



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE QUÍMICA

**“ANÁLISIS POBLACIONAL DE STRS AUTOSÓMICOS EN
POBLACIÓN MEXICANA EMPLEANDO EL *KIT POWERPLEX FUSION*
6 C”**

TESIS

**QUE PARA OBTENER EL TÍTULO DE
QUÍMICA FARMACÉUTICA BIOLÓGICA**

PRESENTA

SHEILA VIVIAM ORTIZ SARABIA



CDMX

2023



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

JURADO ASIGNADO:

PRESIDENTE: Dr. Cruz García Felipe

VOCAL: Dr. Plasencia De la Parra Francisco Javier

SECRETARIO: Dr. Guardado Estrada José Mariano

1er. SUPLENTE: Dra. Dimitrova Dinkova Tzvetanka

2° SUPLENTE: Dra. Bernal Gracida Lilia Angélica

SITIO DONDE SE DESARROLLÓ EL TEMA: Laboratorio de Genética, Instituto de Ciencias Forenses. Tribunal Superior de Justicia de la Ciudad de México.

ASESOR DEL TEMA

Dr. José Mariano Guardado Estrada

SUSTENTANTE

Sheila Viviam Ortiz Sarabia

“La vida no se hace más fácil o más indulgente, nosotros nos hacemos más fuertes y resilientes”.

-Steve Maraboli

INDICE GENERAL

RESUMEN	7
ABREVIATURAS	8
JUSTIFICACIÓN	10
ÍNDICE DE TABLAS	11
ÍNDICE DE FIGURAS	13
1. INTRODUCCIÓN	14
2. MARCO TEÓRICO	16
2.1. ESTRUCTURA Y ORGANIZACIÓN DEL DNA	17
2.2. DOGMA CENTRAL DE LA BIOLOGÍA MOLECULAR	20
2.3. ALELOS Y VARIANTES GENÉTICAS DEL DNA	21
2.4. MICROSATÉLITES (STR)	23
2.4.1. STR AUTOSÓMICOS.....	26
2.4.2. MARCADORES DE LINAJE	27
2.4.2.1. MARCADORES DETERMINANTES DEL SEXO	27
2.4.2.2. Y-STR.....	27
2.4.2.3. X-STR.....	28
2.4.3. NOMENCLATURA.....	28
2.5. ANÁLISIS DE LOS STRs	29
2.6. APLICACIONES DEL ANÁLISIS DE STRS	34
2.7. KITS COMERCIALES DE PCR MÚLTIPLE PARA LA TIPIFICACIÓN DE STR.....	35
2.8. BASES DE DATOS DE DNA POBLACIONALES Y PARA LA IDENTIFICACIÓN HUMANA.....	36
2.9. GENÉTICA POBLACIONAL.....	38
2.9.1. FRECUENCIAS GENOTÍPICAS Y ALÉLICAS	39
2.9.2. PROBABILIDAD DE COINCIDENCIA AL AZAR	40
2.9.3. PODER DE DISCRIMINACIÓN (PD).....	40
2.9.4. HETEROCIGOSIDAD ESPERADA (HE)	41
2.9.5. HETEROCIGOSIDAD OBSERVADA (Ho).....	41
2.9.6. HOMOCIGOSIDAD OBSERVADA (H)	41

2.9.7. CONTENIDO DE INFORMACIÓN POLIMÓRFICA (PIC)	41
2.9.8. PODER DE EXCLUSIÓN (PE)	42
2.9.9. ÍNDICE TÍPICO DE PATERNIDAD (TPI)	42
2.9.10. LEY DEL EQUILIBRIO DE HARDY-WEINBERG.....	42
2.9.10.1. CORRECCIÓN DE BONFERRONI	43
2.9.11 EJEMPLO DE UNA IDENTIFICACIÓN CUANDO LOS PERFILES COINCIDEN.....	44
3. HIPÓTESIS.....	47
4. OBJETIVOS.....	47
4.1. OBJETIVOS GENERALES	47
4.2. OBJETIVOS PARTICULARES.....	47
5. PROCEDIMIENTO EXPERIMENTAL	48
5.1. TIPO DE ESTUDIO.....	48
5.2. POBLACIÓN DE ESTUDIO	48
5.3. CARACTERÍSTICAS DE LAS MUESTRAS	49
5.4. CRITERIOS DE INCLUSIÓN, EXCLUSIÓN Y ELIMINACIÓN	49
5.4.1. INCLUSIÓN	49
5.4.2. EXCLUSIÓN	49
5.4.3. ELIMINACIÓN	49
5.5. MATERIALES Y MÉTODOS.....	49
5.5.1. MATERIAL.....	49
5.5.2. <i>LOC</i> / INCLUIDOS EN ESTE ESTUDIO.....	50
5.5.3. MÉTODOS.....	51
5.5.3.1. OBTENCIÓN DE LA MUESTRA	51
5.5.4. CUANTIFICACIÓN	52
5.5.5. AMPLIFICACIÓN DE LOS STR.....	55
5.5.6. ELECTROFORESIS CAPILAR.....	59
5.5.7. INTERPRETACIÓN DE PERFILES STR.....	60
5.5.8. ANÁLISIS ESTADÍSTICO.....	61
6. RESULTADOS Y DISCUSIÓN	64
7. CONCLUSIONES	80
8. PERSPECTIVAS.....	80
REFERENCIAS.....	81

ANEXO 1.....	89
1.1 Electroferogramas de los controles.....	89
1.2 Matrices para cada <i>locus</i>	94

RESUMEN

La genética forense se puede definir como la especialidad de la genética humana aplicada a la resolución de conflictos jurídicos. La genética forense puede ser utilizada en diferentes casos, algunos de los más comunes son: Identificación del origen biológico de una muestra de una escena del crimen o de una víctima, identificación de cuerpos humanos y estudios de paternidad. Para ello, se analizan diversos marcadores genéticos, de los cuales destacan los microsatélites o STR (del inglés, *short tandem repeats*). Los STR son secuencias de nucleótidos cortas repetidas en tándem, y los que se analizan en la genética forense son aquellos repetidos de cuatro pares de bases, se encuentran distribuidos en todos los cromosomas y son altamente variables. Hoy en día existen kits comerciales para la amplificación de estos marcadores genéticos empleando la reacción en cadena de la polimerasa múltiple para su análisis a través de la electroforesis capilar. En la genética forense se requiere de bases de datos poblacionales de los marcadores genéticos en cuestión, y se utilizan para estimar las probabilidades al azar de coincidencia entre otros parámetros forenses. El objetivo del presente trabajo fue analizar el DNA genómico de 50 individuos mestizos mexicanos de la Ciudad de México empleando el *kit PowerPlex Fusion 6C* que incluye STRs de 23 *loci* autosómicos para fines de identificación humana para conformar una base de datos poblacional. Todos los marcadores analizados en los 50 individuos se encontraron en equilibrio de Hardy Weinberg y la probabilidad de coincidencia al azar combinada fue de 1.36×10^{-25} . Estos resultados indican que pese a ser un tamaño de muestra limitado, con los parámetros forenses estimados, esta base poblacional podría empezar a utilizarse para fines de identificación humana conforme se vaya incrementando en el número de individuos.

ABREVIATURAS

DNA	Ácido desoxirribonucleico
mtDNA	Ácido desoxirribonucleico mitocondrial
RNA	Ácido ribonucleico
mRNA	Ácido ribonucleico mensajero
CE	Electroforesis capilar
CODIS	<i>Combined DNA Index System</i>
dNTP	Deoxi-nucleósido-trifosfato
ESS	<i>European Standard Set</i>
FA	Frecuencia alélica
FBI	Oficina Federal de Investigación
FG	Frecuencia genotípica
FGR	Fiscalía General de la República
GD	Heterocigosidad esperada
HWE	Equilibrio de Hardy-Weinberg
ncRNA	Ácido ribonucleico no codificante
ng	Nanogramos
PCR	Reacción en cadena de la polimerasa
PD	Poder de discriminación

PE	Poder de exclusión
PIC	Contenido de información del polimorfismo
PM	Probabilidad de coincidencia al azar
PPF6C	<i>PowerPlex Fusion 6C</i>
SNV	Variante de un único nucleótido antes SNP
STD	Estandar
STR	Microsatélites / <i>Short Tandem Repeats</i>
TPI	Índice típico de paternidad
VNTR	Repetición en tándem con número variable de copias
μL	Microlitro

JUSTIFICACIÓN

En la actualidad México enfrenta una emergencia forense, se han reportado alrededor de 39,000 cuerpos no identificados en los servicios médicos forenses o inhumados de forma anónima en cementerios públicos, además hasta el 2019 se habían localizado 3,631 fosas clandestinas y hasta el 27 de enero del 2022 se reportaron 97 mil 343 personas desaparecidas. Es por estas cifras que se ha recurrido a revisar el funcionamiento del sistema forense del país (GIZ México *et al.*, 2021; Torres *et al.*, 2022).

Hoy en día la genética forense en México es de gran importancia para la resolución de la crisis forense y en general implica la acreditación de los laboratorios forenses, los recursos humanos, las metodologías y la capacitación del personal dedicado a las áreas involucradas (Torres *et al.*, 2022). Hasta el momento el país no cuenta con una base de datos nacional, además, se desconoce el número de bases de datos que hay en los laboratorios del país, aunque se prevé que, en el 2023 empiece a funcionar una base de datos Nacional, indicado por la ley como obligación de la FGR (Torres *et al.*, 2022). Ya que la creación de bases de datos en las áreas forenses es urgente, el presente proyecto que será de libre acceso y ayudará a los laboratorios que utilicen el sistema PowerPlex Fusion 6C y a la crisis forense en la que se encuentra México.

ÍNDICE DE TABLAS

Tabla 1. Loci STR autosómicos más frecuentes y su ubicación cromosómica.	26
Tabla 2. Ejemplo representativo de la comparación de dos perfiles genéticos con cuatro marcadores STR.....	45
Tabla 3. Ejemplo hipotético del cálculo de PCA y LR en cuatro marcadores STR.	46
Tabla 4. Intervalos de la fuerza del valor del índice de verosimilitud.....	46
Tabla 5. Datos poblacionales de la Ciudad de México y de los Estados Unidos Mexicanos	48
Tabla 6. Diluciones en serie del DNA de control Z1 para la curva estándar.	53
Tabla 7. Mezcla de reacción para cuantificación escalada a ¼ X.	53
Tabla 8. Volumen de agua y DNA para cada pozo.	54
Tabla 9. Condiciones de ciclado de ciclado para el sistema de PCR en tiempo real Applied Biosystems 7500 del fabricante	54
Tabla 10. Volúmenes para la mezcla de amplificación por PCR.	56
Tabla 11. Volúmenes de agua y DNA utilizados para una concentración de 2ng/µL de DNA.....	56
Tabla 12. Volúmenes de agua y DNA utilizados para una concentración de 4ng/µL de DNA.....	57
Tabla 13. Volúmenes de agua y DNA utilizados para una concentración de 8ng/µL de DNA.....	58
Tabla 14. Condiciones de ciclado para PCR múltiplex con el kit PowerPlex Fusion 6C.....	58
Tabla 15. Ajuste del número de ciclos.....	59
Tabla 16. Parámetros utilizados para llevar a cabo la electroforesis capilar (3500 Genetic Analyser).....	60
Tabla 17. Formato de Excel para su análisis en el programa Access.	62
Tabla 18. Ejemplo representativo del formato para su análisis en el programa Access.....	62
Tabla 19. Matriz para cada locus generada por el programa Access.....	62

Tabla 20. Matriz generada en el programa Excel para el cálculo de las frecuencias alélicas.	63
Tabla 21. Frecuencias alélicas obtenidas de los 23 loci autosómicos analizados por el kit PowerPlex Fusion 6C para 50 individuos mestizos mexicanos de la CDMX.	66
Tabla 22. Resultados del equilibrio de Hardy-Weinberg obtenidos con el software STRAF 2.0.8: STR Analysis for Forensics para los 23 <i>loci</i> STR del kit PPF6C.	68
Tabla 23. Parámetros estadísticos calculados para los 23 loci autosómicos estudiados con el kit PPF6C en población mestiza-mexicana de la CDMX.	70
Tabla 24. Resultado de los índices combinados del kit PowerPlex Fusion 6C para 23 loci STR estudiados en una población mestiza-mexicana de la CDMX.	72
Tabla 25. Ejemplos representativo de la comparación entre el kit comercial Investigator 24plex, secuenciación de última generación y el kit PowerPlex Fusion 6C en muestras de 5 individuos.	74
Tabla 26. Comparación de los parámetros estadísticos forenses entre el kit PowerPlex Fusion 6C y el kit Investigator 24Plex.	75
Tabla 27. Comparación de los marcadores incluidos en cada uno de los sistemas con los del presente estudio.	77
Tabla 28. Comparación de los valores de los parámetros estadísticos de interés forense obtenidos en este estudio (kit PPF6C) con los reportados en la literatura.	79

ÍNDICE DE FIGURAS

Figura 1. Estructura de un nucleótido.....	18
Figura 2. Estructura de las bases nitrogenadas.	18
Figura 3. Estructura del DNA.....	19
Figura 4. Dogma central de la biología molecular	21
Figura 5. Representación de STRs.	24
Figura 6. Posición cromosómica de los 13 loci STR básicos de CODIS	25
Figura 7. Proceso de la PCR.....	32
Figura 8. Separación de alelos STR. Pasos de la electroforesis capilar y sus componentes.....	33
Figura 9. Transformación de la información de una muestra en un locus STR	34
Figura 10. Loci de estudio pertenecientes al kit PowerPlex Fusion 6C. .	50
Figura 11. Ejemplo representativo de un electroferograma.....	61
Figura 12. Electroferograma de la muestra LCF015.	65

1. INTRODUCCIÓN

Hoy en día, la identificación humana mediante el uso de metodologías de tipificación del DNA es muy popular para la ciencia forense y en procesos judiciales. Para ello, se analizan marcadores genéticos como las repeticiones cortas en tándem (STR), que constituyen a las secuencias microsatélite, y que son muy fáciles de amplificar por la reacción en cadena de la polimerasa (PCR) (Jamieson & Bader, 2016; Ranjan *et al.*, 2018; Shrivastava *et al.*, 2020). Esto ha sido posible gracias a que, se han creado kits comerciales para analizar estos marcadores empleando la PCR múltiple y la electroforesis capilar, además de que se pueden obtener resultados de manera más rápida, eficiente, así como la amplificación de muestras muy pequeñas (Ranjan *et al.*, 2018). Además, existen varios kits en el mercado aprobados por los órganos regulatorios de la genética forense a nivel mundial; con ellos es posible desarrollar bases de datos para hacer los cálculos necesarios en la comparación de perfiles de STR (Butler, 2005; Wickenheiser, 2022).

Cuando se hace una comparación de perfiles genéticos, ya sea para identificación humana o para relaciones de parentesco biológico, se requiere hacer cálculos que nos permitan identificar si esa coincidencia fue al azar o realmente se trata de una identificación o una prueba de parentesco positiva (Torres *et al.*, 2022). Por esta razón, este tipo de pruebas requiere de bases de datos poblacionales, las cuales contienen las frecuencias alélicas de los marcadores STR analizados. Estas bases de datos son representativas de la población en la que se lleva a cabo el estudio, y a su vez nos permite estimar qué tan frecuente sería cierto perfil genético en esa población. Dicha estimación se basa en la fórmula que se utiliza para saber si dicho marcador se encuentra en equilibrio de Hardy-Weinberg, en la que, a partir de las frecuencias alélicas observadas en cierta población, se pueden estimar la frecuencia esperada para cierto genotipo de determinado STR en la población (Thompson & Black, 2007).

En los laboratorios de genética forense, cuando se va a utilizar un kit comercial, ya sea para pruebas de paternidad o para identificación humana, se requiere hacer una base de datos que contenga las frecuencias alélicas de los STR

estudiados en la población. Por otra parte, para evaluar qué tan útil puede ser determinado kit para fines forenses, se estiman una serie de parámetros forenses cuyo objetivo es determinar el poder de discriminación que presenta el kit. Algunos de los principales parámetros estadísticos que se calculan son: Probabilidad de coincidencia al azar (PM), Poder de discriminación (PD), Heterocigosidad esperada (GD o H), Contenido de información del polimorfismo (PIC), Poder de exclusión (PE) y el índice típico de paternidad (TPI) (Hernández & Trejo, 2014). La aplicación de la estadística permite estimar la probabilidad de que exista relación entre dos perfiles genéticos (Rangel, 2020). De acuerdo con las recomendaciones de la Sociedad Internacional de Genética Forense, las bases de datos poblacionales tienen que cumplir con las siguientes características: utilizar muestras de personas no emparentadas de etnia conocida si es posible, también deben seleccionarse sin conocimiento previo de los genotipos en los *loci* a estudiar para asegurar la aleatoriedad de las muestras, el laboratorio debe tener buenos estándares de calidad para que los resultados sean de confianza y la mayoría de los datos de la población incluyan un orden de 100 a 200 tipos de STR por *locus* por población examinada (Butler, 2005; Butler, 2010). Sin embargo, la obtención de una base de datos que cumpla con dichas características puede ser laboriosa y costosa, lo cual podría resolverse al analizar sistemas que incluyan muchos marcadores en pocas muestras (Butler, 2010; Jamieson *et al.*, 2016). Esto permitiría empezar a analizar los estudios de genética forense con una base de datos poblacional en constante crecimiento sin perder la confiabilidad.

En este estudio se analizó un grupo de 50 muestras de DNA genómico empleando el *kit PowerPlex Fusion 6C* que analiza 23 *loci* autosómicos (CSF1PO, FGA, TH01, vWA, D1S1656, D2S1338, D2S441, D3S1358, D5S818, D7S820, D8S1179, D10S1248, D12S391, D13S317, D16S539, D18S51, D19S433, D21S11, Penta D, Penta E, D22S1045, TPOX y SE33) con el objetivo de conformar una base de datos poblacional que pueda ser útil para la identificación humana. Las muestras empleadas provienen de individuos mexicanos mestizos de la Ciudad de México, seleccionados aleatoriamente y no relacionados biológicamente. Con los datos obtenidos se evaluó si una base

de datos poblacional limitada en tamaño de individuos, puede ser utilizada para fines forenses debido al número de marcadores que analiza.

2. MARCO TEÓRICO

La genética es la ciencia que estudia los genes y la herencia basada en los cambios de la secuencia de DNA heredados de los padres a los hijos (NIH, 2022). La genética de poblaciones es la rama de la genética que estudia de manera teórica y empírica las causas de los cambios en las frecuencias de las variantes genéticas (diferencias en la secuencia de DNA que ocurren entre los individuos de una población determinada) dentro de las poblaciones (Charlesworth y Jensen, 2021; NIH, 2023).

Todos los organismos vivos están compuestos de células que son la unidad básica de la vida, estas producen las materias primas, la energía y tiene la capacidad de eliminar los desechos necesarios para mantener el equilibrio. Un ser humano tiene aproximadamente 100 billones de células y cada una de ellas contiene la misma información genética (Butler, 2010; Karp, 2020).

Las células en su núcleo contienen a los cromosomas y estos a su vez una sustancia química llamada DNA (ácido desoxirribonucleico) o DNA nuclear, que contiene toda la información necesaria para replicar la célula, para la replicación del DNA, transcripción de DNA a RNA y para traducir el RNA en proteína (Butler, 2010; Karp, 2020). También las mitocondrias de las células contienen su genoma, el cual es una molécula circular de DNA de doble cadena de 16,6 kb y es utilizado como marcador de linaje porque es heredado por parte de la madre (Karp, 2020; Ranjan *et al.*, 2018). El cuerpo colectivo de información genética presente en una especie junto con las instrucciones para crear un organismo nuevo se le denomina genoma (Karp, 2020).

Un gen es la secuencia del genoma que codifica para uno o varios productos funcionales, los cuales pueden ser polipéptidos o RNA no codificante (ncRNA), y que ocupa un *locus* en un cromosoma; pueden existir en diferentes formas a las cuales se les denomina alelos (Castillo *et al.*, 2019). Los genes contienen la información necesaria para formar polipéptidos, la expresión génica se refiere a la generación de un producto funcional a partir de la información codificada en

un gen y se lleva a cabo en dos etapas: transcripción y traducción, en las cuales se produce una copia de RNA del gen y la síntesis de una cadena polipeptídica (Castillo *et al.*, 2019; Karp, 2020).

El DNA tiene algunas funciones principales: hacer copias de sí mismo para que las células al dividirse transmitan la misma información; y llevar las instrucciones para que las células puedan construir la maquinaria de la vida (Butler, 2005). Toda la información contenida en el DNA se transmite de generación en generación, en el caso de una persona, la mitad de la información proviene de la madre y la otra mitad proviene del padre (Castillo *et al.*, 2019).

2.1. ESTRUCTURA Y ORGANIZACIÓN DEL DNA

El DNA está compuesto por cadenas de polinucleótidos unidos por enlaces fosfodiéster, cada nucleótido está constituido por: una base nitrogenada, un azúcar de cinco carbonos (pentosa) conocida como desoxirribosa y un grupo fosfato (Castillo *et al.*, 2019; Goodwin *et al.*, 2011). La base nitrogenada (llamada así porque hay átomos de nitrógeno en los anillos de la molécula) imparte la variación en cada nucleótido y el fosfato enlazado con el azúcar constituyen la columna vertebral de la molécula de DNA (Karp, 2020; Mathews *et al.*, 2013). Durante la formación de una cadena de DNA, el grupo hidroxilo unido al carbono 3' de la desoxirribosa de un nucleótido se une mediante un enlace éster al grupo fosfato del carbono 5' del siguiente nucleótido en la cadena, a estos enlaces se les llama 3'-5'- fosfodiéster ya que el fosfato se esterifica con dos átomos de oxígeno uno de cada desoxirribosa adyacente (Karp, 2020; Mathews *et al.*, 2013).

Nucleótido

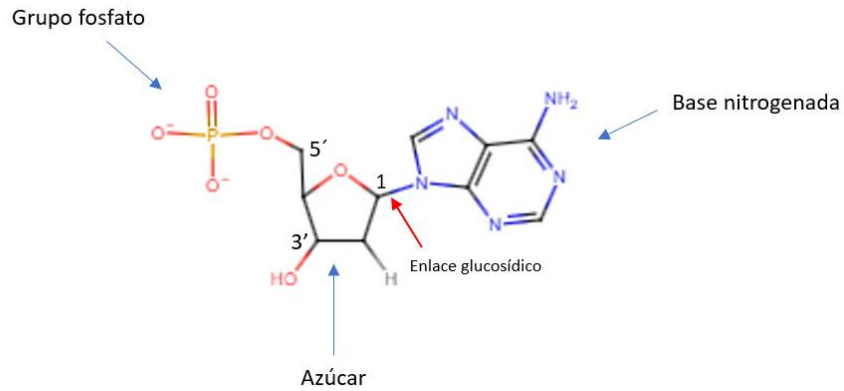


Figura 1. Estructura de un nucleótido. Un nucleótido se compone de: un grupo fosfato, un azúcar en este caso desoxirribosa y una base nitrogenada (adenina). Se muestra la base nitrogenada unida por un enlace N-glucosídico al carbono 1' de la pentosa, cuyo carbono 5' se encuentra enlazado al grupo fosfato.

Las bases nitrogenadas son anillos heterocíclicos, y se dividen en dos clases: purina y de la pirimidina. Las purinas: adenina y guanina, están compuestas por dos anillos, uno de seis átomos enlazado con uno de 5 átomos (difieren en las posiciones de sus dobles enlaces y en los grupos unidos al anillo de seis átomos) y las pirimidinas: timina y citosina están constituidas por un anillo de seis átomos como se muestra en la figura 2 (Castillo *et al.*, 2019; Mathews *et al.*, 2013; Pierce, 2014).

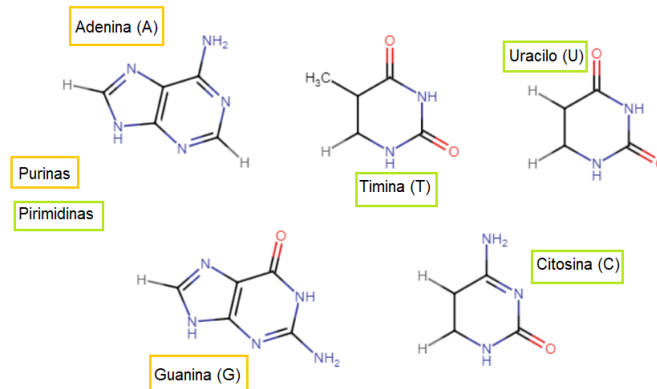


Figura 2. Estructura de las bases nitrogenadas. Las cinco bases nitrogenadas se clasifican en Purinas (A y G) y pirimidinas (T, C y U). La adenina y guanina tienen un doble ciclo mientras que la timina y pirimidina están compuestas por un solo ciclo.

