



**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
FACULTAD DE ESTUDIOS SUPERIORES ZARAGOZA**



**“Análisis Genético de 22 Loci STR en la Población
de la Ciudad de México”**

T E S I S

QUE PARA OBTENER EL TÍTULO DE

QUÍMICO FARMACÉUTICO BIÓLOGO

PRESENTA

ERIKA RAMÍREZ FLORES

DIRECTORA DE TESIS

QFB MA. LOURDES VEGA NAVARRETE

FACULTAD DE ESTUDIOS SUPERIORES ZARAGOZA UNAM

ASESORA DE TESIS

M. EN C. LEONOR AGUILAR SANTELISES

FACULTAD DE ESTUDIOS SUPERIORES ZARAGOZA UNAM

MÉXICO D.F., A 5 DE MAYO DE 2014

Agradecimientos

Quiero agradecer a mis padres por todo el apoyo brindado, no solo a lo largo de mi carrera sino de mi vida.

A mi madre Demetria Flores Pioquinto, quien ha sido mi guía, mi soporte, mi consejera y mi principal apoyo, gracias por siempre darnos a mis hermanos y a mí tu amor, apoyo y confianza, en palabras no podría expresarte la admiración que siento hacia ti, gracias por todo.

A mi padre Félix Ramírez González por apoyarme desde principio a fin y también por depositar su confianza en mí, gracias por tu apoyo.

A mis hermanos que siempre han estado presentes aconsejándome en cada decisión que he tomado. Gracias Erendida por tu confianza y apoyo, Misael por siempre escucharme y Luis (+) por enseñarme lo valioso de la familia, sabes que siempre estas con nosotros.

A mi tía Elena, que me brindo apoyo desde que empecé la carrera, también agradezco a Pas, Abi, a mis sobrinas Stephanie y Danna por siempre sacarme una sonrisa y también a mi abuelita Roberta (+) a quien recuerdo con mucho cariño.

A la maestra Lourdes Vega Navarrete, a quien admiro y respeto. Gracias por brindarme la oportunidad de realizar esta tesis y de seguir aprendiendo, por sus consejos y sobre todo por la confianza brindada.

A la bióloga Dora Villegas Carmona que me guió por este camino y me brindó su amistad, gracias por ser también una maestra para mí.

Al Ing. Jorge A. Zogbi Velásquez por darme la oportunidad de estar en su laboratorio y proporcionarme los medios para que esta tesis se pudiera llevar a cabo.

A la maestra Leonor Aguilar Santelises por apoyarme y brindarme parte de su tiempo en la revisión de esta tesis.

Índice

Índice de abreviaturas.....	3
1. Introducción	4
2. Marco teórico	5
2.1 Componentes y estructura del ADN.....	5
2.2 Organización del ADN.....	8
2.3 Transmisión de la información	9
2.4 Alelos y polimorfismos	10
2.5 Detección de los polimorfismos de longitud	11
2.5.1 Minisatélites	11
2.5.2 Microsatélites	11
2.6 Características deseables de STRs usados en la tipificación de ADN ...	14
2.7 Importancia del análisis estadístico de los STR.....	14
2.7.1 Estudios de paternidad	14
2.7.2 Criminalística	15
2.7.3 Personas desaparecidas.....	15
2.8 Genética de poblaciones.....	16
a) Frecuencias alélicas y genotípicas	16
b) Ley de equilibrio de Hardy Weinberg.....	17
c) Poder de exclusión	17
d) Poder de discriminación.....	17
e) Probabilidad de coincidencia aleatoria.....	17
f) Heterocigosidad.....	18
g) Índice de contenido polimórfico	18
h) Índice de paternidad	18
i) Número efectivo de alelos.....	18
2. 9 Población de la Ciudad de México.....	19
2.10 Loci de estudio	20
3. Planteamiento del problema	21
4. Hipótesis de trabajo	21

5. Objetivos	22
6. Metodología	22
6.1 Tipo de estudio	22
6.2 Población de estudio.....	22
6.3 Criterios de inclusión, exclusión y eliminación	23
6.4 Variables	23
6.5 Material y método	23
6.5.1 Material	23
6.5.2 Método	24
7. Resultados y descripción de resultados.....	26
Figura 7.1 Electroferograma	28
Tabla 7.1 Frecuencias alélicas.....	29
Tabla 7.2 Equilibrio de Hardy-Weinberg	30
Tabla 7.3 Parámetros estadísticos calculados.....	31
Figura 7.2 Heterocigosidad	32
Figura 7.3 Probabilidad de coincidencia aleatoria.....	32
Figura 7.4 Poder de discriminación.....	33
Figura 7.5 Poder de exclusión	33
Figura 7.6 Índice de contenido polimórfico.....	34
Figura 7.7 Índice de paternidad	34
Figura 7.8 Número efectivo de alelos	35
Tabla 7.4 Índices combinados	35
Tabla 7.5 Comparación con los paneles anteriores	36
8. Discusión de resultados	36
9. Conclusiones	39
10. Recomendaciones	40
11. Anexos	41
12. Referencias bibliográficas	42

Índice de abreviaturas

A	Adenina
ADN	Ácido desoxirribonucleico
ARN	Ácido ribonucleico
C	Citosina
CODIS	Sistema de índice Combinado de ADN (Combined DNA Index System)
dNTP's	Desoxinucleótidos trifosfato
EC	Electroforesis capilar
ESS	Conjunto de Estándares Europeos (European Standar Set)
G	Guanina
PCR	Reacción en cadena de la polimerasa
RFLP	Polimorfismo de longitud de fragmentos de restricción (Restriction Fragment Length Polymorphism)
STR	Repeticiones Cortas en Tándem (Short Tándem Repeat)
T	Timina
VNTR	Número variable de repeticiones en tándem (variable number of tandem repeats.)

1.- Introducción

El estudio y caracterización de una población ha tomado una gran importancia, debido al creciente uso de la tecnología de ADN en el sistema de justicia, por lo que es necesaria la creación de una base de datos de las frecuencias alélicas de una población, el objetivo principal al generar una base de datos es encontrar todos los alelos “comunes” y cuántas veces se repiten estos alelos en múltiples muestras, con el fin de estimar con fiabilidad los alelos presentes en la población de estudio y de esta manera establecer el peso estadístico de los perfiles genéticos, y las confrontas genéticas en la identificación de individuos.⁽¹⁻⁴⁾

En la actualidad, se utilizan marcadores STR (Short Tandem Repeat) para la obtención de perfiles genéticos, ya que son fácilmente amplificados por la reacción en cadena de la polimerasa y sin los problemas de una amplificación preferencial. Esto es debido al hecho de que el tamaño los dos alelos de un individuo heterocigoto son similares, y el tamaño de repetición es pequeño. El número de repeticiones en tándem en los marcadores STR puede ser altamente variable entre los individuos, lo que hace a estos marcadores eficaces para fines de identificación humana.⁽⁵⁾

La disponibilidad de kits STR que permiten la amplificación multiplex ha impactado favorablemente la ciencia forense debido a que la probabilidad de coincidencia puede superar una en varios miles de millones y esto es posible en una sola amplificación con 1 ng (o menos) de una muestra de ADN. Actualmente los resultados se pueden obtener en sólo unas cuantas horas en comparación con las semanas que requerían los métodos de polimorfismo de longitud de fragmentos de restricción (RFLP) hace unos pocos años.

En el mercado, existen diferentes paneles que contienen marcadores STR y permiten la amplificación multiplex, tales como los sistemas PowerPlex System y el sistema AmpF/STR® Identifier, sin embargo, se han presentado casos de asociaciones fortuitas o coincidencias aleatorias, por lo que se han desarrollado nuevos paneles con un mayor número de marcadores polimórficos, aumentando de esta manera el poder de discriminación y disminuyendo esas asociaciones.

En este estudio se analizó la capacidad del nuevo panel PowerPlex Fusion System, que integra los loci del CODIS y los loci del conjunto de Estándares Europeos (ESS) dando como resultado un panel con 24 loci (22 STR, 1 marcador de cromosoma “Y” y amelogenina)⁽⁶⁾. El estudio se centró en el análisis genético de 200 muestras de individuos de la Ciudad de México, seleccionados aleatoriamente, no relacionados biológicamente, así como la comparación del potencial del sistema PowerPlex Fusion con paneles anteriores, para observar si es una buena herramienta para el sistema de justicia mexicano.

2.- Marco teórico

La unidad básica de la vida es la célula, la cual produce las materias primas y la energía necesaria para mantener la vida. Un ser humano se compone de miles de millones de células, todas las cuales se originaron a partir de una sola célula. Dentro del núcleo de las células, existe una sustancia conocida como ADN que contiene la información para la replicación de la célula y la construcción de enzimas necesarias. Debido a que este ADN reside en el núcleo de la célula, se refiere a menudo como ADN nuclear.⁽¹⁾

Sin embargo existe otro tipo de ADN, el cual es un genoma circular y se encuentra dentro de las mitocondrias. La cantidad de moléculas de ADN mitocondrial puede ir de una a 15 copias por mitocondria, considerando que cada célula tiene cientos de mitocondrias, entonces tendría miles de copias de este ADN, en comparación con el ADN nuclear que solo hay una copia en el núcleo de cada célula.

El conjunto completo de instrucciones para construir un organismo, es decir, todo el ADN en una célula, se conoce colectivamente como genoma. El ADN tiene dos propósitos principales: 1) hacer copias de sí mismo para que las células puedan dividirse y seguir con la misma información, y 2) llevar las instrucciones para la síntesis de proteínas. La información codificada en la estructura del ADN se transmite de generación en generación, teniendo la mitad de la información genética origen materno y el resto, origen paterno.⁽¹⁾

2.1 Componentes y estructura del ADN

El ADN está constituido por unidades fundamentales, los desoxirribonucleótidos, también conocidos como bases, esto debido a que están constituidos por bases orgánicas cíclicas. Cada nucleótido contiene un fosfato, un azúcar (desoxirribosa) y una base nitrogenada (adenina, guanina, citosina y timina). Los nombres químicos completos de los nucleótidos son 5'-monofosfato de desoxiadenosina (dAMP), 5'-monofosfato desoxiguanosina (dGMP), 5'-monofosfato de desoxicitidina (dCMP) y 5'-monofosfato de timidina (dTMP).⁽⁷⁾

Hay dos tipos de bases nitrogenadas: las purinas (Adenina y Guanina), con un doble anillo de nueve lados, y las pirimidinas (Citosina y Timina), con un anillo de seis lados.⁽⁸⁾

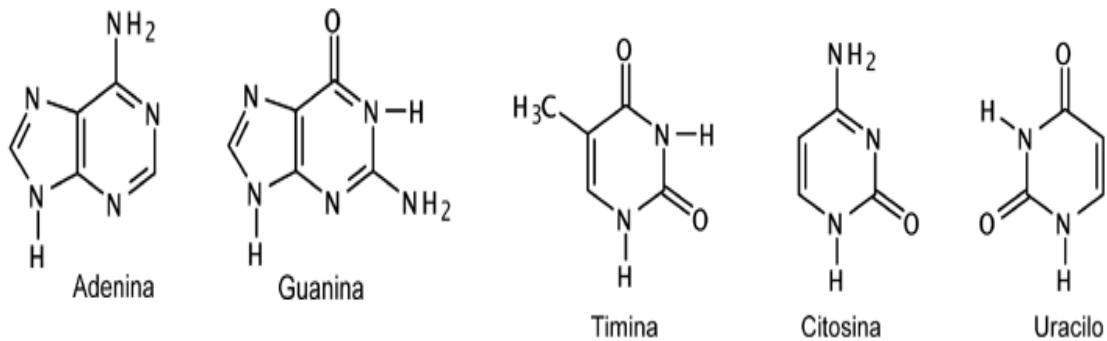


Figura 1. Estructura de las bases nitrogenadas.⁽⁹⁾

James Watson y Francis Crick dedujeron en 1953, con la información disponible hasta entonces, y con la ayuda del análisis cristalográfico por difracción de rayos X, la estructura correcta del ADN, obteniendo conclusiones como:

- Dos largas cadenas polinucleotídicas están enrolladas alrededor de un eje central, formando una doble hélice enrollada hacia la derecha (dextrógira).
- Las dos cadenas son antiparalelas; es decir, la orientación C-5'-C-3' va en direcciones opuestas.
- Las bases de las dos cadenas yacen formando estructuras planas y perpendiculares al eje; están apiladas unas sobre otras, separadas por 3,4 Å (0,34 nm), y se encuentran en el interior de la estructura.
- Las bases nitrogenadas de las cadenas opuestas están apareadas como resultado de la formación de puentes de hidrógeno; en el ADN sólo se permiten los emparejamientos A con T y G con C.
- Cada vuelta de la hélice tiene una longitud de 34 Å (3.4 nm); de este modo, cada vuelta de la cadena contiene 10 bases.
- En cualquier segmento de la molécula, se observa un surco mayor y un surco menor que se alternan a lo largo del eje.
- La doble hélice mide 20 Å (2.0 nm) de diámetro.^(8,10,11)

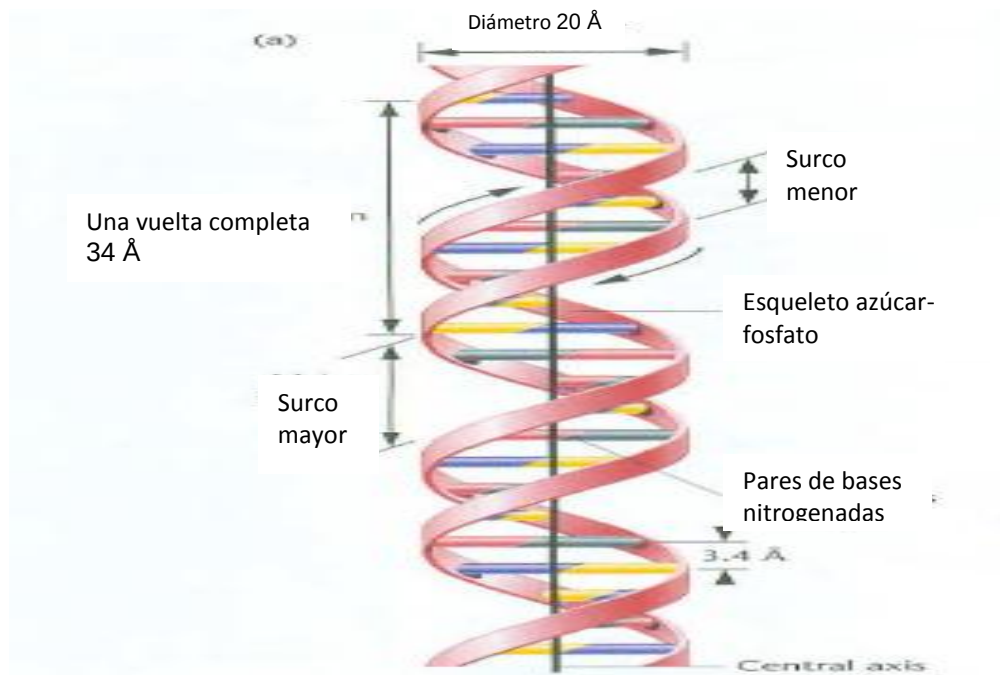


Figura 2.- Modelo de doble hélice propuesto por Watson y Crick. Modificado de Klug SW., Cummings.⁽⁸⁾

