



**UNIVERSIDAD NACIONAL AUTÓNOMA DE
MÉXICO**

FACULTAD DE ESTUDIOS SUPERIORES CUAUTILÁN

Frecuencia alélica de 15 STRs de una
población del centro y sur de México

T E S I S

PARA OBTENER EL TÍTULO DE:
LICENCIADA EN BIOQUÍMICA DIAGNÓSTICA

P R E S E N T A:

MARÍA FERNANDA PÉREZ Y PÉREZ

ASESORA:

M. en C. Maritere Domínguez Rojas

COASESORA:

Dra. Virginia Sánchez Monroy

CUAUTILÁN IZCALLI, ESTADO DE MÉXICO, 2017.



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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



UNIVERSIDAD NACIONAL
AUTÓNOMA DE
MÉXICO

FACULTAD DE ESTUDIOS SUPERIORES CUAUTITLÁN
UNIDAD DE ADMINISTRACIÓN ESCOLAR
DEPARTAMENTO DE EXÁMENES PROFESIONALES

FACULTAD DE ESTUDIOS
SUPERIORES CUAUTITLÁN

ASUNTO: VOTO APROBATORIO



M. en C. JORGE ALFREDO CUÉLLAR ORDAZ
DIRECTOR DE LA FES CUAUTITLÁN
PRESENTE

ATN: I.A. LAURA MARGARITA CORTAZAR FIGUEROA
Jefa del Departamento de Exámenes Profesionales
de la FES Cuautitlán.

Con base en el Reglamento General de Exámenes, y la Dirección de la Facultad, nos permitimos comunicar a usted que revisamos el: Trabajo de Tesis

Frecuencia alélica de 15 STRs en una población del centro y sur de México.

Que presenta la pasante: María Fernanda Pérez y Pérez

Con número de cuenta: 308296934 para obtener el Título de la carrera: Licenciatura en Bioquímica Diagnóstica

Considerando que dicho trabajo reúne los requisitos necesarios para ser discutido en el EXAMEN PROFESIONAL correspondiente, otorgamos nuestro VOTO APROBATORIO.

ATENTAMENTE

“POR MI RAZA HABLARÁ EL ESPÍRITU”

Cuautitlán Izcalli, Méx. a 01 de Agosto de 2017.

PROFESORES QUE INTEGRAN EL JURADO

	NOMBRE	FIRMA
PRESIDENTE	Dra. Sandra Díaz Barriga Arceo	
VOCAL	M. en C. Maritere Domínguez Rojas	
SECRETARIO	Q.F.B. Alejandro Gutiérrez García	
1er. SUPLENTE	L.B.D. Larisa Andrea González Salcedo	
2do. SUPLENTE	L.B.D. Rosa María de los Angeles López Cabrera	

NOTA: los sinodales suplentes están obligados a presentarse el día y hora del Examen Profesional (art. 127).

Frecuencia alélica de 15 STRs en una población del centro y sur de México

Tesis para obtener el título de licenciada en Bioquímica Diagnóstica

**Universidad Nacional Autónoma de México
Facultad de Estudios Superiores Cuautitlán**

Asesora Interna: M. en C. Maritere Dominguéz Rojas
Asesora Externa: Dra. Virginia Sánchez Monroy

**Universidad Nacional Autónoma de México
Facultad de Estudios Superiores Cuautitlán**

*A mi familia
por impulsarme a ser siempre mejor*

Agradecimientos

*Para obrar, el que da debe olvidar pronto, y
el que recibe, nunca.*

Séneca

A la M. en C. Maritere Domínguez Rojas por darme la oportunidad de trabajar con ella, por su invaluable orientación y paciencia.

A la Dra. Virginia Sánchez Monroy por darme las facilidades para desarrollar la parte experimental de mi tesis y por sus acertadas observaciones que me permitieron ir puliendo este trabajo.

A la Dra. Sandra Díaz Barriga Arceo, por compartirme un poco de su experiencia y sabiduría cuando tuve la oportunidad de ser su alumna y por sus valiosas sugerencias e interesantes puntos de vista que enriquecieron de manera importante esta tesis.

Al Q.F.B. Alejandro Gutiérrez García por su invaluable ayuda y orientación oportuna sobre todos los pasos necesarios para realizar el examen profesional, y por el tiempo y esfuerzo invertido en la revisión de esta tesis, a pesar de las tantas actividades que lo ocupan.

Al M. en C. Feliciano Monterrubio por todo su cariño y por el apoyo brindado. Y finalmente a mi familia:

- A mi padre, el Dr. Gerardo Francisco Pérez Sánchez por ser un excelente ser humano y predicar siempre con el ejemplo. En tu caso, hacerte llamar *el magnífico* no es una exageración. Jamás acabaré de agradecerte por todo lo que te debo, empezando por la vida, te amo.
- A mi madre, la Psic. Marina Pérez Rosas, por permitirme aprender de una persona tan inteligente y amorosa como ella. Eres la mejor madre del mundo, te amo muchísimo.
- A mis hermanos: Carlos Pérez y Jacob Pérez, por su cariño incondicional pero sobre todo por su increíble capacidad para inspirarme a ser siempre mejor. Gracias por todas las risas y los buenos momentos, los amo con todas mis fuerzas.

Lista de abreviaturas

- STR- *Short tandem repeats*- Microsatélite
- DNA- *Deoxyribonucleic acid*- Ácido desoxirribonucleico
- RNA-*Ribonucleic acid*- Ácido ribonucleico
- INEGI- Instituto Nacional de Estadística y Geografía
- nm- Nanometros
- HGP- *Human Genome Project*- Proyecto Genoma Humano
- GRC- *Genome Reference Consortium*- Consorcio de Referencia del Genoma
- p- *petite*- pequeño
- q- *quarter*- largo
- tRNA- *transfer RNA*- RNA de transferencia
- rRNA- *ribosomal RNA*- RNA ribosomal
- miRNA- microRNA
- SINE- *Short Interspersed Nuclear Elements*- Elemento intercalado corto
- LINE- *Long interspersed nuclear element*- Elemento intercalado largo
- LTR- *Long terminal repeat*- Repetición terminal larga
- VNTR- *Variable Number of Tandem Repeats*- Número Variable de Repeticiones en Tándem
- SVA- SINE-VNTR-Alus
- HERV- *Human Endogenous Retroviruses* - Retrovirus Endógenos Humanos
- ORF- *Open Reading Frame*- Marco abierto de lectura
- mRNA- *messenger RNA*- RNA mensajero

- TE- *Transposable Elements*- Elementos transponible
- SSR- *Short Sequences Repeat*- Secuencias cortas repetidas
- Kb- Kilobase
- pb- Par de bases
- SNP- *Single Nucleotide Polymorphisms*- Polimorfismo de nucleótido único
- SNV- *Single Nucleotide Variant*- Variante de nucleótido único
- HaMap- *Haplotype Map*- Mapa de haplotipos
- DNA MMR- *DNA Mismatch Repair*- Reparación de DNA a partir de un emparejamiento incorrecto
- DSB- *Double Strand Break*- Ruptura de doble hebra
- PCR- *Polymerase Chain Reaction*- Reacción en cadena de la polimerasa
- μL - Microlitros
- HWE- *Hardy-Weinberg equilibrium*- Equilibrio de Hardy-Weinberg
- PBS- *Phosphate Buffered Saline*- Tampón salino de fosfatos
- rpm- revoluciones por minuto
- H_0 - Hipótesis nula
- H_a - Hipótesis alternativa

Índice

Agradecimientos	VII
Lista de abreviaturas	IX
1. Antecedentes	1
1.1. Introducción	1
2. Marco teórico	3
2.1. Generalidades de los ácidos nucleicos	3
2.2. Estructura y Función del DNA en eucariontes	5
2.3. DNA codificante	8
2.4. DNA no codificante	9
2.4.1. DNA repetitivo disperso	10
2.4.2. DNA repetitivo en tándem	12
2.5. Polimorfismos del DNA	14
2.5.1. Polimorfismos de un solo nucleótido	14
2.5.2. Polimorfismos de mini y microsatélites	16
2.6. Estudio de los microsatélites	19
2.6.1. Clasificación y nomenclatura de los STRs	19
2.6.2. Evolución de los microsatélites	20
2.6.3. Importancia funcional de los microsatélites	23
2.6.4. Aplicaciones de los microsatélites	24
2.7. Genética de Poblaciones y STRs	25
2.7.1. Frecuencias génicas y genotípicas	26
2.7.2. Equilibrio de Hardy-Weinberg	28
3. Planteamiento del problema	31
3.1. Justificación	31
3.2. Hipótesis	31
3.3. Objetivos del trabajo	32
3.3.1. Objetivo General	32
3.3.2. Objetivos Particulares	32

4. Materiales y Métodos	33
4.1. Materiales	33
4.1.1. Cebadores para la PCR	33
4.1.2. Kits comerciales	33
4.1.3. Reactivos	34
4.1.4. Equipos	34
4.2. Metodos	34
4.2.1. Población en estudio	34
4.2.2. Toma de muestra	35
4.2.3. Extracción de DNA	35
4.2.4. Cuantificación de DNA	36
4.2.5. Reacción de PCR	37
4.2.6. Electroforesis capilar	38
4.3. Análisis genético de las muestras	40
5. Resultados	43
5.1. Frecuencias alélicas	43
5.2. Número de alelos	63
5.3. Equilibrio de Hardy-Weinberg	63
5.4. Comparación entre pares usando el estadístico de Wright(F_{ST})	68
6. Análisis de Resultados y Conclusiones	71
6.1. Análisis de resultados	71
6.1.1. Frecuencias alélicas	71
6.1.2. Equilibrio de Hardy Weinberg	71
6.1.3. División de la población usando el estadístico de Wright	72
6.2. Conclusiones	74
Bibliografía	75

Índice de figuras

2.1. Estructuras químicas del ácido ribonucleico RNA y del ácido desoxirribonucleico DNA. Imagen adaptada de Mathews, Holde, y Ahern, 2013	4
2.2. Modelo para el empaquetamiento del DNA en cromatina (Passarge, 2010).	7
2.3. Modelo esquemático del deslizamiento de polimerasa. Modificado de (Jobling et al., 2013)	18
2.4. Modelo esquemático del entrecruzamiento desigual (González-Blanco, 2005).	19
4.1. Pasos a seguir en el protocolo <i>Purification of Total DNA from Animal Tissues (Spin-Column Protocol)</i>	36
4.2. Electroferograma obtenido por el software GeneMapper®ID-X	39
5.1. Representación esquemática de las frecuencias alélicas del locus D8S1179, D21S11 y D7S820.	58
5.2. Representación esquemática de las frecuencias alélicas del locus CSF1PO, D3S1358 y TH01.	59
5.3. Representación esquemática de las frecuencias alélicas del locus D13S317, D16S539 y D2S1338.	60
5.4. Representación esquemática de las frecuencias alélicas del locus D19S433, VWA y POX.	61
5.5. Representación esquemática de las frecuencias alélicas del locus D18S51, D5S818 y FGA.	62
6.1. Localización geográfica de las poblaciones bajo estudio y su agrupación en regiones según el estadístico de Wright	73

Índice de Tablas

2.1. Datos sobre la abundancia de la mayoría de los elementos móviles de genoma humano (Awise, 2010)	11
2.2. Clasificación de los SSRs según su organización (Fernández, 2005).	20
2.3. Frecuencias genotípicas bajo el equilibrio de Hardy-Weinberg (Nielsen y Slatkin, 2013).	28
4.2. Número de individuos estudiados por estado	35
4.3. Reactivos empleados en la preparación de la mezcla de reacción .	37
4.4. Indicaciones para la preparación de muestras	37
4.5. Condiciones de la PCR	38
4.1. Loci y alelos del kit AmpF ℓ STRs®Identifiler®Plus	41
5.1. Frecuencias alélicas encontradas en los 135 individuos de Veracruz.	44
5.2. Frecuencias alélicas encontradas en los 123 individuos de Oaxaca.	45
5.3. Frecuencias alélicas encontradas en los 91 individuos del Estado de México.	46
5.4. Frecuencias alélicas encontradas en los 79 individuos de Chiapas.	47
5.5. Frecuencias alélicas encontradas en los 78 individuos de Puebla. .	48
5.6. Frecuencias alélicas encontradas en los 49 individuos de Guerrero.	49
5.7. Frecuencias alélicas encontradas en los 48 individuos de Hidalgo.	50
5.8. Frecuencias alélicas encontradas en los 31 individuos de Tabasco.	51
5.9. Frecuencias alélicas encontradas en los 20 individuos de la Ciudad de México.	52
5.10. Frecuencias alélicas para el locus D8S1179.	53
5.11. Frecuencias alélicas para el locus D21S11.	53
5.12. Frecuencias alélicas para el locus D7S820.	54
5.13. Frecuencias alélicas para el locus CSF1PO.	54
5.14. Frecuencias alélicas para el locus D3S1358.	54
5.15. Frecuencias alélicas para el locus TH01.	54
5.16. Frecuencias alélicas para el locus D13S317.	55
5.17. Frecuencias alélicas para el locus D16S539.	55
5.18. Frecuencias alélicas para el locus D2S1338.	55

5.19. Frecuencias alélicas para el locus D19S433.	56
5.20. Frecuencias alélicas para el locus VWA.	56
5.21. Frecuencias alélicas para el locus TPOX.	56
5.22. Frecuencias alélicas para el locus D18S51.	57
5.23. Frecuencias alélicas para el locus D5S818.	57
5.24. Frecuencias alélicas para el locus FGA.	57
5.25. Número de alelos diferentes encontrados para cada locus.	63
5.26. Cálculos del HWE para la población de Veracruz.	64
5.27. Cálculos del HWE para la población de Oaxaca.	64
5.28. Cálculos del HWE para la población del Estado de México.	65
5.29. Cálculos del HWE para la población de Chiapas.	65
5.30. Cálculos del HWE para la población de Puebla.	66
5.31. Cálculos del HWE para la población de Guerrero.	66
5.32. Cálculos del HWE para la población de Hidalgo.	67
5.33. Cálculos del HWE para la población de Tabasco.	67
5.34. Cálculos del HWE para la población de CDMX.	68
5.35. Valores de F_{ST} por pares de poblaciones	69
5.36. P- values de F_{ST} entre las 9 poblaciones de estudio	70

Capítulo 1

Antecedentes

Todo comienzo tiene su encanto.

Johann Wolfgang Goethe

Introducción

La genética de poblaciones es la disciplina biológica que suministra los principios teóricos de la evolución. En esta ciencia se parte del hecho de que los cambios evolutivos a pequeña escala se dan en el seno de las poblaciones de las especies (Barbadilla, 2003). Casi todas las especies comprenden una o más poblaciones de individuos que se cruzan entre sí, formando una comunidad de intercambio genético denominada población mendeliana. Una población consiste en un sistema de organismos de una misma especie que conviven en espacio y tiempo y se reproducen entre sí. Para la genética de poblaciones, una población es definida por su reservorio génico, que es el conjunto de todos los alelos de todos los genes de los individuos que la constituyen. El objetivo básico de la genética de poblaciones es caracterizar los reservorios génicos, los cambios en su composición a lo largo del tiempo y del espacio geográfico e investigar los procesos que explican estos cambios (Curtis et al., 2008).

Inicialmente, los estudios de diversidad génica se realizaban a través de electroforesis de proteínas, pero la resolución en los estimados de heterocigosidad era limitada, detectando sólo un 30% de la variación. Con la técnica para secuenciación por dideoxinucleótidos diseñada por Sanger, y el desarrollo de métodos cada vez más novedosos se han ampliado considerablemente el número de marcadores empleados en la genética de poblaciones.

Los STRs, (*del inglés: Short Tandem Repeats*) son unidades de DNA de 1 a 6 pares de bases de longitud que se distribuyen ampliamente por todo el genoma nuclear de eucariontes (Putman y Carbone, 2014). Debido a que son altamente polimórficos, los STRs, se han utilizado ampliamente para realizar estudios de macroevolución, tanto del ser humano como de otras especies debido a que el cambio

en la extensión de un repetido de nucleótidos origina un nuevo alelo y cada repetido se puede considerar como un alelo neutral, por lo tanto la heterocigosidad observada en la muestra se puede adecuar a un modelo de alelos infinitos, partiendo de la teoría neutral. Se han hecho estudios de diversidad genética en la población mexicana mestiza y amerindia, utilizando varios locus de STRs. Los locus que han sido estudiados han sido APO-B, vWA, TH01, CSF1PO, HPRTB, FGA. Los alelos vWA y TH01 pueden distinguir a la población mexicana, respecto a otras poblaciones como caucásicas o africanas (Guardado-Estrada et al., 2008). La población mexicana actual surge desde el contacto europeo con el nuevo mundo, lo que origina una mezcla de dos razas que geográficamente estuvieron separadas por mucho tiempo, la amerindia y la europea. Además de ascendencia europea, en México existe la mezcla de ciertos linajes africanos traídos como esclavos de varias regiones de África. Sin embargo, hay que tener en cuenta que los genes africanos podrían haber llegado a México de los soldados españoles de ascendencia morisca como resultado de la ocupación islámica de la Península Ibérica durante casi ocho siglos (Rubi-Castellanos et al., 2009c). Los 500 años de la mezcla entre los indígenas americanos, esclavos africanos y europeos ha formado a la mayoría de la población actual de México. Esta población de habla hispana es conocida actualmente como mestiza. Para que un individuo sea considerado como mestizo tiene que haber nacido en México, tener un apellido de origen español y ancestros de origen mexicano tres generaciones hacia atrás. Según el INEGI, el 93 % de la población Mexicana es mestiza. A pesar de esto, se han estimado diferentes componentes amerindios en las poblaciones americanas, que van desde el 0.9 hasta el 94.5 % (Bonilla et al., 2005). Debido a esto, la composición génica de los mexicanos es compleja, pues el mestizaje depende de diversos aspectos geográficos, económicos, culturales y educativos, observándose una estratificación genética.

Además de ayudar a entender la composición génica de la actual población mexicana, el conocimiento de la frecuencia alélica de los distintos STRs tiene diversas aplicaciones. En el campo de la medicina forense proporcionan los datos necesarios para mejorar la identificación de individuos; y en la medicina genómica facilitan la predicción de la predisposición de la población a diversas enfermedades asociadas con algún alelo de los STRs estudiados.

Capítulo 2

Marco teórico

Creo que la razón por la que las personas se interesan menos en la ciencia hoy en día que hace cincuenta años atrás es que se ha hecho muy complicada

James Watson

Generalidades de los ácidos nucleicos

Muchas ideas importantes acerca de la genética fueron descubiertas mucho tiempo antes de que existiera cualquier conocimiento sobre la molécula responsable de la transmisión de la herencia (Stram, 2014).

Gracias a científicos como F. Griffith, Avery, MacLeod, McCarty y demás colaboradores, en 1944 se sabía que una cadena larga de química relativamente simple, un ácido nucleico, llevaba la información genética en las bacterias (Passarge, 2010). En 1952, Hershey y Chase probaron que la información genética solo se transfiere mediante DNA y excluyeron otras moléculas. Con este descubrimiento, la pregunta acerca de la estructura de DNA cobró gran importancia en la Biología (Passarge, 2010). Cuando en 1953 Watson y Crick finalmente publicaron sus descubrimientos acerca de la estructura molecular del ácido desorribonucleico o DNA ya existía una gran cantidad de conocimientos acerca de la naturaleza de la herencia en plantas, animales y la especie humana. Las implicaciones inmediatas de este descubrimiento en la herencia mendeliana, en conjunto con los descubrimientos de H.G Khorana, M.Nirenberg y colaboradores respecto a la relación entre los codones del DNA y los aminoácidos, revolucionaron la historia de la genética.

Hoy en día se sabe que existen dos tipos de ácidos nucleicos: el DNA (*del inglés: Deoxyribonucleic Acid*) y el RNA (*del inglés: Ribonucleic Acid*). Cada uno de ellos es una cadena polimérica, en la que las unidades monoméricas están conectadas por enlaces fosfodiéster. En ambos casos la unidad monomérica está formada por un azúcar de cinco carbonos (la ribosa en el RNA y la desoxirribosa en el DNA). La única diferencia entre los dos azúcares radica en el sustituyente

del carbono 2': la ribosa tiene un grupo hidroxilo en esa posición, mientras que la desoxirribosa tiene un hidrógeno (Mathews et al., 2010) (véase figura 2.1).

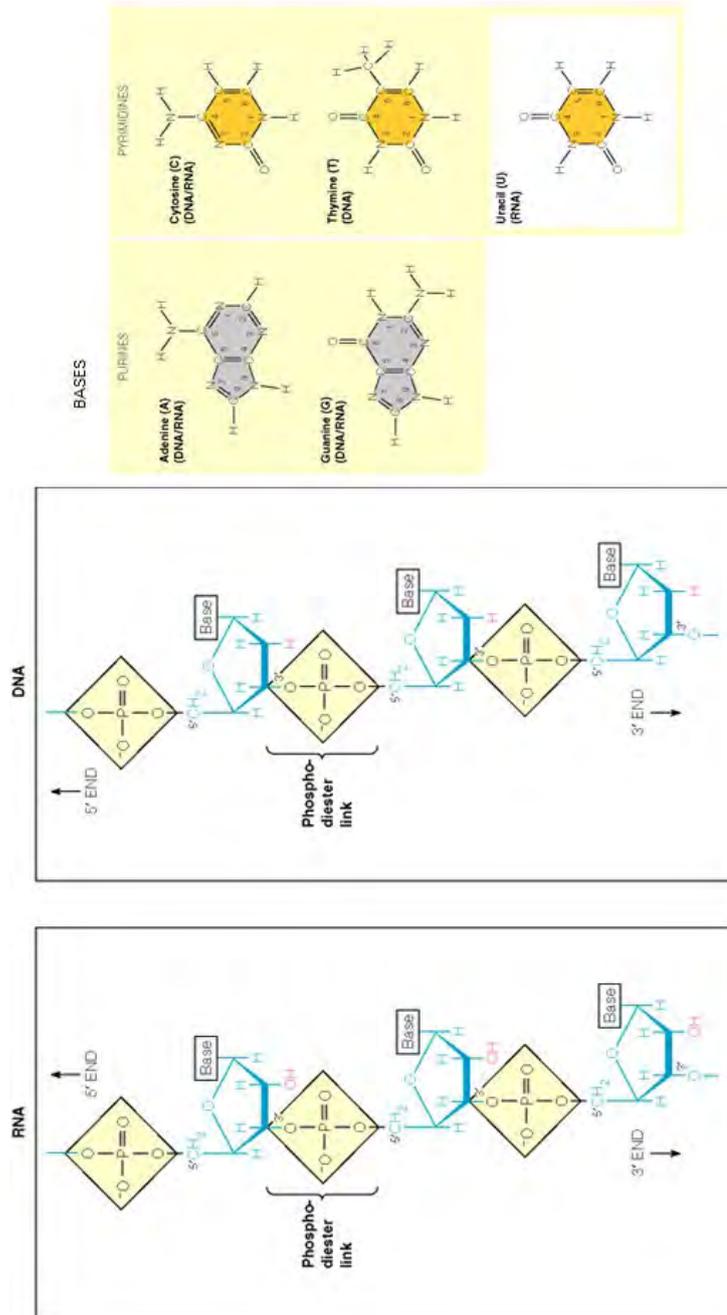


Figura 2.1: Estructuras químicas del ácido ribonucleico RNA y del ácido desoxirribonucleico DNA. Imagen adaptada de Mathews, Holde, y Ahern, 2013

Los residuos de azúcar unidos mediante enlace fosfodiéster constituyen el armazón de los ácidos nucleicos. Cada monómero de la cadena contiene además una base heterocíclica, que siempre va unida al carbono 1' del azúcar por un enlace N-glucosídico (véase figura 2.1). Existen dos tipos de bases heterocíclicas: purinas o pirimidinas. El DNA tiene dos purinas: adenina (A) y guanina (G) y dos pirimidinas: citosina (C) y timina (T). El RNA posee las mismas bases, con la diferencia de que la timina se sustituye por el uracilo (U) (Mathews et al., 2010).

Dado que todos los ácidos nucleicos pueden considerarse polímeros de nucleótidos, se les suele asignar la denominación genérica de polinucleótidos. A los polímeros pequeños, que solo tienen algunos residuos se les denomina oligonucleótidos.

La estructura primaria de los ácidos nucleicos consiste en una cadena polinucleotídica que posee un sentido o direccionalidad. El enlace fosfodiéster entre las unidades monoméricas se produce entre el carbono 3' de un monómero y el carbono 5' del siguiente. Esto provoca que los dos extremos de una cadena polinucleotídica lineal sean diferenciables. Un extremo lleva normalmente un fosfato 5' sin reaccionar y el otro extremo un grupo 3' sin reaccionar. Por convención, la secuencia de una cadena polinucleotídica suele escribirse con el extremo 5' a la izquierda y el 3' a la derecha (Mathews et al., 2010). La estructura secundaria de los ácidos nucleicos consta de dos largas cadenas polinucleotídicas enrolladas alrededor de un eje central, formando una doble hélice de 20 nm de diámetro, enrollada hacia la derecha. Las dos cadenas son antiparalelas, es decir, que van en sentidos opuestos, y las bases de las dos cadenas se encuentran al interior de la estructura, formando escalones planos y perpendiculares al eje, separados por 0.34 nm (Mathews et al., 2010). Existe un apareamiento específico entre las bases A-T y G-C en el DNA ó A-U y G-C en el caso del RNA. Las bases AT y AU se unen mediante dos puentes de hidrógeno, mientras que las bases GC se unen con tres puentes de hidrógeno. Cada vuelta completa de la hélice tiene una longitud de 3.4 nm, de este modo cada vuelta de la cadena contiene 10 bases (Klug et al., 2006).

