



**UNIVERSIDAD CENTRAL  
DE VENEZUELA**

**FACULTAD DE CIENCIAS**

**ESCUELA DE BIOLOGÍA**

**ANÁLISIS DE LA DISTRIBUCIÓN ALÉLICA DE 23  
MICROSATÉLITES DE IMPORTANCIA FORENSE EN  
LA POBLACIÓN VENEZOLANA**

**TRABAJO ESPECIAL DE GRADO**

Presentado ante la Ilustre Universidad Central de Venezuela, por la Br. **Yudis Alejandra Franco Semprun** como requisito parcial para optar al título de Licenciada en Biología

Tutor: Lic. Jorge Castro

Asesor Académico: Dra. Roxana Gajardo

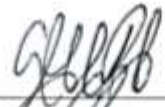
CARACAS, VENEZUELA

FEBRERO - 2018

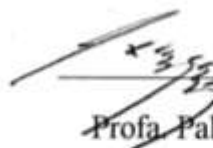
**DEL EXAMEN PÚBLICO Y SOLEMNE DEL TRABAJO ESPECIAL  
DE GRADO DE LA Br. YUDIS ALEJANDRA FRANCO SEMPRUN**

Quienes suscribimos, miembros del jurado evaluador designado por el Consejo de Escuela de Biología de la Facultad de Ciencias de la Universidad Central de Venezuela para examinar el Trabajo Especial de Grado de la Br. Yudis Alejandra Franco Semprun, C.I: 23.691.727, titulado “ANÁLISIS DE LA DISTRIBUCIÓN ALÉLICA DE 23 MICROSATÉLITES DE IMPORTANCIA FORENSE EN LA POBLACIÓN VENEZOLANA”, para optar al título de Licenciada en Biología, considerando que dicho trabajo cumple con los requisitos exigidos en los reglamentos respectivos lo consideramos **APROBADO**.

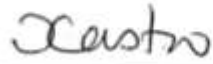
Para dar fe de ello, se levanta la presente acta en Caracas, a los 20 días del mes de febrero del año 2018, dejando constar que la Profa. Roxana Gajardo actuó como coordinadora del jurado examinador.



Prof. Giovanni Angiolillo



Profa. Palmira Guevara



Lic. Jorge Castro  
(Tutor 1)



Profa. Roxana Gajardo  
(Tutora 2- Coordinadora)



Universidad Central de Venezuela  
Facultad de Ciencias  
Escuela de Biología  
Dirección

### CONSTANCIA

El Consejo de la Escuela de Biología, Facultad de Ciencias de la Universidad Central de Venezuela, certifica que en su sesión del día 21/03/2018, acordó otorgar la **Mención Honorífica** al Trabajo Especial de Grado de la bachiller **Franco S., Yudis A.**, titular de la cédula de identidad No. 23.691.727, titulado: **"Análisis de la distribución alélica de 23 microsateélites de importancia forense en la población venezolana"** considerando la originalidad, independencia y creatividad en la elaboración del mismo, así como la sobresaliente calidad del trabajo escrito y la presentación oral.

Constancia que se expide a petición de la parte interesada a los veintitrés días del mes de marzo del año dos mil dieciocho.

  
Prof. Fernando González  
Director (Encargado)

/br.-

## **AGRADECIMIENTOS**

Son muchos a los que les quiero dedicar este increíble momento de mi vida, pero especialmente a mis padres Marcel Franco y Yudis Semprun. Nada de esto hubiera sido posible sin su apoyo, ayuda, consejos, cuidados y amor incondicional. Gracias a ustedes soy quién soy y he podido llegar hasta este momento, el cual estoy muy emocionada y feliz de haber alcanzado. Mis mejores amigos y mis guías, las mejores personas que conozco, este logro es para ustedes. Quiero agradecerles por todo el apoyo brindado, no solo a lo largo de mi carrera sino de mi vida. Gracias por siempre darnos a mi hermano y a mí su amor, apoyo y confianza, en palabras no podría expresarles la admiración que siento hacia ustedes, gracias por todo. A mi hermano, que siempre ha estado presente, gracias Marcel Eduardo por tu apoyo.

Agradezco al Lcdo. Jorge Castro y al Dr. Oscar Feo por la orientación recibida, también por su paciencia y disponibilidad en todos los momentos en que necesité. Infinitamente agradecida por su inestimable ayuda, ánimo y amistad, y por acogerme en su tutoría y haberme permitido realizar este trabajo; asimismo, quisiera agradecerles por su motivación recibida en todos los momentos. Al igual que a los compañeros de laboratorio que siempre estuvieron avocados en ayudarme, María Gabriela Rojas, Marjorie Sayegh y Magaly Salazar, gracias por el cariño con el que me acogieron.

A la Dr. Yarimar Ruiz, a quien admiro y respeto. Gracias por brindarme la oportunidad de realizar esta tesis y de seguir aprendiendo, por tu apoyo y sobre todo por la confianza brindada. Por abrirme las puertas del laboratorio y proporcionarme la ayuda necesaria para culminar este estudio. Al Ministerio Público por darme la oportunidad de estar en el laboratorio de genética forense y proporcionarme los medios para que esta tesis se pudiera llevar a cabo.

Un gran agradecimiento a la Dra. Roxana Gajardo, quién me enseñó y me ayudó. Es una increíble persona y un excelente modelo a seguir, quién me orientó durante todo este camino para cumplir este requisito, de verdad, gracias. A la Dra. Guillermina Alonso,

quién me orientó por este camino de la genética, gracias por ser una profesora tan excelente y digna de admiración, por inculcarme el interés hacia la “molécula perfecta” y por sus sabios consejos. Al igual que a la profesora Victoria Navas, quién despertó un interés en mí que no sabía que tenía, sobre la genética, gracias por enseñarme con paciencia este mundo.

A mi compañero, Daniel Romero, que desde el inicio de la carrera me brindó su ayuda y su apoyo siempre que lo necesité.

A todos aquellos que, aunque no nombrados, me brindaron con sus inestimables ayudas en distintos momentos, mi reconocimiento y afectivo agradecimiento.

Y, sobre todo, a la UCV, la casa que vence las sombras, estoy agradecida de por vida por esta increíble experiencia que comenzó hace unos años atrás.

## ÍNDICE DE CONTENIDO

<b>ÍNDICE DE FIGURAS .....</b>	<b>IV</b>
<b>ÍNDICE DE TABLAS.....</b>	<b>V</b>
<b>ABREVIATURAS.....</b>	<b>VI</b>
<b>RESUMEN .....</b>	<b>VII</b>
<b>INTRODUCCIÓN .....</b>	<b>1</b>
1. EL ADN .....	4
1.1. Componentes y estructura del ADN .....	5
1.2. Organización del ADN .....	8
1.3. Alelos y polimorfismos .....	10
1.4. Polimorfismos de longitud .....	12
1.5. Características deseables de STR utilizados en la tipificación de ADN .....	22
2. IMPORTANCIA DEL ANÁLISIS ESTADÍSTICO DE LOS STR .....	23
2.1. Estudios de paternidad .....	23
2.2. Criminalística .....	23
2.3. Personas desaparecidas .....	25
3. BASES DE DATOS GENÉTICAS .....	26
4. GENÉTICA DE POBLACIONES.....	28
4.1. Frecuencias alélicas y genotípicas .....	30
4.2. Ley de equilibrio de Hardy-Weinberg .....	30

4.3.	Poder de exclusión (PE).....	31
4.4.	Poder de discriminación (PD).....	32
4.5.	Probabilidad de coincidencia aleatoria (Pi).....	32
4.6.	Heterocigosidad (H).....	33
4.7.	Índice de contenido polimórfico (PIC).....	33
5.	LOCOS DE ESTUDIO .....	34
6.	PLANTEAMIENTO DEL PROBLEMA .....	35
<b>ANTECEDENTES .....</b>		<b>37</b>
1.	ESTUDIOS A NIVEL INTERNACIONAL .....	37
2.	ESTUDIOS A NIVEL NACIONAL.....	42
<b>OBJETIVOS.....</b>		<b>45</b>
1.	OBJETIVO GENERAL .....	45
2.	OBJETIVOS ESPECÍFICOS.....	45
<b>MATERIALES Y MÉTODOS.....</b>		<b>46</b>
1.	MUESTRA POBLACIONAL .....	46
2.	CRITERIOS DE INCLUSIÓN.....	49
3.	MUESTRAS BIOLÓGICAS.....	49
4.	ANÁLISIS DE LABORATORIO.....	50
4.1.	Amplificación vía PCR-múltiplex.....	50
4.2.	Detección y asignación de alelos STR .....	53

4.3. Análisis de resultados.....	57
5. ANÁLISIS ESTADÍSTICOS.....	58
6. CONSENTIMIENTO INFORMADO. ....	60
7. COMPARACIÓN DE LA FRECUENCIA ALÉLICA ENTRE LA POBLACIÓN VENEZOLANA EN ESTUDIO Y OTRAS POBLACIONES.....	61
<b>RESULTADOS.....</b>	<b>63</b>
1. FRECUENCIAS ALÉLICAS DE LOS 23 LOCI STR Y PARÁMETROS ESTADÍSTICOS DE LA POBLACIÓN VENEZOLANA .....	63
2. MARCADORES STR ANALIZADOS EN LA POBLACIÓN VENEZOLANA.....	73
<b>DISCUSIÓN DE RESULTADOS.....</b>	<b>76</b>
<b>CONCLUSIONES .....</b>	<b>92</b>
<b>BIBLIOGRAFÍA .....</b>	<b>94</b>
<b>ANEXOS .....</b>	<b>102</b>



## ÍNDICE DE FIGURAS

<b>Figura 1.</b> Estructura química de las bases nitrogenadas .....	6
<b>Figura 2.</b> Modelo de doble hélice propuesto por Watson y Crick.....	7
<b>Figura 3.</b> Genoma Humano .....	9
<b>Figura 4.</b> Organización del ADN total humano, según su estructura y función.....	10
<b>Figura 5.</b> Caracterización del polimorfismo de secuencia y longitud .....	12
<b>Figura 6.</b> Caracterización del ADN repetido en tándem .....	14
<b>Figura 7.</b> Separación y detección de fragmentos de ADN por electroforesis capilar con un analizador genético ABI 3130xl .....	21
<b>Figura 8.</b> Configuración del sistema PowerPlex 16 HS .....	34
<b>Figura 9.</b> Configuración del sistema PowerPlex ESX 17 .....	35
<b>Figura 10.</b> Electroferograma de una muestra utilizando el sistema <i>PowerPlex 16 HS</i> ..	64
<b>Figura 11.</b> Heterocigosidad observada de los marcadores STR estudiados en la población venezolana.....	69
<b>Figura 12.</b> Probabilidad de coincidencia aleatoria de los marcadores STR estudiados en la población venezolana .....	70
<b>Figura 13.</b> Poder de discriminación de los marcadores STR estudiados en la población venezolana .....	71
<b>Figura 14.</b> Índice de contenido polimórfico de los marcadores STR estudiados en la población venezolana.....	72
<b>Figura 15.</b> Poder de exclusión de los marcadores STR estudiados en la población venezolana .....	73

## ÍNDICE DE TABLAS

<b>Tabla 1:</b> 24 loci de estudio incluidos en estuches comerciales <i>PowerPlex® ESX 17</i> y <i>PowerPlex® 16 HS</i> .....	46
<b>Tabla 2.</b> Características de los STR que incluye el sistema <i>PowerPlex® 16 HS</i> .....	47
<b>Tabla 3.</b> Características de los STR que incluye el sistema <i>PowerPlex® ESX 17</i> .....	48
<b>Tabla 4:</b> Composición de la mezcla de reacción de PCR para una amplificación directa a partir de discos de tarjetas FTA® .....	51
<b>Tabla 5:</b> Protocolo del termociclador para el estuche comercial <i>PowerPlex® ESX 17</i> ...	52
<b>Tabla 6:</b> Protocolo del termociclador para el estuche comercial <i>PowerPlex® 16 HS</i> .....	53
<b>Tabla 7.</b> Composición de la mezcla de carga de las muestras en placa en el analizador genético <i>3130xl</i> .....	54
<b>Tabla 8.</b> Condiciones de corrida electroforética en el analizador genético <i>3130xl</i> de <i>Applied Biosystems</i> .....	55
<b>Tabla 9.</b> Frecuencias alélicas obtenidas de los marcadores STR evaluados en la población venezolana.....	65
<b>Tabla 10.</b> Cálculos y resultados para evaluar el equilibrio Hardy-Weinberg entre los 23 loci estudiados en la población venezolana. ....	67
<b>Tabla 11.</b> Parámetros estadísticos de interés forense para los 23 loci estudiados en la población venezolana.....	68
<b>Tabla 12.</b> Número total de alelos observados y alelos más y menos frecuentes para cada locus de estudio en la población venezolana .....	74

## ABREVIATURAS

ADN: Ácido desoxirribonucleico.

CCD: *Charge-coupled device* (dispositivo de carga acoplada).

CODIS: *Combined DNA index system* (Sistema de referencia combinado de ADN).

H<sub>esp</sub>: Heterocigosidad esperada.

H<sub>obs</sub>: Heterocigosidad observada.

pb: Pares de bases.

PCR: *Polymerase chain reaction* (reacción en cadena de la polimerasa).

PD: Poder de discriminación.

PE: Poder de exclusión.

PIC: Contenido de información polimórfica.

Pi: Probabilidad de coincidencia aleatoria.

STR: *Short tandem repeats* (repeticiones cortas en tándem).

RFU: *Relative fluorescence unit* (Unidad de Fluorescencia Relativa).

VNTR: *Variable number of tandem repeat* (número variable de repeticiones en tándem).

## RESUMEN

El avance tecnológico y científico en la definición e identificación de los marcadores genéticos moleculares ha permitido el desarrollo de la genética forense, aplicada en la identificación de personas y restos cadavéricos, el análisis de restos biológicos, la determinación de relaciones de filiación y determinación de las variabilidades genéticas de la población. Los microsatélites o *short tandem repeats* son los marcadores de elección en identificación humana, debido a su elevado polimorfismo, con tasas de mutaciones relativamente bajas y su tamaño pequeño, que permite optimizar su amplificación. Debido a que se encuentran ampliamente distribuidos en el genoma, son la principal herramienta de la genética forense porque permiten trabajar con poca cantidad de ADN, incluso degradado, son manejables para la automatización, altamente discriminatorios y ofrecen resultados fáciles de manejar y comparar. Con el propósito de medir la eficacia forense de los marcadores genéticos analizados es necesario conocer una serie de parámetros estadísticos con los cuales se sustentan los cálculos de probabilidades en los peritajes forenses. El objetivo del presente estudio fue analizar la estructura genética de una muestra poblacional venezolana utilizando 23 loci STR y construir una base de datos poblacional que pueda ser empleada en cálculos estadísticos para identificación humana y pruebas de paternidad. De esta manera, el estudio involucró 134 individuos no emparentados residentes de la región capital de la República Bolivariana de Venezuela con los sistemas *PowerPlex ESX 17* y *PowerPlex 16 HS* de la casa comercial Promega, los cuales co-amplifican, en una estrategia de PCR multiplex, 17 y 16 loci STR, respectivamente. Los datos genéticos fueron obtenidos

mediante el análisis de los productos de PCR por electroforesis capilar, y fueron analizados con diversos software y programas estadísticos, para ser comparados con una población de referencia. Se determinaron las frecuencias alélicas, y la distribución de genotipos de todos los marcadores, los cuales estuvieron en equilibrio Hardy-Weinberg. Se estimaron los siguientes parámetros de interés forense: poder de discriminación (PD), poder de exclusión (PE), heterocigosidad observada ( $H_{obs}$ ), heterocigosidad esperada ( $H_{esp}$ ), contenido de información polimórfica (PIC) y probabilidad de coincidencia aleatoria ( $P_i$ ). Nuestros resultados analizando 23 STR en la población venezolana muestran por primera vez, que los marcadores Penta E y SE33, son los más informativos producto a sus altos índices de PD, PIC, PE, heterocigosidad y mayor variabilidad de alelos, presentando un valor de  $P_i$  menor al resto de los marcadores analizados, constituyendo un incremento importante respecto a estudios previos. La información reportada valida el uso de estos 23 STR en la población venezolana para interpretar de forma confiable perfiles de ADN en pruebas de paternidad y casos forenses.

**Palabras clave:** STR, frecuencia alélica, genética forense, PCR, base de datos poblacional.

## INTRODUCCIÓN

Los avances científicos y técnicos en las herramientas moleculares utilizadas en genética han permitido su aplicación en el campo de la genética forense lo cual ha tenido gran impacto social. La genética forense es una especialidad enfocada en la aplicación de ciertas pruebas cuyos resultados impactan en la toma de decisiones a nivel penal o jurídico. En los laboratorios de genética forense los análisis más frecuentes son los relacionados con investigaciones de paternidad, o comparación de muestras encontradas en escenas de interés, para ser posteriormente identificadas. El objetivo de los genetistas forenses es identificar con la mayor certeza posible el origen de una muestra biológica (Jobling y Gill, 2004).

Estos estudios genéticos se basan en la comparación en conjunto de un número de distintos marcadores de herencia mendeliana, con respecto a una población de referencia. Este estudio comparativo de las muestras en relación a una población de referencia es el que le aporta el peso estadístico a un perfil genético en la identificación de individuos. De esta manera, el estudio y caracterización de una población ha tomado gran importancia en la actualidad, por lo cual surge la necesidad de las denominadas “bases de datos” de frecuencias alélicas de una población, la cual tiene como objetivo principal, recopilar todos los alelos presentes en una población, determinar tanto los más representativos como los menos frecuentes y de esta manera poder caracterizar a dicha población (Butler, 2009; Álvarez-Cubero y col., 2010).

La evolución en el campo de la genética forense comenzó con el descubrimiento de las variantes estudiadas en los grupos de sangre ABO (Landsteiner, 1990), y su posterior aplicación para el estudio de filiaciones, así como en casos criminales, hasta el descubrimiento de loci hipervariables conocidos como minisatélites (Jeffreys y col., 1985) generando los patrones multibanda que hoy día conocemos como huella génica. Posteriormente, tras el descubrimiento de los microsatélites STR (por sus siglas en inglés de *Short Tandem Repeats*) o secuencias repetidas en tándem, junto con las tecnologías de secuenciación automatizada, se han ido consolidando los sistemas para la identificación de individuos, teniendo como principal ventaja el gran poder de discriminación debido a su elevado grado de polimorfismo, así como la sensibilidad en la detección de las secuencias de interés (Jobling y Gill, 2004).

Actualmente, se utilizan marcadores STR para la obtención de perfiles genéticos, debido a las características anteriormente mencionadas y al advenimiento de la reacción en cadena de la polimerasa o PCR (por sus siglas en inglés de *Polymerase Chain Reaction*). La disponibilidad de estuches comerciales para STR que permitan la amplificación en una estrategia multiplex, permitiendo la amplificación de distintos marcadores genéticos simultáneamente, ha impactado favorablemente la ciencia forense debido a que la probabilidad de coincidencia puede superar una en varios miles de millones y esto es posible en una sola amplificación con una cantidad mínima de muestra, haciendo que el procedimiento sea más rápido, fácil y de menor costo. Hoy en día los resultados pueden obtenerse en sólo unas cuantas horas en comparación con las

semanas que requerían los métodos para la detección de los patrones multibanda o huella génica (Butler, 2006; Jobling y Gill, 2004).

A pesar del alto poder de discriminación de los estuches comerciales que se emplean actualmente en los laboratorios forenses del país, existen situaciones en las cuales el número de STR que contienen las bases de datos actuales es insuficiente para resolver algunos casos, como pruebas de paternidad donde no participa la madre, entre otros. Además, en estudios forenses en los cuales se emplea material de origen biológico cuyo ADN se encuentra degradado, se generan informes periciales reportando la imposibilidad de obtener un perfil genético, o que el mismo se ha logrado solo para algunos de los marcadores genéticos, conllevando a valores estadísticos poco significativos. Por tal motivo, se han desarrollado estuches comerciales más potentes, implementando un número mayor de marcadores a analizar. El problema de implementar estos nuevos estuches comerciales es la escasa cantidad de bases de datos en el país que contengan estos nuevos marcadores genéticos.

La inclusión de nuevos marcadores STR para uso forense requiere la caracterización de las poblaciones en las que se aplicarán, por lo tanto, en este trabajo se propone el estudio genético de una población venezolana, lo que permitirá contar con una base de datos para un conjunto de STR, requerida en los cálculos de casos forenses. Esta herramienta efectiva conlleva a mejorar el servicio del laboratorio, en casos donde el ADN se encuentra degradado o en poca cantidad y en los casos de filiación poder contar con marcadores adicionales que permitan alcanzar los valores de probabilidad



requeridos según los criterios de calidad establecidos por la ISFG (International Society for Forensic Genetics) (Martin y col., 2016).

## **1. EI ADN**

La unidad fundamental de la vida es la célula, la cual produce las materias primas y la energía necesaria para mantener la vida. El conjunto completo de instrucciones para construir un organismo, es decir, todo el ADN en una célula se conoce colectivamente como genoma. El ADN tiene dos funciones principales: realizar copias de sí mismo para que las células puedan dividirse y seguir con la información, y llevar las instrucciones para la síntesis de proteínas. La información codificada en la estructura del ADN se transmite de generación en generación, teniendo la mitad de la información genética origen materno y el resto, origen paterno (Butler, 2009).

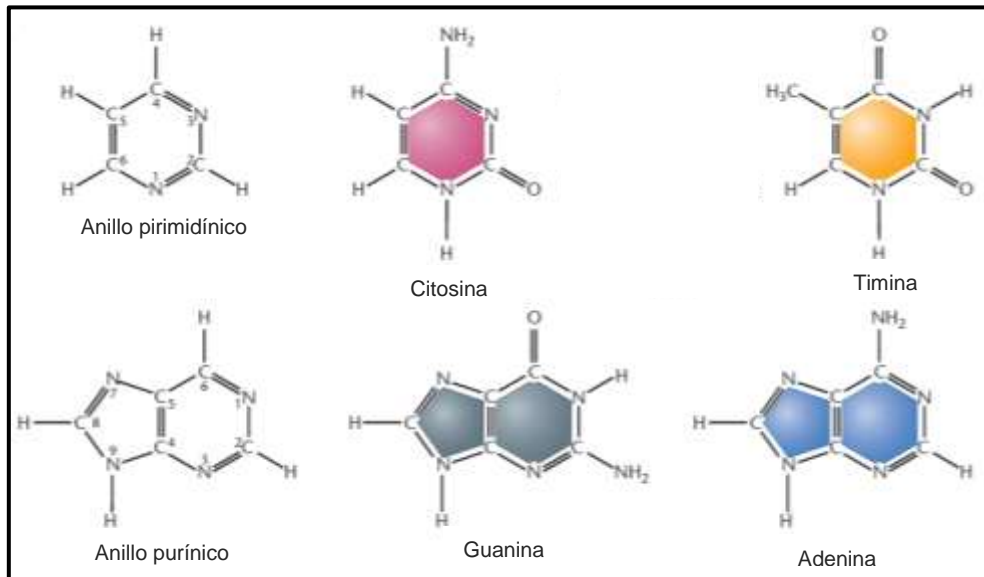
Dentro del núcleo de las células, existe una sustancia conocida como ADN que contiene la información para la replicación de la célula y la construcción de enzimas necesarias. Debido a que este ADN reside en el núcleo de la célula, se refiere como ADN nuclear (Butler, 2009).

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

nuclear que solo hay una copia en el núcleo de cada célula. (Klug y col., 2012). El ADN mitocondrial presenta regiones codificantes y una región hipervariable, además de poseer una herencia exclusivamente materna, por lo cual es muy utilizado también en identificación humana, para fines forenses que involucran tejidos antiguos como huesos, apéndices pilosos y dientes. Ambos genomas tienen interés médico-forense y poseen características que los hacen complementarios a la hora de la identificación e investigación de relaciones familiares, sobre todo en casos en los cuales el ADN nuclear no presente las mínimas condiciones para ser analizado y deba analizarse el ADN mitocondrial (Gill y col., 1994).

### **1.1. Componentes y estructura del ADN**

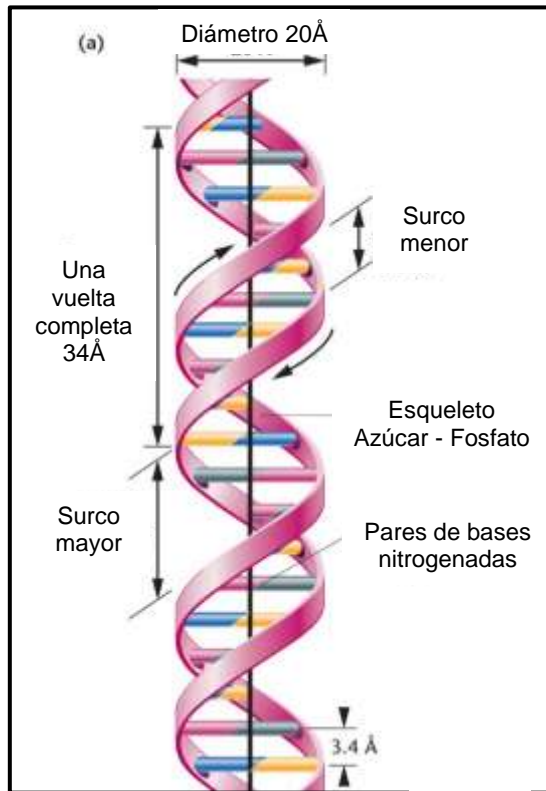
El ADN está constituido por unidades fundamentales, los desoxirribonucleótidos, también conocidos como bases, esto debido a que están constituidos por bases orgánicas cíclicas. Cada nucleótido contiene un fosfato, un azúcar (desoxirribosa) y una base nitrogenada (Griffiths y col., 2000). De esta manera, existen dos tipos de bases nitrogenadas: las purinas (Adenina y Guanina), con un doble anillo de nueve lados, y las pirimidinas (Citosina y Timina), con un anillo de seis lados (Klug y col., 2012) (Figura 1).



**Figura 1.** Estructura química de las bases nitrogenadas. Adaptado de: Klug y col., 2012.

James Watson y Francis Crick dedujeron en 1953, con la información disponible hasta entonces, y con la ayuda del análisis cristalográfico por difracción de rayos X, el modelo más aceptado de la estructura del ADN, destacándose las siguientes conclusiones (Klug y col., 2012) (Figura 2):

- Dos largas cadenas polinucleotídicas enrolladas alrededor de un eje central, formando una doble hélice girando hacia la derecha (dextrógira). Ambas son antiparalelas; es decir, la orientación C-5' - C-3' va en direcciones opuestas.
- Las bases nitrogenadas de las cadenas opuestas están apareadas como resultado de la formación de puentes de hidrógeno; en el ADN sólo se permiten los emparejamientos A con T y G con C.



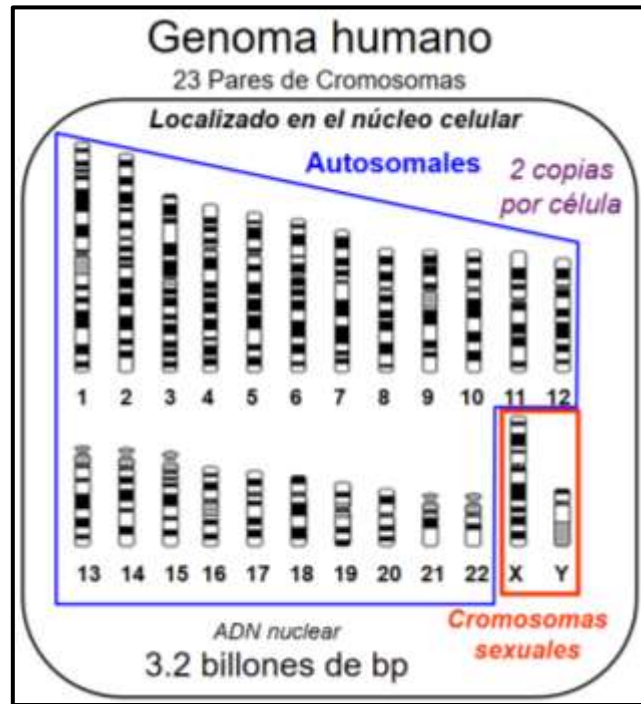
**Figura 2.** Modelo de doble hélice propuesto por Watson y Crick. Adaptado de: Klug y col., 2012.

El ADN es capaz de almacenarse dentro del núcleo gracias a una perfecta organización de su compactación. De esta manera, puede de reducir su tamaño mediante proteínas histonas que se asocian y llevan a cabo la compactación del mismo. Las histonas son proteínas básicas con carga parcial positiva, ricas en residuos de lisina y arginina, las cuales desempeñan un papel importante en la fijación del esqueleto azúcar-fosfato que tiene carga negativa (Griffiths y col., 2000).

## **1.2. Organización del ADN**

El ADN está dividido en cromosomas, en forma de densos paquetes de ADN, el genoma humano consiste en 22 pares de cromosomas autosómicos y un par de cromosomas sexuales. Con base en los cromosomas sexuales, los varones se designan como XY porque contienen una sola copia del cromosoma X y una sola copia del cromosoma Y, mientras que las mujeres tienen dos copias del cromosoma X y se designan como XX (Butler, 2009).

Los cromosomas de las células de todo el cuerpo (somáticas) son diploides; es decir, contienen dos juegos de cada cromosoma. Por otro lado, los gametos (espermatozoides u ovocitos) son haploides; ellos tienen un solo juego de cromosomas. Cuando un ovocito y un espermatozoide se combinan durante la concepción, resulta un cigoto que de nuevo es diploide. Por lo tanto, un cromosoma de cada par cromosomal se deriva de uno de los padres en el momento de la concepción (Butler, 2009) (Figura 3).



