

**ANÁLISIS POBLACIONAL DE MARCADORES AUTOSÓMICOS INDELS EN  
UNA MUESTRA DE INDIVIDUOS DEL DEPARTAMENTO DE SANTANDER,  
COLOMBIA**

**NATHALIA ANDREA TRUJILLO PELAYO**

**UNIVERSIDAD INDUSTRIAL DE SANTANDER  
FACULTAD DE CIENCIAS  
ESCUELA DE BIOLOGIA  
BUCARAMANGA**

**2016**

**ANÁLISIS POBLACIONAL DE MARCADORES AUTOSÓMICOS INDELS EN  
UNA MUESTRA DE INDIVIDUOS DEL DEPARTAMENTO DE SANTANDER,  
COLOMBIA**

**NATHALIA ANDREA TRUJILLO PELAYO**

**Trabajo de grado para optar por el título de:  
Bióloga**

**Directora**

**CLARA INÉS VARGAS CASTELLANOS**

**Médica. Magister en Biología con énfasis en Genética Clínica  
Magister en Epidemiología**

**UNIVERSIDAD INDUSTRIAL DE SANTANDER**

**FACULTAD DE CIENCIAS**

**ESCUELA DE BIOLOGIA**

**BUCARAMANGA**

**2016**

A la memoria de quién fue mi más grande tesoro  
en la vida y quién ahora en la distancia me  
acompaña... mi Madre,  
URANIA

## AGRADECIMIENTOS

No hay palabras para expresar a Dios la gratitud que siento al finalizar esta etapa de mi vida. A El quién ha sido siempre mi fortaleza para continuar y aunque no fue fácil el camino, me sostuvo y puso en mi vida grandes ángeles que de una u otra forma han hecho parte de este lindo culminar. A mi gran ángel celestial, mi querida madre, quién desde siempre creyó en mí, en mis capacidades, quien fue mi motor y mi sustento, quién mantuvo siempre vivo mi deseo de estudiar en la Universidad Industrial de Santander, desde que solo era una niña y a quién le debo todo lo que he llegado a ser. Hoy echo de menos su presencia, pero sé que desde el cielo celebra con orgullo este día.

A mis queridos docentes y amigos, la Doctora Clara Inés Vargas Castellanos y el Doctor Fernando Rondón González, para quiénes tengo un gran afecto. No solo me acogieron como su estudiante, también me brindaron todo su cariño fraternal y con sus enseñanzas me ayudaron a ser mejor cada día, a no dejarme vencer por los obstáculos por más grandes que parecieran. Con sus consejos y su guía logré finalizar este proyecto y son para mí un gran ejemplo a seguir dentro de este lindo campo de la Genética.

A cada una de las personas que hacen parte del grupo de investigación de Genética Humana-UIS, especialmente a las docentes Adriana María Gil Zapata y Adriana Lucía Pico Romero, quienes me acogieron, enseñaron y guiaron mi proceso de aprendizaje en el laboratorio. Un agradecimiento especial a Adriana Gil, quién me guio espiritualmente y fue un pilar fundamental para finalizar este proceso. Nuestras largas jornadas de trabajo en el laboratorio me enseñaron que con firmeza y rigor todo se puede lograr. A la Doctora Adriana Castillo Pico, Viviana, Valentina y Mary, quienes me brindaron todo su apoyo y fueron parte de este sueño que ahora se hace realidad.

A cada uno de mis docentes de la Escuela de Biología, que con sus enseñanzas y rigor ayudaron en mi proceso de formación como Bióloga. A mis compañeros de carrera, especialmente a mi querida amiga Liliana Gutiérrez, quién con sus consejos y apoyo, me ayudo a salir siempre adelante y finalmente hemos logrado terminar esta etapa de la vida.

A mi familia, que con todo su amor me han acompañado y apoyado desde el inicio. Mi padre, hermanas, mi chiquito y esa nueva luz en mi vida que me hacen ser mejor cada día y a tener la fortaleza de continuar sin desfallecer en el intento.

## TABLA DE CONTENIDO

	Página
INTRODUCCIÓN	20
1. OBJETIVOS	22
1.1 OBJETIVO GENERAL	22
1.2 OBJETIVOS ESPECÍFICOS	22
2. MARCO TEÓRICO	23
2.1 POLIMORFISMOS DE ADN	23
2.2 POLIMORFISMOS DE INSERCIÓN-DELECIÓN, ASPECTOS GENERALES	25
2.3 ASPECTOS MOLECULARES Y CAMPOS DE APLICACIÓN	26
2.3.1 Pruebas de paternidad	28
2.3.2 Análisis de muestras degradadas e identificación humana	29
2.3.2.1 Probabilidad de exclusión (PE)	30
2.3.2.2 Poder de discriminación (PD).	30
2.3.2.3 Contenido de información polimórfica (PIC)	30
2.3.2.4 Probabilidad de coincidencia (PM)	30
3. HIPOTESIS DE INVESTIGACIÓN	32

4.	MATERIALES Y MÉTODOS	33
4.1	MUESTREO	33
4.1.1	Población de estudio	33
4.1.2	Población objeto de estudio.	34
4.1.3	Colección de muestras.	35
4.1.4	Criterio de selección de muestras.	35
4.2	FASE DE LABORATORIO	36
4.2.1	Toma de muestra y extracción de ADN.	37
4.2.2	Genotipificación.	37
4.2.3	Selección de marcadores genéticos.	37
4.2.4	Información de los Loci.	38
4.2.5	Amplificación y detección de marcadores.	38
4.2.6	Análisis estadístico.	39
4.3	ASPECTOS ÉTICOS	40
5.	RESULTADOS	41
5.1	BASE DE DATOS	41
5.2	VARIACIÓN GENÉTICA DE LA POBLACIÓN	41
5.2.1	Frecuencias alélicas.	41
5.2.2	Frecuencias genotípicas.	42
5.3	ANÁLISIS DE LA DIVERSIDAD	43
5.4	EVALUACIÓN DE LAS DISTANCIAS GENÉTICAS	44
5.5	EVALUACIÓN DEL EQUILIBRIO DE HARDY WEINBERG	45

5.6 ANÁLISIS DE LA ESTRUCTURA GENÉTICA Y ASIGNACIÓN DE INDIVIDUOS A POBLACIONES	46
5.7 EVALUACIÓN DEL DESEQUILIBRIO DE LIGAMIENTO	48
5.8 ANÁLISIS DE EFICIENCIA FORENSE	49
6. DISCUSIÓN	50
7. CONCLUSIONES	56
8. RECOMENDACIONES	57
BIBLIOGRAFIA	58
ANEXOS	65

## LISTA DE FIGURAS

	Página
<b>Figura 1.</b> División geopolítica del departamento de Santander.	36
<b>Figura 2.</b> Matriz de diferenciación genética ( $F_{ST}$ ).	44
<b>Figura 3.</b> Representaciones gráficas de la matriz de Slatkin ( $F_{ST}$ ), y estimaciones de ancestría entre Santander y poblaciones de referencia (Europa, África, Asia y Rio de Janeiro), obtenidas mediante los SW Arlequín V3.5 y Structure V2.3.3.	48
<b>Anexo D: Figura 4.</b> Electroferograma de amplificación para los 38 sistemas autosómicos de inserción-delección, usando 0.25 ng de ADN de un individuo nativo del departamento de Santander.	67

## LISTA DE TABLAS

Página

<b>Tabla 1.</b> Diversidad y distancia genética de las seis provincias regionales de Santander ( $F_{ST}$ ).	43
<b>Tabla 2.</b> Análisis de Varianza molecular (AMOVA) para la población del departamento de Santander. Estimación de la estructura poblacional y evaluación de la variación genética en la población.	46
<b>Anexo A: Tabla 3.</b> Información específica del panel de 38 marcadores autosómicos INDELS desarrollado por Pereira et al. 2009b.	65
<b>Anexo B: Tabla 4.</b> Protocolo de amplificación mediante el kit de PCR multiplex de Quiagen.	66
<b>Anexo C: Tabla 5.</b> Protocolo de amplificación, técnica de reacción de PCR multiplex (Pereira et al. 2009b).	67
<b>Anexo E: Tabla 6.</b> Perfil genético de las líneas celulares humanas 9947A y 9948 (Promega).	68
<b>Anexo H: Tabla 7.</b> Datos genotípicos poblacionales y parámetros estadísticos forenses usando un panel de 38 marcadores autosómicos de inserción-delección en Santander, Colombia.	73
<b>Anexo I: Tabla 8.</b> Frecuencias alélicas para 38 marcadores autosómicos INDELS evaluadas en cada una de las regiones provinciales del departamento de Santander.	75

**Anexo J: Tabla 9.** Frecuencias genotípicas para 38 marcadores autosómicos INDELS evaluadas en una muestra poblacional del departamento de Santander. 77

**Anexo K: Tabla 10.** Análisis de los estadísticos F y sus respectivos valores de significancia encontrados en una muestra poblacional del departamento de Santander. 78

**Anexo L: Tabla 11.** Evaluación de probabilidades a priori ( $L_nP$ ) y posteriori ( $P_k$ ), en el análisis de asignación de individuos a las poblaciones. 79

## LISTA DE ANEXOS

	Página
<b>Anexo A: Tabla 3.</b> Información específica del panel de 38 marcadores autosómicos INDELS desarrollado por Pereira et al. 2009b.	65
<b>Anexo B: Tabla 4.</b> Protocolo de amplificación mediante el kit de PCR multiplex de Quiagen.	66
<b>Anexo C: Tabla 5.</b> Protocolo de amplificación, técnica de reacción de PCR multiplex (Pereira et al. 2009b).	67
<b>Anexo D: Figura 4.</b> Electroferograma de amplificación para los 38 sistemas autosómicos de inserción-delección, usando 0.25 ng de ADN de un individuo nativo del departamento de Santander.	67
<b>Anexo E: Tabla 6.</b> Perfil genético de las líneas celulares humanas 9947A y 9948 (Promega).	68
<b>Anexo F:</b> Permiso otorgado por el Comité de Ética de la Universidad Industrial de Santander para el desarrollo de la presente investigación.	69
<b>Anexo G:</b> Consentimiento informado de los individuos participantes del presente estudio. Este documento fue avalado por el Comité de Ética de la Universidad Industrial de Santander.	70
<b>Anexo H: Tabla 7.</b> Datos genotípicos poblacionales y parámetros estadísticos forenses usando un panel de 38 marcadores autosómicos de inserción-delección en Santander, Colombia.	73

<b>Anexo I: Tabla 8.</b> Frecuencias alélicas para 38 marcadores autosómicos INDELS evaluadas en cada una de las regiones provinciales del departamento de Santander.	75
<b>Anexo J: Tabla 9.</b> Frecuencias genotípicas para 38 marcadores autosómicos INDELS evaluadas en una muestra poblacional del departamento de Santander.	77
<b>Anexo K: Tabla 10.</b> Análisis de los estadísticos F y sus respectivos valores de significancia encontrados en una muestra poblacional del departamento de Santander.	78
<b>Anexo L: Tabla 11.</b> Evaluación de probabilidades a priori ( $L_nP$ ) y posteriori ( $P_k$ ), en el análisis de asignación de individuos a las poblaciones.	79

## GLOSARIO

**AIMs:** Del ingles “Ancestry Informative Markers”. Polimorfismos que exhiben sustancialmente diferentes frecuencias entre poblaciones y son muy informativos para la asignación de ascendencia.

**AMOVA:** Análisis de Varianza Molecular. Evalúa el grado de estructura genética presente en las poblaciones.

**ADN:** Acido Desoxirribonucleico. Material genético que se encuentra en el núcleo celular.

**ADNmt:** Acido Desoxirribonucleico mitocondrial.

**DANE:** Departamento Nacional Administrativo de Estadística.

**HETEROCIGOSIDAD:** Medida de la variación genética de una población respecto a un locus particular.

**HETEROCIGOTO:** Estado en el que los alelos del mismo locus en los cromosomas homólogos son diferentes.

**HOMOCIGOTO:** Estado en el que los alelos del mismo locus en los cromosomas homólogos son idénticos.

**HW:** Equilibrio de Hardy-Weinberg. Hipótesis que establece que en una población suficientemente grande, los apareamientos se producen al azar y no se encuentra sometida a ninguna fuerza evolutiva.

**INDELS:** Del inglés “Insertion- Deletion Polymorphism”. Polimorfismos de inserción, deleción, o inserción y deleción de bases presentes en regiones codificantes o no codificantes del ADN.

**PCR:** Reacción en cadena de la Polimerasa. Técnica de amplificación de pequeños fragmentos de ADN.

**SNP:** Del inglés “Single Nucleotide Polymorphism”. Variación en la secuencia de ADN que afecta a una sola base de una secuencia del genoma

**STRs:** Del inglés “Short Tandem Repeats”. Secuencias de ADN en las que un fragmento (dos hasta seis pares de bases) se repite de manera consecutiva.

**TÍTULO:** ANÁLISIS POBLACIONAL DE MARCADORES AUTOSÓMICOS INDELS EN UNA MUESTRA DE INDIVIDUOS DEL DEPARTAMENTO DE SANTANDER, COLOMBIA\*

**AUTOR:** NATHALIA ANDREA TRUJILLO PELAYO\*\*

**PALABRAS CLAVES:** Polimorfismos de inserción- delección INDELS, identificación humana, Santander, estructura de la población, mezcla.

## **RESUMEN**

Los polimorfismos de inserción-delección (INDELS), son una forma alternativa de variación genética natural que se han convertido en el foco de numerosas investigaciones durante los últimos años. Los INDELS presentan características potenciales como baja tasa de mutación, alta capacidad multiplex y facilidad de genotipificación de muestras de ADN altamente degradadas que los hacen más adecuados para investigar eventos de mezcla en las poblaciones y una herramienta útil en el campo forense o de identificación humana.

A partir del análisis de 38 INDELS autosómicos, se estableció el grado de estructura genética en Santander a través de la tipificación de 500 individuos sanos no relacionados entre sí. Las frecuencias alélicas y parámetros genéticos tales como heterocigosidad, prueba exacta de Hardy Weinberg, análisis de distancia genética y desequilibrio de ligamiento, fueron reportados. De igual forma, parámetros estadísticos como poder de discriminación, probabilidad de coincidencia y probabilidad de exclusión fueron estimados.

Desviaciones significativas del equilibrio HW fueron encontradas en seis de los 38 marcadores estudiados ( $P < 0,01$ ). El AMOVA indicó que el componente individual de la variación genética representa el 96.08% de la diversidad genética total de la población, lo cual está relacionado con el bajo valor de  $F_{ST}$  obtenido. Además, al estimar el poder de discriminación y exclusión combinado del panel se obtuvieron resultados de 1 y 0,99619147, respectivamente, validando la eficiencia forense del panel.

De acuerdo a lo anterior, Santander sigue siendo la misma unidad genética reproductiva sin subestructura. Así mismo, la eficiencia y facilidad de trabajo de estos marcadores los convierten en una herramienta interesante para futuros estudios de identificación humana y variación genética en poblaciones humanas. Finalmente, la base de datos actual será la primera reportada para Santander y será útil para propósitos forenses y de paternidad.

---

\*Tesis de grado

\*\* Facultad de Ciencias. Escuela de Biología. Directora: Clara Inés Vargas Castellanos, Médico

**TITLE:** POPULATION ANALYSIS OF INDELS MARKERS IN A SAMPLE OF INDIVIDUALS FROM SANTANDER, COLOMBIA\*

**AUTHOR:** NATHALIA ANDREA TRUJILLO PELAYO\*\*

**KEY WORDS:** Insertion-deletion polymorphisms, human identification, Santander, population structure, admixture

### **ABSTRACT/ SUMMARY**

As complementary markers of Short Tandem Repeats (STRs) and Single Nucleotide Polymorphisms (SNPs), Insertion-Deletion polymorphisms (INDELS), an alternative form of natural genetic variation, have become the focus of numerous research during the recent years. INDELS have potential features like low mutation rates, multiplexing capability and the ease of genotyping highly degraded DNA samples, which make them more suitable to investigate population admixture events and a useful tool in the field of human identification.

