



UNIVERSIDAD NACIONAL DE COLOMBIA

Validación y consistencia de información en estudios de diversidad genética humana a partir de marcadores microsatélites.

William Usaquén Martínez

Universidad Nacional de Colombia
Facultad de Ciencias, Departamento de Biología
Bogotá, Colombia
2012

Validación y consistencia de información en estudios de diversidad genética humana a partir de marcadores microsatélites.

William Usaquén Martínez M.Sc.

Tesis o trabajo de investigación presentada(o) como requisito parcial para optar al título de:

Doctor en Ciencias – Biología.

Director.

Luis Fernando García Pinzón Ph.D.
Departamento de Biología.

Director Asociado

Jimmy Corzo Salamanca Dr.rer.nat.
Departamento de Estadística.

Línea de Investigación:

Genética de Poblaciones Humanas

Grupo de Investigación:

Genética de Poblaciones e Identificación

Universidad Nacional de Colombia
Facultad de Ciencias, Departamento de Biología
Bogotá, Colombia
2012

Dedicado ...

A mi maestra Indiana Pura Bustos Bustos, creo que con este documento terminó la tarea querida maestra... o será solo el inicio?. Tu me has enseñado a no estar satisfecho, a buscar algo mas, a no comer entero, he tratado de cultivar la agudeza mental y el buen corazón, lo que ha sido tu marca personal, y que espero sea la mía y la de mis estudiantes, he tratado de hacer la escuela que soñamos... en eso ando, transmitiendo tus enseñanzas de generación en generación.

Con gran cariño por ti Indiana.

A mis estudiantes, tendrán que disculparme por la severidad, por el temperamento, pero bien saben que les he dado lo mejor y lo peor que tengo, deben escoger. Les he presionado y fortificado en lo que se hacer y les he dejado incompletos en mi propia incompetencia, deben superarse. Por ejemplo, los se buenos analistas de datos, pero con necesidad de mas horas de laboratorio, halla paz, ya encontraran sus propios caminos. Deben seguir su educación para marcar a una generación de colombianos ilustres que reoriente el cause de nuestra historia.

A mi esposa, ¿Qué seria de mis días sin tu sonrisa?. Gracias por venir.

A mi gente, lo mas bello de este mundo, siempre me hacen sentir un orgullo profundo...

Bogotá, diciembre de 2011.

Carta de Agradecimientos.

La lectura de esta parte no es necesaria, pero se recomienda por el contexto.

Mis mas profundo agradecimientos a todas las personas que menciono en la siguiente carta. Este trabajo no es mío es de ellos, es una versión ampliada y puesta en papel de las discusiones sostenidas.

En el año 1999, mientras iniciaba tramites para realizar estudios fuera del país, decidí cursar la especialización en estadística en la Universidad Nacional de Colombia. No alcance a percibir lo determinante que seria esta decisión para mi. Algo que inicialmente seria un programa tendiente a mejorar mi nivel en un área normalmente descuidada en los pregrados colombianos, termino rediseñando mi futuro. En los estudios biológicos, con frecuencia surgen múltiples preguntas estadísticas, especialmente sobre el tamaño de muestra, en ese programa tuve algunas primeras discusiones con el profesor Leonardo Bautista, Luis Guillermo Díaz, Jorge Ortiz, charlas que tomaron forma cuando conocí al profesor Jimmy Corzo, y quien me fue mostrando un panorama de la estadística amplio, en el que además percibí como las respuestas sobre el tema serian mucho mas complejas.

Estas discusiones se articulaba con el momento que vivía el país, se estaban haciendo las primeras estandarizaciones para el uso de STR's en el país, específicamente en el Instituto Nacional de Medicina Legal (INML). Como sucede con cualquier novedad, el uso del analizador genético ABI 310, representaba un gran avance tecnológico. Éramos muchos los que esperábamos paciente mente la oportunidad de ponerle las manos encima al nuevo aparato, sin embargo en ese momento gracias al biólogo Fernando Parra Velandia, encontré

un juguete mucho mas atractivo: El lenguaje de programación Visual Basic Applications (VBA), un nuevo cambio en la vida, bien sabe un programador que un nuevo lenguaje es adictivo, con lo cual me aleje por completo del mundo del laboratorio.

Ese momento histórico para el país, también marco para mi una década de actividades: Trabajo en bases de datos con Adriana Castillo y Clara Vargas en la Universidad Industrial de Santander, Adriana Ibarra y Aníbal Gaviria en la Universidad de Antioquia, Julieta Henao y Liliana Porras en la Universidad Tecnológica de Pereira, María Victoria Gómez en la Universidad del Valle, Amparo Acosta en la Universidad del Cauca, Carlos Restrepo en la Universidad del Rosario, Andrés Gutiérrez y Alejandro Giraldo en la Fundación Gillow, Dora Fonseca y Andrea Gómez en Genética molecular de Colombia en Bogotá y a Juan José Builes del Laboratorio Nacional de Referencia. En todos los lugares, analizando tipificaciones genéticas año tras año me dio la experiencia para detectar variantes raras, en mirar lo que pueden ser posibles rutas de migración y poblamiento, en la fantasía de tener un sistema de bases de datos en línea para todo el país y la región andina, en lo bueno que seria hacer estudios de campo.

Luego apareció el Instituto de Genética de la Universidad Nacional. Ya han pasado seis años de permanencia. Muchas Altas y bajas. Aquí he conformado el grupo de Genética de poblaciones e Identificación Humana, donde se ha cocinado lentamente este trabajo.

Miguel Novoa fue la primera persona con quien inicie esta meta, fue mi primer compañero en la exploración de las aplicaciones en genética y el sin numero de supuestos, un compañero incansable, riguroso, trabajar con el significan horas, muchas horas de concentración. Posteriormente, se nos unió Leonardo Eljach, un compañero leal, agudo, buen pensador, he perdido la cuenta del numero de discusiones que tuvimos dando forma al proyecto Guajira. El tiempo que pasamos de viaje, conociendo las rancherías Wayuu, es en mis recuerdos una estupenda experiencia de vida.

Con algunas ideas ya maduras, pero solo eso, ideas, llegaron a nuestro grupo, Luz Ángela Alonso, Madelyn Yiseth Rojas y Vanessa Sarmiento. Empecé el entrenamiento de las chicas superpoderosas desde lo básico, ya han construido sus propios elementos, su propia personalidad científica, pero además dieron orden, forma, ritmo y dinámica al día a día de trabajo. Personas alegres, atentas al detalle, cuidadosas y perseverantes, hicieron realidad ser un grupo registrado. De ese empuje fue cierto ganar convocatorias, ejecutar presupuestos, tener computadores, organizar cursos, participar en eventos, hacer campo para estudios Genético poblacionales. Al fin todo fue cierto.

Ángel Carracedo, alguna vez me dijo que el procuraba vincular a su grupo personas de muy buena energía, he seguido a puño y letra su consejo, fue así como al grupo inicial, de estudiosos en aplicaciones, estadística y poblaciones, en el ultimo tiempo se ha unido Jenny Blanco, mujer única, disciplinada, alegre, espontánea, trabajadora incansable, Verónica Rocha, espiritual, meticulosa, leal, y Sandra Chala, sorprendente, previsiva, una mujer que siempre esta un paso adelante, ellas, dirigidas por Andrea (de quien hablare mas adelante) son nuestro brazo fuerte en las técnicas de laboratorio. Dobereiner Chala, nuestro joven hombre candidato a sabio, nuestra perspectiva en antropología y arqueología y Deicy Amparo Osorio, quien hace justicia a su nombre, ella ese quien nos ampara. Lo que fue un sueño es ahora una empresa científica de la cual ella es la administradora.

Quiero agradecer a mi Hermano, Luis Alejandro Camero Ramos, hermano por afinidad, como ya habrán descubierto los analistas de isonimia. Mis tiempos de madurez en la perspectiva filosófica de estas ideas se las debo a este Biólogo Filósofo, que vive de forma ascética, pero de quien la comunidad científica colombiana debería saber mas.

Finalmente, Lilian Andrea Casas Vargas, quien llego a mitad del proceso, buscando trabajar con un grupo impetuoso, de una manera tranquila, nos ha dado una lección importante sobre disciplina, perseverancia, cariño y crecimiento personal, ejemplo del buen

trabajo, buena amiga, buena compañera, ella es quien le ha dado una articulación y estructuración única al equipo. Tal como lo soñé desde el principio, la parte de identificación no debería estar separada del trabajo en poblaciones, esta charnela es su aporte, Andrea es la unión y el equilibrio.

Este es el grupo con el cual llegamos al final de este viaje, Con este grupo de personas, que en promedio de edad tienen 26 años tienen múltiples bendiciones. En este grupo hay un derroche de energía, no apto para cardiacos, una capacidad de lectura, de abordaje, un ímpetu sin igual, ya al frente de este equipo, de ese tren, es muy difícil bajarse, como en las montañas rusas, solo tengo opción de aferrarme fuerte cerrar los ojos y gritar, porque el viaje empezó.

Si de algo he disfrutado en este proceso es de libertad, esa que solo brindan los verdaderos maestros, Luis Fernando García Pinzón y Jimmy Corzo Salamanca, me han dirigido con aprecio, dedicación, rigor, esmero y paciencia. Han sobrellevado mis limitaciones de una manera sabia. Las muchas horas de trabajo, en sus oficinas o al calor de un café, permitiéndome equivocarme y corrigiéndome, es un gesto para el que solo tengo palabras de aprecio y gratitud. Este trabajo crea entre nosotros un vínculo que agradeceré por el resto de la vida.

Aunque el contacto fue menor, no menos importante, A la profesora Leonor Gusmão, quien en momentos difíciles reviso las principales ideas y me dio un impulso importante para seguir adelante, Al profesor Fabricio Santos quien me dio enfoques importantes en el análisis poblacional, Al profesor Alberto Gómez, Hombre admirable de quien indirectamente he disfrutado sus clases, al profesor Guillermo Barreto, un estudioso de las poblaciones colombianas quien me dio ideas importantes sobre el trabajo con comunidades. Finalmente, agradecemos a todos los colombianos que nos han dado un pedazo de si para lograr esta propuesta metodológica. Sinceramente reiteramos nuestro compromiso por hacer un país mejor a través de nuestros estudios para nuestra filial 1.

Durante todos estos años tuvimos el apoyo de más estudiantes de pregrado y postgrado que participaron en las tomas de muestra, apoyo al trabajo del laboratorio, diseño de material y en las tareas diarias del laboratorio. Entre ellos se destacan, Blanca Schroeder, Rita Baldrich, Wilson Sierra, Natalia Lamprea. Al personal de la división de Investigaciones en Bogotá, a la señora Lucy Gaitán, e Islena Bonilla, a las monitoras Julieth Castiblanco Quiroga y Jenny Barrera Velandia, además a una lista muy grande de personas en las diferente partes donde realizamos las diferentes partes del proyecto.

William Usaquén Martínez.
Estudiante Doctorado en Ciencias.
Universidad Nacional de Colombia.

Resumen

En los últimos veinte años en Colombia se ha dado un gran desarrollo en los estudios poblacionales, en buena parte por el auge de la genética forense de acuerdo con nuestra compleja realidad nacional, como producto derivado se ha dado la importación al país de tecnología suficiente para el desarrollo de nuevas técnicas en biología molecular. Sin embargo dos aspectos metodológicos como los cálculos de tamaño de muestra y la adecuada selección de las muestras no han tenido el mismo desarrollo en los estudios poblacionales, manteniéndose sin criterios específicos, en ocasiones generando resultados pocos claros y con estimaciones confusas sobre nuestras poblaciones.

Buscando realizar un aporte en estas dos direcciones, en el presente trabajo se presenta un nuevo algoritmo para la determinación de tamaños de muestra en estudios poblacionales. Este además arroja estimaciones de frecuencias alélicas asociadas a intervalos de confianza, lo que permite aproximaciones más confiables a las realizadas por conteo directo. Junto a este algoritmo se propone un método fundamentado en el análisis genealógico y demográfico para mejorar la selección muestral. La evaluación del algoritmo, se realizó con la base de datos de los usuarios de pruebas de filiación, mientras que para el análisis demográfico se realizaron estudios de campo en las poblaciones del archipiélago de San Andrés, Providencia y Santa Catalina, en el Caribe colombiano y en la península de la Guajira, población con una fuerte ancestría de la comunidad indígena Wayuu.

Los resultados obtenidos en los estudios muestrales indican tamaños cercanos a las 700 personas en poblaciones grandes y con marcada influencia de múltiples grupos ancestrales, dada las altas de migración que caracteriza la población de Bogotá. Por otra parte los análisis demográficos permitieron la construcción de múltiples grupos

poblacionales, de lo cual se generaron modelos de estructura *a priori* a los obtenidos con la información genética. El uso de variables demográficas permitió una gran definición de las poblaciones en el espacio y en el tiempo, independientes de las básicas clasificaciones político-administrativas, e incluso de las geográficas. Además, las variables demográficas, permitieron evaluar desde otra óptica los efectos de las diferentes fuerzas de cambio evolutivo en las poblaciones estudiadas.

Luego en el trabajo se intentó ir más allá, construyendo dos capítulos de reflexión, el primero dirigido a los principiantes en la genética de poblaciones, en el que se explora el origen de la disciplina desde “El origen de las especies”. El segundo dirigido a los investigadores que incursionan en las poblaciones y a quienes presentamos cinco propuestas sugeridas para diseñar e incorporar en sus estudios. Fue un reto lograr los muestreos en el trabajo de campo, por tanto consideramos que junto al refinamiento demográfico o la precisión matemática, se trata de un acercamiento humano, el cual debe ser cuidadosamente realizado y por sobre todo integrador entre una cultura científica y nuestras culturas ancestrales.

En los capítulos de reflexión me he permitido plantear mi experiencia de trabajo y después de mucho pensarlo sin lugar a duda creo que esta tesis, so pena de ser extensa es el escenario para hacerlo.

Palabras Clave: Intervalo de confianza, Muestreo, Genética de Poblaciones, Frecuencias alélicas.

Abstract

During the last twenty years, Colombia has greatly advanced in population studies, mainly for two reasons: (1) because necessary forensic genetics studies developed as a result of our complex national reality, (2) as a derived product as the country has enhanced in molecular biology tools. Nevertheless two methodological aspects including an adequate sample selection and sample size have not had the same development as the population studies. Instead, non-specific criteria exist, results may not be very clear or confusing estimates are generated about our populations. In order to contribute in further discussion on both issues, this thesis introduces a new algorithm to determine sample size in population studies. The output of the algorithm provides allelic frequencies associated to confidence intervals, which allows more reliable approaches to those performed by classical direct count. In addition, a new method, based on genealogical and demographic analyses is proposed in order to improve the selection of the sample.

To evaluate the algorithm, a database corresponding to paternity and filiation tests were used, where as for the demographic analysis two field works were carried out (1) with human populations of the archipelago of San Andres, Old Providence and Santa Catalina Islands in the Caribbean and (2) with populations of Wayuu, an Amerindian ethnic group with strong genetic ancestry of the Guajira Peninsula in northern Colombia. The results obtained in the study samples suggest samples sizes of about 700 people for large populations influenced by multiple ancestral groups and large migration rates. Demographic studies allowed proposing a multi-population structure, providing different models *a priori*. The use of demographic variables, instead of administrative and political boundaries, provides a good scenario in time and space for population studies. They also provide elements to understand the effect of evolutionary forces and the quality of

sampling.

In this work, we tried to go beyond throughout two additional chapters of discussion and reflection on the issues of population studies. The first one is aimed to those interested in population genetics by exploring the origin of the discipline, even from Darwin's "*the origin of species*". The second one is aimed to researchers in forensic genetics throughout five proposals suggested to incorporate in their studies.

To get the samples for this work was a worthwhile challenge, but together with the demographic refining of the sampling and the proposed mathematical tool, it certainly provides an integrative approach of our scientific culture and our ancestral cultures. In the reflection chapters, ideas and thoughts come from my own work experience for the last 10 years, so I thought this thesis would be a good scenario to write them out.

Key words: Confidence interval, Sampling, Population Genetics, Allele frequencies.

Tabla de Contenido.

Introducción General	9
Capitulo Introductorio: El Origen de las Especies y su relación con el inicio de la actual teoría de la herencia.....	13
Resumen.....	13
Abstract.....	13
El concepto de herencia	14
Antecedentes Históricos	14
El origen de las especies.....	16
El modelo pangenético.....	17
El extraño caso de Gregor Mendel.....	20
Siglo XX.....	21
Conclusión	23
Bibliografía.....	24
Capitulo No. 1: Introducción General sobre Elementos de Muestreo.....	25
Genética de Poblaciones.....	25
Desarrollo histórico de la genética de poblaciones.....	26
Estructura Investigativa de la Genética de Poblaciones.....	27
Modelos en genética de poblaciones.....	28
Algunos estudios relacionados en Colombia	29
Definición de Sistemas Microsatélites	30
Clasificación.....	31
Teoría de Muestreo	31
Conceptos iniciales de teoría de muestreo en genética.....	32
La Investigación por muestreo.	33
Muestreo probabilístico.....	34

Muestreo no probabilístico.....	35
Plan de Muestreo Estadístico.....	36
Pasos para la construcción del plan de muestreo.....	36
Fuentes de sesgos y varianza.....	37
Análisis Multivariado.....	38
Bibliografía.....	40
Capitulo No. 2: Una nueva aproximación al cálculo de tamaño de muestra en estudios genético poblacionales y forenses a partir de marcadores microsatélites.....	43
1. Introducción.....	44
2. Método propuesto.....	46
2.1. Fase 1: Cálculo secuencial de las frecuencias alélicas.....	46
2.2. Fase 2: Cálculo de coeficientes de variación.....	47
2.2.1. Promedios ponderados de frecuencias.....	47
2.2.2. Desviación Estándar.....	49
2.2.3. Coeficientes de variación.....	50
2.2.4. Alelos efectivos y alelos raros.....	50
2.3. Selección de un tamaño de muestra.....	51
2.4. Análisis multi loci.....	51
2.5. Aplicación desarrollada.....	51
2.6. Datos empleados.....	52
3. Resultados.....	52
3.1. Cálculo inicial para un solo sistema STR.....	52
3.2. Tamaño de muestra establecido para el sistema completo.....	54
4. Discusión.....	57
5. Conclusión.....	59
Bibliografía:.....	61
Capitulo No. 3: Análisis de la Estructura de la población de La Guajira: Una visión Genética Demográfica y Genealógica.....	63

Resumen.....	63
1. Introducción.....	64
2. Metodología.....	67
2.1 Análisis Molecular.....	70
2.2 Análisis Demográfico de la Muestra.....	70
2.3 Análisis Genético de las Agrupaciones Genealógicas.....	71
2.4 Análisis de las Agrupaciones Genéticas.....	72
3. Resultados.....	72
3.1 Análisis Demográfico de la Muestra.....	72
3.2 Análisis Genético de las Agrupaciones Genealógicas.....	77
3.3 Análisis Genético de las Agrupaciones Genéticas.....	85
4. Discusión.....	88
6. Conclusión.....	93
8. Bibliografía.....	95
8. Anexos.....	98
Capítulo No. 4: Análisis de apellidos y de cromosoma Y de las poblaciones raizales del archipiélago de San Andrés y Providencia (Colombia).....	103
Introducción.....	104
Materiales y Métodos.....	108
2.1 Población de estudio.....	108
2.2 Otras poblaciones.....	108
2.3 Análisis molecular.....	108
2.4 Análisis Estadísticos.....	111
2.5 Comparación con otras poblaciones.....	111
2.6 Análisis de apellidos.....	112
3. Resultados.....	112
3.1 Diversidad Genética y haplogrupos.....	112
3.2 Análisis de redes.....	114

3.3 Estructura poblacional.....	116
3.4 Comparación con otras poblaciones.....	117
Apellidos.....	119
3.5 Diversidad y número de apellidos.....	119
4.0 Discusión.....	122
5.0 Conclusión.....	125
Capitulo No. 5: Cinco propuestas de unificación entre genética de poblaciones y genética forense para la próxima década.....	133
Resumen.....	133
1. El origen de los estudios en Colombia.....	134
1.1. Semblanza Histórica.....	134
1.2. Un cuestionamiento sobre la actualidad.....	137
2. Fuentes de conocimiento de la Genética de poblaciones.....	137
2.1. Biología Molecular.....	138
2.2. Matemáticas y Estadística.....	140
2.3. El Análisis demográfico.....	142
3. Cinco propuestas de trabajo.....	146
3.1. Un análisis descriptivo de cada polimorfismo.....	146
3.2. Una información demográfica integrada a datos históricos, etnográficos y estadísticos.....	148
3.3. Genética clásica, isonimia.....	151
3.4. Una base de datos nacional: Para uso en genética de poblaciones, forense y para monitoreo en epidemiología.....	154
3.5. Unificación con el sistema de salud y la comunidad.....	158
Conclusión.....	159
Bibliografía.....	159
Anexo No. 1: Guía de Laboratorio para Análisis Genético.....	163

Introducción General.

El trabajo está compuesto por 6 capítulos a través de los cuales se desarrollan nuevos elementos estadísticos y demográficos para proponer mejoras en la investigación sobre genética de poblaciones. Esta disciplina en el país se ha desarrollado de una manera muy cercana a la genética forense, por tanto es frecuente la referencia de una hacia la otra. Los capítulos se pueden dividir en dos tipos: Técnicos, en los que se muestran los nuevos algoritmos y los estudios poblacionales, otros son los artículos de reflexión, tendientes a proponer cambios en las perspectivas de investigación, en los enfoques recomendados para la práctica científica y social. A continuación se presenta un resumen más detallado cada uno de ellos.

Capítulo introductorio: se presenta el concepto de herencia y su importancia en “El origen de las especies”. El gran vacío en la teoría Darwiniana para la explicación de la variabilidad en las poblaciones suscitó una gran carrera hacia la comprensión de la transmisión de caracteres a lo largo de las generaciones, lo que culminó en 1903 en la definición de la Genética como una disciplina nueva, cabe destacar que desde el origen mismo de esta disciplina, una de las principales subdivisiones creadas fue la genética de poblaciones, con la cual se logró la formulación matemática de la selección natural. Este capítulo, surgió como un ejercicio reflexivo del primer año de formación doctoral. Tratando de centrar la propuesta de investigación dentro de la genética de poblaciones actual, se realizó una exhaustiva mirada en el pasado. Puede pensarse que un capítulo de historia es una extravagancia, sin embargo hoy en día considero este elemento fundamental para una propuesta que quiera orientarse hacia el futuro. Es un elemento sumamente enriquecedor en la formación de cualquier investigador, permite evaluar formas de pensar y de proyectar. Algunos cuestionamientos actuales sobre la estructura, tamaño, estado y origen de una población, de hecho están vigentes desde el siglo XIX. En esta parte la conclusión más destacable

es la perspectiva de estar frente a un nuevo modelo de selección natural en las próximas décadas, quizá orientado desde nuevas disciplinas en las que nos gustaría proponer el análisis estadístico y demográfico.

Capítulo No. 1: Se presenta una revisión sobre algunos elementos que son importantes a modo de marco de referencia buscando un mejor avance a lo largo del documento, este parte fue considerada más como un punto de encuentro para contextualizar las principales ideas entre genética y estadística. Para el lector estadístico, ver la relación de las dos disciplinas puede abrir nuevas perspectivas en el modelamiento teórico. Para el Genetista, se trata de encontrar una revisión sobre elementos estadísticos a tener en cuenta en la construcción de los muestreos, sinceramente espero pueda ser una guía en la construcción de sus próximos trabajos de investigación.

Capítulo No. 2: Se presenta un método estadístico nuevo desarrollado para el cálculo de tamaños de muestra asociado a la determinación de intervalos de confianza que permitirán cambiar las estimaciones de modelos puntuales a modelos por intervalos. En este capítulo se logra relacionar el tamaño de un polimorfismo genético con el tamaño de la muestra. Es un capítulo técnico en el cual invertimos cinco años de arduo trabajo, implicó un fuerte esfuerzo de programación. Al final del trabajo, incluso aun merece algunos cambios, sin embargo como procedimiento está completo. Las perspectivas de uso de este algoritmo están en desarrollo. La aparición de las nuevas versiones de software tales como *Arlequín*, o *Structure*, incorporan procedimientos por intervalos de confianza, lo que significó un impulso estimulante para continuar este desarrollo. Creemos que el mundo forense sin lugar a dudas se verá beneficiado por esta metodología e incluso se logren proponer nuevas interpretaciones de una manera más probabilística que determinística en casos de filiación e identificación. **El anexo No. 1** es la guía completa de análisis del sistema CODIS construida con este algoritmo. Esta guía actualmente se utiliza en nuestro laboratorio como base de referencia en la resolución de casos de filiación. Próximamente será distribuida en cuatro laboratorios más del país para su uso y evaluación y luego implementada vía WEB.

Capítulo No. 3: Se presentan los resultados de investigación del proyecto Wayuu, es un proyecto pensado desde el principio para el desarrollo de métodos demográficos como es el uso de encuestas asociadas al análisis genealógico buscando un mejoramiento de los muestreos por conveniencia los cuales son los más utilizados en genética. Este proyecto además tuvo un amplio análisis preliminar de la población, tal que al llegar a campo contábamos con un buen conocimiento de los diferentes grupos humanos presentes en la región lo que redundó ampliamente en la optimización y calidad el muestreo. La delicada toma de muestras, el registro de un amplio número de variables y la eliminación de sesgos debidos al uso de muestras provenientes directamente de campo y no de casos de filiación e identificación, permitieron después utilizar una muy amplia gama de vías y métodos de análisis. Con frecuencia en nuestro grupo de investigación recibimos preguntas sobre como analizar datos, sin embargo la principal limitación que vemos al asesorar, es la poca información con la que se cuenta. Este es un análisis exhaustivo de una región y de un maravilloso grupo humano. Salimos de lo teórico y de la sugerencia para dar paso a la práctica.

Capítulo No. 4: Muestra los resultados del proyecto San Andrés, donde se utilizaron datos del cromosoma Y, nuevamente se utilizó una encuesta demográfica, sin embargo se realizó el abordaje de la información no desde marcadores autosómicos, sino desde la perspectiva de los haplotipos del cromosoma Y. En esta parte se pudo explorar los efectos de la selección de las muestras en procesos ancestrales. Aquí se utilizaron dos de los elementos más importantes de nuestras propuestas: por un lado los análisis de isonimia y por otro los modelos de inferencia haplotípica que están fundamentados en el análisis Bayesiano. Es la unificación de los métodos de muestreo, el análisis de haplotipos de cromosoma Y y el uso de herramientas de genética clásica, otro de los elementos que creemos esencial rescatar para los estudios poblacionales.

Capítulo No. 5: A modo de conclusión es un artículo de reflexión sobre la situación de la disciplina en nuestro país, surge como un modelo propositivo en contraste con los capítulos anteriores que estuvieron centrados en elementos teóricos. Basado en la experiencia, se que una propuesta teórica, de no encontrar un guía o propuesta de implementación, puede ser abandonada

rápidamente. En este capítulo se relaciona el análisis poblacional con los problemas prioritarios en salud. En la actualidad en el país la genética de poblaciones es considerada únicamente como una subdisciplina teórica, cuando realmente brinda los elementos prácticos para la comprensión de la diseminación y alcance de muchas patologías de origen genético e incluso algunas asociadas a procesos ambientales o del desarrollo. También se exploró el panorama de las bases de datos y de la demografía. Es nuestra pretensión a más largo plazo promover una ampliación de las fuentes de investigación en genética de poblaciones, mas parámetros de discusión, buscando salir de una época en la que estamos cayendo en una repetición constante de los trabajos de investigación y en ocasiones de la utilización de categorías y clasificaciones coloniales, que merecen ser repensadas según la América que somos y queremos ser.

Capítulo Introductorio: El Origen de las Especies y su relación con el inicio de la actual teoría de la herencia.

The Origin of Species and its Relation with the Birth of Present Genetic Theory

William Usaquén Martínez, M.Sc.

Instituto de Genética, Universidad Nacional de Colombia. Bogotá D.C., Colombia.
wusaquenm@unal.edu.co

Trabajo publicado en acta Biológica Colombiana: Acta biol. Colomb., Vol. 14 S, 2009 77 - 84

Resumen

En la primera parte se mostrarán algunos elementos históricos relacionados con la “*Philosophie Zoologique*” de Jean-Baptiste Lamarck, obra fundamental de la biología moderna merecedora de una mención especial en su aniversario número 200. Esta obra contiene algunos conceptos importantes para “*The Origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*” el trabajo fundamental de Charles Darwin. Posteriormente se revisará la idea general de la pangénesis y sus diferentes contradicciones en la explicación de la variabilidad necesaria para la acción de la selección natural que culminarían con el nacimiento de la genética. Es importante mostrar la posición particular de los experimentos mendelianos en relación con el paradigma evolutivo causado por el origen de las especies.

Palabras clave: pangénesis, herencia, lamarckismo, experimentos mendelianos, evolución.

Abstract

In the first part I present some historical elements related to the “*Philosophie Zoologique*” by Jean- Baptiste Lamarck, a fundamental work of modern biology. This work deserves a special mention in its 200th anniversary, since it contains some important concepts for “*The Origin of Species by Means of Natural Selection, or the Preservation of favored races in the struggle for life*” the principal work of Charles Darwin. Subsequently we review the pangenesis and various contradictions in the explanation of the variability required for the action of natural selection that culminate with the birth of genetics. In this paper is important show the particular position of Mendelian experiments in relation with the evolutionary paradigm caused by the origin of species.

Key words: pangenesis, heredity, Lamarckism, Mendelian experiments, evolution.

El concepto de herencia

La herencia genética formalmente definida es el proceso por el cual la descendencia de un organismo adquiere las características de su célula u organismo progenitor. Esta herencia genética intervendrá en la regulación de un organismo en cuatro aspectos esenciales: morfología, fisiología, lineamientos generales de desarrollo, y por último, algunos aspectos del comportamiento.

La preocupación por la herencia nace con la humanidad misma, asociada a los fenómenos de similitud y continuidad de la forma en una especie biológica, posteriormente, se incorporó la noción de variabilidad para conformar lo que hoy en día denominamos genética (Jacob, 1988).

Actualmente la genética se constituye en una de las principales disciplinas de las ciencias biológicas, por la amplia gama de explicaciones y predicciones en la expresión genotípica final de los seres vivos a diferentes niveles orgánicos, desde el molecular hasta el poblacional. Otro aspecto a resaltar de la genética es la definición de las fuerzas de cambio evolutivo que tratan de explicar la continuidad de las especies a lo largo del tiempo, manteniendo una alta reproducibilidad en la replicación del material genético, generación tras generación, pero a la vez con modelos que explican la variabilidad como elemento de respuesta a los procesos evolutivos.

Antecedentes Históricos

Jean-Baptiste Lamarck, en la “*Philosophie Zoologique*” plantea por primera vez y de manera formal el concepto de Biología como disciplina científica. Como antecedente a esta obra se encuentra “*Histoire naturelle, générale et particulière*”, escrita entre 1745 a 1788 por Georges Louis Leclerc conde de Buffon, una enciclopedia completa en 37 volúmenes, como producto de un esfuerzo individual. En esta obra Buffon da una primera organización a la embriología, desarrollando principios integradores para la definición de los vertebrados, haciendo énfasis en la anatomía comparada y la importancia de estudiar el medio ambiente de los organismos. Para 1735

Carl von Linné había publicado “*Systema Naturae*”, obra duramente criticada por el conde de Buffon quien sostenía que dicho sistema en lugar de aclarar la clasificación de los organismos servía para conducir a confusiones debido a la extrema complejidad de sus normas. A pesar de los reparos de Buffon, Lamarck utilizó el sistema binomial para realizar la primera clasificación de la flora y fauna francesa, mostrando la versatilidad del mismo, lo que le permite obtener una plaza en el museo nacional de historia natural de Francia.

En el pensamiento de Jean-Baptiste Lamarck confluyen tres elementos esenciales: las ideas políticas revolucionarias que le permiten una mayor libertad de pensamiento lejos del poder monárquico y clerical de la época, un vigoroso proceso de actualización en los diferentes hallazgos de los naturalistas contemporáneos, y finalmente un profundo conocimiento de las obras clásicas griegas, que le llevaron a adoptar diferentes elementos teóricos tales como el sistema atómico planteado por Demócrito y a relacionarlo con el origen de la vida por generación espontánea. En sus trabajos, Jean-Baptiste Lamarck, plantea el estudio de una química diferente basada en las propiedades emergentes de las sustancias producidas por los seres vivos, particularmente de los ácidos, que una vez sintetizados cambiaban su capacidad de corrosión determinando compuestos completamente nuevos. Según Ernest Haeckel, la Filosofía zoológica reúne el primer cuerpo de conocimientos organizados que enlaza el concepto de evolución con el concepto de herencia (Lamarck, 1986).

Los dos postulados referentes a la herencia, en el trabajo de Jean-Baptiste Lamarck fueron: 1. La herencia de caracteres adquiridos. 2. El principio del uso y desuso de órganos. En la primera de ellas, Lamarck da importancia al hábito como una característica heredable. Este principio fue fácilmente aceptado por Darwin en el origen de las especies como parte de su interpretación de la continuidad de las formas vivientes. Por otra parte, Lamarck plantea dentro de su modelo evolutivo una tendencia lineal, un modelo de causas finales hacia el desarrollo y mejoría de los organismos. La linealidad evolutiva, en el pensamiento de Lamarck, no hace referencia a un diseñador preexistente, no existe en la mente de Lamarck un dios que actuase como fundador de

las directrices para el desarrollo orgánico, pero si plantea un concepto teleológico en la evolución natural, asumiendo una posición de vitalista científico, similar al concepto aristotélico. En este punto se da un rompimiento conceptual entre Darwin y Lamarck. Para Darwin el proceso evolutivo no sigue una dirección final determinada. A pesar de estas diferencias, los modelos darwinianos de herencia y evolución están notablemente influenciados por el trabajo de Lamarck, a quien también debemos reconocer, como uno de los principales fundadores de la estructura epistemológica y filosófica de la Biología para llegar a constituirse en disciplina científica. En este punto es prudente recordar la siguiente cita:

“Nadie ignora que toda ciencia debe tener su filosofía, y que sólo por este camino puede hacer progresos reales. En vano consumirán los naturalistas todo su tiempo describiendo nuevas especies y en marcar todos los matices de sus variaciones para aumentar la lista inmensa de las especies inscritas, porque si la filosofía es olvidada sus progresos resultarán sin realidad y la obra entera quedará imperfecta”.

Philosophie Zoologique.

El origen de las especies

Para 1859 la obra ya había pasado por un largo proceso de maduración y muchas de las ideas planteadas por Darwin eran esperadas. Según las referencias históricas, cuando la obra es publicada, en pocos días logra agotarse la primera edición causando diferentes reacciones y una polaridad casi inmediata, desde muchos sectores de la sociedad diferentes a las ciencias naturales. Se dio inicio a múltiples interpretaciones convirtiéndola en obra multipropósito, no sólo científica sino incluso filosófica y religiosa, que rápidamente fue cayendo en la confusión debido a las diferencias en interpretación por parte de diferentes autores.

Se le atribuye al origen de las especies haber causado una división en relación con la Iglesia católica, sin embargo, una lectura cuidadosa de la obra deja ver a un Charles Darwin profundamente religioso, desinteresado en primera instancia por entrar en conflicto con la Iglesia. No en

vano en el último párrafo de su obra en la tercera edición atribuye el origen mismo de la vida y la grandeza de la naturaleza al creador.

Como una de las primeras grandes novedades introducidas en el Origen de las especies, está la proposición de población biológica como unidad de cambio evolutivo, a este nivel orgánico actuaría el recién propuesto principio de selección natural, por tanto, se hace indispensable proponer un mecanismo de herencia, por el cual la variabilidad pueda generarse y mantenerse a lo largo del tiempo geológico para dar origen a nuevas especies. En la edición de 1859 aún no se ha planteado un modelo de herencia, sin embargo para 1868 de una forma provisional, Darwin propone el modelo pangenético. Hoy en día, dicho modelo podría parecer superfluo, sin embargo, se examinarán algunos elementos para demostrar cómo Darwin no tomó a la ligera el problema de la herencia.

El modelo pangenético

La pangénesis en términos generales, describe la migración de partículas provenientes de las diferentes partes del cuerpo y que mantendrían las características estructurales y funcionales de cada órgano para mezclarse en la simiente y dar origen a un nuevo organismo. Es un modelo antiguo descrito en una forma preliminar por Demócrito quien habló de partículas mecánicas. Ya en la versión del siglo XVI, con una marcada influencia de los árabes, a quienes debemos la preservación de una buena parte de las obras de la antigüedad griega, la pangénesis se había reformulado según la teoría de los humores que habla del Atrabilis o bilis negra, la bilis, las flemas y la sangre como las sustancias básicas constitutivas del cuerpo, concordantes con los cuatro elementos clásicos del mundo griego. Del balance de estas cuatro sustancias dentro del organismo dependía la buena salud y el equilibrio orgánico del ser, también del equilibrio en las características de la personalidad. Por el contrario, los procesos patológicos estarían asociados con el déficit o exceso de alguna de estas sustancias. Por muy errada que parezca la pangénesis, le proporcionaba a Darwin un modelo de continuidad en el tiempo con una constante relación entre organismo, ambiente y desarrollo individual, lo que inicialmente es suficientemente coherente para ser aceptado dentro del mecanismo de selección natural como modelo de herencia.

La teoría pangenética también había sido importante para otros pensadores en biología, lo que le da confianza al mismo Darwin. Un ejemplo fue Carl von Linné, quien al proponer el sistema natural, inicialmente no encuentra la manera de clasificar al hombre (Boorstin, 1997). Nueve años después de la primera edición del sistema natural, establece diferentes razas del hombre basado en los tipos de carácter y personalidades humanas descritos por la teoría de los humores cuya herencia a lo largo del tiempo se mantenía por partículas (que posteriormente Darwin denominó pangenes).

En el siglo XVIII y XIX, con el desarrollo de la microscopía se introduce un mayor número de observaciones relacionadas con el funcionamiento del organismo, llevando a una mayor elaboración del modelo pangenético. Se pasa de cuatro elementos esenciales a la interpretación de cinco “tejidos”: sangre, huesos, vísceras, nervios y bilis; tejidos básicos generadores de pangenes, constituyentes de la simiente, que se asocian en una “mezcla” homogénea de las características provenientes por línea materna y paterna, para dar origen a un nuevo organismo. De esta época datan diagramas y atlas histológicos encargados de sustentar la formación de dicha simiente.

Finalmente, el modelo pangenético, conduce a tres problemas biológicos básicos para la articulación entre selección natural y herencia:

1. La selección natural, a partir del cual una especie, puede dar origen a otra, exige un modelo que explique cómo se mantiene la continuidad de las características hereditarias propias de un organismo a lo largo del tiempo, pero a la vez capaz de mantener la “plasticidad” para cambiar la naturaleza de la simiente en el tiempo y constituir una nueva especie.
2. Al aparecer el nivel orgánico denominado población, como la unidad de cambio evolutivo, requiere de un mecanismo generador de variabilidad constante sobre el cual pueda actuar la selección natural.

3. Los pangenes provenientes de cada uno de los tejidos corporales que realizarían una mezcla tendiente a la homogeneidad de la simiente, al cabo de pocas generaciones de apareamiento se llegaría a la uniformidad de las características hereditarias dentro de una población, lo cual limitaría el supuesto de variabilidad requerido para la acción del principio de selección natural. Fenómeno congruente con un esquema lamarckiano de evolución que Darwin adopta en ese momento para explicar la influencia directa con resultados definidos, del medio ambiente sobre los organismos.

Darwin adoptó la pangénesis dentro de su teoría, sin embargo, cabe recordar que para ese momento Matthias Schleiden, Friedrich Theodor Schwann y Rudolf Virchow, habían formulado las bases de la teoría celular con tres principios básicos y generales que permitían interpretar el mundo vegetal y el mundo animal desde una visión unificadora de los seres vivos. Darwin no aceptó completamente la teoría celular, en especial, el segundo principio referente a la preexistencia de cada célula (Ruiz y Ayala, 2002). Para 1864 (un año antes de la aparición de la última edición del origen de las especies), Louis Pasteur, realizó las demostraciones que ponen fin al principio de generación espontánea y que de inmediato apoyan la generalización de la teoría celular.

Al terminar de escribir el origen de las especies Darwin se encuentra enfrentado a una teoría celular cada vez más aceptada y a la abolición por completo del principio de generación espontánea, causando serias inconsistencias para explicar el origen inicial de los seres vivos en su obra. Darwin nos ofrece un mecanismo por el cual una especie se origina a partir de una especie preexistente, pero el inicio mismo de las formas vivas es fundamentalmente una creación divina.

En las notas originales se muestra como Darwin inició una serie de experimentos de acuerdo con los métodos propios de su época: cruzando híbridos de plantas. Se da a la tarea de buscar un mecanismo de herencia para explicar la variabilidad en las poblaciones, pero uno de los problemas fundamentales de sus experimentos, es que fueron desarrollados sobre la observación completa del organismo, enfrentándose de lleno a la herencia de caracteres complejos, que incluso hoy en día es

de difícil abordaje.

Esta serie de contradicciones dentro del pensamiento darwiniano de ninguna manera tienden a opacar la calidad conceptual de su obra, simplemente nos permiten interpretar a un autor dentro de su contexto histórico, puede verse como las contradicciones propias de su teoría fueron planteadas por él mismo. Estas observaciones nos permiten poner en su justa dimensión al origen de las especies y a Charles Darwin como un pensador de una alta capacidad crítica sobre sí mismo.

La contradicción entre un principio que nos explica el proceso de cambio en las especies, fundamentado en la variabilidad de la población, y por otra parte, en un mecanismo de herencia tendiente a disminuir dicha variabilidad, será el caldo de cultivo para generar un nuevo paradigma científico que será el objeto de estudio de las primeras generaciones de biólogos evolutivos a partir de 1868 y que derivará en lo que hoy conocemos como teoría genética.

El extraño caso de Gregor Mendel

Hoy en día se considera a Gregor Mendel como el padre de la genética debido a su trabajo de 1865, titulado “experimentos en híbridos de plantas”, en el cual se describen los principios básicos para la transmisión de caracteres hereditarios, principios ampliamente demostrados en muchas especies, sin embargo, el mismo Mendel no vislumbró la capacidad de generalización de sus trabajos. Su trabajo, una vez publicado y leído, cayó en el completo olvido hasta 1900 cuando Carl Correns, Erich Von Tschermak y Hugo de Vries, formulan los mismos principios planteados 35 años antes. La pregunta es: ¿por qué después de la publicación del origen de las especies en el cual era clara la necesidad de buscar dichos principios, el trabajo de Mendel no tuvo ninguna relevancia?

Mendel había optado por la carrera eclesiástica como vía para acceder a un mayor nivel de formación académica. Luego de su ordenación, fue enviado a la Universidad de Viena durante 1851 a 1853 donde, entre otros, fue discípulo del botánico Franz Andreas Nicolaus Unger y del físico Andreas Christian Doppler. Este último había sido un destacado científico quien introdujo a

Mendel en los rudimentos matemáticos de la experimentación. A su regreso a la abadía Agustina de Brunn, Mendel por su formación en ciencias naturales, fue encargado de resolver varios problemas de horticultura relacionados con la depuración de cepas para la producción de semillas. Luego de 10 años de rigurosa investigación, culminó su trabajo con la formulación de los principios de segregación y segregación independiente, sin embargo, sus experimentos se encontraban lejos del paradigma genético y evolutivo de ese momento. El problema de Mendel era un problema eminentemente práctico, en el que utilizó un detallado análisis matemático, llegando a la resolución y generalización de un problema biológico, incluso propuso un modelamiento teórico para el principio de segregación. Muy a pesar de la suficiencia de sus experimentos, estos carecen por completo de una perspectiva evolutiva, y por tanto fue imposible vincularlos con el origen de las especies.

Sea o no Mendel el padre de la genética, en cualquier caso sus experimentos han sido admirados y enseñados. Por un lado, no llega a comprender que a partir de sus trabajos en híbridos de plantas se deriva una disciplina nueva, diferente de la fisiología y que entraría a encargarse de la herencia y sus mecanismos como tema principal, por otro lado, sus procedimientos, definiciones de cruces, metodologías planteadas, una vez redescubiertas, influenciaron a varias generaciones de investigadores, incluido Thomas Hunt Morgan, posteriormente ganador del premio Nobel por sus descubrimientos que llevaron a la constitución de la teoría cromosómica de la herencia, en la que además, utilizó un modelo como *Drosophila melanogaster*, con el que verificó los principios mendelianos para el mundo animal.

Por el rigor de sus experimentos, por el alcance de los mismos y por su capacidad de plantear modelos matemáticos, Gregor Mendel debería ser también estudiado y reconocido como uno de los precursores del análisis matemático en biología.

Siglo XX

Después del origen de las especies toda una generación de biólogos empieza a adoptar el principio de selección natural dentro de la explicación de diferentes modelos biológicos, uno de los biólogos

más entusiastas de esta época es William Bateson, que se convierte en un darwinista acérrimo, para después encontrar una serie de dificultades dentro del principio de selección natural, llevando incluso al estudio de mecanismos alternativos (Sturtevant, 1965).

Hasta ese momento, la herencia era estudiada como una parte de la fisiología de los órganos reproductores. Los estudios de Bateson, junto con los descubrimientos de Mendel crearon el marco conceptual para estudiar la herencia como una disciplina aparte, con una serie de preguntas independientes. Una disciplina que poco a poco fue reafirmando el principio de selección natural, encargada de estudiar en cada ser vivo sus características hereditarias e interpretando la naturaleza de la población en función de unidades estructurales constantes a lo largo del tiempo, a las que ya en el siglo XX se llamó genes (Fox, 2002). Por tanto, la constitución de la disciplina llamada genética viene a ser una derivación del principio de selección natural. El origen de las especies generó la crisis conceptual y la necesidad de buscar los principios y mecanismos que dieran cuenta de la variabilidad en las poblaciones biológicas, dejando la puerta abierta para la búsqueda de la estructura del gen.

Una vez constituida la disciplina, se inicia un proceso de adaptación de los principios mendelianos a la selección natural terminando hacia 1930 con la formulación del neodarwinismo o la nueva síntesis que es la interpretación numérica y matemática del principio de selección natural (Fisher, 1958).

Más allá de los propios conceptos y teorías biológicas, desde el principio, las teorías genética y evolutiva, han tenido diversas interpretaciones llevando a la justificación de variadas posiciones políticas, científicas y religiosas. Por ejemplo, la sociobiología, que fundamentada en el neodarwinismo ha gozado de gran credibilidad científica, tratando de explicar el funcionamiento y comportamiento de los seres vivos desde un conjunto de reglas rígidas, en buena parte, sustentadas teóricamente por la construcción de modelos matemáticos, llegando a reducir lo viviente sólo a su base genética. En respuesta al reduccionismo genético y de los modelos químicos, pero igualmente nociva, se construyen interpretaciones religiosas, esotéricas y místicas de la evolución, derivando

en corrientes como el diseño inteligente que trata de dar un sustento científico al modelo creacionista clásico de la religión católica.

Indiscutiblemente nuestra comprensión del mundo se ha enriquecido notablemente con todo el desarrollo de la teoría biológica, pero muy por encima de todo nuestro desarrollo técnico y científico, seguimos enfrentados a las viejas preguntas planteadas desde los inicios de la humanidad, relacionadas con nuestro origen y con nuestras diferencias como especie.

Conclusión

En la segunda mitad del siglo XX, se ha dado un cambio radical con el advenimiento de la tecnología del ADN, cambiando la interpretación de la variabilidad poblacional y las relaciones evolutivas entre las especies. Igualmente, se ha dado un crecimiento exponencial en informática, la capacidad de almacenamiento de información, el diseño de nuevas estructuras en sistemas de información y una mayor velocidad de análisis. Los nuevos datos, y la nueva forma de manejarlos, han impulsado el desarrollo teórico en técnicas estadísticas, especialmente en el mundo del análisis multivariado y en el desarrollo de pruebas no paramétricas que dejan la puerta abierta a nuevos tipos de análisis de datos.

Una de las primeras implicaciones que tiene el desarrollo tecnológico y teórico sobre la biología, es un cambio sustancial en los procedimientos experimentales. Clásicamente los seres vivos fueron analizados desde disciplinas y procedimientos separados, por ejemplo, las técnicas de campo en ecología, los análisis de ADN en laboratorio, la taxonomía clásica. Esta individualidad está cambiando a un modelo experimental transversal para las disciplinas tradicionales del saber biológico. La unificación implica la reinterpretación de los fundamentos teóricos y determina nuevos alcances en las teorías evolutiva y genética. Quizás estamos entrando a una época en la cual veremos un nuevo principio de selección natural, más articulado con la biología del desarrollo, interpretado desde una genética más amplia, entendida no solo desde el gen, sino que considere la ecología y la fisiología. Quizás estamos frente a una nueva teoría biológica en cuyo centro no esté en el gen, sino el organismo en sí.

Bibliografía

BOORSTIN D. Los descubridores. 1.a edición. Barcelona: Editorial Grijalbo Mondadori;

DARWIN C. El origen de las especies. España: Editorial Los grandes pensadores; 1983.

FISHER R. The genetical theory of natural selection. 2a edición. New York; 1958.

FOX E. El siglo del gen cien años de pensamiento genético. 1a edición. Barcelona:
Ediciones Península; 2002.

JACOB F. La lógica de lo viviente. 1a edición. Barcelona: Salvat Editores; 1988.

LAMARCK J. Filosofía zoológica. Barcelona: Editorial Ata Fulla; 1986.

RUIZ R, AYALA F. De Darwin al DNA y el origen de la humanidad: la evolución y sus
Polémicas. 1a edición. México: Ediciones Científicas Universitarias; 2002.

STURTEVANT A. A History of genetics. 1a edition. London - Tokyo: A Harper international
edition; 1965.

Capítulo No. 1: Introducción General sobre Elementos de Muestreo.

William Usaquén Martínez, M.Sc.

Instituto de Genética, Universidad Nacional de Colombia. Bogotá D.C., Colombia.
wusaquenm@unal.edu.co

Genética de Poblaciones

La genética de poblaciones es una rama de la genética que se encarga de estudiar los cambios en las frecuencias alélicas de generación en generación y las fuerzas evolutivas responsables de dichos cambios en las especies. En la mayoría de los casos las poblaciones de estudio corresponden a poblaciones silvestres; sin embargo, los modelos genéticos se han extendido a poblaciones humanas (que no son consideradas poblaciones naturales), animales domésticos (como resultado de cruzamientos dirigidos), plantas (poblaciones agrícolas) y poblaciones de microorganismos (Hartl, 1997). Actualmente, los modelos poblacionales incluyen el hábitat de los organismos, considerando variables biológicas propias de regiones terrestres, acuáticas y aéreas. Tal como lo enuncia Gillespie (1998), el genetista de poblaciones intenta describir la estructura genética de una población o teoriza sobre las fuerzas evolutivas que actúan en las poblaciones; por lo tanto, la genética de poblaciones está fundamentada en la construcción de modelos. Si bien, existe un marco general de conocimientos; la biología propia de cada organismo, las condiciones cambiantes de los sistemas biológicos, las complejas relaciones de los organismos con su ambiente y los eventos del desarrollo, hacen que los modelos deban analizarse puntualmente. Por esta razón, la genética de poblaciones en la biología moderna es considerada una disciplina transversal y fundamental para el entendimiento en áreas como biología evolutiva, sistemática, ecología, historia natural y conservación genética. En esta última área, hoy en día es inevitable tomar decisiones, en cuanto al manejo y conservación de una especie sin la adecuada evaluación genético-poblacional. Fuera de las áreas biológicas, la genética de poblaciones interactúa con otras disciplinas entre las que se cuentan: medicina, derecho, biotecnología, sociología y antropología. (Hedrick, 2005).

Desarrollo histórico de la genética de poblaciones.

La genética de poblaciones parte de los principios mendelianos de segregación e independencia para hacer extrapolaciones a los efectos que dichos principios tendrían en una población. Pueden considerarse dos momentos decisivos para el inicio de ésta disciplina, el primero de ellos se da cuando Charles Darwin en 1859 plantea el principio de selección natural como un proceso que afecta a la totalidad de los organismos, iniciando la observación de la especie más allá del nivel orgánico individual. Esta fue la introducción de un nuevo nivel de organización dentro de la teoría biológica. Hacia 1865, el planteamiento de Gregor Mendel sobre las unidades de la herencia como unidades de carácter discreto, inició el desarrollo del análisis matemático de las mismas porque a partir de ese momento las “partículas” de la herencia conceptualmente se hicieron cuantificables. Con el nacimiento mismo de la genética (propuesta como una disciplina nueva en 1903), Godfrey Harold Hardy y Wilhelm Weinberg formularon el principio de equilibrio que se constituye en uno de los pilares para el desarrollo de la genética de poblaciones, debido a que se establece como un punto cero; permitiendo determinar la diferenciación en valores de frecuencias alélicas en una población y a partir de dicha diferencia, establecer los mecanismos de cambio evolutivo en el acervo genético de una población. En la década de 1920 a 1930 Ronald Fisher, J.B.S. Haldane y Sewall Right alcanzaron un entendimiento matemático completo del concepto de selección, planteado por Darwin que constituye el cuerpo formal de la síntesis evolutiva que estableció las principales fuerzas del cambio evolutivo: selección, migración, mutación y deriva genética (Fisher, 1958).

Hasta 1960 la genética de poblaciones tuvo básicamente un gran desarrollo teórico; sin embargo, el uso de aloenzimas y secuencias de aminoácidos a principios de los 70's indujo un profundo cambio, debido a la posibilidad de evaluar la composición de una población de manera directa a partir de polimorfismos genéticos (Lewontin, 1985). A mediados de los 70's se inició el estudio directamente sobre el ADN, lo que ha cambiado completamente la interpretación de la variación poblacional, indicando que ésta es más grande de lo que se había pensado; incluso, al interior de

las especies, llevando nuevamente al cuestionamiento de este concepto. El uso de datos moleculares ha generado un nuevo conjunto de preguntas, que para su consideración han requerido de la competencia de profesionales en áreas más allá de la biología; como ingeniería, estadística, matemáticas, antropología y demografía (Blinj, 1999).

Estructura Investigativa de la Genética de Poblaciones

La genética de poblaciones, en su estructura investigativa, tiene elementos empíricos, experimentales y teóricos. El enfoque empírico de la genética de poblaciones, comprende la observación de la variación genética a partir de caracteres morfológicos, análisis de polimorfismos moleculares y estudios cromosómicos. Actualmente, los modelos estadísticos basados en ADN, se centran en la variación de secuencias a nivel inter e intraespecífico. En cuanto al enfoque experimental, los trabajos clásicos se centran sobre fenómenos que se dan al someter poblaciones a nuevos ambientes, para luego ser comparadas con poblaciones que se han mantenido naturales, y posteriormente determinar el efecto del medio ambiente en la variación genética. El trabajo de laboratorio junto con los modelos computacionales provee soporte para las hipótesis desarrolladas a partir de datos empíricos. En años recientes, la definición de experimento genético se ha ampliado; hoy en día muchos de ellos se centran en la comparación de secuencias de ADN entre organismos que tienen diferentes historias evolutivas, variaciones en las funciones o diferentes procesos en su desarrollo. En cuanto al tiempo, se pueden emplear marcadores moleculares para evaluar una población en el momento actual o para hacer la reconstrucción de su historia a lo largo de un período de tiempo específico (Hartl, 1997).

Los modelos teóricos, proveen un marco general para evaluar el efecto de factores evolutivos en los niveles de variación y patrones de relación genética en diferentes ambientes. Estos modelos permiten además, predecir el cambio futuro de una población en diferentes escenarios. Sin embargo, estos análisis deben extrapolarse con cuidado debido a que las investigaciones teóricas se fundamentan en supuestos específicos y pueden tener solamente sentido en algunas especies o ser útiles solamente en algunos contextos de la realidad biológica (Hedrick, 2005).

La genética de poblaciones según sugieren algunos autores, es única dentro del desarrollo de las ciencias biológicas, debido a que antes de tener datos experimentales que fueran significantes para la construcción de una disciplina científica, fue elaborada como un área teórica. Esto sugiere que el temprano énfasis en la construcción teórica causó un distanciamiento entre la parte teórica, y el intercambio entre los aspectos empíricos y experimentales (Hedrick, 2005).

En la genética de poblaciones actual, los experimentos son de carácter estadístico y deben ser diseñados según las premisas básicas del diseño experimental: primero, la existencia de un concepto de réplica que sea de carácter independiente; segundo, deben desarrollarse los controles experimentales del caso; y tercero, obtener un adecuado tamaño de muestra para disminuir el riesgo de sesgos estadísticos. Este concepto está asociado al poder estadístico, que es la probabilidad de rechazar una hipótesis nula cuando ésta es falsa. Una de las características del experimento estadístico, es la inseguridad que genera en relación a si las muestras poseen el suficiente poder estadístico, de tal forma que sea posible la utilización de herramientas inferenciales; lo que llevaría a rigurosos procedimientos para determinar una muestra al momento de un estudio poblacional (Gale, 1980).

Modelos en genética de poblaciones

De los diferentes tipos de modelos a emplear en genética, los gráficos y las representaciones matemáticas son los más utilizados. Los modelos tienen la ventaja de describir el proceso y la organización de un fenómeno natural. Una de las características más deseables de un modelo es que tenga alta precisión respecto a la descripción de los fenómenos observados y que sea consistente con las observaciones; sin embargo, los supuestos hechos para propósitos matemáticos podrían algunas veces dar una visión imprecisa de la realidad biológica. Para mejorar la calidad de un modelo, se debe realizar una selección de supuestos que puedan ser similares o extrapolables al mundo natural; permitiendo estudiar fenómenos evolutivos entre las especies (Lewontin, 1985).

Los modelos matemáticos dentro de la genética de poblaciones pueden ser determinísticos, es decir, aquellos que dados unos valores particulares de los parámetros, arrojan una estimación puntual que puede llevarse a eventos futuros. Sin embargo, la mayoría de las veces los modelos son de naturaleza estocástica. En estos modelos, los parámetros se introducen asumiendo la existencia de valores de probabilidad para diferentes escenarios teóricos. Como resultado de este fenómeno, no siempre es posible, y de manera exacta predecir un evento futuro. En los modelos estocásticos se trabaja dentro de márgenes de probabilidad con la finalidad de acercarse al verdadero valor de un parámetro en una población. (Hedrick, 2005).

Dado que los modelos estocásticos están asociados a valores de probabilidad, algunas veces puede usarse una distribución de dicha naturaleza para determinar estos valores; sin embargo, si los modelos tienden a ser complejos, se prefiere realizar simulaciones basadas en números aleatorios llamadas simulaciones de Montecarlo que pueden utilizarse para estimación de parámetros; por ejemplo para la estimación de frecuencias alélicas (Guo & Thompson, 1992).

En muchos problemas de la genética de poblaciones, las distribuciones de probabilidad no son normales o no pueden ser asumidas como tal; cuando esto ocurre, es útil estimar intervalos de confianza alrededor de la media, antes de realizar un proceso de aleatorización en el que la probabilidad de los valores más extremos, puede ser calculada exactamente o estimada con una simulación computacional (Manly, 1991).

Algunos estudios relacionados en Colombia

Los primeros trabajos realizados en el país para la investigación biológica de la paternidad, se realizaron entre 1972 y 1995, a partir de grupos sanguíneos y fueron desarrollados por el Instituto Colombiano de Bienestar Familiar (ICBF). Hasta ese momento todas las pruebas antropoheredobiológicas se realizaban en Bogotá, lo que causó grandes represamientos de resultados. Es por esto, que en 1995 el ICBF inició un proceso de descentralización para las pruebas de paternidad, dirigiéndose a la contratación de laboratorios en las universidades públicas

y posteriormente a laboratorios de carácter privado. Esto condujo a un importante desarrollo tecnológico y técnico-científico; se pasó de las pruebas con grupos sanguíneos y polimorfismos de HLA, a la utilización de sistemas con un gran poder de discriminación como son los microsatélites. Estos avances crearon el ambiente apropiado para la discusión y posterior legislación sobre pruebas de filiación e identificación en Colombia (Bravo, 2001; 2003).

En la actualidad, se sigue utilizando esta tecnología con un mayor número de sistemas microsatélites, hasta llegar a utilizar los empleados por el CODIS, desarrollado por el FBI y validado a nivel internacional. Para estos marcadores moleculares, a nivel regional cada laboratorio ha realizado publicaciones particulares de frecuencias alélicas que se utilizan en la práctica cotidiana (Acosta et al., 2002; Benitez et al., 2003; Bustos et al., 2001; Groot et al., 1992; Duran et al., 2000; Rey et al., 2003; Vargas et al., 2002; Paredes et al., 2003; Usaquen, 2006; Gómez et al., 2003).

En el año 2003, ante el gran volumen de trabajo, el ICBF contrató la tipificación de aproximadamente 40.000 personas con laboratorios de todo el país, constituyéndose en una de las tomas de datos más grandes a nivel mundial. Una compilación y análisis de esta información, permitiría ubicar a Colombia a la vanguardia en el mundo de los análisis genético poblacionales; lo cual sería un estímulo para el desarrollo de nuevos métodos de análisis, tarea que a la fecha no ha sido realizada.

Definición de Sistemas Microsatélites

Los microsatélites también conocidos como STR's (*short tandem repeats*) son repeticiones en tándem de unidad pequeña de dos a seis pares de nucleótidos que corresponden aproximadamente al 20% del genoma humano, y son sistemas de herencia codominante que han mostrado ser altamente polimórficos. Hasta el momento no hay claridad sobre su significado funcional y por tanto, se encuentran como sistemas sin función; sin embargo, algunos autores piensan que contienen exones y se hallan asociados con la expresión de algunas enfermedades (Goldstein &

Schlotterer, 1999).

Los microsatélites están distribuidos en todo el genoma eucarionte. Los motivos de repetición a lo largo de los diferentes genomas suelen variar sustancialmente, siendo la repetición AT la de mayor frecuencia (Goldstein & Schlotterer, 1999).

Un aspecto de gran relevancia, para la utilización de microsatélites como marcadores en procesos de identificación, es que no están sometidos a procesos de selección debido a que no presentan una secuencia codificante. Por otra parte, las regiones flanqueantes de los microsatélites son secuencias altamente conservadas, lo que permite su utilización para realizar inferencia filogenética de diferentes especies (Avise, 1996).

Clasificación

La clasificación de los microsatélites se hace de acuerdo con el patrón de ordenamiento de los motivos de repetición (Chambers & MacAvoy, 2000):

- Simple: Un solo motivo repetido n veces en serie. Ej. (AC)⁹
- Simple interrumpido: Un solo motivo repetido n veces, donde se intercalan nucleótidos entre las distintas repeticiones. Ej. (CA)²AA(CA)¹²
- Compuestos: Dos o más motivos repetidos en serie. Ej. (GT)²(TG)¹⁰
- Compuestos interrumpidos: Al menos uno de sus motivos presenta nucleótidos intercalados. Ej. (CT)⁴(GT)²CTAT(GT)¹⁵
- Complejos: Combinaciones entre cualquiera de las clases anteriores, sin ningún patrón de orden definido. Ej. (ACC)⁸TG(GA)¹²(TTA)⁵GC(TTA)⁴.

Teoría de Muestreo

En la investigación genética, es muy frecuente la utilización de estudios por muestreo; sin

embargo, la teoría de muestreo suele emplearse mal en estas áreas, en parte por desconocimiento del investigador; pero también desde un punto de vista biológico, por la imposibilidad técnica de construir los marcos de referencia necesarios para definir un muestreo probabilístico; razón por la cual, la mayoría de investigaciones se concentran sobre muestreos no probabilísticos, por ejemplo, por norma o por conveniencia.

Conceptos iniciales de teoría de muestreo en genética.

Universo: Es el conjunto total de elementos bajo estudio o sobre los cuales se desea efectuar un proceso de inferencia, que permite la proyección de resultados obtenidos de la muestra a toda la población donde se realizó el muestreo. De ahí, la importancia que la muestra represente lo mejor posible a la población objetivo (Ospina, 2001).

Población estadística: Es el conjunto de mediciones que se hacen sobre los elementos muestreados (Ospina, 2001). Un solo conjunto de muestras biológicas podría dar lugar a diferentes poblaciones estadísticas, lo que se denomina población multivariada. En el caso genético, se tiene que cada posible persona de la población se examina a diferentes niveles y para múltiples genes. Cuando se necesita establecer distribuciones de frecuencias genotípicas, cada elemento del universo es una persona; sin embargo, al estudiar frecuencias alélicas se cambia el nivel de observación y se pasa a estudiar alelos individuales. Por otra parte si se desea estudiar marcadores de cromosoma Y para estudios de ancestría, entonces el nivel de observación se transfiere a la observación haplotípica (Ospina, 2001).

Métodos de muestreo: Se refieren al conjunto de técnicas estadísticas que estudian la forma de seleccionar una muestra representativa de una población y cuya información permita inferir las propiedades o características de toda la población asumiendo un error medible y controlable. A pesar de los métodos propuestos por la teoría de muestreo, se han desarrollado alternativas diferentes no siempre basadas en los supuestos formales, lo que puede limitar el proceso de inferencia pero no invalida necesariamente las investigaciones biológicas. Los métodos de

Captura-recaptura y las técnicas de remoción ampliamente utilizados dentro de estudios ecológicos, son ejemplos de métodos que no pueden ser clasificados estrictamente como métodos de muestreo; sin embargo, si se tiene el suficiente cuidado en el diseño experimental y en los procedimientos analíticos, permitirán realizar inferencias acertadas y cálculos de los niveles de confianza para la población (Ospina, 2001).

Estimadores: un estimador es un estadístico (determinado a partir de una muestra), usado para acercarse a un parámetro desconocido de la población. Los estimadores, son funciones matemáticas, que se comportan como variables aleatorias. En general, un buen estimador es aquel que es insesgado, eficiente, convergente, es decir que en los diferentes modelamientos genera un valor dentro de un intervalo de confianza y además robusto. El valor de un estimador, proporciona una estimación puntual del valor del parámetro en estudio. En general, se prefieren las estimaciones mediante intervalos, es decir un intervalo $[a, b]$ dentro del cual se espera encontrar el valor real del parámetro, con un determinado nivel de confianza (López, 2005).

La Investigación por muestreo.

Las investigaciones por muestreo son ampliamente utilizadas debido a la necesidad de obtener conclusiones de una población a partir de un grupo de individuos. Para la realización de una investigación por muestreo, autores como Bautista (1998) proponen los siguientes pasos:

1. Diseño muestral
2. Mediciones a tomar
3. Trabajo de campo
4. Análisis estadístico
5. Diseño Muestral

Esta etapa incluye desarrollar un plan de muestreo y los procedimientos de estimación. Por plan de muestreo, se entiende la metodología utilizada para seleccionar los individuos de la muestra en una

población determinada. Los procedimientos de estimación son los algoritmos o fórmulas usadas para obtener estimaciones de valores poblacionales y su confiabilidad a partir de datos muestrales. La selección de un diseño muestral se basa en tres aspectos básicos: el tipo de variables, las estimaciones requeridas, y los niveles de confiabilidad y validez de acuerdo con los recursos existentes (Bautista, 1998).

La manera como son tomados los datos depende del tipo de muestreo, es decir, si este es probabilístico o no. Una preparación pobre de la información o una deficiente capacitación frecuentemente conduce a fracasos en el estudio. Cuando las mediciones se refieren a unidades diferentes de personas, como animales, plantas o información de laboratorio, el problema es más sencillo de manejar ya que la información no está sujeta a opiniones o situaciones emocionales, como ocurre en el caso de las personas; sin embargo, en estos casos se deben implementar los debidos controles experimentales como una medida de aseguramiento de la reproducibilidad de la información (Bautista, 1998).

Una vez realizado el diseño de los instrumentos, es conveniente en muchos casos tomar una muestra piloto, que es muy reducida y permite probar los instrumentos de medición y eliminar los defectos del proceso a desarrollar. Esto permite además evaluar la presencia de errores tanto de muestreo como de medición.

Muestreo probabilístico

Es aquel en el cual se puede establecer la probabilidad de obtener cada una de las muestras seleccionadas a partir de un marco referencial utilizando un procedimiento aleatorio. Las principales características de lo que se puede definir como un muestreo aleatorio son:

1. Poder definir el conjunto total de muestras posibles que pueden seleccionarse de la población de acuerdo con el procedimiento muestral. Este conjunto determina el marco referencial.

2. Conocer para cada una de las muestras posibles, la probabilidad de ser seleccionada.
3. El procedimiento utilizado debe dar a cada elemento de la población una probabilidad de selección diferente de 0.
4. La selección debe ser aleatoria, esto es, el mecanismo de probabilidad diseñado para la selección debe ser tal que cada muestra tenga la misma probabilidad de selección en cualquier parte del espacio geográfico (Ospina, 2001).

Muestreo no probabilístico.

En general, todo tipo de muestreo que no cumpla con alguna de las condiciones enumeradas anteriormente es un muestreo no probabilístico. Los siguientes son ejemplos de muestreo no probabilísticos frecuentemente empleados.

1. La muestra ha sido restringida a la parte de la población que es fácilmente accesible, como es el caso más común en los análisis genético-poblacionales en humanos, debido a que estos se realizan con los usuarios de los servicios de filiación y genotipificación.
2. La muestra se selecciona teniendo en cuenta el azar más no un procedimiento de aleatorización.
3. Con una población heterogénea y pequeña, el investigador inspecciona la población en general y selecciona una muestra pequeña de unidades tipo.

En condiciones apropiadas, cualquiera de los anteriores métodos puede proporcionar resultados útiles. Sin embargo, en ellos no se puede aplicar estrictamente la teoría de muestreo ya que ésta se basa en el supuesto de selección aleatoria, que es el principal problema del muestreo no probabilístico, de aquí que haya una precisión indeterminada. En este tipo de muestras la precisión se sustenta en la experiencia de quien selecciona la muestra. En el caso de las muestras probabilísticas no se presenta este problema, pues el proceso de selección ayuda a lograr un grado

de precisión (Ospina, 2001).

Plan de Muestreo Estadístico

Este se establece en función de los objetivos, los recursos con los que cuenta el laboratorio y la calidad deseada de los estimadores poblacionales, estos aspectos determinan las decisiones de tipo operativo que paso a paso se van tomando. El plan debe estar acompañado de una estrategia de crítica e imputación a fin de evitar errores que generen sesgos y que pueden llegar a invalidar los resultados globales del estudio (Bautista, 1998).

Pasos para la construcción del plan de muestreo

Los pasos que se siguen para la elaboración del plan de muestreo, son similares a los que deben seguirse para la elaboración de cualquier proyecto de investigación. Se debe iniciar por la definición precisa de todas las variables y el establecimiento de un plan de análisis asociado con los objetivos de la investigación para garantizar que la información necesaria se recolecte y para evitar que mucha información que se recogió con un gran costo quede sin uso. Según Bautista (1998) los siguientes pasos deben seguirse para el diseño de un plan de muestreo:

1. Definición concreta y clara de los objetivos.
2. Traducción de los objetivos en resultados estadísticos.
3. Especificación de las variables y parámetros de la población objetivo.
4. Construcción del marco de muestreo.
5. Inventario de recursos disponibles.
6. Establecimiento de cronogramas.
7. Determinación de la calidad de las estimaciones realizadas.
8. Especificación de los métodos de recolección de información.
9. Especificación del diseño muestral.
10. Determinación de procesamiento de crítica e imputación.
11. Especificación de las fórmulas de estimación y de las medidas de calidad.
12. Planeación del trabajo de campo.

13. Diseño del plan de control y evaluación.

Fuentes de sesgos y varianza

Estos son los dos defectos principales que deben ser evitados a cada paso y en cada decisión del proceso de planeación de un estudio. El sesgo reduce el nivel de confiabilidad del intervalo de confianza, la varianza amplía la longitud del intervalo. Los errores que se pueden cometer y los procesos que deben ser revisados a fin de evitar y reducir la producción de sesgos son (Bautista, 1998):

1. La selección: esta puede generarse debido a la calidad del marco generando problemas de cobertura.
2. Inadecuada recolección debida a errores: en los procesos de entrevistado, falta de motivación, deficiencias en el cuestionario, inadecuado proceso de recolección.
3. Procesamiento: errores en los procesos de codificación, crítica e imputación y finalmente errores en el cálculo de estimativos.
4. Verificación: Validación y consistencia de la información.

Una vez definido el plan completo de crítica se deben realizar las tres operaciones siguientes.

1. Verificación: es el proceso que generalmente acompaña a la grabación de la información y consiste en garantizar que la información almacenada es un fiel reflejo de lo contenido en las encuestas. Generalmente este proceso se realiza como un segundo paso en el proceso de registro de la información.
2. Validación: es el proceso por el cual se determina si los datos cumplen ciertas reglas preestablecidas de aceptabilidad. Estas reglas constituyen la crítica estadística.

3. Consistencia: Se parte del principio de realizar el mínimo cambio posible de datos para obtener una base de datos completa y corregida. Esto requiere del desarrollo de reglas de crítica para determinar los procesos de cambio de información (Bautista, 1998).

Análisis Multivariado.

Las técnicas de análisis multivariado son ampliamente conocidas y utilizadas en la actualidad en los estudios genético-poblacionales debido a la capacidad que tienen para analizar grandes conjuntos de variables, justamente lo que sucede cuando se trabaja con sistemas genéticos polimórficos. En una matriz de datos construida a partir de las distribuciones de frecuencias alélicas de sistemas microsatélites, cada uno de los alelos del sistema se asume como una de las variables en estudio (columnas) y cada uno de los individuos (filas) corresponde a las poblaciones que se analizan. A partir de este modelo se realiza clasificación y comparaciones entre poblaciones, que es uno de los objetivos principales en la caracterización genética de poblaciones. Las técnicas principalmente utilizadas son componentes principales, correspondencias múltiples y escalamientos multidimensionales (Cavalli-Sforza y Bodmer 1999). Pese a que estas técnicas fueron desarrolladas desde principios del siglo XX, su implementación en la genética se inicia hasta los años 60 debido a dos situaciones favorables: Se alcanzó el suficiente conocimiento teórico sobre la naturaleza de los polimorfismos genéticos y el desarrollo de las técnicas de laboratorio adecuadas para obtener datos confiables; en segundo lugar, ya se contaba con las herramientas computacionales que permitían la utilización de algoritmos multivariados, hasta ese momento tenidos como procedimientos bastante complejos y tediosos de desarrollar (Dutta, 2002).

Otra técnica muy utilizada es el análisis de conglomerados; a partir de este se establecen las relaciones de agrupamiento entre poblaciones. Esta técnica es diferente del análisis cladístico y filogenético, en el cual se parte de un grupo ancestral para luego ir reconstruyendo la historia evolutiva de un conjunto de taxones. En este caso, los dendrogramas son representaciones que

carecen de una medida directa de tiempo, por lo tanto son reconstrucciones actuales de poblaciones; sin embargo, como lo sugiere Nei (1987) pueden ser reconstrucciones filogenéticas aceptables, cuando la distancia seleccionada genera unos valores acordes con el pasado evolutivo de los organismos. En el caso de la especie humana se puede hablar también de reconstrucciones filogenéticas cuando los dendrogramas construidos reflejan procesos migratorios, las fuerzas de selección y los procesos de deriva por los que ha pasado una población. En el caso más local, los dendrogramas construidos hasta el momento (Usaquen, 2006) muestran cómo los procesos de poblamiento en Colombia tienen una alta asociación con los procesos históricos regionales.

La base para los análisis de conglomerados son las matrices de distancias; desde el punto de vista estadístico, una distancia es una métrica que cumple con las siguientes condiciones (Díaz, 2002):

1. No negatividad
2. Simetría
3. Desigualdad triangular
4. Identificación de no identidad
5. Identidad

En genética se utilizan distancias como Czekanowski, Manhattan, Rogers que cumplen las condiciones anteriores. Otras distancias muy robustas, pero que no cumplen todos los supuestos anteriores, son por ejemplo la distancia de Mahalanobis utilizada cuando los caracteres a estudiar están correlacionados. Otra distancia desarrollada específicamente para datos genéticos es la distancia de Cavalli-Sforza Edwards. Ésta representa las poblaciones como vectores sobre la superficie de una esfera multidimensional, donde la distancia está dada por los ángulos formados por dichos vectores (Nei, 1987).

Para la obtención de una clasificación juegan un papel importante la distancia seleccionada y la metodología de clasificación propuesta. Una combinación especialmente eficiente es la distancia de Cavalli-Sforza Edwards junto con el método Ward. Éste se ha ensayado para las diferentes

poblaciones obteniéndose clasificaciones muy acordes con procesos de poblamiento histórico (Usaquén, 2006).

Para una buena interpretación multivariada de los datos se sugiere la combinación de diferentes técnicas, buscando encontrar consistencia entre los diferentes ordenamientos, especialmente los ordenamientos producidos por planos factoriales y los métodos de construcción de dendrogramas (Cavalli – Sforza et al., 1994).

La genética de poblaciones en la actualidad combina el desarrollo teórico junto con el desarrollo experimental. Ha pasado a ser una disciplina práctica cuyos resultados son el fundamento para la sistemática molecular, la identificación forense tanto en humanos como en otras especies y la toma de decisiones en la conservación genética (siendo considerada en esta última como una herramienta esencial). Sin embargo, las conclusiones deben estar basadas en análisis estadísticos minuciosos, empezando por el análisis en la consistencia de la muestra analizada, tal que el investigador tenga mejores parámetros para llegar a conclusiones acertadas. En este momento, es necesario el desarrollo de herramientas teóricas y computacionales acorde con los nuevos marcadores moleculares que se están implementando. Esta es la línea de pensamiento planteada en el presente trabajo.

Bibliografía

- ACOSTA, M; BRION, M; LAREU, M; CARRACEDO, A; 2002. Genetic data on eight STRs (D5S818, D7S820, F13B, LPL, TH01, TPOX, VWA31, CSF1PO) from a Colombian population. *Forensic Science International*. V 129. pp 216-218.
- AVISE, J. 1996. *Conservation genetics. Case Histories from Nature*. New York, Kluwer Academic Press.
- BAUTISTA, L. 1998. *Diseño de Muestreos Estadísticos*. Bogotá, Universidad Nacional de Colombia.

- BENÍTEZ, A. & OSSA, H. 2003. Allelic frequencies at 12 STR loci in Colombian population. *Forensic Science International*. V 136. pp 86-88.
- BLINJ, H. & MURPHY, A. 1999. *Human Geography. Culture Society and Space*. Londres: Wiley and Sons.
- BRAVO, M. 2001. *Introducción a la genética forense de las pruebas de paternidad*. Medellín, Universidad de Antioquia.
- BRAVO, M. 2003. *Legislación sobre la investigación científica de la paternidad biológica: ley 721 de 2001*. Medellín, Universidad de Antioquia.
- BUSTOS, I; ACOSTA, M; BRAVO, M; BUILES, J; CARABALLO, L. 2001. Colombia: País genéticamente megadiverso en poblaciones humanas caracterizadas por sistemas STR's. VI Jornadas de Genética forense. GEP-ISFG. Córdoba, Argentina.
- CAVALLI-SFORZA, L.L; MENOZZI, P; PIAZZA, A. 1994. *The history and Geography of Human Genes*. New York. Princeton University Press.
- CAVALLI-SFORZA, L.L; BODMER, W. F. 1999. *The genetics of Human Populations*. New York: Dover Publicatons.
- CHAMBERS, G & MACAVOY, E. 2000 *Microsatellites: consensus and controversy*. *Comparative Biochemistry and Physiology Part B* 126 (2000) 455–476.
- DIAZ, L. G. 2002. *Estadística multivariada: inferencia y métodos*. Bogotá: universidad nacional de Colombia.
- DURAN, R. 2000. *Análisis de las frecuencias alélicas de los loci D1S80, vWA y TH01 en un grupo de individuos de Santa fe de Bogota. Para aplicación en Investigación Criminal*. Tesis Universidad Nacional de Colombia.
- DUTTA, R. 2002. *Patterns of Genetics Diversity at the Nine Forensically Approved STR Loci in the Indian Population*. *Hum Biol*. 2002 Feb;74(1):33-49.
- FISHER, R. 1958. *The Genetical Theory of Natural Selection*. Londres: Oxford University Press.
- GALE, J. S. 1980. *Population genetics*. London: Blackie & son limited.
- GILLESPIE, J. H. 1998. *Population Genetics. A concise guide*. The Johns Hopkins University Press. Baltimore.

- GOLDSTEIN D & SCHLOTTERER C. 1999. *Microsatellites*. Londres: Oxford University Press.
- GÓMEZ, M. V; REYES, M. E; CÁRDENAS, H & GARCÍA, O. 2003. Genetic variation for 12 STRs loci in a Colombian population (Department of Valle del Cauca). *Forensic Science International*. V 137.
- GROOT, H. 1992. Estructura genética de la comisaría de Amazonas. Universidad de Los Andes. 104
- GUO, S. A & THOMPSON, E. A. 1992. Performing the exact test of Hardy Weinberg proportion of multiple alleles. *Biometrics* 48 361-372
- HARTL, D. 1997. *Principles of population genetics*. Sunderland, Massachusetts. Sinauer Associates Inc.
- HEDRICK, P. W. 2005. *Genetics of populations*. Tercera Edición. Jones and Bartlett Publishers. Massachusetts.
- LEWONTIN, R. 1985. Population genetics. *Annual Review of Genetics*. Vol 19.
- LOPEZ, C. 2005. *Muestreo Estadístico, Conceptos y problemas resueltos*. Prentice may. Madrid.
- MANLY, B. 1991. *Randomization Monte Carlo Methods in Biology*. Chapman Londres.
- NEI, M. 1987. *Molecular Evolutionary Genetics*. Columbia University Press. New York
- OSPINA D. 2001. *Introducción Al Muestreo*. Facultad de Ciencias UN.
- PAREDES et al. 2003. Analysis of the CODIS autosomal STR loci in four main Colombian regions. *Forensic Science International*. 137; 67-73.
- REY, M; GUTIERREZ, A; SCHROEDER, B; USAQUEN, W; CARRACEDO, A; BUSTOS, I; GIRALDO, A. 2003. "Allele frequencies for 13 STR's from two Colombian populations: Bogotá and Boyacá". *Forensic Sci Int*. 136:83-85.
- USAQUÉN, W. 2006. *Caracterización genética de 23 poblaciones humanas colombianas a partir de 9 sistemas microsatélites*. Tesis: Universidad Nacional de Colombia.
- VARGAS, CI; CASTILLO, A; GIL, A.M; PICO, A.L; GARCÍA, O. 2002. "Allele frequencies for the AmpFISTR profiler loci in a Colombian population (Department of Boyacá)". *J Forensic Sci*. 47:406-407.

Capítulo No. 2: Una nueva aproximación al cálculo de tamaño de muestra en estudios genético poblacionales y forenses a partir de marcadores microsatélites.

W. Usaquén¹, J. Corzo², L.F. García³, M. Rojas¹, A. Alonso¹, A. Casas¹.

1. Instituto de Genética Universidad Nacional de Colombia, 2. Departamento de Estadística Universidad Nacional de Colombia. 3. Departamento de Biología. Universidad Nacional de Colombia.

Resumen

Una pregunta frecuente en los estudios poblacionales es si el tamaño de muestra utilizado es suficiente para obtener estimadores genéticos confiables, como por ejemplo el cálculo de frecuencias alélicas. El primer problema en la resolución de esta incógnita es el procedimiento para seleccionar las muestras. Los estudios genéticos generalmente utilizan técnicas de muestreos por norma o por conveniencia. Sin embargo por esta vía no es posible la estimación del error muestral, por tanto se hace difícil evaluar si la variación de los estimadores poblacionales es debida a variaciones evolutivas o a muestreos deficientes.

En este trabajo se presenta un procedimiento para determinar tamaños de muestra y evaluar la calidad de un estimador genético en función de su dispersión. Se analizaron frecuencias alélicas de sistemas microsatélites. El procedimiento planteado se fundamenta en el aumento secuencial e iterativo del tamaño de muestra, obteniendo estimadores por intervalos de confianza asociados a la frecuencia de cada variante alélica.

Los resultados indican un tamaño de muestra aproximado de 700 personas para una población con las características de Bogotá, es decir de múltiples orígenes étnicos en los cuales el nivel de polimorfismo se espera sea alto. Para la evaluación de este modelo se utilizaron sistemas genéticos microsatélites (STR's: *Short Tandem Repeats*) utilizados en el *combined DNA index System* CODIS; sin embargo, el uso de esta metodología podría extenderse a cualquier sistema del que se pueda determinar la composición genotípica, como por ejemplo los sistemas SNP's. Los límites de los intervalos de confianza también podrían emplearse en el cálculo de índices de paternidad o en estudios de subestructura genética.

Keywords.

STRs, Paternity, Validation, Sample Size.

1. Introducción.

Al momento de seleccionar un grupo de personas para un estudio genético poblacional, el cálculo de tamaño de muestra suele obviarse, en buena parte por la dificultad de aplicar los supuestos y la metodología de la teoría formal de muestreo. La alternativa más utilizada es seguir las recomendaciones hechas por asociaciones internacionales tales como el “*committee on DNA technology in forensic science*” y la “*International Society for Forensic Genetics*” (ISFG), organizaciones que desde el año 1992 sugerían tamaños de muestra aproximados de 200 individuos (ISFH, 1992). Estos valores se han incrementado, en parte por la acumulación de información en las diferentes bases de datos; pero también por críticas teóricas al reducido tamaño de muestra (Evet y Weir 1998), especialmente sobre los tests para independencia de frecuencias alélicas los cuales podrían tener bajo poder estadístico.

En el informe del año 2002 de la ISFG, también en el 2004, se sugieren tamaños de muestra superiores a 500 personas, sin que se presente una justificación estadística para esta cifra. A pesar de esto, se sigue careciendo de procedimientos para evaluar la exactitud de la información, entendida esta como el cálculo de la variación en los estimadores poblacionales. Una alternativa puede ser el cálculo de intervalos de confianza que permitan un mejor acercamiento a los verdaderos parámetros poblacionales.

Por tanto, los muestreos que se utilizan en los estudios poblacionales corresponden a muestreos no probabilísticos provenientes de servicios de identificación genética. El principal argumento a favor de este tipo de muestreos surge de la neutralidad evolutiva asumida para los sistemas microsatélites, es decir, que al no tener ninguna expresión fenotípica no estarían sometidos a procesos de selección natural directa, y por lo tanto su distribución en la población dependería directamente de su frecuencia poblacional, lo que indicaría independencia de cualquier tipo de

sesgo. Este argumento, ignora el efecto simultáneo de todas las fuerzas de cambio evolutivo causantes de un equilibrio dinámico, además de los fenómenos de endogamia y subestructura que deberían considerarse al momento de realizar estimaciones de frecuencias poblacionales. Por otra parte, algunos autores consideran que los sistemas microsátélites pueden no ser tan neutrales como se ha pensado (Goldstein y Schlotterer, 1999), por tanto la argumentación de selección de muestras no aleatorias, defendida por la neutralidad de los microsátélites, perdería fundamento.

En cuanto a la selección de individuos, un gran número de reportes de frecuencias en el mundo se han realizado con muestras provenientes de bases de datos ya establecidas en laboratorios de análisis genético como aquellos que realizan pruebas y análisis de filiación e identificación, o instituciones gubernamentales relacionadas con criminalística y casos de desaparición forzada. Por la naturaleza de estos datos, la única información demográfica que suele recuperarse es el lugar de nacimiento de las personas relacionadas. En estas bases de datos se procura no incluir individuos emparentados; sin embargo no se realiza ninguna evaluación sobre el efecto de poblaciones cerradas y de tamaño reducido, incrementando la posibilidad de seleccionar individuos relacionados por parentesco dos o tres generaciones atrás.

El problema de muestreo en genética ha sido abordado por algunos autores entre los que se destacan, Kalinowski, (2005); Campbell, (2000); Chakravarty, (2007); Gillespie, (2001); Buttler, (2008); Hey,(2004). Uno de los trabajos más sofisticados es el publicado por Chakraborty, (1992), quien propone determinar el tamaño de muestra a partir de los alelos más frecuentes debido a que estos son los más representativos en la población, en tanto se ignorarían los alelos de baja frecuencia. Este argumento justifica la clasificación en alelos efectivos y no efectivos, tal como lo realiza Buttler, (2008). Otro trabajo importante es el publicado por Harding (1998) en el que se propuso una metodología para la selección de muestras de bases de datos poblacionales; sin embargo, el autor sólo plantea una alternativa gráfica sin presentar ningún desarrollo estadístico que permita determinar un número mínimo de personas para un estudio poblacional. De más reciente aparición se encuentra el trabajo presentado por Medina (2006) basado en técnicas de remuestreo, con la finalidad de obtener un tamaño de muestra idóneo para evitar sesgos al

momento de realizar análisis multivariados de clasificación y agrupamiento de poblaciones.

Dentro de este contexto, este trabajo presenta una nueva aproximación al cálculo de tamaño de muestra en estudios genético poblacionales y forenses a partir de marcadores microsátélites. Un método para determinar intervalos de confianza asociados a la frecuencia de cada variante alélica de un determinado polimorfismo, partiendo de muestreos no probabilísticos. A partir de dichos intervalos de confianza se puede encontrar el margen de variación para estimadores poblacionales y forenses como por ejemplo los índices de paternidad. Los intervalos de confianza pueden ser también utilizados para determinar si dos muestras corresponden a la misma población o no. Hall (1999), Hohls (1998), Wang (2003). La generación de intervalos de confianza diferentes en dos muestreos diferentes sería un claro indicio de procesos de subestructura poblacional.

2. Método propuesto.

El método ha sido organizado en dos fases desarrolladas en forma de algoritmos descritos en los apartes 2.1 y 2.2.

2. 1. Fase 1: Cálculo secuencial de las frecuencias alélicas.

1. Inicialmente se selecciona una muestra m_1 de n personas. Sobre este tamaño se realiza la estimación de frecuencias correspondiente a la distribución de frecuencias que llamaremos d_1 de la muestra m_1 .
2. En el siguiente paso se selecciona una muestra m_2 de la misma población y del mismo tamaño que m_1 , con lo cual se obtiene una muestra de tamaño m_2 .
3. El número de personas n inicial, se mantiene constante a lo largo de los muestreos, cada nueva muestra se suma a la inmediatamente anterior, y de esta manera se continua hasta obtener una sucesión estable en valores de frecuencias alélicas. Con el anterior algoritmo, se construyó la matriz No. 1 de frecuencias para un sistema de k alelos con m_n muestreos.

Sample size.	Allele frequencies				
M_1	f_{11}	f_{12}	f_{13}	...	f_{1k}
M_2	f_{21}	f_{22}	f_{23}	...	f_{2k}
M_3	f_{31}	f_{32}	f_{33}	...	f_{3k}
M_t	f_{t1}	f_{t2}	f_{t3}	...	f_{tk}

Matríz No.1 : Sucesión de distribuciones de frecuencias alélicas a lo largo de los incrementos en el tamaño de la muestra. Con cada nuevo muestreo se obtiene una nueva distribución de frecuencias cuyo incremento es en n individuos.

2.2. Fase 2: Cálculo de coeficientes de variación.

En esta fase se calculan promedios por cada variante alélica desde f_{11} hasta f_{it} , igualmente la matriz de desviaciones estándar también desde f_{11} hasta f_{it} . Posteriormente se calculan los coeficientes de variación. A medida que se da el incremento en el tamaño de muestra se obtienen coeficientes de variación menores lo que permite determinar un punto de corte asociado a un valor específico de tamaño de muestra. Según Bautista, (1998) un valor de coeficiente de variación mínimo en un estudio muestral, es cercano a 0.1. Por tanto cuando se alcanza el valor antes mencionado, se obtiene un tamaño de muestra específico del microsatélite analizado, el cual se considera un tamaño de muestra óptimo, dado que la variación en el valor de frecuencia establecido alcance un valor aceptable en su estimación.

2.2.1. Promedios ponderados de frecuencias.

La matriz de promedio de frecuencias desde f_{11} hasta f_{tk} , es decir hasta el último término de la última variante alélica a lo largo de las diferentes distribuciones de frecuencias inicialmente se calcula desde el promedio entre la 1ª y la 2ª muestra, luego la sucesión empieza en la segunda distribución:

<i>Alelo 1</i>	<i>Alelo 2</i>	...	<i>Alelo k</i>
$\bar{x}_{21} = \frac{f_{11} + f_{21}}{m_1 + m_2}$	$\bar{x}_{22} = \frac{f_{12} + f_{22}}{m_1 + m_2}$...	$\bar{x}_{2k} = \frac{f_{1k} + f_{2k}}{m_1 + m_2}$
$\bar{x}_{31} = \frac{f_{11} + f_{21} + f_{31}}{m_1 + m_2 + m_3}$	$\bar{x}_{32} = \frac{f_{12} + f_{22} + f_{32}}{m_1 + m_2 + m_3}$...	$\bar{x}_{3k} = \frac{f_{1k} + f_{2k} + f_{3k}}{m_1 + m_2 + m_3}$
.	.		.
.	.		.
.	.		.
$\bar{x}_{i1} = \frac{f_{11} + f_{21} + \dots + f_{i1}}{m_1 + m_2 + \dots + m_i}$	$\bar{x}_{i2} = \frac{f_{12} + f_{22} + \dots + f_{i2}}{m_1 + m_2 + \dots + m_i}$...	$\bar{x}_{ik} = \frac{f_{1k} + f_{2k} + \dots + f_{ik}}{m_1 + m_2 + \dots + m_i}$
$\bar{x}_{i1} = \frac{\sum_{t=1}^i f_{t1}}{\sum_{t=1}^i m_t}$	$\bar{x}_{i2} = \frac{\sum_{t=1}^i f_{t2}}{\sum_{t=1}^i m_t}$...	$\bar{x}_{ik} = \frac{\sum_{t=1}^i f_{tk}}{\sum_{t=1}^i m_t}$

Matríz No.2 : Promedios de frecuencias calculados por cada variante .

La forma general empleada para calcular los promedios de frecuencias alélicas sería:

Ecuación No. 1:
$$\bar{x}_{.k} = \frac{1}{t} \sum_{i=1}^t f_{ik} \cdot w_i$$

En la expresión anterior aparece un termino w_i , que corresponde a la ponderación a medida que se da un incremento de los muestreos desarrollados, por tanto w_i corresponde a un promedio ponderado. A continuación se presentan los primeros desarrollos hasta llegar a la forma general calculada de la siguiente forma:

$$w_1 = \frac{1}{2n} + \frac{1}{3n} + \frac{1}{4n}$$

$$w_2 = w_1$$

$$w_3 = \frac{1}{3n} + \frac{1}{4n}$$

$$w_3 = \frac{1}{4n}$$

Considerando este desarrollo para b muestras se tiene que:

$$w_1 = \sum_{k=2}^{b+1} \frac{1}{kn} \quad , \quad w_2 = w_1$$

$$w_3 = \sum_{k=3}^{b+1} \frac{1}{kn} \quad , \quad w_4 = \sum_{k=4}^{b+1} \frac{1}{kn}$$

En la forma general se tendría que:

$$w_t = \frac{1}{n} \sum_{k=t}^{b+1} \frac{1}{k} = \frac{1}{n} \cdot \frac{1}{1 - \left(\frac{1}{k}\right)^{b+1}}$$

Ecuación No. 2:

En la ecuación anterior puede verse como el número de alelos incide sobre la velocidad de convergencia $w_i \rightarrow 1/n$ así:

- Si k , es decir el tamaño del polimorfismo, es pequeño w_i tiende a $1/n$ de una manera más lenta. Es decir que se necesita un mayor tamaño de muestra, por lo tanto b es un valor mucho más grande.
- Si k es un valor grande, entonces el cociente $1/k$, tiende a cero a mayor velocidad, por lo tanto se necesita un menor tamaño de muestra, es decir b es un valor más pequeño.

2.2.2. Desviación Estándar.

Posteriormente, la desviación estándar se determinó utilizando las siguientes ecuaciones:

Calculada para la primera y segunda variante alélica:

$$S_1 = \sqrt{\frac{\sum (\bar{x}_{i1} - \bar{x}_1)^2}{b}}$$

$$S_2 = \sqrt{\frac{\sum (\bar{x}_{i2} - \bar{x}_2)^2}{b}}$$

La forma calculada para el k ésimo alelo es:

$$S_k = \sqrt{\frac{\sum (\bar{x}_{ik} - \bar{x}_k)^2}{b}}$$

Donde S_k corresponde a la desviación estándar a lo largo de los muestreos realizados para el alelo kn . En el muestreo específico m . A partir de estos valores posteriormente se desarrolla una matriz de datos escalonada.

2.2.3. Coeficientes de variación.

Finalmente se determina el coeficiente de variación para la distribución de frecuencias en cada variante alélica: Ecuación No. 3.

Ecuación No. 3:
$$CV_{ik} = \frac{S_{ik}}{x_{ik}}$$

2.2.4. Alelos efectivos y alelos raros.

Los alelos en el polimorfismo se clasifican como alelos efectivos y alelos raros según la ecuación No. 4, basada en los valores de la homocigocidad del sistema:

Ecuación No. 4:
$$A_e = \frac{1}{\sum k_i^2}$$

Debido a su comportamiento diferencial respecto al coeficiente de variación, los alelos en baja frecuencia siguen presentando oscilaciones muy altas que no permiten un acercamiento claro a la definición de un tamaño de muestra, y es por esto que el procedimiento se centra en los alelos efectivos.

2.3. Selección de un tamaño de muestra.

Inicialmente se selecciona un tamaño de muestra por cada sistema microsatélite utilizando dos criterios:

1. Un muestreo m_i en el cual aparece la totalidad de los alelos reportados en el muestreo poblacional, incluidos los alelos en baja frecuencia.
2. Un valor menor o igual a 0.05 en las fluctuaciones de los coeficientes de variación de las frecuencias para los alelos efectivos. Puede utilizarse un segundo criterio para la desviación estándar en el valor de frecuencia alélica inferior a 0.01, que es concordante con la definición de polimorfismo, es decir oscilaciones de la frecuencia de un alelo de 1%.

2.4. Análisis multi loci.

El análisis secuencial se realiza para cada uno de los sistemas en estudio, por tanto se obtiene para cada uno de ellos un tamaño de muestra m_i por sistema, luego el N final corresponderá al mayor valor m reportado en el conjunto total de marcadores evaluados. El valor N final se obtiene a partir de múltiples iteraciones, para estimar la dispersión en los valores de frecuencias alélicas.

2.5. Aplicación desarrollada.

Una aplicación computacional fue desarrollada en Visual Basic Applications (VBA) y los datos manipulados en microsoft Access ® por la facilidad de manejo y de implementación en los

laboratorios de identificación, que suelen trabajar con hojas de cálculo generadas por los diferentes modelos de analizadores genéticos y como herramienta para la utilización y apropiación del método desarrollado. Los algoritmos desarrollados siguen los lineamientos de calidad para un laboratorio de ensayos ISO-17025 e ISO 9001, además de los métodos de cálculo programados para la resolución de los ejercicios de proeficiencia de la ISFG.

2.6. Datos empleados.

Una vez obtenido el algoritmo, este se verificó utilizando los datos poblacionales registrados por el grupo de Genética de Poblaciones e Identificación del Instituto de Genética de la Universidad Nacional de Colombia (GPI-IGUN). La población de Bogotá (Ciudad Capital) es un buen modelo para el análisis de tamaño de muestra y variabilidad de un polimorfismo, cuenta con una población de aproximadamente 7.000.000 millones de habitantes en la que se da un activo proceso de migración desde todo el país e incluso de algunos países andinos cercanos, también recibe influencia de la región Caribe, con una marcada influencia de grupos afro descendientes y en menor proporción de población Europea. También se encuentran grupos de ancestría indígena del centro del país. Todos los datos de perfiles genéticos utilizados cuentan con su respectivo consentimiento informado siguiendo la normatividad respectiva para su utilización. Los sistemas microsatélites empleados son los utilizados por el sistema CODIS, 13 en total. El método se explica en su totalidad utilizando el microsatélite D5S818.

3. Resultados.

3.1. Cálculo inicial para un solo sistema STR.

El procedimiento se explica en detalle con el sistema D5S818:

1. Se utilizó una base de datos con 1.331 personas, sobre la cual se realizaron un total de 100 muestreos (m_{100}). La gráfica No.1 muestra las oscilaciones de cada variante alélica a medida que se incrementa gradualmente el tamaño de muestra.

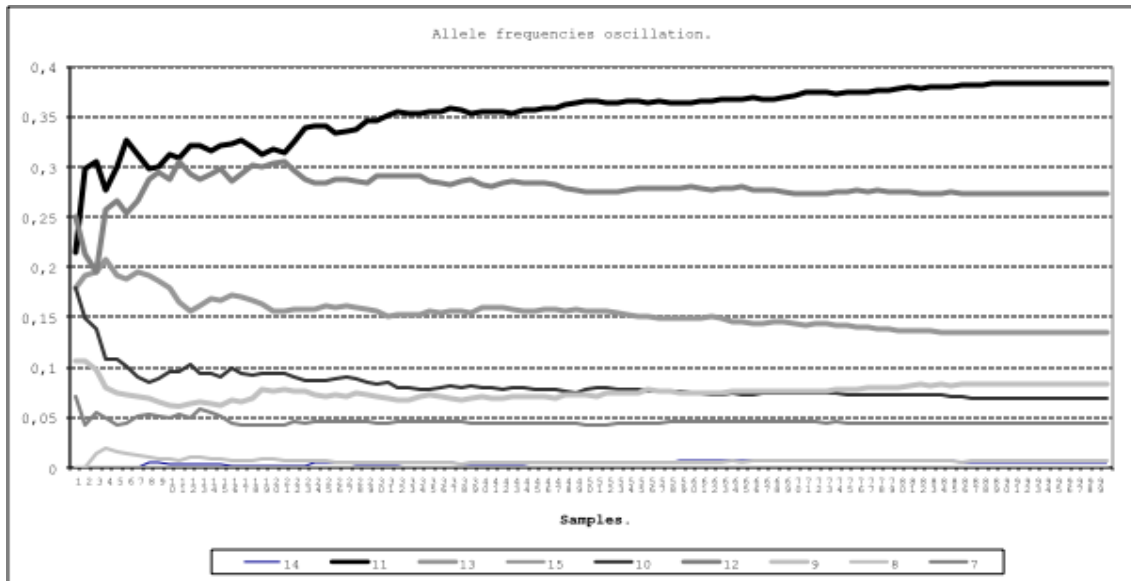


Figura No 1. Comportamiento de las frecuencias alélicas del sistema D5S818 a lo largo de 100 muestreos secuenciales. Cada muestreo adiciona un número n de personas. Las distribuciones iniciales presentan mayores fluctuaciones en el valor de frecuencias entre muestreos, sin embargo poco a poco se tiende a un valor estable para cada variante alélica acercándose al valor real del parámetro poblacional.

El comportamiento de las primeras distribuciones de frecuencias es completamente errático pero poco a poco el valor de cada frecuencia alélica se estabiliza, sin embargo aún no se puede estimar ningún tamaño de muestra óptimo puesto que esta es solo una medida gráfica.

En la siguiente fase del algoritmo, se calcularon los promedios como indica la matriz número 2, también fueron calculadas las desviaciones estándar, dado que son los parámetros para calcular el Coeficiente de variación. A medida que crece la muestra el coeficiente de variación disminuye notablemente, acercándose al valor de 0.01, establecido como el punto en el cual se selecciona una muestra en la distribución de frecuencias d_i asociada a un tamaño de muestra n_i . La gráfica No. 2b muestra como los alelos en baja frecuencia generan mayores oscilaciones en los coeficientes de variación, por tanto fueron retirados para la determinación del tamaño de muestra.

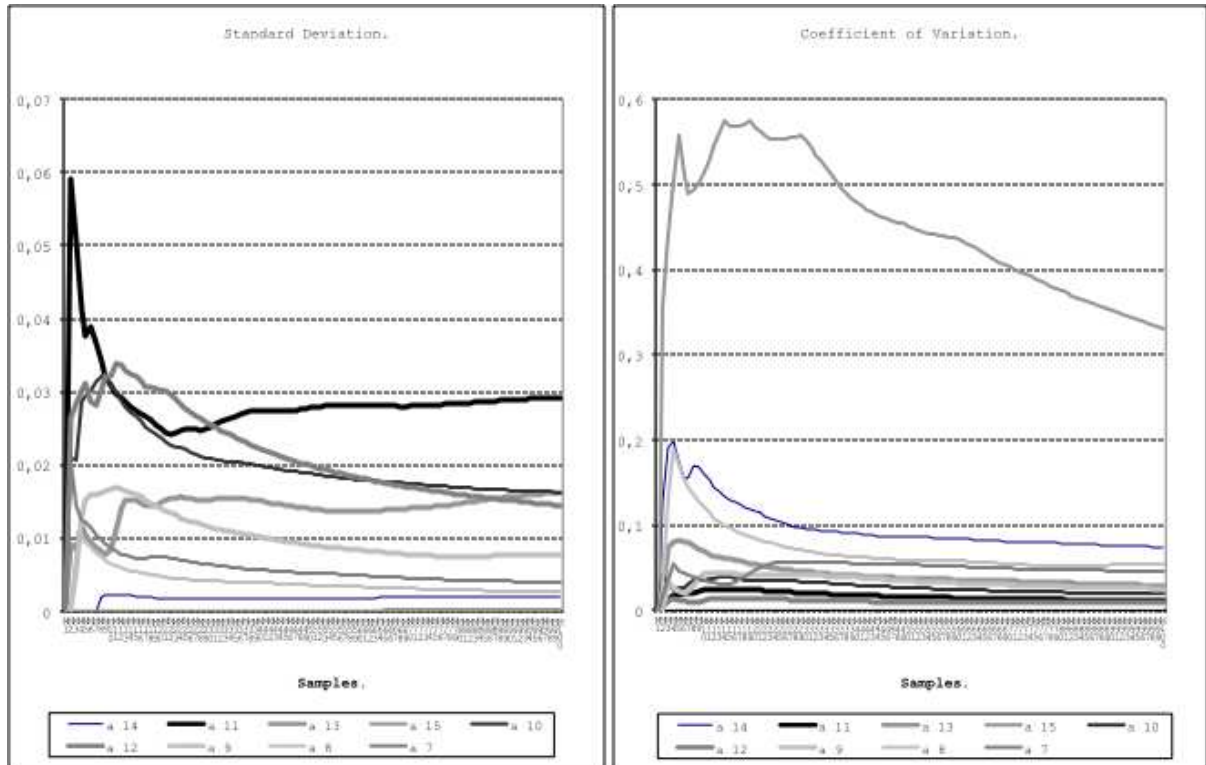


Figura No. 2 Las gráficas anteriores muestran el comportamiento del sistema microsatélite D5S818 en el experimento propuesto. A medida que se incrementan los muestreos, se logra una tendencia a disminuir considerablemente la desviación estándar, la gráfica de coeficientes de variación muestra la misma tendencia, sin embargo se da una separación entre alelos en baja frecuencia y alelos efectivos.

3.2. Tamaño de muestra establecido para el sistema completo.

Aplicando el algoritmo secuencial, se determinaron los siguientes tamaños de muestra para cada uno de los sistemas microsatélites estudiados.

STR'S	Category.	No. Alleles	Alleles found.	% alleles found.	No. Efective alleles.	% efective alleles.	Unique Alleles	Sample	No. Ind.
D5S818	Simple	10	9	90%	4	44%	1	24	308
D16S539	Simple	10	8	80%	5	63%	1	4	56
D13S317	Simple*	14	10	71%	6	60%	2	12	152
CSF1PO	Simple	15	10	67%	4	40%	2	22	250
TPOX	Simple	17	9	53%	3	33%	4	7	86
D8S1179	Compuesto	19	10	53%	5	50%	3	2	28
D7S820	Simple	22	11	50%	5	45%	3	16	213
D18S51	Simple	42	20	48%	8	40%	4	53	709
D3S1358	Compuesto	20	9	45%	4	44%	2	2	29
VWA	Compuesto	28	12	43%	4	33%	4	10	134
TH01	Simple*	21	7	33%	4	57%	0	17	208
D21S11	Complejo	70	23	33%	6	26%	5	48	651
FGA	Compuesto	98	21	21%	8	38%	7	26	339

Tabla No. 1: Cálculo de tamaño de muestra para los sistemas STR's del sistema CODIS. Además del tamaño de muestra se presenta el análisis porcentual de cada sistema en función del número de variantes alélicas reportadas. No.de alelos, es el número de alelos reportados a nivel mundial para el polimorfismo, Alelos encontrados se refiere a los alelos hallados en la muestra estudiada. No. Alelos efectivos es el número de alelos que corresponden a los alelos más representativos de la población según el procedimiento de Buttlar para separar el polimorfismo (Buttler, 2007). Sample es el m_i en el cual el coeficiente de variación alcanza un valor inferior a 0.01.

El tamaño de muestra sugerido por la aplicación es de 709 personas, número elevado en comparación con estudios tradicionalmente desarrollados en poblaciones humanas que están alrededor de 200 muestras. En la población de Bogotá con diferentes orígenes y un activo flujo de migración de diferentes grupos étnicos, se encontró una amplia distribución de los polimorfismos STR's: Un promedio de 52,8% variantes alélicas por sistema sobre el número total reportado en el mundo. Sin embargo, la dispersión en este valor es alta (entre 21% a 90%), por tanto se evaluó mediante un coeficiente de correlación entre el número de variantes reportadas por sistema polimórfico y el número de variantes encontradas en la población, con un valor alto de correlación: 0,8878. La correlación entre el número de alelos encontrados en la población de estudio y el

tamaño de muestra reportado por este método igualmente presentó un valor de 0,8211. Es decir entre más variantes alélicas tiene un sistema, mayor es el tamaño de muestra a emplear. Este comportamiento se dio independientemente de la complejidad estructural de cada sistema microsatélite. Inicialmente se había pensado que los microsatélites simples, deberían arrojar tamaños menores, e ir incrementando según la estructura, esta tendencia no se vio en el estudio, luego la variable complejidad del microsatélite fue descartada de la interpretación del problema.

Si bien, aunque la muestra requerida es de 709 personas, la distribución completa de alelos efectivos en la población se encontró con un número de 230 personas. Por tanto la distribución de frecuencias alélicas para alelos efectivos puede ser similar en una muestra pequeña que en una grande. Este efecto nos indica que aunque los sistemas microsatélites son ampliamente polimórficos, en términos efectivos en la población se presentan unas pocas variantes. No importa si un polimorfismo es grande, importa es que tan representado está en la población.

Después de un tamaño de muestra mínimo (230 personas), que explica en promedio el 90% de los genotipos en la población, se da exclusivamente la aparición de alelos raros los cuales permanecen en baja frecuencia. El ampliar la muestra por encima del tamaño mínimo propio de los alelos efectivos tendría la ventaja de poder representar el polimorfismo completo del sistema microsatélite, lo cual es útil en estudios poblacionales y en los casos de identificación. Una muestra de aproximadamente 200 individuos, simplemente representaría la población efectiva, dejando la distribución de alelos raros menormente representada.

Adicionalmente al tamaño de muestra, se calcularon intervalos de confianza para cada variante alélica y también un intervalo general por marcador, el cual permite hacer una aproximación de la dispersión de una frecuencia alélica para variantes nuevas en la población de referencia. En el anexo No.1 se presenta la estimación de frecuencias alélicas por intervalos de confianza para la población de Bogotá. La siguiente tabla muestra el valor del intervalo de confianza general de cada sistema evaluado.

STR'S	Promedio	Min	Max
D5S818	0.0016	0.0001	0.0029
D16S539	0.0072	0.0009	0.0108
D13S317	0.0009	0.0001	0.0019
CSF1PO	0.0033	0.0005	0.0068
TPOX	0.0141	0.0035	0.0209
D8S1179	0.0011	0.0001	0.0029
D7S820	0.0020	0.0002	0.0040
D18S51	0.0108	0.0019	0.0199
D3S1358	0.0012	0.0001	0.0030
VWA	0.0018	0.0002	0.0039
TH01	0.0002	0.0000	0.0005
D21S11	0.0005	0.0000	0.0015
FGA	0.0002	0.0000	0.0007

Tabla No 2 Intervalos de confianza promedios de cada sistema microsatélite. En el anexo No. 1 se estimó el valor del intervalo de confianza de cada variante alélica reportada, sin embargo se realizó una estimación promedio por sistema microsatélite que es útil en cálculos generales sobre el sistema o como valor de referencia en el caso de la aparición de alelos nuevos.

4. Discusión.

Normalmente en cualquier análisis genético poblacional, el investigador realiza un conteo directo para la determinación de frecuencias alélicas sin considerar una medida de la variación del estimador, esto se debe al uso tradicional de aplicar modelos determinísticos en lugar de usar modelos probabilísticos. Esto es un problema en muestreos con n pequeño, y con criterios poco rigurosos de inclusión. El algoritmo anteriormente planteado, resuelve estos problemas. En primer término al calcular un valor de frecuencia alélica, asociado a un intervalo de confianza, se hace posible determinar la calidad del muestreo realizado en función de la dispersión, obtenida a partir de múltiples repeticiones del algoritmo secuencial.

Es también importante considerar después de cierto punto, en el cual se han alcanzado todos los alelos efectivos de un polimorfismo, el esfuerzo de muestreo que se realice dado que no se lograría un mayor nivel de resolución incrementando sustancialmente la muestra. Con un número mayor se asegura una mejor representatividad de alelos raros, pero no cambios importantes en la distribución de frecuencias.

Algunas ecuaciones permiten calcular intervalos de confianza para distribuciones de frecuencias alélicas, Gillespie, (1998), sin embargo este procedimiento matemático, presenta un argumento circular. Como puede verse en la Ecuación No. 5:

Ecuación No. 5:

$$I.C. = \bar{P} \pm 1.96 \sqrt{P \cdot (1 - P) / n}$$

Se requiere del n muestral que es el número de muestras obtenido por el investigador, pero no corresponde a ningún cálculo, es el valor obtenido. El procedimiento propuesto en este artículo, calcula un valor de intervalo de confianza, pero este valor se reporta cuando el tamaño de muestra y los valores de coeficientes de variación han llegado a un límite inferior. Por tanto el cálculo del tamaño de muestra y la obtención de los intervalos de confianza, son procesos simultáneos, si se da un incremento en el tamaño de muestra se realiza el recálculo de los intervalos de confianza. Intervalos de confianza demasiado grandes indicarían tamaños de muestra insuficientes o con criterios de inclusión poco claros. Calcular tamaños de muestra e intervalos de confianza de esta manera, permiten hacer un cambio conceptual, pasando de ecuaciones establecidas, donde primero se calcula un tamaño de muestra y luego un intervalo, a procedimientos simultáneos que requieren de un cuidadoso análisis de la muestra a medida que ésta es tomada y no tanto del seguimiento de una ecuación inicial para obtener un tamaño específico.

Normalmente en la práctica, se realizan cálculos muestrales *a priori* específicos en cada estudio, sin embargo lo que puede verse al utilizar una metodología *a posteriori*, es que se obtiene un valor del comportamiento de un polimorfismo en la población, no un valor para un estudio específico, corresponde más a un estudio muestral sobre la variabilidad genética de la población,

convirtiéndose en un valor de referencia para otros estudios a realizar y no simplemente en el cálculo de tamaño de muestra.

El cálculo de frecuencias alélicas asociadas a intervalos de confianza implica también el cambio en la forma de estimar los demás estadísticos forenses y genético poblacionales, es decir los estadísticos como los índices y probabilidades de paternidad posiblemente puedan pasar de ser estimadores puntuales a estimadores por intervalos.

En muchas oportunidades al realizar un determinado análisis no sabemos si estamos cercanos a un tamaño óptimo, los cálculos de tamaño de muestra permiten hacer un acercamiento al mejor valor en relación con la variabilidad genética presente en la población. Normalmente se utilizan tamaños pequeños de muestra y se piensa que entre más muestras mejor, sin embargo, números excesivos de muestra terminan generando costos innecesarios y procesos computacionales más complejos de lo requerido.

La clasificación de alelos efectivos y raros arrojaría dos tipos diferentes de información, la correspondiente a los alelos que conforman en mayor medida la población de estudio, y que dan cuenta de la mayoría de genotipos encontrados en la población, y los alelos raros que dan mayor resolución sobre los procesos evolutivos seguidos en el tiempo. La presencia de variantes en baja frecuencia suelen ser más características de la población, especialmente originadas por procesos de migración y mutación. Por lo tanto sugerimos dar un peso diferencial en los análisis de similitud poblacional a los alelos efectivos y alelos en baja frecuencia, tal que las variantes en baja frecuencia podrían ser útiles en los análisis de ancestría y en casos de identificación y filiación.

5. Conclusión.

Este trabajo se ha realizado en base a las frecuencias alélicas, sin embargo el cálculo de intervalos de confianza puede hacerse extensible a todas las demás medidas que suelen calcularse sobre una población, no todos podrían obtenerse directamente por análisis de muestreo secuencial, pero al contar con intervalos de confianza para las frecuencias alélicas, dichos intervalos puede ser

utilizados para calcular la dispersión de los demás estimadores. En particular llama la atención los cálculos por intervalos de confianza de índices de paternidad, que hasta el momento son calculados como estimadores puntuales, sin embargo podrían estimarse como intervalos también con la finalidad de lograr una mejor interpretación de la relación biológica del padre en cuestión con la población de referencia.

Anteriormente se utilizaban límites para la interpretación de las probabilidades de paternidad como sucedía con los enunciados verbales de Hummel, estos fueron eliminados y reemplazados por valores puntuales, el uso de intervalos de confianza, podría permitir nuevamente el uso de enunciados ajustados a los nuevos niveles de resolución molecular, a los métodos de cálculo y la diversidad de los actuales marcadores. Esto podría facilitar la interpretación de autoridades y usuarios, logrando una mejor explicación en términos poblacionales.

Por otra parte, hemos probado el presente algoritmo con marcadores moleculares STR's, sin embargo el comportamiento y los tamaños de muestra reportados, podrían ser menores en marcadores moleculares SNP's, dado el nivel de polimorfismo.

La actual teoría evolutiva arroja valores de variabilidad genética de una magnitud tal que permiten la individualización de cada persona en el planeta, luego los cálculos de tamaño de muestra en estudios con múltiples marcadores genéticos pueden ser muy grandes, creemos que el uso de herramientas complementarias tales como los análisis demográficos y genealógicos pueden ser herramientas útiles para asegurar la calidad de los resultados de investigación, disminuyendo considerablemente los tamaños de muestra. Más que realizar un cálculo de tamaño de muestra, se necesita realizar un estudio muestral en cada población que además considere las características genealógicas de la población como criterios de inclusión en muestras. De esta forma se podría evaluar la distribución en el tiempo y el espacio de diferentes polimorfismos en la población debido a la construcción de mapas y planos factoriales que serían la base para estudios posteriores en genética forense.

Bibliografía.:

- Bautista L. (1998). Diseño de Muestreos Estadísticos. Bogotá: Universidad Nacional de Colombia. 160 p
- Campbell AW, Campbell PJ, Griffin WB, Burritt DJ, and Conner AJ. 2001. Prediction of Sample Size to Maintain Genetic Variation in Doubled-Haploid Populations Following Marker Selection. *Theoretical and Applied Genetics* 103:1028-1036.
- Chakravarty S. 2007. Sample Size Determination for Multinomial Population. Available from: [http://tx.liberal.ntu.edu.tw/~PurpleWoo/Literature/!Methodology/!Distribution_SampleSize/Sample Size Determination for Multinomial Population.pdf](http://tx.liberal.ntu.edu.tw/~PurpleWoo/Literature/!Methodology/!Distribution_SampleSize/Sample%20Size%20Determination%20for%20Multinomial%20Population.pdf)
- Chakraborty R. (1992). Sample Size Requirements for addressing the population Genetic issues of forensic use of DNA Typing. *Human Biology*. Vol 64, No.2.
- Evetts, I. W. & S. Weir. (1998) *Interpreting DNA Evidence*. Sinauer, Sunderland, MA...173 p.
- ISFG 1992. International Society for Forensic Haemogenetics. DNA recommendations — 1992 report concerning recommendations of the DNA Commission of the International Society for Forensic Haemogenetics relating to the use of PCR-based polymorphisms. *International Journal of Legal Medicine*. Vol. 105, No. 1.
- ISFG. 2002. Annual Report Summary for Testing. American Association of Blood Banks. [Internet]:1-62. Available from: http://www.gep-isfg.org/documentos/AABB_2004.pdf
- ISFG. 2004. Annual Report Summary Report Summary for Testing. American Association of Blood Banks [Internet]:1-62. Available from: http://www.gep-isfg.org/documentos/AABB_2004.pdf
- Gillespie J. H. (1998). *Population Genetics. A concise guide*. The Johns Hopkins University Press. Baltimore. Pag 142
- Goldstein D, Schlotterer C. (1999). *Microsatellites*. Londres: Oxford University Press. 352 p.
- Hall P, and Peng L. 1999. On Prediction Intervals Based on Predictive Likelihood or Bootstrap Methods. *Biometrika* 86:871-880.

Harding H. (1998). DNA Database Size. *Journal Forensic Science*. 15:345-349

Hey J, and Nielsen R. 2004. Multilocus Methods for Estimating Population Sizes, Migration Rates and Divergence Time, with Applications to the Divergence of *Drosophila Pseudoobscura* and *D. Persimilis*. *Genetics* 167:747-760.

Hohls T. 1998. Reliability of Confidence Interval Estimators under Various Nested Design Parental Sample Sizes. *Biometrical Journal* 40:85-98.

Kalinowski ST. 2005. Do Polymorphic Loci Require Large Sample Sizes to Estimate Genetic Distances? *Heredity* 94:33-6.

Medina R. (2006). Number of Individuals and molecular markers to use in genetic differentiation studies. *Molecular Ecology* Vol 6.

Wang J, and Whitlock MC. 2003. Estimating Effective Population Size and Migration Rates from Genetic Samples over Space and Time. *Genetics* 163:429-446.

Capítulo No. 3: Análisis de la Estructura de la población de La Guajira: Una visión Genética Demográfica y Genealógica.

M. Rojas¹, A. Alonso¹, L. Eljach¹, V. Sarmiento¹, W. Usaquén¹.

Grupo de Genética de poblaciones e Identificación. Instituto de Genética. Universidad Nacional de Colombia.

Resumen

En este estudio se caracterizó la subestructura genética y la dinámica migracional de la población de La Guajira-Colombia desde un punto de vista genético, demográfico y genealógico. Se analizó una muestra de la población tanto residente como ancestral (Wayúu) que habita actualmente la región. Basados en esos criterios, se plantearon unas categorías genealógicas y un diseño de muestreo, que diera cuenta de la diversidad, subestructura, del efecto de la deriva genética y de la migración de la población. Ambos tipos de variables tanto demográficas como genéticas (marcadores moleculares STRs autosómicos) permitieron dilucidar una estructura genético-poblacional robusta y consistente con la dinámica migracional de la Guajira. Se encontró una subpoblación con ancestría Wayúu con altos niveles de heterocigosidad (H_o 0.728) que se diferencia de los demás grupos genealógicos (Θ 0.024 IC: 0.023 \pm 0.006) exceptuando los individuos con genealogía Wayúu-Guajiro. Para los microsatélites evaluados, no se encontró que estuviera endogámica (F_{st} 0.017). Los datos demográficos mostraron a la población Guajira como joven, en crecimiento y expuesta a un moderado efecto de la deriva genética (N_e 11.903). Geográficamente, los Wayúu coexisten en algunos territorios con los demás grupos y tienen un moderado flujo génico con ellos, en su mayoría con individuos Wayúu-Guajiros. La dirección del flujo génico en su mayoría va en dirección externa a la población Wayúu. En síntesis, la cuidadosa selección de la muestra nos permitió obtener un modelo de subestructura y del patrón de migración, causado por procesos culturales, históricos y geográficos.

Palabras Clave

Wayúu, Subestructura, Migración, Deriva genética, Demografía, Genealogía.

1. Introducción.

La península de La Guajira se encuentra ubicada en las coordenadas 10° 23' y 12° 28' latitud norte, 71° 06' y 73° 39' longitud oeste. Es la región más septentrional de Sudamérica, entre el extremo nororiental de Colombia y el extremo noroccidental de Venezuela, tiene una superficie de 25.000 Km², en su división política pertenece en mayor parte al departamento de La Guajira (20.848 Km²) (Gobernación de la Guajira, 2011). La población colombiana de La Guajira según el DANE (Departamento Nacional de Estadística) es de 846.609 habitantes con una tasa de crecimiento de 1985 a 2005 del 87.95%, donde el 51,9% de los habitantes de la Guajira, viven en las Cabeceras municipales y el 48.1% en región rural. En general la densidad poblacional es 28.37 hab./km², por debajo de la media nacional de 40,7 hab. /km².

En este territorio existe una gran diversidad étnica, representada por pueblos nativo-americanos de las familias lingüísticas Arawak y Chibcha, una población residente afrodescendiente y colombianos de múltiples ancestrías. Por otro lado hay presencia de sirio-libaneses, especialmente en la ciudad de Maicao (CELADE, 2011). Los Wayúu (Arawak) son el pueblo indígena que representa el 44.9% de la población del departamento de la Guajira, siendo a la vez pueblo indígena más grande de Colombia, constituido por más de 144.000 personas que representan el 20.5% de la población indígena nacional (DANE, 2005).

Ante este escenario, se partió de una primera hipótesis: La población Wayúu es una población diferenciada del resto de población residente por su distribución geográfica, composición cultural, y su historia de poblamiento. Primero que todo, la península está constituida geográficamente en el centro y norte por una planicie baja con un clima de estepa árida o semiárida. Esta es considerada la región más seca de Colombia, donde se han asentado los Wayúu por miles de años, estableciendo numerosos clanes o tribus. Por otra parte al sur de la península, se encuentra la Sierra Nevada de Santa Marta, la serranía del Perijá y la planicie aluvial de los ríos Ranchería y

Cesar, con un clima diferente, de selva húmeda de montaña, selva seca y sabana xerófila. En la Sierra Nevada se ubican comunidades indígenas ancestrales como los Kogi, Ika y Kankuamo (Gobernación de la Guajira, 2011).

La adaptación de los Wayuu a las condiciones inhóspitas del desierto, les sirvió como refugio de los europeos durante la colonización, en consecuencia no sufrieron invasión sino hasta después de la independencia de Colombia y Venezuela, manteniendo hasta hace poco una amplia autonomía extralegal. De esta manera, parte de la comunidad Wayuu se ha mantenido relativamente aislada en rancherías lejanas dispuestas irregularmente dentro del territorio (Vásquez S & Correa H, 2000). De esta manera la densidad poblacional de los Wayúu es mayor en Nazaret, Uribí, la Serranía de Jala'ala y las sabanas de Wopu'müin, zonas de la alta Guajira, sin embargo también habitan en las ciudades principales (Programa Presidencial de Derechos Humanos, 2011).

Según el Antropólogo José R. Oliver (1991), el origen de los Wayúu según un criterio lingüístico es Arawak, de la región Amazónica. Se presume que la población migró desde el Río Negro hacia la costa occidental de Venezuela y La Guajira hace 4000 a 5000 años, en esa trayectoria, desde hace 3000 a 2000 años hubo una separación entre Lokono y Wayúu y desde hace 1500 a 1000 años hubo una separación entre Wayúu y Paraujano, siendo posible que estas divergencias ocurrieran en la región que conecta el Orinoco y los llanos con la Amazonia central, por último, avanzaron desde este punto hacia el norte por la costa occidental de Venezuela, adicionalmente, a principios del siglo XVI en la península cohabitaban con otros grupos indígenas de diferente origen lingüístico (Caribes y Chibchas), de múltiples orígenes y con diferentes historias evolutivas. Desde esta época hasta el Siglo XVI y el presente, los distintos pueblos convivieron de manera pacífica compartiendo su territorio, sin embargo muchos clanes desaparecieron unos por fusión y los demás por causas no evidenciadas (Oliver JR, 1991).

La configuración cultural y social del grupo indígena wayuu, ha estado influenciada por su adaptación al ecosistema, es así como los Wayúu suelen no vivir en asentamientos estables, dado que con frecuencia se trasladan dependiendo del régimen de lluvias (Arango, 2004; Ardila G,

1992). Desde una perspectiva social y política, en los Wayúu no existe un poder central, se organizan en clanes o castas matriarcales con matrilocidad, que comparten una condición social y un pasado común, es así como las relaciones de parentesco, el e'irukuu (correspondiente al apellido o casta) se hereda por línea materna al igual que la identidad Wayúu (Arango R, 2004). En Colombia hay 22 castas entre las que se destacan por su tamaño los Epiayú 20.8%, Uriana o Uliana 17.1% Ipuana 16,2%, en particular, el matrimonio suele practicarse entre miembros de diferentes castas, se practica la monogamia y la poligamia en una versión poligínica, donde cada esposa vive con su madre y hermanas. La identidad Wayúu se hereda por línea materna, por lo que algunos individuos de ancestría múltiple (madre Wayuú-padre Arijuna) se consideran Wayúu, mientras que el caso contrario no; por otro lado, los primos maternos no pueden casarse por ser del mismo linaje, pero sí es posible entre primos paternos (Arango R, 2004).

En retrospectiva, la segunda hipótesis se fundamenta en la residencia en la región de migrantes de múltiples ancestrías, principalmente provenientes de la costa Caribe colombiana y algunos del resto del país que se concentran en los asentamientos urbanos principales; aquí, la dinámica migracional regional podría caracterizarse por inmigración, en su mayoría de corta distancia. Adicionalmente en la zona viven los *Guajiros*, personas que se consideran de la región pero que no se identifican como de ancestría Wayúu. Es así como en las zonas urbanas suelen coexistir tanto migrantes, guajiros e individuos Wayúu, compartiendo el territorio y algunas actividades tanto económicas como sociales.

Uno de los supuestos iniciales es el de un moderado flujo bidireccional entre grupos generando diferentes modelos de ancestría. Un ejemplo claro es el matrimonio *Wayúu-Arijuna* (persona no Wayúu), que se da principalmente en enclaves urbanos que sirven como concentración tanto de migrantes e indígenas que están en búsqueda de una fuente de trabajo. Es así como el estudio de variables como el conocimiento del lenguaje (Wayuunaiki) y la pertenencia a una casta, permitirían asociar individuos con ancestría Wayúu a menores niveles de introgresión, dado que este conocimiento suele perderse con la aculturación. Además, no se tiene certeza de que las poblaciones aisladas geográficamente, difieran considerablemente de las que se encuentran en

contacto permanente con poblaciones externas.

Con el fin de obtener este nivel de resolución en el estudio de la subestructura y migración, se ha acudido a las características demográficas, genealógicas y socioculturales propias de las poblaciones, debido a que el estudio de la estructura de edad y sexo, la distribución espacio-temporal, los caracteres sociales, económicos y culturales permiten tener un conocimiento del pasado, presente y futuro cultural, social y biológico de la población de estudio (Albeza MV. et al, 2004), además permiten evaluar el efecto de la deriva genética en la diferenciación poblacional. Es así como la evaluación de elementos como el tamaño efectivo poblacional, el aislamiento y la panmixia junto a los datos genéticos es importante porque influyen directamente en la estructura genética de las poblaciones (Albeza MV. et al, 2004). No obstante, tradicionalmente en los modelos de genética dep se han ignorado estos factores, que permiten obtener hipótesis más acertadas de la dinámica poblacional sobre las fuerzas de cambio evolutivo. (Albeza MV. et al, 2004). Es por esta razón que el objetivo de este estudio fue caracterizar la subestructura genética y la dinámica migracional de la población de La Guajira-Colombia desde un punto de vista genético, demográfico y genealógico; implementando una metodología que nos permitió encontrar una población conservada de ascendencia Wayúu que se diferencia del resto de la muestra, pero que converge en algunas zonas geográficas con la población residente con una dinámica migracional bidireccional activa.

2. Metodología

Se realizó un muestreo dirigido para la población de la Guajira y la población Wayúu con un tamaño de muestra de 290 individuos. El muestreo se lleva a cabo en los hospitales de los

municipios de Riohacha, Maicao y Uribí (Figura 1) y en algunas rancherías de los tres municipios. Previo consentimiento informado, se tomaron muestras de saliva ó 10ml de sangre periférica, cada una fue preservada mediante una tarjeta FTA (Whatman BioScience, USA). El ADN fue extraído tomando un fragmento de la matriz de la tarjeta.



Figura No. 1 Mapa de la Guajira mostrando la localización de las tomas de muestra.

Con el fin de entender la interacción entre la dinámica de estos loci y la organización social, se planteó una encuesta que permitiera indagar sobre la información demográfica de la población y su pertenencia étnica (comunidad indígena Wayúu), al mismo tiempo, se reconstruyó la historia genealógica de cada individuo indagando por su lugar de nacimiento, el de su parentela y la pertenencia a un grupo étnico en las previas tres generaciones, esta información permitió hacer una distinción de la ancestría de cada grupo. Las variables tomadas en la encuesta fueron: género, edad, lugar de nacimiento, número de hijos, edad a la que se tuvo el primer hijo, edad del cónyuge, lugar de nacimiento del cónyuge y la pertenencia étnica del cónyuge (por autodeterminación). Se preguntó el lugar de nacimiento y la pertenencia étnica de tres generaciones anteriores, tanto por línea materna como paterna, por último, si el individuo pertenecía a la etnia Wayúu, se indagó sobre su casta, ranchería y si hablaba o no Wayuunaki.

Aunque el criterio de inclusión al estudio fueron personas residentes del departamento, con la información anterior también se pudo clasificar a los individuos de acuerdo a una reconstrucción genealógica *a priori*; es así como los lugares de nacimiento y los nombres de las castas se usaron como evidencia para la clasificación genealógica: Ubicamos como *Wayúu* a los individuos con ancestría proveniente de la etnia Wayúu tanto por linajes materno como paterno (Figura 2), *Guajiro* a los individuos con ancestría oriunda de este departamento que no pertenece a esta etnia, *Wayúu-Guajiro* a los individuos con ancestría por un lado Wayúu y por otra del departamento de La Guajira no perteneciente a la etnia, los individuos con una ancestría diferente al departamento de La Guajira se denominaron migrantes: *Migrantes 1* fueron los individuos con una ancestría tanto por linaje materno y paterno de un mismo departamento diferente a La Guajira, mientras que *Migrantes 2* fueron los individuos con linajes provenientes de departamentos distintos entre sí y por supuesto que no fueran de La Guajira. Por último los individuos de *ancestría múltiple* se definieron como los individuos con ancestría por un lado del departamento de La Guajira y por otro lado de un migrante. Esta clasificación se basó en la estructura genealógica de los individuos y refleja la dinámica del flujo génico de los diferentes grupos en la población, permitiendo obtener una muestra Wayúu según criterios genealógicos.

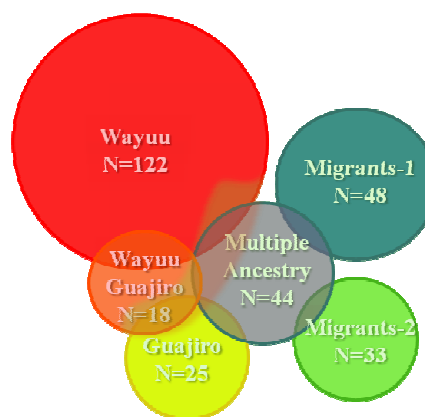


Figura No. 2 Grafica con las agrupaciones genealógicas y los tamaños de muestra

2.1 Análisis Molecular

Se amplificaron 15 loci STR autosómicos por PCR utilizando el kit comercial AmpFISTR® Identifiler™ (Applied Biosystems, Warrington, U.K.), un termociclador programable 9700 (Applied Biosystems, Foster City, California). Las amplificaciones fueron realizadas utilizando los siguientes parámetros por ciclo: denaturación inicial (95°C por 11 min) seguido por 28 ciclos de denaturación de las hebras (94°C por 1 min.), anillamiento de los primers (59°C por 1 min), extensión del DNA (72°C por 1 min), y extensión final (60°C por 60 min). Las alícuotas del ADN fueron amplificadas y marcadas con fluorescencia y mezcladas con formamida (Applied Biosystems, Foster City, California) y LIZ 500 (Applied Biosystems, Warrington, U.K.) y genotipificadas en un analizador genético ABI Prism 310 (Applied Biosystems, Stafford, Texas) utilizando el programa Genescan 3.1.2. Las asignaciones alélicas fueron hechas mediante la comparación de los fragmentos amplificados con escalera alélicas (Applied Biosystems, Warrington, U.K.) que contiene el kit y utilizando el programa Genotyper 2.5.2 (Applied Biosystems, Foster City, California).

2.2 Análisis Demográfico de la Muestra

A partir de la información demográfica se hizo un análisis de la distribución de edad y sexo de la muestra. Se estimó la edad promedio de la población total, masculina y femenina y de la fracción reproductora. Con base a las edades se calculó el tiempo generacional, el tamaño reproductivo de la población (N_r) (varones mayores de 18 años y mujeres entre 18-49 años con al menos un descendiente vivo entre 0-30 años), el tamaño efectivo poblacional (N_e) y tasa de migración efectiva (m_e) estimada a partir de la aproximación de Malécot (Magalhaes y Arce-Gómez 1987a). A partir de estos últimos parámetros se estimó el coeficiente de aislamiento reproductivo ($N_e m_e$) y el coeficiente de endogamia (F). Finalmente se realizó un análisis de la distancia geográfica entre parejas y la dirección del flujo génico estudiando las frecuencias de los matrimonios entre las diferentes clasificaciones genealógicas.

2.3 Análisis Genético de las Agrupaciones Genealógicas

Estas clasificaciones genealógicas *a priori* fueron analizadas mediante las siguientes metodologías. En primer lugar, en el programa Genepop 4.09 (Raymond M. & Rousset F, 1995; Rousset F. 2008) se calculó el promedio de la heterocigosidad (Nei, 1987) por locus usando la corrección de Levene. El test exacto del Equilibrio de Hardy Weinberg fue calculado en este mismo programa con los siguientes parámetros 10.000 periodos o ciclos de demorización, 1000 *batches* y 100 iteraciones. En el programa Fstat 2.9.3.2 (Goudet J, 1995) fueron calculados los estadísticos-F (Capf, Theta, Smallf) (Weir & Cockerham, 1984) por locus y un Fis (Smallf) dentro de los grupos genealógicos. Los intervalos de confianza de los estadísticos-F fueron estimados basados en técnicas de remuestreo jackknifey bootstrapal 95% and 99% sobre los loci. Se estimó el estadístico Rst (Slatkin, 1995).

Por otro lado, en Arlequin 3.5.1.2 (Schneider S. et al, 2000) se estimaron las comparaciones por pares de Fst (theta) y el Rst de las muestras con 10.000 repeticiones y un nivel de significancia de 0.05. Se calculó también, el análisis molecular de varianza (AMOVA) con Fst usando 99999 repeticiones.

Se calculó la distancia de Cavalli-Sforza and Edwards (1967) (Dc) con el método de agrupamiento Neighbor-Joining para dilucidar las diferencias entre las seis agrupaciones genealógicas. Esto mediante el programa Populations 1.2.30 (Langella O, 1999) y con un bootstrapping de 1000 sobre los loci. Por otro lado, para el análisis de los individuos en función de los alelos de los sistemas microsatélites evaluados se acudió a técnicas de análisis multivariado, específicamente el análisis factorial de correspondencias (AFC) en Genetix 4.05 (Belkhir K, 2001).

Se realizó un análisis de asignación de individuos a poblaciones en el programa Structure 2.3.3 (Pritchard et al, 2000) asumiendo un modelo de mezcla y de frecuencias alélicas correlacionadas.

Con un periodo de quemado de 100.000 y 100.000 MCMC repeticiones sobre el periodo de quemado. Al analizar los resultados de Structure se evidenció un $K=2$ consistentes con la agrupación genealógica Wayúu y un grupo externo. Después de esto, basados en la inferencia de ancestría de los individuos con un $K=2$ se calculó el nivel de introgresión en el grupo externo a la población Wayúu. Se utilizó como un estimativo del nivel de introgresión, la diferencia acumulada por locus y por sistema, y el promedio del valor absoluto del delta entre las distribuciones de frecuencias alélicas para todos los sistemas de las frecuencias alélicas entre las dos subpoblaciones.

2.4 Análisis de las Agrupaciones Genéticas

Teniendo en cuenta los valores de Q del análisis de Structure para un $K=2$, las clasificaciones genealógicas y la pertenencia a las castas Wayúu, se realizó un análisis de correspondencias múltiples en SPAD V. 5.6. Adicionalmente se usó el software Bayesass 1.2 (Wilson and Rannala 2003) para estimar la tasa de migración reciente entre los Wayúu y el grupo externo. Se hizo un análisis descriptivo de las subpoblaciones evidenciadas en la muestra (Wayúu-Grupo Externo), calculando heterocigosidad, equilibrio de Hardy Weinberg y el número de migrantes (Barton & Slatkin, 1986) en Genepop 4.0.1. Así mismo, se calcularon los estadísticos-F (F_{st} , F_{is} , F_{it}) (Weir & Cockerham, 1984) locus por locus, el F_{st} (F_{is}) para las dos subpoblaciones. Los intervalos de confianza de los estadísticos-F basados en las técnicas de remuestreo jackknifing y bootstrapping al 95% y 99% sobre los loci y los estadísticos-R (Slatkin, 1995).

3. Resultados

3.1 Análisis Demográfico de la Muestra

En la muestra se obtuvieron 39% hombres y 61% mujeres. Divididos en seis agrupaciones genealógicas: Wayúu 42.1% (122 individuos), Migrantes-1 16.6% (48), Ancestría Múltiple 15.2% (44), Migrantes-2 11.3% (33) Guajiros 8.6% (25) y Wayúu-Guajiros 6.2% (18). En la pirámide poblacional (Figura 3) se muestra las proporciones de sexo de los grupos de edades calculadas como porcentaje sobre la muestra, mostrando una estructura de una población joven y en

crecimiento, dado que los grupos etarios de la base están en mayor frecuencia. La edad promedio de la muestra fue 39.21 años, para los hombres 42.5 años y para las mujeres 36.07. La población entre 18-29 años fue 40.2% mientras que la población mayor de 45 años 33.56%. El tamaño reproductivo (Nr) fue de 173 individuos con una edad media de 45.93 años. Sobre la estructura de edad y de sexo se identificó que la muestra estaba conformada principalmente por personas jóvenes en edad reproductiva que aportan al pool genético de la población. El tiempo generacional de la muestra fue 19.55 +/- 4.8 años, calculado con base a la edad media a la cual las mujeres de la muestra tuvieron su primer hijo. Según la ENDS (Encuesta Nacional de Salud) (Profamilia, 2007) la edad en que las mujeres de la Guajira tienen su primer hijo es alrededor de 21 años y la tasa de fecundidad en la muestra es de 2.87 hijos por mujer.

