



**UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO**

**FACULTAD DE MEDICINA
LICENCIATURA EN CIENCIA FORENSE**

**Análisis poblacional de STRs autosómicos
empleando el kit comercial Investigator® 24plex QS
de la marca Qiagen en 156 individuos mestizo-
mexicanos de la Ciudad de México**

Facultad de Medicina



T E S I S

QUE PARA OBTENER EL TÍTULO DE:

LICENCIADA EN CIENCIA FORENSE

P R E S E N T A:

MARIAN LIZZETTE SALINAS PINEDA



**CIENCIA
FORENSE**
U N A M

DIRECTOR DE TESIS:

Dr. José Mariano Guardado Estrada

Cd. Mx. 2022



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.

Análisis poblacional de STRs autosómicos empleando el kit comercial Investigator® 24plex QS de la marca Qiagen en 156 individuos mestizo-mexicanos de la Ciudad de México

Agradecimientos técnicos

Gracias a las personas que me ayudaron durante mi proceso de titulación:

Al doctor Mariano Guardado por su apoyo incondicional, enseñanzas y guía desde que le planteé por primera vez mi interés en realizar este proyecto de tesis.

A la maestra Carolina León Campos y al maestro Mauro López Armenta por su asesoría y enseñanzas, así como por recibirme en el laboratorio de genética del INCIFO para poder realizar la presente investigación.

A la doctora Sofía Huerta por apoyarme con el análisis de datos.

Agradecimientos personales

A las doctoras: Ana Pamela Romero, Alexa Villavicencio Queijeiro, Laura Espinosa y Patricia López Olvera que me dieron clases en la licenciatura y me inspiraron a seguir en el camino de la ciencia y la investigación.

A mis amigxs que conocí en la preparatoria, en especial a Karla y Brenda por apoyarme desde que tome la decisión de entrar a esta carrera y hasta ahora por escucharme cada que les hablaba acerca de este proyecto.

A mis amigxs de la carrera, Jazmín y Brenda por ser mis amigas y mi equipo durante toda la carrera y a Diana por todas sus enseñanzas y apoyo.

Sobre todo, gracias a mi familia, Rosario, Lázaro y Monse por apoyarme y brindarme todo lo necesario durante la carrera y en mi vida en general para poder lograr la meta que siempre quise de ser científica forense.

Índice

ABREVIATURAS	1
GLOSARIO.....	2
PLANTEAMIENTO DEL PROBLEMA	4
JUSTIFICACIÓN.....	5
OBJETIVOS.....	6
Objetivo general	6
Objetivos particulares.....	6
MARCO TEÓRICO	7
Generalidades del DNA	7
<i>Componentes y estructura</i>	<i>7</i>
<i>Recombinación genética.....</i>	<i>8</i>
<i>Mutaciones.....</i>	<i>8</i>
<i>Variación genética.....</i>	<i>8</i>
Marcadores de linaje.....	9
<i>X-STR.....</i>	<i>9</i>
<i>Y-STR.....</i>	<i>9</i>
<i>DNA mitocondrial.....</i>	<i>9</i>
STRs autosómicos.....	10
<i>Nomenclatura.....</i>	<i>10</i>
Análisis general de muestras biológicas en genética forense.....	11
Técnicas de biología molecular involucradas en el análisis de muestras biológicas en genética forense	12
<i>PCR.....</i>	<i>12</i>
<i>PCR en tiempo real.....</i>	<i>14</i>
<i>PCR multiplex.....</i>	<i>15</i>
<i>Electroforesis capilar</i>	<i>15</i>
Kits comerciales para la amplificación de STRs	17
<i>Investigator® 24plex Qiagen.....</i>	<i>18</i>
Utilidad del análisis poblacional en el estudio de STRs autosómicos con fines forenses.....	19

<i>Frecuencias alélicas</i>	19
<i>Frecuencias genotípicas</i>	19
<i>Equilibrio de Hardy-Weinberg</i>	19
<i>Homocigosidad esperada</i>	20
<i>Heterocigosidad observada</i>	21
<i>Heterocigosidad esperada</i>	21
<i>Probabilidad de coincidencia al azar</i>	21
<i>Poder de discriminación</i>	22
<i>Contenido de información polimórfica</i>	22
<i>Poder de exclusión</i>	23
<i>Índice típico de paternidad</i>	23
<i>Distancias genéticas</i>	23
Pruebas genéticas en el ámbito forense	25
<i>Identificación humana en contextos criminalísticos</i>	25
<i>Identificación humana en contextos de desaparición</i>	26
<i>Relaciones de parentesco</i>	26
METODOLOGÍA	28
Cuantificación de DNA	30
Ensayos para el protocolo de amplificación	31
Amplificación de los marcadores STRs	32
Electroforesis capilar	33
Obtención de los alelos y genotipos	34
Análisis genético poblacional	34
RESULTADOS Y DISCUSIÓN	36
Protocolo de PCR	36
<i>Electroferogramas</i>	37
Frecuencias alélicas	43
Parámetros estadísticos de interés forense de una base de datos	46
<i>Para el sistema 24plex QS población mestizo-mexicana de la CDMX</i>	46
<i>Comparación con estudios similares realizados dentro de la República Mexicana</i>	48

<i>Comparación de las frecuencias alélicas con dos análisis similares realizados dentro de la República Mexicana.....</i>	<i>51</i>
CONCLUSIONES	64
REFERENCIAS.....	65
Consentimiento informado del proyecto “Análisis del genoma mitocondrial para la identificación forense en México”	70
ANEXO 2	79
Matrices por marcador	79

ABREVIATURAS

ddDNTP. Didesoxirribonucleótidos trifosfatos

DNA. Ácido desoxirribonucleico (del inglés *desoxyribonucleic acid*)

He. Heterocigotos esperados.

Ho. Heterocigotos observados.

HW. Hardy-Weinberg.

ITP. Índice típico de paternidad.

LCF. Licenciatura en Ciencia Forense

PCR. Reacción en cadena de polimerasa (del inglés *polymerase chain reaction*).

PD. Poder de discriminación.

PE. Poder de exclusión.

PIC. Contenido de información polimórfica (del inglés *polymorphic information content*).

PM. Probabilidad de coincidencia al azar (del inglés *probability of matching*).

RFU. Unidades de fluorescencia relativa (del inglés *relative fluorescence units*).

STR. Repeticiones cortas en tándem (del inglés *short tandem repeats*)

GLOSARIO

Alelo. Cada una de las versiones de un gen. (Ritu & Mohapatra, 2018)

Allelic ladder o escalera alélica. Mezcla artificial de los alelos comunes presentes en una población para un determinado marcador STR. (Butler J. , 2010)

Base de datos poblacional. Perfiles genéticos almacenados de individuos representativos de la población. Permite obtener frecuencias génicas y otros parámetros poblacionales necesarios para el cálculo del poder de exclusión y los índices forenses. (Giovambattista, Goszczynski, Fernandez, & Peral, 2015)

Cromosoma. Paquete ordenado de DNA que se encuentra en el núcleo de la célula. (NHGRI, 2022)

Diploide. Es un término que se refiere a la presencia de dos conjuntos completos de cromosomas en las células de un organismo, donde cada progenitor aporta un cromosoma a cada par. Los seres humanos son diploides y la mayoría de las células del cuerpo contienen 23 pares de cromosomas. (NGHRI, 2022)

Electroforesis. Técnica utilizada para separar moléculas por su tamaño y carga eléctrica. En genética se utiliza usualmente para separar el DNA, el RNA o proteínas. (NHGRI, 2022)

Fluoróforos. Es una molécula que posee la capacidad de absorber radiación de una longitud de onda determinada, para luego emitir esa energía absorbida de una forma diferente a la utilizada para su excitación, lo cual permite que se identifique. (De la Torre Pretell, Roe Battistini, & Roe Battistini, 2016)

Haplotipo. Agrupación física de polimorfismos que tienden a heredarse juntos. (NHGRI, 2022)

Intrón. Es una región de DNA que reside en el interior de un gen, pero no permanece en la molécula madura final de ARNm después de la transcripción y procesamiento por splicing de ese gen y no codifica para los aminoácidos que confirman la proteína codificada por ese gen. (NGHRI, 2022)

Locus. Lugar específico del cromosoma donde está localizado un gen u otra secuencia de DNA. (NHGRI, 2022)

Loci. Plural de *locus*.

Marcador genético. Secuencia de DNA de interés cuya localización física en el cromosoma es conocida.

Microsatélites. Secuencias cortas repetidas de DNA. (NHGRI, 2022)

Mutación. Cambio en la secuencia del DNA. (NHGRI, 2022)

Nucleótido. Componente básico de los ácidos nucleicos, formado por una molécula de ribosa, unida a un grupo fosfato y una base nitrogenada. (NHGRI, 2022)

Oligonucleótidos. Cadenas de aproximadamente 15 a 30 desoxirribonucleótidos, los cuales son complementarios a una cadena distinta del DNA y colocan un límite al fragmento que se desea amplificar. (Bolívar, Rojas, & García-Lugo, 2014)

Perfil genético. Información contenida en las secuencias de DNA, capaces de caracterizar a una persona. (LabGenetics, 2016)

Polimorfismo. Variantes de una secuencia particular del DNA, que están presentes en al menos el 1% de la población analizada. (NHGRI, 2022)

PLANTEAMIENTO DEL PROBLEMA

La genética forense emplea el análisis de marcadores genéticos en los que se incluyen los de tipo STR para hacer identificaciones humanas y establecer relaciones de parentesco biológico. Estos marcadores genéticos se pueden analizar empleando kits comerciales, los cuales pueden amplificar de forma simultánea varios STRs en una muestra, los cuáles pueden ser analizados a través de la electroforesis capilar o secuenciación de nueva generación y posteriormente se realiza un análisis de los resultados obtenidos. El éxito de estos análisis se basa en que, además de contar con experimentos reproducibles, tienen un soporte estadístico que los respalda. Para poder realizar los cálculos estadísticos en este tipo de análisis se requiere de bases de datos poblacionales, las cuales contienen frecuencias alélicas de los STRs que analiza algún kit en particular; esto para realizar los cálculos de poder de exclusión, así como otros índices de genética forense. (Giovambattista, Goszczyński, Fernández, & Peral, 2015)

Aunque existen diversos reportes de bases de datos poblacionales, en la literatura se menciona que es recomendable que cada laboratorio establezca sus bases de datos propias y representativas de la población (Martínez de Pancorbo, 2012); sin embargo, en la práctica se busca que existan bases de datos representativas de cada población y que éstas sean colaborativas y compartidas por los distintos laboratorios que utilicen el mismo sistema o kit.

En el laboratorio de Genética Forense en investigación de la Licenciatura en Ciencia Forense se han llevado a cabo algunos análisis de STRs autosómicos, principalmente con fines académicos. No obstante, al momento de hacer los análisis de dichos marcadores genéticos fue necesario emplear las bases de datos reportadas en la literatura que más se asemejan a la población mestizo-mexicana de la Ciudad de México y que contuvieran la mayoría de los STRs autosómicos analizados por el kit Investigator® 24plex QS de la marca Qiagen, debido a que no existen reportadas bases de datos en las que se utilice este kit para la población de la Ciudad de México. Por tal motivo, surge como necesidad en el laboratorio de Genética Forense, el crear una base de datos poblacional de STRs autosómicos empleando este kit, que capte la variabilidad genética de los individuos en la región del centro de México para que pueda ser compartida con la comunidad científica y con el área de genética forense.

JUSTIFICACIÓN

De acuerdo con el registro nacional de personas desaparecidas o no localizadas (RNPDNO) en mayo del 2022 México superó las 100,000 personas víctimas registradas a partir del 15 de marzo de 1964 (DW, 2022). Esto a pesar de que las instituciones han reconocido la existencia de más de 52,000 personas fallecidas sin identificar. (Movimiento por nuestros desaparecidos en México, 2021)

La genética forense es una rama de la ciencia forense que junto con otras disciplinas permite llevar a cabo la identificación de individuos. La generación de bases de datos genéticas con las frecuencias alélicas de los marcadores, así como de los estimadores forenses puede contribuir a la aclaración de crímenes violentos, delitos sexuales, establecimiento de relaciones de parentesco e identificación de personas desaparecidas. La aceptación social de las pruebas genéticas en el ámbito forense se debe a que están basadas en ensayos reproducibles en el laboratorio, así como un riguroso análisis estadístico. No obstante, para la realización del análisis con estas pruebas, es necesario contar con bases de datos poblacionales con los marcadores genéticos que serán analizados, ya sea STRs autosómicos, del Cromosoma Y o X para contar con las frecuencias adecuadas. (Alonso, 2004). Debido a que las poblaciones, incluyendo la población mexicana, no son homogéneas, por la subestructura de la población, dichas bases de datos tienen que ser representativas de la población donde se llevarán a cabo estos análisis, con el objetivo de obtener cálculos estadísticos más robustos y confiables. La resolución 10/26 del Consejo de Derechos Humanos de la ONU, de 2009, concluyó que el uso de la genética forense y la creación voluntaria de bancos de datos genéticos tienen un rol crucial en la identificación de víctimas de violaciones a los derechos humanos. (Consejo de Derechos Humanos de la ONU, 2009) y (Movimiento por nuestros desaparecidos en México, 2021).

Por ello la creación de bases de datos en todas las áreas forenses dedicadas a la identificación humana es urgente, para hacerle frente a la crisis forense que se vive en México. Si bien el presente proyecto pertenece a la LCF la base de datos resultante será pública, por lo que, cualquier Fiscalía o institución pública que utilice el sistema de identificación Investigator 24plex QS podrá utilizarla.

OBJETIVOS

Objetivo general

Realizar un análisis genético poblacional del kit Investigator[®] 24plex QS de la marca Qiagen en población mestizo-mexicana de la Ciudad de México para uso forense.

Objetivos particulares

- Realizar ensayos para comprobar la funcionalidad del kit Investigator[®] 24plex QS de la marca Qiagen para la amplificación de DNA utilizando PCR multiplex y para la separación de fragmentos por medio de electroforesis capilar a $1/3$ de reacción.
- Establecer una base de datos poblacional de STRs autosómicos empleando el kit comercial Investigator[®] 24plex QS de la marca Qiagen en población mestizo-mexicana de la Ciudad de México.
- Utilizar softwares y estimadores estadísticos especializados en genética de poblaciones, como lo son Arlequín, STRAF y diversas paqueterías del software Rstudio especializadas en análisis de STRs y análisis descriptivo.
- Realizar un análisis interpoblacional con estudios similares reportados en la literatura para distintas poblaciones de México.

MARCO TEÓRICO

La unidad básica de la vida es la célula, la cual transforma las materias primas y para obtener la energía necesaria para mantener la vida. Un ser humano se compone de miles de millones de células y dentro del núcleo de dichas células existe una molécula denominada DNA que contiene la información necesaria para la duplicación de la célula. Debido a que este DNA reside en el núcleo de la célula, se conoce como DNA nuclear. (Butler J. , 2005)

Es importante mencionar que el DNA nuclear no es el único tipo de DNA que existe ya que también se puede encontrar DNA en las mitocondrias. A diferencia del DNA nuclear el DNA mitocondrial no recombinan, si no que se hereda de manera íntegra a partir de las mitocondrias de la madre.

El DNA nuclear almacena información para efecto de hacer copias de sí mismo (replicación) para que las células puedan dividirse y mantener la misma información y lleva las instrucciones para la traducción. En términos generales la información codificada en el DNA se transmite de generación en generación teniendo la mitad de la información genética de origen materno y el resto de origen paterno, con excepción del mtDNA como previamente se menciona y de los cromosomas X y Y. (Butler J. , 2005)

Generalidades del DNA

Componentes y estructura

El DNA está constituido por unidades fundamentales llamadas desoxirribonucleótidos. Cada desoxirribonucleótido contiene un grupo fosfato, un carbohidrato (desoxirribosa) y una base nitrogenada, la cual puede ser adenina, guanina, citosina o timina. La diversidad biológica entre seres humanos y todos los seres vivos depende/está basada en la secuencia de la información codificada en el DNA. Esto es qué bases están contenidas, la cantidad y el orden en que éstas se encuentran. (Butler J. , 2010)

Cada molécula de DNA es una doble hélice formada por dos cadenas complementarias de nucleótidos que se mantienen unidas de manera complementaria por puentes de hidrógeno entre los pares de base de Guanina-Citosina y Adenina-Timina, (Alberts, Johnson, & Lewis, 2002)

Recombinación genética

La recombinación genética es el proceso en el cual tiene lugar la formación de un nuevo DNA a partir de la combinación de dos secuencias genéticas. La recombinación constituye una fuente de variabilidad genética e intercambio físico de segmentos de DNA. (Universidad de la Habana, 2019)

Existen distintos tipos de recombinación genética, la más importante para el presente estudio es la recombinación homóloga que se lleva a cabo durante la meiosis, en la cual pares de cromosomas de los progenitores masculino y femenino se alinean de forma que secuencias parecidas de DNA puedan cruzarse o intercambiarse, de un cromosoma al otro lo que da como resultado la variabilidad genética. (NHGRI, 2022)

Mutaciones

Las mutaciones génicas son los cambios inesperados en la secuencia de pares de bases del DNA dentro de genes concretos. Cuando las mutaciones no tienen ningún agente causal específico ligadas a ellas, se les conoce como mutaciones espontáneas, en general estas son debidas a cambios aleatorios a de la secuencia nucleotídica de los genes. (Klug, Cummings, & Spencer, 2006)

Las mutaciones génicas son el origen de la mayoría de los alelos nuevos, y así como la recombinación genética es el origen de la variación genética entre individuos, las mutaciones son el origen de la variación genética entre poblaciones. (Klug, Cummings, & Spencer, 2006)

Variación genética

La variación genética es definida como las diferencias genómicas existentes entre individuos dentro o entre poblaciones y puede deberse tanto a mutaciones como a la recombinación genética. (Ritu & Mohapatra, 2018)

En el ámbito forense la variación genética tiene una especial importancia ya que debido a esta es posible realizar identificación de personas y establecer relaciones de parentesco biológico, mediante el uso de diversos marcadores biológicos presentes en el DNA como lo pueden ser los marcadores de linaje como el DNA mitocondrial y los STRs del cromosoma X (X-STR) y del cromosoma Y (Y-STR) o marcadores autosómicos como los son los STRs.

Marcadores de linaje

X-STR

En genética forense la mayor cantidad de análisis de X-STR se lleva a cabo para conocer las relaciones biológicas de paternidad entre padre-hija o maternidad entre madre-hijo (Pereira & Gusmão, 2013) e incluso la relación biológica entre dos hermanas. En el primer caso su utilidad se debe que las personas del sexo femenino heredan de manera íntegra el cromosoma X de su padre, es decir, las personas del sexo masculino transmiten a su descendencia femenina todos los marcadores localizados en su cromosoma X en bloque, por lo que, uno de los cromosomas X de cada mujer es idéntico al de su padre. Esta peculiaridad proporciona una herramienta útil en casos de paternidad con descendencia femenina, así como en estudios de filiación o identificación en los que no se cuentan con muestra del supuesto padre y se puede recurrir a hermanas.

Y-STR

El cromosoma Y es heredado de generación en generación de manera íntegra, por lo que se puede utilizar para la identificación de personas desaparecidas, víctimas de desastres masivos, casos criminales como lo pueden ser delitos sexuales y para establecer relaciones de parentesco biológico. (Butler J. , 2014)

En los casos de delitos de índole sexual entre personas de distinto sexo representa una gran alternativa al análisis de STR autosómicos, ya que en algunos casos en las muestras se encuentran mezclas del DNA de las víctimas y del o los agresores, al analizar Y-STR se asegura que sólo se estudie la aportación de la muestra del hombre. (Butler J. , 2010)

Por el contrario cuando se busca hacer identificación de personas desaparecidas o víctimas de desastres masivos, el análisis de estos marcadores no es recomendable, ya que la presencia de personas relacionadas biológicamente y que provienen del mismo linaje paterno y, por tanto, tienen el mismo cromosoma Y amplía el número de posibles coincidencias.

DNA mitocondrial

El análisis de DNA mitocondrial en el contexto forense se realiza principalmente cuando el indicio biológico se encuentra deteriorado, degradado, en pequeñas cantidades o dañado

intencionalmente por agentes físicos o químicos. (Vázquez Corona & Guardado Estrada, 2021)

La identificación de personas analizando el polimorfismo mitocondrial se realiza comparando los haplotipos mitocondriales obtenidos de la muestra de referencia con los de la muestra problema. Debido a que el DNA mitocondrial se hereda por vía materna, se deben seleccionar como muestras de referencia, aquellos familiares que compartan el mismo linaje. Esta comparación puede dar un resultado excluyente cuando ambos haplotipos son distintos o concluir que la persona quien proporcionó la muestra de referencia comparte el mismo linaje mitocondrial con la persona a la que pertenece la muestra problema. (Vázquez Corona & Guardado Estrada, 2021)

STRs autosómicos

Los microsátélites, denominados *short tandem repeats* están repartidos por todo el DNA nuclear, se ha reportado que pueden encontrarse aproximadamente, cada 10,000 nucleótidos representando aproximadamente el 3 % de dicho genoma.

Los genomas eucariotas contienen secuencias de DNA repetidas, cuyo tamaño puede variar y se designan típicamente por la longitud de la unidad de repetición central y el número de unidades de repetición contiguas o la longitud total de la región repetida. (Butler J. , 2010)

Los microsátélites consisten de unidades de repetición de entre 2 y 7 pb y el número de repeticiones en los marcadores STR puede ser muy variable entre los individuos, lo que hace que estos sean útiles para la identificación humana. (Butler J. , 2010) y (Lareu, 2013)

Nomenclatura

Para facilitar la reproducibilidad y la comparación de datos entre laboratorios, se desarrolló una nomenclatura común en la comunidad de genética forense. (Butler J. , 2010)

Si el STR se encuentra en un gen o forma parte de él, entonces se utiliza el nombre del gen para designar el marcador. Por ejemplo, en el caso del STR **TH01** el gen en el que se encuentra es de la tirosina hidroxilasa, situado en el cromosoma 11 y la parte **01** proviene del hecho de que la región de repetición se encuentra en el intrón 1. (Butler J. , 2005)

Mientras que los marcadores de DNA que se encuentran en regiones no codificantes se designan por su posición cromosómica: la **D** significa DNA, el siguiente carácter es el

número del cromosoma en el que se encuentra, la **S** se refiere a que el marcador es una secuencia de una sola copia y el número final indica el orden en el que el marcador fue descubierto y categorizado para un cromosoma en concreto. (Butler J. , 2005) Por ejemplo, en el STR **D8S1179**:

- **D**: DNA
- **8**: cromosoma 8
- **S**: secuencia de una sola copia
- **1179**: 1179° *locus* descrito en el cromosoma 16

Análisis general de muestras biológicas en genética forense

En la Figura 1 se muestra un diagrama del proceso que regularmente se sigue para el análisis de una muestra biológica. Inicia con la recolección de una muestra que puede ser dubitada o de referencia. Esta se recolecta, traslada y es almacenada hasta su análisis o disposición final. Dependiendo de la procedencia de la muestra de ser necesario se iniciará con la caracterización de esta, para definir la naturaleza del indicio biológico, así como el estado de conservación en el que se encuentra. (Villavicencio Queijeiro, 2022) y (Butler J. , 2010) De acuerdo con el tipo de muestra se debe definir el método que se usará para extracción del DNA, el cual en los casos de las muestras dubitadas es cuantificado , amplificado y separado; mientras que en el caso de las muestras de referencia puede amplificarse directamente sin pasar por un proceso de extracción y cuantificación.

Finalmente, se realiza la interpretación de los resultados y se calculan diversos parámetros estadísticos. Los marcadores que pueden analizarse son de tipo STRs, DNA mitocondrial y/o cromosomas sexuales X y Y. El uso de cada uno de ellos depende del tipo de análisis que desea realizarse, de manera general se utilizan los STRs para las tres aplicaciones que se mencionaron anteriormente. En casos de delitos sexuales que ocurren entre personas de distinto sexo se pueden utilizar los de cromosoma Y para diferenciar el perfil de la víctima del atacante y el ADN mitocondrial se utiliza en casos donde las muestras están muy degradadas o solo se tienen familiares en línea materna para comparar. (Villavicencio Queijeiro, 2022)

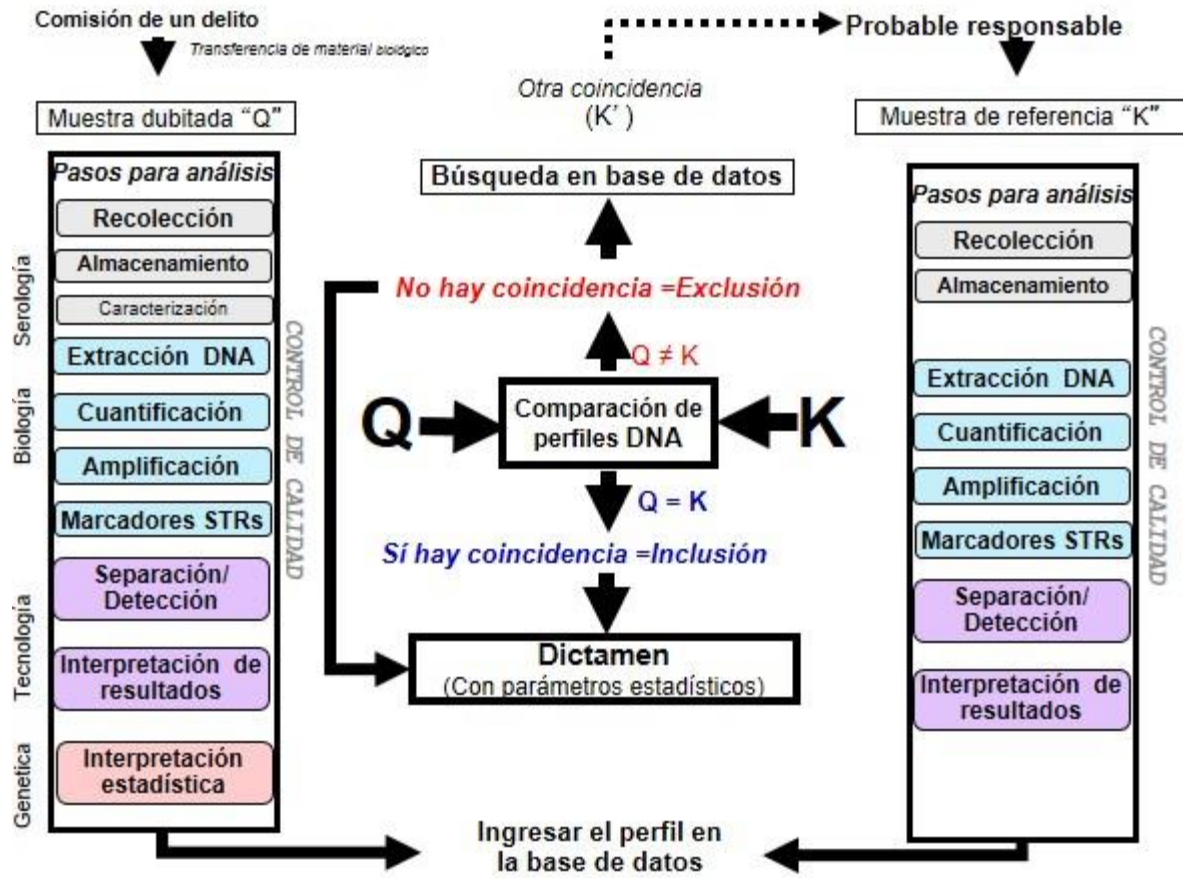


Figura 1. Análisis general que se lleva a cabo en muestras biológicas de interés genético. Modificada de Butler J. , 2010

Técnicas de biología molecular involucradas en el análisis de muestras biológicas en genética forense

PCR

La reacción en cadena de la polimerasa (PCR) es un proceso enzimático que se realiza *in vitro* en el cual una región específica del DNA es amplificada de manera exponencial para obtener muchas copias de una secuencia en particular. (Butler J. M., 2011)

En la PCR se simula el proceso de replicación celular dentro de un tubo, en el cual debe haber una mezcla de varios componentes individuales (Figura 2). Comienza colocando en un tubo la solución amortiguadora que mantiene el pH adecuado para el funcionamiento del DNA polimerasa, por lo general incluye Tris-HCl (pH 8.3) y KCl; una enzima DNA polimerasa termoestable (por lo general se utiliza la enzima *Taq*, llamada así porque se aisló inicialmente del organismo termófilo *Thermus aquaticus*), un cofactor de la enzima, comúnmente $MgCl_2$; los oligonucleótidos, que flanquean la zona de DNA a amplificar y son complementarios a

ella; los desoxinucleótidos (dNTPs) que serán las unidades con las que se formen las nuevas cadenas de DNA; el DNA molde o templado que contiene la región de DNA que se va a amplificar y por último agua libre de nucleasas usada como solvente (Bolivar, Rojas, & García-Lugo, 2014)

Componentes necesarios para la amplificación de PCR	
Reactivo	Concentración óptima
Tris-HCl, pH 8.3 (25 °C)	10-50 mM
MgCl ₂	1.2-2.5 mM
KCl	50 mM
Desoxirribonucleótidos trifosfatos (dNTPs)	200 mM cada dATP, dTTP, dCTP, dGTP
DNA polimerasa termoestable	0.5-5 U
Albúmina de suero bovino (BSA)	100 µg/mL
Oligonucleótidos	0.1-1.0 µM
DNA templado	1-10 ng de DNA genómico

Figura 2. Componentes en la amplificación de PCR. Dichos componentes, excepto el DNA templado, se encuentran en los kits comerciales divididos en *master mix*, oligonucleótidos y agua libre de nucleasas.