De acuerdo al modelo B de Watson y Crick, el DNA tiene una estructura de doble hélice que posee un surco mayor que corresponde al espacio adyacente de cadenas azúcar-fosfato y un surco menor perteneciente al enlace de las

bases, cada una de estas cadenas consta de una serie de moléculas de desoxirribosa enlazadas entre sí mediante un grupo fosfato unido a sus átomos de carbono 3' (tres prima) y 5' (cinco prima) (Mathews *et al.*, 2013; Thompson & Black, 2007). En cada extremo inferior de la cadena se encuentra un átomo de carbono 3' con un grupo hidroxilo libre (-OH) y en el extremo superior se encuentra un carbono 5' con un grupo fosfato libre y esta orientación es inversa en la cadena complementaria, es así como una sección de DNA de doble cadena está constituida por nucleótidos individuales que se aparean con sus bases complementarias a través de puentes de hidrógeno (Alberts *et al.*, 2021; Mathews *et al.*, 2013; Pritchard, 2013). Las reglas de apareamiento de bases son, que la adenina sólo puede hacerlo con la timina y la citosina con la guanina ya que, las distancias entre los carbonos 1' de las porciones de desoxirribosa de adenina-timina y guanina-citosina, son las mismas (1.1nm aproximadamente). Un apareamiento de bases de esta manera permite que la doble hélice tenga un diámetro regular (esto no sería posible si se aparearan bases del mismo tipo (Mathews *et al.*, 2013; Watson *et al.*, 2016).

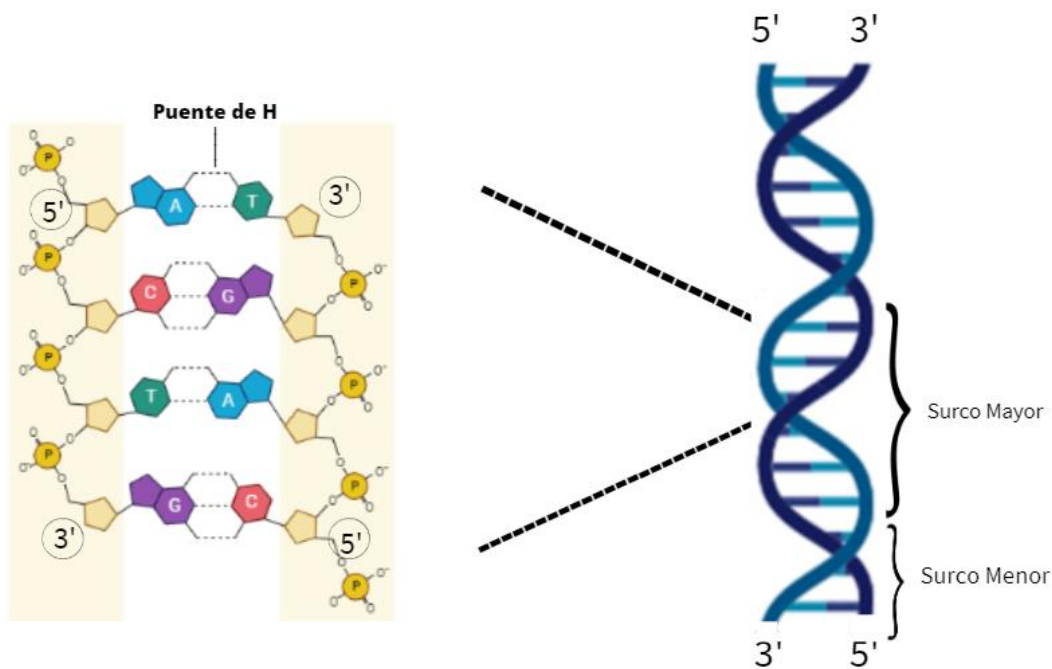


Figura 3. Estructura del DNA (Creado con BioRender.com). Los nucleótidos se unen mediante enlaces fosfodiéster y que forman una doble cadena a partir del apareamiento de sus bases nitrogenadas (A-T y C-G) unidas por puentes de hidrógeno.

El DNA se empaqueta en una célula gracias a *que este se encuentra superenrollado, por lo que se calcula que un humano almacena 3,200,000,000 de pares de bases (pb) de DNA (Goodwin et al., 2011)*. Dentro de la célula, el DNA se encuentra asociado con proteínas a esto se le conoce como cromosoma. La mayor parte de la condensación en las células eucariontes es consecuencia de la asociación del DNA con sus proteínas en su mayoría con unas proteínas llamadas histonas que permiten la formación de estructuras denominadas nucleosomas. Un nucleosoma consta con aproximadamente 150 pb de secuencia de DNA enrollado alrededor de un centro de histonas, estos se pliegan y empaquetan formando un complejo de DNA llamado cromatina. La fibra de cromatina se condensa y da lugar a los cromosomas (Alberts *et al.*, 2021; NIH, 2023; Watson *et al.*, 2016).

El genoma nuclear está dividido en cromosomas que forman el genoma humano constituido por 22 pares autosómicos y dos cromosomas sexuales, por lo tanto, las células humanas tienen en total 46 cromosomas diferentes. Los cromosomas sexuales para el genotipo masculino son X y Y (contienen una sola copia del cromosoma X y otra del cromosoma Y), mientras que el genotipo femenino está definido por XX (heredan un cromosoma X de cada progenitor) (Michaelis *et al.*, 2008).

Los genes están formados por: exones e intrones. Los exones son secuencias de nucleótidos que se conservan en el transcrito primario maduro (RNAm maduro), algunos exones son codificantes, es decir que contienen información para producir una proteína, mientras que otros no son codificantes. Los intrones son secuencias de nucleótidos que son eliminadas del transcrito final durante el procesamiento del RNA heteronuclear, que puede ser no codificante o codificar para un ncRNA (NIH, 2023a; NIH, 2023b).

Las regiones no codificantes son útiles para la identificación humana ya que las variantes genéticas que difieren entre un individuo y otro están ubicados justamente en estas regiones (Butler, 2005).

2.2. DOGMA CENTRAL DE LA BIOLOGÍA MOLECULAR

El dogma central de la biología molecular es el mecanismo para el flujo de la información genética que en un principio indicaba el flujo de la información

comenzaba con el DNA genómico, después se lleva a cabo el proceso de transcripción a RNA y después el RNA dirige la síntesis y secuencia de los polipéptidos. Cabe mencionar que es posible el flujo de información de RNA a DNA un fenómeno conocido como transcripción inversa descubierto en retrovirus que se lleva a cabo gracias a una enzima polimerasa llamada transcriptasa inversa que es capaz de sintetizar DNA de doble cadena a partir de ARN monocatenario (Bhagavan & Ha, 2022; Castillo *et al.*, 2019; Watson *et al.*, 2016).

La información fluye de acuerdo la teoría del dogma central iniciando a partir de los genes que inician una serie de reacciones llamadas expresión génica. Este proceso está constituido por la transcripción y la traducción. En la transcripción se produce una copia del RNA del gen y en la traducción se realiza la síntesis de una cadena polipeptídica y la secuencia de aminoácidos es determinada mediante el código genético a partir de la secuencia de nucleótidos del mRNA (Encina, 2013).

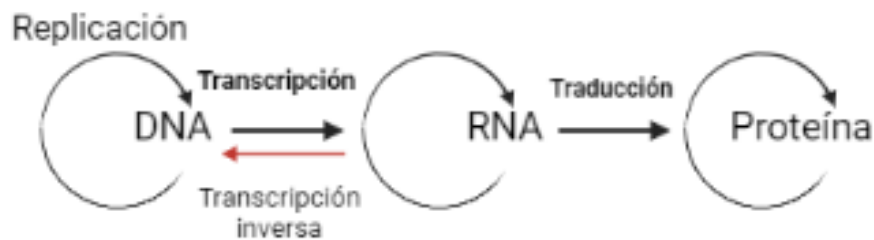


Figura 4. Dogma central de la biología molecular (creado con BioRender.com). En la sección a) se muestra el proceso de replicación que establece un flujo de información genética de DNA a DNA, b) se muestra el proceso de transcripción que establece un flujo de información genética de DNA a RNA, c) muestra el flujo de información genética de RNA a proteína (traducción), e) muestra el proceso de transcripción inversa y d) muestra el proceso de autoreplicación de algunas proteínas específicas (priones) (Herrero *et al.*, 2021).

2.3. ALELOS Y VARIANTES GENÉTICAS DEL DNA

El genoma humano puede experimentar cambios debido a la selección natural y la evolución, provocando modificaciones en la secuencia de nucleótidos del genoma; siendo heredable y permanente a esto se le conoce como mutación (Castillo *et al.*, 2019; Loscalzo *et al.*, 2022).

La secuencia del DNA puede presentarse en diferentes versiones a las cuales se les denominan alelos, todos estos alelos tienen una ubicación específica en

los cromosomas a la que se le denomina *locus* (plural *loci*). Si los alelos de un *locus* son diferentes entonces los individuos se denominan heterocigotos y si ambos alelos son iguales el individuo es homocigoto (Michaelis *et al.*, 2008; Nussbaum *et al.*, 2008). En los *loci* de las variantes genéticas de una población puede haber muchos alelos diferentes. La variación genética entre un individuo y otro, o las diferencias en el genoma entre poblaciones son causadas por mutaciones, segregación y la recombinación meiótica (Castillo *et al.*, 2019; NIH, 2023a; NIH, 2023b; Nussbaum *et al.*, 2008).

Se estima que existe una diferencia de 1000 pb entre cada cromosoma humano, con un promedio de 1.250 pb, lo que significa que compartimos alrededor de un 99.9% de nuestro genoma. Ya que la mayor parte del genoma es común en los humanos no tiene caso analizar todo el DNA sino solamente regiones que varíen entre un individuo y otro (Goodwin *et al.*, 2011).

Los *loci* de las variantes genéticas en una población se pueden utilizar como marcadores genéticos (variantes en las secuencias dentro del DNA y que por lo general se conoce su ubicación en los cromosomas)(NIH, 2023) para la identificación humana, entre ellos están: los minisatélites (VNTR, variable number of tandem repeats), los microsatélites (STR, short tandem repeats), variantes de un solo nucleótido (SNV, single nucleotide variant), inserciones/delecciones o indel (insertion/deletion), entre otros (Castillo *et al.*, 2019).

- Minisatélites (VNTR)

Los minisatélites o también denominados repeticiones en tándem de número variable (VNTR) suelen tener repeticiones de 9 a 80 pb de longitud y suelen repetirse entre 5 y 50 veces. Los minisatélites se encuentran en más de 1000 localizaciones en el genoma humano y presentan una alta tasa de mutación en las poblaciones. Su uso se encuentra limitado por que se necesita una gran cantidad de DNA de buena calidad (Castillo *et al.*, 2019; Ranjan *et al.*, 2018). El procedimiento antiguo para su análisis consistía en utilizar una enzima de restricción para cortar las regiones de DNA que rodeaban a los

VNTR, el cual se sustituyó en el futuro por la técnica de PCR (Shrivastava *et al.*, 2020).

- Variantes de un solo nucleótido (SNV)

Las variantes de un solo nucleótido (SNV), se encuentran cada 300 a 1000 pb en el genoma humano y tienen como ventaja su gran abundancia y su fácil tipificación sin embargo tienen en contra su baja variabilidad lo que complica su uso como marcadores genéticos (Castillo *et al.*, 2019).

- INDEL

Las inserciones y deleciones también llamadas INDEL por su contracción, se refieren a la adición o eliminación de nucleótidos y comprenden desde 1 a 50 nucleótidos. Las variaciones INDEL son importantes ya que pueden alterar características fenotípicas y provocan enfermedades; pueden ser estudiadas a partir de secuenciación de última generación (Castillo *et al.*, 2019).

2.4. MICROSATÉLITES (STR)

En la tipificación del DNA, se examinan múltiples marcadores genéticos y en la mayoría de los laboratorios de genética forense se ocupan los microsatélites o STR (Short tandem repeats), que tienen una secuencia de nucleótidos corta que se repite un número de veces variable entre un individuo y otro (Goodwin *et al.*, 2011). Si cada marcador se hereda independientemente de los demás se puede calcular la frecuencia de un perfil de DNA multiplicando cada una de las de las frecuencias de los genotipos individuales, pero por cuestiones de tiempo no es conveniente analizar toda la secuencia de DNA, así que sólo se evalúan algunas locaciones discretas (Butler, 2005; Goodwin *et al.*, 2011).

Los genomas eucariontes están llenos de secuencias de DNA repetidas, estas secuencias son de diferentes tipos y longitud de pares de bases, de acuerdo con esto es que son clasificadas. Los microsatélites, son secuencias altamente variables de DNA constituidas por repeticiones cortas entre dos y nueve pares

de bases, que varían de longitud entre individuos (Nussbaum *et al.*, 2008; Zauza *et al.*, 2015).

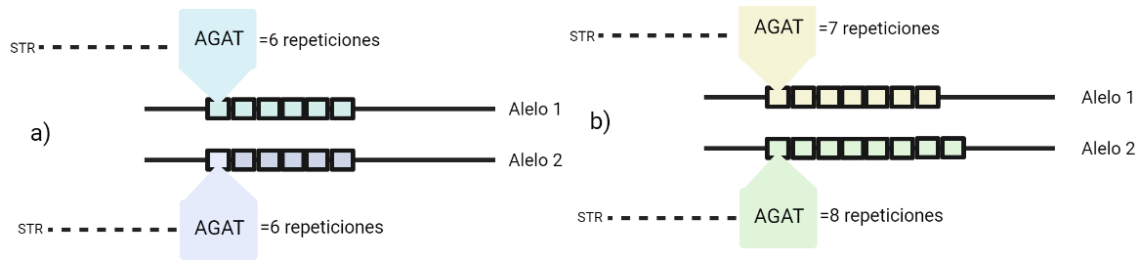


Figura 5. Representación de STRs. En la sección a) se muestra un STR en homocigosis (mismo número de repeticiones); b) se muestra un STR en heterocigosis (diferente número de repeticiones).

Estos marcadores son propensos a mutaciones debido a errores en la replicación del DNA, causado por un fenómeno de deslizamiento de la polimerasa que sufre una detención del copiado, a continuación, una disociación del complejo de replicación y posteriormente la enzima vuelve al sitio donde se realizó la interrupción, pero puede equivocarse formando una repetición previa a la última realizada o posterior a ella (Dieringer y Schlotterer, 2003; Leclercq *et al.*, 2010). Esta tasa de mutación alta hace que los STR tengan una buena capacidad informativa en los estudios genéticos, y especialmente son útiles para investigar la variación genética en las poblaciones (Butler, 2010; Zauza *et al.*, 2015).

Algunas de las ventajas por las que se utilizan los STR son (Goodwin *et al.* 2011; Shrivastava *et al.*, 2020):

- Alelos discretos y distinguibles.
- Proporcionan un nivel alto de discriminación >0.9.
- Heterocigosidad >70%.
- Ausencia de ligamiento genético con otros *loci* analizados.
- Alta variabilidad.
- La capacidad de ser amplificados como parte de una PCR múltiple.

Los marcadores STR se emplearon para la identificación humana a principios de la década de los 90, y desde entonces se ha buscado identificar nuevos *loci* que sean útiles para dicho propósito. Los primeros cuatro *loci* utilizados fueron: TH01, VWA, FES/FPS y F13A1, seguido de un múltiplex de segunda generación formado por: TH01, VWA, FGA, D8S1179, D18S51 y D21S11. El incrementar el número de marcadores permitió establecer una serie de marcadores o *loci* centrales, los cuales además de tener un poder de discriminación alto, facilitan su comparación entre los laboratorios de genética forense. El 13 de octubre de 1998 la Oficina Federal de investigación (FBI) de EE.UU. anunció los 13 *loci* STR básicos de su base de datos de DNA a nivel nacional llamada: Combined DNA Index System (CODIS) con los siguientes marcadores: CSF1PO, FGA, TH01, TPOX, VWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51 y D21S11 (FBI, 2005; Michaelis et al., 2008).

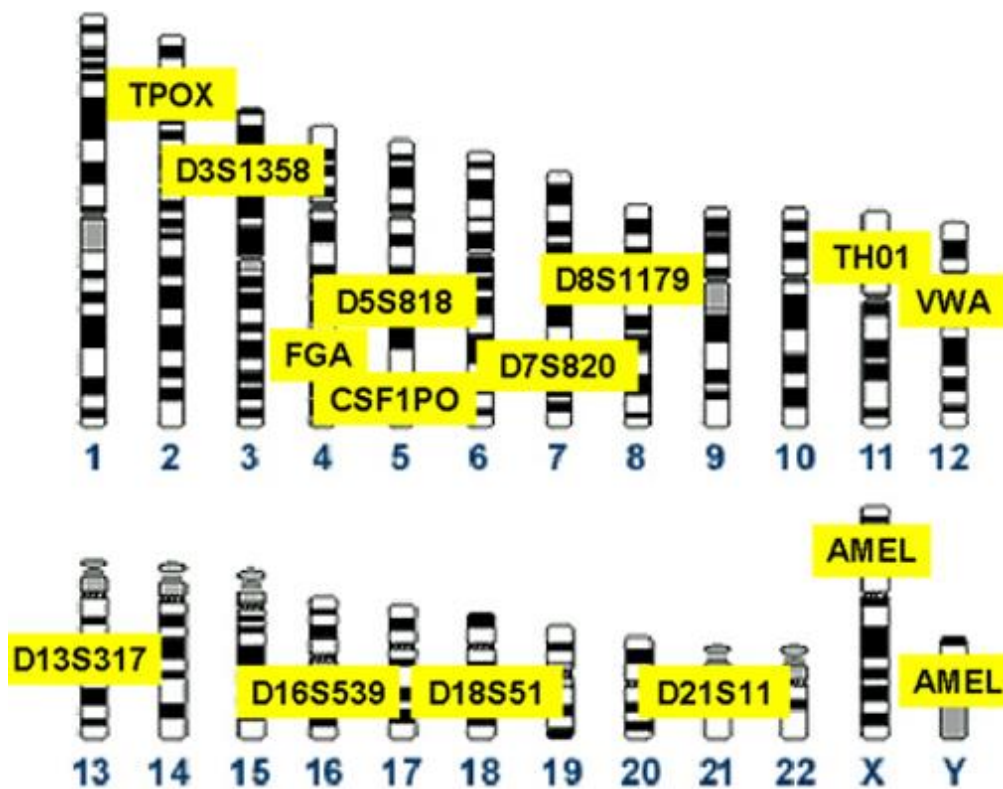


Figura 6. Posición cromosómica de los 13 loci STR básicos de CODIS *modificado de Baggett et al., 2005. Posición cromosómica de los STR aceptados como loci básicos de CODIS: CSF1PO, FGA, TH01, TPOX, VWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11 y un marcador determinante del sexo: Amelogenina (AMEL).*

Los *loci* básicos están ubicados en cromosomas distintos, por lo tanto, se espera que su segregación durante la meiosis sea independiente unos de otros a excepción de algunos (D5S818 y CSF1PO) ligados en el cromosoma 5. Su segregación independiente permite el uso de la regla del producto para determinar las probabilidades de coincidencia al azar en el DNA generada a partir de *loci* STR múltiples (Butler, 2006).

2.4.1. STR AUTOSÓMICOS

Los STR autosómicos como su nombre lo dice se encuentran en alguno de los 22 cromosomas autosómicos humanos y muestran una variación entre individuos, debida a mutaciones, la segregación independiente y a la recombinación meiótica; lo que hace que sean útiles para la identificación humana (Butler, 2010).

Tabla 1. Loci STR autosómicos más frecuentes y su ubicación cromosómica. Se muestra para cada marcador su localización, la secuencia repetida y el rango de alelo (Shrivastava *et al.*, 2020). *Chr= Cromosoma.

Locus	Cromosoma	Localización cromosómica	Secuencia repetida	Rango de alelo
D1S1656	Chr 1	1q42	TAGA	8–20.3
D2S441	Chr 2	2p14	TCTA/TCAA	8–17
TPOX	Chr 2	2p25.3	GAAT	4–16
D2S1338	Chr 2	2q35	TGCC/TTCC	15–28
D3S1358	Chr 3	3p21.31	TCTG/TCTA	8–21
FGA	Chr 4	4q31.3	CTTT/TTCC	12.2–51.2
D5S818	Chr 5	5q23.2	AGAT	7–18
CSF1PO	Chr 5	5q33.1	TAGA	5–16
D6S1043	Chr 6	6q15	AGAT	9–25
SE33	Chr 6	6q14	AAAG	4.2–37
D7S820	Chr 7	7q21.11	GATA	5–16
D8S1179	Chr 8	8q24.13	TCTA/TCTG	7–20
D10S1248	Chr 10	10q26.3	GGAA	7–19
TH01	Chr 11	11p15.5	TCAT	3–14
D12S391	Chr 12	12p13.2	AGAT/AGAC	13–27.2
VWA	Chr 12	12p13.31	TCTG/TCTA	10–25
D13S317	Chr 13	13q31.1	TATC	5–16
Penta E	Chr 15	15q26.2	AAAGA	5–32
D16S539	Chr 16	16q24.1	GATA	5–16
D18S51	Chr 18	18q21.33	AGAA	7–40
D19S433	Chr 19	19q12	AAGG/TAGG	9–17.2
D21S11	Chr 21	21q21.1	TCTA/TCTG	12–41.2
Penta D	Chr 21	21q22.3	AAAGA	1.1–19
D22S1045	Chr 22	22q12.3	ATT	7–20

2.4.2. MARCADORES DE LINAJE

2.4.2.1. MARCADORES DETERMINANTES DEL SEXO

La determinación del sexo en los laboratorios forenses es un proceso rutinario y se hace a partir de diferentes pruebas biológicas, algunos métodos de determinación son a partir de la identificación ósea como el cráneo, pelvis, etc. A veces no es posible la identificación de restos óseos completos por lo que la tipificación por DNA es un método sencillo y rápido (Shrivastava *et al.*, 2020).

El marcador para la identificación de sexo se encuentra en los cromosomas X y Y del individuo es parte del gen que codifica para la amelogenina y no es un STR, hoy en día este marcador está incluido en todos los kits comerciales de tipificación de DNA. El gen *AMEL* se encuentra tanto en el cromosoma X como en el Y en las regiones Xp22.1-p22.3 (longitud de 2,872 pb) y Yp11.2 (longitud de 3,272 pb) respectivamente (Velarde *et al.*, 2008), existen divergencias en longitud entre los alelos codificados en el cromosoma X y el cromosoma Y, los dos genes son homólogos y tienen secuencias altamente conservadas por lo tanto sirven como marcadores fiables para la determinación del sexo (Frances *et al.*, 2008; Rudin & Inman, 2002).

2.4.2.2. Y-STR

Los STR también se encuentran en el cromosoma Y humano, sólo los hombres tienen un cromosoma Y, esto es muy útil para resolver casos forenses ya que tiene secuencias específicas de microsatélites que son denominadas Y-STR (Ranjan *et al.*, 2018).

Ya que este tipo de marcadores solo los tienen los hombres, se pueden realizar estudios en muestras mixtas (procedentes de mujeres y hombres) en el caso de agresiones sexuales o de algún accidente masivo, además, los Y-STR ayudan a un mejor análisis de muestras degradadas y para establecer relaciones de parentesco biológico ya que el cromosoma Y es heredado de manera intacta de generación en generación. En el caso de querer identificar a una persona desaparecida no es conveniente el estudio de solo estos marcadores ya que favorece el aumento de posibles coincidencias porque pueden encontrarse personas del mismo linaje paterno (Rudin & Inman, 2002).

2.4.2.3. X-STR

El análisis de X-STR es utilizado principalmente para la relación biológica entre padre-hija ya que el cromosoma X del padre es heredado de manera íntegra a la hija, además, también es posible llevar a cabo estudios de relación biológica entre medias hermanas paternas y hermanas (Shrivastava et al., 2020).

El estudio de estos marcadores aún tiene limitaciones ya que los kits de estudio aún contienen un número limitado de marcadores y generan amplicones (Fragmento de DNA o RNA que es el producto de amplificación o replicación) de gran tamaño (Shrivastava et al., 2020).

