

Universidad de Costa Rica  
Sistema de Estudios de Posgrado

*Análisis de estructura genética y diversidad de tres  
poblaciones de Costa Rica, México y el Suroeste de Estados  
Unidos relacionadas con la esquizofrenia, utilizando  
marcadores Y-STRs y ADNmt*

Tesis sometida a la consideración de la Comisión del Programa de Estudios de Posgrado  
en Microbiología, Parasitología y Química Clínica para optar por el grado de  
*Magister Scientiae*

**REBECA CAMPOS SÁNCHEZ**

Ciudad Universitaria “Rodrigo Facio” Costa Rica

2005

## DEDICATORIA

Porque mis pensamientos no son los de ustedes, ni sus caminos son los míos

– afirma el Señor –.

Mis caminos y mis pensamientos son más altos que los de ustedes;

¡más altos que los cielos sobre la tierra!

Isaías 55: 8-9

Dios, tú eres mi aliento de vida, por ti es que sigo adelante y cada reto al que me enfrento representa una nueva oportunidad para conocerte más. Este es uno de esos retos que se convierten en victoria por tu fidelidad. Espero honrarte a ti con este trabajo.

Papi y mami, esta es otra evidencia de lo bien que han hecho las cosas. Esto es símbolo de mi esfuerzo que les entrego por todo lo que han dado por mi educación.

## AGRADECIMIENTOS

Al único Dios vivo, que me ha hecho su hija, le agradezco porque me mantuvo en pie, me dio sabiduría y me guió para hacer lo correcto y culminar esta bendición con éxito.

En definitiva, durante todo este proceso no me sentí sola porque muchas personas me acompañaron con sus oraciones y les agradezco inmensamente por esto, que significó un apoyo muy especial para mí.

A la Dra. Sandra Silva, una vez más le agradezco porque su influencia y apoyo en mi vida ha sido único, no solo por su ejemplo, sabiduría y ayuda, sino por el cariño con que me mostró las cosas al guiarme a ser una mejor profesional.

Al Dr. Ramiro Barrantes, porque llegó en el momento en que más lo necesitaba para ampliar mi conocimiento y darme de su tiempo para aconsejarme y guiarme en lo que nadie más podía hacer.

A la Dra. Henriette Raventós, porque me permitió aprender y crecer en conocimiento, y que en medio de su tiempo tan reducido pudiera darme las respuestas que necesitaba, así como la motivación oportuna.

A las personas del CIBCM que durante mi trabajo en el laboratorio fueron mis maestros, mi apoyo y mis ayudantes.

A mis amigos y amigas, que siguieron cada avance de mi estudio con interés genuino para hacerme sentir su apoyo. De verdad, todos son una bendición.

Y a todas las personas que me guiaron en el uso de los programas estadísticos, que me brindaron sus publicaciones y consejos inigualables.

Gracias por su tiempo.

Esta tesis fue aceptada por la Comisión del Programa de Estudios de Posgrado en Microbiología, Parasitología y Química Clínica de la Universidad de Costa Rica, como requisito parcial para optar al grado de *Magíster Scientiae*



Gabriel Macaya Trejos, Ph.D  
Representante del Decano del Sistema de Estudios de Posgrado



Sandra Silva de la Fuente, M.Sc.  
Directora de Tesis



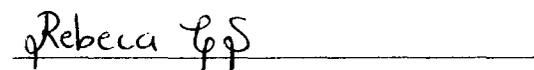
Henriette Raventós Vorst, M.Sc.  
Asesora



Ramiro Barrantes Mesén, Ph.D.  
Asesor



Fernando García Santamaría, Ph.D.  
Director del Programa de Posgrado en Microbiología, Parasitología y Química Clínica



Rebeca Campos Sánchez  
Candidata

## ÍNDICE

	Página
<b>Portada</b>	i
<b>Dedicatoria</b>	ii
<b>Agradecimientos</b>	iii
<b>Hoja de aprobación</b>	iv
<b>Índice</b>	v
<b>Resumen</b>	vii
<b>Abstract</b>	x
<b>Lista de cuadros</b>	xiii
<b>Lista de figuras</b>	xv
<b>Lista de abreviaturas</b>	xvi
<b>Introducción</b>	1
<b>A. Estudios poblacionales</b>	1
<b>B. El cromosoma Y</b>	4
<b>C. El ADN mitocondrial</b>	8
<b>D. Estudios genéticos realizados en las poblaciones de estudio</b>	13
<b>D.1. Costa Rica</b>	13
<b>D.2. México</b>	17
<b>D.3. Estados Unidos de América</b>	19
<b>E. La esquizofrenia</b>	21
<b>Objetivos</b>	25
<b>Materiales y métodos</b>	26
<b>A. Muestras de ADN</b>	26
<b>B. Microsatélites del cromosoma Y</b>	26
<b>C. Secuencia hipervariable I del ADNmt</b>	29
<b>D. Análisis estadístico</b>	31
<b>D.1 Medidas de diversidad genética</b>	31
<b>D.2 Estructura jerárquica de las poblaciones</b>	34
<b>D.3 Relaciones entre poblaciones</b>	36
<b>D.4 Programas estadísticos</b>	39

<b>Resultados</b>	40
<b>A. Diversidad genética de las poblaciones de Costa Rica, México y del Suroeste de EUA relacionadas con la esquizofrenia</b>	40
<b>A.1. Caracterización de la diversidad genética del cromosoma Y</b>	40
<b>A.2. Caracterización de la diversidad genética del ADNmt</b>	51
<b>B. Análisis de la estructura jerárquica de las poblaciones de Costa Rica, México y el Suroeste de EUA relacionadas con la esquizofrenia</b>	62
<b>C. Relaciones entre las poblaciones de Costa Rica, México y el Suroeste de EUA ligadas con la esquizofrenia</b>	67
<b>Discusión</b>	74
<b>A. Diversidad genética de las poblaciones de Costa Rica, México y del Suroeste de EUA relacionadas con la esquizofrenia</b>	74
<b>A.1. Caracterización de la diversidad genética del cromosoma Y</b>	74
<b>A.2. Caracterización de la diversidad genética del ADNmt</b>	78
<b>B. Análisis de la estructura jerárquica de las poblaciones de Costa Rica, México y el Suroeste de EUA relacionadas con la esquizofrenia</b>	80
<b>C. Relaciones entre las poblaciones de Costa Rica, México y el Suroeste de EUA relacionadas con la esquizofrenia</b>	83
<b>Conclusiones</b>	89
<b>Referencias</b>	91
<b>Apéndice</b>	108
<b>A. Haplotipos encontrados en las tres poblaciones de estudio respecto a 12 Y-STR</b>	109
<b>B. Haplotipos del ADNmt para la región HVI en las tres poblaciones de estudio</b>	114
<b>C. Publicación: Genetic structure analysis of three Hispanic populations from Costa Rica, Mexico and the Southwest United States, using Y-STRs markers and mtDNA sequence</b>	120

*Información descriptiva*

Campos Sánchez, Rebeca

Análisis de estructura genética y diversidad de tres poblaciones de Costa Rica, México y el Suroeste de Estados Unidos relacionadas con la esquizofrenia, utilizando marcadores Y-STRs y ADNmt

Tesis de posgrado. – San José, C.R.:

R. Campos S., 2005.

147h.:il. -112 refs.

Se estudiaron 217 individuos masculinos hispanos relacionados con la esquizofrenia provenientes de Costa Rica (CR 120 individuos), el suroeste de Estados Unidos de América (EUA 42 individuos) y México (MX 55 individuos) utilizando 12 Y-STR (DYS19, DYS 385I/II, DYS389I/II, DYS390, DYS391, DYS392, DYS437, DYS438 y DYS439) del kit PowerPlex®Y System (Promega) y la secuencia de la región HVI del ADNmt (imprimadores L15997 y H16401).

Los datos de los microsatélites del cromosoma Y muestran una diversidad génica promedio en Costa Rica de 0,675, en EUA de 0,614 y en México de 0,608. Se obtuvieron 103 haplotipos en CR, dentro de estos se obtuvieron 14 haplotipos compartidos. En la población de EUA se obtuvieron 40 haplotipos, entre ellos sólo 2 compartidos y en la población mexicana se obtuvo un total de 51 haplotipos, de los cuales 2 son compartidos. Para las tres poblaciones la diversidad haplotípica fue de 99%.

El número promedio de diferencias entre individuos (“pairwise differences”) dentro de cada población fue de 8,099, 7,368 y 7,300 para CR, EUA y MX respectivamente. Al graficar la distribución de diferencias (“Mismatch distribution”) se observa una distribución unimodal que muestra una expansión masculina común a las tres poblaciones. No se hallaron haplotipos compartidos entre las tres poblaciones de estudio para los Y-STR, sin embargo, entre EUA y MX se hallaron 4 haplotipos compartidos y 1 entre CR y EUA. Esto es evidencia de migración o de una historia compartida

relacionada con la colonia y el ingreso de españoles y portugueses, principalmente; además de linajes amerindios y mezcla con africanos.

La estructura de las poblaciones para los Y-STR mostró un valor  $R_{st}$  de 0,04 ( $P < 0,001$ ). Específicamente, los marcadores que influyen en este resultado son el DYS392 y el DYS439. Al separar las poblaciones en regiones los resultados fueron significativos ( $R_{st}: 0,038$ ,  $P < 0,01$ ), siendo los marcadores DYS389II, DYS392 y DYS439 relevantes en dicho resultado. A pesar de estos resultados, el porcentaje de diferenciación dentro de las poblaciones siempre supera el 94,8%, lo cual revela que las tres poblaciones relacionadas con la esquizofrenia son más diferentes internamente que entre ellas al evaluar su estructura genética.

Las secuencias del ADNmt presentan principalmente sustituciones respecto a la secuencia Anderson, siendo las transiciones más comunes que las transversiones. En la población costarricense se presentan 52 sitios polimórficos, en EUA 39 y en MX 56, de estos hay un conjunto de polimorfismos únicos en la población, específicamente 18 en CR, 10 en EUA y 19 en MX. Para las tres poblaciones la diversidad de nucleótidos se aproxima a 0,02 y el promedio de diferencias entre individuos (“Pairwise differences”) para CR es de 7,6; para EUA es de 6,8 y para MX es de 8,4. En estos casos cuando se grafica la distribución de diferencias (“Mismatch distribution”), las tres poblaciones revelan un comportamiento bimodal similar, que refleja relación entre ellas y un evento de expansión particular para las tres, confirmado por el índice de desigualdad de Harpending ( $< 0,03$ ).

Los haplotipos del ADNmt obtenidos para CR son 39, revelando una diversidad haplotípica de 0,93. De estos haplotipos, 17 son compartidos dentro de la misma población. En EUA el número de haplotipos es de 33, teniendo 4 compartidos dentro de la población y una diversidad haplotípica de 0,981. En MX se obtuvieron 43 haplotipos, 8 compartidos dentro de la población y la diversidad haplotípica obtenida de 0,989.

Entre las tres poblaciones, se observaron 2 haplotipos compartidos, situación que no se observó con los Y-STR. Particularmente, el haplotipo HM4 se representó en casi el

16% de la población de CR, en 10% de la población de EUA y 7% de la población de MX. Otro haplotipo compartido por las tres poblaciones es el HM10. Por otro lado, hay un haplotipo compartido entre CR y MX, y 6 haplotipos compartidos entre EUA y MX.

El AMOVA realizado para los datos de la región HVI mostró que las tres poblaciones de estudio (CR, EUA y MX) se diferencian mayormente dentro de ellas mismas ( $F_{st}:0,006$ ,  $P:0,162$ ). Sucede lo mismo cuando se separan las poblaciones en regiones ( $F_{st}:0,005$ ,  $P:0,304$ ).

Las distancias genéticas mostraron que para los linajes masculinos (Y-STR) y los linajes femeninos (ADNmt) las tres poblaciones no son diferentes. Y las representaciones gráficas (árbol de agrupamiento de vecinos y el escalamiento multidimensional) revelan un agrupamiento para las regiones más cercanas geográficamente, como era de esperarse por su relación migratoria.

En resumen, el análisis de la región HVI revela una información muy similar a los Y-STR. A pesar de que los datos de los Y-STR reflejan diferenciación entre las poblaciones de CR, EUA y MX, los valores de los mismos no son relevantes para concluir una diferencia determinante entre las poblaciones. Por otro lado, vale la pena resaltar que las poblaciones están comunicadas por las migraciones y esto se evidencia en los haplotipos compartidos, especialmente entre MX y EUA por su cercanía geográfica y relaciones políticas hasta mediados del siglo XIX. Además, estas poblaciones tienen una historia compartida y similar de colonizaciones, mestizaje y migraciones que pueden confirmar dicha relación.

Palabras clave: Y-STR; ADNMT; ESTRUCTURA GENÉTICA; DIVERSIDAD GENÉTICA; COSTA RICA; ESTADOS UNIDOS; MÉXICO; ESQUIZOFRENIA

Director de investigación: Sandra Silva de la Fuente, M.Sc

Unidad Académica: Sistema de Estudios de Posgrado

## ABSTRACT

Two hundred and seventeen Hispanic men related to schizophrenia belonging to Costa Rica (120 individuals), the Southwest United States of America (42 individuals) and Mexico (55 individuals) were studied using PowerPlex®Y System (Promega) which evaluates 12 Y-STRs (DYS19, DYS 385I/II, DYS389I/II, DYS390, DYS391, DYS392, DYS437, DYS438 and DYS439), and the HVI sequence from the mtDNA (primers L15997 and H16401).

Y-chromosome microsatellite data showed a mean genetic diversity of 0.675 for CR, 0.614 for USA, and 0.608 for MX. In CR, 103 haplotypes were identified, from these 14 were shared inside the population. USA presented 40 haplotypes, and just 2 were shared inside the population. MX showed 51 haplotypes, from which 2 were shared inside the population. All three populations showed an haplotype diversity around 99%.

The mean number of pairwise differences between individuals inside each population was 8.099, 7.368, and 7.300 for CR, USA, and MX, respectively. The mismatch distribution was unimodal, implying a common male expansion event for these three populations. None shared haplotypes were found among all three populations, but, USA and MX shared 4 haplotypes, and CR and USA shared 1. This is evidence of migration and/or shared history related to the colonization period (Portuguese and Spanish colonizers) or to Amerindian lineages and African admixture.

Y-STR  $R_{st}$  value was 0.04 ( $P < 0.001$ ) which is significant. Specifically, the markers involved in this difference were DYS392 and DYS439. Also, when the populations were separated into regions there was a significant result ( $R_{st}: 0.038$ ,  $P < 0.01$ ), in this case DYS389II, DYS392, and DYS439 were relevant. Even though, always the percentage of differentiation inside populations overcomes 94.8%, this reveals that the three populations related to schizophrenia are more different internally than between them in their genetic structure.

Substitutions were the most representative polymorphisms in the mitochondrial DNA sequences. Even more common were transitions over transversions. In CR, 52 polymorphic sites were found, 39 in USA and 56 in MX. Some unique polymorphisms were shown, 18 in CR, 10 in USA, and 19 in MX. Nucleotide diversity was 0.02 for all three populations and the mean number of pairwise differences between individuals was 7.6 for CR, 6.8 for USA, and 8.4 for MX. The plot of the mismatch distribution showed a similar bimodal distribution for the three populations which reflects the relationship between the populations and a common expansion event, confirmed by Harpending's raggedness index ( $<0.03$ ).

CR showed 39 mtDNA haplotypes with a diversity of 0.93. From these haplotypes, 17 were shared inside the population. USA achieved 33 haplotypes and a diversity of 0.981, with 4 haplotypes shared inside the population. And MX obtained 43 haplotypes, sharing 8 inside the population, and the haplotype diversity was 0.989.

Two haplotypes were shared among the three populations for mtDNA sequences, contrary to Y-STR haplotypes. The haplotype HM4 was represented in 16% of the CR sample, in 10% of the USA sample, and 7% of the MX sample. Another haplotype shared between these three populations was HM10. There is also one haplotype shared between CR and MX, and 6 haplotypes shared between USA and MX.

The AMOVA based on mtDNA data showed that these three populations are differentiated inside of them ( $F_{st}:0.006$ ,  $P:0.162$ ), predominantly. The same situation was observed when the populations were separated into regions ( $F_{st}:0.005$ ,  $P:0.304$ ).

Genetic distances showed that the three populations are not different for the maternal and paternal lineages. Also, the geographic neighbors were shown together on the graphic representations of this data (neighbor joining tree and multidimensional scaling), as was expected from previous relationships between the regions.

In summary, the analysis of HVI (maternal lineage) revealed a very similar picture compared to Y-STRs (paternal lineage). Though, the Y-chromosome reflected differentiation between CR, USA and MX, the value of this measure is not relevant to

conclude a real difference between these populations. It could be remarked that these populations are related by migration showed by shared haplotypes, especially between MX and USA because of their geographic closeness and political relationship before the mid-19<sup>th</sup> century. Even more, these populations have a common shared history and similar events of colonization, admixture and migrations that can confirm this relationship.

## LISTA DE CUADROS

	Página
<b>Cuadro 1.</b> Secuencia repetitiva, número de alelos y número de acceso en GenBank de los 12 Y-STR analizados en las tres poblaciones de CR, MX y EUA	28
<b>Cuadro 2.</b> Frecuencias y diversidades génicas (DG) obtenidas al analizar 12 Y-STR en Costa Rica (CR), Estados Unidos (EUA) y México (MX)	43
<b>Cuadro 3.</b> Medidas de diversidad dentro de las poblaciones relacionadas con la esquizofrenia de Costa Rica, Estados Unidos y México al analizar 12 Y-STR	46
<b>Cuadro 4.</b> Comparaciones de las poblaciones de Costa Rica, Estados Unidos y México analizadas con 12 Y-STR respecto a otros estudios con microsatélites del cromosoma Y	47
<b>Cuadro 5.</b> Haplotipos compartidos entre y dentro de las tres poblaciones de estudio (Costa Rica, Estados Unidos y México) al analizarse 12 Y-STR	48
<b>Cuadro 6.</b> Número de diferencias entre pares de individuos en Costa Rica, Estados Unidos y México a partir de 12 Y-STR analizados	50
<b>Cuadro 7.</b> Sustituciones presentes en la región HVI del ADNmt para las poblaciones analizadas de Costa Rica, Estados Unidos y México	53
<b>Cuadro 8.</b> Evaluación de la diversidad en la región HVI del ADNmt dentro de la población de Costa Rica, Estados Unidos y México	57
<b>Cuadro 9.</b> Comparaciones con otros estudios al analizar la región control del ADNmt en las poblaciones de Costa Rica, Estados Unidos y México	58
<b>Cuadro 10.</b> Haplotipos compartidos dentro y entre las poblaciones de Costa Rica, Estados Unidos y México al estudiar la región HVI del ADNmt	60
<b>Cuadro 11.</b> Número de diferencias entre individuos en los haplotipos del ADNmt dentro de la muestra de Costa Rica, Estados Unidos y México	61

<b>Cuadro 12.</b> Análisis molecular de varianza en las tres poblaciones de estudio o separadas por regiones al analizar la secuencia HVI del ADNmt	63
<b>Cuadro 13.</b> Análisis molecular de varianza con 12 Y-STR y <i>locus</i> por <i>locus</i> en las tres poblaciones de Costa Rica, México y el Suroeste de Estados Unidos, y separadas por regiones	64
<b>Cuadro 14.</b> Análisis molecular de varianza por cada <i>locus</i> de los 12 Y-STR analizados en Costa Rica, Estados Unidos y México	65
<b>Cuadro 15.</b> Análisis molecular de varianza por cada <i>locus</i> de los 12 Y-STR estudiados en seis regiones de Costa Rica, el Suroeste de Estados Unidos (San Antonio, San Diego y Los Ángeles) y México (Monterrey y Ciudad de México)	66
<b>Cuadro 16.</b> Distancias genéticas para los 12 Y-STR (Nei 1972, arriba de la diagonal) y la región HVI del ADNmt (Tamura y Nei 1993, debajo de la diagonal) en Costa Rica, Estados Unidos y México	68
<b>Cuadro 17.</b> Distancias genéticas para los 12 Y-STR (Nei 1972, arriba de la diagonal) y la región HVI del ADNmt (Tamura y Nei 1993, debajo de la diagonal) en las tres poblaciones de estudio, divididas por regiones	69

## LISTA DE FIGURAS

	<b>Página</b>
<b>Figura 1.</b> Estructura del cromosoma Y y localización de ciertos microsátelites analizados en poblaciones humanas	7
<b>Figura 2.</b> Distribución del genoma del ADNmt, especificando la región control	11
<b>Figura 3.</b> Distribución geográfica de las poblaciones analizadas en el presente estudio	27
<b>Figura 4.</b> Posición de los 12 microsátelites del cromosoma Y analizados en el presente estudio	28
<b>Figura 5.</b> Distribución de diferencias (“Mismatch distribution”) entre pares de individuos para Costa Rica, Estados Unidos y México al analizar 12 Y-STR	51
<b>Figura 6.</b> Distribución de diferencias (“Mismatch distribution”) en las poblaciones de Costa Rica, Estados Unidos y México a partir de la secuencia HVI del ADNmt	62
<b>Figura 7.</b> Árbol de agrupamiento de vecinos basado en el estudio de 12 Y-STR para Costa Rica, el Suroeste de Estados Unidos y México, separado en regiones	70
<b>Figura 8.</b> Escalamiento Multidimensional basado en los datos de 12 Y-STR para las poblaciones hispanas de Costa Rica, el suroeste de Estados Unidos y México separadas por regiones	71
<b>Figura 9.</b> Árbol de agrupamiento de vecinos basado en la región HVI del ADNmt para Costa Rica, Estados Unidos y México, separado en regiones basado en las distancias de Tamura y Nei	72
<b>Figura 10.</b> Escalamiento Multidimensional basado en los datos de la región HVI del ADNmt para las poblaciones hispanas de Costa Rica, el suroeste de Estados Unidos y México separadas por regiones	73

## LISTA DE ABREVIATURAS

A	Adenina	MT	Monterrey
ADN	Ácido desoxiribonucleico	MX	México
ADNmt	ADN mitocondrial	NJ	Árbol de agrupamiento de vecinos ("Neighbor joining")
AMOVA	Análisis Molecular de Varianza	NRY	Región no recombinante del cromosoma Y
ARN	Ácido ribonucleico	OTU	Unidades taxonómicas operacionales
C	Citosina	Pb	Pares de bases
CEPAL	Comisión Económica para América Latina y el Caribe	PCR	Reacción en cadena de la polimerasa
CEPH	Centro de Estudio de los Polimorfismos Humanos	RFLP	Polimorfismos de longitud de fragmentos de restricción
CM	Ciudad de México	R <sub>st</sub>	Medida de diferenciación entre poblaciones
CR	Costa Rica	SA	San Antonio
CRS	Secuencia de referencia Cambridge	SD	San Diego
EUA	Estados Unidos de América	SNP	Polimorfismo de un nucleótido
F <sub>st</sub>	Medida de diferenciación entre poblaciones	SSO	Oligonucleótido de sonda específica
G	Guanina	STR	Microsatélite ("Short Tandem Repeat")
GEP- ISFG	Grupo Español y Portugués de la Sociedad Internacional de Genética Forense	T	Timina
H	Hebra pesada del ADNmt	VCCR	Valle Central de Costa Rica
HM	Haplotipo del ADNmt	YHRD	Base de Datos de Referencia de Haplotipos del Cromosoma Y
HVI	Región Hipervariable I	Y-STR	Microsatélites del cromosoma Y
HVII	Región Hipervariable II	µl	Microlitros
HY	Haplotipo del cromosoma Y	µM	Micromolar
L	Hebra ligera del ADNmt		
LA	Los Ángeles		
LOD	Logaritmo de las probabilidades que favorecen el ligamiento ("Log of the Odds")		
Mb	Mega bases		
MDS	Escalamiento Multidimensional		
MSY1	Minisatélite del cromosoma Y		

## INTRODUCCIÓN

### A. Estudios Poblacionales

Los estudios de la genética de poblaciones involucran la evaluación de una o varias muestras, ya sea de una región geográfica o grupo étnico, para conocer y revelar las características que los marcadores seleccionados manifiesten, indicando eventos pasados que sucedieron por movimientos o comportamientos de la población. Esto ha sido posible gracias a los marcadores moleculares genéticos, que han mostrado importantes respuestas a las preguntas de la evolución humana, migración y las relaciones históricas entre poblaciones separadas (Cavalli-Sforza y Feldman, 2003).

El patrón de variación de las poblaciones modernas depende de la historia demográfica que incluye las fluctuaciones en el tamaño poblacional efectivo, la subestructura y la migración, así como factores específicos de los genes como la mutación y las tasas de recombinación y presión de selección (Tishkoff y Verrelli, 2003). Entonces, las poblaciones naturales se ven determinadas por cuatro factores evolutivos que influirán en la distribución de las frecuencias alélicas los cuales son la selección natural, la deriva genética, la migración y las mutaciones.

La selección natural es el resultado de la variación de la población entre genotipos individuales en sus probabilidades de sobrevivir y reproducirse. Y la deriva genética se debe a un número finito de individuos participando en la formación de la nueva generación. Ambos factores pueden llevar a la eliminación o fijación de un alelo particular. En presencia de mutaciones pero ausencia de selección (es decir, bajo condiciones naturales), la tasa de evolución neutral de una población es igual al recíproco de la tasa de mutación (Cavalli-Sforza y Feldman, 2003; Kimura, 1968).

La migración puede afectar profundamente la variación genómica dentro de una población. El matrimonio entre grupos es un factor de intercambio genético particular, que reduce el aislamiento de muchas poblaciones. La inmigración de un individuo por generación en una población es suficiente para mantener activa la deriva genética y

evita así la fijación completa de los alelos. A veces una población completa o una fracción de ella migra y se establece en otro lugar. Si el grupo migratorio inicial es pequeño pero se expande progresivamente, solo por azar las frecuencias de los alelos entre los fundadores de la nueva población se diferenciarán de los de la población inicial y aún más respecto a los de la población en que se establecieron. En este caso, la migración de grupo tiene un efecto que en ciertos aspectos es opuesta a la migración individual entre poblaciones vecinas, al crear más posibilidades de deriva y por lo tanto de divergencia. El efecto sería la variación intergrupo en las frecuencias alélicas (Cavalli-Sforza y Feldman, 2003).

Toda la historia de la diferenciación de las poblaciones utilizando datos genéticos se infiere, en parte, a partir de árboles filogenéticos y de métodos estadísticos multivariados como el análisis de componentes principales (de los cuales el escalamiento multidimensional es una variante). Los árboles de poblaciones son útiles especialmente para reconstruir la historia, si las poblaciones se originaron por fisión y las diferencias entre ellas ocurrieron al azar en el tiempo y con una tasa constante de evolución. Esto parece ser verdad para los datos de ciertos genes autosómicos en poblaciones grandes que están distantes geográfica y genéticamente. La violación de estas suposiciones, como la presencia de migración o selección, afecta la interpretación de los árboles. Sin embargo, cuando la migración entre vecinos geográficos es frecuente, entonces los componentes principales mostrados en dos dimensiones reflejan la distribución geográfica de las poblaciones. Bajo el modelo evolutivo descrito, los árboles y los componentes dan resultados similares. Para poblaciones que están cercanas geográficamente, las distancias genéticas y geográficas generalmente están correlacionadas (Cavalli-Sforza y Feldman, 2003).

En el caso de los haplotipos de ADNmt y Y-STR se ha visto que son ideales para investigar la estructura geográfica de las poblaciones humanas, debido a que se comportan como marcadores neutros y su rápida tasa de evolución y tamaño pequeño efectivo de la población hace que sean indicadores más sensibles de las diferencias genéticas entre grupos, respecto a los marcadores autosómicos. Aún más, al comparar los patrones de variación del cromosoma Y y el ADNmt se permite adentrar en la

historia materna y paterna de las poblaciones, que pueden diferenciarse, especialmente en las poblaciones mezcladas (Kayser *et al.*, 2003), y su comparación es útil además para estudiar las diferencias en los patrones de migración de hombres y mujeres (Tishkoff y Verrelli, 2003).

Una comparación global del ADNmt y los polimorfismos del cromosoma Y han mostrado una tasa de migración más alta en mujeres que en hombres, asociado a la práctica patrilocal. Sin embargo, los tamaños de la muestra, las poblaciones representadas y los métodos usados para probar la variación genética difieren considerablemente entre estos dos tipos de marcadores, lo que hace problemática una comparación estricta (Pérez-Lezaun *et al.*, 1999). Sin embargo en el 2004, se realizó otro estudio basándose en ciertas secuencias del cromosoma Y (elementos Alu) y la región condificante del ADNmt, observándose que la diferenciación entre poblaciones era similar para ambos sistemas genéticos en todos los niveles geográficos que se probaron, mostrando que los patrones de estructura genética en escalas locales y globales no están determinados por la alta tasa de migración entre mujeres que es similar a la de los hombres (Wilder *et al.*, 2004). Y en todos los casos estos patrones de migración son específicos de la población que se estudie, por lo que una generalización no sería posible (Hamilton, Stoneking y Excoffier, 2005).

Para estimar los patrones de migración relacionados con el sexo, los diferentes marcadores deben ser estudiados en la misma población y aún en los mismos individuos, además el estudio debe restringirse a una región bien definida, debido a que puede haber heterogeneidad en los patrones de migración de las diferentes sociedades humanas (Pérez-Lezaun *et al.*, 1999). De hecho, muchos estudios poblacionales a nivel mundial se han realizado a partir de los marcadores en el cromosoma Y y el ADNmt, buscando evidencia de migración, mezcla, evolución, relaciones parentales, entre otros objetivos, de ahí la utilidad que estos tienen.

## B. El Cromosoma Y

El cromosoma Y es uno de los cromosomas humanos más pequeños, con 60 Mb de tamaño, aproximadamente. Se localiza en el núcleo, presenta muy pocos genes y no participa en la meiosis ni en la mitosis femenina. El brazo largo está constituido principalmente por heterocromatina, la cual constituye más de la mitad del cromosoma en los hombres normales. Esta región posee secuencias altamente repetidas, de tamaño variable en el ámbito de los microsátélites a los minisátélites. Cada *locus* posee distintas tasas de mutación lo que hace posible seleccionar polimorfismos apropiados para estudiar la evolución en diferentes escalas del tiempo (Gusmão, Brión y González-Neira, 2000; Jobling y Tyler-Smith, 1995; Scheffer, 1999). Además, el cromosoma Y está constituido por eucromatina, de tamaño constante, que posee dos regiones homólogas al cromosoma X, secuencias repetitivas específicas y todos los genes funcionales identificados (Gusmão *et al.*, 2000).

El cromosoma Y posee ventajas inherentes a su estructura y función. Debido a que no sufre recombinación, permite la construcción de haplotipos individuales íntegros, se pueden rastrear los ancestros masculinos particulares, facilita la estimación de los eventos de separación en la formación de árboles filogenéticos al usar tasas de mutación calibradas, entre otros aspectos. Esto hace menos compleja la interpretación de la historia genética humana que la deducida del análisis en marcadores autosómicos. Además, la historia paterna revelada con el cromosoma Y unida a la información materna proveniente del ADNmt revela información en aspectos relacionados, pero diferentes, de la misma historia genética (Kayser *et al.*, 2001; Quintana-Murci, Krausz y McElreavey, 2001).

La relevancia de los marcadores polimórficos del cromosoma Y para los estudios forenses, poblacionales y evolutivos, se debe a la especificidad masculina que poseen, a diferencia de los cromosomas autosómicos y el ADNmt (Gusmão *et al.*, 2000). Otro aspecto importante de los marcadores utilizados para estos estudios, es la suposición de neutralidad, por lo tanto las fuerzas de selección no han influido en la distribución alélica de la población general (Quintana-Murci *et al.*, 2001). Por otra parte, el tamaño

efectivo de la población del cromosoma Y es reducido respecto a los autosomas, por lo que los polimorfismos de este cromosoma se ven más afectados por los procesos de deriva génica y pueden ser muy útiles para marcar diferencias genéticas entre poblaciones cercanas muy relacionadas cuyo tiempo de divergencia ha sido relativamente corto (Pérez-Lezaun *et al.*, 1999).

A pesar de la escasez de polimorfismos específicos del cromosoma Y, la región no pseudoautosómica de este cromosoma contiene diversos tipos de polimorfismos (Gusmão *et al.*, 2000; Jobling y Tyler-Smith, 1995; Scheffer, 1999). A partir de estos se han desarrollado marcadores como los marcadores bialélicos, los microsátélites, los minisátélites (por ejemplo MSY1) y el sistema alfoide (González-Neira *et al.*, 2001).

Los microsátélites o STRs definen mejor la variación genética dentro de los haplogrupos y permiten la separación con más exactitud de las poblaciones así como la estimación de los tiempos de unión en haplogrupos definidos (Quintana-Murci *et al.*, 2001; Tishkoff y Verrelli, 2003). Poseen amplias ventajas con respecto a otros marcadores, pues se estudian por PCR (Reacción en cadena de la polimerasa, en inglés) que es una técnica simple, confiable y los productos de amplificación son pequeños por lo que se pueden estudiar en muestras degradadas (Jobling, Pandya y Tyler-Smith, 1997; Roewer *et al.*, 2001).

Los Y-STR son más escasos que los autosómicos y presentan repeticiones di-, tri-, tetra- y pentanucleotídicas, sin embargo debido a sus altos niveles de diversidad, comparados con los polimorfismos bialélicos, son los marcadores más utilizados en genética forense (Gusmão *et al.*, 2000). Además, estos marcadores muestran altos niveles de heterogeneidad en este cromosoma entre y dentro de poblaciones que determinan su importancia en diversas líneas de investigación (de Knijff *et al.*, 1997).

La tasa de mutación de la región no recombinante del cromosoma Y (NRY, en inglés) es comparable con la del ADN nuclear, lo que significa que es más difícil hallar mutaciones pero las genealogías son más fáciles de reconstruir. En promedio la tasa de mutación de los Y-STR se estimó en  $2,8 \times 10^{-3}$  (Gill *et al.*, 2001), con un rango de  $10^{-5}$  a  $10^{-2}$  por generación (Tishkoff y Verrelli, 2003) que es similar a lo observado en los STR

autosómicos. En un contexto evolutivo, esta tasa de mutación es muy alta para ser utilizada por lo que es recomendable recurrir al estudio conjunto de polimorfismos del cromosoma Y más estables, como las sustituciones de bases junto a los Y-STR (de Knijff *et al.*, 1997; Roewer *et al.*, 1996). Aunque el NRY se comporta como un único *locus*, lo cual es insuficiente para análisis evolutivos, ha generado resultados que son consistentes a través de muchos estudios y concordantes con hallazgos arqueológicos (Cavalli-Sforza y Feldman, 2003).

Los datos poblacionales obtenidos a partir del estudio de STRs son importantes para la identificación forense o los casos de investigación de paternidad, también lo son para los análisis de la estructura de migración, asentamiento y apareamiento de la población humana. Los estudios poblacionales basados en haplotipos que incluyen el máximo de polimorfismos posibles parecen ser la mejor estrategia en investigaciones genéticas sobre el origen del hombre y su evolución (de Knijff *et al.*, 1997; Gusmão *et al.*, 2000; Jobling y Tyler-Smith, 1995; Kayser *et al.*, 2001; Previderè *et al.*, 2000).