La PCR se divide en 3 etapas (Figura 3):

Desnaturalización: el primer ciclo de desnaturalización sirve para separar la doble hélice al romper los puentes de hidrógeno entre las bases nitrogenadas. La temperatura y el tiempo de los ciclos se determinan de acuerdo con las características del DNA, cómo lo son el tipo de DNA a amplificar y su composición y de la DNA polimerasa utilizada. (Sierra Palacios, Serrato Díaz, Flores Rentería, & Aportela Cortez, 2014). Por lo general las condiciones oscilan entre 90 °C y 95 °C durante 1 a 5 minutos, esto dependerá de la vida media de la enzima que se esté utilizando. (Bolivar, Rojas, & García-Lugo, 2014)

Alineamiento: se disminuye la temperatura de incubación para que los oligonucleótidos se unan al DNA molde o templado en las zonas 3' de manera complementaria. Ambos oligonucleótidos, llamados forward y reverse, se unen de manera específica y complementaria al DNA flanqueando el fragmento que se quiere amplificar (Sierra Palacios, Serrato Díaz, Flores Rentería, & Aportela Cortez, 2014). La temperatura de alineamiento oscila entre 40 °C y 70 °C durante un tiempo de entre 30 segundos a 2 minutos. Tanto la temperatura como el tiempo de alineamiento dependerán de las características de las secuencias de los oligonucleótidos utilizados, específicamente de la cantidad de nucleótido presente, una forma de calcular la temperatura es la siguiente: $T_m = 4(G+C) + 2(A+T)$ (Bolivar, Rojas, & García-Lugo, 2014)

Extensión: también conocida como elongación, amplificación o polimerización, consiste en la síntesis de la nueva cadena de DNA, a partir del extremo 3' que se encuentra libre del oligonucleótido por acción de la DNA polimerasa, empleando como sustrato los cuatro dNTPs. La etapa de extensión se realiza a 72 °C porque es la temperatura a la cual la *Taq* polimerasa alcanza su mayor actividad. (Sierra Palacios, Serrato Díaz, Flores Rentería, & Aportela Cortez, 2014)

Al terminar los ciclos repetidos se realiza la extensión final, para permitir que la polimerasa termine de sintetizar todos los fragmentos que pueden haber quedado incompletos y se realiza una última extensión de aproximadamente 5 minutos a 72 °C. (Sierra Palacios, Serrato Díaz, Flores Rentería, & Aportela Cortez, 2014)

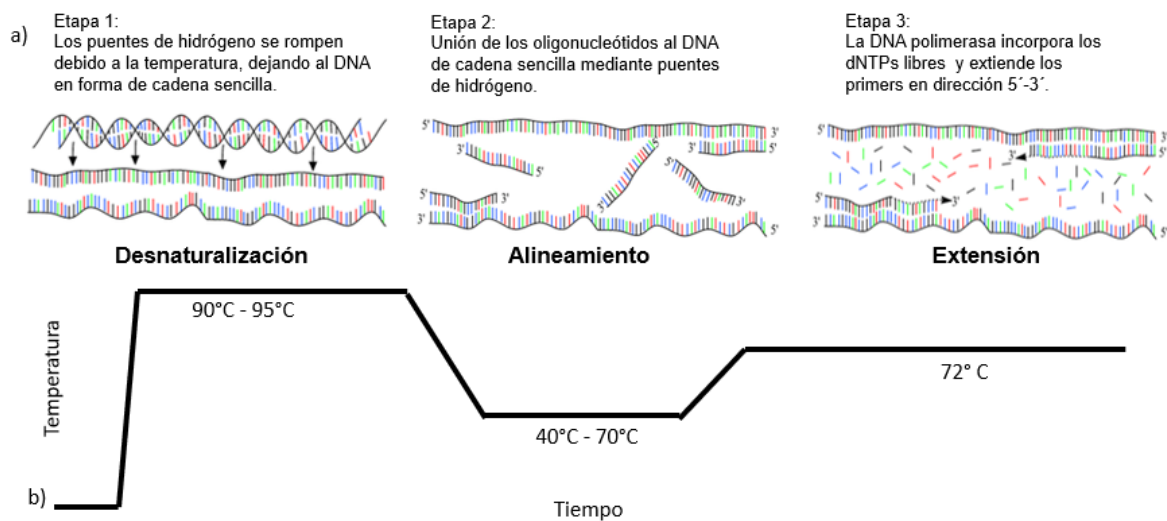


Figura 3. Etapas de un ciclo de amplificación de PCR. Modificada de León, 2017 y Butler J. , 2010. En la sección a) se muestra el proceso que ocurre a nivel molecular, mientras que en el b) se muestra un ejemplo de la variación de las temperaturas a las cuales se da el proceso.

PCR en tiempo real

Fue descrita por primera vez por Higuchi y colaboradores del corporativo Cetus a inicios de la década de 1990, el análisis ciclo a ciclo incrementa la señal de fluorescencia resultante de la amplificación del objetivo secuenciado durante la PCR. (Butler J. M., 2011)

Es una variante de la PCR en la cual la acumulación del DNA amplificado es detectada y cuantificada a medida que la reacción avanza, es decir: “en tiempo real” esto se logra incorporando fluoróforos que se asocian al DNA amplificado, donde el incremento de la

fluorescencia es proporcional al incremento de la cantidad de moléculas de DNA amplificadas en la reacción. (De la Torre Pretell, Roe Battistini, & Roe Battistini, 2016)

PCR multiplex

En el caso de una PCR multiplex, se amplifican simultáneamente en un único tubo distintas regiones específicas, lo cual necesariamente implica que los reactivos mezclados y el programa utilizado sean suficientes y adecuados para permitir la detección de cada fragmento y no inhibir la de las demás. Algunos parámetros, como la concentración de magnesio y de cebadores, el tipo y la cantidad de DNA polimerasa, pueden ajustarse experimentalmente. Para otros, en cambio, debe realizarse un diseño exhaustivo previo. En el caso de los oligonucleótidos y el programa de temperaturas utilizado, es necesario escoger o diseñar oligonucleótidos que no interaccionen entre sí; que tengan temperaturas de alineamiento similares; que cada pareja amplifique una única secuencia diana, y que generen picos de tamaño suficientemente diferente como para poder ser separados y diferenciados tras la amplificación. Actualmente se utilizan kits comerciales para la amplificación de STRs autosómicos, de cromosoma X y de cromosoma Y. (Méndez-Álvarez & Pérez-Roth, 2004) Este tipo de PCR es la más utilizada en genética forense ya que los sistemas de identificación utilizados requieren la amplificación de varios *loci* a la vez.

Electroforesis capilar

La electroforesis es una técnica que se usa para separar moléculas en función de su tamaño y carga eléctrica. Se utiliza una corriente eléctrica para mover las moléculas a través de un medio o matriz que puede ser un gel o un capilar. Los poros de la matriz actúan como un tamiz, lo cual permite que las moléculas más pequeñas se muevan más rápido que las moléculas más grandes. Para determinar el tamaño de las moléculas de una muestra se usan estándares de tamaños conocidos que se separan en la misma matriz y luego se comparan con la muestra. (National Human Genome Research Institute, 2022)

En los laboratorios de genética forense, la electroforesis capilar es una técnica de separación que se utiliza para detectar los alelos de los STR amplificados durante la PCR. (Butler J. M., 2011). Dentro del equipo de electroforesis capilar (Figura 4) se encuentran un cátodo y un ánodo con carga eléctrica, la matriz consistente de un polímero (generalmente dimetil

poliacrilamida) que se encuentra dentro de un capilar; amortiguadores que ayudan a las moléculas de DNA a transportarse a través del polímero; un láser de argón que incide en las moléculas de DNA excitando los fluorocromos que contiene la muestra; la ventana de detección que capta la fluorescencia generada y una cámara CCD que es un dispositivo detector de luz. (Butler J. M., 2011)

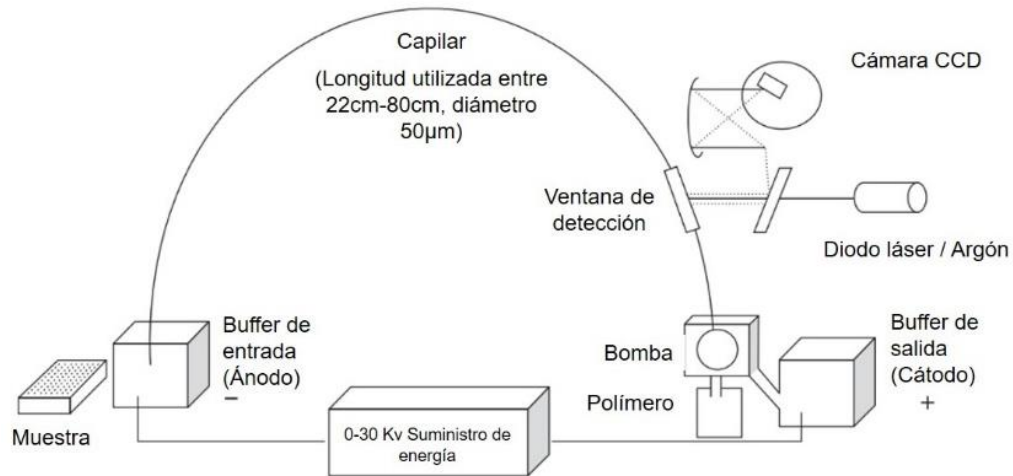


Figura 4. Esquema de un equipo de electroforesis capilar. Modificado de *Butler J. , 2010*

Es necesario diluir los productos de PCR en formamida, la cual: desnaturaliza el DNA y favorecer la migración de las moléculas, impide la dimerización al evitar la unión de cadenas de DNA de diferentes tamaños y estabiliza la migración del DNA. Lo anterior, ayuda al proceso de inyección electrocinética que es el método utilizado para introducir las muestras de DNA en el capilar, la cual consiste en aplicar un voltaje al líquido de la muestra en dónde está inmerso el capilar por un tiempo definido. (Butler J. M., 2011).

En la Figura 5 se muestra un esquema de la separación y detección de fragmentos STR una vez realizada la inyección. Primero la muestra comenzará a migrar por el capilar en dirección al cátodo debido a su carga negativa, en el capilar, que contiene polímero, comenzará la separación de la muestra dependiendo del tamaño de los fragmentos, dichos fragmentos serán incididos por un láser de Argón que excitará los fluorocromos añadidos en la reacción de PCR y ésta será captada por el espectrógrafo del equipo y detectado por el panel CCD con filtros virtuales que es un detector fluorescente fotosensible el cual registra la intensidad de la luz emitida por los fluoróforos y finalmente es interpretado por un software.

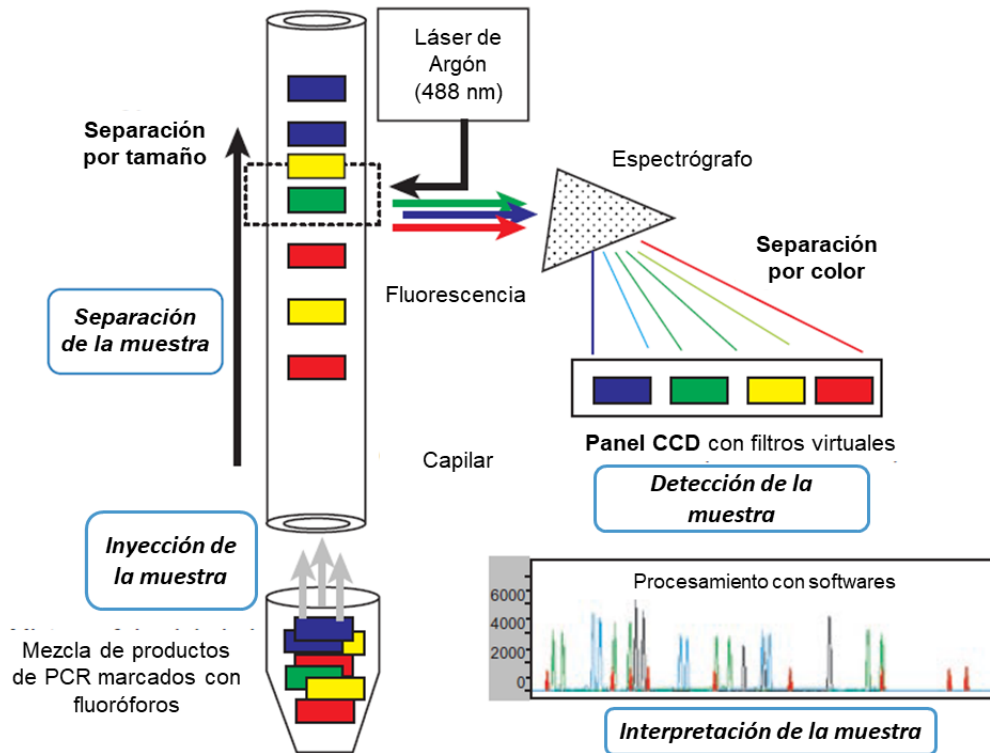


Figura 5. Esquema de la separación y detección de fragmentos STRs en electroforesis capilar. Modificado de Butler M., 2011.

Kits comerciales para la amplificación de STRs

En un esfuerzo por agilizar la amplificación de fragmentos de DNA se han producido una gran variedad kits comerciales que contienen los oligonucleótidos necesarios para poder amplificar varios marcadores simultáneamente.

Los kits contienen 5 componentes: 1) la mezcla de oligonucleótidos marcados con un conjunto de colorantes fluorescentes llamados fluorocromos, que contiene los oligonucleótidos diseñados para la amplificación de un conjunto de *loci* de STRs; 2) el buffer para PCR que incluye los dNTPs, el MgCl₂ que se utiliza como cofactor de la DNA polimerasa (*Taq* polimerasa); 3) el *allelic ladder* o escalera alélica con los alelos observados para cada *loci* que se utiliza para etiquetar el nombre de cada alelo; 4) estándar interno o *internal size standard* el cual se utiliza para la calibración del tamaño de la repetición del alelo y 5) una muestra de DNA control para verificar que los reactivos del kit funcionan correctamente. (Butler J. M., 2011)

Los kits varían entre sí dependiendo de los *loci* que se incluyen, las combinaciones de marcajes fluorescentes que se utilizan, la cadena de DNA que es marcada, el *allelic ladder* presente y lo más importante las secuencias de los oligonucleótidos utilizados para la amplificación por PCR (Butler J. M., 2011).

Investigator® 24plex Qiagen

Para el presente trabajo se utilizó el kit Investigator® 24plex de la marca Qiagen. Dicho kit se utiliza en el análisis forense, para identificación humana y en pruebas de paternidad. En la PCR multiplex se amplifican de manera simultánea el marcador amelogenina (que permite determinar el sexo); 22 marcadores STR polimórficos, de los cuales 21 son autosómicos y 1 del cromosoma Y. Estos 22 STR son los que recomiendan el *Core Loci Working Group* del CODIS, la *European Network of Forensic Science Institutes* y el *European DNA Profiling Group*. (Qiagen, 2021)

El kit también contiene dos controles de PCR internos o sensores de calidad marcados como QS1 y QS2 con los cuales se pueden verificar la eficiencia de los componentes y la presencia de inhibidores en la PCR. (Qiagen, 2021)

Los controles internos antes mencionados, se incluyen en la mezcla de cebadores y se amplifican de forma simultánea con los marcadores STR polimórficos. Estos controles son secuencias de DNA de control interno sintéticas diseñadas con un algoritmo aleatorio, dichas secuencias difieren de todas las secuencias de DNA humano conocidas. Por lo tanto, la posibilidad de una unión no específica en el contexto de una reacción de amplificación mediante PCR multiplex es muy baja. (Qiagen, 2021)

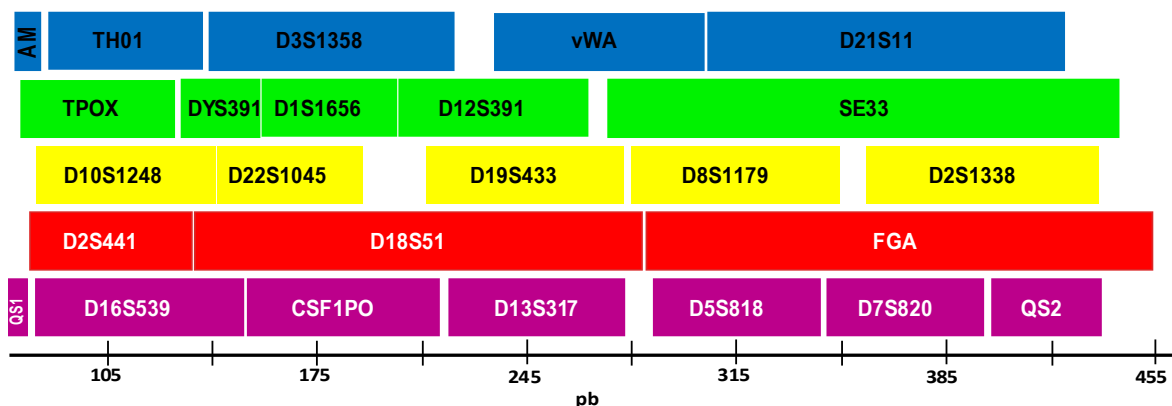


Figura 6: Marcadores STRs del kit Investigator® 24plex Qiagen. Se muestran los *loci* que se amplifican con el kit antes mencionado, en la parte inferior se encuentra la escala de tamaño en pb. Se muestran ordenados tanto por su tamaño como por el color de los fluorocromos utilizados, 6-FAM™ (azul), BTG (verde), BTY (amarillo), BTR2 (rojo) y BTP (violeta).

Utilidad del análisis poblacional en el estudio de STRs autosómicos con fines forenses

La genética de poblaciones es el estudio de la variación heredada y su modulación en tiempo y espacio en una población dada. Esta rama de la genética cuantifica la variación observada dentro de una población o entre diferentes poblaciones, en términos de frecuencias alélicas y genotípicas. (Butler J. , 2005) Igualmente estudia los factores que afectan dichas frecuencias en los distintos *loci* en una población. (Goodwin, Linacre, & Hadi, 2011)

En genética forense se buscan regiones de DNA cuya variabilidad permita encontrar diferencias entre individuos no relacionados biológicamente y que a la vez tenga índices de mutaciones que dificulte ver mutaciones en generaciones contiguas de personas relacionadas biológicamente. (Butler J. , 2010)

Frecuencias alélicas

La frecuencia de un alelo se puede calcular dividiendo el número de copias de un alelo entre el número total de alelos presentes en esa población. (Butler J. , 2005). La fórmula utilizada para calcular las frecuencias alélicas de marcadores diploides como lo son los STRs autosómicos es la siguiente:

$$f(a) = \frac{\text{número de alelos observados}}{2n}$$

Donde n es el tamaño de muestra analizada.

Frecuencias genotípicas

Las frecuencias genotípicas se obtienen dividiendo cada número de combinaciones observadas (genotipo) entre el número total de las muestras analizadas. (Hernández-Rodríguez & Trejo-Medinilla, 2014)

$$f(g) = \frac{\text{número de genotipos observados}}{n}$$

Donde n es el tamaño de muestra analizada.

Equilibrio de Hardy-Weinberg

La ley de Hardy-Weinberg proporciona una representación matemática sencilla de la relación de las frecuencias genotípicas y alélicas en una población ideal y es fundamental para la genética forense. (Goodwin, Linacre, & Hadi, 2011)

El equilibrio de Hardy-Weinberg es una relación matemática la cual establece que, en una población que se reproduce al azar, las frecuencias genotípicas de cualquier *locus* genético permanecen constantes. Cuando una población cumple la ley de Hardy-Weinberg, se dice que está en equilibrio Hardy-Weinberg (HW). Es importante destacar que cuando una población está en HW, las frecuencias genotípicas pueden predecirse a partir de las frecuencias alélicas, (Goodwin, Linacre, & Hadi, 2011). Es decir que, si dos alelos **P** y **Q** ocurren con frecuencias **p** y **q** en la población, entonces el genotipo **PP** (homocigoto) tendrá una frecuencia **p²** y el genotipo **PQ** (heterocigoto) debería aparecer con una frecuencia **2pq**. (Butler J. , 2014)

Para considerar que una población está en equilibrio de Hardy-Weinberg es necesario que la población: sea infinita, se aparee de manera aleatoria, debe estar libre de los efectos de la migración, que no exista la selección natural y que no ocurran mutaciones.

En la práctica las poblaciones humanas no cumplen con todos estos supuestos, sin embargo, es posible realizar una comparación entre los genotipos heterocigotos observados y los esperados, si la población se encontrase en equilibrio. Esto se lleva a cabo mediante la prueba de χ^2 de comparación de hipótesis (Sokal & Rohif, 1979) :

$$x^2 = \sum \frac{(\text{Heterocigotos observados} - \text{heterocigotos esperados})^2}{\text{heterocigotos esperados}}$$

Dando como resultado el valor de un estadístico de prueba, que en caso de ser superior a 0.05, podrá considerarse que las frecuencias observadas se ajustan a las esperadas según el HW es decir que los alelos de los distintos *loci* son independientes entre sí. (Sokal & Rohif, 1979)

Homocigosidad esperada

Es la probabilidad de que un individuo sea homocigoto en un *locus* dado, es decir la suma de los cuadrados de las frecuencias alélicas, y está representada por la siguiente ecuación:

$$He = \left[\sum_{j=1}^n \left(\sum_{i=1}^m p_{ij}^2 \right) \right] / n$$

Dónde *n* es el número de *loci* considerados, *m* el número de alelos en el *locus j* y *p_{ij}* es la frecuencia del alelo *i* en el *locus j*.

Heterocigosidad observada

Es la proporción de individuos heterocigotos en la población. Esto se calcula dividiendo el número de muestras con genotipos heterocigotos entre el número total de muestras. Una alta heterocigosidad significa una mayor diversidad alélica por lo cual hay menos posibilidades de que una muestra aleatoria coincida. (Gouy & Zieger, 2017)

Heterocigosidad esperada

Es la probabilidad de que un individuo sea heterocigoto en un *locus* dado, depende directamente de la diversidad genética en dicho *locus*, que a su vez depende de las frecuencias alélicas de su población y se obtiene utilizando la siguiente fórmula (Gouy & Zieger, 2017):

$$He = \left[\sum_{j=1}^n \left(1 - \sum_{i=1}^m p_{ij}^2 \right) \right] / n$$

Dónde n es el número de *loci* considerados, m el número de alelos en el *locus* j y p_{ij} es la frecuencia del alelo i en el *locus* j .

Probabilidad de coincidencia al azar

La probabilidad de coincidencia al azar en genética forense tiene dos acepciones*, la primera de ellas es la probabilidad de que una muestra al azar de la población tenga el mismo perfil de DNA que una muestra de referencia. Este tipo de probabilidad se calcula utilizando las frecuencias alélicas referidas en bases de datos, dependiendo del genotipo de cada *loci* (Cuadro 1) y multiplicándolas entre sí, a esta multiplicación se le denomina regla del producto y es posible ya que los alelos en cada *locus* se heredan de manera independiente (Goodwin, Linacre, & Hadi, 2011). El resultado es correspondiente a la probabilidad de que esa combinación genética se presente producto del azar (Rangel Villalobos , 2017).

Si el genotipo es	
Homocigoto	Heterocigoto
p^2	$2pq$

Cuadro 1. Fórmulas utilizadas para el cálculo de la probabilidad de coincidencia al azar dependiendo del genotipo; dónde p y q son las frecuencias alélicas referidas en la base de datos que se esté utilizando

La segunda acepción es cuando se utiliza como parámetro de interés forense de una base de datos genética y se define como la probabilidad de que existan dos genotipos iguales en la población de estudio, calculado con la siguiente fórmula (Gouy & Zieger, 2017):

$$PM = \sum_i (p_i)^2$$

Dónde p_i es la frecuencia del genotipo i en un *locus* determinado de la población.

*Nota: para fines prácticos en el presente estudio se hará la distinción utilizando las abreviaturas PM para cuando se refiera a la probabilidad de coincidencia al azar como parámetro de interés forense y PCA cuando se refiera a la probabilidad de coincidencia al azar de dos perfiles idénticos en una población.

Poder de discriminación

Es la probabilidad de que dos individuos no relacionados puedan ser diferenciados genéticamente mediante el análisis de uno o varios marcadores. Depende del número de *loci* analizados, de las variables alélicas encontradas y de las frecuencias de cada *locus* en la población. (Gouy & Zieger, 2017). Se calcula utilizando la fórmula siguiente, en la que PD representa el poder de discriminación y PM la probabilidad de coincidencia al azar

$$PD = 1 - PM$$

Contenido de información polimórfica

El contenido de información polimórfica se refiere a una medida del grado de información que brinda un determinado marcador genético, la cual depende del número de alelos para ese *locus* y de sus frecuencias (González Chavarro, León Lozano, Morillo Coronado, Ochoa, & Morillo Coronado, 2016) para calcularlo se utiliza la siguiente fórmula:

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

Dónde n es el número de alelos, p_i y p_j son las frecuencias alélicas.

Un locus puede considerarse polimórfico cuando la heterocigosidad esperada (H_o) fue superior a 0.1 y altamente polimórfico si H_o fue superior a 0.7 (González Chavarro, León Lozano, Morillo Coronado, Ochoa, & Morillo Coronado, 2016)

Poder de exclusión

Es la fracción de individuos que tienen un perfil de DNA que es diferente del de un individuo seleccionado al azar en un caso típico de paternidad. El valor para cada caso individual variará. El promedio para un *locus* determinado está representado por la siguiente fórmula (Huston, 1998):

$$PE = h^2(1 - 2hH^2)$$

Donde h es el número de individuos heterocigotos y H el número de individuos homocigotos en la muestra de la población.