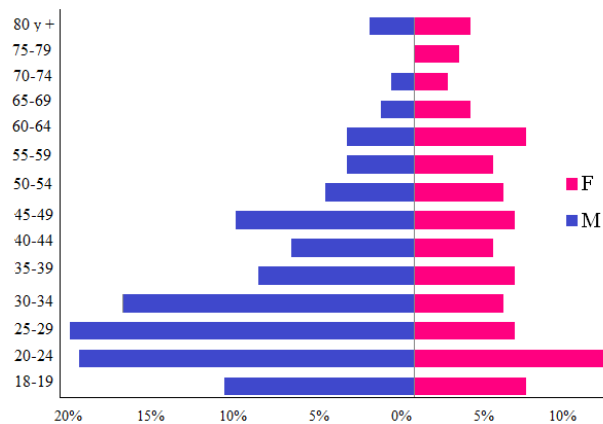


Figura No. 3 Pirámide de la muestra con la estructura de sexo y de edad. Este análisis fue calculado usando el porcentaje sobre la muestra.

Con base a las variables demográficas se evaluó algunos parámetros que involucran el efecto de deriva genética en la población. Encontrando un tamaño reproductivo de la población (Nr) de 155 individuos (53.4%) y en la población Wayúu 56 individuos (45,9%) (Tabla 2). Analizando el Nr en

los diferentes grupos de edades (generaciones de 19 años) se encontró la mayoría de población reproductiva entre 20 y 38 años (50%) para ambos grupos. Según el Nr de la población se estimó el tamaño efectivo en (N_e) 46.129 (13.5%) donde todos los adultos con descendencia, tienen igual probabilidad de ser los padres de cualquier descendiente individual. El N_e de la población total fue 46.129 (13,5%) y de los Wayúu 16.475 (13,5%). La tasa de migración efectiva es un estimativo de contribución de los migrantes al pool genético de la población calculado sobre la muestra. Para estimar este valor se establecieron como referencia a los grupos nativos (Wayúu, Wayúu-Guajiro y Guajiro) y como migrantes a larga distancia a Migrantes-1 y Migrantes-2 y a la agrupación de ancestría múltiple como migrantes a corta distancia. El 61.3% de los migrantes fueron a larga distancia, mientras que el 38.6 fueron a corta distancia. La tasa de migración (m_e) fue 0,258. Con el fin de medir los efectos de la deriva a partir del tamaño efectivo poblacional se calculó el coeficiente de aislamiento reproductivo (N_{me}) (Wright en 1938). El coeficiente de aislamiento reproductivo fue 11.9 lo que implica un efecto de deriva pequeño (Albeza MV. et al, 2004). Finalmente se hizo un estimativo del coeficiente de endogamia (F_{is}) de 0.0129 para la población total.

Parental Generation	Reproductive Individuals No.		(%) Percentage of Reproductive Size	
	Total	Wayuu	Total	Wayuu
First Generation (18-19 years old)	7	4	4.5161	7.1429 *
Second Generation (20-38)	79	28	50.9677	50 *
Third Generation (39-57)	48	14	30.9677	25 *
Fourth Generation (58-76)	18	7	11.6129	12.5 *
Fiveth Generation (77-95)	3	3	1.9355	5.3571 *
Reproductive Population Nr	155	56	53.4483	45.9016
k	4.2194	5.1964		
σ^2k	14.562	27.2516		
Effective Population Ne	46.129	16.475	15.906	13.504 **
Sample N	290	122		
Tasa de Inmigración (me)	0.258		0.5484	Migrants No./Nr
Proportion long-distance migrants (l)	0.1724		0.3226	l No./Nr
Proportion short-distance migrants (s)	0.1069		0.3500	s No./Nr
Isolation index (Neme)	11.9033			
Inbreeding coefficient (F)	0.0129			

Tabla No. 1 Tabla de Generaciones, tamaño reproductivo y aislamiento reproductivo. Tiempo Generacional Aproximado: 19 años. * Porcentaje sobre el tamaño reproductivo Nr. ** Porcentaje sobre el total de la población. Tamaño efectivo de la población $Ne = (2Nr-2) / (k-1) + (\sigma^2k / k)$ donde Nr = Hombres > 15 años y mujeres entre 15 y 49 ascendiente vivo entre 0 y 30 años, k = promedio de número de hijos por familia, σ^2k = varianza de el número de hijos por familia. (me) es la tasa de migración efectiva $(me) = [l(l+2s)] / 2$. (Arce-Gómez Magalhães and 1987). (l) = Proporción de migrantes a larga distancia y (s) = Proporción de migrantes a corta distancia $F = (1-m)^2 / 2N - (2N-1)(1-m)^2$.

Se estudió la dirección del flujo génico entre las parejas, teniendo en cuenta la clasificación genealógica de la persona muestreada y la autodeterminación del conyugue (Tabla 2). Se encontró que el 80% de los Wayúu establecen grupos familiares con individuos Wayúu, sin embargo a una proporción de parejas con Wayúu-Guajiros, Guajiros y hasta con mestizos de ancestría indefinida (múltiples orígenes no caracterizados por criterios genealógicos). A pesar de que definimos varios grupos genealógicos se puede ver cierto nivel de flujo entre ellos. Este flujo entre grupos hace consistente la definición de grupos intermedios (Wayúu-Guajiros y Ancestría Multiple).

En el análisis de las distancias geográficas entre parejas (Figura 4) se encontró que los Wayúu, Guajiros, Wayúu-Guajiros y los de Ancestría Múltiple forman pareja en general más cerca que los grupos migrantes. Los migrantes tienen una dispersión diferencial, indicando que los otros grupos se encuentran en una área geográfica más restringida dentro de la Guajira y coexisten en una misma región geográfica. Este análisis es importante porque da una idea del patrón de migración de los individuos. Los grupos migrantes divergen de los demás en la extensión del radio de migración al cual se puede encontrar su pool génico.

Genealogical Classification	/ Self-Determination	No. couples	Percentage
Wayuu Women	x Wayuu Men	21	40
Wayuu Men	x Wayuu Women	21	40
Wayuu Women	x I.A Men	9	17
Wayuu Men	x I.A Women	1	2
Guajira Women	x Wayuu Men	2	20
Guajira Women	x I.A Men	5	50
Guajira Women	x I.A Women	1	10
Guajira Women	x Afrocolombian Men	1	10
Guajira Men	x Afrocolombian Women	1	10
Wayuu-Guajira Women	x Wayuu Men	2	33
Wayuu-Guajira Men	x Wayuu Women	1	17
Wayuu-Guajira Women	x I.A Men	3	50

Tabla No 2 Tabla de Flujo Génico entre Parejas: Clasificación Genealógica únicamente para (Wayuu, Wayuu-Guajiro, Guajiro) de la persona muestreada con respecto a la autodeterminación en grupos étnicos de la pareja del encuestado I.A: Ancestría Indefinida.

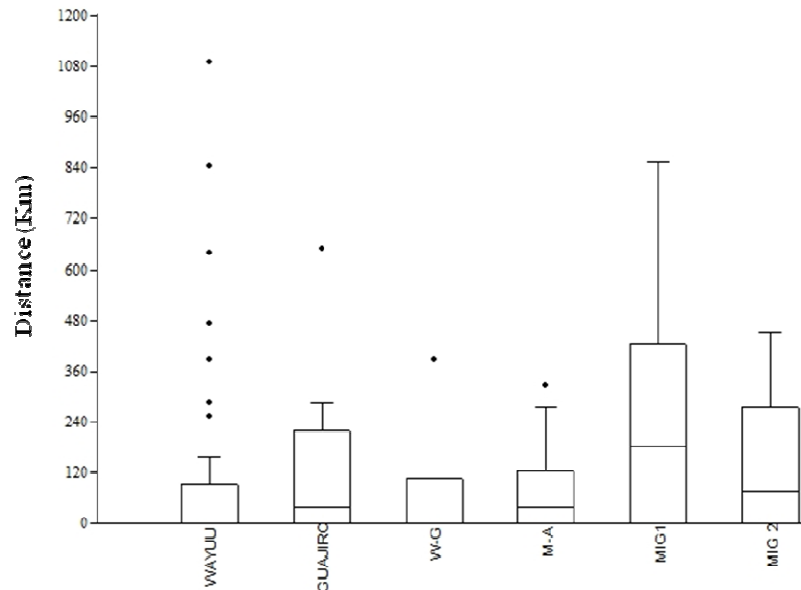


Figura No.4 Distancias Geográficas por parejas. W-G: Wayuu-Guajiro M.A: Ancestría Múltiple.

3.2 Análisis Genético de las Agrupaciones Genealógicas

Para la totalidad de la población se calcularon los estadísticos descriptivos de la población (Tabla 3). La heterocigosidad observada (H_o) total fue 0.773 y la heterocigosidad esperada (H_e) fue 0.780. Se calculó el H_o locus por locus tanto para la población total, como para cada agrupación genealógica. La menor heterocigosidad fue para el grupo Wayúu 0.727. La población Wayúu resultó ser una población con alta diversidad genética, sustentada por la heterocigosidad tanto por sistema como para el total.

Al evaluar el EHW dentro de los grupos genealógicos se encontraron todos en equilibrio: Wayúu con un p-valor de 0.3082, Wayúu-Guajiro 0.8780, Guajiro 0.3051, Ancestría Múltiple 0.8293, Migrantes-1 0.2754 y Migrantes-2 0.066, como se mostró anteriormente la heterocigosidad para estos grupos es alta. Al evaluar el EHW por locus se encontraron la mayoría de sistemas en equilibrio a excepción de los sistemas D8S1179 y el TH01. Como consecuencia se encontró un

desequilibrio global en la población total altamente significativo.

Para la población general se encontró un $Capf$ de 0.026 con un intervalo de confianza de 0.033 ± 0.01 , un Θ de 0.016 IC 0.021 ± 0.005 y un $Smallf$ de 0.011 IC 0.012 ± 0.008 (Tabla 3). Se obtuvo un Rst de 0.0167. Dentro de las agrupaciones se calculó el $Smallf$ para cada locus y global (Tabla 3). Basado en este resultado, se realizó un análisis de AMOVA con Fst (Tabla 4) evaluando tres hipótesis de subestructura: las seis agrupaciones en conjunto, Wayúu y el Grupo Externo y por último Wayúu y Wayúu-Guajiro con el grupo externo. Al evaluar la población general la mayoría de la varianza se encontró entre los individuos, con un Fit de 0.0262 y un p-valor de 0.0037 ± 0.0002 . Un Fst de 0.0105 significativo. Al evaluar la población Wayúu con respecto a los grupos externos, la mayor variación igualmente fue entre los individuos con un Fit de 0.00328. Aun así, se encontró un Fct de 0.01876 con un p-valor de 0.1659 ± 0.0012 y un Fsc de 0.0037 p-valor 0.0167 ± 0.0004 . Del mismo modo, en la agrupación Wayúu, Wayúu-Guajiro y en el grupo externo se encontró un Fct de 0.0193 p-valor 0.0671 ± 0.0008 y un Fsc de 0.003 p-valor de 0.0231 ± 0.0005 .

Se encontró que la agrupación (Wayúu-Wayúu-Guajiro) presentó el Fct mayor con un p-valor significativo, confirmando la hipótesis de 2 grupos.

LocName	All (N=290)					Wayuu (N=122)		Guajiro (N=25)		W-G (N=18)		M. A (N=44)		Mig1(N=48)		Mig2 (N=33)	
	Ho	Theta	Smallf	Capf	Rst	Ho	Smallf	Ho	Smallf	Ho	Smallf	Ho	Smallf	Ho	Smallf	Ho	Smallf
CSF1P0	0.684	0.004	0.071	0.075	0.023	0.699	0.12	0.788	0.187	0.734	0.167	0.759	-0.018	0.733	0.033	0.706	-0.073
D13S31	0.772	0.014	-0.004	0.01	0.008	0.772	-0.073	0.808	-0.089	0.757	0.192	0.822	0.032	0.801	0.09	0.804	0.02
D16S53	0.768	0.006	0.022	0.029	0.008	0.734	0.05	0.767	-0.043	0.792	-0.052	0.776	-0.113	0.785	0.045	0.794	0.16
D18S51	0.869	0.001	0.029	0.029	0.003	0.878	0.01	0.906	-0.06	0.894	-0.057	0.882	0.047	0.88	0.077	0.893	0.117
D19S43	0.87	0.009	-0.016	-0.007	0.006	0.832	0.045	0.844	0.005	0.801	-0.18	0.818	-0.084	0.866	-0.083	0.831	0.015
D21S11	0.82	0.015	0.003	0.019	0.003	0.784	-0.024	0.862	-0.021	0.851	-0.044	0.816	-0.002	0.817	0.057	0.823	0.079
D2S133	0.876	0.001	-0.013	-0.012	0.008	0.846	-0.008	0.889	0.1	0.853	-0.042	0.858	-0.086	0.88	0.006	0.877	-0.037
D3S135	0.727	0.045	-0.025	0.021	0.059	0.55	0.046	0.77	0.065	0.632	-0.054	0.71	-0.024	0.788	-0.189	0.772	-0.021
D5S818	0.702	0.036	0.033	0.068	0.081	0.661	0.069	0.71	0.099	0.665	0.081	0.784	-0.073	0.754	0.089	0.753	-0.086
D7S820	0.784	0.018	-0.017	0.002	0.046	0.685	0.019	0.794	0.043	0.76	-0.024	0.802	-0.02	0.723	0.02	0.801	-0.211
D8S117	0.781	0.003	0.012	0.015	0.003	0.817	-0.044	0.803	0.054	0.822	-0.014	0.812	0.132	0.799	0.062	0.765	-0.03
FGA	0.834	0.021	0.001	0.022	-0.006	0.818	-0.052	0.853	0.25	0.863	0.034	0.885	0.024	0.883	-0.015	0.889	-0.022
TH01	0.708	0.053	0.075	0.124	-0.008	0.599	0.165	0.784	-0.071	0.675	0.012	0.774	-0.058	0.783	0.149	0.783	0.033
TPOX	0.682	0.008	-0.017	-0.01	0.003	0.635	-0.007	0.737	0.023	0.587	-0.136	0.678	-0.14	0.683	0.085	0.655	-0.018
vWA	0.716	0.013	0.009	0.022	0.013	0.605	-0.003	0.792	-0.011	0.735	0.093	0.757	-0.081	0.742	0.13	0.731	-0.037
Overall	0.773	0.011	0.016	0.026	0.017	0.728	0.017	0.807	0.035	0.761	-0.002	0.796	-0.029	0.7945	0.035	0.792	-0.005
CI(Jk)	0.021+/-0.005		0.0120+/-0.080		0.0330+/-0.01												
CI(Bs 95%)	0.010+/-0.031		-0.002+/-0.028		0.016+/-0.054												
CI(Bs 99%)	0.010+/-0.035		-0.005+/-0.033		0.012+/-0.061												

Tabla No. 3 Diversidad y subestructura para las agrupaciones genealógicas: Se reporta la heterocigosidad observada (Ho) y el Smallf (Fis) (Weir & Cockerham, 1984) para los seis grupos genealógicos por sistema y global. Para la población total se reporta la Ho, los estadísticos-F y el Rst por locus y global. Para los estadísticos-F totales se calculó el Intervalo de confianza (CI) por Jackknifing (Jk), Bootstapping (Bs) al 95% y Bootstrapping al 99%.

Source of variation	Total			(Wayú/Remaining)			(Wayú+Wayú-Guajiro/Remaining)		
	% of variation	Stastics-F	P-valor	% of variation	Stastics-F	P-valor	% of variation	Stastics-F	P-valor
Among groups	–	–	–	1.88	Fct: 0.01876	0.16590+ -0.00126	1.93	Fct: 0.0193	0.0671+ -0.0008
Among populations within groups	1.59	Fst: 0.0105	0.0000+0.0000	0.36	Fsc:0.00368	0.01671+0.00040	0.29	Fsc:0.0030	0.0231+0.0005
Among individuals within populati	1.04	Fis: 0.0159	0.1067+ -0.0010	1.03	Fis:0.01052	0.10726+ -0.00097	1.03	Fis:0.0105	0.1079+ -0.0010
Within individuals	97.37	Fit: 0.0262	0.0037+0.0002	96.73	Fit: 0.00328	0.00326+0.00018	96.75	Fit: 0.0325	0.0036+0.0020

Tabla No. 4 Análisis Molecular de Varianza (AMOVA). Se calculan tres hipótesis de agrupamiento entre las poblaciones. Total: Todas en un solo grupo, Wayuu-Remaining: Wayuu en un grupo y las demás en otro y Wayuu+Wayuu-Guajiro/Remaining: Wayuu y Wayuu-Guajiro en un grupo y todo lo demás en otro grupo.

	Wayuu		Guajiro		Wayuu-Guajiro		Multiple-Ancestry		Migrants 1	
	Fst	P-valor	Fst	P-valor	Fst	P-valor	Fst	P-valor	Fst	P-valor
Wayuu	–	–								
Guajiro	0.0342	0.0000+0.0000	–	–						
W-G	0.0050	0.0991+ -0.0370	0.0083	0.0630+ -0.0237	–	–				
M.A	0.0243	0.0000+0.0000	0.0053	0.0540+ -0.0242	0.0057	0.0631+ -0.0305	–	–		
Migrants 1	0.0279	0.0000+0.0000	0.0032	0.1802+ -0.0383	0.0123	0.0000+0.0000	0.0017	0.1712+ -0.0212	–	–
Migrants 2	0.0201	0.0000+0.0000	0.0028	0.1441+ -0.0309	0.0003	0.4595+ -0.0417	0.0009	0.2883+ -0.0297	0.0012	0.2883+ -0.0227

Tabla No. 5 Matriz de comparaciones de Fst por pares de poblaciones

Se hizo una comparación por pares de poblaciones usando los Fst. Se observó una diferencia significativa entre Wayúu y los demás grupos, exceptuando el grupo Wayúu-Guajiro, confirmando la clasificación genealógica. Además, Wayúu-Guajiro mostró ser significativamente diferente al grupo Migrantes-1.

Al calcular las distancias genéticas se obtuvieron los mismos resultados. Usando la distancia de Cavalli-Sforza and Edwards (1967) (Figura 4) se agrupó a Wayúu con Wayúu-Guajiro. Este grupo es congruente con la naturaleza de los grupos de ancestría compartida. Además, se clasificó Guajiro, Ancestría Múltiple, Migrantes-1 y Migrantes-2 en un mismo agrupamiento, mostrando que no hay diferencias significativas entre estos grupos, por lo que las relaciones entre ellos no son claras.

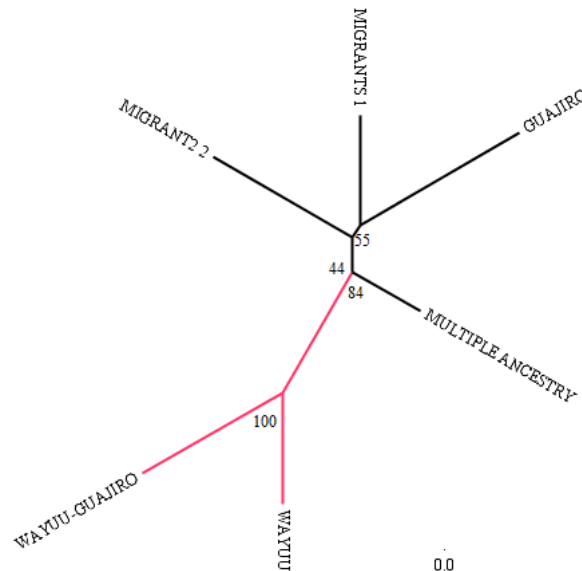
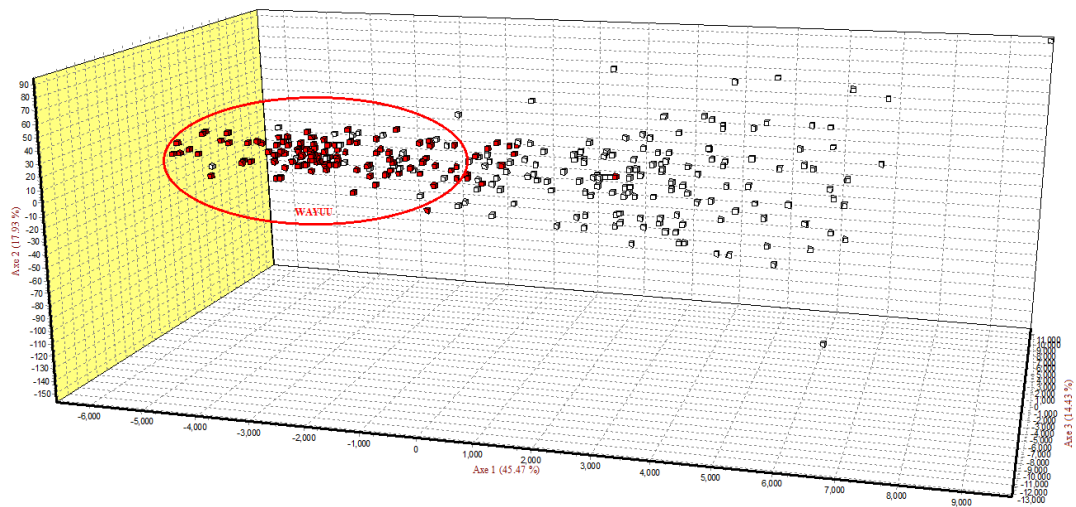


Figura No. 4 Árboles de distancias mostrando las relaciones entre grupos genealógicos Dc-NJ.

Se hizo un análisis factorial de correspondencias (AFC) como indicador de las diferencias entre individuos (Figura 5A) y los grupos (Figura 5B), en función de las variables genéticas evaluadas (Genetix 4.05) (Belkhir et al, 2001). El uso de esta herramienta es posible, dado a que la suma de los valores propios del AFC puede equipararse a los Fst de Robertson, Hill y Guinand (1996). Los cuatro

factores encontrados hicieron una contribución a la inercia total del 91.46%, dividida así (factor-1 45.47%, factor-2 17.93%, factor-3 14.43% y factor-4 13.59%). Se graficaron los tres primeros ejes mostrando un 63.4% de inercia. En el hiperespacio se observó una nube de puntos que muestra diferenciación de los individuos Wayúu, con respecto a los demás. Estas similitudes son causadas por las distancias entre las distribuciones de los alelos, reflejada en los aportes al centro de gravedad del análisis (Barton NH & Slatkin, 1986).

A)



B)

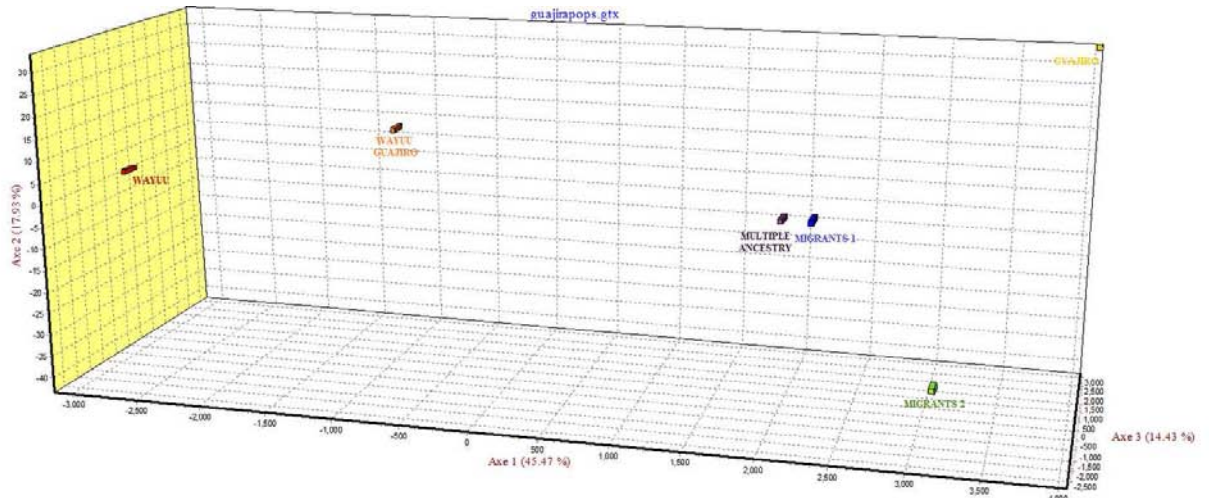


Figura No. 5 Plano Factorial ACM mostrando A: Relaciones entre individuos B: Relaciones entre poblaciones.

Se encontró en el análisis de asignación de individuos a poblaciones de Structure (Figura 6) que el $\ln P(D)$ más cercano a cero fue para un $K=2$ (-14836.6+/-16.46). Al analizar las agrupaciones genealógicas *a priori* se evidenció que el $K=2$ corresponde a una agrupación Wayúu y a las poblaciones (Guajiro, Migrantes-1, Migrantes-2 and Ancestría Múltiple). El valor de alfa de 0.2099 ($\alpha < 1$) mostró que en general los individuos tienen mayoritariamente ancestría de la subpoblación asignada. No obstante se puede identificar un nivel de flujo génico entre las dos subpoblaciones. Como se esperaba Wayúu-Guajiro mostró un gradiente de mezcla entre ambas agrupaciones por su naturaleza de genealogía mixta. El nivel de introgresión del grupo externo a la población Wayúu para toda la distribución de frecuencias alélicas fue de 0.0016+/- 0.0025.

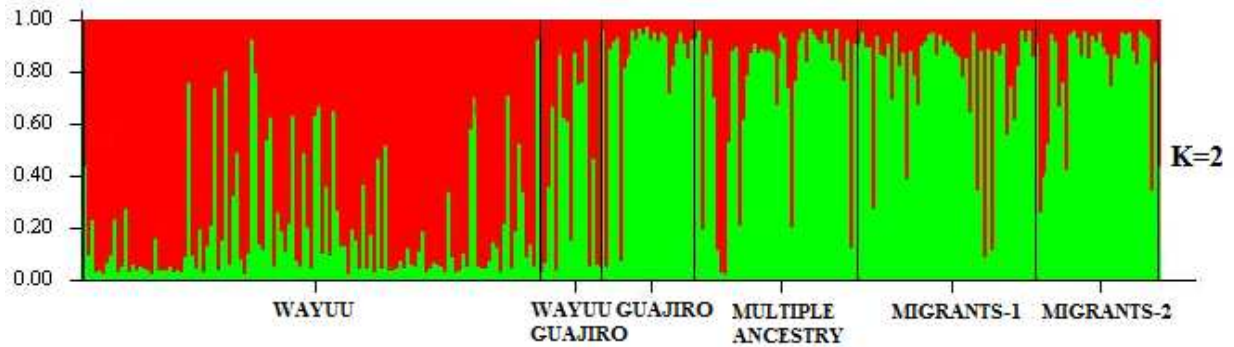


Figura No. 6 Bar Plot mostrando la estructura poblacional infiriendo estructura poblacional por asignación de individuos. *STRUCTURE*K=2.

En el estudio de la población Wayúu una variable importante, como se mencionaba anteriormente es la casta. Debido a que puede usarse como indicador de pertenencia a las familias con ancestría Wayúu. Estas se heredan de la madre a sus hijos, junto con la pertenencia de los Wayúu. Por otro lado es un indicador de mestizaje para los individuos que genealógicamente no son totalmente Wayúu, porque se pierde cuando es heredado por linaje paterno. Con base a esta información, el valor Q de asignación de individuos a poblaciones y la agrupación genealógica, se realizó un análisis de correspondencias múltiples. Se obtuvo un plano principal (Figura 7), donde las castas se asocian a las agrupaciones Wayúu y Wayúu-Guajiro, los grupos con los valores Q más altos encontrados por Structure. Esta asociación se presenta porque los estados de las variables se presentan en los mismos individuos, en ese sentido son similares entre sí.

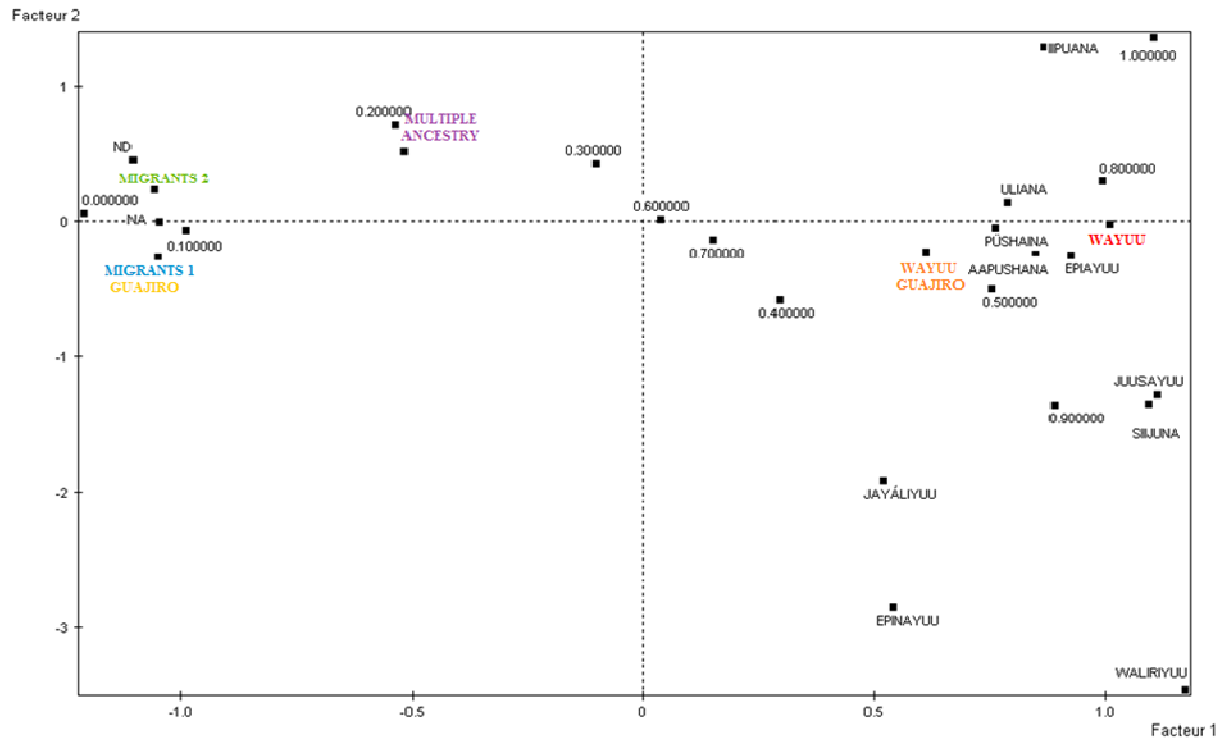


Figura No 7 Grafica de correspondencias múltiples, variables: Castas, Valor Q (Structure), Grupos Genealógicos

3.3 Análisis Genético de las Agrupaciones Genéticas

Dado que ya se evidenció la presencia de dos individuos, se procedió a la caracterización de las subpoblaciones. Primero que todo, se retiraron del análisis los individuos de la agrupación Wayú-Guajiro, con el fin de dilucidar las diferencias entre grupos e identificar unas posibles poblaciones parentales. Otra causa para retirarlo del análisis fue el pequeño tamaño de muestra.

Al evaluar el EHW por locus se encontró igualmente los sistemas D8S1179 p-valor de 0.0432, TH01 0.0110 y el sistema CSF1P0 0.0224 en desequilibrio. También se encontró la población en desequilibrio con un P-valor de 0.0241. En las subpoblaciones se encontró la población Wayú en

equilibrio con un P-valor de 0.2543 y la población externa en desequilibrio con un P-valor de 0.0164. Por otro lado, al asumir dos poblaciones (Tabla 6) se encontró un Capf de 0.036 con IC de 0.036+/- 0.0101. El valor de Theta fue 0.024 CI 0.023 +/- 0.006 y el valor de de Smallf fue 0.012 IC 0.012+/- 0.008(Tabla 6). Estos valores son mayores, aunque los intervalos de confianza se sobrelapan con los de los seis grupos. El valor del Rst fue 0.0254.

LocName	All			Wayuu External Group		
	Capf	Theta	Smallf	Rst	Smallf	Smallf
CSF1P0	0.073	0.0090	0.065	0.044	0.12	0.023
D13S31	0.006	0.0230	-0.017	0.009	-0.073	0.026
D16S53	0.037	0.0080	0.03	0.014	0.05	0.015
D18S51	0.036	0.0030	0.033	0	0.01	0.052
D19S43	0.009	0.0150	-0.006	0.01	0.045	-0.047
D21S11	0.03	0.0210	0.009	0.01	-0.024	0.034
D2S133	-0.009	0.0040	-0.012	-0.003	-0.008	-0.016
D3S135	0.043	0.0640	-0.023	0.093	0.046	-0.063
D5S818	0.084	0.0540	0.032	0.113	0.069	0.006
D7S820	0.011	0.0220	-0.011	0.066	0.019	-0.033
D8S117	0.018	0.0040	0.014	0.019	-0.044	0.063
FGA	0.029	0.0240	0.004	0.001	-0.052	0.046
TH01	0.152	0.0760	0.082	-0.004	0.165	0.031
TPOX	0.001	0.0110	-0.01	0.014	-0.007	-0.011
vWA	0.026	0.0260	0	-0.004	-0.003	0.002
Overall	0.036	0.024	0.012	0.0254	0.017	0.009
CI(Jk)	0.036+-0.010	0.023+-0.006	0.012+-0.008			
CI(Bs 95%)	0.019+-0.057	0.014+-0.035	-0.001+-0.028			
CI(Bs 99%)	0.015+-0.065	0.011+-0.039	-0.022+-0.074			

Tabla No. 6 Diversidad y subestructura para los dos grupos genéticos: Se reporta el Smallf (Fis) (Weir & Cockerham, 1984) para las dos subpoblaciones, retirando los datos para la agrupación Wayuu-Guajira. Para la población total se reportan los estadísticos-F y el Rst por locus y global. Los intervalos de confianza (CI) por Jackknifing (Jk), Bootstapping (Bs)

Finalmente se evaluó la migración entre los dos grupos, calculando el número de migrantes basado en los alelos privados de las dos subpoblaciones, la estimación de las tasas de migración reciente y un

cálculo de asignación de individuos a poblaciones y a estados temporales de migración.

Con base a una frecuencia promedio de los alelos privados de 0.008, se estimó un número de migrantes de 9.5 para la población total. La estimación de las tasas de migración reciente en BayesAss reveló una potencial introgresión de los alelos externos en la población Wayúu ($m=0.1041$; $DS=0.024$) con un intervalo de confianza al 95% (0.062976- 0.156755), mientras que la tasa de no-migración de ese grupo externo fue de ($m=0.960279$; $DS=0.016616$) con un IC (0.9223- 0.987521). Adicionalmente, la tasa de migración de los individuos Wayuu hacia la población externa fue ($m=0.0397$; $DS=0.0166$) con un IC (0.0124, 0.0776) y dentro de la población la tasa de no-migración ($m=0.960279$; $DS 0.016$) con un IC de (0.9223-0.9875). En el cálculo de asignación de individuos a las dos poblaciones y a los siguientes posibles estados temporales (No migrante, migrante y descendiente de migrante), representado en un barplot (Figura 8), se muestra como efectivamente la migración hacia la población externa es mayor que hacia la población Wayúu y que la mayor migración se dio en individuos descendientes de la población hacia poblaciones externas y los migrantes están especialmente categorizados como individuos de ancestría múltiple y guajiros.

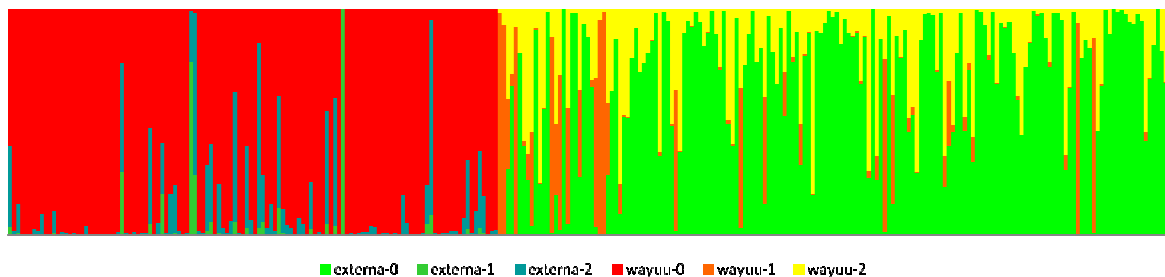


Figura No. 8 Barplot con la asignación de individuos a poblaciones con la estimación de 0: no migrantes, 1: migrante y 2: descendientes de migrantes.

4. Discusión.

En la Guajira, al igual que en otras poblacionales americanas, la dinámica poblacional es compleja debido a su estrepitosa historia de poblamiento, a los múltiples orígenes de sus pueblos y a los recientes hechos de conquista y colonización (500 años un periodo corto en la escala genética).

El Caribe colombiano en particular es una clara muestra de ello. Es una región, cuyos pobladores ancestrales corresponden a comunidades indígenas americanas, las cuales previamente ya habían vivido diferentes guerras, procesos de migración, epidemias y además habían construido complejos procesos sociales y culturales; de los cuales solo tenemos una fracción de la información. Además es un lugar de arribo para migrantes de todas partes del mundo e influenciado por la esclavitud de los siglos XV al XVIII. Tanto la cultura indígena actual como la población residente actual, tiende a conservar muchos de esos elementos, en una mezcla de tradiciones ancestrales, elementos europeos de la conquista y de la colonia y del proceso actual de globalización.

Esta dinámica poblacional esta comenzándose a estudiar en los estudios genético poblacionales. Tradicionalmente estos estudios en Colombia, se han limitado a un número restringido de muestras, tomadas de las comunidades y en su selección se han basado principalmente en criterios de autodeterminación. Si bien en este tipo de estudios los muestreos son por conveniencia, se ha caído en la tendencia de realizarlos sin el mayor rigor metodológico, ni matemático. Nuestro estudio, no se diseñó a criterio del investigador, fue específico para la población Wayúu. Diseñado desde un abordaje tanto genealógico-demográfico como genético. Esto con el propósito de realizar un cuidadoso y riguroso examen de la población.

Estos diseños redundan en mejores tamaños de muestra. En este caso se tomaron 290 personas, de las cuáles 122 son Wayúu de las cuales se tiene certeza de su pertenencia por los criterios genealógicos. Más aún esta es la muestra de comunidades indígenas más grande en el país y no estuvo focalizada en un solo lugar de muestreo, sino que se desarrolló en una amplia región geográfica del departamento de la Guajira. Aun así, cuando fragmentamos la muestra se redujo el tamaño muestral dentro de las

agrupaciones. Pero la selección cuidadosa de los individuos permitió compensar este efecto arrojando resultados significativos.

La encuesta fue útil para el refinamiento del muestreo por conveniencia y fue una herramienta práctica de toma de datos que permitió tener un conocimiento integral del participante (reconstrucción genealógica, información demográfica y de pertenencia étnica). Pero en realidad, esta herramienta permitió definir clasificaciones genealógicas para el estudio de la subestructura y de la dinámica migracional, que reflejan las hipótesis de trabajo del estudio. Estas agrupaciones se evaluaron desde la demografía y desde la genética para estudiar el efecto de la estructura de sexos y edades, el tamaño poblacional, del aislamiento reproductivo, la formación de parejas de modo no aleatorio y la endogamia. Factores insoslayables cuando se estudia la estructura genética de una población (Acreche N. 2004).

La evaluación de parámetros demográficos permitió describir en este sentido la dinámica de crecimiento, el efecto del tamaño reproductivo y la migración en la población sobre el efecto neto de la deriva genética. Además de dar una idea del nivel de flujo génico entre los grupos analizados.

La distribución de edad y sexo mostró que la muestra se asemeja a una pirámide poblacional característica de una población donde la mayoría del pool génico se encuentra en las generaciones jóvenes en edad reproductiva. El 51% de los individuos del tamaño reproductivo de la muestra se encontraron entre 20-38 años. La tasa de fecundidad de 2.87 hijos por mujer es consistente, indicando que la población se encuentra en crecimiento superando la tasa de sustitución de la misma (2 hijos por mujer) (CIA, 2011). Encontrar poblaciones con la mayoría de su pool génico concentrado en los grupos etarios jóvenes, en crecimiento y con un activo proceso reproductivo, podría hacerla sensible al efecto de la deriva genética, generando posibles procesos estocásticos de pérdida y ganancia de variantes alélicas. Percibimos que varias características culturales de la población tienen como consecuencia el activo proceso de reproducción. Sin embargo, esto no implica que la población se encuentre en expansión. Debido a factores compensadores como la alta tasa de mortalidad infantil (38.4 muertes de menores de 5 años por cada mil nacidos vivos) (DANE, 2005) y el efecto del

ecosistema en el crecimiento poblacional. Por lo tanto, 10.000 años de asentamiento en la zona no ha generado una expansión exponencial. Posiblemente el ecosistema de desierto tiene una capacidad de carga menor por la escases de recursos. El DANE reporta al respecto, que el 70% de la población Wayúu tiene las necesidades básicas insatisfechas, en general por la carencia de agua y el debilitamiento de la economía tradicionalmente Wayúu.

Los factores evaluados anteriormente afectan también el tamaño efectivo poblacional dado a que involucra tanto al tamaño reproductivo como al número de hijos por familia. Encontramos para la población total un N_r de 53.4% del N (290) y un N_e del 15.9%) y para la población Wayúu (N_r el 45.9% y N_e del 13.5%). Comparado con otros estudios: Magalhaës y Arce Gómez en 1987 de la población Brazileria aislada de Guaraqueaba, se encontró un N_r menor de 36.6% y un N_e similar al Wayúu de tan solo del 13.2%. Se estiman que para la mayoría de poblaciones humanas el tamaño efectivo es del 25% del N total. Por lo que los valores encontrados parecen ser pequeños, posiblemente afectados por el hecho de no haber tomado individuos menores de 18 años y de no considerar la superposición de generaciones.

Como se mencionaba anteriormente, suponemos que hay un activo proceso de migración hacia los Wayúu de los demás grupos genealógicos. Estas categorías permitieron el estudio de la migración de grupos consistente con el movimiento de diferentes linajes en el territorio. De esta manera los migrantes a larga distancia (61.3%) se tomaron como los individuos con genealogías de lugares de nacimiento fuera de la región geográfica. Mientras que los migrantes a corta distancia fueron los descendientes de Wayúu o Guajiros y migrantes. Esta ordenación, permitió evaluar el movimiento de las variantes alélicas de diferentes genealogías y que caracterizan la subestructura actual.

La tasa de migración junto al tamaño efectivo de la población, permitió evaluar el coeficiente de aislamiento reproductivo, para medir el accionar de la deriva genética. EL valor de 11.9 indicó una moderada sujeción de la deriva genética, posiblemente contrarrestada por el flujo génico de la población. Adicionalmente se encontró un coeficiente de endogamia de 0.0129 lo que indica que no es una población endogámica. En poblaciones humanas estas consideraciones tienen sentido dado a

que presentan características sociales y distribuciones en el espacio que hacen que el sistema de apareamiento no sea al azar, formando “barreras” reproductivas que afectan el flujo génico entre grupos, pero no necesariamente producen endogamia dentro de las poblaciones.

Las diferencias de flujo génico y el efecto de la deriva genética, también están influenciados por la formación de parejas de modo no aleatorio entre grupos en los diferentes espacios geográficos. Los análisis de flujo genético y distancias entre parejas, mostraron que los Wayúu establecen grupos familiares mayoritariamente con personas de la comunidad y que se encuentran a una distancia geográfica menor a comparación de otros grupos (un radio máximo de migración de 290 Km) Es así como la tendencia de los individuos nativos, es a establecerse dentro de las poblaciones en vez migrar, esto producido principalmente por aspectos culturales. Como consecuencia, el flujo génico entre grupos genealógicos es moderado a pesar de que coexistan en las mismas regiones geográficas (51,9% de personas viven en las cabeceras municipales).

La caracterización de la subestructura y los patrones de migración se complementó con el análisis genético de 15 STRs autosómicos. En general, estos permitieron hacer una buena aproximación al modelo de subestructura genética y a los estimadores de diversidad genética de la población.

En primera instancia se encontró que los Wayúu ($H_o = .727$) así como los demás grupos genealógicos presentaron niveles de diversidad genética altos. El hecho de que tenga alta diversidad en comparación con otras poblaciones amerindias Colombianas como Emberá 0.655, Ingano 0.687, Ticuna 0.682 y Zenu 0.666 (Mesa NR et al, 2000), puede ser explicado por los procesos sociales de matrimonios entre personas de diferentes castas o personas externas a la comunidad Wayuu; dinámica que subsecuentemente reduce la endogamia de la población. Adicionalmente el encontrar desequilibrio en la población global posiblemente fue causado por la presencia de diferentes distribuciones genotípicas en la muestra de poblaciones diferentes, dando un inicio de que no hay una sola población en la muestra.

No se encontró un coeficiente de endogamia alto (0.011 con un IC de 0.012+/-0.008) al encontrado

con datos demográficos. Por otro lado, interpretando los estadísticos-F la población tiene una moderada diferenciación genética, pero significativa entre grupos. Comparado con una mayor variación entre individuos. El valor de Theta total (0.016 con IC 0.021+/- 0.005) y por marcador mostró que hay por lo menos un grupo que genera diferenciación genética. El hecho de que los Wayúu presenten un grado de subestructura moderado, no sea una población endogámica y presente alta variabilidad, muestra que la diferenciación puede ser causada por diferencias en su distribución genotípica, con respecto a las poblaciones externas.

La subestructura genética fue corroborada por el AMOVA donde las agrupaciones: Wayúu y (Wayúu, Wayúu-Guajiro) diferían significativamente de las demás agrupaciones genealógicas. Es así como los resultados indicaron la presencia de dos grupos: los Wayuu junto con el grupo intermedio Wayuu-Guajiro y el segundo compuesto por los grupos migrantes, guajiros y el grupo de ancestría múltiple. Estos grupos intermedios son producto del flujo génico entre las diferentes agrupaciones. Por otro lado, la evaluación de la distancia genética, mostró los mismos agrupamientos. Sin embargo es de resaltar, el clúster Wayúu y Wayúu-Guajiro consistente. Finalmente el ACM, también mostró la misma tendencia en los individuos. Y en efecto, se nota claramente que los grupos intermedios no son homogéneos, dado que los individuos Wayúu-Guajiros están más cercanos a Wayúu mientras que los individuos de ancestría múltiple se muestran cercanos a los Migrantes-1.

Para validar el modelo de subestructura genética *a priori* basado en las agrupaciones genealógicas, se evaluó el análisis de asignación de individuos a poblaciones utilizando una distribución *a posteriori* basada en un análisis bayesiano sobre la información genotípica de los individuos. Se encontró que la población se ajusta mejor a un $K=2$ donde Wayúu forma un núcleo diferente al de los demás grupos. Al calcular este nivel de introgresión a la población Wayúu desde los demás grupos se encontró 0.16%. El ACM muestra en general un gradiente de mestizaje desde Wayúu (con los mayores valores Q) hacia los migrantes. Las castas se asociaron tanto a Wayúu como a Wayúu-Guajiro, mostrando que como variable característica de la población, se asocia directamente a los individuos Wayúu, ratificando su pertenencia a los clanes matriarcales y por ende afirma su ancestría. El análisis igualmente permitió dilucidar las diferencias entre los grupos intermedios si se evalúan con respecto a

una población Wayúu parental.

Sobre la subestructura definida, se encontró igualmente un DHW global, pero al obtener desequilibrio en el grupo externo asumimos que no hay homología entre ellos. Al asumir las dos subpoblaciones se recalcularon los estadísticos-F, aumentando el valor Theta a 0.024 con un IC de 0.023. Igualmente el R_{st} aumento a 0.0254. Estos estadísticos mostraron mayor diferenciación entre los grupos sobre la diferencia entre individuos que el que se presentaba con las seis agrupaciones. Finalmente, con el fin de complementar el modelo de dinámica migracional de la población Wayúu que se planteó con demografía, con los datos genéticos se encontró que el número de migrantes basado en alelos privados entre las dos subpoblaciones fue 9.5. Este número es consistente dado que en este modelo de subestructura los Wayúu son una población diferencial con niveles bajos de introgresión en términos del movimiento de alelos. Sin embargo según el análisis de las tasas de migración reciente, muestra que esta pequeña introgresión ($m=0.1041$ $DS=0.0240$) es mayor que la de Wayúu hacia la población externa ($m=0.0397$ $DS=0.0166$), este resultado fue corroborado por la asignación de individuos a las subpoblaciones y a los estados de migración. La migración detectada de la población Wayúu es mayoritariamente de descendientes de Wayúu. Este resultado muestra la dirección de flujo génico entre las dos subpoblaciones, posiblemente generado por la aculturación de la población indígena. Sin embargo, como se mencionaba en el análisis de flujo génico entre parejas la tendencia es a mantenerse dentro de cada subpoblación.

6. Conclusión

En esta investigación, en resumen queremos resaltar que las categorías de clasificación de individuos son primordiales al diseñar el muestreo para detectar la subestructura poblacional. Por lo que al decidir si y cómo utilizar las categorías raciales, étnicas y ancestrales en la investigación, los genetistas nos enfrentamos a escenarios contradictorios. Muchos investigadores han propuesto fuertes argumentos a favor de reducir o incluso eliminar el uso de categorías raciales o étnicas en la investigación genética (Boyer et al 2005), dado que clasificaciones tradicionales que acuñen a los términos de mestizo, europeo y afro-descendiente son imprecisas. Los análisis de poblaciones

humanas requieren de clasificaciones más robustas y cuidadosamente realizadas que permitan hacer inferencias mucho más finas y adecuadas que correspondan a criterios que den cuenta de la dinámica de poblamiento y a los procesos históricos específicos de la población.

La definición de un modelo de clasificación fenotipo- específico de cada uno de los grupos humanos utilizando categorías raciales, para nosotros carece de sentido, dado que encontraremos en los procesos de constitución y de poblamiento una amplia gama de matices y ancestrías propias de poblaciones humanas, como es el este caso de las americanas, en donde se reúnen múltiples ancestrías, que obedecen más a los diferentes procesos de migración que a la constitución de nuevos grupos humanos o a la mezcla de grupos étnicos tradicionales. La clasificación genealógica realizada en este estudio, permite diferenciar a los Wayuu, que puede definirse como una población parental en estudios posteriores. Por ende sugerimos el uso de categorías genealógicas o basadas en características históricas o sociales, específicas para responder las hipótesis en los diferentes escenarios de cada estudio.

Se encontró que la población de la Guajira es una población en términos de su estructura de edad, joven y se encuentra en crecimiento. Además se encuentra bajo un efecto moderado de deriva genética, la población no es homogénea conforme a su historia de poblamiento y composición étnica. Existe subestructura poblacional diferenciando a los individuos Wayúu, de los residentes y migrantes tanto Caribes como de otras regiones de Colombia, esta población indígena mostro ser una población conservada, a pesar de que coexiste y presenta un flujo migracional bidireccional con los demás grupos poblacionales, esta dinámica migracional está representada en grupos de ancestría múltiple con diferencias en su tipo de ascendente Wayúu (Wayúu-Guajiro), donde no todos los individuos se autodeterminan como pertenecientes a la comunidad por razones culturales. No obstante, no se presentan diferencias significativas con el grupo Wayúu sustentando este dato con la información genealógica. Esta población es un ejemplo de un proceso de subestructura a nivel local que en su mayoría no es causado por aislamiento por distancia, si no por diferencias culturales y de parentesco que generan posibles barreras reproductivas. El hecho de que a nivel local se encuentren poblaciones aisladas es importante en estudios posteriores de la etiología de algunas enfermedades y sus

implicaciones en los aspectos prioritarios de salud pública.

7. Agradecimientos

En primer lugar a los habitantes de la población de la Guajira y a la Comunidad Wayuu, por permitirnos conocer un poco más acerca de su historia biológica, así como también a las diferentes instituciones de salud: al Hospital Nuestra Señora del Perpetuo Socorro, al Hospital San José de Maicao, al Hospital Nuestra Señora de los Remedios y a la IPS CAPRECOM quienes prestaron su colaboración en el muestreo. En segundo lugar, a la Dirección de Investigación (DIB) de la Universidad Nacional de Colombia por brindarnos los recursos financieros para realizar el estudio, al Instituto de Genética de la Universidad Nacional y finalmente a los miembros del Grupo de Genética de Poblaciones e Identificación por su apoyo e ideas en la realización de este estudio.

8. Bibliografía

- Albeza MV, Acreche NE, Caruso GB. 2004. Biogeografía en poblaciones de la Albeza MV et al, 2004 (Chañarcito, Santa Rosa de los Pastos Grandes y Olacapato) Salta, Argentina. *Revista de Antropología Chilena* 34 (1): 119-126.
- Arango R, Sánchez E. Los pueblos Indígenas de Colombia en el Umbral del Nuevo Milenio. Departamento Nacional de Planeación de Colombia. Dirección de Desarrollo Territorial Sostenible. Bogotá, 2004.
- Ardila G, Cuando el progreso se enfrenta a la vida: los Wayúu de la Guajira. *Diversidad es Riqueza, ensayos sobre la realidad colombiana*. ICAN, Instituto Colombiano de Cultura y Consejería Presidencial para los Derecho Humanos. Bogotá, 1992.
- Barton NH, and Slatkin M. 1986. A quasi-equilibrium theory of the distribution of rare alleles in a subdivided population. *Heredity* 56 (3):409-15.
- Belkhir K, Borsa P, Chikhi L, Raufaste N, and Bonhomme F. 2001. GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome Populations Interactions.

- CELADE (Comisión Económica para América Latina y el Caribe). Perfil Socio-demográfico básico de la Guajira. www.eclac.cl/celade/noticias/paginas/2/40392/1_La_Guajira.pdf 15 de Noviembre de 2011.
- CIA, 2011 The World Factbook. <https://www.cia.gov/library/publications/the-world-factbook/> 20 de Septiembre de 2011.
- DANE (Departamento Administrativo Nacional de Estadística). Censo Nacional 2005 Nivel Nacional. Dirección de Difusión, Mercadeo y Cultura Estadística. Bogotá.
- Goudet J. 1995. FSTAT (Version 1.2): A computer program to calculate F-statistics. *Journal of Heredity* 86:485-486.
- Gobernación de la Guajira, Pagina web oficial de La Guajira. <http://dev.laguajira.gov.co/joomla/> consultado 10 de septiembre de 2011
- Keyeux G, William UM, Rutas Migratorias hacia Suramérica y poblamiento de las cuencas del rio Amazonas y Orinoco, deducidas a partir de estudios genético moleculares. *Pueblos y paisajes antiguos de la selva amazónica*. Pág. 43-62
- Langella O (2002) populations 1.2.28. Population Genetic Software (Individuals or Populations Distances, Phylogenetic Trees). <http://www.cnrs-gif.fr/pge/bioinfo/populations/index.php>
- Magalhães JC, Arce-Gomez B. 1987 Study on a Brazilian isolate. I. Population structure and random genetic drift. *Hum. Hered.* 37 (5):278-84.
- Mesa NR, Mondragón MC, Soto ID, Parra MV, Duque C, Ortíz-Barrientos D, García LF, Vélez ID, Bravo ML, Múnera JG, et al. 2000. Autosomal, mtDNA, and Y-chromosome diversity in Amerinds: pre- and post-Columbian patterns of gene flow in South America. *American journal of human genetics* 67:1277-86.
- Oliver JR, Reflexiones sobre los posibles orígenes del Wayúu (Guajiro). In: *La Guajira: de la memoria al porvenir. Una visión antropológica*. 1991. Editorial Fondo FEN. 83-135). Bogotá, Colombia.
- Pritchard JK, Stephens M, and Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945-959.
- Programa Presidencial de Derechos Humanos. Diagnóstico de la situación del pueblo indígena

Wayúu.

http://www.derechoshumanos.gov.co/Observatorio/documents/2010/DiagnosticoIndigenas/Diagnostico_WAY%C3%9AU.pdf 15 de Septiembre de 2011.

Profamilia. Encuesta nacional de demografía y salud. (2007) Bogotá.

Pérez L. Los Wayúu: Tiempos, Espacios y Circunstancias. Cuaderno Venezolano de Sociología 15:403-426

Raymond M. & Rousset F, 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *J. Heredity*, 86:248-249.

Rousset, F., 2008. Genepop: a complete reimplementación of the Genepop software for Windows and Linux. *Mol. Ecol. Resources* 8: 103-106.

Schneider S, Roessli D, and Excoffier L. 2000. Arlequin: A software for population genetics data analysis. User manual ver 2:2496–2497.

Slatkin M, 1995. A measure of population subdivision based on microsatellite allele frequencies. 1995 *Genetics* 139(1):457-462.

Vásquez S., Correa H. Los Wayuu, entre Juya (“El que llueve”), Mma (“la tierra”) y el desarrollo urbano regional. *Geografía Humana de Colombia, Nordeste Indígena Tomo II*. <http://www.banrepcultural.org/blaavirtual/geografia/geograf2/wayuu1.htm> consultado el 10 de septiembre de 2011.

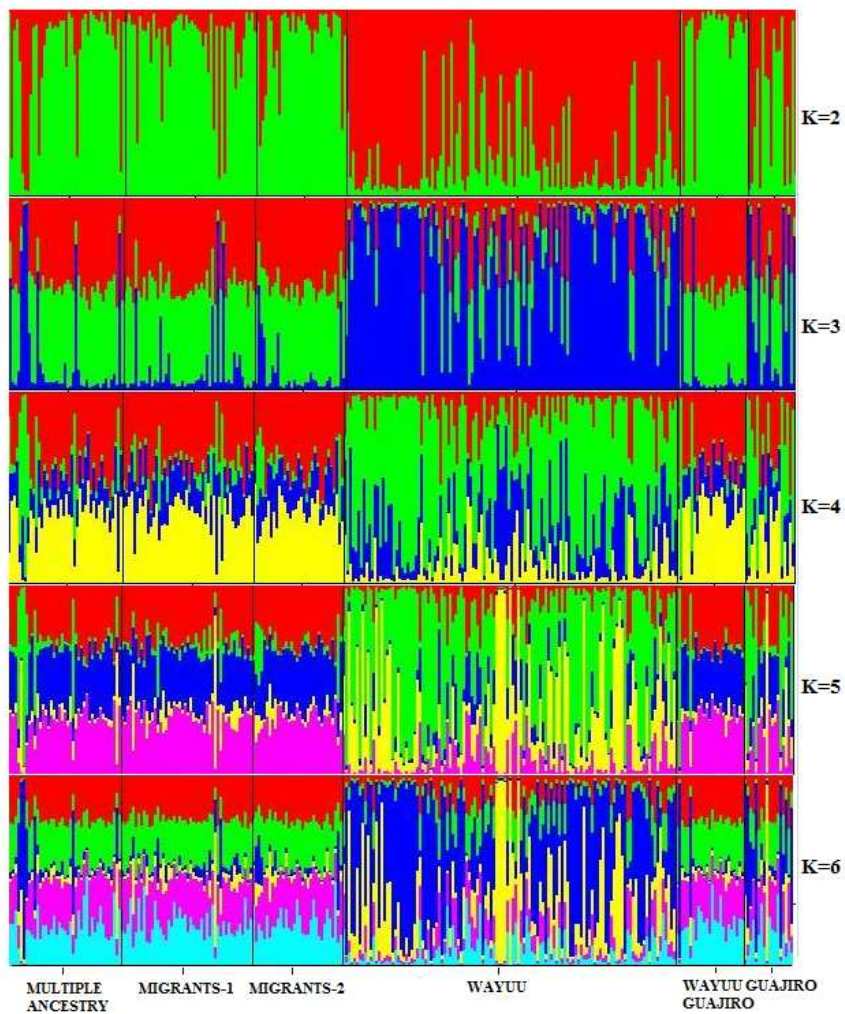
Wilson GA, Rannala B. 2003 Bayesian inference of recent migration rates using multilocus genotypes. *Genetics* 163:1177-1191.

Weir B.S, Cockerham C.C. 1984. Estimating F-statistics for the analysis of population structure. *Evolution*, 38(6):1358:1370.

8. Anexos.

K	Ln P(D)	Var(LnP(D))	α
1	-15150.0	72.0	-
1	-15150.7	72.7	-
2	-14837.7	269.7	0.1997
2	-14836.6	271.2	0.2099
3	-15306.8	1271.3	0.0827
3	-15010.6	627.1	0.0811
4	-15391.7	1574.6	0.0638
4	-15439.2	1635.4	0.0600
5	-15725.6	2382.2	0.0631
5	-15483.1	1972.0	0.0661
6	-16060.2	3103.5	0.0620
6	-15954.6	2956.8	0.0613

Anexo No 1: Tabla de los valores de Ln P(D) para dos repeticiones del análisis mostrando K=1-6 y los valores de varianza y alfa para cada corrida de datos.



Anexo No 3. Grafico de Barras mostrando la estructura poblacional inferida por asignación de individuos a poblaciones STRUCTURE software K=2, 3, 4, 5, 6.

Number	Value	Percentage	Percentage	
1	0.9	10.1	10.1	*****
2	0.5	5.4	15.5	*****
3	0.5	5.4	20.8	*****
4	0.4	4.9	25.8	*****
5	0.4	4.7	30.5	*****
6	0.4	4.6	35.1	*****
7	0.4	4.5	39.6	*****
8	0.4	4.4	44.0	*****
9	0.4	4.2	48.2	*****
10	0.4	4.1	52.3	*****
11	0.4	4.1	56.4	*****
12	0.3	3.9	60.3	*****
13	0.3	3.9	64.1	*****
14	0.3	3.8	68.0	*****
15	0.3	3.7	71.6	*****
16	0.3	3.6	75.2	*****
17	0.3	3.5	78.7	*****
18	0.3	3.3	82.0	*****
19	0.3	3.2	85.2	*****
20	0.3	3.1	88.3	*****
21	0.3	2.9	91.2	*****
22	0.2	2.7	93.9	*****
23	0.2	2.6	96.4	*****
24	0.2	2.3	98.8	*****
25	0.1	0.9	99.6	*****
26	0.0	0.4	100.0	****

Anexo No 4. Histograma de los 26 primeros valores propios del ACM análisis de correspondencias múltiples.



PROJECT: "GENETIC CHARACTERIZATION IN A HUMAN POPULATION FROM THE GUAJIRA DEPARTMENT: CONTRIBUTION TO HISTORICAL AND STRUCTURAL KNOWLEDGE ABOUT COLOMBIAN ETHNIC GROUPS"

1.

Date	DD	June	2009
Sample code			
Sampling Institution			
Sampling officer			
Pollster name			

2. Sampling site: Municipality _____ Specific site _____

3. Personal Information:

Residence Place			Residence Time	
Address		Telephone Number	E-mail	
First name		Middle Name		

4. Surname:

P1 _____ P2 _____ M1 _____ M2 _____

6.

Ethnicity	Caste	Ran cheria	
Age of Mother at first birth	Offspring number		
Wayunaiki	Talking	Yes	Not
	Write	Yes	Not

5. Age LN LN

Anexo No5. Encuesta Genealógica y Demográfica aplicada en el muestreo en el departamento de La Guajira. P1-2 primero y segundo apellidos paternos, M1- primero y segundo apellidos maternos. POP: Población de

Capítulo No. 4: Análisis de apellidos y de cromosoma Y de las poblaciones raizales del archipiélago de San Andrés y Providencia (Colombia)

Ángela Alonso Morales; William Usaquén Martínez. Instituto de Genética. Universidad Nacional de Colombia Sede Bogotá. Bogotá-C. laalonsom@unal.edu.co. Correspondencia: Ciudad Universitaria, Avenida carrera 30 N.º 45-03, Instituto de Genética, oficina 03. Edificio 426. Bogotá, Colombia. wusaquenm@unal.edu.co

Resumen

El archipiélago de San Andrés y Providencia es un departamento de Colombia aislado en las aguas occidentales del mar Caribe. La mayor parte de sus habitantes corresponde al grupo afrocolombiano antillano denominado "raizal", el cual cuenta con características culturales únicas como producto de una mezcla centenaria de diversos orígenes principalmente esclavos africanos y colonos europeos. En la actualidad poco se conoce acerca de la composición y diversidad genética de estas poblaciones. Por ello, el presente estudio pretende investigar la variación y estructura de la población raizal del Archipiélago. Un total de 54 individuos fueron seleccionados mediante un criterio genealógico de tres generaciones de ancestros nacidos en las islas. Se analizaron 17 Y-STRs, incluyendo la información heredada en el primer apellido. De este análisis se encontró una diversidad haplotípica baja en ambas islas, en comparación a lo reportado en otras poblaciones afro descendientes. Se utilizó un análisis molecular de la varianza para evaluar la hipótesis de subestructura genética en el archipiélago, en donde no se encontraron diferencias significativas ($p > 0.05$). Se predijeron los haplogrupos de cromosoma Y presentes en las islas, siendo E1b1a y R1b los más frecuentes. Estos representan más del 80% de la muestra y son los haplogrupos característicos de poblaciones africanas y europeas; correspondiendo con la historia de poblamiento de las islas. Para análisis genéticos comparativos se calcularon distancias genéticas con respecto a otras poblaciones del Caribe, Europeas, Asiáticas, Africanas y Colombianas. Se observó

concordancia entre las poblaciones afrodescendientes en relación a las demás poblaciones evaluadas. El análisis de apellidos mostró que los raizales que tienen el mismo apellido comparten también el haplotipo en un porcentaje mayor al 50%. Esto sugiere una relación de parentesco que se sustenta en la historia de poblamiento y en los niveles de diversidad haplotípica encontrados en las islas.

Key Words: *San Andrés and Providencia, Y-Chromosome, short tandem repeat (STR), Surnames, Colombia.*

Introducción

El mar Caribe o mar de las Antillas es un mar interior rodeado en gran parte por grupos de islas. En épocas prehistóricas fue el camino que comunicó a numerosos pueblos del centro de América con los pueblos de América del Sur. Posteriormente, a la llegada de Cristóbal Colón, el mar Caribe se volvió el eslabón, el paso obligado entre Europa y el nuevo mundo, convirtiéndose en el medio natural de apoyo para la expansión de los pueblos Europeos hacia el Norte, Centro y Sur América. El poblamiento del Caribe ha sido el producto de las rutas originadas en las comunidades indígenas de América, seguidas por los diferentes procesos de conquista, constituyéndose de esta manera en escenario de coexistencia multiétnica, multicultural y multireligiosa (Vollmer, 1997).

El archipiélago de San Andrés y Providencia es una población aislada en la inmensidad de las aguas occidentales del mar Caribe. Se localiza a 180 Km. de las costas centroamericanas, a 400 Km. de Jamaica y a 480 Km. de las costas colombianas. Está formado por las islas de San Andrés, Providencia y Santa Catalina y por una gran cantidad de cayos, islotes y bajos (Figura 1).

Su historia del poblamiento, se ha entendido como la historia de la conformación de un territorio construido a través de la relación de los distintos grupos humanos que allí han interactuado. Desde esta perspectiva, Vollmer (1997) menciona la existencia de seis ciclos de poblamiento que han contribuido a la dinámica poblacional actual de las islas: el primero, el territorio Miskito (no se conoce -1692), donde individuos pertenecientes a la familia Miskito Sumo Matalpaba, transitaron e hicieron del archipiélago parte de su territorio sin establecer ningún asentamiento hasta donde se

conoce. Seguido del periodo denominado: las avanzadas de colonización (1629-1677), cuyo inicio se caracteriza por la llegada de Puritanos ingleses a la isla de San Andrés, que provenían de las Islas Bermudas con propósitos agrícolas, como el cultivo de tabaco y de colonización de territorios americanos. Para 1633 llegaron los primeros esclavos al archipiélago, proviniendo principalmente de la isla Tortuga; con el paso del tiempo el archipiélago, se convirtió en un importante centro de tráfico de esclavos en el Caribe, lo que condujo también, a la llegada de piratas ingleses y holandeses con fines contrabandistas y de asalto a las naves españolas que pasaban por el Caribe occidental, llevando los tesoros robados a los indígenas de América del Sur. Por esta razón, en 1641, España toma la isla de Providencia, lo cual da inicio a un periodo de 36 años de ocupaciones militares españolas e inglesas disputándose la propiedad de las islas. Luego de este proceso, viene el siglo del Olvido (1677-1780) donde a partir del último episodio militar entre España e Inglaterra, el Archipiélago regresa a su absoluta soledad. En esta época hacia 1730, llegaron personas del Caribe anglófono: Jamaica, Barbados y Trinidad y Tobago; otras llegaron directamente de las Islas Británicas, en particular de Escocia e Irlanda, africanos de África Occidental, españoles, franceses, holandeses, los cuáles iban y venían, arribaban, zarpaban y algunos se quedaban por la belleza de las islas. El siguiente periodo, se denomina: el Poblamiento Raizal (1780-1953). Durante este periodo, la base económica en el Archipiélago fue el cultivo de algodón como producto de exportación. Para ello fue necesaria la importación de mano de obra del África occidental y del Caribe anglófono, donde los historiadores calculan que para 1806 habitaban en San Andrés 1.200 personas, de las cuales 800 eran esclavos, y en Providencia no hubo más de 300 personas: algunas de San Andrés, pero la mayoría eran de Jamaica. Para 1822 el Archipiélago fue incorporado al territorio colombiano y en 1847, Phillip Beekman funda en San Andrés la primera iglesia Bautista, lo que condujo que a finales de siglo el 95% de la población fuera batista y el 90% supiera leer y escribir la lengua inglesa. Para 1851 el gobierno Colombiano ordena por ley la abolición de la esclavitud. Para 1953 inicia el quinto periodo donde el Archipiélago fue declarado Puerto libre (1953-1991). Las nuevas oportunidades para el comercio y para el turismo, impulsaron una nueva ola migratoria hacia San Andrés. Dos nuevos grupos de pobladores aparecen: migrantes del medio oriente (sirios, libaneses, palestinos y algunos judíos) y colombianos de origen continental (de los departamentos de Bolívar, Atlántico y Antioquia). Finalmente la Constitución Colombiana, le dio al archipiélago el reconocimiento de Departamento y

les otorga autonomía política, administrativa y fiscal a sus pobladores nativos.

En este estudio, se presenta un análisis genético y de apellidos de la población humana del archipiélago de San Andrés y Providencia (Colombia) que es un microcosmos del crisol cultural del Caribe insular, cuyo origen proviene de una mezcla centenaria de migrantes (indios miskito, puritanos ingleses, piratas holandeses, esclavos jamaquinos y africanos, población caribeña y españoles) que han contribuido de manera diferencial al establecimiento de dos grupos de pobladores nativos: los raizales de la isla de San Andrés y los raizales de la isla de Providencia (Vollmer 1997; Lamprea 2009).

En las sociedades humanas tener un nombre y ser identificado es esencial (King T.E. & Jobling., 2009). Desde el periodo medieval Inglés al presente, cada hijo lleva el apellido de su padre (Sykes & Irven, 2000), lo que conlleva a que tanto los apellidos como el cromosoma Y compartan el mismo mecanismo de herencia. En el archipiélago algunos historiadores atribuyen el origen de los apellidos isleños en casi su totalidad, a la transmisión que hicieron durante la esclavitud, algunos de los amos de su nombre de familia a los esclavos de su propiedad. Pero como menciona Francis C, (1991) la gran mayoría de estos apellidos pertenecen a los descendientes directos de pobladores ingleses, irlandeses, españoles, noruegos y aún norteamericanos que se establecieron en las islas durante y después de la esclavitud. Algunos son de origen oriental (chinos, hindúes) que se sumaron a la población en los primeros años del siglo XX.

Siendo San Andrés y Providencia escenarios de diferentes eventos históricos que involucran poblaciones heterogéneas en cuanto al lugar de procedencia y ancestría, resulta muy interesante estudiar estas islas en relación a un marcador uniparental como el cromosoma Y asociado con la herencia del primer apellido, que se asume como un alelo heredado por vía paterna (Jobling, 2001). Por esta razón el objetivo de este estudio fue determinar la composición genética de los individuos raizales de San Andrés y raizales de Providencia, así como sus posibles relaciones ancestrales con otras poblaciones mediante el uso de marcadores STR del cromosoma Y, asociando también la información heredada en el primer apellido, lo que permite asociar la genética clásica con el análisis demográfico y los análisis de marcadores uniparentales.

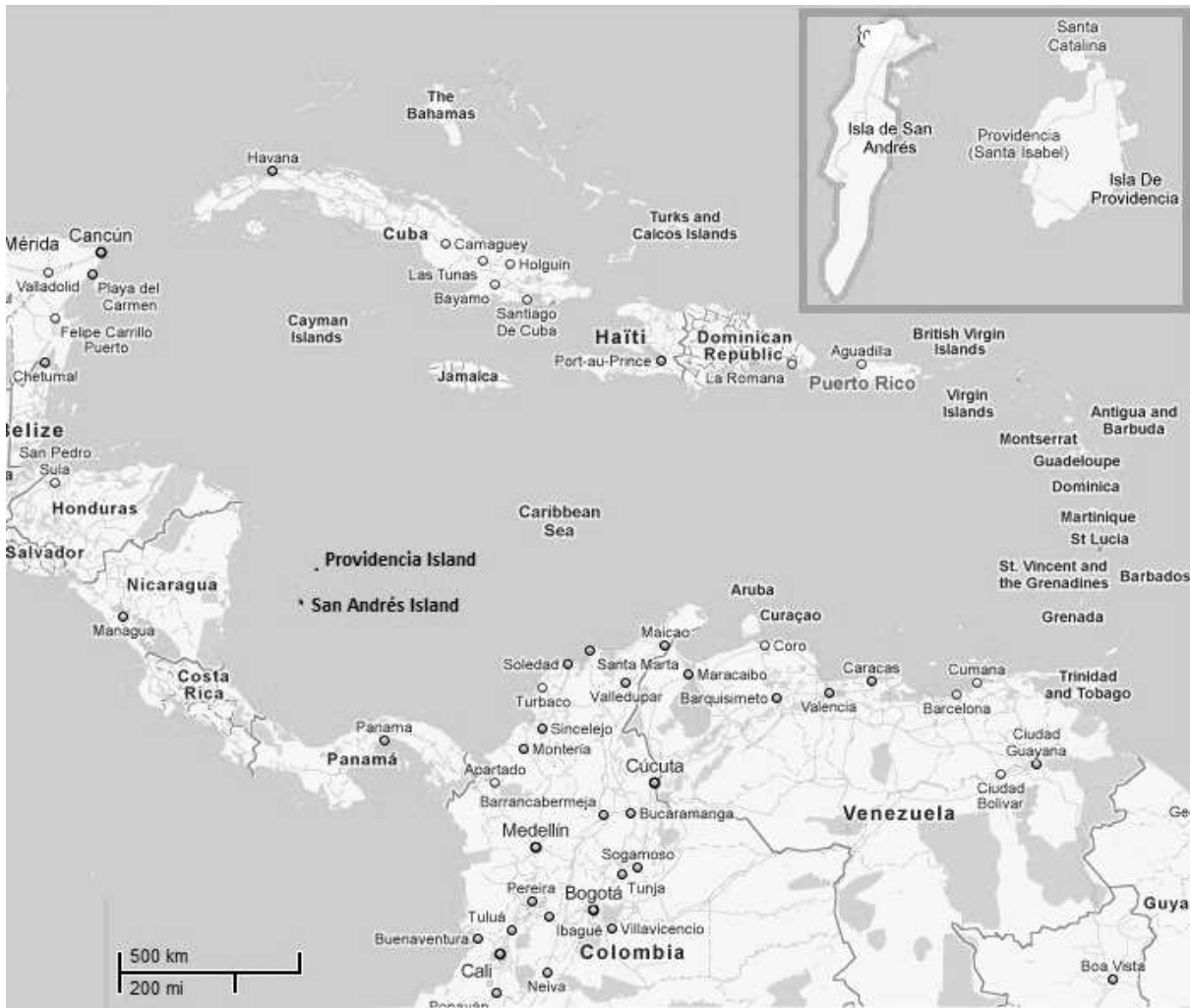


Figura 1. Posición Geográfica del Archipiélago de San Andrés y Providencia (Google Maps - ©2011 Google).

Materiales y Métodos

2.1 Población de estudio

Se recolectaron 54 hombres residentes en las islas (N=23 para la isla de Providencia y N=31 para la isla de San Andrés). Se aplicó una encuesta genealógica para clasificar como “Raizales de San Andrés” y “Raizales de Providencia” a aquellas personas que tenían al menos tres generaciones de ancestros (padre, abuelo paterno y bisabuelo paterno) nacidos en las islas. Según esta encuesta ninguno de los individuos participantes compartía ningún ancestro. Previo consentimiento informado, se tomaron muestras de 10ml de sangre, una fracción fue preservada mediante una tarjeta FTA (Whatman BioScience, USA).

2.2 Otras poblaciones

Con el propósito de comparar las islas con otras poblaciones, se incluyeron 2243 haplotipos publicados para 34 poblaciones. La inclusión de estas poblaciones tuvo como propósito la búsqueda de similitudes y diferencias entre las poblacionales en función de la ancestría y ubicación geográfica (Tabla 1).

2.3 Análisis molecular.

Diez y siete loci Y-STR fueron amplificados por PCR utilizando el kit comercial AmpFISTR Yfiler (Applied Biosystems, Warrington, U.K.), utilizando un termociclador programable 9700 (Applied Biosystems, Foster City, California). Las amplificaciones fueron realizadas utilizando los siguientes parámetros por ciclo: denaturación inicial (95°C por 11 min) seguido por 30 ciclos de denaturación de las hebras (94°C por 1 min), alineamiento de primers (61°C por 1 min), y extensión del ADN (72°C por 1 min), y extensión final (60°C por 60 80 min). Las alícuotas del ADN amplificado y marcado

con fluorescencia fueron mezcladas con formamida (Applied Biosystems, Foster City, California) y LIZ 500 (Applied Biosystems, Warrington, U.K.) y genotipificadas en un analizador genético ABI Prism 310 (Applied Biosystems, Stafford, Texas) utilizando el programa Genescan, versión 3.1.2. Las asignaciones alélicas fueron hechas mediante la comparación de los fragmentos amplificados con la escalera alélica (Applied Biosystems, Warrington, U.K.) que contiene el kit utilizando el programa Genotyper, versión 2.5.2 (Applied Biosystems, Foster City, California).

Biogeographical origin	Population	Sample size	Reference
NEW WORLD			
United States	African-American	10	Kayser et al., 2003
Bahamas	Abaco	41	Simms et al., 2011
	Eleuthera	30	Simms et al., 2011
	Exuma	30	Simms et al., 2011
	Grand Bahama	33	Simms et al., 2011
	Long Island	31	Simms et al., 2011
	New Providence	79	Simms et al., 2011
Caribbean islands	Jamaica	53	Torres et al., 2007
	Dominica	21	Torres et al., 2007
	Grenada	35	Torres et al., 2007
	St. Lucia	24	Torres et al., 2007
	St. Kitts	33	Torres et al., 2007
	St. Vincent	21	Torres et al., 2007
	Trinidad	32	Torres et al., 2007
	St. Thomas	133	Torres et al., 2007
	San Andrés	31	This study
	Providencia	26	This study
Colombia	Afro-Chocó	110	Yunis et al., 2005
	Andes	122	Yunis et al., 2005
	Cartagena	173	Builes et al., 2005
EUROPEAN			
United Kindom	Caucasians	220	Ballard et al., 2005
	Afro caribbean	241	Ballard et al., 2005
	South Asians	227	Ballard et al., 2005
AFRICAN			
western	North western	74	Bosch et al., 2000
	Biaka	8	Tishkoff et al., 2007
	Yoruba	12	Tishkoff et al., 2007
	Guinea Bissau	154	Rosa et al., 2006
eastern	Datog	23	Tishkoff et al., 2007
	Burunge	31	Tishkoff et al., 2007
	Mbugwe	14	Tishkoff et al., 2007
	Turu	20	Tishkoff et al., 2007
	Hadza	54	Tishkoff et al., 2007
	Sandawe	67	Tishkoff et al., 2007
	Sukuma	30	Tishkoff et al., 2007

Tabla 1. Poblaciones utilizadas para los análisis de comparación.

2.4 Análisis Estadísticos

Se estimaron frecuencias haplotípicas y diversidad haplotípica, de acuerdo a lo propuesto por Nei (1987), utilizando la ecuación: $D = (n/(n - 1)(1 - \sum p_i^2))$ donde n es el tamaño de la muestra y p_i es la frecuencia del haplotipo, utilizando el programa Arlequín 3.1 (Excoffier et al. 2005). El mismo programa fue utilizado para realizar el análisis molecular de la varianza (AMOVA) bajo un modelo mutacional paso a paso, para cuantificar la proporción de la varianza genética total entre y dentro de las islas.

Se predijeron los posibles haplogrupos de cromosoma Y presentes en cada isla utilizando los haplotipos encontrados con los 17 STRs utilizando el programa Athey's Haplogroup Predictor, versión 5, el cual se basa en una aproximación bayesiana de frecuencias alélicas (Athey, 2006).

Finalmente se realizó un análisis de redes median-joining utilizando Network, version 4.5.1.0 (Bandelt et al. 1999) (<http://www.fluxus-technology.com>). Los pesos de locus específicos fueron asignados de 2 a 10 de acuerdo con Decker et al. (2008) donde a los loci con mayores tasas de mutación se les asignó el menor peso. El locus DYS385a/b fue excluido. Algunos autores sugieren que este debe ser interpretado como un haplotipo aparte (Butler, 2003).

2.5 Comparación con otras poblaciones

Utilizando seis microsatélites: DYS19, DYS389-I, DYS390, DYS391, DYS392 Y DYS393 se realizó el análisis molecular de la varianza (AMOVA) entre las islas y las 34 poblaciones reportadas para evaluar la hipótesis de distribución al azar de los individuos entre pares de poblaciones. En el análisis se evaluaron diferentes agrupamientos con 10.000 permutaciones utilizando Arlequín 3.11 (Excoffier et al. 2005). Adicionalmente se calcularon distancias genéticas por pares de poblaciones utilizando los valores RST, con la cuales posteriormente se realizó un escalamiento multidimensional (MDS) utilizando el programa estadístico SPSS versión 20 (SPSS Inc., Chicago, IL).

2.6 Análisis de apellidos

Tomando el primer apellido de cada uno de los 54 participantes, se estimaron frecuencias y diversidad por apellidos; adicionalmente se registraron el número de haplotipos de cromosoma Y encontrados para cada uno y su correspondiente haplogrupo predicho. Una segunda línea de análisis, consistió en determinar similitudes y diferencias entre haplotipos, asociando a cada haplotipo su correspondiente apellido y lugar de nacimiento, ello se determinó utilizando un Análisis de Correspondencias Múltiples (ACM) mediante el programa SPAD versión 3.5 (2000) y un Análisis de Componentes Principales (APC) mediante el programa MVSP versión 3.0 (Kovach.,1998).

3. Resultados

3.1 Diversidad Genética y haplogrupos

Utilizando 17 sistemas microsatélites, se encontraron 18 haplotipos únicos para la isla de Providencia 26 y para la Isla de San Andrés (Tablas 1 y 2); no se observaron haplotipos compartidos entre ellas. La diversidad haplotípica calculada en las islas omitiendo el sistema DYS385a/b, que al ser un microsatélite complejo, debe ser interpretado como un haplotipo aparte (Butler, 2003); fue para Providencia 0.9488 y para San Andrés 0.9870 (Tablas suplementarias 1 y 2). La diversidad genética promedio para todos los loci fue en Providencia 0.5990 (+/- 0.1287) y en San Andrés 0.7265 (+/- 0.0756). El sistema DYS385, cuando se considera un genotipo, fue el sistema más diverso en el conjunto de marcadores analizados, con una diversidad génica de 0.8673 en Providencia y en San Andrés de 0.9147. Cabe resaltar la baja variabilidad (menor al 50%) de los sistemas DYS389I (0.3457), DYS437 (0.4012), DYS393 (0.4753) encontrados en Providencia.

Se predijeron los haplogrupos de cromosoma Y presentes, en las tablas 2 y 3 se reportan los haplogrupos predichos para las poblaciones de San Andrés y Providencia con el valor de probabilidad en cada caso. Se encontraron para Providencia cuatro haplogrupos, donde el haplogrupo E1b1a fue el más frecuente 47.8 %, seguido del haplogrupo R1b 39.1%; los otros dos haplogrupos encontrados fueron R1a y E1b1b los cuales representaron el 8.69% y 4.34%, respectivamente. Para San Andrés se

encontraron seis haplogrupos; donde el haplogrupo más frecuente también fue el E1b1a 48.4%, seguido del haplogrupo R1b 35.5%. Los haplogrupos J2a1h, I2b1, I1 y E1b1b representaron el 6.5%, 3.2%, 3.2% y 3.2%, respectivamente. El valor de probabilidad fue mayor al 80%, salvo en el haplotipo H017 de la isla de providencia con una probabilidad igual a 47.9%.

Providencia Island																			
HAPLOTYPE	M	DYS19	DYS185a/b	DYS189	DYS189b	DYS190	DYS191	DYS192	DYS193	DYS413	DYS426	DYS439	DYS448	DYS456	DYS458	DYS615	Y-GATA-H4	Predicted haplogroup	Probability(%)
H001	A	13	11-11	12	28	25	12	13	13	15	12	12	19	15	17	25	13	R1b	100
H002	B	14	11-14	12	28	25	12	13	13	15	12	12	19	15	17	24	12	R1b	100
H003	A	14	11-14	12	28	25	11	13	13	15	12	12	19	15	17	24	12	R1b	100
H004	A	17	17-18	13	30	21	11	11	14	14	11	12	21	17	16	22	11	E1b1a	100
H005	A	17	17-19	13	31	21	10	11	14	14	11	13	22	16	16	21	11	E1b1a	100
H006	A	15	11-14	13	29	25	10	11	13	14	11	11	20	17	16	23	12	R1a	100
H007	B	16	15-19	13	30	21	10	11	14	14	11	12	21	15	19	20	11	E1b1a	100
H008	2	16	15-18	13	30	21	10	11	14	14	11	12	21	15	19	20	11	E1b1a	100
H009	A	16	15-19	12	30	21	10	11	14	14	11	12	21	15	19	20	11	E1b1a	100
H010	A	15	16-17	13	31	21	11	13	13	14	11	11	21	16	16	21	11	E1b1a	100
H011	A	14	12-14	13	29	24	10	13	13	15	12	11	19	15	17	23	12	R1b	100
H012	A	16	15-15	13	30	22	11	11	14	14	11	12	21	17	16	22	11	E1b1a	99.4
H013	A	13	16-16	13	30	24	10	11	13	14	10	11	20	17	16	21	12	E1b1b	100
H014	A	14	11-15	13	29	24	11	13	13	15	12	12	19	16	18	23	11	R1b	100
H015	A	14	11-14	13	29	24	10	13	13	14	10	13	19	15	16	23	13	R1b	99
H016	A	14	11-14	13	29	24	10	13	13	14	10	13	19	15	17	23	13	R1b	99
H017	A	15	11-14	13	29	25	10	11	13	14	11	12	20	17	16	23	12	R1a	47.9
H018	A	17	17-19	13	30	21	10	11	14	14	10	12	21	14	17	21	13	E1b1a	100
Genetic diversity		0.772	0.867284	0.34568	0.69753	0.691195	0.537	0.4938	0.4753	0.4012	0.6235	0.5894	0.6667	0.642	0.6605	0.787	0.6419753		

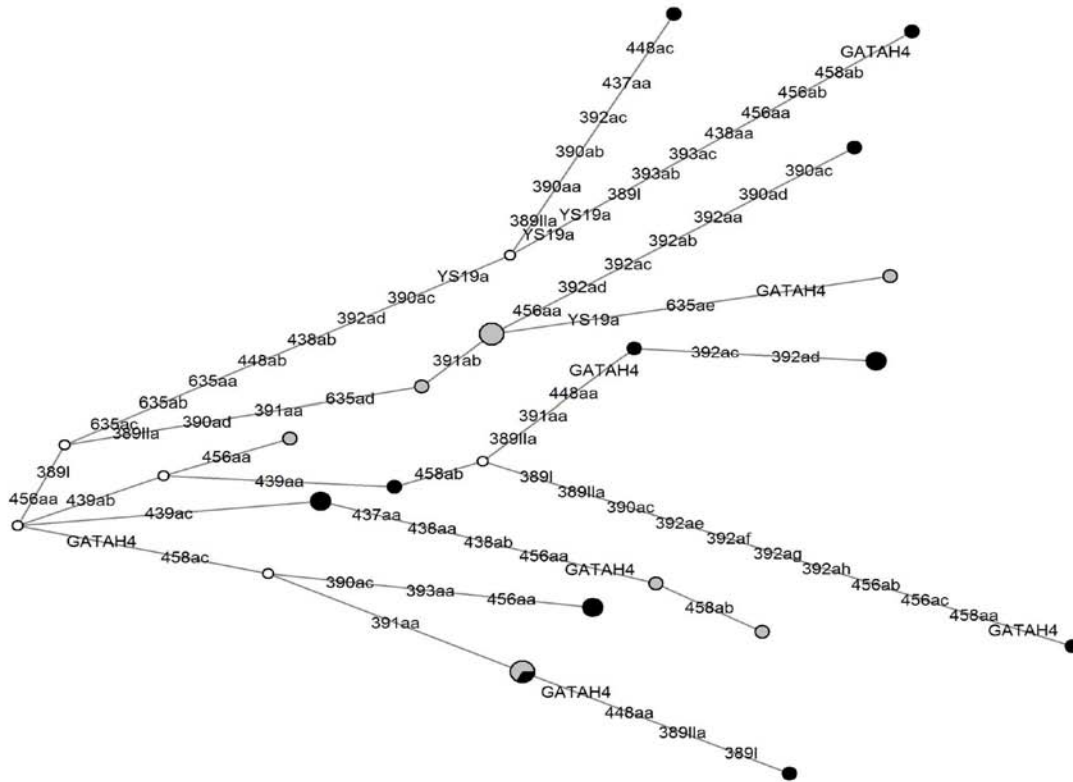
Table 2. Haplogrupos predichos de cromosoma Y y sus valores de probabilidad en la isla de providencia (N=23)

San Andrés Island																	Predicted haplogroup	Probability(%)	
HAPLOTYPE	N	DYS19	DYS385a/b	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS437	DYS438	DYS439	DYS448	DYS456	DYS458	DYS635	Y-GATA-H4		
H001	1	16	16-16	13	31	21	10	11	13	14	13	12	21	15	18	21	13	E1b1a	100
H002	2	14	11-15	13	29	24	10	13	13	15	12	13	19	16	17	23	12	R1b	100
H003	1	14	11-15	13	29	24	10	13	13	15	12	10	19	16	17	23	12	R1b	100
H004	1	17	15-16	13	29	23	10	12	15	15	10	12	20	17	16	20	11	I2b1	100
H005	1	14	11-14	14	30	24	11	13	13	15	12	12	18	16	18	23	12	R1b	100
H006	2	13	13-16	13	29	23	10	9	12	15	9	12	20	15	16	22	12	J2a1h	87.6
H007	1	16	14-20	14	32	21	11	11	14	14	11	12	21	15	16	21	12	E1b1a	100
H008	1	14	11-13	12	28	23	12	9	13	15	12	12	19	16	17	24	12	R1b	100
H009	1	16	17-18	13	30	21	10	13	14	14	11	10	20	16	16	24	11	E1b1a	100
H010	1	14	11-14	13	30	24	11	11	13	15	12	10	18	16	16	23	13	R1b	86.3
H011	1	14	11-14	13	30	24	11	13	13	15	12	10	18	16	16	23	13	R1b	100
H012	1	14	15-15	12	29	22	10	11	13	17	10	12	20	15	14	19	11	I1	100
H013	2	14	11-14	13	29	23	10	13	12	15	12	12	19	15	18	23	11	R1b	100
H014	1	15	16-18	13	31	21	10	11	13	14	11	12	21	18	16	21	11	E1b1a	100
H015	1	16	16-16	13	31	21	10	11	13	14	12	11	21	15	16	21	12	E1b1a	100
H016	1	17	16-16	13	31	21	10	11	13	14	12	11	21	15	16	21	12	E1b1a	100
H017	3	17	17-17	13	30	21	10	11	14	14	11	13	21	16	16	22	11	E1b1a	100
H018	1	13	17-17	13	30	21	10	11	14	14	11	10	21	16	16	22	11	E1b1a	81.6
H019	1	14	11-16	13	29	24	11	13	13	15	12	12	19	16	18	23	11	R1b	100
H020	1	15	16-16	12	30	21	10	11	13	14	11	12	21	15	17	20	12	E1b1a	100
H021	1	16	17-18	13	30	21	10	11	15	14	11	13	21	16	16	21	11	E1b1a	100
H022	1	15	16-16	12	30	21	10	13	13	14	11	13	21	18	18	20	13	E1b1a	100
H023	1	15	16-16	12	30	21	10	11	13	14	12	12	21	15	17	21	12	E1b1a	100
H024	1	14	11-14	12	28	23	10	17	13	15	12	10	19	18	15	23	11	R1b	100
H025	1	13	13-13	13	31	21	10	9	13	14	11	12	21	18	18	23	12	E1b1b	99.6
H026	1	15	16-17	13	31	21	10	11	13	14	11	12	21	16	18	23	12	E1b1a	100
Genetic diversity		0.825	0.9146722	0.62123	0.80229	0.73153	0.5567	0.7461	0.6379	0.6681	0.7399	0.7378	0.7607	0.7482	0.7711	0.821	0.7294485		

Table 3. Haplogrupos predichos de cromosoma Y y sus valores de probabilidad en la isla de San Andrés (N=31)

3.2 Análisis de redes

Con el propósito de tener un mejor conocimiento de las bases moleculares de la variabilidad encontrada en las islas, se construyó una red median-joining con todos los haplotipos pertenecientes a los dos haplogrupos más frecuentes en las islas (E1b1a y R1b), utilizando todos los Y-STRs salvo el locus DYS385a/b. Mediante este análisis solo se observó un haplotipo compartido entre las islas en el haplogrupo R1b (Figura 3). Dentro de las redes fueron también postulados varios vectores propios (haplotipos no muestreados ó haplotipos extintos); 10 en el haplogrupo E1b1a y 6 en el Haplogrupo R1b. Los demás haplotipos se encuentran en su mayoría dispersos, esto como consecuencia de la diversidad genética encontrada en las islas (figuras 2 y 3).



Figuras 2 y 3. Median-joining network de los haplogrupo E1b1a y R1b, respectivamente. Los círculos representan haplotipos con área proporcional a su frecuencia; los colores indican la población de origen (Negro=San Andrés; Gris=Providencia). Los median vectors se encuentran en blanco.

3.3 Estructura poblacional

El porcentaje de variabilidad observado en las islas mostró que la mayoría de la variación se encuentra en el interior de las poblaciones (98.31%), mientras que el porcentaje de variación entre las poblaciones fue de 1.69% no significativo ($p=0.1741$, 10001 permutaciones).

3.4 Comparación con otras poblaciones

En cuanto al porcentaje de variabilidad observado entre 10 regiones (El primer grupo compuesto por Providencia-Colombia, San Andrés-Colombia, Grenada, St. Lucia, St. Kitts, St. Vincent, Trinidad, St. Thomas, Jamaica, Afro Chocó-Colombia, Afro caribbean-UK, Eastern Africa-Sukuma, Eastern Africa-Sandawe, Eastern Africa-Turu, Eastern Africa-Mbugwe, Eastern Africa-Burunge, Eleuthera-Bahamas, Long Island-Bahamas, Exuma-Bahamas, Grand Bahama-Bahamas, New Providence-Bahamas, Abaco-Bahamas; el segundo grupo compuesto por Andes-Colombia, Cartagena-Colombia, Dominica y Caucasians UK; el tercer grupo por la población de South Asians; el cuarto grupo compuesto por Eastern Africa-Datog; el quinto grupo por: Eastern Africa-Hadza; el sexto por la población: Western Africa-Yoruba; el séptimo por Western Africa-Biaka; el octavo por: Western Africa- -Guinea Bissau; el noveno compuesto por African-American y el décimo grupo por North Western Africa); mostró que la mayoría de la variación se encuentra dentro de los grupos poblacionales, pero hay un porcentaje significativo de la variación 27,31% ($p < 0.000$, 10000 permutaciones) que puede ser explicado por diferencias entre grupos de poblaciones.

Para evaluar las diferencias poblacionales se calcularon distancias genéticas RST por pares de poblaciones y se graficaron mediante un escalamiento multidimensional (Figura 4). Mediante este análisis se observan dos conjuntos de poblaciones, principalmente. El primero, en donde se encuentran la mayoría de poblaciones africanas y con ancestría africana, grupo que corresponde con las regiones uno utilizada en el AMOVA. En este grupo se encuentran San Andrés y Providencia más cercanos a las islas del Caribe y a las poblaciones del este de África: Burungé, Mbugue y Turu. En el segundo conjunto se reúnen poblaciones de otros orígenes, Europeo, Andes y Cartagena, entre otros.

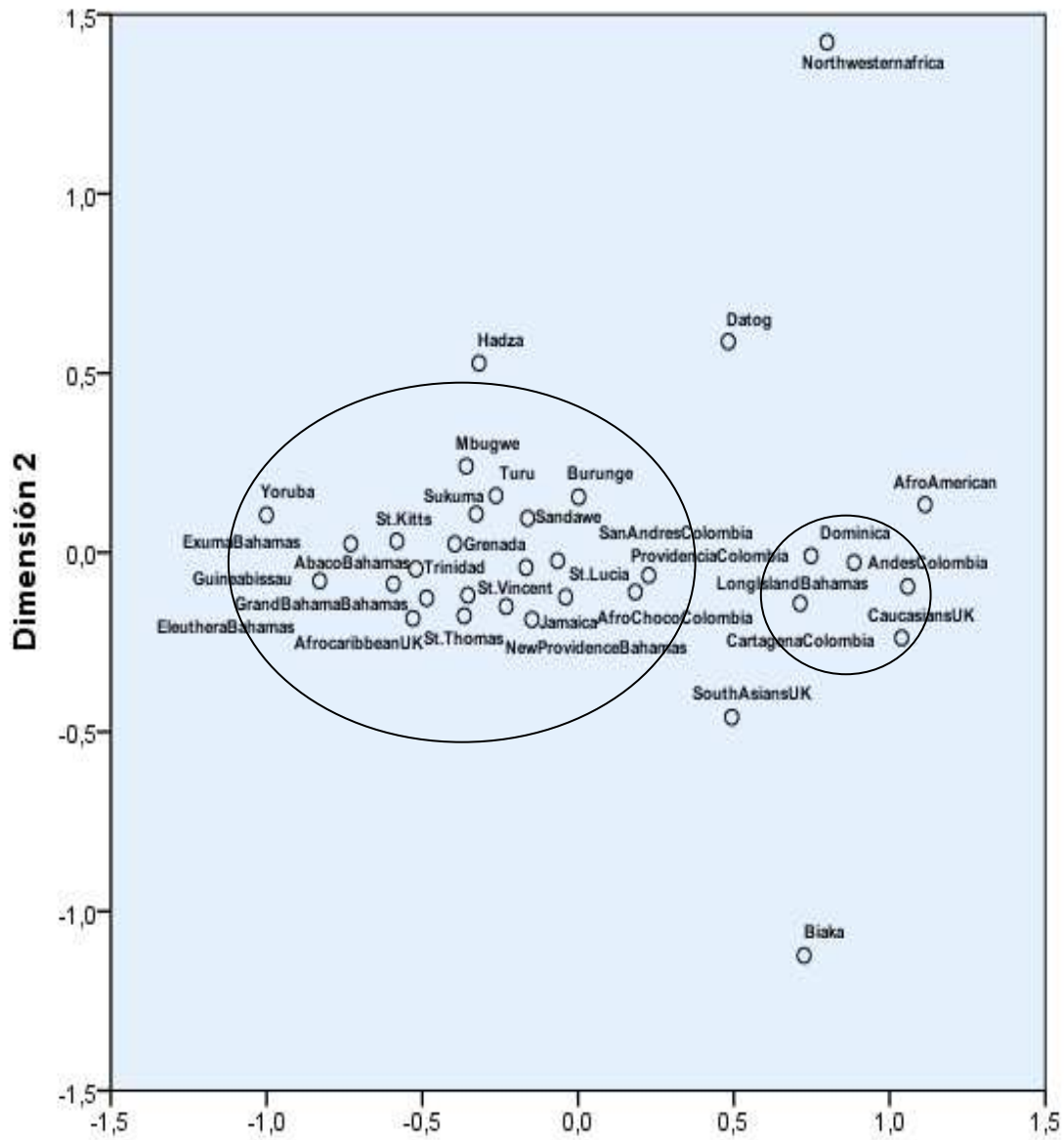


Figura 4. Escalamiento multidimensional basado en la matriz de distancia RST por pares de poblaciones. El valor de stress fue 0.0154.

Apellidos.

3.5 Diversidad y número de apellidos

A partir del primer apellido de cada uno de los participantes se encontraron para la isla de Providencia 11 apellidos con una diversidad de 0,4783, siendo los apellidos más frecuentes Bryan (n=7) y Archbold (n=5) representando el 52,17% del total de apellidos para la isla. Al tomar cada uno de estos apellidos, y comparar sus haplotipos y haplogrupos predichos se encontró que 3 de 4 haplotipos Bryan pertenecen al haplogrupo E1b1a y solo uno al haplogrupo R1a (Tabla 4), en el caso del apellido Archbold los 3 haplotipos de cromosoma Y pertenecen al mismo haplogrupo.

Para la isla de San Andrés se encontraron 18 apellidos, con una diversidad de 0,5806, siendo Martínez (N=4), Bent (N=3) y Pomare (N=3) los apellidos más frecuentes con el 32,26 % del total de los apellidos, presentado los haplogrupos E1B1a y R1b (Tabla 5).

Como se observa en las tablas 4 y 5, el único apellido compartido en ambas islas fue Bent (N=1 en Providencia y N=3 San Andrés), aunque con diferentes haplotipos y haplogrupos: E1B1a en Providencia y R1a en San Andrés.

Al evaluar el número de haplotipos presentes en cada apellido se encontró que para Bryan el 57% comparte el mismo haplotipo, para Archbold el 60% en el caso de Providencia, y para San Andrés, Martínez y Bent con 50 y 66 respectivamente.

Al realizar el Análisis de Correspondencias Múltiples (ACM) y el Análisis de Componentes Principales (PCA), tomando los 17 marcadores STR del cromosoma Y para cada individuo, se determinaron diferencias y semejanzas entre cada uno de los haplotipos, ubicándose cercanos aquellos haplotipos iguales o muy similares en un espacio geográfico. Como cada haplotipo fue asociado con su correspondiente apellido, esto permitió determinar cuáles individuos comparten la información genética en el cromosoma Y, y la información genealógica de los apellidos que tienen esos individuos (Figuras 5 y 6). Se encontraron dos grupos principales de apellidos, que corresponden a los haplogrupos predichos R1b y E1b1a .

Providencia Island

nSurname s	N	Frequency (%)	N° haplotypes	Predicted haplogroup
BRYAN	7	30,43	4	E1b1a/R1a
ARCHBOLD	5	21,74	3	R1b
NEWBALL	2	8,70	2	R1b
ROBINSON	2	8,70	2	R1a/E1b1a
BENT	1	4,35	1	E1b1a

Tabla 4. Frecuencia, número de haplotipos y haplogrupo predicho de los apellidos más frecuentes en la Isla de Providencia. Los valores de probabilidad fueron superiores al 80% salvo en el apellido ROBINSON donde para el haplogrupo R1a el valor encontrado fue de 47.9%.

San Andrés Island

Surnames	N	Frequency (%)	N° haplotypes	Predicted haplogroup
MARTÍNEZ	4	12,90	2	E1b1a
BENT	3	9,68	2	R1a
POMARE	3	9,68	3	E1b1a
JAMES	2	6,45	2	R1b
LIVINGSTONE	2	6,45	2	R1b
MANUEL	2	6,45	2	E1b1a
WALTERS	2	6,45	2	E1b1b-E1b1a

Tabla 5. Frecuencia, número de haplotipos y haplogrupo predicho de los apellidos más frecuentes en la Isla de San Andrés. Los valores de probabilidad en todos los casos fueron superiores al 80%.

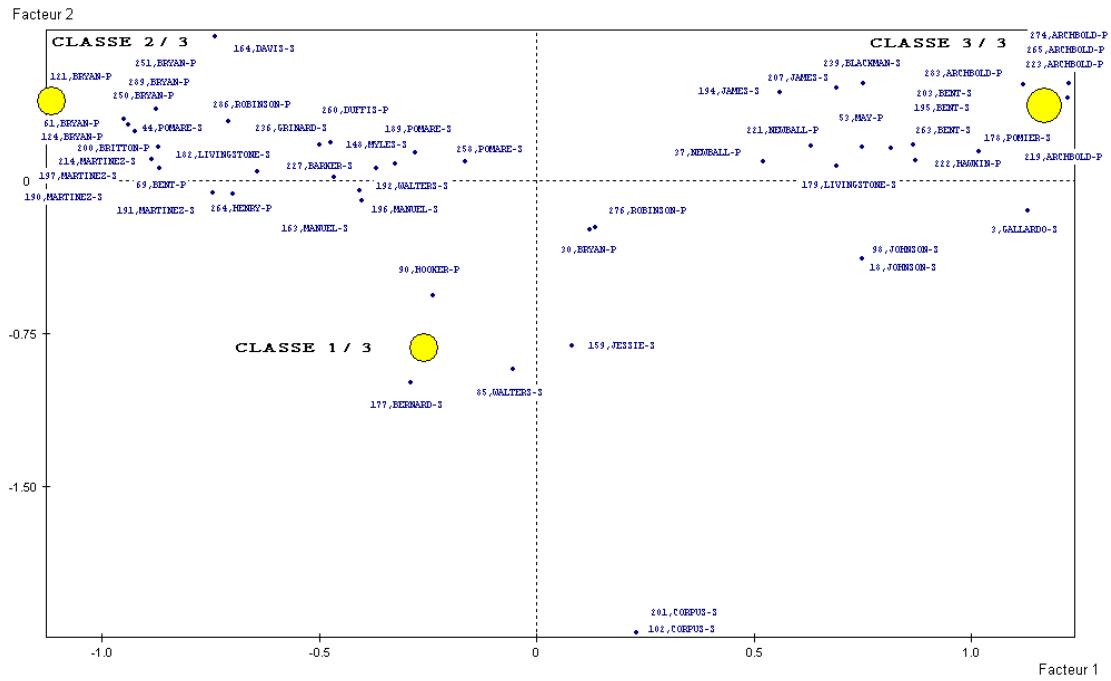


Figura 5. Análisis de Correspondencias Múltiples de apellidos y de datos genéticos para 17 marcadores STR del cromosoma Y en el Archipiélago de San Andrés y Providencia.

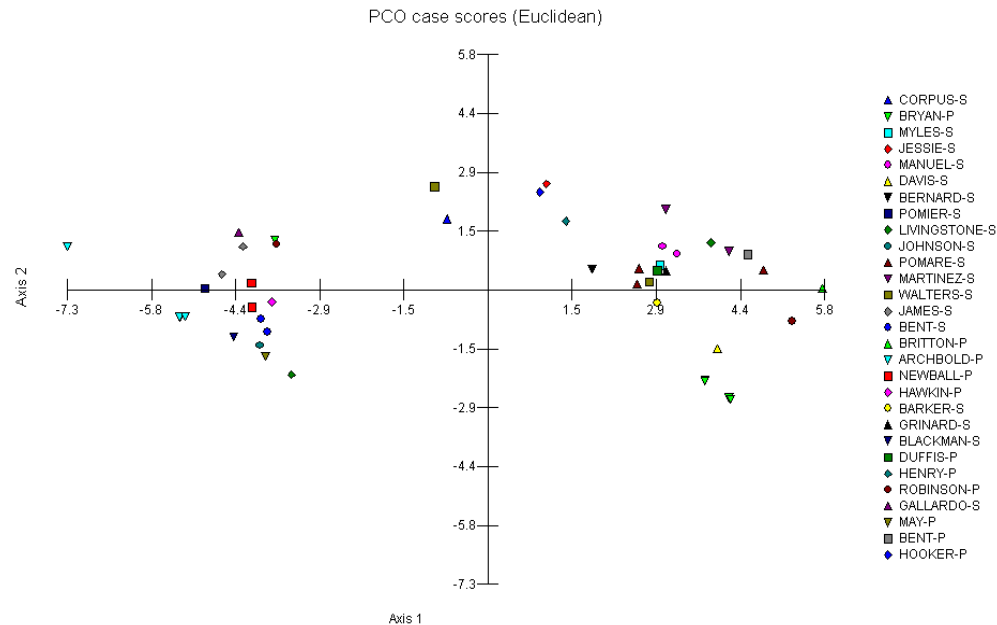


Figura 6. Análisis de Componentes Principales de apellidos y de datos genéticos para 17 marcadores STR del cromosoma Y en el Archipiélago de San Andrés y Providencia.

4.0 Discusión

En el curso de la evolución, las poblaciones varían en el tiempo y en el espacio, pueden mantenerse aisladas durante uno o varios periodos de tiempo, bien sea por la existencia de barreras geográficas, ecológicas o culturales; o por otra parte, tienen la posibilidad de intercambiar genes, dando lugar a poblaciones que presentan mezcla de algunas características y rasgos genéticos (Cavalli-Sforza y Bodmer, 1971). Este contexto no es ajeno a las poblaciones humanas, mucho menos en aquellas que han sido producto de escenarios demográficos tan drásticos como las poblaciones afro descendientes en el mundo, pero principalmente en el continente Americano.

Para la elaboración del mapa cultural del Caribe se partió del reconocimiento de que el proceso de formación de las distintas identidades culturales de la región caribeña está directamente relacionado con la historia de su poblamiento: la historia de los encuentros y desencuentros entre los diferentes

grupos humanos que fueron habitando este territorio (Vollmer 1997). Desde esta óptica, la genética contribuye al conocimiento de la historia de poblamiento a partir de la información contenida en su ADN ligada de manera indistinguible a los procesos culturales que caracterizan determinada población. Adicionalmente pocos son los estudios de cromosoma Y que se han realizado en las poblaciones del Caribe (Torres et al. 2007).

En este estudio se realizó un análisis profundo de la diversidad de Cromosoma Y en las poblaciones raizales de San Andrés y Providencia mediante la utilización de un criterio genealógico, que permitió analizar individuos que tienen por lo menos tres generaciones de ancestros en las islas para 17 marcadores microsatélites y la asociación del primer apellido a esta información genética. De esta manera, se evaluó el nivel de variabilidad y la subestructura genética de estas poblaciones, así como el establecimiento de la contribución genética de las poblaciones africanas y/o afrodescendientes y poblaciones europeas en las islas.

Se encontró que los valores de diversidad haplotípica son bajos (Providencia: 0.9488, San Andrés: 0.9870), con respecto a los ya publicados para otras poblaciones afrodescendientes; por ejemplo, para la población Afro colombiana de Chocó se ha reportado una diversidad haplotípica de 0.9955 (Yunis et al. 2005) mientras que para poblaciones del Caribe anglo parlantes se ha reportado: Dominica un valor de 0.9714, Grenada 0.9937, St. Lucía 0.9964, St. Kitts 0.9886, Trinidad 0.9960, St. Thomas 0.9973 y Jamaica 0.9954 (Torres et al. 2007). La reducida diversidad genética observada en Providencia puede ser el resultado del muestreo de una comunidad pequeña con ancestría genética similar o como resultado de un tamaño efectivo poblacional en la isla, donde los loci del cromosoma Y son susceptibles a fenómenos de deriva, como un cuello de botella reciente (Giroti y Talwar 2010; Torres et al. 2007; De Knijff 2000).

En la isla de Providencia cuatro haplogrupos fueron predichos siendo el haplogrupo E1b1a el más frecuente 47.8 % (N=11), seguido del haplogrupo R1b 39.1% (N=9). Para la isla de San Andrés seis haplogrupos fueron predichos; siendo también el E1b1a 48.4% (N=15), seguido del haplogrupo R1b 35.5% (N=11) los haplogrupos más frecuentes. Este haplogrupo E1b1a alcanza frecuencias de más del 80% en muchas partes de África Occidental, África Central, África Oriental y África del Sur (Trombetta

B et al. 2011) y es el menos común en las poblaciones del Cáucaso (Hammer et al. 2006; Y chromosome Consortium 2002; Quintana-Murci et al. 1999). Por otra parte, más del 50% de los hombres en Europa están asociados con el haplogrupo R (Sims et al. 2007; Jobling & Tyler-Smith 2003) que a su vez se encuentran distribuidos en el occidente, centro y sur de Asia (Chiaroni J et al. 2009).

Otros estudios indican que la mayoría de Africanos americanos y caucásicos en los Estados Unidos pertenecen a uno de los dos principales sub-haplogrupos, E3a(58-62%) y R1b(47-58,3%), respectivamente (Sims et al. 2007; Hammer et al. 2006; Vallone and Butler 2004). En nuestros resultados se observa este mismo fenómeno de conservar ese pasado Africano y Europeo que se dio como producto de las diferentes procesos de colonización y esclavitud entre 1730-1853, caracterizado por la llegada de gente del Caribe anglófono: Jamaica, Barbados, y Trinidad y Tobago, otras llegaron directamente de las islas británicas, en particular de Escocia e Irlanda; y otras, directamente de África occidental. Antes de este periodo las islas permanecieron casi abandonadas (Vollmer 1997).

El AMOVA entre las dos islas utilizando 15 Y-STR's no mostró diferencia significativa $p=0.1741$, a pesar de que entre las dos poblaciones sólo se comparte un haplotipo, sus distribuciones de frecuencias entre sistemas no son muy diferentes y no reflejan procesos de subestructura poblacional, lo cual pudo observarse en la red para cada uno de los haplogrupos donde se encuentran haplotipos de ambos orígenes (Figuras 2 y 3).

Por su parte, el AMOVA expuso la presencia de 10 regiones con un porcentaje de variación de 27,31% ($p<0.000$), cuya representación se observa en la figura 4, donde se ve que las poblaciones afro-descendientes, agrupadas en la región 1 del AMOVA, se ubican en un espacio genético similar respecto a las demás poblaciones; a excepción de la población de Dominica quien se agrupó con las poblaciones de Cartagena, Mestiza Caucásica colombiana y Caucásica del Reino Unido. Esta población en un estudio realizado por Torres et al. (2007) presentó la menor frecuencia de la inserción ALU DYS 287 que define el haplogrupo E africano y se piensa puede tener una mayor contribución genética de otras poblaciones no africanas o afrodescendientes, lo cual podría explicar este agrupamiento.

Los análisis de apellidos evidencian una menor diversidad de apellidos en Providencia en relación a San Andrés, 47% y 58% respectivamente. Adicionalmente se observó en ambas islas, que los raizales que tienen el mismo apellido comparten además el haplotipo en un porcentaje mayor al 50%, lo cual implica que aunque en los criterios de inclusión de los participantes se tuvo en cuenta en base a su genealogía que no compartieran ancestros, hay una alta proporción de personas que comparten el mismo apellido y el mismo haplotipo, es decir en algún momento en su historia familiar tuvieron un ancestro en común. Con base en este mismo criterio los análisis multivariados realizados permitieron observar aquellos apellidos que más se parecen en base a su información genética. Todos los apellidos analizados se encuentran dentro del listado realizado nombres de las familias de mayor arraigo y de vigencia actual en las islas (Francis C., 1991)

5.0 Conclusión

El archipiélago de San Andrés y Providencia presenta en su mayoría un componente ancestral Africano y Europeo, característico de los diferentes ciclos de poblamiento del Caribe insular. Adicionalmente no se encuentran diferencias significativas entre San Andrés y Providencia. Ambas poblaciones se ubican cercanas a poblaciones africanas y afro descendientes. Son similares a otras poblaciones del caribe insular, a otras poblaciones del este de África y no se diferencian de otras poblaciones colombianas afro-descendientes como la del departamento del Chocó. En la reconstrucción de la historia filogenética de las poblaciones interactúan diferentes disciplinas como la antropología, arqueología y lingüística, en donde se obtienen diferentes puntos de vista. Este estudio no pretende cuestionar las características culturales que identifican a cada una de ellas.

Agradecimientos

En primer lugar a los habitantes de la población de San Andrés y Providencia por permitirnos conocer un poco más acerca de su historia biológica, así como también a las diferentes instituciones de salud: CAPRECOM, BIOLAB, OMALINA OWKIN DE GONZÁLEZ y el Hospital Providencia, y a la Bióloga Natalia Lamprea quienes colaboraron en el muestreo. En segundo lugar, a los miembros del Grupo de Genética de Poblaciones e Identificación, a la Universidad Nacional de Colombia sede Bogotá, al Instituto de Genética y a la sede Caribe. Finalmente a Colciencias con su programa Jóvenes Investigadores e Innovadores "Virginia Gutiérrez de Pineda" año 2010 por su apoyo en la realización de este estudio.

Bibliografía

- Athey, T. W. 2006. Haplogroup prediction from Y-STR values using a Bayesian–allele frequency approach. *J. Genet. Geneal.* 2:34–39.
- Ballard, D. J., Phillips, C., Thacker, C. R., Robson, C., Revoir, a P., & Syndercombe Court, D. 2005. Y chromosome STR haplotypes in three UK populations. *Forensic Sci. Int. Genet.* 152(2-3):289-305.
- Bandelt, H. J., Forster, P., & Röhl, a. 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* 16(1): 37-48..
- Builes, J. J., Martínez, B., Gómez, A., Caraballo, L., Espinal, C., Aguirre, D., et al. 2007. Y chromosome STR haplotypes in the Caribbean city of Cartagena Colombia. *Forensic Sci. Int. Genet.* 167(1):62-69.
- Butler, J M. 2003. Recent Developments in Y-Short Tandem Repeat and Y-Single Nucleotide Polymorphism Analysis Recent Developments in Y-Short Tandem Repeat and Y- Single Nucleotide Polymorphism Analysis a. *Forensic Science Review.*15: 91-111.

- Cavalli-Sforza, L and Bodmer, W. 1971. *The Genetics of Human Populations*. W. H. Freeman, San Francisco (reprinted 1999 by Dover Publications).
- Chiaroni, J., Underhill, P. a, & Cavalli-Sforza, L. L. 2009. Y chromosome diversity, human expansion, drift, and cultural evolution. *Proc. Natl. Acad. Sci. U.S.A.* 106(48):20174-20179.
- Consortium, T. Y. C. 2002. A Nomenclature System for the Tree of Human Y-Chromosomal Binary Haplogroups. *Genome Res.* 12(2):339-348.
- Decker, A. E., M. C. Kline, J. W. Redman et al. 2008. Analysis of mutations in father-son pairs with 17 Y-STR loci. *Forensic Sci. Int. Genet.* 2:e31–e35.
- De Knijff, P. 2000. Messages through bottlenecks: On the combined use of slow and fast evolving polymorphic markers on the human Y chromosome. *Am. J. Hum. Genet.* 67(5):1055–1061.
- Excoffier, L., Laval, G., & Schneider, S. 2005. Arlequin version 3.0: an integrated software package for population genetics data analysis. *Evol. Bioinform. Online.* 1:47-50.
- Francis, C. 1991. *Compendio de cultura popular tradicional de las islas de San Andrés y Providencia*. Ediciones archipiélago. San Andrés Isla, Colombia.
- Giroti, R., & Talwar, I. 2010. The most ancient democracy in the world is a genetic isolate: an autosomal and Y-chromosome study of the hermit village of Malana Himachal Pradesh, India. *Hum. Biol.* 82(2):123-141.
- Hammer, M. F., Chamberlain, V. F., Kearney, V. F., Stover, D., Zhang, G., Karafet, T., et al. 2005. Population structure of Y chromosome SNP haplogroups in the United States and forensic implications for constructing Y chromosome STR databases. *Forensic Sci. Int. Genet.* 164(1):45-55.

- Jobling, M a. 2001. Y-chromosomal SNP haplotype diversity in forensic analysis. *Forensic Sci. Int. Genet.* 118(2-3):158-162.
- Jobling, Mark a, & Tyler-Smith, C. 2003. The human Y chromosome: an evolutionary marker comes of age. *Nat. Rev. Genet.* 48: 598-612.
- King T.E. & Jobling M.A. 2009. What's in a name? Y chromosomes, surnames and the genetic genealogy revolution. *Trends Genet.*, 25: 351-360.
- Kovach, W.L., 1998. *MVSP - A Multivariate Statistical Package for Windows, ver. 3.0.* Kovach Computing Services, Pentraeth, Wales, U.K.
- Quintana-Murci, L., Semino, O., Minch, E., Passarimo, G., Brega, a, & Santachiara-Benerecetti, a S. 1999. Further characteristics of proto-European y chromosomes. *Eur. J. Hum. Genet.* 75: 603-608.
- Lamprea, N.2009. Caracterización genética de la población humana de San Andrés y Providencia a partir de los marcadores microsatélites (STR'S) empleados por el combined dna index system (CODIS). Tesis de Maestria, Departamento de Biología, facultad de ciencias. Universidad Nacional de Colombia. Bogotá, Colombia.
- Nei, M. 1987. *Molecular Evolutionary Genetics.* Columbia University Press, New York, USA.
- Sykes B., &Irven C. 2000 Surnames the Y chromosome. *American Journal of Human Genetics.* 66(4):1417–1419.
- Sims, L. M., Garvey, D., & Ballantyne, J. 2007. Sub-populations within the major European and African derived haplogroups R1b3 and E3a are differentiated by previously phylogenetically undefined Y-SNPs. *Hum. Mutat.* 28(1): 97.

- Torres, J., Kittles, R. a, & Stone, a C. 2007. Mitochondrial and Y chromosome diversity in the English-speaking Caribbean. *Ann. Hum. Genet.* 71(Pt 6):782-90
- Trombetta, B., Cruciani, F., Sellitto, D., & Scozzari, R. 2011. A new topology of the human Y chromosome haplogroup E1b1 E-P2 revealed through the use of newly characterized binary polymorphisms. *PLoS ONE.* 6(1): e16073.
- Vallone, P. M., & Butler, John M. 2004. Y-SNP typing of U.S. African American and Caucasian samples using allele-specific hybridization and primer extension. *Journal of forensic sciences,* 49. 723-732.
- Vollmer, L. 1997. *La historia del poblamiento de San Andrés, Vieja Providencia y Santa Catalina.* Ediciones archipiélago. San Andrés Isla, Colombia.
- Yunis, J. J., Acevedo, L. E., Campo, D. S., & Yunis, E. J. 2005. Population data of Y-STR minimal haplotypes in a sample of Caucasian-Mestizo and African descent individuals of Colombia. *Forensic Sci. Int. Genet.* 151:307-313.

Tabla suplementaria 1. Frecuencias haplotípicas de 15 Y-STR para los individuos raizales de la isla de Providencia (N=23).

Providencia Island																	
HAPLOTYPE	N	DYS19	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS437	DYS438	DYS439	DYS448	DYS456	DYS458	DYS635	Y-GATA-H4	Frequency
H001	1	13	12	28	25	12	13	13	15	12	12	19	15	17	25	13	0.0435
H002	3	14	12	28	25	12	13	13	15	12	12	19	15	17	24	12	0.13
H003	1	14	12	28	25	11	13	13	15	12	12	19	15	17	24	12	0.0435
H004	1	17	13	30	21	11	11	14	14	11	12	21	17	16	22	11	0.0435
H005	1	17	13	31	21	10	11	14	14	11	13	22	16	16	21	11	0.0435
H006	1	15	13	29	25	10	11	13	14	11	11	20	17	16	23	12	0.0435
H007	5	16	13	30	21	10	11	14	14	11	12	21	15	19	20	11	0.217
H008	1	16	12	30	21	10	11	14	14	11	12	21	15	19	20	11	0.0435
H009	1	15	13	31	21	11	13	13	14	11	11	21	16	16	21	11	0.0435
H010	1	14	13	29	24	10	13	13	15	12	11	19	15	17	23	12	0.0435
H011	1	16	13	30	22	11	11	14	14	11	12	21	17	16	22	11	0.0435
H012	1	13	13	30	24	10	11	13	14	10	11	20	17	16	21	12	0.0435
H013	1	14	13	29	24	11	13	13	15	12	12	19	16	18	23	11	0.0435
H014	1	14	13	29	24	10	13	13	14	10	13	19	15	16	23	13	0.0435
H015	1	14	13	29	24	10	13	13	14	10	13	19	15	17	23	13	0.0435
H016	1	15	13	29	25	10	11	13	14	11	12	20	17	16	23	12	0.0435
H017	1	17	13	30	21	10	11	14	14	10	12	21	14	17	21	13	0.0435

Tabla suplementaria 2. Frecuencias haplotípicas de 15 Y-STR para los individuos raizales de la isla de San Andrés (N=31).

San Andrés Island																Y-GATA-H4	Frequency
HAPLOTYPE	N	DYS19	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS437	DYS438	DYS439	DYS448	DYS456	DYS458	DYS635		
H001	1	16	13	31	21	10	11	13	14	13	12	21	15	18	21	13	0.0323
H002	2	14	13	29	24	10	13	13	15	12	13	19	16	17	23	12	0.0645
H003	1	14	13	29	24	10	13	13	15	12	10	19	16	17	23	12	0.0323
H004	1	17	13	29	23	10	12	15	15	10	12	20	17	16	20	11	0.0323
H005	1	14	14	30	24	11	13	13	15	12	12	18	16	18	23	12	0.0323
H006	2	13	13	29	23	10	9	12	15	9	12	20	15	16	22	12	0.0645
H007	1	16	14	32	21	11	11	14	14	11	12	21	15	16	21	12	0.0323
H008	1	14	12	28	23	12	9	13	15	12	12	19	16	17	24	12	0.0323
H009	1	16	13	30	21	10	13	14	14	11	10	20	16	16	24	11	0.0323
H010	1	14	13	30	24	11	11	13	15	12	10	18	16	16	23	13	0.0323
H011	1	14	13	30	24	11	13	13	15	12	10	18	16	16	23	13	0.0323
H012	1	14	12	29	22	10	11	13	17	10	12	20	15	14	19	11	0.0323
H013	2	14	13	29	23	10	13	12	15	12	12	19	15	18	23	11	0.0645
H014	1	15	13	31	21	10	11	13	14	11	12	21	18	16	21	11	0.0323
H015	1	16	13	31	21	10	11	13	14	12	11	21	15	16	21	12	0.0323
H016	1	17	13	31	21	10	11	13	14	12	11	21	15	16	21	12	0.0323
H017	3	17	13	30	21	10	11	14	14	11	13	21	16	16	22	11	0.0968
H018	1	13	13	30	21	10	11	14	14	11	10	21	16	16	22	11	0.0323
H019	1	14	13	29	24	11	13	13	15	12	12	19	16	18	23	11	0.0323
H020	1	15	12	30	21	10	11	13	14	11	12	21	15	17	20	12	0.0323
H021	1	16	13	30	21	10	11	15	14	11	13	21	16	16	21	11	0.0323
H022	1	15	12	30	21	10	13	13	14	11	13	21	18	18	20	13	0.0323
H023	1	15	12	30	21	10	11	13	14	12	12	21	15	17	21	12	0.0323
H024	1	14	12	28	23	10	17	13	15	12	10	19	18	15	23	11	0.0323
H025	1	13	13	31	21	10	9	13	14	11	12	21	18	18	23	12	0.0323
H026	1	15	13	31	21	10	11	13	14	11	12	21	16	18	23	12	0.0323

Capítulo No. 5: Cinco propuestas de unificación entre genética de poblaciones y genética forense para la próxima década.

William Usaquén¹, Andrea Casas¹, Madelyn Rojas¹, Döbereiner Chala¹, Ángela Alonso¹, Jenny Blanco¹, Verónica Rocha¹, Jenny Barrera², Julieth Castiblanco², Luis Fernando García².

¹ Instituto de Genética Universidad Nacional de Colombia, ² Departamento de Biología. Universidad Nacional de Colombia.

Resumen.

Colombia es uno de los países más activos en el desarrollo de análisis genéticos de filiación debido a la naturaleza de nuestra constitución, que proclama los derechos de los menores sobre los derechos de los adultos y en la que se establece la asignación de una personalidad jurídica como uno de los principios fundamentales. Por otra parte, nuestra constante guerra civil, en la que se presentan diferentes actores del conflicto, ha dejado aproximadamente 35.000 desaparecidos en las dos últimas décadas. Según establecen las leyes colombianas, la identificación de las personas fallecidas durante el conflicto armado sería de prioridad nacional, como se expresa en la ley 1408 de 2010, por la cual se rinde homenaje a las víctimas del delito de desaparición forzada y se dictan medidas para su localización e identificación. Esto ha generado un desarrollo tecnológico a nivel nacional, que es considerablemente superior respecto a la región andina. En total, en el país existen aproximadamente 25 analizadores genéticos, que tienen funciones relacionadas con identificación y filiación humana, es un número alto, lo que lleva a pensar sobre el volumen de tipificaciones realizadas en el país, y el subsiguiente manejo de las mismas. Un cálculo modesto, es de aproximadamente trescientas mil tipificaciones en los últimos 15 años, sin embargo, a pesar del desarrollo tecnológico y la implementación de nuevos y cada vez más sofisticados marcadores genéticos, no se ha seguido el mismo desarrollo teórico. En los diferentes artículos científicos producidos, ha sido sobresaliente la ausencia de métodos de análisis robustos a nivel poblacional. Otro factor preocupante es la ausencia de programas de formación y capacitación equilibrados entre biología molecular, estadística, informática y

genética de poblaciones. En este trabajo, se realizara primero una breve revisión de nuestro pasado reciente, posteriormente se realizará una propuesta en 5 puntos que deberían ser fundamentales para la unificación de los estudios en genética de poblaciones y en genética forense para el país.

Keywords.

STRs, Paternity, Colombia, Data Bases, genetic Demography.

1. El origen de los estudios en Colombia.

Nadie hubiera imaginado que la ley 75 de 1968, con la cual se da estructura orgánica sobre el derecho familiar y el menor, creando el Instituto Colombiano de Bienestar Familiar (ICBF), tendría un papel determinante en la genética colombiana. Inicialmente, en esta ley se establecen las normas de protección a los menores de edad, indicando la manera como deberían proceder las autoridades frente a las diferentes situaciones sociales, y el abordaje legal conducentes a la asignación de la paternidad. Aunque los primeros casos de filiación fueron realizados desde 1972 por el ICBF, es hasta con el decreto 2388, donde se establece la práctica de los exámenes antropoheredobiológicos, por el Laboratorio de genética del ICBF, a solicitud de la autoridad competente. Estos exámenes se realizaban en forma simultánea a las personas involucradas en el proceso. Con el avance de la tecnología, se fue pasando de las pruebas antropoheredobiologicas, que implicaban medidas morfométricas, a los grupos sanguíneos; lo que para el año 1972, era un gran avance. Durante dos décadas se fueron realizando nuevos desarrollos hasta llegar a los marcadores moleculares actuales. Pasaremos a revisar, que pasa hoy en día con esta problemática, pero antes debemos revisar más en detalle algunos elementos históricos.

1.1. Semblanza Histórica.

El primer estudio genético poblacional realizado en el país, con el objetivo de determinar los principales grupos poblacionales, fue el desarrollado por el profesor Emilio Yunis Turbay en los años

70, a partir de los datos de las pruebas de filiación del Instituto Colombiano de Bienestar Familiar. En este trabajo se evidenció un primer intento por evaluar la distribución de frecuencias de marcadores eritrocitarios en los diferentes departamentos. Este estudio se realizó con muestras de aproximadamente 70,000 personas, lo que implica un enorme esfuerzo de análisis, organización y colección de datos (Sandoval, 1982; López, 1993). Para ese momento, no se habían implementado en el país las herramientas informáticas que permitieran un adecuado almacenamiento de datos perdurables en el tiempo, ni tampoco las técnicas estadísticas multivariadas de clasificación.

En 1995, la Universidad Nacional, en cabeza de la profesora Indiana Bustos Bustos, realizó la sistematización de información de grupos sanguíneos del ICBF desde 1972 hasta 1995, tendiente a establecer valores de referencia actualizados a su momento con los cuales se daría inicio a un proceso descentralizado para el servicio de filiación en el país. Para ese momento, se planteó el uso de los índices y probabilidades de paternidad. Estos cálculos se basaban en los valores de referencia poblacionales establecidos, dando una perspectiva completamente diferente a la identificación en el país.

El éxito de los marcadores STR's en la práctica forense, y los primeros reportes de variantes polimórficas a lo largo del mundo, crearon grandes expectativas en el país, es así como en 1997 se empezaron a traer los primeros analizadores genéticos, creando la necesidad de establecer valores de referencia de frecuencias alélicas para los nuevos marcadores moleculares, en los que se implementarían los estadísticos de identificación y filiación. Hasta ese momento, los valores de referencia provenían de los reportados por las casas comerciales, naturalmente ignorando las variantes alélicas propias de nuestras poblaciones. En la Universidad Nacional de Colombia, se adelantaron los primeros trabajos tendientes a determinar las frecuencias genéticas para la población colombiana con sistemas microsatélites. Estos mismos estudios fueron desarrollados también por el doctor Manuel Paredes, del Instituto Nacional de Medicina legal y Ciencias Forenses. En esos primeros estudios se pudieron detectar diferencias en las frecuencias alélicas reportadas por ambos estudios, que correspondían a dos muestreos completamente distintos sobre una población en la que no se había realizado ningún proceso de caracterización genética; bajo estas circunstancias, es imposible determinar si se trata de errores técnicos, debidos al muestreo, a las tipificaciones, o realmente a

procesos de la historia evolutiva de las poblaciones. En ese primer inicio, las muestras eran tomadas por muestreos a conveniencia, teniendo como marco los primeros casos de filiación realizados.

Contando con estos primeros valores de referencia, se inicia la importación de varios tipos de analizadores genéticos tales como: el ALF express de Pharmacia, Hitachi FMBIO I y II, o el analizador ABI 310 de Applied Biosystems. Todo esto permitió una mayor capacidad para estudios de filiación, y de manera indirecta los estudios genético poblacionales. La mayoría de estos estudios desarrollados para el país se han hecho utilizando entre 200 y 300 muestras aproximadamente, siguiendo los lineamientos de las comisiones extranjeras para el estudio o desarrollo de pruebas de ADN. Por otra parte, la totalidad de estos estudios corresponden a trabajos realizados con muestras provenientes de pruebas de ADN en casos de paternidad, los cuáles no utilizan métodos aleatorios y por tanto no aseguran una condición de equiprobabilidad. A la fecha, Colombia cuenta con aproximadamente 40 reportes de frecuencias, en los cuales el nivel de organización básico corresponde a departamentos, sin embargo, esta clasificación sociopolítica podría no ser la correcta para la interpretación de la estructura del país y es por tanto necesaria una reevaluación.

Este panorama no solo aplica a marcadores microsatélites autosómicos, en nuestro país se han realizado una gran cantidad de estudios genético poblacionales que utilizan la información genética uniparental (DNA mitocondrial y Cromosoma Y). Entre algunos trabajos ya publicados se encuentran: Mesa et al. (2000); Carvajal-Carmona et al. (2000); Keyeux et al. (2002); Rodas et al. (2003); Bedoya et al. (2006); Salas et al. (2008) y más recientemente Rojas et al. (2010). Todos ellos, con propósitos históricos en relación al patrón y dinámica de flujo génico entre diferentes poblaciones, diferenciando esta contribución por sexo. Adicionalmente, desde las perspectivas médica y epidemiológica, se pretende realizar un aporte al conocimiento de la historia demográfica y dinámica de mezcla de las poblaciones, que sirva para el mapeo genético de rasgos complejos en los estudios de asociación clínica (Rojas et al., 2010; Bedoya et al., 2006; Carvajal-Carmona et al., 2000; Mesa et al., 2000). Si bien dichos estudios han realizado aportes importantes al conocimiento de la historia biológica de las poblaciones humanas en Colombia; estos presentan limitaciones en cuanto a criterios de inclusión, tamaño y procedencia de las muestras, y como consecuencia, también las conclusiones a las que se han podido llegar con estos trabajos son restringidas.

1.2. Un cuestionamiento sobre la actualidad.

En este momento estamos dando pasos hacia la utilización de nuevos marcadores moleculares con una gran capacidad de individualización, y estamos en la carrera de implementar nuevas tecnologías. Sin embargo, el desarrollo de sistemas de información conjunta, procedimientos matemáticos y estadísticos tendientes a desarrollar nuevas metodologías, y la evaluación de los tamaños de muestra, aún siguen en el mismo punto de los primeros estudios de 1998. Pero quizás, lo más preocupante es que nuestro nivel de capacitación en matemáticas, estadística y genética de poblaciones, sigue en franco desnivel con el avance tecnológico.

La ausencia de bases de datos es uno de los temas de mayor preocupación, ya que no existe un criterio unánime para el almacenamiento de la información genética. Debido a esto, cada estudio toma datos de tipificaciones y variables diferentes, criterios poblacionales disímiles, con diversas concepciones sobre muestreo; lo que conlleva a la aplicación de técnicas de toma de datos incoherentes. Un elemento adicional, un tanto complejo de analizar, es el desarrollo de trabajos de grado en su mayoría que se reducen a la aplicación de técnicas moleculares, sin ningún tipo de análisis ni de implicaciones estrictas sobre la genética de poblaciones. Además, los estudios parecen corresponder a un modelo repetitivo de investigación sin mayor cuestionamiento.

En la siguiente parte del artículo veremos lo que pasa en el resto del mundo. Haremos una revisión de aspectos teóricos a tener para el desarrollo de la genética de poblaciones.

2. Fuentes de conocimiento de la Genética de poblaciones.

La genética de poblaciones se fundamenta en tres disciplinas para su desarrollo epistemológico: la biología molecular, encargada de las técnicas de laboratorio para el análisis y obtención de nuevos marcadores; los métodos estadísticos y bioinformáticos, cuyo objetivo es el de analizar los grandes conjuntos de datos que se producen actualmente, tratando de encontrar nuevos métodos y procedimientos analíticos; y finalmente las metodologías de campo, guiadas por el análisis demográfico, que permiten realizar una conexión entre la información genética, cultural, histórica y ambiental de la población en estudio; esta última tendencia además permite el uso de herramientas

provenientes de la genética clásica.

2.1. Biología Molecular.

El desarrollo tecnológico en los análisis genéticos de filiación para esclarecer casos de paternidad, con el fin de proteger los derechos de los menores y las desapariciones forzadas en actos violentos, se han convertido en una prioridad para el país.

Desde 2005, la Fiscalía viene realizando una tarea de exhumación de víctimas del conflicto colombiano, permitiendo la recuperación de 1903 cadáveres (hasta el 20 de noviembre de 2008), de los cuales tan sólo el 16% (304) de los cuerpos han sido identificados y entregados a sus familiares, dando cumplimiento al artículo 44.4 de la Ley de Justicia y Paz el cual proclama: “*La búsqueda de los desaparecidos y de los restos de personas muertas, y la ayuda para identificarlos y volverlos a inhumar según las tradiciones familiares y comunitarias*”.

La identificación de desaparecidos, restos cadavéricos y solución de casos por medio de técnicas de análisis de polimorfismos de ADN, se ha convertido en una actividad frecuente dentro de la rutina pericial de muchos laboratorios de identificación genética, las fases analíticas de estas pruebas se basan en fundamentos científicos lo suficientemente rigurosos que requieren el cumplimiento de condiciones específicas de calidad durante el procedimiento técnico (Begoña et al., 2005).

Las investigaciones tecnológicas de los marcadores moleculares han sido enfocadas, precisamente, hacia el perfeccionamiento de procesos aptos para la resolución de casos a partir de muestras difíciles, por lo que se observa un gran desarrollo tecnológico en las herramientas de biología molecular. En 1985 Alec Jeffreys introdujo la “huella genética” empleando una técnica rudimentaria y dispendiosa; contrario a esto, hoy en día contamos con facilidades tecnológicas empleando sistemas *multiplex* que nos permiten tipificar e individualizar a una persona usando marcadores de los cromosomas autosómicos, sexuales y del ADN mitocondrial, a bajos precios y en menor tiempo.

Los principales marcadores con los que se logró analizar varios loci simultáneamente, fueron los STRs (*Short Tandem Repeat*), para los que se requiere una mínima cantidad de muestra y cuyo procesamiento es rápido y automático empleando kits comerciales (Yunis et al., 2002). Para constituir

los datos de las frecuencias poblacionales, inicialmente el FBI (*Federal Bureau of Investigation*) de Estados Unidos, implementó el uso de una base de datos introduciendo 13 marcadores tipo STR llamada CODIS (*Combined DNA Index System*). En muestras parcialmente degradadas se han introducido los miniSTR, los cuales poseen una reducción del tamaño del producto amplificado generando un mayor número de casos resueltos (Butler et al., 2003).

Hoy en día, uno de los marcadores novedosos más prometedores en la práctica forense, son los polimorfismos simples de un solo nucleótido (SNP). Existen ya en la actualidad multiplex de SNPs validados para uso forense (Butler, 2000), permitiendo resolver casos complejos de paternidad y especialmente en criminalística e identificación. Los SNPs han sido considerados como marcadores genéticos ante la comunidad forense por varias razones: la primera y más importante, es que los productos de PCR pueden ser menores a 100pb, lo que significa que estos marcadores tienen la habilidad de recobrar información de muestras de ADN altamente degradadas, mejor que los STRs, los cuales generan amplicones de tamaño superior (300-400pb). En segundo lugar, se pueden manejar set multiplex con una mayor cantidad de regiones analizadas que no estén condicionadas por el número de fluorocromos o rango de tamaño. Finalmente, el procesamiento de la muestra y el análisis de los datos pueden ser automatizados rápidamente debido a que no es necesaria una separación basada en tamaño. Para el campo forense es un reto la necesidad de analizar un gran número de SNPs para que sean informativos y obtener un poder de discriminación razonable. Sin embargo, ya se han realizado algunos multiplex que permiten analizar 50 SNPs simultáneamente, generando un perfil individual, y por tanto, permitiendo la obtención de un poder de discriminación que le facilita competir con los marcadores microsatélites.