Con estos haplotipos, es posible analizar las diferencias en la estructura de la población al comparar la diversidad haplotípica y el número de haplotipos específicos de la población. El uso de estos marcadores permite el análisis simultáneo de poblaciones cercana y lejanamente relacionadas, por medio del análisis de haplotipos compartidos, análisis moleculares de varianza (AMOVA) y árboles filogenéticos (Gusmão *et al.*, 2000; Kayser *et al.*, 2001; Kayser *et al.*, 1997).

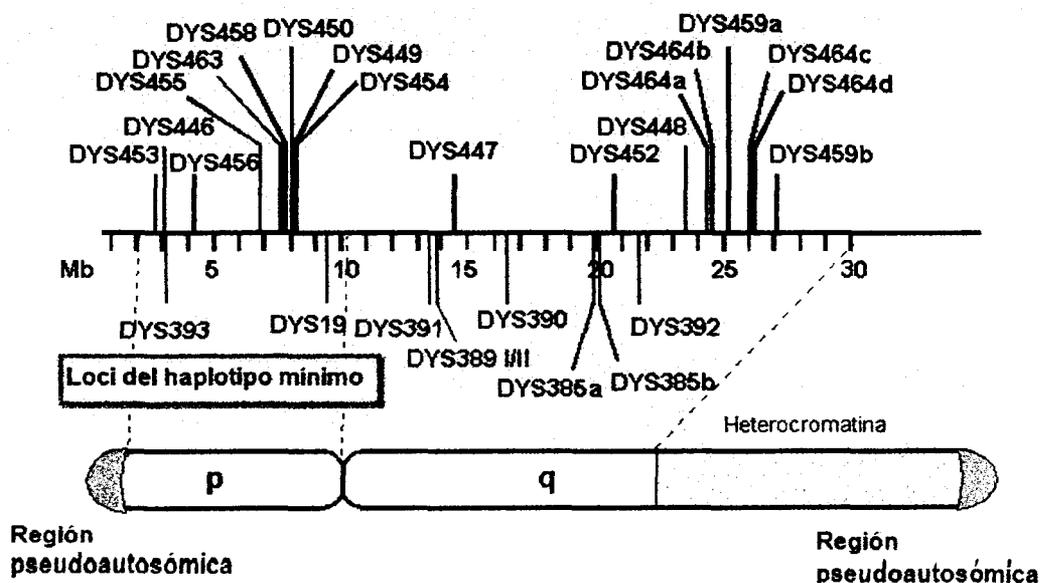
La importancia de los estudios evolutivos se ha visto evidenciada en el aumento de las investigaciones con diversos polimorfismos, especialmente para comprender el rol de los hombres en la migración y la mezcla de grupos (Gusmão *et al.*, 2000; Roewer *et al.*, 1996).

La base de los estudios evolutivos y poblacionales en el cromosoma Y es que su distribución no es al azar en ciertas poblaciones, es decir, las variantes del cromosoma Y tienden a ser más localizadas geográficamente que aquellas del ADNmt o autosómico, lo que brinda claves sobre el origen de la población en estudio, aunque las mezclas y la subestructuración de la población representen un problema, como en los procesos de

migración específica de hombres, los efectos de fundador y la deriva genética al azar que pueden llevar a la sobrerrepresentación de haplotipos específicos en una población (Carracedo *et al.*, 2001; Gusmão *et al.*, 2000; Jobling y Tyler-Smith, 1995; Quintana-Murci *et al.*, 2001). Para esto es importante distinguir entre patrones de distribución alélica debido a la historia de la dispersión demográfica y a la selección natural (Quintana-Murci *et al.*, 2001).

Los STRs que tienen más datos acumulados por ser los más usados en genética forense y poblacional son: DYS19, DYS385, DYS389I/II, DYS390, DYS391, DYS392 y DYS393 (ver figura 1). Estos son compatibles con los patrones del ISFG (“International Society of Forensic Genetics”) para análisis forenses con Y-STR (Roewer *et al.*, 2001). Son los mejor caracterizados por el desempeño de la amplificación y la especificidad, para estrategias de amplificación por múltiplex, en la estructura de su secuencia y nomenclatura, así como disponibilidad de las distribuciones de las frecuencias alélicas mundiales (Beleza *et al.*, 2003; Gusmão y Carracedo, 2003a).

**Figura 1.** Estructura del cromosoma Y y localización de ciertos microsatélites analizados en las poblaciones humanas



Este set central de marcadores ha sido utilizado por muchos laboratorios que realizan diagnóstico forense y pruebas de parentesco así como investigación antropológica y arqueológica. Por lo tanto, ha sido elegido como el fundamento para las bases de datos de Y-STR. Los perfiles obtenidos al estudiar estos STR se les llama haplotipos mínimos, para enfatizar que representan la información mínima para un haplotipo informativo en casos forenses (la diversidad haplotípica general para el set central es 0.9972) (Roewer *et al.*, 2001).

El haplotipo mínimo definido en la Base de Datos de Referencia de los Haplotipos del cromosoma Y (YHRD, en inglés) se ha utilizado para diversos estudios poblacionales en Italia (Presciuttini *et al.*, 2001), Taiwan (Wu y Pu, 2001), Filipinas (Miranda *et al.*, 2001), Alemania y China (Hidding y Schmitt, 2000), Estados Bálticos (Lessig, Edelman y Krawczak, 2001), norte (Gené *et al.*, 1999) y sureste de España (Gamero *et al.*, 2002), entre otras regiones y países.

Sin embargo, al extender el haplotipo mínimo con 8 Y-STR adicionales se lograron grandes beneficios. Estos marcadores fueron el DYS437, DYS438, DYS439, GATA A7.1 (DYS460), GATA A7.2 (DYS461), GATA A10, GATA C4 y GATA H4. Los resultados mostraron que los 16 marcadores en conjunto tienen un mayor poder de discriminación y por lo tanto son más informativos que el set central, aportando información útil adicional a la disponible actualmente (Beleza *et al.*, 2003).

### **C. El ADN mitocondrial**

El ADN mitocondrial (ADNmt) es una molécula circular, cerrada, de cadena doble y de 16569 pares de bases (pb) de longitud. La hebra pesada (H) contiene en mayor proporción guaninas y adeninas, mientras que la hebra ligera (L) es más rica en citosinas y timinas (Tully *et al.*, 2001). La información que contiene esta molécula corresponde a los ARN ribosomales 12S y 16S, 22 ARN de transferencia y ciertas proteínas de la cadena respiratoria (Anderson *et al.*, 1981). Sin embargo, hay una secuencia no codificante de 1122 pb que además presenta de tres a cuatro veces más diversidad que la

región codificante, esta se llama la región control o D-loop (Wallace, Brown y Lott, 1999).

La región control contiene los promotores de la hebra H y L, los elementos reguladores de la transcripción, los orígenes de replicación del ADNmt y la secuencia asociada a la terminación (TAS, en inglés) (Scheffer, 1999).

Las mitocondrias se encuentran en el citoplasma y su forma de transmisión de una generación a otra es a través del citoplasma del oocito (Wallace *et al.*, 1999). Un oocito maduro de humano contiene unas 100.000 moléculas de ADNmt (Chinnery *et al.*, 2000). En las células somáticas cada mitocondria puede contener de 2-10 copias de la molécula entera del ADNmt, inclusive hasta 1000 (Budowle *et al.*, 2003). Mientras que el espermatozoide contiene unos 100 ADNmt, que pueden ingresar durante la fertilización, de modo que su información persistiría en cruces interespecíficos, pero serían eliminados selectivamente en cruces intraespecíficos (Scheffer, 1999). Esto se ha confirmado con la observación de que las mitocondrias en la mitad del espermatozoide son ubiquitinadas para la posterior destrucción en el citoplasma del oocito (Wallace *et al.*, 1999).

El valor de las variaciones en la secuencia del ADNmt se basa en varios hechos. El primero es que hay una estricta herencia materna. Además, la tasa de evolución del ADNmt es de 10 a 20 veces mayor que el ADN nuclear (Seielstad, Minch y Cavalli-Sforza, 1998), esto provee abundantes sitios polimórficos pero crea dificultades en reconstruir genealogías debido a mutaciones repetidas y reversas. Por otro lado, no hay evidencia de recombinación aunque hay una baja frecuencia de rearrreglos en el ADNmt somático en células musculares del corazón (Cavalli-Sforza y Feldman, 2003). Otra ventaja es que toda la molécula ha sido secuenciada (Anderson *et al.*, 1981), es abundante en las células y por lo tanto accesible para experimentación (Cavalli-Sforza y Feldman, 2003; Scheffer, 1999).

La secuencia del ADNmt se interpreta como un único *locus* haploide porque la molécula se hereda como un haplotipo (Tully *et al.*, 2001).

Los tipos de mutación más comunes dentro de la región control son las sustituciones de una base, siendo las transiciones las que sobrepasan a las transversiones en una proporción 40:1. Además, se presentan inserciones y deleciones pequeñas, estas son comunes en las regiones homopoliméricas (poliC entre los nucleótidos 302-310 y 16183-16194), esto se incrementa conforme aumenta la longitud del homopolímero sin interrupciones (Tully *et al.*, 2001).

La nomenclatura estándar universal para el ADNmt se deriva de la primera secuencia completa, que proviene de un individuo de ascendencia europea junto con ciertos fragmentos de secuencia de ADNmt de células HeLa y bovinas (Andrews *et al.*, 1999). Cada una de las bases en el genoma mitocondrial posee un número consecutivo empezando cerca del origen de replicación de la hebra H con el número uno y terminando en 16569, a esta se le llama la Secuencia Anderson (Anderson *et al.*, 1981) o la Secuencia de Referencia Cambridge (CRS, en inglés). Entonces una muestra analizada se puede reportar como una lista de las diferencias respecto a la Secuencia Anderson, aunque ésta no sea la más representativa de las poblaciones mundiales (Tully *et al.*, 2001). La revisión de esta secuencia permitió hallar ciertos errores y polimorfismos extraños respecto a CRS, siendo un 0.07% la frecuencia de error (Andrews *et al.*, 1999).

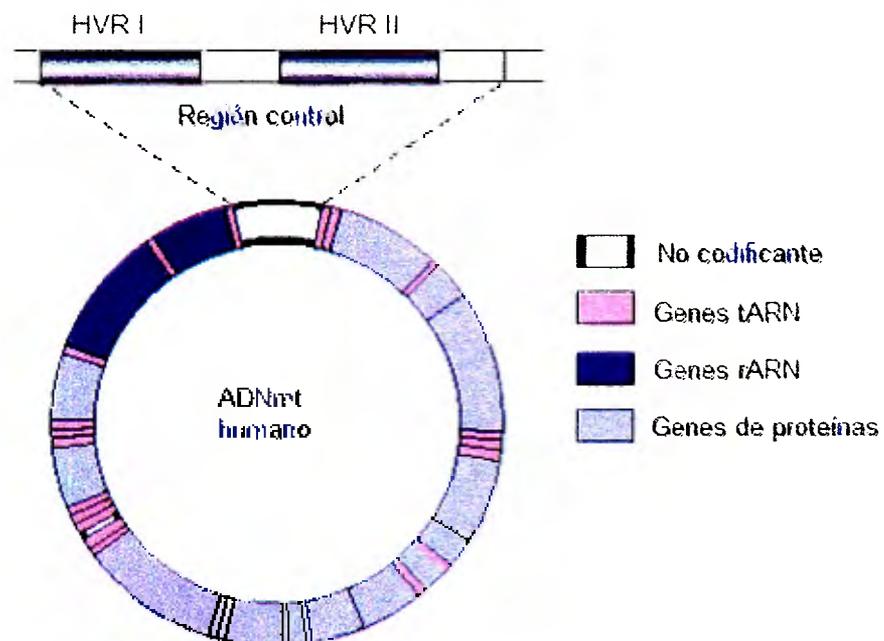
Los segmentos hipervariables del ADNmt (ver figura 2), específicamente HVI y HVII, poseen características que lo hacen valioso para estudiar poblaciones relacionadas cercanamente o que se han separado recientemente. La diversidad se concentra en estos dos segmentos, alcanzando el 92% de los polimorfismos del asa D (Mclean *et al.*, 2003). Además las tasa de sustitución en estos segmentos evoluciona más rápidamente que la región codificante (Di Rienzo y Wilson, 1991; Finnilä, Lehtonen y Majamaa, 2001).

La mayoría de los estudios de secuenciación del ADNmt se han enfocado en HVI y HVII entre las posiciones nucleotídicas 16020-16400 y 48-408, respectivamente. Mientras que HVI se ha utilizado como una fuente de información evolutiva en muchos análisis filogenéticos del ADNmt, también la base de datos de HVII se ha incrementado con datos de poblaciones humanas de todo el mundo disponibles en la red ([www.hvrbase.de](http://www.hvrbase.de)) (Bonné-Tamir *et al.*, 2003). Otra fuente de referencia para datos

poblacionales es Mitomap, que posee información de múltiples poblaciones para ambas secuencias de la región control.

La secuencia del ADNmt ha evolucionado por la acumulación secuencial de sustituciones de bases a través de los linajes maternos, estas mutaciones son vistas en alta frecuencia y corresponden a polimorfismos de secuencia específicos para un continente. Estos polimorfismos se asocian con haplotipos específicos del ADNmt y a grupos de haplotipos relacionados (haplogrupos) (Sherratt *et al.*, 1999). El rápido cambio en los linajes del ADNmt que se observa entre continentes puede haber sido influido por la selección así como por la deriva (Wallace *et al.*, 1999).

**Figura 2.** Distribución del genoma mitocondrial especificando la región control



Tomado de: Hagelberg, 2002. Traducido al español.

Sin embargo, la acumulación de cambios de bases en el ADNmt ocurre también en una escala de tiempo lo suficientemente corta como para generar polimorfismos dentro

de las poblaciones de la misma categoría étnica (por ejemplo los caucásicos); así hay polimorfismos que distinguen los diferentes grupos poblacionales humanos y además hay polimorfismos dentro de cada individuo, a lo que se llama heteroplasmia, que los distinguen. No obstante, la mayoría de los individuos son homoplásmicos (Scheffer, 1999).

Una población que ha existido y permanecido en un lugar por largo tiempo, al ser muestreada, revelará una amplia diversidad de genomas mitocondriales, asumiendo que el reloj molecular estuvo funcionando continuamente. Si una población de origen reciente fue establecida por un número pequeño de fundadores se esperaría que fuera relativamente homogénea respecto a sus secuencias mitocondriales (Scheffer, 1999). Si las diferencias entre un par de poblaciones se dan en un número pequeño de nucleótidos implica que ambas provienen de la misma población, mientras que un alto número de nucleótidos diferentes entre ambas poblaciones implica que deben tener diferentes orígenes (van Holst *et al.*, 1998).

Debido a la hipervariabilidad de la región control, muchos sujetos dentro de un mismo grupo racial tendrán diferencias respecto a la secuencia más representativa dentro de su población y el análisis de pares de diferencias mostrará la cantidad de variabilidad dentro de la población (Sherratt *et al.*, 1999).

El ADNmt es una herramienta accesible para estudios poblacionales y evolutivos intra e interespecíficos (van Holst *et al.*, 1998). En particular, el origen y evolución de los nativos americanos se evaluó inicialmente con RFLPs así como con la secuenciación del segmento hipervariable I (HVI) de la región control del ADNmt. Los resultados mostraron que estos ADNmt descendieron de cinco haplogrupos fundadores, del A al D y el X (Bandelt *et al.*, 2003).

Una ventaja adicional de estudiar el ADNmt, más que el ADN nuclear, es que en muchos lugares (por ejemplo Australia y América) donde los grupos indígenas fueron colonizados por otros pueblos, la historia social ha resultado en que los ancestros indígenas maternos se han mantenido mientras que los linajes paternos se han mezclado. En estos casos, los estudios de ADN nuclear tienen poco valor, mientras que el ADNmt

provee información acerca de las conexiones maternas que algunas veces se pierden con el rompimiento de grupos tradicionales (van Holst *et al.*, 1998).

#### ***D. Estudios genéticos realizados en las poblaciones de estudio para los marcadores del cromosoma Y y ADNmt***

En la actualidad se han realizado múltiples estudios en la población costarricense, estadounidense y mexicana utilizando marcadores para uso forense, para estimar la proporción de mezcla de sus individuos y de sus regiones geográficas, así como procesos evolutivos y migratorios.

##### ***D.1. Costa Rica***

Costa Rica es un país de Centroamérica que recibió varias oleadas de inmigrantes durante la época colonial provenientes de España y Portugal. Durante finales del siglo XVI al XIX estos inmigrantes se establecieron en el Valle Central del país. El movimiento de la población hacia fuera y adentro se estancó por los límites geográficos de las montañas, haciendo del transporte hacia las costas casi imposible. La población creció de un grupo pequeño de inmigrantes de pocos granjeros hasta aproximadamente 4 millones en el año 2001 (DeLisi *et al.*, 2001; Morera, Barrantes y Marín-Rojas, 2003).

Debido a esto, la población costarricense se considera un trihíbrido ya que el contacto no sólo se dio con los europeos sino con Africanos y Amerindios durante la época colonial. Así mismo sucedió en otros países Latinoamericanos (Morera *et al.*, 2003).

Sobre las poblaciones amerindias que habitan el país se han realizado diversos estudios genéticos, no sólo entre ellos y dentro de ellos, sino comparándolos con otros nativos de América y el resto del mundo. En 1994, Santos, Ward y Barrantes publicaron un estudio basado en una muestra de la población amerindia huetar chibcha de Costa Rica utilizando la variación genética del ADNmt en la región control y la distribución de ciertos polimorfismos en proteínas de la sangre. Este estudio reveló la presencia de 11 linajes a partir de los 24 sitios variables hallados en los 27 individuos muestreados. En

conjunto con la información de los polimorfismos de proteínas se observó que las variantes reconocidas son de antigüedad considerable, lo que implica una divergencia del acervo genético fundador para la población, además revela la reducción en el tamaño de la población Chibcha, ligado a la estructura poblacional genética de los huetares.

En otro estudio, cinco Y-STR (DYS391, DYS388, DYS395, DYS394 y DYS390) se utilizaron para evaluar la dispersión de las poblaciones mundiales. En este caso las muestras americanas provenían de indígenas costarricenses (guaymies y bribries), chilenos (pehuenche) y canadienses (dogrib), además de africanos, caucásicos de Alemania y de un cohorte del CEPH (Centro de Estudio de Polimorfismos Humanos), mongoloides y poblaciones de las islas del Pacífico (Deka *et al.*, 1996). En este caso se creó una red de haplotipos, no se observó en ella el agrupamiento de haplotipos de la misma población o de poblaciones vecinas geográficamente, sin embargo sí surgieron agrupamientos relacionados con etnias específicas como las de los nativos americanos, africanos y la muestra de Samoa. Cuando se realiza un árbol “neighbor joining” (agrupamiento de vecinos, en inglés) a partir de distancias genéticas se observan ramas específicas para los nativos americanos, africanos, asiáticos y caucásicos. Parte de las dificultades del estudio fue la resolución de ciertos agrupamientos que podría mejorarse al utilizar más *loci*, además las migraciones pasadas disminuyen la evidencia de distinción entre los agrupamientos, esto aunado a las tasas de mutación que hacen de las estructuras haplotípicas muy azarosas (Deka *et al.*, 1996).

Los cabécar, bribri, boruca y guaymí fueron parte de un estudio, entre otras poblaciones nativas de América, en que se evaluó la migración desde Siberia en la conformación de los pueblos nativos de América (Lell *et al.*, 2002).

A partir de nueve marcadores del cromosoma Y (YAP, sistema *ch*, DYS199, DYS19, DYS389 a y b, DYS390, DYS391 y DYS393) se estudió la variación genética de cinco tribus chibchas de Costa Rica y Panamá (huetar, guaymí, cabécar, bribri y teribe). Se obtuvieron 39 haplotipos en los 93 individuos estudiados para una diversidad haplotípica de 0,937. Al comparar los datos con el ADNmt, se observó que la diversidad en el cromosoma Y fue mayor, posiblemente por la alta tasa de mutación de los microsatélites. Sin embargo, los valores del  $F_{st}$  sugieren que no hay diferencias en la

estructura genética poblacional de hombres y mujeres en estas tribus (Ruiz-Narvaez *et al.*, 2005).

El otro componente de la población costarricense es de origen africano, su ingreso al país fue en condición de esclavos durante la época colonial. Ellos provenían del oeste de África, específicamente de Senegal y Gambia, además de la Costa de Oro, Congo, Angola y otros países (Morera *et al.*, 2003). También hay una pequeña contribución de afrocaribeños que ingresaron al país principalmente de Jamaica en 1872 hasta las primeras décadas del siglo XX. En esta población se han realizado estudios con marcadores del cromosoma Y. Por ejemplo, en el año 2000 se utilizaron los microsatélites DYS19, DYS389 a y b, DYS390, DYS391 y DYS393, además del *locus* bialélico YAP para analizar una muestra afrocaribeña. Los resultados revelaron una diversidad haplotípica de 96,1% en 34 individuos analizados. Al comparar esta población con una muestra del noroeste de África se observa que toda la varianza se concentra dentro de las poblaciones y que los marcadores DYS389 a y b, YAP y DYS393 son relevantes en la diferenciación interpoblacional (Jiménez, 2000).

Una publicación más reciente fue revelada por Gusmão y colaboradores (2003b) respecto a Y-STR en una muestra africana en Costa Rica, su información será comentada adelante.

Los mestizos costarricenses no han sido muy estudiados genéticamente y su expansión progresiva en el territorio de Costa Rica provoca la pérdida gradual de información útil respecto a su variabilidad genética. Esto provoca que la aplicación del conocimiento genético en problemas epidemiológicos y de salud sea restringida, y reduce la disponibilidad de evidencias acerca de su origen y de la contribución de poblaciones ancestrales en su conformación (Morera *et al.*, 2003).

En el estudio llevado a cabo por Carvajal-Carmona *et al.* (2003) se observó que Antioquía (Colombia) y el Valle Central de Costa Rica (VCCR) fueron establecidas por poblaciones parentales relacionadas cercanamente, donde los ancestros femeninos de ambas poblaciones pueden ser trazados hasta poblaciones nativas definidas. El trasfondo amerindio es predominante en el VCCR (83%), Antioquía (90%) y el norte

central de México (90%). El haplogrupo A del ADNmt constituye el 90% de los linajes nativos americanos en Antioquia y el VCCR, mientras que en el norte central de México y Brasil se hallan más distribuidos. El estimado de mezcla mostró que los ancestros masculinos del VCCR son 0,86 europeos, 0,07 africanos occidentales y 0,01 amerindios (Carvajal-Carmona *et al.*, 2003).

Específicamente con el ADNmt, Meléndez (2001) estudió la región HVI y HVII en noventa individuos costarricenses (hombres y mujeres no emparentados por la línea materna) para evaluar la capacidad discriminante con fines de identificación forense y poder de discriminación. De este estudio se obtuvieron 80 haplotipos definidos por 103 sitios polimórficos. Las sustituciones representaron un 72% de las mutaciones presentes, de las cuales 70% fueron transiciones y 2% transversiones. La diversidad genética fue de 0,983 para ambas regiones en conjunto. Esto llevó a la posibilidad de utilizar el ADNmt como herramienta para la identificación forense.

En el 2003, como parte de un estudio colaborativo del Grupo Español y Portugués de la Sociedad Internacional de Genética Forense (GEP-ISFG), se publicaron los resultados del estudio de trece poblaciones con dos tetraplexes de Y-STR, formado en total por ocho *loci* (DYS461, GATA C4, DYS437, DYS438, DYS460, GATA A10, GATA H4 y DYS439). Entre las poblaciones se encuentra una muestra de 102 individuos mestizos y 18 negros costarricenses, además de muestras de Portugal, Argentina, Macao, Brasil, Mozambique y España. Entre los resultados más sobresalientes se encuentra un árbol “neighbor joining” basado en distancias  $\Phi_{st}$  que agrupa en una rama a todas las poblaciones caucásicas de origen Ibérico, como ha sido reportado previamente, en otra rama la muestra China (Macao), cerca de ella la población Lara (amerindios argentinos) y en un extremo la muestra de Mozambique y los negros costarricenses. Cuando se analizan los haplotipos generados se observa que el haplotipo más frecuente nunca es el mismo en las muestras descendientes de Iberia, sin embargo difieren del haplotipo más compartido en una repetición de un microsatélite, lo que podría representar un mismo linaje ancestral (Gusmão *et al.*, 2003b).

## *D.2. México*

Las personas de México son descendientes de ancestros nativos americanos y europeos. Es el resultado de una mezcla que comenzó hace 400 años con el establecimiento de las misiones españolas y continuó durante los siglos XVI al XVIII con el ingreso de esclavos negros en el territorio, en un aproximado de 200 000 a 500 000 individuos. El tamaño de la población, aproximadamente 33 millones, resalta la importancia de aplicar métodos de mapeo de mezcla en esta población. Además muchas enfermedades presentan diferencias de prevalencia en los europeos y amerindios haciendo apropiado un estudio de mezclas en la población mexicana. Aunque es difícil determinar cuales de estas diferencias se deben a la genética o al ambiente, los estudios de migración o los estudios de mezcla relacionados a los riesgos de sufrir una enfermedad sustenta una contribución genética significativa (Collins-Schramm *et al.*, 2004; Green, Derr y Knight, 2000).

En el estudio de Green *et al.* (2000) basándose en el ADNmt de una muestra del norte central de México, se observó la presencia de los cuatro haplogrupos de los nativos americanos en el ADNmt. Las frecuencias fueron: 33,6% haplogrupo A, 26,5% haplogrupo B, 23,3% haplogrupo C y 5,8% haplogrupo D. Estos resultados son consistentes con los estudios en ADNmt de los nativos americanos en América Central. Además, los estimados de mezcla europea son bajos en este estudio debido a que históricamente más hombres de origen español que mujeres colonizaron México.

En otro análisis de mezcla realizado en la población mexicana se observó que individualmente la proporción de los componentes de la mezcla variaron mucho, sin embargo la mayoría de los individuos poseen un 40-80% de ancestría europea (Collins-Schramm *et al.*, 2004).

Por otro lado, se ha evaluado el valor de diversos Y-STR para la población mexicana y su utilidad en la ciencia forense. Estudios realizados por Rangel-Villalobos *et al.* respecto a los marcadores GATA A4, GATA A7.1, GATA A7.2, GATA A10 y GATA A4 (2001a); y DYS19, DYS385, DYS389I, DYS390, DYS391 y DYS393 (2001b) en la

población del noroeste de México revelaron su utilidad en pruebas de análisis forense y para identificar linajes masculinos.

Los amerindios nativos de ciertas zonas de México han sido estudiados en caracterizaciones de todo el continente, como en el estudio de Bianchi *et al.* (1998), revelando semejanzas amplias con el resto de América para definir haplotipos fundadores.

También el componente negro producto de la diáspora Africana fue estudiado tomando en cuenta una muestra de mexicanos, beliceños, dominicanos, colombianos y panameños negros, representando la variabilidad del Caribe y Centroamérica. Junto a ellos se analizaron muestras de estadounidenses y Brasileños, además de una muestra completa de todas las regiones africanas involucradas en el proceso. En este caso se estudió la región hipervariable I del ADNmt, algunos datos de la región II y ciertos RFLP. Se halló y confirmó lo relatado por los historiadores, que las mayores contribuciones de negros en América fueron realizadas del oeste y centro-oeste de África, utilizando estimaciones de mezcla, representaciones de componentes principales y redes filogenéticas (Salas *et al.*, 2004).

Respecto al cromosoma Y, se publicó un estudio en el año 2002 que evaluó el origen de los nativos americanos a partir de Siberia. La muestra de estudio estuvo integrada por catorce poblaciones de Siberia, cuatro de Centroamérica, siete de Suramérica y tres de Norteamérica (Seminoles y Navajo de EUA, y sureños mexicanos nativos). Se utilizaron marcadores bialélicos y microsátélites (DYS19, DYS388, DYS390 y DYS391) que revelaron que el haplogrupo predominante para los nativos americanos fue el M3 y el segundo fue el M45 lo cual es demostrado en parte en el árbol de agrupamiento de vecinos. Se observó que la diversidad en Siberia es mayor que en América, lo que coincide con un efecto fundador durante el poblamiento de América. Además, los haplotipos obtenidos sugieren una sola migración que dio lugar a todas las poblaciones lingüísticamente diversas de los nativos americanos (Lell *et al.*, 2002.)

### ***D.3. Estados Unidos de América***

Este país posee una diversidad genética extraordinaria, con africanos traídos del oeste del continente (de Senegal hasta Angola), europeos, asiáticos y nativos americanos que contribuyeron al acervo genético actual de la población. La existencia de diferencias genéticas significativas entre los subgrupos geográficos de las poblaciones estadounidenses tendría implicaciones importantes para la comunidad forense y en el estudio de enfermedades genéticas (Kayser *et al.*, 2003).

En un estudio realizado en una población del Valle de San Luis en Nuevo México, EUA, se confirmó al estudiar el ADNmt que existía direccionalidad en la formación de parejas donde las mujeres nativas americanas se unieron a los hombres hispanos. Debido a que la variación del ADNmt de los amerindios es limitada y bien definida, la frecuencia de los haplotipos amerindios en una población refleja directamente la frecuencia de los linajes maternos amerindios que contribuyen a dicha población, este hecho representa una situación bien conocida en América producto de la colonización europea (Merriwether *et al.*, 1997).

En el año 2001, se publicó un estudio de las poblaciones norteamericanas respecto a la región control del ADNmt por medio de oligonucleótidos específicos de secuencia (SSO, en inglés). En este caso se analizaron los tres grupos predominantes en la conformación de la población que son los africanos, europeos e hispanos americanos con el objetivo de determinar el poder del ADNmt en las pruebas forenses. El principal resultado fue que no había heterogeneidad significativa entre los tres grupos a nivel de subpoblación y región (Melton *et al.*, 2001).

Por otro lado, se estudió el cromosoma Y y los resultados del AMOVA realizado en los grupos estudiados por Kayser y colaboradores (2003) indican que los haplotipos difieren significativamente entre los americanos africanos, los americanos europeos y los hispanos. El 25% de la varianza genética refleja diferencias entre estas poblaciones, mientras que un 1% refleja diferencias entre los grupos regionales dentro de cada población y 74% refleja la varianza genética dentro de los grupos regionales.

La varianza genética total compartida fue de 98-99% respecto al estudio de ADNmt entre los africanos, europeos e hispanos de EUA, y sólo 0,5-0,8% se debe a diferencias entre grupos regionales dentro de una población (Kayser *et al.*, 2003).

Las estructuras maternas y paternas difieren en el grupo hispano, reflejando una mayor contribución de los cromosomas Y de los americanos europeos que de los haplotipos del ADNmt al acervo genético hispano. Otros estudios han encontrado una mayor contribución de los ADNmt nativos americanos que los genes nucleares a las poblaciones hispanas, lo que apoya los resultados que indican una mayor contribución de los hombres americanos europeos que las mujeres al acervo genético hispano (Kayser *et al.*, 2003).

La investigación indica que no hay estructuración geográfica de los haplotipos basados en Y-STR entre los grupos regionales africanos, europeos e hispanos americanos. Para los africanos e hispanos se halló evidencia de una mayor contribución genética europea de los hombres que de las mujeres (Kayser *et al.*, 2003).

Los Navajo y Seminole de EUA fueron parte de una muestra de estudio para evaluar las migraciones desde Siberia de los primeros pobladores de América. Se analizaron SNP y STR del cromosoma Y entre diversas poblaciones nativas de América y Siberia obteniendo como resultado la presencia de haplogrupos predominantes en América y la sugerencia de una única migración desde Siberia que dio origen a todas las lenguas nativas del continente americano (Lell *et al.*, 2002).

Por otro lado, se estudiaron siete *loci* específicos del cromosoma Y (DYS199, sistema alfoide *ah*, DYS19, DYS389 I y II, DYS390, DYS391, DYS392 y DYS393) en diferentes grupos lingüísticos de América, entre ellos amerindios nativos y Na-Dene de Estados Unidos, además de muestras de México, Colombia, Rusia, Venezuela, Brasil, Paraguay, Chile y Argentina. En general, los linajes Amerindios mostraron haplotipos que derivan de un mismo ancestro, que no está relacionado filogenéticamente con la muestra del CEPH en la comparación. También se logró identificar el haplotipo fundador del linaje DYS199T y siete haplogrupos derivados de este con diferencias de uno a siete pasos mutacionales (Bianchi *et al.*, 1998).

Como se mencionó anteriormente, en la información de la población mexicana, también los negros estadounidenses, entre otras poblaciones, han sido estudiados respecto al ADNmt al evaluar la contribución de las diferentes regiones africanas en la diáspora en América, confirmando lo revelado por los historiadores (Salas *et al.*, 2004).

En un estudio para analizar la afinidad de los africanos americanos con otras regiones del mundo se buscaron nuevos polimorfismos por restricción en la región HVI. La muestra de África provino de Sierra Leona, la americana del Sur de Carolina (Gullah y Charleston) y California (Oakland), otra muestra de negros jamaquinos y de americanos de descendencia europea. Se encontraron tres nuevos sitios de restricción que exhiben un genotipo predominante en los americanos europeos y son polimórficos en las muestras derivadas de África. Los haplotipos obtenidos por estos sitios de restricción poseen un alto contenido de información para estudiar poblaciones mezcladas de americanos europeos y poblaciones parentales africanas siendo útiles para estudiar linajes maternos (Mclean *et al.*, 2003).

Otra investigación desarrollada en la población estadounidense es la elaboración de una base de datos de referencia para los haplotipos del cromosoma Y (YHRD, en inglés) con el uso de nueve Y-STR a partir del estudio de 1705 individuos africanos, europeos e hispanos, todos estadounidenses. Se realizó para poner a disposición de la comunidad forense las frecuencias de los alelos, protocolos y análisis genéticos de la población. Este estudio reveló que no hay subestructuración de las diferentes regiones del país respecto a los tres tipos de poblaciones, sin embargo es apropiado utilizar muestras más grandes para complementar y confirmar la información presente en internet (Kayser *et al.*, 2002).

### **E. La esquizofrenia**

La esquizofrenia es un reto terapéutico de la medicina moderna y una de las últimas fronteras en la investigación del cerebro. Ocurre en un 1% de la población mundial y aunque los medicamentos antipsicóticos han mejorado las vidas de los pacientes, las causas de la enfermedad aún son desconocidas (Abecasis *et al.*, 2004; Seeman y Kapur, 2000).

La esquizofrenia es una enfermedad mental crónica caracterizada por perturbaciones en el pensamiento, percepciones, cognición y voluntad (Abi-Dargham *et al.*, 2000; Deakin y Simpson, 1997; Seeman y Kapur, 2000).

En diversos estudios genéticos y epidemiológicos se ha revelado que para el desarrollo del desorden existe un componente genético grande y que factores ambientales, aún no identificados, también son requeridos (Levinson *et al.*, 2003; Nestler, 1997; Seeman y Kapur, 2000).

En cuanto al componente genético, una evidencia sustancial corresponde a los estudios realizados en gemelos. El diagnóstico concordante es dos veces más alto en gemelos monozigóticos (41%-65% de concordancia) que en gemelos dizigóticos (0%-28% de concordancia) y otros familiares de primer grado de pacientes esquizofrénicos. Por otro lado, se ha definido que la heredabilidad es de 0,70 - 0,85 y que el incremento en el riesgo de hermanos de los probandos es de diez veces, aproximadamente (Lewis *et al.*, 2003). Parece clara la influencia de *loci* de genes múltiples y factores ambientales en la susceptibilidad de sufrir la enfermedad; sin embargo, aún es limitado entender el rol de genes individuales (Abecasis *et al.*, 2004; Levinson *et al.*, 2003).

