



UNIVERSIDAD REGIONAL AMAZÓNICA IKIAM

FACULTAD DE CIENCIAS DE LA VIDA

CARRERA DE INGENIERÍA EN BIOTECNOLOGÍA

**Estudio de genética poblacional: Frecuencias alélicas de marcadores STR
para haplotipos de Cromosoma Y en la población ecuatoriana.**

Proyecto de investigación previo a la obtención del Título de:
INGENIERA EN BIOTECNOLOGÍA

AUTORA: CAROLINA STEFANÍA ROMERO MEJÍA

TUTORA: Ph.D. CAROLINE NICOLE ORIANA BACQUET PÉREZ

Napo - Ecuador

2022

Declaración de derecho de autor, autenticidad y responsabilidad

Tena, 7 de abril de 2022

Yo, Carolina Stefanía Romero Mejía con documento de identidad N° 1726453200, declaro que los resultados obtenidos en la investigación que presento en este documento final, previo a la obtención del título de Ingeniero en Biotecnología son absolutamente inéditos, originales, auténticos y personales.

En virtud de lo cual, el contenido, criterios, opiniones, resultados, análisis, interpretaciones, conclusiones, recomendaciones y todos los demás aspectos vertidos en la presente investigación son de mi autoría y de mi absoluta responsabilidad. Por la favorable atención a la presente, suscribo de usted,

Por la favorable atención a la presente, suscribo de usted,

Atentamente

Firma:

Carolina Stefanía Romero Mejía

1726453200

Certificado de dirección de trabajo de integración curricular

Certifico que el trabajo de integración curricular titulado: “Estudio de genética poblacional: Frecuencias alélicas de marcadores STR para haplotipos de Cromosoma Y en la población ecuatoriana” en la modalidad de: proyecto de investigación en formato tesis, fue realizado por: Carolina Stefanía Romero Mejía, bajo mi dirección.

El mismo ha sido revisado en su totalidad y analizado por la herramienta de verificación de similitud de contenido; por lo tanto, cumple con los requisitos teóricos, científicos, técnicos, metodológicos y legales establecidos por la Universidad Regional Amazónica Ikiám, para su entrega y defensa.

Tena, 7 de abril de 2022



Firmado electrónicamente por:

CAROLINE NICOLE
ORIANA BACQUET
PEREZ

Caroline Nicole Bacquet Perez

CI: 1756516041

Agradecimientos

Agradezco al Servicio Nacional de Medicina Legal y Ciencias Forenses por la apertura brindada para el desarrollo de mi proyecto de tesis previo la obtención del título de Tercer Nivel.

A la gloriosa Universidad Regional Amazónica Ikiam.

A la PhD. Caroline Nicole Bacquet, a la TM Eugenia Osorio y al PhD. Willin Álvarez por la motivación, los consejos, las sugerencias y por ser mi guía en el desarrollo de este trabajo.

Dedicatoria

A mi abuelito Ángel y a mi abuelita Carmen por mostrarme que es la bondad, la empatía y el amor por la familia.

A mis padres Leoncio y Patricia que con su amor, confianza y esfuerzo me han permitido culminar una etapa más en mi vida, por ser el motor que me mueve y por mostrarme que lo malo no es caer sino no levantarse.

A mis hermanos Ana y Brandon por su cariño y apoyo durante todo este proceso. A Kleber por la confianza y ayuda. A toda mi familia porque siempre han estado pendientes de mí y de mi formación académica.

Finalmente, quiero dedicar esta tesis a todos mis amigos, por ser mi segunda familia y apoyarme siempre en momentos difíciles.

Resumen

Los microsatélites se aplican a la ciencia forense, puesto que cada ser humano tiene un patrón único de este tipo de marcadores. El cálculo de frecuencias alélicas de una población permite realizar un cálculo estadístico de verosimilitud de las coincidencias o compatibilidades observadas entre perfiles de ADN. Con este estudio se pretende generar una herramienta técnico-científica que permita dar significancia estadística a los análisis de haplotipo de cromosoma Y realizados en la población ecuatoriana. El análisis de frecuencias se realizó con el software de análisis de genética poblacional GenAlEx-6.5. Para esto no se consideraron los alelos intermedios. La diferencia por regiones se analizó en el software libre R versión 4.1.3. De los 968 perfiles genéticos analizados, se registraron 924 haplotipos de cromosoma Y con 20 marcadores Y-STR. La diversidad haplotípica es de 0.999, el poder de discriminación haplotípica es de 0.954, con su correspondiente probabilidad de coincidencia de 0.046. En el análisis intra-poblacional se encontró que no hay diferencias significativas entre regiones para la mayoría de los marcadores analizados, con excepción de DYS576, DYS389I, DYS389II, DYS19, DYS437, DYS392, DYS393. La información presentada fortalece los estudios e investigaciones previos sobre las frecuencias alélicas del Cromosoma Y en Ecuador y representa una herramienta aplicable al cálculo de frecuencias alélicas en la población ecuatoriana. Además, durante el proceso de desarrollo de este trabajo se generó la primera base de datos de haplotipos de Cromosoma Y del SNMLCF.

Palabras clave: Ecuador, STR, Cromosoma Y, frecuencias alélicas

Abstract

Microsatellites are applied to forensic science, since each human being has a unique pattern of this kind of markers. The calculation of allelic frequencies of a population allows a statistical calculation of the plausibility of the coincidences or compatibilities observed between DNA profiles. The aim of this study is to generate a technical-scientific tool to give statistical significance to the Y chromosome haplotype analysis performed in the Ecuadorian population. The frequency analysis was performed with the population genetics analysis software GenAEx-6.5. For this, intermediate alleles were not considered. Of the 968 genetic profiles analyzed, 924 Y-chromosome haplotypes with 20 Y-STR markers were recorded. The haplotypic diversity is 0.999, the haplotypic discriminatory power is 0.954, with a corresponding matching probability of 0.046. In the intra-population analysis it was found that there are no significant differences between regions for most of the markers analyzed, with the exception of DYS576, DYS389I, DYS389II, DYS19, DYS437, DYS392, DYS393. The information presented strengthens previous studies and research on Y chromosome allele frequencies in Ecuador and represents a tool applicable to the calculation of allele frequencies in the Ecuadorian population. In addition, during the development process of this work, the first SNMLCF Y chromosome haplotype database was generated.

Keywords: Ecuador, STR, Y Chromosome, allele frequencies

ÍNDICE GENERAL

| | |
|--|------------|
| Agradecimientos | iv |
| Dedicatoria | v |
| Resumen | vi |
| Abstract | vii |
| CAPÍTULO I: PLANTEAMIENTO DEL PROBLEMA DE INVESTIGACIÓN | 1 |
| 1.1. Antecedentes..... | 1 |
| 1.1.1 Genética..... | 1 |
| 1.1.1.1 Genética poblacional..... | 1 |
| 1.1.2 Genoma Humano | 2 |
| 1.1.3 ADN..... | 3 |
| 1.1.3.1 Cromosomas | 5 |
| 1.1.3.2 Polimorfismos del ADN | 7 |
| 1.1.4 Identificación humana..... | 7 |
| 1.1.5 Población ecuatoriana..... | 9 |
| 1.2 Planteamiento del problema..... | 10 |
| 1.3 Justificación de la investigación | 11 |
| 1.4 Objetivos de la investigación..... | 12 |
| CAPÍTULO II: MARCO METODOLÓGICO | 13 |
| 2.1 Recopilación de datos | 13 |
| 2.2 Tratamiento de Datos | 14 |
| 2.2.1 Descarte de parentesco en haplotipos duplicados | 14 |
| 2.2.2 Determinación de los marcadores para trabajar | 14 |

| | | |
|--|--|-----------|
| 2.3 | Índices genéticos, probabilidades específicas y análisis estadístico | 15 |
| 2.4 | Base de datos..... | 17 |
| CAPÍTULO III: PRESENTACIÓN DE DATOS Y RESULTADOS..... | | 18 |
| 3.1 | Análisis poblacional..... | 18 |
| 3.2 | Análisis intra-poblacional..... | 23 |
| CAPÍTULO IV: INTERPRETACIÓN Y DISCUSIÓN | | 46 |
| 4.1 | Análisis poblacional..... | 46 |
| 4.2 | Análisis intra-poblacional..... | 47 |
| CAPÍTULO V: CONCLUSIONES Y RECOMENDACIONES | | 49 |
| REFERENCIAS | | 50 |

ÍNDICE DE TABLAS

| | |
|---|----|
| Tabla 1. Marcadores STR de Cromosoma Y: rangos alélico y tamaños de productos de PCR. | 6 |
| Tabla 2.- Determinación de los marcadores con los que se trabajó. | 14 |
| Tabla 3. Frecuencia y porcentaje de haplotipos en la población ecuatoriana.. | 18 |
| Tabla 4. Frecuencias alélicas de la población ecuatoriana para los 20 marcadores Y-STR analizados.... | 19 |
| Tabla 5. Índices genéticos y probabilidades específicas de relevancia en el campo forense. | 22 |
| Tabla 6.- Matriz de población por pares de distancia genética Nei. | 43 |
| Tabla 7.- Rango y grupos resultantes del análisis de Kruskal-Wallis. | 44 |
| Tabla 8.- Alelos intermedios para los diferentes marcadores analizados. | 45 |

ÍNDICE DE FIGURAS

| | |
|---|----|
| Figura 1. Auto identificación étnica de la población ecuatoriana según censo 2001 y 2010..... | 9 |
| Figura 2. Gráficas del marcador DYS576 para las tres regiones de estudio. | 24 |
| Figura 3. Gráficas del marcador DYS389I para las tres regiones de estudio | 25 |
| Figura 4. Gráficas del marcador DYS448 para las tres regiones de estudio | 26 |
| Figura 5. Gráficas del marcador DYS389II para las tres regiones de estudio | 27 |
| Figura 6. Gráficas del marcador DYS19 para las tres regiones de estudio | 28 |
| Figura 7. Gráficas del marcador DYS391 para las tres regiones de estudio | 29 |
| Figura 8. Gráficas del marcador DYS481 para las tres regiones de estudio | 30 |
| Figura 9. Gráficas del marcador DYS533 para las tres regiones de estudio | 31 |
| Figura 10. Gráficas del marcador DYS438 para las tres regiones de estudio | 32 |
| Figura 11. Gráficas del marcador DYS437 para las tres regiones de estudio | 33 |
| Figura 12. Gráficas del marcador DYS570 para las tres regiones de estudio | 34 |
| Figura 13. Gráficas del marcador DYS635 para las tres regiones de estudio | 35 |
| Figura 14. Gráficas del marcador DYS390 para las tres regiones de estudio | 36 |
| Figura 15. Gráficas del marcador DYS439 para las tres regiones de estudio | 37 |
| Figura 16. Gráficas del marcador DYS392 para las tres regiones de estudio | 38 |
| Figura 17. Gráficas del marcador DYS393 para las tres regiones de estudio | 39 |
| Figura 18. Gráficas del marcador DYS458 para las tres regiones de estudio | 40 |
| Figura 19. Gráficas del marcador DYS456 para las tres regiones de estudio | 41 |
| Figura 20. Gráficas del marcador Y-GATA-H4 para las tres regiones de estudio..... | 42 |
| Figura 21. Frecuencia alélica del marcador DYS385 (a . b) para las tres regiones de estudio. | 43 |

CAPÍTULO I: PLANTEAMIENTO DEL PROBLEMA DE INVESTIGACIÓN

1.1. Antecedentes

“DNA testing is to justice what the telescope is for the stars; not a lesson in biochemistry, not a display of the wonders of a magnifying glass, but a way to see things as they really are”.

— Barry Scheck and Peter Neufeld, Actual Innocence

1.1.1 Genética

A lo largo de la historia, el ser humano ha sido testigo de la naturaleza heredable de los rasgos físicos, con lo que han practicado principios de genética desde sus inicios. La genética como ciencia no se delimita a un área específica, por el contrario, está relacionada con todas las áreas de la biotecnología, y algunas ramas del campo médico, ambiental e industrial [1]. Es importante en varios campos de interés humano, entre los cuales tenemos: la agricultura, que tuvo sus inicios cuando el ser humano aplicó las bases de la genética al cultivo para así obtener mejores frutos; la industria farmacéutica, pues la genética ha permitido la producción a gran escala de fármacos a través de organismos genéticamente modificados; y finalmente, el campo médico, donde ha permitido identificar enfermedades y desórdenes de componentes hereditarios[1,2].

La genética es la ciencia que estudia la variación y herencia biológica de los organismos, permitiendo determinar las diferencias entre las distintas especies y las diferencias entre los individuos de una misma especie. Se analiza en base individual, donde los rasgos son heredados de padres a hijos, o con una base global, a través del movimiento de marcadores genéticos en una población [2,3].

1.1.1.1 Genética poblacional

La genética poblacional, es la ciencia que estudia las fuentes y niveles de la variabilidad dentro y entre las poblaciones, explicando esto patrones a partir de las fuerzas evolutivas [4]. Tiene sus inicios en 1908 con Godfrey Harold Hardy y Wilhelm Weinberg, quienes describieron las frecuencias genotípicas de una población estática a partir de las frecuencias alélicas [5].

Las fuerzas que producen variabilidad genética en los humanos son: el flujo génico, se refiere a la transferencia de genes de una población a otra; la deriva génica, es el cambio al azar en las proporciones de los alelos; las mutaciones, que son cambios en el genoma y la selección natural, que implica que individuos que tienen los alelos ventajosos, dejan mayor descendencia, incrementando su proporción en las pozas génicas, mientras que los alelos en desventaja son removidos por la selección natural[4,5].

1.1.2 Genoma Humano

El desarrollo tecnológico de finales del siglo ha permitido que se lleven a cabo varios avances y descubrimientos científicos. El secuenciamiento del genoma humano es uno de los grandes logros de finales de siglo y permite un mejor conocimiento del material genético de la especie humana [6–8].

El Proyecto del Genoma Humano o PGH, es un proyecto que tiene por objetivo secuenciar y ensamblar la totalidad del material genético presente en los 23 cromosomas de la especie *Homo sapiens*; de los cuales 22 son autosómicos (numerados del 1 al 22) y 1 de carácter sexual (XX o XY). Tuvo sus inicios en 1990, como un proyecto internacional de cooperación pública y privada. El 26 de junio del 2000, tras 10 años de trabajo en el proyecto la compañía biotecnológica Celera y el Consorcio Internacional “Proyecto del Genoma Humano”, anuncian que la secuenciación y ensamblaje del genoma humano, se ha alcanzado en un 99%. En febrero del 2001 se publicó los resultados de este proyecto y el primer análisis del genoma humano en revistas Nature y Science [6,9,10].

Tras las grandes expectativas de la secuenciación del genoma humano, los científicos encargados del proyecto declararon estar sorprendidos con los resultados, pues no concordaban con las proyecciones realizadas. El llamado “Libro de la de Vida” está compuesto apenas por aproximadamente entre 20.000 y 25.000 genes. Mucho menos de lo expuesto en declaraciones previas en las que se hablaba de hasta 60 000 genes. Esta secuenciación también arrojó otros resultados sobre el ADN del ser humano. Del total de nuestro genoma el 80 % tiene funciones desconocidas [11].

El ADN, tiene regiones codificantes y no codificantes. Las regiones codificantes corresponden al 4% del genoma humano, son altamente conservadas y son las encargadas

de la producción de proteínas. Por otro lado, las regiones no codificantes, también denominadas ADN basura, corresponden aproximadamente al 96 % del genoma humano y son regiones generalmente variables, y contrario a lo que se pensaba inicialmente, estas regiones cumplen funciones importantes de regulación [2,12].

1.1.3 ADN

El ADN o ácido desoxirribonucleico, es una molécula formada por una doble hélice, que contiene la información genética de todos los seres vivos [13]. Cada hélice está constituida por una larga cadena de nucleótidos, cada uno a su vez formado por tres componentes: un grupo fosfato, una molécula de azúcar desoxirribosa y una base nitrogenada. Las bases nitrogenadas pueden ser de dos tipos según su estructura química: purinas, adenina y guanina y pirimidinas, citosina y timina; A, G, C y T, respectivamente [10].

Se diferencian en que la adenina se une a la timina formando dos enlaces de hidrogeno, mientras que la citosina se une a la guanina formando tres; son estos enlaces los que estabilizan el ADN. De esta manera, siempre se unen en pares A-T y C-G, generando cadenas complementarias [14]. Según su localización, encontramos dos tipos de ADN: nuclear y mitocondrial. El ADN nuclear (ADNn) se encuentra en el núcleo celular, es lineal, está distribuido en 23 cromosomas, una célula puede contener una sola copia del genoma nuclear y proporciona la mayoría de la información genética. Por otro lado, el ADN mitocondrial (ADNmt) se encuentra en las mitocondrias, es circular, está formado por 16.569 pares de bases de ADN, contiene 37 genes activos que codifican 13 proteínas, una célula puede contener miles de copias de su genoma mitocondrial y se hereda sólo de la línea materna [12,15].

ADN codificante

El ADN codificante se encarga de codificar para la síntesis de proteínas y es responsable parcialmente del fenotipo y de las características fisiológicas del individuo [12]. Representa el 4% del genoma humano, está constituido por aproximadamente 20 000 genes; con un promedio de 500 a 1 000 genes por cada cromosoma. El ADN codificante está conformado por secuencias internas no codificantes (intrones), que corresponden al 24% del genoma humano y las secuencias codificantes (exones), en una proporción de secuencias de 24:1,

respectivamente [16,17]. Tras el Proyecto del Genoma Humano, la identificación de genes codificantes es una de las áreas más estudiadas, puesto que permite realizar correlaciones con enfermedades y posibles tratamientos [6].

ADN no codificante

Es el ADN que no sintetiza proteínas, también llamado “ADN basura” debido a que previamente se creía que este no tenía función alguna, sin embargo, actualmente se sabe que este interviene en la regulación de la expresión génica, la estructura de los cromosomas y su segregación y en la división nuclear [16]. Representa el 96% del genoma humano. Según sus repeticiones pueden ser de dos tipos: de copia sencilla o de copia múltiple [17].

El ADN de copia única o no repetitivo, actúa como espaciador entre las regiones codificantes del genoma (intrones). Por otro lado, el ADN de copias múltiples, tiene secuencias repetitivas y existen dos clases: las secuencias repetidas en tándem y las secuencias repetidas intercaladas. Las secuencias repetidas intercaladas representan el 20% del genoma humano, son secuencias que forman unidades sencillas se encuentran intercaladas en diversos puntos del genoma. Y existen dos grupos: SINE, elementos nucleares cortos intercalados y LINES, elementos nucleares largos intercalados [18]. Las secuencias repetitivas en tándem representan del 5-10% del genoma humano, se caracterizan por tener una secuencia que se repite una tras otra, existen dos tipos: tipo I, compuesto por un bloque grande de diversas unidades que se repiten en tándem; y tipo II, que son bloques pequeños con una única unidad sencilla que se repite. Según su longitud encontramos tres grupos: satélite, minisatélites y microsatélites [16].

Las secuencias de ADN satélite pueden estar constituidos por unidades de repetición de una longitud similar a la de los minisatélites y microsatélites (1–30 pb) o mucho mayor (hasta 108 pb), y generalmente se ubican en la heterocromatina, principalmente en los centrómeros. Sin embargo, se diferencian en que los satélites no son tan variables en tamaño dentro de las poblaciones [19].

Las secuencias de ADN microsatélites y minisatélites son repeticiones de 1 a 10 pb y de 10 a 100 pb, y son denominadas secuencias de repeticiones cortas en tándem o STR (Short

Tandem Repeat) y secuencias de número variable de repeticiones en tándem o VNTR (Variable Number of Tandem Repeats), respectivamente. Se encuentran principalmente en grupos más grandes en regiones centroméricas y teloméricas, aunque también se encuentran dispersos a lo largo de todos los cromosomas [20,21]. Tanto microsatélites como minisatélites son altamente polimórficos, pero se estudian con finalidades diferentes. Mientras los minisatélites se aplican para la caracterización citogenética molecular de especies; los microsatélites se aplican a la ciencia forense, puesto que cada ser humano tiene un patrón único de microsatélites [16,20].

1.1.3.1 Cromosomas

Los cromosomas se ubican en el núcleo celular y son estructuras que se pueden encontrar en un super empaquetamiento de ADN durante la mitosis o descondensados durante la interfase [22].

