



Diversidad genética en la población mexicana: Utilización de marcadores de ADN

Mariano Guardado-Estrada,* Gloria Queipo,**
Marco Meraz-Ríos,*** Jaime Berumen-Campos****

RESUMEN

A principios de este siglo inicia lo que se conoce como la era genómica, partiendo de la secuenciación del genoma humano. Al obtener secuencias de individuos de distintas regiones geográficas se definirá con mayor precisión la variación que existe dentro de los seres humanos. Conocer estas variaciones en el genoma humano nos permitirá identificar aquellos genes que pudieran conferir susceptibilidad o protección hacia ciertas enfermedades complejas, identificar la interacción de los mismos con los diversos factores ambientales e incluso predecir una respuesta al tratamiento de la enfermedad. La genética de poblaciones es una herramienta útil que ha adquirido mucho peso al estudiar las variaciones genómicas. La población mexicana actual surge principalmente de la mezcla de dos poblaciones que estuvieron separadas por mucho tiempo, la española y la amerindia, posteriormente con una contribución en menor proporción de la población africana. Conocer la variación genética de la población mexicana, así como la identificación de aquellos factores que han ido actuando sobre las frecuencias alélicas permitirá en un futuro establecer si existe alguna relación entre el polimorfismo de nuestro genoma, el medio ambiente y las enfermedades complejas con una mayor prevalencia en nuestra población con miras hacia una medicina individualizada.

Palabras clave: Alelo, haplotipo, haplogrupo, polimorfismo, microsatélite, SNP.

ABSTRACT

At the beginning of the 21st begins what it is known as the genomic era with the complete sequentiation of the human genome. Obtaining sequences from individuals all around the world could precise define the human variation among individuals. The knowledge of this genome variation will help to identify those genes that confer risk or protection for complex disease, identify the interactions between this genes and environmental factors and even predict the treatment response for that disease. Population Genetics is a useful tool and has an important meaning studying the genomic variation. Mexican population arose principally from the admixture of two populations that were separated long time ago, the Spaniards and the Amerindians, afterwards with a less degree of contribution of Africans. With the knowledge of the genetic variation among the Mexican population and those factors that act over the allelic frequencies could establish whether exists a direct relationship among the polymorphism in our genome, our environment and those complex disease with a higher prevalence in our population looking towards an individualized medicine.

Key words: Allele, haplotype, haplogroup, polymorphism, microsatellite, SNP.

www.medigraphic.com

INTRODUCCIÓN

Los estimados de diversidad genética han sido utilizados para tratar de esclarecer el origen de nuestra especie y para encontrar las diferencias presentes entre individuos de distintas poblaciones que nos permitan diferenciar sujetos en el campo de la

* Laboratorio de Medicina Genómica, Hospital General de México (HGM)-Facultad de Medicina (FM), Universidad Nacional Autónoma de México (UNAM).

** Servicio de Genética, HGM-FM, UNAM.

*** Dirección de Planeación, CINVESTAV.

**** Unidad de Medicina Genómica, HGM-FM, UNAM.

medicina forense o tratar de entender la naturaleza de las enfermedades del ser humano.¹ Con la secuenciación completa del genoma humano a principios de este siglo,^{2,3} se incrementa día con día el conocimiento que arrojan los tres billones de pares de bases, lo que en el futuro nos permitirá esclarecer y definir las diferencias que existen entre los grupos de poblaciones en seres humanos; esto último a través de la segunda fase del HapMap.⁴ Sin embargo, antes de la secuenciación del genoma humano ya se habían hecho estimados de la variabilidad genética en el ser humano, primero a través de proteínas y posteriormente con fragmentos de secuencias, siendo esto posible gracias a la genética de poblaciones. Al utilizar marcadores de ADN como microsátélites, elementos Alu y secuencias de ADN mitocondrial, se ha encontrado que la variación genética que existe entre los distintos grupos poblacionales de los seres humanos es aproximadamente de 5 a 15%, el resto de la variación es observada entre individuos que conforman una población.^{5,6}

La población mexicana actual surge en el año 1521 con la llegada de los conquistadores españoles, lo que origina una mezcla de dos razas que geográficamente estuvieron separadas por mucho tiempo, la amerindia y la europea. Nuestra población contemporánea se considera mestiza, y 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.⁷ La composición genética del mexicano es compleja debido a que por factores sociales, demográficos y económicos la distribución alélica no es ecuánime en nuestra población general y cambia de acuerdo a la región geográfica analizada.⁸ Existen zonas marginadas con poblaciones amerindias que han sufrido una reducción notable en el tamaño de la población casi al punto de extinción y poblaciones mayores donde el flujo genético se ha mantenido constante. El componente amerindio es la base de nuestra población junto con el europeo que trajeron consigo los españoles y una pequeña proporción proveniente de África que trajeron los españoles consigo.⁹ En México, los primeros estudios de diversidad genética que se realizaron en poblaciones indígenas como mestizas utilizando marcadores de grupo sanguíneo, haptoglobinas, albúmina y hemoglobina fueron descritos por Lisker, donde se empiezan a observar estas variaciones en las frecuencias alélicas entre los individuos de distintas regiones analizadas.^{10,11}

De proteínas a secuencias de ADN

Para poder entender los estimados de variación es necesario tener una noción acerca de la estadística básica en la genética de poblaciones. Existen patrones genéticos que no obedecen a las leyes mendelianas, por lo que se necesitaba un modelo para explicar las interacciones entre rasgos genéticos y la población perteneciente; es así como surge la genética de poblaciones. La genética de poblaciones no sólo estudia la composición genética del individuo, sino la composición genética de la población entera y cómo cambia ésta a través de las generaciones.¹² Al estudiar la composición genética en una población podemos establecer la variación de distintos locus, tratar de explicar las causas posibles que originaron esa variación e incluso establecer un modelo o simulación para predecir el futuro de esa variación. La genética de poblaciones se aplica a todos los seres vivos, y no son la excepción los seres humanos. Inicialmente se observó que la distribución de los grupos sanguíneos ABO no era homogénea en distintos grupos étnicos, encontrándose un nivel de variación, en estudios subsecuentes también se ha encontrado variación en otro tipo de marcadores.¹³ Un alelo se define como las variaciones posibles de un gen o secuencia de ADN en la población y cada par de alelos en la célula conformará un genotipo. Cada alelo es segregado en forma aleatoria a través de los gametos para formar nuevos genotipos, una vez llevado a cabo el proceso de fecundación. Las células de los organismos eucariontes presentan dos genomas, uno nuclear y el otro mitocondrial. El genoma de los seres humanos consta de 23 pares de cromosomas contenidos dentro del núcleo y un genoma haploide circular localizado en las mitocondrias dentro del cual están 37 genes necesarios para la cadena respiratoria de la célula.¹⁴ Cada locus dentro de un cromosoma se compone de un par de alelos, heredados uno por cada padre. En el caso del ADN mitocondrial y la mayor parte del cromosoma Y se consideran genomas haploides debido a su modo de transmisión, ya que se heredan en bloque exclusivamente por la madre o por el padre, respectivamente, y para fines de estudio se pueden considerar como un solo alelo.