Si se extendiera la molécula de ADN de todos los cromosomas, su longitud total sería de dos metros. El ADN se puede guardar en el núcleo de las células gracias a una perfecta organización de la compactación de dicho ADN. Esta extraordinaria manera de empaquetar el ADN para reducir su tamaño es posible gracias a unas proteínas llamadas “histonas” que son las encargadas de asociarse al ADN y llevar a cabo la compactación del mismo.⁽¹²⁾

Las histonas son proteínas básicas con carga parcial positiva, ricas en residuos de lisina y arginina. Desempeñan un papel importante en la fijación del esqueleto azúcar-fosfato que tiene carga negativa.^(13,14,15)

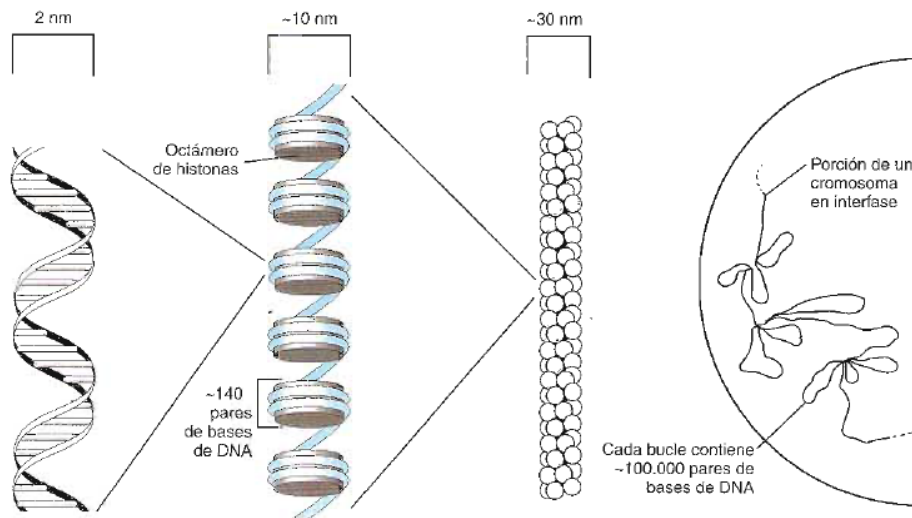


Figura 3.- Niveles jerárquicos de empaquetamiento de la cromatina. Tomado de Nussbaum RL., McInnes RR., Willard HF.⁽¹⁶⁾

2.2 Organización del ADN

El ADN está dividido en cromosomas, en forma de densos paquetes de ADN, el genoma humano consiste de 22 pares de cromosomas autosómicos y un par de cromosomas sexuales. Con base en los cromosomas sexuales, los varones se designan como XY porque contienen una sola copia del cromosoma X y una sola copia del cromosoma Y, mientras que las mujeres tienen dos copias del cromosoma X y se designan como XX.⁽¹⁾

Los cromosomas de las células de todo el cuerpo (somáticas) son diploides; es decir, contienen dos juegos de cada cromosoma. Por otro lado, los gametos (espermatozoides u ovocitos) son haploides; ellos tienen un solo juego de cromosomas. Cuando un ovocito y un espermatozoide se combinan durante la concepción, resulta un cigoto que de nuevo es diploide. Por lo tanto, un cromosoma de cada par cromosomal se deriva de uno de los padres en el momento de la concepción.⁽¹⁾

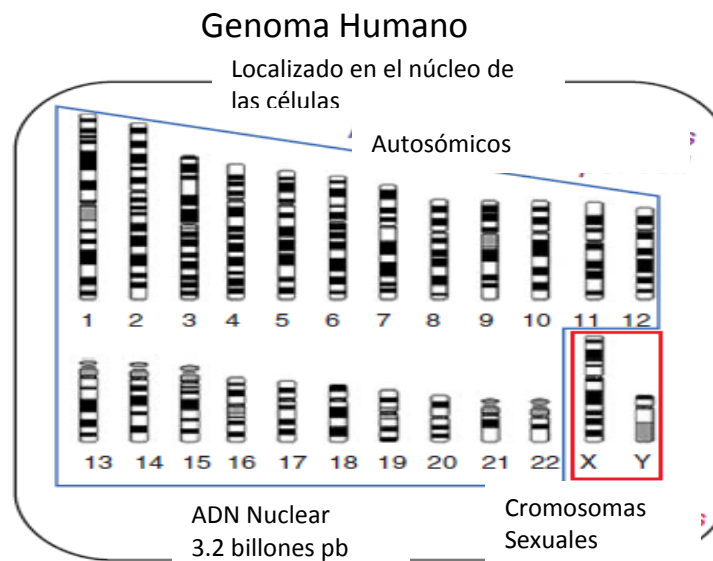


Figura 4.- Genoma humano. Modificado de Butler JM.⁽¹⁾

Además, el ADN en los cromosomas está compuesto de regiones “codificantes” y “no codificantes”. Las regiones codificantes son conocidas como genes y contienen la información necesaria para la síntesis de proteínas, y forman alrededor del 5% del genoma humano. Las regiones no codificantes de ADN forman el resto de nuestro material cromosómico. Debido a que estas regiones no están directamente relacionadas con la producción de proteínas se refieren a veces como ADN basura y son usados como marcadores para pruebas de identificación humana debido a su alto polimorfismo.^(1, 17)

2.3 Transmisión de la información

En 1928, el microbiólogo inglés Frederick Griffith realizó el descubrimiento asombroso de que las cepas no virulentas de las bacterias se tornaban virulentas cuando se mezclaban con sus contrapartidas patógenas muertas por el calor. Esto alertó sobre la posibilidad de que cuando células patógenas se mueren por calor, sus componentes genéticos permanecen sin daños.⁽¹⁸⁾

Varios años después se comprobó que extractos de las bacterias muertas eran capaces de inducir transformaciones hereditarias y se inició la búsqueda de la identidad química del agente transformador. En 1944, Oswald T Avery y sus colegas, anunciaron que el principio genético activo era el ADN, gracias a sus experimentos con desoxirribonucleasa pancreática, una enzima que degrada de manera específica moléculas de ADN hasta sus componentes de nucleótidos y no tiene efecto alguno sobre la integridad de las moléculas proteicas ni sobre el ARN.⁽¹⁹⁾

Otro estudio importante fue hecho por Alfred D. Hershey y Martha Chase, en el cual se marcaba la cubierta proteica de un bacteriófago con isótopo radioactivo ³⁵S y el núcleo de ADN con el isótopo radiactivo ³²P. De este experimento surgió un resultado bien definido; en la progenie del fago se detectó la mayor parte del ácido nucleico del progenitor y nada de las proteínas de éste.⁽²⁰⁾

Con estos experimentos y con el descubrimiento de la doble hélice, había terminado de manera definitiva cualquier controversia acerca de si el ADN era la sustancia genética primaria. La preocupación principal de los genetistas moleculares era conocer la manera en que la información genética del ADN actúa para ordenar los aminoácidos durante la síntesis proteica, ya que el ADN debe aportar la información para ordenar los aminoácidos, era bastante obvio que la doble hélice en sí no podía ser la plantilla para la síntesis de las proteínas, ya que la síntesis de las proteínas en todas las células eucariotas se produce en el citoplasma, que está separado del ADN nuclear por la membrana nuclear.

En consecuencia, tenía que haber una segunda molécula con contenido de información que obtuviese la especificidad genética del ADN y luego esta se trasladará hacia el citoplasma para funcionar como plantilla para la síntesis proteica. Desde un principio, la atención se concentró en la segunda clase de ácidos nucleicos, el ARN.⁽⁸⁾

Con base en estos descubrimientos, se adoptó la hipótesis de trabajo de que el DNA cromosómico funcionaba como plantilla para la molécula de ARN, que con ulterioridad este último se trasladaba al citoplasma, donde determinaban la disposición de los aminoácidos dentro de las proteínas. Francis Crick llamó dogma central a este mecanismo para el flujo de la información genética.^(21, 22)

Propuesta inicial de Crick (1970)

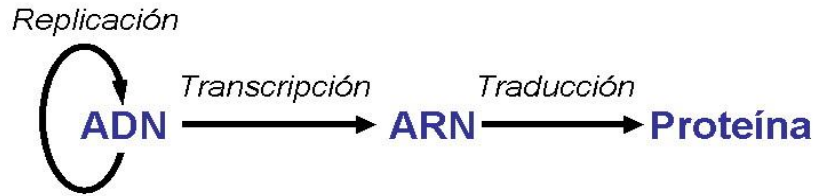


Figura 5. Dogma central de la biología molecular. ⁽²¹⁾

Este esquema central de flujo de la información pronto fue modificado, ya que en algunos virus cuyo material hereditario es ARN, la información se conserva o mantiene mediante replicación del ARN. Además, también se comprobó que la información no va siempre del ADN hacia el ARN (ADN→ARN), en algunos casos la información puede fluir del ARN hacia el ADN (ARN→ADN), es decir sintetizar ADN tomando como molde ARN, teniendo lugar el fenómeno de la *transcripción inversa*.⁽²¹⁾



Figura 6. Modificaciones del "Dogma Central de la Biología Molecular". ⁽²¹⁾

2.4 Alelos y polimorfismos

Alelo es el término que define a formas alternativas de un gen o locus, y locus se utiliza para designar una posición o localización específica de un gen o una secuencia blanco en un cromosoma.⁽²¹⁾

La coexistencia de alelos múltiples en un locus se denomina polimorfismo genético. Por definición, cualquier sitio en el cual existen alelos múltiples como componentes estables de la población, es polimórfico. En otras palabras, el polimorfismo expresa la variabilidad que existe dentro de un fragmento de ADN. Como regla general, cuantos más alelos haya, mayor polimorfismo y por ende, mayor poder de identificación.^(23,24, 25)

Los polimorfismos pueden ir desde la modificación de una sola base, hasta cambios en el número y tamaño de las unidades de repetición:

- ◆ Polimorfismos de secuencia. Se producen por el cambio de uno o más nucleótidos en una secuencia de ADN. Como polimorfismos de secuencia

podemos encontrar aquellos genes que codifican para los grupos sanguíneos, el sistema HLA, el análisis de ADN mitocondrial, por mencionar algunos ejemplos.

- ◆ Polimorfismos de longitud. Se presentan por la inserción o delección de uno o más nucleótidos. Dentro de los polimorfismos de longitud se encuentran principalmente los minisatélites y microsatélites. Ambos se caracterizan por tener una secuencia que se repite varias veces, en tándem. Este tipo de polimorfismo se presenta en regiones no codificantes, lo cual hace que diferencias en las secuencias de ese ADN no sean raras y se transmitan como tal a la descendencia, por lo que son comúnmente utilizados como marcadores genéticos en la identificación humana.^(1,26,27)

2.5 Detección de los polimorfismos de longitud

2.5.1 Minisatélites

Para los minisatélites la unidad de repetición va de 10 a 100 bases. La mayoría de los polimorfismos minisatélites se detectan mediante enzimas de restricción, y se les conoce como VNTR (Número Variable de Repeticiones en Tándem). La identificación de los VNTR se hace con la utilización de sondas complementarias sintéticas que hibridan con la secuencia de ADN, propia del VNTR en estudio (secuencia repetitiva). Estas sondas pueden ser “mono locus”, que hibridan en una sola región de ADN, y las “multi locus”, que hibridan en forma simultánea con varias regiones.

Los análisis con minisatélites contienen varias limitaciones, entre las cuales sobresalen: el tiempo que se invierte en cada estudio (tres semanas), la elevada concentración de ADN que deben tener las muestras, y la poca eficiencia del estudio cuando se trata de muestras con ADN degradado, por el mayor tamaño del ADN que se examina.^(28, 29, 30, 31)

2.5.2 Microsatélites

Los microsatélites contienen el mismo tipo de polimorfismo de longitud, pero varían en el tamaño de la unidad repetitiva. Para los microsatélites, la unidad de repetición es de 2 a 6 nucleótidos, a este tipo de polimorfismo se les conoce también como STR (Short Tandem Repeats). La aparición de los STR en el ámbito forense ocurrió a principios de la década de los 90, gracias al desarrollo del proceso denominado Reacción en Cadena de la Polimerasa (PCR).^(28, 32, 33)

Este tipo de polimorfismo puede ser analizado mediante diversos pasos, los cuales son:

a) Extracción del ADN: Cuando se obtienen muestras biológicas, ya sea una de un lugar de los hechos o muestras de referencia, contienen además del ADN otras sustancias que pueden inhibir la habilidad de analizarlo, como son proteínas celulares que empaquetan y protegen al ADN en el interior de la célula o inhibidores de la polimerasa en el soporte en el que está contenido el indicio. Por lo que, se han desarrollado métodos de extracción del ADN para separar las proteínas y otros materiales que están unidos a las moléculas de ADN. En la actualidad, se utilizan tres principales técnicas para la extracción del ADN en los laboratorios forenses: la extracción orgánica, la extracción con Chelex y la extracción en papel FTA ^(1,34,35), además de otros métodos basados en la aplicación de resinas paramagnéticas cuyo uso se ha extendido en los laboratorios forenses debido al rendimiento de ADN de alta pureza.

b) Cuantificación del ADN: Después de extraer el ADN, es necesario conocer la cantidad presente, así como calidad del mismo. La adición de una cantidad óptima de ADN para PCR producirá mejores resultados en un menor tiempo; la adición de demasiado ADN o una cantidad insuficiente, dará lugar a un perfil que será difícil o incluso imposible de interpretar. Esto es especialmente importante cuando se trata de indicios, cuando es muy difícil conocer el estado de conservación del material biológico y, en muchos casos, es difícil estimar la cantidad que se ha recogido de material celular. Es menos importante cuantificar ADN utilizando muestras de referencia, donde se espera una cantidad similar de ADN en cada extracción, ya que no hay muchas variables que afecten su concentración.⁽³⁶⁾

c) Amplificación: La PCR es un proceso enzimático en el que se replica una región específica de ADN una y otra vez para producir muchas copias de una secuencia particular. Este proceso molecular de "fotocopiado" consiste en calentar y enfriar las muestras en un patrón de ciclos térmicos. Durante cada ciclo, una copia de la secuencia de ADN diana se genera por cada molécula que contiene la secuencia diana. Los límites del producto amplificado se definen por los cebadores de oligonucleótidos que son complementarias a los extremos 3' de la secuencia de interés. Teóricamente, después de 30 ciclos se han generado más de mil millones de copias de la región diana de ADN. Este producto de PCR, a veces referido como un "amplicón", es entonces una cantidad suficiente que se puede medir fácilmente por una variedad de técnicas.^(1, 37, 38)

d) Electroforesis capilar: En la reacción en cadena de la polimerasa (PCR) en la que los alelos STR se amplifican, se produce una mezcla de moléculas de ADN que representa un problema de difícil separación. La PCR multiplex puede producir 20 o más fragmentos de ADN que deben y pueden ser distinguidos uno de otro. Además, la resolución de una sola base requiere que se diferencie entre

los alelos estrechamente espaciados (por ejemplo, alelos TH01 9,3 y 10). Con el fin de observar las diversas moléculas una de otra, el paso de separación requiere mover aparte los fragmentos en diferentes tamaños. La separación se realiza típicamente por un proceso conocido como electroforesis y se lleva a cabo ya sea en gel o en capilar. ^(1, 39)