En la publicación de Lewis *et al.* (2003) se hace un resumen de todos los estudios de ligamiento realizados hasta su publicación, que corresponden a veinte tamizajes genómicos, aproximadamente (Levinson *et al.*, 2003). Se observa, en un amplio número de proyectos de tamizaje del genoma que se han realizado, la presencia de tres regiones cromosómicas que evidencian posibilidad de ligamiento con la enfermedad. En un estudio de genealogías irlandesas se reportó un puntaje LOD de 3,51 en el cromosoma 6p24-22, apoyado por cinco estudios más. Recientemente se presentó evidencia apuntando a la disbindina (DTNBP, 6p22.3) como un gen candidato para la esquizofrenia en esta región. Otra región es la 13q32 en estudios con genealogías ancestrales de europeos y canadienses. Además en la misma muestra canadiense, apoyado por resultados en genealogías de ingleses e islándicos, se asoció a la región 1q21-22. Por otro lado, los cromosomas 8p21-22 y 6q21-22.3 se han encontrado resultados insinuantes de ligamiento, inclusive uno de los grupos presenta evidencia del gen NRG1 (8p21-p12, neuroregulina 1) como un candidato para la esquizofrenia. Todo

esto refleja que los *loci* de susceptibilidad pueden estar presentes en múltiples cromosomas (Abecasis *et al.*, 2004; De Lisi *et al.*, 2001).

Sin embargo, ninguna de estas regiones reportadas han producido evidencia fuerte de ligamiento. Las explicaciones posibles para estas inconsistencias pueden ser: errores en los genotipos y variaciones en el diagnóstico, diferencias en estrategias estadísticas, o diferencias verdaderas en los efectos genéticos de estos y otros *loci* a través de grupos étnicos o tipos de genealogías. Debido al gran número de proyectos y análisis que se han llevado a cabo, se pueden reportar varias veces los mismos errores. Otra posibilidad es que cada uno de estos *loci* tiene un pequeño efecto de susceptibilidad en la población, siendo difícil detectar consistencias sin utilizar muestras grandes (Abecasis *et al.*, 2004; Lewis *et al.*, 2003).

Recientemente, los estudios genéticos de esquizofrenia han utilizado poblaciones aisladas genéticamente como Finlandia, Suecia, Quebec, Costa Rica y dos islas Micronesas: Palau y Kosrae. Es posible que en estas poblaciones hayan menos *loci* y alelos de susceptibilidad, además, las poblaciones fundadoras exhiben menos heterogeneidad ambiental en comparación con otras poblaciones y permiten una investigación genealógica detallada y la reconstrucción de genealogías multigeneracionales extendidas. Un ejemplo de ello es el estudio en la población Africana de Suráfrica, donde se logró trazar los ancestros de 98 probandos y se halló que 87 de ellos descendían de una pareja que inmigró a Suráfrica hace 12,5 generaciones. Esto sugiere que los individuos afectados en esta población podrían compartir un número reducido de alelos enfermos, como se observó en sus resultados, en los que los cromosomas 1, 9 y 13 presentaron evidencia de ligamiento (Abecasis *et al.*, 2004).

Si bien las poblaciones del presente estudio (CR, MX y EUA) han sido estudiadas mayormente respecto a la clínica de la enfermedad (Escamilla *et al.*, 2002; González *et al.*, 2004; Medina *et al.*, 2002; Medina *et al.*, 2003; Salazar *et al.*, 2004; Urraca *et al.*, 2004), un análisis reciente se realizó a nivel de genoma en busca de *loci* de susceptibilidad para la esquizofrenia obteniéndose indicios de ligamiento en los cromosomas 1, 5, 10 y 18 (Escamilla *et al.*, en preparación). Sin embargo, hasta el momento no se han realizado estudios poblacionales genéticos de estas poblaciones para

evaluar su diversidad y estructura genética, obviando información útil confirmatoria de migraciones y relaciones entre las poblaciones hispanas de Costa Rica, EUA y México relacionadas con la esquizofrenia, que pueden aportar indicios de su historia y comportamiento. Esto es importante porque una parte considerable de la medicina genética moderna recae sobre el entendimiento de la historia demográfica humana, por lo que hay una gran demanda de investigación de alta calidad en genética poblacional humana (Hey y Machado, 2003). Además, la confirmación de la relación entre estas poblaciones sería útil para los futuros análisis genéticos que serían más completos y certeros pues se podrían estudiar en conjunto las tres poblaciones.

## OBJETIVOS

### *Objetivos generales:*

- a. Estudiar la variación genética de tres poblaciones hispanas (Costa Rica, México y Estados Unidos) relacionadas con la esquizofrenia, respecto a 12 microsatélites del cromosoma Y y la secuencia hipervariable I (HVI) del ADN mitocondrial.
- b. Analizar el comportamiento demográfico y genético de las poblaciones hispanas de Costa Rica, Estados Unidos y México relacionadas con la esquizofrenia utilizando la variación molecular de 12 Y-STRs y la región HVI del ADNmt.

### *Objetivos específicos*

- a. Analizar y estimar la variación genética por medio de 12 microsatélites del cromosoma Y, que corresponde al haplotipo mínimo extendido, en la muestra costarricense, mexicana y estadounidense.
- b. Analizar y estimar la diversidad interna de las secuencias del ADNmt en la región hipervariable I para las tres poblaciones del presente estudio.
- c. Evaluar la estructura genética, y por lo tanto la diferenciación jerárquica, de las poblaciones aquí analizadas mediante el estudio de la variación dentro y entre las distintas poblaciones.
- d. Investigar las posibles relaciones filogenéticas, a partir de diferencias y semejanzas entre las poblaciones de estudio, así como su representación gráfica a partir del análisis de los microsatélites del cromosoma Y y la secuencia hipervariable I del ADNmt.
- e. Estudiar la demografía histórica de las poblaciones hispanas de Costa Rica, Estados Unidos y México relacionadas con la esquizofrenia usando marcadores moleculares del cromosoma Y y el ADN mitocondrial, y compararlo con datos de las poblaciones naturales correspondientes.

## MATERIALES Y MÉTODOS

### A. Muestras de ADN

El estudio poblacional se llevó a cabo en tres poblaciones de sexo masculino, provenientes de Costa Rica (120 individuos), México (55 individuos en total, de Monterrey 30 y de Ciudad de México 25) y el Suroeste hispano de Estados Unidos (42 individuos en total, de San Antonio 23, San Diego 6 y Los Ángeles 13), que forman parte de dos estudios de esquizofrenia (“Genética de la Esquizofrenia en Poblaciones Latinas” V.I. 800-A0-519 y “Genes de Susceptibilidad a la Esquizofrenia en el Valle Central de Costa Rica” V.I. 801-98-500). Las muestras están compuestas por probandos o familiares no enfermos, ninguno de la misma familia. La distribución geográfica de las regiones se observa en la figura 3. En total se estudiaron 217 individuos. El ADN fue previamente extraído, durante el reclutamiento de los probandos y familiares de los proyectos mencionados anteriormente.

### B. Microsatélites del cromosoma Y

Los doce microsatélites del cromosoma Y (Y-STR) estudiados fueron: DYS19, DYS 385I/II, DYS389I/II, DYS390, DYS391, DYS392, DYS437, DYS438 y DYS439. En la figura 4 se muestra la localización de los *loci* en el cromosoma Y y el cuadro 1 muestra ciertas características de los mismos marcadores. Se utilizó el kit PowerPlex<sup>®</sup>Y System (Promega). La reacción de PCR se realizó en un volumen de 12,5 µl con 2 ng/µl de ADN, STR Buffer 10x (500mM KCl, 100mM Tris-HCl pH 9.0 a 25°C, 15mM MgCl<sub>2</sub>, 1% Triton<sup>®</sup> X-100 y 2mM de cada dNTP, Promega), Power Plex Y 10x Primer Pair Mix, 0,275 µl Taq DNA polymerase 5U/µl (Promega) y agua. El control positivo que se utilizó es el 9948 Male DNA (Promega). El protocolo de amplificación utilizado fue el siguiente: 95°C por 11 min; luego 96°C por 1 min, posteriormente 10 ciclos de 94°C por 30 seg, 60°C por 30 seg y 70°C por 45 seg; 18 a 22 ciclos de 90°C por 30 seg, 58°C por 30 seg y 70°C por 45 seg; por último 60°C por 30 min y 4°C para concluir. Se hizo uso del termociclador Applied Biosystems GeneAmp<sup>®</sup> PCR System 9700.

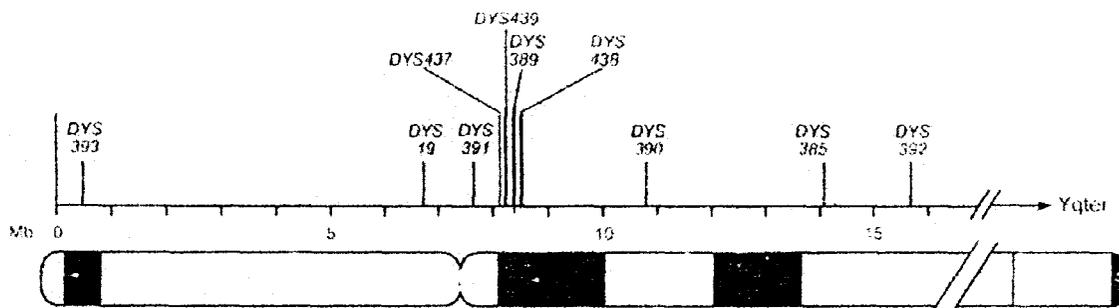
**Figura 3.** Distribución geográfica de las poblaciones analizadas en el presente estudio



Nota: encerrado en cuadros rojos se especifican las posiciones geográficas de las poblaciones

Tomado de: ENCARTA 2004. Modificado.

**Figura 4.** Posición de los 12 microsátelites del cromosoma Y analizados en el presente estudio



Tomado de: Bosch *et al.* 2002. Modificado.

**Cuadro 1.** Secuencia repetitiva, número de alelos y número de acceso en GenBank de los 12 Y-STR analizados en las tres poblaciones de CR, MX y EUA

<i>Locus</i>	Secuencia repetitiva	No de alelos *	No de acceso en GenBank®
<b>DYS 385I/II</b>	GAAA	7-24, 28	Z93950
<b>DYS 389I</b>	(TCTG)(TCTA)	9-17	AF140635
<b>DYS389II</b>	(TCTG)(TCTA)	24-34	AF140635
<b>DYS 390</b>	(TCTG)(TCTA)	17-28	AC011289
<b>DYS 391</b>	TCTA	6-14	G09613
<b>DYS 392</b>	TAT	6-17	G09867
<b>DYS 393</b>	AGAT	9-17	G09601
<b>DYS 437</b>	(TCTA)(TCTG)	13-17	AC002992
<b>DYS 438</b>	TTTTC	6-14	AC002531
<b>DYS 439</b>	GATA	9-14	AC002992
<b>DYS 19</b>	TAGA	10-19	X77751

Tomado de: Powerplex Y System, Promega

\* Información de Short Tandem Repeat DNA Internet Data Base (<http://www.cstl.nist.gov/biotech/strbase/>)

Posteriormente, los productos de amplificación se separaron en un gel de poliacrilamida 5% Long Ranger/6M urea en el secuenciador ABIPrism 377, se tomó 1  $\mu$ l de la reacción junto con 1,5  $\mu$ l de solución de carga azul-dextran y 0,5  $\mu$ l del estándar interno 600 como preparación de la muestra para cargarla en el gel. El PowerPlex<sup>®</sup> Y Allelic Ladder Mix se utilizó como referencia en cada gel para la asignación de los alelos de cada muestra.

El software Genescan Analysis 3.1 (Applied Biosystems) se usó para elaborar las hojas de trabajo y el software Genotyper 2.1 (Applied Biosystems) se usó para registrar los datos. El análisis de los datos se realizó por medio del PowerTyper<sup>™</sup> Y Macro (Promega) y con ello la asignación de los alelos para su análisis estadístico.

### **C. Secuencia Hipervariable I del ADNmt**

La región hipervariable I (HV1) del ADN mitocondrial se amplificó utilizando los primers L15997 (5' CAC CAT TAG CAC CCA AAG CT) y H16401 (5' TGA TTT CAC GGA GGA TGG TG). La amplificación inicial se realizó en volúmenes de 50  $\mu$ l con los siguientes reactivos: 4  $\mu$ l ADN a 2 ng/ $\mu$ l, 2  $\mu$ l de cada primer a 5 uM, 5  $\mu$ l de STR Buffer 10x (500mM KCl, 100mM Tris-HCl pH 9.0 a 25°C, 15mM MgCl<sub>2</sub>, 1% Triton<sup>®</sup> X-100 y 2mM de cada dNTP, Promega), 3  $\mu$ l Taq DNA Polymerase 5 U/ $\mu$ l (Promega) y agua (destilada, desionizada y autoclavada). El control negativo utilizado fue agua para evaluar la calidad de los reactivos y posibles contaminaciones. El protocolo de amplificación que se utilizó en el termociclador Applied Biosystems GeneAmp<sup>®</sup> PCR System 9700 es: 95°C por 5 min; 35 ciclos de 94°C por 45 seg, 65°C por 45 seg y 72°C por 90 seg; seguido de un paso a 72°C por 7 min y al finalizar una incubación sostenida a 4°C.

Para confirmar la amplificación se tomaron 4  $\mu$ l de los productos de amplificación junto con 1  $\mu$ l de buffer de carga 2x y se separaron en un gel de agarosa al 2% con bromuro de etidio, para la posterior visualización con luz ultravioleta. Las condiciones fueron 100V por una hora. El marcador 1kb DNA Ladder (Promega) se utilizó para comparar los tamaños de los productos de amplificación.

Los 46  $\mu\text{l}$  restantes de la reacción se purificaron con el kit Wizard PCR Preps (Promega) según las indicaciones del proveedor. Se tomaron 10  $\mu\text{l}$  del producto purificado para la cuantificación por espectrofotometría en un volumen total de 500  $\mu\text{l}$ . A partir de los datos de absorbancia se obtuvo la concentración aproximada de la solución de ADN purificado.

El ADN purificado se llevó a una concentración aproximada de 20  $\text{ng}/\mu\text{l}$  para continuar con el protocolo BigDye<sup>®</sup> Terminator v3.1 Cycle Sequencing RR-100 (Applied Biosystems) para la secuenciación. La reacción utilizó 8  $\mu\text{l}$  del BigDye<sup>®</sup> Terminator, 3  $\mu\text{l}$  de ADN (20 $\text{ng}/\mu\text{l}$ ), 2  $\mu\text{l}$  de primer a 1.6  $\text{pmol}$  y 8  $\mu\text{l}$  de agua, para un total de 20  $\mu\text{l}$  de reacción. Se preparó una reacción para cada primer por separado. El control positivo utilizó 8  $\mu\text{l}$  del BigDye<sup>®</sup> Terminator, 2  $\mu\text{l}$  de pGEM<sup>®</sup>37f(+), 4  $\mu\text{l}$  M13(-21) primer y 6  $\mu\text{l}$  de agua. El protocolo de amplificación fue el siguiente: 96°C por 1 min; 25 ciclos de 96°C por 10 seg, 50°C por 5 seg y 60°C por 4 min; por último 4°C.

Posteriormente, se procedió a precipitar el ADN con el protocolo de EDTA y etanol. Cada muestra se resuspendió directamente en 5  $\mu\text{l}$  de formamida y 1  $\mu\text{l}$  de buffer de carga concentrado.

Cada muestra se separó en un gel de poliacrilamida 5% LongRanger/6M preparado para el secuenciador ABI Prism 377 de 36 pozos. El software Genescan Analysis 3.1 (Applied Biosystems) se utilizó para elaborar las hojas de trabajo y el software Sequencing Analysis 3.3 (Applied Biosystems) se utilizó para registrar los datos y elaborar los electroferogramas. Las corridas fueron de 7 horas, posterior a esto se evaluó la calidad de las secuencias con Bioedit (Hall, 1999) y se alinearon respecto a la secuencia Cambridge (Anderson *et al.*, 1981) con Clustal.

Del método de estudio se pueden comentar las dificultades causadas por el polimorfismo 16189C y el imprimador L15997. Este polimorfismo implica la extensión del homopolímero de citosinas que a su vez tiene un efecto desestabilizador en la actividad de la ADN polimerasa y con ello provoca la pérdida de la calidad de la secuencia posterior a esta región. De modo que la única forma de recuperar los extremos de la secuencia es con la secuencia complementaria. Sin embargo, el extremo 5' de la

región HVI no pudo ser recuperado en las secuencias que presentaron este polimorfismo, debido a que se pierde el extremo 3' de la hebra L imprimada por el H16401 y el imprimador L15997 es poco eficiente al inicio de la secuencia, perdiéndose la lectura de aproximadamente 50 bases. Es por esta razón que los análisis del presente estudio se realizaron con la secuencia a partir del nucleótido 16047, con un total de 319 nucleótidos, cuando la secuencia analizada comúnmente es de 340 bases.

#### **D. Análisis estadístico**

Los datos obtenidos se analizaron para determinar la diversidad genética, la estructura jerárquica y las relaciones entre las poblaciones ligadas con la esquizofrenia provenientes de Costa Rica, México y el suroeste hispano de EUA, por medio de las siguientes pruebas estadísticas.

##### **D.1. Medidas de diversidad genética**

Estas medidas van a revelar las características genéticas de las poblaciones a nivel interno para posteriormente compararlas entre ellas y revelar semejanzas o diferencias específicas en cada población. Ciertas de estas pruebas estadísticas consideran la variación molecular permitiendo la estimación de tiempos de expansión de las poblaciones, tomando en cuenta las tasas de mutación de los marcadores. En el caso del ADNmt se utilizó un valor gamma A de 0,26 debido a la heterogeneidad en la tasa de mutación, específicamente de la región control I y se utilizó una tasa de transición-transversión de 15:1.

##### *Frecuencias alélicas, nucleotídicas y haplotípicas*

Las frecuencias se estimaron por la fórmula

$$p_i = x_i / n ,$$

Donde  $x_i$  es el número de veces que se observó el alelo  $i$  en los Y-STRs, el nucleótido  $i$  en las secuencias del ADNmt o el haplotipo  $i$  en la población de  $n$  individuos.

Y la varianza de la frecuencia se estima como

$$V(p_i) = p_i(1 - p_i) / (n-1)$$

#### *Diversidad génica y haplotípica*

La diversidad génica es equivalente a la heterocigocidad esperada para datos diploides. Se define como la probabilidad de que dos genes o haplotipos, según la diversidad estimada, tomados al azar sean diferentes en la muestra (Nei, 1987). Aquí se presentan las fórmulas para la estimación de este valor junto con su varianza.

$$DG = (n/n-1) (1 - \sum p_i^2)$$

$$V(DG) = 2/n(n-1) \{2(n-2) [\sum p_i^3 - (\sum p_i^2)^2] + \sum p_i^2 - (\sum p_i^2)^2\}$$

Donde  $n$  es el número de copias del gen o el haplotipo en la muestra y  $p_i$  es la frecuencia en la muestra del gen o haplotipo  $i$ .

#### *Diversidad de nucleótidos*

Es la probabilidad de que dos nucleótidos homólogos tomados al azar sean diferentes. Es equivalente a la diversidad génica a nivel de nucleótidos (Tajima, 1983; Nei, 1987). Las fórmulas que se utiliza para estimar la diversidad de nucleótidos y su respectiva varianza son

$$\pi_n = \sum \sum p_i p_j d_{ij} / L$$

$$V(\pi_n) = [n+1/3(n-1)L] \pi_n + [2(n^2+n+3)/9n(n-1)] \pi_n^2$$

Donde  $p_i$  es la frecuencia poblacional del tipo de secuencia de ADN  $i$  y  $d_{ij}$  es la proporción de nucleótidos diferentes entre la secuencia  $i$  y  $j$ .  $L$  corresponde al número de nucleótidos de la secuencia analizada, eliminando las deleciones e inserciones. Y  $n$  es el número de secuencias analizadas.

*Promedio de diferencias entre pares de individuos ("Pairwise differences")*

Es el número promedio de diferencias entre todos los pares de haplotipos en la muestra (Tajima, 1983; Tajima, 1993). Se estima según la siguiente fórmula

$$\pi = \sum \sum p_i p_j d_{ij}$$

Donde  $d_{ij}$  es el estimado del número de mutaciones que ocurrieron desde la divergencia de los haplotipos  $i$  y  $j$ ,  $p_i$  es la frecuencia del haplotipo  $i$  y  $p_j$  es la frecuencia del haplotipo  $j$ . La varianza total asumiendo neutralidad selectiva y no recombinación entre sitios se obtiene como

$$V(\pi) = [3n(n+1)\pi + 2(n^2+n+3)\pi^2] / 11(n^2 - 7n + 6)$$

*Distribución de diferencias ("Mismatch distribution")*

Es la distribución del número de diferencias observado entre pares de haplotipos dentro de una población. Esta es multimodal en muestras tomadas de poblaciones que se encuentran en equilibrio demográfico, pero es unimodal en poblaciones que han pasado una expansión demográfica reciente (Rogers y Harpending 1992). En el caso de las

secuencias del ADNmt equivale a las diferencias entre nucleótidos en pares de individuos, mientras que para los microsatélites del cromosoma Y equivale a la sumatoria total de las diferencias en el número de repeticiones (Ruiz, 1998).

También se estimó el Índice de Desigualdad de Harpending (Harpending *et al.*, 1993; Harpending, 1994) como:

$$r = \sum (x_i - x_{i-1})^2$$

Donde  $x_i$  es la frecuencia de los pares con  $i$  diferencias. Este índice toma valores altos para distribuciones multimodales comúnmente encontradas en poblaciones estacionarias, y valores bajos para poblaciones que han sufrido una expansión reciente. Harpending *et al.* (1993) estimaron por estudios de simulación un valor límite de  $r=0,03$  para distinguir entre ambos tipos de poblaciones.

## D.2. Estructura jerárquica de las poblaciones

El análisis de la estructura de la población involucra el desglose de la variabilidad y por lo tanto la observación de la diferenciación entre individuos, grupos o poblaciones. Para el presente estudio cumple el objetivo de estudiar y comparar a las poblaciones, para evaluar su conformación y diferenciación respecto a las demás en tres niveles jerárquicos (individuo, grupo y población).

### *Análisis de Varianza Molecular (AMOVA)*

Este método investiga la estructura genética de las poblaciones y es muy similar a los acercamientos basados en el análisis de varianza de frecuencias génicas, pero toma en cuenta el número de mutaciones entre haplotipos moleculares. Un análisis de varianzas jerárquico parte la varianza total en componentes de covarianza debido a diferencias entre poblaciones, además de la variabilidad dentro y entre individuos de las

poblaciones. Los componentes de covarianza se utilizan para estimar los índices de fijación.

La significancia de los índices de fijación se prueba utilizando un acercamiento de permutaciones no paramétricas (Excoffier, Smouse y Quattro, 1992), que consiste en permutar haplotipos, individuos o poblaciones, entre individuos, poblaciones o grupos de poblaciones. Luego de cada ronda de permutación, se vuelven a estimar todos los estadísticos para obtener la distribución nula. Se llevan a cabo diferentes tipos de permutaciones según el tipo de estadístico probado y el diseño jerárquico tomado.

En el caso de datos haploides en los que solo se presenta un grupo de poblaciones, las varianzas se estiman de la siguiente manera:

Fuente de variación	Grados de libertad	Diversidad de suma de cuadrados	Promedios cuadrados esperados
Entre poblaciones	$P - 1$	SC (EP)	$n \sigma_a^2 + \sigma_b^2$
Dentro de poblaciones	$N - P$	SC (DP)	$\sigma_b^2$
Total	$N - 1$	SC (T)	$\sigma_T^2$

Donde  $n$  y  $F_{st}$  se definen como,

$$n = N - (\sum N_p^2 / N) / P - 1$$

$$F_{st} = \sigma_a^2 / \sigma_T^2$$

En este y el siguiente caso  $N$  significa el número total de individuos y  $\sigma$  es el valor de la varianza correspondiente al nivel evaluado.

Para los datos haploides con varios grupos de poblaciones la estimación de las varianzas se realizan de la siguiente manera:

Fuente de variación	Grados de libertad	Desviación de la suma de cuadrados	Promedios cuadrados esperados
Entre grupos	G - 1	SC (EG)	$n'' \sigma_a^2 + n' \sigma_b^2 + \sigma_c^2$
Entre poblaciones / Dentro de grupos	P - G	SC (EP / DG)	$n \sigma_b^2 + \sigma_c^2$
Dentro de poblaciones	N - P	SC (DP)	$\sigma_c^2$
Total	N - 1	SC (T)	$\Sigma_T^2$

Donde las n y los estadísticos F se definen por,

$$S_G = \Sigma \Sigma N_p^2 / N_g$$

$$F_{CT} = \sigma_a^2 / \sigma_T^2$$

$$n = N - S_G / P - G$$

$$F_{SC} = \sigma_b^2 / \sigma_b^2 + \sigma_c^2$$

$$n' = S_G - (\Sigma N_p^2 / N) / G - 1$$

$$F_{ST} = \sigma_a^2 + \sigma_b^2 / \sigma_T^2$$

$$n'' = N - (\Sigma N_g^2 / N) / G - 1$$

El AMOVA *locus por locus* evalúa los componentes de varianza y los estadísticos F para cada *locus* por separado y se reportan en una tabla global. Los diferentes componentes de varianza de diversos niveles se combinan para producir estimaciones sintéticas de los estadísticos F.

### D.3. Relaciones entre poblaciones

Al analizar las relaciones entre las poblaciones lo que se evalúa son las diferencias o afinidades entre las distintas poblaciones, dentro de esto se incluyen los haplotipos compartidos y las distancias genéticas, así como la representación gráfica de las relaciones filogenéticas entre las poblaciones.

### *Distancias genéticas*

Se estimaron dos tipos de distancias genéticas según el tipo de datos a analizar. Para las secuencias de ADNmt se estimó la distancia de Tamura y Nei que brinda el porcentaje corregido de nucleótidos para los que dos haplotipos son diferentes. La corrección permite diferentes tasas de transversión y transición, pero se hace una distinción de las tasas de transición entre purinas y pirimidinas (Tamura y Nei, 1993), que para el presente caso la proporción es de 1:15, respectivamente.

Mientras que para los Y-STR se calcularon las distancias de acuerdo a Nei (1972). Este supone que las diferencias entre las poblaciones se deben a la deriva génica, la tasa de mutación es neutra y cada mutante produce alelos nuevos. Además, se asume que todos los *loci* poseen la misma tasa de mutación neutra y que la variabilidad genética inicial de la población está en equilibrio entre la mutación y la deriva genética, con un tamaño de población efectiva que permanece constante para cada población.

La distancia de Nei se calcula:

$$D = -\ln \left\{ \frac{(\sum_m \sum_i p_{1mi} p_{2mi})}{[(\sum_m \sum_i p_{1mi}^2)^{1/2} (\sum_m \sum_i p_{2mi}^2)^{1/2}]} \right\}$$

Donde  $m$  se suma sobre todos los *loci*,  $i$  sobre los alelos en el *locus*  $m$  y donde  $p_{1mi}$  es la frecuencia del alelo  $i$  en el *locus*  $m$  en la población 1. Debido a las suposiciones anteriores, se espera que la distancia se extienda linealmente en el tiempo para una muestra con muchos *loci* equivalentes.

### *Árbol de agrupamiento de vecinos ("Neighbor joining tree")*

Este método se utiliza para construir árboles filogenéticos a partir de datos de distancias evolutivas. El principio es encontrar pares de unidades taxonómicas operacionales (OTU que corresponden a los vecinos) que minimizen la longitud total de las ramas en cada etapa del agrupamiento. Entonces, se obtiene la topología y la longitud

de las ramas del árbol más parsimonioso utilizando cualquier tipo de datos de distancias evolutivas (Saitou y Nei, 1987).

Un par de vecinos es un par de OTUs conectados por una sola rama interior del árbol. El número de pares de vecinos depende de la topología del árbol, pero el mínimo para N cantidad de OTUs ( $\geq 4$ ) es siempre 2, mientras que el máximo es  $N/2$  cuando N es un número par y  $(N-1)/2$  cuando N es un número impar (Saitou y Nei, 1987).

La topología del árbol se puede definir al unir sucesivamente pares de vecinos y produciendo nuevos pares de vecinos. En general,  $N-2$  pares de vecinos se pueden producir de un árbol bifurcado de N OTUs. Al encontrar sucesivamente esos pares de vecinos se puede obtener la topología del árbol. El método inicia construyendo el árbol a partir de uno en forma de estrella, en el que no hay agrupamientos. Sin embargo, en la práctica algunos pares de OTUs están relacionados más cerca entre ellos que de otros. Entre los posibles pares de OTUs que se obtienen se escoge el que brinda la menor suma de longitud de las ramas. Entonces, este par de OTUs se utiliza como un único OTU, y se continúa escogiendo el siguiente par de OTUs que da la menor suma de longitud de ramas. Esto se continúa hasta obtener  $N-3$  ramas internas en el árbol (Saitou y Nei, 1987).

Algunas longitudes de las ramas pueden ser negativas, por lo que se pueden eliminar de dos maneras. La primera es imponer la condición de que todas las ramas son positivas y luego se reestima la longitud de las ramas. La otra forma es asumir que los estimados negativos se deben a un error de muestreo y que el valor real es cero, lo cual es más razonable pues si se observan los valores absolutos de los estimados negativos generalmente son muy pequeños (Saitou y Nei, 1987).

#### *Escalamiento multidimensional (“Multidimensional scaling”)*

El escalamiento multidimensional (MDS) es una herramienta matemática, alternativa al análisis de factores, que permite representar las similitudes o diferencias de los objetos en un mapa espacial y esto se analiza como matrices de correlación. Los objetos

que son similares resultan en puntos cercanos en el mapa espacial (Schiffman, Reynolds y Young, 1981).

Es una manera de arreglar objetos de una forma eficiente, para alcanzar una configuración que se aproxime a las distancias observadas. La interpretación de las dimensiones es el último paso en el análisis. La orientación de los ejes es arbitraria y pueden ser rotados en cualquier dirección. Primero se deben elaborar los gráficos de dispersión de los objetos en dos planos. Además de las dimensiones importantes, se deben observar los agrupamientos, patrones particulares como círculos o agregados (Schiffman *et al.*, 1981). El valor “stress” es una medida que muestra cuánto se alejan los datos del modelo, entre menor sea este valor la representación gráfica es más adecuada.

#### **D.4. Programas estadísticos**

Los programas estadísticos utilizados son: el Arlequin 2.000 (Schneider, Roessli y Excoffier, 2000) y la versión 1.1 (Schneider *et al.*, 1997), Phylip (Felsestein, 2004) y Statistica (Statsoft, 1995).

Además, el programa Treeview (2001) se utilizó para representar gráficamente los árboles filogenéticos.

## RESULTADOS

### **A. Diversidad genética de las poblaciones de Costa Rica, México y del Suroeste de EUA relacionadas con la esquizofrenia**

#### *A.1. Caracterización de la diversidad genética del cromosoma Y*

El cuadro 2 muestra los datos de las frecuencias alélicas para los 12 microsatélites del cromosoma Y estudiados en las tres poblaciones relacionadas con la esquizofrenia. El marcador DYS385I muestra los alelos de 10 a 18 repeticiones, siendo la población costarricense la única que presenta todas las variantes. Las tres poblaciones presentan el alelo 11 como el más representativo, llegando hasta un 50% en la población de EUA, a un 41% en la población de MX y a un 34% en la población de CR. Este marcador presenta una distribución bimodal en las tres poblaciones.

El DYS385II presentó de 11 a 20 repeticiones en las poblaciones estudiadas, donde ninguna presentó todas las variantes. CR, EUA y MX compartieron, para este marcador, el alelo 14 como el más representativo con un valor de 30%, 47% y 32%, respectivamente. En el estudio de mestizos mexicanos de Rangel-Villalobos *et al.* (2001b), se observa que la combinación 11/14 de los marcadores DYS385I/II es la más frecuente al igual que en el presente estudio.

El alelo 13 del marcador DYS389I fue el más frecuente para las tres poblaciones, con una frecuencia de 54% hasta un 72%. Los alelos encontrados fueron de 12 a 15 repeticiones, siendo la población costarricense la única en presentar las cuatro variantes.

El marcador DYS389II presentó de 27 a 33 repeticiones, siendo el alelo 30 el más representativo en CR y MX con un 28% y 40%, respectivamente. El alelo 29 fue el más frecuente en EUA (42%). Éste marcador y el DYS389I han sido estudiados en la población costarricense (Jiménez, 2000; Ruiz, 1998) y mexicana (Rangel-Villalobos *et al.*, 2001b) por el método de tinción con plata, por lo que la asignación de los alelos se distingue por la resta de los dos fragmentos generados por este marcador (en que el

imprimador delantero imprima dos sitios consecutivos); debido a esto no se pudo comparar dicha información con los presentes datos pues la numeración de los alelos difiere a pesar de ser la misma región amplificada.

El DYS390 mostró de 21 a 25 repeticiones, siendo el alelo 24 el más representativo en las tres poblaciones, con más del 50% de frecuencia hasta un 60%. Este alelo coincide con el estudio en negros (Jiménez, 2000) y amerindios costarricenses (Ruiz, 1998) y con los mestizos mexicanos (Rangel-Villalobos *et al.*, 2001b). Respecto a la Base de Datos de Referencia para EUA del cromosoma Y, se observa que la muestra de dicho país comparte el alelo 24 con los hispanos y europeos americanos, pero no con los africanos (alelo 21 más frecuente). Las 5 variantes se presentaron en las poblaciones de CR y MX.

El marcador DYS391 presentó cuatro repeticiones, de 9 a 12, todas presentes en la población de CR y MX. El alelo 10 fue el más representativo en CR (61%) y MX (47%), mientras que el alelo 11 (47%) lo fue en EUA. En el estudio de Jiménez (2000) en la población negra de CR se observó que el alelo 11 fue el más frecuente, mientras que el segundo en frecuencia fue el alelo 10 (20%). Para los amerindios de la zona costarricense estudiados con este marcador se mostró, al igual que el presente estudio, que el alelo 10 fue el más representativo (Ruiz, 1998). Con respecto a la población mestiza del estudio de Rangel-Villalobos *et al.* (2001b), se observó que el alelo más frecuente fue el 10, al igual que el presente estudio. Los tres grupos étnicos que componen la población estadounidense comparten el alelo 10 como el más representativo, según la Base de Datos de Referencia para EUA del cromosoma Y, situación diferente a este estudio.

El marcador DYS392 presenta un comportamiento bimodal en las poblaciones de CR y EUA. Los alelos más frecuentes son el 11 (40%) y 13 (37%) en CR, y los mismos en EUA pero el 13 es más frecuente (57%), que es la misma situación evidenciada en la Base de Datos de Referencia para EUA del cromosoma Y. En la población mexicana el alelo 13 fue el más representativo (36%). Este marcador mostró de 9 a 16 repeticiones y solo en la población de CR se observan todas las variantes.

Las tres poblaciones comparten el alelo 13 como el más frecuente en el marcador DYS393, en proporciones desde 66% hasta 83%. En este se presentan las repeticiones de 12 hasta 15 veces del motivo AGAT, todas presentes solo en la población CR. Este resultado coincide con los datos de negros (Jiménez, 2000) y amerindios costarricenses (Ruiz, 1998), además de los mestizos mexicanos (Rangel-Villalobos *et al.*, 2001b) y la Base de Datos de Referencia para EUA del cromosoma Y.