Índice típico de paternidad

El índice de paternidad refleja cuántas veces más probable es que la persona analizada sea el padre biológico, en lugar de un individuo seleccionado al azar. El índice de paternidad típico se asigna a un *locus* y no a un caso individual. Por lo general, un ITP inferior a uno es indicativo de falta de parentesco. Este índice está representado por la siguiente ecuación (Huston, 1998):

$$ITP = \frac{1}{2H}$$

Donde H es la proporción de individuos homocigotos en la muestra de la población.

Distancias genéticas

Las distancias genéticas fueron diseñadas para expresar, con un único número, la diferencia entre dos poblaciones. Cuando no hay diferencias entre las poblaciones de estudio, la distancia es 0, mientras que, si estas no tienen alelos en común para ningún locus, la diferencia es 1. (Bruno & Balzarini, 2009)

La distancia de Nei ha sido aplicada a datos genotípicos obtenidos a partir de marcadores STR y sus propiedades han sido evaluadas desde la perspectiva de estudios poblacionales. (Bruno & Balzarini, 2009)

Nei describió varias medidas de distancia genética utilizadas para los datos de frecuencia génica, las cuáles han sido aplicadas a datos genotípicos obtenidos a partir de marcadores STR y sus propiedades ha sido evaluadas desde la perspectiva de estudios poblacionales (Bruno & Balzarini, 2009)

Las distancias genéticas pueden ser divididas en dos grandes tipos la primera refleja únicamente la similitud entre poblaciones con respecto a las frecuencias alélicas presentes en las poblaciones, sin asumir *a priori* ningún supuesto evolutivo (Demarchi, 2009), la distancia de este tipo más utilizada en análisis es la de Nei *et al* (1983) D_a cuya ecuación es: (Takezaki & Nei, 1996)

$$D_a = 1 - \frac{1}{r} \sum_j^r \sum_i^{m_j} \sqrt{x_{ij}y_{ij}}$$

Dónde y_{ij} y x_{ij} son las frecuencias de los alelos i en el locus j en cada población, m_j es el número de alelos en el locus j y r es el número de loci estudiados. (Nei, Tajima, & Tateno, 1983)

Mientras que el segundo tipo de distancia se basa en modelos evolutivos donde la variación es atribuida principalmente a la acción de la mutación y la deriva, asumiendo que los marcadores utilizados no están sujetos a la selección natural (Demarchi, 2009). La distancia de este tipo más utilizada es la de Latter, denominada F_{st} y cuya ecuación es: (Takezaki & Nei, 1996)

$$F_{st} = \frac{(J_x - J_y)/2 - J_{xy}}{1 - J_{xy}}$$

Dónde J es la homocigosidad esperada para las poblaciones x y y .

Como se mencionó anteriormente los valores de las distancias genéticas van de 0 a 1, la interpretación de estos valores puede estar dada como se muestra en el Cuadro 2 :

Valor de la distancia	Diferenciación
De 0 a 0.05	pequeña
De 0.05 a 0.15	moderada
De 0.15 a 0.25	grande
>0.25	muy grande

Cuadro 2. Interpretación de los valores de distancias genéticas.

Pruebas genéticas en el ámbito forense

Identificación humana en contextos criminalísticos

En noviembre de 1997, el FBI seleccionó 13 STRs como un set de marcadores mínimo para generar e intercambiar perfiles genéticos dentro de los Estados Unidos de América (EUA). De forma análoga, en Europa se cuenta con el Set europeo estándar (ESS por sus siglas en inglés *European Standard Set*) de STRs. (Rangel Villalobos , 2017)

En un caso de identificación se necesitan dos perfiles genéticos, el primero de la muestra del indicio que se tiene y el segundo de una muestra de referencia, si son distintos entonces se puede hablar de una exclusión, pero si los perfiles son iguales entonces se genera una inclusión (Thompson & Black, 2007) y se debe realizar el cálculo de probabilidad de coincidencia al azar y la razón de verosimilitud (LR por sus siglas en inglés *likelihood ratio*) (Butler J. , 2014).

Razón de verosimilitud (LR):

Es la comparación de probabilidades de la evidencia bajo dos propuestas alternativas y mutuamente excluyentes: la del fiscal (H_f) contra la de la defensa (H_d). La persona experta forense calcula el peso de la evidencia sin transgredir aquellas áreas reservadas para la jueza o juzgador.

La hipótesis del fiscal representa que el perfil genético del indicio tenga como origen biológico al imputado; por lo tanto, la hipótesis de la fiscalía asume el 100% de la probabilidad ($H_f=1$). En contraparte, la hipótesis de la defensa asume que el perfil genético del indicio coincide con el imputado, pero pertenece a otro individuo de la población de estudio por lo que la probabilidad es calculada a partir de la frecuencia del genotipo del perfil genético particular. Si el genotipo resultante es heterocigoto, entonces esta probabilidad sería $2pq$, donde p es la frecuencia del alelo 1 y q es la frecuencia del alelo 2 dentro de la población para cada uno de los *locus*, pero si genotipo resultante es homocigoto la probabilidad sería p^2 (Goodwin, Linacre, & Hadi, 2011). Tomando en cuenta lo anterior se utiliza la ecuación:

$$LR = \frac{H_f}{H_d} = \frac{1}{PCA}$$

Identificación humana en contextos de desaparición

México atraviesa una crisis forense, especialmente en casos de desaparición, por lo cuál las familias de las personas desaparecidas han tenido que intervenir en la búsqueda. Uno de los principales retos para afrontar esta crisis, radica en la identificación de personas desaparecidas, la cual se realiza mediante pruebas de parentesco biológico, ya que en muchos casos no se cuenta con una muestra de referencia de la persona desaparecida para su análisis.

Relaciones de parentesco

Como se ha mencionado anteriormente, los STR autosómicos están formados por dos alelos uno heredado vía materna y otro vía paterna, gracias a esto es posible realizar pruebas de relación de parentesco ya que cada progenitor debería tener un alelo en común con cada una de sus hijas o hijos, excepto en caso de mutaciones. (Butler J. , 2014)

Los hermanos cuyos ambos progenitores sean los mismos deben compartir un alelo idéntico aproximadamente el 50% de las veces, dos alelos el 25% y ningún alelo el 25% de las veces restante. (Butler J. , 2014) En la Figura 7 se muestra la forma de calcular el índice de paternidad en un caso típico, es decir, cuando se tiene de referencia el genotipo de la madre biológica, en caso de que no se tenga esa referencia, entonces es necesario aplicar las fórmulas que se muestran en la Figura 8.

Genotipo de la madre	Genotipo del hijx	Genotipo del supuesto padre	PI (El supuesto padre es el verdadero padre)
aa	aa	aa	$\frac{1}{P_a}$
ab			
bb			
bc	bb		
aa	aa	ab	$\frac{1}{2P_a}$
ab			
bb			
bc	ab		
bc			
cc			
cd	ac		
ab	ab	aa	$\frac{1}{P_a + P_b}$
		ab	$\frac{1}{2(P_a + P_b)}$
		ac	

Figura 7. Cálculo del índice de paternidad (PI) para todas las combinaciones no excluidas cuando se tienen los genotipos maternos y paternos. Modificada de Buckleton, Taylor, & Bright, 2016

Genotipo del hijx	Genotipo del supuesto padre	PI (El supuesto padre es el verdadero padre)
aa	aa	$\frac{1}{P_a}$
aa	ab	$\frac{1}{2P_a}$
ab	aa	
aa	aa	$\frac{P_a + P_b}{4P_a P_b}$
ab	ac	$\frac{1}{4P_a}$

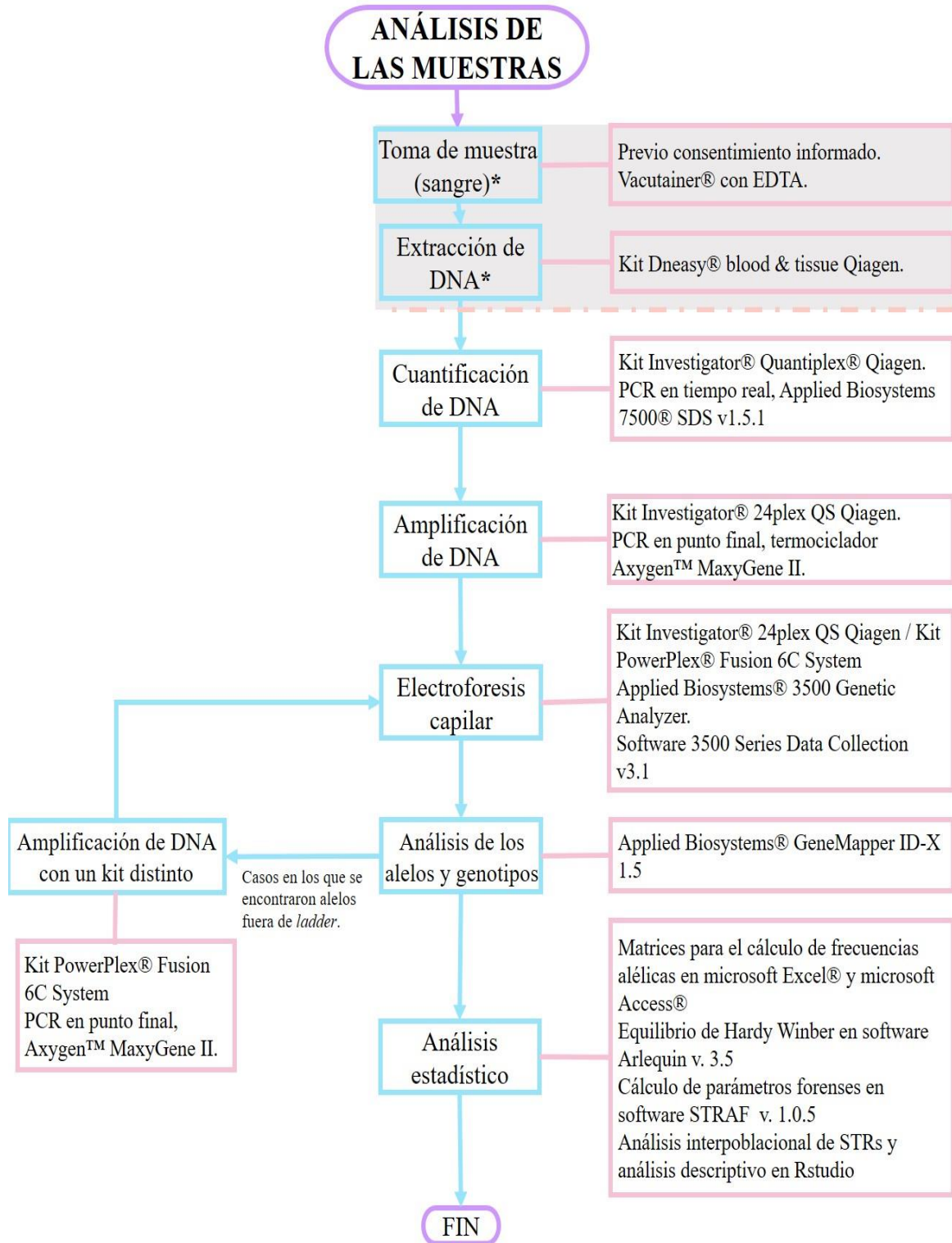
Figura 8. Cálculo del PI para todas las combinaciones no excluidas cuando se tiene el genotipo paterno.
Modificada de Buckleton, Taylor, & Bright, 2016

Desde el punto de vista de la genética, los medios hermanos, las relaciones tío-sobrino y abuelos-nietos son equivalentes y son etiquetadas como parientes de segundo grado. Con cada grado de separación entre dos parientes analizados, hay una mayor probabilidad de segregación alélica o de mutación que dificulta la capacidad de confirmar una relación de parentesco biológica específica. (Buckleton, Taylor, & Bright, 2016)

METODOLOGÍA

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 y el cual fue abalado por la división de investigación y comisión de ética de la Facultad de Medicina bajo el número de registro 65/2022. (Páginas 70 a 78, Anexo 1)

En el Diagrama 1 se muestra la metodología general utilizada para el análisis de las muestras, la cual fue realizada en las instalaciones del laboratorio de genética del Instituto de Servicios Periciales y Ciencias Forenses (INCIFO) de la CDMX.



* Proceso realizado por personal del Laboratorio de Genética de la Licenciatura en Ciencia Forense en el 2019.

Diagrama 1. Metodología general para el análisis de las muestras LCF001-LCF57.

La recolección de las muestras se realizó con el consentimiento informado del proyecto mencionado anteriormente, el cual cuenta con un apartado sobre el almacenamiento y uso futuro de las muestras (Anexo 1, páginas 75 a 78), en el cual se autoriza que las muestras de sangre sean almacenadas y se usen en investigaciones futuras por el laboratorio de genética de la LCF de la UNAM.

Las muestras fueron tomadas de personas mestizo-mexicanas de la Ciudad de México, entendiendo como persona mestiza-mexicana aquella que nació en México, tiene un apellido de origen español y ancestros de origen mexicano tres generaciones hacia atrás según lo reportado por Aguilar Velazquez y Rangel-Villalobos en 2021. Para asegurarse de que las y los participantes cumplieran con los criterios antes mencionados se les preguntó el lugar dónde vivían y el lugar de nacimiento de su madre, padre, abuelos y bisabuelos.

Personal del laboratorio de genética de la LCF realizó la toma de muestras de sangre periférica utilizando tubos vacutainer con EDTA, las muestras fueron identificadas como LCF001-LCF157 y dicha etiqueta se mantuvo para el presente estudio.

La extracción de DNA de las muestras fue realizada por el Dr. Christian Cárdenas y la Dra. Alexa Villavicencio días después de la recolección a partir de sangre total, se realizó una lisis de glóbulos rojos utilizando buffer RBCLA y posteriormente se utilizó el kit DNeasy[®] blood & tissue Qiagen, siguiendo el protocolo del fabricante (Qiagen, 2016) previo al inicio de la presente investigación en el laboratorio de Genética en el área de investigación de la LCF en el año 2019. Las muestras de DNA se resguardaron a -20 °C y se mantuvieron bajo esas condiciones hasta su procesamiento.

Cuantificación de DNA

El volumen de las muestras proporcionadas por el laboratorio de genética de la LCF fue de 5µL por lo que previo a la cuantificación se realizó una dilución con 15 µL de agua libre de nucleasas, para asegurar volúmenes suficientes para el posterior análisis.

Para realizar esta cuantificación se utilizó el kit Investigator[®] Quantiplex[®] (Qiagen, 2018) con el sistema de PCR en tiempo Real Applied Biosystems 7500 y el software SDS v 1.5.1.

Las mezclas de reacción se escalaron a $1/4$, siguiendo el protocolo previamente estandarizado por el laboratorio de genética del INCIFO de la Ciudad de México

Se realizó una curva estándar con el protocolo previamente mencionado (Cuadro 3 y Cuadro 4). Los volúmenes finales utilizados para la cuantificación de las muestras fueron los siguientes:

- Mezcla maestra o *master mix*: 2.5 μL
- Oligonucleótidos: 2.5 μL
- Agua libre de nucleasas: 5.0 μL
- Muestra/DNA: 2 μL

Estándar	Concentración de las diluciones seriales del DNA control (ng/ μL)	DNA (μL)	Buffer (μL)
1	20	No diluido	-
2	5	10	30
3	1.25	10	30
4	0.3125	10	30
5	0.078125	10	30
6	0.01953125	10	30
7	0.0048828125	10	30

Cuadro 3. Diluciones seriadas del DNA control para realizar la curva estándar.

Paso	Temperatura	Tiempo	Número de ciclos
Paso de activación inicial de PCR	95 °C	3 min	-
Paso 2 ciclado			40
Desnaturalización	95 °C	5 s	
Alineamiento y Extensión	60 °C	35 s	

Cuadro 4. Condiciones de ciclado para el sistema de PCR en tiempo real Applied Biosystems 7500.

Ensayos para el protocolo de amplificación

Se realizó la amplificación del DNA control 9948 proporcionado por el fabricante (Electroferograma 1, página 37) y de un control negativo (Electroferograma 2, página 38) para corroborar que todos los reactivos funcionaran de manera adecuada.

Posteriormente se seleccionó de manera aleatoria la muestra LCF 001 para realizar ensayos y establecer el protocolo a seguir para la amplificación utilizando PCR multiplex, la cual se llevó a cabo con el kit Investigator[®] 24plex QS de la marca Qiagen en un termociclador Axygen[™] MaxyGene II.

Para fines prácticos se realizaron pruebas escalando la reacción descrita en el manual del kit a $1/5$, $1/4$ y $1/3$ (Cuadro 5) y con las condiciones de termociclado de acuerdo con el fabricante (Cuadro 6).

Reactivo	1x	¹ / ₅ x	¹ / ₄ x	¹ / ₃ x
Fast	7.5 µL	1.3 µL	1.875 µL	2.5 µL
Oligonucleótidos	2.5 µL	0.5 µL	0.625 µL	0.833 µL
Agua libre de nucleasas	Variable	Variable	Variable	Variable
DNA templado	Variable	Variable	Variable	Variable
Volumen final	25 µL	5 µL	6.25 µL	8.33 µL

Cuadro 5. Reactivos utilizados para la reacción de PCR, escalados a ¹/₅, ¹/₄ y ¹/₃ de reacción.

Temperatura	Tiempo	Número de ciclos
98 °C	30 s	
64 °C	55 s	3
72 °C	5 s	
96 °C	10 s	
61 °C	55 s	24
72 °C	5 s	
68 °C	5 min	
60 °C	5 min	
10 °C	∞	

Cuadro 6. Condiciones de termociclado establecidas en el protocolo del fabricante.

Para descartar potenciales fenómenos de contaminación cruzada por parte de la operadora sobre las muestras, se realizó el procesamiento de una muestra de saliva propia (Electroferograma 6, página 42), con el fin de que, en caso de existir alguna muestra contaminada o mezcla, fuera más sencillo descartar en qué etapa del procesamiento se dio la contaminación y de ser el caso si es viable su repetición.

Amplificación de los marcadores STRs

Con las condiciones de PCR ya estandarizadas (Cuadro 5 y Cuadro 6) se amplificaron las muestras restantes. Para asegurar la dilución de las sales presentes en las muestras que actuaban como inhibidores de la PCR y disminuir el error de pipeteo se decidió tomar un mínimo de 1 µL de cada muestra por lo cual se ajustaron los volúmenes de agua libre de nucleasas y DNA en cada muestra obteniendo concentraciones finales de 2 ng/µL, 4 ng/µL u 8 ng/µL (Cuadro 7).

Cantidad de DNA	Número de ciclos
2 ng/µL	27
4 ng/µL	26
8 ng/µL	25

Cuadro 7. Ciclaje utilizado en PCR de acuerdo con la concentración de DNA.

Electroforesis capilar

Se realizó la electroforesis capilar siguiendo las recomendaciones del fabricante (Qiagen, 2021) en el equipo Applied Biosystems 3500 Genetic Analyzer (Cuadro 8) de 8 capilares, y software 3500 series data collection v3.1.

Material	Especificaciones
Capilar	Matriz de 36 cm
Polímero	POP7®
Buffer	Anode Buffer Container (ABC) 3500 series Cathode Buffer Container (CBC) 3500 series

Cuadro 8. Material utilizado para la electroforesis capilar en el equipo Applied Biosystems 3500.

Se realizó una mezcla de formamida con el marcador de peso molecular (*size standard*) BT6 para cada reacción utilizando:

→ Formamida 12 μ L

→ *Size Standard* 0.5 μ L

Posteriormente en una placa de 96 pozos se cargaron 12 μ L de la mezcla anterior en cada pozo de acuerdo con el número de muestras por corrida y se agregó 1 μ L de escalera alélica en el primer pozo utilizado y 1 μ L de producto de PCR de cada muestra.

La electroforesis capilar se llevó a cabo conforme a los parámetros del protocolo del instrumento recomendados por el fabricante (Qiagen, 2021) descritas en el Cuadro 9

Parámetro	Ajuste para Applied Biosystems 3500
Tipo de aplicación	Fragmentos
Longitud del capilar	36 cm
Polímero	POP7
Juego de colorantes	BT6
Temperatura del horno	60 °C
Voltaje del ciclo	13.0 kV
Voltaje antes del ciclo	15 kV
Voltaje de inyección	1.2 kV
Tiempo del ciclo	1550 s
Tiempo antes del ciclo	180 s
Tiempo de inyección	8 s

Cuadro 9. Parámetros utilizados para realizar la electroforesis capilar en el equipo 3500 Genetic Analyzer.

Obtención de los alelos y genotipos

Los datos resultantes de la electroforesis capilar se analizaron en el software GeneMapper ID-X 1.5 utilizando los archivos recomendados para el kit Investigator® (Cuadro 10)

Tipo de archivo	Nombre
Paneles	24plex_Panels_x
Sets de Bin	24plex_Bins_x
<i>Stutter</i>	24plex_Stutter_x
<i>Size standard</i>	SST-BTO_60-500bp

Cuadro 10. Archivos de plantilla del kit Investigator® recomendados para GeneMapper ID-X v1.5

Las muestras LCF 002, LCF 004, LCF 006, LCF 056, LCF 073, LCF 134, y LCF 142 presentaron alelos fuera de *ladder* en uno de sus marcadores, por lo que siguiendo las recomendaciones de Butler en su libro *Advanced Topics in Forensic DNA Typing: Interpretation* se realizó un segundo análisis utilizando el kit PowerPlex Fusion 6C, para descartar posibles mutaciones.

Análisis genético poblacional

A partir del archivo de texto obtenido del software de GeneMapper ID-X 1.5 se ordenaron los datos de cada perfil genotípico utilizando Excel de acuerdo con el formato necesario para el software que posteriormente se utilizaría.

Muestra	LocusA	LocusA.1	LocusB	LocusB.1
Nombre de la muestra	# de alelo 1	# de alelo 2	# de alelo 1	# de alelo 2
Muestra	TH01	TH01.1	D3S1358	D3S1358.1
LCF 001	6	9.3	14	15

Formato 1. Formato generado en Excel para su posterior análisis en Access. La parte sombreada de la tabla representa un ejemplo de dicho formato.

# de muestra	# de alelo 1 (Locus A)	# de alelo 2 (Locus A)	# de alelo 1 (Locus B)	# de alelo 2 (Locus B)
1	6	9.3	14	15

Formato 2. Formato generado para su posterior análisis en el software Arlequin v3.5. La parte sombreada de la tabla representa un ejemplo de dicho formato.

id	pop	Locus A	Locus A	Locus B	Locus B
# de muestra	# de población	# de alelo 1	# de alelo 2	# de alelo 1	# de alelo 2
id	pop	TH01	TH01	D3S1358	D3S1358
1	Pop1	6	9.3	14	15

Formato 3. Formato generado para su posterior análisis en el software STRAF v1.0.5. La parte sombreada de la tabla representa un ejemplo de dicho formato.

En el programa Access se cargó la matriz correspondiente al Formato 1 para realizar una cuenta del número de genotipos repetidos para cada *locus*, con esta información se generaron

matrices individuales para cada *locus* (Anexo 2, páginas 79 a 87, Matriz 1-Matriz 21) a partir de las cuales se obtuvieron las frecuencias alélicas de cada uno de los alelos observados en los 21 STR autosómicos.

Para el software Arlequín fue necesario guardar la matriz (Formato 2) como texto separado por tabulaciones, para su correcta lectura. Se generó un proyecto con datos de tipo microsatélite, a partir de datos genotípicos, para una sola muestra y *locus* separados por tabulaciones. En este software se calculó el equilibrio de Hardy-Weinberg.

Mientras que en el software STRAF v1.0.5 se calcularon los parámetros estadísticos de interés forense, los cuales fueron: heterocigotos esperados, contenido de información polimórfica, poder de discriminación, poder de exclusión, índice típico de paternidad y probabilidad de coincidencia al azar para cada uno de los STRs autosómicos analizados. Para utilizar este software también fue necesario guardar la matriz generada en Excel (Formato 3) como texto separado por tabulaciones.

Posteriormente se realizó un estudio interpoblacional utilizando la paquetería *hierfstaf* (Goudet, Jombart, Kamvar, Archer, & Hardy, 2022) del software Rstudio, para conocer las distancias genéticas existentes entre poblaciones del mismo país. Para esto se utilizaron dos estudios similares publicados por Melo, et al. en 2022 y Martínez-Cortés, et al. en 2018 en los que se analizaron los mismos 21 STRs autosómicos del kit Investigator® 24plex, para el análisis poblacional del estado de Veracruz y la ciudad de Tijuana respectivamente.

Por último, también utilizando Rstudio, pero con las paqueterías *gplots* (Warnes, et al., 2022) y *hierfstaf* (Goudet, Jombart, Kamvar, Archer, & Hardy, 2022) se realizó un análisis descriptivo para generar gráficos tipo *Heatmaps* en los que se muestra las distintas frecuencias alélicas calculadas en las tres poblaciones antes mencionadas.

RESULTADOS Y DISCUSIÓN

Protocolo de PCR

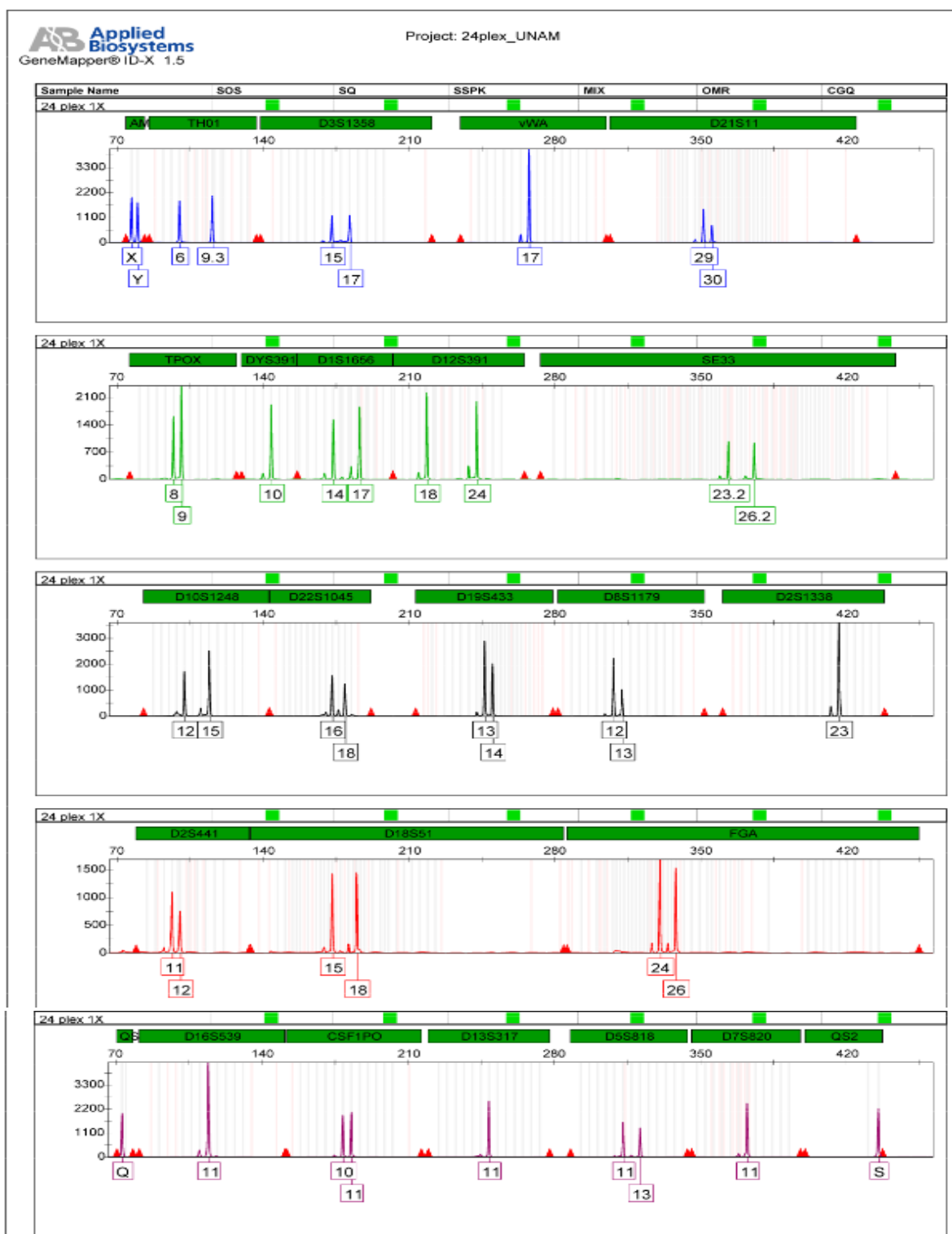
En el Electroferograma 3 resultante de la amplificación a $1/5$ de reacción se puede observar un perfil parcial de la muestra LCF001. Fueron amplificados 12 de los 21 *loci* analizados por el kit, los cuales fueron: D3S1358, D21S11, D22S1045, D8S1179, D2S1338, D2S441, D18S51, FGA, D16S539, D13S317, D5S818 y D7S820. En el electroferograma se puede observar ruido en la amplificación y un desbalance en los picos de los *loci* D8S1179, D16S539, FGA y D2S441 esto puede deberse a inhibidores presentes en la muestra.