La detección de la muestra se lleva a cabo automáticamente por el instrumento de Electroforesis Capilar (EC) a través de la medición del intervalo de tiempo de inyección de la muestra y cuya detección se realiza a través de una CCD camera. La luz de un láser incide en el capilar en una posición fija, en donde se ha quemado una ventana en la cubierta del capilar. Los fragmentos de ADN son iluminados a medida que pasan por la ventana en el capilar. Como en los geles, las moléculas más pequeñas llegan primero al punto de detección, seguidas por las moléculas más grandes. Los datos obtenidos de la separación están representados gráficamente en función de la intensidad relativa de fluorescencia observada a partir de la emisión de fluorescencia de los colores que pasan el detector. Las señales de emisión fluorescente de los colorantes que están unidos a las moléculas de ADN se pueden utilizar para detectar y cuantificar las moléculas de ADN que pasan por el equipo. ^(1,40,41, 42, 43)

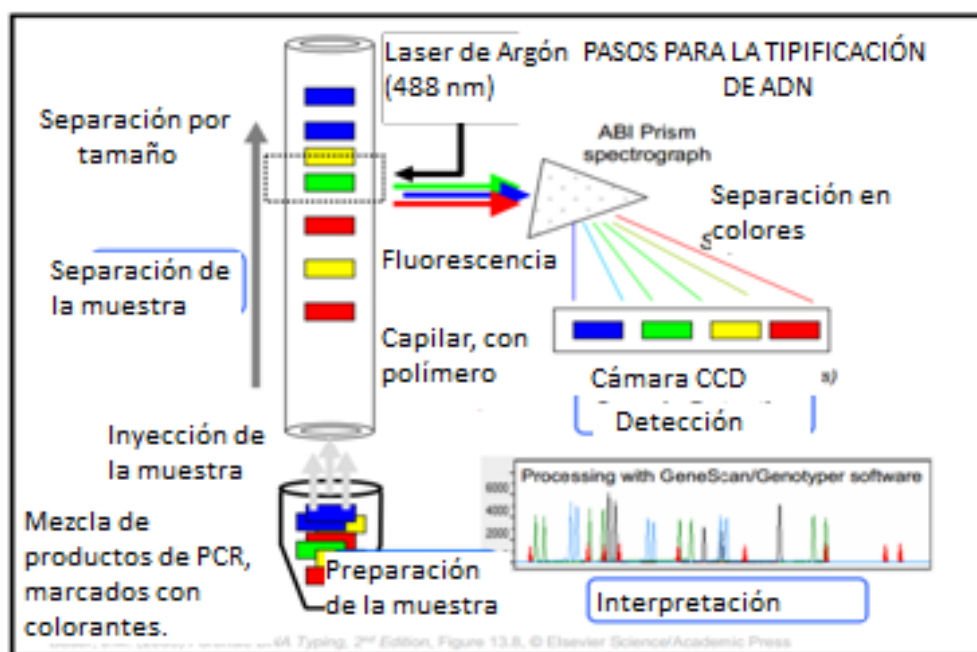


Figura 7. Separación y detección de Fragmentos STRs. Modificado de Butler JM. Forensic DNA Typing: biology, technology and genetics of STR markers. ⁽¹⁾

e).- Análisis de fragmentos mediante el uso de software: Los espectros de los colorantes usados para marcar los productos de la PCR se traslapan y los datos obtenidos contienen picos que se componen de más de una marca fluorescente. El software, ya sea el GeneScan o el GeneMapper, elimina la superposición espectral en el perfil y calcula los tamaños de los fragmentos de ADN

amplificados. El software calcula la cantidad de solapamiento espectral que hay entre cada colorante y resta este de los picos en el perfil, obteniendo de esta manera picos separados que pueden ser interpretados.⁽³⁶⁾

2.6 Características deseables de STRs usados en la tipificación de ADN

Para la identificación humana, es importante tener marcadores de ADN que exhiban la más alta variación posible en la población o al menos un número de marcadores polimórficos para que se puedan combinar con el fin de obtener una capacidad para discriminar entre muestras. Las mezclas prevalecen, así como en algunas muestras forenses, tales como las obtenidas en casos de delitos sexuales que contienen material biológico tanto de la víctima como del perpetrador.

El pequeño tamaño de los alelos STR (100-400pb) comparado con los alelos minisatélites (400-1000pb) hacen que los marcadores STR sean mejores candidatos para usarlos en el sistema forense en donde es común encontrar muestras con ADN degradado. La amplificación por PCR de muestras con ADN degradado puede realizarse mejor con productos de menor tamaño.

Entre los diversos tipos de STR, los tetranucleótidos son más populares que los di o trinucleótidos. Penta y hexanucleótidos son menos comunes en el genoma humano pero están siendo examinados por algunos laboratorios.^(1, 44). Actualmente en los nuevos kits que existen en el mercado han introducido en sus paneles pentanucleótidos (Como el Penta D y el Penta E que están dentro del conjunto de marcadores del PowerPlex Fusion).⁽⁷⁰⁾

2.7 Importancia del análisis estadístico de los STR

2.7.1 Estudios de paternidad

El que cada persona conozca a su padre o a su madre no es sólo normal, sino que además es un derecho que suele ser garantizado por las leyes de muchos países. No obstante, en ciertas circunstancias una persona puede ignorar quién es su padre (o madre en algunos casos) o, tener dudas sobre su paternidad. El análisis de la paternidad se basa en las leyes mendelianas de la herencia, en las que se establece por primera vez que todo el material genético de una persona procede a partes iguales del padre y de la madre. Por lo tanto, si en un hijo encontramos un alelo que no posee el probable padre, la paternidad queda excluida con seguridad absoluta.

Si por lo contrario encontramos que el hijo tiene alelos que pudieron haberse heredado del probable padre, hay que realizar una serie de cálculos estadísticos

con el objetivo de determinar cuántas personas entre la población general, podrían ser también padres potenciales, por tener esos alelos.⁽⁴⁵⁾

2.7.2 Criminalística

El delito en general, surgió con el hombre mismo, y éste, consciente de la punibilidad de su acción, siempre ha intentado ocultar la autoría. Y desde siempre el delito ha venido acompañado de la necesidad de investigarlo, de aclararlo, de buscar y castigar al culpable.

Con el paso del tiempo se ha ido desarrollando la criminalística, que podemos definir como la “ciencia aplicada que de acuerdo al ordenamiento jurídico de cada país, estudia científicamente los indicios y evidencias con objeto de convertirlos en pruebas formales que puedan ser presentadas ante las autoridades judiciales para permitir la identificación de víctimas y de los delincuentes y esclarecer las circunstancias de un probable delito”.⁽⁴⁵⁾

El obtener un perfil es sólo la primera parte del análisis, los perfiles tienen que ser comparados. Por ejemplo, si un perfil obtenido de una evidencia recolectada de un lugar de los hechos es comparado con el perfil del probable responsable, se presentan dos posibles situaciones: los perfiles pueden ser los mismos o pueden ser diferentes. Si los perfiles son diferentes, la interpretación es simple: la evidencia recolectada del lugar de los hechos, no pertenece al probable responsable, y se considera como una exclusión. Si los perfiles son los mismos, entonces es una inclusión y significa que la coincidencia debe ser estimada.

El primer paso es estimar la frecuencia de cada alelo encontrado en una población, con esta información las frecuencias genotípicas pueden ser calculadas. Si tuviéramos una base de datos lo suficientemente grande podríamos hacerlo de la misma manera en que calculamos las frecuencias alélicas. El problema con este enfoque es que la ocurrencia de muchos genotipos es baja y sería difícil hacer estimaciones precisas.

En lugar de conteo directo, se utiliza el modelo de equilibrio de Hardy Weinberg para predecir la proporción de genotipos. Este modelo, describe cómo se comportan los alelos o genes en una población ideal.⁽⁴⁶⁾

2.7.3 Personas desaparecidas

Los conflictos armados, como la guerra u otras situaciones de violencia armada generalizada, así como también catástrofes naturales suelen traer consigo la desaparición de numerosas personas. Cualquiera que sea el motivo de su desaparición, la falta de noticias de sus seres queridos, sumada a la incertidumbre acerca de su paradero, causan sufrimientos indecibles a los familiares afectados.

En estas situaciones, es frecuente que las personas desaparecidas hayan muerto. El único alivio para los familiares es recibir una confirmación fidedigna de la muerte y saber que los restos de sus seres queridos han sido o pueden ser tratados con dignidad y con respeto por su cultura y sus creencias religiosas. Por

ello la identificación adecuada de los restos humanos es parte fundamental del proceso de reparación no sólo para los familiares de personas desaparecidas, sino para comunidades enteras.⁽⁴⁷⁾

La identificación de un cadáver depende básicamente del estado del mismo, si el cuerpo está en condiciones que permitan tomar fotografías, huellas dactilares o datos odontológicos, los datos obtenidos pueden ser de gran ayuda para la identificación. Cuando todas estas posibilidades se agotan, o cuando el cadáver no es sino un esqueleto (a veces incompleto), surge la necesidad de usar algo más específico, a ser posible único, que esté en todos los cadáveres y que se pueda encontrar incluso en los huesos: ese algo es el ADN.^(45, 48)

2.8 Genética de poblaciones

La población se define como un grupo de individuos que comparten un conjunto de genes, que viven en la misma área geográfica y que tienen capacidad de reproducirse.^(31, 49,50, 51)

La genética de poblaciones es el estudio de la variación heredada y su modulación en tiempo y espacio. Es un intento de cuantificar la variación observada dentro de un grupo de población o entre diferentes grupos de población, en términos de frecuencias alélicas y genotípicas.⁽¹⁾

La variabilidad dentro de un locus tiene que ser lo suficientemente estable como para pasar con precisión al alelo a la próxima generación (es decir, que posean una tasa de mutación baja). Sin embargo, tampoco deben ser demasiado estables porque si no sólo unos pocos alelos existirían en el tiempo y el locus no podría aportar tanta información (es decir, no sería útil en pruebas de identificación humana).⁽¹⁾

a) Frecuencias alélicas y genotípicas

La frecuencia genotípica es la cantidad de veces que aparece el genotipo en una población.⁽⁵²⁻⁵⁴⁾

La frecuencia de un alelo se define como el número de copias del alelo en la población, dividido por la suma de todos los alelos en esa población.

El cálculo de las frecuencias alélicas de una población se hace teniendo en cuenta que los homocigotos tienen dos alelos idénticos y los heterocigotos dos alelos diferentes donde: la probabilidad para un heterocigoto, o sea, que tenga ambos alelos en un solo locus es:

$$P=2pq$$

Mientras que para un individuo homocigoto la probabilidad del genotipo es:

$$P=p^2$$

b) Ley de equilibrio de Hardy Weinberg

En grandes poblaciones de apareamiento aleatorio en ausencia de mutación, migración y selección, las frecuencias génicas y genotípicas permanecen constantes de generación en generación, pudiendo predecir sus valores para las futuras generaciones. ^(1, 30, 55-57)

De acuerdo con la Ley de equilibrio de Hardy-Weinberg, las frecuencias genotípicas de la progenie están determinadas por las frecuencias alélicas de sus progenitores. ^(1, 58, 59)

Si la población está en equilibrio, las frecuencias alélicas son iguales en los progenitores y en la progenie, de manera que las frecuencias alélicas observadas en la progenie pueden usarse como si fueran las frecuencias alélicas paternas para calcular las frecuencias genotípicas esperadas por la Ley de Hardy-Weinberg. ⁽¹⁾

c) Poder de exclusión

Es la probabilidad de que un sistema genético específico proporcione evidencias que conduzcan a la exclusión de un probable responsable. ^(52, 53, 60-62)

$$\text{Poder de Exclusión (PE)} = H^2 (1 - (1 - H)H^2)$$

Donde H= Heterocigosidad

d) Poder de discriminación

Es la probabilidad de que dos individuos no relacionados puedan ser diferenciados genéticamente mediante el análisis de uno o varios marcadores. Depende del número de loci analizados, de las variables alélicas encontradas y frecuencias de cada locus en cada población. ^(1, 52, 53, 62)

$$PD=1-P_i$$

P_i = Probabilidad coincidencia aleatoria

e) Probabilidad de coincidencia aleatoria

Es la probabilidad de que dos individuos seleccionados al azar tendrán un genotipo idéntico en el locus sometido a prueba. ^(1, 52, 53, 62)

Es el número de individuos que podrán ser examinados antes de encontrar el mismo patrón de ADN en un individuo seleccionado aleatoriamente.

$$\text{Probabilidad de coincidencia aleatoria} = P_i = \sum_{i=1}^n x_i^2$$

Donde X_i^2 = Frecuencia de los genotipos

f) Heterocigosidad

Es la proporción de individuos heterocigóticos que se espera en la población. Una alta heterocigosidad significa que existe más diversidad alélica y por lo tanto hay una menor probabilidad de coincidencia en muestras. ^(1, 52, 53, 62, 63)

$$\text{Heterocigosidad (H)} = 1 - \text{Homocigosidad} = 1 - \sum_{i=1}^n P_i^2$$

g) Índice de contenido polimórfico

Refleja la capacidad de deducir el origen parental de cada uno de los alelos del genotipo en la descendencia en un locus particular. ^(1, 52, 53, 62, 64, 65)

$$\text{PIC} = 1 - \sum_{i=1}^n P_i^2 - (\sum_{i=1}^n P_i^2)^2 + \sum_{i=1}^n P_i^4$$

h) Índice de paternidad

Es la probabilidad de que los alelos presentes en el niño apoyen la hipótesis de que el hombre examinado es el verdadero padre biológico, en lugar de otro hombre examinado no relacionado biológicamente, tomado aleatoriamente de la población. ^(1, 52, 53, 62)

$$\text{Índice de Paternidad (PI)} = \frac{H+h}{2h} = \frac{(1-h)+h}{2h} = \frac{1}{2h}$$

i) Número efectivo de alelos

Hace referencia a los alelos con capacidad de pasar a la siguiente generación. ^(66, 67)

Esta medida indica el número de alelos esperados en cada locus de una determinada población, y es la inversa de la homocigosidad esperada de un locus.

$$\text{Número efectivo de alelos (n}_e\text{)} = \frac{1}{\text{Homocigosidad}} = \frac{1}{h}$$

2.9 Población de la ciudad de México ⁽⁶⁸⁾

Cuadro 1. Se muestran los datos poblacionales de la Ciudad de México y del país en general.