**Figura 3.** Genoma Humano. Adaptado de: Butler, 2009.

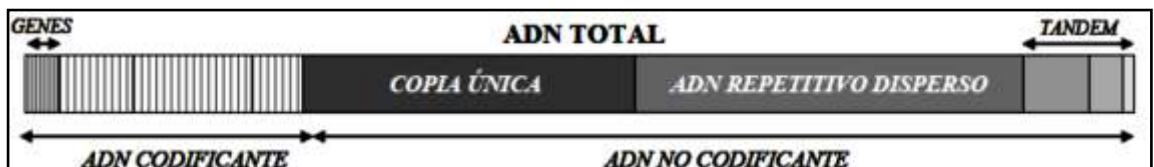
Asimismo, el ADN en los cromosomas se encuentra compuesto de regiones codificantes y no codificantes. Las regiones codificantes son conocidas como genes y contienen la información necesaria para la síntesis de proteínas que actúan a nivel celular y que se expresan en la persona como un carácter individual. Además, forma alrededor del 5% del genoma humano y presenta una escasa variabilidad entre individuos, con la excepción de ciertas regiones, como el sistema de antígenos leucocitarios humanos o HLA (por sus siglas en inglés de *Human Leukocyte Antigen*) (Butler, 2009).

Por su parte, las regiones no codificantes de ADN forman el resto del material cromosómico, el cual desempeña funciones estructurales, reguladoras y, en gran parte, algunas funciones desconocidas. Debido a que estas regiones no están directamente

relacionadas con la síntesis de proteínas, son usados como marcadores para pruebas de identificación humana debido a que contiene la mayor parte de la variabilidad genética entre los individuos, por su alto polimorfismo, ya que las mutaciones no están sometidas a una presión selectiva tan fuerte como la que ocurre en el ADN codificante (Butler, 2009).

En general, el ADN no codificante puede ser de dos tipos (Butler y Hill, 2012) ADN repetitivo de copia sencilla que actúa como ADN espaciador entre regiones codificantes del genoma, entre las que se encuentran los intrones, que no contienen información útil para la síntesis de la cadena polipeptídica y son eliminados durante el proceso de *splicing*.

- ADN repetitivo de copia múltiple, que se caracteriza por la presencia de una secuencia común, repetida en tándem de manera consecutiva, en un segmento de ADN y se distribuye por todo el genoma.



**Figura 4.** Organización del ADN total humano, según su estructura y función. Tomado de: Sánchez, 2010.

### 1.3. Alelos y polimorfismos

Alelo es el término que define a formas alternativas de un gen o locus, y locus se utiliza para designar una posición o localización específica de un gen o una secuencia blanco en un cromosoma (Griffiths y col., 2000).

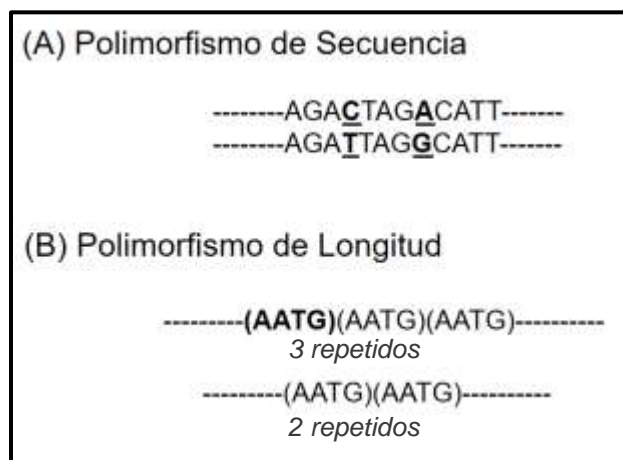
La coexistencia de alelos múltiples en un locus se denomina polimorfismo genético. Por definición, cualquier sitio en el cual existen alelos múltiples como componentes estables de la población, es polimórfico. Es decir, el polimorfismo expresa la variabilidad que existe dentro de un fragmento de ADN. En general, cuantos más alelos haya, mayor polimorfismo y, por tanto, mayor poder de identificación (Butler, 2009).

Por su parte, los polimorfismos se definen como variaciones en la secuencia del ADN, tengan o no consecuencia biológica selectiva alguna, y son detectados en al menos el 1% de los individuos de una población. Para considerar un locus como polimórfico, el alelo más común de dicho locus debe tener una frecuencia poblacional menor al 95% y al menos un 2% de la población debe ser heterocigota para este locus. Los polimorfismos pueden ir desde la modificación de una sola base, hasta cambios en el número y tamaño de las unidades de repetición (Kruglya y Nickerson, 2001; Butler, 2009; Butler, 2011) (Figura 5):

- Polimorfismos de secuencia: se producen por el cambio de uno o más nucleótidos en una secuencia de ADN. Por mencionar algunos ejemplos, se encuentran aquellos genes que codifican para los grupos sanguíneos, el sistema HLA, el análisis de ADN mitocondrial, entre otros.
- Polimorfismo de longitud: se presentan por la inserción o delección de uno o más nucleótidos. Dentro de estos se encuentran principalmente los minisatélites y microsatélites. Ambos se caracterizan por tener una secuencia que se repite varias veces, en tándem. Este tipo de polimorfismo conforma el 10-25% del



genoma total y se presenta en regiones no codificantes, lo cual hace que diferencias en las secuencias de ese ADN no sean raras y se transmitan como tal a la descendencia, de manera mendeliana, por lo que son comúnmente utilizados como marcadores genéticos en la identificación humana.



**Figura 5.** Caracterización del polimorfismo de secuencia y longitud. Adaptado de: Butler, 2009.

## 1.4. Polimorfismos de longitud

### 1.4.1 Minisatélites

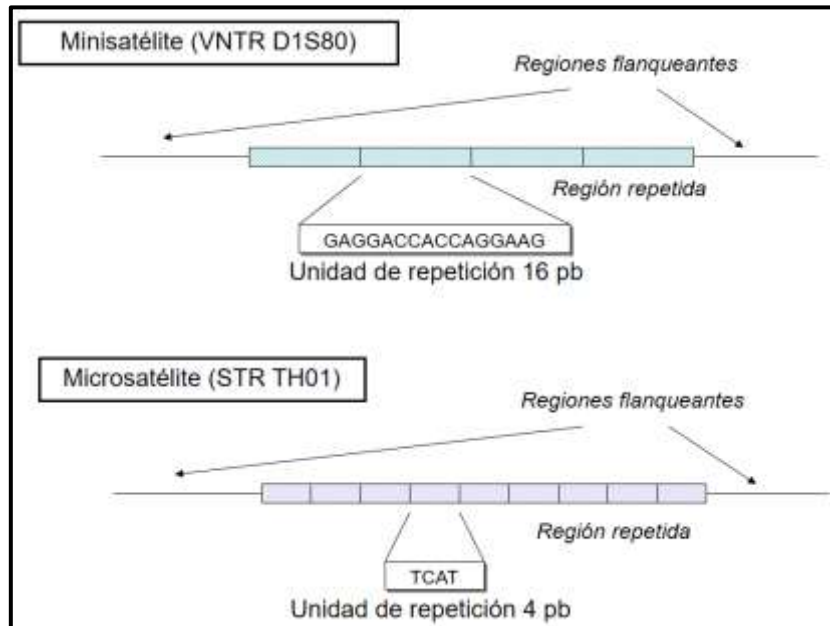
Los minisatélites presentan una unidad de repetición de 10 a 100 bases. El término minisatélite proviene por extensión del de ADN satélite, una parte considerable del genoma eucariota que se puede separar del resto por configuración en gradiente de densidad. La mayoría de estos polimorfismos se detectan mediante enzimas de restricción, y se les conoce como VNTR (por sus siglas en inglés de *Variable Number Of Tandem Repeats*) o número variable de repeticiones en tándem (Figura 6). La identificación

de los VNTR se realiza con el uso de sondas complementarias sintéticas que hibridan con la secuencia de ADN, propia del VNTR en estudio. Estas sondas pueden ser mono-locus, que hibridan en una sola región de ADN, y las multi-locus, que hibridan en forma simultánea con varias regiones (Butler, 2009; Jobling y Gill, 2004).

A partir del descubrimiento de los primeros VNTR se observó que podrían ser una fuente importante de marcadores en medicina forense y podrían sustituir a los marcadores convencionales. Posteriormente se demostró que se podría obtener un patrón de VNTR o “fingerprint” estable y reproducible en manchas secas de sangre de más de 4 años de antigüedad, de manchas de semen, de folículos pilosos y de manchas de sangre expuestas a distintas condiciones ambientales (Jeffreys y col., 1985).

Aunque los análisis con VNTR contienen varias limitaciones, entre las cuales destacan (Jobling y Gill, 2004):

- El tiempo que se invierte en cada estudio (semanas)
- La elevada concentración de ADN que deben presentar las muestras
- La poca eficiencia del estudio al tratarse de muestras con ADN degradado, por el mayor tamaño del ADN que se examina
- No es posible amplificar conjuntamente varios marcadores a la vez (estrategias múltiplex).



**Figura 6.** Caracterización del ADN repetido en tándem. Representando un minisatélite, con un núcleo de repetición de 16 pb en 3 bloques. En la parte inferior, un microsatélite, con un núcleo de 4 pb repetido 9 veces consecutivas. Adaptado de: Butler, 2009.

#### 1.4.2 Microsatélites

Los microsatélites varían en el tamaño de la unidad de repetición, el cual es de 2 a 6 nucleótidos. A este tipo de polimorfismo se le conoce como *short tandem repeats* (STR) y su aparición en el campo forense empieza a principios de la década de los 90 gracias al advenimiento de la PCR (Butler, 2009; Jobling y Gill, 2004) (Figura 6).

A diferencia de los loci VNTR, que se sitúan preferentemente en las regiones teloméricas, los STR parecen ser abundantes por todo el genoma y han sido descritos tanto en regiones génicas como extragénicas. Los STR situados en regiones génicas están presentes no solo en intrones y secuencias flanqueantes, sino también dentro de

regiones codificantes. Los STR están asociados frecuentemente con el extremo 3' de las inserciones Alu, resultando, probablemente, de la degeneración de la secuencia de la cola poli-A (Economou y col., 1990).

Desde su descubrimiento han sido los responsables de un gran avance en el campo de la identificación humana, en el desarrollo de los mapas de ligamiento humano y en nuevos procedimientos de diagnóstico clínico. Así, diversos loci STR han sido utilizados para el diagnóstico predictivo de múltiples enfermedades genéticas, incluyendo distrofia miotónica, fibrosis quística, distrofia muscular de Duchenne/Becker y enfermedad de Huntington (Herráez, 2012).

Cabe destacar que el patrón de nomenclatura de los STR sigue un orden. La primera letra del nombre del marcador es una "D" y significa DNA (ADN). Luego, se tiene un número que representa en cuál cromosoma el marcador está localizado. El tercer elemento del nombre es la letra "S" (*Single Copy Sequence*), que significa secuencia de copia única. Por último, se tiene el número del locus en que el marcador fue descrito (Butler, 2009).

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

**A. Extracción del ADN:** cuando se obtienen muestras biológicas, ya sea una de un lugar de los hechos o muestras de referencia, contienen además del ADN otras sustancias que pueden inhibir la habilidad de analizarlo, como son proteínas celulares que

empaquetan y protegen al ADN en el interior de la célula, DNAsas o inhibidores de la polimerasa en el soporte en el que está contenido el indicio. Por lo que, se han desarrollado métodos de extracción del ADN para separar las proteínas y otros materiales que están unidos a las moléculas de ADN. En la actualidad, se utilizan tres principales técnicas para la extracción del ADN en los laboratorios forenses: la extracción orgánica, la extracción con Chelex y la extracción en papel FTA (Butler, 2007; Butler, 2009), además de otros métodos basados en la aplicación de resinas paramagnéticas cuyo uso se ha extendido en los laboratorios forenses debido al rendimiento de ADN de alta pureza.

**B. Cuantificación de ADN:** luego de la extracción de ADN, es necesario conocer la cantidad obtenida, así como su calidad. La adición de una cantidad óptima de ADN para PCR producirá mejores resultados en un menor tiempo; la adición de excesivo ADN o una cantidad insuficiente daría lugar a un perfil difícil o incluso imposible de interpretar. Esto es importante cuando se trata de indicios, cuando es muy difícil conocer el estado de conservación del material biológico y, en muchas ocasiones, es difícil estimar la cantidad que se ha recogido de material celular. Es menos importante cuantificar ADN utilizando muestras de referencia, donde se espera una cantidad similar de ADN en cada extracción, ya que no hay muchas variables que afecten su concentración (Butler, 2009; Herráez, 2012).

**C. Amplificación:** la PCR es un proceso enzimático en el que se replica una región específica de ADN una y otra vez para producir más de un millón de copias de una secuencia particular.

Para este procedimiento se emplean dos oligonucleótidos sintéticos de unos 15-20 nucleótidos que son complementarios a las zonas flanqueantes de la región que se quiere amplificar. Dichos oligonucleótidos actúan como cebadores para la síntesis in vitro de ADN, la cual, se encuentra catalizada habitualmente por una enzima llamada *Taq polimerasa*. Dicha enzima se aísla de una bacteria termófila, denominada *Thermus aquaticus*, que es capaz de crecer a temperaturas elevadas. A esta temperatura dicha enzima es capaz de mantener una media de extensión de más de 60 nucleótidos por segundo en regiones ricas en uniones G-C. La temperatura óptima a la que actúa esta enzima permite el uso de elevadas temperaturas para la unión de los cebadores y para la extensión, de esta manera se aumenta el nivel de exigencia de la reacción y se reduce la extensión de los cebadores unidos inespecíficamente al ADN (Herráez, 2012).

La reacción se lleva a cabo en una serie de ciclos cada uno de los cuales incluye tres fases (Herráez, 2012):

- a. Desnaturalización: para que ocurra la reacción es necesario que el ADN molde se encuentre en forma de cadena sencilla. Esto se obtiene aplicando temperaturas de 90 a 95°C que producen la ruptura de los puentes de hidrógeno intercatenarios y por lo tanto la separación de ambas cadenas. Para una separación completa de las hebras de toda la muestra, la temperatura debe mantenerse unos minutos. Si el ADN solo se desnaturaliza parcialmente, éste tenderá a renaturalizarse

rápidamente, evitando así una eficiente hibridación de los cebadores y una posterior extensión

- b. Hibridación: una vez que el ADN se encuentra desnaturalizado se disminuye la temperatura hasta un rango comprendido entre 40 y 60°C para permitir la unión de los cebadores a las secuencias flanqueantes del fragmento que se desea amplificar. La temperatura de fusión o *annealing* ( $T_m$ , *melting temperature*) depende de diversos factores y es relativamente específica para cada cebador, ya que si la misma es muy baja la unión se hará de manera inespecífica y si es muy alta no se producirá una unión completa.
- c. Extensión: la *Taq polimerasa* incorpora nucleótidos en el extremo 3' del cebador utilizando como molde la cadena de ADN previamente desnaturalizada. La temperatura a la que se lleva a cabo este paso suele ser de 72°C ya que es la temperatura en la cual la enzima alcanza su máxima actividad. Normalmente una extensión de 20 segundos es suficiente para fragmentos menores de 500 pb, y 40 segundos para fragmentos por encima de 1,2 Kb.

Un factor importante en el transcurso de las distintas fases es el tiempo de rampa, el cual se define como el tiempo invertido en pasar de una temperatura a otra y depende del diseño y de las características del instrumento donde se realiza automáticamente este proceso, el termociclador.

Finalmente es necesario destacar que para poder efectuar esta reacción se requieren de otros componentes básicos (Herráez, 2012):

- a. Buffer de amplificación: normalmente contiene cloruro de potasio, tris y cloruro de magnesio. El  $MgCl_2$  es el componente más influyente en la especificidad y rendimiento de la reacción ya que los iones  $Mg^{2+}$  son necesarios para la actividad de la *Taq polimerasa*, es decir, actúan como cofactores de la polimerasa.
- b. Desoxinucleótidos trifosfatos (dNTP): se utilizan concentraciones aproximadas de 200  $\mu M$  en un volumen de reacción de 25  $\mu l$  para poder sintetizar entre 6-6,5  $\mu g$  de ADN. La concentración de dNTP y  $MgCl_2$  está relacionada ya que el magnesio se une a los dNTP.

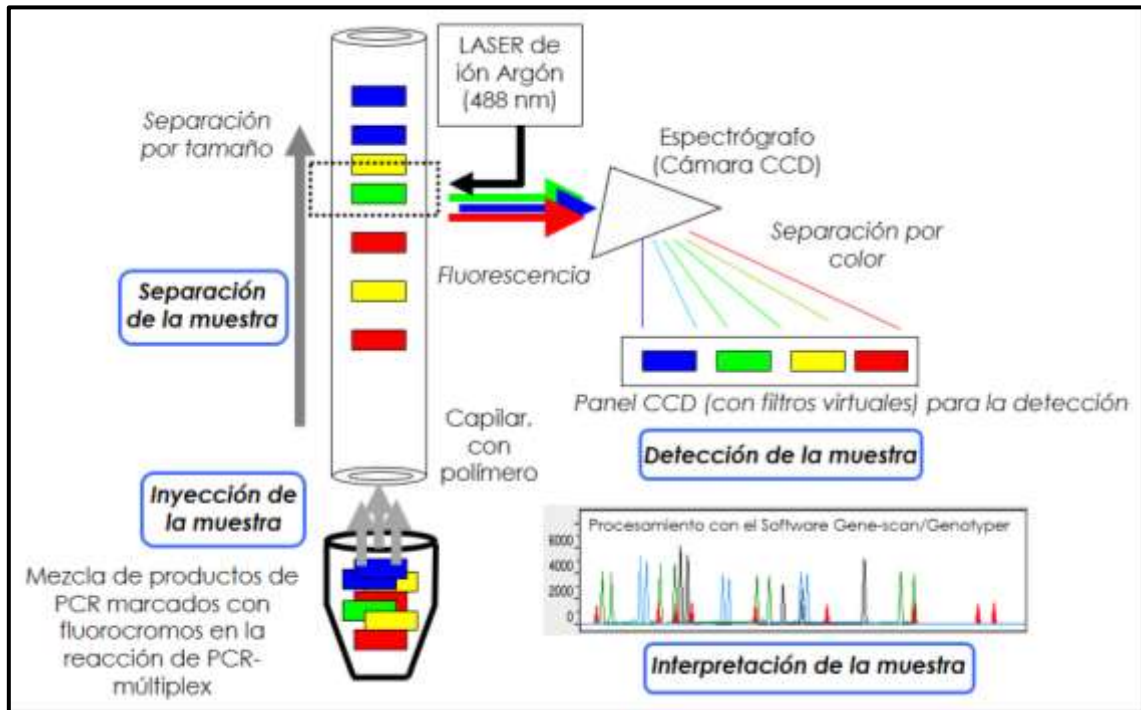
**D. Electroforesis capilar**: durante la PCR en la cual los alelos STR son amplificados, se produce una mezcla de moléculas de ADN que representa un problema de difícil separación. La PCR múltiplex es capaz de producir 20 o más fragmentos de ADN que deben y pueden ser distinguidos uno de otro. Además, la resolución de una sola base requiere que se diferencia entre los alelos estrechamente espaciados (por ejemplo, alelos TH01 9,3 y 10). Con la finalidad de observar diversas moléculas una de otra, el paso de separación requiere mover aparte los fragmentos en diferentes tamaños. La separación se realiza típicamente por un proceso conocido como electroforesis y se lleva a cabo ya sea en gel o en capilar (Butler, 2009; Butler y col., 2004).

La detección de la muestra se lleva a cabo automáticamente por el instrumento de electroforesis capilar a través de la medición del intervalo de tiempo de inyección de la muestra y cuya detección se realiza a través de una cámara CCD (por sus siglas en inglés de *Charge-Coupled Device*). El capilar en el cual se lleva a cabo la



electroforesis se encuentra cubierto por una sustancia opaca, excepto por una pequeña zona denominada “ventana del capilar” la cual es atravesada por el láser y permite que las muestras sean excitadas a su paso por la misma. Como en los geles, las moléculas más pequeñas llegan primero al punto de detección, seguidas por las moléculas más grandes. Los datos obtenidos de la separación están representados gráficamente en función de la intensidad relativa de fluorescencia observada a partir de la emisión de fluorescencia de los colores que pasan por el detector. Las señales de emisión fluorescente de los fluorocromos que están unidos a los iniciadores se pueden utilizar para detectar y cuantificar las moléculas de ADN que pasan por el equipo (Butler, 2009; Butler y col., 2004) (Figura 7).

Además, la asignación de los tamaños de los alelos se lleva a cabo mediante escaleras alélicas que consisten en la mezcla de secuencias amplificadas de varios alelos que se emplean como marcadores para cada loci STR. La comparación de las muestras con la escalera alélica permite una determinación rápida de los tamaños de los alelos.



**Figura 7.** Separación y detección de fragmentos de ADN por electroforesis capilar con un analizador genético ABI 3130xl. Adaptado de: Butler, 2009.

**E. Análisis de fragmentos mediante el uso de software:** los espectros de los fluorocromos utilizados para el marcaje de los productos de PCR se solapan y los datos obtenidos contienen picos que se componen de más de un marcaje fluorescente. El software, ya sea el *GeneScan* o el *GeneMapper*, elimina la superposición espectral en el perfil y calcula los tamaños de los fragmentos de ADN amplificados. El software calcula la cantidad de solapamiento espectral que hay entre cada fluorocromo y resta este de los picos en el perfil, obteniendo de esta manera picos separados que pueden ser interpretados (Butler, 2009).

### **1.5. Características deseables de STR utilizados en la tipificación de ADN**

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

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

Además de la variación en el número de repeticiones, las secuencias STR varían también en la longitud de la unidad de repetición. Entre los distintos tipos de STR, los tetranucleótidos son más populares que los di o trinucleótidos. Penta y hexanucleótidos son menos comunes en el genoma humano, pero están siendo examinados por algunos laboratorios (Butler, 2009; Butler y Hill, 2012). Actualmente, en los nuevos estuches comerciales existentes en el mercado se han introducido en sus sistemas pentanucleótidos (como el Penta D y el Penta E que están dentro del conjunto de marcadores del *PowerPlex® 16 HS* y *PowerPlex® Fusion*) (Promega, 2016; Promega, 2017B).

## **2. Importancia del análisis estadístico de los STR**

### **2.1. Estudios de paternidad**

El que cada persona conozca a su padre o madre es normal y además es un derecho que suele ser garantizado por las leyes de muchos países. Sin embargo, en ciertos casos una persona puede ignorar quien es su padre (o madre en algunos casos) o, tener dudas sobre su paternidad. El análisis de la paternidad se basa en las leyes mendelianas de la herencia, en las que se establece por primera vez que todo el material genético de una persona procede a partes iguales del padre y de la madre. Por lo tanto, si en un hijo se encuentra un alelo que no posee el probable padre, la paternidad podría quedar excluida.

Si por lo contrario se encontrara que el hijo tiene alelos que pudieron haberse heredado del probable padre, hay que realizar una serie de cálculos estadísticos con el objetivo de determinar cuántas personas entre la población general, podría ser también padres potenciales, por tener esos alelos (Ibarra y Arcos, 2000).

### **2.2. Criminalística**

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

Con el pasar del tiempo se ha ido desarrollando la criminalística, que podemos definir como la “ciencia aplicada que, de acuerdo al ordenamiento jurídico de cada país, estudia científicamente los indicios y evidencias con objeto de convertirlos en pruebas formales que puedan ser presentadas ante las autoridades judiciales para permitir la identificación de víctimas y de los delincuentes y esclarecer las circunstancias de un probable delito” (Hernández, 2014).

El obtener un perfil genético es sólo la primera parte del análisis, los perfiles tienen que ser comparados. El análisis estadístico de dichos sistemas está basado en el establecimiento de una probabilidad de identidad entre perfiles de ADN, lo cual se encuentra implementado hoy en día en casuística forense para pruebas de identificación, filiación y criminalística (Evet y Weir, 1998).

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

### **2.3. Personas desaparecidas**

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

En estas circunstancias, es frecuente que las personas desaparecidas hayan fallecido. El único alivio para los familiares es recibir una confirmación fidedigna del fallecimiento y saber que los restos de sus seres queridos han sido o pueden ser tratados con dignidad y con respeto por su cultura y sus creencias religiosas. Por ello la identificación adecuada de restos humanos es parte fundamental del proceso de reparación no solo para los familiares de personas desaparecidas, sino para las comunidades enteras.

La evolución de la ciencia forense y, en particular, de la genética forense mediante el análisis de ADN, ha permitido que muchas familias de personas desaparecidas conozcan el paradero de sus seres queridos desaparecidos, como así también que los restos de las personas desaparecidas fallecidas sean identificado y entregados a sus familiares. Antes de la aparición del análisis de ADN, uno de los elementos utilizados en programas de identificación humana era la hemogenética forense, técnica que se aplicó particularmente en Argentina durante la década de 1980. Sin embargo, dichos estudios

pioneros eran complejos, costosos y carecían del alcance y las facilidades que ofrece hoy en día el análisis forense de ADN (Comité Internacional de la Cruz Roja, 2009).

La identificación de un cadáver depende básicamente de su estado, si el cuerpo se encuentra en condiciones que permitan tomar fotografías, huellas dactilares o datos odontológicos, los datos obtenidos pueden ser de gran ayuda para su identificación. Sin embargo, cuando todas estas posibilidades se agotan, o cuando el cadáver no es sino un esqueleto (a veces incompleto), surge la necesidad de usar algo más específico, a ser posible único, que esté en todos los cadáveres y que se pueda encontrar incluso en los restos óseos: ese algo es el ADN (Hernández, 2014). De esta manera surge la necesidad de los bancos de datos que almacenan información referente a los perfiles genéticos pertenecientes a individuos, los cuales son usados principalmente para investigar delitos e identificar personas sospechosas o restos de personas desaparecidas (Mestres y Vives-Rego, 2009)

### **3. Bases de datos genéticas**

Los avances científicos y técnicos han generado y generan una cantidad de información extraordinaria, la cual ha de almacenarse de modo racional y ordenado para su posterior uso. De esta manera, surge la necesidad de las denominadas “bases de datos”, lugares donde se almacena de modo ordenado y coherente cualquier tipo de información, que luego es rescatada de modo automático de acuerdo con parámetros previamente establecidos.

Una base de datos poblacional es una recopilación de las frecuencias estimadas en una población para cada variante de un marcador genético y para todos los sistemas genéticos considerados. Estas se utilizan para estimar cuán probable es encontrar por azar un perfil concreto en una determinada población. Una base de datos precisa la información genética de una muestra de un número elevado de individuos de una población, recomendando que el tamaño de la muestra sea como mínimo de 100. Además, dichos individuos no deben tener relaciones de parentesco entre ellos, sino, de lo contrario, algunas variantes genéticas estarían más representadas de lo que están en la población y otras poco representadas o incluso ausentes (Mestres y Vives-Rego, 2009).

De nuevo, por ser una base de datos, se refiere a una colección de perfiles de ADN. Sin embargo, esta base de datos poblacional no se utiliza para ningún propósito de comparación de muestras. Más bien, es utilizada para estimar las probabilidades de coincidencias aleatorias basadas en las mediciones de la frecuencia de los alelos de un grupo de, usualmente, 100 o más individuos seleccionados para representar un grupo específico de interés. Cabe destacar que los individuos a partir de los cuales se generan los perfiles genéticos para su uso en una base de datos poblacional son completamente anónimos y solo se clasifican y agrupan por su identidad étnica autoidentificada (Butler, 2009)

Existe un consenso internacional acerca de la necesidad de establecer bancos de datos de perfiles de ADN para fines civiles o criminales. E.E.U.U fue el pionero en crear,



en 1994, el sistema CODIS (*Combined DNA Index System*), desarrollado por el FBI (*Federal Bureau of Investigation*), el cual consistía en 13 marcadores genéticos, que hoy en día forman el sistema índice de cualquier laboratorio forense (Butler y Hill, 2012). Para el año 2015 se actualizó el sistema CODIS al añadirle 7 loci adicionales, D1S1656, D2S441, D2S1338, D10S1248, D12S391, D19S433 y D22S1045, para un total de 20 loci STR, y de esta manera reducir el número de coincidencias accidentales, aumentar la compatibilidad internacional y aumentar el poder de discriminación para casos criminales y de personas desaparecidas (Hares, 2015).

Actualmente, el sistema CODIS es un gran instrumento de trabajo debido a que permite la compatibilidad y el trabajo independiente de distintos laboratorios; posibilita el acúmulo e intercambio de datos a distintos niveles: Local (L-DIS), Estatal (S-DIS) y Nacional (N-DIS) cuya unión conforma el sistema CODIS; contiene diferentes apartados o índices, incluyendo los criminales y civiles de personas desaparecidas; su instalación es gratuita así como el asesoramiento por parte de los EE. UU. a través del FBI a los laboratorios oficiales de los distintos países, siempre que haya una petición oficial y se den una serie de circunstancias favorables (Shastri, 2002).

#### **4. Genética de poblaciones**

La población se define como un grupo de individuos que comparten un conjunto de genes, que viven en la misma área geográfica y que tienen capacidad de reproducirse (IPGRI y Cornell University, 2004).

La genética de poblaciones es el estudio de la variación heredada y su modulación en tiempo y espacio. Es un intento de cuantificar la variación observada dentro de un grupo de población o entre diferentes grupos de población, en términos de frecuencias alélicas y fenotípicas (Butler, 2009).