Estructura y Función del DNA en eucariontes

En los individuos eucariontes el DNA se organiza en cromosomas. Los cromosomas están compuestos de cromatina espiralizada. La cromatina se compone de una molécula de DNA con una cantidad sustancial de proteínas asociadas (Knight, 2009). Estas proteínas se clasifican en histonas y no histonas:

Histonas: Son proteínas con grandes cantidades de lisina y arginina, aminoácidos que están cargados positivamente, lo que les posibilita unirse electrostáticamente a los grupos fosfato con carga negativa de los nucleótidos. Existen cuatro tipos diferentes de histonas: H2A, H2B, H3, H4 y H1.

Proteínas no histonas: Tienen funciones estructurales en la configuración cromosómica, permitiendo la formación de cromátidas.

Un tetrámero es un conjunto de cuatro histonas en el que se enrolla la doble hélice de DNA. Las histonas H2A, H2B, H3 y H4 forman dos tipos de tetrámeros:

$$(H2A)_2 - (H2B)_2$$

y

$$(H3)_2 - (H4)_2$$

Un nucleosoma está formado por un tetrámero de cada tipo, asociado a 147 pares de bases, unido con otros nucleosomas por la histona H1. La formación del nucleosoma representa el primer nivel de empaquetamiento del DNA, creando una fibra de cromatina de 11 nm de diámetro. Esta fibra de 11 nm se empaqueta aún más, generando una fibra más gruesa de 20 nm de ancho, conocida como solenoide, que consiste en seis nucleosomas estrechamente enrollados. Esto es conocido como el segundo nivel de empaquetamiento. La estructura de 30 nm forma una serie de dominios en forma de lazo formados por las proteínas no histonas, que condensan aún más la cromatina, originando estructuras de 300 nm. Estas fibras de 300 nm se enrollan hasta crear una cromátida de 700 nm de diámetro. Un par de cromátidas hermanas constituyen un cromosoma (Passarge, 2010) (véase imagen 2.2).

El genoma humano está compuesto de 22 autosomas y 2 cromosomas sexuales X o Y. La mayoría de las células del cuerpo contienen dos copias del genoma, es decir, son diploides. Una célula diploide normal contiene un total de 46 cromosomas, 22 pares de autosomas homólogos, y un par de cromosomas sexuales; las mujeres tienen el par XX, y los hombres, el XY. Existen también células que contienen una sola copia del genoma; estas células son llamadas haploides y son el óvulo y el espermatozoide, células encargadas de la reproducción (Knight, 2009). Es posible encontrar muchos genes en un sólo cromosoma, aunque cada cromosoma tiene diferente número de genes. El número de genes que tiene cada cromosoma se puede encontrar en la página del *Genome Reference Consortium* (<http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/human/>), aunque es importante considerar que en el caso de los genes autosómicos habrán dos copias de cada gen en una célula diploide, uno por cada cromosoma (Knight, 2009).

Todos los cromosomas tienen un centrómero que se hace evidente durante la duplicación cromosómica antes de la división celular. El centrómero divide al cromosoma en dos brazos. El brazo más corto de los dos es el p *petite* y el más largo es el brazo q *quarter* (Stram, 2014). Las diversas técnicas de bandeado cromosómico han facilitado notablemente la identificación visual y agrupación de cada cromosoma. Además de servir para diferenciar a los cromosomas, las tinciones cromosómicas pusieron de manifiesto que la mayor parte de un cromosoma se encuentra presente como cromatina desespirilizada, mientras que algunas partes permanecen condensadas y se tiñen fuertemente durante la interfase. Este descubrimiento llevó a que en 1928 se acuñaran los términos heterocromatina y eucromatina para describir las partes de los cromosomas que estaban enrolladas o desespirilizadas, respectivamente (Klug et al., 2006). A partir de este descubrimiento, se observó

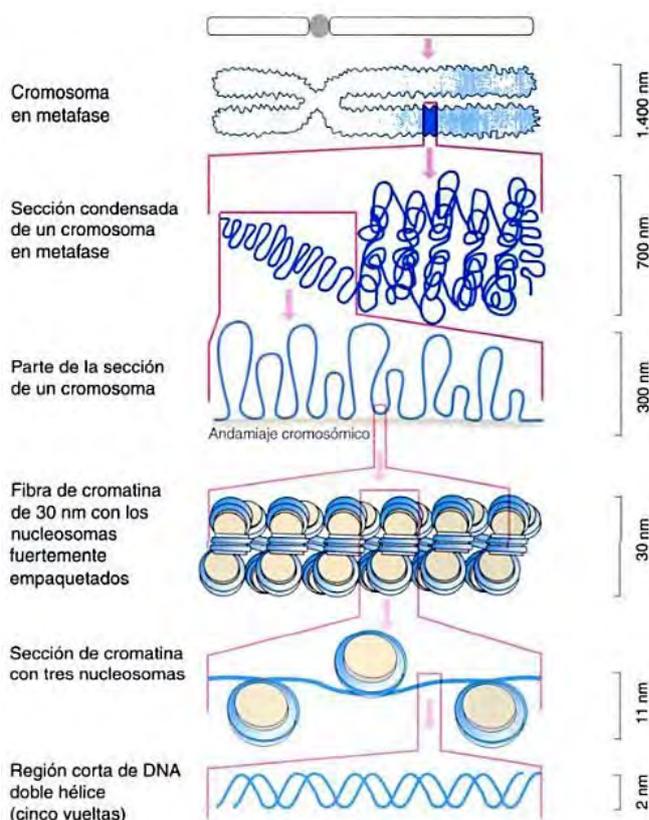


Figura 2.2: Modelo para el empaquetamiento del DNA en cromatina (Passarge, 2010).

que las áreas heterocromáticas son genéticamente inactivas y se replican más tarde que la eucromatina durante la duplicación del DNA en la división celular.

El hecho de que la eucromatina es genéticamente más activa que la heterocromatina llevó al interés de la comunidad científica por esta parte del genoma. Por este motivo, en el año de 1988 inició una investigación de esta parte del genoma, conocida como HGP (*del inglés: Human Genome Project*). Este proyecto se logró gracias a la colaboración de diversos organismos internacionales coordinadas por el Departamento de Energía y el Instituto Nacional de Salud de Estados Unidos de América. Los objetivos principales de este proyecto fueron la determinación de la secuencia de los cromosomas humanos y la identificación y mapeo de genes (ENA, 2015). El HGP usó un enfoque basado en clones para producir los mapas y ensamblar la secuencia del genoma humano. En febrero del 2001 se publicó el primer borrador del genoma humano en la revista *Nature*, con el 90% de la secuencia completa. La secuencia completa se publicó en Abril del 2003, en la misma revista (ENA, 2015). En el año 2008 se formó el GRC, (*del inglés: Genome Reference*

Consortium) encargado de mantener actualizada la información referente al genoma humano. Los objetivos generales de este grupo son: la corrección de ciertas regiones tergiversadas, eliminar las brechas restantes y la producción de secuencias alternativas para los locis con variantes estructurales. El GRC está formado por *The Wellcome Trust Sanger Institute*, *The Genome Center at Washington University*, *The European Bioinformatics Institute* y *The National Center for Biotechnology Information* (ENA, 2015).

Según la versión 78.38 de Ensembl (Ensembl, 2015), el genoma humano tiene un total de 3,381,944,086 pares de bases. En todo el genoma, solo existen 22,469 genes que se convertirán en proteína, conocidos como genes codificantes. La longitud total de todos los exones de los genes codificantes es de aproximadamente 34 Mb, lo que representa aproximadamente un 1.2% del genoma humano (IHGSC, 2004). Así pues, el HGP demostró que menos del 2% del genoma es codificante, mientras que los elementos repetitivos del genoma representan más del 50% del genoma (IHGSC, 2004). El resto del DNA codifica para diferentes tipos de RNA, incluyendo el RNA de transferencia o tRNA, RNA ribosomal o rRNA y microRNAs o miRNA. Además el DNA contiene muchas otras partes que se encuentran involucradas en la regulación génica, como las secuencias promotoras (Stram, 2014).

DNA codificante

La función de los genes codificantes encontrados en el genoma es obvia: la síntesis de proteínas, que realicen diversas funciones en el organismo. Para que un gen codificante pueda convertirse en una cadena de aminoácidos, es necesario en primer lugar que la secuencia se convierta en RNA. Generalmente la secuencia de DNA codificante para una proteína se encuentra contenida en exones, intercalados con secuencias no codificantes, conocidas como intrones (Stram, 2014). El paso de DNA a RNA se conoce como *transcripción*. Este proceso involucra factores de transcripción que reconozcan específicamente las secuencias no codificantes reguladoras y la separación física de las dos hebras de DNA para que la RNA polimerasa (enzima encargada de la transcripción) pueda unir las bases al nuevo RNA sintetizado. Durante la transcripción, el DNA molde es leído por la RNA polimerasa en dirección 3' a 5', de modo que el RNA se va ensamblando en dirección 5' a 3' desde el inicio del primer exón, hasta el último. Los intrones se eliminan del RNA mensajero recién copiado mediante un proceso llamado empalme o *splicing*, formándose un RNA maduro. Muchos genes pasan por un proceso de empalme alternativo, pudiéndose eliminar algunos exones, o incluir trozos de intrones en el RNA maduro. El proceso de empalme alternativo permite que de un mismo gen, se produzcan diferentes proteínas. El último paso para la creación de una proteína se llama *traducción*. La traducción implica la migración del RNA maduro al citoplasma de la célula, para su unión con un complejo de RNAmicrosomal/proteínas, llamado ribosoma. En el ribosoma el RNA mensajero maduro interactúa con mo-

léculas de RNA de transferencia o tRNA; que reconoce los codones (mediante una secuencia antisentido complementaria a los codones de RNA mensajero) y recluta los aminoácidos específicos que se unirán a la cadena polipeptídica en crecimiento (Stram, 2014). Un codón es el conjunto de bases que codifica un aminoácido único. En un principio, muchos investigadores reconocieron que los codones debían de contener un mínimo de tres nucleótidos. Si un codón estuviera formado por un único nucleótido, entonces solo serían posibles 4 codones diferentes, lo que no es suficiente para codificar los 20 aminoácidos distintos. De la misma manera, si un codón estuviera hecho de dos nucleótidos habría sólo 16 codones posibles. En cambio con tres nucleótidos por codón existen 64 codones posibles, suficientes para especificar los 20 aminoácidos (Pierce, 2009). De los 64 codones posibles, tres son codones de terminación, que especifican la finalización de la traducción. Los 61 codones restantes codifican los otros 20 aminoácidos; esto quiere decir que se puede especificar un mismo aminoácido con varios codones. Algunos aminoácidos están codificados hasta por 6 codones, como es el caso de la leucina. Solo el triptófano y la metionina están codificados por un único codón (Pierce, 2009).

DNA no codificante

La función del DNA no codificante no siempre fue clara; de hecho, en un principio los elementos repetitivos del genoma fueron considerados como DNA "basura", sin ninguna función importante (IHGSC, 2004). Sin embargo, conforme se profundizó en el estudio del genoma humano, se descubrió que los genes no codificantes tienen funciones muy importantes. Algunos, como los RNAs no codificantes, se encuentran involucrados en la regulación de la expresión génica. Otros como las secuencias promotoras o los tRNAs se encargan de llevar a cabo correctamente procesos como la transcripción y traducción, respectivamente. Mientras, que los elementos repetitivos del genoma tienen varias funciones: por una parte, las repeticiones constituyen un rico registro paleontológico, que contiene claves cruciales sobre los eventos y fuerzas evolutivas. Así mismo, estas zonas han causado reordenamientos y generando nuevos genes o modificando los ya existentes, lo que se emplea como herramienta para la medicina genómica o para la genética de poblaciones (IHGSC, 2004). Para facilitar el estudio del DNA no codificante, éste se ha clasificado de varias maneras. En este documento, el DNA repetitivo se agrupará en dos grandes rubros:

1. DNA repetido disperso: Comprende a los elementos dispersos por diferentes partes del genoma, en este rubro incluye todos los elementos trasponibles, cuya naturaleza se explicará más adelante.
2. DNA repetido en tándem: Son repetidos de un número particular que se agrupan en ciertas partes del genoma.

DNA repetitivo disperso

La mayoría de las secuencias repetidas en el genoma humano se derivan de los elementos transponibles. Estos elementos abarcan aproximadamente el 45 % de todo el genoma humano (IHGSC, 2004). Un elemento transponible o móvil es una secuencia de nucleótidos con la capacidad de moverse en el genoma, ya sea por sí misma o con la ayuda de secuencias auxiliares (Hackenberg, 2005). En el genoma humano los elementos transponibles se dividen en cuatro clases principales: 1) Elementos cortos de DNA disperso o SINEs (*del inglés: Short Interspersed Nucleic Element*), 2) Elementos largos de DNA disperso o LINEs (*del inglés: Long Interspersed Nucleic Element*), 3) Elementos con repeticiones terminales largas o LTRs (*del inglés: Long Terminal Repeat Elements*) y 4) Transposones de DNA (IHGSC, 2004). Los cuatro elementos transponibles existentes se dividen a su vez en dos grandes categorías dependiendo su método de transposición. Estas categorías son:

- Elementos de clase I o retrotransposones
- Elementos de clase II o transposones de DNA

Los elementos de clase I o retrotransposones llevan a cabo su replicación mediante la transcripción inversa de un RNA intermediario (Avisé, 2010). Dentro de la categoría de los retrotransposones se encuentran tres principales subclases: 1) LINEs, 2) SINEs y 3) LTRs. Cada una de estas subclases se compone a su vez de subfamilias. Entre los LINEs, la subclase más grande es la de los elementos L1. Cada elemento L1 completo llega a medir aproximadamente 6,000 pares de bases, existiendo en el genoma humano más de 500,000 de estos elementos, comprendiendo en conjunto 17 % del genoma humano (IHGSC, 2004). La secuencia más representativa de los SINEs es la Alu, llamada así por tener un patrón de reconocimiento de la enzima de restricción AluI (Hackenberg, 2005). Se estima que la secuencia Alu se encuentra 1,090,000 veces en el genoma humano, y dado que éstas miden aproximadamente 300 pares de bases de longitud, abarcan alrededor del 11 % del genoma (Avisé, 2010). Otra particularidad de las secuencias Alu es su alto contenido en G+C (en torno al 57 %) (Hackenberg, 2005). Los elementos SVA (*SINE-R*, *VNTR*, y *Alu*) son los más representativos de la subfamilia de los SINEs. Cada SVA tiene una longitud aproximada de 1,500 pares de bases y cerca de 3,000 copias en el genoma humano. Por último, por parte de los LTRs, la subclase más abundante son los HERV (Human Endogenous Retroviruses), que solos representan el 8 % del genoma humano (Avisé, 2010). En la tabla 2.1 hay información más detallada sobre la cantidad de elementos móviles que existen en el genoma humano.

Como ya se dijo anteriormente, los elementos de clase I se producen mediante la transcripción inversa de un RNA intermediario. Para transcribirse a un RNA intermediario, estos elementos usan recursos celulares como la polimerasa III (en el caso las Alus) y la polimerasa II (en el caso de los LINEs). El proceso de transcripción inversa a DNA e inserción en el genoma es un poco más complicado. En

Categoría	No. de copias	% del genoma	Tamaño de los elementos básicos (pb)
Transposones	294,000	2.8	2,500
Retrotransposones			
LINEs	868,00	20.4	6,000
SINEs	1,558,00	13.1	300
LTRs	443,000	8.3	300
Total	3,163,000	44.6	

Tabla 2.1: Datos sobre la abundancia de la mayoría de los elementos móviles de genoma humano (Avise, 2010)

general se requieren de dos enzimas: una transcripasa inversa y una endonucleasa encargada del corte de la secuencia donde se produce la inserción. Los elementos LINE1 poseen dos marcos abiertos de lectura u ORFs (*del inglés: Open Reading Frame*) (ORF1 y ORF2). Se ha podido demostrar que el ORF2 codifica para una transcriptasa inversa que además posee un dominio de endonucleasa. Aunque todavía no se entiende completamente el papel que desempeña la proteína del ORF1, parece que su función es fijar el mRNA del elemento LINE (Hackenberg, 2005). Una vez que el RNA del LINE ha sido traducido, este se reúne con sus propias proteínas codificadas y se mueve hacia el núcleo, donde la actividad endonucleasa de la polimerasa realiza una muesca en el DNA monocatenario, mientras que la transcriptasa inversa utiliza el DNA mellado como cebador para realizar la transcripción inversa desde el extremo 3' del RNA perteneciente al LINE (IHGSC, 2004).

Los elementos LINE son autónomos, ya que codifican todas las proteínas necesarias para su propia reproducción. Por el contrario, las Alus no codifican ninguna proteína y por lo tanto dependen de enzimas procedentes de otros elementos. Hay varias evidencias que demuestran que las Alus utilizan la maquinaria enzimática de los elementos LINE para llevar a cabo el proceso de retrotransposición (Hackenberg, 2005). Se cree que la maquinaria enzimática de los LINEs es la responsable de la mayoría de transcripción inversa en el genoma, incluyendo la retrotransposición de varios elementos no autónomos, como los SINEs y la creación de pseudogenes (IHGSC, 2004).

Por otra parte, los elementos de clase II o transposones de DNA no necesitan de un RNA intermediario, pues realizan el proceso de transposición de manera muy similar a los transposones bacterianos: mediante terminales de repeticiones invertidas que codifican a la enzima transposasa. Esta enzima se une cerca de las repeticiones invertidas y media la movilidad del transposón a través de un proceso conocido como *cortado y pegado* (Klug et al., 2006) (IHGSC, 2004).

En general, los elementos transponibles o TEs (*del inglés: Transposable Elements*) son encontrados en diferentes tipos de regiones genómicas, como la cromatina pericentromérica, telómeros, regiones regulatorias, exones e intrones, siendo más frecuentes en la heterocromatina y las regiones reguladoras. Los TEs son encontrados de manera abundante en regiones genómicas cercanas al telómero y a lo largo de los brazos cromosómicos intercalados en la heterocromatina. En teoría

los elementos trasponibles pueden moverse en cualquier parte del genoma, pero se cree que las inserciones de TEs en la heterocromatina son menos perjudiciales que las inserciones en la eucromatina y su concentración en las regiones de baja densidad génica son un mecanismo para evitar la recombinación ectópica (Miller y Capy, 2006).

Desde que se descubrió la existencia de los elementos trasponibles en el genoma humano, los investigadores se han preguntado si éstos tienen alguna función o simplemente son fragmentos de DNA parásito, cuyo único fin es diseminarse por el genoma, generando múltiples copias de sí mismo dentro del genoma del hospedero. En la actualidad es evidente que los genomas pueden beneficiarse de la presencia de elementos trasponibles, pues estos actúan como un factor que acelera la evolución del genoma, en particular cuando estos se enfrentan a estrés poblacional o ambiental (Miller y Capy, 2006).

DNA repetitivo en tándem

El DNA repetitivo en tándem tiene varias características que lo hacen notablemente diferente de las secuencias trasponibles. La característica más importante es que estos tipos de repetidos se encuentran agrupados en cierta región del genoma sin ser separados por ningún otro tipo de secuencia. A este tipo de DNA también se le conoce como secuencias cortas repetidas o SSR (*del inglés: Short Sequence Repeats*), ya que son secuencias relativamente cortas, comparadas con otros elementos (IHGSC, 2004). La longitud de estas secuencias es un criterio que se utiliza para su clasificación. Según este criterio, las SSRs se clasifican en tres grupos:

- **Satélites:** Repetidos comprimidos en arreglos largos, típicamente de 100 kb o varias megabases de tamaño. La unidad de repetición varía entre 5 y 71 pb.
- **Minisatélites:** Repetidos en arreglos de tamaño intermedio (entre 100 pb y 20 kb). Cada unidad de repetición tiene entre 6 y 100 pb de longitud.
- **Microsatélites:** Arreglos cortos de menos de 100 pb de longitud, hechos de repetidos simples de 1 a 6 pb de longitud. Sin embargo, en ocasiones el límite entre microsatélites y minisatélites se ve desdibujado, pues en algunas fuentes se considera que las unidades de repetición de los microsatélites pueden tener una longitud de 6 a 12 pb.

Como ya se mencionó, el DNA satélite comprende matrices muy grandes de DNA no codificantes repetidas en tándem. Estas agrupaciones de megabases de tamaño constituyen el componente principal de la heterocromatina. El DNA satélite es el componente principal de los centrómeros funcionales, así como de la heterocromatina pericentromérica, de las regiones teloméricas y de los brazos cortos de los cromosomas acrocéntricos (Cromosomas 13,14,15,21 y 22) (Knight, 2009). Dado que la heterocromatina se encuentra en un estado sumamente condensado

y empaquetado, aún existen algunas dificultades técnicas en la secuenciación de estas regiones, por lo que la comprensión de la organización precisa y función de este tipo de repetidos aún es imprecisa. Sin embargo, se estima que el DNA satélite representa del 2 % al 6 % del genoma (Pierce, 2009). La composición del DNA satélite es variable y típica de cada especie. Los primeros trabajos realizados para identificar el DNA satélite utilizaron gradientes de densidad, debido a que este tipo de DNA tiene porcentajes diferentes de las bases nitrogenadas y en consecuencia se separan como varias fracciones cuando se centrifugan a velocidades elevadas (Pierce, 2009). Usando la centrifugación en gradiente se pueden distinguir tres tipos de DNA satélite: 1) el DNA Satélite 1, con una unidad de repetición 42 nucleótidos 2) el DNA Satélite 2 cuya unidad de repetición de 10 nucleótidos es (ATTCCATTCG) y 3) el DNA Satélite 3, que contiene repeticiones del pentámero (ATTCC) (Cal-Teba, 2000). Existe además otro tipo de DNA satélite que no es posible separar fácilmente por centrifugación. Este tipo de DNA satélite se caracterizó por primera vez con enzimas de restricción que reconocen ciertas regiones en las unidades de repetición. El satélite alfoide o satélite alfa pertenece a esta categoría. La secuencia alfoide, que se encuentra principalmente en las regiones centroméricas, tiene un tamaño de 171 pb y es rica en AT. Otros tipos de DNA satélite son el satélite beta (repetición de 68 pb) y el satélite gamma (repetición de 220 pb), que también se encuentran en la cromatina centromérica de varios cromosomas (Cal-Teba, 2000) (Knight, 2009) (Klug et al., 2006). Cuando fue identificado por primera vez, el DNA satélite fue considerado como DNA basura, aunque hoy en día se sabe que juega un papel fundamental en la organización y emparejamiento de los cromosomas (Knight, 2009).

Los microsatélites y minisatélites están estrechamente relacionados y son notoriamente diferentes del DNA satélite. Esta clase de repetidos está compuesta por subunidades de menos de 13 pb de tamaño. Como ya se mencionó, existen diferentes criterios acerca del tamaño de las subunidades de micro y minisatélites, y por este motivo puede ser difícil distinguir un grupo de otro basándose únicamente en el tamaño de las unidades de repetición (Salipante, 2013), (Debrauwere et al., 1997) (Knight, 2009). A pesar de la confusión que se puede generar entre estos dos tipos de repetidos, Debrauwere y colaboradores señalan otras características que pueden tomarse en cuenta para distinguir a un grupo de otro: 1) Los minisatélites forman arreglos de más de 100 pb y hasta 20 kb, mientras que los microsatélites forman arreglos pequeños, que como máximo alcanzan los 100 pb. 2) Los minisatélites se encuentran abundantemente en las regiones centroméricas, subteloméricas y en ciertos cromosomas, como el 19, mientras que los microsatélites se encuentran dispersos por todos los cromosomas, en las regiones no codificantes cercanas a las sí codificantes. Además algunos autores consideran como características importantes de los minisatélites el alto contenido en G-C y su inestabilidad en líneas germinales (Debrauwere et al., 1997).

Tanto microsatélites como minisatélites son altamente polimórficos, lo que quiere decir que existen varios alelos (con diferente número de repetidos) para

un solo locus. Esta característica, de la cual se hablará más extensamente en la siguiente sección, es la que los ha convertido en objeto de estudio para distintos campos.

Polimorfismos del DNA

Un polimorfismo se define como una de las múltiples formas de un gen o molécula normal, presente en al menos el 1 % de la población. Es importante diferenciar un polimorfismo de un heteromorfismo. Un heteromorfismo es una región cromosómica microscópicamente visible que varía en tamaño, morfología y tinción en diferentes individuos, siendo esta una forma normal de la variación del genoma humano. Así pues, diferencia principal entre heteromorfismo y polimorfismo es que el polimorfismo se refiere a variantes del genoma que solo pueden ser detectadas a nivel molecular (Wyandt y Tonk, 2012). Existen varios tipos de polimorfismos en el genoma humano, tanto en las regiones codificantes como en las no codificantes. En este trabajo solo se discutirán tres tipos de secuencias polimórficas: Los SNPs, VNTRs y STRs.