Otros marcadores que están empezando a ser analizados e implementados en las ciencias forenses son los polimorfismos generados por la inserción o delección de uno o más nucleótidos, conocidos como *indels*(DIPs). (Weber et al. 2002). Estos marcadores combinan las características óptimas de los SNPs y los STRs como: 1. Se extienden por todo el genoma humano (cromosomas autosómicos y sexuales); 2. Este polimorfismo se deriva de un evento mutacional único; 3. Muestran diferencias significativas en las frecuencias alélicas entre los grupos de población geográficamente separados, por lo tanto, pueden ser utilizados como marcadores informativos ancestrales; 4. Los DIPs generan amplicones

cortos con un máximo de 150pb, lo cual genera beneficios para manejar muestras forenses altamente degradadas; y5. Poseen una mayor capacidad de exclusión que los SNPs y STRs, con respecto a la identificación forense o el análisis de la paternidad (Weber et al. 2002; Pereira et al., 2009). Esto demuestra que los recientes avances en genética forense se han centrado en el desarrollo de ensayos de genotipificación con amplicones cortos, con el fin de mejorar la amplificación para la obtención de perfiles genéticos completos especialmente en muestras degradadas. Además, la genotipificación rápida y rentable con estos marcadores puede ser utilizada para distinguir los principales grupos poblacionales, especialmente los europeos, africanos y americanos nativos. Los *indels* también permiten identificar subestructura en poblaciones mixtas. Actualmente existen paneles de 48 marcadores DIPs que se han empleado en el análisis poblacional, demostrando que en poblaciones mixtas de diferentes grupos ancestrales los *indels* permiten estimar, de manera precisa y fiable, la mezcla interétnica individual y global en relación con los grupos (Weber et al. 2002). Por tanto estos marcadores son altamente competentes y útiles en los campos de análisis forense y poblacional.

Como vemos, los avances en biología molecular han dado grandes saltos tecnológicos, ofreciendo cada vez, un mayor número de métodos automatizados, rápidos, robustos, y económicos para las identificaciones humanas; por lo que en un futuro, sin duda, se podrá secuenciar a gran escala cromosomas enteros o todo el genoma (Xue et al., 2010). El desarrollo de tecnologías de ultrasecuenciación con diseños bioquímicos particulares, induce a pensar que sus potenciales aplicaciones forenses no serán ciencia ficción (Carracedo et al., 2010).

2.2. Matemáticas y Estadística.

Ahora es posible la unión de la biología molecular, aplicada con herramientas de alta tecnología; junto con elementos matemáticos, estadísticos e informáticos. Durante varias décadas, los aportes de naturaleza matemática solamente constituyeron modelos teóricos lejanos a los conceptos biológicos, debido a que no se contaba con la tecnología, las herramientas informáticas ni el conocimiento sobre el genoma humano para realizar esta integración.

Muchas de las disciplinas de la estadística hoy en día tienen que ver con el desarrollo de la genética de poblaciones: A partir de los estadísticos multivariados es posible para nosotros evaluar múltiples

genes y variantes alélicas simultáneamente, para determinar las diferencias y similitudes entre poblaciones, por otra parte, hoy en día es posible gracias al uso de la teoría de probabilidad realizar particiones en espacios muestrales y asignar la probabilidad de aparición de determinadas variantes alélicas, haplotipos y haplogrupos. También es posible la utilización de teoremas de probabilidad, que permiten evaluar la aparición de un determinado paisaje fenotípico teniendo presentes diferentes eventos condicionales. Hoy la genética también se constituye en una herramienta de desarrollo constante la estadística no paramétrica, y también plantea problemas fundamentales a las metodologías y técnicas de muestreo establecidas desde las escuelas tradicionales de pensamiento. En los próximos 50 años, es posible que muchas de las herramientas matemáticas y estadísticas paramétricas actualmente en uso cambien y sean reevaluadas, gracias al desarrollo que han tenido por los constantes cuestionamientos de las disciplinas biológicas. Una de los cambios más determinantes en la genética moderna es el paso de estimadores puntuales a estimadores por intervalos.

Las herramientas matemáticas y estadísticas, asociadas a la genética de poblaciones, han permitido evaluar con mucha mayor precisión la magnitud de las fuerzas de cambio evolutivo, la forma en la que afectan a una población, el grado de las interacciones existentes entre ellas, y han permitido, también, cuantificar y ordenar los diferentes procesos que afectan a las poblaciones. Uno de los elementos que hoy en día vive un gran cuestionamiento es el tamaño de muestra necesaria en un estudio genético poblacional, de tal manera que las conclusiones evolutivas obtenidas a partir de un número limitado de individuos puedan ser extrapoladas a la totalidad de la población. Una revisión detallada de los diferentes estudios genéticos poblacionales, deja ver una gran disparidad en los tamaños de muestra, incluso en las mismas regiones geográficas. La consecuencia inmediata de este fenómeno es la obtención de estimadores genético poblacionales en un rango amplio, que no permiten ver claramente cuáles son los efectos de las fuerzas de cambio evolutivo, y cuáles son variaciones debidas a las disparidades propias de un margen de error por causas del procedimiento establecido para la obtención de las muestras. La comisión internacional que regula los estudios con ADN a lo largo de los últimos 15 años, ha emitido diferentes comunicados, en los cuales se propone cada vez más un mayor número de muestras, sin embargo, no existen estudios estadísticos apropiados que nos indiquen cuáles un número aproximado de muestras a tomar en un estudio genético poblacional.

En este punto, es probable que tengamos la necesidad de desarrollar nuevas herramientas matemáticas y estadísticas, y nuevos algoritmos con lógicas de procedimiento y procesamiento diferentes a los tradicionales, ya que el muestreo dentro de la investigación genético-poblacional, en relación con los métodos tradicionales, introduce una serie de elementos hasta ahora poco utilizados tales como la naturaleza misma del marcador a estudiar. De esta manera, algunos elementos que pueden ser determinantes en el proceso de establecer un adecuado tamaño de muestra son: la identidad de la fuente de información genética; si el marcador es autosómico, es el cromosoma Y o el ADN mitocondrial, que corresponden a linajes uniparentales; ó si corresponde a marcadores sexuales, tales como el cromosoma X, que tiene una serie de particularidades en el proceso de recombinación; o, por ejemplo, el nivel de polimorfismo del marcador, si éste es de naturaleza neutral o si está asociado a una serie de marcadores con acentuados procesos de selección fenotípica. Por otra parte, procesos genético poblacionales relacionados con la historia de poblamiento, la distancia en el tiempo, los cuellos de botella, el nivel de endogamia que pueda sufrir una población, los diferentes modelos de flujo genético que pueden estar afectando la interacción entre poblaciones, o el nivel de barreras geográficas que han determinado las diferentes rutas de poblamiento de una región en particular; son elementos que en la práctica cotidiana y en la teoría formal no son considerados para el establecimiento de un adecuado tamaño de muestra.

2.3. El Análisis demográfico.

La información demográfica se ha aprovechado para involucrar aspectos como la estructura por edad y sexo, ó las tasas de migración, natalidad y mortalidad de las poblaciones a los análisis de la información genética colectada. Este tipo de estudios se dirigen, por ejemplo, al establecimiento de evidencias sobre la variabilidad biológica de algunos grupos que habitan en nuestro país a partir de la demografía, las ancestrías y las frecuencias alélicas de determinados marcadores moleculares (Carvajal - Carmona et al., 2003), o a la identificación de aspectos como la selección y la deriva a partir de movimientos territoriales, nacimientos y defunciones (Acreche et al., 2003). Esta información presenta a su vez la posibilidad de integrar *variables socioeconómicas y espaciales* o análisis de isonimia -tan frecuentemente utilizados por la genética clásica (Crow y Mange, 1965) que soportan desde un punto de vista temporal, espacial y social los análisis realizados, teniendo en cuenta el hecho

de que la genética de poblaciones humanas no se puede abarcar ni en tiempos ni en dinámicas poblacionales constantes o independientes de cualquier aspecto social o cultural.

Precisamente, entre los años 50 y 70, la genética clásica también aprovechaba otro tipo de *variables socioculturales* -de carácter más cualitativo y descriptivo-, que podían junto a la economía y el espacio, llevar a un alcance mucho mayor el tipo de estudios realizados en algunas poblaciones (Salzano 1977, 1979, 1980), la etnografía y la historia serían disciplinas importantes a la hora de realizar abordajes a la población, existiendo así, un trabajo de campo y una serie de métodos de investigación consecuentes, no sólo con preguntas de investigación genéticas sino también antropológicas. Sin embargo, esta tendencia cambió sustancialmente cuando los estudios genéticos empezaron a tener una mayor preferencia por los novedosos análisis a partir de marcadores moleculares, los cuales aunque se seguirían sosteniendo de algunas variables espaciales y demográficas -o en algunos casos menores socioeconómicas- para demostrar variabilidad y factores microevolutivos; no volverían a aprovechar las variables socioculturales, de la manera como se hacía para épocas anteriores; un tanto porque las disciplinas: genética y antropológica, empezarían a especializarse y a utilizar metodologías de investigación más distintas la una de la otra, y también porque los genetistas prefirieron modificar la manera de realizar sus muestreos y de definir los criterios desde los cuales se elegirían las poblaciones sobre las que se iba a trabajar.

Esta separación ha traído una serie de transformaciones a la aplicación de la información demográfica sobre los datos genéticos. Como se venía diciendo en numerales anteriores, se debe tener en cuenta que algunos estudios realizados en nuestro país, han tenido ciertas dificultades en cuanto a la organización por departamentos de las frecuencias reportadas o en los criterios de inclusión y tamaño de las muestras. Para el primer caso, por ejemplo, se ignora el hecho de que un factor, como lo es, el establecimiento de departamentos o de regiones administrativas, no puede coincidir exactamente con el devenir geográfico e histórico de las poblaciones, el cual muy probablemente va más allá de una separación hecha con líneas en un mapa, cosa que hace que nos cuestionemos respecto a la manera como se definen las poblaciones a muestrear, cuando se pretende realizar a un nivel nacional o regional, al mismo tiempo, nos hace pensar respecto a nuevas formas de definir esos criterios con los cuales se escogen las personas que harán parte de un estudio, más allá de su pertenencia a una base de

datos o a una serie de muestras almacenadas masivamente en un laboratorio. A partir de ello, podríamos considerar entonces, que la demografía integrada con otro tipo de información histórica, etnográfica, geográfica y económica dentro de estos nuevos estudios genéticos, tanto en su planteamiento como en la presentación de sus resultados; ofrecería la posibilidad de superar las dificultades antes mencionadas.

Sánchez Compadre (2001), por ejemplo, reconoce que factores microevolutivos como las migraciones, la deriva genética o los patrones de cruzamiento, están determinados por los componentes culturales de nuestras sociedades, implicando el desarrollo de otro tipo de análisis que complementen el trabajo biodemográfico; en este caso, el trabajo de campo se presenta como una oportunidad para que el genetista realice un estudio con unos criterios de selección, un muestreo y un análisis de la información genética concordante con la historia, la geografía y los aspectos culturales de una población. Parece ser que existe cierta conformidad con los simples reportes de las frecuencias alélicas encontradas o con la sistematización de bases de datos, sin contar con que la probabilidad de cometer errores puede ser mucho mayor si no se tienen en cuenta las razones por las cuales se escogió una población, los motivos de selección de los marcadores o el fundamento que sostiene el uso de determinadas ecuaciones.

Muchos genetistas en nuestro país consideran el trabajo de campo como un proceso demasiado difícil y costoso, prefiriendo seguir recurriendo a las muestras almacenadas en casos forenses o de filiación, sin cumplir con los requisitos estadísticos de la equiprobabilidad y sin tener en cuenta los errores que pueden llegar a existir en los resultados presentados, al no haber criterio alguno de selección. Sin embargo, debería considerarse a la demografía más que como una herramienta matemático/estadística para expresar las dinámicas poblacionales sobre las que se fundamentan los estudios genéticos (Grafica 1), como una valiosa fuente de información cuantitativa y cualitativa -al estar integrada con información etnográfica o histórica-, para realizar un análisis más profundo tanto de las dinámicas y estructuras poblacionales, como de las fuerzas evolutivas que inciden sobre los grupos con quienes se trabaja. Se pretende entonces dejar claro que la demografía es una fuente de conocimientos que le presenta una oportunidad a la genética de poblaciones, de integrarse con otro tipo de metodologías tanto cuantitativas como cualitativas para realizar análisis mucho más completos y potentes de la

información genética resultante en el laboratorio y en los software de análisis poblacional. Más adelante se presentará en detalle la manera como se puede aprovechar esa oportunidad y la manera como los aspectos biológicos y antropológicos que alguna vez en la época clásica hicieron parte de la genética podrían ser de nuevo de utilidad en el reconocimiento de la estructura de la población de nuestro país.

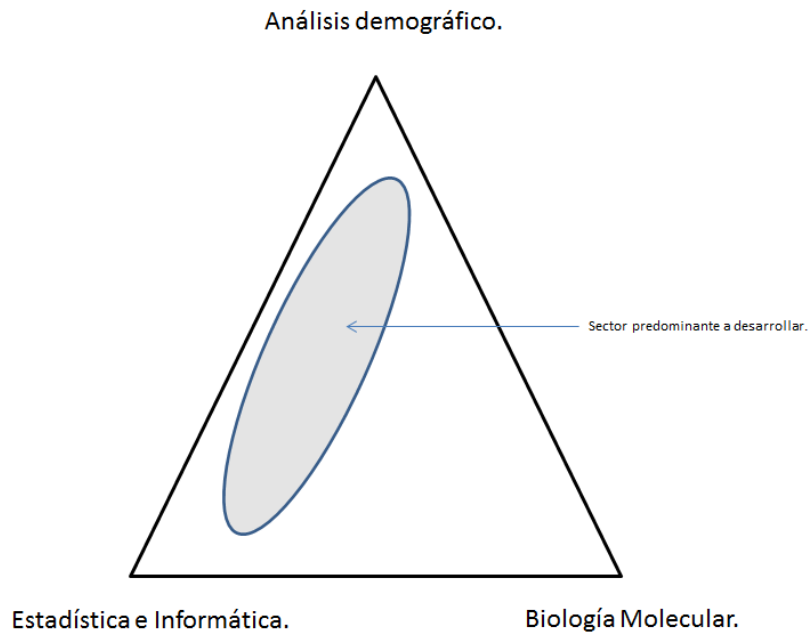


Figura No. 1: Tres fuentes del desarrollo epistemológico de la genética de poblaciones humanas. En azul, la región de importancia a desarrollar en el país.

Basados en lo anterior nos permitimos reorientar las tendencias mundiales para hacer un planteamiento a nivel local, buscando desarrollos concretos en la próxima década de investigación genético poblacional en nuestra región.

3. Cinco propuestas de trabajo.

3.1. Un análisis descriptivo de cada polimorfismo.

Los primeros valores que se obtienen de los análisis genético poblacionales son las frecuencias alélicas, normalmente calculadas por la técnica del conteo directo, a partir de ellas suelen realizarse los demás estimadores de la genética descriptiva, tales como la heterocigocidad, la condición de equilibrio HW; y los principales estimadores forenses, como por ejemplo, el poder de discriminación, el contenido de información polimórfica, y demás valores. Una de las sugerencias es iniciar con el análisis mismo del polimorfismo, empezando por la determinación de alelos efectivos. Éste es un cálculo muy sencillo que permitirá encontrar las variantes alélicas determinantes en aproximadamente el 90% de la población, o visto de otra forma, el número de variantes más frecuentes encontradas en los diferentes perfiles genéticos. Además, este valor va a indicar la verdadera variabilidad del polimorfismo. Un ejemplo sobre nueve sistemas microsatélites y la manera como estos son analizados se presenta en la tabla número uno. Este tipo de análisis simple nos deja ver desde una perspectiva descriptiva la variabilidad y el comportamiento de un polimorfismo e una población específica.

	CSF1P0	D13S317	D3S1358	D5S818	D7S820	FGA	TH01	TPOX	vWA31
No. de alelos reportados	11	11	10	10	10	26	8	8	13
No. Teórico de Genotipos Posibles	66	66	55	55	55	351	36	36	91
No. de genotipos Observados	49	44	36	43	50	128	29	34	61
% de genotipos Observados sobre el total teórico.	74.2%	66.7%	65.5%	78.2%	90.9%	36.5%	80.6%	94.4%	67.0%
No. alelos efectivos	4	5.8	4.0	4.4	4.9	8.3	4.4	3.5	4.6

% de Alelos efectivos respecto al total del sistema.	37%	52%	40%	44%	49%	32%	55%	44%	35%
No. genotipos posibles con el No de alelos efectivos	10	19.5	10.2	11.7	14.3	38.4	11.9	7.8	12.8
% genotipos observados solo con los alelos efectivos sobre el total de la población.	90%	90.20%	96.44%	88.70%	92.01%	91.56%	88.51%	87.38%	91.24%

Tabla No. 1: Análisis descriptivo de nueve polimorfismos genéticos. Algunos estadísticos y valores, que se proponen para el análisis de la utilidad de un marcador genético.

En primer término, aparece el número de variantes alélicas encontradas en la población; luego el número de variantes que son reportadas a nivel mundial para el sistema de estudio; la relación entre estos valores, nos permite observar el porcentaje de variantes sobre el total que han sido encontradas en la población de estudio. Otro porcentaje muy importante es el calculado del número de alelos efectivos sobre el número de variantes encontradas, lo que arroja el porcentaje de alelos efectivos sobre el total de la población, y esto nos permite evaluar de una mejor manera la verdadera eficiencia del polimorfismo. Es frecuente encontrar sistemas con un gran número de variantes alélicas, pero que en términos efectivos, representan un porcentaje pequeño sobre el polimorfismo. En muchas oportunidades, al discutir sobre la calidad de un polimorfismo en la práctica, se entiende que es bueno, si presenta un gran número de variantes genéticas; pero, muchas de estas variantes no son efectivas en la población. Un cálculo interesante a realizar, es el porcentaje de genotipos en la población de estudio conformados por los alelos efectivos; estos en buena parte, corresponden aproximadamente al 90% de las personas en la población. Tan sólo, un 10% restante, será propio de las variantes en baja frecuencia y esto es un indicador directo, es decir, un valor observado de la verdadera variabilidad y utilidad del polimorfismo en la población.

3.2. Una información demográfica integrada a datos históricos, etnográficos y estadísticos.

El análisis demográfico, representa una herramienta de gran valor en la clasificación de datos y en la interpretación de procesos, no sólo de migración, sino también de ancestrías y conformación histórica de las poblaciones. Esta información puede tomarse a través de una encuesta como la presentada en la Gráfica No. 2.

UNIVERSIDAD NACIONAL DE COLOMBIA SEDE BOGOTÁ

PROYECTO: "CARACTERIZACIÓN GENÉTICA DE LA POBLACIÓN HUMANA DEL DEPARTAMENTO DE LA GUAJIRA COMO PARTE DEL ENTENDIMIENTO HISTÓRICO Y ESTRUCTURAL DE LAS ETNIAS COLOMBIANAS"

1. Fecha: DD 30 Junio/09
 Código de la Mue.: 60090320
 Institución Toma mue.: H.N.S.B.
 Persona Toma Mue.:
 Persona Toma Encu.: M.R.

2. Localización de toma de Muestra: Municipio RIOHACHA Loc. Específica HOSPITAL

3. Información Personal: Lugar de Residencia RIOHACHA Tiempo de Residencia 23
 Dirección ND Teléfono ND C-Electrónico NA
 Primer Nombre [Redacted] Segundo Nombre NA

4. Apellidos:
 P1 MARTINEZ LN RIOHACHA POP. MESTIZO
 P2 NAVARRO LN PALOMINO POP. MESTIZO
 M1 MUNOZ LN S.J. CESAR POP. MESTIZO
 M2 DOMINGUEZ LN S.J. CESAR POP. MESTIZO

7. Edad 52 LN RIOHACHA POP. MESTIZO

8. Edad 49 LN SAN JUAN DEL CESAR POP. MESTIZO

5. Edad 23 LN RIOHACHA POP. MESTIZO
 Edad 20 LN BARRANQUILLA POP. MESTIZO

6. Etnia: NA Costa NA Ranchería NA
 Edad/facimientto 1° hijo NA No. hijos NA
 Wayunaiki Habla SI No. Escribe SI No.

9. POP. INDG, AFR, MEST, ARB, MIG

Figura No. 2: Modelo de encuesta genealógica utilizada para complementar los muestreos por conveniencia, haciéndolos independientes de los criterios del investigador.

Habitualmente, en los análisis poblacionales hasta ahora realizados en el país, la única variable con la que se cuenta, es el lugar de nacimiento; que está expuesta a un gran número de problemas, ya que no refleja necesariamente la ancestría de una persona. Precisamente, con el uso de la encuesta, se podrán

detallar no sólo los lugares de nacimiento de la persona encuestada, sino también una serie de informaciones familiares que nos podrán llevar hacia atrás en el tiempo y en el espacio, definiendo de esta manera, los procesos migratorios de las diferentes familias participantes en el muestreo y las configuraciones familiares desde las cuales se pueden emprender estudios de ancestría y de consanguinidad. Esta información permitirá a su vez, la integración con una serie de datos históricos y etnográficos, que podrán validar o describir a mayor profundidad la serie de eventos registrados en la encuesta; con miras, no sólo hacia la vinculación con los análisis genéticos, sino hacia la definición de los criterios de selección de poblaciones para el muestreo, teniendo en cuenta elementos estructurales para un buen cálculo de tamaño de muestra (Grafica No. 3), mucho antes del análisis de laboratorio.

El uso de encuestas, va a permitir además, la clasificación y ordenación de los principales grupos que habitan una región geográfica específica, bajo unos criterios que no se ciñen exclusivamente al espacio o a la división política de un territorio; sino también, a lo que la información histórica, económica y etnográfica logre definir como grupo poblacional. La mayoría de veces, como se mencionó anteriormente, los individuos de una población son clasificados sólo por su lugar de nacimiento y así ingresan a los archivos de datos. Sin embargo, la posibilidad de clasificar a los individuos no por poblaciones geográficas, sino por poblaciones de naturaleza ancestral, justificando esta división con información tanto cuantitativa como cualitativa, logrará una aproximación más precisa de los principales grupos ancestrales y de una serie de características socioeconómicas, espaciales y culturales que componen una población humana en particular.

Aplicaciones que utilizan metodologías bayesianas para la asignación de individuos a poblaciones, como el software Structure 2.3.3 (Pritchard et al., 2000), son ampliamente utilizados, ya que arrojan un valor k correspondiente a los grupos poblacionales presentes en la muestra de estudio; pero, estrictamente, esto debe entenderse como el número de núcleos de inercia o diferentes conformaciones genotípicas presentes en una población. Cuando se ingresan los datos, ordenados por un criterio de ancestría, se tiene mayor posibilidad de encontrar correlación entre los grupos definidos por el software Structure y los núcleos poblacionales descritos al inicio del experimento (la encuesta permite hacer un análisis a priori a la clasificación). Por lo tanto, el análisis demográfico, nos va a

permitir verificar un modelo *a priori* de grupos poblacionales presentes en una región geográfica, y compararlo con el modelo matemáticamente construido *a posteriori*; como un proceso de verificación del primer modelo. En consecuencia, no estamos restringidos estrictamente a los modelos matemáticos, sino que éstos se convierten en una verificación de la información obtenida por clasificación demográfica de la población a estudiar genéticamente; es una manera de tener un segundo punto de control respecto a los análisis matemáticos que se realizan posteriores al análisis en biología molecular. Frecuentemente del laboratorio, se pasa directamente al análisis informático sin contar con un segundo punto de referencia, que permita inferir si estos análisis matemáticos son o no acertados. En el mejor de los panoramas, se realizan diferentes metodologías matemáticas y estadísticas, que se contrastan entre sí, para buscar la coherencia de la información obtenida, con el análisis genealógico. Así, la clasificación de las poblaciones *a priori* antes del análisis en biología molecular, se convierte en una poderosa herramienta que da una amplia perspectiva al análisis evolutivo.



Figura No. 3: Elementos estructurales para un tamaño de muestra.

3.3. Genética clásica, isonimia.

El advenimiento de las técnicas de biología molecular, y el análisis informático sobre el genoma han relegado casi hasta la desaparición muchas de las herramientas del análisis genético clásico, entre ellas los estudios de isonimia. Los apellidos son una característica cultural y la principal forma en la que se identifican relaciones de parentesco entre las personas de una población (Gómez, 2004). En la civilización occidental, se generaron entre la edad media y comienzos de la edad moderna, cuando se hizo necesaria la identificación precisa de las personas con fines sociales y demográficos (Giraldo, 2001). Los colonizadores de España y otras partes de Europa llevaron sus respectivos sistemas de nombrar al Nuevo Mundo, donde acabaron por establecerse tanto entre los supervivientes de los pueblos autóctonos como entre el resto de los recién llegados (Lasker, 1991).

Los apellidos son usados en estudios de biología humana debido a que en muchas culturas, se heredan como los genes; en cada generación, los descendientes tienen la mitad de los apellidos de sus padres y la mitad de sus genes autosómicos (Lasker, 1980). En nuestra civilización, los apellidos se heredan por vía paterna y en el caso de la descendencia masculina, se asemejan a alelos neutros de un gen transmitido por el cromosoma Y (Yashuda & Furusho, 1971; Zei *et al.*, 1983). Con relación a éste evento bio-antropológico como bien lo dicen Pineda *et al.*, (1999), la Isonimia, definida de forma muy general como el estudio de concordancia en los apellidos (iso= igual, nimia=nombre), ha surgido como una herramienta para analizar poblaciones humanas y permitir una aproximación inicial a la estructura genética. Sanz *et al.*, (2003) se refieren a la isonimia (o isonomía) como al estudio de la frecuencia y distribución de apellidos en poblaciones humanas, mediante el cual pueden establecerse relaciones de parentesco y origen. Estudios como el de Giraldo (2001) se refieren a la isonimia como al análisis de los matrimonios de parejas que comparten apellidos, éste concepto está muy relacionado con el establecido por Crow y Mange (1965), quienes la definieron como la frecuencia de matrimonios en los que ambos cónyuges tienen el mismo apellido. Estos autores también sentaron las bases teóricas y los supuestos básicos (Román *et al.*, 2007) para el desarrollo del método isonímico:

- 1) El apellido es hereditario. Es decir, que en la población de estudio, los apellidos han

de ser transmitidos de forma regular de generación en generación y sin modificación desde su creación.

- 2) No hay desproporción numérica en relación a los dos sexos.
- 3) El apellido es monofilético

El uso de apellidos en estudios genéticos data de 1875, cuando George Darwin los usó para estimar la frecuencia de matrimonios entre primos hermanos (Jobling, 2001). Los primeros estudios en recurrir al método isonímico se dieron en el campo de la biología en la década de 1960. Según Lasker (1991), en 1960 Richard Shaw publicó un breve artículo que no hacía referencia a ninguna bibliografía anterior. En él hacía notar que el uso de los apellidos según el sistema español, significaba que los dos miembros de un matrimonio entre primos hermanos comparten un apellido. Shaw señalaba que éste hecho podría ser útil a la hora de distinguir la consanguinidad en los países de lengua española. Sin embargo, la mayoría de publicaciones hacen referencia al trabajo de Crow y Mange, quienes publicaron en 1965 un estudio sobre la estimación de la consanguinidad de las poblaciones por medio de la isonimia (Gómez, *et al.*, 2008).

Los análisis de isonimia se han usado básicamente para determinar tasas de migraciones, relaciones de ancestría y origen de las poblaciones, grado de estructuración genética (Bedoya, *et al.*, 2006), resolución de algunos casos forenses (King & Jobling, 2009), diferenciación étnica (Mateos, 2007); para evaluar procesos de mestizaje (Gottlieb, 1983) y relacionar la incidencia de algunas enfermedades hereditarias con posibles consecuencias del efecto fundador de las poblaciones (Bedoya *et al.*, 2006). La utilidad del método se fundamenta en la rápida obtención de datos, que conllevan a la identificación de características básicas de la población, de la misma forma que se haría mediante el uso de herramientas más costosas y de difícil acceso. Las fuentes de información para obtener datos de apellidos son diversas; según Costa *et al.*, (2000) se han utilizado registros religiosos y civiles de bautismo, matrimonios, defunciones, nacimientos y directorios telefónicos. Para estos autores, la utilización de registros de defunciones para analizar la estructura genética de una población, no es frecuente, y esto podría ser una desventaja. Sin

embargo, Lasker (1969) en un estudio del pueblo pesquero de San José (Perú), empleó cuatro fuentes de datos: entrevistas, lápidas funerarias, registros de nacimiento y registros de defunción; mostrando que no había diferencias estadísticamente significativas entre las fuentes de datos usadas (Lasker, 1991).

En Colombia, las fuentes empleadas corresponden a: bases de datos para pruebas de paternidad, de casos del Instituto Colombiano de Bienestar Familiar, ICBF (Giraldo, 2001; Gómez, 2004), y del Instituto Colombiano de Medicina Legal (Gómez, Ávila y Briceño, 2008); registros de usuarios del Sisbén (Bedoya *et al.*, 2006; Gómez, Hinestroza y Muñetón, 2008), registros de nacimiento (Pineda, Arcos y Bravo, 1999) y parroquiales de matrimonio (Soto *et al.*, 2001).

En la literatura se han descrito diversos coeficientes, los cuales pueden llevar a confusión según la modalidad de traducción y según la letra latina o griega utilizada en su formulación (Giraldo, 2001). A partir de la frecuencia de apellidos es posible calcular estimadores como: la isonimia aleatoria sesgada (Scapoli *et al.*, 2007) y no sesgada (Rodríguez-Larrade, *et al.*, 1993), el coeficiente de endogamia total (Malnar, 2002) y aleatorio (Scapoli *et al.*, 2007), distancia genética (Giraldo, 2001) y los estimadores A, B (Pineda *et al.*, 1999) y C (Bedoya *et al.*, 2006). Además del indicador alfa (α) de Fisher (Barrai, *et al.*, 2007) y del indicador v de Karlin-McGregor (Zei *et al.*, 1983).

Los estudios de isonimia en Colombia no son muy abundantes, sin embargo existen varias publicaciones usando el método con diferentes fines. Básicamente, los análisis de isonimia en el país, han estado centrados sobre poblaciones pequeñas, que muestran señales culturales, étnicas o demográficas de constituir aislados genéticos. Tan solo dos trabajos (Giraldo, 2001; Gómez, 2004) se han realizado en Colombia con una perspectiva nacional y no departamental o municipal. Cabe señalar también, que es Antioquia la zona de mayor aplicación de análisis de isonimia (Pineda *et al.*, 1999; Soto *et al.*, 2001; Bedoya *et al.*, 2006; Gómez *et al.*, 2008; Arias, Rojas & Bedoya, 2010), ya que como lo mencionan algunos autores (Bedoya *et al.*, 2006a; Bedoya *et al.*, 2006b) el poblamiento de este departamento, se originó a partir de dos núcleos fundadores y su población se ha considerado característicamente cerrada, debido a profundas diferencias culturales y a una historia marcada por fenómenos de aislamiento después de la época

colonial. Otros estudios realizados corresponden a un análisis comparativo entre los haplotipos del cromosoma Y y los apellidos en los departamentos del Valle del Cauca, Cauca y Nariño (Gómez et al, 2008).

A pesar de los atributos con los que cuenta el análisis de isonimia, el método también impone ciertas restricciones propias de la técnica y otras que están sujetas al diseño experimental de los investigadores. Los supuestos con los que se aplica el modelo han sido muy controversiales (Rogers, 1991), pues los apellidos en realidad no presentan un origen único, e incluso muchos de ellos fueron impuestos a las personas debido a su condición de esclavitud (Navarrete, 2005). Adicionalmente, la proporcionalidad entre hombres y mujeres no siempre es posible de alcanzar debido a que los procesos de migración, pueden afectar de una manera diferencial a ambos sexos, casos como las guerras y el movimiento en busca de oportunidades laborales, son algunos ejemplos. Aunque, las objeciones del método abundan como consecuencia de los supuestos, es necesario también recordar que todos los modelos planteados en ciencia presentan limitaciones. Por tanto, la manera apropiada de enfrentarlas es a través de un diseño experimental adecuado, iniciando con la determinación de las preguntas de investigación, susceptibles a responder a partir del método de isonimia y con la elección de las poblaciones en las que es posible aplicarlo. Igualmente, la selección de la base de datos, constituye un paso crítico en el establecimiento de estudios de isonimia; y en este sentido, la base de datos más inclusiva de la población debería dar una idea más clara de la dinámica de apellidos; sin embargo, esto dependerá en gran medida de su historia y naturaleza. Una buena opción sería combinar diferentes fuentes de datos de apellidos teniendo la precaución de no sobreestimar la abundancia de los mismos.

3.4. Una base de datos nacional: Para uso en genética de poblaciones, forense y para monitoreo en epidemiología.

En el mundo, a partir de 1950, las tecnologías de la información y la comunicación iniciaron un desarrollo notable. Esto ha repercutido en la transformación de la comunicación y en las relaciones

sociales, económicas y científicas; como consecuencia del incremento en las capacidades para generar, sistematizar, compartir, transmitir, analizar y difundir la información (Michan, 2010). En investigación científica ha habido una producción acelerada de una gran variedad de programas, aplicaciones, herramientas, recursos y servicios electrónicos; además de adoptar el formato electrónico y la Web como medio de comunicación para compartir información. Estos cambios, sin duda han repercutido en la transformación de la práctica científica a varios niveles, debido a que se ha reducido la energía, el costo y el tiempo requeridos para el análisis de la información. Además, ha generado nuevas oportunidades de investigación al incrementar el espectro de opciones técnicas y análisis factibles (Michan, 2010).

Después de tres décadas de trabajo en biología molecular, hay un gran volumen de información biológica generada por toda la comunidad científica. Este crecimiento explosivo de la información ha exigido en primera instancia la creación de bases de datos para su almacenamiento y organización, además del desarrollo de herramientas computacionales para el análisis de la información, que permitan estudiar nuevas hipótesis biológicas de manera lógica y eficiente. Actualmente en internet están disponibles un número importante de bases de datos para distintos marcadores moleculares y para distintos usos, se destacan por ejemplo ALFRED, NCBI, MITOMAP y mtDB.

En Colombia se ha acumulado información de VNTRs, STRs, SNPs y secuencias desde los años 90s, producto de pruebas de paternidad e investigaciones de los principales laboratorios y universidades. Desde el año 2000 hasta la actualidad, se han publicado aproximadamente 40 reportes de frecuencias de diferentes regiones en el país (Gómez et al., 2003; Paredes et al., 2003; Rey et al., 2003; Usaquén, 2006). Aunque dicha información, es valiosa por su tamaño de muestra y representatividad, infortunadamente estos datos se encuentran dispersos; ya que no se cuenta con un sistema que los almacene en una sola fuente de información. Además, para muchos de estos datos se carece de un análisis riguroso genético-poblacional, en buena parte por falta de fuentes de informaciones completas, consolidadas y con adecuados criterios de validación.

Considerando, que han pasado los primeros quince años de investigación en esta área, y partiendo del

poco conocimiento de nuestras poblaciones; el trabajo hasta el momento realizado, puede considerarse adecuado. Con el transcurso de una época en la que se presentaron los trabajos de genética de poblaciones básicas y el desarrollo de una primera revisión de frecuencias alélicas, se obtuvieron valores con los cuales hoy realizamos los cálculos de paternidad y criminalística. Sin embargo, actualmente se calcula que los laboratorios forenses cuentan con aproximadamente 300.000 tipificaciones entre casos de paternidad, así como aproximadamente 80.000 tipificaciones realizadas por los laboratorios privados. Hasta ahora se han hecho caracterizaciones aisladas de diferentes regiones del país, en cuyo análisis es importante la comparación con otras poblaciones relacionadas; pero no se cuenta con los datos actualizados para hacer inferencias de las relaciones con otras poblaciones y con la población general. Por otro lado, en el área forense, la ausencia de reportes de frecuencias alélicas en algunas regiones, puede llevar a un cálculo inexacto de probabilidades de exclusión en la identificación de presuntos padres, agresores, desaparecidos, fallecidos, entre otras.

Es así como la cantidad de información disponible y la solución de estas necesidades requiere fundamentalmente de una nueva generación de herramientas informáticas y plataformas de bases de datos. La propuesta es el desarrollo de una plataforma web de información genética humana nacional para marcadores moleculares, que contemple aspectos en genética forense y epidemiología. Es así, como la construcción de esta aplicación proporcionará una herramienta computacional con amplia capacidad de análisis y organización, que facilitará la recolección de la información. En consecuencia, sería posible construir una muestra referencial actualizada de Colombia para los distintos marcadores moleculares, además de tener acceso a los datos almacenados; con el fin de construir nuevas hipótesis sobre el estado genético de la población colombiana, a nivel regional y local; generando así, estimaciones acertadas de los análisis genético poblacionales, como por ejemplo una estimación robusta de la estructura genético-poblacional del país; por otro lado, puede usarse también, para el almacenamiento de otros tipos de datos, como SNPs y secuencias, permitiendo al mismo tiempo, que la comunidad académica tenga datos disponibles y actualizados. Además, sería importante poder detectar poblaciones atípicas para los estadísticos representativos de los distintos marcadores, con respecto a la muestra referencial, su detección es clave en la búsqueda de factores de riesgo específicos en estudios epidemiológicos.

Es posible así, estudiar fenómenos como selección, subestructura genética y mutación en los fenotipos constitucionales. Además, se pueden evaluar los factores de riesgo presentes en el medio ambiente, que pueden interactuar con la constitución genética de una población (Wyszynski, 1998).

En la construcción de la base de datos, es necesario asegurar la validez y calidad de los datos, debido a su importancia y sensibilidad. Esto es posible implementando mecanismos que evalúen constantemente la información reduciendo el sesgo. La manera de hacerlo, es con una correcta selección de la información, codificación, crítica e imputación, verificación, validación y consistencia de la información; también, asegurando la privacidad de la base, permitiendo un acceso específico y limitado, dependiendo los tipos de usuario. Por otro lado, es importante que se tengan en cuenta los aspectos éticos para las bases de datos, consagrados en la Declaración Internacional sobre los Datos Genéticos Humanos de la UNESCO Paris, 16 de Octubre de 2003.

Nuestro laboratorio ha desarrollado software para manejo de perfiles genéticos, logrando integrar los archivos de resultados de secuenciadores automáticos de la casa Hitachi: Modelo FMBIO y de Applied Biosystems tales como el modelo ABI 310 y toda la serie 31: 3100, 3130 y 3130 xl. Además de haber diseñado, en 1996 la primera base de datos con la que se almacenó información de 104.000 personas provenientes de casos de paternidad, también se diseñó en el año 1998, la aplicación APC para analizar casos de paternidad a partir de grupos sanguíneos y los primeros microsatélites que se usaron en el país. Con base a la aplicación en el año 2003 se analizaron 28000 casos de filiación y posteriormente se recolectaron aproximadamente 80000 tipificaciones que reposan actualmente en el Instituto Colombiano de Bienestar Familiar. Conjuntamente, la integración de nuestro conocimiento; en trabajo de campo y en las diferentes variables genéticas, demográficas, geográficas y modelos matemáticos utilizados en la genética de poblaciones, nos permitirán el desarrollo de una gran bases de datos con una estructura sólida integrada con todos estos aspectos, importantes en el desarrollo de la genética de poblaciones.

3.5. Unificación con el sistema de salud y la comunidad.

La mayoría de los estudios de “asociación genética”, están siendo utilizados para descubrir el componente genético que subyace a las enfermedades comunes de alta prevalencia, como por ejemplo, la diabetes mellitus, la enfermedad coronaria y el cáncer. Estos tipos de estudios, generalmente se trata de cohortes prospectivas o de tipo casos-contrroles, en los cuales se establece el peso relativo del componente genético con respecto a otros factores como el ambiental y el riesgo a desarrollar la enfermedad. Habitualmente, se utilizan como marcadores genéticos a los polimorfismos simples puntuales (SNPs). Estas variaciones pueden ser en sí mismas funcionales y estar relacionadas a la fisiopatología de la enfermedad, pero en la mayoría de los casos, son utilizadas para el mapeo y ubicación de los verdaderos sitios relevantes (Sevilla, 2007). Según estas interpretaciones, los análisis de estos polimorfismos están solamente enfocados a la naturaleza propia del gen con la enfermedad, y no se tienen en cuenta variables externas que afectan el comportamiento del gen en la población en general; ya que se enfocan solamente en muestras de pacientes, sin un mayor análisis demográfico, genealógico, ancestral y poblacional de cada individuo.

Antes de realizar este tipo de estudios tenemos que comprender las fuerzas de cambio evolutivo propias de cada población, como son: el efecto fundador, la deriva, migración, selección, cuello de botella, mutación. Si analizamos la población desde los microsátélites, podemos establecer el tipo de fuerza evolutiva presente en esa población, y de esta manera establecer mapas, planos, rutas y modelos genético poblacionales, para interpretar a mayor escala los datos obtenidos de un estudio de asociación simple a una patología específica.

Es por ello, que antes de realizar un estudio de asociación con una determinada enfermedad, considerada como problema prioritario de salud, es necesario establecer una base poblacional; incluyendo el componente genético, ambiental, cultural y demográfico; y evaluando estas variables en las poblaciones como unidades y no por regiones o etnias, ya que algunas características de la población, estarán condicionadas por la renovación en el cambio de una generación a la siguiente bajo diferentes entornos ambientales, sociales y culturales. De esta manera se puede explicar con certeza el comportamiento de la patología, en una población en particular.

Los estudios en genética de poblaciones se deben integrar totalmente con los estudios de asociación

genética involucrados en la explicación de las enfermedades, considerados problemas prioritarios en salud; aspecto que no ha sido considerado en este tipo de estudios.

Conclusión.

Nuestro mayor interés está centrado hacia un mejor uso de la información genética; seguir desconectados en la práctica diaria de los usuarios de los diferentes servicios, y que nuestros trabajos sigan siendo considerados como valores de referencia simplemente utilizados para realizar cálculos a diferentes niveles, representa un estancamiento de la disciplina, sin que se tenga un mayor impacto en otras áreas de las ciencias biomédicas y mucho menos en la población. Hemos estado abriendo un camino de comunicación entre las comunidades campesinas, indígenas y minorías étnicas, que creemos en un futuro dará a nuestros usuarios, un mejor nivel de comprensión de la genética humana en el país. A diferentes niveles educativos, se considera que uno de nuestros mayores problemas es la falta de identidad que nos impide defender y cuidar nuestro patrimonio, creemos también, que la construcción de sistemas de información, mejores prácticas en el abordaje de nuestros usuarios de filiación e identificación, una mejor interpretación de la variabilidad de nuestras poblaciones y la generación de mejores conjuntos de datos, son un sustrato para hacer de la genética de poblaciones en Colombia, un terreno fértil para el desarrollo de modelos matemáticos y estadísticos, que permitan una mejor utilización de la tecnología en biología molecular que poseemos, y para un acercamiento de los profesionales en ciencias biomédicas a los profesionales de las ciencias sociales.

Bibliografía.

Begoña M, Sánchez D, González A. 2005. El Estudio de Polimorfismos de ADN a Partir de Restos Óseos y Dientes y sus Aplicaciones en la Identificación de Desaparecidos. *Revista Aragonesa de Medicina Legal* 7: 163:184.

Bedoya G, Montoya P, García J, Soto I, Bourgeois S, Carvajal L, Labuda D, Alvarez V, Ospina J, Hedrick PW, et al. 2006. Admixture Dynamics in Hispanics: a Shift in the Nuclear Genetic Ancestry Of A South American Population Isolate. *Proceedings of the National Academy of Sciences of the United States of America* 103:7234-9.

Butler J, Shen Y, McCord B. 2006. The development of reduced size STR amplicons as tool for analysis of degraded DNA. *J Forensic Sci.* 48: 1054-64.

Callegari Jacques SM, and Salzano FM. 1979. Demography and Genetics of the Krahó and Gorotire Indians of Brazil. *Journal of Human Evolution* 8:513-522.

Carracedo A, Salas A, Lareu M. Problemas y retos de futuro de la genética forense en el siglo XXI. *Cuad Med Forense.* 2010;16:31-35.

Carvajal-Carmona LG, Ophoff R, Service S, Hartiala J, Molina J, Leon P, Ospina J, Bedoya G, Freimer N, and Ruiz-Linares A. 2003. Genetic Demography of Antioquia (Colombia) and the Central Valley of Costa Rica. *Human genetics* 112:534-41.

Carvajal-Carmona LG, Soto ID, Pineda N, Ortíz-Barrientos D, Duque C, Ospina-Duque J, McCarthy M, Montoya P, Alvarez VM, Bedoya G, et al. 2000. Strong Amerind/White sex bias and a possible Sephardic contribution among the founders of a population in northwest Colombia. *American journal of human genetics* 67:1287-95.

Crow JF, and Mange AP. 1965. Measurement of Inbreeding from the Frequency of Marriages Between Persons of the Same Surname. *Revista Eugenics Quarterly* 12:199-203.

De G, and Rica C. 1981. Demografía y Genética de la Población Amerindia Guayrní de Limoncito, Costa Rica. *Revista de Biología Tropical* 29:23-31.

Gómez MV, Reyes ME, Cárdenas H, García O. 2003 Genetic variation for 12 STR loci in a Colombian population (Department of Valle del Cauca). *J Forensic Sci.* 48(3): 687-688.

Jeffreys A, Wilson V, Thein S. 1985 Individual-specific "Fingerprints" of Human DNA. *Nature.* 314:67-73.

Jujuy UD. 2004. Cuadernos de la Facultad de Humanidades y Ciencias Sociales. Cuadernos de la Facultad de Humanidades y Ciencias Sociales 22:171-194.

Keyeux G, Rodas C, Gelvez N, and Carter D. 2002. Possible Migration Routes into South America Deduced from Mitochondrial DNA Studies in Colombian Amerindian Populations. *Human Biology* 74:211-233.

López N. Estructura Genética de la Población colombiana segunda muestra. . Tesis Maestría Universidad Nacional de Colombia, Facultad de medicina. 1993

Mesa NR, Mondragón MC, Soto ID, Parra MV, Duque C, Ortiz-Barrientos D, García LF, Velez ID, Bravo ML, Múnera JG, et al. 2000. Autosomal, mtDNA, and Y-chromosome Diversity in Amerinds: pre- and post-Columbian Patterns of Gene flow in South America. *American journal of Human Genetics* 67:1277-86.

Michan L. 12 de octubre 2010. Sesión de genética y tecnología: Del alfabeto del gen a los pixeles en pantalla. Cátedra de Sede José Celestino Mutis. Todo lo que Usted quiere saber de genética y nunca se atrevió a preguntar. Universidad Nacional de Colombia. [Internet]: Available from <http://laylamichanunam.blogspot.com/>

Pérez AA, Salzano FM. 1978. Evolutionary Implications of the Ethnography and Demography of Ayoreo Indians. *Journal of Human Evolution* 7:253-268..

Rodas C, Gelvez N, and Keyeux G. 2003. Mitochondrial DNA Studies Show Asymmetrical Amerindian Admixture in Afro-Colombian and Mestizo Populations. *Human Biology* 75:13-30.

Rodríguez, J. 1994 Introducción a la Antropología Forense: Análisis e interpretación de restos óseos humanos. Universidad Nacional de Colombia. Anaconda Editores. Bogotá.

Rojas W, Parra MV, Campo O, Caro MA, Lopera JG, Arias W, Duque C, Naranjo A, García J, Vergara C, et al. 2010. Genetic Makeup and Structure of Colombian Populations by Means of Uniparental and Biparental DNA Markers. *American journal of physical anthropology* 143:13-20.

Salas A, Acosta A, Alvarez-Iglesias V, Cerezo M, Phillips C, Lareu MV, And Carracedo A. 2008. The mtDNA Ancestry of Admixed Colombian Populations. *American Journal of Human Biology*: The Official Journal of the Human Biology Council 20:584-91.

Sánchez E. 2001. Biodemografía: Una Apuesta para el Estudio Biológico de las Poblaciones. *Revista de Demografía Histórica* XIX: 71-86.

Sandoval C. Estructura Genética de la población Colombiana. . Tesis Maestría Universidad Nacional de Colombia, Facultad de medicina. 1993

Sevilla, S. 2007. Metodología de los estudios de asociación genética. *Insuficiencia Cardíaca*. 2; 3: 111-114.

Solla L. 2004 Aplicaciones forenses del ADN Comisaría General de Policía Científica Laboratorio de ADN. [Internet] Available from: www.cej.justicia.es/pdf/publicaciones/fiscales/FISCAL333.pdf.

Pereira R, Phillip C, Alves C, Amorim A, Carracedo A, Gusmão L 2009 Insertion/Deletion Polymorphisms: a Multiplex Assay and Forensic Applications. *Forensic Sci Int Genetics Suppl* 2:513-515.

QIAGEN. 2010. Investigator DIPplex Handbook Sample and Assay Technologies.

Wyszynski D. 1998 La epidemiología genética: disciplina científica en expansión. *Pan Am J Public Health* 3(1), 26-34

Weber JL, David D, Heil J, Fan Y, Zhao C, Marth G 2002. Human Diallelic Insertion/Deletion Polymorphisms. *Am J Hum Genet* 71:854-862.

XUE Y, Tyler-Smith C. The hare and the tortoise: One small step for four SNPs, one giant leap for SNP-kind. *Forensic Sci Int Genet.* 2010;4:59-61

Yunis E, Yunis J. El ADN en la identificación humana. Editorial Temis S.A. 2002. Bogotá.

Anexo No. 1: Guía de Laboratorio para Análisis Genético.

Validación y consistencia de información en estudios de diversidad genética humana a partir de marcadores microsatélites.

Tesis de Doctorado en Ciencias – Biología.

Estudiante

William Usaquen Martínez M.Sc.
Departamento de Biología.

Director.

Luis Fernando García Pinzón Ph.D.
Departamento de Biología.

Director Asociado

Jimmy Corzo Salamanca Dr.rer.nat.
Departamento de Estadística.

Universidad Nacional de Colombia
2011

Dedicado ...

A mi maestra Indiana Pura Bustos Bustos, creo que con este documento terminó la tarea querida maestra... o será solo el inicio?. Tu me has enseñado a no estar satisfecho, a buscar algo mas, a no comer entero, he tratado de cultivar la agudeza mental y el buen corazón, lo que ha sido tu marca personal, y que espero sea la mía y la de mis estudiantes, he tratado de hacer la escuela que soñamos... en eso ando, transmitiendo tus enseñanzas de generación en generación.

Con gran cariño por ti Indiana.

A mis estudiantes, tendrán que disculparme por la severidad, por el temperamento, pero bien saben que les he dado lo mejor y lo peor que tengo, deben escoger. Les he presionado y fortificado en lo que se hacer y les he dejado incompletos en mi propia incompetencia, deben superarse. Por ejemplo, los se buenos analistas de datos, pero con necesidad de mas horas de laboratorio, halla paz, ya encontraran sus propios caminos. Deben seguir su educación para marcar a una generación de colombianos ilustres que reoriente el cause de nuestra historia.

A mi esposa, ¿Qué seria de mis días sin tu sonrisa?. Gracias por venir.

A mi gente, lo mas bello de este mundo, siempre me hacen sentir un orgullo profundo...

Bogotá, diciembre de 2011.

Carta de Agradecimientos.

La lectura de esta parte no es necesaria, pero se recomienda por el contexto.

Mis mas profundo agradecimientos a todas las personas que menciono en la siguiente carta. Este trabajo no es mío es de ellos, es una versión ampliada y puesta en papel de las discusiones sostenidas.

En el año 1999, mientras iniciaba tramites para realizar estudios fuera del país, decidí cursar la especialización en estadística en la Universidad Nacional de Colombia. No alcance a percibir lo determinante que seria esta decisión para mi. Algo que inicialmente seria un programa tendiente a mejorar mi nivel en un área normalmente descuidada en los pregrados colombianos, termino rediseñando mi futuro. En los estudios biológicos, con frecuencia surgen múltiples preguntas estadísticas, especialmente sobre el tamaño de muestra, en ese programa tuve algunas primeras discusiones con el profesor Leonardo Bautista, Luis Guillermo Díaz, Jorge Ortiz, charlas que tomaron forma cuando conocí al profesor Jimmy Corzo, y quien me fue mostrando un panorama de la estadística amplio, en el que además percibí como las respuestas sobre el tema serian mucho mas complejas.

Estas discusiones se articulaba con el momento que vivía el país, se estaban haciendo las primeras estandarizaciones para el uso de STR's en el país, específicamente en el Instituto Nacional de Medicina Legal (INML). Como sucede con cualquier novedad, el uso del analizador genético ABI 310, representaba un gran avance tecnológico. Éramos muchos los que esperábamos paciente mente la oportunidad de ponerle las manos encima al nuevo aparato, sin embargo en ese momento gracias al biólogo Fernando Parra Velandia, encontré

un juguete mucho mas atractivo: El lenguaje de programación Visual Basic Applications (VBA), un nuevo cambio en la vida, bien sabe un programador que un nuevo lenguaje es adictivo, con lo cual me aleje por completo del mundo del laboratorio.

Ese momento histórico para el país, también marco para mi una década de actividades: Trabajo en bases de datos con Adriana Castillo y Clara Vargas en la Universidad Industrial de Santander, Adriana Ibarra y Aníbal Gaviria en la Universidad de Antioquia, Julieta Henao y Liliana Porras en la Universidad Tecnológica de Pereira, María Victoria Gómez en la Universidad del Valle, Amparo Acosta en la Universidad del Cauca, Carlos Restrepo en la Universidad del Rosario, Andrés Gutiérrez y Alejandro Giraldo en la Fundación Gillow, Dora Fonseca y Andrea Gómez en Genética molecular de Colombia en Bogotá y a Juan José Builes del Laboratorio Nacional de Referencia. En todos los lugares, analizando tipificaciones genéticas año tras año me dio la experiencia para detectar variantes raras, en mirar lo que pueden ser posibles rutas de migración y poblamiento, en la fantasía de tener un sistema de bases de datos en línea para todo el país y la región andina, en lo bueno que seria hacer estudios de campo.

Luego apareció el Instituto de Genética de la Universidad Nacional. Ya han pasado seis años de permanencia. Muchas Altas y bajas. Aquí he conformado el grupo de Genética de poblaciones e Identificación Humana, donde se ha cocinado lentamente este trabajo.

Miguel Novoa fue la primera persona con quien inicie esta meta, fue mi primer compañero en la exploración de las aplicaciones en genética y el sin numero de supuestos, un compañero incansable, riguroso, trabajar con el significan horas, muchas horas de concentración. Posteriormente, se nos unió Leonardo Eljach, un compañero leal, agudo, buen pensador, he perdido la cuenta del numero de discusiones que tuvimos dando forma al proyecto Guajira. El tiempo que pasamos de viaje, conociendo las rancherías Wayuu, es en mis recuerdos una estupenda experiencia de vida.

Con algunas ideas ya maduras, pero solo eso, ideas, llegaron a nuestro grupo, Luz Ángela Alonso, Madelyn Yiseth Rojas y Vanessa Sarmiento. Empecé el entrenamiento de las chicas superpoderosas desde lo básico, ya han construido sus propios elementos, su propia personalidad científica, pero además dieron orden, forma, ritmo y dinámica al día a día de trabajo. Personas alegres, atentas al detalle, cuidadosas y perseverantes, hicieron realidad ser un grupo registrado. De ese empuje fue cierto ganar convocatorias, ejecutar presupuestos, tener computadores, organizar cursos, participar en eventos, hacer campo para estudios Genético poblacionales. Al fin todo fue cierto.

Ángel Carracedo, alguna vez me dijo que el procuraba vincular a su grupo personas de muy buena energía, he seguido a puño y letra su consejo, fue así como al grupo inicial, de estudiosos en aplicaciones, estadística y poblaciones, en el ultimo tiempo se ha unido Jenny Blanco, mujer única, disciplinada, alegre, espontánea, trabajadora incansable, Verónica Rocha, espiritual, meticulosa, leal, y Sandra Chala, sorprendente, previsiva, una mujer que siempre esta un paso adelante, ellas, dirigidas por Andrea (de quien hablare mas adelante) son nuestro brazo fuerte en las técnicas de laboratorio. Dobereiner Chala, nuestro joven hombre candidato a sabio, nuestra perspectiva en antropología y arqueología y Deicy Amparo Osorio, quien hace justicia a su nombre, ella ese quien nos ampara. Lo que fue un sueño es ahora una empresa científica de la cual ella es la administradora.

Quiero agradecer a mi Hermano, Luis Alejandro Camero Ramos, hermano por afinidad, como ya habrán descubierto los analistas de isonimia. Mis tiempos de madurez en la perspectiva filosófica de estas ideas se las debo a este Biólogo Filósofo, que vive de forma ascética, pero de quien la comunidad científica colombiana debería saber mas.

Finalmente, Lilian Andrea Casas Vargas, quien llego a mitad del proceso, buscando trabajar con un grupo impetuoso, de una manera tranquila, nos ha dado una lección importante sobre disciplina, perseverancia, cariño y crecimiento personal, ejemplo del buen trabajo, buena amiga, buena compañera, ella es quien le ha dado una articulación y estructuración única al equipo. Tal como lo soñé desde el principio, la parte de identificación no debería estar separada del trabajo en poblaciones, esta charnela es su aporte, Andrea es la unión y el equilibrio.

Este es el grupo con el cual llegamos al final de este viaje, Con este grupo de personas, que en promedio de edad tienen 26 años tienen múltiples bendiciones. En este grupo hay un derroche de energía, no apto para cardiacos, una capacidad de lectura, de abordaje, un ímpetu sin igual, ya al frente de este equipo, de ese tren, es muy difícil bajarse, como en las montañas rusas, solo tengo opción de aferrarme fuerte cerrar los ojos y gritar, porque el viaje empezó.

Si de algo he disfrutado en este proceso es de libertad, esa que solo brindan los verdaderos maestros, Luis Fernando García Pinzón y Jimmy Corzo Salamanca, me han dirigido con aprecio, dedicación, rigor, esmero y paciencia. Han sobrellevado mis limitaciones de una manera sabia. Las muchas horas de trabajo, en sus oficinas o al calor de un café, permitiéndome equivocarme y corrigiéndome, es un gesto para el que solo tengo palabras de aprecio y gratitud. Este trabajo crea entre nosotros un vínculo que agradeceré por el resto de la vida.

Aunque el contacto fue menor, no menos importante, A la profesora Leonor Gusmão, quien en momentos difíciles reviso las principales ideas y me dio un impulso importante para seguir adelante, Al profesor Fabricio Santos quien me dio enfoques importantes en el análisis poblacional, Al profesor Alberto Gómez, Hombre admirable de quien indirectamente he disfrutado sus clases, al profesor Guillermo Barreto, un estudioso de las poblaciones colombianas quien me dio ideas importantes sobre el trabajo con comunidades.

Finalmente, agradecemos a todos los colombianos que nos han dado un pedazo de si para lograr esta propuesta metodológica. Sinceramente reiteramos nuestro compromiso por hacer un país mejor a través de nuestros estudios para nuestra filial 1.

Durante todos estos años tuvimos el apoyo de más estudiantes de pregrado y postgrado que participaron en las tomas de muestra, apoyo al trabajo del laboratorio, diseño de material y en las tareas diarias del laboratorio. Entre ellos se destacan, Blanca Schroeder, Rita Baldrich, Wilson Sierra, Natalia Lamprea. Al personal de la división de Investigaciones en Bogotá, a la señora Lucy Gaitán, e Islena Bonilla, a las monitoras Julieth Castiblanco Quiroga y Jenny Barrera Velandia, además a una lista muy grande de personas en las diferente partes donde realizamos las diferentes partes del proyecto.

William Usaquén Martínez.
Estudiante Doctorado en Ciencias.
Universidad Nacional de Colombia.

Resumen

En los últimos veinte años en Colombia se ha dado un gran desarrollo en los estudios poblacionales, en buena parte por el auge de la genética forense de acuerdo con nuestra compleja realidad nacional, como producto derivado se ha dado la importación al país de tecnología suficiente para el desarrollo de nuevas técnicas en biología molecular. Sin embargo dos aspectos metodológicos como los cálculos de tamaño de muestra y la adecuada selección de las muestras no han tenido el mismo desarrollo en los estudios poblacionales, manteniéndose sin criterios específicos, en ocasiones generando resultados pocos claros y con estimaciones confusas sobre nuestras poblaciones.

Buscando realizar un aporte en estas dos direcciones, en el presente trabajo se presenta un nuevo algoritmo para la determinación de tamaños de muestra en estudios poblacionales. Este además arroja estimaciones de frecuencias alélicas asociadas a intervalos de confianza, lo que permite aproximaciones más confiables a las realizadas por conteo directo. Junto a este algoritmo se propone un método fundamentado en el análisis genealógico y demográfico para mejorar la selección muestral. La evaluación del algoritmo, se realizó con la base de datos de los usuarios de pruebas de filiación, mientras que para el análisis demográfico se realizaron estudios de campo en las poblaciones del archipiélago de San Andrés, Providencia y Santa Catalina, en el Caribe colombiano y en la península de la Guajira, población con una fuerte ancestría de la comunidad indígena Wayuu.

Los resultados obtenidos en los estudios muestrales indican tamaños cercanos a las 700 personas en poblaciones grandes y con marcada influencia de múltiples grupos ancestrales, dada las altas de migración que caracteriza la población de Bogotá. Por otra parte los análisis demográficos permitieron la construcción de múltiples grupos

poblacionales, de lo cual se generaron modelos de estructura *a priori* a los obtenidos con la información genética. El uso de variables demográficas permitió una gran definición de las poblaciones en el espacio y en el tiempo, independientes de las básicas clasificaciones político-administrativas, e incluso de las geográficas. Además, las variables demográficas, permitieron evaluar desde otra óptica los efectos de las diferentes fuerzas de cambio evolutivo en las poblaciones estudiadas.

Luego en el trabajo se intento ir más allá, construyendo dos capítulos de reflexión, el primero dirigido a los principiantes en la genética de poblaciones, en el que se explora el origen de la disciplina desde “El origen de las especies”. El segundo dirigido a los investigadores que incursionan en las poblaciones y a quienes presentamos cinco propuestas sugeridas para diseñar e incorporar en sus estudios. Fue un reto lograr los muestreos en el trabajo de campo, por tanto consideramos que junto al refinamiento demográfico o la precisión matemática, se trata de un acercamiento humano, el cual debe ser cuidadosamente realizado y por sobre todo integrador entre una cultura científica y nuestras culturas ancestrales.

En los capítulos de reflexión me he permitido plantear mi experiencia de trabajo y después de mucho pensarlo sin lugar a duda creo que esta tesis, so pena de ser extensa es el escenario para hacerlo.

Abstract

During the last twenty years, Colombia has greatly advanced in population studies, mainly for two reasons: (1) because necessary forensic genetics studies developed as a result of our complex national reality, (2) as a derived product as the country has enhanced in molecular biology tools. Nevertheless two methodological aspects including an adequate sample selection and sample size have not had the same development as the population studies. Instead, non-specific criteria exist, results may not be very clear or confusing estimates are generated about our populations. In order to contribute in further discussion on both issues, this thesis introduces a new algorithm to determine sample size in population studies. The output of the algorithm provides allelic frequencies associated to confidence intervals, which allows more reliable approaches to those performed by classical direct count. In addition, a new method, based on genealogical and demographic analyses is proposed in order to improve the selection of the sample.

To evaluate the algorithm, a database corresponding to paternity and filiation tests were used, where as for the demographic analysis two field works were carried out (1) with human populations of the archipelago of San Andres, Old Providence and Santa Catalina Islands in the Caribbean and (2) with populations of Wayuu, an Amerindian ethnic group with strong genetic ancestry of the Guajira Peninsula in northern Colombia. The results obtained in the study samples suggest samples sizes of about 700 people for large populations influenced by multiple ancestral groups and large migration rates. Demographic studies allowed proposing a multi-population structure, providing different models *a priori*. The use of demographic variables, instead of administrative and political boundaries, provides a good scenario in time and space for population studies. They also provide elements to understand the effect of evolutionary forces and the quality of sampling.

In this work, we tried to go beyond throughout two additional chapters of discussion and reflection on the issues of population studies. The first one is aimed to those interested in population genetics by exploring the origin of the discipline, even from Darwin's "*the origin of species*". The second one is aimed to researchers in forensic genetics throughout five proposals suggested to incorporate in their studies.

To get the samples for this work was a worthwhile challenge, but together with the demographic refining of the sampling and the proposed mathematical tool, it certainly provides an integrative approach of our scientific culture and our ancestral cultures. In the reflection chapters, ideas and thoughts come my own work experience for the last 10 years, so I thought this thesis would be a good scenario to write them out.

Tabla de Contenido.

Introducción General.....	9
Capitulo Introductorio: El Origen de las Especies y su relación con el inicio de la actual teoría de la herencia.....	13
Resumen	13
Abstract.....	13
El concepto de herencia.....	14
Antecedentes Históricos.....	14
El origen de las especies.....	16
El modelo pangenético.....	17
El extraño caso de Gregor Mendel.....	20
Siglo XX.....	21
Conclusión.....	23
Bibliografía	24
Capitulo No. 1: Introducción General sobre Elementos de Muestreo.....	25
Genética de Poblaciones.....	25
Desarrollo histórico de la genética de poblaciones.....	26
Estructura Investigativa de la Genética de Poblaciones.....	27
Modelos en genética de poblaciones.....	28
Algunos estudios relacionados en Colombia.....	29
Definición de Sistemas Microsatélites.....	30
Clasificación.....	31
Teoría de Muestreo.....	31
Conceptos iniciales de teoría de muestreo en genética.....	32
La Investigación por muestreo.....	33
Muestreo probabilístico.....	34
Muestreo no probabilístico.....	35

Plan de Muestreo Estadístico	36
Pasos para la construcción del plan de muestreo	36
Fuentes de sesgos y varianza	37
Análisis Multivariado.	38
Bibliografía	40
Capitulo No. 2: Una nueva aproximación al cálculo de tamaño de muestra en estudios genético poblacionales y forenses a partir de marcadores microsatélites.	43
1. Introducción.	44
2. Método propuesto.....	46
2.1. Fase 1: Cálculo secuencial de las frecuencias alélicas.....	46
2.2. Fase 2: Cálculo de coeficientes de variación.	47
2.2.1. Promedios ponderados de frecuencias.....	47
2.2.2. Desviación Estándar.....	49
2.2.3. Coeficientes de variación.	50
2.2.4. Alelos efectivos y alelos raros.....	50
2.3. Selección de un tamaño de muestra.	51
2.4. Análisis multi loci.....	51
2.5. Aplicación desarrollada.....	51
2.6. Datos empleados.	52
3. Resultados.	52
3.1. Cálculo inicial para un solo sistema STR.....	52
3.2. Tamaño de muestra establecido para el sistema completo.....	54
4. Discusión.....	57
5. Conclusión.	59
Bibliografía:	61
Capitulo No. 3: Análisis de la Estructura de la población de La Guajira: Una visión Genética Demográfica y Genealógica.....	63
Resumen	63

1. Introducción	64
2. Metodología	67
2.1 Análisis Molecular.....	70
2.2 Análisis Demográfico de la Muestra.....	70
2.3 Análisis Genético de las Agrupaciones Genealógicas	71
2.4 Análisis de las Agrupaciones Genéticas.....	72
3. Resultados	72
3.1 Análisis Demográfico de la Muestra.....	72
3.2 Análisis Genético de las Agrupaciones Genealógicas	77
3.3 Análisis Genético de las Agrupaciones Genéticas	85
4. Discusión	88
6. Conclusión	93
8. Bibliografía	95
8. Anexos	98
Capítulo No. 4: Análisis de apellidos y de cromosoma Y de las poblaciones raizales del archipiélago de San Andrés y Providencia (Colombia)	103
Introducción	104
Materiales y Métodos	108
2.1 Población de estudio.....	108
2.2 Otras poblaciones.....	108
2.3 Análisis molecular.....	108
2.4 Análisis Estadísticos.....	111
2.5 Comparación con otras poblaciones.....	111
2.6 Análisis de apellidos	112
3. Resultados	112
3.1 Diversidad Genética y haplogrupos	112
3.2 Análisis de redes.....	114
3.3 Estructura poblacional.....	116

3.4 Comparación con otras poblaciones.....	117
Apellidos.....	119
3.5 Diversidad y número de apellidos	119
4.0 Discusión	122
5.0 Conclusión.....	125
Capitulo No. 5: Cinco propuestas de unificación entre genética de poblaciones y genética forense para la próxima década.	133
Resumen.	133
1. El origen de los estudios en Colombia.....	134
1.1. Semblanza Histórica.	134
1.2. Un cuestionamiento sobre la actualidad.....	137
2. Fuentes de conocimiento de la Genética de poblaciones.	137
2.1. Biología Molecular.	138
2.2. Matemáticas y Estadística.	140
2.3. El Análisis demográfico.	142
3. Cinco propuestas de trabajo.....	146
3.1. Un análisis descriptivo de cada polimorfismo.....	146
3.2. Una información demográfica integrada a datos históricos, etnográficos y estadísticos.....	148
3.3. Genética clásica, isonimia.	151
3.4. Una base de datos nacional: Para uso en genética de poblaciones, forense y para monitoreo en epidemiología.....	154
3.5. Unificación con el sistema de salud y la comunidad.	158
Conclusión.....	159
Bibliografía.	159
Anexo No. 1: Guía de Laboratorio para Análisis Genético.	163

Introducción General.

El trabajo está compuesto por 6 capítulos a través de los cuales se desarrollan nuevos elementos estadísticos y demográficos para proponer mejoras en la investigación sobre genética de poblaciones. Esta disciplina en el país se ha desarrollado de una manera muy cercana a la genética forense, por tanto es frecuente la referencia de una hacia la otra. Los capítulos se pueden dividir en dos tipos: Técnicos, en los que se muestran los nuevos algoritmos y los estudios poblacionales, otros son los artículos de reflexión, tendientes a proponer cambios en las perspectivas de investigación, en los enfoques recomendados para la práctica científica y social. A continuación se presenta un resumen más detallado cada uno de ellos.

Capítulo introductorio: se presenta el concepto de herencia y su importancia en “El origen de las especies”. El gran vacío en la teoría Darwiniana para la explicación de la variabilidad en las poblaciones suscitó una gran carrera hacia la comprensión de la transmisión de caracteres a lo largo de las generaciones, lo que culminó en 1903 en la definición de la Genética como una disciplina nueva, cabe destacar que desde el origen mismo de esta disciplina, una de las principales subdivisiones creadas fue la genética de poblaciones, con la cual se logró la formulación matemática de la selección natural. Este capítulo, surgió como un ejercicio reflexivo del primer año de formación doctoral. Tratando de centrar la propuesta de investigación dentro de la genética de poblaciones actual, se realizó una exhaustiva mirada en el pasado. Puede pensarse que un capítulo de historia es una extravagancia, sin embargo hoy en día considero este elemento fundamental para una propuesta que quiera orientarse hacia el futuro. Es un elemento sumamente enriquecedor en la formación de cualquier investigador, permite evaluar formas de pensar y de proyectar. Algunos cuestionamientos actuales sobre la estructura, tamaño, estado y origen de una población, de hecho están vigentes desde el siglo XIX. En esta parte la conclusión más destacable

es la perspectiva de estar frente a un nuevo modelo de selección natural en las próximas décadas, quizá orientado desde nuevas disciplinas en las que nos gustaría proponer el análisis estadístico y demográfico.

Capítulo No. 1: Se presenta una revisión sobre algunos elementos que son importantes a modo de marco de referencia buscando un mejor avance a lo largo del documento, este parte fue considerada más como un punto de encuentro para contextualizar las principales ideas entre genética y estadística. Para el lector estadístico, ver la relación de las dos disciplinas puede abrir nuevas perspectivas en el modelamiento teórico. Para el Genetista, se trata de encontrar una revisión sobre elementos estadísticos a tener en cuenta en la construcción de los muestreos, sinceramente espero pueda ser una guía en la construcción de sus próximos trabajos de investigación.

Capítulo No. 2: Se presenta un método estadístico nuevo desarrollado para el cálculo de tamaños de muestra asociado a la determinación de intervalos de confianza que permitirán cambiar las estimaciones de modelos puntuales a modelos por intervalos. En este capítulo se logra relacionar el tamaño de un polimorfismo genético con el tamaño de la muestra. Es un capítulo técnico en el cual invertimos cinco años de arduo trabajo, implicó un fuerte esfuerzo de programación. Al final del trabajo, incluso aun merece algunos cambios, sin embargo como procedimiento está completo. Las perspectivas de uso de este algoritmo están en desarrollo. La aparición de las nuevas versiones de software tales como *Arlequín*, o *Structure*, incorporan procedimientos por intervalos de confianza, lo que significó un impulso estimulante para continuar este desarrollo. Creemos que el mundo forense sin lugar a dudas se verá beneficiado por esta metodología e incluso se logren proponer nuevas interpretaciones de una manera más probabilística que determinística en casos de filiación e identificación. **El anexo No. 1** es la guía completa de análisis del sistema CODIS construida con este algoritmo. Esta guía actualmente se utiliza en nuestro laboratorio como base de referencia en la resolución de casos de filiación. Próximamente será distribuida en cuatro laboratorios más del país para su uso y evaluación y luego implementada vía WEB.

Capítulo No. 3: Se presentan los resultados de investigación del proyecto Wayuu, es un proyecto pensado desde el principio para el desarrollo de métodos demográficos como es el uso de encuestas asociadas al análisis genealógico buscando un mejoramiento de los muestreos por conveniencia los cuales son los más utilizados en genética. Este proyecto además tuvo un amplio análisis preliminar de la población, tal que al llegar a campo contábamos con un buen conocimiento de los diferentes grupos humanos presentes en la región lo que redundó ampliamente en la optimización y calidad el muestreo. La delicada toma de muestras, el registro de un amplio número de variables y la eliminación de sesgos debidos al uso de muestras provenientes directamente de campo y no de casos de filiación e identificación, permitieron después utilizar una muy amplia gama de vías y métodos de análisis. Con frecuencia en nuestro grupo de investigación recibimos preguntas sobre como analizar datos, sin embargo la principal limitación que vemos al asesorar, es la poca información con la que se cuenta. Este es un análisis exhaustivo de una región y de un maravilloso grupo humano. Salimos de lo teórico y de la sugerencia para dar paso a la práctica.

Capítulo No. 4: Muestra los resultados del proyecto San Andrés, donde se utilizaron datos del cromosoma Y, nuevamente se utilizó una encuesta demográfica, sin embargo se realizó el abordaje de la información no desde marcadores autosómicos, sino desde la perspectiva de los haplotipos del cromosoma Y. En esta parte se pudo explorar los efectos de la selección de las muestras en procesos ancestrales. Aquí se utilizaron dos de los elementos más importantes de nuestras propuestas: por un lado los análisis de isonimia y por otro los modelos de inferencia haplotípica que están fundamentados en el análisis Bayesiano. Es la unificación de los métodos de muestreo, el análisis de haplotipos de cromosoma Y y el uso de herramientas de genética clásica, otro de los elementos que creemos esencial rescatar para los estudios poblacionales.

Capítulo No. 5: A modo de conclusión es un artículo de reflexión sobre la situación de la disciplina en nuestro país, surge como un modelo propositivo en contraste con los capítulos anteriores que estuvieron centrados en elementos teóricos. Basado en la experiencia, se que una propuesta teórica, de no encontrar un guía o propuesta de implementación, puede ser abandonada

rápida. En este capítulo se relaciona el análisis poblacional con los problemas prioritarios en salud. En la actualidad en el país la genética de poblaciones es considerada únicamente como una subdisciplina teórica, cuando realmente brinda los elementos prácticos para la comprensión de la diseminación y alcance de muchas patologías de origen genético e incluso algunas asociadas a procesos ambientales o del desarrollo. También se exploró el panorama de las bases de datos y de la demografía. Es nuestra pretensión a más largo plazo promover una ampliación de las fuentes de investigación en genética de poblaciones, mas parámetros de discusión, buscando salir de una época en la que estamos cayendo en una repetición constante de los trabajos de investigación y en ocasiones de la utilización de categorías y clasificaciones coloniales, que merecen ser repensadas según la América que somos y queremos ser.

Capítulo Introductorio: El Origen de las Especies y su relación con el inicio de la actual teoría de la herencia.

The Origin of Species and its Relation with the Birth of Present Genetic Theory

William Usaquén Martínez, M.Sc.

Instituto de Genética, Universidad Nacional de Colombia. Bogotá D.C., Colombia.
wusaquenm@unal.edu.co

Trabajo publicado en acta Biológica Colombiana: Acta biol. Colomb., Vol. 14 S, 2009 77 - 84

Resumen

En la primera parte se mostrarán algunos elementos históricos relacionados con la “*Philosophie Zoologique*” de Jean-Baptiste Lamarck, obra fundamental de la biología moderna merecedora de una mención especial en su aniversario número 200. Esta obra contiene algunos conceptos importantes para “*The Origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*” el trabajo fundamental de Charles Darwin. Posteriormente se revisará la idea general de la pangénesis y sus diferentes contradicciones en la explicación de la variabilidad necesaria para la acción de la selección natural que culminarían con el nacimiento de la genética. Es importante mostrar la posición particular de los experimentos mendelianos en relación con el paradigma evolutivo causado por el origen de las especies.

Palabras clave: pangénesis, herencia, lamarckismo, experimentos mendelianos, evolución.

Abstract

In the first part I present some historical elements related to the “*Philosophie Zoologique*” by Jean- Baptiste Lamarck, a fundamental work of modern biology. This work deserves a special mention in its 200th anniversary, since it contains some important concepts for “*The Origin of Species by Means of Natural Selection, or the Preservation of favored races in the struggle for life*” the principal work of Charles Darwin. Subsequently we review the pangensis and various contradictions in the explanation of the variability required for the action of natural selection that culminate with the birth of genetics. In this paper is important show the particular position of Mendelian experiments in relation with the evolutionary paradigm caused by the origin of species.

Key words: pangensis, heredity, Lamarckism, Mendelian experiments, evolution.

El concepto de herencia

La herencia genética formalmente definida es el proceso por el cual la descendencia de un organismo adquiere las características de su célula u organismo progenitor. Esta herencia genética intervendrá en la regulación de un organismo en cuatro aspectos esenciales: morfología, fisiología, lineamientos generales de desarrollo, y por último, algunos aspectos del comportamiento.

La preocupación por la herencia nace con la humanidad misma, asociada a los fenómenos de similitud y continuidad de la forma en una especie biológica, posteriormente, se incorporó la noción de variabilidad para conformar lo que hoy en día denominamos genética (Jacob, 1988).

Actualmente la genética se constituye en una de las principales disciplinas de las ciencias biológicas, por la amplia gama de explicaciones y predicciones en la expresión genotípica final de los seres vivos a diferentes niveles orgánicos, desde el molecular hasta el poblacional. Otro aspecto a resaltar de la genética es la definición de las fuerzas de cambio evolutivo que tratan de explicar la continuidad de las especies a lo largo del tiempo, manteniendo una alta reproducibilidad en la replicación del material genético, generación tras generación, pero a la vez con modelos que explican la variabilidad como elemento de respuesta a los procesos evolutivos.

Antecedentes Históricos

Jean-Baptiste Lamarck, en la “*Philosophie Zoologique*” plantea por primera vez y de manera formal el concepto de Biología como disciplina científica. Como antecedente a esta obra se encuentra “*Histoire naturelle, générale et particulière*”, escrita entre 1745 a 1788 por Georges Louis Leclerc conde de Buffon, una enciclopedia completa en 37 volúmenes, como producto de un esfuerzo individual. En esta obra Buffon da una primera organización a la embriología, desarrollando principios integradores para la definición de los vertebrados, haciendo énfasis en la anatomía comparada y la importancia de estudiar el medio ambiente de los organismos. Para 1735

Carl von Linné había publicado “*Systema Naturae*”, obra duramente criticada por el conde de Buffon quien sostenía que dicho sistema en lugar de aclarar la clasificación de los organismos servía para conducir a confusiones debido a la extrema complejidad de sus normas. A pesar de los reparos de Buffon, Lamarck utilizó el sistema binomial para realizar la primera clasificación de la flora y fauna francesa, mostrando la versatilidad del mismo, lo que le permite obtener una plaza en el museo nacional de historia natural de Francia.

En el pensamiento de Jean-Baptiste Lamarck confluyen tres elementos esenciales: las ideas políticas revolucionarias que le permiten una mayor libertad de pensamiento lejos del poder monárquico y clerical de la época, un vigoroso proceso de actualización en los diferentes hallazgos de los naturalistas contemporáneos, y finalmente un profundo conocimiento de las obras clásicas griegas, que le llevaron a adoptar diferentes elementos teóricos tales como el sistema atómico planteado por Demócrito y a relacionarlo con el origen de la vida por generación espontánea. En sus trabajos, Jean-Baptiste Lamarck, plantea el estudio de una química diferente basada en las propiedades emergentes de las sustancias producidas por los seres vivos, particularmente de los ácidos, que una vez sintetizados cambiaban su capacidad de corrosión determinando compuestos completamente nuevos. Según Ernest Haeckel, la Filosofía zoológica reúne el primer cuerpo de conocimientos organizados que enlaza el concepto de evolución con el concepto de herencia (Lamarck, 1986).

Los dos postulados referentes a la herencia, en el trabajo de Jean-Baptiste Lamarck fueron: 1. La herencia de caracteres adquiridos. 2. El principio del uso y desuso de órganos. En la primera de ellas, Lamarck da importancia al hábito como una característica heredable. Este principio fue fácilmente aceptado por Darwin en el origen de las especies como parte de su interpretación de la continuidad de las formas vivientes. Por otra parte, Lamarck plantea dentro de su modelo evolutivo una tendencia lineal, un modelo de causas finales hacia el desarrollo y mejoría de los organismos. La linealidad evolutiva, en el pensamiento de Lamarck, no hace referencia a un diseñador preexistente, no existe en la mente de Lamarck un dios que actuase como fundador de

las directrices para el desarrollo orgánico, pero si plantea un concepto teleológico en la evolución natural, asumiendo una posición de vitalista científico, similar al concepto aristotélico. En este punto se da un rompimiento conceptual entre Darwin y Lamarck. Para Darwin el proceso evolutivo no sigue una dirección final determinada. A pesar de estas diferencias, los modelos darwinianos de herencia y evolución están notablemente influenciados por el trabajo de Lamarck, a quien también debemos reconocer, como uno de los principales fundadores de la estructura epistemológica y filosófica de la Biología para llegar a constituirse en disciplina científica. En este punto es prudente recordar la siguiente cita:

“Nadie ignora que toda ciencia debe tener su filosofía, y que sólo por este camino puede hacer progresos reales. En vano consumirán los naturalistas todo su tiempo describiendo nuevas especies y en marcar todos los matices de sus variaciones para aumentar la lista inmensa de las especies inscritas, porque si la filosofía es olvidada sus progresos resultarán sin realidad y la obra entera quedará imperfecta”.

Philosophie Zoologique.

El origen de las especies

Para 1859 la obra ya había pasado por un largo proceso de maduración y muchas de las ideas planteadas por Darwin eran esperadas. Según las referencias históricas, cuando la obra es publicada, en pocos días logra agotarse la primera edición causando diferentes reacciones y una polaridad casi inmediata, desde muchos sectores de la sociedad diferentes a las ciencias naturales. Se dio inicio a múltiples interpretaciones convirtiéndola en obra multipropósito, no sólo científica sino incluso filosófica y religiosa, que rápidamente fue cayendo en la confusión debido a las diferencias en interpretación por parte de diferentes autores.

Se le atribuye al origen de las especies haber causado una división en relación con la Iglesia católica, sin embargo, una lectura cuidadosa de la obra deja ver a un Charles Darwin profundamente religioso, desinteresado en primera instancia por entrar en conflicto con la Iglesia. No en

vano en el último párrafo de su obra en la tercera edición atribuye el origen mismo de la vida y la grandeza de la naturaleza al creador.

Como una de las primeras grandes novedades introducidas en el Origen de las especies, está la proposición de población biológica como unidad de cambio evolutivo, a este nivel orgánico actuaría el recién propuesto principio de selección natural, por tanto, se hace indispensable proponer un mecanismo de herencia, por el cual la variabilidad pueda generarse y mantenerse a lo largo del tiempo geológico para dar origen a nuevas especies. En la edición de 1859 aún no se ha planteado un modelo de herencia, sin embargo para 1868 de una forma provisional, Darwin propone el modelo pangenético. Hoy en día, dicho modelo podría parecer superfluo, sin embargo, se examinarán algunos elementos para demostrar cómo Darwin no tomó a la ligera el problema de la herencia.

El modelo pangenético

La pangénesis en términos generales, describe la migración de partículas provenientes de las diferentes partes del cuerpo y que mantendrían las características estructurales y funcionales de cada órgano para mezclarse en la simiente y dar origen a un nuevo organismo. Es un modelo antiguo descrito en una forma preliminar por Demócrito quien habló de partículas mecánicas. Ya en la versión del siglo XVI, con una marcada influencia de los árabes, a quienes debemos la preservación de una buena parte de las obras de la antigüedad griega, la pangénesis se había reformulado según la teoría de los humores que habla del Atrabilis o bilis negra, la bilis, las flemas y la sangre como las sustancias básicas constitutivas del cuerpo, concordantes con los cuatro elementos clásicos del mundo griego. Del balance de estas cuatro sustancias dentro del organismo dependía la buena salud y el equilibrio orgánico del ser, también del equilibrio en las características de la personalidad. Por el contrario, los procesos patológicos estarían asociados con el déficit o exceso de alguna de estas sustancias. Por muy errada que parezca la pangénesis, le proporcionaba a Darwin un modelo de continuidad en el tiempo con una constante relación entre organismo, ambiente y desarrollo individual, lo que inicialmente es suficientemente coherente para ser aceptado dentro del mecanismo de selección natural como modelo de herencia.

La teoría pangenética también había sido importante para otros pensadores en biología, lo que le da confianza al mismo Darwin. Un ejemplo fue Carl von Linné, quien al proponer el sistema natural, inicialmente no encuentra la manera de clasificar al hombre (Boorstin, 1997). Nueve años después de la primera edición del sistema natural, establece diferentes razas del hombre basado en los tipos de carácter y personalidades humanas descritos por la teoría de los humores cuya herencia a lo largo del tiempo se mantenía por partículas (que posteriormente Darwin denominó pangenes).

En el siglo XVIII y XIX, con el desarrollo de la microscopía se introduce un mayor número de observaciones relacionadas con el funcionamiento del organismo, llevando a una mayor elaboración del modelo pangenético. Se pasa de cuatro elementos esenciales a la interpretación de cinco “tejidos”: sangre, huesos, vísceras, nervios y bilis; tejidos básicos generadores de pangenes, constituyentes de la simiente, que se asocian en una “mezcla” homogénea de las características provenientes por línea materna y paterna, para dar origen a un nuevo organismo. De esta época datan diagramas y atlas histológicos encargados de sustentar la formación de dicha simiente.

Finalmente, el modelo pangenético, conduce a tres problemas biológicos básicos para la articulación entre selección natural y herencia:

1. La selección natural, a partir del cual una especie, puede dar origen a otra, exige un modelo que explique cómo se mantiene la continuidad de las características hereditarias propias de un organismo a lo largo del tiempo, pero a la vez capaz de mantener la “plasticidad” para cambiar la naturaleza de la simiente en el tiempo y constituir una nueva especie.
2. Al aparecer el nivel orgánico denominado población, como la unidad de cambio evolutivo, requiere de un mecanismo generador de variabilidad constante sobre el cual pueda actuar la selección natural.

3. Los pangenes provenientes de cada uno de los tejidos corporales que realizarían una mezcla tendiente a la homogeneidad de la simiente, al cabo de pocas generaciones de apareamiento se llegaría a la uniformidad de las características hereditarias dentro de una población, lo cual limitaría el supuesto de variabilidad requerido para la acción del principio de selección natural. Fenómeno congruente con un esquema lamarckiano de evolución que Darwin adopta en ese momento para explicar la influencia directa con resultados definidos, del medio ambiente sobre los organismos.

Darwin adoptó la pangénesis dentro de su teoría, sin embargo, cabe recordar que para ese momento Matthias Schleiden, Friedrich Theodor Schwann y Rudolf Virchow, habían formulado las bases de la teoría celular con tres principios básicos y generales que permitían interpretar el mundo vegetal y el mundo animal desde una visión unificadora de los seres vivos. Darwin no aceptó completamente la teoría celular, en especial, el segundo principio referente a la preexistencia de cada célula (Ruiz y Ayala, 2002). Para 1864 (un año antes de la aparición de la última edición del origen de las especies), Louis Pasteur, realizó las demostraciones que ponen fin al principio de generación espontánea y que de inmediato apoyan la generalización de la teoría celular.

Al terminar de escribir el origen de las especies Darwin se encuentra enfrentado a una teoría celular cada vez más aceptada y a la abolición por completo del principio de generación espontánea, causando serias inconsistencias para explicar el origen inicial de los seres vivos en su obra. Darwin nos ofrece un mecanismo por el cual una especie se origina a partir de una especie preexistente, pero el inicio mismo de las formas vivas es fundamentalmente una creación divina.

En las notas originales se muestra como Darwin inició una serie de experimentos de acuerdo con los métodos propios de su época: cruzando híbridos de plantas. Se da a la tarea de buscar un mecanismo de herencia para explicar la variabilidad en las poblaciones, pero uno de los problemas fundamentales de sus experimentos, es que fueron desarrollados sobre la observación completa del organismo, enfrentándose de lleno a la herencia de caracteres complejos, que incluso hoy en día es

de difícil abordaje.

Esta serie de contradicciones dentro del pensamiento darwiniano de ninguna manera tienden a opacar la calidad conceptual de su obra, simplemente nos permiten interpretar a un autor dentro de su contexto histórico, puede verse como las contradicciones propias de su teoría fueron planteadas por él mismo. Estas observaciones nos permiten poner en su justa dimensión al origen de las especies y a Charles Darwin como un pensador de una alta capacidad crítica sobre sí mismo.

La contradicción entre un principio que nos explica el proceso de cambio en las especies, fundamentado en la variabilidad de la población, y por otra parte, en un mecanismo de herencia tendiente a disminuir dicha variabilidad, será el caldo de cultivo para generar un nuevo paradigma científico que será el objeto de estudio de las primeras generaciones de biólogos evolutivos a partir de 1868 y que derivará en lo que hoy conocemos como teoría genética.

El extraño caso de Gregor Mendel

Hoy en día se considera a Gregor Mendel como el padre de la genética debido a su trabajo de 1865, titulado “experimentos en híbridos de plantas”, en el cual se describen los principios básicos para la transmisión de caracteres hereditarios, principios ampliamente demostrados en muchas especies, sin embargo, el mismo Mendel no vislumbró la capacidad de generalización de sus trabajos. Su trabajo, una vez publicado y leído, cayó en el completo olvido hasta 1900 cuando Carl Correns, Erich Von Tschermak y Hugo de Vries, formulan los mismos principios planteados 35 años antes. La pregunta es: ¿por qué después de la publicación del origen de las especies en el cual era clara la necesidad de buscar dichos principios, el trabajo de Mendel no tuvo ninguna relevancia?

Mendel había optado por la carrera eclesiástica como vía para acceder a un mayor nivel de formación académica. Luego de su ordenación, fue enviado a la Universidad de Viena durante 1851 a 1853 donde, entre otros, fue discípulo del botánico Franz Andreas Nicolaus Unger y del físico Andreas Christian Doppler. Este último había sido un destacado científico quien introdujo a

Mendel en los rudimentos matemáticos de la experimentación. A su regreso a la abadía Agustina de Brunn, Mendel por su formación en ciencias naturales, fue encargado de resolver varios problemas de horticultura relacionados con la depuración de cepas para la producción de semillas. Luego de 10 años de rigurosa investigación, culminó su trabajo con la formulación de los principios de segregación y segregación independiente, sin embargo, sus experimentos se encontraban lejos del paradigma genético y evolutivo de ese momento. El problema de Mendel era un problema eminentemente práctico, en el que utilizó un detallado análisis matemático, llegando a la resolución y generalización de un problema biológico, incluso propuso un modelamiento teórico para el principio de segregación. Muy a pesar de la suficiencia de sus experimentos, estos carecen por completo de una perspectiva evolutiva, y por tanto fue imposible vincularlos con el origen de las especies.

Sea o no Mendel el padre de la genética, en cualquier caso sus experimentos han sido admirados y enseñados. Por un lado, no llega a comprender que a partir de sus trabajos en híbridos de plantas se deriva una disciplina nueva, diferente de la fisiología y que entraría a encargarse de la herencia y sus mecanismos como tema principal, por otro lado, sus procedimientos, definiciones de cruces, metodologías planteadas, una vez redescubiertas, influenciaron a varias generaciones de investigadores, incluido Thomas Hunt Morgan, posteriormente ganador del premio Nobel por sus descubrimientos que llevaron a la constitución de la teoría cromosómica de la herencia, en la que además, utilizó un modelo como *Drosophila melanogaster*, con el que verificó los principios mendelianos para el mundo animal.

Por el rigor de sus experimentos, por el alcance de los mismos y por su capacidad de plantear modelos matemáticos, Gregor Mendel debería ser también estudiado y reconocido como uno de los precursores del análisis matemático en biología.

Siglo XX

Después del origen de las especies toda una generación de biólogos empieza a adoptar el principio de selección natural dentro de la explicación de diferentes modelos biológicos, uno de los biólogos

más entusiastas de esta época es William Bateson, que se convierte en un darwinista acérrimo, para después encontrar una serie de dificultades dentro del principio de selección natural, llevando incluso al estudio de mecanismos alternativos (Sturtevant, 1965).

Hasta ese momento, la herencia era estudiada como una parte de la fisiología de los órganos reproductores. Los estudios de Bateson, junto con los descubrimientos de Mendel crearon el marco conceptual para estudiar la herencia como una disciplina aparte, con una serie de preguntas independientes. Una disciplina que poco a poco fue reafirmando el principio de selección natural, encargada de estudiar en cada ser vivo sus características hereditarias e interpretando la naturaleza de la población en función de unidades estructurales constantes a lo largo del tiempo, a las que ya en el siglo XX se llamó genes (Fox, 2002). Por tanto, la constitución de la disciplina llamada genética viene a ser una derivación del principio de selección natural. El origen de las especies generó la crisis conceptual y la necesidad de buscar los principios y mecanismos que dieran cuenta de la variabilidad en las poblaciones biológicas, dejando la puerta abierta para la búsqueda de la estructura del gen.

Una vez constituida la disciplina, se inicia un proceso de adaptación de los principios mendelianos a la selección natural terminando hacia 1930 con la formulación del neodarwinismo o la nueva síntesis que es la interpretación numérica y matemática del principio de selección natural (Fisher, 1958).

Más allá de los propios conceptos y teorías biológicas, desde el principio, las teorías genética y evolutiva, han tenido diversas interpretaciones llevando a la justificación de variadas posiciones políticas, científicas y religiosas. Por ejemplo, la sociobiología, que fundamentada en el neodarwinismo ha gozado de gran credibilidad científica, tratando de explicar el funcionamiento y comportamiento de los seres vivos desde un conjunto de reglas rígidas, en buena parte, sustentadas teóricamente por la construcción de modelos matemáticos, llegando a reducir lo viviente sólo a su base genética. En respuesta al reduccionismo genético y de los modelos químicos, pero igualmente nociva, se construyen interpretaciones religiosas, esotéricas y místicas de la evolución, derivando

en corrientes como el diseño inteligente que trata de dar un sustento científico al modelo creacionista clásico de la religión católica.

Indiscutiblemente nuestra comprensión del mundo se ha enriquecido notablemente con todo el desarrollo de la teoría biológica, pero muy por encima de todo nuestro desarrollo técnico y científico, seguimos enfrentados a las viejas preguntas planteadas desde los inicios de la humanidad, relacionadas con nuestro origen y con nuestras diferencias como especie.

Conclusión

En la segunda mitad del siglo XX, se ha dado un cambio radical con el advenimiento de la tecnología del ADN, cambiando la interpretación de la variabilidad poblacional y las relaciones evolutivas entre las especies. Igualmente, se ha dado un crecimiento exponencial en informática, la capacidad de almacenamiento de información, el diseño de nuevas estructuras en sistemas de información y una mayor velocidad de análisis. Los nuevos datos, y la nueva forma de manejarlos, han impulsado el desarrollo teórico en técnicas estadísticas, especialmente en el mundo del análisis multivariado y en el desarrollo de pruebas no paramétricas que dejan la puerta abierta a nuevos tipos de análisis de datos.

Una de las primeras implicaciones que tiene el desarrollo tecnológico y teórico sobre la biología, es un cambio sustancial en los procedimientos experimentales. Clásicamente los seres vivos fueron analizados desde disciplinas y procedimientos separados, por ejemplo, las técnicas de campo en ecología, los análisis de ADN en laboratorio, la taxonomía clásica. Esta individualidad está cambiando a un modelo experimental transversal para las disciplinas tradicionales del saber biológico. La unificación implica la reinterpretación de los fundamentos teóricos y determina nuevos alcances en las teorías evolutiva y genética. Quizás estamos entrando a una época en la cual veremos un nuevo principio de selección natural, más articulado con la biología del desarrollo, interpretado desde una genética más amplia, entendida no solo desde el gen, sino que considere la ecología y la fisiología. Quizás estamos frente a una nueva teoría biológica en cuyo centro no esté en el gen, sino el organismo en sí.

Bibliografía

BOORSTIN D. Los descubridores. 1.a edición. Barcelona: Editorial Grijalbo Mondadori;

DARWIN C. El origen de las especies. España: Editorial Los grandes pensadores; 1983.

FISHER R. The genetical theory of natural selection. 2a edición. New York; 1958.

FOX E. El siglo del gen cien años de pensamiento genético. 1a edición. Barcelona:
Ediciones Península; 2002.

JACOB F. La lógica de lo viviente. 1a edición. Barcelona: Salvat Editores; 1988.

LAMARCK J. Filosofía zoológica. Barcelona: Editorial Ata Fulla; 1986.

RUIZ R, AYALA F. De Darwin al DNA y el origen de la humanidad: la evolución y sus
Polémicas. 1a edición. México: Ediciones Científicas Universitarias; 2002.

STURTEVANT A. A History of genetics. 1a edition. London - Tokyo: A Harper international
edition; 1965.

Capítulo No. 1: Introducción General sobre Elementos de Muestreo.

William Usaquén Martínez, M.Sc.

Instituto de Genética, Universidad Nacional de Colombia. Bogotá D.C., Colombia.
wusaquenm@unal.edu.co

Genética de Poblaciones

La genética de poblaciones es una rama de la genética que se encarga de estudiar los cambios en las frecuencias alélicas de generación en generación y las fuerzas evolutivas responsables de dichos cambios en las especies. En la mayoría de los casos las poblaciones de estudio corresponden a poblaciones silvestres; sin embargo, los modelos genéticos se han extendido a poblaciones humanas (que no son consideradas poblaciones naturales), animales domésticos (como resultado de cruzamientos dirigidos), plantas (poblaciones agrícolas) y poblaciones de microorganismos (Hartl, 1997). Actualmente, los modelos poblacionales incluyen el hábitat de los organismos, considerando variables biológicas propias de regiones terrestres, acuáticas y aéreas. Tal como lo enuncia Gillespie (1998), el genetista de poblaciones intenta describir la estructura genética de una población o teoriza sobre las fuerzas evolutivas que actúan en las poblaciones; por lo tanto, la genética de poblaciones está fundamentada en la construcción de modelos. Si bien, existe un marco general de conocimientos; la biología propia de cada organismo, las condiciones cambiantes de los sistemas biológicos, las complejas relaciones de los organismos con su ambiente y los eventos del desarrollo, hacen que los modelos deban analizarse puntualmente. Por esta razón, la genética de poblaciones en la biología moderna es considerada una disciplina transversal y fundamental para el entendimiento en áreas como biología evolutiva, sistemática, ecología, historia natural y conservación genética. En esta última área, hoy en día es inevitable tomar decisiones, en cuanto al manejo y conservación de una especie sin la adecuada evaluación genético-poblacional. Fuera de las áreas biológicas, la genética de poblaciones interactúa con otras disciplinas entre las que se cuentan: medicina, derecho, biotecnología, sociología y antropología. (Hedrick, 2005).

Desarrollo histórico de la genética de poblaciones.

La genética de poblaciones parte de los principios mendelianos de segregación e independencia para hacer extrapolaciones a los efectos que dichos principios tendrían en una población. Pueden considerarse dos momentos decisivos para el inicio de ésta disciplina, el primero de ellos se da cuando Charles Darwin en 1859 plantea el principio de selección natural como un proceso que afecta a la totalidad de los organismos, iniciando la observación de la especie más allá del nivel orgánico individual. Esta fue la introducción de un nuevo nivel de organización dentro de la teoría biológica. Hacia 1865, el planteamiento de Gregor Mendel sobre las unidades de la herencia como unidades de carácter discreto, inició el desarrollo del análisis matemático de las mismas porque a partir de ese momento las “partículas” de la herencia conceptualmente se hicieron cuantificables. Con el nacimiento mismo de la genética (propuesta como una disciplina nueva en 1903), Godfrey Harold Hardy y Wilhelm Weinberg formularon el principio de equilibrio que se constituye en uno de los pilares para el desarrollo de la genética de poblaciones, debido a que se establece como un punto cero; permitiendo determinar la diferenciación en valores de frecuencias alélicas en una población y a partir de dicha diferencia, establecer los mecanismos de cambio evolutivo en el acervo genético de una población. En la década de 1920 a 1930 Ronald Fisher, J.B.S. Haldane y Sewall Wright alcanzaron un entendimiento matemático completo del concepto de selección, planteado por Darwin que constituye el cuerpo formal de la síntesis evolutiva que estableció las principales fuerzas del cambio evolutivo: selección, migración, mutación y deriva genética (Fisher, 1958).

Hasta 1960 la genética de poblaciones tuvo básicamente un gran desarrollo teórico; sin embargo, el uso de aloenzimas y secuencias de aminoácidos a principios de los 70's indujo un profundo cambio, debido a la posibilidad de evaluar la composición de una población de manera directa a partir de polimorfismos genéticos (Lewontin, 1985). A mediados de los 70's se inició el estudio directamente sobre el ADN, lo que ha cambiado completamente la interpretación de la variación poblacional, indicando que ésta es más grande de lo que se había pensado; incluso, al interior de las especies, llevando nuevamente al cuestionamiento de este concepto. El uso de datos moleculares ha generado un nuevo conjunto de preguntas, que para su consideración han requerido

de la competencia de profesionales en áreas más allá de la biología; como ingeniería, estadística, matemáticas, antropología y demografía (Blinj, 1999).

Estructura Investigativa de la Genética de Poblaciones

La genética de poblaciones, en su estructura investigativa, tiene elementos empíricos, experimentales y teóricos. El enfoque empírico de la genética de poblaciones, comprende la observación de la variación genética a partir de caracteres morfológicos, análisis de polimorfismos moleculares y estudios cromosómicos. Actualmente, los modelos estadísticos basados en ADN, se centran en la variación de secuencias a nivel inter e intraespecífico. En cuanto al enfoque experimental, los trabajos clásicos se centran sobre fenómenos que se dan al someter poblaciones a nuevos ambientes, para luego ser comparadas con poblaciones que se han mantenido naturales, y posteriormente determinar el efecto del medio ambiente en la variación genética. El trabajo de laboratorio junto con los modelos computacionales provee soporte para las hipótesis desarrolladas a partir de datos empíricos. En años recientes, la definición de experimento genético se ha ampliado; hoy en día muchos de ellos se centran en la comparación de secuencias de ADN entre organismos que tienen diferentes historias evolutivas, variaciones en las funciones o diferentes procesos en su desarrollo. En cuanto al tiempo, se pueden emplear marcadores moleculares para evaluar una población en el momento actual o para hacer la reconstrucción de su historia a lo largo de un período de tiempo específico (Hartl, 1997).

Los modelos teóricos, proveen un marco general para evaluar el efecto de factores evolutivos en los niveles de variación y patrones de relación genética en diferentes ambientes. Estos modelos permiten además, predecir el cambio futuro de una población en diferentes escenarios. Sin embargo, estos análisis deben extrapolarse con cuidado debido a que las investigaciones teóricas se fundamentan en supuestos específicos y pueden tener solamente sentido en algunas especies o ser útiles solamente en algunos contextos de la realidad biológica (Hedrick, 2005).

La genética de poblaciones según sugieren algunos autores, es única dentro del desarrollo de las

ciencias biológicas, debido a que antes de tener datos experimentales que fueran significantes para la construcción de una disciplina científica, fue elaborada como un área teórica. Esto sugiere que el temprano énfasis en la construcción teórica causó un distanciamiento entre la parte teórica, y el intercambio entre los aspectos empíricos y experimentales (Hedrick, 2005).

En la genética de poblaciones actual, los experimentos son de carácter estadístico y deben ser diseñados según las premisas básicas del diseño experimental: primero, la existencia de un concepto de réplica que sea de carácter independiente; segundo, deben desarrollarse los controles experimentales del caso; y tercero, obtener un adecuado tamaño de muestra para disminuir el riesgo de sesgos estadísticos. Este concepto está asociado al poder estadístico, que es la probabilidad de rechazar una hipótesis nula cuando ésta es falsa. Una de las características del experimento estadístico, es la inseguridad que genera en relación a si las muestras poseen el suficiente poder estadístico, de tal forma que sea posible la utilización de herramientas inferenciales; lo que llevaría a rigurosos procedimientos para determinar una muestra al momento de un estudio poblacional (Gale, 1980).

Modelos en genética de poblaciones

De los diferentes tipos de modelos a emplear en genética, los gráficos y las representaciones matemáticas son los más utilizados. Los modelos tienen la ventaja de describir el proceso y la organización de un fenómeno natural. Una de las características más deseables de un modelo es que tenga alta precisión respecto a la descripción de los fenómenos observados y que sea consistente con las observaciones; sin embargo, los supuestos hechos para propósitos matemáticos podrían algunas veces dar una visión imprecisa de la realidad biológica. Para mejorar la calidad de un modelo, se debe realizar una selección de supuestos que puedan ser similares o extrapolables al mundo natural; permitiendo estudiar fenómenos evolutivos entre las especies (Lewontin, 1985).

Los modelos matemáticos dentro de la genética de poblaciones pueden ser determinísticos, es decir, aquellos que dados unos valores particulares de los parámetros, arrojan una estimación

puntual que puede llevarse a eventos futuros. Sin embargo, la mayoría de las veces los modelos son de naturaleza estocástica. En estos modelos, los parámetros se introducen asumiendo la existencia de valores de probabilidad para diferentes escenarios teóricos. Como resultado de este fenómeno, no siempre es posible, y de manera exacta predecir un evento futuro. En los modelos estocásticos se trabaja dentro de márgenes de probabilidad con la finalidad de acercarse al verdadero valor de un parámetro en una población. (Hedrick, 2005).

Dado que los modelos estocásticos están asociados a valores de probabilidad, algunas veces puede usarse una distribución de dicha naturaleza para determinar estos valores; sin embargo, si los modelos tienden a ser complejos, se prefiere realizar simulaciones basadas en números aleatorios llamadas simulaciones de Montecarlo que pueden utilizarse para estimación de parámetros; por ejemplo para la estimación de frecuencias alélicas (Guo & Thompson, 1992).

En muchos problemas de la genética de poblaciones, las distribuciones de probabilidad no son normales o no pueden ser asumidas como tal; cuando esto ocurre, es útil estimar intervalos de confianza alrededor de la media, antes de realizar un proceso de aleatorización en el que la probabilidad de los valores más extremos, puede ser calculada exactamente o estimada con una simulación computacional (Manly, 1991).

Algunos estudios relacionados en Colombia

Los primeros trabajos realizados en el país para la investigación biológica de la paternidad, se realizaron entre 1972 y 1995, a partir de grupos sanguíneos y fueron desarrollados por el Instituto Colombiano de Bienestar Familiar (ICBF). Hasta ese momento todas las pruebas antropoheredobiológicas se realizaban en Bogotá, lo que causó grandes represamientos de resultados. Es por esto, que en 1995 el ICBF inició un proceso de descentralización para las pruebas de paternidad, dirigiéndose a la contratación de laboratorios en las universidades públicas y posteriormente a laboratorios de carácter privado. Esto condujo a un importante desarrollo tecnológico y técnico-científico; se pasó de las pruebas con grupos sanguíneos y polimorfismos de

HLA, a la utilización de sistemas con un gran poder de discriminación como son los microsatélites. Estos avances crearon el ambiente apropiado para la discusión y posterior legislación sobre pruebas de filiación e identificación en Colombia (Bravo, 2001; 2003).

En la actualidad, se sigue utilizando esta tecnología con un mayor número de sistemas microsatélites, hasta llegar a utilizar los empleados por el CODIS, desarrollado por el FBI y validado a nivel internacional. Para estos marcadores moleculares, a nivel regional cada laboratorio ha realizado publicaciones particulares de frecuencias alélicas que se utilizan en la práctica cotidiana (Acosta et al., 2002; Benitez et al., 2003; Bustos et al., 2001; Groot et al., 1992; Duran et al., 2000; Rey et al., 2003; Vargas et al., 2002; Paredes et al., 2003; Usaquen, 2006; Gómez et al., 2003).

En el año 2003, ante el gran volumen de trabajo, el ICBF contrató la tipificación de aproximadamente 40.000 personas con laboratorios de todo el país, constituyéndose en una de las tomas de datos más grandes a nivel mundial. Una compilación y análisis de esta información, permitiría ubicar a Colombia a la vanguardia en el mundo de los análisis genético poblacionales; lo cual sería un estímulo para el desarrollo de nuevos métodos de análisis, tarea que a la fecha no ha sido realizada.

Definición de Sistemas Microsatélites

Los microsatélites también conocidos como STR's (*short tandem repeats*) son repeticiones en tándem de unidad pequeña de dos a seis pares de nucleótidos que corresponden aproximadamente al 20% del genoma humano, y son sistemas de herencia codominante que han mostrado ser altamente polimórficos. Hasta el momento no hay claridad sobre su significado funcional y por tanto, se encuentran como sistemas sin función; sin embargo, algunos autores piensan que contienen exones y se hallan asociados con la expresión de algunas enfermedades (Goldstein & Schlotterer, 1999).

Los microsatélites están distribuidos en todo el genoma eucarionte. Los motivos de repetición a lo largo de los diferentes genomas suelen variar sustancialmente, siendo la repetición AT la de mayor frecuencia (Goldstein & Schlotterer, 1999).

Un aspecto de gran relevancia, para la utilización de microsatélites como marcadores en procesos de identificación, es que no están sometidos a procesos de selección debido a que no presentan una secuencia codificante. Por otra parte, las regiones flanqueantes de los microsatélites son secuencias altamente conservadas, lo que permite su utilización para realizar inferencia filogenética de diferentes especies (Avise, 1996).

Clasificación

La clasificación de los microsatélites se hace de acuerdo con el patrón de ordenamiento de los motivos de repetición (Chambers & MacAvoy, 2000):

- Simple: Un solo motivo repetido n veces en serie. Ej. (AC)⁹
- Simple interrumpido: Un solo motivo repetido n veces, donde se intercalan nucleótidos entre las distintas repeticiones. Ej. (CA)²AA(CA)¹²
- Compuestos: Dos o más motivos repetidos en serie. Ej. (GT)²(TG)¹⁰
- Compuestos interrumpidos: Al menos uno de sus motivos presenta nucleótidos intercalados. Ej. (CT)⁴(GT)²CTAT(GT)¹⁵
- Complejos: Combinaciones entre cualquiera de las clases anteriores, sin ningún patrón de orden definido. Ej. (ACC)⁸TG(GA)¹²(TTA)⁵GC(TTA)⁴.

Teoría de Muestreo

En la investigación genética, es muy frecuente la utilización de estudios por muestreo; sin embargo, la teoría de muestreo suele emplearse mal en estas áreas, en parte por desconocimiento del investigador; pero también desde un punto de vista biológico, por la imposibilidad técnica de

construir los marcos de referencia necesarios para definir un muestreo probabilístico; razón por la cual, la mayoría de investigaciones se concentran sobre muestreos no probabilísticos, por ejemplo, por norma o por conveniencia.

Conceptos iniciales de teoría de muestreo en genética.

Universo: Es el conjunto total de elementos bajo estudio o sobre los cuales se desea efectuar un proceso de inferencia, que permite la proyección de resultados obtenidos de la muestra a toda la población donde se realizó el muestreo. De ahí, la importancia que la muestra represente lo mejor posible a la población objetivo (Ospina, 2001).

Población estadística: Es el conjunto de mediciones que se hacen sobre los elementos muestreados (Ospina, 2001). Un solo conjunto de muestras biológicas podría dar lugar a diferentes poblaciones estadísticas, lo que se denomina población multivariada. En el caso genético, se tiene que cada posible persona de la población se examina a diferentes niveles y para múltiples genes. Cuando se necesita establecer distribuciones de frecuencias genotípicas, cada elemento del universo es una persona; sin embargo, al estudiar frecuencias alélicas se cambia el nivel de observación y se pasa a estudiar alelos individuales. Por otra parte si se desea estudiar marcadores de cromosoma Y para estudios de ancestría, entonces el nivel de observación se transfiere a la observación haplotípica (Ospina, 2001).

Métodos de muestreo: Se refieren al conjunto de técnicas estadísticas que estudian la forma de seleccionar una muestra representativa de una población y cuya información permita inferir las propiedades o características de toda la población asumiendo un error medible y controlable. A pesar de los métodos propuestos por la teoría de muestreo, se han desarrollado alternativas diferentes no siempre basadas en los supuestos formales, lo que puede limitar el proceso de inferencia pero no invalida necesariamente las investigaciones biológicas. Los métodos de Captura-recaptura y las técnicas de remoción ampliamente utilizados dentro de estudios ecológicos, son ejemplos de métodos que no pueden ser clasificados estrictamente como métodos

de muestreo; sin embargo, si se tiene el suficiente cuidado en el diseño experimental y en los procedimientos analíticos, permitirán realizar inferencias acertadas y cálculos de los niveles de confianza para la población (Ospina, 2001).

Estimadores: un estimador es un estadístico (determinado a partir de una muestra), usado para acercarse a un parámetro desconocido de la población. Los estimadores, son funciones matemáticas, que se comportan como variables aleatorias. En general, un buen estimador es aquel que es insesgado, eficiente, convergente, es decir que en los diferentes modelamientos genera un valor dentro de un intervalo de confianza y además robusto. El valor de un estimador, proporciona una estimación puntual del valor del parámetro en estudio. En general, se prefieren las estimaciones mediante intervalos, es decir un intervalo $[a, b]$ dentro del cual se espera encontrar el valor real del parámetro, con un determinado nivel de confianza (López, 2005).

La Investigación por muestreo.

Las investigaciones por muestreo son ampliamente utilizadas debido a la necesidad de obtener conclusiones de una población a partir de un grupo de individuos. Para la realización de una investigación por muestreo, autores como Bautista (1998) proponen los siguientes pasos:

1. Diseño muestral
2. Mediciones a tomar
3. Trabajo de campo
4. Análisis estadístico
5. Diseño Muestral

Esta etapa incluye desarrollar un plan de muestreo y los procedimientos de estimación. Por plan de muestreo, se entiende la metodología utilizada para seleccionar los individuos de la muestra en una población determinada. Los procedimientos de estimación son los algoritmos o fórmulas usadas para obtener estimaciones de valores poblacionales y su confiabilidad a partir de datos muestrales.

La selección de un diseño muestral se basa en tres aspectos básicos: el tipo de variables, las estimaciones requeridas, y los niveles de confiabilidad y validez de acuerdo con los recursos existentes (Bautista, 1998).

La manera como son tomados los datos depende del tipo de muestreo, es decir, si este es probabilístico o no. Una preparación pobre de la información o una deficiente capacitación frecuentemente conduce a fracasos en el estudio. Cuando las mediciones se refieren a unidades diferentes de personas, como animales, plantas o información de laboratorio, el problema es más sencillo de manejar ya que la información no está sujeta a opiniones o situaciones emocionales, como ocurre en el caso de las personas; sin embargo, en estos casos se deben implementar los debidos controles experimentales como una medida de aseguramiento de la reproducibilidad de la información (Bautista, 1998).

Una vez realizado el diseño de los instrumentos, es conveniente en muchos casos tomar una muestra piloto, que es muy reducida y permite probar los instrumentos de medición y eliminar los defectos del proceso a desarrollar. Esto permite además evaluar la presencia de errores tanto de muestreo como de medición.

Muestreo probabilístico

Es aquel en el cual se puede establecer la probabilidad de obtener cada una de las muestras seleccionadas a partir de un marco referencial utilizando un procedimiento aleatorio. Las principales características de lo que se puede definir como un muestreo aleatorio son:

1. Poder definir el conjunto total de muestras posibles que pueden seleccionarse de la población de acuerdo con el procedimiento muestral. Este conjunto determina el marco referencial.
2. Conocer para cada una de las muestras posibles, la probabilidad de ser seleccionada.

3. El procedimiento utilizado debe dar a cada elemento de la población una probabilidad de selección diferente de 0.
4. La selección debe ser aleatoria, esto es, el mecanismo de probabilidad diseñado para la selección debe ser tal que cada muestra tenga la misma probabilidad de selección en cualquier parte del espacio geográfico (Ospina, 2001).

Muestreo no probabilístico.

En general, todo tipo de muestreo que no cumpla con alguna de las condiciones enumeradas anteriormente es un muestreo no probabilístico. Los siguientes son ejemplos de muestreo no probabilísticos frecuentemente empleados.

1. La muestra ha sido restringida a la parte de la población que es fácilmente accesible, como es el caso más común en los análisis genético-poblacionales en humanos, debido a que estos se realizan con los usuarios de los servicios de filiación y genotipificación.
2. La muestra se selecciona teniendo en cuenta el azar más no un procedimiento de aleatorización.
3. Con una población heterogénea y pequeña, el investigador inspecciona la población en general y selecciona una muestra pequeña de unidades tipo.

En condiciones apropiadas, cualquiera de los anteriores métodos puede proporcionar resultados útiles. Sin embargo, en ellos no se puede aplicar estrictamente la teoría de muestreo ya que ésta se basa en el supuesto de selección aleatoria, que es el principal problema del muestreo no probabilístico, de aquí que haya una precisión indeterminada. En este tipo de muestras la precisión se sustenta en la experiencia de quien selecciona la muestra. En el caso de las muestras probabilísticas no se presenta este problema, pues el proceso de selección ayuda a lograr un grado de precisión (Ospina, 2001).

Plan de Muestreo Estadístico

Este se establece en función de los objetivos, los recursos con los que cuenta el laboratorio y la calidad deseada de los estimadores poblacionales, estos aspectos determinan las decisiones de tipo operativo que paso a paso se van tomando. El plan debe estar acompañado de una estrategia de crítica e imputación a fin de evitar errores que generen sesgos y que pueden llegar a invalidar los resultados globales del estudio (Bautista, 1998).

Pasos para la construcción del plan de muestreo

Los pasos que se siguen para la elaboración del plan de muestreo, son similares a los que deben seguirse para la elaboración de cualquier proyecto de investigación. Se debe iniciar por la definición precisa de todas las variables y el establecimiento de un plan de análisis asociado con los objetivos de la investigación para garantizar que la información necesaria se recolecte y para evitar que mucha información que se recogió con un gran costo quede sin uso. Según Bautista (1998) los siguientes pasos deben seguirse para el diseño de un plan de muestreo:

1. Definición concreta y clara de los objetivos.
2. Traducción de los objetivos en resultados estadísticos.
3. Especificación de las variables y parámetros de la población objetivo.
4. Construcción del marco de muestreo.
5. Inventario de recursos disponibles.
6. Establecimiento de cronogramas.
7. Determinación de la calidad de las estimaciones realizadas.
8. Especificación de los métodos de recolección de información.
9. Especificación del diseño muestral.
10. Determinación de procesamiento de crítica e imputación.
11. Especificación de las fórmulas de estimación y de las medidas de calidad.
12. Planeación del trabajo de campo.
13. Diseño del plan de control y evaluación.

Fuentes de sesgos y varianza

Estos son los dos defectos principales que deben ser evitados a cada paso y en cada decisión del proceso de planeación de un estudio. El sesgo reduce el nivel de confiabilidad del intervalo de confianza, la varianza amplía la longitud del intervalo. Los errores que se pueden cometer y los procesos que deben ser revisados a fin de evitar y reducir la producción de sesgos son (Bautista, 1998):

1. La selección: esta puede generarse debido a la calidad del marco generando problemas de cobertura.
2. Inadecuada recolección debida a errores: en los procesos de entrevistado, falta de motivación, deficiencias en el cuestionario, inadecuado proceso de recolección.
3. Procesamiento: errores en los procesos de codificación, crítica e imputación y finalmente errores en el cálculo de estimativos.
4. Verificación: Validación y consistencia de la información.

Una vez definido el plan completo de crítica se deben realizar las tres operaciones siguientes.

1. Verificación: es el proceso que generalmente acompaña a la grabación de la información y consiste en garantizar que la información almacenada es un fiel reflejo de lo contenido en las encuestas. Generalmente este proceso se realiza como un segundo paso en el proceso de registro de la información.
2. Validación: es el proceso por el cual se determina si los datos cumplen ciertas reglas preestablecidas de aceptabilidad. Estas reglas constituyen la crítica estadística.
3. Consistencia: Se parte del principio de realizar el mínimo cambio posible de datos para obtener una base de datos completa y corregida. Esto requiere del desarrollo

de reglas de crítica para determinar los procesos de cambio de información (Bautista, 1998).

Análisis Multivariado.

Las técnicas de análisis multivariado son ampliamente conocidas y utilizadas en la actualidad en los estudios genético-poblacionales debido a la capacidad que tienen para analizar grandes conjuntos de variables, justamente lo que sucede cuando se trabaja con sistemas genéticos polimórficos. En una matriz de datos construida a partir de las distribuciones de frecuencias alélicas de sistemas microsatélites, cada uno de los alelos del sistema se asume como una de las variables en estudio (columnas) y cada uno de los individuos (filas) corresponde a las poblaciones que se analizan. A partir de este modelo se realiza clasificación y comparaciones entre poblaciones, que es uno de los objetivos principales en la caracterización genética de poblaciones. Las técnicas principalmente utilizadas son componentes principales, correspondencias múltiples y escalamientos multidimensionales (Cavalli-Sforza y Bodmer 1999). Pese a que estas técnicas fueron desarrolladas desde principios del siglo XX, su implementación en la genética se inicia hasta los años 60 debido a dos situaciones favorables: Se alcanzó el suficiente conocimiento teórico sobre la naturaleza de los polimorfismos genéticos y el desarrollo de las técnicas de laboratorio adecuadas para obtener datos confiables; en segundo lugar, ya se contaba con las herramientas computacionales que permitían la utilización de algoritmos multivariados, hasta ese momento tenidos como procedimientos bastante complejos y tediosos de desarrollar (Dutta, 2002).

Otra técnica muy utilizada es el análisis de conglomerados; a partir de este se establecen las relaciones de agrupamiento entre poblaciones. Esta técnica es diferente del análisis cladístico y filogenético, en el cual se parte de un grupo ancestral para luego ir reconstruyendo la historia evolutiva de un conjunto de taxones. En este caso, los dendrogramas son representaciones que carecen de una medida directa de tiempo, por lo tanto son reconstrucciones actuales de poblaciones; sin embargo, como lo sugiere Nei (1987) pueden ser reconstrucciones filogenéticas aceptables, cuando la distancia seleccionada genera unos valores acordes con el pasado evolutivo

de los organismos. En el caso de la especie humana se puede hablar también de reconstrucciones filogenéticas cuando los dendrogramas construidos reflejan procesos migratorios, las fuerzas de selección y los procesos de deriva por los que ha pasado una población. En el caso más local, los dendrogramas construidos hasta el momento (Usaquen, 2006) muestran cómo los procesos de poblamiento en Colombia tienen una alta asociación con los procesos históricos regionales.

La base para los análisis de conglomerados son las matrices de distancias; desde el punto de vista estadístico, una distancia es una métrica que cumple con las siguientes condiciones (Díaz, 2002):

1. No negatividad
2. Simetría
3. Desigualdad triangular
4. Identificación de no identidad
5. Identidad

En genética se utilizan distancias como Czekanowski, Manhattan, Rogers que cumplen las condiciones anteriores. Otras distancias muy robustas, pero que no cumplen todos los supuestos anteriores, son por ejemplo la distancia de Mahalanobis utilizada cuando los caracteres a estudiar están correlacionados. Otra distancia desarrollada específicamente para datos genéticos es la distancia de Cavalli-Sforza Edwards. Ésta representa las poblaciones como vectores sobre la superficie de una esfera multidimensional, donde la distancia está dada por los ángulos formados por dichos vectores (Nei, 1987).

Para la obtención de una clasificación juegan un papel importante la distancia seleccionada y la metodología de clasificación propuesta. Una combinación especialmente eficiente es la distancia de Cavalli-Sforza Edwards junto con el método Ward. Éste se ha ensayado para las diferentes poblaciones obteniéndose clasificaciones muy acordes con procesos de poblamiento histórico (Usaquén, 2006).

Para una buena interpretación multivariada de los datos se sugiere la combinación de diferentes técnicas, buscando encontrar consistencia entre los diferentes ordenamientos, especialmente los ordenamientos producidos por planos factoriales y los métodos de construcción de dendrogramas (Cavalli – Sforza et al., 1994).

La genética de poblaciones en la actualidad combina el desarrollo teórico junto con el desarrollo experimental. Ha pasado a ser una disciplina práctica cuyos resultados son el fundamento para la sistemática molecular, la identificación forense tanto en humanos como en otras especies y la toma de decisiones en la conservación genética (siendo considerada en esta última como una herramienta esencial). Sin embargo, las conclusiones deben estar basadas en análisis estadísticos minuciosos, empezando por el análisis en la consistencia de la muestra analizada, tal que el investigador tenga mejores parámetros para llegar a conclusiones acertadas. En este momento, es necesario el desarrollo de herramientas teóricas y computacionales acorde con los nuevos marcadores moleculares que se están implementando. Esta es la línea de pensamiento planteada en el presente trabajo.

Bibliografía

- ACOSTA, M; BRION, M; LAREU, M; CARRACEDO, A; 2002. Genetic data on eight STRs (D5S818, D7S820, F13B, LPL, TH01, TPOX, VWA31, CSF1PO) from a Colombian population. *Forensic Science International*. V 129. pp 216-218.
- AVISE, J. 1996. *Conservation genetics. Case Histories from Nature*. New York, Kluwer Academic Press.
- BAUTISTA, L. 1998. *Diseño de Muestreos Estadísticos*. Bogotá, Universidad Nacional de Colombia.
- BENÍTEZ, A. & OSSA, H. 2003. Allelic frequencies at 12 STR loci in Colombian population. *Forensic Science International*. V 136. pp 86-88.
- BLINJ, H. & MURPHY, A. 1999. *Human Geography. Culture Society and Space*. Londres: Wiley and Sons.

- BRAVO, M. 2001. Introducción a la genética forense de las pruebas de paternidad. Medellín, Universidad de Antioquia.
- BRAVO, M. 2003. Legislación sobre la investigación científica de la paternidad biológica: ley 721 de 2001. Medellín, Universidad de Antioquia.
- BUSTOS, I; ACOSTA, M; BRAVO, M; BUILES, J; CARABALLO, L. 2001. Colombia: País genéticamente megadiverso en poblaciones humanas caracterizadas por sistemas STR's. VI Jornadas de Genética forense. GEP-ISFG. Córdoba, Argentina.
- CAVALLI-SFORZA, L.L; MENOZZI, P; PIAZZA, A. 1994. The history and Geography of Human Genes. New York. Princeton University Press.
- CAVALLI-SFORZA, L.L; BODMER, W. F. 1999. The genetics of Human Populations. New York: Dover Publicatons.
- CHAMBERS, G & MACAVOY, E. 2000 Microsatellites: consensus and controversy. Comparative Biochemistry and Physiology Part B 126 (2000) 455–476.
- DIAZ, L. G. 2002. Estadística multivariada: inferencia y métodos. Bogotá: universidad nacional de Colombia.
- DURAN, R. 2000. Análisis de las frecuencias alélicas de los loci D1S80, vWA y TH01 en un grupo de individuos de Santa fe de Bogota. Para aplicación en Investigación Criminal. Tesis Universidad Nacional de Colombia.
- DUTTA, R. 2002. Patterns of Genetics Diversity at the Nine Forensically Approved STR Loci in the Indian Population. Hum Biol. 2002 Feb;74(1):33-49.
- FISHER, R. 1958. The Genetical Theory of Natural Selection. Londres: Oxford University Press.
- GALE, J. S. 1980. Population genetics. London: Blackie & son limited.
- GILLESPIE, J. H. 1998. Population Genetics. A concise guide. The Johns Hopkins University Press. Baltimore.
- GOLDSTEIN D & SCHLOTTERER C. 1999. Microsatellites. Londres: Oxford University Press.

- GÓMEZ, M. V; REYES, M. E; CÁRDENAS, H & GARCÍA, O. 2003. Genetic variation for 12 STRs loci in a Colombian population (Department of Valle del Cauca). *Forensic Science International*. V 137.
- GROOT, H. 1992. Estructura genética de la comisaría de Amazonas. Universidad de Los Andes. 104
- GUO, S. A & THOMPSON, E. A. 1992. Performing the exact test of Hardy Weinberg proportion of multiple alleles. *Biometrics* 48 361-372
- HARTL, D. 1997. Principles of population genetics. Sunderland, Massachusetts. Sinauer Associates Inc.
- HEDRICK, P. W. 2005. Genetics of populations. Tercera Edición. Jones and Bartlett Publishers. Massachusetts.
- LEWONTIN, R. 1985. Population genetics. *Annual Review of Genetics*. Vol 19.
- LOPEZ, C. 2005. Muestreo Estadístico, Conceptos y problemas resueltos. Prentice may. Madrid.
- MANLY, B. 1991. Randomization Monte Carlo Methods in Biology. Chapman Londres.
- NEI, M. 1987. Molecular Evolutionary Genetics. Columbia University Press. New York
- OSPINA D. 2001. Introducción Al Muestreo. Facultad de Ciencias UN.
- PAREDES et al. 2003. Analysis of the CODIS autosomal STR loci in four main Colombian regions. *Forensic Science International*. 137; 67-73.
- REY, M; GUTIERREZ, A; SCHROEDER, B; USAQUEN, W; CARRACEDO, A; BUSTOS, I; GIRALDO, A. 2003. "Allele frequencies for 13 STR's from two Colombian populations: Bogotá and Boyacá". *Forensic Sci Int*. 136:83-85.
- USAQUÉN, W. 2006. Caracterización genética de 23 poblaciones humanas colombianas a partir de 9 sistemas microsatélites. Tesis: Universidad Nacional de Colombia.
- VARGAS, CI; CASTILLO, A; GIL, A.M; PICO, A.L; GARCÍA, O. 2002. "Allele frequencies for the AmpFISTR profiler loci in a Colombian population (Department of Boyacá)". *J Forensic Sci*. 47:406-407.

Capítulo No. 2: Una nueva aproximación al cálculo de tamaño de muestra en estudios genético poblacionales y forenses a partir de marcadores microsatélites.

W. Usaquén¹, J. Corzo², L.F. García³, M. Rojas¹, A. Alonso¹, A. Casas¹.

1. Instituto de Genética Universidad Nacional de Colombia, 2. Departamento de Estadística Universidad Nacional de Colombia. 3. Departamento de Biología. Universidad Nacional de Colombia.

Resumen

Una pregunta frecuente en los estudios poblacionales es si el tamaño de muestra utilizado es suficiente para obtener estimadores genéticos confiables, como por ejemplo el cálculo de frecuencias alélicas. El primer problema en la resolución de esta incógnita es el procedimiento para seleccionar las muestras. Los estudios genéticos generalmente utilizan técnicas de muestreos por norma o por conveniencia. Sin embargo por esta vía no es posible la estimación del error muestral, por tanto se hace difícil evaluar si la variación de los estimadores poblacionales es debida a variaciones evolutivas o a muestreos deficientes.

En este trabajo se presenta un procedimiento para determinar tamaños de muestra y evaluar la calidad de un estimador genético en función de su dispersión. Se analizaron frecuencias alélicas de sistemas microsatélites. El procedimiento planteado se fundamenta en el aumento secuencial e iterativo del tamaño de muestra, obteniendo estimadores por intervalos de confianza asociados a la frecuencia de cada variante alélica.

Los resultados indican un tamaño de muestra aproximado de 700 personas para una población con las características de Bogotá, es decir de múltiples orígenes étnicos en los cuales el nivel de polimorfismo se espera sea alto. Para la evaluación de este modelo se utilizaron sistemas genéticos microsatélites (STR's: *Short Tandem Repeats*) utilizados en el *combined DNA index System* CODIS; sin embargo, el uso de esta metodología podría extenderse a cualquier sistema del que se pueda determinar la composición genotípica, como por ejemplo los sistemas SNP's. Los límites de los intervalos de confianza también podrían emplearse en el cálculo de índices de paternidad o en estudios de subestructura genética.

Keywords.

STRs, Paternity, Validation, Sample Size.

1. Introducción.

Al momento de seleccionar un grupo de personas para un estudio genético poblacional, el cálculo de tamaño de muestra suele obviarse, en buena parte por la dificultad de aplicar los supuestos y la metodología de la teoría formal de muestreo. La alternativa más utilizada es seguir las recomendaciones hechas por asociaciones internacionales tales como el “*committee on DNA technology in forensic science*” y la “*International Society for Forensic Genetics*” (ISFG), organizaciones que desde el año 1992 sugerían tamaños de muestra aproximados de 200 individuos (ISFH, 1992). Estos valores se han incrementado, en parte por la acumulación de información en las diferentes bases de datos; pero también por críticas teóricas al reducido tamaño de muestra (Evet y Weir 1998), especialmente sobre los tests para independencia de frecuencias alélicas los cuales podrían tener bajo poder estadístico.

En el informe del año 2002 de la ISFG, también en el 2004, se sugieren tamaños de muestra superiores a 500 personas, sin que se presente una justificación estadística para esta cifra. A pesar de esto, se sigue careciendo de procedimientos para evaluar la exactitud de la información, entendida esta como el cálculo de la variación en los estimadores poblacionales. Una alternativa puede ser el cálculo de intervalos de confianza que permitan un mejor acercamiento a los verdaderos parámetros poblacionales.

Por tanto, los muestreos que se utilizan en los estudios poblacionales corresponden a muestreos no probabilísticos provenientes de servicios de identificación genética. El principal argumento a favor de este tipo de muestreos surge de la neutralidad evolutiva asumida para los sistemas microsatélites, es decir, que al no tener ninguna expresión fenotípica no estarían sometidos a procesos de selección natural directa, y por lo tanto su distribución en la población dependería directamente de su frecuencia poblacional, lo que indicaría independencia de cualquier tipo de sesgo. Este argumento, ignora el efecto simultáneo de todas las fuerzas de cambio evolutivo causantes de un equilibrio dinámico, además de los fenómenos de endogamia y subestructura que

deberían considerarse al momento de realizar estimaciones de frecuencias poblacionales. Por otra parte, algunos autores consideran que los sistemas microsátélites pueden no ser tan neutrales como se ha pensado (Goldstein y Schlotlerer, 1999), por tanto la argumentación de selección de muestras no aleatorias, defendida por la neutralidad de los microsátélites, perdería fundamento.

En cuanto a la selección de individuos, un gran número de reportes de frecuencias en el mundo se han realizado con muestras provenientes de bases de datos ya establecidas en laboratorios de análisis genético como aquellos que realizan pruebas y análisis de filiación e identificación, o instituciones gubernamentales relacionadas con criminalística y casos de desaparición forzada. Por la naturaleza de estos datos, la única información demográfica que suele recuperarse es el lugar de nacimiento de las personas relacionadas. En estas bases de datos se procura no incluir individuos emparentados; sin embargo no se realiza ninguna evaluación sobre el efecto de poblaciones cerradas y de tamaño reducido, incrementando la posibilidad de seleccionar individuos relacionados por parentesco dos o tres generaciones atrás.

El problema de muestreo en genética ha sido abordado por algunos autores entre los que se destacan, Kalinowski, (2005); Campbell, (2000); Chakravarty, (2007); Gillespie, (2001); Buttler, (2008); Hey,(2004). Uno de los trabajos más sofisticados es el publicado por Chakraborty, (1992), quien propone determinar el tamaño de muestra a partir de los alelos más frecuentes debido a que estos son los más representativos en la población, en tanto se ignorarían los alelos de baja frecuencia. Este argumento justifica la clasificación en alelos efectivos y no efectivos, tal como lo realiza Buttler, (2008). Otro trabajo importante es el publicado por Harding (1998) en el que se propuso una metodología para la selección de muestras de bases de datos poblacionales; sin embargo, el autor sólo plantea una alternativa gráfica sin presentar ningún desarrollo estadístico que permita determinar un número mínimo de personas para un estudio poblacional. De más reciente aparición se encuentra el trabajo presentado por Medina (2006) basado en técnicas de remuestreo, con la finalidad de obtener un tamaño de muestra idóneo para evitar sesgos al momento de realizar análisis multivariados de clasificación y agrupamiento de poblaciones.

Dentro de este contexto, este trabajo presenta una nueva aproximación al cálculo de tamaño de muestra en estudios genético poblacionales y forenses a partir de marcadores microsátélites. Un método para determinar intervalos de confianza asociados a la frecuencia de cada variante alélica de un determinado polimorfismo, partiendo de muestreos no probabilísticos. A partir de dichos intervalos de confianza se puede encontrar el margen de variación para estimadores poblacionales y forenses como por ejemplo los índices de paternidad. Los intervalos de confianza pueden ser también utilizados para determinar si dos muestras corresponden a la misma población o no. Hall (1999), Hohls (1998), Wang (2003). La generación de intervalos de confianza diferentes en dos muestreos diferentes sería un claro indicio de procesos de subestructura poblacional.

2. Método propuesto.

El método ha sido organizado en dos fases desarrolladas en forma de algoritmos descritos en los apartes 2.1 y 2.2.

2. 1. Fase 1: Cálculo secuencial de las frecuencias alélicas.

1. Inicialmente se selecciona una muestra m_1 de n personas. Sobre este tamaño se realiza la estimación de frecuencias correspondiente a la distribución de frecuencias que llamaremos d_1 de la muestra m_1 .
2. En el siguiente paso se selecciona una muestra m_2 de la misma población y del mismo tamaño que m_1 , con lo cual se obtiene una muestra de tamaño m_2 .
3. El número de personas n inicial, se mantiene constante a lo largo de los muestreos, cada nueva muestra se suma a la inmediatamente anterior, y de esta manera se continua hasta obtener una sucesión estable en valores de frecuencias alélicas. Con el anterior algoritmo, se construyó la matriz No. 1 de frecuencias para un sistema de k alelos con m_n muestreos.

Sample size.	Allele frequencies				
M_1	f_{11}	f_{12}	f_{13}	...	f_{1k}
M_2	f_{21}	f_{22}	f_{23}	...	f_{2k}
M_3	f_{31}	f_{32}	f_{33}	...	f_{3k}
M_t	f_{t1}	f_{t2}	f_{t3}	...	f_{tk}

Matríz No.1 : Sucesión de distribuciones de frecuencias alélicas a lo largo de los incrementos en el tamaño de la muestra. Con cada nuevo muestreo se obtiene una nueva distribución de frecuencias cuyo incremento es en n individuos.

2.2. Fase 2: Cálculo de coeficientes de variación.

En esta fase se calculan promedios por cada variante alélica desde f_{11} hasta f_{it} , igualmente la matriz de desviaciones estándar también desde f_{11} hasta f_{it} . Posteriormente se calculan los coeficientes de variación. A medida que se da el incremento en el tamaño de muestra se obtienen coeficientes de variación menores lo que permite determinar un punto de corte asociado a un valor específico de tamaño de muestra. Según Bautista, (1998) un valor de coeficiente de variación mínimo en un estudio muestral, es cercano a 0.1. Por tanto cuando se alcanza el valor antes mencionado, se obtiene un tamaño de muestra específico del microsatélite analizado, el cual se considera un tamaño de muestra óptimo, dado que la variación en el valor de frecuencia establecido alcance un valor aceptable en su estimación.

2.2.1. Promedios ponderados de frecuencias.

La matriz de promedio de frecuencias desde f_{11} hasta f_{it} , es decir hasta el último término de la última variante alélica a lo largo de las diferentes distribuciones de frecuencias inicialmente se calcula desde el promedio entre la 1ª y la 2ª muestra, luego la sucesión empieza en la segunda distribución:

<i>Alelo 1</i>	<i>Alelo 2</i>	...	<i>Alelo k</i>
$\bar{x}_{21} = \frac{f_{11} + f_{21}}{m_1 + m_2}$	$\bar{x}_{22} = \frac{f_{12} + f_{22}}{m_1 + m_2}$...	$\bar{x}_{2k} = \frac{f_{1k} + f_{2k}}{m_1 + m_2}$
$\bar{x}_{31} = \frac{f_{11} + f_{21} + f_{31}}{m_1 + m_2 + m_3}$	$\bar{x}_{32} = \frac{f_{12} + f_{22} + f_{32}}{m_1 + m_2 + m_3}$...	$\bar{x}_{3k} = \frac{f_{1k} + f_{2k} + f_{3k}}{m_1 + m_2 + m_3}$
.	.		.
.	.		.
.	.		.
$\bar{x}_{i1} = \frac{f_{11} + f_{21} + \dots + f_{i1}}{m_1 + m_2 + \dots + m_i}$	$\bar{x}_{i2} = \frac{f_{12} + f_{22} + \dots + f_{i2}}{m_1 + m_2 + \dots + m_i}$...	$\bar{x}_{ik} = \frac{f_{1k} + f_{2k} + \dots + f_{ik}}{m_1 + m_2 + \dots + m_i}$
$\bar{x}_{i1} = \frac{\sum_{t=1}^i f_{t1}}{\sum_{t=1}^i m_t}$	$\bar{x}_{i2} = \frac{\sum_{t=1}^i f_{t2}}{\sum_{t=1}^i m_t}$...	$\bar{x}_{ik} = \frac{\sum_{t=1}^i f_{tk}}{\sum_{t=1}^i m_t}$

Matríz No.2 : Promedios de frecuencias calculados por cada variante .

La forma general empleada para calcular los promedios de frecuencias alélicas sería:

Ecuación No. 1:
$$\bar{x}_{.k} = \frac{1}{t} \sum_{i=1}^t f_{ik} \cdot w_i$$

En la expresión anterior aparece un termino w_i , que corresponde a la ponderación a medida que se da un incremento de los muestreos desarrollados, por tanto w_i corresponde a un promedio ponderado. A continuación se presentan los primeros desarrollos hasta llegar a la forma general calculada de la siguiente forma:

$$w_1 = \frac{1}{2n} + \frac{1}{3n} + \frac{1}{4n}$$

$$w_2 = w_1$$

$$w_3 = \frac{1}{3n} + \frac{1}{4n}$$

$$w_3 = \frac{1}{4n}$$

Considerando este desarrollo para b muestras se tiene que:

$$w_1 = \sum_{k=2}^{b+1} \frac{1}{kn} \quad , \quad w_2 = w_1$$

$$w_3 = \sum_{k=3}^{b+1} \frac{1}{kn} \quad , \quad w_4 = \sum_{k=4}^{b+1} \frac{1}{kn}$$

En la forma general se tendría que:

$$w_i = \frac{1}{n} \sum_{k=i}^{b+1} \frac{1}{k} = \frac{1}{n} \cdot \frac{1}{1 - \left(\frac{1}{k}\right)^{b+1}}$$

Ecuación No. 2:

En la ecuación anterior puede verse como el número de alelos incide sobre la velocidad de convergencia $w_i \rightarrow 1/n$ así:

- Si k , es decir el tamaño del polimorfismo, es pequeño w_i tiende a $1/n$ de una manera más lenta. Es decir que se necesita un mayor tamaño de muestra, por lo tanto b es un valor mucho más grande.
- Si k es un valor grande, entonces el cociente $1/k$, tiende a cero a mayor velocidad, por lo tanto se necesita un menor tamaño de muestra, es decir b es un valor más pequeño.