Existen cuatro fuerzas a considerar que pueden cambiar la frecuencia alélica y son la selección, deriva génica, mutación y migración.¹⁵ La selección es la principal fuerza que puede actuar sobre un gen en particular y está relacionado con adaptación y ventaja para el individuo, pero requiere de mucho tiempo para que los efectos sean visibles; las otras fuerzas

requieren menor tiempo, dependen de otros factores como el tamaño de la población, o mejor dicho tamaño efectivo de la población, y de la tasa de mutación. La frecuencia de un genotipo, ya sea homocigoto o heterocigoto, cambia como es de esperarse, al cambiar las frecuencias alélicas. Cuando encontramos un locus frecuentemente homocigoto, la variación de este locus en la población es casi nula; sin embargo, la aparición de nuevos alelos formará nuevos genotipos y aumentará la heterocigosidad en el locus o proporción genotipos heterocigotos. Un alelo nuevo se introduce en la población gracias a la mutación, es decir, cuando existe un evento de replicación de ADN, el resultado son dos copias idénticas al antecesor, aunque existe una probabilidad de 10^{-9} en los seres humanos de presentarse una mutación o error en la replicación, lo que originaría un nuevo alelo.¹⁶ Cuando existen más de dos alelos de un mismo gen, ese gen se considera como polimórfico y cada nuevo alelo que surge en la población aumenta el polimorfismo del gen o secuencia de ADN en estudio. No obstante, para que realmente se considere un gen polimórfico, la frecuencia del alelo más común debe ser de menos del 99% y para que se considere un alelo como común cuando menos debe superar el 0.005% de frecuencia en la población; los alelos que no alcancen dicha frecuencia son considerados como raros y, a menos que no actúe una fuerza sobre ellos, tienden a desaparecer.¹

A principios del siglo XX se establece la Ley del Equilibrio de Hardy-Weinberg y hoy en nuestros días se aplica como una estadística estándar al estudiar un locus diploide.¹⁵ El equilibrio en la ley de Hardy Weinberg ocurre en una población que suponemos se mezcla al azar y no ocurren las fuerzas de migración, deriva génica o selección en una población infinita; el resultado que obtenemos de un locus en equilibrio es que las frecuencias alélicas no cambian de generación en generación,¹⁷ por lo que las frecuencias genotípicas también se mantienen constantes, por lo cual la suma de las frecuencias genotípicas de un locus así como de los alelos siempre debe ser uno. Una desviación del equilibrio de Hardy Weinberg donde exista una variación en las frecuencias alélicas indica que existen y actúan sobre el alelo alguna de las cuatro fuerzas ya mencionadas; pero también nos habla de una población en la que no hay un apareamiento al azar o que la población recientemente acaba de pasar por un cuello de botella que disminuyó de manera drástica el número de individuos en la población.

Sabemos que el tamaño poblacional de los seres humanos no es infinito, más bien está delimitado a un

número. En poblaciones con un número finito de integrantes existe variación en las frecuencias alélicas que no dependen de la selección; más bien está en relación de procesos estocásticos, es decir, está en proporción del tamaño efectivo de la población que se refiere al número de individuos que se aparean en cada generación.^{15,17} Cuando surge un nuevo alelo debido a la mutación, el destino final del mismo será la extinción o fijación como resultado de si el portador del alelo se apareará, perpetuando el alelo en la próxima generación hasta que éste se fije o, caso contrario, si no se aparea. A esta fuerza se le conoce como deriva génica.¹⁸ En el caso de poblaciones pequeñas esta fuerza es muy importante; sin embargo, las poblaciones con un número elevado de individuos se vuelven insensibles a ella.

Inicialmente el polimorfismo se estudiaba a través de la electroforesis de proteínas, pero la resolución en los estimados de heterocigosidad era limitada, detectando sólo un 30% de la variación.¹⁹ Con la técnica para secuenciación por díoxinucleótidos diseñada por Sanger,²⁰ y el desarrollo de métodos cada vez más novedosos, es muy común obtener secuencias de ADN en el campo de la investigación. Mediante la secuenciación de ADN se permite definir mejor los polimorfismos en cualquier genoma. Al analizar una secuencia de ADN y compararla con otras, podemos encontrar que en la mayor parte de los casos existen diferencias entre una secuencia y otra. La teoría neutral de la evolución desarrollada por Kimura establece que la mayor parte de las mutaciones encontradas son silenciosas o neutras, es decir, la mayoría no originan un cambio en la secuencia de aminoácidos; al conocer esta tasa de variación tan baja en las proteínas se han hecho aproximados del tiempo de divergencia entre dos especies como si se tratase de un reloj molecular;²¹ la probabilidad de que una mutación se pueda originar en cualquier sitio del genoma y que la combinación de ellas cree nuevos alelos es el principio del modelo de sitios infinitos.¹² Uno de los estimados de variación bajo este modelo es el cálculo de θ , donde $\theta = 4Nu$, esta fórmula depende del tamaño efectivo de la población (N) y la tasa de mutación (u), por lo que nos da un estimado de variación. Cada cambio de base originado por una mutación se considera como un sitio segregante, al comparar cada una de las secuencias obtenemos la diferencia promedio de nucleótidos por par. Con los sitios segregantes y la heterocigosidad de nuestra muestra podremos calcular θ y π , respectivamente; la diferencia entre estos dos nos sirve para calcular la D de Tajima, con valores en un intervalo entre -2 y 2, que nos da un

aproximado de variaciones demográficas como son cuellos de botella en resultados negativos o crecimientos en la población si son positivos.²²

Utilizando el número de sitios segregantes en varias secuencias de ADN, se pueden realizar estimados de coalescencia, así como el tiempo requerido para la misma.²³ La coalescencia nos permite estimar la genealogía de un grupo de alelos que pudieran diferir entre sí; pero mediante simulaciones calculadas al azar, se puede recorrer generaciones hacia atrás hasta encontrar un antecesor común (MRCA).^{24,25} Cada alelo actual puede originarse mediante dos procesos estocásticos, la mutación y la deriva genética; sin embargo, nuestro modelo de secuencia tiene que cumplir con ciertas características, tienen que ser individuos de preferencia haploides y de preferencia no presentar recombinación, ya que la recombinación *per se* introduce mayor variación al momento de realizar las genealogías, aunque finalmente llegaremos al MCRA.¹⁷

Finalmente, sabemos que la raza humana no se ha mantenido estática, ya que desde su origen, hace aproximadamente 200,000 años en África, ha migrado constantemente para poblar la mayor parte del globo terráqueo. Las migraciones traen consigo un flujo genético, lo que origina que las frecuencias alélicas se mantengan constantes entre las poblaciones e impide los efectos de la deriva génica.²⁶ Un modelo de isla establecido por Wright sirve para ilustrarnos cómo se mantienen las frecuencias alélicas a manera de representación entre una supuesta isla mayor o principal rodeado de pequeñas islas en las cuales existe un determinado número de individuos que se aparean al azar.²⁷ Sin embargo, estas pequeñas islas reciben cierto número de inmigrantes provenientes de la isla mayor, lo cual introduce nuevos alelos a cada una de las islas, hasta que las frecuencias alélicas en las pequeñas sean similares a las de la isla mayor. Este modelo lo podemos aplicar a la raza humana a partir de que, a lo largo de la historia, han existido pequeños grupos colonizadores que se han asentado en distintas partes del mundo y que posteriormente han recibido inmigrantes. Los grupos colonizadores por lo general son reducidos en número, y en estos grupos la frecuencia alélica varía de acuerdo a la deriva génica y es probable que ciertos alelos se fijen en esos grupos, razón por la cual las frecuencias alélicas llegan a ser distintas entre un grupo y otro. El efecto fundador disminuye la heterogeneidad de alelos en la población y la migración la aumenta.²⁸

En la meiosis, las cromátides presentan intercambio de material genético con otras cromátides herma-

nas. Estos eventos de recombinación son comunes en el genoma humano y la tasa de recombinación varía de una región a otra con sitios más propensos a recombinar.²⁹ El proceso de recombinación origina nuevas cromátides e incrementa, junto con la mutación, la diversidad genética, por lo que cambian las frecuencias de los gametos.³⁰ En este momento ya no sólo nos enfocamos a un solo locus, sino a varios que a su vez forman haplotipos. Imaginemos un locus con dos alelos posibles A_1 y A_2 en un cromosoma, en el mismo cromosoma existe otro locus con los alelos B_1 o B_2 . En un cromosoma con el alelo A_1 puede estar con mayor frecuencia ligado al alelo B_1 más que al alelo B_2 , esto se conoce como desequilibrio de ligamiento. Al calcular el desequilibrio de ligamiento podremos identificar si un alelo está ligado a otro en una región cromosómica, lo que depende de su cercanía física;³¹ a menor distancia física entre alelos, menor es la probabilidad de que ocurra un evento de recombinación entre ellos, por lo que se heredan en bloque. Al decir que un alelo está en equilibrio de ligamiento aumenta la probabilidad de encontrar distintos haplotipos por la alta recombinación que existe entre ellos. Si existe selección sobre un alelo en particular, se puede presentar una fuerza conocida como hitchhiking donde la selección actúa sobre las regiones circundantes, lo cual disminuye la heterocigosidad en la región hasta que surja una recombinación que libere a los alelos en cuestión.