Los cromosomas se encuentran formados por dos cromátidas hermanas, que se unen en un punto de constricción llamado centrómero; telómeros, que son tramos repetitivos de ADN ubicados en los extremos del cromosoma e impiden el desempaquetamiento del ADN, y las histonas que son proteínas en forma de hilo sobre las cuales se encuentra el ADN enrollado [23]. La estructura previamente descrita para los cromosomas ocurre sólo después de la fase S de la mitosis y hasta la metafase, después las cromátidas se separan y los cromosomas se descondensan. Es al conjunto de cromosomas característicos de una especie a lo que denominamos cariotipo.

El ser humano tiene 46 cromosomas, organizados en 23 pares; heredades de padre (50%) y madre (50%). El último par (23), también denominados cromosomas sexuales, determinan el sexo del individuo. Las mujeres tienen dos cromosomas X; mientras que, los hombres tienen un cromosoma X y un cromosoma Y [24].

El cromosoma Y humano es considerado un marcador de linaje (un haplotipo) que es transferido directamente de padre a hijo. Es por la naturaleza uniparental de este marcador que resulta útil en circunstancias como: agresión sexual; personas desaparecidas; identificación de víctimas de desastres; análisis de parentesco complejo; e inferencia de población [25,26]. En el análisis del Cromosoma Y se pueden lograr poderes de

discriminación bastante significativos cuando se usan muchos marcadores Y STR, los marcadores necesarios para formar el haplotipo mínimo son: DYS19, DYS385a/b, DYS389I, DYS389II, DYS390, DYS391, DYS392 y DYS393 [27].

Tabla 1. Marcadores STR de Cromosoma Y: rangos alélico y tamaños de productos de PCR. Los rangos de tamaño proporcionados se basan en GenBank. Secuencias de cebadores para la amplificación. Las temperaturas de melting (Tm) se calcularon utilizando una concentración total de cebador = 0,05 μ M y [Na+] = 50 μ M (modificada de [27–29]).

| STR LOCUS | RANGO ALÉLICO | RANGO DE TAMAÑO [PB] | SECUENCIA DE CEBADORES | | | | | |
|-----------|---------------|----------------------|----------------------------------|----------|---------|-----------------------------------|----------|---------|
| | | | Forward | μ M] | Tm (oC) | Reverse | μ M] | Tm (oC) |
| DYS 19 | 10-19 | 233-269 | ACTACTGAGTTTCT GTTATAGTGTTTTT | 1.8 | 55 | GTCAATCTCT GCACCTGGAA AT | 1.8 | 60.5 |
| DYS 385 | 7-23 | 242-306 | AGCATGGGTGACA GAGCTA | 0.6 | 56.9 | GCCAATTACA TAGTCCTCCT TTC | 0.6 | 54.7 |
| DYS 388 | 10-18 | 151-175 | GAATTCATGTGAGT TAGCCGTTTAGC | 1.8 | 63.4 | GAGGCGGAG CTTTTAGTGA G | 1.8 | 59.2 |
| DYS 389I | 9-17 | 143-175 | CCAACTCTCATCTG TATTATCTATG | 1.3 | 54.2 | GTTATCCCTG AGTAGTAGA AGAATG | 1.3 | 59 |
| DYS 389II | 26-34 | 263-295 | CCAACTCTCATCTG TATTATCTATG | 1.3 | 54.2 | GTTATCCCTG AGTAGTAGA AGAATG | 1.3 | 59 |
| DYS 390 | 17-28 | 189-295 | TATATTTTACACATT TTGGGCC | 0.2 | 57.2 | GTGACAGTAA AATGAAAACA TTGC | 0.2 | 57 |
| DYS 391 | 7-14 | 93-121 | TTCAATCATAACACC CATATCTGTC | 0.2 | 57.9 | GATAGAGGG ATAGGTAGG CAGGC | 0.2 | 60 |
| DYS 392 | 6-16 | 290-320 | TAGAGGCAGTCATC GCAGTG | 1.8 | 60.2 | GACCTACCAA TCCCATTCTT | 1.8 | 57.3 |
| DYS 393 | 9-16 | 109-133 | GTGGTCTTCTACTT GTGTCAATAC | 0.4 | 54.7 | GAATCAAGT CCAAAAATG AGG | 0.4 | 57.7 |
| DYS 426 | 10-12 | 92-98 | CTCAAAGTATGAAA GCATGACCA | 0.2 | 59.3 | GGTGACAAG ACGAGACTTT GTG | 0.2 | 59.8 |
| DYS 437 | 14-17 | 186-198 | GACTATGGGCGTG AGTGCAT | 0.2 | 61.1 | GAGACCCTGT CATTACAGA TGA | 0.2 | 59.6 |
| DYS 438 | 6-13 | 300-335 | CCAAAATTAGTGG GGAATAGTTG | 0.2 | 58.9 | GATCACCCAG GGTCTGGAGT T | 0.2 | 62.6 |

| | | | | | | | | |
|----------------|-------|---------|------------------------------|-----|------|-------------------------------------|-----|------|
| DYS 439 | 16-24 | 210-230 | TCGAGTTGTTATGG TTTTAGGTCT | 0.2 | 58.3 | GTGGCTTGGA ATTCTTTTACC C | 0.2 | 60.3 |
| DYS 447 | 22-29 | 207-242 | GGTCACAGCATGG CTTGTT | 0.7 | 63.4 | GGGCTTGCTT TGCATTATCT CT | 0.7 | 64 |
| DYS 448 | 20-26 | 299-335 | TGGGAGAGGCAAG GATCCAA | 1.1 | 65.2 | GTCATATTTTC TGGCCGGTCT GG | 1.1 | 64.3 |
| DYS 460 | 7-12 | 101-121 | GAGGAATCTGACA CCTCTGACA | 0.7 | 59.3 | GTCCATATCA TCTATCCTCT GCCTA | 0.7 | 59.1 |
| H4 | 8-13 | 122-142 | ATGCTGAGGAGAA TTTCCAA | 0.4 | 57.3 | GCTATTCATC CATCTAATCT ATCCATT | 0.4 | 56.2 |
| YCAII | 11-24 | 135-161 | TGTCAAAATTTAAC CCACAATCA | 0.9 | 59.3 | GCAGTCTTTC ACCATAAGGT TAGC | 0.9 | 57.2 |

1.1.3.2 Polimorfismos del ADN

Los polimorfismos del ADN son variaciones en la secuencia de ADN presentes en al menos el 1% de una población y pueden ocurrir tanto en regiones codificantes como en regiones no codificantes [30].

Si bien algunos polimorfismos del ADN humano se basan en inserciones, eliminaciones y reordenamientos de secuencias; la mayoría implican variaciones de un solo nucleótido, denominado polimorfismo de un solo nucleótido o SNP (Single-nucleotide polymorphism), o en variaciones en el número de repeticiones en tándem. Los polimorfismos de un solo nucleótido son muy abundantes en el genoma humano y ocurren una vez cada 200-500 pb; de igual manera, las variaciones en el número de repeticiones en tándem son abundantes con al menos decenas de miles de sitios polimórficos. El estudio de los polimorfismos de ADN ha permitido la construcción de mapas de ligamiento y el mapeo de genes de enfermedades genéticas [31,32].

1.1.4 Identificación humana

El ser humano desde siempre ha tenido la necesidad de tener una identidad, algo que lo represente y haga único. A lo largo de la historia, se han utilizado varios métodos para identificar a las personas relacionadas con crímenes, logrando atribuir y responsabilizar a un individuo de un delito [33]. Las primeras técnicas utilizadas consistían en un marcaje

que permitía reconocer reincidentes, como las marcas de fuego en partes visibles del cuerpo, se seccionaban partes del cuerpo o se aplicaban tatuajes [34]. Al mismo tiempo, se desarrollaron métodos como la descripción personal y reconocimiento, el retrato hablado, fotografía, desarrollo del sistema antropométrico, la dactiloscopia, radiología y odontología forense; hasta llegar a la genética forense [33,34].

La genética forense empieza con el estudio de grupos sanguíneos. El primer grupo estudiado fueron los antígenos del sistema ABO, que se encuentran en la superficie de los glóbulos rojos, mientras que los anticuerpos correspondientes se encuentran en el plasma (inmunoglobulinas); con una probabilidad de exclusión de aproximadamente el 75%. Luego se desarrolló el análisis de las proteínas plasmáticas, enzimas eritrocitarias y de los antígenos de histocompatibilidad (HLA), con una probabilidad de exclusión cercana al 71, 61 y 95% respectivamente, hasta llegar al estudio de la huella genética [34].

DNA fingerprinting o huella genética es un método de identificación humana encontrado casualmente por Jeffreys y colaboradores en 1984 al estudiar el gen de la globina; ellos notaron una secuencia de 33 pb con un patrón repetitivo en las muestras [35]. Nos referimos como huella genética al conjunto de polimorfismos propio de cada persona, únicos e invariables a lo largo de la vida. La huella genética se utilizó por primera vez con fines legales apenas un año después, en 1985 para resolver un caso de inmigración [35,36]. En 1987 ya se admitió como prueba de base científica en países como Inglaterra y EEUU [37]. Los marcadores genéticos se encuentran en el ADN no codificante, también se denominan polimorfismos debido a su alta variabilidad. Un perfil genético es el resultado del análisis de varios marcadores genéticos, generando combinaciones únicas (en caso de seres humanos no relacionados) que nos permiten identificar a los individuos [38]. La técnica utilizada inicialmente para tipificar polimorfismos de ADN fue polimorfismo de longitud de fragmentos de restricción (RFLP), el análisis se realizaba por transferencia Southern. Esta técnica presentaba algunas complicaciones en el ámbito forense pues su aplicabilidad requiere ADN en grandes cantidades y de buena calidad. Actualmente, se utilizan marcadores de repeticiones cortas en tándem (STR), puesto que se requieren cantidades menores de ADN y tienen buena susceptibilidad de amplificación por PCR [35,38].

1.1.5 Población ecuatoriana

La población ecuatoriana actual es de aproximadamente 18 millones, y tiene una tasa de crecimiento menor a la registrada en décadas pasadas. Entre los censos de 1950 a 1962 se calcula una tasa de crecimiento de 2.96 %, mientras que, entre los censos del 2001 y 2010 fue de 1.95% [39,40]. Los censos del 2001 y 2010 sobre la auto identificación étnica de la población arrojaron que al menos un 70% de los habitantes de Ecuador se autodefinen como mestizos. El resto de categorías no superan en 7.5% cada una: montubios 7.4%, indígenas 7%, afro ecuatorianos 7.2%, blancos 6,1 % y otros con 0,3% [39]. Las variaciones en las categorías presentadas se exponen en el gráfico.

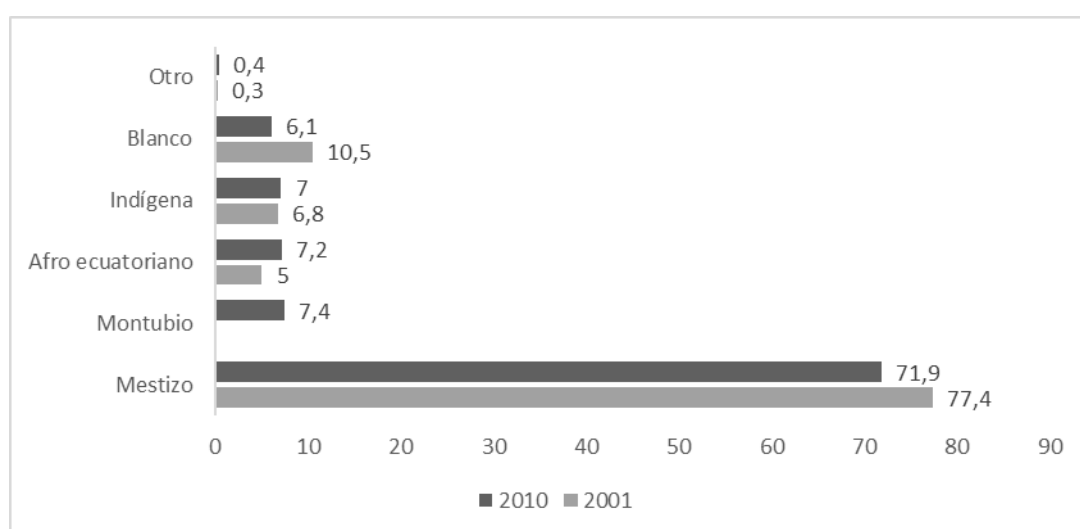


Figura 1. Auto identificación étnica de la población ecuatoriana según censo 2001 y 2010.

Fuente: INEC, Censo de Población y Vivienda 2001 y 2010.

Elaboración: Autoría propia.

El Centro de Investigación de Genética y Genómica (CIGG) UTE, mediante una investigación sobre el origen de los ecuatorianos, en la que se analizaron tres huellas génicas de ADN: ADN nuclear, mitocondrial (heredado por línea materna) y del cromosoma Y (heredado por línea paterna), confirmó que la población ecuatoriana mayoritariamente es mestiza. Se utilizaron 46 AIM-InDels (Ancestry Informative Insertion/Deletion Markers). Los resultados de esta investigación demostraron que los mestizos tienen 61% de proporción de ascendencia indígena, 32% caucásica, y 7% africana. Mientras que los afrodescendientes tienen 65% de africana, 23% de caucásico y 12% de indígena. Finalmente, los

indoamericanos (indígenas) tienen una proporción de ascendencia de 90% indígenas, 7% caucásico y 3% africanos [41,42]. Lo que confirma los datos censales del INEC, presentados previamente.

La composición de la población ecuatoriana también se ve afectada por la entrada y salida de personas al país (migración y emigración), el INEC informa que para el año 2018 existía un crecimiento de estos dos tipos de movimientos. Las entradas se incrementaron en un 376,5% y las salidas en un 418,9% desde el año de 1997 hasta el 2018. Sin embargo, para el año 2020, por la pandemia el flujo migratorio general registró un decrecimiento del 71% con respecto al año 2019. La población inmigrante en Ecuador proviene principalmente de Venezuela, Colombia, Estados Unidos y Perú; mientras que los países destino que más reciben ecuatorianos son Estados Unidos, Perú Colombia y México [43–45].

1.2 Planteamiento del problema

Los marcadores STR han sido ampliamente utilizados en la genética forense por ser un método sensible y altamente discriminatorio [46]. Han permitido la resolución de crímenes que requieren identificación de personas o aquellos en los que están involucrados vestigios biológicos como muestras de sangre, semen, pelos, saliva, tejidos o restos óseos [47]. Los STR del cromosoma Y no se recombinan durante la meiosis, y por lo tanto siguen las reglas de la herencia mendeliana simple, es decir que el haplotipo (combinación de estados alélicos de marcadores) se transmite patrilinealmente, dicho en otras palabras, de generación en generación en los integrantes del sexo masculino de una familia [48,49].

En Ecuador, la mayoría de estudios realizados de frecuencias alélica de cromosoma Y, se refieren a una parte específica de la población: mestizos, kichwas, afro ecuatorianos waoranis, todos con una población de estudio que no supera los 550 individuos [50]. Existe un único estudio enfocado la caracterización y análisis de haplotipos de 11 loci Y-STR que se refiere a toda la población ecuatoriana, con un muestreo de 206 individuos [51].

Desde la mitad de la década de 1990, en países como Estados Unidos las bases de datos informáticas de perfiles genéticos de muestras de escenas de crimen han hecho posible conectar delitos con delincuentes. Estas bases de datos recolectan información del perfil genético de individuos condenados, en procesos legales e incluso de aquellos arrestados por un delito [52]. Esta tecnología permite que miles de crímenes sean resueltos, sobre

todo los cometidos por delincuentes reincidentes [52,53]. En Ecuador, la ausencia de una base de datos de perfiles genéticos, estudios o investigaciones sobre las frecuencias alélicas del cromosoma Y representan una debilidad en el Sistema de Justicia Nacional.

1.3 Justificación de la investigación

Los marcadores STR aplicados al Cromosoma Y tienen aplicaciones especializadas. Al identificar solo el componente masculino en una mezcla, se obtiene directamente el haplotipo característico del individuo, sin necesidad de realizar diferenciación de células masculinas y femeninas [49]. Son útiles en casos de violencia sexual ya que permiten identificar ADN masculino en mezclas de fluidos corporales, identificando uno o varios individuos. Lo que permite determinar la cantidad de personas aportantes en una mezcla [26].

Permiten identificar a los individuos en casos de filiación paterna, ya que se transmiten en bloque de padre a hijo, representa una ventaja en el caso de medio hermanos por línea materna y de ausencia del presunto padre [54]. Además, se aplican ampliamente en el caso de desastres masivos, donde el estado de los restos mortales no permite determinar la cuarteta básica de cada individuo (estatura, edad, raza y sexo), permitiendo correlación de grupos familiares [55].

La aplicación de marcadores del cromosoma Y en el campo forense requiere de la identificación de loci polimórficos, el desarrollo de métodos de tipificación y de la construcción de bases de datos específicas de cada población [56,57]. Se debe considerar además que el resultado de este análisis es el perfil genético de un individuo que es parte de una población y que una correcta interpretación de este perfil genético está sujeta a estudios de genética poblacional [49].

El cálculo de frecuencias alélicas de una población permite realizar un cálculo estadístico de verosimilitud de las coincidencias o compatibilidades observadas entre perfiles de ADN [58]. Con este estudio se pretende generar una herramienta técnico-científica que permita dar significancia estadística a los análisis de haplotipo de cromosoma Y realizados en la población ecuatoriana.

1.4 Objetivos de la investigación

General

- Determinar la frecuencia alélica para marcadores STRs del haplotipo de cromosoma Y de la población ecuatoriana para fines de identificación humana en genética forense.

Específicos

- Desarrollar la primera base de datos del Servicio Nacional de Ciencias Forenses (SNMLCF) de marcadores SRT de Cromosoma Y.
- Determinar los alelos más y menos frecuentes en la población ecuatoriana.
- Determinar parámetros de estadística forense y genética que permitan una mejor presentación de los resultados.

CAPÍTULO II: MARCO METODOLÓGICO

Para el desarrollo de este estudio, primero se entregó un proyecto de tesis al Servicio Nacional de Medicina Legal y Ciencias Forenses (SNMLCF). La aprobación del mismo se llevó a cabo el 12 de febrero de 2021.

Realizado por el SNMLCF

Obtención de los perfiles genéticos

Todos los perfiles genéticos se obtuvieron de muestras referencia (directamente obtenidas de individuos y no de indicios criminales). La recolección de material biológico (sangre) se realizó en tarjetas FTA. Se inició con la lisis celular que requiere se perfora con un punch la mancha en la tarjeta FTA, obteniendo un círculo de hasta 1 x 0,5 pulgadas que contiene aproximadamente 50 μ L de sangre y se coloca en un tubo de PCR de 2 ml. Se agregó 500 μ L de lisis (LBB) y 30 μ L de Proteinasa K y se sometió a agitación durante 5 min a velocidad moderada. La extracción de ADN se realizó siguiendo el protocolo de la química PrepFiler *Express*[™] [59] para el sistema de extracción de ADN forense AutoMate *Express*[™]. La amplificación de ADN se realizó siguiendo el protocolo de la química PowerPlex[®] Y23 de Promega Corporation [60] y de la química Yfiler[™] Plus de Thermo Fisher Scientific [61], para finalmente pasar al analizador genético HITACHI ABI PRISM 3130xl de Applied Biosystems [62].

Realizado por la autora

2.1 Recopilación de datos

La recopilación de datos se realizó en un formato de base de datos para perfiles genéticos previamente diseñado por la tesista y aprobado por la tutora PhD. Caroline Bacquet. El tiempo a tratar corresponde a los años 2014-2020. La recolección de la información para la base de datos se inició el 9 de agosto del 2021 hasta el 10 de septiembre del 2021 en Quito y se culminó en Guayaquil desde el 13 de septiembre del 2021 hasta el 18 de septiembre del 2021. Se recopilaron 1000 perfiles genéticos, con lo que se procedió al tratamiento de datos y el análisis estadístico de genética poblacional.

2.2 Tratamiento de Datos

2.2.1 Descarte de parentesco en haplotipos duplicados

Los haplotipos duplicados se encontraron por programación VBA en Excel 2016, y se descartó parentesco por apellidos en el archivo físico del SNMLCF, de los 90 perfiles duplicados 32 fueron eliminados por relación de parentesco, y 58 se conservaron. Se trabajó con 968 perfiles.