2.2.2. Desviación Estándar.

Posteriormente, la desviación estándar se determinó utilizando las siguientes ecuaciones:

Calculada para la primera y segunda variante alélica:

$$S_1 = \sqrt{\frac{\sum (x_{i1} - \bar{x}_1)^2}{b}}$$

$$S_2 = \sqrt{\frac{\sum (x_{i2} - \bar{x}_2)^2}{b}}$$

La forma calculada para el k ésimo alelo es:

$$S_k = \sqrt{\frac{\sum (x_{ik} - \bar{x}_k)^2}{b}}$$

Donde S_k corresponde a la desviación estándar a lo largo de los muestreos realizados para el alelo kn . En el muestreo específico m . A partir de estos valores posteriormente se desarrolla una matriz de datos escalonada.

2.2.3. Coeficientes de variación.

Finalmente se determina el coeficiente de variación para la distribución de frecuencias en cada variante alélica: Ecuación No. 3.

Ecuación No. 3:
$$CV_{ik} = \frac{S_{ik}}{x_{ik}}$$

2.2.4. Alelos efectivos y alelos raros.

Los alelos en el polimorfismo se clasifican como alelos efectivos y alelos raros según la ecuación No. 4, basada en los valores de la homocigocidad del sistema:

Ecuación No. 4:
$$A_e = \frac{1}{\sum k_i^2}$$

Debido a su comportamiento diferencial respecto al coeficiente de variación, los alelos en baja frecuencia siguen presentando oscilaciones muy altas que no permiten un acercamiento claro a la definición de un tamaño de muestra, y es por esto que el procedimiento se centra en los alelos efectivos.

2.3. Selección de un tamaño de muestra.

Inicialmente se selecciona un tamaño de muestra por cada sistema microsatélite utilizando dos criterios:

1. Un muestreo m_i en el cual aparece la totalidad de los alelos reportados en el muestreo poblacional, incluidos los alelos en baja frecuencia.
2. Un valor menor o igual a 0.05 en las fluctuaciones de los coeficientes de variación de las frecuencias para los alelos efectivos. Puede utilizarse un segundo criterio para la desviación estándar en el valor de frecuencia alélica inferior a 0.01, que es concordante con la definición de polimorfismo, es decir oscilaciones de la frecuencia de un alelo de 1%.

2.4. Análisis multi loci.

El análisis secuencial se realiza para cada uno de los sistemas en estudio, por tanto se obtiene para cada uno de ellos un tamaño de muestra m_i por sistema, luego el N final corresponderá al mayor valor m reportado en el conjunto total de marcadores evaluados. El valor N final se obtiene a partir de múltiples iteraciones, para estimar la dispersión en los valores de frecuencias alélicas.

2.5. Aplicación desarrollada.

Una aplicación computacional fue desarrollada en Visual Basic Applications (VBA) y los datos manipulados en microsoft Access ® por la facilidad de manejo y de implementación en los laboratorios de identificación, que suelen trabajar con hojas de cálculo generadas por los diferentes modelos de analizadores genéticos y como herramienta para la utilización y apropiación del

método desarrollado. Los algoritmos desarrollados siguen los lineamientos de calidad para un laboratorio de ensayos ISO-17025 e ISO 9001, además de los métodos de cálculo programados para la resolución de los ejercicios de proeficiencia de la ISFG.

2.6. Datos empleados.

Una vez obtenido el algoritmo, este se verificó utilizando los datos poblacionales registrados por el grupo de Genética de Poblaciones e Identificación del Instituto de Genética de la Universidad Nacional de Colombia (GPI-IGUN). La población de Bogotá (Ciudad Capital) es un buen modelo para el análisis de tamaño de muestra y variabilidad de un polimorfismo, cuenta con una población de aproximadamente 7.000.000 millones de habitantes en la que se da un activo proceso de migración desde todo el país e incluso de algunos países andinos cercanos, también recibe influencia de la región Caribe, con una marcada influencia de grupos afro descendientes y en menor proporción de población Europea. También se encuentran grupos de ancestría indígena del centro del país. Todos los datos de perfiles genéticos utilizados cuentan con su respectivo consentimiento informado siguiendo la normatividad respectiva para su utilización. Los sistemas microsatélites empleados son los utilizados por el sistema CODIS, 13 en total. El método se explica en su totalidad utilizando el microsatélite D5S818.

3. Resultados.

3.1. Cálculo inicial para un solo sistema STR.

El procedimiento se explica en detalle con el sistema D5S818:

1. Se utilizó una base de datos con 1.331 personas, sobre la cual se realizaron un total de 100 muestreos (m_{100}). La gráfica No.1 muestra las oscilaciones de cada variante alélica a medida que se incrementa gradualmente el tamaño de muestra.

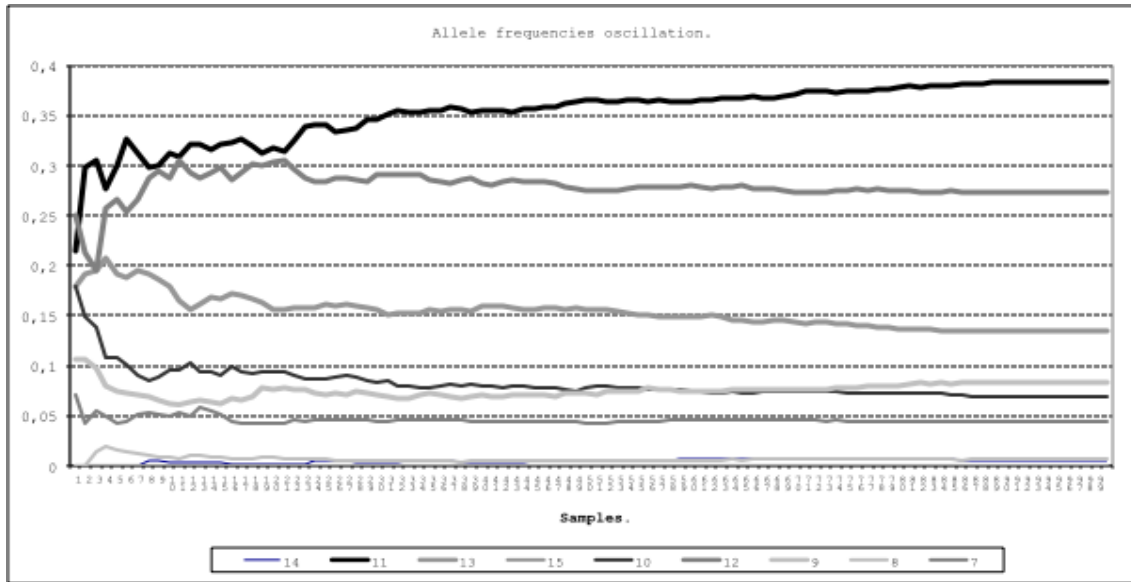


Figura No 1. Comportamiento de las frecuencias alélicas del sistema D5S818 a lo largo de 100 muestreos secuenciales. Cada muestreo adiciona un número n de personas. Las distribuciones iniciales presentan mayores fluctuaciones en el valor de frecuencias entre muestreos, sin embargo poco a poco se tiende a un valor estable para cada variante alélica acercándose al valor real del parámetro poblacional.

El comportamiento de las primeras distribuciones de frecuencias es completamente errático pero poco a poco el valor de cada frecuencia alélica se estabiliza, sin embargo aún no se puede estimar ningún tamaño de muestra óptimo puesto que esta es solo una medida gráfica.

En la siguiente fase del algoritmo, se calcularon los promedios como indica la matriz número 2, también fueron calculadas las desviaciones estándar, dado que son los parámetros para calcular el Coeficiente de variación. A medida que crece la muestra el coeficiente de variación disminuye notablemente, acercándose al valor de 0.01, establecido como el punto en el cual se selecciona una muestra en la distribución de frecuencias d_i asociada a un tamaño de muestra n_i . La gráfica No. 2b muestra como los alelos en baja frecuencia generan mayores oscilaciones en los coeficientes de variación, por tanto fueron retirados para la determinación del tamaño de muestra.

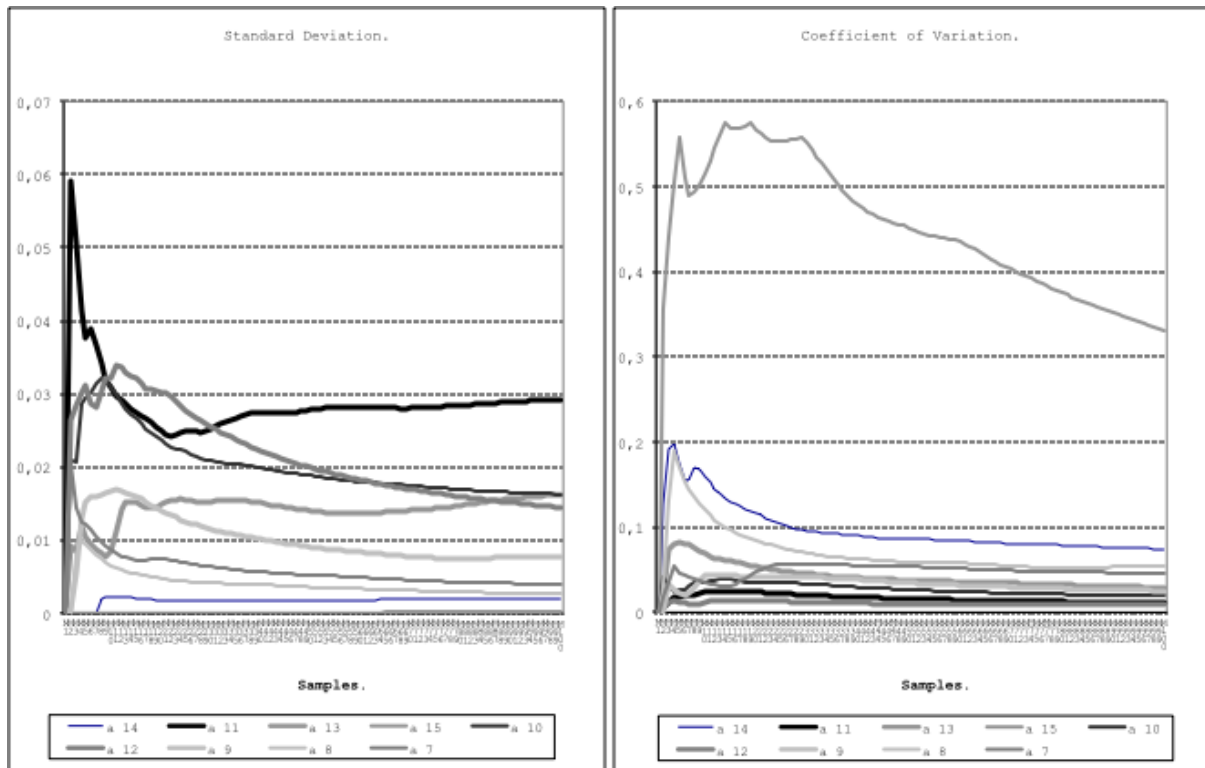


Figura No. 2 Las gráficas anteriores muestran el comportamiento del sistema microsátélite D5S818 en el experimento propuesto. A medida que se incrementan los muestreos, se logra una tendencia a disminuir considerablemente la desviación estándar, la gráfica de coeficientes de variación muestra la misma tendencia, sin embargo se da una separación entre alelos en baja frecuencia y alelos efectivos.

3.2. Tamaño de muestra establecido para el sistema completo.

Aplicando el algoritmo secuencial, se determinaron los siguientes tamaños de muestra para cada uno de los sistemas microsátélites estudiados.

STR'S	Category.	No. Alleles	Alleles found.	% alleles found.	No. Efective alleles.	% efective alleles.	Unique Alleles	Sample	No. Ind.
D5S818	Simple	10	9	90%	4	44%	1	24	308
D16S539	Simple	10	8	80%	5	63%	1	4	56
D13S317	Simple*	14	10	71%	6	60%	2	12	152
CSF1PO	Simple	15	10	67%	4	40%	2	22	250
TPOX	Simple	17	9	53%	3	33%	4	7	86
D8S1179	Compuesto	19	10	53%	5	50%	3	2	28
D7S820	Simple	22	11	50%	5	45%	3	16	213
D18S51	Simple	42	20	48%	8	40%	4	53	709
D3S1358	Compuesto	20	9	45%	4	44%	2	2	29
VWA	Compuesto	28	12	43%	4	33%	4	10	134
TH01	Simple*	21	7	33%	4	57%	0	17	208
D21S11	Complejo	70	23	33%	6	26%	5	48	651
FGA	Compuesto	98	21	21%	8	38%	7	26	339

Tabla No. 1: Cálculo de tamaño de muestra para los sistemas STR's del sistema CODIS. Además del tamaño de muestra se presenta el análisis porcentual de cada sistema en función del número de variantes alélicas reportadas. No. de alelos, es el número de alelos reportados a nivel mundial para el polimorfismo, Alelos encontrados se refiere a los alelos hallados en la muestra estudiada. No. Alelos efectivos es el número de alelos que corresponden a los alelos más representativos de la población según el procedimiento de Buttler para separar el polimorfismo (Buttler, 2007). Sample es el m_i en el cual el coeficiente de variación alcanza un valor inferior a 0.01.

El tamaño de muestra sugerido por la aplicación es de 709 personas, número elevado en comparación con estudios tradicionalmente desarrollados en poblaciones humanas que están alrededor de 200 muestras. En la población de Bogotá con diferentes orígenes y un activo flujo de migración de diferentes grupos étnicos, se encontró una amplia distribución de los polimorfismos STR's: Un promedio de 52,8% variantes alélicas por sistema sobre el número total reportado en el mundo. Sin embargo, la dispersión en este valor es alta (entre 21% a 90%), por tanto se evaluó mediante un coeficiente de correlación entre el número de variantes reportadas por sistema polimórfico y el número de variantes encontradas en la población, con un valor alto de correlación: 0,8878. La correlación entre el número de alelos encontrados en la población de estudio y el

tamaño de muestra reportado por este método igualmente presentó un valor de 0,8211. Es decir entre más variantes alélicas tiene un sistema, mayor es el tamaño de muestra a emplear. Este comportamiento se dio independientemente de la complejidad estructural de cada sistema microsatélite. Inicialmente se había pensado que los microsatélites simples, deberían arrojar tamaños menores, e ir incrementando según la estructura, esta tendencia no se vio en el estudio, luego la variable complejidad del microsatélite fue descartada de la interpretación del problema.

Si bien, aunque la muestra requerida es de 709 personas, la distribución completa de alelos efectivos en la población se encontró con un número de 230 personas. Por tanto la distribución de frecuencias alélicas para alelos efectivos puede ser similar en una muestra pequeña que en una grande. Este efecto nos indica que aunque los sistemas microsatélites son ampliamente polimórficos, en términos efectivos en la población se presentan unas pocas variantes. No importa si un polimorfismo es grande, importa es que tan representado está en la población.

Después de un tamaño de muestra mínimo (230 personas), que explica en promedio el 90% de los genotipos en la población, se da exclusivamente la aparición de alelos raros los cuales permanecen en baja frecuencia. El ampliar la muestra por encima del tamaño mínimo propio de los alelos efectivos tendría la ventaja de poder representar el polimorfismo completo del sistema microsatélite, lo cual es útil en estudios poblacionales y en los casos de identificación. Una muestra de aproximadamente 200 individuos, simplemente representaría la población efectiva, dejando la distribución de alelos raros menormente representada.

Adicionalmente al tamaño de muestra, se calcularon intervalos de confianza para cada variante alélica y también un intervalo general por marcador, el cual permite hacer una aproximación de la dispersión de una frecuencia alélica para variantes nuevas en la población de referencia. En el anexo No.1 se presenta la estimación de frecuencias alélicas por intervalos de confianza para la población de Bogotá. La siguiente tabla muestra el valor del intervalo de confianza general de cada sistema evaluado.

STR'S	Promedio	Min	Max
D5S818	0.0016	0.0001	0.0029
D16S539	0.0072	0.0009	0.0108
D13S317	0.0009	0.0001	0.0019
CSF1PO	0.0033	0.0005	0.0068
TPOX	0.0141	0.0035	0.0209
D8S1179	0.0011	0.0001	0.0029
D7S820	0.0020	0.0002	0.0040
D18S51	0.0108	0.0019	0.0199
D3S1358	0.0012	0.0001	0.0030
VWA	0.0018	0.0002	0.0039
TH01	0.0002	0.0000	0.0005
D21S11	0.0005	0.0000	0.0015
FGA	0.0002	0.0000	0.0007

Tabla No 2 Intervalos de confianza promedios de cada sistema microsatélite. En el anexo No. 1 se estimó el valor del intervalo de confianza de cada variante alélica reportada, sin embargo se realizó una estimación promedio por sistema microsatélite que es útil en cálculos generales sobre el sistema o como valor de referencia en el caso de la aparición de alelos nuevos.

4. Discusión.

Normalmente en cualquier análisis genético poblacional, el investigador realiza un conteo directo para la determinación de frecuencias alélicas sin considerar una medida de la variación del estimador, esto se debe al uso tradicional de aplicar modelos determinísticos en lugar de usar modelos probabilísticos. Esto es un problema en muestreos con n pequeño, y con criterios poco rigurosos de inclusión. El algoritmo anteriormente planteado, resuelve estos problemas. En primer término al calcular un valor de frecuencia alélica, asociado a un intervalo de confianza, se hace posible determinar la calidad del muestreo realizado en función de la dispersión, obtenida a partir de múltiples repeticiones del algoritmo secuencial.

Es también importante considerar después de cierto punto, en el cual se han alcanzado todos los alelos efectivos de un polimorfismo, el esfuerzo de muestreo que se realice dado que no se lograría un mayor nivel de resolución incrementando sustancialmente la muestra. Con un número mayor se asegura una mejor representatividad de alelos raros, pero no cambios importantes en la distribución de frecuencias.

Algunas ecuaciones permiten calcular intervalos de confianza para distribuciones de frecuencias alélicas, Gillespie, (1998), sin embargo este procedimiento matemático, presenta un argumento circular. Como puede verse en la Ecuación No. 5:

Ecuación No. 5:

$$I.C. = \bar{P} \pm 1.96 \sqrt{\frac{P \cdot (1 - P)}{n}}$$

Se requiere del n muestral que es el número de muestras obtenido por el investigador, pero no corresponde a ningún cálculo, es el valor obtenido. El procedimiento propuesto en este artículo, calcula un valor de intervalo de confianza, pero este valor se reporta cuando el tamaño de muestra y los valores de coeficientes de variación han llegado a un límite inferior. Por tanto el cálculo del tamaño de muestra y la obtención de los intervalos de confianza, son procesos simultáneos, si se da un incremento en el tamaño de muestra se realiza el recálculo de los intervalos de confianza. Intervalos de confianza demasiado grandes indicarían tamaños de muestra insuficientes o con criterios de inclusión poco claros. Calcular tamaños de muestra e intervalos de confianza de esta manera, permiten hacer un cambio conceptual, pasando de ecuaciones establecidas, donde primero se calcula un tamaño de muestra y luego un intervalo, a procedimientos simultáneos que requieren de un cuidadoso análisis de la muestra a medida que ésta es tomada y no tanto del seguimiento de una ecuación inicial para obtener un tamaño específico.

Normalmente en la práctica, se realizan cálculos muestrales *a priori* específicos en cada estudio, sin embargo lo que puede verse al utilizar una metodología *a posteriori*, es que se obtiene un valor del comportamiento de un polimorfismo en la población, no un valor para un estudio específico, corresponde mas a un estudio muestral sobre la variabilidad genética de la población,

convirtiéndose en un valor de referencia para otros estudios a realizar y no simplemente en el cálculo de tamaño de muestra.

El cálculo de frecuencias alélicas asociadas a intervalos de confianza implica también el cambio en la forma de estimar los demás estadísticos forenses y genético poblacionales, es decir los estadísticos como los índices y probabilidades de paternidad posiblemente puedan pasar de ser estimadores puntuales a estimadores por intervalos.

En muchas oportunidades al realizar un determinado análisis no sabemos si estamos cercanos a un tamaño óptimo, los cálculos de tamaño de muestra permiten hacer un acercamiento al mejor valor en relación con la variabilidad genética presente en la población. Normalmente se utilizan tamaños pequeños de muestra y se piensa que entre más muestras mejor, sin embargo, números excesivos de muestra terminan generando costos innecesarios y procesos computacionales más complejos de lo requerido.

La clasificación de alelos efectivos y raros arrojaría dos tipos diferentes de información, la correspondiente a los alelos que conforman en mayor medida la población de estudio, y que dan cuenta de la mayoría de genotipos encontrados en la población, y los alelos raros que dan mayor resolución sobre los procesos evolutivos seguidos en el tiempo. La presencia de variantes en baja frecuencia suelen ser más características de la población, especialmente originadas por procesos de migración y mutación. Por lo tanto sugerimos dar un peso diferencial en los análisis de similitud poblacional a los alelos efectivos y alelos en baja frecuencia, tal que las variantes en baja frecuencia podrían ser útiles en los análisis de ancestría y en casos de identificación y filiación.

5. Conclusión.

Este trabajo se ha realizado en base a las frecuencias alélicas, sin embargo el cálculo de intervalos de confianza puede hacerse extensible a todas las demás medidas que suelen calcularse sobre una población, no todos podrían obtenerse directamente por análisis de muestreo secuencial, pero al contar con intervalos de confianza para las frecuencias alélicas, dichos intervalos puede ser

utilizados para calcular la dispersión de los demás estimadores. En particular llama la atención los cálculos por intervalos de confianza de índices de paternidad, que hasta el momento son calculados como estimadores puntuales, sin embargo podrían estimarse como intervalos también con la finalidad de lograr una mejor interpretación de la relación biológica del padre en cuestión con la población de referencia.

Anteriormente se utilizaban límites para la interpretación de las probabilidades de paternidad como sucedía con los enunciados verbales de Hummel, estos fueron eliminados y reemplazados por valores puntuales, el uso de intervalos de confianza, podría permitir nuevamente el uso de enunciados ajustados a los nuevos niveles de resolución molecular, a los métodos de cálculo y la diversidad de los actuales marcadores. Esto podría facilitar la interpretación de autoridades y usuarios, logrando una mejor explicación en términos poblacionales.

Por otra parte, hemos probado el presente algoritmo con marcadores moleculares STR's, sin embargo el comportamiento y los tamaños de muestra reportados, podrían ser menores en marcadores moleculares SNP's, dado el nivel de polimorfismo.

La actual teoría evolutiva arroja valores de variabilidad genética de una magnitud tal que permiten la individualización de cada persona en el planeta, luego los cálculos de tamaño de muestra en estudios con múltiples marcadores genéticos pueden ser muy grandes, creemos que el uso de herramientas complementarias tales como los análisis demográficos y genealógicos pueden ser herramientas útiles para asegurar la calidad de los resultados de investigación, disminuyendo considerablemente los tamaños de muestra. Más que realizar un cálculo de tamaño de muestra, se necesita realizar un estudio muestral en cada población que además considere las características genealógicas de la población como criterios de inclusión en muestras. De esta forma se podría evaluar la distribución en el tiempo y el espacio de diferentes polimorfismos en la población debido a la construcción de mapas y planos factoriales que serían la base para estudios posteriores en genética forense.

Bibliografía.:

- Bautista L. (1998). Diseño de Muestreos Estadísticos. Bogotá: Universidad Nacional de Colombia. 160 p
- Campbell AW, Campbell PJ, Griffin WB, Burritt DJ, and Conner AJ. 2001. Prediction of Sample Size to Maintain Genetic Variation in Doubled-Haploid Populations Following Marker Selection. *Theoretical and Applied Genetics* 103:1028-1036.
- Chakravarty S. 2007. Sample Size Determination for Multinomial Population. Available from: [http://tx.liberal.ntu.edu.tw/~PurpleWoo/Literature/!Methodology/!Distribution_SampleSize/Sample Size Determination for Multinomial Population.pdf](http://tx.liberal.ntu.edu.tw/~PurpleWoo/Literature/!Methodology/!Distribution_SampleSize/Sample%20Size%20Determination%20for%20Multinomial%20Population.pdf)
- Chakraborty R. (1992). Sample Size Requirements for addressing the population Genetic issues of forensic use of DNA Typing. *Human Biology*. Vol 64, No.2.
- Evetts, i. W. Y b. S. Weir. (1998) *Interpreting DNA Evidence*. Sinauer, Sunderland, MA...173 p.
- ISFG 1992. International Society for Forensic Haemogenetics. DNA recommendations — 1992 report concerning recommendations of the DNA Commission of the International Society for Forensic Haemogenetics relating to the use of PCR-based polymorphisms. *International Journal of Legal Medicine*. Vol. 105, No. 1.
- ISFG. 2002. Annual Report Summary for Testing. American Association of Blood Banks. [Internet]:1-62. Available from: [http://www.gep-isfg.org/documentos/AABB 2004.pdf](http://www.gep-isfg.org/documentos/AABB%2004.pdf)
- ISFG. 2004. Annual Report Summary Report Summary for Testing. American Association of Blood Banks [Internet]:1-62. Available from: [http://www.gep-isfg.org/documentos/AABB 2004.pdf](http://www.gep-isfg.org/documentos/AABB%2004.pdf)
- Gillespie J. H. (1998). *Population Genetics. A concise guide*. The Johns Hopkins University Press. Baltimore. Pag 142
- Goldstein d, Schlotterer c. (1999). *Microsatellites*. Londres: Oxford University Press. 352 p.
- Hall P, and Peng L. 1999. On Prediction Intervals Based on Predictive Likelihood or Bootstrap Methods. *Biometrika* 86:871-880.

- Harding H. (1998). DNA Database Size. *Journal Forensic Science*. 15:345-349
- Hey J, and Nielsen R. 2004. Multilocus Methods for Estimating Population Sizes, Migration Rates and Divergence Time, with Applications to the Divergence of *Drosophila Pseudoobscura* and *D. Persimilis*. *Genetics* 167:747-760.
- Hohls T. 1998. Reliability of Confidence Interval Estimators under Various Nested Design Parental Sample Sizes. *Biometrical Journal* 40:85-98.
- Kalinowski ST. 2005. Do Polymorphic Loci Require Large Sample Sizes to Estimate Genetic Distances? *Heredity* 94:33-6.
- Medina R. (2006). Number of Individuals and molecular markers to use in genetic differentiation studies. *Molecular Ecology* Vol 6.
- Wang J, and Whitlock MC. 2003. Estimating Effective Population Size and Migration Rates from Genetic Samples over Space and Time. *Genetics* 163:429-446.

Capítulo No. 3: Análisis de la Estructura de la población de La Guajira: Una visión Genética Demográfica y Genealógica.

M. Rojas¹, A. Alonso¹, L. Eljach¹, V. Sarmiento¹, W. Usaquén¹.

Grupo de Genética de poblaciones e Identificación. Instituto de Genética. Universidad Nacional de Colombia.

Resumen

En este estudio se caracterizó la subestructura genética y la dinámica migracional de la población de La Guajira-Colombia desde un punto de vista genético, demográfico y genealógico. Se analizó una muestra de la población tanto residente como ancestral (Wayúu) que habita actualmente la región. Basados en esos criterios, se plantearon unas categorías genealógicas y un diseño de muestreo, que diera cuenta de la diversidad, subestructura, del efecto de la deriva genética y de la migración de la población. Ambos tipos de variables tanto demográficas como genéticas (marcadores moleculares STRs autosómicos) permitieron dilucidar una estructura genético-poblacional robusta y consistente con la dinámica migracional de la Guajira. Se encontró una subpoblación con ancestría Wayúu con altos niveles de heterocigosidad (H_o 0.728) que se diferencia de los demás grupos genealógicos (Θ 0.024 IC: 0.023 \pm 0.006) exceptuando los individuos con genealogía Wayúu-Guajiro. Para los microsatélites evaluados, no se encontró que esta fuera endogámica ($SmallF$ 0.017). Los datos demográficos mostraron a la población Guajira como joven, en crecimiento y expuesta a un moderado efecto de la deriva genética (N_{me} 11.903). Geográficamente, los Wayúu coexisten en algunos territorios con los demás grupos y tienen un moderado flujo génico con ellos, en su mayoría con individuos Wayúu-Guajiros. La dirección del flujo génico en su mayoría va en dirección externa a la población Wayúu. En síntesis, la cuidadosa selección de la muestra nos permitió obtener un modelo de subestructura y del patrón de migración, causado por procesos culturales, históricos y geográficos.

Palabras Clave

Wayúu, Subestructura, Migración, Deriva genética, Demografía, Genealogía.

1. Introducción.

La península de La Guajira se encuentra ubicada en las coordenadas 10° 23' y 12° 28' latitud norte, 71° 06' y 73° 39' longitud oeste. Es la región más septentrional de Sudamérica, entre el extremo nororiental de Colombia y el extremo noroccidental de Venezuela, tiene una superficie de 25.000 Km², en su división política pertenece en mayor parte al departamento de La Guajira (20.848 Km²) (Gobernación de la Guajira, 2011). La población colombiana de La Guajira según el DANE (Departamento Nacional de Estadística) es de 846.609 habitantes con una tasa de crecimiento de 1985 a 2005 del 87.95%, donde el 51,9% de los habitantes de la Guajira, viven en las Cabeceras municipales y el 48.1% en región rural. En general la densidad poblacional es 28.37 hab./km², por debajo de la media nacional de 40,7 hab. /km².

En este territorio existe una gran diversidad étnica, representada por pueblos nativo-americanos de las familias lingüísticas Arawak y Chibcha, una población residente afrodescendiente y colombianos de múltiples ancestrías. Por otro lado hay presencia de sirio-libaneses, especialmente en la ciudad de Maicao (CELADE, 2011). Los Wayúu (Arawak) son el pueblo indígena que representa el 44.9% de la población del departamento de la Guajira, siendo a la vez el pueblo indígena más grande de Colombia, constituido por más de 144.000 personas que representan el 20.5% de la población indígena nacional (DANE, 2005).

Ante este escenario, se partió de una primera hipótesis: La población Wayúu es una población diferenciada del resto de población residente por su distribución geográfica, composición cultural, y su historia de poblamiento. Primero que todo, la península está constituida geográficamente en el centro y norte por una planicie baja con un clima de estepa árida o semiárida. Esta es considerada la región más seca de Colombia, donde se han asentado los Wayúu por miles de años, estableciendo numerosos clanes o tribus. Por otra parte al sur de la península, se encuentra la Sierra Nevada de Santa Marta, la serranía del Perijá y la planicie aluvial de los ríos Ranchería y

Cesar, con un clima diferente, de selva húmeda de montaña, selva seca y sabana xerófila. En la Sierra Nevada se ubican comunidades indígenas ancestrales como los Kogi, Ika y Kankuamo (Gobernación de la Guajira, 2011).

La adaptación de los Wayuu a las condiciones inhóspitas del desierto, les sirvió como refugio de los europeos durante la colonización, en consecuencia no sufrieron invasiones sino hasta después de la independencia de Colombia y Venezuela, manteniendo hasta hace poco una amplia autonomía extralegal. De esta manera, parte de la comunidad Wayuu se ha mantenido relativamente aislada en rancherías lejanas dispuestas irregularmente dentro del territorio (Vásquez S & Correa H, 2000). De esta manera la densidad poblacional de los Wayúu es mayor en Nazaret, Uribía, la Serranía de Jala'ala y las sabanas de Wopu'müin, zonas de la alta Guajira, sin embargo también habitan en las ciudades principales (Programa Presidencial de Derechos Humanos, 2011).

Según el Antropólogo José R. Oliver (1991), el origen de los Wayúu según un criterio lingüístico es Arawak, de la región Amazónica. Se presume que la población migró desde el Río Negro hacia la costa occidental de Venezuela y La Guajira hace 4000 a 5000 años, en esa trayectoria, desde hace 3000 a 2000 años hubo una separación entre Lokono y Wayúu y desde hace 1500 a 1000 años hubo una separación entre Wayúu y Paraujano, siendo posible que estas divergencias ocurrieran en la región que conecta el Orinoco y los llanos con la Amazonia central, por último, avanzaron desde este punto hacia el norte por la costa occidental de Venezuela, adicionalmente, a principios del siglo XVI en la península cohabitaban con otros grupos indígenas de diferente origen lingüístico (Caribes y Chibchas), de múltiples orígenes y con diferentes historias evolutivas. Desde esta época hasta el Siglo XVI y el presente, los distintos pueblos convivieron de manera pacífica compartiendo su territorio, sin embargo muchos clanes desaparecieron unos por fusión y los demás por causas no evidenciadas (Oliver JR, 1991).

La configuración cultural y social del grupo indígena wayuu, ha estado influenciada por su adaptación al ecosistema, es así como los Wayúu suelen no vivir en asentamientos estables, dado que con frecuencia se trasladan dependiendo del régimen de lluvias (Arango, 2004; Ardila G,

1992). Desde una perspectiva social y política, en los Wayúu no existe un poder central, se organizan en clanes o castas matriarcales con matrilocidad, que comparten una condición social y un pasado común, es así como las relaciones de parentesco, el e'irukuu (correspondiente al apellido o casta) se hereda por línea materna al igual que la identidad Wayúu (Arango R, 2004). En Colombia hay 22 castas entre las que se destacan por su tamaño los Epiayú 20.8%, Uriana o Uliana 17.1% Ipuana 16,2%, en particular, el matrimonio suele practicarse entre miembros de diferentes castas, se practica la monogamia y la poligamia en una versión poligínica, donde cada esposa vive con su madre y hermanas. La identidad Wayúu se hereda por línea materna, por lo que algunos individuos de ancestría múltiple (madre Wayuú-padre Arijuna) se consideran Wayúu, mientras que el caso contrario no; por otro lado, los primos maternos no pueden casarse por ser del mismo linaje, pero sí es posible entre primos paternos (Arango R, 2004).

En retrospectiva, la segunda hipótesis se fundamenta en la residencia en la región de migrantes de múltiples ancestrías, principalmente provenientes de la costa Caribe colombiana y algunos del resto del país que se concentran en los asentamientos urbanos principales; aquí, la dinámica migracional regional podría caracterizarse por inmigración, en su mayoría de corta distancia. Adicionalmente en la zona viven los *Guajiros*, personas que se consideran de la región pero que no se identifican como de ancestría Wayúu. Es así como en las zonas urbanas suelen coexistir tanto migrantes, guajiros e individuos Wayúu, compartiendo el territorio y algunas actividades tanto económicas como sociales.

Uno de los supuestos iniciales es el de un moderado flujo bidireccional entre grupos generando diferentes modelos de ancestría. Un ejemplo claro es el matrimonio *Wayúu-Arijuna* (persona no Wayúu), que se da principalmente en enclaves urbanos que sirven como concentración tanto de migrantes e indígenas que están en búsqueda de una fuente de trabajo. Es así como el estudio de variables como el conocimiento del lenguaje (Wayuunaiki) y la pertenencia a una casta, permitirían asociar individuos con ancestría Wayúu a menores niveles de introgresión, dado que este conocimiento suele perderse con la aculturación. Además, no se tiene certeza de que las poblaciones aisladas geográficamente, difieran considerablemente de las que se encuentran en

contacto permanente con poblaciones externas.

Con el fin de obtener este nivel de resolución en el estudio de la subestructura y migración, se ha acudido a las características demográficas, genealógicas y socioculturales propias de las poblaciones, debido a que el estudio de la estructura de edad y sexo, la distribución espacio-temporal, los caracteres sociales, económicos y culturales permiten tener un conocimiento del pasado, presente y futuro cultural, social y biológico de la población de estudio (Albeza MV. et al, 2004), además permiten evaluar el efecto de la deriva genética en la diferenciación poblacional. Es así como la evaluación de elementos como el tamaño efectivo poblacional, el aislamiento y la panmixia junto a los datos genéticos es importante porque influyen directamente en la estructura genética de las poblaciones (Albeza MV. et al, 2004). No obstante, tradicionalmente en los modelos de genética dep se han ignorado estos factores, que permiten obtener hipótesis más acertadas de la dinámica poblacional sobre las fuerzas de cambio evolutivo. (Albeza MV. et al, 2004). Es por esta razón que el objetivo de este estudio fue caracterizar la subestructura genética y la dinámica migracional de la población de La Guajira-Colombia desde un punto de vista genético, demográfico y genealógico; implementando una metodología que nos permitió encontrar una población conservada de ascendencia Wayúu que se diferencia del resto de la muestra, pero que converge en algunas zonas geográficas con la población residente con una dinámica migracional bidireccional activa.

2. Metodología

Se realizó un muestreo dirigido para la población de la Guajira y la población Wayúu con un tamaño de muestra de 290 individuos. El muestreo se llevó a cabo en los hospitales de los

municipios de Riohacha, Maicao y Uribía (Figura 1) y en algunas rancherías de los tres municipios. Previo consentimiento informado, se tomaron muestras de saliva ó 10 ml de sangre periférica, cada una fue preservada mediante una tarjeta FTA (Whatman BioScience, USA). El ADN fue extraído tomando un fragmento de la matriz de la tarjeta.



Figura No. 1 Mapa de la Guajira mostrando la localización de las tomas de muestra.

Con el fin de entender la interacción entre la dinámica de estos loci y la organización social, se planteó una encuesta que permitiera indagar sobre la información demográfica de la población y su pertenencia étnica (comunidad indígena Wayúu), al mismo tiempo, se reconstruyó la historia genealógica de cada individuo indagando por su lugar de nacimiento, el de su parentela y la pertenencia a un grupo étnico en las previas tres generaciones, esta información permitió hacer una distinción de la ancestría de cada grupo. Las variables tomadas en la encuesta fueron: género, edad, lugar de nacimiento, número de hijos, edad a la que se tuvo el primer hijo, edad del cónyuge, lugar de nacimiento del cónyuge y la pertenencia étnica del cónyuge (por autodeterminación). Se preguntó el lugar de nacimiento y la pertenencia étnica de tres generaciones anteriores, tanto por línea materna como paterna, por último, si el individuo pertenecía a la etnia Wayúu, se indagó sobre su casta, ranchería y si hablaba o no Wayuunaki.

Aunque el criterio de inclusión al estudio fueron personas residentes del departamento, con la información anterior también se pudo clasificar a los individuos de acuerdo a una reconstrucción genealógica *a priori*; es así como los lugares de nacimiento y los nombres de las castas se usaron como evidencia para la clasificación genealógica: Ubicamos como *Wayúu* a los individuos con ancestría proveniente de la etnia *Wayúu* tanto por linajes materno como paterno (Figura 2), *Guajiro* a los individuos con ancestría oriunda de este departamento que no pertenece a esta etnia, *Wayúu-Guajiro* a los individuos con ancestría por un lado *Wayúu* y por otra del departamento de La Guajira no perteneciente a la etnia, los individuos con una ancestría diferente al departamento de La Guajira se denominaron migrantes: *Migrantes 1* fueron los individuos con una ancestría tanto por linaje materno y paterno de un mismo departamento diferente a La Guajira, mientras que *Migrantes 2* fueron los individuos con linajes provenientes de departamentos distintos entre sí y por supuesto que no fueran de La Guajira. Por último los individuos de *ancestría múltiple* se definieron como los individuos con ancestría por un lado del departamento de La Guajira y por otro lado de un migrante. Esta clasificación se basó en la estructura genealógica de los individuos y refleja la dinámica del flujo génico de los diferentes grupos en la población, permitiendo obtener una muestra *Wayúu* según criterios genealógicos.

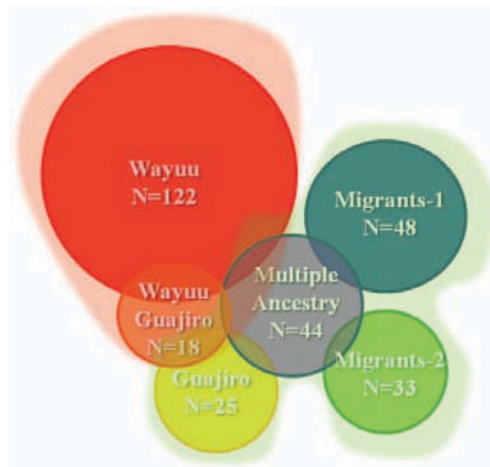


Figura No. 2 Grafica con las agrupaciones genealógicas y los tamaños de muestra

2.1 Análisis Molecular

Se amplificaron 15 loci STR autosómicos por PCR utilizando el kit comercial AmpFISTR® Identifier™ (Applied Biosystems, Warrington, U.K.), un termociclador programable 9700 (Applied Biosystems, Foster City, California). Las amplificaciones fueron realizadas utilizando los siguientes parámetros por ciclo: denaturación inicial (95°C por 11 min) seguido por 28 ciclos de denaturación de las hebras (94°C por 1 min.), anillamiento de los primers (59°C por 1 min), extensión del DNA (72°C por 1 min), y extensión final (60°C por 60 min). Las alícuotas del ADN fueron amplificadas y marcadas con fluorescencia y mezcladas con formamida (Applied Biosystems, Foster City, California) y LIZ 500 (Applied Biosystems, Warrington, U.K.) y genotipificadas en un analizador genético ABI Prism 310 (Applied Biosystems, Stafford, Texas) utilizando el programa Genescan 3.1.2. Las asignaciones alélicas fueron hechas mediante la comparación de los fragmentos amplificados con escalera alélicas (Applied Biosystems, Warrington, U.K.) que contiene el kit y utilizando el programa Genotyper 2.5.2 (Applied Biosystems, Foster City, California).

2.2 Análisis Demográfico de la Muestra

A partir de la información demográfica se hizo un análisis de la distribución de edad y sexo de la muestra. Se estimó la edad promedio de la población total, masculina y femenina y de la fracción reproductora. Con base a las edades se calculó el tiempo generacional, el tamaño reproductivo de la población (N_r) (varones mayores de 18 años y mujeres entre 18-49 años con al menos un descendiente vivo entre 0-30 años), el tamaño efectivo poblacional (N_e) y tasa de migración efectiva (m_e) estimada a partir de la aproximación de Malécot (Magalhaës y Arce-Gómez 1987a). A partir de estos últimos parámetros se estimó el coeficiente de aislamiento reproductivo ($N_e m_e$) y el coeficiente de endogamia (F). Finalmente se realizó un análisis de la distancia geográfica entre parejas y la dirección del flujo génico estudiando las frecuencias de los matrimonios entre las diferentes clasificaciones genealógicas.

2.3 Análisis Genético de las Agrupaciones Genealógicas

Estas clasificaciones genealógicas *a priori* fueron analizadas mediante las siguientes metodologías. En primer lugar, en el programa Genepop 4.09 (Raymond M. & Rousset F, 1995; Rousset F. 2008) se calculó el promedio de la heterocigosidad (Nei, 1987) por locus usando la corrección de Levene. El test exacto del Equilibrio de Hardy Weinberg fue calculado en este mismo programa con los siguientes parámetros 10.000 periodos o ciclos de demorización, 1000 *batches* y 100 iteraciones. En el programa Fstat 2.9.3.2 (Goudet J, 1995) fueron calculados los estadísticos-F (Capf, Theta, Smallf) (Weir & Cockerham, 1984) por locus y un Fis (Smallf) dentro de los grupos genealógicos. Los intervalos de confianza de los estadísticos-F fueron estimados basados en técnicas de remuestreo jackknife y bootstrap al 95% and 99% sobre los loci. Se estimó el estadístico Rst (Slatkin, 1995).

Por otro lado, en Arlequin 3.5.1.2 (Schneider S. et al, 2000) se estimaron las comparaciones por pares de Fst (theta) y el Rst de las muestras con 10.000 repeticiones y un nivel de significancia de 0.05. Se calculó también, el análisis molecular de varianza (AMOVA) con Fst usando 99999 repeticiones.

Se calculó la distancia de Cavalli-Sforza and Edwards (1967) (Dc) con el método de agrupamiento Neighbor-Joining para dilucidar las diferencias entre las seis agrupaciones genealógicas. Esto mediante el programa Populations 1.2.30 (Langella O, 1999) y con un bootstrapping de 1000 sobre los loci. Por otro lado, para el análisis de los individuos en función de los alelos de los sistemas microsatélites evaluados se acudió a técnicas de análisis multivariado, específicamente el análisis factorial de correspondencias (AFC) en Genetix 4.05 (Belkhir K, 2001).

Se realizó un análisis de asignación de individuos a poblaciones en el programa Structure 2.3.3 (Pritchard et al, 2000) asumiendo un modelo de mezcla y de frecuencias alélicas correlacionadas.

Con un periodo de quemado de 100.000 y 100.000 MCMC repeticiones sobre el periodo de quemado. Al analizar los resultados de Structure se evidenció un $K=2$ consistentes con la agrupación genealógica Wayúu y un grupo externo. Después de esto, basados en la inferencia de ancestría de los individuos con un $K=2$ se calculó el nivel de introgresión en el grupo externo a la población Wayúu. Se utilizó como un estimativo del nivel de introgresión, la diferencia acumulada por locus y por sistema, y el promedio del valor absoluto del delta entre las distribuciones de frecuencias alélicas para todos los sistemas de las frecuencias alélicas entre las dos subpoblaciones.

2.4 Análisis de las Agrupaciones Genéticas

Teniendo en cuenta los valores de Q del análisis de Structure para un $K=2$, las clasificaciones genealógicas y la pertenencia a las castas Wayúu, se realizó un análisis de correspondencias múltiples en SPAD V. 5.6. Adicionalmente se usó el software Bayesass 1.2 (Wilson and Rannala 2003) para estimar la tasa de migración reciente entre los Wayúu y el grupo externo. Se hizo un análisis descriptivo de las subpoblaciones evidenciadas en la muestra (Wayúu-Grupo Externo), calculando heterocigosidad, equilibrio de Hardy Weinberg y el número de migrantes (Barton & Slatkin, 1986) en Genepop 4.0.1. Así mismo, se calcularon los estadísticos-F (F_{ST} , F_{IS} , F_{IT}) (Weir & Cockerham, 1984) locus por locus, el F_{ST} (Fis) para las dos subpoblaciones. Los intervalos de confianza de los estadísticos-F basados en las técnicas de remuestreo jackknifing y bootstrapping al 95% y 99% sobre los loci y los estadísticos-R (Slatkin, 1995).

3. Resultados

3.1 Análisis Demográfico de la Muestra

En la muestra se obtuvieron 39% hombres y 61% mujeres. Divididos en seis agrupaciones genealógicas: Wayúu 42.1% (122 individuos), Migrantes-1 16.6% (48), Ancestría Múltiple 15.2% (44), Migrantes-2 11.3% (33) Guajiros 8.6% (25) y Wayúu-Guajiros 6.2% (18). En la pirámide poblacional (Figura 3) se muestra las proporciones de sexo de los grupos de edades calculadas como porcentaje sobre la muestra, mostrando una estructura de una población joven y en

crecimiento, dado que los grupos etarios de la base están en mayor frecuencia. La edad promedio de la muestra fue 39.21 años, para los hombres 42.5 años y para las mujeres 36.07. La población entre 18-29 años fue 40.2% mientras que la población mayor de 45 años 33.56%. El tamaño reproductivo (Nr) fue de 173 individuos con una edad media de 45.93 años. Sobre la estructura de edad y de sexo se identificó que la muestra estaba conformada principalmente por personas jóvenes en edad reproductiva que aportan al pool genético de la población. El tiempo generacional de la muestra fue 19.55 +/- 4.8 años, calculado con base a la edad media a la cual las mujeres de la muestra tuvieron su primer hijo. Según la ENDS (Encuesta Nacional de Salud) (Profamilia, 2007) la edad en que las mujeres de la Guajira tienen su primer hijo es alrededor de 21 años y la tasa de fecundidad en la muestra es de 2.87 hijos por mujer.

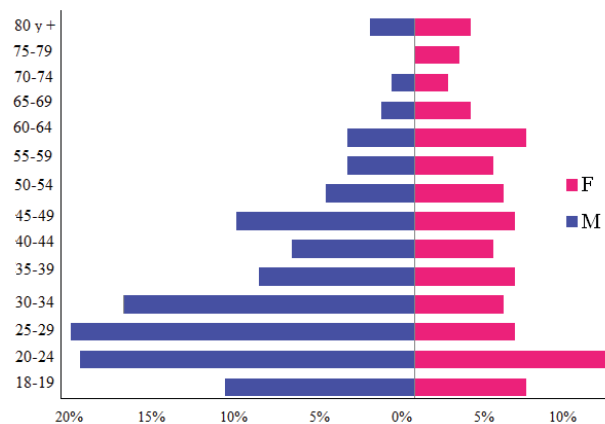


Figura No. 3 Pirámide de la muestra con la estructura de sexo y de edad. Este análisis fue calculado usando el porcentaje sobre la muestra.

Con base a las variables demográficas se evaluó algunos parámetros que involucran el efecto de deriva genética en la población. Encontrando un tamaño reproductivo de la población (Nr) de 155 individuos (53.4%) y en la población Wayúu 56 individuos (45,9%) (Tabla 2). Analizando el Nr

en los diferentes grupos de edades (generaciones de 19 años) se encontró la mayoría de población reproductiva entre 20 y 38 años (50%) para ambos grupos. Según el Nr de la población se estimó el tamaño efectivo en (N_e) 46.129 (13.5%) donde todos los adultos con descendencia, tienen igual probabilidad de ser los padres de cualquier descendiente individual. El N_e de la población total fue 46.129 (13,5%) y de los Wayúu 16.475 (13,5%). La tasa de migración efectiva es un estimativo de contribución de los migrantes al pool genético de la población calculado sobre la muestra. Para estimar este valor se establecieron como referencia a los grupos nativos (Wayúu, Wayúu-Guajiro y Guajiro) y como migrantes a larga distancia a Migrantes-1 y Migrantes-2 y a la agrupación de ancestría múltiple como migrantes a corta distancia. El 61.3% de los migrantes fueron a larga distancia, mientras que el 38.6 fueron a corta distancia. La tasa de migración (m_e) fue 0,258. Con el fin de medir los efectos de la deriva a partir del tamaño efectivo poblacional se calculó el coeficiente de aislamiento reproductivo (N_{em_e}) (Wright en 1938). El coeficiente de aislamiento reproductivo fue 11.9 lo que implica un efecto de deriva pequeño (Albeza MV. et al, 2004). Finalmente se hizo un estimativo del coeficiente de endogamia (F_{is}) de 0.0129 para la población total.

Parental Generation	Reproductive Individuals No.		(%) Percentage of Reproductive Size	
	Total	Wayuu	Total	Wayuu
First Generation (18-19 years old)	7	4	4.5161	7.1429 *
Second Generation (20-38)	79	28	50.9677	50 *
Third Generation (39-57)	48	14	30.9677	25 *
Fourth Generation (58-76)	18	7	11.6129	12.5 *
Fiveth Generation (77-95)	3	3	1.9355	5.3571 *
Reproductive Population Nr	155	56	53.4483	45.9016
k	4.2194	5.1964		
σ^2k	14.562	27.2516		
Effective Population Ne	46.129	16.475	15.906	13.504 **
Sample N	290	122		
Tasa de Inmigración (me)	0.258		0.5484	Migrants No./Nr
Proportion long-distance migrants (l)	0.1724		0.3226	l No./Nr
Proportion short-distance migrants (s)	0.1069		0.3500	s No./Nr
Isolation index (Neme)	11.9033			
Inbreeding coefficient (F)	0.0129			

Tabla No. 1 Tabla de Generaciones, tamaño reproductivo y aislamiento reproductivo. Tiempo Generacional Aproximado: 19 años. * Porcentaje sobre el tamaño reproductivo Nr. ** Porcentaje sobre el total de la población. Tamaño efectivo de la población $N_e = (2Nr-2) / (k-1) + (\sigma^2k / k)$ donde Nr = Hombres > 15 años y mujeres entre 15 y 49 ascendiente vivo entre 0 y 30 años, k = promedio de número de hijos por familia, σ^2k = varianza de el número de hijos por familia. (me) es la tasa de migración efectiva $(me) = [l(l + 2s)] / 2$. (Arce-Gómez Magalhães and 1987). (l) = Proporción de migrantes a larga distancia y (s) = Proporción de migrantes a corta distancia $F = (1-m)^2 / 2N - (2N-1)(1-m)^2$.

Se estudió la dirección del flujo génico entre las parejas, teniendo en cuenta la clasificación genealógica de la persona muestreada y la autodeterminación del conyugue (Tabla 2). Se encontró que el 80% de los Wayúu establecen grupos familiares con individuos Wayúu, sin embargo a una proporción de parejas con Wayúu-Guajiros, Guajiros y hasta con mestizos de ancestría indefinida (múltiples orígenes no caracterizados por criterios genealógicos). A pesar de que definimos varios grupos genealógicos se puede ver cierto nivel de flujo entre ellos. Este flujo entre grupos hace consistente la definición de grupos intermedios (Wayúu-Guajiros y Ancestría Multiple).

En el análisis de las distancias geográficas entre parejas (Figura 4) se encontró que los Wayúu, Guajiros, Wayúu-Guajiros y los de Ancestría Múltiple forman pareja en general más cerca que los grupos migrantes. Los migrantes tienen una dispersión diferencial, indicando que los otros grupos se encuentran en una área geográfica más restringida dentro de la Guajira y coexisten en una misma región geográfica. Este análisis es importante porque da una idea del patrón de migración de los individuos. Los grupos migrantes divergen de los demás en la extensión del radio de migración al cual se puede encontrar su pool génico.

<u>Genealogical Classification</u> / <u>Self-Determination</u>	<u>No. couples</u>	<u>Percentage</u>
Wayuu Women x Wayuu Men	21	40
Wayuu Men x Wayuu Women	21	40
Wayuu Women x I.A Men	9	17
Wayuu Men x I.A Women	1	2
Guajira Women x Wayuu Men	2	20
Guajira Women x I.A Men	5	50
Guajira Women x I.A Women	1	10
Guajira Women x Afrocolombian Men	1	10
Guajira Men x Afrocolombian Women	1	10
Wayuu-Guajira Women x Wayuu Men	2	33
Wayuu-Guajira Men x Wayuu Women	1	17
Wayuu-Guajira Women x I.A Men	3	50

Tabla No 2 *Tabla de Flujo Génico entre Parejas: Clasificación Genealógica únicamente para (Wayuu, Wayuu-Guajiro, Guajiro) de la persona muestreada con respecto a la autodeterminación en grupos étnicos de la pareja del encuestado I.A: Ancestría Indefinida.*

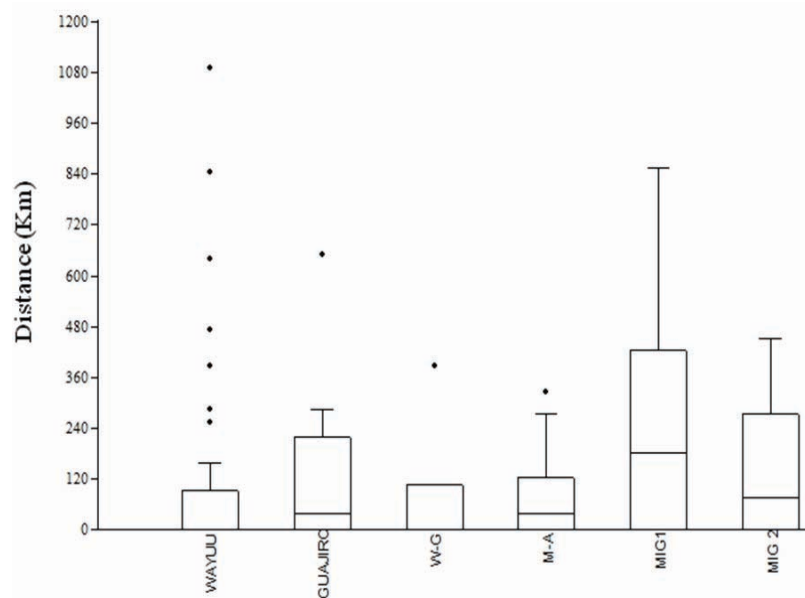


Figura No.4 Distancias Geográficas por parejas. W-G: Wayuu-Guajiro M.A: Ancestría Multiple.

3.2 Análisis Genético de las Agrupaciones Genealógicas

Para la totalidad de la población se calcularon los estadísticos descriptivos de la población (Tabla 3). La heterocigosidad observada (H_o) total fue 0.773 y la heterocigosidad esperada (H_e) fue 0.780. Se calculó el H_o locus por locus tanto para la población total, como para cada agrupación genealógica. La menor heterocigosidad fue para el grupo Wayúu 0.727. La población Wayúu resultó ser una población con alta diversidad genética, sustentada por la heterocigosidad tanto por sistema como para el total.

Al evaluar el EHW dentro de los grupos genealógicos se encontraron todos en equilibrio: Wayúu con un p-valor de 0.3082, Wayúu-Guajiro 0.8780, Guajiro 0.3051, Ancestría Múltiple 0.8293, Migrantes-1 0.2754 y Migrantes-2 0.066, como se mostró anteriormente la heterocigosidad para estos grupos es alta. Al evaluar el EHW por locus se encontraron la mayoría de sistemas en equilibrio a excepción de los sistemas D8S1179 y el TH01. Como consecuencia se encontró un

desequilibrio global en la población total altamente significativo.

Para la población general se encontró un $Capf$ de 0.026 con un intervalo de confianza de 0.033 ± 0.01 , un $Theta$ de 0.016 IC 0.021 ± 0.005 y un $Smallf$ de 0.011 IC 0.012 ± 0.008 (Tabla 3). Se obtuvo un Rst de 0.0167. Dentro de las agrupaciones se calculó el $Smallf$ para cada locus y global (Tabla 3). Basado en este resultado, se realizó un análisis de AMOVA con Fst (Tabla 4) evaluando tres hipótesis de subestructura: las seis agrupaciones en conjunto, Wayúu y el Grupo Externo y por último Wayúu y Wayúu-Guajiro con el grupo externo. Al evaluar la población general la mayoría de la varianza se encontró entre los individuos, con un Fit de 0.0262 y un p-valor de 0.0037 ± 0.0002 . Un Fst de 0.0105 significativo. Al evaluar la población Wayúu con respecto a los grupos externos, la mayor variación igualmente fue entre los individuos con un Fit de 0.00328. Aun así, se encontró un Fct de 0.01876 con un p-valor de 0.1659 ± 0.0012 y un Fsc de 0.0037 p-valor 0.0167 ± 0.0004 . Del mismo modo, en la agrupación Wayúu, Wayúu-Guajiro y en el grupo externo se encontró un Fct de 0.0193 p-valor 0.0671 ± 0.0008 y un Fsc de 0.003 p-valor de 0.0231 ± 0.0005 .

Se encontró que la agrupación (Wayúu-Wayúu-Guajiro) presentó el Fct mayor con un p-valor significativo, confirmando la hipótesis de 2 grupos.

LocName	All (N=290)					Wayuu (N=122)		Guajiro (N=25)		W-G (N=18)		M. A (N=44)		Mig1(N=48)		Mig2 (N=33)	
	Ho	Theta	Smallf	Capf	Rst	Ho	Smallf	Ho	Smallf	Ho	Smallf	Ho	Smallf	Ho	Smallf	Ho	Smallf
CSF1P0	0.684	0.004	0.071	0.075	0.023	0.699	0.12	0.788	0.187	0.734	0.167	0.759	-0.018	0.733	0.033	0.706	-0.073
D13S31	0.772	0.014	-0.004	0.01	0.008	0.772	-0.073	0.808	-0.089	0.757	0.192	0.822	0.032	0.801	0.09	0.804	0.02
D16S53	0.768	0.006	0.022	0.029	0.008	0.734	0.05	0.767	-0.043	0.792	-0.052	0.776	-0.113	0.785	0.045	0.794	0.16
D18S51	0.869	0.001	0.029	0.029	0.003	0.878	0.01	0.906	-0.06	0.894	-0.057	0.882	0.047	0.88	0.077	0.893	0.117
D19S43	0.87	0.009	-0.016	-0.007	0.006	0.832	0.045	0.844	0.005	0.801	-0.18	0.818	-0.084	0.866	-0.083	0.831	0.015
D21S11	0.82	0.015	0.003	0.019	0.003	0.784	-0.024	0.862	-0.021	0.851	-0.044	0.816	-0.002	0.817	0.057	0.823	0.079
D2S133	0.876	0.001	-0.013	-0.012	0.008	0.846	-0.008	0.889	0.1	0.853	-0.042	0.858	-0.086	0.88	0.006	0.877	-0.037
D3S135	0.727	0.045	-0.025	0.021	0.059	0.55	0.046	0.77	0.065	0.632	-0.054	0.71	-0.024	0.788	-0.189	0.772	-0.021
D5S818	0.702	0.036	0.033	0.068	0.081	0.661	0.069	0.71	0.099	0.665	0.081	0.784	-0.073	0.754	0.089	0.753	-0.086
D7S820	0.784	0.018	-0.017	0.002	0.046	0.685	0.019	0.794	0.043	0.76	-0.024	0.802	-0.02	0.723	0.02	0.801	-0.211
D8S117	0.781	0.003	0.012	0.015	0.003	0.817	-0.044	0.803	0.054	0.822	-0.014	0.812	0.132	0.799	0.062	0.765	-0.03
FGA	0.834	0.021	0.001	0.022	-0.006	0.818	-0.052	0.853	0.25	0.863	0.034	0.885	0.024	0.883	-0.015	0.889	-0.022
TH01	0.708	0.053	0.075	0.124	-0.008	0.599	0.165	0.784	-0.071	0.675	0.012	0.774	-0.058	0.783	0.149	0.783	0.033
TPOX	0.682	0.008	-0.017	-0.01	0.003	0.635	-0.007	0.737	0.023	0.587	-0.136	0.678	-0.14	0.683	0.085	0.655	-0.018
vWA	0.716	0.013	0.009	0.022	0.013	0.605	-0.003	0.792	-0.011	0.735	0.093	0.757	-0.081	0.742	0.13	0.731	-0.037
Overall	0.773	0.011	0.016	0.026	0.017	0.728	0.017	0.807	0.035	0.761	-0.002	0.796	-0.029	0.7945	0.035	0.792	-0.005
CI(Jk)	0.021+/-0.005		0.0120+/-0.080		0.0330+/-0.01												
CI(Bs 95%)	0.010+/-0.031		-0.002+/-0.028		0.016+/-0.054												
CI(Bs 99%)	0.010+/-0.035		-0.005+/-0.033		0.012+/-0.061												

Tabla No. 3 Diversidad y subestructura para las agrupaciones genealógicas: Se reporta la heterocigosidad observada (Ho) y el Smallf (Fis) (Weir & Cockerham, 1984) para los seis grupos genealógicos por sistema y global. Para la población total se reporta la Ho, los estadísticos-F y el Rst por locus y global. Para los estadísticos-F totales se calculó el Intervalo de confianza (CI) por Jackknifing (Jk), Bootstapping (Bs) al 95% y Bootstrapping al 99%.

Source of variation	Total			(Wayuú/Remaining)			(Wayuú+Wayuú-Guajiro/Remaining)		
	% of variation	Stastics-F	P-valor	% of variation	Stastics-F	P-valor	% of variation	Stastics-F	P-valor
Among groups	–	–	–	1.88	Fct: 0.01876	0.16590+ -0.00126	1.93	Fct: 0.0193	0.0671+ -0.0008
Among populations within groups	1.59	Fst: 0.0105	0.0000+-0.0000	0.36	Fsc:0.00368	0.01671+-0.00040	0.29	Fsc:0.0030	0.0231+-0.0005
Among individuals within populatic	1.04	Fis: 0.0159	0.1067+ -0.0010	1.03	Fis:0.01052	0.10726+ -0.00097	1.03	Fis:0.0105	0.1079+ -0.0010
Within individuals	97.37	Fit: 0.0262	0.0037+-0.0002	96.73	Fit: 0.00328	0.00326+-0.00018	96.75	Fit: 0.0325	0.0036+-0.0020

Tabla No. 4 Análisis Molecular de Varianza (AMOVA). Se calculan tres hipótesis de agrupamiento entre las poblaciones. Total: Todas en un solo grupo, Wayuu-Remaining: Wayuu en un grupo y las demás en otro y Wayuu+Wayuu-Guajiro/Remaining: Wayuu y Wayuu-Guajiro en un grupo y todo lo demás en otro grupo.

	Wayuu		Guajiro		Wayuu-Guajiro		Multiple-Ancestry		Migrants 1	
	Fst	P-valor	Fst	P-valor	Fst	P-valor	Fst	P-valor	Fst	P-valor
Wayuu	–	–								
Guajiro	0.0342	0.0000+-0.0000	–	–						
W-G	0.0050	0.0991+ -0.0370	0.0083	0.0630+ -0.0237	–	–				
M.A	0.0243	0.0000+-0.0000	0.0053	0.0540+ -0.0242	0.0057	0.0631+ -0.0305	–	–		
Migrants 1	0.0279	0.0000+-0.0000	0.0032	0.1802+ -0.0383	0.0123	0.0000+-0.0000	0.0017	0.1712+ -0.0212	–	–
Migrants 2	0.0201	0.0000+-0.0000	0.0028	0.1441+ -0.0309	0.0003	0.4595+ -0.0417	0.0009	0.2883+ -0.0297	0.0012	0.2883+ -0.0227

Tabla No. 5 Matriz de comparaciones de Fst por pares de poblaciones

Se hizo una comparación por pares de poblaciones usando los F_{st} . Se observó una diferencia significativa entre Wayúu y los demás grupos, exceptuando el grupo Wayúu-Guajiro, confirmando la clasificación genealógica. Además, Wayúu-Guajiro mostró ser significativamente diferente al grupo Migrantes-1.

Al calcular las distancias genéticas se obtuvieron los mismos resultados. Usando la distancia de Cavalli-Sforza and Edwards (1967) (Figura 4) se agrupó a Wayúu con Wayúu-Guajiro. Este grupo es congruente con la naturaleza de los grupos de ancestría compartida. Además, se clasificó a Guajiro, Ancestría Múltiple, Migrantes-1 y Migrantes-2 en un mismo agrupamiento, mostrando que no hay diferencias significativas entre estos grupos, por lo que las relaciones entre ellos no son claras.

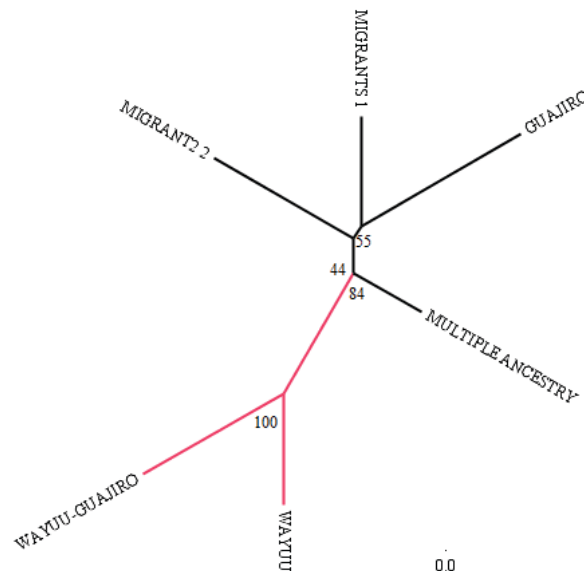
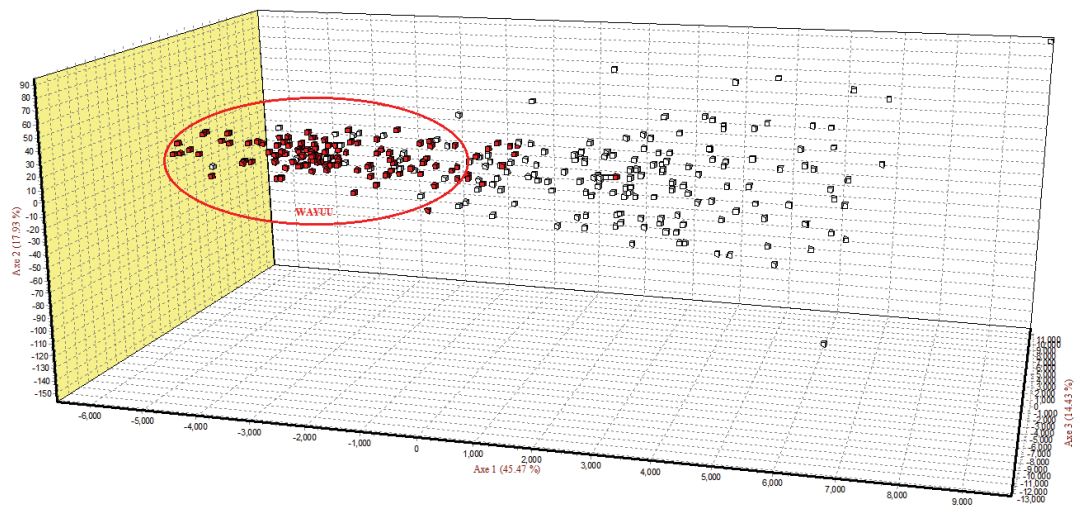


Figura No. 4 Árboles de distancias mostrando las relaciones entre grupos genealógicos Dc-NJ.

Se hizo un análisis factorial de correspondencias (AFC) como indicador de las diferencias entre individuos (Figura 5A) y los grupos (Figura 5B), en función de las variables genéticas evaluadas (Genetix 4.05) (Belkhir et al, 2001). El uso de esta herramienta es posible, dado a que la suma de los valores propios del AFC puede equipararse a los F_{st} de Robertson, Hill y Guinand (1996). Los cuatro

factores encontrados hicieron una contribución a la inercia total del 91.46%, dividida así (factor-1 45.47%, factor-2 17.93%, factor-3 14.43% y factor-4 13.59%). Se graficaron los tres primeros ejes mostrando un 63.4% de inercia. En el hiperespacio se observó una nube de puntos que muestra diferenciación de los individuos Wayúu, con respecto a los demás. Estas similitudes son causadas por las distancias entre las distribuciones de los alelos, reflejada en los aportes al centro de gravedad del análisis (Barton NH & Slatkin, 1986).

A)



B)

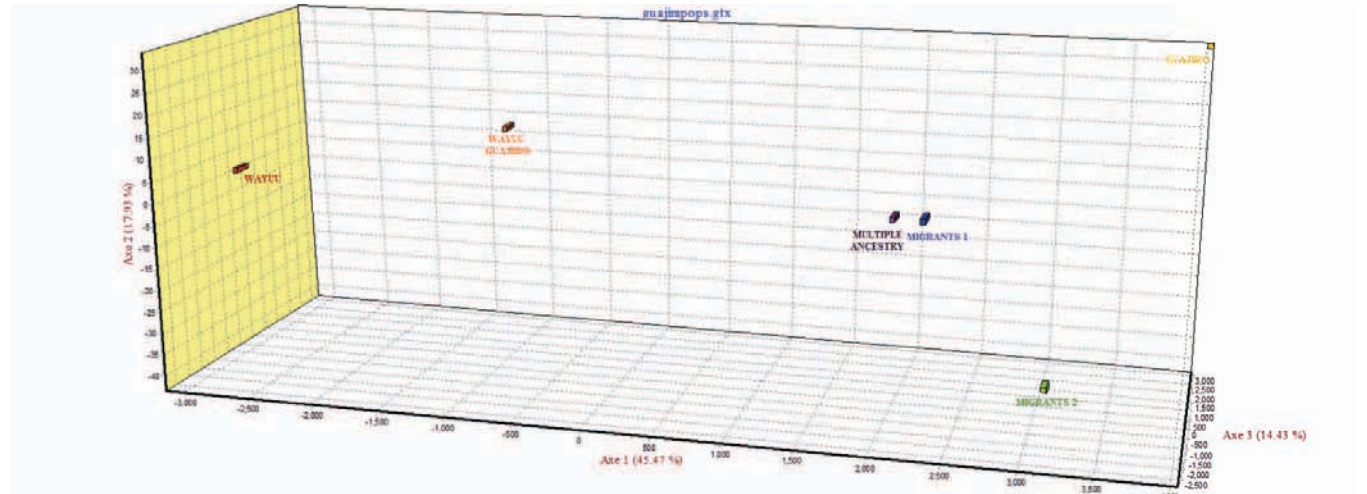


Figura No. 5 Plano Factorial ACM mostrando A: Relaciones entre individuos B: Relaciones entre poblaciones.

Se encontró en el análisis de asignación de individuos a poblaciones de Structure (Figura 6) que el $\ln P(D)$ más cercano a cero fue para un $K=2$ (-14836.6 \pm 16.46). Al analizar las agrupaciones genealógicas *a priori* se evidenció que el $K=2$ corresponde a una agrupación Wayúu y a las poblaciones (Guajiro, Migrantes-1, Migrantes-2 and Ancestría Múltiple). El valor de alfa de 0.2099 ($\alpha < 1$) mostró que en general los individuos tienen mayoritariamente ancestría de la subpoblación asignada. No obstante se puede identificar un nivel de flujo génico entre las dos subpoblaciones. Como se esperaba Wayúu-Guajiro mostró un gradiente de mezcla entre ambas agrupaciones por su naturaleza de genealogía mixta. El nivel de introgresión del grupo externo a la población Wayúu para toda la distribución de frecuencias alélicas fue de 0.0016 \pm 0.0025.

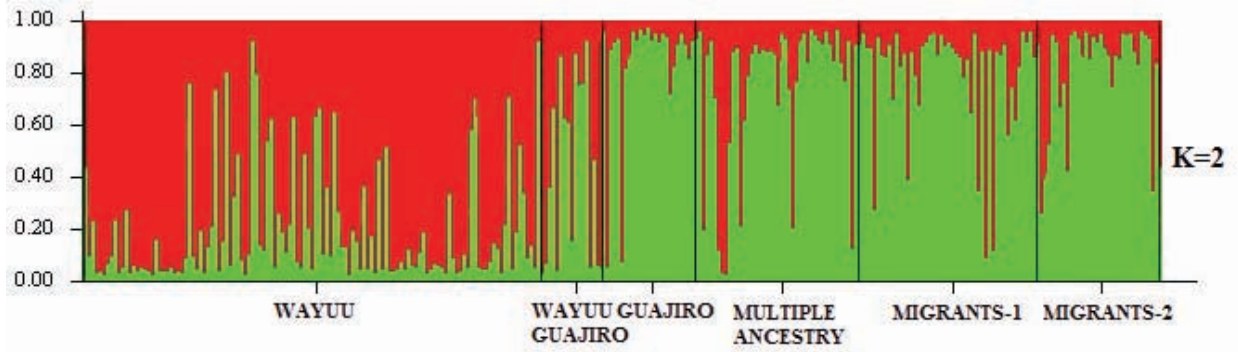


Figura No. 6 Bar Plot mostrando la estructura poblacional infiriendo estructura poblacional por asignación de individuos. *STRUCTUREK=2*.

En el estudio de la población Wayúu una variable importante, como se mencionaba anteriormente es la casta. Debido a que puede usarse como indicador de pertenencia a las familias con ancestría Wayúu. Estas se heredan de la madre a sus hijos, junto con la pertenencia de los Wayúu. Por otro lado es un indicador de mestizaje para los individuos que genealógicamente no son totalmente Wayúu, porque se pierde cuando es heredado por linaje paterno. Con base a esta información, el valor Q de asignación de individuos a poblaciones y la agrupación genealógica, se realizó un análisis de correspondencias múltiples. Se obtuvo un plano principal (Figura 7), donde las castas se asocian a las agrupaciones Wayúu y Wayúu-Guajiro, los grupos con los valores Q más altos encontrados por Structure. Esta asociación se presenta porque los estados de las variables se presentan en los mismos individuos, en ese sentido son similares entre sí.

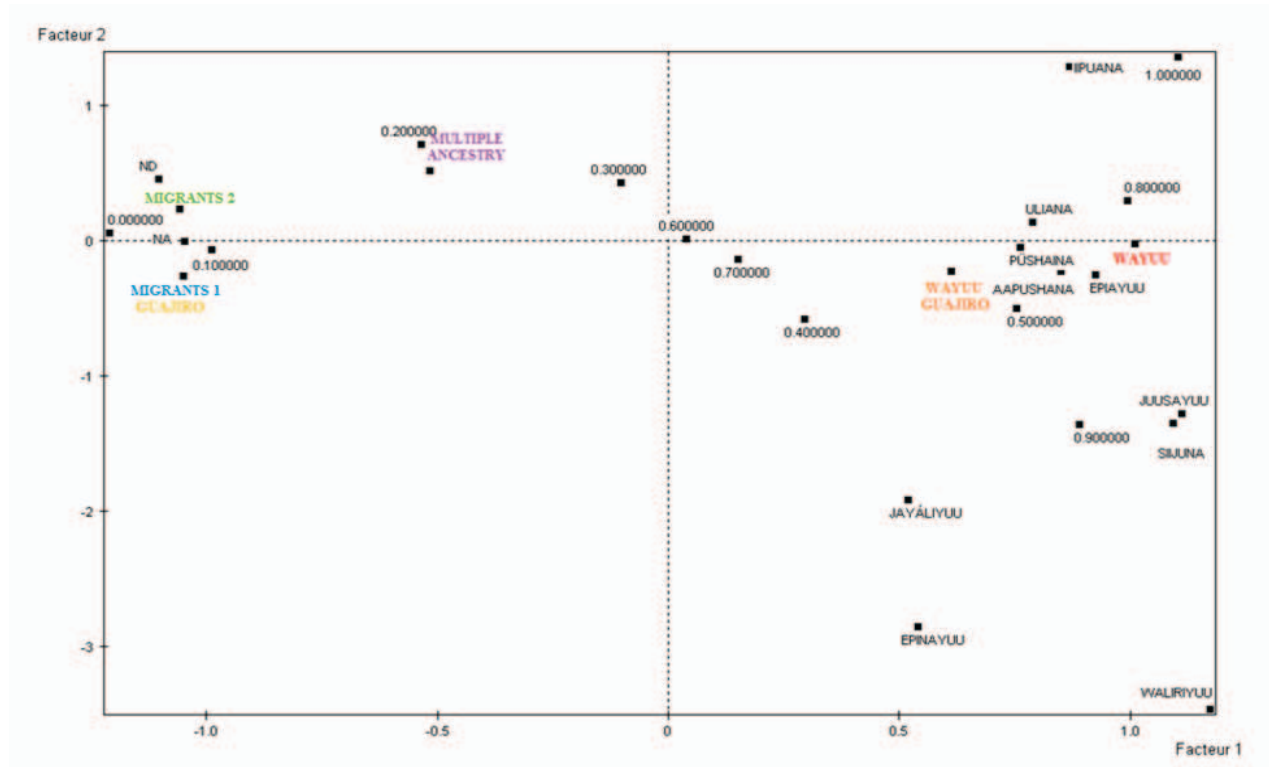


Figura No 7 Grafica de correspondencias múltiples, variables: Castas, Valor Q (Structure), Grupos Genealógicos

3.3 Análisis Genético de las Agrupaciones Genéticas

Dado que ya se evidenció la presencia de dos individuos, se procedió a la caracterización de las subpoblaciones. Primero que todo, se retiraron del análisis los individuos de la agrupación Wayú-Guajiro, con el fin de dilucidar las diferencias entre grupos e identificar unas posibles poblaciones parentales. Otra causa para retirarlo del análisis fue el pequeño tamaño de muestra.

Al evaluar el EHW por locus se encontró igualmente los sistemas D8S1179 p-valor de 0.0432, TH01 0.0110 y el sistema CSF1P0 0.0224 en desequilibrio. También se encontró la población en desequilibrio con un P-valor de 0.0241. En las subpoblaciones se encontró la población Wayú en

equilibrio con un P-valor de 0.2543 y la población externa en desequilibrio con un P-valor de 0.0164. Por otro lado, al asumir dos poblaciones (Tabla 6) se encontró un Capf de 0.036 con IC de 0.036+/- 0.0101. El valor de Theta fue 0.024 CI 0.023 +/- 0.006 y el valor de de Smallf fue 0.012 IC 0.012+/- 0.008(Tabla 6). Estos valores son mayores, aunque los intervalos de confianza se sobrelapan con los de los seis grupos. El valor del Rst fue 0.0254.

LocName	All			Rst	Wayuu	External Group
	Capf	Theta	Smallf		Smallf	Smallf
CSF1P0	0.073	0.0090	0.065	0.044	0.12	0.023
D13S31	0.006	0.0230	-0.017	0.009	-0.073	0.026
D16S53	0.037	0.0080	0.03	0.014	0.05	0.015
D18S51	0.036	0.0030	0.033	0	0.01	0.052
D19S43	0.009	0.0150	-0.006	0.01	0.045	-0.047
D21S11	0.03	0.0210	0.009	0.01	-0.024	0.034
D2S133	-0.009	0.0040	-0.012	-0.003	-0.008	-0.016
D3S135	0.043	0.0640	-0.023	0.093	0.046	-0.063
D5S818	0.084	0.0540	0.032	0.113	0.069	0.006
D7S820	0.011	0.0220	-0.011	0.066	0.019	-0.033
D8S117	0.018	0.0040	0.014	0.019	-0.044	0.063
FGA	0.029	0.0240	0.004	0.001	-0.052	0.046
TH01	0.152	0.0760	0.082	-0.004	0.165	0.031
TPOX	0.001	0.0110	-0.01	0.014	-0.007	-0.011
vWA	0.026	0.0260	0	-0.004	-0.003	0.002
Overall	0.036	0.024	0.012	0.0254	0.017	0.009
CI(Jk)	0.036+/-0.010	0.023+/-0.006	0.012+/-0.008			
CI(Bs 95%)	0.019+/-0.057	0.014+/-0.035	-0.001+/-0.028			
CI(Bs 99%)	0.015+/-0.065	0.011+/-0.039	-0.022+/-0.074			

Tabla No. 6 Diversidad y subestructura para los dos grupos genéticos: Se reporta el Smallf (Fis) (Weir & Cockerham, 1984) para las dos subpoblaciones, retirando los datos para la agrupación Wayuu-Guajira. Para la población total se reportan los estadísticos-F y el Rst por locus y global. Los intervalos de confianza (CI) por Jackknifing (Jk), Bootstapping (Bs)

Finalmente se evaluó la migración entre los dos grupos, calculando el número de migrantes basado en los alelos privados de las dos subpoblaciones, la estimación de las tasas de migración reciente y un

cálculo de asignación de individuos a poblaciones y a estados temporales de migración.

Con base a una frecuencia promedio de los alelos privados de 0.008, se estimó un número de migrantes de 9.5 para la población total. La estimación de las tasas de migración reciente en BayesAss reveló una potencial introgresión de los alelos externos en la población Wayúu ($m=0.1041$; $DS=0.024$) con un intervalo de confianza al 95% (0.062976- 0.156755), mientras que la tasa de no-migración de ese grupo externo fue de ($m=0.960279$; $DS=0.016616$) con un IC (0.9223- 0.987521). Adicionalmente, la tasa de migración de los individuos Wayuu hacia la población externa fue ($m=0.0397$; $DS=0.0166$) con un IC (0.0124, 0.0776) y dentro de la población la tasa de no-migración ($m=0.960279$; $DS 0.016$) con un IC de (0.9223-0.9875). En el cálculo de asignación de individuos a las dos poblaciones y a los siguientes posibles estados temporales (No migrante, migrante y descendiente de migrante), representado en un barplot (Figura 8), se muestra como efectivamente la migración hacia la población externa es mayor que hacia la población Wayúu y que la mayor migración se dio en individuos descendientes de la población hacia poblaciones externas y los migrantes están especialmente categorizados como individuos de ancestría múltiple y guajiros.

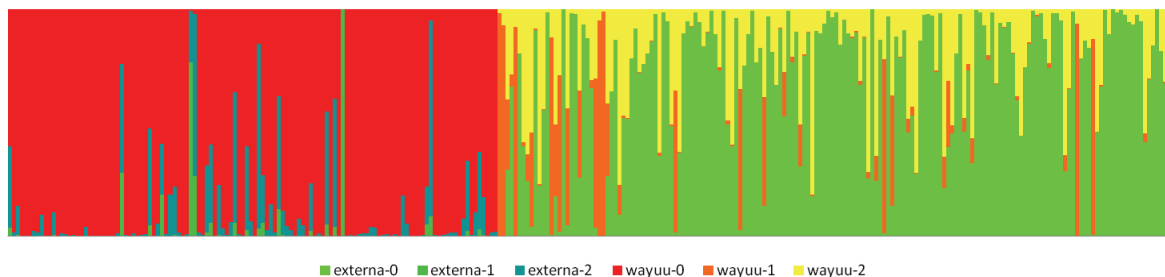


Figura No. 8 Barplot con la asignación de individuos a poblaciones con la estimación de 0: no migrantes, 1: migrante y 2: descendientes de migrantes.

4. Discusión.

En la Guajira, al igual que en otras poblacionales americanas, la dinámica poblacional es compleja debido a su estrepitosa historia de poblamiento, a los múltiples orígenes de sus pueblos y a los recientes hechos de conquista y colonización (500 años un periodo corto en la escala genética).

El Caribe colombiano en particular es una clara muestra de ello. Es una región, cuyos pobladores ancestrales corresponden a comunidades indígenas americanas, las cuales previamente ya habían vivido diferentes guerras, procesos de migración, epidemias y además habían construido complejos procesos sociales y culturales; de los cuales solo tenemos una fracción de la información. Además es un lugar de arribo para migrantes de todas partes del mundo e influenciado por la esclavitud de los siglos XV al XVIII. Tanto la cultura indígena actual como la población residente actual, tiende a conservar muchos de esos elementos, en una mezcla de tradiciones ancestrales, elementos europeos de la conquista y de la colonia y del proceso actual de globalización.

Esta dinámica poblacional esta comenzándose a estudiar en los estudios genético poblacionales. Tradicionalmente estos estudios en Colombia, se han limitado a un número restringido de muestras, tomadas de las comunidades y en su selección se han basado principalmente en criterios de autodeterminación. Si bien en este tipo de estudios los muestreos son por conveniencia, se ha caído en la tendencia de realizarlos sin el mayor rigor metodológico, ni matemático. Nuestro estudio, no se diseñó a criterio del investigador, fue específico para la población Wayúu. Diseñado desde un abordaje tanto genealógico-demográfico como genético. Esto con el propósito de realizar un cuidadoso y riguroso examen de la población.

Estos diseños redundan en mejores tamaños de muestra. En este caso se tomaron 290 personas, de las cuáles 122 son Wayúu de las cuales se tiene certeza de su pertenencia por los criterios genealógicos. Más aún esta es la muestra de comunidades indígenas más grande en el país y no estuvo focalizada en un solo lugar de muestreo, sino que se desarrolló en una amplia región geográfica del departamento de la Guajira. Aun así, cuando fragmentamos la muestra se redujo el tamaño muestral dentro de las

agrupaciones. Pero la selección cuidadosa de los individuos permitió compensar este efecto arrojando resultados significativos.

La encuesta fue útil para el refinamiento del muestreo por conveniencia y fue una herramienta práctica de toma de datos que permitió tener un conocimiento integral del participante (reconstrucción genealógica, información demográfica y de pertenencia étnica). Pero en realidad, esta herramienta permitió definir clasificaciones genealógicas para el estudio de la subestructura y de la dinámica migracional, que reflejan las hipótesis de trabajo del estudio. Estas agrupaciones se evaluaron desde la demografía y desde la genética para estudiar el efecto de la estructura de sexos y edades, el tamaño poblacional, del aislamiento reproductivo, la formación de parejas de modo no aleatorio y la endogamia. Factores insoslayables cuando se estudia la estructura genética de una población (Acreche N. 2004).

La evaluación de parámetros demográficos permitió describir en este sentido la dinámica de crecimiento, el efecto del tamaño reproductivo y la migración en la población sobre el efecto neto de la deriva genética. Además de dar una idea del nivel de flujo génico entre los grupos analizados.

La distribución de edad y sexo mostró que la muestra se asemeja a una pirámide poblacional característica de una población donde la mayoría del pool génico se encuentra en las generaciones jóvenes en edad reproductiva. El 51% de los individuos del tamaño reproductivo de la muestra se encontraron entre 20-38 años. La tasa de fecundidad de 2.87 hijos por mujer es consistente, indicando que la población se encuentra en crecimiento superando la tasa de sustitución de la misma (2 hijos por mujer) (CIA, 2011). Encontrar poblaciones con la mayoría de su pool génico concentrado en los grupos etarios jóvenes, en crecimiento y con un activo proceso reproductivo, podría hacerla sensible al efecto de la deriva genética, generando posibles procesos estocásticos de pérdida y ganancia de variantes alélicas. Percibimos que varias características culturales de la población tienen como consecuencia el activo proceso de reproducción. Sin embargo, esto no implica que la población se encuentre en expansión. Debido a factores compensadores como la alta tasa de mortalidad infantil (38.4 muertes de menores de 5 años por cada mil nacidos vivos) (DANE, 2005) y el efecto del

ecosistema en el crecimiento poblacional. Por lo tanto, 10.000 años de asentamiento en la zona no ha generado una expansión exponencial. Posiblemente el ecosistema de desierto tiene una capacidad de carga menor por la escases de recursos. El DANE reporta al respecto, que el 70% de la población Wayúu tiene las necesidades básicas insatisfechas, en general por la carencia de agua y el debilitamiento de la economía tradicionalmente Wayúu.

Los factores evaluados anteriormente afectan también el tamaño efectivo poblacional dado a que involucra tanto al tamaño reproductivo como al número de hijos por familia. Encontramos para la población total un N_r de 53.4% del N (290) y un N_e del 15.9%) y para la población Wayúu (N_r el 45.9% y N_e del 13.5%). Comparado con otros estudios: Magalhaës y Arce Gómez en 1987 de la población Brazilerá aislada de Guaraqueaba, se encontró un N_r menor de 36.6% y un N_e similar al Wayúu de tan solo del 13.2%. Se estiman que para la mayoría de poblaciones humanas el tamaño efectivo es del 25% del N total. Por lo que los valores encontrados parecen ser pequeños, posiblemente afectados por el hecho de no haber tomado individuos menores de 18 años y de no considerar la superposición de generaciones.

Como se mencionaba anteriormente, suponemos que hay un activo proceso de migración hacia los Wayúu de los demás grupos genealógicos. Estas categorías permitieron el estudio de la migración de grupos consistente con el movimiento de diferentes linajes en el territorio. De esta manera los migrantes a larga distancia (61.3%) se tomaron como los individuos con genealogías de lugares de nacimiento fuera de la región geográfica. Mientras que los migrantes a corta distancia fueron los descendientes de Wayúu o Guajiros y migrantes. Esta ordenación, permitió evaluar el movimiento de las variantes alélicas de diferentes genealogías y que caracterizan la subestructura actual.

La tasa de migración junto al tamaño efectivo de la población, permitió evaluar el coeficiente de aislamiento reproductivo, para medir el accionar de la deriva genética. EL valor de 11.9 indicó una moderada sujeción de la deriva genética, posiblemente contrarrestada por el flujo génico de la población. Adicionalmente se encontró un coeficiente de endogamia de 0.0129 lo que indica que no es una población endogámica. En poblaciones humanas estas consideraciones tienen sentido dado

a que presentan características sociales y distribuciones en el espacio que hacen que el sistema de apareamiento no sea al azar, formando “barreras” reproductivas que afectan el flujo génico entre grupos, pero no necesariamente producen endogamia dentro de las poblaciones.

Las diferencias de flujo génico y el efecto de la deriva genética, también están influenciados por la formación de parejas de modo no aleatorio entre grupos en los diferentes espacios geográficos. Los análisis de flujo genético y distancias entre parejas, mostraron que los Wayúu establecen grupos familiares mayoritariamente con personas de la comunidad y que se encuentran a una distancia geográfica menor a comparación de otros grupos (un radio máximo de migración de 290 Km) Es así como la tendencia de los individuos nativos, es a establecerse dentro de las poblaciones en vez migrar, esto producido principalmente por aspectos culturales. Como consecuencia, el flujo génico entre grupos genealógicos es moderado a pesar de que coexistan en las mismas regiones geográficas (51,9% de personas viven en las cabeceras municipales).

La caracterización de la subestructura y los patrones de migración se complementó con el análisis genético de 15 STRs autosómicos. En general, estos permitieron hacer una buena aproximación al modelo de subestructura genética y a los estimadores de diversidad genética de la población.

En primera instancia se encontró que los Wayúu ($H_o = .727$) así como los demás grupos genealógicos presentaron niveles de diversidad genética altos. El hecho de que tenga alta diversidad en comparación con otras poblaciones amerindias Colombianas como Emberá 0.655, Ingano 0.687, Ticuna 0.682 y Zenu 0.666 (Mesa NR et al, 2000), puede ser explicado por los procesos sociales de matrimonios entre personas de diferentes castas o personas externas a la comunidad Wayuu; dinámica que subsecuentemente reduce la endogamia de la población. Adicionalmente el encontrar desequilibrio en la población global posiblemente fue causado por la presencia de diferentes distribuciones genotípicas en la muestra de poblaciones diferentes, dando un inicio de que no hay una sola población en la muestra.

No se encontró un coeficiente de endogamia alto (0.011 con un IC de 0.012+/-0.008) al encontrado

con datos demográficos. Por otro lado, interpretando los estadísticos-F la población tiene una moderada diferenciación genética, pero significativa entre grupos. Comparado con una mayor variación entre individuos. El valor de Theta total (0.016 con IC 0.021+/- 0.005) y por marcador mostró que hay por lo menos un grupo que genera diferenciación genética. El hecho de que los Wayúu presenten un grado de subestructura moderado, no sea una población endogámica y presente alta variabilidad, muestra que la diferenciación puede ser causada por diferencias en su distribución genotípica, con respecto a las poblaciones externas.

La subestructura genética fue corroborada por el AMOVA donde las agrupaciones: Wayúu y (Wayúu, Wayúu-Guajiro) diferían significativamente de las demás agrupaciones genealógicas. Es así como los resultados indicaron la presencia de dos grupos: los Wayuu junto con el grupo intermedio Wayuu-Guajiro y el segundo compuesto por los grupos migrantes, guajiros y el grupo de ancestría múltiple. Estos grupos intermedios son producto del flujo génico entre las diferentes agrupaciones. Por otro lado, la evaluación de la distancia genética, mostró los mismos agrupamientos. Sin embargo es de resaltar, el clúster Wayúu y Wayúu-Guajiro consistente. Finalmente el ACM, también mostró la misma tendencia en los individuos. Y en efecto, se nota claramente que los grupos intermedios no son homogéneos, dado que los individuos Wayúu-Guajiros están más cercanos a Wayúu mientras que los individuos de ancestría múltiple se muestran cercanos a los Migrantes-1.

Para validar el modelo de subestructura genética *a priori* basado en las agrupaciones genealógicas, se evaluó el análisis de asignación de individuos a poblaciones utilizando una distribución *a posteriori* basada en un análisis bayesiano sobre la información genotípica de los individuos. Se encontró que la población se ajusta mejor a un $K=2$ donde Wayúu forma un núcleo diferente al de los demás grupos. Al calcular este nivel de introgresión a la población Wayúu desde los demás grupos se encontró 0.16%. El ACM muestra en general un gradiente de mestizaje desde Wayúu (con los mayores valores Q) hacia los migrantes. Las castas se asociaron tanto a Wayúu como a Wayúu-Guajiro, mostrando que como variable característica de la población, se asocia directamente a los individuos Wayúu, ratificando su pertenencia a los clanes matriarcales y por ende afirma su ancestría. El análisis igualmente permitió dilucidar las diferencias entre los grupos intermedios si se evalúan con respecto a

una población Wayúu parental.

Sobre la subestructura definida, se encontró igualmente un DHW global, pero al obtener desequilibrio en el grupo externo asumimos que no hay homología ente ellos. Al asumir las dos subpoblaciones se recalcularon los estadísticos-F, aumentando el valor Theta a 0.024 con un IC de 0.023. Igualmente el R_{st} aumento a 0.0254. Estos estadísticos mostraron mayor diferenciación entre los grupos sobre la diferencia entre individuos que el que se presentaba con las seis agrupaciones. Finalmente, con el fin de complementar el modelo de dinámica migracional de la población Wayúu que se planteó con demografía, con los datos genéticos se encontró que el número de migrantes basado en alelos privados entre las dos subpoblaciones fue 9.5. Este número es consistente dado que en este modelo de subestructura los Wayúu son una población diferencial con niveles bajos de introgresión en términos del movimiento de alelos. Sin embargo según el análisis de las tasas de migración reciente, muestra que esta pequeña introgresión ($m=0.1041$ $DS=0.0240$) es mayor que la de Wayúu hacia la población externa ($m=0.0397$ $DS=0.0166$), este resultado fue corroborado por la asignación de individuos a las subpoblaciones y a los estados de migración. La migración detectada de la población Wayúu es mayoritariamente de descendientes de Wayúu. Este resultado muestra la dirección de flujo génico entre las dos subpoblaciones, posiblemente generado por la aculturación de la población indígena. Sin embargo, como se mencionaba en el análisis de flujo génico entre parejas la tendencia es a mantenerse dentro de cada subpoblación.

6. Conclusión

En esta investigación, en resumen queremos resaltar que las categorías de clasificación de individuos son primordiales al diseñar el muestreo para detectar la subestructura poblacional. Por lo que al decidir si y cómo utilizar las categorías raciales, étnicas y ancestrales en la investigación, los genetistas nos enfrentamos a escenarios contradictorios. Muchos investigadores han propuesto fuertes argumentos a favor de reducir o incluso eliminar el uso de categorías raciales o étnicas en la investigación genética (Boyer et al 2005), dado que clasificaciones tradicionales que acuñen a los términos de mestizo, europeo y afro-descendiente son imprecisas. Los análisis de poblaciones

humanas requieren de clasificaciones más robustas y cuidadosamente realizadas que permitan hacer inferencias mucho más finas y adecuadas que correspondan a criterios que den cuenta de la dinámica de poblamiento y a los procesos históricos específicos de la población.

La definición de un modelo de clasificación fenotipo- específico de cada uno de los grupos humanos utilizando categorías raciales, para nosotros carece de sentido, dado que encontraremos en los procesos de constitución y de poblamiento una amplia gama de matices y ancestrías propias de poblaciones humanas, como es el este caso de las americanas, en donde se reúnen múltiples ancestrías, que obedecen más a los diferentes procesos de migración que a la constitución de nuevos grupos humanos o a la mezcla de grupos étnicos tradicionales. La clasificación genealógica realizada en este estudio, permite diferenciar a los Wayuu, que puede definirse como una población parental en estudios posteriores. Por ende sugerimos el uso de categorías genealógicas o basadas en características históricas o sociales, específicas para responder las hipótesis en los diferentes escenarios de cada estudio.

Se encontró que la población de la Guajira es una población en términos de su estructura de edad, joven y se encuentra en crecimiento. Además se encuentra bajo un efecto moderado de deriva genética, la población no es homogénea conforme a su historia de poblamiento y composición étnica. Existe subestructura poblacional diferenciando a los individuos Wayúu, de los residentes y migrantes tanto Caribes como de otras regiones de Colombia, esta población indígena mostro ser una población conservada, a pesar de que coexiste y presenta un flujo migracional bidireccional con los demás grupos poblacionales, esta dinámica migracional está representada en grupos de ancestría múltiple con diferencias en su tipo de ascendente Wayúu (Wayúu-Guajiro), donde no todos los individuos se autodeterminan como pertenecientes a la comunidad por razones culturales. No obstante, no se presentan diferencias significativas con el grupo Wayúu sustentando este dato con la información genealógica. Esta población es un ejemplo de un proceso de subestructura a nivel local que en su mayoría no es causado por aislamiento por distancia, si no por diferencias culturales y de parentesco que generan posibles barreras reproductivas. El hecho de que a nivel local se encuentren poblaciones aisladas es importante en estudios posteriores de la etiología de algunas enfermedades y sus

implicaciones en los aspectos prioritarios de salud pública.

7. Agradecimientos

En primer lugar a los habitantes de la población de la Guajira y a la Comunidad Wayuu, por permitirnos conocer un poco más acerca de su historia biológica, así como también a las diferentes instituciones de salud: al Hospital Nuestra Señora del Perpetuo Socorro, al Hospital San José de Maicao, al Hospital Nuestra Señora de los Remedios y a la IPS CAPRECOM quienes prestaron su colaboración en el muestreo. En segundo lugar, a la Dirección de Investigación (DIB) de la Universidad Nacional de Colombia por brindarnos los recursos financieros para realizar el estudio, al Instituto de Genética de la Universidad Nacional y finalmente a los miembros del Grupo de Genética de Poblaciones e Identificación por su apoyo e ideas en la realización de este estudio.

8. Bibliografía

- Albeza MV, Acreche NE, Caruso GB. 2004. Biogeografía en poblaciones de la Albeza MV et al, 2004 (Chañarcito, Santa Rosa de los Pastos Grandes y Olacapato) Salta, Argentina. *Revista de Antropología Chilena* 34 (1): 119-126.
- Arango R, Sánchez E. Los pueblos Indígenas de Colombia en el Umbral del Nuevo Milenio. Departamento Nacional de Planeación de Colombia. Dirección de Desarrollo Territorial Sostenible. Bogotá, 2004.
- Ardila G, Cuando el progreso se enfrenta a la vida: los Wayúu de la Guajira. *Diversidad es Riqueza, ensayos sobre la realidad colombiana*. ICAN, Instituto Colombiano de Cultura y Consejería Presidencial para los Derechos Humanos. Bogotá, 1992.
- Barton NH, and Slatkin M. 1986. A quasi-equilibrium theory of the distribution of rare alleles in a subdivided population. *Heredity* 56 (3):409-15.
- Belkhir K, Borsa P, Chikhi L, Raufaste N, and Bonhomme F. 2001. GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome Populations Interactions.

- CELADE (Comisión Económica para América Latina y el Caribe). Perfil Socio-demográfico básico de la Guajira. www.eclac.cl/celade/noticias/paginas/2/40392/1_La_Guajira.pdf 15 de Noviembre de 2011.
- CIA, 2011 The World Factbook. <https://www.cia.gov/library/publications/the-world-factbook/> 20 de Septiembre de 2011.
- DANE (Departamento Administrativo Nacional de Estadística). Censo Nacional 2005 Nivel Nacional. Dirección de Difusión, Mercadeo y Cultura Estadística. Bogotá.
- Goudet J. 1995. FSTAT (Version 1.2): A computer program to calculate F-statistics. *Journal of Heredity* 86:485-486.
- Gobernación de la Guajira, Pagina web oficial de La Guajira. <http://dev.laguajira.gov.co/joomla/> consultado 10 de septiembre de 2011
- Keyeux G, William UM, Rutas Migratorias hacia Suramérica y poblamiento de las cuencas del rio Amazonas y Orinoco, deducidas a partir de estudios genético moleculares. *Pueblos y paisajes antiguos de la selva amazónica*. Pág. 43-62
- Langella O (2002) populations 1.2.28. Population Genetic Software (Individuals or Populations Distances, Phylogenetic Trees).<http://www.cnrs-gif.fr/pge/bioinfo/populations/index.php>
- Magalhães JC, Arce-Gomez B. 1987 Study on a Brazilian isolate. I. Population structure and random genetic drift. *Hum. Hered.* 37 (5):278-84.
- Mesa NR, Mondragón MC, Soto ID, Parra MV, Duque C, Ortíz-Barrientos D, García LF, Vélez ID, Bravo ML, Múnera JG, et al. 2000. Autosomal, mtDNA, and Y-chromosome diversity in Amerinds: pre- and post-Columbian patterns of gene flow in South America. *American journal of human genetics* 67:1277-86.
- Oliver JR, Reflexiones sobre los posibles orígenes del Wayúu (Guajiro). In: *La Guajira: de la memoria al porvenir. Una visión antropológica*. 1991. Editorial Fondo FEN. 83-135). Bogotá, Colombia.
- Pritchard JK, Stephens M, and Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945-959.
- Programa Presidencial de Derechos Humanos. Diagnóstico de la situación del pueblo indígena

Wayúu.

http://www.derechoshumanos.gov.co/Observatorio/documents/2010/DiagnosticoIndigenas/Diagnostico_WAY%C3%9AU.pdf 15 de Septiembre de 2011.

Profamilia. Encuesta nacional de demografía y salud. (2007) Bogotá.

Pérez L. Los Wayúu: Tiempos, Espacios y Circunstancias. Cuaderno Venezolano de Sociología 15:403-426

Raymond M. & Rousset F, 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *J. Heredity*, 86:248-249.

Rousset, F., 2008. Genepop: a complete reimplementation of the Genepop software for Windows and Linux. *Mol. Ecol. Resources* 8: 103-106.

Schneider S, Roessli D, and Excoffier L. 2000. Arlequin: A software for population genetics data analysis. User manual ver 2:2496–2497.

Slatkin M, 1995. A measure of population subdivision based on microsatellite allele frequencies. 1995 *Genetics* 139(1):457-462.

Vásquez S., Correa H. Los Wayuu, entre Juya (“El que llueve”), Mma (“la tierra”) y el desarrollo urbano regional. *Geografía Humana de Colombia, Nordeste Indígena Tomo II*. <http://www.banrepcultural.org/blaavirtual/geografia/geograf2/wayuu1.htm> consultado el 10 de septiembre de 2011.

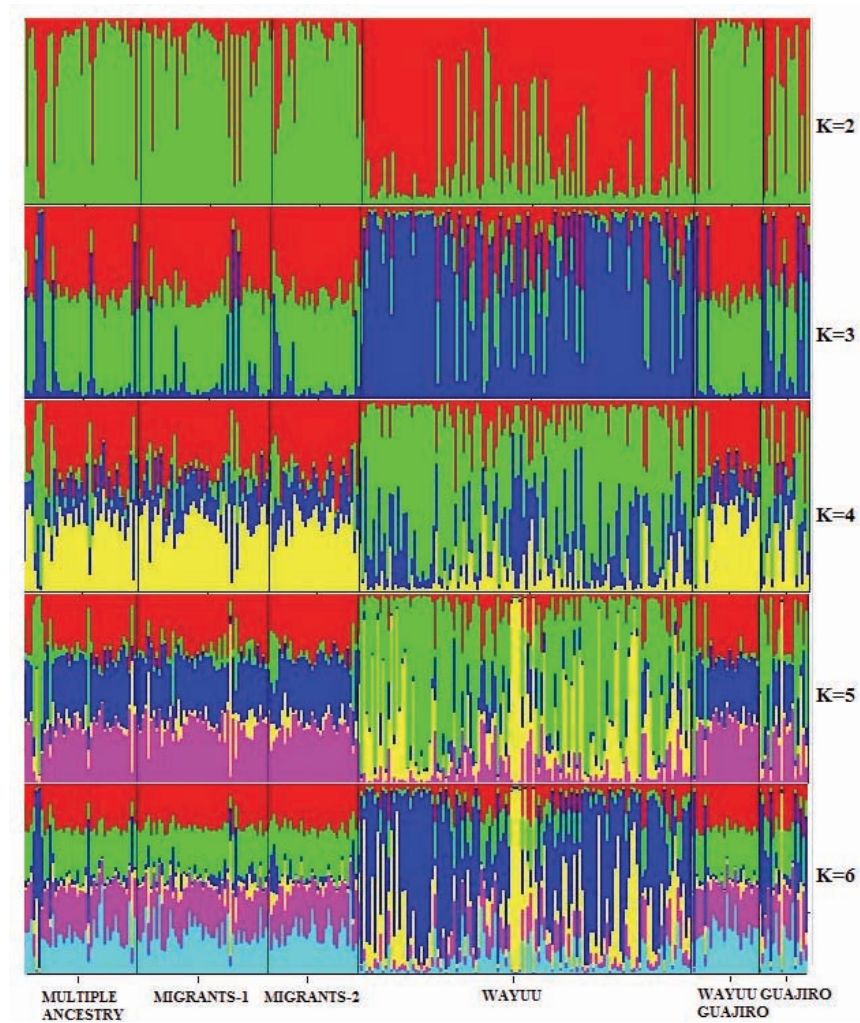
Wilson GA, Rannala B. 2003 Bayesian inference of recent migration rates using multilocus genotypes. *Genetics* 163:1177-1191.

Weir B.S, Cockerham C.C. 1984. Estimating F-statistics for the analysis of population structure. *Evolution*, 38(6):1358:1370.

8. Anexos.

K	Ln P(D)	Var(LnP(D))	α
1	-15150.0	72.0	-
1	-15150.7	72.7	-
2	-14837.7	269.7	0.1997
2	-14836.6	271.2	0.2099
3	-15306.8	1271.3	0.0827
3	-15010.6	627.1	0.0811
4	-15391.7	1574.6	0.0638
4	-15439.2	1635.4	0.0600
5	-15725.6	2382.2	0.0631
5	-15483.1	1972.0	0.0661
6	-16060.2	3103.5	0.0620
6	-15954.6	2956.8	0.0613


Anexo No 1: Tabla de los valores de Ln P(D) para dos repeticiones del análisis mostrando K=1-6 y los valores de varianza y alfa para cada corrida de datos.



Anexo No 3. Grafico de Barras mostrando la estructura poblacional inferida por asignación de individuos a poblaciones STRUCTURE software K=2, 3, 4, 5, 6.

Number	Value	Percentage	Percentage	
1	0.9	10.1	10.1	*****
2	0.5	5.4	15.5	*****
3	0.5	5.4	20.8	*****
4	0.4	4.9	25.8	*****
5	0.4	4.7	30.5	*****
6	0.4	4.6	35.1	*****
7	0.4	4.5	39.6	*****
8	0.4	4.4	44.0	*****
9	0.4	4.2	48.2	*****
10	0.4	4.1	52.3	*****
11	0.4	4.1	56.4	*****
12	0.3	3.9	60.3	*****
13	0.3	3.9	64.1	*****
14	0.3	3.8	68.0	*****
15	0.3	3.7	71.6	*****
16	0.3	3.6	75.2	*****
17	0.3	3.5	78.7	*****
18	0.3	3.3	82.0	*****
19	0.3	3.2	85.2	*****
20	0.3	3.1	88.3	*****
21	0.3	2.9	91.2	*****
22	0.2	2.7	93.9	*****
23	0.2	2.6	96.4	*****
24	0.2	2.3	98.8	*****
25	0.1	0.9	99.6	*****
26	0.0	0.4	100.0	****

Anexo No 4. Histograma de los 26 primeros valores propios del ACM análisis de correspondencias múltiples.



UNIVERSIDAD NACIONAL DE COLOMBIA SEDE BOGOTÁ

PROJECT: "GENETIC CHARACTERIZATION IN A HUMAN POPULATION FROM THE GUAJIRA DEPARTMENT: CONTRIBUTION TO HISTORICAL AND STRUCTURAL KNOWLEDGE ABOUT COLOMBIAN ETHNIC GROUPS"

1.

Date	DD	June	2009
Sample code			
Sampling Institution			
Sampling officer			
Pollster name			

2. Sampling site:

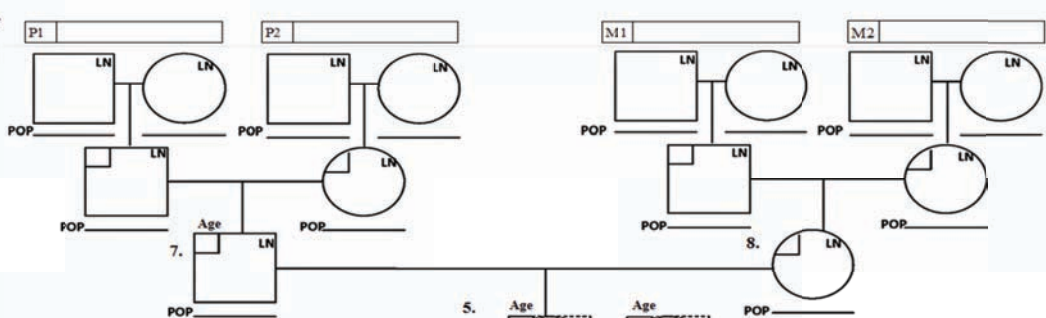
Municipality	Specific site
--------------	---------------

3. Personal Information:

Residence Place	Telephone Number	Residence Time
Adress		E-mail
First name	Middle Name	

4. Surname:

P1 <input style="width: 80%;" type="text"/>	P2 <input style="width: 80%;" type="text"/>	M1 <input style="width: 80%;" type="text"/>	M2 <input style="width: 80%;" type="text"/>
---------------------------------------------	---------------------------------------------	---------------------------------------------	---------------------------------------------



6.

Ethnicity	Caste	Rancheria				
Age of Mother at first birth	Offspring number					
Wayunaiki	Talking	Yes	Not	Write	Yes	Not

Anexo No5. Encuesta Genealógica y Demográfica aplicada en el muestreo en el departamento de La Guajira. P1-2 primero y segundo apellidos paternos, M1- primero y segundo apellidos maternos. POP: Población de

Capítulo No. 4: Análisis de apellidos y de cromosoma Y de las poblaciones raizales del archipiélago de San Andrés y Providencia (Colombia)

Ángela Alonso Morales; William Usaquén Martínez. Instituto de Genética. Universidad Nacional de Colombia Sede Bogotá. Bogotá-C. laalonsom@unal.edu.co. Correspondencia: Ciudad Universitaria, Avenida carrera 30 N.º 45-03, Instituto de Genética, oficina 03. Edificio 426. Bogotá, Colombia. wusaquenm@unal.edu.co

Resumen

El archipiélago de San Andrés y Providencia es un departamento de Colombia aislado en las aguas occidentales del mar Caribe. La mayor parte de sus habitantes corresponde al grupo afrocolombiano antillano denominado "raizal", el cual cuenta con características culturales únicas como producto de una mezcla centenaria de diversos orígenes principalmente esclavos africanos y colonos europeos. En la actualidad poco se conoce acerca de la composición y diversidad genética de estas poblaciones. Por ello, el presente estudio pretende investigar la variación y estructura de la población raizal del Archipiélago. Un total de 54 individuos fueron seleccionados mediante un criterio genealógico de tres generaciones de ancestros nacidos en las islas. Se analizaron 17 Y-STRs, incluyendo la información heredada en el primer apellido. De este análisis se encontró una diversidad haplotípica baja en ambas islas, en comparación a lo reportado en otras poblaciones afro descendientes. Se utilizó un análisis molecular de la varianza para evaluar la hipótesis de subestructura genética en el archipiélago, en donde no se encontraron diferencias significativas ($p > 0.05$). Se predijeron los haplogrupos de cromosoma Y presentes en las islas, siendo E1b1a y R1b los más frecuentes. Estos representan más del 80% de la muestra y son los haplogrupos característicos de poblaciones africanas y europeas; correspondiendo con la historia de poblamiento de las islas. Para análisis genéticos comparativos se calcularon distancias genéticas con respecto a otras poblaciones del Caribe, Europeas, Asiáticas, Africanas y Colombianas. Se observó concordancia entre las poblaciones afrodescendientes en relación a las demás poblaciones evaluadas. El análisis de apellidos mostró que los raizales que tienen el mismo apellido

comparten también el haplotipo en un porcentaje mayor al 50%. Esto sugiere una relación de parentesco que se sustenta en la historia de poblamiento y en los niveles de diversidad haplotípica encontrados en las islas.

Key Words: *San Andrés and Providencia, Y-Chromosome, short tandem repeat (STR), Surnames, Colombia.*

Introducción

El mar Caribe o mar de las Antillas es un mar interior rodeado en gran parte por grupos de islas. En épocas prehistóricas fue el camino que comunicó a numerosos pueblos del centro de América con los pueblos de América del Sur. Posteriormente, a la llegada de Cristóbal Colón, el mar Caribe se volvió el eslabón, el paso obligado entre Europa y el nuevo mundo, convirtiéndose en el medio natural de apoyo para la expansión de los pueblos Europeos hacia el Norte, Centro y Sur América. El poblamiento del Caribe ha sido el producto de las rutas originadas en las comunidades indígenas de América, seguidas por los diferentes procesos de conquista, constituyéndose de esta manera en escenario de coexistencia multiétnica, multicultural y multireligiosa (Vollmer, 1997).

El archipiélago de San Andrés y Providencia es una población aislada en la inmensidad de las aguas occidentales del mar Caribe. Se localiza a 180 Km. de las costas centroamericanas, a 400 Km. de Jamaica y a 480 Km. de las costas colombianas. Está formado por las islas de San Andrés, Providencia y Santa Catalina y por una gran cantidad de cayos, islotes y bajos (Figura 1).

Su historia del poblamiento, se ha entendido como la historia de la conformación de un territorio construido a través de la relación de los distintos grupos humanos que allí han interactuado. Desde esta perspectiva, Vollmer (1997) menciona la existencia de seis ciclos de poblamiento que han contribuido a la dinámica poblacional actual de las islas: el primero, el territorio Miskito (no se conoce -1692), donde individuos pertenecientes a la familia Miskito Sumo Matalpaba, transitaron e hicieron del archipiélago parte de su territorio sin establecer ningún asentamiento hasta donde se conoce. Seguido del periodo denominado: las avanzadas de colonización (1629-1677), cuyo inicio se caracteriza por la llegada de Puritanos ingleses a la isla de San Andrés, que provenían de las Islas

Bermudas con propósitos agrícolas, como el cultivo de tabaco y de colonización de territorios americanos. Para 1633 llegaron los primeros esclavos al archipiélago, proviniendo principalmente de la isla Tortuga; con el paso del tiempo el archipiélago, se convirtió en un importante centro de tráfico de esclavos en el Caribe, lo que condujo también, a la llegada de piratas ingleses y holandeses con fines contrabandistas y de asalto a las naves españolas que pasaban por el Caribe occidental, llevando los tesoros robados a los indígenas de América del Sur. Por esta razón, en 1641, España toma la isla de Providencia, lo cual da inicio a un periodo de 36 años de ocupaciones militares españolas e inglesas disputándose la propiedad de las islas. Luego de este proceso, viene el siglo del Olvido (1677-1780) donde a partir del último episodio militar entre España e Inglaterra, el Archipiélago regresa a su absoluta soledad. En esta época hacia 1730, llegaron personas del Caribe anglófono: Jamaica, Barbados y Trinidad y Tobago; otras llegaron directamente de las Islas Británicas, en particular de Escocia e Irlanda, africanos de África Occidental, españoles, franceses, holandeses, los cuáles iban y venían, arribaban, zarpaban y algunos se quedaban por la belleza de las islas. El siguiente periodo, se denomina: el Poblamiento Raizal (1780-1953). Durante este periodo, la base económica en el Archipiélago fue el cultivo de algodón como producto de exportación. Para ello fue necesaria la importación de mano de obra del África occidental y del Caribe anglófono, donde los historiadores calculan que para 1806 habitaban en San Andrés 1.200 personas, de las cuales 800 eran esclavos, y en Providencia no hubo más de 300 personas: algunas de San Andrés, pero la mayoría eran de Jamaica. Para 1822 el Archipiélago fue incorporado al territorio colombiano y en 1847, Phillip Beekman funda en San Andrés la primera iglesia Bautista, lo que condujo que a finales de siglo el 95% de la población fuera batista y el 90% supiera leer y escribir la lengua inglesa. Para 1851 el gobierno Colombiano ordena por ley la abolición de la esclavitud. Para 1953 inicia el quinto periodo donde el Archipiélago fue declarado Puerto libre (1953-1991). Las nuevas oportunidades para el comercio y para el turismo, impulsaron una nueva ola migratoria hacia San Andrés. Dos nuevos grupos de pobladores aparecen: migrantes del medio oriente (sirios, libaneses, palestinos y algunos judíos) y colombianos de origen continental (de los departamentos de Bolívar, Atlántico y Antioquia). Finalmente la Constitución Colombiana, le dio al archipiélago el reconocimiento de Departamento y le otorga autonomía política, administrativa y fiscal a sus pobladores nativos.

En este estudio, se presenta un análisis genético y de apellidos de la población humana del

archipiélago de San Andrés y Providencia (Colombia) que es un microcosmos del crisol cultural del Caribe insular, cuyo origen proviene de una mezcla centenaria de migrantes (indios miskito, puritanos ingleses, piratas holandeses, esclavos jamaquinos y africanos, población caribeña y españoles) que han contribuido de manera diferencial al establecimiento de dos grupos de pobladores nativos: los raizales de la isla de San Andrés y los raizales de la isla de Providencia (Vollmer 1997; Lamprea 2009).

En las sociedades humanas tener un nombre y ser identificado es esencial (King T.E. & Jobling., 2009). Desde el periodo medieval Inglés al presente, cada hijo lleva el apellido de su padre (Sykes & Irven, 2000), lo que conlleva a que tanto los apellidos como el cromosoma Y compartan el mismo mecanismo de herencia. En el archipiélago algunos historiadores atribuyen el origen de los apellidos isleños en casi su totalidad, a la transmisión que hicieron durante la esclavitud, algunos de los amos de su nombre de familia a los esclavos de su propiedad. Pero como menciona Francis C, (1991) la gran mayoría de estos apellidos pertenecen a los descendientes directos de pobladores ingleses, irlandeses, españoles, noruegos y aún norteamericanos que se establecieron en las islas durante y después de la esclavitud. Algunos son de origen oriental (chinos, hindúes) que se sumaron a la población en los primeros años del siglo XX.

Siendo San Andrés y Providencia escenarios de diferentes eventos históricos que involucran poblaciones heterogéneas en cuanto al lugar de procedencia y ancestría, resulta muy interesante estudiar estas islas en relación a un marcador uniparental como el cromosoma Y asociado con la herencia del primer apellido, que se asume como un alelo heredado por vía paterna (Jobling, 2001). Por esta razón el objetivo de este estudio fue determinar la composición genética de los individuos raizales de San Andrés y raizales de Providencia, así como sus posibles relaciones ancestrales con otras poblaciones mediante el uso de marcadores STR del cromosoma Y, asociando también la información heredada en el primer apellido, lo que permite asociar la genética clásica con el análisis demográfico y los análisis de marcadores uniparentales.

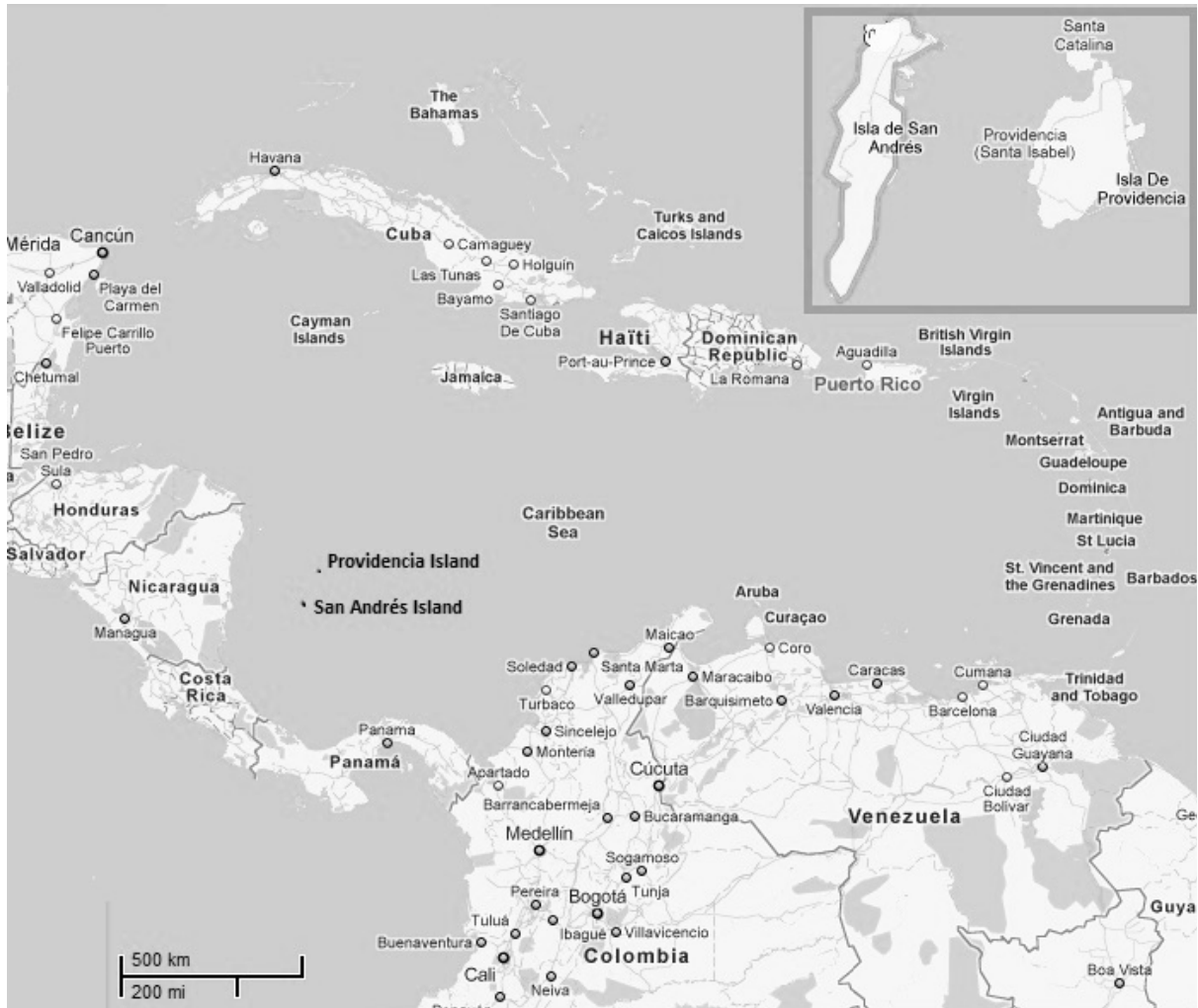


Figura 1. Posición Geográfica del Archipiélago de San Andrés y Providencia (Google Maps - ©2011 Google).

Materiales y Métodos

2.1 Población de estudio

Se recolectaron 54 hombres residentes en las islas (N=23 para la isla de Providencia y N=31 para la isla de San Andrés). Se aplicó una encuesta genealógica para clasificar como “Raizales de San Andrés” y “Raizales de Providencia” a aquellas personas que tenían al menos tres generaciones de ancestros (padre, abuelo paterno y bisabuelo paterno) nacidos en las islas. Según esta encuesta ninguno de los individuos participantes compartía ningún ancestro. Previo consentimiento informado, se tomaron muestras de 10 ml de sangre, una fracción fue preservada mediante una tarjeta FTA (Whatman BioScience, USA).

2.2 Otras poblaciones

Con el propósito de comparar las islas con otras poblaciones, se incluyeron 2243 haplotipos publicados para 34 poblaciones. La inclusión de estas poblaciones tuvo como propósito la búsqueda de similitudes y diferencias entre las poblacionales en función de la ancestría y ubicación geográfica (Tabla 1).

2.3 Análisis molecular.

Diez y siete loci Y-STR fueron amplificados por PCR utilizando el kit comercial AmpFISTR Yfiler (Applied Biosystems, Warrington, U.K.), utilizando un termociclador programable 9700 (Applied Biosystems, Foster City, California). Las amplificaciones fueron realizadas utilizando los siguientes parámetros por ciclo: denaturación inicial (95°C por 11 min) seguido por 30 ciclos de denaturación de las hebras (94°C por 1 min), alineamiento de primers (61°C por 1 min), y extensión del ADN (72°C por 1 min), y extensión final (60°C por 60 80 min). Las alícuotas del ADN amplificado y marcado con fluorescencia fueron mezcladas con formamida (Applied Biosystems, Foster City, California) y LIZ 500 (Applied Biosystems, Warrington, U.K.) y genotipificadas en un analizador genético ABI

Prism 310 (Applied Biosystems, Stafford, Texas) utilizando el programa Genescan, versión 3.1.2. Las asignaciones alélicas fueron hechas mediante la comparación de los fragmentos amplificados con la escalera alélica (Applied Biosystems, Warrington, U.K.) que contiene el kit utilizando el programa Genotyper, versión 2.5.2 (Applied Biosystems, Foster City, California).

Biogeographical origin	Population	Sample size	Reference	
NEW WORLD				
	United States	African-American	10 Kayser et al. 2003	
	Bahamas	Abaco	41	Simms et al., 2011
		Eleuthera	30	Simms et al., 2011
		Exuma	30	Simms et al., 2011
		Grand Bahama	33	Simms et al., 2011
		Long Island	31	Simms et al., 2011
		New Providence	79	Simms et al., 2011
	Caribbean islands	Jamaica	53	Torres et al.,2007
		Dominica	21	Torres et al.,2007
		Grenada	35	Torres et al.,2007
		St. Lucia	24	Torres et al.,2007
		St.Kitts	33	Torres et al.,2007
		St.Vincent	21	Torres et al.,2007
		Trinidad	32	Torres et al.,2007
		St.Thomas	133	Torres et al.,2007
		San Andrés	31	This study
		Providencia	26	This study
	Colombia	Afro-Chocó	110	Yunis et al.,2005
		Andes	122	Yunis et al.,2005
Cartagena		173	Builes et al., 2005	
EUROPEAN				
United Kindom	Caucasians	220	Ballard et al., 2005	
	Afro caribbean	241	Ballard et al., 2005	
	South Asians	227	Ballard et al., 2005	
AFRICAN				
western	North western	74	Bosch et el.,2000	
	Biaka	8	Tishkoff et al., 2007	
	Yoruba	12	Tishkoff et al., 2007	
	Guinea Bissau	154	Rosa et al., 2006	
	eastern	Datog	23	Tishkoff et al., 2007
		Burunge	31	Tishkoff et al., 2007
		Mbugwe	14	Tishkoff et al., 2007
		Turu	20	Tishkoff et al., 2007
		Hadza	54	Tishkoff et al., 2007
		Sandawe	67	Tishkoff et al., 2007
		Sukuma	30	Tishkoff et al., 2007

Tabla 1. Poblaciones utilizadas para los análisis de comparación.

2.4 Análisis Estadísticos

Se estimaron frecuencias haplotípicas y diversidad haplotípica, de acuerdo a lo propuesto por Nei (1987), utilizando la ecuación: $D = (n/(n - 1)(1 - \sum p_i^2))$ donde n es el tamaño de la muestra y p_i es la frecuencia del haplotipo, utilizando el programa Arlequín 3.1 (Excoffier et al. 2005). El mismo programa fue utilizado para realizar el análisis molecular de la varianza (AMOVA) bajo un modelo mutacional paso a paso, para cuantificar la proporción de la varianza genética total entre y dentro de las islas.

Se predijeron los posibles haplogrupos de cromosoma Y presentes en cada isla utilizando los haplotipos encontrados con los 17 STRs utilizando el programa Athey's Haplogroup Predictor, versión 5, el cual se basa en una aproximación bayesiana de frecuencias alélicas (Athey, 2006).

Finalmente se realizó un análisis de redes median-joining utilizando Network, version 4.5.1.0 (Bandelt et al. 1999) (<http://www.fluxus-technology.com>). Los pesos de locus específicos fueron asignados de 2 a 10 de acuerdo con Decker et al. (2008) donde a los loci con mayores tasas de mutación se les asignó el menor peso. El locus DYS385a/b fue excluido. Algunos autores sugieren que este debe ser interpretado como un haplotipo aparte (Butler, 2003).

2.5 Comparación con otras poblaciones

Utilizando seis microsatélites: DYS19, DYS389-I, DYS390, DYS391, DYS392 Y DYS393 se realizó el análisis molecular de la varianza (AMOVA) entre las islas y las 34 poblaciones reportadas para evaluar la hipótesis de distribución al azar de los individuos entre pares de poblaciones. En el análisis se evaluaron diferentes agrupamientos con 10.000 permutaciones utilizando Arlequín 3.11 (Excoffier et al. 2005). Adicionalmente se calcularon distancias genéticas por pares de poblaciones utilizando los valores RST, con la cuales posteriormente se realizó un escalamiento multidimensional (MDS) utilizando el programa estadístico SPSS versión 20 (SPSS Inc., Chicago, IL).

2.6 Análisis de apellidos

Tomando el primer apellido de cada uno de los 54 participantes, se estimaron frecuencias y diversidad por apellidos; adicionalmente se registraron el número de haplotipos de cromosoma Y encontrados para cada uno y su correspondiente haplogrupo predicho. Una segunda línea de análisis, consistió en determinar similitudes y diferencias entre haplotipos, asociando a cada haplotipo su correspondiente apellido y lugar de nacimiento, ello se determinó utilizando un Análisis de Correspondencias Múltiples (ACM) mediante el programa SPAD versión 3.5 (2000) y un Análisis de Componentes Principales (APC) mediante el programa MVSP versión 3.0 (Kovach.,1998).

3. Resultados

3.1 Diversidad Genética y haplogrupos

Utilizando 17 sistemas microsatélites, se encontraron 18 haplotipos únicos para la isla de Providencia 26 y para la Isla de San Andrés (Tablas 1 y 2); no se observaron haplotipos compartidos entre ellas. La diversidad haplotípica calculada en las islas omitiendo el sistema DYS385a/b, que al ser un microsatélite complejo, debe ser interpretado como un haplotipo aparte (Butler, 2003); fue para Providencia 0.9488 y para San Andrés 0.9870 (Tablas suplementarias 1 y 2). La diversidad genética promedio para todos los loci fue en Providencia 0.5990 (+/- 0.1287) y en San Andrés 0.7265 (+/- 0.0756). El sistema DYS385, cuando se considera un genotipo, fue el sistema más diverso en el conjunto de marcadores analizados, con una diversidad génica de 0.8673 en Providencia y en San Andrés de 0.9147. Cabe resaltar la baja variabilidad (menor al 50%) de los sistemas DYS389I (0.3457), DYS437 (0.4012), DYS393 (0.4753) encontrados en Providencia.

Se predijeron los haplogrupos de cromosoma Y presentes, en las tablas 2 y 3 se reportan los haplogrupos predichos para las poblaciones de San Andrés y Providencia con el valor de probabilidad en cada caso. Se encontraron para Providencia cuatro haplogrupos, donde el haplogrupo E1b1a fue el más frecuente 47.8 %, seguido del haplogrupo R1b 39.1%; los otros dos haplogrupos encontrados fueron R1a y E1b1b los cuales representaron el 8.69% y 4.34%, respectivamente. Para San Andrés se

encontraron seis haplogrupos; donde el haplogrupo más frecuente también fue el E1b1a 48.4%, seguido del haplogrupo R1b 35.5%. Los haplogrupos J2alh, I2b1, I1 y E1b1b representaron el 6.5%, 3.2%, 3.2% y 3.2%, respectivamente. El valor de probabilidad fue mayor al 80%, salvo en el haplotipo H017 de la isla de providencia con una probabilidad igual a 47.9%.

Providencia Island																			
HAPLOTYPE	N	DYS19	DYS19a/b	DYS389	DYS389b	DYS390	DYS391	DYS392	DYS393	DYS417	DYS426	DYS439	DYS448	DYS456	DYS458	DYS635	Y-GATA-10a	Predicted haplogroup	Probability(%)
H001	1	13	11-11	12	28	25	12	13	13	15	12	12	19	15	17	25	13	R1b	100
H002	3	14	11-14	12	28	25	12	13	13	15	12	12	19	15	17	24	12	R1b	100
H003	1	14	11-14	12	28	25	11	13	13	15	12	12	19	15	17	24	12	R1b	100
H004	1	17	17-18	13	30	21	11	11	14	14	11	12	21	17	16	22	11	E1b1a	100
H005	1	17	17-19	13	31	21	10	11	14	14	11	13	22	16	16	21	11	E1b1a	100
H006	1	15	11-14	13	29	25	10	11	13	14	11	11	20	17	16	23	12	R1a	100
H007	3	16	15-19	13	30	21	10	11	14	14	11	12	21	15	19	20	13	E1b1a	100
H008	2	16	15-18	13	30	21	10	11	14	14	11	12	21	15	19	20	11	E1b1a	100
H009	1	16	15-19	12	30	21	10	11	14	14	11	12	21	15	19	20	11	E1b1a	100
H010	1	15	16-17	13	31	21	11	13	13	14	11	11	21	16	16	21	11	E1b1a	100
H011	1	14	12-14	13	29	24	10	13	13	15	12	11	19	15	17	23	12	R1b	100
H012	1	16	15-15	13	30	22	11	11	14	14	11	12	21	17	16	22	11	E1b1a	99.4
H013	1	13	16-16	13	30	24	10	11	13	14	10	11	20	17	16	21	12	E1b1b	100
H014	1	14	11-15	13	29	24	11	13	13	15	12	12	19	16	18	23	11	R1b	100
H015	1	14	11-14	13	29	24	10	13	13	14	10	13	19	15	16	23	13	R1b	99
H016	1	14	11-14	13	29	24	10	13	13	14	10	13	19	15	17	23	13	R1b	99
H017	1	15	11-14	13	29	25	10	11	13	14	11	12	20	17	16	23	12	R1a	47.9
H018	1	17	17-19	13	30	21	10	11	14	14	10	12	21	14	17	21	13	E1b1a	100
Genetic diversity		0.772	0.867284	0.34568	0.69753	0.69136	0.537	0.4938	0.4753	0.4012	0.6235	0.5494	0.4667	0.642	0.6605	0.787	0.6419753		

Table 2. Haplogrupos predichos de cromosoma Y y sus valores de probabilidad en la isla de providencia (N=23)

San Andrés Island																	Predicted haplogroup	Probability(%)	
HAPLOTYPE	N	DYS19	DYS385a/b	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS437	DYS438	DYS439	DYS448	DYS456	DYS458	DYS635			Y-GATA-H4
H001	1	16	16-16	13	31	21	10	11	13	14	13	12	21	15	18	21	13	E1b1a	100
H002	2	14	11-15	13	29	24	10	13	13	15	12	13	19	16	17	23	12	R1b	100
H003	1	14	11-15	13	29	24	10	13	13	15	12	10	19	16	17	23	12	R1b	100
H004	1	17	15-16	13	29	23	10	12	15	15	10	12	20	17	16	20	11	I2b1	100
H005	1	14	11-14	14	30	24	11	13	13	15	12	12	18	16	18	23	12	R1b	100
H006	2	13	13-16	13	29	23	10	9	12	15	9	12	20	15	16	22	12	J2a1h	87.6
H007	1	16	14-20	14	32	21	11	11	14	14	11	12	21	15	16	21	12	E1b1a	100
H008	1	14	11-13	12	28	23	12	9	13	15	12	12	19	16	17	24	12	R1b	100
H009	1	16	17-18	13	30	21	10	13	14	14	11	10	20	16	16	24	11	E1b1a	100
H010	1	14	11-14	13	30	24	11	11	13	15	12	10	18	16	16	23	13	R1b	86.3
H011	1	14	11-14	13	30	24	11	13	13	15	12	10	18	16	16	23	13	R1b	100
H012	1	14	15-15	12	29	22	10	11	13	17	10	12	20	15	14	19	11	I1	100
H013	2	14	11-14	13	29	23	10	13	12	15	12	12	19	15	18	23	11	R1b	100
H014	1	15	16-18	13	31	21	10	11	13	14	11	12	21	18	16	21	11	E1b1a	100
H015	1	16	16-16	13	31	21	10	11	13	14	12	11	21	15	16	21	12	E1b1a	100
H016	1	17	16-16	13	31	21	10	11	13	14	12	11	21	15	16	21	12	E1b1a	100
H017	3	17	17-17	13	30	21	10	11	14	14	11	13	21	16	16	22	11	E1b1a	100
H018	1	13	17-17	13	30	21	10	11	14	14	11	10	21	16	16	22	11	E1b1a	81.6
H019	1	14	11-16	13	29	24	11	13	13	15	12	12	19	16	18	23	11	R1b	100
H020	1	15	16-16	12	30	21	10	11	13	14	11	12	21	15	17	20	12	E1b1a	100
H021	1	16	17-18	13	30	21	10	11	15	14	11	13	21	16	16	21	11	E1b1a	100
H022	1	15	16-16	12	30	21	10	13	13	14	11	13	21	18	18	20	13	E1b1a	100
H023	1	15	16-16	12	30	21	10	11	13	14	12	12	21	15	17	21	12	E1b1a	100
H024	1	14	11-14	12	28	23	10	17	13	15	12	10	19	18	15	23	11	R1b	100
H025	1	13	13-13	13	31	21	10	9	13	14	11	12	21	18	18	23	12	E1b1b	99.6
H026	1	15	16-17	13	31	21	10	11	13	14	11	12	21	16	18	23	12	E1b1a	100
Genetic diversity	0.825	0.9146722	0.62123	0.80229	0.73153	0.5567	0.7461	0.6379	0.6681	0.7399	0.7378	0.7607	0.7482	0.7711	0.821	0.7294485			

Table 3. Haplogrupos predichos de cromosoma Y y sus valores de probabilidad en la isla de San Andrés (N=31)

3.2 Análisis de redes

Con el propósito de tener un mejor conocimiento de las bases moleculares de la variabilidad encontrada en las islas, se construyó una red median-joining con todos los haplotipos pertenecientes a los dos haplogrupos más frecuentes en las islas (E1b1a y R1b), utilizando todos los Y-STRs salvo el locus *DYS385a/b*. Mediante este análisis solo se observó un haplotipo compartido entre las islas en el haplogrupo R1b (Figura 3). Dentro de las redes fueron también postulados varios vectores propios (haplotipos no muestreados ó haplotipos extintos); 10 en el haplogrupo E1b1a y 6 en el Haplogrupo R1b. Los demás haplotipos se encuentran en su mayoría dispersos, esto como consecuencia de la diversidad genética encontrada en las islas (figuras 2 y 3).

3.4 Comparación con otras poblaciones

En cuanto al porcentaje de variabilidad observado entre 10 regiones (El primer grupo compuesto por Providencia-Colombia, San Andrés-Colombia, Grenada, St. Lucia, St. Kitts, St. Vincent, Trinidad, St. Thomas, Jamaica, Afro Chocó-Colombia, Afro caribbean-UK, Eastern Africa-Sukuma, Eastern Africa-Sandawe, Eastern Africa-Turu, Eastern Africa-Mbugwe, Eastern Africa-Burunge, Eleuthera-Bahamas, Long Island-Bahamas, Exuma-Bahamas, Grand Bahama-Bahamas, New Providence-Bahamas, Abaco-Bahamas; el segundo grupo compuesto por Andes-Colombia, Cartagena-Colombia, Dominica y Caucasians UK; el tercer grupo por la población de South Asians; el cuarto grupo compuesto por Eastern Africa-Datog; el quinto grupo por: Eastern Africa-Hadza; el sexto por la población: Western Africa-Yoruba; el séptimo por Western Africa-Biaka; el octavo por: Western Africa- -Guinea Bissau; el noveno compuesto por African-American y el décimo grupo por North Western Africa); mostró que la mayoría de la variación se encuentra dentro de los grupos poblacionales, pero hay un porcentaje significativo de la variación 27,31% ($p < 0.000$, 10000 permutaciones) que puede ser explicado por diferencias entre grupos de poblaciones.