La variabilidad dentro de un locus debe ser lo suficientemente estable como para pasar con precisión el alelo a la próxima generación, es decir, que posean una tasa de mutación baja. Sin embargo, tampoco deben ser demasiado estables porque si no sólo unos pocos alelos existirían en el tiempo y el locus no podría aportar tanta información, es decir, no sería útil en pruebas de identificación humana (Butler, 2009).

La implementación de marcadores STR por los laboratorios forenses, requieren de la realización de estudios poblacionales que permitan estimar parámetros estadísticos de interés, como frecuencias alélicas, y comprobar que la población de estudio se encuentre en equilibrio Hardy-Weinberg, lo cual permitiría estimar la frecuencia de genotipos a partir de frecuencias alélicas. Los parámetros forenses y la validación poblacional permiten utilizar de manera confiable estos sistemas para estimar la frecuencia de un perfil genético en un caso criminal, o calcular la probabilidad de paternidad cuando coinciden un supuesto padre y el hijo en una prueba de paternidad. Por este motivo, se han realizado estudios poblacionales y se han generado bases de datos con STR a nivel mundial (Butler, 2009).

#### **4.1. Frecuencias alélicas y genotípicas**

La frecuencia genotípica es la cantidad de veces que aparece el genotipo en una población. Mientras que la frecuencia alélica es la medida de la proporción relativa de alelos de una población específica, expresadas en tantos por ciento o en tantos por uno, y se calculan mediante el recuento de cada alelo y dividiendo este número por el número total de alelos analizados (Cavalli-Sforza y col., 1981).

$$p = \frac{\# \text{ alelos observados de un locus}}{\# \text{ total de alelos}}$$

#### **4.2. Ley de equilibrio de Hardy-Weinberg**

La Ley de Hardy-Weinberg fue postulada en 1908 por George Hardy y Wilhelm Weinberg. Según ella “en toda población suficientemente grande, en la que los apareamientos se producen al azar y que no está sujeta a ninguna fuerza evolutiva como mutación, selección y migración, las frecuencias génicas y genotípicas se mantienen constantes de generación en generación”. De esta manera, una población que cumpla dichas características se dice que está en equilibrio de Hardy-Weinberg. En este contexto, las frecuencias alélicas y genotípicas de una determinada población permanecen constantes, generación tras generación, en una población grande con apareamiento al azar y en ausencia de mutaciones, migraciones y selección, pudiendo predecir sus valores para las futuras generaciones (Stark, 2005).

De acuerdo con la Ley de equilibrio de Hardy-Weinberg, las frecuencias genotípicas de la progenie están determinadas por las frecuencias alélicas de sus progenitores. Si la población está en equilibrio, las frecuencias alélicas son iguales en los progenitores y en la progenie, de manera que las frecuencias alélicas observadas en la progenie pueden utilizarse como si fueran las frecuencias alélicas paternas para calcular las frecuencias genotípicas esperadas por la Ley de Hardy-Weinberg (Stark, 2005).

#### **4.3. Poder de exclusión (PE)**

Es la probabilidad de excluir un individuo que esté involucrado en algún caso forense, permitiendo establecer la proporción de individuos falsamente implicados en un peritaje. En otras palabras, corresponde a la probabilidad de que un sistema genético específico proporcione evidencias que conduzcan a la exclusión de un probable responsable. El Grupo Español y Portugués de la Sociedad Internacional de Genética Forense (GEP-ISFG) considera que estos índices deben estar por encima del 99,9% (Rosado, 2008).

Éste es un estadístico que predice cómo se producirá la identificación en el análisis a posteriori, es decir, los análisis estadísticos correspondientes a la identificación de perfiles genéticos. Depende del número de alelos identificados y de la distribución de las frecuencias alélicas. Corresponde a un valor porcentual que es función directa del polimorfismo de un sistema. Cuánto más polimórfico éste es y más equilibradas estén las frecuencias de sus alelos, mayor será su PE, lo que lo hace ser un parámetro de gran valor médico-legal, especialmente en determinaciones biológicas de paternidad.

$$PE = H^2(1 - (1 - H)H^2) \quad \text{donde } H = \text{Heterocigosidad}$$

#### 4.4. Poder de discriminación (PD)

Es la probabilidad de que dos individuos no relacionados y seleccionados al azar puedan ser diferenciados genéticamente mediante el análisis de uno o varios marcadores. Depende del número de loci analizados, de las variables alélicas encontradas y frecuencias de cada locus en cada población y es una medida relativa de la eficacia del sistema o sistemas analizados. Un PD cercano a 1 sería, una situación deseable para establecer una identificación (Smalldon y Moffat, 1973).

$$PD = 1 - Pi \quad \text{donde } Pi = \text{Probabilidad de coincidencia aleatoria}$$

#### 4.5. Probabilidad de coincidencia aleatoria (Pi)

Es la probabilidad de que dos individuos no relacionados y seleccionados al azar compartan un genotipo idéntico en el locus sometido a prueba.

Corresponde al número de individuos que podrán ser examinados antes de encontrar el mismo patrón de ADN en un individuo seleccionado aleatoriamente (Jones, 1972).

$$Pi = \sum_{i=1}^n x_i^2 \quad \text{donde } x_i^2 = \text{frecuencia genotípica}$$

#### 4.6. Heterocigosidad (H)

Probabilidad de que dos alelos del mismo locus, tomados al azar, sean diferentes.  $H_{obs}$  refleja la frecuencia de individuos heterocigotos encontrados en la población de estudio, mientras que  $H_{esp}$  es una estimación más precisa de la proporción de individuos heterocigotos que se espera en la población. Una alta heterocigosidad implica que existe más diversidad alélica y por lo tanto hay una menor probabilidad de coincidencia en las muestras (Crow y Kimura, 1970).

$$H_{obs} = 1 - \sum_{i=1}^n P_i$$

$$H_{esp} = 1 - \text{Homocigosidad} = 1 - \sum_{i=1}^n P_i^2$$

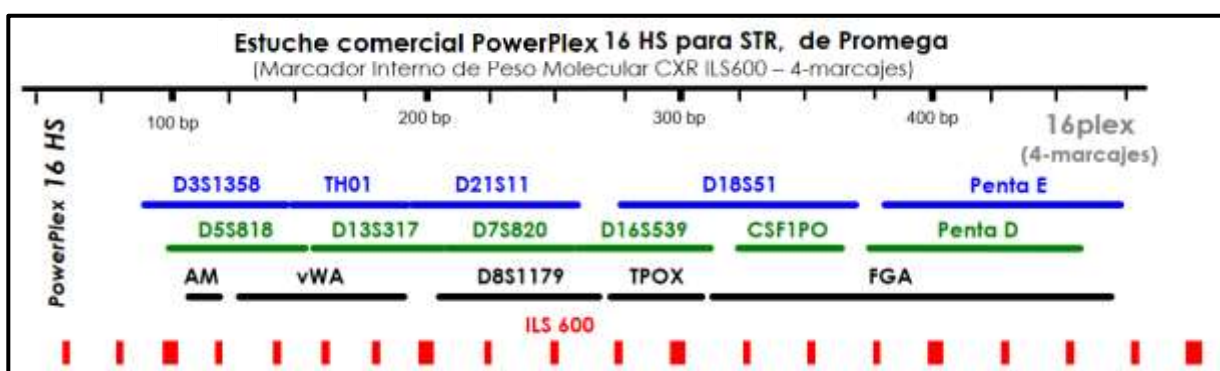
#### 4.7. Índice de contenido polimórfico (PIC)

Es un índice que evalúa la información de un marcador en la población de acuerdo con las frecuencias de los alelos. Refleja la capacidad de deducir el origen parental de cada uno de los alelos del genotipo en la descendencia en un locus particular. Un STR es considerado como un buen marcador molecular cuando presenta un número de alelos superior a 5 y el PIC con índices por encima de 0,75 (Botstein y col., 1980).

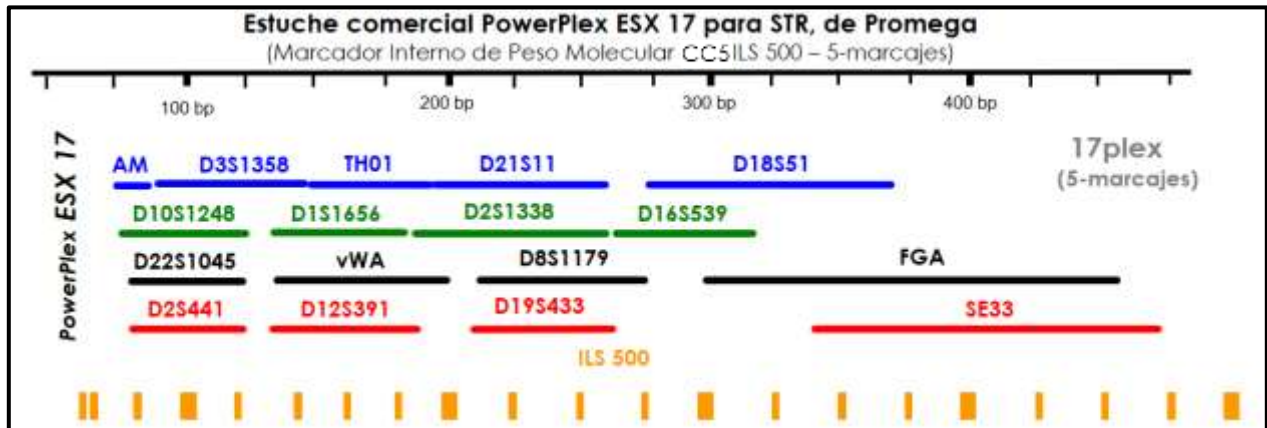
$$PIC = 1 - \sum_{i=1}^n P_i^2 - \left( \sum_{i=1}^n P_i^2 \right)^2 + \sum_{i=1}^n P_i^4$$

## 5. Loci de estudio

En este estudio se analizaron los sistemas *PowerPlex*<sup>®</sup> 16 HS (**Figura 8**) y *PowerPlex*<sup>®</sup> ESX 17 (**Figura 9**), que permiten la co-amplificación y detección fluorescente de 16 loci (15 STR y amelogenina) y 17 loci (16 STR y amelogenina), respectivamente, incluyendo los loci del CODIS y los loci del conjunto de Estándares Europeos (ESS).



**Figura 8.** Configuración del sistema PowerPlex 16 HS. Este sistema permite la co-amplificación de D3S1358, TH01, D21S11, D18S51 y Penta E marcados con fluoresceína; D5S818, D13S317, D7S820, D16S539, CSF1PO y Penta D marcados con JOE; y Amelogenina, vWA, 18S1179, TPOX y FGA marcados con TMR-ET. El estándar interno (ILS 600) está marcado con el colorante CXR-ET y contiene 22 fragmentos de ADN de 60, 80, 100, 120, 140, 160, 180, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475, 500, 550 y 600 bases. Adaptado de: Butler, 2015.



**Figura 9.** Configuración del sistema PowerPlex ESX 17. Este sistema permite la co-amplificación de Amelogenina, D3S1358, TH01, D21S11 y D18S51 marcados con fluoresceína; D10S1248, D1S1656, D2S1338 y D16S539 marcados con JOE; D22S1045, vWA, D8S1179 y FGA marcados con TMR-ET; y D2S441, D12S391, D19S433 y SE33 marcados con CXR-ET. El estándar interno (ILS 500) está marcado con CC5 y contiene 21 fragmentos de ADN de 60, 65, 80, 100, 120, 140, 160, 180, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475 y 500 bases. Adaptado de: Butler, 2015.

## 6. Planteamiento del problema

El uso de la tecnología del ADN se ha vuelto más frecuente por lo que, cada vez se exigen resultados más confiables y precisos que eviten asociaciones fortuitas y nos permitan enfrentar los desafíos que representa el trabajar muestras severamente comprometidas, ya que de esto dependen situaciones importantes como el hallazgo de una persona desaparecida entre una gran cantidad de restos humanos. Se conocen actualmente sistemas que incluyen un mayor número de marcadores con el fin de aumentar el poder de discriminación y disminuir la probabilidad de coincidencia aleatoria. En este estudio se determinaron diversos parámetros estadísticos con respecto a los alelos de los sistemas *PowerPlex*<sup>®</sup> 16 HS y *PowerPlex*<sup>®</sup> ESX 17 para la población venezolana con el fin de demostrar si estos sistemas son buenas herramientas para atender a las demandas que impone la actual situación de nuestro país. Así también se



determinó si las frecuencias alélicas de todos los loci están en equilibrio Hardy-Weinberg para que puedan ser utilizados en los cálculos estadísticos.

## ANTECEDENTES

En los últimos años ha habido un gran avance en las técnicas de detección de individuos por su huella genética. Con el tiempo, las instituciones encargadas de impartir justicia en cada país van recopilando cada vez más huellas genéticas, por lo que es importante organizar toda esta información a fin de que sea útil para tiempos posteriores. Debido a esto, nacieron los bancos de ADN.

Dado que la primera base de datos de ADN nacional fue establecida en Reino Unido en abril de 1995, las bases de datos de ADN alrededor del mundo han revolucionado la habilidad de usar la información del perfil de ADN para vincular evidencias de escenas del crimen con perpetradores. Por lo cual, se ha producido un incremento en el número de países que han desarrollado una legislación específica para dar cobertura legal a la creación de bases de datos de ADN nacionales (Butler, 2009).

### 1. Estudios a nivel internacional

El 10 de abril de 1995 en Reino Unido surge la primera base de datos nacional de ADN a nivel mundial, usualmente referida como NDNAD (*National DNA Databank*), la cual almacenó originalmente datos de solo seis loci STR del SGM (*Second Generation Multiplex*) que consistían en FGA, TH01, VWA, D8S1179, D18S51, y D21S11. Para el año 1999, se realizó una ampliación del número de loci a 10 (los seis loci SGM más D3S1358, D16S539, D2S1338, y D19S433) utilizando el estuche comercial SGM Plus de la casa comercial *Applied Biosystems*. Para esa fecha se tenían planes para una posible

base de datos paneuropea que probablemente ampliarían la cantidad de loci necesarios hasta un número de 15 STR. A partir de esta base de datos pionera con fines criminalísticos se ha producido un incremento en el número de países que han desarrollado una legislación específica para dar cobertura legal a la creación de bases de datos de ADN (Butler, 2009), así como el incremento en el número de marcadores analizados, a modo de conocer sus frecuencias alélicas para uso forense.

Luego, la comunidad europea adoptó un conjunto estándar mínimo de loci, conocido como la ESS (*European Standard Set*), el cual incluía los siete loci FGA, TH01, VWA, D3S1358, D8S1179, D18S51, y D21S11. Para el año 2009, el Consejo de la Unión Europea extendió el ESS a un conjunto de 12 loci (los siete STR originales de la ESS, en adición de los 5 loci D1S1656, D2S441, D10S1248, D12S391, y D22S1045) para permitir el intercambio de datos de ADN entre las distintas bases de datos nacionales de ADN de los Estados miembros de la Unión Europea, y de esta manera aumentar la compatibilidad en las regiones STR estudiadas por los distintos laboratorios de genética forense europeos y reducir la posibilidad de coincidencias fortuitas. Asimismo, la Interpol ha adoptado el conjunto de loci ESS junto con el marcador de tipificación del sexo, la amelogenina, como información adicional opcional (Butler, 2009; García y col., 2012).

En Estados Unidos de Norteamérica, durante la reunión del Proyecto STR en noviembre de 1997, un núcleo de 13 loci STR fueron escogidos para ser la base de una futura base de datos nacional de ADN, el CODIS (*Combined DNA Index System*). Es entonces cuando el 13 de octubre de 1998, el FBI (*Federal Bureau of Investigation*)

oficialmente estableció su base de datos de ADN a nivel nacional para reforzar a las agencias que participan en NDIS (*U.S National DNA Index System*), la cual es una base de datos del CODIS que contiene más de 6,5 millones de perfiles de STR y vincula a los 50 estados en E.E.U.U de Norteamérica con la capacidad de buscar perfiles de ADN criminal. El CODIS consiste en una base y banco de datos electrónico de perfiles de ADN, la cual también está implementada en otros países (Butler, 2009).

Este núcleo de 13 loci incluyen a CSF1PO, FGA, TH01, TPOX, VWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, y D21S11, que al ser analizados presentan una probabilidad de coincidencia aleatoria promedio de más de uno en un trillón, entre individuos no relacionados. Estos loci son internacionalmente reconocidos como el estándar para la identificación en humanos. De estos 13 STR, los tres marcadores más polimórficos son FGA, D18S51 y D21S11, mientras que TPOX muestra la menor variación entre individuos (Butler, 2009).

Recientemente, Moretti y col. (2016) estudiaron la frecuencia alélica de 23 loci STR autosómicos en 11 poblaciones (caucásicos, hispanos del suroeste y del sudeste, afroamericanos, bahameños, jamaicanos, trinitenses, chamorros, filipinos, apaches y navajos), con la finalidad de expandir los datos proporcionados por el CODIS para análisis forenses en EE. UU. de Norteamérica.

Durante muchos años, lo laboratorios forenses de ADN europeos han estado utilizando muchos de los mismos loci STR utilizados por E.E.U.U de Norteamérica. En 2006, nuevos loci STR europeos fueron recomendados para ser incluidos en futuros

estuches comerciales de tipificación de STR, los cuales incluyen a D2S441, D10S1248, D22S1045, D1S1656 y D12S391 (Butler, 2009).

Los mismos marcadores STR se utilizan en muchos casos, por ejemplo, hay ocho loci STR (FGA, TH01, VWA, D3S1358, D8S1179, D16S539, D18S51, y D21S11) que superponen el conjunto de loci de las bases de datos de ADN europeas y estadounidenses. Este hecho permite la colaboración internacional en casos que lo justifiquen, aunque tener más loci en común sería beneficioso ya que las diversas bases de datos de ADN continúan creciendo en tamaño. Sin embargo, algunos países, como Alemania, utilizan marcadores STR (por ejemplo, el locus SE33, altamente polimórfico) que no se aplican tan ampliamente en otras bases de datos nacionales de ADN. Por lo tanto, la información de marcadores genéticos adicionales como el SE33 no puede ser utilizado para vincular la delincuencia transfronteriza potencial, ya que no estaría disponible en el conjunto de datos de ambos países (Butler, 2009).

En el año 2011, el proyecto IDNADEX (*Improving DNA Data Exchange*) valida dos sistemas para el análisis de perfiles de ADN con una compatibilidad global con las bases de datos nacionales de ADN a nivel internacional. Este sistema global comprende 21 STR que incluye tanto los 13 STR del sistema CODIS como los 12 STR del ESS para mejorar el poder de discriminación de las comparaciones genéticas en la base nacional de perfiles de ADN. Este proyecto se convirtió en el sistema de análisis de STR de uso forense con la mayor discriminación genética, así como el más compatible a nivel mundial (García y col.,2012)

Por su parte, Filho (2012) determinó la frecuencia alélica de 13 loci STR, y otros parámetros estadísticos (PD, PIC, PE,  $H_{obs}$ ,  $H_{esp}$ ,  $P-value$  y  $P_i$ ) de una población brasileña, con una muestra de 792 individuos. De esta manera reportó dos alelos que aún no han sido descritos en la base de datos del NIST (*The National Institute of Standards and Technology*), los cuales son el 24.3 del loci D21S11 y el 17 del loci D13S317. Además, comprobó que 6 de los loci STR estudiados estaban en equilibrio Hardy-Weinberg, y podrían ser incorporados en base de datos de frecuencias haplotípicas de la población brasileña, para su aplicación en casos forenses. Al realizar comparaciones de esta población con la población española, confirma que la población brasileña presenta índices de heterocigosidad ligeramente por encima de los valores de la población española, demostrando que la población brasileña presenta una diversidad génica mayor que las poblaciones con menor dinámica geográfica. Concluyendo que, el conjunto de los 13 loci STR presentan un alto poder de discriminación en ambas poblaciones y pueden ser aplicados en casos forenses.

Hernández y Trejo (2014) estudiaron las frecuencias alélicas de 15 marcadores STR presentes en la población de Zacatecas para aplicaciones forenses. De esta manera, crearon una base de datos, obteniendo perfiles genéticos para así determinar las frecuencias alélicas y genotípicas, encontrando alelos que no se habían reportado en poblaciones similares, empleando estos datos en casos forenses de manera exitosa y con resultados más exactos.

Seguidamente, Martínez y col. (2015) validaron el sistema de análisis genético *PowerPlex® 21* en 374 individuos mestizos de la región occidental de México, determinando las frecuencias alélicas y la distribución genotípica de estos los marcadores, los cuales se encontraban en equilibrio Hardy-Weinberg. Además, realizaron la prueba de desequilibrio de ligamiento para descartar asociaciones entre los pares de loci. También estimaron parámetros de interés forense, como la heterocigosidad, PIC e índice de paternidad típico, con un PD y PE combinado de -100% y 99,999999473%, respectivamente, lo cual constituye un incremento importante respecto a lo ofrecido por sistemas tradicionales con 15 STR para la misma población.

## **2. Estudios a nivel nacional**

En Venezuela también se han realizado estudios sobre estos marcadores genéticos para usos en genética forense. Chiurillo y col. (2003) analizaron la distribución alélica de 15 STR en una población caraqueña, que comprendía 255 sujetos no relacionados, 206 con los cuatro abuelos nativos de Venezuela y 49 muestreados al azar sin considerar el origen de sus abuelos. De esta manera no encontraron desviaciones significativas con la población de Maracaibo, Brasil ni Colombia. Sin embargo, al comparar con la base datos hispana de EE. UU. de Norteamérica se observaron desviaciones en las frecuencias alélicas de los loci D13S317 y D2S1338. Al igual que en un estudio realizado en Argentina, el loci D13S313 también presenta desviación en la frecuencia alélica. Para ese momento, las frecuencias alélicas de dos de los STR estudiados (D2S1338 y

D19S433) no habían sido reportados anteriormente, ni en ninguna otra población latinoamericana. De esta manera ellos concluyen que estos STR estudiados son marcadores útiles en pruebas de paternidad y casos forenses, además que la base de datos que construyeron podría ser una buena representación de la población total venezolana.

En el año 2006 también se analizaron los mismos 15 STR en 203 individuos no relacionados nacidos en Maracaibo, al noroeste de Venezuela, con al menos un padre nativo de la misma localidad. Las frecuencias alélicas fueron comparadas con otras poblaciones, sin encontrar diferencias estadísticamente significativas con respecto a la población caraqueña, aunque sí se encontraron diferencias al comparar con poblaciones peruana, mexicana y brasileña (Pineda y col., 2006).

A partir del estudio de Chiurillo y colaboradores, Lander y col., (2006) reportaron un nuevo alelo del marcador D21S11 en la población venezolana, el cual fue encontrado en un sujeto de sexo femenino. De esta manera se obtuvieron muestras de sangre de la familia del sujeto en estudio, se amplificó mediante el estuche comercial *AmpF/STR Identifiler* de la casa comercial *Applied Biosystems* y luego se procedió a secuenciar la región de interés que correspondía al loci D21S11. El estudio reveló que el linaje paterno contribuía al nuevo alelo.

De acuerdo con estos estudios se destaca la importancia de realizar el análisis de las frecuencias alélicas para los nuevos marcadores que se deseen incorporar en la base de datos poblacional con fines criminalísticos, debido a las diferencias que podría presentar



nuestra población con respecto a otras, y la presencia de alelos nuevos no descritos anteriormente en la literatura.

## **OBJETIVOS**

### **1. Objetivo General**

Analizar la estructura genética de una muestra poblacional venezolana utilizando 23 STR y construir una base de datos poblacional que pueda ser empleada en cálculos estadísticos para identificación humana y pruebas de paternidad.

### **2. Objetivos Específicos**

1. Obtener los perfiles de identidad genéticos característicos de una muestra de una población venezolana.
2. Determinar las frecuencias alélicas de 23 loci STR: D18S51, D21S11, TH01, D3S1358, D16S539, D2S1338, D1S1656, D10S1248, FGA, D8S1179, vWA, D22S1045, SE33, D19S433, D12S391, D2S441, Penta E, TPOX, Penta D, CSF1PO, D7S820, D13S317 y D5S818, en una población venezolana.
3. Estimar los parámetros de genética de poblaciones: poder de discriminación (PD), contenido de información polimórfica (PIC), poder de exclusión (PE), heterocigosidad observada ( $H_{obs}$ ), heterocigosidad esperada ( $H_{esp}$ ), equilibrio de Hardy-Weinberg y probabilidad de coincidencia ( $P_i$ ), para determinar el comportamiento de los marcadores en la población venezolana.
4. Determinar los alelos más y menos frecuentes en esta población venezolana.
5. Comparar la frecuencia alélica de los 23 loci STR obtenidos de la población venezolana con la frecuencia alélica de otra población estudiada.

## MATERIALES Y MÉTODOS

### 1. Muestra poblacional

En este trabajo se analizaron individuos no emparentados de la región del Distrito Capital y el Estado Miranda de Venezuela. Entre ellos, se eligió, de manera aleatoria, una muestra de voluntarios de diferente localización geográfica que fueron sometidos a una prueba para la toma de su ADN. La población estudiada comprendió un tamaño de 134 muestras de individuos.

El tamaño de muestra de 134 individuos corresponde a la información suministrada por la literatura en la cual, la mayoría de los datos poblacionales publicados incluyen el orden de 100 a 200 individuos por locus por población examinada.

Para la determinación de las frecuencias alélicas y parámetros estadísticos forenses de la población venezolana, se utilizaron un total de 23 loci STR, incluidos en los estuches comerciales *PowerPlex® ESX 17* y *PowerPlex® 16 HS* de la casa comercial *Promega Corporations*, como se muestran en la Tabla 1:

**Tabla 1:** 24 loci de estudio incluidos en estuches comerciales *PowerPlex® ESX 17* y *PowerPlex® 16 HS*

<b>PowerPlex® ESX 17</b>	<b>PowerPlex® 16 HS</b>	<b>ambos estuches comerciales</b>
D2S1338, D10S1248, SE33, D19S433, D12S391 y D2S441	D1S1656, Penta E, TPOX, Penta D, CSF1PO, D7S820, D13S317, D5S818	D18S51, D21S11, TH01, vWA, D3S1358, FGA, D8S1179, D16S539 y amelogenina

A continuación, se muestra cada uno de los loci amplificados con algunas características:

**Tabla 2.** Características de los STR que incluye el sistema *PowerPlex® 16 HS*

<b>Color</b>	<b>Locus</b>	<b>Localización en el cromosoma</b>	<b>Rango de tamaño en pb</b>	<b>Unidad de repetición</b>	<b>Número de alelos existentes</b>	<b>Variación alélica</b>
<i>Azul</i>	D3S1358	3p21.31	97-147	TCTA TCTG	24	8 a 20
	TH01	11p.15.5	146-274	AATG	21	3 a 14
	D21S11	21q21.1	138-273	TCTA TCTG	60	12 a 41.2
	D18S51	18q21.33	262- 396,37	AGAA	55	7 a 39.2
	Penta E	15q26.2	379-484	AAAGA	40	5 a 26
<i>Verde</i>	D5S818	5q23.2	115-178	AGAT	16	6 a 18
	D13S317	13q31.1	157-241	TATC	19	5 a 17
	D7S820	7q21.11	194-293	GATA	32	5 a 16
	D16S539	16q24.1	129-308	GATA	21	4 a 16
	CSF1PO	5q33.1	277-361	AGAT	23	5 a 16
	Penta D	21q22.3	370-454	AAAGA	40	1.1 a 18
<i>Amarillo</i>	vWA	12p13.31	122-212	TCTA TCTG	22	10 a 25
	D8S1179	8q24.13	123-255	TATC	19	7 a 20
	TPOX	2p25.3	98-302	AATG	17	4 a 16
	FGA	4q28	158-464	TTTC CTTT TTCC	97	12.2 a 51.2

Adaptado de: [http://strbase.nist.gov/str\\_fact.htm](http://strbase.nist.gov/str_fact.htm) [Consulta: 24 de Octubre de 2017]

**Tabla 3.** Características de los STR que incluye el sistema *PowerPlex® ESX 17*

<b>Color</b>	<b>Locus</b>	<b>Localización en el cromosoma</b>	<b>Rango de tamaño en pb</b>	<b>Unidad de repetición</b>	<b>Número de alelos existentes</b>	<b>Variación alélica</b>
<i>Azul</i>	D3S1358	3p21.31	97-147	TCTA TCTG	24	8 a 20
	TH01	11p.15.5	146-274	AATG	21	3 a 14
	D21S11	21q21.1	138-273	TCTA TCTG	60	12 a 41.2
	D18S51	18q21.33	262- 396,37	AGAA	55	7 a 39.2
<i>Verde</i>	D10S1248	10q26.3	79-271	GGAA	12	8 a 19
	D1S1656	1q42	121-169	TAGA TG	22	9 a 21
	D2S1338	2q35	277-359	TGCC TTCC	21	11 a 28
	D16S539	16q24.1	129-308	GATA	21	4 a 16
<i>Amarillo</i>	D22S1045	22q12.3	79-165	ATT	13	8 a 20
	vWA	12p13.31	122-212	TCTA TCTG	22	10 a 25
	D8S1179	8q24.13	123-255	TATC	19	7 a 20
	FGA	4q28	158-464	TTTC CTTT TTCC	97	12.2 a 51.2
<i>Rojo</i>	D2S441	2p14	78-114	TCTA	14	8 a 17
	D12S391	12p*	209-253	AGAT AGAC	15	15 a 26
	D19S433	19q12	92-150	AAGG AAAG TAGG	30	5.2 a 20
	SE33	6q14	197-343	AAAG	72	3 a 39.2

Adaptado de: [http://strbase.nist.gov/str\\_fact.htm](http://strbase.nist.gov/str_fact.htm) [Consulta: 24 de Octubre de 2017]

\*No se encuentra señalado en la referencia.

