



**CENTRO DE INVESTIGACIÓN Y DE ESTUDIOS
AVANZADOS DEL INSTITUTO POLITÉCNICO
NACIONAL**

UNIDAD ZACATENCO

DEPARTAMENTO DE BIOMEDICINA MOLECULAR

**“Diversidad genética patrilineal de las poblaciones Amerindias
(familias: Uto-Aztecas, Oto-Mangue y Tarascas) mediante marcadores
uniparentales de cromosoma Y: implicaciones biomédicas para futuros
estudios de asociación.”**

T E S I S

Que presenta

Biol. Aida Cristina Coss Rojas

Para obtener el grado de

MAESTRA EN CIENCIAS

EN LA ESPECIALIDAD DE BIOMEDICINA MOLECULAR

Directores de la Tesis: **Dr. Marco Antonio Meraz Ríos**

Dra. María del Rocío Gómez Ortega

Ciudad de México

Agosto, 2016

Agradecimientos

A los grupos Amerindios que participaron en este proyecto y que tan amablemente aportaron más que muestras a nuestro trabajo y a nuestra historia. Al CDI, ya que sin su logística e información no hubiera sido posible la realización de esta tesis.

Este trabajo se realizó gracias al apoyo del Consejo Nacional de Ciencia y Tecnología (CONACYT) con el financiamiento de Ciencia Básica 178239, otorgado a la Dra. María del Rocío Gómez Ortega. También agradezco al CONACYT por la beca con el número de apoyo 396913.

A *National Geographic* por el financiamiento al proyecto *Genographic*, del cual es parte este proyecto. AL departamento de Antropología de la Universidad de Pensilvania, en específico al Dr. Theodore G. Schurr, ya que es parte importante del proyecto.

Este trabajo se realizó en el laboratorio 14 “Genómica Poblacional” del Departamento de Toxicología y en el laboratorio del Dr. Marco Antonio Meraz en el Departamento de Biomedicina Molecular, ambos de CINVESTAV. También agradezco el apoyo del LaNSE para la genotipificación de las muestras.

Quiero agradecer también la guía de mis tutores, su apoyo y su paciencia, sobretodo esta última, porque sé que la necesitaron. Rocío, por abrirme las puertas del laboratorio y de este mundo, por tu apoyo y todo lo demás.

A mi familia, la columna que siempre me ha sostenido y que siempre me han apoyado. Mamá, que hay más que pueda decirte. Mis abuelos, porque forman parte de todo lo que soy. Los amo.

A mis amigos, que siempre estuvieron ahí, a pesar de la distancia. Caro, porque siempre me recibes con una doble cuando te visito para escuchar las penas, sabes que te amodoro. A las personas que conocí y que hicieron más divertido este viaje. Karla, Isma, Alex, Sandy, gracias por escucharme cuando invadía su laboratorio. Ale, llegaste en la recta final, pero llegaste para quedarte, con todo tu amor y todo tu apoyo, siendo como sólo tú sabes ser, forever.

Índice

Resumen.....	1
Abstract	2
Glosario.....	3
Introducción	4
Tipos de variación genética en las poblaciones.....	5
Estructura genética de las poblaciones.....	8
Determinación de la ancestría genética de una población	11
Cromosoma Y	13
Implicaciones del Linaje en los EAG	18
Panorama en México	19
Justificación.....	23
Hipótesis	24
Objetivo general.....	24
Objetivos particulares	24
Metodología	25
□ Población de estudio.....	25
□ Determinación del haplotipo mediante en PCR punto final.....	27
□ Parámetros de Genética Poblacional	27
Determinación de los macro-haplogrupos del Cromosoma Y	27
Parámetros de diversidad.....	28
□ Caracterización de los marcadores de alta definición por medio de PCR en tiempo real.....	29
□ Realización de redes	29
Comparación con otras poblaciones.....	30
Resultados.....	31
□ Determinación de los haplotipos por medio de marcadores tipo STR.....	31
Macro-haplogrupos del cromosoma Y.....	31
□ Determinación los HG específicos de los individuos que pertenecen al macro-HG Q por medio de marcadores tipo SNP.....	32
□ Parámetros de genética poblacional y diversidad.	32
Q-M3.....	33

Q-L54	36
□ Análisis de redes filogenéticas	37
Comparación con otras poblaciones	39
Discusión	42
Conclusión	52
Perspectivas	52
Bibliografía	53
Texto suplementario	A
Características demográficas de las poblaciones muestreadas	A
Bibliografía del texto suplementario	B
Imágenes Suplementarias	D
Tablas Suplementarias	G

Resumen

En los últimos años se ha detectado que uno de los principales problemas al realizar Estudios de Asociación Genética es la falta de reproducibilidad entre las poblaciones humanas. Esta dificultad se ha asociado a la complejidad en la estructura genética de las poblaciones mestizadas, debido a procesos demográficos recientes. Este estudio tuvo como objetivo caracterizar la diversidad genética y haplotípica del Cromosoma Y en poblaciones Amerindias. Estas poblaciones son antiguas, aproximadamente 13,500 años, lo cual ha permitido que su trasfondo genético sea menos complejo que el de una población joven, como la Mestiza Mexicana. Este trasfondo, de cierta manera homogeneizado por el tiempo, hace que estas poblaciones sean un buen modelo para la búsqueda de biomarcadores de riesgo para enfermedades complejas. Una forma de explorar este trasfondo, es el análisis de la ancestría, lo cual permite agrupar individuos en linajes específicos que, probablemente, compartirán características genéticas debido a su historia en común.

Para ello analizamos 17 marcadores hipervariables en la región no recombinante del Cromosoma Y, así como marcadores bialélicos para definir los linajes específicos de 122 individuos de 9 comunidades en Michoacán (Mazahua y P'urépecha), y en Nayarit (Nayeri y Wirráríka). Se encontró que estas poblaciones no están exentas de mestizaje, la ancestría dominante es la Amerindia, presente en una frecuencia del 80%. Asimismo, los niveles de diversidad genética son bajos, en contraposición de la alta diversidad haplotípica. Dentro de este linaje Amerindio fue interesante notar la presencia de una estructura representada por los principales sublinajes Americanos, QM3 y QL54, los cuales reflejan la historia demográfica de las poblaciones Nativas en su migración para poblar América, y los movimientos y flujos secundarios, que tuvieron lugar antes de la Conquista. Sin embargo, es necesario aumentar el número de muestra y de marcadores bialélicos para poder definir con mayor precisión la estructura genética del Cromosoma Y.

Abstract

In the last years, it has been detected that, one of the most important problems while performing Association Genetic Studies is the lack of reproducibility between human populations. These discrepancies have been associated with the complexity of the genetic structure of admixed populations, caused by recent demographic processes. This study aimed to portray the genetic and haplotypic diversity of Y Chromosome in Amerindian populations. Populations as old as these, about 13,500 years, have a less complex genetic background compared to younger populations as Mexican Mestizos. This background, homogenized, makes these populations to be good models of the search of risk biomarkers for complex diseases. One way to explore this background, is the ancestry analysis, which allows to group individuals to specific lineages that, probably, share genetic traits consequence of a common history.

In order to characterize the genetic structure as to define the specific lineages, we used 17 hypervariable markers and biallelic markers in the Y-Chromosome non-recombining region to analyze 122 males, of nine communities in Michoacán (Mazahua and P'urhépecha), and Nayarit (Nayeri and Wirrárika). We found that these populations have a low level of admixture, being Amerindian the dominant ancestry, with a frequency of 80%. Also, genetic diversity is low, contrasting with the high haplotypic diversity. Inside the Amerindian legacy, it was interesting to find a substructure represented by the two principal American lineages, QM3 and QL54. These lineages reflect the demographic history of the Native populations during the entrance to America, and the secondary movements that happened before the arrival of Spaniard conquerors. However, it is necessary to increase the size of the sample and the number of biallelic markers in order to get a clear characterization the genetic structure of Y-Chromosome.

Glosario

AIMs: Marcadores Informativos de Ancestría.

DNA: Ácido Desoxirribonucleico.

PCR: Reacción en cadena de la polimerasa.

EAG: Estudio de Asociación Genética.

FTDNA: Family Tree DNA.

h: Diversidad.

HG: Haplogrupo.

Ht: Haplotipo.

INDELS: Inserción-Delección.

k: Número de alelos.

MSY: Región específica masculina.

mtDNA: DNA mitocondrial.

SNP: Polimorfismo de un solo nucleótido.

SRY: Región determinante del sexo del Cromosoma Y.

STR: Repetido corto en tándem.

VNTR: Número variable de repeticiones en tándem.

YSTR: Repetido corto en tándem del cromosoma Y.

Introducción

La diversidad genética, es una medida que cuantifica la magnitud de la variación genética que existe en una población; esta variación es la base por medio de la cual las fuerzas evolutivas actúan permitiendo la adaptación de las especies a su ambiente ¹.

En la especie humana, los patrones actuales de diversidad genética permiten entender la base genética de las adaptaciones fenotípicas que les permitieron la colonización del planeta ². Estos patrones son, a su vez, el resultado de diferentes eventos demográficos y evolutivos (migraciones, eventos de selección, deriva génica, etc.) así como de eventos culturales (religión, conquistas, migraciones, etc.)²⁻⁴.

La determinación del nivel de variación en el genoma humano y las bases por las cuales esta variación se relaciona con la salud humana es uno de los principales desafíos de las ciencias biomédicas ⁵. Para ello, se han utilizado modelos (animales y/o celulares), cuyo concepto base es el de una analogía de los sucesos biológicos entre estos y una especie “meta”. No obstante, aunque el valor predictivo de los resultados obtenidos en estos modelos pueda ser alto, la extrapolación total de éstos hacia los humanos no es completamente real, ya que estos modelos están definidos genéticamente y sus patrones son uniformes, es decir, son idénticos genéticamente. Los humanos, por su parte, son genéticamente muy diversos, influenciados por diferencias culturales, dietéticas y ambientales, entre muchas otras ⁶. Consecuentemente, al hacer estudios en poblaciones humanas, los resultados pueden no ser consistentes entre diferentes poblaciones, esto debido a que podrían existir variantes específicas de las poblaciones o diferentes interacciones gen-gen o gen-ambiente ⁷.

Gracias a tecnologías como la reacción en cadena de la polimerasa (por sus siglas en Inglés *Polymerase Chain Reaction*, PCR) y la secuenciación, se han

comenzado a determinar estos niveles de variación genética entre los humanos, permitiendo comparar la diversidad de los individuos intra e interpoblacionalmente ^{3,8,9}. Con estos datos se puede realizar la búsqueda de biomarcadores de susceptibilidad o genes responsables de las enfermedades más comunes presentes en las poblaciones humanas ^{8,10}. Sin embargo, para lograr una determinación y comparación confiable de las variantes de riesgo se debe estudiar la estructura genética y ancestral de las poblaciones ¹¹.

Tipos de variación genética en las poblaciones

A lo largo de los tres mil millones de nucleótidos que conforman el genoma humano existen diferentes cambios, estas variaciones en las diferentes regiones del DNA principalmente son referidas como comunes o raras ⁵. Las primeras que son sinónimo de polimorfismos se definen como variantes con una frecuencia alélica de al menos el 1% de la población ¹².

Las características de estas variantes son diferentes a lo largo del genoma, por lo que su uso para la determinación de la estructura de las poblaciones depende de ciertos criterios ^{13,14}:

- Ser altamente diversos, es decir que entre los individuos exista una amplia diversidad alélica.
- Caracterización fácil y de bajo costo.
- Ser reproducibles dentro de la población y que los resultados obtenidos sean de fácil interpretación.
- Ser neutros, es decir que no estén bajo presión selectiva y/o afecten algún rasgo fenotípico.
- Tener una baja tasa de mutación, es decir, que no existan modificaciones de una generación a la siguiente (25 años).

También se puede hablar de estas variantes en términos de su composición nucleotídica, en donde se reconocen dos tipos ⁵:

- Polimorfismos de un solo nucleótido
- Variantes estructurales: inserciones-delecciones (INDELs), sustituciones, inversiones y número variable de copias (Figura 1).

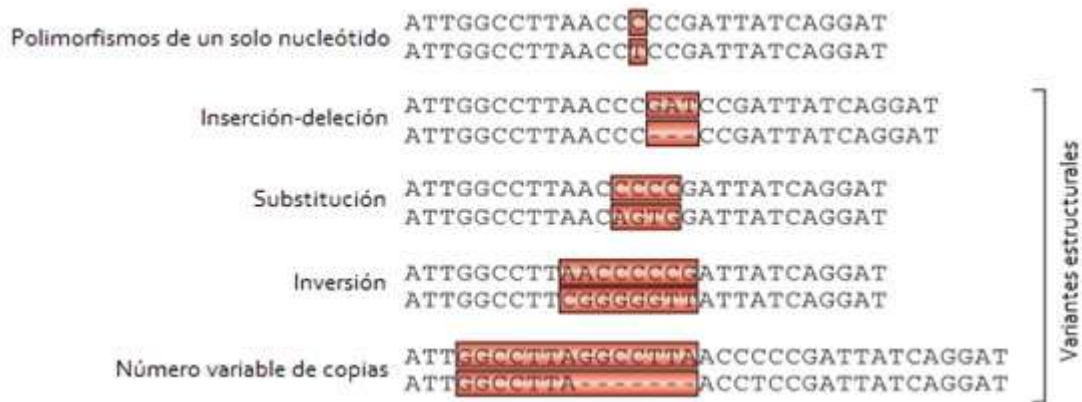


Figura 1. Tipos de variantes genéticas. Modificado de Frazer et al, 2009.

Entre estas variantes las más estudiadas son:

- Polimorfismos de un solo nucleótido** (SNPs, por sus siglas en inglés, *Single Nucleotide Polymorphism*). Representan los polimorfismos más abundantes en el genoma humano. Son cambios de una sola base en la secuencia de DNA, usualmente con alelos en una posición dada, es decir son bialélicos ¹⁵. Tienen como ventaja que presentan una baja tasa de mutación (aproximadamente 2.5×10^{-8} por nucleótido) y debido a su número es fácil analizarlos y crear mapas genéticos de alta densidad, ya que se estima que existen aproximadamente 1.42 millones de SNPs en el genoma ¹⁶⁻¹⁸. Sin embargo se requiere estudiar un gran número de ellos (50 o más para igualar el uso de 12 microsatélites) para poder tener un buen poder estadístico ¹⁷.

b. **Repetidos en tándem.** Estos marcadores son secuencias de uno o más nucleótidos (*motif* o motivos) que se repiten de forma adyacente, de manera que se encuentran en tándem y varían en número, lo cual los hace altamente polimórficos ¹⁹. Se pueden identificar dos tipos de repetidos en tándem: Los *minisatélites* o VNTRs (por sus siglas en inglés, *Variable Number Tandem Repeat*) y los *microsatélites* o STRs (por sus siglas en inglés, *Short Tandem Repeat*) ¹⁴. Los primeros se encuentran principalmente en zonas subteloméricas y son motivos de 6 a 100 nucleótidos; mientras que los *microsatélites*, son motivos de 1 a 6 bases que van de 50 a 300 pb, localizados a lo largo de todo el genoma, con un alto grado de variación; por su tamaño, estos últimos son los más utilizados y se pueden clasificar dependiendo del tamaño del motivo o su estructura (Tablas 1 y 2) ^{14,20,21}.

Tabla 1 Clasificación de los microsatélites dependiendo de su tamaño. Tóth et al., 2000.

Tipo de microsatélite	Ejemplo
Mononucleótidos	(AAAAAAAAAAAA) / (A) ₁₃
Dinucleótidos	(CACACACACACA) / (CA) ₆
Trinucleótidos	(CAG) ₁₂
Tetranucleótidos	(GAAG) ₈
Pentanucleótidos	(TACGT) ₉
Hexanucleótidos	(TGCTAG) ₇

Tabla 2 Clasificación de los microsatélites dependiendo de su estructura. Tóth et al., 2000.

Tipo de Microsatélite	Ejemplo
Puros o perfectos	(AC) _n
Puro interrumpido	(CA) _n AA (CA) _n
Compuestos	(GT) _n (TG) _n
Compuestos interrumpidos	(CT) _n (GT) _n CTAT(GT) _n
Complejos	(ACC) _n TG(GA) _n (TTA) _n

La presencia y frecuencia de los diferentes alelos de cada uno de estos polimorfismos puede cambiar entre poblaciones e incluso dentro de una misma población, lo que contribuye a la diversidad genética de las mismas ^{22,23}. Esto genera que las poblaciones sean diferentes e incluso dentro de ellas pueden existir subdivisiones relacionadas con su trasfondo genético, esto es conocido como estructura genética, la que en los últimos años, se ha convertido en un factor importante para los Estudios de Asociación Genética (EAG) y estudios relacionados con la estructura y validación de polimorfismos en las poblaciones ²⁴⁻²⁶.

Estructura genética de las poblaciones

Principalmente las poblaciones humanas se definen con base en su afinidad geográfica y cultural, pero esto no es absoluto, sobre todo cuando se determina y compara la estructura genética de los individuos que la conforman ²⁷. La presión de las fuerzas evolutivas, como la mutación, la selección y la deriva, hacen que el componente genético de las poblaciones sea altamente dinámico, a lo que se suman aspectos demográficos y culturales, como las conquistas, migraciones o las jerarquías sociales y religiosas ²².

Se ha identificado al *admixture* o mestizaje, como uno de los procesos más importantes por el cual se genera una sub-estructuración en las poblaciones (Figura 2) ^{25,28}. Este mestizaje se ve reflejado genéticamente en los alelos presentes en las poblaciones jóvenes que son el resultado de intensos procesos demográficos recientes, como conquistas o migraciones, y que descienden de dos o más poblaciones ancestrales, como es el caso de las poblaciones Latinoamericanas ²⁹.

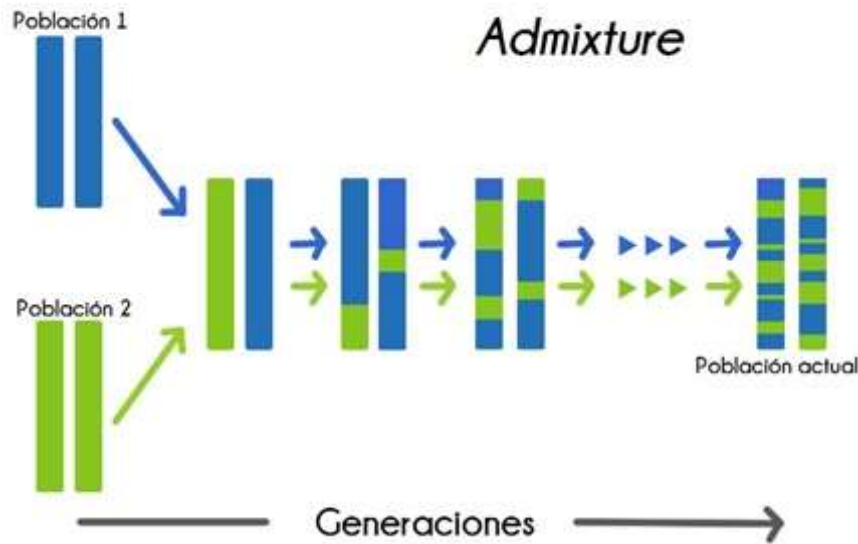


Figura 2. Esquemización del proceso de *admixture*. Las frecuencias alélicas cambian de generación en generación debido a procesos como la recombinación, la selección, la deriva génica y la mutación.

Analizando lo anterior, podemos distinguir dos escenarios poblacionales dependiendo de la antigüedad de las poblaciones y de los procesos que les han dado origen:

Población homogénea. Estas son poblaciones antiguas (de más de 100 generaciones) que ha mantenido un flujo genético constante entre ellas a lo largo del tiempo bajo la presión de ambientes similares. Consecuentemente, las frecuencias alélicas en estas poblaciones tienen una distribución similar en los individuos que las conforman (Figura 3A)³⁰. Es necesario aclarar que una población homogénea **no** es una población endogámica, ya que en esta última la frecuencia de ciertos alelos se ven aumentadas mientras que las de otros son muy bajas o nulas (Figura 3B).

Población heterogénea. Es una población formada por la mezcla reciente de diferentes poblaciones parentales y consecuentemente es muy joven (menos de 15 generaciones). Aunado a este hecho, este tipo de poblaciones han pasado por eventos demográficos que han modificado sus frecuencias alélicas (cuellos de botella, efecto fundador, etc.)^{28,31}. Es por ello

que los individuos que las integran presentan diferentes grados de mestizaje, provocando que existan diferentes estratos genéticos, también llamados subpoblaciones (Figura 4) ³².

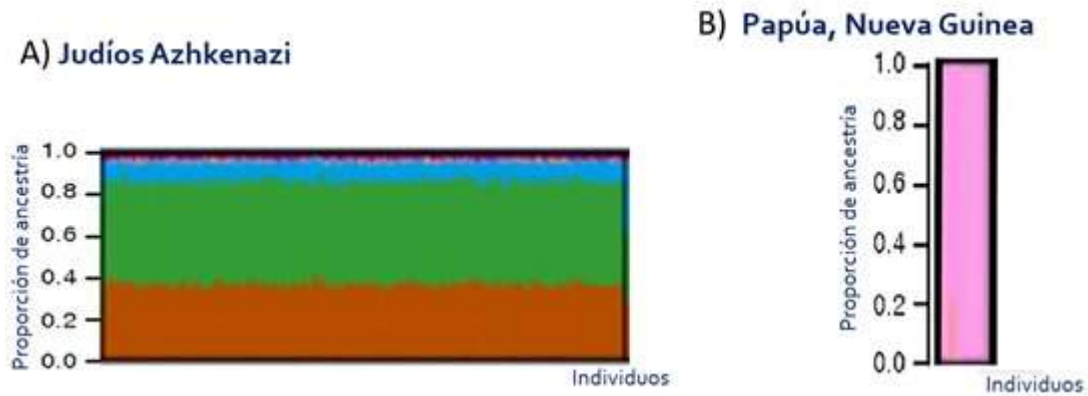


Figura 3. La arquitectura genética en una población homogénea es muy similar entre los individuos que la conforman. Como se puede observar en el análisis realizado con el programa STRUCTURE, a pesar de tener diferentes orígenes ancestrales, los individuos de éstas poblaciones tienen patrones similares de diversidad. En los judíos Ashkenazi (A), que son una población homogénea, se tiene la misma probabilidad de encontrar uno u otro alelo al muestrear a un individuo de esa población, esto no sucede en una población endogámica como la de Papúa (B), donde todos los individuos presentan, con altas frecuencias, algunos alelos, porque estas poblaciones son poco diversas, al contrario de las poblaciones homogéneas que son altamente diversas.

