

ESTRUCTURA GENÉTICA DE LA POBLACIÓN PANAMEÑA

*Genes Ancestrales e Implicaciones
Biomédicas*

EDGARDO CASTRO PÉREZ

MAGALI DE CHIAL

OLGA CHEN

BRENDA DE MAYORGA

CARLOS RAMOS DELGADO

ESTRUCTURA GENÉTICA DE LA POBLACIÓN PANAMEÑA

Genes Ancestrales e Implicaciones Biomédicas

© EDGARDO CASTRO PÉREZ / MAGALI DE CHIAL/ OLGA CHEN/BRENDA DE MAYORGA y
CARLOS RAMOS DELGADO

1era Edición 2025

<https://doi.org/10.5281/zenodo.14052378>

575.15 Castro Pérez, Edgardo
C279 Estructura genética de la población panameña : genes
ancestrales e implicaciones biomédicas / Edgardo Castro
Pérez, Magali De Chial, Olga Chen, Brenda Mayorga y Carlos
Ramos Delgado. -- Panamá: Editorial Portobelo, 2024.
248 páginas ; 25 cm.

ISBN 978-9962-15-187-6

1. GENÉTICA HUMANA

2. GENÉTICA DE LA POBLACIÓN HUMANA I. TÍTULO

Diagramación:

Silvio Sequera

Imagen de Portada:

Dr. Carlos Ramos

Montaje de Portada:

Yessenia de Gracia

Edición e impresión al cuidado de:

Cultural Portobelo

Vía Argentina No. 83 - Edificio Tang - Planta baja

Panamá, República de Panamá

Teléfonos: (507) 269-9493/6678-7983

Dirección electrónica: www.culturalportobelo.com

info@culturalportobelo.com

Facebook: [@culturalportobelo](https://www.facebook.com/culturalportobelo) / Instagram: [@culturalportobelo](https://www.instagram.com/culturalportobelo)

Todos los derechos reservados.

Queda prohibida, sin autorización escrita de los titulares de los derechos de autor, bajo las sanciones contempladas por las leyes, la reproducción parcial o total de esta obra por cualquier medio o procedimiento, incluida la fotocopia, el procesamiento informático y la distribución de ejemplares de esta obra mediante alquiler o préstamo.

Hecho el depósito legal.

**ESTRUCTURA
GENÉTICA DE
LA POBLACIÓN
PANAMEÑA**

*Genes Ancestrales e Implicaciones
Biomédicas*

CONTENIDO

Capítulo 1

- Ascendencia genética de la población panameña:
Estructura polimórfica, genes amerindios chibchas y
perspectivas biológicas sobre las enfermedades.** 17

*Edgardo Castro-Pérez, Diomedes E Trejos, Tomas Hrbek,
Vijayasarah Setaluri y Carlos Ramos Delgado.*

Capítulo 2

- Polimorfismo genético y parámetros forenses de
nueve loci de repeticiones cortas en tándem en
amerindios Ngöbe y Emberá de Panamá.** 53

*Edgardo Castro-Pérez, Diomedes E. Trejos, Vicente Berovides-
Alvarez, Tomás D. Arias y Carlos Ramos Delgado.*

Capítulo 3

- Análisis de 30 marcadores INDELS polimórficos en la
población panameña: estimaciones de mezcla racial,
estructura poblacional y parámetros forenses.** 75

*Carlos Ramos Delgado, Edgardo Castro-Pérez, Concepción Molina-
Jirón C y Diomedes E. Trejos.*

Capítulo 4

Asociación clínica inmunológica y molecular de los alelos DRB1, DQA1 y DQB1 con la diabetes mellitus tipo 1 en la población de Panamá. 97

Brenda de Mayorga y Olga Chen.

Capítulo 5

Polimorfismo genético y actividad de la paraoxonase en la población panameña. 113

Ana E. Tejada E, Tomás A. Diez, Gonzalo Carrasco y Carlos Ramos Delgado.

Capítulo 6

Determinación de alelos HLA-DRB1 en pacientes panameños con artritis reumatoidea severa. 131

Diana Mora, Orlando Montenegro y Carlos Ramos Delgado.