Based on the analysis of 38 autosomal INDELS, we established the degree of genetic structure in Santander through the typing of 500 healthy unrelated individuals. We reported allele frequencies and genetic parameters such as heterozygosity, Hardy Weinberg exact test, genetic distance analysis and linkage disequilibrium. We also, evaluated statistical parameters such as discrimination power, match probability and exclusion probability in order to detect the forensic efficiency of this panel.

Significant deviation from Hardy–Weinberg expectations were found in six of the 38 markers studied ( $P < 0.01$ ). The AMOVA indicated that the individual component of genetic variation represents 96.08% of the total genetic diversity which is related to the low value of  $F_{ST}$  found in the population. Also, the combined discrimination power and power of exclusion for the whole panel were 1 and 0.99619147, respectively.

Santander remains as the same reproductive genetic unit with no substructure. The efficiency and ease of these markers make them a useful and interesting tool for future studies of human identification and genetic variation in natural populations. The present database will be the first database released for Santander, Colombia with these types of markers and will be useful for forensic and paternity purposes.

---

\* Bachelor Thesis

\*\* Faculty of Science. Biology School. Tutor: Clara Inés Vargas Castellanos, Medico

## INTRODUCCIÓN

Durante los últimos años, múltiples investigaciones han sido realizadas a partir del análisis de polimorfismos de inserción-delección (INDELs), convirtiéndose en un foco de interés dentro de la investigación científica por sus potenciales características moleculares tales como baja tasa de mutación, capacidad multiplex y facilidad de genotipificación de muestras de ADN altamente degradadas. Estas características los hacen más adecuados en la evaluación de estudios como eventos de mezcla de poblaciones y una herramienta novedosa en el campo de la identificación humana.

A partir de esto, diferentes investigaciones como la evaluación del grado de estructura genética en poblaciones humanas (Rosenberg et al. 2005; Tishkoff et al. 2009), su uso como marcadores informativos de ascendencia (Yang et al. 2005; Freitas et al. 2010), diseño de paneles (Ribeiro-Rodrigues et al. 2009; Freitas et al. 2010) y estudios de identificación humana (Pereira et al. 2009b; Pimenta et al. 2010) han sido realizadas mediante el uso de dichos marcadores.

En Colombia, la evaluación del grado de estructura genética en poblaciones humanas ha sido ampliamente estudiada mediante el uso de marcadores tipo SNPs y STRs, entre los cuales se destacan las investigaciones realizadas por Yunis et al. (2000); Peña (2007); Rojas et al. (2011); Castillo et al. (2013 y 2015); Ibarra et al. (2014). Sin embargo, información que reporte la identificación y distribución de formas alternativas de variación genética natural en la población colombiana, como lo son los polimorfismos de inserción-delección, aún se mantiene escasa.

Por otra parte, Santander es uno de los departamentos con mayor número de habitantes en el país, localizado en la región noreste de Colombia (DANE 2005) y su población resultó principalmente de la mezcla entre amerindios que habitaban la

región y europeos quienes llegaron durante el periodo de la colonización (Builes et al. 2006). Debido a esta condición de mezcla en la población, Santander ha sido objeto de estudios anteriores de tipo genético poblacional, a partir del análisis de marcadores microsatélites y SNPs. Hincapié et al. 2009, Rondón et al. 2009 y más recientemente Castillo et al. 2013; 2015, lograron estudiar el grado de variabilidad genética mediante el análisis de dichos sistemas observados en la población, sin embargo, hasta el momento no se reporta información de alguna forma alterna de variación genética, como los INDELS, presente en la población.

La necesidad de generar nueva información acerca de la evaluación del grado de la estructura genética en la población mediante el análisis de dichos polimorfismos, creó un primer acercamiento al conocimiento de la distribución genética de 38 marcadores autosómicos INDELS y permitió establecer la caracterización genética de la población santandereana estableciendo su relación con otros grupos poblacionales a partir del análisis de la distribución de las frecuencias alélicas.

Este estudio contribuyó en la generación de la primera base de datos reportada para el departamento y de las pocas obtenidas en el país. Así mismo, la evaluación de la eficiencia forense de este panel de marcadores, permitió implementarlos como una herramienta valiosa para la identificación de perfiles genéticos en muestras degradadas y una herramienta novedosa dentro de los campos de estudio clínico y forense en Colombia.

## **1. OBJETIVOS**

### **1.1 OBJETIVO GENERAL**

Establecer la estructura genética de un grupo poblacional procedente de las seis provincias regionales del departamento de Santander a través de la tipificación de 38 marcadores autosómicos de Inserción-Delección (INDELS).

### **1.2 OBJETIVOS ESPECÍFICOS**

Evaluar el equilibrio Hardy Weinberg y comparar las frecuencias alélicas esperadas y observadas en una muestra poblacional de 500 individuos de las seis provincias regionales del departamento de Santander.

Ampliar el conocimiento de la diversidad genética de la población Santandereana mediante la estimación de la estructura y distancias genéticas, así como también mediante el análisis de las heterocigosidades esperadas y observadas de las seis provincias regionales del departamento de Santander.

Establecer los parámetros de importancia forense que permitan a futuro usar estos marcadores en identificación humana.

## **2. MARCO TEÓRICO**

A partir de la década de los 90'S, los avances en la investigación de la genética de poblaciones y evolutiva han sido consecuencia del advenimiento de nuevas herramientas moleculares y la mejora continua en la manera en que similitudes o disimilitudes genéticas entre genomas ha sido evaluada (Valí et al. 2008). Visto desde una perspectiva temporal, el uso de marcadores genéticos ha pasado de centrarse en los fenotipos a los genotipos, y después de su introducción, al estudio de las poblaciones naturales mediante la evaluación de las frecuencias alélicas y la distribución que presentan varios marcadores moleculares en estas (Rondón y Barreto 2013).

### **2.1 POLIMORFISMOS DE ADN**

La secuencia de ADN humano abarca distintos tipos de variación genética, cada uno presentando características deseables para abordar diferentes aspectos biológicos. Por ejemplo, en el ADN nuclear encontramos los STRs que corresponden a regiones de ADN repetitivo de 2 a 7 pares de bases (pb), repartidas en todo el genoma y presentan un alto grado de polimorfismo de longitud cuya base molecular es la variación en el número de regiones génicas como extragénicas (Ellegren 2004). Debido a que son muy diversos, este tipo de marcadores genéticos son muy informativos y ampliamente utilizados en la genética de poblaciones y forense (Pereira et al. 2009b). Los microsatélites (STRs) han sido los marcadores genéticos base para muchas investigaciones en las que se busca la relación entre los individuos, poblaciones o especies (Ellegren 2004).

De igual forma, los polimorfismos de un único nucleótido (SNPs), son marcadores de secuencia que se producen por el cambio de un nucleótido debido a una

mutación puntual en una secuencia de ADN (Rondón y Barreto 2013). Además, los SNPs son polimorfismos abundantes y frecuentes y sus características los convierten en marcadores deseables para estudios de análisis de diversidad genética, el establecimiento de componentes de mezcla y la relación que guardan con enfermedades genéticas, dada la estabilidad que presentan por su baja tasa de mutación (Santos y Tyler 1996; Sherry et al. 1999).

Por su parte, el DNA<sub>mt</sub> se caracteriza por su herencia uniparental, lo que permite determinar el posible origen matrilineo extracontinental de una muestra particular (Rondón y Barreto 2013). Esta molécula se puede analizar mediante la evaluación de RFLP-PCR, un marcador molecular que mediante el uso de enzimas de restricción logra reconocer secuencias específicas en el ADN y produce fragmentos comparables en tamaño (Botstein et al. 1980) permitiendo identificar la presencia o ausencia de mutaciones asociadas al origen de las muestras (Rondón y Barreto 2013).

No obstante, dichos marcadores exhiben ciertas dificultades técnicas y analíticas, a pesar de su utilidad. El patrón de mutación complejo y heterogéneo de los STRs podría presentar ambigüedades para el análisis detallado de los datos (Ellegren 2004), por ejemplo, algunos errores de genotipificación podrían ocurrir debido a la presencia de stutter bands y artefactos técnicos tales como alelos nulos o falsos, los cuales son bastante comunes en la genotipificación de dichos marcadores (Syvänen 2001). Por su parte los SNPs, presentan limitaciones como la necesidad de tener cerca de 60 SNPs balanceados para lograr obtener un poder de discriminación similar a los kits multiplex de STRs empleados en genética forense (Ibarra et al. 2014). Además, a pesar de los diferentes métodos de genotipado disponibles (Syvänen 2001), la mayoría de ellos son relativamente costosos a pequeñas o medianas escalas, y requiere de equipos especiales para un genotipado de alto rendimiento (Valí et al. 2008).

Por otra parte, la introducción de nuevos marcadores genéticos para el estudio de las poblaciones naturales ha seguido generalmente un desarrollo metodológico (Valí et al. 2008). Por ejemplo, estudios realizados por Weber et al. (2002) y Mills et al. (2006) contribuyeron fundamentalmente a nuestro conocimiento sobre los INDELS, destacando su utilidad en la mayoría de estudios genéticos, dada su abundancia en el genoma y su facilidad de análisis.

En este trabajo se dedica especial atención a los polimorfismos de inserción-delección, en particular a su eficiencia en estudios de análisis de variación genética e investigaciones de casos forenses

## **2.2 POLIMORFISMOS DE INSERCIÓN-DELECCIÓN, ASPECTOS GENERALES**

En el 2002, Weber identificó y caracterizó ~ 2000 INDELS bialélicos humanos con diferente variación de longitud y medidas de las frecuencias alélicas de individuos europeos, africanos, japoneses y americanos nativos. Más tarde, en el 2006, se informó del primer mapa de variación polimórfica tipo INDEL en el genoma humano (Mills et al. 2006), que comprendió más de 415.000 polimorfismos. Actualmente, existe un creciente interés de dichos marcadores dentro de la investigación genómica (Bhangale et al. 2005; Mills et al. 2006).

Este tipo de polimorfismos, son abundantes en los genomas de organismos modelos de investigación genética y de igual forma en el genoma humano. Por ejemplo, estudios de variación genética en *Drosophila melanogaster* y *Caenorhabditis elegans* han mostrado que los INDELS representan entre el 16% y 25% de todos los polimorfismos en estas especies (Berger et al. 2001; Wicks et al. 2001). Por otra parte, un estudio de variación genética del cromosoma 22 sugirió que el genoma humano contiene niveles similares de variación de este tipo de polimorfismo (los INDELS representaron el 18% de los polimorfismos para este cromosoma; Dawson et al. 2001). Teniendo en cuenta estas estimaciones, los

INDELs representarían entre el 16% y 25% de todos los polimorfismos del genoma humano (Mills et al. 2006) siendo altamente informativos.

Durante los últimos años, mayor información ha sido generada sobre estos marcadores, y su uso dentro de la investigación científica ha crecido como una herramienta valiosa en el análisis de estudios de variabilidad genética, enfermedades genéticas complejas, relaciones ancestrales extracontinentales e identificación humana.

### **2.3 ASPECTOS MOLECULARES Y CAMPOS DE APLICACIÓN**

Teniendo en cuenta aspectos mencionados anteriormente, se pone de manifiesto que algunos tipos de INDELs combinan características interesantes de otros marcadores genéticos comúnmente utilizados en medicina forense como los STRs y SNPs (Pereira y Gusmão 2012). Sin embargo, los INDELs exhiben atractivas particularidades como:

- Son abundantes y ampliamente distribuidos a lo largo del genoma
- Los INDELs muy pequeños, pueden ser analizados en amplicones cortos, mejorando la eficiencia de las amplificaciones de ADN degradado
- Su genotipado, aun con INDELs muy pequeños, es relativamente fácil y económico
- Pequeños INDELs también son adecuados para el análisis con tecnologías de alto rendimiento (Pereira y Gusmão 2012)
- Dada su baja tasa de mutación, estos marcadores permiten resolver casos de paternidad complejos con muy pocas inconsistencias alélicas que podrían ser explicadas por eventos mutacionales (3-5 artículo 1 fuera de la carpeta, Shedlock y Okada 2000).

- Presentan frecuencias alélicas significativamente diferentes en poblaciones geográficamente separadas, lo cual los hace ser utilizados como marcadores de información ancestral (AIMs) (Santos et al. 2010).

Gracias a sus características los INDELs presentan una amplia utilidad, desde la identificación de variantes genéticas asociadas a enfermedades complejas, hasta la evaluación de análisis de variabilidad genética, estudios de ancestría e identificación humana, según su tipo.

En el campo clínico, los estudios de asociación genética se realizan frecuentemente en poblaciones de individuos no relacionados para la identificación de loci susceptibles a patologías complejas (Saloum, et al. 2013b). Actualmente, los INDELs están revelando fuentes previamente desconocidas de la diversidad genética que pueden ser factor importante en la determinación de enfermedades hereditarias en los seres humanos (Weber et al. 2002; Mills et al. 2011).

Por otra parte, los principios genéticos y estadísticos subyacentes a la genética de poblaciones son en sus mayores partes simples y directos (Rondón y Barreto 2013). El análisis de las frecuencias alélicas y genotípicas permite verificar el grado de polimorfismo de los marcadores, caracterizando así la variabilidad genética de las poblaciones. Sewall Wright logró demostrar que en el estudio de la estructura genética poblacional, la variación de la frecuencia génica puede ser evaluada mediante el uso de los índices de fijación ( $F$ ) (Wright 1965), y la medida del grado de diferenciación genética entre poblaciones respecto a la población total por el índice de subestructuración poblacional ( $F_{ST}$ ) (Nei 1973).

En el campo de la genética de poblaciones, los INDELs tipo AIMs son usados en la estimación de la proporción de ancestría en poblaciones mezcladas y también evalúa la estructura genética de dichas poblaciones (Saloum et al. 2013b). La investigación de perfiles de ascendencia genética en poblaciones humanas, es una

herramienta de gran impacto para tratar de entender la dinámica de los procesos de migración y colonización. Así mismo, facilitan la determinación de los patrones de mezcla dentro de las poblaciones (Saloum et al. 2013b).

La evaluación de las estimaciones de mezcla es una medida importante en los estudios de genética de poblaciones, particularmente en el contexto de las poblaciones altamente mezclados como son la mayoría de las que se encuentran en los países de América. Algunos estudios se han centrado en la investigación de proporción de ascendencia en las poblaciones a partir de datos cromosómicos uniparentales. Sin embargo, los estudios que utilizan marcadores autosómicos, reuniendo la historia de ambos linajes, siguen siendo escasos (Lins et al. 2011; Santos et al. 2010; Ibarra et al. 2014).

Finalmente, dentro del campo forense, los INDELS se han convertido en una herramienta útil y novedosa para la investigación o estudios de identificación humana y pruebas de paternidad.