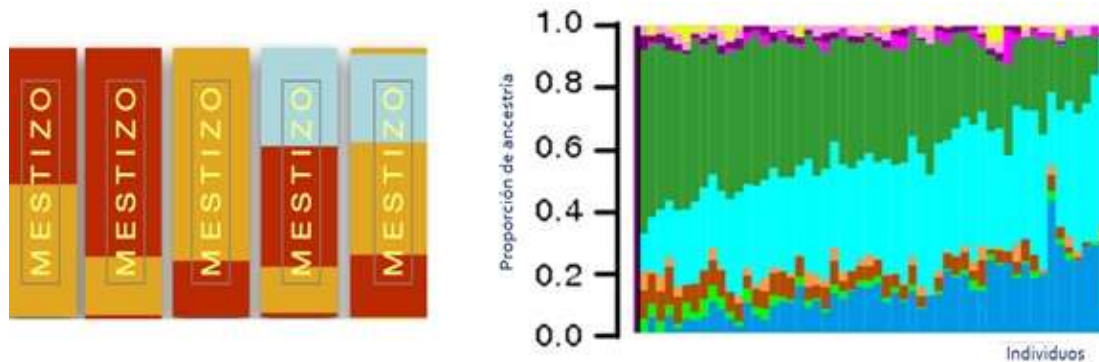


Figura 4. En las poblaciones heterogéneas las proporciones alélicas varían, cada uno de los individuos que las conforman pueden tener diferentes proporciones de las características genéticas de las poblaciones parentales (izq.). Como se puede observar en la gráfica de STRUCTURE (der.), se observa una población heterogénea en donde cada individuo presenta proporciones diferentes de las ancestrías, por lo que la probabilidad de tomar un alelo cambia de un individuo a otro.

Una manera por medio de la cual se ha propuesto determinar la estructura genética de una población es mediante el estudio de su ancestría. Estos estudios permiten conocer los linajes ancestrales que conformaron a una población particular, lógicamente, los individuos de una población

antigua y por lo tanto homogénea comparten cargas ancestrales similares ³³. Además, la especificidad con la cual se define la ancestría permite a su vez diferenciar poblaciones genéticamente divergentes, que han estado aisladas por cierto tiempo y que poseen un trasfondo genético específico y características alélicas únicas ^{34,35}.

Determinación de la ancestría genética de una población

Metodológicamente, para determinar la ancestría de un individuo, se pueden utilizar marcadores diploides (marcadores informativos de ancestría o AIMS por sus siglas en inglés *Ancestry Informative Markers*) o, haploides (DNA mitocondrial y cromosoma Y), cada uno de ellos proporciona información complementaria para poder determinar la estructura de la población.

- **AIMs:** Estos marcadores autosomales representan una gran proporción del genoma; cada uno de los marcadores tipo SNP que se utilizan para este tipo de estudio presenta diferentes frecuencias entre poblaciones ya que se ha demostrado que estos SNPs tienen diferencias interétnicas ³⁶. Sin embargo al estar sujetos a procesos como recombinación y presentar diferentes tasas de mutación se requiere de utilizar un gran número de ellos (entre 2,000 a 5,000), para disminuir el sesgo y tener una identificación más precisa de los orígenes ancestrales de las poblaciones, lo que encarece el procedimiento ³⁶⁻³⁹. Además, estos marcadores solo indican la proporción de cada ancestría con respecto a los cuatro grupos poblacionales más importantes (Africanos, Amerindios, Asiáticos y Europeos); no obstante, estos marcadores son incapaces de asignar el clan ancestral específico de un individuo ⁴⁰.

- **DNA mitocondrial (mtDNA):** Es una molécula de DNA circular de doble cadena, de 16,569 pares de bases que contiene 37 genes que codifican para componentes de la cadena respiratoria, rRNAs y tRNAs ⁴¹. En la célula, el mtDNA se encuentra en gran número, lo cual permite su fácil obtención; se hereda de manera materna por lo que no lleva a cabo recombinación y posee una alta tasa de mutación mayor (0.043 mutaciones por generación) a la del DNA autosómico (2.5×10^{-8} mutaciones por generación), lo cual permite utilizarlo para estudios de estructura genética ^{41,42}.
- **Cromosoma Y:** Es el segundo cromosoma más pequeño de todos, con un tamaño promedio de 60 Mb, localizado en el núcleo, contiene los genes necesarios para la determinación del sexo masculino (SRY, por sus siglas en inglés *Sex-determining Region Y*) por lo que su herencia es paterna, siendo la contraparte natural del mtDNA ⁴³.

Estos últimos dos marcadores son conocidos como marcadores uniparentales, los cuales ofrecen la ventaja de que al no recombinar, se heredan con pocas o nulas modificaciones a lo largo del tiempo; esto permite que su evolución sea descrita a través de genealogías (árboles filogenéticos) agrupándolos por sus afinidades moleculares y, además, siguiendo patrones geográficos. Finalmente y dado que las mutaciones se acumulan secuencialmente, a través de los marcadores uniparentales se puede determinar la antigüedad de los linajes; en resumen, permiten asignar un clan ancestral a un individuo, al que se correlaciona con orígenes genéticos y geográficos; además de su bajo costo y rápido análisis (Figura 5) ^{44,45}.

Estos marcadores, por lo tanto, son de gran utilidad en el estudio de poblaciones complejas ya que proveen evidencia sobre las historias demográficas de las poblaciones ⁴⁰. Inclusive, se propone que el cromosoma Y es el más útil en el estudio de estas poblaciones, ya que movimientos

migratorios, cambios culturales, conquistas, entre otros procesos demográficos, son llevados a cabo, principalmente, por el género masculino dejando una huella genética en este marcador ⁴⁶⁻⁴⁹.

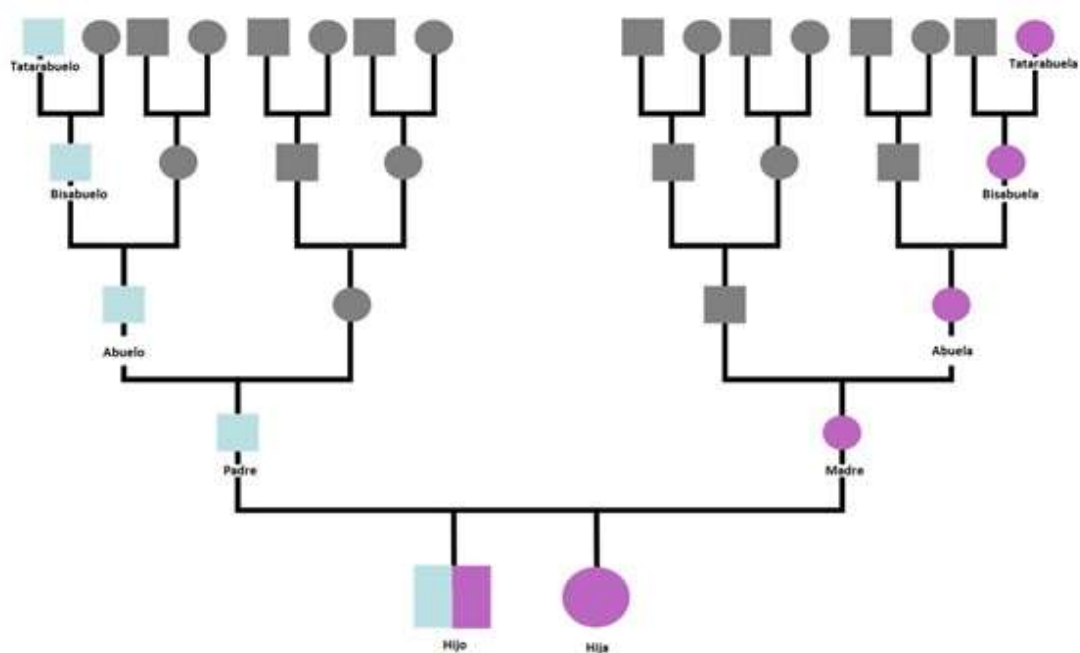


Figura 5. Por medio de los marcadores uniparentales se pueden asignar clanes ancestrales, los cuales debido a los mínimos cambios que ocurren, serán los mismos entre padres e hijos. Por ejemplo, para el mtDNA, el clan de la hija será el mismo que el de la madre y la abuela; los mismo ocurre con el cromosoma Y, pero en el género masculino.

Cromosoma Y

En el humano, el cromosoma Y determina al sexo masculino, ya que contiene el gen SRY el cual genera la cascada de señalización necesaria para el desarrollo del fenotipo masculino⁵⁰. Se cree que en un principio era homólogo del cromosoma X, pero existieron eventos de inversión cromosómica que disminuyeron y finalmente suprimieron casi en su totalidad esta homología, a excepción de dos segmentos pseudo-autosómicos ⁴⁸.

Este cromosoma contiene 27 genes que codifican para proteínas, dos segmentos pseudo-autosómicos (2.6 Mb) que pueden recombinar con el cromosoma X y una región de gran tamaño no recombinante o región específica masculina (MSY, por sus siglas en inglés *Male-specific region of the Y*

chromosome), que abarca un 95% del tamaño total del cromosoma y que posee una estructura genética compleja como se muestra en la Figura 6 ⁵¹.

Una de las características más importantes de este cromosoma como marcador es su modo de herencia paterna que, aunada a la ausencia de recombinación, permite que la MSY se transmita de manera casi idéntica a través de las generaciones, por lo que la única forma de diversificación de las variantes contenidas en esta zona es por medio de las mutaciones ⁴⁸. En la MYS existen diversos tipos de polimorfismos, tanto SNPs y STRs, permitiendo la realización de estudios de diversidad, la determinación de estados ancestrales y finalmente el establecimiento de linajes ^{44,52-54}.

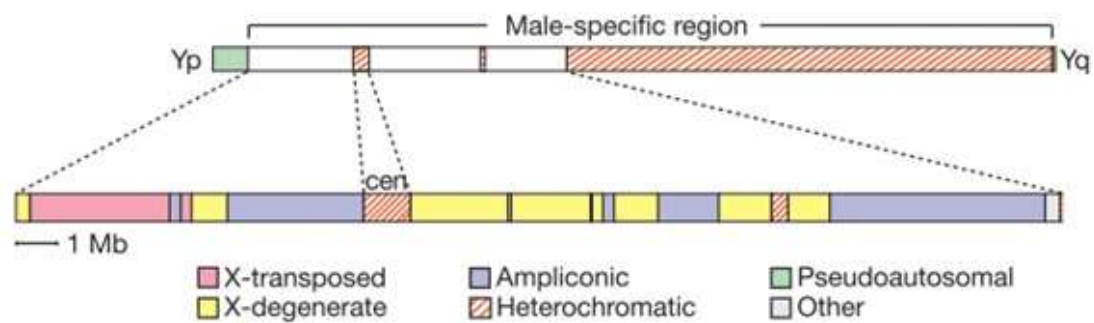


Figura 6. Estructura de la región específica masculina del cromosoma Y. Tomado de Skaletsky et al, 2003.

Los estudios de ancestría se inician principalmente al determinar el haplotipo (Ht), que es el conjunto de marcadores contenidos en la MYS y que se define por medio de STRs, como se mencionó anteriormente, estos polimorfismos son multialélicos por lo que se puede encontrar una gran diversidad entre los individuos de una población lo que depende de las características de la misma ^{47,55}. Esta diversidad puede ser consistente con movimientos demográficos, como fragmentaciones, aislamientos y re-expansiones de las poblaciones ⁴⁰.

Para determinar el linaje específico, también llamado haplogrupo (HG), se utilizan los marcadores tipo SNP, ya que su baja tasa de mutación permite la agrupación de cromosomas Y en diferentes linajes que comparten estados derivados de estos marcadores ⁵⁶. Una vez identificados estos últimos, se utiliza un método y una nomenclatura establecida por el Consorcio del Cromosoma Y, así finalmente se determina su pertenencia a un clado (linaje) ancestral ⁵⁵. Estos HG también nos permite conocer movimientos migratorios, lo que se puede observar en la forma en que los éstos se distribuyen a lo largo de los continentes. Además, los linajes ancestrales también nos dan información de las poblaciones divergentes, lo que se logra por medio de la formación de sub-haplogrupos, incluso se han encontrado diferentes sub-haplogrupos específicos en poblaciones que estuvieron aisladas (Figura 7)⁴⁰.

Esta nomenclatura se realizó a partir de la comparación con grupos externos como chimpancés (*Pan sp.*), gorilas (*Gorilla sp.*) y/o orangutanes (*Pongo sp.*), para determinar los estados ancestrales de los marcadores ^{56,57}. Los clados principales, o macro-haplogrupos, se nombraron con letras mayúsculas (de la A a la R) (Figura 7). Dentro de éstos existen sub-haplogrupos, que son nombrados utilizando un sistema alfanumérico alternado; por ejemplo, dentro del macro-HG Q, existe el HG Q1, y dentro de éste otros sub-HG denominados Q1a, y Q1b; derivados de éstos existe Q1a1, Q1a2, Q1a3 y así sucesivamente hasta la punta de las ramificaciones ^{56,58}. Alternativamente, se pueden nombrar a los HG por las mutaciones que los definen, es decir, si al HG Q lo define la mutación M242, entonces se podría llamar también Q-M242; finalmente se denota la existencia de un paragrupo al colocar un asterisco (*) (Figura 8) ^{56,58}.

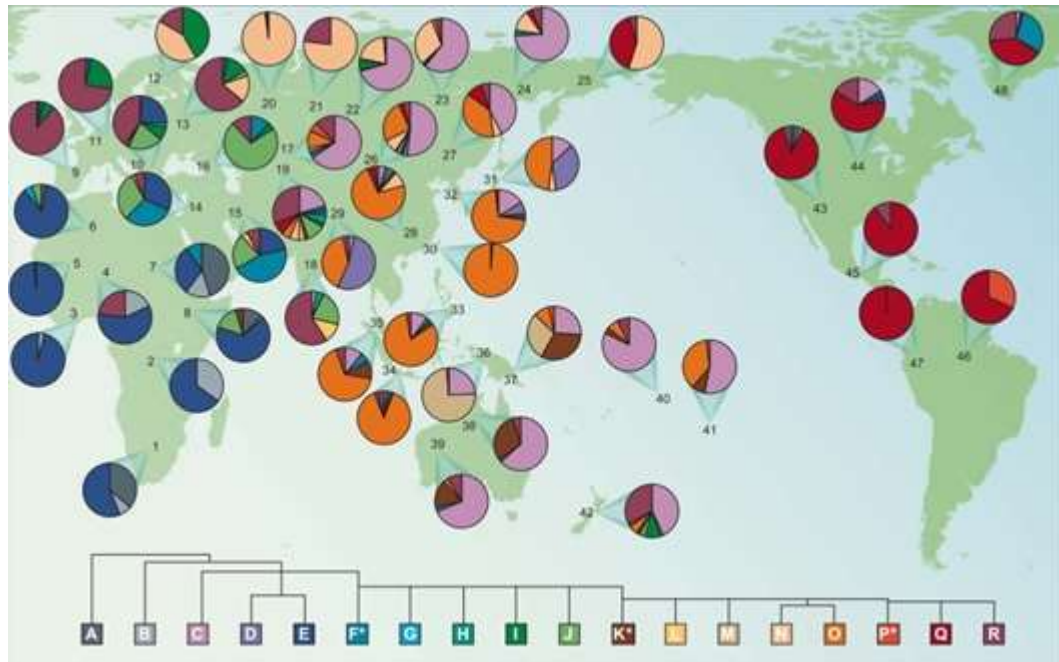


Figura 7. Distribución global de los macro-HG del cromosoma Y. Cada círculo muestra la frecuencia de cada uno de los HG en diferentes zonas geográficas. Tomado de Jobling et al, 2003.

Las relaciones entre los HG son representados por árboles filogenéticos, que permiten observar las relaciones genealógicas y evolutivas de los marcadores del cromosoma Y (Figura 8) ⁵⁶. Las agrupaciones de cromosomas Y que comparten una serie de estados derivados, son conocidos como grupos monofiléticos. No obstante, a la fecha muchos de ellos no están definidos, pues se desconocen los cambios de los posibles estados derivados, estos son conocidos como paragrupos (Fig. 7) ⁵⁶.

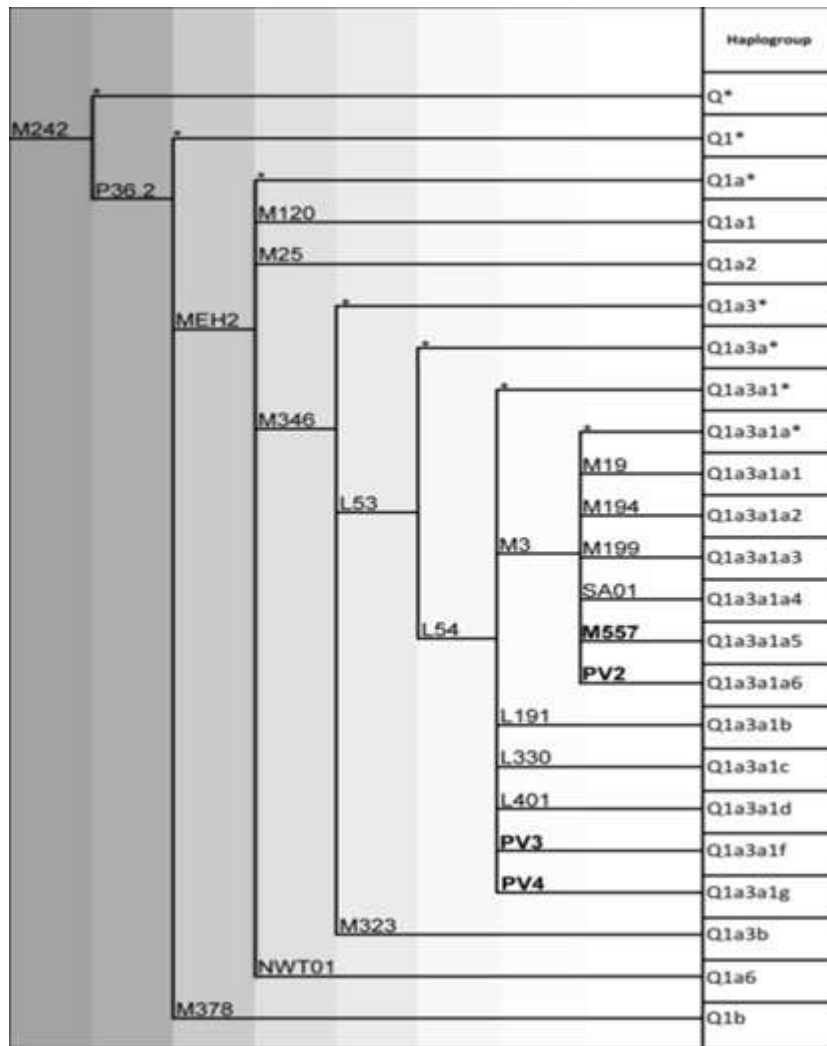


Figura 8, Árbol filogenético resumido del macro HG Q. Los nombres de los sub-HG se encuentran en las puntas. Los SNPs que definen cada clado se indican en las ramas. * indica la presencia de un parágrafo, es decir que no se han encontrado los marcadores para definir a los sub-HG. Tomado de Battaglia et al, 2013.

Estos análisis comenzaron a ser usados de manera generalizada a finales de los años noventa, cuando se logró identificar una gran cantidad de polimorfismos con las características adecuadas³⁸. Sin embargo no se han utilizado de manera extensiva para identificar la estructura genética de las poblaciones por medio de su ancestría⁶⁰.

Implicaciones del Linaje en los EAG

Los EAG buscan una correlación entre una enfermedad y una variante genética, con el fin de identificar genes candidatos o regiones genómicas que contribuyan a esa enfermedad en particular ⁶¹. El diseño experimental más utilizado es el de casos y controles, en donde los cambios de frecuencia entre variantes alélicas en un grupo que presenta la enfermedad (casos) y uno que no la presenta (controles) se toma como evidencia de una posible asociación con la enfermedad causal⁶². No obstante, al realizar este tipo de comparaciones es posible que las diferencias estén reflejando la arquitectura genética de estratos diferentes dentro de una población y no una asociación con el fenotipo de la enfermedad ²⁵. Es por ello que la sub-estructuración de las poblaciones juega un papel importante ya que es la principal fuente de confusión, generando asociaciones espurias (falsos-positivos o falsos-negativos) o resultados no reproducibles entre poblaciones ²⁶.

Del mismo modo, la falta de reproducibilidad entre las poblaciones es otro de las problemáticas de los EAG, y es que, a diferencia de las enfermedades mendelianas donde la presencia de un gen determina la enfermedad, la mayoría de las enfermedades “modernas”, más conocidas como complejas, se relacionan con una cantidad, hasta el momento, indefinida de genes ⁶³.

Para evitar estas situaciones, se emplean poblaciones “modelo” que por definición deben ser genéticamente homogéneas. Entre las más empleadas se encuentran las poblaciones: Askenazi, Han, Vasca y, Yoruba, que debido a que sus características genéticas disminuyen notablemente los efectos provocados por la estratificación genética, y cuyos patrones genéticos están extensamente estudiados, lo cual permite que las asociaciones sean más evidentes (Figura 9) ⁶⁴⁻⁶⁶.

La característica principal por la cual estas poblaciones son consideradas “modelos” es la distribución de sus frecuencias alélicas, las que son proporcionales para cada alelo en toda la población. Lo anterior es consecuencia de un aislamiento cultural, así como de la antigüedad de la población, además del mantenimiento de un flujo génico constante - pero no intenso - con otras poblaciones cercanas con factores ambientales en común y por lo tanto bajo las mismas presiones selectivas; características principalmente encontradas en las poblaciones nativas de las diferentes regiones ⁶⁷.

