

## Análisis de la estructura genética en una muestra poblacional de Bucaramanga, departamento de Santander\*

MARTHA LUCÍA HINCAPIÉ, MSc<sup>1</sup>, ADRIANA MARÍA GIL, MSc<sup>2</sup>, ADRIANA LUCÍA PICO, MSc<sup>2</sup>,  
LEONOR GUSMÃO, PhD<sup>3</sup>, FERNANDO RONDÓN, PhD<sup>4</sup>, CLARA INÉS VARGAS, MSc<sup>2</sup>,  
ADRIANA CASTILLO, MSc<sup>2</sup>

### RESUMEN

**Introducción:** El fenómeno de sub-estructura en las poblaciones ha tenido desde hace varios años un abordaje amplio, que se enfocó, entre otros, en la identificación y cuantificación de la mezcla étnica presente en estudios de mapeo asociativo, para comprobar la asociación de marcadores polimórficos en el desarrollo de enfermedades comunes complejas, como responsable de falsos positivos. No obstante el reconocimiento de este problema, no se tiene suficiente información genética en el contexto nacional ni local que permita determinar la posible diferenciación de subgrupos poblacionales en cada región en particular.

**Objetivo:** Determinar la estructura genética en una muestra poblacional de la ciudad de Bucaramanga, a partir del análisis de 19 marcadores microsatélites autosómicos en distintos subgrupos poblacionales.

**Metodología:** De la base de datos del Laboratorio de Genética Humana de la Universidad Industrial de Santander, se seleccionaron aleatoriamente 350 muestras de ADN, y se amplificaron 19 marcadores autosómicos Short Tandem Repeat mediante los «kits Powerplex<sup>®</sup> 16 y FFFL (Promega)».

**Resultados:** En el análisis de equilibrio Hardy Weinberg, no se obtuvieron diferencias estadísticamente significativas en 18 de 19 marcadores Short Tandem Repeat autosómicos analizados en la población de Bucaramanga. El único marcador que mostró no estar en equilibrio Hardy Weinberg en la población de Bucaramanga fue el F13B (valor de significancia de  $p=0.00264$ , después de aplicar la corrección de Bonferroni).

**Discusión:** Las poblaciones representadas en los seis estratos socioeconómicos mostraron alta diversidad genética intragrupos, que ratificó una alta variabilidad entre los individuos de la ciudad de Bucaramanga, acorde con el bajo valor de  $F_{ST}$  entre distintos grupos, determinado en el análisis molecular de varianza con base en frecuencias alélicas observadas para los 19 Short Tandem Repeat analizados.

**Conclusión:** La alta diversidad genética y el análisis molecular de varianza mostraron una baja diferenciación entre los seis estratos socioeconómicos en la ciudad de Bucaramanga, y evidenciaron a la población como una misma unidad genética, no sub-estructurada.

**Palabras clave:** STRs; Sub-estructura poblacional; Frecuencias alélicas; Bucaramanga; Estudios de asociación.

*An assessment of genetic structure in a Bucaramanga, department of Santander, population sample*

### SUMMARY

**Introduction:** The phenomenon of substructure in the populations has been greatly analyzed for several years, and it has been focused especially on the identification and quantification of ethnic mixture present in studies of associative mapping

---

\* Estudio financiado por la Vicerrectoría de Investigación y Extensión de la Universidad Industrial de Santander, Bucaramanga, código interno 5266 y por IPATIMUP, apoyado parcialmente por la Fundação para la Ciência e Tecnologia, Programa Operacional Ciencia e Innovación 2010 (POCI2010), Porto, Portugal.

1. Docente, Escuela de Bacteriología, Facultad de Salud, Universidad Industrial de Santander, Bucaramanga, Colombia. e-mail: maluhincapie@yahoo.com
2. Docente, Escuela de Medicina, Laboratorio de Genética Humana, Facultad de Salud, Universidad Industrial de Santander, Bucaramanga, Colombia. e-mail: labgen2@uis.edu.co, cvargas@uis.edu.co, castillo@uis.edu.co
3. Investigadora Senior, Instituto de Patología e Inmunología Molecular, Universidad de Porto (IPATIMUP), Porto, Portugal. e-mail: lgusmao@ipatimup.pt
4. Docente, Escuela de Biología, Facultad de Salud, Universidad Industrial de Santander, Bucaramanga, Colombia. e-mail: fercho\_gen@yahoo.com

Recibido para publicación marzo 16, 2009 Aceptado para publicación septiembre 30, 2009

to verify the association of polymorphic markers in the development of complex and common diseases responsible for false positives. Nevertheless, despite the recognition of this issue, there is insufficient genetic information within the national or local contexts that allow assessing the possible differentiation of population sub-groups in each particular region.

**Objective:** To determine the genetic structure in the city of Bucaramanga through the analysis of 19 autosomal microsatellite markers in different subgroups of the population.

**Methodology:** A total of 350 DNA samples were randomly selected from the database of the Human Genetic Laboratory at Universidad Industrial de Santander by using Epi Info version 6.04 2001. Also, 19 Short Tandem Repeat markers were amplified using «kits Powerplex® 16 and FFFL (Promega)».

**Results:** In the Hardy Weinberg equilibrium analysis (100 steps in Markov chain and 1000 dememorization steps), no statistically significant differences in 18 out of the 19 analyzed STRs markers in the population of Bucaramanga were obtained. A unique marker that proved not present in HWE in the population of Bucaramanga was the F13B (for a significance value of  $p=0.00264$ , after applying the Bonferroni correction).

**Discussion:** The populations represented in the six socioeconomic levels presented high genetic diversity intragroups, which ratified the high variability among the individuals in this city according to the low value of *FST* for different groups, determined via the molecular analysis of variance based on the allelic frequencies observed for the 19 analyzed Short Tandem Repeats.

**Conclusions:** The high genetic diversity, as well as the analysis of molecular variance displayed low divergence for each of the six socioeconomic levels in the city of Bucaramanga. Therefore, it is suggested that this population is an equivalent genetic unit and not substructured.