Polimorfismos de un solo nucleótido

El polimorfismo más común del genoma humano, que se encuentra tanto en secuencias codificantes como en no codificantes, es conocido como polimorfismo de un sólo nucleótido o SNP (*del inglés: Single Nucleotide Polymorphism*). Como su nombre lo dice, un SNP consiste en el polimorfismo de una sola base en una posición específica del DNA, con el cambio medido en relación con alguna secuencia consenso existente (Stram, 2014). En ocasiones las variaciones de un solo nucleótido no pueden considerarse como un SNP, pues no se presentan en el 1 % de la población, en cuyo caso, son llamados variante de un sólo nucleótido o SNV, (*del inglés: Single Nucleotide Variant*). En ocasiones la distinción entre un SNV y un SNP se hace de manera arbitraria, pues esta depende del tamaño de la población en estudio. Un SNV puede originarse cuando ocurre un error en la copia de un solo nucleótido durante la replicación del DNA. Este error puede ocurrir durante la mitosis o la meiosis. Sin embargo, si ocurriera durante la mitosis, no sería heredado de generación en generación y solo se presentaría en un reducido grupo de células que descendieran directamente de la célula con el SNV. En cambio, si el error ocurriera durante la meiosis, el SNV se encontraría presente en el DNA del óvulo o espermatozoide, y podría heredarse. Es por esto, que se ha propuesto que un SNP se origina gracias a una mutación puntual del DNA durante la meiosis. Es importante señalar que la tasa de mutación en el proceso de meiosis es de 10^{-8} , y que además existen varios mecanismos mediante los cuales el DNA dañado se repara completamente. Debido a que las mutaciones puntuales en el genoma son raras, se sostiene la teoría de que la gran mayoría de SNPs que hay en el DNA de una persona fueron heredados de uno o ambos padres (Stram, 2014). Actual-

mente se han encontrado 104,000,000 SNPs en todo el genoma humano, pero se ha estimado que existen aproximadamente 20 millones de estos en todo el genoma, lo que llevaría a una distribución de un SNP por cada 100 a 300 nucleótidos (Ramírez-Bello et al., 2013). Debido a su amplia distribución, estos polimorfismos se localizan en cualquier parte del genoma, generando diversos efectos en el fenotipo de quien los posee. Con base en esto, los SNPs se pueden clasificar de acuerdo con su importancia funcional. Los SNPs *no funcionales* son aquellos que no generan ningún cambio en la cantidad o actividad de las proteínas codificadas en los respectivos genes. Esta clase de SNP suele encontrarse fuera de las regiones codificantes, aunque también podrían encontrarse dentro de algún gen, cambiando el codón, pero no el aminoácido resultante, debido a la degeneración del código genético. Por otra parte, los SNPs *funcionales* desempeñan un papel biológico importante en el desarrollo de enfermedades comunes debido a que pueden afectar al gen, al RNAm de genes que sintetizan proteínas o incluso a las mismas proteínas (Ramírez-Bello et al., 2013). Los SNPs funcionales a su vez se subdividen en:

- rSNP: Este tipo de SNP se encuentran en los promotores de los genes que sintetizan proteínas y afectan a los niveles de expresión génica.
- srSNP: SNP que puede localizarse en el RNAm maduro o en el inmaduro, en regiones no traducidas, regiones intrónicas y codificantes llegando a afectar a la estructura y función normal de los mRNA, incluyendo el corte y empalme, la regulación de la traducción de los mRNA a proteínas, la estabilidad de los mRNA, la poliadenilación de los mRNA, la funcionalidad de las proteínas y otros procesos biológicos normales.
- cSNP: SNP que se localiza dentro de los exones de un gen y cuyo cambio de nucleótido genera el cambio de un aminoácido, modificando la secuencia de aminoácidos de la proteína resultante.

El estudio de los SNPs tiene múltiples aplicaciones en diversas áreas de la ciencia. Es posible utilizar a los SNPs para estudiar la variación de la secuencia de DNA entre especies, ayudando a localizar los puntos de ramificación de cada especie, lo que ayudaría a la comprensión de como ha evolucionado el genoma moderno (Shastry, 2002). Se han iniciado grandes consorcios, como el HaMap (*del inglés: The haplotype map*) con el único fin de identificar todos los SNPs de genoma humano y caracterizar las diferencias de éstos en diferentes poblaciones (Guardado-Estrada et al., 2008). Además, los SNPs se han convertido en herramientas sumamente útiles para las empresas farmacéuticas, pues permiten determinar si un individuo responderá a cierto medicamento o experimentará efectos secundarios graves. Además, algunos SNPs suelen asociarse a ciertas enfermedades multigénicas a partir del conocimiento del significado funcional que tiene el SNP en algún proceso biológico. Sin embargo, aún no está claro como una variación genética puede determinar la susceptibilidad de un individuo a alguna enfermedad, pues la

mayoría de los rasgos fenotípicos comunes son el resultado de la interacción a largo plazo de los factores genéticos y no genéticos.

Polimorfismos de mini y microsatélites

Los minisatélites, también conocidos como repetidos en tandem de número variable o VNTRs (*del inglés: Variable Number of Tandem Repeats*) son muy frecuentes en el genoma humano. Se estima que existen alrededor de 3000 minisatélites en el genoma humano, y que cada locus contiene una unidad de repetición distintiva, de tamaño y secuencia diferente al resto de los minisatélites. Los minisatélites se consideran mitóticamente estables, pues solo mutan en raras ocasiones en los tejidos somáticos, sin embargo las tasas de mutación de los locis pueden variar significativamente en la línea germinal. En general, la frecuencia de mutación de la mayoría de los minisatélites se estima en el orden de 10^{-3} . Aproximadamente 300 minisatélites se han examinado en pedigreos humanos y sólo el 3% de estos se definen como ‘hipermutables’, lo que quiere decir que muestran en promedio una tasa de mutación mayor a 5×10^{-3} en la línea germinal. El minisatélite más hipervariable conocido es el CEB1, cuya frecuencia de mutación meiótica es de $1,3 \times 10^{-1}$ (Salipante, 2013). Además la frecuencia de mutación de los minisatélites puede variar de persona a persona, dependiendo de si está expuesta a radiación ionizante o agentes genotóxicos, pues se sabe que estos aumentan la inestabilidad de los minisatélites (Bois, 2003).

Los microsatélites o STRs se encuentran por todo el genoma humano, con mayor presencia en regiones subtelerómicas. Sin embargo la densidad de STRs varía ampliamente de un cromosoma al otro. En humanos, el cromosoma 19 es el que tiene la mayor densidad de STRs (Fan y Chu, 2007). En promedio, hay un STR por cada 2000 pb en el genoma humano. Los STRs tienen una tasa de mutación significativamente mayor que la observada en otras partes del genoma (10^{-9} contra 10^{-3} a 10^{-5}). Estas diferencias se han estudiado empleando diferentes técnicas, encontrándose que existen grandes diferencias en las tasas de mutación de un loci a otro (Fan y Chu, 2007).

La elevada tasa de mutación de STRs y VNTRs se puede explicar mediante varios mecanismos, dos de ellos se han considerado como los más importantes: el deslizamiento de la polimerasa durante la síntesis de DNA y el entrecruzamiento cromosómico desigual.

El deslizamiento de la polimerasa es el fenómeno más aceptado para la formación y evolución de STRs y VNTRs (Jobling et al., 2013). Este mecanismo se produce durante la replicación del DNA, como una consecuencia del desapareamiento de una o más unidades de repetición entre la hebra naciente y la hebra molde. El fragmento desapareado se ve forzado a generar una horquilla sobre el sitio de replicación. Si la síntesis de DNA continúa, entonces el número de repetidos cambia (véase figura 2.3). La horquilla formada es estable cuando se forma un arreglo de 3 a 4 unidades de repetición.

Las regiones del DNA que contienen repetidos en tándem son más propensas a la mutación, por lo que la estabilidad de los microsatélites es controlada *in vivo* mediante el mecanismo de reparación del DNA MMR, (*del inglés: DNA mismatch repair*). Si este sistema fuera defectuoso, las secuencias codificantes con repetidos en tándem serían objeto de mutación, como ocurre en tejidos tumorales. Por lo tanto, la inestabilidad de los repetidos en tándem desempeña un papel fundamental en la carcinogénesis. Se ha observado que un sistema MMR funcional reduce la frecuencia de mutación entre 100 y 1000 veces (Fan y Chu, 2007). Los genes mayores y menores del sistema MMR contienen extensiones cortas de $(A)_n$ dentro de su región codificante. Estas regiones son susceptibles de una delección espontánea o de la inserción de mutaciones, lo que podría resultar en la inactivación de estos genes, y por lo tanto la deficiencia de el sistema MMR (Chistiakov et al., 2006). Además de la efectividad del sistema MMR, otros factores pueden afectar la frecuencia, velocidad y dirección del desplazamiento. Por ejemplo, la tasa de deslizamiento es más alta en los dinucleótidos y más baja en los tetranucleótidos, pues las unidades de repetición más largas requieren que la hebra se deslice más antes de que las bases puedan emparejarse correctamente de nuevo. Otro factor que afecta el deslizamiento de la polimerasa es el número de repetidos; en humanos, la tasa de deslizamiento aumenta exponencialmente con el incremento de unidades de repetición (Fan y Chu, 2007).

En STRs la posible existencia de otros mecanismos de mutación sigue siendo un debate, pues teóricamente las mutaciones de STRs en la línea germinal humana podrían formarse por el entrecruzamiento desigual o la conversión génica (Debrauwere et al., 1997) (Lopes et al., 2002). Sin embargo, existen diversas evidencias de que los eventos de recombinación no tienen un papel cualitativamente importante en la mutación de STRs. La evidencia clave incluye el hecho de que la tasa y patrón de mutación de STRs no difiere entre cromosomas hemiciótopos (como el cromosoma Y) y cromosomas pares o diploides, lo que sugiere que los eventos de mutación no requieren contacto entre cromosomas homólogos. Otra pista importante proviene de los experimentos *in vitro*, pues se ha observado que los microsatélites son propensos a sufrir resbalamiento de la polimerasa. No ocurre lo mismo para VNTRs, pues el modelo del deslizamiento de la polimerasa no puede explicar algunos eventos de mutación de los minisatélites, pues los VNTRs más inestables no son necesariamente las secuencias más homogéneas (Debrauwere et al., 1997).

El entrecruzamiento desigual o *crossing over* fue propuesto por diversos autores para explicar la evolución de cierto tipo de VNTRs. Así, en ocasiones las moléculas de DNA se alinean de manera errónea, dando lugar a duplicaciones o delecciones, que conducen a un alargamiento o acortamiento de la secuencia repetitiva. En la figura 2.4 se muestra un esquema de este proceso.

Además del resbalamiento de la polimerasa y el entrecruzamiento desigual, se han descrito otros mecanismos que pueden generar polimorfismos en VNTRs, como la transferencia no recíproca de DNA de una molécula a su homóloga, o la hibridación de cadena sencilla en la que se inicia la mutación por una o varias

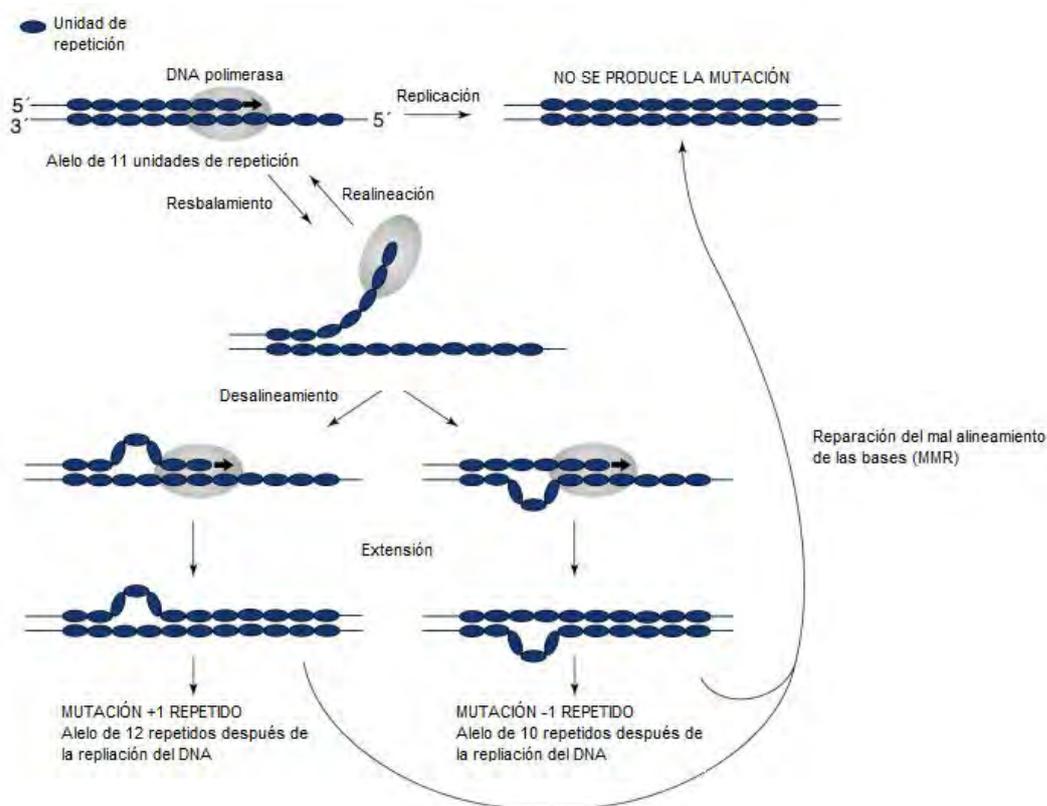


Figura 2.3: Modelo esquemático del deslizamiento de polimerasa. Modificado de (Jobling et al., 2013)

rupturas de la doble hebra o DSB (*del inglés: (Double-strand Break)*) (Fernández, 2005).

La mutación de VNTRs durante la meiosis se explica mejor mediante un modelo dependiente de las DSB. Los lugares del genoma donde ocurre este fenómeno son conocidos como *hotspots* o sitios calientes de recombinación. La estructura molecular de los *hotspots* ha sido ampliamente estudiada en diversos genomas, tanto eucariotas como procariontas, encontrándose similitudes grandes entre los sitios calientes de recombinación de roedores y humanos, e incluso, existe una similitud importante entre los *hotspots* de mamíferos y la secuencia Chi, conocido punto caliente de recombinación en eucariontes (Bois, 2003). La hipótesis de que los minisatélites tienen mecanismos de mutación diferentes en células somáticas y germinales se refuerza por el hecho de que la tasa de mutación de minisatélites humanos es diferente entre las células germinales de hombres y mujeres, lo que recuerda las diferencias observadas entre patrones específicos de sexo en la recombinación meiótica (Fernández, 2005). El desarrollo de nuevas técnicas para



Figura 2.4: Modelo esquemático del entrecruzamiento desigual (González-Blanco, 2005).

el estudio de la inestabilidad de microsatélites, tales como la PCR, ha aumentado nuestra comprensión de como ocurren estos procesos. Sin embargo, aún es necesario identificar varios compuestos intermediarios que ayuden a caracterizar de mejor manera las vías de inestabilidad de SSRs (Vergnaud y Denoeud, 2000).

Estudio de los microsatélites

Existen varias razones importantes para interesarse en los microsatélites. Primero, la abundancia y alto nivel de variación alélica de estas secuencias en varios organismos los ha convertido en marcadores genéticos populares. Además, la distancia genética entre los microsatélites se puede utilizar para responder a preguntas sobre la estructura de una población y su divergencia. Finalmente, la expansión de microsatélites con tripletes conduce a varios trastornos genéticos humanos, tales como el síndrome de X frágil y la distrofia miotónica. Estas características, aunadas a su fácil amplificación mediante PCR, convierten a los STRs en secuencias que son empleadas como herramientas en diversos campos de la ciencia. Por este motivo los STRs han sido objeto de estudios detallados que nos han hecho comprender mejor su naturaleza. Los aspectos de más relevancia para esta tesis se desglosan a continuación.

Clasificación y nomenclatura de los STRs

Para facilitar el estudio de los STRs estos se han clasificado de distintas formas. En este apartado se mencionan los criterios de clasificación más importantes, y la nomenclatura que se ha consensado para nuevos STRs.

Una de las clasificaciones de los STRs se basa en el orden y número de re-

peticiones de los STRs. Según esta clasificación, los STRs se pueden dividir en cinco rubros: puros, interrumpidos puros, compuestos, interrumpidos compuestos y complejos. Un STR es considerado puro cuando se componen de una sola unidad de repetición copiada cierto número de veces. Cuando un STR puro se interrumpe por un único motivo pequeño que divide el tándem de repetidos en dos partes, se le denomina interrumpido compuesto. Un STR compuesto es aquel que se forma por repeticiones de dos motivos diferentes. Si un STR compuesto es interrumpido por una sola secuencia pequeña que corta el tandem en dos partes, el STR resultante se denomina interrumpido compuesto. Finalmente, un STR complejo se forma de más de dos unidades repetitivas, que pueden estar interrumpidas por varias secuencias cortas de copia única (tabla 2.2) (Fernández, 2005).

Tipo de SSR	No. de UR diferentes	Ejemplo
Puro	1	-(TAGGG) ₁₂ -
Puro interrumpido	1	-(TAGGG) ₈ - (TGAGG) - (TAGGG) ₃ -
Compuesto	2	-(TAGGG) ₉ - (TGAAC) ₅ -
Compuesto interrumpido	2	-(TAGGG) ₉ - (TAAGG) - (TGAAC) ₄ -
Complejo	Mayor a 2	-(TAGGG) ₄ - (TGAAG) ₇ - (CAATG) - (TTGGG) ₃ -

Tabla 2.2: Clasificación de los SSRs según su organización (Fernández, 2005).

Otra clasificación común de los STRs es la que se hace según el número de pares de bases que componen la unidad de repetición. Como ya se sabe, un STR se forma por unidades de repetición de hasta 6 pb, por lo que los microsatélites se clasifican en: mononucleótidos, dinucleótidos, trinucleótidos, tetranucleótidos, pentanucleótidos y hexanucleótidos. La abundancia de cada tipo de STR ha sido ampliamente estudiada en el genoma, obteniendo varias observaciones importantes. En el caso de los mononucleótidos, la densidad de los repetidos poli(A) y poli(T) es mucho mayor que la de poli(G) o poli(C). De manera similar, los dinucleótidos AC y AT son más abundantes que GC. Siguiendo este patrón, los trímeros, tetrámeros, pentámeros y hexámeros más comunes son aquellos con mayor contenido de adenina: AAT, AAAT, AAAAT, AAAAAG. Esto puede explicarse tomando en cuenta que las bases A/T se separan con mayor facilidad de la cadena de DNA en comparación con las bases GC (Subramanian et al., 2003).

Además de las diferentes clasificaciones existentes para los STRs, estos tienen una nomenclatura específica, que depende de sus características. Por ejemplo, en el locus D3S1266 la letra D representa el DNA, el número 3 indica que este STR se localiza en el cromosoma 3, la letra S significa que es un STR y el número 1266 es el identificador único, número exclusivo de cada STR (Fan y Chu, 2007).

Evolución de los microsatélites

La pregunta de cómo se originaron los STRs en los genomas eucariontes, ha sido un tema controversial. Sin embargo, la mayoría de los autores conciden en el hecho de que el origen de un nuevo STR requiere la generación de un 'proto-STRó

‘proto-minisatélite’. Un proto-STR es una secuencia corta, la cual mutará generando un nuevo STR. La principal característica de un proto-STR es su capacidad de mutación. Existe un número mínimo de repeticiones necesarias para una mayor expansión; es de cuatro a cinco repeticiones para un STR dinucleótido y dos repeticiones para un tetranucleótido

Existen diversas hipótesis sobre el origen de los microsatélites en los genomas. Una de estas hipótesis sugiere que los microsatélites se originaron a partir de la degeneración de secuencias de copia única, cuyos repetidos se originaban mediante inserciones, sustituciones o ambas (Fernández, 2005). Un ejemplo de esta hipótesis se observa en especies que tienen al primate como ancestro común (es decir, gorilas, chimpances y humanos). Una sustitución de G por A en el locus de la η globina del primate cambió la secuencia ATGTGTGT a ATGTATGT, creandose así el microsatélite $(ATGT)_2$ que se convirtió en $(ATGT)_4$ en monos africanos, y finalmente en $(ATGT)_5$ en humanos (Oliveira et al., 2006). Otra hipótesis plantea que el origen de algunas secuencias repetidas puede deberse a algún elemento retrotasponible que contendría en su secuencia lugares proclives a la formación de este tipo de secuencias repetitivas (Fernández, 2005). Se especula que los microsatélites ricos en adenina pudieron generar una extensión del extremo terminal de un elemento trasponible, pues el extremo 3' de dichos elementos contienen un segmento muy similar al mecanismo de poliadenilación de mRNAs (Fan y Chu, 2007). Esta teoría se formuló después de encontrar relación entre los STRs más comunes (aquellos con alto contenido de A) y ciertos elementos trasponibles. La secuencia Alu, perteneciente a la familia de los SINEs, esta distribuida ampliamente en el genoma de los primates, y es posiblemente la que contribuyo al origen de microsatélites, debido a la presencia de regiones ricas en adenina en el extremo terminal 3' y dentro de dicha secuencia. La asociación entre microsatélites y elementos Alu puede ser explicada en tres mecanismos: 1) los elementos Alu se integran a un microsatélite preexistente, dando como resultado que el microsatélite preexistente flanqueó el elemento insertado, 2) elementos Alu con mutaciones en el extremo terminal 3' son integrados en el transcrito primario durante la transcripción inversa y la mutación actúa como un núcleo para la génesis de microsatélites y 3) La acumulación de mutaciones aleatorias en la cola de poli A de los elementos Alu, seguido por la expansión de esta región por deslizamiento o recombinación intra-alélica produce nuevos microsatélites (Oliveira et al., 2006). A pesar de que esta hipótesis puede aplicarse a un gran número de organismos, una gran cantidad de elementos trasponibles no siempre concide con una alta densidad de STRs, por lo que será necesario realizar estudios más detallados para dilucidar si es este realmente un mecanismo para la mutación de STRs (Fan y Chu, 2007).

Una vez que se ha producido el proto-STR, este sufrirá variaciones en su longitud, por pérdida y ganancia de unidades repetitivas. Las tasas de mutación de los STRs varía de manera considerable, no solo entre especies, sino también entre loci diferentes de una misma especie (Fernández, 2005). Se ha demostrado en diversos estudios que uno de los factores más determinantes en la tasa de mutación de un

microsatélite es la longitud de este. En la especie humana, el promedio de unidades de repetición que tenga un STR es directamente proporcional a su grado de polimorfismo, observándose que los loci más largos mutan más a menudo. Es posible que la correlación positiva entre el grado de mutación de un STR y el tamaño de este se pueda explicar por el hecho de que el resbalamiento de la polimerasa ocurre en más lugares en un STR con muchas unidades repetitivas. Se ha propuesto que la tasa de mutación de un STR puede variar debido a otros factores, diferentes de la longitud del loci. Uno de estos podría ser la secuencia que flanquea al microsatélite. Otro factor importante podría ser la diferencia entre los alelos que presente un individuo o el rango entre alelos. Aún no existe evidencia experimental sólida que respalde alguna de estas teorías (Ellegren, 2000).

Es poco común que los microsatélites alcancen longitudes mayores a unas cuantas decenas de unidades de repetición. Esta observación ha provocado una discusión sobre qué factores impiden el crecimiento infinito de un microsatélite. Así, en humanos se ha observado que mientras la tasa de expansión de un STR es lineal y constante, la tasa de contracción aumenta exponencialmente a medida que aumenta el número de unidades repetitivas del microsatélite (Xu et al., 2000). Mientras más crece un microsatélite, más probable es que pierda un buen número de unidades de repetición, lo que evita su crecimiento infinito. Otro mecanismo que evita el crecimiento infinito de un STR es la ocurrencia de mutaciones puntuales. Como cualquier tipo de secuencia, los microsatélites están expuestos a los eventos de mutación puntual. Si un STR puro (es decir, compuesto de un número entero de unidades de repetición) sufriera una mutación puntual, daría como resultado un STR interrumpido, lo que disminuiría su tasa de mutación, pues los microsatélites interrumpidos son menos variables que los puros. Se ha demostrado que una sola interrupción en un microsatélite resulta en la disminución de la tasa de mutación de dos a seis veces, en comparación con los microsatélites puros (Ananda et al., 2014). Esto se debe a que la mutación puntual genera una división de los repetidos perfectos, teniendo el más largo de estos una mayor tasa de mutación, pues el polimorfismo de una región repetitiva depende de la cantidad de unidades repetitivas continuas, no de la longitud total de la región repetitiva. Hay indicios de que estas mutaciones puntuales tienen cierta polaridad, efecto que también se ha observado en minisatélites, aunque las causas de esta polaridad se desconocen (Ellegren, 2000).

En algunos procesos patológicos, los microsatélites tienden inevitablemente a la expansión, a pesar de los mecanismos que regulan el crecimiento descontrolado de los STRs. Por ejemplo, algunos STRs con trinucleótidos implicados en las enfermedades de expansión de tripletes (como CCG, CGC, AGG, CCT y otros) tienen una alta probabilidad de formar horquillas estables durante el resbalamiento de la polimerasa, lo que hace que se expandan sin importar la longitud y naturaleza del alelo (Chistiakov et al., 2006).

De acuerdo a lo dicho anteriormente, la evolución de los microsatélites es un proceso complejo, donde algunos alelos son expandidos o reducidos de manera

arbitraria por el resbalamiento de la polimerasa, hasta el punto en que una mutación corta la matriz de repeticiones perfectas en dos tramos más cortos. El conjunto de todos estos procesos degenerarían la secuencia original, lo que podría resultar en la formación de microsatélites nuevos (Ellegren, 2004).

Importancia funcional de los microsatélites

Algunos microsatélites tienen un efecto neutro en el genoma, mientras que algunos otros tienen funciones importantes. Evidencia reciente sugiere que los STRs pueden modular una gran variedad de procesos biológicos, tanto de organismos maduros como en organismos en desarrollo (Hannan, 2010). El hecho de que algunos STRs se encuentren altamente conservados entre diferentes especies sugiere que estos microsatélites podrían cumplir funciones importantes. Tal es el caso de los STRs que están implicados en la organización de la estructura cromosómica. Por ejemplo, la secuencia hexamérica (*TTAGGG*)_n es reconocida por la ribonucleoproteína de la polimerasa, que es una telomerasa encargada de sintetizar repeticiones teloméricas en la parte terminal de los cromosomas para superar la pérdida de secuencias de DNA durante la replicación, lo que le confiere estabilidad a los cromosomas (Chistiakov et al., 2006). Algunos reportes indican que los microsatélites también se asocian con la regulación y/o función de genes. Se ha demostrado que los microsatélites ricos en C y T adyacentes al punto de inicio de la transcripción funcionan como secuencias diana para la unión de los factores de transcripción y podrían ser importantes en la frecuencia con que esta inicia. En muchos casos, el número de repeticiones podría influir significativamente en el nivel de la expresión génica. Los microsatélites ubicados en las regiones no traducidas del extremo 5' podrían formar estructuras específicas e inusuales de DNA. En este caso, la longitud de la región de repetición podría afectar el nivel de traducción del mRNA diana (Sharma et al., 2007). Los microsatélites en la región no traducida 3' pueden afectar la expresión génica a través de la influencia de la estabilidad de los productos transcritos. Se ha demostrado que aún los STRs en regiones intrónicas pueden afectar la transcripción génica (Chistiakov et al., 2006). Los STRs también pueden regular la función génica mediante epigenética. Recientemente se ha observado la metilación del DNA mediante regulación epigenética asociada al trinucleótido CCG (Rama-moorthy et al., 2014). Además los STRs pueden influir en la replicación del DNA, formando bucles que sirven como señal de parada para la DNA polimerasa. Los genes humanos que codifican factores de fidelidad y crecimiento importantes (como BCL2, IGF2R, BRCA2 O TGFβ-2) contienen STRs, cuya mutación (ya sea inserciones o deleciones) afectan a estos genes y como consecuencia se inicia el proceso de tumorigénesis (Chistiakov et al., 2006). La evidencia anterior sugiere que la variación de los microsatélites puede ocasionar cambios importantes en la expresión génica. Debido a la gran abundancia y alta mutabilidad de los STRs, los cambios en estos son importantes en la adaptación y supervivencia de una población (Oliveira et al., 2006).