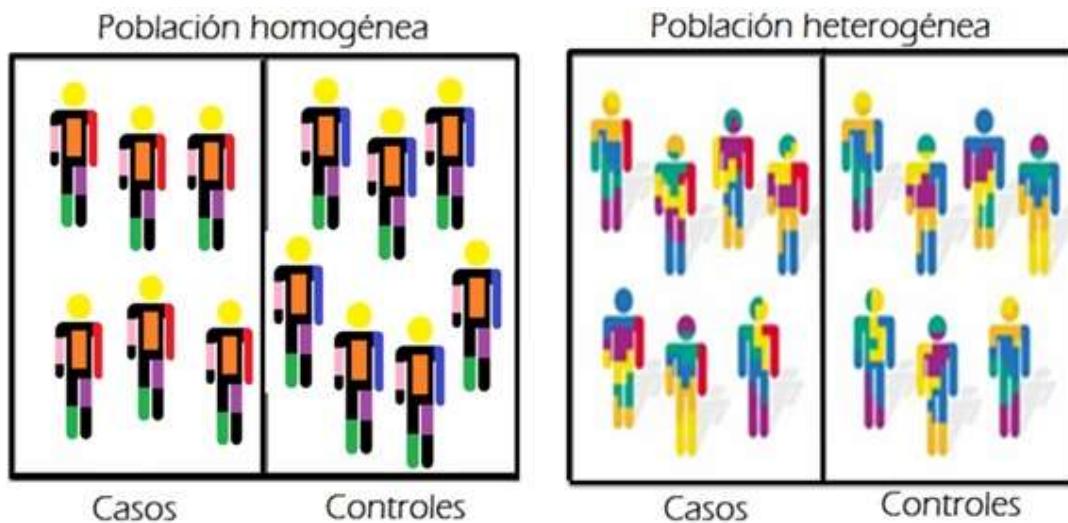


Figura 9. En las poblaciones homogéneas, una variación en las frecuencias entre los casos y los controles es más evidente que en las poblaciones heterogéneas. Tomada y modificada de la Universidad Hebrea de Jerusalén, 2014.

Panorama en México

La población mexicana es un claro ejemplo de una población compleja. Esta población se formó a partir de un intenso proceso de mestizaje durante la conquista, por lo que principalmente posee ancestrías Europea, Africana y Amerindia; e incluso existe evidencia de ancestría Asiática y del Medio Oriente; este patrón genético se vio moldeado también por el sistema de castas instaurado posteriormente y por pequeñas guerras y migraciones ^{69,70}.

Los estudios de marcadores uniparentales realizados en México han revelado que el 93.3% de la población Mestiza presenta una ancestría Amerindia analizada por medio de mtDNA⁷¹. Por el contrario, la ancestría patrilineal ha mostrado una mayor heterogeneidad, relacionada con el gran número de hombres de, diferentes orígenes, que migraron al país durante la conquista (Figura 10) ^{49,72}.

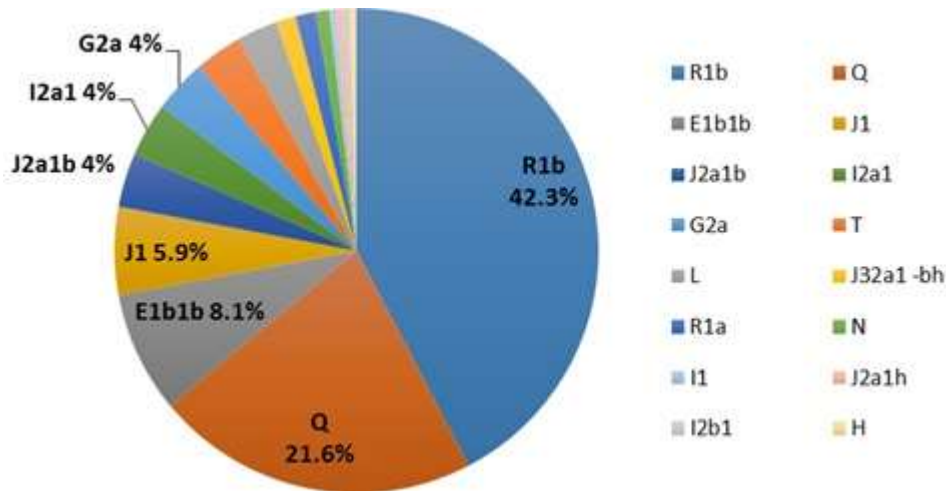


Figura 10. Tipos de ancestría paterna presentes en la población Mestiza mexicana por medio del análisis del cromosoma Y. Siendo las más comunes la Europea (R1b), la Amerindia (Q), la Africana (E1b1b) y la del Medio Oriente (J1). Tomado de Santana et al, 2014.

En México habitan más de 10 millones de Nativos Americanos a lo largo del país, los que se distribuyen en diferentes comunidades, relativamente aisladas de los grupos Mestizos debido a causas culturales o económicas ⁷³. Estas etnias se concentran principalmente en la región centro y sur del país (90%) aunque existen algunas comunidades en el norte (Figura 11) ⁷³.

El macro-HG específico de la ancestría paterna amerindia, es el macro-HG Q definido por el marcador M242 (rs8179021 C>T), proveniente del centro de Asia, cuya llegada al continente Americano ocurrió aproximadamente hace 15,000 años ⁴. Este HG tiene al menos tres representaciones importantes en el continente, las que probablemente correspondan a las olas migratorias durante el poblamiento de America.

Entre ellos destacan, Q-MEH2 (rs4252209 G>T), un HG minoritario (frecuencia 0.065) que se distribuye principalmente en el norte del continente y que ha sido relacionado con las tribus del Este de Siberia ^{58,74}.



Figura 11. Distribución de las poblaciones indígenas en México. Tomado del Instituto Nacional de las Mujeres.

Q-M3 (rs3894 C>T) es el HG mayoritario del continente con una distribución clinal Norte-Sur. QM3 es considerado el HG exclusivo del Continente Americano y se piensa que apareció mientras los migrantes Asiáticos se establecieron en la parte Noroeste del continente antes de pasar por Norteamérica hacia el sur. Se distribuye a lo largo de todo América y es el más frecuente en las poblaciones Amerindias (76.8%) (Figura 12) ⁵⁸. De descubrimiento reciente, QL54 (rs34954951 G>A), es el menos estudiado de los HG, pero se sabe que proviene de la actual República de Altai (al Sur de Rusia, Oeste de Mongolia, Norte de China y Este de Kazahstan); de la región

de los Montes Sayanes (localizados en la frontera norte de la meseta de Mongolia, separándola de Siberia), y de Tuvá (Rusia) ⁷⁵. QL54 se distribuye de Mesoamérica hacia el sur, probablemente porque los migrantes que provenían de Siberia llegaron por medio de una ruta costera. Este HG tiene una frecuencia de 0.167 en las poblaciones Amerindias y a la fecha no está totalmente dilucidada su diversidad ^{58,76-78}.

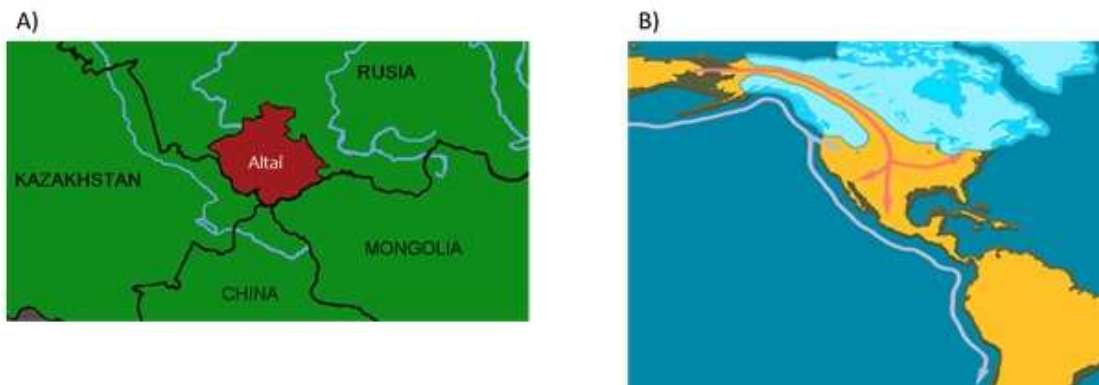


Figura 12. Los migrantes Asiáticos que colonizaron América provenían de la actual República de Altai (A), y se cree que hubo posibles diferentes entradas al Continente Americano, por lo que en (B) se muestran las posibles rutas migratorias. Tomado de Pennsylvania Historical & Museum Commission.

En México, recientemente se ha reportado la presencia de QM3 (0.69) y QL54 (0.30). Sin embargo, los estudios filogenéticos sugieren la existencia de una gran diversidad conformada por diferentes paragrupos que no han sido muy bien definidos a la fecha. Asimismo, los estudios realizados en el sur del continente han mostrado la presencia de diferente sub-HGs, los que no han sido analizados en las poblaciones Amerindias mexicanas ⁵⁸.

Desde el punto de vista biomédico, se ha observado que los grupos Amerindios presentan enfermedades complejas como las enfermedades cardiovasculares, metabólicas, cáncer y pulmonares; las que a su vez son problemas de salud a nivel nacional⁸⁰. Muchas de estas poblaciones comparten características con las poblaciones homogéneas por lo que pueden ser utilizadas como modelos de estudio en la búsqueda de marcadores de susceptibilidad, además de la posible presencia de variantes regionales o

específicas de estas poblaciones, que podrían estar compartidas con la población Mestiza. Pero para llegar a ese punto, primero es necesario evaluar la diversidad y estructura genética de estas poblaciones Nativas con el fin de determinar si cumplen con los criterios para ser homogéneas. En este punto y dado que la ancestría paterna es la más heterogénea, el estudio de la diversidad en el cromosoma Y sería el más informativo.

Justificación

Las poblaciones Nativas Americanas de México son poblaciones antiguas algunas de las cuales han mantenido un aislamiento de las poblaciones modernas en un proceso de reafirmación étnica. Como toda población, las poblaciones Nativas son dinámicas y han pasado por diferentes procesos demográficos, por lo que la caracterización de sus estructuras genéticas, así como el análisis de sus linajes ancestrales, permitirá evaluar el valor biológico que pueden tener en el desarrollo de “modelos poblacionales” potenciales que puedan ser usados para futuros estudios de asociación genética.

Hipótesis

Las poblaciones Nativas Americanas, por sus características demográficas poseen una estructura genética particular, que puede ser usada como modelo poblacional en estudios de asociación genética.

Objetivo general

Determinar la diversidad genética de las poblaciones Nativas Americanas de México usando la metodología del análisis de polimorfismos multialélicos tipo STR, en combinación con la evaluación de marcadores linaje específico.

Objetivos particulares

1. Determinar la diversidad haplotípica y los datos genéticos descriptivos de cuatro poblaciones Amerindias: Mazahua, Nayeri, P'urhépecha y Wirráríka.
2. Realizar el análisis de la diversidad de los diferentes haplogrupos presentes en las poblaciones mencionadas.
3. Dentro los diferentes subhaplogrupos en los distintos linajes Nativos Americanos, mediante el uso de marcadores linaje específico.
4. Determinar las relaciones filogenéticas entre y dentro de las poblaciones analizadas.

Metodología

- **Población de estudio**

Se realizó un muestreo por conveniencia de sangre y/o saliva de individuos masculinos pertenecientes a grupos los étnicos de los estados de Michoacán y Nayarit, los que cumplieron con los siguientes criterios (Figura 13):

Inclusión: Individuos del sexo masculino, que se reconozcan asímismos como pertenecientes a un grupo Nativo Americano y que hablen el idioma Tarasco, Nayeri, Mazahua o Wirrárika.

Exclusión: Individuos del grupo Mestizo, que no hablen ninguna de las lenguas antes mencionadas.

Eliminación: Individuos de los cuales no se pueda obtener la información requerida.

Cada participante firmó una carta de consentimiento aprobada por el Comité de ética de la Universidad de Pensilvania con el número 803115 correspondiente al Proyecto Genográfico: Análisis genético y molecular de los pueblos indígenas de América del Norte. La carta no sólo consentía la toma de muestra, sino que además contenía un breve cuestionario para conocer algunos datos genealógicos para detectar la existencia de individuos emparentados.

El número de localidades y el tamaño de muestra total de las poblaciones de estudio se muestran en la Tabla 3.



Figura 13. Zonas de muestreo en Michoacán y Nayarit.

Tabla 3. Número de individuos y localidades muestreadas para este estudio.

Población	Número de individuos	Localidad
P'urhépecha	57	Ihuatzio, Mich.
		Ichupio, Mich.
		Ucazanatacua, Mich.
		San Pedro Zipiejo, Mich.
		La Tenencia de Crecencio, Mich.
Mazahuas	26	El Rincón, Mich.
		El Capulín, Mich.
Nayeri	20	Santa Cruz de Guaybel, Nay.
Wurrárikas	19	Potrero de la Palmita, Nay.

Estas poblaciones fueron seleccionadas ya que poseen características demográficas diferentes: se tratan de grupos lingüísticos diferentes, con localizaciones geográficas diferentes y diferentes historias demográficas durante la conquista (Material suplementario).

- **Determinación del haplotipo mediante en PCR punto final**

El DNA genómico fue extraído a partir de leucocitos de sangre periférica y/o saliva con el kit *PureGene Genomic DNA Purification Kit* (Qiagen, Genra System, Minneapolis MN, EUA) el cual está basado en el proceso de precipitación salina o *salting out*.

La región específica masculina del cromosoma Y se caracterizó en cada hombre por medio de 17 STRs (DYS19, DYS385a/b, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS635 y GATA-H4) usando el kit comercial de amplificación AmpF/STR Yfiler™ PCR (Applied Biosystems, Carlsbad, CA, USA), de acuerdo con las especificaciones del proveedor y empleando un termociclador Veriti de 96 pozos (Applied Biosystems, Carlsbad, CA, USA) (Tabla 4). Los amplicones resultantes fueron analizados por medio de electroforesis capilar en un secuenciador ABI Prism 3130XL utilizando el *software* GeneMapper ID v.3.2. (Applied Biosystems, Carlsbad, CA, USA).

- **Parámetros de Genética Poblacional**

Determinación de los macro-haplogrupos del Cromosoma Y

El Ht de cada muestra fue introducido al *software online Haplogroup Predictor* (<http://www.hprg.com/hapest5/>) (Athey, 2012) el que por medio de una inferencia Bayesiana con los datos de *Family Tree DNA* (FTDNA) y de *Y-Search*, correlaciona el Ht con un macro-HG en particular (Figura 14).

Tabla 4 Estructura de los YSTRs que se analizaron en este proyecto.

STR	Estructura	Alelos	Tamaño (pb)
DYS19	(TAGA) _n	10-19	175-225
DYS385 a/b	(GAAA) _n	7-28	240-330
DYS389I	(TCTG) _n (TCTA) _n	9-17	140-190
DYS389II	(TCTG) _n (TCTA) _n (TCTG) _n (TCTA) _n	9-17	255-315
DYS390	(TCTG) _n (TCTA) _n ACTA(TCTA) _n (TCTG) _n (TCTA) _n	17-28	195-250
DYS391	(TCTA) _n	6-14	150-200
DYS392	(TAT) _n	6-17	300-350
DYS393	(AGAT) _n	9-17	100-150
DYS437	(TCTA) _n (TCTG) _n (TCTA) _n	13-17	160-220
DYS438	(TTTTTC) _n	6-14	225-270
DYS439	(GATA) _n	9-14	200-240
DYS448	(AGAGAT) _n N ₄₂ (AGAGAT) _n	17-24	280-340
DYS456	(AGAT) _n	12-18	90-130
DYS458	(GAAA) _n	13-20	125-160
DYS635	(TCTA) _n (TGTA) _n (TCTA) _n (TGTA) _n (TCTA) _n (TGT A) _n (TCTA) _n TC-A (TCTA) _n	17-26	250-290
GATA-H4	(TAGA) _n N ₁₂ (GATC) ₂ AA (TAGA) ₄	8-13	120-155

Parámetros de diversidad

Las frecuencias haplotípicas y la diversidad haplotípica (HD) fueron estimados por medio del *software* Arlequin v. 3.5⁸². El número de haplotipos únicos fue estimado por un conteo directo. La comparación entre las poblaciones se analizó por medio de distancias genéticas (R_{ST}) y análisis de varianza molecular (AMOVA) con el *software* Arlequin v. 3.5.

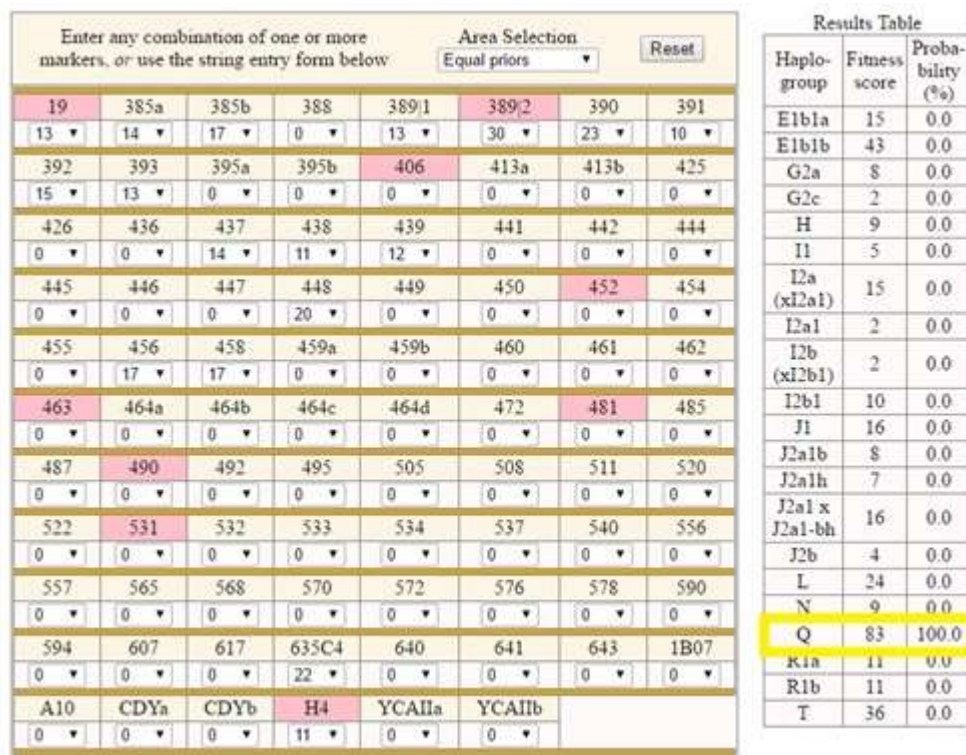


Figura 14. Ejemplo del análisis de un haplotipo por medio del software Haplogroup Predictor. El *Fitness score* indica que existe una elevada probabilidad de que este Ht pertenezca al linaje Q.

- **Caracterización de los marcadores de alta definición por medio de PCR en tiempo real**

Para determinar los HG Amerindios encontrados, se emplearon cuatro marcadores tipo SNP. Estos marcadores incluyeron: M242, MEH2, L54 y M3. La genotipificación se llevó a cabo en un termociclador C1000 (BioRad Life Science, Hercules, CA, USA) por medio de ensayos TaqMan® (Applied Biosystems, Carlsbad, CA, USA). La combinación de los marcadores STR y SNP determinaron los linajes paternos en los individuos estudiados.

- **Realización de redes**

Con el fin de conocer la relación filogenética de los Ht y complementar la información acerca de los patrones de diversidad, se realizaron redes filogenéticas con el software *Network 4.613* utilizando el método *median-joining* (MD)⁸³. Los haplotipos utilizados consistieron de 15-YSTRs, el locus

DYS385a/b fue excluido dado que representa un locus STR duplicado. El peso que se le dio a cada marcador fue basado en los reportes previos de Roewer et al, y corresponden al inverso de la tasa de mutación de cada marcador, permitiendo al *software* tomar en cuenta que carácter es más importante que otro. El aumento de este valor indica eventos (cambios) con una menor probabilidad de ocurrir y por lo tanto son significativos cuando suceden; de la misma manera un valor menor indica que el evento tiene una mayor probabilidad de ocurrencia ^{4,84}.

Comparación con otras poblaciones

Para comparar nuestros datos con otras poblaciones, se obtuvo información haplotípica de reportes previos de poblaciones del Sur y Norte de América, así como de poblaciones de Siberia y de la zona de Altai (Tabla S11, Material suplementario).

Resultados

- **Determinación de los haplotipos por medio de marcadores tipo STR**

Los haplotipos de los 17-YSTR en los 122 individuos analizados se muestran en las tablas suplementarias S1-S4. El 80% de los individuos estudiados no mostraron ninguna relación de parentesco aparente, por lo que solo este porcentaje fue empleado para realizar los análisis posteriores. De los 98 haplotipos analizados, 79 (81%) fueron haplotipos únicos (*singletons*) y ninguno compartido entre las poblaciones. El número de haplotipos únicos y los valores de diversidad (h) de cada población se muestran en la Tabla 5.

Tabla 5. Número de haplotipos estudiados, singletons y valores de diversidad de las poblaciones estudiadas.

Población	Número total de haplotipos	Haplotipos únicos	Diversidad (h)
Mazahua	23	18	0.560 ± 0.052
Nayeri	13	10	0.517 ± 0.052
P'urhépecha	46	38	0.554 ± 0.052
Wirrárika	16	13	0.472 ± 0.052

Macro-haplogrupos del cromosoma Y

Por medio del uso del software *Haplogroup Predictor* encontramos que el HG más frecuente fue Q (0.74), seguido por los linajes R1b (0.09), E1b1b (0.07), L (0.03), E1b1a (0.02), T (0.01), J2a1b (0.01), J1 (0.01) y G2a (0.01) (Tabla 6).

Los siguientes análisis solo se realizaron con las muestras cuyo HG fue Q, ya que el objetivo era caracterizar población Amerindia.

Tabla 6. Frecuencias de los diferentes HG encontrados en las poblaciones Amerindias analizadas

Población	Q	E1b1a	E1b1b	G2a	J1	J2a1b	L	R1b	T
Mazahua	0.7		0.04			0.04		0.22	
Nayeri	0.54		0.23		0.08			0.15	
P'urhépecha	0.87	0.043					0.22	0.043	0.22
Wirrárika	0.63	0.06	0.13	0.06			0.13		

- **Determinación los HG específicos de los individuos que pertenecen al macro-HG Q por medio de marcadores tipo SNP.**

Una vez que encontramos las muestras con mayor probabilidad de pertenecer al macro-HG Q realizamos PCR en tiempo real para confirmar el linaje Nativo Americano y los diferentes sublinajes reportados (QM242, QMEH2, QL54, o QM3).