**Keywords:** STRs; Population substructure; Allelic frequencies; Bucaramanga; Association studies.

Los rápidos avances en el campo molecular y el estudio del Proyecto Genoma Humano<sup>1</sup>, facilitan cada vez más la búsqueda de genes de susceptibilidad a las enfermedades mediante el método de casos y controles, donde se emplea tanto el enfoque poblacional como el familiar. Este método es uno de los más usados en estudios de asociación, por las facilidades que presenta en el reclutamiento de individuos.

En los estudios poblacionales los métodos de casos y controles se utilizan para: i). Análisis de los determinantes de mutaciones humanas. ii). Evaluar el papel que desempeñan los indicadores genéticos inespecíficos,

como la consanguinidad y la mezcla de individuos con diversos orígenes de antepasados en la etiología de las enfermedades y iii). Valor del papel etiológico de los rasgos genéticos<sup>2</sup>.

Este desarrollo, en los últimos años, ha permitido realizar numerosos estudios de mapeo asociativo, que buscan relacionar una enfermedad con un marcador de ADN localizado cerca de un gen de interés o gen candidato. El hallazgo de una asociación se toma como evidencia de la posibilidad que un marcador esté ligado a un gen que produce la enfermedad, aunque pueda existir asociación en ausencia de ligamiento. Por esto, además de una prueba de asociación, es recomendable determinar la relación del marcador con el gen productor de la susceptibilidad a la enfermedad<sup>3</sup>.

Las asociaciones genéticas en marcadores no ligados surgen con frecuencia en poblaciones que han sufrido mezcla reciente, asociada con una subestructuración poblacional y con grupos que comparten relaciones de herencia o cuando hay consanguinidad unida con numerosos factores de índole demográfico, cultural, religioso y geográfico, que se podrían relacionar con la enfermedad y, por consiguiente, convertirse en factores potenciales responsables de la detección de «falsas asociaciones».

Y son entonces las «asociaciones falsas» el resultado de fenómenos de mezcla, estratificación o subdivisión<sup>2</sup>, con los que se denota la estructura genética presente en una población, al mostrar diferencias en las frecuencias genotípicas y alélicas que suelen ser divergentes de una región geográfica a otra; este tipo de asociaciones pueden darse cuando se hace el muestreo sin considerar la etnicidad (pertenencia de individuos de una población a un grupo étnico), ni la herencia (origen histórico) de la población en estudio.

Si la estratificación estuviera presente en la población que se analiza, tanto los marcadores no ligados como los ligados podrían mostrar asociación con determinado fenotipo<sup>2,3</sup>. Por este motivo, para inferir la estratificación presente en la población se pueden emplear técnicas bastante precisas, como el método llamado control genómico (CG), propuesto por Devlin *et al.*<sup>4</sup> que se basa en la implementación de marcadores bialélicos, utilizados para corregir el efecto debido a la presencia de mezcla en la población y el de asociación estructurada (AS) propuesto por Pritchard & Rosenberg<sup>3</sup>, donde se usan datos genotípicos de

multilocus a fin de relacionar en detalle la estructura de la población bajo estudio. Este último presenta ventajas sobre el CG, pues permite conocer más datos acerca de los fenómenos que afectan la(s) población(es) estudiada(s), siendo los patrones de migración y colonización, coalescencia y deriva, algunos de ellos<sup>3</sup>.

La evaluación de la estructura genética poblacional a partir de datos genotípicos polimórficos, se realiza con los estimadores de sub-estructura poblacional  $F_{ST}$  y  $R_{ST}$ . El estimador  $F_{ST}$  se ha implementado durante años para medir la diferenciación entre poblaciones<sup>5</sup>, pues evalúa el déficit en la proporción de genotipos heterocigotos, que determina la presencia de subestructuras<sup>2</sup>; y por el contrario el  $R_{ST}$  asume el modelo de mutación gradual de los microsatélites, definido como una fracción de la varianza total del tamaño de los alelos entre poblaciones<sup>5</sup>. Con estos estimadores se establece la estructura genética presente en una población<sup>3</sup>.

Los fenómenos de la estructura genética en las poblaciones enfrentan inconvenientes como la presencia de sub-estructura críptica<sup>2</sup>, que podría presentarse cuando los tamaños de muestra grandes donde la sub-estructura de la población puede ir en aumento, y facilitar el hallazgo de «falsas asociaciones». En este mismo sentido, los estudios de meta-análisis implican sesgos significativos, debido al considerable número de individuos en el análisis, sumado a la gran diversidad de criterios de selección de los mismos y a la divergencia del fondo genético que puede existir entre las distintas poblaciones estudiadas<sup>3</sup>.

Si se considera que en el contexto nacional<sup>6-8</sup>, y local<sup>8</sup> son pocos los estudios que se centran en determinar la diferenciación genética de las poblaciones, se diseñó un estudio con una muestra de subgrupos poblacionales pertenecientes a los distintos estratos socioeconómicos de la ciudad de Bucaramanga, para tipificar 19 marcadores microsatélites autosómicos. Lo obtenido constituye un primer resultado de una serie de investigaciones proyectadas en líneas de carácter biomédico y evolutivo, cuya pretensión, como objetivo central, es suministrar información que soporte los efectos de estudios de asociación hechos en muestras provenientes de la población santandereana, para así evitar los sesgos ocasionados por «falsas» asociaciones y falta de reproducibilidad en los hallazgos de los modelos que estudian las enfermedades complejas<sup>3</sup>.

## MATERIALES Y MÉTODOS

**Población de estudio.** El área urbana de la ciudad de Bucaramanga, capital del Departamento de Santander, se halla en el noreste de Colombia. Su población se clasifica en seis estratos socioeconómicos (SES), definidos por la disponibilidad de servicios públicos y la calidad de las unidades habitacionales en el vecindario<sup>9</sup>.

