencia o datos de los haplotipos, en lugar de los datos de distancia, y generan sus propias distancias genéticas de estos datos (Huson 1998).

El network Median Joining (MJ) identifica relaciones e introduce una hipótesis sobre el origen ancestral que pueda unir los distintos tipos en un árbol o red (Bandelt *et.al.*, 1999).

El análisis de Network utilizando el software Network 4.1 (Fluxus Technology Ltd., Clare, Suffolk, UK) se utilizó para situar la mayor parte de los haplotipos observados en las muestras Colón/Colombo. El network MJ fue calculado colocando ϵ igual a cero, con el fin de reducir el tiempo de run del análisis y evitar network complejos y difíciles de interpretar.

3.6.1.3.- OTROS PARÁMETROS ESTADÍSTICOS DE INTERÉS FORENSE.

La frecuencia de un alelo en una población se denomina frecuencia génica o alélica y se puede definir como el cociente resultante de dividir el número de alelos iguales en una población por el número total de alelos.

Para la aplicación de cualquier marcador genético en Genética Forense es fundamental determinar las frecuencias alélicas del sistema polimórfico en la población, realizando una estima de dichas frecuencias a través de un fenotipado muestral representativo, dada la imposibilidad de fenotipar a todos y cada uno de los integrantes de la población para obtener su valor exacto.

Los siguientes parámetros estadísticos nos indican la utilidad a priori de los marcadores:

✚ **Probabilidad de exclusión a priori (CE)** de un marcador en una población determinada. Se define como la probabilidad de exclusión de un individuo como padre en un análisis de paternidad.

Para sistemas autosómicos se calcula mediante la siguiente fórmula (Ohno *et al.*, 1982):

$$CE = \sum_{i=1}^n p_i (1 - p_i)^2 (1 - p_i + p_i^2) + \sum_{i=1}^{n-1} \sum_{j=i+1}^n p_i p_j (p_i + p_j)(1 - p_i - p_j)^2$$

siendo p_i , p_j frecuencias génicas respectivas.

Para sistemas de cromosoma Y ésta coincidiría con $1 - \sum p_i^2$ siendo $\sum p_i^2$ la probabilidad de inclusión, es decir, que el padre sea efectivamente el padre.

El CE depende del número de alelos identificados y de la distribución de las frecuencias alélicas. Es un valor porcentual que es función directa del polimorfismo de un sistema. Cuanto más polimórfico ser éste y más equilibradas estén las frecuencias de sus alelos, tanto mayor será su CE, lo que lo hace ser un parámetro de gran valor médico-legal, especialmente en determinaciones biológicas de paternidad.

✚ **Poder de discriminación (PD)**. Es una medida relativa de la eficacia de un sistema, sobre todo con fines de identificación forense. Se define como la probabilidad de que dos individuos tomados al azar, pertenecientes a la misma población, posean diferentes fenotipos. Se calcula mediante la siguiente fórmula (Fisher, 1951):

$$PD = 1 - \sum_{i=1}^n p_i^2$$

siendo p_i frecuencias genotípicas esperadas, que para cromosoma Y coincidan con las frecuencias alélicas, siendo $\sum p_i^2$ la probabilidad de coincidencia o "matching".

✚ **Probabilidad de coincidencia (PM)** o "matching probability". Se define como la probabilidad de que dos individuos tomados al azar de la misma población coincidan en su genotipo para ese locus. Esta probabilidad fue descrita por Jones (1972) y se calcula a partir de la siguiente fórmula:

$$PM = \sum_{i=1}^n p_i^2$$

siendo p_i la frecuencia esperada de cada genotipo (para cromosoma Y, frecuencias alélicas).

Como vemos, el poder de discriminación y la probabilidad de coincidencia son conceptos opuestos ($PD = 1 - PM$).

✚ **Diversidad génica y diversidad haplotípica.** Para STRs de cromosoma Y, los valores de diversidad génica (D) o haplotípica en su caso, tienen el mismo valor que el poder de discriminación y la probabilidad de exclusión, de tal manera que:

$$PD = CE = 1 - PM = 1 - \sum p_i^2$$

Siendo p_i las frecuencias alélicas o haplotípicas.



Gran Via de Colón
AUTOR: José Martínez Rioboó

Fundación Rodríguez-Acosta
Donación Martínez Sóla.

4.-RESULTADOS

En este capítulo se procederá a la exposición ordenada y esquematizada de todos los resultados obtenidos. Es decir, se estudiarán los resultados obtenidos para el estudio poblacional de los loci: DYS 456, DYS 389I, DYS 390, DYS 389II, DYS 458, DYS 19, DYS 385, DYS 393, DYS 391, DYS 439, DYS 635, DYS 392, Y-GATA H4, DYS 437, DYS 438 y DYS 448. Además compararemos los resultados obtenidos en cada una de las zonas analizadas.

4.1.- RESULTADOS DEL ESTUDIO POBLACIONAL.

4.1.1.- VISUALIZACIÓN DE LOS RESULTADOS OBTENIDOS.

Tras el análisis de las muestras con el programa Genotyper, éste asignó automáticamente los alelos a cada marcador y los resultados se visualizaron como se muestra a continuación.

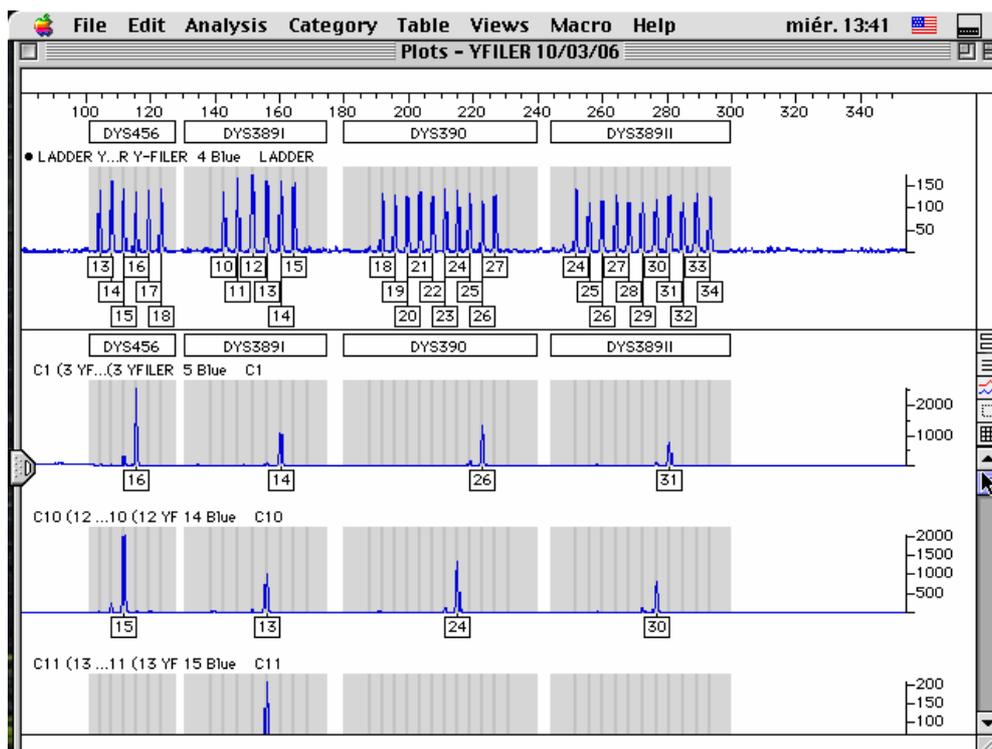


Imagen 4.1. Electroferograma en el que se observan los resultados obtenidos para los loci marcados con el fluorocromo azul (6-FAM): DYS 456, DYS 389I, DYS 390 y DYS 389II.

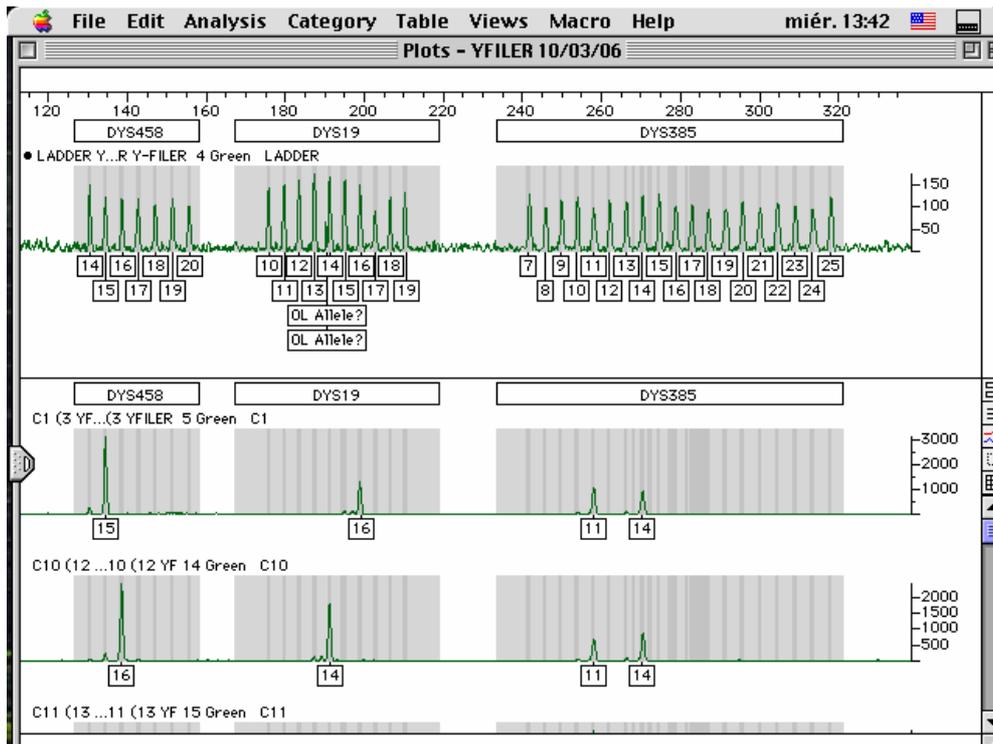


Imagen 4.2. Electroferograma en el que se observan los resultados obtenidos para los loci marcados con el fluorocromo verde (VIC): DYS 458, DYS 19 y DYS 385.

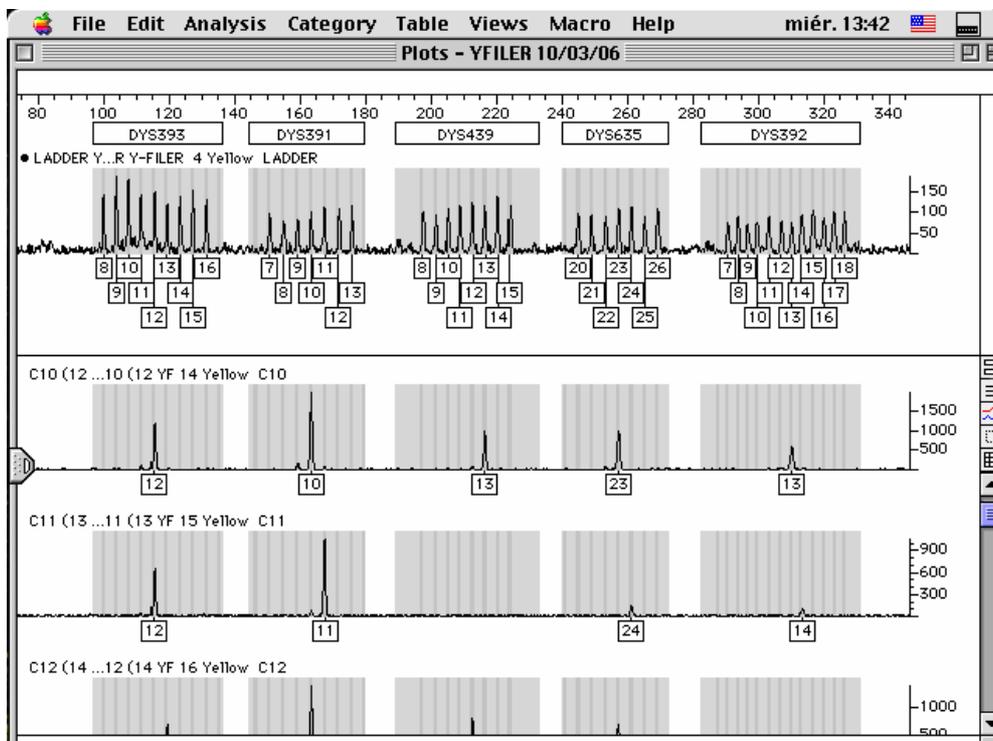


Imagen 4.3. Electroferograma en el que se observan los resultados obtenidos para los loci marcados con el fluorocromo amarillo (NED): DYS 393, DYS 391, DYS 439, DYS 635 y DYS 392.

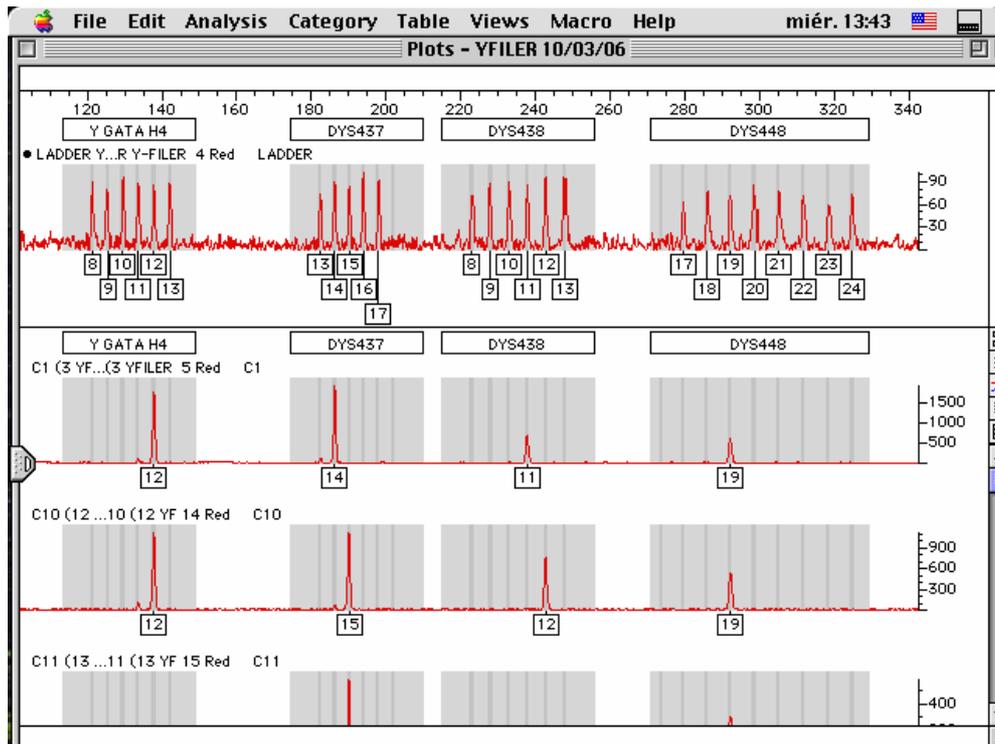


Imagen 4.4. Electroferograma en el que se observan los resultados obtenidos para los loci marcados con el fluorocromo rojo (PET): Y-GATA H4, DYS 437, DYS 438 y DYS 448.

4.1.2.- RESULTADOS OBTENIDOS CON EL KIT Y-FILER.

Pasamos a exponer las frecuencias alélicas así como la diversidad génica, de cada uno de los loci, empezando por el DYS 456 de cada una de las poblaciones estudiadas, en primer lugar los apellidados Colón, para poder observar mejor las diferencias entre los distintos grupos poblacionales. Así pues, son 125 individuos catalanes, 50 individuos de la población balear, 44 valencianos, 35 individuos del sur de Francia, 46 individuos de Liguria (Génova y sus alrededores), 50 de Lombardía (Milán y sus alrededores) y, finalmente, 13 individuos de Piamonte (Turín y sus alrededores):

Tabla 4.1. Frecuencias alélicas del sistema DYS 456 para la población catalana.

DYS 456	Recuento	Frecuencia alélica	s.d.	Diversidad génica
<13	3	0.024	0.013	0.5952
13	4	0.032	0.015	
14	4	0.032	0.015	
15	58	0.464	0.044	
16	54	0.432	0.044	
17	2	0.016	0.011	

Diversidad génica computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.2. Frecuencias alélicas del sistema DYS 456 de la población balear.

DYS 456	Recuento	Frecuencia alélica	s.d	Diversidad génica
14	2	0.04	0.027	0.4024
15	37	0.74	0.062	
16	11	0.22	0.058	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.3. Frecuencias alélicas del sistema DYS 456 de la población valenciana.

DYS 456	Recuento	Frecuencia alélica	s.d	Diversidad génica
13	5	0.1136	0.047	0.6125
15	18	0.4091	0.074	
16	20	0.4545	0.075	
17	1	0.0227	0.022	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.4. Frecuencias alélicas del sistema DYS 456 de la población francesa.

DYS 456	Recuento	Frecuencia alélica	s.d	Diversidad génica
<13	1	0.0285	0.028	0.4947
14	4	0.1143	0.053	
15	24	0.6857	0.078	
16	5	0.1428	0.059	
17	1	0.0285	0.028	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.5. Frecuencias alélicas del sistema DYS 456 de la población italiana de Liguria.

DYS 456	Recuento	Frecuencia alélica	s.d	Diversidad génica
13	3	0.0652	0.036	0.6862
14	3	0.0652	0.036	
15	21	0.4565	0.073	
16	13	0.2826	0.066	
17	6	0.1304	0.049	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.6. Frecuencias alélicas del sistema DYS 456 de la población italiana de Lombardía.

DYS 456	Recuento	Frecuencia alélica	s.d	Diversidad génica
14	9	0.18	0.054	0.7000
15	20	0.4	0.069	
16	16	0.32	0.065	
17	3	0.06	0.033	
18	2	0.04	0.069	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.7. Frecuencias alélicas del sistema DYS 456 de la población italiana de Piamonte.

DYS 456	Recuento	Frecuencia alélica	s.d	Diversidad génica
15	5	0.3846	0.134	0.4734
16	8	0.6154	0.134	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

DYS 456

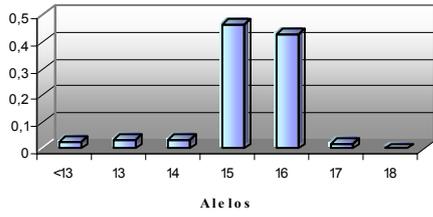


Fig. 4.1: Frecuencias alélicas de la población Catalana

DYS 456

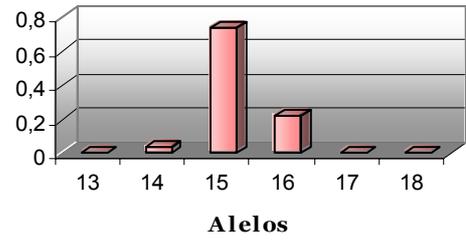


Fig. 4.2: Frecuencias alélicas de la población balear.

DYS 456

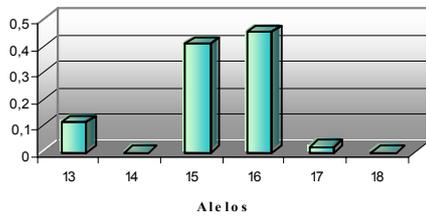


Fig. 4.3: Frecuencias alélicas de la población valenciana

DYS 456

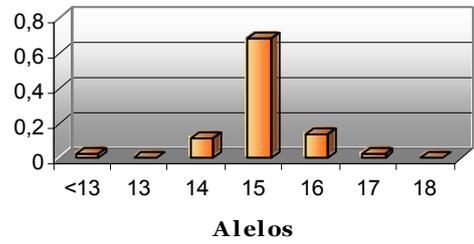


Fig. 4.4: Frecuencias alélicas de Francia.

DYS 456

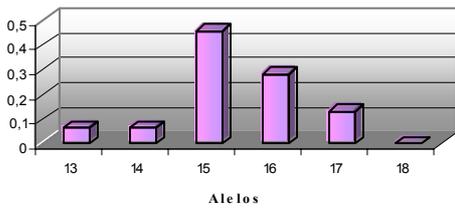


Fig.a 4.5: Frecuencias alélicas de Liguria.

DYS 456

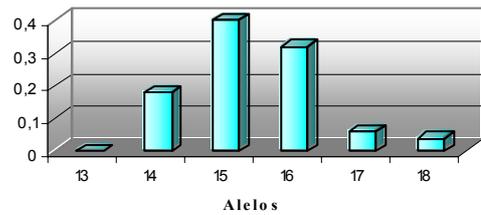


Fig.4.6: Frecuencias alélicas de Lombardía.

DYS 456

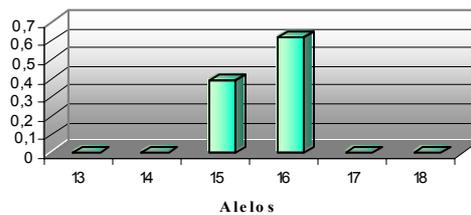


Fig. 4.7: Frecuencias alélicas de Piamonte

Tabla 4.8. Frecuencias alélicas del sistema DYS 389I para la población catalana.

DYS 389I	Recuento	Frecuencia alélica	s.d	Diversidad génica
11	2	0.016	0.011	0.5154
12	19	0.152	0.032	
13	82	0.656	0.042	
14	22	0.176	0.034	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.9. Frecuencias alélicas del sistema DYS 389I para la población balear.

DYS 389I	Recuento	Frecuencia alélica	s.d	Diversidad génica
12	1	0.02	0.019	0,2152
13	44	0.88	0.045	
14	5	0.1	0.042	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.10 Frecuencias alélicas del sistema DYS 389I para la población valenciana.

DYS 389I	Recuento	Frecuencia alélica	s.d	Diversidad génica
12	5	0.1136	0.047	0.2779
13	37	0.8409	0.055	
14	2	0.0454	0.031	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.11. Frecuencias alélicas del sistema DYS 389I para la población francesa.

DYS 389I	Recuento	Frecuencia alélica	s.d	Diversidad génica
12	8	0.2285	0.070	0.6010
13	19	0.5428	0.084	
14	8	0.2285	0.070	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.12. Frecuencias alélicas del sistema DYS 389I para la población italiana de Liguria.

DYS 389I	Recuento	Frecuencia alélica	s.d	Diversidad génica
12	8	0.1739	0.055	0.4434
13	33	0.7174	0.066	
14	5	0.1087	0.045	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.13. Frecuencias alélicas del sistema DYS389I para la población italiana de Lombardía.

DYS 389I	Recuento	Frecuencia alélica	s.d	Diversidad génica
12	8	0.16	0.051	0.4416
13	36	0.72	0.063	
14	6	0.12	0.045	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.14. Frecuencias alélicas del sistema DYS 389I para la población italiana de Piamonte.

DYS 389I	Recuento	Frecuencia alélica	s.d	Diversidad génica
12	3	0.2307	0.116	0.6037
13	7	0.5384	0.138	
14	3	0.2307	0.116	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

DYS 389I

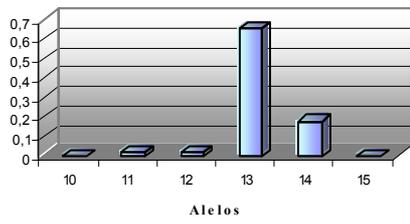


Figura 4.8: Frecuencias alélicas de Cataluña.

DYS 389I

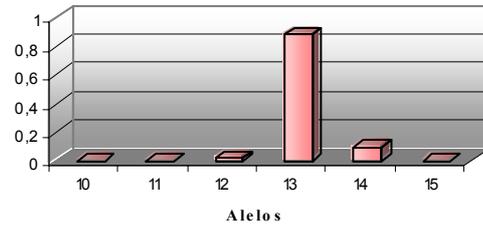


Figura 4.9: Frecuencias alélicas de Baleares.

DYS 389I

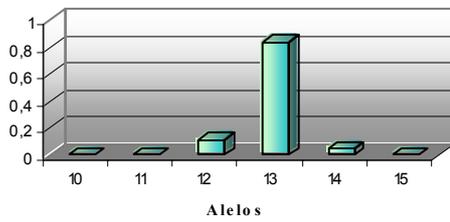


Figura 4.10: Frecuencias alélicas de Valencia.

DYS 389I

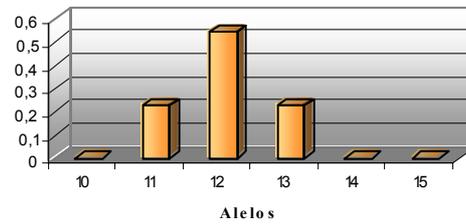


Figura 4.11: Frecuencias alélicas de Francia.

DYS 389I

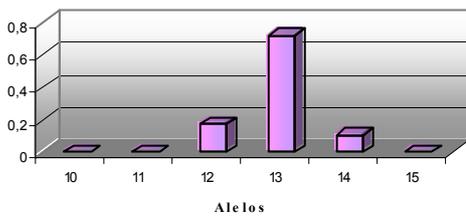


Figura 4.12: Frecuencias alélicas de Liguria.

DYS 389I

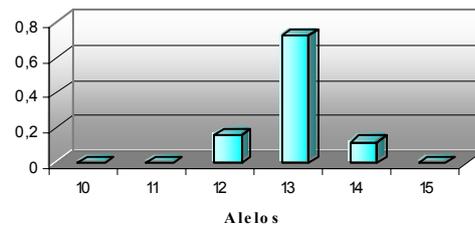


Figura 4.13: Frecuencias alélicas de Lombardía.

DYS 389I

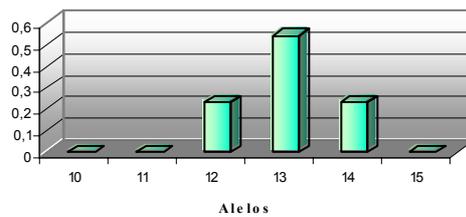


Figura 4.14: Frecuencias alélicas de Piemonte

Tabla 4.15. Frecuencias alélicas del sistema DYS 390 para la población catalana.

DYS 390	Recuento	Frecuencia alélica	s.d	Diversidad génica
21	2	0.016	0.011	0.5088
22	6	0.048	0.019	
23	23	0.184	0.034	
24	84	0.672	0.041	
25	5	0.04	0.017	
26	5	0.04	0.017	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.16. Frecuencias alélicas del sistema DYS 390 para la población balear.

DYS 390	Recuento	Frecuencia alélica	s.d	Diversidad génica
22	3	0.06	0.033	0.2528
23	3	0.06	0.033	
24	43	0.86	0.049	
25	1	0.02	0.019	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.17. Frecuencias alélicas del sistema DYS 390 para la población valenciana.

DYS 390	Recuento	Frecuencia alélica	s.d	Diversidad génica
23	30	0.6818	0.070	0.4340
24	14	0.3181	0.070	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.18. Frecuencias alélicas del sistema DYS 390 para la población francesa.

DYS 390	Recuento	Frecuencia alélica	s.d	Diversidad génica
22	3	0.0857	0.047	0.4588
23	5	0.1428	0.059	
24	25	0.7143	0.076	
25	2	0.0571	0.039	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.19. Frecuencias alélicas del sistema DYS 390 para la población italiana de Liguria.

DYS 390	Recuento	Frecuencia alélica	s.d	Diversidad génica
22	3	0.0652	0.041	0.6267
23	13	0.2826	0.076	
24	24	0.5217	0.084	
25	6	0.1304	0.056	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.20. Frecuencias alélicas del sistema DYS 390 para la población italiana de Lombardía.

DYS 390	Recuento	Frecuencia alélica	s.d	Diversidad génica
22	4	0.08	0.038	0.5896
23	12	0.24	0.060	
24	29	0.58	0.069	
25	5	0.1	0.042	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.21. Frecuencias alélicas del sistema DYS 390 para la población italiana de Piemonte.

DYS 390	Recuento	Frecuencia alélica	s.d	Diversidad génica
22	2	0.1538	0.100	0.6155
23	6	0.4615	0.138	
24	5	0.3846	0.134	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

DYS390

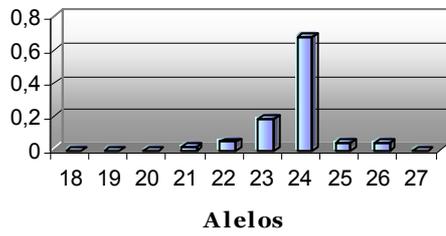


Figura 4.15: Frecuencias alélicas de Cataluña

DYS390

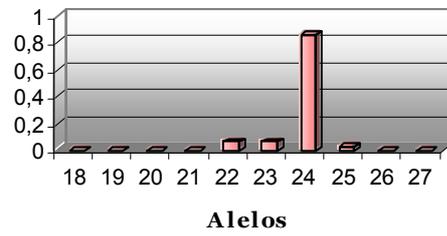


Figura 4.16: Frecuencias alélicas de Baleares

DYS390

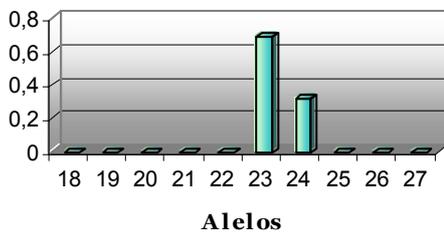


Figura 4.17: Frecuencias alélicas de Valencia

DYS390

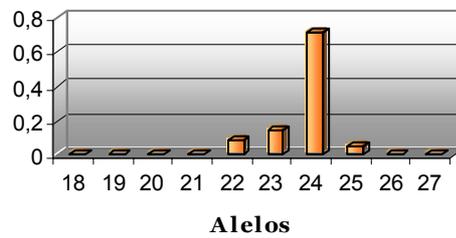


Figura 4.18: Frecuencias alélicas de Francia

DYS390

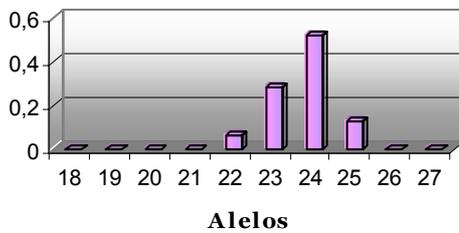


Figura 4.19: Frecuencias alélicas de Liguria

DYS390

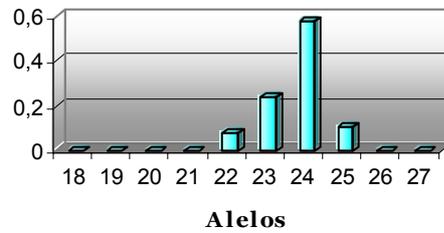


Fig. 4.20: Frecuencias alélicas de Lombardia

DYS390

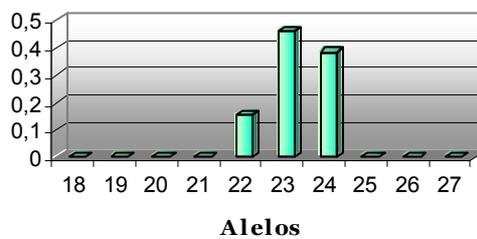


Figura 4.21: Frecuencias alélicas de Piemonte

Tabla 4.22. Frecuencias alélicas del sistema DYS 389II para la población catalana.

DYS 389II	Recuento	Frecuencia alélica	s.d	Diversidad génica
27	2	0.016	0.011	0.7022
28	30	0.24	0.038	
29	42	0.336	0.042	
30	44	0.352	0.042	
31	7	0.056	0.020	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.23. Frecuencias alélicas del sistema DYS 389II para la población balear.

DYS 389II	Recuento	Frecuencia alélica	s.d	Diversidad génica
29	45	0.9	0.042	0.1864
30	2	0.04	0.027	
31	2	0.04	0.027	
32	1	0.02	0.019	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.24. Frecuencias alélicas del sistema DYS 389II para la población valenciana.

DYS 389II	Recuento	Frecuencia alélica	s.d	Diversidad génica
28	19	0.4318	0.074	0.6396
29	16	0.3636	0.072	
30	9	0.2045	0.060	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.25. Frecuencias alélicas del sistema DYS 389II para la población francesa.

DYS 389II	Recuento	Frecuencia alélica	s.d	Diversidad génica
28	3	0.0857	0.047	0.6204
29	16	0.4571	0.084	
30	14	0.4	0.082	
31	2	0.0571	0.039	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.26. Frecuencias alélicas del sistema DYS 389II para la población italiana de Liguria.

DYS 389II	Recuento	Frecuencia alélica	s.d	Diversidad génica
28	7	0.1521	0.052	0.5898
29	27	0.5869	0.072	
30	9	0.1956	0.058	
31	3	0.0652	0.036	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.27. Frecuencias alélicas del sistema DYS389II para la población italiana de Lombardía.

DYS 389II	Recuento	Frecuencia alélica	s.d	Diversidad génica
28	5	0.1	0.042	0.6400
29	25	0.5	0.070	
30	15	0.3	0.064	
31	5	0.1	0.042	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.28. Frecuencias alélicas del sistema DYS 389II para la población italiana de Piamonte.

DYS 389II	Recuento	Frecuencia alélica	s.d	Diversidad génica
28	3	0.2307	0.116	0.7340
29	4	0.3077	0.128	
30	4	0.3077	0.128	
32	2	0.1538	0.100	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

DYS389II

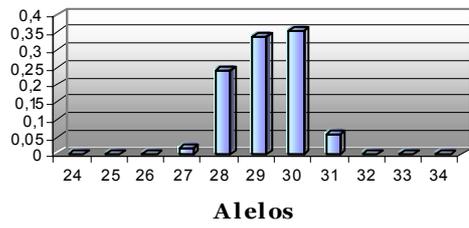


Figura 4.22: Frecuencias alélicas de Cataluña

DYS389II

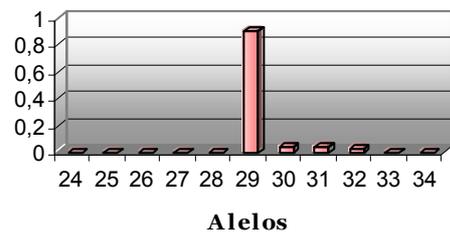


Figura 4.23: Frecuencias alélicas de Baleares

DYS389II

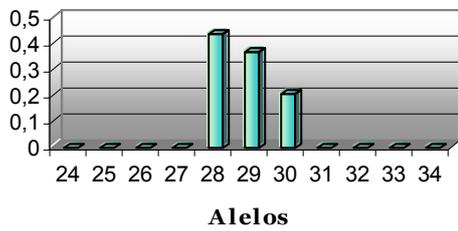


Figura 4.24: Frecuencias alélicas de Valencia

DYS389II

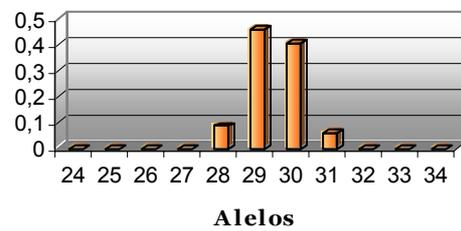


Figura 4.25: Frecuencias alélicas de Francia

DYS389II

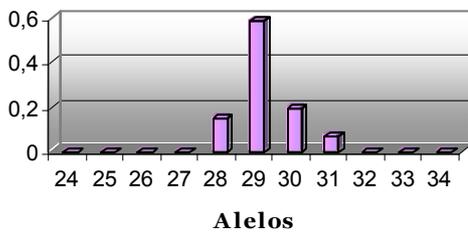


Figura 4.26: Frecuencias alélicas de Liguria

DYS389II

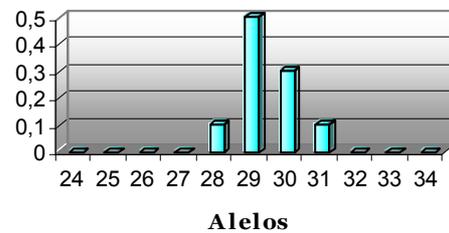


Figura 4.27: Frecuencias alélicas de Lombardia

DYS389II

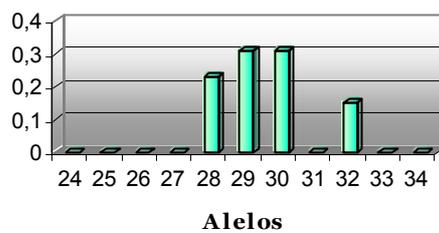


Figura 4.28: Frecuencias alélicas de Piemonte

Tabla 4.29. Frecuencias alélicas del sistema DYS 458 para la población catalana.

DYS 458	Recuento	Frecuencia alélica	s.d	Diversidad génica
15	26	0.208	0.036	0.6814
16	59	0.472	0.044	
17	25	0.2	0.035	
18	14	0.112	0.028	
19	1	$8 \cdot 10^{-3}$	$7.96 \cdot 10^{-3}$	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.30. Frecuencias alélicas del sistema DYS 458 para la población balear.

DYS 458	Recuento	Frecuencia alélica	s.d	Diversidad génica
15	3	0.06	0.033	0.3760
16	39	0.78	0.058	
17	5	0.1	0.042	
18	2	0.04	0.027	
19	1	0.02	0.019	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.31. Frecuencias alélicas del sistema DYS 458 para la población valenciana.

DYS 458	Recuento	Frecuencia alélica	s.d	Diversidad génica
15	2	0.0454	0.031	0.7779
16	11	0.25	0.065	
16.2	14	0.3181	0.070	
17	6	0.1363	0.051	
18	8	0.1818	0.058	
19	3	0.0682	0.038	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.32. Frecuencias alélicas del sistema DYS 458 para la población francesa.

DYS 458	Recuento	Frecuencia alélica	s.d	Diversidad génica
<14	1	0.0285	0.028	0.6890
14	1	0.0285	0.028	
15	3	0.0857	0.047	
16	13	0.3714	0.081	
17	14	0.4	0.082	
18	2	0.0571	0.039	
20	1	0.0285	0.028	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.33. Frecuencias alélicas del sistema DYS 458 para la población italiana de Liguria.

DYS 458	Recuento	Frecuencia alélica	s.d	Diversidad génica
<14	1	0.0217	0.021	0.7335
14	2	0.0435	0.030	
15	8	0.1739	0.055	
16	3	0.0652	0.036	
17	20	0.4348	0.073	
17.2	1	0.0217	0.021	
18	9	0.1956	0.058	
19	2	0.0435	0.030	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.34. Frecuencias alélicas del sistema DYS 458 para la población italiana de Lombardía.

DYS 458	Recuento	Frecuencia alélica	s.d	Diversidad génica
14	3	0.06	0.033	0.7368
15	10	0.2	0.056	
16	9	0.18	0.054	
17	21	0.42	0.069	
17.2	1	0.02	0.019	
18	5	0.1	0.042	
18.2	1	0.02	0.019	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.35. Frecuencias alélicas del sistema DYS 458 para la población italiana de Piemonte.

DYS 458	Recuento	Frecuencia alélica	s.d	Diversidad génica
15	4	0.3077	0.128	0.6628
16	6	0.4615	0.138	
17	2	0.1538	0.100	
18	1	0.0769	0.073	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

DYS 458

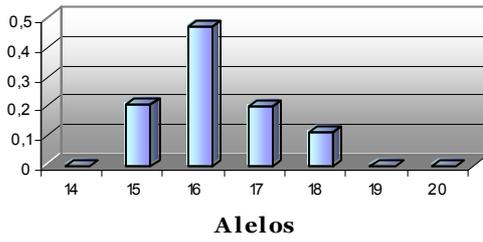


Figura 4.29: Frecuencias alélicas de Cataluña

DYS 458

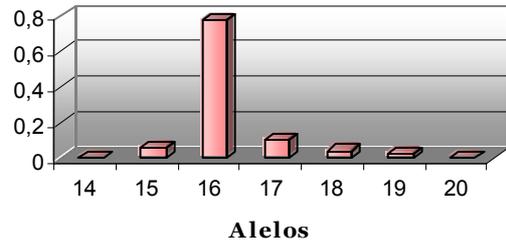


Figura 4.30: Frecuencias alélicas de Baleares

DYS 458

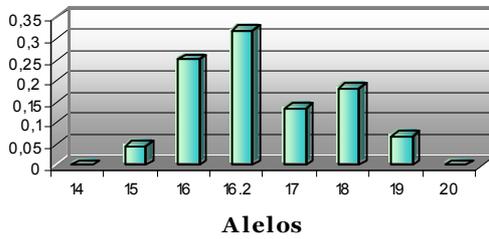


Figura 4.31: Frecuencias alélicas de Valencia

DYS 458

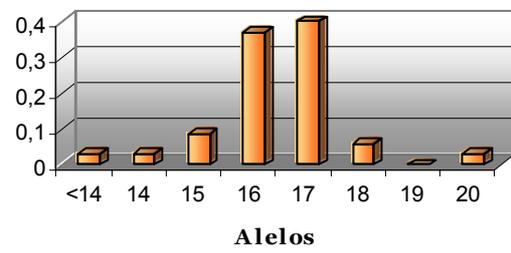


Figura 4.32: Frecuencias alélicas de Francia

DYS 458

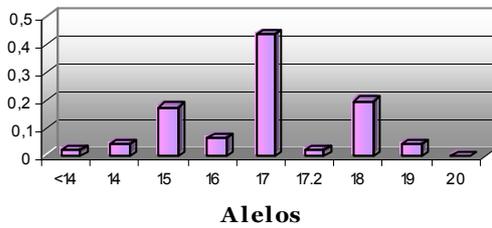


Figura 4.33: Frecuencias alélicas de Liguria

DYS 458

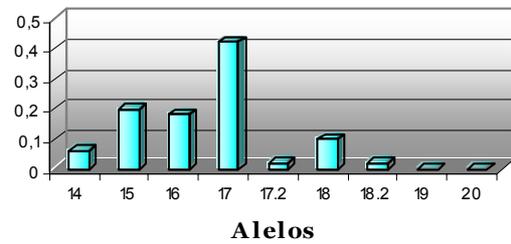


Figura 4.34: Frecuencias alélicas de Lombardía

DYS 458

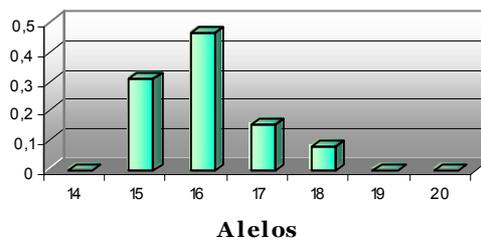


Figura 4.35: Frecuencias alélicas de Piemonte

Tabla 4.36. Frecuencias alélicas del sistema DYS 19 para la población catalana.

DYS 19	Recuento	Frecuencia alélica	s.d	Diversidad génica
13	6	0.048	0.019	0.4855
14	85	0.68	0.041	
15	27	0.216	0.036	
16	7	0.056	0.020	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.37. Frecuencias alélicas del sistema DYS 19 para la población balear.

DYS 19	Recuento	Frecuencia alélica	s.d	Diversidad génica
13	2	0.04	0.027	0.1848
14	45	0.9	0.042	
15	3	0.06	0.033	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.38. Frecuencias alélicas del sistema DYS 19 para la población valenciana.

DYS 19	Recuento	Frecuencia alélica	s.d	Diversidad génica
14	34	0.7727	0.063	0.3512
15	10	0.2273	0.063	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.39. Frecuencias alélicas del sistema DYS 19 para la población francesa.

DYS 19	Recuento	Frecuencia alélica	s.d	Diversidad génica
12	1	0.0285	0.028	0.3870
13	2	0.0571	0.039	
14	27	0.7714	0.070	
15	4	0.1143	0.053	
17	1	0.0285	0.028	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.40. Frecuencias alélicas del sistema DYS 19 para la población italiana de Liguria.

DYS 19	Recuento	Frecuencia alélica	s.d	Diversidad génica
13	5	0.1087	0.045	0.5369
13.2	1	0.0217	0.021	
14	30	0.6521	0.070	
15	7	0.1521	0.052	
16	1	0.0217	0.021	
17	2	0.0435	0.030	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.41. Frecuencias alélicas del sistema DYS 19 para la población italiana de Lombardía.

DYS 19	Recuento	Frecuencia alélica	s.d	Diversidad génica
13	5	0.1	0.042	0.5416
14	32	0.64	0.067	
15	9	0.18	0.054	
16	4	0.08	0.038	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.42. Frecuencias alélicas del sistema DYS 19 para la población italiana de Piamonte.