Como se puede observar en la Tabla 7, el linaje más prominente entre los Nativos Americanos de México fue de QM3 (0.77) seguido por QL54 (0.23). No obstante, cada población presentó sus particularidades.

Tabla 7. Proporción de los sub-haplogrupos en las poblaciones Amerindias analizadas.

Población	Q-M3	Q-L54	Población	Q-M3	Q-L54
P'urhépechas	0.65	0.35	Nayeri	1.00	0
Mazahuas	0.75	0.25	Wirrárikas	1.00	0

- **Parámetros de genética poblacional y diversidad.**

Dado que los dos linajes encontrados dentro del HG Q provienen de migraciones diferentes que tuvieron procesos demográficos diferentes, los fueron realizados separando a Q-M3 de Q-L54 ⁷⁷.

Q-M3

La diversidad alélica que existe dentro de estas poblaciones fue examinada por medio de las frecuencias alélicas, así como por el número de alelos diferentes (k) en cada locus (Tablas S5-S8). Los marcadores más informativos fueron variables entre poblaciones, en donde DYS458 ($k= 5, h= 0.8$) fue el más informativo para la población P'urhépecha, mientras que DYS385b lo fue ($k= 6, h= 0.848$) para la población Mazahua. Las poblaciones Nayeri y Wirrárika mostraron poca diversidad ($k= 3, h= 0.762$ y $k=3, h= 0.711$, respectivamente).

La diversidad haplotípica se realizó utilizando 15 YSTRs en lugar de los 17, por lo que ya fue mencionado con anterioridad respecto al marcador DYS385a/b. Como se puede observar en la Tabla 8, los grados de diversidad haplotípica fueron muy cercanos a 1.00, el que corresponde a la mayor diversidad, lo anterior es congruente con los valores de *mean pairwise difference*, que también fueron muy altos indicando las diferencias entre los Ht.

Tabla 8. Diversidad haplotípica y *pairwise differences* de los Ht Q-M3.

Población	n	Ht diferentes	Diversidad Haplotípica	<i>Mean pairwise difference</i>
P'urhépecha	26	22	0.9846 ± 0.0160	7.073846 ± 3.430798
Mazahua	12	10	0.9545 ± 0.0569	6.606061 ± 3.356741
Nayeri	7	5	0.9048 ± 0.1033	7.190476 ± 3.838721
Wirrárika	10	7	0.9111 ± 0.0773	4.622222 ± 2.476494

Posteriormente se realizó un Análisis Molecular de Varianza (AMOVA), que permitió analizar las variaciones moleculares dentro del grupo. Los resultados mostraron una elevada variación genética dentro de

las poblaciones (79%), lo que es congruente con los datos de diversidad encontrados, mientras que la variación entre las poblaciones fue de 21%. Esta variación es estadísticamente significativa ($P=0.00059$) mostrando una diferencia importante entre las poblaciones analizadas (Tabla 9A).

Tabla 9. Porcentajes y fuentes de variación obtenidos por AMOVA. A. Poblaciones sin criterios de agrupación. B. Poblaciones agrupadas por características lingüísticas.

A

Fuente de variación	Porcentaje de Variación	P
Entre poblaciones	20.62	0.00059
Dentro de las poblaciones	79.38	0.00059

B

Fuente de variación	Porcentaje de Variación	P
Entre grupos lingüísticos	9.02	0.33743
Entre las poblaciones dentro de los grupos	12.09	<0.00001
Dentro de las poblaciones	78.89	0.00030

Se ha demostrado que los patrones genéticos en algunas poblaciones se distribuyen de acuerdo a los patrones lingüísticos de las poblaciones, por ello se jerarquizaron los grupos de acuerdo a la lengua: P'urhépecha (Tarasca), Mazahua (Oto-Mangue) y Nayeri y Wirrárrika (Uto-Azteca) ⁴. El valor de la variación se mantiene entre los individuos de cada población (~79%), mientras que la variación por las características lingüísticas no es significativa por lo que los cambios en los haplotipos no dependen de este rasgo (Tabla 9B).

Con la finalidad de soportar estos hallazgos, se realizó un cálculo de distancia genética empleando la distancia de Slatkin que incorpora los procesos de *stepwise mutation* en los microsatelites R_{ST} y mide la

diferenciación poblacional en función de sus frecuencias alélicas ⁸⁵. El rango de valores que puede existir es de 0 a 1, en donde el valor 0 nos indica que no existe una diferencia entre las poblaciones, mientras que un valor de 1 implica la completa diferencia entre poblaciones. Los valores intermedios se interpretan como sigue:

- De 0 a 0.05 indican poca diferencia
- De 0.05 a 0.15 indican una diferencia moderada
- De 0.15 a 0.25 una alta diferencia
- Valores mayores a 0.25 una diferencia muy alta

En la Tabla 10 se observan los valores de R_{ST} , todas las poblaciones divergen entre ellas, destacándose la población Wirrárika, la cual tiene los valores más altos. En el caso de los Nayeri y los Mazahua, existe un valor negativo, estos valores se toman como 0, pero esto se debe al número de muestra.

Tabla 10. Valores R_{ST} que indican una cuantificación de la variación de las frecuencias haplotípicas dada la distancia genética que existe entre las poblaciones. Entre paréntesis se muestra la significancia estadística.

Población	P'urhépecha	Mazahua	Nayeri	Wirrárika
P'urhépecha	0.000			
Mazahua	0.169 (P= 0.01574)	0.000		
Nayeri	0.106 (P= 0.11227)	-0.012 (P= 0.45065)	0.000	
Wirrárika	0.261 (P= 0.00366)	0.393 (P< 0.00001)	0.354 (P= 0.00178)	0.000

De manera más gráfica, los valores de R_{ST} pueden observarse por medio de la intensidad de colores en una gráfica como la Figura 15, en donde a mayor intensidad del color azul significa una mayor diferenciación de la población.

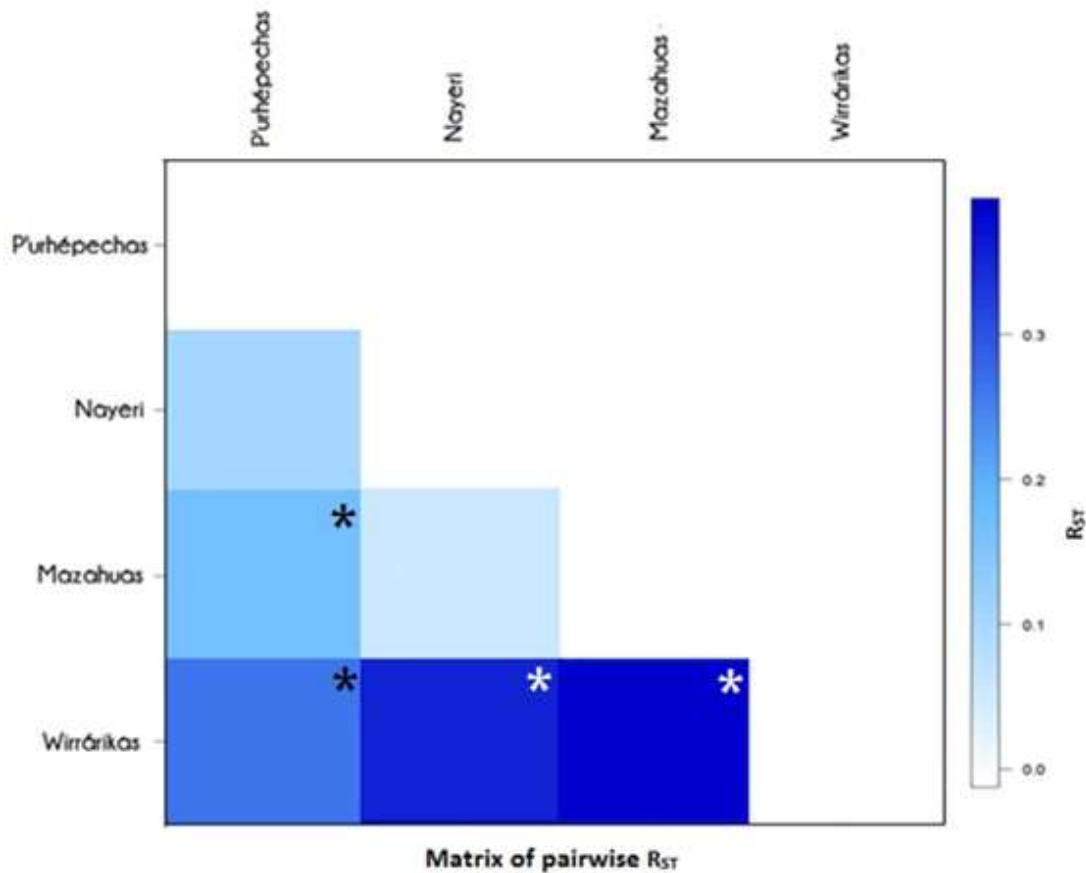


Figura 15. Gráfica de calor que representan los valores de R_{ST} . El asterisco (*) resalta las diferencias significativas ($P < 0.05$).

Q-L54

Este HG fue encontrado solamente en los P'urhépechas y en los Mazahuas. Las frecuencias alélicas se muestran en las tablas S9 y S10 del material suplementario. De entre los marcadores analizados, DYS458 fue, en los P'urhépecha, el más informativo ($k= 5$, $h= 0.769$), mientras que, en los Mazahua, fue DYS89II ($k= 3$, $h= 1$). Hay que tomar con cuidado estos datos debido al bajo número de muestra.

Los análisis indican una alta diversidad a nivel haplotípico. Para el caso de los P'urhépechas, éstos presentaron una diversidad haplotípica incluso mayor a la encontrada en las muestras Q-M3 de la misma población, así como un mayor número de *pairwise differences* entre los haplotipos de esta

población. Llama la atención de que, a pesar del bajo número de muestra, la población Mazahua parece ser muy diversa (Tabla 11).

Tabla 11. Diversidad haplotípica y pairwise differences en las muestras Q-L54 de este estudio.

Población	n	Ht diferentes	Diversidad Haplotípica	Mean pairwise difference
P'urhépecha	14	10	0.9231±0.0604	7.4835±3.7206
Mazahua	3	3	1.0±0.2722	7.0±4.527693

Tanto en los análisis de R_{ST} , como en los de AMOVA, los valores obtenidos no fueron significativos, por lo que no se muestran los resultados de estos análisis.

- **Análisis de redes filogenéticas**

Usualmente el uso de marcadores genéticos genera filogenias conflictivas, debido a que en el genoma existen patrones de duplicación, perdidas, recombinación e intercambio, sobre todo en eventos de expansión poblacional ⁸⁶. Las redes son representaciones filogenéticas que permiten el análisis de la relación evolutiva, la gran ventaja es que se pueden representar eventos de reticulación, es decir cambios paralelos independientes, muy característico de la evolución del genoma ^{87,88}.

Las redes son interpretadas de la siguiente manera:

- Cada Ht está representado por un círculo, el tamaño del círculo representa la frecuencia de ese Ht en la población.
- Los Ht están unidos entre sí por ramas, el largo de éstas representa la distancia (medida en número de mutaciones) entre los Ht.
- Existen nodos, que son vectores de unión entre diferentes Ht, este puede ser tomado como un Ht hipotético que se requiere para que se

siga el principio de parsimonia, lo que significa el menor número de cambios de un Ht al otro.

De manera general, las redes de ambos HGs mostraron una alta ramificación, lo cual corresponde a la elevada diversidad encontrada en estas poblaciones, pero también puede sugerir la influencia de eventos demográficos recientes (cuellos de botella, aislamiento, deriva) en estas poblaciones. Lo anterior está relacionado con la gran cantidad de cambios (mayores a 10), indicando una diferenciación en cada una de las etnias (Fig. 16-17 y Fig. S1-S6 del Material suplementario). También es importante observar que no hay un patrón característico con respecto a la etnia, la zona geográfica o la lengua, lo cual indica que estas etnias comparten un ancestro común. Interesantemente, las redes de QM3 muestran la formación de agrupaciones, conocidas como paragrupos, que se separan del resto y que indican la presencia de sublinajes dentro de este HG. Los paragrupos están enmarcados dentro de los círculos azules y marcados con asterisco (Fig. 17).

En ambos linajes, se vuelve a destacar la gran presencia de Ht únicos separados por varias mutaciones (Fig. 16).

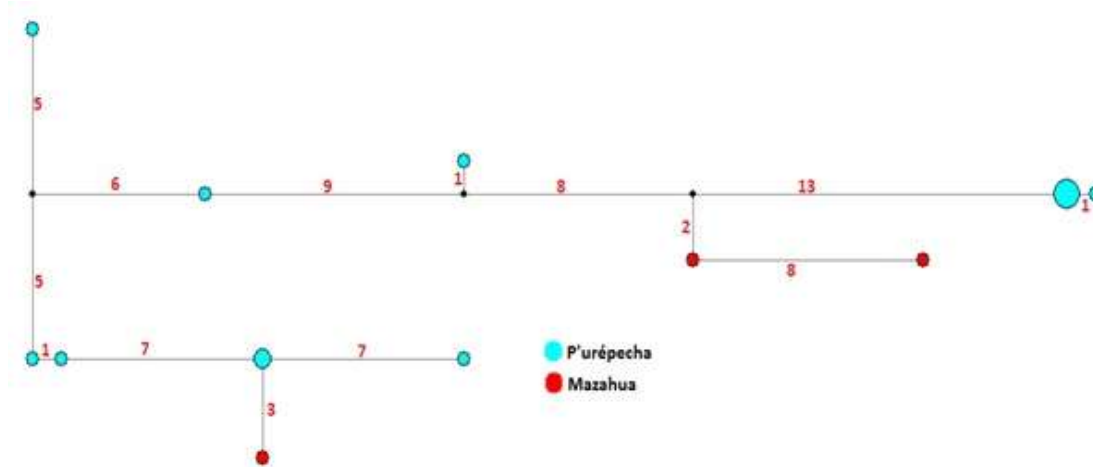


Figura 16. Red del HG Q-L54 de las poblaciones Amerindias utilizando 15 Y-STRs.

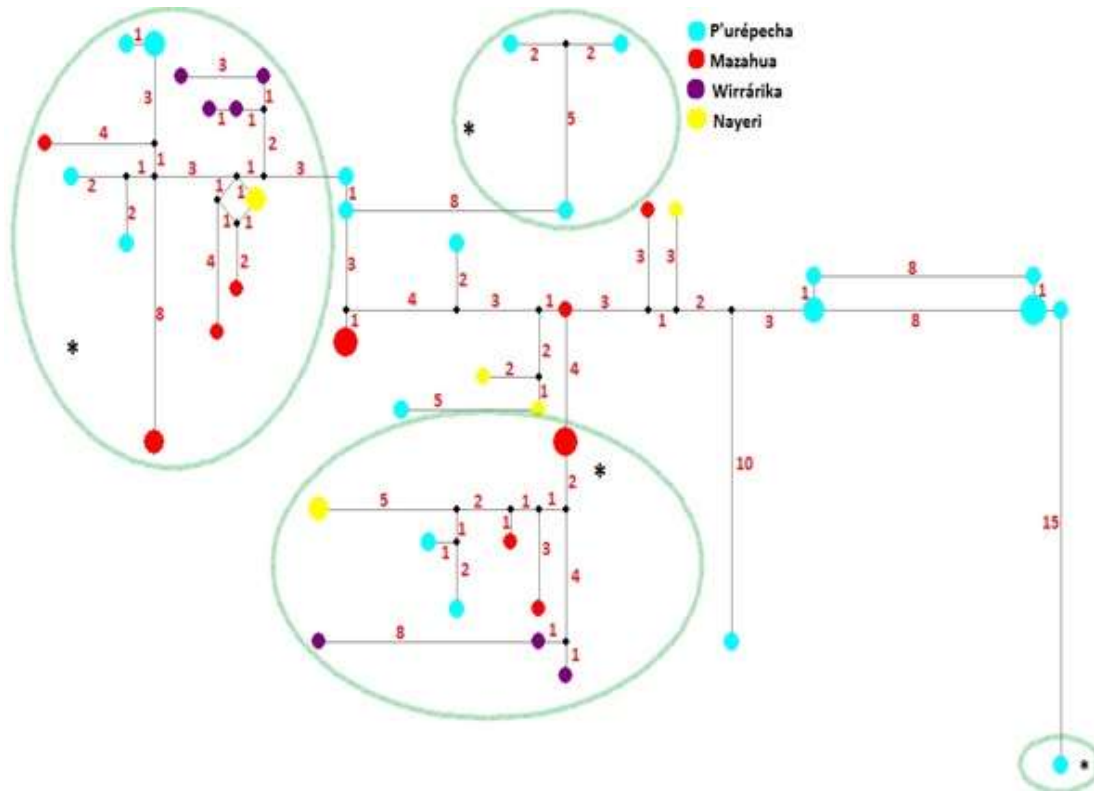


Figura 17. Red del HG Q-M3 de las poblaciones amerindias utilizando 15 Y-STRs. En verde y con asteriscos se marcan los posibles paragrupos.

Comparación con otras poblaciones

Para conocer la relación de nuestras poblaciones con otras poblaciones que comparten ancestría, se compararon nuestros datos con datos previamente reportados, incluyendo poblaciones Amerindias y Asiáticas (Tabla S11 del Material suplementario).

En el análisis de redes que realizamos para Q-M3 se puede observar la gran diversidad de los grupos Nativos de México y la conexión ancestral con las poblaciones del norte del continente. Interesantemente, se muestra la presencia de dos probables paragrupos; de arriba hacia abajo, el paragrupo más pequeño sugiere una relación ancestral estrecha entre los Pimas (Arizona y Sonora), Nahuas, Mazahuas, P'urhépechas y Jemez (Nuevo México). Mientras que el paragrupo inferior esta conformado por las poblaciones Mixteco (Guerrero, Oaxaca y Puebla), Sioux (Norte de E.U. y Sur de Cánada),

Metis (Oeste de C nada), Nahuas, P'urh pechas, Mazahuas, Nayeri y Wirr rikas (Fig. 18).

De la misma forma, las muestras Q-L54 se asociaron principalmente con poblaciones del Norte de Am rica, como los Cheyenne, los Cherokee y los Sioux, y con las poblaciones de Altai, mostrando una mayor distancia con las poblaciones de Sudam rica, Siberia y la Rep blica de Tuva (Sur de Rusia, norte de Mongolia) (Fig. 19). Llama la atenci n el alto n mero de reticulaciones, lo cual indica una falta de definici n de los marcadores moleculares en la red.

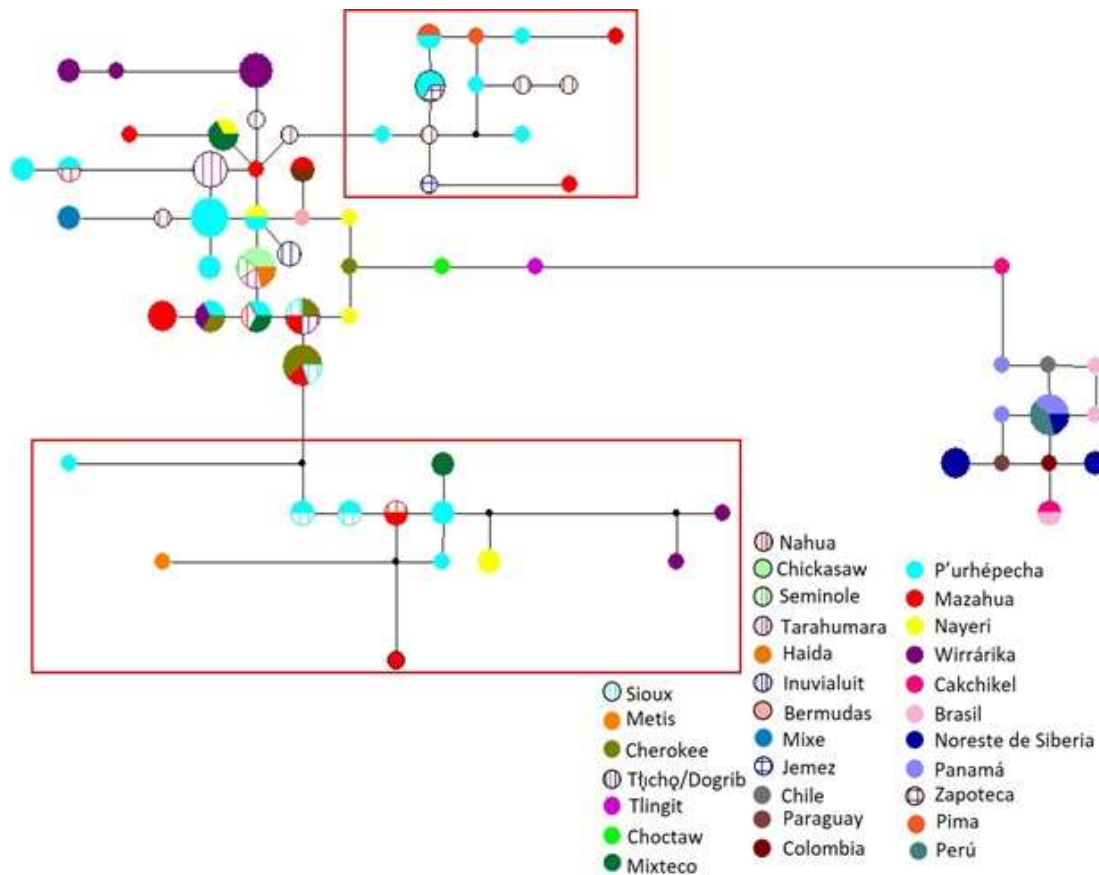


Figura 18. Comparaci n de poblaciones Q-M3. En los rect ngulos rojos se marca la posible existencia de paragrupos de Q-M3.