**Muestras de la población.** Fueron seleccionadas aleatoriamente en total 350 muestras de ADN (Epi Info versión 6.04 2001)<sup>10</sup>, de la base de datos del Laboratorio de Genética Humana de la Universidad Industrial de Santander (GH-UIS); además, estas muestras cumplieron con los criterios de selección, con un poder de 80% y un error tipo alfa de 0.05. El número de muestras por estrato socioeconómico, fue tomado con base en la distribución porcentual de la totalidad de habitantes de la ciudad, que corresponde a 509,918 personas, de acuerdo con los datos registrados en el Departamento Nacional Administrativo de Estadística (DANE), para el último censo del 2005. El número de individuos y su respectiva distribución porcentual por estrato, con base en ese censo<sup>11</sup>, se resumen en el Cuadro 1. Esta investigación fue clasificada como un estudio «de riesgo mínimo», según el numeral b del artículo 10, de la resolución 008430 de 1993 del Ministerio de Salud de la República de Colombia.

**Extracción de ADN.** El ADN fue extraído con el método descrito por Miller *et al.*<sup>12</sup> en muestras de sangre de individuos sanos no emparentados nacidos en Bucaramanga, guardando proporción entre hombres y mujeres. Todo participante del estudio firmó un consentimiento informado, que aprobó el comité de bioética de investigaciones científicas de la Facultad de Salud de la UIS, para autorizar la toma de muestra sanguínea. Fue obtenido y clasificado el nivel socioeconómico (seis niveles) de acuerdo con el criterio del DANE<sup>11</sup>. Asimismo, cada persona diligenció una encuesta, donde informaba su etnicidad según el color de su piel y otros factores adicionales (Cuadro 1).

**Amplificación por PCR y genotipificación de STRs.** Fueron amplificados 19 marcadores autosómicos STRs (D3S1358, HUMTH01, D21S11, D18S51, D5S818, D13S317, D7S820, D16S539, CSF1PO, HUMvWA, D8S1179, HUMTPOX, FGA, LPL, F13B, FESFPS, F13A01, PENTA D y PENTA E), con los

**Cuadro 1**  
**Población y muestreo de la ciudad de Bucaramanga en los SES<sup>a</sup>**

<b>Estratos</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>Total</b>
Porcentaje <sup>b</sup>	8	18	30	29	4	11	100
Número de individuos	28	63	105	102	14	38	350
<b>Etnicidad autorreportada</b>							
Color de piel mestiza	<b>14</b>	<b>49</b>	<b>83</b>	<b>83</b>	<b>11</b>	<b>24</b>	<b>264</b>
blanca	<b>14</b>	<b>13</b>	<b>25</b>	<b>17</b>	<b>3</b>	<b>13</b>	<b>85</b>
negra	<b>0</b>	<b>1</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>1</b>

a. Seis estratos socioeconómicos

b. Fuente: Departamento Nacional de Planeación. Estratificación Socioeconómica. Manual General. 1994<sup>9</sup>  
Departamento Administrativo Nacional de Estadística, DANE. 2005<sup>11</sup>

«kits Powerplex<sup>®</sup> 16 y FFFL (Promega)» según las instrucciones de los fabricantes (Promega Corporation, Madison, Wisconsin, USA, Cat. A1120).

Los productos amplificados fueron separados mediante electroforesis capilar en un analizador genético ABI PRISM 310 (PE-Biosystems, Foster City, CA) y la tipificación fue realizada por comparación con una escalera alélica de referencia provista por el fabricante (Promega). La asignación de alelos fue hecha según las recomendaciones de la Comisión de ADN de la Sociedad Internacional de Genética Forense<sup>13</sup>.

**Análisis de datos.** Para el análisis, los datos obtenidos fueron organizados al considerar tres abordajes diferentes, de los SES representados en la ciudad de Bucaramanga.

**Abordaje 1.** Aquí los datos fueron reunidos en dos grupos: BAJO-MEDIO (incluye los estratos 1, 2 y 3; N=196) y MEDIO-ALTO (estratos 4, 5 y 6; N=154). Esta estrategia se basó en la selección de individuos incluidos en el proyecto «Conjunto de Acciones para la Reducción Multifactorial de Enfermedades no Transmisibles (CARMEN)», hecho en la ciudad de Bucaramanga para estudiar la conexión entre factores de riesgo y enfermedades crónicas no transmisibles, en estratos socioeconómicos 2 y 3<sup>14</sup>. Por tanto, en la categoría BAJO-MEDIO estaría la población estudiada en CARMEN, con lo cual el resto de la población debía quedar concentrada en la categoría MEDIO-ALTO.

**Abordaje 2.** En éste la distribución siguió las clasificaciones del Departamento Nacional de Planeación, DNP<sup>9</sup>, que divide a los habitantes de Bucaramanga en

tres grupos: BAJO (incluye estratos 1 y 2; N=91), MEDIO (estratos 3 y 4; N=207) y ALTO (con los estratos 5 y 6; N=52).

**Abordaje 3.** Éste incluyó cuatro grupos, fue planteado para equilibrar el número de individuos que se asignaron a cada uno de los grupos establecidos en el abordaje 2. El nivel BAJO (incluye los estratos 1 y 2; N=91), BAJO-MEDIO (constituido únicamente por el estrato 3; N=105), MEDIO (estrato 4; N=102) y ALTO (con los estratos 5 y 6; N=52).