Los STRs también están implicados en muchas enfermedades monogénicas

que afectan principalmente al sistema nervioso, incluyendo la enfermedad de Huntington, varias ataxias espinocerebelosas, enfermedades poliglutámicas, ataxia de Friedreich, síndrome de X frágil, la epilepsia mioclónica, trastornos polialanina y distrofia miotónica. La naturaleza exacta de cada trastorno depende de la ubicación de la repetición en tándem, así como el patrón de expresión espacio-temporal y la función de cada gen de la enfermedad (Siwach y Ganesh, 2008).

Aplicaciones de los microsatélites

Después de que en 1986 Kary Mullis desarrollara la PCR, la genética empezó a utilizar esta técnica para el estudio de la variabilidad genética a nivel molecular. Así comenzó la utilización de los microsatélites como marcadores útiles para toda clase de estudios genéticos, pues sus características los convierten en marcadores valiosos (Aranguren-Méndez et al., 2005). El mapeo genético representa uno de los principales campos de investigación en los que se han aplicado los STRs como marcadores. El mapeo genético consiste en el uso de técnicas genéticas para construir mapas que muestran las posiciones de los genes. Para desarrollar un mapa genético es necesario el uso de marcadores. Un marcador genético útil tiene al menos dos alelos segregados de manera independiente, además de ser altamente polimórfico y fácil de identificar. Los STRs cumplen con todas estas características, por lo que se han convertido en marcadores de elección para la construcción de mapas genéticos (Brown, 2002).

Los microsatélites son una herramienta muy útil en el campo de la criminalística, empleándose para la identificación forense. Además, los STRs se utilizan en la determinación de paternidad en humanos, plantas y animales (delaRosa et al., 2004) (Chistiakov et al., 2006). A pesar de que es posible identificar individuos usando minisatélites y MLPs, el uso de minisatélites no resulta útil en la identificación forense, ya que para obtener una huella genética confiable son necesarios varios μg de DNA genómico de buena calidad, y comúnmente las muestras forenses contienen DNA escaso y/o degradado. El análisis de microsatélites es la técnica más sensible que existe en la identificación forense, pues es posible realizar la identificación incluso si se tiene una sola célula, lo que hace que en la práctica forense se prefiera el uso de STRs en lugar de VNTRs. (Tamakia y Jeffreys, 2005). Los mejores marcadores para estimar la paternidad o parentesco de dos individuos son sin duda los microsatélites, pues son codominantes, abundantes en varios genomas y lo suficientemente hipervariables para distinguir entre individuos relacionados. El hecho de que se usen STRs en vez de VNTRs en los estudios de paternidad se debe principalmente a que estos son amplificables por PCR a diferencia de los VNTRs (Oliveira et al., 2006).

Una parte importante de las pruebas de paternidad es la tasa de mutación de los STRs empleados, que deberá tomarse en cuenta para saber que tan posible es que este mute al transmitirse (Asicioglu et al., 2004). Es por este motivo que los STRs han demostrado ser más eficientes que los VNTRs en pruebas de paternidad, pues muestran un polimorfismo considerable, pero son mucho menos variables que

la mayoría de los minisatélites. (Tamakia y Jeffreys, 2005).

Los microsatélites también se emplean para evaluar la estructura genética de las poblaciones. Durante los últimos años se han realizado muchos estudios que han utilizado los microsatélites para análisis filogenéticos, concluyendo que, con un buen número de loci analizados y con apropiada tasas de mutación, los microsatélites pueden dar una buena aproximación de la filogenia (Aranguren-Méndez et al., 2005). Un factor importante a considerar durante un estudio de filogenia es el número de microsatélites a estudiar, pues aumentar moderadamente la cantidad de loci empleados potenciará la estimación de las distancias génicas y la exactitud de los filogramas (Koskinen et al., 2004). Sin embargo, es importante tomar en cuenta que los STRs poseen ciertos atributos que limitan su utilidad en esta área. Por ejemplo, el rango de tamaño de los alelos puede ser limitado y con una alta tasa de mutación. Como consecuencia de la combinación de estas propiedades el potencial de divergencia génica puede ser limitado, especialmente en grandes poblaciones (Richard y Thorpe, 2001). En los estudios de genética de poblaciones, los STRs permiten la identificación de cada alelo por locus, la obtención de datos poblacionales y el cálculo de las frecuencias alélicas. De esta manera podemos estimar las distancias genéticas entre poblaciones o entre individuos logrando determinar la estructura de la población. Las aplicaciones de los STRs en la genética de poblaciones se desglosarán con más detalle en la siguiente sección.

Genética de Poblaciones y STRs

La genética de poblaciones puede ser definida simplemente como la parte de la genética que se ocupa de la descripción de las características heredables observadas o inferidas de las poblaciones a través del espacio y el tiempo (Amorim, 2013). En otras palabras, la genética de poblaciones trata de explicar el proceso de adaptación en términos cuantitativos.

El proceso de adaptación es uno de los componentes fundamentales de la evolución. La adaptación y la especiación son los principales fenómenos que han cambiado la herencia de las poblaciones a través del tiempo. La especiación es el proceso mediante el cual se originan nuevas entidades evolutivas y el causante de la gran biodiversidad, mientras que la adaptación es el mecanismo que habilita a las poblaciones para resolver problemas que les presenta el ambiente, permitiéndoles adquirir características favorecedoras a la sobrevivencia y fertilidad con respecto a un medio dado (Eguiarte, 2009). En 1859 C. Darwin, en su libro *El origen de las especies*, proponía la teoría de que la evolución orgánica se da gracias a la selección natural, definiéndose a la selección natural como la sobrevivencia y reproducción diferencial de organismos diversos. Cincuenta años después de que Darwin publicara su teoría, R. Fisher, J. Haldane y S. Wright conjugaron la genética y la teoría de Darwin, en un nuevo enfoque conocido como genética de poblaciones, mostrando cómo podía actuar la selección natural en un mundo mendeliano

La genética de poblaciones presenta dos aspectos fundamentales: el aspecto

predictivo que se encarga de predecir la composición futura de cierta población, y el aspecto *retrospectivo*, que se encarga de entender los factores que determinaron la composición actual de la población en estudio. El aspecto predictivo de la genética de poblaciones se desarrolló en el siglo XX, proponiendo cambios en las frecuencias de ciertos alelos y genotipos bajo ciertas condiciones, mientras que el enfoque retrospectivo, es más reciente, y se enfoca en la ancestría de los genes (Nielsen y Slatkin, 2013).

Dado que la genética de poblaciones es una generalización de la teoría mendeliana, a continuación se dará una breve recapitulación de los conceptos más importantes de la genética mendeliana. Los genes suelen presentarse en versiones diferentes denominadas alelos. Un locus (o loci en plural) es la posición del genoma en la que puede haber uno o más alelos segregantes. El genotipo es el conjunto de alelos que posee un organismo individual. Un organismo diploide que posee dos alelos idénticos es homocigoto para ese locus, mientras que uno que posee dos alelos diferentes es heterocigoto para ese locus (Pierce, 2009).

Frecuencias génicas y genotípicas

Una frecuencia es simplemente una proporción o porcentaje, expresado de manera habitual como una fracción decimal.

Por lo tanto, una frecuencia génica o frecuencia alélica no es más que el número de copias de cierto alelo dividido por el número total de copias génicas en la población. En una población diploide, los individuos contienen dos copias de un mismo gen, una en cada cromosoma, lo que implica que si en cierta población hay N individuos, hay entonces $2N$ copias génicas (Nielsen y Slatkin, 2013). Para simplificar el análisis de las frecuencias genéticas, comúnmente se utilizan modelos donde solamente existen dos alelos de un mismo gen, por ejemplo, el alelo A y a . Estos modelos son llamados modelos dialélicos. En este tipo de modelos la frecuencia de los alelos A y a suele representarse con los símbolos p y q , respectivamente.

Entonces, para calcular la frecuencia de los genes A y a , se emplean las fórmulas:

$$p = f_A = \frac{N_A}{2N} \quad (2.1)$$

y

$$q = f_a = \frac{N_a}{2N} \quad (2.2)$$

Donde N_A y N_a son el número de alelos de A y a , respectivamente. Dado que se está manejando un modelo dialélico, la suma de las frecuencias de A y a deben dar como resultado el total de la población, es decir $f_A + f_a = 1$ o $p + q = 1$.

Aunque los locus dialélicos son los más simples de estudiar, en ocasiones es necesario calcular la frecuencia de locus con varios alelos diferentes. Un locus que

tiene k alelos diferentes, donde $k \geq 1$ es cualquier número natural, es comúnmente conocido como un locus k -alélico. Es posible encontrar expresiones para calcular la frecuencia alélica de locus k -alélicos; pues dichas expresiones son muy similares a las utilizadas para locis dialélicos. Para cada alelo $i \in 1, 2, \dots, k$ con N_i copias en la población, la frecuencia génica es:

$$f_i = \frac{N_i}{2N} \quad (2.3)$$

La frecuencia genotípica es la proporción de individuos que poseen cierto genotipo en una población. Para calcular una frecuencia genotípica sumamos la cantidad de individuos que poseen cierto genotipo y dividimos este número por el número total de individuos en la muestra o N . Para un locus dialélico existen tres posibles genotipos. Suponiendo que representamos al gen dominante con la letra A y al recesivo con a , los genotipos probables serían: AA , Aa y aa . La frecuencia de cada genotipo se calcularía con las formulas:

$$\begin{aligned} f_{AA} &= \frac{n_{AA}}{N} \\ f_{Aa} &= \frac{n_{Aa}}{2N} \\ f_{aa} &= \frac{n_{aa}}{N} \end{aligned} \quad (2.4)$$

Notese que la suma de la frecuencia de todos los genotipos debe ser 1, es decir $f_{AA} + f_{Aa} + f_{aa} = 1$ o $p^2 + 2pq + q^2 = 1$.

Las frecuencias alélicas también pueden calcularse a partir de las frecuencias genotípicas. Un individuo diploide tiene dos copias de un mismo gen, pudiendo ser estas diferentes o iguales, así pues, la frecuencia del genotipo homocigoto más la mitad de la frecuencia del heterocigoto darán como resultado la frecuencia alélica. Con base en este razonamiento, la frecuencia del alelo A y a podría calcularse de la siguiente manera:

$$\begin{aligned} p = f_A &= \frac{2n_{AA} + n_{Aa}}{2N} = f_{AA} + \frac{f_{Aa}}{2} \\ q = f_a &= \frac{2n_{aa} + n_{Aa}}{2N} = f_{aa} + \frac{f_{Aa}}{2} \end{aligned} \quad (2.5)$$

Donde n_{AA} representa el número de individuos con el genotipo AA , n_{Aa} el número de individuos con el genotipo Aa y n_{aa} el número de individuos con el genotipo aa .

En locis con alelos múltiples o k -alélicos también es posible calcular las frecuencias génicas a partir de las frecuencias genotípicas. El procedimiento es similar al empleado para locis dialélicos: se suma la frecuencia del genotipo homocigoto, más la mitad de las frecuencias heterocigotas. La expresión matemática se muestra a continuación:

$$f_i = f_{ii} + \sum_{j:j \neq i}^{\infty} f_{ij}/2 \quad (2.6)$$

Equilibrio de Hardy-Weinberg

En la sección anterior se explicó como calcular las frecuencias alélicas a partir de las frecuencias genotípicas, sin embargo no se mostró el proceso inverso. ¿Es posible calcular las frecuencias genotípicas a partir de las frecuencias alélicas? La respuesta a esta pregunta es afirmativa, siempre y cuando se asuman ciertas condiciones.

El cálculo de frecuencias genotípicas a partir de las frecuencias génicas se hizo por primera vez en 1908, cuando el matemático inglés Godfrey H. Hardy y el físico alemán Whihelm Weinberg descubrieran independientemente la relación entre las frecuencias alélicas y genotípicas (Chen, 2010). El equilibrio de Hardy-Weinberg o HWE (*del inglés: Hardy-Weinberg equilibrium*) afirma que bajo ciertas condiciones, las frecuencias genotípicas de un locus son determinadas por las frecuencias alélicas, y que ambas frecuencias (genotípicas y alélicas) permanecerán constantes en el futuro (Wigginton et al., 2005). Las condiciones que son necesarias para mantener el HWE son:

1. Que la población sea grande
2. Que los apareamientos se realicen al azar y no exista la endogamia
3. Que la población no esté sujeta a selección natural, mutacion o migración
4. Que las poblaciones no estén solapadas

Esto quiere decir que, si una población se encuentra bajo el HWE, sea cual sea la frecuencia genotípica inicial, la generación siguiente presentará las frecuencias genotípicas que se observan en la tabla 2.3.

Genotipo	AA	Aa	aa
	pp	pq	qq
Frecuencia	f_A^2	$2f_A f_a$	f_a^2
	p^2	pq	q^2

Tabla 2.3: Frecuencias genotípicas bajo el equilibrio de Hardy-Weinberg (Nielsen y Slatkin, 2013).

La versión del HWE que se presenta en la tabla 2.3 es la más sencilla posible: aquella que involucra únicamente un locus con dos alelos. Sin embargo, el HWE también puede usarse para un loci multialélico. Imaginemos un locus k -alélico con las frecuencias alélicas f_1, f_2, \dots, f_k . Después de una generación de apareamiento

aleatorio, las frecuencias genotípicas pueden ser obtenidas mediante la multiplicación de los alelos adecuados. Es decir, la frecuencia genotípica esperada de los individuos homocigotos con el genotipo ii es f_i^2 para cualquier alelo i , mientras que la frecuencia genotípica para los individuos heterocigotos con el genotipo ij es $2f_i f_j$, para cualquier par de alelos i y j .

En términos generales es difícil pensar que las cuatro condiciones del HWE puedan cumplirse siempre en todas las poblaciones. Basta con que una no se cumpla para que las frecuencias alélicas cambien, y en consecuencia también lo hagan las frecuencias genotípicas. Así pues, analicemos las diferentes violaciones a las condiciones del HWE. La primera condición es el tamaño de la población. En teoría, el HWE requiere de una población que tenga un tamaño infinitamente grande, aunque en la práctica, no existe ninguna población que tenga un tamaño infinito (Amorim, 2013). A pesar de esto, muchas poblaciones cumplen las proporciones esperadas por el HWE, y las desviaciones importantes sólo surgen cuando el tamaño de la población es más bien pequeño (Pierce, 2009). La violación a la condición de que la población tenga un tamaño grande se llama **deriva génica**, y va ser más importante entre más pequeña sea la población (Eguiarte, 2009).

Una segunda presunción del HWE es que los individuos no están expuestos al **apareamiento selectivo**. El apareamiento ocurre cuando los individuos son más propensos a aparearse con individuos del mismo genotipo o de uno similar (Nielsen y Slatkin, 2013). Si el proceso de apareamiento selectivo continua durante varias generaciones, esta se verá empobrecida de individuos heterocigotos. La **endogamia** tiene efectos similares a los del apareamiento selectivo, con la diferencia de que la endogamia afecta al genoma entero, mientras que el apareamiento selectivo solo afecta aquellos loci que determinan el rasgo o rasgos que afectan la preferencia del apareamiento.

El tercer supuesto del HWE es que las frecuencias alélicas de la población no son afectadas por la selección natural, la migración o la mutación. Si bien en toda población aparecen mutaciones, su tasa es tan baja que tiene poco efecto en las predicciones del HWE, y por lo general tienen efecto sólo tras largos periodos de tiempo, o cuando no hay otras fuerzas en acción. Por otra parte la selección natural y la migración son factores significativos que hay que considerar en las poblaciones reales (Pierce, 2009). La **selección natural** ocurre cuando existe una diferencia de supervivencia o reproducción entre individuos debido a su genotipo, mientras que la **migración** consiste en el movimiento de individuos (y por tanto sus alelos) entre poblaciones. Tanto la selección natural como la migración causan el cambio de las frecuencias alélicas y genotípicas de manera importante, motivo por el cual existen diversas maneras de calcular sus efectos sobre una población.

Por último, el supuesto de que no existan poblaciones solapadas se refiere a la estructura de una población. En ocasiones las muestras recolectadas al azar de una población no pertenecen a una, sino a varias subpoblaciones. Al ignorar la estructura de la población, se pasa por alto el principio del apareamiento al azar, pues los individuos de una población no siempre pueden aparearse con los de otra

(Chen, 2010).

A pesar de que el HWE se ha empleado durante más de un siglo como herramienta para entender mejor las características de una población, en ocasiones no es muy sensible a ciertos tipos de desviaciones a las condiciones explicadas anteriormente, e incluso en ocasiones los efectos de las desviaciones de varios supuestos pueden anularse entre sí (Wittke-Thompson et al., 2005). Por lo tanto la violación a estas presunciones no necesariamente da lugar a una desviación observable en el HWE. Sin embargo, una desviación en el HWE, sugiere fuertemente que al menos uno de los supuestos se viola. Es debido a este motivo, y a su elegancia e importancia teórica que el HWE se ha convertido en un punto de partida importante para las distintas investigaciones genéticas (Wigginton et al., 2005).

Capítulo 3

Planteamiento del problema

La formulación de un problema es más importante que su solución.

Albert Einstein

Justificación

Determinar la frecuencia génica de los STRs en población mexicana, ayuda a la mejora de la identificación de individuos, pues las frecuencias alélicas y las tasas de mutación de ciertos STRs cambia de población en población. Establecer la frecuencia y número de alelos distintos de estos STRs en individuos de diferentes estados del centro y sur de México, permite convenir los STRs más informativos para fines criminalísticos.

Además, el conocimiento de estas frecuencias también puede ser aprovechado en el campo de la genética de poblaciones, ayudando a definir la estructura de las poblaciones estudiadas y si estas se desvían del HWE. En caso de que una población se encuentre fuera del HWE, podría investigarse más a fondo los motivos a los que obedece este desequilibrio.

Actualmente numerosos artículos han descrito las frecuencias alélicas de los 15 STRs que detecta el kit AmpF/STR®Identifiler®Plus en población del centro y sur de la República Mexicana. Sin embargo, son pocos los investigadores que consideran el origen de los padres y abuelos de un sujeto para considerarlo parte de una población o no. Esta diferencia podría hacer que los resultados previamente reportados sean distintos de los obtenidos en este trabajo, no sólo en efectos de frecuencias alélicas, sino también en la distancia genética entre poblaciones.

Hipótesis

Si los sujetos de las poblaciones evaluadas no son endogámicos, ni están expuestos a migración, mutación o selección natural, entonces las frecuencias alélicas

de los 15 STRs analizados deben cumplir con las frecuencias genotípicas del HWE.

Objetivos del trabajo

Objetivo General

Determinar la frecuencia de los alelos D8S1179, D21S11, D7S820, CSF1PO, D19S433, vWA, TPOX, D18S51, D3S1358, TH01, D13S317, D16S539, D2S1338, D5S818, y FGA en una población del centro y sur de México, mediante electroforesis capilar utilizando el kit AmpF ℓ STR®Identifiler®Plus PCR identificando los alelos más comunes en cada estado.

Objetivos Particulares

- Comparar las frecuencias de los 15 alelos analizados en los 9 estados de la República Mexicana muestreados para determinar si existen diferencias significativas entre un estado y otro, lo que permitirá identificar si existen alelos característicos de algún estado en particular.
- Contrastar las frecuencias alélicas obtenidas en este estudio con las reportadas anteriormente para la población del centro y sur de México, analizando las discrepancias entre los datos obtenidos y los reportados, si las hubiera.

Capítulo 4

Materiales y Métodos

Hay tres principales medios de adquirir conocimiento: la observación de la naturaleza, la reflexión y la experimentación. La observación recoge datos, la reflexión los combina; la experimentación verifica el resultado de esa combinación

Denis Diderot

Materiales

Cebadores para la PCR

Los cebadores o *primers* empleados para la amplificación de los 15 STRs y el marcador de género Amelogenina provienen del kit AmpF ℓ STR $\text{\textcircled{R}}$ Identifiler $\text{\textcircled{R}}$ Plus PCR, fabricado por Applied Biosystems. Se emplearon enlazadores no nucleotídicos para la síntesis de los primers para los loci: CSF1PO, D13S317, D16S539, D2S1338, y TPOX para permitir el posicionamiento reproducible de los alelos, facilitando la separación interlocus. Además, el kit utiliza 5 colorantes fluorescentes o *5-dye set*, que permiten la amplificación simultánea y eficiente separación de los 15 loci de STR y Amelogenina durante el análisis automático de los fragmentos de DNA (Biosystems, 2015). La tabla 4.1 muestra los locis amplificados, su localización cromosómica y el colorante fluorescente con el que está marcado cada primer.

Kits comerciales

- DNeasy $\text{\textcircled{R}}$ Blood&Tissue (Qiagen), para la extracción de DNA
- AmpF ℓ STR $\text{\textcircled{R}}$ Identifiler $\text{\textcircled{R}}$ Plus PCR (Applied Biosystems $\text{\textcircled{R}}$)

Reactivos

- Etanol anhidro desnaturalizado de pureza grado biología molecular
- Alcohol isopropílico de pureza grado biología molecular
- Agua ultra pura MiliQ®
- AmpliTaq Gold™DNA Polymerase
- Hi-Di™Formamida

Equipos

- Centrífuga
- Microcentrífuga
- Centrífuga para microplacas
- Vórtex
- Espectrofotómetro NanoDrop 1000 V3.7
- Termociclador GeneAmp®PCR System de Applied Biosystems
- Sistema de electroforesis capilar ABI PRISM®3100/3100-Avant Genetic Analyzer de Applied Biosystems
- Micropipetas Rainin Classic™ de 0.5 a 10 μ L, 2 a 20 μ L, y 100 a 1000 μ L de Mettler Toledo
- Puntas para pipeta de 10, 20 y 100 μ L
- Tubos eppendorf de 1.5 mL y 0.6 mL

Metodos

Población en estudio

La población analizada tiene un total de 656 individuos no emparentados entre sí provenientes de los estados de Veracruz, Oaxaca, Estado de México, Chiapas, Puebla, Guerrero, Hidalgo, Tabasco y Ciudad de México. Todos los individuos incluidos en el estudio tienen padres y abuelos originarios del mismo estado. En la tabla 4.2 se muestra el número de individuos estudiados por estado.

Estado	Número de individuos
Veracruz	135
Oaxaca	134
Estado de México	91
Chiapas	79
Puebla	78
Guerrero	49
Hidalgo	48
Tabasco	31
Ciudad de México	20
Total	656

Tabla 4.2: Número de individuos estudiados por estado

Consideraciones éticas

Las muestras usadas en este trabajo fueron obtenidas de personas que dieron su autorización mediante una carta de consentimiento informado. Todos los datos personales de los sujetos se mantendrán confidenciales.

Toma de muestra

El material biológico utilizado fueron células de la mucosa oral, obtenidas por raspado del carrillo bucal con un cepillo citológico estéril. Una vez obtenidas las células, éstas se conservaron en 2 mL de buffer de fosfatos salino o PBS (*por sus siglas del inglés: Phosphate Buffered Saline*)

Extracción de DNA

La extracción de DNA se llevo a cabo utilizando el kit DNeasy ®Blood&Tissue de Qiagen. El protocolo utilizado se llama *Purification of Total DNA from Animal Tissues (Spin-Column Protocol)* y viene descrito en el manual FGA DNeasy ®Blood&Tissue Handbook. (Qiagen, 2006). Los pasos de este procedimiento son:

1. Centrifugar la muestra a 13,200 rpm durante 5 minutos.
2. Descartar 700 μ L de sobrenadante y agregar 120 mAU de Proteinasa K.
3. Agitar en el vórtex por 15 segundos. Añadir 200 μ L de buffer AL a la muestra.
4. Agitar en el vórtex de 15 a 20 segundos y posteriormente incubar a una temperatura de 70 °C durante 30 minutos.

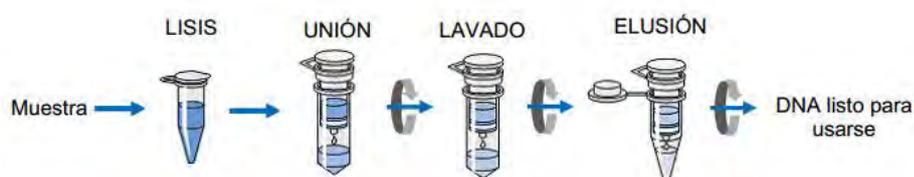


Figura 4.1: Pasos a seguir en el protocolo *Purification of Total DNA from Animal Tissues (Spin-Column Protocol)*.