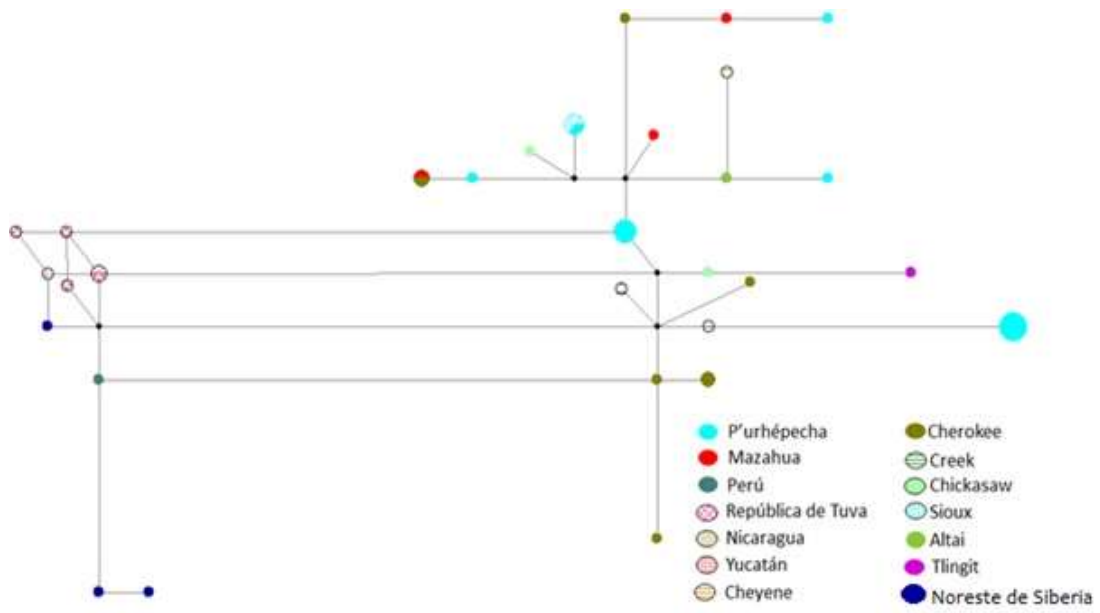


Figura 19. Comparación de poblaciones Q-L54.

Discusión

La población Mexicana tiene una estructura genética compleja causada por su historia demográfica, por el proceso de conquista y el reciente mestizaje. Estas características tienen como consecuencia que la búsqueda de genes candidatos para la prevención y tratamiento de enfermedades comunes se vea afectada por factores de confusión como la estructura genética. En este trabajo, se caracterizó la diversidad genética y ancestral de cuatro poblaciones Nativas Americanas con el fin de ser empleadas como posibles modelos para la realización de estudios de asociación genética de enfermedades complejas.

Como se mencionó en la introducción, la actual población mestiza en México fue originada durante la conquista, hace aproximadamente 500 años, donde los conquistadores Españoles, y los esclavos Asiáticos y Africanos, contribuyeron a los patrones genéticos al mezclarse con las poblaciones Amerindias residentes ⁸⁹. Este mestizaje fue asimétrico, observándose en las distribuciones de la ancestría materna y paterna, siendo que la gran mayoría de la población pertenece a linajes Amerindios maternos ($\approx 93\%$) ⁹⁰. De manera diferente, en la ancestría paterna los linajes ancestrales son diversos, llegándose a detectar hasta dieciséis linajes diferentes, sólo en el centro del país, de acuerdo a Santana et al, en el 2014; en cambio, en el resto del país, Martínez-Cortés et al, en el 2012, encontraron dieciocho linajes diferentes, indicando no sólo una enorme diversidad sino a su vez la distribución diferencial de los linajes, con diferentes frecuencias en las poblaciones mestizas ^{91,92}.

Nuestros resultados evidencian también esta incorporación de linajes del cromosoma Y en las poblaciones nativas, las cuales no están exentas de mestizaje, como se había mencionado en los reportes anteriores de Paéz-Riberos et al, en 2006; Rangel et al, en 2008; y Sandoval et al, en 2012; donde,

aunque la ancestría Amerindia es la más frecuente, hay presencia de otros linajes ^{78,93,94}.

Evidencia de este mestizaje es la presencia de linajes Europeos, como R1b, HG existente en alta frecuencia en España, concordante con la llegada de los conquistadores, y su relación con las mujeres Amerindias en el continente ^{89,95}. La alta frecuencia de HG españoles en las poblaciones mestizas y en los grupos étnicos, tiene que ver con el poderío que ejercían los españoles sobre las poblaciones Nativas, lo que incluía las violaciones como parte de las tácticas de terror, y, posteriormente a la conquista, la práctica del derecho de pernada (derecho del señor para pasar la noche de bodas con la nueva esposa de su siervo), ejercida desde la época Babilónica y que se extendió, en México, hasta inicios de la Revolución ^{78,89,95-103}. Esta ancestría Europea también está representada por el HG G2a, correlacionada con la presencia de los comerciantes Vascos en los puertos de la Nueva España ^{89,104}.

Aunado a esto, individuos con ancestría del Medio Oriente formaban parte de estos grupos de conquistadores Españoles, demostrado por la presencia de HG como J21a y J1, debido a la presencia de cripto-judíos que escapaban de la inquisición Española ^{89,104,105}. Otra razón de la presencia de individuos del Medio Oriente junto con los Españoles, es la invasión Árabe durante el s. VIII a la Península Ibérica, que probablemente introdujo HGs como E1*, también encontrado en nuestras muestras ¹⁰⁴. Inclusive, la presencia de la ancestría del Medio Oriente puede deberse a la migración de población Libanesa a México a finales del s. XIX, que se dio debido a la escasez de recursos en Líbano causada por el régimen de Mutasarifiya, iniciado en 1861 ¹⁰⁶.

Por otro lado, la ancestría de Asia, como la India o Filipinas, representada por el HG L en nuestras muestras, pudo haber llegado junto con los sirvientes personales de los Españoles, que eran esclavos asiáticos cuya

compra solía ser más barata que la de los Africanos, y que llegaron a la Nueva España desde 1570 ⁷⁰. Estos esclavos, entraban por el Pacífico, principalmente por el puerto de Acapulco y, se cree, trabajaban también en las minas debido a la gran mortandad que existía entre los esclavos Nativos Americanos; en otras poblaciones de Argentina, también se han encontrado HG ligados a estas zonas como lo son el HG K y F ^{70,89,96}.

Esta disminución de los esclavos Nativos, también disparó el aumento del tráfico de esclavos Africanos, cuya incorporación a la población Nativa se evidencia con la presencia de HGs E1b1a, E1b1b y T, de la parte central y Oeste de África, llegando a los puertos de la Nueva España u otras partes del continente, por lo que también se encuentra en otras poblaciones del Sur como Colombia, Argentina, Puerto Rico y Perú ^{92,96,107-109}.

Enfocándonos en la ancestría Amerindia, esta es la segunda en frecuencia en las poblaciones mestizas, con $\approx 21\%$ ⁴⁹. Lo contrario sucede en nuestras muestras donde es la ancestría dominante ($\approx 80\%$), al igual que en otras poblaciones Nativas de México como los Triquis, Tarahumaras, Pimas, Tzotziles, Mixes, Mixtecos, Zapotecos y Tlapanecos (70-100%), y en menor medida en poblaciones nativas urbanizadas como los Nahuas o los Otomíes ^{78,92,94}. Esto podría significar que las poblaciones que conservan una mayor frecuencia de ancestría paterna Amerindia, son poblaciones que se encontraban en regiones inaccesibles para los conquistadores y que además están aisladas de las poblaciones mestizas modernas ^{89,110}.

Dentro de este linaje Amerindio, se observa una estructura interesante dada por los linajes representativos de los Nativos Americanos QL54 y QM3, siendo el último el más frecuente en general (0.77). Esta estructura tiene que ver con su historia demográfica, la cual tiene puntos en común. En el 2012, Dulik et al, mencionan que QM3 es un linaje derivado de QL54, siendo este último proveniente de la zona sur de la República de Altai, las planicies del

Norte de Mongolia y la República de Tuva ⁷⁵. Posteriormente se cree que portadores de QL54/QMEH2 migraron hacia Beringia, estableciéndose en la zona, donde también, hace ≈ 13.4 miles de años, aparecieron los portadores de QM3 ^{75,111,112}. Esta relación con las poblaciones ancestrales de la zona de Altai, se puede observar en nuestra red de QL54, estando cercanas a nuestras muestras P'urhépecha y Mazahua, así como la cercanía de poblaciones de Norteamérica como los Cheyenne y los Cherokee, en coherencia con los modelos migratorios propuestos de Norte a Sur ¹¹³⁻¹¹⁵. Esta cercanía no se observa en la red de QM3, en cambio, en esta, existe una gran distancia de nuestras muestras con las poblaciones del Noreste de Siberia. Algunos autores indican posibles contactos posteriores a la primera entrada a América, entre poblaciones de Siberia y Amerindias, debido al comercio y al tráfico de esclavos esquimales, de ahí, es posible que esta distancia se deba a que son Ht con historias demográficas diferentes a las de nuestras muestras¹¹⁶.

Además, el número de olas y las vías migratorias hacia América no se han definido por completo, sin embargo, se cree que no fue solo una. Se estima que una de las rutas, el corredor continental estuvo cerrado hasta hace $\approx 13,500$ años, lo cual no concuerda con algunas zonas arqueológicas, como la de Monte Verde en Chile (18,500-14,500 años), Cactus Hill, E.U. (18,000-15,070 años), o Vale da Pedra Furada en Brasil ($\approx 20,000$ años) ¹¹⁷⁻¹²⁰. Por lo que la alternativa es una ruta costera, donde la evidencia geológica indica que el paso por la costa Noroeste fue posible hace $\approx 20,000$, pero, con menor riesgo y biológicamente posible hace $\approx 17,000$ años ¹²¹. Esta ruta está apoyada además, por puntas talladas encontradas en costas del Pacífico Asiático y Americano, registros paleontológicos de la flora y fauna, modelaje de los niveles del mar, y registros paleoclimáticos ¹²¹⁻¹²⁴. Finalmente, simulaciones de la colonización de América, sugieren que los patrones de diversidad genética en el continente

se relacionan más con una primera expansión en Sudamérica, por la vía costera, seguida de una migración poblacional en el norte, asociada con la ruta continental ^{113,125}.

Esto apoya las observaciones en nuestras redes conjuntas, tanto de QM3 como de QL54, donde hay una gran distancia que separa a las poblaciones de Sudamérica de nuestras poblaciones, debido a las diferentes rutas que siguieron estas poblaciones durante la entrada a América. Además, estas diferencias en las rutas de entrada pueden ser parte de los factores que determinaron la frecuencia diferencial de estos haplogrupos en Latinoamérica. De acuerdo a Battaglia et al, en 2013, las distribuciones de los HG QL54 y QM3 son diferentes en Sudamérica; sus resultados muestran que QM3 es el HG con mayor frecuencia (79%) y distribuido por toda esta zona ⁵⁸. El HG QL54 por otro lado, sólo se encontró en las muestras localizadas de Mesoamérica hacia el Sur en menor frecuencia (18%), es por esta razón probablemente que no encontrásemos QL54 en las muestras Nayeri y Wirrárika debido a que su localización original era fuera de Mesoamérica ⁵⁸.

La entrada a América a su vez, marcó los patrones genéticos de las poblaciones Amerindias, debido a la presencia de las fuerzas evolutivas. Haciendo una comparación a nivel global, se ha observado que los niveles de diversidad genética disminuyen conforme existe una mayor distancia de África, lugar de origen de la especie humana y, por lo tanto, base de la diversidad genética de la misma ⁹. Este fenómeno se adjudica a las migraciones por las cuales se pobló la Tierra, lo que conllevó a una serie de eventos como la fragmentación y disminución de las poblaciones, lo que a su vez se ve afectado por las fuerzas evolutivas, como la deriva génica, resultando en un efecto consecutivo de cuellos de botella, cuyo último paso, por así decirlo, en el poblamiento del Planeta, fue la entrada a América ^{9,126-128}.

Caracterizando la diversidad dentro de estos HGs, alélicamente tienen una tendencia hacia una baja diversidad en todas nuestras poblaciones. Decimos esto, ya que al compararlos con poblaciones mestizas, los alelos informativos de nuestra muestra tienen un menor número de alelos (k); la causa de una mayor k es el mestizaje, lo cual, en este caso, aumentó la cantidad de cromosomas Y en *pool* genético de las poblaciones mestizas^{49,126}. Este fenómeno también se observa en poblaciones Amerindias más mestizadas como los Nahuas, mientras que en las poblaciones menos mestizadas, la k es similar a la de nuestras muestras, indicando una pérdida de diversidad genética en las poblaciones Amerindias^{93,94}. En el 2012, Gorostiza et al, asumen un modelo de diversidad genética en donde la población inicial crece dando lugar a poblaciones derivadas y aisladas^{111,129}. Este modelo, de inicio, implica la pérdida de diversidad, como lo observan Wang et al, en 2007, y Wall et al, en el 2011, a nivel continental y como lo observamos nosotros a nivel alélico^{128,130}.

Sin embargo, la zona de Mesoamérica ha sido reconocida como un sitio de diversificación genética, de acuerdo a los resultados de Sandoval et al, en 2012, de Regueiro et al, en 2013 y de Battaglia et al, en el mismo año^{58,78,111}. Similar a esta observación, nuestros datos muestran una alta diversidad haplotípica, siendo mayor en los Mazahua (0.95) y P'urhépecha (0.98) (localizadas dentro de Mesoamérica) y menor en poblaciones al norte de Mesoamérica, los Nayeri (0.90) y los Wirrárikas (0.91). Esto también se evidencia en las redes de nuestras poblaciones debido al gran número de *singletons*.

Mientras que las ramificaciones en nuestras poblaciones indican la existencia de expansiones poblacionales. Continuando con el modelo propuesto por Gorostiza et al, en 2012, ellos mencionan que a nivel regional no se observa una pérdida de diversidad a nivel haplotípico, al igual que

nuestras muestras, debido a esta serie de expansiones poblacionales, junto con movimientos secundarios y flujos génicos que aumentaron y diversificaron el *pool* haplotípico ¹²⁹.

Con respecto a las expansiones poblacionales, Regueiro et al, en el 2013, sugieren que la introducción de la agricultura, y otras tecnologías, permitieron esta expansión poblacional asociada a su vez con la diversificación haplotípica ¹¹¹. En México, esta expansión se observa en poblaciones con grandes distribuciones como los Aztecas, P'urépechas, y Wirrarikás, que interactuaban, por medio de la guerra y/o arreglos comerciales con otras culturas, como los Mazahua y los Nayeri ^{89,131,132}. Esto conllevó a una estructura en las poblaciones Amerindias como la reportada por Moreno et al, en el 2014, donde existe un matiz de la variación a lo largo del país, evidenciado por agrupaciones de poblaciones Amerindias asociadas a su localización y a las interacciones previas entre ellas ¹³³.

Esta estructura a su vez se refleja en el número de haplotipos únicos de nuestras poblaciones, dando un alto porcentaje de variación dentro de las poblaciones de acuerdo a nuestros resultados del AMOVA (79.38%, $P=0.00059$). Sin embargo, al no compartir H_t , puede indicar la divergencia y existencia de características genéticas específicas de las poblaciones ¹³³. Esta divergencia la observamos por medio del AMOVA, ya que la variación entre las poblaciones era alta (20.62%, $P=0.00059$), confirmado por lo valores de las distancias genéticas dadas por el R_{ST} (0.169-0.393, $P<0.01$), indicando una heterogeneidad entre ellas.

La existencia de las grandes culturas conlleva a la aparición de diferentes factores, como la cultura (el lenguaje o las jerarquías religiosas) o la localización (aislamiento o cercanía a rutas comerciales), que pueden ser los causantes de la diferencia o similitud genética. En 2013, Roewer et al, describen que, en muchos casos, los patrones genéticos siguen una

correlación geográfica y lingüística, ya que son vehículos por los cuales existe un flujo entre las poblaciones, viéndose limitados por factores como el tiempo, las fragmentaciones y la hegemonía de algunas poblaciones ^{3,4}. Ellos encontraron que en Sudamérica esta correlación no existe, más que a pocos Km de distancia, adjudicándose a los factores antes mencionados, aunados a la baja cantidad de individuos Amerindios que existen en la actualidad; además de ser la causa por la que se encuentra una alta heterogeneidad entre las poblaciones Nativas ⁴.

Las lenguas en América tienden a clasificarse en tres grandes familias, la Esquimo-Aleutino, la Na-Dené y la Amerindia, las primeras dos se distribuyen principalmente en Canadá mientras que la Amerindia está caracterizada por una serie de sub-clasificaciones a lo largo del continente, que pueden llegar a ser muy diferentes entre sí ¹³⁴. Los Nayeri y Wirrárika, del tronco Uto-Azteca, formado por dos grandes ramas, la Meridional, y la Septentrional, habladas en el sur de E.U. México y zonas de Mesoamérica; estas dos poblaciones, particularmente, forman una rama conocida como Cora-chol, por lo que podríamos suponer una cercanía genética, de acuerdo a Roewer et al ^{4,135,136}. Por otro lado, el tronco Oto-Mangue, al cual pertenecen los Mazahua, está compuesta por ocho ramas, incluyendo lenguas aisladas, y se distribuye en zonas de Veracruz, San Luis Potosí, Guerrero y Oaxaca ¹³⁷. Finalmente, la lengua Tarasca o *P'urhé*, no tienen parentesco con ninguna de las lenguas habladas en México, pero se le ha asociado con la lengua hablada por los Zuñi, del Suroeste de E.U. ^{93,138}.

Aunque el AMOVA no fue significativo para las variaciones dadas las afiliaciones lingüísticas, estas nos pueden dar una idea de las relaciones ancestrales entre las poblaciones Amerindias. Un ejemplo son las distancias cortas en nuestras redes conjuntas entre las poblaciones del tronco Uto-Azteca (Tarahumaras, Nahuas, y Jemez), las cuales tienen relación con

poblaciones del Suroeste de E.U. también parte de este tronco ^{135,139,140}. Inclusive, la presencia en estas agrupaciones de los P'urhépechas, que, aunque su lengua es aislada, tiene ligeras semejanzas con las de poblaciones de esta misma zona ^{93,138}.

Otra de las razones subyacentes de la heterogeneidad entre las poblaciones son las distancias geográficas, que están asociadas con el aislamiento como en la sierra Nayarita correspondiente a nuestros valores altos de R_{ST} en los Wurrárikas. Anteriormente, había sido mencionada en estudios como el de Regueiro et al, en el 2013, y Sandoval et al, en el 2012, donde reportan una diferencia entre las poblaciones que habitaban la zona de Mesoamérica y las que no la habitaban, y que es el punto de quiebre entre la diversidad de Norteamérica y Sudamérica ^{78,111}. Dado que la distribución original de los Nayeri y los Wurrárikas era fuera de la región de Mesoamérica, agrupamos a las poblaciones por su localización fuera (Nayeri y Wurrárika) y dentro (P'urhépechas y Mazahuas) de Mesoamérica y realizamos un AMOVA, pero tampoco fue significativo.

Aunque esta separación geográfica no es la razón de la estructura y distancia entre las poblaciones, el aislamiento que esto provoca, puede generar la diversificación de las poblaciones de manera local como lo menciona Battaglia et al, en 2013, para las poblaciones de Sudamérica ⁵⁸. Ellos observan esta diversificación en forma de paragrupos en sus redes, de la misma manera que nosotros los observamos en las nuestras, principalmente en QM3, y sugerido por las grandes distancias en la de QL54. En el 2012, Sandoval et al, observó también la presencia de paragrupos en muestras Nativas Mexicanas, aunque estos podrían estar representando a QL54, dado que en ese momento aún no se identificaba ese marcador ⁷⁸.

La especificación de estos paragrupos suele verse limitada por la falta de recursos o por la falta de marcadores que los definan. Por ejemplo en los

últimos años se han encontrado marcadores tipo SNP que definen sub-HG como SA01, derivado de QM3, encontrado por Jota et al, en 2011; PV2 y M557, también derivados de QM3, y PV3 y PV4 clados hermanos de QM3, descubiertos por Battaglia et al, en 2013; y finalmente, este año Jota et al, encontraron los marcadores SA02-05 y SA29, derivadas de QL54, pronosticando aún la necesidad de buscar más marcadores de este tipo ^{58,78,93,94,141,142}.

Estos datos recabados acerca de la estructura de las poblaciones Nativas de México, son importantes para su uso como modelos en la búsqueda de genes candidatos para enfermedades complejas. Las poblaciones modelo no están exentas de presentar cierta estructura genética, como se ha demostrado en los Chinos Han, sin embargo el trasfondo ancestral y los tiempos de homogeneización, han permitido que esta estructura sea menos compleja que la que ocurre en poblaciones jóvenes con procesos demográficos recientes e intensos ^{67,143}.

Aunque no existen datos epidemiológicos completos de la incidencia de enfermedades complejas en poblaciones Amerindias, ya se han determinado que algunas de estas enfermedades pueden estar asociadas a la etnicidad tal como la diabetes en los Pimas, la masa corporal con un trasfondo Amerindio, e incluso rasgos biomédicos como el Volumen Espiratorio Forzado ¹⁴⁴⁻¹⁴⁶.

No obstante, nuestros datos deben ser tomados con cautela dado el bajo número de muestra. A su vez también es necesario el aumento del número de marcadores que permitan definir los sublinajes asociados a estas expansiones locales, para obtener una mayor precisión de la estructura genética de estas poblaciones.

Conclusión

En este estudio encontramos que, a pesar del proceso de mestizaje ocurrido, estas poblaciones presentan una alta frecuencia de linajes ancestrales paternos Amerindios, aunque cabe mencionar la presencia de sub-linajes que generan una estructura genética dentro de estas poblaciones.

Acerca de la utilidad biomédica, el trasfondo ancestral Amerindio tanto paterno como materno, permite proponerlas como modelos para la búsqueda de biomarcadores de susceptibilidad. Esto tomando en cuenta la estructura genética, y considerando el aumento en el número de muestra. Finalmente, la divergencia entre estas poblaciones puede sugerirnos la existencia de alelos regionales que afecten rasgos biomédicos y que por lo tanto estén asociados a la etnicidad Amerindia.

Perspectivas

- Aumentar el número de muestra y de marcadores específicos, para precisar más la estructura de estas poblaciones.
- Realizar un análisis en marcadores autosómicos en estas poblaciones para comparar los resultados de la estructura genética del cromosoma Y con estos.
- Realizar análisis de este tipo en otras poblaciones indígenas aisladas.
- Realizar un análisis piloto de búsqueda de genes candidatos.