DYS 19	Recuento	Frecuencia alélica	s.d	Diversidad génica
13	1	0.0769	0.073	0.6036
14	7	0.5384	0.1382	
15	4	0.3076	0.128	
16	1	0.0769	0.073	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

DYS 19

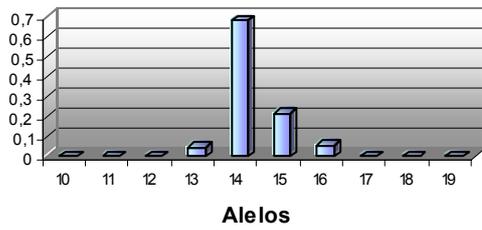


Figura 4.36: Frecuencias alélicas de Cataluña

DYS 19

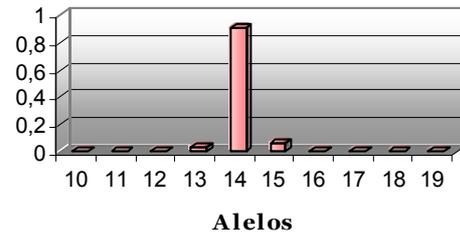


Figura 4.37: Frecuencias alélicas de Baleares

DYS 19

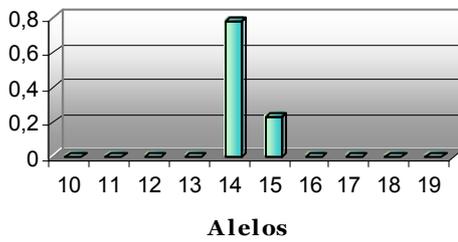


Figura 4.38: Frecuencias alélicas de Valencia

DYS 19

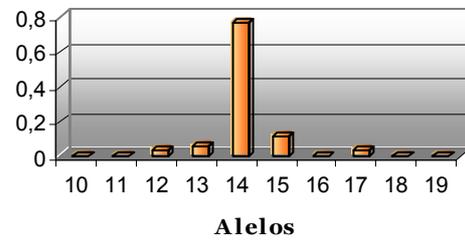


Figura 4.39: Frecuencias alélicas de Francia

DYS 19

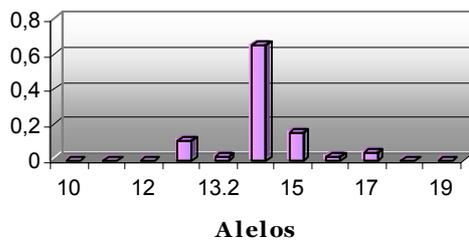


Figura 4.40: Frecuencias alélicas de Liguria

DYS 19

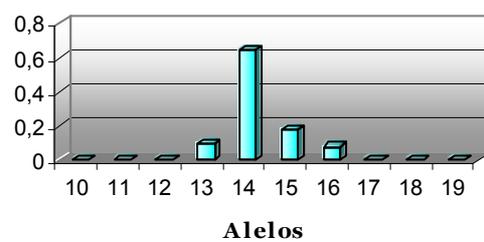


Figura 4.41: Frecuencias alélicas de Lombardía

DYS 19

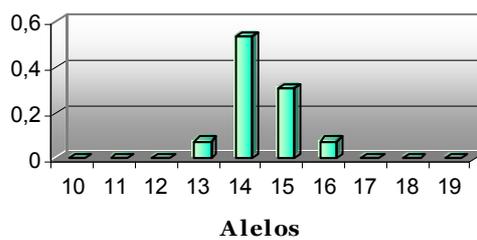


Figura 4.42: Frecuencias alélicas de Piemonte

Tabla 4.43. Frecuencias alélicas del sistema DYS 385 para la población catalana.

DYS 385	Recuento	Frecuencia alélica	s.d	Diversidad génica
10	15	0.06	0.015	0.7718
11	71	0.284	0.028	
12	2	$8 \cdot 10^{-3}$	$5.63 \cdot 10^{-3}$	
13	24	0.096	0.018	
14	87	0.348	0.030	
15	21	0.084	0.017	
16	19	0.076	0.016	
17	6	0.024	$9.67 \cdot 10^{-3}$	
18	2	$8 \cdot 10^{-3}$	$5.63 \cdot 10^{-3}$	
20	3	0.012	$6.88 \cdot 10^{-3}$	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.44. Frecuencias alélicas del sistema DYS 385 para la población balear.

DYS 385	Recuento	Frecuencia alélica	s.d	Diversidad génica
10	2	0.02	0.014	0.6242
11	40	0.4	0.048	
12	2	0.02	0.014	
13	1	0.01	$9.94 \cdot 10^{-3}$	
14	46	0.46	0.049	
15	5	0.05	0.021	
16	2	0.02	0.014	
18	2	0.02	0.014	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.45. Frecuencias alélicas del sistema DYS 385 para la población valenciana.

DYS 385	Recuento	Frecuencia alélica	s.d	Diversidad génica
10	2	0.0227	0.015	0.7658
11	22	0.25	0.046	
12	1	0.0113	0.011	
13	8	0.0909	0.030	
14	32	0.3636	0.051	
15	4	0.0454	0.022	
17	5	0.0568	0.024	
18	14	0.1591	0.038	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.46. Frecuencias alélicas del sistema DYS 385 para la población francesa.

DYS 385	Recuento	Frecuencia alélica	s.d	Diversidad génica
11	23	0.3285	0.056	0.7242
12	2	0.0285	0.019	
13	2	0.0285	0.019	
14	27	0.3857	0.058	
15	8	0.1143	0.038	
16	4	0.0571	0.027	
17	1	0.0143	0.014	
19	2	0.0285	0.019	
20	1	0.0142	0.014	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.47. Frecuencias alélicas del sistema DYS 385 para la población italiana de Liguria.

DYS 385	Recuento	Frecuencia alélica	s.d	Diversidad génica
11	17	0.1848	0.040	0.8446
12	10	0.1087	0.032	
13	19	0.2065	0.042	
14	20	0.2174	0.043	
15	8	0.0869	0.029	
16	7	0.0761	0.027	
17	6	0.0652	0.025	
18	4	0.0434	0.021	
20	1	0.0108	0.010	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.48. Frecuencias alélicas del sistema DYS 385 para la población italiana de Lombardía.

DYS 385	Recuento	Frecuencia alélica	s.d	Diversidad génica
10	1	0.01	9.94 10⁻³	0.8158
11	27	0.27	0.044	
12	6	0.06	0.023	
13	9	0.09	0.028	
14	28	0.28	0.044	
15	11	0.11	0.031	
15.1	1	0.01	9.94 10⁻³	
16	6	0.06	0.023	
17	4	0.04	0.019	
18	6	0.06	0.023	
21	1	0.01	9.94 10⁻³	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.49. Frecuencias alélicas del sistema DYS 385 para la población italiana de Piamonte.

DYS 385	Recuento	Frecuencia alélica	s.d	Diversidad génica
11	5	0.1923	0.077	0.8048
12	2	0.0769	0.052	
14	7	0.2692	0.086	
15	6	0.2307	0.082	
16	4	0.1538	0.070	
17	1	0.0384	0.037	
19	1	0.0384	0.037	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

DYS385

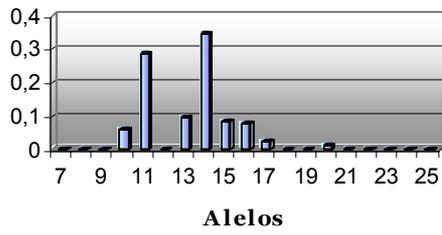


Figura 4.43: Frecuencias alélicas de Cataluña

DYS385

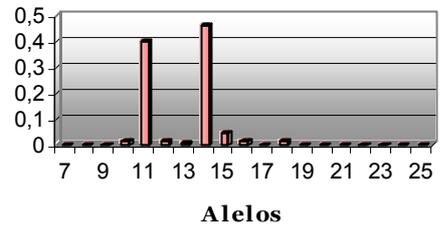


Figura 4.44: Frecuencias alélicas de Baleares

DYS385

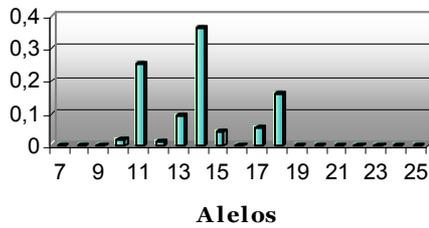


Figura 4.45: Frecuencias alélicas de Valencia

DYS385

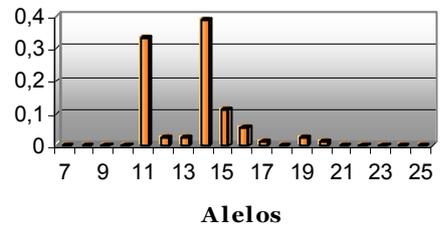


Figura 4.46: Frecuencias alélicas de Francia

DYS385

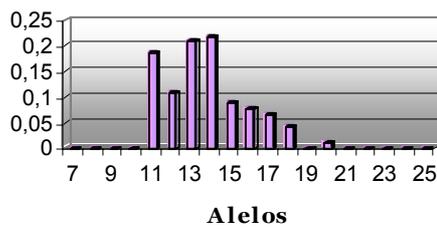


Figura 4.47: Frecuencias alélicas de Liguria

DYS385

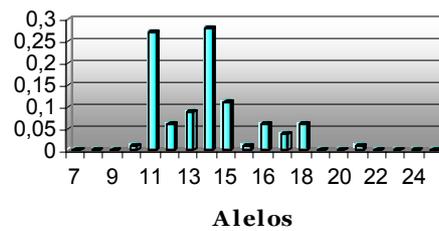


Figura 4.48: Frecuencias alélicas de Lombardía

DYS385

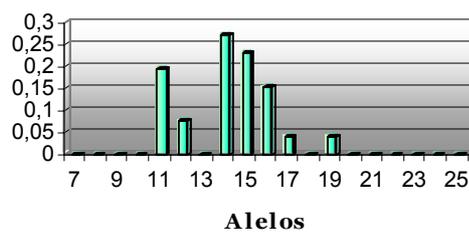


Figura 4.49: Frecuencias alélicas de Piemonte

Tabla 4.50. Frecuencias alélicas del sistema DYS 393 para la población catalana.

DYS 393	Recuento	Frecuencia alélica	s.d	Diversidad génica
12	30	0.24	0.038	0.5184
13	80	0.64	0.042	
14	15	0.12	0.029	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.51. Frecuencias alélicas del sistema DYS 393 para la población balear.

DYS 393	Recuento	Frecuencia alélica	s.d	Diversidad génica
12	2	0.04	0.027	0.2208
13	44	0.88	0.045	
14	2	0.04	0.027	
15	2	0.04	0.027	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.52. Frecuencias alélicas del sistema DYS 393 para la población valenciana.

DYS 393	Recuento	Frecuencia alélica	s.d	Diversidad génica
12	19	0.4318	0.074	0.4908
13	25	0.5682	0.074	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.53. Frecuencias alélicas del sistema DYS 393 para la población francesa.

DYS 393	Recuento	Frecuencia alélica	s.d	Diversidad génica
12	3	0.0857	0.047	0.2988
13	29	0.8285	0.063	
14	3	0.0857	0.047	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.54. Frecuencias alélicas del sistema DYS 393 para la población italiana de Liguria.

DYS 393	Recuento	Frecuencia alélica	s.d	Diversidad génica
11	1	0.0217	0.021	0.5322
12	8	0.1739	0.055	
13	30	0.6521	0.070	
14	3	0.0652	0.036	
15	4	0.0869	0.041	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.55. Frecuencias alélicas del sistema DYS 393 para la población italiana de Lombardía.

DYS 393	Recuento	Frecuencia alélica	s.d	Diversidad génica
12	10	0.02	0.019	0.5304
13	32	0.64	0.067	
14	7	0.14	0.049	
15	1	0.02	0.019	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.56. Frecuencias alélicas del sistema DYS 393 para la población italiana de Piemonte.

DYS 393	Recuento	Frecuencia alélica	s.d	Diversidad génica
12	1	0.0769	0.073	0.5208
13	8	0.6154	0.134	
14	4	0.3077	0.128	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

DYS 393

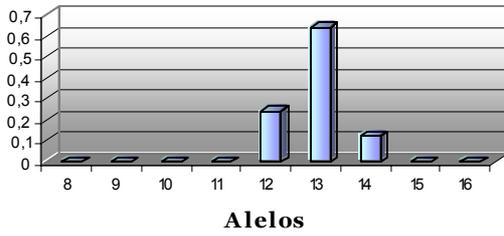


Figura 4.50: Frecuencias alélicas de Cataluña

DYS 393

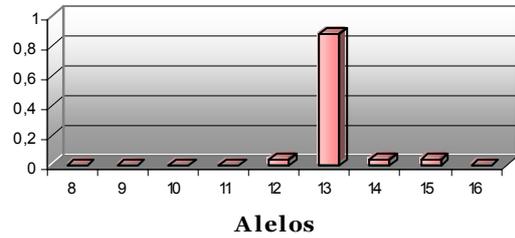


Figura 4.51: Frecuencias alélicas de Baleares

DYS 393

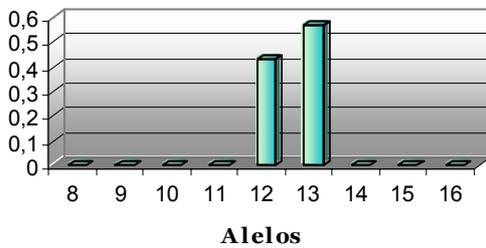


Figura 4.52: Frecuencias alélicas de Valencia

DYS 393

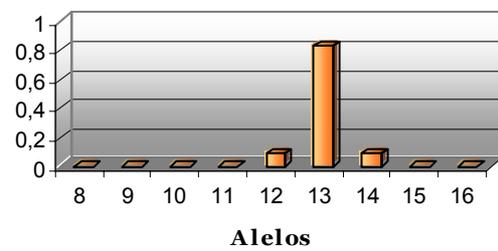


Figura 4.53: Frecuencias alélicas de Francia

DYS 393

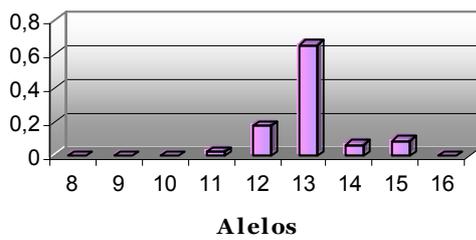


Figura 4.54: Frecuencias alélicas de Liguria

DYS 393

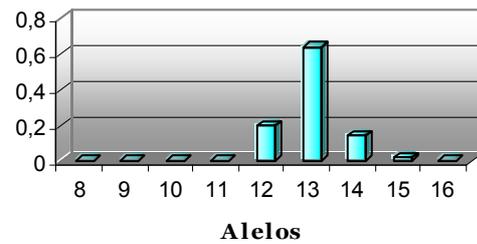


Figura 4.55: Frecuencias alélicas de Lombardía

DYS 393

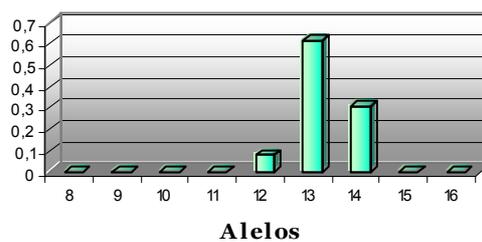


Figura 4.56: Frecuencias alélicas de Piemonte

Tabla 4.57. Frecuencias alélicas del sistema DYS 391 para la población catalana.

DYS 391	Recuento	Frecuencia alélica	s.d	Diversidad génica
10	89	0.712	0.040	0.4102
11	36	0.288	0.040	

Diversidad génica computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.58. Frecuencias alélicas del sistema DYS 391 para la población balear.

DYS 391	Recuento	Frecuencia alélica	s.d	Diversidad génica
10	6	0.12	0.045	0.2112
11	44	0.88	0.045	

Diversidad génica computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.59. Frecuencias alélicas del sistema DYS 391 para la población valenciana.

DYS 391	Recuento	Frecuencia alélica	s.d	Diversidad génica
10	27	0.6136	0.073	0.4743
11	17	0.3863	0.073	

Diversidad génica computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.60. Frecuencias alélicas del sistema DYS 391 para la población francesa.

DYS 391	Recuento	Frecuencia alélica	s.d	Diversidad génica
9	1	0.0285	0.028	0.5209
10	19	0.5428	0.084	
11	15	0.4285	0.083	

Diversidad génica computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.61. Frecuencias alélicas del sistema DYS 391 para la población italiana de Liguria.

DYS 391	Recuento	Frecuencia alélica	s.d	Diversidad génica
10	28	0.6087	0.071	0.5293
11	14	0.3043	0.067	
12	4	0.0869	0.041	

Diversidad génica computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.62. Frecuencias alélicas del sistema DYS 391 para la población italiana de Lombardía.

DYS 391	Recuento	Frecuencia alélica	s.d	Diversidad génica
9	1	0.02	0.019	0.5696
10	21	0.42	0.069	
11	25	0.5	0.070	
12	3	0.06	0.033	

Diversidad génica computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.63. Frecuencias alélicas del sistema DYS 391 para la población italiana de Piamonte.

DYS 391	Recuento	Frecuencia alélica	s.d	Diversidad génica
10	10	0.7692	0.116	0.3551
11	3	0.2307	0.116	

Diversidad génica computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

DYS 391

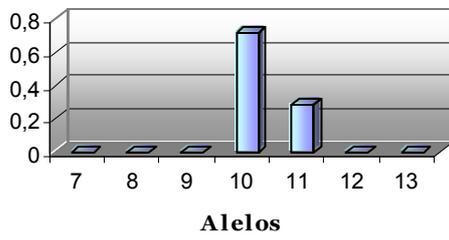


Figura 4.57: Frecuencias alélicas de Cataluña

DYS 391

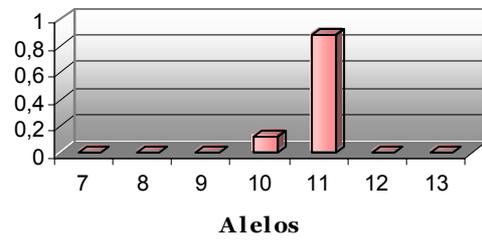


Figura 4.58: Frecuencias alélicas de Balears

DYS 391

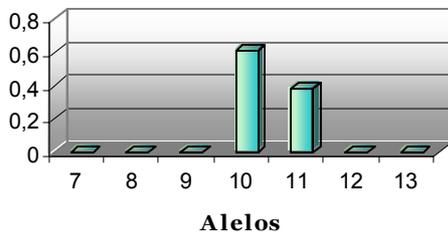


Figura 4.59: Frecuencias alélicas de Valencia

DYS 391

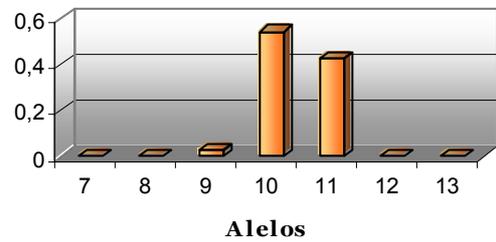


Figura 4.60: Frecuencias alélicas de Francia

DYS 391

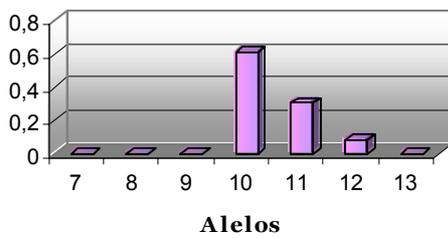


Figura 4.61: Frecuencias alélicas de Liguria

DYS 391

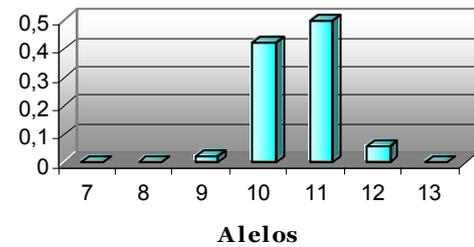


Figura 4.62: Frecuencias alélicas de Lombardía

DYS 391

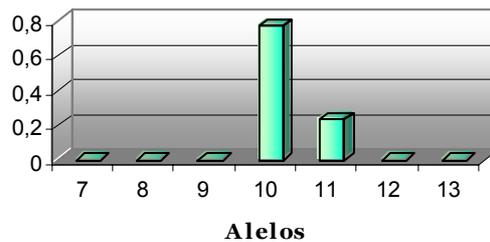


Figura 4.63: Frecuencias alélicas de Piemonte

Tabla 4.64. Frecuencias alélicas del sistema DYS 439 para la población catalana.

DYS 439	Recuento	Frecuencia alélica	s.d	Diversidad génica
10	3	0.024	0.013	0.6647
11	33	0.264	0.039	
12	54	0.432	0.044	
13	35	0.28	0.040	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.65. Frecuencias alélicas del sistema DYS 439 para la población balear.

DYS 439	Recuento	Frecuencia alélica	s.d	Diversidad génica
10	1	0.02	0.019	0.2152
11	5	0.1	0.042	
12	44	0.88	0.045	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.66. Frecuencias alélicas del sistema DYS 439 para la población valenciana.

DYS 439	Recuento	Frecuencia alélica	s.d	Diversidad génica
10	4	0.0909	0.043	0.4948
11	29	0.6591	0.071	
12	11	0.25	0.065	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.67. Frecuencias alélicas del sistema DYS 439 para la población francesa.

DYS 439	Recuento	Frecuencia alélica	s.d	Diversidad génica
10	1	0.0285	0.028	0.5927
11	11	0.3143	0.078	
12	19	0.5428	0.084	
13	4	0.1143	0.053	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.68. Frecuencias alélicas del sistema DYS 439 para la población italiana de Liguria.

DYS 439	Recuento	Frecuencia alélica	s.d	Diversidad génica
11	12	0.2608	0.064	0.4991
12	30	0.6521	0.070	
13	4	0.0869	0.041	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.69. Frecuencias alélicas del sistema DYS 439 para la población italiana de Lombardía.

DYS 439	Recuento	Frecuencia alélica	s.d	Diversidad génica
10	1	0.02	0.019	0.5856
11	15	0.3	0.064	
12	28	0.56	0.070	
13	5	0.1	0.042	
14	1	0.02	0.019	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.70. Frecuencias alélicas del sistema DYS 439 para la población italiana de Piemonte.

DYS 439	Recuento	Frecuencia alélica	s.d	Diversidad génica
11	6	0.4615	0.138	0.5681
12	6	0.4615	0.138	
13	1	0.0769	0.073	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

DYS 439

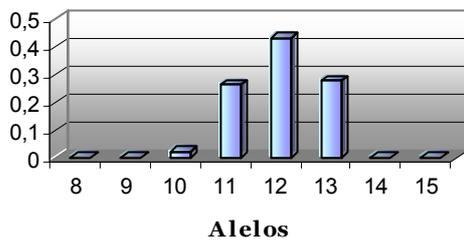


Figura 4.64: Frecuencias alélicas de Cataluña

DYS 439

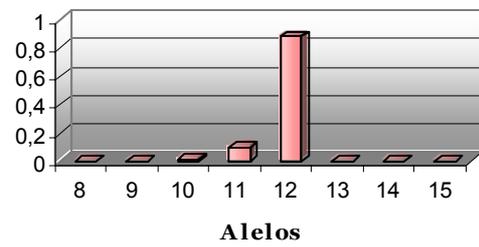


Figura 4.65: Frecuencias alélicas de Baleares

DYS 439

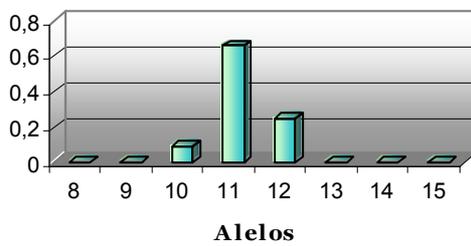


Figura 4.66: Frecuencias alélicas de Valencia

DYS 439

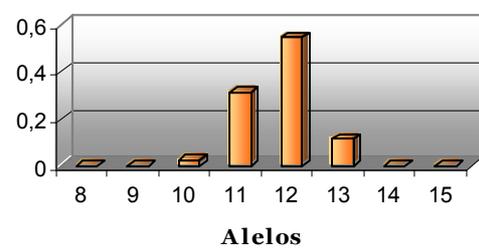


Figura 4.67: Frecuencias alélicas de Francia

DYS 439

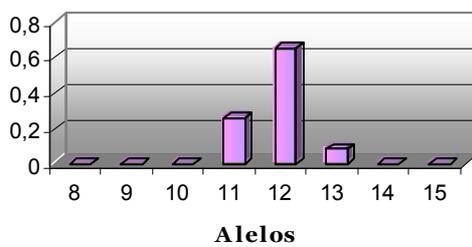


Figura 4.68: Frecuencias alélicas de Liguria

DYS 439

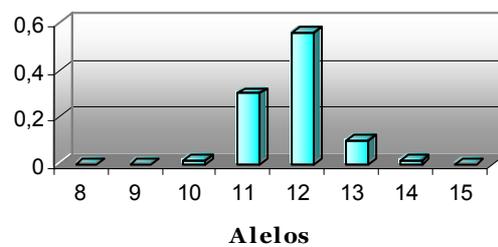


Figura 4.69: Frecuencias alélicas de Lombardía

DYS 439

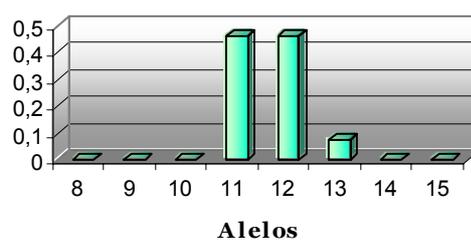


Figura 4.70: Frecuencias alélicas de Piemonte

Tabla 4.71. Frecuencias alélicas del sistema DYS 635 para la población catalana.

DYS 635	Recuento	Frecuencia alélica	s.d	Diversidad génica
<20	2	0.016	0.011	0.6051
20	3	0.024	0.013	
21	25	0.2	0.035	
22	5	0.04	0.017	
23	72	0.576	0.044	
24	18	0.144	0.031	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.72. Frecuencias alélicas del sistema DYS 635 para la población balear.

DYS 635	Recuento	Frecuencia alélica	s.d	Diversidad génica
20	4	0.08	0.038	0.2848
21	2	0.04	0.027	
23	42	0.84	0.051	
24	2	0.04	0.027	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.73. Frecuencias alélicas del sistema DYS 635 para la población valenciana.

DYS 635	Recuento	Frecuencia alélica	s.d	Diversidad génica
<20	5	0.1136	0.047	0.6116
20	14	0.3182	0.070	
23	23	0.5227	0.075	
24	1	0.0227	0.022	
25	1	0.0227	0.022	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.74. Frecuencias alélicas del sistema DYS 635 para la población francesa.

DYS 635	Recuento	Frecuencia alélica	s.d	Diversidad génica
20	1	0.0285	0.028	0.5926
21	4	0.1143	0.053	
22	5	0.1428	0.059	
23	21	0.6	0.082	
24	4	0.1143	0.053	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.75. Frecuencias alélicas del sistema DYS 635 para la población italiana de Liguria.

DYS 635	Recuento	Frecuencia alélica	s.d	Diversidad génica
20	3	0.0652	0.036	0.6484
21	9	0.1956	0.058	
22	4	0.0869	0.041	
23	25	0.5435	0.073	
24	3	0.0652	0.036	
25	2	0.0435	0.030	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.76. Frecuencias alélicas del sistema DYS 635 para la población italiana de Lombardía.

DYS 635	Recuento	Frecuencia alélica	s.d	Diversidad génica
20	3	0.06	0.033	0.5136
21	4	0.08	0.038	
22	5	0.1	0.042	
23	34	0.68	0.065	
24	3	0.06	0.033	
25	1	0.02	0.019	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.77. Frecuencias alélicas del sistema DYS 635 para la población italiana de Piamonte.

DYS 635	Recuento	Frecuencia alélica	s.d	Diversidad génica
19	1	0.0769	0.073	0.7694
20	2	0.1538	0.100	
21	2	0.1538	0.100	
22	2	0.1538	0.100	
23	5	0.3846	0.134	
25	1	0.0769	0.073	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

DYS 635

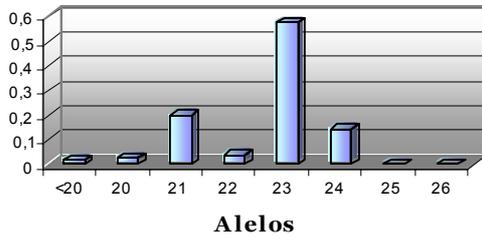


Figura 4.71: Frecuencias alélicas de Cataluña

DYS 635

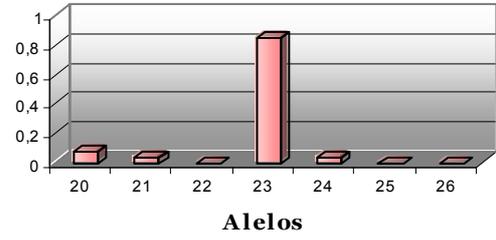


Figura 4.72: Frecuencias alélicas de Baleares

DYS 635

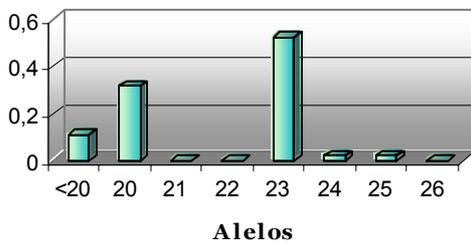


Figura 4.73: Frecuencias alélicas de Valencia

DYS 635

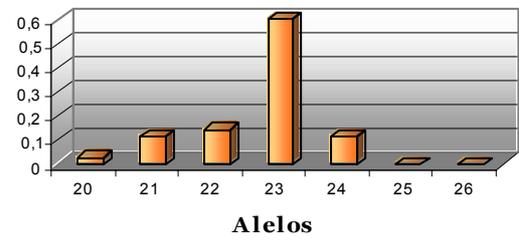


Figura 4.74: Frecuencias alélicas de Francia

DYS 635

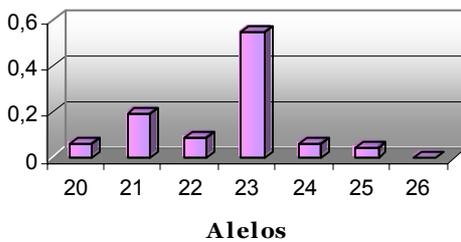


Figura 4.75: Frecuencias alélicas de Liguria

DYS 635

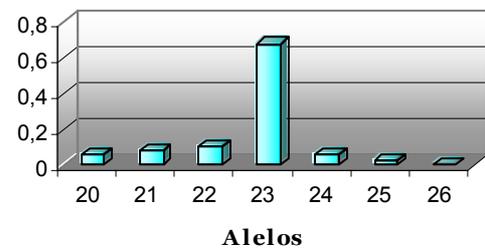


Figura 4.76: Frecuencias alélicas de Lombardía

DYS 635

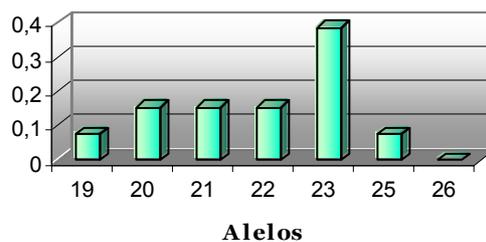


Figura 4.77: Frecuencias alélicas de Piamonte

Tabla 4.78. Frecuencias alélicas del sistema DYS 392 para la población catalana.

DYS 392	Recuento	Frecuencia alélica	s.d	Diversidad génica
11	40	0.32	0.041	0.4770
12	3	0.024	0.013	
13	81	0.648	0.042	
14	1	$8 \cdot 10^{-3}$	$7.96 \cdot 10^{-3}$	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.79. Frecuencias alélicas del sistema DYS 392 para la población balear.

DYS 392	Recuento	Frecuencia alélica	s.d	Diversidad génica
11	5	0.1	0.042	0.2152
12	1	0.02	0.019	
13	44	0.88	0.045	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.80. Frecuencias alélicas del sistema DYS 392 para la población valenciana.

DYS 392	Recuento	Frecuencia alélica	s.d	Diversidad génica
11	19	0.4318	0.074	0.4908
13	25	0.5682	0.074	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.81. Frecuencias alélicas del sistema DYS 392 para la población francesa.

DYS 392	Recuento	Frecuencia alélica	s.d	Diversidad génica
11	9	0.2571	0.073	0.4228
13	25	0.7143	0.076	
14	1	0.0285	0.028	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.82. Frecuencias alélicas del sistema DYS 392 para la población italiana de Liguria.

DYS 392	Recuento	Frecuencia alélica	s.d	Diversidad génica
11	17	0.3695	0.071	0.6579
12	3	0.0652	0.036	
13	20	0.4348	0.073	
14	5	0.1087	0.045	
16	1	0.0217	0.021	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.83. Frecuencias alélicas del sistema DYS 392 para la población italiana de Lombardía.

DYS 392	Recuento	Frecuencia alélica	s.d	Diversidad génica
11	16	0.32	0.065	0.5336
12	3	0.06	0.033	
13	30	0.6	0.069	
13.1	1	0.02	0.019	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.84. Frecuencias alélicas del sistema DYS 392 para la población italiana de Piamonte.

DYS 392	Recuento	Frecuencia alélica	s.d	Diversidad génica
11	6	0.4615	0.138	0.6155
12	2	0.1538	0.100	
13	5	0.3846	0.134	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

DYS 392

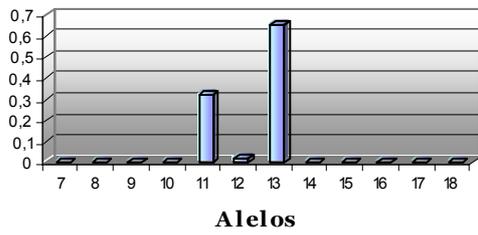


Figura 4.78: Frecuencias alélicas de Cataluña

DYS 392

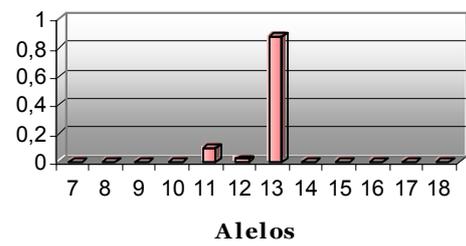


Figura 4.79: Frecuencias alélicas de Balears

DYS 392

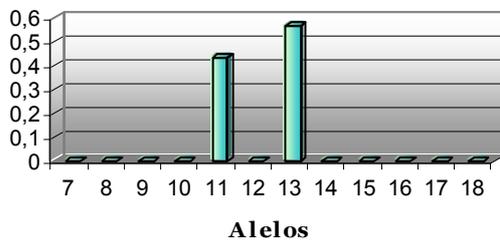


Figura 4.80: Frecuencias alélicas de Valencia

DYS 392

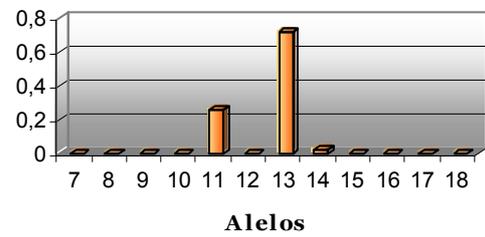


Figura 4.81: Frecuencias alélicas de Francia

DYS 392

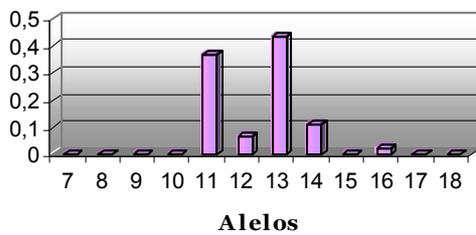


Figura 4.82: Frecuencias alélicas de Liguria

DYS 392

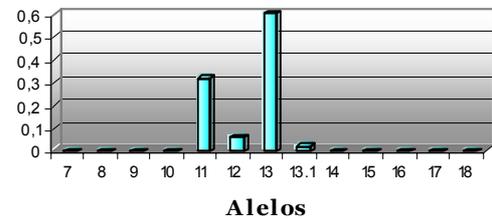


Figura 4.83: Frecuencias alélicas de Lombardia

DYS 392

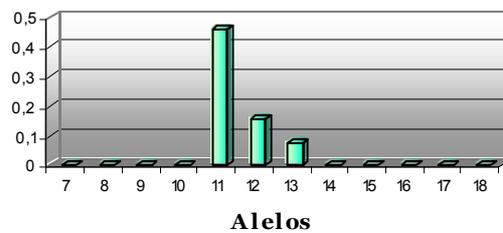


Figura 4.84: Frecuencias alélicas de Piemonte

Tabla 4.85. Frecuencias alélicas del sistema Y GATA H4 para la población catalana.

Y GATA H4	Recuento	Frecuencia alélica	s.d	Diversidad génica
11	57	0.456	0.044	0.5047
12	67	0.536	0.044	
13	1	$8 \cdot 10^{-3}$	$7.96 \cdot 10^{-3}$	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.86. Frecuencias alélicas del sistema Y GATA H4 para la población balear.

Y GATA H4	Recuento	Frecuencia alélica	s.d	Diversidad génica
10	1	0.02	0.019	0.3336
11	40	0.8	0.056	
12	8	0.16	0.051	
13	1	0.02	0.019	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.87. Frecuencias alélicas del sistema Y GATA H4 para la población valenciana.

Y GATA H4	Recuento	Frecuencia alélica	s.d	Diversidad génica
11	31	0.7045	0.068	0.4165
12	13	0.2954	0.068	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.88. Frecuencias alélicas del sistema Y GATA H4 para la población francesa.

Y GATA H4	Recuento	Frecuencia alélica	s.d	Diversidad génica
11	19	0.5428	0.084	0.5208
12	15	0.4285	0.083	
13	1	0.0285	0.028	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.89. Frecuencias alélicas del sistema Y GATA H4 para la población italiana de Liguria.

Y GATA H4	Recuento	Frecuencia alélica	s.d	Diversidad génica
9	1	0.0217	0.021	0.6115
10	4	0.0869	0.041	
11	24	0.5217	0.073	
12	15	0.3261	0.069	
13	2	0.0435	0.030	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.90. Frecuencias alélicas del sistema Y GATA H4 para la población italiana de Lombardía.

Y GATA H4	Recuento	Frecuencia alélica	s.d	Diversidad génica
10	5	0.1	0.042	0.6088
11	12	0.24	0.060	
12	28	0.56	0.070	
13	5	0.1	0.042	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.91. Frecuencias alélicas del sistema Y GATA H4 para la población italiana de Piamonte.

Y GATA H4	Recuento	Frecuencia alélica	s.d	Diversidad génica
11	7	0.5384	0.138	0.5563
12	5	0.3846	0.134	
13	1	0.0769	0.073	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Y GATA H4

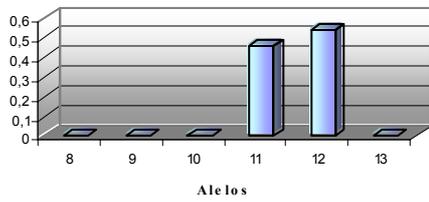


Figura 4.85: Frecuencias alélicas de Cataluña

Y GATA H4

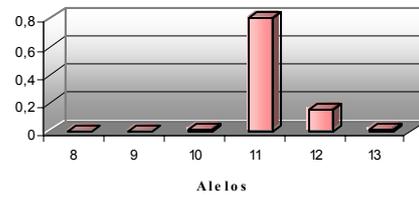


Figura 4.86: Frecuencias alélicas de Baleares

Y GATA H4

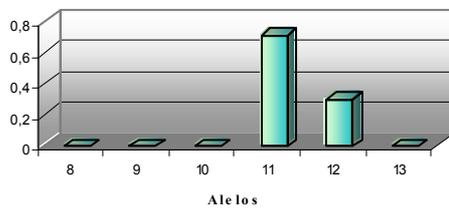


Figura 4.87: Frecuencias alélicas de Valencia

Y GATA H4

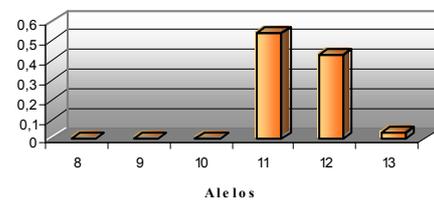


Figura 4.88: Frecuencias alélicas de Francia

Y GATA H4

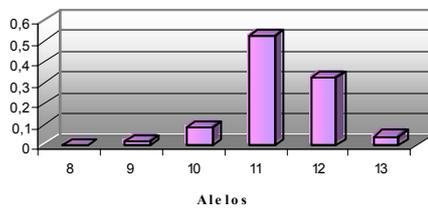


Figura 4.89: Frecuencias alélicas de Liguria

Y GATA H4

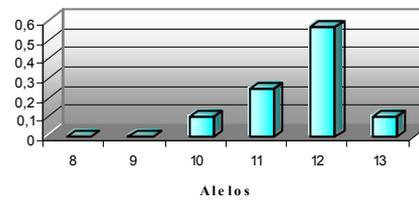


Figura 4.90: Frecuencias alélicas de Lombardía

Y GATA H4

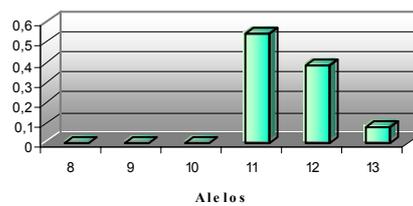


Figura 4.91: Frecuencias alélicas de Piemonte

Tabla 4.92. Frecuencias alélicas del sistema DYS 437 para la población catalana.

DYS 437	Recuento	Frecuencia alélica	s.d	Diversidad génica
14	46	0.368	0.043	0.5817
15	65	0.52	0.044	
16	14	0.112	0.028	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.93. Frecuencias alélicas del sistema DYS 437 para la población balear.

DYS 437	Recuento	Frecuencia alélica	s.d	Diversidad génica
14	44	0.88	0.045	0.2184
15	3	0.06	0.033	
16	3	0.06	0.033	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.94. Frecuencias alélicas del sistema DYS 437 para la población valenciana.

DYS 437	Recuento	Frecuencia alélica	s.d	Diversidad génica
14	26	0.5909	0.074	0.5507
15	13	0.2954	0.068	
16	5	0.1136	0.053	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.95. Frecuencias alélicas del sistema DYS 437 para la población francesa.

DYS 437	Recuento	Frecuencia alélica	s.d	Diversidad génica
14	11	0.3143	0.078	0.6074
15	18	0.5142	0.084	
16	6	0.1714	0.063	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.96. Frecuencias alélicas del sistema DYS 437 para la población italiana de Liguria.

DYS 437	Recuento	Frecuencia alélica	s.d	Diversidad génica
14	12	0.2608	0.064	0.5823
15	26	0.5652	0.073	
16	8	0.1739	0.055	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.97. Frecuencias alélicas del sistema DYS 437 para la población italiana de Lombardía.

DYS 437	Recuento	Frecuencia alélica	s.d	Diversidad génica
14	15	0.3	0.064	0.5768
15	28	0.56	0.070	
16	7	0.14	0.049	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.98. Frecuencias alélicas del sistema DYS 437 para la población italiana de Piemonte.

DYS 437	Recuento	Frecuencia alélica	s.d	Diversidad génica
14	6	0.4615	0.138	0.6155
15	5	0.3846	0.134	
16	2	0.1538	0.100	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

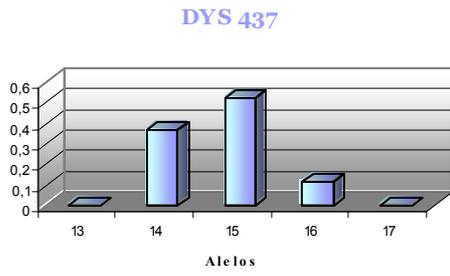


Figura 4.92: Frecuencias alélicas de Cataluña

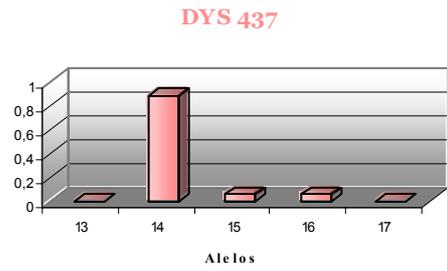


Figura 4.93: Frecuencias alélicas de Baleares

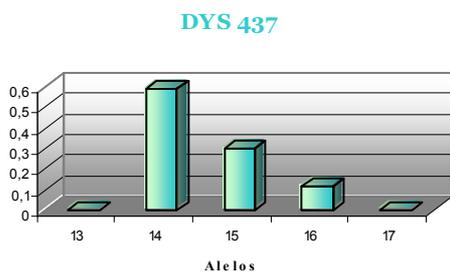


Figura 4.94: Frecuencias alélicas de Valencia

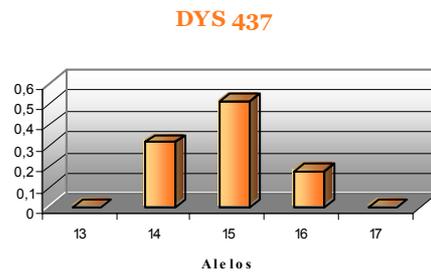


Figura 4.95: Frecuencias alélicas de Francia

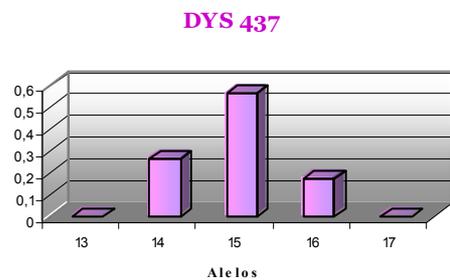


Figura 4.96: Frecuencias alélicas de Liguria

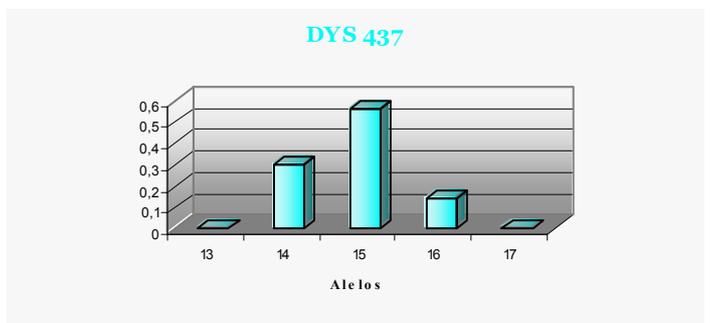


Figura 4.97: Frecuencias alélicas de Lombardía

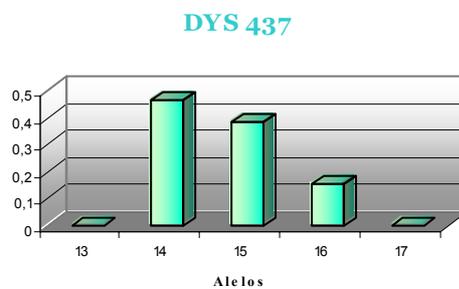


Figura 4.98: Frecuencias alélicas de Piemonte

Tabla 4.99. Frecuencias alélicas del sistema DYS 438 para la población catalana.