De acuerdo con el abordaje obtenido, fueron construidas diferentes bases de datos duplicadas en Excel<sup>®</sup>, para validarlas con la herramienta Validate del software Epi Info versión 6.04 2001<sup>10</sup>. El cálculo de las frecuencias alélicas, el «test» de equilibrio de Hardy-Weinberg (HW), la corrección de Bonferroni para «tests» múltiples, que consiste en hacer contrastes *a posteriori* precisando las muestras con diferencias significativas<sup>15</sup>, que provocan el rechazo de la hipótesis nula, el «test» exacto de diferenciación poblacional descrito por Raymond & Rousset (1995)<sup>16</sup>, la estimación de las distancias genéticas poblacionales «pairwise»  $F_{ST}$ , y el análisis de varianza molecular (AMOVA) efectuados con el software Arlequin 3.11<sup>17</sup>. Adicionalmente fue realizado un análisis de asignación de las muestras por grupos de membresía a partir de los genotipos individuales obtenidos con los 19 microsatélites caracterizados, mediante el software STRUCTURE v. 2.0<sup>18</sup>. La subestructuración fue evaluada por comparación de los diversos grupos socioeconómicos, si se tiene en cuenta que puede ser un factor importante que contribuye a la

formación de sub-estructura genética en algunas áreas urbanas de la ciudad.

## RESULTADOS

Las frecuencias alélicas de la población de Bucaramanga fueron calculadas para 19 STRs autosómicos (Cuadro 2).

La diversidad genética promedio sobre todos los loci evaluados fue  $0.772 \pm 0.383$ . El sistema más diverso fue Penta E (0.88857) y el menos diverso fue FESFPS (0.65429). De los 19 STRs analizados, el más polimórfico fue Penta E con 19 alelos y los sistemas TH01 y F13B fueron los menos polimórficos porque presentaron 6 alelos para el total de la población estudiada.

Los análisis jerárquicos de varianza molecular, basados en 100,000 permutaciones, revelaron los componentes de varianza sobre la base de dos coeficientes de diferenciación genética entre poblaciones ( $F_{ST}$  y  $R_{ST}$ )<sup>5</sup>. En general, estos fueron bajos entre los SES que se analizaron en los diferentes abordajes, sin presencia de aislamiento por distancia genética (los datos no se muestran). No hubo diferencias estadísticas significativas dentro ni entre las distintas agrupaciones, confirmando que la mayor variación fue encontrada entre los individuos dentro de los grupos. Por tanto, los productos reflejarían que los SES representados en la población de Bucaramanga y en particular, la tercera estrategia para agrupar los habitantes, se encontraban homogéneamente distribuidos desde el punto de vista genético (Cuadro 3).

Al emplear cualquiera de las tres estrategias de agrupamiento del total de muestras analizadas, aproximadamente 98% de la variabilidad genética encontrada correspondió a variación entre individuos dentro de los grupos. No se observó diferenciación genética significativa entre los distintos grupos; esto refleja que la población de Bucaramanga se comporta como una unidad panmíctica, desde el punto de vista genético y permite inferir que las uniones entre los individuos ocurren en forma aleatoria<sup>3</sup>. Los resultados también fueron confirmados mediante el análisis con el software STRUCTURE v. 2.0<sup>18</sup>, realizando corridas desde  $K=2$  hasta  $K=6$  en cada una de las estrategias de agrupamiento de los SES, y en todos los casos se obtuvo la superposición de las diferentes muestras analizadas para cada estrategia de agrupamiento y para cada valor de  $K$  en el rango mencionado (Figura 1).

En cuanto al análisis de equilibrio HW (100,000 pasos de dememorización en la cadena de Markov y 1,000 pasos de memorización), no se obtuvieron desvíos significativos, para 18 de los 19 marcadores STRs autosómicos analizados (Cuadro 2). El único marcador que mostró diferencias genotípicas significativamente diferentes a las esperadas, fue el F13B (incluso después de aplicar la corrección de Bonferroni para pruebas múltiples;  $p < 0.00264$ )<sup>19</sup>.

## DISCUSIÓN

Este estudio es pionero en el análisis de la estructura poblacional a partir de muestras de individuos nacidos en Bucaramanga. Como no hay informes sobre este tipo de investigaciones en la ciudad, los resultados de la tipificación de los STRs en esta población fueron comparados con estudios que incluían el departamento de Santander y por tanto algunas muestras de esta ciudad. Hay tres trabajos para el departamento y, como el 60% de la población de Santander se concentra en la capital, Bucaramanga<sup>11</sup>, se puede inferir que los resultados reflejan de una manera representativa la población de esta ciudad.

Por tanto, fueron comparados los datos del presente estudio, mediante AMOVA pareada con los reportes de Vargas *et al.*<sup>19,20</sup>, Paredes *et al.*<sup>21</sup> y Yunis *et al.*<sup>8</sup> La distribución de las frecuencias alélicas estimadas en el presente estudio (Cuadro 2a, 2b) fue concordante con los hallazgos de estos autores para los sistemas STRs autosómicos: CFS1PO, TH01, TPOX, vWA, D13S317, D7S820, D16S539, D3S1358, D21S11, D18S51, PENTA E y PENTA D ( $p > 0.082$ ) pero no con los sistemas D5S818, FGA y D8S1178 ( $p < 0.05$ ).

El presente trabajo mostró al sistema TH01 como el menos polimórfico, hallazgo concordante con Vargas *et al.*<sup>19</sup>, Paredes *et al.*<sup>21</sup> y Yunis *et al.*<sup>8</sup> El sistema más polimórfico, Penta E, coincidió con los datos de Vargas *et al.*<sup>20</sup> y Yunis *et al.*<sup>8</sup>

El análisis para el HWE indicó que 18 de los 19 sistemas STRs se encontraron en equilibrio, concordante con los resultados de los autores mencionados para los 15 marcadores STRs incluidos en estos estudios. Hay así una evidencia para afirmar que la población de Santander en general se comporta como una misma unidad genética (Figura 1).