Bibliografía

1. Hughes, A. R., Inouye, B. D., Johnson, M. T. J., Underwood, N. & Vellend, M. Ecological consequences of genetic diversity. *Ecol. Lett.* **11**, 609–623 (2008).
2. Balaresque, P. L., Ballereau, S. J. & Jobling, M. A. Challenges in human genetic diversity: Demographic history and adaptation. *Hum. Mol. Genet.* **16**, 134–139 (2007).
3. Cavalli, L., Menozzi, P. & Piazza, A. *The History and Geography of Human Genes*. (Princeton University Press, 1994).
4. Roewer, L. *et al.* Continent-Wide Decoupling of Y-Chromosomal Genetic Variation from Language and Geography in Native South Americans. *PLoS Genet.* **9**, e1003460 (2013).
5. Frazer, K., Murray, S., Schork, N. & Topol, E. Human genetic variation and its contribution to complex traits. *Nat. Rev. Genet.* **10**, 241–251 (2009).
6. Hau, J. in *Sourcebook of Models for Biomedical Research* (ed. Conn, P. M.) 762 (Humana Press, 2008). doi:10.1007/978-1-59745-285-4
7. Ioannidis, J. P. A. Non-replication and inconsistency in the genome-wide association setting. *Hum. Hered.* **64**, 203–213 (2007).
8. Collins, F., Green, E., Guttmacher, A. & Guyer, M. A vision for the future of genomics research. *Nature* **422**, 835–847 (2003).
9. Tishkoff, S. a & Kidd, K. K. Implications of biogeography of human populations for ‘race’ and medicine. *Nat. Genet.* **36**, S21–S27 (2004).
10. Liu, R., Wang, X., Aihara, K. & Chen, L. Early Diagnosis of Complex Diseases by Molecular Biomarkers, Network Biomarkers, and Dynamical Network Biomarkers. *Med. Res. Rev.* **34**, 455–478 (2014).
11. Wacholder, S., Rothman, N. & Caporaso, N. Population stratification in epidemiologic studies of common genetic variants and cancer: quantification of bias. *J. Natl. Cancer Inst.* **92**, 1151–1158 (2000).
12. Cavalli-Sforza, L. L. & Bodmer, W. *The Genetics of Human Populations*. (W. H. Freeman, 1971).
13. Kiaris, H. *Genes, Polymorphisms, and the Making of Societies: How Genetic Behavioral Traits Influence Human Cultures*. (Universal Publishers, 2012).

14. Goodwin, W., Linacre, A. & Hadi, S. *An Introduction to Forensic Genetics*. (John Wiley & Sons, 2007).
15. Vignal, A., Milan, D., San Cristobal, M. & Eggen, A. A review on SNP and other types of molecular markers and their use in animal genetics. *Genet. Sel. Evol.* **40**, 241–264 (2002).
16. Jehan, T. & Lakhanpaul, S. Single nucleotide polymorphism (SNP)-methods and applications in plant genetics: a review. *Indian J. Biotechnol.* **5**, 435 (2006).
17. Sobrino, B., Brión, M. & Carracedo, A. SNPs in forensic genetics: A review on SNP typing methodologies. *Forensic Sci. Int.* **154**, 181–194 (2005).
18. Nachman, M. & Crowell, S. Estimate of the Mutation Rate per Nucleotide in Humans. *Genetics* **156**, 297–304 (2000).
19. Jeffreys, A. J., Wilson, V. & Thein, S. L. Hypervariable minisatellite regions in human DNA. *Nature* **314**, 67–73 (1985).
20. Zane, L., Bargelloni, L. & Patarnello, T. Strategies for microsatellite isolation: A review. *Mol. Ecol.* **11**, 1–16 (2002).
21. Tóth, G., Gáspári, Z. & Jurka, J. Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Res.* **10**, 967–981 (2000).
22. Tishkoff, S. a & Verrelli, B. C. Patterns of human genetic diversity: implications for human evolutionary history and disease. *Annu. Rev. Genomics Hum. Genet.* **4**, 293–340 (2003).
23. Balloux, F. & Lugon-Moulin, N. The estimation of population differentiation with microsatellite markers. *Mol. Ecol.* **11**, 155–165 (2002).
24. Mueller, L. in *Encyclopedia of Genetics* 1519–1523 (Academic Press, 2001).
25. Hu, D. & Ziv, E. in *Pharmacogenomics in Drug Discovery and Development* (ed. Yan, Q.) 487 (Humana Press, 2008).
26. Little, J. *et al.* STrengthening the REporting of genetic association studies (STREGA)- An extension of the STROBE statement. *Genet. Epidemiol.* **33**, 581–598 (2009).
27. Jobling, M., Hollox, E., Hurles, M., Kivisild, T. & Tyler-Smith, C. *Human evolutionary genetics*. (Garland Science, 2014).

28. Ziv, E. & González, E. Human population structure and genetic association studies. *Pharmacogenomics* **4**, 431–441 (2003).
29. Salzano, B. F. M. & Bortolini, M. C. The Evolution and Genetics of Latin American Populations. *Blood* **37**, 151–170 (2002).
30. Tian, C., Gregersen, P. & Seldin, M. Accounting for ancestry: Population substructure and genome-wide association studies. *Hum. Mol. Genet.* **17**, 143–150 (2008).
31. Rellethford, J. H., Colledge, Y. & York, N. in *Encyclopedia of Life Sciences* 1–6 (John Wiley & Sons, 2001). doi:10.1038/npg.els.0001470
32. Long, J. The genetic structure of admixed populations. *Genetics* **127**, 417–428 (1991).
33. Klyosov, A. DNA Genealogy, Mutation Rates, and Some Historical Evidence Written in the Y-Chromosome: I. Basic Principles and the Method. *J. Genet. Genealogy* **5**, 1–113 (2009).
34. Rebbeck, T. & Sankar, P. Ethnicity, ancestry, and race in molecular epidemiologic research. *Cancer Epidemiol. Biomarkers Prev.* **14**, 2467–2471 (2005).
35. Ali-Khan, S. E., Krakowski, T., Tahir, R. & Daar, A. S. The use of race, ethnicity and ancestry in human genetic research. *Hugo J.* **5**, 47–63 (2011).
36. Royal, C. D. *et al.* Inferring Genetic Ancestry: Opportunities, Challenges, and Implications. *Am. J. Hum. Genet.* **86**, 661–673 (2010).
37. Tian, C. *et al.* A genomewide single-nucleotide-polymorphism panel with high ancestry information for African American admixture mapping. *Am. J. Hum. Genet.* **79**, 640–649 (2006).
38. Underhill, P. *et al.* Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. *Genome Res.* **7**, 996–1005 (1997).
39. Kosoy, R. *et al.* Ancestry Informative Marker Sets for Determining Continental Origin and Admixture Proportions in Common Populations in America. *Hum. Mutat.* **30**, 69–78 (2009).
40. Underhill, P. & Kivisild, T. Use of Y chromosome and Mitochondrial DNA. Population Structure in Tracing Human Migrations. *Annu. Rev. Genet.* **41**,

539–564 (2007).

41. Pakendorf, B. & Stoneking, M. Mitochondrial Dna and Human Evolution. *Annu. Rev. Genomics Hum. Genet.* **6**, 165–183 (2005).
42. Sigurðardóttir, S., Helgason, A., Gulcher, J. R., Stefansson, K. & Donnelly, P. The Mutation Rate in the Human mtDNA Control Region. *Am. J. Hum. Genet.* **66**, 1599–1609 (2000).
43. Sandoval, K. Ethnicity, Linguistics, and Genetic Diversity in Native Mexicans: Reconstructing the Population History of Mesoamerica. (Universidad Pompeu Fabra, 2010).
44. Jobling, M. & Tyler-Smith, C. Fathers and sons: the Y chromosome and human evolution. *Trends Genet.* **11**, 449–456 (1995).
45. Scozzari, R. *et al.* An unbiased resource of novel SNP markers provides a new chronology for the human Y chromosome and reveals a deep phylogenetic structure in Africa. *Genome Res.* **24**, 535–544 (2014).
46. Karmin, M. *et al.* A recent bottleneck of Y chromosome diversity coincides with a global change in culture. *Genome Res.* **25**, 459–466 (2015).
47. Jobling, M. & Tyler, C. The human Y chromosome: an evolutionary marker comes of age. *Nat. Rev. Genet.* **4**, 598–612 (2003).
48. Blanco, A. Linajes del Cromosoma Y humano: Aplicaciones genético-poblacionales y forenses. (Universidad de Santiago Compostela, 2008).
49. Santana, C. *et al.* Genetic analysis of 17 Y-STRs in a Mestizo population from the Central Valley of Mexico. *Hum. Biol. Open Access Pre-Prints* **63**, (2014).
50. Koopman, P. Sry and Sox9: Mammalian testis-determining genes. *Cell. Mol. Life Sci.* **55**, 839–856 (1999).
51. Skaletsky, H. *et al.* The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* **423**, 825–837 (2003).
52. Vilar, M. G. *et al.* Genetic diversity in Puerto Rico and its implications for the peopling of the Island and the West Indies. *Am. J. Phys. Anthropol.* **155**, 352–368 (2014).
53. Bosch, E. *et al.* Variation in short tandem repeats is deeply structured by

- genetic background on the human Y chromosome. *Am. J. Hum. Genet.* **65**, 1623–1638 (1999).
54. Krausz, C., Quintana-Murci, L. & Forti, G. Y chromosome polymorphisms in medicine. *Ann. Med.* **36**, 573–583 (2004).
 55. de Knijff, P. Messages through bottlenecks: on the combined use of slow and fast evolving polymorphic markers on the human Y chromosome. *Am. J. Hum. Genet.* **67**, 1055–1061 (2000).
 56. Consortium, T. Y. C. A Nomenclature System for the Tree of Human Y-Chromosomal Binary Haplogroups. *Genome Res.* **12**, 339–348 (2002).
 57. The Chimpanzee Sequencing and Analysis Consortium. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437**, 69–87 (2005).
 58. Battaglia, V. *et al.* The first peopling of South America: new evidence from Y-chromosome haplogroup Q. *PLoS One* **8**, e71390 (2013).
 59. Achilli, A. *et al.* Reconciling migration models to the Americas with the variation of North American native mitogenomes. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 14308–13 (2013).
 60. Vieira, a R., Karras, J. C., Orioli, I. M., Castilla, E. E. & Murray, J. C. Genetic origins in a South American clefting population. *Clin. Genet.* **62**, 458–463 (2002).
 61. Lewis, C. M. & Knight, J. Introduction to genetic association studies. *Cold Spring Harb. Protoc.* **7**, 297–306 (2012).
 62. Hirschhorn, J. N., Lohmueller, K., Byrne, E. & Hirschhorn, K. A comprehensive review of genetic association studies. *Genet. Med.* **4**, 45–61 (2002).
 63. Primrose, S. & Twyman, R. *Principles of Genome Analysis and Genomics.* (Blackwell Publishing, 2003).
 64. Dekker, M., Nelson, L. & Van Duijn, C. in *Neuroepidemiology : From Principles to Practice* (eds. Nelson, L., Tanner, C. & Van Den, S.) 480 (Oxford University Press, 2003).
 65. Lao, O. *et al.* Correlation between Genetic and Geographic Structure in Europe. *Curr. Biol.* **18**, 1241–1248 (2008).

66. Losos, J. *et al.* *The Princeton Guide to Evolution*. (Princeton University Press, 2014).
67. Taniguchi, N. K., Taulii, M. & Maddock, J. A Comparative Analysis of Indigenous Research Guidelines to Inform Genomic Research in Indigenous Communities. *Int. Indig. Policy J.* **3**, 1–20 (2012).
68. Universidad Hebrea de Jerusalén. The Value of Homogeneity for Association Studies. *The Hebrew University Genetic Resource* (2014). Available at: <http://hugr.huji.ac.il/homogeneity.aspx>. (Accessed: 26th May 2015)
69. López, C. & García, V. Aproximaciones científicas al mestizo mexicano. *História, Ciências, Saúde* **20**, 391–410 (2013).
70. Oropeza, D. La esclavitud asiática en el virreinato de la Nueva España. *Hist. Mex.* **61**, 5–57 (2011).
71. Guardado, M. *et al.* A great diversity of Amerindian mitochondrial DNA ancestry is present in the Mexican mestizo population. *J. Hum. Genet.* **54**, 695–705 (2009).
72. López, C. Sangre y temperamento pureza y mestizajes en las sociedades de castas americanas. *Saberes locales ensayos sobre Hist. la Cienc. en América Lat.* 289–342 (2008).
73. CDI. *Indicadores Sociodemográficos de la Población Indígena 2000-2005*. (2006).
74. Olofsson, J. K., Pereira, V., Børsting, C. & Morling, N. Peopling of the North Circumpolar Region – Insights from Y Chromosome STR and SNP Typing of Greenlanders. *PLoS One* **10**, e0116573 (2015).
75. Dulik, M. C. *et al.* Mitochondrial DNA and Y chromosome variation provides evidence for a recent common ancestry between Native Americans and Indigenous Altaians. *Am. J. Hum. Genet.* **90**, 229–246 (2012).
76. Dulik, M. C., Osipova, L. P. & Schurr, T. G. Y-chromosome variation in Altaian Kazakhs reveals a common paternal gene pool for Kazakhs and the influence of Mongolian expansions. *PLoS One* **6**, e17548 (2011).
77. Reich, D. *et al.* Reconstructing native American population history. *Nature* **488**, 370–374 (2012).
78. Sandoval, K. *et al.* Y-chromosome diversity in Native Mexicans reveals

- continental transition of genetic structure in the Americas. *Am. J. Phys. Anthropol.* **148**, 395–405 (2012).
79. Pennsylvania Archaeology. Settling of the New World. *Paleoindian Period* (2015). Available at: <http://www.phmc.state.pa.us/portal/communities/archaeology/native-american/paleoindian-period.html>. (Accessed: 14th July 2016)
 80. Salud, S. de. Salud y nutrición para población marginada, rural e indígena. (2001). Available at: <http://www.salud.gob.mx/unidades/dgpfs/pueblosindigenas/PueblosIndigenas.pdf>. (Accessed: 23rd July 2015)
 81. Athey, W. Haplogroup prediction from Y-STR values using a Bayesian-allele frequency approach. *J. Genet. Genalogy* **2**, 34–39 (2006).
 82. Excoffier, L. & Lischer, L. Arlequin suite ver 3.5: A newseries of programs to perform population genetic analyses under Linux and Window. *Mol. Ecol. Resour.* **10**, 564–567 (2010).
 83. Bandelt, H.-J., Forster, P. & Röhl, A. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* **16**, 37–48 (1999).
 84. Fluxus Technology. *Network 4.6.1.1. User Guide*. (2012).
 85. Slatkin, M. Inbreeding coefficients and coalescence times. *Genet. Res.* **58**, 167–75 (1991).
 86. Gabaldón, T. Filogenómica y evolución del genoma. *Nuevas vías* (2011). Available at: http://instituto-roche.es/Biotecnologia/NuevasVias/62/Filogenomica_y_evolucion_del_genoma. (Accessed: 23rd May 2016)
 87. Posada, D. & Crandall, K. A. Intraspecific gene genealogies: trees grafting into networks. *Trends Ecol. Evol.* **16**, 37–45 (2001).
 88. Nakhleh, L. in *Problem Solving Handbook in Computational Biology and Bioinformatics* (eds. Heath, L. & Ramakrishnan, N.) 125–158 (Springer Science, 2011). doi:10.1007/978-0-387-09760-2
 89. Beezley, W. *Mexico in World History*. (Oxford University Press, 2011).
 90. Guardado-Estrada, M. *et al.* A great diversity of Amerindian mitochondrial DNA ancestry is present in the Mexican mestizo population. *J. Hum. Genet.*

- 54, 695–705 (2009).
91. Martínez-Cortés, G. *et al.* Admixture and population structure in Mexican-Mestizos based on paternal lineages. *J. Hum. Genet.* **57**, 568–574 (2012).
 92. Santana, C. *et al.* Genetic analysis of 17 y-STRs in a mestizo population from the Central Valley of Mexico. *Hum. Biol.* **86**, 289–312 (2015).
 93. Páez-Riberos, L. a. *et al.* Y-linked haplotypes in Amerindian chromosomes from Mexican populations: Genetic evidence to the dual origin of the Huichol tribe. *Leg. Med.* **8**, 220–225 (2006).
 94. Rangel, H. *et al.* Genetic admixture, relatedness, and structure patterns among Mexican populations revealed by the Y-chromosome. *Am. J. Phys. Anthropol.* **135**, 448–461 (2008).
 95. Ambrosio, B. *et al.* Y-STR genetic diversity in autochthonous Andalusians from Huelva and Granada provinces (Spain). *Forensic Sci. Int. Genet.* **6**, (2012).
 96. Bailliet, G. *et al.* Antecedentes y nuevos aportes en el estudio del Cromosoma Y en poblaciones humanas sudamericanas. *J. Basic Appl. Genet.* **22**, 1–9 (2011).
 97. Mendizabal, I. *et al.* Genetic origin, admixture, and asymmetry in maternal and paternal human lineages in Cuba. *BMC Evol. Biol.* **8**, 213 (2008).
 98. Klyosov, A. *DNA Genealogy, Mutation Rates, and Some Historical Evidence Written in the Y-Chromosome: I. Basic Principles and the Method.* *Journal of Genetic Genealogy* **5**, (2009).
 99. Young, K. L. *The Basques in the Genetic Landscape of Europe.* (University of Kansas, 2009).
 100. Caballos, E. Terror, violación y pederastia en la Conquista de América : el caso de Lázaro Fonte. *Anu. Hist. Am. Lat.* **44**, 37–66 (2007).
 101. Lugo, C. Machismo y violencia. *Nueva Soc.* 40–47 (1985).
 102. Wettlaufer, J. The jus primae noctis as a male power display: A review of historic sources with evolutionary interpretation. *Evol. Hum. Behav.* **21**, 111–123 (2000).
 103. Villanueva, É. El fin del oro verde. Confl icto social y movimiento

- campesino 1960-2008. *Rev. Latinoam. Derecho Soc.* **17**, 333–338 (2013).
104. Adams, S. M. *et al.* The Genetic Legacy of Religious Diversity and Intolerance: Paternal Lineages of Christians, Jews, and Muslims in the Iberian Peninsula. *Am. J. Hum. Genet.* **83**, 725–736 (2008).
 105. Kaminer, J. Los judíos y su presencia en México desde el siglo XVI. *Diario de la vida judía en México y en el mundo* (2010). Available at: <http://diariojudio.com/opinion/los-judios-y-su-presencia-en-mexico-desde-el-siglo-xvi/5594/>. (Accessed: 1st June 2016)
 106. Embajada de Líbano en México. Emigración Libanesa a México. Available at: http://www.embajadadelibano.org.mx/emigracion_libanesa.php. (Accessed: 1st June 2016)
 107. Elhaik, E. *et al.* Geographic population structure analysis of worldwide human populations infers their biogeographical origins. *Nat. Commun.* **5**, 3513 (2014).
 108. Gómez-Camargo, D. E. *et al.* Genetic Structure of Cartagena de Indias Population Using Hypervariable Markers of Y Chromosome. *Open J. Genet.* **5**, 27–41 (2015).
 109. Wheat, D. The first great waves: African provenance zones for the transatlantic slave trade to Carctagena de Indias, 1570–1640. *J. Afr. Hist.* **52**, 1–22 (2011).
 110. Janssen, E., Casas, R. M. & Shaw, G. B. Una propuesta para estimar la población indígena en México a partir de los datos censales. *Estud. Demogr. Urbanos Col. Mex.* **21**, 457–471 (2006).
 111. Regueiro, M., Alvarez, J., Rowold, D. & Herrera, R. J. On the origins, rapid expansion and genetic diversity of native Americans from hunting-gatherers to agriculturalists. *Am. J. Phys. Anthropol.* **150**, 333–348 (2013).
 112. Dillehay, T. D. Probing deeper into first American studies. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 971–978 (2009).
 113. Schurr, T. G. & Sherry, S. T. Mitochondrial DNA and Y chromosome diversity and the peopling of the Americas: Evolutionary and demographic evidence. *Am. J. Hum. Biol.* **16**, 420–439 (2004).
 114. Verdu, P. *et al.* Patterns of Admixture and Population Structure in Native Populations of Northwest North America. *PLoS Genet.* **10**, (2014).

115. Mizuno, F. *et al.* Complete mitogenome analysis of indigenous populations in Mexico: its relevance for the origin of Mesoamericans. *J. Hum. Genet.* 1–9 (2014). doi:10.1038/jhg.2014.35
116. Karafet, T. & *et al.* Y chromosome markers and trans-Bering Strait dispersals. *Am. J. Phys. Anthropol.* **102**, 301–314 (1997).
117. Schurr, T. G. The Peopling of the New World: Perspectives from Molecular Anthropology. *Annu. Rev. Anthropol.* **33**, 551–583 (2004).
118. Dillehay, T. D. *et al.* New Archaeological Evidence for an Early Human Presence at Monte Verde, Chile. *PLoS One* **10**, e0141923 (2015).
119. McAvoy, J. & McAvoy, L. *Archaeological Investigations of Site 44SX202, Cactus Hill, Sussex County, Virginia.* (Department of Historic Resources, 1997).
120. Boëda, E. *et al.* A new late Pleistocene archaeological sequence in South America: the Vale da Pedra Furada (Piauí, Brazil). *Antiquity* **88**, 927–941 (2014).
121. Dixon, E. J. in *Mobility and Ancient Society in Asia and the Americas* (eds. Frachetti, D. & Spengler, R. N.) 169–184 (Springer, 2015).
122. Erlandson, J. & Braje, T. in *Mobility and Ancient Society in Asia and the Americas2* (eds. Frachetti, M. D. & Spengler, R. N.) 202 (Springer, 2015).
123. Erlandson, J. *et al.* The Kelp Highway Hypothesis: Marine Ecology, the Coastal Migration Theory, and the Peopling of the Americas. *J. Isl. Coast. Archaeol.* **2**, 161–174 (2007).
124. Renato Kipnis. Early hunter-gatherers in the Americas: perspectives from central Brazil. *Antiquity* **72**, 581–592 (1998).
125. Fix1, A. Colonization Models and Initial Genetic Diversity in the Americas. *Hum. Biol.* **74**, 1–10 (2002).
126. Ellegren, H. & Ellegren, N. Determinants of genetic diversity. *Nat Rev Genet* **advance on**, 422–433 (2016).
127. Tishkoff, S. A. & Verrelli, B. C. Patterns of Human Genetic Diversity : Implications for Human Evolutionary History and Disease. *Annu. Rev. Genomics Hum. Genet.* **4**, 293–340 (2003).