2.2.2 Determinación de los marcadores para trabajar

El SNMLCF tanto en Quito como Guayaquil trabaja con dos químicas para la obtención de haplotipo de Cromosoma Y, se usa Yfiler™ Plus de Applied Biosystems y PowerPlex® Y23 System de Promega Corporation, cada uno con marcadores diferentes. Se determinó que se trabajará con los 20 marcadores compartidos por las dos químicas, **Tabla 2**.

Tabla 2. Marcadores de Yfiler™ Plus de Applied Biosystems y PowerPlex® Y23 System de Promega Corporation, determinación de los marcadores con los que se trabajó.

| MARCADORES Y-STR | KIT COMERCIAL Y-STR | | A TRATAR |
|---------------------|-----------------------------|---------------------------|-------------|
| | PowerPlex® Y23 ^a | Yfiler™ Plus ^b | |
| DYS19 | ✓ | ✓ | ✓ |
| DYS385A/B | ✓ | ✓ | ✓ |
| DYS389I | ✓ | ✓ | ✓ |
| DYS389II | ✓ | ✓ | ✓ |
| DYS390 | ✓ | ✓ | ✓ |
| DYS391 | ✓ | ✓ | ✓ |
| DYS392 | ✓ | ✓ | ✓ |
| DYS393 | ✓ | ✓ | ✓ |
| DYS437 | ✓ | ✓ | ✓ |
| DYS438 | ✓ | ✓ | ✓ |
| DYS439 | ✓ | ✓ | ✓ |
| DYS448 | ✓ | ✓ | ✓ |
| DYS456 | ✓ | ✓ | ✓ |
| DYS458 | ✓ | ✓ | ✓ |
| DYS635 | ✓ | ✓ | ✓ |
| Y-GATA-H4 | ✓ | ✓ | ✓ |
| DYS481 | ✓ | ✓ | ✓ |
| DYS533 | ✓ | ✓ | ✓ |
| DYS549 | ✓ | | |
| DYS570 | ✓ | ✓ | ✓ |

| | | | |
|----------------------------|---|---|---|
| DYS576 | ✓ | ✓ | ✓ |
| DYS643 | ✓ | | |
| DYS449 | | ✓ | |
| DYS460 | | ✓ | |
| DYS518 | | ✓ | |
| DYS627 | | ✓ | |
| DYF387S1A/B | | ✓ | |
| a=PROMEGA | | | |
| b=THERMO FISHER SCIENTIFIC | | | |

2.3 Índices genéticos, probabilidades específicas y análisis estadístico

El análisis estadístico para frecuencias se realizó con el software de análisis genético GenAlEx-6.5 [63,64]. Para esto no se consideraron los alelos intermedios. El marcador DYS385 a/b, se consideró como alelo único para el cálculo de frecuencias y se excluyó para el análisis de similitud entre las poblaciones (índice de Nei). La diferencia por regiones se analizó en el software libre R versión 4.1.3 a través de la prueba de Kruskal-Wallis.

- **Frecuencias Alélicas:** Es una probabilidad de encontrar un alelo determinado respecto a una muestra, esta refleja una relación entre el número de alelos observados y el número total de alelos. Este valor de puede representar por unidad, véase **Ecuación 1**, o en porcentaje, véase **Ecuación 2**.

$$f(A) = \frac{\text{Alelos observados}}{\text{Total de alelos}}$$

Ecuación 1. Ecuación para el cálculo de frecuencias alélicas

$$f\%(A) = \frac{\text{Alelos observados}}{\text{Total de alelos}} \times 100$$

Ecuación 2. Ecuación para el cálculo del porcentaje de frecuencias alélicas

- **Diversidad génica:** Al ser haploide el Cromosoma Y, los valores de diversidad génica y diversidad haplotípica se calculan con la misma ecuación, véase **Ecuación 3 y 4**.
Diversidad génica

$$DG = \frac{\frac{N}{N-1}}{(1 - \sum xi^2)}$$

Ecuación 3. Ecuación para el cálculo de la diversidad génica. Donde: N es el tamaño de la muestra y xi es la frecuencia del ith alelo.

Diversidad haplotípica

$$DH = \frac{\frac{N}{N-1}}{(1 - \sum xi^2)}$$

Ecuación 4. Ecuación para el cálculo de la diversidad génica. Donde: N es el tamaño de la muestra y xi es la frecuencia del ith haplotipo.

- **Poder de discriminación por haplotipo (PDh):** Se refiere a la probabilidad de seleccionar dos individuos al azar de la población estudiada y que sus haplotipos sean diferentes, se calcula al dividir el número de haplotipos observados sobre el número de individuos analizados.

$$PDh = \frac{\text{Haplotipos observados}}{\text{Muestras}}$$

Ecuación 5.- Ecuación para el cálculo del poder de discriminación.

- **Poder de discriminación por locus (PD):** Se refiere a la probabilidad de seleccionar dos individuos al azar de la población estudiada y que sus alelos del mismo locus sean diferentes, se calcula al restar la suma de las frecuencias alélicas de la población al cuadrado de la unidad.

$$PDg = (1 - \sum xi^2)$$

Ecuación 6.- Ecuación para el cálculo del poder de discriminación por locus. Donde xi es la frecuencia del ith alelo.

- **Probabilidad de coincidencia (PM):** Representa la probabilidad de que dos personas tomadas al azar posean el mismo haplotipo, esta probabilidad es descrita para la población de estudio.

$$PM = 1 - PD$$

Ecuación 7.- Ecuación para el cálculo de la probabilidad de coincidencia. Donde PD es el poder de discriminación.

- **Índice de Shannon (I):** Es una forma de medir la diversidad de especies en una comunidad.

$$I = (-1) * (\sum xi * \ln(xi))$$

Ecuación 8.- Ecuación para el cálculo del índice de Shannon. Donde xi es la frecuencia del ith alelo.

- **Kruskall-Wallis:** Es un método de estadística no paramétrica, que permite probar si diferentes grupos de datos provienen de la misma población.

2.4 Base de datos

La base de datos resultado de este estudio se entregó al SNMLCF en digital en formato Excel para facilidad de manipulación, y en formato MySQL.

CAPÍTULO III: PRESENTACIÓN DE DATOS Y RESULTADOS

3.1 Análisis poblacional

De los 968 perfiles genéticos analizados, se registraron 924 haplotipos de Cromosoma Y con 20 marcadores Y-STR. De los cuales, 890 son haplotipos únicos (96.30%), 25 duplicados (2.70%), 8 triplicados (0.90%) y 1 cuadruplicado (0.10%). La diversidad haplotípica es de 0.999, el poder de discriminación haplotípica es de 0.954, con su correspondiente probabilidad de coincidencia de 0.046; véase **Tabla 3**.

Tabla 3. Frecuencia y porcentaje de haplotipos en la población ecuatoriana. Donde: DH es la diversidad haplotípica, PD es el poder de discriminación y PM la probabilidad de coincidencia.

| HAPLOTIPOS | N | FRECUENCIA | PORCENTAJE (%) |
|---------------|-------|------------|----------------|
| ÚNICOS | 890 | 0,963 | 96,30 |
| DUPLICADOS | 25 | 0,027 | 2,70 |
| TRIPLICADOS | 8 | 0,009 | 0,90 |
| CUADRUPLICADO | 1 | 0,001 | 0,10 |
| TOTAL | 924 | 1.00 | 100 |
| DH | 0.999 | | |
| PD | 0.954 | | |
| PM | 0.046 | | |

La **Tabla 4**, presenta las frecuencias alélicas de la población ecuatoriana para los 20 marcadores Y-STR analizados. La población ecuatoriana presenta una escalera alélica conformada por 28 alelos, del 7 al 34. De estos, los alelos más frecuentes para los distintos marcadores son: 11, 12,13 y 14. Por otro lado, los alelos menos frecuentes son: 7, 8, 33 y 34.

Tabla 4. Frecuencias alélicas de la población ecuatoriana para los 20 marcadores Y-STR analizados. Donde: A= alelo, M= marcador, Frec = frecuencia.

| A / M | DYS 576 | DYS 389I | DYS 448 | DYS 389II | DYS 19 | DYS 391 | DYS 481 | DYS 533 | DYS 438 | DYS 437 | DYS 570 | DYS 635 | DYS 390 | DYS 439 | DYS 392 | DYS 393 | DYS 458 | DYS 456 | Y-GATA-H4 | DYS 385 | | |
|-----------|---------|----------|---------|-----------|--------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|-----------|----------------|--------------|-------|
| N | 955 | 965 | 965 | 965 | 966 | 966 | 956 | 953 | 967 | 966 | 929 | 967 | 967 | 964 | 935 | 940 | 912 | 938 | 940 | 938 | | |
| 7 | | | | | | 0,002 | | | | | | | | | 0,001 | | | | | A (a,b) | Frec | |
| 8 | | | | | | 0,003 | | | 0,011 | | | | | 0,001 | | | | | | | 10,11 | 0,001 |
| 9 | | | | | | 0,064 | | 0,015 | 0,062 | | | | | 0,008 | 0,001 | 0,001 | | | | 0,009 | 10,14 | 0,001 |
| 10 | 0,001 | 0,003 | | | | 0,610 | | 0,040 | 0,172 | | | | | 0,028 | 0,005 | | | | | 0,031 | 10,15 | 0,002 |
| 11 | | 0,001 | | | 0,001 | 0,303 | | 0,320 | 0,440 | | 0,001 | | | 0,256 | 0,230 | 0,003 | | | | 0,452 | 10,17 | 0,001 |
| 12 | | 0,151 | | | 0,016 | 0,017 | | 0,554 | 0,300 | | 0,001 | | | 0,508 | 0,029 | 0,113 | 0,001 | 0,001 | 0,447 | | 11 | 0,005 |
| 13 | 0,006 | 0,620 | | | 0,410 | | | 0,068 | 0,014 | 0,026 | 0,001 | | | 0,176 | 0,310 | 0,651 | 0,007 | 0,009 | 0,056 | | 11,12 | 0,003 |
| 14 | 0,021 | 0,221 | | | 0,380 | | | 0,002 | 0,001 | 0,686 | 0,014 | | | 0,021 | 0,338 | 0,177 | 0,032 | 0,078 | 0,004 | | 11,13 | 0,021 |
| 15 | 0,040 | 0,004 | | | 0,118 | 0,001 | | | | 0,245 | 0,042 | | | 0,001 | 0,066 | 0,054 | 0,175 | 0,529 | | | 11,14 | 0,181 |
| 16 | 0,107 | | | | 0,049 | | | | | 0,039 | 0,140 | | | | 0,017 | 0,001 | 0,289 | 0,281 | 0,001 | | 11,15 | 0,031 |
| 17 | 0,191 | | 0,002 | | 0,026 | | 0,001 | | | 0,003 | 0,370 | 0,001 | | | 0,002 | | 0,310 | 0,082 | | | 11,18 | 0,006 |
| 18 | 0,320 | | 0,089 | | 0,001 | | | 0,001 | | | 0,230 | 0,001 | | | | | 0,139 | 0,014 | | | 12 | 0,013 |
| 19 | 0,186 | | 0,366 | | | | | | | | 0,130 | 0,001 | 0,001 | | | | 0,041 | 0,003 | | | 12,13 | 0,003 |
| 20 | 0,098 | | 0,384 | | | | 0,013 | | | | 0,033 | 0,030 | 0,005 | | | | 0,002 | 0,003 | | | 12,14 | 0,026 |
| 21 | 0,027 | | 0,148 | | | | 0,050 | | | | 0,011 | 0,163 | 0,050 | | | | 0,003 | | | | 12,15 | 0,007 |
| 22 | 0,001 | | 0,010 | | | | 0,256 | | | | 0,016 | 0,436 | 0,060 | | | | | | | | 12,16 | 0,005 |
| 23 | 0,001 | | | | | | 0,130 | | | | 0,009 | 0,312 | 0,247 | | | | | | | | 12,17 | 0,021 |
| 24 | | | | | | | 0,236 | | | | | 0,046 | 0,515 | | | | | | | | 12,18 | 0,012 |
| 25 | | | | | | | 0,179 | | | | 0,001 | 0,006 | 0,105 | | | | | | | | 12,19 | 0,009 |
| 26 | | | | 0,001 | | | 0,066 | | | | | 0,003 | 0,016 | | | | | | | | 12,20 | 0,004 |

| | | | | | |
|-----------|-------|-------|-------|--------------|-------|
| 27 | 0,006 | 0,042 | 0,001 | 12,21 | 0,001 |
| 28 | 0,084 | 0,019 | | 13 | 0,007 |
| 29 | 0,267 | 0,008 | | 13,14 | 0,021 |
| 30 | 0,375 | | | 13,15 | 0,023 |
| 31 | 0,198 | | | 13,16 | 0,029 |
| 32 | 0,059 | | | 13,17 | 0,039 |
| 33 | 0,008 | | | 13,18 | 0,034 |
| 34 | 0,001 | | | 13,19 | 0,019 |
| | | | | 13,20 | 0,010 |
| | | | | 13,21 | 0,002 |
| | | | | 13,22 | 0,001 |
| | | | | 13,36 | 0,001 |
| | | | | 14 | 0,027 |
| | | | | 14,15 | 0,017 |
| | | | | 14,16 | 0,049 |
| | | | | 14,17 | 0,058 |
| | | | | 14,18 | 0,027 |
| | | | | 14,19 | 0,018 |
| | | | | 14,20 | 0,004 |
| | | | | 14,21 | 0,002 |
| | | | | 14,22 | 0,001 |
| | | | | 15 | 0,022 |
| | | | | 15,16 | 0,048 |
| | | | | 15,17 | 0,049 |
| | | | | 15,18 | 0,020 |

| | | | |
|--|--|--------------|-------|
| | | 15,19 | 0,011 |
| | | 16 | 0,018 |
| | | 16,17 | 0,029 |
| | | 16,18 | 0,006 |
| | | 16,19 | 0,006 |
| | | 16,20 | 0,002 |
| | | 17 | 0,003 |
| | | 17,18 | 0,014 |
| | | 17,19 | 0,005 |
| | | 17,20 | 0,002 |
| | | 18 | 0,010 |
| | | 18,19 | 0,003 |
| | | 18,20 | 0,001 |
| | | 19 | 0,001 |
| | | 19,20 | 0,002 |

Como complemento a la **Tabla 4**, los cálculos de índices genéticos y probabilidades específicas de relevancia en el campo forense se presentan para cada marcador en la **Tabla 5**. Se presenta el número de alelos, número de alelos efectivos, índice de información de Shannon, poder de discriminación, probabilidad de coincidencia y diversidad génica. Los valores promedio de este análisis son: 953, 11,3.936, 1.423, 0.674, 0.326, 0.674 respectivamente.

Se encontró que los marcadores más informativos son DYS576, DYS389II, DYS481, DYS570, DYS392, DYS458 y DYS385; debido a que tanto el índice de Shannon, el PD y la DG en estos marcadores tienen valores mayores a 1.47, 0.731 y 0.732 respectivamente. De estos, DYS437 mostró la diversidad de genes más baja (0.467), mientras que DYS481 mostró la diversidad de genes más alta (0.821) en una sola copia Y-STR, y DYS385 mostró la diversidad de genes más alta (0.945).

Tabla 5. Índices genéticos y probabilidades específicas de relevancia en el campo forense para cada marcador analizado.

| LOCUS | N | NA | NE | I | PD | PC | DG |
|----------|-----|----|-------|-------|-------|-------|-------|
| DYS576 | 955 | 12 | 5.060 | 1.822 | 0.802 | 0.198 | 0.803 |
| DYS389I | 965 | 6 | 2.195 | 0.964 | 0.544 | 0.456 | 0.545 |
| DYS448 | 965 | 6 | 3.209 | 1.294 | 0.688 | 0.312 | 0.689 |
| DYS389II | 965 | 9 | 3.816 | 1.502 | 0.738 | 0.262 | 0.739 |
| DYS19 | 966 | 8 | 3.034 | 1.306 | 0.670 | 0.330 | 0.671 |
| DYS391 | 966 | 7 | 2.136 | 0.946 | 0.532 | 0.468 | 0.532 |
| DYS481 | 956 | 11 | 5.569 | 1.902 | 0.820 | 0.180 | 0.821 |
| DYS533 | 953 | 7 | 2.405 | 1.086 | 0.584 | 0.416 | 0.585 |
| DYS438 | 967 | 7 | 3.157 | 1.317 | 0.683 | 0.317 | 0.684 |
| DYS437 | 966 | 5 | 1.874 | 0.843 | 0.467 | 0.533 | 0.467 |
| DYS570 | 929 | 14 | 4.343 | 1.739 | 0.770 | 0.230 | 0.771 |
| DYS635 | 967 | 10 | 3.148 | 1.338 | 0.682 | 0.318 | 0.683 |
| DYS390 | 967 | 9 | 2.909 | 1.348 | 0.656 | 0.344 | 0.657 |
| DYS439 | 964 | 8 | 2.806 | 1.233 | 0.644 | 0.356 | 0.644 |
| DYS392 | 935 | 10 | 3.719 | 1.475 | 0.731 | 0.269 | 0.732 |
| DYS393 | 940 | 7 | 2.124 | 1.023 | 0.529 | 0.471 | 0.530 |
| DYS458 | 912 | 10 | 4.292 | 1.614 | 0.767 | 0.233 | 0.768 |
| DYS456 | 938 | 9 | 2.689 | 1.242 | 0.628 | 0.372 | 0.629 |

| | | | | | | | |
|------------------|-------|-------|--------|-------|-------|-------|-------|
| Y-GATA-H4 | 940 | 7 | 2.449 | 1.059 | 0.592 | 0.408 | 0.592 |
| DYS385 | 938 | 59 | 17.781 | 3.412 | 0.944 | 0.056 | 0.945 |
| PROMEDIO | 953 | 11 | 3.936 | 1.423 | 0.674 | 0.326 | 0.674 |
| SE | 3.638 | 2.572 | 0.762 | 0.123 | 0.026 | 0.026 | 0.026 |

Na = Número de alelos diferentes; **Ne** = Número de alelos efectivos = $1 / (\text{Suma } \pi^2)$; **I** = Índice de información de Shannon = $-1 * \text{Suma } (\pi * \ln(\pi))$; **PD** = Poder de discriminación = $1 - \text{Suma } \pi^2$; **PM** = Probabilidad de coincidencia = $1 - \text{PD}$; **DG** = Diversidad génica = $(N / (N-1)) * h$ Donde π es la frecuencia del i-ésimo alelo para la población y $\text{Suma } \pi^2$ es la suma de las frecuencias alélicas de la población al cuadrado; **SE** = error estándar.

3.2 Análisis intra-poblacional

Para este análisis los 968 perfiles genéticos recopilados se diferenciaron por región, agrupados de la siguiente manera: Costa 340 perfiles, Sierra 516 perfiles y Oriente 112 perfiles. Los resultados de la diferenciación de frecuencias alélicas por regiones se presenta gráficas para cada marcador, se encontró que para la mayoría de marcadores en las tres regiones se presenta una distribución unimodal con excepción de DYS481, DYS392 que presentan una distribución multimodal.

El marcador DYS576 registra una escalera alélica que va desde el alelo 10 hasta el alelo 23. El alelo con mayor frecuencia para este marcador es el 18 en las tres regiones: Costa 0.264, Sierra 0.346 y Oriente 0.375. Por otro lado, el alelo menos frecuente es el 23, que está presente únicamente en la Sierra con una frecuencia de 0.002, véase **Figura 2A**. La exploración gráfica realizada con un diagrama de cajas del marcador DYS576 para las distintas regiones muestra que aparentemente no hay diferencia en sus medianas. Se evidencia que dos de las tres regiones presentan datos atípicos, mientras la Sierra presenta valores atípicos altos y bajos, la Costa presenta únicamente valores atípicos bajos, véase **Figura 2B**.