2.4.3. NOMENCLATURA

Para facilitar la reproducibilidad entre laboratorios se ha desarrollado una nomenclatura en común en el ámbito del DNA forense. Si un marcador forma parte de un gen, se utiliza el nombre del gen en la designación, un ejemplo es el marcador TH01 que procede del intrón uno del gen humano que codifica a la hidroxilasa de tirosina situado en el cromosoma 11. En el caso del marcador TPOX este se localiza en el intrón diez del gen de la peroxidasa tiroidea (*TPO*). Algunas veces se agrega el prefijo HUM- al principio del nombre del *locus*; sin embargo, esto no es necesario si son pruebas para la identificación humana (Butler, 2010).

Los marcadores de DNA que quedan fuera de las regiones génicas pueden designarse por su posición cromosómica. Un ejemplo es el marcador D5S818, la "D" significa DNA, el siguiente carácter se refiere al número de cromosoma en el que se encuentra, en este caso "5", la "S" indica que el marcador de DNA es una secuencia específica de una sola copia por genoma haploide y el número final indica el orden en que se descubrió y se clasificó el marcador para ese cromosoma en este caso el 818. Los números secuenciales se utilizan para dar un carácter único a cada marcador de DNA identificado (Shrivastava et al., 2020).

2.5. ANÁLISIS DE LOS STRs

Muchos marcadores STR se han logrado caracterizar por laboratorios académicos y comerciales para su uso en identificación de enfermedades, elaboración de kits comerciales y creación de bases de datos.

El proceso de tipificación de marcadores STR involucra varios pasos, empezando por la recolección de la muestra biológica, la extracción del DNA, su cuantificación de DNA, amplificación por PCR múltiple, la electroforesis capilar de los productos de la amplificación (para separar y caracterizar los alelos del STR), la genotipificación del STR y por último la interpretación del perfil (Shrivastava *et al.*, 2020).

La recolección de muestra depende del material biológico que se desea analizar, el DNA se puede obtener de diferentes fuentes como: huesos, sangre, saliva, semen, dientes, entre otras, pero los materiales más comunes son: la sangre y el semen. Cabe mencionar que las muestras deben recolectarse, conservarse, almacenarse y transportarse de manera cuidadosa para evitar contaminaciones antes de realizar el análisis del DNA en el laboratorio forense (Goodwin *et al.*, 2011).

Como se busca obtener las muestras de la manera más rápida e indolora posible, el método más sencillo es la recolección de células de descamación de la mucosa bucal o saliva, aunque algunas veces se opta por usar la toma de muestras de sangre capilar para poder utilizar tarjetas FTA que están elaboradas de un papel absorbente a base de celulosa que se impregna con sustancias químicas y permite el almacenamiento de DNA) (Goodwin *et al.*, 2011; Shrivastava *et al.*, 2020).

Después de llevar a cabo la recolección, la muestra debe someterse a un método de extracción ya que la molécula de DNA se encuentra acompañada por algunas proteínas y materiales celulares que pueden inhibir la capacidad de analizarla, por lo tanto, se han desarrollado diferentes métodos de extracción de DNA. Entre los más comunes se encuentran: la extracción orgánica, extracción con Chelex (es una suspensión en agua de pequeñas partículas de copolímeros de estireno-divinil-benceno uniéndose a los iones de metales

polivalentes como: Mg^{2+} , Ca^{2+} y Fe^{2+}) y extracción con tarjeta FTA, el método se escoge dependiendo de la muestra biológica que se necesite analizar (Linacre & Tobe, 2013).

Independientemente del método que se utilice en la extracción, esta debe hacerse cuidadosamente para evitar contaminaciones y resultados erróneos. El DNA extraído puede ser almacenado a $-20^{\circ}C$ o a $-80^{\circ}C$ si se requiere conservar por mayor tiempo, esto se hace para evitar la actividad de nucleasas que son proteínas que degradan al DNA para permitir el reciclaje de los componentes de los nucleótidos. Otra manera de inhibir a las nucleasas es el uso de tubos con EDTA que es una sustancia quelante que se une al magnesio libre (cofactor de las nucleasas) (Goodwin *et al.*, 2011; NIH, 2022; Salazar & Guerrero, 2016).

Al terminar la extracción de DNA se cuantifica porque la determinación de la cantidad de DNA en una muestra es esencial para la mayoría de los ensayos basados en la PCR (es un proceso enzimático en el cual una región específica del DNA se replica una y otra vez para producir muchas copias de una secuencia) (Ranjan *et al.*, 2018), ya que una concentración demasiado alta puede dar lugar a picos fuera de la escala de medición para la técnica y una concentración muy pequeña puede provocar la pérdida de alelos porque no es posible amplificar el DNA (Butler, 2005). En el caso de los kits comerciales funcionan mejor con un rango de concentración específico que ya se ha estandarizado y validado anteriormente (Butler, 2005; Shrivastava *et al.*, 2020). La cuantificación a través del método de PCR en tiempo real se ha convertido en uno de los métodos más comunes y consiste en una variante de la PCR, en donde la cantidad de producto de DNA se va midiendo conforme avanza la reacción cíclica gracias a fluoróforos que se le unen, donde el incremento de la fluorescencia es proporcional al aumento de la cantidad de moléculas de DNA (Thompson & Black, 2007)

Después de realizar la cuantificación, se realiza la amplificación del DNA mediante la técnica de PCR que consiste en una serie de cambios de temperatura que se repiten un promedio de 30 veces, generando

aproximadamente un billón de copias de la muestra (Thompson & Black, 2007), cada ciclo consta de tres pasos: desnaturalización, alineamiento y extensión.

Para poder llevar a cabo la PCR se necesita DNA molde (aislado de una muestra), cebadores, que son secuencias cortas de DNA monocatenario sintetizado, que se unen a regiones específicas del DNA molde, es así como se definen la región que se va a amplificar (Thompson & Black, 2007), una enzima llamada Taq polimerasa que es termoestable, cloruro de magnesio ($MgCl_2$) que es cofactor de la Taq polimerasa, una solución tampón que mantiene el pH óptimo para la enzima Taq polimerasa que contiene cloruro de potasio (KCl) junto con Tris-HCl (pH=8.4 a temperatura ambiente), nucleótidos (*dNTP*: Trifosfatos de desoxinucleótidos) con los que se construirán las nuevas cadenas de DNA y por último agua libre de nucleasas como solvente (Moreno *et al.*, 2014; Thompson & Black, 2007). Ya que se tienen todos los componentes necesarios en un único tubo, la mezcla se somete a varios ciclos específicos de temperaturas, estos ciclos se clasifican en tres fases:

Desnaturalización: En esta etapa se calienta la mezcla de reacción a una temperatura aproximada de 95°C (esta temperatura dependerá de las características del DNA molde) la cual permite que el DNA se separe en dos cadenas sencillas (Bolivar *et al.*, 2014; Thompson & Black, 2007).

Alineamiento: La temperatura se disminuye entre 50 y 60°C, esta temperatura depende de varios factores: la concentración de cebadores, el número de pares de bases de los cebadores, el porcentaje de guaninas y citosinas, lo que permitirá que los cebadores se unan a la cadena molde de DNA monocatenario antes que la molécula de DNA pueda renaturalizarse (Serrato *et al.*, 2014; Thompson & Black, 2007).

Extensión: Ya que los cebadores se han unido al DNA monocatenario, se aumenta la temperatura a 72°C, que es la temperatura óptima para que la enzima Taq polimerasa comience a añadir los nucleótidos a los extremos 3' de los cebadores utilizando la cadena de DNA original como molde. Como resultado se obtiene una copia exacta de la molécula de

DNA (amplicón) definida por los cebadores (Bolívar *et al.*, 2014; Thompson & Black, 2007).

La amplificación depende de la cantidad de DNA añadido a la reacción por lo que cada muestra debe ajustarse a un nivel que permita que la PCR funcione de manera óptima. Para llevar a cabo la amplificación de manera más rápida se puede utilizar un kit comercial de PCR (Butler, 2007; Hawkins & Guest, 2017)(Jamieson & Bader, 2016).

Los kits comerciales que usan como marcadores a los *loci* de STR utilizan la PCR múltiple, que es la amplificación de dos o más regiones de DNA de manera simultánea. Para que esta PCR funcione, sus cebadores deben ser compatibles y además el diseño de estos para los STR es un reto ya que se deben ajustar para generar una buena separación entre los *loci* marcados con el mismo color, además deben producir picos de buen tamaño por lo que en general la optimización de este tipo de PCR es muy compleja (Ranjan *et al.*, 2018).

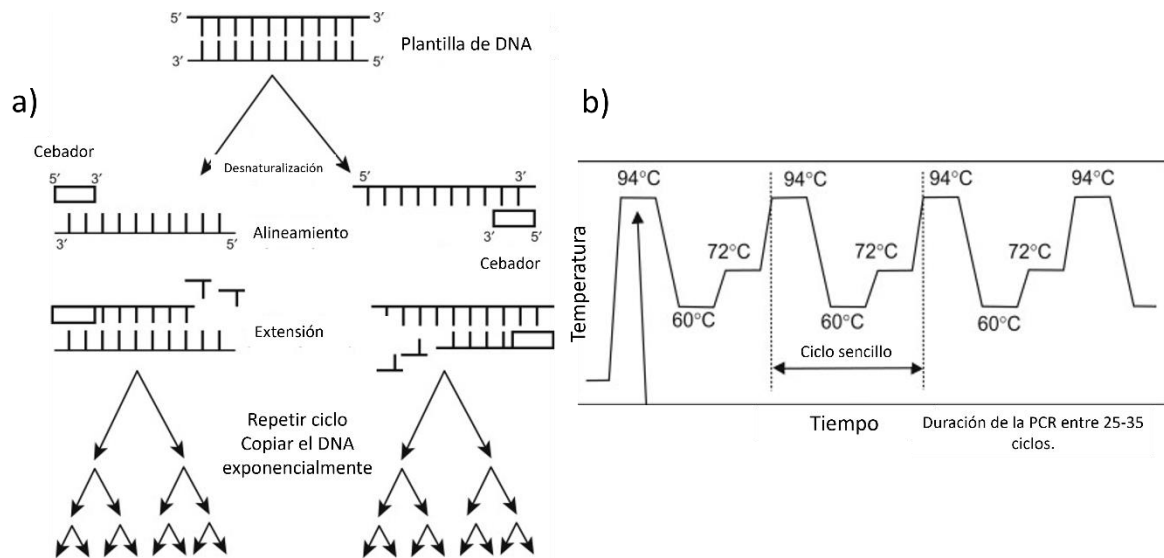


Figura 7. Proceso de la PCR. a) Proceso de amplificación del DNA con la reacción en cadena de la polimerasa, b) Perfil de temperaturas de termociclado para PCR. Modificado de Butler, 2010.

Para finalizar se mide la longitud del amplicón para determinar el número de repeticiones en cada alelo, ya que la PCR múltiple produce una mezcla de moléculas de DNA, la técnica más utilizada para hacer esta separación es la electroforesis capilar (CE), que ofrece como ventajas: velocidad, resolución y

sensibilidad (Smith & Nelson, 2003). El proceso de separación ocurre dentro de un capilar de sílice fundida en presencia de un voltaje alto. Los capilares se llenan con un gel de poliacrilamida que permite que los fragmentos se separen en función de su tamaño, la señal es transmitida a través de una sección del capilar que se inserta en la trayectoria de luz del detector UV (Smith & Nelson, 2003). Al realizar la CE se registra el color del tinte (cada amplicón es marcado con fluorescencia durante la PCR esto se debe a que los cebadores están unidos a un colorante fluorescente) y el tiempo de migración de los fragmentos de DNA en relación con un estándar interno, así se puede determinar el tamaño de cada alelo STR (Imle , 2005; Smith & Nelson, 2003).

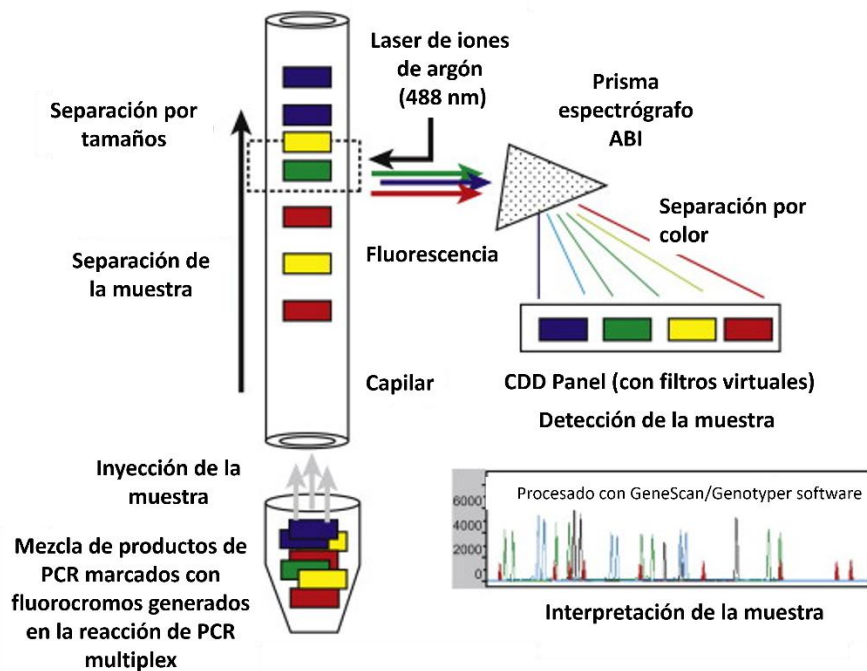


Figura 8. Separación de alelos STR. Pasos de la electroforesis capilar y sus componentes. Modificada de Butler, 2012.

Al terminar la CE es posible hacer el análisis de los datos mediante algunos *softwares* como *GeneMapper* o *GeneScan* que convierten un archivo de computadora a una representación electrónica de la información del DNA obtenida, estos facilitan la identificación de picos primarios, el cálculo del tamaño de los picos identificados y la asignación de los alelos (Imle , 2005). La tipificación de los STR por lo general se lleva a cabo con escaleras alélicas y un estándar de tamaño que poseen los alelos más comunes y que se han secuenciado para revelar el número real de repeticiones (Butler, 2006). Los

datos obtenidos son guardados en un archivo y luego la información se recopila para cada locus para determinar el perfil general de los marcadores STR (Butler, 2015).

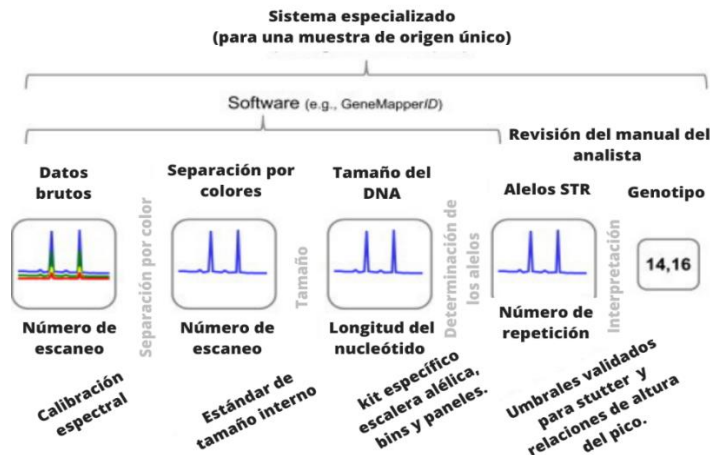


Figura 9. Transformación de la información de una muestra en un locus STR, modificado de Butler, 2015.

2.6. APLICACIONES DEL ANÁLISIS DE STRS

Para identificar los marcadores STR algunas compañías como Promega Corporation y Applied Biosystems, entre otras, han desarrollado kits comerciales que permiten la amplificación de algunos marcadores STR en específico (marcados con fluorescencia), y que hacen posible obtener resultados de manera rápida con reacciones de PCR múltiples, que permiten la amplificación de los *loci* de STR en una sola reacción junto con algunos marcadores adicionales como los de AMELXY (marcador de tipificación sexual) (Butler, 2007).

El análisis de STR permite determinar parentesco biológico como en el caso de los estudios de paternidad o maternidad. En este tipo de estudios solo pueden existir dos tipos de resultados: compatibilidad o no compatibilidad, ya que un hijo hereda la mitad de los alelos de la madre y la otra mitad del padre (Vázquez, 2022). En el caso de la identificación de personas se puede llevar a cabo a través de la comparación de un perfil obtenido de un indicio biológico (muestras de cabello, semen, sangre, saliva, tejido, etc.) y compararlo con una muestra de referencia. Es importante mencionar que a través de estudios de parentesco biológico también se pueden llevar a cabo identificaciones

humanas. En el caso de identificación de restos humanos, por el tipo de muestra que se analiza, puede ser compleja la identificación principalmente porque se pueden obtener perfiles parciales o incompletos (Vázquez, 2022). También el análisis de STR permite la creación de bases de datos poblacionales, lo que permite ser emplearlas para hacer los cálculos de los parámetros estadísticos forenses, y poder llevar a cabo las identificaciones (Mendoza & Pérez, 2023; Torres *et al.*, 2022; Vázquez, 2022).

2.7. KITS COMERCIALES DE PCR MÚLTIPLE PARA LA TIPIFICACIÓN DE STR.

Los kits comerciales de PCR múltiple (permite la amplificación de dos o más regiones simultáneamente añadiendo más de un conjunto de cebadores a la reacción (Jamieson & Bader, 2016), proporcionan cebadores premezclados (cebadores específicos de las regiones de los STRs que se quieren analizar) y una mezcla maestra que contiene: una DNA polimerasa, los cuatro dNTPs, tampón de PCR, y cebadores , además contiene una escalera alélica y un estándar de tamaño que simplifican la generación de los perfiles de STR (Bolivar *et al.*, 2014; Shrivastava *et al.*, 2020).

Los laboratorios prefieren utilizar los kits comerciales en vez de ensayos internos porque tienen diversas ventajas como:

- Simplifican y estandarizan los procedimientos.
- Eliminan el trabajo arduo del control de calidad de los componentes de la PCR.
- La obtención de resultados es muy rápida.
- Se pueden realizar amplificaciones con 1ng o menos de DNA.
- Permiten la consistencia en el uso de marcadores y la nomenclatura de alelos en el laboratorio.
- Permiten compartir datos entre laboratorios.
- Ayudan a mejorar el control de calidad.

- Permiten la amplificación simultánea de más de 20 *loci* de STR (Butler, 2010).

Además, los kits con la escalera alélica y el estándar de tamaño, obtenidos de una secuenciación de DNA, logran calibrar los tamaños de los productos de PCR al número de repeticiones del STR para hacer un genotipado y se puede realizar un análisis de resultados a través de algunos *softwares* desarrollados especialmente para ellos (Butler, 2007; Shrivastava *et al.*, 2020).

Los kits de PCR múltiple de STR varían según los cebadores que utilizan para la amplificación, así como los *loci* STR incluidos, las combinaciones de colorantes fluorescentes y la escalera alélica utilizada (Jamieson & Bader, 2016; Shrivastava *et al.*, 2020).

En la tipificación del DNA, se examinan diferentes marcadores, entre más marcadores se comparen, mayor será la probabilidad de que dos individuos no relacionados tengan genotipos diferentes y por el contrario cada dato que coincida aumenta la confianza de que dos perfiles sean del mismo individuo (Shrivastava *et al.*, 2020; Vázquez, 2022).

La mayoría de los kits comerciales para la tipificación de STR usan como base los 13 *loci* STR de CODIS (lo que les permite el intercambio nacional e internacional de los perfiles de DNA (Jamieson & Bader, 2016; Shrivastava *et al.*, 2020), aunque esta cantidad de marcadores no es suficiente para lograr discriminar entre perfiles genéticos ya que entre más marcadores de DNA examinados y comparados, mayor es la probabilidad de que dos individuos no emparentados tengan genotipos diferentes, por lo que las compañías han desarrollado *kits* con más de 20 marcadores que permitan tener un mayor poder de discriminación (Butler, 2010; Shrivastava *et al.*, 2020).

2.8. BASES DE DATOS DE DNA POBLACIONALES Y PARA LA IDENTIFICACIÓN HUMANA

Las bases de datos en genética forense se apoyan de programas de cómputo, los cuales son capaces de almacenar perfiles genéticos y permitir que sean consultados rápidamente según las características que se deseen (Torres *et*

al., 2022). En la genética forense existen bases de datos forenses y bases de datos poblacionales.

Existen algunas bases de datos forenses importantes entre las cuales se encuentran: CODIS (Combined DNA Index System) perteneciente al laboratorio del FBI y ESS (European Standard Set) en donde se establecieron algunos de los marcadores STR básicos ya que cuando se analizan los 13 *loci* de STR de CODIS la probabilidad media de coincidencia al azar es inferior a una entre un trillón entre individuos no relacionados (Butler, 2005; Shrivastava *et al.*, 2020). Últimamente se ha mejorado la eficacia de las bases de datos expandiendo los perfiles genéticos, no solo limitándose a la comparación con los *loci* centrales como CODIS, sino incluyendo repeticiones cortas en tándem del cromosoma Y (Y-STR), DNA mitocondrial (mtDNA), variantes de nucleótido único, etc. (Butler, 2006; Shrivastava *et al.*, 2020).

Hoy en día las bases de datos de DNA se han convertido en una herramienta muy importante para la genética forense, ya que permiten la comparación de perfiles de DNA de escenas del crimen, personas desaparecidas o para algún otro caso de investigación, además permiten a los laboratorios compartir la información de todos estos perfiles (Vázquez, 2022; Wickenheiser, 2022). El objetivo de las bases de datos forenses es la identificación humana, permitiendo así vincular la evidencia biológica encontrada en una escena del crimen con el perfil de un sospechoso. Cuando la evidencia y el perfil sospechoso coinciden puede apelarse a que alguien más podría tener el mismo perfil de DNA, por esta razón es necesaria la evaluación estadística la cual permite evaluar la frecuencia del perfil genético en la población y así evitar asociaciones al azar (Rangel, 2020). Para ello, se requieren de bases de datos poblacionales.

Las bases de datos poblacionales se utilizan para estimar probabilidades de coincidencias al azar a través de frecuencias alélicas con un aproximado de entre 100 a 200 muestras de individuos que son seleccionados para representar a un grupo de interés en específico y también permiten la caracterización genética de una región geográfica (Butler, 2010). Las bases de datos poblacionales proporcionan un respaldo estadístico ya que las

frecuencias alélicas permiten el cálculo de parámetros estadísticos como: Probabilidad de coincidencia al azar (PM), Poder de discriminación (PD), Heterocigosidad esperada (GD o H), Contenido de información del polimorfismo (PIC), Poder de exclusión (PE) y el índice típico de paternidad (TPI) (Butler, 2010; Gouy & Zieger, 2021).

La eficacia de una base de datos poblacional aumenta proporcionalmente a medida que se incremente el número de individuos en esta. Sin embargo, es importante considerar que su conformación requiere de mucho tiempo y una inversión de dinero elevada. La base de datos “ideal” sería una que esté constituida por el perfil de cada individuo de la sociedad (Butler, 2005; Wickenheiser, 2022).

2.9. GENÉTICA POBLACIONAL

Una población es un grupo de individuos de la misma especie que viven en la misma área geográfica por un tiempo determinado donde cada uno puede aparearse aleatoriamente y tener descendencia fértil (Castillo *et al.*, 2019). La genética de poblaciones estudia en forma cuantitativa las variaciones genéticas, los cambios en las frecuencias alélicas y genotípicas en las poblaciones. Además, tiene como objetivo caracterizar la diversidad genética de un conjunto de poblaciones observando que tan relacionados están los individuos en función de las variaciones en su genoma (Nussbaum *et al.*, 2008).

Se puede utilizar la genética de poblaciones para investigaciones criminales (haciendo coincidir muestras de DNA), para comprobar parentesco como en las pruebas de paternidad o también para la identificación de una persona extraviada (Gouy & Zieger, 2021).

Para la genética de poblaciones es importante obtener una buena caracterización de los alelos en diferentes *loci* en la población de interés para poder obtener los parámetros forenses mencionados anteriormente (Hernández & Trejo, 2014).

La importancia de calcular los parámetros estadísticos forenses anteriores radica en que, aunque dos perfiles de DNA sean idénticos se puede considerar

como hipótesis alternativa que la muestra no sea parte de la misma persona y que provenga de otro individuo de la población. Lo mismo ocurre en los estudios de paternidad en los que se puede considerar que los alelos paternos del hijo no provienen del padre dubitado sino de otro individuo de la población. En estos casos se debe estimar la probabilidad de que el perfil genético o los alelos pertenezca a la población en general (Peral *et al.*, 2015).