5. Agregar 200 μL de etanol anhidro y mezclar inmediatamente en el vórtex por 15 segundos.
6. Tomar 700 μL de la mezcla realizada (incluyendo cualquier precipitado) y colocarlo en la columna *DNeasy Mini spin colum* colocada dentro de un tubo colector de 2 mL. Centrifugar a 8,000 rpm durante 2 minutos.
7. Descartar el sobrenadante y colocar la columna en un tubo eppendorf de 1.5 mL limpio. Adicionar a la columna 500 μL de solución AW1. Centrifugar a 8,000 rpm durante 2 minutos.
8. Colocar la columna en un tubo eppendorf de 1.5 mL nuevo y adicionar 500 μL de solución AW2. Centrifugar a 13,200 rpm durante 3 minutos.
9. Poner la columna en un tubo colector limpio y agregar 100 μL de buffer AE.
10. Dejar incubando 10 minutos a temperatura ambiente. Centrifugar a 8,000 rpm durante 2 minutos.
11. Añadir a la columna otros 100 μL de buffer AE. Dejar reposando 10 minutos a temperatura ambiente.
12. Finalmente centrifugar a 8,000 rpm. Desechar la columna. El DNA extraído se encuentra en el tubo colector.

El procedimiento mencionado anteriormente se resume en la imagen 4.1

Cuantificación de DNA

Una vez obtenido el DNA, se cuantificó y midió su pureza mediante el espectrofotómetro NanoDrop 1000 V3.7. El procedimiento fue el siguiente:

1. Prender el equipo, es decir, espectrofotómetro y la computadora asociada a este.
2. Realizar un lavado, colocando en el espectro 3 μL de agua ultra pura MilliQ®.

3. Calibrar el espectro con 3 μL de buffer de elusión AE. Esta solución se utiliza como blanco, por lo que su absorbancia debe ser ajustada a 0 $\text{ng}/\mu\text{L}$.
4. Verificar que la longitud de onda del espectro sea de 260 nm.
5. Realizar la lectura de cada muestra, colocando en el NanoDrop 3 μL de cada una.

Reacción de PCR

Preparación de la mezcla de reacción

Para preparar la mezcla de reacción es necesario calcular el volumen de cada componente necesario para preparar las reacciones, usando como referencia la tabla 4.3

Reactivo	Volumen por reacción (μL)
AmpF ℓ STR®Identifiler®Plus Master Mix	10
AmpF ℓ STR®Identifiler®Plus Primer Set	5
AmpliTaq Gold ®DNA polymerase	0.5

Tabla 4.3: Reactivos empleados en la preparación de la mezcla de reacción

Nota: Incluir una o dos reacciones adicionales al momento de preparar la mezcla de reacción, para prevenir la pérdida de volumen que ocurre durante el pipeteo de los reactivos.

Una vez que se pipetean los volúmenes requeridos de los dos componentes de la mezcla de reacción en un tubo eppendorf de tamaño adecuado, se prosigue a mezclar en el vórtex durante 3 segundos. Finalmente, centrifugar brevemente el tubo en una microcentrifuga para remover cualquier líquido de la tapa.

Preparación de las muestras

La preparación de las muestras se lleva a cabo según lo indicado en la tabla 4.4

Tipo de muestra	Preparación
Control Negativo	10 μL de agua MiliQ®
Muestra	10 μL de DNA diluido, con una concentración de 0.1 $\text{ng}/\mu\text{L}$
Control positivo	10 μL de DNA control 9947A (0.1 $\text{ng}/\mu\text{L}$)

Tabla 4.4: Indicaciones para la preparación de muestras

Nota: Es importante que se utilice agua MiliQ® para realizar la dilución de las muestras.

Una vez preparada, se debe agregar a cada muestra 15 μL de la mezcla de reacción. El volumen final de la reacción será de 25 μL . Antes de realizar la PCR se debe agitar el tubo con la muestra y mezcla de reacción durante 3 segundos y después centrifugar a 3000 rpm durante 20 segundos.

Programar el termociclador GeneAmp®PCR System de Applied Biosystems con las condiciones de la tabla 4.5

Incubación inicial	Ciclos (28)			Extensión final	Contención final
	Mantener	Desnaturalización	Alineamiento		
95°C 11 min	94°C 1 min	59°C 1 min	72°C 1 min	60°C 60 min	4°C ∞

Tabla 4.5: Condiciones de la PCR

Electroforesis capilar

Preparación de las muestras para la electroforesis

Las muestras para la electroforesis se preparan directamente en los pocillos de la placa 3100/3100-Avant. La preparación de las muestras tiene los siguientes pasos:

1. Agregar en cada pocillo 17 μL de Hi-Di™ Formamida, 0.6 μL de GenScan 500 LIZ Size Standar y 2 μL del producto de la PCR.
Nota: Es importante que, una vez agregados los tres reactivos en el pocillo, estos se mezclen varias veces con la micropipeta.
2. Preparar escalera alélica (por cada 15 muestras), agregando en un pocillo 18 μL de Hi-Di™ Formamida y 2 μL del reactivo que contiene la escalera alélica o *allelic ladder*.
3. Sellar la placa y centrifugar brevemente.
4. Calentar la placa en un termociclador durante 3 minutos a 95°C.
5. Inmediatamente después, colocar la placa en hielo durante 3 minutos.

La placa con las muestras preparadas se coloca en la bandeja de muestras en el inyector automático del equipo 3130 Genetic Analyzer de Applied Biosystems. El polímero empleado en el equipo es POP-4™ y los arreglos capilares empleados miden 36 cm.

Los resultados obtenidos por el equipo se analizan en la computadora acoplada a este, con el software GeneMapper®ID V3.2. El software se encarga de la identificación, separación por color y cuantificación de los picos. Además compara los picos obtenidos con los de la escalera alélica para así poder generar los genotipos. Dando como resultado un gráfico conocido como **electroferograma** (Figura 4.2).

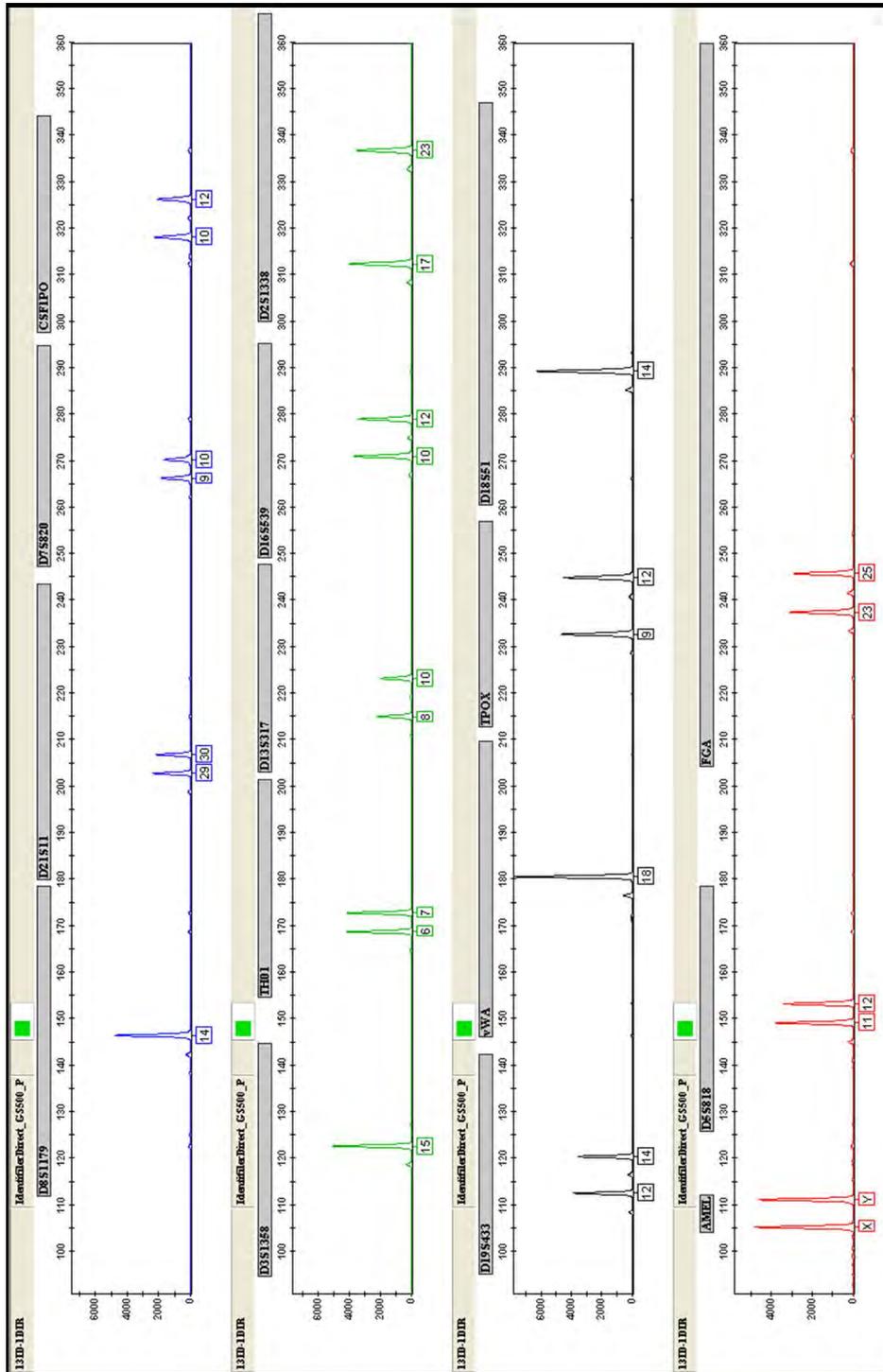


Figura 4.2: Electroferograma obtenido por el software GeneMapper®ID-X. Cada STR es separado por color, según el fluoróforo con el que fue marcado en la reacción de PCR (tabla 4.1). En el eje de las y se grafica la fluorescencia relativa, mientras que en el eje de las x se grafica el tamaño del producto de la PCR en pares de bases. Nótese que en el caso de los genotipos homocigotos, la fluorescencia relativa aumenta significativamente en comparación a la señal de un alelo heterocigoto.

Análisis genético de las muestras

Una vez obtenidos los genotipos de todos los individuos estudiados se obtuvieron los parámetros mencionados a continuación, empleando el programa Arlequin Ver 3.5.1.

- Frecuencias alélicas: Son un indicador de diversidad genética, pues reflejan la cantidad de alelos existentes en una población (Eguiluz Moya, 2014).
- Heterocigosis observada y heterocigosis esperada: La heterocigosis es una medida de la variación genética dentro de una población. Altos valores de heterocigosis pueden darse por una selección natural a largo plazo, o a apareamientos al azar y entre diferentes poblaciones, mientras que un bajo nivel de heterocigosis puede significar el aislamiento de una población. La heterocigosis observada (H_o), es el cálculo directo de la frecuencia de individuos o genotipos heterocigotos.

Por otra parte, la heterocigosis esperada (H_e), también llamada diversidad génica (Nei, 1973) se calcula de las frecuencias alélicas individuales y se define como la probabilidad de que dos alelos escogidos al azar en una población sean diferentes. En términos prácticos, este término equivale a la heterocigosis que se esperaría en condiciones de HWE.

- Equilibrio de Hardy-Weinberg (HWE): La mayoría de las deducciones sobre las poblaciones y la genética cuantitativa dependen de este parámetro. Las características que debe tener una población para encontrarse en HWE se discuten ampliamente en la sección 2.7 de este trabajo. La desviación de una población del HWE se puede probar usando diferentes métodos. En esta tesis para detectar una desviación importante en el HWE se utilizó una prueba análoga a la prueba exacta de Fisher para tablas de contingencia (Excoffier y Lischer, 2010). Además, como los locis son multialélicos, se empleó un algoritmo de cadena de Markov que se utilizó para obtener una estimación imparcial de la probabilidad exacta de equivocarse al rechazar la hipótesis nula. Este método se emplea en muestras con tamaños pequeños (menos de 50 muestras) y locus multialélicos, ya que la prueba de Chi cuadrada no es válida en estos casos.
- Estadístico de Wright y comparación entre pares de poblaciones: El estadístico de Wright es la medida más comúnmente utilizada para cuantificar las posibles subdivisiones de una población y puede ser utilizado como una medida de las distancias genéticas entre poblaciones a corto plazo. La significancia de los valores de F_{ST} calculados es evaluada permutando los haplotipos entre poblaciones. La comparación entre pares usando el valor de F_{ST} se da en forma de matriz, y el valor de P es la proporción de permutaciones que dan como resultado un valor de F_{ST} mayor o igual al observado. Los valores de P también se dan en forma de matriz (Excoffier y Lischer, 2010).

Tabla 4.1: Loci y alelos del kit AmpF/STRs®Identifiler®Plus

Nombre del locus	Localización cromosómica	Alelos incluidos en el ladder alelico	Colorante	Alelos del DNA control
D8S1179	8	8,9,10,11,12,13,14,15,16,17,18,19		13
D21S11	21q11.2-q21	24,24.2,25,26,27,28,28.2,29,29.2,30,30.2,31,31.2,32,32.2,33,33.2,34,34.2,35,35.2,36,37,38	FAM	30
D7S820	7q11.21-22	6,7,8,9,10,11,12,13,14,15		10,11
CSF1PO	5q33.3-34	6,7,8,9,10,11,12,13,14,15		10,12
D3S1358	3p	12,13,14,15,16,17,18,19	VIC	14,15
TH01	11p15.5	4,5,6,7,8,9,9.3,10,11,13,3		8,9,3
DI3S317	13q22-31	8,9,10,11,12,13,14,15		11
DI6S539	16q24-qter	5,8,9,10,11,12,13,14,15		11,12
D2S1338	2q35-37.1	15,16,17,18,19,20,21,22,23,24,25,26,27,28		19,23
D19S433	19q12-13.1	9,10,11,12,12.2,13,13.2,14,14.2,15,15.2,16,16.2,17,17.2	NED	14,15
vWA	12p12-pter	11,12,13,14,15,16,17,18,19,20,21,22,23,24		17,18
TPOX	2p23-2per	6,7,8,9,10,11,12,13		8
DI8S51	18q21.3	7,9,10,10.2,11,12,13,13.2,14,14.2,15,16,17,18,19,20,21,22,23,24,25,26,27		15,19
Amelogenina	X: p22.1-22.3\Y: p11.2	X,Y	PET	X
D5S818	5q21-31	7,8,9,10,11,12,13,14,15,16		11
FGA	4q28	17,18,19,20,21,22,23,24,25,26,26.2,27,28,29,30,30.2,31.2,32.2,33.2,42.2,43.2,44.2,45.2,46.2,47.2,48.2,50.2,51.2		23,24

Capítulo 5

Resultados

*Si buscas resultados distintos, no hagas
siempre lo mismo*

Albert Einstein

Frecuencias alélicas

Una vez obtenidos los electroferogramas, se prosiguió a capturar los datos en el software Arlequin V 3.5.2, obteniéndose así las frecuencias alélicas de cada locus en las diferentes poblaciones. Estos datos se muestran a continuación en las tablas 5.1 a la 5.9. Para facilitar el análisis de las frecuencias alélicas se realizaron las tablas 5.10 a la 5.24, en las que se analizan las frecuencias alélicas de un solo locus los nueve estados muestreados.

Alelo	D8S1179	D21S11	D7S820	CSFIPO	D3S1358	TH01	D13S317	D16S539	D2S1338	D19S433	VWA	TPOX	D18S51	D5S818	FGA
6						0.42593						0.00370			
7			0.00741			0.37407								0.11481	
8	0.00370		0.04444			0.01481	0.02963	0.00370				0.51852			
9	0.00370		0.03333	0.03333		0.03704	0.33704	0.08148				0.04074		0.07407	
9.3						0.14815									
10	0.10370		0.22222	0.26667			0.16667	0.28519	0.00370			0.00370	0.00370	0.03704	
11	0.04074		0.32963	0.25926			0.11852	0.24074				0.26296		0.58148	
12	0.12222		0.32963	0.40370			0.14815	0.31111	0.01111			0.17037	0.09630	0.15556	
12.2									0.03704						
13	0.34074		0.02963	0.03704	0.01852		0.14815	0.07778	0.17037	0.00370			0.11852	0.03704	
13.2									0.17037						
14	0.23704		0.00370		0.03704		0.04815		0.21111	0.01481			0.20370		
14.2									0.08148						
15	0.12222				0.53704				0.14074	0.05926			0.17407		
15.2									0.11111						
16	0.01481				0.26296				0.03704	0.44815			0.08889		
16.2									0.00370						
17	0.01111				0.10370				0.01852				0.19259		
17.2									0.00370						
18					0.02963		0.00370		0.06667		0.14815		0.04815		0.00741
19					0.00370				0.28148		0.04074		0.02963		0.09259
20					0.00741				0.12963		0.01111		0.01852		0.04444
21									0.02593				0.01111		0.05185
22									0.11481				0.01111		0.05556
22.2															
23									0.22593				0.00370		0.01111
23.2															
24									0.04074						0.00370
25									0.00741						0.23333
26		0.00370													0.23704
27		0.00370													0.11481
28		0.04074													0.03704
29		0.18889													0.03704
29.2		0.00370													0.00741
30		0.28519													
30.2		0.01111													
31		0.07407													
31.2		0.11852													
32		0.01481													
32.2		0.16667													
33.2		0.08148													
34.2		0.00370													
35		0.00370													

Tabla 5.1: Frecuencias alélicas encontradas en los 135 individuos de Veracruz.

Alelo	D8S1179	D21S11	D7S820	CSFIPO	D3S1358	TH01	D13S317	D16S539	D2S1338	D19S433	VWA	TPOX	D18S51	D5S818	FGA
4						0.00813									
6						0.33333						0.00407			
7			0.00407			0.41057								0.06911	
8			0.04878	0.00407		0.00813	0.04878	0.00407				0.47154			
9			0.01626	0.01220		0.03659	0.26829	0.06911				0.00813		0.06098	
9.3						0.19919									
10	0.10163		0.20732	0.25203		0.00407	0.16260	0.29675				0.00407		0.03252	
11	0.02033		0.39024	0.24797		0.15447	0.23984		0.02439			0.27236	0.00407	0.58537	
12	0.13415		0.30081	0.42683		0.13415	0.32520		0.00407			0.22764	0.09350	0.21138	
12.2									0.12602						
13	0.35772		0.03252	0.05691		0.15854	0.04878		0.19106			0.01220	0.12602	0.04065	
13.2									0.19106						
14	0.22764				0.03252		0.06911	0.01220	0.00407	0.22358			0.02033	0.19512	
14.2									0.06911						
15	0.10976				0.60976		0.00407	0.00407	0.15447		0.05285		0.16260		
15.2									0.13821						
16	0.04878				0.28455				0.00813	0.02846	0.45935		0.04472		
16.2									0.02439						
17					0.05285				0.08537	0.00407	0.23984		0.21951		0.00407
18					0.02033				0.04065	0.01220	0.17073		0.08130		0.10569
19									0.24797		0.04878		0.01626		0.02846
20									0.16667		0.00813		0.00813		0.09350
21									0.02033				0.03659		0.07724
22									0.08943				0.00813		0.13008
23									0.28049				0.00407		0.02033
23.2															0.15041
24		0.00407							0.04472						0.18699
25									0.01220						0.13821
26		0.00407													0.04878
27		0.00407													0.00813
28		0.03252													
29		0.15447													
30		0.28862													
30.2		0.00407													
31		0.05691													0.00407
31.2		0.15447													
32		0.00813													
32.2		0.20732													
33		0.00407													
33.2		0.08130													
34.2															0.00407

Tabla 5.2: Frecuencias alélicas encontradas en los 123 individuos de Oaxaca.

Alelo	D8S11179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D13S317	D16S539	D2S1338	D19S433	VWA	TPOX	DI8S51	D5S818	FGA
6						0.28177									
7			0.00552	0.00549		0.41989						0.52747		0.08242	
8	0.00549		0.05525	0.01099		0.00552	0.06044	0.00549				0.02198	0.00549	0.06044	
9	0.00549		0.03867	0.01099		0.07735	0.35165	0.07692							
9.3						0.20994									
10	0.09890		0.24862	0.19780			0.10989	0.29121		0.00549		0.03297	0.00549	0.06044	
10.2															
10.3						0.00552	0.10989								
11	0.04396		0.39779	0.27473	0.00549		0.19231	0.26923				0.26923	0.00549	0.52747	
12	0.08791		0.17127	0.40659			0.13736	0.26923		0.02747		0.13736	0.10989	0.17582	
12.2										0.03846					
13	0.34615		0.07182	0.07692	0.00549		0.10440	0.07692		0.17033		0.01099	0.09341	0.08791	
13.2										0.15385					
14	0.29670		0.01105	0.01099	0.03297		0.04396	0.00549		0.19780	0.03297		0.10440	0.00549	
14.2										0.06593					
15	0.09890				0.53846			0.00549		0.11538	0.03297		0.21978		
15.2										0.13187					
16	0.01648				0.27473				0.01657	0.04945	0.42308		0.08791		
16.2										0.03846					
16.3										0.00549					
17					0.09890				0.08287		0.31868		0.18681		
18					0.03846				0.04972		0.14286		0.08242		
19					0.00549				0.34807		0.04945		0.01648		
20				0.00549					0.16022				0.03297		0.07692
21									0.01105				0.02198		0.03297
22									0.09945				0.00549		0.07143
23									0.14917				0.00549		0.12088
24									0.05525				0.00549		0.17033
25									0.02210				0.00549		0.12088
26									0.00552				0.00549		0.20330
27	0.01099												0.00549		0.18132
28	0.03297														0.02198
29	0.19231														
29.2	0.00549														
30	0.27473														
30.2	0.01648														
31	0.11538														
31.2	0.14286														
32	0.01099														
32.2	0.15385														
33.2	0.03297														
33.3	0.00549														
34.2	0.00549														
35															

Tabla 5.3: Frecuencias alélicas encontradas en los 91 individuos del Estado de México.

Alelo	D8S1179	D21S11	D7S820	CSFIPO	D3S1358	TH01	D13S317	D16S539	D2S1338	D19S433	VWA	TPOX	DI8S51	D5S818	FGA
6						0.50633									
7				0.00633		0.25949		0.01266						0.06329	
8			0.03165			0.01899	0.06329					0.41772			
9			0.03165	0.01899		0.03797	0.33544	0.11392				0.00633			
9.3						0.17089									
10	0.04430		0.29114	0.22152		0.00633	0.10759	0.24684	0.00633			0.02532	0.00633	0.09494	0.05063
11	0.01899		0.28481	0.26582			0.13291	0.27848				0.3354	0.01266	0.51266	
12	0.12658		0.32278	0.36076			0.13291	0.29114	0.00633			0.21519	0.06962	0.18987	
12.2									0.02532						
13	0.34810		0.03165	0.10127			0.14557	0.05696	0.25316	0.00633			0.10759	0.08228	
13.2									0.20253						
14	0.29747		0.00633	0.02532	0.04430		0.07595		0.18354	0.04430			0.24051	0.00633	
14.2									0.05063						
15	0.15823				0.56962		0.00633		0.12025	0.05063			0.13924		
15.2									0.06962						
16	0.00633				0.24051				0.04430	0.41139			0.11392		
16.2									0.03165						
17					0.12025				0.05063	0.00633	0.30380		0.13924		0.03165
18					0.02532				0.05696		0.10759		0.08228		0.06962
19									0.31013		0.05696		0.03165		0.03165
20									0.18354		0.01899				0.07595
21									0.00633				0.02532		0.04430
22									0.05696				0.00633		0.09494
23									0.27848				0.01899		0.00633
23.2															0.17089
24									0.05063						0.18987
25									0.00633						0.20253
26															0.08228
27		0.00633													
28		0.07595													
29		0.27848													
29.2		0.00633													
30		0.20253													
31		0.04430													
31.2		0.19620													
32.2		0.10759													
33.2		0.06962													
34.2		0.00633													
35.2		0.00633													

Tabla 5.4: Frecuencias alélicas encontradas en los 79 individuos de Chiapas.

Alelo	D8S1179	D21S11	D7S820	CSFIPO	D3S1358	TH01	D13S317	D16S539	D2S1338	D19S433	VWA	TPOX	DI8S51	D5S818	FGA
6						0.42949									
7			0.01282			0.35256						0.00641		0.07051	
8			0.06410			0.01282						0.57051			
9	0.01282		0.03846	0.03205		0.03846		0.08974				0.00641		0.05128	
9.3						0.17949									
10	0.14744		0.25641	0.24359			0.13462	0.18590				0.00641		0.04487	
11	0.04487		0.33974	0.25000	0.00641		0.13462	0.30769				0.24359		0.56410	
11.2									0.01282						
12	0.111538		0.23077	0.42949			0.13462	0.38462	0.00641			0.16667		0.18590	
12.2									0.04487						
13	0.333333		0.05128	0.03205	0.01923		0.15385	0.03205	0.15385				0.10256	0.08333	
13.2							0.22436		0.22436						
14	0.25000		0.00641	0.00641	0.03205		0.21154	0.00641	0.21154				0.21154		
14.2							0.04487		0.04487						
15	0.07051			0.00641	0.53205		0.15385	0.05128	0.15385				0.15385		
15.2							0.11538		0.11538						
16	0.01923				0.32692		0.01282	0.46154	0.01282				0.06410		
16.2							0.01923		0.01923						
17	0.00641								0.05128				0.22436		
18					0.07051				0.07692		0.30769		0.07692		
19					0.00641				0.33333		0.15385		0.01923		0.10256
20					0.00641				0.16667		0.01923		0.01923		0.01282
21									0.03205				0.01282		0.07051
22									0.08333				0.01282		0.10256
23									0.20513				0.00641		0.08333
23.2									0.03846				0.00641		0.00641
24									0.00641				0.00641		0.21154
25									0.00641						0.23718
26									0.00641						0.12821
27															0.03205
28															0.01282
29		0.05769													
30		0.19231													
30.2		0.32051													
31		0.00641													
31.2		0.08333													
32		0.12821													
32.2		0.01282													
33.2		0.12821													
34.2		0.04487													
35		0.01923													
		0.00641													

Tabla 5.5: Frecuencias alélicas encontradas en los 78 individuos de Puebla.