	DISTRITO FEDERAL	MÉXICO (TOTAL)
Población total, 2010	8,851,080 (1/12)	112,336,538
Población total hombres, 2010	4,233,783	54,855,231
Población total mujeres, 2010	4,617,297	57,481,307
Porcentaje de población emigrante, 2010	-9.30	No aplica
Porcentaje de población inmigrante (migración reciente), 2010	3.00	No aplica
Porcentaje de población no nativa (migración acumulada), 2010	18.97	No aplica
Densidad de población (Habitantes por kilómetro cuadrado), 2010	5,920.45	57.30
Tasa de crecimiento total, 2010	0.30 $\approx$ 1/5	1.40

Distrito Federal
División Municipal

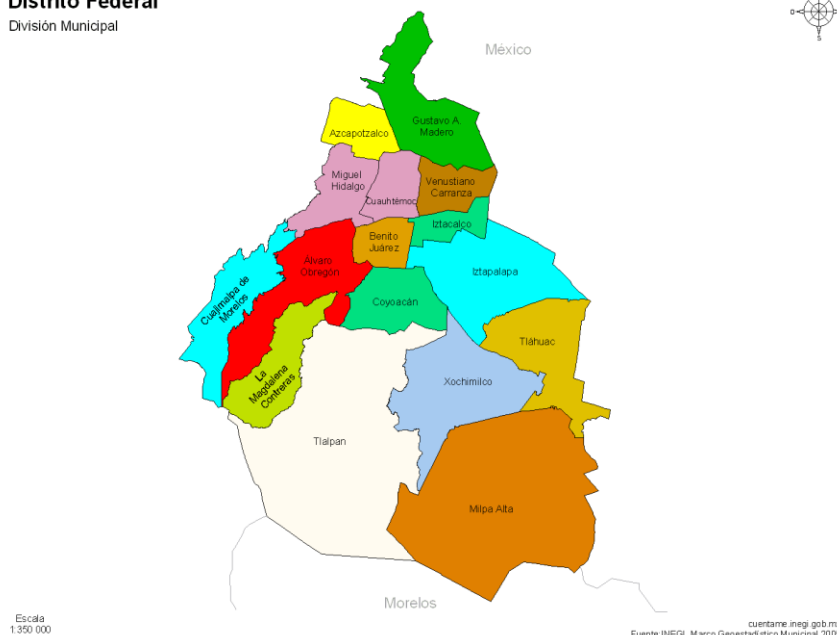


Figura 8. Distribución geográfica de la Ciudad de México. El Distrito Federal colinda al norte, este y oeste con el estado de México y al sur con el estado de Morelos. ⁽⁶⁹⁾

2.10 Loci de estudio

En este estudio se analizará el potencial del sistema PowerPlex Fusion que permite la co-amplificación y detección fluorescente de 24 loci (23 STR y amelogenina), incluyendo los loci del CODIS y los loci del conjunto de Estándares Europeos (ESS).⁽⁷⁰⁾

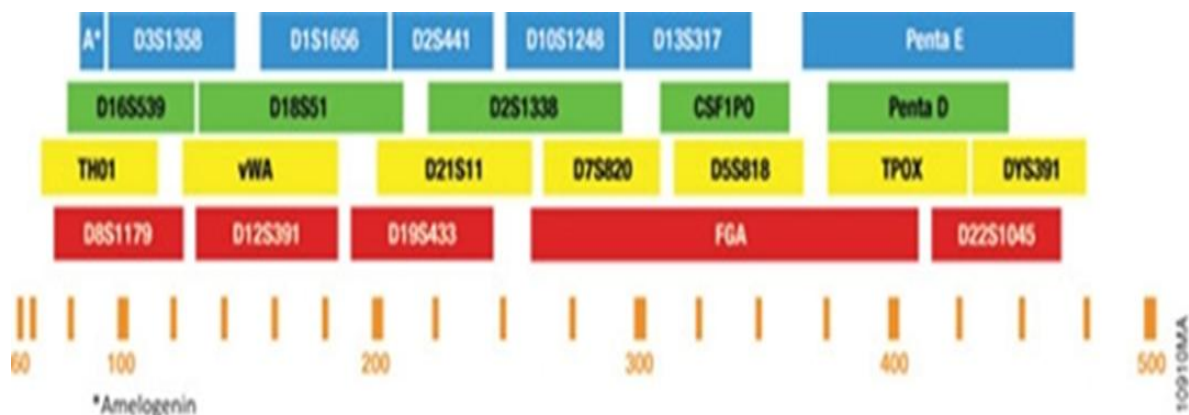


Figura 9. Configuración del sistema PowerPlex Fusion. Este sistema permite la co-amplificación de Amelogenina, D3S1358, D1S1656, D2S441, D10S1248, D13S317 y Penta E marcados con fluoresceína; D16S539, D18S51, D2S1338, CSF1PO y Penta D marcados con JOE; TH01, vWA, D21S11, D7S820, D5S818, TPOX y DYS391 marcados con TMR-ET; y D8S1179, D12S391, D19S443, FGA y D22S1045 marcados con CXR-ET. El estándar interno (ILS 500) está marcado con el colorante CC5 y contiene 21 fragmentos de ADN de 60, 65, 80, 100, 120, 140, 160, 180, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475 y 500 bases.⁽³⁰⁾

3.- Planteamiento del problema

El uso de la tecnología del ADN se ha vuelto más frecuente por lo que, cada vez se exigen resultados más confiables y precisos que eviten asociaciones fortuitas y nos permitan enfrentar los desafíos que representa el trabajar muestras severamente comprometidas, ya que de esto dependen situaciones importantes como el hallazgo de una persona desaparecida entre una gran cantidad de restos humanos. Se conocen actualmente sistemas que incluyen un mayor número de marcadores con el fin de aumentar el poder de discriminación y disminuir la probabilidad de coincidencia aleatoria. En este estudio se determinaron diversos parámetros como el poder de discriminación, la probabilidad de coincidencia aleatoria, Heterocigosidad, índice de contenido polimórfico, índice de paternidad y número efectivo de alelos del panel PowerPlex Fusion para la población de la Ciudad de México con el objetivo de demostrar si este panel es una buena herramienta para atender las demandas que impone la actual situación de nuestro país. Así también se determinó si las frecuencias alélicas de todos los loci están en equilibrio de Hardy-Weinberg para que puedan ser utilizadas en los cálculos estadísticos.

Por lo que el planteamiento central es si las frecuencias alélicas están en equilibrio y pueden ser utilizadas al determinar parámetros estadísticos y si el panel Powerplex Fusion, el cual es el más reciente sistema de identificación humana lanzado al mercado, provee más certeza estadística al momento de realizar las confrontas de perfiles genéticos.

4.- Hipótesis

Al tener un mayor número de marcadores autosómicos polimórficos, se espera que el panel *PowerPlex Fusion* tenga un mayor poder de discriminación para que de esta manera aumente la certeza estadística en la identificación humana, y se eviten coincidencias aleatorias entre perfiles genéticos, en comparación con los paneles PowerPlex 16HS y PowerPlex 21 System,. También se espera que las frecuencias alélicas de la muestra poblacional estén en equilibrio de Hardy-Weinberg en todos los loci del sistema para que puedan ser utilizados en los cálculos estadísticos y que la heterocigosidad sea mayor o igual al 80% en la mayoría de los loci del panel.

5.- Objetivos

- ④ Determinar las frecuencias alélicas de los 22 marcadores autosómicos polimórficos del panel PowerPlex Fusion en la población de la Ciudad de México.
- ④ Determinar si la población está en equilibrio de Hardy-Weinberg.
- ④ Determinar los parámetros estadísticos: Heterocigosidad, Poder de discriminación, Poder de Exclusión, Índice de contenido polimórfico, Índice de paternidad, Número efectivo de alelos y Probabilidad de Coincidencia aleatoria de cada locus del panel PowerPlex Fusion en la población de la Cd. de México.
Comparar los parámetros estadísticos Heterocigosidad, Poder de discriminación, Poder de Exclusión, Índice de paternidad y Probabilidad de Coincidencia aleatoria del Kit PowerPlex Fusion con los Kits PowerPlex 16 HS y PowerPlex 21.

6.- Metodología

6.1 Tipo de estudio

- ◆ Observacional
- ◆ Prospectivo
- ◆ Transversal
- ◆ Descriptivo

6.2 Población de estudio

- ◆ Tamaño: 200 muestras
- ◆ Tipo: Muestras sanguíneas
- ◆ Soporte: Tarjetas FTA
- ◆ Población: Individuos no relacionados biológicamente de la Ciudad de México

6.3 Criterios de inclusión, exclusión y eliminación

INCLUSIÓN:

- Muestras sanguíneas de individuos no relacionados biológicamente de la Cd. de México.

EXCLUSIÓN:

- Individuos relacionados biológicamente.

ELIMINACIÓN:

- Muestras contaminadas.

6.4 Variables

INDEPENDIENTE:

- ◆ El genotipo de cada individuo de la muestra poblacional.

DEPENDIENTES:

- ◆ Las frecuencias alélicas.
- ◆ Índice de heterocigosidad y homocigosidad.
- ◆ Poder de discriminación del panel.
- ◆ Poder de exclusión del panel
- ◆ Contenido de índice polimórfico

6.5 Material y método

6.5.1 Material

- Analizador automático de biología molecular 3100-Avant Genetic Analyzer de AppliedBiosystem.
- Termociclador GENEAMP® SYSTEM 9700.
- Campana de Bioseguridad Air Clean600 PCR Workstation
- Micropipeta Eppendorf de 10µL.
- Micropipeta Eppendorf de 100µL.
- Micropipeta Eppendorf de 1000µL.
- Vortex, Sarstedt.
- Puntas para pipeta de 10µL.
- Puntas para pipeta de 100µL.
- Puntas para pipeta de 1000µL.
- Harris Micro-Punch para papel FTA
- Guantes de nitrilo.
- Tubos Eppendorf de 1.5mL.

- Placa de 96 pozos
- Tapas de goma, para charola de 96 pozos. (septa)
- Tapas para charola de amplificación
- Lancetas
- Papel FTA
- Torundas
- PowerPlex® Fusion System
- Hi-Di™ formamide
- Pop-4
- Cubrebocas
- Micropipeta multicanal Eppendorf
- Congelador Revco
- Refrigerador Revco
- Alcohol etílico al 70%
- Encendedor

6.5.2 Método

Obtención de la muestra

Se pidió a cada donante que rotulara una tarjeta FTA, posteriormente se limpió el área de punción para la obtención de sangre con una torunda con alcohol al 70%. Se Puncionó con una lanceta y se depositaron aproximadamente 5 gotas de sangre sobre la tarjeta FTA

Amplificación

El panel utilizado fue el PowerPlex® Fusion System, el Kit se mantuvo en congelación hasta el momento de su uso.

Se determinó el número de reacciones incluyendo el control positivo y negativo, se preparó la mezcla de reacción en un tubo eppendorf (dando vortex por 15 segundos a cada uno de los tubos de reactivo).

La mezcla se agitó y se fraccionó en una placa de 96 pozos, de acuerdo al número de muestras a amplificar, posteriormente se cortó un disco de papel FTA con muestra con ayuda del perforador Mini-Punch Harris de 1.2 mm de diámetro y se colocó el disco en el pozo correspondiente.

La amplificación se llevó a cabo en el termociclador GENEAMP® SYSTEM 9700 siguiendo el protocolo recomendado por el fabricante, el cual se muestra a continuación:

Protocolo de Amplificación

96°C por 1 minuto

94°C por 10 segundos

59°C por 1 minuto

72°C por 30 segundos

Por 25 ciclos, después:

60°C por 20 minutos

4°C (Se conserva a 4°C hasta el momento de su análisis).

Electroforesis Capilar

El Kit se conservó en el congelador hasta el momento de su uso. Se preparó una mezcla de Formamida e ILS en un tubo eppendorf, las cantidades de cada reactivo fueron de acuerdo al número de muestras, la proporción de cada reactivo se calculó de la siguiente manera:

$[(1.0\mu\text{l CC5 ILS 500}) \times (\# \text{muestras})] + [(10.0\mu\text{l Hi-Di}^{\text{TM}} \text{ formamida}) \times (\# \text{muestras})]$

La mezcla se agitó durante 15 segundos y enseguida se fraccionó en una placa de 96 pozos. Posteriormente se colocó 1.5 μL de cada producto de PCR en el pozo correspondiente y 1 μL of PowerPlex® Fusion Allelic Ladder Mix en cada uno de los pozos destinados para este. La placa se selló con una septa y se procedió a realizar la hoja de corrida en el software Data Collection, y comenzar con el corrimiento electroforético.

Los resultados fueron analizados con ayuda del software GeneMapper ID V.3.2

Análisis Estadístico

En el presente estudio, además de analizar el potencial del panel PowerPlex Fusion, se realizó la base de datos de la Ciudad de México; con los datos obtenidos, y mediante el método de conteo de genes se determinaron las frecuencias alélicas. Una vez obtenidos los datos de las frecuencias alélicas se aplicó la prueba de bondad de ajuste (chi cuadrada) para determinar si los loci se encontraban dentro del equilibrio de Hardy-Weinberg. Adicionalmente se determinaron los parámetros estadísticos: poder de discriminación, poder de exclusión, índice de contenido polimórfico, índice de paternidad, probabilidad de coincidencia aleatoria, número efectivo de alelos, heterocigosidad y homocigosidad.

Además se compararon los parámetros con paneles anteriores para determinar el avance o las diferencias que tiene entre sí.

7. Resultados

Descripción de resultados

En la figura 7.1 se observa un electroferograma obtenido de una de las muestras analizadas, en él se encuentran los 22 marcadores autosómicos polimórficos, un marcador de cromosoma Y y amelogenina (para determinar el sexo). En cada locus se observan picos que corresponden a los alelos. Los dos alelos (provenientes uno de origen paterno y el otro de origen materno) pueden ser iguales (homocigoto) o diferentes (heterocigoto).

En la tabla 7.1 se muestran las frecuencias obtenidas de los 22 marcadores STR. Se observan los diferentes valores obtenidos de las frecuencias alélicas, así como también los alelos de cada marcador, algunos marcadores exhiben más alelos como el Penta E. Estas frecuencias son utilizadas para calcular diversos parámetros que son de importancia en la genética forense.

En la tabla 7.2 se observa que 21 de los 22 marcadores se encuentra en equilibrio de Hardy-Weinberg, lo que nos indica que las frecuencias alélicas obtenidas en estos marcadores se pueden utilizar para el cálculo de diversos parámetros del área forense. El marcador D19S433, estuvo en desequilibrio, pero en un estudio ulterior se aplicó la corrección de Bonferroni, con lo cual este marcador resulto en equilibrio de Hardy-Weinberg.⁽⁷¹⁾

En la tabla 7.3 se muestran los valores obtenidos de manera individual de diversos parámetros de los 22 marcadores STR. Como se observa el marcador Penta E fue el que tuvo índices más altos, por lo que la presencia de este locus en el panel proporciona mucho peso estadístico al momento de realizar los cálculos de una confronta de perfiles.