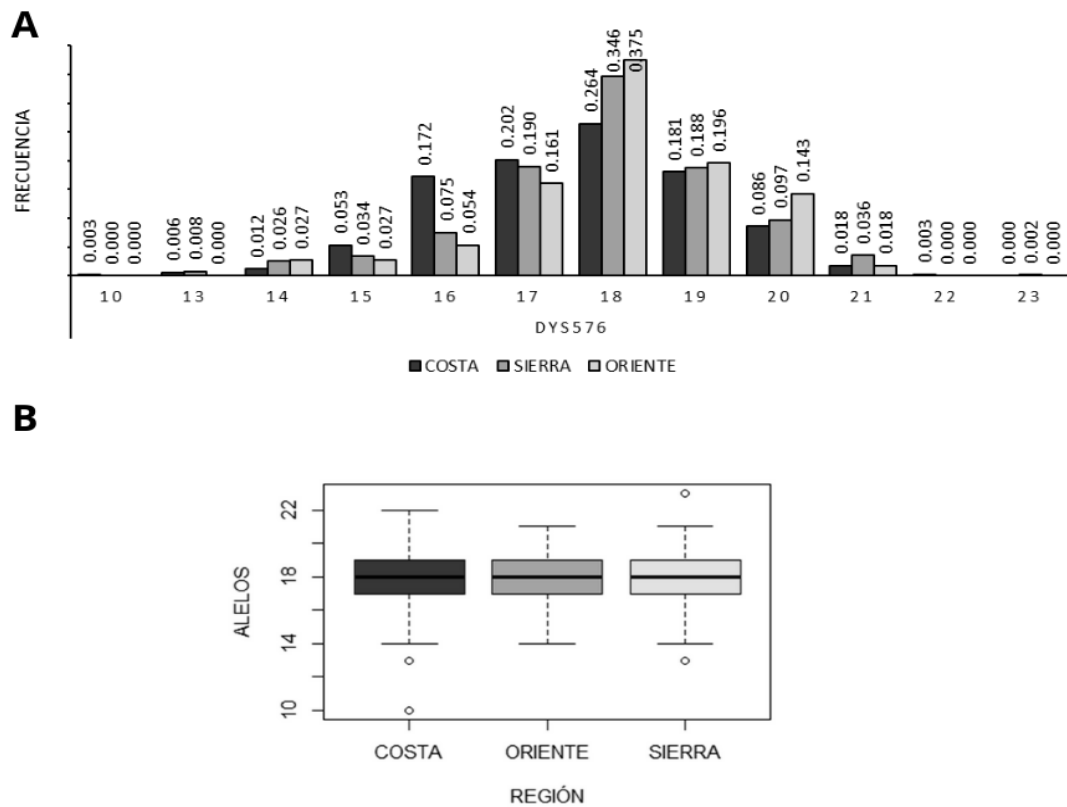


Figura 2. Gráficas del marcador DYS576 para las tres regiones de estudio. A: frecuencias alélicas, B: alelos.

El marcador DYS389I registra una escalera alélica que va desde el alelo 10 hasta el alelo 15. El alelo con mayor frecuencia para este marcador es el 13 en las tres regiones: Costa 0.652, Sierra 0.576 y Oriente 0.723. Por otro lado, el alelo menos frecuente es el 11, que está presente únicamente en la Sierra con una frecuencia de 0.002, véase **Figura 3A**. La exploración gráfica realizada con un diagrama de cajas del marcador DYS389I para las distintas regiones muestra que aparentemente no hay diferencia en sus medianas. Se evidencia que las tres regiones presentan datos atípicos, mientras la Costa y el Oriente presentan valores atípicos altos y bajos, la Sierra presenta únicamente valores atípicos bajos, véase **Figura 3B**.

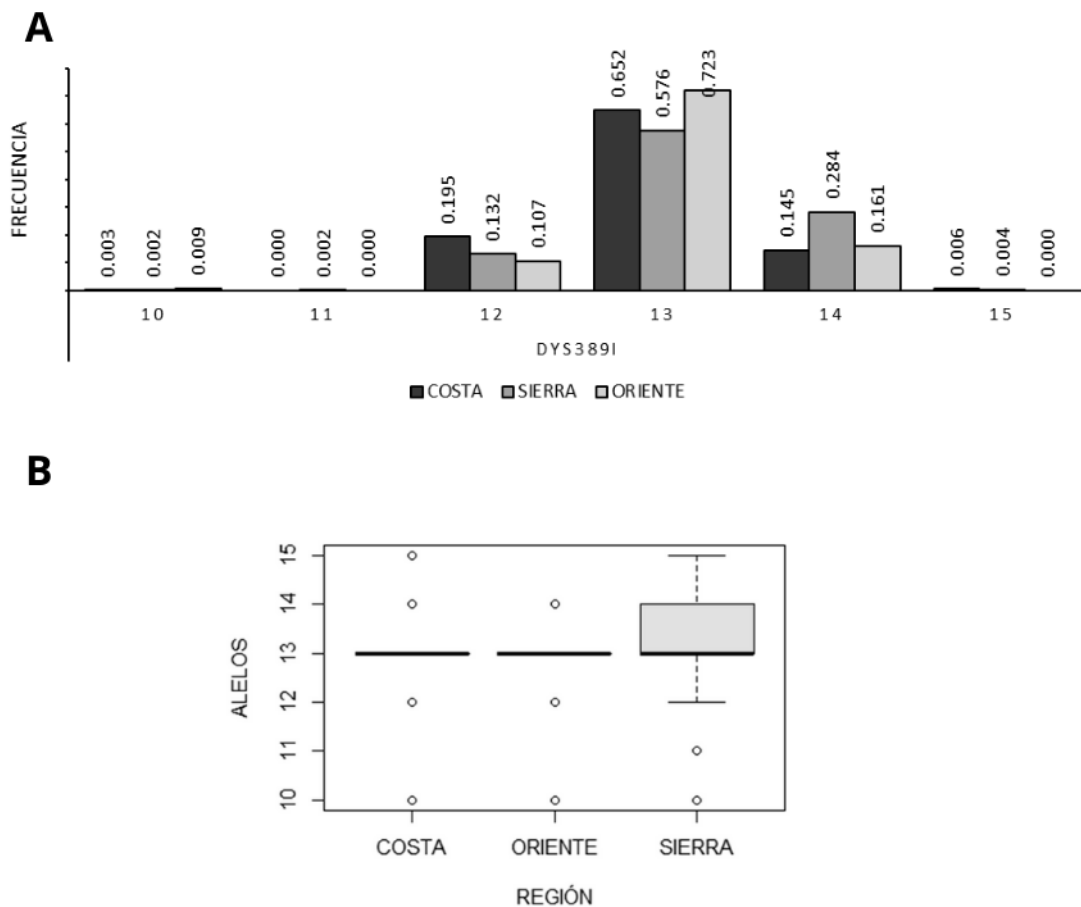


Figura 3. Gráficas del marcador DYS389I para las tres regiones de estudio. **A:** frecuencias alélicas, **B:** alelos.

El marcador DYS448 registra una escalera alélica que va desde el alelo 17 hasta el alelo 22. Este marcador tiene dos alelos frecuentes en las tres regiones: alelo 19 y alelo 20. Las frecuencias para el alelo 19 son: Costa 0.386, Sierra 0.350 y Oriente 0.375. Mientras que, las frecuencias para el alelo 20 son: Costa 0.339, Sierra 0.409 y Oriente 0.411. Por otro lado, el alelo menos frecuente es el 17, que está presente únicamente en la Sierra con una frecuencia de 0.004, véase **Figura 4A**. La exploración gráfica realizada con un diagrama de cajas del marcador DYS448 para las distintas regiones muestra que aparentemente no hay diferencia en sus medianas. Se evidencia que las tres regiones presentan datos atípicos, mientras la Costa y el Oriente presentan valores atípicos altos, la Sierra presenta valores atípicos altos y bajos, véase **Figura 4B**.

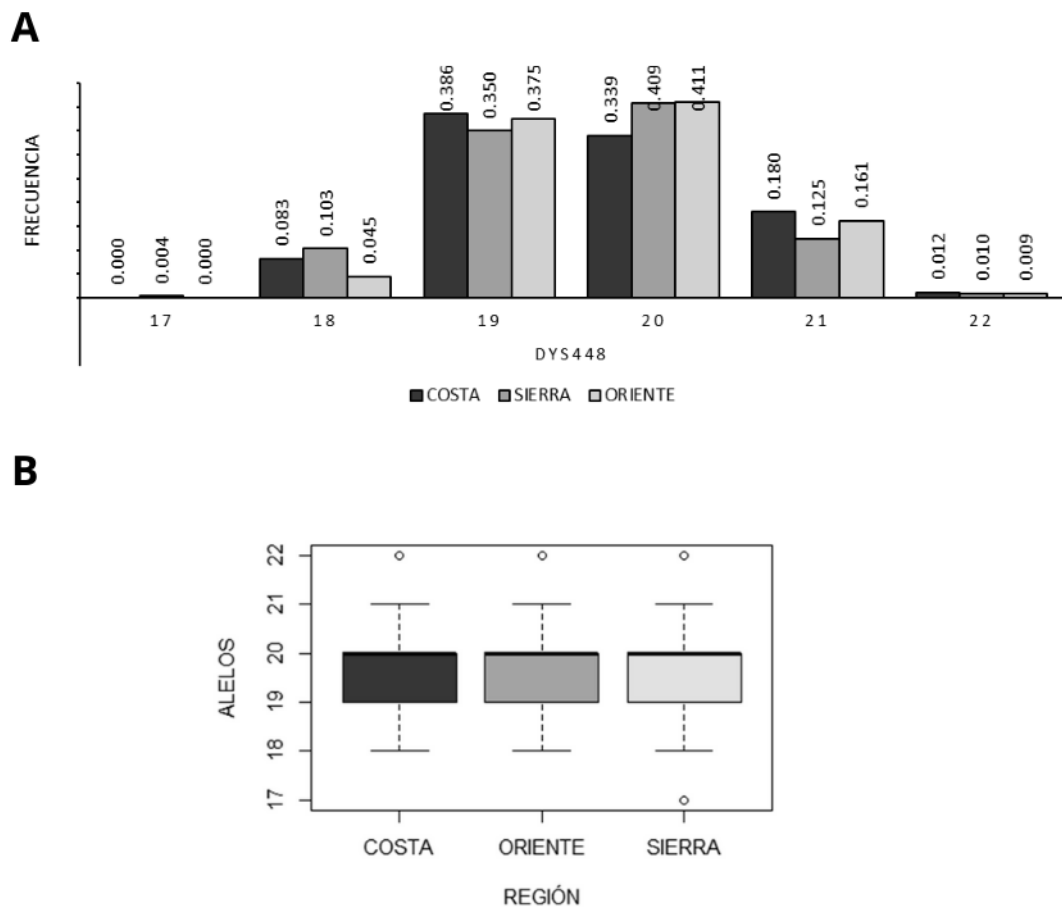


Figura 4. Gráficas del marcador DYS448 para las tres regiones de estudio. **A:** frecuencias alélicas, **B:** alelos.

El marcador DYS389II registra una escalera alélica que va desde el alelo 26 hasta el alelo 34. El alelo con mayor frecuencia para este marcador es el 30 en las tres regiones: Costa 0.347, Sierra 0.378 y Oriente 0.446. Por otro lado, los alelos menos frecuentes son el 26 y el 34, que está presentes únicamente en la Sierra con una frecuencia de 0.002 cada uno, véase **Figura 5A**. La exploración gráfica realizada con un diagrama de cajas del marcador DYS389II para las distintas regiones muestra que aparentemente no hay diferencia en sus medianas. Se evidencia que dos de las tres regiones presentan datos atípicos, mientras el Oriente presenta valores atípicos altos, la Costa presenta valores atípicos altos y bajos, véase **Figura 5B**.

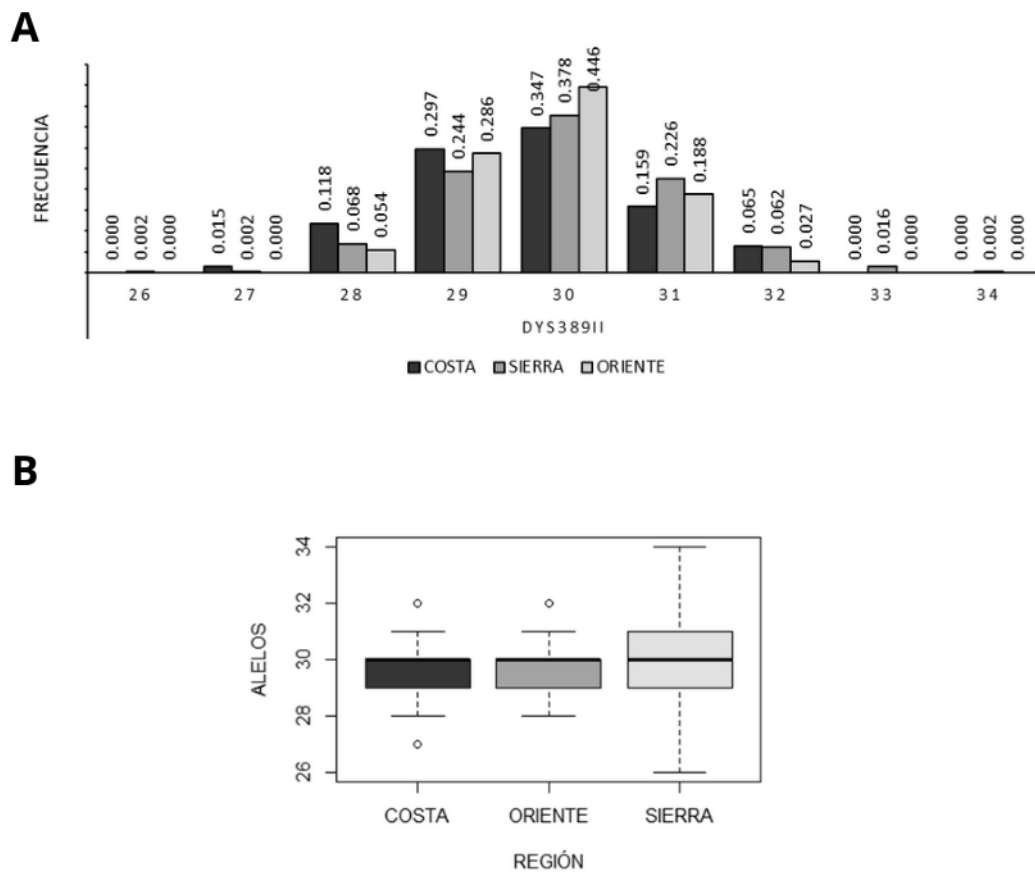


Figura 5. Gráficas del marcador DYS389II para las tres regiones de estudio. **A:** frecuencias alélicas, **B:** alelos.

El marcador DYS19 registra una escalera alélica que va desde el alelo 11 hasta el alelo 18. Este marcador tiene dos alelos frecuentes en las tres regiones: alelo 13 y alelo 14. Las frecuencias para el alelo 13 son: Costa 0.299, Sierra 0.463 y Oriente 0.500. Mientras que, las frecuencias para el alelo 14 son: Costa 0.441, Sierra 0.355 y Oriente 0.313. Por otro lado, los alelos menos frecuentes son el 11 y el 18, que está presentes únicamente en la Sierra con una frecuencia de 0.002 cada uno, véase **Figura 6A** La exploración gráfica realizada con un diagrama de cajas del marcador DYS19 para las distintas regiones muestra que aparentemente no hay diferencia en sus medianas al comparar Costa y Sierra. Sin embargo, hay una clara diferencia en la mediana de la región Oriente. Se evidencia que dos de las tres regiones presentan datos atípicos, mientras el Oriente presenta valores atípicos altos, la Sierra presenta valores atípicos altos y bajos, véase **Figura 6B**.

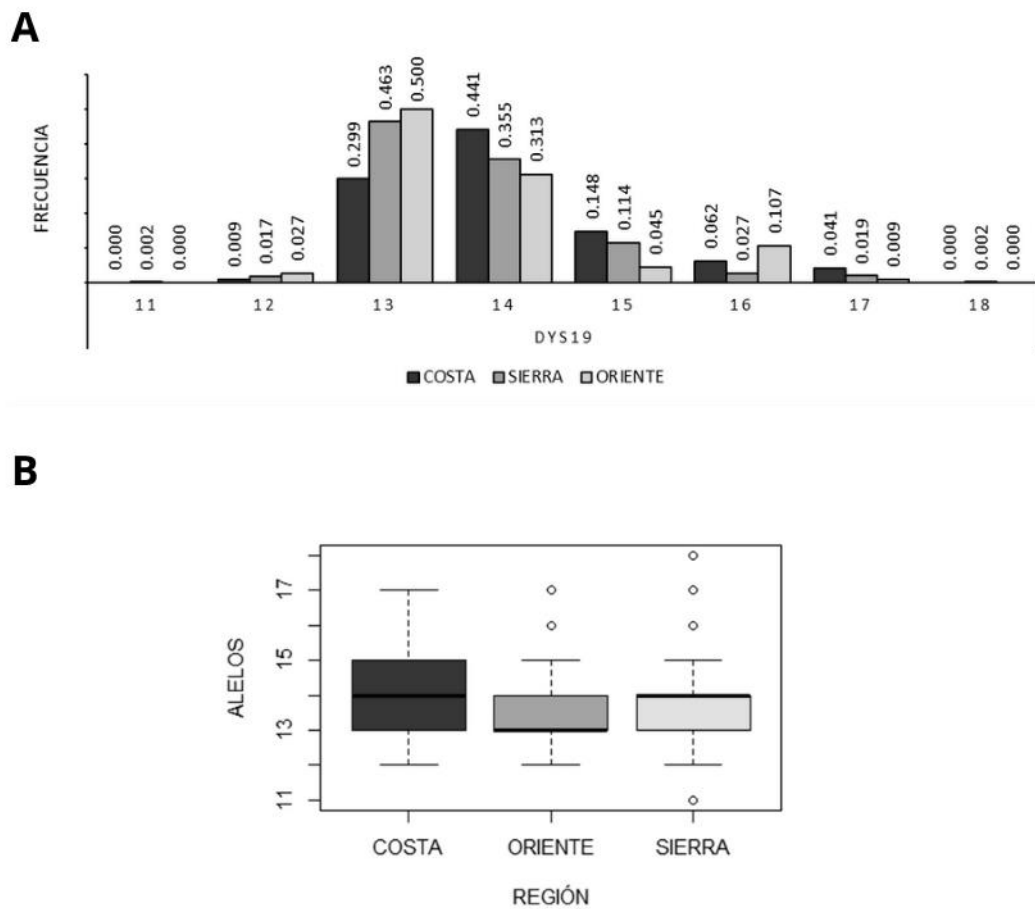


Figura 6. Gráficas del marcador DYS19 para las tres regiones de estudio. **A:** frecuencias alélicas, **B:** alelos.

El marcador DYS391 registra una escalera alélica que va desde el alelo 7 hasta el alelo 15. El alelo con mayor frecuencia para este marcador es el 10 en las tres regiones: Costa 0.572, Sierra 0.637 y Oriente 0.598. Por otro lado, los alelos menos frecuentes son el 7 y el 15, que está presentes únicamente en la Sierra con una frecuencia de 0.002 cada uno, véase **Figura 7A**. La exploración gráfica realizada con un diagrama de cajas del marcador DYS391 para las distintas regiones muestra que aparentemente no hay diferencia en sus medianas. Se evidencia que las tres regiones presentan datos atípicos, mientras la Costa y el Oriente presentan valores atípicos bajos, la Sierra presenta valores atípicos altos y bajos, véase **Figura 7B**.