El marcador DYS437 es el que presentó menos repeticiones, con los alelos 14 al 16, siendo el único caso en que los tres alelos estuvieron presentes en las tres poblaciones. El alelo 14 fue el más frecuente en las tres poblaciones (CR 43%, EUA 52% y MX 61%). Este dato concuerda con la información de negros costarricenses pero difiere para los mestizos costarricenses, según el estudio de Gusmão *et al.* (2003b), en que el alelo más frecuente fue el 15 y el segundo en frecuencia fue el 14 (34%).

El marcador DYS438 mostró de 9 a 13 repeticiones del motivo TTTTC, las 5 variantes estuvieron presentes en la población de CR y MX. El alelo 12 fue el más frecuente para las tres poblaciones, con 38%, 54% y 45% de frecuencia en CR, EUA y MX, respectivamente. En este caso la coincidencia es con los mestizos pero difiere con los negros costarricenses, en que el alelo 11 es el más representativo, y los alelos 10 y 13 presentan igual frecuencia (Gusmão *et al.*, 2003b).

Los datos del marcador DYS439 mostraron que las tres poblaciones comparten el alelo 12 como el más frecuente, aunque la población de EUA presenta el alelo 11 con igual frecuencia al 12 (40%). Este marcador mostró de 9 a 14 repeticiones, todas presentes solo en la población de MX. Al comparar esta información con los datos de Gusmão *et al.* (2003b), se observa que este comportamiento concuerda con la muestra de negros y mestizos costarricenses de su estudio.

Por último, el marcador DYS19 mostró de 12 a 17 repeticiones, siendo el alelo 14 el más representativo en las tres poblaciones (45% CR, 57% EUA y 54% MX). A diferencia del resultado presente, en el estudio de negros (Jiménez, 2000) y amerindios (Ruiz, 1998) costarricenses el alelo más frecuente fue el 13 y en segundo lugar se halló el 14 con una frecuencia de 29% y 10% para negros y amerindios respectivamente. No

obstante, en la muestra de mestizos mexicanos sí hubo coincidencia en el alelo 14 como el más frecuente (Rangel-Villalobos *et al.*, 2001b) y para la Base de Datos de Referencia para EUA del cromosoma Y en hispanos y europeos americanos. Para este marcador, la población de EUA fue la única que no presentó todas las variantes alélicas.

En resumen, se observa que para 9 marcadores las tres poblaciones comparten el alelo más frecuente, indicando la relación genética de los hombres de las poblaciones estudiadas, información que será confirmada con los siguientes análisis.

**Cuadro 2.** Frecuencias y diversidades génicas (DG) obtenidas al analizar 12 Y-STR en Costa Rica (CR), Estados Unidos (EUA) y México (MX)

Y-STR	CR	EUA	MX	Y-STR	CR	EUA	MX
<b>DYS 385I</b>				<b>DYS 385II</b>			
10	0.008 (0.008)	0.024 (0.024)	0.036 (0.025)	11	0.017 (0.012)		
11	<b>0.342 (0.043)</b>	<b>0.500 (0.078)</b>	<b>0.418 (0.067)</b>	12	0.025 (0.014)	0.071 (0.040)	0.018 (0.018)
12	0.083 (0.025)	0.048 (0.033)	0.055 (0.031)	13	0.033 (0.016)	0.024 (0.024)	0.036 (0.025)
13	0.167 (0.034)	0.167 (0.058)	0.109 (0.042)	14	<b>0.308 (0.042)</b>	<b>0.476 (0.078)</b>	<b>0.327 (0.064)</b>
14	0.192 (0.036)	0.190 (0.061)	0.109 (0.042)	15	0.217 (0.038)	0.190 (0.061)	0.200 (0.054)
15	0.058 (0.021)	0.048 (0.033)	0.182 (0.052)	16	0.067 (0.023)	0.095 (0.046)	0.055 (0.031)
16	0.083 (0.025)		0.073 (0.035)	17	0.217 (0.038)	0.071 (0.040)	0.127 (0.045)
17	0.058 (0.021)	0.024 (0.024)	0.018 (0.018)	18	0.108 (0.028)	0.048 (0.033)	0.109 (0.042)
18	0.008 (0.008)			19		0.024 (0.024)	0.127 (0.045)
DG	0,798	0,680	0,758	20	0.008 (0.008)		
				DG	0,793	0,714	0,804
<b>DYS 389I</b>				<b>DYS 389II</b>			
12	0.217 (0.038)	0.190 (0.061)	0.109 (0.042)	27		0.024 (0.024)	0.018 (0.018)
13	<b>0.542 (0.046)</b>	<b>0.619 (0.076)</b>	<b>0.727 (0.061)</b>	28	0.175 (0.035)	0.119 (0.051)	0.109 (0.042)
14	0.225 (0.038)	0.190 (0.061)	0.164 (0.050)	29	0.258 (0.040)	<b>0.429 (0.077)</b>	0.327 (0.064)
15	0.017 (0.012)			30	<b>0.283 (0.041)</b>	0.333 (0.074)	<b>0.400 (0.067)</b>
DG	0,609	0,544	0,432	31	0.225 (0.038)	0.095 (0.046)	0.127 (0.045)
				32	0.050 (0.020)		0.018 (0.018)
				33	0.008 (0.008)		
				DG	0,769	0,681	0,704
<b>DYS 390</b>				<b>DYS 391</b>			
21	0.100 (0.028)		0.018 (0.018)	9	0.008 (0.008)	0.095 (0.046)	0.055 (0.031)
22	0.125 (0.030)	0.119 (0.051)	0.018 (0.018)	10	<b>0.617 (0.045)</b>	0.429 (0.077)	<b>0.473 (0.068)</b>
23	0.183 (0.035)	0.238 (0.067)	0.182 (0.052)	11	0.333 (0.043)	<b>0.476 (0.078)</b>	0.455 (0.068)
24	<b>0.508 (0.046)</b>	<b>0.595 (0.077)</b>	<b>0.600 (0.067)</b>	12	0.042 (0.018)		0.018 (0.018)
25	0.083 (0.025)	0.048 (0.033)	0.182 (0.052)	DG	0,507	0,580	0,567
DG	0,675	0,573	0,573				

## Continuación Cuadro 2

<b>DYS 392</b>				<b>DYS 393</b>			
9	0.008 (0.008)			12	0.150 (0.033)	0.167 (0.058)	0.109 (0.042)
10	0.017 (0.012)	0.024 (0.024)		13	<b>0.667 (0.043)</b>	<b>0.762 (0.067)</b>	<b>0.836 (0.050)</b>
11	<b>0.400 (0.045)</b>	0.238 (0.067)	0.182 (0.052)	14	0.175 (0.035)	0.071 (0.040)	0.055 (0.031)
12	0.008 (0.008)	0.024 (0.024)		15	0.008 (0.008)		
13	0.375 (0.044)	<b>0.571 (0.077)</b>	<b>0.364 (0.065)</b>	DG	<i>0,502</i>	0,387	0,286
14	0.108 (0.028)	0.048 (0.033)	0.327 (0.064)				
15	0.075 (0.024)	0.024 (0.024)	0.055 (0.031)				
16	0.008 (0.008)	0.071 (0.040)	0.073 (0.035)				
DG	0,682	0,608	0,719				
<b>DYS 437</b>				<b>DYS 438</b>			
14	<b>0.433 (0.045)</b>	<b>0.524 (0.078)</b>	<b>0.618 (0.066)</b>	9	0.067 (0.023)	0.095 (0.046)	0.018 (0.018)
15	0.392 (0.045)	0.405 (0.077)	0.327 (0.064)	10	0.267 (0.041)	0.190 (0.061)	0.182 (0.052)
16	0.175 (0.035)	0.071 (0.040)	0.055 (0.031)	11	0.275 (0.041)	0.167 (0.058)	0.327 (0.064)
DG	<i>0,628</i>	0,557	0,508	12	<b>0.383 (0.045)</b>	<b>0.548 (0.078)</b>	<b>0.455 (0.068)</b>
				13	0.008 (0.008)		0.018 (0.018)
				DG	<i>0,702</i>	0,627	0,653
<b>DYS 439</b>				<b>DYS 19</b>			
9		0.024 (0.024)	0.018 (0.018)	12	0.008 (0.008)		0.018 (0.018)
10	0.017 (0.012)	0.071 (0.040)	0.055 (0.031)	13	0.192 (0.036)	0.262 (0.069)	0.309 (0.063)
11	0.333 (0.043)	<b>0.405 (0.077)</b>	0.218 (0.056)	14	<b>0.458 (0.046)</b>	<b>0.571 (0.077)</b>	<b>0.545 (0.068)</b>
12	<b>0.433 (0.045)</b>	<b>0.405 (0.077)</b>	<b>0.618 (0.066)</b>	15	0.217 (0.038)	0.143 (0.055)	0.055 (0.031)
13	0.167 (0.034)	0.095 (0.046)	0.055 (0.031)	16	0.092 (0.026)	0.024 (0.024)	0.036 (0.025)
14	0.050 (0.020)		0.036 (0.025)	17	0.033 (0.016)		0.036 (0.025)
DG	<i>0,671</i>	0,658	0,563	DG	<i>0,697</i>	0,584	0,601

Nota: Entre paréntesis se encuentran las desviaciones estándar de las frecuencias. En itálicas se resalta la DG más alta entre las tres poblaciones, en negrita se resalta la frecuencia alélica más alta en cada población. Cálculos realizados con Arlequín 2.0.

Las diversidades génicas son una medida de variabilidad de cada marcador dentro de una población, para el presente estudio se presentan en el cuadro 2.

La población de Costa Rica muestra un rango de 0,798 (DYS385I) a 0,502 (DYS393). Especialmente en esta población nueve marcadores presentan el valor más alto de DG respecto a EUA y MX. En la población estadounidense las DG estuvieron entre 0,714 (DYS385II) a 0,387 (DYS393) y de 0,804 (DYS385II) a 0,286 (DYS393) en

la población mexicana. Esto evidencia que el marcador menos informativo es el DYS393 y los más informativos son el DYS385I y el DYS385II para las tres poblaciones en general.

Cuando se comparan estos datos con los de otras poblaciones se observa que: en el estudio de Gusmão *et al.* (2003b) los mestizos de CR presentan la misma DG para el DYS439 que en esta población (67%), a diferencia del DYS438 que mostró un 59% de DG y el DYS437 con 56%. Por otro lado, el estudio de Rangel-Villalobos *et al.* (2001b) con mestizos mexicanos mostró similitud en DG para los marcadores DYS19 (66%), DYS390 (66,4%) y DSY391 (53%). Pero se mostraron diferencias grandes, para los marcadores DYS393 (51,4%) y DYS385I/II (92,5%).

Las DG promedio (cuadro 3) mostraron que las tres poblaciones superaron el 60% (CR 67,5%, EUA 61,4% y MX 60,8%), indicando que estos marcadores genéticos son altamente informativos ya que presentan una alta variabilidad dentro de las poblaciones. Esta medida depende de los marcadores estudiados y el tamaño de la muestra analizada, sin embargo cuando se promedia se observa que en general la combinación de más de cinco microsatélites del cromosoma Y lleva a un valor superior al 50% (cuadro 4).

La diversidad haplotípica es una medida de la probabilidad de encontrar dos haplotipos diferentes en una misma población. Para las poblaciones analizadas en el presente estudio este valor es superior al 99% (CR 99,7%, EUA 99,8% y MX 99,5%) indicando que los haplotipos generados al estudiar estos 12 Y-STR aportan suficiente información para diferenciar individuos dentro de una misma población y por lo tanto para encontrar diferencias entre poblaciones. Esto se puede observar en el Cuadro 3.

Según publicaciones y trabajos previos en poblaciones relacionadas a las del presente estudio, estas diversidades haplotípicas son similares cuando se trata de mestizos y negros CR; africanos, europeos e hispanos de EUA y mestizos de MX, teniendo en consideración que no se utilizaron los mismos 12 marcadores del cromosoma Y del presente estudio. Y sí se hallan diferencias respecto a los amerindios, mestizos y negros de CR (ver cuadro 4).

**Cuadro 3.** Medidas de diversidad dentro de las poblaciones relacionadas con la esquizofrenia de Costa Rica, Estados Unidos y México al analizar 12 Y-STR

	<b>CR</b>	<b>EUA</b>	<b>MX</b>
<b>No. individuos</b>	120	42	55
<b>No. haplotipos</b>	103	40	51
<b>No. haplotipos compartidos</b>	14	2	2
<b>No. haplotipos únicos</b>	89	38	49
<b>Diversidad haplotípica</b>	0,997 (0,002)	0,998 (0,006)	0,995 (0,005)
<b>Diversidad génica promedio</b>	0,675 (0,349)	0,614 (0,324)	0,608 (0,321)
<b>No. promedio de diferencias entre individuos</b>	8,099 (3,786)	7,368 (3,516)	7,300 (3,470)
<b>Promedio de la distribución de diferencias</b>	8.099	7.368	7.300
<b>Varianza de la distribución de diferencias</b>	5.238	6.265	5.102

Nota: Entre paréntesis se encuentra la desviación estándar de los datos. Cálculos realizados con Arlequín 2.0.

**Cuadro 4.** Comparaciones de las poblaciones de Costa Rica, Estados Unidos y México analizadas con 12 Y-STR respecto a otros estudios con microsatélites del cromosoma Y

Población	N	Diversidad haplotípica	Promedio o rango de diversidad génica	Número promedio de diferencias
<b>CR</b>				
Mestizos ■	120	0,997	0,675	8,019
Mestizos *	102	0,993	0,595	
Negros *	18	0,994	0,614	
Amerindios ♦	93	0,937	0,434	3,797 ○
Valle Central ∞	89	0,950		
Negros □	34	0,961	0,576	5,740
<b>EUA</b>				
Mestizos ■	42	0,998	0,614	7,368
Africanos ♠	599	0,998		10,760
Europeos ♠	628	0,996		8,610
Hispanos ♠	478	0,995		10,540
<b>MX</b>				
Mestizos ■	55	0,995	0,608	7,300
Mestizos ●	107	0,993	0,925-0,514	
Mestizos †	101	0,990	0,747-0,571	

■ Presente estudio

\* *Loci* DYS460, DYS461, GATA A10, GATA C4, GATA H4, DYS437, DYS438 y DYS439. Gusmão *et al.*, 2003b.

♦ *Loci* YAP, DYS199, ah, DYS19, DYS389a/b, DYS390, DYS391 y DYS393. Tribus Cabécar, Bribri, Teribe, Huetar y Guaymí. Ruiz, 1998.

○ Equivale al número de repeticiones por STR.

∞ *Loci* DYS19, DYS388, DYS391, DYS392 y DYS393. Carvajal-Carmona *et al.*, 2003.

□ *Loci* DYS19, DYS389a/b, DYS390, DYS391 y YAP. Jiménez, 2000.

♠ *Loci* DYS19, DYS389I/II, DYS390, DYS391, DYS392, DYS393 y DYS385I/II. Kayser *et al.*, 2003.

● *Loci* DYS19, DYS385, DYS389I/II, DYS390, DYS391 y DYS393. Rangel-Villalobos *et al.*, 2001b.

† *Loci* GATA A4, GATA A7.1 (DYS460), GATA A7.2 (DYS461), GATA A10 y GATA C4. Rangel-Villalobos *et al.*, 2001a.

Al evaluar los haplotipos generados por los 12 Y-STR se observó que en la muestra costarricense se obtuvieron 103 haplotipos, a partir de 120 individuos. De estos, 89 haplotipos fueron únicos dentro de la población y 14 fueron compartidos (11 compartidos por 2 individuos y 3 haplotipos compartidos por 3 individuos).

La población estadounidense presentó un total de 40 haplotipos al analizar los 42 individuos. Se presentaron dos haplotipos compartidos que los presentan dos individuos dentro de la población.

Por último, la muestra mexicana de 55 individuos presentó 51 haplotipos, de estos 49 fueron únicos y 2 fueron compartidos, uno por 4 personas y el otro por 2 individuos. Esta información se observa en los Cuadros 3 y 5.

La composición de los haplotipos obtenidos se observa en el apéndice A.

**Cuadro 5.** Haplotipos compartidos entre y dentro de las tres poblaciones de estudio (Costa Rica, Estados Unidos y México) al analizarse 12 Y-STR

Haplotipo	CR	EUA	MX	Haplotipo	EUA	MX
HY5	2 (0.017)			HY69	2 (0.017)	
HY10	2 (0.017)			HY79	2 (0.017)	
HY14	2 (0.017)			HY83	2 (0.017)	
HY25	3 (0.025)	2 (0.048)		HY91	2 (0.017)	
HY29	2 (0.017)			HY106	1 (0.024)	1 (0.018)
HY30	3 (0.025)			HY109	2 (0.048)	
HY38	2 (0.017)			HY119	1 (0.024)	1 (0.018)
HY41	2 (0.017)			HY125	1 (0.024)	4 (0.073)
HY52	2 (0.017)			HY135	1 (0.024)	1 (0.018)
HY66	3 (0.025)			HY181		2 (0.036)

Nota: entre paréntesis se encuentran las frecuencias de los haplotipos dentro de las poblaciones. Cálculos realizados con Arlequín 2.0.

Al evaluar el promedio de diferencias entre dos haplotipos, es decir, el número de repeticiones que separan a dos haplotipos dentro de la población, promediado para cada par de comparaciones; se observa que para la población costarricense es de 8.099, mientras que en México y EUA es 7,3 (ver cuadro 3).

El cuadro 6 muestra que en la población de CR a las 8 diferencias se presenta un acumulado de casi el 50% de todas las diferencias dentro de la población, mientras que para EUA y MX esta proporción se alcanza en las 7 diferencias. El máximo de diferencias para las poblaciones son 12 para los Y-STR.

La figura 5 muestra la distribución de diferencias para cada población y se observa un comportamiento similar para las tres poblaciones, presentándose un pico correspondiente al promedio de diferencias. Este comportamiento revela un evento de expansión de las poblaciones.

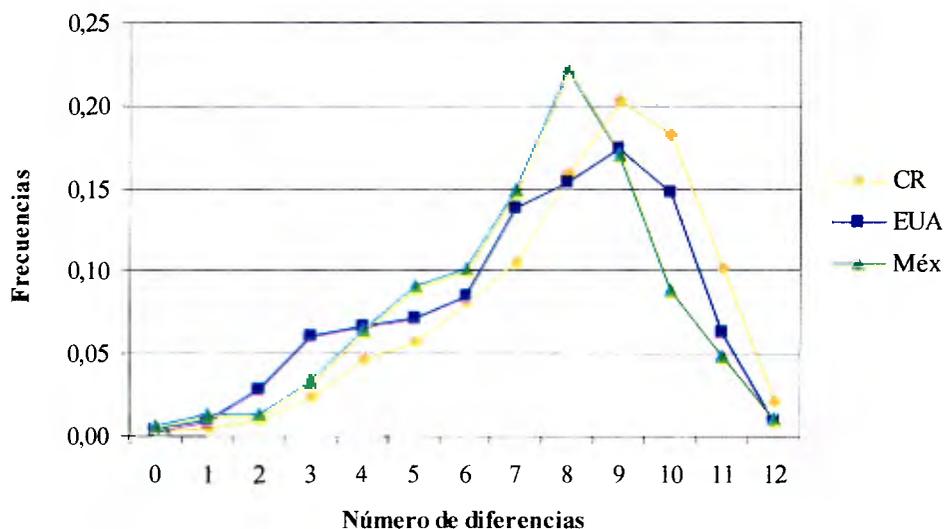
El estudio de Ruiz (1998) muestra que para la población amerindia analizada, el comportamiento de la distribución de diferencias es similar al presente estudio, en cuanto a la presencia de una moda, pero el promedio de diferencias se presenta en 3,797. El caso de la muestra de negros limonenses de Jiménez (2000) muestra un promedio de diferencias de 5,74. Y según el estudio de Kayser *et al.* (2003) el promedio para la muestra de EUA está entre 8,61 y 10,76 (ver cuadro 4).

**Cuadro 6.** Número de diferencias entre pares de individuos en Costa Rica, Estados Unidos y México a partir de los 12 Y-STR analizados

<b>No. de diferencias</b>	<b>CR</b>	<b>EUA</b>	<b>MX</b>
0	0.003 (0,003)	0.002 (0,002)	0.005 (0,005)
1	0.005 (0,008)	0.008 (0,010)	0.011 (0,016)
2	0.009 (0,017)	0.028 (0,038)	0.012 (0,028)
3	0.024 (0,041)	0.059 (0,097)	0.032 (0,060)
4	0.047 (0,088)	0.066 (0,163)	0.064 (0,124)
5	0.057 (0,145)	0.071 (0,234)	0.090 (0,214)
6	0.082 (0,227)	0.084 (0,318)	0.100 (0,314)
7	0.105 (0,332)	0.137 (0,455)	0.149 (0,463)
8	0.160 (0,492)	0.153 (0,608)	0.222 (0,685)
9	0.204 (0,696)	0.173 (0,781)	0.170 (0,855)
10	0.182 (0,878)	0.148 (0,929)	0.088 (0,943)
11	0.101 (0,979)	0.063 (0,992)	0.048 (0,991)
12	0.022 (1,001)	0.008 (1,000)	0.009 (1,000)

Nota: Entre paréntesis se especifica la frecuencia acumulada. Cálculos realizados con Arlequín 2.0.

**Figura 5.** Distribución de diferencias (“Mismatch distribution”) entre pares de individuos para Costa Rica, Estados Unidos y México al analizar 12 Y-STR



Nota: elaborado con Excel 2003.

### *A.2. Caracterización de la diversidad genética del ADNmt*

El análisis de la región HVI del ADNmt (del nucleótido 16047 al 16365, 319 nucleótidos en total) en las tres poblaciones mostró la presencia de 52, 39 y 56 sitios polimórficos en CR, EUA y MX, respectivamente. De los cuales 18, 10 y 19 son únicos para las mencionadas poblaciones. Las transiciones fueron más comunes (48 en CR, 36 en EUA y 53 en MX) que las transversiones (4 en CR y MX, 3 en EUA), aproximando una razón 12,41:1. Esto se observa en el cuadro 8.

Se presentaron en total 88 polimorfismos de una base, de los cuales sólo 18 polimorfismos se compartieron entre las tres poblaciones. Hay 4 polimorfismos que se comparten entre CR y EUA, 12 que se comparten entre CR y MX, y 7 que se comparten entre EUA y MX.

En general, el nucleótido de la secuencia Anderson es más frecuente (>80%) que el polimorfismo, sin embargo en las posiciones 16111T, 16183C, 16189C, 16223T,

16290T, 16319A y 16362C para las tres poblaciones, y la 16217C en CR y MX el polimorfismo presenta más del 20% hasta un 71% de frecuencia. Esta información se observa en el cuadro 7.

Al comparar estos datos con el estudio de Santos *et al.* (1994) se observa que los polimorfismos reportados para la muestra Huetar (amerindios) están presentes en la población costarricense.

Al realizar una comparación similar con el estudio de los mestizos costarricenses (Meléndez, 2001), se observan mayores diferencias. De los 53 polimorfismos encontrados en la región HVI del estudio mencionado se comparten 37 polimorfismos. Además, hay 16 polimorfismos presentes en el estudio de mestizos que no están presentes en el actual estudio, y 15 polimorfismos que presentan el caso contrario.

Los datos de la población mexicana se compararon con el estudio de originarios de México respecto a la región HVI (Green *et al.*, 2000) y se observó que se comparten 34 polimorfismos de un total de 71 encontrados en dicha publicación. Se observaron 37 polimorfismos no presentes en el actual estudio y 23 ausentes.

De igual forma se comparan los datos actuales con una publicación que estudia las regiones HVI y HVII con oligonucleótidos específicos de secuencia (SSO, por sus siglas en inglés) en la población estadounidense (Melton *et al.*, 2001). Específicamente se utilizaron 21 sondas que reconocen 13 polimorfismos para generar los haplotipos. Respecto al presente análisis, están representados 6 de los 7 polimorfismos de la región HVI estudiados anteriormente.

También se presentaron mutaciones de longitud del tipo inserciones en las posiciones 16188.1C, 16192.1C y 16193.1C, que no fueron estudiadas en el análisis poblacional (Cordaux *et al.*, 2003; Salas *et al.*, 2004).

La base de datos Mitomap registra el total de polimorfismos reportados hasta el 16 de setiembre del 2004 respecto a la región control del ADNmt. Cuando se compara con la presente información se observaron 4 polimorfismos no reportados previamente (16175G, 16204A, 16287G, 16348T).

**Cuadro 7.** Sustituciones presentes en la región HVI del ADNmt para las poblaciones analizadas de Costa Rica, Estados Unidos y México

Posición de referencia	CR		EUA		MX		
16051					A (0.891)	G (0.091)*	
16069	C (0.967)	T (0.025 *)	C (0.976)	T (0.024)	C (0.982)	T (0.018)	
16086	T (0.992)	C (0.008)	T (0.976)	C (0.024)			
16091	A (0.992)	T (0.008)					
16092			T (0.976)	C (0.024)			
16093	T (0.983)	C (0.017)			T (0.927)	C (0.073)	
16102			T (0.976)	C (0.024)			
16104					C (0.982)	T (0.018)	
16105	T (0.992)	G (0.008)					
16111	C (0.450)	T (0.550)	C (0.595)	T (0.405)	C (0.509)	T (0.491)	
16114	C (0.983)	T (0.017)					
16124					T (0.982)	C (0.018)	
16126	T (0.967)	C (0.033)			T (0.964)	C (0.036)	
16129	G (0.942)	A (0.058)	G (0.976)	A (0.024)	G (0.946)	A (0.036)	C (0.018)
16131					T (0.982)	C (0.018)	
16134					C (0.982)	T (0.018)	
16136	T (0.958)	C (0.042)			T (0.964)	C (0.036)	
16145	G (0.992)	A (0.008)			G (0.982)	A (0.018)	
16148					C (0.964)	T (0.036)	
16150	C (0.983)	T (0.017)					
16155	A (0.992)	G (0.008)					
16168	C (0.983)	T (0.017)					
16169	C (0.992)	T (0.008)					
16172	T (0.983)	C (0.017)			T (0.964)	C (0.036)	
16174					C (0.982)	T (0.018)	
16175	A (0.967)	G (0.033)	A (0.976)	G (0.024)			
16176			C (0.976)	T (0.024)			
16177	A (0.992)	G (0.008)					
16182	A (0.917)	C (0.083)	A (0.857)	C (0.143)	A (0.855)	C (0.146)	
16183	A (0.683)	C (0.317)	A (0.786)	C (0.214)	A (0.691)	C (0.309)	
16185	C (0.992)	T (0.008)					

*Continuación Cuadro 7*

Posición de referencia	CR		EUA		MX	
16186					C (0.982)	T (0.018)
16187	C (0.925)	T (0.075)				
16188			C (0.952)	T (0.048)	C (0.964)	T (0.036)
16189	T (0.567)	C (0.433)	T (0.786)	C (0.214)	T (0.636)	C (0.364)
16201	C (0.992)	T (0.008)				
16204					G (0.964)	A (0.036)
16209	T (0.925)	C (0.075)			T (0.982)	C (0.018)
16214			C (0.976)	T (0.024)		
16217	T (0.725)	C (0.275)	T (0.857)	C (0.143)	T (0.764)	C (0.236)
16222					C (0.982)	T (0.018)
16223	C (0.375)	T (0.625)	C (0.286)	T (0.714)	C (0.346)	T (0.655)
16224	T (0.992)	C (0.008)	T (0.976)	C (0.024)	T (0.982)	C (0.018)
16233			A (0.976)	G (0.024)		
16239					C (0.909)	T (0.091)
16240			A (0.976)	G (0.024)	A (0.982)	G (0.018)
16242			C (0.976)	T (0.024)		
16247			A (0.976)	G (0.024)	A (0.982)	T (0.018)
16249	T (0.992)	C (0.008)				
16256			C (0.976)	T (0.024)	C (0.982)	T (0.018)
16257	C (0.992)	T (0.008)			C (0.982)	T (0.018)
16259					C (0.964)	T (0.036)
16260					C (0.982)	T (0.018)
16261	C (0.983)	T (0.017)	C (0.976)	T (0.024)	C (0.982)	T (0.018)
16264	C (0.992)	T (0.008)				
16271					T (0.982)	C (0.018)
16274			G (0.976)	A (0.024)	G (0.982)	A (0.018)
16278	C (0.967)	T (0.033)			C (0.927)	T (0.073)
16284			A (0.976)	G (0.024)		
16287			C (0.976)	G (0.024)		
16288					T (0.982)	C (0.018)
16290	C (0.450)	T (0.550)	C (0.452)	T (0.548)	C (0.436)	T (0.564)
16292	C (0.967)	T (0.033)			C (0.982)	T (0.018)
16293	A (0.992)	G (0.008)				

*Continuación Cuadro 7*

Posición de referencia	CR		EUA		MX	
16294	C (0.992)	T (0.008)			C (0.982)	T (0.018)
16295			C (0.976)	T (0.024)	C (0.964)	T (0.036)
16298	T (0.933)	C (0.067)	T (0.857)	C (0.143)	T (0.982)	C (0.018)
16299	A (0.967)	G (0.033)				
16300	A (0.983)	G (0.017)				
16304			T (0.976)	C (0.024)		
16309					A (0.982)	G (0.018)
16311	T (0.967)	C (0.033)	T (0.952)	C (0.048)	T (0.909)	C (0.091)
16319	G (0.458)	A (0.542)	G (0.452)	A (0.548)	G (0.418)	A (0.582)
16320	C (0.992)	T (0.008)	C (0.976)	T (0.024)	C (0.964)	T (0.036)
16324	T (0.992)	C (0.008)			T (0.982)	C (0.018)
16325	T (0.908)	C (0.092)	T (0.857)	C (0.143)	T (0.873)	C (0.127)
16327	C (0.925)	T (0.075)	C (0.881)	T (0.119)	C (0.927)	T (0.073)
16335	A (0.992)	G (0.008)			A (0.982)	G (0.018)
16343					A (0.982)	G (0.018)
16344					C (0.982)	T (0.018)
16348					C (0.982)	T (0.018)
16352	T (0.925)	C (0.075)	T (0.976)	C (0.024)		
16354			C (0.976)	T (0.024)		
16355	C (0.992)	T (0.008)	C (0.952)	T (0.048)		
16356	T (0.958)	C (0.042)				
16357			T (0.905)	C (0.095)	T (0.964)	C (0.036)
16360	C (0.825)	T (0.175)				
16362	T (0.425)	C (0.575)	T (0.381)	C (0.619)	T (0.418)	C (0.582)

\* La frecuencia no suma 1 porque una muestra no pudo ser secuenciada en dicha región.

Nota: La primera posición corresponde al nucleótido de la secuencia Anderson, la segunda es la sustitución en la población. Entre paréntesis se indica la frecuencia en que está presente la sustitución en la población respectiva. Cálculos realizados con Arlequín 2.0.

La diversidad de nucleótidos revela la heterocigosidad a nivel de nucleótidos y para las tres poblaciones se presenta un valor similar de 0.02 (ver cuadro 8). En organismos

eucariotas se ha estimado que varía de 0,002 a 0,019 y es muy similar para genes nucleares como para mitocondriales (Nei, 1987).

Estos datos se comparan con la información del ADNmt en otras publicaciones como un estudio en Asia Central (Comas *et al.*, 1998) cuyas poblaciones estudiadas mostraron un rango de diversidad de 0,016-0,018. En el estudio de Bonatto y Salzano (1997b) de nativos americanos los valores son muy reducidos (0,80-0,96%) al analizar los haplogrupos fundadores. Un estudio realizado por Bonatto y Salzano (1997a) con amerindios del haplogrupo A mostró que la diversidad de nucleótidos dentro de las poblaciones estuvo entre 0,64-1,38%. Al analizar los linajes maternos de los pobladores del sur de Siberia se observó una diversidad de 0,012-0,017 (Derenko *et al.*, 2003). Y los valores de diversidad en un estudio de la India mostraron un rango de 0,005-0,023 (Cordaux *et al.*, 2003). Aún en una población más relacionada como la amerindia costarricense del estudio de Santos *et al.* (1994) se observa que la diversidad es muy reducida (0,004). Este valor tan variado en los estudios mencionados, depende de las poblaciones estudiadas y su origen genético.

Al igual que en los Y-STR, la diversidad haplotípica implica la probabilidad de encontrar dos haplotipos diferentes dentro de la población. El valor más bajo lo presentó la población de CR con un 93%. Y las poblaciones mexicana y estadounidense obtuvieron un valor similar de 98%. Estos datos se observan en el cuadro 8.

Si se observan estos datos comparados con otras publicaciones se notan diferencias, especialmente determinadas por la población estudiada (ver cuadro 9). En general, los amerindios presentan diversidades más bajas y el resto de poblaciones muestran valores muy similares a los del presente estudio, sugiriendo heterogeneidad interna en las poblaciones respecto al estudio del ADNmt.

**Cuadro 8.** Evaluación de la diversidad en la región HVI del ADNmt dentro de la población de Costa Rica, Estados Unidos y México

	<b>CR</b>	<b>EUA</b>	<b>MX</b>
<b>No. individuos</b>	120	42	55
<b>Sustituciones</b>	52	39	57
<b>Transiciones</b>	48	36	53
<b>Transversiones</b>	4	3	4
<b>No. sitios polimórficos</b>	52	39	56
<b>No. polimorfismos únicos</b>	18	10	19
<b>Diversidad de nucleótidos</b>	0.024 (0.012)	0.022 (0.012)	0.026 (0.014)
<b>No. haplotipos totales</b>	39	33	43
<b>No. haplotipos únicos</b>	22	29	35
<b>No. haplotipos compartidos</b>	17	4	8
<b>Diversidad haplotípica</b>	0.930 (0.012)	0.981 (0.011)	0.989 (0.007)
<b>No. promedio de diferencias entre individuos</b>	7.657 (3.595)	6.881 (3.303)	8.445 (3.967)
<b>Promedio de la distribución de diferencias</b>	6.369	5.869	7.009
<b>Varianza de la distribución de diferencias</b>	14.493	10.296	12.461
<b>Índice de desigualdad de Harpending (P)</b>	0,009 (0,740)	0,012 (0,819)	0,012 (0,610)

Nota: Entre paréntesis se encuentra la desviación estándar, con excepción del índice de Harpending que indica la P. Cálculos realizados con Arlequín 2.0.

**Cuadro 9.** Comparaciones con otros estudios al analizar la región control del ADNmt en las poblaciones de Costa Rica, Estados Unidos y México

<b>Población</b>	<b>N</b>	<b>Región</b>	<b>Diversidad haplotípica</b>	<b>Promedio en el número de diferencias</b>
<b>CR</b>				
Mestizos *	120	HVI	0,930	7,657
Valle Central ♦	58	HVI	0,870	
Mestizos □	90	HVI	0,991	12,730 ►
		HVII	0,960	
Amerindios ■	108	Restricción	0,764	
Amerindios ■	91	HVI	0,782	
Negros Centroamérica ◇	223	HVI	0,981	7,200
Amerindios ●	27	HVI-HVII	0,840 □	
<b>EUA</b>				
Mestizos *	42	HVI	0,981	6,881
Africanos ◇	101	HVI	0,991	7,800
Africanos ♦	805	HVI-HVII	0,983	
Europeos ♦	922	HVI-HVII	0,964	
Hispanos ♦	461	HVI-HVII	0,967	
<b>MX</b>				
Mestizos *	55	HVI	0,989	8,445
Norte central ∞	63	HVI	0,960	

\* Presente estudio

♦ Carvajal-Carmona *et al.*, 2003.