DYS 438	Recuento	Frecuencia alélica	s.d	Diversidad génica
9	22	0.176	0.034	0.5808
10	14	0.112	0.028	
11	9	0.072	0.023	
12	76	0.608	0.043	
13	4	0.032	0.015	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.100. Frecuencias alélicas del sistema DYS 438 para la población balear.

DYS 438	Recuento	Frecuencia alélica	s.d	Diversidad génica
10	6	0.12	0.045	0.2456
12	43	0.86	0.049	
13	1	0.02	0.019	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.101. Frecuencias alélicas del sistema DYS 438 para la población valenciana.

DYS 438	Recuento	Frecuencia alélica	s.d	Diversidad génica
9	5	0.1136	0.047	0.2456
10	14	0.3182	0.070	
12	25	0.5682	0.074	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.102. Frecuencias alélicas del sistema DYS 438 para la población francesa.

DYS 438	Recuento	Frecuencia alélica	s.d	Diversidad génica
9	1	0.0285	0.028	0.4228
10	9	0.2571	0.073	
12	25	0.7143	0.076	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.103. Frecuencias alélicas del sistema DYS 438 para la población italiana de Liguria.

DYS 438	Recuento	Frecuencia alélica	s.d	Diversidad génica
9	8	0.1739	0.055	0.6276
10	12	0.2608	0.064	
11	2	0.0435	0.030	
12	24	0.5217	0.073	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.104. Frecuencias alélicas del sistema DYS 438 para la población italiana de Lombardía.

DYS 438	Recuento	Frecuencia alélica	s.d	Diversidad génica
8	2	0.04	0.027	0.6424
9	3	0.06	0.033	
10	14	0.28	0.063	
11	2	0.04	0.027	
12	26	0.52	0.070	
13	1	0.02	0.019	
>13	2	0.04	0.027	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.105. Frecuencias alélicas del sistema DYS 438 para la población italiana de Piemonte.

DYS 438	Recuento	Frecuencia alélica	s.d	Diversidad génica
9	1	0.0769	0.073	0.6747
10	5	0.3846	0.0134	
11	2	0.1538	0.0100	
12	5	0.3846	0.0134	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

DYS 438

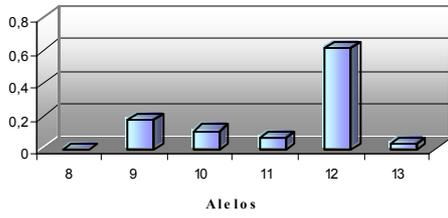


Figura 4.99: Frecuencias alélicas de Cataluña

DYS 438

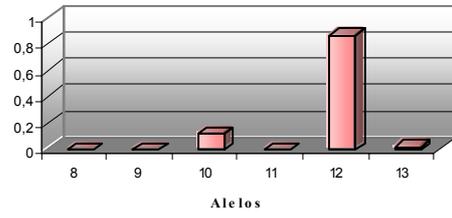


Figura 4.100: Frecuencias alélicas de Baleares

DYS 438

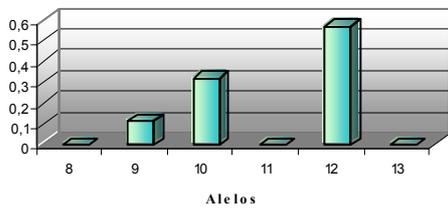


Figura 4.101: Frecuencias alélicas de Valencia

DYS 438

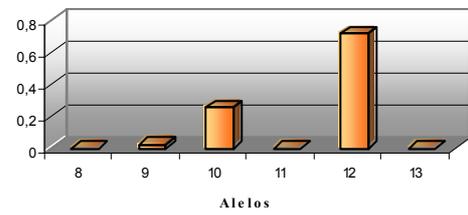


Figura 4.102: Frecuencias alélicas de Francia

DYS 438

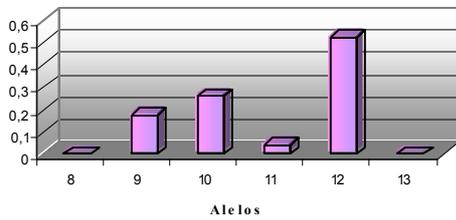


Figura 4.103: Frecuencias alélicas de Liguria

DYS 438

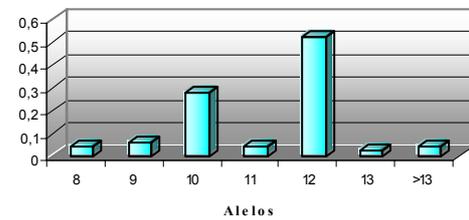


Figura 4.104: Frecuencias alélicas de Lombardía

DYS 438

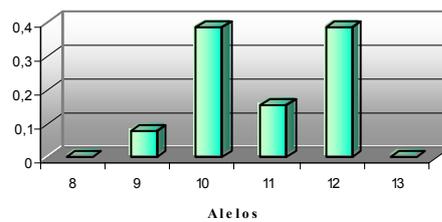


Figura 4.105: Frecuencias alélicas de Piemonte

Tabla 4.106. Frecuencias alélicas del sistema DYS 448 para la población catalana.

DYS 448	Recuento	Frecuencia alélica	s.d	Diversidad génica
18	19	0.152	0.032	0.5652
19	77	0.616	0.043	
20	8	0.064	0.021	
21	21	0.168	0.033	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.107. Frecuencias alélicas del sistema DYS 448 para la población balear.

DYS 448	Recuento	Frecuencia alélica	s.d	Diversidad génica
18	41	0.82	0.054	0.3112
19	6	0.12	0.045	
20	1	0.02	0.019	
21	2	0.04	0.027	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.108. Frecuencias alélicas del sistema DYS 448 para la población valenciana.

DYS 448	Recuento	Frecuencia alélica	s.d	Diversidad génica
18	12	0.2727	0.067	0.6571
19	18	0.4091	0.074	
20	14	0.3182	0.070	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.109. Frecuencias alélicas del sistema DYS 448 para la población francesa.

DYS 448	Recuento	Frecuencia alélica	s.d	Diversidad génica
17	1	0.0285	0.028	0.6563
18	7	0.2	0.067	
19	17	0.4857	0.084	
20	9	0.2571	0.073	
21	1	0.0285	0.028	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.110. Frecuencias alélicas del sistema DYS 448 para la población italiana de Liguria.

DYS 448	Recuento	Frecuencia alélica	s.d	Diversidad génica
18	6	0.1304	0.049	0.5889
19	27	0.5869	0.072	
20	10	0.2174	0.060	
21	2	0.0435	0.030	
22	1	0.0217	0.021	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.111. Frecuencias alélicas del sistema DYS 448 para la población italiana de Lombardía.

DYS 448	Recuento	Frecuencia alélica	s.d	Diversidad génica
19	31	0.62	0.068	0.5376
20	13	0.26	0.062	
21	5	0.1	0.042	
22	1	0.02	0.019	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.112. Frecuencias alélicas del sistema DYS 448 para la población italiana de Piemonte.

DYS 448	Recuento	Frecuencia alélica	s.d	Diversidad génica
19	6	0.4615	0.138	0.6392
20	4	0.3077	0.128	
21	3	0.2307	0.116	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

DYS 448

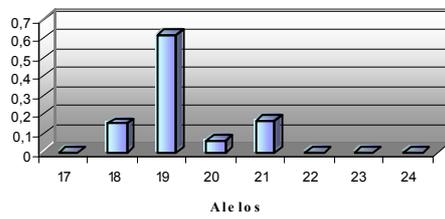


Figura 4.106: Frecuencias alélicas de Cataluña

DYS 448

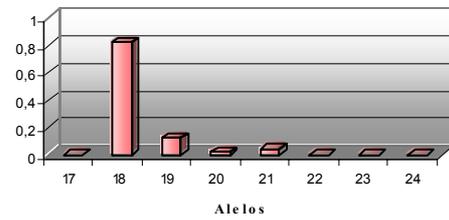


Figura 4.107: Frecuencias alélicas de Baleares

DYS 448

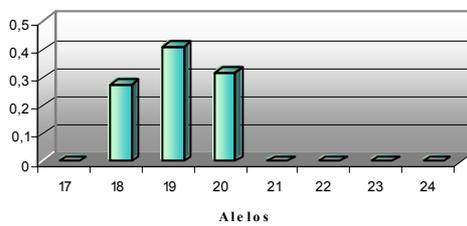


Figura 4.108: Frecuencias alélicas de Valencia

DYS 448

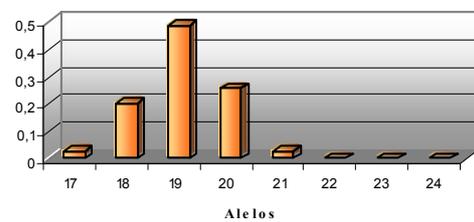


Figura 4.109: Frecuencias alélicas de Francia

DYS 448

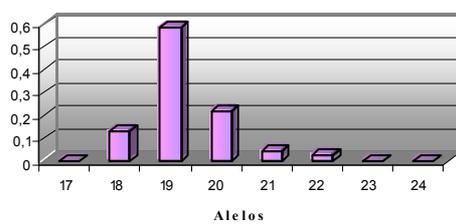


Figura 4.110: Frecuencias alélicas de Liguria

DYS 448

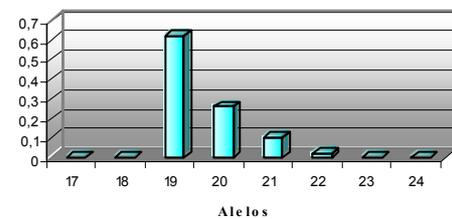


Figura 4.111: Frecuencias alélicas de Lombardia

DYS 448

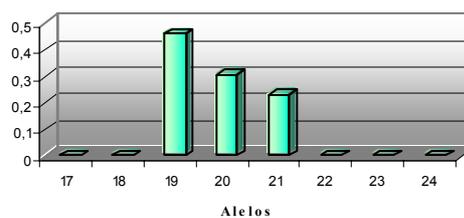


Figura 4.112: Frecuencias alélicas de Piemonte

Las frecuencias alélicas, así como la diversidad génica, de cada uno de los loci, del análisis genético-poblacional de los 58 individuos no emparentados, llevado a cabo en las zonas españolas anteriormente mencionadas, tomados de referencia, ya que no poseen el apellido Colón, junto con los 49 individuos de las zonas italianas se exponen en las siguientes tablas:

Tabla 4.113. Frecuencias alélicas del sistema DYS 456 para la población española.

DYS 456	Recuento	Frecuencia alélica	s.d	Diversidad génica
13	2	0.0345	0.023	0.6666
14	7	0.1207	0.042	
15	27	0.4655	0.065	
16	18	0.3103	0.060	
17	4	0.0689	0.033	

Diversidad génica computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.114. Frecuencias alélicas del sistema DYS 456 para la población italiana.

DYS 456	Recuento	Frecuencia alélica	s.d	Diversidad génica
14	8	0.1632	0.052	0.6747
15	24	0.4898	0.071	
16	11	0.2245	0.059	
17	4	0.0816	0.039	
18	2	0.0408	0.028	

Diversidad génica computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

DYS 456

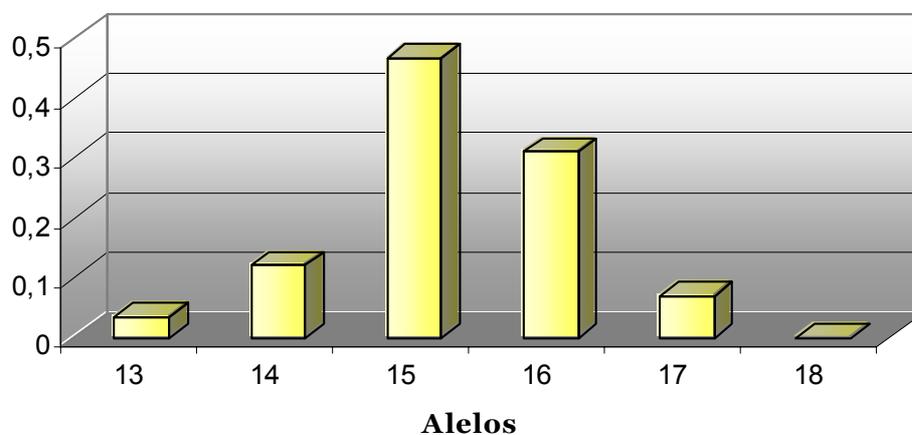


Figura 4.113: Frecuencias alélicas de la población de referencia española

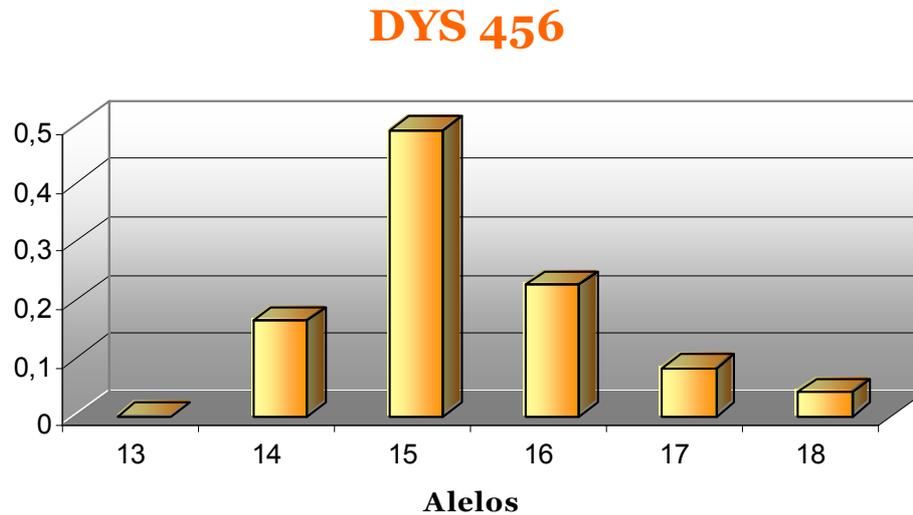


Figura 4.14: Frecuencias alélicas de la población de referencia italiana

Tabla 4.115. Frecuencias alélicas del sistema DYS 389I para la población española.

DYS 389I	Recuento	Frecuencia alélica	s.d	Diversidad génica
11	1	0.0172	0.017	0.5583
12	8	0.1379	0.045	
13	35	0.6034	0.064	
14	14	0.2413	0.056	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.116. Frecuencias alélicas del sistema DYS 389I para la población italiana.

DYS 389I	Recuento	Frecuencia alélica	s.d	Diversidad génica
11	1	0.0204	0.020	0.5231
12	4	0.0816	0.039	
13	32	0.6530	0.068	
14	10	0.2041	0.057	
15	2	0.0408	0.028	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

DYS 389I

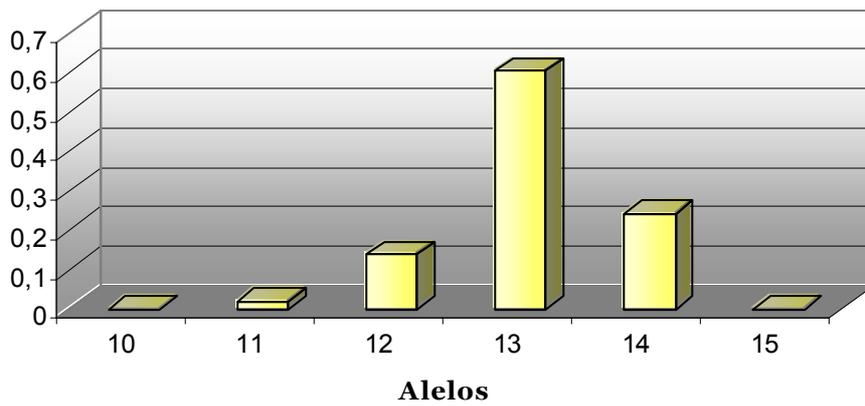


Figura 4.115: Frecuencias alélicas de la población de referencia española

DYS 389I

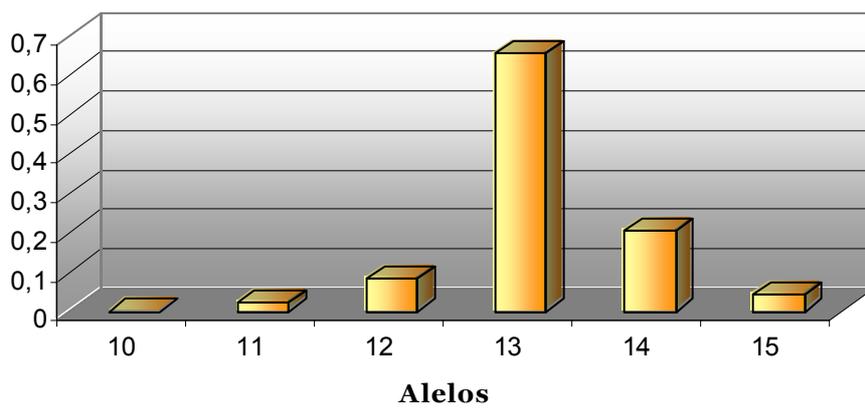


Figura 4.116: Frecuencias alélicas de la población de referencia italiana

Tabla 4.117. Frecuencias alélicas del sistema DYS 390 para la población española.

DYS 390	Recuento	Frecuencia alélica	s.d	Diversidad génica
21	1	0.0172	0.017	0.6397
22	2	0.0345	0.023	
23	15	0.2586	0.057	
24	30	0.5172	0.065	
25	9	0.1551	0.047	
27	1	0.0172	0.017	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.118. Frecuencias alélicas del sistema DYS 390 para la población italiana.

DYS 390	Recuento	Frecuencia alélica	s.d	Diversidad génica
22	6	0.1224	0.046	0.5741
23	12	0.2449	0.061	
24	29	0.5918	0.070	
25	1	0.0204	0.020	
26	1	0.0204	0.020	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

DYS 390

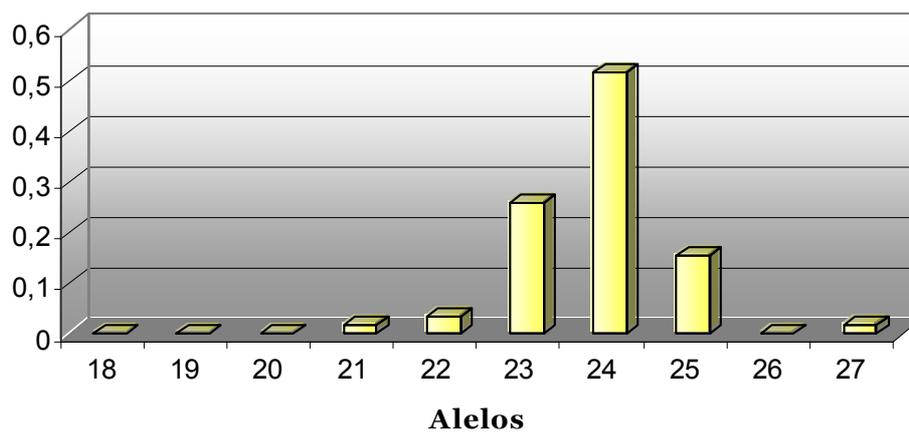


Figura 4.117: Frecuencias alélicas de la población de referencia española

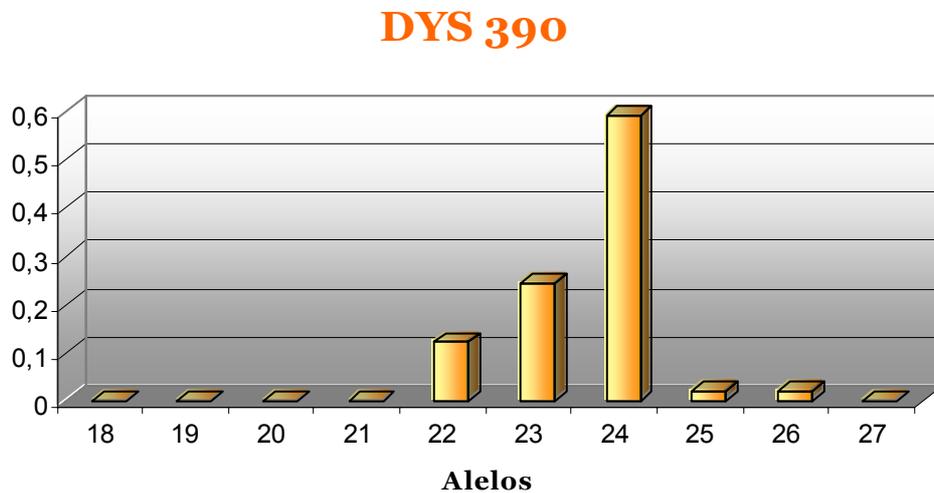


Figura 4.118: Frecuencias alélicas de la población de referencia italiana

Tabla 4.119. Frecuencias alélicas del sistema DYS 389II para la población española.

DYS 389II	Recuento	Frecuencia alélica	s.d	Diversidad génica
27	1	0.0172	0.017	0.6434
28	7	0.1207	0.042	
29	30	0.5172	0.065	
30	15	0.2586	0.057	
31	5	0.0862	0.036	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.120. Frecuencias alélicas del sistema DYS 389II para la población italiana.

DYS 389II	Recuento	Frecuencia alélica	s.d	Diversidad génica
27	1	0.0204	0.018	0.6865
28	3	0.0612	0.034	
29	23	0.4693	0.071	
30	13	0.2653	0.063	
31	6	0.1224	0.046	
32	3	0.0612	0.034	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

DYS 389II

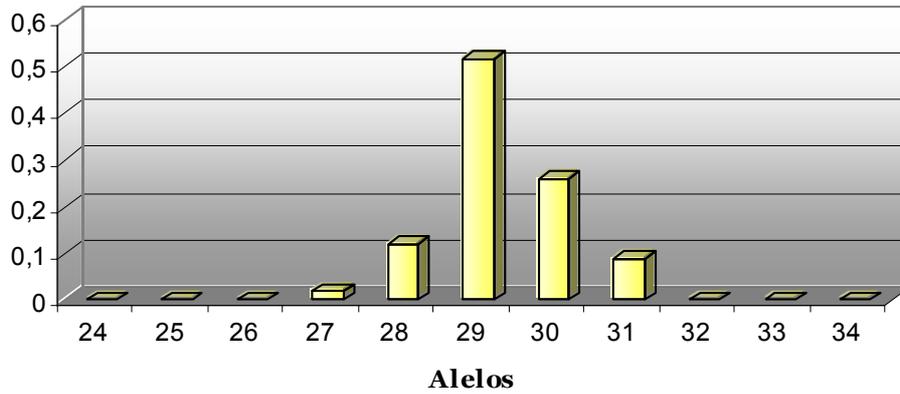


Figura 4.119: Frecuencias alélicas de la población de referencia española

DYS 389II

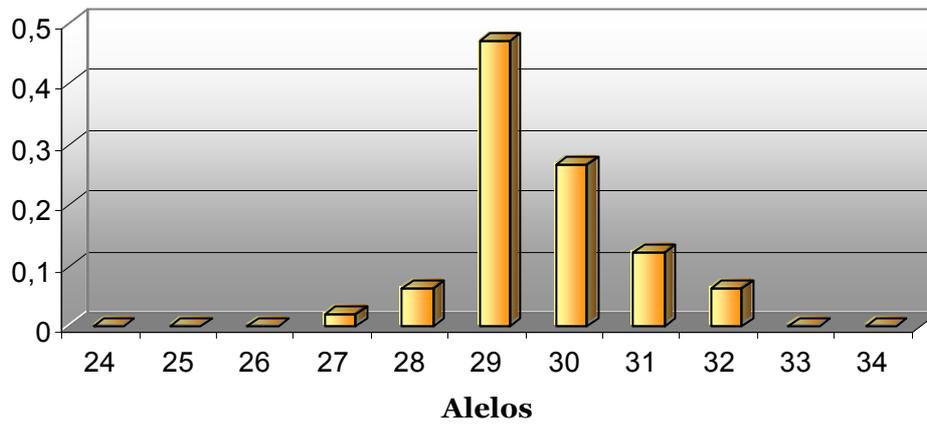


Figura 4.120: Frecuencias alélicas de la población de referencia italiana

Tabla 4.121. Frecuencias alélicas del sistema DYS 458 para la población española.

DYS 458	Recuento	Frecuencia alélica	s.d	Diversidad génica
15	9	0.1551	0.047	0.7461
16	11	0.1896	0.051	
17	19	0.3275	0.061	
18	17	0.2931	0.059	
19	1	0.0172	0.017	
>19	1	0.0172	0.017	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.122. Frecuencias alélicas del sistema DYS 458 para la población italiana.

DYS 458	Recuento	Frecuencia alélica	s.d	Diversidad génica
15	4	0.0816	0.039	0.7239
16	12	0.2449	0.061	
17	20	0.4081	0.070	
18	10	0.2041	0.057	
19	1	0.0204	0.020	
19.2	1	0.0204	0.020	
>20	1	0.0204	0.020	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

DYS 458

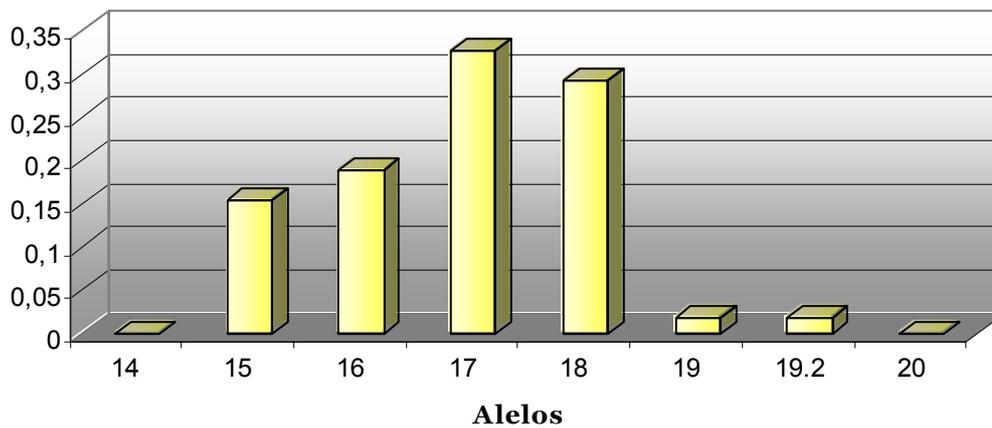


Figura 4.121: Frecuencias alélicas de la población de referencia española

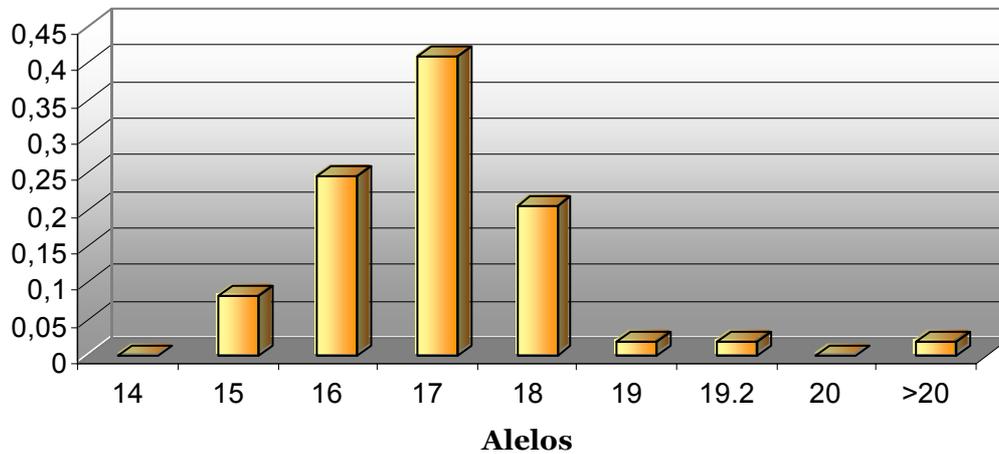
DYS 458

Figura 4.122: Frecuencias alélicas de la población de referencia italiana

Tabla 4.123. Frecuencias alélicas del sistema DYS 19 para la población italiana.

DYS 19	Recuento	Frecuencia alélica	s.d	Diversidad génica
12	1	0.0172	0.017	0.4127
13	2	0.0345	0.023	
14	43	0.7413	0.057	
15	11	0.1896	0.051	
16	1	0.0172	0.017	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.124. Frecuencias alélicas del sistema DYS 19 para la población italiana.

DYS 19	Recuento	Frecuencia alélica	s.d	Diversidad génica
13	5	0.1020	0.043	0.5007
14	33	0.6734	0.066	
15	9	0.1836	0.055	
16	2	0.0408	0.028	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

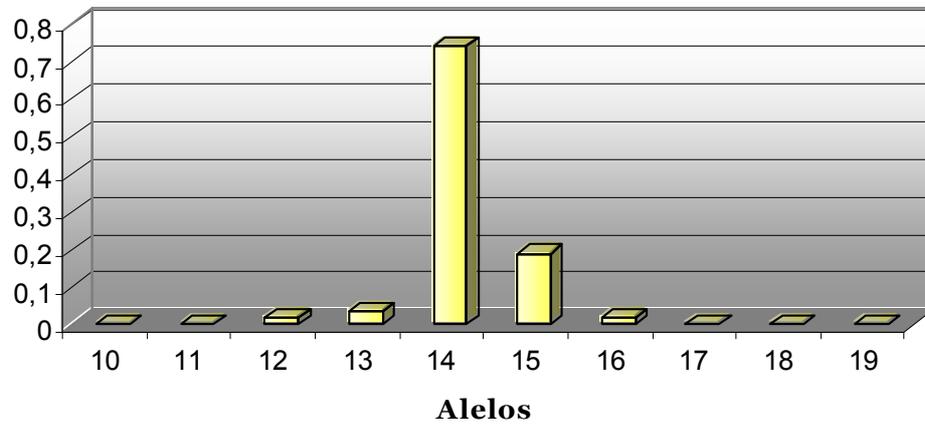
DYS 19

Figura 4.123: Frecuencias alélicas de la población de referencia española

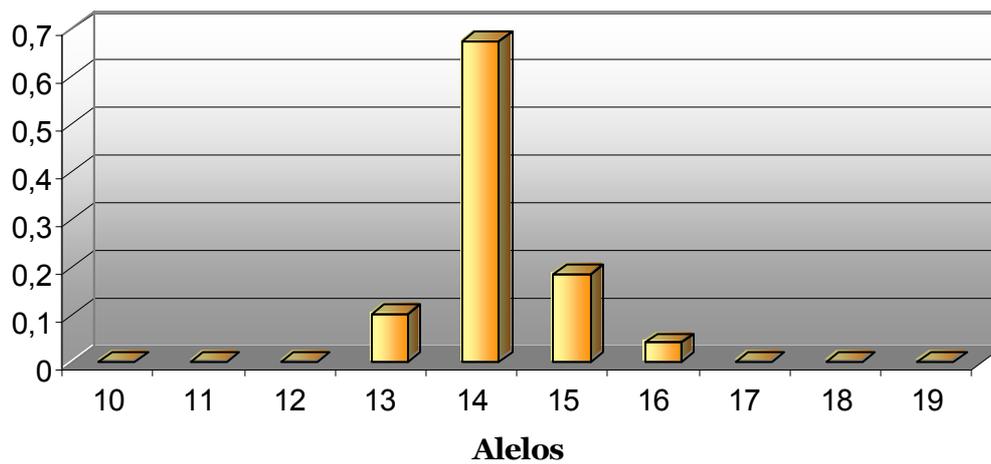
DYS 19

Figura 4.124: Frecuencias alélicas de la población de referencia italiana

Tabla 4.125. Frecuencias alélicas del sistema DYS 385 para la población española.

DYS 385	Recuento	Frecuencia alélica	s.d	Diversidad génica
10	1	$8.6207 \cdot 10^{-3}$	$8.583 \cdot 10^{-3}$	0.7314
11	39	0.3362	0.043	
12	5	0.0431	0.018	
13	10	0.0862	0.026	
14	43	0.3707	0.044	
15	8	0.0689	0.023	
16	7	0.0603	0.022	
17	1	$8.6207 \cdot 10^{-3}$	$8.583 \cdot 10^{-3}$	
18	2	0.0172	0.012	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.126. Frecuencias alélicas del sistema DYS 385 para la población italiana.

DYS 385	Recuento	Frecuencia alélica	s.d	Diversidad génica
10	3	0.0306	0.017	0.7960
11	26	0.2653	0.044	
12	7	0.0714	0.026	
13	6	0.0612	0.024	
14	32	0.3265	0.047	
15	8	0.0816	0.027	
16	9	0.0918	0.029	
17	4	0.0408	0.019	
19	2	0.0204	0.014	
20	1	0.0102	0.010	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

DYS 385

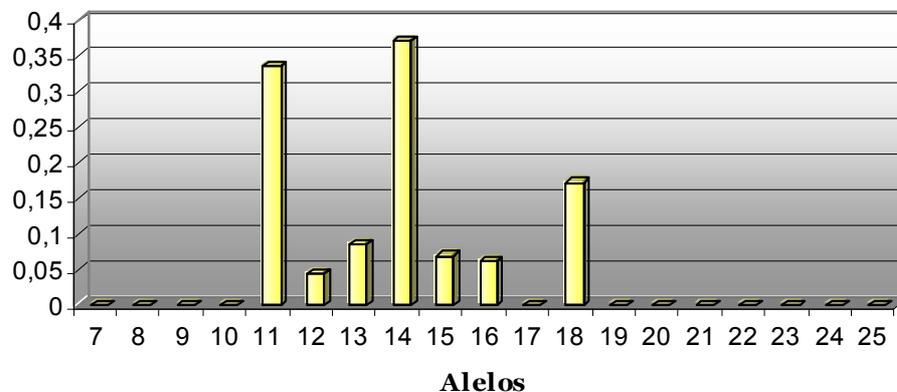


Figura 4.125: Frecuencias alélicas de la población de referencia española

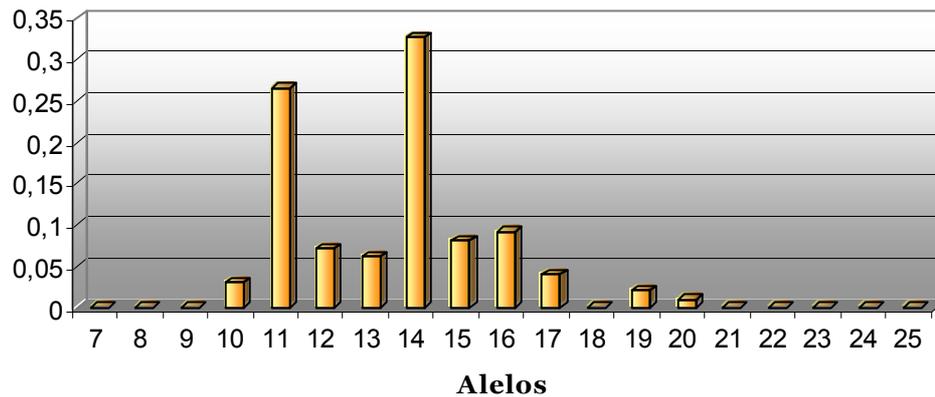
DYS 385

Figura 4.126: Frecuencias alélicas de la población de referencia italiana

Tabla 4.127. Frecuencias alélicas del sistema DYS 393 para la población española.

DYS 393	Recuento	Frecuencia alélica	s.d	Diversidad génica
10	1	0.0172	0.017	0.4454
12	6	0.1034	0.039	
13	42	0.7241	0.058	
14	8	0.1379	0.045	
15	1	0.0172	0.017	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.128. Frecuencias alélicas del sistema DYS 393 para la población italiana.

DYS 393	Recuento	Frecuencia alélica	s.d	Diversidad génica
12	8	0.1632	0.052	0.4266
13	36	0.7347	0.063	
14	4	0.0816	0.039	
15.1	1	0.0204	0.020	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

DYS 393

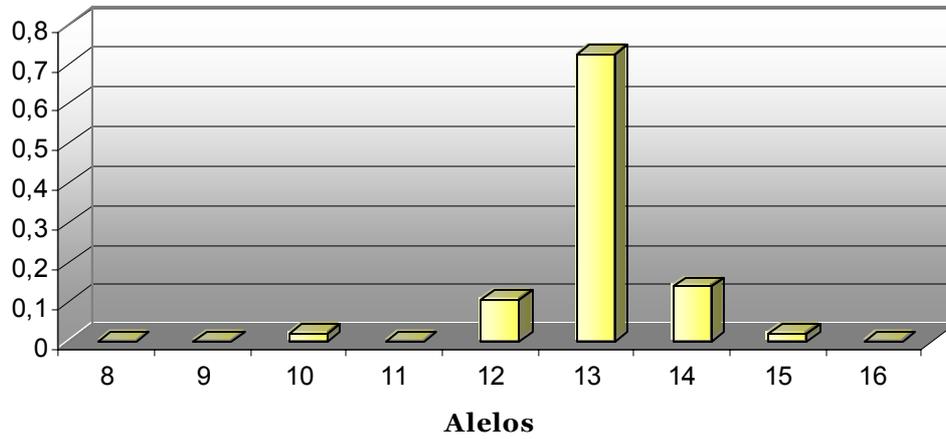


Figura 4.127: Frecuencias alélicas de la población de referencia española

DYS 393

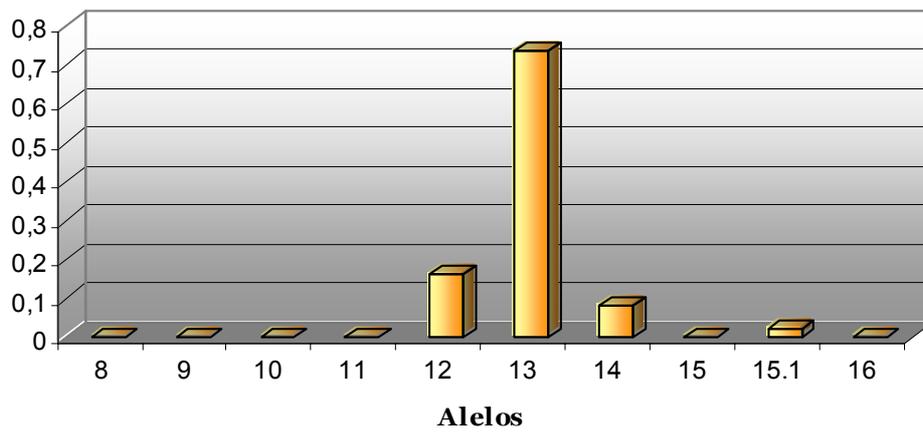


Figura 4.128: Frecuencias alélicas de la población de referencia italiana

Tabla 4.129. Frecuencias alélicas del sistema DYS 391 para la población española.

DYS 391	Recuento	Frecuencia alélica	s.d	Diversidad génica
9	2	0.0345	0.023	0.5452
10	19	0.3276	0.061	
11	34	0.5862	0.064	
12	3	0.0517	0.029	

Diversidad génica computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.130. Frecuencias alélicas del sistema DYS 391 para la población italiana.

DYS 391	Recuento	Frecuencia alélica	s.d	Diversidad génica
9	2	0.0408	0.028	0.5365
10	22	0.4489	0.071	
11	25	0.5102	0.071	

Diversidad génica computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

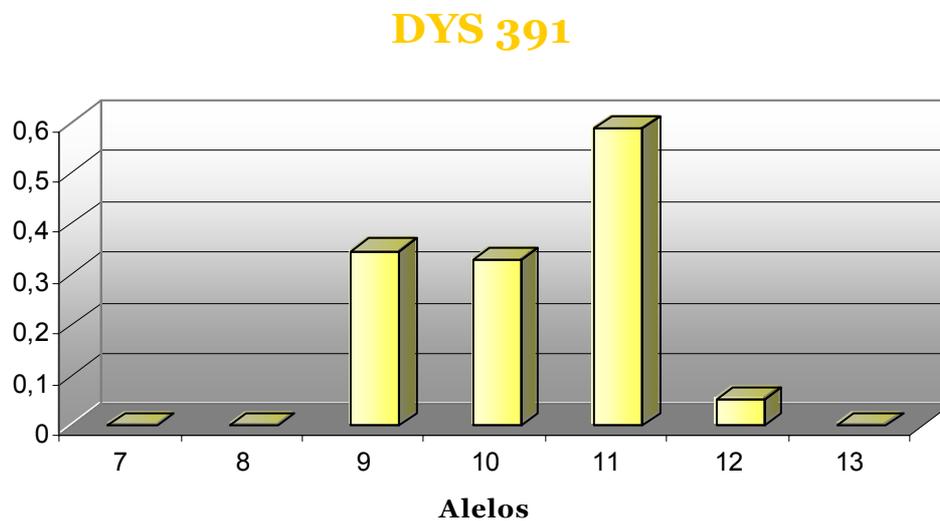


Figura 4.129: Frecuencias alélicas de la población de referencia española

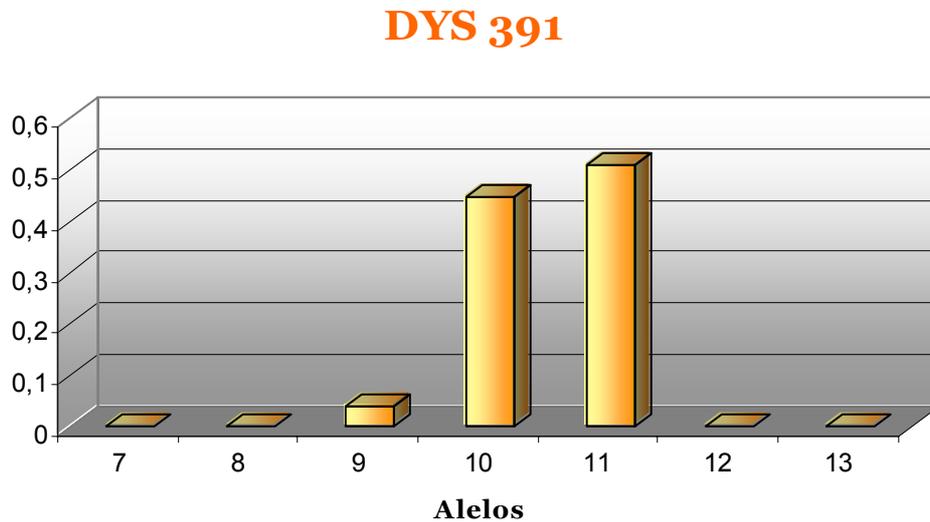


Figura 4.130: Frecuencias alélicas de la población de referencia italiana

Tabla 4.131. Frecuencias alélicas del sistema DYS 439 para la población española.

DYS 439	Recuento	Frecuencia alélica	s.d	Diversidad génica
10	1	0.0172	0.017	0.6279
11	17	0.2931	0.059	
12	29	0.5	0.065	
13	11	0.1896	0.051	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.132. Frecuencias alélicas del sistema DYS 439 para la población italiana.