**2.3.1 Pruebas de paternidad.** Debido a la baja tasa mutacional que presentan los INDELS, estos marcadores se han convertido en una herramienta muy útil en la resolución de casos complejos de pruebas de paternidad. La evaluación de dichas pruebas, puede generar incertidumbre debido a las incompatibilidades de herencia mendeliana que pueden ser observadas a partir del análisis de marcadores STRs, pues estos exhiben una alta tasa de mutación. Sin embargo, estas dificultades pueden evitarse mediante el uso de un conjunto de INDELS, ya que estos son marcadores genéticos más estables que los STRs (Pereira et al. 2009a). La caracterización de un alto número de marcadores puede mostrar de forma eficiente las nuevas incompatibilidades de transmisión, lo que favorece la hipótesis de exclusión o en el caso opuesto, la de inclusión, al incrementar los valores del índice

de paternidad a niveles más confiables que apoyan la paternidad (Pereira y Gusmão 2012).

**2.3.2 Análisis de muestras degradadas e identificación humana.** Un reciente interés por los INDELs como herramienta en el campo de aplicación forense, se hace relevante por la posibilidad de analizar amplificadores mucho más cortos en comparación con los STRs estándar. Esta es sin duda la razón más importante para la introducción de los SNPs en la práctica forense, con el fin de sobrepasar el inconveniente principal que presentan los STRs, y es su desempeño débil en la amplificación de muestras altamente degradadas (Pereira y Gusmão 2012).

Los INDELs con pequeña variación de longitud entre los alelos (2-5 pb), que al igual que los SNPs pueden ser analizados en amplicones muy cortos (menos de 160 pb), presentan una ventaja sobre estos últimos, ya que la detección alélica se lleva a cabo simplemente mediante la separación de fragmentos con electroforesis capilar, contrario a los protocolos laboriosos de tipificación de SNPs. En resumen, los INDELs permiten la combinación de una estrategia de amplicón corto con un simple flujo de trabajo analítico (Pereira y Gusmão 2012).

Además de esto, los marcadores de inserción-delección son de gran interés en el campo de aplicación forense, ya que presentan todo el potencial característico para proporcionar nuevas herramientas de inteligencia penal. Es decir, al tener un conjunto de marcadores que ayuden a la identificación humana como determinado elemento de prueba, podrían ayudar a dirigir el curso de la Investigación Penal (Pereira et al. 2009a, b; Fondevila et al. 2012).

Entre los parámetros estadísticos de relevancia en Genética Forense se tienen la probabilidad de exclusión, el poder de discriminación, la probabilidad de coincidencia y el contenido de información polimórfica. Estos parámetros son

valores estadísticos que permiten evaluar el poder informativo a priori de un marcador genético o un conjunto de marcadores en casos de filiación o investigación criminalística (Ibarra et al. 2014).

**2.3.2.1 Probabilidad de exclusión (PE).** Se define como la probabilidad de que un sistema genético muestre evidencias que conduzcan a la exclusión de un individuo involucrado en algún caso forense. Mediante este parámetro se puede establecer la proporción de individuos falsamente incluidos en un peritaje (Pereira y Gusmao 2012).

**2.3.2.2 Poder de discriminación (PD).** El poder de discriminación se determina mediante la probabilidad a priori de dos individuos no relacionados y seleccionados al azar entre la población, puedan diferenciarse genéticamente en uno o más loci, mediante el análisis de un marcador o un conjunto de marcadores (Pereira y Gusmao 2012).

**2.3.2.3 Contenido de información polimórfica (PIC).** Índice que evalúa la informatividad de un marcador en la población de acuerdo con las frecuencias de los alelos. Este valor fue descrito por Botstein et al. (1980). Según su clasificación, marcadores que precisen valores de PIC superiores a 0,5 se consideran altamente informativos, valores entre 0,25 y 0,50 medianamente informativos y valores inferiores a 0,25, muy poco informativos (Botstein et al. 1980).

**2.3.2.4 Probabilidad de coincidencia (PM).** Se define como la probabilidad que dos individuos elegidos al azar tengan genotipos idénticos en uno o más sistemas

genéticos, es decir, es la probabilidad, que dos individuos no relacionados compartan un mismo perfil de ADN (Pereira y Gusmao 2012).

A partir de lo mencionado anteriormente, se espera comprender de una mejor forma los resultados obtenidos en el análisis de los marcadores de inserción-delección considerados en el presente estudio.

### 3. HIPOTESIS DE INVESTIGACIÒN

**H<sub>0</sub>:** Los municipios que conforman en su totalidad el departamento de Santander se comportan como una misma unidad genética, analizados a través de la tipificación de 38 marcadores de Inserción - Delección (INDELs).

**H<sub>a</sub>:** Los municipios que conforman en su totalidad el departamento de Santander no se comportan como una misma unidad genética, analizados a través de la tipificación de 38 marcadores de Inserción - Delección (INDELs).

## 4. MATERIALES Y MÉTODOS

### 4.1 MUESTREO

**4.1.1 Población de estudio.** Debido a su posición geográfica Colombia ha recibido la influencia directa de tendencias culturales procedentes de Mesoamérica, el Caribe, Inca y civilizaciones amazónicas. Esta es la razón por la que a finales del siglo XV los conquistadores que vinieron de Europa, encontraron una amplia variedad de grupos nativos, que eventualmente se convirtió en el principal pool genético de las actuales poblaciones mestizas colombianas (Ossa et al. 2015).

Santander es el sexto departamento en número de habitantes y se ubica en el sector nororiental de los Andes colombianos (Castillo et al. 2015). Limita al norte con los departamentos de Cesar y Norte de Santander, al oriente con Norte de Santander y Boyacá, al sur con Boyacá y al occidente con Antioquia y Bolívar. Este departamento representa una gran afluencia histórica del país. Antes de la aparición de influencias europeas en Santander, diferentes poblaciones nativas como Guanes, Laques, Chitareros, Carares, Opores y Yariguíes se establecían en la región y durante el periodo de la colonización europea, los españoles se fundieron en una nueva raza con los escasos sobrevivientes de dichos pueblos indígenas (Academia de historia de Santander, Gobernación de Santander 2012).

Finalmente, con el advenimiento de la república, la industria y el comercio creció y durante el siglo XIX individuos no solo de España, también de Alemania, Inglaterra, Irlanda, Dinamarca y Asia se establecieron en el departamento y sus influencias se prolongaron hasta finales del siglo XX, contribuyendo a las tradiciones, etnias,

cultura y el actual pool genético de la población (Academia de historia de Santander, Gobernación de Santander 2012; Martínez et al. 2015).

El departamento de Santander se encuentra dividido desde 1910 en seis provincias regionales denominadas Soto, Guané, García Rovira, Vélez, Mares y Guanentá, las cuales se componen de 87 municipios y 2 corregimientos (Academia de historia de Santander, Gobernación de Santander 2012).

Los cambios demográficos sucedidos durante el siglo XIX, detuvieron el crecimiento de las provincias de García Rovira, Guanentá y Comunera, mientras que Soto creció convirtiéndose en la cabecera municipal del departamento. Debido a la expansión cafetera, se produjo un boom poblacional para Bucaramanga, Piedecuesta, Floridablanca, Girón, Rionegro y Lebrija, municipios que iniciaron su irrefrenable crecimiento. Su aglomeración demográfica atribuye a este conjunto una cifra de 1.072.587 habitantes, lo que representa un poco más de la mitad del total de la población del departamento de Santander (Oviedo 1990; DANE 2005, Hincapié et al. 2009). En la actualidad, dichos aspectos sociales, culturales y geopolíticos se reflejan en las seis provincias regionales en que se encuentra dividido el departamento.

**4.1.2 Población objeto de estudio.** La población objeto de estudio fue conformada por 500 individuos nativos del departamento de Santander, seleccionados del banco de muestras del Laboratorio de Genética Humana de la Universidad Industrial de Santander. Los donantes de las muestras son individuos no relacionados entre sí, procedentes de las seis provincias regionales del departamento (Soto, Mares, García Rovira, Guanentá, Comunera y Vélez), quienes previamente bajo consentimiento informado, autorizaron el uso de su material genético en estudios de investigación genética poblacional.

**4.1.3 Colección de muestras.** Durante este estudio, un total de 500 muestras fueron seleccionadas de forma aleatoria (epi info v. 6.04-2001) a partir de las muestras de individuos no relacionados entre sí, provenientes de las seis provincias regionales del departamento de Santander.

Las muestras seleccionadas fueron obtenidas de la genoteca del Laboratorio de Genética Humana de la Universidad Industrial de Santander (GH-UIS). Estas muestras estaban debidamente ordenadas e identificadas mediante códigos, lo cual permitió una selección de tipo aleatoria. El uso de los códigos permitió conservar y respetar el anonimato de los individuos.

Para cada uno de los 87 municipios del departamento se realizaron cálculos de tamaño de muestra con base en el número de habitantes por municipio según el último reporte censal de 2005 (DANE 2005) del Departamento Administrativo Nacional de Estadística (DANE).

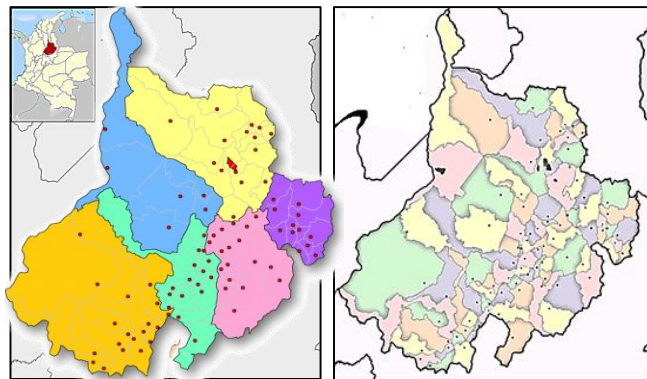
**4.1.4 Criterio de selección de muestras.** Las muestras seleccionadas corresponden a hombres y mujeres sanos, mayores de edad, no emparentados entre sí, nacidos en el departamento de Santander. Información sobre el color de la piel o etnicidad de los individuos no fue considerada dentro de los propósitos de selección de muestras.

**4.1.5 Información de individuos muestreados.** Una base de datos fue construida en Excel ® a partir de la información colectada para cada uno de los individuos participantes en la investigación. La identificación de los individuos se realizó

mediante el uso de un código evitando así la violación de confidencialidad de la información.

Se consideró para el análisis, variables demográficas tales como procedencia, lugar de origen, sexo definido como hombre o mujer y las variables genéticas como el genotipo de cada individuo obtenido a partir de la caracterización de los 38 marcadores autosómicos INDELS, todo esto fue incluido en la base de datos construida para tal fin.

**Figura 1.** División geopolítica del departamento de Santander.



Definición territorial de las seis provincias regionales de estudio: Amarillo oscuro, **Vélez**; Verde, **Comunera**; Rosa, **Guanentá**; Púrpura, **García Rovira**; Amarillo claro, **Soto**; Azul, **Mares**, y los 87 municipios muestreados

## 4.2 FASE DE LABORATORIO

Todas las muestras de ADN seleccionadas, fueron obtenidas del banco de muestras del Laboratorio de Genética Humana de la Universidad Industrial de Santander.

**4.2.1 Toma de muestra y extracción de ADN.** Las muestras analizadas correspondían a ADN, el cual fue extraído de un volumen final de 5 mL de sangre colectado mediante procedimiento de venopunción, previa firma del consentimiento informado.

Este procedimiento se realizó siguiendo el protocolo estándar de SaltingOut (Miller et al. 1988), y/o a partir de una mancha de sangre capilar recolectada en papel filtro estéril, extraído por el método de Chelex. Estas muestras permanecen conservadas a -70 °C en el banco de muestras del Laboratorio de Genética Humana de la Universidad Industrial de Santander (GH-UIS).

La cuantificación de ADN fue determinada por espectrofotometría mediante el uso del equipo NanoDrop 2000 Spectrophotometer (Thermo Scientific)

**4.2.2 Genotipificación.** La genotipificación de los polimorfismos de Inserción-Delección se realizó en el Laboratorio de Genética-UIS, mediante el uso de un panel de 38 marcadores genéticos ubicados a lo largo de todo el genoma. Este panel fue diseñado por Pereira et al. (2009b).

**4.2.3 Selección de marcadores genéticos.** El grupo inicial de marcadores candidatos para este estudio se basó en el trabajo previo de Weber et al. (2002). Partiendo de una lista de 4000 INDELS bialélicos previamente confirmados y caracterizados en los principales grupos poblacionales, Pereira et al. (2009a) lograron seleccionar 38 nuevos marcadores de inserción-delección de acuerdo a los siguientes criterios: (I) no codificantes, autosómicos y bialélicos (II) frecuencia alélica mínima de 0.25 para los grupos poblacionales de Europa, África, y Asia; (III) heterocigosidad promedio de 0.40; y (IV) variación alélica en longitud de 2 a 5 pb, a

fin de lograr una distribución uniforme por todo el genoma, y una distancia suficiente entre los marcadores en un mismo cromosoma para evitar el desequilibrio de ligamiento (ANEXO A, tabla 1).

Todos los marcadores se organizaron por longitud de amplificación esperada y se asignaron a cuatro fluorocromos diferentes a fin de lograr un ensayo de genotipado desde una única PCR y separación electroforética (Pereira et al. 2009b).

**4.2.4 Información de los Loci.** Los 38 INDELS distribuidos a lo largo de los 22 autosomas fueron amplificados eficientemente en una sola reacción de genotipificación (Pereira et al. 2009b). Información sobre cada uno de los 38 marcadores seleccionados para esta investigación se presenta en el ANEXO A, tabla 1.

**4.2.5 Amplificación y detección de marcadores.** La amplificación de los 38 marcadores se realizó mediante la técnica de PCR multiplex, usando el kit de PCR multiplex Qiagen a una concentración 2x, mezcla de primers 10x y 0.25 ng de ADN genómico en 2.5 uL de volumen final de reacción (ANEXO B, tabla 2). Las condiciones de termociclado consistieron inicialmente en un período de incubación a 95°C por 15 min; 10 ciclos a 94°C por 30 s, 60°C por 90 s and 72°C por 60 s; 20 ciclos a 94°C por 30 s, 58°C por 90 s and 72°C por 60 s; con una extensión final de 72°C por 60 min (Pereira et al. 2009b), dichas condiciones se encuentran en la (ANEXO C, tabla 3).

Finalmente, los productos de PCR fueron detectados mediante electroforesis capilar (ANEXO D, figura 2), añadiendo 0.8 uL de cada amplificado a 12.3 uL de mezcla de formamida desionizada Hi-Di (Applied Biosystems) y marcador de peso molecular

LIZ 500 (Applied Biosystems) utilizando el polímero POP-4 (Applied Biosystems). La detección alélica se realizó en un secuenciador ABI PRISM 310 (Applied Biosystems) siguiendo los siguientes parámetros de corrida:

Filtro GS STR POP-4 (1 ml); tiempo de inyección de muestra de 5 minutos; temperatura de corrida de 60°C a 15.0 Kv durante 30 minutos.

Las muestras fueron tipificadas mediante el programa GeneMapper v3.2 (Applied Biosystems). Las líneas celulares humanas 9947A y 9948 (Promega) ® fueron usadas como control positivo durante el desarrollo del protocolo de genotipado multiplex (ANEXO E, tabla 4).

**4.2.6 Análisis estadístico.** La estimación de las frecuencias alélicas, la evaluación de heterocigosidad observada y esperada, y la condición de equilibrio de Hardy-Weinberg en la población, así como el Análisis de Varianza Molecular, distancias genéticas y desequilibrio de ligamiento fue realizada mediante el software Arlequín v3.5 (Excoffier y Lischer 2010). Los parámetros usados para su evaluación locus por locus correspondieron a un número de 10.000 permutaciones, 1.000.000 de número de pasos en la cadena de Markov y un número de pasos de dememorización de 1.000.000 con un nivel de significancia de 0.01.