Para evaluar las diferencias poblacionales se calcularon distancias genéticas RST por pares de poblaciones y se graficaron mediante un escalamiento multidimensional (Figura 4). Mediante este análisis se observan dos conjuntos de poblaciones, principalmente. El primero, en donde se encuentran la mayoría de poblaciones africanas y con ancestría africana, grupo que corresponde con las regiones uno utilizada en el AMOVA. En este grupo se encuentran San Andrés y Providencia más cercanos a las islas del Caribe y a las poblaciones del este de África: Burungé, Mbugue y Turu. En el segundo conjunto se reúnen poblaciones de otros orígenes, Europeo, Andes y Cartagena, entre otros.

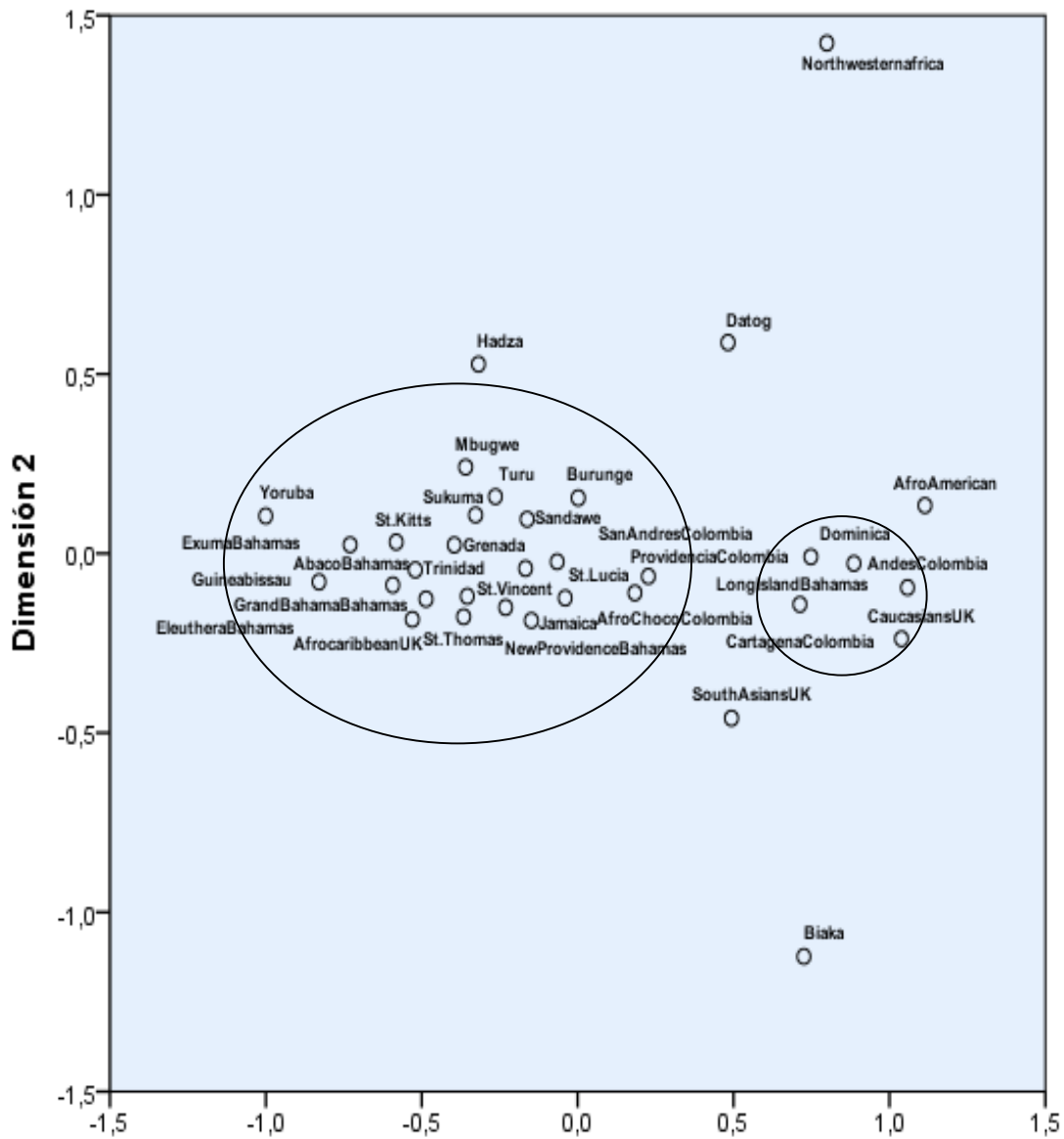


Figura 4. Escalamiento multidimensional basado en la matriz de distancia RST por pares de poblaciones. El valor de stress fue 0.0154.

Apellidos.

3.5 Diversidad y número de apellidos

A partir del primer apellido de cada uno de los participantes se encontraron para la isla de Providencia 11 apellidos con una diversidad de 0,4783, siendo los apellidos más frecuentes Bryan (n=7) y Archbold (n=5) representando el 52,17% del total de apellidos para la isla. Al tomar cada uno de estos apellidos, y comparar sus haplotipos y haplogrupos predichos se encontró que 3 de 4 haplotipos Bryan pertenecen al haplogrupo E1b1a y solo uno al haplogrupo R1a (Tabla 4), en el caso del apellido Archbold los 3 haplotipos de cromosoma Y pertenecen al mismo haplogrupo.

Para la isla de San Andrés se encontraron 18 apellidos, con una diversidad de 0,5806, siendo Martínez (N=4), Bent (N=3) y Pomare (N=3) los apellidos más frecuentes con el 32,26 % del total de los apellidos, presentado los haplogrupos E1B1a y R1b (Tabla 5).

Como se observa en las tablas 4 y 5, el único apellido compartido en ambas islas fue Bent (N=1 en Providencia y N=3 San Andrés), aunque con diferentes haplotipos y haplogrupos: E1B1a en Providencia y R1a en San Andrés.

Al evaluar el número de haplotipos presentes en cada apellido se encontró que para Bryan el 57% comparte el mismo haplotipo, para Archbold el 60% en el caso de Providencia, y para San Andrés, Martínez y Bent con 50 y 66 respectivamente.

Al realizar el Análisis de Correspondencias Múltiples (ACM) y el Análisis de Componentes Principales (PCA), tomando los 17 marcadores STR del cromosoma Y para cada individuo, se determinaron diferencias y semejanzas entre cada uno de los haplotipos, ubicándose cercanos aquellos haplotipos iguales o muy similares en un espacio geográfico. Como cada haplotipo fue asociado con su correspondiente apellido, esto permitió determinar cuáles individuos comparten la información genética en el cromosoma Y, y la información genealógica de los apellidos que tienen esos individuos (Figuras 5 y 6). Se encontraron dos grupos principales de apellidos, que corresponden a los haplogrupos predichos R1b y E1b1a .

Providencia Island

nSurname s	N	Frequency (%)	N° haplotypes	Predicted haplogroup
BRYAN	7	30,43	4	E1b1a/R1a
ARCHBOLD	5	21,74	3	R1b
NEWBALL	2	8,70	2	R1b
ROBINSON	2	8,70	2	R1a/E1b1a
BENT	1	4,35	1	E1b1a

Tabla 4. Frecuencia, número de haplotipos y haplogrupo predicho de los apellidos más frecuentes en la Isla de Providencia. Los valores de probabilidad fueron superiores al 80% salvo en el apellido ROBINSON donde para el haplogrupo R1a el valor encontrado fue de 47.9%.

San Andrés Island

Surnames	N	Frequency (%)	N° haplotypes	Predicted haplogroup
MARTÍNEZ	4	12,90	2	E1b1a
BENT	3	9,68	2	R1a
POMARE	3	9,68	3	E1b1a
JAMES	2	6,45	2	R1b
LIVINGSTONE	2	6,45	2	R1b
MANUEL	2	6,45	2	E1b1a
WALTERS	2	6,45	2	E1b1b-E1b1a

Tabla 5. Frecuencia, número de haplotipos y haplogrupo predicho de los apellidos más frecuentes en la Isla de San Andrés. Los valores de probabilidad en todos los casos fueron superiores al 80%.

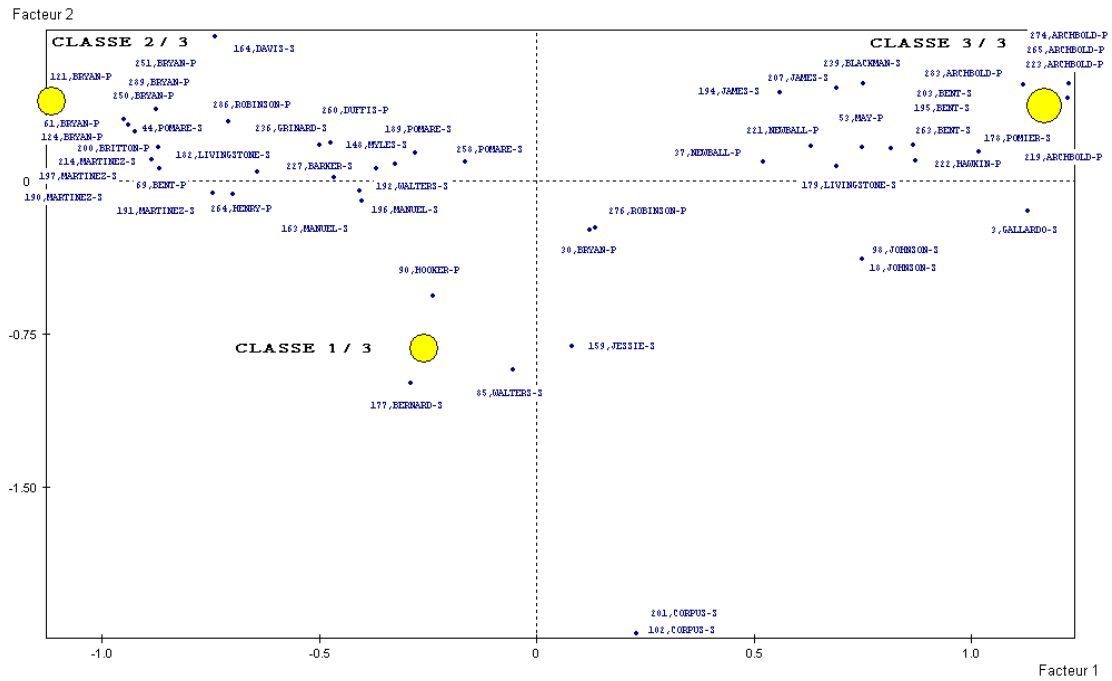


Figura 5. Análisis de Correspondencias Múltiples de apellidos y de datos genéticos para 17 marcadores STR del cromosoma Y en el Archipiélago de San Andrés y Providencia.

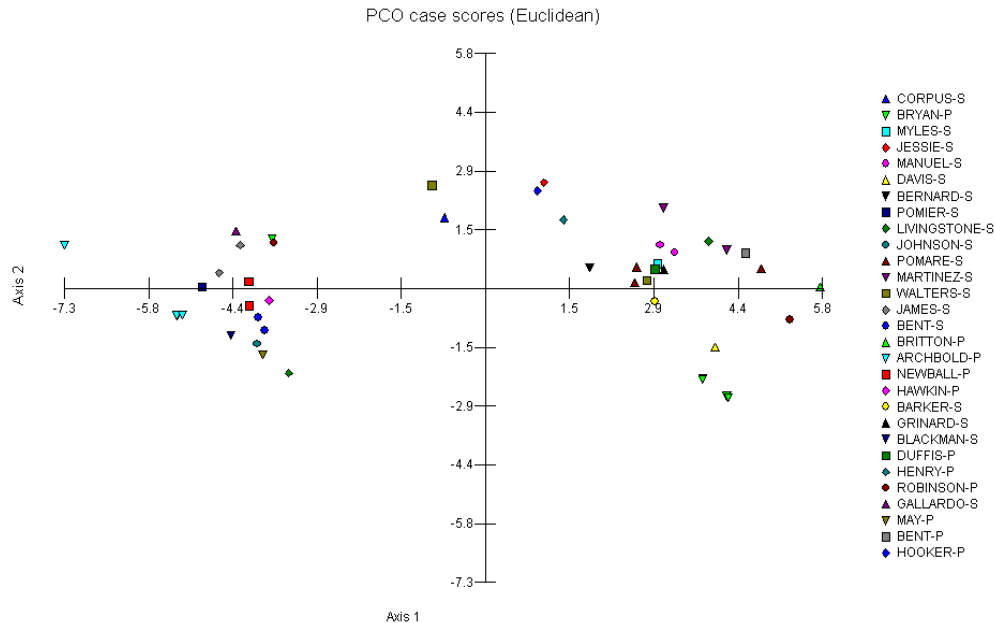


Figura 6. Análisis de Componentes Principales de apellidos y de datos genéticos para 17 marcadores STR del cromosoma Y en el Archipiélago de San Andrés y Providencia.

4.0 Discusión

En el curso de la evolución, las poblaciones varían en el tiempo y en el espacio, pueden mantenerse aisladas durante uno o varios periodos de tiempo, bien sea por la existencia de barreras geográficas, ecológicas o culturales; o por otra parte, tienen la posibilidad de intercambiar genes, dando lugar a poblaciones que presentan mezcla de algunas características y rasgos genéticos (Cavalli-Sforza y Bodmer, 1971). Este contexto no es ajeno a las poblaciones humanas, mucho menos en aquellas que han sido producto de escenarios demográficos tan drásticos como las poblaciones afro descendientes en el mundo, pero principalmente en el continente Americano.

Para la elaboración del mapa cultural del Caribe se partió del reconocimiento de que el proceso de formación de las distintas identidades culturales de la región caribeña está directamente relacionado con la historia de su poblamiento: la historia de los encuentros y desencuentros entre los diferentes

grupos humanos que fueron habitando este territorio (Vollmer 1997). Desde esta óptica, la genética contribuye al conocimiento de la historia de poblamiento a partir de la información contenida en su ADN ligada de manera indistinguible a los procesos culturales que caracterizan determinada población. Adicionalmente pocos son los estudios de cromosoma Y que se han realizado en las poblaciones del Caribe (Torres et al. 2007).

En este estudio se realizó un análisis profundo de la diversidad de Cromosoma Y en las poblaciones raizales de San Andrés y Providencia mediante la utilización de un criterio genealógico, que permitió analizar individuos que tienen por lo menos tres generaciones de ancestros en las islas para 17 marcadores microsatélites y la asociación del primer apellido a esta información genética. De esta manera, se evaluó el nivel de variabilidad y la subestructura genética de estas poblaciones, así como el establecimiento de la contribución genética de las poblaciones africanas y/o afrodescendientes o poblaciones europeas en las islas.

Se encontró que los valores de diversidad haplotípica son bajos (Providencia: 0.9488, San Andrés: 0.9870), con respecto a los ya publicados para otras poblaciones afrodescendientes; por ejemplo, para la población Afro colombiana de Chocó se ha reportado una diversidad haplotípica de 0.9955 (Yunis et al. 2005) mientras que para poblaciones del Caribe anglo parlantes se ha reportado: Dominica un valor de 0.9714, Grenada 0.9937, St. Lucía 0.9964, St. Kitts 0.9886, Trinidad 0.9960, St. Thomas 0.9973 y Jamaica 0.9954 (Torres et al. 2007). La reducida diversidad genética observada en Providencia puede ser el resultado del muestreo de una comunidad pequeña con ancestría genética similar o como resultado de un tamaño efectivo poblacional en la isla, donde los loci del cromosoma Y son susceptibles a fenómenos de deriva, como un cuello de botella reciente (Giroti y Talwar 2010; Torres et al. 2007; De Knijff 2000).

En la isla de Providencia cuatro haplogrupos fueron predichos siendo el haplogrupo E1b1a el más frecuente 47.8 % (N=11), seguido del haplogrupo R1b 39.1% (N=9). Para la isla de San Andrés seis haplogrupos fueron predichos; siendo también el E1b1a 48.4% (N=15), seguido del haplogrupo R1b 35.5% (N=11) los haplogrupos más frecuentes. Este haplogrupo E1b1a alcanza frecuencias de más del 80% en muchas partes de África Occidental, África Central, África Oriental y África del Sur

(Trombetta B et al. 2011) y es el menos común en las poblaciones del Cáucaso (Hammer et al. 2006; Y chromosome Consortium 2002; Quintana-Murci et al. 1999). Por otra parte, más del 50% de los hombres en Europa están asociados con el haplogrupo R (Sims et al. 2007; Jobling & Tyler-Smith 2003) que a su vez se encuentran distribuidos en el occidente, centro y sur de Asia (Chiaroni J et al. 2009).

Otros estudios indican que la mayoría de Africanos americanos y caucásicos en los Estados Unidos pertenecen a uno de los dos principales sub-haplogrupos, E3a (58-62%) y R1b (47-58,3%), respectivamente (Sims et al. 2007; Hammer et al. 2006; Vallone and Butler 2004). En nuestros resultados se observa este mismo fenómeno de conservar ese pasado Africano y Europeo que se dio como producto de las diferentes procesos de colonización y esclavitud entre 1730-1853, caracterizado por la llegada de gente del Caribe anglófono: Jamaica, Barbados, y Trinidad y Tobago, otras llegaron directamente de las islas británicas, en particular de Escocia e Irlanda; y otras, directamente de África occidental. Antes de este periodo las islas permanecieron casi abandonadas (Vollmer 1997).

El AMOVA entre las dos islas utilizando 15 Y-STR's no mostró diferencia significativa $p=0.1741$, a pesar de que entre las dos poblaciones sólo se comparte un haplotipo, sus distribuciones de frecuencias entre sistemas no son muy diferentes y no reflejan procesos de subestructura poblacional, lo cual pudo observarse en la red para cada uno de los haplogrupos donde se encuentran haplotipos de ambos orígenes (Figuras 2 y 3).

Por su parte, el AMOVA expuso la presencia de 10 regiones con un porcentaje de variación de 27,31% ($p<0.000$), cuya representación se observa en la figura 4, donde se ve que las poblaciones afro-descendientes, agrupadas en la región 1 del AMOVA, se ubican en un espacio genético similar respecto a las demás poblaciones; a excepción de la población de Dominica quien se agrupó con las poblaciones de Cartagena, Mestiza Caucásica colombiana y Caucásica del Reino Unido. Esta población en un estudio realizado por Torres et al. (2007) presentó la menor frecuencia de la inserción ALU DYS 287 que define el haplogrupo E africano y se piensa puede tener una mayor contribución genética de otras poblaciones no africanas o afrodescendientes, lo cual podría explicar este agrupamiento.

Los análisis de apellidos evidencian una menor diversidad de apellidos en Providencia en relación a San Andrés, 47% y 58% respectivamente. Adicionalmente se observó en ambas islas, que los raizales que tienen el mismo apellido comparten además el haplotipo en un porcentaje mayor al 50%, lo cual implica que aunque en los criterios de inclusión de los participantes se tuvo en cuenta en base a su genealogía que no compartieran ancestros, hay una alta proporción de personas que comparten el mismo apellido y el mismo haplotipo, es decir en algún momento en su historia familiar tuvieron un ancestro en común. Con base en este mismo criterio los análisis multivariados realizados permitieron observar aquellos apellidos que más se parecen en base a su información genética. Todos los apellidos analizados se encuentran dentro del listado realizado nombres de las familias de mayor arraigo y de vigencia actual en las islas (Francis C., 1991)

5.0 Conclusión

El archipiélago de San Andrés y Providencia presenta en su mayoría un componente ancestral Africano y Europeo, característico de los diferentes ciclos de poblamiento del Caribe insular. Adicionalmente no se encuentran diferencias significativas entre San Andrés y Providencia. Ambas poblaciones se ubican cercanas a poblaciones africanas y afro descendientes. Son similares a otras poblaciones del caribe insular, a otras poblaciones del este de África y no se diferencian de otras poblaciones colombianas afro-descendientes como la del departamento del Chocó. En la reconstrucción de la historia filogenética de las poblaciones interactúan diferentes disciplinas como la antropología, arqueología y lingüística, en donde se obtienen diferentes puntos de vista. Este estudio no pretende cuestionar las características culturales que identifican a cada una de ellas.

Agradecimientos

En primer lugar a los habitantes de la población de San Andrés y Providencia por permitirnos conocer un poco más acerca de su historia biológica, así como también a las diferentes instituciones de salud: CAPRECOM, BIOLAB, OMALINA OWKIN DE GONZÁLEZ y el Hospital Providencia, y a la Bióloga Natalia Lamprea quienes colaboraron en el muestreo. En segundo lugar, a los miembros del Grupo de Genética de Poblaciones e Identificación, a la Universidad Nacional de Colombia sede Bogotá, al Instituto de Genética y a la sede Caribe. Finalmente a Colciencias con su programa Jóvenes Investigadores e Innovadores "Virginia Gutiérrez de Pineda" año 2010 por su apoyo en la realización de este estudio.

Bibliografía

- Athey, T. W. 2006. Haplogroup prediction from Y-STR values using a Bayesian–allele frequency approach. *J. Genet. Geneal.* 2:34–39.
- Ballard, D. J., Phillips, C., Thacker, C. R., Robson, C., Revoir, a P., & Syndercombe Court, D. 2005. Y chromosome STR haplotypes in three UK populations. *Forensic Sci. Int. Genet.* 152(2-3):289-305.
- Bandelt, H. J., Forster, P., & Röhl, a. 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* 16(1): 37-48..
- Builes, J. J., Martínez, B., Gómez, A., Caraballo, L., Espinal, C., Aguirre, D., et al. 2007. Y chromosome STR haplotypes in the Caribbean city of Cartagena Colombia. *Forensic Sci. Int. Genet.* 167(1):62-69.
- Butler, J M. 2003. Recent Developments in Y-Short Tandem Repeat and Y-Single Nucleotide Polymorphism Analysis Recent Developments in Y-Short Tandem Repeat and Y- Single Nucleotide Polymorphism Analysis a. *Forensic Science Review.*15: 91-111.

- Cavalli-Sforza, L and Bodmer, W. 1971. *The Genetics of Human Populations*. W. H. Freeman, San Francisco (reprinted 1999 by Dover Publications).
- Chiaroni, J., Underhill, P. a, & Cavalli-Sforza, L. L. 2009. Y chromosome diversity, human expansion, drift, and cultural evolution. *Proc. Natl. Acad. Sci. U.S.A.* 106(48):20174-20179.
- Consortium, T. Y. C. 2002. A Nomenclature System for the Tree of Human Y-Chromosomal Binary Haplogroups. *Genome Res.* 12(2):339-348.
- Decker, A. E., M. C. Kline, J. W. Redman et al. 2008. Analysis of mutations in father-son pairs with 17 Y-STR loci. *Forensic Sci. Int. Genet.* 2:e31–e35.
- De Knijff, P. 2000. Messages through bottlenecks: On the combined use of slow and fast evolving polymorphic markers on the human Y chromosome. *Am. J. Hum. Genet.* 67(5):1055–1061.
- Excoffier, L., Laval, G., & Schneider, S. 2005. Arlequin version 3.0: an integrated software package for population genetics data analysis. *Evol. Bioinform. Online.* 1:47-50.
- Francis, C. 1991. *Compendio de cultura popular tradicional de las islas de San Andrés y Providencia*. Ediciones archipiélago. San Andrés Isla, Colombia.
- Giroti, R., & Talwar, I. 2010. The most ancient democracy in the world is a genetic isolate: an autosomal and Y-chromosome study of the hermit village of Malana Himachal Pradesh, India. *Hum. Biol.* 82(2):123-141.
- Hammer, M. F., Chamberlain, V. F., Kearney, V. F., Stover, D., Zhang, G., Karafet, T., et al. 2005. Population structure of Y chromosome SNP haplogroups in the United States and forensic implications for constructing Y chromosome STR databases. *Forensic Sci. Int. Genet.* 164(1):45-55.

- Jobling, M a. 2001. Y-chromosomal SNP haplotype diversity in forensic analysis. *Forensic Sci. Int. Genet.* 118(2-3):158-162.
- Jobling, Mark a, & Tyler-Smith, C. 2003. The human Y chromosome: an evolutionary marker comes of age. *Nat. Rev. Genet.* 48: 598-612.
- King T.E. & Jobling M.A. 2009. What's in a name? Y chromosomes, surnames and the genetic genealogy revolution. *Trends Genet.*, 25: 351-360.
- Kovach, W.L., 1998. *MVSP - A Multivariate Statistical Package for Windows, ver. 3.0*. Kovach Computing Services, Pentraeth, Wales, U.K.
- Quintana-Murci, L., Semino, O., Minch, E., Passarimo, G., Brega, a, & Santachiara-Benerecetti, a S. 1999. Further characteristics of proto-European y chromosomes. *Eur. J. Hum. Genet.* 75: 603-608.
- Lamprea, N.2009. Caracterización genética de la población humana de San Andrés y Providencia a partir de los marcadores microsatélites (STR'S) empleados por el combined dna index system (CODIS). Tesis de Maestria, Departamento de Biología, facultad de ciencias. Universidad Nacional de Colombia. Bogotá, Colombia.
- Nei, M. 1987. *Molecular Evolutionary Genetics*. Columbia University Press, New York, USA.
- Sykes B., & Irven C. 2000 Surnames the Y chromosome. *American Journal of Human Genetics.* 66(4):1417–1419.
- Sims, L. M., Garvey, D., & Ballantyne, J. 2007. Sub-populations within the major European and African derived haplogroups R1b3 and E3a are differentiated by previously phylogenetically undefined Y-SNPs. *Hum. Mutat.* 28(1): 97.

- Torres, J., Kittles, R. a, & Stone, a C. 2007. Mitochondrial and Y chromosome diversity in the English-speaking Caribbean. *Ann. Hum. Genet.* 71(Pt 6):782-90
- Trombetta, B., Cruciani, F., Sellitto, D., & Scozzari, R. 2011. A new topology of the human Y chromosome haplogroup E1b1 E-P2 revealed through the use of newly characterized binary polymorphisms. *PLoS ONE.* 6(1): e16073.
- Vallone, P. M., & Butler, John M. 2004. Y-SNP typing of U.S. African American and Caucasian samples using allele-specific hybridization and primer extension. *Journal of forensic sciences,* 49. 723-732.
- Vollmer, L. 1997. *La historia del poblamiento de San Andrés, Vieja Providencia y Santa Catalina.* Ediciones archipiélago. San Andrés Isla, Colombia.
- Yunis, J. J., Acevedo, L. E., Campo, D. S., & Yunis, E. J. 2005. Population data of Y-STR minimal haplotypes in a sample of Caucasian-Mestizo and African descent individuals of Colombia. *Forensic Sci. Int. Genet.* 151:307-313.

Tabla suplementaria 1. Frecuencias haplotípicas de 15 Y-STR para los individuos raizales de la isla de Providencia (N=23).

Providencia Island																	
HAPLOTYPE	N	DYS19	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS437	DYS438	DYS439	DYS448	DYS456	DYS458	DYS635	Y-GATA-H4	Frequency
H001	1	13	12	28	25	12	13	13	15	12	12	19	15	17	25	13	0.0435
H002	3	14	12	28	25	12	13	13	15	12	12	19	15	17	24	12	0.13
H003	1	14	12	26	25	11	13	13	15	12	12	19	15	17	24	12	0.0435
H004	1	17	13	30	21	11	11	14	14	11	12	21	17	16	22	11	0.0435
H005	1	17	13	31	21	10	11	14	14	11	13	22	16	16	21	11	0.0435
H006	1	15	13	29	25	10	11	13	14	11	11	20	17	16	23	12	0.0435
H007	5	16	13	30	21	10	11	14	14	11	12	21	15	19	20	11	0.217
H008	1	16	12	30	21	10	11	14	14	11	12	21	15	19	20	11	0.0435
H009	1	15	13	31	21	11	13	13	14	11	11	21	16	16	21	11	0.0435
H010	1	14	13	29	24	10	13	13	15	12	11	19	15	17	23	12	0.0435
H011	1	16	13	30	22	11	11	14	14	11	12	21	17	16	22	11	0.0435
H012	1	13	13	30	24	10	11	13	14	10	11	20	17	16	21	12	0.0435
H013	1	14	13	29	24	11	13	13	15	12	12	19	16	18	23	11	0.0435
H014	1	14	13	29	24	10	13	13	14	10	13	19	15	16	23	13	0.0435
H015	1	14	13	29	24	10	13	13	14	10	13	19	15	17	23	13	0.0435
H016	1	15	13	29	25	10	11	13	14	11	12	20	17	16	23	12	0.0435
H017	1	17	13	30	21	10	11	14	14	10	12	21	14	17	21	13	0.0435

Tabla suplementaria 2. Frecuencias haplotípicas de 15 Y-STR para los individuos raizales de la isla de San Andrés (N=31).

San Andrés Island																	
HAPLOTYPE	N	DYS19	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS437	DYS438	DYS439	DYS448	DYS456	DYS458	DYS635	Y-GATA-H4	Frequency
H001	1	16	13	31	21	10	11	13	14	13	12	21	15	18	21	13	0.0323
H002	2	14	13	29	24	10	13	13	15	12	13	19	16	17	23	12	0.0645
H003	1	14	13	29	24	10	13	13	15	12	10	19	16	17	23	12	0.0323
H004	1	17	13	29	23	10	12	15	15	10	12	20	17	16	20	11	0.0323
H005	1	14	14	30	24	11	13	13	15	12	12	18	16	18	23	12	0.0323
H006	2	13	13	29	23	10	9	12	15	9	12	20	15	16	22	12	0.0645
H007	1	16	14	32	21	11	11	14	14	11	12	21	15	16	21	12	0.0323
H008	1	14	12	28	23	12	9	13	15	12	12	19	16	17	24	12	0.0323
H009	1	16	13	30	21	10	13	14	14	11	10	20	16	16	24	11	0.0323
H010	1	14	13	30	24	11	11	13	15	12	10	18	16	16	23	13	0.0323
H011	1	14	13	30	24	11	13	13	15	12	10	18	16	16	23	13	0.0323
H012	1	14	12	29	22	10	11	13	17	10	12	20	15	14	19	11	0.0323
H013	2	14	13	29	23	10	13	12	15	12	12	19	15	18	23	11	0.0645
H014	1	15	13	31	21	10	11	13	14	11	12	21	18	16	21	11	0.0323
H015	1	16	13	31	21	10	11	13	14	12	11	21	15	16	21	12	0.0323
H016	1	17	13	31	21	10	11	13	14	12	11	21	15	16	21	12	0.0323
H017	3	17	13	30	21	10	11	14	14	11	13	21	16	16	22	11	0.0968
H018	1	13	13	30	21	10	11	14	14	11	10	21	16	16	22	11	0.0323
H019	1	14	13	29	24	11	13	13	15	12	12	19	16	18	23	11	0.0323
H020	1	15	12	30	21	10	11	13	14	11	12	21	15	17	20	12	0.0323
H021	1	16	13	30	21	10	11	15	14	11	13	21	16	16	21	11	0.0323
H022	1	15	12	30	21	10	13	13	14	11	13	21	18	18	20	13	0.0323
H023	1	15	12	30	21	10	11	13	14	12	12	21	15	17	21	12	0.0323
H024	1	14	12	28	23	10	17	13	15	12	10	19	18	15	23	11	0.0323
H025	1	13	13	31	21	10	9	13	14	11	12	21	18	18	23	12	0.0323
H026	1	15	13	31	21	10	11	13	14	11	12	21	16	18	23	12	0.0323

Capítulo No. 5: Cinco propuestas de unificación entre genética de poblaciones y genética forense para la próxima década.

William Usaquén¹, Andrea Casas¹, Madelyn Rojas¹, Döbereiner Chala¹, Ángela Alonso¹, Jenny Blanco¹, Verónica Rocha¹, Jenny Barrera², Julieth Castiblanco², Luís Fernando García².

¹ Instituto de Genética Universidad Nacional de Colombia, ² Departamento de Biología. Universidad Nacional de Colombia.

Resumen.

Colombia es uno de los países más activos en el desarrollo de análisis genéticos de filiación debido a la naturaleza de nuestra constitución, que proclama los derechos de los menores sobre los derechos de los adultos y en la que se establece la asignación de una personalidad jurídica como uno de los principios fundamentales. Por otra parte, nuestra constante guerra civil, en la que se presentan diferentes actores del conflicto, ha dejado aproximadamente 35.000 desaparecidos en las dos últimas décadas. Según establecen las leyes colombianas, la identificación de las personas fallecidas durante el conflicto armado sería de prioridad nacional, como se expresa en la ley 1408 de 2010, por la cual se rinde homenaje a las víctimas del delito de desaparición forzada y se dictan medidas para su localización e identificación. Esto ha generado un desarrollo tecnológico a nivel nacional, que es considerablemente superior respecto a la región andina. En total, en el país existen aproximadamente 25 analizadores genéticos, que tienen funciones relacionadas con identificación y filiación humana, es un número alto, lo que lleva a pensar sobre el volumen de tipificaciones realizadas en el país, y el subsiguiente manejo de las mismas. Un cálculo modesto, es de aproximadamente trescientas mil tipificaciones en los últimos 15 años, sin embargo, a pesar del desarrollo tecnológico y la implementación de nuevos y cada vez más sofisticados marcadores genéticos, no se ha seguido el mismo desarrollo teórico. En los diferentes artículos científicos producidos, ha sido sobresaliente la ausencia de métodos de análisis robustos a nivel poblacional. Otro factor preocupante es la ausencia de programas de formación y capacitación equilibrados entre biología molecular, estadística, informática y

genética de poblaciones. En este trabajo, se realizara primero una breve revisión de nuestro pasado reciente, posteriormente se realizará una propuesta en 5 puntos que deberían ser fundamentales para la unificación de los estudios en genética de poblaciones y en genética forense para el país.

Keywords.

STRs, Paternity, Colombia, Data Bases, genetic Demography.

1. El origen de los estudios en Colombia.

Nadie hubiera imaginado que la ley 75 de 1968, con la cual se da estructura orgánica sobre el derecho familiar y el menor, creando el Instituto Colombiano de Bienestar Familiar (ICBF), tendría un papel determinante en la genética colombiana. Inicialmente, en esta ley se establecen las normas de protección a los menores de edad, indicando la manera como deberían proceder las autoridades frente a las diferentes situaciones sociales, y el abordaje legal conducentes a la asignación de la paternidad. Aunque los primeros casos de filiación fueron realizados desde 1972 por el ICBF, es hasta con el decreto 2388, donde se establece la práctica de los exámenes antropoheredobiológicos, por el Laboratorio de genética del ICBF, a solicitud de la autoridad competente. Estos exámenes se realizaban en forma simultánea a las personas involucradas en el proceso. Con el avance de la tecnología, se fue pasando de las pruebas antropoheredobiologicas, que implicaban medidas morfométricas, a los grupos sanguíneos; lo que para el año 1972, era un gran avance. Durante dos décadas se fueron realizando nuevos desarrollos hasta llegar a los marcadores moleculares actuales. Pasaremos a revisar, que pasa hoy en día con esta problemática, pero antes debemos revisar más en detalle algunos elementos históricos.

1.1. Semblanza Histórica.

El primer estudio genético poblacional realizado en el país, con el objetivo de determinar los principales grupos poblacionales, fue el desarrollado por el profesor Emilio Yunis Turbay en los años

70, a partir de los datos de las pruebas de filiación del Instituto Colombiano de Bienestar Familiar. En este trabajo se evidenció un primer intento por evaluar la distribución de frecuencias de marcadores eritrocitarios en los diferentes departamentos. Este estudio se realizó con muestras de aproximadamente 70,000 personas, lo que implica un enorme esfuerzo de análisis, organización y colección de datos (Sandoval, 1982; López, 1993). Para ese momento, no se habían implementado en el país las herramientas informáticas que permitieran un adecuado almacenamiento de datos perdurables en el tiempo, ni tampoco las técnicas estadísticas multivariadas de clasificación.

En 1995, la Universidad Nacional, en cabeza de la profesora Indiana Bustos Bustos, realizó la sistematización de información de grupos sanguíneos del ICBF desde 1972 hasta 1995, tendiente a establecer valores de referencia actualizados a su momento con los cuales se daría inicio a un proceso descentralizado para el servicio de filiación en el país. Para ese momento, se planteó el uso de los índices y probabilidades de paternidad. Estos cálculos se basaban en los valores de referencia poblacionales establecidos, dando una perspectiva completamente diferente a la identificación en el país.

El éxito de los marcadores STR's en la práctica forense, y los primeros reportes de variantes polimórficas a lo largo del mundo, crearon grandes expectativas en el país, es así como en 1997 se empezaron a traer los primeros analizadores genéticos, creando la necesidad de establecer valores de referencia de frecuencias alélicas para los nuevos marcadores moleculares, en los que se implementarían los estadísticos de identificación y filiación. Hasta ese momento, los valores de referencia provenían de los reportados por las casas comerciales, naturalmente ignorando las variantes alélicas propias de nuestras poblaciones. En la Universidad Nacional de Colombia, se adelantaron los primeros trabajos tendientes a determinar las frecuencias genéticas para la población colombiana con sistemas microsatélites. Estos mismos estudios fueron desarrollados también por el doctor Manuel Paredes, del Instituto Nacional de Medicina legal y Ciencias Forenses. En esos primeros estudios se pudieron detectar diferencias en las frecuencias alélicas reportadas por ambos estudios, que correspondían a dos muestreos completamente distintos sobre una población en la que no se había realizado ningún proceso de caracterización genética; bajo estas circunstancias, es imposible determinar si se trata de errores técnicos, debidos al muestreo, a las tipificaciones, o realmente a

procesos de la historia evolutiva de las poblaciones. En ese primer inicio, las muestras eran tomadas por muestreos a conveniencia, teniendo como marco los primeros casos de filiación realizados.

Contando con estos primeros valores de referencia, se inicia la importación de varios tipos de analizadores genéticos tales como: el ALF express de Pharmacia, Hitachi FMBIO I y II, o el analizador ABI 310 de Applied Biosystems. Todo esto permitió una mayor capacidad para estudios de filiación, y de manera indirecta los estudios genético poblacionales. La mayoría de estos estudios desarrollados para el país se han hecho utilizando entre 200 y 300 muestras aproximadamente, siguiendo los lineamientos de las comisiones extranjeras para el estudio o desarrollo de pruebas de ADN. Por otra parte, la totalidad de estos estudios corresponden a trabajos realizados con muestras provenientes de pruebas de ADN en casos de paternidad, los cuáles no utilizan métodos aleatorios y por tanto no aseguran una condición de equiprobabilidad. A la fecha, Colombia cuenta con aproximadamente 40 reportes de frecuencias, en los cuales el nivel de organización básico corresponde a departamentos, sin embargo, esta clasificación sociopolítica podría no ser la correcta para la interpretación de la estructura del país y es por tanto necesaria una reevaluación.

Este panorama no solo aplica a marcadores microsatélites autosómicos, en nuestro país se han realizado una gran cantidad de estudios genético poblacionales que utilizan la información genética uniparental (DNA mitocondrial y Cromosoma Y). Entre algunos trabajos ya publicados se encuentran: Mesa et al. (2000); Carvajal-Carmona et al. (2000); Keyeux et al. (2002); Rodas et al. (2003); Bedoya et al. (2006); Salas et al. (2008) y más recientemente Rojas et al. (2010). Todos ellos, con propósitos históricos en relación al patrón y dinámica de flujo génico entre diferentes poblaciones, diferenciando esta contribución por sexo. Adicionalmente, desde las perspectivas médica y epidemiológica, se pretende realizar un aporte al conocimiento de la historia demográfica y dinámica de mezcla de las poblaciones, que sirva para el mapeo genético de rasgos complejos en los estudios de asociación clínica (Rojas et al., 2010; Bedoya et al., 2006; Carvajal-Carmona et al., 2000; Mesa et al., 2000). Si bien dichos estudios han realizado aportes importantes al conocimiento de la historia biológica de las poblaciones humanas en Colombia; estos presentan limitaciones en cuanto a criterios de inclusión, tamaño y procedencia de las muestras, y como consecuencia, también las conclusiones a las que se han podido llegar con estos trabajos son restringidas.

1.2. Un cuestionamiento sobre la actualidad.

En este momento estamos dando pasos hacia la utilización de nuevos marcadores moleculares con una gran capacidad de individualización, y estamos en la carrera de implementar nuevas tecnologías. Sin embargo, el desarrollo de sistemas de información conjunta, procedimientos matemáticos y estadísticos tendientes a desarrollar nuevas metodologías, y la evaluación de los tamaños de muestra, aún siguen en el mismo punto de los primeros estudios de 1998. Pero quizás, lo más preocupante es que nuestro nivel de capacitación en matemáticas, estadística y genética de poblaciones, sigue en franco desnivel con el avance tecnológico.

La ausencia de bases de datos es uno de los temas de mayor preocupación, ya que no existe un criterio unánime para el almacenamiento de la información genética. Debido a esto, cada estudio toma datos de tipificaciones y variables diferentes, criterios poblacionales disímiles, con diversas concepciones sobre muestreo; lo que conlleva a la aplicación de técnicas de toma de datos incoherentes. Un elemento adicional, un tanto complejo de analizar, es el desarrollo de trabajos de grado en su mayoría que se reducen a la aplicación de técnicas moleculares, sin ningún tipo de análisis ni de implicaciones estrictas sobre la genética de poblaciones. Además, los estudios parecen corresponder a un modelo repetitivo de investigación sin mayor cuestionamiento.

En la siguiente parte del artículo veremos lo que pasa en el resto del mundo. Haremos una revisión de aspectos teóricos a tener para el desarrollo de la genética de poblaciones.

2. Fuentes de conocimiento de la Genética de poblaciones.

La genética de poblaciones se fundamenta en tres disciplinas para su desarrollo epistemológico: la biología molecular, encargada de las técnicas de laboratorio para el análisis y obtención de nuevos marcadores; los métodos estadísticos y bioinformáticos, cuyo objetivo es el de analizar los grandes conjuntos de datos que se producen actualmente, tratando de encontrar nuevos métodos y procedimientos analíticos; y finalmente las metodologías de campo, guiadas por el análisis demográfico, que permiten realizar una conexión entre la información genética, cultural, histórica y ambiental de la población en estudio; esta última tendencia además permite el uso de herramientas

provenientes de la genética clásica.

2.1. Biología Molecular.

El desarrollo tecnológico en los análisis genéticos de filiación para esclarecer casos de paternidad, con el fin de proteger los derechos de los menores y las desapariciones forzadas en actos violentos, se han convertido en una prioridad para el país.

Desde 2005, la Fiscalía viene realizando una tarea de exhumación de víctimas del conflicto colombiano, permitiendo la recuperación de 1903 cadáveres (hasta el 20 de noviembre de 2008), de los cuales tan sólo el 16% (304) de los cuerpos han sido identificados y entregados a sus familiares, dando cumplimiento al artículo 44.4 de la Ley de Justicia y Paz el cual proclama: “*La búsqueda de los desaparecidos y de los restos de personas muertas, y la ayuda para identificarlos y volverlos a inhumar según las tradiciones familiares y comunitarias*”.

La identificación de desaparecidos, restos cadavéricos y solución de casos por medio de técnicas de análisis de polimorfismos de ADN, se ha convertido en una actividad frecuente dentro de la rutina pericial de muchos laboratorios de identificación genética, las fases analíticas de estas pruebas se basan en fundamentos científicos lo suficientemente rigurosos que requieren el cumplimiento de condiciones específicas de calidad durante el procedimiento técnico (Begoña et al., 2005).

Las investigaciones tecnológicas de los marcadores moleculares han sido enfocadas, precisamente, hacia el perfeccionamiento de procesos aptos para la resolución de casos a partir de muestras difíciles, por lo que se observa un gran desarrollo tecnológico en las herramientas de biología molecular. En 1985 Alec Jeffreys introdujo la “huella genética” empleando una técnica rudimentaria y dispendiosa; contrario a esto, hoy en día contamos con facilidades tecnológicas empleando sistemas *multiplex* que nos permiten tipificar e individualizar a una persona usando marcadores de los cromosomas autosómicos, sexuales y del ADN mitocondrial, a bajos precios y en menor tiempo.

Los principales marcadores con los que se logró analizar varios loci simultáneamente, fueron los STRs (*Short Tandem Repeat*), para los que se requiere una mínima cantidad de muestra y cuyo procesamiento es rápido y automático empleando kits comerciales (Yunis et al., 2002). Para constituir los datos de las frecuencias poblacionales, inicialmente el FBI (*Federal Bureau of Investigation*) de

Estados Unidos, implementó el uso de una base de datos introduciendo 13 marcadores tipo STR llamada CODIS (*Combined DNA Index System*). En muestras parcialmente degradadas se han introducido los miniSTR, los cuales poseen una reducción del tamaño del producto amplificado generando un mayor número de casos resueltos (Butler et al., 2003).

Hoy en día, uno de los marcadores novedosos más prometedores en la práctica forense, son los polimorfismos simples de un solo nucleótido (SNP). Existen ya en la actualidad multiplex de SNPs validados para uso forense (Butler, 2000), permitiendo resolver casos complejos de paternidad y especialmente en criminalística e identificación. Los SNPs han sido considerados como marcadores genéticos ante la comunidad forense por varias razones: la primera y más importante, es que los productos de PCR pueden ser menores a 100pb, lo que significa que estos marcadores tienen la habilidad de recobrar información de muestras de ADN altamente degradadas, mejor que los STRs, los cuales generan amplicones de tamaño superior (300-400 pb). En segundo lugar, se pueden manejar set multiplex con una mayor cantidad de regiones analizadas que no estén condicionadas por el número de fluorocromos o rango de tamaño. Finalmente, el procesamiento de la muestra y el análisis de los datos pueden ser automatizados rápidamente debido a que no es necesaria una separación basada en tamaño. Para el campo forense es un reto la necesidad de analizar un gran número de SNPs para que sean informativos y obtener un poder de discriminación razonable. Sin embargo, ya se han realizado algunos multiplex que permiten analizar 50 SNPs simultáneamente, generando un perfil individual, y por tanto, permitiendo la obtención de un poder de discriminación que le facilita competir con los marcadores microsatélites.

Otros marcadores que están empezando a ser analizados e implementados en las ciencias forenses son los polimorfismos generados por la inserción o delección de uno o más nucleótidos, conocidos como *indels* (DIPs). (Weber et al. 2002). Estos marcadores combinan las características óptimas de los SNPs y los STRs como: 1. Se extienden por todo el genoma humano (cromosomas autosómicos y sexuales); 2. Este polimorfismo se deriva de un evento mutacional único; 3. Muestran diferencias significativas en las frecuencias alélicas entre los grupos de población geográficamente separados, por lo tanto, pueden ser utilizados como marcadores informativos ancestrales; 4. Los DIPs generan amplicones cortos con un máximo de 150pb, lo cual genera beneficios para manejar muestras forenses

altamente degradadas; y 5. Poseen una mayor capacidad de exclusión que los SNPs y STRs, con respecto a la identificación forense o el análisis de la paternidad (Weber et al. 2002; Pereira et al., 2009). Esto demuestra que los recientes avances en genética forense se han centrado en el desarrollo de ensayos de genotipificación con amplicones cortos, con el fin de mejorar la amplificación para la obtención de perfiles genéticos completos especialmente en muestras degradadas. Además, la genotipificación rápida y rentable con estos marcadores puede ser utilizada para distinguir los principales grupos poblacionales, especialmente los europeos, africanos y americanos nativos. Los *indels* también permiten identificar subestructura en poblaciones mixtas. Actualmente existen paneles de 48 marcadores DIPs que se han empleado en el análisis poblacional, demostrando que en poblaciones mixtas de diferentes grupos ancestrales los *indels* permiten estimar, de manera precisa y fiable, la mezcla interétnica individual y global en relación con los grupos (Weber et al. 2002). Por tanto estos marcadores son altamente competentes y útiles en los campos de análisis forense y poblacional.

Como vemos, los avances en biología molecular han dado grandes saltos tecnológicos, ofreciendo cada vez, un mayor número de métodos automatizados, rápidos, robustos, y económicos para las identificaciones humanas; por lo que en un futuro, sin duda, se podrá secuenciar a gran escala cromosomas enteros o todo el genoma (Xue et al., 2010). El desarrollo de tecnologías de ultrasecuenciación con diseños bioquímicos particulares, induce a pensar en que sus potenciales aplicaciones forenses no serán ciencia ficción (Carracedo et al., 2010).

2.2. Matemáticas y Estadística.

Ahora es posible la unión de la biología molecular, aplicada con herramientas de alta tecnología; junto con elementos matemáticos, estadísticos e informáticos. Durante varias décadas, los aportes de naturaleza matemática solamente constituyeron modelos teóricos lejanos a los conceptos biológicos, debido a que no se contaba con la tecnología, las herramientas informáticas ni el conocimiento sobre el genoma humano para realizar esta integración.

Muchas de las disciplinas de la estadística hoy en día tienen que ver con el desarrollo de la genética de poblaciones: A partir de los estadísticos multivariados es posible para nosotros evaluar múltiples

genes y variantes alélicas simultáneamente, para determinar las diferencias y similitudes entre poblaciones, por otra parte, hoy en día es posible gracias al uso de la teoría de probabilidad realizar particiones en espacios muestrales y asignar la probabilidad de aparición de determinadas variantes alélicas, haplotipos y haplogrupos. También es posible la utilización de teoremas de probabilidad, que permiten evaluar la aparición de un determinado paisaje fenotípico teniendo presentes diferentes eventos condicionales. Hoy la genética también se constituye en una herramienta de desarrollo constante la estadística no paramétrica, y también plantea problemas fundamentales a las metodologías y técnicas de muestreo establecidas desde las escuelas tradicionales de pensamiento. En los próximos 50 años, es posible que muchas de las herramientas matemáticas y estadísticas paramétricas actualmente en uso cambien y sean reevaluadas, gracias al desarrollo que han tenido por los constantes cuestionamientos de las disciplinas biológicas. Una de los cambios más determinantes en la genética moderna es el paso de estimadores puntuales a estimadores por intervalos.

Las herramientas matemáticas y estadísticas, asociadas a la genética de poblaciones, han permitido evaluar con mucha mayor precisión la magnitud de las fuerzas de cambio evolutivo, la forma en la que afectan a una población, el grado de las interacciones existentes entre ellas, y han permitido, también, cuantificar y ordenar los diferentes procesos que afectan a las poblaciones. Uno de los elementos que hoy en día vive un gran cuestionamiento es el tamaño de muestra necesaria en un estudio genético poblacional, de tal manera que las conclusiones evolutivas obtenidas a partir de un número limitado de individuos puedan ser extrapoladas a la totalidad de la población. Una revisión detallada de los diferentes estudios genéticos poblacionales, deja ver una gran disparidad en los tamaños de muestra, incluso en las mismas regiones geográficas. La consecuencia inmediata de este fenómeno es la obtención de estimadores genético poblacionales en un rango amplio, que no permiten ver claramente cuáles son los efectos de las fuerzas de cambio evolutivo, y cuáles son variaciones debidas a las disparidades propias de un margen de error por causas del procedimiento establecido para la obtención de las muestras. La comisión internacional que regula los estudios con ADN a lo largo de los últimos 15 años, ha emitido diferentes comunicados, en los cuales se propone cada vez más un mayor número de muestras, sin embargo, no existen estudios estadísticos apropiados que nos indiquen cuál es un número aproximado de muestras a tomar en un estudio genético poblacional.

En este punto, es probable que tengamos la necesidad de desarrollar nuevas herramientas matemáticas y estadísticas, y nuevos algoritmos con lógicas de procedimiento y procesamiento diferentes a los tradicionales, ya que el muestreo dentro de la investigación genético-poblacional, en relación con los métodos tradicionales, introduce una serie de elementos hasta ahora poco utilizados tales como la naturaleza misma del marcador a estudiar. De esta manera, algunos elementos que pueden ser determinantes en el proceso de establecer un adecuado tamaño de muestra son: la identidad de la fuente de información genética; si el marcador es autosómico, es el cromosoma Y o el ADN mitocondrial, que corresponden a linajes uniparentales; ó si corresponde a marcadores sexuales, tales como el cromosoma X, que tiene una serie de particularidades en el proceso de recombinación; o, por ejemplo, el nivel de polimorfismo del marcador, si éste es de naturaleza neutral o si está asociado a una serie de marcadores con acentuados procesos de selección fenotípica. Por otra parte, procesos genético poblacionales relacionados con la historia de poblamiento, la distancia en el tiempo, los cuellos de botella, el nivel de endogamia que pueda sufrir una población, los diferentes modelos de flujo genético que pueden estar afectando la interacción entre poblaciones, o el nivel de barreras geográficas que han determinado las diferentes rutas de poblamiento de una región en particular; son elementos que en la práctica cotidiana y en la teoría formal no son considerados para el establecimiento de un adecuado tamaño de muestra.

2.3. El Análisis demográfico.

La información demográfica se ha aprovechado para involucrar aspectos como la estructura por edad y sexo, ó las tasas de migración, natalidad y mortalidad de las poblaciones a los análisis de la información genética colectada. Este tipo de estudios se dirigen, por ejemplo, al establecimiento de evidencias sobre la variabilidad biológica de algunos grupos que habitan en nuestro país a partir de la demografía, las ancestrías y las frecuencias alélicas de determinados marcadores moleculares (Carvajal - Carmona et al., 2003), o a la identificación de aspectos como la selección y la deriva a partir de movimientos territoriales, nacimientos y defunciones (Acreche et al., 2003). Esta información presenta a su vez la posibilidad de integrar *variables socioeconómicas y espaciales* o análisis de isonimia -tan frecuentemente utilizados por la genética clásica (Crow y Mange, 1965) que soportan desde un punto de vista temporal, espacial y social los análisis realizados, teniendo en cuenta

el hecho de que la genética de poblaciones humanas no se puede abarcar ni en tiempos ni en dinámicas poblacionales constantes o independientes de cualquier aspecto social o cultural.

Precisamente, entre los años 50 y 70, la genética clásica también aprovechaba otro tipo de *variables socioculturales* -de carácter más cualitativo y descriptivo-, que podían junto a la economía y el espacio, llevar a un alcance mucho mayor el tipo de estudios realizados en algunas poblaciones (Salzano 1977, 1979, 1980), la etnografía y la historia serían disciplinas importantes a la hora de realizar abordajes a la población, existiendo así, un trabajo de campo y una serie de métodos de investigación consecuentes, no sólo con preguntas de investigación genéticas sino también antropológicas. Sin embargo, esta tendencia cambió sustancialmente cuando los estudios genéticos empezaron a tener una mayor preferencia por los novedosos análisis a partir de marcadores moleculares, los cuales aunque se seguirían sosteniendo de algunas variables espaciales y demográficas -o en algunos casos menores socioeconómicas- para demostrar variabilidad y factores microevolutivos; no volverían a aprovechar las variables socioculturales, de la manera como se hacía para épocas anteriores; un tanto porque las disciplinas: genética y antropológica, empezarían a especializarse y a utilizar metodologías de investigación más distintas la una de la otra, y también porque los genetistas prefirieron modificar la manera de realizar sus muestreos y de definir los criterios desde los cuales se elegirían las poblaciones sobre las que se iba a trabajar.

Esta separación ha traído una serie de transformaciones a la aplicación de la información demográfica sobre los datos genéticos. Como se venía diciendo en numerales anteriores, se debe tener en cuenta que algunos estudios realizados en nuestro país, han tenido ciertas dificultades en cuanto a la organización por departamentos de las frecuencias reportadas o en los criterios de inclusión y tamaño de las muestras. Para el primer caso, por ejemplo, se ignora el hecho de que un factor, como lo es, el establecimiento de departamentos o de regiones administrativas, no puede coincidir exactamente con el devenir geográfico e histórico de las poblaciones, el cual muy probablemente va más allá de una separación hecha con líneas en un mapa, cosa que hace que nos cuestionemos respecto a la manera como se definen las poblaciones a muestrear, cuando se pretende realizar a un nivel nacional o regional, al mismo tiempo, nos hace pensar respecto a nuevas formas de definir esos criterios con los cuales se escogen las personas que harán parte de un estudio, más allá de su pertenencia a una base de

datos o a una serie de muestras almacenadas masivamente en un laboratorio. A partir de ello, podríamos considerar entonces, que la demografía integrada con otro tipo de información histórica, etnográfica, geográfica y económica dentro de estos nuevos estudios genéticos, tanto en su planteamiento como en la presentación de sus resultados; ofrecería la posibilidad de superar las dificultades antes mencionadas.

Sánchez Compadre (2001), por ejemplo, reconoce que factores microevolutivos como las migraciones, la deriva genética o los patrones de cruzamiento, están determinados por los componentes culturales de nuestras sociedades, implicando el desarrollo de otro tipo de análisis que complementen el trabajo biodemográfico; en este caso, el trabajo de campo se presenta como una oportunidad para que el genetista realice un estudio con unos criterios de selección, un muestreo y un análisis de la información genética concordante con la historia, la geografía y los aspectos culturales de una población. Parece ser que existe cierta conformidad con los simples reportes de las frecuencias alélicas encontradas o con la sistematización de bases de datos, sin contar con que la probabilidad de cometer errores puede ser mucho mayor si no se tienen en cuenta las razones por las cuales se escogió una población, los motivos de selección de los marcadores o el fundamento que sostiene el uso de determinadas ecuaciones.

Muchos genetistas en nuestro país consideran el trabajo de campo como un proceso demasiado difícil y costoso, prefiriendo seguir recurriendo a las muestras almacenadas en casos forenses o de filiación, sin cumplir con los requisitos estadísticos de la equiprobabilidad y sin tener en cuenta los errores que pueden llegar a existir en los resultados presentados, al no haber criterio alguno de selección. Sin embargo, debería considerarse a la demografía más que como una herramienta matemático/estadística para expresar las dinámicas poblacionales sobre las que se fundamentan los estudios genéticos (Grafica 1), como una valiosa fuente de información cuantitativa y cualitativa -al estar integrada con información etnográfica o histórica-, para realizar un análisis más profundo tanto de las dinámicas y estructuras poblacionales, como de las fuerzas evolutivas que inciden sobre los grupos con quienes se trabaja. Se pretende entonces dejar claro que la demografía es una fuente de conocimientos que le presenta una oportunidad a la genética de poblaciones, de integrarse con otro tipo de metodologías tanto cuantitativas como cualitativas para realizar análisis mucho más completos y potentes de la

información genética resultante en el laboratorio y en los software de análisis poblacional. Más adelante se presentará en detalle la manera como se puede aprovechar esa oportunidad y la manera como los aspectos biológicos y antropológicos que alguna vez en la época clásica hicieron parte de la genética podrían ser de nuevo de utilidad en el reconocimiento de la estructura de la población de nuestro país.

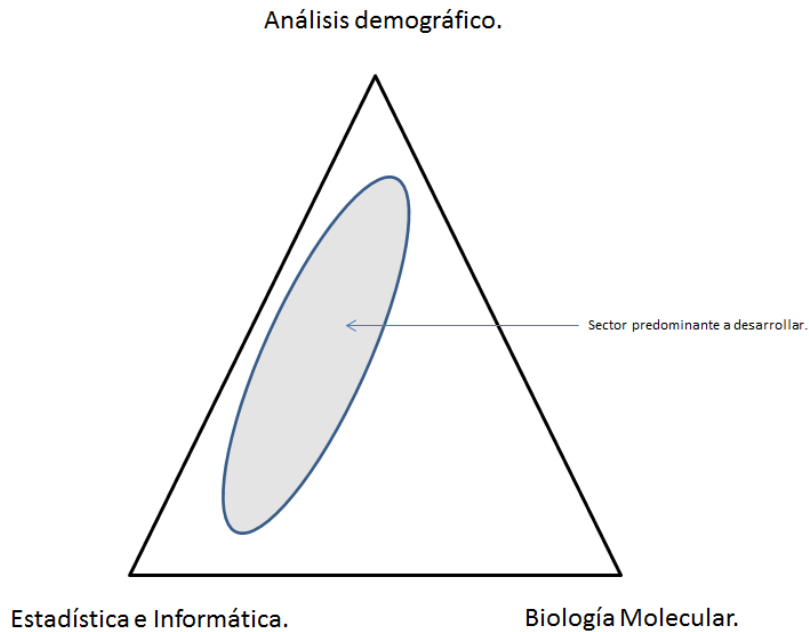


Figura No. 1: Tres fuentes del desarrollo epistemológico de la genética de poblaciones humana. En azul, la región de importancia a desarrollar en el país.

Basados en lo anterior nos permitimos reorientar las tendencias mundiales para hacer un planteamiento a nivel local, buscando desarrollos concretos en la próxima década de investigación genético poblacional en nuestra región.

3. Cinco propuestas de trabajo.

3.1. Un análisis descriptivo de cada polimorfismo.

Los primeros valores que se obtienen de los análisis genético poblacionales son las frecuencias alélicas, normalmente calculadas por la técnica del conteo directo, a partir de ellas suelen realizarse los demás estimadores de la genética descriptiva, tales como la heterocigocidad, la condición de equilibrio HW; y los principales estimadores forenses, como por ejemplo, el poder de discriminación, el contenido de información polimórfica, y demás valores. Una de las sugerencias es iniciar con el análisis mismo del polimorfismo, empezando por la determinación de alelos efectivos. Éste es un cálculo muy sencillo que permitirá encontrar las variantes alélicas determinantes en aproximadamente el 90% de la población, o visto de otra forma, el número de variantes más frecuentes encontradas en los diferentes perfiles genéticos. Además, este valor va a indicar la verdadera variabilidad del polimorfismo. Un ejemplo sobre nueve sistemas microsatélites y la manera como estos son analizados se presenta en la tabla número uno. Este tipo de análisis simple nos deja ver desde una perspectiva descriptiva la variabilidad y el comportamiento de un polimorfismo e una población específica.

	CSF1P0	D13S317	D3S1358	D5S818	D7S820	FGA	TH01	TPOX	vWA31
No. de alelos reportados	11	11	10	10	10	26	8	8	13
No. Teórico de Genotipos Posibles	66	66	55	55	55	351	36	36	91
No. de genotipos Observados	49	44	36	43	50	128	29	34	61
% de genotipos Observados sobre el total teórico.	74.2%	66.7%	65.5%	78.2%	90.9%	36.5%	80.6%	94.4%	67.0%
No. alelos efectivos	4	5.8	4.0	4.4	4.9	8.3	4.4	3.5	4.6

% de Alelos efectivos respecto al total del sistema.	37%	52%	40%	44%	49%	32%	55%	44%	35%
No. genotipos posibles con el No de alelos efectivos	10	19.5	10.2	11.7	14.3	38.4	11.9	7.8	12.8
% genotipos observados solo con los alelos efectivos sobre el total de la población.	90%	90.20%	96.44%	88.70%	92.01%	91.56%	88.51%	87.38%	91.24%

Tabla No. 1: Análisis descriptivo de nueve polimorfismos genéticos. Algunos estadísticos y valores, que se proponen para el análisis de la utilidad de un marcador genético.

En primer término, aparece el número de variantes alélicas encontradas en la población; luego el número de variantes que son reportadas a nivel mundial para el sistema de estudio; la relación entre estos valores, nos permite observar el porcentaje de variantes sobre el total que han sido encontradas en la población de estudio. Otro porcentaje muy importante es el calculado del número de alelos efectivos sobre el número de variantes encontradas, lo que arroja el porcentaje de alelos efectivos sobre el total de la población, y esto nos permite evaluar de una mejor manera la verdadera eficiencia del polimorfismo. Es frecuente encontrar sistemas con un gran número de variantes alélicas, pero que en términos efectivos, representan un porcentaje pequeño sobre el polimorfismo. En muchas oportunidades, al discutir sobre la calidad de un polimorfismo en la práctica, se entiende que es bueno, si presenta un gran número de variantes genéticas; pero, muchas de estas variantes no son efectivas en la población. Un cálculo interesante a realizar, es el porcentaje de genotipos en la población de estudio conformados por los alelos efectivos; estos en buena parte, corresponden aproximadamente al 90% de las personas en la población. Tan sólo, un 10% restante, será propio de las variantes en baja frecuencia y esto es un indicador directo, es decir, un valor observado de la verdadera variabilidad y utilidad del polimorfismo en la población.

3.2. Una información demográfica integrada a datos históricos, etnográficos y estadísticos.

El análisis demográfico, representa una herramienta de gran valor en la clasificación de datos y en la interpretación de procesos, no sólo de migración, sino también de ancestrías y conformación histórica de las poblaciones. Esta información puede tomarse a través de una encuesta como la presentada en la Gráfica No. 2.

UNIVERSIDAD NACIONAL DE COLOMBIA SEDE BOGOTÁ

PROYECTO: "CARACTERIZACIÓN GENÉTICA DE LA POBLACIÓN HUMANA DEL DEPARTAMENTO DE LA GUAJIRA COMO PARTE DEL ENTENDIMIENTO HISTÓRICO Y ESTRUCTURAL DE LAS ETNIAS COLOMBIANAS"

1. Fecha: 30 Junio/09
 Código de la Mue.: 600903020
 Institución Toma mue.: H.N.S.R.
 Persona Toma Mue.:
 Persona Toma Encu.: HR

2. Localización de toma de Muestra: Municipio: RIOHACHA Loc. Específica: HOSPITAL

3. Información Personal: Lugar de Residencia: RIOHACHA Tiempo de Residencia: 23
 Dirección: ND Teléfono: ND C-Electrónico: NA
 Primer Nombre: [REDACTED] Segundo Nombre: NA

4. Apellidos:
 P1 MARTINEZ (RIOHACHA LN) POP. MESTIZO
 P2 NAVARRO (PALONINO LN) POP. MESTIZO
 M1 MUÑOZ (S.J. CESAR LN) POP. MESTIZO
 M2 DOMINGUEZ (S.J. CESAR LN) POP. MESTIZO

5. Edad: 23 (RIOHACHA LN) POP. MESTIZO; Edad: 20 (BAERINQUA LN) POP. MESTIZO

6. Etnia: NA Casta: NA Ranchería: NA
 Edad nacimiento 1º hijo: NA No. hijos: NA
 Wayunaiki Habla: SI Escribe: SI X

7. Edad: 52 (RIOHACHA LN) POP. MESTIZO

8. Edad: 49 (SAN JUAN DEL CESAR LN) POP. MESTIZO

9. POP: INDG, AFR, MEST (X), ARB, MIG

Figura No. 2: Modelo de encuesta genealógica utilizada para complementar los muestreos por conveniencia, haciéndolos independientes de los criterios del investigador.

Habitualmente, en los análisis poblacionales hasta ahora realizados en el país, la única variable con la que se cuenta, es el lugar de nacimiento; que está expuesta a un gran número de problemas, ya que no refleja necesariamente la ancestría de una persona. Precisamente, con el uso de la encuesta, se podrán

detallar no sólo los lugares de nacimiento de la persona encuestada, sino también una serie de informaciones familiares que nos podrán llevar hacia atrás en el tiempo y en el espacio, definiendo de esta manera, los procesos migratorios de las diferentes familias participantes en el muestreo y las configuraciones familiares desde las cuales se pueden emprender estudios de ancestría y de consanguinidad. Esta información permitirá a su vez, la integración con una serie de datos históricos y etnográficos, que podrán validar o describir a mayor profundidad la serie de eventos registrados en la encuesta; con miras, no sólo hacia la vinculación con los análisis genéticos, sino hacia la definición de los criterios de selección de poblaciones para el muestreo, teniendo en cuenta elementos estructurales para un buen cálculo de tamaño de muestra (Grafica No. 3), mucho antes del análisis de laboratorio.

El uso de encuestas, va a permitir además, la clasificación y ordenación de los principales grupos que habitan una región geográfica específica, bajo unos criterios que no se ciñen exclusivamente al espacio o a la división política de un territorio; sino también, a lo que la información histórica, económica y etnográfica logre definir como grupo poblacional. La mayoría de veces, como se mencionó anteriormente, los individuos de una población son clasificados sólo por su lugar de nacimiento y así ingresan a los archivos de datos. Sin embargo, la posibilidad de clasificar a los individuos no por poblaciones geográficas, sino por poblaciones de naturaleza ancestral, justificando esta división con información tanto cuantitativa como cualitativa, logrará una aproximación más precisa de los principales grupos ancestrales y de una serie de características socioeconómicas, espaciales y culturales que componen una población humana en particular.

Aplicaciones que utilizan metodologías bayesianas para la asignación de individuos a poblaciones, como el software Structure 2.3.3 (Pritchard et al., 2000), son ampliamente utilizados, ya que arrojan un valor k correspondiente a los grupos poblacionales presentes en la muestra de estudio; pero, estrictamente, esto debe entenderse como el número de núcleos de inercia o diferentes conformaciones genotípicas presentes en una población. Cuando se ingresan los datos, ordenados por un criterio de ancestría, se tiene mayor posibilidad de encontrar correlación entre los grupos definidos por el software Structure y los núcleos poblacionales descritos al inicio del experimento (la encuesta permite hacer un análisis a priori a la clasificación). Por lo tanto, el análisis demográfico, nos va a

permitir verificar un modelo *a priori* de grupos poblacionales presentes en una región geográfica, y compararlo con el modelo matemáticamente construido *a posteriori*; como un proceso de verificación del primer modelo. En consecuencia, no estamos restringidos estrictamente a los modelos matemáticos, sino que éstos se convierten en una verificación de la información obtenida por clasificación demográfica de la población a estudiar genéticamente; es una manera de tener un segundo punto de control respecto a los análisis matemáticos que se realizan posteriores al análisis en biología molecular. Frecuentemente del laboratorio, se pasa directamente al análisis informático sin contar con un segundo punto de referencia, que permita inferir si estos análisis matemáticos son o no acertados. En el mejor de los panoramas, se realizan diferentes metodologías matemáticas y estadísticas, que se contrastan entre sí, para buscar la coherencia de la información obtenida, con el análisis genealógico. Así, la clasificación de las poblaciones *a priori* antes del análisis en biología molecular, se convierte en una poderosa herramienta que da una amplia perspectiva al análisis evolutivo.



Figura No. 3: Elementos estructurales para un tamaño de muestra.

3.3. Genética clásica, isonimia.

El advenimiento de las técnicas de biología molecular, y el análisis informático sobre el genoma han relegado casi hasta la desaparición muchas de las herramientas del análisis genético clásico, entre ellas los estudios de isonimia. Los apellidos son una característica cultural y la principal forma en la que se identifican relaciones de parentesco entre las personas de una población (Gómez, 2004). En la civilización occidental, se generaron entre la edad media y comienzos de la edad moderna, cuando se hizo necesaria la identificación precisa de las personas con fines sociales y demográficos (Giraldo, 2001). Los colonizadores de España y otras partes de Europa llevaron sus respectivos sistemas de nombrar al Nuevo Mundo, donde acabaron por establecerse tanto entre los supervivientes de los pueblos autóctonos como entre el resto de los recién llegados (Lasker, 1991).

Los apellidos son usados en estudios de biología humana debido a que en muchas culturas, se heredan como los genes; en cada generación, los descendientes tienen la mitad de los apellidos de sus padres y la mitad de sus genes autosómicos (Lasker, 1980). En nuestra civilización, los apellidos se heredan por vía paterna y en el caso de la descendencia masculina, se asemejan a alelos neutros de un gen transmitido por el cromosoma Y (Yashuda & Furusho, 1971; Zei *et al.*, 1983). Con relación a éste evento bio-antropológico como bien lo dicen Pineda *et al.*, (1999), la Isonimia, definida de forma muy general como el estudio de concordancia en los apellidos (iso= igual, nimia=nombre), ha surgido como una herramienta para analizar poblaciones humanas y permitir una aproximación inicial a la estructura genética. Sanz *et al.*, (2003) se refieren a la isonimia (o isonomía) como al estudio de la frecuencia y distribución de apellidos en poblaciones humanas, mediante el cual pueden establecerse relaciones de parentesco y origen. Estudios como el de Giraldo (2001) se refieren a la isonimia como al análisis de los matrimonios de parejas que comparten apellidos, éste concepto está muy relacionado con el establecido por Crow y Mange (1965), quienes la definieron como la frecuencia de matrimonios en los que ambos cónyuges tienen el mismo apellido. Estos autores también sentaron las bases teóricas y los supuestos básicos (Román *et al.*, 2007) para el desarrollo del método isonímico:

- 1) El apellido es hereditario. Es decir, que en la población de estudio, los apellidos han

de ser transmitidos de forma regular de generación en generación y sin modificación desde su creación.

- 2) No hay desproporción numérica en relación a los dos sexos.
- 3) El apellido es monofilético

El uso de apellidos en estudios genéticos data de 1875, cuando George Darwin los usó para estimar la frecuencia de matrimonios entre primos hermanos (Jobling, 2001). Los primeros estudios en recurrir al método isonímico se dieron en el campo de la biología en la década de 1960. Según Lasker (1991), en 1960 Richard Shaw publicó un breve artículo que no hacía referencia a ninguna bibliografía anterior. En él hacía notar que el uso de los apellidos según el sistema español, significaba que los dos miembros de un matrimonio entre primos hermanos comparten un apellido. Shaw señalaba que éste hecho podría ser útil a la hora de distinguir la consanguinidad en los países de lengua española. Sin embargo, la mayoría de publicaciones hacen referencia al trabajo de Crow y Mange, quienes publicaron en 1965 un estudio sobre la estimación de la consanguinidad de las poblaciones por medio de la isonimia (Gómez, *et al.*, 2008).

Los análisis de isonimia se han usado básicamente para determinar tasas de migraciones, relaciones de ancestría y origen de las poblaciones, grado de estructuración genética (Bedoya, *et al.*, 2006), resolución de algunos casos forenses (King & Jobling, 2009), diferenciación étnica (Mateos, 2007); para evaluar procesos de mestizaje (Gottlieb, 1983) y relacionar la incidencia de algunas enfermedades hereditarias con posibles consecuencias del efecto fundador de las poblaciones (Bedoya *et al.*, 2006). La utilidad del método se fundamenta en la rápida obtención de datos, que conllevan a la identificación de características básicas de la población, de la misma forma que se haría mediante el uso de herramientas más costosas y de difícil acceso. Las fuentes de información para obtener datos de apellidos son diversas; según Costa *et al.*, (2000) se han utilizado registros religiosos y civiles de bautismo, matrimonios, defunciones, nacimientos y directorios telefónicos. Para estos autores, la utilización de registros de defunciones para analizar la estructura genética de una población, no es frecuente, y esto podría ser una desventaja. Sin

embargo, Lasker (1969) en un estudio del pueblo pesquero de San José (Perú), empleó cuatro fuentes de datos: entrevistas, lápidas funerarias, registros de nacimiento y registros de defunción; mostrando que no había diferencias estadísticamente significativas entre las fuentes de datos usadas (Lasker, 1991).

En Colombia, las fuentes empleadas corresponden a: bases de datos para pruebas de paternidad, de casos del Instituto Colombiano de Bienestar Familiar, ICBF (Giraldo, 2001; Gómez, 2004), y del Instituto Colombiano de Medicina Legal (Gómez, Ávila y Briceño, 2008); registros de usuarios del Sisbén (Bedoya *et al*, 2006; Gómez, Hinestroza y Muñetón, 2008), registros de nacimiento (Pineda, Arcos y Bravo, 1999) y parroquiales de matrimonio (Soto *et al*, 2001).

En la literatura se han descrito diversos coeficientes, los cuales pueden llevar a confusión según la modalidad de traducción y según la letra latina o griega utilizada en su formulación (Giraldo, 2001). A partir de la frecuencia de apellidos es posible calcular estimadores como: la isonimia aleatoria sesgada (Scapoli *et al*, 2007) y no sesgada (Rodríguez-Larrade, *et al.*, 1993), el coeficiente de endogamia total (Malnar, 2002) y aleatorio (Scapoli *et al*, 2007), distancia genética (Giraldo, 2001) y los estimadores A, B (Pineda *et al*, 1999) y C (Bedoya *et al*, 2006). Además del indicador alfa (α) de Fisher (Barrai, *et al.*, 2007) y del indicador ν de Karlin-McGregor (Zei *et al*, 1983).

Los estudios de isonimia en Colombia no son muy abundantes, sin embargo existen varias publicaciones usando el método con diferentes fines. Básicamente, los análisis de isonimia en el país, han estado centrados sobre poblaciones pequeñas, que muestran señales culturales, étnicas o demográficas de constituir aislados genéticos. Tan solo dos trabajos (Giraldo, 2001; Gómez, 2004) se han realizado en Colombia con una perspectiva nacional y no departamental o municipal. Cabe señalar también, que es Antioquia la zona de mayor aplicación de análisis de isonimia (Pineda *et al*, 1999; Soto *et al*, 2001; Bedoya *et al*, 2006; Gómez *et al*, 2008; Arias, Rojas & Bedoya, 2010), ya que como lo mencionan algunos autores (Bedoya *et al*, 2006a; Bedoya *et al*, 2006b) el poblamiento de este departamento, se originó a partir de dos núcleos fundadores y su población se ha considerado característicamente cerrada, debido a profundas diferencias culturales y a una historia marcada por fenómenos de aislamiento después de la época

colonial. Otros estudios realizados corresponden a un análisis comparativo entre los haplotipos del cromosoma Y y los apellidos en los departamentos del Valle del Cauca, Cauca y Nariño (Gómez et al, 2008).

A pesar de los atributos con los que cuenta el análisis de isonimia, el método también impone ciertas restricciones propias de la técnica y otras que están sujetas al diseño experimental de los investigadores. Los supuestos con los que se aplica el modelo han sido muy controversiales (Rogers, 1991), pues los apellidos en realidad no presentan un origen único, e incluso muchos de ellos fueron impuestos a las personas debido a su condición de esclavitud (Navarrete, 2005). Adicionalmente, la proporcionalidad entre hombres y mujeres no siempre es posible de alcanzar debido a que los procesos de migración, pueden afectar de una manera diferencial a ambos sexos, casos como las guerras y el movimiento en busca de oportunidades laborales, son algunos ejemplos. Aunque, las objeciones del método abundan como consecuencia de los supuestos, es necesario también recordar que todos los modelos planteados en ciencia presentan limitaciones. Por tanto, la manera apropiada de enfrentarlas es a través de un diseño experimental adecuado, iniciando con la determinación de las preguntas de investigación, susceptibles a responder a partir del método de isonimia y con la elección de las poblaciones en las que es posible aplicarlo. Igualmente, la selección de la base de datos, constituye un paso crítico en el establecimiento de estudios de isonimia; y en este sentido, la base de datos más inclusiva de la población debería dar una idea más clara de la dinámica de apellidos; sin embargo, esto dependerá en gran medida de su historia y naturaleza. Una buena opción sería combinar diferentes fuentes de datos de apellidos teniendo la precaución de no sobreestimar la abundancia de los mismos.

3.4. Una base de datos nacional: Para uso en genética de poblaciones, forense y para monitoreo en epidemiología.

En el mundo, a partir de 1950, las tecnologías de la información y la comunicación iniciaron un desarrollo notable. Esto ha repercutido en la transformación de la comunicación y en las relaciones

sociales, económicas y científicas; como consecuencia del incremento en las capacidades para generar, sistematizar, compartir, transmitir, analizar y difundir la información (Michan, 2010). En investigación científica ha habido una producción acelerada de una gran variedad de programas, aplicaciones, herramientas, recursos y servicios electrónicos; además de adoptar el formato electrónico y la Web como medio de comunicación para compartir información. Estos cambios, sin duda han repercutido en la transformación de la práctica científica a varios niveles, debido a que se ha reducido la energía, el costo y el tiempo requeridos para el análisis de la información. Además, ha generado nuevas oportunidades de investigación al incrementar el espectro de opciones técnicas y análisis factibles (Michan, 2010).

Después de tres décadas de trabajo en biología molecular, hay un gran volumen de información biológica generada por toda la comunidad científica. Este crecimiento explosivo de la información ha exigido en primera instancia la creación de bases de datos para su almacenamiento y organización, además del desarrollo de herramientas computacionales para el análisis de la información, que permitan estudiar nuevas hipótesis biológicas de manera lógica y eficiente. Actualmente en internet están disponibles un número importante de bases de datos para distintos marcadores moleculares y para distintos usos, se destacan por ejemplo ALFRED, NCBI, MITOMAP y mtDB.

En Colombia se ha acumulado información de VNTRs, STRs, SNPs y secuencias desde los años 90s, producto de pruebas de paternidad e investigaciones de los principales laboratorios y universidades. Desde el año 2000 hasta la actualidad, se han publicado aproximadamente 40 reportes de frecuencias de diferentes regiones en el país (Gómez et al., 2003; Paredes et al., 2003; Rey et al., 2003; Usaquén, 2006). Aunque dicha información, es valiosa por su tamaño de muestra y representatividad, infortunadamente estos datos se encuentran dispersos; ya que no se cuenta con un sistema que los almacene en una sola fuente de información. Además, para muchos de estos datos se carece de un análisis riguroso genético-poblacional, en buena parte por falta de fuentes de informaciones completas, consolidadas y con adecuados criterios de validación.

Considerando, que han pasado los primeros quince años de investigación en esta área, y partiendo del

poco conocimiento de nuestras poblaciones; el trabajo hasta el momento realizado, puede considerarse adecuado. Con el transcurso de una época en la que se presentaron los trabajos de genética de poblaciones básicas y el desarrollo de una primera revisión de frecuencias alélicas, se obtuvieron valores con los cuales hoy realizamos los cálculos de paternidad y criminalística. Sin embargo, actualmente se calcula que los laboratorios forenses cuentan con aproximadamente 300.000 tipificaciones entre casos de paternidad, así como aproximadamente 80.000 tipificaciones realizadas por los laboratorios privados. Hasta ahora se han hecho caracterizaciones aisladas de diferentes regiones del país, en cuyo análisis es importante la comparación con otras poblaciones relacionadas; pero no se cuenta con los datos actualizados para hacer inferencias de las relaciones con otras poblaciones y con la población general. Por otro lado, en el área forense, la ausencia de reportes de frecuencias alélicas en algunas regiones, puede llevar a un cálculo inexacto de probabilidades de exclusión en la identificación de presuntos padres, agresores, desaparecidos, fallecidos, entre otras.

Es así como la cantidad de información disponible y la solución de estas necesidades requiere fundamentalmente de una nueva generación de herramientas informáticas y plataformas de bases de datos. La propuesta es el desarrollo de una plataforma web de información genética humana nacional para marcadores moleculares, que contemple aspectos en genética forense y epidemiología. Es así, como la construcción de esta aplicación proporcionará una herramienta computacional con amplia capacidad de análisis y organización, que facilitará la recolección de la información. En consecuencia, sería posible construir una muestra referencial actualizada de Colombia para los distintos marcadores moleculares, además de tener acceso a los datos almacenados; con el fin de construir nuevas hipótesis sobre el estado genético de la población colombiana, a nivel regional y local; generando así, estimaciones acertadas de los análisis genético poblacionales, como por ejemplo una estimación robusta de la estructura genético-poblacional del país; por otro lado, puede usarse también, para el almacenamiento de otros tipos de datos, como SNPs y secuencias, permitiendo al mismo tiempo, que la comunidad académica tenga datos disponibles y actualizados. Además, sería importante poder detectar poblaciones atípicas para los estadísticos representativos de los distintos marcadores, con respecto a la muestra referencial, su detección es clave en la búsqueda de factores de riesgo específicos en estudios epidemiológicos.

Es posible así, estudiar fenómenos como selección, subestructura genética y mutación en los fenotipos constitucionales. Además, se pueden evaluar los factores de riesgo presentes en el medio ambiente, que pueden interactuar con la constitución genética de una población (Wyszynski, 1998).

En la construcción de la base de datos, es necesario asegurar la validez y calidad de los datos, debido a su importancia y sensibilidad. Esto es posible implementando mecanismos que evalúen constantemente la información reduciendo el sesgo. La manera de hacerlo, es con una correcta selección de la información, codificación, crítica e imputación, verificación, validación y consistencia de la información; también, asegurando la privacidad de la base, permitiendo un acceso específico y limitado, dependiendo los tipos de usuario. Por otro lado, es importante que se tengan en cuenta los aspectos éticos para las bases de datos, consagrados en la Declaración Internacional sobre los Datos Genéticos Humanos de la UNESCO Paris, 16 de Octubre de 2003.

Nuestro laboratorio ha desarrollado software para manejo de perfiles genéticos, logrando integrar los archivos de resultados de secuenciadores automáticos de la casa Hitachi: Modelo FMBIO y de Applied Biosystems tales como el modelo ABI 310 y toda la serie 31: 3100, 3130 y 3130 xl. Además de haber diseñado, en 1996 la primera base de datos con la que se almacenó información de 104.000 personas provenientes de casos de paternidad, también se diseñó en el año 1998, la aplicación APC para analizar casos de paternidad a partir de grupos sanguíneos y los primeros microsatélites que se usaron en el país. Con base a la aplicación en el año 2003 se analizaron 28000 casos de filiación y posteriormente se recolectaron aproximadamente 80000 tipificaciones que reposan actualmente en el Instituto Colombiano de Bienestar Familiar. Conjuntamente, la integración de nuestro conocimiento; en trabajo de campo y en las diferentes variables genéticas, demográficas, geográficas y modelos matemáticos utilizados en la genética de poblaciones, nos permitirán el desarrollo de una gran bases de datos con una estructura sólida integrada con todos estos aspectos, importantes en el desarrollo de la genética de poblaciones.

3.5. Unificación con el sistema de salud y la comunidad.

La mayoría de los estudios de “asociación genética”, están siendo utilizados para descubrir el componente genético que subyace a las enfermedades comunes de alta prevalencia, como por ejemplo, la diabetes mellitus, la enfermedad coronaria y el cáncer. Estos tipos de estudios, generalmente se trata de cohortes prospectivas o de tipo casos-controles, en los cuales se establece el peso relativo del componente genético con respecto a otros factores como el ambiental y el riesgo a desarrollar la enfermedad. Habitualmente, se utilizan como marcadores genéticos a los polimorfismos simples puntuales (SNPs). Estas variaciones pueden ser en sí mismas funcionales y estar relacionadas a la fisiopatología de la enfermedad, pero en la mayoría de los casos, son utilizadas para el mapeo y ubicación de los verdaderos sitios relevantes (Sevilla, 2007). Según estas interpretaciones, los análisis de estos polimorfismos están solamente enfocados a la naturaleza propia del gen con la enfermedad, y no se tienen en cuenta variables externas que afectan el comportamiento del gen en la población en general; ya que se enfocan solamente en muestras de pacientes, sin un mayor análisis demográfico, genealógico, ancestral y poblacional de cada individuo.

Antes de realizar este tipo de estudios tenemos que comprender las fuerzas de cambio evolutivo propias de cada población, como son: el efecto fundador, la deriva, migración, selección, cuello de botella, mutación. Si analizamos la población desde los microsatélites, podemos establecer el tipo de fuerza evolutiva presente en esa población, y de esta manera establecer mapas, planos, rutas y modelos genético poblacionales, para interpretar a mayor escala los datos obtenidos de un estudio de asociación simple a una patología específica.

Es por ello, que antes de realizar un estudio de asociación con una determinada enfermedad, considerada como problema prioritario de salud, es necesario establecer una base poblacional; incluyendo el componente genético, ambiental, cultural y demográfico; y evaluando estas variables en las poblaciones como unidades y no por regiones o etnias, ya que algunas características de la población, estarán condicionadas por la renovación en el cambio de una generación a la siguiente bajo diferentes entornos ambientales, sociales y culturales. De esta manera se puede explicar con certeza el comportamiento de la patología, en una población en particular.

Los estudios en genética de poblaciones se deben integrar totalmente con los estudios de asociación

genética involucrados en la explicación de las enfermedades, considerados problemas prioritarios en salud; aspecto que no ha sido considerado en este tipo de estudios.

Conclusión.

Nuestro mayor interés está centrado hacia un mejor uso de la información genética; seguir desconectados en la práctica diaria de los usuarios de los diferentes servicios, y que nuestros trabajos sigan siendo considerados como valores de referencia simplemente utilizados para realizar cálculos a diferentes niveles, representa un estancamiento de la disciplina, sin que se tenga un mayor impacto en otras áreas de las ciencias biomédicas y mucho menos en la población. Hemos estado abriendo un camino de comunicación entre las comunidades campesinas, indígenas y minorías étnicas, que creemos en un futuro dará a nuestros usuarios, un mejor nivel de comprensión de la genética humana en el país. A diferentes niveles educativos, se considera que uno de nuestros mayores problemas es la falta de identidad que nos impide defender y cuidar nuestro patrimonio, creemos también, que la construcción de sistemas de información, mejores prácticas en el abordaje de nuestros usuarios de filiación e identificación, una mejor interpretación de la variabilidad de nuestras poblaciones y la generación de mejores conjuntos de datos, son un sustrato para hacer de la genética de poblaciones en Colombia, un terreno fértil para el desarrollo de modelos matemáticos y estadísticos, que permitan una mejor utilización de la tecnología en biología molecular que poseemos, y para un acercamiento de los profesionales en ciencias biomédicas a los profesionales de las ciencias sociales.

Bibliografía.

Begoña M, Sánchez D, González A. 2005. El Estudio de Polimorfismos de ADN a Partir de Restos Óseos y Dientes y sus Aplicaciones en la Identificación de Desaparecidos. *Revista Aragonesa de Medicina Legal* 7: 163:184.

Bedoya G, Montoya P, García J, Soto I, Bourgeois S, Carvajal L, Labuda D, Alvarez V, Ospina J, Hedrick PW, et al. 2006. Admixture Dynamics in Hispanics: a Shift in the Nuclear Genetic Ancestry Of A South American Population Isolate. *Proceedings of the National Academy of Sciences of the United States of America* 103:7234-9.

Butler J, Shen Y, McCord B. 2006. The development of reduced size STR amplicons as tool for analysis of degraded DNA. *J Forensic Sci.* 48: 1054-64.

Callegari Jacques SM, and Salzano FM. 1979. Demography and Genetics of the Krahó and Gorotire Indians of Brazil. *Journal of Human Evolution* 8:513-522.

Carracedo A, Salas A, Lareu M. Problemas y retos de futuro de la genética forense en el siglo XXI. *Cuad Med Forense.* 2010;16:31-35.

Carvajal-Carmona LG, Ophoff R, Service S, Hartiala J, Molina J, Leon P, Ospina J, Bedoya G, Freimer N, and Ruiz-Linares A. 2003. Genetic Demography of Antioquia (Colombia) and the Central Valley of Costa Rica. *Human genetics* 112:534-41.

Carvajal-Carmona LG, Soto ID, Pineda N, Ortíz-Barrientos D, Duque C, Ospina-Duque J, McCarthy M, Montoya P, Alvarez VM, Bedoya G, et al. 2000. Strong Amerind/White sex bias and a possible Sephardic contribution among the founders of a population in northwest Colombia. *American journal of human genetics* 67:1287-95.

Crow JF, and Mange AP. 1965. Measurement of Inbreeding from the Frequency of Marriages Between Persons of the Same Surname. *Revista Eugenics Quarterly* 12:199-203.

De G, and Rica C. 1981. Demografía y Genética de la Población Amerindia Guayrní de Limoncito, Costa Rica. *Revista de Biología Tropical* 29:23-31.

Gómez MV, Reyes ME, Cárdenas H, García O. 2003 Genetic variation for 12 STR loci in a Colombian population (Department of Valle del Cauca). *J Forensic Sci.* 48(3): 687-688.

Jeffreys A, Wilson V, Thein S. 1985 Individual-specific “Fingerprints” of Human DNA. *Nature.* 314:67-73.

Jujuy UD. 2004. Cuadernos de la Facultad de Humanidades y Ciencias Sociales. Cuadernos de la Facultad de Humanidades y Ciencias Sociales 22:171-194.

Keyeux G, Rodas C, Gelvez N, and Carter D. 2002. Possible Migration Routes into South America Deduced from Mitochondrial DNA Studies in Colombian Amerindian Populations. *Human Biology* 74:211-233.

López N. Estructura Genética de la Población colombiana segunda muestra. . Tesis Maestría Universidad Nacional de Colombia, Facultad de medicina. 1993

Mesa NR, Mondragón MC, Soto ID, Parra MV, Duque C, Ortiz-Barrientos D, García LF, Velez ID, Bravo ML, Múnera JG, et al. 2000. Autosomal, mtDNA, and Y-chromosome Diversity in Amerinds: pre- and post-Columbian Patterns of Gene flow in South America. *American journal of Human Genetics* 67:1277-86.

Michan L. 12 de octubre 2010. Sesión de genética y tecnología: Del alfabeto del gen a los pixeles en pantalla. Cátedra de Sede José Celestino Mutis. Todo lo que Usted quiere saber de genética y nunca se atrevió a preguntar. Universidad Nacional de Colombia. [Internet]: Available from <http://laylamichanunam.blogspot.com/>

Pérez AA, Salzano FM. 1978. Evolutionary Implications of the Ethnography and Demography of Ayoreo Indians. *Journal of Human Evolution* 7:253-268..

Rodas C, Gelvez N, and Keyeux G. 2003. Mitochondrial DNA Studies Show Asymmetrical Amerindian Admixture in Afro-Colombian and Mestizo Populations. *Human Biology* 75:13-30.

Rodríguez, J. 1994 *Introducción a la Antropología Forense: Análisis e interpretación de restos óseos humanos*. Universidad Nacional de Colombia. Anaconda Editores. Bogotá.

Rojas W, Parra MV, Campo O, Caro MA, Lopera JG, Arias W, Duque C, Naranjo A, García J, Vergara C, et al. 2010. Genetic Makeup and Structure of Colombian Populations by Means of Uniparental and Biparental DNA Markers. *American journal of physical anthropology* 143:13-20.

Salas A, Acosta A, Alvarez-Iglesias V, Cerezo M, Phillips C, Lareu MV, And Carracedo A. 2008. The mtDNA Ancestry of Admixed Colombian Populations. *American Journal of Human Biology : The Official Journal of the Human Biology Council* 20:584-91.

Sánchez E. 2001. Biodemografía: Una Apuesta para el Estudio Biológico de las Poblaciones. *Revista de Demografía Histórica* XIX: 71-86.

Sandoval C. Estructura Genética de la población Colombiana. . Tesis Maestría Universidad Nacional de Colombia, Facultad de medicina. 1993

Sevilla, S. 2007. Metodología de los estudios de asociación genética. *Insuficiencia Cardíaca*. 2; 3: 111-114.

Solla L. 2004 Aplicaciones forenses del ADN Comisaría General de Policía Científica Laboratorio de ADN. [Internet] Available from: www.cej.justicia.es/pdf/publicaciones/fiscales/FISCAL333.pdf.

Pereira R, Phillip C, Alves C, Amorim A, Carracedo A, Gusmão L 2009 Insertion/Deletion Polymorphisms: a Multiplex Assay and Forensic Applications. *Forensic Sci Int Genetics Suppl* 2:513-515.

QIAGEN. 2010. Investigator DIPplex Handbook Sample and Assay Technologies.

Wyszynski D. 1998 La epidemiología genética: disciplina científica en expansión. *Pan Am J Public Health* 3(1), 26-34

Weber JL, David D, Heil J, Fan Y, Zhao C, Marth G 2002. Human Diallelic Insertion/Deletion Polymorphisms. *Am J Hum Genet* 71:854-862.

XUE Y, Tyler-Smith C. The hare and the tortoise: One small step for four SNPs, one giant leap for SNP-kind. *Forensic Sci Int Genet.* 2010;4:59-61

Yunis E, Yunis J. El ADN en la identificación humana. Editorial Temis S.A. 2002. Bogotá.

Anexo No. 1: Guía de Laboratorio para Análisis Genético.

Summary Guide

Population	Bogotá												
System	D8S1179	D21S11	D7S820	CSF1PO	D16S539	D3S1358	TH01	D13S317	TPOX	D18S51	VWA	FGA	D5S818
Alleles number	10	22	10	10	8	7	7	10	9	20	12	21	9
Effective alleles number	5	6	5	4	5	4	4	6	3	8	4	8	4
CV	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
Sample number	2	48	16	22	4	2	17	12	7	53	10	26	24
Sample size	28	650	213	250	56	29	208	152	86	709	134	339	308
Polimorphism alleles total	19	70	22	15	10	20	21	14	17	42	28	98	10

Table N°1. Alleles number, effective alleles, coefficient of variation, sample number, sample size and polymorphism alleles total for 13 STRs loci.

STR'S	Category	No. Alleles	Alleles found	% alleles found.	No. Effective alleles	% Effective alleles	Unique Alleles	No. Ind.	Promedio	Min	Max
D8S1179	Compuesto	19	10	52.63%	5	50.00%	1	28	0.0108	0.0019	0.0199
D21S11	Complejo	70	22	31.43%	6	27.27%	3	650	0.0002	0.0000	0.0007
D7S820	Simple	22	10	45.45%	5	50.00%	2	213	0.0013	0.0001	0.0030
CSF1PO	Simple	15	10	66.67%	4	40.00%	1	250	0.0011	0.0001	0.0029
D16S539	Simple	10	8	80.00%	5	62.50%	1	56	0.0072	0.0009	0.0108
D3S1358	Compuesto	20	7	35.00%	4	57.14%	0	29	0.0141	0.0035	0.0209
TH01	Simple*	21	7	33.33%	4	57.14%	1	208	0.0016	0.0001	0.0029
D13S317	Simple*	14	10	71.43%	6	60.00%	2	152	0.0020	0.0002	0.0040
TPOX	Simple	17	9	52.94%	3	33.33%	3	86	0.0033	0.0005	0.0068
D18S51	Simple	42	20	47.62%	8	40.00%	3	709	0.0002	0.0000	0.0005
VWA	Compuesto	28	12	42.86%	4	33.33%	2	134	0.0018	0.0002	0.0039
FGA	Compuesto	98	21	21.43%	8	38.10%	5	339	0.0005	0.0000	0.0015
D5S818	Simple	10	9	90.00%	4	44.44%	1	308	0.0009	0.0001	0.0019

Table N°2. Information of Category and Confidence Interval for each STR system

Summary Guide

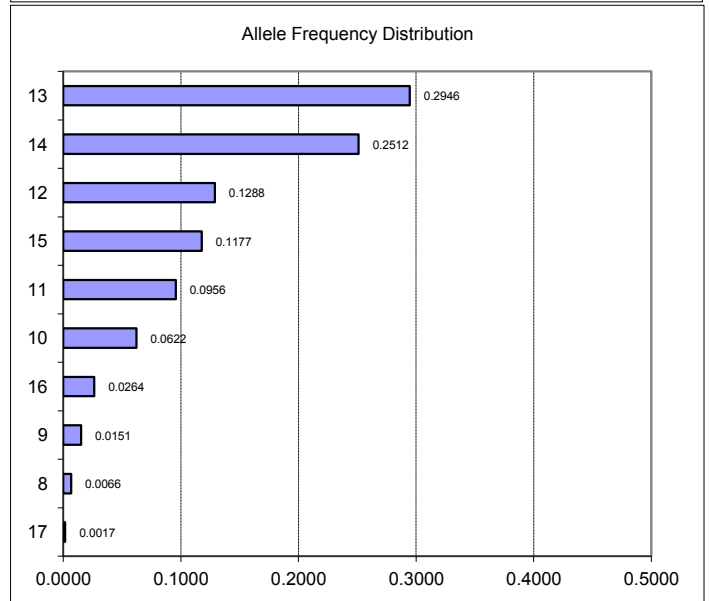
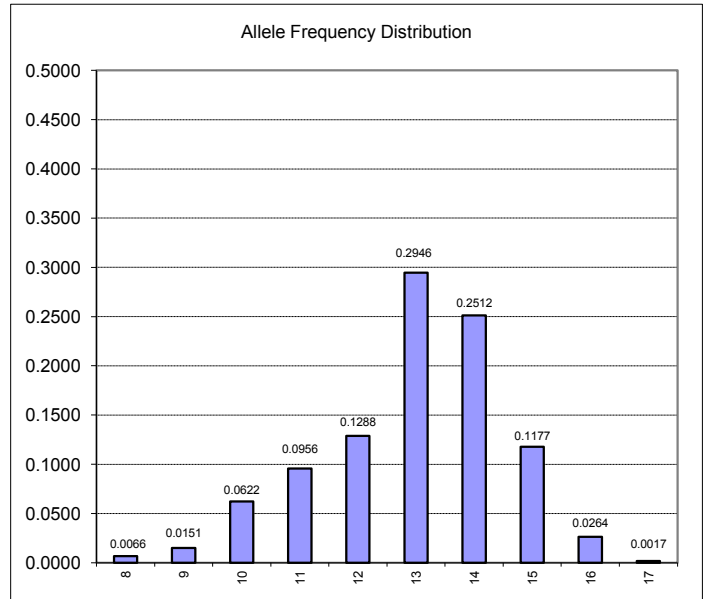
Allele	System												
	D8S1179	D21S11	D7S820	CSF1PO	D16S539	D3S1358	TH01	D13S317	TPOX	D18S51	VWA	FGA	D5S818
a4							0.0006						
a6			0.0009				0.3493	0.0005	0.0047				
a7			0.0177	0.0083			0.2510	0.0003	0.0018				0.0450
a8	0.0087		0.1223	0.0087	0.0136		0.0867	0.1101	0.4570				0.0077
a9	0.0153		0.0859	0.0188	0.1426		0.1279	0.1615	0.0693				0.0848
a9,3							0.1773						
a10	0.0638		0.2554	0.2389	0.1663		0.0071	0.0640	0.0481	0.0060			0.0718
a10,2				0.0008						0.0013			
a11	0.1053		0.2830	0.2715	0.2431			0.2023	0.2931	0.0108	0.0015		0.3794
a12	0.1252		0.1915	0.3652	0.2626			0.2641	0.1228	0.1031	0.0006		0.2727
a13	0.2827		0.0349	0.0753	0.1407	0.0072		0.1291	0.0023	0.1255	0.0023		0.1316
a14	0.2617		0.0080	0.0120	0.0298	0.1049		0.0657	0.0009	0.1650	0.0713		0.0067
a14,2										0.0007			
a15	0.1123			0.0007	0.0014	0.3683		0.0023		0.1494	0.0966		0.0003
a15,2											0.0009		
a16	0.0232					0.2484				0.1374	0.3221		
a17	0.0017					0.1503				0.1418	0.2724	0.0042	
a18						0.1167				0.0570	0.1651	0.0070	
a19						0.0043				0.0439	0.0611	0.0700	
a20										0.0325	0.0056	0.0810	
a20,2										0.0004			
a21										0.0125	0.0006	0.1175	
a22										0.0066		0.1284	
a22,2										0.0003			
a23										0.0033		0.1304	
a24										0.0024		0.1653	
a25		0.0016										0.1616	
a26		0.0024										0.0989	
a27		0.0165								0.0004		0.0227	
a28		0.1054										0.0056	
a29		0.2042										0.0018	
a30		0.2705											
a31		0.0772											
a32		0.0224											
a33		0.0045											
a34		0.0020											
a35		0.0048											
a36		0.0006											
a37		0.0004											

Table N°3. Allele frequencies for 13 STRs loci in a population from Bogotá

Population	Bogotá
System	D8S1179
Alleles number	10
Effective alleles	5

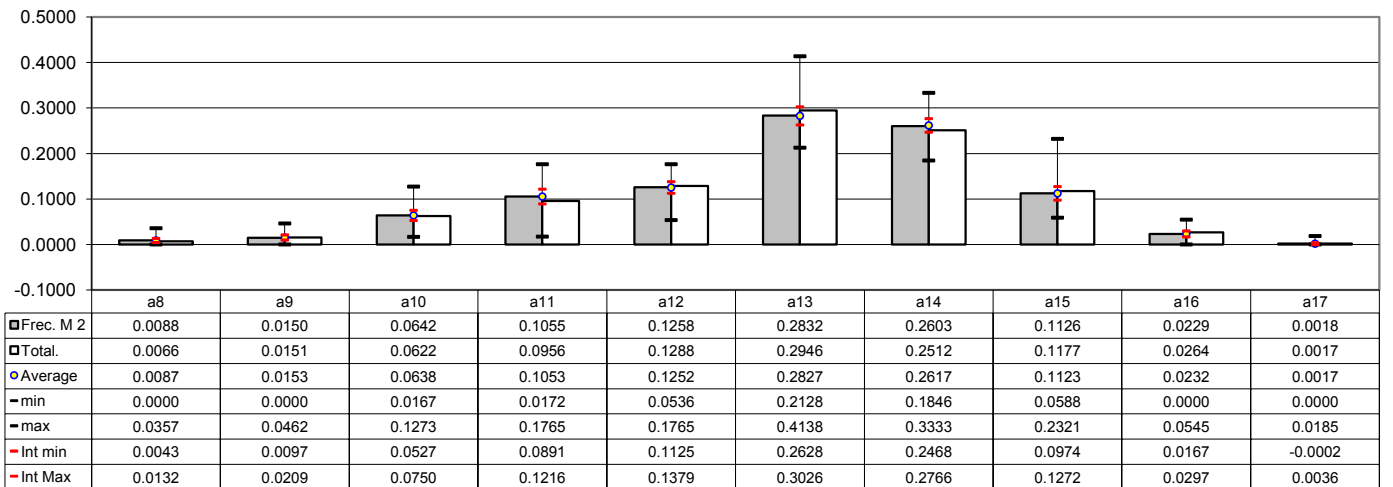
Final frequency.

Allele	M.Weight.	S.D.	Frequency
8	123.29	0.07	0.0066
9	127.32	0.05	0.0151
10	131.41	0.05	0.0622
11	135.49	0.04	0.0956
12	139.73	0.04	0.1288
13	144.25	0.03	0.2946
14	148.71	0.06	0.2512
15	153.16	0.07	0.1177
16	157.51	0.07	0.0264
17	161.72	0.05	0.0017
18	165.84	0.07	
19	169.93	0.05	



Single Alleles	
In the Database	In the Sample
	17

Confidence Intervals to Frequency Allele Distribution

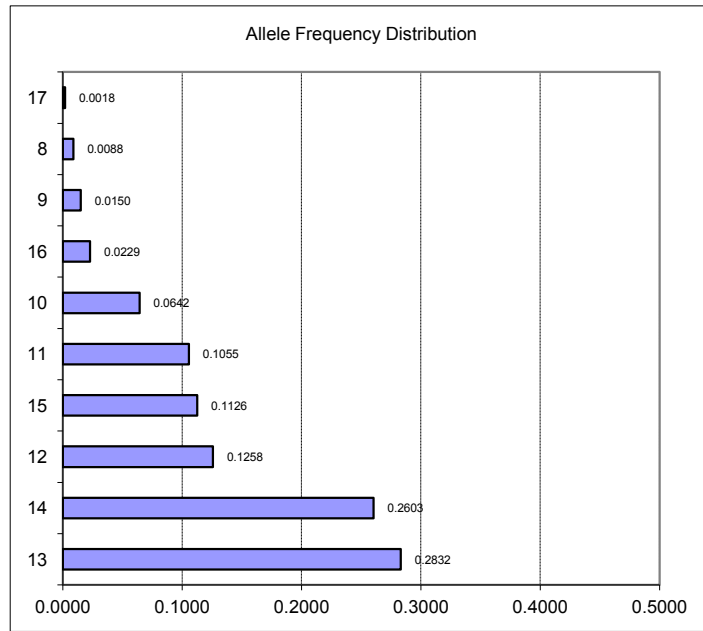


Specific sample analysis.

Sampling Specific.	
Population	Bogotá
System	D8S1179
Alleles number	10
Effective alleles number	5
CV	0.05
Sample number	2
Sample size	28
Polimorphism alleles total	19

EXPERIMENT	(Todas)
Sample number	2

Allele frequency summary	
Alleles	Total
13	0.2832
14	0.2603
12	0.1258
15	0.1126
11	0.1055
10	0.0642
16	0.0229
9	0.0150
8	0.0088
17	0.0018



Single Alleles in the Sampling

EXPERIMENT	(Todas)
Sample number	2

Max	Total
Alleles	Total
17	1
8	2
16	3
9	3
10	7
11	9
12	12
15	13
14	21
13	24

Sampling and CV.

Sample selection
Analysis of CV.

Alleles	Sample No.
14	2
11	2
13	2
15	2
10	3
12	2
17	0
16	44
9	26
8	0

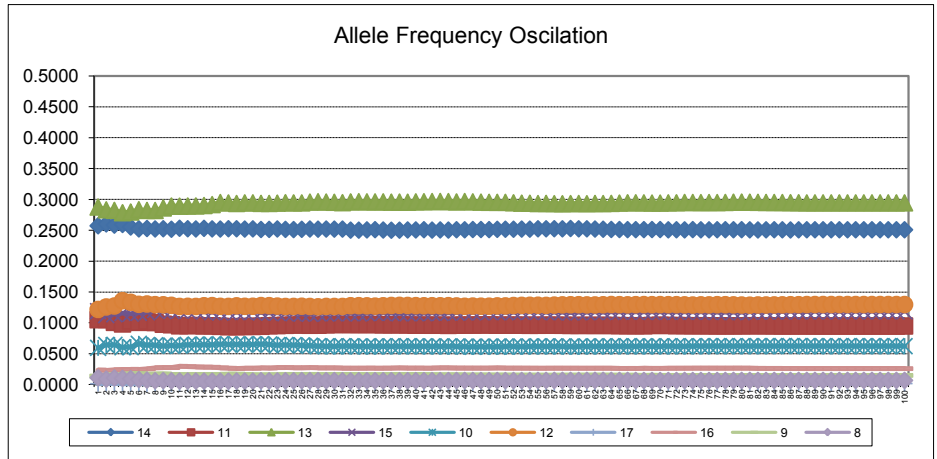
Allelic Distribution Order

Population	Bogotá
System	D8S1179

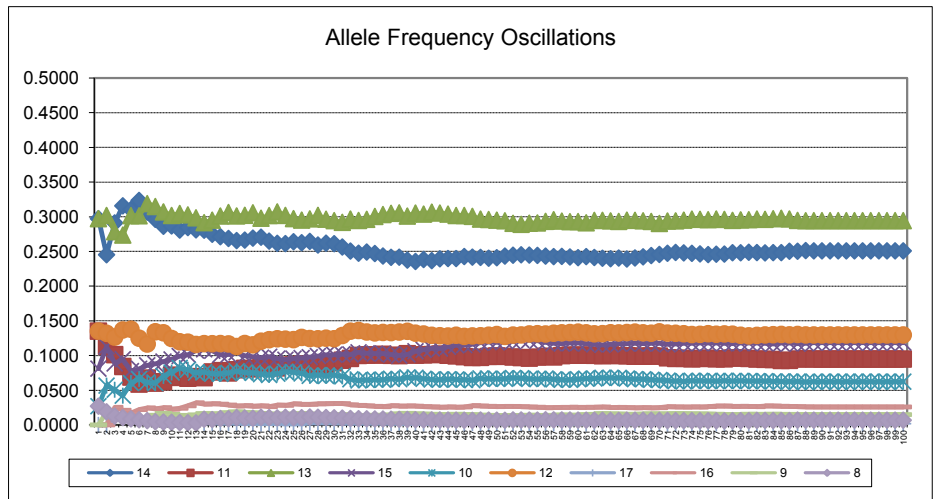
EXPERIMENT	(Todas)
Sample number	(Todas)

Allele frequency summary		
Alleles		Total
	13	0.2946
	14	0.2512
	12	0.1288
	15	0.1177
	11	0.0956
	10	0.0622
	16	0.0264
	9	0.0151
	8	0.0066
	17	0.0017

Compiled Allele Frequency



Allele frequency for a specific experiment.



Analysis for Specific Sampling

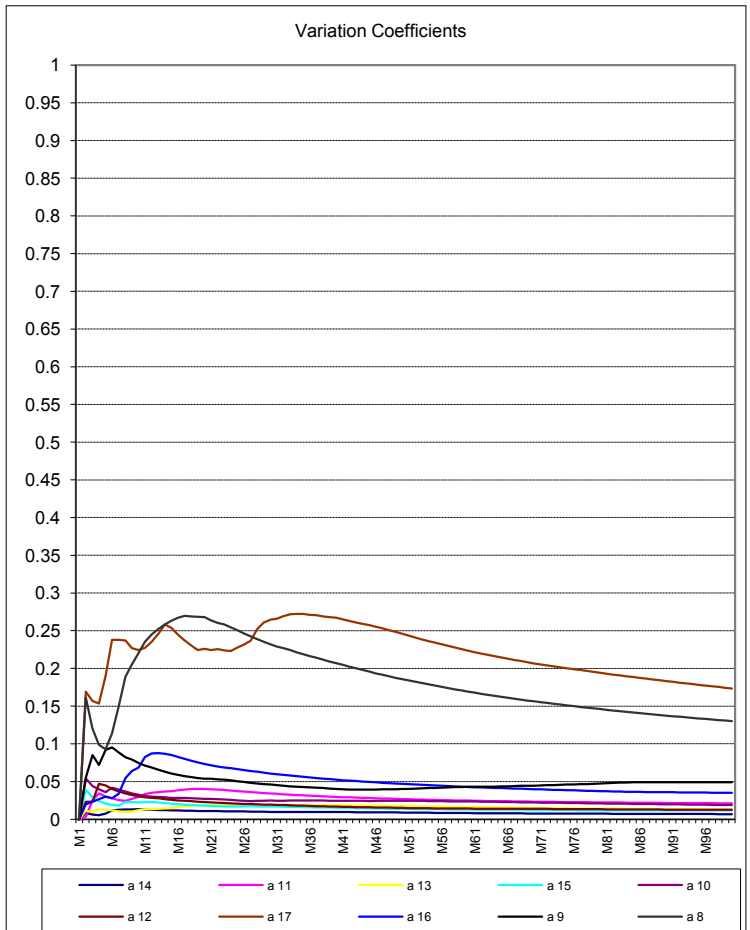
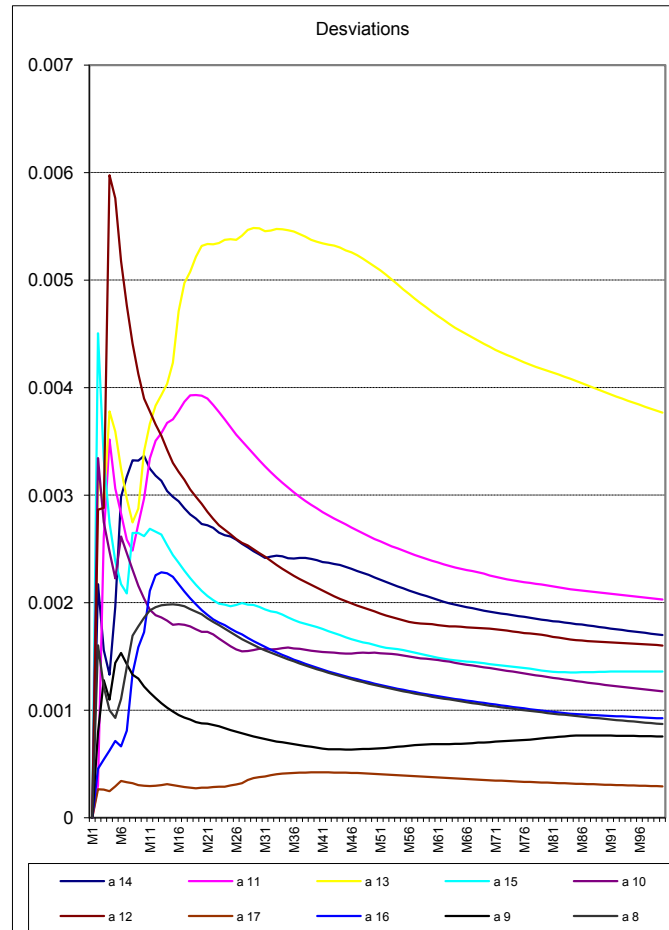
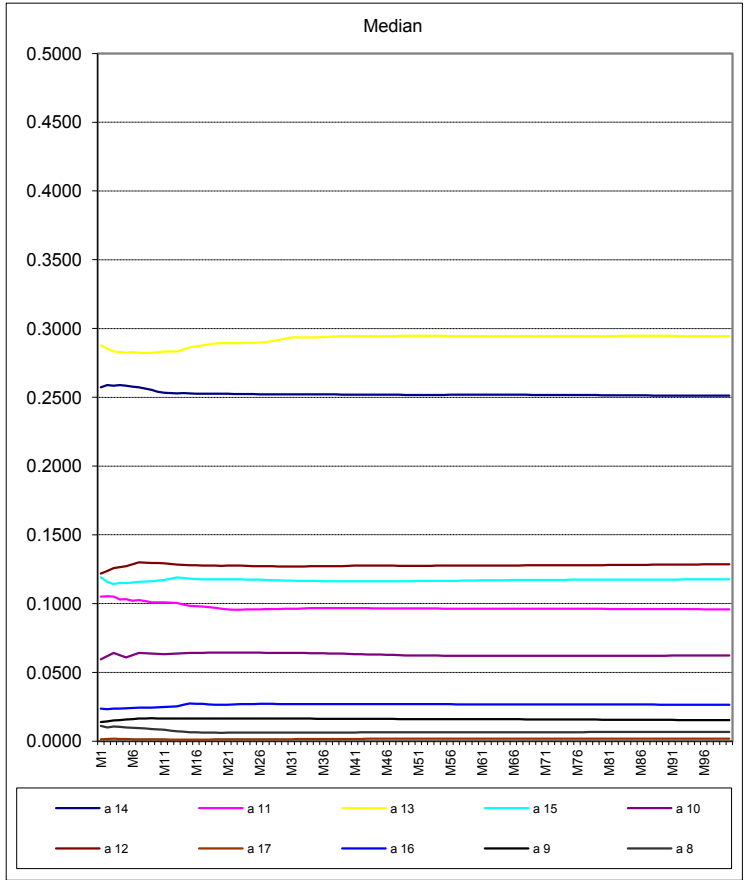
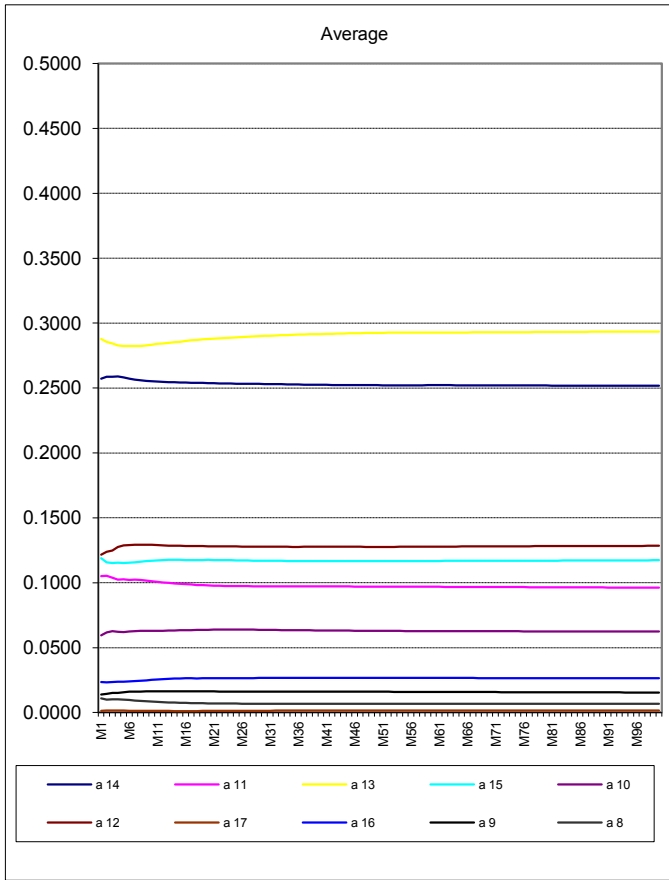
Sample number	2
Sample size	28

Average D	Min	Max
0.0108	0.0019	0.0199

Alleles	Average	min	max	Int min	Int Max	Median	S.D	IC
a8	0.0087	0.0000	0.0357	0.0043	0.0132	0.0000	0.0121	0.0045
a9	0.0153	0.0000	0.0462	0.0097	0.0209	0.0170	0.0152	0.0056
a10	0.0638	0.0167	0.1273	0.0527	0.0750	0.0577	0.0301	0.0112
a11	0.1053	0.0172	0.1765	0.0891	0.1216	0.1095	0.0438	0.0162
a12	0.1252	0.0536	0.1765	0.1125	0.1379	0.1266	0.0343	0.0127
a13	0.2827	0.2128	0.4138	0.2628	0.3026	0.2647	0.0537	0.0199
a14	0.2617	0.1846	0.3333	0.2468	0.2766	0.2697	0.0402	0.0149
a15	0.1123	0.0588	0.2321	0.0974	0.1272	0.1062	0.0403	0.0149
a16	0.0232	0.0000	0.0545	0.0167	0.0297	0.0215	0.0176	0.0065
a17	0.0017	0.0000	0.0185	-0.0002	0.0036	0.0000	0.0052	0.0019

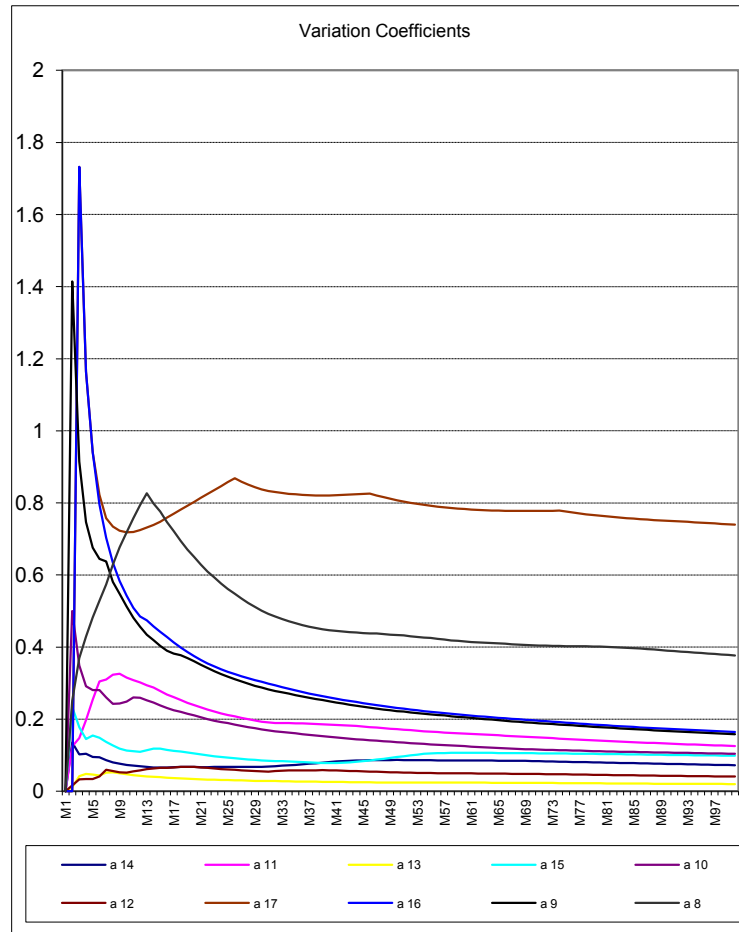
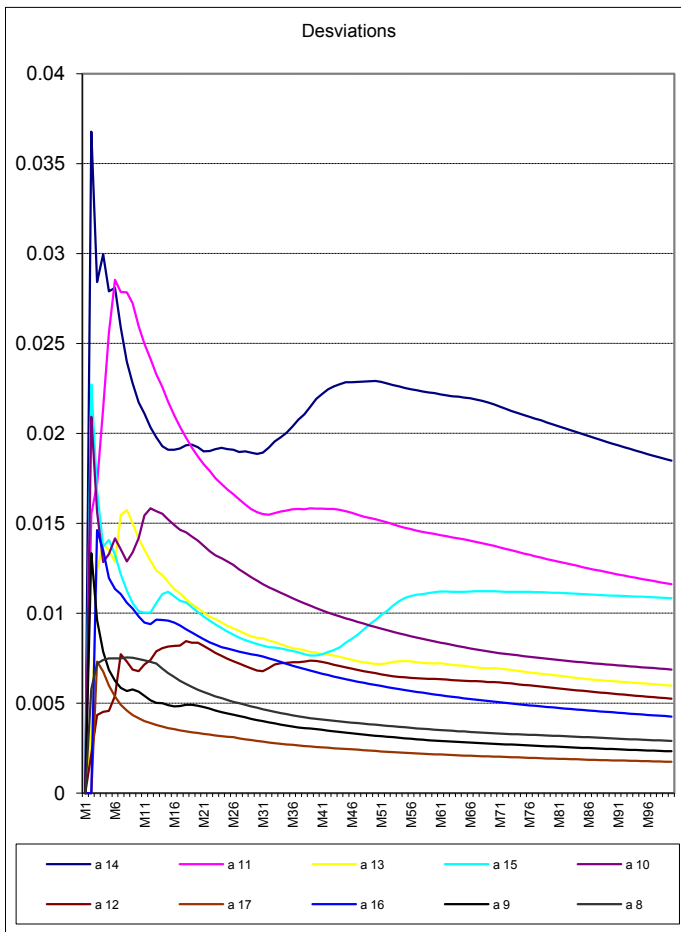
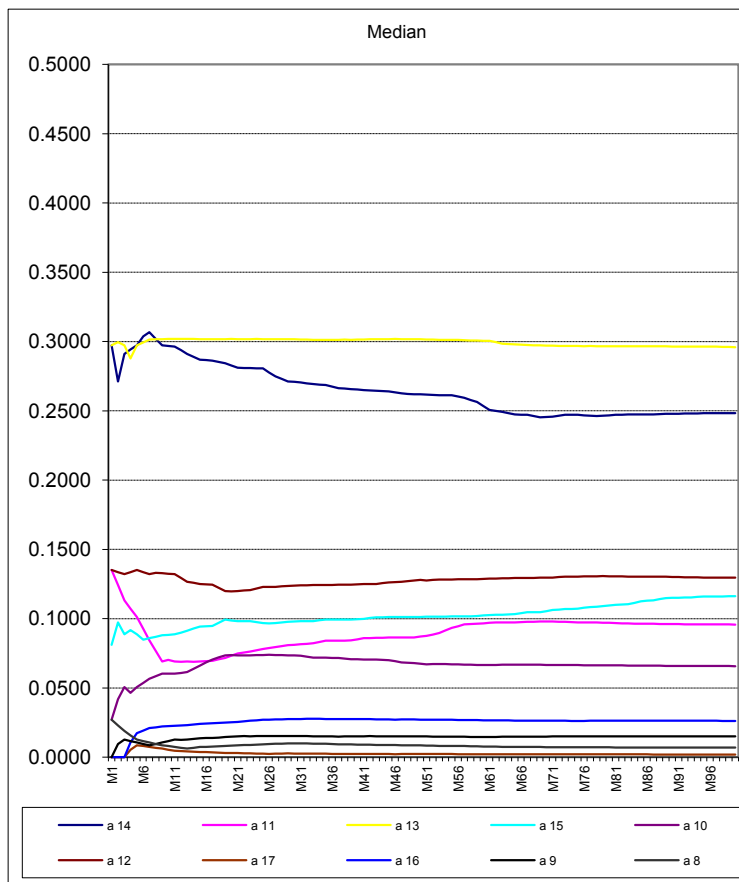
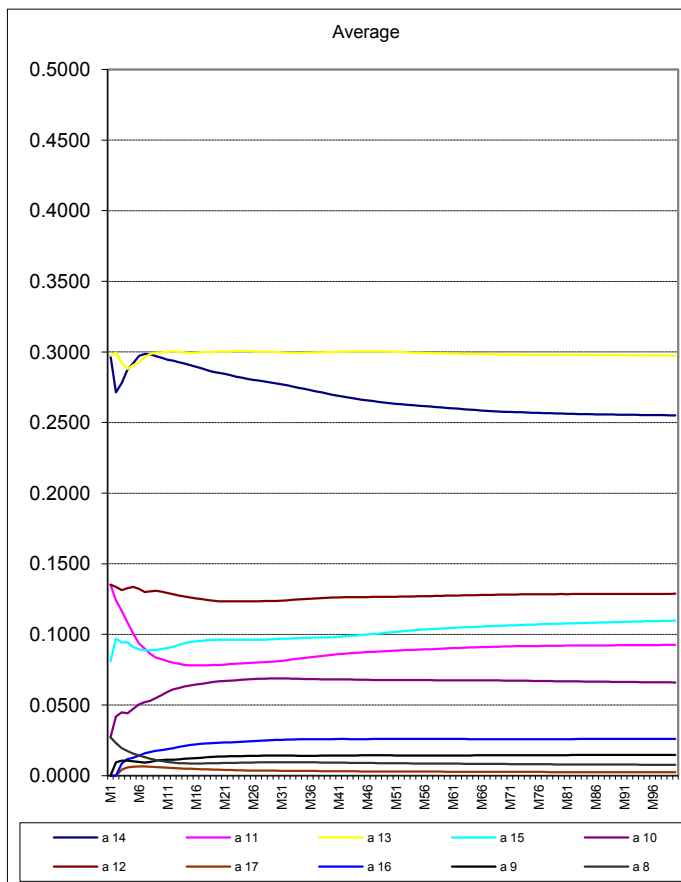
Graphic analysis of polymorphisms.

System D8S1179 Pop. Bogotá



Graphic analysis of polymorphisms.

System D8S1179 Pop. Bogotá



Graphic analysis of polymorphisms.

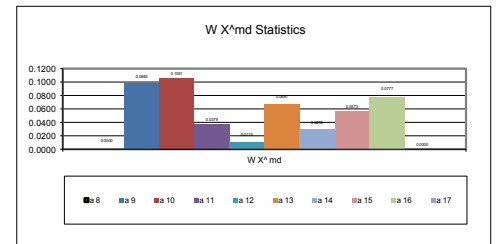
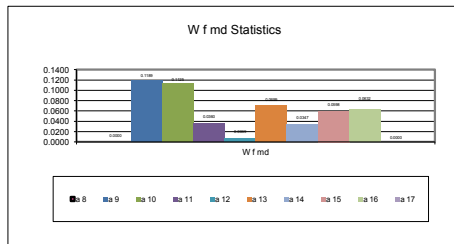
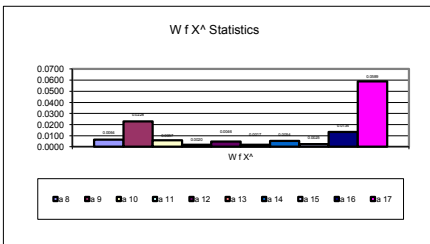
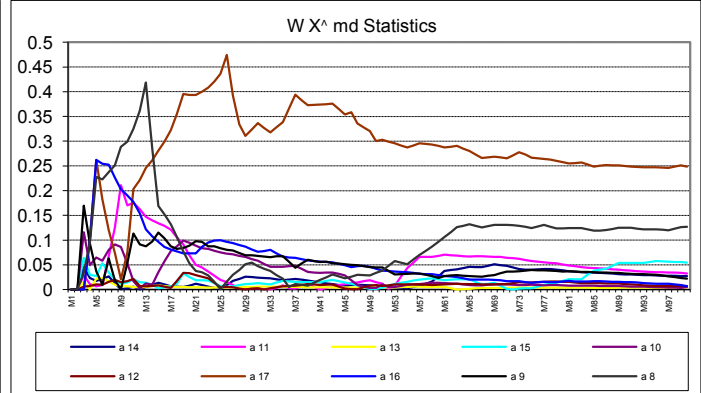
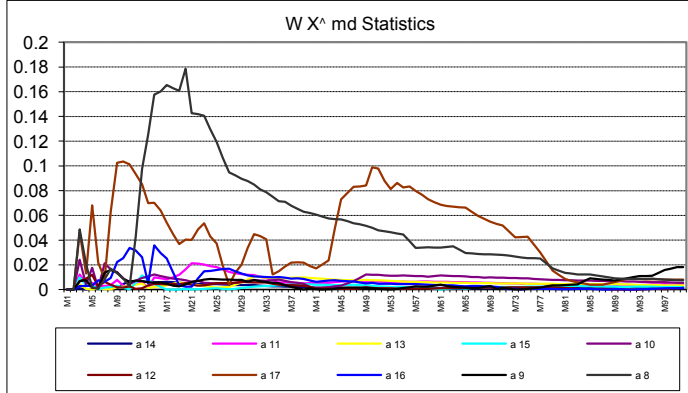
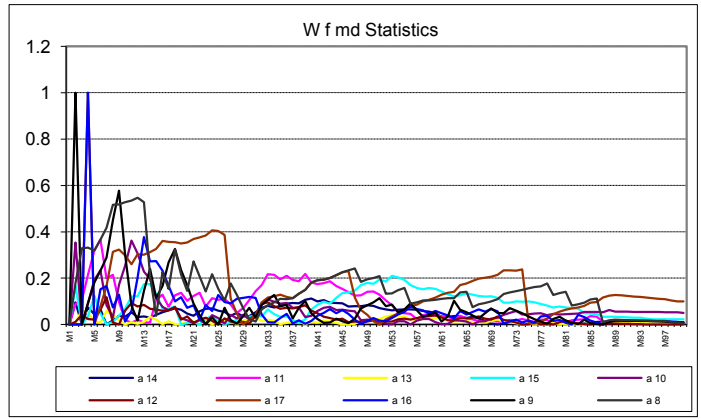
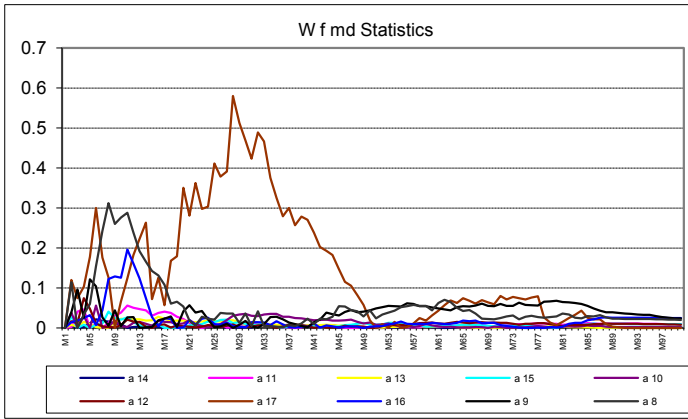
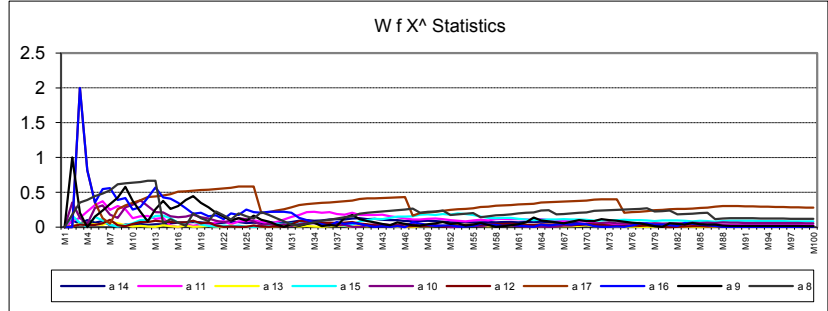
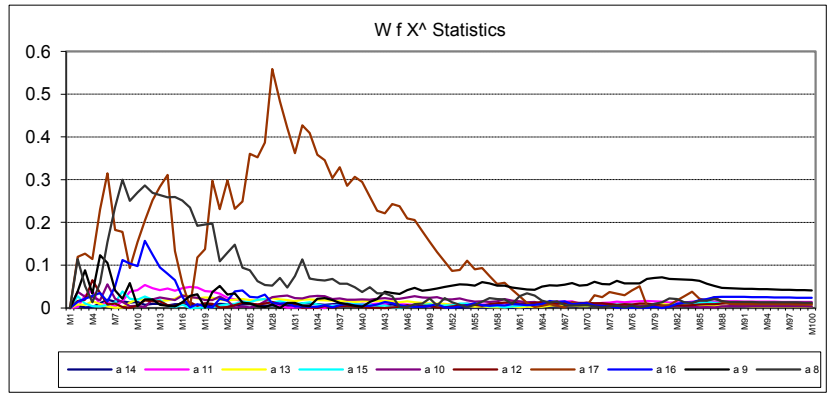
System	D8S1179	Pop.	Bogotá
--------	---------	------	--------

W Statistics to Specific sample

Alleles	W f X [^]	W f md	W X [^] md
a 8	0.0064		
a 9	0.0228	0.1189	0.0983
a 10	0.0057	0.1125	0.1061
a 11	0.0020	0.0360	0.0379
a 12	0.0046	0.0069	0.0115
a 13	0.0017	0.0699	0.0681
a 14	0.0054	0.0347	0.0295
a 15	0.0025	0.0598	0.0572
a 16	0.0134	0.0632	0.0777
a 17	0.0589		

Indeterminate value (Division by zero)

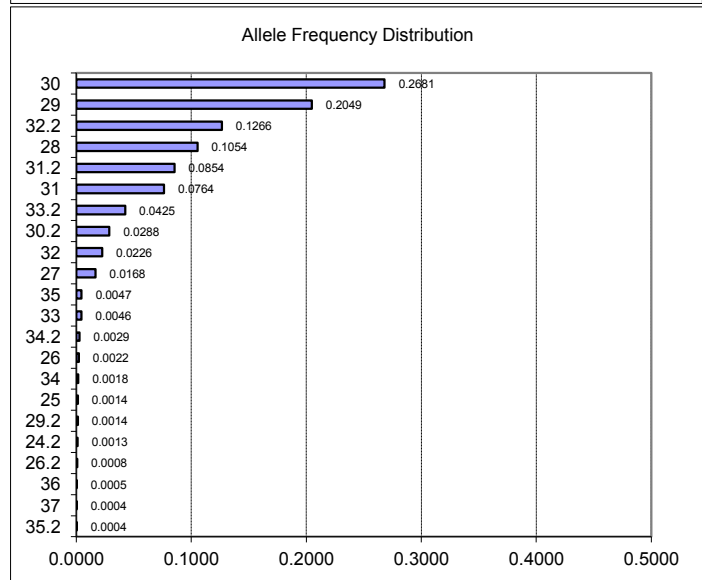
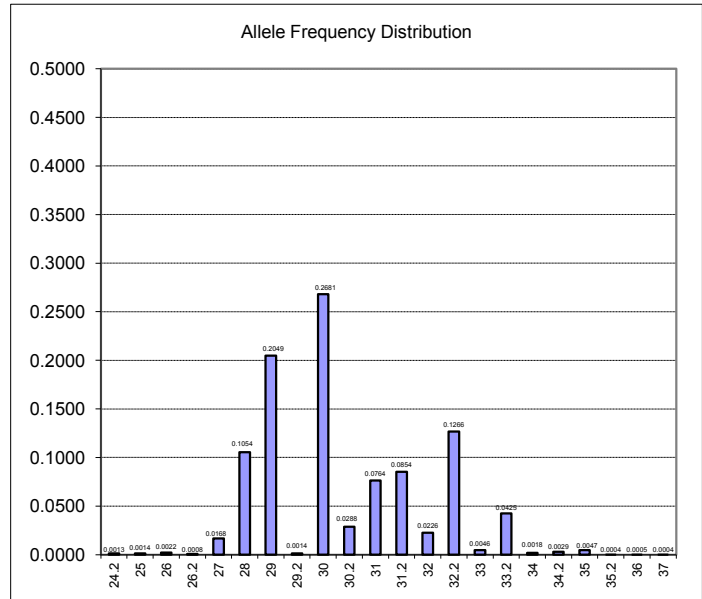
0.0589	0.1189	0.1061
Sampling without repetition		



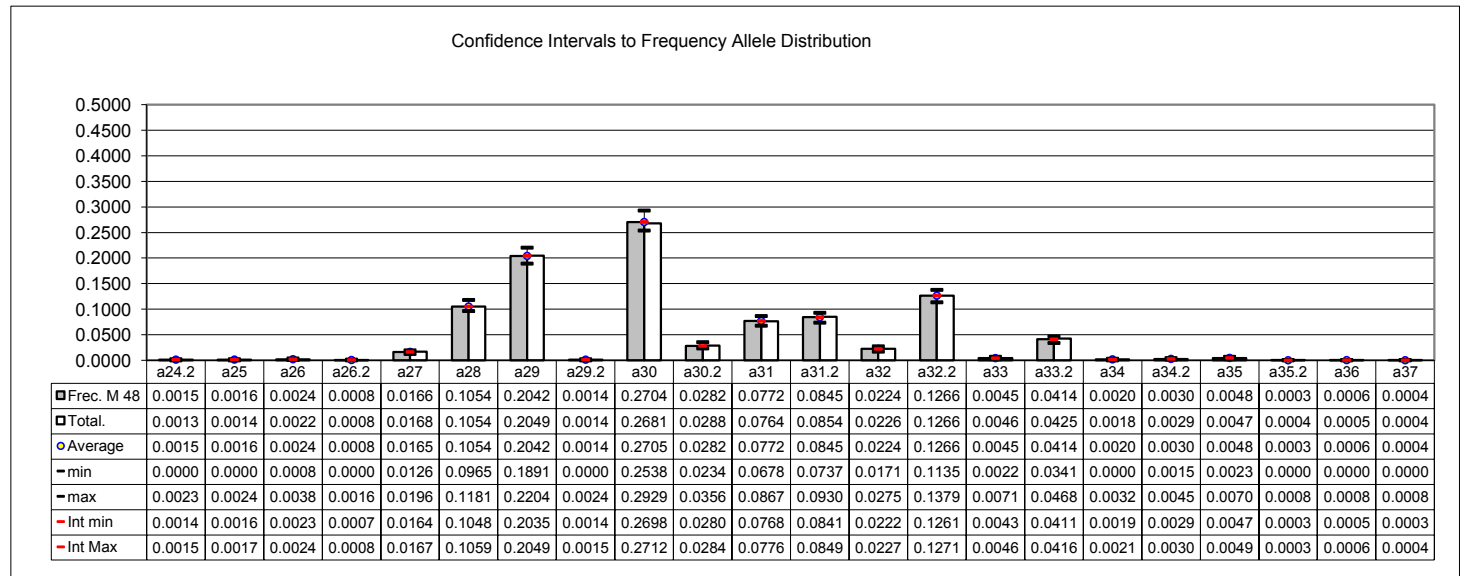
Population	Bogotá
System	D21S11
Alleles number	22
Effective alleles	6

Final frequency.