DYS 439	Recuento	Frecuencia alélica	s.d	Diversidad génica
10	1	0.0204	0.020	0.6207
11	13	0.2653	0.063	
12	26	0.5306	0.071	
13	8	0.1632	0.052	
14	1	0.0204	0.020	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

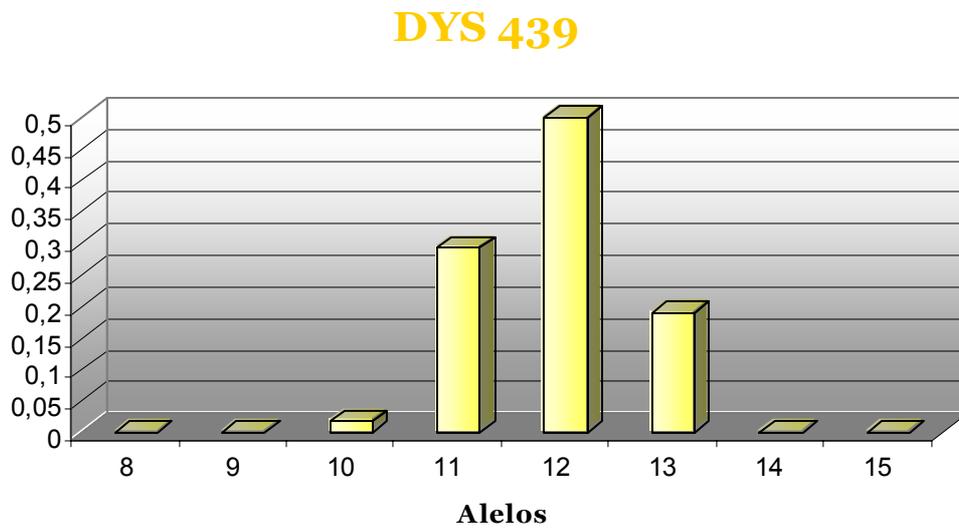


Figura 4.131: Frecuencias alélicas de la población de referencia española

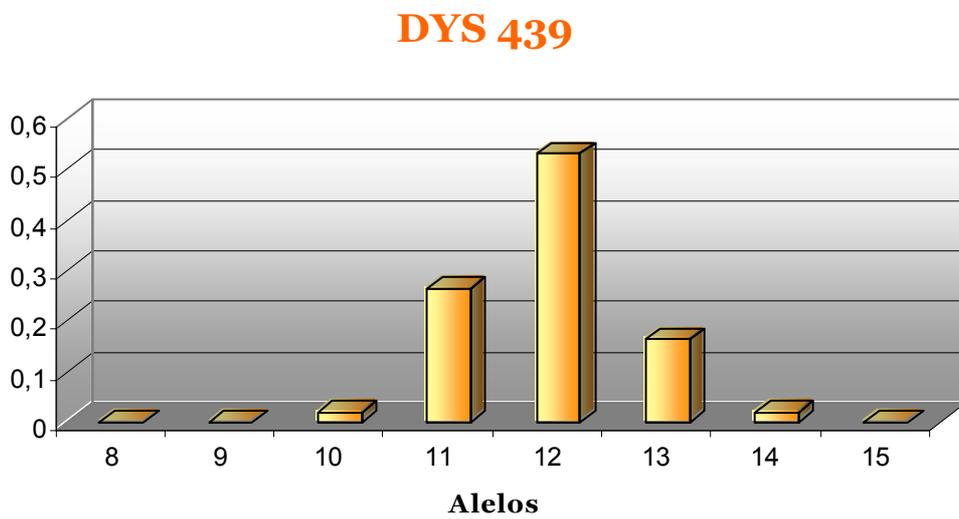


Figura 4.132: Frecuencias alélicas de la población de referencia italiana

Tabla 4.133. Frecuencias alélicas del sistema DYS 635 para la población española.

DYS 635	Recuento	Frecuencia alélica	s.d	Diversidad génica
20	2	0.0345	0.023	0.5749
21	10	0.1724	0.049	
22	3	0.0517	0.029	
23	36	0.6207	0.063	
24	4	0.0689	0.033	
25	2	0.0345	0.023	
26	1	0.0172	0.017	

Diversidad génica computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.134. Frecuencias alélicas del sistema DYS 635 para la población italiana.

DYS 635	Recuento	Frecuencia alélica	s.d	Diversidad génica
21	9	0.1836	0.055	0.5490
22	5	0.1020	0.043	
23	31	0.6326	0.068	
24	4	0.0816	0.039	

Diversidad génica computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

DYS 635

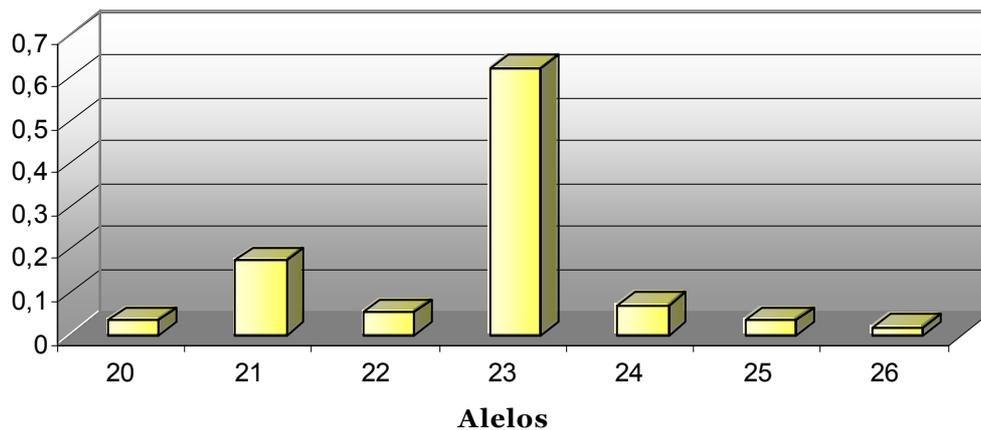


Figura 4.133: Frecuencias alélicas de la población de referencia española

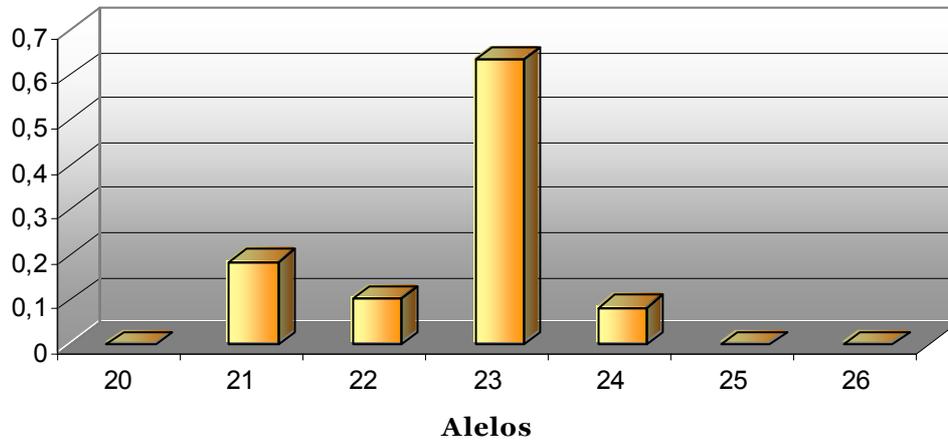
DYS 635

Figura 4.134: Frecuencias alélicas de la población de referencia italiana

Tabla 4.135. Frecuencias alélicas del sistema DYS 392 para la población española.

DYS 392	Recuento	Frecuencia alélica	s.d	Diversidad génica
11	13	0.2241	0.054	0.4240
12	2	0.0345	0.023	
13	42	0.7241	0.058	
15	1	0.0172	0.017	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.136. Frecuencias alélicas del sistema DYS 392 para la población italiana.

DYS 392	Recuento	Frecuencia alélica	s.d	Diversidad génica
11	15	0.3061	0.065	0.6139
12	3	0.0612	0.034	
13	26	0.5306	0.071	
14	4	0.0816	0.039	
15	1	0.0204	0.020	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

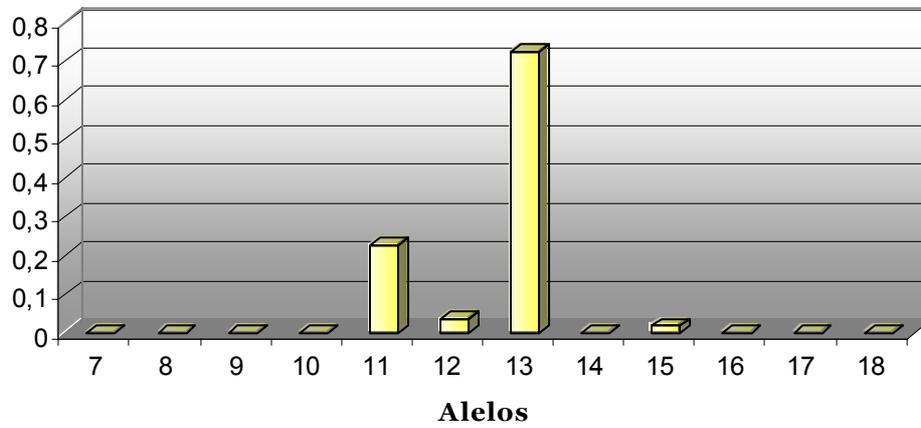
DYS 392

Figura 4.135: Frecuencias alélicas de la población de referencia española

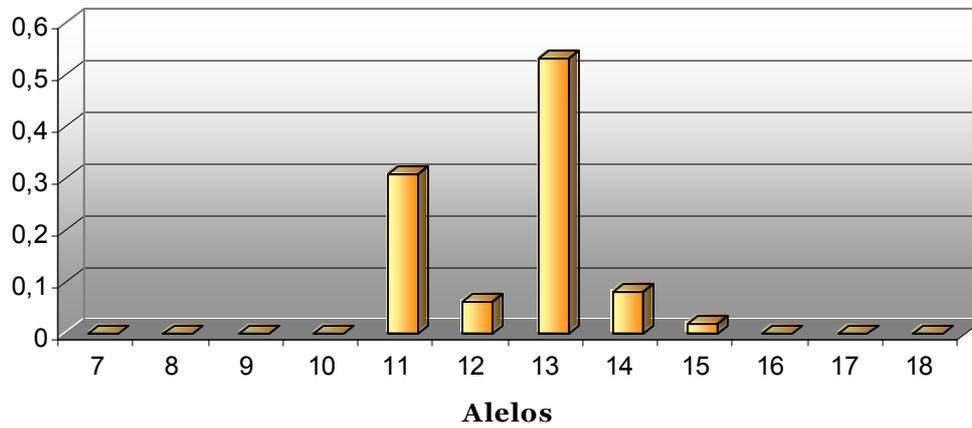
DYS 392

Figura 4.136: Frecuencias alélicas de la población de referencia italiana

Tabla 4.137. Frecuencias alélicas del sistema Y GATA H4 para la población española.

Y GATA H4	Recuento	Frecuencia alélica	s.d	Diversidad génica
10	3	0.0517	0.029	0.5720
11	23	0.3965	0.064	
12	30	0.5172	0.065	
13	1	0.0172	0.017	
>13	1	0.0172	0.017	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.138. Frecuencias alélicas del sistema Y GATA H4 para la población italiana.

Y GATA H4	Recuento	Frecuencia alélica	s.d	Diversidad génica
10	1	0.0204	0.020	0.5915
11	21	0.4285	0.070	
12	23	0.4694	0.071	
13	3	0.0612	0.034	
>13	1	0.0204	0.020	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Y GATA H4

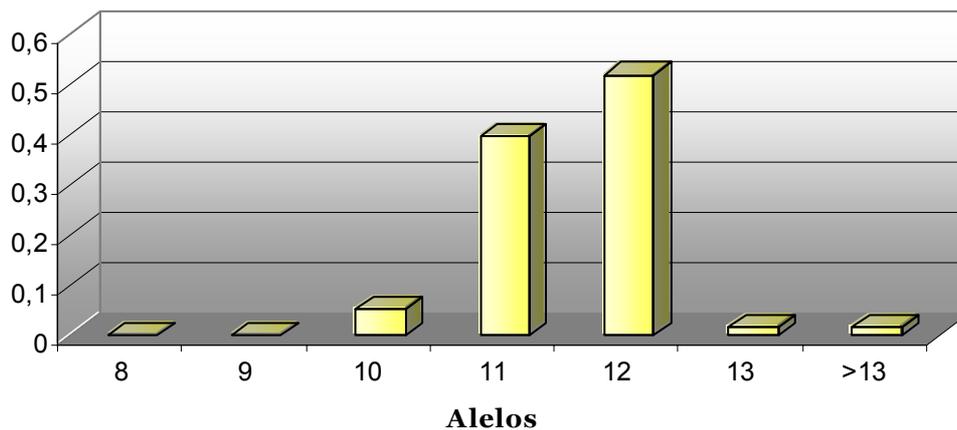


Figura 4.137: Frecuencias alélicas de la población de referencia española

Y GATA H4

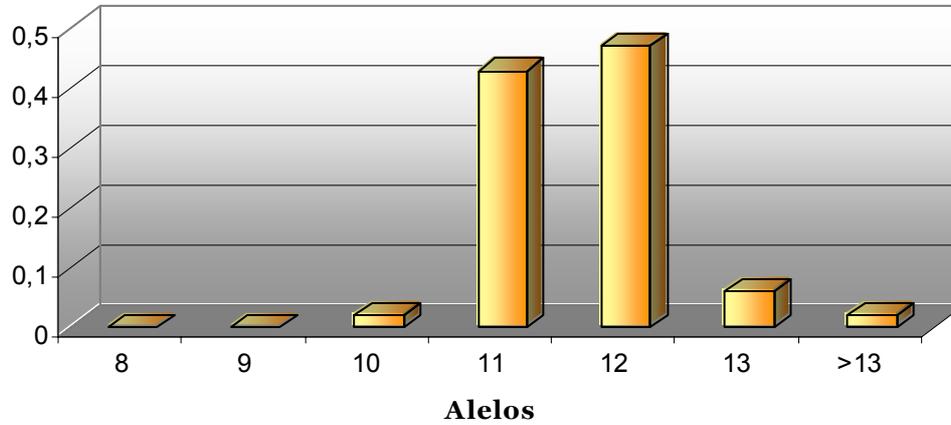


Figura 4.138: Frecuencias alélicas de la población de referencia italiana

Tabla 4.139. Frecuencias alélicas del sistema DYS 437 para la población española.

DYS 437	Recuento	Frecuencia alélica	s.d	Diversidad génica
14	14	0.2413	0.056	0.5018
15	38	0.6552	0.062	
16	6	0.1034	0.039	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.140. Frecuencias alélicas del sistema DYS 437 para la población italiana.

DYS 437	Recuento	Frecuencia alélica	s.d	Diversidad génica
14	20	0.4081	0.070	0.5864
15	24	0.4898	0.071	
16	4	0.0816	0.039	
17	1	0.0204	0.020	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

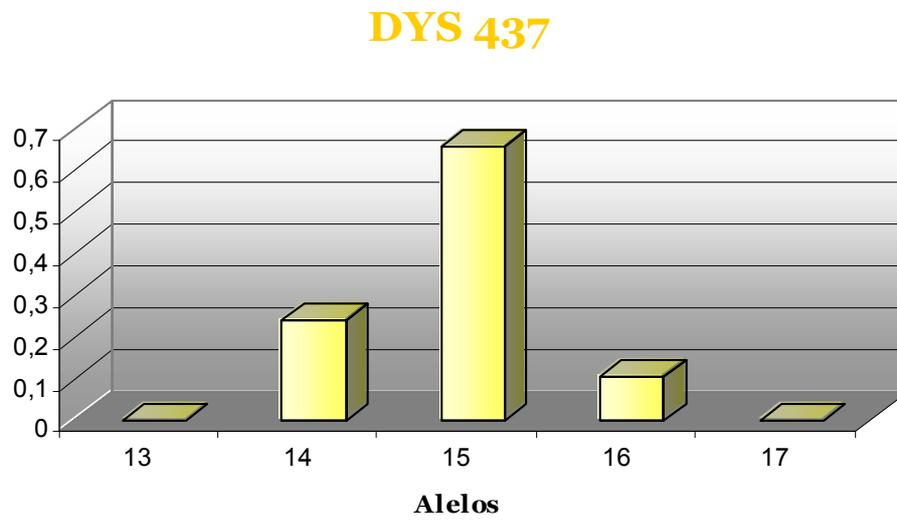


Figura 4.139: Frecuencias alélicas de la población de referencia española

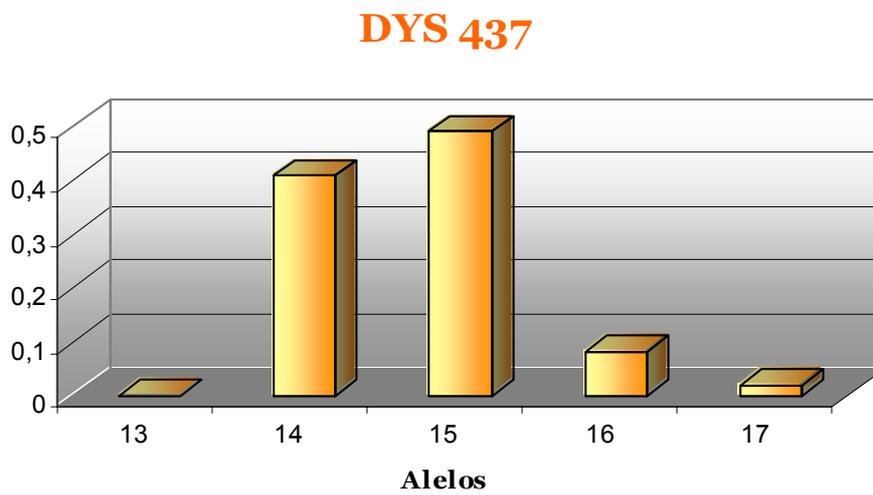


Figura 4.140: Frecuencias alélicas de la población de referencia italiana

Tabla 4.141. Frecuencias alélicas del sistema DYS 438 para la población española.

DYS 438	Recuento	Frecuencia alélica	s.d	Diversidad génica
9	4	0.0689	0.033	0.4917
10	9	0.1551	0.047	
11	3	0.0517	0.029	
12	40	0.6896	0.060	
13	2	0.0345	0.023	

Diversidad génica computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.142. Frecuencias alélicas del sistema DYS 438 para la población italiana.

DYS 438	Recuento	Frecuencia alélica	s.d	Diversidad génica
9	7	0.1428	0.049	0.5691
10	12	0.2449	0.061	
12	29	0.5918	0.070	
13	1	0.0204	0.020	

Diversidad génica computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

DYS 438

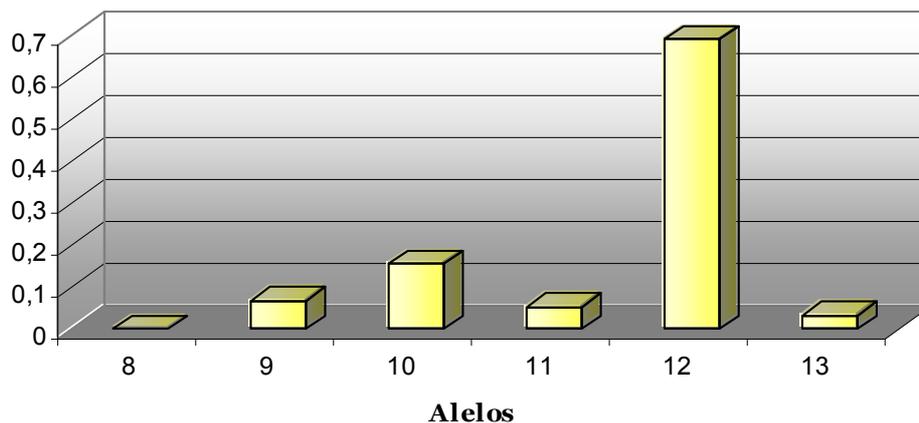


Figura 4.141: Frecuencias alélicas de la población de referencia española

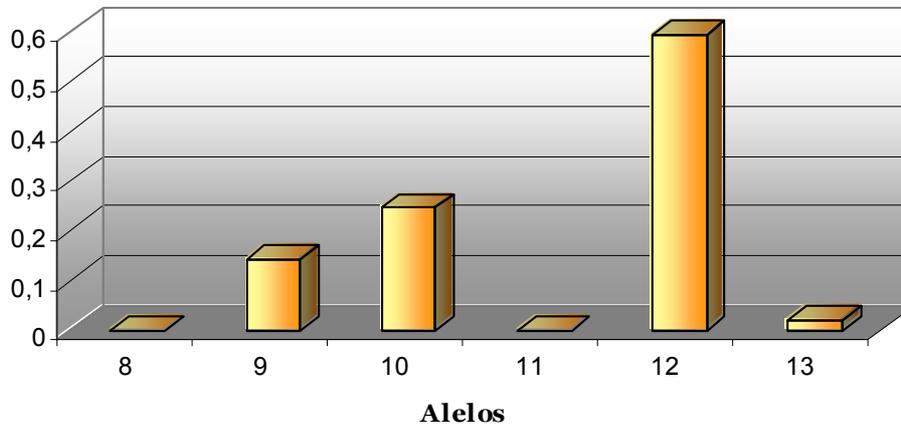
DYS 438

Figura 4.142: Frecuencias alélicas de la población de referencia italiana

Tabla 4.143. Frecuencias alélicas del sistema DYS 448 para la población española.

DYS 448	Recuento	Frecuencia alélica	s.d	Diversidad génica
17	1	0.0172	0.017	0.6118
18	10	0.1724	0.049	
19	33	0.5689	0.065	
20	10	0.1724	0.049	
21	4	0.0689	0.033	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

Tabla 4.144. Frecuencias alélicas del sistema DYS 448 para la población italiana.

DYS 448	Recuento	Frecuencia alélica	s.d	Diversidad génica
18	8	0.1632	0.052	0.6765
19	24	0.4898	0.071	
20	10	0.2041	0.057	
21	6	0.1224	0.046	
22	1	0.0204	0.020	

Diversidad genética computada como $1 - \sum p_i^2$ (p_i Frecuencias alélicas).

DYS 448

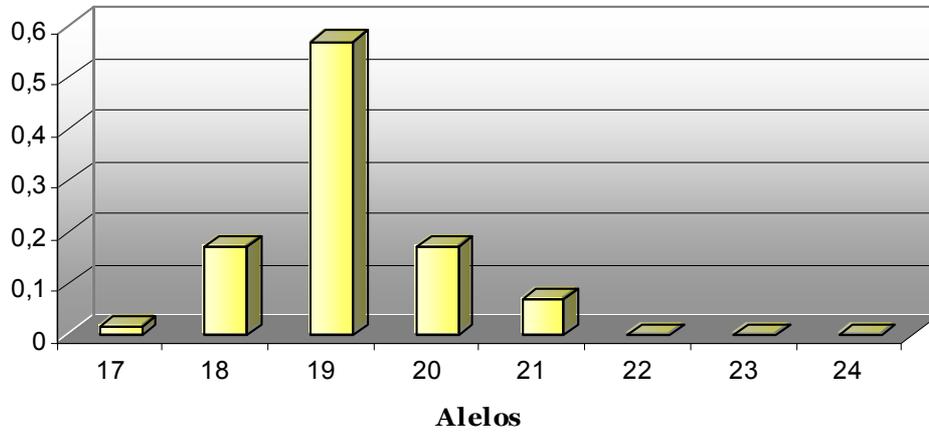


Figura 4.143: Frecuencias alélicas de la población de referencia española

DYS 448

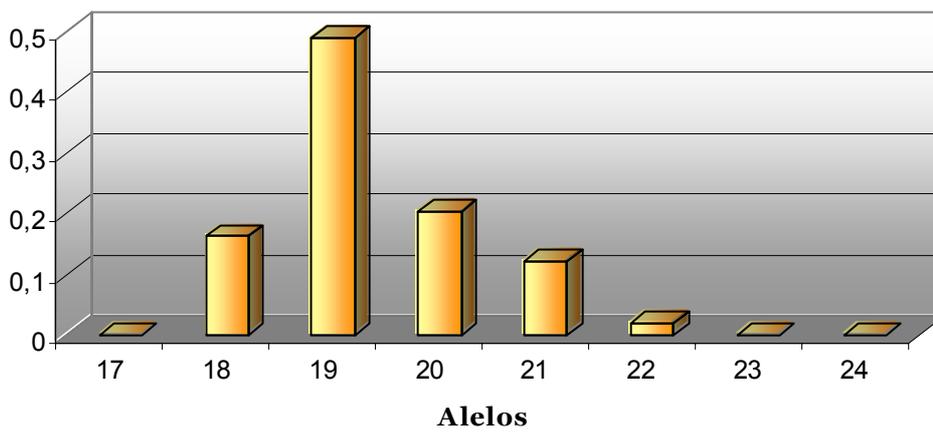


Figura 4.144: Frecuencias alélicas de la población de referencia italiana

4.1.3.- ANÁLISIS HAPLOTÍPICO DE INDIVIDUOS APELLIDADOS COLON

Estudio haplotípico de las muestras de los varones apellidados Colon, obtenidas de individuos que habitan en la zona noroeste de la cuenca mediterránea. Y grupos control (no sesgados) de individuos no relacionados entre si de estas mismas zonas.

Resultados obtenidos mediante el análisis de 16 loci tipo STR, pertenecientes al Kit de identificación genética Y-filer®.

Se han estudiado 477 muestras, para los loci: *DYS456*, *DYS389I*, *DYS390*, *DYS389II*, *DYS458*, *DYS19*, *DYS385*, *DYS393*, *DYS391*, *DYS439*, *DYS635*, *DYS392*, Y *GATA H4*, *DYS437*, *DYS438* y *DYS44*.

ZONAS ESPAÑOLAS ESTUDIADAS:



CATALUÑA:

Se han estudiado 125 muestras de esta zona, obtenidas de individuos apellidados Colon, intentando en la medida de lo posible que no exista relación directa con otros donantes, de estas se han obtenido 125 perfiles genéticos. Como era de suponer, se han encontrado perfiles genéticos idénticos, en todos los marcadores. De los 125 individuos estudiados hemos encontrado 54 perfiles genéticos diferentes.

El haplotipo predominante en la zona, le hemos denominado Haplotipo C-I, que comparten 14 individuos del estudio.

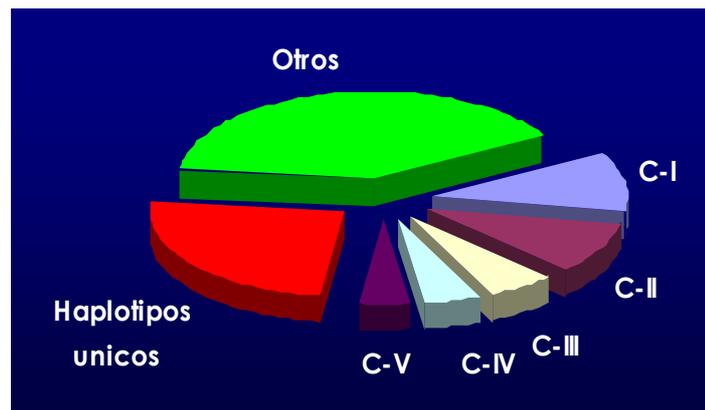
Sample Info	<i>DYS456</i>	<i>DYS389I</i>	<i>DYS390</i>	<i>DYS389II</i>	<i>DYS458</i>	<i>DYS19</i>	<i>DYS385</i>	<i>DYS393</i>	<i>DYS391</i>	<i>DYS439</i>	<i>DYS635</i>	<i>DYS392</i>	<i>YGATA H4</i>	<i>DYS437</i>	<i>DYS438</i>	<i>DYS448</i>
Haplotipo C-I	15	14	24	30	16	14	11 14	13	10	12	23	13	12	15	12	19

Además de este haplogrupo mayoritario, 11%, existe un segundo grupo C-II con 12 individuos. Otros tres grupos con menor representación serían C-III, C-IV y C-V, con 7, 6 y 5 individuos cada uno respectivamente. Además de estos grupos se han encontrado: 5 haplotipos compartidos por cuatro individuos, 4 por tres y 9 haplotipos que cada uno lo comparten dos individuos. Además de estos hay 31 haplotipos individuales, que solo poseen 1 individuo de la muestra.

		DYS 456	DYS 389 I	DYS 390	CYS 389 II	DYS 458	DYS 19	DYS 385	DYS 393	DYS 391	DYS 439	DYS 635	DYS 392	Y GATA H4	DYS 437	DYS 438	DYS 448	
C-II	9,60%	16	13	24	28	16	14	10	14	13	10	13	24	13	11	14	12	19
C-III	5,60%	15	13	24	30	16	14	11	14	12	10	13	23	13	12	15	12	19
C-IV	4,80%	16	13	23	29	15	15	13	16	12	10	11	21	11	11	15	9	21
C-V	4%	15	13	23	29	15	15	13	15	12	10	11	21	11	11	15	9	21

Los 5 haplotipos mayoritarios suponen más del treinta y cinco por ciento del grupo estudiado.

Distribución
haplotípica:



HAPLOTIPOS		N	F	F ²
C-I	1	14	0.112	0.0125
C-II	1	12	0.096	9.216x10 ⁻³
C-III	1	7	0.056	3.136x10 ⁻³
C-IV	1	6	0.048	2.304x10 ⁻³
C-V	1	5	0.040	1.600x10 ⁻³
C-VI, C-VII, C-VIII, C-IX y C-X	5x	4	0.032	(1.024x10 ⁻³)x5
C-XI, C-XII, C-XIII y C-XIV	4x	3	0.024	5.760x10 ⁻⁴)x4
C-XV, C-XVI, C-XVII, C-XVIII, C-XIX, C-XX, C-XXI, C-XXII y C- XXIII	9x	2	0.016	(2.560x10 ⁻⁴)x9
Haplotipos únicos	31x	1	8x10 ⁻³	(6,400x10 ⁻⁴)x31
Total de individuos		125		ΣF ² =0.0404
Probabilidad de coincidencia		0.0404		
Diversidad haplotípica		0.9673		(n/n-1)(1-Σpi ²).
Capacidad de discriminación		43.20%		Haplotipos/total muestrax100

BALEARES:

Las muestras se recogieron a 50 voluntarios de las islas, apellidados colon, de los que se realizó el análisis. De los cincuenta análisis realizados la mitad de ellos comparten el mismo perfil genético para estos loci. A este grupo lo hemos llamado B-I. Pero si sumamos, los individuos que solo se diferencian en un locus del haplotipo principal, B-I mut, obtenemos un 70% de la muestra.

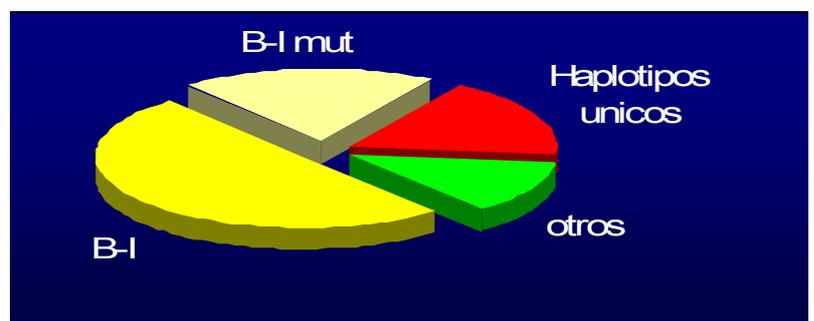
Sample Info	DYS456	DYS389I	DYS390	DYS389II	DYS458	DYS19	DYS385	DYS393	DYS391	DYS439	DYS635	DYS392	YGATA H4	DYS437	DYS438	DYS448
Haplotipo B-I	15	13	24	29	16	14	11 14	13	11	12	23	13	11	14	12	18

		DYS 456	DYS 389 I	DYS 390	CYS 389 II	DYS 458	DYS 19	DYS 385	DYS 393	DYS 391	DYS 439	DYS 635	DYS 392	Y GATA H4	DYS 437	DYS 438	DYS 448
B-I mut	8,00%	16	13	24	29	16	14	11 14	13	11	12	23	13	11	14	12	18
B-I mut'	4,00%	15	13	24	29	17	14	11 14	13	11	12	23	13	11	14	12	18
B-I mut''	2,00%	15	13	23	29	16	14	11 14	13	11	12	23	13	11	14	12	18
B-I mut'''	2,00%	15	13	24	29	16	14	11 15	13	11	12	23	13	11	14	12	18
B-I mut''''	2,00%	15	13	24	29	16	14	11 14	13	11	12	23	13	11	14	13	18
B-I mut'''''	2,00%	15	13	24	29	16	14	12 14	13	11	12	23	13	11	14	12	18

*Mutación por inserción o deleción de un STR del perfil principal.

En total se han encontrado 19 perfiles genéticos distintos en los 50 individuos estudiados. Además del grupo B-1 e individuos relacionados, encontramos 9 individuos que no comparten su perfil con ningún otro y 3 haplotipos compartidos por 2 individuos

Distribución haplotípica:



HAPLOTIPOS		N	F	F ²
B-I	1	25	0.5	0.25
B-I mut.	1	4	0.08	6.4 x10 ⁻³
B-I mut'.	1	2	0.04	1.6 x10 ⁻³
B-I mut'', B-I mut''', B-I mut'''' y B-I mut''''''	4x	1	0.02	(4.0x10 ⁻⁴)x4
B-II, B-III, B-IV	3x	2	0.04	1.6 x10 ⁻³)x2
Haplotipos únicos	9x	1	0.02	(4.0x10 ⁻⁴)x9
Total de individuos	50			ΣF ² =0.0404
Probabilidad de coincidencia	0.268			
Diversidad haplotípica	0.7469		(n/n-1)(1-Σpi ²).	
Capacidad de discriminación	38%		Haplotipos/total muestrax100	

VALENCIA:

Cuarenta y cinco muestras se recogieron en la zona de Valencia. Hemos encontrado 20 haplotipos distintos. De los perfiles que encontramos repetidos tenemos que distinguir 4 grupos mayoritarios: V-I, que si sumamos los individuos idénticos con los que poseen una mutación alcanza el 31% de las muestras, el V-II, 11%, el V-III, 8% y si añadimos los que acumulan una mutación un 16% y ocurre lo mismo con el cuarto grupo V-IV donde un 8% comparte igual perfil y un 15% si añadimos los mutados. Al igual que en el grupo anterior solo nueve individuos poseen un haplotipo único.

Sample Info	DYS456	DYS389I	DYS390	DYS389II	DYS458	DYS19	DYS385	DYS393	DYS391	DYS439	DYS635	DYS392	Y GATA H4	DYS437	DYS438	DYS448
Haplotipo V-I	15	13	23	28	16.2	14	14 18	12	10	11	20	11	11	14	10	20
Haplotipo V-II	13	12	24	28	17	15	14 17	12	10	12	<20	11	11	16	9	19
Haplotipo V-III	16	13	23	29	16	15	11 13	13	11	11	23	13	12	15	12	19
Haplotipo V-IV	16	13	23	30	18	14	11 14	13	10	11	23	13	11	14	12	18

		DYS 456	DYS 389 I	DYS 390	CYS 389 II	DYS 458	DYS 19	DYS 385	DYS 393	DYS 391	DYS 439	DYS 635	DYS 392	Y GATA H4	DYS 437	DYS 438	DYS 448
V-I	15,56%	15	13	23	28	16.2	14	14 18	12	10	11	20	11	11	14	10	20
V-I mut	8,89%	15	13	23	28	16.2	14	14 18	12	10	10	20	11	11	14	10	20
V-I mut'	4,44%	16	13	23	28	16.2	14	14 18	12	10	11	20	11	11	14	10	20
V-I mut''	2,22%	15	13	23	28	16	14	14 18	12	10	11	20	11	11	14	10	20

31% DE LA POBLACIÓN ESTUDIADA

		DYS 456	DYS 389 I	DYS 390	CYS 389 II	DYS 458	DYS 19	DYS 385	DYS 393	DYS 391	DYS 439	DYS 635	DYS 392	Y GATA H4	DYS 437	DYS 438	DYS 448	
V-III	6,67%	16	13	23	29	16	14	11	13	13	11	11	23	13	12	15	12	19
V-III mut	8,89%	16	13	23	29	16	15	11	13	13	11	11	23	13	12	15	12	19
V-III mut'	2,22%	16	13	23	29	16	14	11	13	13	11	11	23	13	12	15	12	18

17,78% DE LA POBLACIÓN ESTUDIADA

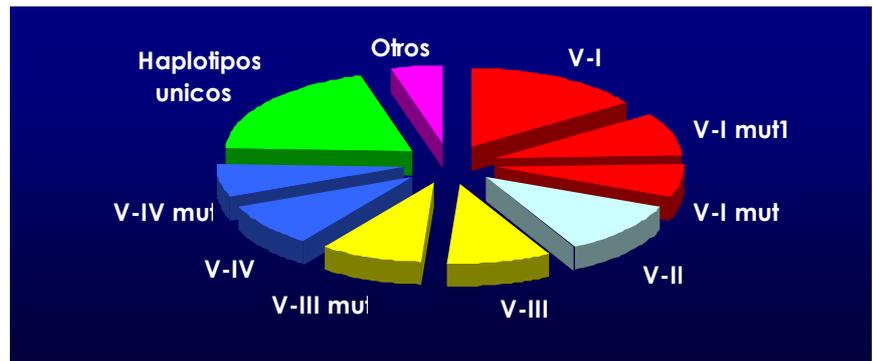
		DYS 456	DYS 389 I	DYS 390	CYS 389 II	DYS 458	DYS 19	DYS 385	DYS 393	DYS 391	DYS 439	DYS 635	DYS 392	Y GATA H4	DYS 437	DYS 438	DYS 448	
V-IV	8,89%	16	13	23	30	18	14	11	14	13	10	11	23	13	11	14	12	18
V-IV mut	6,67%	16	13	23	30	19	14	11	14	13	10	11	23	13	11	14	12	18

15,56% DE LA POBLACIÓN ESTUDIADA

4 Haplotipos predominantes= 75,56% del grupo.

Después de observar estos cuatro haplotipos predominantes, y los procedentes de una mutación puntual, nos damos cuenta que más de un setenta y cinco por ciento encajan con estos cuatro perfiles raíz.

Distribución haplotípica:



HAPLOTIPOS		N	F	F ²
V-I	1	7	0.150	0.0225
V-I mut	1	4	0.088	7.901x10 ⁻³
V-I mut'	1	2	0.044	1.936x10 ⁻³
V-I mut''	1	1	0.022	4.938x10 ⁻⁴
V-II	1	5	0.111	0.01232
V-III	1	3	0.067	4.444x10 ⁻³
V-III mut	1	4	0.089	7.901x10 ⁻³
V-III mut'	1	1	0.022	4.938x10 ⁻⁴
V-IV	1	4	0.089	7.901x10 ⁻³
V-IV mut	1	3	0.067	4.444x10 ⁻³
V-V	1	2	0.044	1.936x10 ⁻³
Haplotipos únicos	9x	1	0.022	(4.938x10 ⁻⁴)x9
Total de individuos		45		
Probabilidad de coincidencia		0.0765		ΣF ² =0.0765
Diversidad haplotípica		0.9445		(n/n-1)(1-Σpi ²).
Capacidad de discriminación		44.44%		Haplotipos/total muestrasx100

ZONA FRANCESA ESTUDIADA:

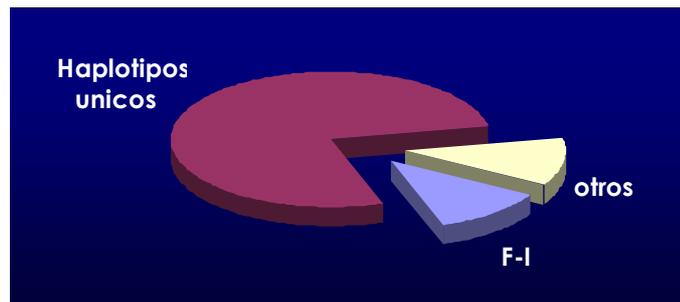


En la zona del sur de Francia estudiada, localizamos, dentro de los individuos apellidados Colon 30 haplotipos, tras estudiar un total de treinta y cinco muestras.

Existe un haplotipo mayoritario que se repite en cuatro de las muestras que corresponde a un 11,40% de la población, haplotipo F-1. También existen dos haplotipos que se repiten en 2 muestras, cada uno y el resto son haplotipos individuales.

Sample Info	DYS456	DYS389I	DYS390	DYS389II	DYS458	DYS19	DYS385	DYS393	DYS391	DYS439	DYS635	DYS392	YGATA H4	DYS437	DYS438	DYS448
Haplotipo F-I	15	14	24	30	16	14	11 14	13	10	12	23	13	12	15	12	19

Distribución haplotípica:



HAPLOTIPOS		N	F	F ²
F-I	1	4	0.1142	0.01306
F-II, F-III	2X	2	0.05714	(3.265x10 ⁻³)x2
Haplotipos únicos	27x	1	0.02857	(8.163x10 ⁻⁴)x27
Total de individuos		35		
Probabilidad de coincidencia		0.04163		ΣF ² =0.04163
Diversidad haplotípica		0.9864		(n/n-1)(1-Σpi ²).
Capacidad de discriminación		85.71%		Haplotipos/total muestrax100

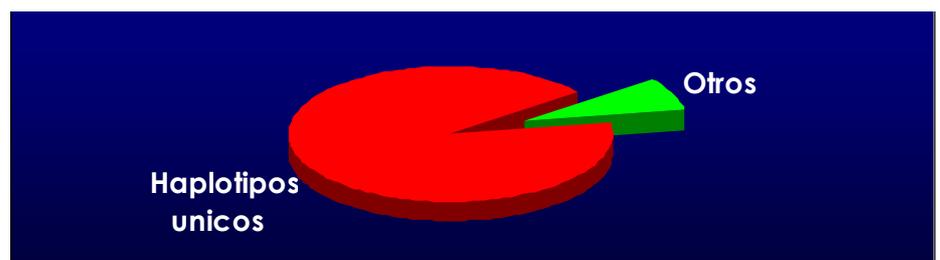
ZONAS ITALIANAS ESTUDIADAS:



LOMBARDIA:

En el estudio de esta zona han participado 52 individuos con apellido colon, tras el análisis de las muestras, se han encontrado 50 haplotipos. Por lo que el 92,31% de los individuos estudiados poseen un perfil que los individualiza. Como podemos ver es el grupo mas divergente, no ha influido que se hayan tomado individuos con el mismo apellido, ya que los resultados no convergen en haplotipos predominantes, solamente hemos encontrado 2 haplotipos compartidos por 2 individuos, cercano a lo que puede aparecer en un estudio poblacional, con el muestreo no condicionado.

Distribución haplotípica:

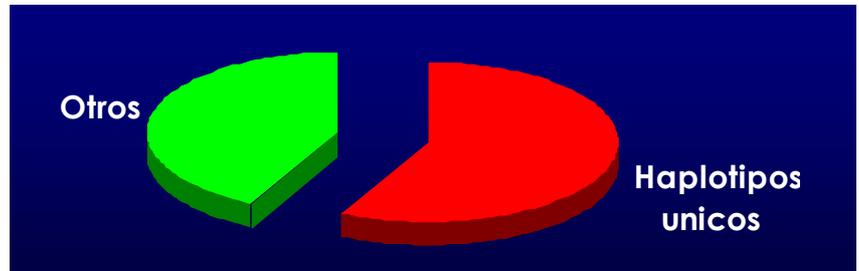


HAPLOTIPOS		N	F	F ²
LO-I, LO-II	2X	2	0.0.3846	(1.4792x10 ⁻³)x2
Haplotipos únicos	48x	1	0.01923	(3.698x10 ⁻⁴)x48
Total de individuos		52		ΣF ² =0.02071
Probabilidad de coincidencia		0.02071		
Diversidad haplotípica		0.9985		(n/n-1)(1-Σpi ²).
Capacidad de discriminación		96.15%		Haplotipos/total muestrax100

LIGURIA:

Cuarenta y siete muestras fueron analizadas de Liguria, en esta zona se encontraron treinta y seis haplotipos individuales, siete que comparten dos individuos y dos que comparten tres individuos, pero no se encontró tampoco al igual que en Lombardía ningún grupo con gran representación.

Distribución haplotípica:



Aparecen mas haplotipos repetidos que en Lombardia pero aun así la diversidad es bastante alta, sin llegar a ser tan alta como la de

HAPLOTIPOS		N	F	F ²
LI-I, L-II	2X	3	0.06382	$(4.074 \times 10^{-3}) \times 2$
LI-III, LI-IV, LI-V, LI-VI, LI-VII, LI-VIII, LI-IX	7X	2	0.04255	$(1.8107 \times 10^{-3}) \times 7$
Haplotipos únicos	27x	1	0.02127	$(4.527 \times 10^{-4}) \times 27$
Total de individuos	47			$\Sigma F^2 = 0.03304$
Probabilidad de coincidencia	0.03304			
Diversidad haplotípica	0.9879		$(n/n-1)(1-\Sigma \pi_i^2)$.	
Capacidad de discriminación	76.59%		Haplotipos/total muestrax100	

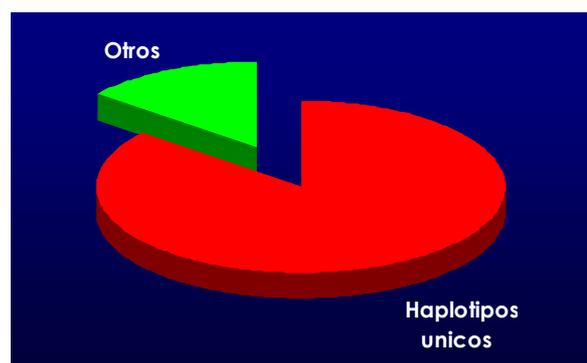
Lombardia, pero si dista mucho de lo que observábamos en Baleares, Valencia o Cataluña.

PIAMONTE:

Este es el grupo mas pequeño que se ha estudiado, y comparte en cierta manera características con las otras 2 zonas italianas estudiadas.

Se analizaron 14 muestras, y de ellas dos individuos comparten un mismo perfil.

Distribución haplotípica:



Los valores estudiados de divergencia y coincidencia se sitúan en una posición intermedia con respecto a los otros dos grupos italianos.

El resto de los Valores estudiados, así como la no aparición de algún haplotipo más numeroso, coincide con lo estudiado en los otros dos grupos de Italia.

HAPLOTIPOS		N	F	F ²
PI-I	1	2	0.1428	0.02040
Haplotipos únicos	12x	1	0.0714	(5.102x10 ⁻³)x12
Total de individuos	14			ΣF ² =0.08162
Probabilidad de coincidencia	0.08162			
Diversidad haplotípica	0.9890		(n/n-1)(1-Σpi ²).	
Capacidad de discriminación	92.85%		Haplotipos/total muestrax100	

GRUPOS CONTROL:

Muestreo de individuos no relacionados entre si de las mismas zonas estudiadas para los apellidos Colon.