Dos historias distintas

El ADN mitocondrial y la región no recombinante del cromosoma Y presentan características peculiares que se adaptan a los modelos de coalescencia y a la teoría neutral de la evolución de una manera relativamente simple.¹⁷ El ADN mitocondrial es un genoma extracromosomal que consta de 16 kb que contiene 37 genes necesarios para la fosforilación oxidativa y una región de control no codificante dividida en dos regiones hipervariables,³² se hereda exclusivamente por vía materna, tiene una tasa de mutación elevada respecto al ADN genómico, es un genoma haploide y no presenta recombinación.³³ En el ADN nuclear existe un cromosoma que tiene características diferentes a las de los demás cromosomas, el cromosoma Y. El cromosoma Y se compone de una región no recombinante (NRY por las siglas en inglés)³⁴ que abarca el 95% del cromosoma y contiene los genes para diferenciación masculina y pseudogenes homólogos a ciertos genes del cromosoma X.³⁵ El cromosoma Y mantiene una herencia patrilineal debido a

que se transmite de padres a hijos exclusivamente. Este modo de herencia y la nula recombinación fue motivación para realizar estudios de diversidad genética con el polimorfismo observado en la región.³⁶ Sin embargo, la región no recombinante de cromosoma Y presenta una tasa de mutación baja comparada con el ADN mitocondrial (1.24×10^{-9} versus 3.7×10^{-3}).^{37,38} Es considerado el cromosoma Y como el cromosoma con el nivel más bajo de polimorfismos, esto se explica por estimados hechos a través de secuenciación que han calculado un polimorfismo de un solo nucleótido por cada 10,000 pares de bases,³⁹ a diferencia de los autosomas donde existe un polimorfismo cada 500 a 1,000 pares de bases aproximadamente.^{40,41} Los estimados de variabilidad tanto en el ADN mitocondrial como en la región NRY del cromosoma Y se han hecho a través de enzimas de restricción, identificando RFLP (del inglés *Restriction Length Fragment Polymorphism*),⁴² secuenciación directa⁴³ y microsatélites.³⁸ Como ya se ha mencionado, a partir de los polimorfismos en secuencias de ADN tomados de individuos no relacionados entre sí, y calculando las distancias entre ellos se puede hacer un estimado de la genealogía del gen en particular hasta llegar a un antecesor común (MRCA) y calcular el tiempo para llegar a ese ancestro común (TMRCA) utilizando la coalescencia.⁴⁴ Al realizar los estimados de coalescencia, por lo general se han comparado los resultados de ambos locus con el objetivo de establecer similitudes en los mismos.⁴⁵ A principios del 2000 se propone el origen del ser humano moderno hace 171,000 años en África al secuenciar el genoma completo del ADN mitocondrial, y pudiendo separar los grupos poblacionales africanos de los no africanos a través de la genealogía del ADN mitocondrial.^{37,46} La teoría causó mucha controversia al punto de rebatirla,^{16,47} aunque estudios posteriores, secuenciando el genoma completo y excluyendo las regiones hipervariables, coinciden en un solo origen del ser humano, situando la raíz profunda de la genealogía en África.^{48,49} Al estudiar la región no recombinante del cromosoma Y³⁴ para definir el origen del hombre moderno, también ha podido ser situado en África. El tiempo de coalescencia estimado con el cromosoma Y suele ser casi la mitad que el del locus mitocondrial con un promedio de 50,000 años, aproximadamente, lo que probablemente se deba a un tamaño menor efectivo en la población.^{44,50} Este tamaño efectivo menor se explica por una varianza mayor en el éxito reproductivo de los varones, selección natural o estocasticidad en el proceso evolutivo.⁵⁰

Aunque el ADN mitocondrial fue el primer genoma en ser secuenciado por completo,³² los primeros estudios de diversidad se hicieron utilizando enzimas de restricción para detectar el polimorfismo de un solo nucleótido en la región D loop, así como en todo el genoma.⁴² Posteriormente se observó que existen polimorfismos, los cuales son compartidos entre individuos de una misma región geográfica.^{51,52} Al agrupar estos polimorfismos del ADN mitocondrial que son compartidos por grupos étnicos, se han identificado los polimorfismos ancestrales que son utilizados para definir haplogrupos poblacionales.^{49,53} Inicialmente, estudios realizados en poblaciones amerindias que incluyen a los mayas encontraron una alta diversidad genética al secuenciar la región D loop del ADN mitocondrial con una tasa de diversidad por nucleótidos desde 0.71% hasta 1.95%, un número elevado de haplotipos únicos que junto con los más comunes se podían clasificar en cuatro grupos principales.^{54,55} Estos datos indican, como más adelante se verá, que existe la presencia de grupos constituidos por haplotipos de individuos pertenecientes a una misma población; a estos grupos nos referiremos como haplogrupos poblacionales. Los haplogrupos poblacionales actualmente se designan por letras del alfabeto.^{37,46,49} Con esta nueva clasificación tenemos que en el Continente Americano encontramos a los haplogrupos A, B, C, D y en menor frecuencia el X. Estos haplogrupos son considerados como los haplogrupos fundadores que trajeron los primeros pobladores al continente.^{56,57} Los polimorfismos que definen a los haplogrupos amerindios están caracterizados mediante enzimas de restricción en el genoma mitocondrial; sin embargo, también se han identificado sitios característicos que definen a estos haplogrupos en la región no codificante del ADN mitocondrial (*Cuadro I*). Datos antropológicos coinciden a favor de que los fundadores amerindios emigraron desde Asia a través del Estrecho de Bering en el pleistoceno, para posteriormente colonizar el resto del Continente Americano.^{13,58} Mediante los mismos

Cuadro I. Polimorfismos en la región D loop del ADN mitocondrial asociados a haplogrupos amerindios.

<i>Polimorfismos</i>	<i>Haplogrupo</i>
16,223:T, 16,290:T, 16,319:A	A
16,189:C, 16,217:C, 16,223:C	B
16,223:T, 16,298:C, 16,327:T	C
16,223:T, 16,325:C, 16,362:C	D

datos antropológicos se sugiere que la entrada de los primeros pobladores amerindios fue hace aproximadamente 15,000 años.¹³ Al analizar la gran diversidad genética que existe en los grupos poblacionales amerindios actuales y con estimados de coalescencia, inicialmente se llegó a calcular la entrada de los primeros pobladores al continente hace 40,000 años aproximadamente.⁵⁹ Sin embargo, cuando se secuenció la región hipervariable I y la región codificante se redujo el tiempo calculado hasta hace 16,000 años.⁶⁰ Estos resultados han variado debido a la región del ADN mitocondrial analizada y al método para detectar la diversidad genética en las poblaciones, además de que no se tiene la certeza de que si los tiempos estimados de coalescencia correspondan al surgimiento de estos haplogrupos en Asia o su posterior dispersión hacia América; finalmente es importante considerar el número de haplotipos fundadores en cada haplogrupo, pues si hay más de un haplotipo fundador, aumentará la diversidad genética inicial y la edad calculada para estos linajes será sobreestimada. Aunque en las distintas tribus amerindias analizadas se encuentran prácticamente los cuatro haplogrupos fundadores, la distribución de los haplogrupos no es uniforme a lo largo de todo el Continente Americano. En las tribus de América Central y México se ha encontrado con mayor frecuencia el haplogrupo A; después en orden descendente tenemos al B, C y en mucho menor frecuencia el D.^{61,62} Como ya se ha mencionado, los estudios de ADN mitocondrial en poblaciones amerindias han sido realizados en tribus con un supuesto componente ancestral amerindio, como lo son los grupos mixtecas o mayas; sin embargo, la población mexicana actual es considerada mestiza, producto del flujo genético amerindio, europeo y, en mucho menor proporción, el componente africano como ya se había mencionado previamente. Pocos son los estudios de ADN mitocondrial que se han hecho en nuestra población mestiza. De acuerdo al modo de herencia del genoma mitocondrial y la historia del mestizaje en nuestro país producto de la mezcla de mujeres amerindias con los conquistadores españoles, se espera que las frecuencias de los haplogrupos amerindios no varíen en nuestra población; estos haplogrupos se mantienen como la base de la población mexicana. La distribución de los haplogrupos amerindios varía muy poco y, por lo general, se mantiene hasta en un 90% en la población general.⁴³ De ese 90% de los haplogrupos amerindios encontrados, el haplogrupo que se mantiene con mayor frecuencia es el A con una distribución que va desde el 33% hasta el 53%; esta variación está rela-