Allele	M.Weight.	S.D.	Frequency
24	184.86	0.04	
24.2	186.82	0.02	0.0013
25	188.77	0.03	0.0014
26	192.69	0.05	0.0022
26.2	194.69		0.0008
27	196.56	0.04	0.0168
28	200.41	0.05	0.1054
28.2	202.36	0.05	
29	204.32	0.03	0.2049
29.2	206.31	0.02	0.0014
30	208.39	0.07	0.2681
30.2	210.24	0.05	0.0288
31	212.23	0.05	0.0764
31.2	214.14	0.06	0.0854
32	216.14	0.04	0.0226
32.2	218.10	0.04	0.1266
33	220.14	0.05	0.0046
33.1	221.14		0.0020
33.2	222.07	0.04	0.0425
34	224.10	0.07	0.0018
34.2	226.20	0.06	0.0029
35	228.70	0.06	0.0047
35.2	230.01	0.07	0.0004
36	232.04	0.07	0.0005
37	236.00	0.03	0.0004
38	239.94	0.08	



Single Alleles	
In the Database	In the Sample
35.2	35.2
36	36
37	37

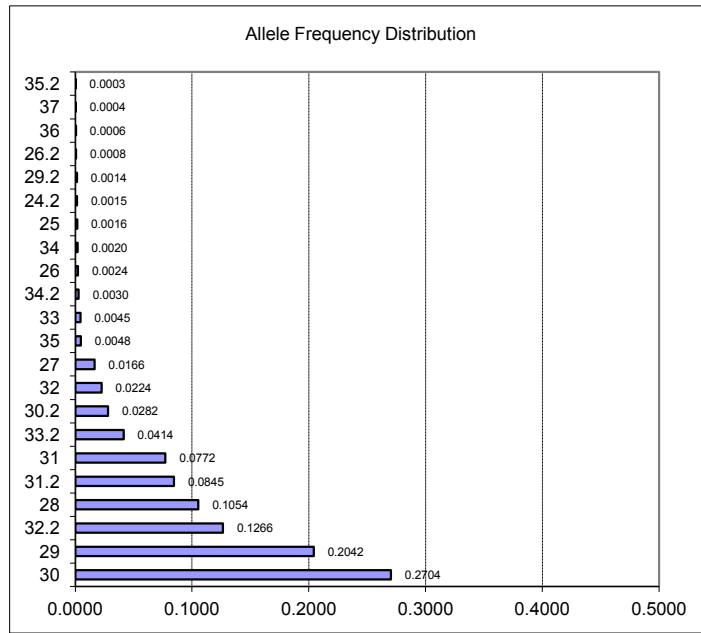


Specific sample analysis.

Sampling Specific.	
Population	Bogotá
System	D21S11
Alleles number	22
Effective alleles number	6
CV	0.05
Sample number	48
Sample size	650
Polimorphism alleles total	70

EXPERIMENT	(Todas)
Sample number	48

Allele frequency summary	
Alleles	Total
30	0.2704
29	0.2042
32.2	0.1266
28	0.1054
31.2	0.0845
31	0.0772
33.2	0.0414
30.2	0.0282
32	0.0224
27	0.0166
35	0.0048
33	0.0045
34.2	0.0030
26	0.0024
34	0.0020
25	0.0016
24.2	0.0015
29.2	0.0014
26.2	0.0008
36	0.0006
37	0.0004
35.2	0.0003



Single Alleles in the Sampling

EXPERIMENT	(Todas)
Sample number	48

Max	Alleles	Total
	37	1
	35.2	1
	36	1
	26.2	2
	29.2	3
	25	3
	24.2	3
	34	4
	26	5
	34.2	6
	33	9
	35	9
	27	26
	32	37
	30.2	47
	33.2	62
	31	115
	31.2	123
	28	159
	32.2	181
	29	289
	30	368

Sampling and CV.

Sample selection
Analysis of CV.

Alleles	Sample No.
24.2	0
25	0
26	0
26.2	0
27	0
28	3
29	2
29.2	0
30	2
30.2	85
31	6
31.2	2
32	83
32.2	48
33	0
33.2	36
34	0
34.2	0
35	0
35.2	0
37	0
36	0

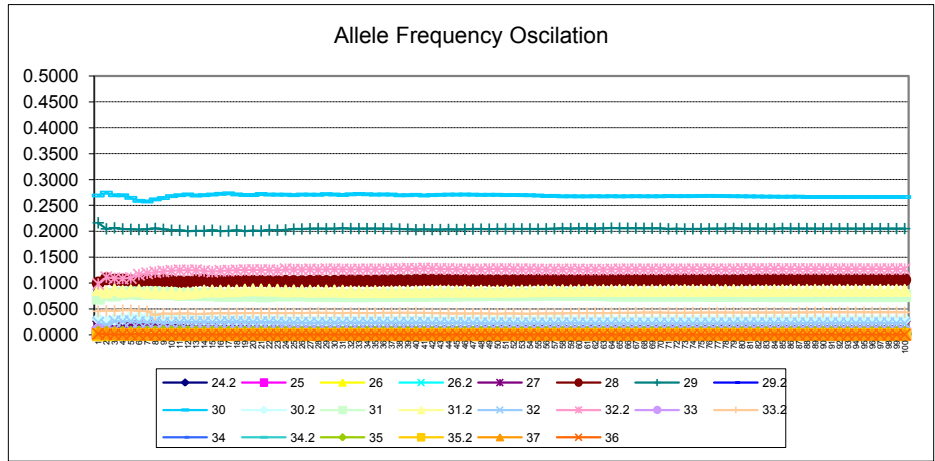
Allelic Distribution Order

Population	Bogotá
System	D21S11

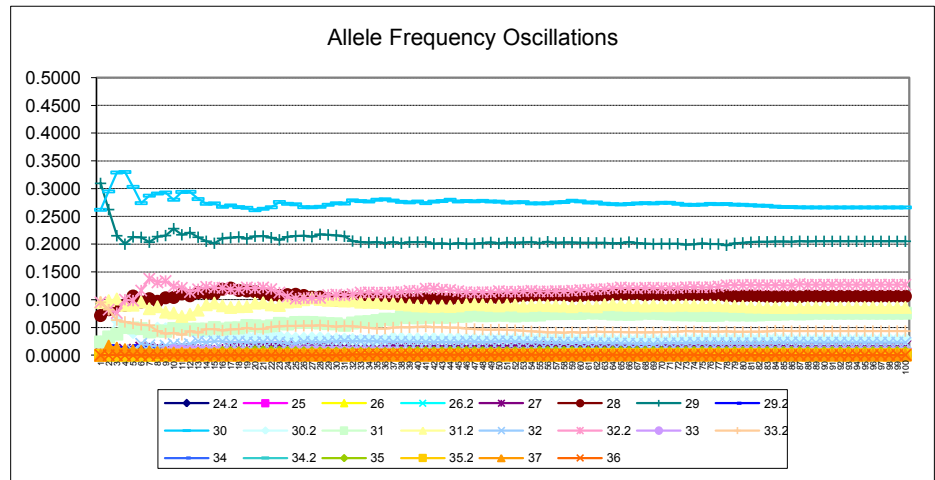
EXPERIMENT	(Todas)
Sample number	(Todas)

Alleles frequency summary	
Alleles	Total
30	0.2681
29	0.2049
32.2	0.1266
28	0.1054
31.2	0.0854
31	0.0764
33.2	0.0425
30.2	0.0288
32	0.0226
27	0.0168
35	0.0047
33	0.0046
34.2	0.0029
26	0.0022
34	0.0018
25	0.0014
29.2	0.0014
24.2	0.0013
26.2	0.0008
36	0.0005
37	0.0004
35.2	0.0004

Compiled Allele Frequency



Allele frequency for a specific experiment.



Analysis for Specific Sampling

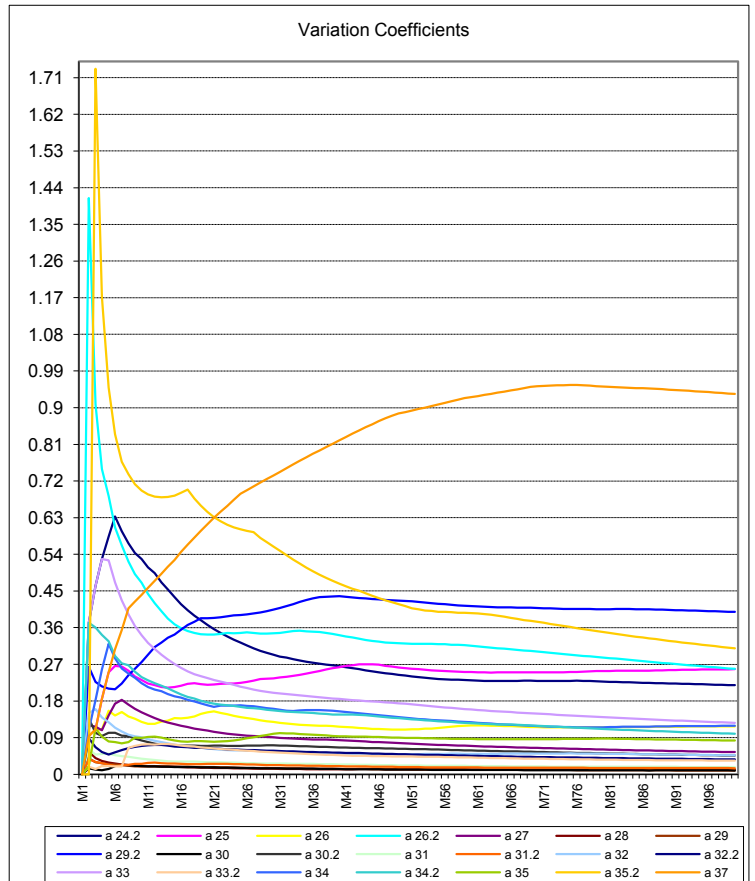
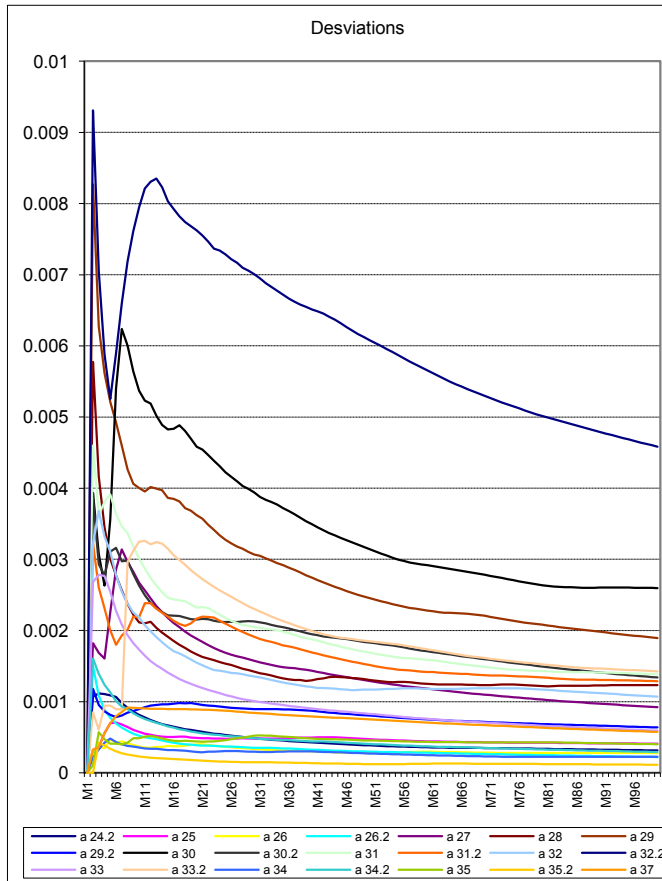
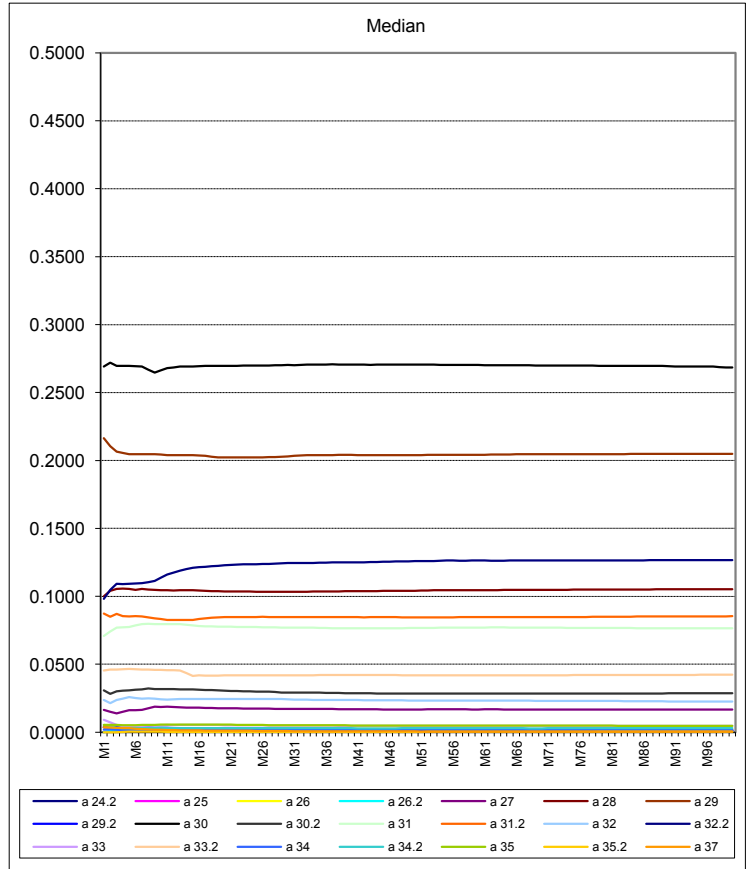
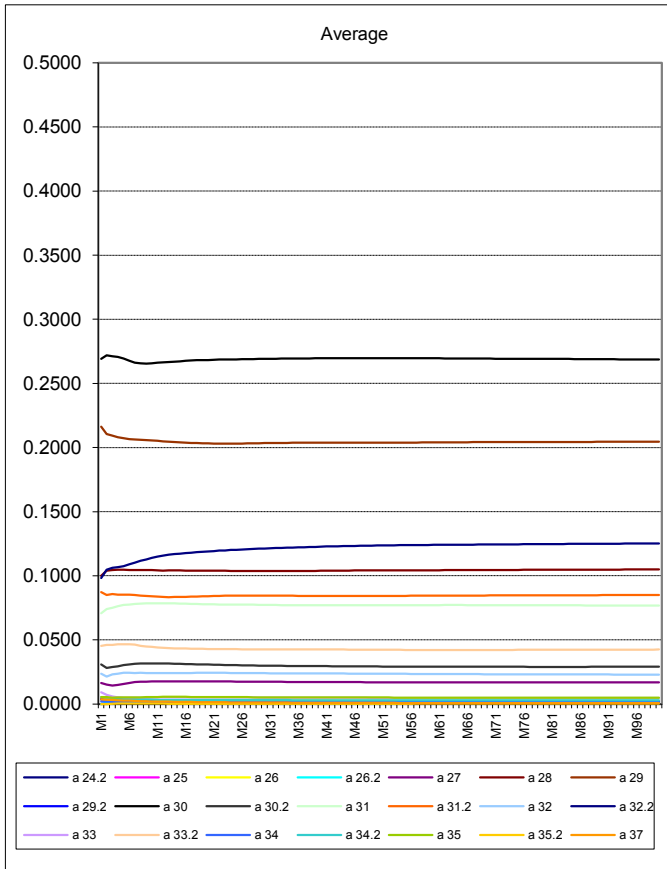
Sample number	48
Sample size	650

Average D	Min	Max
0.0002	0.0000	0.0007

Alleles	Average	min	max	Int min	Int Max	Median	S.D	IC
a24.2	0.0015	0.0000	0.0023	0.0014	0.0015	0.0016	0.0008	0.0001
a25	0.0016	0.0000	0.0024	0.0016	0.0017	0.0015	0.0007	0.0001
a26	0.0024	0.0008	0.0038	0.0023	0.0024	0.0023	0.0008	0.0001
a26.2	0.0008	0.0000	0.0016	0.0007	0.0008	0.0008	0.0004	0.0000
a27	0.0165	0.0126	0.0196	0.0164	0.0167	0.0170	0.0022	0.0002
a28	0.1054	0.0965	0.1181	0.1048	0.1059	0.1042	0.0070	0.0005
a29	0.2042	0.1891	0.2204	0.2035	0.2049	0.2059	0.0087	0.0007
a29.2	0.0014	0.0000	0.0024	0.0014	0.0015	0.0015	0.0006	0.0000
a30	0.2705	0.2538	0.2929	0.2698	0.2712	0.2679	0.0089	0.0007
a30.2	0.0282	0.0234	0.0356	0.0280	0.0284	0.0279	0.0030	0.0002
a31	0.0772	0.0678	0.0867	0.0768	0.0776	0.0774	0.0051	0.0004
a31.2	0.0845	0.0737	0.0930	0.0841	0.0849	0.0847	0.0054	0.0004
a32	0.0224	0.0171	0.0275	0.0222	0.0227	0.0227	0.0032	0.0002
a32.2	0.1266	0.1135	0.1379	0.1261	0.1271	0.1274	0.0061	0.0005
a33	0.0045	0.0022	0.0071	0.0043	0.0046	0.0045	0.0015	0.0001
a33.2	0.0414	0.0341	0.0468	0.0411	0.0416	0.0417	0.0035	0.0003
a34	0.0020	0.0000	0.0032	0.0019	0.0021	0.0022	0.0009	0.0001
a34.2	0.0030	0.0015	0.0045	0.0029	0.0030	0.0031	0.0009	0.0001
a35	0.0048	0.0023	0.0070	0.0047	0.0049	0.0047	0.0012	0.0001
a35.2	0.0003	0.0000	0.0008	0.0003	0.0003	0.0000	0.0004	0.0000
a36	0.0006	0.0000	0.0008	0.0005	0.0006	0.0008	0.0004	0.0000
a37	0.0004	0.0000	0.0008	0.0003	0.0004	0.0000	0.0004	0.0000

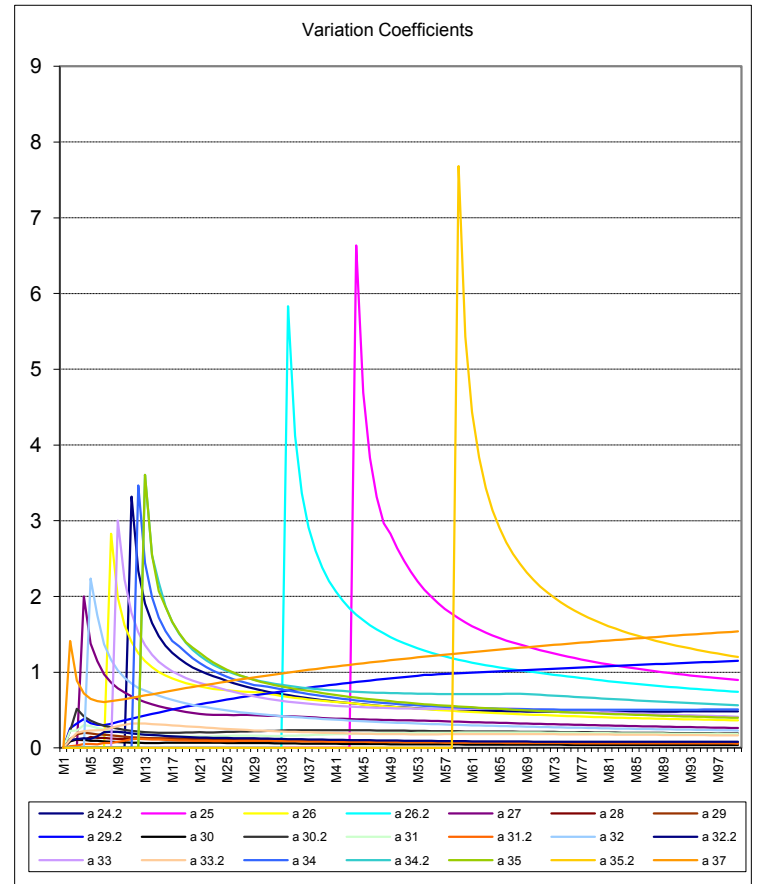
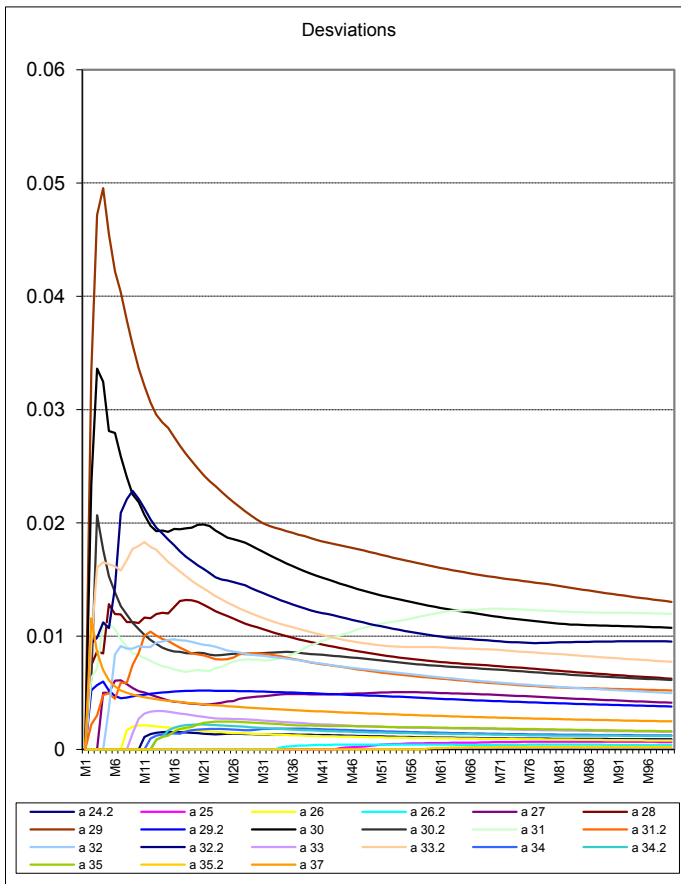
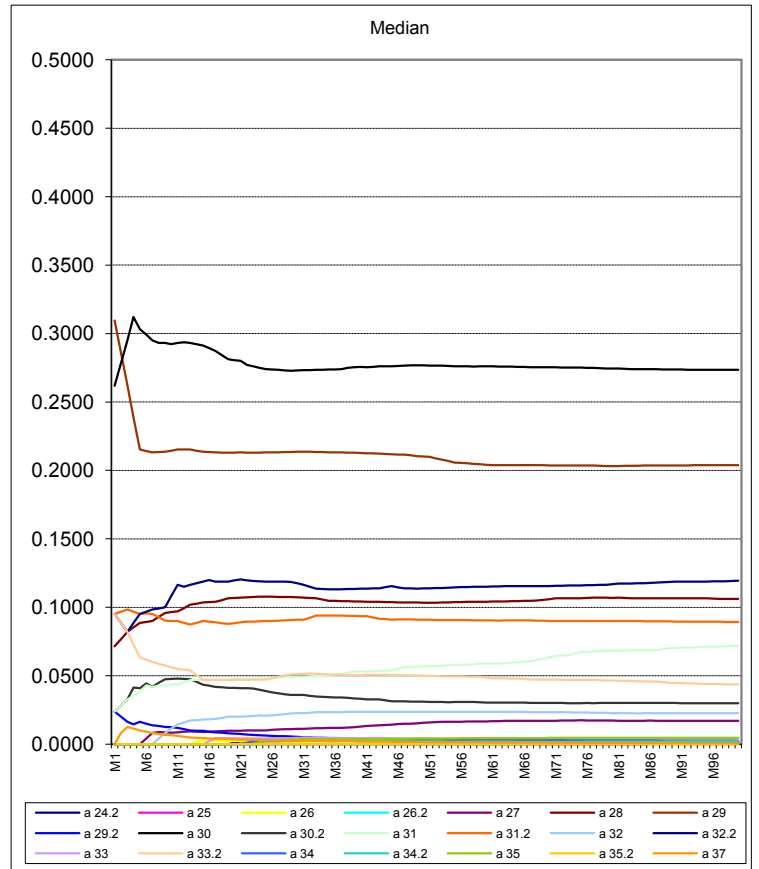
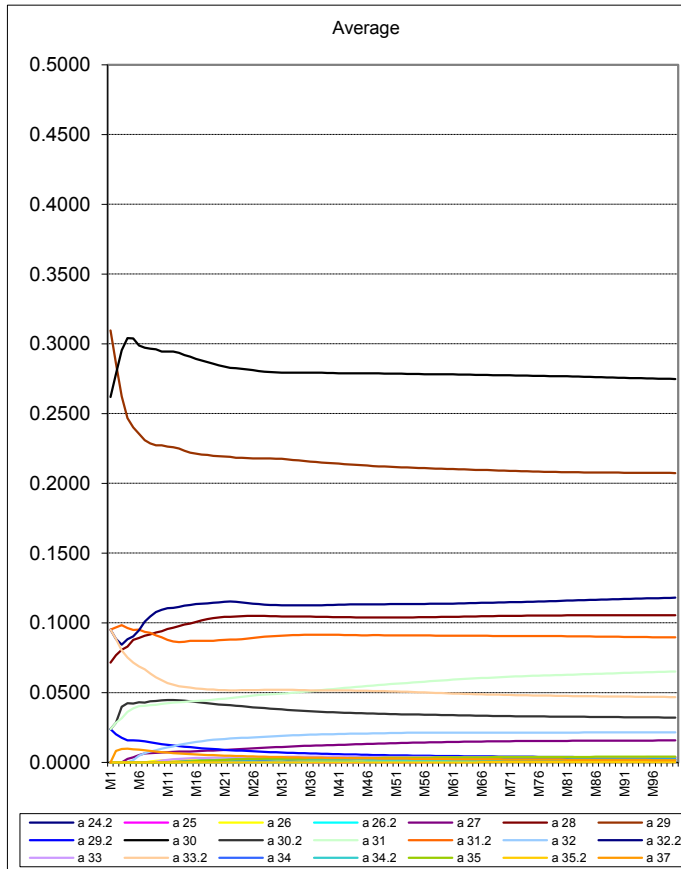
Graphic analysis of polymorphisms.

System D21S11 Pop. Bogotá



Graphic analysis of polymorphisms.

System	D21S11	Pop.	Bogotá
--------	--------	------	--------



Graphic analysis of polymorphisms.

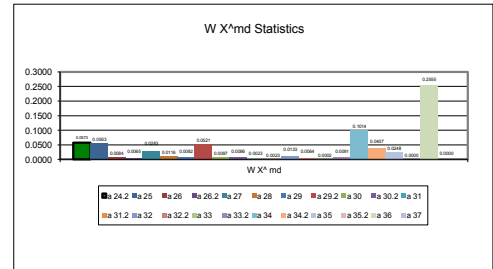
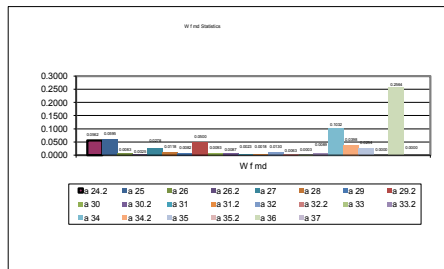
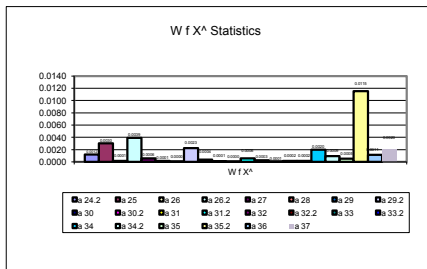
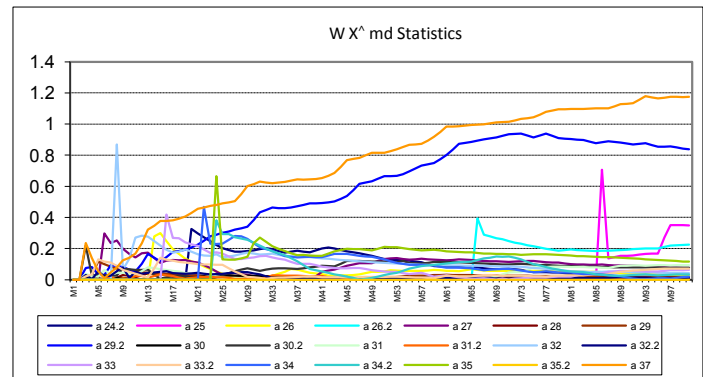
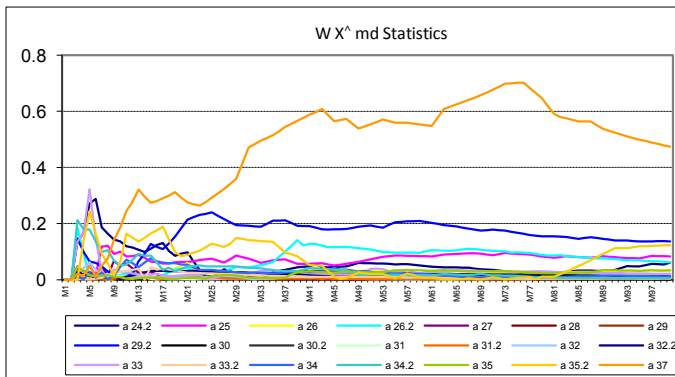
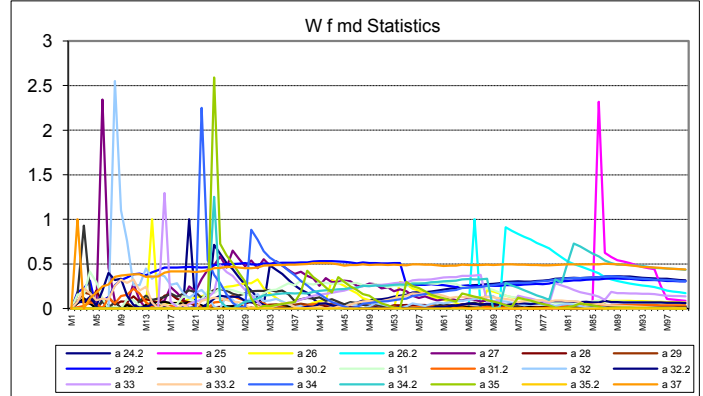
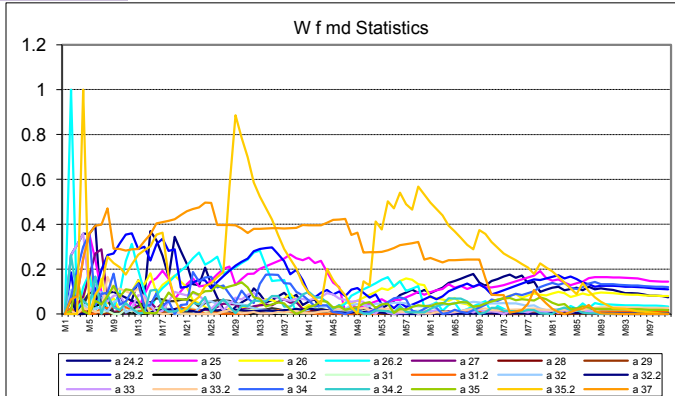
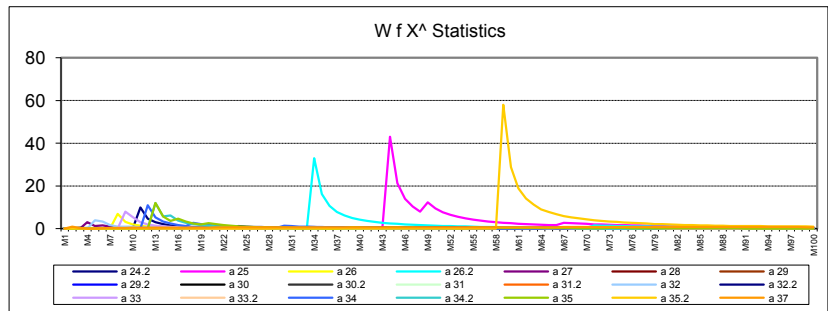
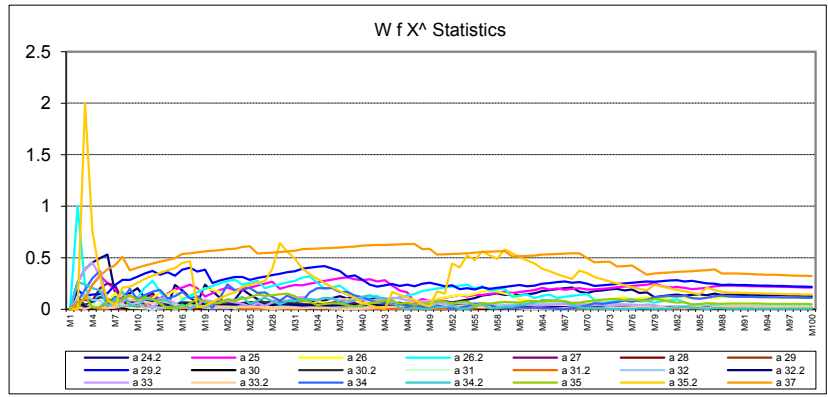
System	D21S11	Pop.	Bogotá
--------	--------	------	--------

W Statistics to Specific sample

Alleles	W f X [^]	W f md	W X [^] md
a 24.2	0.0012		
a 25	0.0030	0.0595	0.0563
a 26	0.0001	0.0083	0.0084
a 26.2	0.0039	0.0025	0.0065
a 27	0.0006	0.0278	0.0283
a 28	0.0001	0.0118	0.0116
a 29	0.0000	0.0082	0.0082
a 29.2	0.0023	0.0500	0.0521
a 30	0.0004	0.0093	0.0097
a 30.2	0.0001	0.0087	0.0086
a 31	0.0000	0.0023	0.0023
a 31.2	0.0006	0.0018	0.0023
a 32	0.0003	0.0130	0.0133
a 32.2	0.0001	0.0063	0.0064
a 33	0.0002	0.0003	0.0002
a 33.2	0.0002	0.0089	0.0091
a 34	0.0020	0.1032	0.1014
a 34.2	0.0009	0.0398	0.0407
a 35	0.0005	0.0254	0.0248
a 35.2	0.0115		
a 36			
	0.0115	0.1032	0.1014

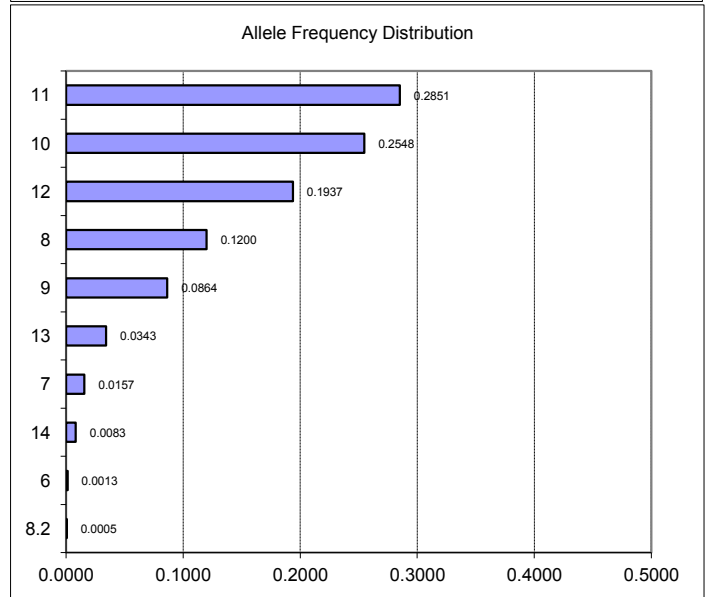
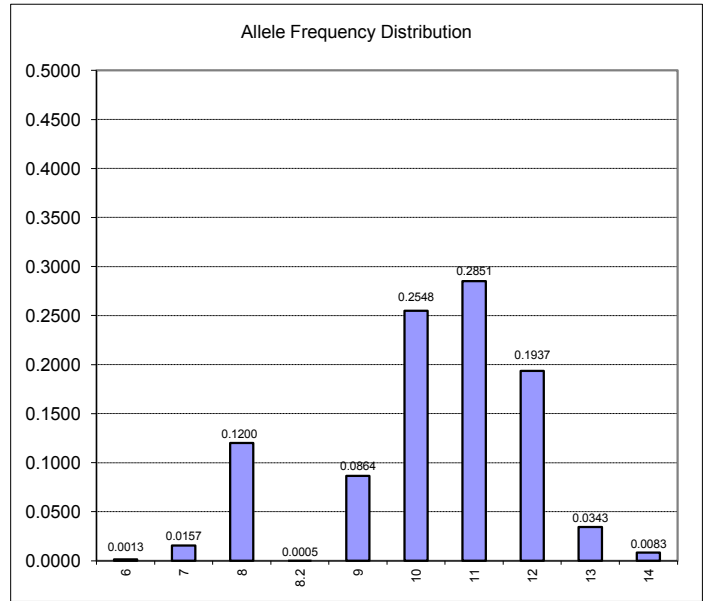
Sampling without repetition

Indeterminate value (Division by zero)

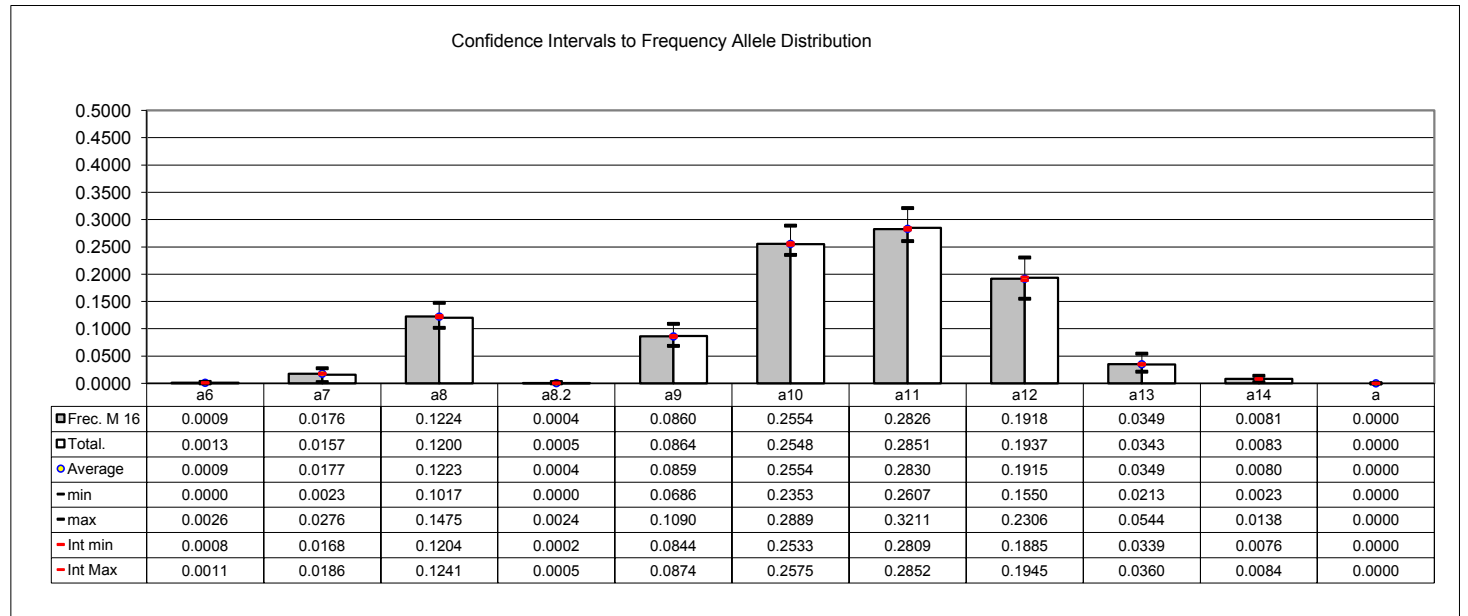


Population	Bogotá
System	D7S820
Alleles number	10
Effective alleles	5

Allele	M.Weight.	S.D.	Frequency
5	253.00		0.0010
6	255.15	0.08	0.0013
7	259.21	0.07	0.0157
8	263.24	0.07	0.1200
8.2	267.00		0.0005
9	267.26	0.09	0.0864
10	271.32	0.08	0.2547
11	275.35	0.06	0.2850
12	279.42	0.07	0.1936
13	283.42	0.06	0.0342
14	287.48	0.10	0.0083
15	291.58	0.06	



Single Alleles	
In the Database	In the Sample
8.2	6
	8.2

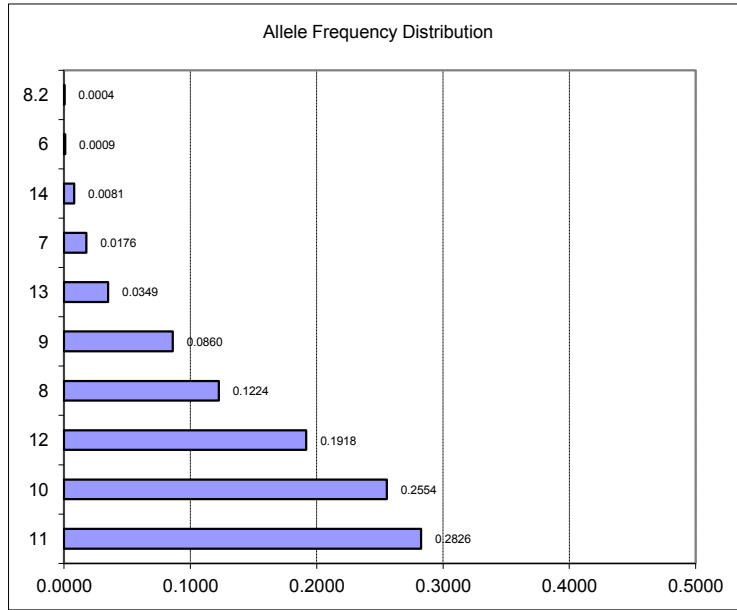


Specific sample analysis.

Sampling Specific.	
Population	Bogotá
System	D7S820
Alleles number	10
Effective alleles number	5
CV	0.05
Sample number	16
Sample size	213
Polimorphism alleles total	22

EXPERIMENT	(Todas)
Sample number	16

Allele frequency summary	
Alleles	Total
11	0.2826
10	0.2554
12	0.1918
8	0.1224
9	0.0860
13	0.0349
7	0.0176
14	0.0081
6	0.0009
8.2	0.0004



Single Alleles in the Sampling

EXPERIMENT	(Todas)
Sample number	16

Max	Total
Alleles	Total
8.2	1
6	1
14	6
7	12
13	23
9	47
8	68
12	98
10	128
11	130

Sampling and CV.

Sample selection
Analysis of CV.

Alleles	Sample No.
14	0
11	2
13	17
10	2
12	2
9	15
8	16
7	0
6	0
8.2	0

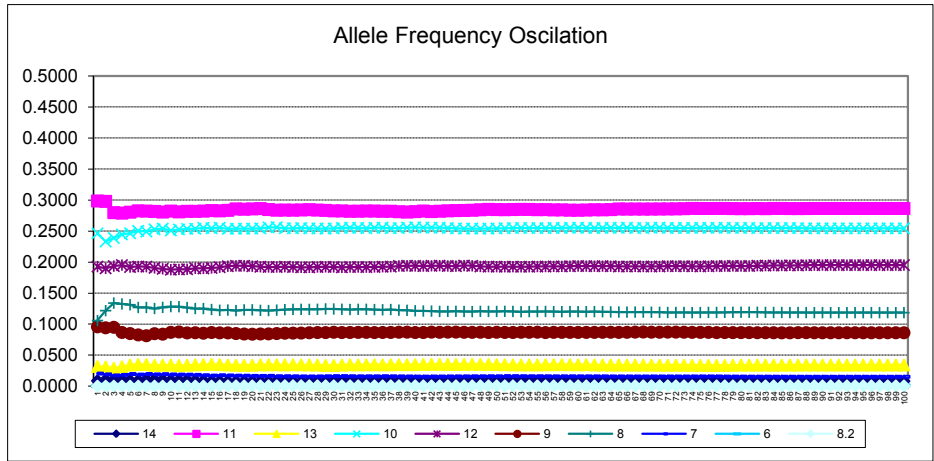
Allelic Distribution Order

Population	Bogotá
System	D7S820

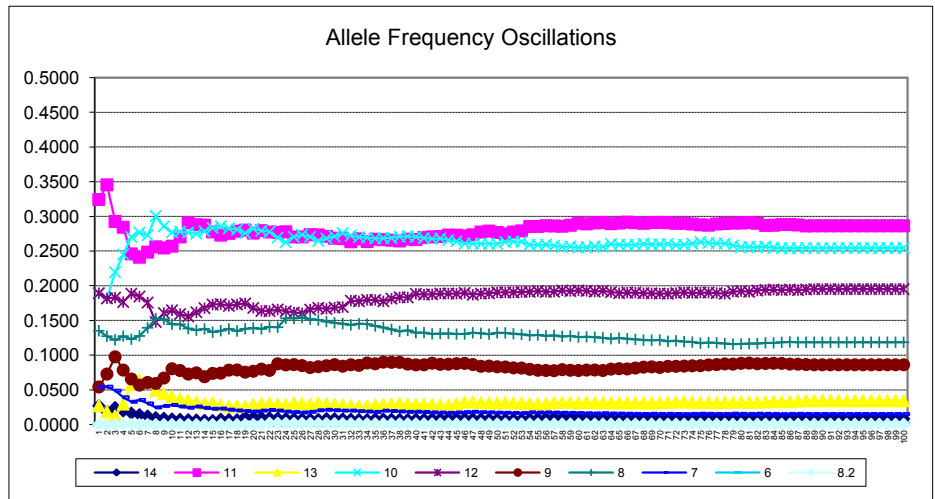
EXPERIMENT	(Todas)
Sample number	(Todas)

Allele frequency summary		
Alleles		Total
	11	0.2851
	10	0.2548
	12	0.1937
	8	0.1200
	9	0.0864
	13	0.0343
	7	0.0157
	14	0.0083
	6	0.0013
	8.2	0.0005

Compiled Allele Frequency



Allele frequency for a specific experiment.



Analysis for Specific Sampling

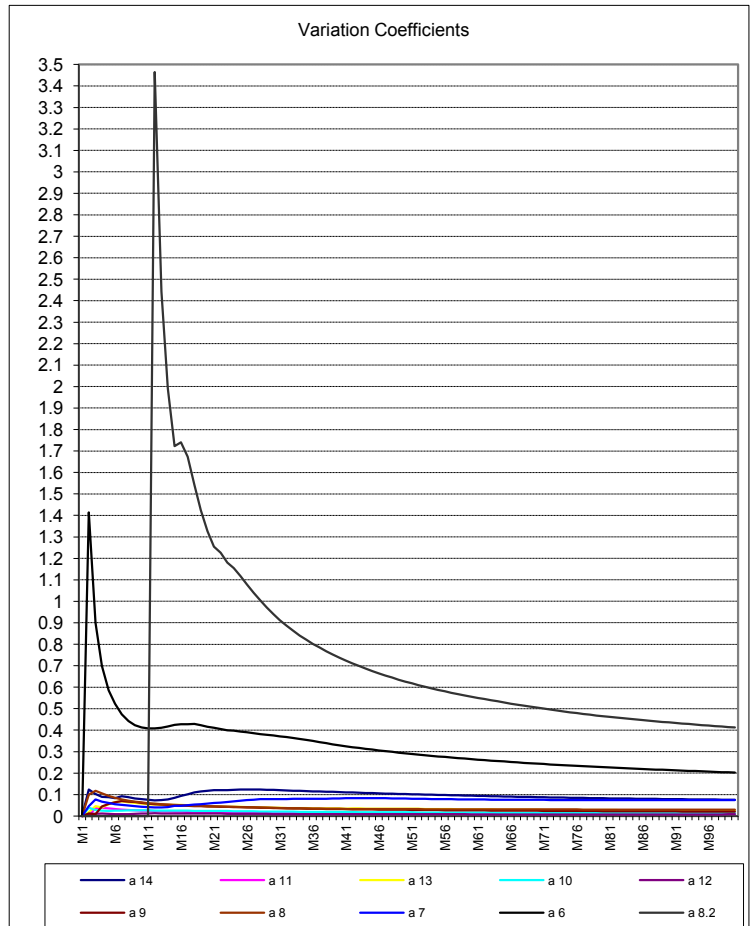
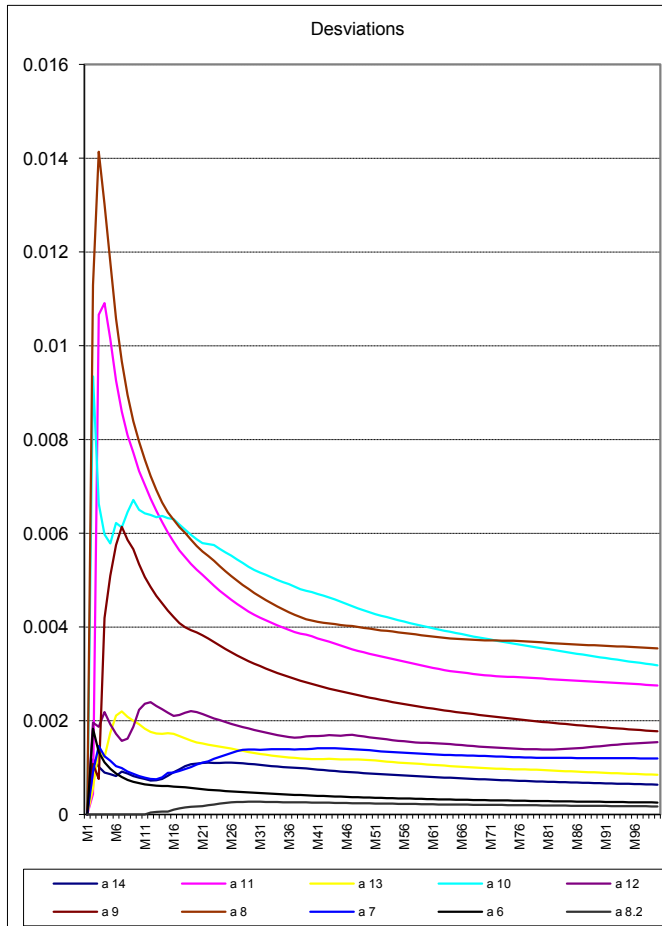
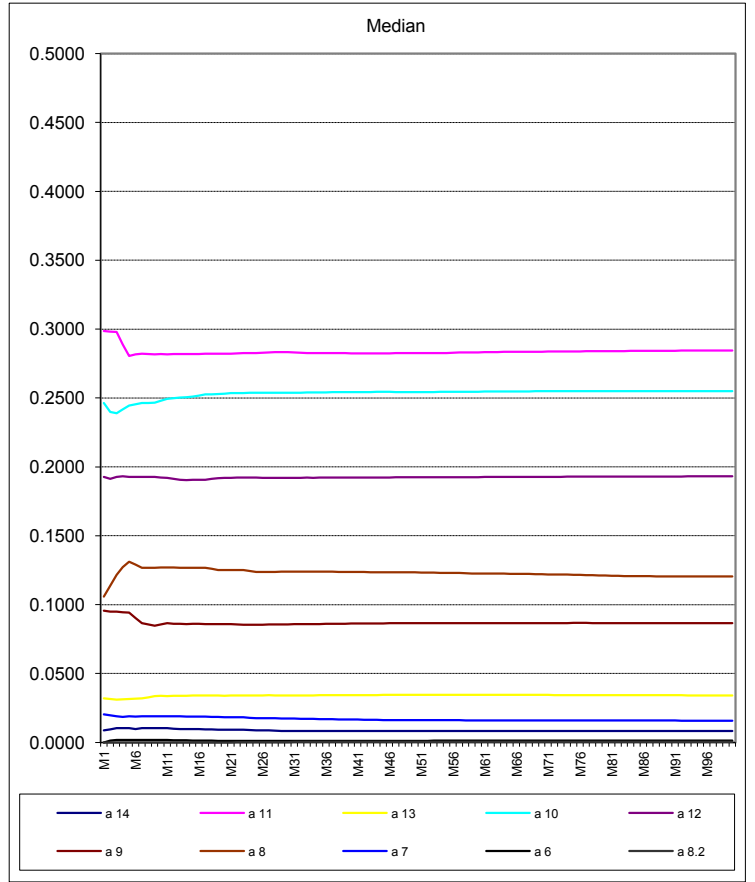
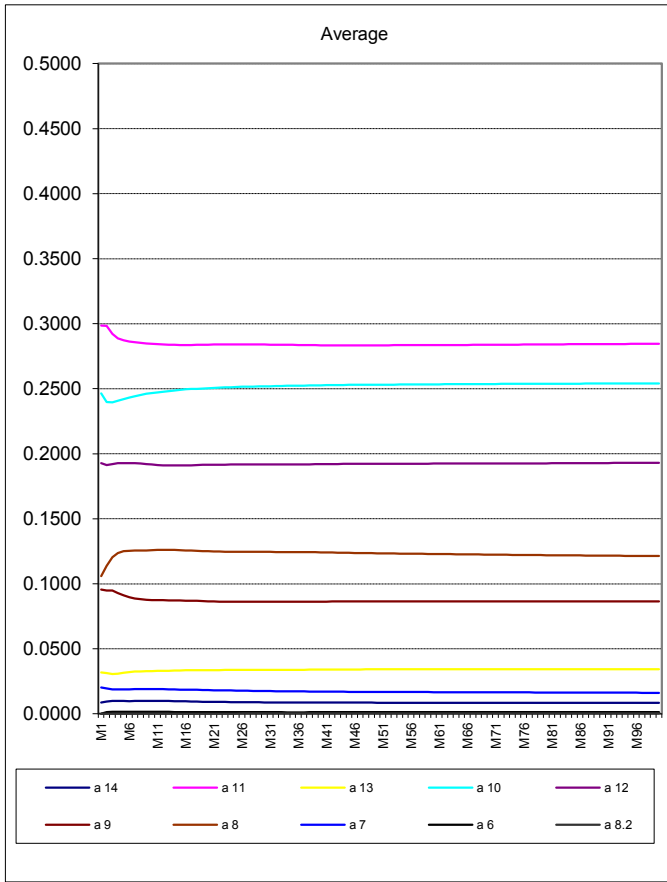
Sample number	16
Sample size	213

Average D	Min	Max
0.0013	0.0001	0.0030

Alleles	Average	min	max	Int min	Int Max	Median	S.D	IC
a6	0.0009	0.0000	0.0026	0.0008	0.0011	0.0000	0.0012	0.0002
a7	0.0177	0.0023	0.0276	0.0168	0.0186	0.0191	0.0066	0.0009
a8	0.1223	0.1017	0.1475	0.1204	0.1241	0.1202	0.0136	0.0018
a8.2	0.0004	0.0000	0.0024	0.0002	0.0005	0.0000	0.0009	0.0001
a9	0.0859	0.0686	0.1090	0.0844	0.0874	0.0874	0.0111	0.0015
a10	0.2554	0.2353	0.2889	0.2533	0.2575	0.2538	0.0158	0.0021
a11	0.2830	0.2607	0.3211	0.2809	0.2852	0.2770	0.0161	0.0022
a12	0.1915	0.1550	0.2306	0.1885	0.1945	0.1933	0.0224	0.0030
a13	0.0349	0.0213	0.0544	0.0339	0.0360	0.0348	0.0078	0.0010
a14	0.0080	0.0023	0.0138	0.0076	0.0084	0.0086	0.0030	0.0004

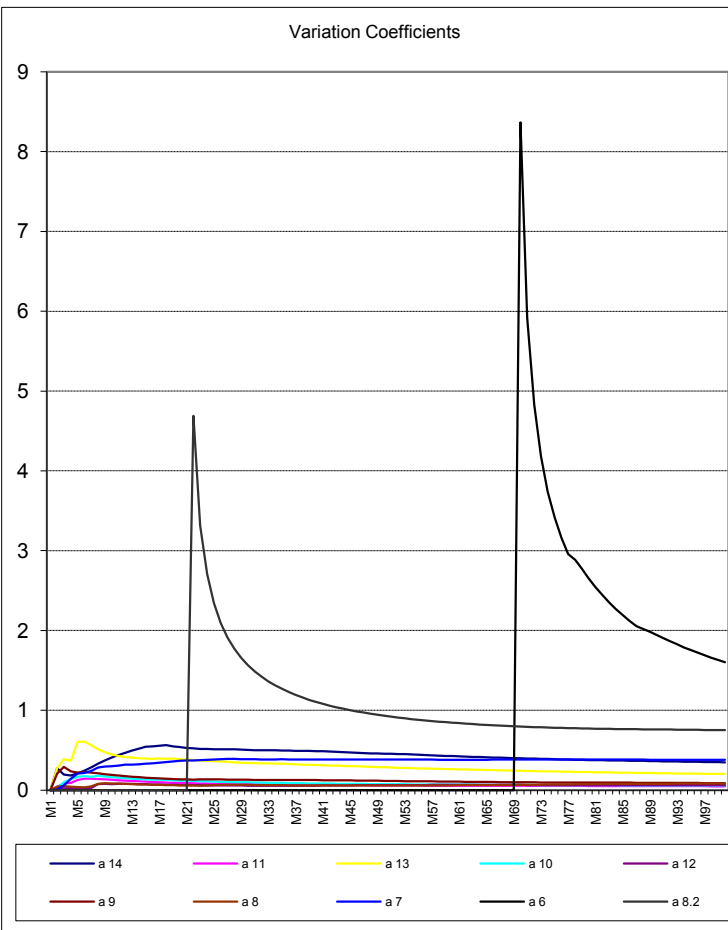
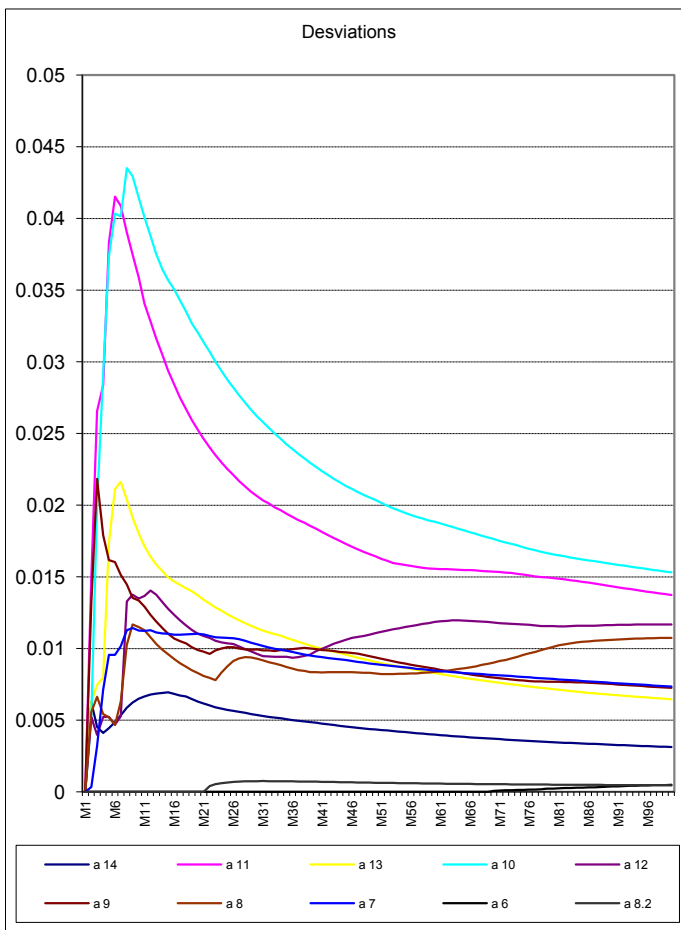
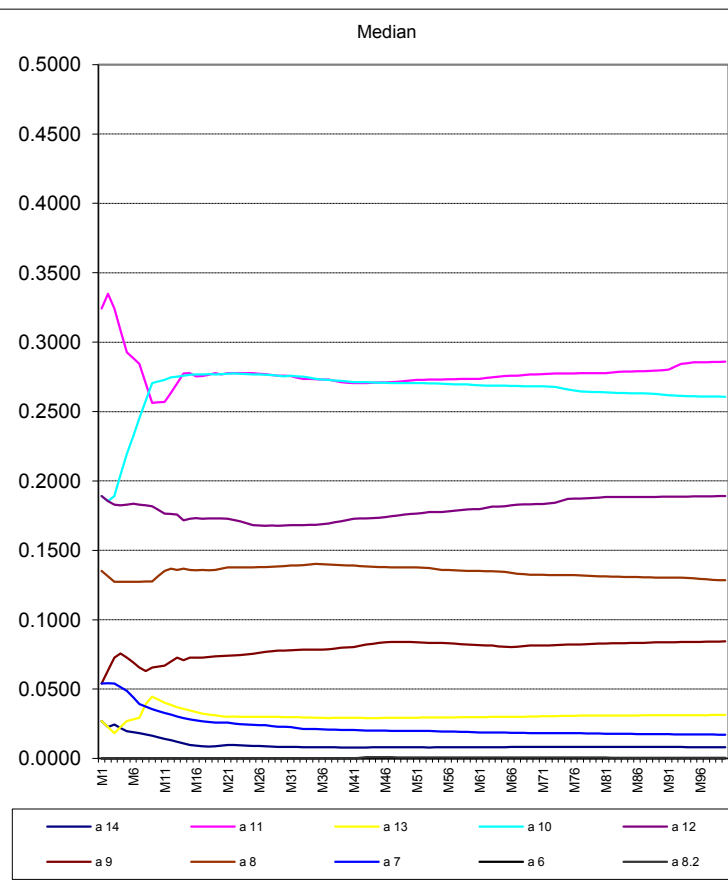
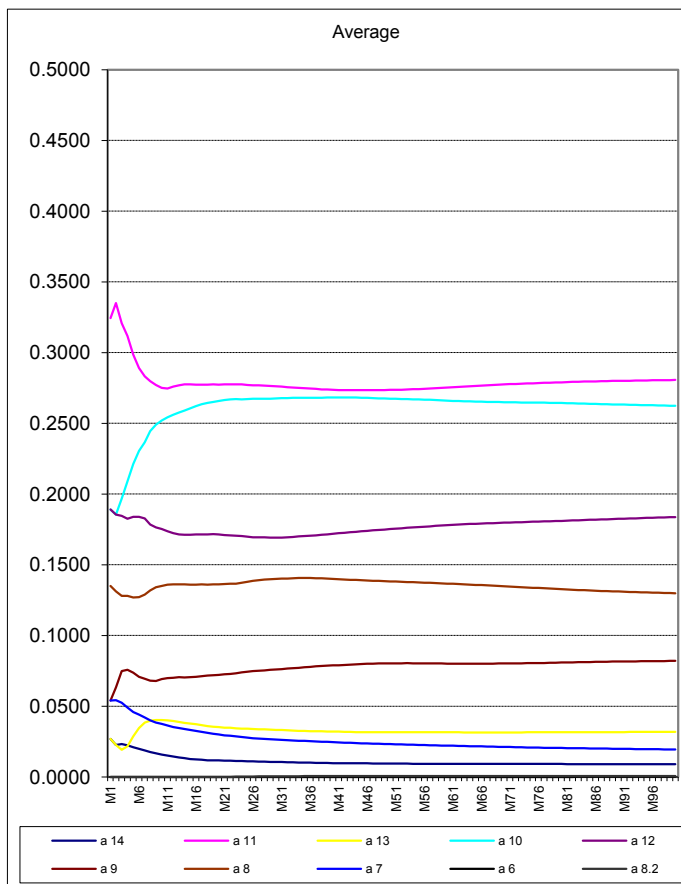
Graphic analysis of polymorphisms.

System D7S820 Pop. Bogotá



Graphic analysis of polymorphisms.

System D7S820 Pop. Bogotá

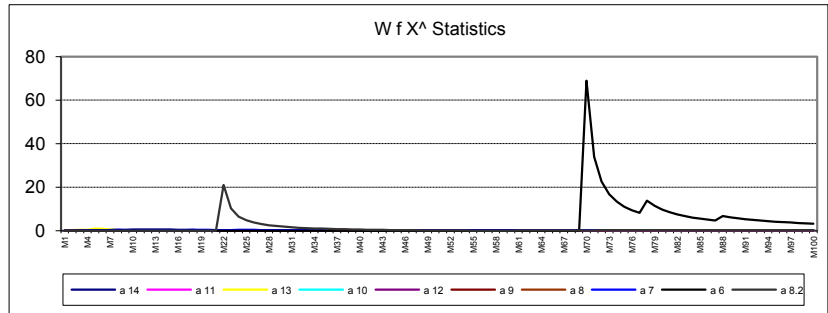
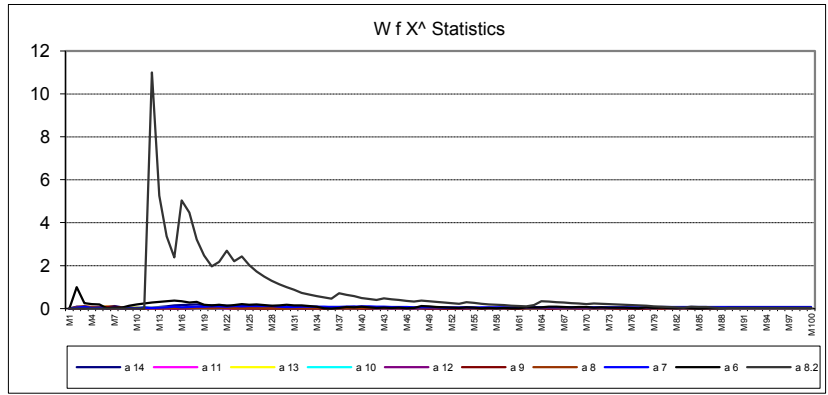


Graphic analysis of polymorphisms.

System	D7S820	Pop.	Bogotá
--------	--------	------	--------

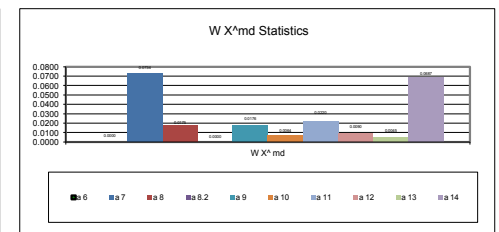
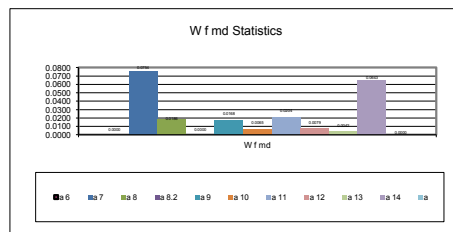
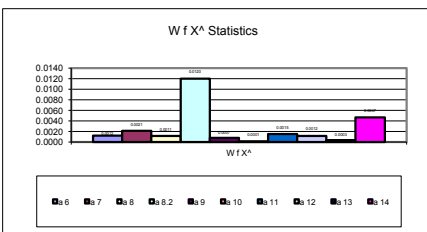
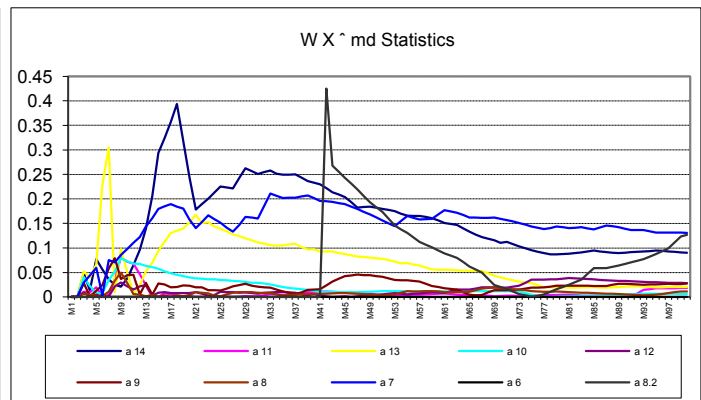
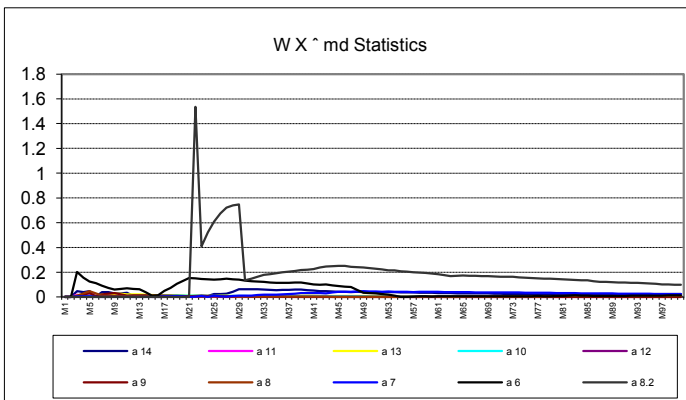
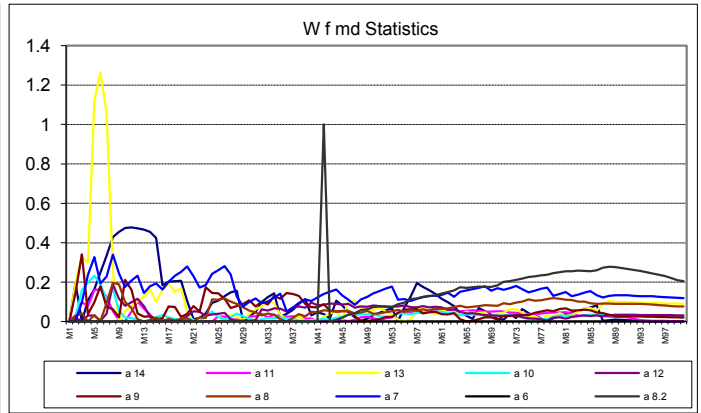
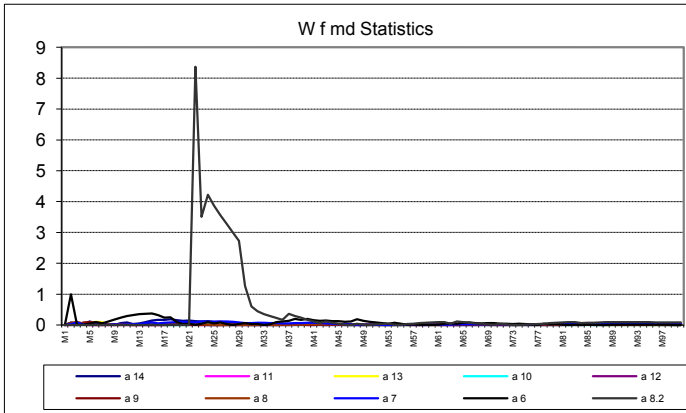
W Statistics to Specific sample

Alleles	W f X [^]	W f md	W X [^] md
a 6	0.0012		
a 7	0.0021		
a 8	0.0011	0.0186	0.0175
a 8.2	0.0120		
a 9	0.0007		
a 10	0.0001	0.0065	0.0064
a 11	0.0015	0.0204	0.0220
a 12	0.0012	0.0079	0.0090
a 13	0.0003	0.0042	0.0045
a 14	0.0047	0.0643	0.0687



Indeterminate value (Division by zero)

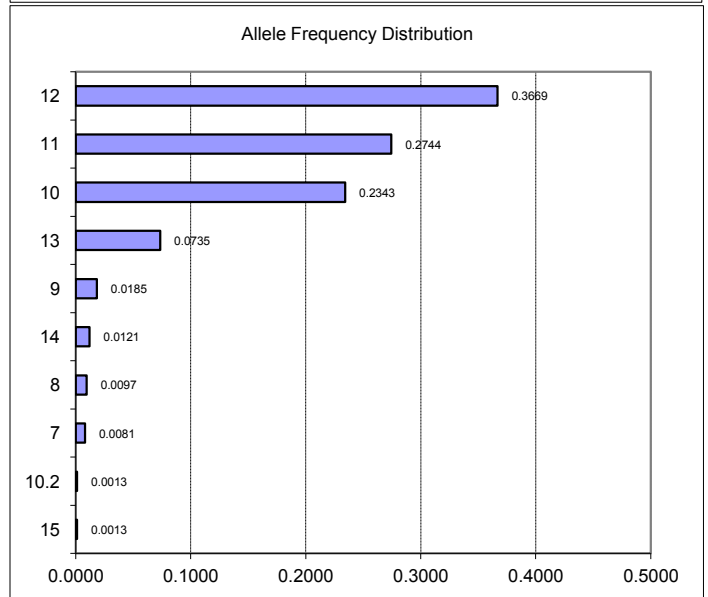
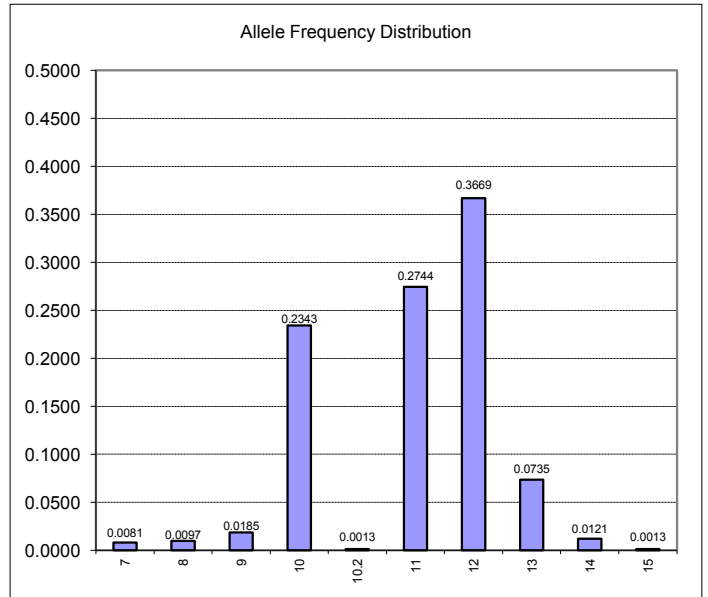
0.0120	0.0643	0.0687
Sampling without repetition		



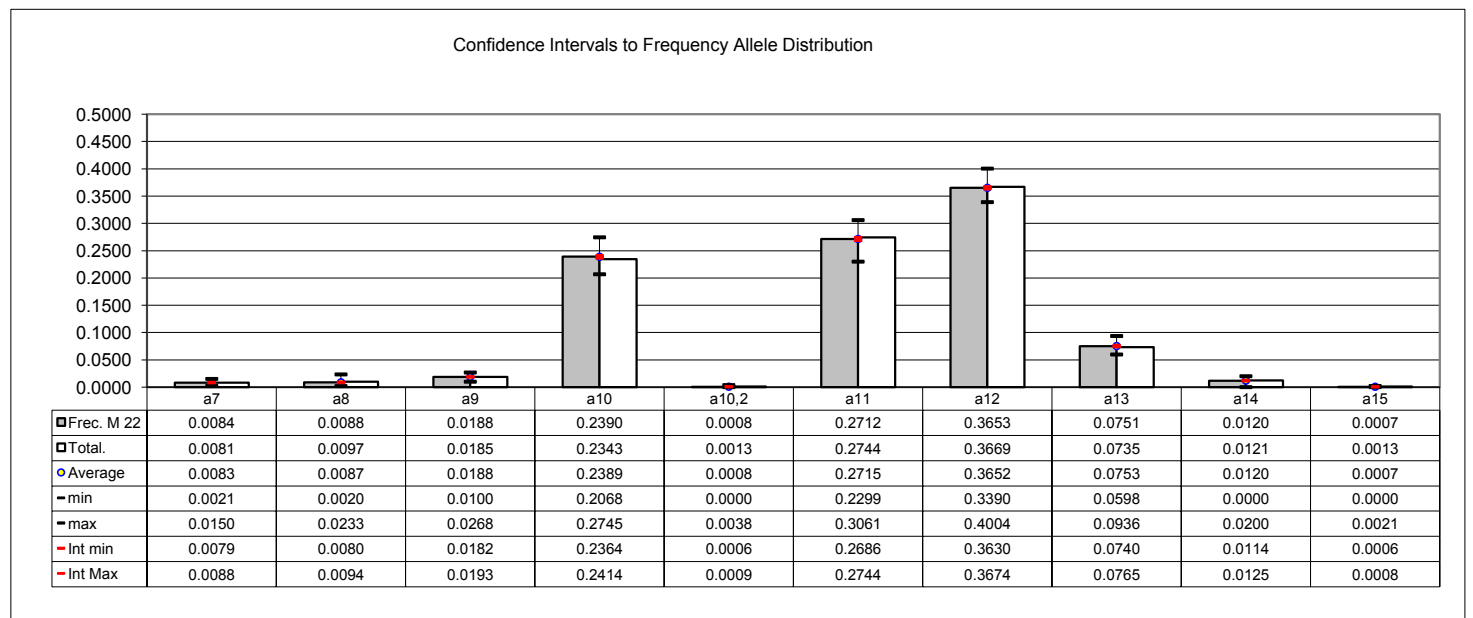
Population	Bogotá
System	CSF1PO
Alleles number	10
Effective alleles	4

Final frequency.

Allele	M.Weight.	S.D.	Frequency
6	304.69	0.08	0.0010
7	309.01	0.10	0.0081
8	313.30	0.10	0.0097
9	317.55	0.11	0.0185
10	321.97	0.12	0.2343
10.2	323.00		0.0013
11	325.86	0.11	0.2744
12	329.97	0.13	0.3669
13	334.00	0.10	0.0735
14	338.04	0.11	0.0121
15	341.84	0.08	0.0013



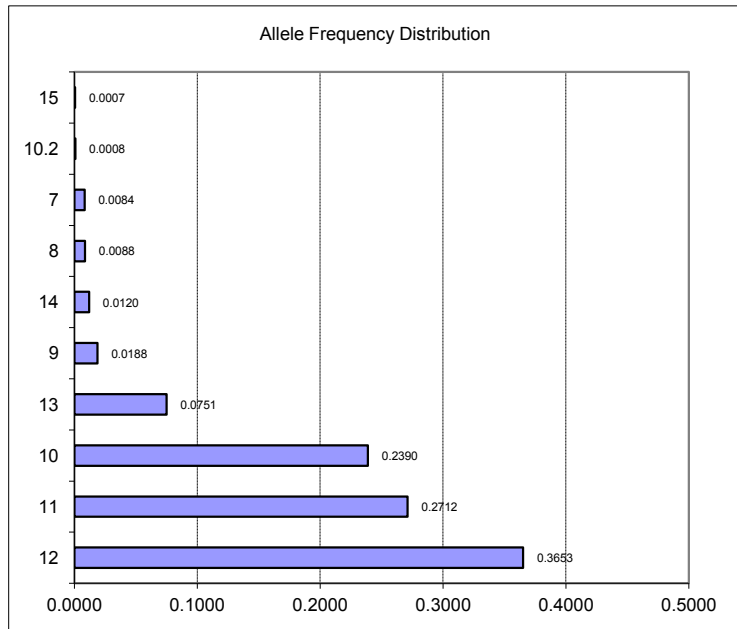
Single Alleles	
In the Database	In the Sample
	15



Sampling Specific.	
Population	Bogotá
System	CSFIPO
Alleles number	10
Effective alleles number	4
CV	0.05
Sample number	22
Sample size	250
Polimorphism alleles total	15

EXPERIMENT	(Todas)
Sample number	22

Allele frequency summary	
Alleles	Total
12	0.3653
11	0.2712
10	0.2390
13	0.0751
9	0.0188
14	0.0120
8	0.0088
7	0.0084
10.2	0.0008
15	0.0007



Single Alleles in the Sampling

EXPERIMENT	(Todas)
Sample number	22

Alleles	Total
15	1
10.2	2
7	8
14	10
8	11
9	13
13	47
10	154
11	164
12	213

Sampling and CV.

Sample selection
Analysis of CV.

Alleles	Sample No.
14	0
11	2
13	2
15	0
10	2
12	2
9	22
8	0
7	65
10.2	0

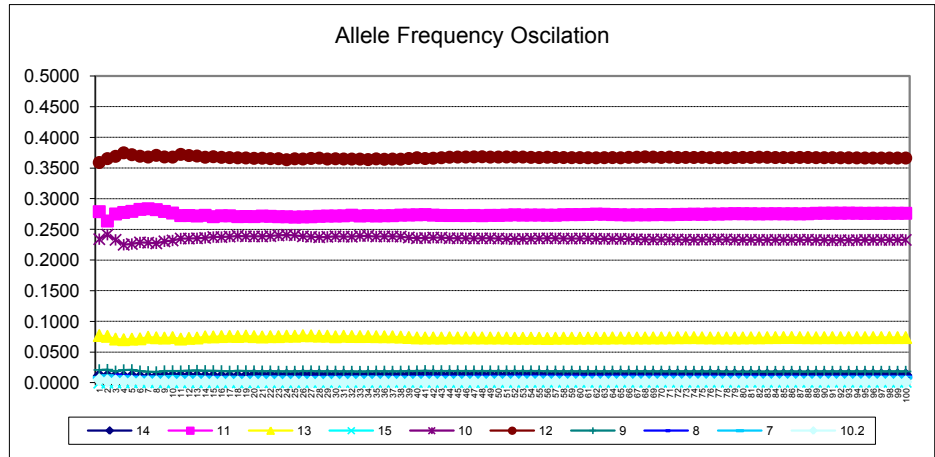
Allelic Distribution Order

Population	Bogotá
System	CSF1PO

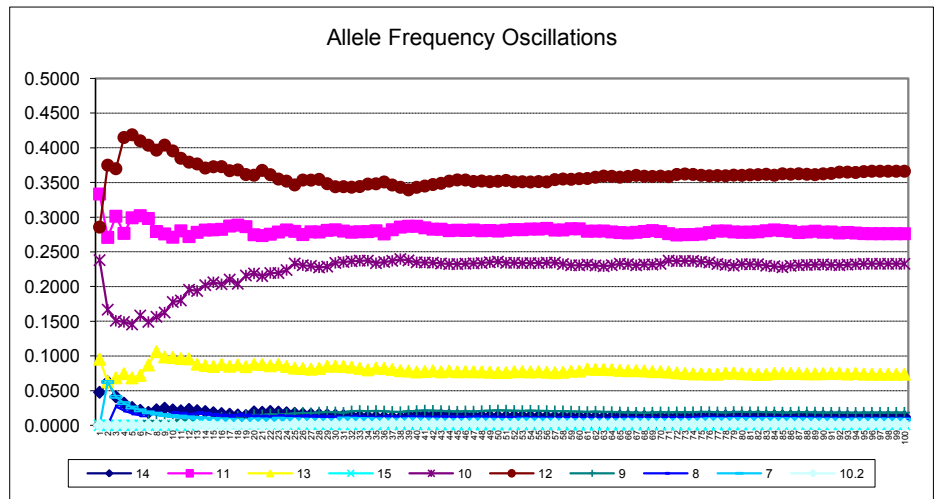
EXPERIMENT	(Todas)
Sample number	(Todas)

Allele frequency summary		
Alleles		Total
	12	0.3669
	11	0.2744
	10	0.2343
	13	0.0735
	9	0.0185
	14	0.0121
	8	0.0097
	7	0.0081
	10.2	0.0013
	15	0.0013

Compiled Allele Frequency



Allele frequency for a specific experiment.



Analysis for Specific Sampling

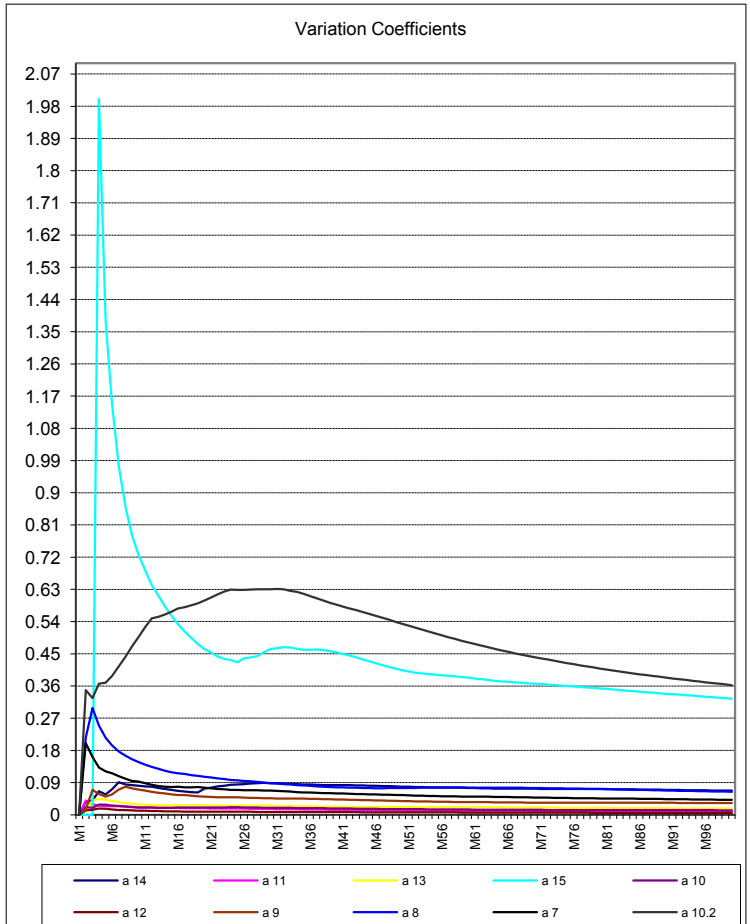
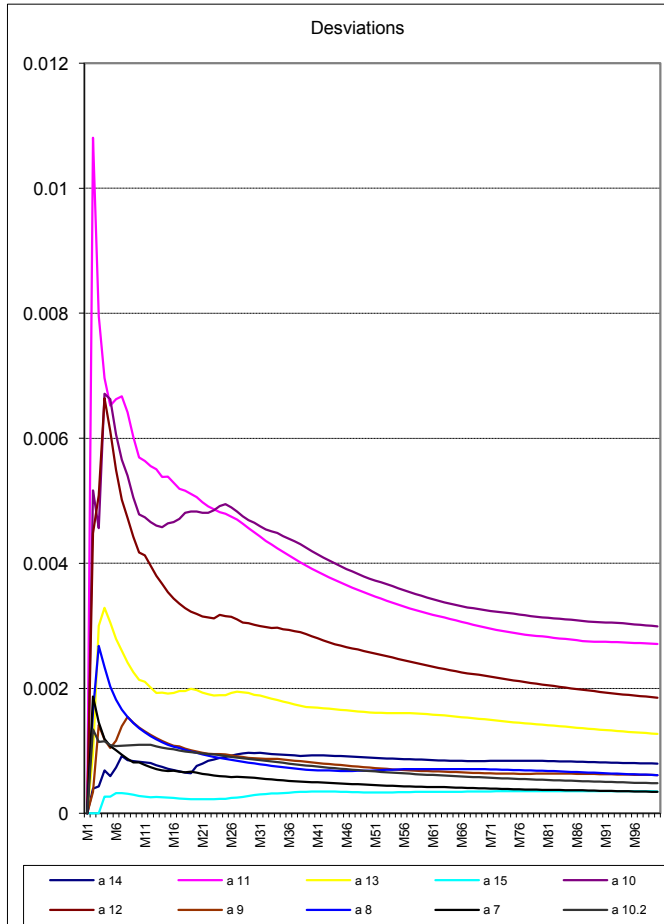
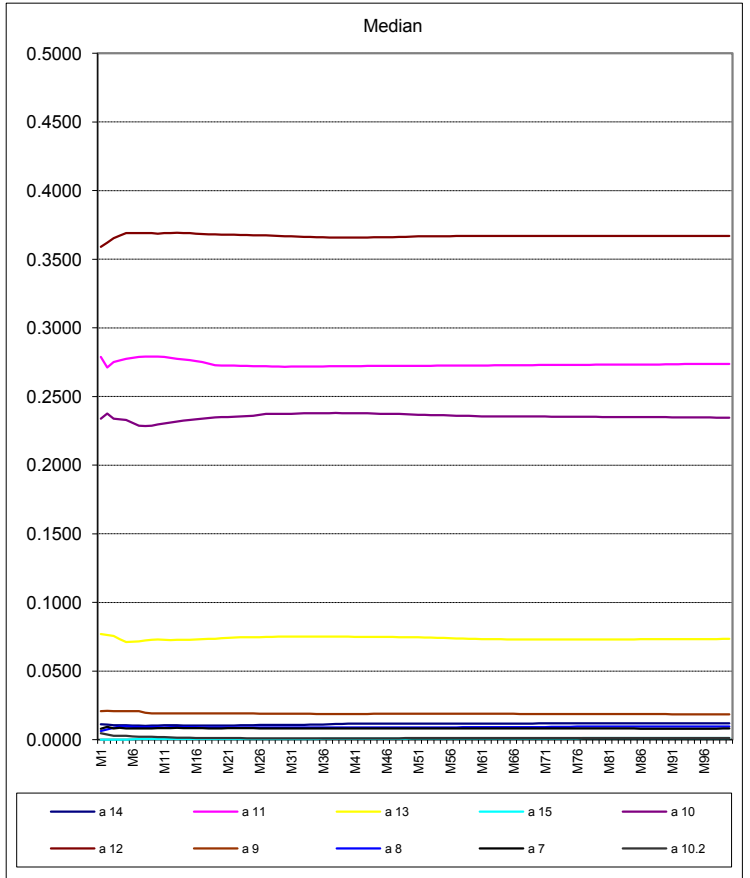
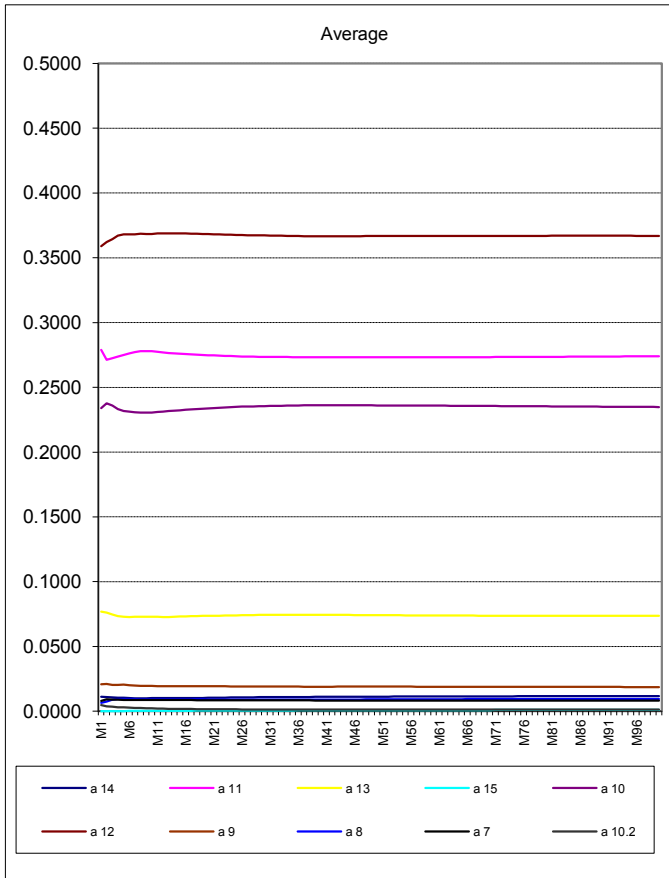
Sample number	22
Sample size	250

Average D	Min	Max
0.0011	0.0001	0.0029

Alleles	Average	min	max	Int min	Int Max	Median	S.D	IC
a7	0.0083	0.0021	0.0150	0.0079	0.0088	0.0080	0.0037	0.0005
a8	0.0087	0.0020	0.0233	0.0080	0.0094	0.0062	0.0055	0.0007
a9	0.0188	0.0100	0.0268	0.0182	0.0193	0.0188	0.0044	0.0005
a10	0.2389	0.2068	0.2745	0.2364	0.2414	0.2363	0.0201	0.0025
a10,2	0.0008	0.0000	0.0038	0.0006	0.0009	0.0000	0.0011	0.0001
a11	0.2715	0.2299	0.3061	0.2686	0.2744	0.2737	0.0234	0.0029
a12	0.3652	0.3390	0.4004	0.3630	0.3674	0.3628	0.0177	0.0022
a13	0.0753	0.0598	0.0936	0.0740	0.0765	0.0742	0.0101	0.0012
a14	0.0120	0.0000	0.0200	0.0114	0.0125	0.0124	0.0046	0.0006
a15	0.0007	0.0000	0.0021	0.0006	0.0008	0.0000	0.0010	0.0001

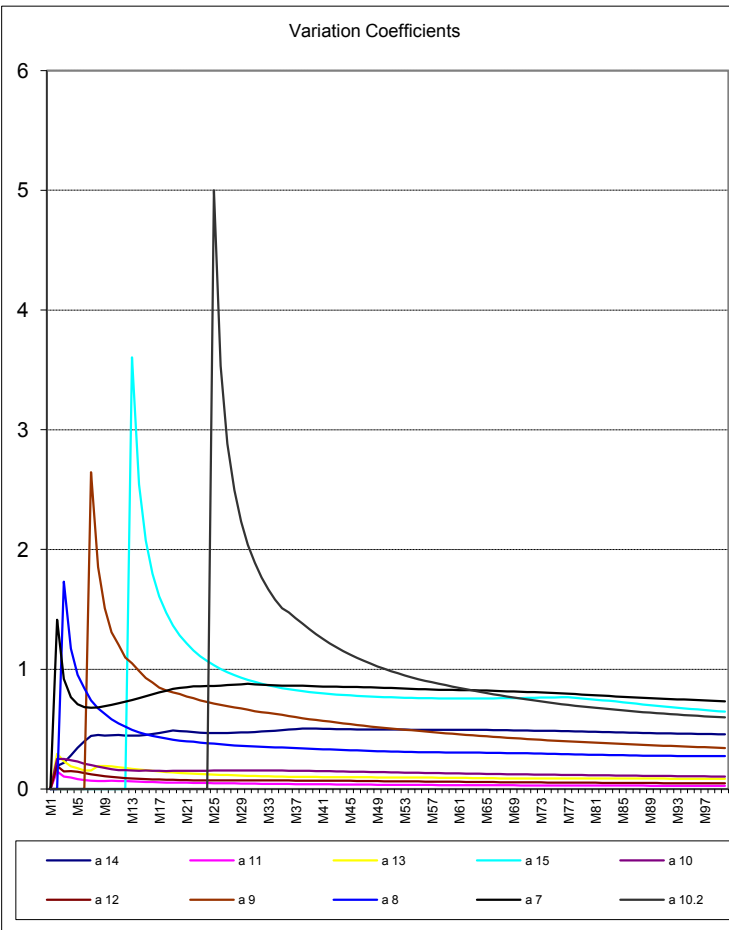
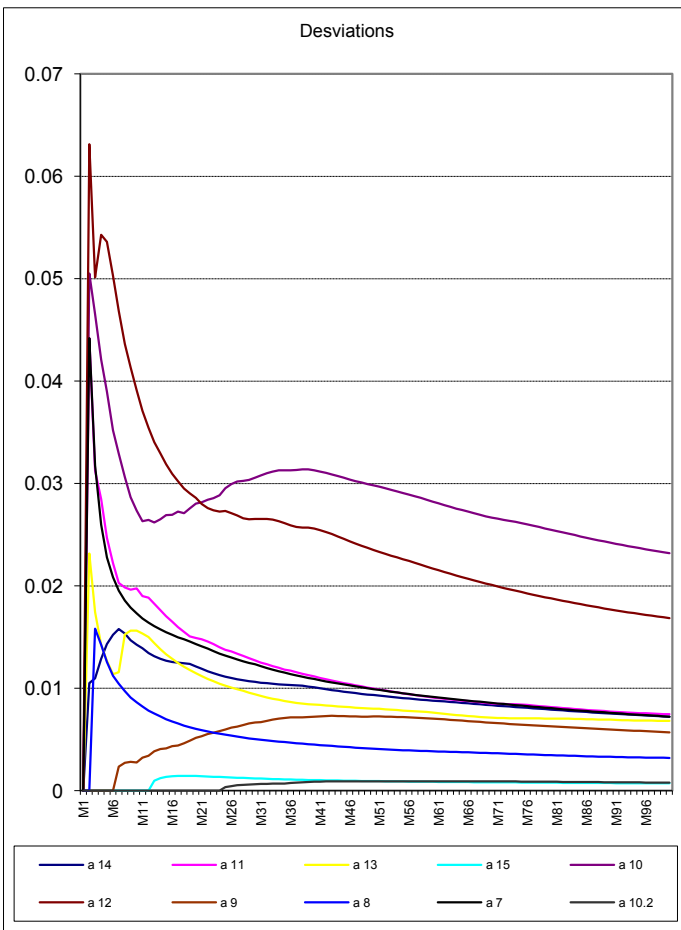
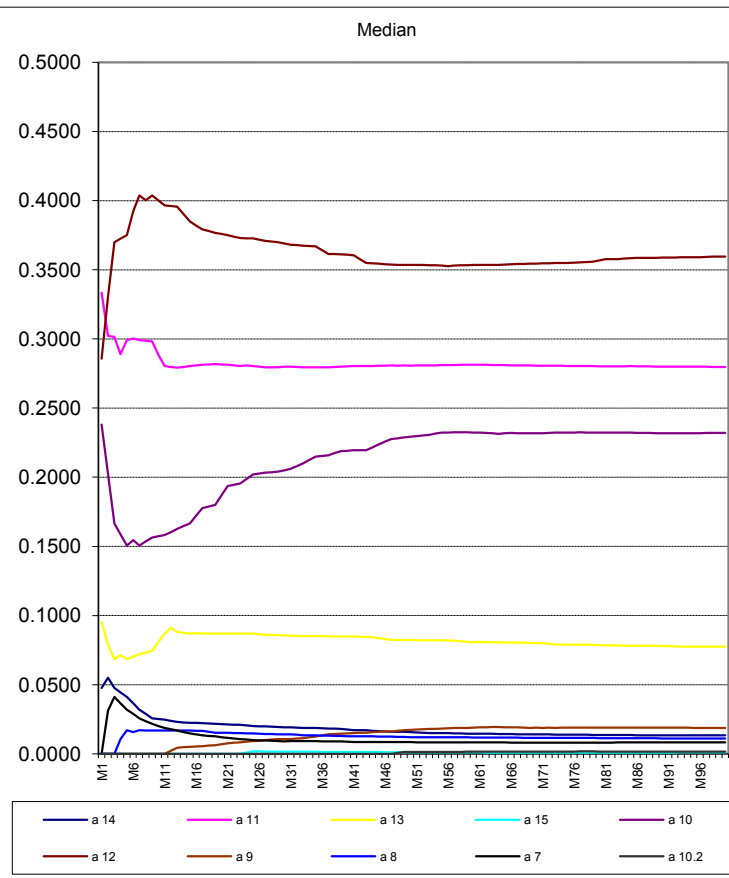
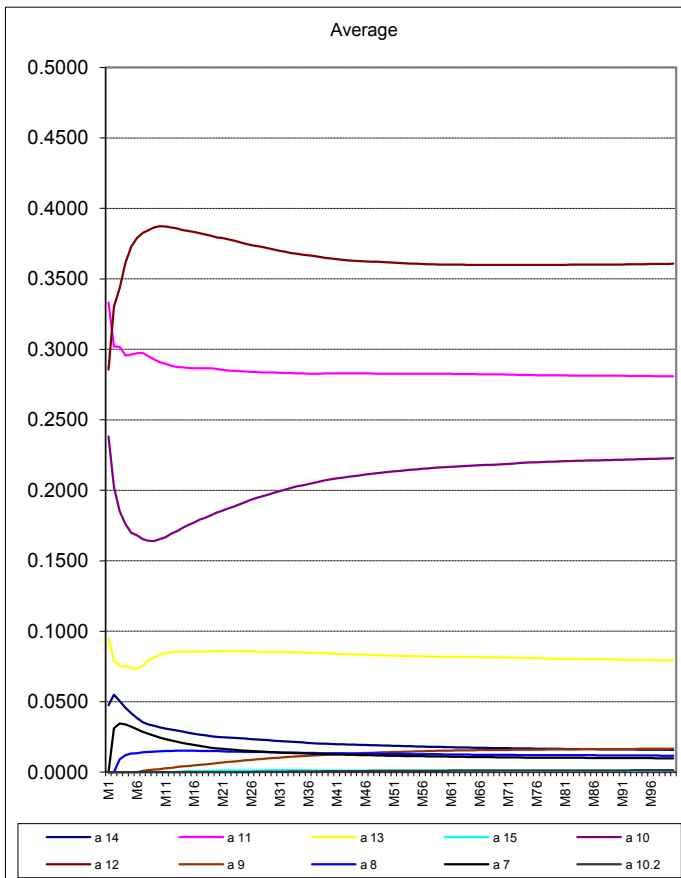
Graphic analysis of polymorphisms.

System CSF1PO Pop. Bogotá



Graphic analysis of polymorphisms.

System	CSF1PO	Pop.	Bogotá
--------	--------	------	--------

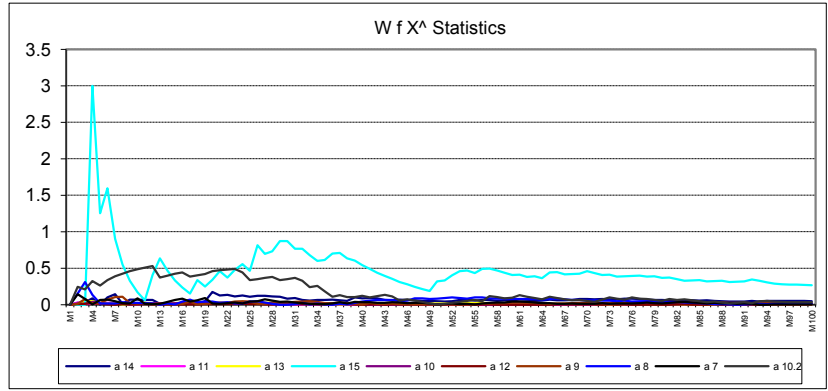


Graphic analysis of polymorphisms.

System	CSF1PO	Pop.	Bogotá
--------	--------	------	--------

W Statistics to Specific sample

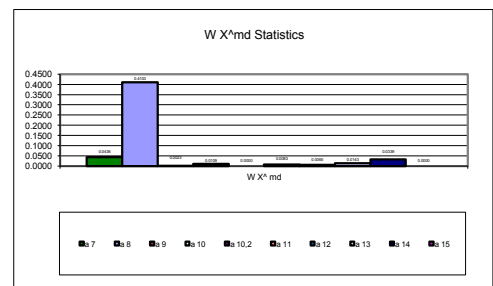
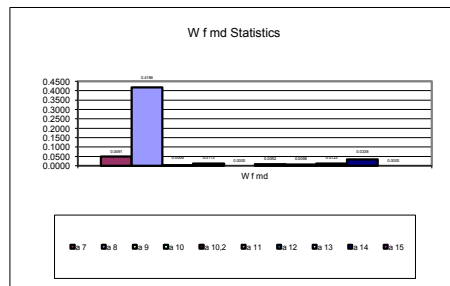
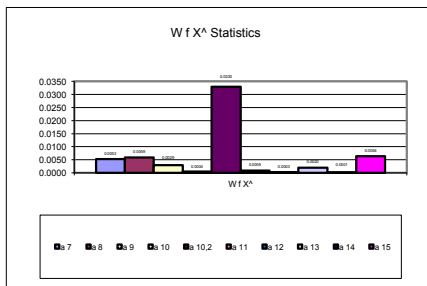
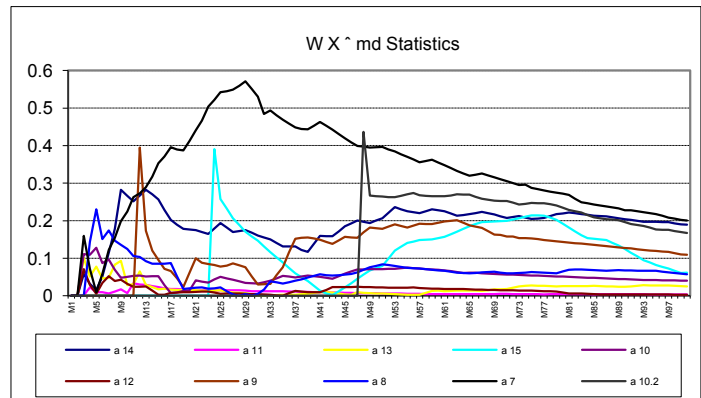
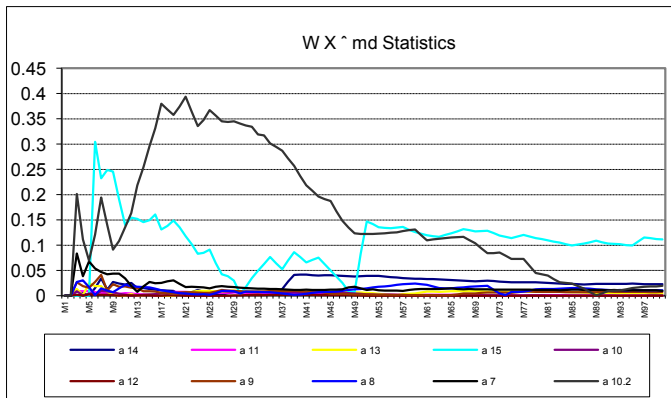
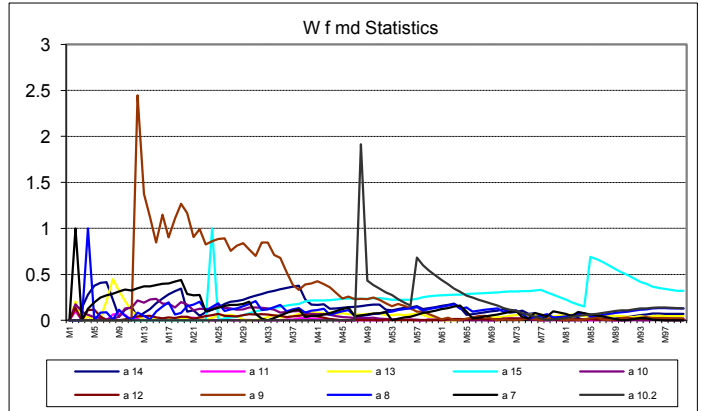
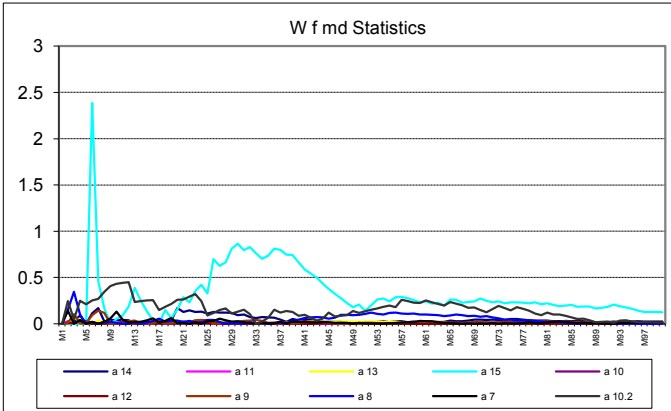
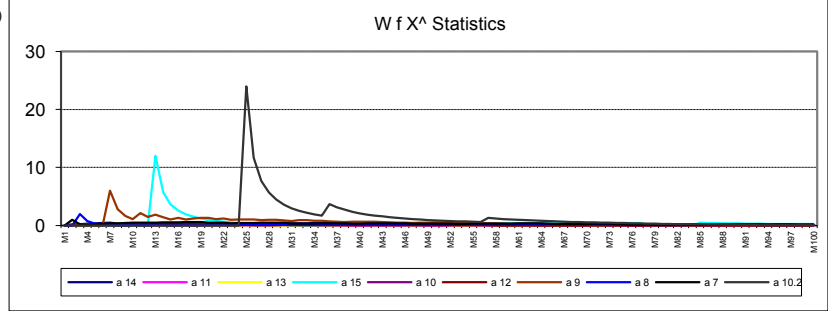
Alleles	W f X^	W f md	W X^ md
a 7	0.0053	0.0491	0.0436
a 8	0.0059	0.4186	0.4103
a 9	0.0029	0.0006	0.0023
a 10	0.0004	0.0113	0.0109
a 10,2	0.0330		
a 11	0.0009	0.0092	0.0083
a 12	0.0003	0.0068	0.0065
a 13	0.0020	0.0123	0.0143
a 14	0.0001	0.0338	0.0339
a 15	0.0064		



Indeterminate value (Division by zero)

0.0330 0.4186 0.4103

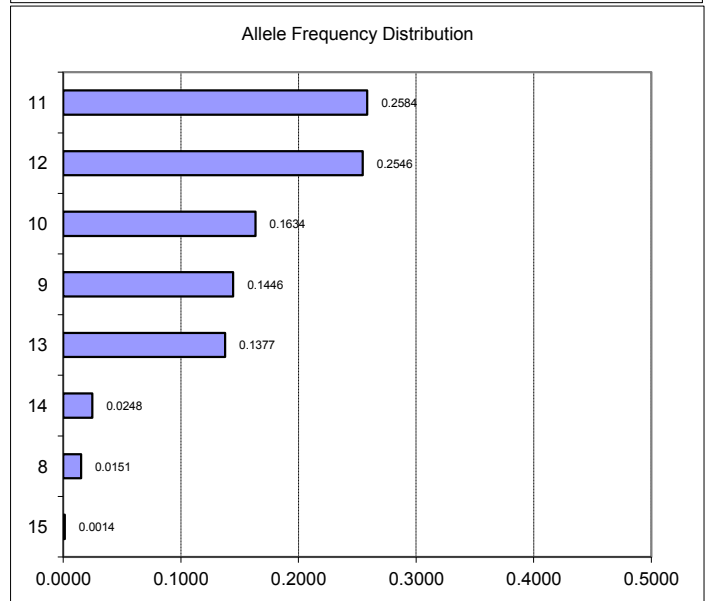
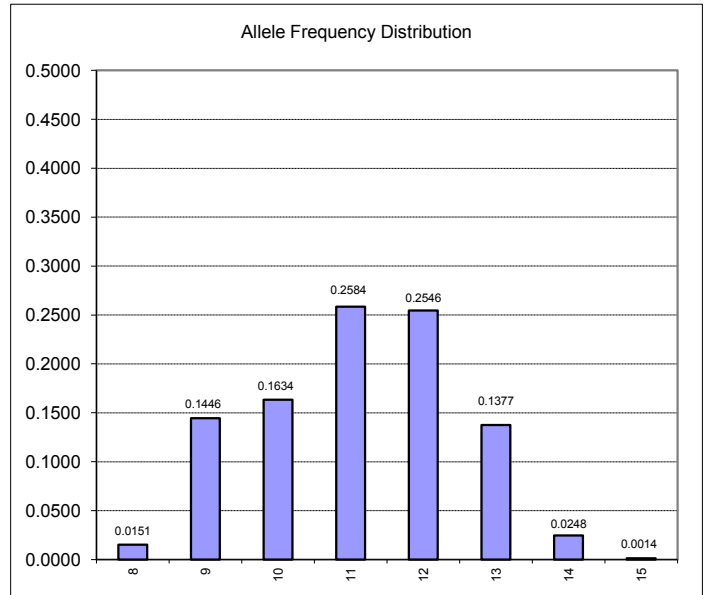
Sampling without repetition



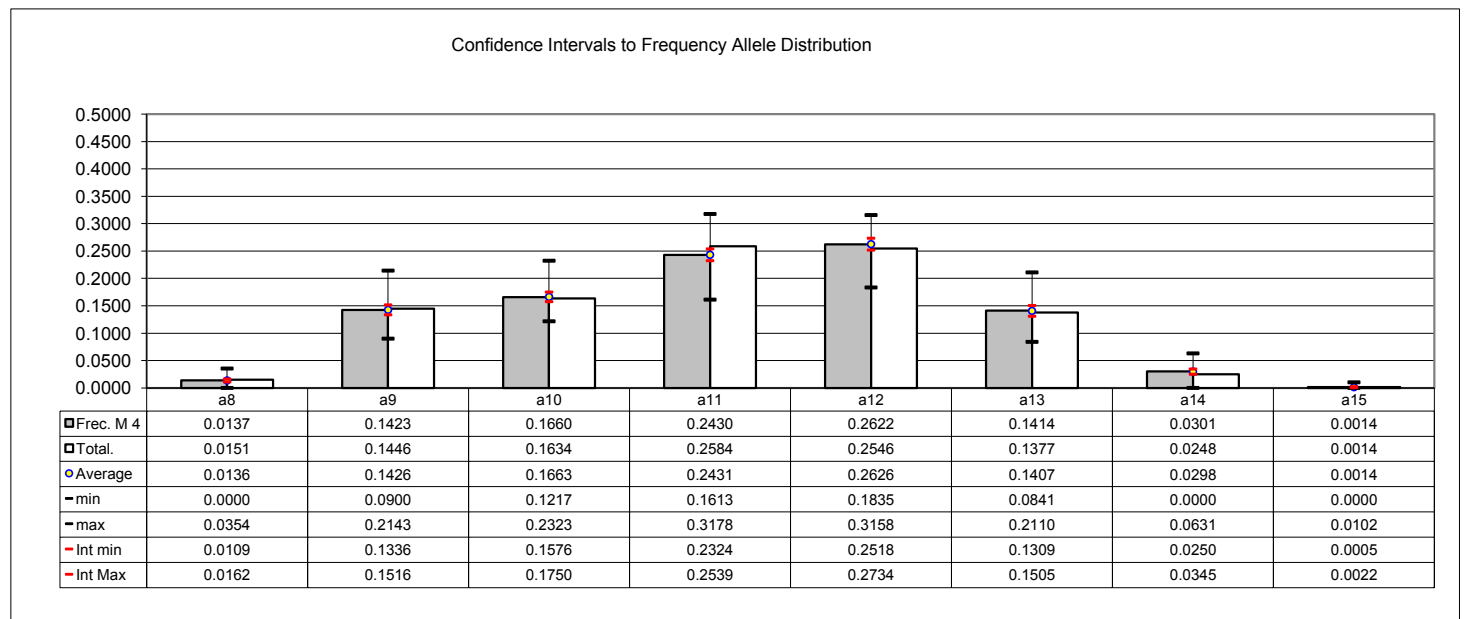
Population	Bogotá
System	D16S539
Alleles number	8
Effective alleles	5

Final frequency.

Allele	M.Weight.	S.D.	Frequency
5	252.37	0.08	Az,Ori0.001
	252.37	0.08	Caribe0.002
8	264.30	0.07	0.0151
9	268.32	0.08	0.1446
10	272.32	0.06	0.1634
11	276.37	0.07	0.2584
12	280.37	0.09	0.2546
13	284.34	0.07	0.1377
14	288.44	0.09	0.0248
15	292.51	0.07	0.0014



Single Alleles	
In the Database	In the Sample
	15

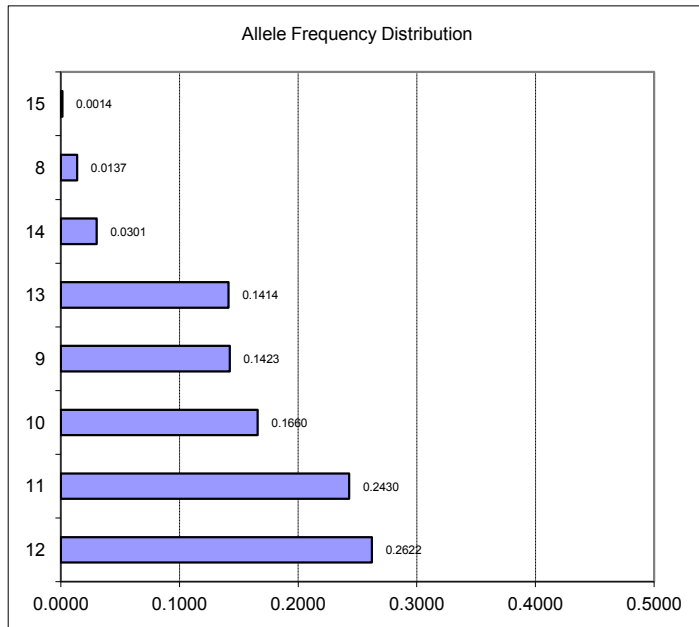


Specific sample analysis.

Sampling Specific.	
Population	Bogotá
System	D16S539
Alleles number	8
Effective alleles number	5
CV	0.05
Sample number	4
Sample size	56
Polimorphism alleles total	10

EXPERIMENT	(Todas)
Sample number	4

Allele frequency summary	
Alleles	Total
12	0.2622
11	0.2430
10	0.1660
9	0.1423
13	0.1414
14	0.0301
8	0.0137
15	0.0014



Single Alleles in the Sampling

EXPERIMENT	(Todas)
Sample number	4

Máx	Alleles	Total
	15	1
	8	4
	14	7
	9	22
	13	25
	10	27
	11	36
	12	36

Sampling and CV.

Sample selection
Analysis of CV.

Alleles	Sample No.
14	0
11	2
13	2
15	0
10	4
12	2
9	2
8	5

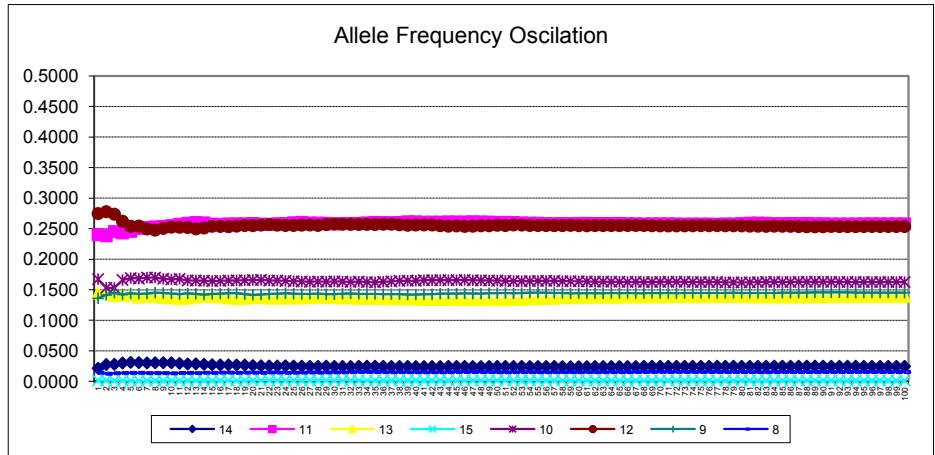
Allelic Distribution Order

Population	Bogotá
System	D16S539

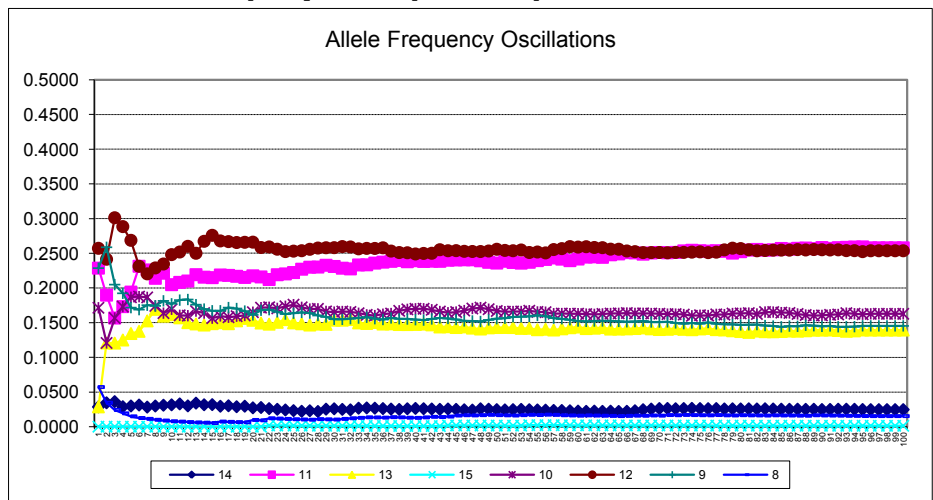
EXPERIMENT	(Todas)
Sample number	(Todas)

Allele frequency summary	
Alleles	Total
11	0.2584
12	0.2546
10	0.1634
9	0.1446
13	0.1377
14	0.0248
8	0.0151
15	0.0014

Compiled Allele Frequency



Allele frequency for a specific experiment.



Analysis for Specific Sampling

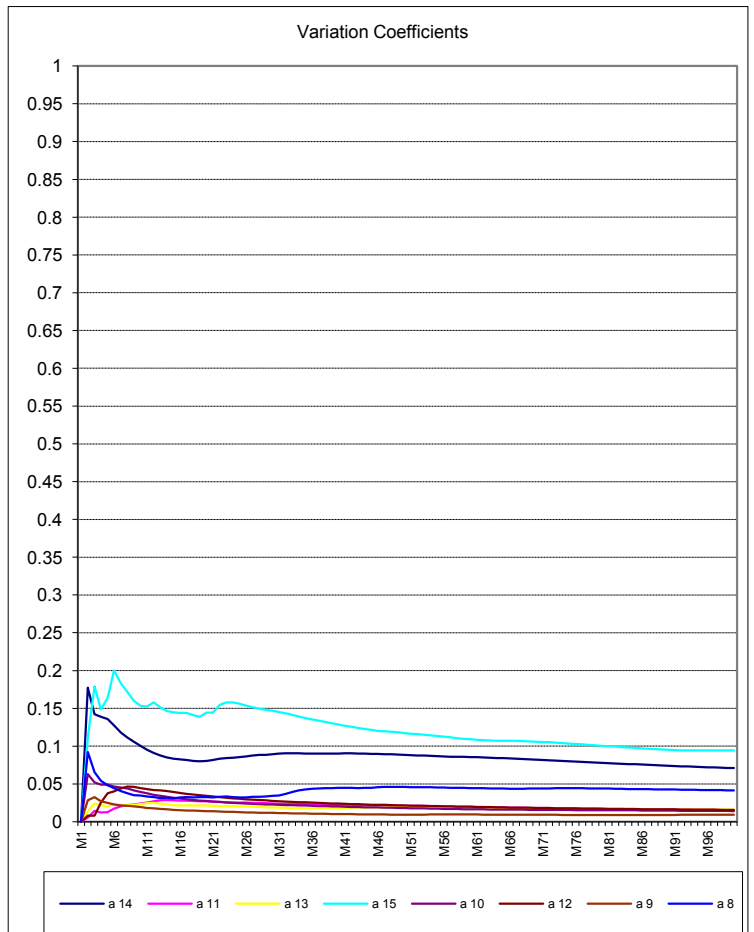
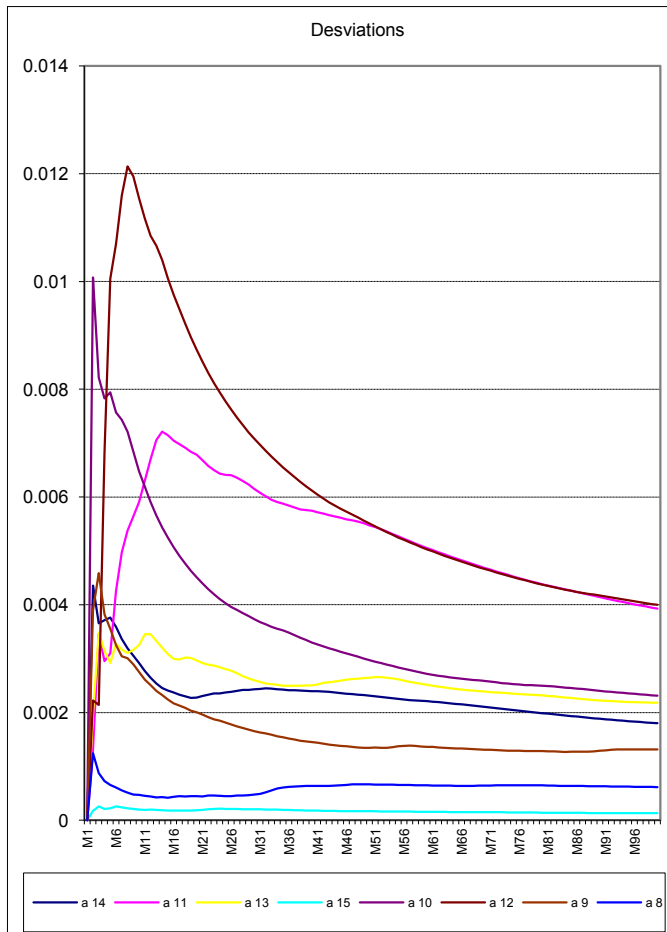
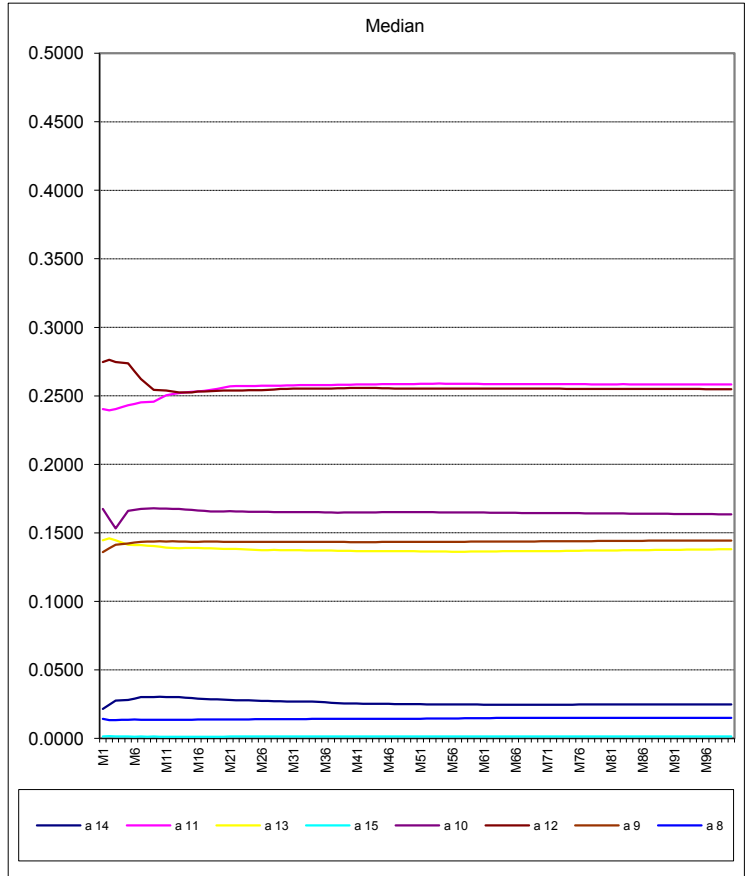
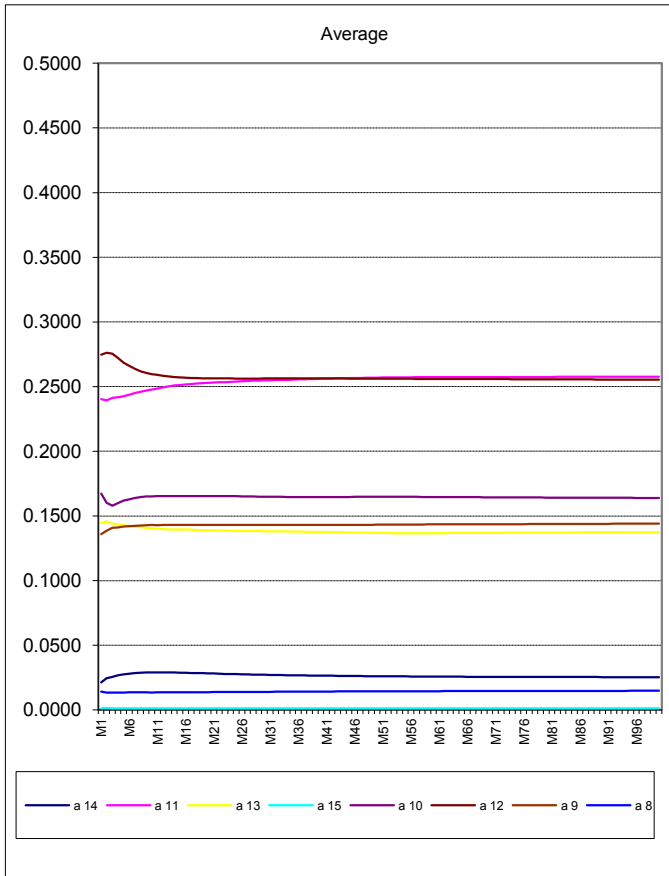
Sample number	4
Sample size	56

Average D	Min	Max
0.0072	0.0009	0.0108

Alleles	Average	min	max	Int min	Int Max	Median	S.D	IC
a8	0.0136	0.0000	0.0354	0.0109	0.0162	0.0104	0.0102	0.0027
a9	0.1426	0.0900	0.2143	0.1336	0.1516	0.1376	0.0344	0.0090
a10	0.1663	0.1217	0.2323	0.1576	0.1750	0.1683	0.0333	0.0087
a11	0.2431	0.1613	0.3178	0.2324	0.2539	0.2489	0.0411	0.0108
a12	0.2626	0.1835	0.3158	0.2518	0.2734	0.2747	0.0413	0.0108
a13	0.1407	0.0841	0.2110	0.1309	0.1505	0.1324	0.0376	0.0098
a14	0.0298	0.0000	0.0631	0.0250	0.0345	0.0284	0.0182	0.0048
a15	0.0014	0.0000	0.0102	0.0005	0.0022	0.0000	0.0034	0.0009

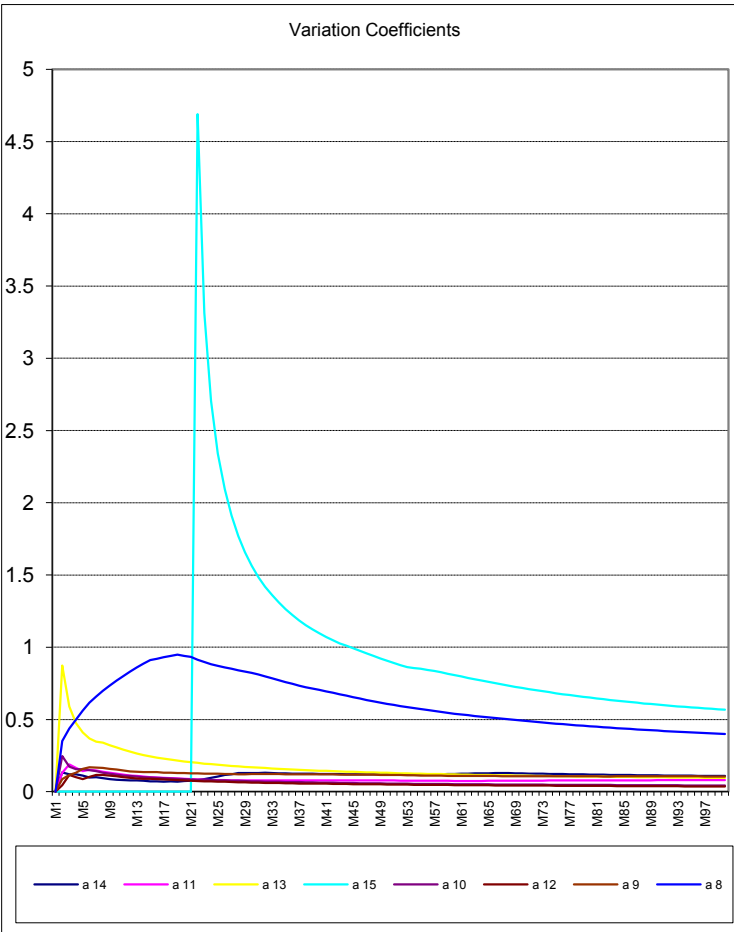
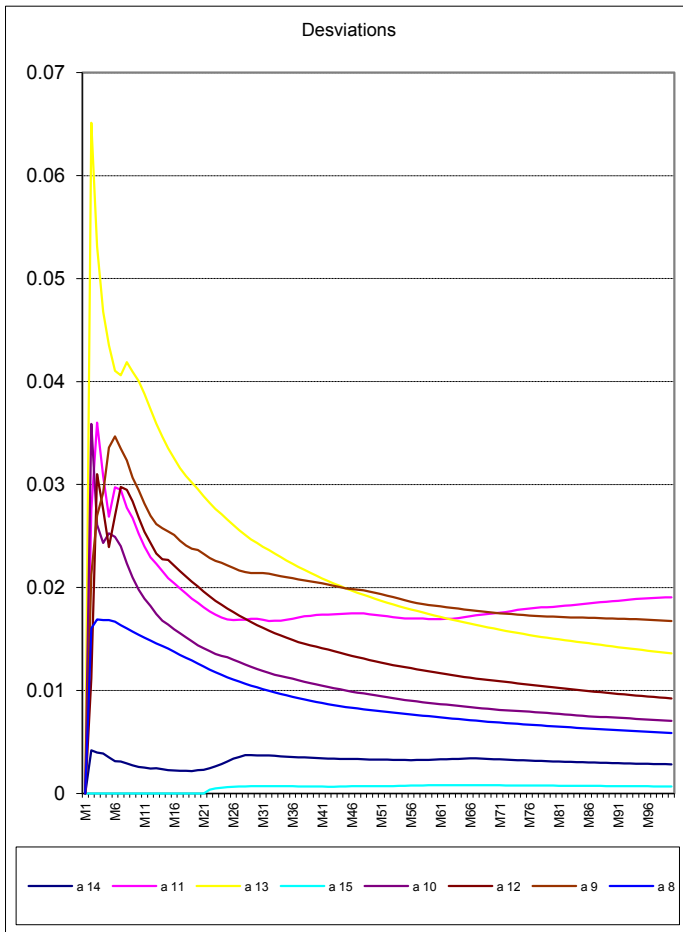
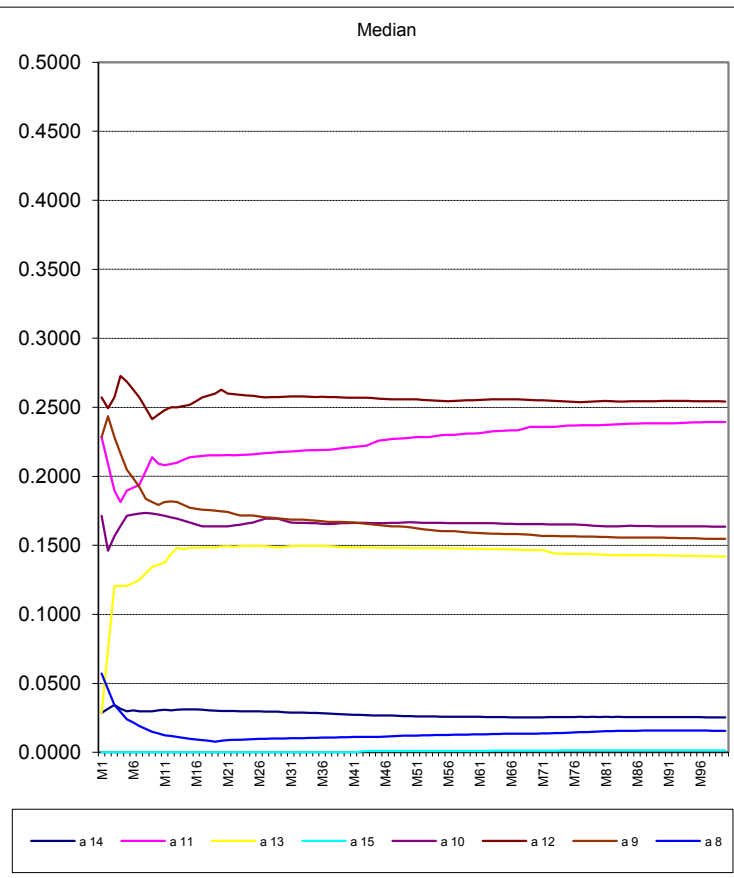
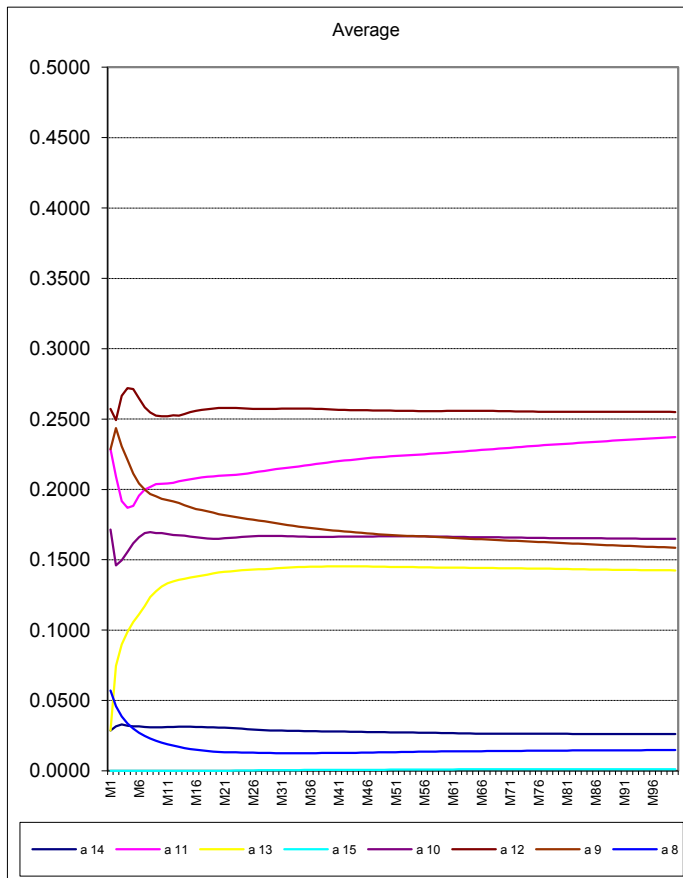
Graphic analysis of polymorphisms.

System D16S539 Pop. Bogotá



Graphic analysis of polymorphisms.

System D16S539 Pop. Bogotá



Graphic analysis of polymorphisms.