## **2. Criterios de inclusión.**

Los individuos no presentaban relaciones de parentesco biológico y fueron seleccionados al azar. Además, procedían de al menos dos generaciones anteriores originarias de Venezuela, es decir, ambos padres y los cuatro abuelos. Asimismo, se excluían a individuos que no dieron su consentimiento, que hayan recibido un trasplante de médula ósea y que no cumplieran con los criterios de inclusión.

## **3. Muestras biológicas.**

Se tomaron muestras de sangre periférica en soportes de tarjetas *Whatman FTA® Classic card*, luego de que los usuarios firmaron el respectivo formato de consentimiento autorizando el uso anónimo de sus muestras. Cada muestra tuvo asignado un código de números con el cual podría ser relacionado en el futuro.

Antes de obtener la muestra sanguínea mediante una punción distal con una lanceta, se limpió el área del dedo con un algodón impregnado con etanol al 70%. Luego, se depositaron aproximadamente 3-5 gotas de sangre sobre la tarjeta FTA.

Las muestras fueron trasladadas al Laboratorio de Genética Forense de la Unidad Criminalística Contra la Vulneración de Derechos Fundamentales, Área Metropolitana, del Ministerio Público, para su procesamiento.

## 4. Análisis de laboratorio.

### 4.1. Amplificación vía PCR-múltiplex

La amplificación de las muestras sobre tarjetas *FTA*<sup>®</sup> se realizó de manera directa, a partir de discos de 1,2 mm de diámetro de cada muestra, los cuales se obtuvieron con ayuda de un perforador *Mini-Punch Harris*. La amplificación del ADN se efectuó mediante PCR, siguiendo el protocolo del estuche comercial *PowerPlex*<sup>®</sup> *ESX 17* y *PowerPlex*<sup>®</sup> *16 HS*, ambos de la casa comercial *Promega Co.* (Promega, 2016; Promega, 2017A), los cuales permiten la co-amplificación de un total de 17 y 16 loci STR diferentes, respectivamente. Cabe destacar que ambos estuches comerciales se mantuvieron en un congelador a -20°C hasta el momento de su uso.

Las condiciones de la PCR, según el estuche comercial utilizado, se presentan en la tabla 2, para un volumen total de reacción de 12,5 µL por muestra, es decir, mitad de la mezcla de reacción de acuerdo con el protocolo, para el ahorro de los reactivos, lo cual ha sido previamente validado por estudios previos (Hoffman y Fenger, 2010). Previamente, todos los reactivos fueron descongelados, centrifugados brevemente y agitados por 15 segundos.

**Tabla 4:** Composición de la mezcla de reacción de PCR para una amplificación directa a partir de discos de tarjetas FTA®

<i>PowerPlex® ESX 17</i>	<i>PowerPlex® 16 HS</i>	<b>Volumen por reacción</b>
Agua grado biología molecular	Agua grado biología molecular	6,25 µL
<i>PowerPlex® ESX 5X Master Mix</i>	<i>PowerPlex® 16 HS 5X Master Mix</i>	2,50 µL
<i>PowerPlex® ESX 17 10X Primer Pair Mix</i>	<i>PowerPlex® 16 HS 10X Primer Pair Mix</i>	1,75 µL
<i>5X AmpSolution™ Reagent</i>	<i>5X AmpSolution™ Reagent</i>	2,50 µL
<b>Volumen total</b>		<b>12,5 µL</b>

Las concentraciones y componentes de cada reactivo son desconocidos.

Luego se agitó la mezcla de amplificación por 10 segundos, y se agregaron 12,5 µL de la misma en cada tubo de reacción, para luego añadir el disco de FTA. Para el control positivo de amplificación, se agregó a un tubo de reacción, con la mezcla de amplificación, 1 µL (5 ng) de ADN Control 2800M (10 ng/ µL) de la casa comercial *Promega Co.*, previamente diluido a una concentración de 5 ng/ µL. Mientras que, para el control negativo, se reservó un tubo de reacción conteniendo sólo la mezcla de amplificación.

Las amplificaciones se llevaron a cabo empleando un termociclador *MasterCycle Pro S* de *Eppendorff*, previamente verificado y calibrado, con los tiempos y temperaturas de ciclado presentados en las siguientes tablas (5 y 6), de acuerdo con el procedimiento recomendado por el estuche comercial empleado:



**Tabla 5:** Protocolo del termociclador para el estuche comercial *PowerPlex® ESX 17*.

<b>Temperatura</b>	<b>Tiempo</b>	<b>Número de ciclos</b>
96°C	2 min	1 ciclo
94°C	30 seg	
59°C	2 min	24 ciclos
72°C	90 seg	
60°C	45 min	1 ciclo
4°C	Mantener hasta el momento de su análisis	

**Tabla 6:** Protocolo del termociclador para el estuche comercial *PowerPlex® 16 HS*.

<b>Temperatura</b>	<b>Tiempo</b>	<b>Número de ciclos</b>
96°C	2 min	1 ciclo
94°C	30 seg	
60°C	30 seg	10 ciclos
70°C	45 seg	
90°C	30 seg	
60°C	30 seg	17 ciclos
70°C	45 seg	
60°C	30 min	1 ciclo
4°C	Mantener hasta el momento de su análisis	

Finalmente, todos los productos de amplificación obtenidos de la reacción en cadena de la polimerasa presentaron un marcaje en uno de sus cebadores, con fluorocromos, que posteriormente fueron detectados en el analizador genético.

#### **4.2. Detección y asignación de alelos STR**

El análisis de los amplicones se llevó a cabo mediante la separación por tamaño de los fragmentos de ADN por electroforesis capilar, en el equipo *Genetic Analyzer*

(analizador genético) *ABI 3130xl* de la casa comercial *Applied Biosystems®*, con *multiarrrays* de 16 capilares simultáneos.

Se preparó una mezcla de formamida desionizada (*Ultra Pure Grade*) y del marcador de tamaño o ILS (*Internal Line Standard*), correspondiente a cada estuche comercial y la proporción de cada reactivo se preparó de la siguiente manera:

**Tabla 7.** Composición de la mezcla de carga de las muestras en placa en el analizador genético *3130xl*.

<b>Componentes</b>	<b>Volumen por reacción</b>	
	<b>PowerPlex® 16 HS</b>	<b>PowerPlex® ESX 17</b>
<b>CXR ILS 600</b>	0,5 µL	--
<b>CC5 ILS 500</b>	--	1,0 µL
<b>Hi-Di Formamida</b>	9,5 µL	10,0 µL

Ambos estuches comerciales se conservaron en el congelador a  $-20^{\circ}\text{C}$  hasta el momento de su uso, todos los reactivos fueron descongelados, centrifugados brevemente y agitados por 15 segundos, previo a su uso.

Luego, la mezcla de carga se agitó durante 15 segundos y enseguida se dispensó en una placa de 96 pozos. Posteriormente se agregaron 1,0 µL de cada producto de PCR en el pozo correspondiente y 1,0 µL de la escalera alélica correspondiente a cada estuche comercial, en cada uno de los pozos destinados para el mismo.

Al llevar cada muestra un estándar de tamaño se evitan errores en la asignación del tamaño de los fragmentos analizados debidos a diferencias en la movilidad electroforética

entre las distintas inyecciones (ya sea por variaciones en la temperatura, de concentración de polímero o en las condiciones de electroforesis). Con el estándar de tamaño utilizado para el sistema *PowerPlex® 16 HS* es posible analizar fragmentos de entre 60-600 pb, mientras que con el sistema *PowerPlex® ESX 17* es posible analizar fragmentos de entre 60-500 pb.

Una vez preparadas las muestras en la placa, se selló con una septa y se desnaturalizó a 95°C durante 3 minutos, se colocó en hielo por 3 minutos más y se centrifugó para luego colocar la placa en su posición adecuada dentro del soporte que va acoplado al distribuidor de muestras del analizador genético. Luego se procedió a realizar la hoja de corrida en el software *Data Collection* de la firma *ABI*.

Antes de comenzar la corrida electroforética se establecieron los siguientes parámetros:

**Tabla 8.** Condiciones de corrida electroforética en el analizador genético *3130xl* de *Applied Biosystems*.

<b>Parámetros</b>	<b>PowerPlex® 16 HS</b>	<b>PowerPlex® ESX 17</b>
<i>Tiempo de inyección</i>	5 seg	2 seg
<i>Voltaje de inyección</i>	3,0 KVolt	3,0 KVolt
<i>Voltaje de electroforesis</i>	15,0 KVolt	15,0 KVolt
<i>Temperatura de electroforesis</i>	60°C	60°C
<i>Tiempo de electroforesis</i>	4.000 seg	3.200 seg

- El voltaje y el tiempo de inyección pueden ser variados para regular la cantidad de ADN que entra en el capilar. Así, en caso de obtener una señal débil en una muestra es posible aumentar el tiempo de inyección para permitir que se inyecte más cantidad de producto amplificado al capilar y que aumente la señal de fluorescencia. No obstante, un tiempo de inyección excesivo puede originar una pérdida de resolución.
- El voltaje de electroforesis para fragmentos analizados utilizando un polímero *POP-4* es de 319 V/cm. Para un capilar de 50 cm esto supone un voltaje de aproximadamente 16.000 V y la corriente a este voltaje es de alrededor de 9  $\mu$ A. bajo estas condiciones electroforéticas el tiempo necesario para que un fragmento de 400 pb llegue a la ventana de detección es de unos 25 minutos aproximadamente. Si se aumenta el voltaje, el tiempo requerido es menor, pero se pierde resolución.
- La temperatura de electroforesis para detección fragmentos es de 60°C, distinta a la temperatura de secuenciación que es de 50°C.
- Por último, el tiempo de electroforesis puede acortarse cuando solo se requiere información de productos de amplificación de pequeño tamaño. Generalmente, este tiempo programado es un 10% mayor a la media del tiempo necesario para que el fragmento de mayor tamaño migre lo suficiente.

Una vez alcanzada la temperatura de inyección, la electroforesis inicia.

Las señales generadas por la excitación de moléculas de fluorocromos que se encuentran acoplados químicamente a estos amplicones fueron detectadas a través de un láser de Argón acoplado a una cámara CCD, que a su vez realiza la transmisión de

los datos desde el equipo a una computadora, los cuales son registrados a través del software *Data Collection*.

### **4.3. Análisis de resultados**

El análisis de resultados consistió en evaluar la calidad de los electroferogramas en términos de intensidad de la señal registrada (unidades de fluorescencia, UF), tiempo de lectura y tamaño de los amplicones detectados. Posterior a esto, los datos brutos o *raw data* fueron procesados a través de un segundo software, *Gene Mapper ID-X v3.2* de la casa comercial *Applied Biosystems*, que permitió la cuantificación en pares de bases de todos los picos. Como cada muestra llevó un estándar interno de tamaño, éste migra a la vez que el producto amplificado y bajo las mismas condiciones, evitando el problema de las variaciones entre muestras.

Asimismo, el software permite la asignación de los alelos de acuerdo con el sistema STR correspondiente y su reporte final como perfil genético, mediante la comparación con los alelos de la escalera. Cada estuche comercial contiene un *allelic ladder* o escalera alélica que permite la asignación precisa de los alelos (aceptándose una variación aproximada  $\pm 0,5\text{pb}$ ).

El software genera una línea (*dye / lane list*) para cada fluorocromo en la que están incluidos el tamaño, cantidad y la información de todos los fragmentos marcados con un mismo fluorocromo, representado gráficamente en un electroferograma.

Cabe destacar que, antes de realizar estos experimentos, se empleó el protocolo de validación interna (Promega, 2013) de ambos sistemas de amplificación autosomal (*PowerPlex® 16 HS* y *PowerPlex® ESX 17*), de manera que se estimó que con un mínimo de 500 pg de ADN por reacción era óptimo para trabajar ya que todos los alelos se encontraban dentro de los parámetros de sensibilidad y además, se determinó un umbral analítico de 52 RFU para el estuche comercial *PowerPlex® 16 HS* y de 129 RFU para el estuche comercial *PowerPlex® ESX 17* esto con el propósito de considerar los picos como alelos y no como ruido.

## **5. Análisis estadísticos.**

En el presente estudio, se analizaron las frecuencias alélicas y genotípicas, permitiendo verificar el grado de polimorfismo de los STR estudiados, caracterizando la variabilidad genética de las poblaciones.

De este modo, para determinar la variabilidad de los marcadores genéticos de la población venezolana, se calcularon, además de las frecuencias alélicas, otros parámetros estadísticos forenses. Estos índices de diversidad han sido computados con el programa *Genepop® 4.2* (Rousset, 2008).

El objetivo de *Genepop®* es proporcionar el medio, en genética poblacional, con un amplio conjunto de métodos y pruebas estadísticas para poder obtener toda la información sobre aspectos genéticos y demográficos, en un conjunto de muestras

poblacionales. Así pues, con *Genepop*® se pueden tratar una amplia gama de datos, como pueden ser secuencias de ADN, RFLP, STR, frecuencias alélicas, entre otros.

Los siguientes parámetros estadísticos indican la utilidad *a priori* de los STR estudiados:

### **5.1. Frecuencias alélicas**

Para determinar este parámetro, se utilizó el programa bioinformático *Power Stats V12.xls*® (*Promega Corporation*®, *Madison, USA*).

### **5.2. Poder de discriminación (PD)**

En este trabajo, PD fue calculado con el programa bioinformático *Power Stats V12.xls*® (*Promega Corporation*®, *Madison, USA*).

### **5.3. Contenido de Información Polimórfica (PIC)**

En este trabajo PIC fue calculado con el programa bioinformático *Power Stats V12.xls*® (*Promega Corporation*®, *Madison, USA*).

### **5.4. Poder de Exclusión (PE)**

En este estudio, PE fue calculado con el programa bioinformático *Power Stats V12.xls*® (*Promega Corporation*®, *Madison, USA*).

### **5.5. Heterocigosidad**

En este estudio,  $H_{obs}$  fue calculado utilizando el programa bioinformático *Power Stats V12.xls*® (*Promega Corporation*®, *Madison, USA*), mientras que, para  $H_{esp}$  el programa *Genepop*® (Rousset, 2008).



## **5.6. Equilibrio de Hardy-Weinberg**

En este trabajo, el desvío del equilibrio de Hardy-Weinberg fue calculado mediante el test exacto (H1: exceso de heterocigotos) del programa bioinformático *Genepop*® (Rousset, 2008), aplicando el procedimiento de Cadenas de Markov (Guo y Thompson, 1992) con 1.000 dememorizaciones, 100 lotes y 1.000 iteraciones por lote para determinar las probabilidades.

## **5.7. Probabilidad de coincidencia aleatoria (Pi)**

En este trabajo, Pi fue calculado con el programa bioinformático *Power Stats V12.xls*® (*Promega Corporation*®, *Madison, USA*).

## **6. Consentimiento informado.**

Todos los datos provenientes de los sujetos de la muestra poblacional como el sexo, origen geográfico del individuo y perfiles genéticos son anónimos en todos los documentos publicados. Se utilizó un formato de consentimiento informado, donde se autorizó la toma voluntaria de su muestra sanguínea y el uso de su perfil genético para estudios genético-poblacionales, una vez informado del compromiso institucional de mantener los datos anónimos. Cabe destacar que las muestras forman parte exclusivamente del presente trabajo.

En resumen, se les explicó a las personas voluntarias de este estudio, en términos sencillos, la razón por la cual se recolecta el material biológico, de manera que comprendieran los siguientes aspectos:

- El por qué se están recolectando muestras.
- Aspectos prácticos de la participación en el estudio.
- El modo en que se manejan y utilizarán los datos relacionados con los individuos y el hecho de que se respetarán los principios de la protección de datos personales.
- Detalles del formulario de consentimiento que llenaron.
- Por último, que la participación es voluntaria y que, si cambian de idea más adelante, podrían retirarse del estudio sin estar obligados a dar explicaciones.

El propósito del consentimiento informado tiene por objeto impedir que las personas proporcionen muestras bajo coacción (Anexo 1).

## **7. Comparación de la frecuencia alélica entre la población venezolana en estudio y otras poblaciones**

La comparación de las frecuencias alélicas entre la población venezolana en estudio y otra población fue realizada con el fin de detectar la existencia de diferencias entre ambas poblaciones para cada marcador analizado. De este modo, los resultados de las frecuencias alélicas y de los parámetros estadísticos forenses obtenidos en este estudio para la población venezolana fueron comparados con otros estudios realizados para la

población hispana de Estados Unidos de Norteamérica. Para esta comparación se utilizó el siguiente estudio publicado en una revista científica: Hill y col. (2013), quienes analizaron 29 marcadores STR de varias poblaciones para un número total de 1036 individuos, de los cuales 236 eran hispanos. De esta manera, los parámetros estadísticos comparados entre la población venezolana e hispana norteamericana fueron los siguientes: heterocigosidad observada ( $H_{obs}$ ), probabilidad de coincidencia ( $P_i$ ), probabilidad de exclusión ( $PE$ ) y el test exacto de equilibrio Hardy-Weinberg ( $p$ -value). Además, se comparó la variabilidad de alelos de cada marcador. Cabe destacar que de los 29 loci STR estudiados por estos autores, la comparación solo se hizo con los 23 marcadores analizados en el presente estudio.

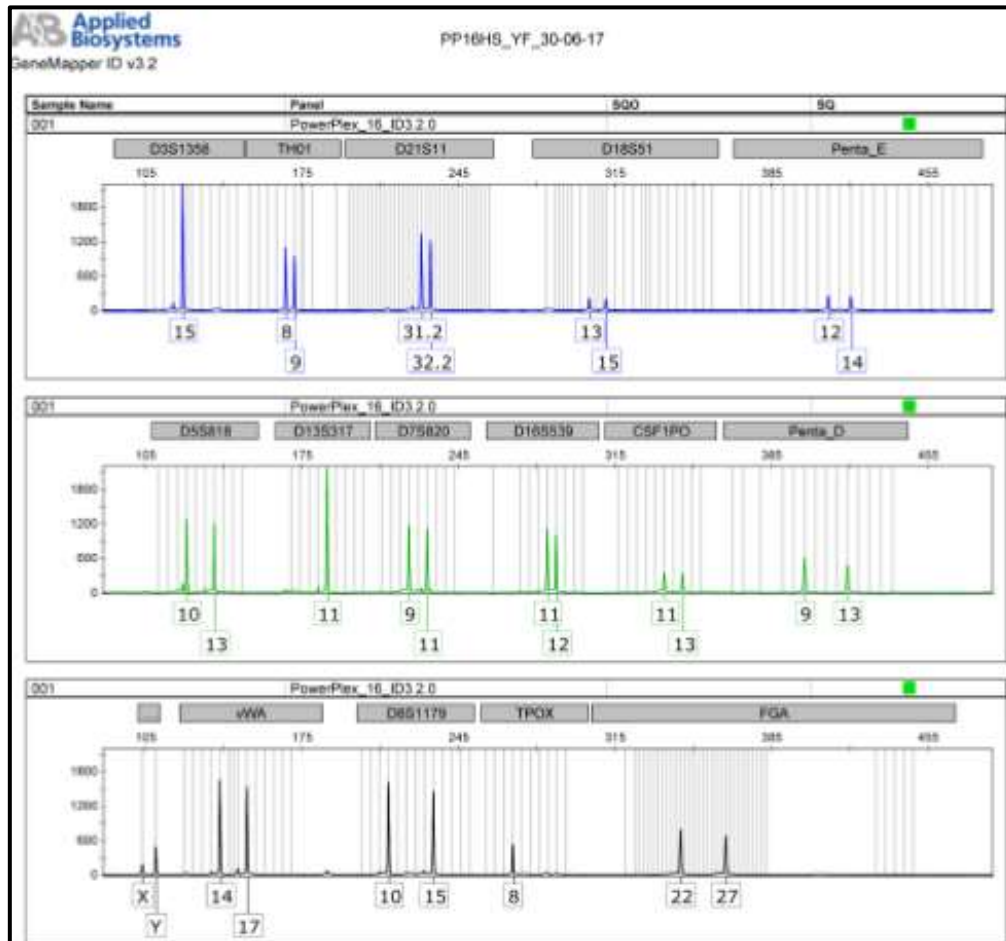
## RESULTADOS

### 1. Frecuencias alélicas de los 23 loci STR y parámetros estadísticos de la población venezolana

En esta sección del trabajo se procede a la exposición ordenada y esquematizada de todos los resultados obtenidos para el estudio poblacional de los loci: D18S51, D21S11, TH01, D3S1358, D16S539, D2S1338, D1S1656, D10S1248, FGA, D8S1179, vWA, D22S1045, SE33, D19S433, D12S391, D2S441, Penta E, TPOX, Penta D, CSF1PO, D7S820, D13S317 y D5S818.

Cabe destacar que en 1992 Ranajit Chakraborty concluyó que de 100 a 150 individuos por población podrían proporcionar un muestreo adecuado para un locus genético siempre que las frecuencias alélicas por debajo de 0,01 no fueran utilizadas en cálculos forenses. Otros han llegado a conclusiones similares, que 100 a 120 individuos por locus por población son suficientes para los cálculos de probabilidades. La recopilación de información de más muestras solo mejora la precisión de las estimaciones de frecuencia para alelos raros (Butler, 2015).

En la figura 10 se observa un electroferograma obtenido, a partir del sistema *PowerPlex® 16 HS*, de una de las muestras analizadas, en él se encuentran 15 marcadores autosómicos polimórficos y amelogenina (para la determinación del sexo). En cada locus se observan picos que corresponden a los alelos. Los dos alelos (provenientes uno de origen paterno y el otro de origen materno) pueden ser iguales (homocigoto) o diferentes (heterocigoto).



**Figura 10.** Electroferograma de una muestra utilizando el sistema *PowerPlex 16 HS*.

En la tabla 9 se muestran las frecuencias alélicas observadas para los 23 marcadores STR evaluados en la población venezolana. Se observan los distintos valores obtenidos de las frecuencias alélicas, los cuales se encuentran en un rango de 0,0037 y 0,4179, así como también los alelos de cada marcador, algunos marcadores exhiben más alelos con respecto al resto, como es el caso del marcador SE33. Estas frecuencias son utilizadas para calcular diversos parámetros que son de importancia en la genética forense.

**Tabla 9.** Frecuencias alélicas obtenidas de los marcadores STR evaluados en la población venezolana.

Frecuencias alélicas	CSFIPO	D13S317	D16S539	D18S51	D21S11	D3S1358	D5S818	D7S820	D8S1179	FGA	Penta D	Penta E	TH01	TPOX	vWA	D10S1248	D12S391	D19S433	D1S1656	D22S1045	D2S1338	D2S441	SE33
2.2											0.0155												
5											0.0194	0.0672											
6													0.2331	0.0112									
7		0.0075					0.0373	0.0299			0.0155	0.1119	0.2406										
7.1			0.0037																				
8		0.0037	0.1269	0.0112			0.0037	0.1381	0.0075		0.0233	0.0448	0.1278	0.4179		0.0037				0.0037		0.0037	
9		0.0261	0.1082	0.1866	0.0037		0.0448	0.0896	0.0037		0.2248	0.0224	0.2030	0.1343								0.0037	
9.3													0.1955										
10		0.2425	0.0560	0.1418	0.0037		0.0522	0.2687	0.0821		0.1977	0.0522		0.0597				0.0042	0.0042	0.0149		0.2612	
10.3								0.0037															
11		0.2873	0.2575	0.2351	0.0037		0.3694	0.2649	0.0709		0.1434	0.0933		0.2649		0.0336		0.0250	0.0375	0.0746		0.3582	
11.2																		0.0125					
11.3																						0.0410	
12		0.3843	0.2425	0.2873	0.1119		0.3582	0.1716	0.1679		0.1705	0.1978		0.1045		0.0634		0.0875	0.1167	0.0224		0.0485	
12.2																		0.0042					0.0053
13		0.0336	0.1493	0.1231	0.1157	0.0075	0.1269	0.0299	0.2537		0.1279	0.1082		0.0075	0.0037	0.2052		0.2500	0.0750		0.0112	0.0053	
13.2																		0.0792					
14		0.0149	0.0485	0.0149	0.1828	0.1194	0.0075	0.0037	0.2910	0.0543	0.0634				0.0858	0.3433		0.3042	0.1292	0.0448	0.2463	0.0368	
14.3																		0.0083					
15		0.0075		0.1418		0.3358			0.0821			0.0672			0.1045	0.2201	0.0610	0.1542	0.2333	0.3843	0.0261	0.0579	
15.2																		0.0458					
15.3																			0.0750				
16				0.1418		0.2687			0.0336	0.0078	0.0410				0.2985	0.1194	0.0203	0.0333	0.1208	0.3246	0.0448	0.0895	
16.3																						0.0050	
17				0.1381		0.1567			0.0037			0.0522			0.2500	0.0112	0.0894		0.0500	0.1231	0.1866	0.0842	
17.3																	0.0041		0.0917				
18				0.0522		0.0933			0.0037	0.0075	0.0187				0.1716		0.2033			0.0075	0.0597	0.1263	
18.2										0.0038													
18.3																							
19				0.0522		0.0149				0.0714	0.0187				0.0709		0.1911		0.0583		0.1791	0.1053	
20				0.0261		0.0037				0.1090	0.0224				0.0149		0.1992				0.1306	0.0789	
20.2																						0.0053	
21				0.0187						0.1353	0.0149						0.0650				0.0299	0.0368	
21.2				0.0037						0.0038													
22				0.0037						0.1429							0.0691				0.1045	0.0158	
22.2																						0.0105	
23										0.1053	0.0037						0.0610				0.1231		
24										0.1128							0.0122				0.0821		
24.2																							0.0263
25										0.1617							0.0081				0.0522		
25.2																							0.0632
25.3										0.0038													
26										0.0865							0.0041				0.0075		
26.2																							0.0421
27					0.0354					0.0414													
27.2																							0.0632
28					0.1566					0.0075													
29					0.2222					0.0038													
29.2																							0.0947
30					0.2828					0.0038													
30.2					0.0253																		0.0526
31					0.0455																		
31.2					0.1667																		
33.1					0.0051																		
33.2					0.0556																		
35					0.0051																		

En la tabla 10 se reportan los *p-value* calculados al analizar los datos obtenidos para los 23 STR, para lo cual se obtuvo que todos los marcadores evaluados se encuentran en equilibrio Hardy-Weinberg, lo que indica que las frecuencias alélicas obtenidas en estos marcadores se pueden utilizar para el cálculo de diversos parámetros del área forense. Los resultados de un test exacto para determinar el equilibrio Hardy-Weinberg indicaron que ninguno de los marcadores presentó una desviación significativa del equilibrio Hardy-Weinberg (*p-value* < 0,05). Todos los loci STR presentaron valores de *P* que oscilaron en un rango de 0,0795 - 0,9856, indicando el equilibrio Hardy-Weinberg para estos marcadores.

Los valores de *p-value* mayores a 0,05 indican la aceptación de la hipótesis nula que implica el cumplimiento del modelo Hardy-Weinberg, es decir, no existe diferencias significativas en las frecuencias alélicas, ya que el valor obtenido se encuentra dentro del intervalo de confianza.

**Tabla 10.** Cálculos y resultados para evaluar el equilibrio Hardy-Weinberg entre los 23 loci estudiados en la población venezolana.

<b>Locus</b>	<b>P-value</b>	<b>Resultado</b>
<b>CSF1PO</b>	0,9284	Equilibrio
<b>D13S317</b>	0,9175	Equilibrio
<b>D16S539</b>	0,4409	Equilibrio
<b>D18S51</b>	0,6151	Equilibrio
<b>D21S11</b>	0,2379	Equilibrio
<b>D3S1358</b>	0,8933	Equilibrio
<b>D5S818</b>	0,9856	Equilibrio
<b>D7S820</b>	0,5594	Equilibrio
<b>D8S1179</b>	0,1318	Equilibrio
<b>FGA</b>	0,0795	Equilibrio
<b>Penta D</b>	0,2771	Equilibrio
<b>Penta E</b>	0,9574	Equilibrio
<b>TH01</b>	0,0829	Equilibrio
<b>TPOX</b>	0,1072	Equilibrio
<b>vWA</b>	0,6098	Equilibrio
<b>D10S1248</b>	0,6569	Equilibrio
<b>D12S391</b>	0,6222	Equilibrio
<b>D19S433</b>	0,6896	Equilibrio
<b>D1S1656</b>	0,1768	Equilibrio
<b>D22S1045</b>	0,9418	Equilibrio
<b>D2S1338</b>	0,0984	Equilibrio
<b>D2S441</b>	0,3014	Equilibrio
<b>SE33</b>	0,1647	Equilibrio

En la tabla 11 se muestran los valores obtenidos de manera individual de los parámetros de interés en genética forense para cada uno de los 23 marcadores STR. Como se observa los marcadores SE33 y Penta E fueron los que tuvieron índices más altos, por lo que la presencia de estos loci en el sistema proporciona mucho peso estadístico al momento de realizar los cálculos de una confronta de perfiles.