Alelo	D8S1179	D21S11	D7S820	CSFIPO	D3S1358	TH01	D13S317	D16S539	D2S1338	D19S433	VWA	TPOX	DI8S51	D5S818	FGA
6						0.43878						0.01020			
7						0.34694								0.07143	
8			0.111224			0.04082	0.01020			0.47959				0.01020	
9			0.06122	0.01020		0.04082	0.30612	0.12245				0.04082		0.04082	
9.3						0.12245									
10	0.08163		0.25510	0.15306		0.01020	0.18367	0.25510				0.01020		0.03061	
11	0.01020		0.33673	0.27551		0.11224	0.11224	0.29592				0.14286	0.01020	0.61224	
12	0.10204		0.19388	0.41837		0.16327	0.26531		0.03061			0.30612	0.08163	0.19388	
12.2									0.03061						
13	0.33673		0.03061	0.13265		0.13265	0.06122		0.08163			0.01020	0.13265	0.04082	
13.2									0.25510						
14	0.30612		0.01020	0.01020	0.10204		0.09184		0.24490				0.15306		
14.2									0.04082		0.04082				
15	0.10204				0.50000				0.14286		0.09184		0.20408		
15.2									0.10204						
16	0.05102				0.25510				0.03061		0.38776		0.13265		
16.2									0.03061						
16.3															
17	0.01020				0.11224				0.13265		0.28571		0.20408		
17.2										0.01020					
18					0.02041				0.05102		0.11224		0.01020		0.01020
19					0.01020				0.23469		0.08163		0.03061		0.12245
20									0.15306						0.06122
21									0.01020						0.08163
22									0.10204				0.01020		0.19388
23									0.20408				0.02041		0.08163
24									0.07143						0.14286
25									0.03061						0.15306
26									0.01020				0.01020		0.11224
27	0.01020														0.02041
28	0.03061														0.02041
29	0.12245														
30	0.33673														
30.2															
31															
31.2															
32															
32.2															
33.2															
36															

Tabla 5.6: Frecuencias alélicas encontradas en los 49 individuos de Guerrero.

Alelo	D8S1179	D21S11	D7S820	CSFIPO	D3S1358	TH01	D13S317	D16S539	D2S1338	D19S433	VWA	TPOX	D18S51	D5S818	FGA
6						0.32292						0.01042			
7						0.38542								0.06250	
8			0.07447	0.01042		0.01042	0.02083	0.01042				0.47917			
9			0.02128	0.01042		0.03125	0.32292	0.10417				0.01042		0.09375	
9.3						0.25000									
10	0.07292		0.25532	0.27083			0.17708	0.27083						0.03125	
11	0.05208		0.24468	0.26042			0.14583	0.21875				0.26042		0.55208	
12	0.12500		0.31915	0.39583			0.17708	0.33333		0.01042		0.23958	0.06250	0.17708	
12.2										0.01042					
13	0.28125		0.06383	0.03125			0.07292	0.06250		0.12500			0.07292	0.08333	
13.2										0.26042					
14	0.28125		0.02128	0.01042	0.02083		0.08333			0.26042			0.13542		
14.2										0.05208	0.03125				
15	0.14583			0.01042	0.54167					0.12500	0.04167		0.12500		
15.2										0.08333					
16	0.04167				0.26042				0.01042	0.04167	0.33333		0.15625		0.01042
16.2										0.02083					
17					0.14583				0.09375	0.01042	0.36458		0.28125		
18					0.03125				0.03125		0.17708		0.07292		
19									0.36458		0.05208		0.02083		0.06250
19.3															0.01042
20									0.17708				0.04167		0.01042
21									0.07292				0.01042		0.01042
22									0.18750						0.07292
23									0.18750				0.02083		0.11458
23.2															0.10417
24									0.05208						0.01042
25									0.01042						0.18750
26															0.20833
27															0.17708
28		0.04167													0.02083
29		0.22917													0.01042
30		0.22917													0.18750
31		0.10417													0.20833
31.2		0.11458													0.17708
32		0.01042													0.02083
32.2		0.18750													0.01042
33.2		0.08333													0.01042

Tabla 5.7: Frecuencias alélicas encontradas en los 48 individuos de Hidalgo.

Alelo	D8S1179	D21S11	D7S820	CSFIPO	D3S1358	TH01	D13S317	D16S539	D2S1338	D19S433	VWA	TPOX	DI8S51	D5S818	FGA
6						0.50000						0.01613			
7				0.01613		0.30645								0.16129	
8	0.01613		0.08475	0.03226		0.03226	0.08065	0.32258				0.35484			
9			0.05085	0.03226		0.08065	0.25806	0.16129				0.06452			
9.3						0.08065									
10	0.09677		0.27119	0.16129		0.11290						0.08065	0.01613	0.03226	
11	0.01613		0.33898	0.20968		0.08065						0.24194		0.67742	
11.2	0.12903		0.25424	0.45161		0.17742			0.01613			0.24194	0.08065	0.08065	
12.2										0.06452					
13	0.27419			0.09677	0.01613	0.14516	0.03226			0.14516	0.01613		0.11290	0.04839	
13.2										0.12903	0.09677		0.17742		
14	0.27419				0.06452	0.12903			0.01613	0.33871	0.09677				
14.2										0.08065			0.27419		
15	0.17742				0.51613	0.01613	0.01613	0.01613	0.01613	0.16129	0.03226				
15.2										0.03226					
16	0.01613				0.25806					0.03226	0.45161		0.06452		
17					0.09677				0.11290		0.17742		0.19355		
18					0.04839				0.01613		0.17742		0.03226		
19									0.35484		0.04839		0.03226		0.04839
20									0.09677						
21									0.01613						0.04839
22									0.12903						0.04839
23									0.22581						0.12903
24									0.01613				0.01613		0.09677
25															0.17742
26															0.19355
27		0.01613													0.25806
28		0.06452													0.04839
29		0.30645													0.04839
30		0.22581													0.04839
30.2		0.04839													0.04839
31		0.03226													0.04839
31.2		0.14516													0.04839
32.2		0.14516													0.04839
33.2		0.04839													0.04839
34.2		0.01613													0.04839

Tabla 5.8: Frecuencias alélicas encontradas en los 31 individuos de Tabasco.

Alelo	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D13S317	D16S539	D2S1338	D19S433	VWA	TPOX	D18S51	D5S818	FGA
6						0.27500									
7						0.27500									
8			0.10000				0.10000					0.50000		0.07500	
9			0.05000	0.02500		0.02500	0.10000	0.05000				0.05000		0.02500	
9.3						0.42500									
10	0.15000		0.25000	0.35000			0.17500	0.25000				0.02500		0.12500	
11	0.05000		0.35000	0.32500			0.22500	0.42500				0.27500		0.40000	
12	0.10000		0.22500	0.27500			0.20000	0.20000	0.02500			0.15000	0.12500	0.22500	
13	0.27500		0.02500	0.02500			0.15000	0.07500	0.17500			0.10000	0.10000	0.15000	
13.2									0.07500						
14	0.35000				0.07500		0.05000		0.22500				0.27500		
14.2									0.10000						
15	0.07500				0.45000				0.15000	0.15000			0.07500		
15.2									0.15000						
16					0.30000				0.02500	0.02500	0.35000		0.12500		
16.2									0.07500						
17					0.15000				0.12500		0.32500		0.22500		
18									0.10000		0.12500		0.02500		
19					0.02500				0.27500	0.05000	0.05000		0.02500		
20									0.15000				0.05000		0.12500
21									0.05000				0.05000		0.07500
22									0.02500						0.27500
23									0.15000						0.12500
24									0.02500						0.10000
25									0.05000						0.07500
26									0.02500						0.12500
28		0.07500													
29		0.22500													
29.2		0.02500													
30		0.25000													
31.2		0.15000													
32.2		0.15000													
33.2		0.07500													

Tabla 5.9: Frecuencias alélicas encontradas en los 20 individuos de la Ciudad de México.

Alelos	Frecuencias								
	Veracruz	Oaxaca	Estado de México	Chiapas	Puebla	Guerrero	Hidalgo	Tabasco	Ciudad de México
8	0.00370	0	0.00549	0	0	0	0	0.01613	0
9	0.00370	0	0.00549	0	0.01282	0	0	0	0
10	0.10370	0.10163	0.09890	0.04430	0.14744	0.08163	0.07292	0.09677	0.15000
11	0.04074	0.02033	0.04396	0.01899	0.04487	0.01020	0.05208	0.01613	0.05000
12	0.12222	0.13415	0.08791	0.12658	0.11538	0.10204	0.12500	0.12903	0.10000
13	0.34074	0.35772	0.34615	0.34810	0.33333	0.33673	0.28125	0.27419	0.27500
14	0.23704	0.22764	0.29670	0.29747	0.25000	0.30612	0.28125	0.27419	0.35000
15	0.12222	0.10976	0.09890	0.15823	0.07051	0.10204	0.14583	0.17742	0.07500
16	0.01481	0.04878	0.01648	0.00633	0.01923	0.05102	0.04167	0.01613	0
17	0.01111	0	0	0	0.00641	0.01020	0	0	0

Tabla 5.10: Frecuencias alélicas para el locus D8S1179.

Alelos	Frecuencias								
	Veracruz	Oaxaca	Estado de México	Chiapas	Puebla	Guerrero	Hidalgo	Tabasco	Ciudad de México
25	0	0.00407	0	0	0	0	0	0	0
26	0.00370	0	0	0	0	0	0	0	0
27	0.00370	0.00407	0.01099	0.00633	0	0.01020	0	0.01613	0
28	0.04074	0.03252	0.03297	0.07595	0.05769	0.03061	0.04167	0.06452	0.07500
29	0.18889	0.15447	0.19231	0.27848	0.19231	0.12245	0.22917	0.30645	0.22500
29.2	0.00370	0	0.00549	0.00633	0	0	0	0	0.02500
30	0.28519	0.28862	0.27473	0.20253	0.32051	0.33673	0.22917	0.22581	0.25000
30.2	0.01111	0.00407	0.01648	0	0.00641	0.02041	0	0.04839	0
31	0.07407	0.05691	0.11538	0.04430	0.08333	0.06122	0.10417	0.03226	0.05000
31.2	0.11852	0.15447	0.14286	0.19620	0.12821	0.15306	0.11458	0.14516	0.15000
32	0.01481	0.00813	0.01099	0	0.01282	0.04082	0.01042	0	0
32.2	0.16667	0.20732	0.15385	0.10759	0.12821	0.18367	0.18750	0.09677	0.15000
33	0	0.00407	0	0	0	0	0	0	0
33.2	0.08148	0.08130	0.03297	0.06962	0.04487	0.03061	0.08333	0.04839	0.07500
33.3	0	0	0.00549	0	0	0	0	0	0
34.2	0.00370	0	0.00549	0.00633	0.01923	0	0	0.01613	0
35	0.00370	0	0	0	0.00641	0	0	0	0
35.2	0	0	0	0.00633	0	0	0	0	0
36	0	0	0	0	0	0.01020	0	0	0

Tabla 5.11: Frecuencias alélicas para el locus D21S11.

Alelos	Frecuencias								
	Veracruz	Oaxaca	Estado de México	Chiapas	Puebla	Guerrero	Hidalgo	Tabasco	Ciudad de México
7	0.00741	0.00407	0.00552	0	0.01282	0	0	0	0
8	0.04444	0.04878	0.05525	0.03165	0.06410	0.11224	0.07447	0.08475	0.10000
9	0.03333	0.01626	0.03867	0.03165	0.03846	0.06122	0.02128	0.05085	0.05000
10	0.22222	0.20732	0.24862	0.29114	0.25641	0.25510	0.25532	0.27119	0.25000
11	0.32963	0.39024	0.39779	0.28481	0.33974	0.33673	0.24468	0.33898	0.35000
12	0.32963	0.30081	0.17127	0.32278	0.23077	0.19388	0.31915	0.25424	0.22500
13	0.02963	0.03252	0.07182	0.03165	0.05128	0.03061	0.06383	0	0.02500
14	0.00370	0	0.01105	0.00633	0.00641	0.01020	0.02128	0	0

Tabla 5.12: Frecuencias alélicas para el locus D7S820.

Alelos	Frecuencias								
	Veracruz	Oaxaca	Estado de México	Chiapas	Puebla	Guerrero	Hidalgo	Tabasco	Ciudad de México
7	0	0	0.00549	0.00633	0	0	0	0.01613	0
8	0	0.00407	0.01099	0	0	0	0.01042	0.03226	0
9	0.03333	0.01220	0.01099	0.01899	0.03205	0.01020	0.01042	0.03226	0.02500
10	0.26667	0.25203	0.19780	0.22152	0.24359	0.15306	0.27083	0.16129	0.35000
11	0.25926	0.24797	0.27473	0.26582	0.25000	0.27551	0.26042	0.20968	0.32500
12	0.40370	0.42683	0.40659	0.36076	0.42949	0.41837	0.39583	0.45161	0.27500
13	0.03704	0.05691	0.07692	0.10127	0.03205	0.13265	0.03125	0.09677	0.02500
14	0	0	0.01099	0.02532	0.00641	0.01020	0.01042	0	0
15	0	0	0	0	0.00641	0	0.01042	0	0
19	0	0	0.00549	0	0	0	0	0	0

Tabla 5.13: Frecuencias alélicas para el locus CSF1PO.

Alelos	Frecuencias								
	Veracruz	Oaxaca	Estado de México	Chiapas	Puebla	Guerrero	Hidalgo	Tabasco	Ciudad de México
11	0	0	0.00549	0	0.00641	0	0	0	0
13	0.01852	0	0.00549	0	0.01923	0	0	0.01613	0
14	0.03704	0.03252	0.03297	0.04430	0.03205	0.10204	0.02083	0.06452	0.07500
15	0.53704	0.60976	0.53846	0.56962	0.53205	0.50000	0.54167	0.51613	0.45000
16	0.26296	0.28455	0.27473	0.24051	0.32692	0.25510	0.26042	0.25806	0.30000
17	0.10370	0.05285	0.09890	0.12025	0.07051	0.11224	0.14583	0.09677	0.15000
18	0.02963	0.02033	0.03846	0.02532	0.00641	0.02041	0.03125	0.04839	0
19	0.00370	0	0.00549	0	0	0.01020	0	0	0.02500
20	0.00741	0	0	0	0.00641	0	0	0	0

Tabla 5.14: Frecuencias alélicas para el locus D3S1358.

Alelos	Frecuencias								
	Veracruz	Oaxaca	Estado de México	Chiapas	Puebla	Guerrero	Hidalgo	Tabasco	Ciudad de México
4	0	0.00813	0	0	0	0	0	0	0
6	0.42593	0.33333	0.28177	0.50633	0.42949	0.43878	0.32292	0.50000	0.27500
7	0.37407	0.41057	0.41989	0.25949	0.35256	0.34694	0.38542	0.30645	0.27500
8	0.01481	0.00813	0.00552	0.01899	0.01282	0.04082	0.01042	0.03226	0
9	0.03704	0.03659	0.07735	0.03797	0.02564	0.04082	0.03125	0.08065	0.02500
9.3	0.14815	0.19919	0.20994	0.17089	0.17949	0.12245	0.25000	0.08065	0.42500
10	0	0.00407	0	0.00633	0	0.01020	0	0	0
10.3	0	0	0.00552	0	0	0	0	0	0

Tabla 5.15: Frecuencias alélicas para el locus TH01.

Alelos	Frecuencias								
	Veracruz	Oaxaca	Estado de México	Chiapas	Puebla	Guerrero	Hidalgo	Tabasco	Ciudad de México
8	0.02963	0.04878	0.06044	0.06329	0.03846	0.01020	0.02083	0.08065	0.10000
9	0.33704	0.26829	0.35165	0.33544	0.33333	0.30612	0.32292	0.25806	0.10000
10	0.16667	0.16260	0.10989	0.10759	0.13462	0.18367	0.17708	0.11290	0.17500
11	0.11852	0.15447	0.19231	0.13291	0.13462	0.11224	0.14583	0.08065	0.22500
12	0.14815	0.13415	0.13736	0.13291	0.13462	0.16327	0.17708	0.17742	0.20000
13	0.14815	0.15854	0.10440	0.14557	0.15385	0.13265	0.07292	0.14516	0.15000
14	0.04815	0.06911	0.04396	0.07595	0.07051	0.09184	0.08333	0.12903	0.05000
15	0	0.00407	0	0.00633	0	0	0	0.01613	0
18	0.00370	0	0	0	0	0	0	0	0

Tabla 5.16: Frecuencias alélicas para el locus D13S317.

Alelos	Frecuencias								
	Veracruz	Oaxaca	Estado de México	Chiapas	Puebla	Guerrero	Hidalgo	Tabasco	Ciudad de México
7	0	0	0	0.01266	0	0	0	0	0
8	0.00370	0.00407	0.00549	0	0	0	0.01042	0	0
9	0.08148	0.06911	0.07692	0.11392	0.08974	0.12245	0.10417	0.16129	0.05000
10	0.28519	0.29675	0.29121	0.24684	0.18590	0.25510	0.27083	0.32258	0.25000
11	0.24074	0.23984	0.26923	0.27848	0.30769	0.29592	0.21875	0.30645	0.42500
12	0.31111	0.32520	0.26923	0.29114	0.38462	0.26531	0.33333	0.16129	0.20000
13	0.07778	0.04878	0.07692	0.05696	0.03205	0.06122	0.06250	0.03226	0.07500
14	0	0.01220	0.00549	0	0	0	0	0.01613	0
15	0	0.00407	0.00549	0	0	0	0	0	0

Tabla 5.17: Frecuencias alélicas para el locus D16S539.

Alelos	Frecuencias								
	Veracruz	Oaxaca	Estado de México	Chiapas	Puebla	Guerrero	Hidalgo	Tabasco	Ciudad de México
14	0	0.00407	0	0	0	0	0	0.01613	0
15	0	0	0	0	0	0	0	0.01613	0
16	0.00370	0.00813	0.01657	0	0	0	0.01042	0	0.02500
17	0.10370	0.08537	0.08287	0.05063	0.05128	0.13265	0.09375	0.11290	0.12500
18	0.06667	0.04065	0.04972	0.05696	0.07692	0.05102	0.03125	0.01613	0.10000
19	0.28148	0.24797	0.34807	0.31013	0.33333	0.23469	0.36458	0.35484	0.27500
20	0.12963	0.16667	0.16022	0.18354	0.16667	0.15306	0.17708	0.09677	0.15000
21	0.02593	0.02033	0.01105	0.00633	0.03205	0.01020	0	0.01613	0.05000
22	0.11481	0.08943	0.09945	0.05696	0.08333	0.10204	0.07292	0.12903	0.02500
23	0.22593	0.28049	0.14917	0.27848	0.20513	0.20408	0.18750	0.22581	0.15000
24	0.04074	0.04472	0.05525	0.05063	0.03846	0.07143	0.05208	0.01613	0.02500
25	0.00741	0.01220	0.02210	0.00633	0.00641	0.03061	0.01042	0	0.05000
26	0	0	0.00552	0	0.00641	0.01020	0	0	0.02500

Tabla 5.18: Frecuencias alélicas para el locus D2S1338.

Alelos	Frecuencias								
	Veracruz	Oaxaca	Estado de México	Chiapas	Puebla	Guerrero	Hidalgo	Tabasco	Ciudad de México
10	0.00370	0	0	0.00633	0	0	0	0	0
10.2	0	0	0.00549	0	0	0	0	0	0
11.2	0	0	0	0	0.01282	0	0	0.01613	0
12	0.01111	0.02439	0.02747	0.00633	0.00641	0.03061	0.01042	0	0.02500
12.2	0.03704	0.00407	0.03846	0.02532	0.04487	0.03061	0.01042	0.06452	0
13	0.17037	0.12602	0.17033	0.25316	0.15385	0.08163	0.12500	0.14516	0.17500
13.2	0.17037	0.19106	0.15385	0.20253	0.22436	0.25510	0.26042	0.12903	0.07500
14	0.21111	0.22358	0.19780	0.18354	0.21154	0.24490	0.26042	0.33871	0.22500
14.2	0.08148	0.06911	0.06593	0.05063	0.04487	0.04082	0.05208	0.08065	0.10000
15	0.14074	0.15447	0.11538	0.12025	0.15385	0.14286	0.12500	0.16129	0.15000
15.2	0.11111	0.13821	0.13187	0.06962	0.11538	0.10204	0.08333	0.03226	0.15000
16	0.03704	0.02846	0.04945	0.04430	0.01282	0.03061	0.04167	0.03226	0.02500
16.2	0.01852	0.02439	0.03846	0.03165	0.01923	0.03061	0.02083	0	0.07500
16.3	0	0	0.00549	0	0	0	0	0	0
17	0.00370	0.00407	0	0.00633	0	0	0.01042	0	0
17.2	0.00370	0	0	0	0	0.01020	0	0	0
18	0	0.01220	0	0	0	0	0	0	0

Tabla 5.19: Frecuencias alélicas para el locus D19S433.

Alelos	Frecuencias								
	Veracruz	Oaxaca	Estado de México	Chiapas	Puebla	Guerrero	Hidalgo	Tabasco	Ciudad de México
13	0.00370	0	0	0.00633	0	0	0	0.01613	0
14	0.01481	0.02033	0.03297	0.04430	0.00641	0.04082	0.03125	0.09677	0
15	0.05926	0.05285	0.03297	0.05063	0.05128	0.09184	0.04167	0.03226	0.15000
16	0.44815	0.45935	0.42308	0.41139	0.46154	0.38776	0.33333	0.45161	0.35000
17	0.27407	0.23984	0.31868	0.30380	0.30769	0.28571	0.36458	0.17742	0.32500
18	0.14815	0.17073	0.14286	0.10759	0.15385	0.11224	0.17708	0.17742	0.12500
19	0.04074	0.04878	0.04945	0.05696	0.01923	0.08163	0.05208	0.04839	0.05000
20	0.01111	0.00813	0	0.01899	0	0	0	0	0

Tabla 5.20: Frecuencias alélicas para el locus VWA.

Alelos	Frecuencias								
	Veracruz	Oaxaca	Estado de México	Chiapas	Puebla	Guerrero	Hidalgo	Tabasco	Ciudad de México
6	0.00370	0.00407	0	0	0	0	0.01042	0.01613	0
7	0	0	0	0	0.00641	0.01020	0	0	0
8	0.51852	0.47154	0.52747	0.41772	0.57051	0.47959	0.47917	0.35484	0.50000
9	0.04074	0.00813	0.02198	0.00633	0.00641	0.04082	0.01042	0.06452	0.05000
10	0.00370	0.00407	0.03297	0.02532	0.00641	0.01020	0	0.08065	0.02500
11	0.26296	0.27236	0.26923	0.33544	0.24359	0.14286	0.26042	0.24194	0.27500
12	0.17037	0.22764	0.13736	0.21519	0.16667	0.30612	0.23958	0.24194	0.15000
13	0	0.01220	0.01099	0	0	0.01020	0	0	0

Tabla 5.21: Frecuencias alélicas para el locus TPOX.

Alelos	Frecuencias								
	Veracruz	Oaxaca	Estado de México	Chiapas	Puebla	Guerrero	Hidalgo	Tabasco	Ciudad de México
10	0.00370	0	0.00549	0.00633	0	0	0	0.01613	0
11	0	0.00407	0.00549	0.01266	0	0.01020	0	0	0
11.2	0	0	0	0	0.00641	0	0	0	0
12	0.09630	0.09350	0.10989	0.06962	0.09615	0.08163	0.06250	0.08065	0.12500
13	0.11852	0.12602	0.09341	0.10759	0.10256	0.13265	0.07292	0.11290	0.10000
14	0.20370	0.19512	0.10440	0.24051	0.21154	0.15306	0.13542	0.17742	0.27500
15	0.17407	0.16260	0.21978	0.13924	0.15385	0.20408	0.12500	0.27419	0.07500
16	0.08889	0.04472	0.08791	0.11392	0.06410	0.13265	0.15625	0.06452	0.12500
17	0.19259	0.21951	0.18681	0.13924	0.22436	0.20408	0.28125	0.19355	0.22500
18	0.04815	0.08130	0.08242	0.08228	0.07692	0.01020	0.07292	0.03226	0.02500
19	0.02963	0.01626	0.01648	0.03165	0.01923	0.03061	0.02083	0.03226	0
20	0.01852	0.00813	0.03297	0	0.01923	0	0.04167	0	0.05000
21	0.01111	0.03659	0.02198	0.02532	0.01282	0	0.01042	0	0
22	0.01111	0.00813	0.00549	0.00633	0	0.01020	0	0	0
23	0.00370	0.00407	0.00549	0.01899	0.00641	0.02041	0.02083	0	0
24	0	0	0.00549	0	0.00641	0	0	0.01613	0
25	0	0	0.00549	0	0	0	0	0	0
26	0	0	0	0	0	0.01020	0	0	0

Tabla 5.22: Frecuencias alélicas para el locus D18S51.

Alelos	Frecuencias								
	Veracruz	Oaxaca	Estado de México	Chiapas	Puebla	Guerrero	Hidalgo	Tabasco	Ciudad de México
7	0.11481	0.06911	0.08242	0.06329	0.07051	0.07143	0.06250	0.16129	0.07500
8	0	0	0	0	0	0.01020	0	0	0
9	0.07407	0.06098	0.06044	0.09494	0.05128	0.04082	0.09375	0	0.02500
10	0.03704	0.03252	0.06044	0.05063	0.04487	0.03061	0.03125	0.03226	0.12500
11	0.58148	0.58537	0.52747	0.51266	0.56410	0.61224	0.55208	0.67742	0.40000
12	0.15556	0.21138	0.17582	0.18987	0.18590	0.19388	0.17708	0.08065	0.22500
13	0.03704	0.04065	0.08791	0.08228	0.08333	0.04082	0.08333	0.04839	0.15000
14	0	0	0.00549	0.00633	0	0	0	0	0

Tabla 5.23: Frecuencias alélicas para el locus D5S818.