Así mismo, se evaluó la estructura genética de la población y se realizó una estimación de análisis ancestral individual y global mediante el software STRUCTURE v2.3.3 (Pritchard et al. 2000) usando muestras de referencia de poblaciones Hispanas, Europeas, Africanas y Asiáticas reportadas por Pereira et al. (2009b), asumiendo un modelo de mezcla con un período de quemado de 10.000 y 10.000 MCMC repeticiones sobre el período de quemado.

La estimación de los parámetros de importancia forense como probabilidad de exclusión (PE), poder de discriminación (PD), contenido de información polimórfica (PIC) y probabilidad de coincidencia (PM), fueron realizados con el software Power Stats V1.2. (Promega Corporation) ®. Todos estos parámetros fueron calculados para determinar la eficiencia de estos marcadores en la resolución de casos de identificación humana y forense.

### **4.3 ASPECTOS ÉTICOS**

Todas las muestras en este estudio fueron colectadas bajo consentimiento informado de participantes no emparentados, para su uso de forma anónima y responsable según la ley estatutaria N° 1581 de 2012, decreto 1377 de 2013. Así mismo, la propuesta de investigación y consentimiento informado fueron evaluados y aprobados por el comité ético de la Universidad Industrial de Santander bajo la protección de datos personales, resolución N° 1227 de 2013. Este proyecto de investigación no tiene ninguna orientación comercial. Durante la investigación se preservó los principios de beneficencia, no maleficencia, autonomía y justicia. De la misma forma, este estudio no generó impactos negativos sobre la población. El manejo de residuos se realizó de acuerdo a lo estipulado en el Decreto 351 de 2014, protocolo PGIR UIS. Este estudio fue clasificado de riesgo mínimo de acuerdo con la resolución 008430 de 1993 del Ministerio de salud de la República de Colombia.

Acta de aval del proyecto de investigación otorgada por el Comité de Ética de la Universidad Industrial de Santander y consentimiento informado de los individuos participes de este estudio (ANEXOS F y G, respectivamente).

## 5. RESULTADOS

### 5.1 BASE DE DATOS

Este estudio logró construir la primera base de datos de INDELs para Santander, donde se reporta la tipificación de 38 marcadores autosómicos de inserción-delección, obteniendo el perfil genético de cada uno de los individuos participantes de esta investigación para dichos marcadores.

Esta base de datos fue construida teniendo en cuenta la distribución política del departamento de Santander representado por las seis provincias regionales que lo conforman (Figura 1).

### 5.2 VARIACIÓN GENÉTICA DE LA POBLACIÓN

La variación genética de los 38 INDELs se estudió mediante la evaluación de las frecuencias alélicas y genotípicas de la población Santandereana y sus seis provincias regionales.

**5.2.1 Frecuencias alélicas.** A partir de los resultados obtenidos de la tipificación de 38 INDELs autosómicos en la muestra poblacional de estudio, fueron calculadas las frecuencias alélicas tanto para el departamento de Santander (ANEXO H, tabla 5) como para sus seis regiones provinciales teniendo en cuenta la evaluación de cada uno de los loci (ANEXO I, tabla 6).

El análisis de las frecuencias reportadas para cada loci en Santander, mostró que de los 38 polimorfismos bialélicos estudiados, sólo 4 de estos marcadores (B6, G5, Y4, y R9) exhibieron un valor de frecuencia alélica menor a 0.25 para su alelo corto

(delección), los marcadores R9 y G5 presentaron las frecuencias más bajas (0.19 y 0.20 respectivamente), de todo el panel (ANEXO H, tabla 5).

De igual forma, teniendo en cuenta el análisis realizado para cada región del departamento, se encontró que dichos marcadores (R9 y G5), mostraron una frecuencia menor a 0.25 para las seis provincias regionales, seguido del marcador Y4 el cual presentó una frecuencia baja de su alelo corto en 4 de las 6 regiones (Comunera, Mares, Soto y Vélez, con frecuencias de 0.16; 0.15; 0.23; 0.22 respectivamente), mientras que el marcador B6 sólo mostró este comportamiento en 3 de las 6 regiones (Comunera, Mares y Soto, con frecuencias de 0.20; 0.23; 0.23 respectivamente) (ANEXO I, tabla 6)

Por otra parte, 30 de los 38 marcadores que conforman el panel de estudio exhibieron una frecuencia alélica por encima de 0.45 para el Departamento de Santander. Así mismo, estos valores fueron consistentes con las frecuencias alélicas encontradas para cada provincia.

**5.2.2 Frecuencias genotípicas.** La muestra poblacional de estudio estuvo constituida por 500 individuos procedentes de los 87 municipios que conforman Santander. A partir de la tipificación de 38 sistemas autosómicos INDELS analizados en el total de la muestra, se logró calcular las frecuencias genotípicas poblacionales para Santander (ANEXO J, tabla 7).

Los genotipos heterocigotos estuvieron menos representados en la población total de Santander en comparación con los genotipos homocigotos (ANEXO J, tabla 7). Por otra parte, el genotipo 2 para el sistema R9 exhibió la frecuencia más alta de la población (0.666), seguido de los genotipos 2 para los sistemas G5, B6 y G9 con frecuencias de 0.636; 0.576; 0.576 respectivamente y el genotipo 1 para el sistema

R6 con frecuencia de 0.568. De igual forma, el genotipo 1 del sistema G5, fue el menos representado en la población con una frecuencia de 0.048.

### 5.3 ANÁLISIS DE LA DIVERSIDAD

La diversidad genética promedio a partir del cálculo de las frecuencias alélicas de los 38 marcadores INDELS fue de  $0.43 \pm 0.06$ . Con base en este valor, la provincia regional de García Rovira presentó la diversidad genética más baja ( $0.41 \pm 0.09$ ), mientras que la región provincial de Vélez mostró la mayor diversidad ( $0.44 \pm 0.09$ ) (Tabla 1).

Por otra parte, los sistemas más diversos para la muestra poblacional de Santander fueron los marcadores B2 y G8 con una heterocigosidad observada de 0.53 en ambos casos. Sin embargo, al evaluar la diversidad genética de cada marcador para las diferentes regiones de Santander, se apreció una diversidad genética menor a 0.25 para algunos de los sistemas evaluados.

**Tabla 1.** Diversidad y distancia genética de las seis provincias regionales de Santander ( $F_{ST}$ ).

Población	H	s.d.	Comunera	García Rovira	Guanentá	Mares	Soto	Vélez
Comunera	0,4386	0.0901	-	475,9089	Inf.	121,4905	Inf.	200,0758
García Rovira	0,4091	0.0956	0.0010	-	195,2716	57,7963	393,8201	216,1239
Guanentá	0,4148	0.1104	-0.0012	0.0025	-	3252,1883	Inf.	Inf.
Mares	0,4308	0.0787	0.0041	0.0086	0.0001	-	1663,7978	131,3721
Soto	0,4240	0.0688	0.0007	0.0025	-0.0020	0.0008	-	501,5999
Vélez	0,4416	0.0980	0.0025	0.0023	-0.0009	0.0038	0.0023	-
Santander	0,4259	0.0644	$F_{ST}$					

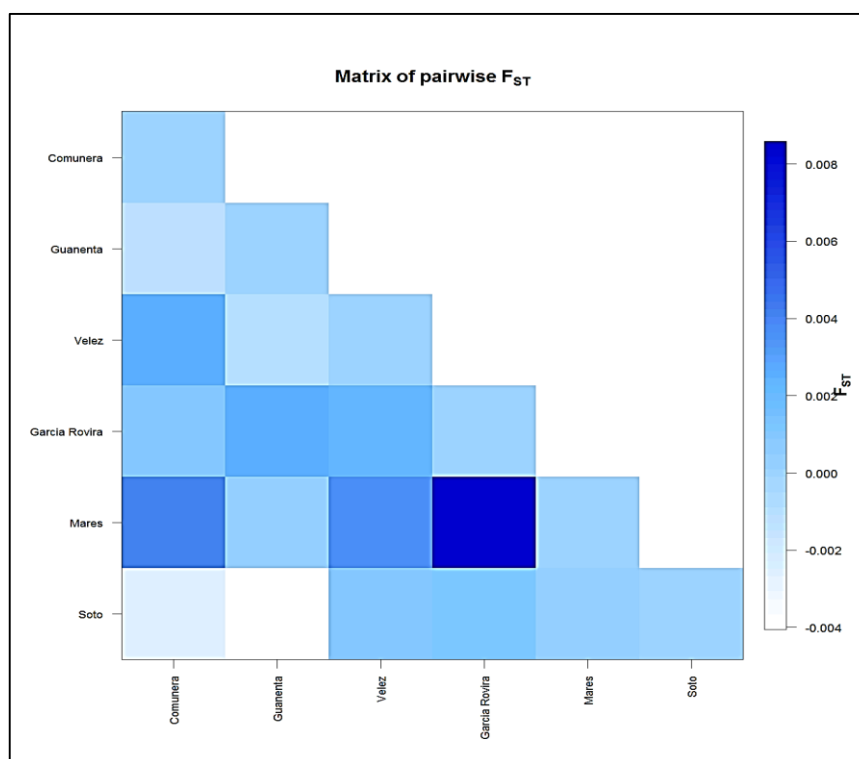
**H:** Diversidad genética promedio presente en cada población; **s.d.:** Desviación estándar. **Inf.:** Valores infinitos. **Valores de  $F_{ST}$**  bajo la diagonal. **Valores de tasa efectiva de migración**, sobre la diagonal.

#### 5.4 EVALUACIÓN DE LAS DISTANCIAS GENÉTICAS

La comparación entre poblaciones se realizó mediante el análisis de distancia genética, basado en el estimador  $F_{ST}$ , el cual reveló una tasa de variación genética baja entre las seis provincias regionales de Santander y una alta proporción de individuos migrantes efectivos por generación entre las poblaciones (Tabla 8).

Sin embargo, las provincias regionales de Mares y García Rovira mostraron una muy ligera diferenciación genética entre ellas ( $p < 0.01$ ), como puede ser observado en la matriz de  $F_{ST}$ , obtenida mediante el software Arlequín V.3.5. (Figura 2).

**Figura 2.** Matriz de diferenciación genética ( $F_{ST}$ ).



Representación gráfica de la diferenciación genética ( $F_{ST}$ ) entre las seis provincias regionales de Santander.

## 5.5 EVALUACIÓN DEL EQUILIBRIO DE HARDY WEINBERG

Al evaluar el equilibrio de Hardy Weinberg en la población total, se evidenció que seis de los 38 loci analizados se encontraban fuera de HW. Estos marcadores fueron, B7 ( $p= 0.0000 \pm 0.0000$ ), G4 ( $p= 0.0000 \pm 0.0000$ ), G6 ( $p= 0.0000 \pm 0.0000$ ), G7 ( $p= 0.0000 \pm 0.0042$ ), Y8 ( $p= 0.0000 \pm 0.0000$ ) y R9 ( $p= 0.0068$ ) (ANEXO H, tabla 5).

Una vez realizado este mismo análisis, para cada provincia regional del departamento, se logró observar que en su mayoría los marcadores anteriormente mencionados, presentaban de igual forma diferencias significativas, hallándose

fuera del equilibrio HW para las regiones de Guanentá (Y8,  $p=0.0012 \pm 0.0000$ ), Mares (B7,  $p=0.0035 \pm 0.0001$ ), Soto (B7,  $p= 0.0000 \pm 0.0000$ ; G4,  $p= 0.0001 \pm 0.0000$ ; G6,  $p= 0.0015 \pm 0.0000$ ; Y8,  $p= 0.0011 \pm 0.0000$ ) y Vélez (Y9,  $p= 0.0012 \pm 0.0000$ ). Por otra parte, todos los sistemas evaluados se encontraron bajo equilibrio de HW ( $p<0.01$ ) para las provincias regionales de García Rovira y Comunera. Ninguna corrección estadística fue aplicada a los resultados obtenidos durante este análisis.

## 5.6 ANÁLISIS DE LA ESTRUCTURA GENÉTICA Y ASIGNACIÓN DE INDIVIDUOS A POBLACIONES

La evaluación del análisis de varianza molecular realizado mostró que el componente de variación genética individual representó el 96.1% de la diversidad genética total de la población, lo cual indicó que solo el 0.08% de variación genética se encontró representada entre las poblaciones (Tabla 2).

**Tabla 2.** Análisis de Varianza molecular (AMOVA) para la población del departamento de Santander. Estimación de la estructura poblacional y evaluación de la variación genética en la población.

Fuente de Variación	Suma de cuadrados	Componentes de varianza	% Variación
Entre las poblaciones	47.408	0.0069	<b>0.0808</b>
Entre los individuos dentro de las poblaciones	2.493.493	0.3265	<b>3.8306</b>
Dentro de individuos	2.358.500	8.1892	<b>96.0885</b>
<b>Total</b>	4.899.401	8.5226	

Además, al realizar la evaluación de los estadísticos F, los resultados obtenidos fueron estadísticamente significativos ( $p < 0.01$ ) tanto para la población total  $F_{IT}$ , como para los individuos dentro de las poblaciones  $F_{IS}$ , lo que sugiere una posible tasa de endogamia baja presente en las poblaciones (ANEXO K, tabla 10). El valor de diferenciación poblacional encontrado  $F_{ST}$  fue de 0.0008 ( $p = 1.0000 \pm 0.0000$ ). Así mismo, estos resultados fueron corroborados mediante el Análisis de Varianza Molecular Locus por Locus, donde los sistemas B7 ( $p = 0.0000$ ), G4 ( $p = 0.0050$ ), G6 ( $p = 0.0018$ ), Y7 ( $p = 0.0015$ ) y R6 ( $p = 0.0072$ ) exhibieron diferencias significativas, para el índice  $F_{IS}$ .

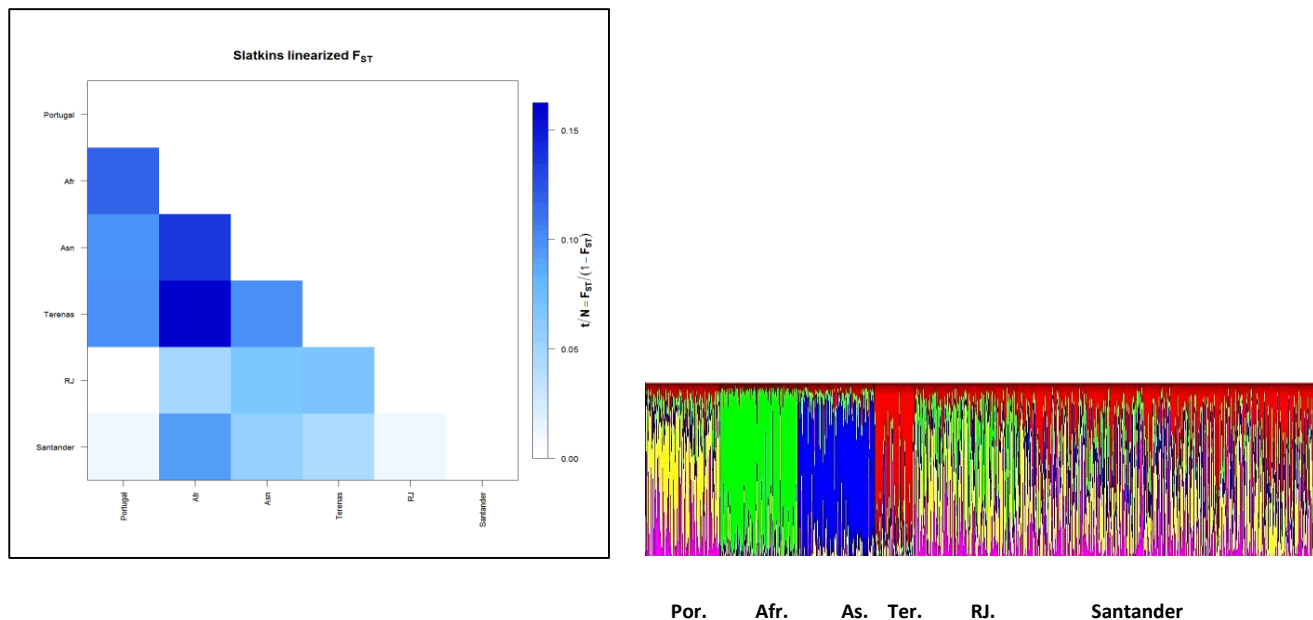
Por otra parte, los resultados obtenidos mediante el software Arlequín V3.5 fueron confirmados usando el software Structure v2.3.3, realizando un análisis de asignación de individuos a poblaciones teniendo en cuenta la información genotípica de cada sujeto. Bajo un modelo de mezcla, se calculó la distribución de las frecuencias alélicas a posteriori asumiendo que la población se encontraba en HW. Valores de probabilidad a priori ( $LnP$ ) y posteriori ( $P_K$ ) fueron evaluados (Pritchard et al. 2000) (ANEXO M, tabla 11). Un valor de ( $P_K$ ) = 1, correspondió al mejor ajuste del número de poblaciones que aportan al actual pool genético de Santander ( $K = 5$ )

Los datos obtenidos, también fueron usados para calcular las distancias genéticas ( $F_{ST}$ ) entre Santander y las demás poblaciones de referencia, África, Europa (Portugal), Asia, Nativos Americanos (Terenas) y Rio de Janeiro.