*Capítulo 7***Asociación del polimorfismo pro12ala en el receptor activado por proliferadores de peroxisomas gamma-2 (PPAR γ 2) en una muestra de la población panameña.** 171

Evelyn Visuetti, Tomás A. Diez, Magali de Chial, Edgardo Castro-Pérez y Carlos Ramos Delgado.

*Capítulo 8***Secuenciación genómica del componente ancestral amerindio de la población panameña revela factores de riesgo genético a enfermedades no transmisibles.** 223

Edgardo Castro-Pérez, Carlos Ramos Delgado, Lucía Jorge-Nebert, Beatrice Di Biase, Kathia Canto-Díaz, Madián Poveda, Germán Espino, Vijayasradhi Setaluri, Carmen Mela-Sánchez, Magaly de Chial, Olga Chen, Julio C. Castillo, Carlos M. Restrepo, Rolando Gittens, Diomedes E. Trejos, Jorge Lasso De La Vega, Brenda de Mayorga, Gumercindo Lorenzo, Ovidio Durán.

Capítulo 9

Identificación de dos SNPs en los genes MTHFR (rs1801133) y TP53 (rs1042522) como factores de riesgo a cánceres y otras enfermedades en amerindios Ngöbe de Panamá: significancia biomédica en la población mestiza panameña.

211

Edgardo Castro-Pérez, Carlos Ramos Delgado, Germán Espino, Beatrice Di Biase, Kathia Canto-Díaz, Madián Poveda, Caleb Herrera, Nathan Herrera, Vijayasaradhi Setaluri, Reynaldo Arosemena, Carmen Mela-Sánchez, Marina Pacheco, Jorge Lasso De La Vega, Pablo Noel, Fernando Díaz, Idalina Cubilla-Batista, Isabel Blanco José Robinson-Duggon.

PRÓLOGO

El universo de la genética es un ámbito muy especializado y por tanto limitado a un selecto y preparado grupo de investigadores, que nos ayudan a dilucidar los orígenes de nuestro grupos étnicos.

Es reconfortante encontrar que contamos con investigadores panameños que han dedicado su tiempo y su esfuerzo, en navegar por las difíciles aguas de los marcadores genéticos de nuestros grupos antropológicos, para definir caminos que nos permitan conocer nuestro pasado y dirigir nuestro presente y nuestro futuro.

Más importante aún, es que dichos esfuerzos, tienen una aplicación práctica inmediata. No se quedan en hipótesis potenciales, sino que muestran sus primeros resultados aplicables, develando la relación de la proporción de genes ancestrales y la incidencia de ciertas enfermedades en nuestra población.

Fueron necesarias más de dos décadas de investigación, aplicando las herramientas más novedosas en el campo de la genética, para lograr estos documentos de gran valía para nuestro mundo académico e investigativo. Con gran detalle, los investigadores utilizando marcadores moleculares, determinaron la composición genética de la población panameña y la proporción de genes ancestrales en las diferentes provincias.

Pero no solamente se estudió la genética de los grupos ancestrales, pues basados en la hipótesis de un aporte complementario de los genes europeos y el posterior aporte de los grupos africanos, se logra hacer un mapa detallado de los aportes de cada uno en esta mezcla de razas, que se han ido uniendo a los grupos originarios panameños, para crear la multiplicidad de grupos étnicos que conviven en nuestro país.

Un logro muy importante, además de analizar los grupos originarios, es que también se estudia y se llega a conclusiones válidas, sobre el aporte de grupos, en este interesante cóctel de razas, que por siglos se han ido integrando a la población panameña, tan diferente de los demás países del continente.

En su momento el famoso intelectual Demetrio Porras ya había planteado que la población panameña se ha ido enriqueciendo con los constantes “traumas” históricos que ha enfrentado. Como lugar de tránsito, nuestro país ha recibido oleadas de inmigrantes por múltiples razones, pero que han sembrado genes en momentos cumbres que han definido su identidad

actual: el descubrimiento de América por los españoles; la llegada de un grupo inmenso de nativos africanos para trabajar en las minas y las plantaciones; los aportes de los ciudadanos colombianos cuando Panamá formó parte de la Gran Colombia y la Nueva Granada; el descubrimiento del oro en California que convirtió en realidad el construir el ferrocarril y que conllevó la llegada de norteamericanos, chinos, caribeños y de muchos países; la construcción del canal francés que implicó sangre europea nueva además de los grupos anteriores del ferrocarril; y finalmente, el aporte norteamericano y de muchos otros países para la construcción del Canal de Panamá.