El número total de muestras analizadas fueron 109 para estos grupos control.

GRUPO DE REFERENCIA ITALIA:

Para realizar este estudio se tomaron 50 muestras de individuos no relacionados, que viven en la Zona de Lombardia, Liguria y Piamonte.

Los resultados obtenidos en este estudio haplotípico son los esperados para un kit de identificación con un numero tan grande de marcadores, si la muestra se toma al azar.

HAPLOTIPOS		N	F	F ²
Haplotipos únicos	50x	1	0.02	(4.00x10 ⁻⁴)x50
Total de individuos	50			ΣF ² =0.02
Probabilidad de coincidencia	0.02			
Diversidad haplotípica	0.99999		(n/n-1)(1-Σpi ²).	
Capacidad de discriminación	100%		Haplotipos/total muestrax100	

Como podemos ver ninguno de los haplotipos se repite en ninguna ocasión, los grandes grupos familiares que observamos al condicionar el muestreo, desaparecen en el grupo control.

Esto valida los resultados obtenidos en la zona estudiada.

GRUPO DE REFERENCIA ESPAÑA:

Para realizar este estudio se tomaron 59 muestras de individuos no relacionados, que viven en la Zona de Cataluña, Valencia y Baleares.

Los resultados obtenidos en este estudio haplotípico, al igual que al ocurrido con el grupo de referencia anterior, son los esperados al utilizar 16 marcadores STR.

HAPLOTIPOS		N	F	F²
Haplotipos únicos	59x	1	0.01694	$(2.873 \times 10^{-4}) \times 59$
Total de individuos		59		$\Sigma F^2 = 0.01694$
Probabilidad de coincidencia		0.01694		
Diversidad haplotípica		1	$(n/n-1)(1-\Sigma p_i^2)$.	
Capacidad de discriminación		100%	Haplotipos/total muestrax100	

Los resultados obtenidos, tienen iguales características que en el grupo de referencia Italiano, siendo aun mas altos los valores de discriminación, pero se debe a un aumento en el numero de muestras estudiadas, no a que exista mayor diversidad, al igual que en Italia todos los haplotipos son únicos. Si que difieren en gran medida de los estudios realizados a los individuos apellidados Colon de esta zona, donde la diversidad hemos visto que disminuye mucho, así como el resto de los valores estudiados.

Al igual que hemos visto en el grupo anterior, estos resultados validan los resultados obtenidos en los grupos de individuos de las zonas españolas con apellido Colon. Para finalizar este apartado, simplemente indicar, que hemos encontrado algunos haplotipos que se repiten en distintas poblaciones. Por ejemplo, un haplotipo de Italia de referencia lo hemos encontrado en España de referencia, dos de Italia de referencia en Cataluña, uno de Francia en Baleares, cuatro de Francia en Cataluña y otro en Valencia, uno de Piamonte en Liguria, dos de Baleares en Cataluña y otros dos en Valencia y finalmente cinco de Cataluña en Valencia.

4.1.4.- ANÁLISIS DE LOS HAPLOGRUPOS DE INDIVIDUOS APELLIDADOS COLON

Cada una de las muestras estudiadas, se pueden ir encuadrando en un haplogrupo determinado, según el haplotipo que posean. De este modo, hemos ido estudiando que haplogrupo posee cada muestras, y cuál de ellos es más frecuente en cada una de las zonas o regiones estudiadas. Los haplogrupos pueden ser designados en base a los SNPs.

El método está implementado en un programa basado en Excel, donde se introduce un número de marcadores STRs de cromosoma Y y éste devuelve una puntuación sobre 10 acerca de la bondad de ajuste para los haplogrupos E3a, E3b, G2, I1a, I1b, I1c, J2, N3, R1a y R1b.

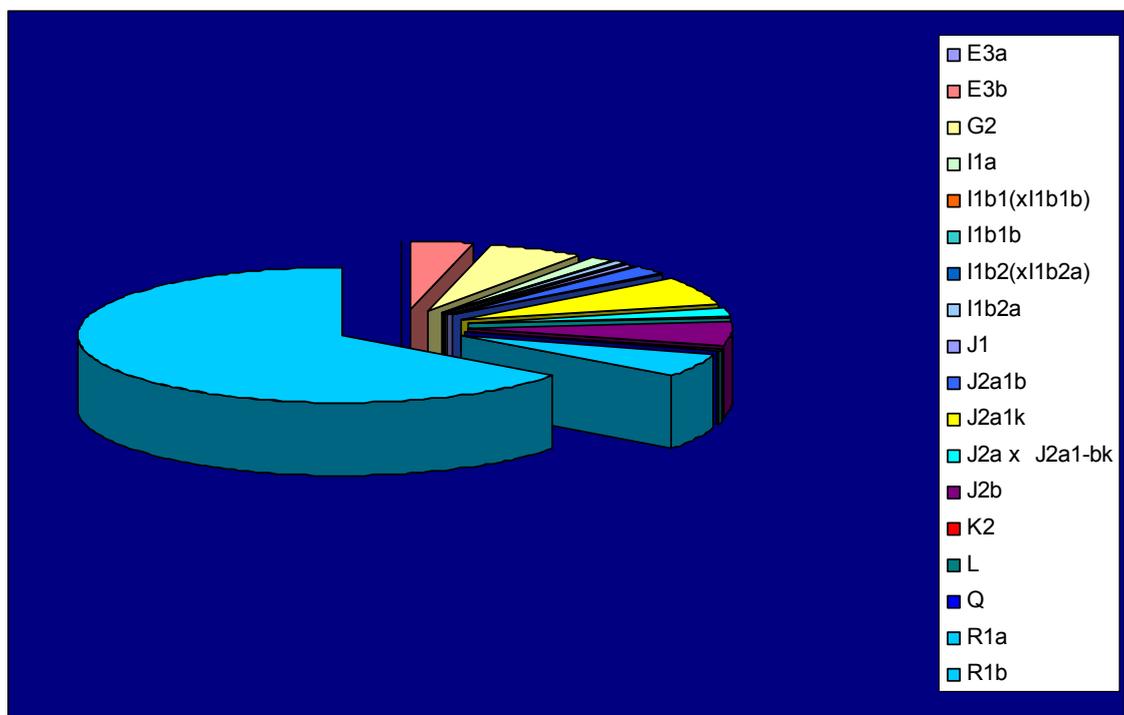
Una de las principales compañías de pruebas de ADN, Family Tree DNA (FTDNA) en cooperación con la Universidad de Arizona usa su propio algoritmo para predecir el haplogrupo de personas que tienen sus valores de Y-STR. El algoritmo de predicción no se ha publicado pero parece estar basado en la distancia genética del haplotipo en cuestión con otros haplotipos de la base de datos de la Universidad de Arizona.

La aproximación de la FTDA ha sido bastante exitosa y probablemente el 80% de los usuarios consiguen una predicción del haplogrupo. La desventaja de esta aproximación es que si no se puede hacer una predicción, entonces el usuario no obtiene ninguna información, incluso estando claro que algunos haplogrupos puedan descartarse, o que el haplotipo pertenezca a un número de posibles haplogrupos. Los alelos más representativos para pertenecer a uno u otro grupo son:

Haplogr	DYS456	DYS389I	DYS390	DYS458	DYS19	DYS385a	DYS385b	DYS393	DYS391	DYS392	YGATA H4	DYS437	DYS438	DYS448
L			29			7		9						
G2								17	7					23
R1B	20	9	19			11	10,11						12,13,14	14,15
E3b					9,10,11	16,17,18,19					27			11
Q							23,24			17,18		11		
J2b	11,12,13					8								
J1				22,23							7			
J2a1b													7	
K2														16
I1b1b							12							
I1b2 (xI1b2a)			28											
H					20									
N					25									
E3a						20								
R1a											14			

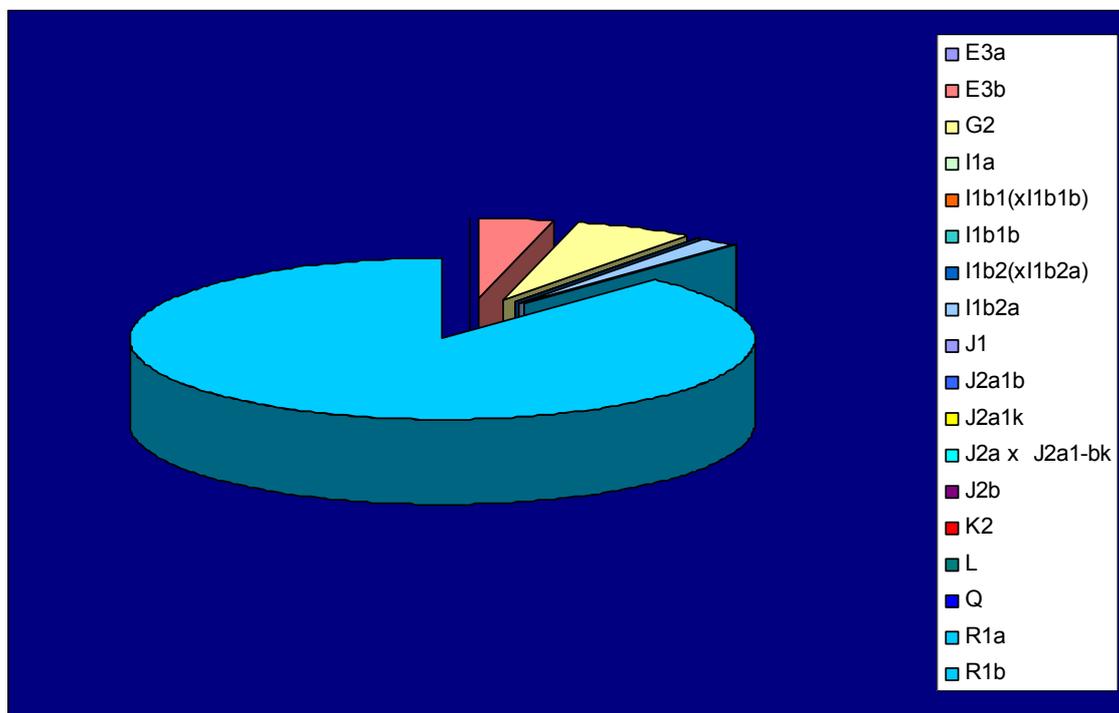
CATALUÑA:

HAPLOGRUPOS	N	F	F ²
E3b	5	0.04	1.6 10 ⁻³
G2	7	0.056	3.136 10 ⁻³
I1a	2	0.016	2.56 10 ⁻³
I1b2a	1	8 10 ⁻³	6.4 10 ⁻³
J1	1	8 10 ⁻³	6.4 10 ⁻³
J2a1k	9	0.072	5.184 10 ⁻³
J2axJ2a1-bk	5	0.04	1.6 10 ⁻³
J2b	7	0.056	3.136 10 ⁻³
R1b	81	0.648	0.4199
R1a	7	0.056	3.136 10 ⁻³
Total de individuos	125		ΣF ² =0.4530
Probabilidad de coincidencia	0.4530		
Diversidad haplotípica	0.5514	(n/n-1)(1-Σpi ²).	

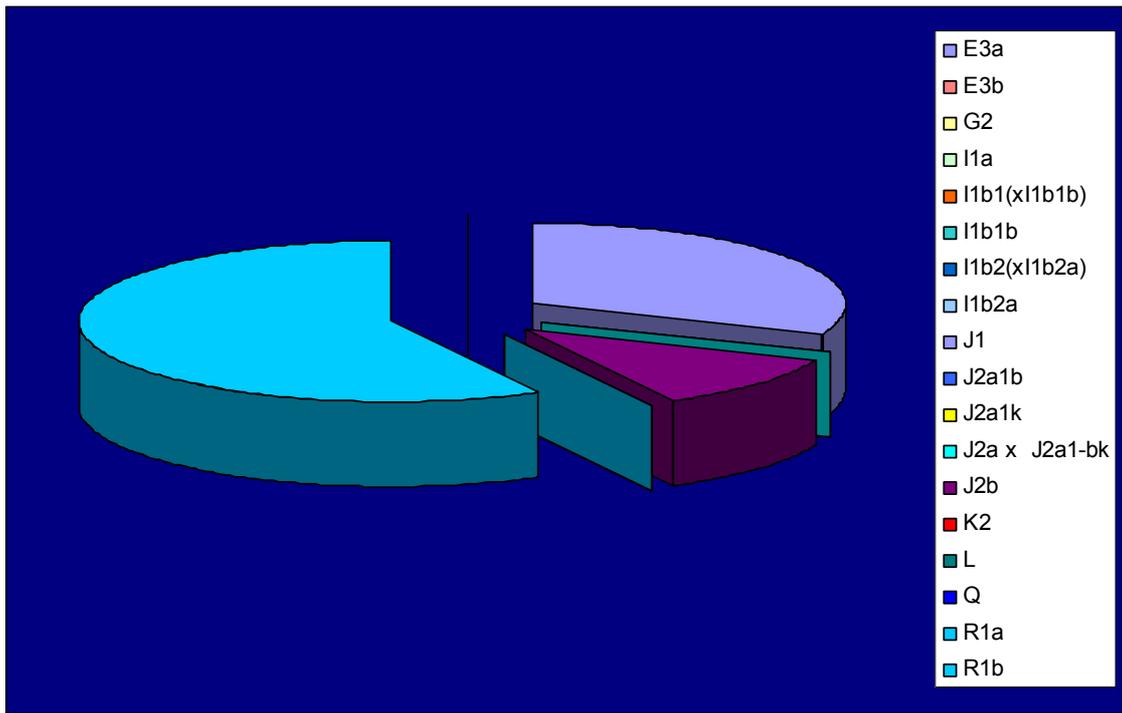


BALEARES:

HAPLOGRUPOS	N	F	F ²
E3b	2	0.04	1.6 10 ⁻³
G2	3	0.06	3.6 10 ⁻³
I1b2a	1	0.02	4 10 ⁻³
R1b	44	0.88	0.7744
Total de individuos	50		ΣF ² =0.7836
Probabilidad de coincidencia	0.7836		
Diversidad haplotípica	0.2208	(n/n-1)(1-Σpi ²).	

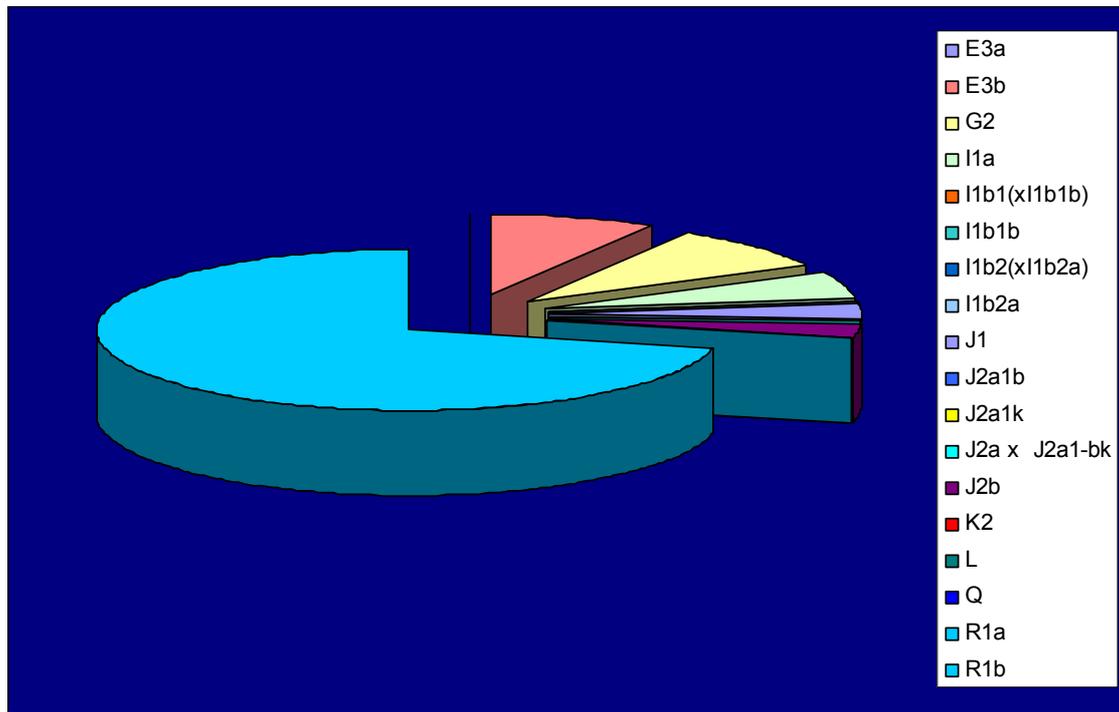
**VALENCIA:**

HAPLOGRUPOS	N	F	F ²
J1	14	0.3111	0.0968
J2b	5	0.1111	0.0123
R1b	26	0.5777	0.3338
Total de individuos	45		ΣF ² =0.4429
Probabilidad de coincidencia	0.4429		
Diversidad haplotípica	0.5697	(n/n-1)(1-Σpi ²).	



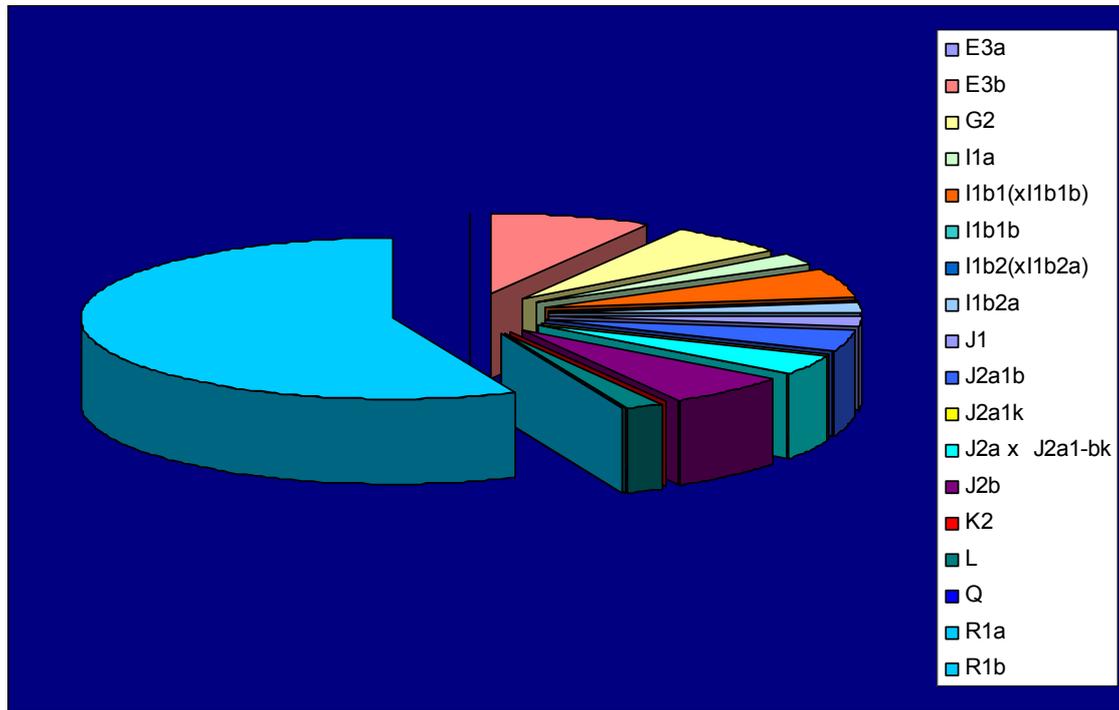
FRANCIA:

HAPLOGRUPOS	N	F	F ²
E3b	3	0.085	7.346 10 ⁻³
G2	3	0.085	7.346 10 ⁻³
I1a	2	0.057	3.265 10 ⁻³
J1	1	0.028	8.163 10 ⁻³
J2b	1	0.028	8.163 10 ⁻³
R1b	25	0.714	0.5102
Total de individuos	35		ΣF ² =0.5444
Probabilidad de coincidencia	0.5444		
Diversidad haplotípica	0.4689	(n/n-1)(1-Σpi ²).	



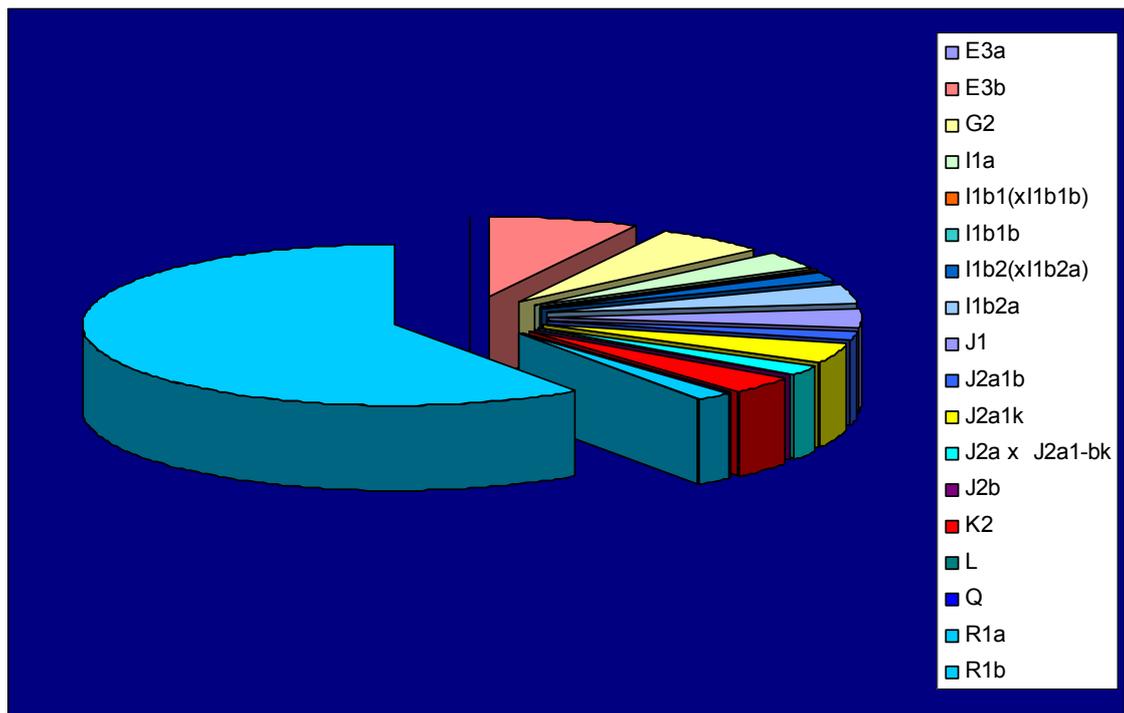
LIGURIA:

HAPLOGRUPOS	N	F	F ²
E3b	4	0.0833	6.944 10 ⁻³
G2	3	0.0625	3.906 10 ⁻³
I1a	1	0.0208	4.340 10 ⁻³
I1b2a	1	0.0208	4.340 10 ⁻³
I1b1(xI1b1b)	3	0.0625	3.906 10 ⁻³
J1	1	0.0208	4.340 10 ⁻³
J2a1b	3	0.0625	3.906 10 ⁻³
J2axJ2a1-bk	1	0.0208	4.340 10 ⁻³
J2b	3	0.0625	3.906 10 ⁻³
L	1	0.0208	4.340 10 ⁻³
R1b	27	0.5625	0.3164
Total de individuos	48		ΣF ² =0.3608
Probabilidad de coincidencia	0.3608		
Diversidad haplotípica	0.6528	(n/n-1)(1-Σpi ²).	



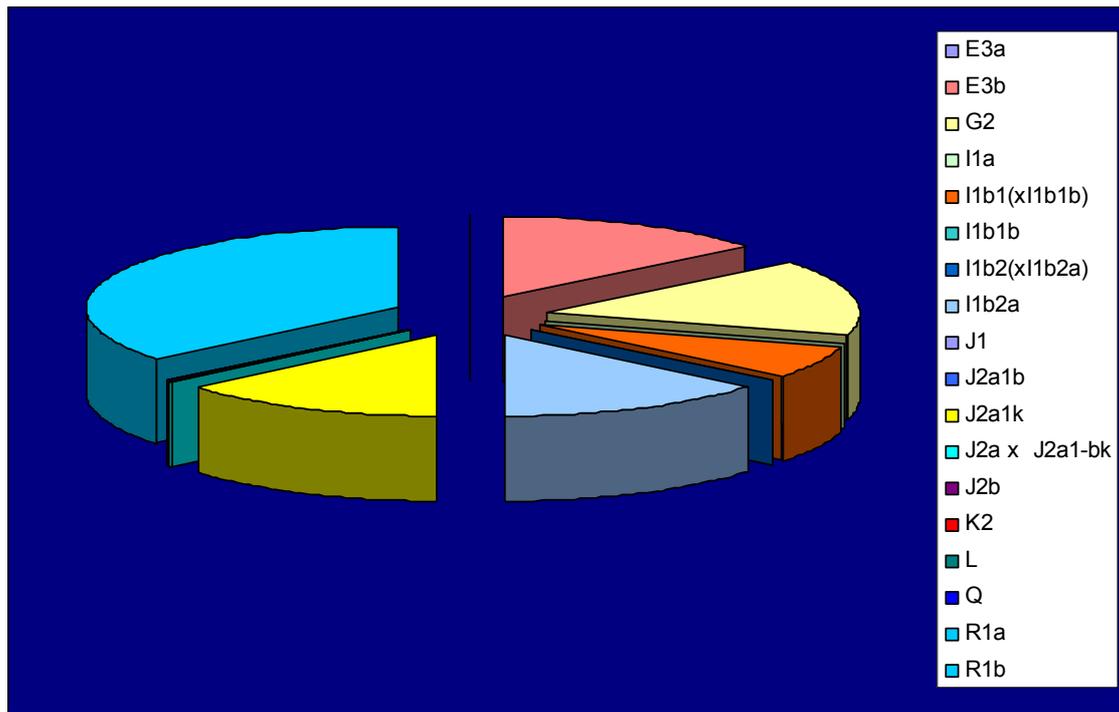
LOMBARDÍA:

HAPLOGRUPOS	N	F	F ²
E3b	4	0.0769	5.917 10 ⁻³
G2	3	0.0577	3.328 10 ⁻³
I1a	2	0.0384	1.478 10 ⁻³
I1b2a	2	0.0384	1.478 10 ⁻³
I1b2(xI1b2a)	1	0.0192	3.698 10 ⁻³
J1	2	0.0384	1.478 10 ⁻³
J2a1b	1	0.0192	3.698 10 ⁻³
J2a1k	2	0.0384	1.478 10 ⁻³
J2axJ2a1-bk	1	0.0192	3.698 10 ⁻³
K2	2	0.0384	1.478 10 ⁻³
R1b	32	0.6154	0.3786
Total de individuos	52		ΣF ² =0.4063
Probabilidad de coincidencia	0.4063		
Diversidad haplotípica	0.6053	(n/n-1)(1-Σpi ²).	



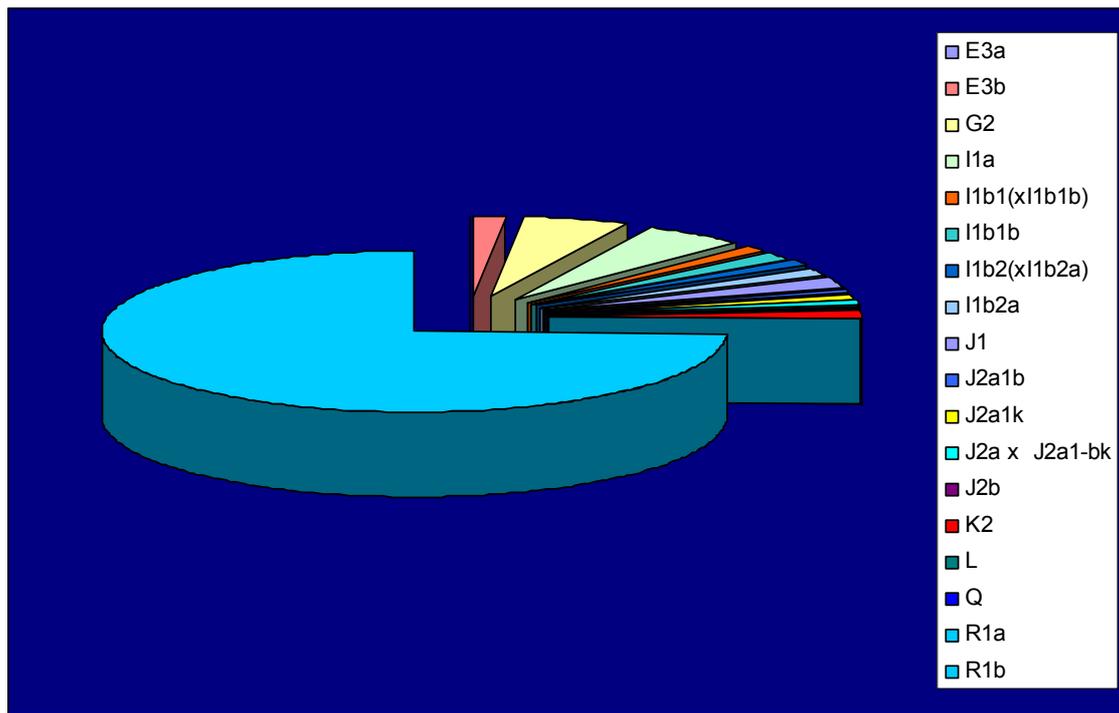
PIAMONTE:

HAPLOGRUPOS	N	F	F ²
E3b	2	0.1428	0.0204
G2	2	0.1428	0.0204
I1b2a	2	0.1428	0.0204
I1b(xI1b1b)	1	0.0714	5.102 10 ⁻³
J2a1k	2	0.1428	0.0204
R1b	5	0.3571	0.1275
Total de individuos	14		ΣF ² =0.2142
Probabilidad de coincidencia	0.2142		
Diversidad haplotípica	0.8462	(n/n-1)(1-Σpi ²).	



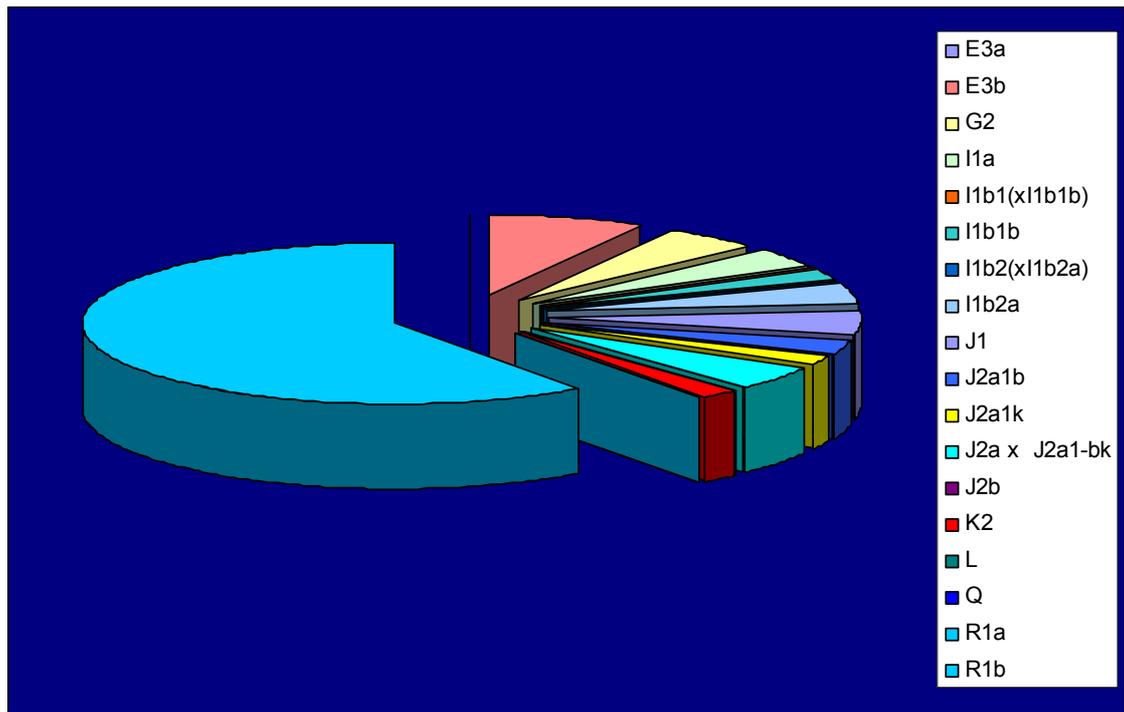
ESPAÑA REFERENCIA:

HAPLOGRUPOS	N	F	F ²
E3b	1	0.0169	2.872 10 ⁻³
G2	3	0.0508	2.585 10 ⁻³
I1a	3	0.0508	2.585 10 ⁻³
I1b1b	1	0.0169	2.872 10 ⁻³
I1b2(xI1b2a)	1	0.0169	2.872 10 ⁻³
I1b2a	1	0.0169	2.872 10 ⁻³
I1b(xI1b1b)	1	0.0169	2.872 10 ⁻³
J1	1	0.0169	2.872 10 ⁻³
J2a1k	1	0.0169	2.872 10 ⁻³
J2axJ2a1-bk	1	0.0169	2.872 10 ⁻³
K2	1	0.0169	2.872 10 ⁻³
R1b	44	0.7457	0.5561
Total de individuos	59		ΣF ² =0.5871
Probabilidad de coincidencia	0.5871		
Diversidad haplotípica	0.4200	(n/n-1)(1-Σpi ²).	



ITALIA REFERENCIA:

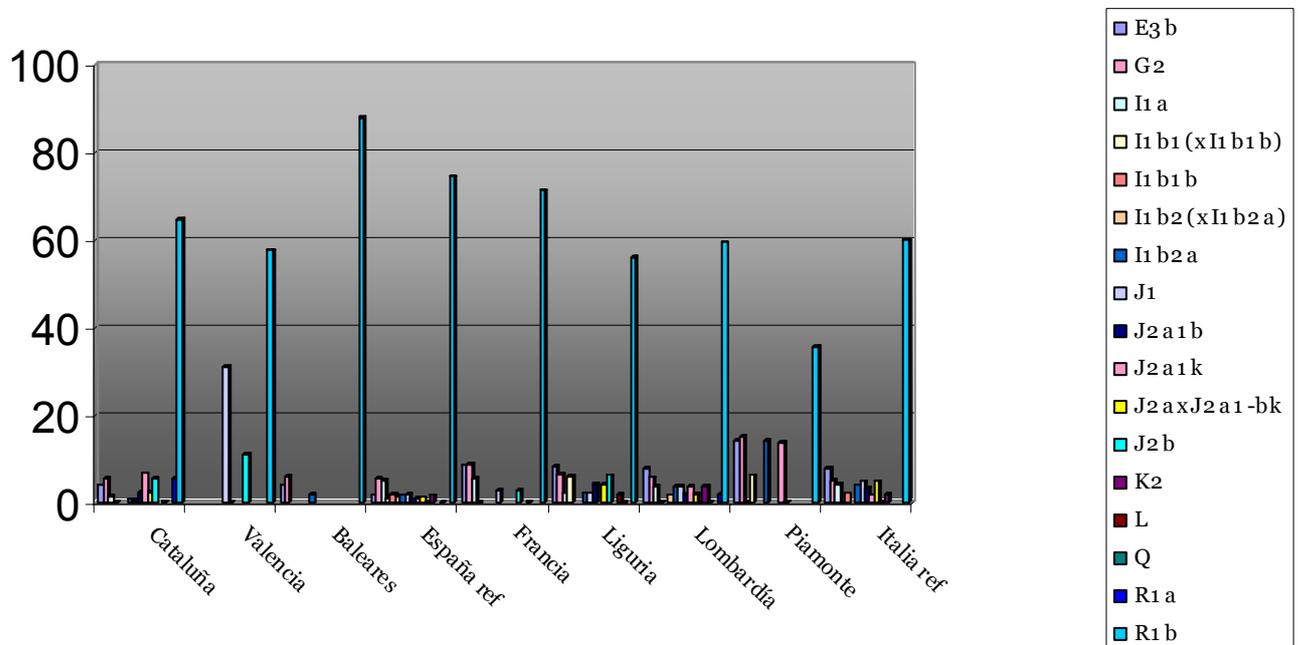
HAPLOGRUPOS	N	F	F ²
E3b	4	0.08	6.4 10 ⁻³
G2	3	0.06	3.6 10 ⁻³
I1a	2	0.04	1.6 10 ⁻³
I1b1b	1	0.02	4 10 ⁻⁴
I1b2a	2	0.04	1.6 10 ⁻³
J1	2	0.04	1.6 10 ⁻³
J2a1b	1	0.02	4 10 ⁻³
J2a1k	1	0.02	4 10 ⁻⁴
J2axJ2a1-bk	3	0.06	3.6 10 ⁻³
K2	1	0.02	4 10 ⁻⁴
R1b	30	0.6	0.36
Total de individuos	50		ΣF ² =0.38
Probabilidad de coincidencia	0.38		
Diversidad haplotípica	0.6326	(n/n-1)(1-Σpi ²).	



Es curioso, que la mayor diversidad haplotípica, según los haplogrupos, la tenga Piamonte, que es la población de la que menos individuos disponíamos para estudiar. Por otra parte, y como cabe esperar, la población que menor diversidad haplotípica tiene es Baleares, ya que al tratarse de una isla, es lógico que tenga una menor diversidad.

A continuación se expone una tabla con un resumen de todos los datos obtenidos de los haplogrupos, con el porcentaje de cada uno de ellos según la población:

Haplogrupo	Italia Ref	España Ref	Francia Col.	Liguria Col.	Lombardía Col.	Piamonte Col.	Cataluña Col.	Valencia Col.	Baleares Col.
E3a	0.004%				0.006%				
E3b	7.974%	1.759%	8.574%	8.314%	7.710%	14.28%	4.001%		4%
G2	4.972%	5.517%	8.811%	6.46%	5.767%	15.08%	5.599%		6%
I1a	4.308%	5.097%	5.474%	2.125%	3.848%	0.078%	1.6%		
I1b1(xI1b1b)	0.01%	1.114%	0.006%	5.904%	0.202%	6.271%	0.002%		
I1b1b	2.21%	1.829%							
I1b2(xI1b2a)		1.4322%			1.704%				
I1b2a	3.99%	1.695%		2.083%	3.825%	14.28%	0.797%		2%
J1	4.864%	2.041%	2.831%	2.131%	3.835%		0.475%	31.078%	
J2a1b	3.148%	0.753%		4.229%	2.063%		2.316%	0.016%	
J2a1k	1.652%	1.056%			3.829%	13.88%	6.624%		
J2a x J2a1-bk	4.822%	1.175%		4.091%	1.806%	0.264%	2.26%		
J2b	0.03%	0.263%	2.866%	6.25%	0.006%		5.586%	11.112%	
K2	2%	1.695%			3.840%				
L			0.006%	2.025%	0.006%		0.337%		
Q		0.027%		0.077%					
R1a					1.925%		5.6%		
R1b	60%	74.549%	71.428%	56.25%	59.613%	35.71%	64.8%	57.778%	88%





Hospital Real; concurso de carrozas (1913)
AUTOR: José Martínez Rioboó

Fundación Rodríguez-Acosta
Donación Martínez Sóna.

5.-DISCUSIÓN

El genoma humano contiene un gran número de segmentos de ADN, caracterizado por secuencias cortas de un número variable de repeticiones en tándem (VNTRs). Los primeros métodos para la caracterización de las regiones VNTRs se basaban en el análisis mediante Southern blot, pero a pesar de ser éste un método bien establecido y una técnica estándar en biología molecular (Schneider *et al.*, 1991), conlleva una serie de desventajas en su aplicación a la identificación genética de vestigios biológicos de interés forense. Este método no posee la resolución suficiente para la detección de pequeñas diferencias en el tamaño de una o dos unidades de repetición entre dos fragmentos de restricción relativamente grandes. Además, se requiere una cantidad de ADN de alto peso molecular de más de 50 ng para el análisis con sondas específicas, que en muchos casos es difícil de obtener cuando se trabaja con vestigios biológicos.

Estos problemas han sido solucionados con el desarrollo de las técnicas de amplificación de loci VNTRs y STRs mediante PCR. Esta técnica debido a su gran sensibilidad, no requiere que el ADN sea de alto peso molecular, característica importante ya que los vestigios biológicos frecuentemente se encuentran en un avanzado estado de degradación. Esta alta sensibilidad hace también que se requieran cantidades mínimas de ADN lo que hace posible el estudio de muestras muy pequeñas y la realización de contra-análisis o contraperitaciones, esto posibilita que los análisis puedan ser repetidos en caso de fallos técnicos o cuando se necesita una confirmación de los resultados obtenidos. En general para la aplicación de estas técnicas se obtendrá suficiente ADN para poder realizar un segundo análisis o contraperitación en otro laboratorio forense. Por el contrario, el tipado de ADN mediante el método de Southern requiere mayor cantidad de ADN y de alto peso molecular, por lo que en la mayoría de los casos la repetición de los análisis no es posible. Por tanto, gracias a esta técnica se ha incrementado el éxito de los análisis de ADN de muestras forenses, demostrando ser un método sencillo para el estudio de los polimorfismos de ADN, que reduce considerablemente el tiempo necesario para concluir los análisis. Otra ventaja importante es que mediante PCR se obtienen distribuciones discretas de alelos de manera que las frecuencias alélicas estimadas son más correctas y por tanto más exacto el valor de probabilidad de coincidencia calculado (Budowle *et al.*, 1991^a; Chakraborty *et al.*, 1991; Deka *et al.*, 1992a, b; Rand *et al.*, 1992; Edwards *et al.*, 1992).

El análisis de las regiones microsatélites o STRs aunque contienen menor número de alelos que las regiones minisatélites o VNTRs amplificables, poseen otras ventajas sobre estos últimos. Los STRs debido al pequeño tamaño de las unidades de repetición (de uno a cinco nucleótidos) son más fácilmente amplificables y no

muestran amplificación preferencial de alelos debido a la pequeña diferencia de tamaños que existen entre ellos (Kimpton *et al.*, 1993). La similar eficiencia en la amplificación de los diferentes alelos supone una ventaja en las aplicaciones forenses, especialmente en los casos de agresiones sexuales donde las muestras forenses muestran mezcla celular (víctima y agresor). Por otra parte el pequeño tamaño de los alelos de los STRs hace posible también el diseño conjunto de primers, de tal manera que, el tamaño de los productos de amplificación no se superpongan en el transcurso de la electroforesis (Edwards *et al.*, 1991), y cuando las condiciones de PCR son idénticas, pueden ser amplificados simultáneamente múltiples loci STRs en una única reacción (Kimpton *et al.*, 1993). Como resultado, con estos procedimientos se aumenta la información obtenida, reduciendo la cantidad de ADN, el tiempo y el trabajo necesario para los estudios de identificación genética.

5.1.- CONSIDERACIONES METODOLÓGICAS.

5.1.1.- SOBRE EL MÉTODO DE EXTRACCIÓN DE ADN.

Hemos utilizados dos tipos distintos de extracción según la muestra de la que disponíamos. El método de extracción con resinas quelantes (Chelex[®] 100 Resin, Bio Rad Labs, Richmond, CA) es más rápido, sencillo y barato que el método fenol-cloroformo y minimiza la pérdida de material por manipulación, pero su principal inconveniente es que el ADN se conserva menos tiempo y contiene más impurezas, por lo que sólo lo hemos utilizado en muestras de sangre, ya que ésta contiene mucha cantidad de ADN y en buenas condiciones.

Sin embargo, para muestras de saliva se utilizó el método fenol-cloroformo (Valverde *et al.*, 1993) por presentar las siguientes ventajas:

- Se realiza todo el proceso en microtubos desechables.
- El proceso se acelera mediante digestión a alta temperatura, obteniéndose un buen rendimiento en la cantidad de ADN.
- El ADN extraído conservado a 4°C no plantea problemas de degradación. Además puede ser almacenado a -20°C durante largo periodo de tiempo sin perder cualidades.
- Se obtiene un ADN con menos impurezas, evitando así sustancias inhibitoras de la PCR.

5.1.2.- SOBRE EL MÉTODO DE CUANTIFICACIÓN.

La cuantificación se ha realizado tras la extracción mediante geles de agarosa al 0.8%. Esta técnica presenta el inconveniente de que el ADN que se cuantifica no es necesariamente humano, ya que también se reflejaría ADN de bacterias, hongos, etc... Existen otros métodos más específicos para cuantificar sólo y exclusivamente ADN humano pero son técnicas más costosas y laboriosas.

Como nuestras muestras se extraen justo después de haber sido obtenidas, no hay tiempo para la contaminación con microorganismos como ocurre en las muestras que se recogen en la escena del crimen, o muestras que hayan estado a la intemperie mucho tiempo. Por ello, en nuestro caso se pueden utilizar geles de agarosa al 0.8%.

5.1.3.- SISTEMA ELECTROFORÉTICO.