Para poder calcular la probabilidad de que un individuo seleccionado al azar de la población tenga el mismo perfil genético que el de una muestra de DNA problema se necesita conocer la frecuencia del genotipo en la población. Sin embargo, es poco factible conocer la frecuencia de cada genotipo en una población, por lo que se llevan a cabo estimaciones basadas en las frecuencias alélicas observadas en una muestra poblacional. Para ello se emplea un estimado como la fórmula del equilibrio de Hardy Weinberg (Peral *et al.*, 2015). A partir de las frecuencias obtenidas se pueden calcular parámetros forenses que respaldan estadísticamente los resultados obtenidos. Un ejemplo es la probabilidad de coincidencia al azar (PM), que es la probabilidad de que dos muestras elegidas al azar de la población tengan el mismo perfil genético, al obtener un índice pequeño en este parámetro indicaría que la probabilidad de que dos perfiles genéticos coincidan disminuye (Peral *et al.*, 2015).

2.9.1. FRECUENCIAS GENOTÍPICAS Y ALÉLICAS

La **frecuencia alélica (FA)** es la proporción relativa de los alelos de una población y se expresa en porcentaje o en la unidad (Hernández & Trejo, 2014), y se puede calcular con la siguiente fórmula:

$$FA = \frac{\text{número de alelos observados}}{2n}$$

Donde, n es el tamaño de la muestra analizada.

La **frecuencia genotípica (FG)** es la cantidad de veces que aparece un genotipo en la población (Hernández & Trejo, 2014) y se obtiene a partir de la siguiente fórmula:

$$FG = \frac{\text{número de genotipos observados}}{n}$$

Donde n, es el tamaño de muestra analizada.

2.9.2. PROBABILIDAD DE COINCIDENCIA AL AZAR

La probabilidad de coincidencia al azar en genética forense puede tener dos significados:

1. La probabilidad de coincidencia al azar (PCA) para dos perfiles idénticos: es la probabilidad de observar una coincidencia aleatoria entre dos individuos que no están relacionados y se puede obtener a través de la fórmula del equilibrio de Hardy Weinberg para homocigotos y heterocigotos, y después multiplicando las frecuencias alélicas referidas en bases de datos (regla del producto), esto es posible porque los alelos de cada locus se heredan de manera independiente.
2. La probabilidad de coincidencia al azar (PM) como parámetro forense es la probabilidad de que existan dos genotipos iguales en la población y se calcula con la siguiente fórmula:

$$PM = \sum_{i=a}^n xi^2$$

Donde, xi^2 es la frecuencia del genotipo i en un locus (Gouy & Zieger, 2021).

2.9.3. PODER DE DISCRIMINACIÓN (PD)

El poder de discriminación se define como la probabilidad de discriminar entre dos individuos no relacionados y se puede obtener con la siguiente formula:

$$PD = 1 - PM$$

Donde, PM= probabilidad de coincidencia al azar (Gouy & Zieger, 2021).

2.9.4. HETEROCIGOSIDAD ESPERADA (HE)

La heterocigosidad esperada es la probabilidad de que un individuo sea heterocigoto en un *locus* dado, esta depende de las frecuencias alélicas de la población. Se puede obtener a partir de la siguiente fórmula:

$$He = \frac{n}{n-1} \left(1 - \sum_{i=1}^m x_{ij}^2\right)$$

Donde, n=número de *loci* considerados, m= número de alelos en el locus j y x_{ij} son las frecuencias del alelo i en el locus j (Gouy & Zieger, 2021).

2.9.5. HETEROCIGOSIDAD OBSERVADA (Ho)

Es la proporción de individuos heterocigotos en la población (Gouy & Zieger, 2021). Se calcula de la siguiente manera:

$$Ho = \frac{\text{número de muestras con genotipos heterocigotos}}{\text{número total de muestras}}$$

2.9.6. HOMOCIGOSIDAD OBSERVADA (H)

Es la proporción de individuos homocigotos en la población (Gouy & Zieger, 2021). Esta se calcula de la siguiente manera:

$$h = \frac{\text{número de muestras con genotipos homocigotos}}{\text{número total de muestras}}$$

2.9.7. CONTENIDO DE INFORMACIÓN POLIMÓRFICA (PIC)

El contenido de información polimórfica es la probabilidad de que los alelos maternos y paternos de un individuo sean deducibles o la probabilidad de poder deducir qué alelo ha transmitido un padre al hijo. El PIC se puede obtener a partir de la siguiente fórmula:

$$PIC = 1 - \sum_{i=1}^n PM^2 - \left(\sum_{i=1}^n PM^2\right)^2 \sum_{i=1}^n PM^4$$

Donde, PM= probabilidad de coincidencia al azar (Gouy & Zieger, 2021).

2.9.8. PODER DE EXCLUSIÓN (PE)

El poder de exclusión se define como la fracción de individuos que tienen un perfil de DNA que es diferente al de un individuo seleccionado al azar y se puede obtener a partir de la siguiente fórmula:

$$\text{Poder de exclusión} = Ho^2 (1 - 2Hoh^2)$$

Donde Ho= Heterocigosidad observada y h= Homocigosidad observada (Gouy & Zieger, 2021).

2.9.9. ÍNDICE TÍPICO DE PATERNIDAD (TPI)

El índice típico de paternidad mide cuántas veces es más probable que un posible padre sea el padre real que un hombre seleccionado al azar en cierta población (Gouy & Zieger, 2021).

$$TPI = \frac{1}{2h}$$

Donde h= Homocigotos observados en la población.

2.9.10. LEY DEL EQUILIBRIO DE HARDY-WEINBERG

El equilibrio de Hardy-Weinberg (HWE), es un concepto fundamental en genética de poblaciones y la biología evolutiva porque describe la relación de las frecuencias de alelos y genotipos y establece que las frecuencias de los genotipos en una población permanecen constantes entre generaciones en ausencia de perturbaciones por factores externos como: mutaciones, selección natural, deriva génica, entre otros (Abramovs *et al.*, 2020; Lachance, 2016). Este principio se puede llevar a cabo si se tienen las siguientes características: a) el apareamiento aleatorio, b) la ausencia de selección natural, c) una población muy grande (la deriva genética es insignificante), d) no hay flujo de migración de genes, e) no hay mutación y f) el *locus* es autosómico, si se llegan a violar estas condiciones entonces se dice que el equilibrio de Hardy-Weinberg ha tenido una desviación (Lachance, 2016). Es evidente que ninguna población humana cumple con todos los criterios, aunque la mayoría de las poblaciones se comportan como si estuvieran cerca del HWE (aunque no cumplan todos los supuestos). Si la población se encuentra en equilibrio de

Hardy-Weinberg es posible realizar una comparación entre los genotipos heterocigotos observados y los esperados (Goodwin *et al.*, 2011).

Cuando una población se encuentra en equilibrio de Hardy-Weinberg, las frecuencias genotípicas pueden predecirse a partir de las frecuencias alélicas. La frecuencia del genotipo de un homocigoto puede calcularse utilizando p^2 y la de los heterocigotos puede calcularse utilizando $2pq$ (Lachance, 2016). Cuando ya se tienen los valores de los genotipos esperados y observados se puede evaluar si el locus se encuentra en HWE, esto se lleva a cabo a partir de la prueba de chi-cuadrada (χ^2) siguiendo la siguiente formula:

$$\chi^2 = \sum \left(\frac{\text{Heterocigotos observados} - \text{Heterocigotos esperados}}{\text{Heterocigotos esperados}} \right)^2$$

El objetivo de la prueba de χ^2 es determinar si el conjunto de datos observados difiere significativamente del conjunto de datos esperado, ya obtenido el resultado estadístico este debe ser mayor al nivel de significancia (α) que generalmente es un valor de 0.05, en este caso se considera que las frecuencias observadas se ajustan a las frecuencias esperadas (los alelos de los *loci* son independientes entre sí) (Butler, 2010; Goodwin *et al.*, 2001; Linacre & Tobe, 2013).

2.9.10.1. CORRECCIÓN DE BONFERRONI

Antes de aplicar la corrección de Bonferroni se deben de conocer las pruebas de hipótesis para la significación estadística. En el caso de una identificación forense la hipótesis nula (H_0), indicaría que el perfil genético si pertenece al individuo, mientras que la hipótesis alternativa (H_1), indica que el perfil genético pertenece a otro individuo dentro de la población, estas hipótesis se plantean para representar condiciones excluyentes por lo que, si la hipótesis nula es rechazada, la hipótesis alternativa debe ser aceptada (Goodwin *et al.*, 2011).

En las hipótesis planteadas también se pueden generar errores al ser aceptadas o rechazadas, hay dos tipos de errores: el error tipo I, en el que se rechaza la hipótesis nula cuando en realidad es verdadera generando un falso negativo y el error tipo II, que consiste en aceptar la hipótesis nula cuando en realidad es falsa generando un falso positivo (Goodwin *et al.*, 2011).

El nivel de significancia representado por α , representa la probabilidad de cometer un error tipo I, los valores de alfa comúnmente utilizados son: 0.05 que equivale a un límite de confianza del 95% y 0.01 que equivale a un límite de confianza del 99%. Al seleccionar el nivel de significancia estamos también estableciendo la probabilidad de cometer un error de tipo I. De acuerdo con lo anterior, la obtención de un valor $p < 0.05$ provocará el rechazo de la hipótesis nula, mientras que los valores de $p > 0.05$ conducen a su aceptación (Butler, 2010; Goodwin *et al.*, 2011).

Cuando se realizan análisis múltiples simultáneamente para varias pruebas estadísticas independientes se puede aplicar la corrección de Bonferroni que se basa básicamente en disminuir un valor alfa determinado, ya que un valor alfa para cada comparación individual es suficiente, pero no para todo un conjunto de datos. Se establece que, si una hipótesis nula es verdadera y se usa $p < 0.05$ como criterio para todas las pruebas, se observa una diferencia significativa de 1 en cada 20 pruebas (Armstrong, 2014). El valor de alfa se reduce teniendo en cuenta el número de comparaciones que se realizan dividiendo n (número de comparaciones) entre el valor alfa de cada comparación, con esto el nivel de significancia ajustado queda de la siguiente manera:

$$1 - (1 - \alpha)^{\frac{1}{n}} = \alpha/n$$

Por lo tanto, en el análisis de datos para el equilibrio de Hardy-Weinberg la corrección de Bonferroni ayuda a eliminar algún *locus* que esté por debajo del nivel de significancia (Butler, 2005; Lesack & Naugler, 2011).

2.9.11 EJEMPLO DE UNA IDENTIFICACIÓN CUANDO LOS PERFILES COINCIDEN.

Tenemos un perfil genético que coincide con un perfil obtenido de un indicio biológico en un lugar de investigación, la coincidencia se establece realizando una comparación entre los alelos de cada marcador (esto se hace de manera automática en la base de datos), a continuación, en la Tabla 2 se muestra un ejemplo representativo:

Tabla 2. Ejemplo representativo de la comparación de dos perfiles genéticos con cuatro marcadores STR.

Marcador	Muestra 1	Muestra 2
TH01	10 , 6	10 , 6
D8S1179	10	10
vWA	19 , 14	19 , 14
TPOX	8 , 13	8 , 13

Al encontrar una coincidencia entre los dos perfiles se calcula la razón de verosimilitud o *likelihood ratio* que es la relación entre dos hipótesis que compiten entre sí (Vázquez, 2022).

H_0 = El DNA presente en la evidencia pertenece al individuo (Hipótesis de la fiscalía).

H_1 = El DNA presente en la evidencia pertenece a otro individuo de la población (Hipótesis de la defensa).

Y se define como la probabilidad (Pr) de que las pruebas de DNA dada la hipótesis planteada por la fiscalía (H_0) o por la defensa (H_1) y se representa de la siguiente manera (Vázquez, 2022).

$$\frac{\Pr(H_0)}{\Pr(H_1)}$$

La probabilidad de ocurrencia bajo la hipótesis de la defensa es igual a la probabilidad de observar el perfil si su origen fuera alguien distinto del individuo sospechoso en la población a la que pertenece este. Es así que la razón de verosimilitud se calcula a partir del inverso de la probabilidad de coincidencia al azar (PCA) (Goodwin *et al.*, 2011; Vázquez, 2022):

$$Likelihood Ratio = \frac{1}{PCA}$$

Tabla 3. Ejemplo hipotético del cálculo de PCA y LR en cuatro marcadores STR. Las frecuencias utilizadas se encuentran reportadas por Hernández y Trejo, 2014.

Marcador STR	TH01	D8S1179	vWA	TPOX			
Genotipo	10 , 6	10	19 , 14	8 , 13			
HWE	2pq	p ²	2pq	2pq			
cálculo de HWE	2(0.0111)(0.2491)	(0.1439) ²	2(0.0535)(0.0738)	2(0.5055)(0.0018)			
Proporción del genotipo	0.00553	0.02071	0.00790	0.00182			
Regla del producto			Probabilidad de coincidencia al azar (PM) =			1.64555E-09	
Índice de verosimilitud (LR)			LR= 1/PM =			607698144	Se expresa como 1 en 607698144

Y para calcular la razón de la fuerza de verosimilitud en términos de la fortaleza se consideran los intervalos de la Tabla 4.

Tabla 4. Intervalos de la fuerza del valor del índice de verosimilitud (Butler, 2005).

Valor de índice de verosimilitud	Fortaleza de la evidencia
1 -- 10	Soporte bajo
11 -- 100	Soporte moderado
101 -- 1000	Soporte fuerte
> 1000	Soporte muy fuerte

El resultado se reporta en un dictamen de la siguiente manera:

“Se determina que el individuo “x” no se excluye como el origen biológico de la muestra de DNA, por lo que se realizó un análisis estadístico para estimar la frecuencia de los perfiles genéticos dentro de la población, así como el índice de verosimilitud para contrastar la hipótesis nula y la hipótesis alternativa. Con base en lo anterior, se establece que comparten información genética entre sí,

además de presentar la misma frecuencia estadística siendo esta de 1.6455×10^{-9} dentro de la población de estudio, con un índice de verosimilitud de 607698144 que indica que la fortaleza de la prueba genética “Soporta muy fuertemente” la hipótesis de que el individuo “x” es el origen biológico de la muestra de DNA obtenida” (Butler, 2010; Goodwin *et al.*, 2011).

3. HIPÓTESIS

La conformación de una base de datos poblacional, basada en STRs autosómicos de una población de 50 individuos empleando el kit PowerPlex Fusion 6C, permitirá calcular parámetros forenses confiables para llevar a cabo identificaciones humanas.

4. OBJETIVOS

4.1. OBJETIVOS GENERALES

Realizar un estudio poblacional de STR autosómicos en población mestizo-mexicana empleando el kit PowerPlex Fusion 6C.

4.2. OBJETIVOS PARTICULARES

- Determinar las frecuencias alélicas de los 23 marcadores autosómicos del kit comercial PowerPlex Fusion 6C en 50 individuos mestizos mexicanos de la ciudad de México.
- Llevar a cabo los cálculos de genética de poblaciones empleados en la genética forense para evaluar la utilidad de este estudio poblacional en la identificación humana.
- Comparación de los perfiles obtenidos a través del kit PowerPlex fusión 6C con otros sistemas de perfilado de STR por electroforesis capilar y por secuenciación de última generación.

5. PROCEDIMIENTO EXPERIMENTAL

5.1. TIPO DE ESTUDIO

Se trata de un estudio observacional, prospectivo, transversal y comparativo

5.2. POBLACIÓN DE ESTUDIO

En el presente estudio se obtuvieron muestras de sangre periférica de mujeres y hombres mestizos mexicanos, no relacionados biológicamente, pertenecientes a la Ciudad De México.

Tabla 5. Datos poblacionales de la Ciudad de México y de los Estados Unidos Mexicanos (INEGI, 2021a; INEGI, 2021b; INEGI, 2021c; INEGI, 2021d).

	Ciudad De México	Estados Unidos Mexicanos
Población total (2020)	9,209,944	126 014 024
Población total hombres (2020)	4,404,927	64 540 634
Población total mujeres (2020)	4,805,017	61 473 390
Porcentaje de población emigrante (2020)	557 181	21,611,963
Porcentaje de población inmigrante (2020)	308 686	21,611,963
Densidad de población (habitantes por kilómetro cuadrado, 2020)	6 163.3	64
Tasa de crecimiento total (2020)	0.4	1.1

5.3. CARACTERÍSTICAS DE LAS MUESTRAS

Las muestras utilizadas para este estudio (50 muestras de sangre obtenidas a partir de punción venosa) pertenecen a la genoteca resguardada en el laboratorio de Genética en Investigación de la Licenciatura de Ciencia Forense, creada para el proyecto “Análisis del genoma mitocondrial para la identificación forense en México”, cuyo responsable es el Dr. José Mariano Guardado Estrada y el cual fue avalado por la división de investigación y comisión de ética de la Facultad de Medicina bajo el número de registro 65/2022 (Anexo 1).

5.4. CRITERIOS DE INCLUSIÓN, EXCLUSIÓN Y ELIMINACIÓN

5.4.1. INCLUSIÓN

Muestras de sangre periférica de mujeres y hombres no relacionados biológicamente, pertenecientes a la Ciudad De México.

5.4.2. EXCLUSIÓN

Individuos relacionados biológicamente

5.4.3. ELIMINACIÓN

Muestras contaminadas

5.5. MATERIALES Y MÉTODOS.

5.5.1. MATERIAL

- Termociclador Axygen® MaxyGene II
- Analizador genético 3500 Hitachi
- Campana de bioseguridad
- Micropipeta Eppendorf de 10 µL
- Micropipeta Eppendorf de 200 µL
- Micropipeta Eppendorf 1000 µL
- Puntas para micropipeta (10 µL, 200 µL y 1000 µL)
- Placa para PCR de 96 pozos

- Vortex
- Guantes de nitrilo
- Tubos Eppendorf de 0.5 ml y 1.5 ml
- Polímero POP-7
- PowerPlex® Fusion 6C System
- Hi-Di™ Formamide
- Centrífuga

5.5.2. LOC/INCLUIDOS EN ESTE ESTUDIO

El *kit PowerPlex® Fusion 6C* es un sistema de 6 colores que permite la co-amplificación y la detección fluorescente de 18 *loci* autosómicos pertenecientes a la base de datos CODIS (CSF1PO, FGA, TH01, vWA, D1S1656, D2S1338, D2S441, D3S1358, D5S818, D7S820, D8S1179, D10S1248, D12S391, D13S317, D16S539, D18S51, D19S433 y D21S11), así como la amelogenina y el DYS391 que se utilizan para la determinación del sexo. También se encuentran los marcadores: Penta D, Penta E, D22S1045, TPOX y SE33, que se incluyen para aumentar la discriminación y permitir la búsqueda en bases de datos que incluyan perfiles con estos *loci*. Por último, contiene *loci* Y-STR: DYS570 y DYS576. Este panel está diseñado para satisfacer las recomendaciones del CODIS y del ESS (Promega, 2018). Para el desarrollo de este análisis solo se utilizarán los *loci* autosómicos siguientes:

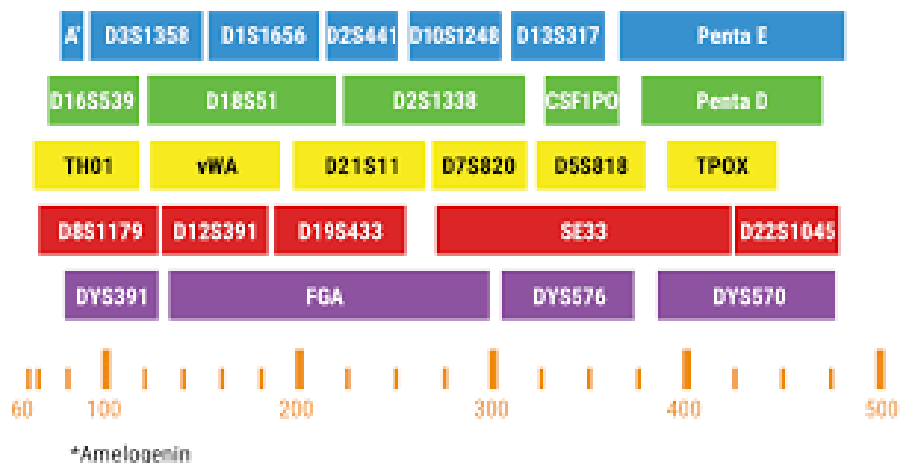


Figura 10. Loci de estudio pertenecientes al *kit PowerPlex Fusion 6C*. En la parte inferior se muestra la escala en pares de bases (pb). Los marcadores STR están ordenados por tamaño y color de los fluorocromos: FL-6C (azul), JOE-6C (verde), TMR-6C (amarillo), CXR-6C (rojo) y TOM-6C (violeta).

5.5.3. MÉTODOS

5.5.3.1. OBTENCIÓN DE LA MUESTRA

Las muestras utilizadas para este estudio provienen de la genoteca resguardada en el laboratorio de genética en investigación de la Licenciatura en Ciencia Forense, creada a partir del proyecto “Análisis del genoma mitocondrial para la identificación forense en México”, cuyo responsable es el Dr. José Mariano Guardado Estrada.

La recolección de las muestras se realizó bajo el consentimiento informado del proyecto anteriormente mencionado, en el cual se autoriza que las muestras pueden ser almacenadas y utilizadas en investigaciones futuras por el laboratorio de genética de la licenciatura en ciencia forense (LCF) de la UNAM.

Las muestras fueron obtenidas de personas mestizo-mexicanas que son personas nacidas en México, resultado de la mezcla entre nativos americanos y españoles, con apellido de origen español y antepasados mexicanos tres generaciones atrás (Aguilar & Rangel , 2020; Rangel *et al.*, 2010) de la Ciudad de México por personal del laboratorio de genética de la LCF. Para asegurarse de que los sujetos de estudio cumplieran con los criterios antes mencionados se les preguntó el lugar en donde vivían y el lugar de nacimiento de madre, padre, abuelos y bisabuelos.

Se realizaron tomas de muestra de sangre periférica en tubos vacutainer con EDTA y se utilizaron las muestras LCF001-LCF050.

La extracción del DNA genómico de las muestras fue realizada por el Dr. Christian Cárdenas y la Dra. Alexa Villavicencio a partir de sangre total, realizando una lisis de glóbulos rojos utilizando el tampón RBCLA y posteriormente el kit DNeasy blood & tissue (Qiagen, Ciudad De México), siguiendo el protocolo del fabricante (Qiagen, 2020). Estas muestras fueron almacenadas a -20 °C.

El desarrollo de la metodología a partir de la cuantificación se llevó a cabo en el laboratorio de genética del Instituto de Servicios Periciales y Ciencias forenses (INCIFO) de la ciudad de México.

5.5.4. CUANTIFICACIÓN

El volumen de las muestras proporcionado por el laboratorio de genética de la LCF fue inicialmente de 5 μ L, al cual se le realizó una dilución con 15 μ L de agua libre de nucleasas, para asegurar que el volumen fuera suficiente para el análisis completo de la muestra. Posteriormente, se realizó una cuantificación de DNA utilizando el *kit investigator Quantiplex* (Qiagen, Ciudad De México), siguiendo los siguientes pasos de acuerdo con el protocolo (Qiagen, 2012):

1. Primero se descongeló la solución tampón para la dilución de ácidos nucleicos *QuantiTect* (Qiagen, Ciudad De México). Antes de utilizarse debe mezclarse muy bien.
2. Ya que las muestras de DNA se encontraban en almacenamiento en refrigeración, se sacaron para descongelarlas.
3. Se prepararon diluciones en serie frescas del DNA control Z1 conforme a la tabla 6, pertenecientes a la curva estándar (se agrega por duplicado en la placa), se mezclaron mediante agitación en vórtex por al menos 5 segundos y después se centrifugaron brevemente antes de extraer una cantidad de alícuota para la siguiente dilución (se debe utilizar una nueva punta de pipeta para la siguiente dilución).

Tabla 6. Diluciones en serie del DNA de control Z1 para la curva estándar. *NTC (control negativo, solo se le agrega tampón) (Qiagen, 2012).

Estándar (STD)	Dilución en serie del DNA de control Z1 (ng/ μ L)	DNA de control Z1 (μ L)	Tampón para dilución de ácidos nucleicos QuantiTect (μ L)
STD1	20	DNA no diluido	-
STD2	5	10	30
STD3	1.25	10	30
STD4	0.3125	10	30
STD5	0.078125	10	30
STD6	0.01953125	10	30
STD7	0.004882813	10	30
NTC	-	-	2

- Después se descongelaron los ácidos nucleicos de plantilla y se mezclaron la mezcla de reacción FQ y la mezcla de cebador IC YQ invirtiendo los tubos varias veces.
- Se preparó una mezcla maestra conforme a la tabla 7, con un volumen 10% mayor al que se requiere para el número total de reacciones (incluyendo los controles) para considerar el error de pipeteo.