cionada con la proporción de los otros haplogrupos, siendo el segundo haplogrupo más frecuente el B (20.0-33.3%), seguido por los haplogrupos C y D (6.7-23.3% y 2.9-5.8%, respectivamente). El haplogrupo D puede estar ausente en la población mexicana. Los haplogrupos no amerindios cuyo porcentaje es menor al 10%, corresponden al flujo genético materno inmigrante que ha recibido la población mexicana.^{63,64} Una de las fuentes de este flujo genético corresponde a Europa, probablemente de España. De los haplogrupos europeos que se encuentran en la población mexicana, el H es el más frecuente, cuya frecuencia en Europa es del 40%.⁶⁵ También podemos encontrar otros haplogrupos europeos como el K, J, V, U. El segundo continente que proporciona haplogrupos es el africano, representado por los haplogrupos L1 y L2, cuya frecuencia en la región Subsahara de África es de 70% hasta 100%.⁶⁵

En un principio se sabía poco o no se habían estudiado a profundidad los polimorfismos presentes en la región NRY del cromosoma Y, no fue hasta que en 1997 se introduce la técnica de cromatografía líquida desnaturante de alta resolución y se han identificado 200 polimorfismos que comprenden SNPs, inserciones Alu y deleciones con un bajo porcentaje de mutaciones paralelas o reversas.⁶⁶ La diversidad genética de la región no recombinante del cromosoma Y es también clasificada en haplogrupos, a través de polimorfismos binarios que se encuentran en las regiones SMCY, DFFRY, UTY y DBY del NRY.⁶⁷ Sin embargo, al identificar globalmente los polimorfismos, existieron distintas nomenclaturas establecidas por varios autores para definir a los haplogrupos.⁶⁸⁻⁷⁰ Para tener una nomenclatura uniforme de los haplogrupos, basándose en los marcadores genéticos binarios, se estableció un consorcio internacional del cromosoma Y.⁷¹ Un haplogrupo se compone de polimorfismos binarios en un SNPs, se llama binario debido a que es un polimorfismo bialélico, y se designa con el prefijo M; al realizar la filogenia cada haplogrupo pertenece junto con otros a grupos designados por letras del alfabeto que van de la letra A a la R, por ejemplo el haplogrupo G-M201; un haplotipo lo compone los polimorfismos binarios del haplogrupo al que pertenece y el polimorfismo que presenta en un STR. El cromosoma Y en la población mexicana la podemos dividir en una etapa previa a la colonización, donde encontramos exclusivamente el haplogrupo amerindios y una etapa después de la llegada de los españoles, donde se introduce a la población el haplogrupo europeo que predomina en España.

Existen cinco haplogrupos reportados que se encuentran en las poblaciones amerindias Q-M3 o DSY199T*, R1a1-M17, P-M45, F-M89, y C-M130 de acuerdo a la clasificación actual.⁷² Sin embargo, sólo los haplogrupos Q-M3 y P-M45 son los más frecuentes en estos grupos poblacionales con una frecuencia mayor al 5% (~70% y ~30%, respectivamente).^{73,74} Datos sobre el origen de los distintos haplogrupos encontrados en amerindios favorecen la hipótesis de una sola migración de Asia al Continente Americano.^{60,75} El tiempo de entrada de los primeros pobladores al Continente Americano se espera menor al del ADN mitocondrial por las características previamente descritas; sin embargo, utilizando la variación de los STRs para calcular la tasa de mutación, se ha estimado la entrada de los primeros pobladores a través del haplogrupo Q-M3 de 10,100 a 17,200 años.⁷⁵ El polimorfismo DSY199T* y como ya se ha mencionado es el haplogrupo más frecuente en el Continente Americano; en las poblaciones mixteca, zapoteca, maya y mixe lo encontramos con una frecuencia hasta del 85%.^{76,77} Como es de esperarse, sobreviene una disminución en la proporción de este alelo en la población mexicana tras la llegada de los españoles, producto de la reducción de los amerindios. En la población mestiza mexicana que encontramos principalmente en las grandes ciudades, la proporción de individuos con el alelo DSY199T es de 26.9%; la proporción de este alelo aumentará significativamente en poblaciones rurales donde se presume que existe un mayor número de individuos de origen amerindio de hasta del 60.4%. Utilizando la fórmula de Bernstein para calcular la contribución paterna europea a nuestra población se estima de 67% en población urbana y en la población rural de 14.1%. Esta diferencia en la contribución europea del alelo DSY199T* por región geográfica en nuestro país, indica que el mestizaje no fue uniforme en el país. Cuando llegaron los conquistadores españoles, se conoce que las uniones tanto reconocidas como no reconocidas comúnmente fueron entre varones españoles y mujeres indígenas.^{63,78,79} Los datos moleculares de marcadores mitocondriales y del cromosoma Y concuerdan con el mestizaje que ha ocurrido en nuestra historia, encontrándose una mayor proporción de haplogrupos mitocondriales amerindios en la población mexicana y una mayor proporción del marcador europeo M170 en cromosoma Y en los varones. Bonilla y colaboradores comentan que las diferencias en el mestizaje observadas por un aumento en la proporción del alelo amerindio DYS199T pudiera ser resultado de las variaciones que ocurrieron en el mestizaje

en las haciendas en el siglo XVII, cuando la corona española otorga el título de cacique a algunos indígenas que pertenecían a la nobleza en el periodo precolombino, quienes a su vez tienen bajo su control a grupos de indígenas en sus tierras, siendo el entorno de la «hacienda» donde se llevó a cabo la mezcla entre españoles e indígenas en aquella época.

Otros marcadores de ADN

Cuando en el 2001 se completa el primer borrador del genoma humano y se dio a la tarea de la búsqueda e identificación de genes en esos tres billones de pares de bases, sorprende el resultado en el que se logra identificar no más de 30,000 genes aproximadamente. La región codificante abarca sólo el 1.5% del genoma humano, y el resto del genoma no codificante en un principio se consideró como ADN basura sin función alguna.^{80,81} Sin embargo, ese ADN «basura» o «junkie» es el que ha sido recientemente utilizado para realizar estudios de evolución y diversidad genética, en el que podemos adaptar modelos de sitios infinitos.^{82,83} Gran parte de este ADN no codificante está compuesto por secuencias repetitivas que, dependiendo del número de repetidos, se conocen como satélites cuya longitud llega a ser de megabases, minisatélites de 0.1 a 20 kilobases o microsatélites es menor a los 100 pares de bases.⁸⁴