□ Meléndez, 2001.

► Incluye las sustituciones, deleciones e inserciones encontradas en las secuencias.

■ Ruiz, 1998.

◇ Incluye 8 negros de México, además de negros de Belice, República Dominicana, Colombia y Panamá. Salas *et al.*, 2004.

● Santos *et al.*, 1994.

□ No incluye las mutaciones de longitud.

♦ Detección de la variación con sondas SSO ("Sequence Specific Oligonucleotides"). Melton *et al.*, 2001

∞ Green *et al.*, 2000. Citado por Carvajal-Carmona *et al.*, 2003.

El número de haplotipos obtenidos para el ADNmt es reducido comparado al cromosoma Y. En este caso para la población de CR se obtuvieron 39 haplotipos de los 120 individuos estudiados, de estos 22 son únicos. Dentro de la población se comparten 17 haplotipos, 7 de estos son compartidos por dos individuos, 3 son compartidos por tres individuos y los restantes 7 son compartidos por 4, 6, 7, 8, 11, 19 y 20 individuos (ver cuadros 8 y 10). Es importante resaltar que el haplotipo HM12 es igual a la secuencia Anderson y está representado en dos individuos de la población costarricense y ausente en la muestra mexicana y estadounidense.

En EUA se hallaron 33 haplotipos compartidos, de los cuales 29 son únicos. Los haplotipos compartidos están presentes 2 en 4 individuos, 1 en tres y 1 en dos individuos, sumando 4 haplotipos y 13 individuos (ver cuadro 8 y 10).

En México se observaron 5 haplotipos compartidos por dos individuos, 2 haplotipos compartidos por 3 individuos y 1 compartido por 4 individuos. En total son 43 haplotipos, siendo 35 únicos y 8 compartidos dentro de la población (cuadro 8 y 10).

La composición de los haplotipos obtenidos por el estudio de las secuencias de la región HVI del ADNmt se encuentran en el Apéndice B.

Las tres poblaciones mostraron desigualdades en las diferencias entre pares de individuos que es una medida de diversidad molecular, ya que la población mexicana presenta en promedio 8,445 nucleótidos diferentes entre dos haplotipos, mientras que la costarricense difiere en 7,657 nucleótidos y la estadounidense en 6,881 (ver cuadro 8). De igual forma y proporción varió el promedio de la distribución de diferencias con los valores 7,009; 6,369 y 5,869; respectivamente (ver cuadro 8 y 11). Esto implica que aproximadamente en este valor se alcanza la mitad de las diferencias entre los individuos dentro de cada población.

Esta información se observa más claramente en la distribución de diferencias (“Mismatch distribution”), que es un reflejo de la expansión de las poblaciones y por lo tanto de su historia. En las tres poblaciones se distinguió un comportamiento bimodal, mostrando algún evento de reducción de las poblaciones (cuello de botella), que se comentará más adelante (ver figura 6). Es importante resaltar que la primera moda en la

**Cuadro 10.** Haplotipos compartidos dentro y entre las poblaciones de Costa Rica, Estados Unidos y México al estudiar la región HVI del ADNmt

Haplotipo	CR	EUA	MX	Haplotipo	CR	EUA	MX
HM1	7 (0.058)			HM25	8 (0.067)		
HM 4	19 (0.158)	4 (0.095)	4 (0.073)	HM26	2 (0.017)		
HM5	2 (0.017)			HM33	2 (0.017)		
HM7	11 (0.092)			HM39	1 (0.008)		1 (0.018)
HM8	3 (0.025)			HM43		1 (0.024)	1 (0.018)
HM10	6 (0.050)	2 (0.048)	1 (0.018)	HM49		4 (0.095)	1 (0.018)
HM12	2 (0.017)			HM53		1 (0.024)	1 (0.018)
HM13	20 (0.167)			HM54		3 (0.071)	3 (0.055)
HM16	3 (0.025)			HM57		1 (0.024)	1 (0.018)
HM18	2 (0.017)			HM69		1 (0.024)	2 (0.036)
HM19	3 (0.025)			HM74			3 (0.055)
HM20	1 (0.008)		1 (0.018)	HM83			2 (0.036)
HM21	2 (0.017)			HM85			2 (0.036)
HM22	2 (0.017)			HM97			2 (0.036)
HM24	4 (0.033)			HM100			2 (0.036)

Nota: Entre paréntesis se encuentra la frecuencia con que está representado el haplotipo en la respectiva población. Cálculos realizados con Arlequín 1.1.

distribución se encuentra entre la diferencia 1 y 2, mostrando una semejanza en la historia de las tres poblaciones.

Otro indicador del comportamiento de expansión de las poblaciones es el índice de desigualdad de Harpending (cuadro 8). Cuando este es menor a 0,03 confirma que las poblaciones sufrieron un crecimiento, y esto es evidente para las poblaciones del presente estudio (0,009 CR; y 0,012 MX y EUA).

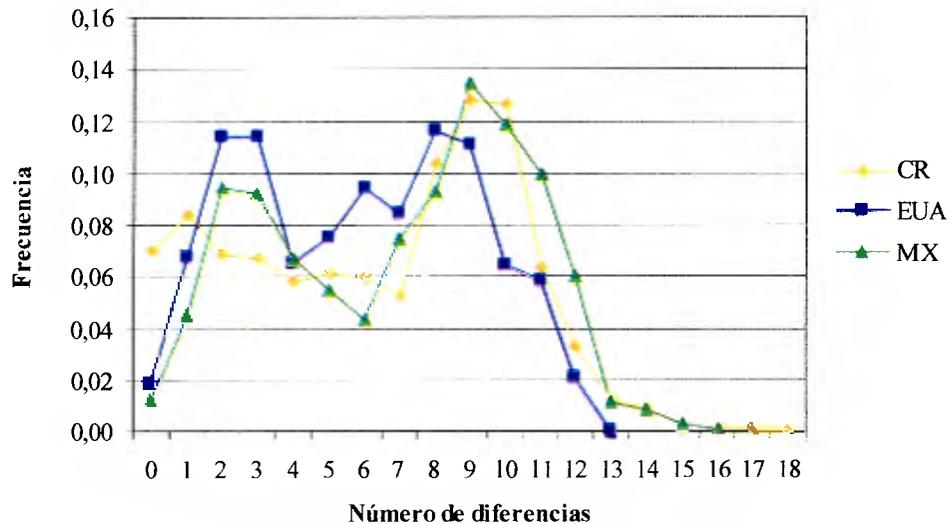
El máximo de diferencias entre pares de individuos para CR es de 18, para EUA es de 13 y para MX es de 16 (ver cuadro 11).

**Cuadro 11.** Número de diferencias entre individuos en los haplotipos del ADNmt dentro de la muestra de Costa Rica, Estados Unidos y México

<b>No. de diferencias</b>	<b>CR</b>	<b>EUA</b>	<b>MX</b>
<b>0</b>	0.070 (0,070)	0.019 (0,019)	0.011 (0,011)
<b>1</b>	0.084 (0,154)	0.067 (0,086)	0.044 (0,055)
<b>2</b>	0.069 (0,223)	0.114 (0,2)	0.094 (0,149)
<b>3</b>	0.067 (0,29)	0.114 (0,314)	0.092 (0,241)
<b>4</b>	0.058 (0,348)	0.065 (0,379)	0.066 (0,307)
<b>5</b>	0.061 (0,409)	0.074 (0,453)	0.054 (0,361)
<b>6</b>	0.060 (0,469)	0.094 (0,547)	0.042 (0,403)
<b>7</b>	0.053 (0,522)	0.084 (0,631)	0.074 (0,477)
<b>8</b>	0.104 (0,626)	0.116 (0,747)	0.092 (0,569)
<b>9</b>	0.128 (0,754)	0.110 (0,857)	0.134 (0,703)
<b>10</b>	0.127 (0,881)	0.064 (0,921)	0.118 (0,821)
<b>11</b>	0.063 (0,944)	0.058 (0,979)	0.099 (0,92)
<b>12</b>	0.033 (0,977)	0.021 (1,00)	0.059 (0,979)
<b>13</b>	0.012 (0,989)	0.000	0.011 (0,99)
<b>14</b>	0.009 (0,998)		0.007 (0,997)
<b>15</b>	0.001 (0,999)		0.002 (0,999)
<b>16</b>	0.002 (1,001)		0.000
<b>17</b>	0.001 (1,002)		
<b>18</b>	0.000		

Nota: entre paréntesis se indica la frecuencia acumulada. Cálculos realizados con Arlequín 2.0.

**Figura 6.** Distribución de diferencias (“Mismatch distribution”) en las poblaciones de Costa Rica, Estados Unidos y México a partir de la secuencia HVI del ADNmt



Nota: elaborado con Excel 2003.

### **B. Análisis de la estructura jerárquica de las poblaciones de Costa Rica, México y el Suroeste de EUA relacionadas con la esquizofrenia**

El AMOVA realizado para evaluar la diferenciación de las poblaciones a partir del ADNmt, revela que no hay diferencia entre las tres poblaciones ( $F_{st}$ : 0.006,  $P > 0.05$ ), pues el 99,36% de la variación se encuentra dentro de las poblaciones (ver cuadro 12).

Si se estudian las poblaciones separándolas por las seis regiones de muestreo (Costa Rica CR, San Antonio SA, San Diego SD, Los Ángeles LA, Monterrey MT y Ciudad de México CM) se obtiene un  $F_{st}$  de 0.005 ( $P > 0.304$ ) con un porcentaje de variación dentro de poblaciones de 99,51%, lo que implica que no hay diferencia significativa entre las regiones (ver cuadro 12).

**Cuadro 12.** Análisis molecular de varianza en las tres poblaciones de estudio o separadas por regiones al analizar la secuencia HVI del ADNmt

	Varianza	F <sub>st</sub>	P
<b>3 poblaciones *</b>			
Entre poblaciones	0.640	0.006	0.162
Dentro de poblaciones	99.360		
<b>6 regiones **</b>			
Entre poblaciones	0.960	0.005	0.304
Entre poblaciones dentro grupos	-0.470		
Dentro de poblaciones	99.510		

\* Costa Rica, Estados Unidos y México

\*\* Costa Rica, San Diego, San Antonio, Los Ángeles, Monterrey y Ciudad de México

Nota: Cálculos realizados con Arlequín 2.0.

Al analizar la diferenciación de las tres poblaciones por medio de un AMOVA para la información de los Y-STR, éste revela que el 96% de las diferencias se debe a la variación interna de la población ( $R_{st}$ : 0,04;  $P < 0,001$ ) y se obtiene un resultado significativo al separar la variación por cada *locus* ( $R_{st}$ : 0,016;  $P < 0,001$ ) (ver cuadro 13). Específicamente para el marcador DYS392 ( $R_{st}$ : 0,051;  $P < 0,001$ ) en que la variación entre poblaciones es de 5,148% y dentro de poblaciones 94,852%. Un caso similar se obtiene con el DYS439 con 97,8% de variación dentro de la población con una probabilidad de 0,038; implicando diferenciación pero no en un grado contundente (ver cuadro 14).

Un resultado similar se obtiene al estudiar la muestra de EUA y MX dividida por zonas de muestreo o regiones (SD, LA, SA, MT y CM) donde el 96,25% de la diferenciación se debe a variaciones internas dentro de la muestra ( $R_{st}$ : 0,038;  $P < 0,01$ , ver cuadro 12) y se obtiene un resultado significativo al separar la variación por cada *locus* ( $R_{st}$ : 0,019;  $P < 0,001$ ). Estas diferencias se deben a los marcadores DYS392 ( $R_{st}$ :

0,044;  $P < 0,01$ ), DYS439 ( $R_{st}$ : 0,043;  $P < 0,01$ ) y DYS 389II ( $R_{st}$ : 0,037;  $P < 0,01$ ) (ver cuadro 15).

Cuando se presentan varianzas negativas en los AMOVA se debe a que son realmente covarianzas. Esto sucede cuando no hay estructura genética, debido a que el valor verdadero del parámetro a estimar es cero. Su significado biológico puede ser que los genes de poblaciones diferentes pueden estar más relacionados entre ellos que con los genes de la misma población.

**Cuadro 13.** Análisis molecular de varianza con 12 Y-STR y *locus por locus* en las tres poblaciones de Costa Rica, México y el Suroeste de Estados Unidos, y separadas por regiones

	12 Y-STR			<i>Locus por locus</i>	
	Varianza	$R_{st}$	P	$R_{st}$	P
<b>3 poblaciones *</b>					
Entre poblaciones	4.000	0.040	0.001	0.016	< 0.001
Dentro de poblaciones	96.000				
<b>6 regiones **</b>					
Entre poblaciones	4.530	0.038	0.008	0.019	< 0.001
Entre poblaciones dentro grupos	-0.770				
Dentro de poblaciones	96.250				

\* Costa Rica, Suroeste de Estados Unidos y México

\*\* Costa Rica, San Diego, San Antonio, Los Ángeles, Monterrey y Ciudad de México

Nota: Cálculos realizados con Arlequín 2.0.

**Cuadro 14.** Análisis molecular de varianza por cada *locus* de los 12 Y-STR analizados en Costa Rica, el Suroeste de Estados Unidos y México

<b>Marcador</b>	<b>Varianza</b>	<b>R<sub>st</sub></b>	<b>P</b>	<b>Marcador</b>	<b>Varianza</b>	<b>R<sub>st</sub></b>	<b>P</b>
<b>DYS 385I</b>				<b>DYS 392</b>			
EP	0.737	0.007	0.147	EP	5.148	0.051	0.001
DP	99.263			DP	94.852		
<b>DYS 385II</b>				<b>DYS 393</b>			
EP	0.974	0.010	0.098	EP	1.900	0.019	0.069
DP	99.026			DP	98.100		
<b>DYS 389I</b>				<b>DYS 437</b>			
EP	1.095	0.011	0.147	EP	1.332	0.013	0.123
DP	98.905			DP	98.668		
<b>DYS 389II</b>				<b>DYS 438</b>			
EP	1.003	0.010	0.125	EP	0.735	0.007	0.190
DP	98.997			DP	99.265		
<b>DYS 390</b>				<b>DYS 439</b>			
EP	0.809	0.008	0.157	EP	2.192	0.022	0.038
DP	99.191			DP	97.808		
<b>DYS 391</b>				<b>DYS 19</b>			
EP	2.278	0.023	0.066	EP	1.202	0.012	0.104
DP	97.722			DP	98.798		

Nota: EP: variación entre poblaciones, DP: variación dentro de poblaciones. Cálculos realizados con Arlequín 2.0.

**Cuadro 15.** Análisis molecular de varianza por cada *locus* de los 12 Y-STR estudiados en seis regiones de Costa Rica, el Suroeste de Estados Unidos (San Antonio, San Diego y Los Ángeles) y México (Monterrey y Ciudad de México)

<b>Marcador</b>	<b>Varianza</b>	<b>R<sub>st</sub></b>	<b>P</b>	<b>Marcador</b>	<b>Varianza</b>	<b>R<sub>st</sub></b>	<b>P</b>
<b>DYS 385I</b>				<b>DYS 392</b>			
EP	2.001	0.002	0.045	EP	6.767	0.044	0.010
EP-DG	-1.847			EP-DG	-2.354		
DP	99.846			DP	95.587		
<b>DYS 385II</b>				<b>DYS 393</b>			
EP	0.712	0.011	0.145	EP	2.975	0.014	0.206
EP-DG	0.383			EP-DG	-1.569		
DP	98.905			DP	98.593		
<b>DYS 389I</b>				<b>DYS 437</b>			
EP	-1.260	0.022	0.088	EP	3.185	0.005	0.385
EP-DG	3.438			EP-DG	-2.705		
DP	97.821			DP	99.520		
<b>DYS 389II</b>				<b>DYS 438</b>			
EP	-4.897	0.037	0.003	EP	0.097	0.010	0.214
EP-DG	8.616			EP-DG	0.932		
DP	96.281			DP	98.971		
<b>DYS 390</b>				<b>DYS 439</b>			
EP	3.196	-0.003	0.619	EP	-2.486	0.043	0.007
EP-DG	-3.486			EP-DG	6.823		
DP	100.290			DP	95.663		
<b>DYS 391</b>				<b>DYS 19</b>			
EP	1.847	0.025	0.114	EP	-0.839	0.021	0.069
EP-DG	0.628			EP-DG	2.980		
DP	97.525			DP	97.859		

Nota: EP: entre poblaciones, EP-DG: entre poblaciones dentro de grupos, DP: dentro de poblaciones. Cálculos realizados con Arlequín 2.0.

### **C. Relaciones entre las poblaciones de Costa Rica, México y el Suroeste de EUA relacionadas con la esquizofrenia**

A diferencia de los Y-STR, en las secuencias del ADNmt se encuentran dos haplotipos compartidos por las tres poblaciones, en los que particularmente uno de ellos (HM4) está representado 19 veces en la población costarricense (frecuencia 0,158), 4 en la mexicana (frecuencia 0,073) y 4 en la estadounidense (frecuencia 0,095). Las sustituciones que componen este haplotipo corresponden a polimorfismos de alta frecuencia entre las poblaciones estudiadas (16111T, 16223T, 16290T, 16319A y 16362C, ver cuadro 7 y apéndice B).

El otro haplotipo compartido es el HM10 presente en 6 individuos de CR, 2 de EUA y 1 de MX, compuesto por los polimorfismos 16182C, 16183C, 16189C y 16217C (ver apéndice B).

Además, se presentaron 2 haplotipos compartidos entre la población de CR y la MX (HM20 y HM39), y 6 haplotipos compartidos entre EUA y MX, entre los cuales uno está representado 4 veces en EUA y 1 en MX (HM49), otro se encuentra en 3 individuos en cada población (HM54) y el HM69 se encuentra en 2 individuos de MX y 1 de EUA. Los otros 3 se encuentran en un individuo de cada población (HM43, HM53 y HM57).

Por otro lado, al analizar los haplotipos compartidos obtenidos por los Y-STR se observó que entre la población costarricense y estadounidense se comparte un haplotipo representado 3 y 2 veces en las respectivas poblaciones (HY25), mientras que MX y EUA comparten 4 haplotipos (HY106, HY119, HY125 y HY135, ver cuadro 5 y apéndice A), uno de los cuales es el más representado en MX (HY125 con una frecuencia de 0,073).

Cuando se estudian las relaciones entre las poblaciones de Costa Rica, Estados Unidos y México por distancias genéticas se observa que para el ADNmt y el cromosoma Y no son diferenciables, como se nota en el cuadro 16.

Sin embargo, en baja proporción, se puede observar una mayor distancia entre CR y EUA respecto a los dos sistemas genéticos analizados. Y una mayor cercanía entre MX y EUA, especialmente para el ADNmt. Como se explicará más adelante, este puede ser un efecto de la cercanía geográfica que favorece las migraciones entre vecinos.

**Cuadro 16.** Distancias genéticas para los 12 Y-STR (Nei 1972, arriba de la diagonal) y la región HVI del ADNmt (Tamura y Nei 1993, debajo de la diagonal) en Costa Rica, Estados Unidos y México

ADNmt \ Y-STR	CR	EUA	MX
CR		0,075	0,073
EUA	0,015		0,072
MX	0,003	-0,002	

Nota: Cálculos realizados con Arlequín 2.0. y Phylip versión 3.6.

En el cuadro 17, se evalúan las distancias genéticas entre las regiones estudiadas. Nuevamente no se observan diferencias significativas respecto al ADNmt, que son más claras al ver las representaciones gráficas de estos datos. Sin embargo, es evidente la separación de CR y SD del resto de regiones.

Por otra parte, los datos del cromosoma Y son más extremos, especialmente los datos de Ciudad de México y San Diego respecto a las demás regiones, y la distancia es aún mayor entre estas dos regiones. Este detalle es enmascarado al analizar los datos en las tres poblaciones, y de esta forma revela información importante del comportamiento de las poblaciones humanas relacionadas a la distancia geográfica y la historia que comparten. Sin embargo, estos resultados deben tomarse con cautela debido al tamaño reducido de las muestras, con excepción de la muestra costarricense.

**Cuadro 17.** Distancias genéticas para los 12 Y-STR (Nei 1972, arriba de la diagonal) y la región HVI del ADNmt (Tamura y Nei 1993, debajo de la diagonal) en las tres poblaciones de estudio, divididas por regiones

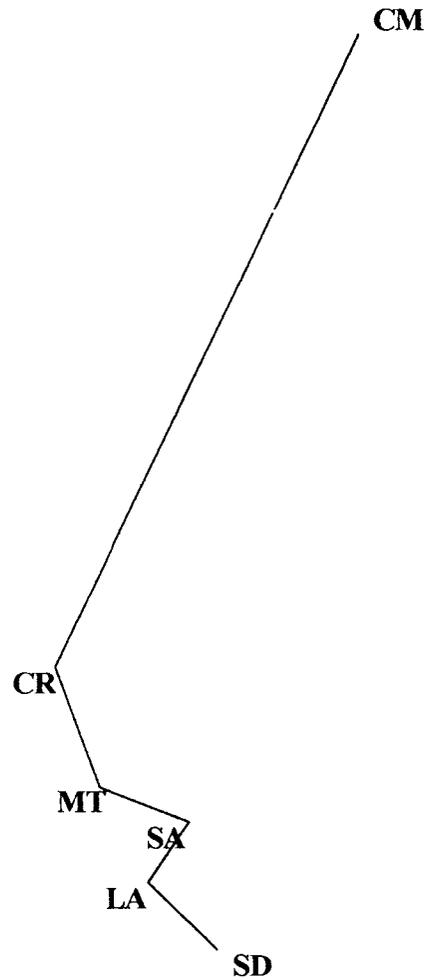
ADNmt \ Y-STR	CR	SA	SD	LA	MT	CM
CR		0,064	0,233	0,064	0,080	2,730
SA	0,004		0,277	0,080	0,062	3,540
SD	0,050	0,052		0,167	0,253	9,014
LA	0,019	-0,036	0,107		0,102	3,526
MT	-0,003	-0,012	-0,008	0,010		2,825
CM	0,004	-0,021	0,062	-0,018	-0,013	

Nota: CR: Costa Rica, SA: San Antonio, SD: San Diego, LA: Los Ángeles, MT: Monterrey, CM: Ciudad de México. Cálculos realizados con Arlequín 2.0. y Phylip versión 3.6.

Los datos de entrada para la elaboración de las representaciones gráficas son las distancias genéticas analizadas previamente. En estos casos lo que hacen estas figuras es confirmar la información que se argumenta con los datos. Ambas representaciones son útiles, aunque el escalamiento multidimensional es más adecuado cuando se trata de vecinos geográficos.

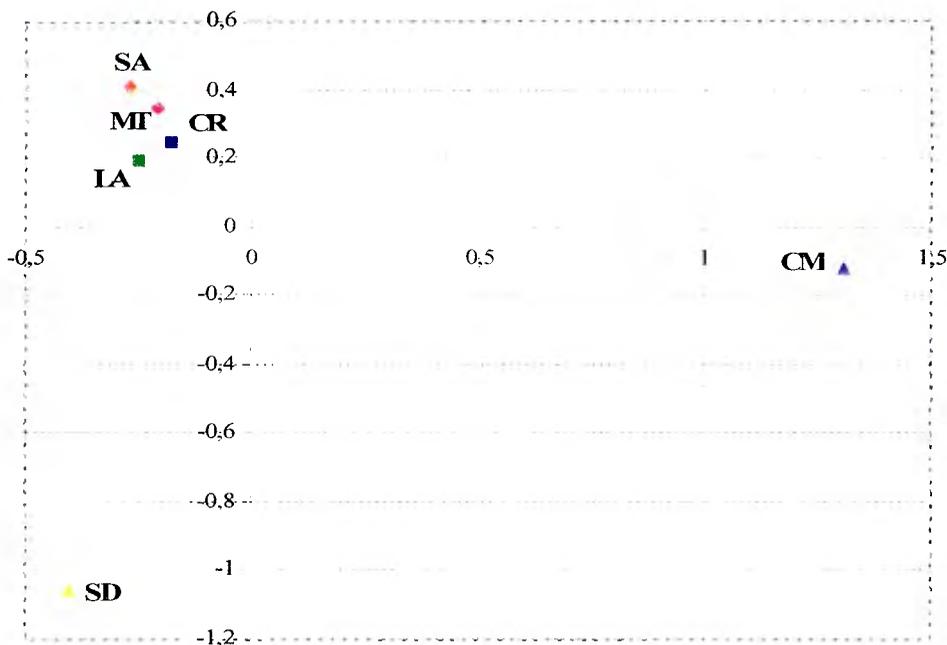
Como se dijo anteriormente, es evidente la separación de San Diego y Ciudad de México respecto a las demás regiones, que se encuentran aglomeradas. En el árbol de agrupamiento de vecinos (figura 7) se observa más claramente la distancia entre San Diego y Ciudad de México, encontrándose San Diego más cercano al agrupamiento, que se puede observar también en el escalamiento multidimensional (figura 8).

**Figura 7.** Árbol de agrupamiento de vecinos basado en el estudio de 12 Y-STR para Costa Rica, el Suroeste de Estados Unidos y México, separado en regiones



Nota: elaborado con Phylip versión 3.6 y Treeview.

**Figura 8.** Escalamiento Multidimensional basado en los datos de 12 Y-STR para las poblaciones hispanas de Costa Rica, el suroeste de Estados Unidos y México separadas por regiones

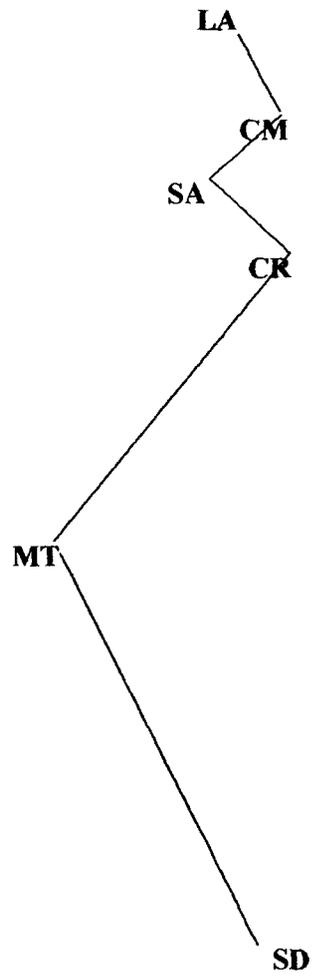


Nota: elaborado con Statistica y Excel 2003.

Es relevante en estas representaciones del ADNmt que para las regiones de CR y MT la distribución no es similar, pues en el árbol de agrupamiento de vecinos (figura 9), MT presenta una posición intermedia entre CR y SD. Mientras que en el escalamiento multidimensional forma parte del agrupamiento del resto de regiones.

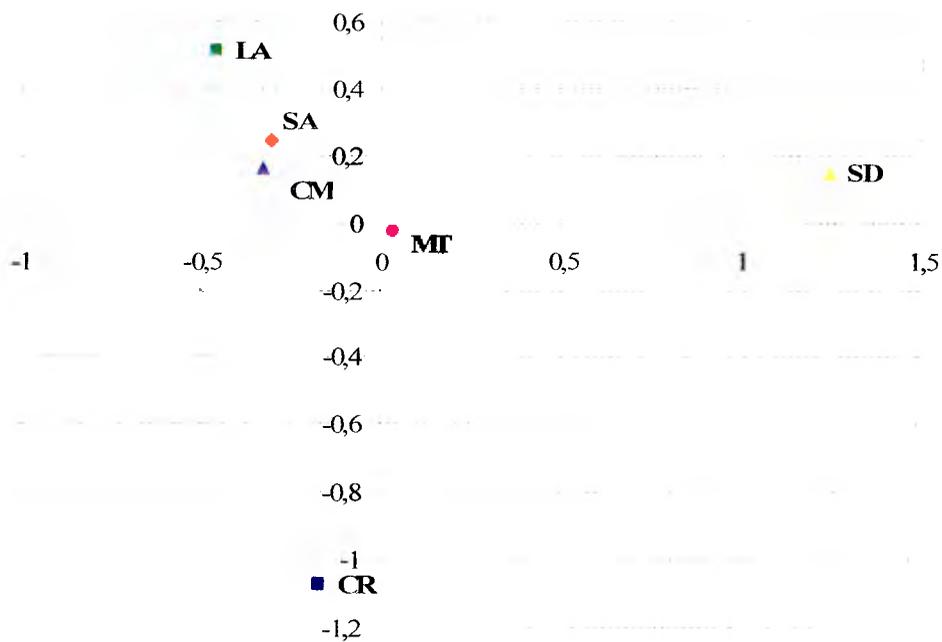
La observación de este agrupamiento en el MDS es evidente, conformado por LA, SA, CM y MT; así como la separación de CR y SD que será explicado adelante.

**Figura 9.** Árbol de agrupamiento de vecinos basado en la región HVI del ADNmt para Costa Rica, el Suroeste de Estados Unidos y México, separado en regiones basado en las distancias de Tamura y Nei



Nota: elaborado con Phylip versión 3.6 y Treeview.

**Figura 10.** Escalamiento Multidimensional basado en los datos de la región HVI del ADNmt para las poblaciones hispanas de Costa Rica, el suroeste de Estados Unidos y México separadas por regiones



Nota: elaborado con Statistica y Excel 2003.

## DISCUSIÓN

### **A. Diversidad genética de las poblaciones de Costa Rica, México y del Suroeste de EUA relacionadas con la esquizofrenia**

#### *A.1. Caracterización de la diversidad genética del cromosoma Y*

El estudio de los 12 microsatélites del cromosoma Y en las poblaciones del presente estudio (CR, EUA y MX) fue una elección adecuada, debido a la inmensa variabilidad encontrada dentro de las poblaciones favoreciendo todos los análisis posteriores a partir de las frecuencias alélicas. Además, la utilización del kit Powerplex Y System (Promega) agilizó el trabajo al identificar los alelos y economizar ciertos reactivos al realizar una sola reacción para los 12 Y-STR.

Al comparar las tres poblaciones estudiadas se observó la relación respecto al alelo más frecuente para 9 marcadores, este es un fuerte indicio de relación genética por la historia que comparten los linajes paternos de estas regiones (Bianchi *et al.*, 1998), incluyendo todo Centroamérica y parte de Suramérica. Historia relacionada con nativos provenientes de las mismas regiones de Asia y colonizaciones provenientes de Europa (Bortolini *et al.*, 2003). Sin embargo, estas comparaciones y análisis de frecuencias no son concluyentes.

Si se comparan las poblaciones con otros estudios a partir de los mismos marcadores puede brindarse una señal de la representación que puedan tener las poblaciones relacionadas con la esquizofrenia respecto a las poblaciones sanas. Así, se observa que la población mestiza mexicana estudiada por Rangel-Villalobos *et al.* (2001b), coincide en los alelos más representativos de 6 Y-STR analizados aquí. Esto lleva a la idea de una representatividad de la presente muestra de hombres respecto a las poblaciones sanas del noroeste de México. Además es algo de esperar, ya que se ha probado que el cromosoma Y no tiene ninguna influencia genética en la esquizofrenia, por lo que las distribuciones

no deben ser diferentes del resto de la población, a menos que halla uno o varios linajes asociados, que aún no hayan sido identificados.

En el caso de la población de CR relacionada con la enfermedad, se observa que en dos marcadores coinciden los alelos más frecuentes con la muestra de mestizos (Gusmão *et al.*, 2003b), en cuatro marcadores coinciden con dos estudios de negros costarricenses (Gusmão *et al.*, 2003b; Jiménez, 2000) y en tres marcadores coinciden con los amerindios (Ruiz, 1998). Esta información no es suficiente para alcanzar conclusiones acerca de las relaciones de la presente muestra respecto a los demás grupos que conforman el país, por la falta de marcadores estudiados en conjunto y por lo tanto no se puede definir si esta muestra es representativa de la población mestiza natural del país. A futuro, se puede pensar en completar la información con los marcadores que faltan en la muestra de Gusmão *et al.* (2003b) y así concluir estadísticamente una relación válida.

De igual forma, la población de EUA se comparó con la Base de Datos de Referencia del Cromosoma Y resultando en la identidad de cuatro marcadores para los grupos de hispanos y europeos americanos, y en dos marcadores con los negros americanos. Esta información al igual que en los casos anteriores no es suficiente para alcanzar conclusiones importantes, aunque sí es evidencia de similitud entre la muestra relacionada con la esquizofrenia y la población natural.

Para confirmar estadísticamente esta información para las tres poblaciones sería necesario un análisis de distancias genéticas y de haplotipos compartidos, inclusive de haplotipos específicos del cromosoma Y, con los datos de una muestra aleatoria de la población sana respecto a la relacionada con la enfermedad utilizando los mismos marcadores genéticos, pues no se pueden alcanzar conclusiones acertadas solo comparando frecuencias alélicas.

Respecto al DYS385I/II la asignación de los alelos se hizo otorgando el alelo de menos repeticiones al DYS385I y el de más al DYS385II, aunque se sugiere analizar el marcador como un genotipo, Kayser *et al.* (2003) encontró que el método utilizado aquí no genera discordancias en los resultados.

En ciertos marcadores se observó un comportamiento bimodal en las frecuencias alélicas (DYS385I para CR, EUA y MX; DYS385II para MX; DYS392 para CR y EUA), esto significa que los alelos más frecuentes en la población no son los que se diferencian en tan solo una unidad repetitiva como se espera según el modelo de mutación de un paso para microsatélites (Cavalli-Sforza, 1998). Este comportamiento de las frecuencias alélicas se ha sugerido es una marca de la historia de las poblaciones que implica eventos de migración o fundadores que involucran la entrada de nuevas variantes que con el tiempo aumentan sus proporciones en la población, es decir la existencia de dos poblaciones con trasfondos genéticos diferentes, como sucede en Finlandia (Carracedo *et al.*, 2001). Para las poblaciones de este estudio esta explicación es probable ya que en el caso de MX y CR se ha comprobado la composición trihíbrida de las poblaciones (Green *et al.*, 2000; Morera *et al.*, 2003) y para EUA es aún más diversa aunque fragmentada geográficamente (Kayser *et al.*, 2003).

Si se trata de explicar este fenómeno según la herencia de las poblaciones europeas y negras se pueden alcanzar ideas interesantes, pues el componente de herencia masculina es mayor respecto a estos dos grupos que de poblaciones amerindias. Al comparar las frecuencias de las poblaciones de España y Portugal (Carracedo *et al.*, 2001; Gusmão *et al.*, 2003b), que son las de mayor contribución masculina a América durante la colonia, se observa que ambas poblaciones comparten los mismos alelos frecuentes para 8 Y-STR. La comparación de esta información con las tres poblaciones de este estudio revela que en 6 marcadores comparten el alelo más frecuente, y además EUA lo comparte con el DYS389II, aunque MX y CR comparten el alelo de segunda frecuencia de este marcador. Esto indica relación entre las cinco poblaciones comparadas respecto a los linajes paternos (Bianchi *et al.*, 1998).

El otro caso corresponde a los linajes negros, especialmente para la población costarricense que se ha estudiado previamente (Gusmão *et al.* 2003b, Jiménez 2000), se observa que las poblaciones negras analizadas concuerdan en cuatro marcadores respecto al alelo más frecuente en la población del presente estudio, revelando de igual forma relación producto de la hibridación en la conformación de la población.