En la figura 7.2 se muestran los valores obtenidos de la Heterocigosidad, se observa que el marcador Penta E es el que presenta la mayor Heterocigosidad, aportando un mayor poder de discriminación al panel. Al contrario, el marcador D22S1045 presenta el valor más bajo

La figura 7.3 muestra los valores de la probabilidad de coincidencia aleatoria, el cual indica la probabilidad de que dos individuos seleccionados al azar tengan un genotipo idéntico en el locus probado. Se observa que el marcador Penta E es el que posee el valor más bajo, haciendo que disminuyan las asociaciones fortuitas.

Así mismo en la gráfica 7.4 se observa que todos los marcadores del panel presentaron un valor alto de discriminación, los cuales al combinarse dan como resultado un enorme poder de discriminación, de tal manera que dos personas puedan ser fácilmente diferenciados genéticamente, al disminuir al mínimo la posibilidad de que existan asociaciones fortuitas.

En la figura 7.5 se presenta los valores obtenidos del poder de exclusión, el cual nos indica la posibilidad de excluir a una persona falsamente imputada. Al igual que en parámetros anteriores, el Penta E, fue quien tuvo el valor más alto, comprobando una vez más la utilidad de este marcador para el panel.

En la imagen 7.6 se exhiben los valores obtenidos del índice de contenido polimórfico, el cual expresa el grado de diversidad biológica que posee cada marcador. El locus Penta E, es el que posee un mayor polimorfismo, y el que ofrece una mayor información. Como se puede observar los valores entre los marcadores son muy similares.

En la figura 7.7 se muestran los valores del índice de paternidad. Este valor apoya la hipótesis de que el individuo analizado sea el verdadero padre biológico del menor en vez de otro individuo no relacionado de la población. En este parámetro también el Penta E fue quien presentó el valor más alto.

En la figura 7.8 se observan los valores obtenidos del número efectivo de alelos. Este valor hace referencia a los alelos con capacidad de pasar a la siguiente generación, y los cuales son necesarios para que exista una diversidad biológica. Se observa que el marcador Penta E es que presenta el valor más alto, comprobando una vez más su gran utilidad dentro del panel.

En la tabla 7.4 se muestran los valores combinados de los 22 marcadores autosómicos polimórficos del panel PowerPlex Fusion. Se observa que el panel en general tiene una heterocigosidad del 80%, así como un alto poder de discriminación, lo cual hace a este panel una gran herramienta para el sistema de justicia mexicano.

Finalmente en la tabla 7.5 se muestra la comparación de los parámetros estadísticos Heterocigosidad, Poder de Discriminación, Poder de Exclusión, Probabilidad de Coincidencia Aleatoria e Índice de Paternidad entre el panel PowerPlex® Fusion, PowerPlex® 21 y PowerPlex® 16. Se observa que el panel PowerPlex® Fusion presenta índices más altos que sus antecesores.

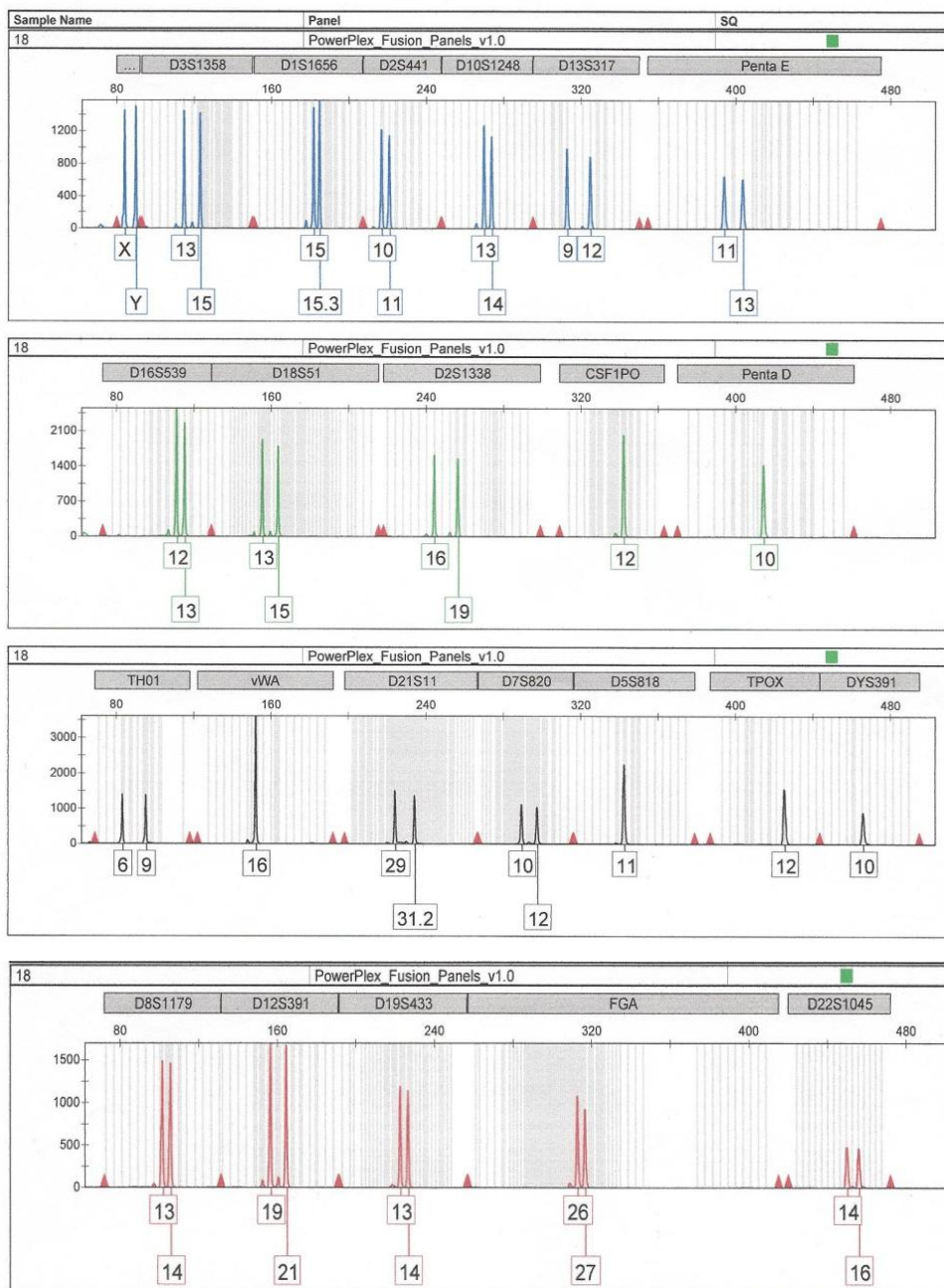


Figura 7.1. Electroferograma.

Para la obtención de las frecuencias alélicas se utilizó la siguiente fórmula:

$$\text{Frec. alélicas} = \frac{\text{No. de veces que fue observado el alelo}}{\text{Total de alelos}}$$

Tabla 7.1. Frecuencias alélicas obtenidas de los 22 Loci del Panel PowerPlex Fusion

LOCUS ALELO	D3S1338	D1S1656	D2S441	D10S1248	D13S317	PENTA E	D16S539	D18S51	D2S1338	CSF1PO	PENTA D	TH01	Vwa	D21S11	D7S820	D5S818	TPOX	D8S1179	D12S391	D19S433	FGA	D22S1045
2.2											0.01											
5						0.0325					0.0025											
6												0.3125										
7						0.055					0.0075	0.3575			0.005	0.065						
8					0.075	0.0275	0.01			0.005	0.0125	0.045			0.105		0.535	0.0125				
8.3												0.0025										
9			0.005	0.005	0.2425	0.005	0.0975			0.0175	0.195	0.06			0.0375	0.06	0.06	0.005				
9.3												0.2175										
10			0.4675		0.1075	0.035	0.2			0.2525	0.245	0.005			0.2575	0.035	0.03	0.1375		0.0025		
11		0.0325	0.295	0.0025	0.1725	0.0425	0.2875	0.0025		0.2875	0.1225				0.315	0.51	0.2375	0.0475		0.01		0.04
11.2																				0.0025		
11.3			0.025																			
11.4						0.005																
12		0.065	0.04	0.0325	0.2075	0.16	0.28	0.1325		0.3775	0.1825				0.2325	0.2525	0.1375	0.1125		0.05		0.005
12.1						0.0025																
12.2																				0.015		
13	0.01	0.0925	0.01	0.23	0.1275	0.0575	0.1125	0.0975		0.055	0.165				0.045	0.075		0.2825		0.2175		0.005
13.2																				0.115		
14	0.065	0.1175	0.135	0.415	0.0675	0.095	0.01	0.1725		0.005	0.055		0.055		0.0025	0.0025		0.23		0.2375		0.0325
14.2																				0.0375		
14.3		0.0025																				
15	0.455	0.1375	0.02	0.21		0.125	0.0025	0.17			0.0025		0.06					0.14	0.005	0.1475		0.52
15.2	0.0025																			0.0975		
15.3		0.03																				
16	0.2925	0.2125	0.0025	0.085		0.105		0.105	0.0325				0.385					0.0275	0.0325	0.0425		0.35
16.2																				0.0225		
16.3		0.04																				
16.4						0.005																
17	0.1325	0.065		0.02		0.095		0.185	0.1725				0.2875					0.005	0.06	0.0025	0.0025	0.0425
17.3		0.1575																	0.005			
17.4						0.0025																
18	0.04	0.0025				0.0225		0.05	0.04				0.175							0.165		0.0125
18.3		0.04																		0.0075		0.0025
19	0.0025					0.0325		0.03	0.2325				0.035							0.3025		0.0725
19.3		0.005																		0.0175		0.0025
20						0.0425		0.015	0.1225				0.0025							0.23		0.085
21						0.025		0.02	0.0375											0.07		0.0875
22						0.02		0.01	0.08											0.05		0.15
22.2																						0.0025
23						0.005		0.005	0.2075											0.0275		0.135
24						0.0025		0.0025	0.0325											0.0175		0.155
25								0.0025	0.03											0.0075		0.1425
26									0.0075											0.0025		0.105
27									0.005						0.0075							0.0425
28															0.09							0.005
29															0.1725							
29.2															0.005							
30															0.285							0.0025
30.2															0.01							
31															0.0675							
31.2															0.1275							
32															0.025							
32.2															0.1375							
33.1															0.0025							
33.2															0.065							
34.2															0.005							

Con las frecuencias alélicas se aplicó la prueba de chi cuadrada para observar si los loci se encontraban en equilibrio de Hardy-Weinberg, teniendo resultados de la tabla 7.2:

Tabla 7.2. Resultados de la prueba de X^2 del sistema PowerPlex® Fusion

LOCUS	X^2 CALCULADA	X^2 TABLAS ($\alpha=0.05$)	RESULTADO
D3S1358	2.2688	5.991	EQUILIBRIO
D1S1656	8.3912	16.919	EQUILIBRIO
D2S441	5.6324	5.991	EQUILIBRIO
D10S1248	1.2311	5.991	EQUILIBRIO
D13S317	6.4105	11.07	EQUILIBRIO
PENTA E	12.3258	16.919	EQUILIBRIO
D16S539	5.2684	11.07	EQUILIBRIO
D18S51	11.2583	16.919	EQUILIBRIO
D2S1338	4.8988	11.07	EQUILIBRIO
CSF1PO	1.8181	5.991	EQUILIBRIO
PENTA D	8.7925	11.07	EQUILIBRIO
TH01	4.2692	5.991	EQUILIBRIO
vWA	8.1320	11.07	EQUILIBRIO
D21S11	4.1221	11.07	EQUILIBRIO
D7S820	3.9262	5.991	EQUILIBRIO
D5S818	4.6719	5.991	EQUILIBRIO
TPOX	3.1679	5.991	EQUILIBRIO
D8S1179	9.4771	16.919	EQUILIBRIO
D12S391	7.1826	11.07	EQUILIBRIO
D19S433	12.2282	11.07	DESEQUILIBRIO*
FGA	13.2356	16.919	EQUILIBRIO
D22S1045	5.0588	11.07	EQUILIBRIO

Tabla 7.3 Parámetros estadísticos calculados.

	D3S135 8	D1S165 6	D2S44 1	D10S124 8	D13S31 7	PENTA E	D16S53 9	D18S5 1	D2S133 8	CSF1P O	PENT A D	vWA	TH01	D21S1 1	D7S82 0	D5S81 8	TPOX	D8S117 9	D12S39 1	D19S43 3	FGA	D22S104 5
PD	0.8562	0.9698	0.8478	0.8811	0.9467	0.9816	0.9175	0.9643	0.9557	0.8562	0.9458	0.891	0.852 9	0.9508	0.907	0.8393	0.814 2	0.9343	0.9321	0.9511	0.967 8	0.7864
PE	0.3691	0.7346	0.3031	0.4283	0.6177	0.857	0.4928	0.7346	0.7246	0.4928	0.5446	0.451 8	0.562 5	0.7246	0.5535	0.4518	0.376 2	0.6558	0.6367	0.7346	0.836 4	0.2509
PC	0.1438	0.0303	0.1522	0.1189	0.0533	0.0184	0.0825	0.0357	0.0443	0.1438	0.0542	0.109	0.147 1	0.0493	0.093	0.1607	0.185 9	0.0657	0.068	0.049	0.032 2	0.2136
PIC	0.6345	0.8628	0.6225	0.6793	0.8083	0.9078	0.7415	0.85	0.8277	0.6535	0.7989	0.687 8	0.672 2	0.8175	0.7284	0.6186	0.585 5	0.788	0.7924	0.8275	0.869 9	0.5328
Ho	0.66	0.87	0.61	0.7	0.81	0.93	0.74	0.87	0.865	0.74	0.77	0.715	0.78	0.865	0.775	0.715	0.665	0.83	0.82	0.87	0.92	0.565
He	0.6856	0.8773	0.6751	0.7239	0.8325	0.9163	0.7785	0.867	0.8477	0.7095	0.8251	0.732 5	0.723 4	0.8382	0.7678	0.6631	0.635 6	0.815	0.8168	0.8474	0.884	0.6041
NEA	31.638	8.008	30.622	35.979	58.958	116.29 6	44.758	74.002	64.777	34.211	56.517	35.91 6	37.12 5	61.027	42.721	29.539	27.31 9	53.451	53.988	64.615	84.58 4	25.162
IP	15.819	4.004	15.311	17.989	29.479	58.148	22.379	37.001	32.388	17.106	28.258	17.95 8	18.56 2	30.513	23.161	14.769	1.366	26.725	26.994	32.308	42.29 2	12.581

Ho: Heterocigosidad Observada, He: Heterocigosidad Esperada, PD: Poder de Discriminación, PE: Poder de Exclusión, PIC: Índice de Contenido Polimórfico, PC: Probabilidad de Coincidencia, NEA: Número Efectivo de Alelos, IP: Índice de Paternidad