Los microsatélites o secuencia de repetidos simples usualmente son secuencias menores de 10 pares de bases, que se encuentran distribuidas en todo el genoma humano en regiones ricas de heterocromatina, y representan el 2% del genoma.⁸⁵ Es frecuente encontrar repetidos de di, tri o tetranucleótidos. Los microsatélites de ADN surgen como errores en la replicación de ADN donde la polimerasa, al pasar por estas regiones en el templado de ADN, sufre de un resbalamiento donde incorpora de más o deja fuera alguno de estos repetidos y otro mecanismo propuesto es el de la recombinación meiótica que deja secuencias de repetidos dispares. Aunque los STRs los encontramos en regiones no codificantes, pueden abarcar regiones codificantes de genes. Existe un debate controversial acerca de la función de STRs localizados en exones, pero se sabe que cuando en estos repetidos ocurren cambios en la extensión puede originar proteínas anómalas que se traduzcan en enfermedades neurodegenerativas como la enfermedad de Huntington o la ataxia espinocerebelar. También se sabe que existen tumores que presentan una inestabilidad en los microsatélites como los tumores gástricos, colon, endometrio o pulmón.^{86,87} Este tipo

de tumores presenta mutaciones en el sistema de reparación de ADN, el cual va a afectar al momento de replicación el largo de las secuencias de repetidos; las cuales, al estar en regiones codificantes, originarán proteínas alteradas con capacidades supresoras de tumores. Los repetidos simples de ADN, como ya se ha mencionado, 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, la heterocigosidad observada en la muestra se puede adecuar a un modelo de alelos infinitos, partiendo de la teoría neutral. Uno de los usos principales que se le da en la actualidad a la genotipificación de estos alelos es para la identificación de individuos.⁸⁸ La gran variabilidad de estos repetidos de nucleótidos entre un individuo y otro permite diferenciarlos completamente entre sí, por lo que se han utilizado ampliamente en el campo de la medicina forense, estableciendo una huella digital de ADN de cada individuo. Se han hecho estudios de diversidad genética en la población mexicana mestiza como 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.⁸ Al estudiar los alelos previamente mencionados en poblaciones indígenas tarahumaras, huicholes y purépechas se encontró una disminución en la heterocigosidad y un número menor de alelos por locus a diferencia de poblaciones mestizas, resultado de un aislamiento geográfico y la deriva génica.⁸⁹ En poblaciones urbanas como la Ciudad de México se han incluido los siguientes STRs D3S1358, D8S1179, D21S11, D18S51, D5S818, D13S317 y D7S820, donde existe una diferencia en la frecuencia alélica respecto a otras regiones del país;⁹⁰ al ampliar el estudio a una región más extensa de la zona central del país, como es la Ciudad de México con zonas aledañas, se mantiene la diferencia en las frecuencias alélicas respecto al norte del país,⁹¹ siendo más marcada en los alelos TH01 y D3S1358; diferencias que pueden estar relacionadas con el tipo de mestizaje que varía en distintas regiones del país;⁹² la misma tendencia se mantiene al comparar varios estados del centro de México.^{93,94} Un locus que también puede proporcionar datos tanto de mestizaje como de origen geográfico es el STR DXYS15695. Al analizar las diferencias en las frecuencias alélicas, utilizando los STRs arrojan datos

que indica un mestizaje no uniforme a lo largo del país con un componente caucásico mayor en las regiones del norte de México y el mestizaje amerindio en la zona sur y centro del país. Los STRs también se correlacionan con los eventos demográficos en poblaciones indígenas como ha sido un aislamiento geográfico y una reducción en el tamaño de la población, lo que se traduce en una disminución de heterocigotos.

Los polimorfismos de un simple nucleótido

El polimorfismo más frecuente y de menor tamaño que encontramos en el genoma humano es el SNP (del inglés *Single Nucleotide Polymorphism* y se pronuncia SNIp), cada SNP se considera como un polimorfismo bialélico. Se estima que existe un SNP por cada 1,000 pares de bases en el genoma humano y que existen probablemente más de 10 millones de SNPs en común en la población general, lo que constituiría el 90% de variación genética en la población. Una vez cumplida la meta del Proyecto del Genoma Humano de secuenciar el genoma humano por completo, se lograron identificar 1.42 millones de SNPs,³⁹ por lo que se propone un consorcio internacional, mejor conocido como HapMap, para identificar nuevos SNPs y las variaciones comunes de estos SNPs en distintas poblaciones de las cuales inicialmente se incluyeron poblaciones de origen caucásico, chino, y africano.⁹⁶ Otra meta del HapMap es el establecimiento de una base de datos de dominio público la cual constantemente se irá enriqueciendo añadiendo nuevos polimorfismos, incluyendo nuevos grupos poblacionales o étnicos (www.hapmap.org). También podemos encontrar otras bases de datos como la del NCBI (*Nacional Center of Biotechnology Information*)⁹⁷ o la JSNP (*Japanese Single Nucleotide Polymorphism database*).⁹⁸ En México se están realizando trabajos que permitan identificar SNPs en grupos mestizos, para poder desarrollar un haplotipo que defina a la población mexicana actual.⁹⁹ Recientemente se han desarrollado plataformas novedosas que permiten genotipificar los SNPs a gran escala de todo el genoma humano,¹⁰⁰ por lo que en un solo experimento se pueden genotipificar hasta 500,000 SNPs, reduciendo relativamente el costo y el tiempo en comparación con otras tecnologías para genotipificar individuos.¹⁰¹ Aunque en México ya se cuenta con la tecnología en varios grupos de investigación, siendo uno de ellos la Facultad de Medicina de la UNAM en conjunto con el Hospital General de México, los costos de genotipificación mediante estas

plataformas siguen siendo elevados. Debido al creciente número de proyectos para identificar estos marcadores; surge lo que se puede considerar como el sucesor del proyecto del Genoma Humano: el Proyecto del Varioma Humano, donde se intenta agrupar y recolectar a manera de catálogo las variaciones comunes en la población y los fenotipos que en ella representan, de una manera organizada y universal en la que cualquier grupo pueda aportar sus hallazgos y compartirlos con la comunidad científica.¹⁰²

Los SNPs tienen ciertas peculiaridades, por lo que recientemente se ha hecho un gran esfuerzo para la identificación y genotipificación de estos marcadores. Al tener una mayor resolución de estos marcadores, se puede revelar de una forma completa la variación que existe entre los seres humanos y distintas poblaciones; sobre todo identificar las variantes raras de SNPs cuya frecuencia alélica se encuentra dentro del intervalo del 0.5-5%.¹⁰³ A su vez, cada grupo de SNPs en un individuo forma haplotipos en fase, es decir un bloque de SNPs que es heredado por el padre y otro por la madre. La frecuencia de estos haplotipos se espera que varíe también entre los seres humanos, y es aquí donde entra en juego la recombinación y el desequilibrio de ligamiento. Viendo el genoma humano desde un punto de vista abstracto, éste se hereda como un conjunto de bloques y la presencia de cada bloque en el conjunto depende de la recombinación que pueda o no suceder en el genoma. La identificación de SNPs a gran escala en varios individuos permite calcular el desequilibrio de ligamiento D' y r^2 , este último es un coeficiente al cuadrado de correlación entre dos marcadores; se prefiere utilizar r^2 por tener mayor poder estadístico en los marcadores, sobre todo en los estudios de asociación que la simple D' .¹⁰⁴ Una de las metas ya completadas del HapMap fue identificar el mayor número de SNPs en distintas poblaciones con el objetivo de calcular el desequilibrio de ligamiento entre cada SNP en cuatro poblaciones (Yoruba, Han, Japonesa y Utha) y poder identificar haplotipos.³¹ El desequilibrio de ligamiento está sujeto a fuerzas como la mutación y la selección, y existen zonas específicas en los cromosomas más propensas a la recombinación;¹⁰⁵ además de que el LD no es uniforme a lo largo del genoma en distintos grupos poblacionales, por lo que al comparar individuos de origen africano existen regiones de tamaño menor sin la presencia de desequilibrio de ligamiento a diferencia de grupos caucásicos cuya distancia aproximada para cada sitio sujeto a recombinación es hasta de 60kb, manteniendo mayor número de haplotipos comunes.¹⁰⁶