A pesar de que la información individual de cada marcador es importante, la forma de analizar los datos es en haplotipos. Una medida de la cantidad de información que en conjunto aportan los marcadores es la diversidad haplotípica, que revela para las tres poblaciones un 99% aproximadamente, un valor suficientemente alto para encontrar variabilidad y por lo tanto diferenciación dentro de cada población de estudio, como ha sido probado en múltiples investigaciones a nivel mundial. Este valor se ve confirmado por los haplotipos compartidos dentro de las poblaciones, siendo en proporciones de 2 y 3 individuos, y un solo caso de 4 individuos dentro de la población mexicana. Esta variabilidad reflejada por estos datos explica los valores obtenidos al efectuar los posteriores análisis de varianza y distancias genéticas, ya que si hay dos poblaciones poco diversas internamente posiblemente sean semejantes al compararlas, mientras que la diversidad garantiza que si hay semejanzas no se deberá a la homogeneidad de las muestras.

Otro análisis de los datos que muestran la diversidad existente dentro de las muestras es el estudio de las diferencias entre pares de individuos (Kayser *et al.*, 2003). Los valores menores implican pérdida de diversidad (Zegura *et al.*, 2004), situación contraria a las poblaciones aquí estudiadas debido a la mezcla de la cual son producto. Además, se observó que en proporción este valor indicó una semejanza para las tres poblaciones, además esto se ve confirmado por la distribución de diferencias que muestra un comportamiento similar, en forma de campana. Específicamente, la distribución de diferencias está señalando un evento de expansión antiguo común a las tres poblaciones (Rogers y Harpending, 1992), posiblemente, revelando una vez más la similitud en la historia paterna de las tres poblaciones.

Las comparaciones que se realizaron con otras publicaciones en las mismas poblaciones del presente estudio muestran variaciones del promedio en el número de diferencias, ya que se trata de poblaciones con historias diferentes como es el caso de los amerindios que residen en Costa Rica (Ruiz, 1998) y los negros de CR (Jiménez, 2000), a diferencia de las poblaciones mestizas con un alto contenido europeo, como las muestras del presente estudio. En el caso de la muestra de EUA se presentan más

semejanzas con la muestra europea de Texas que la hispana, resaltando el componente europeo de la población mestiza estudiada aquí (Kayser *et al.*, 2003).

#### *A.2. Caracterización de la diversidad genética del ADNmt*

Los resultados del estudio de la región HVI del ADNmt concuerdan con múltiples estudios mundiales acerca de la proporción de transiciones y transversiones, y polimorfismos más comunes como lo son 16129, 16223, 16311 y 16362 (Hagelberg, 2002), como se comentará más adelante.

La presencia de 18 polimorfismos compartidos entre las tres poblaciones es un indicio de la relación histórica de las líneas maternas, aún más para las poblaciones de CR y MX que presentan 12 polimorfismos en común, y entre EUA y MX que comparten 7 polimorfismos. Es interesante observar que 5 de los polimorfismos compartidos por las tres poblaciones presentan frecuencias altas respecto a la secuencia Anderson (de origen europeo), y que además están presentes en poblaciones amerindias y mestizas de CR y MX (Green *et al.*, 2000; Meléndez, 2001; Santos *et al.*, 1994). Esto puede reflejar el contenido genético de las poblaciones estudiadas y la semejanza entre los linajes maternos de las mismas poblaciones.

Por otro lado, la diversidad de nucleótidos también confirma la semejanza de las poblaciones estudiadas, al obtener el mismo valor de 0,02. Este valor implica que al tomar dos secuencias y comparar dos nucleótidos homólogos al azar existe un 2% de probabilidad de que sean diferentes. Siempre hay que tomar con cautela los datos individuales de los polimorfismos debido a la alta tasa de mutación de la región control permitiendo la presencia de mutaciones reversas en sitios altamente mutables (Tully *et al.*, 2001), lo que dificulta el análisis de eventos filogenéticos.

Las diversidades haplotípicas obtenidas, aunque menos diversas que los Y-STR, muestran una alta variabilidad que es importante para hallar diferencias dentro de las poblaciones. Además, son datos similares a otras investigaciones realizadas que confirman a su vez los datos obtenidos (Carvajal-Carmona *et al.*, 2003; Meléndez, 2001;

Melton *et al.*, 2001; Salas *et al.*, 1994), a excepción de los datos en amerindios que ciertamente son poblaciones con determinada estructura genética y muestran poca mezcla (Ruiz, 1998; Santos *et al.*, 1994).

Es evidente que a pesar de que la población de CR es más del doble que la MX y EUA, la cantidad de haplotipos encontrados no es proporcional, ya que hay una sobre representación de algunos (tres haplotipos con 20, 19 y 11 individuos que los comparten). Los haplotipos compartidos dentro de las poblaciones son datos que revelan específicamente la homogeneidad interna de las poblaciones. Es evidente que el linaje materno de los individuos relacionados con la esquizofrenia es más homogéneo en CR que en MX y EUA.

Es recomendable realizar un análisis estadístico comparativo de la muestra actual de CR con una selección aleatoria de la población humana natural de CR y estimar si las diferencias son significativas para los dos grupos de poblaciones. De ser así, esto implicaría que la muestra relacionada con la esquizofrenia es más homogénea, ya sea por un efecto de muestreo o un efecto biológico, que implique a un número reducido de linajes o haplotipos en relación con la historia de la enfermedad.

Es muy particular, la evidencia que aportan las diferencias entre individuos para las tres poblaciones estudiadas. Estas varían proporcionalmente y las distribuciones de diferencias confirman la similitud de las mismas, pues el comportamiento de las tres poblaciones es muy parecido implicando a su vez eventos pasados compartidos en los linajes maternos de las poblaciones relacionadas con la esquizofrenia. La presencia de bimodalidad en la distribución contrasta con el índice de desigualdad de Harpending, señalando que las tres poblaciones sufrieron una expansión demográfica reciente (Rogers y Harpending, 1992).

Recapitulando, la caracterización interna de las poblaciones reveló que son altamente diversas para los dos sistemas de marcadores genéticos estudiados. Además, se observó que son muy similares a las poblaciones naturales reportadas para los mismos marcadores, que deberá ser corroborado posteriormente con análisis estadísticos apropiados. En respuesta a la suposición, se observa que las tres poblaciones hispanas de

CR, MX y el suroeste de EUA relacionadas con la esquizofrenia son muy similares entre sí, confirmado con el resto de estimaciones a partir de los datos.

### **B. Análisis de la estructura jerárquica de las poblaciones de Costa Rica, México y el Suroeste de EUA relacionadas con la esquizofrenia**

En la actualidad se sabe que la mayoría de las poblaciones mundiales son producto de una mezcla por movimientos de poblaciones vecinas o lejanas, ya sea por eventos naturales como las heladas de la prehistoria, impulsos económicos y de dominio como lo ejemplifica el período de la colonia, o por comportamiento de los individuos que incluye matrimonios, oferta de trabajo, demanda laboral, situación política del país entre otros factores (CEPAL, 1999). Si bien esto fue y sigue siendo generalizado, no todos los grupos mundiales han sido absorbidos por la diversidad, ejemplo de esto son los amerindios y las tribus africanas muy aisladas y apartadas del entorno, conservando altamente su contenido genético ancestral, siendo diferenciables de los grupos mezclados.

Este efecto se puede estudiar por el Análisis Molecular de Varianza (AMOVA), que evalúa la diferenciación o la estructuración jerárquica genética de las poblaciones. En general, para las poblaciones mundiales se ha evidenciado estadísticamente que la diferenciación se debe a variaciones internas más que a diferencias estrictas entre las poblaciones, en proporciones menores al 15% de diferenciación entre poblaciones para los *loci* neutros, debido a que los humanos son muy homogéneos y sus diferencias genéticas son triviales (Cavalli-Sforza, 1998; Harpending y Rogers, 2000; Relethford, 2001). Sin embargo, cuando las diferencias entre las poblaciones son altas se debe a eventos determinados de los grupos estudiados, por distancias geográficas que aíslan a los individuos de las poblaciones siendo esto un ejemplo de deriva genética, aunque este efecto es muy reducido en la actualidad por los movimientos acelerados de las poblaciones humanas.

El presente caso, es un ejemplo más de que las diferencias entre las poblaciones son muy reducidas. Particularmente para el ADNmt, se demostró que CR, MX y EUA son altamente diferenciables internamente evidenciado en las elevadas diversidades haplotípicas, pero no entre ellas por efecto de la migración y los linajes ancestrales en común. Estas tres poblaciones son descendientes de cinco haplogrupos mitocondriales principalmente, de carácter amerindio (A, B, C, D y X; Bianchi *et al.*, 1998), que aún en la actualidad guardan su identidad. Al menos en la población mestiza mexicana se comprobó la presencia de estos cuatro haplogrupos que equivale a un 90% de ascendencia amerindia, además se encontraron otros de ascendencia europea y africana, pero en proporciones muy reducidas (Green *et al.*, 2000). En 59 mestizos del Valle Central de Costa Rica se estimó la mayor proporción del haplogrupo A (63%), seguido por el B (33%) y en menor frecuencia el C (4%), lo que equivale a un total de 83% de trasfondo amerindio en esta población (Carvajal-Carmona *et al.*, 2003). Específicamente en las poblaciones analizadas en el presente estudio se reveló que estas proporciones se conservan, siendo predominante el haplogrupo A y el B para las tres poblaciones. Todo esto es razón que confirma, adicionalmente, las relaciones entre las poblaciones en cuanto a su linaje materno.

Una situación similar refleja el estudio del cromosoma Y, sin embargo, este lo que evidenciará en mayor proporción es la mezcla que originaron los colonizadores europeos. Respecto a la Base de Datos de Referencia para los Haplotipos del Cromosoma Y para Europa fundamentada en el estudio de 8592 individuos, se observó que para cuatro marcadores las tres poblaciones conservaban el mismo alelo predominante al europeo (DYS19, DYS390, DYS393 y DYS389I). Para el DYS391 solo CR y MX compartieron con Europa el alelo más representativo, para el DYS392 solo CR para el más frecuente y MX para el segundo más frecuente. Y para el DYS389II solo EUA comparte con Europa el alelo 29 como el más frecuente, aunque CR y MX su alelo más frecuente es el segundo en Europa. Sin embargo, cuando el análisis se limita solo a la información de los países de España y Portugal, que estuvieron directamente involucrados en la colonia, los resultados son más interesantes pues evidencian identidad en el análisis de 6 Y-STR (Carracedo *et al.*, 2001; Gusmão *et al.*, 2003b), mostrando

indicios de relaciones históricas. También se debe tomar en cuenta el componente amerindio, identificado por el alelo ah II y el 13 del DYS19 que son predominantes en poblaciones de nativos americanos (Bianchi *et al.*, 1998). Este último marcador, analizado en el presente estudio no se observó en altas proporciones en las poblaciones mestizas, sugiriendo asimismo una menor proporción de linajes amerindios en estas poblaciones, efecto que no podría deberse necesariamente a las mutaciones pues su tasa de mutación es baja comparado con otros Y-STR (Kayser *et al.*, 2000). En este sentido, no se debe olvidar el componente negro en la mezcla de estas poblaciones, siendo predominantemente masculina su influencia por las migraciones hacia estas poblaciones, sin embargo, la falta de información disponible específica de estas poblaciones dificulta la confirmación de estas suposiciones.

Por otra parte, el efecto de similitud entre las poblaciones a partir de los Y-STR puede estar también influido por los factores de migración y mutación, antes explicados para las medidas de diversidad.

Es interesante observar que aunque el AMOVA para los Y-STR es significativo, ya sea para las tres poblaciones o las seis regiones, el porcentaje de diferenciación entre poblaciones es muy reducido para el presente estudio (4-4,5%), por lo que es un dato razonable para discutir, pues la evidencia biológica indica que hay relación entre ellas. Y esto es indicio de que las variantes del cromosoma Y están más localizadas geográficamente que el ADNmt debido en parte a que la migración de los hombres es más limitada que las mujeres (Cavalli-Sforza, 1998). Por esto se separó la variación por cada *locus* estudiado para así hallar señales de marcadores específicos que estuvieran haciendo la diferencia. Así es como se observó que tres microsatélites (DYS392, DYS439 y DYS389II) estudiados fueron determinantes en este aspecto, y esto se relaciona directamente con las frecuencias de los alelos más frecuentes, donde en el DYS392 solo EUA y MX comparten el alelo pero con frecuencias diferentes hasta un 20,7%; en el caso del DYS439 aunque sí comparten el alelo más frecuente la diferencia en frecuencias llega hasta un 21,3%; además EUA presenta también el alelo 11 con igual frecuencia. Por último, en el DYS389II solo CR y MX comparten el alelo más frecuente pero con una diferencia de 11,7%; además de la diferencia en el número de alelos. El

efecto de este último marcador se puede deber a la alta tasa de mutación ( $4,71 \times 10^{-3}$ ; Kayser *et al.*, 2000) favoreciendo en este la presencia de diferentes proporciones de los alelos respecto a las poblaciones originales de la región.

De igual forma a las medidas de diversidad analizadas anteriormente para los Y-STR y el ADNmt, el AMOVA confirma la semejanza de las poblaciones de CR, MX y EUA relacionadas con la esquizofrenia, ya que no presentan estructuración jerárquica en su conformación.

### **C. Relaciones entre las poblaciones de Costa Rica, México y el Suroeste de EUA relacionadas con la esquizofrenia**

Esta es una de las principales evidencias de relación entre las poblaciones, ya que los haplotipos compartidos son producto de migraciones, historias compartidas o la eventualidad de mutaciones paralelas (identidad por estado), provocando la identidad de los haplotipos no debido a la ancestría común (identidad por descendencia) y, por lo tanto sin relación hereditaria directa.

En el caso del cromosoma Y, la presencia de más haplotipos compartidos entre MX y EUA es parte de la evidencia del conocido comportamiento migratorio de estas poblaciones, favorecido por las distancias geográficas reducidas entre las regiones del presente estudio. Por ejemplo, dos de los haplotipos compartidos entre MX y EUA son específicamente de Monterrey y San Antonio, regiones muy cercanas geográficamente (ver figura 3).

El no encontrar más haplotipos compartidos entre las poblaciones, a pesar de que se conoce la relación entre los grupos, se puede deber al efecto de las mutaciones de los microsatélites, pues con un solo paso mutacional se pueden alterar los haplotipos ancestrales, evidenciando diferencias no fundamentadas en el comportamiento de las sociedades, sino en las propiedades biológicas de los sistemas genéticos susceptibles a errores. Si se analizaran los marcadores más estables, es decir con baja tasa de mutación, pero con alta variabilidad, se seguirían observando diversidades haplotípicas altas que

resaltan la gran variabilidad interna de las poblaciones, pero se lograría determinar una mayor proporción de haplotipos compartidos por descendencia, ya que los alelos heredados serían susceptibles a mutaciones pero en menor grado, permitiendo la conservación de la información genética de los linajes masculinos. Esta sería una prueba contundente de la relación de las poblaciones. Una forma de realizar esto es analizando los polimorfismos de un nucleótido (SNP) o también llamados bialélicos, cuya tasa de mutación es menor que la de los microsatélites evitando las mutaciones reversas y convergentes, además algunos de estos polimorfismos evidencian los haplogrupos de origen de las poblaciones estudiadas, implicando la identidad por descendencia (Carvajal-Carmona *et al.*, 2000; Tishkoff y Verrelli, 2003), de modo que si se logra identificar el origen predominante de las muestras relacionadas con la esquizofrenia tal vez se pueda rastrear también el trasfondo poblacional más influyente en la enfermedad. Sin embargo, en el presente estudio a pesar de la alta variabilidad y las tasas de mutación de los microsatélites, se hallaron evidencias de eventos que relacionan a las tres poblaciones.

Otra razón por la que no se hallaron más semejanzas puede ser la composición genética de las poblaciones. Siendo mestizas su contenido genético es una amalgama constituida por los componentes amerindio, negro y europeo, y en EUA también asiático (Kayser *et al.*, 2003). Las tres poblaciones comparten un contenido amerindio muy similar proveniente de Asia, aunque la reducción drástica de las poblaciones en la época de la colonia pudo tener efectos de deriva y haber influido en la presencia y proporción de ciertos haplotipos que quizás en algún momento evidenciaban claramente su relación. Por otro lado, el contenido negro proveniente principalmente del Caribe, África Occidental y Central en diferentes tiempos de la historia aporta mayor variabilidad al ser poblaciones más diversas, pero el contenido en las poblaciones de estudio resulta en proporciones muy similares. Y el contenido europeo es en general proveniente de España y Portugal, siendo en mayor proporción hombres de estos países los que entraron a territorio americano, en diferentes tiempos durante la colonia y a diferentes regiones del territorio, ejerciendo cierta diferenciación. Es necesario resaltar que las personas de EUA analizadas en el presente estudio están fuertemente relacionadas con los

mexicanos, debido a que estas regiones fueron parte de México hasta mediados del siglo XIX y por las migraciones recientes, por esto el componente genético de las mismas es muy similar. Si bien todas estas variantes históricas ejercen cierta distinción en las poblaciones mestizas estudiadas, el efecto no es tan relevante como lo evidencia el análisis de varianza molecular.

Por otro lado, el haplotipo compartido entre CR y EUA, específicamente proviene de San Diego (California), siendo una situación particular pues no son vecinos geográficos. No obstante, se da la posibilidad que dicho haplotipo corresponda a un linaje antiguo conservado entre las dos poblaciones hasta la actualidad, que sea un efecto de mutaciones convergentes o un evento migratorio reciente (CEPAL, 1999). Pues ha sido reportado que las migraciones a nivel de América fueron muy comunes a partir de los años 70 hasta los 90 en que se intensificaron, especialmente de Centroamérica hacia EUA, y de México respecto al suroeste de EUA (CEPAL, 1999).

Aún más relevante en este sentido es la información emanada del ADNmt, ya que sí se presentan haplotipos compartidos entre las tres poblaciones, en este caso puede deberse a migraciones, mutaciones paralelas o linajes maternos históricos, que podría comprobarse analizando los haplogrupos amerindios del ADNmt en estas mismas poblaciones. Es por esto que se determinó solo con los polimorfismos de la región HVI (Green *et al.*, 2000; Santos *et al.*, 1994) la distribución de estos haplogrupos en las tres poblaciones de estudio aportando información valiosa. El haplotipo HM4 corresponde al haplogrupo A, el más representado en las tres poblaciones del presente estudio (CR 54%, EUA 52,4% y MX 54,5%) y el haplotipo HM10 corresponde al haplogrupo B, el segundo más representado (CR 27,5%, EUA 14,3% y MX 23,6%). Las proporciones de los haplogrupos obtenidas son representativas de las poblaciones mestizas de México (Green *et al.*, 2000) y amerindias de Costa Rica (Santos *et al.*, 1994). Esto es indicio de que MX y CR presentan un linaje amerindio muy similar en proporciones, inclusive en la población relacionada con la esquizofrenia.

Al igual que con los haplotipos del Y, en el ADNmt se presentan más haplotipos compartidos entre MX y EUA, que entre CR y MX. Una posible explicación de esta situación, además del efecto de los linajes amerindios, es la distancia geográfica que

puede favorecer las migraciones que culminaron en el asentamiento de los individuos y la posterior expansión de los haplotipos (CEPAL, 1999). En este caso las migraciones serían de mujeres, en algunos casos hacia los lugares de residencia de sus esposos, como se ha estimado previamente para ciertas poblaciones (Pérez-Lezaun *et al.*, 1997), aunque desde antes de 1970 hasta 1980 la predominancia de las migraciones de México y Centroamérica hacia EUA fueron de mujeres por razones como la segmentación laboral que favorecía el empleo de mujeres, la reunificación familiar y los desplazamientos forzados (principalmente de grupos familiares), teniendo como efecto la disminución de las diferencias entre las poblaciones y la reducción de las distancias genéticas (CEPAL, 1999).

En este sentido, también las mutaciones paralelas, por la alta tasa de mutación de la región control del ADNmt (Tully *et al.*, 2001), puede ser un factor clave en la similitud o diferencia entre las poblaciones, por lo que no debe obviarse su posible influencia en las distribuciones de los haplotipos. Esto no puede ser comprobado con los análisis realizados.

El análisis poblacional a partir de distancias genéticas es indispensable cuando se analizan relaciones de grupos cercanos o relativamente distantes geográficamente, para observar afinidades y diferencias analizadas bajo un contexto evolutivo-filogenético. El presente caso es representativo de esta situación. Aunque las distancias genéticas no son significativas estadísticamente, sí se observa un efecto de diferenciación. Cuando se desglosa la información en las regiones estudiadas, las distancias tampoco son significativas para ninguno de los marcadores genéticos estudiados. Sin embargo, se pueden analizar ciertos detalles mostrados más claramente en las representaciones gráficas de las distancias genéticas.

Es necesario especificar que ambos métodos, los árboles filogenéticos y el escalamiento multidimensional, muestran resultados muy similares. No obstante la suposición de los árboles es que las poblaciones que se analizan son producto de fisiones, mientras que el escalamiento multidimensional es una mejor estrategia para observar el comportamiento de poblaciones cercanas geográficamente (Cavalli-Sforza y Feldman, 2003), como es el presente estudio. Por esto se prefiere el estudio de los

resultados a partir del escalamiento multidimensional, donde dos poblaciones que se encuentren cercanas implicarían que son genéticamente similares (Harpending y Rogers, 2000; Schiffman *et al.*, 1981).

Sobre los Y-STR se observa la formación de un agrupamiento de las regiones de San Antonio, Monterrey, Los Ángeles y Costa Rica, que se encuentran apartadas por completo de Ciudad de México y San Diego. Es evidente que la relación de San Antonio y Monterrey sea directa, pues comparten haplotipos más que otras regiones en el presente estudio, y como se ha explicado se debe a efectos migratorios naturales.

El efecto de San Diego se debe a la reducida muestra analizada (sólo 6 individuos), que repercute directamente en las frecuencias alélicas, en la proporción de haplotipos encontrados y en los haplotipos compartidos con otras poblaciones. Aunque se asume es muy similar a su vecino, Los Ángeles, con este no presenta ningún haplotipo compartido, aunque sí se observa relación en el árbol NJ. Respecto a Costa Rica, San Diego comparte dos haplotipos, este efecto se puede deber al tamaño de la muestra o a una migración previa evidente en la actualidad, como se ha reportado en un análisis de las Naciones Unidas donde la población costarricense durante los años 1970 - 1990 mostró junto con el resto de países de Centroamérica un alto índice de migración hacia EUA, aproximadamente nueve millones de inmigrantes regionales de los cuales el 90% eran de México y Centroamérica (CEPAL, 1999). La proporción de individuos de CR analizados permite encontrar gran diversidad de haplotipos, como no se logra con las otras cinco regiones estudiadas, esto aumenta la probabilidad de encontrar haplotipos particulares, como los de San Diego.

El ADNmt reveló un patrón similar de agrupamiento en el MDS, mostrando la cercanía de LA, SA y CM que se observa también en el árbol NJ. Además SD se encuentra apartado, posiblemente por la misma razón anterior respecto al tamaño de la muestra que evidencia poca variabilidad. Por esto es poco confiable su localización según las distancias genéticas respecto a las demás regiones.

Resumiendo, respecto a las distancias genéticas y las representaciones gráficas para las seis regiones es necesario ser muy cauteloso con los resultados, debido a las

diferentes proporciones de los tamaños de las muestras en el análisis de los Y-STR y el ADNmt que causan sesgos o errores en la interpretación. Sin embargo, en gran parte la información revelada concuerda con lo conocido de las interacciones entre estas poblaciones y regiones de CR, EUA (hispanos del Suroeste) y MX, además las distancias genéticas mostraron diferencias pequeñas lo que implica que aunque se observen contrastes grandes entre las distribuciones espaciales de ciertas regiones, estas observaciones realmente no implican diferencias determinantes en la relación entre ellas. Lo que implica que no hay diferenciación entre estas poblaciones relacionadas con la esquizofrenia, que pueden ser estudiadas como un grupo en los posteriores análisis genéticos para aportar mayor robustez a los resultados y conclusiones que se obtengan.

## CONCLUSIONES

A partir del estudio previo de los datos y el análisis de la información se pueden obtener un conjunto de conclusiones y sugerencias útiles para posteriores evaluaciones.

1. Las medidas de diversidad estudiadas en el cromosoma Y reflejan una alta variabilidad dentro de las muestras, ya sea al analizar cada uno de los 12 marcadores o como haplotipos. La comparación de estos datos entre las poblaciones hispanas de Costa Rica, México y el Suroeste de Estados Unidos reflejan similitudes entre ellas, siendo esto indicio de una relación genética debido a una historia similar de colonizaciones europeas, migraciones africanas y componentes amerindios.
2. Por otro lado, las medidas de diversidad analizadas a partir de las secuencias de la región hipervariable I del ADNmt para cada una de las poblaciones de estudio, muestran nuevamente similitudes relevantes, correspondientes a polimorfismos comunes en proporciones similares y un comportamiento semejante en la distribución de diferencias. Sin embargo, la relación a partir de estos datos es más relevante entre México y Estados Unidos, favorecido por su cercanía geográfica.
3. La presencia de dos haplotipos compartidos entre las tres poblaciones para el ADNmt son evidencia de una historia compartida por los linajes maternos ancestrales, así como eventos de migración reciente. Es importante tomar en cuenta que las mutaciones paralelas son posibles y podrían incrementar o disminuir la similitud entre las poblaciones. La relación es más intensa entre México y Estados Unidos, pues presentan más haplotipos en común.
4. El caso del cromosoma Y fue más restrictivo respecto a los haplotipos compartidos, pues no se presenta ninguno entre las tres poblaciones, pero sí hay relación entre México y Estados Unidos, específicamente entre Monterrey y San Antonio, esto debido a su cercanía geográfica y a la relación política hasta mediados del siglo XIX.

5. Las poblaciones hispanas de Costa Rica, Estados Unidos y México relacionadas con la esquizofrenia no presentan diferenciación entre ellas en la región HVI del ADNmt pero sí para los Y-STR. Esto no es determinante conociendo el trasfondo biológico de las relaciones entre ellas y la información de las medidas de diversidad que confirman la similitud genética en los linajes masculinos y femeninos.
6. Igualmente, las distancias genéticas revelan que no hay diferencia significativa entre las poblaciones estudiadas ni cuando se separan por regiones, recalcando su semejanza genética por eventos históricos comunes.
7. Todo lo anterior confirma la similitud entre las poblaciones costarricense, mexicana y estadounidense (suroeste hispano), lo que brinda una justificación científica para agrupar a los individuos pertenecientes a dichas regiones en futuros estudios genéticos de ligamiento y asociación para obtener resultados más robustos, en este caso relacionado con una enfermedad psiquiátrica como la esquizofrenia.
8. Es recomendable hacer un estudio similar al presente comparando las poblaciones relacionadas con la esquizofrenia respecto a las poblaciones sanas y así determinar si hay algún indicio de la historia paterna o materna en particular que indique descendencia de un linaje específico. Aunque idealmente, la mejor estrategia sería la evaluación de las poblaciones relacionadas con la esquizofrenia respecto a casos control.
9. Además, es conveniente analizar los linajes amerindios y negros en las presentes poblaciones para confirmar la relación materna y la probabilidad de asociar algún linaje o haplotipo con la población relacionada con la esquizofrenia.

**REFERENCIAS**

Abecasis GR, Burt RA, Hall D, Bochum S, Doheny KF, Lundy SL, Torrington M, Roos JL, Gogos JA, Karayiorgou M. Genome scan in families with schizophrenia from the founder population of Afrikaners reveals evidence for linkage and uniparental disomy on chromosome 1. *Am J Hum Genet* 2004; 74: 403-417.

Abi-Dargham A, Rodenhiser J, Printz D, Zea-Ponce Y, Gil R, Kegeles LS, Weiss R, Cooper TB, Mann JJ, Van Heertum RL, Gorman JM, Laruelle M. Increased baseline occupancy of D<sub>2</sub> receptors by dopamine in schizophrenia. *Proc Natl Acad Sci USA* 2000; 97: 8104-8109.

Anderson S, Bankier AT, Barrel BG, Bruijn MHL, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJH, Staden R, Young IG. Sequence and organization of the human mitochondrial genome. *Nature* 1981; 290: 457-465.

Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* 1999; 23:147.

Bandelt HJ, Herrnstadt C, Yao YG, Kong QP, Kivisild T, Rengo C, Scozzari R, Richards M, Villems R, Macaulay V, Howell N, Torroni A, Zhang YP. Identification of native american founder mtDNAs through the analysis of complete mtDNA sequences: some caveats. *Ann Hum Genet* 2003; 67: 512-524.

- Beleza S, Alves C, González-Neira A, Lareu M, Amorim A, Carracedo A, Gusmão L. Extending STR markers in Y chromosome haplotypes. *Int J Legal Med* 2003; 117: 27-33.
- Bianchi NO, Catanesi CI, Bailliet G, Martínez-Marignac VL, Bravi CM, Vidal-Rioja LB, Herrera RJ, López-Camelo JS. Characterization of ancestral and derived Y-chromosome haplotypes of New World Native populations. *Am J Hum Genet* 1998; 63: 1862-1871.
- Bonato SL, Salzano FM. Diversity and age of the four major mtDNA haplogroups, and their implications for the peopling of the new world. *Am J Hum Genet* 1997 A; 61: 1413-1423.
- Bonato SL, Salzano FM. A single and early migration for the peopling of the Americas supported by mitochondrial DNA sequence data. *Proc Natl Acad Sci USA* 1997 B; 94: 1866-1871.
- Bonné-Tamir B, Korostishevsky M, Redd AJ, Pel-Or Y, Kaplan ME, Hammer MF. Maternal and paternal lineages of the Samaritan isolate: mutation rates and time to the most recent common male ancestor. *Ann Hum Genet* 2003; 67: 153-164.
- Bortolini MC, Salzano FM, Thomas MG, Stuart S, Nasanen SPK, Bau CHD, Hutz MH, Layrisse Z, Petlz-Erler ML, Tsuneto LT, Hill K, Hurtado AM, Castro-de-Guerra D, Torres MM, Groot H, Michalski R, Nymadawa P, Bedoya G, Bradman N, Labuda D, Ruiz-Linares A. Y-chromosome evidence for differing ancient demographic histories in the Americas. *Am J Hum Genet* 2003; 73: 524-539.
- Bosch E, Lee A, Calafell F, Arroyo E, Henneman P, de Knijff P, Jobling M. High resolution Y chromosome typing: 19 STRs amplified in three multiplex reactions. *Forensic Sci Int* 2002; 125: 42-51.

- Budowle B, Allard MW, Wilson MR, Chakraborty R. Forensic and mitochondrial DNA: applications, debates, and foundations. *Annu Rev Genomics Hum Genet* 2003; 4: 119-141.
- Carracedo A, Beckmann A, Bengs A, Brinkman B, Caglia A, Capelli C, Gill P, Gusmão L, Hagelberg G, Hohoff C, Hoste B, Kihlgren A, Kloosterman A, Dupuy BM, Morling N, O'donnell G, Parson W, Phillips C, Pouwells M, Scheithauer R, Schmitter H, Schneider P, Schumm J, Skitsa I, Stradmann-Bellinghausen B, Stuart M, Syndercombe D, Vide C. Results of a collaborative study of the EDNAP group regarding the reproducibility and robustness of the Y-chromosome STRs DYS19, DYS389 I and II, DYS390 and DYS393 in a PCR pentaplex format. *Forensic Sci Int* 2001; 119: 28-41.
- Carvajal-Carmona LG, Ophoff R, Service S, Hartiala J, Molina J, León P, Ospina J, Bedoya G, Freimer N, Ruiz-Linares A. Genetic demography of Antioquia (Colombia) and the Central Valley of Costa Rica. *Hum Genet* 2003; 112: 534-541.
- Carvajal-Carmona LG, Soto ID, Pineda N, Ortíz-Barrientos D, Duque C, Ospina-Duque J, McCarthy M, Montoya P, Alvarez VM, Bedoya G, Ruiz-Linares A. Strong Amerind/White sex bias and possible Sephardic contribution among founders of a population in Northwest Colombia. *Am J Hum Genet* 2000; 67:1287-1295.
- Cavalli-Sforza LL, Feldman MW. The application of molecular genetic approaches to the study of human evolution. *Nat Genet supplement* 2003; 33: 266-274.
- Cavalli-Sforza LL. The DNA revolution in population genetics. *Trends Genet* 1998; 14: 60-65.

CEPAL (Comisión Económica para América Latina y el Caribe). Migración y desarrollo en América del Norte y Centroamérica: una visión sintética. Santiago, Chile: Naciones Unidas; 1999. 51 p.

Chinnery PF, Thorburn DR, Samuels DC, White SL, Dahl HHM, Turnbull DM, Lightowlers RN, Howell N. The inheritance of mitochondrial DNA heteroplasmy: random drift, selection or both? *Trends Genet* 2000; 16: 500-505.

Collins-Schramm HE, Chima B, Morii T, Wah K, Figueroa Y, Criswell LA, Hanson RL, Knowler WC, Silva G, Belmont JW, Seldin MF. Mexican American ancestry - informative markers: examination of population structure and marker characteristics in European Americans, Mexican Americans, Amerindians and Asians. *Hum Genet* 2004; 114: 263-271.

Comas D, Calafell F, Mateu E, Pérez-Lezaun A, Bosch E, Martínez-Arias R, Clarimon J, Facchini F, Fiori G, Luiselli D, Pettener D, Bertranpetit J. Trading genes along the Silk Road: mtDNA sequences and the origin of Central Asian populations. *Am J Hum Genet* 1998; 63: 1824-1838.

Cordaux R, Saha N, Bentley GR, Aunger R, Sirajuddin SM, Stoneking M. Mitochondrial DNA analysis reveals diverse histories of tribal populations from India. *Eur J Hum Genet* 2003; 11: 253-264.

de Knijff P, Kayser M, Corach D, Fretwell N, Gehrig C, Graziosi G, Heidorn F, Herrman S, Herzog B, Hidding M, Honda K, Jobling M, Krawczak M, Leim K, Meuser S, Meyer E, Oesterreich W, Pandya A, Parson W, Penacino G, Perez-Lezaun A, Piccinini A, Prinz M, Schmitt C, Schneider P, Szibor R, Teifel-Greding J, Weichhold G, Roewer L. Chromosome Y microsatellites: population genetic and evolutionary aspects. *Intl J Legal Med* 1997; 110: 134-140.

- De Lisi LE, Mesen A, Rodríguez C, Bertheau A, LaPrade B, Llach M, Riondet S, Razi K. Clinical characteristics of schizophrenia in multiply affected Spanish origin families from Costa Rica. *Psychiatr Genet* 2001; 11: 145-152.
- Deakin JFW, Simpson MDC. A two-process theory of schizophrenia: evidence from studies in post-mortem brain. *J Psychiat Res* 1997; 31: 277-295.
- Deka R, Jin L, Shriver MD, Yu LM, Saha N, Barrantes R, Chakraborty R, Ferrel RE. Dispersion of human Y chromosome haplotypes based on five microsatellites in global populations. *Genome Res* 1996; 6: 1177-1184.
- Derenko MV, Grzybowski T, Malyarchuk BA, Dambueva IK, Denisova GA, Czarny J, Dorzhu CM, Kakpakov VT, Miścicka-Śliwka D, Woźniak M, Zakharov IA. Diversity of mitochondrial DNA lineages in South Siberia. *Ann Hum Genet* 2003; 67: 391-411.
- Di Rienzo A, Wilson AC. Branching pattern in the evolutionary tree of human mitochondrial DNA. *Proc Natl Acad Sci USA* 1991; 88: 1597-1601.
- Encarta Enciclopedia. Microsoft Corporation; 2004.
- Escamilla M, Ontiveros A, Nicolini H, Raventós H, Mendoza R, Delgado A, Esparza E, Montero P, Camacho A, Munoz R, Dassori AM, Armas R, Medina RA, Contreras SA, Cano E. Clinical characteristics of subjects from a Multi-site genetic study of schizophrenia (SC) in Latin American families. *Am J Med Genet* 2002; 114: 846-847.
- Escamilla M, Ontiveros A, Nicolini H, Raventós H, Mendoza R, Medina R, Muñoz R, Levinson D, Peralta JM, Dassori A, Almasly L. A genome-wide scan for

schizophrenia susceptibility *loci* in families of Mexican and Central American ancestry. 2005; en preparación.