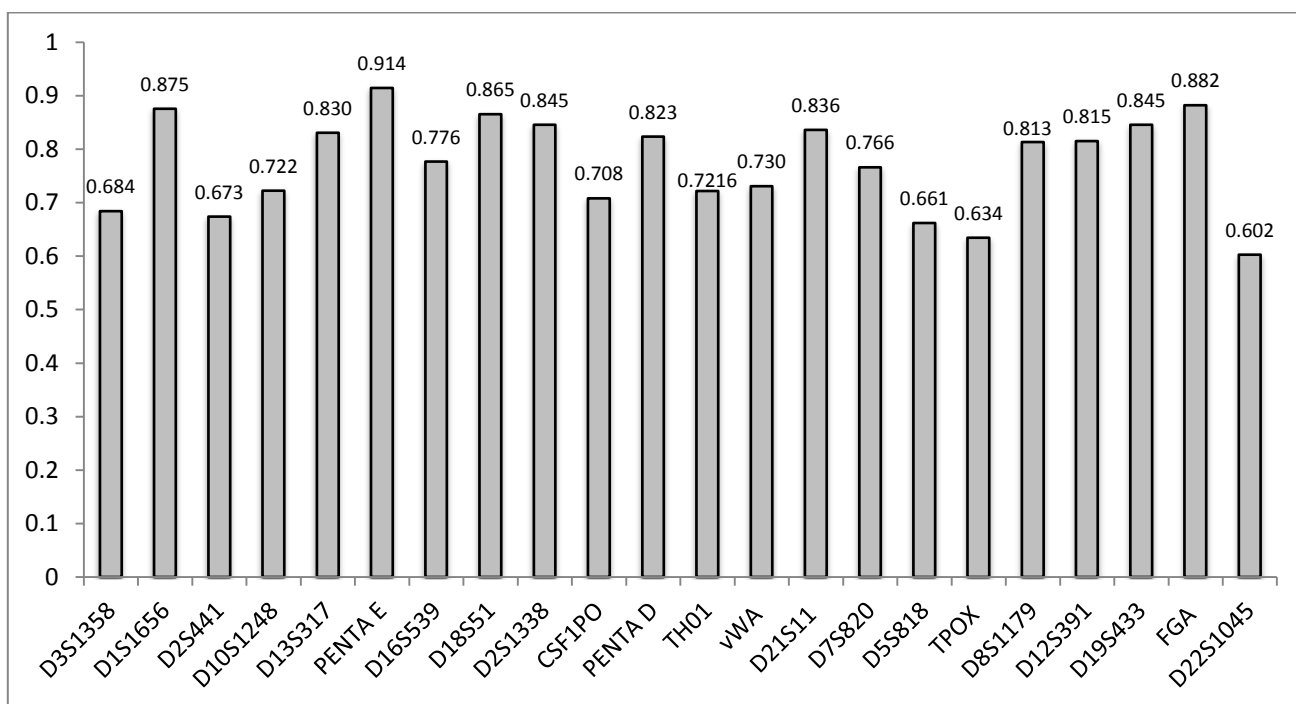


Figura 7.2. Heterocigosidad.

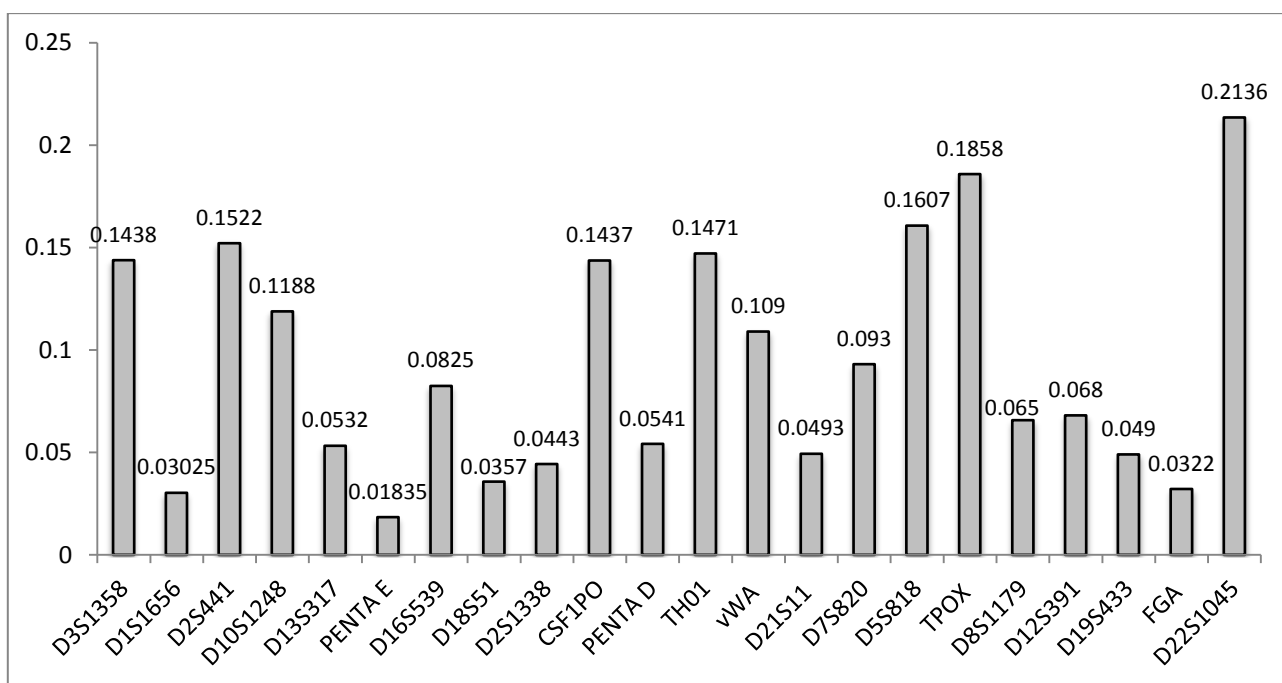


Figura 7.3. Probabilidad de Coincidencia Aleatoria.

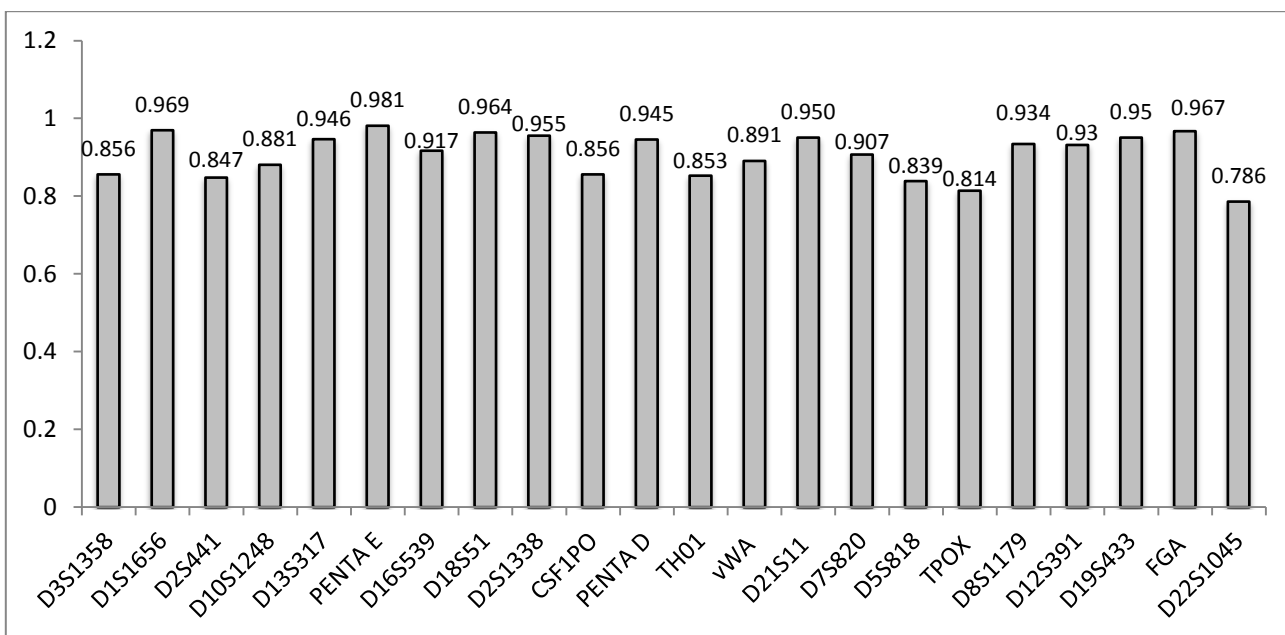


Figura 7.4. Poder de discriminación.

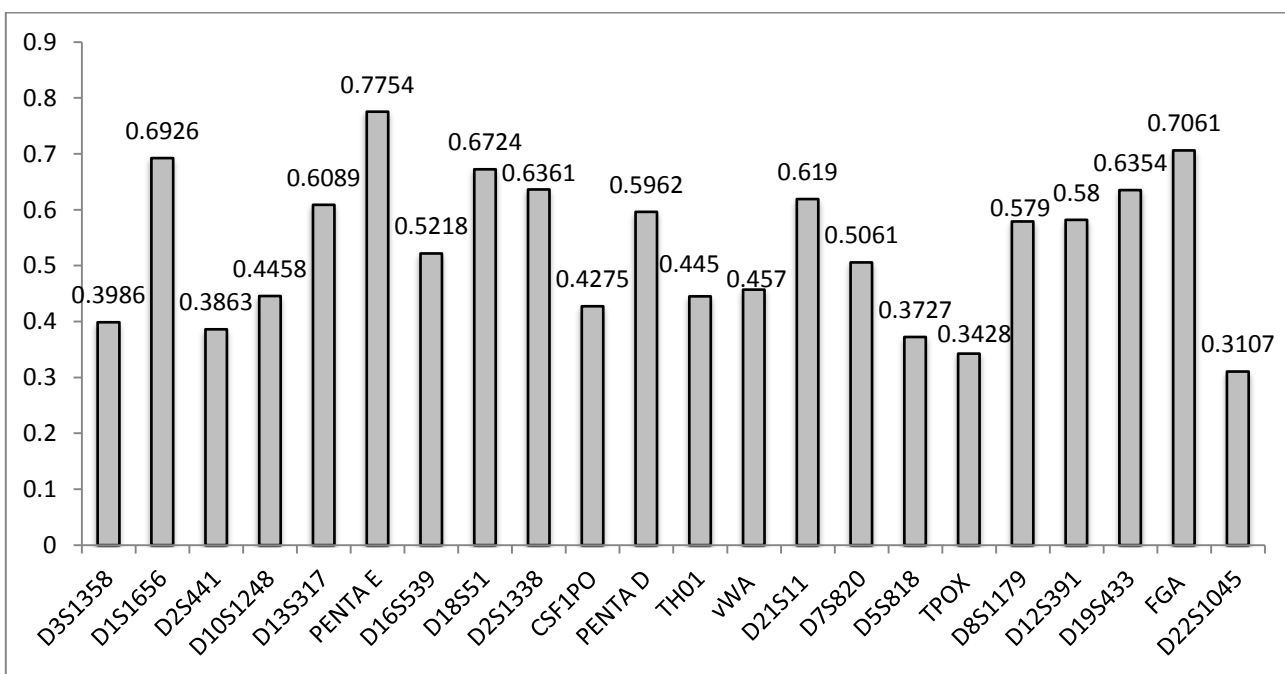


Figura 7.5. Poder de exclusión.

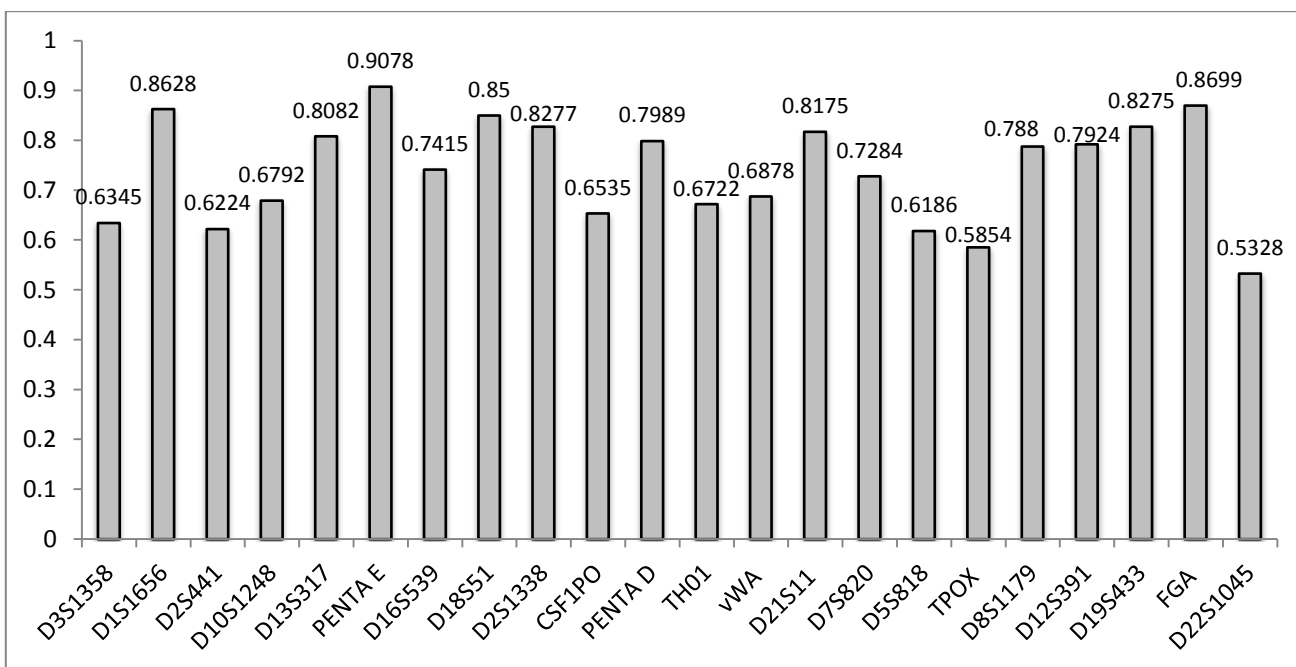


Figura 7.6. Índice de contenido polimórfico.