Tabla 7. Mezcla de reacción para cuantificación escalada a $\frac{1}{4}$ X. Se escaló la reacción a $\frac{1}{4}$ (ya estandarizada en el laboratorio).

Concentración final	Volumen de Mezcla de reacción FQ (μ L)	Volumen de Mezcla de cebadores IC YQ (μ L)	Volumen total de la mezcla (μ L)
1X	9	9	18
$\frac{1}{4}$ X	2.5	2.5	5

- Se agitó en vórtex la mezcla maestra y se virtieron 5 μ L en los pocillos de una placa de PCR.

7. Después se añadieron 2 μL de DNA control o 2 μL de DNA muestra a los pocillos individuales, las cantidades de agua libre de nucleasas indicadas en la tabla 8 y 2 μL de tampón solo para los pocillos de NTC.

Tabla 8. Volumen de agua y DNA para cada pozo.

Concentración final	Volumen de agua (μL)	Volumen de DNA (μL)
1X	-	2
1/4 X	5	2

8. Se mezclaron meticulosamente los componentes de la placa y se cerró.
9. Cuando la placa estaba lista se analizó utilizando el *Applied Biosystems 7500 Real-Time PCR System (Thermo Fisher, Ciudad De México)*, siguiendo las condiciones del protocolo de ciclado del fabricante que se encuentran en la tabla 9. Después se realizó el análisis de los resultados con el *software HID Real-Time PCR Analysis Software v1.5.1.(Thermo Fisher, Ciudad De México)*.

Tabla 9. Condiciones de ciclado de ciclado para el sistema de PCR en tiempo real Applied Biosystems 7500 del fabricante.

Condiciones de ciclado			
Paso	Temperatura $^{\circ}\text{C}$	Tiempo	número de ciclos
Paso inicial de activación de la PCR	95 $^{\circ}\text{C}$	3 min	-
Desnaturalización	95 $^{\circ}\text{C}$	5 s	40 ciclos
Alineación/extensión combinada	60 $^{\circ}\text{C}$	35 s	

Para identificar o descartar alguna contaminación de las muestras de estudio por parte del analista se obtuvo su perfil genético de los marcadores STR utilizados (Anexo 1) (Qiagen, 2012).

5.5.5. AMPLIFICACIÓN DE LOS STR

Para realizar la amplificación se utilizó el *kit PowerPlex® Fusion 6C System* de Promega (Ciudad de México). Para comenzar se sacaron las muestras de refrigeración y se les realizó una centrifugación y agitación en vórtex para que la cantidad de muestra que se encuentre en la tapa y en las paredes del tubo no se pierda. Después se rotularon tubos eppendorf de 0.5 µL para cada muestra de DNA y uno de 1.5 µL para preparar la mezcla de los reactivos (Promega, 2018).

Se determinó el número de reacciones junto con el control positivo (muestra de DNA ya estandarizada por el laboratorio) y negativo (agua de grado de amplificación), considerando una reacción más por el error de pipeteo. Al terminar los cálculos, se sacaron los reactivos del *kit* de refrigeración y se agitaron en un vórtex por aproximadamente 15 segundos.

En un tubo eppendorf de 1.5 µL se preparó una mezcla entre el PowerPlex® Fusion 6C 5X Master Mix y el PowerPlex® Fusion 6C 5X Primer Pair Mix y al terminar se les agitó en vortex. Después se preparó la mezcla de reacción, se escaló a ¼ de reacción (ya estandarizado en el laboratorio) de acuerdo con la tabla 10, y se añadieron las cantidades de DNA y agua según las tablas 11,12 y 12. Se agitó cada tubo en vórtex, se centrifugaron y se revisaron que estuvieran cerrados correctamente. Acto seguido se metieron los tubos al termociclador bajo las condiciones indicadas en las tablas 14 y 15 (Promega, 2018).

Tabla 10. Volúmenes para la mezcla de amplificación por PCR.

	1X	1/4 X
Componente de mezcla de amplificación	Volumen por reacción	Volumen por reacción
A un volumen final de 25		
Agua grado de amplificación	µL	Variable
PowerPlex® Fusion 6C 5X Master Mix	5 µL	1.2 µL
PowerPlex® Fusion 6C 5X Primer Pair Mix	5 µL	1.2 µL
DNA	Hasta 15 µL	Variable
Volumen total de reacción	25 µL	6.4 µL

Tabla 11. Volúmenes de agua y DNA utilizados para una concentración de 2ng/µL de DNA.

Concentración de 2 ng/µL			
Muestra	Volumen de agua (µL)	Volumen de DNA (µL)	Número de ciclos
LCF 033	1.5	2.5	27 ciclos
LCF 035	1.5	2.5	
LCF 036	2	2	
LCF 037	2.9	1.1	
LCF 046	2.2	1.8	
LCF 047	2.9	1.1	
LCF 048	2.6	1.4	
LCF 049	2.6	1.4	
LCF 050	2	2	

Tabla 12. Volúmenes de agua y DNA utilizados para una concentración de 4ng/μL de DNA.

Concentración de 4 ng/μL			
Muestra	Volumen de agua (μL)	Volumen de DNA (μL)	Número de ciclos
LCF 006	0	4	27 ciclos
LCF 010	0	4	
LCF 023	2.4	1.6	
LCF 031	2	2	
LCF 032	2	2	
LCF 034	1.5	2.5	
LCF 038	2.3	1.7	
LCF 041	2.3	1.7	
LCF 043	1.6	2.4	
LCF 044	2.3	1.7	
LCF 045	2.5	1.5	

Tabla 13. Volúmenes de agua y DNA utilizados para una concentración de 8ng/μL de DNA.

Concentración de 8 ng/μL			
Muestra	Volumen de agua (μL)	Volumen de DNA (μL)	Número de ciclos
LCF 001	3	1	25 ciclos
LCF 002	2	2	
LCF 003	2	2	
LCF 004	3	1	
LCF 005	3.1	0.9	
LCF 007	2.6	1.4	
LCF 008	2.5	1.5	
LCF 009	1.9	2.1	
LCF 011	2.9	1.1	
LCF 012	2.5	1.5	
LCF 013	3	1	
LCF 014	2.6	1.4	
LCF 015	2.9	1.1	
LCF 016	2.6	1.4	
LCF 017	2.7	1.3	
LCF 018	2.6	1.4	
LCF 019	2.8	1.2	
LCF 020	1.6	2.4	
LCF 021	2.3	1.7	
LCF 022	1.4	2.6	
LCF 024	1.5	2.5	
LCF 025	2.4	1.6	
LCF 026	2.4	1.6	
LCF 027	2.1	1.9	
LCF 028	2.5	1.5	
LCF 029	2.3	1.7	
LCF 030	2	2	
LCF 039	1.9	2.1	
LCF 040	1.9	2.1	
LCF 042	2	2	

Tabla 14. Condiciones de ciclado para PCR múltiplex con el *kit PowerPlex Fusion 6C*.

Incubación inicial	Desnaturalización/Alineación		Extensión	Tiempo sostenido	Número de ciclos
96 °C	96 °C	60°C	60°C	4°C	Variable
1 min	5s	1 min	10 min	∞	

Tabla 15. Ajuste del número de ciclos

Concentración de DNA (ng)	Número de ciclos
1 ng	28
2ng	27
4 ng	26
8 ng	25

5.5.6. ELECTROFORESIS CAPILAR

Después de realizar la amplificación las muestras pueden llevarse a refrigeración o pasar directamente al paso de electroforesis capilar. Se rotuló una placa de 96 pozos y se calcularon las cantidades de reactivos de acuerdo con el protocolo que indica el fabricante realizando una mezcla entre *WEN ILS 500* (estándar de tamaño interno) (Promega, Ciudad De México) y formamida de la siguiente manera:

$$[(0.5\mu\text{l WEN ILS 500}) \times (\# \text{ muestras})] + [(9.5\mu\text{l formamida}) \times (\# \text{ muestras})]$$

Después se agitó en vórtex la mezcla (con cantidades ya estandarizadas anteriormente en el laboratorio) y se agregó en cada pocillo 12 μL de la mezcla de formamida/ estándar interno, en el primer pocillo se agregó 1 μL de PowerPlex® Fusion 6C Allelic Ladder Mix y en los pocillos restantes 1 μL de las muestras amplificadas de DNA, evitando en todo momento hacer burbujas dentro de la placa y por último se cerró la placa correctamente (Promega, 2018).

Después se realizó una hoja de corrida en el software Data Collection v3 y la placa se mete al analizador genético 3500 Hitachi para comenzar con el corrimiento electroforético (realizado bajo las condiciones que marca el fabricante) (Promega, 2018). Al terminar, los resultados fueron analizados con el software *GeneMapper® ID-X Software Version 1.5*.

Tabla 16. Parámetros utilizados para llevar a cabo la electroforesis capilar (3500 Genetic Analyser).

Parámetro	Ajustes para Applied Biosystems 3500
Tipo de aplicación	Fragmentos
Longitud del capilar	36 cm
Polímero	POP7
Temperatura de horno	60°C
Tiempo de inyección	3 s
Voltaje de inyección	1,2 kV
Tiempo del ciclo	1,500 s

5.5.7. INTERPRETACIÓN DE PERFILES STR

El software *GeneMapper*® ID-X Software Version 1.5., remueve la superposición espectral en los perfiles debido a que los espectros de los colorantes utilizados para etiquetar a los productos de PCR se traslapan y se obtienen como datos brutos picos compuestos por más de un colorante (Goodwin *et al.*, 2011; Ranjan *et al.*, 2020). *GeneMapper* permite calcular el tamaño de los picos generados por los fragmentos amplificados (Ranjan *et al.*, 2020). Para poder calcular el tamaño, altura y área de los picos, se añade a los productos de PCR un estándar de tamaño antes de someterlos a la electroforesis, este también se encuentra marcado por un colorante fluorescente, y el programa hace una comparación del estándar con los productos de PCR. La altura de los picos se mide en unidades de fluorescencia relativa (RFU), y la altura del pico dependerá de la cantidad de producto de PCR que se detecta (Thompson & Black, 2007).

El software también es capaz de asignar a cada alelo un número que representa la cantidad de repeticiones que contiene y esto lo logra a partir de un reactivo llamado escalera alélica, que contiene la mayoría de los alelos posibles que puede contener un *loci* (Ranjan *et al.*, 2020). Para que los alelos sean nombrados deben caer dentro de una ventana de 1pb, es decir en ± 0.5 pb del tamaño de la escalera, si los alelos difieren en más de este valor se clasifican como fuera de la escalera (OL) (Butler, 2010; Goodwin *et al.*, 2011). Los resultados obtenidos por *GeneMapper* son los electroferogramas de los

perfiles genéticos de las muestras y además un documento de texto (.txt) que contiene los alelos para cada marcador genético.

Los perfiles genéticos de STR deberían generar las mismas designaciones de alelos en cualquier laboratorio independientemente de las variaciones en la metodología del laboratorio en el que se realizó el análisis, además los *loci* marcadores incluidos en diferentes *kits* comerciales también deberán producir resultados idénticos (Goodwin *et al.*, 2011).

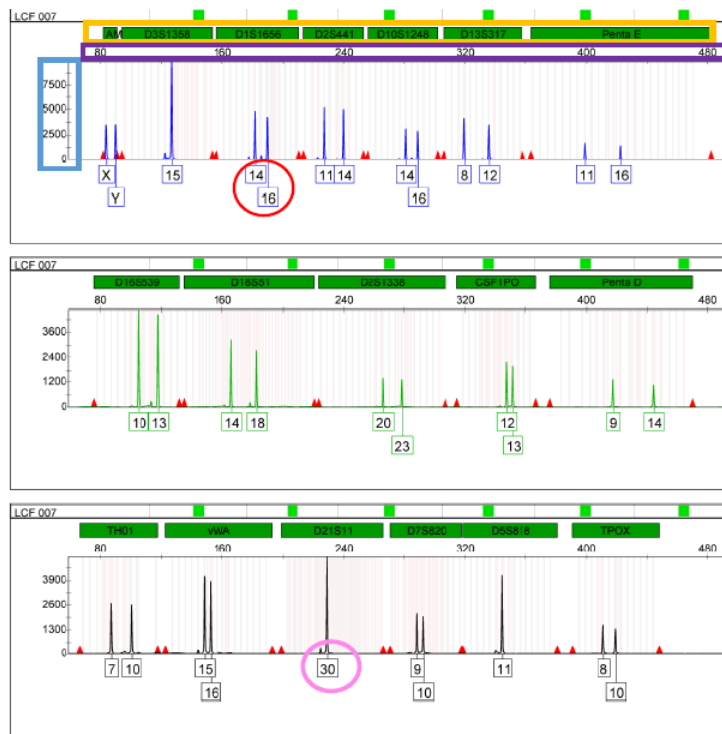


Figura 11. Ejemplo representativo de un electroferograma. El nombre de los marcadores STR están señalados en el recuadro amarillo, las unidades relativas de fluorescencia (RFUs) (Eje Y) se encuentran señaladas con azul, la escala de pares de bases (pb) (eje X) con la que se mide el tamaño de los fragmentos, se encuentra señalada con el recuadro morado y los alelos se encuentran señalados en rojo (heterocigoto) y rosa (homocigoto). Electroferograma obtenido en este proyecto.

5.5.8. ANÁLISIS ESTADÍSTICO

Se realizó una base de datos y se analizó la capacidad del *kit PowerPlex fusión 6C* como herramienta para la identificación humana, obteniendo las frecuencias alélicas y genotípicas. Para esto se ordenaron los datos obtenidos con el *software GeneMapper* en el programa Excel, creando una matriz general de todos los alelos obtenidos para cada *locus* de cada muestra (tabla 17 y tabla 18), esta matriz se analizó en el programa Access para realizar la cuenta de los

genotipos repetidos para cada locus y se obtuvieron matrices individuales para cada uno de los marcadores (tabla 19) a partir de las cuales se obtuvieron las frecuencias alélicas (en el programa Excel) para cada uno de los alelos de las muestras (tabla 20).

Tabla 17. Formato de Excel para su análisis en el programa Access.

Muestra	LocusX	LocusX.1	LocusY	LocusY.1	LocusZ	LocusZ.1
Nombre de la muestra	n° del alelo 1	n° del alelo 2	n° del alelo 1	n° del alelo 2	n° del alelo 1	n° del alelo 2

Tabla 18. Ejemplo representativo del formato para su análisis en el programa Access.

Muestra	TPOX	TPOX	vWA	vWA	SE33	SE33
LCF001	8	8	16	16	18	28.2
LCF002	8	11	17	17	28.2	31.2

Tabla 19. Matriz para cada locus generada por el programa Access. Datos obtenidos en este proyecto.

vWA	vWA.1	Cuenta De Muestra
14	15	1
14	16	1
14	17	1
15	16	4
15	17	1
15	18	1
16	16	7
16	17	9
16	18	10
16	19	1
17	17	7
17	18	4
17	19	1
18	18	2
Total		50

Tabla 20. Matriz generada en el programa Excel para el cálculo de las frecuencias alélicas. Matriz para el marcador vWA generada en el presente estudio.

vWA	Allele	14	15	16	17	18	19	Cuenta	Alelos totales	Frecuencia
	14			1	1	1			3	3
15				4	1	1		6	7	0.07
16				7	9	10	1	27	39	0.39
17					7	4	1	12	30	0.3
18						2		2	19	0.19
19								0	2	0.02
Total		0	1	12	18	17	2	50	100	1

Después a través del software STRAF 2.0.8: *STR Analysis for Forensics* a partir de la matriz generada por el programa Access para todos los marcadores en formato de texto separado por tabuladores, se obtuvieron los siguientes parámetros forenses para cada uno de los STR autosómicos analizados: Probabilidad de coincidencia al azar (PM), Poder de discriminación (PD), Heterocigosidad esperada (GD o H), Contenido de información del polimorfismo (PIC), Poder de exclusión (PE) y el índice típico de paternidad (TPI) y se determinó si las muestras se encontraban dentro del equilibrio de Hardy-Weinberg. Por último, se realizó una comparación cualitativa con los datos obtenidos en las mismas muestras a partir de secuenciación de última generación que se llevó a cabo en el laboratorio de genética de la Licenciatura de Ciencia Forense de la UNAM y un el *kit* comercial *Investigator 24 Plex* de Qiagen (Ciudad de México) para determinar si hay diferencias entre estos sistemas.

6. RESULTADOS Y DISCUSIÓN

La figura 12 presenta uno de los 50 electroferogramas obtenidos donde se observan los 23 marcadores autosómicos analizados en este estudio, además cuatro marcadores para determinar el sexo (tres STRY y la amelogenina). En cada locus se detectan picos que corresponden a los dos alelos, uno proveniente de la madre y otro del padre. Si se obtuvo como resultado un solo pico, nos indica que el individuo es homocigoto en ese locus y si se obtuvieron dos picos, es heterocigoto (Butler, 2015).

En los electroferogramas para cada una de las muestras, se observa que en la muestra LCF 002 hay un alelo “fuera de la escalera alélica” en el marcador D12S391 (Anexo 1), esto se debe a que es un alelo que no cae en la escala específica del *locus* correspondiente a la escalera alélica determinada por el *kit PowerPlex Fusion 6C* y para verificarlo se pueden analizar nuevamente con un *kit* diferente al utilizado. En este estudio se realizó la comparación con el *kit Investigator 24 Plex de Qiagen* donde también se obtuvo este alelo “fuera de la escalera alélica”, pero al realizar la secuenciación de última generación se logró obtener el alelo correspondiente (Barbosa *et al.*, 2010; Jamieson & Bader, 2016). También en el electroferograma de la muestra LCF 050 (anexo 1) se obtuvo un alelo nulo en el marcador Penta E, esto se debe a que son alelos presentes en el DNA que no fueron amplificados debido a que probablemente existan mutaciones en las regiones que flanquean a los STR y es dónde se unen los cebadores de los marcadores genéticos y esto provoca que no se lleve a cabo la amplificación de uno o de ambos alelos presentes en ese *locus*. Los alelos nulos pueden comprobarse con una amplificación con un nuevo juego de cebadores de otro *kit* comercial (Butler, 2006). En el caso de este estudio, no fue posible comprobar con otro juego de cebadores, ya que el *kit* con el que se compararon los datos no contiene el marcador Penta E.

Se obtuvieron las frecuencias alélicas a partir de las matrices generadas por el programa Access, estas se presentan en la tabla 21, donde se observa que el marcador con la más alta variedad de alelos fue el SE33, seguido de penta E.

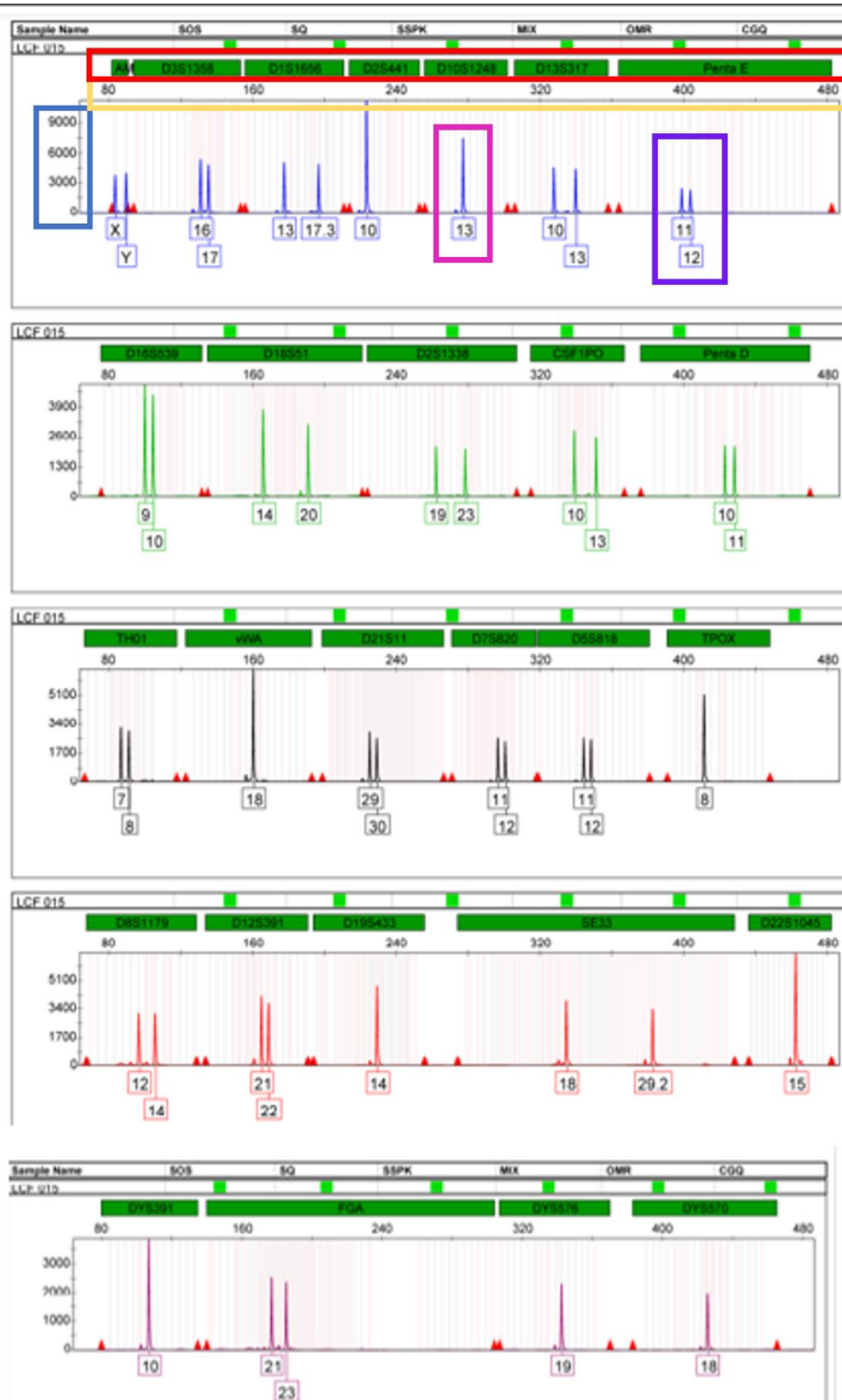


Figura 12. Electroferograma de la muestra LCF015. Ejemplo representativo para una de las 50 muestras analizadas. El nombre de los marcadores STR están señalados en el recuadro rojo, las unidades relativas de fluorescencia (RFUs) (Eje Y) se encuentran señaladas con azul, la escala de pares de bases (pb) (eje X) con la que se mide el tamaño de los fragmentos, se encuentra señalada con el recuadro amarillo y los alelos se encuentran señalados en morado (heterocigoto) y rosa (homocigoto). Electroferograma obtenido en este proyecto.

Tabla 21. Frecuencias alélicas obtenidas de los 23 *loci* autosómicos analizados por el *kit PowerPlex Fusion 6C* para 50 individuos mestizos mexicanos de la CDMX.