System	D16S539	Pop.	Bogotá
--------	---------	------	--------

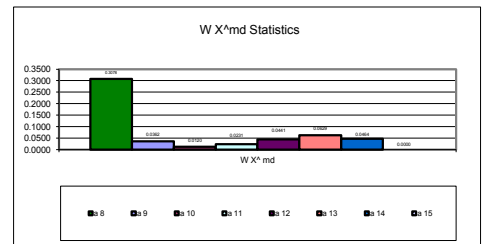
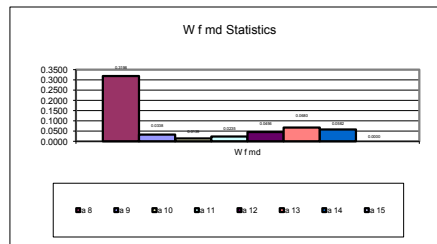
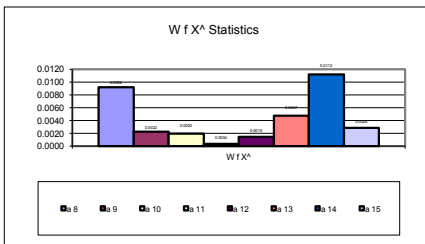
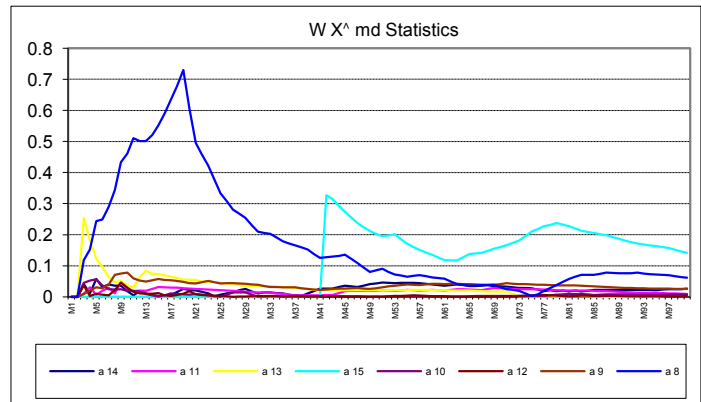
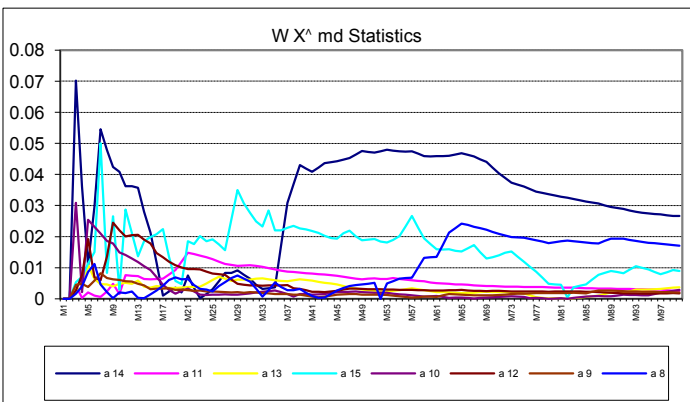
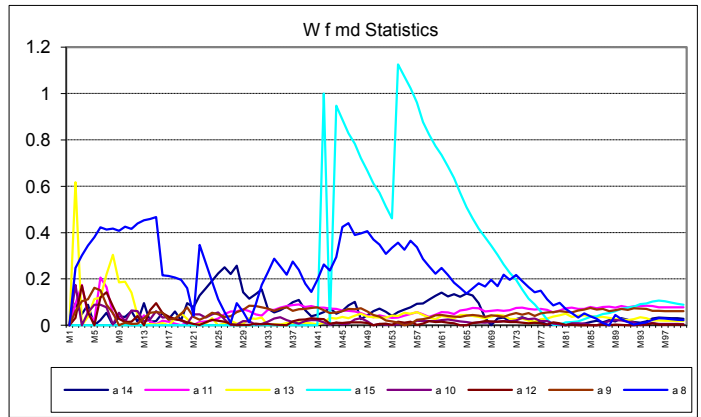
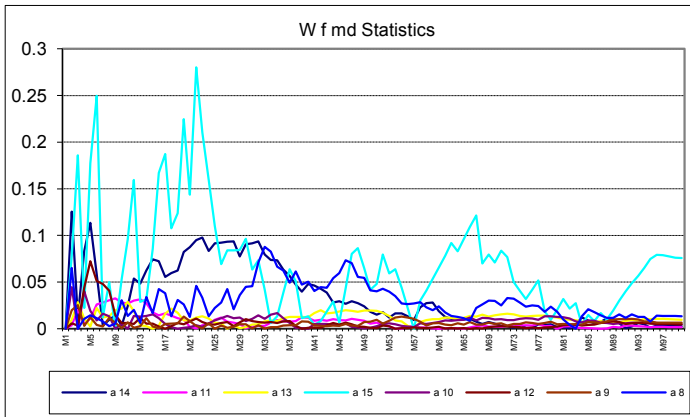
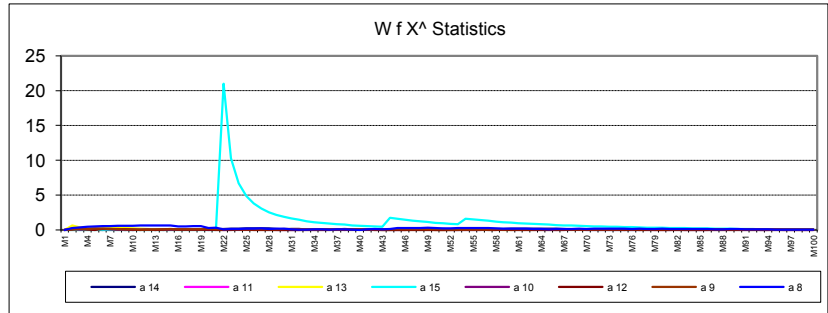
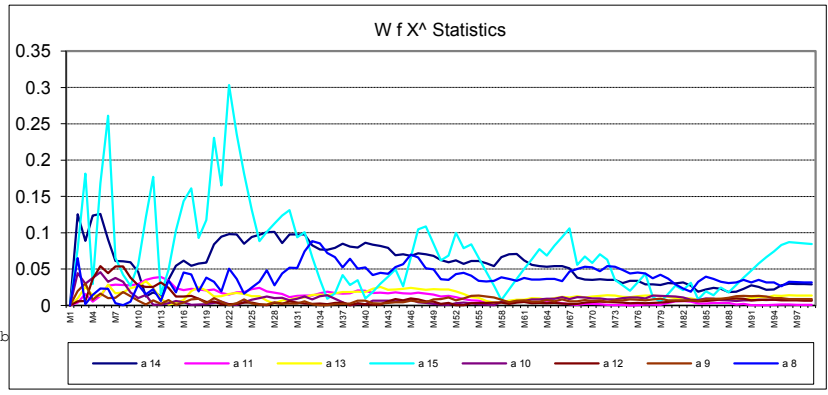
W Statistics to Specific sample

Alleles	W f X [^]	W f md	W X [^] md
a 8	0.0092	0.3198	0.3078
a 9	0.0022	0.0338	0.0362
a 10	0.0020	0.0139	0.0120
a 11	0.0004	0.0235	0.0231
a 12	0.0015	0.0456	0.0441
a 13	0.0047	0.0680	0.0629
a 14	0.0112	0.0582	0.0464
a 15	0.0028		

Indeterminate value (Division b

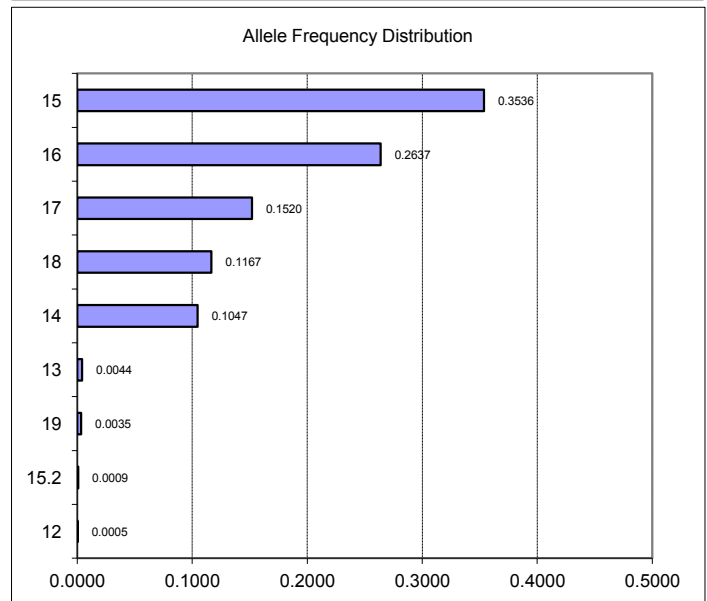
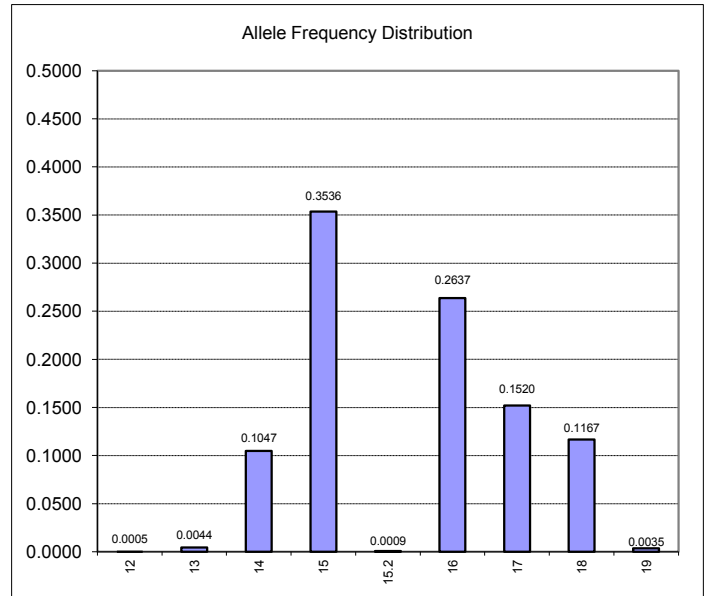
0.0112 0.3198 0.3078

Muestras sin repeticiones



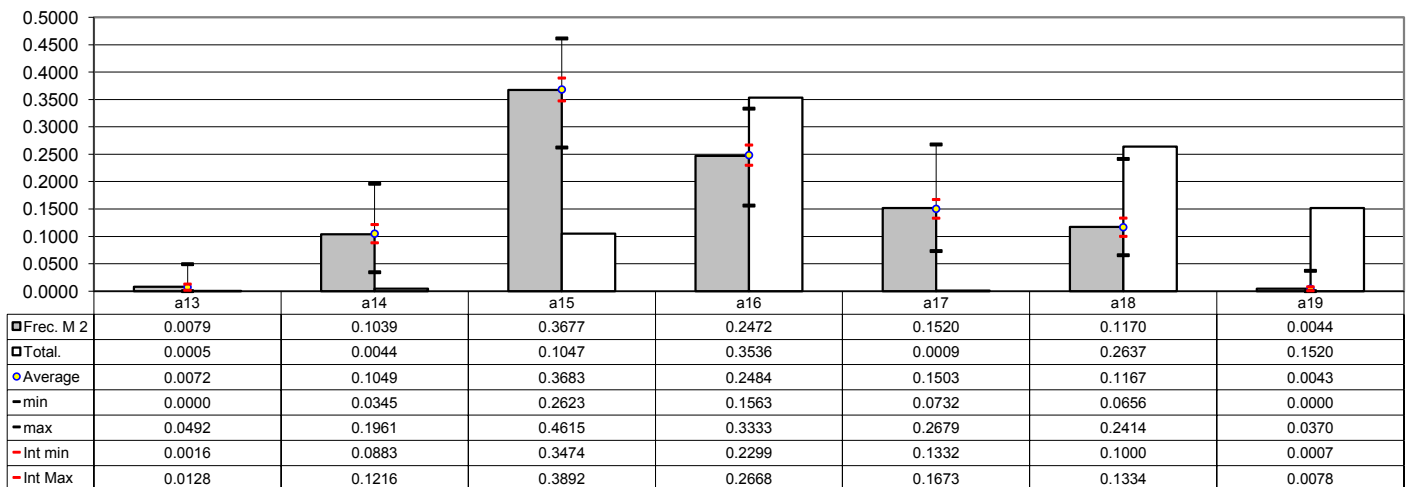
Population	Bogotá
System	D3S1358
Alleles number	9
Effective alleles	4

Allele	M.Weight.	S.D.	Frequency
12	111.96	0.06	0.0005
13	116.04	0.04	0.0044
14	119.99	0.04	0.1047
15	123.89	0.02	0.3536
15.2	127.00		0.0009
16	128.06	0.05	0.2637
17	132.24	0.05	0.1520
18	136.30	0.06	0.1167
19	140.43	0.03	0.0035
20	145.00		0.0010



Single Alleles	
In the Database	In the Sample
12	

Confidence Intervals to Frequency Allele Distribution

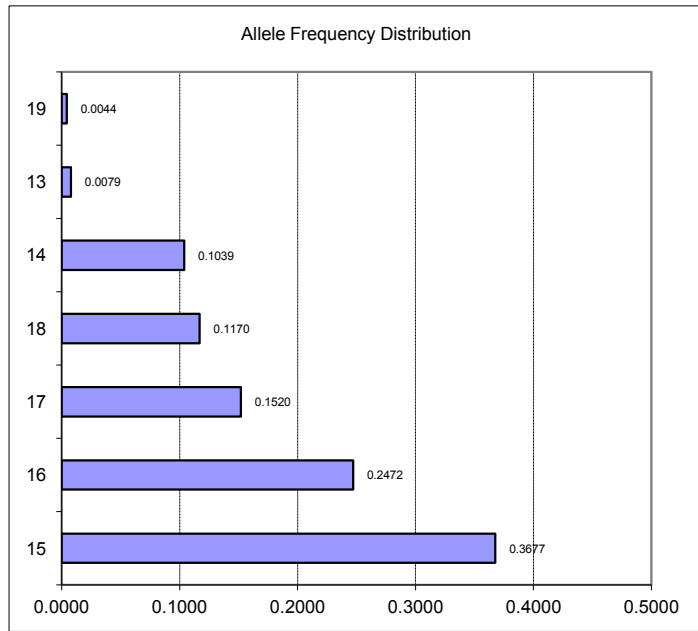


Specific sample analysis.

Sampling Specific.	
Population	Bogotá
System	D3S1358
Alleles number	7
Effective alleles number	4
CV	0.05
Sample number	2
Sample size	29
Polimorphism alleles total	20

EXPERIMENT	(Todas)
Sample number	2

Allele frequency summary	
Alleles	Total
15	0.3677
16	0.2472
17	0.1520
18	0.1170
14	0.1039
13	0.0079
19	0.0044



Single Alleles in the Sampling

EXPERIMENT	(Todas)
Sample number	2

Max	Total
Alleles	
19	2
13	3
14	11
18	14
17	15
16	20
15	28

Sampling and CV.

Sample selection
Analysis of CV.

Alleles	Sample No.
14	2
13	0
15	2
12	0
17	2
16	2
19	0
18	2
15.2	0

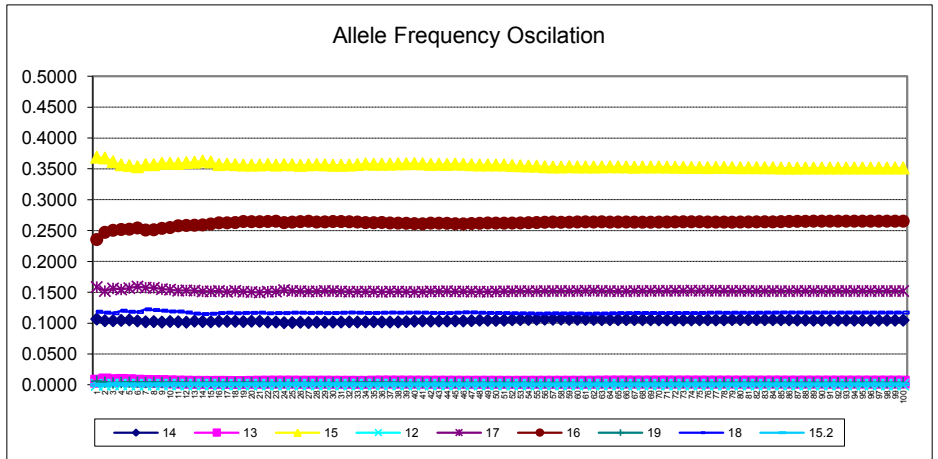
Allelic Distribution Order

Population	Bogotá
System	D3S1358

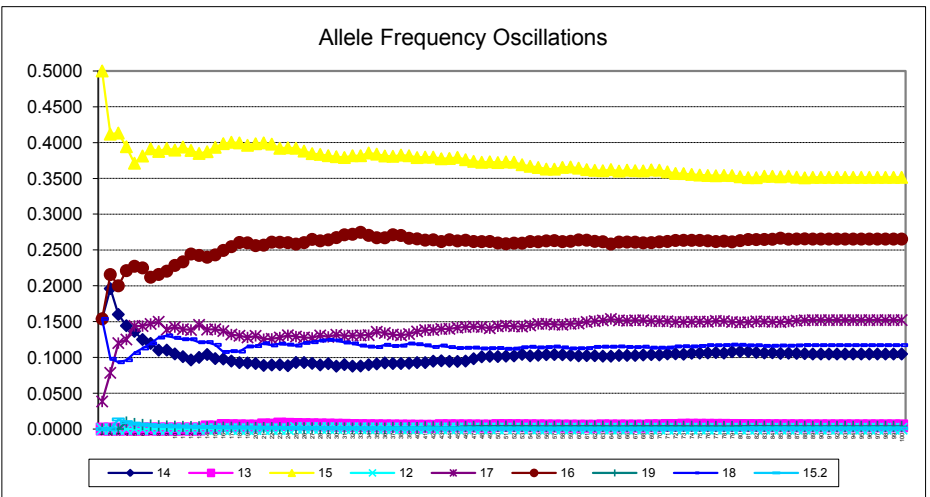
EXPERIMENT	(Todas)
Sample number	(Todas)

Allele frequency summary		
Alleles		Total
	15	0.3536
	16	0.2637
	17	0.1520
	18	0.1167
	14	0.1047
	13	0.0044
	19	0.0035
	15.2	0.0009
	12	0.0005

Compiled Allele Frequency



Allele frequency for a specific experiment.



Analysis for Specific Sampling

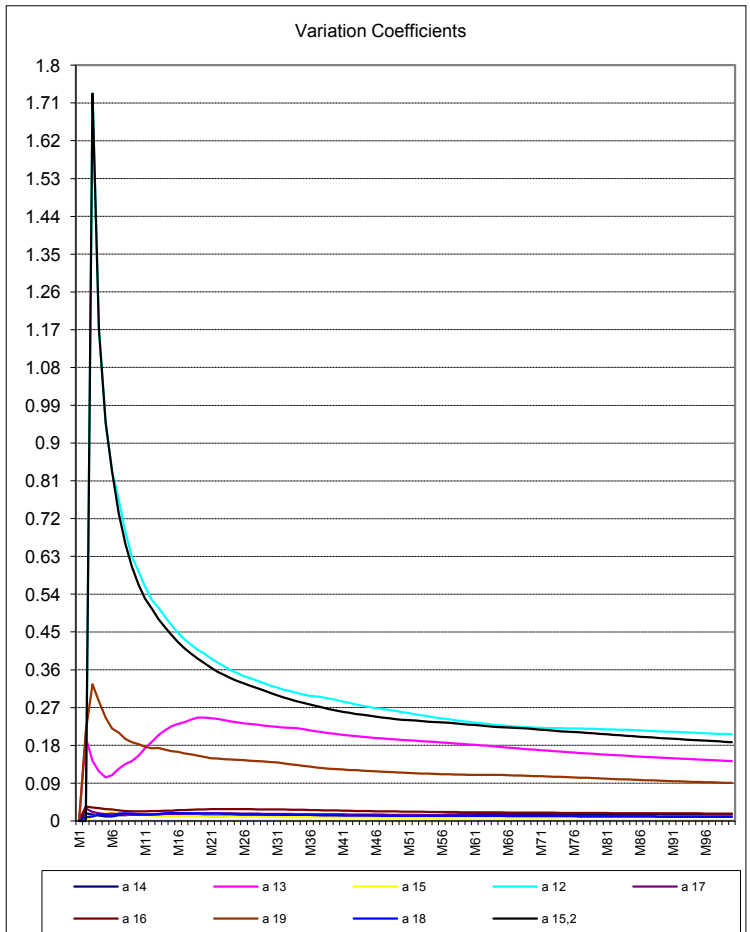
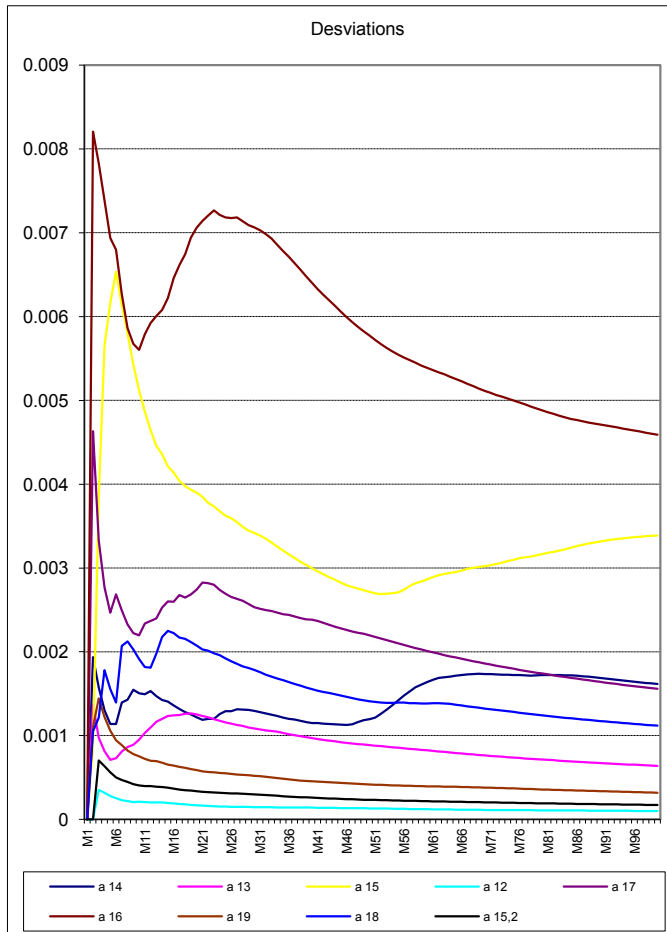
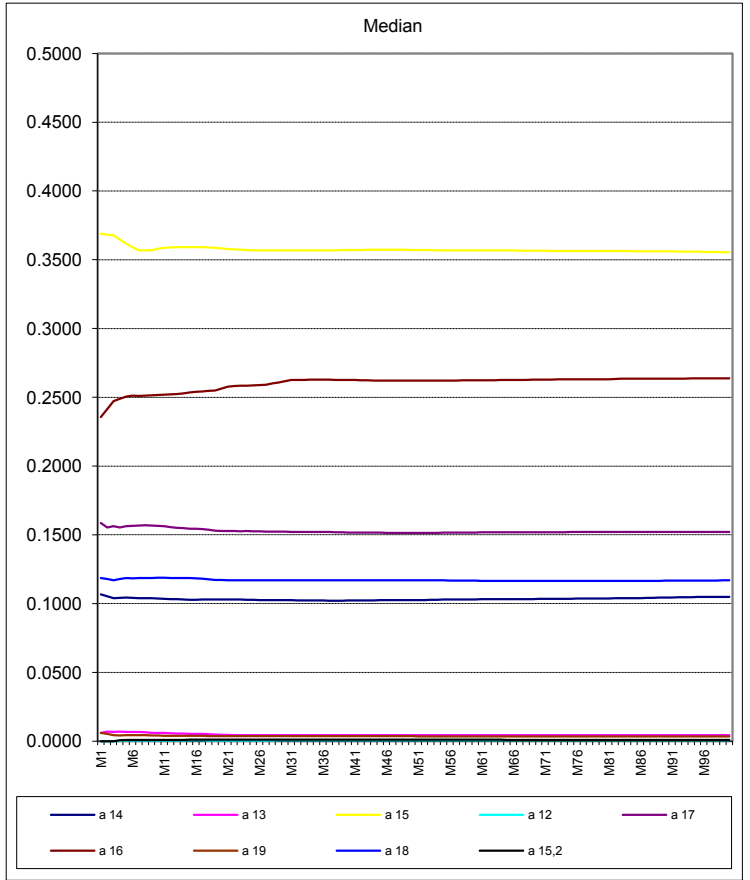
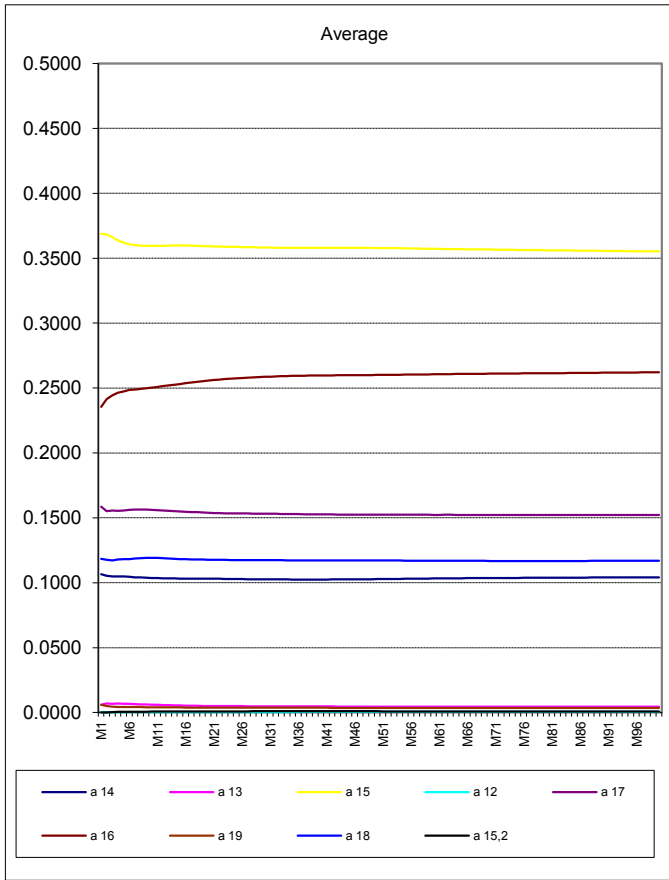
Sample number	2
Sample size	29

Average D	Min	Max
0.0141	0.0035	0.0209

Alleles	Average	min	max	Int min	Int Max	Median	S.D	IC
a13	0.0072	0.0000	0.0492	0.0016	0.0128	0.0000	0.0154	0.0056
a14	0.1049	0.0345	0.1961	0.0883	0.1216	0.1053	0.0457	0.0166
a15	0.3683	0.2623	0.4615	0.3474	0.3892	0.3793	0.0573	0.0209
a16	0.2484	0.1563	0.3333	0.2299	0.2668	0.2578	0.0508	0.0185
a17	0.1503	0.0732	0.2679	0.1332	0.1673	0.1484	0.0469	0.0171
a18	0.1167	0.0656	0.2414	0.1000	0.1334	0.1038	0.0459	0.0167
a19	0.0043	0.0000	0.0370	0.0007	0.0078	0.0000	0.0097	0.0035

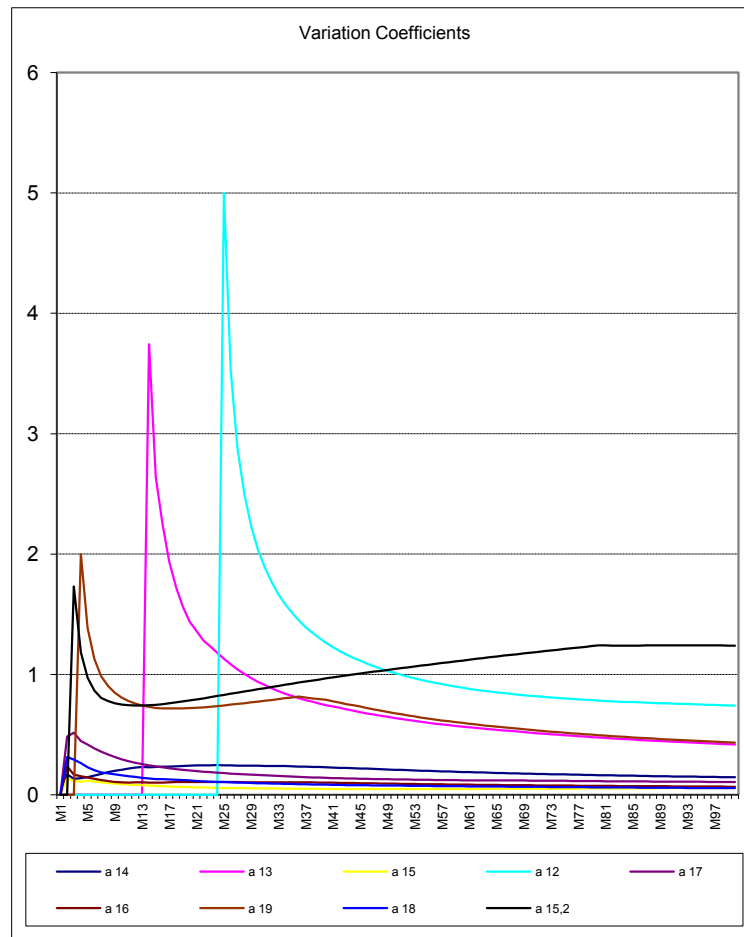
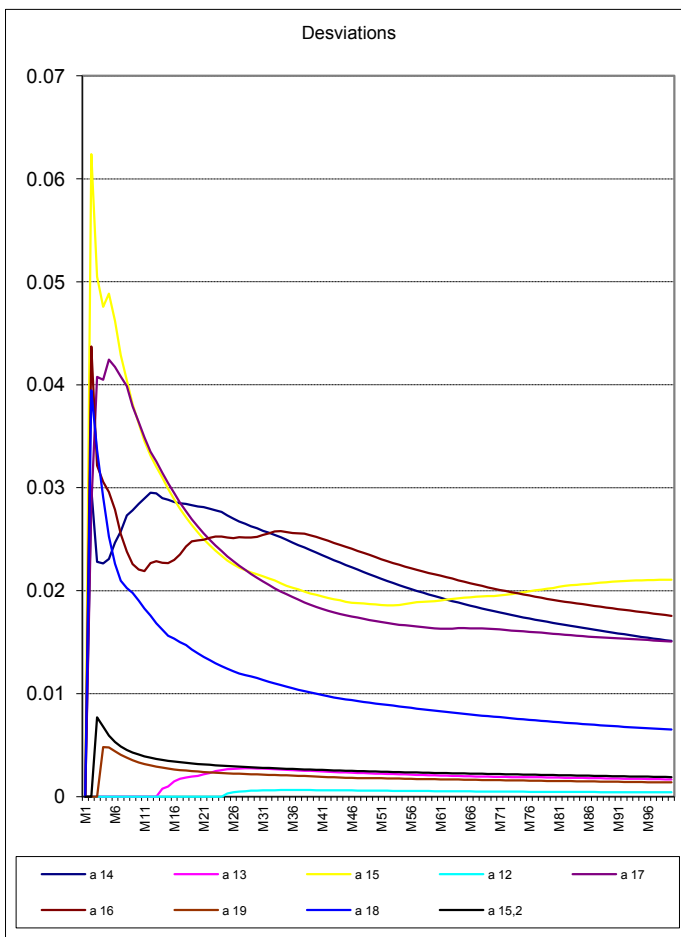
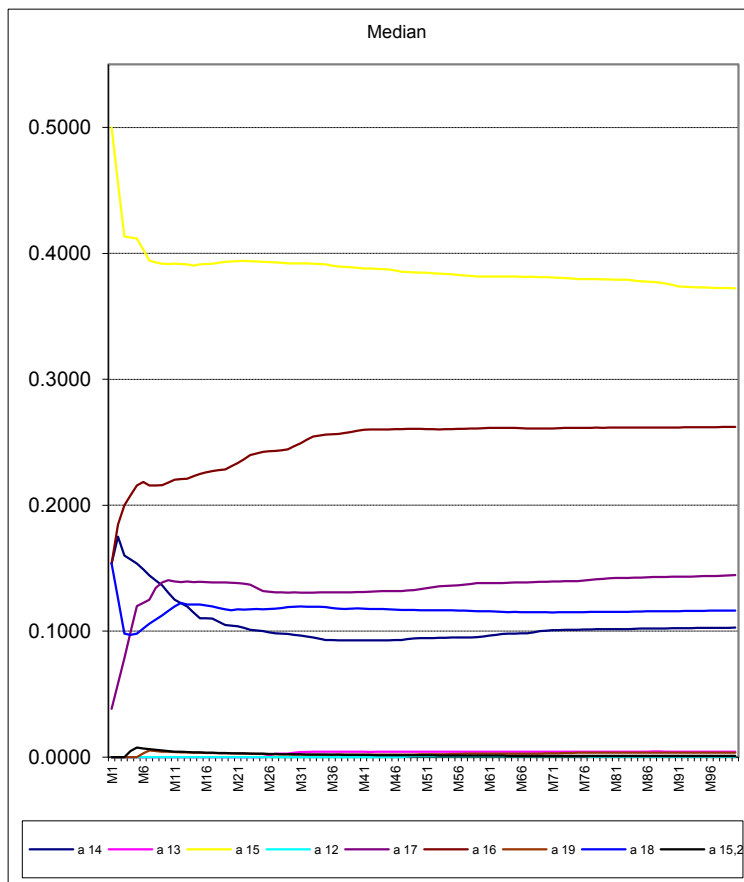
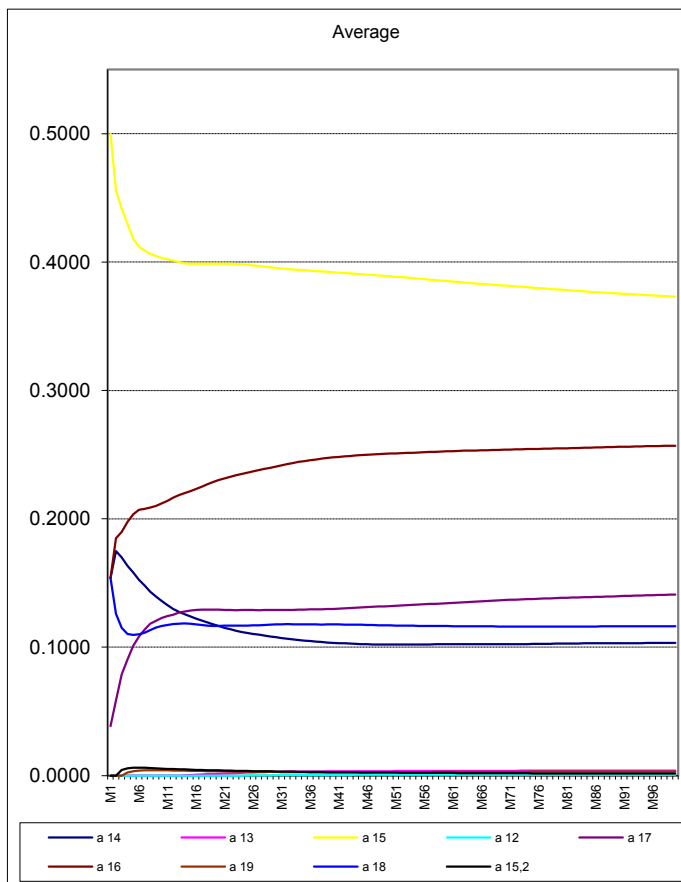
Graphic analysis of polymorphisms.

System D3S1358 Pop. Bogotá



Graphic analysis of polymorphisms.

System D3S1358 Pop. Bogotá



Graphic analysis of polymorphisms.

System	D3S1358	Pop.	Bogotá
--------	---------	------	--------

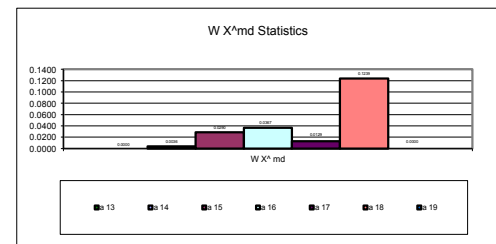
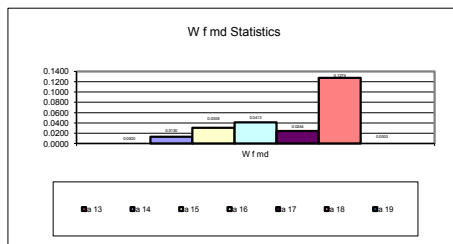
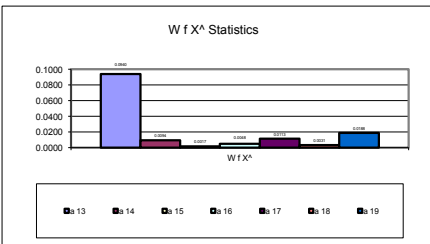
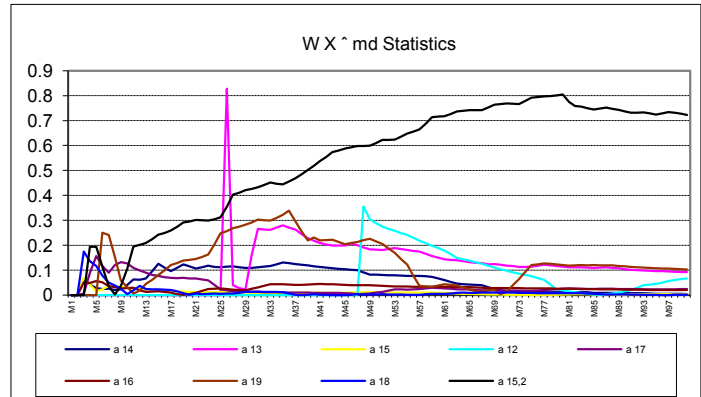
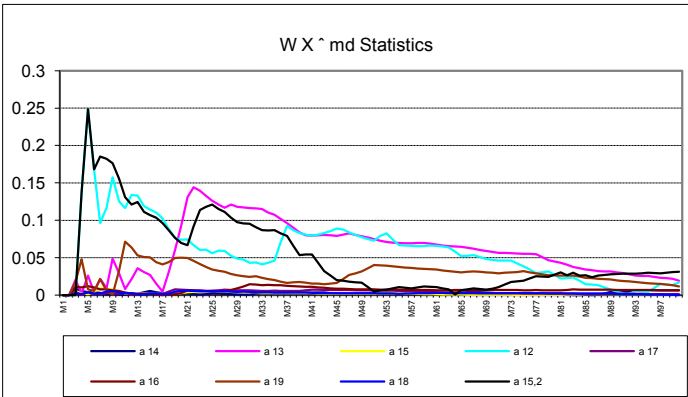
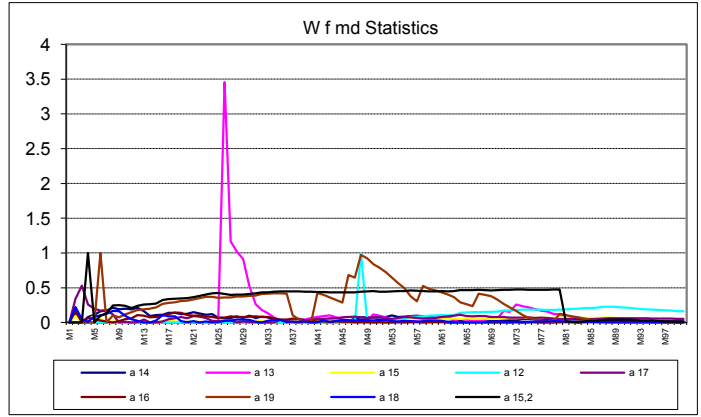
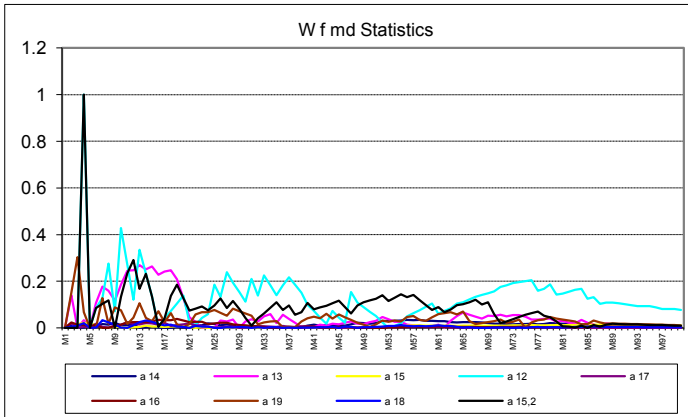
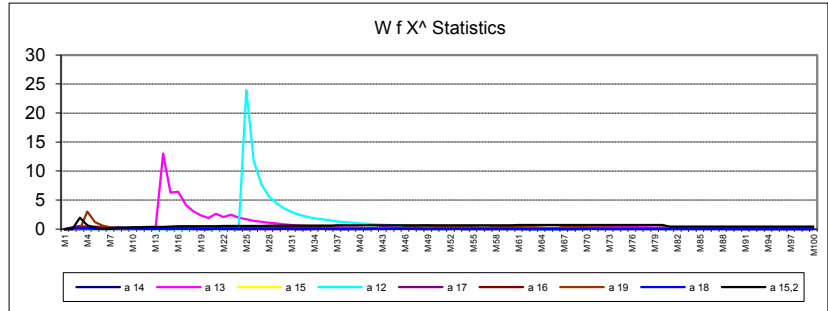
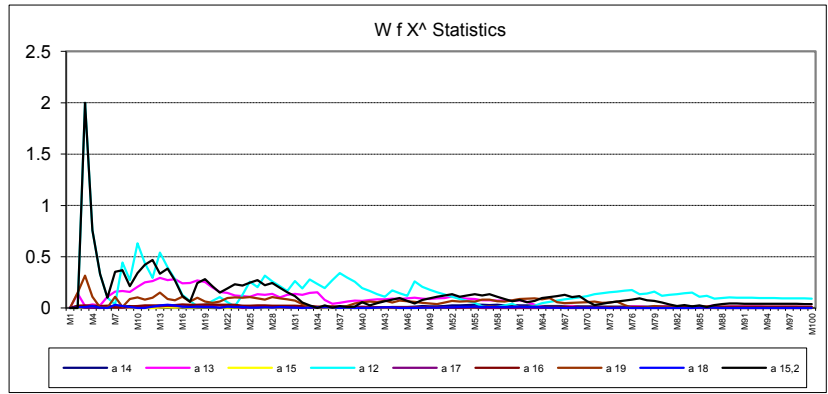
W Statistics to Specific sample

Alleles	W f X [^]	W f md	W X [^] md
a 13	0.0940		
a 14	0.0094	0.0130	0.0036
a 15	0.0017	0.0306	0.0290
a 16	0.0048	0.0413	0.0367
a 17	0.0113	0.0244	0.0129
a 18	0.0031	0.1274	0.1239
a 19	0.0188		

Indeterminate value (Division by zero)

0.0940 0.1274 0.1239

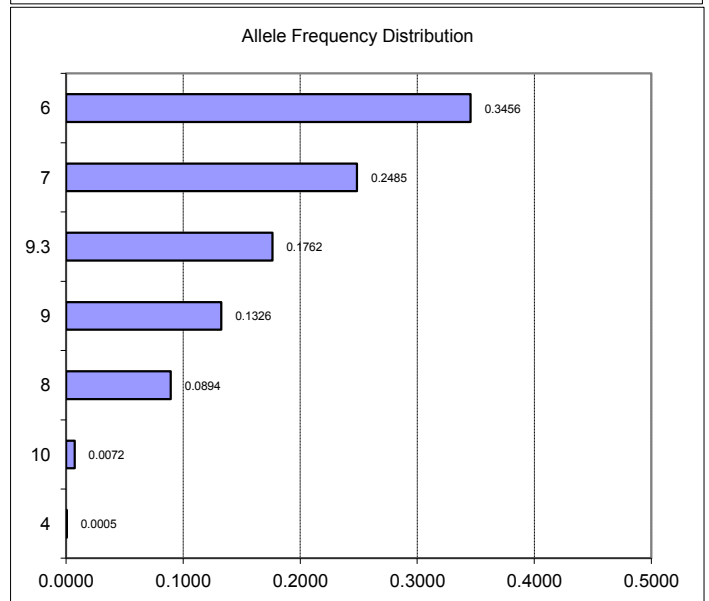
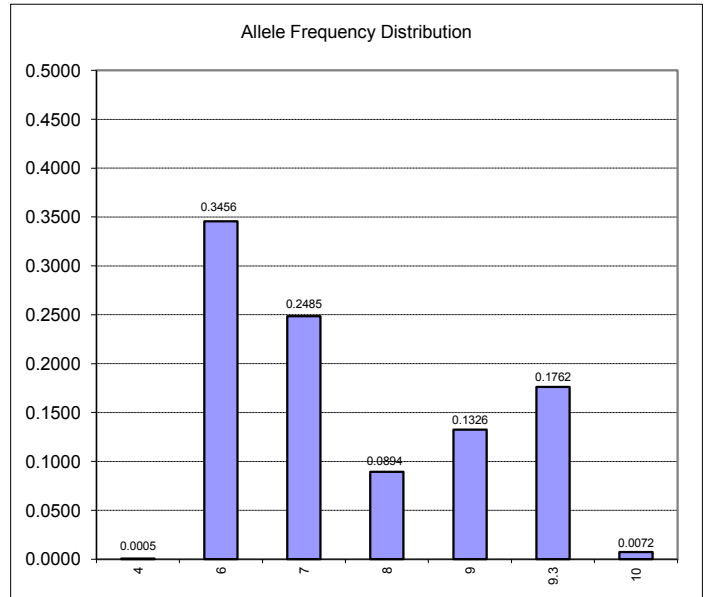
Sampling without repetition



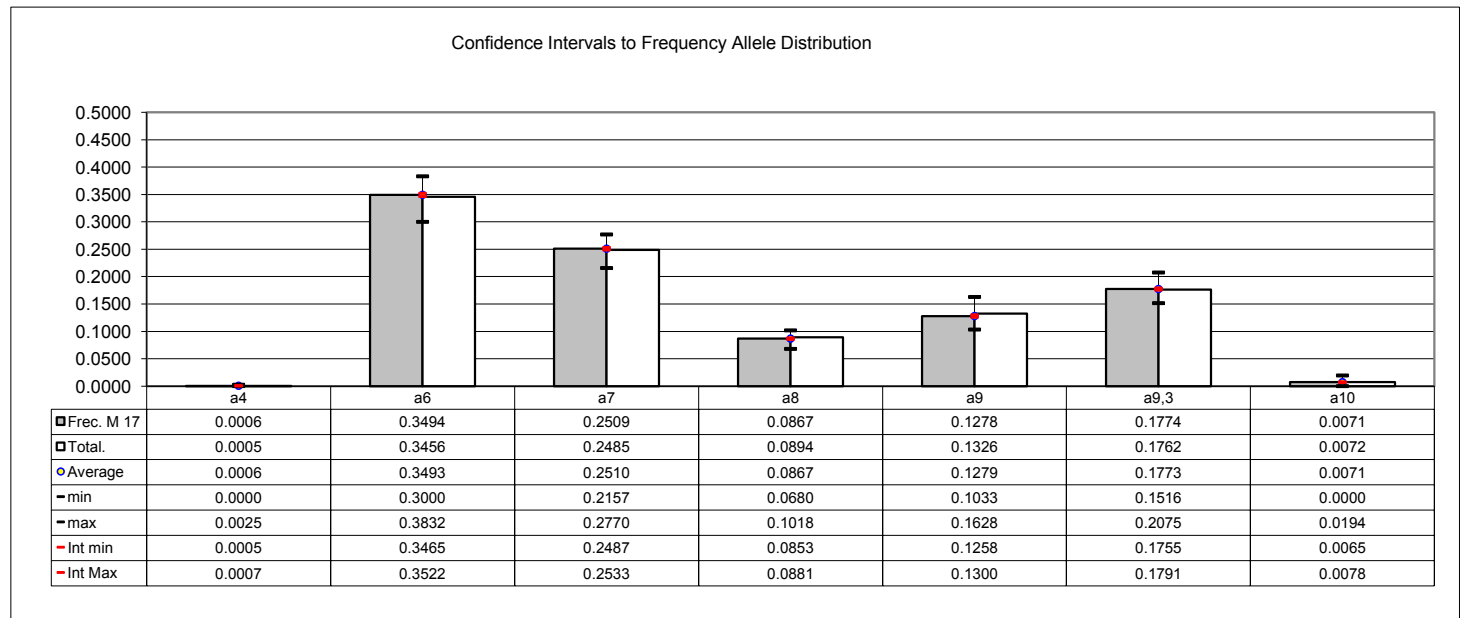
Population	Bogotá
System	TH01
Alleles number	7
Effective alleles	4

Final frequency.

Allele	M.Weight.	S.D.	Frequency
4	163.29	0.04	0.0005
5	167.36	0.03	Caribe:0.0020
	167.36	0.03	Chocó: 0.0040
6	171.40	0.05	0.3456
7	175.40	0.03	0.2485
8	179.38	0.04	0.0894
9	183.36	0.05	0.1326
9.3	186.93	0.02	0.1762
10	187.29	0.04	0.0072
11	191.23	0.03	
13.3	201.94	0.05	



Single Alleles	
In the Database	In the Sample
4	4

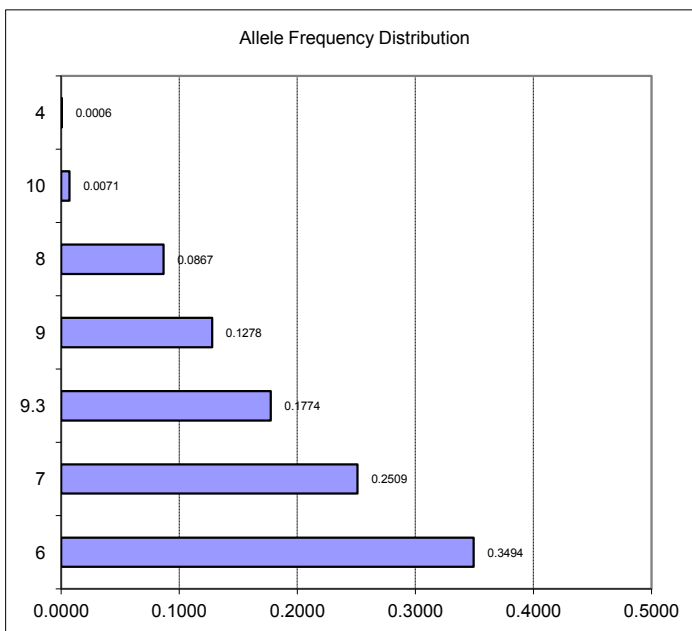


Specific sample analysis.

Sampling Specific.	
Population	Bogotá
System	TH01
Alleles number	7
Effective alleles number	4
CV	0.05
Sample number	17
Sample size	208
Polimorphism alleles total	21

EXPERIMENT	(Todas)
Sample number	17

Allele frequency summary	
Alleles	Total
6	0.3494
7	0.2509
9.3	0.1774
9	0.1278
8	0.0867
10	0.0071
4	0.0006



Single Alleles in the Sampling

EXPERIMENT	(Todas)
Sample number	17

Max	Total
Alleles	Total
4	1
10	8
8	42
9	64
9.3	87
7	120
6	157

Sampling and CV.

Sample selection
Analysis of CV.

Alleles	Sample No.
10	0
9	3
8	17
6	2
7	2
9.3	2
4	0

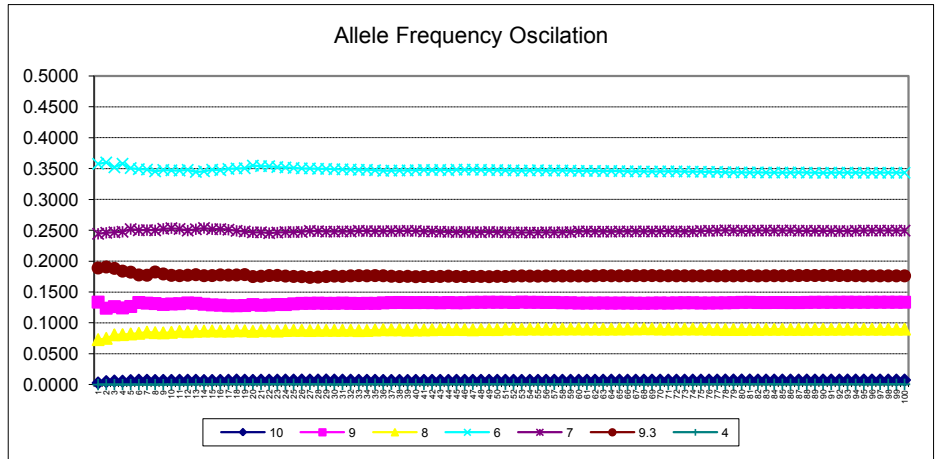
Allelic Distribution Order

Population	Bogotá
System	TH01

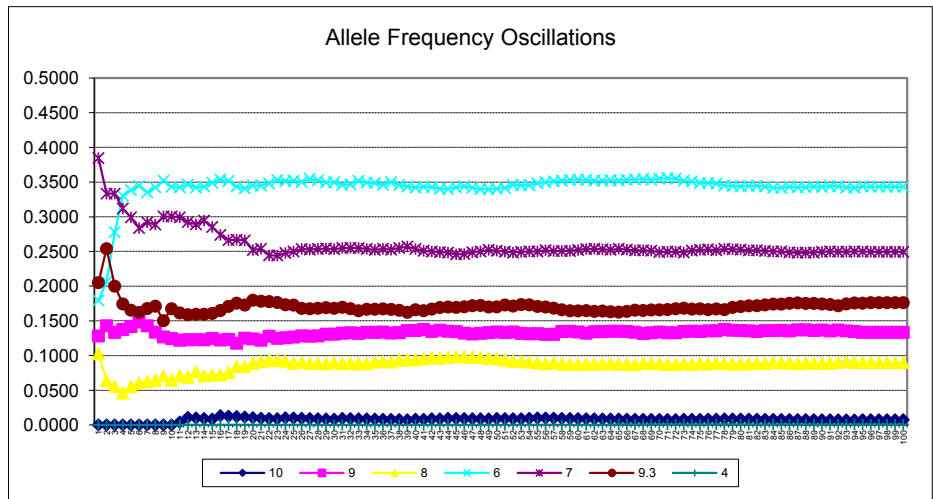
EXPERIMENT	(Todas)
Sample number	(Todas)

Allele frequency summary		Total
Alleles		
	6	0.3456
	7	0.2485
	9.3	0.1762
	9	0.1326
	8	0.0894
	10	0.0072
	4	0.0005

Compiled Allele Frequency



Allele frequency for a specific experiment.



Analysis for Specific Sampling

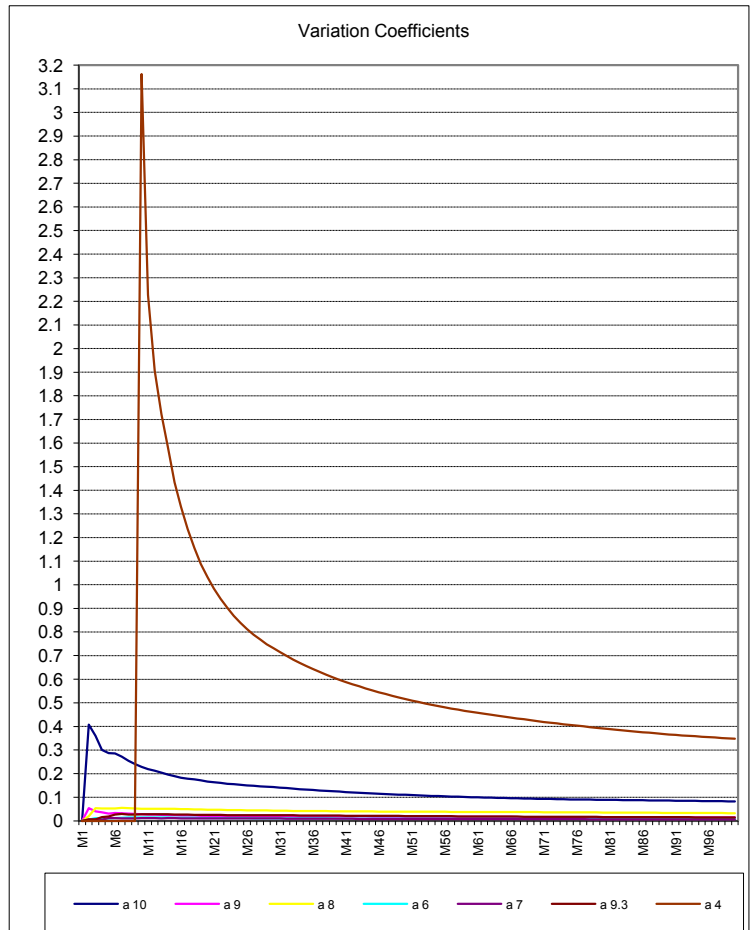
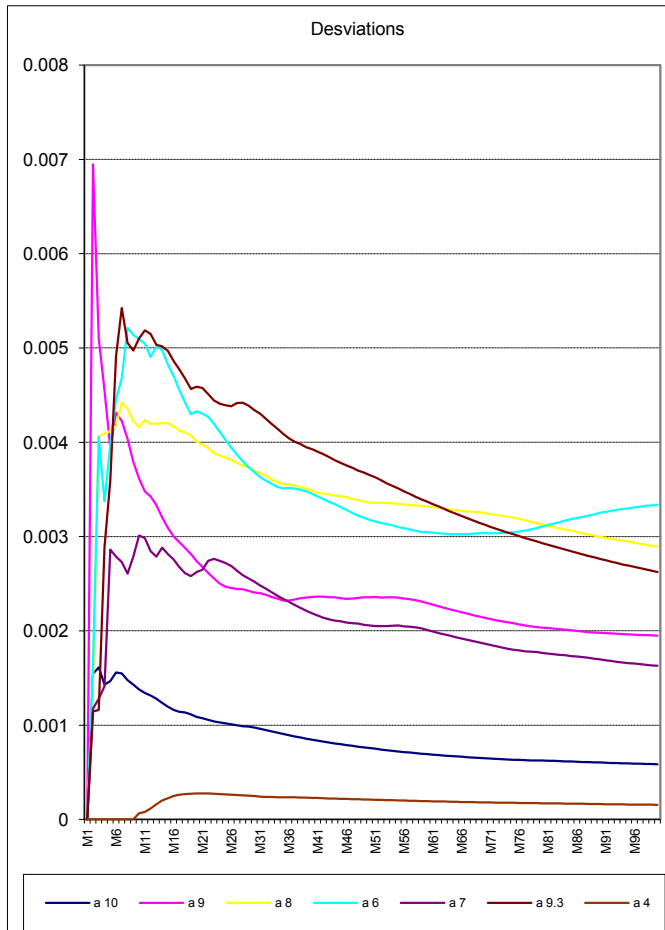
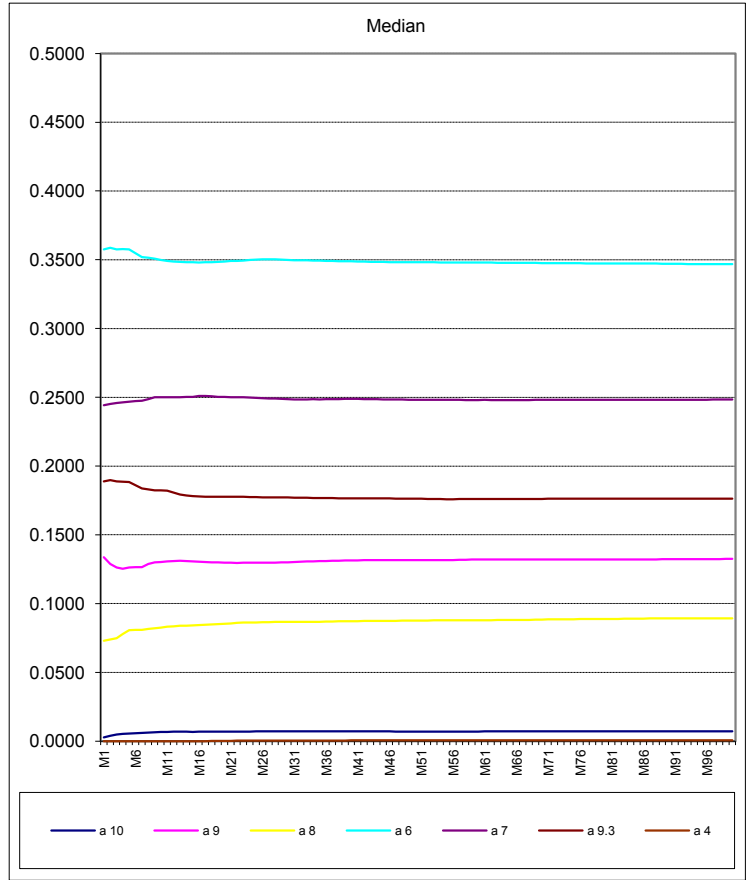
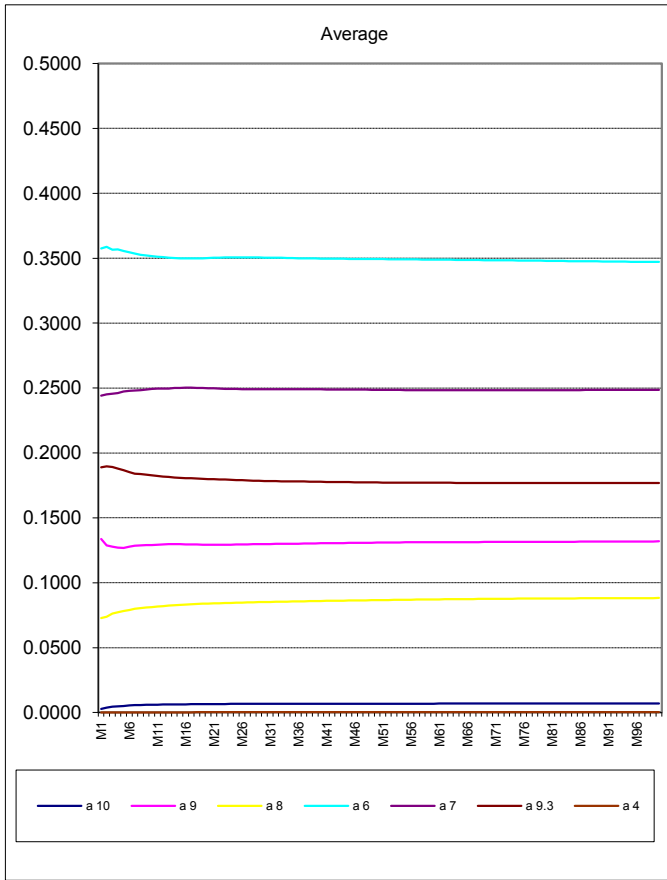
Sample number	17
Sample size	208

Average D	Min	Max
0.0016	0.0001	0.0029

Alleles	Average	min	max	Int min	Int Max	Median	S.D	IC
a4	0.0006	0.0000	0.0025	0.0005	0.0007	0.0000	0.0011	0.0001
a6	0.3493	0.3000	0.3832	0.3465	0.3522	0.3508	0.0211	0.0029
a7	0.2510	0.2157	0.2770	0.2487	0.2533	0.2500	0.0170	0.0023
a8	0.0867	0.0680	0.1018	0.0853	0.0881	0.0865	0.0101	0.0014
a9	0.1279	0.1033	0.1628	0.1258	0.1300	0.1293	0.0153	0.0021
a9,3	0.1773	0.1516	0.2075	0.1755	0.1791	0.1782	0.0133	0.0018
a10	0.0071	0.0000	0.0194	0.0065	0.0078	0.0058	0.0049	0.0007

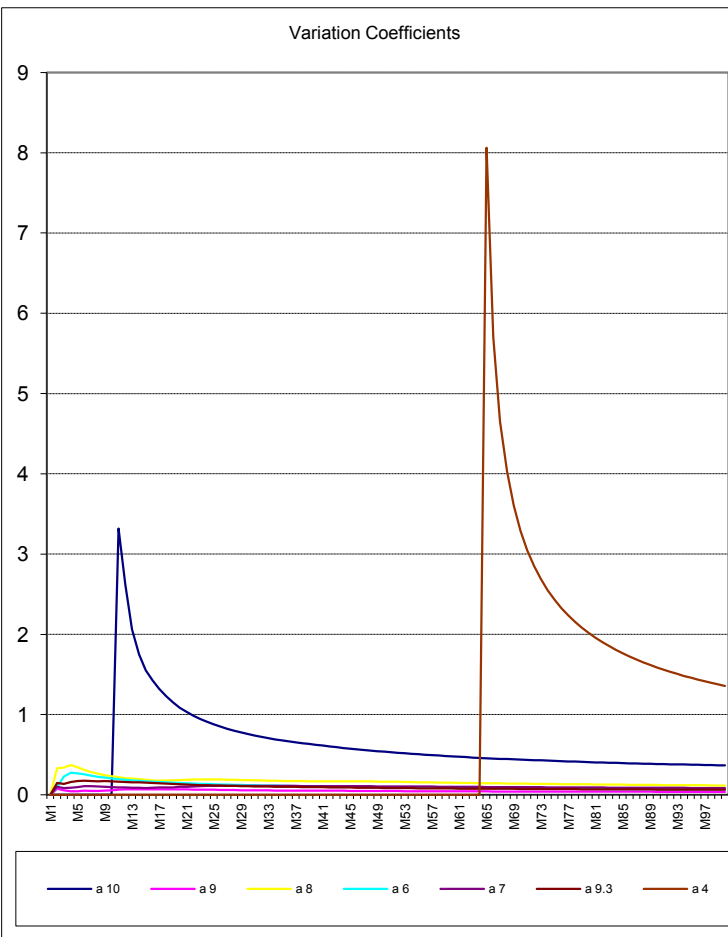
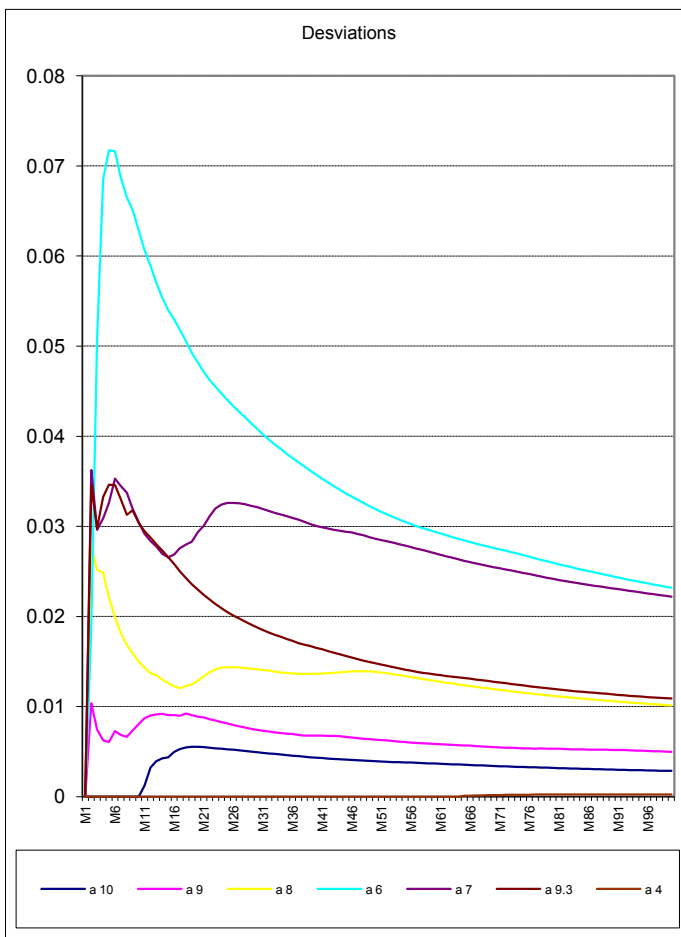
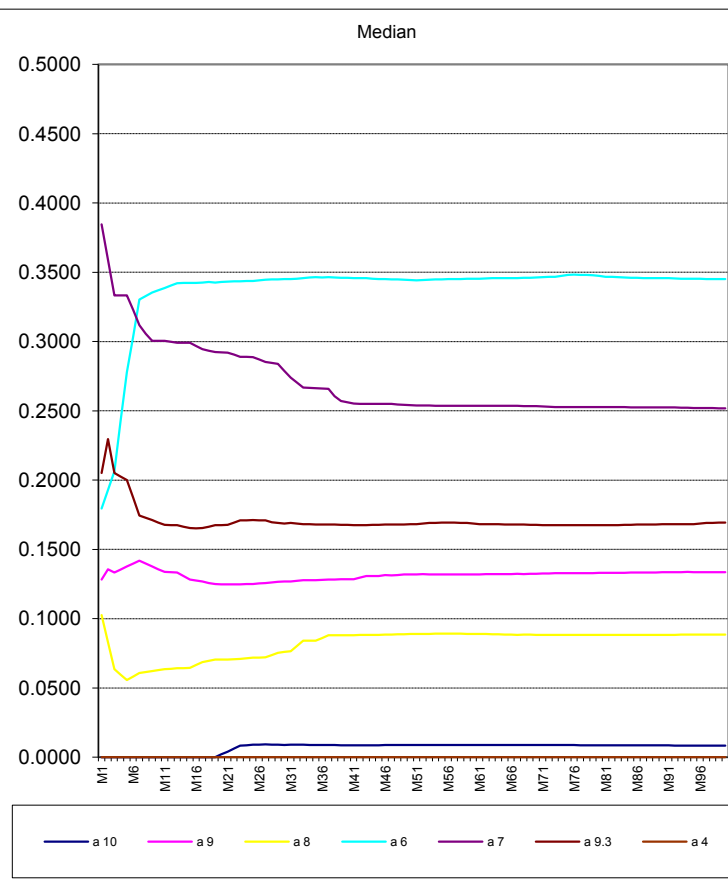
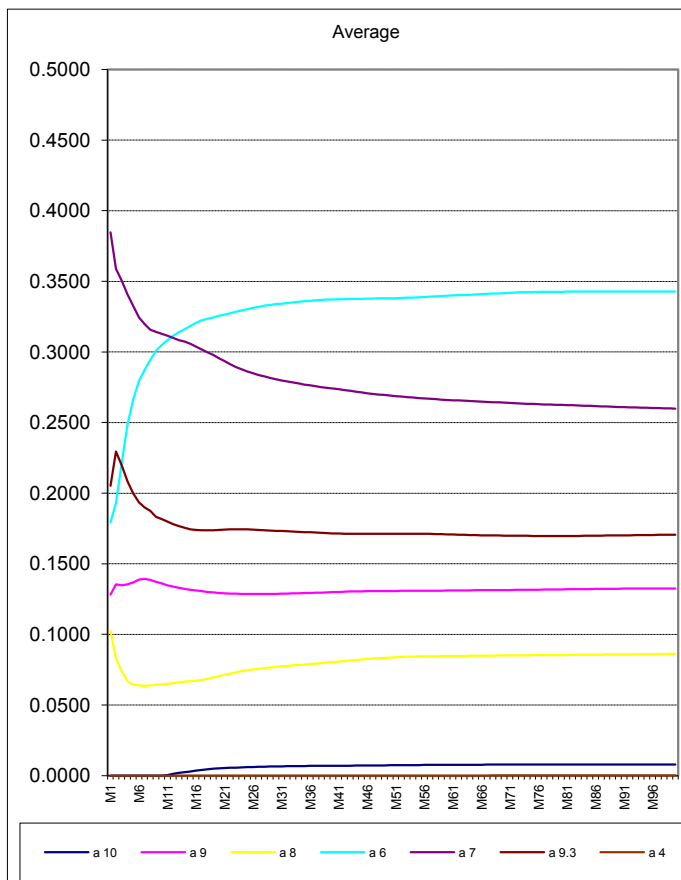
Graphic analysis of polymorphisms.

System TH01 Pop. Bogotá



Graphic analysis of polymorphisms.

System	TH01	Pop.	Bogotá
--------	------	------	--------



Graphic analysis of polymorphisms.

System	TH01	Pop.	Bogotá
--------	------	------	--------

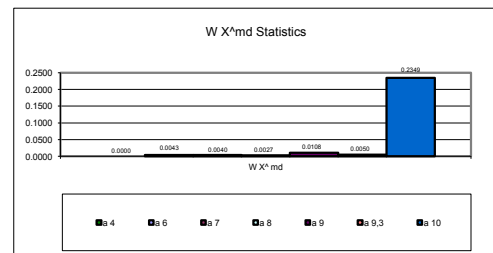
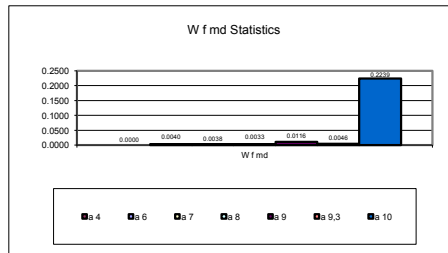
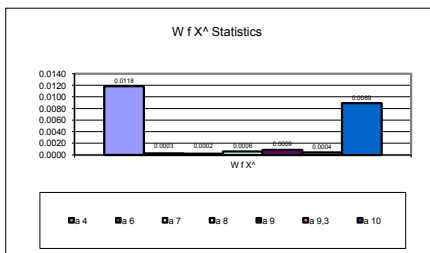
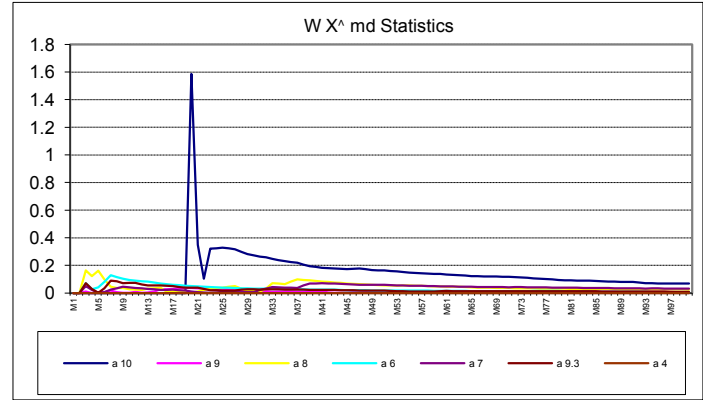
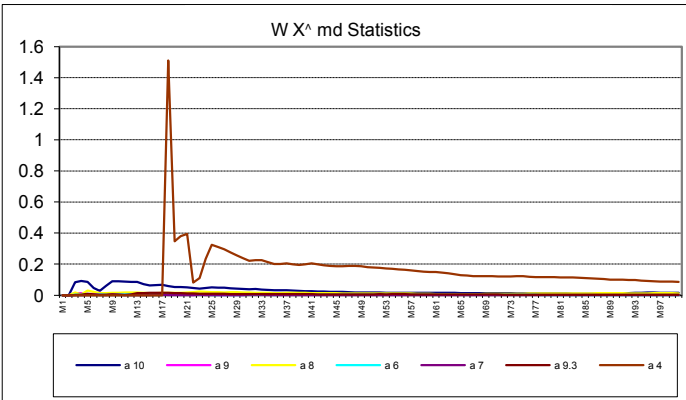
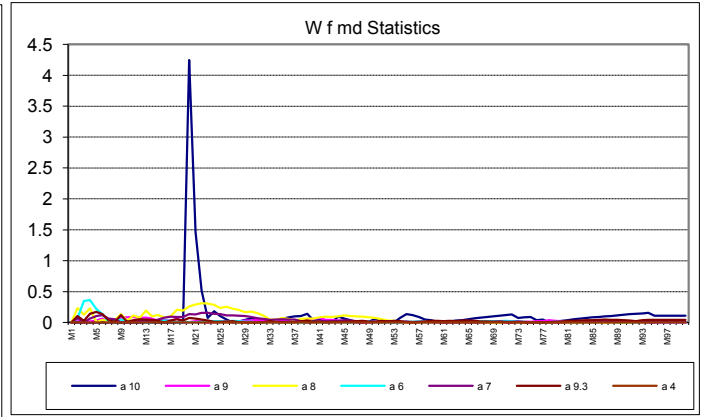
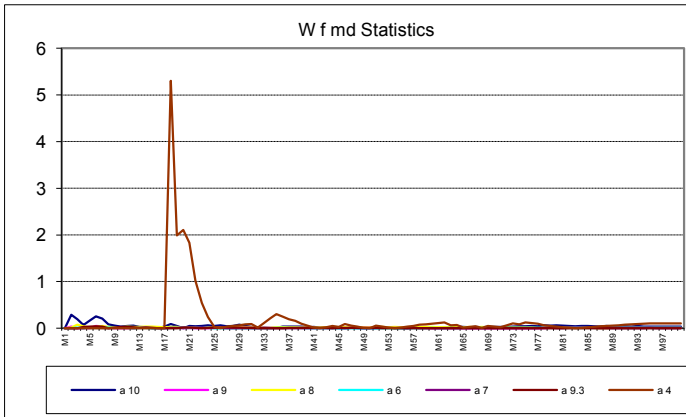
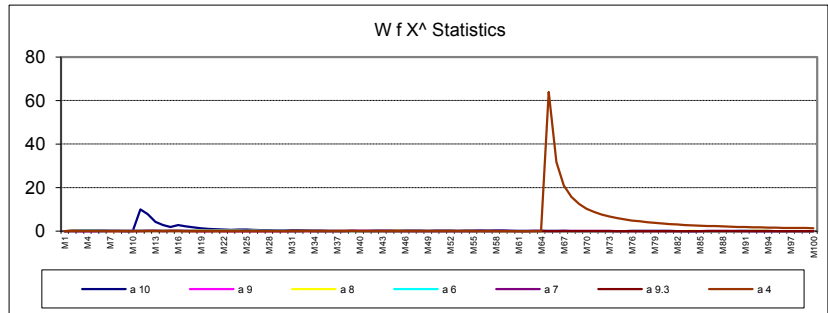
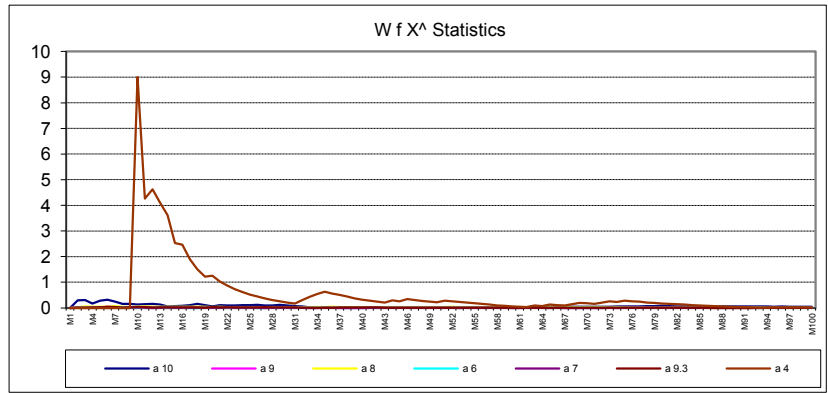
W Statistics to Specific sample

Alleles	W f X^	W f md	W X^ md
a 4	0.0118		
a 6	0.0003	0.0040	0.0043
a 7	0.0002	0.0038	0.0040
a 8	0.0006	0.0033	0.0027
a 9	0.0009	0.0116	0.0108
a 9,3	0.0004	0.0046	0.0050
a 10	0.0089	0.2239	0.2349

Indeterminate value (Division by zero)

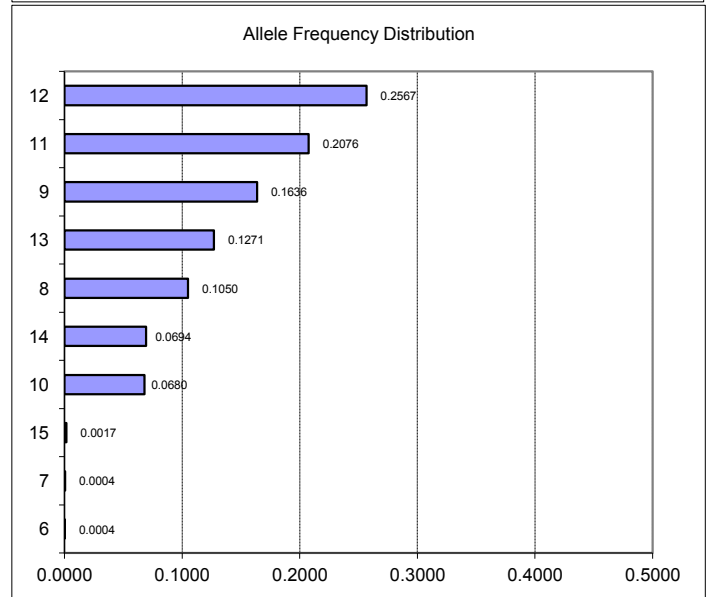
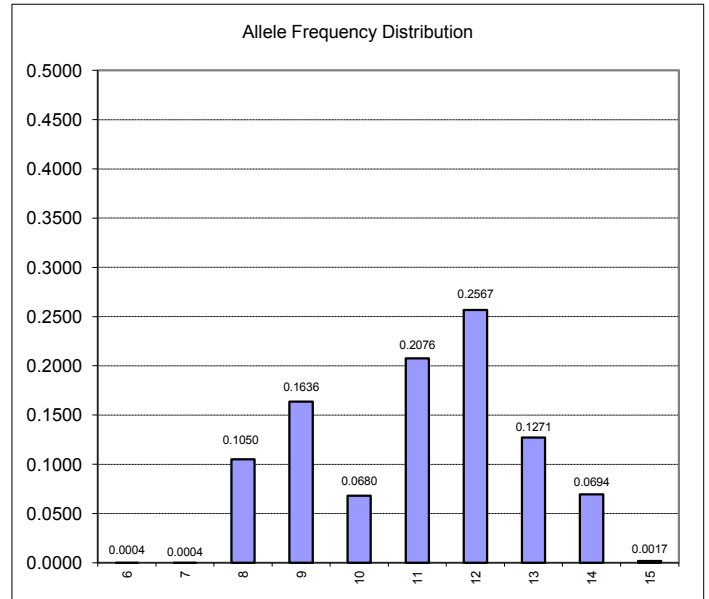
0.0118 0.2239 0.2349

Sampling without repetition

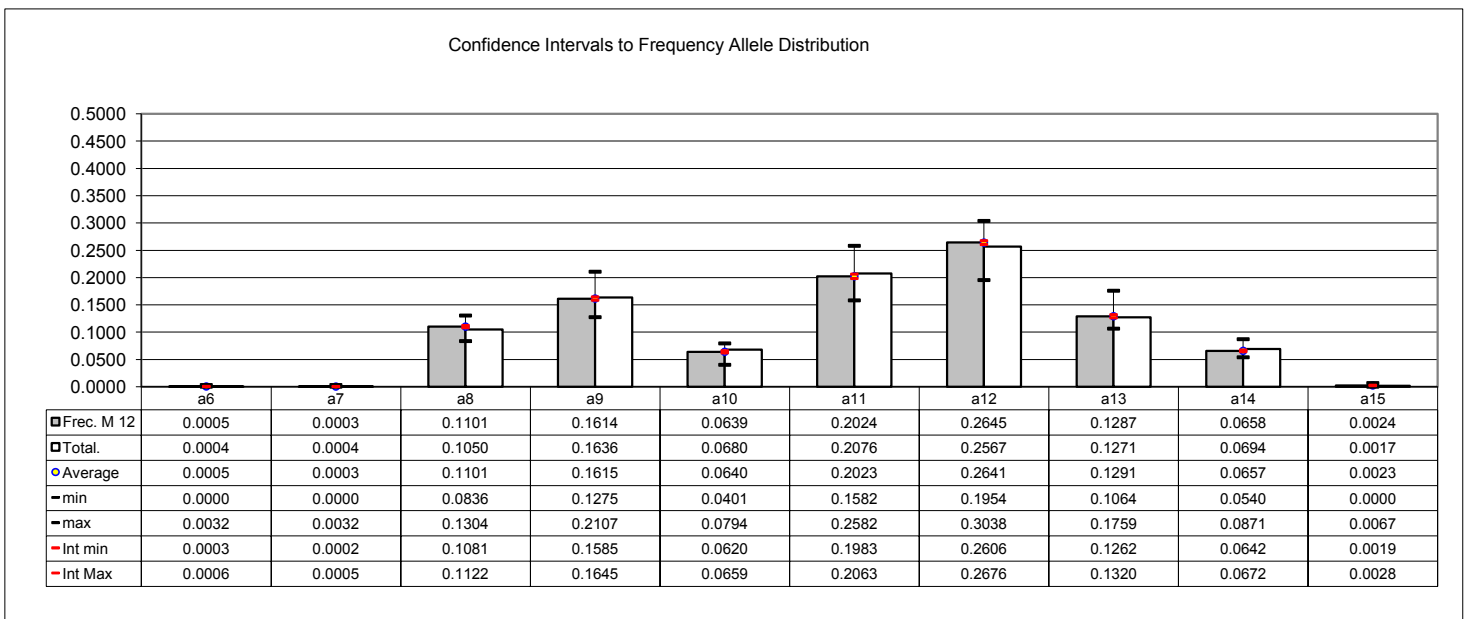


Population	Bogotá
System	D13S317
Alleles number	10
Effective alleles	6

Allele	M.Weight.	S.D.	Frequency
6	197.00		0.0004
7	201.00		0.0004
8	216.87	0.05	0.1050
9	220.83	0.05	0.1636
10	224.77	0.07	0.0680
11	228.88	0.07	0.2076
12	232.81	0.05	0.2567
13	236.68	0.07	0.1271
14	240.69	0.06	0.0694
14.2	Not reported		0.0020
15	244.68	0.09	0.0017



Single Alleles	
In the Database	In the Sample
6	6
7	7

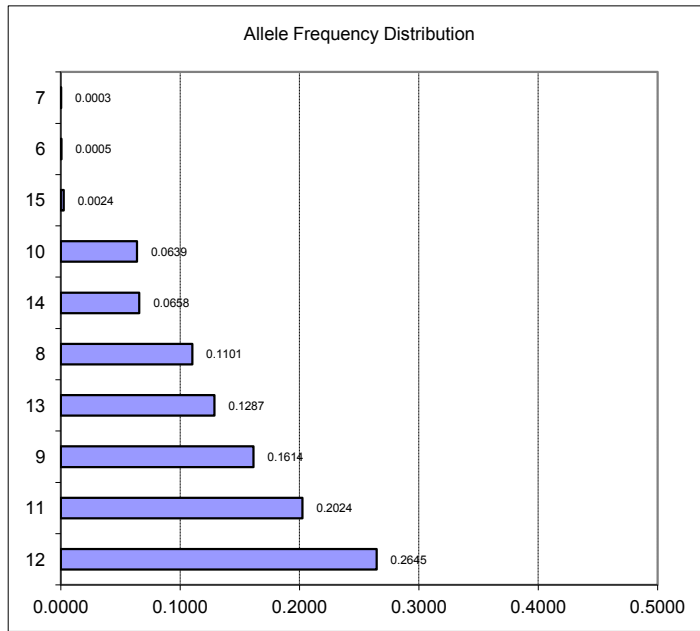


Specific sample analysis.

Sampling Specific.	
Population	Bogotá
System	D13S317
Alleles number	10
Effective alleles number	6
CV	0.05
Sample number	12
Sample size	152
Polimorphism alleles total	14

EXPERIMENT	(Todas)
Sample number	12

Allele frequency summary		
Alleles		Total
12		0.2645
11		0.2024
9		0.1614
13		0.1287
8		0.1101
14		0.0658
10		0.0639
15		0.0024
6		0.0005
7		0.0003



Single Alleles in the Sampling

EXPERIMENT	(Todas)
Sample number	12

Max	
Alleles	Total
6	1
7	1
15	2
10	25
14	29
8	42
13	54
9	63
11	81
12	103

Sampling and CV.

Sample selection
Analysis of CV.

Alleles	Sample No.
14	69
11	2
13	12
15	0
10	2
12	2
9	2
8	2
6	0
7	0

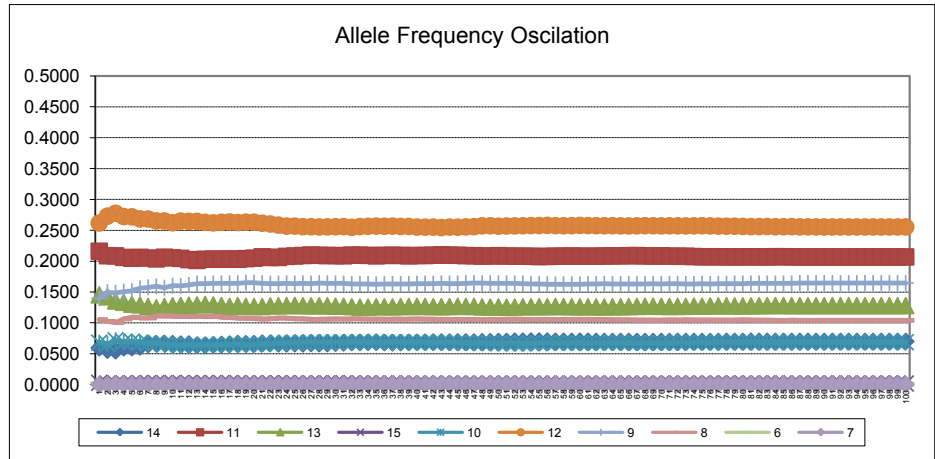
Allelic Distribution Order

Population	Bogotá
System	D13S317

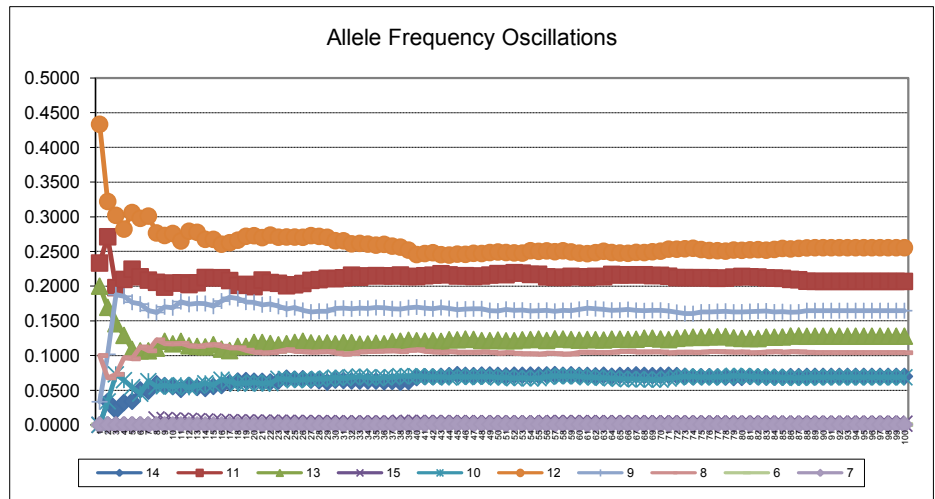
EXPERIMENT	(Todas)
Sample number	(Todas)

Allele frequency summary		
Alleles		Total
	12	0.2567
	11	0.2076
	9	0.1636
	13	0.1271
	8	0.1050
	14	0.0694
	10	0.0680
	15	0.0017
	7	0.0004
	6	0.0004

Compiled Allele Frequency



Allele frequency for a specific experiment.



Analysis for Specific Sampling

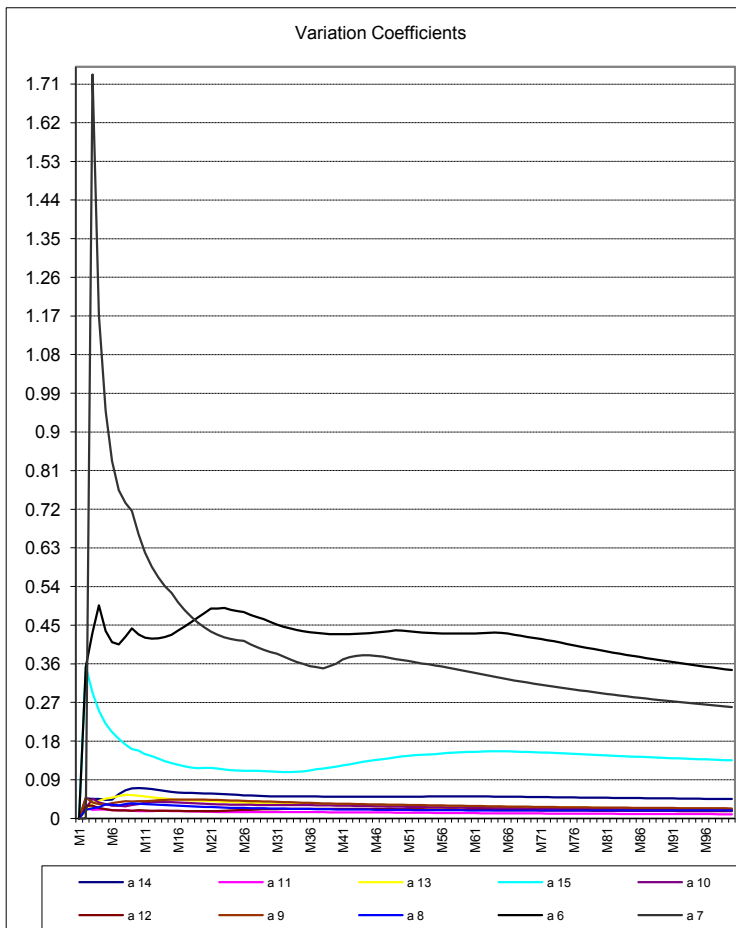
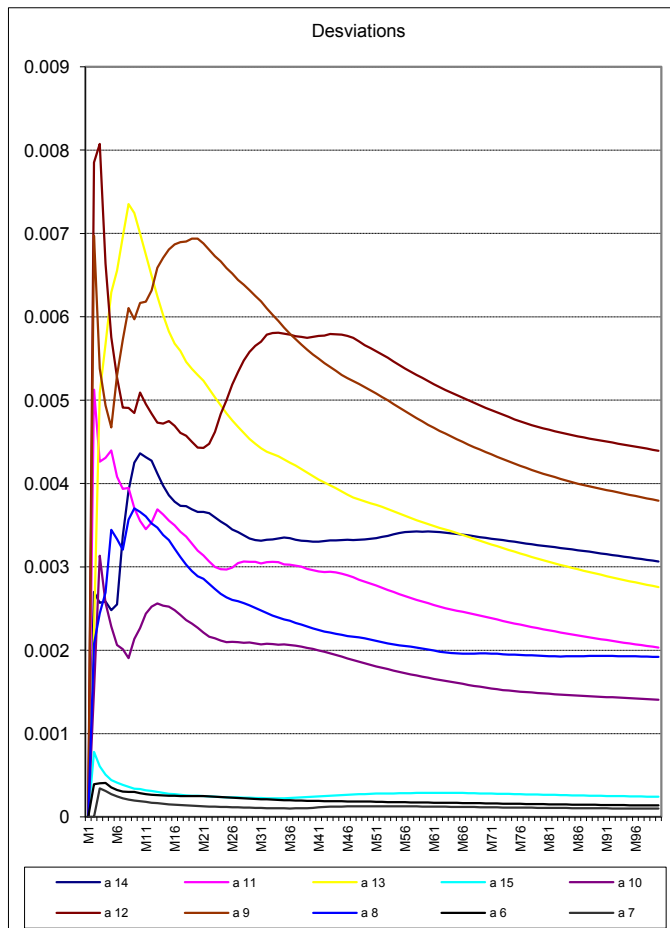
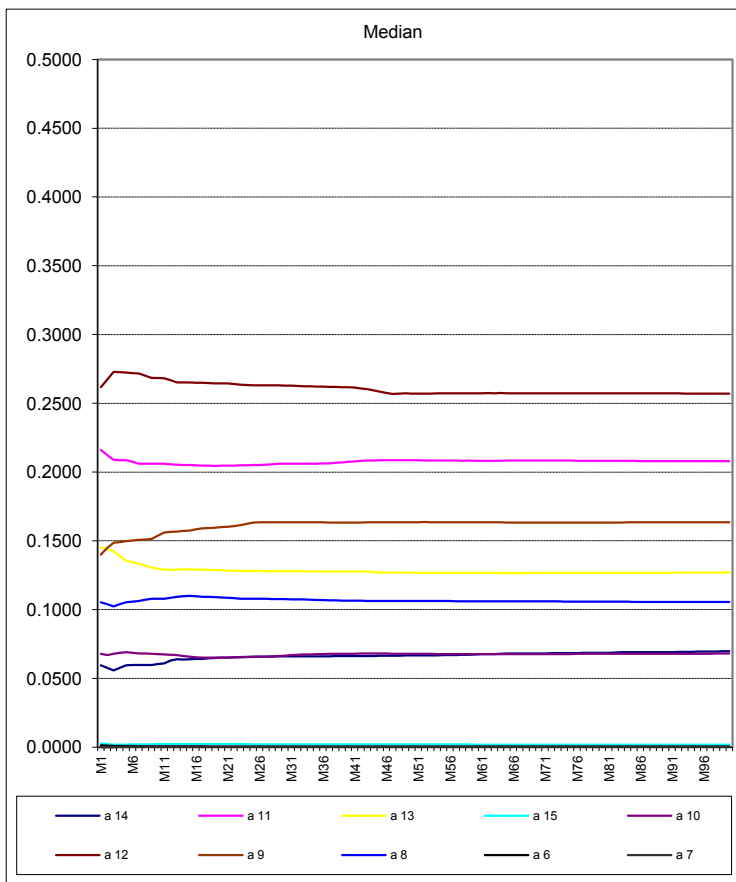
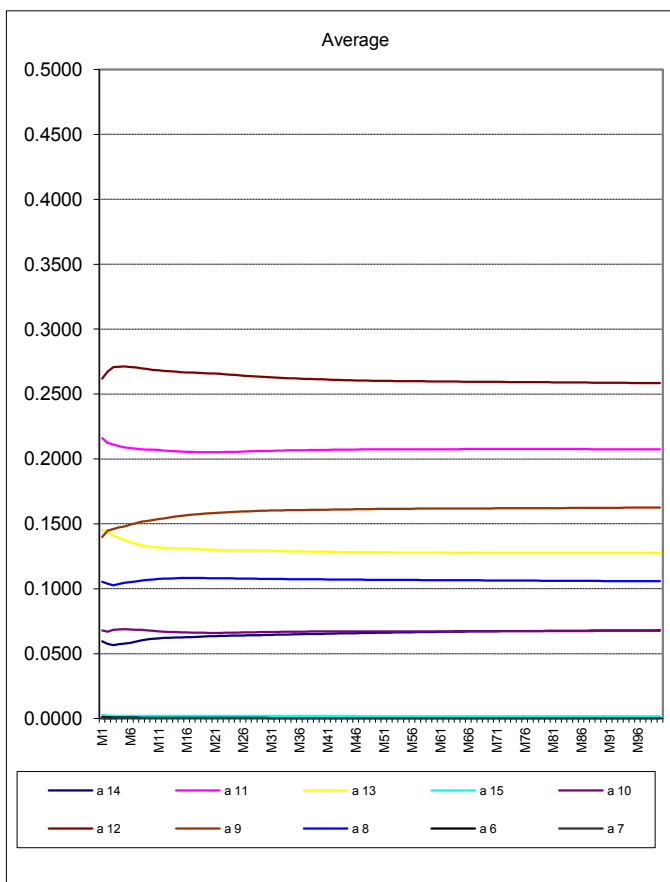
Sample number	12
Sample size	152

Average D	Min	Max
0.0020	0.0002	0.0040

Alleles	Average	min	max	Int min	Int Max	Median	S.D	IC
a6	0.0005	0.0000	0.0032	0.0003	0.0006	0.0000	0.0011	0.0002
a7	0.0003	0.0000	0.0032	0.0002	0.0005	0.0000	0.0010	0.0002
a8	0.1101	0.0836	0.1304	0.1081	0.1122	0.1122	0.0131	0.0021
a9	0.1615	0.1275	0.2107	0.1585	0.1645	0.1635	0.0188	0.0030
a10	0.0640	0.0401	0.0794	0.0620	0.0659	0.0668	0.0123	0.0020
a11	0.2023	0.1582	0.2582	0.1983	0.2063	0.2008	0.0252	0.0040
a12	0.2641	0.1954	0.3038	0.2606	0.2676	0.2642	0.0219	0.0035
a13	0.1291	0.1064	0.1759	0.1262	0.1320	0.1237	0.0184	0.0029
a14	0.0657	0.0540	0.0871	0.0642	0.0672	0.0655	0.0095	0.0015
a15	0.0023	0.0000	0.0067	0.0019	0.0028	0.0015	0.0026	0.0004

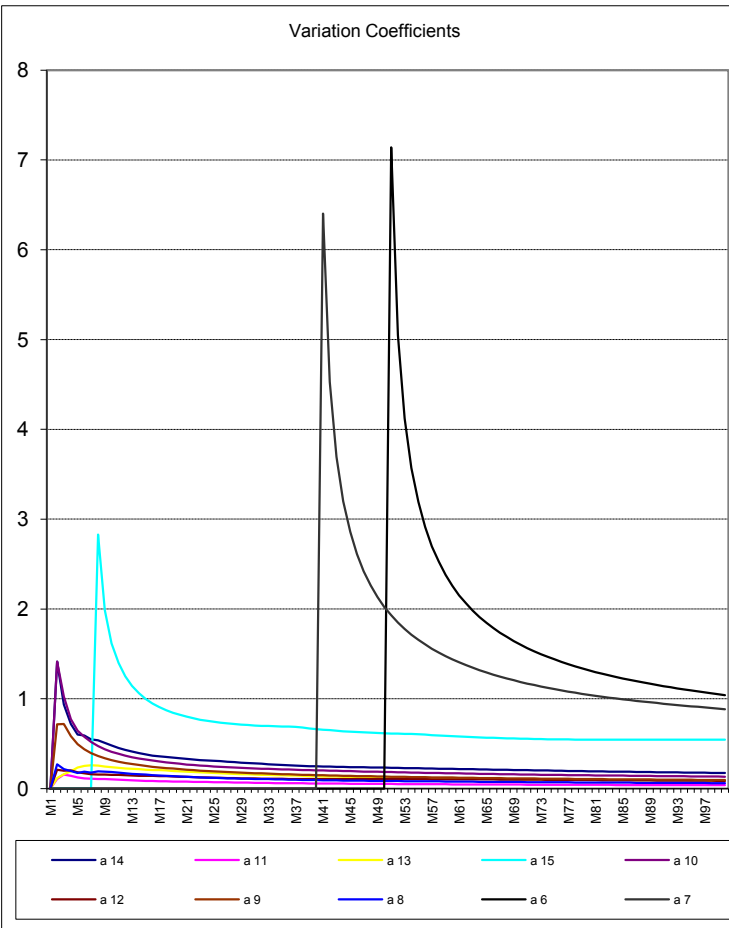
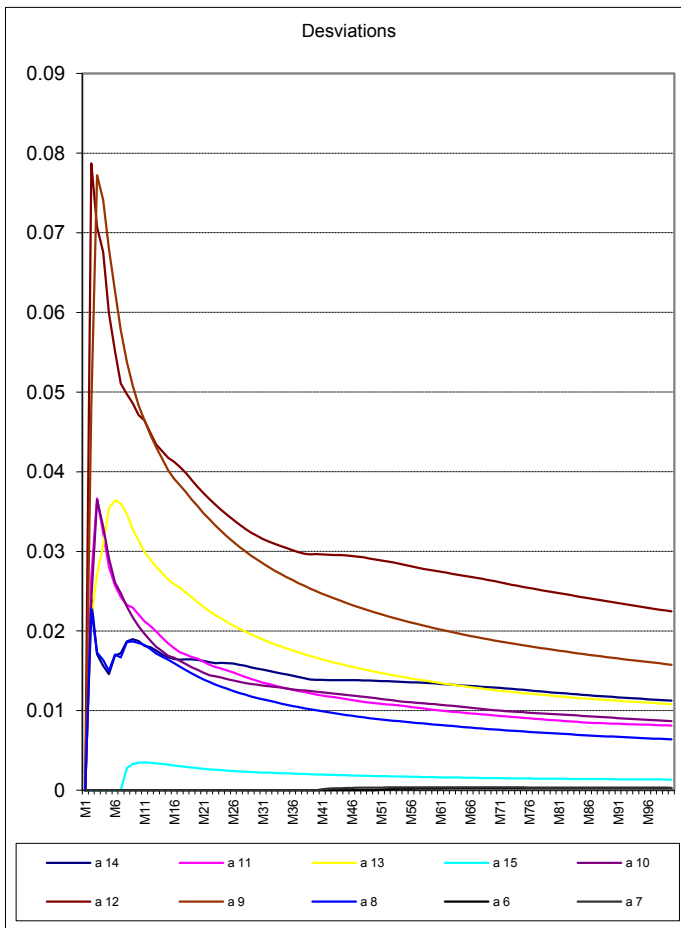
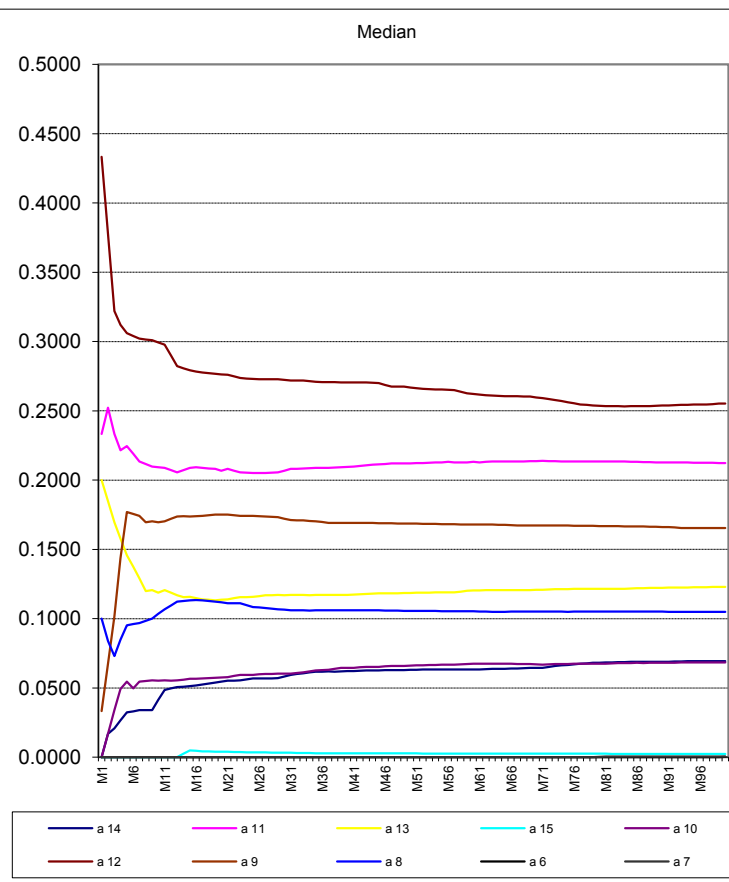
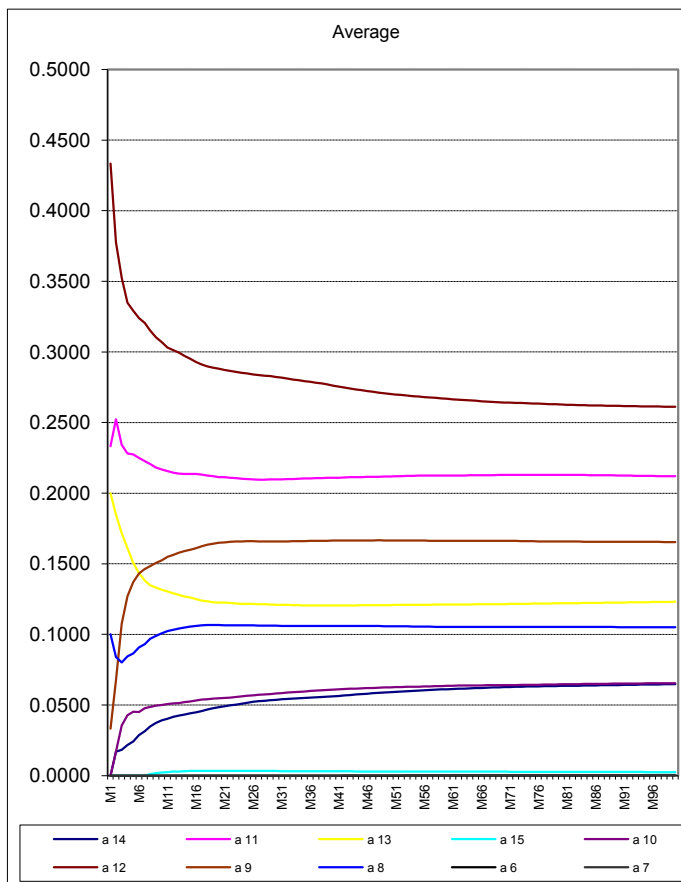
Graphic analysis of polymorphisms.

System D13S317 Pop. Bogotá



Graphic analysis of polymorphisms.

System D13S317 Pop. Bogotá



Graphic analysis of polymorphisms.

System	D13S317	Pop.	Bogotá
--------	---------	------	--------

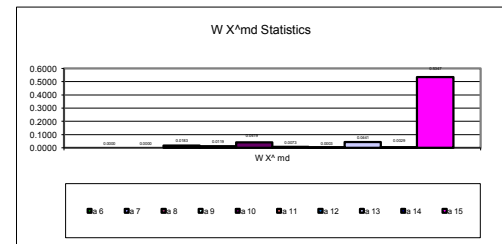
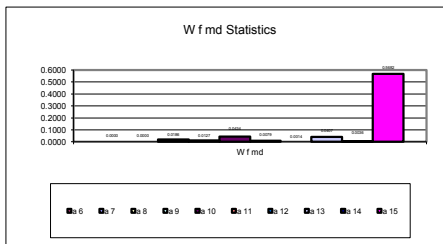
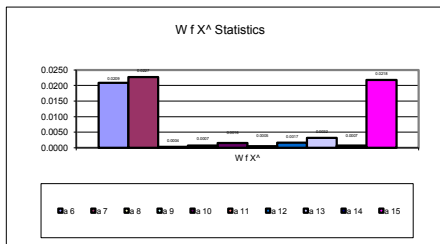
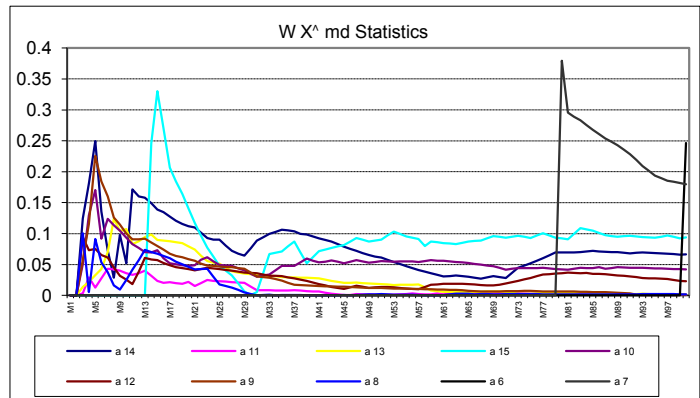
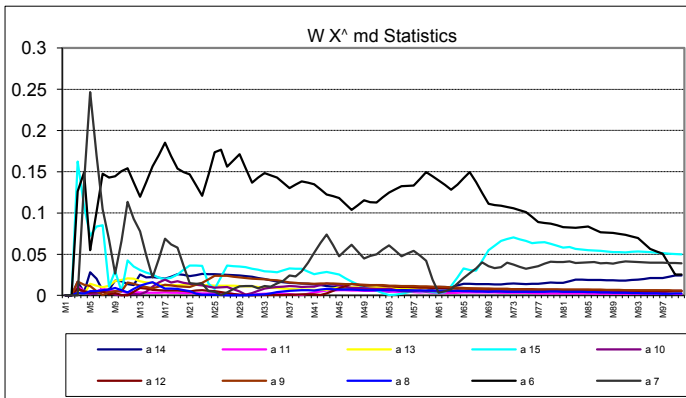
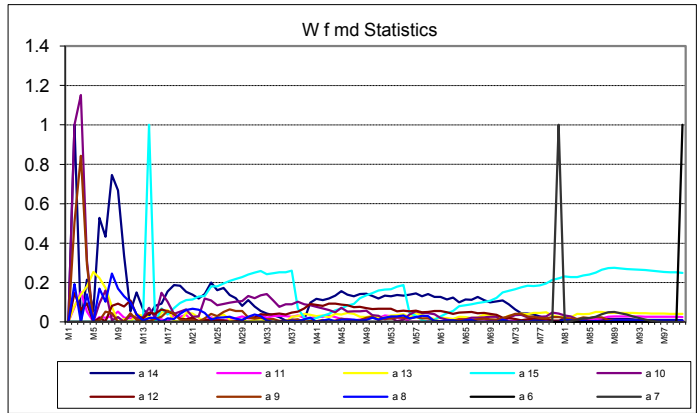
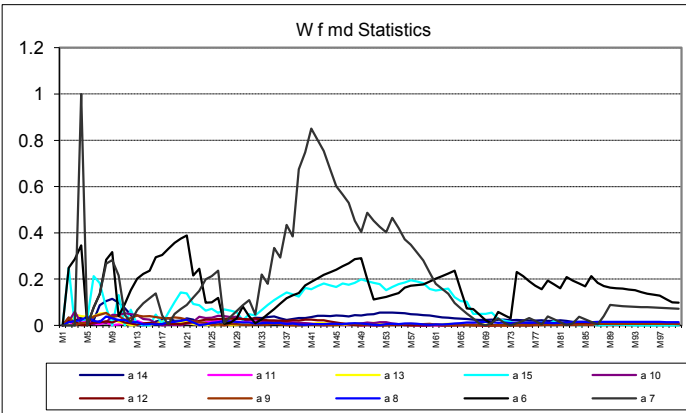
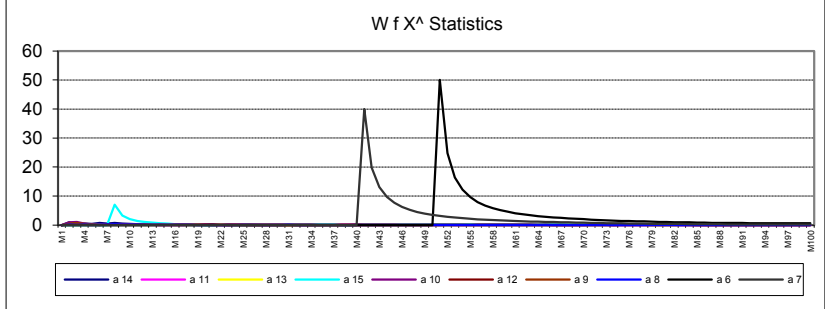
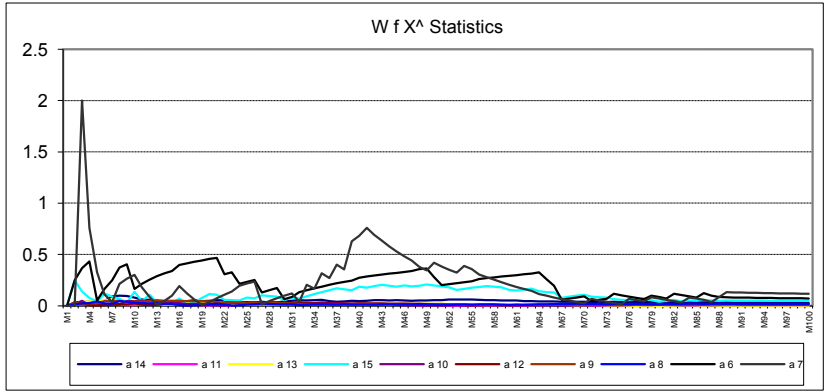
W Statistics to Specific sample

Alleles	W f X [^]	W f md	W X [^] md
a 6	0.0209		
a 7	0.0227		
a 8	0.0004	0.0186	0.0183
a 9	0.0007	0.0127	0.0119
a 10	0.0016	0.0434	0.0419
a 11	0.0005	0.0079	0.0073
a 12	0.0017	0.0014	0.0003
a 13	0.0032	0.0407	0.0441
a 14	0.0007	0.0036	0.0029
a 15	0.0218	0.5682	0.5347

Indeterminate value (Division by zero)

0.0227 0.5682 0.5347

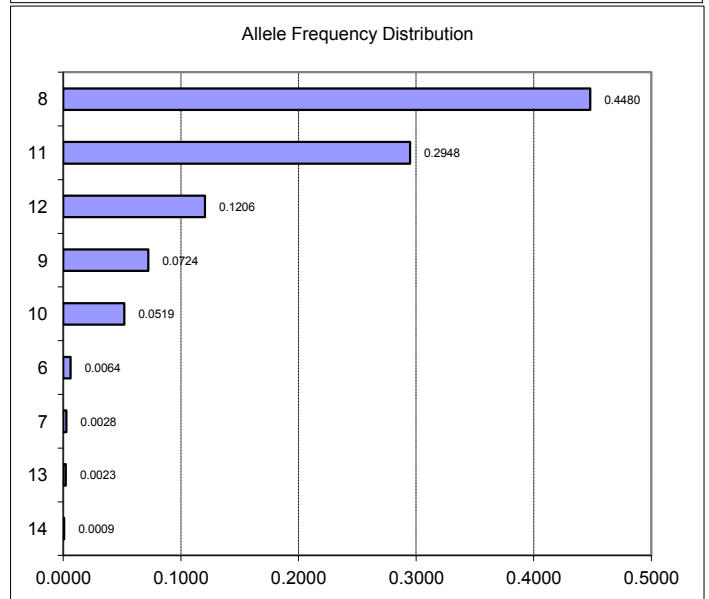
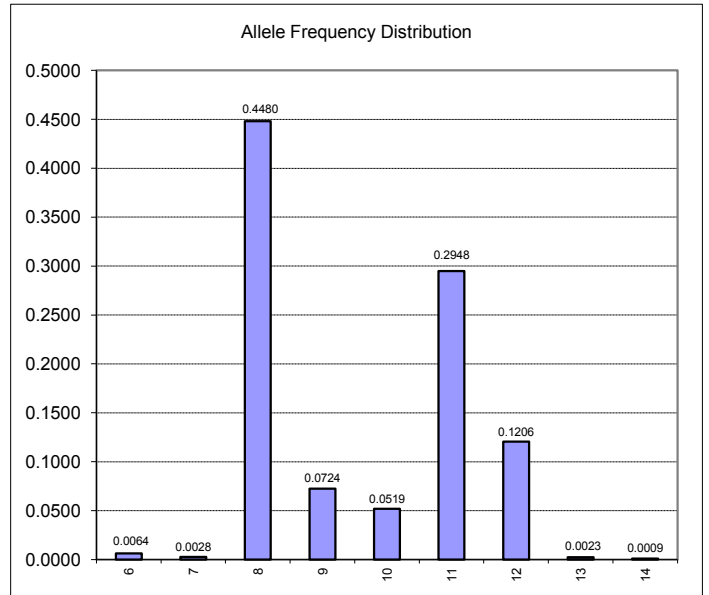
Muestras sin repeticiones



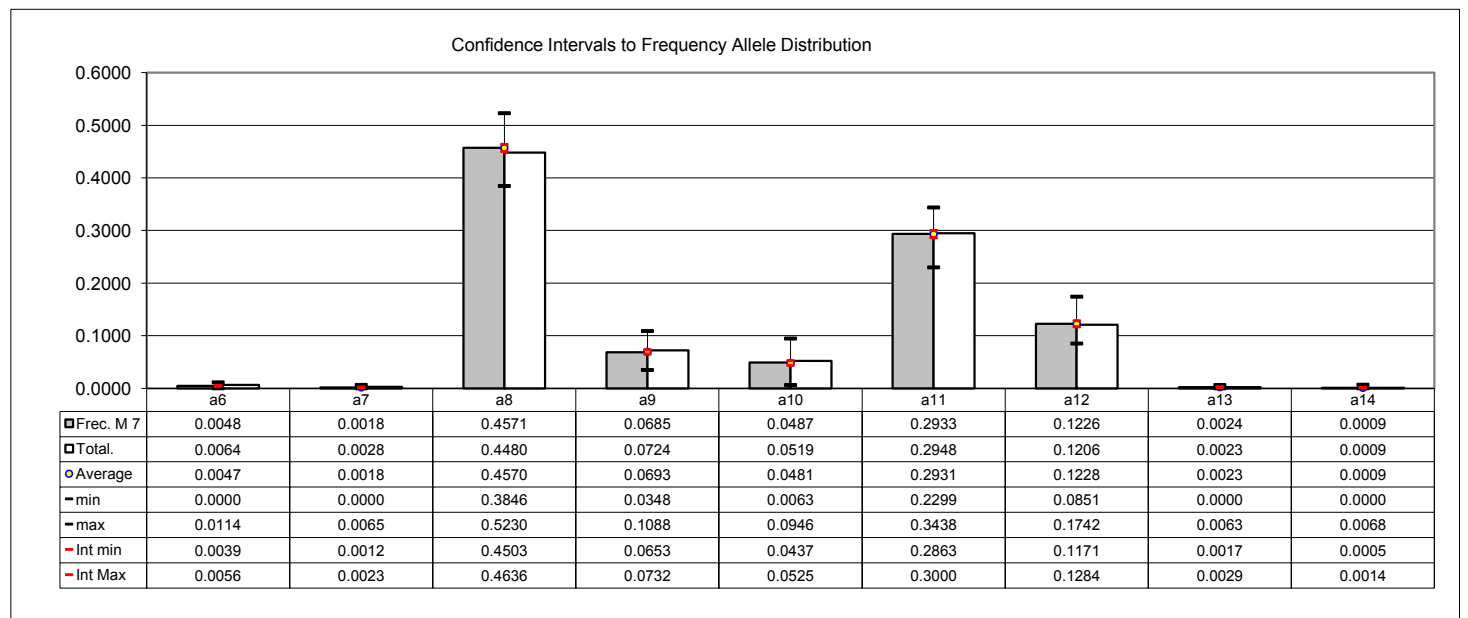
Population	Bogotá
System	TPOX
Alleles number	9
Effective alleles	3

Final frequency.

Allele	M.Weight.	S.D.	Frequency
6	222.07	0.04	0.0064
7	226.02	0.06	0.0028
8	229.91	0.03	0.4480
9	233.86	0.06	0.0724
10	237.88	0.07	0.0519
11	241.83	0.06	0.2948
12	245.77	0.07	0.1206
13	249.78	0.08	0.0023
14	253.00		0.0009



Single Alleles	
In the Database	In the Sample
	7
	13
	14

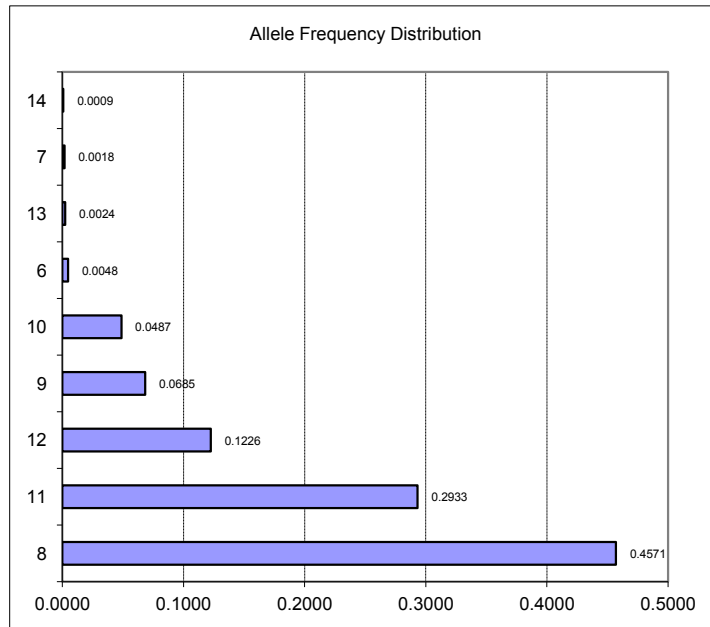


Specific sample analysis.

Sampling Specific.	
Population	Bogotá
System	TPOX
Alleles number	9
Effective alleles number	3
CV	0.05
Sample number	7
Sample size	86
Polimorphism alleles total	17

EXPERIMENT	(Todas)
Sample number	7

Allele frequency summary	
Alleles	Total
8	0.4571
11	0.2933
12	0.1226
9	0.0685
10	0.0487
6	0.0048
13	0.0024
7	0.0018
14	0.0009



Single Alleles in the Sampling

EXPERIMENT	(Todas)
Sample number	7

Max	
Alleles	Total
13	1
7	1
14	1
6	2
10	14
9	17
12	31
11	60
8	93

Sampling and CV.

Sample selection
Analysis of CV.

Alleles	Sample No.
14	0
11	2
13	0
10	2
12	2
9	7
8	2
6	0
7	0

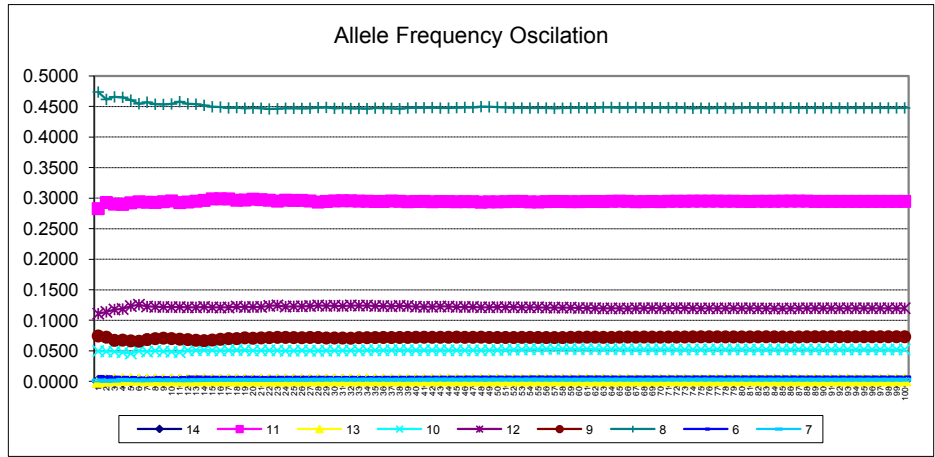
Allelic Distribution Order

Population	Bogotá
System	TPOX

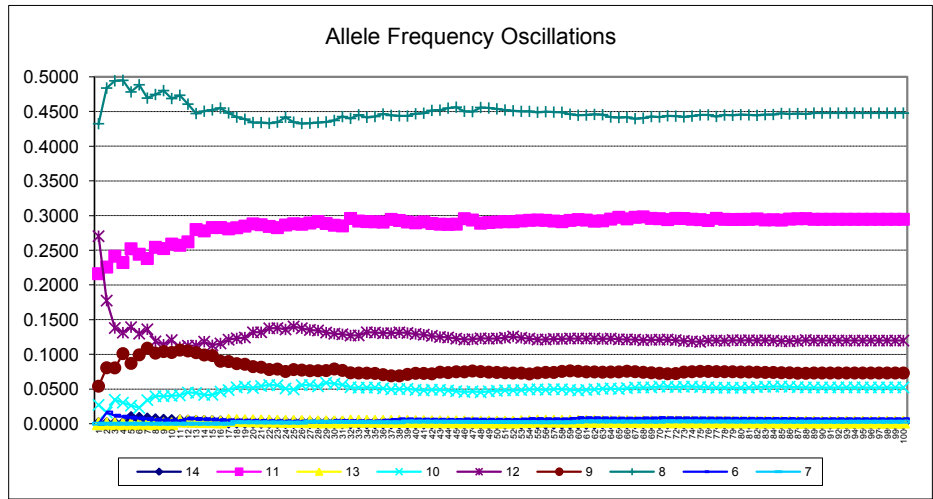
EXPERIMENT	(Todas)
Sample number	(Todas)

Allele frequency summary	
Alleles	Total
8	0.4480
11	0.2948
12	0.1206
9	0.0724
10	0.0519
6	0.0064
7	0.0028
13	0.0023
14	0.0009

Compiled Allele Frequency



Allele frequency for a specific experiment.



Analysis for Specific Sampling

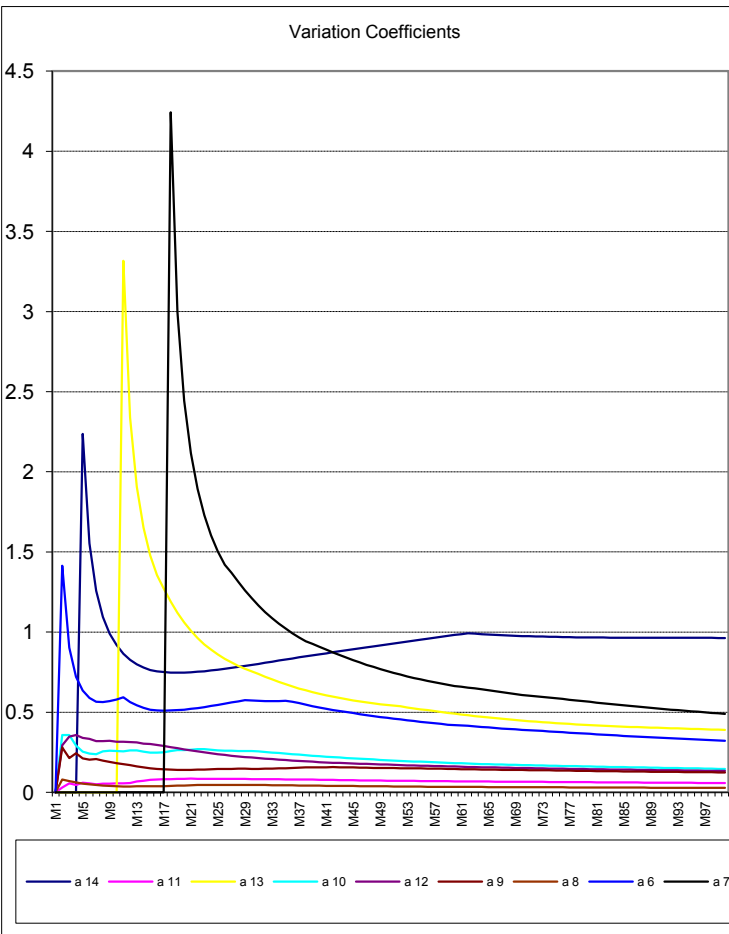
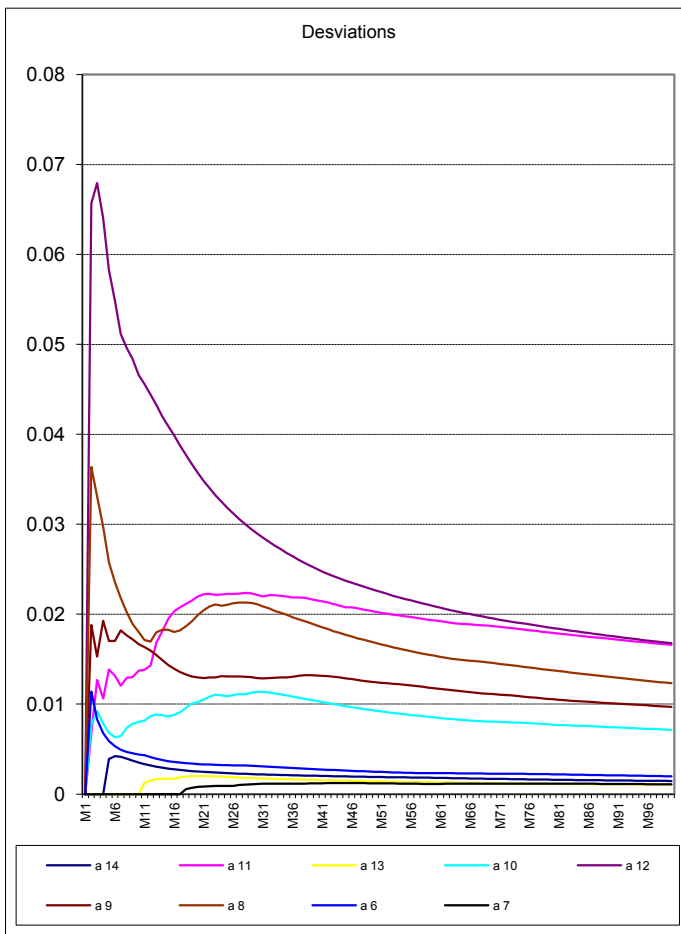
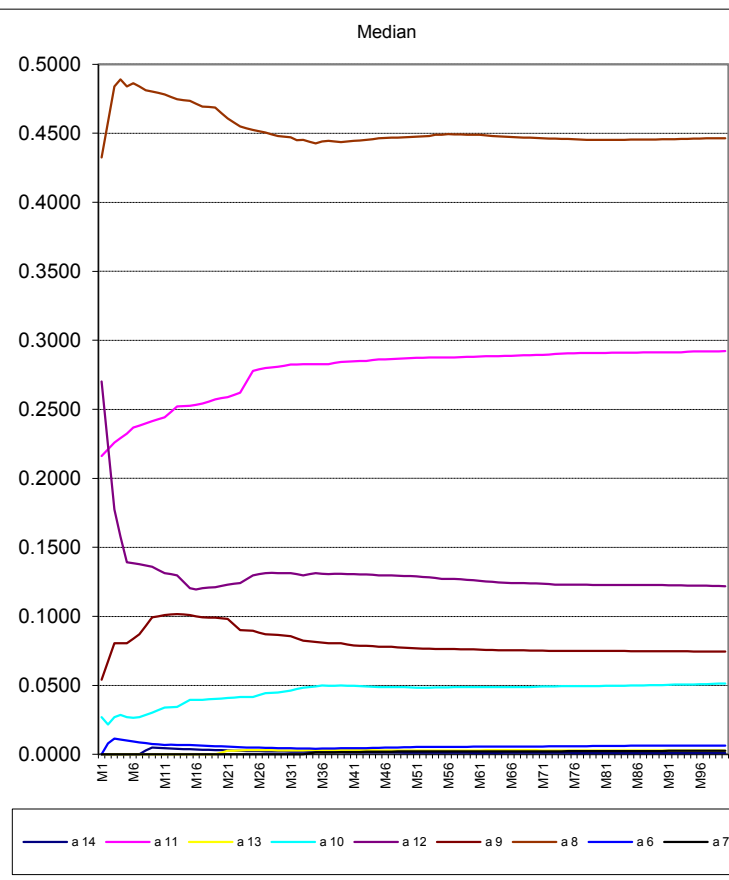
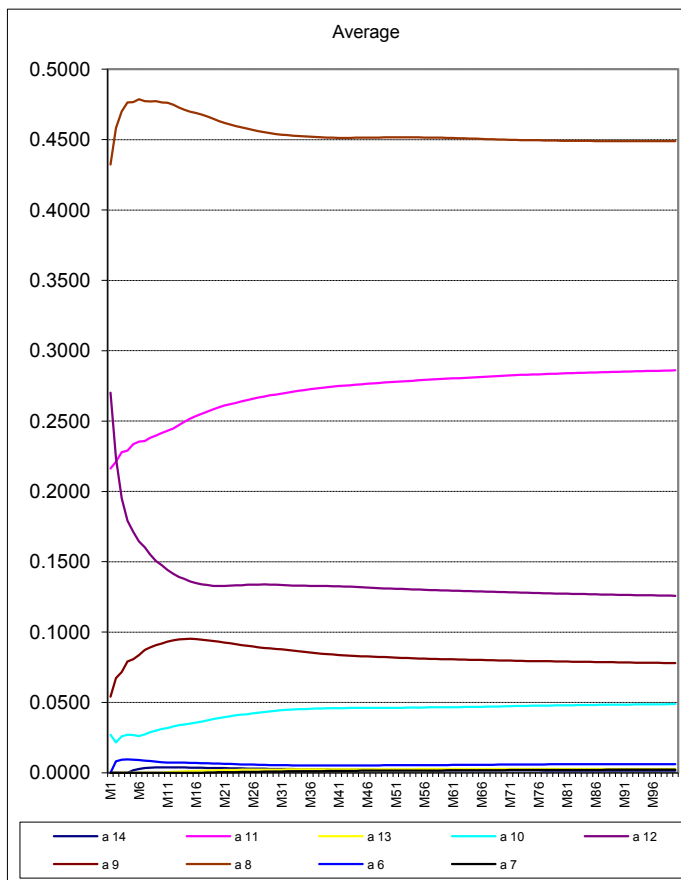
Sample number	7
Sample size	86

Average D	Min	Max
0.0033	0.0005	0.0068

Alleles	Average	min	max	Int min	Int Max	Median	S.D	IC
a6	0.0047	0.0000	0.0114	0.0039	0.0056	0.0060	0.0039	0.0008
a7	0.0018	0.0000	0.0065	0.0012	0.0023	0.0000	0.0028	0.0006
a8	0.4570	0.3846	0.5230	0.4503	0.4636	0.4542	0.0314	0.0066
a9	0.0693	0.0348	0.1088	0.0653	0.0732	0.0674	0.0188	0.0040
a10	0.0481	0.0063	0.0946	0.0437	0.0525	0.0483	0.0209	0.0044
a11	0.2931	0.2299	0.3438	0.2863	0.3000	0.2901	0.0322	0.0068
a12	0.1228	0.0851	0.1742	0.1171	0.1284	0.1176	0.0268	0.0057
a13	0.0023	0.0000	0.0063	0.0017	0.0029	0.0000	0.0029	0.0006
a14	0.0009	0.0000	0.0068	0.0005	0.0014	0.0000	0.0023	0.0005

Graphic analysis of polymorphisms.

System TPOX Pop. Bogotá



Graphic analysis of polymorphisms.

System	TPOX	Pop.	Bogotá
--------	------	------	--------

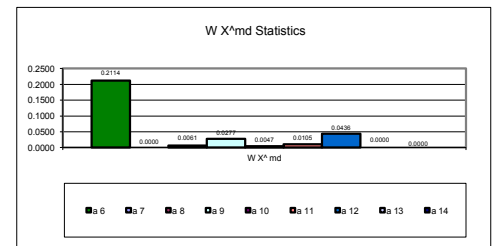
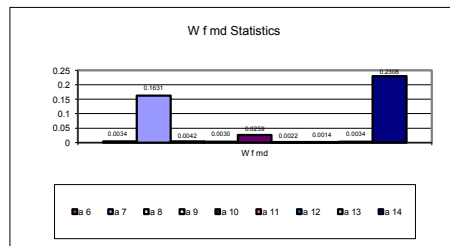
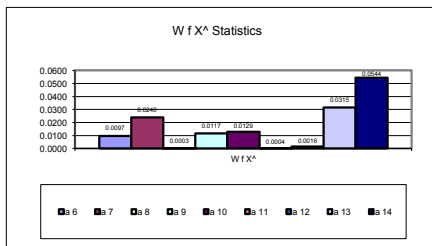
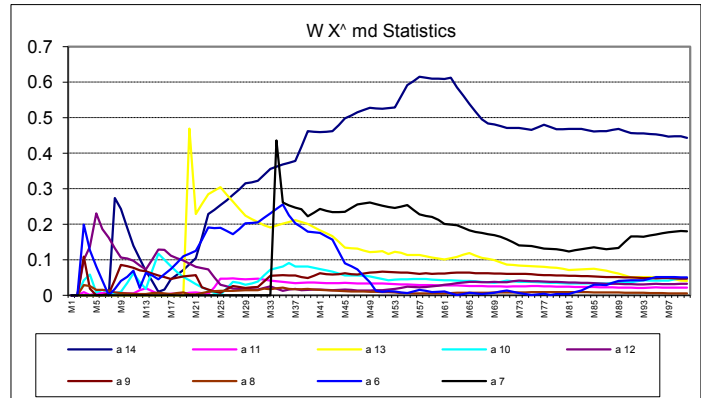
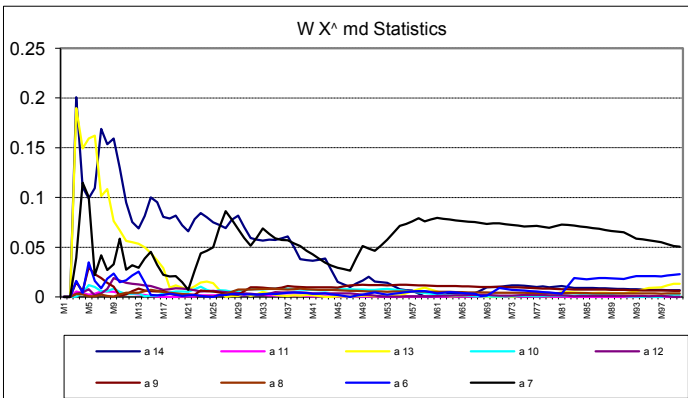
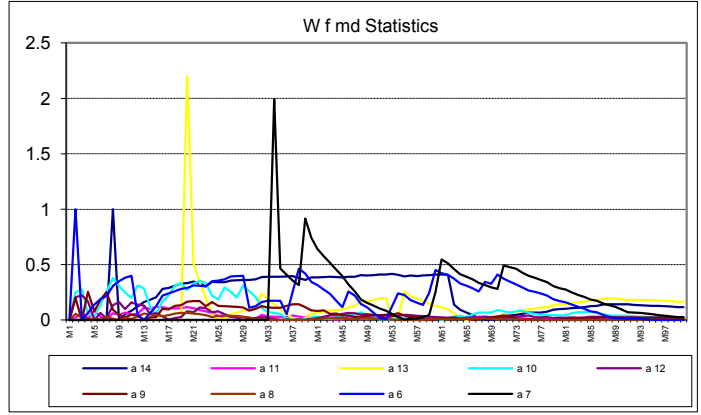
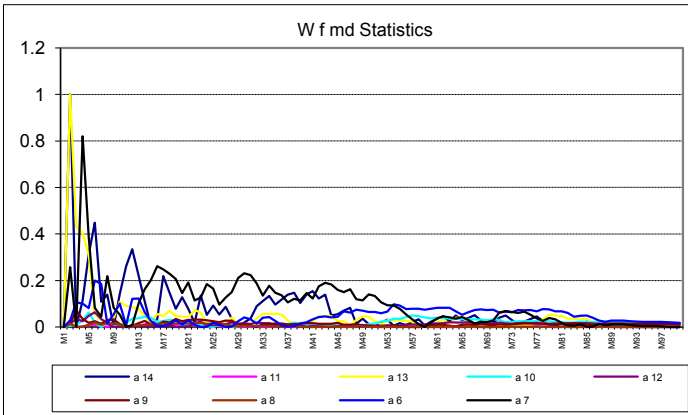
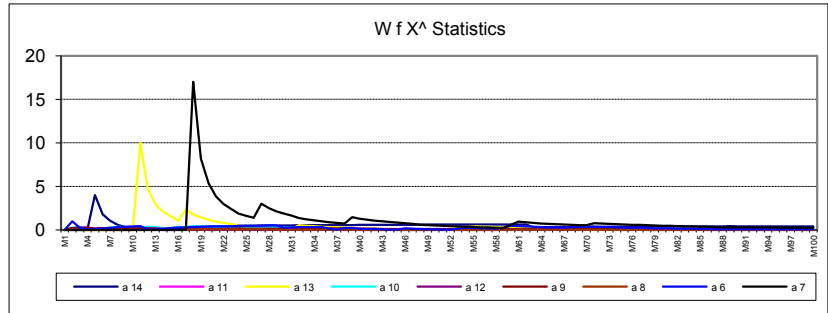
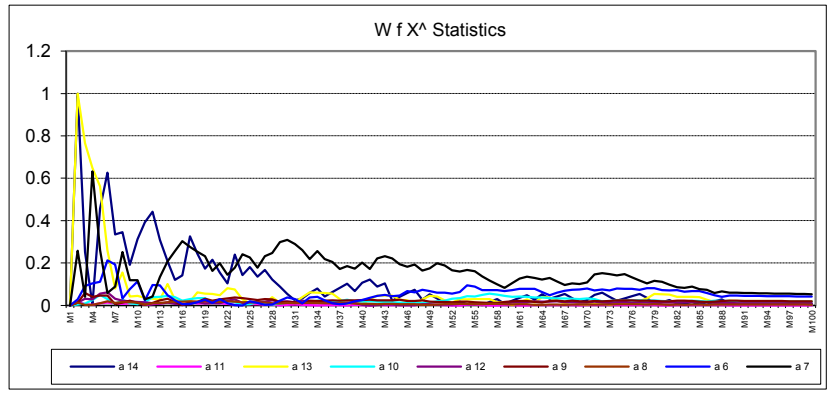
W Statistics to Specific sample

Alleles	W f X [^]	W f md	W X [^] md
a 6	0.0097	0.2038	0.2114
a 7	0.0240		
a 8	0.0003	0.0064	0.0061
a 9	0.0117	0.0158	0.0277
a 10	0.0129	0.0082	0.0047
a 11	0.0004	0.0109	0.0105
a 12	0.0016	0.0420	0.0436
a 13	0.0315		
a 14	0.0544		

Indeterminate value (Division by zero)

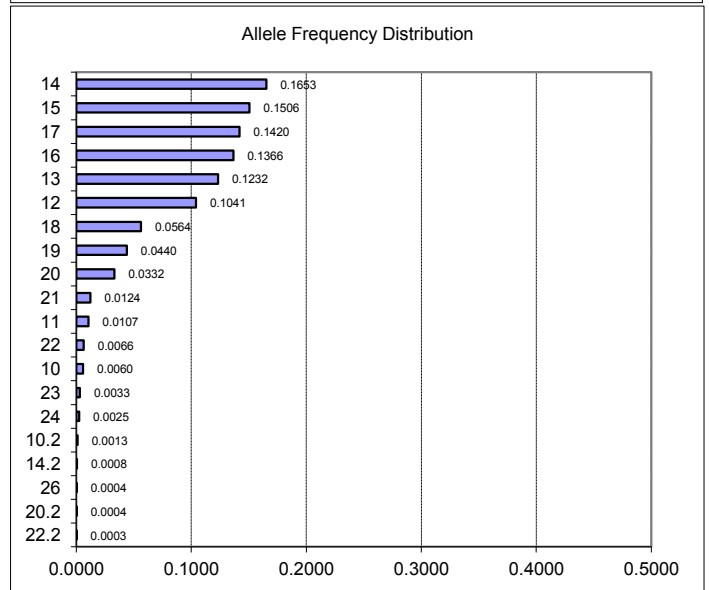
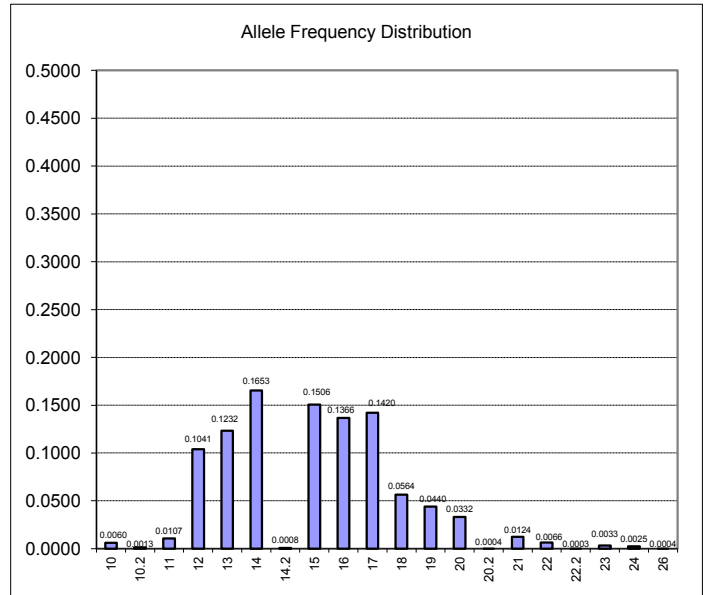
0.0544 0.2038 0.2114

Sampling without repetition

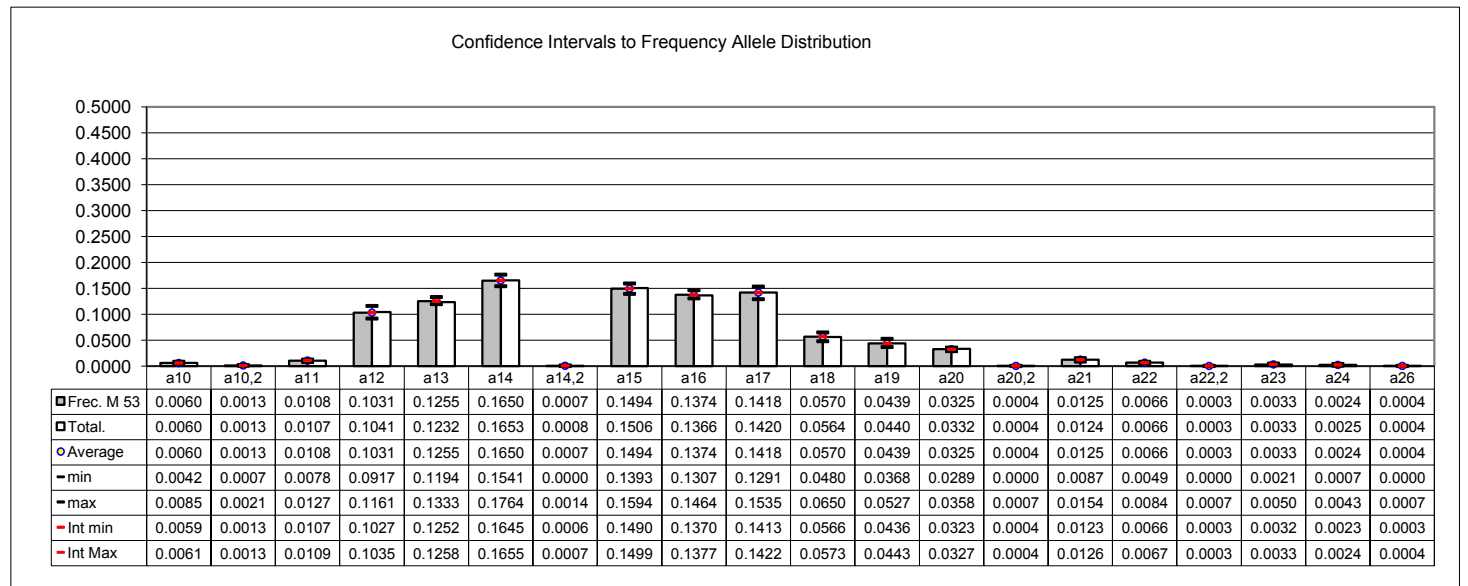


Population	Bogotá
System	D18S51
Alleles number	20
Effective alleles	8

Allele	M.Weight.	S.D.	Frequency
7	262.07	0.08	
9	270.22	0.06	
10	274.34	0.09	0.0060
10.2	276.36	0.06	0.0013
11	278.41	0.08	0.0107
12	282.49	0.05	0.1041
13	286.57	0.06	0.1232
13.2	288.63	0.05	0.0040
14	290.77	0.04	0.1653
14.2	292.78	0.05	0.0008
15	294.91	0.07	0.1506
16	299.07	0.06	0.1366
17	303.50	0.07	0.1420
18	307.94	0.09	0.0564
19	312.40	0.11	0.0440
20	316.71	0.09	0.0332
20.2	318.00		0.0004
21	320.99	0.14	0.0124
22	325.24	0.11	0.0066
22.2	327.00		0.0003
23	329.40	0.11	0.0033
24	333.54	0.15	0.0025
26	341.56	0.09	0.0004
27	345.24	0.08	



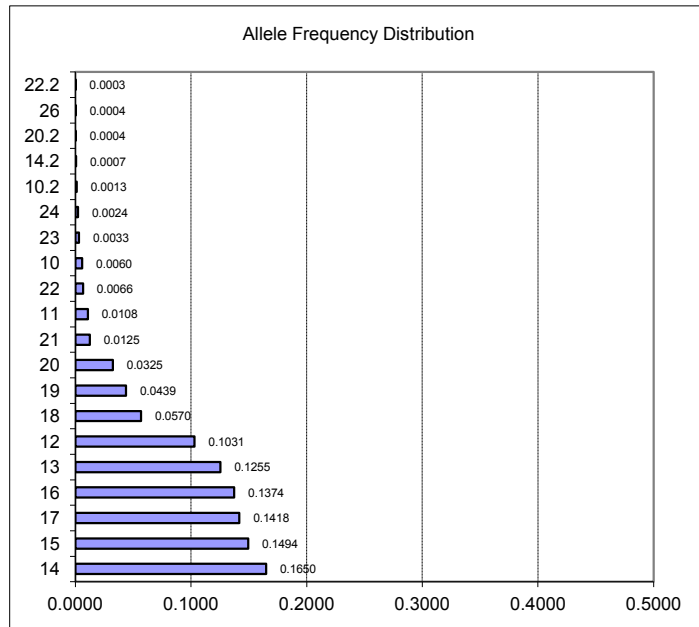
Single Alleles	
In the Database	In the Sample
20.2	20.2
22.2	22.2
26	26



Sampling Specific.	
Population	Bogotá
System	D18S51
Alleles number	20
Effective alleles number	8
CV	0.05
Sample number	53
Sample size	709
Polimorphism alleles total	42

EXPERIMENT	(Todas)
Sample number	53

Allele frequency summary	
Alleles	Total
14	0.1650
15	0.1494
17	0.1418
16	0.1374
13	0.1255
12	0.1031
18	0.0570
19	0.0439
20	0.0325
21	0.0125
11	0.0108
22	0.0066
10	0.0060
23	0.0033
24	0.0024
10.2	0.0013
14.2	0.0007
20.2	0.0004
26	0.0004
22.2	0.0003



Single Alleles in the Sampling

EXPERIMENT	(Todas)
Sample number	53

Max	Total
Alleles	Total
26	1
20.2	1
22.2	1
14.2	2
10.2	3
24	6
23	7
10	12
22	12
11	18
21	22
20	51
19	75
18	91
12	165
13	189
16	210
17	221
15	224
14	248

Sampling and CV.