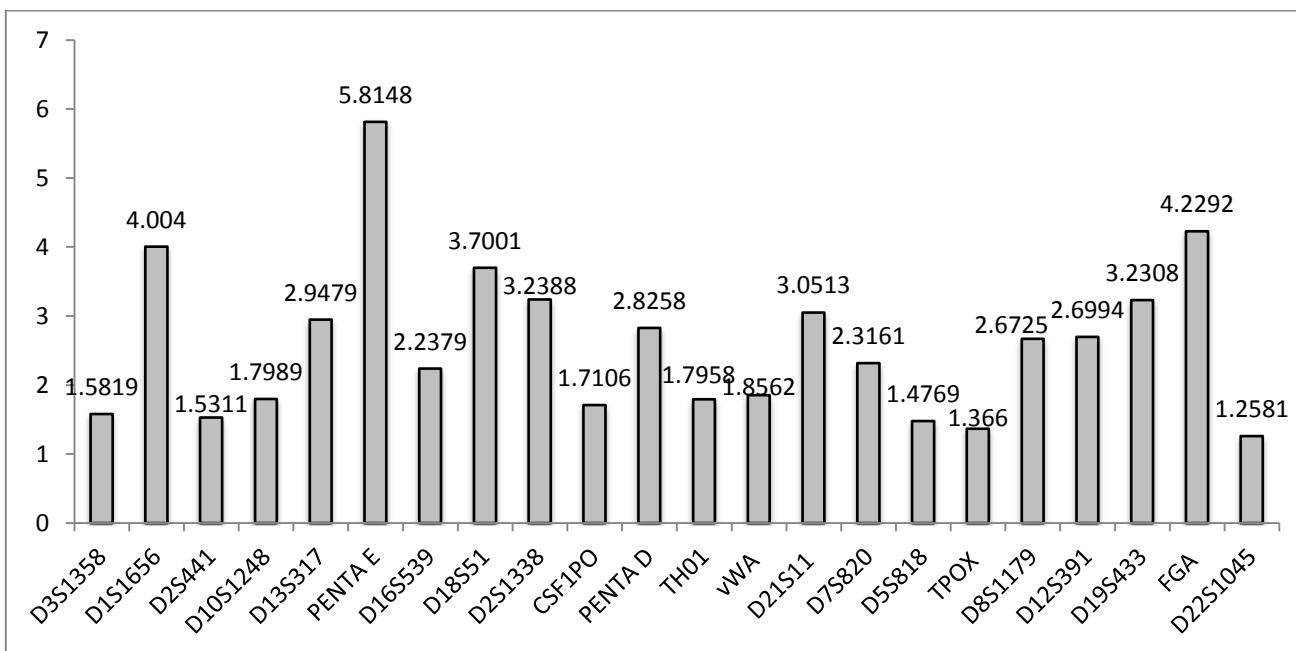


Figura 7.7. Índice de paternidad.

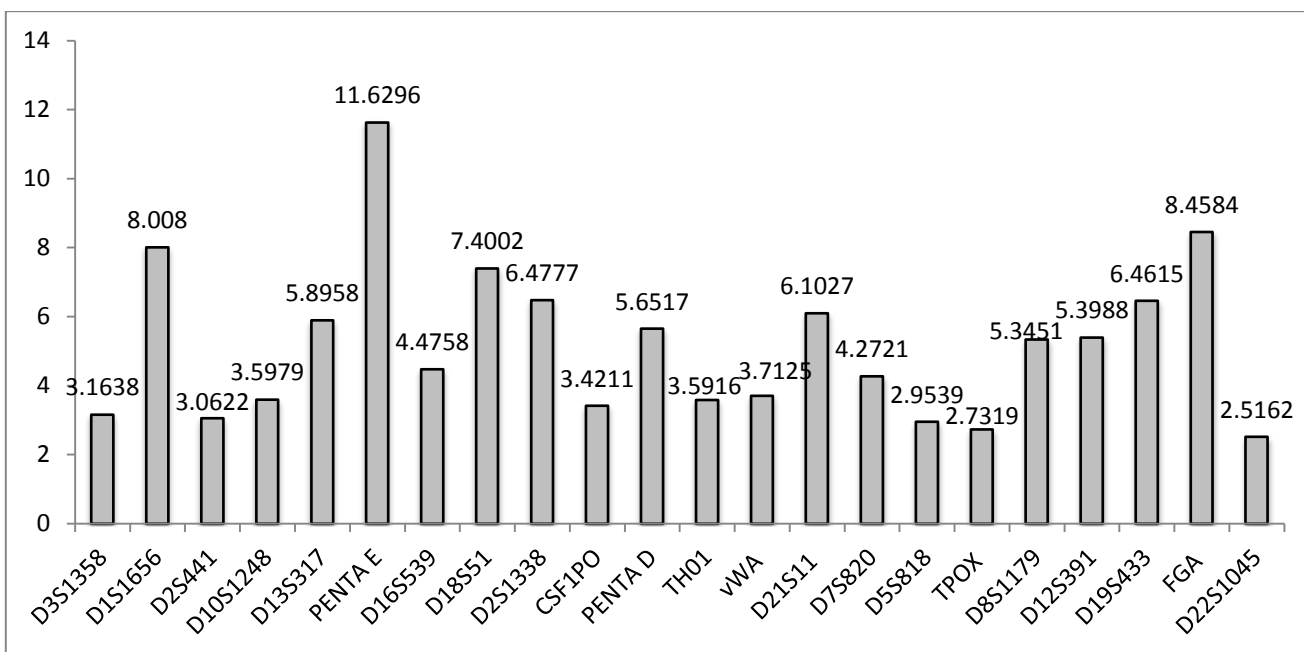


Figura 7.8. Número efectivo de alelos.

Tabla 7.4. Resultados de los índices combinados del panel PowerPlex Fusion.

Índice	Valor
Het	0.8
PD	3.88×10^{24}
PE	0.99999998
Probabilidad de coincidencia aleatoria	2.57396×10^{-25}
Índice de Paternidad	228,483,328.8

Het: Heterocigosidad, PD: Poder de Discriminación, PE: Poder de Exclusión

Tabla 7.5. Comparación de índices obtenidos del panel PowerPlex Fusion con paneles anteriores como el PowerPlex 21 y PowerPlex 16.

	PowerPlex Fusion (Cd de México)	PowerPlex 21 (Cd. De México)	PowerPlex 16 (Hispanos)
No. Marcadores	22	20	15
Het	0.8	0.8	No se cuenta con el dato
PD	3.88x10²⁴	2.698x10²¹	2.94x10¹⁷
PE	0.99999998	0.99999997	0.9999983
Probabilidad de coincidencia	2.57396E-25	3.706E-22	2.82E-19
Índice de Paternidad	228,483,328.8	No se cuenta con el dato	5,220,000

Het: Heterocigosidad, PD: Poder de Discriminación, PE: Poder de Exclusión

En la tabla se muestra la comparación de los parámetros estadísticos Heterocigosidad, Poder de Discriminación, Poder de Exclusión, Probabilidad de Coincidencia Aleatoria e Índice de Paternidad entre el panel PowerPlex® Fusion, PowerPlex® 21 y PowerPlex® 16. Se observa que el panel PowerPlex® Fusion presenta índices más altos que sus antecesores.

8. Discusión de resultados

Debido a que en los últimos años se ha incrementado la aplicación de la genética forense para el esclarecimiento de problemas de índole civil y penal, mismos que impactan de manera importante la vida de las personas involucradas, es imperativo contar con los métodos apropiados para resolver correcta y oportunamente las controversias.⁽³⁾ Gran parte de las conclusiones de un dictamen en materia de genética forense recaen en el tratamiento estadístico de los datos obtenidos de la evidencia, del probable responsable y de la víctima; así como del probable padre y del menor. Para conocer qué tan raro es un perfil genético en la población de estudio, se requiere de la base de datos de las frecuencias alélicas de dicha población. Es decir, la base de datos es uno de los pilares del análisis genético para establecer el peso estadístico de los perfiles genéticos analizados. También es necesario analizar un mayor número de marcadores con el fin de aumentar el poder de discriminación y por lo tanto, también aumenten los índices estadísticos a un nivel de certidumbre científica razonable.

Para que las frecuencias alélicas que se obtengan de la población puedan ser utilizadas en el análisis genético, es necesario que estén en equilibrio de Hardy-Weinberg. Esta ley define el estado genético de una población mendeliana en ausencia de toda fuerza evolutiva, y además relaciona las frecuencias alélicas con las frecuencias genotípicas, de manera que se puedan utilizar las primeras para calcular el valor de las segundas, usando la fórmula p^2 si se trata de individuos homocigotos y $2pq$ si se trata de individuos heterocigotos.⁽¹⁾

En estos estudios poblacionales se utilizan las frecuencias alélicas en vez de las genotípicas, debido a que es prácticamente imposible tener representados a todos los genotipos de una población, además de que no es costeable realizar un estudio de esta magnitud, considerando el número de habitantes de la ciudad de México, en lugar de eso resulta más conveniente tener representados a los alelos y con base en sus frecuencias determinar la frecuencia de cualquier genotipo encontrado.

El equilibrio de Hardy-Weinberg nos indica que si las frecuencias genotípicas observadas son iguales o muy cercanas a las frecuencias genotípicas esperadas (es decir, que no exista diferencia significativa), entonces la población está en equilibrio y por lo tanto se podrán utilizar las frecuencias alélicas para predecir las frecuencias genotípicas en las siguientes generaciones, y con base a estas frecuencias calcular diversos parámetros estadísticos al momento de confrontar dos perfiles y definir o identificar a la fuente de este perfil. Esto es posible ya que al estar en equilibrio dicha población sus frecuencias alélicas serán constantes a través del tiempo y podrán ser utilizadas en las generaciones siguientes.

De los 22 loci analizados, 21 estuvieron en equilibrio de Hardy-Weinberg, el locus D19S433 resultó en desequilibrio (Tabla 2), pero en un estudio ulterior⁷¹ se aplicó la corrección de Bonferroni, con lo cual este locus también estuvo en equilibrio. Este desequilibrio exhibido se puede deber a que se hizo una comparación múltiple al fundirse las celdas de las frecuencias esperadas de los alelos menos frecuentes o bien a que no todos los alelos pudieron estar bien representados.

En cuanto a la heterocigosidad del panel, esta fue del 80%, una alta heterocigosidad significa que existe más diversidad alélica, por lo que habrá un mayor número de genotipos posibles, aumentando el poder de discriminación y disminuyendo la probabilidad de coincidencia en muestras. En cuanto a la heterocigosidad individual, el marcador Penta E tuvo el valor más alto, lo cual lo hace uno de los marcadores que proporciona un mayor poder de discriminación al panel, al contrario el marcador D22S1045 presenta el valor más bajo. Sin embargo dado que el marcador D22S1045 es un miniSTR, se obtienen fragmentos muy pequeños, siendo muy útil cuando se trabaja con muestras severamente comprometidas o dañadas. El marcador Penta E sería equiparable al FGA en poblaciones de Guatemala en cuanto a su Heterocigosidad y su poder de discriminación.⁽⁵¹⁾ En otro estudio reportado para población mestiza del occidente de México, se estudiaron 5 marcadores STR, los cuales estuvieron en equilibrio de Hardy Weinberg, uno de esos marcadores era el Penta E, el cual también exhibió el índice más alto de Heterocigosidad.⁽⁵³⁾

El poder de discriminación indica la probabilidad de que dos individuos no relacionados sean diferenciados genéticamente. El valor de este parámetro de todo el panel fue de 3.88×10^{24} , con lo cual se puede decir que un genotipo analizado con este panel puede ser encontrado uno en 3.8 cuatrillones de individuos (cuando la población mundial es de siete mil millones (7×10^9)).

Otro parámetro estudiado fue el poder de exclusión, que nos indica con que probabilidad un individuo falsamente imputado puede ser excluido. Con este poder de exclusión los resultados son emitidos con un 99.999998%, de certeza estadística, por lo que con este panel se disminuye la probabilidad de que existan asociaciones fortuitas.

El índice de contenido polimórfico, expresa el grado de diversidad biológica que posee cada locus, así como la capacidad para conocer el origen parental de cada uno de los alelos en la descendencia. Como se observa en la gráfica 5 el locus más polimórfico fue el Penta E, con lo cual es uno de los locus que son de gran utilidad, ya que al tener un número grande de alelos bien distribuidos, las frecuencias de éstos son pequeñas, y por lo tanto aumentan el poder de discriminación.

La probabilidad de coincidencia aleatoria es la probabilidad de que dos individuos seleccionados al azar tengan un genotipo idéntico, mientras menor es el valor de este parámetro existen menos posibilidades de que existan asociaciones fortuitas. Al igual que en otros parámetros, el marcador Penta E fue el que tuvo el menor valor. La probabilidad de coincidencia aleatoria combinada del panel fue de 2.57396×10^{-25} , lo cual se puede traducir en que la posibilidad de que dos individuos analizados tengan el mismo genotipo es de 2.57396×10^{-25} , lo cual es un valor extremadamente pequeño lo que hace del panel una gran herramienta para proporcionar resultados exactos.

El índice de paternidad es el valor que apoya la hipótesis de que el individuo examinado sea el verdadero padre biológico del menor, frente a que si fuera otro individuo no relacionado biológicamente con el hombre señalado de la población de estudio, el índice combinado fue de 228,483,328.8. En este parámetro, el marcador Penta E exhibió el valor más alto (Gráfica 6), y por lo tanto el que aporta mayor peso a la hipótesis que apoya que el hombre examinado sea el verdadero padre biológico del menor.

El número efectivo de alelos, es el número de alelos igualmente frecuentes que se necesitaría para alcanzar un determinado grado de diversidad genética y que pasan a la siguiente generación. Este valor disminuye al incrementar la frecuencia de solo uno de los alelos. Si todos los alelos presentes en un locus tuvieran la misma frecuencia alélica, el número efectivo de alelos se podría utilizar para inferir el nivel de heterocigosidad esperada. Como se observa en la gráfica 7 el locus Penta E, nuevamente exhibió el valor más alto en este parámetro, lo que nos indica que es el que posee mayor polimorfismo, y por lo tanto como se mencionó anteriormente uno de los marcadores más importantes del panel en términos de identificación humana y análisis de parentesco. El marcador con el valor más bajo fue el D22S1045, con lo cual podemos observar que las frecuencias que contiene este marcador no son equilibradas, hay alelos que exhiben un valor más alto que otros, aun así este marcador en conjunto con los demás nos proporciona un poder de discriminación muy elevado.

Algunos de los marcadores analizados del panel, también han sido analizados en otras poblaciones, por ejemplo en Colombia, la conclusión de estos estudios reportados, es que estos marcadores (CSF1PO, TH01, TPOX, D16S539, D7S820, D13S317, vWA, D8S1179, D5S818, PENTA E, D18S51 y D3S1358.) ofrecen un alto grado de confiabilidad para la identificación humana.⁽⁵³⁾

Los parámetros obtenidos se compararon con los valores obtenidos de paneles anteriores, como se puede ver en la tabla 5, los índices que presenta el panel PowerPlex Fusion es mayor que el PowerPlex 16 HS y que el PowerPlex 21, haciendo de este panel una gran herramienta en el área forense.

Actualmente no se han reportado estudios similares utilizando el panel PowerPlex Fusion, ni tampoco realizados en la Ciudad de México, pero si existen estudios poblacionales utilizando paneles con menos marcadores y estudiando otras poblaciones. El estudio de estas publicaciones se realizó de manera muy similar al presentado en este trabajo, y se obtuvieron conclusiones parecidas, por lo que este estudio en particular es de utilidad, ya que los datos obtenidos pueden ser utilizados para el cálculo de diversos parámetros al momento de la confronta de perfiles genéticos.

9.- Conclusiones

Las frecuencias alélicas de los 22 marcadores STR del panel PowerPlex® Fusion se encuentran en equilibrio de Hardy-Weinberg, lo cual fue comprobado mediante la prueba de chi cuadrada.