Esta diferencia puede ser debida a recientes cuellos de botella que se presentaron después de la migración del hombre moderno fuera de África. Los estudios genéticos han permitido develar enfermedades con patrón de herencia mendeliana. Sin embargo, la naturaleza de las enfermedades comunes es más complejo y, aunque el factor ambiental juega un papel importante, se cree que existen alelos en la población que en cierta proporción pudiera predisponer a este tipo de enfermedades. La distribución de los padecimientos comunes no es uniforme en las poblaciones con distinto origen; por ejemplo, en individuos de origen hispano que actualmente residen en los Estados Unidos son frecuentes las enfermedades infecciosas, hepáticas y la diabetes, a diferencia de las personas de origen caucásico entre quienes predominan las enfermedades cardíacas.¹⁰⁷ Este fenómeno sugiere que pudiesen existir alelos en ciertas poblaciones cuyo efecto aditivo predispongamos a este tipo de enfermedades. Para lo que se han desarrollado dos teorías opuestas entre sí. La primera de ellas trata acerca de que existen variantes de alelos comunes que probablemente se originaron de forma ancestral previa a la dispersión del hombre moderno en África y cuyo efecto se traduce en la predisposición para las enfermedades comunes.^{108,109} Esta teoría es llamada variantes comunes/enfermedades comunes y trata de explicar la prevalencia de enfermedades comunes en sujetos no relacionados entre sí que comparten estas variantes. La teoría de múltiples variables raras o hipótesis de heterogeneidad es opuesta en extremo a la teoría de variables comunes, donde se postula que la predisposición a las enfermedades es el resultado de la combinación de estas variantes raras en distintos individuos y cuya frecuencia en la población es de menos del 0.01.¹⁰⁹ Ambos modelos pueden correlacionarse con varias enfermedades y está en debate si una sola teoría pudiera explicar la presencia de ellas, por lo que la tendencia actual es seguir un modelo arbitrario donde se puedan combinar tanto la teoría de variantes comunes como la de variantes raras. Para poder identificar genes candidatos que predisponga a una enfermedad en particular y con la genotipificación de SNPs a gran escala se están realizando actualmente estudios de asociación de SNPs causales para varias enfermedades en grupos de casos y controles; con el desequilibrio de ligamiento entre cada SNP permite identificar o marcar los SNPs que representan las variaciones más comunes, lo que se conoce como SNP tagging.^{104,110} Una vez identificada la posición del SNP y si está en desequilibrio de ligamiento con otros

marcadores no conocidos, se puede hacer una relación directa con el fenotipo en el grupo de casos.

Existen dificultades al momento de diseñar y realizar este tipo de estudios. Conforme avanza la tecnología para la búsqueda de SNPs y su asociación a enfermedades, los métodos estadísticos se vuelven más complejos y surgen nuevos problemas para darle validez al estudio.¹¹¹ Uno de los problemas es que el número de muestras requeridas para hacer un estudio de casos y controles deba ser exageradamente grande, por ejemplo 1,000 individuos por grupo para obtener frecuencias alélicas diferenciales mayores y que la razón de momios sea por lo menos del 1.3, esto sin tomar en cuenta los costos del estudio. Como ya se ha mencionado, las frecuencias alélicas de los SNPs no son uniformes en las distintas poblaciones, por lo que se pueden originar en este tipo de estudios errores de tipo I, si no se toma en cuenta el origen poblacional de cada grupo de individuos, a esto se le conoce como estructura de la población.¹⁰⁴ Finalmente, en las poblaciones donde existe mezcla de individuos, como la mexicana, aumenta el desequilibrio de ligamiento cuando se presenta el mestizaje, por lo que los cromosomas constan de bloques de haplotipos que son heredados por los componentes ancestrales, a mayor tiempo transcurrido desde el mestizaje menor es el tamaño de estos bloques ancestrales debido a la recombinación. Un abordaje que permite rastrear estos alelos ancestrales en este tipo de poblaciones es el mapeo de mezcla o del inglés *admixture mapping*.¹¹² El objetivo es identificar la presencia de alelos comunes entre los individuos y correlacionarlos con la ancestría de la población. El mapeo de mezcla suena promisorio para identificar alelos de susceptibilidad a enfermedades en poblaciones mestizas, en México se han utilizado estos marcadores ancestrales para identificar factores de riesgo en diabetes mellitus tipo 2.⁶³

La genética de poblaciones se mantiene actualmente vigente y quizá con mayor fuerza debido a los avances agigantados que ha tenido la ciencia en la secuenciación e identificación de genes y regiones no codificantes en los seres humanos. Los estudios realizados mediante la genética de poblaciones han sido y seguirán siendo un instrumento útil en distintos campos. A través de los marcadores encontrados en el ADN y su variación, podemos dilucidar la historia del ser humano y la dinámica que han presentado nuestras poblaciones. Los datos que han arrojado los distintos locus analizados explican o refuerzan la historia de la población mexicana desde el origen de los primeros pobladores en el continente y posterior a la

llegada de los españoles que culminan con el surgimiento de la población mexicana actual. Con los nuevos avances tecnológicos en búsqueda de variación genética y entendiendo las fuerzas que la originan, en un futuro lejano nos permitirá comprender mejor las interacciones que presenta el genoma del ser humano con el medio ambiente y de cierta forma tratar de predecir de qué manera afectan a la salud del ser humano. Se prevé que en un futuro no muy lejano la medicina como la conocemos hoy cambie radicalmente gracias a los estudios genómicos

BIBLIOGRAFÍA

- Hartl DL. Principles of population genetics. 3rd ed. Sinauer Associates; 1997.
- Consortium. TIHGM. A physical map of the human genome. *Nature* 2001; 409 (6822): 934-941.
- Venter JC, Adams MD, Myers EW et al. The sequence of the human genome. *Science* 2001; 291 (5507): 1304-1351.
- Abecasis G, Bustamante CD, Ostrander EA, Scherer SW, Chanock SJ, Kwok PY, Brookes AJ. Human Genome Variation 2006: Emerging views on structural variation and large-scale SNP analysis. *Nature Genetics* 2007; 39 (2): 153-155.
- Jorde LB, Watkins WS, Bamshad MJ et al. The distribution of human genetic diversity: A comparison of mitochondrial, autosomal, and y-chromosome data. *Am J Hum Genet* 2000; 66: 979-988.
- Jorde LB, Wooding SP. Genetic variation, classification and 'race'. *Nat Genet* 2004.
- Gorodezky C AC, Vazquez-Garcia MN, de la Rosa G, Infante E, Balladares S, Toribio R, Perez-Luque E, Muñoz L. The genetic structure of Mexican Mestizos of different locations: Tracking back their origins through MHC genes, blood group systems, and microsatellites. *Human Immunology* 2001; 62 (9): 979-991.
- Rangel-Villalobos, Rivas, Torres-Rodriguez et al. Allele frequency distributions of six Amp-FLPS (D1S80, APO-B, VWA, TH01, CSF1PO and HPRTB) in a Mexican population. *Forens Sci Int* 1999; 105: 125-129.
- Beltran A. The slave trade in Mexico. *Hispanic Am Hist Rev* 1944; 24: 412-431.
- Lisker R, Ramirez E, Babinsky. V. Genetic structure of autochthonous populations of Meso-America: Mexico. *Hum Biol* 1996; 68 (3): 395-404.
- Lisker R. Estructura genética de la población mexicana. Mexico City; 1981.
- Gillespie J. Population genetics: A Concise Guide. 2nd ed. Press TJHU, 2004.
- Cavalli-Sforza L et al. The history and geography of human genes: Princeton University Press; 1996.
- Strachan T, Read AP. Human Molecular Genetics. 3rd ed. Garland; 2004.
- Templeton AR. Population genetics and microevolutionary theory; 2006.
- Templeton. Haplotype trees and modern human origins. *Am J Phys Anthropol* 2005; suppl 41: 33-59.
- Rosenberg N, Nordborg M. Genealogical trees, coalescent theory and the analysis of genetic polymorphism. *Nature Rev Gen* 2002; 3: 380-390.