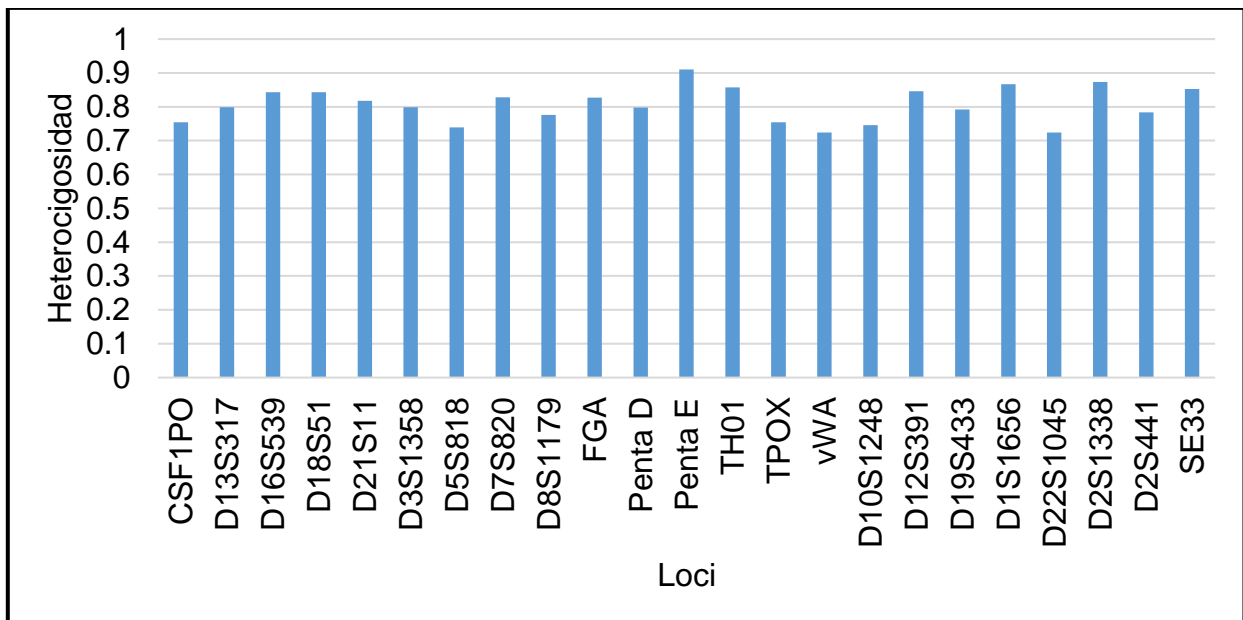
Analizando los índices de heterocigosidad observada ( $H_{obs}$ ) y esperada ( $H_{esp}$ ), se observó que, para todos los marcadores, los índices de  $H_{esp}$  están muy próximos a aquellos observados.

**Tabla 11.** Parámetros estadísticos de interés forense para los 23 loci estudiados en la población venezolana.

<b>Locus</b>	<b><math>H_{esp}</math></b>	<b><math>H_{obs}</math></b>	<b><math>P_i</math></b>	<b><math>PD</math></b>	<b><math>PIC</math></b>	<b><math>PE</math></b>
<b>CSF1PO</b>	0,712	0,754	0,146	0,854	0,66	0,516
<b>D13S317</b>	0,823	0,799	0,058	0,942	0,80	0,596
<b>D16S539</b>	0,795	0,843	0,084	0,916	0,76	0,682
<b>D18S51</b>	0,878	0,843	0,033	0,967	0,86	0,682
<b>D21S11</b>	0,805	0,818	0,068	0,932	0,79	0,633
<b>D3S1358</b>	0,771	0,799	0,091	0,909	0,73	0,596
<b>D5S818</b>	0,716	0,739	0,137	0,863	0,67	0,491
<b>D7S820</b>	0,800	0,828	0,075	0,925	0,77	0,653
<b>D8S1179</b>	0,806	0,776	0,067	0,933	0,78	0,555
<b>FGA</b>	0,885	0,827	0,030	0,970	0,87	0,650
<b>Penta D</b>	0,840	0,798	0,051	0,949	0,82	0,596
<b>Penta E</b>	0,907	0,910	0,020	0,980	0,90	0,817
<b>TH01</b>	0,795	0,857	0,090	0,910	0,76	0,709
<b>TPOX</b>	0,725	0,754	0,134	0,866	0,68	0,516
<b>vWA</b>	0,799	0,724	0,067	0,933	0,77	0,466
<b>D10S1248</b>	0,775	0,746	0,091	0,909	0,74	0,503
<b>D12S391</b>	0,850	0,846	0,041	0,959	0,84	0,686
<b>D19S433</b>	0,790	0,792	0,068	0,932	0,78	0,584
<b>D1S1656</b>	0,863	0,867	0,036	0,964	0,86	0,728
<b>D22S1045</b>	0,726	0,724	0,123	0,877	0,68	0,466
<b>D2S1338</b>	0,877	0,873	0,039	0,961	0,86	0,741
<b>D2S441</b>	0,735	0,784	0,132	0,868	0,69	0,569
<b>SE33</b>	0,898	0,853	0,020	0,980	0,92	0,700

$H_{esp}$ : Heterocigosidad esperada,  $H_{obs}$ : Heterocigosidad observada,  $P_i$ : Probabilidad de coincidencia aleatoria,  $PD$ : Poder de discriminación,  $PIC$ : Índice de contenido polimórfico,  $PE$ : Poder de exclusión.

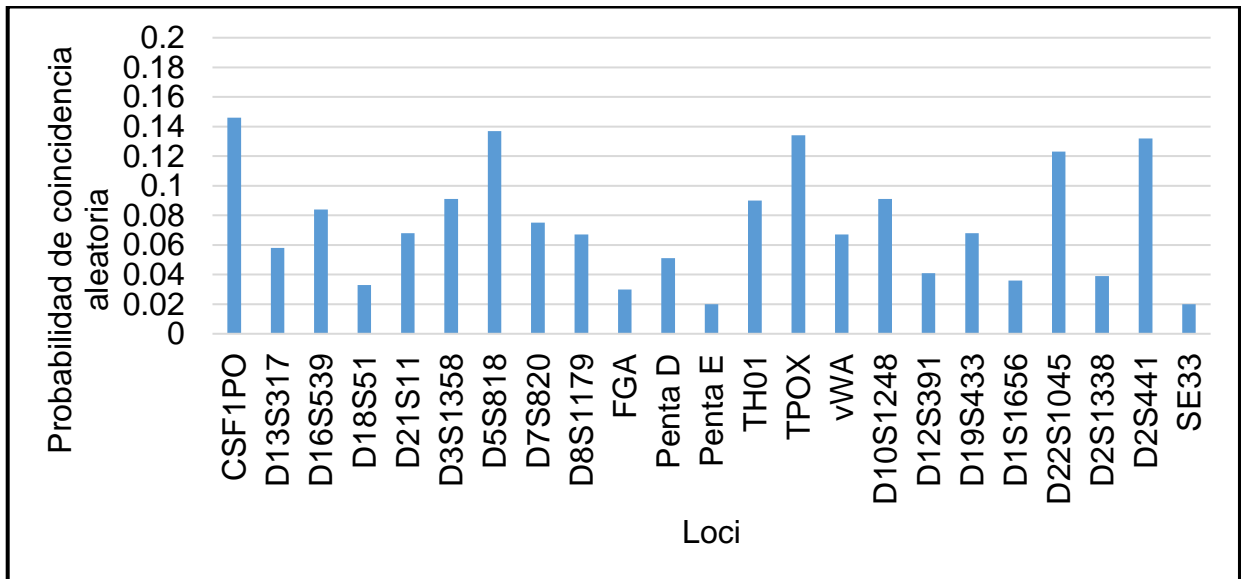
En la figura 11 se muestran los valores obtenidos de la Heterocigosidad observada, en la cual el marcador Penta E (0,910) es el que presenta la mayor heterocigosidad observada, aportando un mayor poder de discriminación al sistema, seguido del D2S1338 (0,873) y del D1S1656 (0,867). Al contrario, los marcadores vWA y D22S1045 (0,724) presentan el valor más bajo, seguido del D5S818 (0,738) y del D10S1248 (0,746). Asimismo, el valor promedio de heterocigosidad observada para la población total es de 0,807.



**Figura 11.** Heterocigosidad observada de los marcadores STR estudiados en la población venezolana.

La figura 12 muestra los valores de la probabilidad de coincidencia aleatoria, que indica la probabilidad de que dos individuos seleccionados al azar tengan un genotipo idéntico en el locus probado. Se observa que los marcadores Penta E y SE33 (0,020) son

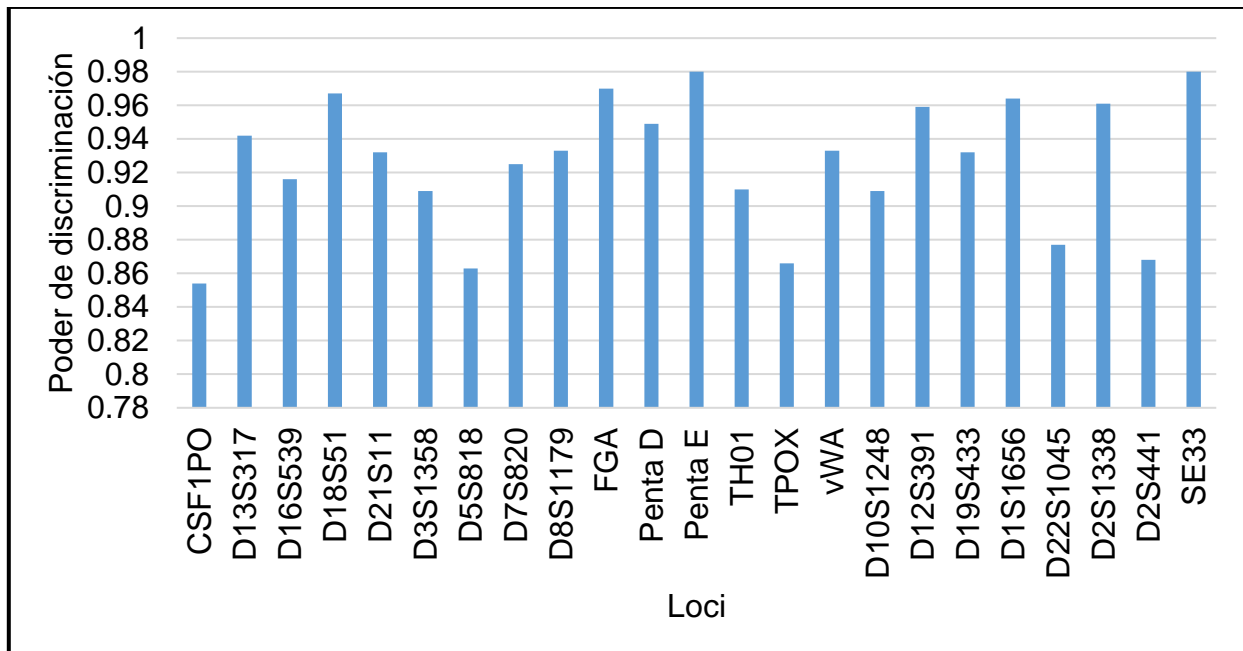
los que presentan el valor más bajo, haciendo que disminuyan las asociaciones fortuitas, seguidos del FGA (0,030) y del D18S51 (0,033). Mientras que el marcador CSF1PO (0,146) presenta el valor más alto, seguido del D5S818 (0,137) y del TPOX (0,134).



**Figura 12.** Probabilidad de coincidencia aleatoria de los marcadores STR estudiados en la población venezolana.

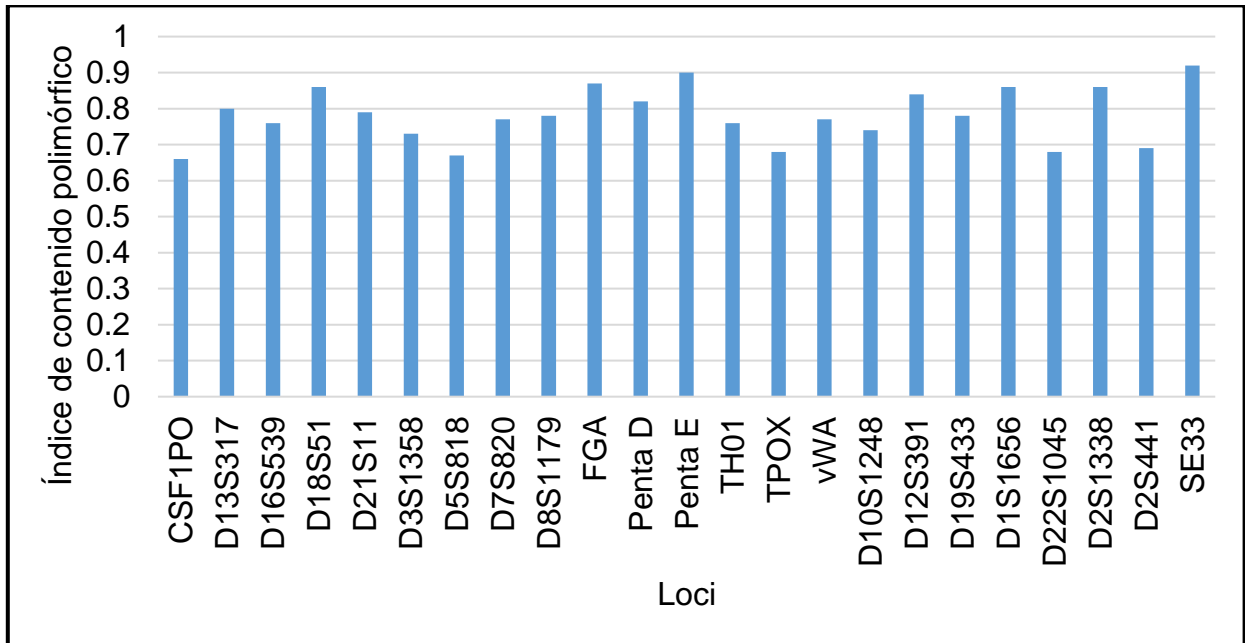
Asimismo, en la figura 13 se muestra el poder de discriminación de los marcadores STR estudiados en la población venezolana, los cuales presentaron valores altos y al combinarse dan como resultado un enorme poder de discriminación, de tal manera que dos personas puedan ser fácilmente diferenciados genéticamente, al disminuir al mínimo la posibilidad de que existan asociaciones fortuitas. Los marcadores Penta E y SE33 (0,980) presentaron el índice más alto, seguidos del FGA (0,970) y del D18S51 (0,967). Mientras que el marcador CSF1PO (0,854) presentó el índice de PD más bajo, seguido

del D5S818 (0,863) y del TPOX (0,866). Asimismo, el valor promedio de poder de discriminación observada para la población total es de 0,926.



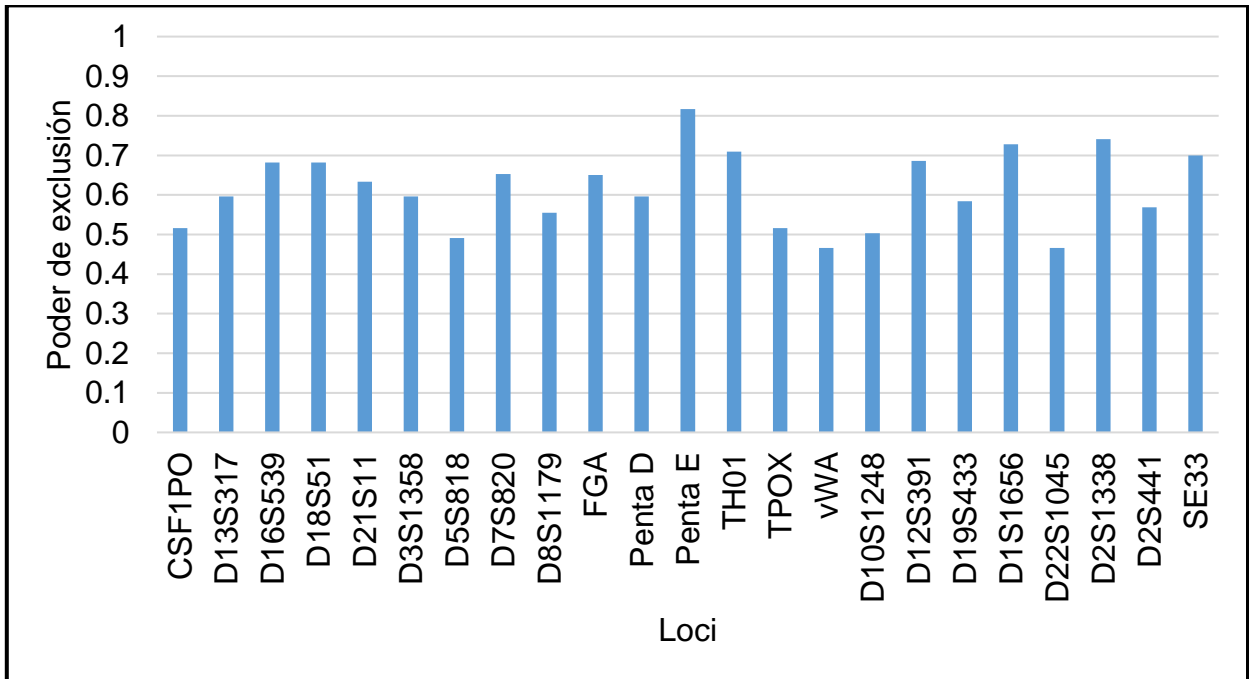
**Figura 13.** Poder de discriminación de los marcadores STR estudiados en la población venezolana.

Por otra parte, en la figura 14 se presentan los valores obtenidos del índice de contenido polimórfico, el cual expresa el grado de diversidad biológica que posee cada marcador. Para la población venezolana, el valor promedio de contenido de información polimórfica observado fue de 0,782. Asimismo, el locus SE33 (0,92) es el que posee un mayor polimorfismo, y el que ofrece una mayor información, seguido del Penta E (0,90) y del FGA (0,87). Al contrario, el marcador CSF1PO (0,66) presenta el valor más bajo, seguido del D5S818 (0,67) y del D22S1045 (0,68). Como se puede observar los valores entre los marcadores son similares.



**Figura 14.** Índice de contenido polimórfico de los marcadores STR estudiados en la población venezolana.

Finalmente, en la figura 15 se exhiben los valores obtenidos del poder de exclusión, el cual indica la posibilidad de excluir a una persona falsamente imputada. Para la población venezolana, el valor promedio de poder de exclusión observado fue de 0,615. Al igual que en parámetros anteriores, el marcador más informativo fue el Penta E (0,817), quien tuvo el valor más alto de PE, comprobando una vez más la utilidad de este marcador para el sistema, seguido del D2S1338 (0,741) y del D1S1656 (0,728). Mientras que los marcadores menos informativos son vWA y D22S1045 (0,466) los cuales obtuvieron el valor más bajo, seguidos del D5S818 (0,491) y del D10S1248 (0,503).



**Figura 15.** Poder de exclusión de los marcadores STR estudiados en la población venezolana.

## 2. Marcadores STR analizados en la población venezolana.

En la tabla 12 se expone el número total de alelos para cada locus y los alelos más y menos frecuentes para cada uno, así como sus frecuencias alélicas respectivas.

**Tabla 12.** Número total de alelos observados y alelos más y menos frecuentes para cada locus de estudio en la población venezolana

<b>Locus</b>	<b>Número total de alelos observados</b>	<b>Alelos más frecuentes</b>	<b>Alelos menos frecuentes</b>
<b>CSF1PO</b> (Anexo 2)	8	12 (0,384) 11 (0,287)	8 (0,004) 7 (0,007)
<b>D13S317</b> (Anexo 3)	9	11 (0,257) 12 (0,243)	7.1 (0,004) 15 (0,007)
<b>D16S539</b> (Anexo 4)	7	12 (0,287) 11 (0,235)	8 (0,011) 14 (0,015)
<b>D18S51</b> (Anexo 5)	15	14 (0,183) 15 y 16 (0,142)	9, 10, 11, 21.2 y 22 (0,004)
<b>D21S11</b> (Anexo 6)	10	30 (0,283) 29 (0,222)	33.1 y el 35 (0,005)
<b>D3S1358</b> (Anexo 7)	8	15 (0,336) 16 (0,269)	20 (0,004) 13 (0,007)
<b>D5S818</b> (Anexo 8)	8	11 (0,369) 12 (0,358)	8 (0,004) 14 (0,007)
<b>D7S820</b> (Anexo 9)	9	10 (0,269) 11 (0,265)	10.3 y el 14 (0,004)
<b>D8S1179</b> (Anexo 10)	11	14 (0,291) 13 (0,254)	9, 17 y 18 (0,004)
<b>FGA</b> (Anexo 11)	16	25 (0,162) 22 (0,143)	18.2, 21.2, 25.3, 29 y 30 (0,004)
<b>Penta D</b> (Anexo 12)	11	9 (0,225) 10 (0,198)	16 (0,008) 2.2 y 7 (0,016)
<b>Penta E</b> (Anexo 13)	17	12 (0,198) 7 (0,112)	23 (0,004) 21 (0,015)

<b>Locus</b>	<b>Cantidad de alelos presentes</b>	<b>Alelos más frecuentes</b>	<b>Alelos menos frecuentes</b>
<b>TH01</b> (Anexo 14)	5	7 (0,241) 6 (0,233)	8 (0,128) 9.3 (0,195)
<b>TPOX</b> (Anexo 15)	7	8 (0,241) 11 (0,265)	13 (0,007) 6 (0,011)
<b>vWA</b> (Anexo 16)	8	16 (0,299) 17 (0,250)	13 (0,004) 20 (0,015)
<b>D10S1248</b> (Anexo 17)	8	14 (0,343) 15 (0,220)	8 (0,004) 17 (0,011)
<b>D12S391</b> (Anexo 18)	14	18 (0,203) 20 (0,199)	17.3 y el 26 (0,004)
<b>D19S433</b> (Anexo 19)	11	14 (0,304) 13 (0,250)	10 y el 12.2 (0,004)
<b>D1S1656</b> (Anexo 20)	12	15 (0,233) 14 (0,129)	10 (0,004) 14.3 (0,008)
<b>D22S1045</b> (Anexo 21)	9	15 (0,384) 16 (0,325)	8 (0,004) 18 (0,007)
<b>D2S1338</b> (Anexo 22)	11	17 (0,187) 19 (0,179)	26 (0,007) 21 (0,030)
<b>D2S441</b> (Anexo 23)	9	11 (0,358) 10 (0,261)	8 y 9 (0,004)
<b>SE33</b> (Anexo 24)	20	18 (0,126) 19 (0,105)	12.2, 13, 16.3 y 20.2 (0,005)



## DISCUSIÓN DE RESULTADOS

Debido a que en los últimos años se ha incrementado la aplicación de la genética forense para el esclarecimiento de problemas de índole civil y penal, mismos que impactan de manera importante la vida de las personas involucradas, es imperativo contar con los métodos apropiados para resolver correcta y oportunamente las controversias (Álvarez-Cubero y col., 2010). La mayor parte de las conclusiones de un dictamen en materia de genética forense recaen en el tratamiento estadístico de los datos obtenidos de la evidencia, del probable responsable y de la víctima; así como del probable padre y del menor. Para conocer qué tan raro es un perfil genético en la población de estudio, se requiere de la base de datos de las frecuencias alélicas de dicha población. Es decir, la base de datos es uno de los pilares del análisis genético para establecer el peso estadístico de los perfiles genéticos analizados. También es necesario analizar un mayor número de marcadores con el fin de aumentar el poder de discriminación y, por lo tanto, también aumenten los índices estadísticos a un nivel de certidumbre científica razonable. En este contexto, al trabajar casos forenses en los cuales se pueden presentar muestras en mal estado de conservación en las cuales se extrae ADN muy degradado, es necesario estudiar un mayor número de marcadores que permitan obtener un mayor poder de discriminación.

En este estudio, el análisis de las frecuencias alélicas fue calculado con el propósito de inferir el grado de polimorfismo de los microsatélites, caracterizando la variabilidad genética de la población venezolana. Agrawal y col. (2004) describen que

dos individuos, cuando son seleccionados aleatoriamente, son idénticos genéticamente en un 99,9% (aproximadamente), mientras que el 0,1% otorga la individualidad de una persona. Asimismo, el estudio de la frecuencia alélica entre las poblaciones posibilita la creación de un banco de datos que puede ser comparado al sistema CODIS o a cualquier otro banco de datos.

Para que las frecuencias alélicas que se obtengan de la población puedan ser utilizadas en el análisis genético, es necesario que estén en equilibrio de Hardy-Weinberg. Esta ley define el estado genético de una población mendeliana en ausencia de toda fuerza evolutiva. (Butler, 2011).

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

Los resultados de las frecuencias alélicas obtenidas en el presente estudio son similares a los descritos por otros autores (Chiurillo y col. (2003) y Pineda y col. (2006)), quienes calcularon las frecuencias alélicas de 15 marcadores STR de la población venezolana de diversas regiones de Venezuela, región central y Maracaibo, respectivamente (Anexos 2,3, 4, 5, 6, 7, 8, 9, 10, 11, 14, 15, 16, 19 y 22). Cabe destacar que dentro del núcleo de 15 STR evaluados en ambas publicaciones, no se incluyen los

marcadores Penta D, Penta E, D10S1248, D12S391, D1S1656, D22S1045, D2S441 y SE33, los cuales son estudiados por primera vez en la población venezolana, en el presente estudio.

En relación con la variabilidad de los marcadores de la población venezolana, de los 23 marcadores STR analizados se observó que los marcadores SE33, Penta E y FGA fueron los que presentaban una mayor variabilidad en cuanto al número de alelos. De esta forma, se detectó que el marcador SE33 presentaba 20 alelos distintos, el Penta E presentaba 17 y el FGA presentaba 16 alelos distintos. En cambio, los marcadores TH01, TPOX y D16S539 presentaban menor variabilidad, detectando para el marcador TH01 5 alelos distintos, y para TPOX y D16S539 7 alelos distintos.

Al comparar la variabilidad de los alelos encontrados en este estudio con los estudios llevados a cabo por Chiurillo y col. (2003), los marcadores con mayor variabilidad en cuanto al número de alelos fueron D18S51, D21S11, FGA y D19S433 con 16 alelos distintos. En otro estudio (Pineda y col. (2006)) el marcador D18S51 fue el más variable, presentando 14 alelos distintos, seguido de los marcadores FGA y D2S1338 con 13 alelos distintos. Además, en otros estudios (Chiurillo y col. (2003) y Pineda y col. (2006)) el marcador con menor variabilidad fue el TH01 con 7 alelos, seguido del TPOX con 8 alelos, semejante a los resultados del presente estudio. Con respecto a los alelos más frecuentes obtenidos, se puede observar que los presentes datos concuerdan con los descritos por ambos autores previamente mencionados.

Por lo tanto, en base a los datos obtenidos en dichos estudios previos y los datos del presente estudio, podemos decir que el marcador FGA es el que presenta mayor variabilidad en relación con el número de alelos en la población venezolana. No es posible discutir acerca de los marcadores SE33 y Penta E ya que no fueron estudiados en estudios previos, por lo tanto son datos novedosos para la población venezolana e importantes ya que presenta aún mayor variabilidad con respecto al marcador FGA.

Respecto a los parámetros estadísticos forenses como el poder de discriminación (PD), el contenido de información polimórfica (PIC), el poder de exclusión (PE), la heterocigosidad observada ( $H_{obs}$ ), la heterocigosidad esperada ( $H_{esp}$ ), el equilibrio Hardy-Weinberg (*P-value*) y la probabilidad de coincidencia ( $P_i$ ), es importante mencionar que para el empleo de los marcadores STR en identificación humana, estos deben presentar un poder de discriminación superior al 0,90 y heterocigosidad observada superior a 0,70 (Butler, 2009).

De esta manera, el poder de discriminación (PD) fue calculado para verificar la probabilidad de que dos individuos no relacionados sean diferenciados genéticamente, es decir, la probabilidad de diferenciación de una muestra con las demás (Desmarais y col., 1998). Por lo tanto, el marcador que presentó el índice más alto de PD fueron el Penta E y SE33 (0,980) seguidos del FGA (0,970) y el D18S51 (0,967). Estos resultados fueron los mismos que los detectados en otros estudios (Chiurillo y col. (2003) y Pineda y col. (2006)), en los cuales el marcador con el índice de PD más alto fueron el FGA y D18S51. De esta forma, queda comprobada la capacidad de identificación forense de los

marcadores SE33, Penta E, FGA y D18S51 y su empleo en análisis de vínculo genético en la población venezolana.

En cambio, el marcador con índice más bajo de PD en el presente estudio fue el CSF1PO (0,854), seguido del D5S818 (0,863) y del TPOX (0,866). Al comparar estos resultados con los detectados en otros estudios (Chiurillo y col. (2003) y Pineda y col. (2006)), se confirmó que los marcadores con índice más bajo de PD en dichos estudios previos también fueron TPOX, CSF1PO y D5S818, cuyos índices se encuentran por debajo de 0,90. En este sentido, dichos marcadores no presentan un índice de poder de discriminación deseable (mayor que 0,90) (Butler, 2006).

Teniendo en cuenta que el PD es proporcional a las frecuencias alélicas y que los marcadores con mayor índice de PD son aquellos que presentan alelos con baja frecuencias, los datos presentes en este estudio indicarían que el marcador CSF1PO se encuentra muy poco distribuido en relación a sus frecuencias alélicas, una vez que apenas tres alelos (10, 11 y 12) obtuvieron una frecuencia sumada de 0,8863 (88,63%), siendo los cinco alelos restantes presentando una frecuencia muy baja con respecto a las presentadas por dichos tres alelos.