La heterocigosidad del panel que fue del 80%, en combinación con su poder de exclusión del 99.999997%, poder de discriminación de 3.88×10^{24} y de su probabilidad de coincidencia aleatoria del 2.57396×10^{-25} , hacen del panel PowerPlex® Fusion una gran herramienta para el sistema de justicia en México

Además el panel PowerPlex® Fusion contiene al locus Penta E, que es uno de los marcadores que aportan mucho peso estadístico en los parámetros calculados, siendo uno de los más importantes del panel en términos de la identificación humana y pruebas de parentesco.

De acuerdo con estos parámetros estadísticos, los datos obtenidos pueden ser útiles para la investigación genética de poblaciones, para la identificación humana y para pruebas de parentesco.

El Sistema PowerPlex Fusion reúne las características que lo hacen ser buen candidato para lidiar con las nuevas demandas de identificación humana que impone la situación actual de nuestro país.

10.- Recomendaciones

El incremento de la violencia en el país ha cobrado la vida de muchas personas, de manera que cada vez se ha intensificado el uso de la genética forense. Los perfiles genéticos obtenidos de las víctimas son almacenados en bases de datos con el fin de establecer la confronta con muestras de referencia de familiares de la persona desaparecida para identificar potenciales relaciones de parentesco. A medida que crece la base de datos, las asociaciones fortuitas se hacen más frecuentes, fenómeno que se busca eliminar mediante el estudio de objetos personales del desaparecido, el estudio de otros miembros de su familia, el análisis de DNAm y cromosoma Y o bien, utilizando paneles autosómicos con un mayor número de loci.

Es por eso que se recomienda en lo posible, realizar los estudios de genética con paneles que posean un número mayor de marcadores, para que de esta manera incremente el poder de discriminación y disminuyan la probabilidad de una coincidencia aleatoria.

Además se recomienda que la base de datos poblacional se realice para cada población o estado de la República, considerando las variaciones genéticas que pudieran existir en determinada región y para obtener resultados robustos y confiables.

11. Anexos

DECLARACIÓN DE CONSENTIMIENTO INFORMADO

C. _____ de _____ años de edad y que es habitante de la Ciudad de México; manifiesta que ha sido informado/a sobre los beneficios que podría suponer la extracción de sangre en papel FTA para cubrir los objetivos del Proyecto de Investigación titulado "Análisis genético de 22 Loci STR en la Población de la Ciudad de México" con el fin de mejorar el análisis forense en el sistema de justicia mexicano.

He sido también informado/a de que mis datos personales serán protegidos y no se utilizarán con ningún otro fin que no sea el de la investigación, antes mencionada.

Tomando ello en consideración, OTORGO mi CONSENTIMIENTO a que esta extracción tenga lugar y sea utilizada para cubrir los objetivos especificados en el proyecto.

México DF, 2013

Donante

12. -Referencias bibliográficas

1. - Butler JM. Forensic DNA Typing: biology, technology and genetics of STR markers. 2a Ed. New York: ELSEVIER, 2005.
- 2.- Berumen CJ. El análisis del ADN en la Identificación de individuos. México: Ciencia y Desarrollo, 1993.
- 3.- Lorente JA., Vega ML, Rosas GO. Genética Forense: la Ciencia al Servicio de la Justicia. Mens Bioquím 2007 (31).
- 4.- Álvarez MJ, Martínez LJ, Saiz M, Álvarez JC, Lorente JA. Nuevas aplicaciones en identificación genética. Cuad Med Foren 2010; 16 (1-2): 5-18.
5. - Butler JM. Genetics and Genomics of Core Short Tandem Repeat Loci Used in Human Identity Testing. J Forensic Sci 2006; 51 (2).
6. - Gill P, Federay L, Morling N, Schneider PM. The evolution of DNA databases-Recommendations for new European STR loci. Forensic Sci Int 2006; 156: 242-244.
7. - Griffiths JF, Miller JH, Suzuki DT, Lewontin RC, Gelbart WM. Genética. 7° ed. McGRAW-HILL-Interamericana.
- 8.- Klug SW, Cummings MR, Spencer CA. Conceptos de Genética. 8° ed. España: Prentice Hall, 2008.
- 9.- Ácidos nucleicos. [En línea]. Disponible en: <<http://toxamb.pharmacy.arizona.edu/c1-1-1-3.html>>. [Consulta: 10-05-2013]
- 10.- Passarge E. Color Atlas of Genetics. 2° ed. New York: Ed Thieme, 2001.
11. - Watson JD, Crick FH. Molecular Structure of nucleic acids: a structure for deoxirribosenucleid Acid. Nature 1953; 171: 737-738
- 12.- Guías didácticas para el desarrollo de experimentos. [En línea]. Disponible en: <<http://www.eez.csic.es/files/GuiasDidacticas.pdf>> [Consulta: 02-06-2013]
- 13.- Lewin B. Genes. 2° ed. Barcelona: Reverte, 1996.
- 14.- D' Robertis. Biología molecular. 13° ed. Argentina: El ateneo, 2000.
- 15.- Luque J, Herraes A. Biología molecular e ingeniería genética. USA: Harcourt, 2001.
16. - Nussbaum RL, McInnes RR, Willard HF. Genética en Medicina. 7° ed. España: Elsevier, 2008.
17. - Vera LF. Doble hélice, genes y cromosomas. Rev. R. Acad Cienc Exact Fis Nat 2003; 97 (2): 203-222.
18. - Griffith F. The serological classification on *Streptococcus pyogenes*. MS 1934;10 (11).
19. - Oswald TA., MacLeod M, McCarty M. Studies on the chemical nature of the substance inducing transformation on pneumococcal types. J of Exp Med. 1944; 79.
- 20.- Hershey AD, Chase M. Independent functions of viral protein and nucleic acid in growth of bacteriophage. J of Gral Phys 1952.

- 21.- Procesos Genéticos de la síntesis de proteínas: La transcripción. [En línea]. Disponible en: <http://pendientedemigracion.ucm.es/info/genetica/grupod/Transcripcion/Transcripcion.htm>. [Consulta: 14-06-2013]
22. - Crick F. Central Dogma of Molecular Biology. Nature: 1970;227.
- 23.- Lorente JA, Lorente AM. El ADN en la identificación criminal y en la paternidad biológica. España: Comares, 1995.
- 24.-Guizar VJ. Genética Médica. 2º ed. México: el manual moderno, 1994.
- 25.-Aranguren JA, Román R, Isea W, Villasmil Y, Jordana J. Los microsatélites (STR's), marcadores moleculares de ADN por excelencia para programas de conservación: una revisión. Arch Latinoam Prod Anim 2005; 13 (1): 1-6.
- 26.- Butler JM. Advanced topics in forensic ADN typing: methodology. USA San Diego California: ELSEVIER, 2011
- 27.- Mendez AJA, Bravo RR, Isea W. Los microsatélites (STR's), marcadores moleculares de ADN por excelencia para programas de conservación: una revisión. ALPA. 2005; 13(1): 30-42
- 28.-Yunis EJ, Yunis J. El ADN en la identificación humana. Bogotá: Ed Temis, 2002.
29. - Coble MD, Hill CR, Vallone PM, Butler JM. Characterization and performance of new MiniSTR loci for typing degraded samples. Intern Congr Ser 2006; 1288: 504-506.
30. - Hill CR, Kline MC, Coble MD, Butler JM. Characterization of 26 MiniSTR Loci for Improved Analysis of Degraded DNA Samples. J Forensic Sci 2008; 53 (1).
- 31.- Ibarra AA, Burgos OM. Discriminación genética de individuos y su uso en medicina forense y filiación natural. Med & Lab 2000; 9:3-4.
32. - Kimpton CP. et al. Automated DNA Profiling Employing Multiplex Amplification of Short Tandem Repeat Loci. Foren Sci Serv 1993; 3: 3-13.
- 33.- Rey M, Usaquen W, Acosta E, Parra J. Datos poblacionales de los Microsatélites de ADN Humano D2S1338, D19S433, PENTA D, PENTA E y SE-33. Acta biol Colomb 2009; 14 (2): 161-168.
34. - Chakraborty R, Kidd KK. The utility of ADN typing in forensic work. JSTOR 1991; 254: 1735-1739
35. - Butler JM. Short tandem repeat typing technologies used in human identity testing. Biotechn 2007; 43 (4): 1-5
36. - Goodwin W, Linacre A, Hadi S. An introduction to forensic genetics. 2º ed. Englad: Wiley and sons, 2011.
- 37.- Pinila GB, Cubillos K, Rodríguez M. Bodas de plata de la reacción en cadena de la polimerasa (PCR). Cienc biom 2008; 6 (9): 65-75.
- 38.- Torres AG, Baca BE. Reacción en cadena de la polimerasa. Elementos 1995; 3 (23):16-21.
39. - Butler JM, Buel E, Crivellente F, McCord BR. Forensic DNA typing by capillary electrophoresis: Using the ABI Prism 310 and 3100 Genetic Analyzers for STR Analysis. [En línea]. Disponible en: http://www.cstl.nist.gov/strbase/pub_pres/Butler2004a.pdf [Consulta: 25-11-2013].

- 40.- Castagnino JM. Electroforesis capilar. Asoc Mex de Bioquim Clín 2000; 25 (1): 13-32.
41. - Mitchelson KR, Cheng J. Capillary Electrophoresis of Nucleic Acids. New Jersey: Ed Humana Press, 2001.
42. - Castagnino JM. Electroforesis capilar. Bioquím 2000; 25 (1): 13-32.
- 43.- Butler JM, Buel E, Crivellente F, McCord BR. Forensic DNA typing by capillary electrophoresis using the ABI Prism 310 and 3100 genetic analyzers for STR analysis. Electrophoresis: 2004; 25: 1391-1412.
- 44.- Butler JM, Hill CR. Biology and Genetics of New Autosomal STR Loci Useful for Forensic DNA Analysis. Forensic Sci Rev 2012; 24 (1).
- 45.- Lorente JA. Un detective llamado ADN. Madrid: Ed Temas de Hoy, 2004.
- 46.- Garay AL. Genética de poblaciones y evolución. México: Universidad Autónoma de Puebla, 1988.
- 47.- Comité Internacional de la Cruz Roja. Personas desaparecidas, análisis forense de ADN e identificación de restos humanos. 2ºed. Ginebra: CICR, 2009.
48. - Parsons TJ, Huel R, Davoren J, Katzmarzyk C, Milos A, Selmanovic A, Et Al. Application of novel "mini-amplicon" STR multiplexes to high volume casework on degraded skeletal remains. Forensic Sci Intern 2007; 1: 175-179.
- 49.- Genética de poblaciones. [En línea]. Disponible en: <<http://geneticabioterio.files.wordpress.com/2012/09/genc3a9tica-de-poblaciones.pdf>> [Consulta: 13-05-2013]
- 50.- Conceptos Básicos de Genética de Poblaciones. [En línea]. Disponible en: <http://biologiamyblog.files.wordpress.com/2010/03/ii-gen_tica_de_poblaciones1.pdf> [Consulta: 15-05-2013]
- 51.- Martínez LJ. Estructura genética de la población de Guatemala, Aplicaciones en el campo antropológico y forense. Editorial de la Universidad de Granada. 2011
- 52.- Quintero Ramos A. Datos poblacionales de cinco STRs de la serie INTERPOL en una población mestiza del Occidente de México. Rev Inv Clín 2009; 61 (2): 104-109.
- 53.- Ossa RH, Tascón EP, Moreno HA, Horta JS, Moreno GR. Frecuencias alélicas de 14 STR's autosómicos en una población de Antioquia, Colombia. ISSN 2010; 8 (13): 25-29.
54. - Edelmann J, Hering S, Michael M, Lessig R, Deichsel D, Meier-Sundhausen G, Et Al. 16 X-chromosome STR loci frequency data from a German population. Forensic Sci Intern 2001; (124) 215-218.
55. - Wigginton JE, Cutler DJ, Abecasis GR. A note on exact tests of Hardy Weinberg Equilibrium. Am J Hum Genet 2005; 76: 887-883.
- 56.- Hamilton MB. Population genetics. UK: Ed Wiley-Blackwell, 2009.
- 57.- Morera B, Marín RR, Barrantes R. Análisis de varios marcadores genéticos clásicos en la población de Costa Rica. Rev Biol Trop 2001; 49: 3-4.
58. - Wigginton JE, Cutler DJ, Abecasis GR. A Note on Exact Tests of Hardy-Weinberg Equilibrium. Am. J. Hum. Genet. 2005; 76: 887-883.

59. - Wittke-Thompson JK, Pluzhnikov A, Cox NJ. Rational Inferences about Departures from Hardy Weinberg Equilibrium. Am J Hum Genet 2005; 76: 967-986.
- 60.-Kamfung W, Yue-ying H. Statistical ADN Forensics: theory methods and computation. UK: Wiley and sons, 2008
- 61- Emigh TH. A comparison of tests for Hardy-Weinberg equilibrium. Biometrics 1980; 36: 627-642.
62. - Huston Kimberly A. Statistical Analysis of STR data. Gene Print 1998; 1: 14-15
- 63.- Budowle B, Shea B, Niezgoda S, Chakraborty R. CODIS STR Loci Data from 41 Sample Population. J Forensic Sci 2001; 46 (3): 453-489.
64. - Botstein D, White RL, Skolnick M, Davis RW. Construction of a Genetic Linkage Map in Man Using Restriction Fragment Length Polymorphisms. Am J Hum Genet 1980; 32: 314-331.
65. - Frankham R, Ballou JD, Briscoe DA, McInnes KH. A Primer of Conservation Genetics. New York: Cambridge university press, 2004.
- 66.- Evaluación e interpretación de resultados en caracterización molecular. [En línea]. Disponible en: http://www3.unileon.es/personal/wwdiafga/web_mex12/presentaciones teoria/leccion_4.pdf [Consulta: 03-05-2013]
67. - Kimura M, Crow JF. The Number of Alleles that can be Maintained in a Finite Population. Genetics 1964; 49: 725-738.
- 68.- INEGI censo 2010. [En línea] Disponible en: <http://www3.inegi.org.mx/sistemas/mexicocifras/default.aspx?e=9> [Consulta: 14-05-2013]
- 69.- Mapa Cd de México INEGI. [En línea]. Disponible en: http://mapserver.inegi.gob.mx/geografia/espanol/datosgeogra/basicos/estados/df_geo.cfm [Consulta: 02-05-2013]
70. - Promega. Technical Manual PowerPlex® Fusion System.[En línea]. Disponible en: <http://www.promega.com/~media/Files/Resources/Protocols/Technical%20Manuals/101/PowerPlex%20Fusion%20System%20Protocol.pdf> [Consulta: 10-04-2013]
- 71.- Ramírez FE., Villegas CD., Vega ML. Genetic variation of 23 STR loci in Mexican mestizo population from Mexico D.F. Forensic Sci Intern 2014
- 72.- Gonzalez-HL, Vega NL, Roche CCc, Canto HJ, Virgen PD, Moscoso CG et al. Forensic Parameters and Genetic Variation of 15 Autosomal STR Loci in Mexican Mestizo Populations from the States of Yucatan and Nayarit. The Open Foren Sce J 2010; 3: 57-63.