Excoffier L, Smouse PE, Quattro JM. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 1992; 131: 479-491.

Felsenstein J. PHYLIP (Phylogeny Inference Package) version 3.6. Department of Genome Sciences, University of Washington, Seattle. 2004.

Finnilä S, Lehtonen MS, Majamaa K. Phylogenetic network for European mtDNA. *Am J Hum Genet* 2001; 68: 1475-1484.

Gamero JJ, Romero JL, González JL, Carvalho M, Anjos MJ, Corte Real F, Vide MC. Y-Chromosome STR haplotypes in a southwest Spain population sample. *Forensic Sci Int* 2002; 125: 86-89.

Gené M, Borrego N, Xifró A, Piqué E, Moreno P. Haplotype frequencies of eight Y-chromosome STR *loci* in Barcelona (North-East Spain). *Int J Legal Med* 1999; 112: 403-405.

Gill P, Brenner C, Brinkmann B, Budowle B, Carracedo A, Jobling MA, de Knijff P, Kayser M, Krawczak M, Mayr WR, Morling N, Olaisen B, Pascali V, Prinz M, Roewer L, Schneider PM, Sajantila A, Tyler-Smith C. DNA Commission of the International Society of Forensic Genetics: recommendations on forensic analysis using Y-chromosome STRs. *Int J Legal Med* 2001; 114: 305-309.

González JR, Escamilla MA, Ontiveros A, Medina RA, Raventos H, Contreras SA, Nicolini H, Dassori AM, Mendoza R, Camacho A, Ramírez M, Cano E, Armas R,

Esparza E, Munoz R, Salazar R, Montero AP. Diagnostic interview for genetic studies (DIGS): factor structure of the delusion and hallucination items and degree of familial aggregation of the factor scores. *Am J Med Genet* 2004; 130B (1): 57.

González-Neira A, Elmoznino M, Lareu MV, Sánchez-Diz P, Gusmão L, Prinz M, Carracedo A. Sequence structure of 12 novel Y chromosome microsatellites and PCR amplification strategies. *Forensic Sci Int* 2001; 122: 19-26.

Green LD, Derr JN, Knight A. mtDNA affinities of the peoples of North-Central Mexico. *Am J Hum Genet* 2000; 66: 989-998.

Gusmão L, Brión M, González-Neira A. The human Y-chromosome: Male-specific polymorphisms and forensic genetics. En: Maciej J Bogusz, editor. *Handbook of Analytical Separations*. Netherland: Elsevier Science; 2000. p 721-733.

Gusmão L, Carracedo A. Y chromosome specific STRs. *Profiles in DNA (Promega)* 2003 A; Marzo: 3-6.

Gusmão L, Sánchez-Diz P, Alves C, Quintáns B, García-Poveda E, Geda H, Raimondi E, Silva de la Fuente SM, Conceição Vide, M, Whittle MR, Zarrabeitia MT, Carvalho M, Negreiros V, Prieto Solla L, Riancho JA, Campos-Sánchez R, Vieira-Silva C, Toscanini U, Amorim A, Carracedo A, GEP-ISFG. Results of the GEP-ISFG collaborative study on the Y-chromosome STRs GATA A10, GATA C4, GATA H4, DYS437, DYS438, DYS439, DYS460 and DYS461: population data. *Forensic Sci Int* 2003 B; 135: 150-157.

Hagelberg E. Recombination or mutation rate heterogeneity? Implications for mitochondrial Eve. *Trends Genet* 2002; 19: 84-90.

- Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser* 1999; 41: 95-98.
- Hamilton G, Stoneking M, Excoffier L. Molecular analysis reveals tighter social regulation of immigration in patrilocal populations than in matrilocal populations. *Proc Natl Acad Sci USA* 2005; 102: 7476-7480.
- Harpending H, Rogers A. Genetic perspectives on human origins and differentiation. *Annu Rev Genomics Hum Genet* 2000; 1: 361-385.
- Harpending HC. Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Hum Biol* 1994; 66: 591-600.
- Harpending HC, Sherry ST, Rogers AR, Stoneking M. Genetic structure of ancient human populations. *Curr Anthropol* 1993; 34: 483-496.
- Hey J, Machado CA. The study of structured populations - new hope for a difficult and divided science. *Nat Rev Genet* 2003; 4: 535-543.
- Hidding M, Schmitt C. Haplotype frequencies and population data of nine Y-chromosomal STR polymorphisms in a German and a Chinese population. *Forensic Sci Int* 2000; 113: 47-53.
- Jiménez Arce G. Variación del cromosoma Y en la población afrocostarricense de Limón, Costa Rica (Tesis de Licenciatura). Costa Rica: Universidad de Costa Rica; 2000.
- Jobling M.A, Pandya A, Tyler-Smith C. The Y chromosome in forensic analysis and paternity testing. *Int J Legal Med* 1997; 110: 118-124.

Jobling M.A, Tyler-Smith C. Fathers and sons: The Y chromosome and human evolution. *Trends Genet* 1995, 11: 449-456.

Kayser M, Brauer S, Schädlich H, Prinz M, Batzer MA, Zimmerman PA, Boatman BA, Stoneking M. Y chromosome STR haplotypes and the genetic structure of U.S. populations of African, European, and Hispanic ancestry. *Genome Res* 2003; 13: 624-634.

Kayser M, Brauer S, Willuweit S, Schädlich H, Batzer MA, Zawacki J, Prinz M, Roewer L, Stoneking M. Online Y-chromosomal short tandem repeat reference database (YHRD) for U.S. populations. *J Forensic Sci* 2002; 47: 513-519.

Kayser M, Caglià A, Corach D, Fretwell N, Gehrig C, Graziosi G, Heidorn F, Herrmann S, Herzog B, Hiding M, Honda K, Jobling M, Krawczak M, Leim K, Meuser S, Meyer E, Oesterreich W, Pandya A, Parson W, Penacino G, Pérez-Lezaun A, Piccinini A, Prinz M, Schmitt C, Weichhold G, de Knijff P, Roewer L. Evaluation of Y-chromosomal STRs: A multicenter study. *Int J Legal Med* 1997; 110: 125-133.

Kayser M, Krawczak M, Excoffier L, Dieltjes P, Corach D, Pascali V, Gehrig C, Bernini L, Jespersen J, Bakker E, Roewer L, de Knijff P. An extensive analysis of Y-chromosomal microsatellite haplotypes in globally dispersed human populations. *Am J Hum Genet* 2001; 68: 990-1018.

Kayser M, Roewer L, Hedman M, Henke L, Henke J, Brauer S, Krüger C, Krawczak M, Nagy M, Dobosz T, Szibor R, de Knijff P, Stoneking M, Sajantila A. Characteristics and frequency of germline mutations at microsatellite *loci* from the human Y chromosome, as revealed by direct observation in father/son pairs. *Am J Hum Genet* 2000; 66: 1580-1588.

Kimura M. Evolutionary rate at the molecular level. *Nature* 1968; 217: 624-626.

Lell JT, Sukernik RI, Starikovskaya YB, Su B, Jin L, Schurr TG, Underhill PA, Wallace DC. The dual origin and Siberian affinities of Native American Y chromosomes. *Am J Hum Genet* 2002; 70: 192-206.

Lessig R, Edelman J, Krawczak M. Population genetics of Y-chromosomal microsatellites in Baltic males. *Forensic Sci Int* 2001; 118: 153-157.

Levinson DF, Levinson MD, Segurado R, Lewis CM. Genome scan meta-analysis of schizophrenia and bipolar disorder, part I: methods and power analysis. *Am J Hum Genet* 2003; 73: 17-33.

Lewis CM, Levinson DF, Wise LH, DeLisi LE, Straub RE, Hovatta I, Williams NM, Schwab SG, Pulver AE, Faraone SV, Brzustowicz LM, Kaufmann CA, Garver DL, Gurling HMD, Lindholm E, Coon H, Moises HW, Byerley W, Shaw SH, Mesen A, Sherrington R *et al.* Genome scan meta-analysis of schizophrenia and bipolar disorder, part II: schizophrenia. *Am J Hum Genet* 2003; 73: 34-48.

McLean DC, Spuill I, Gevao S, Morrison E, Bernard OS, Argyropoulos G, Garvey WT. Three novel mtDNA restriction site polymorphisms allow exploration of population affinities of African Americans. *Hum Biol* 2003; 75:147-161.

Medina R, Escamilla MA, Dassori AM, Contreras SA, Levinson D, Raventós H, Montero P, Balderas T, Ontiveros A, Nicolini H, Mendoza R, Camacho A, Delgado A. Cluster analysis of schizophrenia patients using lifetime dimensions of psychosis. *Am J Med Genet* 2002; 114: 859.

Medina RA, Dassori AM, Escamilla MA, Contreras SA, Armas R, Montero P, Raventos H, Ontiveros A, Nicolini H, Mendoza R, Munoz R, Jerez AA. Naturalistic heterogeneity of subjects diagnosed in DSM schizophrenia: factor-cluster analysis of lifetime dimensions of psychosis. *Am J Med Genet* 2003; 122B (1): 175-176.

- Meléndez EL. Polimorfismos del ADN mitocondrial en una muestra general de la población costarricense y su aplicabilidad en la identificación individual (Tesis de Maestría). Costa Rica: Universidad de Costa Rica; 2001.
- Melton T, Clifford S, Kayser M, Nasidze I, Batzer M, Stoneking M. Diversity and heterogeneity in mitochondrial DNA of North American populations. *J Forensic Sci* 2001; 46: 46-52.
- Merriwether DA, Huston S, Iyengar S, Hamman R, Norris JM, Shetterly SM, Kamboh MI, Ferrell RE. Mitochondrial versus nuclear admixture estimates demonstrate a past history of directional mating. *Am J Phys Anthropol* 1997; 102: 153-159.
- Miranda JJ, Benecke M, Hidding M, Schmitt C. Y-chromosomal short tandem repeat haplotypes at the *loci* DYS393, DYS19, DYS392, DYS385-I/II, DYS390, DYS389-I/II, and DYS391 in a Filipino population sample. *J Forensic Sci* 2001; 46: 1250-1253.
- Morera B, Barrantes R, Marín-Rojas R. Gene admixture in the Costa Rican population. *Ann Hum Genet* 2003; 67: 71-80.
- Nei M. *Molecular Evolutionary Genetics*. New York , USA: Columbia University Press; 1987. 512 p.
- Nei, M. Genetic distance between populations. *Am Nat* 1972; 106: 283-292.
- Nestler EJ. An emerging pathophysiology. *Nature* 1997; 385: 578-579.
- Pérez-Lezaun A, Calafell F, Comas D, Mateu E, Bosch E, Martínez-Arias R, Clarimón J, Fiori G, Luiselli D, Facchini F, Pettener D, Bertranpetit J. Sex-specific migration

- patterns in Central Asian populations, revealed by analysis of Y-chromosome short tandem repeats and mtDNA. *Am J Hum Genet* 1999; 65: 208-219.
- Presciuttini S, Caglià A, Alù M, Asmundo, Dusceci L, Caenazzo L, Carnevali E, Carra E, De Battisti Z, De Stefano F, Domenici R, Piccinini A, Resta N, Ricci, Pascali VL. Y-chromosome haplotypes in Italy: The GEFI collaborative database. *Forensic Sci Int* 2001; 122: 184-188.
- Previderè C, Tyler-Smith C, Grignani P, Peloso G. Y-chromosomal haplotypes and haplogroups in male identification. En: George F. Sensabaugh, Patrick J. Lincoln, Bjornar Olaisen, editores. *Progress in Forensic Genetics 8*. Netherland: Elsevier Science; 2000. p. 257-259.
- Quintana-Murci L, Krausz C, McElreavey K. The human Y chromosome: function, evolution and disease. *Forensic Sci Int* 2001; 113: 169-181.
- Rangel-Villalobos H, Jaloma-Cruz AR, Cerda-Aguilar L, Ríos Angulo CD, Mendoza-Carrera F, Patiño-García B, Sandoval-Ramírez L, Figuera-Villanueva LE. La huella genética del DNA en varones: haplotipos del cromosoma Y en una población mexicana analizando cinco STRs. *Rev Invest Clín* 2001 A; 53: 401-406.
- Rangel-Villalobos H, Jaloma-Cruz AR, Sandoval-Ramírez L, Velarde-Félix JS, Gallegos-Arreola MP, Figuera LE. Y-Chromosome Haplotypes for Six Short Tandem Repeats (STRs) in a Mexican population. *Arch Med Res* 2001 B; 32: 232-237.
- Relethford JH. Genetic history of the human species. En: Balding DJ, Bishop M, Cannings C, editores. *Handbook of Statistical Genetics*. England, West Sussex: John Wiley & Sons, Ltd; 2001. p. 56-59.

Roewer L, Kayser M, Dieltjes P, Nagy M, Bakker E, Krawczak M, de Knijff P. Analysis of molecular variance (AMOVA) of Y-chromosome specific microsatellites in two closely related human population. *Hum Mol Genet* 1996; 5: 1029-1033.

Roewer L, Krawczak M, Willuweit S, Nagy M, Alves C, Amorim A, Anslinger K, Augustin C, Betz A, Bosch E, Cagliá A, Carracedo A, Corach D, Dekairelle AF, Dobosz T, Dupuy BM, Fűredi S, Gehrig C, Gusmão L, Henke J, Henke L. Online reference database of European Y-chromosomal short tandem repeat (STR) haplotypes. *For Sci Int* 2001; 118: 106-13.

Rogers AR, Harpending H. Population growth makes waves in the distribution of pairwise genetic differences. *Mol Biol Evol* 1992; 9: 552-569.

Ruiz-Narvaez EA, Santos FR, Carvalho-Silva DR, Azofeifa J, Barrantes R, Pena SDJ. Genetic variation of the Y-chromosome in Chibcha-speaking Amerindians of Costa Rica and Panama. *Hum Biol* 2005; 77: 71-91.

Ruiz-Narvaez EA. Variación genética del cromosoma Y en las poblaciones amerindias de Costa Rica y Panamá (Tesis de Maestría). Costa Rica: Universidad de Costa Rica; 1998.

Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 1987; 4: 406-425.

Salas A, Richards M, Lareu MV, Scozzari R, Coppa A, Torroni A, Macaulay V, Carracedo A. The African Diaspora: mitochondrial DNA and the Atlantic slave trade. *Am J Hum Genet* 2004; 74: 454-465.

- Salazar R, Escamilla MA, Mendoza R, Esparza E, Montero AP, Ontiveros A, Nicolini H, Medina RA, Saucedo E, Herrera C, Lizano A, Contreras SA, Dassori AM. Use of the diagnostic interview for genetic studies (DIGS) in an International Genetics study of Latino subjects from the US, Mexico and Costa Rica: training and reliability issues. *Am J Med Genet* 2004; 130B (1): 131.
- Santos M, Ward RH, Barrantes R. mtDNA variation in the Chibcha Amerindian Huetar from Costa Rica. *Hum Biol* 1994; 66: 963-977.
- Scheffer IE. *Mitochondria*. USA: A Wiley-Liss publication; 1999. p. 326-341.
- Schneider S, Kueffer JM, Roessli D, Excoffier L. *Arlequin version 1.1*. Geneva. 1997.
- Schneider S, Roessli D, Excoffier L. *Arlequin version 2.000*. A Software for Population Genetics Data Analysis, University of Geneva, Geneva. 2000.
- Schiffman SS, Reynolds ML, Young FW. *Introduction to multidimensional scaling: theory, methods and applications*. Orlando Florida: Academic Press Inc.; 1981. 413 p.
- Seeman P, Kapur S. Schizophrenia: more dopamine, more D<sub>2</sub> receptors. *Proc Natl Acad Sci USA* 2000; 97: 7673-7675.
- Seielstad MT, Minch E, Cavalli-Sforza LL. Genetic evidence for a higher female migration rate in humans. *Nat Genet* 1998; 20: 278-280.
- Sherrat EJ, Thomas AW, Gill-Randall R, Alcolado JC. Phylogenetic analysis of mitochondrial DNA in type 2 diabetes. *Diabetes* 1999; 48: 628-634.
- StatSoft Inc. *STATISTICA for Windows [Computer program manual]*. Tulsa, OK: StatSoft, Inc; 1995.

- Tajima F. Evolutionary relationship of DNA sequences in finite populations. *Genetics* 1983; 105: 437-460.
- Tajima F. Measurement of DNA polymorphism. En: Takahata N, Clark AG, editores. *Mechanisms of Molecular Evolution. Introduction to Molecular Paleopopulation Biology*. Tokyo, Sunderland, MA: Japan Scientific Societies Press, Sinauer Associates, Inc; 1993. p. 37-59.
- Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol* 1993; 10: 512-526.
- Tishkoff SA, Verrelli BC. Patterns of human genetic diversity: implications for human evolutionary history and disease. *Annu Rev Genomics Hum Genet* 2003; 4: 293-340.
- Treeview (Win32) 1.6.6. Roderic D. M. Page, UK; 2001.
- Tully G, Brinkmann B, Carracedo A, Gill P, Morling N, Parson W, Schneider P. Considerations by the European DNA profiling (EDNAP) group on the working practices, nomenclature and interpretation of mitochondrial DNA profiles. *For Sci Int* 2001; 124: 83-91.
- Urraca N, Fresan A, Camarena B, Aguilar-Garcia A, Apiquian R, De la Fuente-Sandoval C, Meyenberg N, Nicolini H, Escamilla MA, Raventos H, Medina RA, Mendoza R, Munoz R, Ontiveros A, Canive J, Jerez A. Lack of association of the promoter region of the monoamine oxidase-A gene (MAO-UVNTR) with violent behaviour in schizophrenia. *Am J Med Genet* 2004; 130B (1): 80.

- Van Holst Pellekaan SM, Frommer M, Sved JA, Boettcher B. Mitochondrial control-region sequence variation in aboriginal australians. *Am J Hum Genet* 1998; 62: 435-449.
- Wallace DC, Brown MD, Lott MT. Mitochondrial DNA variation in human evolution and disease. *Gene* 1999; 238: 211-230.
- Wilder JA, Kingan SB, Mobasher Z, Pilkington MM, Hammer MF. Global patterns of human mitochondrial DNA and Y-chromosome structure are not influenced by higher migration rates of females versus males. *Nat Genet* 2004; 36: 1122-1124.
- Wu FC, Pu CE. Multiplex DNA typing of short tandem repeat loci on Y chromosome of Chinese population in Taiwan. *Forensic Sci Int* 2001; 120: 213-222.
- Zegura SL, Karafet TM, Zhivotovsky LA, Hammer MF. High-resolution SNPs and microsatellite haplotypes point to a single, recent entry of native american Y chromosomes into the Americas. *Mol Biol Evol* 2004; 21: 164-175.

**Referencias electrónicas:**

MITOMAP: A Human Mitochondrial Genome Database. <http://www.mitomap.org>

Y Chromosome Haplotype Reference Database (YHRD) for Europe,  
<http://www.yst.org/europe>

Y Chromosome Haplotype Reference Database (YHRD) for USA,  
<http://www.ystr.org/usa>

Short Tandem Repeat DNA Internet Data Base. <http://www.cstl.nist.gov/biotech/strbase/>)

# APÉNDICE

**Apéndice A.** Haplotipos encontrados en las tres poblaciones de estudio respecto a 12 Y-STR

Haplotipo	Origen *	DYS385 I	DYS385 II	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS437	DYS438	DYS439	DYS19
HY1	C	11	14	13	30	23	11	13	13	14	12	12	14
HY2	C	14	17	14	31	25	10	16	13	14	11	12	13
HY3	C	14	17	13	32	24	10	14	14	14	12	12	13
HY4	C	11	13	14	31	24	11	13	14	15	12	12	14
HY5	C	11	14	15	31	24	11	13	13	15	12	11	14
HY6	C	11	15	13	29	24	11	13	13	14	12	13	13
HY7	C	11	15	13	29	24	10	13	13	15	12	11	14
HY8	C	11	14	13	29	25	10	14	13	15	12	12	14
HY9	C	12	18	13	29	23	10	11	12	14	10	11	14
HY10	C	12	14	14	30	24	11	13	13	15	12	13	14
HY11	C	12	15	12	29	21	10	11	14	16	10	11	15
HY12	C	14	17	12	28	24	12	11	12	16	9	12	15
HY13	C	11	14	13	29	24	11	13	13	15	12	11	14
HY14	C	17	17	13	32	24	10	11	13	14	10	14	14
HY15	C	13	17	13	28	25	10	11	12	16	9	11	13
HY16	C	14	17	13	31	24	10	14	14	14	11	12	13
HY17	C	13	15	12	28	22	10	11	12	15	10	11	16
HY18	C	16	16	13	30	21	11	11	13	14	11	12	15
HY19	C	14	18	13	30	23	10	15	13	14	11	12	13
HY20	C	11	14	13	30	24	11	13	13	15	11	13	14
HY21	C	12	14	13	29	24	11	13	12	15	12	11	14
HY22	C	14	14	14	30	25	9	11	13	14	10	10	13
HY23	C	14	15	12	29	22	10	11	14	16	10	11	15
HY24	C	11	15	14	31	24	11	13	13	14	12	12	14
HY25	CE	11	14	14	30	24	11	13	13	15	12	11	14
HY26	C	14	14	12	28	22	10	11	13	16	10	11	16
HY27	C	11	14	13	29	24	10	13	13	15	12	14	14
HY28	C	13	14	12	29	23	10	11	13	16	10	12	14
HY29	C	13	13	14	30	22	11	11	14	16	10	12	15
HY30	C	16	17	13	32	24	10	11	13	14	10	14	14
HY31	C	17	17	13	31	21	10	11	13	14	11	11	15
HY32	C	11	15	13	29	23	11	13	13	15	12	12	14
HY33	C	11	14	12	28	24	11	14	13	14	12	12	14
HY34	C	17	18	13	31	24	10	11	13	14	10	11	14
HY35	C	14	18	14	30	23	10	13	13	14	9	12	14
HY36	C	12	14	13	29	24	10	13	13	15	12	12	14
HY37	C	12	12	13	28	23	10	11	13	15	10	11	17
HY38	C	14	15	12	28	22	11	11	14	16	11	11	16

## Continuación Apéndice A

Haplotype	Origen *	DYS385 I	DYS385 II	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS437	DYS438	DYS439	DYS19
HY39	C	11	14	13	29	24	11	13	14	15	12	12	14
HY40	C	13	15	13	30	25	10	14	12	14	11	12	13
HY41	C	11	14	14	30	24	11	13	13	15	12	12	14
HY42	C	11	15	13	29	24	12	13	13	15	12	11	14
HY43	C	15	18	12	29	21	10	11	14	14	11	13	15
HY44	C	11	13	13	29	24	11	13	13	15	12	13	16
HY45	C	14	17	14	31	25	10	15	13	14	11	12	14
HY46	C	15	15	13	30	23	10	15	13	14	11	13	13
HY47	C	16	17	14	33	21	10	11	13	14	11	12	15
HY48	C	15	15	13	29	24	11	15	12	14	11	12	13
HY49	C	12	12	13	28	23	10	9	13	15	10	11	17
HY50	C	14	16	13	30	23	10	11	12	15	9	13	14
HY51	C	18	18	13	30	21	10	11	15	14	11	11	17
HY52	C	13	14	12	28	22	10	11	12	15	10	11	16
HY53	C	16	17	13	31	23	10	14	13	14	12	12	13
HY54	C	13	18	14	30	24	11	13	14	14	10	12	15
HY55	C	14	15	12	28	22	10	13	13	14	10	12	14
HY56	C	10	15	14	30	24	10	13	13	15	12	12	14
HY57	C	11	14	13	29	25	10	13	13	15	12	12	14
HY58	C	15	15	12	29	22	10	10	14	16	10	12	16
HY59	C	14	17	13	31	24	10	15	13	14	11	12	13
HY60	C	11	15	12	29	24	10	13	13	14	12	12	14
HY61	C	11	14	14	29	24	11	13	12	15	12	13	14
HY62	C	11	11	13	29	25	10	13	13	15	11	12	13
HY63	C	16	18	13	30	21	10	11	13	14	11	12	15
HY64	C	15	16	12	31	22	10	10	14	16	10	12	15
HY65	C	14	16	14	31	24	10	15	13	14	11	12	13
HY66	C	16	18	13	31	21	10	11	13	14	11	12	15
HY67	C	11	14	13	29	24	10	14	13	15	12	12	14
HY68	C	11	14	13	29	24	10	13	13	15	12	13	14
HY69	C	13	17	13	31	23	11	14	14	14	11	12	13
HY70	C	17	18	13	31	21	10	11	13	14	11	11	15
HY71	C	15	17	14	31	23	10	14	12	15	11	13	13
HY72	C	11	14	13	30	23	12	13	13	15	12	13	15
HY73	C	14	16	13	31	25	10	15	13	14	11	12	13
HY74	C	14	14	12	28	22	10	12	14	16	11	11	16
HY75	C	13	17	12	29	23	10	11	12	14	10	11	14
HY76	C	14	14	12	28	22	10	11	14	16	11	11	16
HY77	C	11	11	14	30	23	10	13	13	15	12	13	14
HY78	C	11	15	14	30	24	10	13	13	14	12	13	14

## Continuación Apéndice A

Haplotipo	Origen *	DYS385 I	DYS385 II	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS437	DYS438	DYS439	DYS19
HY79	C	13	17	12	28	24	10	11	13	16	9	11	15
HY80	C	11	16	13	30	24	10	11	13	15	11	10	17
HY81	C	12	12	13	28	22	10	11	13	15	10	12	16
HY82	C	13	20	13	30	24	10	11	12	14	10	13	15
HY83	C	13	15	13	30	24	11	13	13	15	12	12	14
HY84	C	15	17	13	31	23	10	15	13	14	11	13	13
HY85	C	11	14	13	28	24	11	13	13	16	13	11	15
HY86	C	14	17	14	31	24	10	15	13	14	12	12	13
HY87	C	11	14	14	30	24	10	13	13	15	12	12	15
HY88	C	14	17	12	28	24	11	11	12	16	9	12	15
HY89	C	13	15	13	31	23	10	11	12	14	10	13	14
HY90	C	11	15	14	31	23	11	13	13	14	12	12	14
HY91	C	11	14	13	29	24	11	14	12	15	11	13	14
HY92	C	13	17	13	31	24	10	11	14	14	10	11	13
HY93	C	11	14	13	29	24	12	13	13	15	12	13	14
HY94	C	14	14	12	28	23	10	11	13	16	10	11	15
HY95	C	11	15	13	29	24	12	13	13	14	12	11	14
HY96	C	17	18	12	30	25	10	11	13	14	10	11	13
HY97	C	14	16	12	30	21	10	11	14	16	10	11	15
HY98	C	17	17	13	31	24	10	11	14	14	10	11	13
HY99	C	13	16	12	28	24	10	11	13	16	9	11	15
HY100	C	11	15	13	29	24	11	13	13	15	12	11	14
HY101	C	11	15	14	30	24	11	14	13	14	12	12	14
HY102	C	12	15	13	29	23	11	13	13	15	12	12	14
HY103	C	11	14	13	30	24	10	13	13	15	12	11	12
HY104	E	15	15	12	29	22	10	10	14	16	10	11	15
HY105	E	11	14	13	29	24	11	13	13	14	12	13	14
HY106	EM	11	14	13	29	25	11	13	13	14	12	12	14
HY107	E	11	14	12	28	24	11	13	13	15	12	12	14
HY108	E	11	14	14	31	24	11	13	13	15	12	12	14
HY109	E	11	12	13	29	23	11	13	13	14	12	13	15
HY110	E	15	16	13	31	24	10	16	13	14	11	11	13
HY111	E	17	18	13	30	25	9	12	13	14	10	12	13
HY112	E	14	17	12	29	24	10	13	13	14	11	12	13
HY113	E	11	14	13	28	24	11	13	13	15	12	12	14
HY114	E	12	12	13	27	22	10	11	12	15	10	12	16
HY115	E	13	15	13	29	24	9	11	13	14	10	10	13
HY116	E	13	14	12	28	24	9	11	13	14	10	10	13
HY117	E	14	16	13	29	23	11	11	12	15	9	12	14
HY118	E	14	15	13	29	23	10	15	13	14	11	11	13
HY119	EM	13	14	14	30	24	9	11	13	14	10	10	13

## Continuación Apéndice A

Haplotipo	Origen *	DYS385 I	DYS385 II	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS437	DYS438	DYS439	DYS19
HY120	E	11	14	13	29	23	10	13	13	15	12	12	14
HY121	E	14	16	13	31	23	10	13	13	14	9	11	15
HY122	E	13	14	12	28	22	10	11	13	16	10	12	14
HY123	E	11	13	13	29	24	11	14	14	15	11	12	13
HY124	E	11	14	13	30	24	11	13	13	15	12	11	14
HY125	EM	11	15	13	29	24	11	13	13	15	12	12	14
HY126	E	11	14	14	30	24	11	13	13	14	12	12	14
HY127	E	12	15	13	30	23	11	13	13	14	12	11	14
HY128	E	11	14	14	30	24	10	13	13	15	12	11	14
HY129	E	13	17	13	30	23	10	11	12	15	9	11	14
HY130	E	14	17	12	28	22	10	16	13	14	11	9	13
HY131	E	13	15	13	30	23	10	11	12	15	9	11	15
HY132	E	11	15	14	30	24	10	13	13	14	12	11	14
HY133	E	14	15	13	30	22	10	11	14	16	10	11	15
HY134	E	11	14	13	29	24	11	13	12	14	12	11	14
HY135	EM	11	14	14	30	24	10	13	13	14	12	12	14
HY136	E	11	14	12	29	24	11	11	13	15	12	12	14
HY137	E	14	18	12	29	24	11	13	13	14	12	11	13
HY138	E	14	16	13	30	24	10	14	13	14	11	12	13
HY139	E	13	19	13	31	24	10	16	13	14	11	11	14
HY140	E	11	14	13	29	23	11	13	13	14	12	13	14
HY141	E	10	14	13	29	24	10	13	12	15	12	11	14
HY142	E	11	14	13	29	24	11	13	12	15	12	12	14
HY143	M	11	14	13	29	24	11	13	13	15	12	12	14
HY144	M	11	14	13	29	25	11	14	13	14	12	12	14
HY145	M	11	14	13	30	24	10	13	13	15	12	12	14
HY146	M	11	14	13	29	25	11	14	12	15	12	14	14
HY147	M	16	17	13	30	23	10	16	13	14	11	12	13
HY148	M	12	15	13	30	23	9	11	13	14	10	12	13
HY149	M	17	17	13	30	24	10	14	13	14	10	12	15
HY150	M	16	17	13	29	23	11	14	13	14	10	12	14
HY151	M	11	14	14	30	24	11	13	13	14	12	11	14
HY152	M	15	18	13	30	24	11	16	13	14	11	12	17
HY153	M	15	19	13	30	24	12	14	13	14	11	12	13
HY154	M	12	12	14	29	23	10	11	13	15	10	14	16
HY155	M	11	14	12	27	24	10	14	12	14	12	11	14
HY156	M	16	17	13	31	23	10	11	13	14	10	12	13
HY157	M	11	15	12	28	24	11	14	13	16	12	13	14
HY158	M	11	15	13	29	25	11	13	13	15	12	12	14
HY159	M	11	13	13	29	24	11	13	13	15	12	12	14
HY160	M	14	19	13	31	24	11	16	12	14	11	12	13

## Continuación Apéndice A

Haplotipo	Origen *	DYS385 I	DYS385 II	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS437	DYS438	DYS439	DYS19
HY161	M	12	14	13	29	24	10	14	13	15	12	11	14
HY162	M	14	19	14	32	25	10	15	13	14	11	12	13
HY163	M	11	14	13	29	25	11	14	14	14	12	12	14
HY164	M	13	15	13	30	24	11	14	13	14	11	12	14
HY165	M	15	17	13	30	24	10	15	13	14	11	12	14
HY166	M	15	17	13	31	24	10	11	13	14	10	12	13
HY167	M	11	14	13	30	23	10	13	13	15	12	11	14
HY168	M	13	17	13	31	22	10	14	13	14	11	12	14
HY169	M	11	14	12	28	24	11	13	13	15	12	13	14
HY170	M	15	16	12	28	21	10	11	14	14	11	12	17
HY171	M	14	18	13	30	24	11	15	13	14	11	11	13
HY172	M	10	15	13	30	24	10	13	13	15	12	12	14
HY173	M	11	15	14	30	24	10	13	13	15	12	12	14
HY174	M	11	14	13	29	24	10	13	13	14	13	12	14
HY175	M	15	18	13	30	24	11	16	13	14	11	12	13
HY176	M	14	18	14	31	25	10	14	13	14	11	11	13
HY177	M	11	14	13	29	25	10	11	13	14	11	10	16
HY178	M	14	18	13	31	25	11	14	14	14	11	12	13
HY179	M	13	15	12	28	23	10	11	13	16	10	11	14
HY180	M	15	18	13	30	24	10	11	13	14	10	12	13
HY181	M	15	19	13	30	24	11	14	12	14	11	12	13
HY182	M	14	19	14	31	24	10	14	13	14	12	11	13
HY183	M	10	14	13	30	24	11	13	13	15	12	11	14
HY184	M	16	16	13	29	25	10	14	13	14	11	10	14
HY185	M	13	16	13	29	23	10	11	12	16	9	11	15
HY186	M	11	14	13	30	24	11	13	13	14	12	11	14
HY187	M	15	19	12	28	24	9	13	13	14	11	13	13
HY188	M	13	14	13	28	23	10	13	13	15	12	11	15
HY189	M	11	13	14	30	23	10	14	13	15	10	9	12

\* C: Costa Rica, E: Estados Unidos, M: México









Posición Variable	Secuencia Anderson	Origen *	HM70	HM71	HM72	HM73	HM74	HM75	HM76	HM77	HM78	HM79	HM80	HM81	HM82	HM83	HM84	HM85	HM86	HM87	HM88	HM89	HM90	HM91	HM92	HM93	HM94	HM95	HM96	HM97	HM98	HM99	HM100	HM101	HM102	HM103		
16047	A																																					
16065	C																																					
16082	T																																					
16087	A																																					
16088	T																																					
16089	T																																					
16098	T																																					
16100	C																																					
16101	T																																					
16107	C																																					
16110	C																																					
16120	T																																					
16122	T																																					
16125	G																																					
16127	T																																					
16130	C																																					
16132	T																																					
16141	G																																					
16144	C																																					
16146	C																																					
16151	A																																					
16164	C																																					
16165	C																																					
16168	T																																					
16170	C																																					
16171	A																																					
16172	C																																					
16173	A																																					
16178	A																																					
16179	A																																					
16181	C																																					
16182	C																																					
16183	C																																					
16184	C																																					
16185	T																																					
16201	C																																					
16204	G																																					
16209	T																																					
16214	C																																					
16217	T																																					
16222	C																																					
16223	C																																					
16224	T																																					
16233	A																																					



*Apéndice C.* Publicación sometida a la revista “Human Biology”

**Genetic structure analysis of three Hispanic populations from Costa Rica, Mexico and the Southwest United States, using Y-STRs markers and mtDNA sequence**

Rebeca Campos-Sánchez<sup>1</sup>, Ramiro Barrantes<sup>2</sup>, Sandra Silva<sup>1</sup>, Michael Escamilla<sup>3</sup>, Alfonso Ontiveros<sup>4</sup>, Humberto Nicolini<sup>5</sup>, Ricardo Mendoza<sup>6</sup>, Rodrigo Munoz<sup>7</sup>, Henriette Raventos<sup>1</sup>

<sup>1</sup>Universidad de Costa Rica, Centro de Investigación en Biología Celular y Molecular;

<sup>2</sup>Universidad de Costa Rica, Escuela de Biología; <sup>3</sup>Department of Psychiatry, University of Texas Health Science Center at San Antonio; <sup>4</sup>INFOSAME, Monterrey México;

<sup>5</sup>Grupo de Estudios Médicos y Familiares Carracci S.C., Ciudad de México; <sup>6</sup>David Geffen UCLA School of Medicine; <sup>7</sup>Family Health Centers of San Diego.