Este análisis reveló una baja diferenciación entre Santander y las poblaciones ancestrales de Europa y Nativos Americanos ( $F_{ST} = 0.0116$  p-valor  $0.0000 \pm 0.0000$ ;  $F_{ST} = 0.0398$  p-valor  $0.0000 \pm 0.0000$ , respectivamente). Sin embargo, Asia y África presentaron una diferenciación genética mayor en relación a la población santandereana (El valor hallado para Asia de  $F_{ST}$  fue de 0.0527 p-valor  $0.0000 \pm 0.0000$  y para África su valor de  $F_{ST}$  fue de 0.0838 p-valor  $0.0000 \pm 0.0000$ ).

Las poblaciones de Rio de Janeiro y Santander mostraron la más baja diferenciación genética entre ellas ( $F_{ST} = 0.0113$  p-valor  $0.0000+-0.0000$ ), lo cual podría ser atribuido a una menor presencia de barreras geográficas entre éstas (Figura 3).

**Figura 3.** Representaciones gráficas de la matriz de Slatkin ( $F_{ST}$ ), y estimaciones de ancestría entre Santander y poblaciones de referencia (Europa, África, Asia y Rio de Janeiro), obtenidas mediante los SW Arlequín V3.5 y Structure V2.3.3.



Muestras de Referencia, **Por.** Portugal (Europa); **Áfr.** África; **As.** Asia; **Ter.** Terenas (Nativos Americanos); **RJ.** Rio de Janeiro; Población de estudio, Santander.

### 5.7 EVALUACIÓN DEL DESEQUILIBRIO DE LIGAMIENTO

Una vez realizado el test exacto de desequilibrio de ligamiento entre pares de marcadores INDELS en el mismo cromosoma para la muestra poblacional total de Santander, se encontraron valores significativos ( $p < 0.01$ ) para algunas parejas de marcadores evaluados.

Con este análisis se logró observar para cada una de las provincias regionales que ciertas parejas de marcadores, al igual que el análisis realizado en la muestra poblacional total, se encontraban en desequilibrio de ligamiento ( $p > 0.01$ ), resaltando que dos de estos marcadores fueron hallados tanto en las regiones de Mares y García Rovira (B4 y G7), como en las de Soto y García Rovira (G6 y Y9).

## **5.8 ANÁLISIS DE EFICIENCIA FORENSE**

En cuanto a la eficiencia forense de este panel, los 38 INDELS fueron evaluados mediante el uso de parámetros de importancia forense (ANEXO H, tabla 5).

El estudio individual de cada marcador permitió evaluar su eficiencia, donde se encontró que el sistema B7 mostró el menor valor de probabilidad de coincidencia (0.334). Por el contrario, el sistema G5 presentó la mayor probabilidad de coincidencia (0.507), lo cual indica que la probabilidad de encontrar dos personas con el mismo genotipo para este sistema es del 33% y 50% respectivamente.

Por su parte, el mayor poder de discriminación lo mostró el sistema B7 (0.666), mientras que el sistema R9 presentó el menor (0.477). De igual forma, el PIC reveló un rango de información polimórfica de los 38 marcadores estudiados siendo el más bajo de 0.27 (sistema R9) a 0.37 (en su mayoría los demás sistemas que lo conforman).

Así mismo, los índices de paternidad fueron calculados para cada marcador mostrando valores pequeños de poder de exclusión e índice típico de paternidad. Los parámetros estadísticos de Poder de discriminación y exclusión acumulados para todo el panel fueron calculados, obteniendo resultados de 0.9999 y 0.9962, respectivamente (ANEXO H, tabla 5).

## 6. DISCUSIÓN

Al no existir reportes anteriores sobre el estudio de la estructura poblacional a partir del análisis de marcadores autosómicos INDELs para la población santandereana, los resultados hallados en la presente investigación fueron comparados con estudios previos que han incluido dentro de su población de estudio algunas muestras de individuos del departamento de Santander. De igual forma, la eficiencia del panel se evaluó y fue comparada con diversas investigaciones forenses reportadas para poblaciones de mezcla. En este sentido se encontraron publicados los trabajos de Hincapié et al. (2009); Pereira et al. (2009); Castillo et al. (2013); Martínez et al (2013); Ibarra et al. (2014); Martínez et al. (2015) que presentan resultados para el departamento de Santander y se podría inferir que los resultados obtenidos en este estudio reflejan de una manera representativa a la población de santandereana.

Esta investigación, pionera en el análisis del grado de estructura poblacional y la evaluación de parámetros forenses mediante el uso de 38 marcadores INDELs para el Departamento de Santander, permitió establecer el grado de diversidad genética presente en la población a partir de muestras de individuos nacidos en esta región. Al evaluar la diversidad genética, se encontró que la heterocigosidad media del panel de marcadores INDELs fue de  $0.43 \pm 0.06$ , confirmando las elevadas características polimórficas e informativas de este panel para la población santandereana. Por otra parte, Martínez et al. (2013), evaluando dichos marcadores para la población de San Basilio de Palenque, Colombia encontró heterocigosidades con valores desde 0.10 a 0.57 en la población y corroboró las características polimórficas del panel de estudio.

De igual forma, al realizar el análisis del equilibrio de HW, desviaciones significativas se encontraron en seis de los 38 loci en la población estudiada, observándose un ligero déficit de heterocigotos en la población. Estas variaciones podrían estar relacionados con las desviaciones del equilibrio Hardy-Weinberg debidas a efectos del tamaño de muestreo.

El desequilibrio de Hardy-Weinberg asociado a un exceso de homocigotos podría representar la presencia de alelos silenciosos para ciertos marcadores (marcadores en desequilibrio) en la población santandereana, razón que no puede ser excluida sin la realización de futuras investigaciones.

Así mismo, la provincia de Soto, fue la región con mayor número de marcadores fuera de HW, lo cual podría estar relacionado con un elevado flujo genético entre poblaciones como consecuencia de la relación sociodemográfica y económica que se genera en esta área del departamento. Esta región cuenta entre sus municipios con las ciudades de Bucaramanga, capital del departamento, Floridablanca, Girón y Piedecuesta, que se constituyen como ciudades de amplio desarrollo económico del país y por ende de gran afluencia poblacional.

Las comparaciones entre poblaciones se realizaron mediante el análisis de distancia genética ( $F_{ST}$ ). Este análisis no reveló diferenciación genética entre las regiones provinciales de Santander, a excepción de las regiones de Mares y García Rovira, que presentaron una ligera diferenciación genética entre ellas. Dicha diferenciación puede estar relacionada con la presencia de barreras geográficas y bajos niveles de la tasa de migración efectiva por generación ( $Nm$ .) hallada entre estas dos poblaciones, en contraste con las altas tasas de  $Nm$ . observadas en las demás regiones. Otro aspecto que podría reflejar la ligera diferenciación genética entre estas poblaciones puede estar relacionado a un bajo tamaño efectivo de las poblaciones que, sumado al aislamiento geográfico, estaría promoviendo deriva

genética y por ende una pérdida de diversidad y mayor diferenciación genética entre ellas, respecto a las demás regiones.

Por otra parte, el análisis de varianza molecular reveló que el mayor componente de variación genética se encuentra dentro de individuos y representa el 96.1% de la diversidad genética total en la población. Por lo tanto, lo anterior constituye una evidencia que podría soportar el comportamiento de la población de Santander como una sola unidad genética reproductiva, sin subestructura, basado en el análisis de los 38 INDELS autosómicos estudiados. Estos resultados se corroboran con estudios previamente realizados para la población de Santander mediante el análisis de STRs, Hincapié et al. (2009) y Castillo et al. (2015) en sus investigaciones, logran concluir que Santander no presenta estructura genética en la población y por ende se mantiene como una sola unidad panmíctica reproductiva. Además, el test de diferenciación poblacional (Raymond & Rousset, 1995), confirmó los resultados obtenidos con la evaluación de las distancias genéticas previamente analizadas, logrando establecer que no existen diferencias significativas entre las regiones ( $p < 0.01$ ). Las seis provincias regionales estudiadas componen la división política del departamento, y en términos de estructura genética en Santander, el comportamiento del estadístico  $F_{ST}$  logra evidenciar una tendencia a la homogenización de las regiones, permitiendo agruparlas en una única unidad que representa al Departamento de Santander. Cabe resaltar que la alta variación genética encontrada dentro de individuos confirma la fiabilidad del panel de estudio en la resolución de casos forenses como identificación humana.

Durante el análisis de varianza molecular Locus por Locus, los sistemas B7 ( $p = 0.0000$ ), G4 ( $p = 0.0050$ ), G6 ( $p = 0.0018$ ), Y7 ( $p = 0.0015$ ) y R6 ( $p = 0.0072$ ) exhibieron diferencias significativas, para el índice  $F_{IS}$ , indicando que dichos marcadores aportan a la muy ligera tasa de endogamia observada en la población. Adicionalmente, un análisis específico de índices  $F_{IS}$  fue aplicado a las poblaciones, para corroborar los resultados anteriormente obtenidos. Este análisis mostró que

las provincias de Guantán y García Rovira presentaban un comportamiento endogamo, sin embargo, el p-valor del índice  $F_{IS}$  encontrado para ellas, no fue estadísticamente significativo ( $p > 0.01$ ), por lo cual no se logra afirmar que estas poblaciones efectivamente presenten dicho comportamiento que pudiera estar afectando la diversidad de la población.

El análisis de distancia genética evaluado mediante Arlequín®, fue consistente con los resultados obtenidos por el SW Structure®. Este último análisis se realizó a partir de la base de datos genotípica generada en este estudio y datos publicados previamente con muestras de referencia procedentes de África, Asia, Europa (Portugal), Nativos Americanos (Terenas) y Río de Janeiro (Pereira et al. 2009a). Aunque el panel de marcadores aquí evaluado no es específico para estudios de ascendencia, contribuyó en la estimación de la proporción de ancestría de los individuos para cada una de las poblaciones evaluadas.

Teniendo en cuenta la formación histórica de la población de Santander, se asumió una contribución de cinco poblaciones ( $K=5$ ) al pool genético actual de la población. Mediante este análisis, se logró inferir que la población santandereana presenta una mayor contribución a su pool genético a partir de las poblaciones Europeas seguido de los Nativos Americanos. Ibarra et al. (2014), a partir de 32 marcadores-X INDELS realizó un análisis de mezcla para la población Colombiana y encontró un alto componente genético dado por las poblaciones de referencia de Europa, Nativos Americanos y África presentes en la región andina del país. Sin embargo, dicha región exhibía una mayor contribución a su pool génico de las dos primeras poblaciones de referencia mencionadas. Con lo cual se corroboran los resultados obtenidos en el presente estudio. La diversidad genética encontrada está de acuerdo con reportes previos que describen un gradiente de poblaciones nativas principalmente de origen europeo y Nativos Americanos, seguido de un componente asiático y africano que conforman la población santandereana (Ibarra et al. 2014).

Como se muestra en la figura 6, las poblaciones de Santander y Río de Janeiro presentan una amplia diversidad genética y albergan contribuciones similares de Europa y América nativa, que pueden estar relacionados con la cercana posición geográfica a la que se encuentran y las similitudes culturales entre ellas. Lo anterior confirma la presencia de una alta variabilidad genética en el departamento de Santander como una población de mezcla.

De igual forma, los resultados obtenidos durante la evaluación del desequilibrio de ligamiento, confirman la asociación entre algunos pares de loci, lo cual podría atribuirse a la gran cantidad de mezcla que compone a la población. Estos resultados obtenidos difieren a lo reportado por Pereira et al. (2009b), donde todos los marcadores se encontraban en equilibrio de ligamiento. Sin embargo, cabe resaltar que las poblaciones de estudio en este caso fueron muestras de referencia de poblaciones nativas de Europa, África, Asia y Nativos Americanos, donde el componente de mezcla en dichas poblaciones es muy bajo en comparación con aquellas poblaciones mezcladas como Santander, la cual presenta una alta diversidad y mayor probabilidad de encontrar distintos loci en desequilibrio por recombinación. Cabe mencionar que aquellos loci en desequilibrio de ligamiento se relacionan con los marcadores observados fuera de HW, por ende, no se puede descartar que sea este el mecanismo por el cual se hayan encontrado estos marcadores en desequilibrio de ligamiento en la población.

Cabe resaltar que una de las parejas de marcadores evaluados (B4 y G7) se encontraba en desequilibrio tanto en la región de Mares como en la de García Rovira, es decir en aquellas regiones donde se observó una tasa de migración efectiva por generación muy baja, lo cual podría deberse a los pocos individuos migrantes efectivos entre las poblaciones, y esto reflejarse posiblemente en que esta pareja de marcadores durante el proceso de recombinación está prevaleciendo entre las poblaciones.

Finalmente, los índices de paternidad fueron evaluados mostrando valores bajos de poder de exclusión e índice típico de paternidad para ciertos marcadores. Sin embargo, al evaluar el panel en conjunto, los parámetros estadísticos de Poder de discriminación y exclusión acumulados fueron de 0.9999 y 0.9962, respectivamente, lo que permite inferir una alta discriminación entre individuos y la alta fiabilidad de excluir padres no biológicos, convirtiéndolos en una valiosa herramienta para resolución de casos complejos de paternidad o identificación humana.

Lo anterior se corrobora con las investigaciones reportadas por Pereira et al. (2009b) y Martínez et al. (2013), donde logran establecer poderes de discriminación y exclusión acumulados semejantes a los obtenidos en el presente estudio, concluyendo que este panel de marcadores autosómicos INDELS es lo suficientemente informativo en la resolución de casos forenses y de gran interés para el análisis de muestras altamente degradadas (Pereira et al. 2009b). Con base en los resultados obtenidos en el presente estudio, se puede afirmar que este panel se constituye como una herramienta novedosa dentro del campo forense y estudios de variabilidad genética en Santander.

## 7. CONCLUSIONES

Este primer reporte del análisis de 38 marcadores autosómicos INDELS en la población santandereana permitió estudiar la estructura genética de la población, la cual debe ser tomada en cuenta para la realización de bases de datos forenses. De igual forma, provee frecuencias alélicas y parámetros de eficiencia forense que son relevantes para la validación y la aplicabilidad de dichos marcadores en la resolución de casos forenses como identificación humana y pruebas de paternidad en Santander.