Por otra parte, el contenido de información polimórfica (PIC), calculado para verificar la capacidad informativa de los marcadores en los estudios genéticos, se ha demostrado bastante informativo en el presente estudio, presentando valores significativos (mayores que 0,50) para todos los marcadores. Este parámetro expresa el grado de diversidad biológica que posee cada locus, así como la capacidad para conocer

el origen parental de cada uno de los alelos de la descendencia (Botstein y col., 1980). El promedio de todos los marcadores para este parámetro fue de (0,782), mientras que, de forma individual, se obtuvo una variación de entre 0,66 (CSF1PO) y 0,92 (SE33), destacando otros marcadores que presentaron un índice elevado como fueron el FGA (0,87) y Penta E (0,90) y de manera contraria, los marcadores D5S818 (0,67), TPOX y D22S045 (0,68) presentaron un PIC menor. Dichos resultados concuerdan de manera similar con los obtenidos en otros estudios (Chiurillo y col. (2003) y Pineda y col. (2006)), en los que observaron también que el marcador FGA presenta el mayor índice de PIC. Asimismo, uno de los índices PIC más bajos observado en el presente estudio fue para los marcadores CSF1PO y TPOX, del mismo modo que en los estudios anteriormente mencionados. Que el marcador SE33 haya obtenido un alto contenido de información polimórfica implica que es un locus de gran utilidad, ya que, al tener un número grande de alelos bien distribuidos, las frecuencias de éstos son pequeñas, y por lo tanto aumentan el poder de discriminación.

Como se ha descrito anteriormente, el promedio de todos los marcadores para el PIC detectado en el presente estudio fue de 0,782, mientras que en otras publicaciones el promedio del PIC fue de 0,772 (Chiurillo y col. (2003)), y 0,765 (Pineda y col. (2006)). Por lo tanto, estas publicaciones y el presente estudio obtuvieron valores similares de contenido en información polimórfica en respectivos estudios para la población venezolana. Así, en concordancia con estudios previos, los valores obtenidos en el

presente estudio demuestran un buen contenido de información polimórfica y, en consecuencia, un alto poder discriminativo del conjunto de microsatélites estudiados.

Sin embargo, cuando se consideran los parámetros estadísticos forenses de manera individual para cada marcador, y teniendo en cuenta que un microsatélite es considerado un excelente marcador molecular cuando presenta un número de alelos superior a 5 y el PIC con un índice por encima de 0,75 (Foissac y col., 1997), se puede demostrar, en el presente caso, que el marcador SE33 es un excelente marcador molecular, mientras que el marcador CSF1PO, en cambio, parece no presentar índices suficientes para ser caracterizado como excelente. Además, en el presente estudio, los marcadores CSF1PO, D3S1358, D5S818, TPOX, D10S1248, D22S1045 y D2S441 presentaron índices menores que 0,75, resultado similarmente observado en otros estudios (Chiurillo y col. (2003) y Pineda y col. (2006)) para los marcadores D3S1358, D5S818, TPOX y CSF1PO, por lo tanto, no considerados buenos marcadores moleculares para la población venezolana.

El poder de exclusión (PE) se define como la probabilidad de que un sistema genético específico muestre evidencias que conduzcan a la exclusión de un sospechoso. El PE se encuentra relacionado directamente con el grado de polimorfismo del marcador y, consecuentemente, también con la distribución de las frecuencias alélicas. Por lo tanto, observando los índices obtenidos para el PE, el marcador con el índice más alto fue el Penta E (0,817), seguido del D2S1338 (0,741) y del D1S1656 (0,728). Asimismo, el marcador D2S1338 fue uno de los marcadores con el mayor índice PE en otros estudios

(Chiurillo y col. (2003) y Pineda y col. (2006)). En cambio, el marcador con el índice más bajo en el presente estudio fueron para los marcadores vWA y D22S1045 (0,466), seguidos del D5S818 (0,491) y del D10S1248 (0,503). En los estudios anteriormente mencionados, los marcadores vWA y D5S818 fueron uno de los marcadores con menor índice PE. En este sentido, los marcadores vWA, D22S1045, D5S818 y D10S1248 no presentan un poder de exclusión deseable para la población venezolana.

Según Brenner (2004), los índices de contenido de información polimórfica (PIC), poder de discriminación (PD) y poder de exclusión (PE) son utilizados para conferir mayor claridad en los análisis de vínculo genético y estudios poblacionales. Por lo tanto, en el presente estudio, los marcadores con los índices más altos de PD, PIC y PE fueron Penta E, SE33 y FGA. Con respecto al marcador FGA, también obtuvo los índices más altos en otros estudios (Chiurillo y col. (2003) y Pineda y col. (2006)). Los índices de PIC, PD y PE demuestran la superioridad de los marcadores microsatélites para el mapeo genético, así como su eficiencia en análisis de vínculos genéticos.

Por otro lado, la heterocigosidad observada ( $H_{obs}$ ) refleja la frecuencia de individuos heterocigotos encontrados en la población, mientras la heterocigosidad esperada ( $H_{esp}$ ), o también conocida como diversidad génica, representa una estimación más precisa de la heterocigosidad en la población. La  $H_{esp}$  es un buen estimador de la variabilidad. Además, índices por debajo del 0,50 para la  $H_{esp}$  se considera insuficiente (Crow y Kimura, 1970).



En el presente estudio, el marcador que obtuvo el índice más alto de  $H_{\text{esp}}$  fue el Penta E (0,907), seguido del SE33 (0,898) y del FGA (0,885). Los índices más bajos fueron observados en el marcador CSF1PO (0,712), seguido del TPOX (0,725) y del D22S1045 (0,726). Por otro lado, el marcador que obtuvo el índice más alto de  $H_{\text{obs}}$  fue el Penta E (0,910), seguido del D2S1338 (0,873) y del D1S1656 (0,867). Los índices más bajos fueron observados en los marcadores vWA y D22S1045 (0,724), seguidos del D5S818 (0,738) y del D10S1248 (0,746). Sin embargo, dado que el marcador D22S1045 es un miniSTR, cuyos iniciadores amplifican para que su producto de ADN resultante sea más pequeño, permiten que este marcador sea muy útil cuando se trabaja con muestras severamente comprometidas o dañadas. Comparando los índices de  $H_{\text{obs}}$  y  $H_{\text{esp}}$  del presente trabajo con otros estudios (Chiurillo y col. (2003) y Pineda y col. (2006)), se observó que el valor más alto para  $H_{\text{esp}}$  fue para el marcador FGA y el menor valor de este índice lo obtuvieron los marcadores TPOX y CSF1PO, mientras que el valor más alto de  $H_{\text{obs}}$  se observó para el marcador D2S1338 y el menor valor de este parámetro estadístico lo obtuvo el marcador D5S818, obteniendo resultados semejantes con el presente estudio. Así, teniendo en cuenta que una mayor diversidad alélica existente significa un alto índice de heterocigosidad, y cuanto más altas son las frecuencias alélicas menor es la tasa de heterocigosidad (Butler, 2009), la presencia de un 0,807 (80,70%) de heterocigotos en el presente estudio, indica que los marcadores analizados presentan una gran diversidad genética, y confirma, por lo tanto, lo que se ha mencionado anteriormente. Una alta heterocigosidad del panel significa que existe más diversidad

alélica, por lo que habrá un mayor número de genotipos posibles, aumentando el poder de discriminación y disminuyendo la probabilidad de coincidencia en muestras.

Otro parámetro estadístico forense analizado fue la probabilidad de coincidencia ( $P_i$ ), o también conocida como probabilidad de *match*. Este parámetro fue calculado para verificar la probabilidad de que dos individuos no relacionados compartan un mismo perfil de ADN, mientras menor es el valor de este parámetro existen menos posibilidades de que existan asociaciones fortuitas. En el presente trabajo, el marcador con el índice  $P_i$  más alto fue el CSF1PO (0,146), seguido del D5S818 (0,137) y del TPOX (0,134). En cambio, los marcadores con el índice más bajo fueron el SE33 y Penta E (0,020), seguido por el FGA (0,030) y del D18S51 (0,033). En otras publicaciones (Chiurillo y col. (2003) y Pineda y col. (2006)), no se reportaron resultados con diferencias significativas puesto que para ambos estudios los marcadores con mayor índice  $P_i$  fueron también CSF1PO, D5S818 y TPOX, mientras que los que obtuvieron los menores valores del índice  $P_i$  fueron, también, los marcadores FGA y D18S51. De esta forma, se observa que los marcadores SE33, Penta E, FGA y D18S51 son los que presentan la menor probabilidad de coincidencia para la población venezolana.

Para que un marcador STR pueda ser utilizado en la práctica forense, es necesario que la población se encuentre en equilibrio de Hardy-Weinberg (Szibor y col., 2003). El equilibrio Hardy-Weinberg nos indica que si las frecuencias genotípicas observadas son iguales o muy cercanas a las frecuencias genotípicas esperadas (es decir, que no exista diferencia significativa), entonces la población está en equilibrio y por lo tanto se podrán

utilizar las frecuencias alélicas para predecir las frecuencias genotípicas en las siguientes generaciones, y con base a estas frecuencias calcular diversos parámetros estadísticos al momento de comparar dos perfiles y definir o identificar a la fuente de este perfil. Esto es posible ya que al estar en equilibrio dicha población sus frecuencias alélicas serán constantes a través del tiempo y podrán ser utilizadas en las generaciones siguientes, asumiendo que la población cumpla con los siguientes supuestos: una población de tamaño infinito con presencia de apareamiento aleatorio y ausencia de selección, mutación o migración que introduzca nuevos alelos, lo cual, por supuesto, no existe en poblaciones humanas reales (Butler, 2015).

Las frecuencias alélicas son utilizadas para generar las frecuencias genotípicas esperadas que son luego comparadas con las frecuencias genotípicas observadas. Si los valores observados y esperados son similares entonces se asume que los alelos dentro de un locus son estables a través de generaciones o, en otras palabras, están en “equilibrio”. De esta manera, el equilibrio Hardy-Weinberg predice la estabilidad de las frecuencias alélicas y genotípicas de una generación a la siguiente. El propósito principal en las pruebas para el equilibrio Hardy-Weinberg es determinar si los alelos dentro de un locus son independientes entre sí. Así, las frecuencias no deberían cambiar a lo largo de muchas generaciones si el locus es genéticamente estable. Sin embargo, las poblaciones naturales generalmente violan el equilibrio Hardy-Weinberg en cierto modo y, por lo tanto, causan que las frecuencias alélicas cambien con el tiempo (Butler, 2015).

Los resultados obtenidos en el presente estudio no evidencian ninguna desviación significativa del equilibrio Hardy-Weinberg ( $p\text{-value} < 0,05$ ) para los loci estudiados, lo cual satisface al equilibrio Hardy-Weinberg y, por tanto, no se presenta un déficit en el número de heterocigotos. Estos resultados implican que los alelos de los loci estudiados se heredan de manera independiente entre sí, por lo cual hacen posible que sean estables en la población, de una generación a la siguiente.

Con estos resultados, el conjunto de marcadores STR autosómicos de la población venezolana confirmaron un elevado grado polimórfico (78,22%), elevado poder de discriminación (92,60%), elevado poder de exclusión (61,46%) y alta heterocigosidad (80,70%). Según los resultados presentados, los marcadores D18S51, D21S11, TH01, D3S1358, D16S539, D2S1338, D1S1656, D10S1248, D8S1179, vWA, D22S1045, SE33, D19S433, D12S391, FGA, D2S441, Penta E, TPOX, Penta D, CSF1PO, D7S820, D13S317 y D5S818 se encuentran en equilibrio de Hardy-Weinberg. De este modo, podrían ser perfectamente incorporados para la creación de una base de datos genética para la población venezolana. Asimismo, el marcador Penta E fue el que obtuvo los mejores índices forenses.

Posteriormente, al comparar la población venezolana estudiada con la hispana norteamericana (Anexo 25) se han detectado pocas diferencias entre las frecuencias alélicas y parámetros estadísticos forenses. En el presente estudio, los 23 loci STR autosómicos fueron analizados con el propósito de estimar la diversidad genética, las

frecuencias alélicas y para valorar la similitud o diferencia genética entre la población venezolana y la hispana norteamericana.

Asimismo, comparando la variabilidad de los alelos encontrados en el presente estudio. Los marcadores CSF1PO, D16S539, D5S818, Penta D, Penta E, TH01, D12S391, SE33 y D1S1656 presentaron más alelos en la población hispana con respecto a la venezolana, presentando mayor variabilidad en dicha población. Cabe destacar que el marcador SE33 presenta 16 alelos más en la población hispana norteamericana, con respecto a la población venezolana, por lo cual corresponde al marcador más variable al presentar 46 alelos. Por el contrario, los marcadores D13S317, D10S1248 y D7S820 presentan mayor variabilidad en la población venezolana al presentar 1 alelo más diferente al de la población hispana. Asimismo, los marcadores D3S1358, D8S1179 y D2S1338 no presentaron diferencias entre ambas poblaciones. No obstante, los marcadores que obtuvieron menor variabilidad en cuanto al número de alelos para ambas poblaciones fueron TH01 y TPOX. Así, se ha podido observar que los marcadores Penta E y SE33 son los que presentan mayor variabilidad en cuanto al número de alelos, tanto en la población venezolana como en la hispana norteamericana.

Respecto a los parámetros estadísticos forenses obtenidos en el presente estudio, los índices comparados entre la población venezolana e hispana norteamericana fueron el poder de exclusión (PE), heterocigosidad observada ( $H_{obs}$ ) y probabilidad de coincidencia ( $P_i$ ). Al comparar estos parámetros podemos observar que también fueron muy similares a los obtenidos por otros autores Hill y col. (2013)

De esta manera, al comparar los índices del poder de exclusión (PE) de la población venezolana con la hispana norteamericana, se puede observar que el marcador con el índice más alto encontrado en la población venezolana fue el Penta E, mientras que en la población hispana norteamericana fue el marcador SE33 (0,879) seguido también del Penta E (0,820). Por otro lado, el marcador con el índice más bajo en el presente estudio fue el D22S1045, igualmente observado en los estudios llevados a cabo por Hill y col. (2013).

No obstante, comparando los índices de heterocigosidad observada promedio de la población total venezolana, que fue de 0,807, con el promedio de la población hispana norteamericana, que fue de 0,812, se puede observar que la población hispana norteamericana está ligeramente por encima de los valores de la población venezolana. Esto muestra que en general, las poblaciones sometidas a continuos procesos de mezcla, como son las hispanas, tienen una diversidad génica mayor que las poblaciones con una menor dinámica geográfica.

En este sentido, los marcadores STR que más aportan para los estudios genéticos para la población venezolana e hispana norteamericana, debido a su alto poder de exclusión y alta heterocigosidad, son los Penta E y SE33.

Por último, al comparar la probabilidad de coincidencia ( $P_i$ ) de la población venezolana con la hispana norteamericana, se puede observar que el marcador con el índice más alto en la población venezolana fue el CSF1PO. Sin embargo, en los estudios llevados a cabo por Hill y col. (2013), el marcador con el índice más alto fue D22S1045

(0,174), seguido de TPOX (0,155). En cambio, los marcadores con el índice más bajo de  $P_i$  en el presente estudio fueron el Penta E y SE33, igualmente observado en la población hispana norteamericana con un índice de 0,012 y 0,010, respectivamente. De esta forma se observa que el marcador Penta E presenta un bajo índice de probabilidad de coincidencia para la población venezolana e hispana norteamericana, siendo también el marcador más informativo.

El análisis de los marcadores microsatélites ha permitido establecer, que los mismos son elementos extraordinariamente útiles en la identificación humana y en el mapeo genético. La genética forense ha alcanzado en los últimos años, un gran avance con el desarrollo de técnicas moleculares que actualmente nos permiten analizar evidencias con una cantidad mínima de ADN.

Como se ha descrito, los STR son la principal herramienta de la genética forense (Butler y Hill, 2012) ya que son marcadores altamente polimórficos y capaces de tipificar muestras con muy poco material biológico a través de una estrategia de PCR *multiplex*.

En el campo de la criminalística, la importancia del análisis de las regiones polimórficas del ADN aumenta el conjunto de marcadores disponibles para la resolución de casos criminales en términos estadísticos. Además, se aplican protocolos similares a otros estuches comerciales, lo cual significa que es de fácil y de rápida implementación en los laboratorios de análisis moleculares y forenses.

Asimismo, la tipificación de un mayor número de marcadores polimórficos por muestra aumenta de forma notoria el poder discriminativo y el potencial informativo de estos microsátélites (Butler, 2009).

De manera general, los marcadores STR presentaron una gran variabilidad entre la población venezolana e hispana norteamericana. Además, la realización de estudios tratando de obtener la frecuencia alélica poblacional es extremadamente importante ya que, de esta forma, la divulgación de los resultados a través de la literatura científica y bases de datos hace posible su aplicación en la investigación forense.

De esta manera, el presente estudio representa un gran aporte para la genética forense en el país, ya que se redujo el número de coincidencias accidentales y se aumentó el poder discriminativo de pruebas de identificación al aumentar el número de marcadores analizados, los cuales pueden ser incluidos en las bases de datos criminalísticas actuales, permitiendo una comparación de perfiles genéticos obtenidos por los laboratorios forenses venezolanos con cualquier base de datos internacional, como el sistema CODIS. Además, se describen las frecuencias alélicas de marcadores que son altamente polimórficos con respecto a los marcadores publicados para la población venezolana en estudios previos, como los loci Penta E y SE33, de los cuales no se han descrito en la literatura y cuya utilización es recomendable para aumentar el poder de discriminación de los análisis genéticos.



## CONCLUSIONES

1. Se obtuvieron los perfiles de identidad genéticos para 23 STR en 134 individuos del territorio central venezolano que participaron en el estudio, a partir de los cuales se pudo calcular las frecuencias alélicas. De esta manera, se comprobó que los 23 marcadores se encuentran en equilibrio Hardy-Weinberg, por lo cual sus alelos se heredan de manera independiente entre sí. Por lo tanto podrían ser incorporados en la base de datos de frecuencias de la población venezolana, para su aplicación en casos forenses.
2. Se observó que de los 23 loci STR autosómicos analizados para la población venezolana, el locus Penta E es el más informativo y que más aporta para las pruebas de identificación debido a su alto índice de poder de discriminación (PD), contenido de información polimórfica (PIC), poder de exclusión (PE), heterocigosidad, mayor variabilidad de alelos y menor probabilidad de coincidencia (Pi).
3. Se detectó una concordancia de las frecuencias entre la población venezolana analizada y estudios previos de Chiurillo y col. (2003) y Pineda y col. (2006). Se presentan similitudes en cuanto a los marcadores más variables y al número de alelos que presentan. El marcador más variable en los tres estudios realizados en la población venezolana fue el locus FGA, ya que los marcadores SE33 y Penta E, los cuales fueron los que presentaron mayor variabilidad, no se encuentran dentro de los estudios previos, lo cual implica un aporte importante del presente trabajo.
4. Se pudo comprobar que los marcadores Penta E y SE33 son los más informativos en la población venezolana, lo cual concuerda con los resultados obtenidos para la

población hispana de Estados Unidos de Norteamérica. Ambos marcadores poseen un alto poder de exclusión y heterocigosidad, lo que demuestra que presentan índices discriminativos satisfactorios en estudios de casos forenses para ambas poblaciones.

5. Al comparar los índices de heterocigosidad se puede observar que la población hispana norteamericana está ligeramente por encima de los valores de la población venezolana.
6. Por último, se logró incrementar el poder discriminativo para pruebas de identificación forense en el país, al incluir nuevos marcadores en las bases de datos actuales. De esta forma, se reduce el número de coincidencias accidentales y aumenta la compatibilidad internacional y el poder de discriminación para casos criminales y de personas desaparecidas.

## BIBLIOGRAFÍA

Agrawal, S., Khan, F. y Nityanand, S. 2004. Short tandem repeat technology has diverse applications: individual identification, phylogenetic reconstruction and chimerism based post haematopoietic stem cell transplantation graft monitoring. *Indian J Med Sci.* **58**:297-304.

Álvarez-Cubero, M., Martínez, L., Saiz, M., Álvarez, J. y Lorente, J. 2010. Nuevas aplicaciones en identificación genética. *Cuad. Med. Forense.* **16(1-2)**:5-18.

Botstein, D., White, R., Skolnick, M. y Davis, R. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* **32**:314-331.

Brenner, C. 2004. Multiple mutations, covert mutations and false exclusions in paternity casework. *Int. Cong. Series.* **1261**:112-114.

Butler, J. 2006. Genetics and genomics of core short tandem repeat loci used in human identity testing. *J. Forensic Sci.* **51(2)**:253-65.

Butler, J. 2007. Short tandem repeat typing technologies used in human identity testing. *Biotechn.* **43(4)**:1-5.

Butler, J. 2009. Fundamentals of Forensic DNA typing. Academic Press, Primera Edición, Maryland, E.E.U.U.

- Butler, J. 2011. *Advanced Topics in Forensic DNA Typing: Methodology*. Academic Press, Primera Edición, Maryland, E.E.U.U.
- Butler, J. 2015. *Advanced Topics in Forensic DNA Typing: Interpretation*. Academic Press, Primera Edición, Maryland, E.E.U.U.
- Butler, J. y Hill, C. 2012. Biology and genetics of new autosomal STR loci useful for forensic DNA analysis. *Forensic Sci. Rev.* **24**:15-26.
- Butler, J., Buel, E., Crivellente, F. y McCord, B. 2004. Forensic DNA typing by capillary electrophoresis using the ABI Prism 310 and 3100 genetic analyzers for STR analysis. *Electrophoresis.* **25**:1397-1412.
- Cavalli-Sforza, L., Bodmer, W. y Walter, F. 1981. *Genética de poblaciones humanas*. Ediciones Omega, Primera Edición, Barcelona, España.
- Chiurillo, M., Morales, A., Mendes, A., Lander, N., Tovar, F., Fuentes, A. y Ramírez, J. 2003. Genetic profiling of a central Venezuelan population using 15 STR markers that may be of forensic importance. *Forensic Sci. Int.* **136**:99-101.
- Comité Internacional de la Cruz Roja. 2009. *Personas desaparecidas, análisis forense de ADN e identificación de restos humanos*. CICR, Segunda Edición, Ginebra, Suiza.
- Crow, J. y Kimura, M. 1970. *An introduction to population genetics theory*. Harper and Row Publishers, New York, E.E.U.U.

Desmarais, D., Zhong, Y., Chakraborty, R., Perreault, C. y Busque, L. 1998. Development of a highly polymorphic STR marker for identity testing purposes at the human androgen receptor gene (HUMARA). *J. Forensic Sci.* **43**:1046-49.

Economou, E., Bergen, A., Warren, A. y Antonarakis, S. 1990. The polydeoxyadenylate tract of Alu repetitive elements is polymorphic in the human genome. *Proc. Natl. Acad. Sci.* **79**:6465-6469.

Evett, I. y Weir, B. 1998. *Interpreting DNA evidence: statistical genetics for forensic scientists*. Sinauer Associates, Inc., Sunderland, Massachusetts, E.E.U.U.

Filho, J. 2012. Comparación de la frecuencia alélica de 13 loci STRs de la población brasileña y española para fines de identificación humana en genética forense. Tesis doctoral, Universidad de Valladolid, Valladolid, España.

Foissac, A., Crouau-Roy, B., Faure, S., Thomsen, M. y Cambon-Thomsen, A. 1997. Review article Microsatellites in the HLA region: an overview. *Tissue Antigens.* **49**:197-214.

García, O., Alonso, J., Cano, J., García, R., Luque, G., Martín, P., De Yuso, I. y colaboradores. 2012. Population genetic data and concordance study for the kits Identifiler, NGM, PowerPlex ESX 17 System and Investigator ESSplex in Spain. *Forensic Sci. Int. Genet.* **6(2)**:e78-9.

Gill, P., Ivanov, P., Kimpton, C., Piercy, R., Benson, N., Tully, G., Evett, I. y colaboradores. 1994. Identification of the remains of the Romanov family by DNA analysis. *Nature*. **6**:130-5.

Griffiths, A., Miller, J. y Gelbart, W. 2000. Genética Moderna. McGraw-Hill / Interamericana de España, Primer Edición, Barcelona, España.

Guo, S. y Thompson, E. 1992. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics*. **48**:361-372.

Hares, D. 2015. Selection and implementation of expanded CODIS core loci in the United States. *Forensic Sci. Int.* **17**:33-34

Hernández, M. 2014. Fundamentos de medicina legal. McGraw Hill, Primera Edición, D. F., México.

Hernández, A. y Trejo, F. 2014. Estudio genético poblacional de frecuencias alélicas para 15 marcadores STR presentes en la población del Estado de Zacatecas aplicado a la práctica forense. *Arch. de Medi.* **10**:1-24

Herráez, Á. 2012. Biología molecular e ingeniería genética. Elsevier, Segunda Edición, Madrid, España.

Hill, C., Duewer, D., Kline, M., Coble, M. y Butler, J. 2013. U.S. population data for 29 autosomal STR loci. *Forensic Sci. Int. Genet.* **7**:e82-e83.

- Hoffman, N. y Fenger, T. 2010. Validation of half-reaction amplification using Promega PowerPlex 16\*. *J. Forensic Sci.* **55(4)**:1044-49.
- Ibarra, A. y Arcos, O. 2000. Discriminación genética de individuos y su uso en medicina forense y filiación natural. *Med. & Lab.* **9**:3-4.
- IPGRI y Cornell University. Conceptos básicos de genética de poblaciones. [En línea]. 2004. [Consulta 13 de Octubre de 2017]. Disponible de: [https://biologiamyblog.files.wordpress.com/2010/03/ii-gen\\_tica\\_de\\_poblaciones1.pdf](https://biologiamyblog.files.wordpress.com/2010/03/ii-gen_tica_de_poblaciones1.pdf)
- Jeffreys, A., Wilson, V. y Thein, S. 1985. Individual-specific 'fingerprints' of human DNA. *Nature.* **316(6023)**:76-9.
- Jobling, M. y Gill, P. 2004. Encoded evidence: DNA in forensic analysis. *Nat. Rev. Genet.* **5(10)**:739-751.
- Jones, D. 1972. Blood samples: probability of discrimination. *J. Forensic Sci. Soc.* **12**:355-359.
- Klug, W., Cummings, M., Spencer, C. y Palladino, M. 2012. Concepts of Genetics. Pearson, Décima Edición, California, E.E.U.U.
- Kruglya, L. y Nickerson, D. 2001. Variation in the spice of life. *Nat. Genet.* **27**:234-236.
- Lander, N., Tovar, F., Chiurillo, M. y Ramírez, J. 2006. A New Allele of the Short Tandem Repeat Locus D21S11 in a Venezuelan Population. *J. Forensic Sci.* **51(3)**:695.

Landsteiner, K. 1990. The specificity of serological reactions. Dover Publication, Inc., New York, E.E.U.U.

Martin, B., Bastisch, I., Butler, J., Fimmers, R., Gill, P., Gusmao, L., Morling, N. y colaboradores. 2016. Recommendations of the DNA Commission of the International Society for Forensic Genetics (ISFG) on quality control of autosomal short tandem repeat allele frequency databasing (STRidER). *Forensic Sci. Int.* **24**:97-102.

Martínez, V., Aguilar, J., Inclán, A., Romero, O., Díaz, X., Saiz, M. y Rangel, H. 2015. Parámetros forenses del sistema PowerPlex 21 (Promega Corp.) en población mestiza del occidente de México. *Rev. Esp. Med. Legal.* **42(1)**:10-16.

Mestres, F. y Vives-Rego, J. 2009. Bancos y bases de datos genéticos para usos forenses. *Revista Poder Judicial.* **89**:239-263.

Moretti, T., Moreno, L., Smerick, J., Pignone, M., Hizon, R., Buckleton, J., Bright, J. y colaboradores. 2016. Population data on the expanded CODIS core STR loci for eleven populations of significance for forensic DNA analyses in the United States. *Forensic Sci. Int. Genet.* **25**:175-181.

National Institute of Standards and Technology. STRBase: Overview of STR Fact Sheets. [En línea]. 2017. [Consulta 24 de Octubre de 2017]. Disponible de: [http://strbase.nist.gov/str\\_fact.htm](http://strbase.nist.gov/str_fact.htm)

Pineda, L., Borjas, L., Zabala, W., Portillo, M., Fernández, E., Delgado, W., Tovar, F. y colaboradores. 2006. *Forensic Sci. Int.* **161**:60-63.



Promega. 2013. Internal validation guide of autosomal STR systems for forensic laboratories. Reference Manual.

Promega. 2016. PowerPlex 16 HS System Technical Manual. Wisconsin, E.E.U.U.

Promega. 2017A. PowerPlex ESX 17 System Technical Manual. Wisconsin, E.E.U.U.

Promega. 2017B. PowerPlex Fusion System for Use on the Applied Biosystems Genetic Analyzers Technical Manual. Wisconsin, E.E.U.U.

Rosado, F. (2008). Análisis de 15 loci tipo short tandem repeats (STR) en la población de Paraguay para su uso en identificación humana. Tesis doctoral, Universidad de Granada, Granada, España.

Rousset, F. 2008. Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. *Mol. Ecol. Resources*. **8**:103-106.

Sánchez, M. 2010. Estudio de polimorfismos genéticos en la evolución clínica de pacientes sometidos a transplante alogénico de progenitores hematopoyéticos. Tesis Doctoral, Universidad de Salamanca, Salamanca, España.



Shastri, B. 2002. SNP Alleles in human disease and evolution. *J. Hum. Genet.* **47(11)**:561-6.

Smalldon, K. y Moffat, A. 1973. The calculation of discriminating power for a series of correlated attributes. *J. Forensic. Sci.* **13**:291-295.

Stark, A. 2005. The hardy-Weinberg principle. *Genet. Mol. Biol.* **38**:485.

Szibor, R., Krawczak, M., Hering, S., Edelmann, J., Kuhlisch, E. y Krause, D. 2003. Use of X-linked markers for forensic purposes. *Int. J. Leg. Med.* **117**:67-74.