Para comprobar el resultado de la amplificación se utilizó el instrumento de electroforesis capilar ABI Prism 310, como ya se describió en materiales y métodos. La mayoría de los laboratorios forenses utilizan, actualmente, secuenciadores automáticos de ADN para la lectura de STRs, ya que estos sistemas electroforéticos van a facilitar enormemente la estandarización de los resultados. Los patrones obtenidos se representan por picos, una vez captada la señal fluorescente por el ordenador en tiempo real.

Una vez ha terminado la electroforesis, se observa que ésta ha transcurrido correctamente mediante el GeneScan, donde a parte de ver el tamaño en pares de bases de los picos obtenidos, se comprueba que la migración de los fragmentos ha sido normal observando la migración del estándar de tamaño que migra con las muestras. Posteriormente, con el Genotyper se asignarán los alelos correspondientes según los ladders.

Es muy importante saber reconocer los picos anómalos (que representan bandas electroforéticas) que no corresponden a alelos y que pueden deberse a distintos artefactos. Entre ellos destacan las starter band, debido a su alta frecuencia, que son unas bandas de poca intensidad que aparecen inmediatamente antes de un alelo. Un ejemplo de este artefacto lo tenemos delante del alelo 13 de esta imagen:



Las características de los STRs utilizados en nuestro estudio, junto con un buen rendimiento de la amplificación, hacen que la incidencia de estos fenómenos haya sido mínima, sin llegar a plantear dificultades.

5.2.- ESTUDIO GENÉTICO-POBLACIONAL.

5.2.1.- CARACTERIZACIÓN DE LAS FRECUENCIAS ALÉLICAS DE LAS POBLACIONES ESTUDIADAS.

Hemos llevado a cabo un análisis genético poblacional de 125 individuos apellidados Colón y no emparentados de la población catalana, 50 individuos apellidados Colón de la población balear, 44 individuos apellidados Colón de la población valenciana, 35 individuos apellidados Colón del sur de Francia, 46 individuos apellidados Colón de Liguria, 50 individuos apellidados Colón de Lombardía, 13 individuos apellidados Colón de Piamonte, 58 individuos no apellidados Colón de la población española y 49 individuos no apellidados Colón de la población italiana. Todos estos individuos se han estudiado para los sistemas DYS456, DYS389I, DYS390, DYS389II, DYS458, DYS19, DYS385, DYS393, DYS391, DYS439, DYS635, DYS392, Y GATA H4, DYS437, DYS438 y DYS448.

- ✚ **Sistema DYS456:** Hemos encontrado, en un bajo porcentaje de la población catalana y de la población francesa un alelo no contemplado en el ladder (<13). En todas la demás poblaciones, incluidas las de referencia todos los picos hallados se encuentran contemplados en el ladder. La población que más diversidad genética tiene para este sistema es la población apellidada Colón de Lombardía, por encima incluso de las poblaciones de referencia española e italiana. La población que menos diversidad genética tiene para este sistema es la población apellidada Colón de las Baleares. Los alelos más representados en todas la poblaciones son el 15 y el 16, siendo además los únicos encontrados en la población de Piamonte. El alelo menos representado en la población, ni siquiera ha sido hallado en todas las poblaciones, es el 13.
- ✚ **Sistema DYS389I:** Todos los alelos hallados en este sistema han sido descritos en el ladder. La población que más diversidad génica para este sistema es la de los apellidados Colón de Piamonte, seguida de la población francesa y de las dos poblaciones de referencia (italiana y española). La

población que menos diversidad génica para este sistema es la población apellidada Colón de las Baleares. El alelo mejor representado en todas las poblaciones es el 13. Mientras que el alelo menos representado es el 11, que de hecho, sólo se encuentra en la población catalana y en las dos de referencia.

- ✚ **Sistema DYS390:** Al igual que en el sistema anterior, todos los alelos encontrados están representados en el ladder. La población que más diversidad génica presenta para este sistema es la población de referencia española, seguida muy de cerca de la población italiana de Liguria y Piamonte. La población que menos diversidad génica presenta para este sistema, de nuevo, es la población balear. Los alelos mejor representado en todas las poblaciones son el 23 y 24, que de hecho son los dos únicos hallados en la población valenciana. El alelo peor representado en las poblaciones es el 21, que en la mayoría de las poblaciones está ausente (sólo se encuentra en la población catalana y en la de referencia española).
- ✚ **Sistema DYS389II:** Todos los alelos encontrados en las distintas poblaciones han sido descritos en el ladder. La población que mayor diversidad génica presenta para este sistema es la población de Piamonte, seguida muy de cerca por la población catalana y la italiana de referencia. La población que menos diversidad génica presenta para este sistema es la población balear. El alelo mejor representado en todas las poblaciones es el 29, si bien de forma particular, en la población catalana es superado por el 30, y en la población valenciana por el 28. El alelo peor representado es el 27, que sólo lo hemos encontrado en la población catalana y en las dos de referencia, al igual que el 32 que sólo lo encontramos en la balear y en la referencia italiana.
- ✚ **Sistema DYS458:** Para este sistema, encontramos, en seis de las nueve poblaciones diferentes, ocho alelos fuera del ladder. En primer lugar, en la población valenciana encontramos un pico que se encuentra situado entre el 16 y el 17, restando los tamaños de estos dos picos y dividiéndolo entre el número de nucleótidos que se repiten deducimos que el pico en cuestión es el 16.2. Esto sólo se puede realizar cuando el pico que no se encuentra descrito en el ladder se encuentra entre dos picos que sí lo están, es decir, ni al principio del mismo ni al final. En segundo lugar encontramos, en la población francesa y en la población italiana de Liguria un pico por debajo del 14, que como es el primero lo denominamos <14 (ya que no podemos calcularlo). En tercer lugar, encontramos, en la población española de referencia un pico por encima del 19, que al igual

que en el caso anterior no podemos calcular y lo denominamos >19. En cuarto lugar, encontramos en la población italiana de Liguria el 17.2 y en la población italiana de Lombardía encontramos dos alelos el 17.2 y el 18.2. Finalmente, en la población italiana de referencia encontramos dos alelos fuera del ladder que son el 19.2 y el >20. La población que más diversidad génica posee para este sistema es la valenciana, siendo no obstante, bastante mayor que para el resto de los sistemas vistos anteriormente. La población que menos diversidad génica para este sistema es la población balear. El alelo menos representado en las poblaciones es el 19, que, además, no está presente en todas. Los alelos que no se encuentran recogidos en el ladder, también se presentan en una baja proporción a excepción del 16.2 encontrado en la población valenciana, que de hecho en dicha población es el pico mejor representado. En el resto de poblaciones, los alelos mejor representados son el 16 y 17.

- ✚ **Sistema DYS19:** En este sistema, de nuevo encontramos un alelo que no contiene el ladder. Concretamente, en la población italiana de Liguria encontramos un pico entre el 13 y el 14, que siguiendo los cálculos descritos anteriormente, averiguamos que se trata del 13.2. Todos los demás alelos encontrados en el resto de las poblaciones están contemplados en el ladder. La población que mayor diversidad génica presenta para este sistema es la población italiana de Piamonte, resultando curioso, ya que es la población, que por las circunstancias menos representantes tenemos. La población que menos diversidad génica posee es, como de costumbre la balear. El alelo mejor representado en todas la poblaciones es el 14, mientras que los peor representados son el 12, 13 y 16, que no se encuentran en todas las poblaciones.
- ✚ **Sistema DYS385:** Todos los alelos encontrados, para este sistema, en las distintas poblaciones están contenidos en el ladder. La población que mayor diversidad génica presenta para este sistema, con diferencia, es la población italiana de Liguria. Mientras que la población que menos diversidad posee para este sistema es la balear. El alelo peor representado en las poblaciones es el 20, que de hecho sólo se encuentra en la población catalana y la italiana de referencia, mientras que los más representado en las distintas poblaciones son el 11 y el 14.
- ✚ **Sistema DYS393:** En este sistema, encontramos, de nuevo, un alelo que no está contemplado en el ladder. Concretamente, lo encontramos en la población italiana de referencia y es el 15.1. La población que mayor diversidad génica presenta es

Liguria, seguida muy de cerca de Lombardía. La población que menos diversidad génica contiene es la balear. El alelo mejor representado en las poblaciones es el 13, mientras que el peor representado es el 15, que no se encuentra en todas las poblaciones, y en las que lo hemos encontrado es en muy baja proporción.

- ✚ **Sistema DYS391:** Todos los alelos encontrados en las distintas poblaciones, para este sistema, se encuentran contenidos en el ladder, es decir, no encontramos ninguna variante alélica. La población que mayor diversidad génica posee, para este sistema, es la población de Lombardía. La población que menor diversidad génica posee es la población balear. Los alelos mejor representados en las distintas poblaciones, para este sistema, son el 10 y el 11, siendo los únicos encontrados en algunas de las poblaciones. Mientras que los alelos peor representados en las distintas poblaciones son el 1 y el 12, no encontrándose en todas ellas.
- ✚ **Sistema DYS439:** Todos los alelos encontrados en estas poblaciones, se encuentran contenidos en el ladder, no encontramos ninguna variante alelica. La población que mayor diversidad génica presenta para este sistema es la catalana, seguida por las dos poblaciones de referencia, española e italiana. La población que menos diversidad génica presenta para este sistema, al igual que para los sistemas anteriores, es la población balear. El alelo mejor representado para estas poblaciones es el 12, a excepción de la población valenciana que es el 11. Los alelos peor representados son el 10 y el 14, que no se encuentran en todas las poblaciones.
- ✚ **Sistema DYS635:** En este sistema encontramos dos variantes alélicas en dos de las poblaciones, concretamente en la población catalana y la población valenciana. Esta variante alelica, está por debajo del alelo 20 por lo que lo denominamos <20. La población que mayor diversidad génica presenta para este sistema es la de Piamonte. Mientras que la población que menos diversidad génica presenta para este sistema es la balear. El alelo mejor representado para las distintas poblaciones es 23. El alelo peor representado es el 26 que sólo lo encontramos en la población española de referencia, seguido del alelo 25.
- ✚ **Sistema DYS392:** En este sistema, de nuevo, encontramos una variante alélica. Concretamente, en la población italiana de Lombardía, encontramos el alelo 13.1. La población que mayor diversidad génica presenta para este sistema es Liguria. Por el

contrario, la población que menor diversidad génica presenta para este sistema es la balear. El alelo mejor representado para este sistema es el 13, aunque en la población de Piamonte es superado por el 11. Mientras que el alelo peor representado para este sistema es el 16 que sólo lo encontramos en Liguria, seguido de 15 y el 14.

- ✚ **Sistema Y GATA H4:** En este caso, volvemos a encontrar dos variantes alélicas en las poblaciones de referencia italiana y española. En este caso se trata de un alelo de tamaño superior al 13, por tanto, lo denominamos >13. La población que mayor diversidad génica posee para este sistema es la de Liguria, seguida de Lombardía. La población que menor diversidad génica posee es la de Baleares. Los alelos mejor representados en estas poblaciones son el 11 y el 12. Mientras que el alelo peor representado en estas poblaciones es el 9, que sólo lo hemos hallado en la población italiana de Liguria, seguido por el 10 y por supuesto la variante alélica.
- ✚ **Sistema DYS437:** En este sistema no encontramos ninguna variante alélica. La población que mayor diversidad génica presenta para este sistema es la de Piamonte. Por el contrario, la población que menor diversidad génica posee para este sistema es la de Baleares, siendo muy inferior al resto de las poblaciones estudiadas. Los alelos mejor representados son el 14 y 15. Mientras que el alelo peor representado para este sistema es el 17, que sólo lo hemos encontrado en la población italiana de referencia.
- ✚ **Sistema DYS438:** En este sistema, de nuevo, encontramos una variante alélica. En este caso se trata de un alelo de mayor tamaño que el 13 (>13), hallado en la población italiana de Lombardía. La población que mayor diversidad génica posee es la población de Piamonte. Por el contrario, la población que menor diversidad génica posee es, de nuevo, la población balear. El alelo mejor representado en este sistema es el 12, mientras que el peor representado es el 8 y >13 que sólo se encuentran en Lombardía, seguidos por el 9.
- ✚ **Sistema DYS448:** En este último sistema estudiado, tampoco encontramos ninguna variante alélica. La población que mayor diversidad génica posee es la población italiana de referencia, seguida de la población valenciana. La población que menor diversidad génica posee es de nuevo la balear. El alelo mejor representado para este sistema es el 19, a excepción de la población balear, en la cual se encuentra superado por el 18. Los alelos peor representados para este sistema son el 21 y 22.

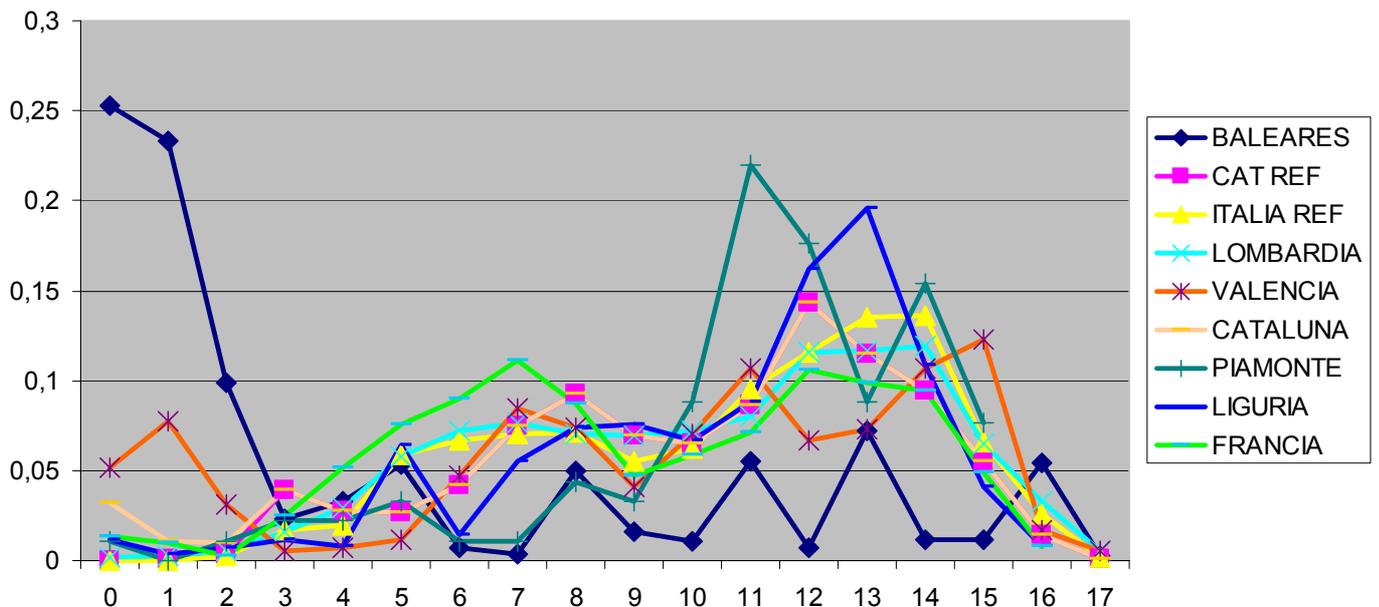
En resumen, a continuación se muestra una tabla donde se puede observar las poblaciones con mayor y con menor diversidad para cada uno de los marcadores:

Marcadores	Mayor diversidad	Menor diversidad
DYS456	Lombardía	Baleares
DYS389I	Piamonte	Baleares
DYS390	España ref	Baleares
DYS389II	Piamonte	Baleares
DYS458	Valencia	Baleares
DYS19	Piamonte	Baleares
DYS385	Liguria	Baleares
DYS393	Liguria	Baleares
DYS391	Lombardía	Baleares
DYS439	Cataluña	Baleares
DYS635	Piamonte	Baleares
DYS392	Italia y España ref	Baleares
Y GATA H4	Liguria	Baleares
DYS437	Piamonte	Baleares
DYS438	Piamonte	Baleares
DYS448	Italia Ref	Baleares

Nos llama la atención que, por una parte las poblaciones que presentan una mayor diversidad para cada uno de los marcadores está muy repartido. La población que es más divergente para un mayor número de marcadores es Piamonte, que curiosamente, es la población de la que menos individuos pudimos conseguir. Esto podría llevarnos a pensar que si tuviéramos un mayor número de individuos podrían aumentar en el número de marcadores, pudiendo llegar a conseguirlo en todos los marcadores. Por otra parte, la población que presente menos diversidad en todos los marcadores es Baleares. Esto podría ser explicado por su situación geográfica, ya que al tratarse de una isla, posee mayores barreras geográficas que el resto de poblaciones estudiadas.

Posteriormente, procedemos a hacer una representación gráfica con los datos obtenidos en Arlequín de las mismach. Esto es comúnmente usado como medio de representar la diversidad de diferentes datos moleculares, conociendo, también, la pairwise differences. La mismatch distribution es apropiada para datos donde las diferencias discretas de los alelos pueden ser contadas. La distribución del número de estas diferencias entre los alelos nos indica la diversidad genética. La diversidad que aparece de cada una

de las muestras, y la forma de la distribución de la misma (representación gráfica en un histograma) nos indica la historia poblacional y concretamente los episodios de expansión popular. Así pues, la forma de la distribución es muy informativa. Procedamos a ver qué ocurre con nuestras poblaciones:



La gráfica anterior representa la distribución del número medio de diferencias entre parejas para los marcadores de cromosoma Y estudiados en los 9 grupos poblacionales realizados en el presente estudio. Como se observa todas las poblaciones presentan una distribución unimodal, signo de expansión demográfica, y, característico de las poblaciones Europeas, a excepción de la población balear. Como podemos observar la distribución de Baleares tiene muchas subidas y bajadas bruscas. Esto indica que tiene mayor consanguinidad que el resto de las poblaciones, lo que también nos explicaría por qué es la población que menos diversidad alélicas presenta en cada uno de los sistemas. Este mayor grado de consanguinidad, podría ser debido a que se trata de una Isla, donde las barreras geográficas son mayores.

A partir de las frecuencias alélicas presentes en las distintas poblaciones se calculó, también, la distancia genética de Reynolds (Reynolds *et al.* 1983). Esta distancia asume que no existe mutación y que la deriva genética es el único factor responsable del cambio en las frecuencias de los alelos. A diferencia de otras distancias, esta

medida no asume un tamaño efectivo de la población constante e igual para todas las poblaciones. De hecho, las fluctuaciones en el tamaño poblacional juegan un importante papel en la magnitud de la distancia puesto que pueden modificar el efecto de la deriva genética.

El cálculo de la matriz de distancias de Reynolds fue calculada con el programa Arlequín. A partir de la matriz de distancias de Reynolds, entre todas las poblaciones estudiadas generada con el programa Phylip se construyeron árboles sin raíz, utilizando los métodos de reconstrucción filogenético de Neighbor-Joining (Saitou y Nei 1987).

El método de Neighbor-Joining se basa en el agrupamiento sucesivo de las diferentes líneas para la construcción del árbol, asignando la longitud de las ramas a medida que éstas se van uniendo. Este método de reconstrucción no asume reloj molecular, por lo que los árboles generados con él no presentan raíz. Para la construcción de un árbol de este tipo a partir de la matriz de distancias de Reynolds entre las diferentes poblaciones, se empleó el programa Neighbor del paquete Phylip 3.5c (Felsenstein 1993).

Al igual que se hizo con la matriz de Reynolds, se realizó también un árbol filogenético a partir de la matriz de distancia de Cavalli.

A continuación se muestran los árboles que obtuvimos con los datos de las poblaciones estudiadas.

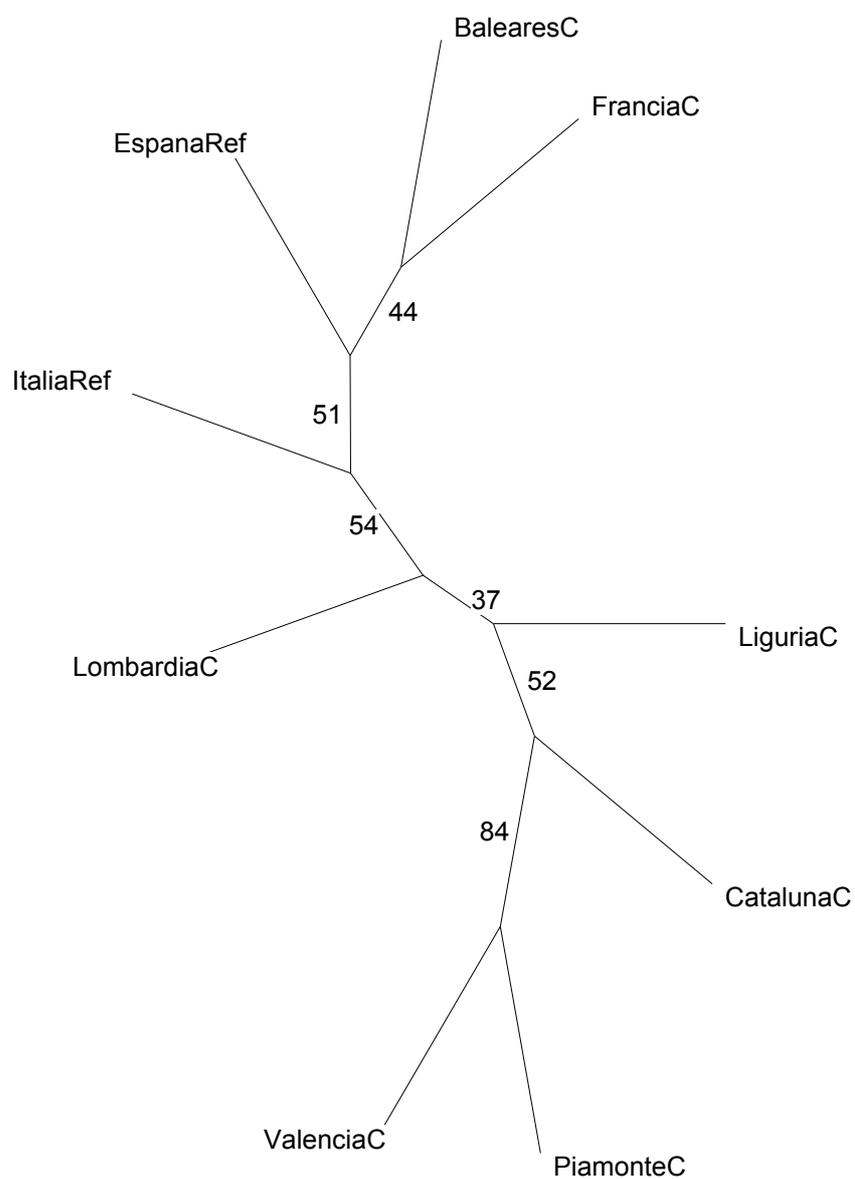


Figura 5.1: Árbol obtenido a partir de la matriz de distancias de Reynolds.

La figura anterior, sería el árbol construido con la matriz de Reynolds. Ahora vamos a ver qué ocurre con el árbol construido a partir de la matriz de distancia de Cavalli.

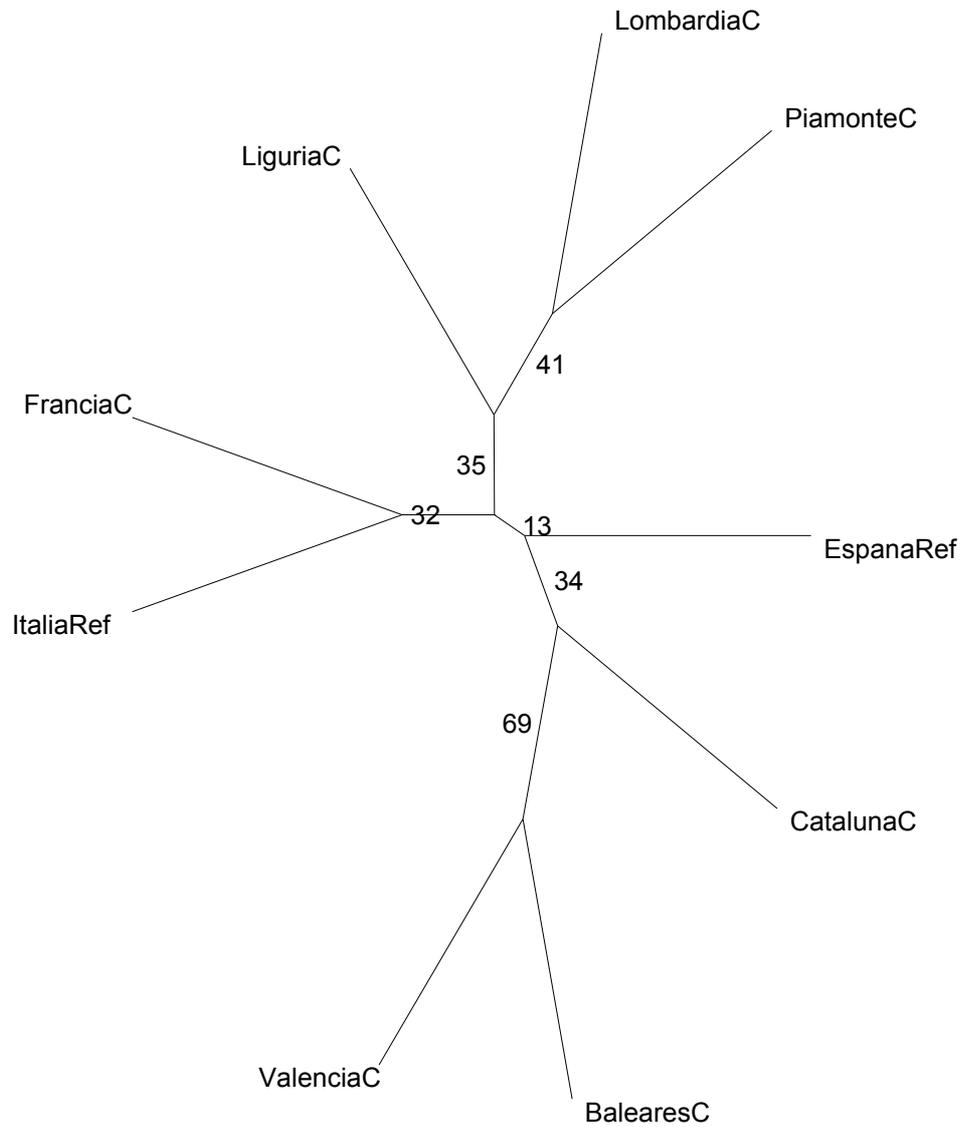


Figura 5.2: Árbol obtenido a partir de la matriz de distancias de Cavalli-Sforza.

En el primer árbol, que es el de Reynolds, vemos que por un lado está Valencia, Piamonte y Cataluña. Liguria y Lombardía son parecidos, al igual que las dos poblaciones de referencia. Baleares está cerca de Francia y justo en el lado opuesto a Valencia, Piamonte y Cataluña. Los números (valores de bootstrap) indican la fracción de las 100 replicaciones de bootstrap que respaldan cada nudo, por tanto, indican el grado de confianza estadístico de los árboles.

En el segundo árbol, que es el de Cavalli, observamos que Valencia y Baleares están por un lado, por otro Lombardía y Piamonte. Francia es más similar a Italia, y, España está en el lado opuesto.

Como podemos ver, ambos árboles son distintos. Cuando los valores del árbol no están por encima de 90%, no son significativos, por eso salen distintos.

Las relaciones genéticas entre las agrupaciones realizadas en nuestro estudio han sido evidenciadas por los análisis de las distancias genéticas y a través de técnicas de análisis multivariante como el escalamiento multidimensional (multidimensional scaling o MDS). A continuación se muestra la representación, en dos dimensiones de las frecuencias alélicas.

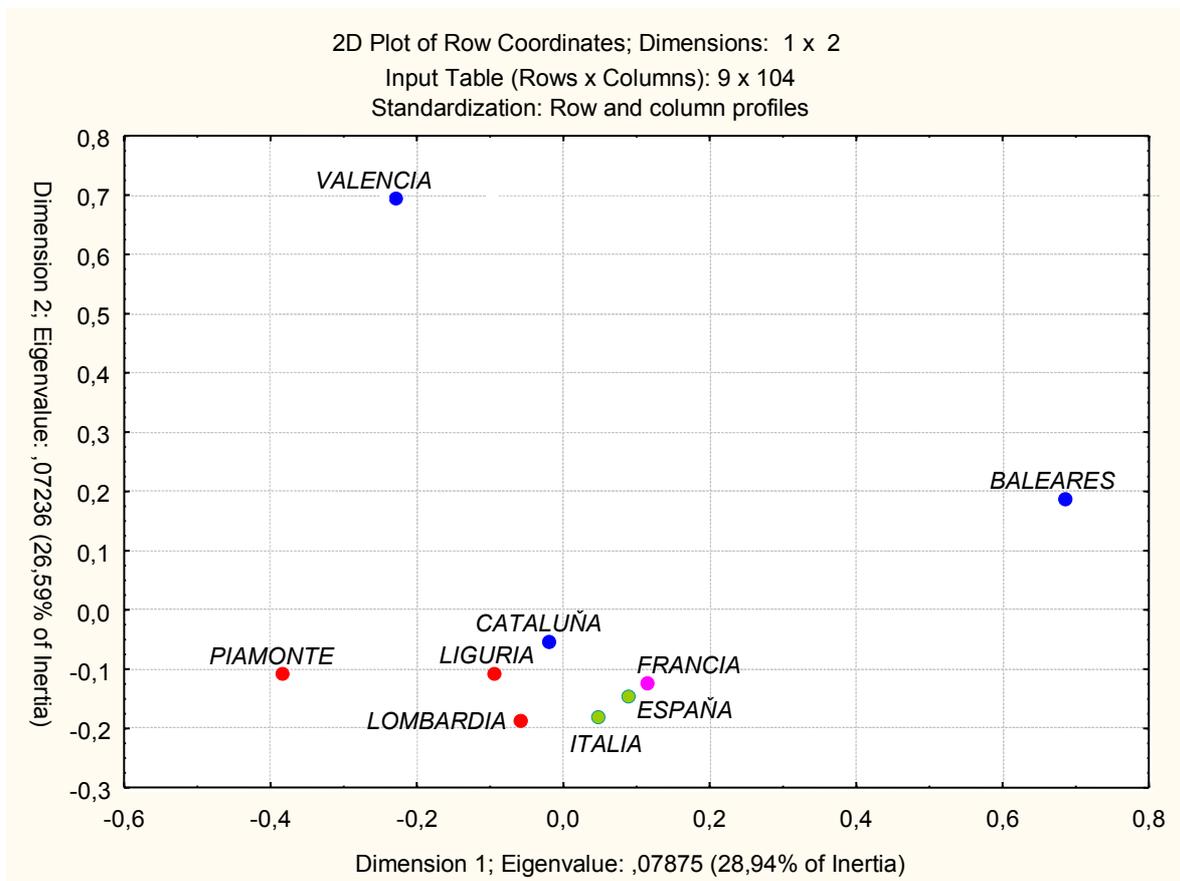


Figura 5.3: MDS de las poblaciones estudiadas con respecto a las frecuencias alélicas.

El análisis de los componentes principales se hace con todos los alelos de todas las poblaciones. Como podemos observar, Valencia, Baleares y Piamonte salen separadas del resto de las poblaciones, que están todas agrupadas la misma zona del gráfico.

Con esto habríamos acabado el análisis con respecto a las frecuencias alélicas de nuestras poblaciones.

5.2.2.- CARACTERIZACIÓN DE LOS HAPLOTIPOS Y HAPLOGRUPOS DE LAS POBLACIONES ESTUDIADAS.

Podemos empezar este apartado estudiando las diferencias, si las hay, en la diversidad en los haplotipos y haplogrupos para las distintas poblaciones. A continuación exponemos una tabla con estos datos.

REGION	DIVERSIDAD HAPLOTIPICA/HAPLOTIPOS	DIVERSIDAD HAPLOTIPICA/HAPLOGRUPOS
Baleares	0.7479	0.2208
España Ref	1	0.4200
Francia	0.9864	0.4689
Cataluña	0.9673	0.5514
Valencia	0.9445	0.5697
Lombardía	0.9985	0.6053
Italia Ref	0.9999	0.6326
Liguria	0.9879	0.6528
Piamonte	0.9890	0.8462

En esta tabla podemos observar que:

- ✚ En primer lugar, la población que presenta una menor diversidad, lo hace tanto para los haplotipos como los haplogrupos y es Baleares. Como recordaremos es la población que menor diversidad presentaba también en relación a los alelos.
- ✚ La población que presenta una mayor diversidad con respecto a los haplotipos, que es Italia de referencia, no es la que lo presenta para los haplogrupos, que sería Piamonte.
- ✚ Aunque existe una cierta correlación entre los valores de diversidad con respecto a haplotipos y haplogrupos excepto en una de las poblaciones, en la cuál, hay unas diferencias importantes. Esta población es España de referencia, en la cual, no coincide ninguno de los haplotipos (son todos distintos entre

sí), pero sin embargo, la mayoría de las muestras pertenecen al mismo haplogrupo, que por otra parte, como veremos más adelante, es de los más comunes por esta zona.

Compararemos las distintas zonas para ver si los distintos grupos poblacionales comparten haplotipos.

Al realizar de nuevo un estudio de comparación haplotípica general, vemos que es menor el numero de coincidencias fuera de los grupos, que dentro de los grupos, como era de esperar.

Encontramos una relación clara entre las muestras de Cataluña y Francia, donde tres haplotipos de las muestras francesas coinciden con otros tres haplotipos catalanes, en un caso haplotipos individuales, en otro, una muestra francesa comparte con un haplotipo compartido por tres muestras catalanas, pero lo que las hace más coincidentes son el mayor grupo catalán (14 indiv.) que es igual que el mayor grupo francés (4 indiv.)

Dos haplotipos son compartidos entre individuos catalanes y valencianos, siendo uno de ellos uno de los grupos mayoritarios de Valencia.

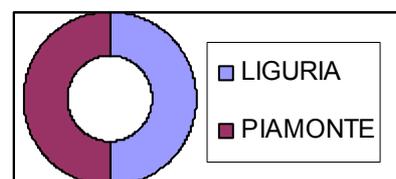
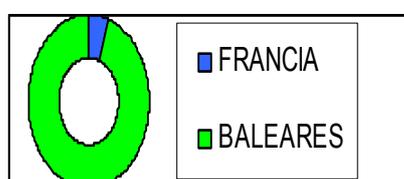
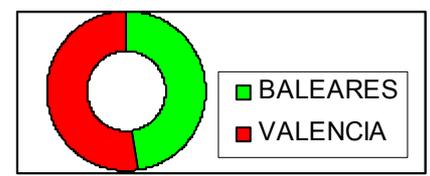
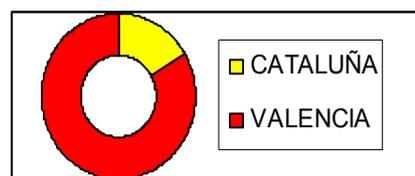
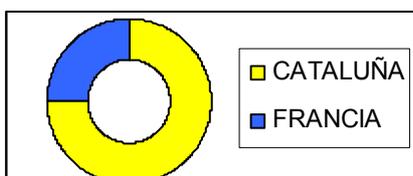
También encontramos coincidencias entre Baleares con Valencianos y Franceses, siendo una muestra francesa coincidente con el gran haplotipo Balear.

CI-F-I
C-29-F14
C-54-87-89-F-16

C-21-V-III
C-49-V-3
B-39 V-12

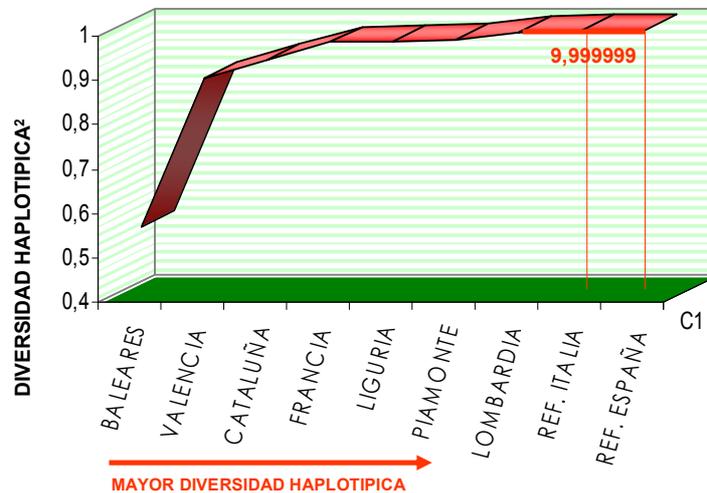
F-36-B-I
LI-37-PI-11

No encontramos relaciones entre las muestras italianas, con las francesas o españolas, pero tampoco aparecen coincidencias entre sus grupos, salvo un haplotipo compartido entre Liguria y Piamonte



En todos los casos disminuye la diversidad al sesgar la muestra. Como era de esperar al condicionar la muestra a personas que se apelliden Colon, van a aparecer más haplotipos coincidentes que en los grupos control de referencia.

Las poblaciones con menor divergencia son Baleares, Valencia y Cataluña (en este orden).



Las poblaciones con mayor diversidad, eliminando los grupos de referencia de España e Italia (que son los que mayor diversidad poseen) son: Lombardia, Piamonte y Liguria (en este orden).

Los haplotipos mayoritarios se van acentuando (al igual que disminuye la diversidad haplotípica) según un eje que podríamos situar desde Lombardia, hasta Valencia, acentuándose aun más en el grupo de Baleares (donde existe un haplotipo que engloba al 70% de las muestras y la diversidad haplotípica disminuye hasta 0.7469 o 0.5938, si incluimos los mutados).



Los grupos españoles junto con el grupo francés comparten, entre ellos, varios de sus haplotipos. Lo que no ocurre con los grupos italianos que no coinciden con Españoles y Franceses, pero tampoco entre ellos.

Siguiendo el linaje paterno del apellido Colon nos hemos encontrado con grandes diferencias de distribución, dependiendo de la zona geográfica en las que nos hallemos. Los componentes paternos representados por los linajes y haplotipos del cromosoma Y presentan proporciones distintas entre las poblaciones de Cataluña, Baleares, Valencia, Francia, Lombardia, Liguria y Piamonte.

La presencia de estas grandes diferencias encontradas en las distintas poblaciones, repartidas en un mismo eje, nos pueden llevar a pensar que ha ocurrido una secuencia de efectos fundadores.

El esquema con el que se representa la disminución de diversidad de haplotipos en la zona geográfica estudiada nos puede hacer pensar que el flujo génico ha aparecido en el norte de Italia y se ha distribuido por la cuenca mediterránea hasta llegar a Valencia, y se ve mas marcado este flujo en el efecto fundador que se puede ver en Baleares.

Esto nos lleva a la conclusión de que el apellido Colon habría aparecido en Italia, y después emigrado al resto de las zonas.

Pero esta teoría no podemos sustentarla al no compartirse entre las distintas poblaciones los perfiles genéticos, y no poder rastrear estos perfiles.

También hay que tener en cuenta que las oscilaciones aleatorias que pueden aparecer pueden ser producto de la mutación, que es muy marcada en los STR estudiados del Cromosoma Y, al haber pasado como mínimo mas de 600 años de la aparición del apellido, las diferencias se pueden marcar mucho mas.

Otro problema de base es si el apellido Colon español tiene un origen común al Colombo italiano.

Por tanto tenemos que esperar al análisis estadístico de los SNPs (que nos aportaran mayor información) para confirmar los resultados de este estudio.

Los análisis que hemos aplicado a las frecuencias alélicas los vamos a aplicar ahora pero con respecto a los genotipos. En primer lugar, ciñéndonos a las poblaciones que hemos estudiado en este trabajo. Posteriormente, sirviéndonos de la bibliografía, vamos a comparar nuestras poblaciones con otras que geográficamente están muy alejadas. De esta forma, empezamos representando nuestras poblaciones en un árbol de distancias con la matriz de Reynolds.

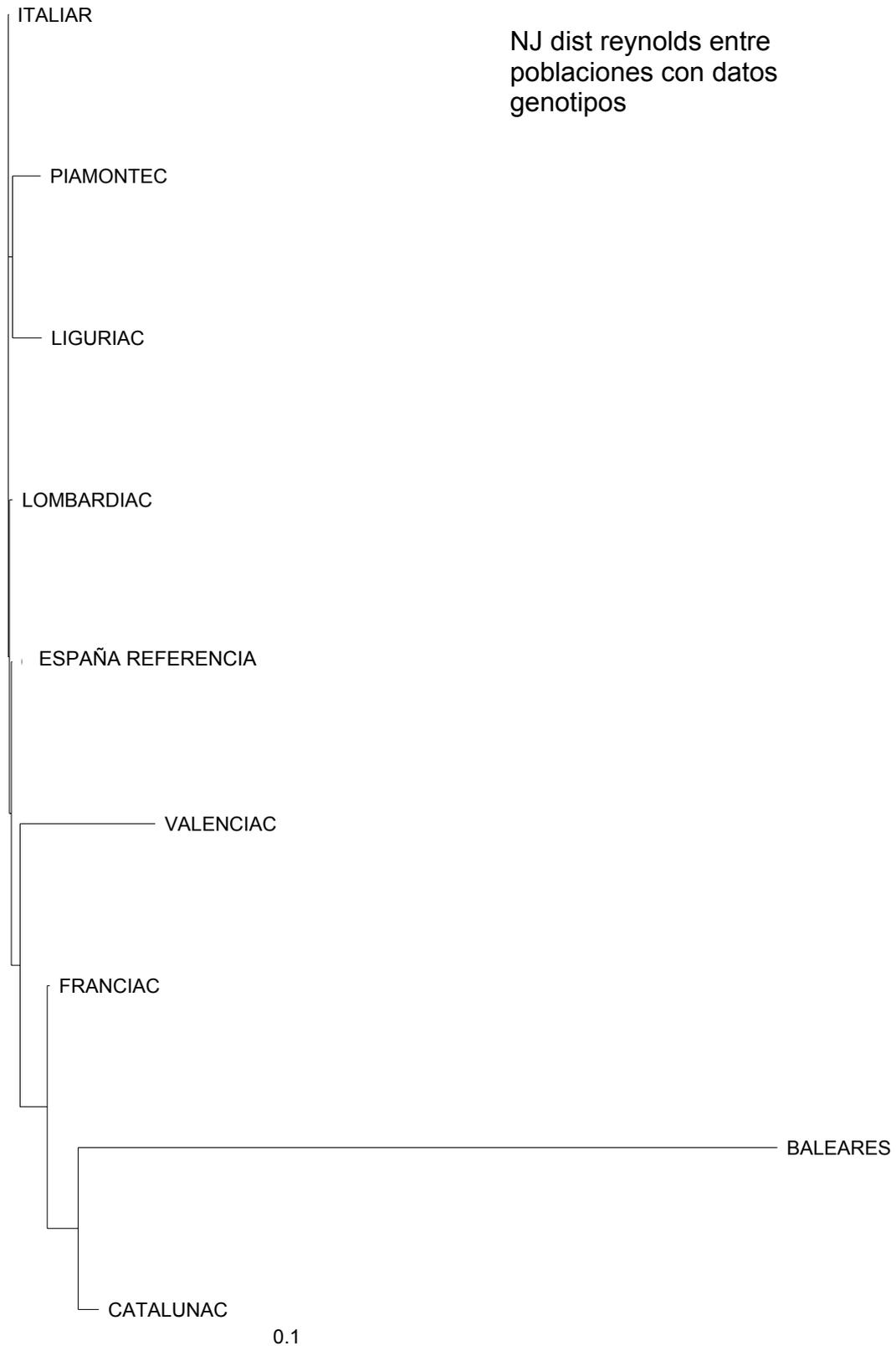


Figura 5.4: Árbol de distancias de Reynolds con respecto a los datos haplotípicos

Como podemos observar, las poblaciones italianas salen parecidas entre sí, y más variables que las españolas. España de referencia sale separando las poblaciones italiana y las españolas (apellidadas Colón), y francia se nos queda en medio de las españolas. Finalmente, podemos observar que, la población que más se nos separa de todo el grupo es Baleares.

Posteriormente, vamos a realizar el análisis de correspondencias, pero en este caso según los genotipos (anteriormente lo habíamos hecho según las frecuencias alélicas). Además esta representación la vamos a realizar en tres dimensiones.