La alta diversidad genética y el análisis de varianza molecular sugieren que Santander sigue comportándose como una misma unidad genética reproductiva sin subestructura. Lo cual está relacionado con la baja diferenciación genética obtenida entre las seis provincias regionales de la metapoblación.

Se logró determinar una variabilidad genética alta dentro de individuos, por lo cual se podría inferir que los apareamientos entre los individuos de Santander se dan de forma aleatoria y confirma la fiabilidad del panel de 38 INDELS como buenos marcadores en la resolución de casos forenses como de identificación humana.

La eficiencia y la facilidad de trabajo con este panel de 38 INDELS desarrollado por Pereira et al. 2009b, lo convierte en una herramienta útil y novedosa para estudios futuros de identificación humana y variación genética en la población Santandereana.

Este estudio permitió un primer acercamiento a la comprensión de la distribución genética de los marcadores autosómicos INDELS en el departamento de Santander, siendo un trabajo piloto para estudios posteriores en Santander y Colombia.

## 8. RECOMENDACIONES

Los resultados de las diferencias significativas encontradas en el análisis de HW para algunas de las provincias regionales de Santander, podrían ser explorados mucho más a fondo en futuras investigaciones, que permitiría llevar a la explicación de posibles fenómenos de migración, divergencia o endogamia que pudieran estar afectando la población.

Futuras investigaciones podrían evaluar la ancestría del departamento de Santander mediante paneles de marcadores INDELS específicos para esto, confirmando o reportando nueva información sobre el grado de estructura genética presente en la población santandereana y su relación con la presencia de enfermedades genéticas complejas que pudieran afectarla.

## BIBLIOGRAFIA

Academia de Historia de Santander. 2012 “Descripción Histórica de Santander.”  
Gobernación de Santander. Recuperado de  
<http://www.santander.gov.co/index.php/gobernacion/departamento/historia-de-santander>

Berger J., Suzuki T., Senti K.A., Stubbs J., Schaffner G., and Dickson B.J. 2001. “Genetic Mapping with SNP Markers in *Drosophila*.” *Nature Genetics* 29 (4): 475–81. doi:10.1038/ng773.

Bhangale T.R., Rieder M.J., Livingston R.J., and Nickerson D.A. 2005. “Comprehensive Identification and Characterization of Diallelic Insertion-Deletion Polymorphisms in 330 Human Candidate Genes.” *Human Molecular Genetics* 14 (1): 59–69. doi:10.1093/hmg/ddi006.

Botstein D., White R.L., Skolnick M., Davis R.W. 1980. “Construction of a genetic linkage map in man using restriction fragment length polymorphisms”. *Am J Human Genet* 32:314-31.

Builes J.J., Bravo M.L, Gómez C. 2006. “Y-chromosome STRs in an Antioquian (Colombia) population sample”. *Forensic Science International* 164: 79–86.

Castillo A., Gil A., Pico A., Vargas C., Yurrebaso I., García O. 2013. “Genetic variation for 20 STR loci in a northeast Colombian population (Department of Santander)”. *Forensic Science International: Genetics Supplement Series* 4 (1): e298 - e299

Castillo A., Rueda K., Pico A., Gil A., Vargas C., Rondón F., Gusmão L. 2015. “Comparing different population groups in Santander—Colombia through Y-STR

haplotype analysis”. *Forensic Science International: Genetics Supplement Series 5*: e482–e483

Dawson E., Chen Y., Hunt S., Smink L.J., Hunt A., Rice K., Livingston S., et al. 2001. “A SNP Resource for Human Chromosome 22: Extracting Dense Clusters of SNPs from the Genomic Sequence.” *Genome Research* 11 (1): 170–78. doi:10.1101/gr.156901.

Departamento Administrativo Nacional de Estadística (DANE). 2005. Datos del censo de Colombia del 2005. *Memorias. Santa Fe de Bogotá, República de Colombia*. Disponible en: <http://www.dane.gov.co>.

Ellegren, H. 2004. “Microsatellites: Simple Sequences with Complex Evolution.” *Nature Reviews. Genetics* 5 (6): 435–45. doi:10.1038/nrg1348.

Excoffier, L. and Lischer, H.E. L. 2010 “Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows”. *Molecular Ecology Resources*. 10: 564-567.

Fondevila M., Phillips C., Santos C., Pereira R., Gusmão L., Carracedo A., Butler J.M, Lareu M.V., and Vallone P.M. 2012. “Forensic Performance of Two Insertion-Deletion Marker Assays.” *International Journal of Legal Medicine* 126 (5): 725–37. doi:10.1007/s00414-012-0721-7.

Freitas S., Natalle C., Resque R.L., Ribeiro-Rodrigues E.M., Guerreiro J.F., Santos N.P., Ribeiro-Dos-Santos A., and Santos S. 2010. “X-Linked Insertion/deletion Polymorphisms: Forensic Applications of a 33-Markers Panel.” *International Journal of Legal Medicine* 124 (6): 589–93. doi:10.1007/s00414-010-0441-9.

Ibarra A., Restrepo T., Rojas W., Castillo A., Amorim A., Martínez B., Burgos G., et al. 2014. “Evaluating the X Chromosome-Specific Diversity of Colombian

Populations Using Insertion/deletion Polymorphisms.” *PLoS ONE* 9 (1). doi: 10.1371/journal.pone.0087202.

Hincapié M.L., Gil A., Pico A., Gusmão L., Rondón F., Vargas C., y Castillo A. 2009. “Análisis de La Estructura Genética En Una Muestra Poblacional de Bucaramanga, Departamento de Santander.” *Colombia Médica* 40 (4): 361–72.

Lins T.C., Vieira R.G., Abreu B.S., Gentil P., Moreno-Lima R., Oliveira R.J., and Pereira R. 2011. “Genetic Heterogeneity of Self-Reported Ancestry Groups in an Admixed Brazilian Population.” *Journal of Epidemiology / Japan Epidemiological Association* 21 (4): 240–45. doi:10.2188/jea.JE20100164.

Martínez B., Builes J.J., Gaviria A., Burgos G., Manrique A., Aguirre D.P., Mendoza L., et al. 2013. “Population Genetic Data of 38 Autosomal InDels in San Basilio de Palenque, the First Free Town in America.” *Forensic Science International: Genetics Supplement Series* 4 (1): 611–12. doi:10.1016/j.fsigss.2013.10.037.

Martínez B., Builes J.J., Aguirre D., Mendoza L., Afanador C.H., Meza C., Marrugo J. 2015. “Ancestry background of a population sample from Bolivar Department, Colombia” *Forensic Science International: Genetics Supplement Series* 5 (2015) e170–e171

Miller S.A., Dykes D.D., and Polesky H.F. 1988. “A Simple Salting out Procedure for Extracting DNA from Human Nucleated Cells.” *Nucleic Acids Research* 16 (3): 1215. doi:10.1093/nar/16.3.1215.

Mills R.E., Luttig C.T., Larkins C.E., Beauchamp A., Tsui C., Stephen Pittard W., and Devine S.E. 2006. “An Initial Map of Insertion and Deletion (INDEL) Variation in the Human Genome.” *Genome Research* 16 (9): 1182–90. doi:10.1101/gr.4565806.

Mills R.E., Stephen Pittard W., Mullaney J.M., Farooq U., Creasy T.H., Mahurkar A.A., Kemeza D.M., et al. 2011. “Natural Genetic Variation Caused by Small

Insertions and Deletions in the Human Genome.” *Genome Research* 21 (6): 830–39. doi:10.1101/gr.115907.110.

Nei M. 1973. “Analysis of Gene Diversity in Subdivided Populations”. *Proc Nat AcadSci* 70(1):3321-23.

Ossa H., Aquino J., Sierra S., Ramírez A., Carvalho E.F., Gusmão L. 2015. “Analysis of admixture in Native American populations from Colombia” *Forensic Science International: Genetics Supplement Series* 5 (2015) e332–e334

Oviedo B.V., de. 1990. “Cualidades y riquezas del Nuevo Reino de Granada. Gobernacion de Santander. EN: Colección Memoria Regional, Bucaramanga. Volumen 4: 34-40

Peña A.V. 2007. “Caracterización de la estructura y diversidad genética de tres poblaciones del sur-occidente colombiano utilizando cuatro sistemas de marcadores moleculares microsatélites” [disertación]. *Santiago de Cali (Va). Universidad del Valle.*

Pereira R., Phillips C., Alves C., Amorim A., Carracedo A., and Gusmão L. 2009a. “Insertion/deletion Polymorphisms: A Multiplex Assay and Forensic Applications.” *Forensic Science International: Genetics Supplement Series* 2 (1): 513–15. doi:10.1016/j.fsigss.2009.09.005.

Pereira R, Phillips C., Alves C., Amorim A., Carracedo A., and Gusmao L. 2009b. “A New Multiplex for Human Identification Using Insertion/deletion Polymorphisms.” *Electrophoresis* 30 (21): 3682–90. doi:10.1002/elps.200900274.

Pereira R., y Gusmao L. 2012. “Capillary electrophoresis of 38 noncoding biallelic mini-Indels for degraded samples and as complementary tool in paternity testing”. *Methods in Molecular Biology* 830:141-57. doi: 10.1007/978-1-61779-461-2\_10.

Pimenta J.R., and Pena S.D.. 2010. "Efficient Human Paternity Testing with a Panel of 40 Short Insertion-Deletion Polymorphisms." *Genetics and Molecular Research: GMR* 9 (1): 601–7. doi:10.4238/vol9-1gmr838.

Pritchard J.K, Stephens M., and Donnelly P. 2000. "Inference of Population Structure Using Multilocus Genotype Data." *Genetics* 155 (2): 945–59. doi:10.1111/j.1471-8286.2007.01758.x.

Ribeiro-Rodrigues E.M., Pereira N., Dos Santos C., Ribeiro Dos Santos A.K.C., Pereira R., Amorim A., Gusmão L., Zago M.A., and Batista Dos Santos S.E. 2009. "Assessing Interethnic Admixture Using an X-Linked Insertion-Deletion Multiplex." *American Journal of Human Biology* 21 (5): 707–9. doi:10.1002/ajhb.20950.

Rojas K., Roa M., Briceño I., Guaneme G., Gómez A. 2011. "Polimorfismos de 17 marcadores STRs del cromosoma Y en una muestra poblacional del altiplano cundiboyacense". Colombia. *Médica*. 42: 88–97.

Rondón F. 2009. "Estudio de la variabilidad genética en poblaciones humanas del centro y suroccidente colombiano mediante el uso de marcadores moleculares" [disertación]. *Santiago de Cali (Va). Universidad del Valle*.

Rondón F., Barreto G. 2013. "Estructura genética, ancestralidad y su relación con los estudios en salud humana" *Medicas UIS*. Vol. 26, núm. 1.

Rosenberg N.A., Mahajan S., Ramachandran S., Zhao C., Pritchard J.K., and Feldman M.W. 2005. "Clines, Clusters, and the Effect of Study Design on the Inference of Human Population Structure." *PLoS ONE* 1 (6): e70. doi:10.1371/journal.pgen.0010070.

Saloum de Neves Manta F., Pereira R., Vianna R., Beuttenmüller de Araújo A.R., Góes Gitaí D.L., da Silva D.A., de Vargas Wolfgramm A., et al. 2013b. "Revisiting

the Genetic Ancestry of Brazilians Using Autosomal AIM-InDels.” *PLoS ONE* 8 (9). doi:10.1371/journal.pone.0075145.

Santos F.R., Tyler-Smith C. 1996. “Reading the human Y chromosome: the emerging DNA markers and human genetic history”. *Braz J Genet* 19:665-70.

Santos Ney P.C., Ribeiro-Rodrigues E.M., Ribeiro-dos-Santos A.K.C., Pereira R., Gusmão L., Amorim A., Guerreiro J.F., et al. 2010. “Assessing Individual Interethnic Admixture and Population Substructure Using a 48-Insertion-Deletion (INSEL) Ancestry-Informative Marker (AIM) Panel.” *Human Mutation* 31 (2): 184–90. doi:10.1002/humu.21159.

Shedlock A.M.; Okada N. 2000. “SINE Insertions: Powerful Tools for Molecular Systematics.” *BioEssays*. doi:10.1002/(SICI)1521-1878(200002)22:2<148::AID-BIES6>3.0.CO;2-Z.

Sherry S.T., Ward M., Sirotkin K. 1999. “dbSNP-database for Single Nucleotide Polymorphisms and other classes of minor genetic variation”. *Genom Res* 9:667-79.

Syvänen A.C. 2001. “Assessing Genetic Variation: Genotyping Single Nucleotide Polymorphisms.” *Nature Reviews. Genetics* 2 (12): 930–42. doi:10.1038/35103535.

Tishkoff S.A., Reed F.A., Friedlaender F.R., Ehret C., Ranciaro A., Froment A., Hirbo J.B, et al. 2009. “The Genetic Structure and History of Africans and African Americans.” *Science (New York, N.Y.)* 324 (5930): 1035–44. doi:10.1126/science.1172257.

Weber J.L, Donna D., Heil J., Fan Y., Zhao C., and Marth G. 2002. “Human Diallelic Insertion/deletion Polymorphisms.” *American Journal of Human Genetics* 71 (4): 854–62. doi:10.1086/342727.

Wicks S.R, Yeh R.T., Gish W.R., Waterston R.H., and Plasterk R.H. 2001. "Rapid Gene Mapping in *Caenorhabditis Elegans* Using a High Density Polymorphism Map." *Nature Genetics* 28 (2): 160–64. doi:10.1038/88878.

Wright S. 1965. "The interpretation of population structure by F-statistics with special regard to systems of mating". *Evolution* 19:395-420.

Yang N., Li H., Criswell L.A., Gregersen P.K., Alarcon-Riquelme M.E, Russell Shigeta R.K., et al. 2005. "Examination of Ancestry and Ethnic Affiliation Using Highly Informative Diallelic DNA Markers: Application to Diverse and Admixed Populations and Implications for Clinical Epidemiology and Forensic Medicine." *Human Genetics* 118 (3-4): 382–92. doi:10.1007/s00439-005-0012-1.

Yunis J.J., García O., Baena A., Arboleda G., Uriarte I., Yunis. E.J. 2000. "Population Frequency of Short Tandem Repeat Loci D18S849, D3S1744 and D12S1090 in Caucasian-Mestizo and African Descent Population of Colombia". *Journal of Forensic. Science* 45:437-9.

## ANEXOS

**Anexo A: Tabla 3.** Información específica del panel de 38 marcadores autosómicos INDELS desarrollado por Pereira et al. 2009b.