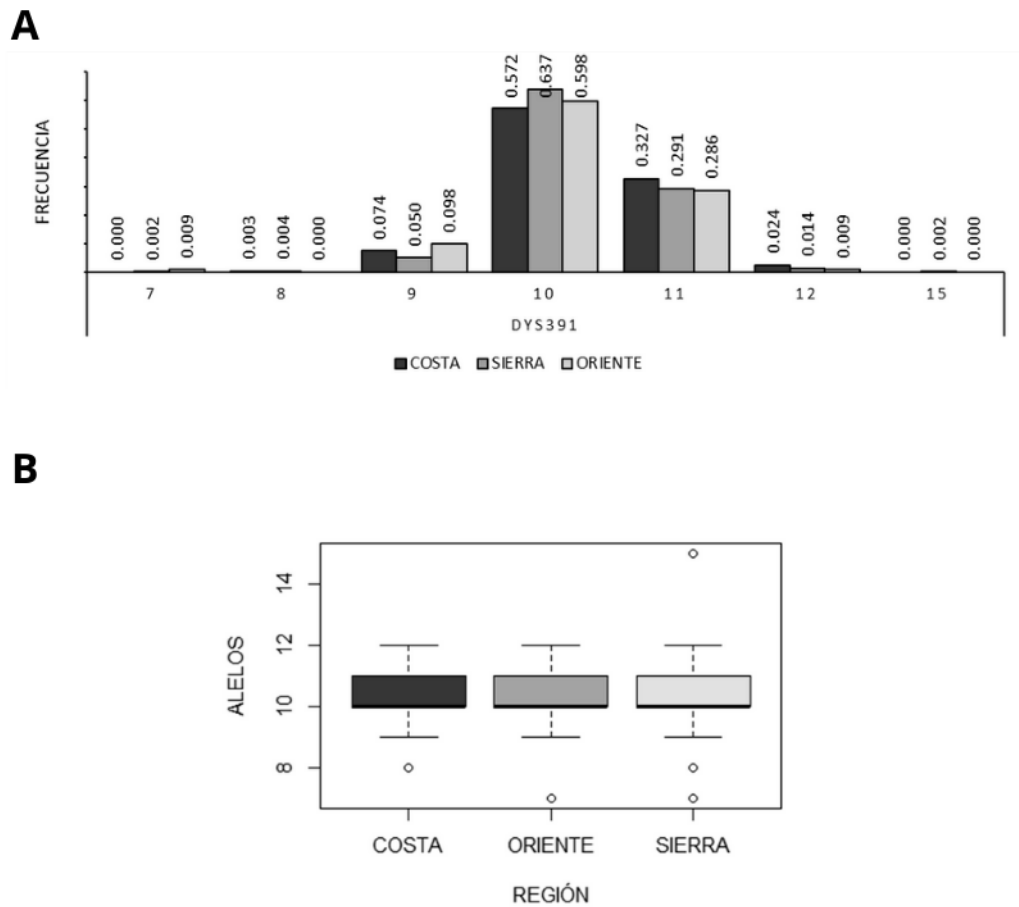


Figura 7. Gráficas del marcador DYS391 para las tres regiones de estudio. **A:** frecuencias alélicas, **B:** alelos.

El marcador DYS481 registra una escalera alélica que va desde el alelo 17 hasta el alelo 29. Este marcador tiene dos alelos frecuentes en las tres regiones: alelo 22 y alelo 24. Las frecuencias para el alelo 2 son: Costa 0.270, Sierra 0.252 y Oriente 0.232. Mientras que, las frecuencias para el alelo 24 son: Costa 0.208, Sierra 0.208 y Oriente 0.277. Por otro lado, el alelo menos frecuente es el 17, que está presente únicamente en la Costa con una frecuencia de 0.003, véase **Figura 8A**. La exploración gráfica realizada con un diagrama de cajas del marcador DYS481 para las distintas regiones muestra que aparentemente no hay diferencia en sus medianas. Se evidencia que únicamente la Costa presenta datos atípicos (bajos), véase **Figura 8B**.

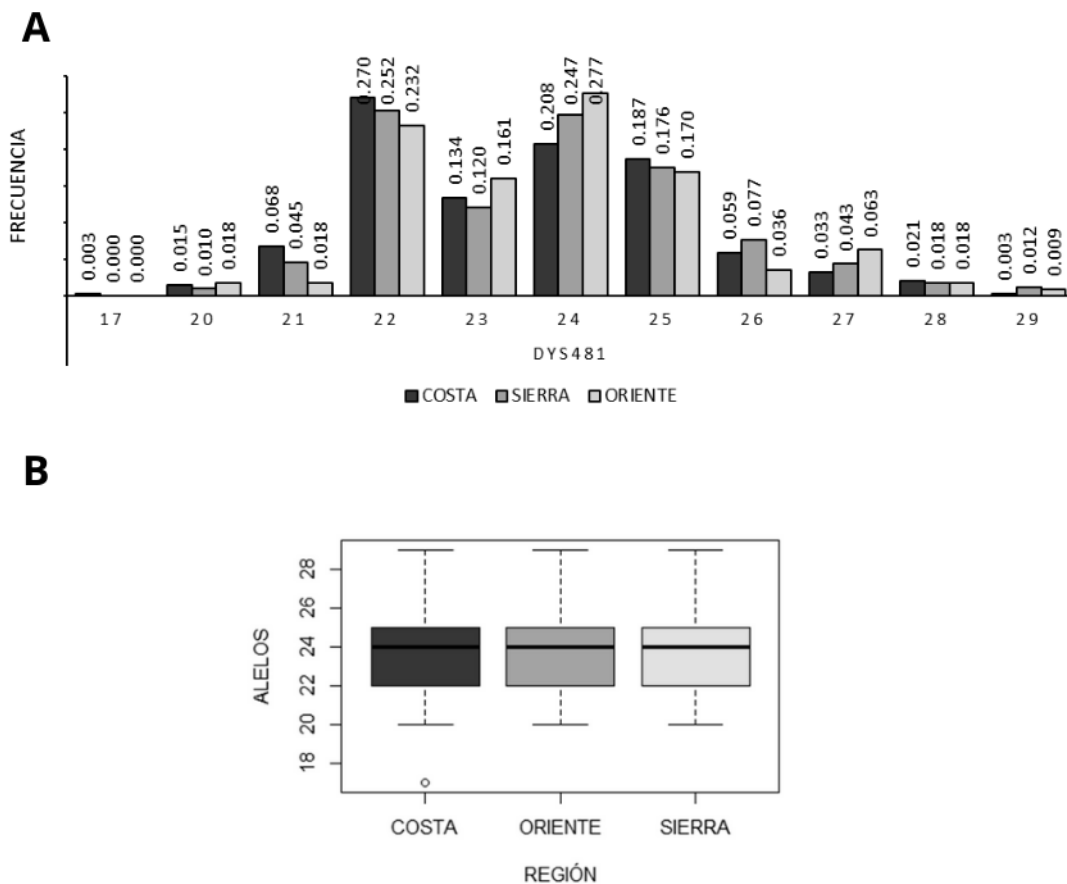


Figura 8. Gráficas del marcador DYS481 para las tres regiones de estudio. **A:** frecuencias alélicas, **B:** alelos.

El marcador DYS533 registra una escalera alélica que va desde el alelo 9 hasta el alelo 18. El alelo con mayor frecuencia para este marcador es el 12 en las tres regiones: Costa 0.522, Sierra 0.576 y Oriente 0.550. Por otro lado, el alelo menos frecuente es el 18, que está presente únicamente en la Costa con una frecuencia de 0.003, véase **Figura 9A**. La exploración gráfica realizada con un diagrama de cajas del marcador DYS533 para las distintas regiones muestra que aparentemente no hay diferencia en sus medianas. Se evidencia que las tres regiones presentan datos atípicos, mientras la Sierra presenta valores atípicos bajos, el Oriente presenta valores atípicos altos y la Costa presenta valores atípicos altos y bajos, véase **Figura 9B**.

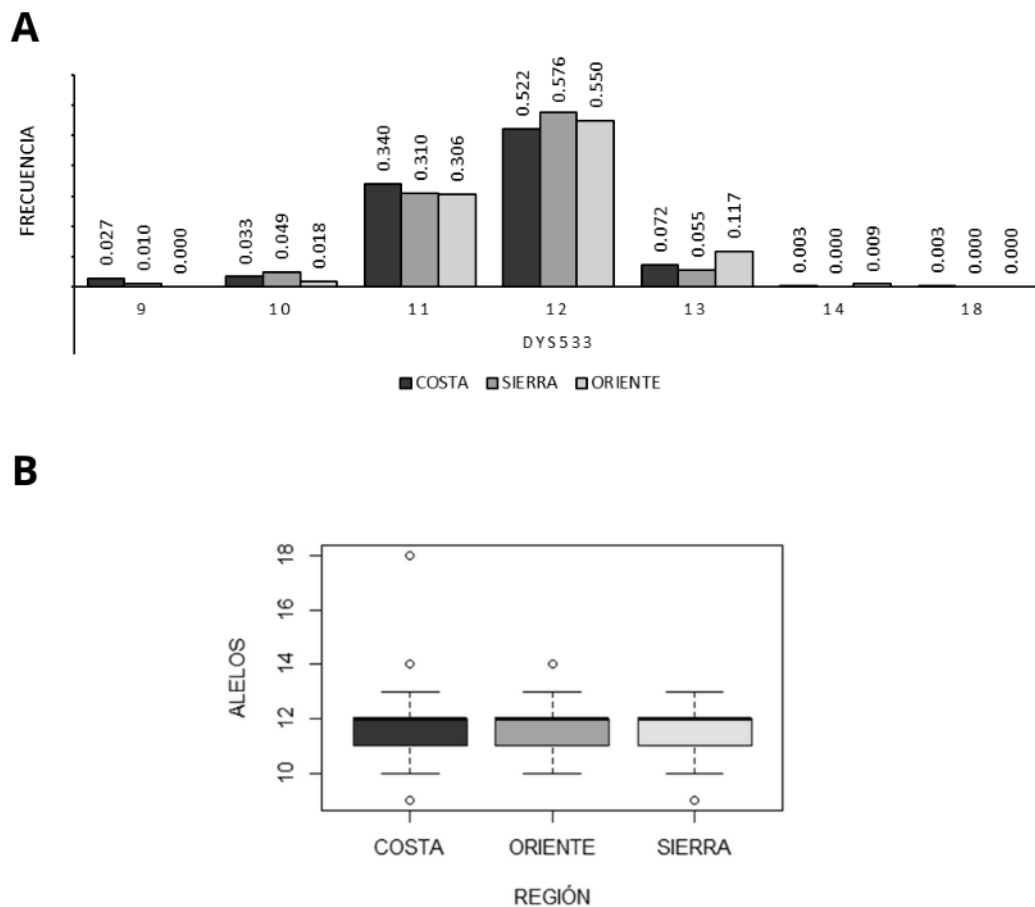


Figura 9. Gráficas del marcador DYS533 para las tres regiones de estudio. **A:** frecuencias alélicas, **B:** alelos.

El marcador DYS438 registra una escalera alélica que va desde el alelo 8 hasta el alelo 14. El alelo con mayor frecuencia para este marcador es el 11 en las tres regiones: Costa 0.362, Sierra 0.474 y Oriente 0.518. Por otro lado, el alelo menos frecuente es el 14, que está presente únicamente en la Costa con una frecuencia de 0.003, véase **Figura 10A**. La exploración gráfica realizada con un diagrama de cajas del marcador DYS438 para las distintas regiones muestra que aparentemente no hay diferencia en sus medianas. Se evidencia que únicamente la Sierra y el Oriente presentan valores atípicos (bajos), véase **Figura 10B**.

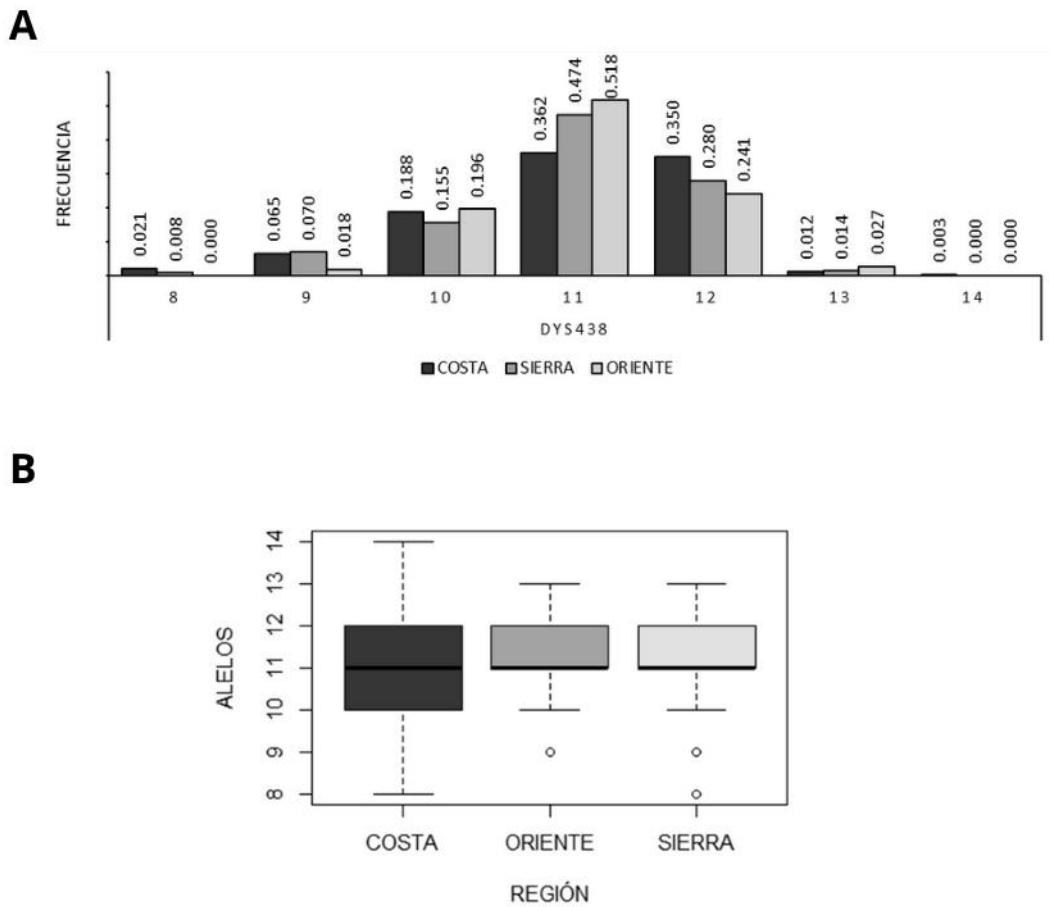


Figura 10. Gráficas del marcador DYS438 para las tres regiones de estudio. **A:** frecuencias alélicas, **B:** alelos.

El marcador DYS437 registra una escalera alélica que va desde el alelo 13 hasta el alelo 17. El alelo con mayor frecuencia para este marcador es el 14 en las tres regiones: Costa 0.625, Sierra 0.718 y Oriente 0.723. Por otro lado, el alelo menos frecuente es el 17, que está presente únicamente en la Costa y la Sierra con una frecuencia de 0.003 y 0.004, respectivamente, véase **Figura 11A**. La exploración gráfica realizada con un diagrama de cajas del marcador DYS437 para las distintas regiones muestra que aparentemente no hay diferencia en sus medianas. Se evidencia que las tres regiones presentan datos atípicos, mientras la Sierra y la Costa presentan únicamente valores atípicos altos, el Oriente presenta valores atípicos altos y bajos, véase **Figura 11B**.

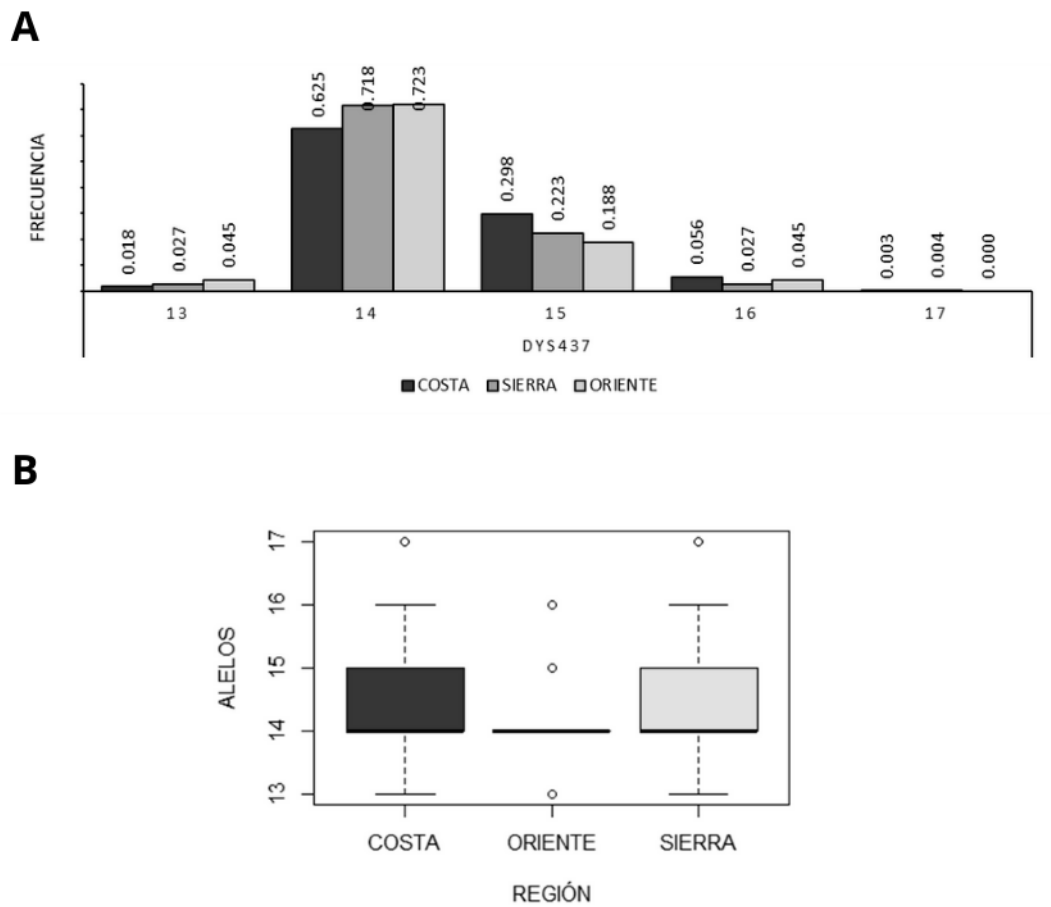


Figura 11. Gráficas del marcador DYS437 para las tres regiones de estudio. **A:** frecuencias alélicas, **B:** alelos.

El marcador DYS570 registra una escalera alélica que va desde el alelo 11 hasta el alelo 25. El alelo con mayor frecuencia para este marcador es el 17 en las tres regiones: Costa 0.355, Sierra 0.374 y Oriente 0.396. Por otro lado, el alelo menos frecuente es el 25, que está presente únicamente en la Sierra con una frecuencia de 0.002, véase **Figura 12A**. La exploración gráfica realizada con un diagrama de cajas del marcador DYS570 para las distintas regiones muestra que aparentemente no hay diferencia en sus medianas. Se evidencia que las tres regiones presentan datos atípicos (altos y bajos), véase **Figura 12B**.

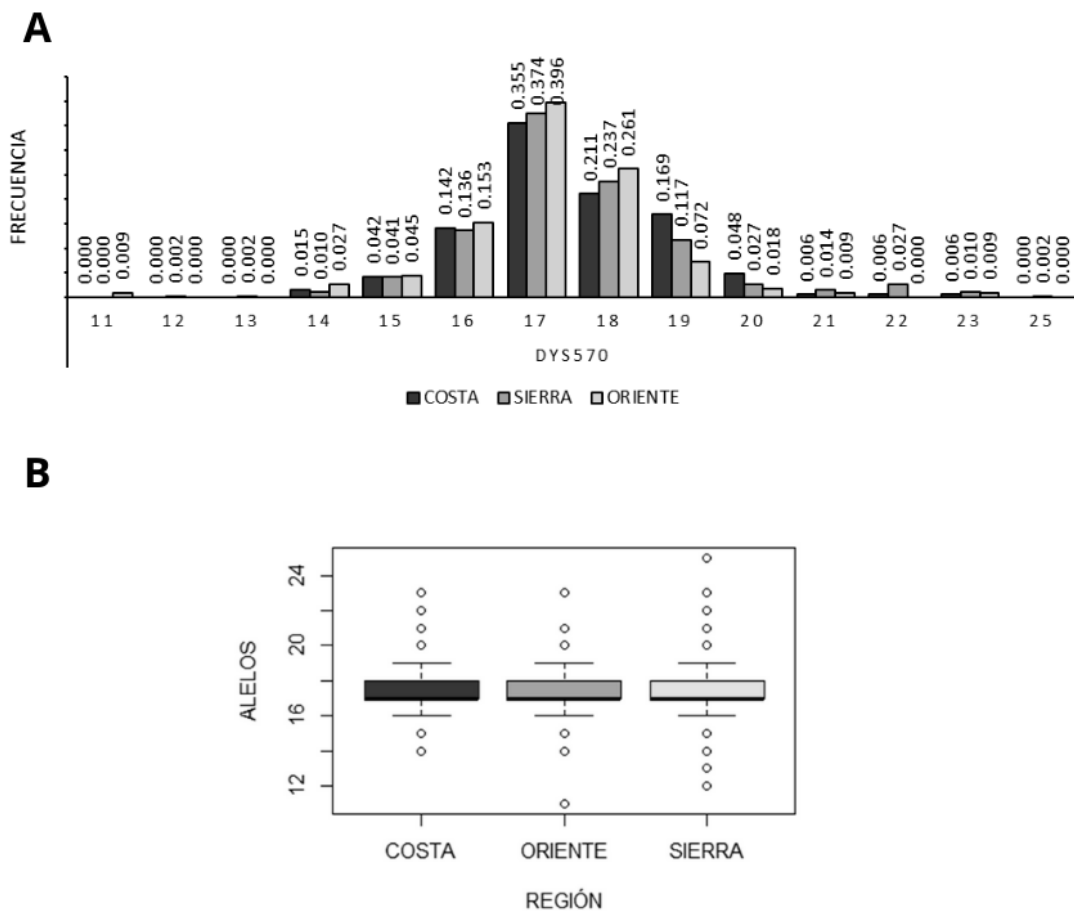


Figura 12. Gráficas del marcador DYS570 para las tres regiones de estudio. **A:** frecuencias alélicas, **B:** alelos.