## ANEXOS

 <p>República Bolivariana de Venezuela Ministerio Público Unidad Criminalística Contra la Vulneración de Derechos Fundamentales Área Metropolitana de Caracas</p>	 <p>UNIVERSIDAD CENTRAL CARACAS - VENEZUELA</p>
<b>Cuestionario y consentimiento informado</b>	
<b>DATOS GENERALES</b>	N° de Control Interno UCCVDF-AMC _____ (FTA Card)
Fecha: _____.	
Lugar de nacimiento: _____.	
Teléfono: _____.	
<b>DATOS GENEALÓGICOS</b>	
Lugar de origen del padre: _____.	
Lugar de origen de la madre: _____.	
Lugar de origen de la abuela materna: _____.	
Lugar de origen del abuelo materno: _____.	
Lugar de origen de la abuela paterna: _____.	
Lugar de origen del abuelo paterno: _____.	
<b>DATOS ADICIONALES</b>	
¿Donaría Ud. Una muestra de sangre para contribuir a la elaboración de una base de datos que permita caracterizar genéticamente a la población venezolana? Si: ___ No: ___	
¿Confirma que se presta Ud. Como voluntario para este estudio? Si: ___ No: ___	
<a href="http://www.ministeriopublico.gob.ve">www.ministeriopublico.gob.ve</a>	
Formato 15/03/2017 BS/EP/JC/OF/YF/franco	Formato 15/03/2017 BS/EP/JC/OF/YF/franco

Yo, \_\_\_\_\_, mayor de edad y portador de la C.I. N° \_\_\_\_\_, hago constar que deseo participar de manera voluntaria en la el estudio "Base de de datos de frecuencias alélicas de la población venezolana", realizado en el Laboratorio de Genética Forense del Ministerio Público.

He sido informado y he entendido que el objetivo de este proyecto es realizar una base de datos de STR autosómicos para la población venezolana, la cual mejorará los valores probabilísticos en las pruebas genéticas de paternidad y forense.

El material donado consiste en: *Sangre periférica*

Este material podrá usarse para estudios de otros marcadores genéticos de interés forense y el resultado obtenido será completamente privado. En caso que así lo decida, tengo derecho a retirarme del estudio, siempre y cuando sea posible.

Participante: \_\_\_\_\_.

Firma: \_\_\_\_\_.

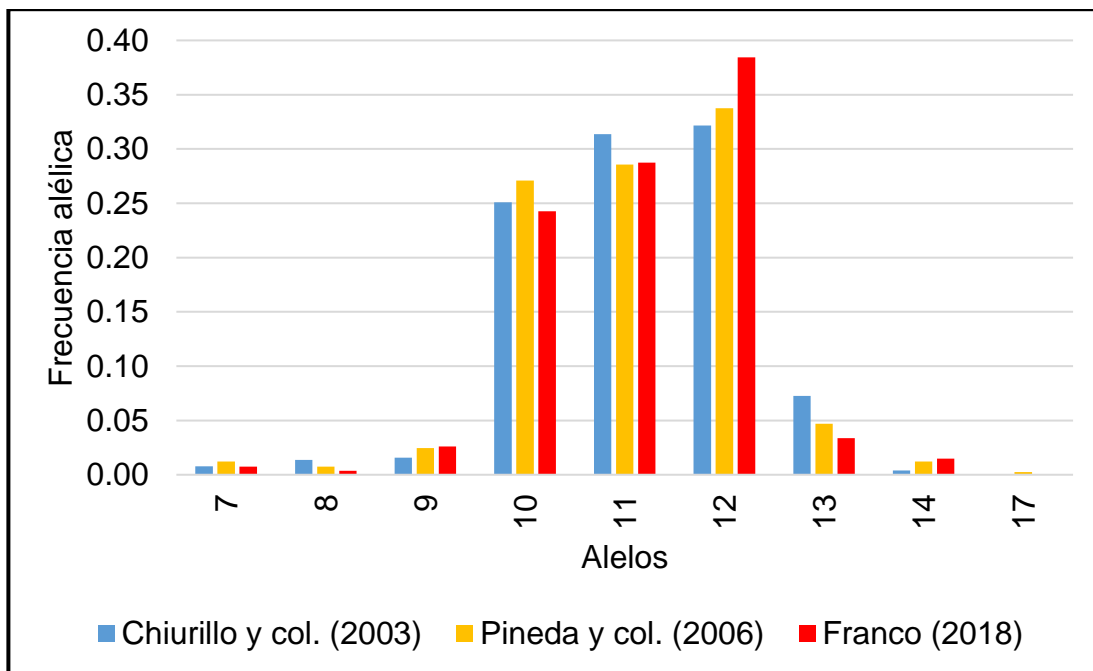
Investigador: \_\_\_\_\_.

Firma: \_\_\_\_\_.

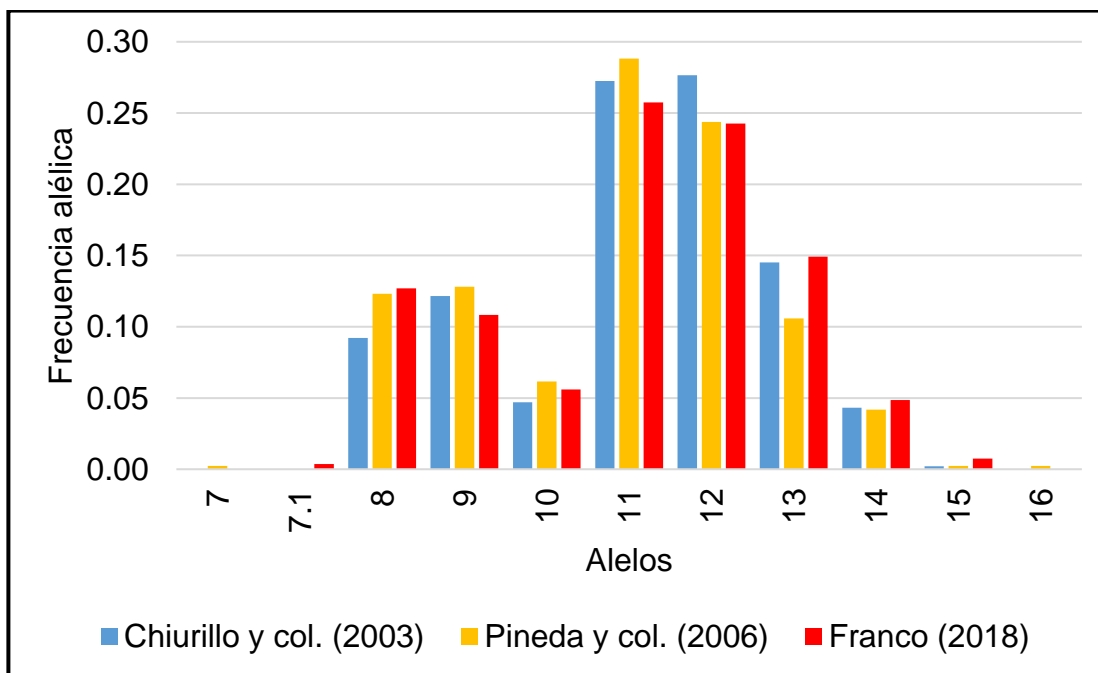
Funcionario testigo: \_\_\_\_\_.

Firma: \_\_\_\_\_.

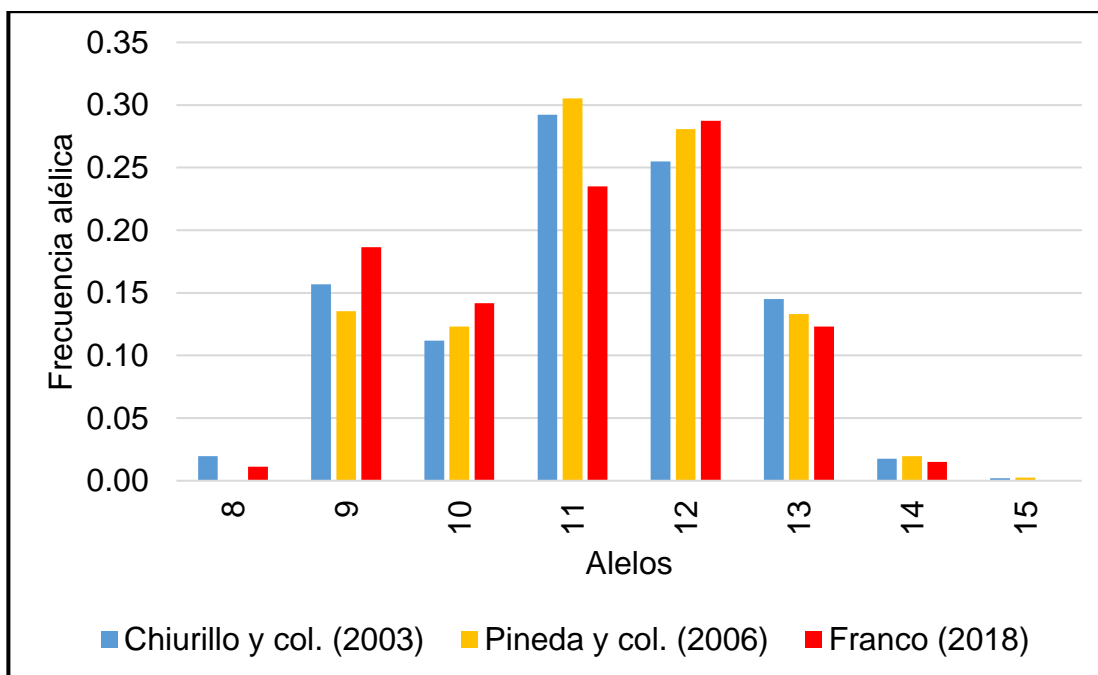
Anexo 1. Formato de consentimiento informado para toma de muestras.



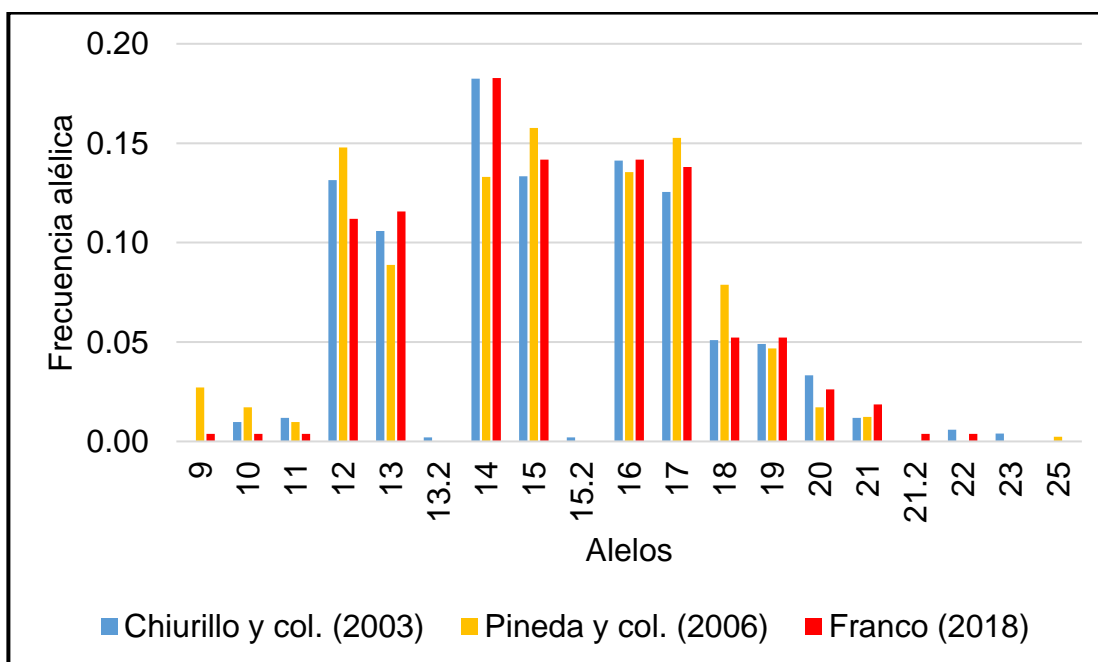
**Anexo 2.** Distribución de las frecuencias alélicas para el marcador CSF1PO de distintas bases de datos venezolanas. La base de datos “Franco (2018)” es la exhibida en el presente estudio.



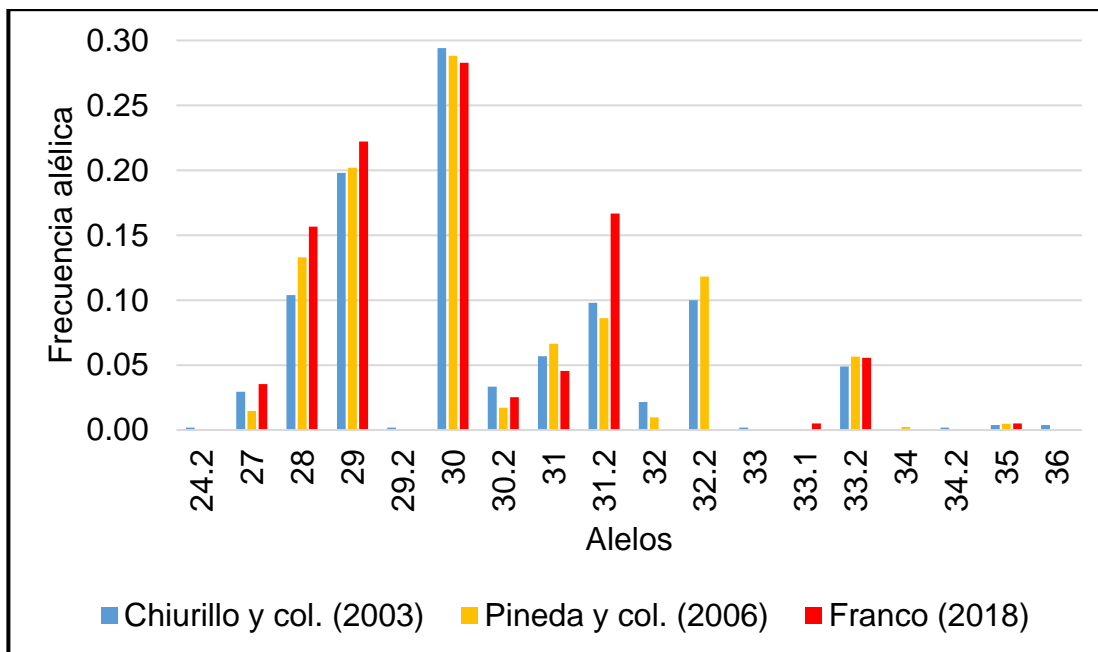
**Anexo 3.** Distribución de las frecuencias alélicas para el marcador D13S317 de distintas bases de datos venezolanas. La base de datos “Franco (2018)” es la exhibida en el presente estudio.



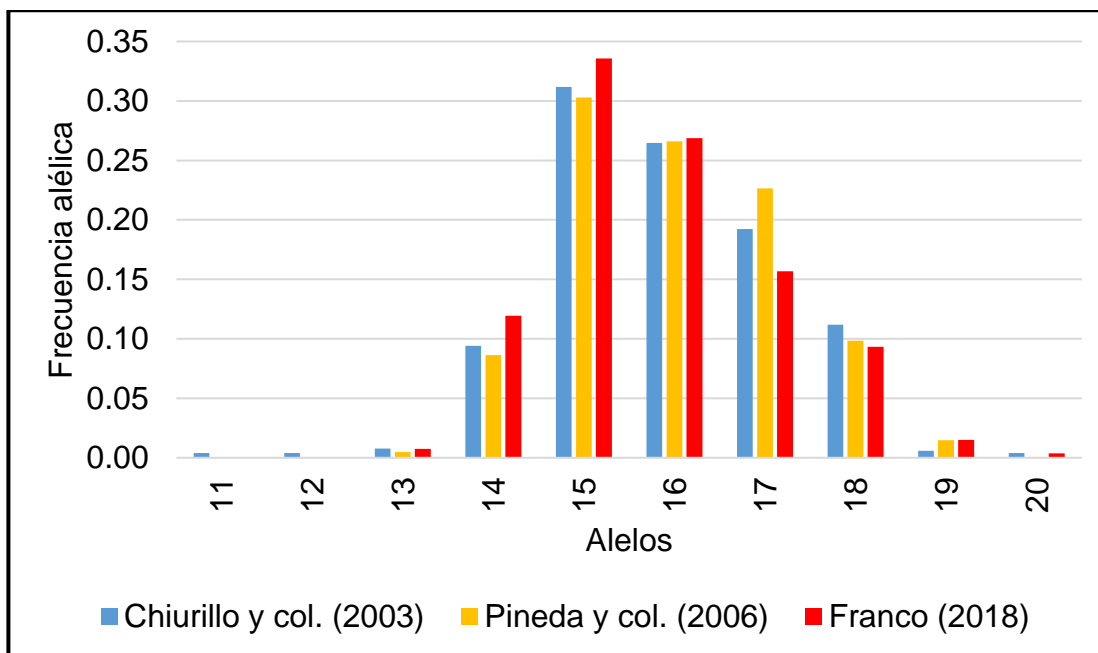
**Anexo 4.** Distribución de las frecuencias alélicas para el marcador D16S539 de distintas bases de datos venezolanas. La base de datos “Franco (2018)” es la exhibida en el presente estudio.



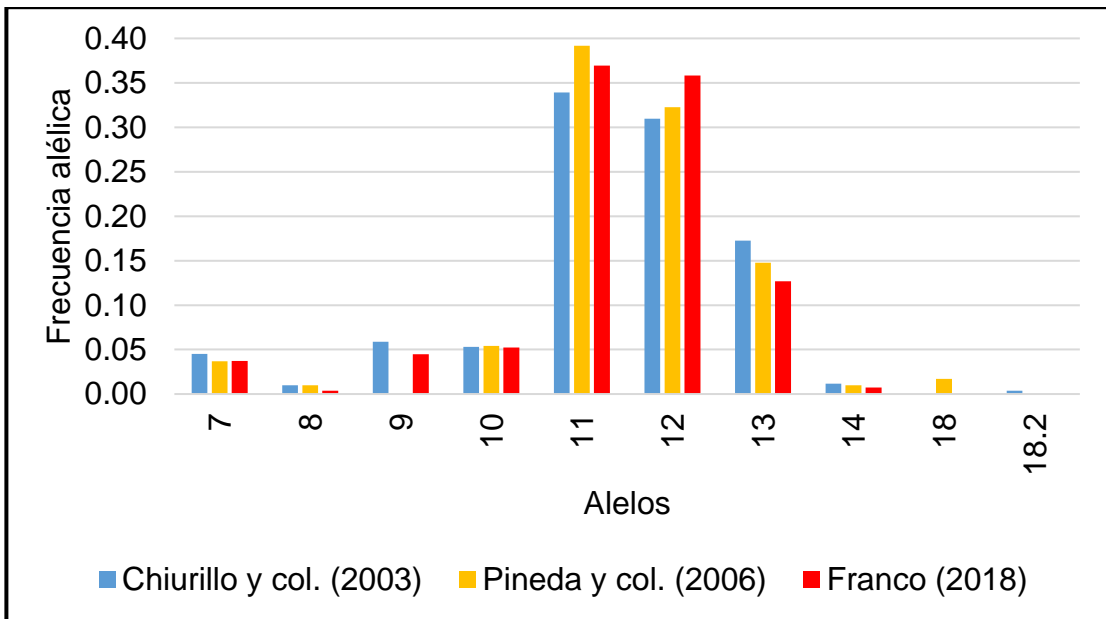
**Anexo 5.** Distribución de las frecuencias alélicas para el marcador D18S51 de distintas bases de datos venezolanas. La base de datos “Franco (2018)” es la exhibida en el presente estudio.



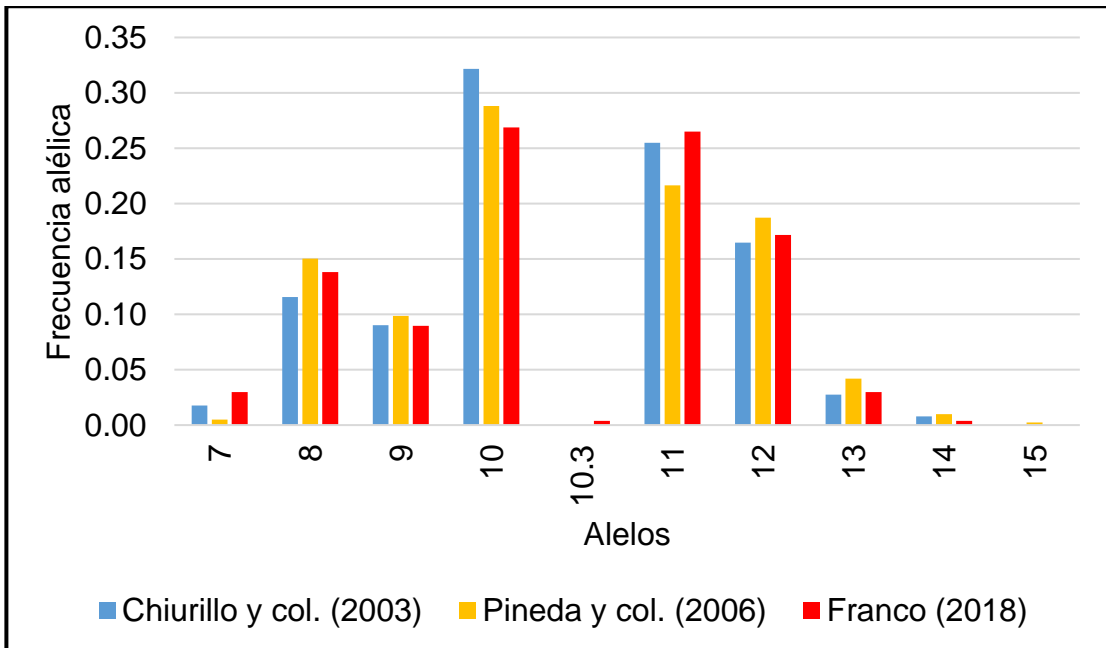
**Anexo 6.** Distribución de las frecuencias alélicas para el marcador D21S11 de distintas bases de datos venezolanas. La base de datos “Franco (2018)” es la exhibida en el presente estudio.



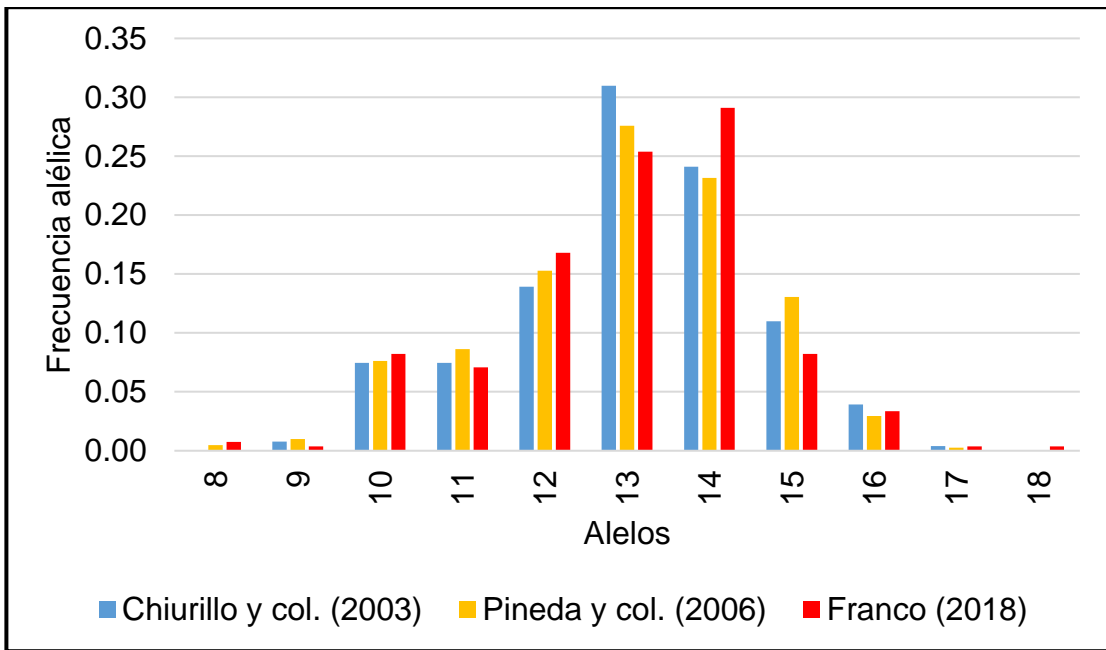
**Anexo 7.** Distribución de las frecuencias alélicas para el marcador D3S1358 de distintas bases de datos venezolanas. La base de datos “Franco (2018)” es la exhibida en el presente estudio.



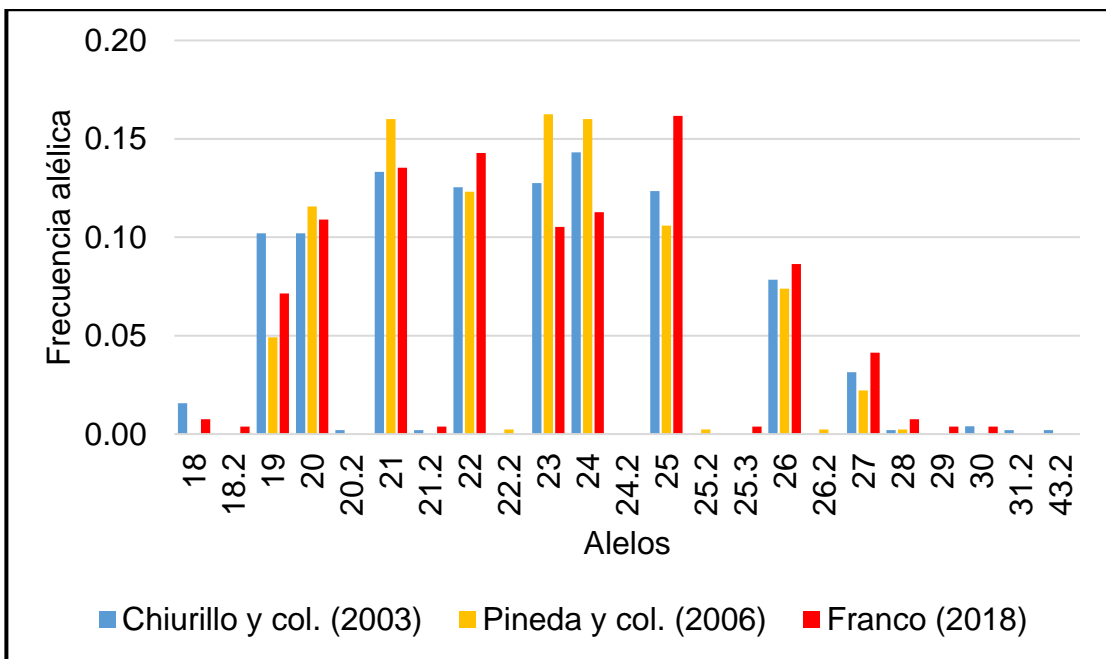
**Anexo 8.** Distribución de las frecuencias alélicas para el marcador D5S818 de distintas bases de datos venezolanas. La base de datos “Franco (2018)” es la exhibida en el presente estudio.



**Anexo 9.** Distribución de las frecuencias alélicas para el marcador D7S820 de distintas bases de datos venezolanas. La base de datos “Franco (2018)” es la exhibida en el presente estudio.

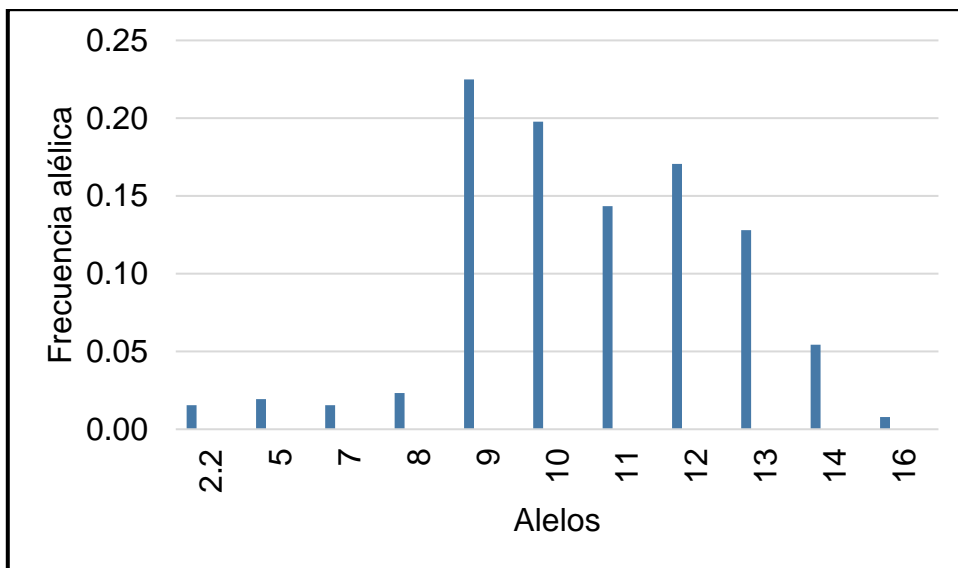


**Anexo 10.** Distribución de las frecuencias alélicas para el marcador D8S1179 de distintas bases de datos venezolanas. La base de datos “Franco (2018)” es la exhibida en el presente estudio.