Alelo	D3S1358	D1S1656	D2S441	D10S1244	D13S317	Penta E	D16S539	D18S51	D2S1338	CSF1PO	Penta D	TH01	vWA	D21S11	D7S820	D5S818	TPOX	D8S1179	D12S391	D19S433	SE33	D22S1044	FGA
2.2											0.01												
6													0.3					0.02					
7						0.06					0.01		0.26			0.01	0.03	0.01					
8					0.06	0.01				0.01			0.09			0.07	0.01	0.48	0.01				
9						0.22		0.1	0.01		0.02	0.26	0.09			0.11	0.07	0.05					
9.3													0.25										
10			0.41	0.01	0.11					0.3	0.23				0.26	0.1	0.06	0.08					
11		0.01	0.34	0.01	0.21	0.07	0.21	0.01		0.26	0.16				0.29	0.36	0.28	0.06					0.08
11.3			0.03																				
12		0.1	0.04	0.03	0.23	0.2	0.31	0.07		0.3	0.12				0.21	0.31	0.1	0.1		0.05			0.01
13	0.01	0.16	0.03	0.2	0.08	0.11	0.11	0.15		0.09	0.17				0.05	0.11		0.28		0.17	0.01		0.01
13.2																					0.13		
14	0.04	0.1	0.14	0.39	0.08	0.14	0.04	0.16	0.01	0.01	0.03		0.03			0.01		0.34		0.23	0.03		0.02
14.2																					0.06		
15	0.48	0.19	0.01	0.3	0.01	0.11		0.12			0.01		0.07					0.1	0.03	0.22	0.03		0.48
15.2																					0.08		
15.3		0.01																					
16	0.23	0.16		0.05		0.09		0.17	0.02				0.39					0.02		0.03	0.04		0.35
16.2																					0.02		
16.3		0.02																					
17	0.17	0.03		0.01		0.02		0.17	0.21	0.01			0.3					0.01	0.08	0.01	0.06		0.04
17.3		0.17																		0.02	0.01		
18	0.06	0.01				0.05		0.05	0.04				0.19							0.15	0.13	0.01	0.01
18.2																							0.01
18.3		0.03																		0.01			
19	0.01					0.05		0.06	0.17				0.02							0.34	0.12		0.07
19.3		0.01																					
20						0.04		0.02	0.13											0.16	0.07		0.09
21						0.01		0.01	0.03											0.11	0.01		0.07
21.2																					0.02		
22						0.01			0.09											0.05	0.02		0.12
23						0.01			0.18											0.04			0.18
24									0.06														0.15
24.2																					0.04		
25									0.05											0.01			0.14
25.2																					0.05		
26														0.01									0.12
26.2																					0.05		
27									0.01					0.03									0.04
27.2																							
28																						0.1	
28.2																					0.05		
29														0.16									
29.2																					0.12		
30														0.31									
30.2														0.02							0.01		
31														0.06									
31.2														0.14							0.02		
32.2														0.19							0.01		
33.2														0.05									

Para que las frecuencias alélicas obtenidas permitan desarrollar una base de datos deben estar en equilibrio de Hardy-Weinberg ya que este modelo propone el comportamiento “ideal” de una población y permite la relación entre las frecuencias alélicas con las genotípicas, generalmente se utilizan las frecuencias alélicas en vez de las genotípicas ya que, es complicado hacer la estimación de todos los genotipos de una población. En cuanto a las frecuencias alélicas si se encuentran en HWE entonces serán constantes y podrán ser utilizadas en las generaciones futuras (Thompson & Black, 2007).

De los 23 *loci* autosómicos para los que se calculó el equilibrio de HWE como se muestra en la Tabla 22, dos no se encontraron en equilibrio: D2S1338 y D10S1248, pero al aplicar la corrección de Bonferroni ($P > 0.0021$) se perdieron las desviaciones del HWE. Se aplica dicha corrección cuando se realizan simultáneamente varias pruebas estadísticas independientes ya que, un valor alfa determinado (en este caso $\alpha = 0.05$) puede ser adecuado para comparaciones individuales, pero no para todo un conjunto, por lo que el valor alfa debe disminuirse para tener en cuenta el número de comparaciones que se realizan. La aplicación de esta corrección elimina el estigma de que un *locus* que esté por debajo del umbral del 5% (Butler, 2005).

Tabla 22. Resultados del equilibrio de Hardy-Weinberg obtenidos con el software STRAF 2.0.8: STR Analysis for Forensics para los 23 *loci* STR del kit PPF6C. *Se aplicó la corrección de Bonferroni ($p > 0.0021$)

Locus	HWE ($\alpha = 0.05$)	Resultado
CSF1PO	0.479	Equilibrio
D10S1248	0.050	Desequilibrio*
D12S391	0.421	Equilibrio
D13S317	0.789	Equilibrio
D16S539	0.917	Equilibrio
D18S51	0.596	Equilibrio
D19S433	0.943	Equilibrio
D1S1656	0.179	Equilibrio
D21S11	0.205	Equilibrio
D22S1045	0.559	Equilibrio
D2S1338	0.013	Desequilibrio*
D2S441	0.779	Equilibrio
D3S1358	0.864	Equilibrio
D5S818	0.926	Equilibrio
D7S820	0.104	Equilibrio
D8S1179	0.693	Equilibrio
FGA	0.890	Equilibrio
Penta D	0.131	Equilibrio
Penta E	0.692	Equilibrio
SE33	0.053	Equilibrio
TH01	0.339	Equilibrio
TPOX	0.978	Equilibrio
vWA	0.699	Equilibrio

Dado que, en este estudio los 23 *loci* autosómicos se encontraron en equilibrio HW (Tabla 22) se pueden utilizar sus frecuencias alélicas para predecir las frecuencias genotípicas para las futuras generaciones y con esto calcular los parámetros forenses para comparar dos perfiles de DNA (Thompson & Black, 2007).

Las frecuencias alélicas también fueron utilizadas para calcular los parámetros forenses que se muestran en la Tabla 23, que son: Probabilidad de coincidencia al azar (PM), Poder de discriminación (PD), Heterocigosidad observada (H_o), Homocigosidad observada (H), Heterocigosidad esperada (GD), Contenido de

información del polimorfismo (PIC), Poder de exclusión (PE) y el índice típico de paternidad (TPI).

En la Tabla 23 se muestra que, el poder de discriminación más alto lo obtuvo el marcador SE33 lo que indica que, este marcador brinda una alta probabilidad de discriminación entre individuos empleando el *kit PowerPlex Fusion 6C (PPF6C)*. También podemos observar que este mismo marcador obtuvo el valor más alto de poder de exclusión, lo que le confiere al *kit PPF6C* una disminución de la probabilidad de que existan asociaciones con un perfil al azar dentro de la población. Para la probabilidad de coincidencia al azar, los valores más bajos son para el marcador SE33, seguido del Penta E, apoyando también a que estos disminuyan la asociación con algún perfil al azar. De acuerdo con los datos para el contenido de información del polimorfismo (PIC) el valor más alto lo obtuvo el marcador SE33 (Gouy & Zieger, 2021). En esta tabla también se pueden observar los valores para el índice típico de paternidad (ITP) que mide cuántas veces es más probable que un individuo sea el padre real que un hombre seleccionado al azar en la población (Gouy & Zieger, 2021). Para este parámetro el marcador SE33 obtuvo el valor más alto.

El locus SE33 obtuvo los valores más representativos para todos los parámetros de interés forense, lo que indica que es el marcador más informativo para el sistema, todo esto puede deberse a que es el marcador con la mayor cantidad de alelos. Por el contrario, el marcador con valores menos representativos para la mayoría de los parámetros estudiados fue el TPOX, esto también puede deberse a la poca cantidad de alelos que tenía el locus.

Tabla 23. Parámetros estadísticos calculados para los 23 loci autosómicos estudiados con el kit PPF6C en población mestiza-mexicana de la CDMX.

Locus	Probabilidad de coincidencia al azar (PM)	Poder de discriminación (PD)	heterocigosidad observada (Ho)	Homocigotos observados (h)	poder de exclusión (PE)	índice típico de paternidad (ITP)	Heterocigosidad esperada (He)	Contenido de información polimórfica (PIC)
CSF1PO	0.1208	0.8792	0.76	0.24	0.527030	2.083	0.7511	0.6986946
D10S1248	0.1424	0.8576	0.66	0.34	0.369131	1.471	0.7214	0.66535986
D12S391	0.0688	0.9312	0.78	0.22	0.562463	2.273	0.8204	0.7916533
D13S317	0.0592	0.9408	0.78	0.22	0.562463	2.273	0.8343	0.80305108
D16S539	0.0864	0.9136	0.78	0.22	0.562463	2.273	0.7911	0.75042516
D18S51	0.0464	0.9536	0.88	0.12	0.754774	4.167	0.8768	0.8536588
D19S433	0.0544	0.9456	0.88	0.12	0.754774	4.167	0.8475	0.81940194
D1S1656	0.0536	0.9464	0.88	0.12	0.754774	4.167	0.8699	0.84558552
D21S11	0.0712	0.9288	0.82	0.18	0.636671	2.778	0.8224	0.8014485
D22S1045	0.1976	0.8024	0.64	0.36	0.341652	1.389	0.6448	0.57577956
D2S1338	0.0624	0.9376	0.84	0.16	0.675254	3.125	0.8691	0.8451156
D2S441	0.1424	0.8576	0.68	0.32	0.398004	1.563	0.7002	0.64108308
D3S1358	0.1512	0.8490	0.78	0.22	0.562463	2.273	0.6893	0.63826356
D5S818	0.1064	0.8936	0.78	0.22	0.562463	2.273	0.7537	0.70808818
D7S820	0.1024	0.8976	0.80	0.20	0.599040	2.500	0.7925	0.7519669
D8S1179	0.0872	0.9128	0.76	0.24	0.527030	2.083	0.7832	0.74471886
FGA	0.0384	0.9616	0.86	0.14	0.714667	3.571	0.8859	0.8643421
PentaD	0.0824	0.9176	0.80	0.20	0.599040	2.500	0.8176	0.78213858
PentaE	0.0371	0.9629	0.88	0.12	0.749830	4.083	0.8969	0.87780000
SE33	0.0320	0.9680	0.90	0.10	0.795420	5.000	0.9309	0.91631676
TH01	0.1168	0.8832	0.82	0.18	0.636671	2.778	0.7713	0.72442228
TPOX	0.1640	0.8360	0.78	0.22	0.562463	2.273	0.6814	0.62806494
vWA	0.1288	0.8712	0.68	0.32	0.398004	1.563	0.7228	0.66727924

Además, en la Tabla 24 se muestran los valores combinados de los parámetros forenses para el *kit*. En cuanto al índice típico de paternidad combinado, este fue de 2, 370,927,252, de acuerdo con lo establecido para los laboratorios en México, la paternidad biológica se considera acreditada cuando la paternidad es mayor a 99.9% (Probabilidad de paternidad= $(ITP/ITP+1) \times 100$), el resultado obtenido para el kit PPF6C fue de 99.9999% lo que indica que la base de datos generada brinda una alta probabilidad de que un posible padre sea el padre real antes que un hombre seleccionado al azar en la población estudiada (Easy DNA México, 2020; Gouy & Zieger, 2021).

La probabilidad de coincidencia al azar (PM) es la probabilidad de que dos muestras elegidas al azar de la población tengan el mismo perfil genético (Gouy & Zieger, 2021; Hernández & Trejo, 2014; Michaelis *et al.*, 2008), el valor combinado para el kit fue de 1.36195×10^{-25} lo que indica entonces que la probabilidad de que existan asociaciones con un perfil al azar de la población estudiada es de 1.36195×10^{-25} por lo tanto la probabilidad disminuye, al utilizar la base de datos generada en este estudio además, cuantos más *loci* de utilizan para la base de datos, más pequeño se vuelve el valor del PM (Craig, 2010), por lo que se puede decir que el número de marcadores utilizados por el kit PPF6C le confiere a la base de datos una alta probabilidad de discriminación (Butler, 2010).

El poder de discriminación es la probabilidad de que dos muestras elegidas al azar no tengan el mismo perfil genético (Craig, 2010), ya que el poder de discriminación combinado obtenido fue de 89.85% indica que la base de datos generada brinda una alta probabilidad de discriminación entre individuos. Por otro lado, el Poder de exclusión es la probabilidad de excluir un perfil seleccionado al azar dentro de la población estudiada de un perfil genético de estudio (Jamieson & Bader, 2016), mayores probabilidades de exclusión reflejan que las coincidencias son menos probables por lo que el resultado no será por mera casualidad (Jamieson & Bader, 2016). Ya que el valor obtenido para el poder de exclusión combinado fue de 99.99% entonces se puede decir que al utilizar la base de datos generada en el estudio se disminuye la probabilidad de que existan asociaciones con perfiles al

azar de la población de estudio y demuestra la capacidad del *kit PowerPlex Fusion 6C* como herramienta para la identificación humana.

Tabla 24. Resultado de los índices combinados del *kit PowerPlex Fusion 6C* para 23 loci STR estudiados en una población mestiza-mexicana de la CDMX. Parámetros estadísticos de interés forense (PEIF), Ho= heterocigotos observados, He= Heterocigotos esperados, PD= poder de discriminación, PE= poder de exclusión, PM= probabilidad de coincidencia al azar, ITP= índice típico de paternidad.

Kit	PPF6C
PEIF	Valor combinado
Ho	0.004250939
He	0.004534955
PD	0.101494889
PE	3.1630E-06
PM	1.36195E-25
ITP	2370927252

Se compararon los 50 perfiles obtenidos de manera cualitativa con el *kit Investigator 24 Plex* (ya validado en el laboratorio y utilizado satisfactoriamente para desarrollar una base de datos) y con los resultados de la secuenciación de última generación para las primeras 32 muestras, obteniendo los resultados que se presentan en la Tabla 25 donde, no hubo diferencias entre los dos análisis con los kit comerciales y la secuenciación de última generación, lo que demuestra la capacidad del *kit PPF6C* como una herramienta para la identificación humana en un laboratorio forense de manera preliminar, mientras logra acoplarse al número de muestras requeridas por la mayoría de los datos poblacionales publicados anteriormente (González *et al.*, 2019; Hernández & Trejo, 2014; Martínez *et al.*, 2019; Melo *et al.*, 2022; Ramírez, 2014). Además, el objetivo de estas comparaciones también es solucionar problemas como los presentados en el caso de la muestra LCF002 (alelo fuera de la escalera alélica) y LCF050 (alelo nulo en el Locus Penta E) (Anexo 1), que como ya se mencionó anteriormente pueden resolverse utilizando un *kit* con una escalera alélica diferente o cebadores diferentes, en este caso la secuenciación de última generación es muy útil ya que permite la lectura de millones de secuencias de DNA en forma masiva y paralela

(Rubio *et al.*, 2020). Como se puede observar en la tabla 25, se logró obtener el alelo correspondiente a la muestra LCF002 del locus D12S391 (23, 28), esto no se pudo realizar para la muestra LCF050 ya que solo se realizó la secuenciación de última generación hasta la muestra LCF032.

Tabla 25. Ejemplo representativo de la comparación entre el kit comercial Investigator 24plex, secuenciación de última generación y el kit PowerPlex Fusion 6C en muestras de 5 individuos.

Locus	Muestra 001			Muestra 002			Muestra 003			Muestra 004			Muestra 005			
Kit	SECM	PPF6C	24Plex	SECM	PPF6C	24Plex	SECM	PPF6C	24Plex	SECM	PPF6C	24Plex	SECM	PPF6C	24Plex	Sin marcador
D1S1656	11,13	11,13	11,13	13,18.3	13,18.3	13,18.3	12,15	12,15	12,15	12,14	12,14	12,14	14,17.3	14,17.3	14,17.3	Coinciden
TPOX	8,8	8,8	8,8	8,11	8,11	8,11	8,11	8,11	8,11	11,12	11,12	11,12	11,11	11,11	11,11	No coinciden
D2S441	10,15	10,15	10,15	11,12	11,12	11,12	11,14	11,14	11,14	10,11	10,11	10,11	10,11.3	10,11.3	10,11.3	alelo Off ladder
D2S1338	24,24	24,24	24,24	19,24	19,24	19,14	17,25	17,25	17,25	18,23	18,23	18,23	17,23	17,23	17,23	
D3S1358	14,15	14,15	14,15	17,17	17,17	17,17	16,18	16,18	16,18	14,15	14,15	14,15	16,17	16,17	16,17	
D4S2408	9,9			11,12			10,11			9,11			9,9			
FGA	21,24	21,24	21,24	23,24	23,24	23,24	24,26	24,26	24,26	23,25	23,25	23,25	22,25	22,25	22,25	
D5S818	11,13	11,13	11,13	12,12	12,12	12,12	9,11	9,11	9,11	10,12	10,12	10,12	11,13	11,13	11,13	
CSF1PO	11,12	11,12	11,12	12,12	12,12	12,12	10,11	10,11	10,11	10,11	10,11	10,11	11,12	11,12	11,12	
D6S1043	11,14			12,14			11,13			11,12			12,18			
D7S820	10,11	10,11	10,11	10,13	10,13	10,13	8,11	8,11	8,11	12,13	12,13	12,13	10,10	10,10	10,10	
D8S1179	10,14	10,14	10,14	13,15	13,15	13,15	12,14	12,14	12,14	13,13	13,13	13,13	13,14	13,14	13,14	
D9S1122	13,13			13,13			13,13			12,13			12,13			
D10S1248	13,15	13,15	13,15	13,14	13,14	13,14	13,14	13,14	13,14	14,15	14,15	14,15	13,16	13,16	13,16	
TH01	6,9.3	6,9.3	6,9.3	7,9	7,9	7,9	7,9.3	7,9.3	7,9.3	8,9	8,9	8,9	7,9	7,9	7,9	
vWA	16,16	16,16	16,16	17,17	17,17	17,17	14,17	14,17	14,17	17,18	17,18	17,18	17,18	17,18	17,18	
D12S391	18.3,20	18.3,20	18.3,20	23,28	23,OL	23,OL	19,21	19,21	19,21	15,18	15,18	15,18	20,20	20,20	20,20	
D13S317	9,9	9,9	9,9	9,13	9,13	9,13	10,11	10,11	10,11	8,8	8,8	8,8	9,9	9,9	9,9	
PentaE	11,14	11,14		14,16	14,16		12,15	12,15		7,11	7,11		13,19	13,19		
D16S539	11,12	11,12	11,12	12,12	12,12	12,12	11,12	11,12	11,12	9,11	9,11	9,11	13,14	13,14	13,14	
D17S1301	11,13			13,13			12,12			11,11			11,12			
D18S51	13,14	13,14	13,14	16,17	16,17	16,17	12,16	12,16	12,16	17,19	17,19	17,19	16,20	16,20	16,20	
D19S433	14.2,15	14.2,15	14.2,15	13,13.2	13,13.2	13,13.2	13,13	13,13	13,13	13,15	13,15	13,15	13,15	13,15	13,15	
D20S482	14,15			13,13			13,14			14,14			13,14			
D21S11	31.2,33.2	31.2,33.2	31.2,33.2	30,31.2	30,31.2	30,31.2	29,33.2	29,33.2	29,33.2	28,30	28,30	28,30	26,31.2	26,31.2	26,31.2	
PentaD	10,13	10,13		9,15	9,15		9,10	9,10		12,13	12,13		10,10	10,10		
D22S1045	15,16	15,16	15,16	16,17	16,17	16,17	15,16	15,16	15,16	15,16	15,16	15,16	16,16	16,16	16,16	
SE33		18,28.2	18,28.2		28.2,31.2	28.2,31.2		18,22	18,22		20,26.2	20,26.2		18,29.2	18,29.2	

En la Tabla 26, se compararon los parámetros del *kit PowerPlex Fusion 6C* con los valores del kit Investigator 24 Plex y como se puede observar los índices que obtiene el *kit PowerPlex Fusion 6C* tienen el mismo orden de magnitud, incluso en el parámetro PM en el sistema PPF6C fue de una magnitud más, esto puede deberse a que el kit PPF6C tiene un mayor número de marcadores (23 marcadores autosómicos) que el kit Investigator 24 Plex (21 marcadores autosómicos), lo que permite que el sistema PPF6C tenga valores más representativos para los parámetros de interés forense (Butler, 2007; Vázquez, 2022), ya que los índices combinados del kit PPF6C son muy parecidos a los del kit 24 Plex, se puede decir que constituye una buena herramienta para la identificación humana y que se puede utilizar satisfactoriamente para predecir de forma fiable las frecuencias alélicas y genotípicas de la población de manera provisional mientras se aumenta su número de muestras. Además, algunos estudios mencionan que las frecuencias alélicas entre las bases de datos poblacionales de tamaño pequeño y grande (para la misma población) rara vez difieren significativamente para los alelos comunes (Butler, 2006).

Tabla 26. Comparación de los parámetros estadísticos forenses entre el *kit PowerPlex Fusion 6C* y el kit Investigator 24Plex. Parámetros estadísticos de interés forense (PEIF), Ho= heterocigotos observados, He= Heterocigotos esperados, PD= poder de discriminación, PE= poder de exclusión, PM= probabilidad de coincidencia al azar, ITP= índice típico de paternidad.

Kit	PPF6C	Investigator 24 Plex
PEIF	Valor combinado	Valor combinado
Ho	0.004250939	0.004568885
He	0.004534955	0.005138751
PD	0.101494889	0.130394753
PE	3.1630E-06	3.43513E-06
PM	1.36195E-25	2.29375E-24
ITP	2370927252	127030890.8

Se realizó una comparación entre el sistema utilizado en el presente estudio: PowerPlex Fusion 6C, con cinco reportados en la literatura utilizados en población mexicana. En la Tabla 27 se muestra la comparación de los marcadores utilizados en cada sistema en la cual se observa en primera instancia que, en el estudio de Melo *et al.*, 2022 (Investigator 24 Plex GO!) se analizó un marcador que no tiene ningún

otro sistema (D8S133879), además contiene dos marcadores autosómicos menos que el kit PowerPlex Fusión 6C y no analiza los marcadores: Penta E, Penta D y D8S1179.

En el estudio de Hernández y Trejo, 2014, en el que utilizaron el kit AmpFISTR Identifiler, analizan solo los 15 marcadores, que tiene en común con el kit PPF6C, los *loci* que no analizaron son: D3S1358, D1S1656, D2S441, D10S1248, Penta E, Penta D, D12S391, SE33 y D22S1045. En cuanto a los estudios de González *et al.*, 2019 y Ramírez, 2014, utilizan el mismo kit: PowerPlex Fusion, que estudia un marcador menos que el kit PPF6C, este marcador es el: SE33. Finalmente, en el estudio de Martínez *et al.*, 2019, se utilizó el kit GlobalFiler y se analizaron dos *loci* STR menos que los del kit PPF6C, los cuales fueron: Penta E y Penta D.

Tabla 27. Comparación de los marcadores incluidos en cada uno de los sistemas con los del presente estudio.

Locus	PowerPlex Fusion 6C	Investigator 24 Plex GO!	AmpFISTR Identifiler	PowerPlex Fusion (Yucatán)	PowerPlex Fusion (CDMX)	GlobalFiler
D3S1358	*	*	*	*	*	*
D1S1656	*	*		*	*	*
D2S441	*	*		*	*	*
D10S1248	*	*		*	*	*
D13S317	*	*	*	*	*	*
Penta E	*			*	*	
D16S539	*	*	*	*	*	*
D18S51	*	*	*	*	*	*
D2S1338	*	*	*	*	*	*
CSF1PO	*	*	*	*	*	*
Penta D	*			*	*	
TH01	*	*	*	*	*	*
vWA	*	*	*	*	*	*
D21S11	*	*	*	*	*	*
D7S820	*	*	*	*	*	*
D5S818	*	*	*	*	*	*
TPOX	*	*	*	*	*	*
D8S1179	*		*	*	*	*
D12S391	*	*		*	*	*
D19S433	*	*	*	*	*	*
SE33	*	*				*
D22S1045	*	*		*	*	*
FGA	*	*	*	*	*	*
D8S133879		*				

Comparación del estudio con el kit PPF6C (23 STR autosómicos) en población mestiza-mexicana con los sistemas: Melo et al. 2022 (Investigator 24 Plex GO!) (21 STR autosómicos) en la población del estado de Veracruz, González et al. 2019 (PowerPlex® Fusion) (22 STR autosómicos) en una población mestiza de la península de Yucatán, Hernández y Trejo, 2014 (kit AmpFISTR Identifiler) (15 STR autosómicos) en una población del estado de Zacatecas, Ramírez, 2014 (PowerPlex Fusion) (22 STR autosómicos) en población de la Ciudad de México y Martínez et al. 2019 (21 *loci* STR autosómicos) (GlobalFiler) en dos poblaciones mexicano-mestizas del noroeste de México. * El kit contiene el locus. Los recuadros seleccionados en gris indican que el locus no está presente en el kit comercial.

En la tabla 28, se realizó la comparación de los parámetros estadísticos de interés forense obtenidos en este estudio para el kit PowerPlex Fusión 6C con los reportados en otros estudios en mexicanos en la literatura, en donde se detecta que, aunque en este estudio (kit PPF6C) se utilizó un menor número de muestras se encuentra en el mismo orden de magnitud que los reportados en la literatura. En

cuanto al estudio con el kit 24Plex GO!, el kit PPF6C obtuvo un orden de magnitud más en el PM y dos órdenes de magnitud para el ITP, esto puede deberse a que el sistema de 24 Plex GO! no analiza los marcadores Penta E y Penta D que son *loci* con alta variabilidad (Butler, 2006). En los estudios realizados con el kit PowerPlex Fusion en CDMX y en Yucatán se obtuvieron parámetros forenses en el mismo orden de magnitud a pesar de que el estudio realizado en Yucatán se analizaron mayor número de muestras (733 individuos). El estudio de GlobalFiler se encuentra en el mismo orden de magnitud que el del sistema PPF6C a pesar de que no analiza los marcadores Penta E y Penta D. Por último, el kit AmpFISTR obtuvo siete órdenes de magnitud menos en el parámetro PM y dos menos en el parámetro ITP, esto puede deberse a que el kit AmpFISTR analiza ocho marcadores menos que el kit PPF6C incluyendo los marcadores Penta E, Penta D y SE33, que son marcadores con variabilidad alta (Butler, 2006).