Alelos	Frecuencias								
	Veracruz	Oaxaca	Estado de México	Chiapas	Puebla	Guerrero	Hidalgo	Tabasco	Ciudad de México
16	0	0	0	0	0	0	0.01042	0	0
18	0.00741	0.00407	0	0.03165	0	0.01020	0	0	0
19	0.09259	0.10569	0.07692	0.06962	0.10256	0.12245	0.06250	0.04839	0.12500
19.3	0	0	0	0	0	0	0.01042	0	0
20	0.04444	0.02846	0.03297	0.03165	0.01282	0.06122	0.01042	0	0.07500
21	0.05185	0.09350	0.07143	0.07595	0.07051	0.08163	0.07292	0.04839	0.27500
22	0.05556	0.07724	0.12088	0.04430	0.10256	0.19388	0.11458	0.12903	0.12500
22.2	0.01111	0	0	0	0	0	0	0	0
23	0.10370	0.13008	0.17033	0.09494	0.08333	0.08163	0.10417	0.09677	0.10000
23.2	0.00370	0.02033	0	0.00633	0.00641	0	0.01042	0	0
24	0.23333	0.15041	0.12088	0.17089	0.21154	0.14286	0.18750	0.17742	0.07500
25	0.23704	0.18699	0.20330	0.18987	0.23718	0.15306	0.20833	0.19355	0.12500
26	0.11481	0.13821	0.18132	0.20253	0.12821	0.11224	0.17708	0.25806	0.10000
27	0.03704	0.04878	0.02198	0.08228	0.03205	0.02041	0.02083	0.04839	0
28	0.00741	0.00813	0	0	0	0.02041	0.01042	0	0
29	0	0	0	0	0.01282	0	0	0	0
32.2	0	0.00407	0	0	0	0	0	0	0
34.2	0	0.00407	0	0	0	0	0	0	0

Tabla 5.24: Frecuencias alélicas para el locus FGA.

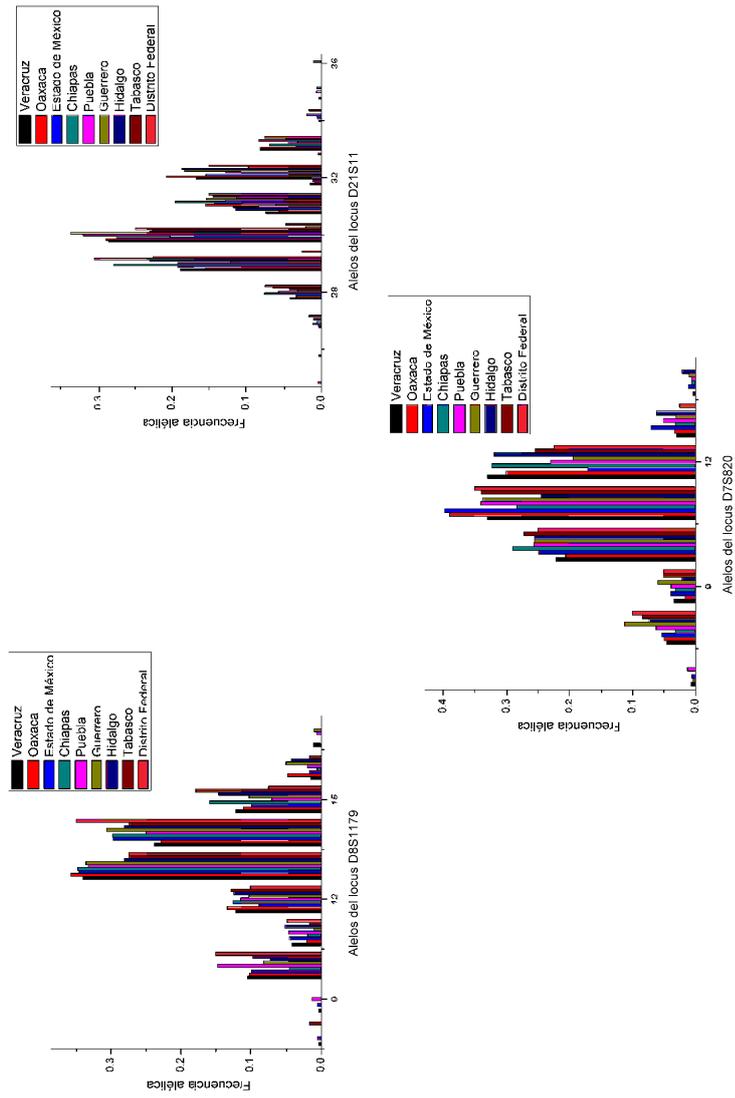


Figura 5.1: Representación esquemática de las frecuencias alélicas del locus D8S1179, D21S11 y D7S820.

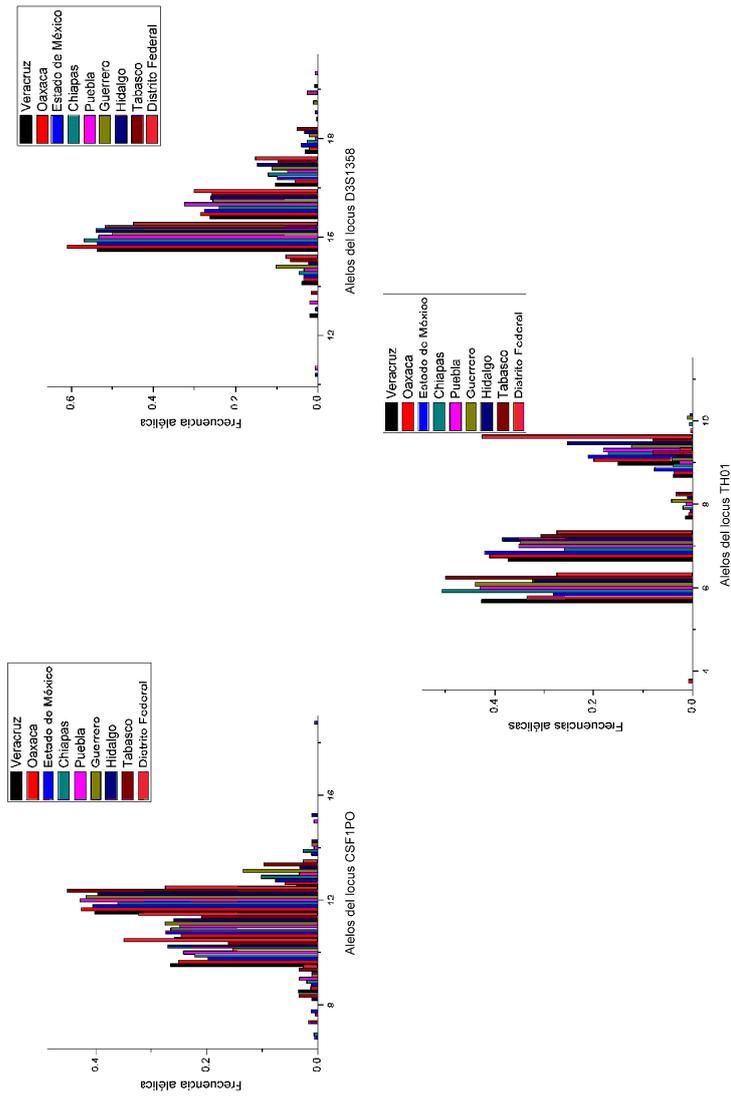


Figura 5.2: Representación esquemática de las frecuencias alélicas del locus CSF1PO, D3S1358 y TH01.

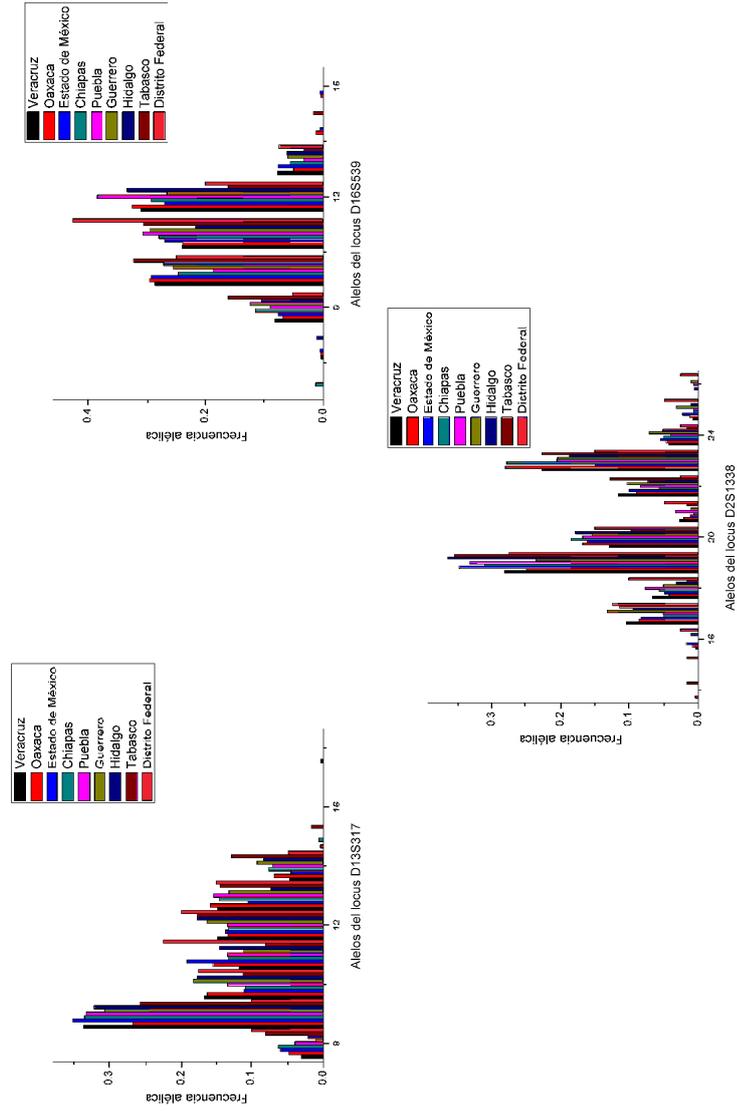


Figura 5.3: Representación esquemática de las frecuencias alélicas del locus D13S317, D16S539 y D2S1338.

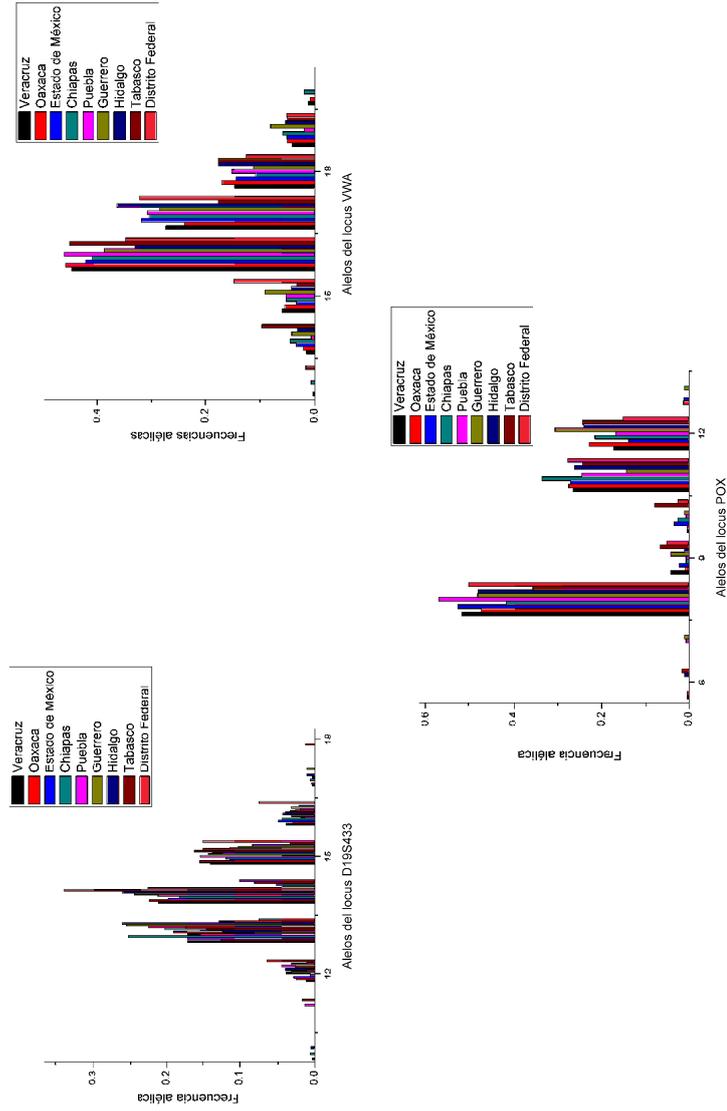


Figura 5.4: Representación esquemática de las frecuencias alélicas del locus D19S433, VWA y POX.

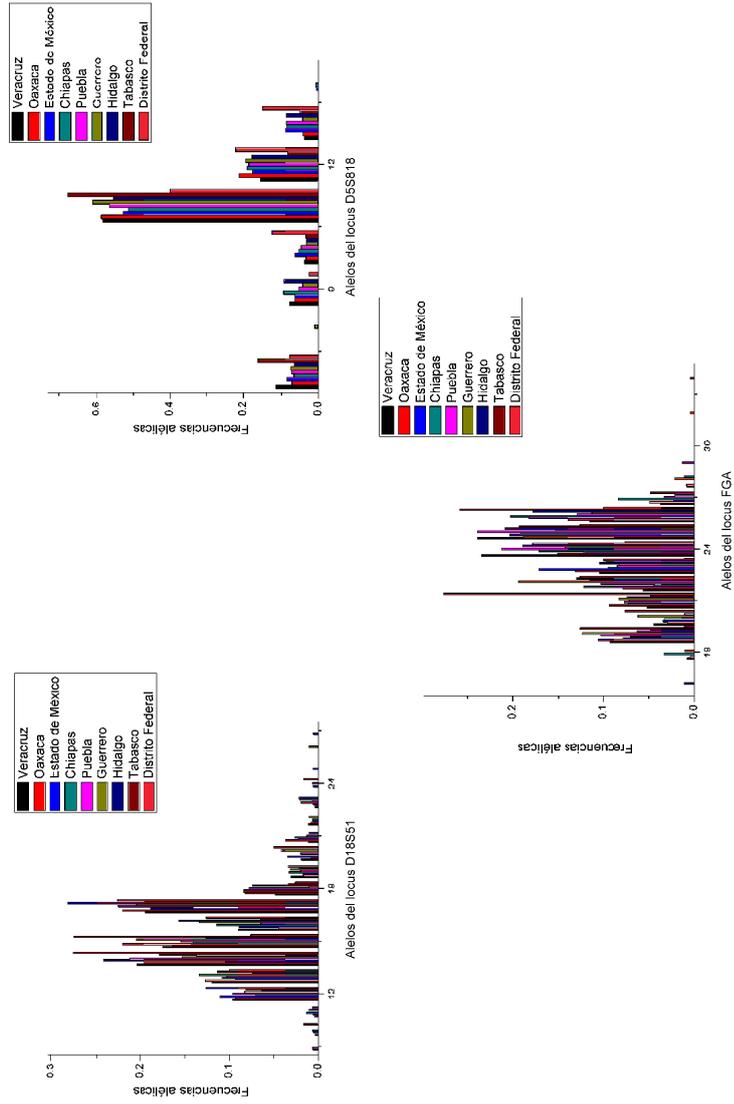


Figura 5.5: Representación esquemática de las frecuencias alélicas del locus D18S51, D5S818 y FGA.

Número de alelos

En la tabla 5.25 se muestran cuántos alelos se encontraron por locus. Este parámetro se calculó con el software Arlequin ver 3.5.2.

Locus	VERACRUZ	OAXACA	EDO. MEX.	CHIAPAS	PUEBLA	GUERRERO	HIDALGO	TABASCO	CDMX
D8S1179	10	7	9	7	9	8	7	8	6
D21S11	14	12	13	11	11	11	8	10	8
D7S820	8	7	8	7	8	7	7	5	6
CSF1PO	5	6	9	7	7	6	8	7	5
D3S1358	8	5	8	5	8	6	5	6	5
TH01	5	7	6	6	5	6	5	5	4
D13S317	8	8	7	8	7	7	7	8	7
D16S539	6	8	8	6	5	5	6	6	5
D2S1338	10	11	11	9	10	10	9	10	11
D19S433	13	12	12	12	11	11	11	9	9
VWA	8	7	6	8	6	6	6	7	5
POX	6	7	6	5	6	7	5	6	5
D18S51	13	13	18	14	13	12	11	10	8
D5S818	6	6	7	7	6	7	6	5	6
FGA	13	14	9	11	11	11	13	8	8

Tabla 5.25: Número de alelos diferentes encontrados para cada locus.

Equilibrio de Hardy-Weinberg

Para poder probar la existencia del HWE en los diferentes locus de las poblaciones se planteó una prueba de hipótesis bilateral, que se muestra a continuación:

$$\begin{aligned}
 H_0 &= H_{esperada} = H_{observada} \\
 H_a &= H_{esperada} \neq H_{observada}
 \end{aligned}
 \tag{5.1}$$

El nivel de confianza de la prueba es del 95 %, por lo tanto, el valor de $\alpha = 0,05$. Sin embargo, hay que considerar que en este trabajo se probaron múltiples pruebas de hipótesis simultáneamente, por lo tanto es necesario realizar la corrección de Bonferroni (Weisstein, 2004). El nuevo valor de α para todas las comparaciones se calculó usando la siguiente fórmula: $\alpha_{EW} = \alpha/\kappa$. Lo que nos da como resultado un valor de $\alpha = 0,05/15 = 0,003$. En las tablas 5.26 a la 5.34, se encuentran los valores para la H_e y H_o y el p-value. El locus fuera del HWE se muestran en negritas (tabla 5.29).

Locus	Núm. de genotipos	Het.Obs.	Het.Esp.	p-value
D8S1179	135	0.74074	0.78797	0.64000
D21S11	135	0.84444	0.83004	0.85220
D7S820	135	0.69630	0.73198	0.08692
CSF1PO	135	0.69630	0.69880	0.41414
D3S1358	135	0.64444	0.63136	0.13040
TH01	135	0.60000	0.65755	0.09111
D13S317	135	0.80000	0.80044	0.81708
D16S539	135	0.77037	0.75401	0.77448
D2S1338	135	0.79259	0.82520	0.16224
D19S433	135	0.81481	0.85852	0.49478
VWA	135	0.71111	0.69916	0.79376
POX	135	0.62963	0.63362	0.59465
D18S51	135	0.89630	0.85926	0.63754
D5S818	135	0.60000	0.61856	0.36340
FGA	135	0.77778	0.85064	0.01987

Tabla 5.26: Cálculos del HWE para la población de Veracruz.

Locus	Núm. de genotipos	Het.Obs.	Het.Esp.	p-value
D8S1179	123	0.78862	0.78022	0.84394
D21S11	123	0.79675	0.81828	0.39309
D7S820	123	0.70732	0.71342	0.13296
CSF1PO	123	0.70732	0.69222	0.15867
D3S1358	123	0.53659	0.54518	0.06967
TH01	123	0.70732	0.68193	0.14487
D13S317	123	0.78862	0.83079	0.60328
D16S539	123	0.78049	0.74435	0.78636
D2S1338	123	0.79675	0.81580	0.88407
D19S433	123	0.81301	0.85117	0.66104
VWA	123	0.73984	0.69952	0.12680
POX	123	0.60976	0.65406	0.19435
D18S51	123	0.82927	0.85578	0.13141
D5S818	123	0.54472	0.60392	0.08640
FGA	123	0.88618	0.88037	0.35263

Tabla 5.27: Cálculos del HWE para la población de Oaxaca.

Locus	Núm. de genotipos	Het.Obs.	Het.Esp.	p-value
D8S1179	91	0.81319	0.76680	0.63114
D21S11	91	0.76923	0.83195	0.12147
D7S820	90	0.77778	0.74389	0.50805
CSF1PO	91	0.73626	0.71769	0.82060
D3S1358	91	0.63736	0.62558	0.71397
TH01	90	0.70000	0.69547	0.64692
D13S317	91	0.83516	0.79631	0.21934
D16S539	91	0.79121	0.76249	0.85885
D2S1338	90	0.84444	0.81397	0.11797
D19S433	91	0.86813	0.87171	0.43629
VWA	91	0.64835	0.69826	0.27372
POX	91	0.62637	0.63220	0.70266
D18S51	91	0.89011	0.87329	0.49450
D5S818	91	0.65934	0.67270	0.96056
FGA	91	0.91209	0.85969	0.57974

Tabla 5.28: Cálculos del HWE para la población del Estado de México.

Locus	Núm. de genotipos	Het.Obs.	Het.Esp.	p-value
D8S1179	79	0.77215	0.75167	0.54216
D21S11	79	0.82278	0.82383	0.44563
D7S820	79	0.67089	0.73152	0.23498
CSF1PO	79	0.75949	0.74353	0.74514
D3S1358	79	0.60759	0.60445	0.74961
TH01	79	0.62025	0.64936	0.03247
D13S317	79	0.82278	0.81472	0.58060
D16S539	79	0.82278	0.76522	0.35763
D2S1338	79	0.72152	0.78586	0.89905
D19S433	79	0.87342	0.84093	0.25810
VWA	79	0.70886	0.72329	0.29609
POX	79	0.63291	0.67024	0.34195
D18S51	79	0.83544	0.87043	0.35528
D5S818	79	0.62025	0.68306	0.00269
FGA	79	0.89873	0.86882	0.69392

Tabla 5.29: Cálculos del HWE para la población de Chiapas.

Locus	Núm. de genotipos	Het.Obs.	Het.Esp.	p-value
D8S1179	78	0.65385	0.78883	0.12614
D21S11	78	0.84615	0.81977	0.26441
D7S820	78	0.78205	0.76203	0.22533
CSF1PO	78	0.66667	0.69603	0.58163
D3S1358	78	0.61538	0.60744	0.92376
TH01	78	0.55128	0.66245	0.33148
D13S317	78	0.73077	0.80959	0.02796
D16S539	78	0.65385	0.71836	0.17002
D2S1338	78	0.74359	0.80612	0.49735
D19S433	78	0.87179	0.84491	0.06012
VWA	78	0.75641	0.66989	0.29342
POX	78	0.56410	0.59107	0.79480
D18S51	78	0.83333	0.85591	0.03879
D5S818	78	0.65385	0.63474	0.66934
FGA	78	0.85897	0.85368	0.70033

Tabla 5.30: Cálculos del HWE para la población de Puebla.

Locus	Núm. de genotipos	Het.Obs.	Het.Esp.	p-value
D8S1179	49	0.73469	0.77046	0.86282
D21S11	49	0.85714	0.81485	0.34800
D7S820	49	0.75510	0.77446	0.47739
CSF1PO	49	0.65306	0.71513	0.15332
D3S1358	49	0.71429	0.66821	0.69765
TH01	49	0.61224	0.67557	0.21792
D13S317	49	0.77551	0.81548	0.65558
D16S539	49	0.81633	0.76604	0.97204
D2S1338	49	0.85714	0.85167	0.76123
D19S433	49	0.83673	0.84052	0.43844
VWA	49	0.73469	0.74627	0.29187
POX	49	0.71429	0.66064	0.97362
D18S51	49	0.97959	0.85841	0.90142
D5S818	49	0.67347	0.58405	0.22496
FGA	49	0.87755	0.88197	0.21632

Tabla 5.31: Cálculos del HWE para la población de Guerrero.

Locus	Núm. de genotipos	Het.Obs.	Het.Esp.	p-value
D8S1179	48	0.77083	0.80351	0.47754
D21S11	48	0.87500	0.83575	0.26866
D7S820	46	0.76087	0.77090	0.82363
CSF1PO	48	0.81250	0.70811	0.34277
D3S1358	48	0.58333	0.62259	0.28809
TH01	48	0.54167	0.69079	0.04559
D13S317	48	0.77083	0.80746	0.37601
D16S539	48	0.68750	0.76075	0.39799
D2S1338	48	0.79167	0.79079	0.03825
D19S433	48	0.89583	0.82961	0.73567
VWA	48	0.70833	0.72675	0.23135
POX	48	0.64583	0.65175	0.20277
D18S51	48	0.83333	0.85417	0.21987
D5S818	48	0.66667	0.65000	0.92870
FGA	48	0.89583	0.86491	0.32237

Tabla 5.32: Cálculos del HWE para la población de Hidalgo.

Locus	Núm. de genotipos	Het.Obs.	Het.Esp.	p-value
D8S1179	31	0.80645	0.80434	0.19027
D21S11	31	0.80645	0.82760	0.43715
D7S820	28	0.75000	0.73831	0.04082
CSF1PO	31	0.70968	0.72607	0.07143
D3S1358	31	0.70968	0.66155	0.88718
TH01	31	0.64516	0.65256	0.90462
D13S317	31	0.80645	0.85193	0.04133
D16S539	31	0.77419	0.76097	0.77575
D2S1338	31	0.77419	0.79588	0.49601
D19S433	31	0.70968	0.82179	0.00430
VWA	31	0.64516	0.73189	0.04428
POX	31	0.67742	0.75833	0.22162
D18S51	31	0.87097	0.84347	0.37193
D5S818	31	0.51613	0.51348	0.61835
FGA	31	0.80645	0.84506	0.61520

Tabla 5.33: Cálculos del HWE para la población de Tabasco.

Locus	Núm. de genotipos	Het.Obs.	Het.Esp.	p-value
D8S1179	20	0.75000	0.78077	0.65443
D21S11	20	0.80000	0.84872	0.68573
D7S820	20	0.65000	0.77051	0.39722
CSF1PO	20	0.70000	0.71282	1.00000
D3S1358	20	0.60000	0.69615	0.32073
TH01	20	0.75000	0.68462	1.00000
D13S317	20	0.85000	0.85513	0.81523
D16S539	20	0.65000	0.72692	0.72187
D2S1338	20	0.80000	0.86795	0.19529
D19S433	20	0.85000	0.87308	0.44544
VWA	20	0.55000	0.75000	0.51805
POX	20	0.60000	0.66538	0.73465
D18S51	20	0.95000	0.84487	0.99642
D5S818	20	0.80000	0.76410	0.29108
FGA	20	0.90000	0.86795	0.97362

Tabla 5.34: Cálculos del HWE para la población de CDMX.