18. Weir BS. Statistical analysis of DNA sequences. *J Natl Cancer Inst* 1988; 80 (6): 395-406.
19. Kimura M, Ohta T. Protein polymorphism as a phase of molecular evolution. *Nature* 1971; 229 (5285): 467-469.
20. Sanger F, Nicklen S, Coulson AR. DNA Sequencing with Chain-Terminating Inhibitors. *PNAS* 1977; 74 (12): 5463-5467.
21. Kimura M. The rate of molecular evolution considered from the standpoint of population genetics. *PNAS* 1969; 63 (4): 1181-1188.
22. Tajima F. Statistical analysis of DNA polymorphism. *Jap J Genet* 1993; 68 (6): 567-95.
23. Innan H, Zhang K, Marjoram P, Tavaré S, Rosenberg NA. Statistical tests of the coalescent model based on the haplotype frequency distribution and the number of segregating sites. *Genetics* 2005; 169 (3): 1763-1777.
24. Rosenberg NA, Hirsh AE. On the use of star-shaped genealogies in inference of coalescence times. *Genetics* 2003; 164 (4): 1677-1682.
25. Degnan JH, Rosenberg NA. Discordance of species trees with their most likely gene trees. *PLoS Genetics* 2006; 2 (5): e68.
26. Hey J, Machado CA. The study of structured populations - new hope for a difficult and divided science. *Nature Rev Genetics* 2003; 4 (7): 535-543.
27. Wright S. Evolution in mendelian populations. *Genetics* 1931; 16 (2): 97-159.
28. Theodorou K, Couvet D. Genetic load in subdivided populations: Interactions between the migration rate, the size and the number of subpopulations. *Heredity* 2005; 96 (1): 69-78.
29. Kong A, Gudbjartsson DF, Sainz J et al. A high-resolution recombination map of the human genome. *Nat Genet* 2002; 31 (3): 241-247.
30. Nachman MW, Bauer VL, Crowell SL, Aquadro CF. DNA variability and recombination rates at X-linked loci in humans. *Genetics* 1998; 150 (3): 1133-1141.
31. Dawson E, Abecasis GR, Bumpstead S et al. A first-generation linkage disequilibrium map of human chromosome 22. *Nature* 2002; 418 (6897): 544-548.
32. Anderson S, Bankier AT, Barrell BG et al. Sequence and organization of the human mitochondrial genome. *Nature* 1981; 290 (5806): 457-465.
33. Taanman J. The mitochondrial genome: Structure, transcription, translation and replication. *Biochim Biophys Acta* 1999; 1410 (2): 103-123.
34. Harpending H, Rogers A. Genetic perspectives on human origins and differentiation. In; 2000: 361-385.
35. Lahn BT, Page DC. Functional coherence of the human Y chromosome. *Science* 1997; 278 (5338): 675-680.
36. Casanova M, Leroy P, Boucekkine C et al. A human Y-linked DNA polymorphism and its potential for estimating genetic and evolutionary distance. *Science* 1985; 230 (4732): 1403-1406.
37. Ingman MKM, Paabo S, Gyllensten U. Mitochondrial genome variation and the origin of modern humans. *Nature* 2000; 408 (7): 708-713.
38. Thomson R, Pritchard JK, Shen P, Oefner PJ, Feldman MW. Recent common ancestry of human Y chromosomes: Evidence from DNA sequence data. *Proc Natl Acad Sci USA* 2000; 97 (13): 7360-7365.
39. Group TISMW. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 2001; 409 (6822): 928-933.
40. Hammer MF, Blackmer F, Garrigan D, Nachman MW, Wilder JA. Human population structure and its effects on sampling Y chromosome sequence variation. *Genetics* 2003; 164 (4): 1495-1509.
41. Whitfield L, Sulston J, Goodfellow P. Sequence variation of the human Y chromosome. *Nature* 1995; 378 (6555): 379-380.
42. Brown W. Polymorphism in mitochondrial DNA of humans as revealed by restriction endonuclease analysis. *Proc Natl Acad Sci USA* 1980; 77 (6): 3605-3609.
43. Green LD DJ, Knight A. mtDNA affinities of the peoples of North-Central Mexico. *Am J Hum Genet* 2000; 66: 989-998.
44. Garrigan D HM. Reconstructing human origins in the genomic era. *Nature Genetics* 2006; 7: 669-680.
45. Wilkins J. Unraveling male and female histories from human genetic data. *Curr Op Genet Dev* 2006; 16: 611-617.
46. Cann RL SM, Wilson AC. Mitochondrial DNA and human evolution. 1987; 325 (6099): 31-36.
47. Templeton. Out of Africa? What do genes tell us? *Curr Opin Genet Dev* 1997 7 (6): 841-847.
48. Ingman M, Gyllensten U. Analysis of the complete human mtDNA genome: Methodology and inferences for human evolution. In; 2001: 454-461.
49. Maca-Meyer N, Gonzalez A, Larruga J, Flores C, Cabrera V. Major genomic mitochondrial lineages delineate early human expansions. *BMC Genetics* 2001; 2 (1): 13.
50. Bertranpetit J. Genome, diversity, and origins: The Y chromosome as a storyteller. *PNAS* 2000; 97 (13): 6927-6929.
51. Aquadro CF, Greenberg BD. Human mitochondrial DNA variation and evolution: Analysis of nucleotide sequences from seven individuals. *Genetics* 1983; 103 (2): 287-312.
52. Cann R, Wilson A. *Genetics* 1983; 109: 699-711
53. Maca-Meyer N CV, Arnay M, Flores C, Fregel R, Gonzalez AM, Larruga JM. Mitochondrial DNA diversity in 17th-18th century remains from Tenerife (Canary Islands). *Am J Phys Anthropol* 2005; 127 (4): 418-426.
54. Horai SKR, Nakagawa-Hattori Y, Hayashi S, Sonoda S, Tajima K. Peopling of the Americas, founded by four major lineages of mitochondrial DNA. *Mol Biol Evol* 1993; 10 (1): 23-47.
55. Ward RH, Frazier BL, Dew-Jager K, Paabo S. Extensive mitochondrial diversity within a single Amerindian Tribe. *Proc Natl Acad Sci USA* 1991; 88 (19): 8720-8724.
56. Bonatto SL SF. Diversity and age of the four major mtDNA haplogroups, and their implications for the peopling of the New World. *Am J Hum Genet* 1997; 61 (6): 1413-1423.
57. Bonatto SL, Salzano FM. A single and early migration for the peopling of the Americas supported by mitochondrial DNA sequence data. *Proc Natl Acad Sci USA* 1997; 94 (5): 1866-1871.
58. Mulligan CJ, Hunley K, Cole S, Long JC. Population genetics, history, and health patterns in native Americans. In; 2004: 295-315.
59. Torroni A, Schurr TG, Yang CC et al. Native American mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were founded by two independent migrations. *Genetics* 1992; 130 (1): 153-162.
60. Schurr TG BS, Gan YY, Hodge JA, Merriwether DA, Lawrence DN, Knowler WC, Weiss KM, Wallace DC. Amerindian mitochondrial DNAs have rare Asian mutations at high frequencies, suggesting they derived from four primary maternal lineages. *Am J Hum Genet* 1990; 46 (3): 613-623.
61. Salzano FM. Molecular variability in Amerindians: Widespread but uneven information. *An Acad Bras Ciênc* 2002; 74: 223-263.