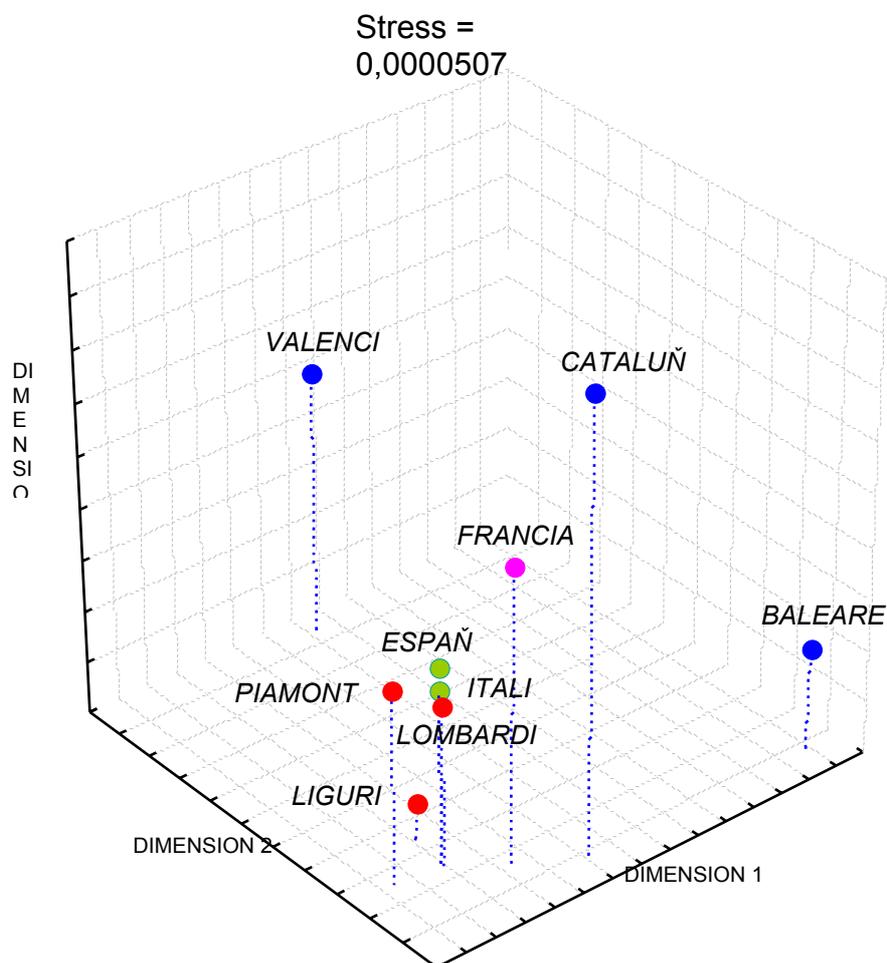


Figura 5.5: MDS de las poblaciones estudiadas con respecto a los haplotipos.

En este caso, las poblaciones que más se alejan del resto son Valencia, Cataluña y Baleares, es decir, las tres poblaciones españolas apellidadas Colón. Sin embargo, España de referencia se sitúa con todas las demás poblaciones estudiadas, que se encuentran muy cercanas unas de otras. De este grupo de poblaciones, la que, en cierto modo, se aleja un poco sería Francia que se sitúa entre las tres poblaciones españolas y el grupo de las italiana y España de referencia.

También podríamos observar las relaciones que existen entre las poblaciones que hemos estudiado y los haplogrupos que hemos obtenido de estas poblaciones. Esto nos daría una gráfica tal que así:

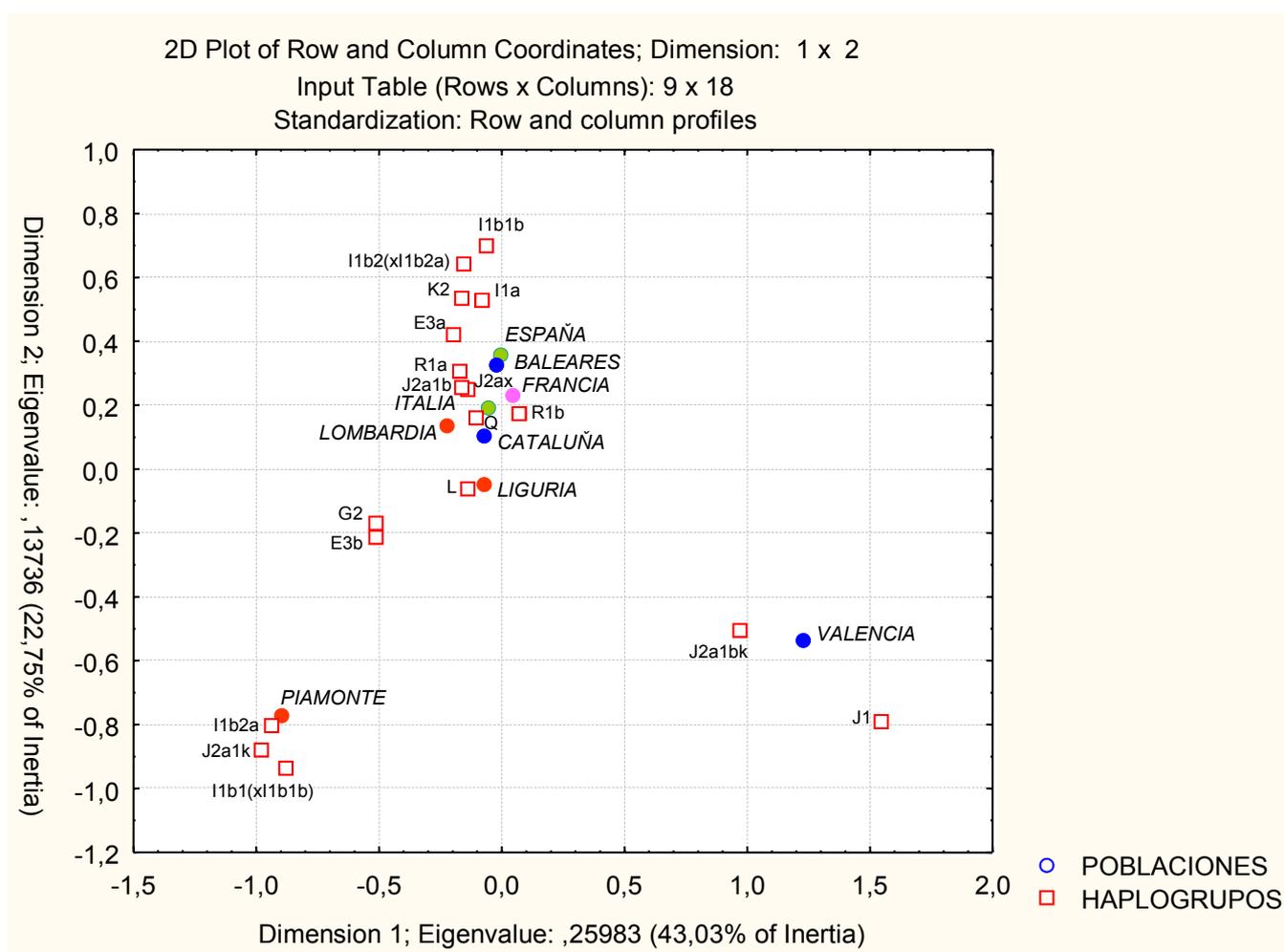


Figura 5.6: MDS de los haplotipos y haplogrupos de nuestras poblaciones.

Si tenemos en cuenta que, el haplogrupo mayoritario es el R1b, podemos observar que dicho haplogrupo se encuentra entre la mayoría de las poblaciones. Este haplogrupo se encuentra un poco más alejado de valencia donde un 57% de la población es R1b pero un 31% de la población es J1 (por tanto aunque es mayoritario

también cuenta con un alto porcentaje de segundo haplogrupo). Al igual ocurre con Piamonte donde un 35% de la población es R1b, un 15% es G2, un 14% es E3b, un 14% es I1b2a y un 13% es J2a1k, por tanto, aunque es el mayoritario, un alto porcentaje de la población no lo posee. En todas las demás poblaciones el porcentaje de individuos R1b, se aleja mucho más del segundo haplogrupo más común.

Finalmente, sometemos a nuestras poblaciones a distintos análisis de la AMOVA. Vamos a intentar ver si existen diferencias culturales, geográficas etc... Para ello, seleccionamos distintos grupos y realizamos el análisis. Los distintos grupos seleccionados y los resultados obtenidos son los siguientes:

- ✚ Se realiza un solo grupo con Cataluña, Valencia y Baleares (es decir, todos los individuos apellidados Colón de España). Los resultados obtenidos fueron los siguientes:

Fuente de variación	Porcentaje de variación
Entre poblaciones	9.21
Dentro de las poblaciones	90.79

Observamos que las mayores variaciones son dentro de las mismas poblaciones que entre ellas.

- ✚ En segundo lugar se hace también un grupo, pero en este caso es de todos los individuos analizados de España, es decir, Cataluña, Valencia, Baleares y España de referencia.

Fuente de variación	Porcentaje de variación
Entre poblaciones	6.89
Dentro de las poblaciones	93.11

En este caso ocurre igual que el anterior, las mayores variaciones se producen dentro de las poblaciones, y se reducen en gran medida entre las mismas.

- ✚ En tercer lugar se compone también un solo grupo que contiene los individuos italianos apellidados Colombo, es decir, Liguria, Lombardía y Piamonte.

Fuente de variación	Porcentaje de variación
Entre poblaciones	0.73
Dentro de las poblaciones	99.27

Ocurre igual que en los casos anteriores, pero además podemos observar que, las diferencias que se producen entre las distintas poblaciones italianas son menores que las que se producen entre las distintas poblaciones españolas (un 0.73% frente a un 9.21%). También son superiores las diferencias ocurridas dentro de las mismas poblaciones.

- ✚ En cuarto lugar se realiza un solo grupo compuesto por todos los individuos italianos, es decir, Liguria, Lombardía, Piamonte y Italia de referencia.

Fuente de variación	Porcentaje de variación
Entre poblaciones	0.51
Dentro de las poblaciones	99.49

En este caso el porcentaje de variación es todavía menor que en el caso de los individuos italianos apellidados Colombo.

- ✚ El siguiente estudio que se realiza es un solo grupo compuesto por todos los individuos apellidados Colón o Colombo de España, Italia y Francia.

Fuente de variación	Porcentaje de variación
Entre poblaciones	5.23
Dentro de las poblaciones	94.77

Ocurre igual que en los casos anteriores las variaciones mayores ocurren dentro de las distintas poblaciones y bajan considerablemente entre ellas.

- ✚ Ahora probamos a hacer dos grupos, el primero compuesto por todos los individuos apellidados Colón de España, Francia e Italia y un segundo grupo compuesto por las poblaciones de referencia de España e Italia.

Fuente de variación	Porcentaje de variación
Entre poblaciones	0.87
Entre las poblaciones del grupo	4.42
Dentro de las poblaciones	96.45

Incluso al hacer dos grupos las mayores variaciones ocurren dentro de las mismas poblaciones, disminuyen un poco entre las poblaciones del mismo grupo y disminuyen todavía más entre todas las poblaciones.

- Finalmente realizamos el análisis componiendo tres grupos. El primer grupo está compuesto por todos los individuos italianos, incluida la población de referencia, el segundo grupo está compuesto por los individuos franceses y el tercer grupo por todos los individuos españoles, incluida la población de referencia.

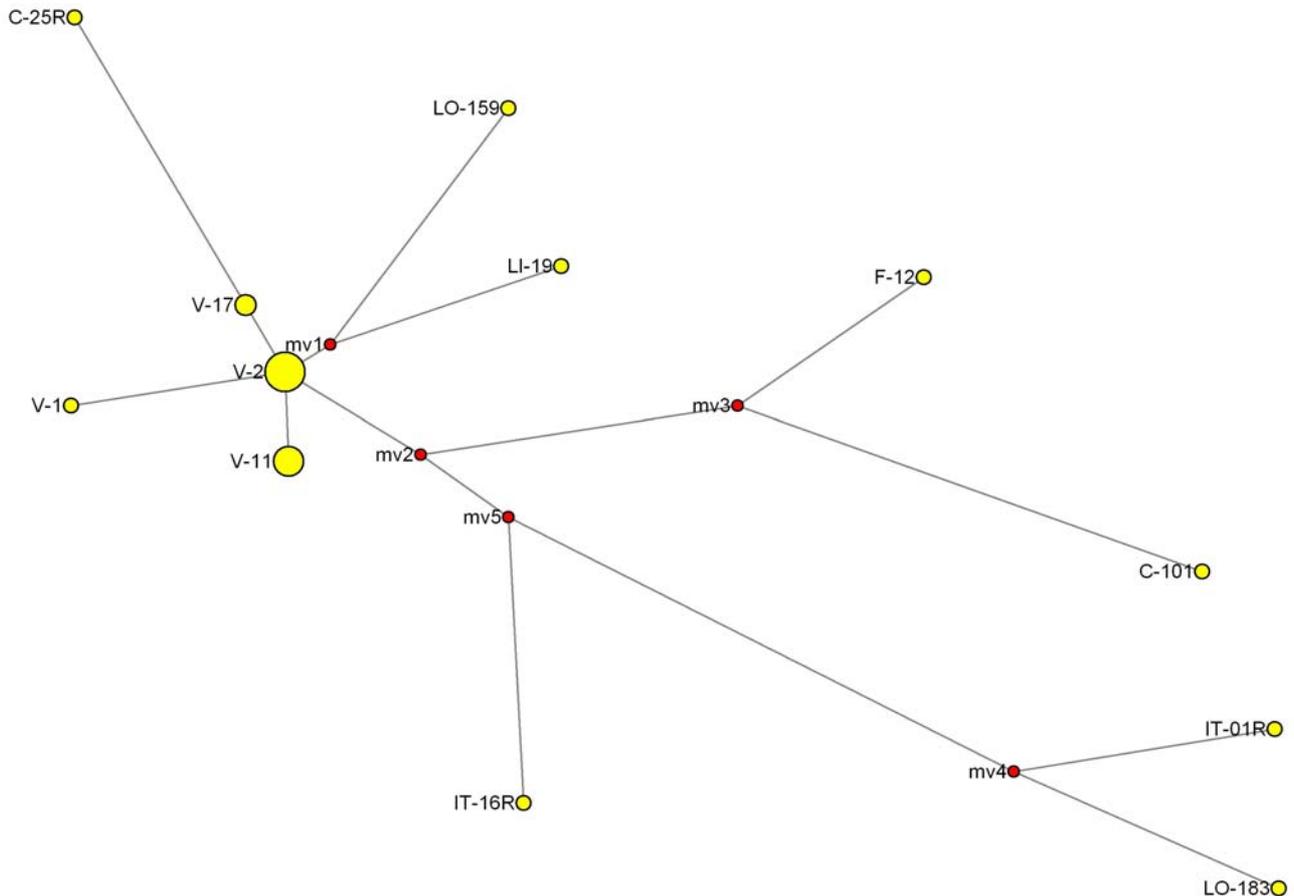
Fuente de variación	Porcentaje de variación
Entre poblaciones	0.75
Entre las poblaciones del grupo	4.53
Dentro de las poblaciones	96.22

En este caso las diferencias se mantienen muy parecidas al caso anterior, de hecho sólo varían un poco en los decimales.

En resumen, la variabilidad más alta en todos los casos es dentro de los mismos grupos. Mientras que la variabilidad más baja, siempre, es entre los distintos grupos (que en ambos casos sale 0). Sin embargo, existe una mayor variabilidad entre los apellidados Colon españoles (Cataluña, Valencia y Baleares), que los apellidados Colon Italianos (Liguria, Lombardía y Piamonte), concretamente 9.21% contra 0.73%.

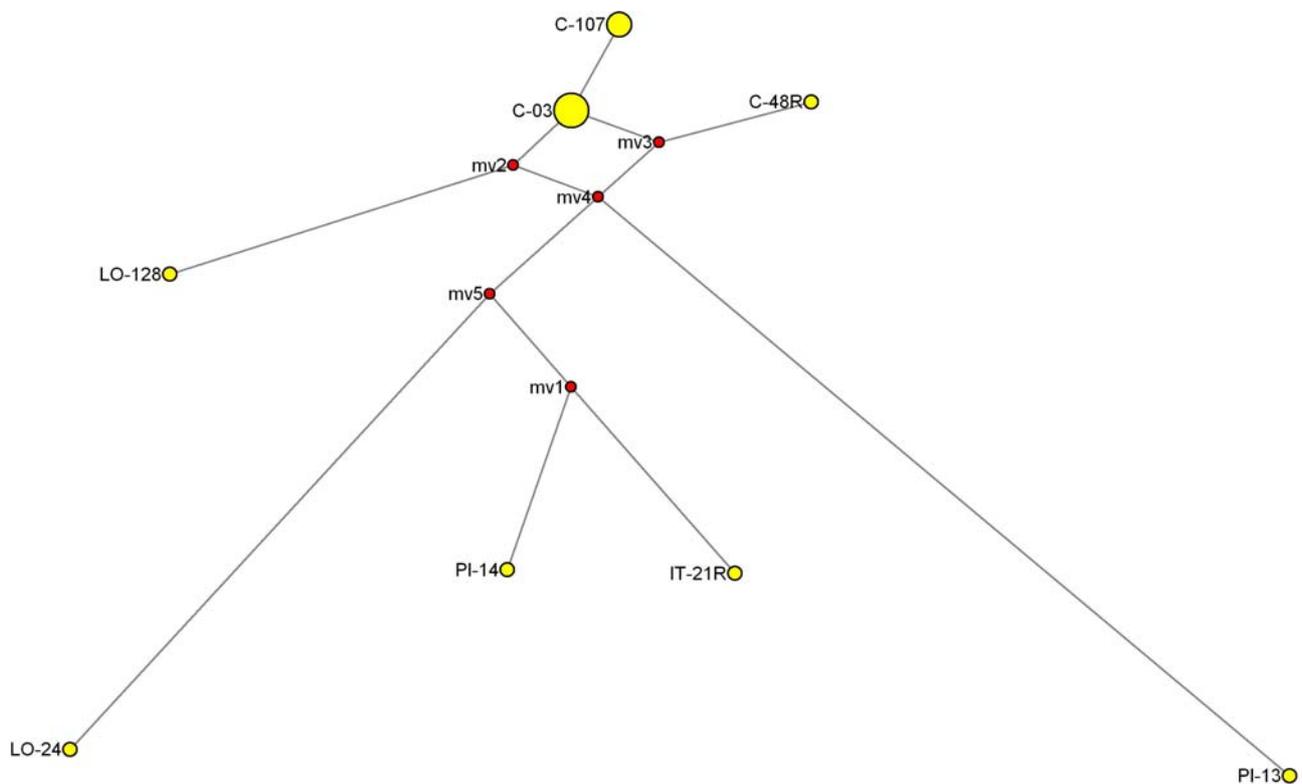
Finalmente, lo que se ha hecho es una comparación de las poblaciones según su haplogrupo, es decir, se han construido una serie de redes (network) para cada uno de los haplogrupos para intentar dilucidar cuales son los haplotipos ancestrales y cuales los derivados, con lo que se podría ver cuales son los Colones más antiguos. Los network se han hecho para todos los haplogrupos excepto para los que están compuestos por menos de cuatro individuos.

Haplogrupo J1:



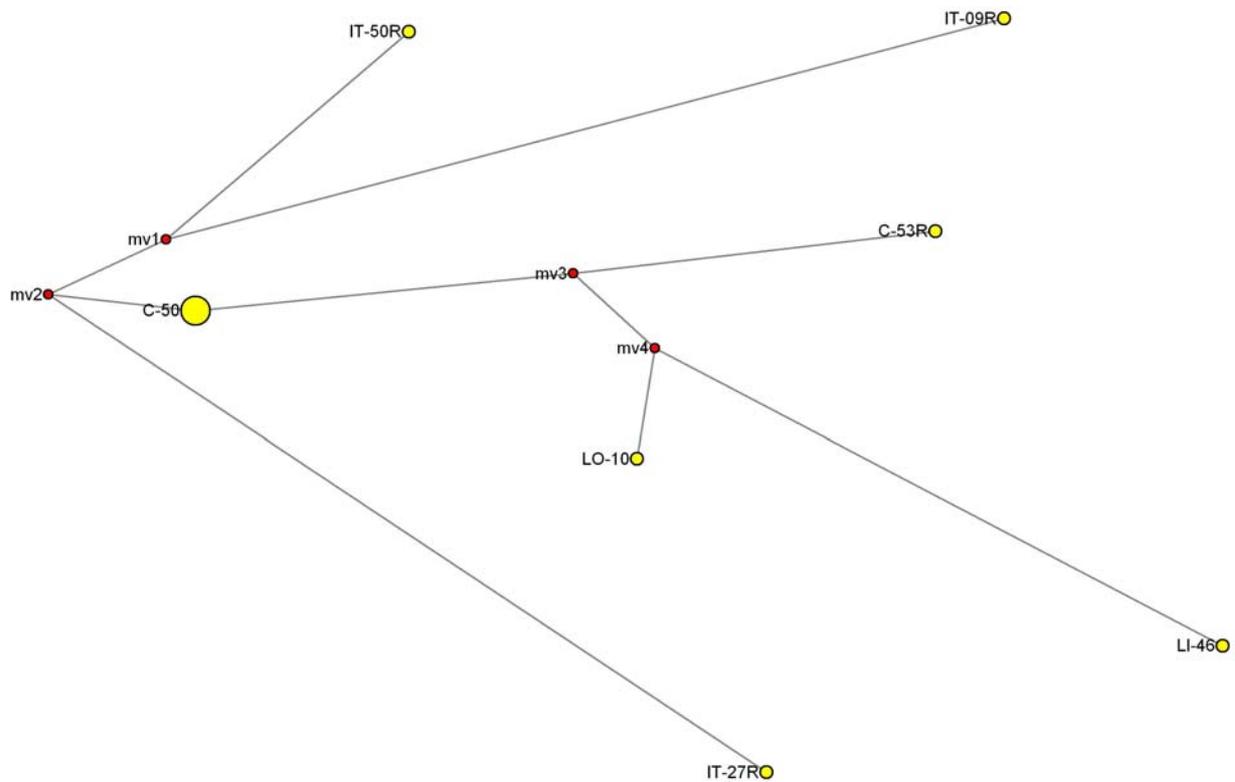
Es importante tener en cuenta que el tamaño de los círculos va en relación al número de individuos representados, es decir, se han agrupado los individuos que tienen igual haplotipo. En este network, observamos, en primer lugar que, a partir del origen o ancestro común (mv1) surgen tres muestras, la primera compuesta por varios individuos, ya que el tamaño del círculo es mayor, es la V-2 (individuos valencianos), al igual que uno de Lombardía (Lo-159) y uno de Liguria (Li-19). A partir de la muestra V-2, surgen varios haplogrupos valencianos que serían V-17, V-1 y V-11. A partir del V-17 surge uno haplotipo de la población española de referencia que sería el C-25R. Una serie de transformaciones o mutaciones en el grupo de individuos incluidos en V-2, van a ir dando lugar a todas las demás muestras de este haplogrupo, que son dos italianas de referencia (IT-16R, IT-01R), una francesa (F-12), una de Lombardía (Lo-183) y una de Cataluña (C-101).

Haplogrupo J2a1k:



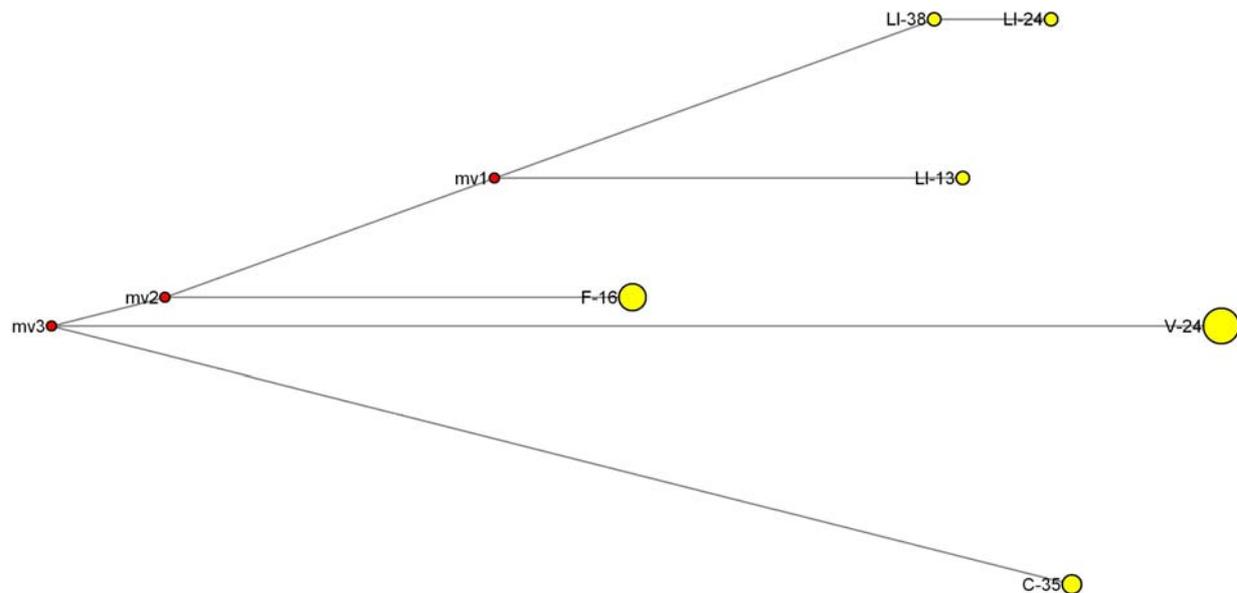
En este haplogrupo, podemos observar que, a partir del origen o ancestro común (mv1) surgen dos muestras ambas italianas, una es de Piamonte (Pi-14) y otra italiana de referencia (IT-21R). Una serie de mutaciones dan lugar a un conjunto de muestras de cataluña (ya que el tamaño del círculo es mayor) que sería C-03, y esta da lugar a otro grupo de individuos catalanes que son C-107. Posteriormente, una serie de mutaciones van a dar lugar a dos de Lombardía (Lo-24 y Lo-128), y a una de Piamonte que es Pi-13. La que más cambios ha sufrido desde el punto de partida es Lo-24, ya que es la que surge del último nudo que es mv5.

Haplogrupo J2axJ2a1-bk:



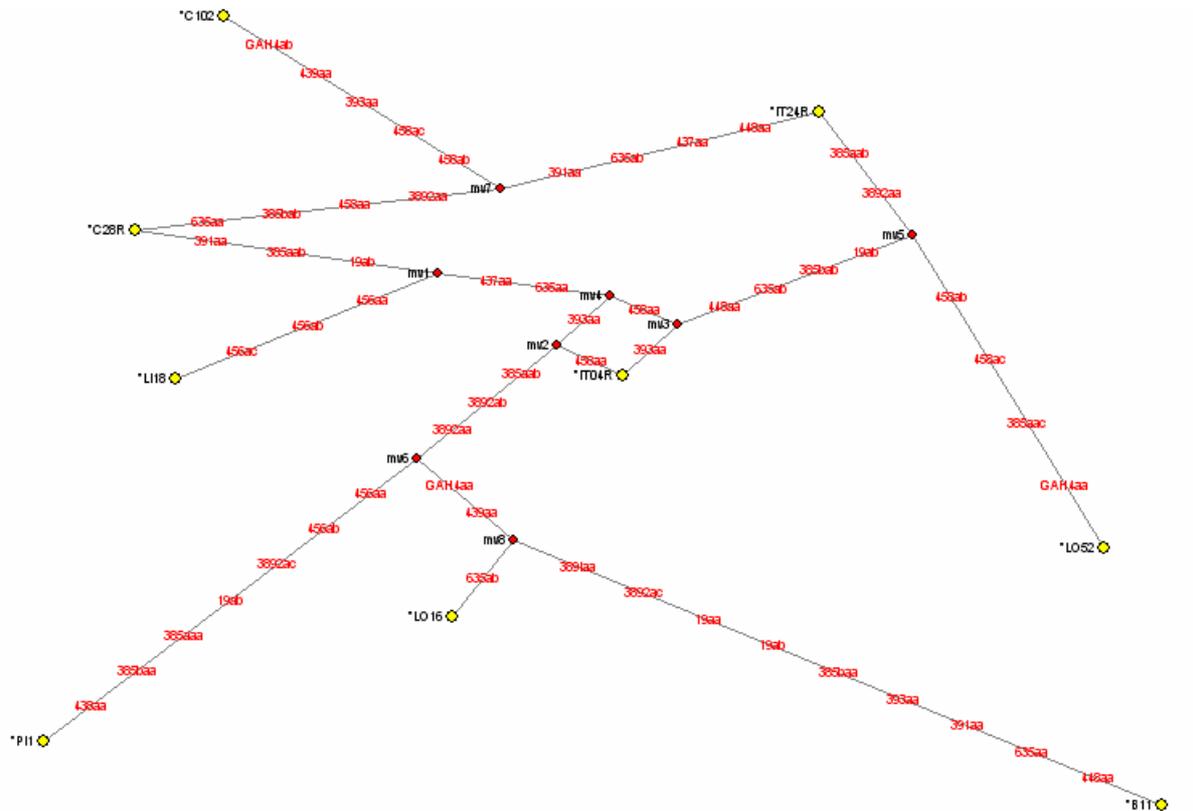
En este network podemos observar que, a partir del ancestro común surgen directamente dos muestras, ambas italianas de referencia que son IT-09R y IT-50R. Posteriormente, una serie de mutaciones distintas a las anteriores originan un conjunto de individuos catalanes que están contemplados en C-50, y una italiana de referencia que es IT-27R. Una serie de mutaciones ocurridas en C-50 van a dar lugar a una de Lombardía Lo-10, una de Liguria Li-46 y una española de referencia que es C-53R.

Haplogrupo J2axJ2a1-bk:



En este network podemos observar que, por un lado las tres muestras de Liguria, se encuentran agrupadas y cercanas a mv1. De este nudo surgen directamente Li-13 y Li-38, y a partir de Li-38 surge Li-24. Posteriormente, otras mutaciones diferentes a las anteriores van a dar lugar a un francesa (F-16), una valenciana (V-24) y una catalana (C-35). Las dos muestras españolas surgen del mismo nudo que es mv3.

Haplogrupo I1b2a:

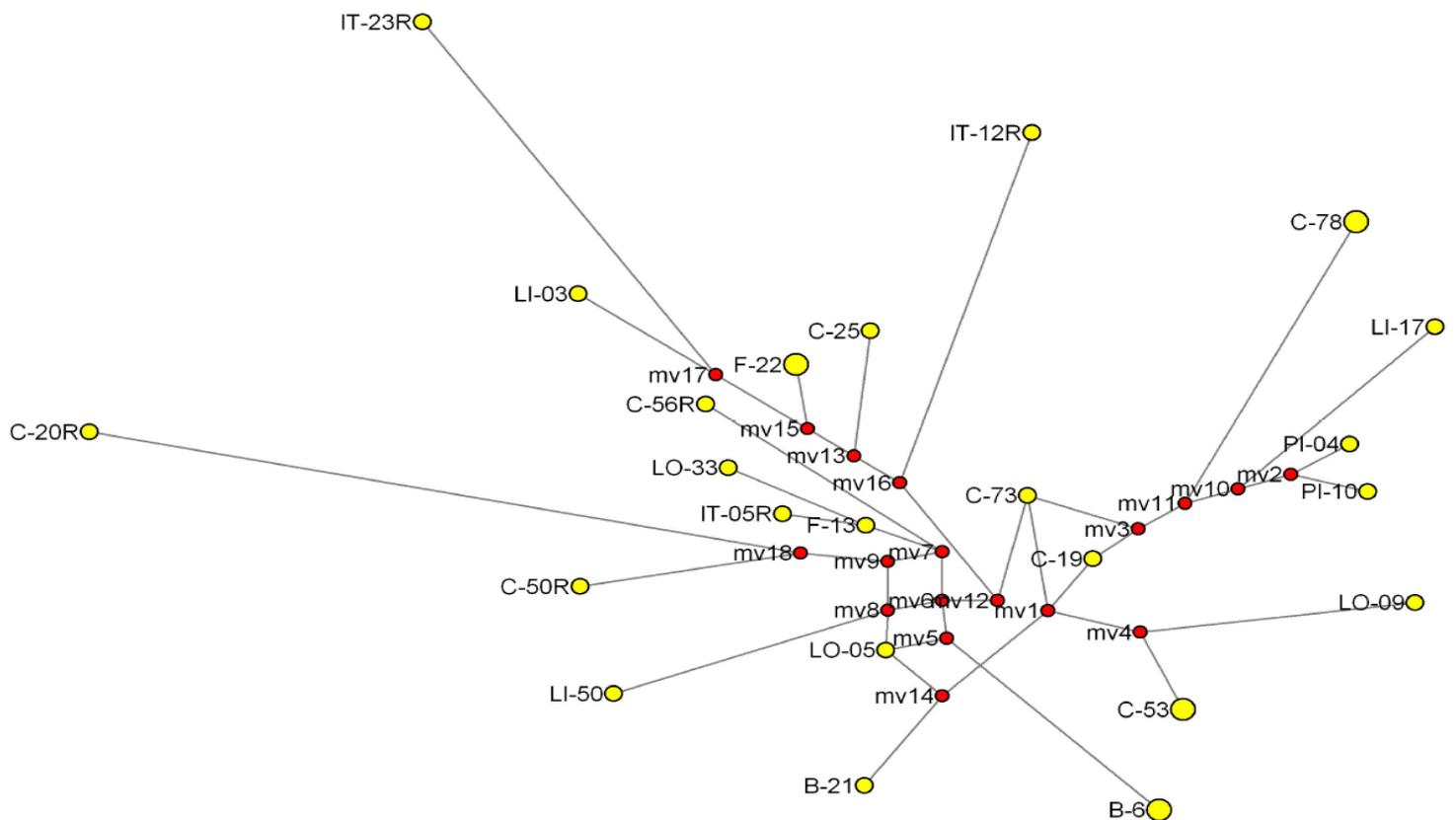


En este network podemos observar que, del ancestro o nudo de origen salen directamente dos muestras una de Liguria (Li-18) y una de España de referencia que es la C-26R. Una serie de mutaciones en ésta última van a dar lugar a otras dos, una de Cataluña (C-102) y otra de Italia de referencia (IT-24R). Otra serie de mutaciones van a dar lugar a (IT-04R).

En general podemos ver que las muestras que más se alejan del resto en esta representación son las muestras de Lombardía Lo-52 y Lo-16, Piamonte Pi-11 y una de Baleares B-11, que parten de

los últimos nudos creados por el programa, lo cual nos indicaría que son las que más mutaciones han sufrido desde el ancestro común.

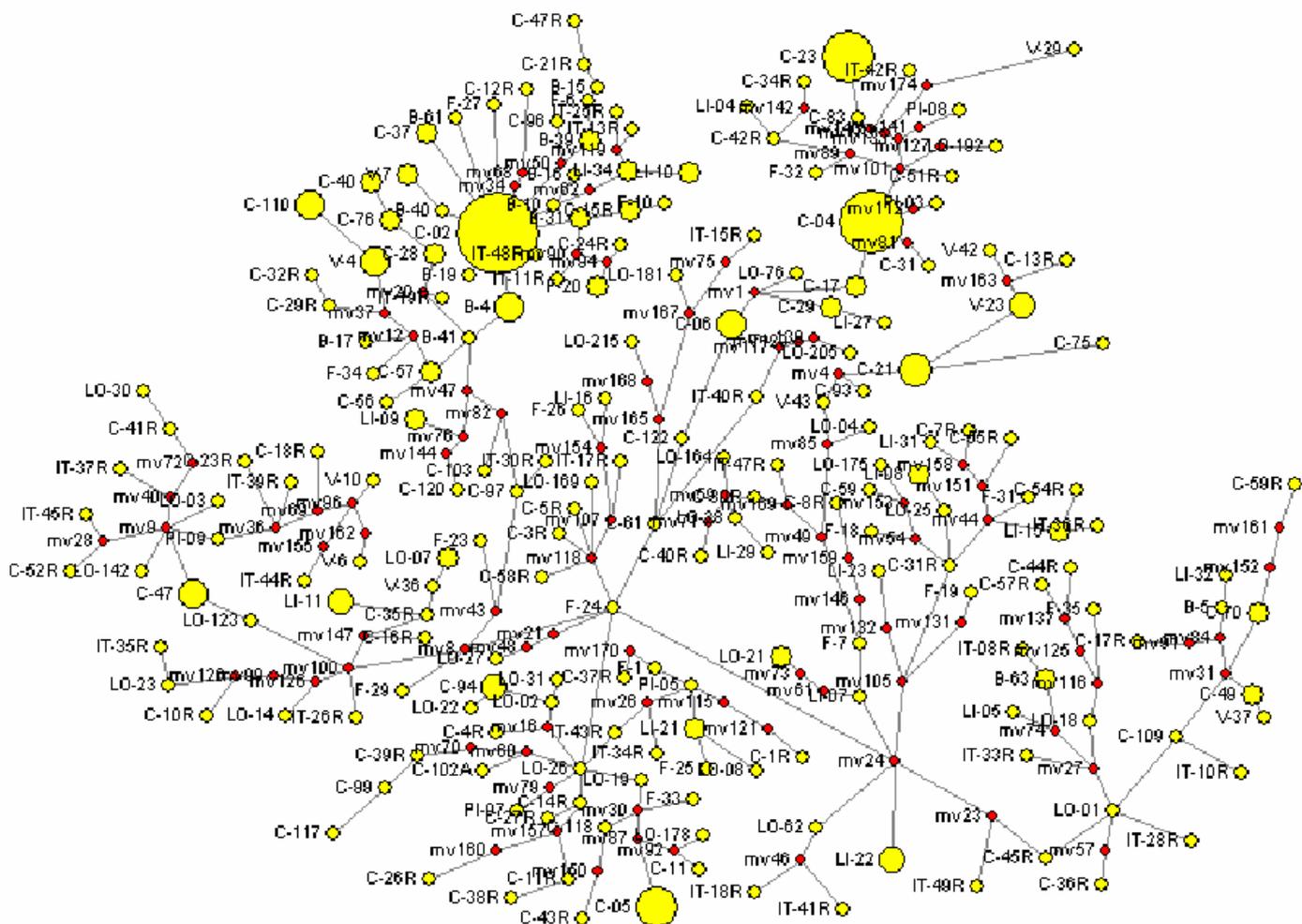
Haplogrupo G2:



En el caso de la representación obtenida con los individuos pertenecientes al haplogrupo G2, observamos que las muestras más distantes al los nudos creados por el programa son las muestras IT-23R, que parte del nudo 17, las C-20R y C-50R, que parten del nudo 18 y la muestra IT-12R que parte del nudo 16, por lo tanto son las más distintas del ancestro común. El primer nudo creado se rodea de las muestras catalanas C-19, C-73 y C-53, que serían las menos diferentes al ancestro común. Tras una serie de mutaciones se originan la C-78. También aparece en el nudo 10 la muestra catalana C-78. El segundo nudo creado agrupa a dos muestras de Piamonte, PI-04 y PI-10. En el mismo ramal que une el nudo 1 y 2 aparece la muestra LI-17. A partir del nudo 7 aparecen muestras de Lombardía, Cataluña y Francia LO-33, C-56 y F-13 ésta última da lugar a la muestra de referencia italiana IT-05R. En el nudo

En cuanto a las francesas se encuentran en un punto algo más intermedio, por un lado F-21 surge del nudo mv4, F-5 lo hace del nudo mv10 y finalmente F-17 surge de mv11.

Haplogrupo R1b:

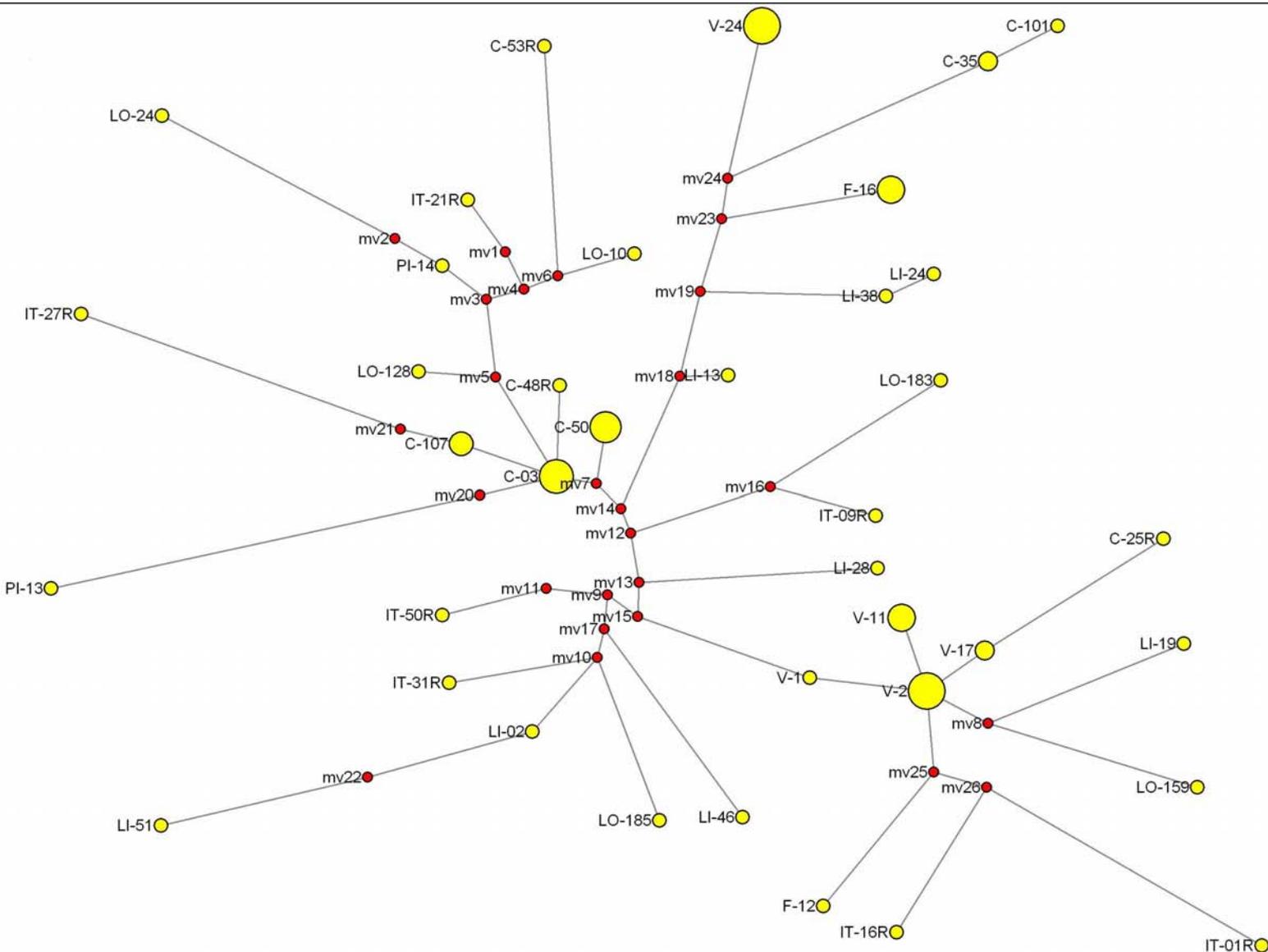


Ésta, que sería la representación del haplogrupo R1b, sería la más compleja de explicar, ya que es la que más individuos contiene. Por un lado, podemos observar tres grandes círculos, en los cuales su tamaño nos indica que contienen bastantes individuos con el mismo haplotipo. El primero C-02 sale del nudo mv34, el segundo C-04 que

sale del nudo mv101 y finalmente C-23 que sale del nudo mv143, las tres muestras son de los individuos apellidados Colón catalanes. Luego, hay otra serie de círculos algo más intermedios, es decir contienen menos individuos que los anteriores que sería C-05 que sale del nudo mv92, C-21 que ha sufrido menos transformaciones desde el nudo de origen ya que sale del nudo mv4, C-47 que sale del nudo mv9, C-05 que sale del nudo mv87. Del nudo de origen salen tres muestras que son tres de cataluña (C-17, C-29 y C-06, ésta última representa a más de un individuo) y una de Lombardía (Lo-76), éstas muestras sería las más parecidas al ancestro común. Las muestras que salen de los nudos más alejados sería V-29 (Valencia) que sale del nudo mv174, dos de Lombardía Lo-215 y Lo-215. Como podemos observar, no parece que dependa mucho el lugar de origen de las muestras para haber sufrido más o menos mutaciones desde el ancestro común.

Lo lógico, según nuestros estudios anteriores, en los que decíamos que según la variabilidad de las muestras, los apellidados Colón franceses y españoles provenían de Italia, sería que las muestras más cercanas al nudo de origen fueran las Italianas, de los nudos intermedios las francesas y de los nudos más lejanos al origen las españolas. Esto nos indicaría que los más parecidos al ancestro serían los Italianos, es decir que el apellido Colón proviene de Italia, pero esto no ocurre así, en los primeros nudos lo mismo hay muestras italianas que españolas. Claro que tampoco podemos afirmar que el apellido Colón tenga el mismo origen en los tres países.

Haplogrupo J1 y J2:



En este esquema podemos observar que, de los primeros nudos salen las muestras italianas IT-21R (italiana de referencia), Lo-24, Lo-10 y Lo-128 (Lombardía), Pi-14 (Piamonte) y C-53R (española de referencia). De los nudos intermedios salen algunas muestras como C-50 y C-03 (Cataluña) que representan a varios individuos, V-1, V-2, V-11 y V-17 (Valencia), IT-31R, IT-50R (italianas de referencia), Li-02 y Li-46 (Liguria) y Lo-185 (Lombardía). De los últimos nudos salen las muestras V-24 y V-2 (Valencia), que representan a más de un individuo, IT-01R y IT-16R (italianas de referencia), F-12 (Francia), C-35 y C-101 (Cataluña) y Li-51 (Liguria).

5.2.3.- COMPARACIÓN DE NUESTRAS POBLACIONES CON OTRAS.