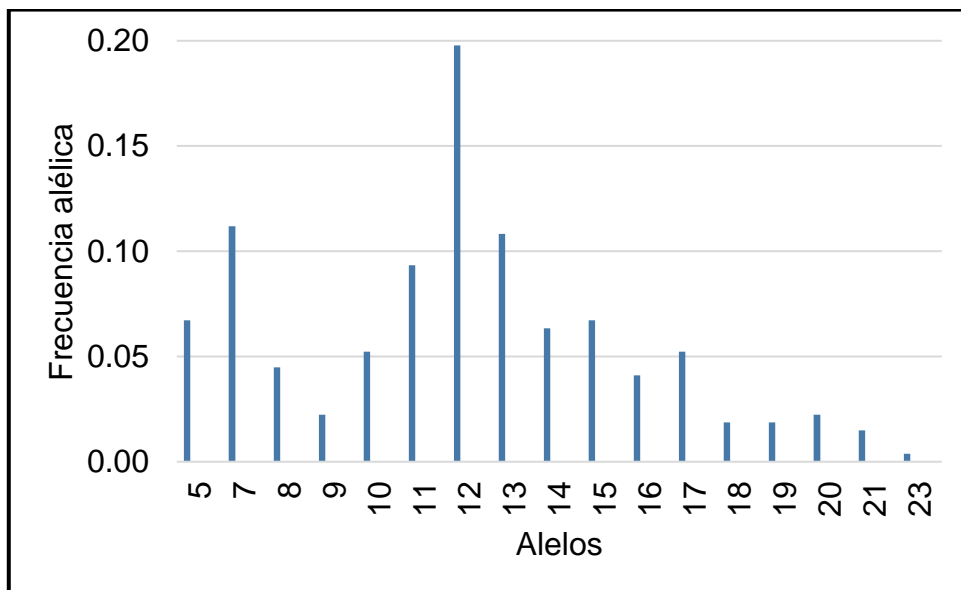


**Anexo 11.** Distribución de las frecuencias alélicas para el marcador FGA de distintas bases de datos venezolanas. La base de datos “Franco (2018)” es la exhibida en el presente estudio.

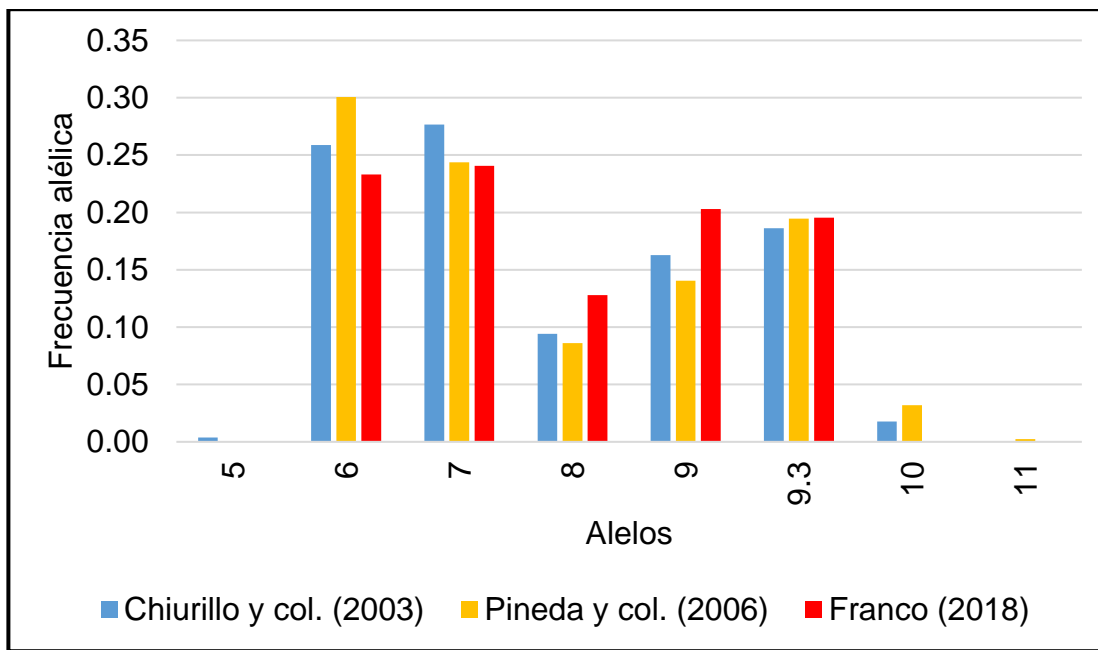




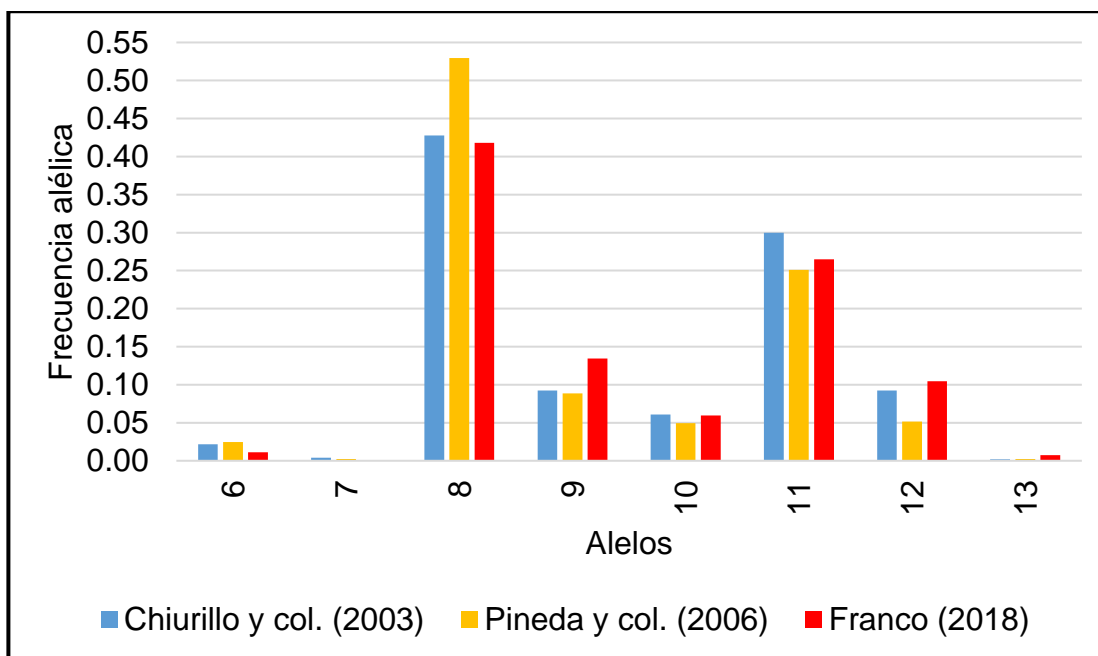
**Anexo 12.** Distribución de las frecuencias alélicas para el marcador Penta D en la población venezolana analizada en el presente estudio.



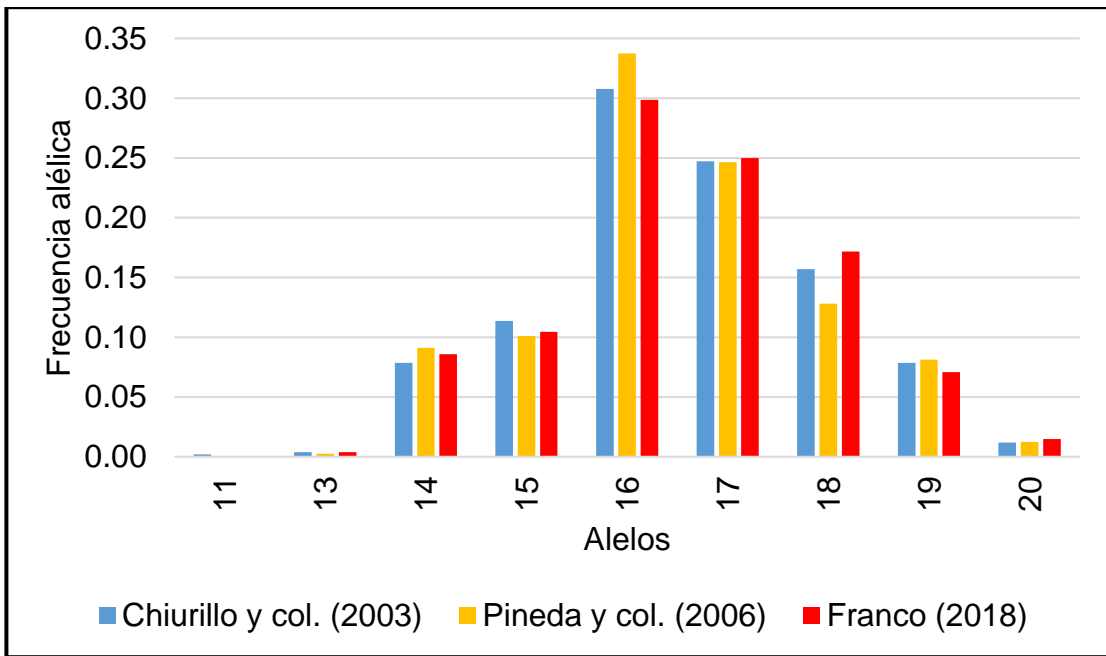
**Anexo 13.** Distribución de las frecuencias alélicas para el marcador Penta E en la población venezolana analizada en el presente estudio.



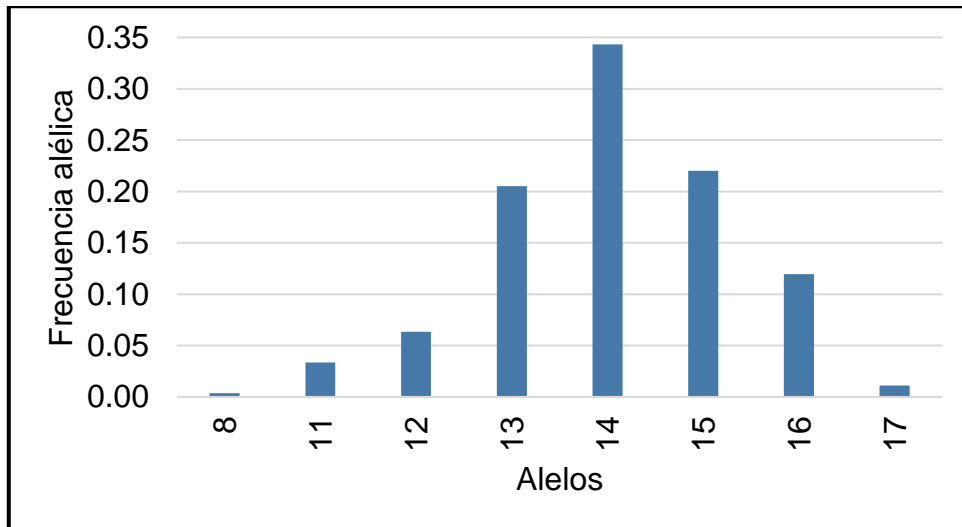
**Anexo 14.** Distribución de las frecuencias alélicas para el marcador TH01 de distintas bases de datos venezolanas. La base de datos “Franco (2018)” es la exhibida en el presente estudio.



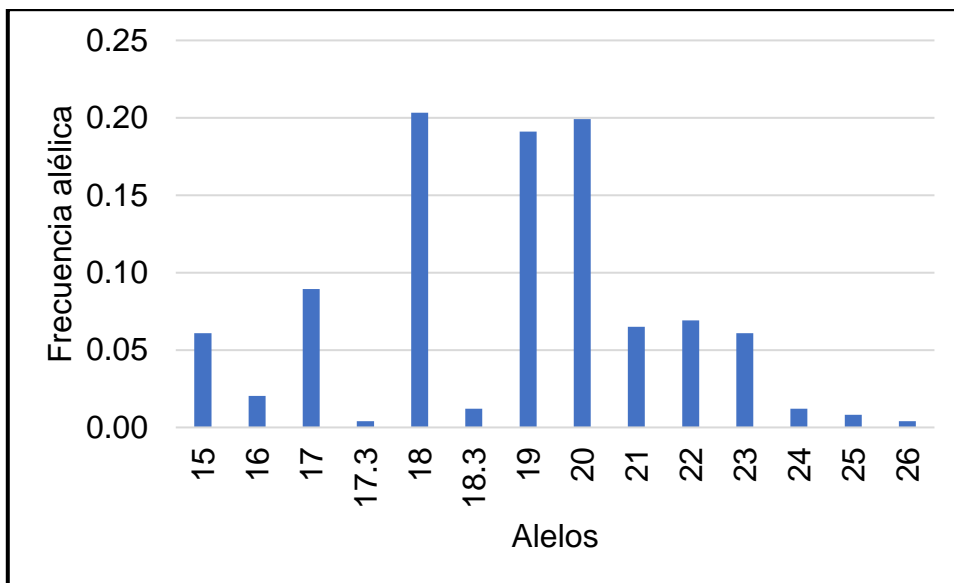
**Anexo 15.** Distribución de las frecuencias alélicas para el marcador TPOX de distintas bases de datos venezolanas. La base de datos “Franco (2018)” es la exhibida en el presente estudio.



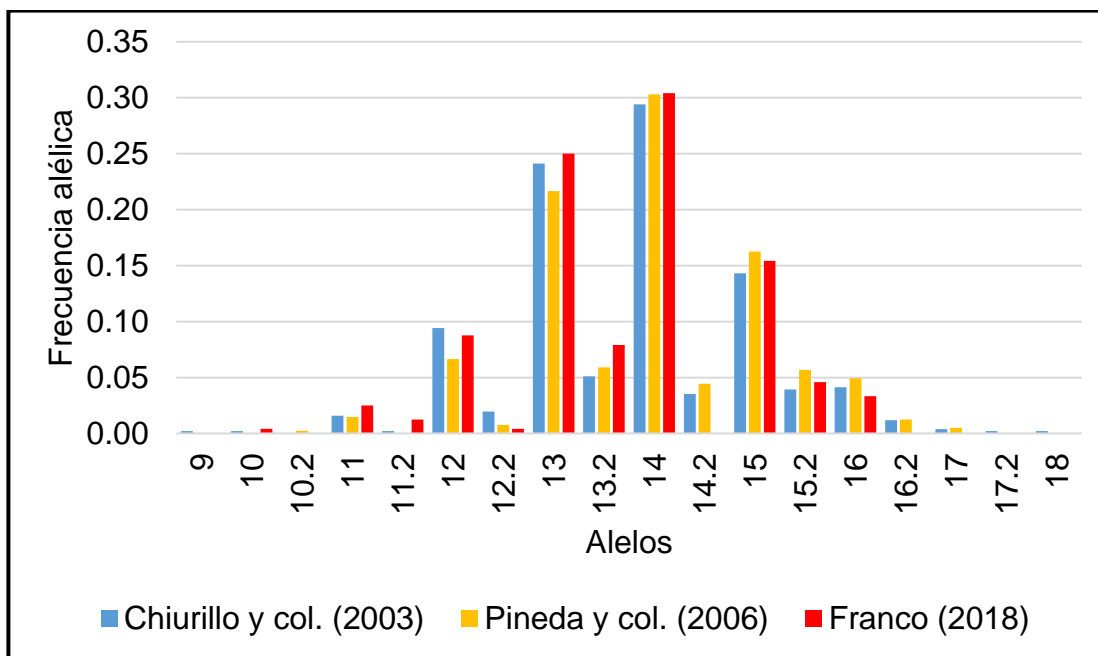
**Anexo 16.** Distribución de las frecuencias alélicas para el marcador vWA de distintas bases de datos venezolanas. La base de datos “Franco (2018)” es la exhibida en el presente estudio.



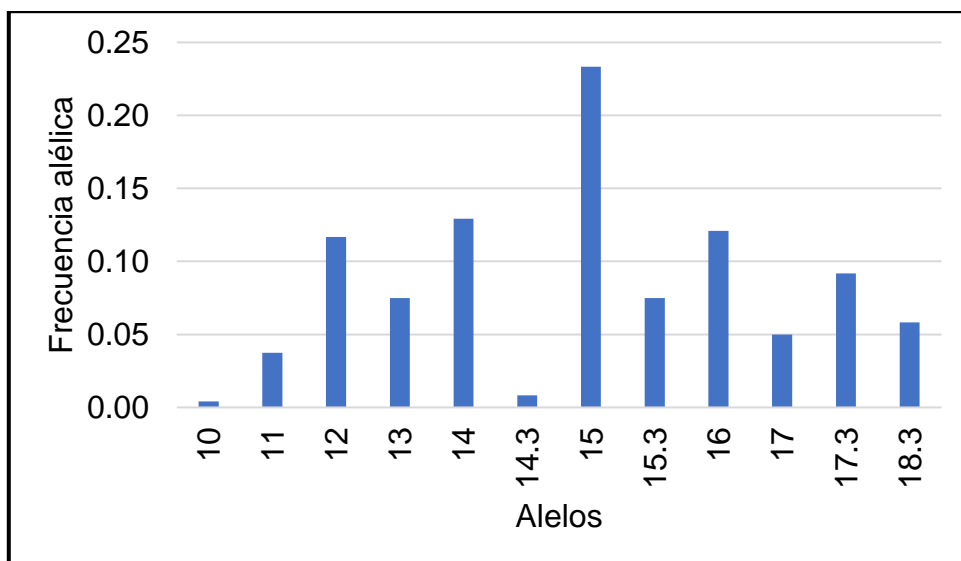
**Anexo 17.** Distribución de las frecuencias alélicas para el marcador D10S1248 en la población venezolana analizada en el presente estudio.



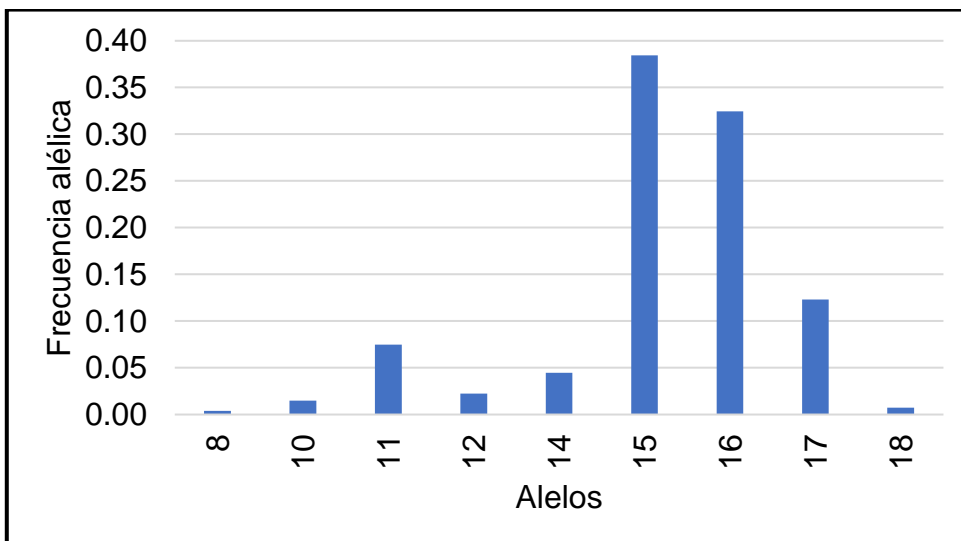
**Anexo 18.** Distribución de las frecuencias alélicas para el marcador D12S391 en la población venezolana analizada en el presente estudio.



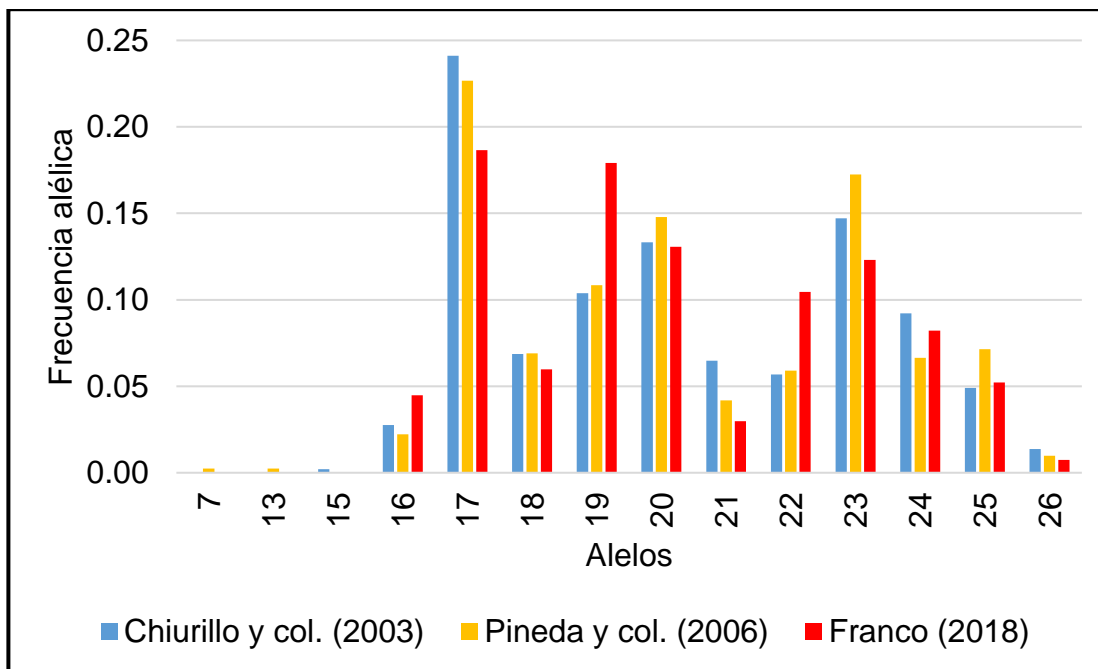
**Anexo 19.** Distribución de las frecuencias alélicas para el marcador D19S433 de distintas bases de datos venezolanas. La base de datos “Franco (2018)” es la exhibida en el presente estudio.



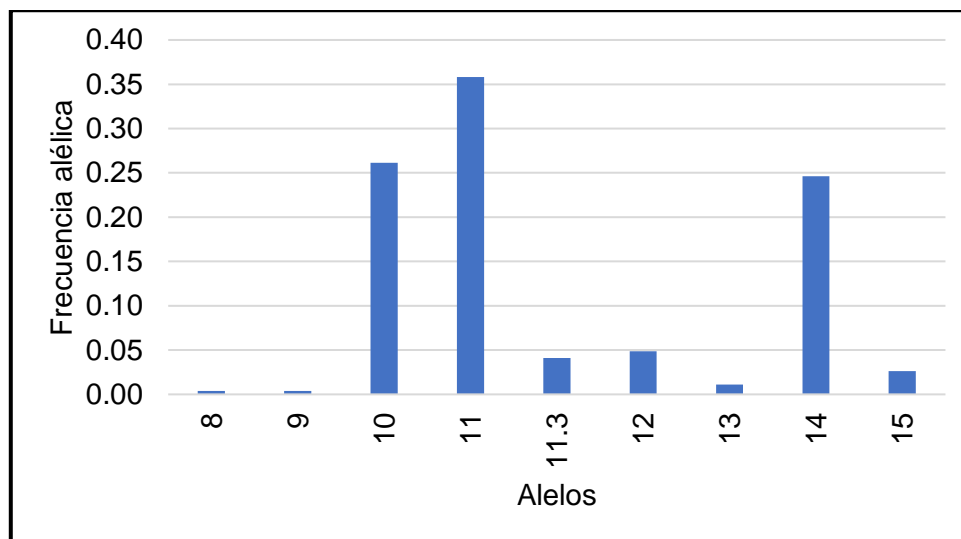
**Anexo 20.** Distribución de las frecuencias alélicas para el marcador D1S1656 en la población venezolana analizada en el presente estudio.



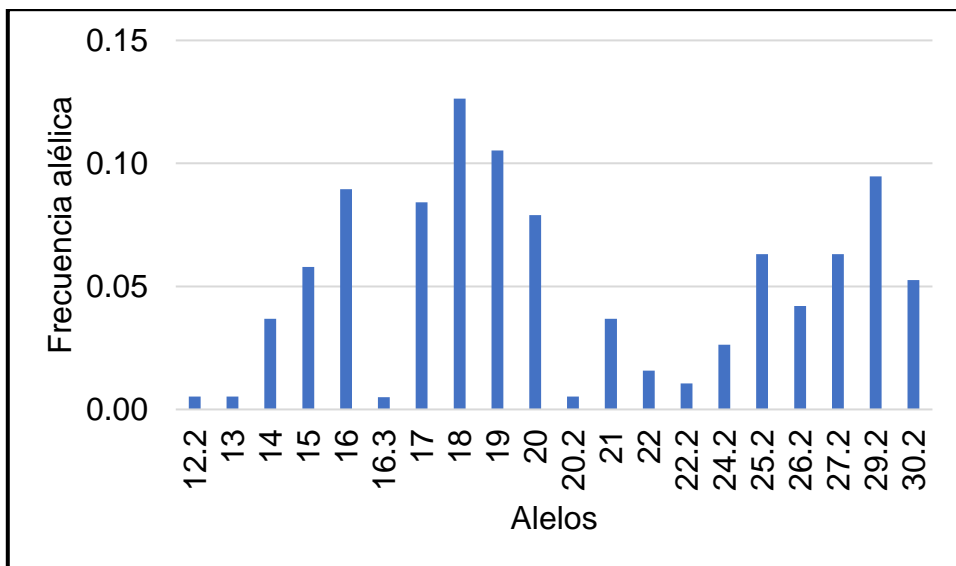
**Anexo 21.** Distribución de las frecuencias alélicas para el marcador D22S1045 en la población venezolana analizada en el presente estudio.



**Anexo 22.** Distribución de las frecuencias alélicas para el marcador D2S1338 de distintas bases de datos venezolanas. La base de datos “Franco (2018)” es la exhibida en el presente estudio.



**Anexo 23.** Distribución de las frecuencias alélicas para el marcador D2S441 en la población venezolana analizada en el presente estudio.



**Anexo 24.** Distribución de las frecuencias alélicas para el marcador SE33 en la población venezolana analizada en el presente estudio.

	CSF1PO	D13S317	D16S539	D18S51	D21S11	D3S1358	D5S818	D7S820	D8S1179	FGA	Penta_D	Penta_E	TH01	TPOX	VWA	D10S1248	D12S391	D19S433	D1S1656	D22S1045	D2S1338	D2S441	SE33
<b>H<sub>obs</sub></b>	0,729	0,852	0,788	0,907	0,864	0,733	0,737	0,822	0,784	0,881	0,868	0,898	0,759	0,674	0,788	0,763	0,890	0,780	0,881	0,733	0,852	0,754	0,928
<b>P<sub>i</sub></b>	0,126	0,053	0,076	0,030	0,047	0,090	0,146	0,087	0,062	0,028	0,050	0,018	0,085	0,155	0,073	0,105	0,033	0,068	0,028	0,174	0,030	0,110	0,010
<b>PE</b>	0,476	0,666	0,592	0,756	0,683	0,543	0,483	0,576	0,630	0,759	0,678	0,820	0,562	0,445	0,601	0,529	0,756	0,617	0,768	0,439	0,747	0,525	0,879

*Hobs: Heterocigosidad observada. Pi: Probabilidad de coincidencia aleatoria. PE: Poder de exclusión*

**Anexo 25.** Parámetros estadísticos de interés forense para algunos de los loci estudiados en la población hispana norteamericana. Adaptado de Hill y col. (2003)




REPÚBLICA BOLIVARIANA DE VENEZUELA  
 MINISTERIO DEL PODER POPULAR PARA LA DEFENSA  
 VICEMINISTERIO DE SERVICIOS, PERSONAL Y LOGÍSTICA  
 DIRECCIÓN GENERAL DE SALUD  
 HOSPITAL MILITAR "DR. CARLOS ARVELO"  
 COMITÉ DE ÉTICA

**APROBACIÓN DEL COMITÉ DE ÉTICA E INVESTIGACIÓN**

*TÍTULO DEL ESTUDIO:*

**"ANÁLISIS DE LA DISTRIBUCIÓN ALÉLICA DE 23 MICROSATÉLITES DE IMPORTANCIA FORENSE EN LA POBLACIÓN VENEZOLANA"**

El Comité de Ética del Hospital Militar "Dr. Carlos Arvelo", luego de analizar el proyecto del Trabajo ANÁLISIS DE LA DISTRIBUCIÓN ALÉLICA DE 23 MICROSATÉLITES DE IMPORTANCIA FORENSE EN LA POBLACIÓN VENEZOLANA, y de acuerdo a los Principios Internacionales y de Buena Práctica Clínica y las regulaciones locales, resuelve autorizar el Trabajo Especial de Grado solicitado por Yudis Alejandra Franco Semprun CI: V- 23.691.727 para optar al Título de Licenciada en Biología en la Universidad Central de Venezuela.

Firma  Dr. José Lisandro Pabón Fecha 16/02/18  
 Nombre: CF. Pabón Dávila, José Lisandro  
 Cargo: Presidente (E) del Comité de Ética.  
 Institución: Hospital Militar "Dr. Carlos Arvelo"  
 Dirección de la Institución: Av. José Ángel Lamas, San Martín, Caracas, Venezuela.

**Lista de los Miembros del Comité de Ética:**

Nombre	Profesión	Cargo dentro del C.E
CF. Dr. José Lisandro Pabón	Médico Psiquiatra Farmacólogo	Presidente (E)
TCNEL. Dr. Ramón Manzano	Traumatólogo	Miembro
TCNEL. Arelis Suárez	Farmacéutico Hospitalario	Miembro
Dr. Joao Mauricio De Sousa	Cirujano General	Miembro
Abg. Douglas Aldana	Abogado	Miembro
Lic. Nelva Orozco	Lic. Enfermería	Miembro