Numero rs*	Cromosoma	Posición (pb)	Alelos	Tamaño esperado amplificado (S/L)
rs3047269	1	161077452	-/CTGA	126–130
rs2307579	1	245878706	-/ATG	104–107
rs16624	2	234681130	-/GT	65–67
rs2308242	3	8591709	-/CT	106–108
rs2308026	4	119404855	-/CA	83–85
rs2307526	5	5178112	-/ACAC	95–99
rs1160956	5	65414216	-/AGA	128–131
rs1610871	5	171020572	-/TAGG	61–65
rs2307710	6	47929222	-/AGGA	92–96
rs2307839	6	117200251	-/GA	152–154
rs2308137	6	149655891	-/GA	61–63
rs2307978	7	83121850	-/GA	156–158
rs35769550	8	76681235	-/TGAC	89–93
rs5895447	8	138489776	-/CA	128–130
rs16402	9	38396788	-/TTAT	150–154
rs2067294	9	70504241	-/CTT	80–83
rs2307580	9	104626014	-/AATT	120–124
rs140809	10	6027167	-/CAA	115–118
rs1160886	10	54112392	-/ACT	75–78
rs10688868	11	258180	-/CT	81–83
rs34811743	11	30134266	-/TG	108–110
rs33972805	11	125794082	-/CT	135–137
rs1610919	12	14801263	-/AT	142–144
rs2067238	12	113772931	-/GCT	71–74

rs2308171	13	43778155	-/TCTG	135-139
rs2308189	14	28106508	/AACTA	119-124
rs2308020	15	51268809	-/TT	127-129
rs2067208	16	83139788	/GCCAG	93-98
rs3051300	17	10076666	-/GTAT	63-67
rs3080855	18	21507205	-/AATT	133-137
rs34511541	18	34677042	-/CTCTT	143-148
rs36040336	19	1353662	-/AT	65-67
rs2307689	19	48896180	-/TTC	74-77
rs33917182	20	11643625	-/CA	142-144
rs34541393	20	30165066	-/AACT	57-61
rs35605984	21	14556736	/TAAAG	151-156
rs10629077	21	30294208	-/AT	74-76
rs2307700	22	25120901	/TCAC	101-105

\*Marcadores seleccionados para el estudio genético poblacional y forense del departamento de Santander.

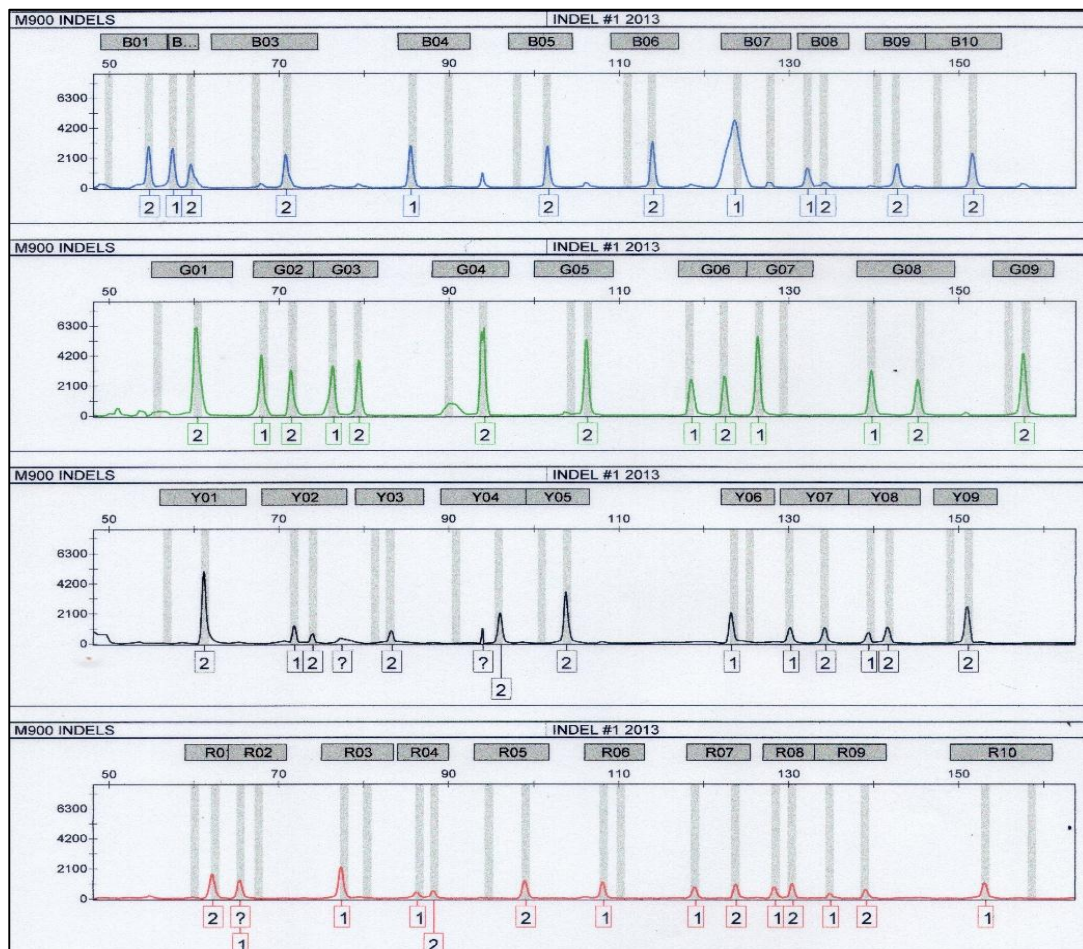
**Anexo B: Tabla 4.** Protocolo de amplificación mediante el kit de PCR multiplex de Quiagen.

<b>Plex-38 INDELS</b>	<b>Volumen por muestra</b>
Master Mix PCR Multiplex 2X Quiagen	1.25 uL
Mezcla de Primers 10X	0.25 uL
H <sub>2</sub> O	0.5 uL
ADN (0.25 ng/uL)	0.5 uL
Volumen Final de Reacción	2.5 uL

**Anexo C: Tabla 5.** Protocolo de amplificación, técnica de reacción de PCR multiplex (Pereira et al. 2009b).

Periodo de Incubación	10 Ciclos			20 Ciclos			Extensión Final	Período Final
	Denaturación	Anillamiento	Extensión	Denaturación	Anillamiento	Extensión		
95°C	94°C	60°C	72°C	94°C	58°C	72°C	72°C	4°C
15 min	30 s	90 s	60 s	30 s	90 s	60 s	60 min	Infinito

**Anexo D: Figura 4.** Electroferograma de amplificación para los 38 sistemas autosómicos de inserción-delección, usando 0.25 ng de ADN de un individuo nativo del departamento de Santander.



La figura muestra un ADN template (0.25 ng) que fue amplificado usando el kit de PCR multiplex Qiagen. Los productos amplificados fueron mezclados con un marcador interno

(Internal Lane Standard 500) y evaluado en un analizador Applied Biosystems 310 Genetic Analyzer. Las muestras fueron tipificadas mediante el programa GeneMapper v3.2 (Applied Biosystems).

**Anexo E: Tabla 6.** Perfil genético de las líneas celulares humanas 9947A y 9948 (Promega).

		<b>9947A</b>									
<b>DNA marker</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>	
B	SS	SL	SL	SL	SL	LL	SL	LL	SL	SL	
G	SL	SL	SL	LL	SL	SL	SS	SL	LL		
Y	LL	LL	SL	SL	SL	SL	SL	SL	SS		
R	SS	SL	SL	LL	LL	SS	SL	SL	SL	SL	

		<b>9948</b>									
<b>DNA marker</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>	
B	SL	SS	SS	SS	SL	SL	SL	SL	SL	LL	
G	LL	SL	LL	SL	LL	SS	SS	SL	LL		
Y	SL	SS	LL	LL	LL	SL	SL	SL	SL		
R	SL	SS	LL	SL	SL	SS	SL	SS	LL	SL	

*S* Short allele; *L* long allele

Control positivo durante el desarrollo del protocolo de genotipado multiplex (Pereira y Gusmão 2012).

**Anexo F:** Permiso otorgado por el Comité de Ética de la Universidad Industrial de Santander para el desarrollo de la presente investigación.



Código 7083 *A02.76*

Bucaramanga, *21 OCT 2013*

*2013 - 13108*

Doctora  
**CLARA INÉS VARGAS CASTELLANOS**  
Profesora Departamento de Ciencias Básicas  
Decana (e)  
Facultad de Salud  
UIS Presente

**Asunto: Aval Comité de Ética Proyecto: "Análisis poblacional de marcadores Indels en una muestra de individuos del Departamento de Santander"**

Cordial Saludo. El Comité de Ética en Investigación Científica de la Universidad Industrial de Santander (CEINCI-UIS) en reunión realizada el 18 de octubre de 2013, según consta en el acta No. 13 de 2013, presenta comunicación enviada por la DIFE donde informan que en el Comité Operativo de Investigación y Extensión (COIE), se designo como nueva directora del proyecto del asunto y teniendo en cuenta que éste, según consta en el acta No. 11 de septiembre 27 de 2013 se había avalado porque cumplía con todos los requerimientos del CEINCI, el Comité acuerda AVALAR este nuevo cambio.

**FRANCISCO ESPINEL CORREAL**  
Presidente (e)  
CEINCI- UIS

**DORA INES PARRA**  
Secretaria  
CEINCI-UIS

Copia: Dra.  Adriana Castillo, Secretaria General de la Universidad Industrial de Santander  
Dra. Lina María Vera. Directora de Investigación y Extensión Facultad de Salud  
Archivo Comité de Ética

Omaira M

UNIVERSIDAD INDUSTRIAL DE SANTANDER SECRETARIA GENERAL	
22 OCT. 2013	
Entregó: <i>Freddy</i>	Hora: <i>4:30 PM</i>
Recibió: <i>Elga</i>	Fecha: <i>1</i>

**Anexo G:** Consentimiento informado de los individuos partícipes del presente estudio. Este documento fue avalado por el Comité de Ética de la Universidad Industrial de Santander.



**Proyecto: Estudio genético poblacional a partir del análisis de marcadores de ADN en una muestra de individuos nacidos en el Departamento de Santander.**

## **CONSENTIMIENTO INFORMADO**

De acuerdo con los principios establecidos en la Resolución 008430 de 4 de Octubre de 1993 por la cual se establecen las normas para la investigación en salud, específicamente en el Artículo 15, en lo relacionado con el Consentimiento Informado, usted deberá conocer acerca de esta investigación y aceptar participar en ella si lo considera conveniente. Por favor lea con cuidado y haga las preguntas que desee hasta su total comprensión.

El grupo de investigación en Genética Humana de la Universidad Industrial de Santander se encuentra realizando el estudio de la población nacida en el departamento de Santander a partir del análisis de marcadores genéticos con alta variabilidad, para establecer la diversidad genética de las personas nacidas en nuestro departamento y conocer el componente ancestral que cada uno posee de acuerdo con el/los grupo/s étnicos poblacionales de los que proviene.

Si usted acepta participar le será tomada una gota de sangre con una lanceta nueva y estéril para ser depositada en un papel de filtro estéril. Este procedimiento no tiene ningún riesgo o complicación para usted. Con su muestra realizaremos la extracción del material genético llamado Acido desoxirribonucleico (ADN) y determinaremos su perfil genético para los marcadores poblacionales analizados. Los resultados de este trabajo son de gran importancia para nuestro departamento y país pues permitirán conocer los componentes genéticos que caracterizan nuestra población y que permiten diferenciarnos de otras y darnos autenticidad.

Si usted tiene alguna pregunta sobre la toma de muestra o las características del trabajo, por favor infórmenos para poderle brindar una explicación más detallada y si lo desea puede retirarse del estudio en cualquier momento sin que por ello se creen prejuicios para usted.

Su muestra será identificada con un código para asegurar que no se conocerá su información privada, tal como su nombre, documento de identidad etc. y que se mantendrá la confidencialidad total de sus datos personales en caso de realizar la publicación científica de los resultados.

Su participación en este estudio no tiene ningún costo, estos serán asumidos por la investigación, de igual manera si usted autoriza, sus muestras y datos obtenidos en este estudio, pueda ser utilizados en otros estudios, previa aprobación del Comité de Ética para la Investigación Científica de la Facultad de Salud de la UIS para realizar dichos estudios.

Autoriza \_\_\_\_ No autoriza \_\_\_\_\_

\_\_\_\_\_  
Firma de autorización



*Adriana Castillo Pico*

Adriana Castillo Pico

Nombre del Investigador principal

Firma

Para cualquier información relacionada con este proyecto usted puede comunicarse al Laboratorio de Genética UIS. Teléfonos: Directo: 6356641 PBX: 6344000 Ext. 3 117 o 3131.

**Anexo H: Tabla 7.** Datos genotípicos poblacionales y parámetros estadísticos forenses usando un panel de 38 marcadores autosómicos de inserción-delección en Santander, Colombia.

CI	Numero rs.	Frecuencias Alélicas		HO	HE	EHW	Parámetros Forenses				
		Alelo 1	Alelo 2				PC	CIP	IP	PD	PE
B1	rs34541393	0,429	0,571	0,44200	0,49041	0,02811	0,361	0,37	0,90	0,639	0,142
B2	rs16624	0,548	0,452	0,53200	0,49589	0,10602	0,397	0,37	1,07	0,603	0,217
B3	rs2307689	0,411	0,589	0,47000	0,48464	0,51946	0,377	0,37	0,94	0,623	0,163
B4	rs35769550	0,529	0,471	0,45000	0,49882	0,03197	0,355	0,37	0,91	0,645	0,147
B5	rs2307700	0,455	0,545	0,49800	0,49645	1,00000	0,378	0,37	1,00	0,622	0,186
B6	rs140809	0,243	0,757	0,36200	0,36827	0,71400	0,467	0,30	0,78	0,533	0,092
B7	rs3047269	0,523	0,477	0,33800	0,49944	0,00000	0,334	0,37	0,75	0,666	0,079
B8	rs33972805	0,475	0,525	0,49400	0,49925	0,85701	0,373	0,37	0,99	0,627	0,182
B9	rs33917182	0,673	0,327	0,41400	0,44058	0,18574	0,403	0,34	0,85	0,597	0,123
B10	rs16402	0,312	0,688	0,43200	0,42974	1,00000	0,419	0,34	0,88	0,581	0,135
G1	rs1610871	0,437	0,563	0,47400	0,49255	0,41271	0,371	0,37	0,95	0,629	0,166
G2	rs2067238	0,566	0,434	0,51600	0,49178	0,27535	0,392	0,37	1,03	0,608	0,202
G3	rs2067294	0,393	0,607	0,49400	0,47758	0,45421	0,394	0,36	0,98	0,606	0,181
G4	rs2307710	0,335	0,665	0,35000	0,44600	0,00000	0,388	0,35	0,77	0,612	0,086
G5	rs2308242	0,206	0,794	0,31600	0,32746	0,49460	0,507	0,27	0,73	0,493	0,070
G6	rs2307580	0,595	0,405	0,39000	0,48243	0,00003	0,356	0,37	0,82	0,644	0,107
G7	rs1160956	0,624	0,376	0,40800	0,46972	0,00418	0,372	0,36	0,84	0,628	0,119
G8	rs34511541	0,541	0,459	0,53000	0,49714	0,15028	0,395	0,37	1,06	0,605	0,215
G9	rs2307978	0,253	0,747	0,34200	0,37836	0,03255	0,455	0,31	0,76	0,545	0,082

Y1	rs3051300	0,346	0,654	0,43200	0,45302	0,32364	0,395	0,35	0,88	0,605	0,135
Y2	rs10629077	0,299	0,701	0,43000	0,41962	0,59613	0,428	0,33	0,88	0,572	0,133
Y3	rs10688868	0,307	0,693	0,39800	0,42593	0,14370	0,414	0,33	0,83	0,586	0,113
Y4	rs2067208	0,226	0,774	0,34800	0,35020	0,89815	0,484	0,29	0,77	0,516	0,085
Y5	rs2307579	0,316	0,684	0,39600	0,43272	0,06155	0,407	0,34	0,83	0,593	0,112
Y6	rs2308020	0,633	0,367	0,46200	0,46509	0,92356	0,394	0,36	0,93	0,606	0,156
Y7	rs3080855	0,375	0,625	0,45800	0,46922	0,63203	0,388	0,36	0,92	0,612	0,153
Y8	rs1610919	0,519	0,481	0,40200	0,49978	0,00000	0,341	0,37	0,84	0,659	0,115
Y9	rs2307839	0,272	0,728	0,36400	0,39643	0,06952	0,439	0,32	0,79	0,561	0,093
R1	rs2308137	0,27	0,73	0,37600	0,39459	0,30909	0,442	0,32	0,80	0,558	0,100
R2	rs36040336	0,647	0,353	0,46200	0,45724	0,84339	0,401	0,35	0,93	0,599	0,156
R3	rs1160886	0,434	0,566	0,47200	0,49178	0,41515	0,371	0,37	0,95	0,629	0,164
R4	rs2308026	0,402	0,598	0,48800	0,48127	0,77609	0,388	0,37	0,98	0,612	0,177
R5	rs2307526	0,349	0,651	0,45800	0,45485	0,92166	0,402	0,35	0,92	0,598	0,153
R6	rs34811743	0,75	0,25	0,36400	0,37538	0,55176	0,460	0,30	0,79	0,540	0,093
R7	rs2308189	0,517	0,483	0,49800	0,49992	1,00000	0,375	0,37	1,00	0,625	0,186
R8	rs5895447	0,274	0,726	0,36000	0,39825	0,03331	0,437	0,32	0,78	0,563	0,091
R9	rs2308171	0,196	0,804	0,27600	0,31548	0,00676	0,523	0,27	0,69	0,477	0,054
R10	rs35605984	0,551	0,449	0,49000	0,49529	0,85765	0,375	0,37	0,98	0,625	0,179
		<b>Heterocigosidad</b>									
		<b>Media</b>		0,42595	0,44849		<b>PD. Acumulado</b>		1,0000		
		<b>DE.</b>		0,06442	0,05303		<b>PE. Acumulado</b>		0,99619		

**CI.** Código interno; **HO.** Heterocigosidad observada; **HE.** Heterocigosidad esperada; **EHW.** Equilibrio de Hardy-Weinberg; **PC.** Probabilidad de coincidencia; **CIP.** Contenido de información polimórfica; **IP.** Índice de paternidad; **PD.** Poder de discriminación; **PE.** Probabilidad de exclusión; **DE.** Desviación estándar.