La prueba para diferenciar poblaciones contra la

**Cuadro 2 a**  
**Frecuencias alélicas de 19 STRs en una muestra poblacional de Bucaramanga, Santander, Colombia**

Alelo	D3S1358	THO1	D13S317	D7S820	D16S539	CSF1PO	VWA	D8S1179	TPOX
2.2	–	–	–	–	–	–	–	–	–
3	–	–	–	–	–	–	–	–	–
4	–	–	–	–	–	–	–	–	–
5	–	–	–	–	–	–	–	–	–
6	–	–	–	–	–	–	–	–	0,004
7	–	–	–	0,013	–	0,001	–	–	–
8	–	–	0,097	0,106	0,011	0,009	–	0,01	0,487
9	–	–	0,134	0,086	0,136	0,024	–	0,003	0,069
9.3	–	0,171	–	–	–	–	–	–	–
10	–	0,001	0,053	0,271	0,187	0,217	–	0,057	0,041
11	–	0,414	0,231	0,246	0,279	0,289	–	0,096	0,307
12	0,001	0,190	0,29	0,246	0,230	0,401	–	0,137	0,087
13	0,003	0,089	0,126	0,027	0,136	0,054	0,003	0,321	0,004
14	0,081	0,134	0,066	0,006	0,021	0,003	0,091	0,246	–
15	0,36	–	0,003	–	–	0,001	0,076	0,103	–
16	0,269	–	–	–	–	–	0,331	0,026	–
17	0,154	–	–	–	–	–	0,287	0,001	–
18	0,129	–	–	–	–	–	0,166	–	–
19	0,003	–	–	–	–	–	0,044	–	–
20	–	–	–	–	–	–	0,001	–	–
21	–	–	–	–	–	–	–	–	–
22	–	–	–	–	–	–	–	–	–
22.2	–	–	–	–	–	–	–	–	–
23	–	–	–	–	–	–	–	–	–
24	–	–	–	–	–	–	–	–	–
25	–	–	–	–	–	–	–	–	–
26	–	–	–	–	–	–	–	–	–
27	–	–	–	–	–	–	–	–	–
28	–	–	–	–	–	–	–	–	–
29	–	–	–	–	–	–	–	–	–

**Cuadro 2 a**  
**Frecuencias alélicas de 19 STRs en una muestra poblacional de Bucaramanga, Santander, Colombia**  
 (continuación)

Alelo	D3S1358	THO1	D13S317	D7S820	D16S539	CSF1PO	VWA	D8S1179	TPOX
29.2	-	-	-	-	-	-	-	-	-
30	-	-	-	-	-	-	-	-	-
30.2	-	-	-	-	-	-	-	-	-
31	-	-	-	-	-	-	-	-	-
31.2	-	-	-	-	-	-	-	-	-
32	-	-	-	-	-	-	-	-	-
32.2	-	-	-	-	-	-	-	-	-
33	-	-	-	-	-	-	-	-	-
33.2	-	-	-	-	-	-	-	-	-
34	-	-	-	-	-	-	-	-	-
34.2	-	-	-	-	-	-	-	-	-
35	-	-	-	-	-	-	-	-	-
Ho	0.757	0.711	0.791	0.760	0.829	0.666	0.774	0.806	0.663
He	0.752	0.738	0.813	0.787	0.798	0.706	0.765	0.795	0.655
P	0.509	0.378	0.189	0.356	0.323	0.572	0.390	0.513	0.961

**Cuadro 2 b**  
**Frecuencias alélicas de 19 STRs en una muestra poblacional de Bucaramanga, Santander, Colombia**

Alelo	FGA	LPL	F13B	FESFPS	F13A01	D21S11	D18S51	D5S818	PENTA D	PENTA E
2.2	-	-	-	-	-	-	-	-	0.004	-
3	-	-	-	-	0.241	-	-	-	-	-
4	-	-	-	-	0.109	-	-	-	-	-
5	-	-	-	-	0.191	-	-	-	0.006	0.033
6	-	-	0.094	-	0.173	-	-	-	-	0.001
7	-	0.001	0.029	0.007	0.261	-	-	0.027	0.006	0.077
8	-	-	0.149	0.014	0.006	-	-	0.009	0.013	0.011
9	-	0.029	0.300	0.013	0.001	-	-	0.070	0.187	0.016
9.3	-	-	-	-	-	-	-	-	-	-
10	-	0,473	0.417	0.251	-	-	0.001	0.053	0.177	0.051

**Cuadro 2 b**  
**Frecuencias alélicas de 19 STRs en una muestra poblacional de Bucaramanga, Santander, Colombia**  
 (continuación)

Alelo	FGA	LPL	F13B	FESFPS	F13A01	D21S11	D18S51	D5S818	PENTA D	PENTA E
11	–	0.201	0.011	0.473	–	–	0.004	0.391	0.211	0.106
12	–	0.249	–	0.160	0.001	–	0.093	0.330	0.164	0.217
13	–	0.046	–	0.079	0.004	–	0.089	0.109	0.150	0.070
14	–0.001	–	0.003	0.001	–	0.199	0.011	0.057	0.103	–
15	–	–	–	0.006	–	0.169	–	0.021	0.104	–
16	–	–	–	0.003	–	0.111	–	0.003	0.051	–
17	0.001	–	–	–	0.001	–	0.171	–	–	0.033
18	0.016	–	–	–	–	–	0.073	–	–	0.027
19	0.063	–	–	–	–	–	0.036	–	–	0.041
20	0.080	–	–	–	–	–	0.027	–	–	0.033
21	0.121	–	–	–	–	–	0.020	–	–	0.020
22	0.154	–	–	–	–	–	0.004	–	–	0.009
22.2	0.003	–	–	–	–	–	–	–	–	–
23	0.141	–	–	–	–	–	0.003	–	–	–
24	0.157	–	–	–	–	–	0.003	–	–	–
25	0.146	–	–	–	–	–	–	–	–	–
26	0.099	–	–	–	–	–	–	–	–	–
27	0.017	–	–	–	–	0.021	–	–	–	–
28	0.001	–	–	–	–	0.057	–	–	–	–
29	–	–	–	–	0.243	–	–	–	–	–
29.2	–	–	–	–	0.001	–	–	–	–	–
30	–	–	–	–	0.320	–	–	–	–	–
30.2	–	–	–	–	0.026	–	–	–	–	–
31	–	–	–	–	0.049	–	–	–	–	–
31.2	–	–	–	–	0.093	–	–	–	–	–
32	–	–	–	–	0.020	–	–	–	–	–
32.2	–	–	–	–	0.120	–	–	–	–	–
33	–	–	–	–	0.001	–	–	–	–	–
33.2	–	–	–	–	0.040	–	–	–	–	–