128. Wall, J. D. *et al.* Genetic variation in native Americans, inferred from latino SNP and resequencing data. *Mol. Biol. Evol.* **28**, 2231–2237 (2011).
129. Gorostiza, A. *et al.* Reconstructing the History of Mesoamerican Populations through the Study of the Mitochondrial DNA Control Region. *PLoS One* **7**, (2012).
130. Wang, S. *et al.* Genetic variation and population structure in Native Americans. *PLoS Genet.* **3**, 2049–2067 (2007).
131. Portilla, M. L., Robles, N., Rojas, J. & CONACULTA. *Pasajes de la Historia.* (México Desconocido, 2003).
132. Uribe, E. El Medio Aborigen Nayarita. *Rev. Mex. Sociol.* **12**, 211–223 (1950).
133. Moreno-Estrada, A. *et al.* The genetics of Mexico recapitulates Native American substructure and affects biomedical traits. *Science* (80-.). **344**, 1280–5 (2014).
134. Promotora Española de Lingüística. *Lenguas Amerindias. Lenguas del Mundo* (2013). Available at: <http://www.proel.org/index.php?pagina=mundo/amerindia>. (Accessed: 2nd August 2016)
135. Promotora Española de Lingüística. *Tronco Uto-Azteca. Lenguas Amerindias* (2013). Available at: <http://www.proel.org/index.php?pagina=mundo/amerindia/uto>. (Accessed: 2nd August 2016)
136. Promotora Española de Lingüística. *Grupo Cora-Chol. Lenguas Amerindias* (2013). Available at: <http://www.proel.org/index.php?pagina=mundo/amerindia/uto/corac hol>. (Accessed: 2nd August 2016)
137. Promotora Española de Lingüística. *Tronco Oto-Mangue. Lenguas Amerindias* (2013). Available at: <http://www.proel.org/index.php?pagina=mundo/amerindia/otomangu e>. (Accessed: 2nd August 2016)
138. Promotora Española de Lingüística. *Lengua Purépecha. Lenguas Aisladas* (2013). Available at: <http://www.proel.org/index.php?pagina=mundo/aisladas/tarasca>. (Accessed: 2nd August 2016)

139. Wheat, J. Mogollon Culture Prior to A.D. 1000. *Mem. Soc. Am. Archaeol.* 1-242 (1955).
140. Rinaldo, J. Notes on the Origins of Historic Zuni Culture. *Kiva* **29**, 86-98 (1964).
141. Jota, M. S. *et al.* A new subhaplogroup of native American Y-Chromosomes from the Andes. *Am. J. Phys. Anthropol.* **146**, 553-559 (2011).
142. Jota, M. S. *et al.* New native South American Y chromosome lineages. *J. Hum. Genet.* 1-11 (2016). doi:10.1038/jhg.2016.26
143. Xu, S. *et al.* Genomic Dissection of Population Substructure of Han Chinese and Its Implication in Association Studies. *Am. J. Hum. Genet.* **85**, 762-774 (2009).
144. Schulz, L. & Chaudari, L. High-Risk Populations: The Pimas of Arizona and Mexico. *Curr. Obes. Rep.* **4**, 92-98 (2015).
145. Moreno, A. *et al.* The genetics of Mexico recapitulates Native American substructure and affects biomedical traits. *Science (80-.)*. **344**, 1280-1285 (2014).
146. Klimentidis, Y. C., Miller, G. F. & Shriver, M. D. The relationship between European genetic admixture and body composition among Hispanics and Native Americans. *Am. J. Hum. Biol.* **21**, 377-382 (2009).

Texto suplementario

Características demográficas de las poblaciones muestreadas

- **P'urhépechas (Familia lingüística: Tarasco):** Se encuentran en la zona lacustre y montañosa del centro de Michoacán ¹⁴⁷. El actual grupo P'urhépecha deriva de una mezcla de grupos Chichimecas, Nahuas y pre-Tarascos¹⁴⁷. Tras la caída de Tenochtitlan, aun cuando existieron conflictos con los emisarios, y el conquistador Nuño Guzmán; los españoles entraron al territorio Tarasco sin resistencia, lo cual evitó la disminución de la población P'urhépecha. Posterior a la muerte del último gobernante Tarasco, se interrumpió la interacción entre las tierras altas y bajas donde habitaban, por lo que los individuos de las tierras altas se mantuvieron alejados del mestizaje con los españoles y africanos ¹³¹. En la actualidad existe un proceso de auto-reafirmación étnica que mantiene muchas comunidades cerradas al mestizaje con poblaciones mestizas actuales ¹⁴⁷.
- **Mazahuas (Familia lingüística: Oto-Mangue):** La región Mazahua está situada en la parte noroeste del Estado de México y en una pequeña área del oriente del estado de Michoacán¹⁴⁸. Esta población nunca ha estado aislada por completo, se sabe que provienen de poblaciones Chichimecas y Tultecas; habitaban una zona de rutas comerciales; a su vez que formaban parte del señorío de Atzacotalco, el cual era pluricultural y por lo cual se relacionaban con individuos Nahuas, Otomíes y P'urhépechas, estos últimos eran los que dominaban la zona habitada por los Mazahuas ^{149,150}. Posterior al dominio español, han mantenido sus costumbres y lengua debido a que se les reconoce una gran capacidad de adaptación, se cree que

probablemente estén altamente mestizados, aunque los individuos mestizados suelen salir de los asentamientos Mazahuas ¹⁴⁸.

- **Nayeri o Coras (Familia lingüística: Uto-Azteca):** Los Coras ocupan una porción noreste del estado de Nayarit; que por el norte limita con el estado de Durango. Antes de la conquista mantenían relaciones con los O'dam (antes llamados Tepehuanos) y los WIRRÁRIKAS (antes llamados Huicholes)¹⁵¹. La entrada y conquista de la sierra del Nayar por parte de los españoles fue difícil, tanto que incluso en el Siglo XVIII aún no era sometido por completo e incluso la primera carretera para llegar a estos territorios fue en los años 80¹⁵¹. En la actualidad presentan un aislamiento cultural de las poblaciones Mestizas ¹⁵¹.
- **WIRRÁRIKAS (Familia lingüística: Uto-Azteca):** Antes conocidos como Huicholes, habitan principalmente en municipios del estado de Jalisco, y de Nayarit, pero existen grupos minoritarios en los estados de Zacatecas y Durango ¹⁵². Se cree que este grupo deriva de Chichimecas, Nahuas y pre-Tarascos; a pesar de su amplia distribución nunca fueron sometidos por los Mexicas y poseían una alta reafirmación étnica por lo que no tenían relaciones con otras poblaciones ¹⁵². Durante la Conquista, los sobrevivientes huyeron a la sierra del Nayar la cual, por su difícil acceso, no fue conquistada. Los WIRRÁRIKAS han logrado impedir el asentamiento de Mestizos al interior de sus comunidades, además de que los matrimonios con individuos Mestizos están mal vistos ¹⁵².

Bibliografía del texto suplementario

147. CDI. Purépechas - P'urhépecha. *Comisión Nacional para el desarrollo de los Pueblos Indígenas* (2009). Available at: http://www.cdi.gob.mx/index.php?option=com_content&view=article&id=604:purepechas-purhepecha&catid=54:monografias-de-los-pueblos-indigenas&Itemid=62.

148. CDI. Mazahuas. *Comisión Nacional para el desarrollo de los Pueblos Indígenas* (2009). Available at: http://www.cdi.gob.mx/index.php?option=com_content&view=article&id=616:mazahuas&catid=54:monografias-de-los-pueblos-indigenas&Itemid=62.
149. Oehmichen, C. *Identidad, género y relaciones interétnicas. Mazahuas en la ciudad de México. Journal of Chemical Information and Modeling* **53**, (Instituto de Investigaciones Antropológicas, 2015).
150. Guzmán, M. Otomíes y mazahuas de Michoacán, siglos XV-XVII. Trazos de una historia. *Rev. Estud. Históricas* 11-74 (2012).
151. CDI. Coras - Nayeri. *Comisión Nacional para el desarrollo de los Pueblos Indígenas* (2009). Available at: http://www.cdi.gob.mx/index.php?option=com_content&view=article&id=591:coras-nayeri&catid=54:monografias-de-los-pueblos-indigenas&Itemid=62.
152. CDI. Huicholes - Wirraritari o Wirrarika. *Comisión Nacional para el desarrollo de los Pueblos Indígenas* (2009). Available at: http://www.cdi.gob.mx/index.php?option=com_content&view=article&id=596:huicholes-wirraritari-o-wirrarika-&catid=54:monografias-de-los-pueblos-indigenas&Itemid=62.

Imágenes Suplementarias

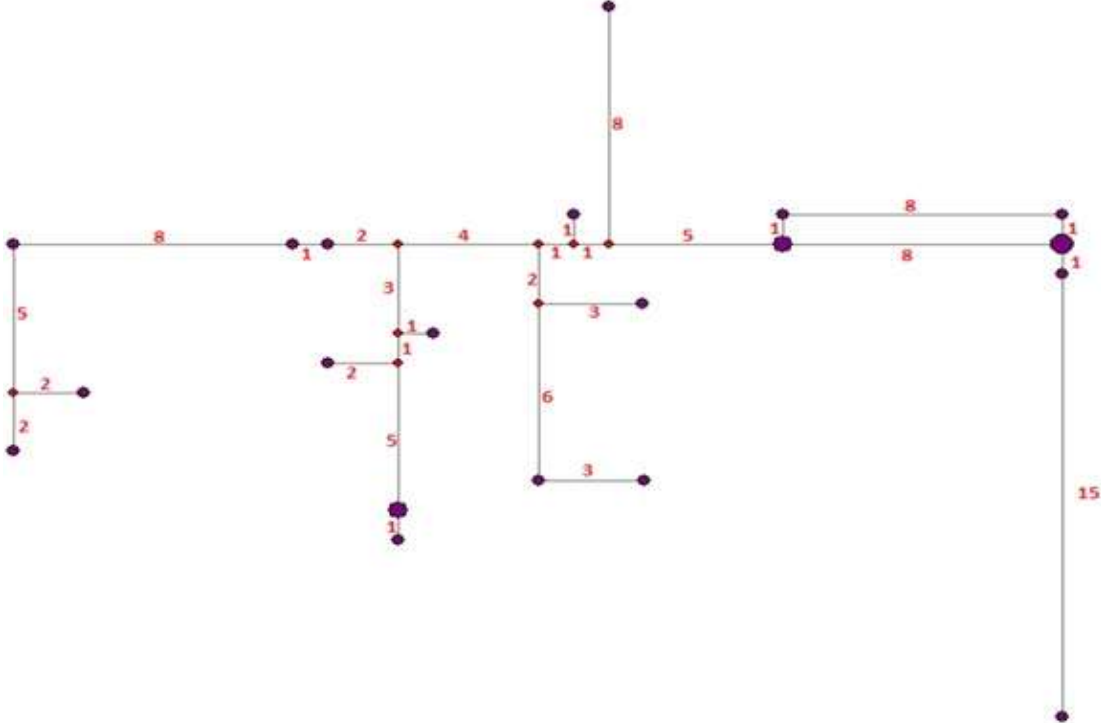


Figura S1. Red filogenética de la población P'urhépecha perteneciente a Q-M3.

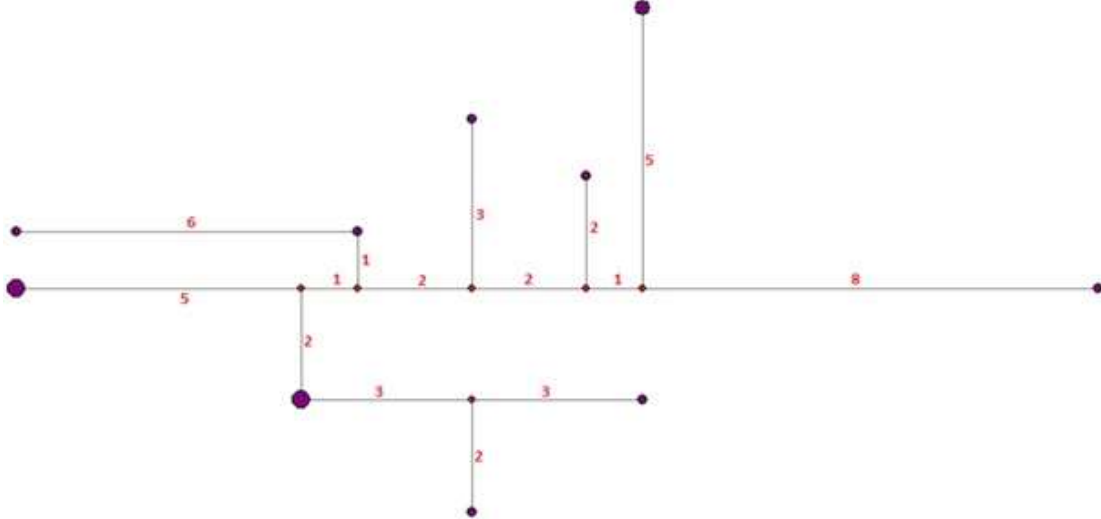


Figura S2. Red filogenética de la población Mazahua perteneciente a Q-M3.

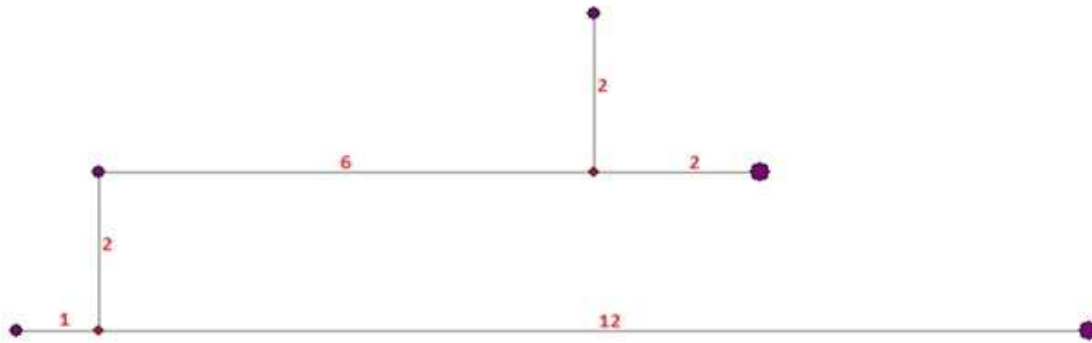


Figura S3. Red filogenética de la población Nayeri perteneciente a Q-M3.

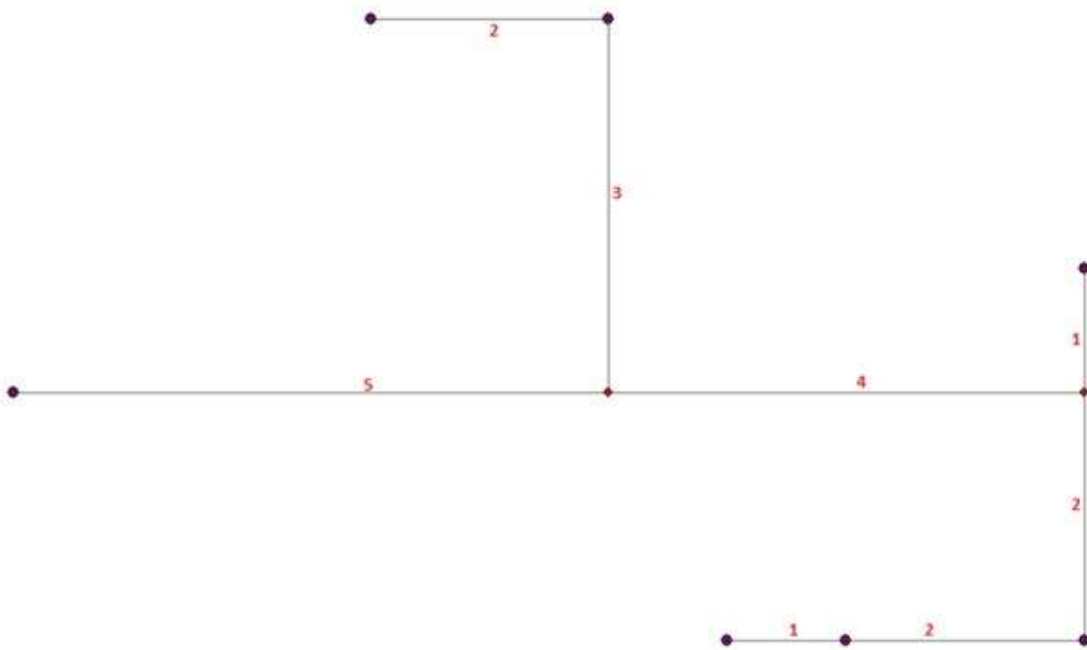


Figura S4. Red filogenética de la población Wirríríka perteneciente a Q-M3.

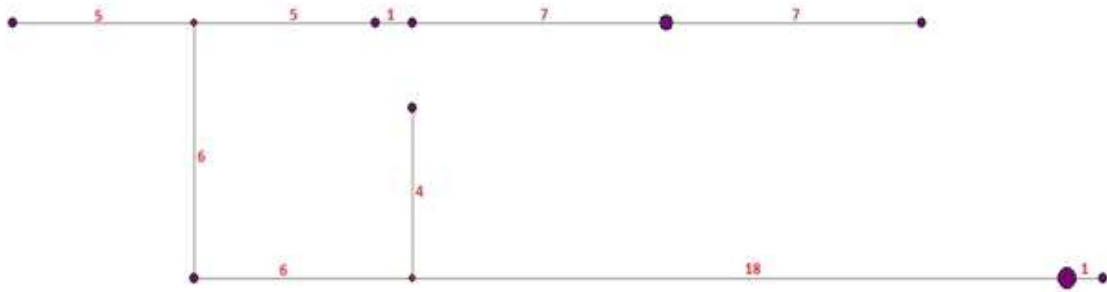


Figura S5. Red filogenética de la población P'urhépecha perteneciente a Q-L54.

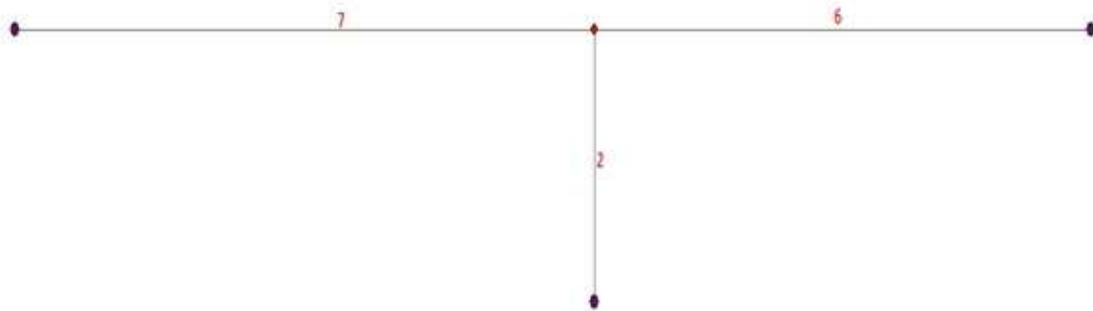


Figura S6. Red filogenética de la población Mazahua perteneciente a Q-L54.

Tablas Suplementarias

Tabla S1. Haplotipos y linajes encontrados en la población P'urhépecha. En colores se encuentran los individuos relacionados. HG: Haplogrupo. FS: Fitness Score.