Debido a que nuestras poblaciones salían bastante parecidas entre sí (no hay grandes diferencias), obtuvimos los datos haplotípicos de distintas poblaciones a las nuestras (cerca y muy alejadas geográficamente) y realizamos un análisis de correspondencias, para poder ver si existían mayores diferencias, ya que las muestras que nosotros analizamos en este trabajo están relativamente cercanas geográficamente, y, su parecido podría ser debido a este hecho. Las poblaciones, concretamente, fueron: Austria, Barcelona, Norte de Italia, Polonia, Portugal, Norte de Portugal, Bereberes, Bubi, Chinos, Brasil, Fangs, Japón, India, Malay, Tibet-Lassa, Tibet-China y Tibet-Quinghai.

De este análisis obtuvimos que, algunos haplotipos coincidían entre las distintas poblaciones, al igual, que ocurrió con las nuestras. Concretamente:

- ✚ 3 brasileños igual que 3 españoles.
- ✚ 1 español igual que 1 del norte de Portugal.
- ✚ 1 francés igual 1 brasileño.
- ✚ 2 de Lombardía igual a 2 del norte de Portugal.
- ✚ 2 bubi igual que 2 fangs.
- ✚ 1 austriaco igual que 1 brasileño.
- ✚ 1 austriaco igual que 1 de Polonia.
- ✚ 1 austriaco igual que 1 italiano.
- ✚ 6 brasileños igual a 6 norte de Portugal.
- ✚ 1 brasileño igual a 1 de Polonia.
- ✚ 1 brasileño igual a 1 italiano.
- ✚ 3 de Barcelona igual a 3 portugueses.
- ✚ 1 de Malay igual a 1 chino.
- ✚ 1 chino igual a 1 de India.
- ✚ 1 de Tibet-Lassa igual a 1 de Tibet-Quinghai.
- ✚ 3 de Tibet-Quinghai igual a 3 de Tibet-China.

Posteriormente, cuando hicimos la representación gráfica del análisis de las correspondencias con respecto a los haplotipos (igual que hicimos anteriormente con nuestras poblaciones), obtuvimos un gráfico donde las poblaciones salen también excesivamente parecidas y por tanto la representación gráfica no es nada clara, por ese motivo no se ha incluido en este trabajo dicha gráfica.

Esto nos indica que el cromosoma Y, no es útil para estudios antropológicos pero tiene una gran utilidad en el campo forense, sobretodo como se explicó en anteriores apartados de este trabajo,

para estudios de criminalística como en casos de abusos sexuales donde el porcentaje de células femeninas es muy superior al de células masculinas. También es muy útil en estudios de paternidad donde el padre ha fallecido y podemos disponer de tíos o abuelos paternos.

5.3.-EFICACIA DE LOS SISTEMAS Y SU UTILIDAD MÉDICO-LEGAL.

La utilidad de los polimorfismos de cromosoma Y en las ciencias forenses está refrendada por múltiples trabajos (Roewer *et al.*, 1996; Cooper *et al.*, 1996; Jobling *et al.*, 1997; Prinz *et al.*, 1997; Kayser *et al.*, 1997; de Knijff *et al.*, 1997, etc...). La utilización de ladders alélicos (Roewer *et al.*, 1996; Heyer *et al.*, 1997; Kayser *et al.*, 1997; de Knijff *et al.*, 1997) y estudios de colaboración, como los llevados a cabo por Kayser *et al.*, 1997 y de Knijff (1997), han contribuido en gran medida a potenciar su introducción en el mundo forense.

Los diecisiete STRs descritos en nuestro estudio conforman haplotipos de cromosoma Y informativos desde el punto de vista forense, con valores de diversidad genética de 0.9945. Además, los sistemas que hemos puesto a punto podemos considerarlos específicamente masculinos, ya que tras amplificar ADN de algunas mujeres no se ha detectado producto amplificado.

En nuestra experiencia hemos obtenido excelentes resultados en casos forenses críticos, especialmente para la identificación de ADN masculino en mezclas varón/mujer (e incluso varón/varón) muy degradadas en casos de violación, incluso múltiple, y en el análisis de paternidad. Cuando se parte de mezclas de material biológico masculino y femenino la detección de microsatélites de cromosoma Y proporciona mayor sensibilidad que la de microsatélites autosómicos, al aumentar la capacidad de identificación del componente masculino, aunque exista una gran desproporción entre ambos (Prinz *et al.*, 1997).

Los STRs de cromosoma Y que hemos estudiado pueden constituir un método de screening rápido y fiable, ya que permite la exclusión de sospechosos, pero no pueden ser utilizados como criterio de inclusión en casos de criminalística: siempre hay que tener presente que todos los familiares varones por línea paterna comparten el mismo haplotipo. Esto constituye una importante limitación, tanto en casos de paternidad o violación con implicados pertenecientes a la misma familia como cuando se trata, por ejemplo, de identificar a familiares varones víctimas de catástrofes.

De lo expuesto anteriormente podemos inferir que, polimorfismos informativos de cromosoma Y, conformados como haplotipos, pueden ser perfectamente válidos en algunos casos de rutina forense (obviamente en aquellos casos con implicación de uno o más varones), pero no podemos hacer una identificación individualizada basada exclusivamente en ellos, por lo que de momento sigue siendo conveniente combinar STRs de cromosoma Y con marcadores autosómicos para proporcionar a ambos más opciones útiles en estudios forenses y pruebas de paternidad.



Reyes Católicos
AUTOR: José Martínez Rioboó

Fundación Rodríguez-Acosta
Donación Martínez Sóna.

5.-CONCLUSIONES

PRIMERA: La electroforesis capilar con fluorescencia inducida por láser en equipos automatizados, así como la amplificación mediante kit multiplex permiten un análisis eficiente y rápido de las muestras, minimizando el riesgo de contaminación.

SEGUNDA: Tras el análisis estadístico de los datos obtenidos se encuentra que en una de las poblaciones se produce, lo que en antropología se denomina el "efecto isla". Esta población es la de Baleares, donde debido a las barreras geográficas, la diversidad de la población para el cromosoma Y disminuye considerablemente.

TERCERA: De todas la poblaciones estudiadas, las que cuentan con una mayor diversidad, para los marcadores estudiados de cromosoma Y, son las pertenecientes a Italia (Liguria, Lombardía y Piamonte).

CUARTA: De todas las poblaciones estudiadas, las que cuentan con menor diversidad, para los marcadores de cromosoma Y estudiados, son las pertenecientes a España (Baleares, Valencia y Cataluña).

QUINTA: Los haplogrupos encontrados en las poblaciones estudiadas, son similares a los de otras poblaciones Europeas analizadas.

SEXTA: El estudio estadístico de las muestras estudiadas, junto con las muestras encontradas en la bibliografía, nos indica que los marcadores estudiados de cromosoma Y, no se pueden utilizar para estudios antropológicos.

SEPTIMA: Se puede afirmar que estos marcadores son muy útiles para el análisis de indicios forenses, especialmente donde existe una mezcla de dos o más muestras masculinas junto con células femeninas, aunque dichas mezclas se encuentren en unas concentraciones desproporcionadas.

OCTAVA: Los marcadores estudiados son, también muy útiles para estudios de linajes paternos, tales como estudios de paternidad donde no disponemos de muestras del padre.

NOVENA: Habida cuenta de la importancia que en Genética Forense supone la búsqueda de nuevos polimorfismos de ADN, cuyo análisis sea rápido y sencillo, podemos afirmar que los loci analizados en este trabajo de tesis cumplen todos los requisitos necesarios para su aplicación en identificación genética humana.

Las técnicas utilizadas para el análisis de los mismos, son el reflejo de la adaptación de la Genética Forense al desarrollo y evolución de la Biología Molecular, desarrollo que está haciendo posible el estudio conjunto de múltiples loci tipo microsatélite de una manera sencilla y rápida, disminuyendo los errores de interpretación de los resultados y permitiendo alcanzar valores de probabilidad, con una fiabilidad cada vez más cercana al 100%.



Incendio de la Iglesia de la Virgen de las Angustias (noche 1916)

Fundación Rodríguez-Acosta Donación Martínez Sóla.



Día despues del incendio

Fundación Rodríguez-Acosta Donación Martínez Sóla.

6.-BIBLIOGRAFÍA

1. **Adams D.E., Presley L.A., Baumstark A.L., Hensley K.W., Hill A.L., Anoe K.S.** 1991. Deoxyribonucleic acid (DNA) analysis by restriction fragment length polymorphisms of blood and other body fluids stains subjected to contamination and environmental insults. *J. Forensic Sci.* **36**: 1284-1298.
2. **Allshire R.C., Dempster M., Hastie N.D.** 1989. Human telomeres contain at least three types of G-rich repeat distributed non-randomly. *Nucleic Acids Res.* **17**(12):4611-27.
3. **Alves C., Gusmão L., Barbosa J., Amorim.** 2003. Evaluating the informative power of Y-STRs: a comparative study using European and new African haplotype data. *Forensic Sci. Int.* **134**: 126-133.
4. **Alves C., Gomes V., Prata M.J., Amorim A., Gusmão L.** 2007. Population data for Y-chromosome haplotypes defined by 17 STRs (AmpFISTR Yfiler) in Portugal. *Forensic Sci. Int.* **171**: 250-255.
5. **Amorim A., Pereira L.** 2005. Pros and cons in the use of SNPs in forensic kinship investigation: A comparative analysis with STRs. *Forensic Sci. Int.* **150**: 17-21.
6. **Anderson S., Bankier A.T., Barrell B.G., deBruijin M.H.L., Coulson A.R., Drouin J., Eperon I.C., Nierlich D.P., Roe B.A., Sanger F., Schreier P.H., Smith A.H.J., Staden R., Young I.G.** 1981. Sequence and organisation of the human mitochondrial genome. *Nature.* **290**: 457-465.
7. **Andreassen R., Olaisen B.** 1997. Length and sequence variation in D7S22 (g3) alleles studied by high resolution length measurements and nucleotide sequencing. *Electrophoresis.* **18**(5):675-81.
8. **Anker R., Steinbrueck T., Donniss-Keller H.** 1992. Tetranucleotide repeat polymorphism at the human thyroid peroxidase (hTPO) locus. *Hum. Mol. Genetic.* **1**: 137-140.
9. **Armour J.A., Harris P.C., Jeffreys A.J.** 1993. Allelic diversity at minisatellite MS205 (D16S309): evidence for polarized variability. *Hum. Mol. Genetics.* **2**(8):1137-45.

10. **Arroyo-Pardo E., Gusmão L., López-Parra A.M., Baeza C., Mesa M.S., Amorin A.** 2005. Genetic variability of 16 Y-chromosome STRs in a sample from Equatorial Guinea (Central Africa). *Forensic Sci. Int.* **149**: 109-113.
11. **Barrot C., Sánchez C., Xifró A., Ortega M., Mas J., Huguet E., Corbella J., Gené M.** 2007. Data for Y-chromosome haplotypes in Fang and Bubi populations from Bioko (Equatorial Guinea). *Forensic Sci. Int.* **168**: e10-e12.
12. **Bassam J.S., Caetano-Anollés G., Gresshoff P.M.** 1991. Fast and sensitive silver staining of DNA in polyacrylamide gels. *Anal. Biochem.* **196**: 80-83.
13. **Beckman J.S., Weber J.L.** 1992. Survey of human and rat microsatellites. *Genomics.* **12**: 627-631.
14. **Bell G., Selby M., Rutter W.** 1982. The highly polymorphic region near the human insulin gene is composed of simple tandemly repeating sequences. *Nature.* **295**(5844):31-35.
15. **Bellas S., Soto J.L., Carracedo A., Rodríguez-Calvo M.S.** 1997. Minisatellite variant repeat coding using semiautomatic system. *BioTechniques.* **23**:44-46.
16. **Berger B., Lindinger A., Niederstätter H., Grubwieser P., Parson W.** 2006. Y-STR typing of an Austrian population sample using a 17-locus multiplex PCR assay. *Int. J. Legal Med.* **119**: 241-246.
17. **Biondo R., Caglià A., Asili P., D'Agostaro G., D'Agostino G., Spinella A.** 2004. Typing of 20 Y-chromosome STRs in the Italian population. *Forensic Sci. Int.* **146S**: S135-S138.
18. **Bosch E., Rosser Z.H., Norby S., Lynnerup N., Jobling M.A.** 2003. Y-chromosomal STR haplotypes in Inuit and Danish population samples. *Forensic Sci. Int.* **132**: 228-232.

19. **Brinkmann B.** 1992. The use of STRs in stain analysis. In: Proceedings from the Third International Symposium of Human Identification, Promega Corporation, Madison, U.S.A., pp: 357-374.
20. **Brinkmann B., Sajantila A., Goedde H.W., Matsumoto H., Nishi K., Wiegand P.** 1996. Population genetic comparisons among eight populations using allele frequency and sequence data from three microsatellite loci. *Eur. J. Hum. Genet.* **4**: 175-182.
21. **Brinkmann B., Klitschar M., Neuhuber F., Hunh J., Rolf B.** 1998. Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. *Am. J. Hum. Genet.* **62**: 1408-1415.
22. **Budowle B., Chakraborty R., Giusti A.M., Eisenberg A.J., Allen R.C.** 1991. Analysis of the VNTR locus D1S80 by the PCR followed by high-resolution PAGE. *Am. J. Hum. Genet.* **48**(1):137-44.
23. **Budowle B., Moretti T., Keys M.** 1997. Validation studies of the CTT STR Multiplex System. *J. Forensic Sci.* **42**(2):701-707.
24. **Builes J.J., Martínez B., Gómez A., Caraballo L., Espinal C., Aguirre D., Montoya A., Moreno M., Amorim A., Gusmão L., Bravo M.L.** 2007. Y chromosome STR haplotypes in the Caribbean city of Cartagena (Colombia). *Forensic Sci. Int.* **167**: 62-69.
25. **Butler J.M., Shen Y., McCord B.R.** 2003. The development of reduced size STR amplicons as tools for analysis of degraded DNA. *J. Foren. Sci.* **48**: 1054-1064.
26. **Capon D.J., Seeburg P.H., McGrath J.P., Hayflick J.S., Edman U., Levinson A.D., Goeddel D.V.** 1983. Activation of Ki-ras2 gene in human colon and lung carcinomas by two different point mutations. *Nature.* **304**(5926):507-13.
27. **Chakraborty R.** 1991. Statistical interpretation of DNA typing data [letter; comment]. *Am. J. Hum. Genet.* **49**(4):895-7, 899-903.

28. **Chakraborty R., Kidd K.K.** 1991. The utility of DNA typing in forensic work .Sci. **254**(5039):1735-9.
29. **Chakraborty R., Stivers D. N., Su B., Zhong Y., Budowle B.** 1999. The utility of short tandem repeat loci beyond human identification: implications for development of new DNA typing systems. Electrophoresis **20**: 1682-1696.
30. **Chang Y.M., Perumal R., Keat P.Y., Kuehn D.L.C.** 2007. Haplotype diversity of 16 Y-chromosomal STRs in three main ethnic populations (Malays, Chinese and Indians) in Malaysia. Forensic Sci. Int. **167**: 70-76.
31. **Cooper G., Amos W., Hoffman D., Rubinsztein D.C.** 1996. Network analysis of human Y microsatellite haplotypes. Hum. Mol. Genet. **5**: 1759-1766.
32. **Coyne V.E., James M.D., Reid S.J., Rybicki E.P.** 1996 Molecular Biology Techniques Manual. Third Edition. Ed. Rybicki.
33. **Deka R., Chakraborty R., De Croo S., Rothhammer F., Barton S.A., Ferrell R.E.** 1992a. Characteristics of polymorphism at VNTR locus 3' to the apolipoprotein B gene in five human populations. Am. J. Hum. Genet. **51**: 1325-1333.
34. **Deka R., De Croo S., Yu L.M., Ferrell R.E.** 1992b. Variable number of tandem repeat (VNTR) polymorphism at locus D17S5 (YNZ22) in four ethnically defined human populations. Hum. Genet. **90**: 86-90.
35. **D'Esposito M.A., Ciccodicola A., Gianfrancesco F., Esposito T., Flagiello L., Mazzarella R., Schlessinger D., D'Urso M.** 1996. A synaptobrevin-like gene in the Xq28 pseudoautosomal region undergoes X-inactivation. Nature Genet. **13**: 227-229.
36. **Dupuy B.M., Stenersen M., Egeland T., Olaisen B.** 2004. Y-chromosomal microsatellite mutation rates: differences in mutation rate between and within loci. Hum. Mutat. **23**: 117-124.

37. **Economou E.P., Bergen A.W., Warren A.C., Antonarakis S.E.** 1990. The polydeoxyadenylate tract of Alu repetitive elements is polymorphic in the human genome. *Proc. Natl. Acad. Sci. U.S.A.* **79**: 6465-6469.
38. **Edwards A., Civitello A., Hammond H., Caskey C.** 1991. DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *Am. J. Hum. Genet.* **49**(4):746-56.
39. **Edwards A., Hammond H., Jin L., Caskey C., Chakraborty R.** 1992. Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. *Genomics.* **12**(2):241-53.
40. **Ellis N., Goodfellow P.N.** 1989a. The mammalian pseudoautosomal region. *Trends Genet.* **5**: 406-410.
41. **Ellison J., Passage M., Yu L.C., Yen P., Mohandas T.K., Shapiro L.** 1992. Directed isolation of human genes that escape X inactivation. *Somat. Cell Mol. Genet.* **18**: 256-268.
42. **Epstein N., Nahor O., Silver J.** 1990. The 3' ends of alu repeats are highly polymorphic. *Nucleic Acids Res.* **18**(15):4634.
43. **Fisher R.A.** 1951. Standard calculations for evaluating a blood-group systems. *Heredity.* **5**: 95-102.
44. **Fisher E.M.C., Beer-Romero P., Brown L.G., Ridley A., McNeil J.A., Lawrence J.B., Willard H.F., Bierber F.R., Page D.C.** 1990. Homologous ribosomal protein genes on the human X and Y chromosomes: escape from inactivation and possible implications for Turner syndrome. *Cell.* **63**: 1205-1218.
45. **Freije D., Helms C., Watson M.S., Donis-Keller.** 1992. Identification of a second pseudoautosomal region near the Xq and Yq telomeres. *Sci.* **258**: 1784-1787.
46. **Frigi S., Pereira F., Pereira L., Yacoubi B., Gusmão L., Alves C., el Khil H.K., Cherni L., Amorin A., el Gaaied A.** 2006. Data for Y-

chromosome haplotypes defined by 17 STRs (AmpFISTR Yfiler) in two Tunisian Berber communities. *Forensic Sci. Int.* **160**: 80-83.

47. **Gabriel S.B., Schaffner S.F., Nguyen H., Moore J.M., Roy J., Blumenstiel B., Higgins J., DeFelice M., Lochner A., Faggart M., Liu-Cordero S.N., Rotimi C., Adeyemo A., Cooper R., Ward R., Lander E.S., Daly M.J., Altshuler D.** 2002. The structure of haplotype blocks in the human genome. *Sci.* **296**: 2225-2229.
48. **Gill P., Jeffreys A.C., Werrett D.J.** 1985. Forensic application of DNA "fingerprints". *Nature.* **318**: 577-579.
49. **Gill P.** 2001. An assessment of the utility of single nucleotide polymorphism (SNPs) for forensic purposes. *Int. J. Legal. Med.* **114**: 204-210.
50. **Gill P., Werrett D.J., Budowle B., 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**: 51-53
51. **Giusti A., Baird M., Pasquale S., Balazs I., Glassberg J.** 1986. Application of deoxyribonucleic acid (DNA) polymorphisms to the analysis of DNA recovered from sperm. *J. Forensic. Sci.* **31**: 409-417.
52. **Goodfellow P., Banting G., Sheer D., Ropers H.H., Caine A., Ferguson-Smith M.A., Povey S., Voss R.** 1983. Genetic-evidence that a Y-linked gene in man is homologous to a gene on the X-chromosome. *Nature.* **302**: 346-349.
53. **Gough N.M., Gearing D.P., Nicola N.A., Baker E., Pritchard M., Callen D.F., Sutherland G.R.** 1990. Localization of the human GM-CSF receptor gene to the X-Y pseudoautosomal region. *Nature.* **345**: 734-736.

-
54. **Hammond H.A., Jin L., Zhong Y., Caskey C.T., Chakraborty R.** 1994. Evaluation of 13 short tandem repeat loci for use in personal identification applications. *Am. J. Hum. Genet.* **55**: 175-189.
55. **Hara M., Kido A., Takada A., Adachi N., Saito K.** 2007. Genetic data for 16 Y-chromosomal STR loci in Japanese. *Legal Med.* **9**: 161-170.
56. **Hassold T.J., Sherman S.L., Pettay D., Page D.C., Jacobs P.A.** 1991. XY-chromosome non disjunction in man associated with diminished recombination in the pseudoautosomal region. *Am. J. Hum. Genet.* **49**: 253-260.
57. **Heyer E., Puymirat J., Dieltjes P., Bakker E., de Knijff P.** 1997. Estimating Y chromosome specific microsatellite mutation frequencies using deep rooting pedigrees. *Hum. Mol. Genet.* **6**: 799-803.
58. **Hopkins B., Williams N.J., Webb M.B., Debenham P.G., Jeffreys A.J.** 1994. The use of minisatellite variant repeat-polymerase chain reaction (MVR-PCR) to determine the source of saliva on a used postage stamp. *J. Forensic Sci.* **39**(2):526-31.
59. **Huang Q.Y., Xu F.H., Shen H., Deng H.Y., Liu Y.J., Liu Y.Z., Li J.L., Recker R.R., Deng H.W.** 2002. Mutation patterns at dinucleotide microsatellite loci in humans. *Am. J. Hum. Genet.* **70**: 625-634.
60. **Huson D.H.** 1998. Splits Tree: analyzing and visualizing evolutionary data. *Bioinformatics (Oxford, England).* **14**: 68-73.
61. **Jarman A.P., Nicholls R.D., Weatherall D.J., Clegg J.B., Higgs D.R.** 1986. Molecular characterisation of a hypervariable region downstream of the human alpha-globin gene cluster. *Embo Journal.* **5**(8):1857-63.
62. **Jeffreys A., Wilson V., Thein S.** 1985a. Hypervariable 'minisatellite' regions in human DNA. *Nature.* **314**(6006):67-73.
63. **Jeffreys A.J., Wilson V., Thein S.L.** 1985b. Individual-specific 'fingerprints' of human DNA. *Nature.* **316**(6023):76-9.

-
64. **Jeffreys A., MacLeod A., Tamaki K., Neil D., Monckton D.** 1991. Minisatellite repeat coding as a digital approach to DNA typing [see comments]. *Nature*. **354**(6350):204-9.
65. **Jeffreys A.J., Monckton D.G., Tamaki K., Neil D.L., Armour J.A., MacLeod A., Collick A., Allen M., Jobling M.** 1993. Minisatellite variant repeat mapping: application to DNA typing and mutation analysis. *Exs.* **67**:125-39.
66. **Jobling M.A., Tyler-Smith C.** 1995. Fathers and sons: the Y chromosome and human evolution. *Trends Genet.* **11**: 449-456.
67. **Jobling M.A., Pandya A., Tayler-Smith C.** 1997. The Y chromosome in forensic and paternity testing. *Int. J. Legal Med.* **110**: 118-124.
68. **Jones D.A.** 1972. Blood samples: probability of discrimination. *J. Forensic Sci. Soc.* **12**: 355-359.
69. **Kanter E., Baird M., Shaler R., Balazs I.** 1986. Analysis of restriction fragment length polymorphisms in deoxyribonucleic acid (DNA) recovered from dried bloodstains. *J. Forensic. Sci.* **31**: 403-408.
70. **Kao F.T.** 1985. Human genome structure. *International Review of Cytology.* **96**:51-88.
71. **Kayser M., Caglia A., Corach D., Fretwell N., Gehrig C., Graziosi G.** 1997. Evaluation of Y-chromosomal STRs: a multicenter study. *Int. J. Legal Med.* **110**: 125-133.
72. **Kayser M., Rower L., Hedman M.** 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**: 1580-1588.

-
73. **Kayser M., Kittler R., Erler A., Hedman M., Lee A.C., Mohyuddin A.** 2004. A comprehensive survey of human Y-chromosomal microsatellite. *Am. J. Hum. Genet.* **74**: 1183-1197.
74. **Kermouni A., Vanroost E., Arden K.C., Vermeesch J.R., Weiss S., Godelaine D., Flint J., Lurquin C., Szikora J.P., Higgs D.R., Marynen P., Renaud J.C.** 1995. The IL-9 receptor gene (IL-9R) genomic structure, chromosomal localization in the pseudoautosomal region of the long arm of the sex-chromosomes, and identification of IL-9R pseudogenes at 9qter, 10pter, 16pter and 18pter. *Genomics.* **29**: 371-382.
75. **Kimpton C., Walton A., Gill P.** 1992. A further tetranucleotide repeat polymorphism in the vWF gene. *Hum. Mol. Genetics.* **1**(4):287.
76. **Kimpton C., Gill P., Walton A., Urquhart A., Millican E.S. Adams M.** 1993. Automated DNA profiling employing multiplex amplification of short tandem repeat loci. *PCR Methods and Applications.* **3**: 13-22.
77. **Kimpton C., Gill P., D'Aloja E., Andersen J.F., Bar W., Holgersson S.** 1995. Report on the second EDNAP collaborative STR exercise. *Forensic Sci. Int.* **71**: 137-152.
78. **Kimura M.** 1983. The neutral theory of molecular evolution. Cambridge, Cambridge University Press.
79. **de Knijff P., Kayser M., Caglia A., Corach D., Fretwell N., Gehrig C.** 1997. Chromosome Y microsatellite: population genetic and evolutionary aspects. *Int. J. Legal Med.* **110**: 134-140.
80. **Kremer E., Baker E., Dandrea R.J., Slim R., Phillips H., Moretti P., López A.F., Petit C., Vadas M.A., Sutherland G.R., Goodall G.J.** 1993. A cytokine receptor gene-cluster in the X-Y pseudoautosomal region. *Blood.* **82**: 22-28.
81. **Kumagai R., Kumagai A., Saigusa K., Aoki Y.** 2007. Haplotype analysis of 17 Y-STR loci in a Japanese population. *Forensic Sci. Int.* **172**: 72-78.

-
82. **Lahn B., Page D.** 1997. Functional coherence of the human Y chromosome. *Sci.* **278**: 675-680.
83. **Lee H.Y., Park M.J., Chung U., Lee H.Y., Yang W.I., Cho S.H., Shin K.J.** 2007. Haplotypes and mutation analysis of 22 Y-chromosomal STRs in Korean father-son pairs. *Int. J. Legal Med.* **121**: 128-135.
84. **Lessig R., Edelmann J.** 1998. Y chromosome polymorphisms and haplotypes in West Saxonia (Germany). *Int. J. Legal. Med.* **111**: 215-218.
85. **Li K.L., Kai L., Yuming M.** 2007. Y chromosome STR haplotypes of Tibetan Living Tibet Lassa. *Forensic Sci. Int.* **172**: 79-83.
86. **Litt M., Luty J.A.** 1989. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Am. J. Hum. Genet.* **44**: 397-401.
87. **Lo Y., Waincoat J.S., Gillmer G., Patel P., Sampietro M., Fleming K.A.** 1989. Prenatal sex determination by DNA amplification from maternal peripheral blood. *Lancet.* **9**: 1363-1365.
88. **López A.M., Álvarez S., Gusmão L., Alves C., Mesa M.S., Albentosa A., Arribas G., López R., Barrio P.A., Amorin A., Arroyo-Pardo E.** 2004. Population data for 16 Y-chromosome STRs in four populations from Pyrenees (Spain). *Forensic Sci. Int.* **140**: 125-129.
89. **Lorente J.A., Lorente M.** 1995. El ADN y la identificación en la investigación criminal y en la paternidad biológica. Editorial Comares. Granada.
90. **Martín P., García-Hirschfeld J., García O., Gusmão L., García P., Albarrán C., Sancho M., Alonso A.** 2004. A Spanish population study of 17 Y chromosome STR loci. *Forensic Sci. Int.* **139**: 231-235
91. **McNally L., Shaler R.C., Baird M., Balazs I., de Forest P., Kobilinsky L.** 1989a. Evaluation of deoxyribonucleic acid (DNA)

- isolated from human bloodstains exposed to ultraviolet light, heat, humidity and soil contaminatios. *J. Forensic Sci.* **34**: 1059-1069.
92. **McNally L., Shaler R.C., Baird M., Balazs I., Kobilinsky L., de Forest P.** 1989b. The effects of environment and substrates on DNA: the use od casework samples from New York city. *J. Forensic Sci.* **34**: 1070-1077.
93. **Morral N., Estivill X.** 1992. Multiplex PCR amplification of three microsatellites within the CFTR gene. *Genomics.* **13**(4):1362-4.
94. **Morton N.E.** 1991. Parameters of the human genome. *Proc. Natl. Acad. Sci. USA.* **88**: 7474-7476.
95. **Mulero J.J., Chang C.W., Calandro L.M., Green R.L., Li Y., Johnson C.L., Hennessy L.K.** 2006. Development and validation of the AmpFSTR Y-Filer PCR amplification Kit: A Male Specific, single amplification 17 Y-STR multiplex system. *J. Forensic Sci.* **51**: 61-71.
96. **Mullis K., Faloona F., Scharf S., Saiki R., Horn G., Erlich H.** 1986. Specific enzymatic amplification of DNA in vitro: The polimerase chain reaction. *Cold Spring Harber Symp Quant. Biol.* **51**: 263-273.
97. **Mullis K.B.** 1990. Target amplification for DNA analysis by the polymerase chain reaction. *Annales De Biologie Clinique.* **48**(8):579-82.
98. **Mully J., Gedeon A, White S., Haan E., Richards R.** 1991. Predictive diagnosis of myotonic dystrophy with flanking microsatellite markers. *J. Med. Genetics.* **28**:448-452.
99. **Nakahori Y., Takenaka O., Nakagome Y.** 1991. A human X-Y homologous region encodes Amelogenin. *Genomics.* **9**: 364-369.
100. **Nakamura Y., Leppert M., O'Connell P., Wolf R., Holm T., Culver M.** 1987. Variable number of tandem repeat (VNTR) markers for human gene mapping. *Sci.* **235**: 1616-1622.

101. **Neil D.L., Jeffreys A.J.** 1993. Digital DNA typing at a second hypervariable locus by minisatellite variant repeat mapping. *Hum. Mol. Genetics*. **2**(8):1129-35.
102. **Nishimura D.Y., Murray J.C.** 1992. A tetranucleotide repeat for the F13B locus. *Nucleic Acids Res.* **20**: 1167.
103. **Ohno S.** 1984. Repeats of base oligomers as the primordial coding sequences of the primeval earth and their vestiges in modern genes. *J. Mol. Evolution*. **20**(3-4):313-21.
104. **Oudet C., Heilig R., Hanauer A., Mandel J.L.** 1991. Nonradioactive assay for new microsatellite polymorphisms at the 5' end of the dystrophin gene, and estimation of intragenic recombination. *Am. J. Hum. Genetics*. **49**(2):311-9.
105. **Owerbach D., Aagaard L.** 1984. Analysis of a 1963-bp polymorphic region flanking the human insulin gene. *Gene*. **32**(3):475-9.
106. **Page D.C., Mosher R., Simpson E.M., Fisher C., Mardon G., Pollack J., McGillivray B., de la Chapelle A., Brown L.G.** 1987. The sex-determining region of the human Y chromosome encodes a finger protein. *Cell*. **51**: 1091-1104.
107. **Park M.J., Lee H.Y., Yoo J.E., Chung U., Lee S.Y., Shin K.J.** 2005. Forensic evaluation and haplotypes of 19 Y-chromosomal STR loci in Koreans. *Forensic Sci. Int.* **152**: 133-147.
108. **Pereira R.W., Monteiro E.H.G., Hirschfeld G.C.R., Wang A.Y., Grattapaglia D.** 2007. Haplotype diversity of 17 Y-chromosome STRs in Brazilians. *Forensic Sci. Int.* **171**: 226-236.
109. **Perkin Elmer Corporation.** 1996. GeneScan Analysis Software, P/N 904435.
110. **Perkin Elmer Corporation.** 1996. Genotyper. Applications Tutorials, P/N 904649.

-
111. **Perkin Elmer Corporation.** 1997. GeneScan Reference Guide, P/N 4303189 rev A.
112. **Perkin Elmer Corporation.** 1998. ABI PRISM™310 Genetic Analyzer User's Manual, P/N 903565 rev B.
113. **Polymeropoulos M.H., Xlao H., Rath D.S., Merrill C.R.** 1991a. Tetranucleotide repeat polymorphism at the human tyrosine hidrolase gene (TH) *Nucleic Acids Res.* **19**: 3753.
114. **Polymeropoulos M.H., Rath D.S., Xlao H., Merrill C.R.** 1991b. Tetranucleotide repeat polymorphism at the c-fes/fps proto-oncogen (FES). *Nucleic Acids Res.* **19**: 4018.
115. **Polymeropoulos M.H., Rath D.S., Xlao H., Merrill C.R.** 1991c. Tetranucleotide repeat polymorphism at the human coagulation factor XIII A subunit gene (F13A1). *Nucleic Acids Res.* **19**: 4036.
116. **Pontes M.L., Cainé L., Abrantes D., Lima G., Pinheiro M.F.** 2007. Allele frequencies and population data for 17 Y-STR loci (AmpI STR Y-Filer) in a Northern Portuguese population sample. *Forensic Sci. Int.* **170**: 62-67.
117. **Prinz M., Boll K., Baun H., Schaler B.** 1997. Multiplexing of Y chromosome specific STRs and performance for mixed samples. *Forensic Sci. Int.* (en prensa).
118. **Proudfoot N.J., Gil A., Maniatis T.** 1982. The structure of the human zeta-globin gene and a closely linked, nearly identical pseudogene. *Cell.* **31**(3 Pt 2):553-63.
119. **Rand S., Puers C., Skowasch K., Wiegand P., Budowle B., Brinkmann B.** 1992. Population genetics and forensic efficiency data of 4 AMPFLP's. *Int. J. Legal Med.* **104**(6):329-33.

120. Recommendations of the DNA Commission of the International Society for Forensic Haemogenetics relating to the use of PCR-based polymorphisms. *Forensic Sci. Int.* 1992. **55**: 1-3
121. **Reich D.E., Schaffner S.F., Daly M.J. McVean G., Mullikin J.C., Higgins J.M., Richter D.J., Lander E.S., Altshuler D.** 2002. Human genome sequence variation and the influence of gene history, mutation and recombination. *Nat. Genet.* **32**: 135-140.
122. **Richards R.I., Sutherland G.R.** 1992. Heritable unstable DNA sequences [news]. *Nature Genetics.* **1**(1):7-9.
123. **Richards M.B., Macaulay V.A., Bandelt H.J., Sykes B.C.** 1998. Phylogeography of mitochondrial DNA in western Europe. *An. Human Genet.* **62**: 241-260.
124. **Rodríguez-Alarcón J., Martínez de Pancorbo M., Santillana L., Castro A., Melchor J.C., Linares A., Fernández-Llebrez L., Aranguren G.** 1996. La "huella de ADN" en lugar de la "huella plantar" en la identificación neonatal. *Medicina Clínica* (en prensa).
125. **Rodríguez-Calvo M.S., Bellas S., Soto J.L., Barros F., Carracedo A.** 1996. Comparison of different electrophoretic methods for digital typing of the MS32 (D1S8) locus. *Electrophoresis.* **17**(8):1294-8.
126. **Roewer L., Kayser M., Dieltjes P., Nagy M., Bakker E., Krawezak M.** 1996. Analysis of molecular variance (AMOVA) of Y-chromosome specific microsatellites in two closely related human populations. *Hum. Mol. Genet.* **5**: 1029-1033.
127. **Roewer L.** 1998. The Y chromosome. Forensic application and evolutionary aspects, in: B. Olaisen, B. Brinkmann, P.J. Lincoln (Eds.), *Progress in Forensic Genetics*, Vol. **7**, Elsevier.
128. **Rogers A.R., Harpending H.** 1992. Population growth makes waves in the distribution of pairwise genetic differences. *Mol. Biol. Evol.* **9**: 552-569.

129. **Royle N.J., Clarkson R.E., Wong Z., Jeffreys A.J.** 1988. Clustering of hypervariable minisatellites in the proterminal regions of human autosomes. *Genomics*. **3**: 352-360.
130. **Saitou N., Nei M.** 1987. The neighbor-joining methods: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**: 406-425.
131. **Sambrook J., Fritsch E., Maniatics T.** 1989. *Molecular Cloning. A Laboratory Manual.*, Second ed, vol. I, II, III. Cold Spring Harbor Laboratory Press, USA.
132. **Sanchez J.J., Borsting C., Hallenberg C., Buchard A., Hernandez A., Morling N.** 2003: Multiplex PCR and minisequencing of SNPs - a model with 35 Y chromosome SNPs. *Foren. Sci. Int.* **137**: 74-84.
133. **Sánchez C., Barrot C., Xifró A., Ortega M., Gómez de Aranda I., Huguet E., Corbella J., Gené M.** 2007. Haplotype frequencies of 16 Y-chromosome STR loci in the Barcelona metropolitan area population using Y-Filer kit. *Forensic Sci. Int.* **172**: 211-217.
134. **Schneider S., Excoffier L.** 1999. Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA. *Gene*. **152**: 1079-1089.
135. **Scheinfeldt L., Friedlaender F., Friedlaender J., Latham K., Koki G., Karafet T., Hammer M., Lorenz J.** 2006. Unexpected NR Y-chromosome variation in Northern Island Melanesia. *Mol. Biol. Evol.* **23 (8)**: 1628-1641.
136. **Schwartz L.S., Tarleton J., Popovich B., Seltzer W.K., Hoffman E.P.** 1992. Fluorescent multiplex linkage analysis and carrier detection for Duchenne/Becker muscular dystrophy. *Am. J. Hum. Genet.* **51**: 721-729.
137. **Sinclair A.H., Berta P., Palmer M.S., Hawkins J.R., Griffiths B., Smith M.J., Foster J.W., Frischauf A.M., Lovell-Bradge R., Goodfellow P.N.** 1990. A gene from the human sex-determining

- region encodes a protein with homology to a conserved DNA-binding motif. *Nature*. **346**: 240-244.
138. **Singer M.** 1982a. Highly repeated sequences en mammalian genomes. *Int Rev Cytol*. **76**:67-112.
139. **Soltyszewki I., Pepinski W., Spolnicka M., Kartasinska E., Konarzewska M., Janica J.** 2007. Y-chromosomal haplotypes for the AmplSTR Yfiler PCR amplification kit in a population sample from Central Poland. *Forensic Sci. Int.* **168**: 61-67
140. **Souza Góes A.C., Carvalho E.F., Gomes I., da Silva D.A., Gil E.H.F., Amorim A., Gusmão L.** 2005. Population and mutation analysis of 17 Y-STR loci from Rio de Janeiro (Brazil). *Int. J. Legal Med.* **119**: 70-76.
141. **Sullivan K.M., Pope S., Gill P., Robertson J.M.** 1992. Automated DNA profiling by fluorescent labelling of PCR products. *PCR Methods Applic.* **2**: 34-40.
142. **Suzumori K., Adachi R., Okada S., Narukawa T., Yagami Y., Sonta S.** 1992. Fetal cells in the maternal circulation: Detection of Y-Sequence by gene amplification. *Obstetrics & Gynecology*. **80**:150-154.
143. **Tamaki K., Monckton D.G., MacLeod A., Neil D.L., Allen M., Jeffreys A.J.** 1992. Minisatellite variant repeat (MVR) mapping: analysis of 'null' repeat units at D1S8 [published erratum appears in *Hum Mol Genet* 1992 Oct;1(7):558]. *Hum. Mol. Genetics*. **1**(6):401-6.
144. **Tamaki K., Huang X.L., Yamamoto T., Uchihi R., Nozawa H., Katsumata Y.** 1995. Applications of minisatellite variant repeat (MVR) mapping for maternal identification from remains of an infant and placenta. *J. Forensic Sci.* **40**(4):695-700.
145. **Turrina S., Atzei R., de Leo D.** 2006. Y-chromosomal STR haplotypes in a Northeast Italian population sample using 17plex loci PCR assay. *Int. J. Legal Med.* **120**: 56-59.

146. **Ulrich A., Dull T.J., Gray A., Philips J.A. III, Peter S.** 1982. Variation in the sequence and modification state of the human insulin gene flanking regions. *Nucleic Acids Res.* **10**: 2225-2240.
147. **Underhill P.A., Jin L., Zemmans R., Oefner P.J., Cavalli-Sforza L.L.** 1996. A precolumbian Y chromosome-specific transition and its implications for human evolutionary history. *Proc. Natl. Acad. Sci. USA.* **93**: 196-200.
148. **Urquhart A., Kimpton C., Downes T., Gill P.** 1994. Variation in short tandem repeat sequences--a survey of twelve microsatellite loci for use as forensic identification markers. *Int J Legal Med.* **107**(1):13-20.
149. **Valverde E., Cabrero C., Cao R., Rodríguez-Calvo M.S., Díez A., Barros F., Alemany J., Carracedo A.** 1993. Population genetics of three VNTR polymorphisms in two different Spanish populations. *Int. J. Legal. Med.* **105**: 251-256.
150. **Vosberg H. P.** 1989. The polymerase chain reaction: an improved method for the analysis of nucleic acids. *Hum. Genetics.* **83**(1):1-15.
151. **Walsh P., Metzger D., Higuchi R.** 1991. Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *Biotechniques.* **10**(4):506-13.
152. **Watchel S.S., Ohno S., Koo G.C., Boyse E.A.** 1975. Possible role of H-Y antigen in the primary determination of sex. *Nature.* **257**: 235-236.
153. **Weber J.L.** 1990. Informativeness of human (dC-dA)_n.(dG-dT)_n polymorphisms. *Genomics.* **7**(4):524-30.
154. **Weber J., May P.** 1989. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am. J. Hum. Genetic.* **44**:338-396.

-
155. **Weber B., Riess O., Wolff G., Andrew S., Collins C., Graham R., Theilmann J., Hayden M.R.** 1992. Delineation of a 50 kilobase DNA segment containing the recombination site in a sporadic case of Huntington's disease. *Nature Genetics*. **2**(3):216-22.
156. **Weber J.L., Wong C.** 1993. Mutation of human short tandem repeats. *Hum. Mol. Genet.* **2**: 1123-1128.
157. **Wiegand P., Bajanowski T., Brinkmann B.** 1993a. DNA typing of debris from fingernails. *Int. J. Leg. Med.* **106**: 81-83.
158. **Wiegand P., Budowle B., Rand S., Brinkmann B.** 1993b. Forensic validation of the STR systems SE 33 and TC11. *Int. J. Leg. Med.* **105**: 315-320.
159. **Willard H.F., Waje J.S.** 1987. Chromosome-specific subsets of human alpha satellite DNA: analysis of sequence divergence within and between chromosomal subsets and evidence for an ancestral pentameric repeat. *J. Mol. Evolution.* **25**(3):207-14.
160. **Wolf U., Schempp W., Scherer G.** 1992. Molecular biology of the human Y chromosome. *Reviews in Physiology, Biochemistry and Pharmacology.* **121**: 147-213.
161. **Wong Z., Wilson V., Patel I., Povey S., Jeffreys A.J.** 1987. Characterization of a panel of highly variable minisatellites cloned from human DNA. *Annals Of Human Genetics.* **51 (Pt 4)**:269-88.
162. **Wyman A.R., White R.** 1980. A highly polymorphic locus in human DNA. *Proceedings Of The National Academy Of Sciences Of The United States Of America.* **77**(11):6754-8.
163. **Yamamoto T., Tamaki K., Kojima T., Uchihi R., Katsumata Y.** 1994. Potential forensic applications of minisatellite variant repeat (MVR) mapping using the polymerase chain reaction (PCR) at D1S8. *J. Forensic Sci.* **39**(3):743-50.

164. **Yan J., Tang H., Liu Y., Jing Y., Jiao Z., Zhang Q., Gao J., Shang L., Guo H., Yu J.** 2007. Genetic polymorphisms of 17 Y-STRs haplotypes in Chinese Han population residing in Shandong province of China. *Legal Med.* **9**: 196-202.
165. **Yi H.F., Donohue S.J., Klein D.C., McBride O.W.** 1993. Localization of the hydroxyindole-o-methyltransferase gene to the pseudoautosomal region-implications for mapping of the psychiatric-desorders. *Hum. Mol. Genet.* **2**: 127-131.
166. **Zang Q., Yan J., Tang H., Jiao Z., Liu Y.** 2006. Genetic polymorphisms of 17 Y-STRs haplotypes in Tibetan ethnic minority group of China. *Legal Med.* **8**: 300-305.
167. **Zhu B., Wu Y., Shen C., Yang T., Deng Y., Xu X., Tian Y., Yan J., Li T.** 2007. Genetic analysis of 17 Y-chromosomal STRs haplotypes of Chinese Tibetan ethnic group residing in Qinghai province of China. *Forensic Sci. Int.* **10**: 1016-1022.
168. **Ziegle J.S., Su Y., Corcoran K.P., Nie L., Mayrand P.E., Hoff L.B., McBride L.J., Kronick M.N., Diehi S.R.** 1992. Application of automated DNA sizing technology for genotyping microsatellite loci. *Genomics.* **14**: 1026-1031.
169. **Zuliani G., Hobbs H.H.** 1990. Tetranucleotide repeat polymorphism in the LPL gene. *Nucleic Acids Res.* **18**: 4958-4961.