Posteriormente se encuentra el Electroferograma 4 perteneciente a la muestra LCF001 amplificado a $1/4$ de reacción del cual no fue posible ningún *loci* amplificado.

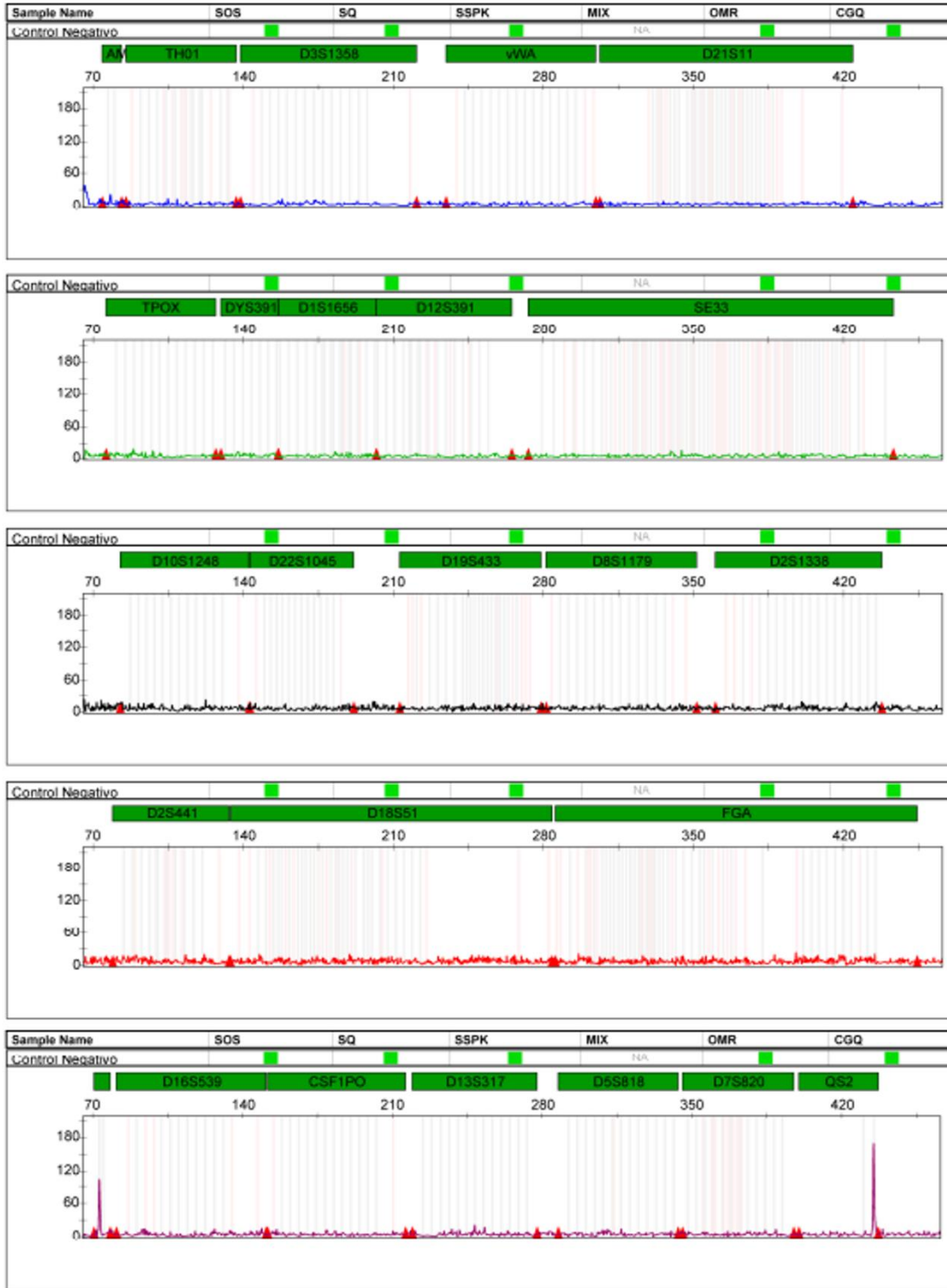
Mientras que en el Electroferograma 5, de la amplificación a $1/3$ de reacción se puede observar un perfil completo, dado que en la amplificación del DNA control 9948 realizada con el protocolo del fabricante se obtuvo un perfil completo (Electroferograma 1) se descarta el mal funcionamiento del kit, por lo tanto, se considera que no fue posible obtener perfiles completos de la muestra LCF001 con las reacciones escaladas a $1/5$ y a $1/4$ debido a inhibidores presentes en la muestra.

Debido a los resultados previamente mencionados se utilizaron tres muestras más para verificar la reproducibilidad de la amplificación a $1/3$ de reacción; en dichas muestras se obtuvieron los perfiles completos y con picos por arriba de los 1000 RFUs. Dado que con ese nivel de señal los picos están por encima del ruido de fondo y del umbral analítico, pero no son demasiado altos como para tener que preocuparse por los datos fuera de escala y contaminación en los canales de color adyacente (Butler J. , 2014) se decidió procesar las muestras restantes con este protocolo de amplificación.

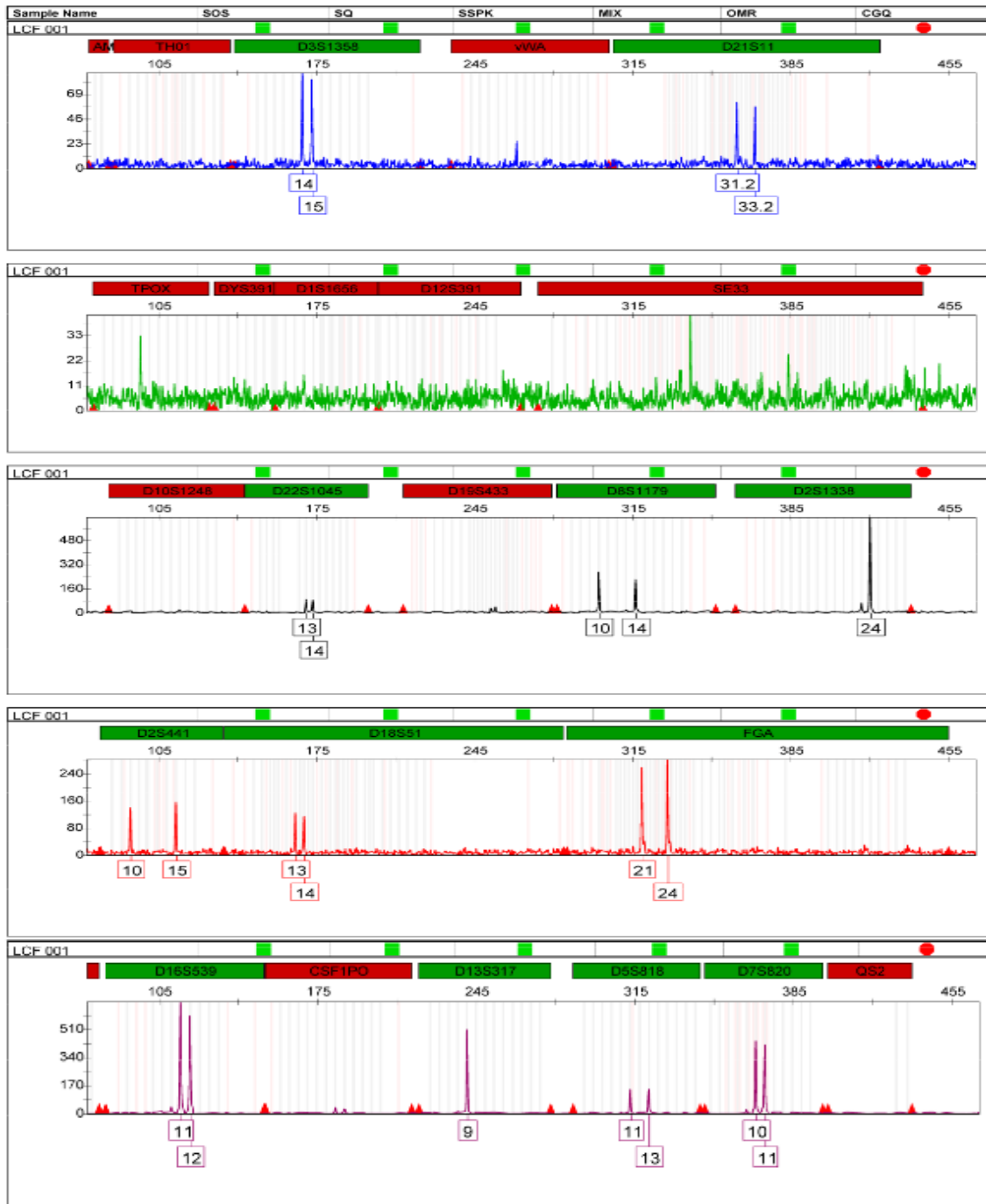
Electroferogramas



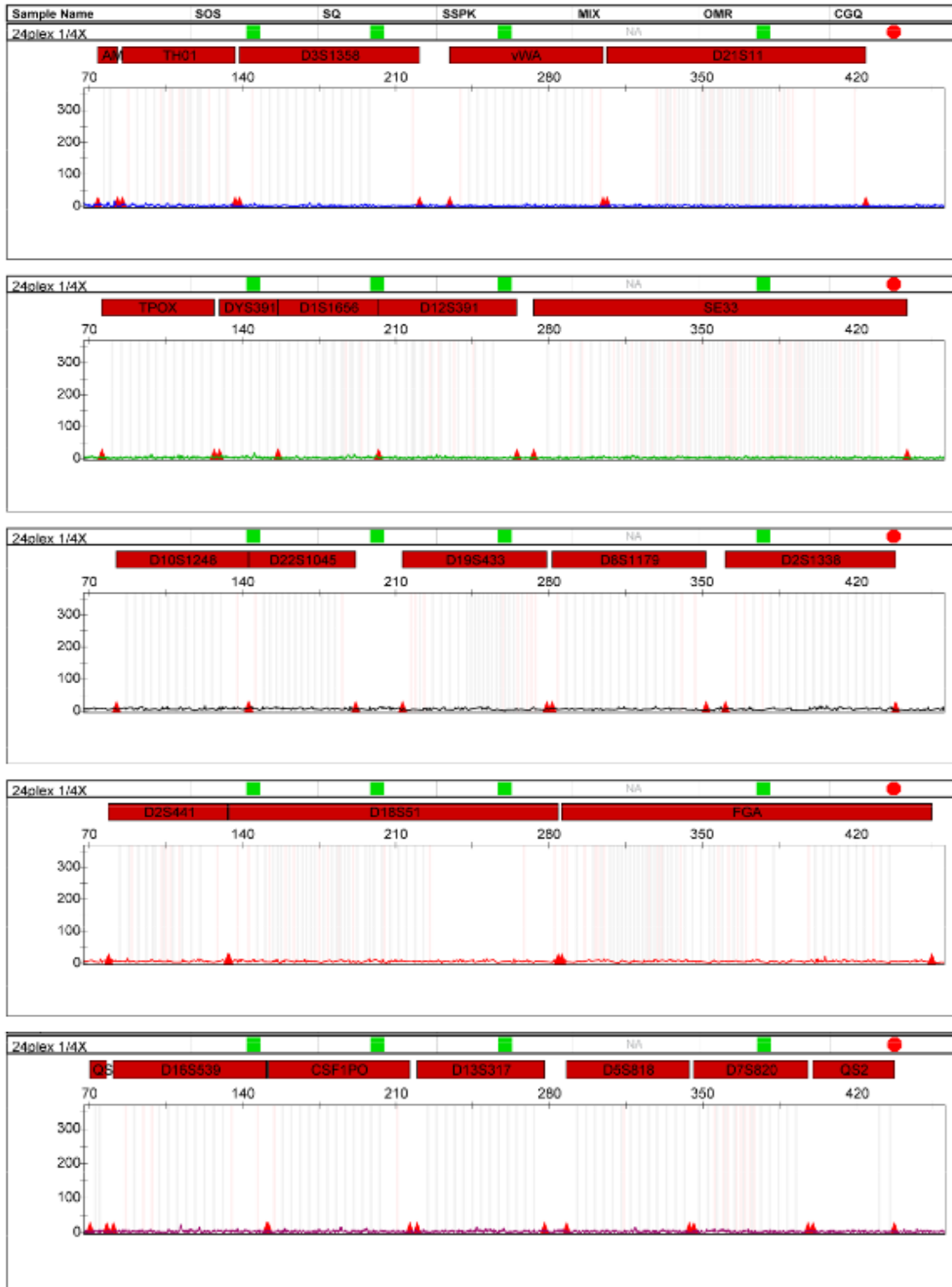
Electroferograma 1. Electroferograma resultante de la amplificación con reacción completa del DNA control 9948 proporcionado por el kit Investigator® 24plex Qiagen.



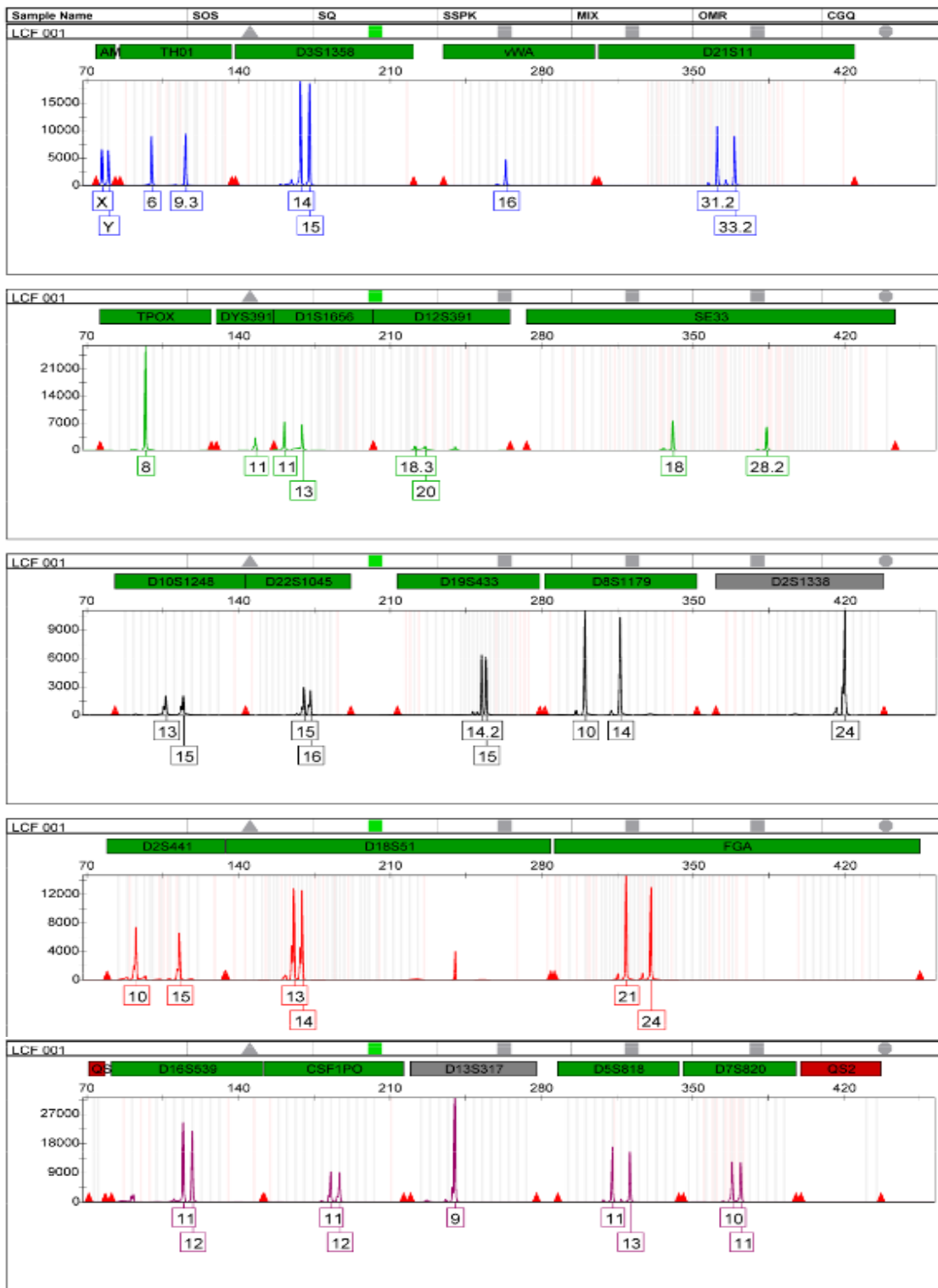
Electroferograma 2. Electroferograma resultante de la amplificación con reacción completa del control negativo el kit Investigator® 24plex Qiagen.



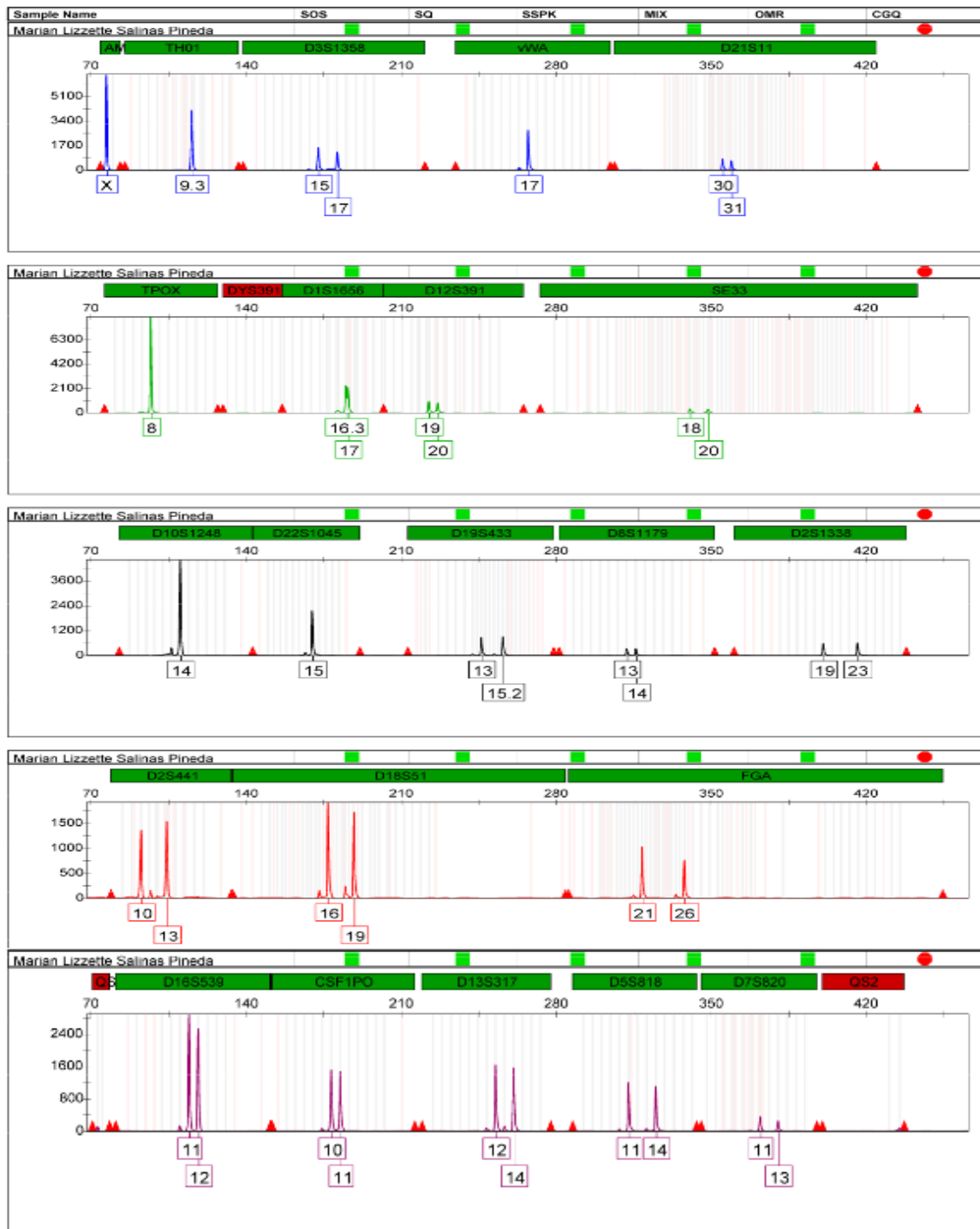
Electroferograma 3. Electroferograma resultante de la amplificación a $1/5$ de reacción de la muestra LCF001.



Electroferograma 4. Electroferograma resultante de la amplificación a 1/4 de reacción de la muestra LCF001.



Electroferograma 5. Electroferograma resultante de la amplificación a $\frac{1}{3}$ de reacción de la muestra LCF001.



Electroferograma 6. Electroferograma resultante de la amplificación del DNA de la persona responsable del procesamiento de las muestras.

Frecuencias alélicas

Como resultado del análisis en Microsoft Excel de cada uno de los alelos observados en los 21 STR autosómicos del kit Investigator[®] 24plex QS de Qiagen (Anexo 2, páginas 79 a 87 Matriz 1-Matriz 21) se obtuvieron las frecuencias alélicas presentes en la muestra de población mestiza-mexicana de la CDMX, las cuales fueron concentradas en las Tabla 1 y Tabla 2 que se muestran a continuación.

Alelo	TH01	D3S1358	vWA	D21S11	TPOX	D1S1656	D12S391	SE33	D10S1248	D22S1045	D19S433
4	0.0032										
6	0.2756				0.0096						
7	0.3590				0.0032						
8	0.0865				0.4904						
9	0.0673				0.0449						
9.3	0.2051										
10	0.0032				0.0641				0.0032	0.0032	
11					0.2436	0.0321			0.0032	0.0641	
11.2											0.0064
11.3											
12					0.1410	0.0865			0.0224	0.0064	0.0385
12.2											0.0032
13		0.0064			0.0032	0.1058		0.0032	0.2276	0.0064	0.2019
13.2											0.1090
13.3								0.0032			
14		0.0513	0.0321			0.1122		0.0160	0.3878	0.0128	0.2724
14.2											0.0449
14.3						0.0064					
15		0.4808	0.0705			0.1538	0.0160	0.0513	0.2660	0.4551	0.1667
15.2											0.0897
15.3						0.0224					
16		0.2340	0.3942			0.1731	0.0160	0.0417	0.0769	0.3686	0.0417
16.2								0.0032			0.0192
16.3						0.0321					
17		0.1250	0.3077			0.0737	0.0449	0.0962	0.0128	0.0737	0.0064
17.3						0.1410	0.0128	0.0032			
18		0.0994	0.1442			0.0064	0.1667	0.0929		0.0032	
18.2											
18.3						0.0417	0.0064				
19		0.0032	0.0513				0.3013	0.0865		0.0064	
19.3						0.0128					
20							0.2179	0.0417			
20.2								0.0064			
21							0.0897	0.0096			
21.2								0.0192			
22							0.0449	0.0064			
22.2								0.0032			
23							0.0353				
23.2								0.0160			
24							0.0321				
24.2				0.0064				0.0160			
25							0.0128				
25.2								0.0481			
26				0.0032				0.0032			
26.2								0.0994			
27				0.0160							
27.2								0.0865			
28				0.0705			0.0032				
28.2								0.0641			
29				0.2276							
29.2				0.0064				0.0897			
30				0.2404							
30.2				0.0192				0.0513			
31				0.0609							
31.2				0.1314				0.0288			
32				0.0064							
32.2				0.1282				0.0096			
33.1				0.0096							
33.2				0.0673							
34								0.0032			
34.2				0.0032							
35				0.0032							

Tabla 1. Frecuencias alélicas calculadas para los *loci* TH01, D3S1358, vWA, D21S11, TPOX, D1S1656, D12S391, SE33, D10S1248 y D22S1045 obtenidas con el kit Investigator® 24plex QS Qiagen.

Alelo	D8S1179	D2S1338	D2S441	D18S51	FGA	D16S539	CSF1PO	D13S317	D5S818	D7S820
4										
6										
7						0.0064			0.0481	0.0128
8	0.0096					0.0032	0.0064	0.0641	0.0064	0.0801
9	0.0032			0.0032		0.0865	0.0096	0.2372	0.0545	0.0705
9.3										
10	0.0994		0.4038			0.2019	0.2885	0.0962	0.0833	0.2212
11	0.0449		0.3526	0.0128		0.1955	0.2628	0.1635	0.4295	0.3429
11.2										
11.3			0.0160							
12	0.1186		0.0321	0.0897		0.3718	0.3558	0.2404	0.2756	0.2372
12.2										
13	0.3237		0.0160	0.1410		0.1186	0.0609	0.1186	0.0962	0.0321
13.2										
13.3										
14	0.2724		0.1571	0.1410		0.0160	0.0160	0.0769	0.0064	0.0032
14.2			0.0192							
14.3										
15	0.1026			0.1635				0.0032		
15.2										
15.3										
16	0.0224	0.0288	0.0032	0.1122						
16.2										
16.3										
17	0.0032	0.1474		0.1667						
17.3										
18		0.0513		0.0673	0.0160					
18.2					0.0032					
18.3										
19		0.1987		0.0353	0.0865					
19.3										
20		0.1538		0.0321	0.0705					
20.2										
21		0.0288		0.0160	0.0641					
21.2					0.0064					
22		0.0833		0.0096	0.1410					
22.2										
23		0.1955		0.0064	0.1218					
23.2					0.0032					
24		0.0513			0.1667					
24.2					0.0032					
25		0.0545		0.0032	0.1603					
25.2										
26		0.0032			0.1250					
26.2										
27		0.0032			0.0256					
27.2										
28					0.0032					
28.2										
29										
29.2										
30					0.0032					
30.2										
31										
31.2										
32										
32.2										
33.1										
33.2										
34										
34.2										
35										

Tabla 2. Frecuencias alélicas calculadas para los *loci* D19S433, D8S1179, D2S1338, D2S441, D18S51, FGA, D16S539, CSF1PO, D13S317, D5S818 y D7S820 obtenidas con el kit Investigator® 24plex QS Qiagen.

Parámetros estadísticos de interés forense de una base de datos

Para el sistema 24plex QS población mestizo-mexicana de la CDMX

En la Tabla 3 y Tabla 4 se muestran los resultados del cálculo del equilibrio de Hardy-Weinberg (HW) y los siguientes parámetros estadísticos de interés forense: heterocigotos observados (Ho); heterocigotos esperados (He); contenido de información polimórfica (PIC); poder de discriminación (PD); poder de exclusión (PE); probabilidad de coincidencia al azar (PM); índice típico de paternidad (ITP) para cada uno de los STRs autosómicos analizados.