Sample selection
Analysis of CV.

Alleles	Sample No.
26	0
14	2
11	0
13	5
15	2
10	0
12	2
17	7
16	2
19	2
18	53
20	0
22	0
21	0
22.2	0
24	0
23	0
20.2	0
10.2	0
14.2	0

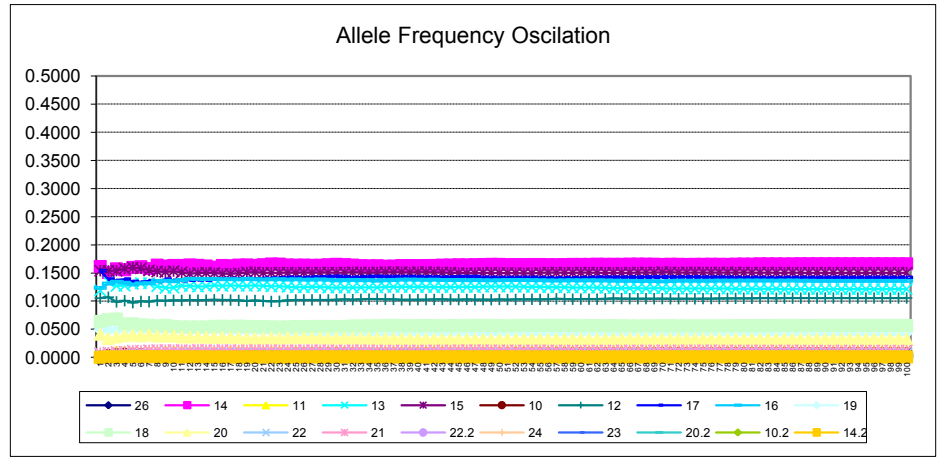
Allelic Distribution Order

Population	Bogotá
System	D18S51

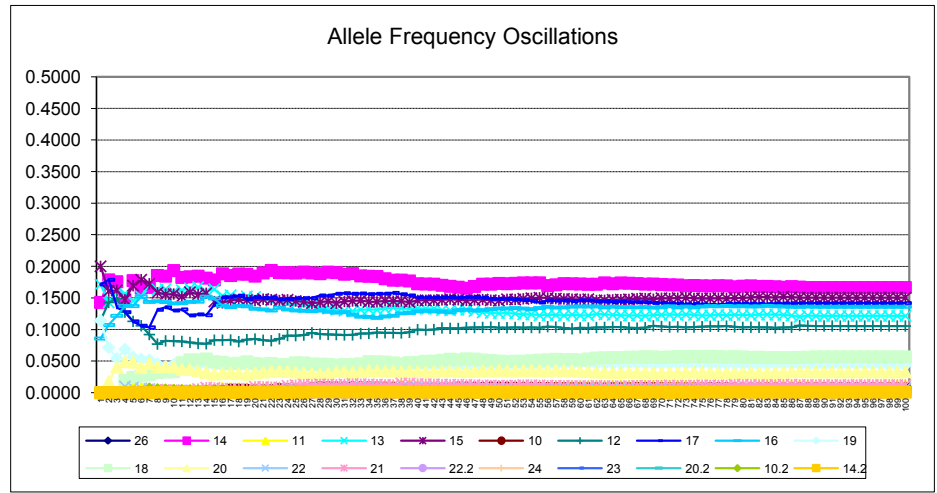
EXPERIMENT	(Todas)
Sample number	(Todas)

Allele frequency summary		
Alleles		Total
	14	0.1653
	15	0.1506
	17	0.1420
	16	0.1366
	13	0.1232
	12	0.1041
	18	0.0564
	19	0.0440
	20	0.0332
	21	0.0124
	11	0.0107
	22	0.0066
	10	0.0060
	23	0.0033
	24	0.0025
	10.2	0.0013
	14.2	0.0008
	26	0.0004
	20.2	0.0004
	22.2	0.0003

Compiled Allele Frequency



Allele frequency for a specific experiment.



Analysis for Specific Sampling

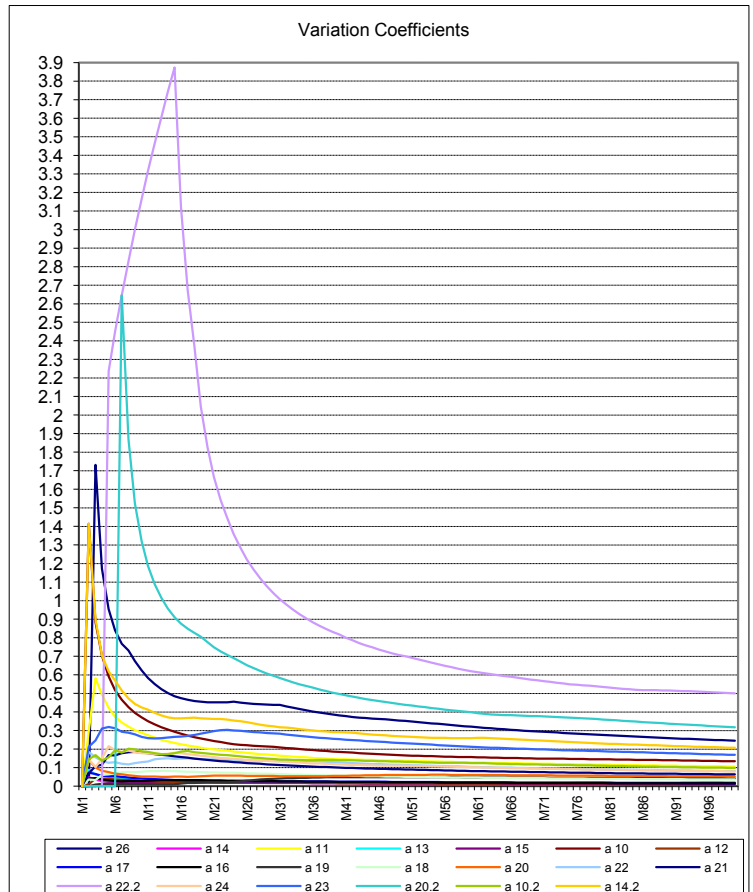
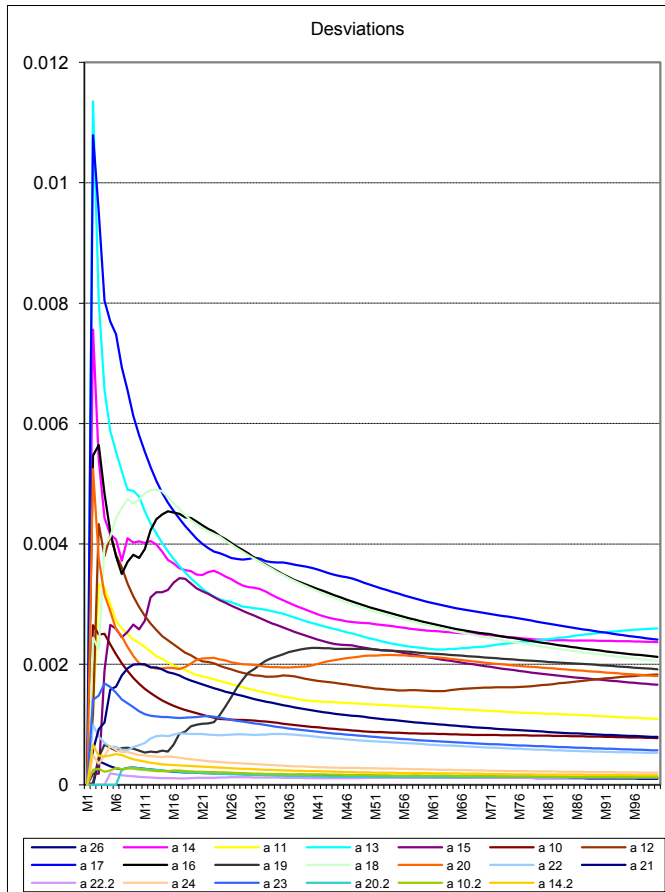
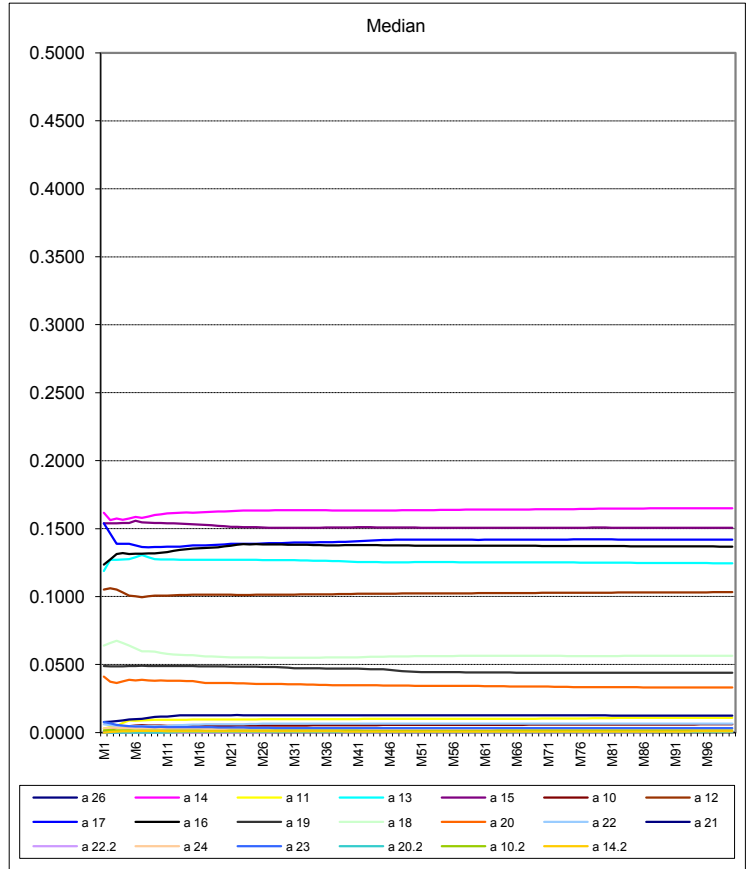
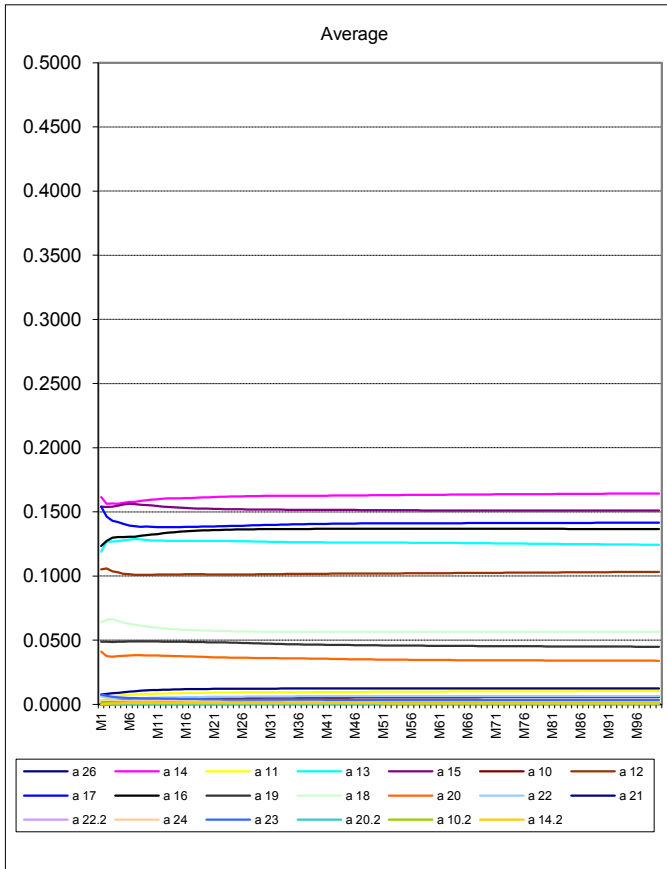
Sample number	53
Sample size	709

Average D	Min	Max
0.0002	0.0000	0.0005

Alleles	Average	min	max	Int min	Int Max	Median	S.D	IC
a10	0.0060	0.0042	0.0085	0.0059	0.0061	0.0057	0.0013	0.0001
a10,2	0.0013	0.0007	0.0021	0.0013	0.0013	0.0014	0.0004	0.0000
a11	0.0108	0.0078	0.0127	0.0107	0.0109	0.0111	0.0015	0.0001
a12	0.1031	0.0917	0.1161	0.1027	0.1035	0.1030	0.0052	0.0004
a13	0.1255	0.1194	0.1333	0.1252	0.1258	0.1244	0.0043	0.0003
a14	0.1650	0.1541	0.1764	0.1645	0.1655	0.1632	0.0072	0.0005
a14,2	0.0007	0.0000	0.0014	0.0006	0.0007	0.0007	0.0005	0.0000
a15	0.1494	0.1393	0.1594	0.1490	0.1499	0.1502	0.0059	0.0004
a16	0.1374	0.1307	0.1464	0.1370	0.1377	0.1360	0.0049	0.0004
a17	0.1418	0.1291	0.1535	0.1413	0.1422	0.1413	0.0059	0.0004
a18	0.0570	0.0480	0.0650	0.0566	0.0573	0.0572	0.0045	0.0003
a19	0.0439	0.0368	0.0527	0.0436	0.0443	0.0434	0.0047	0.0003
a20	0.0325	0.0289	0.0358	0.0323	0.0327	0.0327	0.0025	0.0002
a20,2	0.0004	0.0000	0.0007	0.0004	0.0004	0.0007	0.0004	0.0000
a21	0.0125	0.0087	0.0154	0.0123	0.0126	0.0121	0.0018	0.0001
a22	0.0066	0.0049	0.0084	0.0066	0.0067	0.0069	0.0009	0.0001
a22,2	0.0003	0.0000	0.0007	0.0003	0.0003	0.0000	0.0004	0.0000
a23	0.0033	0.0021	0.0050	0.0032	0.0033	0.0035	0.0009	0.0001
a24	0.0024	0.0007	0.0043	0.0023	0.0024	0.0021	0.0010	0.0001
a26	0.0004	0.0000	0.0007	0.0003	0.0004	0.0007	0.0004	0.0000

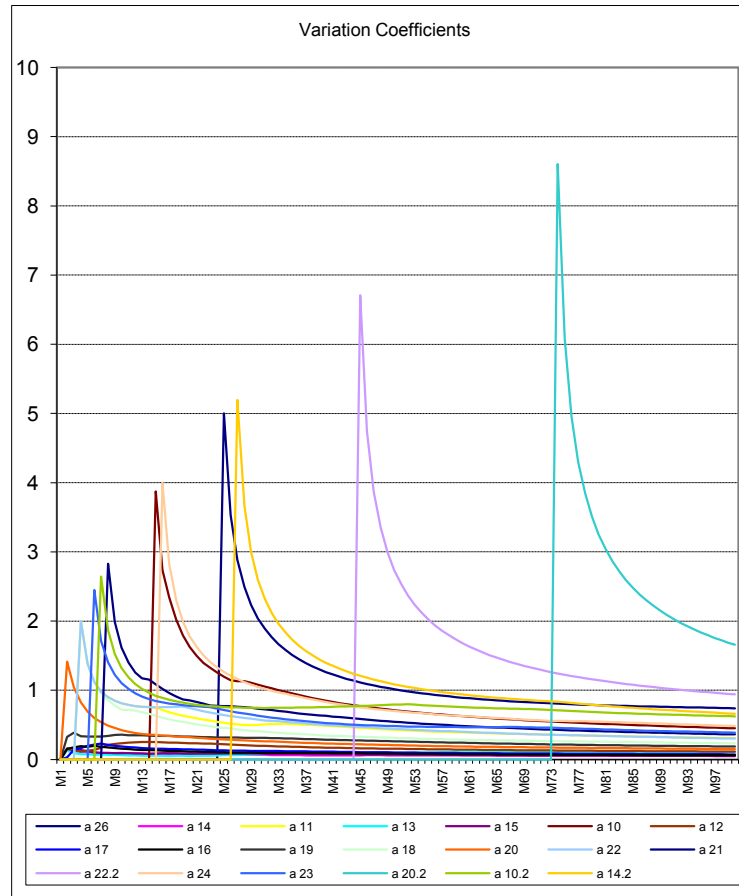
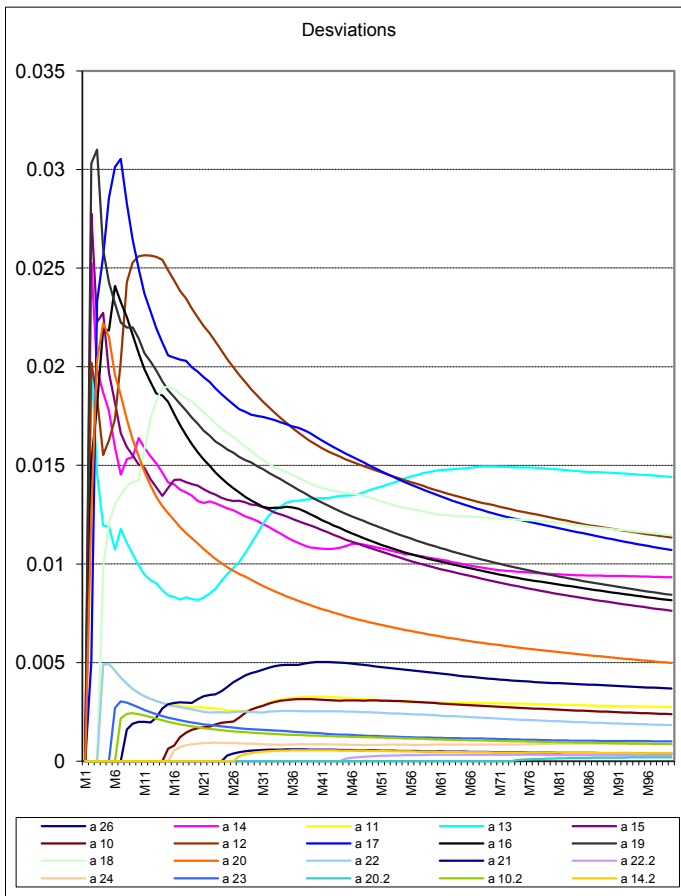
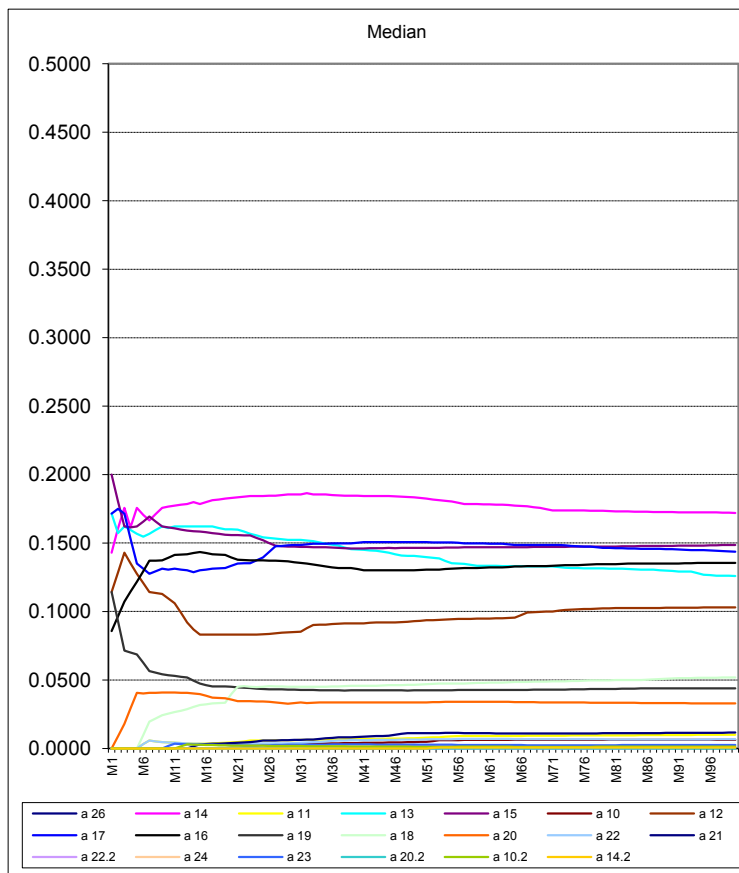
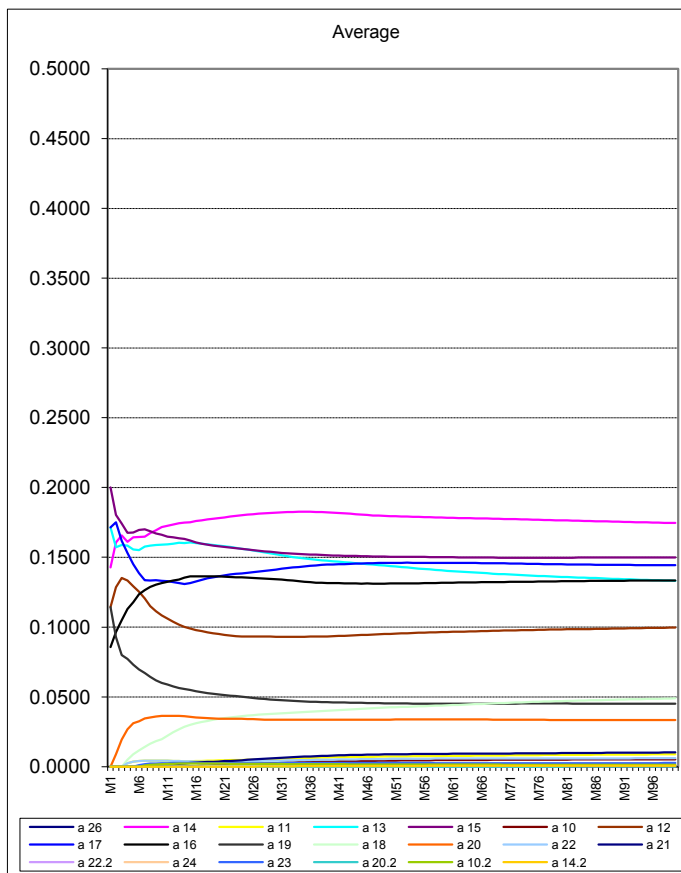
Graphic analysis of polymorphisms.

System D18S51 Pop. Bogotá



Graphic analysis of polymorphisms.

System D18S51 Pop. Bogotá



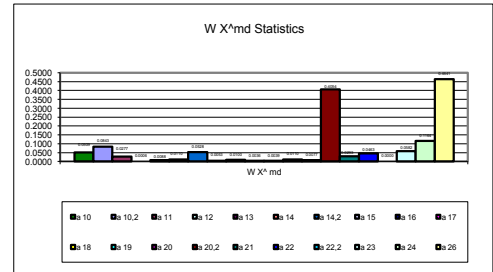
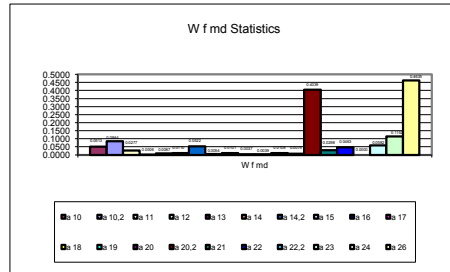
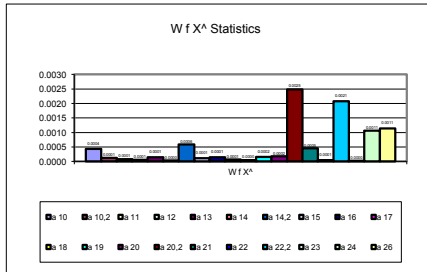
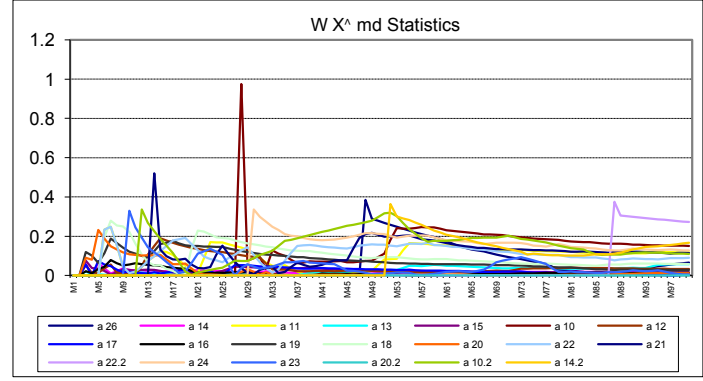
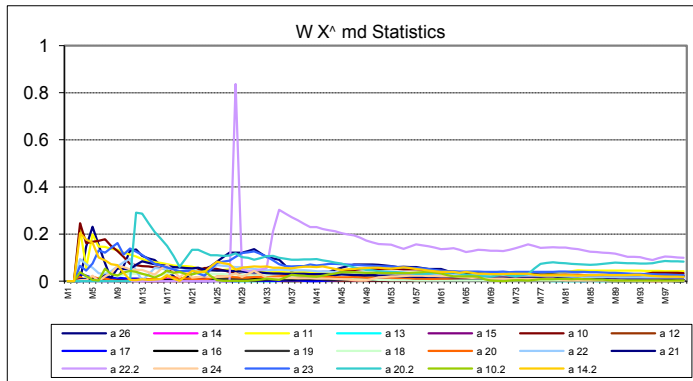
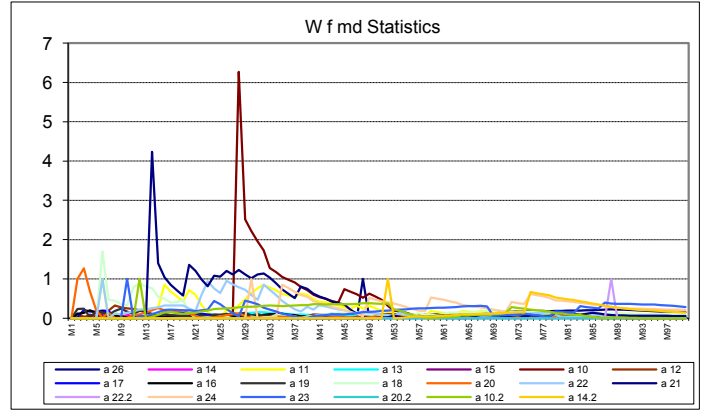
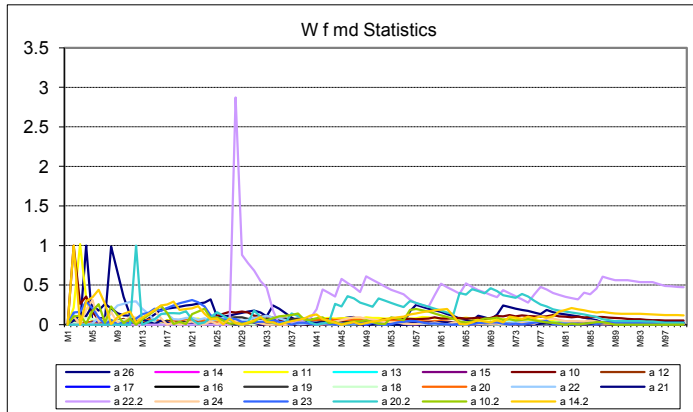
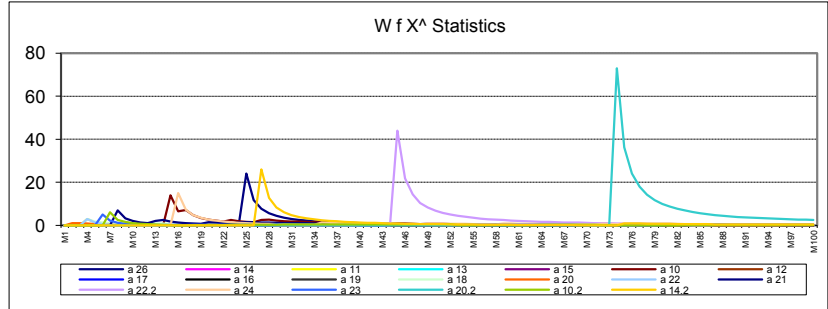
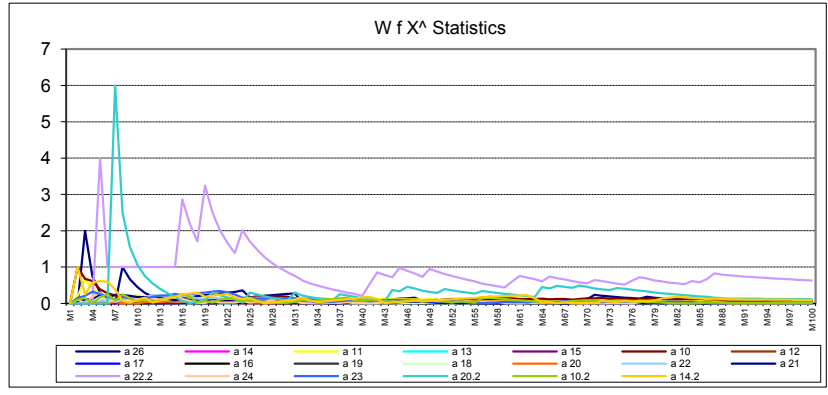
Graphic analysis of polymorphisms.

System	D18S51	Pop.	Bogotá
--------	--------	------	--------

W Statistics to Specific sample

Alleles	W f X [^]	W f md	W X [^] md
a 10	0.0004	0.0513	0.0509
a 10,2	0.0001	0.0844	0.0843
a 11	0.0001	0.0277	0.0277
a 12	0.0001	0.0006	0.0006
a 13	0.0001	0.0087	0.0088
a 14	0.0000	0.0110	0.0110
a 14,2	0.0006	0.0522	0.0528
a 15	0.0001	0.0054	0.0053
a 16	0.0001	0.0101	0.0100
a 17	0.0001	0.0037	0.0036
a 18	0.0000	0.0039	0.0039
a 19	0.0002	0.0108	0.0110
a 20	0.0002	0.0075	0.0077
a 20,2	0.0025		
a 21	0.0005	0.0298	0.0293
a 22	0.0001	0.0463	0.0463
a 22,2	0.0021		
a 23	0.0000	0.0582	0.0582
a 24	0.0011	0.1152	0.1164
a 26	0.0011		

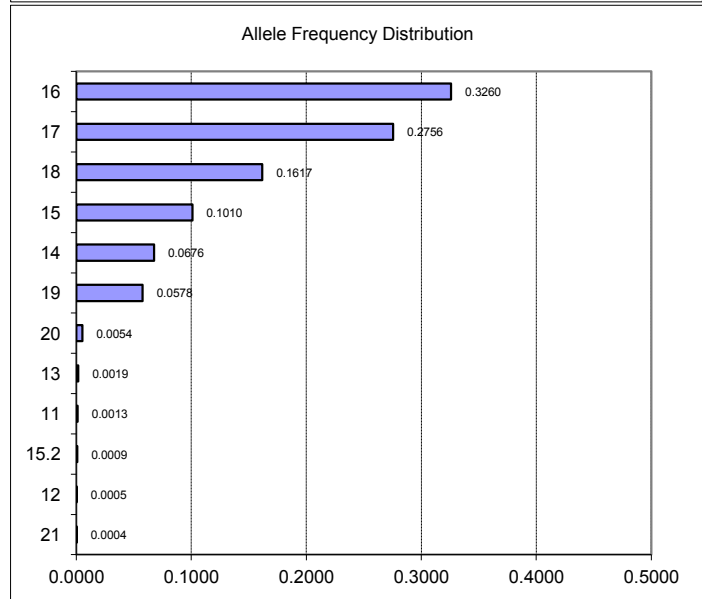
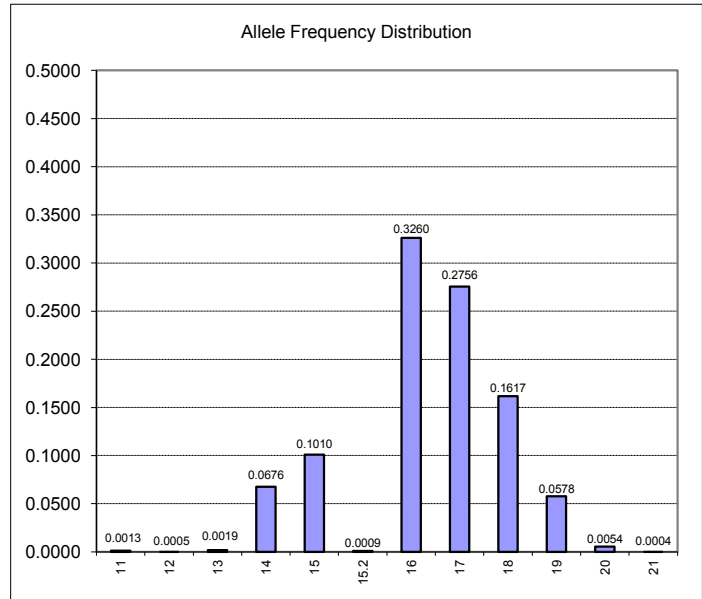
0.0025 0.1152 0.1164
Sampling without repetition
 Indeterminate value (Division by zero)



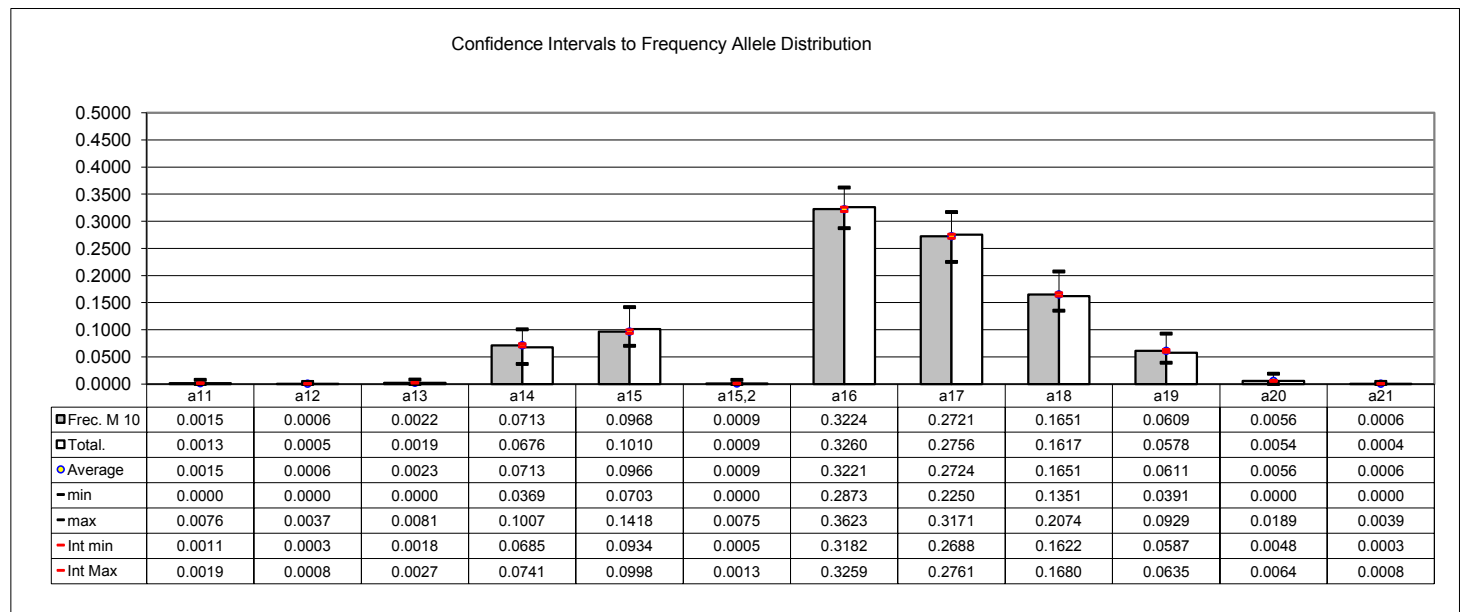
Population	Bogotá
System	VWA
Alleles number	12
Effective alleles	4

Final frequency.

Allele	M.Weight.	S.D.	Frequency
11	154.59	0.08	0.0013
12	158.87	0.07	0.0005
13	163.00	0.05	0.0019
14	167.27	0.05	0.0676
15	171.15	0.05	0.1010
15.2	173.15		0.0009
16	175.15	0.04	0.3260
17	179.15	0.04	0.2756
18	183.08	0.04	0.1617
19	187.00	0.04	0.0578
20	190.93	0.05	0.0054
21	194.80	0.05	0.0004
22	198.62	0.06	
23	202.44	0.05	
24	206.69	0.08	



Single Alleles	
In the Database	In the Sample
12	12
21	21

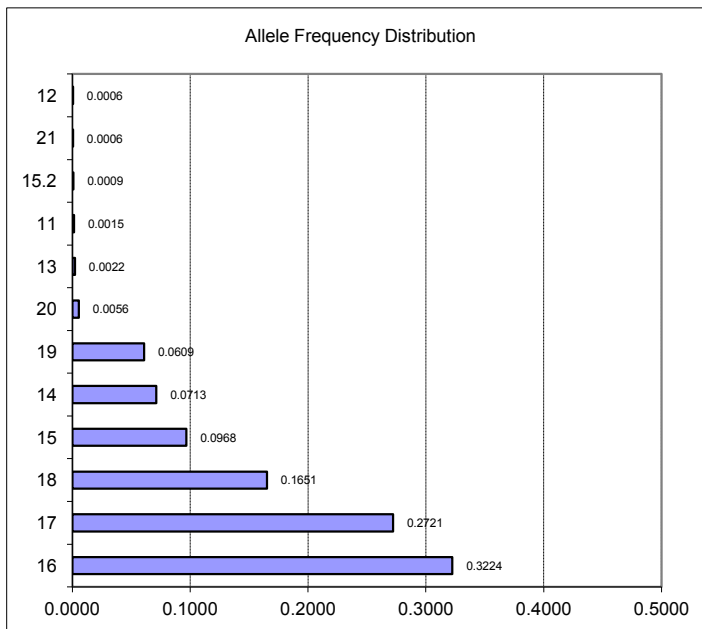


Specific sample analysis.

Sampling Specific.	
Population	Bogotá
System	VWA
Alleles number	12
Effective alleles number	4
CV	0.05
Sample number	10
Sample size	134
Polimorphism alleles total	28

EXPERIMENT	(Todas)
Sample number	10

Allele frequency summary	
Alleles	Total
16	0.3224
17	0.2721
18	0.1651
15	0.0968
14	0.0713
19	0.0609
20	0.0056
13	0.0022
11	0.0015
15.2	0.0009
21	0.0006
12	0.0006



Single Alleles in the Sampling

EXPERIMENT	(Todas)
Sample number	10

Sampling and CV.

Sample selection
Analysis of CV.

Max	Total
Alleles	
12	1
21	1
13	2
15.2	2
11	2
20	5
19	26
14	27
15	38
18	56
17	84
16	100

Alleles	Sample No.
14	4
11	0
13	0
15	2
12	0
17	2
16	2
19	21
18	2
20	0
21	0
15.2	0

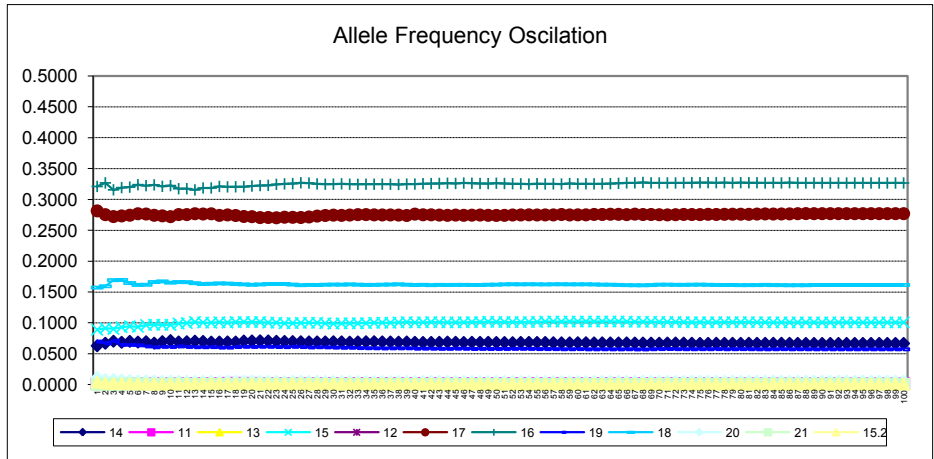
Allelic Distribution Order

Population	Bogotá
System	VWA

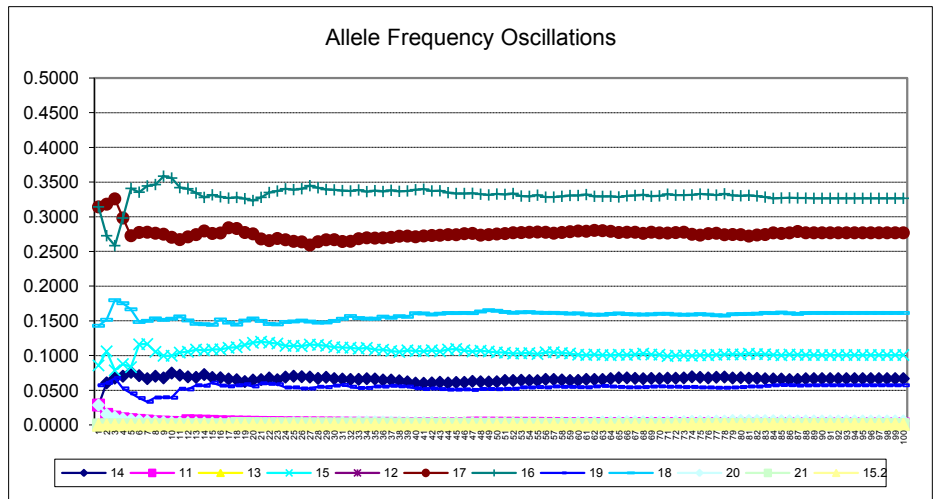
EXPERIMENT	(Todas)
Sample number	(Todas)

Allele frequency summary		
Alleles		Total
	16	0.3260
	17	0.2756
	18	0.1617
	15	0.1010
	14	0.0676
	19	0.0578
	20	0.0054
	13	0.0019
	11	0.0013
	15.2	0.0009
	12	0.0005
	21	0.0004

Compiled Allele Frequency



Allele frequency for a specific experiment.



Analysis for Specific Sampling

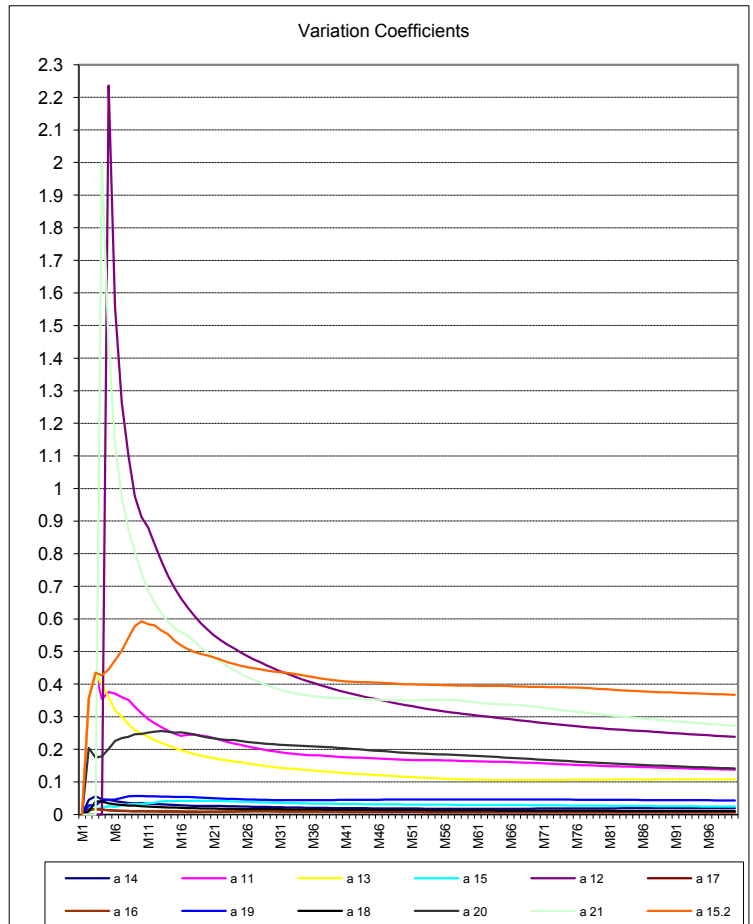
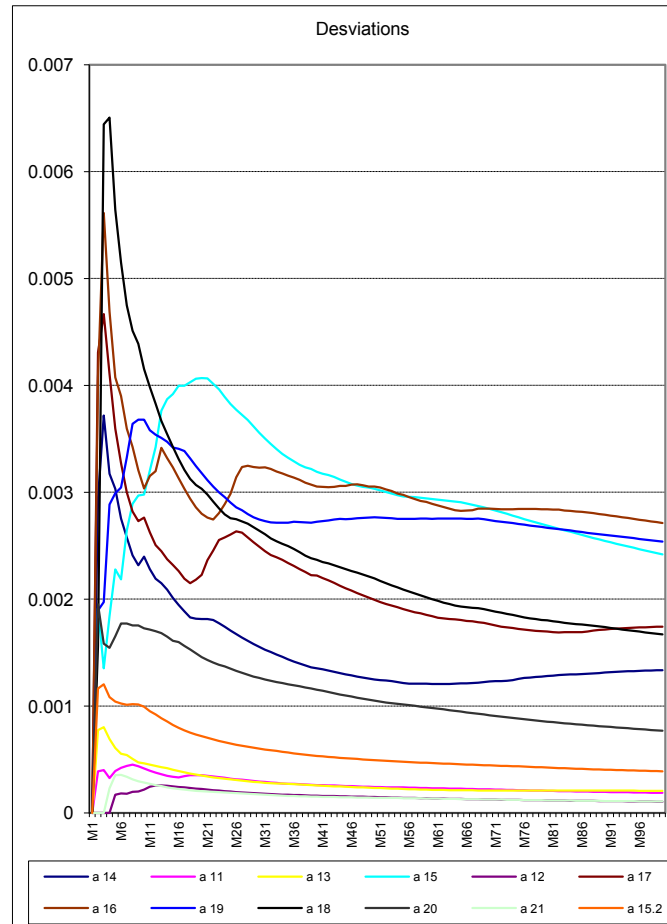
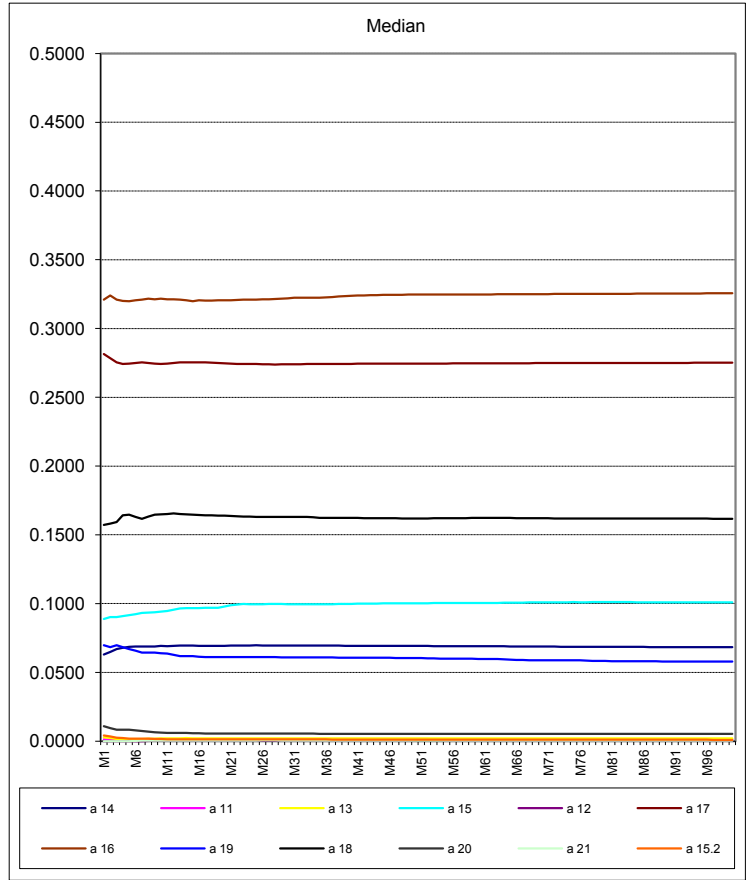
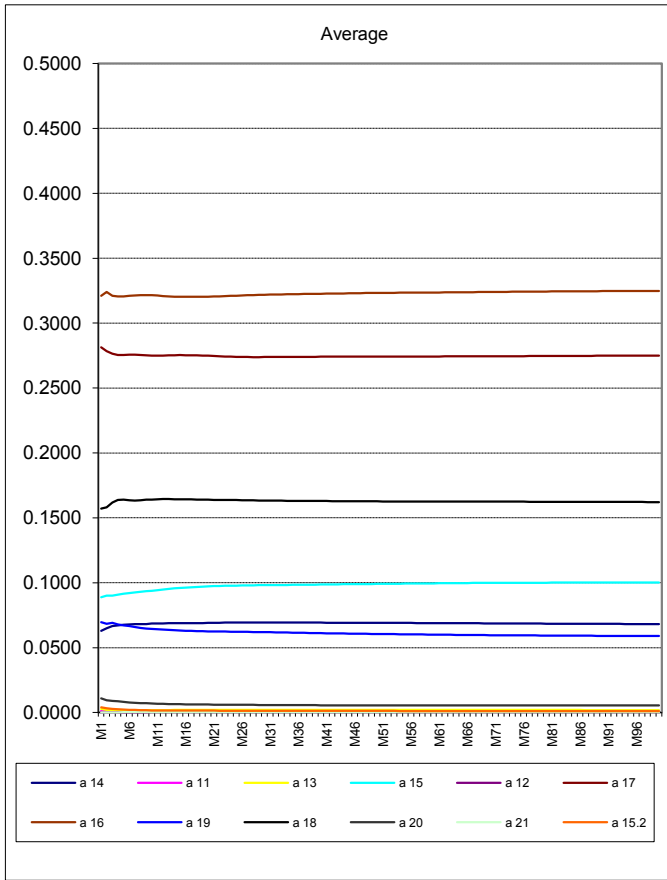
Sample number	10
Sample size	134

Average D	Min	Max
0.0018	0.0002	0.0039

Alleles	Average	min	max	Int min	Int Max	Median	S.D	IC
a11	0.0015	0.0000	0.0076	0.0011	0.0019	0.0000	0.0023	0.0004
a12	0.0006	0.0000	0.0037	0.0003	0.0008	0.0000	0.0013	0.0002
a13	0.0023	0.0000	0.0081	0.0018	0.0027	0.0000	0.0029	0.0005
a14	0.0713	0.0369	0.1007	0.0685	0.0741	0.0724	0.0165	0.0028
a15	0.0966	0.0703	0.1418	0.0934	0.0998	0.0919	0.0187	0.0032
a15,2	0.0009	0.0000	0.0075	0.0005	0.0013	0.0000	0.0024	0.0004
a16	0.3221	0.2873	0.3623	0.3182	0.3259	0.3259	0.0229	0.0039
a17	0.2724	0.2250	0.3171	0.2688	0.2761	0.2750	0.0216	0.0036
a18	0.1651	0.1351	0.2074	0.1622	0.1680	0.1631	0.0173	0.0029
a19	0.0611	0.0391	0.0929	0.0587	0.0635	0.0603	0.0144	0.0024
a20	0.0056	0.0000	0.0189	0.0048	0.0064	0.0037	0.0047	0.0008
a21	0.0006	0.0000	0.0039	0.0003	0.0008	0.0000	0.0014	0.0002

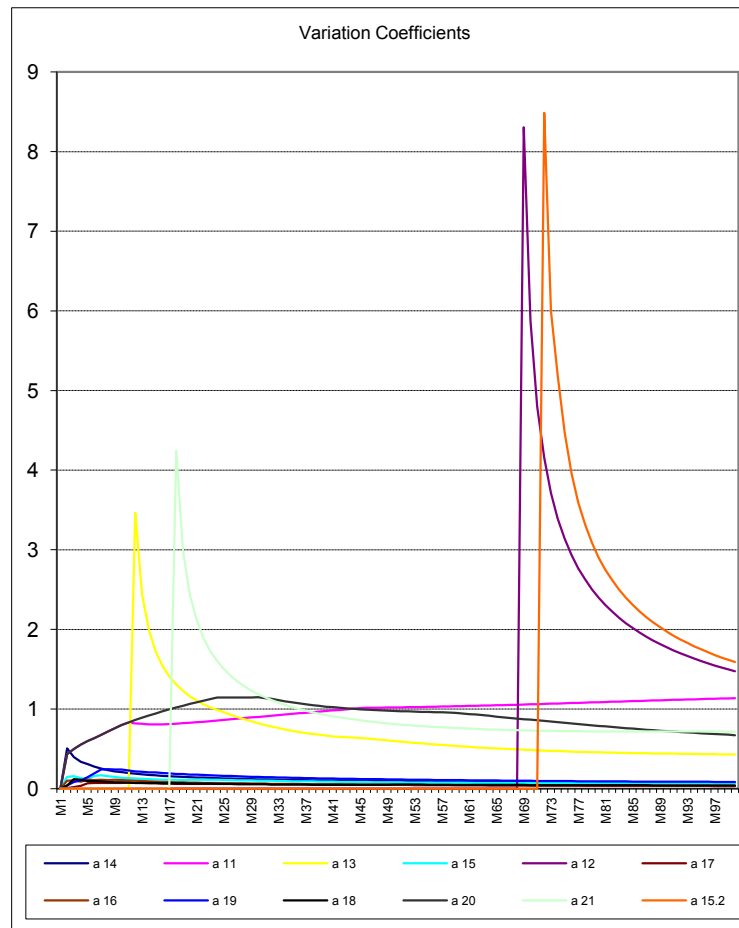
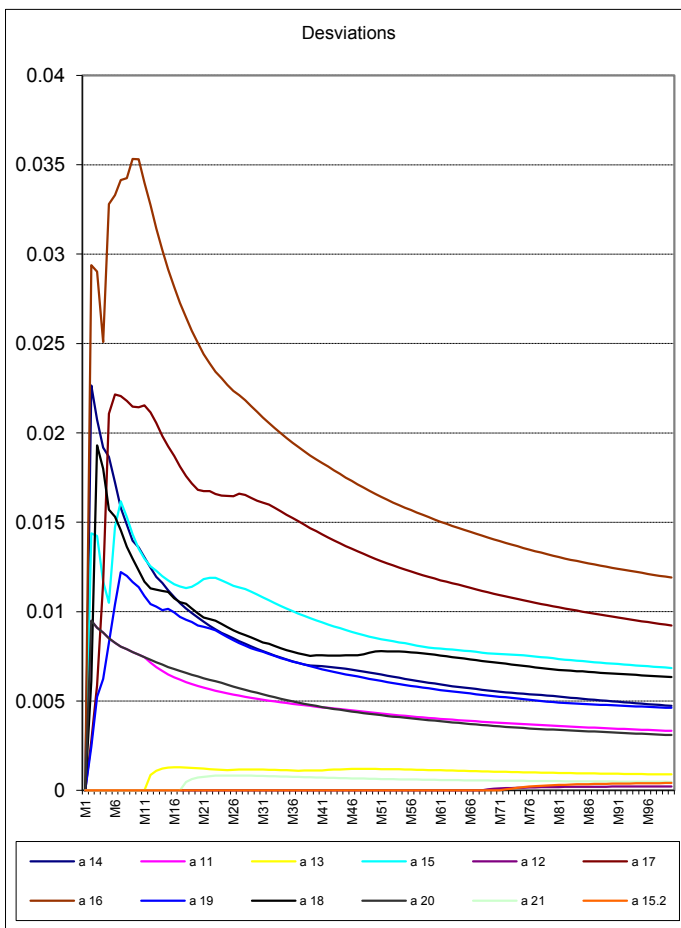
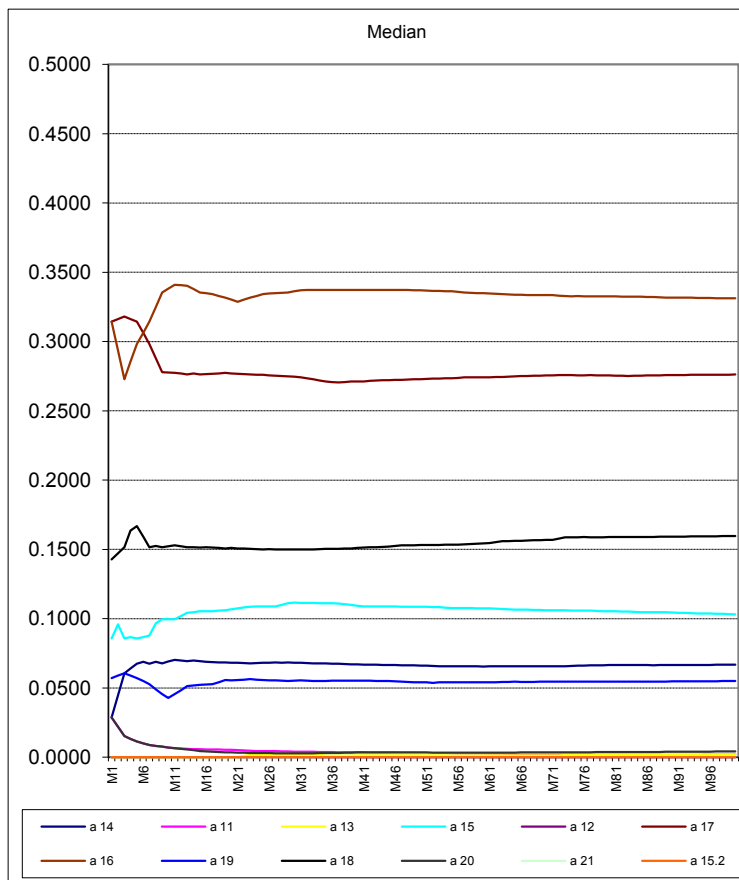
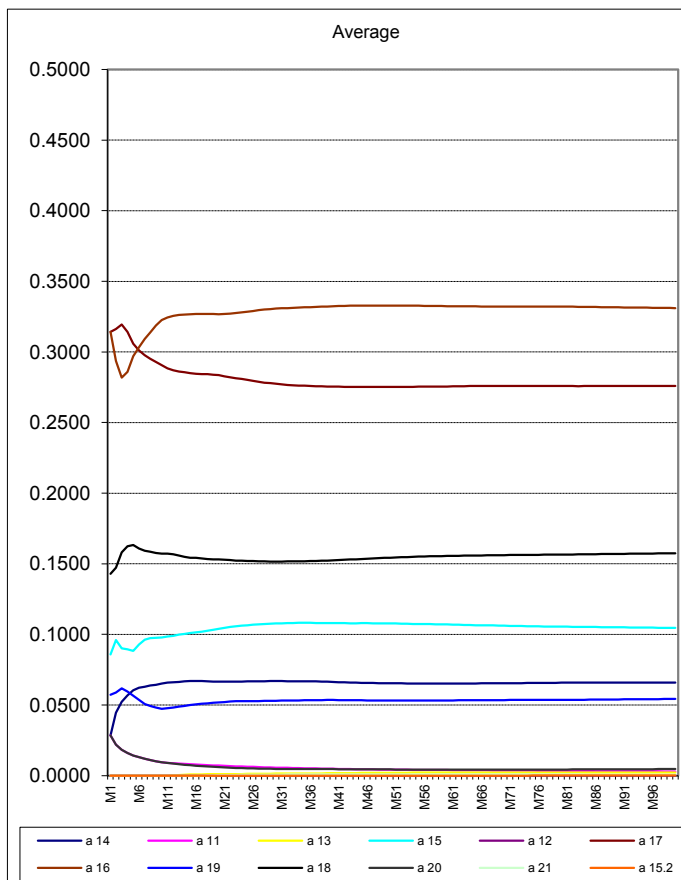
Graphic analysis of polymorphisms.

System VWA Pop. Bogotá



Graphic analysis of polymorphisms.

System	VWA	Pop.	Bogotá
--------	-----	------	--------



Graphic analysis of polymorphisms.

System	VWA	Pop.	Bogotá
--------	-----	------	--------

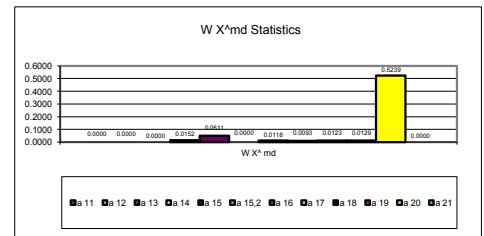
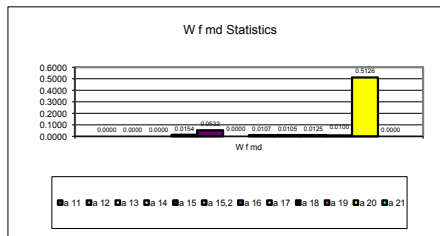
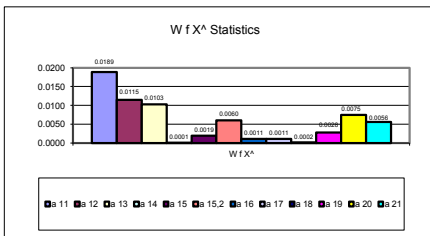
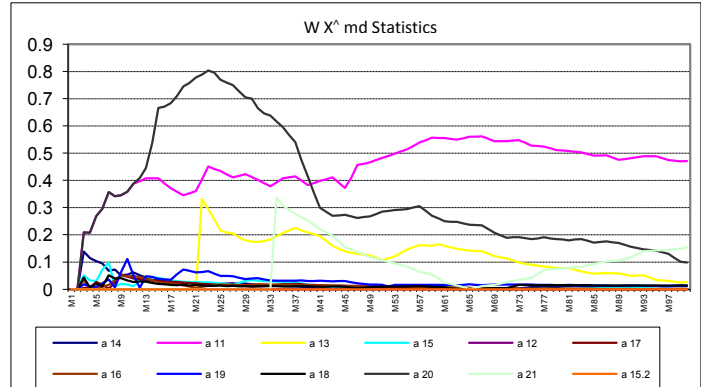
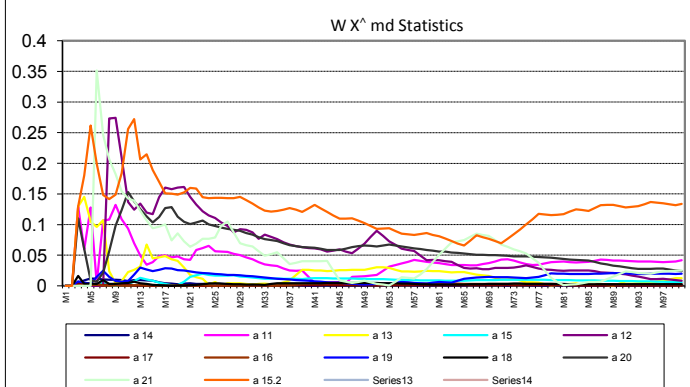
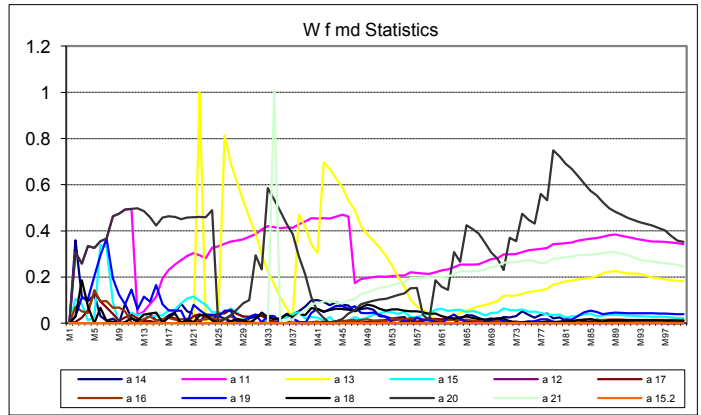
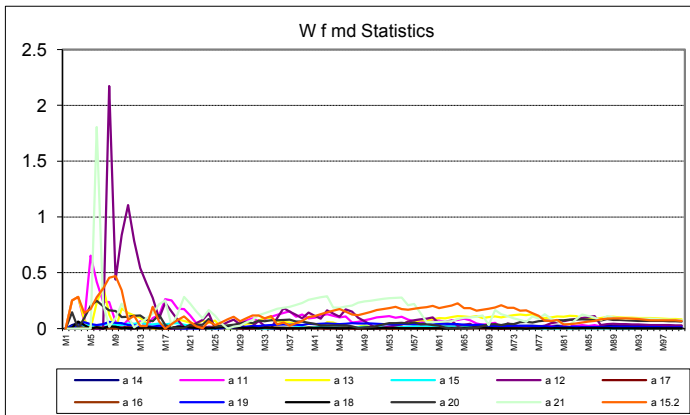
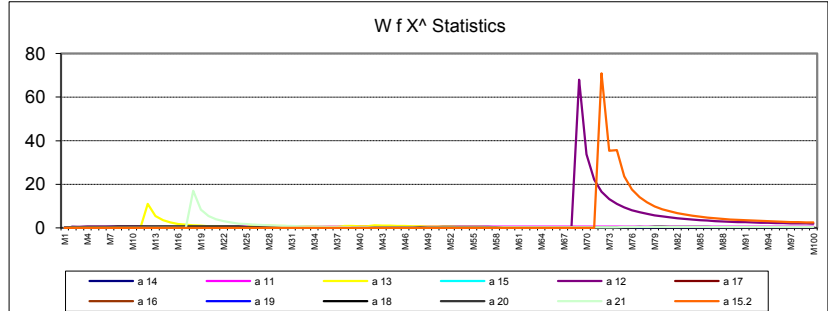
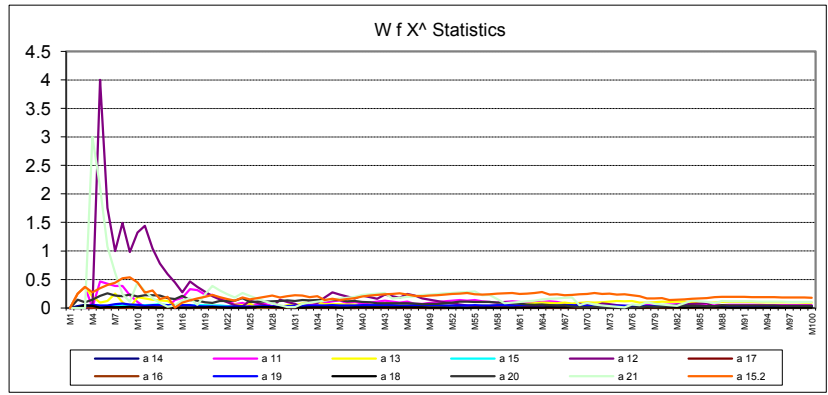
W Statistics to Specific sample

Alleles	W f X^	W f md	W X^ md
a 11	0.0189		
a 12	0.0115		
a 13	0.0103		
a 14	0.0001	0.0154	0.0152
a 15	0.0019	0.0532	0.0511
a 15,2	0.0060		
a 16	0.0011	0.0107	0.0118
a 17	0.0011	0.0105	0.0093
a 18	0.0002	0.0125	0.0123
a 19	0.0028	0.0100	0.0129
a 20	0.0075	0.5126	0.5239
a 21	0.0056		

Indeterminate value (Division by zero)

0.0189 0.5126 0.5239

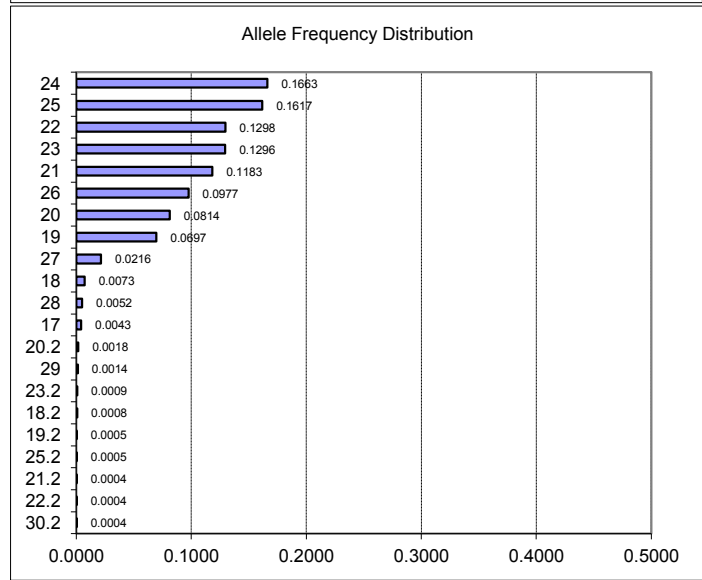
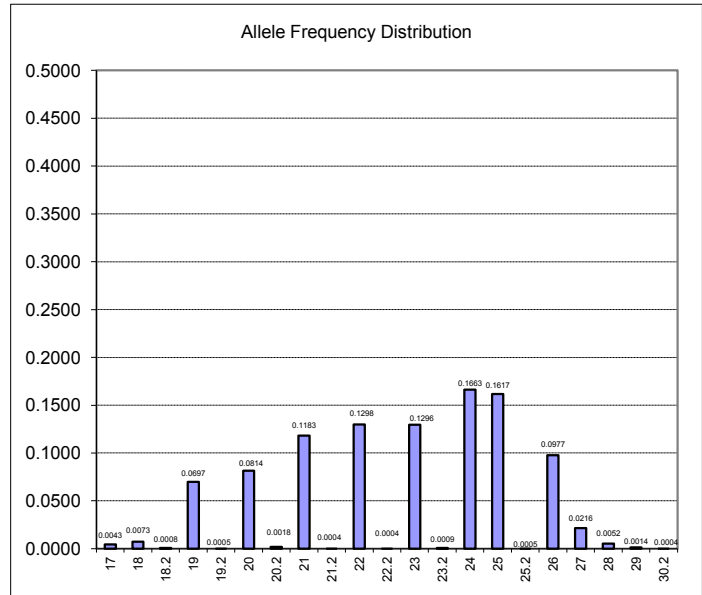
Sampling without repetition



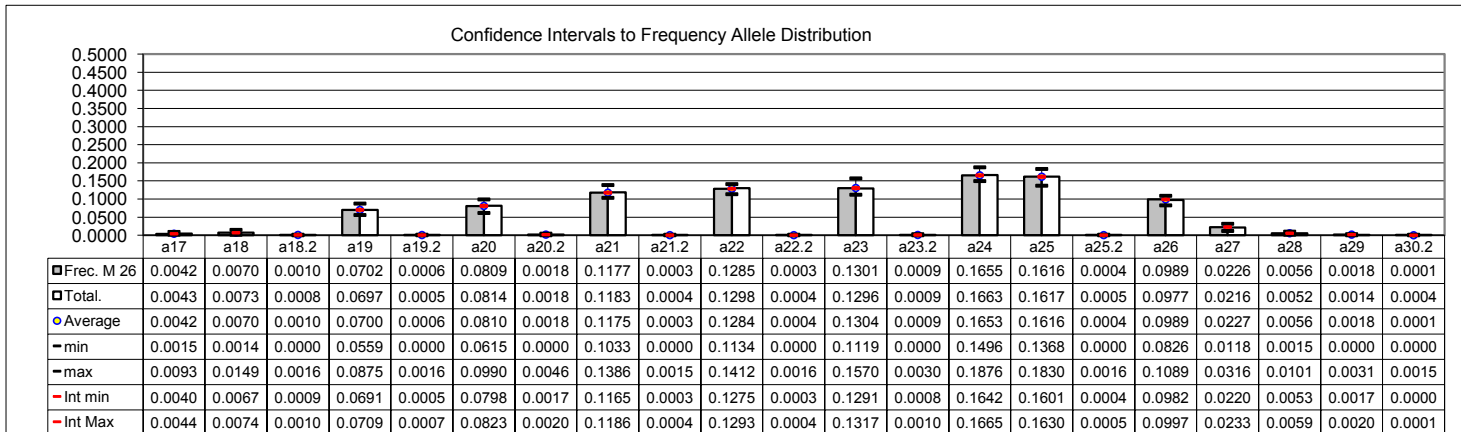
Population	Bogotá
System	FGA
Alleles number	21
Effective alleles	8

Final frequency.

Allele	M.Weight.	S.D.	Frequency
16	210.00	*	0.0040
17	214.81	0.07	0.0043
18	218.80	0.06	0.0073
18.2	220.00	*	0.0008
19	222.79	0.07	0.0697
19.2	224.00	*	0.0005
20	226.81	0.06	0.0814
20.2	228.00	*	0.0018
21	230.76	0.08	0.1183
21.2	232.00	*	0.0004
22	234.78	0.07	0.1298
22.2	236.00	*	0.0004
23	238.81	0.05	0.1296
23.2	240.00	*	0.0009
24	242.83	0.07	0.1663
24.2	244.00	*	0.0020
25	246.88	0.06	0.1617
25.2	248.00	*	0.0005
26	250.96	0.06	0.0977
26.2	252.00	0.09	
27	254.97	0.08	0.0216
28	259.02	0.10	0.0052
29	263.12	0.08	0.0014
30	267.26	0.09	
30.2	269.07	0.10	0.0004
31.2	273.17	0.09	
32.2	277.24	0.08	
33.2	281.33	0.09	
42.2	319.83	0.14	
43.2	324.04	0.14	
44.2	328.26	0.13	
45.2	332.42	0.16	
46.2	336.43	0.14	
47.2	340.42	0.14	
48.2	344.15	0.10	
50.2	351.45	0.05	
51.2	355.13	0.05	



Single Alleles		* Ref. 363 (Identifiler)
In the Database		In the Sample
	19.2	18.2
	21.2	19.2
	22.2	21.2
	25.2	22.2
	30.2	25.2
		30.2

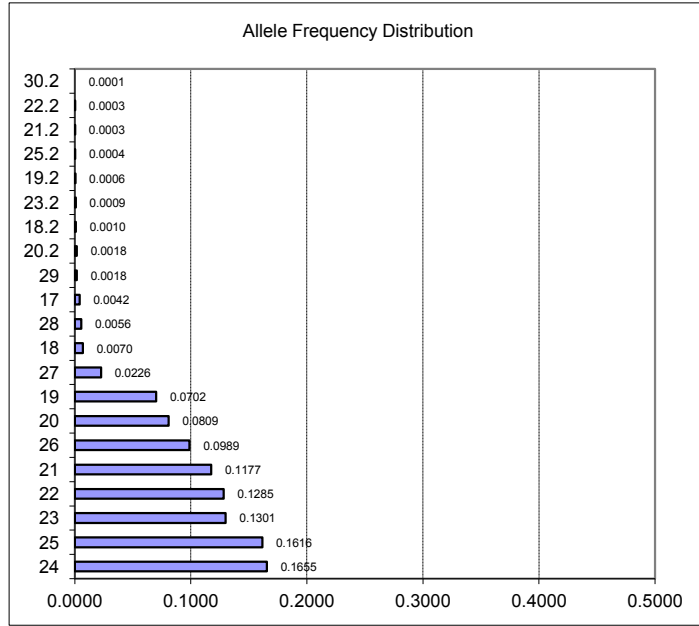


Specific sample analysis.

Sampling Specific.	
Population	Bogotá
System	FGA
Alleles number	21
Effective alleles number	8
CV	0.05
Sample number	26
Sample size	339
Polimorphism alleles total	98

EXPERIMENT	(Todas)
Sample number	26

Allele frequency summary	
Alleles	Total
24	0.1655
25	0.1616
23	0.1301
22	0.1285
21	0.1177
26	0.0989
20	0.0809
19	0.0702
27	0.0226
18	0.0070
28	0.0056
17	0.0042
29	0.0018
20.2	0.0018
18.2	0.0010
23.2	0.0009
19.2	0.0006
25.2	0.0004
21.2	0.0003
22.2	0.0003
30.2	0.0001



Single Alleles in the Sampling

EXPERIMENT	(Todas)
Sample number	26

Max	Alleles	Total
	25.2	1
	21.2	1
	22.2	1
	30.2	1
	19.2	1
	18.2	1
	29	2
	23.2	2
	20.2	3
	17	6
	28	7
	18	10
	27	21
	19	61
	20	69
	26	77
	21	94
	22	100
	23	106
	25	127
	24	136

Sampling and CV.

Sample selection
Analysis of CV.

Alleles	Sample No.
25	9
26	14
27	0
28	0
29	0
30.2	0
17	0
19	26
20	4
21	2
22	2
23	5
23.2	0
24	2
18	0
18.2	0
20.2	0
19.2	0
22.2	0
25.2	0
21.2	0

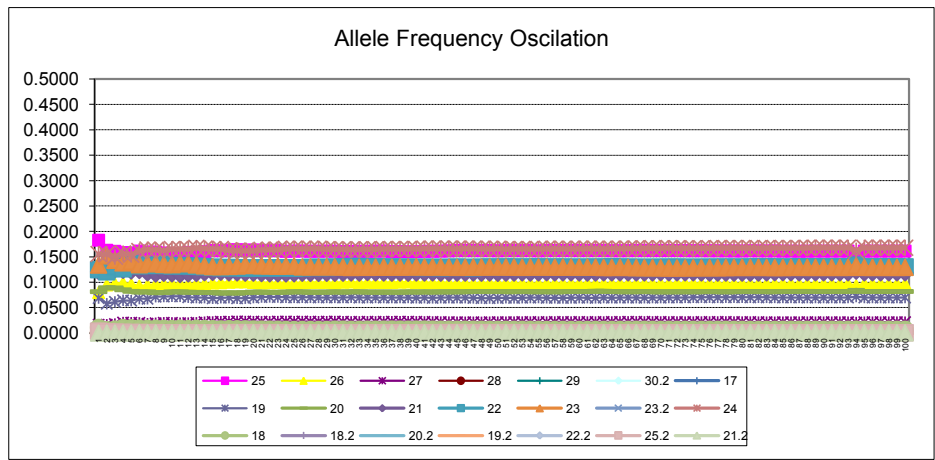
Allelic Distribution Order

Population	Bogotá
System	FGA

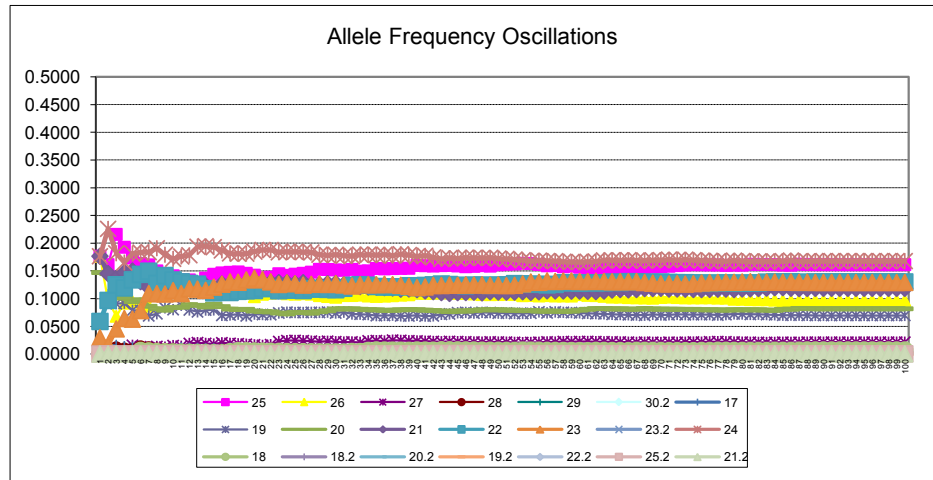
EXPERIMENT	(Todas)
Sample number	(Todas)

Allele frequency summary	
Alleles	Total
24	0.1663
25	0.1617
22	0.1298
23	0.1296
21	0.1183
26	0.0977
20	0.0814
19	0.0697
27	0.0216
18	0.0073
28	0.0052
17	0.0043
20.2	0.0018
29	0.0014
23.2	0.0009
18.2	0.0008
19.2	0.0005
25.2	0.0005
21.2	0.0004
22.2	0.0004
30.2	0.0004

Compiled Allele Frequency



Allele frequency for a specific experiment.



Analysis for Specific Sampling

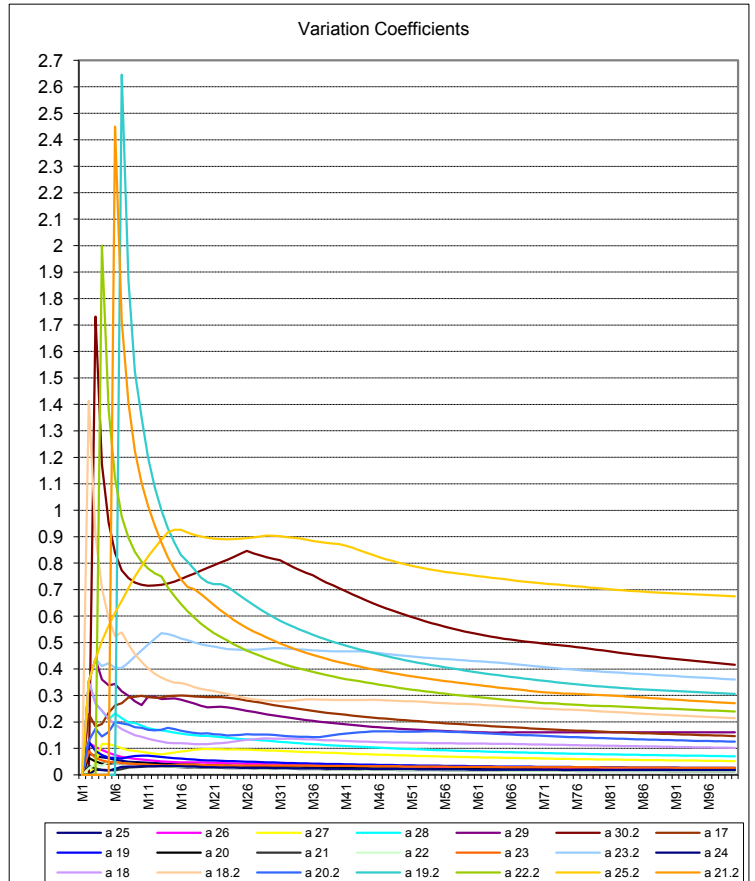
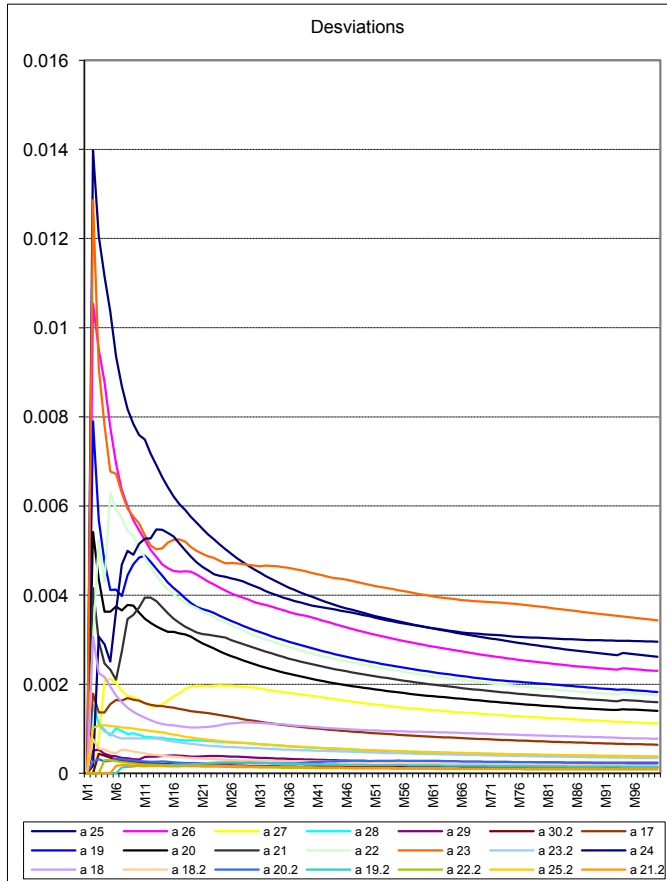
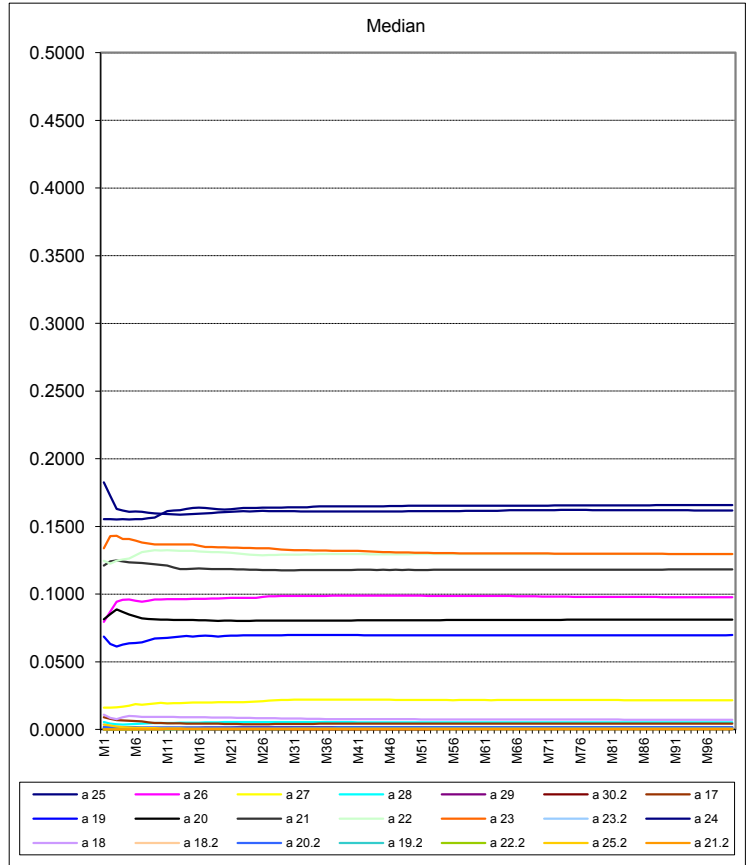
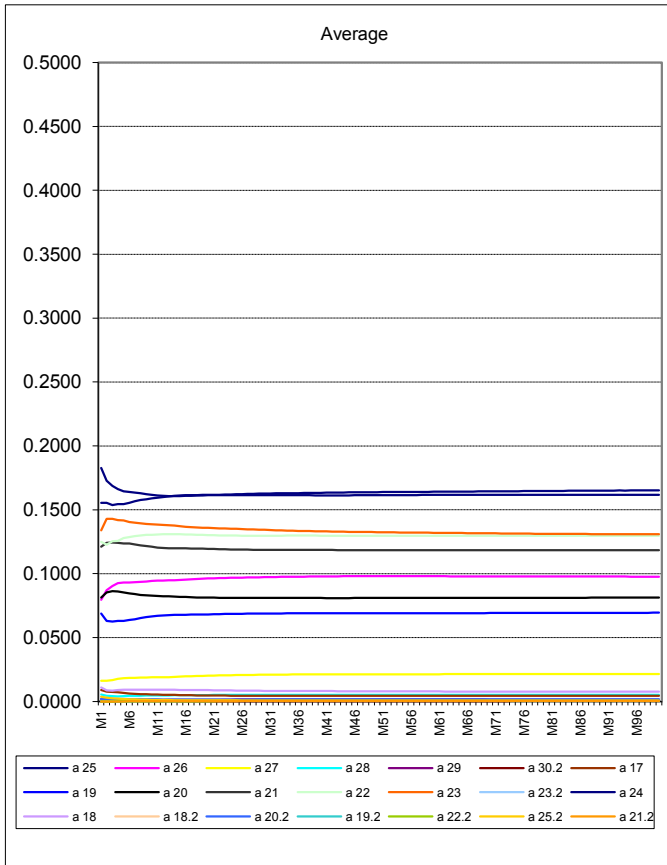
Sample number	26
Sample size	339

Average D	Min	Max
0.0005	0.0000	0.0015

Alleles	Average	min	max	Int min	Int Max	Median	S.D	IC
a17	0.0042	0.0015	0.0093	0.0040	0.0044	0.0043	0.0019	0.0002
a18	0.0070	0.0014	0.0149	0.0067	0.0074	0.0061	0.0032	0.0003
a18.2	0.0010	0.0000	0.0016	0.0009	0.0010	0.0014	0.0007	0.0001
a19	0.0700	0.0559	0.0875	0.0691	0.0709	0.0708	0.0085	0.0009
a19.2	0.0006	0.0000	0.0016	0.0005	0.0007	0.0000	0.0008	0.0001
a20	0.0810	0.0615	0.0990	0.0798	0.0823	0.0817	0.0118	0.0013
a20.2	0.0018	0.0000	0.0046	0.0017	0.0020	0.0015	0.0014	0.0002
a21	0.1175	0.1033	0.1386	0.1165	0.1186	0.1166	0.0099	0.0010
a21.2	0.0003	0.0000	0.0015	0.0003	0.0004	0.0000	0.0006	0.0001
a22	0.1284	0.1134	0.1412	0.1275	0.1293	0.1285	0.0087	0.0009
a22.2	0.0004	0.0000	0.0016	0.0003	0.0004	0.0000	0.0007	0.0001
a23	0.1304	0.1119	0.1570	0.1291	0.1317	0.1265	0.0121	0.0013
a23.2	0.0009	0.0000	0.0030	0.0008	0.0010	0.0014	0.0009	0.0001
a24	0.1653	0.1496	0.1876	0.1642	0.1665	0.1629	0.0109	0.0012
a25	0.1616	0.1368	0.1830	0.1601	0.1630	0.1629	0.0137	0.0015
a25.2	0.0004	0.0000	0.0016	0.0004	0.0005	0.0000	0.0007	0.0001
a26	0.0989	0.0826	0.1089	0.0982	0.0997	0.0966	0.0069	0.0007
a27	0.0227	0.0118	0.0316	0.0220	0.0233	0.0232	0.0059	0.0006
a28	0.0056	0.0015	0.0101	0.0053	0.0059	0.0046	0.0027	0.0003
a29	0.0018	0.0000	0.0031	0.0017	0.0020	0.0015	0.0011	0.0001
a30.2	0.0001	0.0000	0.0015	0.0000	0.0001	0.0000	0.0004	0.0000

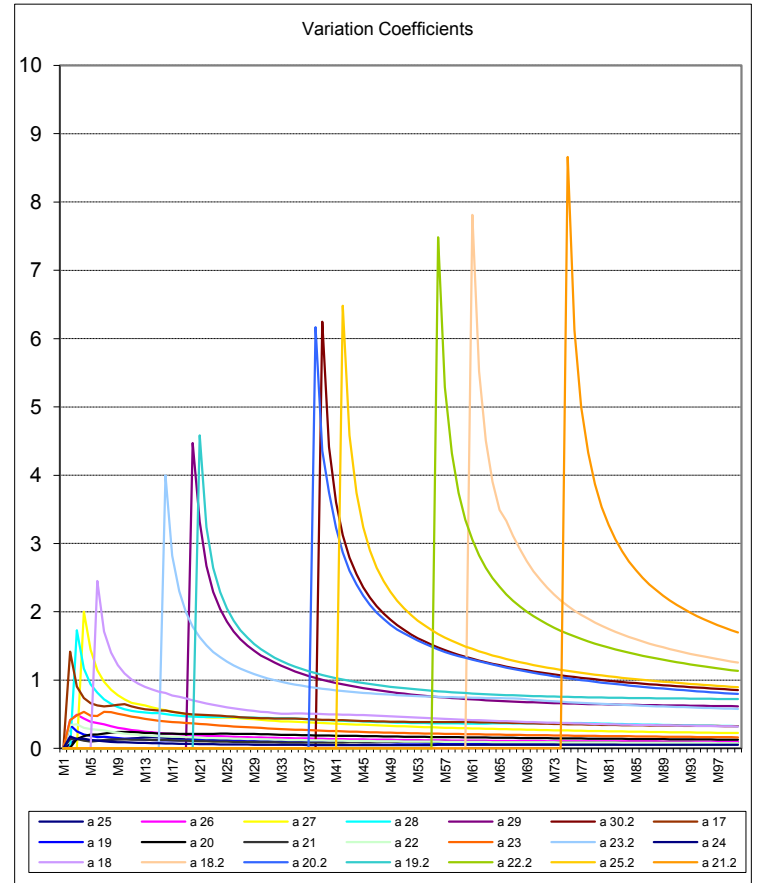
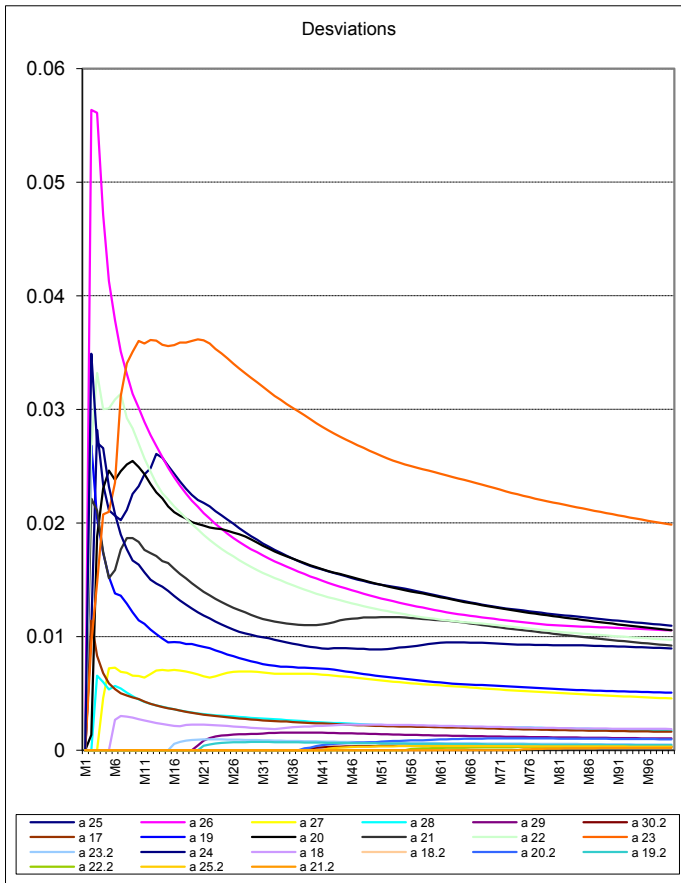
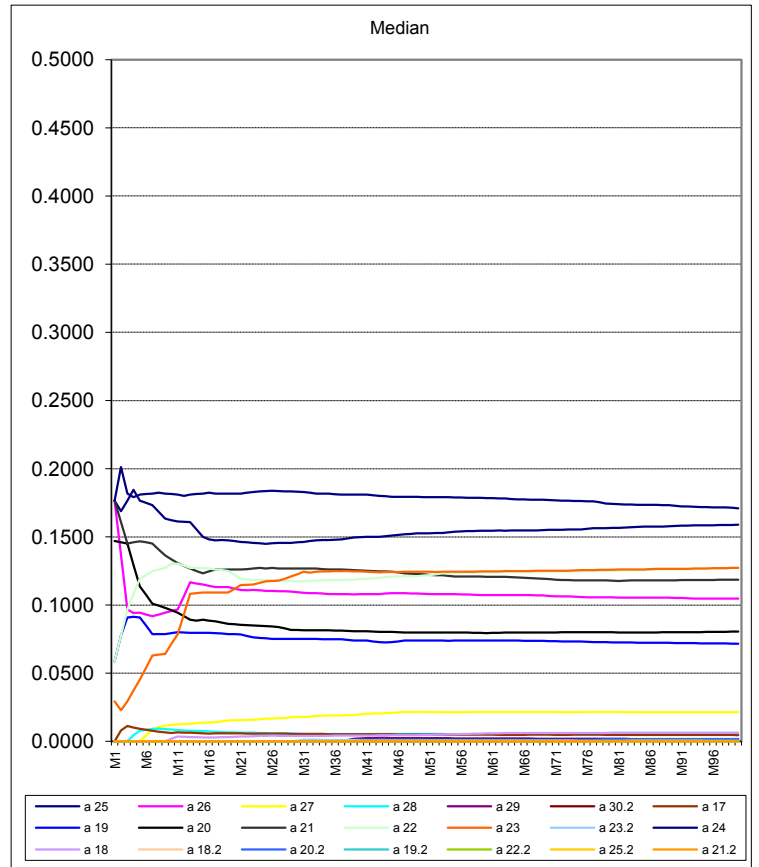
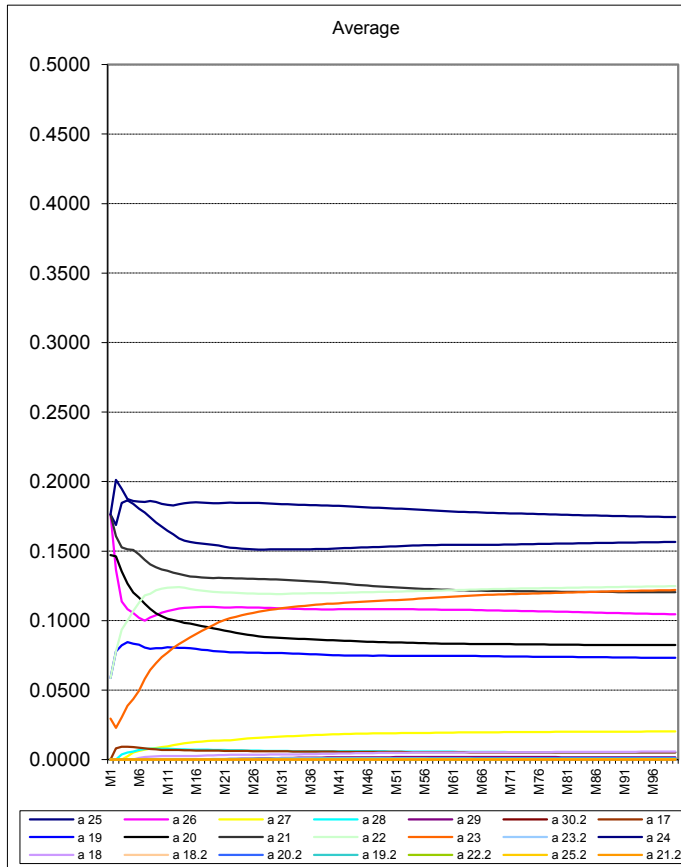
Graphic analysis of polymorphisms.

System FGA Pop. Bogotá



Graphic analysis of polymorphisms.

System	FGA	Pop.	Bogotá
--------	-----	------	--------



Graphic analysis of polymorphisms.

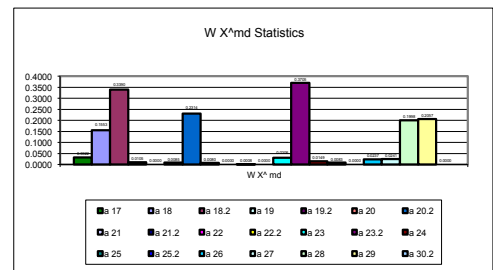
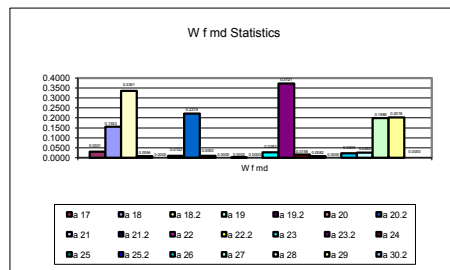
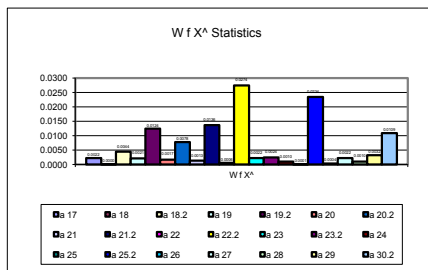
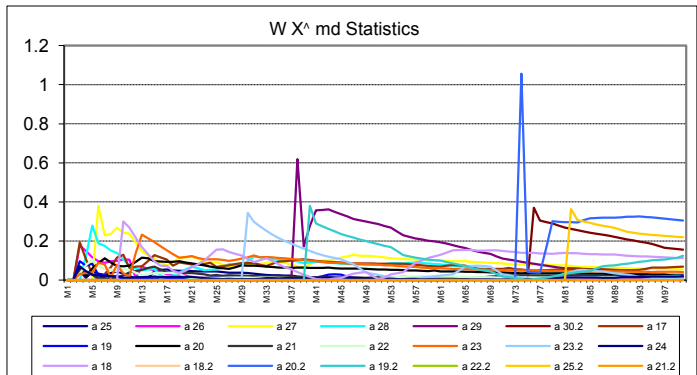
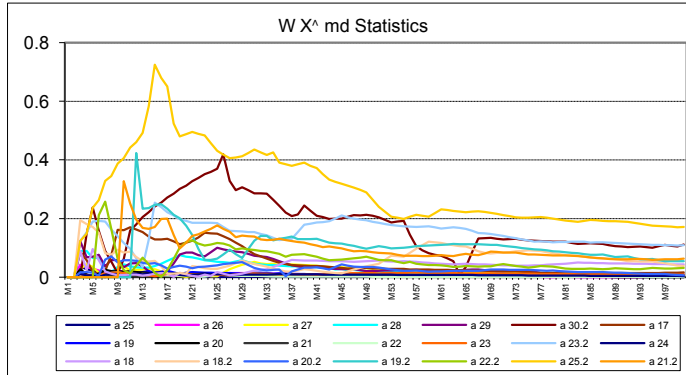
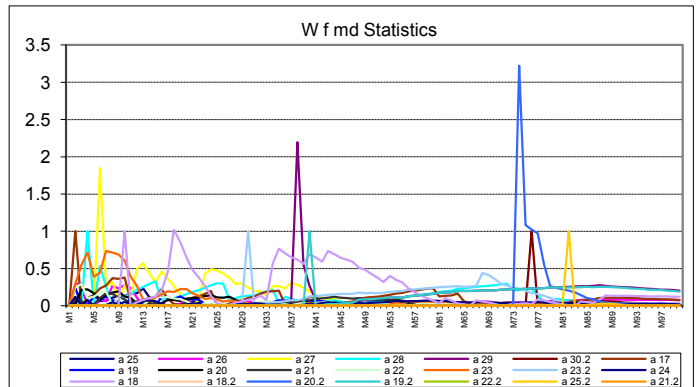
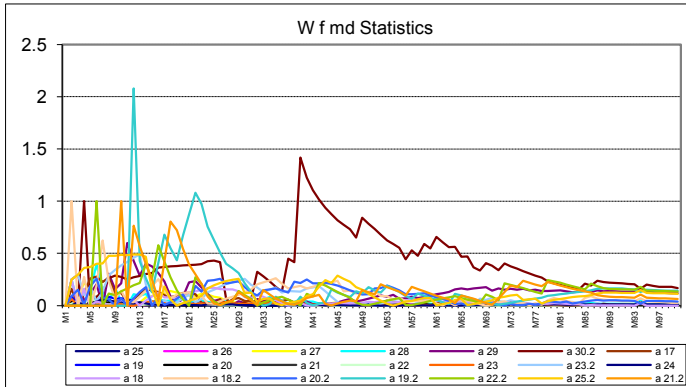
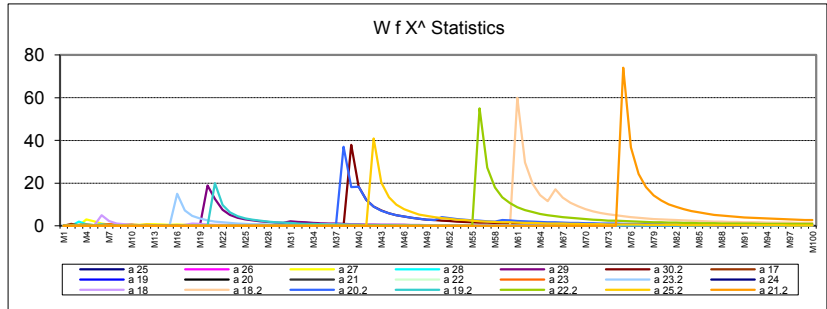
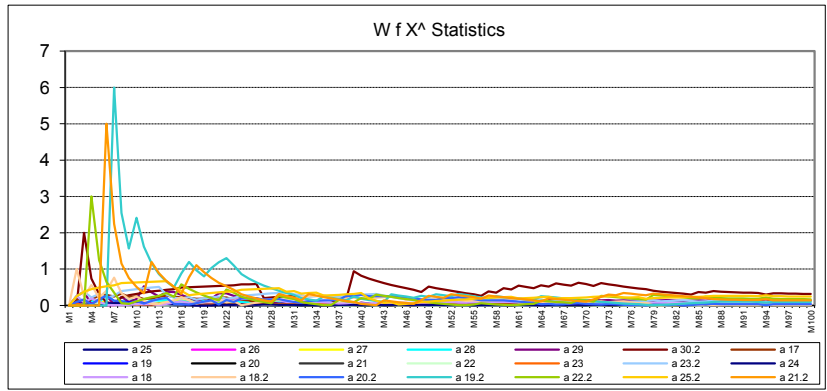
System	FGA	Pop.	Bogotá
--------	-----	------	--------

W Statistics to Specific sample

Alleles	W f X [^]	W f md	W X [^] md
a 17	0.0022	0.0301	0.0322
a 18	0.0000	0.1553	0.1553
a 18.2	0.0044	0.3361	0.3390
a 19	0.0021	0.0084	0.0105
a 19.2	0.0124		
a 20	0.0017	0.0102	0.0085
a 20.2	0.0078	0.2219	0.2314
a 21	0.0013	0.0093	0.0080
a 21.2	0.0136		
a 22	0.0006	0.0002	0.0008
a 22.2	0.0274		
a 23	0.0022	0.0283	0.0306
a 23.2	0.0024	0.3721	0.3705
a 24	0.0010	0.0158	0.0149
a 25	0.0001	0.0082	0.0083
a 25.2	0.0234		
a 26	0.0004	0.0233	0.0237
a 27	0.0022	0.0262	0.0241
a 28	0.0010	0.1986	0.1998
a 29	0.0032	0.2018	0.2057
a 30.2	0.0109		
	0.0274	0.3721	0.3705

Sampling without repetition

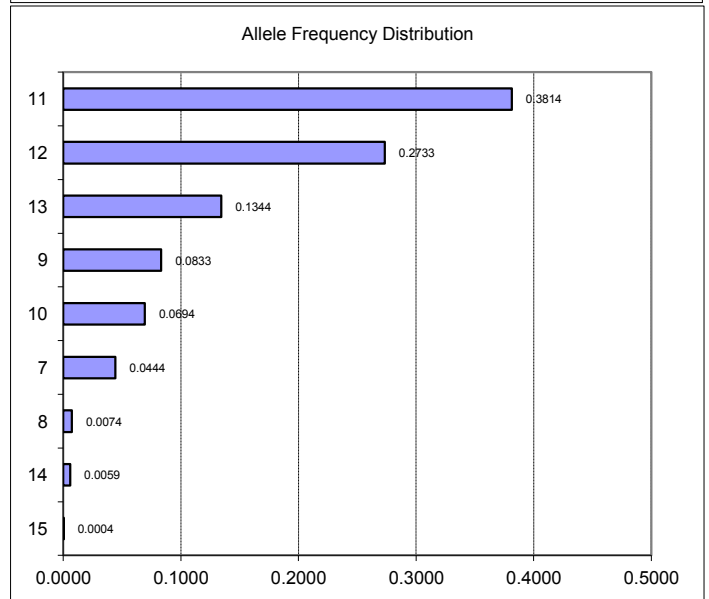
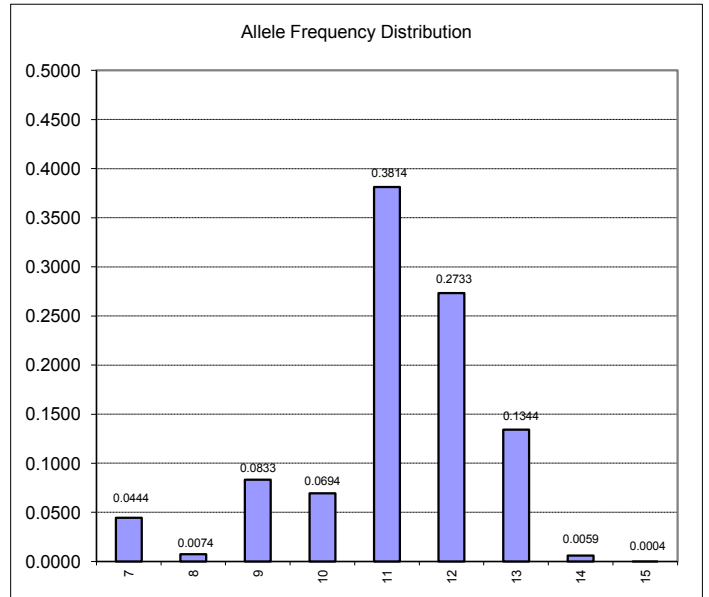
Indeterminate value (Division by zero)



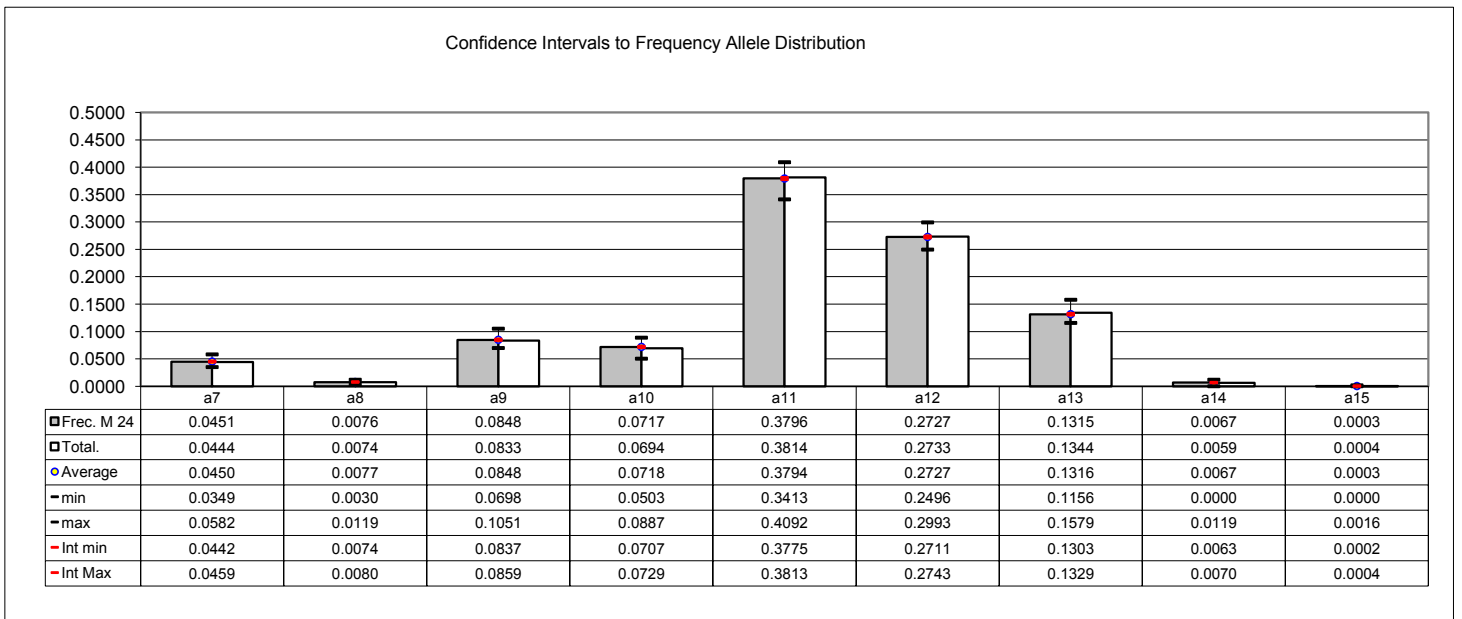
Population	Bogotá
System	D5S818
Alleles number	9
Effective alleles	4

Final frequency.

Allele	M.Weight.	S.D.	Frequency
7	134.14	0.05	0.0444
8	138.21	0.04	0.0074
9	142.56	0.04	0.0833
10	147.02	0.06	0.0694
11	151.31	0.01	0.3814
12	155.63	0.05	0.2733
13	159.81	0.06	0.1344
14	164.04	0.07	0.0059
15	167.95	0.05	0.0004
16	172.09	0.05	0.0010



Single Alleles	
In the Database	In the Sample
15	15

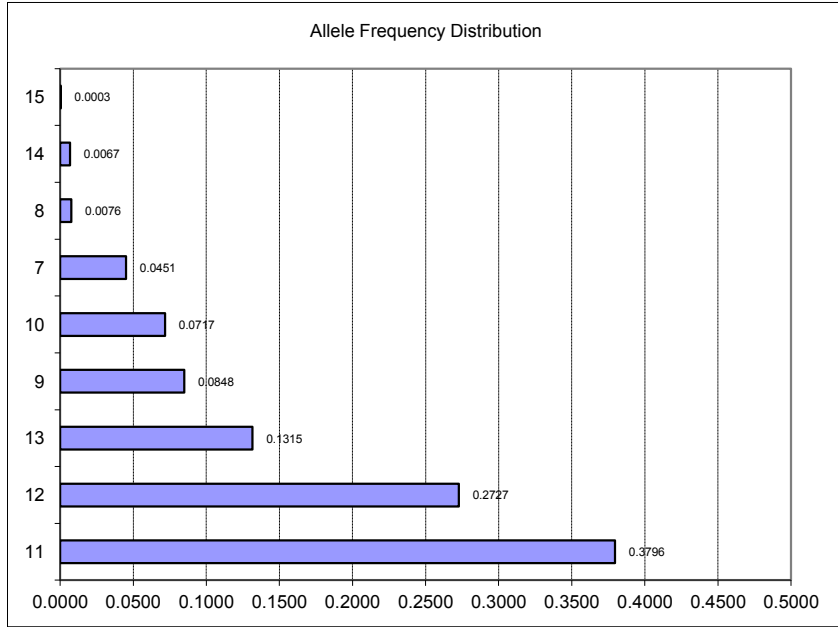


Specific sample analysis.

Sampling Specific.	
Population	Bogotá
System	D5S818
Alleles number	9
Effective alleles number	4
CV	0.05
Sample number	24
Sample size	308
Polimorphism alleles total	10

EXPERIMENT	(Todas)
Sample number	24

Allele frequency summary	
Alleles	Total
11	0.3796
12	0.2727
13	0.1315
9	0.0848
10	0.0717
7	0.0451
8	0.0076
14	0.0067
15	0.0003



Single Alleles in the Sampling

EXPERIMENT	(Todas)
Sample number	24

Max	Alleles	Total
	15	1
	14	7
	8	7
	7	36
	10	54
	9	64
	13	99
	12	192
	11	265

Sampling and CV.

Sample selection
Analysis of CV.

Alleles	Sample No.
14	0
11	2
13	24
15	0
10	2
12	2
9	2
8	0
7	66

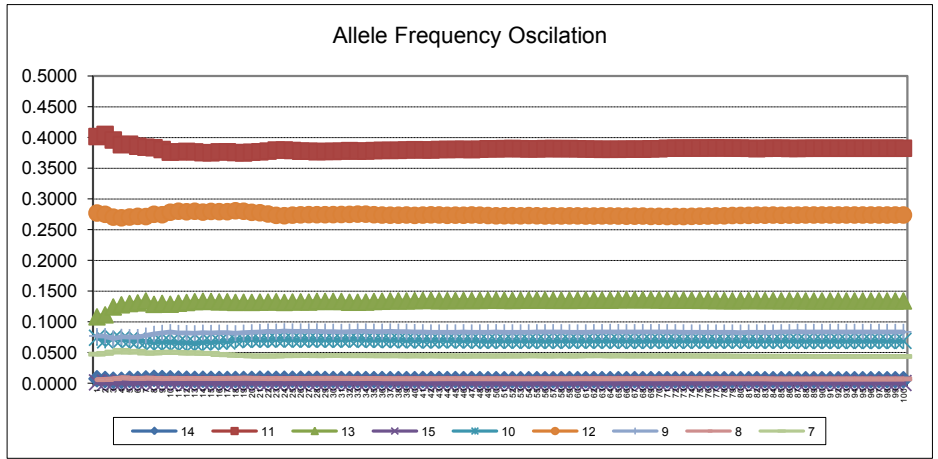
Allelic Distribution Order

Population	Bogotá
System	D5S818

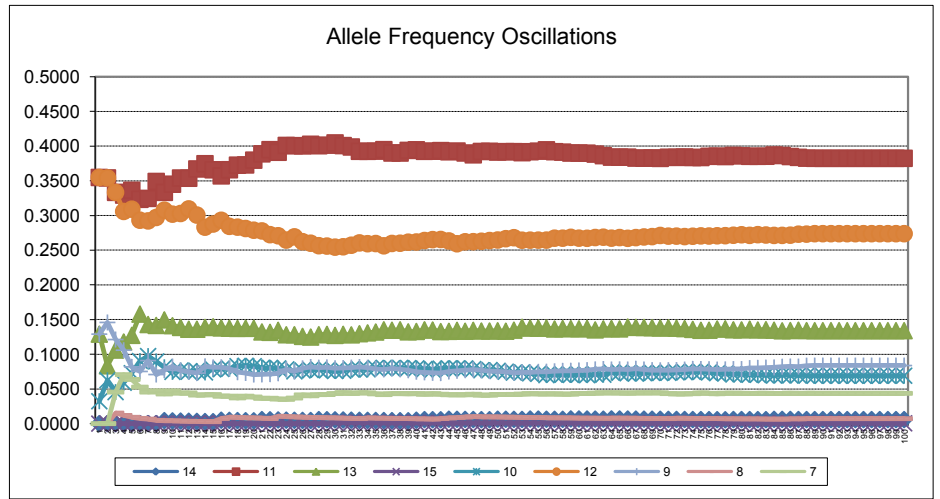
EXPERIMENT	(Todas)
Sample number	(Todas)

Allele frequency summary		
Alleles		Total
	11	0.3814
	12	0.2733
	13	0.1344
	9	0.0833
	10	0.0694
	7	0.0444
	8	0.0074
	14	0.0059
	15	0.0004

Compiled Allele Frequency



Allele frequency for a specific experiment.



Analysis for Specific Sampling

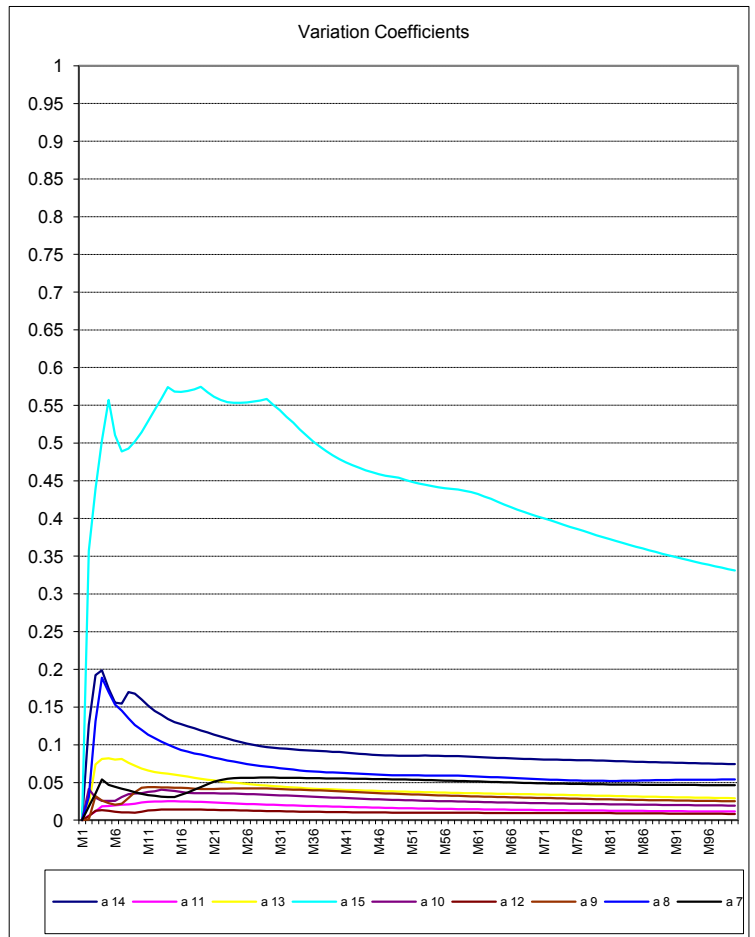
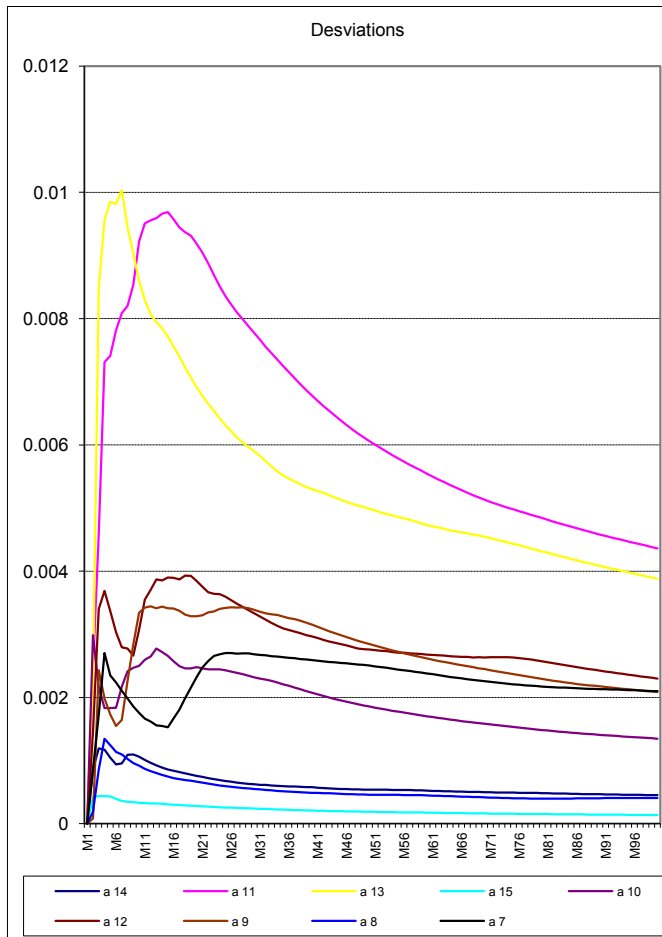
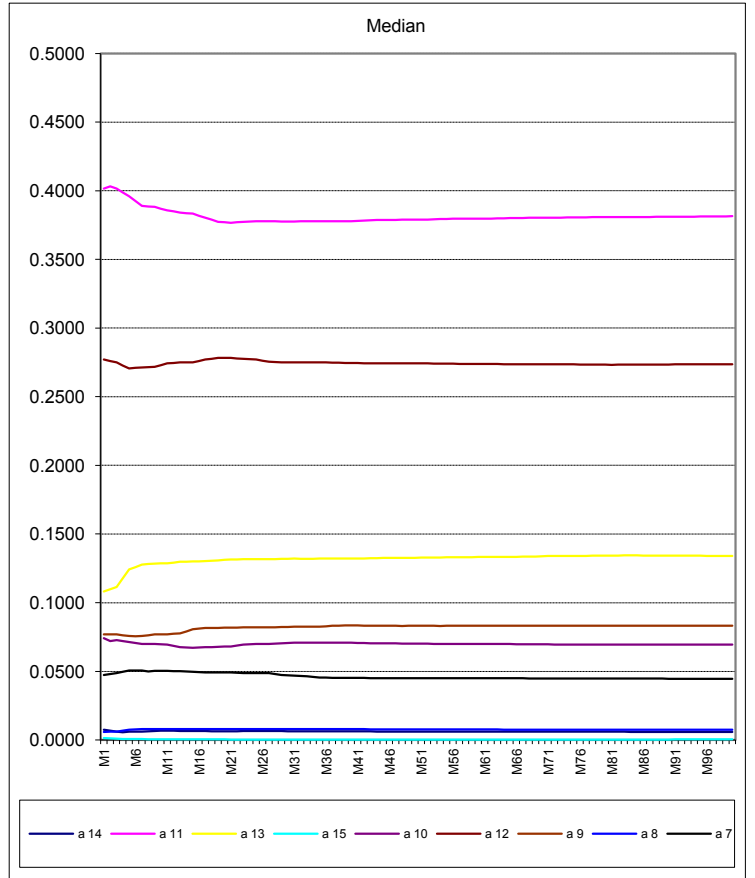
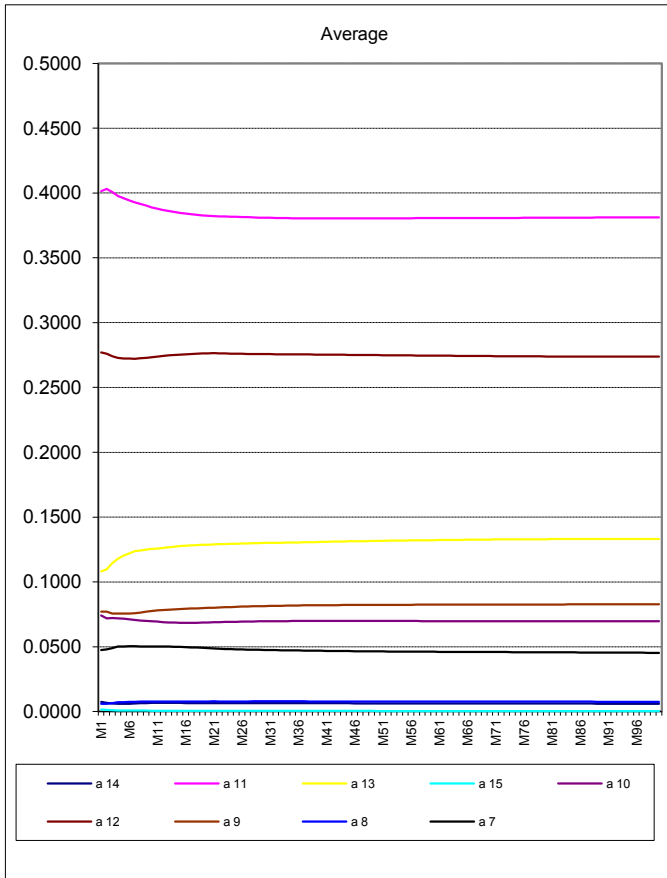
Sample number	24
Sample size	308

Average D	Min	Max
0.0009	0.0001	0.0019

Alleles	Average	min	max	Int min	Int Max	Median	S.D	IC
a7	0.0450	0.0349	0.0582	0.0442	0.0459	0.0454	0.0074	0.0008
a8	0.0077	0.0030	0.0119	0.0074	0.0080	0.0080	0.0025	0.0003
a9	0.0848	0.0698	0.1051	0.0837	0.0859	0.0822	0.0095	0.0011
a10	0.0718	0.0503	0.0887	0.0707	0.0729	0.0730	0.0098	0.0011
a11	0.3794	0.3413	0.4092	0.3775	0.3813	0.3794	0.0169	0.0019
a12	0.2727	0.2496	0.2993	0.2711	0.2743	0.2715	0.0145	0.0016
a13	0.1316	0.1156	0.1579	0.1303	0.1329	0.1283	0.0116	0.0013
a14	0.0067	0.0000	0.0119	0.0063	0.0070	0.0066	0.0032	0.0004
a15	0.0003	0.0000	0.0016	0.0002	0.0004	0.0000	0.0007	0.0001

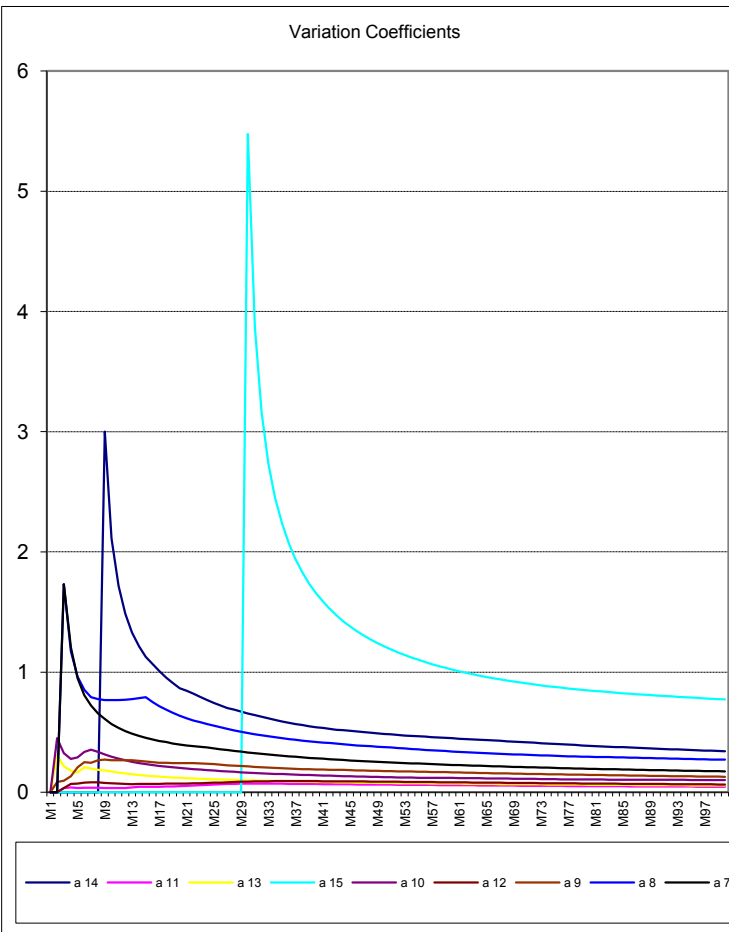
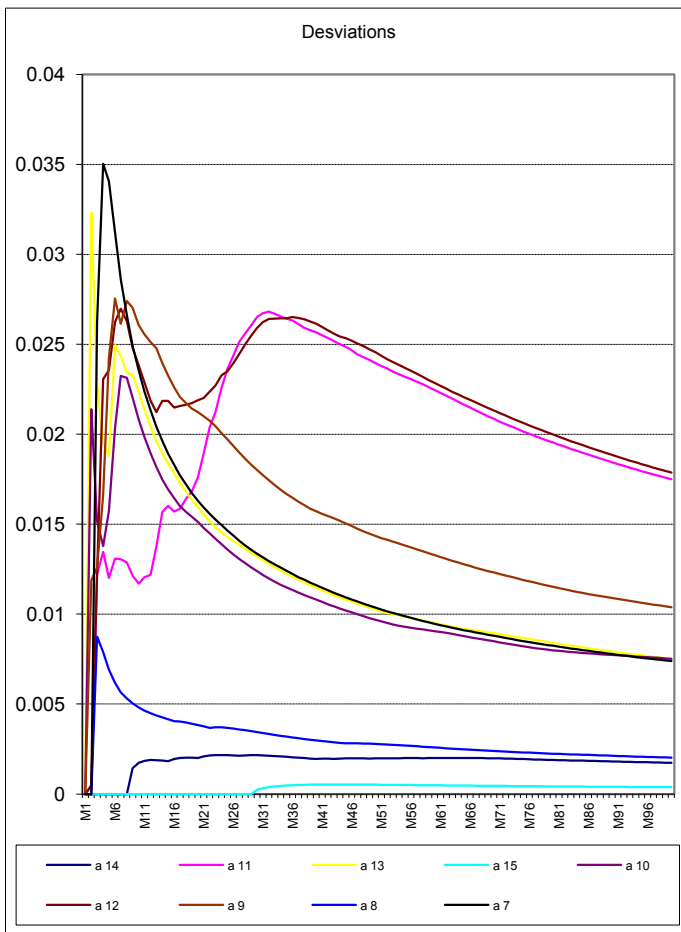
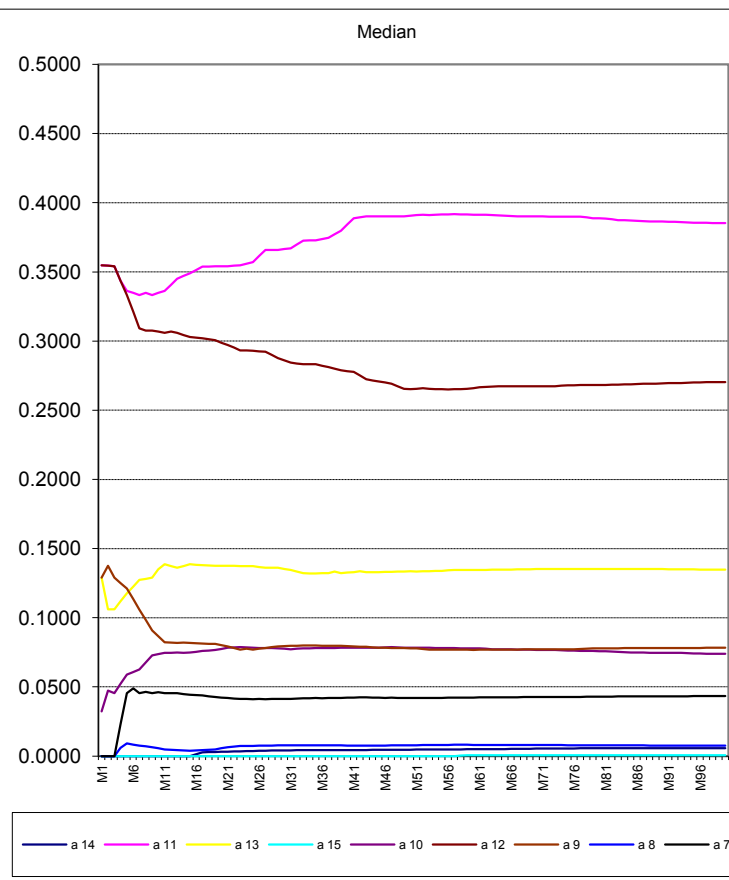
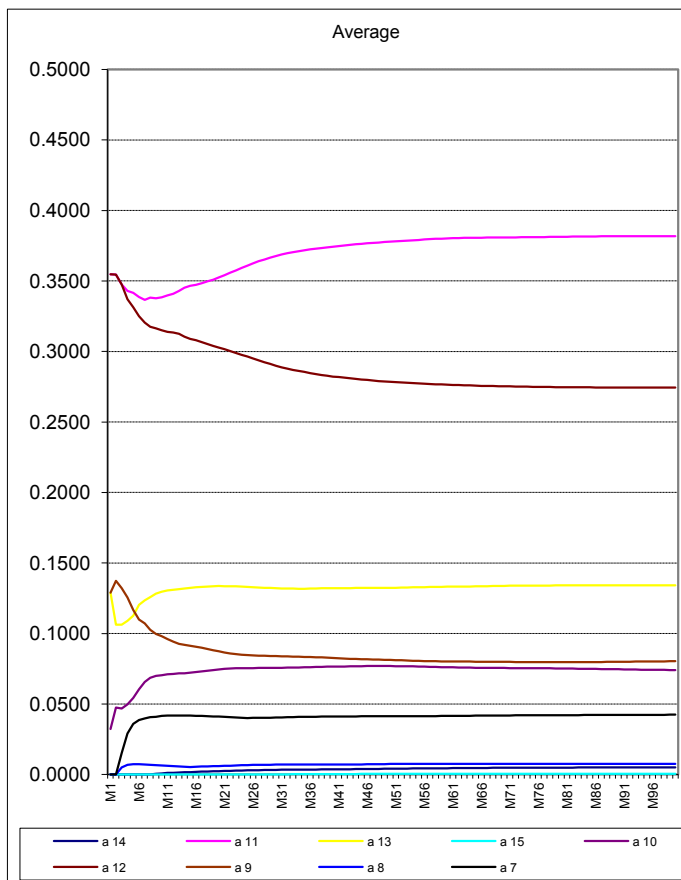
Graphic analysis of polymorphisms.

System D5S818 Pop. Bogotá



Graphic analysis of polymorphisms.

System D5S818 Pop. Bogotá



Graphic analysis of polymorphisms.

System	D5S818	Pop.	Bogotá
--------	--------	------	--------

W Statistics to Specific sample

Alleles	W f X [^]	W f md	W X [^] md
a 7	0.0004	0.0085	0.0088
a 8	0.0048	0.0438	0.0392
a 9	0.0001	0.0316	0.0316
a 10	0.0017	0.0174	0.0158
a 11	0.0005	0.0006	0.0002
a 12	0.0000	0.0044	0.0045
a 13	0.0002	0.0249	0.0251
a 14	0.0009	0.0150	0.0159
a 15	0.0080		

Indeterminate value (Division by zero)

0.0080 0.0438 0.0392

Sampling without repetition