A pesar de que el estudio con el *kit PowerPlex Fusion 6C* tiene una muestra poblacional reducida ($n= 50$), obtuvo valores de PD= 89.85051115% y de PE= 90.9996837 que se encuentran en el mismo orden de magnitud que los de todos los sistemas reportados en la literatura con los que se compararon (González *et al.*, 2019; Hernández & Trejo, 2014; Martínez *et al.*, 2019; Melo *et al.*, 2022; Ramírez, 2014).

Tabla 28. Comparación de los valores de los parámetros estadísticos de interés forense obtenidos en este estudio (kit PPF6C) con los reportados en la literatura. Parámetros estadísticos de interés forense (PEIF), N°=número, Ho= heterocigotos observados, He= Heterocigotos esperados, PD= poder de discriminación, PE= poder de exclusión, PM= probabilidad de coincidencia al azar, ITP= índice típico de paternidad.

Kit	PPF6C	24PLEX GO!	AmpFISTR	PowerPlex Fusion (Yucatán)	PowerPlex Fusion (CDMX)	GlobalFiler
PEIF	Valor combinado	Valor combinado	Valor combinado	Valor combinado	Valor combinado	Valor combinado
N° muestras	50	234	271	733	200	381
Ho	0.00425094	0.00385079	0.018338034	0.00336705	0.003206581	0.00441602
He	0.00453496	0.00437562	0.028627736	No reportado	0.003251877	0.0053595
PD	0.10149489	0.12548056	0.27603488	0.14220866	0.10858606	0.15513883
PE	3.1630×10^{-6}	2.2573×10^{-6}	No reportado	1.67422×10^{-6}	1.46257×10^{-6}	5.9208×10^{-6}
PM	1.3620×10^{-25}	1.5337×10^{-24}	6.7965×10^{-18}	No reportado	2.5740×10^{-25}	1.2364×10^{-25}
ITP	2370927252	60354023.9	453618.6006	297881540.5	228483328.8	295689749

7. CONCLUSIONES

Las frecuencias alélicas de los 23 marcadores STR autosómicos estudiados con el kit comercial PowerPlex Fusion 6C en 50 individuos de la Ciudad de México, están en equilibrio de Hardy-Weinberg.

Los parámetros estadísticos forenses para cada locus del kit PPF6C indican una heterocigosidad observada del 99.57%, una probabilidad de coincidencia al azar de 1.3620×10^{-25} , un poder de discriminación del 89.85051115%, una heterocigosidad esperada del 99.54650445%, un poder de exclusión de 99.9996837% y un índice típico de paternidad de 2370927252, lo que hacen al kit y a la base de datos, una herramienta útil para un laboratorio forense.

De acuerdo con los parámetros forenses obtenidos y debido a que las frecuencias alélicas de los 23 *loci* autosómicos se encuentran en equilibrio de Hardy-Weinberg, se llegó a la conclusión de que el *kit PowerPlex Fusion 6C* es una herramienta útil de manera preliminar para la identificación humana en un laboratorio forense, con una muestra poblacional de 50 individuos.

Los resultados con el *kit PowerPlex Fusion 6C* con 50 individuos son comparables con aquellos obtenidos para el kit Investigator 24 Plex y la secuenciación de última generación hasta la muestra LCF032, demostrando que los resultados son similares.

Los parámetros forenses obtenidos con el kit PPF6C son comparables con el kit Investigator 24 Plex que en ambos casos obtienen índices forenses altos, lo que hace al kit PPF6C factible para la identificación humana en un laboratorio forense.

8. PERSPECTIVAS

Este trabajo abre las puertas a otros proyectos, ya que podría ser útil para los laboratorios forenses en la identificación humana, ya que permitiría comenzar con una base de datos pequeña, mientras logran aumentar el tamaño de muestra recomendado en los estudios poblacionales (150 a 200 individuos), para tener un soporte estadístico más robusto, con mayor poder de discriminación y evitando asociaciones entre perfiles al azar y recabar más información para las bases de datos en México.

REFERENCIAS

- Abramovs, N., Brass, A., & Tassabehji, M. (2020). Hardy-Weinberg Equilibrium in the Large Scale Genomic Sequencing Era. *Frontiers in Genetics*, 1-11.
- Aguilar, J., & Rangel, H. (2020). Distribution of paternal lineages in Mestizo populations throughout Mexico: an in silico study based on Y-STR haplotypes. *Journal of Human Genetics*, 327-332.
- Alberts, B., Hopkin, K., Johnson, A., Morgan, D., Raff, M., Roberts, K., & Walter, P. (2021). *Introducción a la biología celular*. México: Panamericana.
- Armstrong, R. (2014). When to use the Bonferroni correction. *Ophthalmic and Physiological Optics*, 502-508.
- Baggett, D., Folliard, J., & Kluza, J. (2005). DNA Fingerprinting. (*Tesina de licenciatura*). Worcester Polytechnic Institute, Massachusetts.
- Barbosa, M., Salas, O., Posada, Y., Rodríguez, A., & Álvarez, T. (2010). Reporte de casos: alelos nulos y variantes Alélicas detectados en el laboratorio de identificación genética identigen de la universidad de Antioquia entre los años 2008 y 2010. *Iatreia*, S-123.
- Bhagavan, N., & Ha, C. (2022). *Essentials of medical biochemistry : with clinical cases*. San Diego: ELSEVIER.
- Bolivar, A., Rojas, A., & Garcia, P. (2014). PCR y PCR-Múltiple: parámetros críticos y protocolo de estandarización. *Avances en biomedicina*, 25-33.
- Butler, J. (2005). *Forensic DNA Typing: Biology, Technology, and Genetics of STR Markers*. Londres: Elsevier Academic Press.
- Butler, J. (2006). Genetics and Genomics of Core Short Tandem Repeat Loci Used in Human Identity Testing. *J Forensic Sci*, 253-265.
- Butler, J. (2006). Genetics and Genomics of Core Short Tandem Repeat Loci Used in Human Identity Testing. *J Forensic Sci*, 253-265.

- Butler, J. (2007). Short tandem repeat typing technologies used in human identity testing. *BioTechniques*, 2-5.
- Butler, J. (2010). *Fundamentals of forensic DNA typing*. California: Elsevier.
- Butler, J. (2012). *Advanced Topics in Forensic DNA Typing: Methodology*. California: Elsevier.
- Butler, J. (2015). *Advanced Topics in Forensic DNA Typing: Interpretation*. San Diego: Elsevier.
- Castillo, V., Uranga, R., & Zafra, G. (2019). *Genética Clínica*. México: Manual moderno.
- Charlesworth, B., & Jensen, J. (2021). Population Genetics. *Reference Module in Life Sciences*, 2-17.
- Craig, A. (2010). *Essential Mathematics and Statistics for Forensic Science*. Hoboken: WILEY-BLACKWELL.
- Dieringer, D., & Schlotterer, C. (2003). Two Distinct Modes of Microsatellite Mutation Processes: Evidence From the Complete Genomic Sequences of Nine Species. *Genome Research*, 2242-2251.
- Easy DNA México. (2020). *Índice de Paternidad*. Obtenido de EasyDNA: <https://www.easydna.com.mx/terminos-del-glosario/indice-paternidad/#>
- Encina, G. (2013). Biología molecular en oncología: lo que un clínico debiera saber. *Revista Médica Clínica Las Condes*, 563-570.
- FBI. (2005). *CODIS*. Obtenido de Fedeark bureau of investigation: <https://www.fbi.gov/how-we-can-help-you/dna-fingerprint-act-of-2005-expungement-policy/codis-and-ndis-fact-sheet#CODIS>
- Frances, F., Castello, A., & Verdu, F. (2008). El diagnóstico genético del sexo mediante el test de la amelogenina: Métodos y posibles fuentes de error. *Cuadernos de Medicina Forense*, 119-125.

- GIZ México, ONU-DH, & CICR. (22 de Junio de 2021). *Ante la emergencia forense en México, organismos internacionales lanzan identificacionhumana.mx*. Obtenido de Identificación humana.mx: https://www.identificacionhumana.mx/wp-content/uploads/2021/06/20210620-Comunicado-de-prensa-identificacionhumana.mx_.docx.pdf
- González, L., Sosa, J., López, P., López, M., Gamboa, R., Herrera, R., . . . Rangel, H. (2019). Forensic statistical parameters of 22 autosomal STRs in Mestizos from the Peninsula of Yucatán, México. *Forensic Science International: Genetics Supplement Series*, 491-493.
- Goodwin, W., Linacre, A., & Hadi, S. (2011). *An Introduction to Forensic Genetics*. Oxford: WILEY.
- Gouy, A., & Zieger, M. (4 de Octubre de 2021). *Statistical forensics made easy | The STRAF Book*. Obtenido de STRAF: <https://agouy.github.io/straf/>
- Hawkins, S., & Guest, P. (2017). Multiplex analyses using real-time quantitative PCR. *Methods in Molecular Biology*, 125-133.
- Hernández, A., & Trejo, F. (2014). Estudio Genético Poblacional de Frecuencias Alélicas para 15 marcadores STR presentes en la Población del Estado de Zacatecas Aplicado a la Práctica Forense. *iMedPub Journals*, 1-24.
- Herrero , J., Fernández, E., Iborra, F., & Ribera, J. (2021). *esenciales., Biología Celular: Conceptos*. México: Editorial Médica Panamericana.
- Imle , P. (2005). Fluorescence-based fragment size analysis. *Methods in molecular biology*, 139-146.
- INEGI. (29 de Enero de 2021). *En la ciudad de México somos 9209944 habitantes: censo de población y vivienda 2020*. Obtenido de INEGI: https://www.inegi.org.mx/contenidos/saladeprensa/boletines/2021/EstSociodem/ResultCenso2020_CdMx.pdf
- INEGI. (25 de Enero de 2021). *En México somos 126014024 habitantes; censo de población y vivienda 2020*. Obtenido de INEGI:

https://www.inegi.org.mx/contenidos/saladeprensa/boletines/2021/EstSociodem/ResultCenso2020_Nal.pdf

INEGI. (29 de Enero de 2021). *Número de habitantes. Ciudad de México*. Obtenido de INEGI: <https://cuentame.inegi.org.mx/monografias/informacion/df/poblacion/default.aspx?tema=me&e=09>

INEGI. (25 de Enero de 2021). *Número de habitantes. Cuéntame de México*. Obtenido de INEGI: <https://cuentame.inegi.org.mx/poblacion/habitantes.aspx?tema=P>

Jamieson, A., & Bader, S. (2016). *A guide to forensic DNA profiling*. Chichester: WILEY.

Karp, G. (2020). *Biología celular y molecular. Conceptos y experimentos*. México: Mc Graw Hill.

Lachance, J. (2016). Hardy–Weinberg Equilibrium and Random Mating. *Encyclopedia of Evolutionary Biology*, 208-211.

Leclercq, S., Rivals, E., & Jarne, P. (2010). DNA Slippage Occurs at Microsatellite Loci without Minimal Threshold Length in Humans: A Comparative Genomic Approach. *Genome Biology And Evolution*, 325-335.

Lesack, K., & Naugler, C. (2011). An open-source software program for performing Bonferroni and related corrections for multiple comparisons. *Journal of Pathology Informatics*, 2:52.

Linacre, A., & Tobe, S. (2013). *Wildlife DNA analysis : applications in forensic science*. Oxford: WILEY-BLACKWELL.

Loscalzo, J., Fauci, A., Kasper, D., Hauser, S., Longo, D., & Jameson, L. (2022). *Principios de Medicina Interna*. México: Mc Graw Hill.

Martínez, G., Zuñiga, F., Celorio, A., Ruiz, E., Antelo, A., Dalpozzo, V., . . . Rangel, H. (2019). Population data for 21 autosomal STR loci (GlobalFiler kit) in two

Mexican-Mestizo population from the northwest, Mexico. *International journal of legal medicine*, 781-783.

Mathews, C., Holde, V., & Appling, D. (2013). *Bioquímica*. Madrid: Pearson.

Melo, G., Uscanga, K., López, M., Avilés, D., Hernández, P., Rangel, H., . . . López, N. (2022). Use of Investigator 24plex GO! To analyse allele frequencies of 21 autosomal STRs in the population of Veracruz state, Mexico. *Annals of human biology*, 164-169.

Mendoza, M., & Pérez, A. (2023). Herramientas de almacenamiento, cotejo y análisis de datos genéticos para la identificación forense en México. *Revista Digital de Ciencia Forense*, 61-81.

Michaelis, R., Flanders, R., & Wulff, P. (2008). *A litigator's guide to DNA from the laboratory to the courtroom*. California: Elsevier.

Moreno, J., Bolivar, A., Rojas, A., Garcia, P., & Torrealba, J. (2014). PCR y PCR-Múltiple: parámetros críticos y protocolo de estandarización. *Avances en biomedicina*, 25-33.

NIH. (5 de Abril de 2022). *¿Qué es la genética?* Obtenido de National Institute of General Medical Sciences : <https://nigms.nih.gov/education/factsheets/Pages/genetics-spanish.aspx>

NIH. (2022 de Marzo de 2022). *EDTA*. Obtenido de Instituto Nacional Del Cáncer: <https://www.cancer.gov/espanol/publicaciones/diccionarios/diccionario-cancer/def/edta>

NIH. (16 de Mayo de 2023). *EXÓN*. Obtenido de National Human Genome Research Institute: <https://www.genome.gov/es/genetics-glossary/Exon>

NIH. (16 de Mayo de 2023). *Intrón*. Obtenido de National Human Genome Research Institute: <https://www.genome.gov/es/genetics-glossary/Intron>

- NIH. (16 de Mayo de 2023). *Marcador Genético*. Obtenido de National Human Genome Research Institute: <https://www.genome.gov/es/genetics-glossary/Marcador-genetico>
- NIH. (12 de Mayo de 2023). *NIH*. Obtenido de NIH: <https://www.genome.gov/genetics-glossary/Genomic-Variation>
- NIH. (16 de Mayo de 2023). *Nucleosoma*. Obtenido de National Human Genome Research Institute: <https://www.genome.gov/es/genetics-glossary/Nucleosoma>
- NIH. (16 de Mayo de 2023). *Recombinación homóloga* . Obtenido de National Human Genome Research Institute: <https://www.genome.gov/es/genetics-glossary/Recombinacion-homologa>
- NIH. (16 de Mayo de 2023). *Variabilidad genética*. Obtenido de National Human Genome Research Institute : <https://www.genome.gov/es/genetics-glossary/Variabilidad-genetica>
- Nussbaum, R., McInnes, R., & Willard, H. (2008). *Thompson & Thompson. Genética en medicina*. Barcelona: Elsevier.
- Peral, P., Giovambattista, G., & Ripoli, V. (2015). *Genética forense no-humana*. Buenos aires: Editorial de la Universidad de La Plata.
- Pierce, B. (2014). *Genética. Un enfoque conceptual*. New York: Panamericana.
- Pritchard, D. (2013). *Genética Médica / Lo esencial de un vistazo*. México: Panamericana.
- Promega. (2018). *PowerPlex® Fusion 6C System for Use on the Applied Biosystems® Genetic Analyzers Instructions for Use of Products DC2705, DC2720 and DC2780*. Madison: Promega.
- Qiagen. (Agosto de 2012). *Kits Investigator Quantiplex*. Obtenido de Qiagen: file:///C:/Users/satellite/Downloads/1073542_HB_AT_QuantiplexHYres_0812_WW_ES.pdf

- Qiagen. (Julio de 2020). *DNeasy Blood & Tissue Kits*. Obtenido de Qiagen: file:///C:/Users/satellite/Downloads/HB-2061-003_HB_DNY_Blood_Tissue_0720_WW.pdf
- Ramírez, E. (2014). Análisis Genético de 22 Loci STR en la Población. (*Tesis de licenciatura*). Universidad Nacional Autónoma De México Facultad De Estudios Superiores Zaragoza, Ciudad de México.
- Rangel , H., Castellanos, R., Morales, M., Molina, V., Licea, R., Rizzo, S., . . . Muñoz, J. (2010). Admixture estimates and statistical parameters of forensic importance based on PowerPlex® 16 system in Mexican-Mestizos from the States of Guanajuato (Center) and Veracruz (East). *Forensic Science International: Genetics*, 271-272.
- Rangel, H. (2020). Importancia de la interpretación estadística en ciencias forenses. *Identificación humana.mx*, 1-6.
- Ranjan, H., Shrivastava, P., & Das, S. (2020). *Principles and Practices of DNA Analysis: A laboratory Manual for Forensic DNA Typing*. New York: Humana Press.
- Ranjan, H., Shrivastava, P., Mohapatra, B., & Das, S. (2018). *DNA Fingerprinting: Advancements and Future endeavors*. Singapore: Springer.
- Rubio , S., Pacheco, R., Gómez, A., Perdomo, S., & García , R. (2020). Secuenciación de nueva generación (NGS) de ADN: presente y futuro en la práctica clínica. *Universitas Médica*, 1-15.
- Rudin, N., & Inman, K. (2002). *An introduction to forensic DNA analysis*. Florida: CRC PRESS.
- Salazar , A., & Guerrero , J. (2016). *Biología Molecular. Fundamentos y aplicaciones en las ciencias de la salud*. México: Mc Graw Hill.
- Serrato , A., Flores, L., Aportela, J., & Sierra, E. (2014). PCR: reacción en cadena de la polimerasa. *ResearchGate*, 53-73.

- Shrivastava, P., Ranjan, H., Mohapatra, B., & Das, S. (2020). *Forensic DNA Typing: Principles, Applications and Advancements*. Singapore: Springer.
- Smith, A., & Nelson, R. (2003). Capillary Electrophoresis of DNA. *Current Protocols in Nucleic Acid Chemistry*, 10.9.1-10.9.16.
- Thompson, T., & Black, S. (2007). *Forensic human identification : an introduction*. New York: CRC Press.
- Torres, G., López, M., & Villavicencio, A. (2022). Análisis del DNA y su importancia en la crisis forense de identificación de cadáveres desconocidos-Identificación humana. *Identificación humana.mx*, 1-10.
- Vázquez, C. (2022). *Manual de prueba pericial*. México: Suprema Corte de Justicia de la Nación.
- Velarde, S., Molina, C., Solorzano, S., Cázarez, G., Rendón, H., Murillo, J., & Ríos, J. (2008). Identificación del sexo mediante análisis molecular del gen de la amelogenina. *Revista Mexicana de Patología Clínica y Medicina de Laboratorio*, 12-20.
- Watson, J., Baker, T., Bell, S., Gann, A., Levine, M., & Losick, R. (2016). *Biología molecular del gen*. México: Panamericana.
- Wickenheiser, R. (2022). Expanding DNA database effectiveness. *Forensic Science International: Synergy*, 1-17.
- Zauza, M., Rosenfeld, F., & Estrada, H. (2015). Análisis de las variaciones en el número de repeticiones de 5 marcadores ancestrales en donadores recurrentes en México. *Perinatología y Reproducción Humana*, 152-156.

ANEXO 1

1.1 Electroferogramas de los controles.

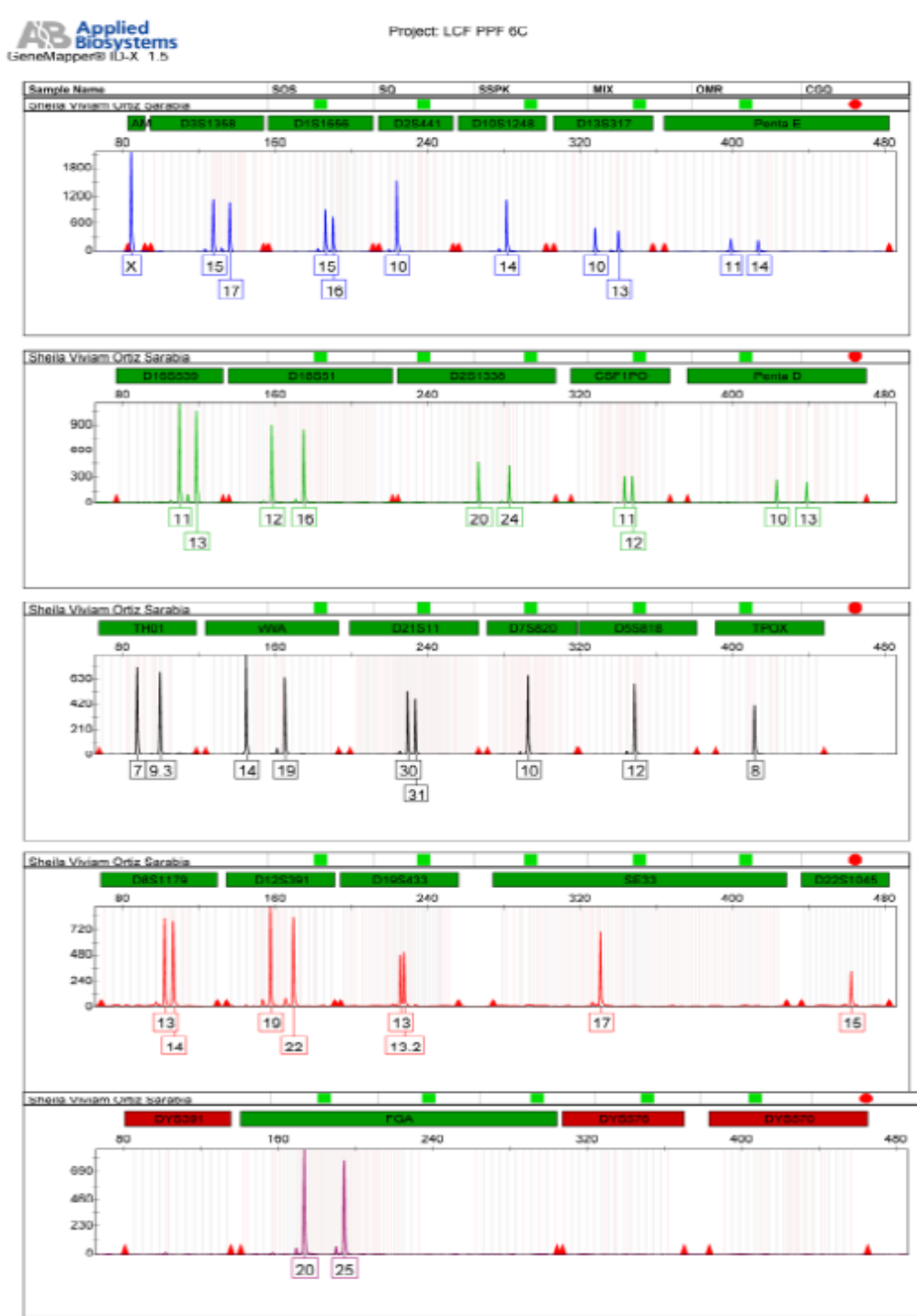


Figura 1. Perfil genético de marcadores STR del analista (electroferograma).