62. Peñalosa-Espinosa. R, Arenas-Aranda D, Cerda-Flores R et al. Characterization of mtDNA haplogroups in 14 Mexican indigenous populations. *Hum Biol* 2007; 79 (3): 313-320.
63. Bonilla C GG, Parra EJ, Kline C, Shriver MD. Admixture analysis of a rural population of the state of Guerrero, Mexico. *Am J Phys Anthropol* 2005; 128 (4): 861-869.
64. Martinez-Marignac VL, Cameron E, Chan A, Perera A, Globus-Goldberg R, Wachter N, Kumate J, McKeigue P, O'donnell D, Shriver MD, Cruz M, Parra EJ. Admixture in Mexico City: Implications for admixture mapping of type 2 diabetes genetic risk factors. *Hum Genet* 2007; 120 (6): 807-819.
65. Torroni A, Huoponen K, Francalacci P et al. Classification of European mtDNAs from an analysis of three European populations. *Genetics* 1996; 144 (4): 1835-1850.
66. Underhill PA, Jin L, Lin AA et al. Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. *Genome Res* 1997; 7 (10): 996-1005.
67. Underhill PA, Passarino G, Lin AA et al. The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Ann Hum Genet* 2001; 65 (1): 43-62.
68. Hammer MF, Karafet TM, Redd AJ et al. Hierarchical patterns of global human Y-chromosome diversity. *Mol Biol Evol* 2001; 18 (7): 1189-1203.
69. Karafet T, Xu L, Du R et al. Paternal population history of East Asia: Sources, patterns, and microevolutionary processes. *Am J Hum Genet* 2001 69 (3): 615-28.
70. Semino O, Passarino G, Oefner PJ et al. The Genetic legacy of paleolithic *Homo sapiens* in extant Europeans: A Y chromosome perspective. *Science* 2000; 290 (5494): 1155-1159.
71. Consortium TYC. A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res* 2002; 12 (2): 339-348.
72. Bianchi NO, Catanesi CI, Baillet G et al. Characterization of ancestral and derived Y-chromosome haplotypes of New World native populations. *Am J Hum Genet* 1998 63 (6): 1862-1871.
73. Lell JT, Sukernik RI, Starikovskaya YB et al. The dual origin and Siberian affinities of Native American Y chromosomes. *Am J Hum Genet* 2002; 70 (1): 192-206.
74. Vallinoto AC et al. Heterogeneity of Y chromosome markers among Brazilian Amerindians. *Am J Hum Biol* 1999; 11 (4): 481-487.
75. Zegura SL, Karafet TM, Zhivotovsky LA, Hammer MF. High-Resolution SNPs and microsatellite haplotypes point to a single, recent entry of Native American Y chromosomes into the Americas. *Mol Biol Evol* 2004; 21 (1): 164-175.
76. Karafet T, Zegura SL, Vuturo-Brady J et al. Y chromosome markers and Trans-Bering Strait dispersals. *Am J Phys Anthropol* 1997 102 (3): 301-314.
77. Lell JT, Brown MD, Schurr TG et al. Y chromosome polymorphisms in native American and Siberian populations: Identification of native American Y chromosome haplotypes. *Hum Genet* 1997 100 (5-6): 536-543.
78. Vieira AR, Karras JC, Orioli IM, Castilla EE, Murray JC. Genetic origins in a South American clefting population. *Clin Genet* 2002; 62 (6): 458-463.
79. Martinez Marignac VL, Bertoni B, Parra EJ, No B. Characterization of admixture in an urban sample from Buenos Aires, Argentina, using uniparentally and biparentally inherited genetic markers. *Hum Biol* 2004 76 (4): 543-557.
80. Doolittle. W, Sapienza. C. Selfish genes, the phenotype paradigm and genome evolution. *Nature* 1980; 284. (5757): 601-603.
81. Orgel L, Crick F. Selfish DNA: The ultimate parasite. *Nature* 1980 284 (5757): 604-607.
82. Kimmel M, Chakraborty R, Stivers DN, Deka R. Dynamics of repeat polymorphisms under a forward-backward mutation model: Within- and between-population variability at microsatellite loci. *Genetics* 1996; 143 (1): 549-555.
83. Pluzhnikov A, Di Rienzo A, Hudson RR. Inferences about human demography based on multilocus analyses of non-coding sequences. *Genetics* 2002; 161 (3): 1209-1218.
84. Li Y-C, Korol AB, Fahima T, Nevo E. Microsatellites within genes: Structure, function, and evolution. *Mol Biol Evol* 2004; 21 (6): 991-1007.
85. Li Y-C, Korol AB, Fahima T, Beiles A, Nevo E. Microsatellites: Genomic distribution, putative functions and mutational mechanisms: A review. In; 2002: 2453-2465.
86. Miozzo M, Sozzi G, Musso K et al. Microsatellite alterations in bronchial and sputum specimens of lung cancer patients. *Cancer Res* 1996; 56 (10): 2285-2288.
87. Young J, Leggett B, Gustafson C et al. Genomic instability occurs in colorectal carcinomas but not in adenomas. *Hum Mutat* 1993; 2 (5): 351-354.
88. Smouse PE, Chevillon C. Analytical aspects of population-specific DNA fingerprinting for individuals. *J Hered* 1998; 89 (2): 143-150.
89. Rangel-Villalobos H, Rivas F, Sandoval L et al. Genetic variation among four Mexican populations (Huichol, Purepecha, Tarahumara, and Mestizo) revealed by two VNTRs and four STRs. *Hum Biol* 2000; 72 (6): 983-1005.
90. Luna-Vazquez A, Vilchis-Dorantes G, Paez-Riberos LA, Muñoz-Valle F, Gonzalez-Martin A, Rangel-Villalobos H. Population data of nine STRs of Mexican-mestizos from Mexico City. *Forensic Sci Internat* 2003; 136 (1-3): 96-98.
91. Gutierrez A, Moguel T, Leon J, Cuellar N, Rangel V. Allele and haplotype distribution for 16 Y-STRs (AmpFISTR® Y-filerT kit) in the state of Chihuahua at North Center of Mexico. *Legal Medicine* 2007; 9 (3): 154-157.
92. Luna-Vazquez. A, G. Vilchis-Dorantes, M.O. Aguilar-Ruiz et al. Population data for 15 loci (Identiflerr Kit) in a sample from the Valley of Mexico. *Legal Medicine* 2005.
93. Hernandez-Gutierrez S, Hernandez-Franco P, Martinez-Tripp S, Ramos-Kuri M, Rangel-Villalobos. H. STR data for 15 loci in a population sample from the central region of Mexico. *Forensic Sci Internat* 2005; 151 (1): 97-100.
94. Licea C, Rizzo J, Muñoz L, Paez R, Rangel V. Population data of nine STRs of Mexican-mestizos from Veracruz (Central South-Eastern, Mexico). *Legal Medicine* 2006; 8 (4): 251-252.
95. Torres R, Martinez C, Paez R et al. Forensic potential of the STR DXYS156 in Mexican populations: Inference of X-linked allele null. *Legal Medicine* 2006; 8 (1): 52-54.
96. The International HapMap Project. *Nature* 2003; 426 (6968): 789-796.
97. Sherry ST, Ward MH, Kholodov M et al. dbSNP: The NCBI database of genetic variation. *Nucl Acids Res* 2001; 29 (1): 308-311.
98. Hirakawa M, Tanaka T, Hashimoto Y, Kuroda M, Takagi T, Nakamura Y. JSNP: A database of common gene variations in the Japanese population. *Nucl Acids Res* 2002; 30 (1): 158-q62.
99. Jimenez-Sanchez G. Developing a platform for genomic medicine in Mexico. *Science* 2003; 300 (5617): 295-296.

100. Matsuzaki H, Loi H, Dong S et al. Parallel genotyping of over 10 000 SNPs using a one primer assay on a high density oligonucleotide array. *Genome Res* 2004; 14: 414-425.
101. Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet* 2005; 6 (2): 95-108.
102. Editorial. What is the Human Variome Project? *Nat Genet*; 39: 423.
103. Gibbs R. Deeper into the genom. *Nature* 2005; 437: 1233-1234.
104. Balding DJ. A tutorial on statistical methods for population association studies. *Nat Rev Genet* 2006; 7 (10): 781-791.
105. Kato M, Sekine A, Ohnishi Y et al. Linkage disequilibrium of evolutionarily conserved regions in the human genome. In; 2006: 326.
106. Gil McVean CCAS, Raphaelle Chaix. Perspective on human genetic variation from the HapMap Project. *PLOS GENETICS* 2005; 1 (4): e54.
107. T.A LV. Why we should continue to study race... but do a better job: An essay on race, racism and health. *Ethn Dis* 1996; 6: 21-29.
108. Race EaGWG. The use of racial, ethnic, and ancestral categories in human genetics research. In; 2005: 519-532.
109. Wang DG, Fan J-B, Siao C-J et al. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* 1998; 280 (5366): 1077-1082.
110. Carlson CS, Eberle M, Rieder M, Yi Q, Kruglyak L, Nickerson D. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *Am J Hum Genet* 2004; 74: 106-120.
111. Weir BS, Anderson A, Hepler A. Genetic relatedness analysis: Modern data and new challenges. *Nature Rev Gen* 2006; 7: 771-780.
112. Shriver M, Mei R, Parra E et al. Largescale SNP analysis reveals clustered and continuous patterns of human genetic variation. *Hum Genomics* 2005; 2 (2): 81-89.

Correspondencia:

Mariano Guardado-Estrada
Hospital General de México
Medicina Genómica,
Facultad de Medicina, UNAM.
Tel y fax: 2789 2000 ext: 1280
E-mail: jos_mariano@hotmail.com