**Cuadro 2 b**  
**Frecuencias alélicas de 19 STRs en una muestra poblacional de Bucaramanga, Santander, Colombia**  
 (continuación)

Alelo	FGA	LPL	F13B	FESFPS	F13A01	D21S11	D18S51	D5S818	PENTA D	PENTA E
34	–	–	–	–	–	0.001	–	–	–	–
34.2	–	–	–	–	–	0.006	–	–	–	–
35	–	–	–	–	–	0.001	–	–	–	–
Ho	0.877	0.671	0.680	0.654	0.783	0.803	0.849	0.723	0.834	0.889
He	0.876	0.672	0.705	0.682	0.796	0.808	0.867	0.718	0.837	0.900
P	0.241	0.667	0.000	0.013	0.085	0.445	0.061	0.574	0.800	0.649

Ho: heterocigosidad observada, He: heterocigosidad esperada, p: test exacto de equilibrio HW, basado en 2000 permutaciones, error standard <0.01. Corrección de Bonferroni (0.05/19= 0.00263)

**Cuadro 3**  
**Resultados del análisis de distribución de la diversidad genética (AMOVA) con base**  
**en diferentes hipótesis de agrupamiento de los SES en la ciudad de Bucaramanga**

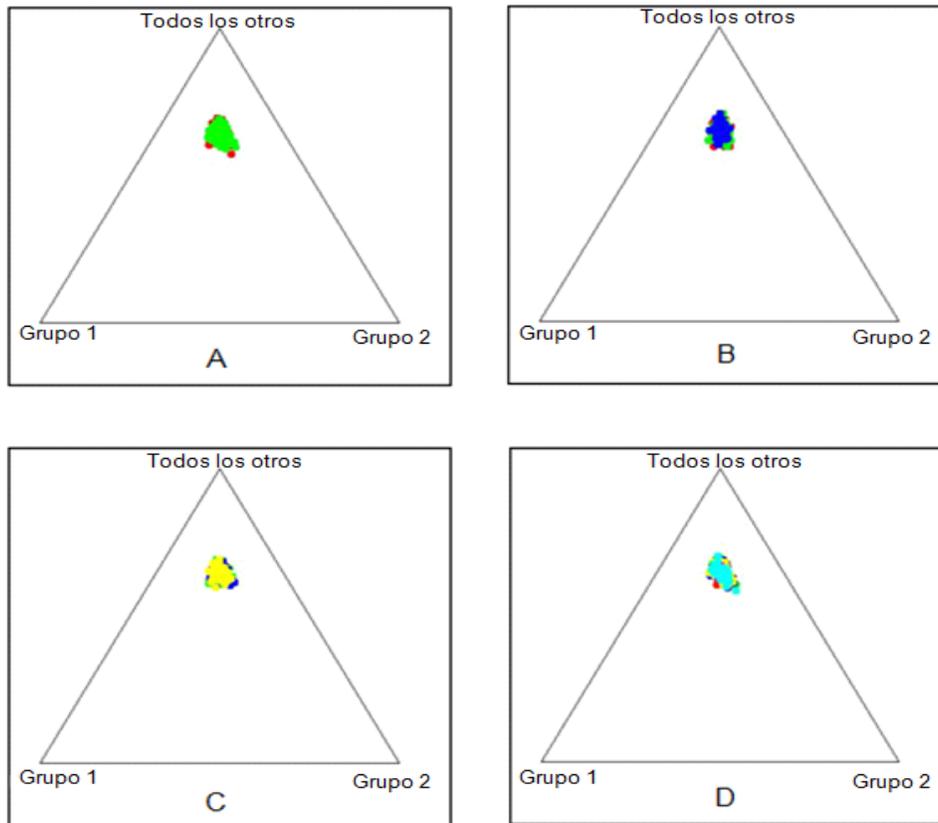
Estrategia de agrupamiento	$F_{ST}$ (%varianza)	p	$R_{ST}$ (%varianza)	p
Dos grupos:				
Alto-Medio/Bajo-Medio	0.01	0.3856	0.01	0.4023
Tres grupos:				
Alto/Medio/Bajo	-0.01	0.6075	-0.12	0.7684
Cuatro grupos:				
Alto/Medio/Medio-Bajo/Bajo	0.01	0.4584	-0.07	0.6596
Seis grupos*:				
1, 2, 3, 4, 5, 6	0.02	0.4689	-0.03	0.5439

\* Los seis grupos corresponden a cada uno de los SES

hipótesis nula de uniformidad genética, corregida por el método de Bonferroni ( $p < 0.00264$ ), mostró que el sistema F13B (valor  $p = 0.00005 \pm 0.00001$ ) (Cuadro 2), tuvo desviaciones de sus frecuencias respecto a HWE debido a un déficit de heterocigotos, siendo el único sistema STR que aportó significativamente a la poca diferenciación detectada, pese a no existir sub-estructura en la muestra total del estudio. Este estudio representa un nuevo aporte a la caracterización de STRs autosómicos en muestras de la ciudad de Bucaramanga, debido al incremento en el tamaño de la muestra con base en el

porcentaje de habitantes por estrato socioeconómico y en el número de otros marcadores analizados adicionalmente: sistemas PENTA D, PENTA E, LPL, F13B, F13A01 y FESFPS, y es el primer informe para las frecuencias de estos últimos cuatro marcadores en poblaciones de Santander.