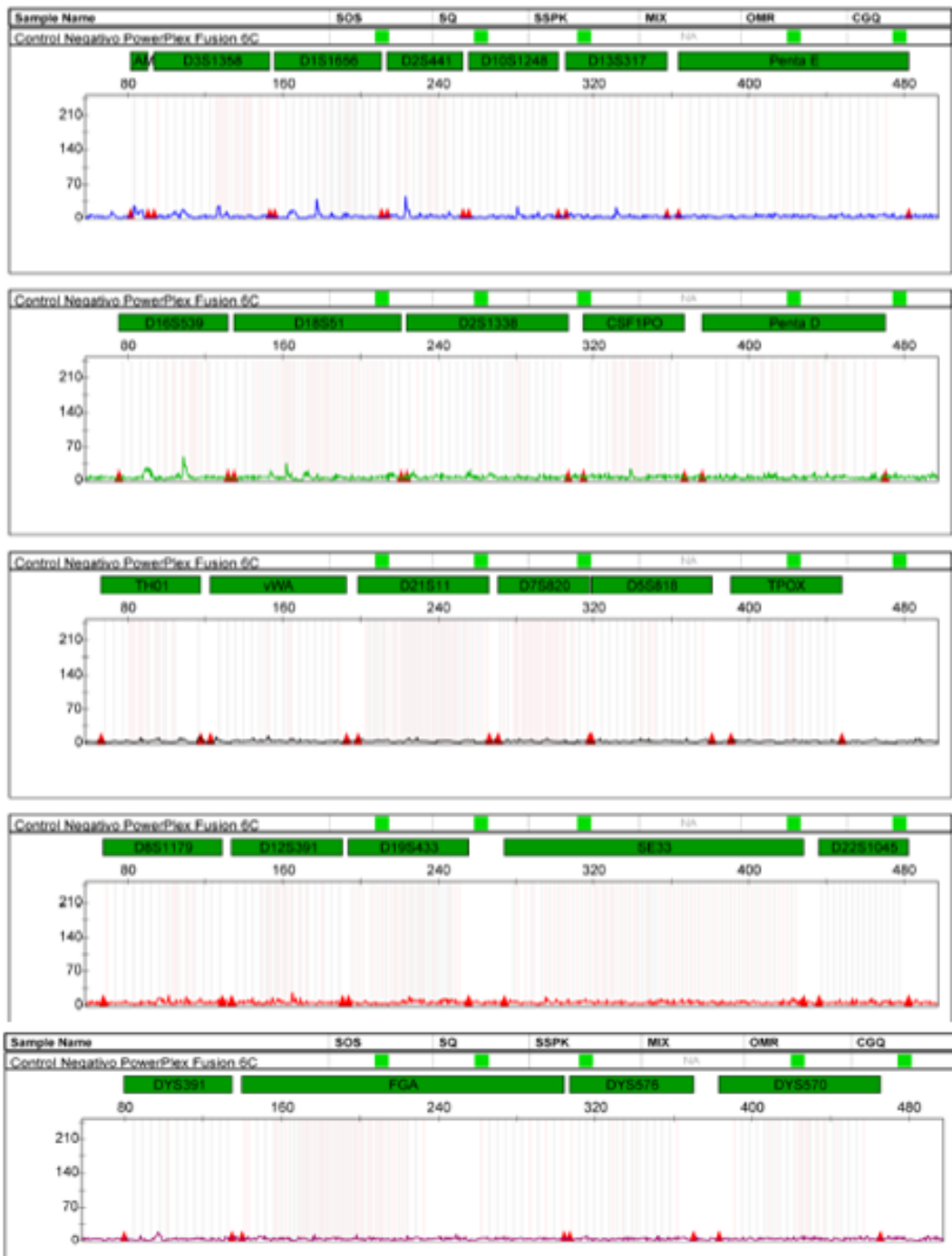


Figura 2. Electroferograma del control negativo proporcionado por el kit PPF6C.

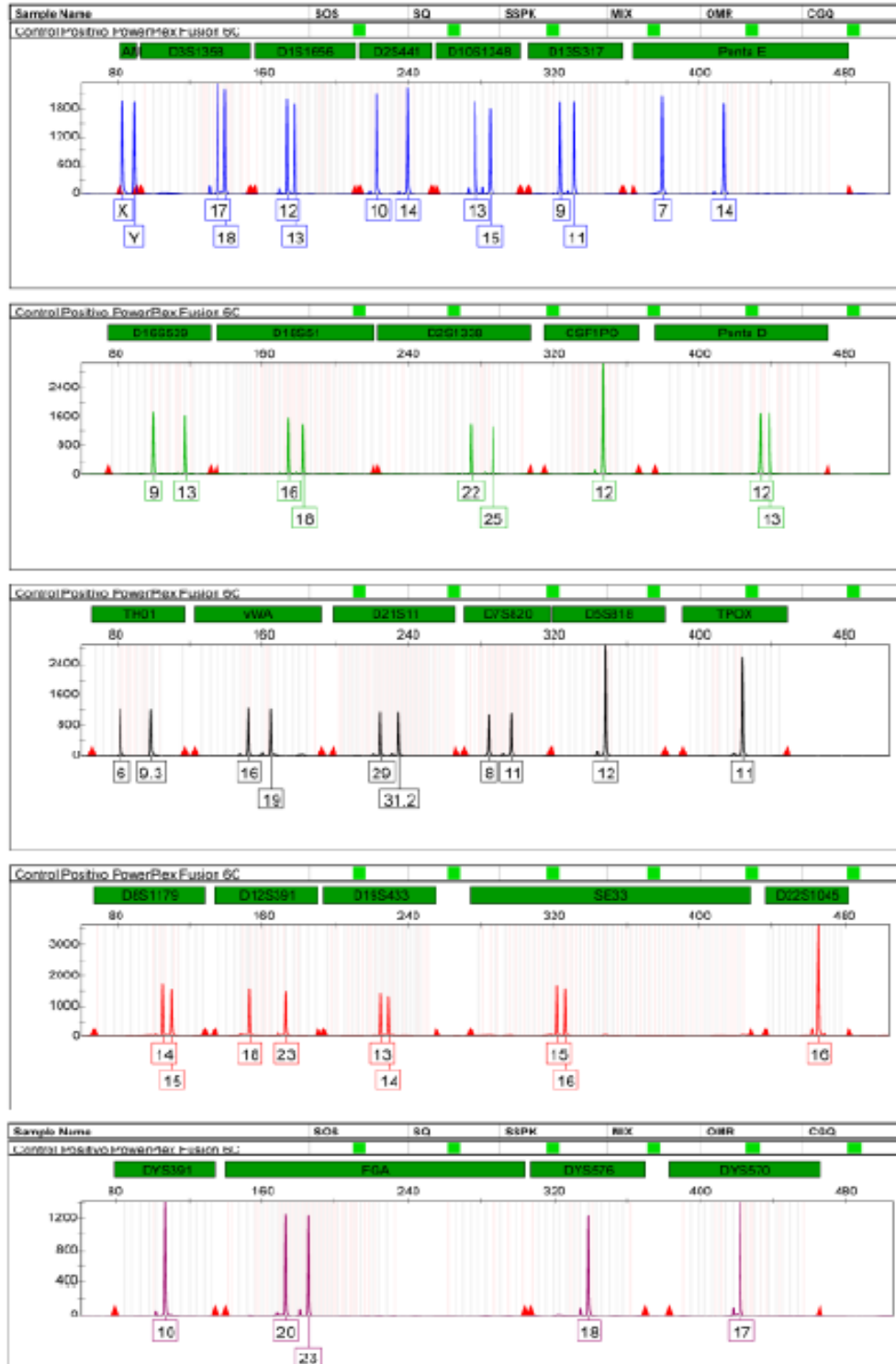


Figura 3. Electroferograma del DNA control 2800M proporcionado por el kit PPF6C.

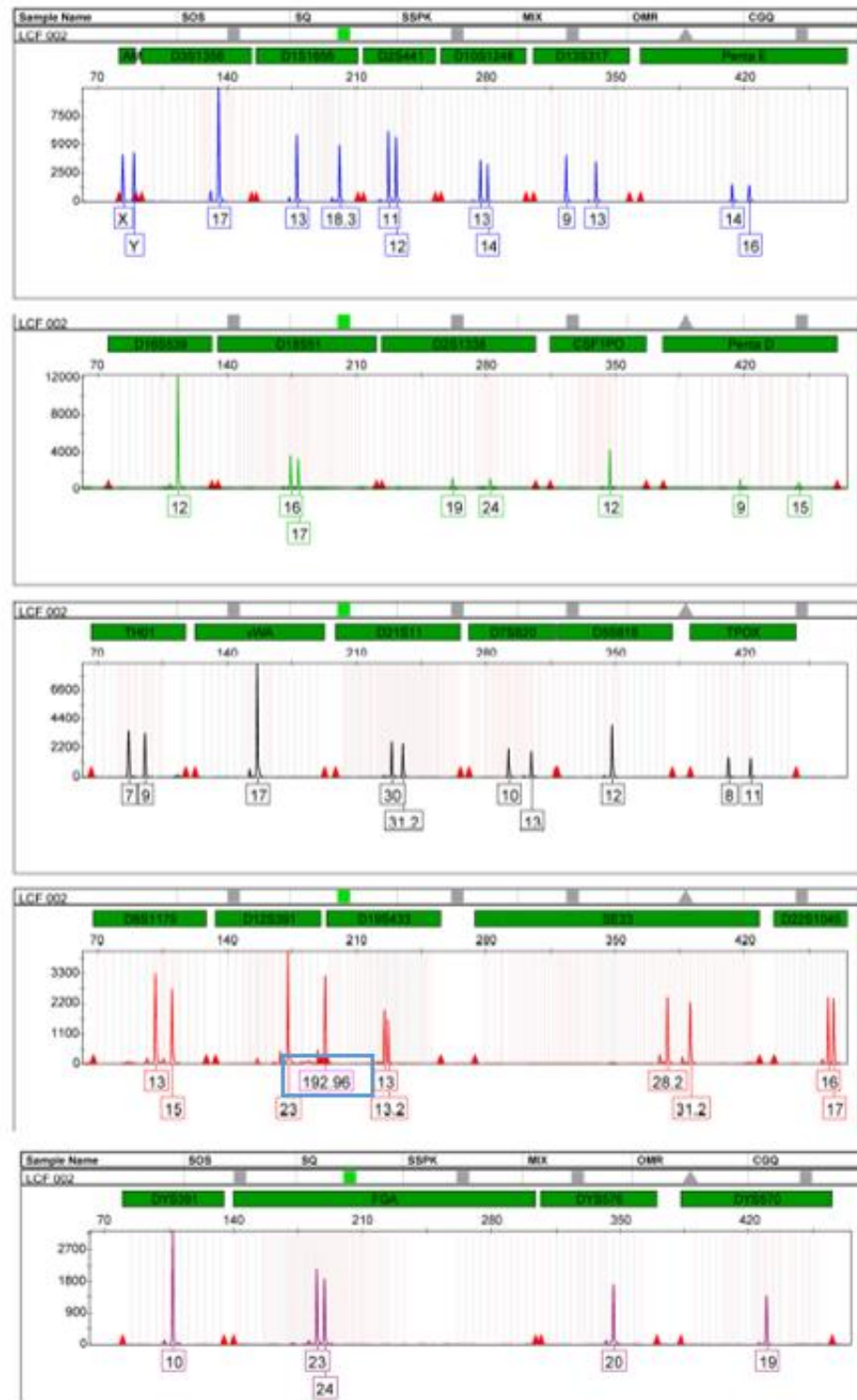


Figura 4. Electroferograma de la muestra LCF 002. Se muestra encerrado con el recuadro azul un alelo fuera de escalera alélica en el marcador D12S391.

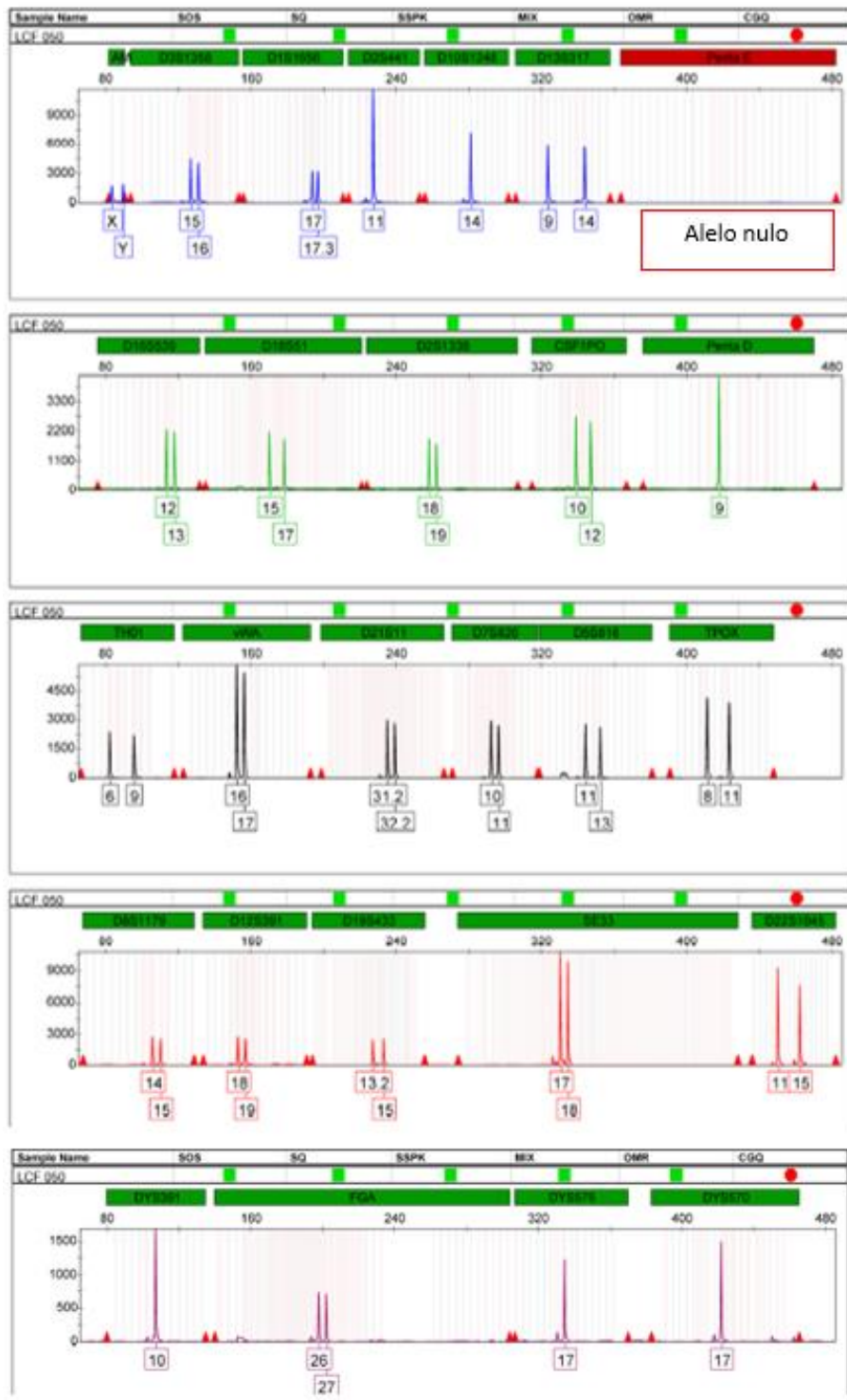


Figura 5. Electroferograma de la muestra LCF 050. Se obtuvo un alelo nulo en el marcador penta E que se puede observar en la zona marcada con el recuadro rojo.

1.2 Matrices para cada locus.

Alelo	13	14	15	16	17	18	19	n	Frecuencia
13			1					1	0.01
14			3	1				4	0.04
15			9	12	8	5	1	48	0.48
16				1	7	1		23	0.23
17					1			17	0.17
18								6	0.06
19								1	0.01

Matriz 1. Matriz para el locus D3S1358.

Alelo	11	12	13	14	15	15.3	16	16.3	17	17.3	18	18.3	19.3	n	Frecuencia
11			1											1	0.01
12			2	1	2		2	1		1			1	10	0.1
13			1	1	5					3		2		16	0.16
14							4			4				10	0.1
15					2	1	2			4	1			19	0.19
15.3														1	0.01
16								3			2			16	0.16
16.3									1					2	0.02
17											2			3	0.03
17.3												1		17	0.17
18														1	0.01
18.3														3	0.03
19.3														1	0.01

Matriz 2. Matriz para el locus D1S1656.

Alelo	10	11	11.3	12	13	14	15	n	frecuencia
10	10	11	2	2	2	3	1	41	0.41
11		5	1	2	1	9		34	0.34
11.3								3	0.03
12								4	0.04
13								3	0.03
14							1	14	0.14
15								1	0.01

Matriz 3. Matriz para el *locus* D2S441.

Alelo	10	11	12	13	14	15	16	17	n	Frecuencia
10							1		1	0.01
11								1	1	0.01
12					1	1		1	3	0.03
13				2	6	9	1		20	0.2
14					10	10	2		39	0.39
15						5			30	0.3
16									5	0.05
17									1	0.01

Matriz 4. Matriz para el *locus* D10S1248.

Alelo	8	9	10	11	12	13	14	15	n	Frecuencia
8	1	2			2				6	0.06
9		3	2	2	6	1	2	1	22	0.22
10			1	2	3	1	1		11	0.11
11				4	4	4	1		21	0.21
12					2	1	3		23	0.23
13							1		8	0.08
14									8	0.08
15									1	0.01

Matriz 5. Matriz para el *locus* D13S317.

Alelo	7	8	11	12	13	14	15	16	17	18	19	20	21	22	23	n	frecuencia
7			1	1		2	1					1				6	0.06122449
8				1												1	0.01020408
11				2	2	1		1								7	0.07142857
12				1	2	5	3	1	1	1	1					20	0.20408163
13					2	1		1			1					11	0.1122449
14						1		1		1	1					14	0.14285714
15							1	2				1	1		1	11	0.1122449
16									1	1					1	9	0.09183673
17																2	0.02040816
18										1						5	0.05102041
19												2				5	0.05102041
20																4	0.04081633
21																1	0.01020408
22																1	0.01020408
23																1	0.01020408

Matriz 6. Matriz para el *locus* Penta E.

Alelo	9	10	11	12	13	14	n	frecuencia
9		3	2	2	2	1	10	0.1
10		3	5	7	2		23	0.23
11			2	8	2		21	0.21
12				5	2	2	31	0.31
13					1	1	11	0.11
14							4	0.04

Matriz 7. Matriz para el *locus* D16S539.

Alelo	9	11	12	13	14	15	16	17	18	19	20	21	n	Frecuencia
9							1						1	0.01
11								1					1	0.01
12				1	1		2	1		2			7	0.07
13				1	5	2	3		1	1			15	0.15
14						3	1	2	2		1	1	16	0.16
15						1	2	3					12	0.12
16							1	4		1	1		17	0.17
17								2		2			17	0.17
18									1				5	0.05
19													6	0.06
20													2	0.02
21													1	0.01

Matriz 8. Matriz para el *locus* D18S51.

Alelo	14	16	17	18	19	20	21	22	23	24	25	27	n	Frecuencia
14			1										1	0.01
16	1						1						2	0.02
17		1			1	2		2	1	1	1	1	21	0.21
18			1					1	2				4	0.04
19				1		6	1		7	1			17	0.17
20					1			2	2	1			13	0.13
21						1			1				3	0.03
22							1				2		9	0.09
23								1	1	2			18	0.18
24									1				6	0.06
25										1			5	0.05
27											1		1	0.01

Matriz 9. Matriz para el *locus* D2S1338.

Alelo	8	9	10	11	12	13	14	17	n	Frecuencia
8					1				1	0.01
9					1	1			2	0.02
10			5	9	9	2			30	0.3
11				3	8	3			26	0.26
12					4	3			30	0.3
13									9	0.09
14								1	1	0.01
17									1	0.01

Matriz 10. Matriz para el *locus* CSF1PO.

Alelo	2.2	7	9	10	11	12	13	14	15	n	Frecuencia
2.2					1					1	0.01
7			1							1	0.01
9			6	3	4	1	2	2	1	26	0.26
10				1	8	4	5	1		23	0.23
11						1	2			16	0.16
12							1	4		12	0.12
13								2		17	0.17
14										3	0.03
15										1	0.01

Matriz 11. Matriz para el *locus* Penta D.

Alelo	6	7	8	9	9.3	10	n	Frecuencia
6	4	8	2	1	11		30	0.3
7		2	5	3	5	1	26	0.26
8				2			9	0.09
9					3		9	0.09
9.3					3		25	0.25
10							1	0.01

Matriz 12. Matriz para el *locus* TH01.

Alelo	14	15	16	17	18	19	n	Frecuencia
14		1	1	1			3	0.03
15			4	1	1		7	0.07
16			7	9	10	1	39	0.39
17				7	4	1	30	0.3
18					2		19	0.19
19							2	0.02

Matriz 13. Matriz para el *locus* vWA.

Alelo	26	27	28	29	30	30.2	31	31.2	32.2	33.2	n	Frecuencia
26								1			1	0.01
27					2			1			3	0.03
28					1		1		1		3	0.03
29				1	8	1		1	4		16	0.16
30					4		3	3	5	1	31	0.31
30.2										1	2	0.02
31								1	1		6	0.06
31.2								1	4	1	14	0.14
32.2									2		19	0.19
33.2										1	5	0.05

Matriz 14. Matriz para el *locus* D21S11.

Alelo	7	8	9	10	11	12	13	n	Frecuencia
7					1			1	0.01
8			1		3	2		7	0.07
9				1	7	2		11	0.11
10					4	6	4	26	0.26
11						2	10	29	0.29
12							2	21	0.21
13								5	0.05

Matriz 15. Matriz para el *locus* D7S820.

Alelo	7	8	9	10	11	12	13	14	n	Frecuencia
7				1	2				3	0.03
8						1			1	0.01
9				1	2	3	1		7	0.07
10				1	4	2			10	0.1
11					5	11	6	1	36	0.36
12						5	4		31	0.31
13									11	0.11
14									1	0.01

Matriz 16. Matriz para el *locus* D5S818.

Alelo	6	7	8	9	10	11	12	n	Frecuencia
6			2					2	0.02
7			1					1	0.01
8			8	3	4	16	6	48	0.48
9						2		5	0.05
10						1	1	6	0.06
11						3	3	28	0.28
12								10	0.1

Matriz 17. Matriz para el *locus* TPOX.

Alelo	8	10	11	12	13	14	15	16	17	n	Frecuencia
8						1				1	0.01
10		1		2	1	1	2			8	0.08
11				1	1	2	1	1		6	0.06
12					2	5				10	0.1
13					5	9	3	1	1	28	0.28
14						6	4			34	0.34
15										10	0.1
16										2	0.02
17										1	0.01

Matriz 18. Matriz para el locus D8S1179.

Alelo	15	17	17.3	18	18.3	19	20	21	22	23	25	n	Frecuencia
15				1		1			1			3	0.03
17		1	1	1		4						8	0.08
17.3						1						2	0.02
18				2		6	1	1	1			15	0.15
18.3							1					1	0.01
19						4	7	5	2			34	0.34
20							2	1		1	1	16	0.16
21								1	1	1		11	0.11
22												5	0.05
23										1		4	0.04
25												1	0.01

Matriz 19. Matriz para el *locus* D12S391

Alelo	12	13	13.2	14	14.2	15	15.2	16	16.2	17	n	Frecuencia
12		1	2		1		1				5	0.05
13		1	3	3		4	2	1	1		17	0.17
13.2				3		4	1				13	0.13
14				4	1	5	2	1			23	0.23
14.2						3	1				6	0.06
15						1	1	1	1	1	22	0.22
15.2											8	0.08
16											3	0.03
16.2											2	0.02
17											1	0.01

Matriz 20. Matriz para el *locus* D19S433.

Alelo	13	14	15	16	17	18	19	20	21	21.2	22	24.2	25.2	26.2	27.2	28.2	29.2	30.2	31.2	32.2	17.3	n	Frecuencia	
13					1																		1	0.01
14		1								1													3	0.03
15			1					1						1									3	0.03
16				1					1						1								4	0.04
17					1					1					1					1			6	0.06
18						1	1				1				1	2	1	4					13	0.13
19							2	2						1		3					1		12	0.12
20															1								7	0.07
21																1							1	0.01
21.2																							2	0.02
22																	1						2	0.02
24.2																2		1					4	0.04
25.2																							5	0.05
26.2																							5	0.05
27.2																							10	0.1
28.2																							5	0.05
29.2																							12	0.12
30.2																							1	0.01
31.2																							2	0.02
32.2																							1	0.01
17.3																							1	0.01

Matriz 21. Matriz para el *locus* SE33.

Alelo	11	12	13	14	15	16	17	18	n	Frecuencia
11					3	5			8	0.08
12						1			1	0.01
13					1				1	0.01
14					1	1			2	0.02
15					13	16	1		48	0.48
16						5	2		35	0.35
17								1	4	0.04
18									1	0.01

Matriz 22. Matriz para el *locus* D22S1045.

Alelo	18	18.2	19	20	21	22	23	24	25	26	27	n	Frecuencia
18								1				1	0.01
18.2				1								1	0.01
19			1		1	1	1		1	1		7	0.07
20					1	1	2	2	1	1		9	0.09
21							1	1	1	1	1	7	0.07
22						1	1	2	4	1		12	0.12
23							3	3	2	1	1	18	0.18
24									3	3		15	0.15
25									1			14	0.14
26										1	2	12	0.12
27												4	0.04

Matriz 23. Matriz para el *locus* FGA.