El marcador DYS635 registra una escalera alélica que va desde el alelo 17 hasta el alelo 26. El alelo con mayor frecuencia para este marcador es el 22 en las tres regiones: Costa 0.312, Sierra 0.482 y Oriente 0.607. Por otro lado, los alelos menos frecuentes son el 17 y el 19, que están presentes únicamente en la Sierra con una frecuencia de 0.002 para ambos alelos, véase **Figura 13A**. La exploración gráfica realizada con un diagrama de cajas del marcador DYS635 para las distintas regiones muestra que aparentemente no hay diferencia en sus medianas. Se evidencia que dos de las tres regiones presentan datos atípicos, mientras el Oriente presenta únicamente valores atípicos bajos, la Sierra presenta valores atípicos altos y bajos, véase **Figura 13B**.

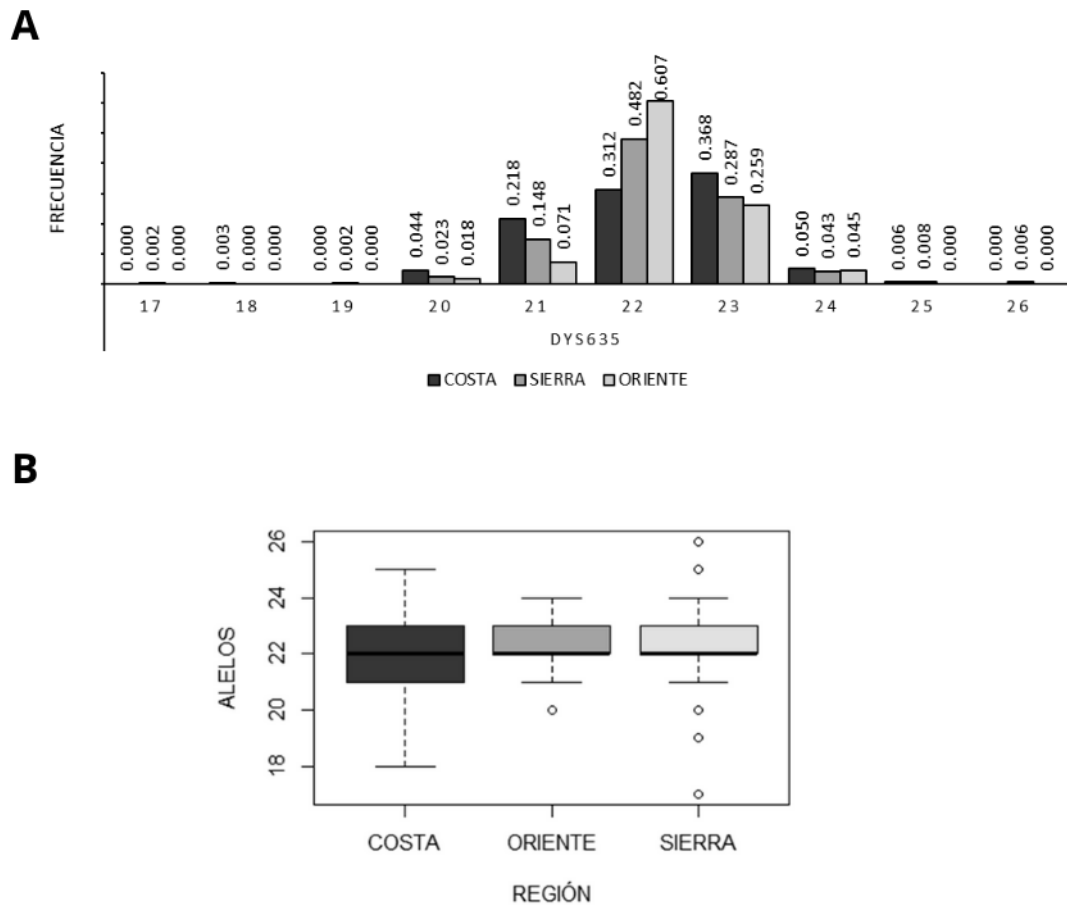


Figura 13. Gráficas del marcador DYS635 para las tres regiones de estudio. **A:** frecuencias alélicas, **B:** alelos.

El marcador DYS390 registra una escalera alélica que va desde el alelo 19 hasta el alelo 27. El alelo con mayor frecuencia para este marcador es el 24 en las tres regiones: Costa 0.484, Sierra 0.519 y Oriente 0.589. Por otro lado, los alelos menos frecuentes son el 19 y el 27, que están presentes únicamente en la Costa con una frecuencia de 0.003 para ambos alelos, véase **Figura 14A**. La exploración gráfica realizada con un diagrama de cajas del marcador DYS390 para las distintas regiones muestra que aparentemente no hay diferencia en sus medianas. Se evidencia que las tres regiones presentan datos atípicos, mientras el Oriente presenta únicamente valores atípicos altos, la Sierra y la Costa presentan valores atípicos altos y bajos, véase **Figura 14B**.

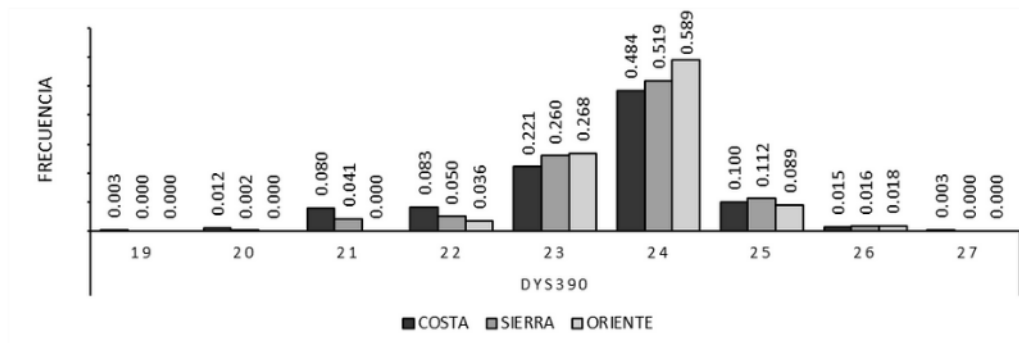
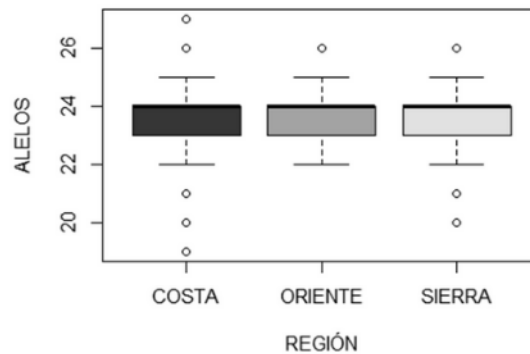
A**B**

Figura 14. Gráficas del marcador DYS390 para las tres regiones de estudio. **A:** frecuencias alélicas, **B:** alelos.

El marcador DYS439 registra una escalera alélica que va desde el alelo 8 hasta el alelo 15. El alelo con mayor frecuencia para este marcador es el 12 en las tres regiones: Costa 0.541, Sierra 0.488 y Oriente 0.500. Por otro lado, el alelo menos frecuente es el 15, que están presentes únicamente en la Sierra con una frecuencia de 0.002, véase **Figura 15A**. La exploración gráfica realizada con un diagrama de cajas del marcador DYS439 para las distintas regiones muestra que aparentemente no hay diferencia en sus medianas. Se evidencia que las tres regiones presentan datos atípicos (altos y bajos), véase **Figura 15B**.

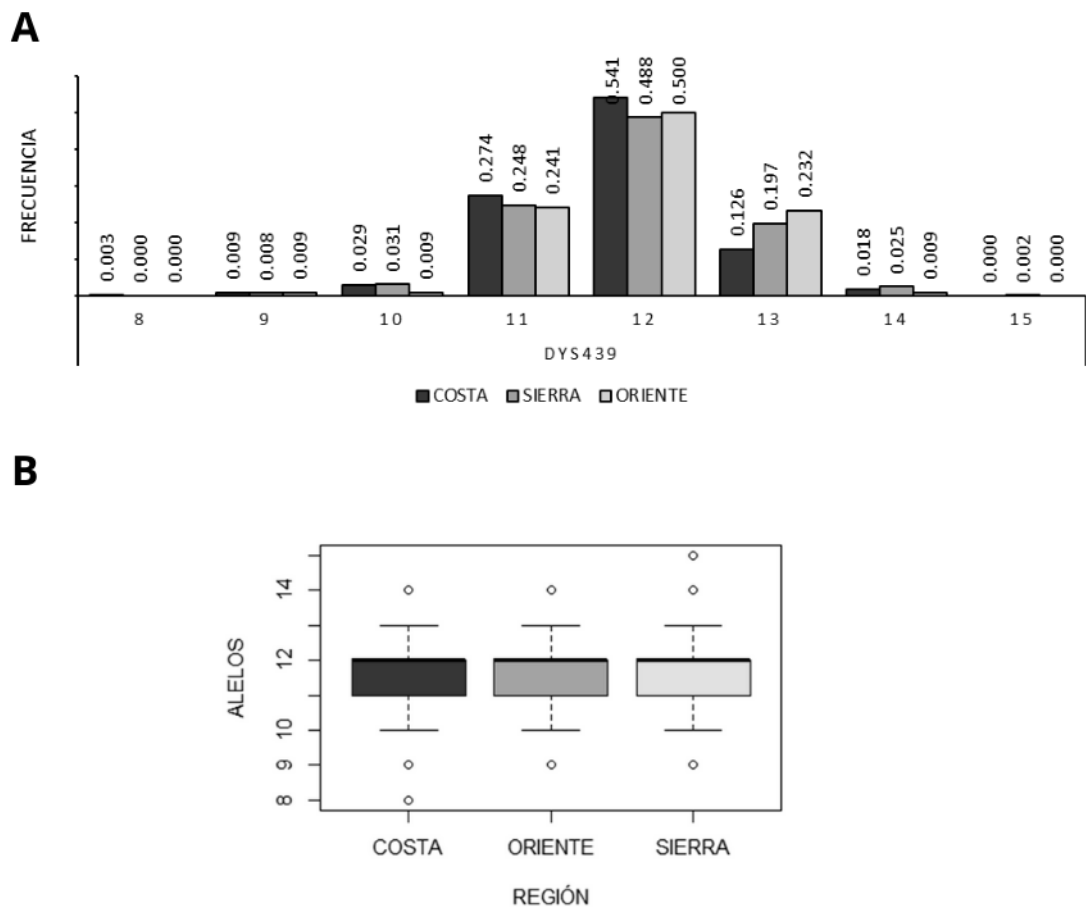


Figura 15. Gráficas del marcador DYS439 para las tres regiones de estudio. **A:** frecuencias alélicas, **B:** alelos.

El marcador DYS392 registra una escalera alélica que va desde el alelo 7 hasta el alelo 17. Este marcador tiene dos alelos frecuentes en las tres regiones: alelo 13 y alelo 14. Las frecuencias para el alelo 13 son: Costa 0.369, Sierra 0.289 y Oriente 0.259. Mientras que, las frecuencias para el alelo 14 son: Costa 0.224, Sierra 0.372 y Oriente 0.500. Por otro lado, los alelos menos frecuentes son el 7 y el 9, que está presente únicamente en la Costa con una frecuencia de 0.003 cada uno, véase **Figura 16A**. La exploración gráfica realizada con un diagrama de cajas del marcador DYS392 para las distintas regiones muestra que aparentemente no hay diferencia en sus medianas al comparar Costa y Sierra. Sin embargo, hay una clara diferencia en la mediana de la región Oriente. Se evidencia que únicamente el Oriente y la Sierra presentan valores atípicos (altos y bajos), véase **Figura 16B**.

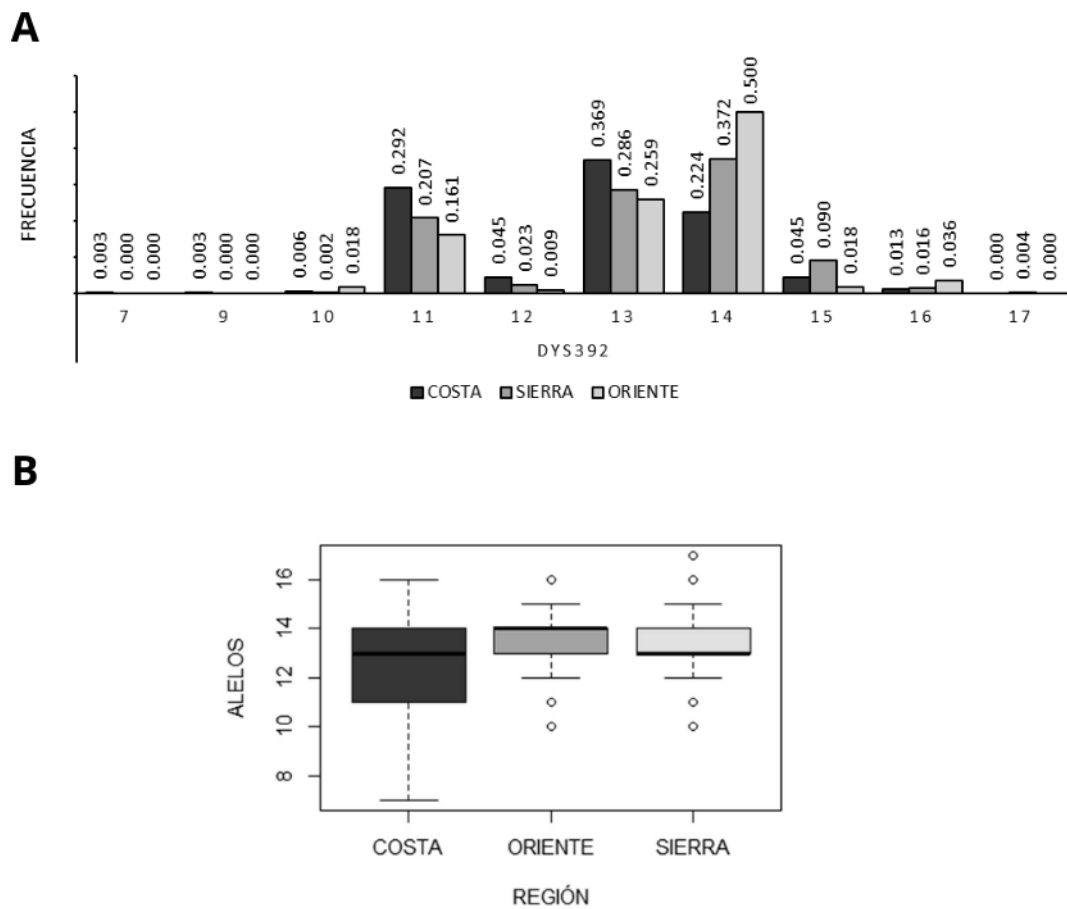


Figura 16. Gráficas del marcador DYS392 para las tres regiones de estudio. **A:** frecuencias alélicas, **B:** alelos.

El marcador DYS393 registra una escalera alélica que va desde el alelo 9 hasta el alelo 16. El alelo con mayor frecuencia para este marcador es el 13 en las tres regiones: Costa 0.554, Sierra 0.684 y Oriente 0.768. Por otro lado, el alelo menos frecuente es el 16, que están presentes únicamente en la Sierra con una frecuencia de 0.002, véase **Figura 17A**. La exploración gráfica realizada con un diagrama de cajas del marcador DYS393 para las distintas regiones muestra que aparentemente no hay diferencia en sus medianas. Se evidencia que las tres regiones presentan datos atípicos, mientras la Costa presenta únicamente valores atípicos bajos, la Sierra y el Oriente presentan valores atípicos altos y bajos, véase **Figura 17B**.

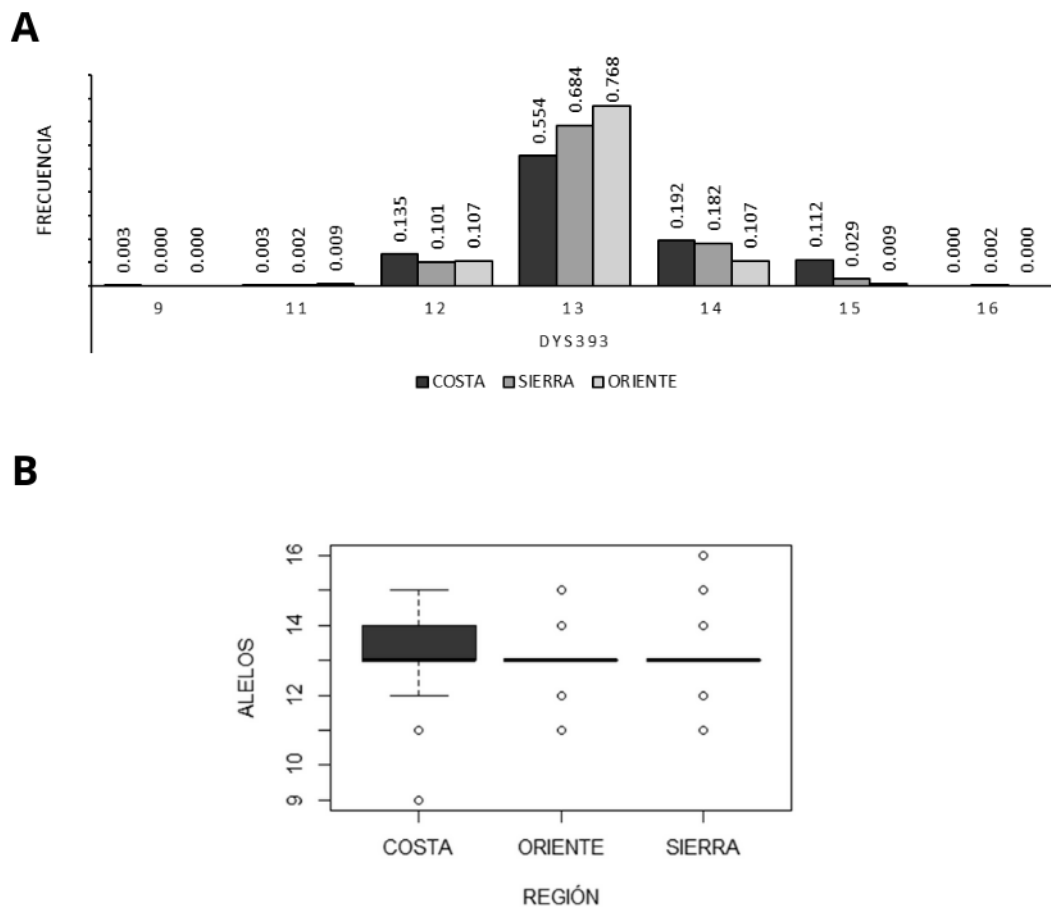


Figura 17. Gráficas del marcador DYS393 para las tres regiones de estudio. **A:** frecuencias alélicas, **B:** alelos.