Los SES analizados conforman la organización demográfica y la división política de la ciudad<sup>11</sup>; el comportamiento de los valores  $F_{ST}$  y  $R_{ST}$  a medida que se subdivide la muestra (Cuadro 3), evidencian que no hay diferencias en los SES en términos de estructura genética



**Figura 1. Resultados obtenidos del análisis con el software STRUCTURE v.2.0.**  
**A: Dos grupos (alto-medio/bajo-medio); B: Tres grupos (alto/medio/bajo);**  
**C: Cuatro grupos (alto/medio/medio-bajo/bajo); D: Seis grupos (1, 2, 3, 4, 5, 6)**

de Bucaramanga, según lo obtenido con STRUCTURE (Figura 1). El uso del  $R_{ST}$  en los casos donde se analizan sistemas de marcadores multilocus, se justifica por el hecho que es un estimador útil de la estructura genética poblacional, cuando las mutaciones han contribuido de manera sustancial a las diferencias alélicas que hay entre poblaciones; mientras que el  $F_{ST}$  además de ser un parámetro de distribución estadística, revela algunas propiedades del proceso evolutivo que han permitido la divergencia entre poblaciones<sup>5</sup>. Pero esto no fue un objetivo del presente estudio.

Los individuos dentro de cada estrato socioeconómico se distribuyen según una investigación con muestras divididas por estratos socioeconómicos de Caracas, donde se demostró la asimetría presente en dicha población, debida a los aportes ancestrales diferenciales en cada estrato socioeconómico, y así se pudo

verificar la sub-estructuración genética en esa población venezolana<sup>22</sup>.

Dentro del rigor metodológico en los estudios de casos y controles, se debe detectar si las frecuencias genotípicas se encuentran en equilibrio HW y si adicionalmente, la población analizada cumple con el principio de no presentar sub-estructuración genética<sup>5</sup>; este último aspecto no se analizó en el trabajo de Arenas *et al.*<sup>23</sup> (donde se evaluó el nivel de la proteína C-Reactiva en el suero asociado con la presencia del gen de angiotensina convertidora de la enzima I/D); tampoco en el de Bautista *et al.*<sup>24</sup> (sobre el polimorfismo del gen de esta misma enzima y el riesgo de infarto de miocardio en Colombia); o en el de Bautista *et al.*<sup>14</sup> (para establecer la prevalencia e impacto de los factores de riesgo cardiovascular en Bucaramanga); ni finalmente, en el estudio de Bautista *et al.*<sup>25</sup> (que analizó el

polimorfismo del gen de renina angiotensina buscando asociarlo con hipertensión arterial esencial entre hispanoamericanos). Todos estos estudios incluyeron en el análisis muestras de participantes de Bucaramanga. Las conclusiones observadas por estos investigadores coincidieron en afirmar que los resultados se ajustaron de acuerdo con la etnicidad auto-reportada por los individuos<sup>24,25</sup>. La auto-asignación de la etnicidad por parte de las personas incluidas en los estudios, podría conducir a hallazgos donde el índice de la enfermedad puede variar, porque la frecuencia alélica difiere entre el caso y el control debido a proporciones de herencia diferenciales (sub-estructura críptica), y es necesario calcular este efecto con metodologías que asuman de forma objetiva, a través de abordajes moleculares el ajuste según el patrimonio étnico, tanto de los casos como de los controles<sup>2</sup>, a fin de reducir la probabilidad de aceptar la hipótesis nula (error tipo I) cuando la alterna es verdadera<sup>2,3</sup>.

En este estudio se controlaron los sesgos, al tomar un número de muestras similares entre poblaciones, donde para cada SES se calcularon tamaños cercanos de muestras, que se tipificaron y validaron con controles internos y externos. En el análisis de los datos obtenidos de tales muestras se siguieron varios modelos para organizar las poblaciones con diferentes abordajes, a fin de definir los grupos objeto de las comparaciones en el análisis.

## CONCLUSIONES

Las poblaciones representadas en los SES exhibieron alta diversidad genética y el análisis molecular de varianza mostró una baja diferenciación entre los SES de Bucaramanga. Esto demuestra a la población como una misma unidad genética. Asimismo, la presencia de equilibrio HW en la población permitió concluir con alta probabilidad, que los hallazgos con los trabajos de asociación, proporcionan resultados válidos al evaluar las interacciones entre los polimorfismos y los factores de riesgo analizados en la población del estudio. Estas razones contribuirán a obviar, en lo que se refiere a trabajos hechos en esta región geográfica, el inconveniente de aparentes asociaciones validadas o ajustadas, a débiles inferencias de la ausencia de subestructura o la falta de reproducibilidad en los hallazgos de los estudios de enfermedades complejas, en las investiga-

ciones de carácter biomédico y evolutivo.

También es importante considerar la habilidad de los investigadores, al escoger apropiadamente los sujetos en una población objeto de estudios epidemiológicos moleculares, donde se debe conocer los orígenes de la muestra bajo estudio para tener un adecuado criterio de selección, permitiendo la interpretación sólida de los resultados y su relación con los procesos genéticos históricos involucrados en enfermedades humanas complejas. Es relevante el hecho de verificar los análisis de asociaciones de alelos con respecto a una enfermedad dentro de investigaciones de casos y controles que involucran poblaciones mezcladas (por la dificultad de asignar individuos a un grupo o a otro, basado en uno o más marcadores). En concordancia con lo anterior, los estudios de asociación se destacan cada vez más en el campo de las investigaciones biomédicas para poblaciones heterogéneas como Bucaramanga, cuantificando la magnitud de riesgos asociados con la enfermedad con alelos específicos en términos relativos, absolutos y atribuibles al grupo poblacional, disminuyendo así el impacto de las variables de confusión.