La tabla presenta las formas bialélicas de cada uno de los 38 INDELs tipificados en la población santandereana y su distribución alélica. Así mismo se presentan los resultados obtenidos durante la evaluación de parámetros forenses para dichos marcadores, evaluación de Hardy-Weinberg y heterocigosidades observadas y esperadas en la población. Los colores representan el fluorocromo con el cual cada INDEL está marcado.

**Anexo I: Tabla 8.** Frecuencias alélicas para 38 marcadores autosómicos INDELS evaluadas en cada una de las regiones provinciales del departamento de Santander.

Numero rs.	Cl.	Alelos	Regiones Provinciales de Santander					
			Comunera	García Rovira	Guanentá	Mares	Soto	Vélez
rs34541393	B1	A1	0,33333	0,56897	0,42105	0,4359	0,39894	0,57609
		A2	0,66667	0,43103	0,57895	0,5641	0,60106	0,42391
rs16624	B2	A1	0,5	0,58621	0,48684	0,5	0,57092	0,54348
		A2	0,5	0,41379	0,51316	0,5	0,42908	0,45652
rs2307689	B3	A1	0,44444	0,51724	0,40789	0,40385	0,40071	0,40217
		A2	0,55556	0,48276	0,59211	0,59615	0,59929	0,59783
rs35769550	B4	A1	0,64815	0,55172	0,52632	0,50641	0,52305	0,52174
		A2	0,35185	0,44828	0,47368	0,49359	0,47695	0,47826
rs2307700	B5	A1	0,44444	0,44828	0,48684	0,38462	0,4805	0,40217
		A2	0,55556	0,55172	0,51316	0,61538	0,5195	0,59783
rs140809	B6	A1	0,2037	0,31034	0,26316	0,23718	0,23404	0,27174
		A2	0,7963	0,68966	0,73684	0,76282	0,76596	0,72826
rs3047269	B7	A1	0,57407	0,5	0,59211	0,51282	0,52305	0,46739
		A2	0,42593	0,5	0,40789	0,48718	0,47695	0,53261
rs33972805	B8	A1	0,5	0,58621	0,43421	0,50641	0,47163	0,3913
		A2	0,5	0,41379	0,56579	0,49359	0,52837	0,6087
rs33917182	B9	A1	0,62963	0,72414	0,65789	0,67308	0,67376	0,67391
		A2	0,37037	0,27586	0,34211	0,32692	0,32624	0,32609
rs16402	B10	A1	0,2037	0,25862	0,32895	0,29487	0,3227	0,3587
		A2	0,7963	0,74138	0,67105	0,70513	0,6773	0,6413
rs1610871	G1	A1	0,57407	0,5	0,44737	0,39103	0,41844	0,5
		A2	0,42593	0,5	0,55263	0,60897	0,58156	0,5
rs2067238	G2	A1	0,53704	0,56897	0,5	0,53205	0,56738	0,68478
		A2	0,46296	0,43103	0,5	0,46795	0,43262	0,31522
rs2067294	G3	A1	0,42593	0,32759	0,39474	0,37179	0,40071	0,40217
		A2	0,57407	0,67241	0,60526	0,62821	0,59929	0,59783
rs2307710	G4	A1	0,38889	0,24138	0,38158	0,37821	0,33333	0,26087
		A2	0,61111	0,75862	0,61842	0,62179	0,66667	0,73913
rs2308242	G5	A1	0,16667	0,15517	0,19737	0,24359	0,21454	0,15217
		A2	0,83333	0,84483	0,80263	0,75641	0,78546	0,84783
rs2307580	G6	A1	0,53704	0,60345	0,61842	0,57051	0,59752	0,63043
		A2	0,46296	0,39655	0,38158	0,42949	0,40248	0,36957
rs1160956	G7	A1	0,66667	0,72414	0,65789	0,53205	0,61702	0,70652
		A2	0,33333	0,27586	0,34211	0,46795	0,38298	0,29348

rs34511541	G8	A1	0,51852	0,56897	0,55263	0,55128	0,53191	0,56522
		A2	0,48148	0,43103	0,44737	0,44872	0,46809	0,43478
rs2307978	G9	A1	0,24074	0,18966	0,13158	0,26923	0,27482	0,23913
		A2	0,75926	0,81034	0,86842	0,73077	0,72518	0,76087
rs3051300	Y1	A1	0,42593	0,36207	0,42105	0,26282	0,33865	0,41304
		A2	0,57407	0,63793	0,57895	0,73718	0,66135	0,58696
rs10629077	Y2	A1	0,31481	0,25862	0,26316	0,27564	0,31383	0,29348
		A2	0,68519	0,74138	0,73684	0,72436	0,68617	0,70652
rs10688868	Y3	A1	0,33333	0,27586	0,30263	0,26282	0,31915	0,31522
		A2	0,66667	0,72414	0,69737	0,73718	0,68085	0,68478
rs2067208	Y4	A1	0,16667	0,27586	0,30263	0,15385	0,23582	0,22826
		A2	0,83333	0,72414	0,69737	0,84615	0,76418	0,77174
rs2307579	Y5	A1	0,2963	0,24138	0,38158	0,36538	0,29965	0,33696
		A2	0,7037	0,75862	0,61842	0,63462	0,70035	0,66304
rs2308020	Y6	A1	0,74074	0,62069	0,60526	0,67949	0,62057	0,59783
		A2	0,25926	0,37931	0,39474	0,32051	0,37943	0,40217
rs3080855	Y7	A1	0,27778	0,27586	0,40789	0,4359	0,38298	0,31522
		A2	0,72222	0,72414	0,59211	0,5641	0,61702	0,68478
rs1610919	Y8	A1	0,48148	0,55172	0,51316	0,55128	0,51773	0,47826
		A2	0,51852	0,44828	0,48684	0,44872	0,48227	0,52174
rs2307839	Y9	A1	0,2963	0,24138	0,28947	0,28846	0,26064	0,30435
		A2	0,7037	0,75862	0,71053	0,71154	0,73936	0,69565
rs2308137	R1	A1	0,18519	0,2069	0,22368	0,33974	0,25887	0,34783
		A2	0,81481	0,7931	0,77632	0,66026	0,74113	0,65217
rs36040336	R2	A1	0,7037	0,56897	0,65789	0,66026	0,63652	0,69565
		A2	0,2963	0,43103	0,34211	0,33974	0,36348	0,30435
rs1160886	R3	A1	0,53704	0,58621	0,32895	0,39103	0,43794	0,41304
		A2	0,46296	0,41379	0,67105	0,60897	0,56206	0,58696
rs2308026	R4	A1	0,5	0,36207	0,38158	0,39103	0,39362	0,45652
		A2	0,5	0,63793	0,61842	0,60897	0,60638	0,54348
rs2307526	R5	A1	0,2963	0,39655	0,30263	0,35256	0,37057	0,25
		A2	0,7037	0,60345	0,69737	0,64744	0,62943	0,75
rs34811743	R6	A1	0,68519	0,75862	0,82895	0,71154	0,75887	0,72826
		A2	0,31481	0,24138	0,17105	0,28846	0,24113	0,27174
rs2308189	R7	A1	0,46296	0,60345	0,5	0,49359	0,5266	0,48913
		A2	0,53704	0,39655	0,5	0,50641	0,4734	0,51087
rs5895447	R8	A1	0,27778	0,41379	0,27632	0,26923	0,26064	0,27174
		A2	0,72222	0,58621	0,72368	0,73077	0,73936	0,72826

rs2308171	R9	A1	0,09259	0,17241	0,18421	0,25	0,19149	0,21739
		A2	0,90741	0,82759	0,81579	0,75	0,80851	0,78261
rs35605984	R10	A1	0,57407	0,63793	0,56579	0,48718	0,5656	0,48913
		A2	0,42593	0,36207	0,43421	0,51282	0,4344	0,51087
<b>HMO.</b>			0,4386	0,40926	0,41482	0,43084	0,42404	0,44165
<b>HME</b>			0,43996	0,44075	0,44801	0,45121	0,44883	0,44796
<b>D.E</b>			0.09012	0.09559	0.11037	0.07867	0.06877	0.09802

**CI.** Código interno; **HMO.** Heterocigosidad media observada; **HME.** Heterocigosidad media esperada; **DE.** Desviación estándar.

La tabla presenta las formas bialélicas de cada uno de los 38 INDELS tipificados para las seis regiones provinciales del departamento de Santander y su distribución alélica. Así mismo se presentan los resultados obtenidos durante la evaluación de las heterocigosidades observadas y esperadas en la población. Los colores representan el fluorocromo con el cual cada INDEL está marcado.

**Anexo J: Tabla 9.** Frecuencias genotípicas para 38 marcadores autosómicos INDELS evaluadas en una muestra poblacional del departamento de Santander.

Genotipo	Frec B1	Frec B2	Frec B3	Frec B4	Frec B5	Frec B6	Frec B7	Frec B8	Frec B9	Frec B10
11	0,208	0,282	0,176	0,304	0,206	0,062	0,354	0,228	0,466	0,096
12	0,442	0,532	0,47	0,45	0,498	0,362	0,336	0,494	0,414	0,432
22	0,35	0,186	0,354	0,246	0,296	0,576	0,31	0,278	0,12	0,472
<b>Homoc.</b>	<b>55,8%</b>	<b>46,8%</b>	<b>53,0%</b>	<b>55,0%</b>	<b>50,2%</b>	<b>63,8%</b>	<b>66,4%</b>	<b>50,6%</b>	<b>58,6%</b>	<b>56,8%</b>
<b>Hetero.</b>	<b>44,2%</b>	<b>53,2%</b>	<b>47,0%</b>	<b>45,0%</b>	<b>49,8%</b>	<b>36,2%</b>	<b>33,6%</b>	<b>49,4%</b>	<b>41,4%</b>	<b>43,2%</b>

Genotipo	Frec G1	Frec G2	Frec G3	Frec G4	Frec G5	Frec G6	Frec G7	Frec G8	Frec G9	
11	0,2	0,308	0,146	0,16	0,048	0,402	0,42	0,276	0,082	
12	0,474	0,516	0,492	0,35	0,316	0,388	0,408	0,53	0,342	
22	0,326	0,176	0,362	0,49	0,636	0,21	0,172	0,194	0,576	
<b>Homoc.</b>	<b>52,6%</b>	<b>48,4%</b>	<b>50,8%</b>	<b>65,0%</b>	<b>68,4%</b>	<b>61,2%</b>	<b>59,2%</b>	<b>47,0%</b>	<b>65,8%</b>	
<b>Heteroc.</b>	<b>47,4%</b>	<b>51,6%</b>	<b>49,2%</b>	<b>35,0%</b>	<b>31,6%</b>	<b>38,8%</b>	<b>40,8%</b>	<b>53,0%</b>	<b>34,2%</b>	

Genotipo	Frec Y1	Frec Y2	Frec Y3	Frec Y4	Frec Y5	Frec Y6	Frec Y7	Frec Y8	Frec Y9	
11	0,13	0,084	0,108	0,052	0,118	0,402	0,146	0,318	0,09	
12	0,432	0,43	0,398	0,348	0,396	0,462	0,456	0,402	0,364	
22	0,438	0,486	0,494	0,6	0,486	0,136	0,398	0,28	0,546	
Homoc.	56,8%	57,0%	60,2%	65,2%	60,4%	53,8%	54,2%	59,8%	63,6%	
Heteroc.	43,2%	43,0%	39,8%	34,8%	39,6%	46,2%	45,8%	40,2%	36,4%	

Genotipo	Frec R1	Frec R2	Frec R3	Frec R4	Frec R5	Frec R6	Frec R7	Frec R8	Frec R9	Frec R10
11	0,082	0,416	0,198	0,158	0,12	0,568	0,268	0,094	0,058	0,306
12	0,376	0,462	0,472	0,488	0,458	0,364	0,498	0,36	0,276	0,49
22	0,542	0,122	0,33	0,354	0,422	0,068	0,234	0,546	0,666	0,204
Homoc.	62,4%	53,8%	52,8%	51,2%	54,2%	63,6%	50,2%	64,0%	72,4%	51,0%
Heteroc.	37,6%	46,2%	47,2%	48,8%	45,8%	36,4%	49,8%	36,0%	27,6%	49,0%

**Frec.** Frecuencia genotípica para cada marcador analizado; **Homoc.** Porcentaje de homocigosidad presente en la población; **Heteroc.** Porcentaje de heterocigosidad presente en la población.

La tabla presenta la distribución de las frecuencias de los genotipos encontrados en la población para cada uno de los 38 INDELS tipificados en la muestra poblacional del departamento de Santander. Los colores representan el fluorocromo con el cual cada INDEL está marcado.

**Anexo K: Tabla 10.** Análisis de los estadísticos F y sus respectivos valores de significancia encontrados en una muestra poblacional del departamento de Santander.

Índices de Fijación		P- Valor
FIS	0.03834	0.00832
FST	0.00081	1.000
FIT	0.03911	0.00733

**Anexo L: Tabla 11.** Evaluación de probabilidades a priori (LnP) y posteriori ( $P_K$ ), en el análisis de asignación de individuos a las poblaciones.

K	Ln	Mean Ln	P(K)
1	-48378	-48368,4	#¡NUM!
2	-47918	-47604,7	4,2658E-213
3	-47746	-47185,7	7,8407E-139
4	-47500	-46810,8	5,38019E-32
5	-47430	-46588,4	1
6	-47755	-46464,2	9,6762E-143

La tabla muestra valores de probabilidad a priori (LnP) y posteriori ( $P_K$ ), evaluados mediante el software Structure v2.3.3 (Pritchard et al. 2000). Un valor de ( $P_K$ ) = 1, correspondió al mejor ajuste del número de poblaciones que contribuyen al pool genético de la población santandereana (K= 5).