En los resultados obtenidos para el equilibrio de Hardy-Weinberg utilizando el valor usual de $p=0.05$ (Budowle, et al., 1995), 20 de los 21 *loci* analizados se encontraban en equilibrio, sin embargo, se encontró que la distribución de genotipos del *locus* D10S1248 se desviaba del HW teniendo un valor de 0.0249. Si bien se considera que la existencia de desequilibrio de Hardy-Weinberg en un *locus* tiene un efecto muy pequeño en el perfil de frecuencias multilocus estimadas para propósitos forenses (Budowle, et al., 1995) se realizó una corrección de Bonferroni (Butler J. , 2014) quedando un valor de $p= 0.0023$ esto se aplicó a los 21 STRs autosómicos del sistema 24plex QS y dio como resultado que todos los *loci* estuvieran en equilibrio de Hardy-Weinberg.

El *locus* SE33 es el que tiene los valores más representativos de todos los parámetros utilizados, esto puede deberse a que también es el que cuenta con la mayor variedad de alelos analizados (28).

PEIF	Locus										
	TH01	D3S1358	vWA	D21S11	TPOX	D1S1656	D12S391	SE33	D10S1248	D22S1045	D19S433
HW	0.3096	0.4692	0.8984	0.2717	0.3827	0.1625	0.5343	0.1292	0.0249	0.646	0.859
Ho	0.7308	0.7051	0.6923	0.8077	0.7500	0.8654	0.8141	0.9359	0.7179	0.6987	0.8205
He	0.7434	0.6881	0.7228	0.8445	0.6762	0.8881	0.8213	0.9353	0.7228	0.6492	0.8343
PIC	0.6982	0.6441	0.6759	0.8235	0.6296	0.8743	0.7972	0.9282	0.6726	0.5841	0.8114
PD	0.8858	0.8594	0.8854	0.9532	0.8339	0.9704	0.9408	0.9848	0.8669	0.8004	0.9502
PE	0.4774	0.4362	0.4165	0.6134	0.5098	0.7254	0.6255	0.8692	0.4566	0.4263	0.6376
PM	0.1142	0.1406	0.1146	0.0468	0.1661	0.0296	0.0592	0.0152	0.1331	0.1996	0.0498
ITP	1.8571	1.6957	1.6250	2.6000	2.0000	3.7143	2.6897	7.8000	1.7727	1.6596	2.7857

Tabla 3. Equilibrio de Hardy-Weinberg y parámetros estadísticos de interés forense para los *locus* TH01, D3S1358, vWA, D21S11, TPOX, D1S1656, D12S391, SE33, D10S1248 y D22S1045. PEIF, Parámetro estadístico de interés forense; HW, valor de p del equilibrio de Hardy-Weinberg; Ho, heterocigotos observados; He, heterocigotos esperados; PIC, contenido de información polimórfica; PD, poder de discriminación; PE, poder de exclusión; PM, probabilidad de coincidencia al azar; ITP, índice típico de paternidad.

PEIF	Locus									
	D8S1179	D2S1338	D2S441	D18S51	FGA	D16S539	CSF1PO	D13S317	D5S818	D7S820
HW	0.2962	0.2717	0.0713	0.5695	0.8815	0.9046	0.4755	0.5472	0.3115	0.0956
Ho	0.7115	0.8590	0.6474	0.9038	0.8718	0.7628	0.7372	0.8397	0.7115	0.7436
He	0.7864	0.8628	0.6882	0.8805	0.8814	0.7634	0.7194	0.8286	0.7203	0.7671
PIC	0.7541	0.8446	0.6301	0.8653	0.8663	0.7262	0.6647	0.8031	0.6784	0.7287
PD	0.9242	0.9562	0.8473	0.9672	0.9680	0.9080	0.8563	0.9412	0.8850	0.9042
PE	0.4463	0.7126	0.3517	0.8033	0.7382	0.5320	0.4881	0.6748	0.4463	0.4989
PM	0.0758	0.0438	0.1527	0.0328	0.0320	0.0920	0.1437	0.0588	0.1150	0.0958
ITP	1.7333	3.5455	1.4182	5.2000	3.9000	2.1081	1.9024	3.1200	1.7333	1.9500

Tabla 4. Equilibrio de Hardy-Weinberg y parámetros estadísticos de interés forense para los *locus* D19S433, D8S1179, D2S1338, D2S441, D18S51, FGA, D16S539, CSF1PO, D13S317, D5S818 y D7S820. PEIF, Parámetro estadístico de interés forense; HW, valor de p del equilibrio de Hardy-Weinberg; Ho, heterocigotos observados; He, heterocigotos esperados; PIC, contenido de información polimórfica; PD, poder de discriminación; PE, poder de exclusión; PM, probabilidad de coincidencia al azar; ITP, índice típico de paternidad.

También se calculó el valor combinado de los parámetros de heterocigotos observados, heterocigotos esperados, poder de discriminación, poder de exclusión, probabilidad de coincidencia al azar e índice típico de paternidad; esto para conocer los valores de los parámetros para toda la base de datos generada (Tabla 5).

PEIF	Valor combinado
Ho	0.004568885
He	0.005138751
PD	0.130394753
PE	3.43513×10^{-06}
PM	2.29375×10^{-24}
ITP	127,030,890.8

Tabla 5. Valores combinados para cada parámetro estadístico de interés forense. PEIF, Parámetro estadístico de interés forense; 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.

Dando como resultado un poder de exclusión combinado de 0.99999656487 es decir, este será el valor medio para cada análisis de perfiles genéticos completos (considerando los 21 *loci*) en los que se utilice la base de datos.

En el caso de la probabilidad de coincidencia al azar se obtuvo un valor combinado de 2.29375×10^{-24} , al igual que con el poder de exclusión, este es el valor medio que se puede esperar al utilizar la base de datos, cuando el análisis es entre perfiles genéticos completos.

Mientras que el valor de ITP reportado indica que cuando se utilice la base de datos en un caso típico de parentesco biológico (Figura 7), se obtendrá en promedio que es 127,030,890.8 más probable que el posible padre sea el padre biológico que un hombre seleccionado al azar en la población.

Comparación con estudios similares realizados dentro de la República Mexicana

Se realizó una comparación entre el sistema utilizado en el presente estudio (Investigator 24plex QS) y con cuatro reportados en la literatura utilizados en población mexicana. En la Tabla 6 se muestra la comparación de los marcadores utilizados en cada sistema, se puede observar que en los estudios de Melo et al 2022 (24 plex GO!) y Martínez-Cortés et al 2018 (GlobalFiler) se utilizaron los mismos 21 STRs autosómicos, aun cuando los sistemas son distintos.

En el estudio de Gonzalez-Herrera et al 2010 en el que utilizaron el sistema PowerPlex Fusion se utilizan 22 STRs autosómicos, es decir un marcador más que en el Investigator 24plex QS. Dentro de los 22 marcadores utilizados se encuentran los PENTA E y PENTA D, mientras que se excluye el SE33 incluido en el sistema del presente estudio.

Mientras que en el estudio de Ramírez Flores et al en el 2014 se realizó una combinación entre los sistemas PowerPlex Fusion y PowerPlex 21, dando como resultado el análisis de 23 STRs autosómicos, en los cuales se incluyen los marcadores PENTA E, PENTA D y D6S1043 que no se analizan en el sistema Investigator 24plex QS y se excluye al igual que en el caso anterior al marcador SE33.

La mayor diferencia se encuentra con el sistema PowerPlex 16 utilizado por Rangel Villalobos et al en 2010 ya que en este sistema sólo se analizan 15 marcadores incluyendo los PENTA E y PENTA D, y no se analizan los marcadores D1S1656, D12S391, SE33, D10S1248, D22S1045, D19S433, D2S1338 y D2S441 que se encuentran presentes en el sistema 24plex QS, como se puede observar en la siguiente tabla (Tabla 6):

<i>Locus</i>	24plex QS	24plex GO!	GlobalFiler	PowerPlex Fusion	PowerPlex Fusion + PowerPlex21	PowerPlex 16
TH01	✓	✓	✓	✓	✓	✓
D3S1358	✓	✓	✓	✓	✓	✓
vWA	✓	✓	✓	✓	✓	✓
D21S11	✓	✓	✓	✓	✓	✓
TPOX	✓	✓	✓	✓	✓	✓
D1S1656	✓	✓	✓	✓	✓	
D12S391	✓	✓	✓	✓	✓	
SE33	✓	✓	✓			
D10S1248	✓	✓	✓	✓	✓	
D22S1045	✓	✓	✓	✓	✓	
D19S433	✓	✓	✓	✓	✓	
D8S1179	✓	✓	✓	✓	✓	✓
D2S1338	✓	✓	✓	✓	✓	
D2S441	✓	✓	✓	✓	✓	
D18S51	✓	✓	✓	✓	✓	✓
FGA	✓	✓	✓	✓	✓	✓
D16S539	✓	✓	✓	✓	✓	✓
CSF1PO	✓	✓	✓	✓	✓	✓
D13S317	✓	✓	✓	✓	✓	✓
D5S818	✓	✓	✓	✓	✓	✓
D7S820	✓	✓	✓	✓	✓	✓
PENTA E				✓	✓	✓
PENTA D				✓	✓	✓
D6S1043					✓	

Tabla 6. Comparación de los marcadores incluidos en cada uno de los sistemas con los cuales se va a comparar el presente estudio. El sistema 24plex GO! utilizado en Veracruz por Melo, *et al* en 2022, GlobalFiler utilizado en Tijuana por Martínez-Cortés, *et al* en 2018, PowerPlex Fusion utilizado en Yucatán por Gonzalez-Herrera, *et al* en 2019, PowerPlex Fusión + PowerPlex21 utilizado en CDMX por Ramírez Flores, *et al* en 2014 y PowerPlex 16 utilizado en Guanajuato por Rangel-Villalobos, *et al* en 2010.

En la Tabla 7 se puede observar que los valores para los 5 parámetros analizados de los sistemas 24plex GO! en población Veracruzana, PowerPlex Fusion en población de Yucatán, y GlobalFiler en población de Tijuana son similares con el sistema utilizado 24plex QS en población mestiza de la CDMX, mientras que con el sistema PowerPlex 16 en población de Guanajuato se observa diferencia de una orden de magnitud en cada uno de los parámetros, esto debido a que el número de marcadores analizados en PowerPlex16 es menor que en los sistemas restantes.

PEIF	24plex QS	24plex GO!	GlobalFiler	PowerPlex Fusion	PowerPlex Fusion + Powerplex21	PowerPlex 16
n	156	234	381	733	200	344
Ho	0.004568885	0.00385079	0.004416021	0.004449672	0.00452126	0.02883618
He	0.005138751	0.00437562	0.005359504	No reportado	0.00431113	0.030364772
PD	0.130394753	0.12548056	0.145770948	0.152907803	0.12011782	0.224907416
PE	3.44×10^{-06}	2.2573×10^{-06}	3.17313×10^{-06}	3.11021×10^{-06}	1.03189×10^{-06}	9.39689×10^{-05}
PM	2.29×10^{-24}	1.5337×10^{-24}	3.15499×10^{-24}	No reportado	2.5234×10^{-22}	1.21966×10^{-14}
ITP	127,030,890.8	60,354,023.9	108,137,606	297,881,540.5	No reportado	14,868.83566

Tabla 7. Comparación de los valores de los parámetros estadísticos de interés forense obtenidos para el kit y la población utilizados en el presente estudio y los reportados en la literatura. PEIF, Parámetro estadístico de interés forense; n, tamaño de muestra; 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. El sistema 24plex GO! utilizado en Veracruz por Melo, *et al* en 2022, GlobalFiler utilizado en Tijuana por Martínez-Cortés, *et al* en 2018, PowerPlex Fusion utilizado en Yucatán por Gonzalez-Herrera, *et al* en 2019, PowerPlex Fusión + PowerPlex21 utilizado en CDMX por Ramírez Flores, *et al* en 2014 y PowerPlex 16 utilizado en Guanajuato por Rangel-Villalobos, *et al* en 2010.

A pesar de que el presente estudio tiene un tamaño de muestra menor (n=156) en comparación con los estudios mencionados, el sistema 24plex QS, presenta un poder de exclusión de 99.99966% el cual se encuentra dentro de la misma orden de magnitud que el PE reportado para los primeros cuatro sistemas con los que se compara, los cuáles actualmente son utilizados en el ámbito forense.

Al comparar los valores de PE del sistema Investigator 24plex QS con el del Powerplex 16 se observa que el último presenta un PE de una orden de magnitud menor, esto debido a que el número de marcadores analizados en el sistema Investigator 24plex QS es mayor.

Es importante mencionar que, aunque se encuentren en la misma orden de magnitud la combinación de los sistemas PowerPlex Fusion + PowerPlex 21 (Ramírez Flores, *et al.*, 2014) tiene un PE mayor, ya que se analizan dos marcadores más que en el sistema Investigator 24plex QS.

Para el sistema PowerPlex Fusion en el estudio de Gonzalez-Herrera, *et al.* en 2019 se reportó un ITP que, si bien se encuentra en la misma orden de magnitud que el del presente estudio tiene un valor del doble, esto puede deberse a que se utiliza un marcador más, y también a que en este kit se utilizan los marcadores PENTA D y PENTA E en los cuáles se encuentra una mayor variación de alelos (Gonzalez-Herrera, *et al.*, 2019).

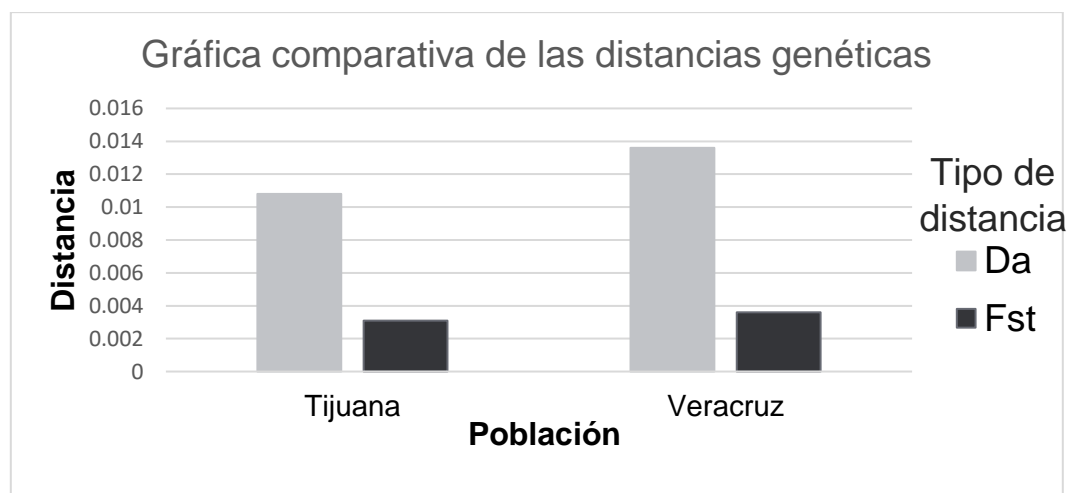
En la Tabla 7 se observa que el ITP combinado calculado para presente estudio y el del sistema 24plex GO!, varía por una orden de magnitud, esto debido que los ITP calculados

por *locus* en el marcador SE33 del estudio del sistema 24plex QS en población de la ciudad de México (ITP=7.8) es casi el doble que el reportado para el mismo marcador en el sistema 24plex GO! (ITP= 4.6) en población del estado de Veracruz, es decir, es menos probable que en Veracruz utilizando solo el marcador SE33 se pueda discernir entre que el posible padre sea el padre real y no un hombre seleccionado al azar en la población; esto debido a que el número de genotipos homocigotos para este marcador es mayor en el estado de Veracruz que en la Ciudad de México.

Comparación de las frecuencias alélicas con dos análisis similares realizados dentro de la República Mexicana.

Se realizó un análisis inter poblacional basado en los STRs del kit Investigator® 24plex QS Qiagen, con población del estado de Veracruz (Melo, et al., 2022) y de la ciudad de Tijuana (Martínez-Cortés, et al., 2018). Se calcularon los dos tipos de distancia previamente descritos, es decir la distancia D_a (Takezaki & Nei, 1996) y la F_{st} (Nei, Tajima, & Tatenno, 1983) entre las dos poblaciones con la del presente trabajo (Gráfica 1) dando como resultado entre la ciudad de México y Tijuana una $D_a = 0.0108$ y $F_{st} = 0.36\%$ y entre ciudad de México y Veracruz $D_a = 0.0136$ y $F_{st} = 0.31\%$. Como se indicó antes estas distancias indican que existe una pequeña diferencia entre las poblaciones tanto entre sus frecuencias alélicas como en la variación de la subestructura de la población.

Estos valores indican que ciudad de México es más distante a Veracruz que a Tijuana en ambos tipos de distancia.



Gráfica 1. Gráfica de las distancias D_a y F_{st} que tiene la Ciudad de México con la ciudad de Tijuana y Veracruz respectivamente.

En comparación con la distancia F_{st} reportada por Martínez-Cortés, et al. en el 2018 para tres poblaciones del Norte de México ($F_{st}= 0.0969\%$), existe una mayor distancia genética entre la población de Ciudad de México, con la ciudad de Tijuana y el estado de Veracruz, que entre las poblaciones del Norte del país. Esto puede deberse a la migración que se vive actualmente dentro del país y a la cercanía geográfica entre las poblaciones del Norte que se analizaron.

A continuación, se muestra en la Figura 9 un escalamiento multidimensional realizado con la distancia F_{st} en el cual se pueden observar de manera gráfica las distancias calculadas entre las poblaciones.

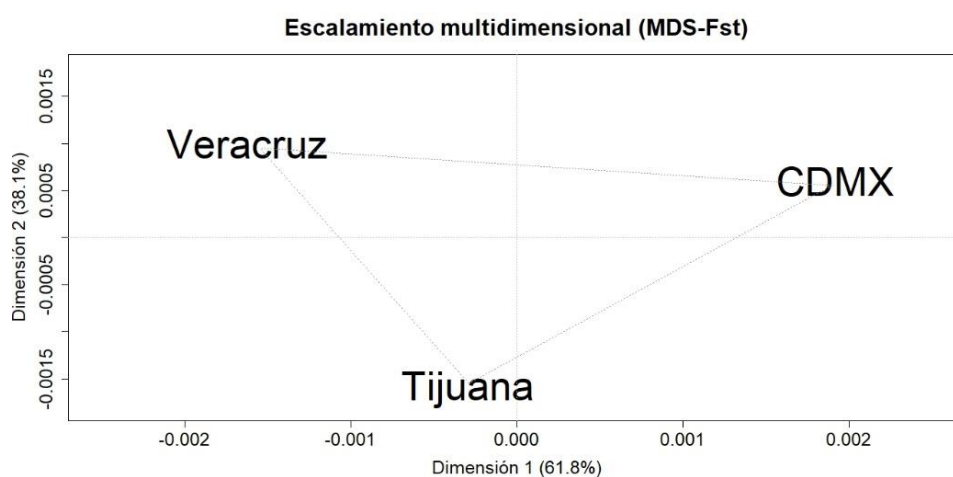
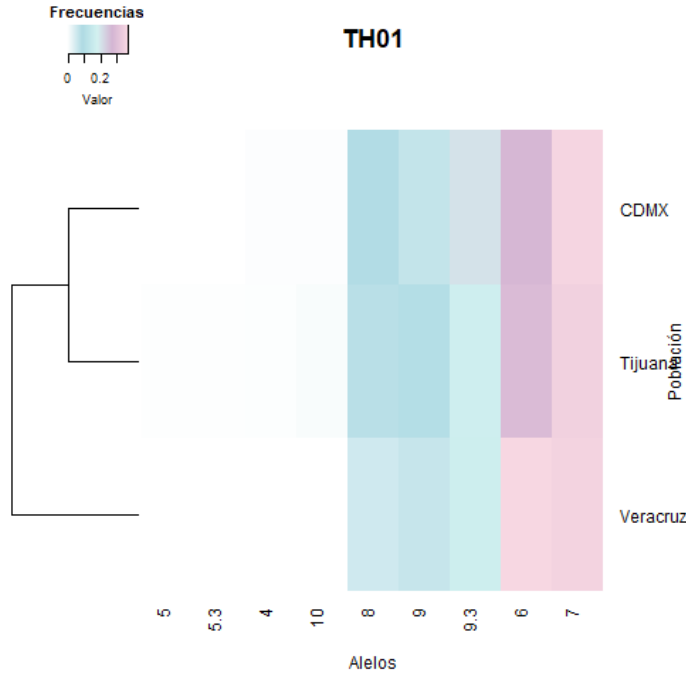


Figura 9. Modelo de escalamiento multidimensional entre las poblaciones de Ciudad de México, Tijuana y Veracruz basado en la distancia F_{st} .

Debido a que la existencia de la distancia D_a encontrada, dentro de las poblaciones que a pesar de ser muy baja es significativa, se realizó una comparación de las frecuencias alélicas de cada uno de los 21 STR autosómicos obtenidos con el sistema 24plex QS en población mestiza de la CDMX con las reportadas en los estudios previamente mencionados con el objetivo de conocer las diferencias de frecuencias entre poblaciones del mismo país, se utilizaron *Heatmaps* para representar las frecuencias de las tres poblaciones estudiadas.

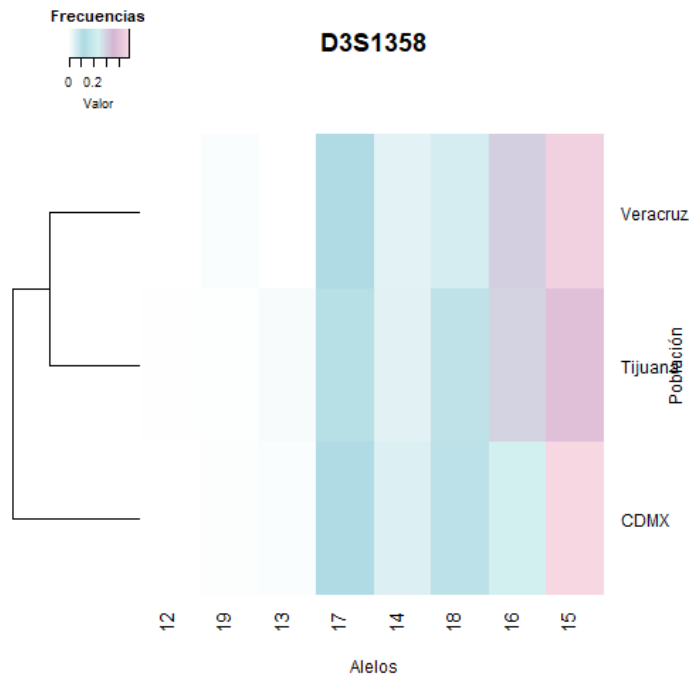
En dichos *Heatmaps* se puede observar que existen alelos en ciertos *locus* que solo están presentes en una o dos de las poblaciones; si bien esta información no es suficiente para poder definir un origen biogeográfico sí se puede observar la necesidad de tener bases de datos representativas de cada población ya que la variabilidad de alelos existentes en cada una es distinta.



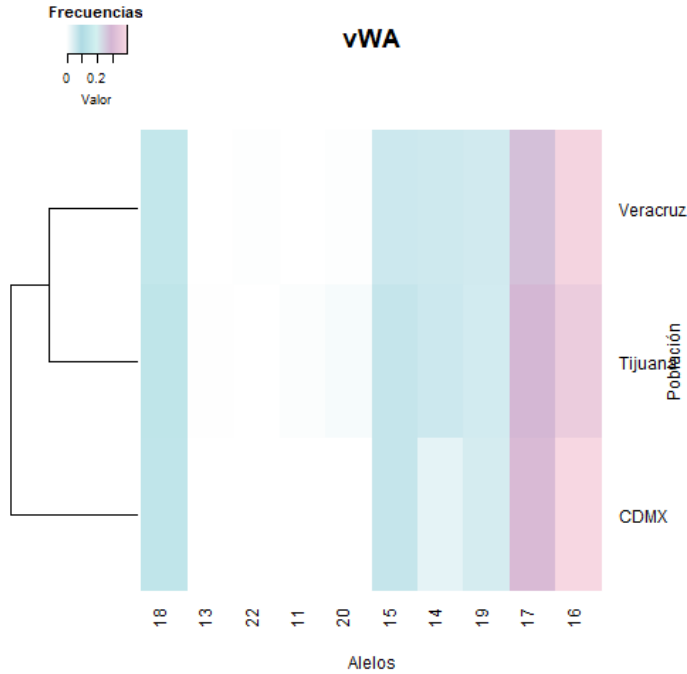
En el *Heatmap 1*, perteneciente al *locus* TH01 se observaron los alelos 4 y 10 en la población de CDMX y Tijuana, pero no en la del estado de Veracruz. Únicamente en la población de ciudad de Tijuana se encontraron los alelos 5 y 5.3.

Heatmap 1. Heatmap generado para el *locus* TH01 a partir de las frecuencias alélicas presentes en la Ciudad de México, la ciudad de Tijuana y el estado de Veracruz. En la parte superior se encuentra una escala de color con un valor entre 0 y 1 que indica las frecuencias observadas en las poblaciones.

En el *locus* D3S1358 (*Heatmap 2*) se observó el alelo 13 en la población de CDMX y Tijuana, pero no en la del estado de Veracruz. Únicamente en la población de ciudad de Tijuana se encontró el alelo 12.



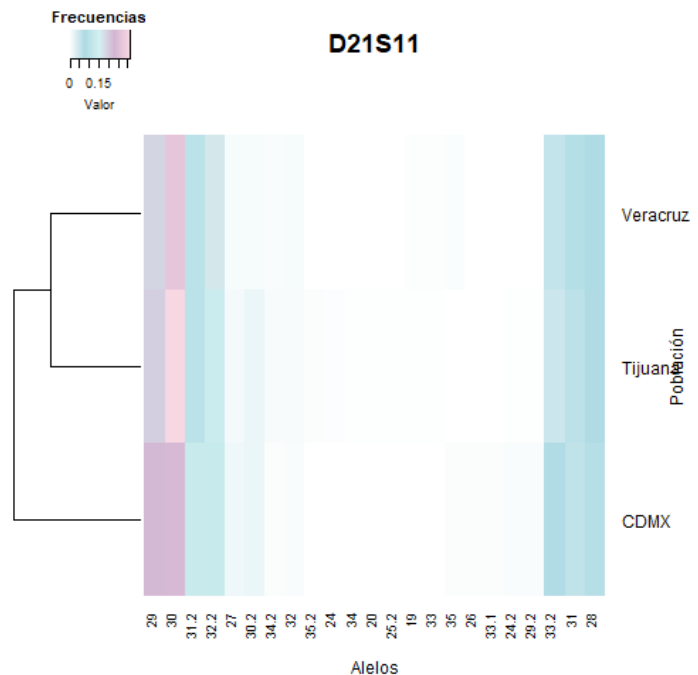
Heatmap 2. Heatmap generado para el *locus* D3S1358 a partir de las frecuencias alélicas presentes en la Ciudad de México, la ciudad de Tijuana y el estado de Veracruz. En la parte superior se encuentra una escala de color con un valor entre 0 y 1 que indica las frecuencias observadas en las poblaciones.