El marcador DYS458 registra una escalera alélica que va desde el alelo 12 hasta el alelo 21. Este marcador tiene dos alelos frecuentes en las tres regiones: alelo 16 y alelo 17. Las frecuencias para el alelo 16 son: Costa 0.264, Sierra 0.309 y Oriente 0.273. Mientras que, las frecuencias para el alelo 17 son: Costa 0.320, Sierra 0.295 y Oriente 0.355. Por otro lado, el alelo menos frecuente es el 12 que está presente únicamente en la Costa con una frecuencia de 0.003, véase **Figura 18A**. La exploración gráfica realizada con un diagrama de cajas del marcador DYS458 para las distintas regiones muestra que aparentemente no hay diferencia en sus medianas al comparar Costa y Sierra. Sin embargo, hay una clara diferencia en la mediana de la región Oriente. Se evidencia que las tres regiones presentan datos atípicos (altos y bajos), véase **Figura 18B**.

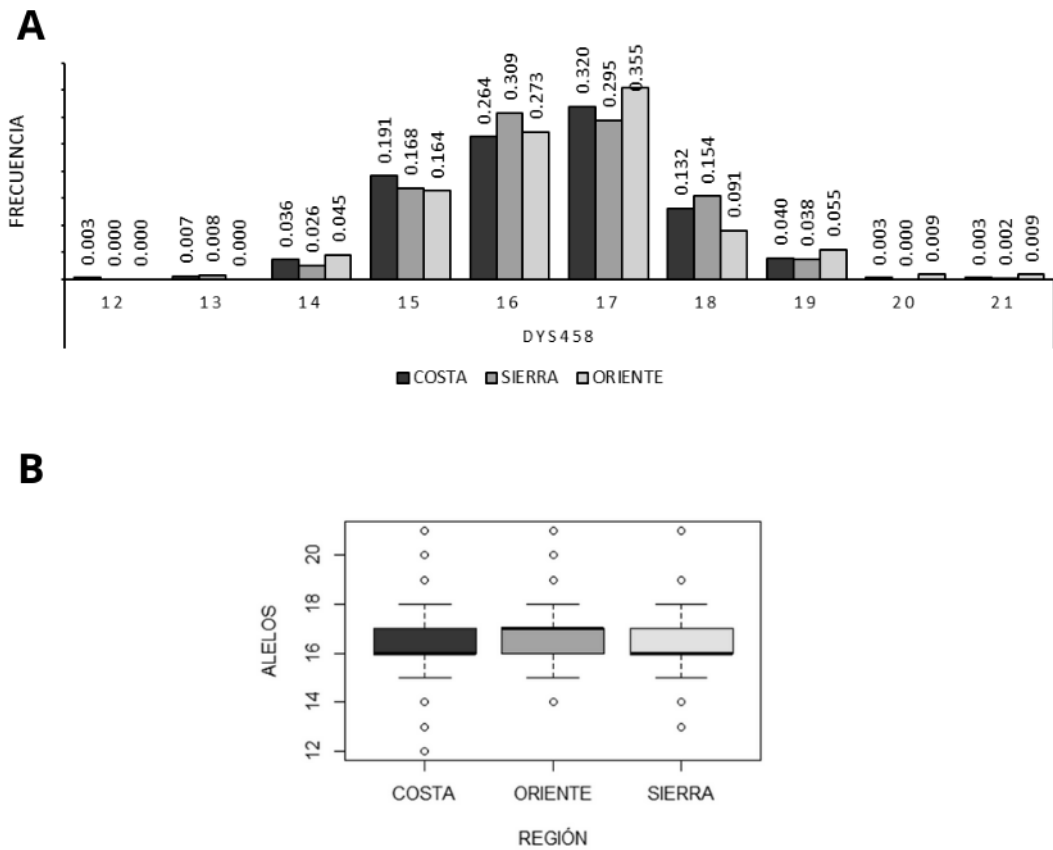


Figura 18. Gráficas del marcador DYS458 para las tres regiones de estudio. **A:** frecuencias alélicas, **B:** alelos.

El marcador DYS456 registra una escalera alélica que va desde el alelo 12 hasta el alelo 20. El alelo con mayor frecuencia para este marcador es el 15 en las tres regiones: Costa 0.545, Sierra 0.515 y Oriente 0.550. Por otro lado, el alelo menos frecuente es el 12, que está presente únicamente en la Sierra con una frecuencia de 0.002, véase **Figura 19A**. La exploración gráfica realizada con un diagrama de cajas del marcador DYS456 para las distintas regiones muestra que aparentemente no hay diferencia en sus medianas. Se evidencia que las tres regiones presentan datos atípicos, mientras el Oriente presenta únicamente valores atípicos altos, la Sierra y la Costa presentan valores atípicos altos y bajos, véase **Figura 19B**.

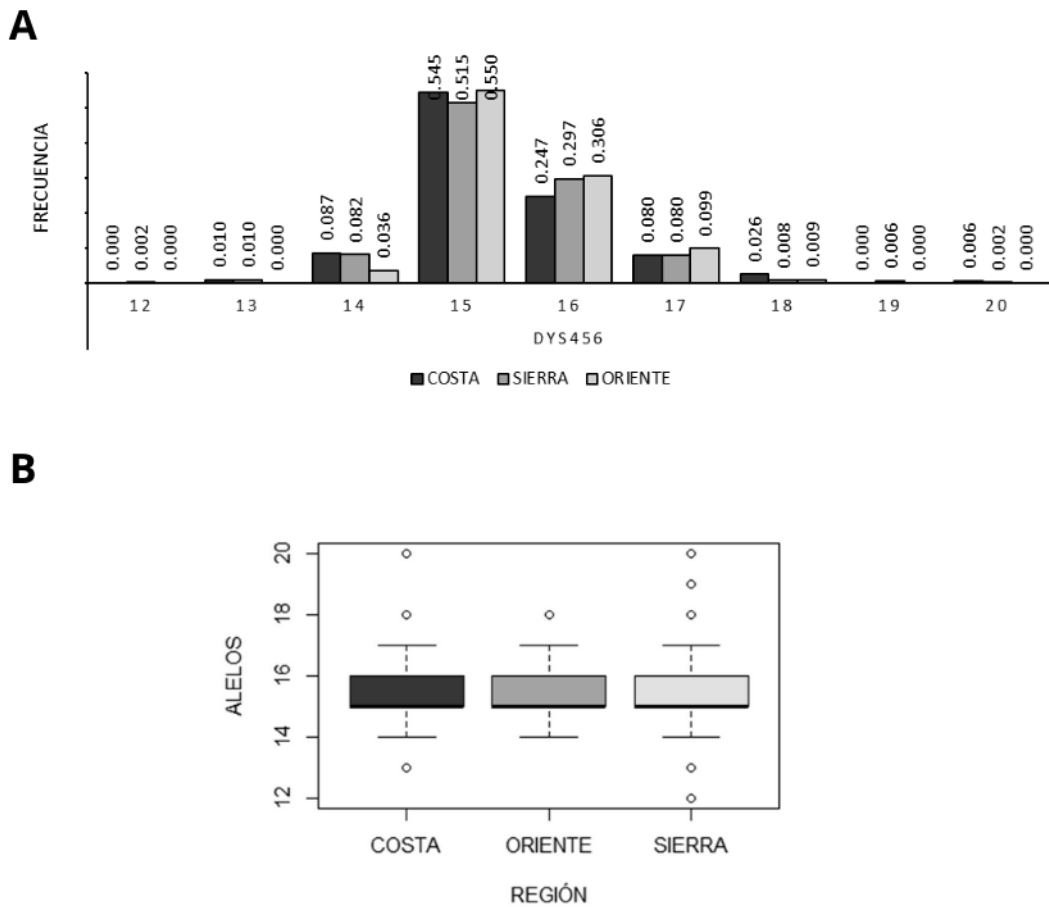


Figura 19. Gráficas del marcador DYS456 para las tres regiones de estudio. **A:** frecuencias alélicas, **B:** alelos.

El marcador Y-GATA-H4 registra una escalera alélica que va desde el alelo 9 hasta el alelo 16. Este marcador tiene dos alelos frecuentes en las tres regiones: alelo 11 y alelo 12. Las frecuencias para el alelo 11 son: Costa 0.481, Sierra 0.426 y Oriente 0.491. Mientras que, las frecuencias para el alelo 12 son: Costa 0.433, Sierra 0.469 y Oriente 0.384. Por otro lado, el alelo menos frecuente es el 16 que está presente únicamente en la Costa con una frecuencia de 0.003, véase **Figura 20A**. La exploración gráfica realizada con un diagrama de cajas del marcador Y-GATA-H4 para las distintas regiones muestra que aparentemente no hay diferencia en sus medianas al comparar Costa y Oriente. Sin embargo, hay una clara diferencia en la mediana de la región Sierra. Se evidencia que únicamente la Costa y la Sierra presentan datos atípicos (altos y bajos), véase **Figura 20B**.

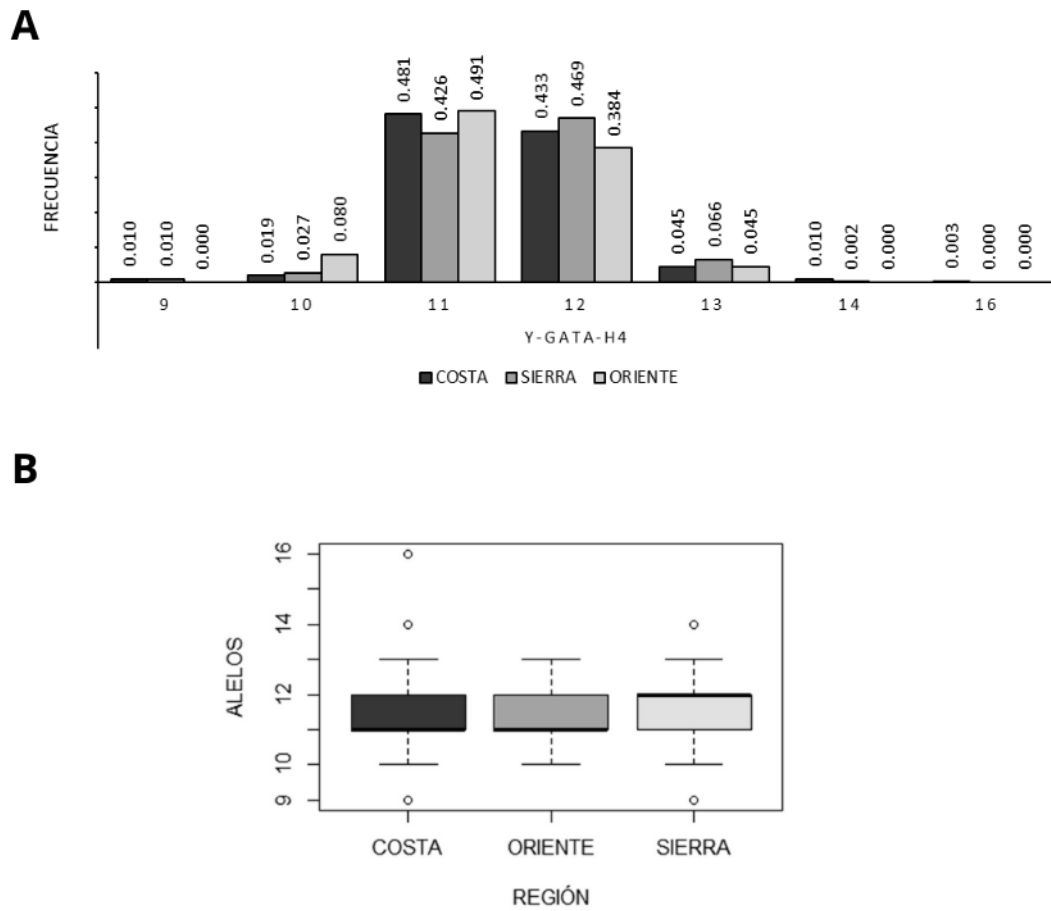


Figura 20. Gráficas del marcador Y-GATA-H4 para las tres regiones de estudio. **A:** frecuencias alélicas, **B:** alelos.

El marcador DYS385 registra una escalera alélica que va desde el alelo 10 hasta el alelo 20. En este marcador la combinación más frecuente es 11,13 (a,b), en las tres regiones: Costa 0.0225, Sierra 0.0195 y Oriente 0.0268, véase **Figura 21**.

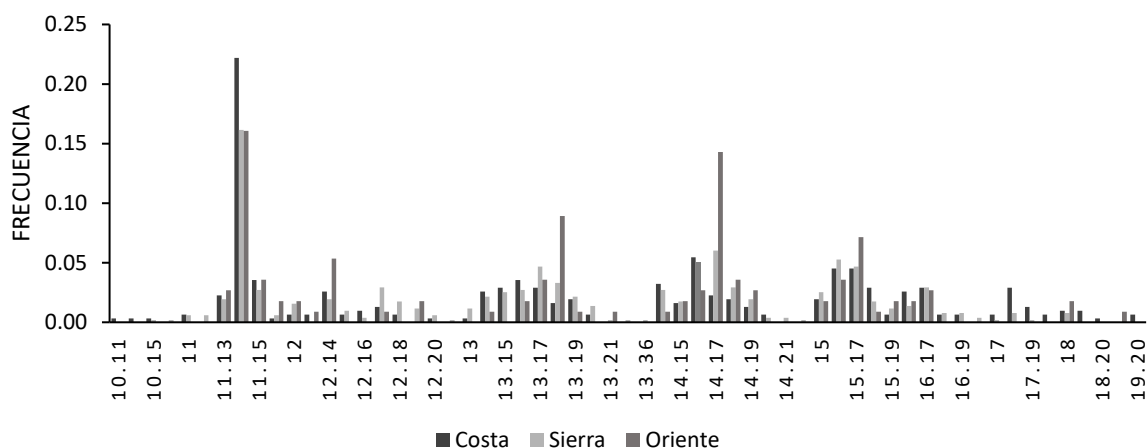


Figura 21. Frecuencia alélica del marcador DYS385 (a . b) para las tres regiones de estudio.

La relación entre las regiones de estudio se estableció a través del índice de Nei, el que nos indica que la distancia genética entre regiones es pequeña, ya que el valor más alto alcanzado es de 0.041 al relacionar la región Costa con la región Oriente, véase **Tabla 6**.

Tabla 6. Matriz de población por pares de distancia genética Nei.

| | COSTA | SIERRA | ORIENTE |
|----------------|--------------|---------------|----------------|
| COSTA | 0,000 | | |
| SIERRA | 0,022 | 0,000 | |
| ORIENTE | 0,041 | 0,013 | 0,000 |

La prueba de Kruskal-Wallis nos muestra una diferencia significativa entre los rangos con un nivel de confianza del 95%. Se evidencia de esta manera que la mayoría de marcadores no presentan diferencia significativa entre regiones, con excepción de: DYS576, DYS389I, DYS389II, DYS19, DYS437, DYS392, DYS393. Los marcadores DYS576, DYS19, DYS437, DYS392 muestran que el Oriente y la Sierra no presentan diferencias significativas en la información alélica recopilada; mientras que, la Costa, mantiene una amplia diferencia de rango al compararse con las otras regiones, siendo el grupo significativamente diferente. Tanto el marcador DYS389I como el DYS389II muestran que hay diferencia significativa entre la Costa y la Sierra; y que el Oriente resulta ser una combinación de las otras dos regiones de estudio. Finalmente, el marcador DYS393 muestra que hay diferencia

significativa entre la Costa y el Oriente; y que la Sierra resulta ser una combinación de las otras dos regiones de estudio, véase **Tabla 7**.

Tabla 7. Rango y grupos resultantes del análisis de Kruskal-Wallis para los diferentes marcadores analizados. Donde: R= región, M= marcador.

| | R / M | DYS576 | DYS389I | DYS448 | DYS389II | DYS19 | DYS391 | DYS481 |
|---------------|---------|----------|----------|----------|----------|-----------|--------------|----------|
| RANGO | Costa | 437.7463 | 437.3997 | 489.2625 | 441.1382 | 547.0991 | 495.3909 | 456.2107 |
| | Sierra | 494.0168 | 515.7296 | 473.0798 | 512.3138 | 451.75 | 481.1903 | 490.3521 |
| | Oriente | 526.7589 | 470.817 | 509.5714 | 475.8125 | 437.8438 | 458.1295 | 491.9152 |
| GRUPOS | Costa | a | a | a | a | a | a | a |
| | Sierra | b | b | a | b | b | a | a |
| | Oriente | b | ab | a | ab | b | a | a |
| | R / M | DYS533 | DYS438 | DYS437 | DYS570 | DYS635 | DYS390 | DYS439 |
| RANGO | Costa | 467.1791 | 491.4426 | 519.7035 | 474.5105 | 483.4206 | 458.5236 | 455.9059 |
| | Sierra | 473.8501 | 478.9796 | 466.466 | 469.0998 | 482.7398 | 493.6793 | 493.9512 |
| | Oriente | 521.027 | 484.4911 | 452.2455 | 418.6036 | 491.5536 | 516.5179 | 510.8839 |
| GRUPOS | Costa | a | a | a | a | a | a | a |
| | Sierra | a | a | b | a | a | a | a |
| | Oriente | a | a | b | a | a | a | a |
| | R / M | DYS392 | DYS393 | DYS458 | DYS456 | Y-GATA-H4 | DYS385 (a,b) | |
| RANGO | Costa | 397.7163 | 496.1330 | 450.3317 | 459.8045 | 464.4792 | 468.1447 | |
| | Sierra | 499.4149 | 465.8614 | 459.1463 | 469.5408 | 484.3663 | 467.4319 | |
| | Oriente | 520.4598 | 420.4643 | 461.4864 | 496.5631 | 423.3884 | 478.5714 | |
| GRUPOS | Costa | a | a | a | a | a | a | |
| | Sierra | b | ab | a | a | a | a | |
| | Oriente | b | b | a | a | a | a | |

Alelos intermedios

El estudio realizado registró que de los 20 marcadores con los que se trabajó, los marcadores DYS389I, DYS448, DYS391, DYS533, DYS437, DYS570, DYS635, DYS392, DYS393, DYS458, DYS385 a/b, DYS456 poseían alelos intermedios. El marcador con más alelos intermedios fue el DYS458, con 29 alelos intermedios, véase **Tabla 8**.

Tabla 8.- Alelos intermedios para los diferentes marcadores analizados.

| MARCADOR | REPETICIÓN | ALELO INTERMEDIO |
|--------------|------------|------------------|
| DYS389I | 1 | 13.3 |
| DYS448 | 1 | 18.2 |
| DYS391 | 1 | 8.3 |
| DYS533 | 1 | 12.3 |
| DYS437 | 1 | 13.2 |
| DYS570 | 1 | 18.3 |
| DYS635 | 1 | 21.3 |
| DYS392 | 1 | 13.2 |
| DYS393 | 1 | 12.1 |
| | 13 | 17.2 |
| | 4 | 14.1 |
| | 4 | 18.2 |
| DYS458 | 3 | 16.2 |
| | 2 | 15.1 |
| | 2 | 19.2 |
| | 1 | 16.3 |
| DYS385 (a,b) | 2 | 11/13.2 |
| | 1 | 11/13.3 |
| DYS456 | 1 | 16.3 |
| | 2 | 15.3 |

CAPÍTULO IV: INTERPRETACIÓN Y DISCUSIÓN

4.1 Análisis poblacional

En este estudio, se compiló y analizó 968 haplotipos de marcadores STR del Cromosoma Y de la población ecuatoriana, siendo el primer estudio de esta magnitud realizado en Ecuador enfocado en toda la población y no en grupos específicos [65–70]. De éstos 890 son haplotipos únicos (96.30%), 25 duplicados (2.70%), 8 triplicados (0.90%) y 1 cuadruplicado (0.10%), (**Tabla 3**). Estos valores coinciden con lo observado por Adnan, *et al.* [71] en un estudio de magnitud similar, donde analizó 855 haplotipos de los cuales apenas 835 (94.99%) eran únicos y un 20 (5.01%) mostraban repeticiones. Por otro lado, la diversidad haplotípica (DH) fue de 0.999, el poder de discriminación haplotípico (PD) fue de 0.954, con su correspondiente probabilidad de coincidencia de 0.046 (**Tabla 3**), lo cual también coincide con estudios similares [69,71].