**Keywords:** genetic structure, Y-STR, mtDNA, Costa Rica, Mexico, United States of America

**Abstract**

A total of 217 male subjects from Costa Rica, Mexico and the Hispanic population of the Southwest United States of America were studied. 12 Y-STRs and the HVI sequence of the mtDNA were analyzed to describe their genetic structure and compare maternal and paternal lineages. All subjects are part of two NIMH funded studies to localize schizophrenia susceptibility genes in Hispanic populations of Mexican and Central American ancestry. We showed that these three populations are very similar when we compare their internal genetic characteristics, as revealed by analyses of mtDNA and Y-STR diversity. These populations are related through their maternal lineage in a stronger way than the paternal, as a higher number of shared haplotypes and polymorphisms is seen in the mtDNA (compared to Y-STRs). These results provide evidence of previous contact between these three populations and shared histories. An analysis of molecular variance (AMOVA) revealed no genetic differentiation for the mtDNA for the three populations but does for the Y-STRs. Genetic distance analysis showed that these three populations are closely related, probably due to migration between close neighbors, as indicated by shared haplotypes. Our results suggest that the populations studied are genetically related and can be used as a group in mapping and association studies of genetic illnesses found in the Hispanic people of Mexico, the Southwest United States and Costa Rica.

Population genetic studies can be used to compare relationships and differences between populations and to determine genetic structures of naturally occurring populations (Hey and Machado 2003). Molecular analyses of Y-STR markers and mtDNA haplotypes have been shown to be ideal in determining the geographic structure of human populations, because they act as neutral markers. Their fast evolutionary rate and small effective population size make them more sensitive to genetic differences among groups than autosomal markers. Moreover, comparing the patterns of variation from the Y-chromosome and the mtDNA allows one to understand the paternal and maternal history of the populations. This is especially useful for admixed populations (Kayser et al. 2003).

The origins of many of the populations from South, Central and North America including Costa Rica, Mexico and USA can be traced to ancestral migrations from Asia through the Bering Strait (the Amerindian heritage) (Collins-Schramm et al. 2004) and posterior migrations and admixture during the colonization period of the 15<sup>th</sup> – 16<sup>th</sup> century (Collins-Schramm et al. 2004; Green et al. 2000). Using microsatellite markers from the non-recombining region of the Y-chromosome and sequencing the HVI region of the mtDNA, we studied the genetic structure of three Hispanic populations from Mexico, the Southwest USA and Costa Rica, in order to characterize their genetic relationships. The sample utilized was drawn from two NIMH (National Institute of Mental Health) funded studies to map schizophrenia susceptibility genes (a multi-center NIMH funded initiative to map schizophrenia susceptibility genes in the Hispanic Population and a population based study of schizophrenia in Costa Rica, grants R01-MH61884 and RO1-MH60881).

This study reports the maternal and paternal relationships between these three groups, based on the analysis of 12 Y-STRs and HVI mtDNA sequences. Similarities between these populations have not been previously reported. Our results demonstrate that these three populations are clearly genetically related to each other, which allows them to be analyzed together in studies which tend to identify schizophrenia susceptibility genes. These results also suggest that these three populations can rationally be studied together in future genetic mapping and association studies of other complex inherited disorders.

## **Materials and Methods**

### **Population Samples**

All subjects were recruited as part of NIMH funded studies to map schizophrenia susceptibility genes, for which informed consent was obtained. 217 DNA samples from unrelated males from Costa Rica (CR 120 samples), Mexico (MX 55 samples from Mexico City and Monterrey) and the Southwest United States (USA 42 samples from the Hispanic populations of Los Angeles CA, San Diego CA and San Antonio TX) were studied. Subjects recruited from the United States had to have some family ancestry from Mexico or Central America. Some were probands (i.e. had a previous diagnosis of schizophrenia) and others were family members but none of the 217 subjects were related to each other. The geographical coordinates of the regions of study are: Central Valley of Costa Rica 9.45-10.15 N 83.45-84.30 W, San Antonio Texas 29.25 N 98.29 W, Los Angeles California 34.03 N 118.14 W, San Diego California 32.42 N 117.09 W, Monterrey 25.40 N 100.19 W, Mexico City 19.24 N 99.09 W.

### Y-STR Typing

Twelve microsatellites from the Y-chromosome were amplified by the PowerPlex Y-System kit (Promega) in a 12.5 ul reaction volume, as specified by the manufacturer. The Y-STRs studied were: DYS19, DYS385I/II, DYS389I/II, DYS390, DYS391, DYS392, DYS437, DYS438, and DYS439. Detection of the amplified fragments was done using the ABIPrism 377. PowerTyper™ Y Macro (Promega) was used to assign the alleles.

### Sequencing of the mtDNA

Hypervariable segment I of the mitochondrial DNA (HVI) was amplified in a PCR reaction volume of 50 ul using primers L15997 (5' CAC CAT TAG CAC CCA AAG CT) and H16401 (5' TGA TTT CAC GGA GGA TGG TG). The first amplification was done in a volume of 50 ul with: 4 ul DNA (2 ng/ul), 2 ul of each primer (5 uM), 5 ul of STR Buffer 10x (Promega), 3 ul Taq DNA Polymerase (5 U/ul) (Promega) and water. The amplification protocol used was: 95°C for 5 min; 35 cycles of 94°C for 45 sec, 65°C for 45 sec and 72°C for 90 sec; followed by 72°C for 7 min and a final hold at 4°C. To confirm amplification, the products were separated on a 2% agarose gel with TBE 1X buffer. PCR products were purified with Wizard PCR Preps kit (Promega), according to the manufacturer, and quantified through spectrophotometry. Double stranded sequencing was carried out by using BigDye® Terminator v3.1 Cycle Sequencing RR-100 (Applied Biosystems) as specified in the manual. The products were precipitated with EDTA and ethanol, and resuspended in formamide and blue dextran, before the

separation on the ABIPrism 377 sequencer. The Sequencing Analysis 3.3 software (Applied Biosystems) was used to register the data and obtain the electropherograms. The quality of the sequences was evaluated with Bioedit (Hall 1999) and then was aligned with the Cambridge Reference Sequence (Anderson et al. 1981) with Clustal. A total of 319 nucleotides from 16047 to 16365 were used for all the statistical analysis.

### Data Analysis

For the Y-STR data, allele and haplotype frequencies were estimated through gene counting. Gene and haplotype diversities were estimated according to Nei (1987). Polymorphism frequencies and nucleotide diversities were estimated for the mtDNA sequences. Also, mean pairwise differences for Y-STRs (based on the number of alleles) and for mtDNA (Tamura and Nei 1993) were studied. Shared haplotypes were used to determine relation between populations.

To analyze the hierarchy genetic structure of the three populations, an analysis of molecular variance (AMOVA) was performed among and within populations. It compares average gene frequencies among populations and differences (mutations) between alleles (Excoffier 2001). For the Y microsatellite data,  $R_{st}$  was estimated and for mtDNA,  $F_{st}$  was calculated. Also, genetic distances were estimated for the three populations according to Tamura and Nei (1993) for mtDNA and according to Nei (1972) for Y-STRs. This method analyzes the average genetic differences between local, regional and continental populations, by computing distances between pairs of populations and then analyzing them to determine how past evolutionary events could

explain the patterns of genetic distance observed in present day populations (Relethford 2001).

All mtDNA analysis was performed using a gamma A value (heterogeneity of mutation rates) of 0.26 and a transition-transversion ratio of 15. The statistical package Arlequin 2.000 (Schneider et al. 2000) and Arlequin version 1.1 (Schneider et al. 1997) were used for all calculations.

## Results

### Y-STR Variability

Table 1 shows the gene diversities for all 12 Y-STRs in CR, USA and MX. For 9 of the 12 markers, CR shows a slightly higher variability inside the population. When allele frequencies were compared between these populations, all three populations shared the most common allele for 9/12 markers (DYS385I/II 11/14, DYS389I 13, DYS390 24, DYS393 13, DYS437 14, DYS438 12, DYS439 12, and DYS19 14, data not shown).

The range of gene diversities was 0.798 (DYS385I) to 0.502 (DYS393) for CR, 0.714 (DYS385II) to 0.387 (DYS393) for USA, and 0.804 (DYS385II) to 0.286 (DYS393) in MX. On average, gene diversities were 0.675, 0.614, and 0.608 for CR, the USA and MX, respectively (table 2). The most informative marker for all three populations was DYS385I/II and the least informative was DYS393.

For the 12 Y-STRs, 103 haplotypes were obtained with a diversity of 99.7% for CR, 40 haplotypes with a diversity of 99.8% for the USA and 51 haplotypes with a diversity of 99.5% for MX (table 2). Although the populations showed high diversities, some shared haplotypes were found inside populations. 14 haplotypes were shared inside CR, 11 by 2 individuals and 3 by 3 individuals. Two haplotypes were shared by 2 individuals in USA, and in MX one haplotype was shared by 4 individuals and a second haplotype was shared by 2 individuals (data not shown). The mean number of pairwise differences between haplotypes was 8.099, 7.368 and 7.300, for CR, USA and MX, respectively (table 2).

## HVI Variability

The analysis of the mtDNA (16047 to 16365 nucleotide position) in the three populations showed 52, 39 and 56 polymorphic sites in CR, USA and MX, respectively (table 3), of which, 18, 10 and 19 were unique for the respective populations. Transitions were more common (48 in CR, 36 in USA and 53 in MX) than transversions (4 in CR and MX, 3 in USA). A total of 88 polymorphisms were present in the three populations together.

There were 18 polymorphisms shared among the three populations, and 4 more were shared between CR and USA, 12 between CR and MX; and 7 between USA and MX. There were also length variation polymorphisms in the nucleotide positions 16188.1C, 16192.1C and 16193.1C that were not introduced in the sequence analysis (Cordaux et al. 2003; Salas et al. 2004). Nucleotide diversity reveals heterozygosity at the nucleotide level, and was approximately 0.02 for all 3 populations, showing another similarity between them (table 3).

In the CR sample, 39 mtDNA haplotypes were found with a haplotype diversity of 93%, sharing a total of 17 haplotypes. The presence of 6 haplotypes shared by 20, 19, 11, 7, and 6 individuals is remarkable. This suggests homogeneity inside the sample. In the USA sample, 33 haplotypes were found for a diversity of 98%, and 4 of them were shared (2 shared by 4 individuals, 1 by 3 individuals and 1 by 2 individuals). In the MX sample, a total of 43 haplotypes were found, for a diversity of 98.9%, with 1 haplotype shared by 4 individuals, 2 haplotypes shared by 3 individuals and 5 haplotypes shared by 2 individuals (table 3). MX showed a mean pairwise difference of 8.445 nucleotides, CR of 7.657, and USA of 6.881 (Table 3).

### Shared Haplotypes between Populations

MX and USA share 4 Y-STR haplotypes, one of which is the most represented in MX (7.3%). Just one haplotype was shared between CR and USA, 3 and 2 times respectively. None of the haplotypes was shared between MX and CR or between all three populations (table 4) for the Y-STRs.

In contrast to the Y-STR data, there were 2 mtDNA haplotypes shared across all the three populations (CR, USA and MX). One of them (HM4) was represented 19 times in CR, 4 times in MX and 4 times in USA (table 5). There were also 2 additional haplotypes shared between subjects from CR and MX, and 6 additional haplotypes shared by individuals from the USA and MX.

### AMOVA

The analysis of the populations by an AMOVA for the Y-STR data reveals that 96% of the differences are the result of variations within populations ( $R_{st}$ : 0.04,  $P < 0.001$ ). This is also significant when the variation is separated per locus (table 6). For DYS392 ( $R_{st}$ : 0.051,  $P < 0.001$ ), the difference between populations is only 5.149%, and for DYS439, 97.8% of the variation is within populations with a probability of 0.038 (data not shown).

When the sample is subdivided in the six regions from CR, USA and MX, similar results are obtained (96.25% of the variation is within the samples,  $R_{st}$ : 0.038,  $P < 0.01$ , table 6). This is also significant when we separate the variation by locus ( $R_{st}$ : 0.019,  $P < 0.001$ ), for DYS392 ( $R_{st}$ : 0.044,  $P < 0.01$ ), DYS439 ( $R_{st}$ : 0.043,  $P < 0.01$ ) and DYS389II ( $R_{st}$ : 0.037,  $P < 0.01$ ).

For the mtDNA, the AMOVA did not show differences between the populations ( $F_{st}$ : 0.006,  $P>0.05$ ) or between the six regions ( $F_{st}$ : 0.005,  $P>0.304$ ).

### Genetic Distances

Analysis of genetic distances showed that the three populations are not distinguishable by either Y-STR data or mtDNA sequences (table 7), allowing us to infer a close relationship between them. By simple observation, CR is farther from USA, and MX is closer to USA, with an intermediate value between CR and MX for both types of data.

### Discussion

The Y-STR results confirmed the assumption that these populations are related because of shared history and migration (Melton et al. 2001), more so between MX and USA as geographic neighbors with a common history (the Southwest United States was part of Mexico until the mid-19<sup>th</sup> century). Frequent allele sharing is evidence of genetic relatedness (Bianchi et al. 1998), so the comparison of this data with previous studies would allow us to answer how representative the present samples are to the natural populations they belong. Two recent publications of Mexican mestizos show identity for six Y-STRs (Rangel-Villalobos et al. 2001a, Rangel-Villalobos et al. 2001b) with the current sample from MX. Previous studies for the admixed populations from CR and the USA with Y-STRs are insufficient to demonstrate the representativeness of the present sample from CR and USA, so more extensive research on this issue should be done.

The genetic similarity of these three populations could be the result of shared inheritance from Europe (Spain and Portugal) during the colonial time and the slave trade from Africa. This was evaluated by simple comparison of allele frequencies from Spain and Portugal (Carracedo et al. 2001, Gusmão et al. 2003), the main genetic European contributors to the gene pool of America. In 6 Y-STRs, these European populations share the most frequent allele with the current populations, showing an important relationship between these five populations in their paternal lineages. Gusmão et al. (2003) grouped CR mestizos with the Iberian rooted populations (including Spain and Portugal) in a neighbor joining tree by Y-STRs, showing evidence of this relation. Also, USA population shares four of the most frequent alleles with the Y-chromosome Reference Database for European and Hispanic Americans, that reports five Y-STR markers. Moreover, African inheritance is evident in the CR population that shares 4 frequent alleles with CR Africans (Gusmão et al. 2003; Jiménez 2000). This information shows evidence of the admixing of these populations.

Five of the eighteen shared polymorphisms in the mtDNA between CR, USA and MX have higher frequencies than 54.2% and are also present in Amerindian and mestizo populations from Mexico and Costa Rica (Meléndez 2001; Santos et al. 1994; Green et al. 2000). This may indicate part of the genetic admixed content of the studied populations and the similarity between their maternal lineages, also supported by their similar nucleotide diversity values. This data can be extrapolated to the USA Hispanic population because of shared history (Maza 1992).

The CR sample shows an over representation of some haplotypes (HM4, HM7 and HM13) which could be evidence of a more homogenous maternal lineage in the Costa

Rican samples. Nonetheless, further studies are needed to answer whether this result can be generalized to the Costa Rican population or is an effect of the selection of the sample for the presence of schizophrenia in a family member.

Shared haplotypes could be the result of relationship between populations due to migrations and/or shared backgrounds or parallel mutations (identity by state). The sharing of more Y-STRs haplotypes between Mexico and the Hispanic population from the Southwest USA is explained by known migration behaviors of these populations favored by reduced geographic distances and shared political history through the mid-nineteenth century (CEPAL 1999, Melton et al. 2001). A possible explanation for the lack of shared haplotypes across all the three populations may be secondary to a reduction of particular Y-chromosomes due to native population bottlenecks in the three populations. Another explanation is the distinctive waves of colonization that occurred during several periods throughout history, causing heterogeneity in different regions of America. A third explanation could be the admixture among Amerindians, Europeans and Africans inside each population, increasing heterogeneity. Lastly, mutations make founding haplotypes differ in their descendents, thus increasing diversity.

The shared haplotypes among these three populations for the mtDNA is relevant because it provides direct evidence of shared ancestral relationship between these groups. Again, this situation could be the result of migrations, parallel mutations and ancestral female lineages that can be tested by analyzing the presence and proportion of Amerindian haplogroups. In the Costa Rican population from the present study the proportion of Amerind haplogroups is 84% (Haplogroup A: 52.5%, Haplogroup B: 27.5% and Haplogroup C: 4.1%) as was previously reported (Carvajal-Carmona et al. 2003). For

the Mexican samples, 59% of mitochondrial haplogroups corresponded to Amerind haplogroups (Haplogroup A: 40%, Haplogroup B: 25.5% and Haplogroup C: 1.8%) which is much less than what has been seen in previous reports (Green et al. 2000). For the USA sample, 67.2% of the sample corresponds to Amerindian lineages (Haplogroup A: 33%, Haplogroup B: 14.3% and Haplogroup C: 11.9%). In proportion these percentages agree with the genetic history of the Americas, in which haplogroup A is most frequent among Amerinds. The current results, which show increased mtDNA contribution from Amerinds, show that directional mating occurred in all 3 of these populations. Specifically the haplotypes shared by the three populations are part of haplogroup A in the case of HM4 and haplogroup B for HM10, assigned by polymorphisms in the HVI region, which should be confirmed with polymorphisms in the coding region (Green et al. 2000). These two haplogroups are represented in mestizo Mexicans (Green et al. 2000) and Costa Rican Amerindians (Santos et al. 1994) in similar proportions to the present data, indicating a direct relationship between Costa Rica and Mexico through maternal Amerindian lineages.

Patrilocal behavior has been used to explain the reduction in mitochondrial differences between populations. This custom, in which the wife goes to live in the husband's location, causes a higher rate of migration in women than men (Cavalli-Sforza 1998; Pérez-Lezaun et al. 1999), and the spread of maternal mitochondrial haplotypes. Recently, however, it has been shown that this rate is not so different from men (Wilder et al. 2004) and depends on the social and mating system of the human population under study (Hamilton et al 2005).

The mutation rate of mitochondrial DNA is high, especially in the control region. Mutations can either falsely elevate relationships, not due to ancestors, or, conversely, can lead to differences between persons with the same ancestral mtDNA (Relethford 2001). This may influence the distribution of the haplotypes (Cavalli-Sforza 1998).

The three populations studied descend predominantly from 5 mitochondrial native Amerindian haplogroups (A, B, C, D and X; Bianchi et al. 1998), which maintain their identity to the present day. In the mestizo Mexican population 59% (current study) to 90% (Green et al. 2000) of the mtDNA is of Amerindian ancestry, compared to 83% (Carvajal-Carmona et al. 2003) to 84% (current study) in the Costa Rican population. This confirms the relationship of the populations through their maternal lineages. The Y-chromosome reflects the predominantly European background of these populations. The most common allele for 6 Y-STRs was shared between these three populations with Spain and Portugal (Carracedo et al. 2001; Gusmão et al. 2003). All three groups (CR, MX and USA) therefore show mtDNA inherited primarily from Amerinds and Y-chromosome DNA inherited primarily from Spanish/Portuguese ancestors.

Although AMOVA was significant for the Y-STRs in the three populations and six regions, the percentage of differentiation was low (4-4.5%). This significance was due primarily to the results of 3 markers (DYS392, DYS439 and DYS389II), in which the populations did not share the most common allele, the difference between the frequencies of the most representative allele was extreme and the number of alleles present was different.

As expected from previous studies, the genetic distances based on Y-STR and mtDNA data showed small differences. The geographic proximity, shared political history, and

frequent migration could explain the similarity between these populations that is revealed by the gene frequencies (Cavalli-Sforza 1998; Cavalli-Sforza and Feldman 2003; CEPAL 1999).

In conclusion, this study describes in detail the genetic relationship between the Hispanic populations of CR, MX and the Southwest USA. As previously reported, the main contribution to the male lineage is of Spanish/Portuguese ancestry and the main contribution to the maternal lineage comes from a probable small effective group of Amerindian ancestors. A higher variation was found for the Y-chromosome than for the mtDNA, suggesting increased migration and mutation events that coalesce the maternal lineages.

### **Acknowledgments**

We are indebted to the patients and family members who participated in this study. This work was supported in part by National Institute of Mental Health grants R01-MH61884 (ME) and RO1-MH60881 (ME), the International Center for Genetic Engineering and Biotechnology Project CRP/COS98-01 (HR). The data is available upon request to HR.

### **Electronic database information**

- Y Chromosome Reference Database for USA, <http://www.ystr.org/usa> (for USA population data on Y-STRs)

**Literature cited**

- Anderson S, Bankier AT, Barrel BG, Bruijn MHL, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJH, Staden R, Young IG. 1981. Sequence and organization of the human mitochondrial genome. *Nature* 290: 457-465.
- Bianchi NO, Catanesi CI, Bailliet G, Martínez-Marignac VL, Bravi CM, Vidal-Rioja LB, Herrera RJ, López-Camelo JS. 1998. Characterization of ancestral and derived Y-chromosome haplotypes of New World Native populations. *Am. J. Hum. Genet.* 63: 1862-1871.
- Carracedo A, Beckmann A, Bengs A, Brinkman B, Caglia A, Capelli C, Gill P, Gusmão L, Hagelberg G, Hohoff C, Hoste B, Kihlgren A, Kloosterman A, Dupuy BM, Morling N, O'donnell G, Parson W, Phillips C, Pouwells M, Scheithauer R, Schmitter H, Schneider P, Schumm J, Skitsa I, Stradmann-Bellinghausen B, Stuart M, Syndercombe D, Vide C. 2001. Results of a collaborative study of the EDNAP Group regarding the reproducibility and robustness of the Y-chromosome STRs DYS19, DYS389 I and II, DYS390 and DYS393 in a PCR pentaplex format. *Forensic Sci. Int.* 119: 28-41.
- Carvajal-Carmona LG, Ophoff R, Service S, Hartiala J, Molina J, León P, Ospina J, Bedoya G, Freimer N, Ruiz-Linares A. 2003. Genetic demography of Antioquia (Colombia) and the Central Valley of Costa Rica. *Hum. Genet.* 112: 534-541.
- Cavalli-Sforza LL. 1998. The DNA revolution in population genetics. *Trends Genet.* 14: 60-65.

- Cavalli-Sforza LL, Feldman MW. 2003. The application of molecular genetic approaches to the study of human evolution. *Nat. Genet. supplement* 33: 266-274.
- CEPAL (Comisión Económica para América Latina y el Caribe). 1999. *Migración y desarrollo en América del Norte y Centroamérica: una visión sintética*. Santiago, Chile: Naciones Unidas.
- Collins-Schramm HE, Chima B, Morii T, Wah K, Figueroa Y, Criswell LA, Hanson RL, Knowler WC, Silva G, Belmont JW, Seldin MF. 2004. Mexican American ancestry-informative markers: examination of population structure and marker characteristics in European Americans, Mexican Americans, Amerindians and Asians. *Hum. Genet.* 114: 263-271.
- Cordaux R, Saha N, Bentley GR, Aunger R, Sirajuddin SM, Stoneking M. 2003. Mitochondrial DNA analysis reveals diverse histories of tribal populations from India. *Eur. J. Hum. Genet.* 11: 253-264.
- Excoffier L. 2001. Analysis of population subdivision. In *Handbook of Statistical Genetics*, Balding DJ, Bishop M, Cannings C, eds. Chichester, England: John Wiley & Sons Ltd, 271-307.
- Green LD, Derr JN, Knight A. 2000. mtDNA affinities of the peoples of North-Central Mexico. *Am. J. Hum. Genet.* 66: 989-998.
- Gusmão L, Sánchez-Diz P, Alves C, Quintáns B, García-Poveda E, Geada H, Raimondi E, Silva de la Fuente SM, Conceição Vide, M, Whittle MR, Zarrabeitia MT, Carvalho M, Negreiros V, Prieto Solla L, Riancho JA, Campos-Sánchez R, Vieira-Silva C, Toscanini U, Amorim A, Carracedo A, GEP-ISFG. 2003. Results of the GEP-ISFG collaborative study on the Y-chromosome STRs GATA A10, GATA C4, GATA H4,

- DYS437, DYS438, DYS439, DYS460 and DYS461: population data. *Forensic Sci. Int.* 135:150-157.
- Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids Symp. Ser.* 41:95-98.
- Hamilton G, Stoneking M, Excoffier L. 2005. Molecular analysis reveals tighter social regulation of immigration in patrilocal populations than in matrilineal populations. *Proc. Natl. Acad. Sci. USA* 102: 7476-7480.
- Hey J, Machado CA. 2003. The study of structured populations - new hope for a difficult and divided science. *Nat. Rev. Genet.* 4: 535-543.
- Jiménez-Arce G. 2000. Variación del cromosoma Y en la población afrocostarricense de Limón, Costa Rica (dissertation). Costa Rica: Universidad de Costa Rica.
- Kayser M, Brauer S, Schädlich H, Prinz M, Batzer MA, Zimmerman PA, Boatman BA, Stoneking M. 2003. Y chromosome STR haplotypes and the genetic structure of U.S. populations of African, European, and Hispanic ancestry. *Genome Res.* 13: 624-634.
- Maza-Zavala DF. 1992. Hispanoamérica-Angloamérica: causas y factores de su diferente evolución. Madrid, España: Editorial Mapfre.
- Meléndez EL. 2001. Polimorfismos del ADN mitocondrial en una muestra general de la población costarricense y su aplicabilidad en la identificación individual (dissertation). Costa Rica: Universidad de Costa Rica.
- Melton T, Clifford S, Kayser M, Nasidze I, Batzer M, Stoneking M. 2001. Diversity and heterogeneity in mitochondrial DNA of North American populations. *J. Forensic Sci.* 46: 46-52.
- Nei M. 1972. Genetic distance between populations. *Amer. Natur.* 106: 283-292.

- Nei M. 1987. *Molecular Evolutionary Genetics*. New York: Columbia University Press.
- Pérez-Lezaun A, Calafell F, Comas D, Mateu E, Bosch E, Martínez-Arias R, Clarimón J, Fiori G, Luiselli D, Facchini F, Pettener D, Bertranpetit J. 1999. Sex-specific migration patterns in Central Asian populations, revealed by analysis of Y-chromosome short tandem repeats and mtDNA. *Am. J. Hum. Genet.* 65: 208-219.
- Rangel-Villalobos H, Jaloma-Cruz AR, Cerda-Aguilar L, Ríos Angulo CD, Mendoza-Carrera F, Patiño-García B, Sandoval-Ramírez L, Figuera-Villanueva LE. 2001a. La huella genética del DNA en varones: haplotipos del cromosoma Y en una población mexicana analizando cinco STRs. *Rev. Invest. Clín.* 53: 401-406.
- Rangel-Villalobos H, Jaloma-Cruz AR, Sandoval-Ramírez L, Velarde-Félix JS, Gallegos-Arreola MP, Figuera LE. 2001b. Y-Chromosome Haplotypes for Six Short Tandem Repeats (STRs) in a Mexican Population. *Arch. Med. Res.* 32: 232-237.
- Relethford JH. 2001. Genetic history of the human species. In *Handbook of Statistical Genetics*, Balding DJ, Bishop M, Cannings C, eds. Chichester, England: John Wiley & Sons Ltd, 813-846.
- Salas A, Richards M, Lareu MV, Scozzari R, Coppa A, Torroni A, Macaulay V, Carracedo A. 2004. The African Diaspora: mitochondrial DNA and the Atlantic slave trade. *Am. J. Hum. Genet.* 74: 454-465.
- Santos M, Ward RH, Barrantes R. 1994. mtDNA variation in the Chibcha Amerindian Huetar from Costa Rica. *Hum. Biol.* 66: 963-977.
- Schneider S, Kueffer JM, Roessli D, Excoffier L. 1997. *Arlequin version 1.1*. University of Geneva, Geneva.

Schneider S, Roessli D, Excoffier L. 2000. Arlequin version 2.000. A Software for Population Genetics Data Analysis, University of Geneva, Geneva.

Tamura K, Nei M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* 10: 512-526.

Wilder JA, Kingan SB, Mobasher Z, Pilkington MM, Hammer MF. 2004. Global patterns of human mitochondrial DNA and Y-chromosome structure are not influenced by higher migration rates of females versus males. *Nat. Genet.* 36:1122-1124.

Table 1. Gene diversity for each of the 12 Y-STRs analyzed in the Hispanic populations related to schizophrenia from Costa Rica, the Southwest United States and Mexico. Standard deviations are in parentheses.

<b>Markers</b>	<b>CR</b>	<b>USA</b>	<b>MX</b>
DYS 19	0,697	0,584	0,601
DYS 385I	0,798	0,680	0,758
DYS 385II	0,793	0,714	0,804
DYS 389I	0,609	0,544	0,432
DYS 389II	0,769	0,681	0,704
DYS 390	0,675	0,573	0,573
DYS 391	0,507	0,580	0,567
DYS 392	0,682	0,608	0,719
DYS 393	0,502	0,387	0,286
DYS 437	0,628	0,557	0,508
DYS 438	0,702	0,627	0,653
DYS 439	0,671	0,658	0,563
Mean gene diversity	0,675 (0,349)	0,614 (0,324)	0,608 (0,321)

Table 2. Diversity measures obtained by the study of 12 Y-STRs in the Hispanic populations of Costa Rica, the Southwest United States and Mexico. N is the number of individuals. Standard deviations are in parenthesis.

	<b>CR</b>	<b>USA</b>	<b>MX</b>
<b>N</b>	120	42	55
<b>No. of haplotypes</b>	103	40	51
<b>Shared haplotypes</b>	14	2	2
<b>Haplotypic diversity</b>	0,997 (0,002)	0,998 (0,006)	0,995 (0,005)
<b>Mean number of pairwise differences</b>	8,099 (3,786)	7,368 (3,516)	7,300 (3,470)

Table 3. Diversity measures in the mtDNA sequences from the Hispanic populations of Costa Rica, the Southwest United States and Mexico. N is the number of individuals. Standard deviations are in parenthesis

	<b>CR</b>	<b>USA</b>	<b>MX</b>
<b>N</b>	120	42	55
<b>No of polymorphic sites</b>	52	39	56
<b>No of unique polymorphisms</b>	18	10	19
<b>Nucleotide diversity</b>	0.024 (0.012)	0.022 (0.012)	0.026 (0.014)
<b>No of haplotypes</b>	39	33	43
<b>Shared haplotypes</b>	17	4	8
<b>Haplotype diversity</b>	0.930 (0.012)	0.981 (0.011)	0.989 (0.007)
<b>Mean number of pairwise differences</b>	7.657 (3.595)	6.881 (3.303)	8.445 (3.967)

Table 4. Shared haplotypes and their description between the Hispanic populations from Costa Rica, the Southwest United States and Mexico with 12 Y-STRs data. Haplotype frequencies are in parenthesis for each population. \* Order of the haplotypes: DYS385I/II, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439 and DYS19.

<b>Haplotype</b>	<b>CR</b>	<b>USA</b>	<b>MX</b>	<b>Description*</b>
<b>HY25</b>	3 (0.025)	2 (0.048)		11/14, 14, 30, 24, 11, 13, 13, 15, 12, 11, 14
<b>HY106</b>		1 (0.024)	1 (0.018)	11/14, 13, 29, 25, 11, 13, 13, 14, 12, 12, 14
<b>HY119</b>		1 (0.024)	1 (0.018)	13/14, 14, 30, 24, 9, 11, 13, 14, 10, 10, 13
<b>HY125</b>		1 (0.024)	4 (0.073)	11/15, 13, 29, 24, 11, 13, 13, 15, 12, 12, 14
<b>HY135</b>		1 (0.024)	1 (0.018)	11/14, 14, 30, 24, 10, 13, 13, 14, 12, 12, 14

Table 5. Shared haplotypes and their description between the Hispanic populations from Costa Rica, the Southwest United States and Mexico for the mtDNA HVI region. The frequency of the haplotype in the population is specified in parenthesis. The description of the haplotype includes the substitution and the position minus 16000.

<b>Haplotype</b>	<b>CR</b>	<b>USA</b>	<b>MX</b>	<b>Description (number minus 16000)</b>
<b>HM 4</b>	19 (0.158)	4 (0.095)	4 (0.073)	111T, 223T, 290T, 319A, 362C
<b>HM10</b>	6 (0.050)	2 (0.048)	1 (0.018)	182C, 183C, 189C, 217C
<b>HM20</b>	1 (0.008)		1 (0.018)	111C, 209C, 223T, 290T, 319A, 360T, 362C
<b>HM39</b>	1 (0.008)		1 (0.018)	183C, 189C, 217C, 278T,
<b>HM43</b>		1 (0.024)	1 (0.018)	111T, 223T, 290T, 362C
<b>HM49</b>		4 (0.095)	1 (0.018)	223T, 298C, 325C, 327T
<b>HM53</b>		1 (0.024)	1 (0.018)	111T, 183C, 189C, 217C
<b>HM54</b>		3 (0.071)	3 (0.055)	223T, 290T, 319A, 362C
<b>HM57</b>		1 (0.024)	1 (0.018)	182C, 183C, 189C, 224C, 311C
<b>HM69</b>		1 (0.024)	2 (0.036)	182C, 183C, 189C, 217C, 295T

Table 6. AMOVA results in the three Hispanic populations or separated by regions for the 12 Y-STRs and the HVI sequence analyzed. The three populations are † Costa Rica, the Southwest United States and Mexico. The six regions are †† Costa Rica, San Diego, San Antonio, Los Angeles, Monterrey, and Mexico City.

	12 Y-STRs			mtDNA	
	Variance	R <sub>st</sub>	R <sub>st</sub> per locus	Variance	F <sub>st</sub>
<b>3 populations †</b>					
Between populations	4.00	0.040*	0.016**	0.64	0.006
Inside populations	96.00			99.36	
<b>6 regions ††</b>					
Between populations	4.53	0.038*	0.019**	0.96	0.005
Between populations inside regions	-0.77			-0.47	
Inside populations	96.25			99.51	

\* P<0.01 \*\* P<0.001

Table 7. Genetic distances between the Hispanic populations from Costa Rica, the Southwest United States and Mexico for the 12 Y-STRs (Nei 1972, above diagonal) and the HVI sequence (Tamura and Nei 1993, below diagonal) of the mtDNA

<b>mtDNA\Y-STR</b>	<b>CR</b>	<b>USA</b>	<b>MX</b>
<b>CR</b>		0,075	0,073
<b>USA</b>	0,015		0,072
<b>MX</b>	0,003	-0,002	