Finalmente, es importante resaltar, que los resultados de este trabajo deben ser tenidos en cuenta en el planteamiento de estudios de casos y controles bajo las condiciones actuales de la población de Bucaramanga. La realización de investigaciones a futuro debería incluir el análisis de otros marcadores biparentales en esta población, por ejemplo análisis de SNPs, con las cuales se realizarían asignaciones más precisas de la diversidad presente en condiciones genéticas heredables implicadas en enfermedades y además permitirían verificar la incidencia de los diferentes componentes poblacionales de una región geográfica en particular. Todo lo anterior justificaría diseñar paneles de marcadores moleculares con los cuales se pueda verificar la ancestría, confirmar los análisis realizados de la estructura genética en poblaciones mezcladas y validar las conclusiones planteadas en los estudios de casos y controles realizados hasta ahora en estas poblaciones.

**Conflicto de intereses.** Los autores declaran que no hay conflicto de intereses en el presente manuscrito.

## AGRADECIMIENTOS

Los autores manifiestan su agradecimiento a la Vicerrectoría de Investigación y Extensión de la Uni-

versidad Industrial Santander (UIS) y al Instituto de Patología e Inmunología Molecular de la Universidad de Porto (Portugal) IPATIMUP.

## REFERENCIAS

1. The human Genome Project where did it come from, where is it going? *Am J Hum Genet.* 1992; 51: 1-6.
2. Cordell HJ, Clayton DG. Genetic association studies. *Lancet.* 2005; 366: 1121-31.
3. Pritchard JK, Rosenberg NA. Use of unlinked genetic markers to detect population stratification in association studies. *Am J Hum Gen.* 1999; 65: 220-8.
4. Devlin B, Roeder K, Wasserman L. Genomic control, a new approach to genetic based association studies. *Theor Popul Biol.* 2001; 60: 155-66.
5. Holsinger KE, Weir BS, Genetic in geographically structured populations: defining, estimating and interpreting  $F_{ST}$ . *Nat Rev Genet.* 2009; 10: 639-50.
6. Gómez MV, Reyes ME, Cárdenas H, García O. Genetic variation for 7 STR loci in a Colombian population (Department of Valle del Cauca). *J Forensic Sci.* 2003; 48: 687-8.
7. Bravo ML, Moreno MA, Builes JJ, Salas A, Lareu MV, Carracedo A. Autosomal STR genetic variation in negroid Chocó and Bogotá populations. *Int J Legal Med.* 2001; 115: 102-4.
8. Yunis JJ, García O, Cuervo A, Guio E, Pineda C, Yunis EJ. Population data for powerplex 16 in thirteen departments and the capital city of Colombia. *J Forensic Sci.* 2005; 50: 1-18.
9. Departamento Nacional de Planeación. *Estratificación socioeconómica. Manual general. Cabeceras municipales tipo 3 y localidades o centros poblados hasta con tres mil habitantes.* Bogotá: Departamento Nacional de Población. 1994.
10. Dean AG, Dean JA, Coulumbier D, Dicker RC, Sullivan KM, Fagan RF. *Epi Info ver. 6.04.* Un programa procesador de texto, bases de datos y estadísticas para salud pública con micro-computadores compatibles IBM. Georgia: *Center for Disease Control Atlanta*; 2001
11. Departamento Administrativo Nacional de Estadística (DANE). Los grupos étnicos de Colombia en el censo de 1993 (in Spanish). Memorias. Santa Fe de Bogotá, República de Colombia. 2005. DANE, datos del censo de Colombia del 2005; [Artículo en internet] 2008 [fecha de acceso 10 septiembre 2009]. Disponible en: <http://www.dane.gov.co>
12. Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acid Res.* 1998; 16: 1215-6.
13. Bär W, Brinkmann B, Budowle B, Carracedo A, Gill P, Lincoln P, et al. DNA recommendations. Further report of the DNA Commission of the ISFG regarding the use of short tandem repeat systems. *Forensic Sci Int.* 1997; 87: 174-9.
14. Bautista LE, Oróstegui M, Vera LM, Prada GE, Orozco LC, Herrán OF. Prevalence and impact of cardiovascular risk factors in Bucaramanga, Colombia: results from the Countrywide Integrated Noncommunicable Disease Intervention Programme (CINDI/CARMEN) baseline survey. *Eur J Cardiovasc Prev Rehabil.* 2006; 13: 769-75.
15. Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. *BMJ.* 1995; 310: 170-3.
16. Raymond ML, Rousset F. An exact test for population differentiation. *Evolution.* 1995; 49: 1280-3.
17. Excoffier L, Laval G, Schneider S. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evol Bioinform.* Online. 2005; 1: 47-50.
18. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics.* 2000; 155: 945-59.
19. Vargas CI, Castillo A, Gil AM, Pico AL, García O. Population genetic data for 13 STR loci in a northeast Colombian (department of Santander) population. *J Forensic Sci: Int Congress Series.* 2003; 1239: 197-200.
20. Vargas CI, Castillo A, Gil AM, Pico AL, García O. Genetic variation for Penta D and Penta E in a Northeast Colombian population (department of Santander). *J Forensic Sci.* 2006; 51: 438-9.
21. Paredes M, Galindo A, Bernal M, Ávila S, Andrade D, Vergara C, et al. Analysis of the CODIS autosomal STR loci in four main Colombian regions. *J Forensic Sci Int.* 2003; 137: 67-73.
22. Martínez H, Rodríguez-Larralde A, Izaguirre MH, Castro D. Admixture estimates for Caracas, Venezuela, based on autosomal, Y-chromosome, and mtDNA markers. *Hum Biol.* 2007; 79: 201-13.
23. Arenas IA, Vargas CI, Davidge ST, Bautista LE. Angiotensin converting enzyme I/D gene polymorphism influences serum C-reactive protein level. *Can J Cardiol.* 2006; 21: 112-6.
24. Bautista LE, Ardila ME, Gamarra G, Vargas CI, Arenas IA. Angiotensin converting enzyme gene polymorphism and risk of myocardial infarction in Colombia. *Med Sci Monit.* 2004; 10: 143-9.
25. Bautista LE, Vargas CI, Oróstegui M, Gamarra G. Population-based case-control study of renin-angiotensin system genes polymorphisms and hypertension among Hispanics. *Hypert Res.* 2008; 31: 401-7.