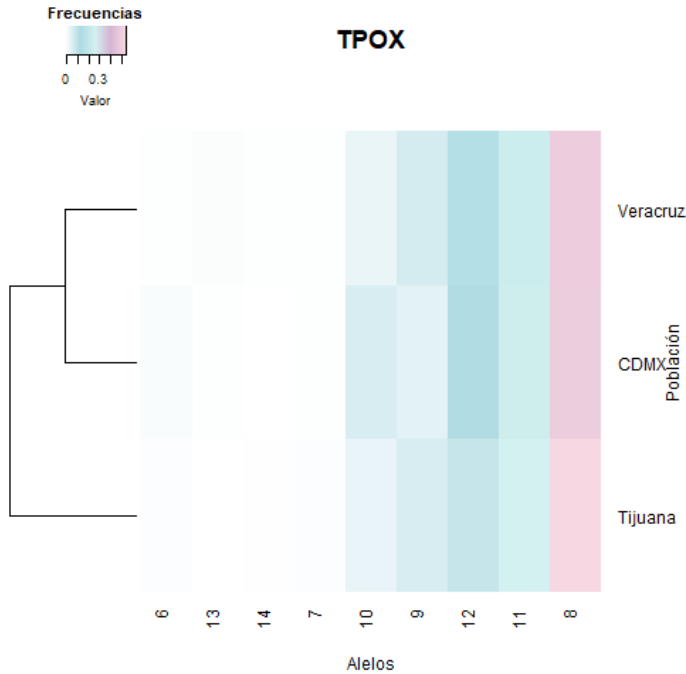
En el *locus* vWA (*Heatmap* 3) Solo en la población del estado de Veracruz se observó el alelo 22 y únicamente en la población de ciudad de Tijuana se encontraron los alelos 11, 13, y 20.

Heatmap 3. *Heatmap* generado para el *locus* vWA a partir de las frecuencias alélicas presentes en la Ciudad de México, la ciudad de Tijuana y el estado de Veracruz. En la parte superior se encuentra una escala de color con un valor entre 0 y 1 que indica las frecuencias observadas en las poblaciones.

En el *locus* D21S11 (*Heatmap* 4) se observó el alelo 24.2 en la población de CDMX y Tijuana, pero no en la del estado de Veracruz. Los alelos 26 y 33.1 se observaron solo en población de la CDMX. El alelo 19 está presente en las poblaciones del estado de Veracruz y de Tijuana. Únicamente en la población de ciudad de Tijuana se encontraron los alelos 20, 24, 25.2, 29.2, 34 y 35.2.



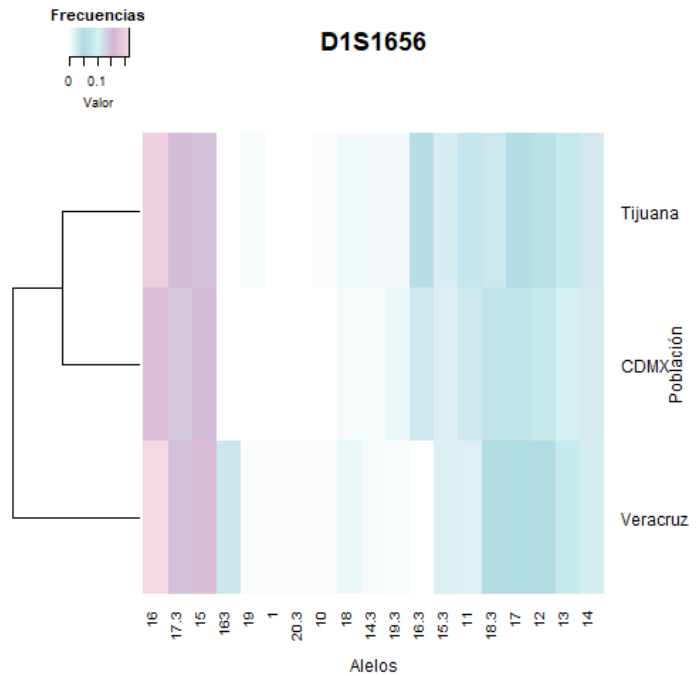
Heatmap 4. *Heatmap* generado para el *locus* D21S11 a partir de las frecuencias alélicas presentes en la Ciudad de México, la ciudad de Tijuana y el estado de Veracruz. En la parte superior se encuentra una escala de color con un valor entre 0 y 1 que indica las frecuencias observadas en las poblaciones.



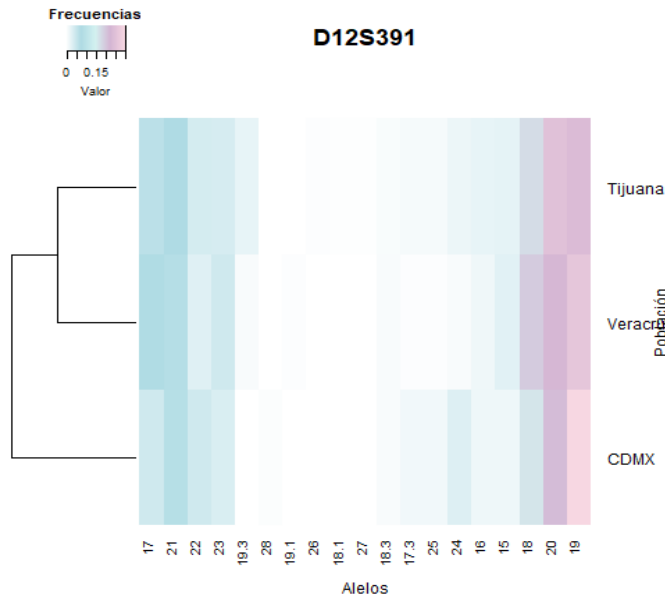
En el *locus* TPOX (*Heatmap* 5) se observó el alelo 14 en población del estado de Veracruz y de la ciudad de Tijuana.

Heatmap 5. *Heatmap* generado para el *locus* TPOX a partir de las frecuencias alélicas presentes en la Ciudad de México, la ciudad de Tijuana y el estado de Veracruz. En la parte superior se encuentra una escala de color con un valor entre 0 y 1 que indica las frecuencias observadas en las poblaciones.

En el *locus* D1S1656 (*Heatmap* 6) alelo 14.3 se obtuvo la misma frecuencia en la población de la CDMX y en la del estado de Veracruz. Los alelos 10 y 19 están presentes en población del estado de Veracruz y de la ciudad de Tijuana. Solo en la población del estado de Veracruz se observó alelo 20.3.



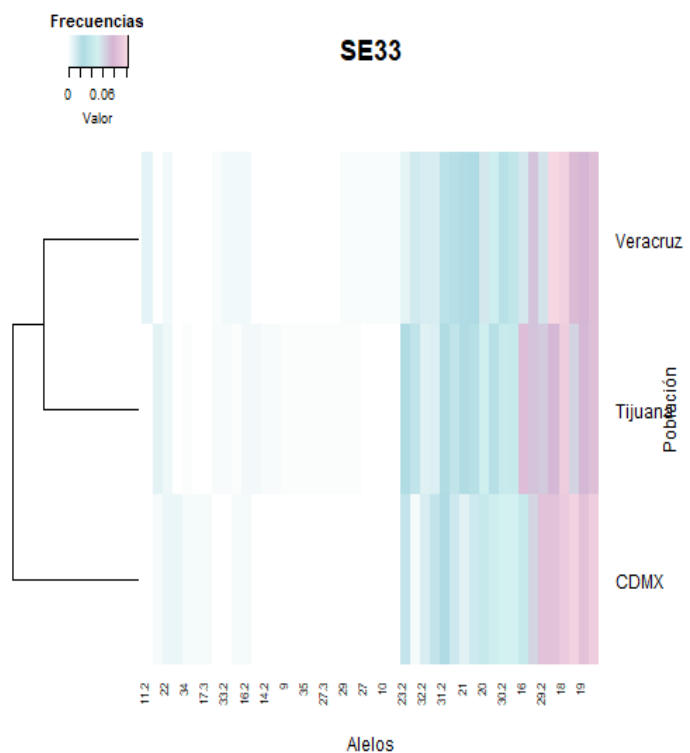
Heatmap 6: *Heatmap* generado para el *locus* D1S1656 a partir de las frecuencias alélicas presentes en la Ciudad de México, la ciudad de Tijuana y el estado de Veracruz. En la parte superior se encuentra una escala de color con un valor entre 0 y 1 que indica las frecuencias observadas en las poblaciones.



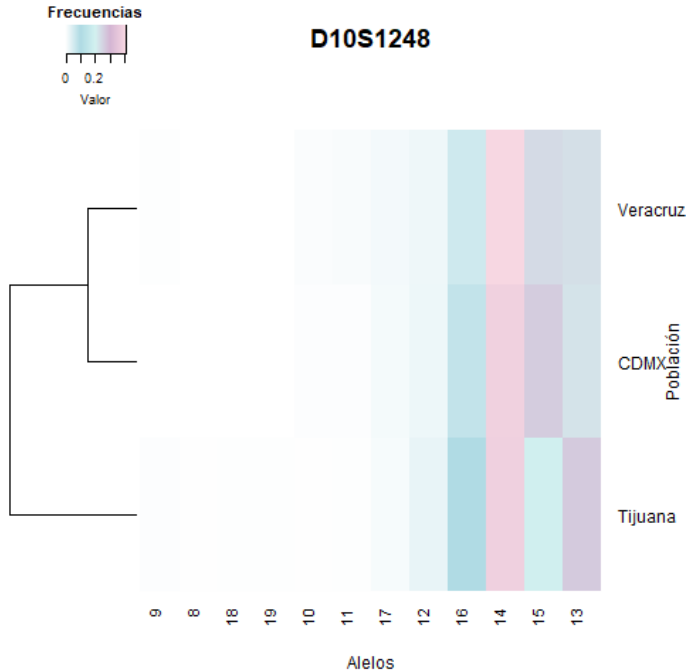
En el *locus* D12S391 (*Heatmap* 7) se obtuvo la misma frecuencia en el alelo 18.3 en la población de la CDMX y en la del estado de Veracruz. El alelo 19.3 está presente en población del estado de Veracruz y de la ciudad de Tijuana. Solo en la población del estado de Veracruz alelo 19.1. Únicamente en la población de ciudad de Tijuana se encontraron los alelos 18.1, 26 Y 27.

Heatmap 7. Heatmap generado para el *locus* D12S391 a partir de las frecuencias alélicas presentes en la Ciudad de México, la ciudad de Tijuana y el estado de Veracruz. En la parte superior se encuentra una escala de color con un valor entre 0 y 1 que indica las frecuencias observadas en las poblaciones.

En el *locus* SE33 (*Heatmap* 8) se observaron los 13 y 34 en SE33 en la población de CDMX y Tijuana, pero no en la del estado de Veracruz. Los alelos 13.3, 17.3 y 20.2 se observaron únicamente en población de la CDMX; mientras que los alelos 12.2, 33 y 33.2 están presentes en población del estado de Veracruz y de la ciudad de Tijuana. Solo en la población del estado de Veracruz se observaron los alelos 10, 11.2, 23, 23.3 y 27. Únicamente en la población de ciudad de Tijuana se encontraron los alelos 8, 9, 14.2, 19.2, 27.3, 28, 29, 31, 34.2 y 35.



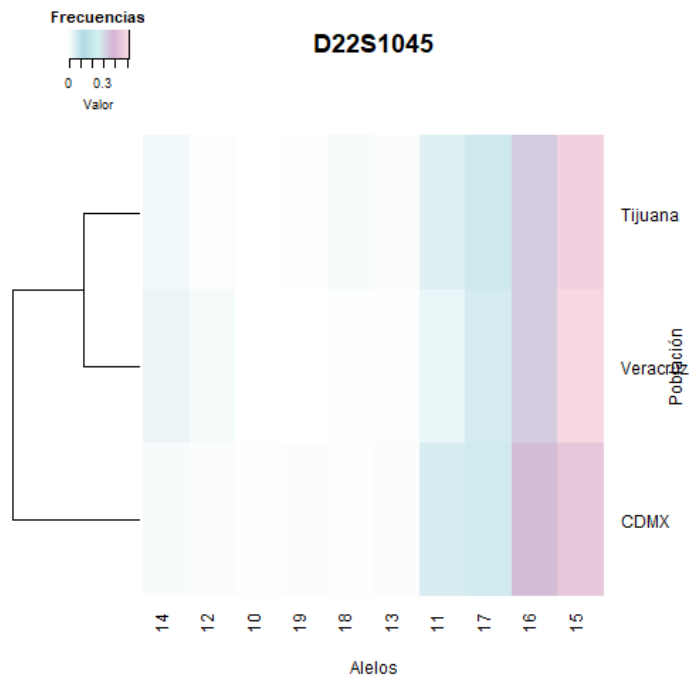
Heatmap 8. Heatmap generado para el *locus* SE33 a partir de las frecuencias alélicas presentes en la Ciudad de México, la ciudad de Tijuana y el estado de Veracruz. En la parte superior se encuentra una escala de color con un valor entre 0 y 1 que indica las frecuencias observadas en las poblaciones.



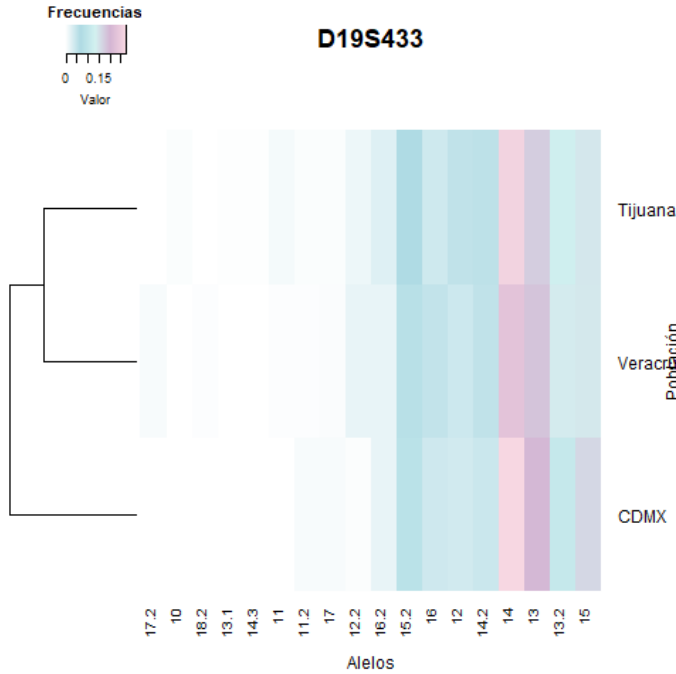
En el *locus* D10S1248 (*Heatmap* 9) el alelo 9 está presente en población del estado de Veracruz y de la ciudad de Tijuana. Únicamente en la población de ciudad de Tijuana se encontraron los alelos 8 y 18.

Heatmap 9. *Heatmap* generado para el *locus* D10S1248 a partir de las frecuencias alélicas presentes en la Ciudad de México, la ciudad de Tijuana y el estado de Veracruz. En la parte superior se encuentra una escala de color con un valor entre 0 y 1 que indica las frecuencias observadas en las poblaciones.

En el *locus* D22S1045 (*Heatmap* 10) se observó el alelo 19 en la población de CDMX y Tijuana, pero no en la del estado de Veracruz. El alelo 10 se observó únicamente en población de la CDMX.



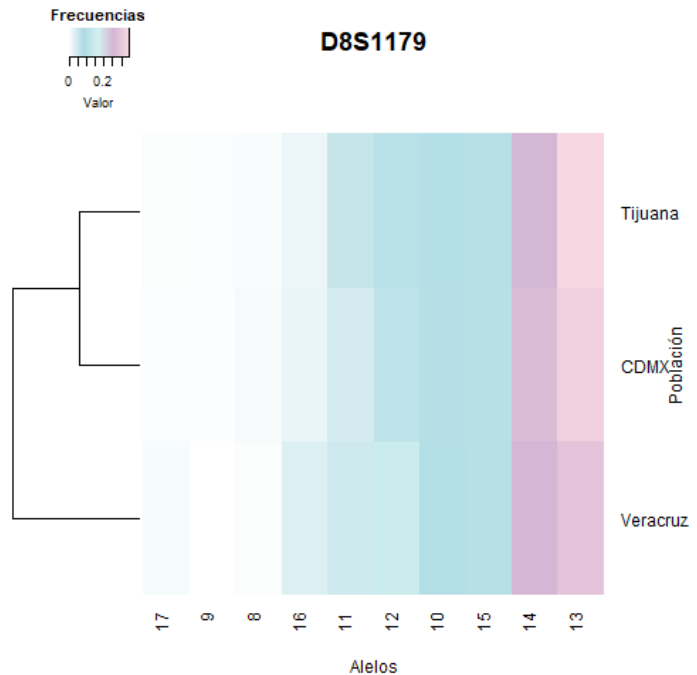
Heatmap 10. *Heatmap* generado para el *locus* D22S1045 a partir de las frecuencias alélicas presentes en la Ciudad de México, la ciudad de Tijuana y el estado de Veracruz. En la parte superior se encuentra una escala de color con un valor entre 0 y 1 que indica las frecuencias observadas en las poblaciones.



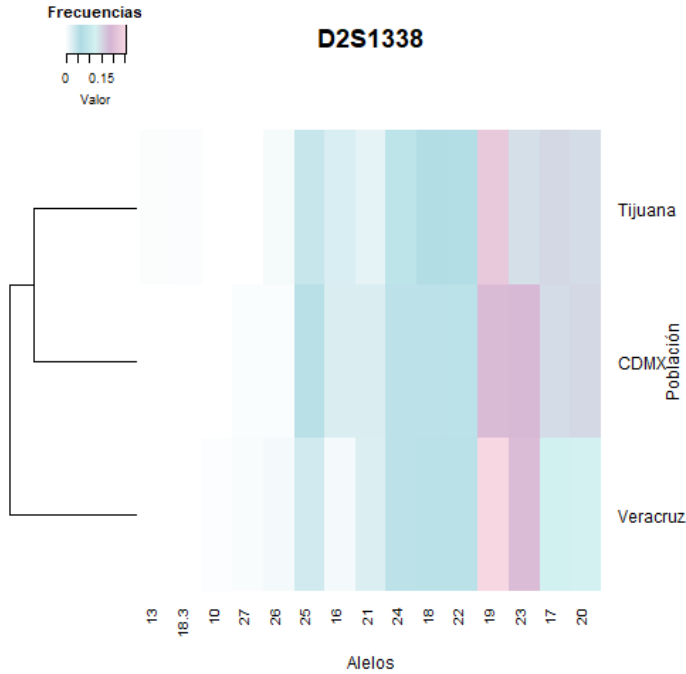
En el *locus* D19S433 (*Heatmap* 11) en el alelo 16.2 se obtuvo la misma frecuencia en las poblaciones de la CDMX y del estado de Veracruz. El alelo 11 está presente en población del estado de Veracruz y de la ciudad de Tijuana. Solo en la población del estado de Veracruz se observaron los alelos 17.2 y 18.2;

Heatmap 11. Heatmap generado para el *locus* D19S433 a partir de las frecuencias alélicas presentes en la Ciudad de México, la ciudad de Tijuana y el estado de Veracruz. En la parte superior se encuentra una escala de color con un valor entre 0 y 1 que indica las frecuencias observadas en las poblaciones.

En el *locus* D8S1179 (*Heatmap* 12) se observó el alelo 9 en la población de CDMX y Tijuana, pero no en la del estado de Veracruz.



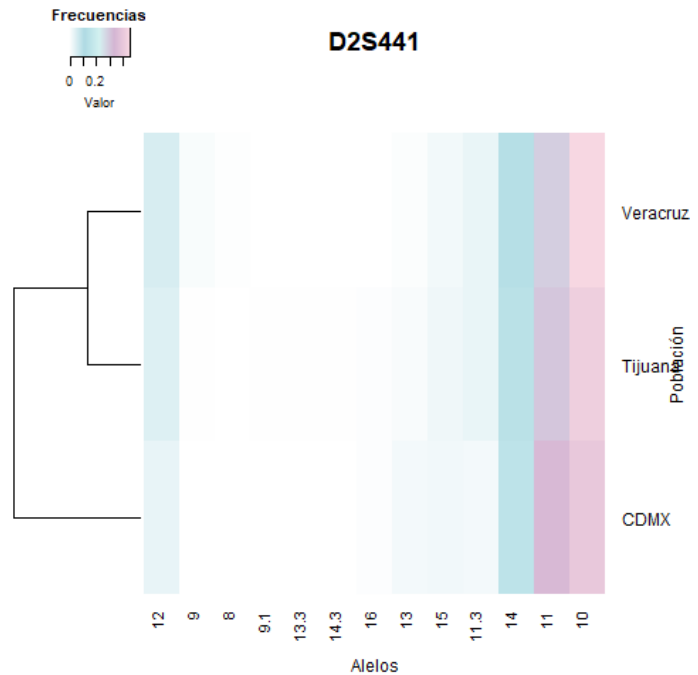
Heatmap 12. Heatmap generado para el *locus* D8S1179 a partir de las frecuencias alélicas presentes en la Ciudad de México, la ciudad de Tijuana y el estado de Veracruz. En la parte superior se encuentra una escala de color con un valor entre 0 y 1 que indica las frecuencias observadas en las poblaciones.



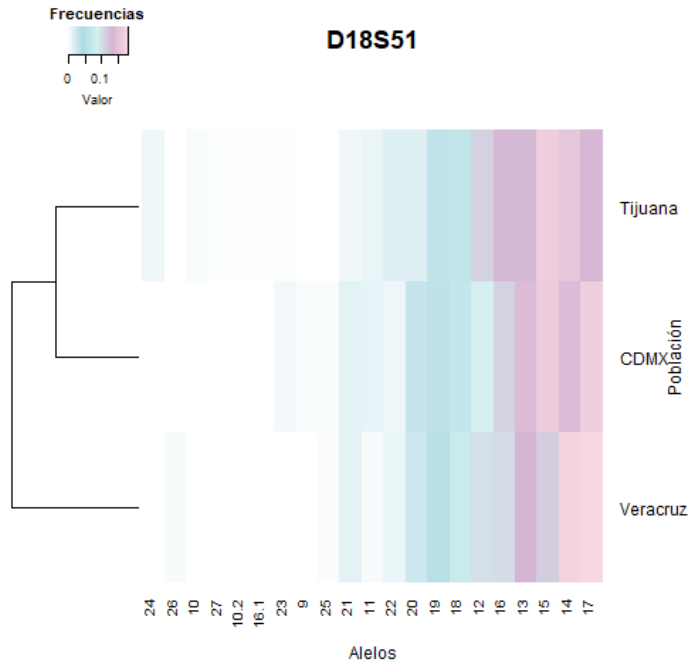
En el *locus* D2S1338 (*Heatmap* 13) alelo 24 se obtuvo la misma frecuencia en las poblaciones de la CDMX y del estado de Veracruz. El alelo 27 observó únicamente en población de la CDMX. Solo en la población del estado de Veracruz se observó el alelo 27.2. Únicamente en la población de ciudad de Tijuana se encontraron los alelos 13,18.3 y 22.

Heatmap 13. *Heatmap* generado para el *locus* D2S1338 a partir de las frecuencias alélicas presentes en la Ciudad de México, la ciudad de Tijuana y el estado de Veracruz. En la parte superior se encuentra una escala de color con un valor entre 0 y 1 que indica las frecuencias observadas en las poblaciones.

En el *locus* D2S441 (*Heatmap* 14) se observó el alelo 16 en la población de CDMX y Tijuana, mientras que el alelo 9 está presente en las poblaciones del estado de Veracruz y Tijuana. Solo en la población del estado de Veracruz se observó el alelo 8. Únicamente en la población de ciudad de Tijuana se encontraron los alelos 9.1 y 14.3.



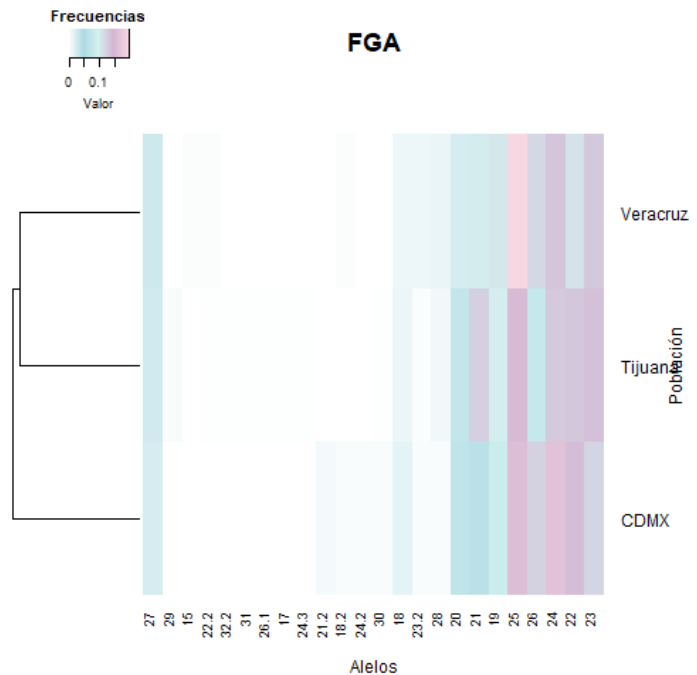
Heatmap 14. *Heatmap* generado para el *locus* D2S441 a partir de las frecuencias alélicas presentes en la Ciudad de México, la ciudad de Tijuana y el estado de Veracruz. En la parte superior se encuentra una escala de color con un valor entre 0 y 1 que indica las frecuencias observadas en las poblaciones.



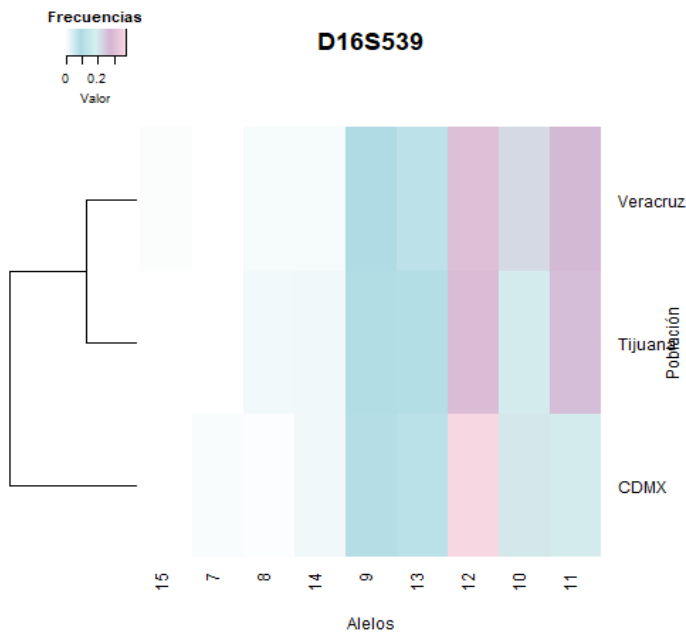
En el *locus* D18S51 (*Heatmap* 15) se observó el alelo 21 en la población de CDMX y Tijuana, pero no en la del estado de Veracruz. El alelo 9 se observó únicamente en población de la CDMX. Solo en la población del estado de Veracruz se observó el alelo 26. Únicamente en la población de ciudad de Tijuana se encontraron los alelos 10, 10.2, 16.1, 24 y 27.

Heatmap 15. *Heatmap* generado para el *locus* D18S51 a partir de las frecuencias alélicas presentes en la Ciudad de México, la ciudad de Tijuana y el estado de Veracruz. En la parte superior se encuentra una escala de color con un valor entre 0 y 1 que indica las frecuencias observadas en las poblaciones.

En el *locus* FGA (*Heatmap* 16) se observaron los alelos 21.2 y 30 en la población de CDMX y Tijuana, pero no en la del estado de Veracruz. El alelo 24.2 se observó únicamente en población de la CDMX; mientras que alelo 22.2 solo está presente en las poblaciones del estado de Veracruz y de la ciudad de Tijuana. Solo en la población del estado de Veracruz se observó el alelo 15. Únicamente en la población de ciudad de Tijuana se encontraron los alelos 17, 24.3, 26.1, 29, 31 y 32.2.