	19	385a	385b	389I	389II	390	391	392	393	437	438	439	448	456	458	635	Y GATA H4	HG	FS	Prob.	Parentesco	Frecuencia relativa	
Pu68	16	15	18	14	31	21	10	11	14	14	11	12	12	15	18	22	11	E1b1a	32	1		1	
Pu77	13	13	19	13	30	25	10	11	13	14	10	12	12	16	16	22	12	E1b1b	31	1		1	
Pu85	14	11	16	13	29	23	10	13	13	15	12	13	19	14	19	21	12	L	36	0.66		1	
Pu180	14	11	14	13	30	24	11	12	14	15	12	12	12	15	17	23	12	R1b	24	1		1	
Pu148	14	14	14	13	29	24	11	13	13	15	12	12	19	15	18	23	13	R1b	52	1		1	
Pu100	17	14	16	12	26	24	10	14	13	14	11	11	19	16	15	22	12	T	27	0.93		1	
Pu154	13	13	18	13	30	24	10	14	13	14	11	12	19	14	16	23	11	Q	66	1		1	
Pu151	13	13	18	13	30	25	11	15	13	14	11	12	20	17	17	22	13	Q	46	0.98		1	
Pu169	13	13	18	13	30	26	11	15	13	14	11	12	12	17	17	22	13	Q	18	0.99		1	
Pu123	13	13	18	13	30	26	11	15	13	14	11	12	20	17	17	22	13	Q	37	0.97		1	
Pu167	13	13	18	14	31	24	10	14	13	14	12	12	12	15	14	22	12	Q	30	1		1	
Pu157	13	13	18	14	31	25	10	14	13	14	12	12	12	15	14	22	12	Q	28	1			
Pu161	13	13	18	14	31	25	10	14	13	14	12	12	12	15	14	22	12	Q	28	1	Relacionados		
Pu164	13	13	18	14	31	25	10	14	13	14	12	12	12	15	14	22	12	Q	28	1			3
Pu170	13	13	18	14	31	25	10	14	13	14	12	12	12	15	14	22	12	Q	28	1	Relacionados		
Pu173	13	13	18	14	31	25	10	14	13	14	12	12	12	15	14	22	12	Q	28	1			
Pu177	13	13	18	14	31	25	10	14	13	14	12	12	12	15	14	22	12	Q	28	1			
Pu140	13	13	18	14	31	25	10	14	13	14	12	12	20	15	14	22	12	Q	55	1		2	
Pu104	13	13	18	14	31	25	10	14	13	14	12	12	20	15	14	22	12	Q	55	1			
Pu158	13	13	18	14	31	25	10	14	14	14	12	12	12	15	14	22	12	Q	25	1		1	
Pu156	13	13	18	14	31	25	10	14	14	14	12	12	20	15	14	22	12	Q	49	1		1	
Pu79	13	14	16	13	31	24	11	17	13	14	11	12	12	15	17	22	13	Q	20	1		1	
Pu99	13	14	17	12	28	25	9	13	13	14	11	12	20	15	16	23	12	Q	46	0.99		1	
Pu88	13	14	17	13	29	23	11	14	13	14	12	14	19	16	16	22	11	Q	65	1		1	
Pu71	13	14	17	13	30	24	11	16	13	14	11	12	12	15	18	22	12	Q	29	1		1	
Pu126	13	14	17	13	30	25	10	14	13	14	11	12	21	16	15	22	11	Q	74	1		1	
Pu133	13	14	18	12	28	25	9	13	14	14	11	12	19	15	16	22	12	Q	49	1		1	
Pu96	13	14	19	13	30	24	10	16	13	14	11	13	21	17	17	22	12	Q	54	1	Relacionados		
Pu86	13	14	19	13	30	24	10	16	13	14	11	13	21	17	17	22	12	Q	54	1			1
Pu181	13	15	16	13	29	23	10	14	13	14	12	12	12	16	16	22	11	Q	38	1		1	
Pu160	13	15	16	13	29	23	10	14	13	14	12	12	12	17	16	22	11	Q	38	1		1	
Pu138	13	15	16	13	29	23	10	14	13	14	12	12	19	17	16	22	11	Q	83	1	Relacionados		
Pu141	13	15	16	13	29	23	10	14	13	14	12	12	19	17	16	22	11	Q	83	1			2

Pu146	13	15	16	13	29	23	10	14	13	14	12	12	19	17	16	22	11	Q	83	1				
Pu149	13	15	16	13	29	23	10	14	13	14	12	12	19	17	16	22	11	Q	83	1	Relacionados			
Pu152	13	15	16	13	29	23	10	14	13	14	12	12	19	17	16	22	11	Q	83	1				
Pu139	13	15	17	13	30	24	10	17	13	14	11	11	20	15	18	22	12	Q	52	1			2	
Pu147	13	15	17	13	30	24	10	17	13	14	11	11	20	15	18	22	12	Q	52	1				
Pu119	13	15	17	13	30	24	11	17	13	14	11	11	20	15	18	22	12	Q	45	1		1		
Pu90	13	15	17	13	31	24	10	16	13	14	11	11	21	18	17	22	12	Q	46	1	Relacionados			
Pu92	13	15	17	13	31	24	10	16	13	14	11	11	21	18	17	22	12	Q	46	1			1	
Pu94	13	15	17	13	31	24	10	16	13	14	11	11	21	18	17	22	12	Q	46	1				
Pu110	13	15	18	12	28	25	9	13	13	14	11	12	20	17	16	23	12	Q	40	0.79		1		
Pu111	13	15	18	13	29	23	11	14	12	14	12	11	19	15	17	23	11	Q	53	1		1		
Pu115	13	16	18	12	28	24	9	14	13	14	11	11	20	15	18	22	12	Q	47	0.99		1		
Pu113	13	16	18	12	28	24	9	14	13	14	11	12	20	15	18	22	12	Q	50	0.99		1		
Pu69	14	13	15	13	30	24	11	14	13	14	11	12	12	14	19	22	10	Q	21	0.95		1		
Pu93	14	13	15	14	32	25	11	14	13	14	11	12	18	14	19	22	10	Q	32	0.78		1		
Pu182	14	15	16	13	30	23	10	14	13	15	11	11	11	17	18	23	11	Q	28	0.99		1		
Pu155	16	13	16	12	26	24	10	14	13	14	11	11	11	16	15	22	12	Q	15	0.77		1		
Pu143	16	13	16	12	26	24	10	14	13	14	11	11	20	16	15	22	12	Q	30	0.79	Relacionados			
Pu144	16	13	16	12	26	24	10	14	13	14	11	11	20	16	15	22	12	Q	30	0.79			1	
Pu150	16	13	16	12	26	24	10	14	13	14	11	11	20	16	15	22	12	Q	30	0.79				
Pu112	16	14	16	12	26	24	10	14	13	14	11	11	19	16	15	22	12	Q	34	0.79				
Pu129	16	14	16	12	26	24	10	14	13	14	11	11	19	16	15	22	12	Q	34	0.79		4		
Pu142	16	14	16	12	26	24	10	14	13	14	11	11	19	16	15	22	12	Q	34	0.79				
Pu145	16	14	16	12	26	24	10	14	13	14	11	11	19	16	15	22	12	Q	34	0.79				
																					Total	46		

Tabla S212. Haplotipos y linajes encontrados en la población Mazahua. En colores se encuentran los individuos relacionados. HG: Haplogrupo. FS: Fitness Score.

	19	2385a	2385b	389I	389II	390	391	392	393	437	438	439	448	456	458	635	Y GATA H4	HG	FS	Prob.	Parentesco	Frecuencia relativa
Maz12	13	15	17	13	30	21	10	12	13	14	11	13	20	14	16	22	11	E1b1b	42	0.76		1
Maz53	14	13	15	13	29	23	10	11	12	14	9	12	20	15	16	21	13	J2a1b	65	0.69		1
Maz5	14	11	14	13	30	25	11	13	13	14	12	12	19	16	16	23	12	R1b	62	1		1
Maz29	14	11	14	14	31	24	11	13	12	14	12	13	19	15	16	23	13	R1b	40	1		
Maz37	14	11	14	14	31	24	11	13	12	14	12	13	19	15	16	23	13	R1b	40	1		
Maz40	14	11	14	14	31	24	11	13	12	14	12	13	19	15	16	23	13	R1b	40	1		4
Maz43	14	11	14	14	31	24	11	13	12	14	12	13	19	15	16	23	13	R1b	40	1		
Maz16	13	13	14	14	32	24	10	14	13	13	11	11	21	15	15	22	12	Q	43	1		1
Maz31	13	14	16	13	30	23	10	14	13	14	13	13	19	17	16	22	11	Q	68	1		1
Maz38	13	14	16	13	30	24	10	14	13	14	11	11	21	15	17	22	12	Q	70	1		1
Maz2	13	14	18	12	28	23	10	13	14	14	11	12	20	14	18	22	12	Q	49	1		1
Maz25	13	14	20	13	29	24	10	13	13	14	11	11	20	15	18	22	12	Q	61	1		
Maz26	13	14	20	13	29	24	10	13	13	14	11	11	20	15	18	22	12	Q	61	1	Relacionados	1
Maz32	13	14	20	13	29	24	10	13	13	14	11	11	20	15	18	22	12	Q	61	1		
Maz28	13	14	20	13	29	24	10	13	13	14	11	12	20	15	18	22	12	Q	65	1		1
Maz8	13	14	20	13	30	26	10	14	13	14	11	12	20	16	18	22	13	Q	42	1		
Maz24	13	14	20	13	30	26	10	14	13	14	11	12	20	16	18	22	13	Q	42	1		3
Maz35	13	14	20	13	30	26	10	14	13	14	11	12	20	16	18	22	13	Q	42	1		
Maz1	13	15	17	12	28	24	9	13	13	14	11	12	20	15	19	22	12	Q	48	0.99		1
Maz39	13	15	19	14	31	24	10	15	13	14	11	12	20	15	16	22	12	Q	60	0.96		1
Maz30	13	16	17	13	31	23	10	17	13	14	11	12	20	18	17	22	12	Q	41	0.99		1
Maz34	14	14	18	13	29	25	11	16	13	14	11	11	20	15	18	22	11	Q	49	1		
Maz36	14	14	18	13	29	25	11	16	13	14	11	11	20	15	18	22	11	Q	49	1	Relacionados	1
Maz22	14	15	17	13	30	24	10	15	13	14	11	12	19	16	18	22	12	Q	68	1		1
Maz33	14	16	16	14	32	24	10	14	13	14	10	12	20	15	16	22	12	Q	41	0.87		1
Maz47	14	16	17	13	29	24	11	14	13	14	10	12	20	17	16	22	11	Q	46	0.97		1
																					Total	23

Tabla S3. Haplotipos y linajes encontrados en la población Cora. En colores se encuentran los individuos relacionados. HG: Haplogrupo. FS: Fitness Score.

	19	2385a	2385b	389I	389II	390	391	392	393	437	438	439	448	456	458	635	Y GATA H4	HG	FS	Prob.	Parentesco	Frecuencia relativa																			
Co4	13	15	19	12	28	25	9	13	12	14	11	11	20	12	17	22	12	E1b1b	16	0.43		2																			
Co6	13	15	19	12	28	25	9	13	12	14	11	11	20	12	17	22	12	E1b1b	16	0.43																					
Co47	13	17	18	12	29	24	10	13	15	14	11	12	20	15	20	22	12	E1b1b	30	0.91		1																			
Co33	14	15	19	12	29	24	9	13	12	14	11	11	20	12	17	22	12	J1	17	0.65	Relacionados	1																			
Co49	14	15	19	12	29	24	9	13	12	14	11	11	20	12	17	22	12	J1	17	0.65																					
Co43	14	11	14	14	31	24	11	13	12	14	12	13	19	15	16	23	13	R1b	40	100		1																			
Co1	14	12	14	14	30	24	11	13	13	14	12	12	18	15	17	23	11	R1b	62	100		1																			
Co3	13	15	17	14	30	24	10	15	13	14	11	12	20	17	16	22	12	Q	64	0.99	Relacionados	2																			
Co21	13	15	17	14	30	24	10	15	13	14	11	12	20	17	16	22	12	Q	64	0.99																					
Co29	13	15	17	14	30	24	10	15	13	14	11	12	20	17	16	22	12	Q	64	0.99																					
Co37	13	15	17	14	30	24	10	15	13	14	11	12	20	17	16	22	12	Q	64	0.99																					
Co9	13	15	17	14	30	24	10	15	13	14	11	12	20	17	16	22	12	Q	64	0.99	Relacionados																				
Co26	13	15	17	14	30	24	10	15	13	14	11	12	20	17	16	22	12	Q	64	0.99																					
Co36	13	15	17	14	30	24	10	15	13	14	11	12	20	17	16	22	12	Q	64	0.99																					
Co16	13	15	18	14	31	24	11	14	13	14	11	12	21	17	16	22	12	Q	52	1		1																			
Co22	13	15	19	12	28	25	9	13	12	14	11	11	20	12	17	22	12	Q	18	0.48																					
Co24	13	15	19	12	28	25	9	13	12	14	11	11	20	12	17	22	12	Q	18	0.48	Relacionados	2																			
Co45	13	15	19	12	28	25	9	13	12	14	11	11	20	12	17	22	12	Q	18	0.48																					
Co15	14	11	17	14	31	24	10	14	13	14	11	11	20	15	17	22	11	Q	59	1		1																			
Co10	14	16	17	13	30	24	10	14	13	14	11	11	19	15	17	22	11	Q	72	1		1																			
Total																																									12

Tabla S4. Haplotipos y linajes encontrados en la población Wirrárika. En colores se encuentran los individuos relacionados. HG: Haplogrupo. FS: Fitness Score.

	19	2385a	2385b	389I	389II	390	391	392	393	437	438	439	448	456	458	635	Y GATA H4	HG	FS	Prob	Parentesco	Frecuencia relativa
	17	15	16	12	29	21	11	11	11	14	11	14	19	14	16	21	11	E1b1a	31	0.99		1
Wi85	13	15	17	14	31	24	10	15	10	14	11	12	21	16	12	22	12	E1b1b	13	0.63	Relacionados	1
Wi92	13	15	17	14	31	24	10	15	10	14	11	12	21	16	12	22	12	E1b1b	13	0.63		
Wi81	13	15	18	14	31	24	11	15	11	14	11	12	20	16	12	22	12	E1b1b	18	0.88		1
Wi77	14	13	16	13	29	24	10	14	10	16	11	11	20	15	16	22	11	L	50	1	Relacionados	1
Wi89	14	13	16	13	29	24	10	14	10	16	11	11	20	15	16	22	11	L	50	1		
Wi73	14	15	17	13	30	24	10	14	10	15	11	11	20	15	17	22	11	L	39	0.97		1
Wi52	15	14	14	12	31	21	10	11	10	14	11	11	21	17	20	21	11	G2a	16	0.95		1
Wi75	13	14	18	12	29	24	10	13	10	14	11	12	20	16	17	22	12	Q	28	0.94		1
Wi82	13	14	18	12	29	24	11	13	11	14	11	12	20	16	17	22	12	Q	38	0.99		1
Wi55	13	14	21	13	30	25	10	14	10	14	11	13	20	15	17	24	12	Q	22	0.99	Relacionados	1
Wi90	13	14	21	13	30	25	10	14	10	14	11	13	20	15	17	24	12	Q	22	0.99		
Wi70	13	15	17	13	29	24	11	15	11	14	11	12	20	17	17	22	12	Q	42	0.99		2
Wi88	13	15	17	13	29	24	11	15	11	14	11	12	20	17	17	22	12	Q	42	0.99		
Wi51	13	15	17	13	30	24	11	15	11	14	11	12	20	17	17	22	12	Q	41	0.99		1
Wi64	13	15	17	14	31	24	10	15	10	14	11	12	20	17	16	22	12	Q	28	0.84		
Wi71	13	15	17	14	31	24	10	15	10	14	11	12	20	17	16	22	12	Q	28	0.84		3
Wi93	13	15	17	14	31	24	10	15	10	14	11	12	20	17	16	22	12	Q	28	0.84		
Wi53	13	15	17	14	31	24	11	15	11	14	11	12	20	17	17	22	12	Q	39	0.99		1
																					Total	16

Tabla S5. Frecuencias alélicas de los 17 marcadores YSTR en la población P'urhépecha Q-M3.

P'urhépechas (n=26)	DYS19	DYS2385a	DYS2385b	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS437	DYS438	DYS439	DYS448	DYS456	DYS458	DYS635	Y GATA H4
Alelo																	
9							0.192										
10							0.577										
11							0.231				0.692	0.231	0.038				0.077
12				0.231							0.308	0.731	0.308				0.769
13	0.962	0.500		0.462				0.115	0.885			0.038					0.154
14		0.231		0.308				0.500	0.115	1.000				0.038	0.308		
15		0.192						0.115						0.654	0.077		
16	0.038	0.077	0.077					0.115						0.077	0.154		
17				0.269				0.154						0.192	0.231		
18				0.615										0.038	0.231		
19			0.038										0.077				
20													0.462				
21													0.115				
22																0.885	
23																0.115	
24							0.462										
25							0.462										
26					0.038	0.077											
28					0.192												
29																	
30						0.385											
31						0.385											
k	2	4	4	3	4	3	3	5	2	1	2	3	5	5	5	2	3
k= 56																	

Tabla S6. Frecuencias alélicas de los 17 marcadores YSTR en la población Mazahua Q-M3.

Mazahuas (n=12)	DYS19	DYS2385a	DYS2385b	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS437	DYS438	DYS439	DYS448	DYS456	DYS458	DYS635	Y GATA H4
Alelo																	
9							0.083										
10							0.833										
11							0.083				1.000	0.333					0.083
12				0.167								0.667					0.667
13	0.833	0.083		0.667				0.250	0.917	0.083							0.250
14	0.167	0.583	0.083	0.167				0.417	0.083	0.917				0.083			
15		0.250						0.167						0.500	0.083		
16		0.083	0.083					0.083						0.333	0.083		
17			0.250					0.083									0.167
18			0.167											0.083	0.583		
19			0.083										0.083		0.083		
20			0.333											0.750			
21													0.167				
22																1.000	
23						0.167											
24						0.500											
25						0.083											
26						0.250											
28					0.167												
29					0.167												
30					0.417												
31					0.167												
32					0.083												
k	2	4	6	3	5	4	3	5	2	2	1	2	3	4	5	1	3
k= 55																	

Tabla S7. Frecuencias alélicas de los 17 marcadores YSTR en la población Cora Q-M3.

Coras (n=7)	DYS19	DYS2385a	DYS2385b	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS437	DYS438	DYS439	DYS448	DYS456	DYS458	DYS635	Y GATA H4
Alelo																	
9							0.286										
10							0.571										
11		0.143					0.143				1.000	0.571					0.286
12				0.286					0.286			0.429		0.286			0.714
13	0.714			0.143				0.286	0.714								
14	0.286			0.571				0.429		1.000							
15		0.714						0.286						0.286			
16		0.143														0.429	
17			0.571											0.429	0.571		
18			0.143														
19			0.286										0.143				
20													0.714				
21													0.143				
22																1.000	
23																	
24						0.714											
25						0.286											
26																	
28					0.286												
29																	
30					0.429												
31					0.286												
32																	
<i>k</i>	2	3	3	3	3	2	3	3	2	1	1	2	3	3	2	1	2
<i>k=39</i>																	

Tabla S8. Frecuencias alélicas de los 17 marcadores YSTR en la población Wírrárika Q-M3.

Wírrárikas (n=10)	DYS19	DYS2385a	DYS2385b	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS437	DYS438	DYS439	DYS448	DYS456	DYS458	DYS635	Y GATA H4
Alelo																	
10							0.500		0.500								
11							0.500		0.500	1.000							
12				0.200								0.900					1.000
13	1.000			0.400				0.200				0.100					
14		0.300		0.400				0.100		1.000							
15		0.700						0.700									
16														0.100			
17														0.200	0.300		
18				0.200													
19																	
20													1.000				
21			0.100														
22																	0.900
23																	
24							0.900										0.100
25							0.100										
26																	
28																	
29							0.400										
30							0.200										
31							0.400										
32																	
k	1	2	3	3	3	2	2	3	2	1	1	2	1	3	2	2	1
k=34																	

Tabla S9. Frecuencias alélicas de los 17 marcadores YSTR en la población P'urhépecha Q-L54.

P'urhépechas (n=14)	DYS19	DYS2385a	DYS2385b	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS437	DYS438	DYS439	DYS448	DYS456	DYS458	DYS635	Y GATA H4
Alelo																	
10							0.714										0.143
11							0.286				0.571	0.500	0.071				0.500
12				0.357					0.071		0.429	0.429	0.214				0.357
13	0.429	0.214		0.571					0.929								
14	0.214	0.357		0.071				1.000		0.929		0.071		0.143			
15		0.429	0.143							0.071				0.071	0.357		
16	0.357		0.714											0.500	0.357		
17			0.071											0.286	0.071		
18			0.071										0.071		0.071		
19													0.571		0.143		
20													0.071				
22																	0.857
23						0.500											0.143
24						0.429											
25						0.071											
26					0.357												
29					0.429												
30					0.143												
32					0.071												
k	3	3	4	3	4	3	2	1	2	2	2	3	5	4	5	2	3
k=51																	

Tabla S10. Frecuencias alélicas de los 17 marcadores YSTR en la población Mazahua Q-L54.

Mazahuas (n=14)	DYS19	DYS2385a	DYS2385b	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS437	DYS438	DYS439	DYS448	DYS456	DYS458	DYS635	Y GATA H4
Alelo																	
10						0.667					0.667						
11						0.333											0.667
12												0.667					0.333
13	0.333			0.667					1.000		0.333	0.333					
14	0.667	0.333		0.333				1.000		1.000							
15														0.333			
16		0.667	0.667												1.000		
17			0.333											0.667			
19													0.333				
20													0.667				
22																1.000	
23						0.333											
24						0.667											
29					0.333												
30					0.333												
32					0.333												
k	2	2	2	2	3	2	2	1	1	1	2	2	2	2	1	1	2
k=30																	

Tabla S11. Poblaciones utilizadas con fines de comparación.

Población	Origen	Clave	Análisis	Referencia
Amerindia/México	Mazahua	Mz	Red	Este estudio
	Nayeri	Co	Red	
	P'urhépecha	Pu	Red	
	Wirrárika	Wi	Red	
	Zapoteca	Zap	Red	
	Mixe	Mix	Red	Kemp et al., 2010
	Tarahumara	Tar	Red	
	Pima	Pim	Red	
	Mixteco	Mit	Red	
	Nahua	Nah	Red	
	Amerindios de México	Mex	Red	Sandoval et al., 2012
Yucatán (Maya)	Yuc	Red	Battaglia et al., 2013	
Amerindia/Estados Unidos	Cherokee	Che	Red	Bonick, Bolnick, & Glenn-Smith, 2006
	Chickasaw	Chi	Red	
	Choctaw	Cho	Red	
	Seminole	Sem	Red	
	Sioux	Sio	Red	
	Creek	Cre	Red	
	Cheyene	Che	Red	
	Jemez	Jem	Red	Kemp et al., 2010
Amerindia/Canadá, Alaska	Metis canadiense	Met	Red	Dulik et al., 2012
	Tlicho	Ttj	Red	
	Tlingit	Tli	Red	
Amerindia/Sudamérica, Caribe	Cakchikel (Maya)	Cak	Red	Regueiro et al., 2013
	Panamá	Pan	Red	Battaglia et al., 2013
	Brasil	Bra	Red	
	Chile	Chl	Red	
	Colombia	Col	Red	
	Perú	Per	Red	
	Bermudas	Ber	Red	Dulik et al., 2012
Asia/Siberia	Evens	Eve	Red	Malyarchuk et al., 2011
	Noroeste de Siberia	Nes	Red	Regueiro et al., 2013
	Altai	Alt	Red	Dulik et al., 2012

Bibliografía de Tablas Suplementarias

153. Kemp, B. *et al.* Evaluating the Farming/Language Dispersal Hypothesis with genetic variation exhibited by populations in the Southwest and Mesoamerica. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 6759–6764 (2010).
154. Sandoval, K. *et al.* Y-chromosome diversity in Native Mexicans reveals continental transition of genetic structure in the Americas. *Am. J. Phys. Anthropol.* **148**, 395–405 (2012).
155. Bonick, D. A., Bolnick, D. I. & Glenn-Smith, D. Asymmetric Male and Female Genetic Histories among Native Americans from Eastern North America. *Mol. Biol. Evol.* **23**, 2161–2174 (2006).
156. Dulik, M. C. *et al.* Y-chromosome analysis reveals genetic divergence and new founding native lineages in Athapaskan- and Eskimoan-speaking populations. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 8471–8476 (2012).
157. Malyarchuk, B. *et al.* Ancient links between Siberians and Native Americans revealed by subtyping the Y chromosome haplogroup Q1a. *J. Hum. Genet.* **56**, 583–588 (2011).