Si bien esta probabilidad de coincidencia puede resultar baja al considerar que una paternidad biológica acreditada requiere una probabilidad mayor al 99,9% [72], se debe considerar que el valor obtenido hace referencia general al estudio realizado y a la población analizada, mientras que el análisis de paternidad considera la frecuencia de cada alelo para los marcadores analizados, las tasas de mutación y al grado de parentesco [69].

Por otro lado, es importante considerar que la cantidad de marcadores analizados está directamente relacionada con la capacidad de discriminación, cómo lo demuestra Adnan, *et al.* [71] en un estudio comparativo del sistema Goldeneye[®] de 20 Y-STR que alcanza un PD de 97,27 % con Y-filer[®] de 17 Y-STR en donde alcanza un PD de 96,93 %, con lo que concluye que el uso de un conjunto extendido de marcadores mejora el análisis.

Las frecuencias alélicas de la población ecuatoriana para los 20 marcadores STR de Cromosoma Y analizados se presentan en la **Tabla 4**. La escalera alélica presente para cada marcador fue comparada con la información de la base de datos de referencia de haplotipos de Cromosoma Y o YHRD (<https://yhrd.org>), validando así la información generada. El rango de alelos para cada marcador va desde 5 para DYS437 hasta 14 para DYS570. De estos, los alelos más frecuentes para los distintos marcadores son: 11 y 12;

alelos que ya han sido reportados como los más frecuentes tanto para la población Amerindia [48] como para países extranjeros como Iraq [74]; por otro lado, los alelos menos frecuentes son: 7, 8, de igual manera coincidentes con los alelos menos frecuentes de la población Amerindia [48].

Los valores estadísticos de relevancia en el campo forense se presentan para cada marcador en **la Tabla 5**. El número de alelos efectivos se refiere al número de alelos de igual frecuencia que serían necesarios para obtener la misma heterocigosidad que en la matriz real de alelos con representación (frecuencia) desigual. El índice de información de Shannon, proporciona una medida conveniente de diferenciación entre poblaciones, y también se puede utilizar para diferenciar frecuencias de alelos en cada locus. La diversidad génica se refiere a la diversidad de genes, o el número de alelos por locus. Además, se presenta el número de alelos, el poder de discriminación y la probabilidad de coincidencia. Se encontró que los marcadores más informativos son DYS576, DYS389II, DYS481, DYS570, DYS392, DYS458 y DYS385, que superan un poder de discriminación (PD) de 0.70. Sin embargo, de estos marcadores se destaca el DYS385 un poder de discriminación de 0.944. Este marcador ha sido descrito como el más informativo por estudios previos, tanto en Ecuador como en países extranjeros, esto debido a que se encuentra duplicado [64,65,67].

4.2 Análisis intra-poblacional

De los 968 perfiles genético recopilados se diferenciaron por región de la siguiente manera: Costa 340 perfiles, Sierra 516 perfiles y Oriente 112 perfiles. Los diagramas de barras presentados para cada marcador muestran que para la mayoría de marcadores se presenta una distribución unimodal (en las tres regiones), con excepción de DYS481 y DYS392 que presentan una distribución multimodal. Por otro lado, los diagramas de cajas permiten observar que en la mayoría de marcadores no hay diferencia en sus medianas, con excepción de DYS19, DYS458 Y Y-GATA-H4; y que todos los marcadores presentan datos atípicos. Sin embargo, estas distribuciones y datos atípicos, ya se han reportado previamente y se consideran normales dentro de una población como lo reporta Díaz-Sarmiento [47]. Estas variaciones se atribuyen a la evolución biológica, que a su vez está sujeta a las mutaciones, la deriva génica, la selección natural y a la migración [75].

El análisis de distancia genética entre las regiones de estudio se realizó a través del índice de Nei. El índice de Nei permite analizar la distancia genética entre poblaciones por pares y varía de 0 a 1 [76]. La matriz triangular que se obtuvo del análisis indica que existe una distancia genética mínima entre las regiones estudiadas, ya que el valor más alto alcanzado es de 0.041 al relacionar la región Costa con la región Oriente, mientras que el valor menor de diferenciación es de 0.013 al relacionar la región Sierra con la región Oriente, **Tabla 6**. Para añadirle significancia estadística a esta matriz se aplicó el método de Kruskal-Wallis para cada marcador, obteniendo que para la mayoría de marcadores no hay diferencia significativa entre regiones. Un estudio previo realizado en Ecuador encontró haplotipos compartidos entre mestizos, afroecuatorianos y kichwas[70], lo que confirma lo previamente establecido. Sin embargo, se encontró que los marcadores DYS576, DYS389I, DYS389II, DYS19, DYS437, DYS392, DYS393 registran diferencias significativas entre regiones. Mientras para los marcadores DYS576, DYS19, DYS437, DYS392 el grupo significativamente diferente fue la Costa; para los marcadores DYS389I y DYS389II el Oriente resulta ser una combinación de las otras dos regiones de estudio; y para el marcador DYS393 la Sierra resulta ser una combinación de la Costa y el Oriente, **Tabla 7**. De esta manera encontramos que la variación encontrada por el índice de Nei, se debe a estos marcadores.

Los alelos intermedios registrados, **Tabla 8**, fueron comparados con la información de la base de datos de referencia de haplotipos de Cromosoma Y o YHRD (<https://yhrd.org>), todos han sido previamente reportados.

CAPÍTULO V: CONCLUSIONES Y RECOMENDACIONES

La información presentada genera una herramienta que permite otorgar significancia estadística a los análisis de haplotipo de Cromosoma Y realizados en la población ecuatoriana. Fortalece los estudios e investigaciones previos sobre las frecuencias alélicas del Cromosoma Y en Ecuador. Además, se generó la primera base de datos de haplotipos de Cromosoma Y del SNMLCF, lo sienta una base para futuras investigaciones como los análisis comparativos de frecuencias alélicas con los países con los que existe flujo migratorio y para la generación de una amplia base de STR autosómicos.

REFERENCIAS

1. Sanchez-Castro EE, Pajuelo-Reyes C. Importancia de la genética como ciencia en relación a la pandemia de Covid-19. *Revista de la Facultad de Medicina Humana*. 2020;20: 690–695. Doi:10.25176/RFMH.V20I4.2944
2. Pierce BA. *Genética: Un enfoque conceptual*. 2009.
3. Sturtevant A. *An introduction to genetics*. Philadelphia; London: W.B. Saunders Company; 1939.
4. Eguiarte LE, Aguirre-Liguori JA, Jardón-Barbolla L, Aguirre-Planter E, Souza V. Genómica de Poblaciones: Nada en Evolución va a tener sentido si no es a la luz de la Genómica, y nada en Genómica tendrá sentido si no es a la luz de la Evolución. *TIP*. 2013;16: 42–56. Doi:10.1016/S1405-888X(13)72077-1
5. Herrera Paz EF. La genética de poblaciones y el origen de la diversidad humana. *Revista Médica Hondureña*. 2013.
6. Almarza J. *El Genoma Humano*. Mérida; 2004.
7. Gibbs RA. The Human Genome Project changed everything. *Nature Reviews Genetics* 2020 21:10. 2020;21: 575–576. Doi:10.1038/s41576-020-0275-3
8. Craig-Venter J, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. *Science*. 2001;291: 1304–1351. Doi:10.1126/SCIENCE.1058040
9. Williams CR. *El Genoma Humano*. *Revista Médica de Honduras*. 2002.
10. Riesgo JR. *Qué es el genoma humano*. *Ciencia*. 2002.
11. Ross DM, Álvarez-Álvarez L, Chávez-Viamontes JÁ, Pérez LM, Alberro M, Góngora OL. *Genoma humano. Actualidades y perspectivas bioéticas. (Ensayo I)*. 2002.
12. Burke B, Zonno K, Farrell J, Ingham C, Larson F, Lavochkin M. *UNDERSTANDING GENETICS*. 2007.
13. Schleif R. *Genetics and Molecular Biology*. 1993.

14. National Human Genome Research Institute. ADN (Ácido Desoxirribonucleico). 2019.
15. Ingman M, Kaessmann H, Pääbo S, Gyllensten U. Mitochondrial genome variation and the origin of modern humans. *Nature*. 2000;408: 708–713. Doi:10.1038/35047064
16. Copelli S. *Genética: Desde la herencia a la manipulación de genes*. 1a ed. Buenos Aires: Fundación de Historia Natural Félix de Azara; 2010.
17. Valdespino Gómez VM, Valdespino Castillo PM, Valdespino Castillo VE. Organización estructural y funcional del genoma humano: variación en el número de copias predisponentes de enfermedades degenerativas. *Gaceta Mexicana de Oncología*. 2013.
18. Weiner AM. SINEs and LINEs: the art of biting the hand that feeds you *Current opinion in cell biology*. 2002;14: 343–350. Doi:10.1016/S0955-0674(02)00338-1
19. Debrauwere H, Gendrel CG, Lechat S, Dutreix M. Differences and similarities between various tandem repeat sequences: Minisatellites and microsatellites. *Biochimie*. 1997;79: 577–586. Doi:10.1016/S0300-9084(97)82006-8
20. Liehr T. *Cytogenomics*. 2020; 409.
21. Liehr T. *Benign and Pathological Chromosomal Imbalances: Microscopic and Submicroscopic Copy Number Variations (CNVs) in Genetics and Counseling*. 2014.
22. Esparza-García E, Cárdenas-Conejo A, Huicochea-Montiel J, Aráujo-Solís MA. *Cromosomas, cromosomopatías y su diagnóstico*. Instituto Mexicano del Seguro Social, México. 2017;84: 30–39.
23. National Human Genome Research Institute. *Cromosomas*. 2019.
24. Bueno ML. *CROMOSOMAS, VEHÍCULOS EN LA ORGANIZACIÓN Y TRANSMISIÓN DE LOS CARACTERES*. 2011.
25. Kayser M. Forensic use of Y-chromosome DNA: a general overview. *Human Genetics*. 2017;136: 621. Doi:10.1007/S00439-017-1776-9

26. Court DS. The Y chromosome and its use in forensic DNA analysis. *Emerging Topics in Life Sciences*. 2021;5: 427. Doi:10.1042/ETLS20200339
27. Kayser M, Kittler R, Eler A, Hedman M, Lee AC, Mohyuddin A, et al. A Comprehensive Survey of Human Y-Chromosomal Microsatellites. *American Journal of Human Genetics*. 2004;74: 1183. Doi:10.1086/421531
28. Butler JM, Schoske R, Vallone PM, Kline MC, Redd AJ, Hammer MF. A novel multiplex for simultaneous amplification of 20 Y chromosome STR markers. *The New England journal of medicine* 2002;129: 10–24. Doi:10.1016/S0379-0738(02)00195-0
29. Butler JM, Devaney JM, Marino MA, Vallone PM. Quality control of PCR primers used in multiplex STR amplification reactions. *Forensic science international*. 2001;119: 87–96. Doi:10.1016/S0379-0738(00)00412-6
30. Housman D. Human DNA polymorphism. *The New England journal of medicine*. 1995;332: 318–320. Doi:10.1056/NEJM199502023320508
31. Weber JL. Human DNA polymorphisms and methods of analysis. *Current opinion in biotechnololy*. 1990;1: 166–171. Doi:10.1016/0958-1669(90)90026-H
32. Trent RJ. Genes, Environment and Inheritance. *Molecular Medicine*. 2012; 39–79. Doi:10.1016/B978-0-12-381451-7.00002-5
33. Quesada-García C, López-Palafox J. HISTORIA DE LA IDENTIFICACIÓN PERSONAL: DESDE EL RECONOCIMIENTO FACIAL HASTA EL ADN DENTAL. *Biociencias*. 2019.
34. Silveyra JO, Silveyra JO, Silveyra P. *Sistemas de identificación humana*. Buenos Aires: La Rocca; 2006.
35. Aaronson SA, Abedon ST, Adams DJ, Adhya S, Aguilera A, Ahn Y, et al. *Brenner's Encyclopedia of Genetics*. Segunda Edición. 2013.
36. Ely Hepfer C, Piperberg JB, Farganis GM. An Introduction to DNA Fingerprinting. *American Biology Teacher*. 1993;55: 216–221. Doi:10.2307/4449636

37. Hernández-Rodríguez AW, Trejo-Medinilla F de M. Estudio Genético Poblacional de Frecuencias Alélicas para 15 marcadores STR presentes en la Población del Estado de Zacatecas Aplicado a la Práctica Forense. *iMedPub Journals*. 2014.
38. Bravo-Aguilar MLJ. La verdad genética de la paternidad. Primera. 2009.
39. Villacís B, Carrillo D. País atrevido: la nueva cara sociodemográfica del Ecuador. Edición especial. Quito; 2012.
40. Tapia-Ortiz MM. POBLACIÓN EN EL ESPACIO RURAL Y URBANO CON BASE EN EL CRECIMIENTO DEMOGRÁFICO Y SU IMPACTO EN EL TERRITORIO. 2015.
41. Zambrano AK, Gaviria A, Cobos-Navarrete S, Gruezo C, Rodríguez-Pollit C, Armendáriz-Castillo I, et al. The three-hybrid genetic composition of an Ecuadorian population using AIMs-InDels compared with autosomes, mitochondrial DNA and Y chromosome data. *Scientific Reports*. 2019;9: 1–8. Doi:10.1038/s41598-019-45723-w
42. UTE. El origen genético de los ecuatorianos se investiga en la UTE. 2018.
43. INEC. Registro Estadístico de Entradas y Salidas Internacionales 2019. 2020.
44. INEC. Registro Estadístico de Entradas y Salidas Internacionales. 2021.
45. INEC. Registro Estadístico de Entradas y Salidas Internacionales 2018. 2019.
46. Delgado E, Neyra CD. Allele frequencies of 21 autosomal STR markers in a mixed race Peruvian population applied to forensic practice. *Spanish Journal of Legal Medicine*. 2019;45: 92–97. Doi: 10.1016/J.REMLE.2018.09.001
47. Díaz-Sarmiento LF. Análisis de 17 loci de STR de cromosoma y en las poblaciones de Bogotá y Santander con fines genético poblacionales y forenses. 2010.
48. López-González PW. Detección de la variabilidad genética en la población de Puno utilizando marcadores STR DYS390, DYS391, DYS392 del cromosoma Y. 2004.

49. Rojas KM, Roa M, Briceño I, Guaneme C, Gómez A. Polimorfismos de 17 marcadores STR del cromosoma-Y en una muestra poblacional del altiplano cundiboyacense. Colombia Médica. 2010.
50. González-Andrade F, Bolea M, Martínez-Jarreta B, Sánchez D. Genetic analysis of autosomic and Y-chromosome STR-PCR polymorphisms of the Huaoranis, the last nomad tribe in Ecuador. International Congress Series. 2006;1288: 331–333. Doi: 10.1016/J.ICS.2005.12.045
51. Gaviria A, Sánchez ME, Morejón G, Vela M, Aguirre V, Burgos G, et al. Characterization and Haplotype analysis of 11 Y-STR loci in Ecuadorian population. Forensic Science International: Genetics Supplement Series. 2013;4: e310–e311. Doi: 10.1016/J.FSIGSS.2013.10.158
52. Johnson P, Williams R, Martin P. Genetics and Forensics: Making the National DNA Database. Science studies (Helsinki, Finland). 2003;16: 22.
53. Hombreirol L. The national database of genetic profiles. Regulation, functioning and operating. 2014.
54. Concheiro L. Tipos de polimorfismos. In: Instituto de Ciencias Forenses.
55. ICRC. Missing People, DNA Analysis and Identification of Human Remains: A Guide to Best Practice in Armed Conflicts and Other Situations of Armed Violence | International Committee of the Red Cross. 2009.
56. Gusmão L, Butler JM, Carracedo A, Gill P, Kayser M, Mayr WR, et al. DNA Commission of the International Society of Forensic Genetics (ISFG): an update of the recommendations on the use of Y-STRs in forensic analysis. International journal of legal medicine. 2006;120: 191–200. Doi:10.1007/S00414-005-0026-1
57. Roewer L, Andersen MM, Ballantyne J, Butler JM, Caliebe A, Corach D, et al. DNA commission of the International Society of Forensic Genetics (ISFG): Recommendations on the interpretation of Y-STR results in forensic analysis. Forensic Science International: Genetics. 2020;48: 102308. Doi: 10.1016/J.FSIGEN.2020.102308

58. CICR. Directrices para el uso de la genética forense en investigaciones sobre Derechos Humanos y derecho internacional humanitario. 2015.
59. Thermo Fisher Scientific. PrepFiler Express™ and PrepFiler Express BTA™ Forensic DNA Extraction Kits USER GUIDE. 2017.
60. Promega Corporation. PowerPlex® Y23 System for Use on the Applied Biosystems® Genetic Analyzers. 2021.
61. Thermo Fisher Scientific. Yfiler™ Plus PCR Amplification Kit USER GUIDE. 2016.
62. Applied Biosystems. Applied Biosystems 3130 and 3130xl Genetic Analyzers. 2010.
63. Peakall R, Smouse PE. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* 6, 288-295. 2006.
64. Peakall R, Smouse PE. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research – an update. *Bioinformatics*. *Bioinformatics* 28, 2537-2539. 2012.
65. Baeza C, Guzmán R, Tirado M, López-Parra AM, Rodríguez T, Mesa MS, et al. Population data for 15 Y-chromosome STRs in a population sample from Quito (Ecuador). *Forensic Science International*. 2007;173: 214–219. Doi: 10.1016/J.FORSCIINT.2006.09.011
66. Villaescusa P, Seidel M, Nothnagel M, Pinotti T, González-Andrade F, Alvarez-Gila O, et al. A Y-chromosomal survey of Ecuador's multi-ethnic population reveals new insights into the tri-partite population structure and supports an early Holocene age of the rare Native American founder lineage C3-MPB373. *Forensic Science International: Genetics*. 2021;51. Doi: 10.1016/J.FSIGEN.2020.102427
67. González-Andrade F, Bolea M, Martínez-Jarreta B, Sánchez D. Genetic analysis of autosomic and Y-chromosome STR-PCR polymorphisms of the Huaorani, the last nomad tribe in Ecuador. *International Congress Series*. 2006;1288: 331–333. Doi: 10.1016/J.ICS.2005.12.045

68. González-Andrade F, Sánchez-Q D, Martínez-Jarreta B. Genetic profile of the Ecuadorian Mestizo population (Ecuador-South America) by using the Power Plex 16 System Kit. *Forensic Science International*. 2003. Doi:10.1016/S0379-0738(03)00165-8
69. González Andrade F. Análisis molecular de variación de polimorfismos STR autosómicos y de cromosoma «Y» en grupos étnicos de Ecuador - Dialnet. 2006.
70. González-Andrade F, Roewer L, Willuweit S, Sánchez D, Martínez-Jarreta B. Y-STR variation among ethnic groups from Ecuador: Mestizos, Kichwas, Afro-Ecuadorians and Waoranis. *Forensic Science International Genetics*. 2009;3. Doi: 10.1016/J.FSIGEN.2008.08.003
71. Adnan A, Kasimu K, Rakha A, He G, Yang T, Wang CC, et al. Comprehensive genetic structure analysis of Han population from Dalian City revealed by 20 Y-STRs. *Molecular genetics & genomic medicine*. 2020;8. Doi:10.1002/MGG3.1149
72. Lagos M, Poggi H, Mellado C. Conceptos básicos sobre el estudio de paternidad. *LABORATORIO CLÍNICO Revista Med Chile*. 2011;139: 542–547.
73. Purps J, Siegert S, Willuweit S, Nagy M, Alves C, Salazar R, et al. A global analysis of Y-chromosomal haplotype diversity for 23 STR loci. *Forensic Science International: Genetics*. 2014;12: 12–23. Doi: 10.1016/J.FSIGEN.2014.04.008
74. Hameed IH, Ommer AJ, Murad AF, Mohammed GJ. Allele frequency data of 21 autosomal short tandem repeat loci in Mesan and Basra provinces in South Iraq. *Egyptian Journal of Forensic Sciences*. 2015;5: 150–156. Doi: 10.1016/J.EJFS.2014.10.003
75. Charlesworth B, Charlesworth D. *Elements of Evolutionary Genetics*. 2010.
76. Nei M. Genetic distance between populations. *The American Naturalist*. 1972;106: 283–292.