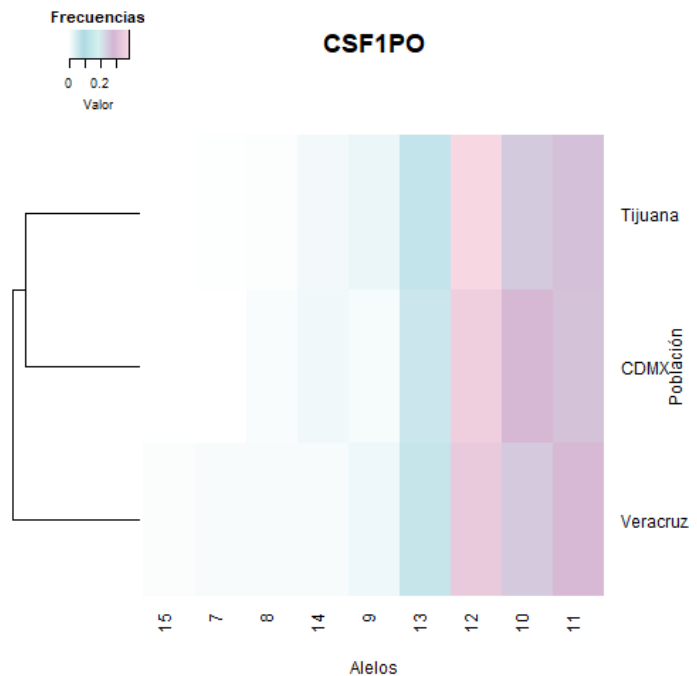
Heatmap 16. *Heatmap* generado para el *locus* FGA a partir de las frecuencias alélicas presentes en la Ciudad de México, la ciudad de Tijuana y el estado de Veracruz. En la parte superior se encuentra una escala de color con un valor entre 0 y 1 que indica las frecuencias observadas en las poblaciones.



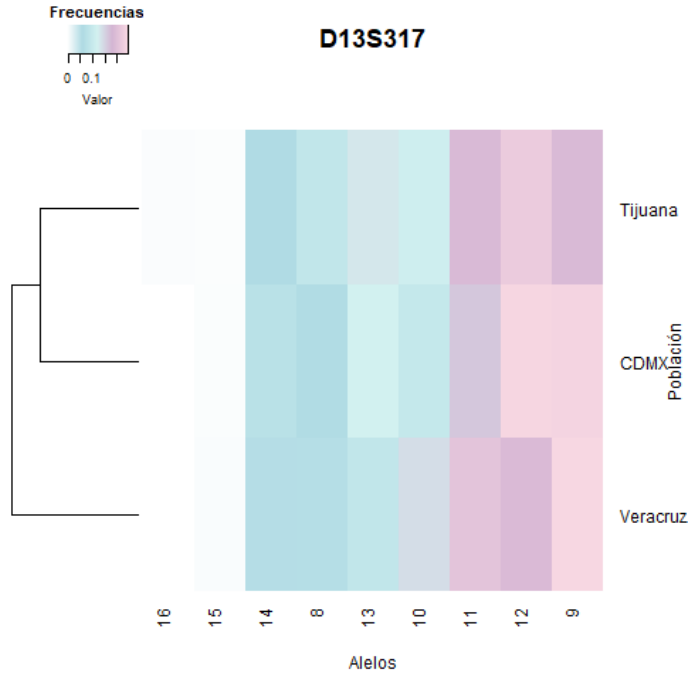
En el *locus* D16S539 (*Heatmap* 17) se observó el alelo 7 en población de la CDMX. Solo en la población del estado de Veracruz se observó el alelo 15.

Heatmap 17. *Heatmap* generado para el *locus* D16S539 a partir de las frecuencias alélicas presentes en la Ciudad de México, la ciudad de Tijuana y el estado de Veracruz. En la parte superior se encuentra una escala de color con un valor entre 0 y 1 que indica las frecuencias observadas en las poblaciones.

En el *locus* CSF1PO (*Heatmap* 18) el alelo 7 está presente en población del estado de Veracruz y de la ciudad de Tijuana. Solo en la población del estado de Veracruz se observó el alelo 15.



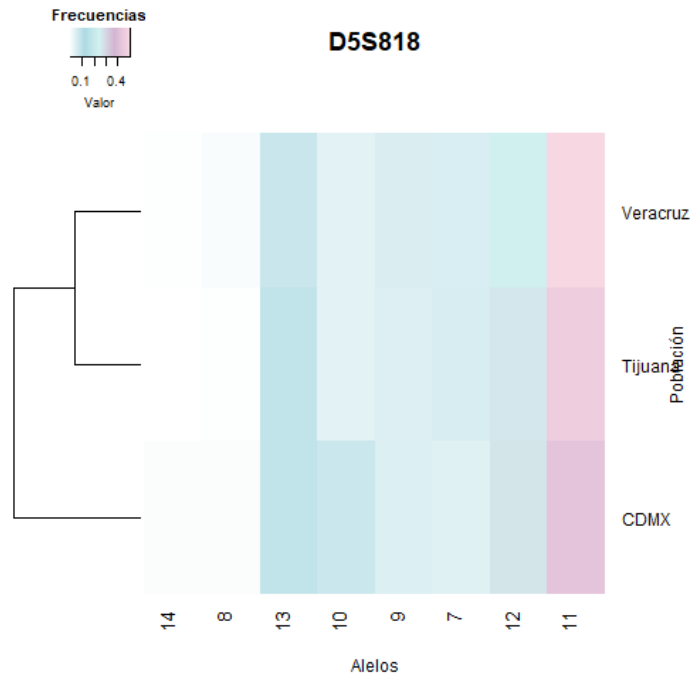
Heatmap 18. *Heatmap* generado para el *locus* CSF1PO a partir de las frecuencias alélicas presentes en la Ciudad de México, la ciudad de Tijuana y el estado de Veracruz. En la parte superior se encuentra una escala de color con un valor entre 0 y 1 que indica las frecuencias observadas en las poblaciones.



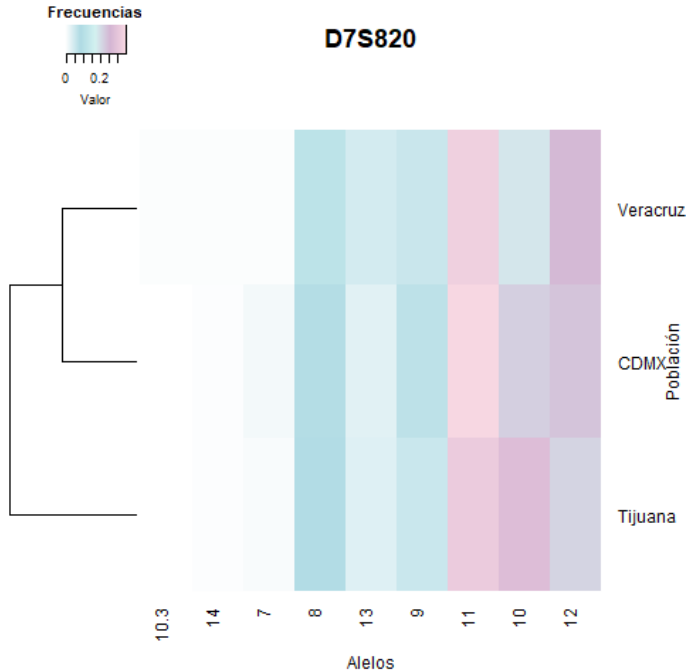
En el *locus* D13S317 (*Heatmap* 19) el alelo 16 únicamente se observó en la población de ciudad de Tijuana.

Heatmap 19. *Heatmap* generado para el *locus* D13S317 a partir de las frecuencias alélicas presentes en la Ciudad de México, la ciudad de Tijuana y el estado de Veracruz. En la parte superior se encuentra una escala de color con un valor entre 0 y 1 que indica las frecuencias observadas en las poblaciones.

En el *locus* D5S818 (*Heatmap* 20) se observaron los mismos alelos en las tres poblaciones, la mayor frecuencia en las tres poblaciones es en el alelo 11, mientras que los menos frecuentes son los alelos 8 y 14, también coincidiendo en las tres poblaciones.



Heatmap 20. *Heatmap* generado para el *locus* D5S818 a partir de las frecuencias alélicas presentes en la Ciudad de México, la ciudad de Tijuana y el estado de Veracruz. En la parte superior se encuentra una escala de color con un valor entre 0 y 1 que indica las frecuencias observadas en las poblaciones.



En el *locus* D7S820 (*Heatmap* 21) se observó el alelo 10.3 únicamente en la población del estado de Veracruz. El alelo 11 es el que se encuentra con mayor frecuencia en las tres poblaciones, mientras que el 14 es el menos frecuente.

Heatmap 21. *Heatmap* generado para el *locus* D7S820 a partir de las frecuencias alélicas presentes en la Ciudad de México, la ciudad de Tijuana y el estado de Veracruz. En la parte superior se encuentra una escala de color con un valor entre 0 y 1 que indica las frecuencias observadas en las poblaciones.

En los *loci* TH01 (*Heatmap* 1), D1S1656 (*Heatmap* 6), D2S1338 (*Heatmap* 13), D18S51 (*Heatmap* 15), CSF1PO (*Heatmap* 18) y D13S317 (*Heatmap* 19), Ciudad de México y Tijuana son más parecidos entre sí que con Veracruz.

La población de Veracruz y Ciudad de México tiene una mayor similitud en los *loci* TPOX (*Heatmap* 5), D10S1248 (*Heatmap* 9) y D7S820 (*Heatmap* 21).

Mientras que en los *loci* D3S1358 (*Heatmap* 2), vWA (*Heatmap* 3), D21S11 (*Heatmap* 4), D12S391 (*Heatmap* 7), SE33 (*Heatmap* 8), D22S1045 (*Heatmap* 10), D19S433 (*Heatmap* 11), D2S441 (*Heatmap* 14), FGA (*Heatmap* 16) y D5S818 (*Heatmap* 20) la Ciudad de México es diferente con ambas poblaciones.

CONCLUSIONES

De acuerdo con el análisis poblacional realizado, los datos obtenidos por el sistema 24plex QS pueden ser útiles para la investigación en genética de poblaciones, para la identificación humana y para pruebas de parentesco.

Se comprobó la funcionalidad del kit Investigator[®] 24plex QS de la marca Qiagen para la amplificación y electroforesis capilar aplicable para muestras con 2 ng, 4 ng u 8 ng de DNA utilizando el kit, escalado a $\frac{1}{3}$ de reacción.

Se estableció la primera base de datos poblacional de STRs autosómicos empleando el kit comercial Investigator[®] 24plex QS Qiagen en la Ciudad de México, utilizando una muestra de 156 individuos, para uso forense. La cuál estará disponible para su uso, bajo petición al laboratorio de genética del área de investigación de la LCF. En principio la base de datos creada es útil para el uso forense en la Ciudad de México, sin embargo, en futuros proyectos se podría incrementar el tamaño de muestra, así como la procedencia de estas para crear una base de datos con mayor representatividad del territorio mexicano.

Con los resultados obtenidos a partir de softwares y estimadores estadísticos especializados en genética de poblaciones, como lo son Arlequín, STRAF y diversas paqueterías del software Rstudio especializadas en análisis de STRs y análisis descriptivo; se realizó un análisis interpoblacional con estudios similares reportados en la literatura para distintas poblaciones de México y se observó lo siguiente:

- ✓ A pesar del tamaño de la muestra la base de datos cuenta con un poder de exclusión combinado de 0.99999656, similar a lo reportado en la literatura para muestras con más del doble de muestras, por lo tanto, se puede concluir que, si bien el tamaño de la muestra es importante para poder considerarla representativa de la población, la cantidad de marcadores analizados tiene un mayor peso estadístico que el tamaño de muestra utilizado.
- ✓ Las distancias genéticas encontradas entre las poblaciones de la Ciudad de México, del Estado de Veracruz y de la ciudad de Tijuana reafirman la necesidad de contar con bases de datos representativas de cada población

REFERENCIAS

- Aguilar Velazquez, J., & Rangel-Villalobos, H. (septiembre de 2021). Distribution of paternal lineages in Mestizo populations throughout Mexico: an in silico study based on Y-STR haplotypes. *Journal of human genetics*, 66, 327-332. Obtenido de <https://www.nature.com/articles/s10038-020-00824-7>
- Alberts, B., Johnson, A., & Lewis, J. (2002). *Molecular Biology of the Cell*. New York: Garland Science. Obtenido de <https://www.ncbi.nlm.nih.gov/books/NBK26821/>
- Bolivar, A. M., Rojas, A., & García-Lugo, P. (2014). PCR y PCR-Múltiple: parámetros críticos y protocolo de estandarización. *Avances en Biomedicina*, 3(1), 25-33. Obtenido de <https://Dialnet-PCRYPCRMultiple-4796903.pdf>
- Bruno, C., & Balzarini, M. (2009). Distancias genéticas entre perfiles moleculares obtenidos desde marcadores multilocus multialélicos. *FCA UNCuyo, Tomo XLI*(3), 43-54. Retrieved from https://www.researchgate.net/publication/286930825_Distancias_geneticas_entre_perfiles_moleculares_obtenidos_desde_marcadores_multilocus-multialelicos
- Buckleton, J., Taylor, D., & Bright, J. (2016). Parentage Testing. In *Forensic DNA evidence interpretation* (Second ed., pp. 353-395). doi:10.1201/b19680-12
- Budowle, B., Baechtel, F., Smerick, J., Presley, K., Giusti, A., Parsons, G., . . . Chakraborty, R. (1995). D1S80 population data in African. *J. Forensic Sci.*(40), 38-44.
- Butler, J. (2005). *Forensic DNA Typing Biology, Technology and Genetics of STR Markers* (Segunda ed.). Maryland, USA: Elsevier.
- Butler, J. (2005). *Forensic DNA Typing Biology, Technology and Genetics of STR Markers* (Segunda ed.). Maryland, USA: Elsevier.
- Butler, J. (2010). *Fundamental of Forensic DNA Typing*. England: Academic Press: Elsevier.
- Butler, J. (2014). *Advanced Topics in Forensic DNA Typing: Interpretation*. USA: Elsevier.
- Butler, J. (2014). Statistical Interpretation Overview. En J. Butler, *Advanced Topics in Forensic DNA Typing Interpretation*. Maryland, USA: Elsevier.
- Butler, J. M. (2011). *Advanced topics in Forensic DNA typing: methodology*. Maryland, Gaithersburg, USA: Elsevier.
- Consejo de Derechos Humanos de la ONU. (2009). *Resolución 10/26. Genética forense y derechos humanos*. Obtenido de https://ap.ohchr.org/documents/S/HRC/resolutions/A_HRC_RES_10_26.pdf

- De la Torre Pretell, J. C., Roe Battistini, C., & Roe Battistini, E. (2016). *Diagnóstico molecular de enfermedades infecciosas*. Santiago de Surco-Lima-Perú: Biblioteca Nacional de Perú.
- Demarchi, D. A. (2009). Microsatelites, distancias genéticas y estructura de poblaciones natias sudamericanas. *Revista Argentina de antropología biológica*, 11(1), 73-88. Obtenido de <https://revistas.unlp.edu.ar/raab/article/download/269/126/>
- DW. (17 de Mayo de 2022). México supera las 100.000 personas desaparecidas, según datos oficiales. Ciudad de México, México. Obtenido de <https://www.dw.com/es/m%C3%A9xico-supera-las-100000-personas-desaparecidas-seg%C3%BAAn-datos-oficiales/a-61819701#:~:text=El%20comit%C3%A9%20hizo%20un%20an%C3%A1lisis,pro%20medio%20de%2028%20personas%20diarias%20.>
- Giovambattista, G., Goszczynski, D., Fernandez, M., & Peral, P. (2015). Capítulo 4 Bases de datos en Genética Forense. En *Gnética Forense no humana* (págs. 102-114).
- González Chavarro, C., León Lozano, M., Morillo Coronado, A., Ochoa, E., & Morillo Coronado, Y. (2016). Caracterización molecular de palma de aceite *Elaeis guineensis* Jacq., procedente de diferentes orígenes (Zaire y Camerún) usando marcadores microsatélites. *Acta Agronómica*, 65(3), 279-283. doi:10.15446/acag.v65n3.48413
- Gonzalez-Herrera, L., Sosa-Escalante, J., López-González, P., López-González, M., Gamboa-Magaña, R., Herrera-Díaz, R., . . . Rangel-Villalobos, 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. Retrieved from <https://dimygen.com/forensic-statistical-parameters-of-22-autosomal-STRs-in-mestizos-from-the-peninsula-of-yucatan-mexico.pdf>
- Goodwin, W., Linacre, A., & Hadi, S. (2011). *An Introduction to Forensic Genetics* (Segunda ed.). Oxford, Reino Unido: John Wiley & Sons Ltd.
- Goudet, J., Jombart, T., Kamvar, Z., Archer, E., & Hardy, O. (2022, Mayo 03). *Estimation and Tests of Hierarchical F-Statistics*. Retrieved from <https://cran.r-project.org/web/packages/hierfstat/index.html>
- Gouy, A., & Zieger, M. (15 de julio de 2017). STRAF—A convenient online tool for STR data evaluation in forensic genetics. *Forensic Science International: Genetics*, 30, 148-151. doi:<http://dx.doi.org/10.1016/j.fsigen.2017.07.007>
- Guo, X., & Elston, R. (1999). Linkage Information Content of Polymorphic Genetic Markers. *Human Heredity*, 49, 112-118.
- Hernández-Rodríguez, A., & Trejo-Medinilla, 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. *Archivos de Medicina*, 10(11). Obtenido

de <https://www.archivosdemedicina.com/medicina-de-familia/estudio-gentico-poblacional-de-frecuencias-allicas-para-15-marcadores-str-presentes-en-la-poblacion-del-estado-de-zacatecas-aplicado-a-la-prctica-forense.pdf>

- Huston, K. (1998). *Statistical Analysis of STR Data*. Obtenido de Promega: https://www.promega.com/-/media/files/resources/profiles-in-dna/103/statistical-analysis-of-str-data.pdf?rev=0a1558d02bb74805947f121ddd1c5848&sc_lang=en
- Klug, W. S., Cummings, M. R., & Spencer, C. A. (2006). *Conceptos de Genética* (8 ed.). (J. Mensua, & D. I Torrens, Trads.) Marid: Pearson.
- LabGenetics. (2016, Octubre 14). *Laboratorio de Genética Clínica S.L*. Retrieved from <https://www.labgenetics.es/que-es-perfil-genetico/>
- Lareu, M. (2013). Short Tandem Repeats. En P. J. Jay A. Siegel (Ed.), *Encyclopedia of Forensic Sciences* (Segunda ed., págs. 219-226). Academic Press. Obtenido de <https://www.sciencedirect.com/science/article/pii/B9780123821652000404>
- León, I. (24 de Mayo de 2017). *¿Cuáles son las diferencias entre PCR, RT PCR, QPCR y RT QPCR?* Obtenido de All Science Ciencia, Tecnología y Ambiente: <https://www.e-allscience.com/blogs/articulos/cuales-son-las-diferencias-entre-pcr-rt-pcr-qpcr-y-rt-qpcr>
- Martínez de Pancorbo, M. (2012). *Bases de datos de frecuencias alélicas de marcadores microsatélites en población residente de Barcelona, Madrid y Sevilla*. . Obtenido de Plataforma en Red Banco Nacional de ADN Carlos III: <https://www.bancoadn.org/proyectos/bases-de-datos-de-frecuencias-alelicas-de-marcadores-microsatelites-en-poblacion-residente-de-barcelona-madrid-y-sevilla.html>
- Martínez-Cortés, G., Zuñiga-Chiquette, F., Celorio-Sanchez, A., Ruíz García, E., Antelo-Figueroa, A., Dalpozzo-Valenzuela, V., . . . Rangel-Villalobos, H. (2018). Population data for 21 autosomal STR loci (GlobalFiler kit) in two Mexican-Mestizo population from the northwest, Mexico. *International Journal of Legal Medicine*. doi:<https://doi.org/10.1007/s00414-018-1950-1>
- Melo, G., Uscanga, K., López-Armenta, M., Avilés-Villada, D., Henández-Romano, P., Rangel-Villalobos, H., . . . Denis, P. (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*. doi:10.1080/03014460.2022.2062050
- Méndez-Álvarez, S., & Pérez-Roth, E. (2004). La PCR múltiple en microbiología clínica. *Enfermedades infecciosas y microbiología clínica*, 22(3), 183-192. Retrieved from <https://www.elsevier.es/es-revista-enfermedades-infecciosas-microbiologia-clinica-28-articulo-la-pcr-multiple-microbiologia-clinica-13058027>
- Movimiento por nuestros desaparecidos en México. (2021). *La crisis ofrense en México: más de 52,000 personas fallecidas sin identificar*. Obtenido de

<https://movndmx.org/wp-content/uploads/2021/08/Informe-La-Crisis-Forense-en-Me%CC%81xico.pdf>

- National Human Genome Research Institute. (22 de Junio de 2022). *NIH*. Obtenido de <https://www.genome.gov/es/genetics-glossary/Electroforesis>
- Nei, M., Tajima, F., & Tatenó, Y. (1983). Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. *J Mol Evol*, 19(2), 153-170. doi: 10.1007/BF02300753
- NGHRI. (17 de Octubre de 2022). *Glosario: Diploide*. Obtenido de National Human Genome Research Institute: <https://www.genome.gov/es/genetics-glossary/Diploide#:~:text=Definici%C3%B3n,contienen%2023%20pares%20de%20cromosomas>.
- NGHRI. (17 de Octubre de 2022). *National Genome Human Research Institute*. Obtenido de Glosario: Intrón: <https://www.genome.gov/es/genetics-glossary/Intron>
- NHGRI. (22 de junio de 2022). *Glosario parlante de términos genómicos y genéticos*. Obtenido de National Human Genome Research Institute: <https://www.genome.gov/es/genetics-glossary>
- NHGRI. (05 de Octubre de 2022). *Recombinación homóloga*. Obtenido de National Human Genome Research Institute: <https://www.genome.gov/es/genetics-glossary/Recombinacion-homologa>
- Pereira, V., & Gusmão, L. (2013). X-Chromosome markers. En M. M. Houck (Ed.), *Forensic Biology*. Porto, Portugal: Elsevier.
- Qiagen. (2016). *Quick-Start Protocol. DNeasy Blood & Tissue kit*.
- Qiagen. (2018). *Investigator® Quantiplex® Pro RGQ Kit Handbook*.
- Qiagen. (2021). *Investigator® 24plex QS*. Obtenido de <https://www.qiagen.com/nl/resources/download.aspx?id=debe09ab-5483-478b-aeb3-e5c128e78a92&lang=en>
- Ramírez Flores, E., Saiz, M., Villegas-Carmona, D., Álvarez-Cubero, M. J., Álvarez, J. C., & Vega-Navarrete, L. (2014, Mayo 01). Genetic variation of 24 STR loci in a Mexican Mestizo population from Mexico D.F. *Forensic Science International: Genetics*, 10, E4-E6. Retrieved from [https://www.fsigenetics.com/article/S1872-4973\(14\)00027-1/fulltext](https://www.fsigenetics.com/article/S1872-4973(14)00027-1/fulltext)
- Rangel Villalobos, H. (2017). Las pruebas de ADN en el contexto forense II Interpretación estadística. *Revista de Ciencias Forenses de Honduras*, 3, 28-38.
- Rangel-Villalobos, H., Rubi-Castellanos, R., Morales-Vallejo, M., Molina-Araujo, V., Licea-Cadena, R., Rizzo-Juárez, S., . . . Muñoz-Valle, J. (2010, Julio 01). 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*, 4, 271-272. Retrieved from [https://www.fsigenetics.com/article/S1872-4973\(09\)00142-2/fulltext](https://www.fsigenetics.com/article/S1872-4973(09)00142-2/fulltext)
- Ritu, & Mohapatra, B. (2018). Genetic Variation. *Vonk J., Shackelford T. (eds) Encyclopedia of Animal Cognition and Behavior*. doi:https://doi.org/10.1007/978-3-319-47829-6_20-1
- Rodríguez Sánchez , I., & Barrera Saldaña, H. (2004). La reacción en cadena de la polimerasa a dos décadas de su invención. *Ciencia UANL*, VII(3). Obtenido de https://eprints.uanl.mx/1584/1/art_cadena.pdf
- Sierra Palacios, E., Serrato Díaz, A., Flores Rentería, L., & Aportela Cortez, J. (2014). *Herramientas moleculares aplicadas en ecología: aspectos teóricos y prácticos*. Obtenido de INECC: [https://www2.congreso.gob.pe/sicr/cendocbib/con4_uibd.nsf/770DBBBD5ADF759505257D4900580FE6/\\$FILE/HerramientasMolecularesAplicadasEcolog%C3%ADa.pdf](https://www2.congreso.gob.pe/sicr/cendocbib/con4_uibd.nsf/770DBBBD5ADF759505257D4900580FE6/$FILE/HerramientasMolecularesAplicadasEcolog%C3%ADa.pdf)
- Sokal, R., & Rohlf, F. (1979). *Biometría. Principios y métodos estadísticos en la investigación biológica*. Madrid: H. Blume Ediciones.
- Takezaki, N., & Nei, M. (1996). Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics*, 144(1), 389-399. doi:10.1093/genetics/144.1.389
- Thompson, T., & Black, S. (2007). *Forensic Human Identification*. Taylor & Francis Group, LLC.
- Universidad de la Habana. (2019). *Recombinación del DNA*. Obtenido de <http://www.fbio.uh.cu/sites/genmol/confs/conf5/index.htm>
- Vázquez Corona, S., & Guardado Estrada, M. (2021). El ADN mitocondrial: una alternativa útil para la identificación forense. *Identificación humana.mx*. Obtenido de [https://www.identificacionhumana.mx/el-adn-mitocondrial-una-alternativa-util-para-la-identificacion-forense/#:~:text=El%20ADN%20mitocondrial%20\(ADNmt\)%20es,de%20herencia%20materna%20\(4\)](https://www.identificacionhumana.mx/el-adn-mitocondrial-una-alternativa-util-para-la-identificacion-forense/#:~:text=El%20ADN%20mitocondrial%20(ADNmt)%20es,de%20herencia%20materna%20(4)).
- Villavicencio Queijeiro, A. (2022). *Guías para la Valoración Judicial de la Prueba Pericial al Sistema de Justicia Penal*. Ciudad de México. Obtenido de <https://dif.slp.gob.mx/wp-content/uploads/2022/07/Gui%CC%81as-Valoracio%CC%81n-Judicial-Prueba-Pericial.pdf>
- Warnes, G., Bolker, B., Bonebakker, L., Gentleman, R., Huber, W., Liaw, A., . . . Galili, T. (2022, Abril 25). *gplots: Various R Programming Tools for Plotting Data*. Retrieved from <https://cran.r-project.org/web/packages/gplots/index.html>

ANEXO 1

Consentimiento informado del proyecto “Análisis del genoma mitocondrial para la identificación forense en México”



FACULTAD DE MEDICINA
DIVISIÓN DE INVESTIGACIÓN
COMISIÓN DE ÉTICA

CONSENTIMIENTO INFORMADO PARA PARTICIPAR EN UN ESTUDIO DE INVESTIGACIÓN MÉDICA

Título del protocolo: Análisis del genoma mitocondrial para la identificación forense en México

Investigador principal: Dr. José Mariano Guardado Estrada

Sede donde se realizará el estudio: Licenciatura en Ciencia Forense Nombre del individuo:

Clave _____

A usted se le está invitando a participar en este estudio de investigación médica. Le pedimos que antes de decidir si participa o no, revise cada uno de los siguientes apartados y comprenda cada uno de ellos. Siéntase con absoluta libertad para preguntar sobre cualquier aspecto que le ayude a aclarar sus dudas al respecto. Este proceso se conoce como consentimiento informado. Una vez que haya comprendido el estudio y si usted desea participar, entonces se le pedirá que firme esta forma de consentimiento, de la cual se le entregará una copia firmada y fechada.