Comparación entre pares usando el estadístico de Wright (F_{ST})

Los valores de F_{ST} entre pares fueron obtenidos en forma de matriz y se muestran en la tabla 5.35.

Para poder probar que los valores de F_{ST} son diferentes de una población a otra se utilizó una prueba de hipótesis como la siguiente:

$$\begin{aligned} H_0 &= F_{ST} \text{ pobl,1} = F_{ST} \text{ pobl,2} \\ H_a &= F_{ST} \text{ pobl,1} \neq F_{ST} \text{ pobl,2} \end{aligned} \quad (5.2)$$

El nivel de confianza de la prueba es del 95 %, por lo tanto, $\alpha = 0,05$. Una vez realizada la corrección de Bonferroni (Weisstein, 2004) $\alpha = 0,05/9 = 0,005$, pues se hicieron nueve comparaciones con las diferentes poblaciones. Por lo tanto:

Si el $p - value \leq 0,005$ se rechaza H_0 , es decir, el valor de F_{ST} de la población 1 no es significativamente diferente del de la población 2.

En cambio, si $p - value > 0,003$ no se rechaza H_0 , es decir, el valor de F_{ST} de la población 1 es diferente al de la población 2.

El valor de la probabilidad (P-value) también se dan en forma de matriz. Los valores de F_{ST} los P-values de F_{ST} se obtuvieron usando el software Arlequin ver 3.5.2. Los P-values F_{ST} significativos se muestran en negritas, en la tabla 5.36.

	Ver	Oax	EdoMex	Chis	Pue	Gro	Hgo	Tab	D.F.
Ver	0.00000								
Oax	0.00001	0.00000							
EdoMex	0.00310	0.00260	0.00000						
Chis	0.00319	0.00624	0.00718	0.00000					
Pue	-0.00116	0.00113	0.00137	0.00351	0.00000				
Gro	0.00155	0.00247	0.00281	0.00628	0.00151	0.00000			
Hgo	-0.00017	0.00151	0.00001	0.00141	0.00032	0.00156	0.00000		
Tab	0.00338	0.00651	0.00690	0.00424	0.00652	0.00112	0.00523	0.00000	
D.F.	0.01023	0.00879	0.00502	0.01084	0.00784	0.00813	0.00422	0.01490	0.00000

Tabla 5.35: Valores de F_{ST} por pares de poblaciones

	Ver	Oax	EdoMex	Chis	Pue	Gro	Hgo	Tab	D.F.
Ver	0.47748+-0.0526								
Oax	0.00000+-0.00000	*							
EdoMex	0.00901+-0.0091	0.00901+-0.0091	*						
Chis	0.00000+-0.00000	0.00000+-0.00000	0.00000+-0.00000	*					
Pue	0.90090+-0.0272	0.16216+-0.0485	0.15315+-0.0360	0.00000+-0.00000	*				
Gro	0.15315+-0.0333	0.03604+-0.0148	0.02703+-0.0139	0.00000+-0.00000	0.27928+-0.0286	*			
Hgo	0.50450+-0.0309	0.19820+-0.0424	0.39640+-0.0338	0.24324+-0.0430	0.40541+-0.0365	0.19820+-0.0402	*		
Tab	0.09910+-0.0286	0.00000+-0.00000	0.02703+-0.0139	0.05405+-0.0242	0.02703+-0.0194	0.34234+-0.0297	0.02703+-0.0139	*	
D.F.	0.00000+-0.00000	0.00901+-0.0091	0.05405+-0.0148	0.00000+-0.00000	0.01802+-0.0121	0.00000+-0.00000	0.20721+-0.0430	0.00901+-0.0091	*

Tabla 5.36: P- values de F_{ST} entre las 9 poblaciones de estudio

Capítulo 6

Análisis de Resultados y Conclusiones

Nada tiene sentido en evolución si no es a la luz de la Genética de poblaciones

Michael Lynch

Análisis de resultados

Frecuencias alélicas

Las frecuencias alélicas representan la incidencia de una variante de un gen. En una población las frecuencias alélicas son el reflejo de diversidad genética. En este trabajo, los loci con mayor número de alelos diferentes fueron el D21S11, D18S51 y el FGA. Estos datos corresponden con lo reportado anteriormente, donde los locus más polimórficos son el D18S1 y el FGA para poblaciones del centro, norte, sur y occidente de México (Rubi-Castellanos et al., 2009a) (Salazar-Flores et al., 2015) (Juárez-Cedillo et al., 2008). Algunos alelos de los loci D21S11, CSF1PO, TH01, D13S317, D2S1338, D19S433, D18S51, D5S818 y FGA sólo se encontraron en una de las nueve poblaciones analizadas. Estos alelos, podrían considerarse exclusivos de cierta región, sin embargo, este efecto puede deberse al azar, pues sólo se muestreó una pequeña parte de la población y es factible que dichos alelos se encuentren en otras poblaciones en individuos que no fueron muestreados.

Equilibrio de Hardy Weinberg

Respecto al HWE, después de evaluar los valores de p una vez aplicada la corrección de Bonferroni, únicamente el locus D5S18 en la población de Chiapas no se encuentra en equilibrio. Esta desviación del HWE puede deberse a numerosas causas. La primera de ellas es que no se cumplan las presunciones establecidas por HWE y que el locus analizado en esta población se encuentre bajo selección,

mutación o deriva génica. Esto parece improbable, ya que si así fuera, otros loci de la misma población no cumplirían con el HWE. La causa más probable de la desviación encontrada son los errores de genotipificación. Estos errores ocurren cuando el genotipo reportado de un individuo no coincide con el genotipo verdadero. Las principales causas de estos errores fueron agrupadas en tres categorías por Pompanon y colaboradores. Estas categorías son: 1) variaciones en la secuencia de DNA, 2) baja calidad o cantidad de DNA y 3) errores humanos. Es posible que alguno o algunos de estos errores pudieran surgir en la realización de este trabajo. En los estudios de genética de poblaciones un error de genotipificación tiene como resultado el reconocimiento incorrecto de un alelo, resultando en estimados de F_{ST} incorrectos (Pompanon et al., 2005). Como sugieren Sen y colaboradores, los errores de genotipificación comúnmente no son estocásticos, y son más comunes en ciertos loci. Si cierto locus es propenso a un error de genotipificación entonces la desviación del HWE se hará más grande conforme el tamaño de muestra aumente, lo que no ocurrirá si el error no existe o es de origen aleatorio (Sen y Burmeister, 2008). Por este motivo, se sugeriría aumentar el número de individuos muestreados para evaluar si en verdad existe un error de genotipificación que afecte el EHW del locus D5S818 en la población de Chiapas.

División de la población usando el estadístico de Wright

El estadístico de Wright se ha usado muy ampliamente para cuantificar la subdivisión de una población, usando como guía el parámetro F_{ST} , que se define como la correlación que existe entre alelos escogidos de manera aleatoria dentro de la misma población, relativo a la población entera, o dicho de otra manera, la proporción de la diversidad genética debido a las diferencias de frecuencias alélicas entre poblaciones (Holsinger y Weir, 2009).

Si las frecuencias alélicas son las mismas en todas las poblaciones $F_{ST} = 0$. Si por el contrario las frecuencias alélicas están diferenciadas de manera máxima en las diferentes poblaciones (ciertos alelos han sido fijados) entonces $F_{ST} = 1$. Valores negativos de F_{ST} se toman como cero.

Una guía para la interpretación de los valores de F_{ST} sugiere que valores entre 0 y 0.05 indican nula o mínima diferenciación, valores entre 0.15 y 0.25 indican gran diferenciación y valores mayores de 0.25 indican una diferenciación excesivamente grande (Nielsen y Slatkin, 2013).

A pesar de que estos valores dan una referencia de la diferenciación de las poblaciones, es necesario hacer una prueba estadística adicional para determinar si éstos son significativamente diferentes de una población a otra. Por esta razón se calcularon los valores de P (P-value). Después de realizar la corrección de Bonferroni ($P < 0,0056$) los valores significativamente diferentes se muestran en la tabla 5.36 en negritas. La población de Chiapas es la que presenta más diferenciación con respecto a las demás regiones del país, siendo diferente a casi todas las poblaciones estudiadas, excepto a Tabasco e Hidalgo. Es muy probable que la similitud de la población de Chiapas con la de Tabasco se deba principalmente a su cercanía.

Por otra parte, Chiapas no se diferenció de la población de Hidalgo, a pesar de que se encuentran a una distancia considerablemente mayor. Interesantemente, la población de Hidalgo no se diferenció de ninguna de las poblaciones de este estudio. Este fenómeno se había documentado anteriormente, en el estudio realizado por Castellanos y colaboradores en el 2009, en donde se estudiaron 10 poblaciones diferentes y todas mostraron diferencias con una o más poblaciones, a excepción de la población de Hidalgo. Es muy probable que esto se deba a la presencia de alelos nulos, en vez de que no exista una diferencia real entre esta población y otras de la República Mexicana (Rubi-Castellanos et al., 2009b).



Figura 6.1: Localización geográfica de las poblaciones bajo estudio y su agrupación en regiones según el estadístico de Wright

Según los resultados de este trabajo, las poblaciones estudiadas se agruparon en cuatro regiones diferentes, basándose en la diferenciación que existe entre ellas (figura 6.1):

- Centro Sur: Estado de México y Distrito Federal
- Centro Oriente: Puebla, Veracruz e Hidalgo
- Sureste: Chiapas y Tabasco
- Suroeste: Guerrero y Oaxaca

Según lo reportado anteriormente por Castellanos y colaboradores en 2 artículos diferentes (Rubi-Castellanos et al., 2009a); (Rubi-Castellanos et al., 2009b), las poblaciones de México se pueden dividir de manera general en dos grandes regiones: la región noreste (Chihuahua, Jalisco y Nuevo León) y la región centro-sureste (Ciudad de México, Puebla, Veracruz, y Campeche). En el caso del presente trabajo, no fue posible evaluar la región noreste, ya que no se cuentan con muestras de la región norte del país. Por otra parte, a diferencia de lo encontrado por Castellanos, nosotros encontramos una fuerte diferenciación entre la región centro y sur de la República. Estas diferencias pueden deberse a los distintos criterios con los que se seleccionaron los individuos pertenecientes a la población, pues como se indica en el capítulo 4, todos los individuos incluidos en este estudio tienen padres y abuelos originarios del mismo estado, criterio que no se utilizó en los estudios de Castellanos anteriormente citados.

Nuestro trabajo coincide con los encontrados en la literatura en el hecho de que Puebla es la población menos diferenciada del centro de la República, presentando similitudes con la mayoría de las poblaciones del centro y suroeste del país. La explicación más probable para este fenómeno tiene que ver con la cercanía de este estado con muchos de centro sur y centro oriente, tales como Hidalgo, Veracruz, Oaxaca, Guerrero, Morelos y el Estado de México (Figura 6.1).

Conclusiones

Los alelos más polimórficos fueron el D21S11, D18S51 y el FGA, lo que coincide con otros trabajos de población mestiza mexicana.

La mayoría de los loci se encuentran en HWE, a excepción del locus D5S818 para la población de Chiapas.

Usando el estadístico de F_{ST} se pudieron calcular las distancias genéticas entre las diferentes poblaciones en estudio, dando como resultado la agrupación de estas en cuatro regiones principales: Centro Sur, Centro Oriente, Sureste y Suroeste. La población más diferenciada fue la de Chiapas, mientras que la población de Hidalgo no fue significativamente diferente de ninguna otra, posiblemente a consecuencia de alelos nulos.

Bibliografía

*Libros, caminos y días dan al hombre
sabiduría.*

Proverbio árabe

- AMORIM, A. Population genetics. En *Brenner's Encyclopedia of Genetics* (editado por S. M. Hughes). Academic Press, Porto, segunda edición, 2013.
- ANANDA, G., HILE, S., BRESKI, A., WANG, Y., KELKAR, Y., MAKOVA, K. y ECKERT, K. Microsatellite interruptions stabilize primate genomes and exist as population-specific single nucleotide polymorphisms within individual human genomes. *PLoS Genetics*, vol. 10(7), 2014.
- ARANGUREN-MÉNDEZ, J. A., ROMÁN-BRAVO, R., ISEA, W., VILLASMIL, Y. y JORDANA, J. Los microsatélites (STRs), marcadores moleculares de ADN por excelencia para programas de conservación: una revisión. *Asociación Latinoamericana de Producción Animal*, vol. 13(1), 2005.
- ASICIOGLU, F., OGUZ-SAVRANB, F. y OZBEK, U. Mutation rate at commonly used forensic str loci: Paternity testing experience. *Disease Markers*, vol. 20(1), 2004.
- AVISE, J. *Inside the Human Genome: A Case for Non-Intelligent Design*. Oxford Scholarship Online, 2010.
- BARBADILLA, A. La genética de poblaciones. 2003. Disponible en <http://biologia.uab.es/divulgacio/genpob.html> (último acceso, Marzo,2015).
- BIOSYSTEMS, A. *AmpFI STR Identifier Plus PCR Amplification Kit User Guide*. Applied Biosystems and HITACHI, Foster City, California, 2015.
- BOIS, P. R. Hypermutable minisatellites, a human affair? *Genomics*, vol. 81(4), 2003.
- BONILLA, C., GUTIÉRREZ, G., PARRA, E., KLINE, C. y SHRIVER, M. Admixture analysis of a rural population of the state of Guerrero, Mexico. *American journal of physical anthropology*, vol. 128(4), 2005.

- BROWN, T. *Genomes*. Oxford: Wiley-Liss, 2002.
- CAL-TEBA, M. L. *Análisis de polimorfismos de ADN microsatélite del cromosoma Y*. Phd, Universidad de Santiago de Compostela, 2000.
- CHEN, J. J. The Hardy-Weinberg principle and its applications in modern population genetics. *Frontiers in Biology*, vol. 5(4), 2010.
- CHISTIAKOV, D. A., HELLEMANS, B. y VOLCKAERT, F. A. Microsatellites and their genomic distribution, evolution, function and applications: A review with special reference to fish genetics. *Aquaculture*, vol. 255(1), 2006.
- CURTIS, H., BOLAÑOS, P., ARÁOZ, J. y BEHRENS, V. *Biología*. Médica Panamericana, 2008.
- DEBRAUWERE, H., GENDREL, C., LECHAT, S. y DUTREIX, M. Differences and similarities between various tandem repeat sequences: Minisatellites and microsatellites. *Biochimie*, vol. 79(10), 1997.
- DELAROSA, R., JAMES, C. M. y TOBUTT, K. R. Using microsatellites for paternity testing in olive progenies. *Hort Science*, vol. 39(2), 2004.
- EGUIARTE, L. Nueva guía para principiantes en genética de poblaciones. *Evolución Biológica. Una visión actualizada de la revista Ciencias, Morrone, JJ, Magaña, P.,(eds.). Facultad de Ciencias, UNAM. México, 2009.*
- EGUILUZ MOYA, M. L. *Identificación de marcadores microsatelites para el estudio de la diversidad genética de Taenia solium*. Universidad Peruana Cayetano Heredia, 2014.
- ELLEGREN, H. Microsatellite mutations in the germline: Implications for evolutionary inference. *Trends in genetics*, vol. 16(12), 2000.
- ELLEGREN, H. Microsatellites: simple sequences with complex evolution. *Nature Reviews Genetics*, vol. 5(1), 2004.
- ENA. The human genome project, currently maintained by the genome reference consortium (grc). 2015. Disponible en http://www.ebi.ac.uk/ena/data/view/GCA_000001405.15 (último acceso, Enero,2015).
- ENSEMBL. Human assembly and gene annotation. 2015. Disponible en http://www.ensembl.org/Homo_sapiens/Info/Annotation#assembly (último acceso, Enero,2015).
- EXCOFFIER, L. y LISCHER, H. E. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under linux and windows. *Molecular ecology resources*, vol. 10(3), 2010.

- FAN, H. y CHU, J.-Y. A brief review of short tandem repeat mutation. *Geno. Prot. Bioinfo*, vol. 5(1), 2007.
- FERNÁNDEZ, P. B. *El minisatélite MsH43: una sorprendente secuencia del genoma humano*. Phd, Universidad Santiago de Compostela, 2005.
- GONZALÉZ-BLANCO, M. *Nuevas aportaciones al estudio del minisatélite humano MsH42: Filogenia, Digestión y Paradojas*. Phd, Universidad Santiago de Compostela, 2005.
- GUARDADO-ESTRADA, M., QUEIPO, G., MERAZ-RÍOS, M. y BERUMEN-CAMPOS, J. Diversidad genética en la población mexicana: Utilización de marcadores de adns. *Rev Med Hosp Gen Mex.*, vol. 71(3), 2008.
- HACKENBERG, M. *Dinámica evolutiva de los retrotrasposonas Alu en el genoma humano*. Phd, Universidad de Granada, 2005.
- HANNAN, A. Trping up the genome: Tandem repeat polymorphisms as dynamic sources of genetic variability in health and disease. *Discovery Medicine*, vol. 10(53), 2010.
- HOLSINGER, K. E. y WEIR, B. S. Genetics in geographically structured populations: defining, estimating and interpreting FST. *Nature Reviews Genetics*, vol. 10(9), 2009.
- IHGSC. Finishing the euchromatic sequence of the human genome. *Nature*, vol. 431(7011), 2004.
- JOBLING, M., HURLES, M. y TYLER-SMITH, C. *Human Evolutionary Genetics: Origins, Peoples and Disease*. Garland Science, 2013.
- JUÁREZ-CEDILLO, T., ZUÑIGA, J., ACUÑA-ALONZO, V., PÉREZ-HERNÁNDEZ, N., RODRÍGUEZ-PÉREZ, J. M., BARQUERA, R., GALLARDO, G. J., SÁNCHEZ-ARENAS, R., DEL CARMEN GARCÍA-PEÑA, M., GRANADOS, J. ET AL. Genetic admixture and diversity estimations in the mexican mestizo population from Mexico City using 15 STR polymorphic markers. *Forensic Science International: Genetics*, vol. 2(3), páginas e37–e39, 2008.
- KLUG, W., CUMMINGS, M. y SPENCER, C. *Conceptos de Genética*. Pearson Education, 2006.
- KNIGHT, J. *Human Genetic Diversity: Functional Consequences for Health and Disease*. Oxford University Press, 2009.
- KOSKINEN, M. T., HIRVONEN, H., LANDRY, P.-A. y PRIMMER, C. R. The benefits of increasing the number of microsatellites utilized in genetic population studies: an empirical perspective. *Hereditas*, vol. 141(1), 2004.

- LOPES, J., DEBRAUWERE, H., BURAD, J. y NICOLAS, A. Instability of the human minisatellite *ceb1* in *rad27* and *dna2-1* replication-deficient yeast cells. *The EMBO Journal*, vol. 21(12), 2002.
- MATHEWS, C., HOLDE, K. y AHERN, K. *Bioquímica*. Pearson education S.A., 2010.
- MILLER, W. y CAPY, P. Applying mobile genetic elements for genome analysis and evolution. *Molecular Biotechnology*, vol. 33(2), 2006.
- NEI, M. Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences*, vol. 70(12), 1973.
- NIELSEN, R. y SLATKIN, M. *A introduction to Population Genetics: Theory and Applications*. Sinauer Associates, Inc., 2013.
- OLIVEIRA, E., PADUA, J., ZUCCHI, M., R.VENCOVSKY y VIEIRA, M. Origin, evolution and genome distribution of microsatellites. *Genetics and Molecular Biology*, vol. 29(2), 2006.
- PASSARGE, E. *Genética, texto y atlas*. Médica Panamericana, 2010.
- PIERCE, B. *Genética: Un enfoque conceptual*. Médica Panamericana, 2009.
- POMPANON, F., BONIN, A., BELLEMAIN, E. y TABERLET, P. Genotyping errors: causes, consequences and solutions. *Nature Reviews Genetics*, vol. 6(11), 2005.
- PUTMAN, A. y CARBONE, I. Challenges in analysis and interpretation of microsatellite data for population genetic studies. *Ecology and Evolution*, vol. 4(22), 2014.
- QIAGEN. *DNeasy Blood and Tissue Handbook*. Qiagen, Austin, Texas, 2006.
- RAMAMOORTHY, S., GARAPATI, H. S. y MISHRA, R. K. Length and sequence dependent accumulation of simple sequence repeats in vertebrates: Potential role in genome organization and regulation. *Gene*, vol. 551(1), 2014.
- RAMÍREZ-BELLO, J., VARGAS-ALARCÓN, G., TOVILLA-ZÁRATE, C. y FRAGOSO, J. M. Polimorfismos de un solo nucleótido (snp): implicaciones funcionales de los snp reguladores (rsnp) y de los snp-arn estructurales (srsnp) en enfermedades complejas. *Gaceta Médica de México*, vol. 149, 2013.
- RICHARD, M. y THORPE, R. S. Can microsatellites be used to infer phylogenies? evidence from population affinities of the western canary island lizard (*Gallotia galloti*). *Mol Phylogenet Evol*, vol. 20(3), 2001.
- RUBI-CASTELLANOS, R., ANAYA-PALAFIX, M., MENA-ROJAS, E., BAUTISTA-ESPAÑA, D., MUÑOZ-VALLE, J. y RANGEL-VILLALOBOS, H. Genetic data of 15 autosomal STRs (identifiler kit) of three mexican mestizo

- population samples from the states of Jalisco (west), Puebla (center), and Yucatan (southeast). *Forensic Science International: Genetics*, vol. 193(3), 2009a.
- RUBI-CASTELLANOS, R., MARTÍNEZ-CORTÉS, G., FRANCISCO MUÑOZ-VALLE, J., GONZÁLEZ-MARTÍN, A., CERDA-FLORES, R. M., ANAYA-PALAFIX, M. y RANGEL-VILLALOBOS, H. Pre-hispanic mesoamerican demography approximates the present-day ancestry of mestizos throughout the territory of Mexico. *American Journal of Physical Anthropology*, vol. 139(3), 2009b.
- RUBI-CASTELLANOS, R., MARTÍNEZ-CORTÉS, G., MUÑOZ-VALLE, J., GONZÁLEZ-MARTÍN, A., CERDA-FLORES, R., ANAYA-PALAFIX, M. y RANGEL-VILLALOBOS, H. Pre-hispanic mesoamerican demography approximates the present-day ancestry of mestizos throughout the territory of Mexico. *American journal of physical anthropology*, vol. 193(3), 2009c.
- SALAZAR-FLORES, J., ZUÑIGA-CHIQUETTE, F., RUBI-CASTELLANOS, R., ÁLVAREZ-MIRANDA, J., ZETINA-HÉRNANDEZ, A., MARTÍNEZ-SEVILLA, V., GONZÁLEZ-ANDRADE, F., CORACH, D., VULLO, C., ÁLVAREZ, J. ET AL. Admixture and genetic relationships of Mexican mestizos regarding Latin American and Caribbean populations based on 13 CODIS-STRs. *HOMO-Journal of Comparative Human Biology*, vol. 66(1), 2015.
- SALIPANTE, S. Minisatellite. En *Brenner's Encyclopedia of Genetics* (editado por S. M. Hughes). Academic Press, San Diego, segunda edición, 2013.
- SEN, S. y BURMEISTER, M. Hardy-Weinberg analysis of a large set of published association studies reveals genotyping error and a deficit of heterozygotes across multiple loci. *Human genomics*, vol. 3(1), 2008.
- SHARMA, V. K., KUMAR, N., BRAHMACHARI, S. K. y RAMACHANDRAN, S. Abundance of dinucleotide repeats and gene expression are inversely correlated: a role for gene function in addition to intron length. *Physiological Genomics*, vol. 31(1), 2007.
- SHASTRY, B. S. SNP alleles in human disease and evolution. *J Hum Genet*, vol. 47, 2002.
- SIWACH, P. y GANESH, S. Tandem repeats in human disorders: mechanisms and evolution. *Frontiers in Bioscience*, vol. 13(1), 2008.
- STRAM, D. O. *Design, Analysis, and Interpretation of Genome-Wide Association Scans*. Versión electrónica, 2014.
- SUBRAMANIAN, S., MISHRA, R. y SINGH, L. Genome-wide analysis of microsatellite repeats in humans: their abundance and density in specific genomic regions. *Genome Biol.*, vol. 4(2), 2003.

- TAMAKIA, K. y JEFFREYS, A. J. Human tandem repeat sequences in forensic dna typing. *Legal Medicine*, vol. 7(1), 2005.
- VERGNAUD, G. y DENOEUDE, F. Minisatellites: Mutability and genome architecture. *Genome Research*, vol. 10(1), 2000.
- WEISSTEIN, E. W. *Bonferroni correction*. Wolfram Research, Inc., 2004.
- WIGGINTON, J. E., CUTLER, D. J. y ABECASIS, G. R. A note on exact tests of hardy-weinberg equilibrium. *The American Journal of Human Genetics*, vol. 76(5), 2005.
- WITTKÉ-THOMPSON, J. K., PLUZHNIKOV, A. y COX, N. J. Rational inferences about departures from hardy-weinberg equilibrium. *The American Journal of Human Genetics*, vol. 76(6), 2005.
- WYANDT, H. E. y TONK, V. S. *Human Chromosome Variation: Heteromorphism and Polymorphism*. Springer Netherlands, 2012.
- XU, X., PENG, M., FANG, Z. y XU, X. The direction of microsatellite mutations is dependent upon allele length. *Nature Genetics*, vol. 24, 2000.

*–¿Qué te parece desto, Sancho? – Dijo Don Quijote –
Bien podrán los encantadores quitarme la ventura,
pero el esfuerzo y el ánimo, será imposible.*

*Segunda parte del Ingenioso Caballero
Don Quijote de la Mancha
Miguel de Cervantes*

*–Buena está – dijo Sancho –; fírmela vuestra merced.
–No es menester firmarla – dijo Don Quijote–,
sino solamente poner mi rúbrica.*

*Primera parte del Ingenioso Caballero
Don Quijote de la Mancha
Miguel de Cervantes*