1. JUSTIFICACIÓN DEL ESTUDIO.

El objetivo principal del presente estudio se centra en el análisis del polimorfismo de ADN mitocondrial. El ADN mitocondrial es una molécula que tienen todas nuestras células y está directamente relacionado con la producción de energía a través de las mitocondrias. Este ADN mitocondrial se encuentra en muchas cantidades por cada célula y se hereda exclusivamente por vía materna. Esto quiere decir que nosotros heredamos el ADN mitocondrial de nuestras madres, el cual fue heredado por su madre y así sucesivamente. El análisis del ADN mitocondrial permite conocer un poco acerca de nuestra ancestría genética. En México la mayor parte de la gente tiene ADN mitocondrial proveniente de nuestros ancestros Amerindios. Por otra parte, el ADN mitocondrial también se analiza para realizar algunas identificaciones humanas. Esto es posible gracias a que se puede extraer de tejidos que son difíciles de analizar como hueso, diente o cabello sin raíz. La identificación del ADN mitocondrial se realiza analizando dos regiones que se conocen



como región hipervariable 1 y 2 a través de su secuenciación. Sin embargo, cuando se encuentran dos individuos con secuencias idénticas de estas regiones no indica que están relacionados biológicamente sino más bien que pertenecen al mismo linaje mitocondrial (por ejemplo hijos con madre, hijos con abuela materna o tíos maternos, etc). Aún así el análisis del ADN mitocondrial es útil en muchos casos de índole forense. Para que se pueda analizar adecuadamente el genoma mitocondrial se requiere una base de datos poblacional que incluya las secuencias mitocondriales de muchos individuos. Esto permitirá hacer estimaciones de que tan frecuente puede ser una secuencia de ADN mitocondrial en una población en caso que esta coincida con otra muestra analizada. Esto es útil pues con estos casos se pueden identificar personas desaparecidas o personas que hayan fallecido por desastres masivos. Otro de los objetivos de este estudio es encontrar otras regiones que nos ayuden a identificar mejor en el caso de personas que compartan la misma secuencia de ADN mitocondrial. Esto será útil pues encontrar dos secuencias iguales de ADN mitocondrial no siempre indicará que tengan un parentesco biológico.

2. OBJETIVO DEL ESTUDIO.

A usted se le está invitando a participar en un estudio de investigación que tiene como objetivos:

- Analizar el polimorfismo del genoma mitocondrial en un subgrupo de 100 individuos mestizo mexicanos empleando la tecnología de secuenciación masiva.
- Realizar una base de datos poblacional con los datos obtenidos útiles para la identificación forense.

3. BENEFICIOS DEL ESTUDIO.

El participar en este estudio no tendrá un beneficio directo para usted. Sin embargo, los resultados obtenidos del presente estudio permitirán en un futuro que diversos organismos o dependencias encargadas de realizar la identificación humana tengan mejores herramientas para llevar a cabo esta tarea.

4. PROCEDIMIENTOS DEL ESTUDIO.

En caso de aceptar participar en el estudio se le realizarán algunas preguntas sobre usted y sus antecedentes médicos. Una vez que responda al cuestionario, se le tomarán dos muestras de sangre, en dos tubos de color morado, especiales para la toma de muestras sanguíneas. La cantidad de sangre requerida total para el estudio es de 10 ml (lo equivalente a dos cucharadas). Para la toma de la muestra se requiere que la persona se



presente con un ayuno de al menos 8 horas. La muestra se tomará de la vena cefálica (porción anterior del codo) de cualquier brazo posterior a la limpieza con una torunda con alcohol. Para la toma de muestras sanguíneas se utilizarán agujas vacutainer estériles y tubos con anticoagulante (morados) también estériles. Toda muestra obtenida será etiquetada, transportada y procesada bajo los cuidados de cualquier muestra potencialmente infecto-contagiosa. Las muestras serán procesadas en el laboratorio de genética de la licenciatura en Ciencia Forense de la Facultad de Medicina en la UNAM. Hasta este momento concluye su participación en el estudio, aunque es probable que se le contacte en un futuro.

5. RIESGOS ASOCIADOS CON EL ESTUDIO.

El riesgo asociado al presente estudio es mínimo y pudiera ser debido a la toma de muestras sanguíneas. En algunos casos se pueden presentar ciertas molestias en el sitio de la toma de muestra como dolor o equimosis (moretón), las cuales ceden en una semana aproximadamente y no requieren manejo alguno. En caso de que usted desarrolle algún efecto adverso secundario o requiera otro tipo de atención, ésta se le brindará en los términos que siempre se le ha ofrecido y como se establece en la sección de aclaraciones.

6. ACLARACIONES

- Su decisión de participar en el estudio es completamente voluntaria.
- No habrá ninguna consecuencia desfavorable para usted, en caso de no aceptar la invitación.
- Si decide participar en el estudio puede retirarse en el momento que lo desee, aun cuando el investigador responsable no se lo solicite, pudiendo informar o no, las razones de su decisión, la cual será respetada en su integridad.
- No tendrá que hacer gasto alguno durante el estudio.
- No recibirá pago por su participación.
- En el transcurso del estudio usted podrá solicitar información actualizada sobre el mismo, al investigador responsable.
- La información obtenida en este estudio, utilizada para la identificación de cada paciente, será mantenida con estricta confidencialidad por el grupo de investigadores.
- En caso de que usted desarrolle algún efecto adverso secundario no previsto, tiene derecho a una indemnización, siempre que estos efectos sean consecuencia de su participación en el estudio.
- Usted también tiene acceso a las Comisiones de Investigación y de Ética de la Facultad de Medicina de la UNAM en caso de que tenga dudas sobre sus derechos como participante del estudio: Teléfono: 5623 2136
- Si considera que no hay dudas ni preguntas acerca de su participación, puede, si así lo desea, firmar la Carta de Consentimiento Informado que forma parte de este documento.



7. CARTA DE CONSENTIMIENTO INFORMADO.

Yo, _____ he leído y comprendido la información anterior y mis preguntas han sido respondidas de manera satisfactoria. He sido informado y entiendo que los datos obtenidos en el estudio pueden ser publicados o difundidos con fines científicos. Convengo en participar en este estudio de investigación. Recibiré una copia firmada y fechada de esta forma de consentimiento.

Firma del participante o del padre o tutor Fecha

Testigo 1 _____ Fecha

Testigo 2 _____ Fecha

He explicado al Sr(a). _____ la naturaleza y los propósitos de la investigación; le he explicado acerca de los riesgos y beneficios que implica su participación. He contestado a las preguntas en la medida de lo posible y he preguntado si tiene alguna duda. Acepto que he leído y conozco la normatividad correspondiente para realizar investigación con seres humanos y me apego a ella. Una vez concluida la sesión de preguntas y respuestas, se procedió a firmar el presente documento.

Firma del investigador Fecha



8. REVOCACIÓN DEL CONSENTIMIENTO

Título del protocolo: Análisis del genoma mitocondrial para la identificación forense en México

Investigador principal:

Sede donde se realizará el estudio:

Nombre del participante:

Por este conducto deseo informar mi decisión de retirarme de este protocolo de investigación por las siguientes razones: (Este apartado es opcional y puede dejarse en blanco si así lo desea el paciente)

Si el paciente así lo desea, podrá solicitar que le sea entregada toda la información que se haya recabado sobre él, con motivo de su participación en el presente estudio.

Firma del participante o Fecha

Testigo Fecha

Testigo Fecha c.c.p El paciente.



CONSENTIMIENTO PARA EL ALMACENAMIENTO Y USO FUTURO DE SU MUESTRA

Consentimiento Adicional al protocolo: Análisis del genoma mitocondrial para la identificación forense en México

Este documento de Consentimiento consta de dos partes; la información acerca de las muestras sin usar y el Formulario de Consentimiento para registrar su acuerdo. A usted se le dará una copia del documento completo del consentimiento.

INFORMACIÓN.

El estudio “Análisis del genoma mitocondrial para la identificación forense en México” busca establecer una base de datos poblacional que se útil para hacer identificación humana empleando el ADN mitocondrial. Para ello se secuenciará la región de control del ADN mitocondrial y en algunos caso es posible que se secuencie todo su genoma mitocondrial. Al momento de participar en el estudio, sus datos personales se mantendrán privados, además que las muestras que se obtengan, se almacenarán hasta su uso para el presente estudio. Dada la naturaleza de la muestra, es posible su uso futuro en trabajos investigación de este grupo de trabajo. El consentimiento que usted firmó nos permite emplear la muestra para el análisis previamente descrito, por lo que le solicitamos nos indique si está de acuerdo en que el remanente de su muestra se resguarde una vez que se hayan concluido las pruebas.

Actualmente, existe una gran cantidad de información acerca de la importancia del genoma (conjunto de genes del individuo). Existe, además del ADN mitocondrial otro tipo de variaciones genéticas, las cuales pueden ser estudiadas y analizadas para buscar nuevos marcadores de identificación o aquellos que estén relacionados con algún rasgo en particular. Los genes difieren entre las personas, y gracias a estas diferencias, se deben algunas variaciones físicas de la población (color de ojos, color de cabello y tipos de sangre).

Su muestra se emplearía únicamente para realizar estudios e investigaciones en genética, con el fin de identificar posibles biomarcadores de DNA, RNA y proteínas para otros proyectos. Sus muestras de sangre no se usarán para generar células madre ya sea permanentes o inmortales (células que han sido modificadas a fin de multiplicarlas ilimitadamente). Si usted no desea que su muestra se emplee para alguno de los fines anteriores o desea que no se use para nada puede señalarlo con libertad.

Algunos de sus datos médicos recabados en el estudio pueden ser relevantes para comprender mejor los resultados de los análisis genéticos. Para proteger su privacidad, su muestra se mantendrá anónima, ya que se identificará con un código. A este código estarán asociados también los datos biológicos y médicos de su muestra. Sin embargo, la información concerniente a sus datos personales (nombre, dirección y teléfono), pueden



estar o no disponibles. En caso que acepte que su muestra sea almacenada, indique si está de acuerdo también en almacenar sus datos personales. Los datos genéticos que resulten de este proyecto de Investigación no se convertirán en parte de sus registros médicos ni se entregarán a ningún otro médico, a su empleador, a su familia, ni a ninguna compañía de seguros. Es importante señalar que los datos generados de los datos genéticos y resultados en general de estos estudios son generados únicamente para propósito de investigación.

DERECHO A REHUSAR Y RECHAZAR

Usted puede revocar su consentimiento para participar en esta parte del Proyecto de Investigación en cualquier momento sin dar una razón. Esto no afectará su participación continua en el estudio o su tratamiento médico. Si usted decide retirar su participación en el estudio, por favor informe al Investigador. Después de su retiro, sus muestras serán destruidas y el material biológico relacionado a ellas, por lo tanto, no se obtendrán más datos genéticos. Los resultados que hayan sido generados en el Proyecto de Investigación antes de su retiro, pueden ser conservados y utilizados para propósitos del Proyecto de Investigación. Si tiene alguna duda, con respecto al uso y almacenamiento de su muestra, puede contactar al investigador principal:

Dr. José Mariano Guardado Estrada

Licenciatura en Ciencia Forense, Laboratorio de Genética.

Circuito de la Investigación Científica s/n UNAM, Ciudad de México

Tel 56 23 23 00 ext 81933



CONFIDENCIALIDAD DE SUS DATOS.

Para evitar el acceso no autorizado a sus muestras de sangre y a sus datos médicos, su información será codificada como se describió antes. Estas medidas fueron aprobadas por el comité de ética y la junta de revisión institucional correspondientes. Usted tiene el derecho a controlar el uso de sus datos personales de acuerdo con la Ley Federal de Protección de Datos Personales en Posesión de los Particulares. Además, usted tiene derecho a solicitar el acceso, corrección, cancelación/oposición de su información personal, y la solicitud será procesada de acuerdo con las regulaciones vigentes de protección de datos.

FORMULARIO DE CONSENTIMIENTO

En caso de que alguna muestra de sangre que he proporcionado para este proyecto de investigación queda sin usar o sobra, una vez que he completado el proyecto decido

(Marque una opción).

Deseo que mi muestra de sangre sea destruida de inmediato.

Deseo que mi muestra de sangre se destruya después de ____ años.

Autorizo a que mi muestra de sangre sea almacenada indefinidamente

Si decido que la muestra sea almacenada:

Autorizo a que mi muestra de sangre sea almacenada y se use en investigación futura, pero sólo con el mismo objetivo del proyecto de investigación actual: Análisis del genoma mitocondrial para la identificación forense en México.

Autorizo a que mi muestra de sangre sea almacenada y se use en investigaciones futuras, excepto en investigación sobre generación de células madre ya sea permanentes o inmortales (células que han sido modificadas a fin de multiplicarlas ilimitadamente).

He leído la información o me la han leído. He tenido la oportunidad de hacer preguntas acerca de ello y mis preguntas han sido respondidas satisfactoriamente. Consiento voluntariamente y entiendo que tengo el



derecho de retirar mi consentimiento sin que esto afecte el estudio de investigación actual o mi atención médica.

Deseo que mi identidad sea eliminada de mi muestra de sangre.

Deseo que mi identidad se mantenga con mi muestra de sangre

Nombre del Participante (con letra de imprenta):

Firma del Participante: _____

Fecha _____

(Día/mes/año)

En caso de que el paciente no sepa leer y escribir.

Un testigo que sepa leer y escribir debe firmar (si fuera posible, esta persona debe ser seleccionada por el participante y no debe tener ninguna conexión con el equipo de investigación). Los participantes analfabetos deben incluir también el impreso de su huella digital.

He presenciado la lectura exacta del documento de consentimiento al participante potencial, y el individuo ha tenido la oportunidad de hacer preguntas. Confirmando que el individuo ha dado el consentimiento libremente.

Nombre del Testigo (con letra de imprenta):

Firma del Testigo: _____

Fecha : _____

(Día/mes/año)

Nombre del Investigador (con letra de imprenta): _____

Firma del Investigador: _____

Fecha _____

(Día/mes/año)

ANEXO 2

Matrices por marcador

		Alelo							n	Frec.
		4	6	7	8	9	9.3	10		
Alelo	4		1						1	0.0032
	6		10	33	7	3	22		86	0.2756
	7			21	13	7	16	1	112	0.3590
	8				1	3	2		27	0.0865
	9					1	6		21	0.0673
	9.3						9		64	0.2051
	10								1	0.0032

Matriz 1. Matriz con la que se calcularon las frecuencias alélicas correspondientes al *locus* TH01

		Alelo							n	Frec.
		13	14	15	16	17	18	19		
Alelo	13			2					2	0.0064
	14		1	8	3		3		16	0.0513
	15			35	34	16	19	1	150	0.4808
	16				7	15	7		73	0.2340
	17					3	2		39	0.1250
	18								31	0.0994
	19								1	0.0032

Matriz 2. Matriz con la que se calcularon las frecuencias alélicas correspondientes al *locus* D3S1358.

		Alelo						n	Frec.	
		14	15	16	17	18	19			
Alelo	14		1	4	3	1	1	10	0.0321	
	15			11	5	5		22	0.0705	
	16				26	31	18	7	123	0.3942
	17					19	13	6	96	0.3077
	18						3	2	45	0.1442
	19								16	0.0513

Matriz 3. Matriz con la que se calcularon las frecuencias alélicas correspondientes al *locus* vWA

		Alelo														n	Frec.				
		24.2	26	27	28	29	29.2	30	30.2	31	31.2	32	32.2	33.1	33.2	34.2	35				
Alelo	24.2				1						1							2	0.0064		
	26										1							1	0.0032		
	27					1		2			2							5	0.0160		
	28				1	5		5		2	2		3		2			22	0.0705		
	29					10		18	2	1	9		6		9			71	0.2276		
	29.2							2										2	0.0064		
	30								11		8	5		9	1	3		75	0.2404		
	30.2											1		1		2		6	0.0192		
	31												4		2		2	19	0.0609		
	31.2													2		7	3	1	1	41	0.1314
	32															2		2	0.0064		
	32.2																5		40	0.1282	
	33.1																		1	0.0032	
	33.2																	1	23	0.0737	
	34.2																		1	0.0032	
35																			1	0.0032	

Matriz 4. Matriz con la que se calcularon las frecuencias alélicas correspondientes al locus D21S11.

		Alelo							n	Frec.			
		6	7	8	9	10	11	12	13				
Alelo	6			3						3	0.0096		
	7				1					1	0.0032		
	8					30	6	9	45	28	1	153	0.4904
	9						1	2	4			14	0.0449
	10							1	4	3		20	0.0641
	11								6	11		76	0.2436
	12									1		44	0.1410
	13											1	0.0032

Matriz 5. Matriz con la que se calcularon las frecuencias alélicas correspondientes al locus TPOX.

		Alelo													n	Frec.			
		11	12	13	14	14.3	15	15.3	16	16.3	17	17.3	18	18.3			19.3		
Alelo	11		1	1	1				3	1	2	1				10	0.0321		
	12			1	5	4		2	1	4	1		4			3	27	0.0865	
	13				2	2	1	8	1	1		2	6		2		33	0.1058	
	14					1		4	1	11	2	2	6				35	0.1122	
	14.3														1		2	0.0064	
	15							6	2	6	1	2	7	1	3		48	0.1538	
	15.3										1		1				7	0.0224	
	16										7	2	5	6	1		54	0.1731	
	16.3												2	1			10	0.0321	
	17												2	4			23	0.0737	
	17.3													1		5	1	44	0.1410
	18																	2	0.0064
	18.3															1		13	0.0417
	19.3																	4	0.0128

Matriz 6. Matriz con la que se calcularon las frecuencias alélicas correspondientes al locus D1S1656.

		Alelo													n	Frec.		
		15	16	17	17.3	18	18.3	19	20	21	22	23	24	25			28	
Alelo	15					1		2	1		1					5	0.0160	
	16							1	1		2	1				5	0.0160	
	17				1	1	1		5	3	1		1			14	0.0449	
	17.3								3							4	0.0128	
	18					4			17	14	4	3	1	2	1	52	0.1667	
	18.3									2						2	0.0064	
	19							15	20	9	3	3	1			94	0.3013	
	20								7	6	1	1	3	2		68	0.2179	
	21									2	2	1	1			28	0.0897	
	22												1	1		14	0.0449	
	23													2		1	11	0.0353
	24																10	0.0321
	25																4	0.0128
	28																1	0.0032

Matriz 7. Matriz con la que se calcularon las frecuencias alélicas correspondientes al locus D12S391.

		Alelo																												n	Frec.	
		13	13.3	14	15	16	16.2	17	17.3	18	19	20	20.2	21	21.2	22	22.2	23.2	24.2	25.2	26	26.2	27.2	28.2	29.2	30.2	31.2	32.2	34			
Alelo	13							1																						1	0.0032	
	13.3									1																					1	0.0032
	14			1		1									1										1						5	0.0160
	15				1			1	1		1				2					1		6	1		1	1			1	16	0.0513	
	16					1				1	1													2		4	3			13	0.0417	
	16.2						1																		1					1	0.0032	
	17							1			3				1	2						5	2	4	4	1	3			30	0.0962	
	17.3											1																		1	0.0032	
	18									1	2	1				1					2	2	3	1	5	3	1		1	29	0.0929	
	19										2	2				1				1	3	2	4	3	2	1			1	27	0.0865	
	20											2	1	1				1										1	1	13	0.0417	
	20.2												1									1								2	0.0064	
	21																							1						3	0.0096	
	21.2																													6	0.0192	
	22																								1					2	0.0064	
	22.2																													1	0.0032	
	23.2																					1		1	2					5	0.0160	
	24.2																				2		1	1	1		1			5	0.0160	
	25.2																				1		1	2		2				15	0.0481	
	26																					1								1	0.0032	
	26.2																					1	3	1	1	1	1	2		31	0.0994	
	27.2																						1	4	1	1				27	0.0865	
	28.2																								1			1		20	0.0641	
	29.2																								1	2	1			28	0.0897	
	30.2																									1				16	0.0513	
	31.2																													9	0.0288	
	32.2																													3	0.0096	
	34																													1	0.0032	

Matriz 8. Matriz con la que se calcularon las frecuencias alélicas correspondientes al locus SE33.

		Alelo							n	Frec.		
		10	11	12	13	14	15	16			17	
Alelo	10							1			1	0.0032
	11								1		1	0.0032
	12					2	3		1	1	7	0.0224
	13				8	26	22		7		71	0.2276
	14					26	33		8		121	0.3878
	15						9		6	1	83	0.2660
	16										24	0.0769
	17										4	0.0128

Matriz 9. Matriz con la que se calcularon las frecuencias alélicas correspondientes al locus D10S1248.

		Alelo										n	Frec.	
		10	11	12	13	14	15	16	17	18	19			
Alelo	10						1						1	0.0032
	11						9	11					20	0.0641
	12							2					2	0.0064
	13						1	1					2	0.0064
	14						2	2					4	0.0128
	15						31	57	9			1	142	0.4551
	16							15	11			1	115	0.3686
	17								1	1			23	0.0737
	18												1	0.0032
	19												2	0.0064

Matriz 10. Matriz con la que se calcularon las frecuencias alélicas correspondientes al locus D22S1045

		Alelo											n	Frec.	
		11.2	12	12.2	13	13.2	14	14.2	15	15.2	16	16.2			17
Alelo	11.2					1	1							2	0.0064
	12				3	3	2	2			2			12	0.0385
	12.2								1					1	0.0032
	13				6	6	17	3	9	4	6	3		63	0.2019
	13.2					2	6		8	4	1	1		34	0.1090
	14						15	4	12	9	3	1		85	0.2724
	14.2								4	1				14	0.0449
	15								4	5	2	1	2	52	0.1667
	15.2									1	1			28	0.0897
	16													13	0.0417
	16.2													6	0.0192
	17													2	0.0064

Matriz 11. Matriz con la que se calcularon las frecuencias alélicas correspondientes al locus D19S433.

		Alelo										n	Frec.	
		8	9	10	11	12	13	14	15	16	17			
Alelo	8				1			2					3	0.0096
	9						1						1	0.0032
	10				2	1	4	8	6	6	2		31	0.0994
	11					1	4	5	1	1			14	0.0449
	12						2	15	9	3	1		37	0.1186
	13							22	20	7	1	1	101	0.3237
	14								17	9			85	0.2724
	15									2	2		32	0.1026
	16												7	0.0224
	17												1	0.0032

Matriz 12. Matriz con la que se calcularon las frecuencias alélicas correspondientes al locus D8S1179.

		Alelo												n	Frec.
		16	17	18	19	20	21	22	23	24	25	26	27		
Alelo	16		4	1	1	1	1			1				9	0.0288
	17			7	5	5	2	3	8	3	1		1	46	0.1474
	18				7	1	1	3	3					16	0.0513
	19					4	10	2	4	19	3	3		62	0.1987
	20						4		7	11	3	2		48	0.1538
	21							1	1				1	9	0.0288
	22								1	4		2		26	0.0833
	23									4	2	5		61	0.1955
	24										1	2		16	0.0513
	25											1		17	0.0545
	26													1	0.0032
	27													1	0.0032

Matriz 13. Matriz con la que se calcularon las frecuencias alélicas correspondientes al locus D2S1338.

		Alelo							n	Frec.		
		10	11	11.3	12	13	14	15			16	
Alelo	10		28	42	3	3	4	17	1		126	0.4038
	11			22	1	5	1	17			110	0.3526
	11.3								1		5	0.0160
	12							1		1	10	0.0321
	13										5	0.0160
	14							5	4		49	0.1571
	15										6	0.0192
	16										1	0.0032

Matriz 14. Matriz con la que se calcularon las frecuencias alélicas correspondientes al locus D2S441.

		Alelo														n	Frec.		
		9	11	12	13	14	15	16	17	18	19	20	21	22	23			25	
Alelo	9							1										1	0.0032
	11						1		3									4	0.0128
	12				1	4	3	3	3	5		2	4		1	1		28	0.0897
	13				1	8	9	8	6	3	1		2	1				44	0.1410
	14					3	8	3	7	3	1	2	2				1	44	0.1410
	15						2	5	10	8	1	1		1				51	0.1635
	16							2	6	2	1	1				1		35	0.1122
	17								5	1	3	1						52	0.1667
	18									1	2							21	0.0673
	19																	11	0.0353
	20													1				10	0.0321
	21																	5	0.0160
	22																	3	0.0096
	23																	2	0.0064
25																	1	0.0032	

Matriz 15. Matriz con la que se calcularon las frecuencias alélicas correspondientes al locus D18S51.

		Alelo														n	Frec.			
		18	18.2	19	20	21	21.2	22	23	23.2	24	24.2	25	26	27			28	30	
Alelo	18				1			1			2		1					5	0.0160	
	18.2				1													1	0.0032	
	19				2	1	2		6	3		3		4	4			27	0.0865	
	20				1	2			3	3		3		3	2	1		22	0.0705	
	21								2	4		5		2	2	1		20	0.0641	
	21.2											1						2	0.0064	
	22								3	6		9		7	2	1		44	0.1410	
	23									4		5		3	3	2		1	38	0.1218
	23.2													1					1	0.0032
	24												2		14	6			52	0.1667
	24.2															1			1	0.0032
	25													4	7				50	0.1603
	26														4	3	1		39	0.1250
	27																		8	0.0256
28																		1	0.0032	
30																		1	0.0032	

Matriz 16. Matriz con la que se calcularon las frecuencias alélicas correspondientes al locus FGA.

		Alelo								n	Frec.	
		7	8	9	10	11	12	13	14			
Alelo	7					1	1				2	0.0064
	8					1					1	0.0032
	9				1	7	3	11	3	1	27	0.0865
	10				8	10	23	7			63	0.2019
	11					8	23	7			61	0.1955
	12						19	17	3		116	0.3718
	13							1	1		37	0.1186
	14										5	0.0160

Matriz 17. Matriz con la que se calcularon las frecuencias alélicas correspondientes al *locus* D16S539.

		Alelo							n	Frec.	
		8	9	10	11	12	13	14			
Alelo	8				1	1				2	0.0064
	9				1	1	1			3	0.0096
	10				15	19	35	5	1	90	0.2885
	11				8	35	7	3		82	0.2628
	12					17	4	1		111	0.3558
	13						1			19	0.0609
	14									5	0.0160

Matriz 18. Matriz con la que se calcularon las frecuencias alélicas correspondientes al *locus* CSF1PO.

		Alelo								n	Frec.	
		8	9	10	11	12	13	14	15			
Alelo	8										20	0.0641
	9										74	0.2372
	10										30	0.0962
	11										51	0.1635
	12										75	0.2404
	13										37	0.1186
	14										24	0.0769
	15										1	0.0032

Matriz 19. Matriz con la que se calcularon las frecuencias alélicas correspondientes al *locus* D13S317.

		Alelo								n	Frec.	
		7	8	9	10	11	12	13	14			
Alelo	7				2	10	2				15	0.0481
	8						2				2	0.0064
	9				1	1	8	3	3		17	0.0545
	10					2	14	4	1		26	0.0833
	11						25	34	17	2	134	0.4295
	12							17	7		86	0.2756
	13								1		30	0.0962
	14										2	0.0064

Matriz 20. Matriz con la que se calcularon las frecuencias alélicas correspondientes al locus D5S818.

		Alelo								n	Frec.	
		7	8	9	10	11	12	13	14			
Alelo	7					2	2				4	0.0128
	8				4	11	6				25	0.0801
	9				2	12	4	2			22	0.0705
	10					8	23	10	3	1	69	0.2212
	11						17	29	4		107	0.3429
	12							11	3		74	0.2372
	13										10	0.0321
	14										1	0.0032

Matriz 21. Matriz con la que se calcularon las frecuencias alélicas correspondientes al locus D7S820.