Pero no nos alejemos del tema principal del libro. La mayor preocupación de los investigadores fue hacer una comprobación científica de la genética ancestral, con los recurrentes síntomas de enfermedades que afectan a grupos humanos específicos. Es así, como la mayor parte de las investigaciones reunidas en este libro, van dirigidas a identificar variantes genéticas que pudieran estar asociadas en la población panameña a ciertas enfermedades.

En especial las diferencias en incidencia de enfermedades como el cáncer, la diabetes, la artritis reumatoidea y muchas otras enfermedades no transmisibles, podrían ser explicadas sobre la base de las diferencias en la composición genética.

Como vemos, hay mucho terreno que cubrir, y el esfuerzo de estos excelentes investigadores, es una muestra de todo el camino que nos queda por recorrer. Ojalá que su lectura sea un incentivo para nuevas y provechosas investigaciones que incrementen el conocimiento científico de nuestros orígenes. En cuanto a los autores, les pedimos que, este aporte, sea solo el primer escalón de nuevos proyectos en beneficio de nuestra gran Nación.

COLABORADORES

- Edgardo Castro-Pérez | Centro de Biología Celular y Molecular de las Enfermedades, INDICASAT-AIP, Clayton, Ciudad de Panamá, Panamá. • Departamento de Genética y Biología Molecular, Facultad de Ciencias Naturales, Exactas y Tecnología, Universidad de Panamá, Ciudad de Panamá, Panamá. • Sistema Nacional de Investigación (SNI), SENACYT, Clayton, Ciudad de Panamá, Panamá.
- Magaly de Chial | Departamento de Genética y Biología Molecular, Facultad de Ciencias Naturales Exactas y Tecnología, Universidad de Panamá, Ciudad de Panamá, Panamá.
- Olga Chen | Departamento de Genética y Biología Molecular, Facultad de Ciencias Naturales Exactas y Tecnología, Universidad de Panamá, Ciudad de Panamá, Panamá.
- Brenda de Mayorga | Departamento de Microbiología y Parasitología, Facultad de Ciencias Naturales Exactas y Tecnología, Universidad de Panamá, Ciudad de Panamá, Panamá.
- Carlos Ramos Delgado | Departamento de Genética y Biología Molecular, Facultad de Ciencias Naturales Exactas y Tecnología, Universidad de Panamá, Ciudad de Panamá, Panamá.
- Ana E. Tejada E | Departamento de Bioquímica, Facultad de Medicina, Universidad de Panamá, Ciudad de Panamá, Panamá.
- Beatrice Di Biase | Departamento de Genética y Biología Molecular, Facultad de Ciencias Naturales Exactas y Tecnología, Universidad de Panamá, Ciudad de Panamá, Panamá.
- Caleb Herrera | Departamento de Genética y Biología Molecular, Facultad de Ciencias Naturales, Exactas y Tecnología, Universidad de Panamá, Ciudad de Panamá, Panamá.
- Carlos M. Restrepo | Centro de Biología Celular y Molecular de las Enfermedades, INDICASAT-AIP, Clayton, Ciudad de Panamá, Panamá. Sistema Nacional de Investigación (SNI), SENACYT, Clayton, Ciudad de Panamá, Panamá.
- Carmen Mela-Sánchez | Centro de Biología Celular y Molecular de las Enfermedades, INDICASAT-AIP, Clayton, Ciudad de Panamá, Panamá.

- Concepción Molina-Jirón C | Departamento de Bioquímica, Facultad de Ciencias Naturales Exactas y Tecnología, Universidad de Panamá, Ciudad de Panamá, Panamá.
- Diana Mora | Bio Labs International S.A., Ciudad de Panamá, Panamá.
- Diomedes E. Trejos | Laboratorio de Análisis Biomolecular, Instituto de Medicina Legal y Ciencias Forenses, Ministerio Público, Clayton, Ciudad de Panamá, Panamá.
- Evelyn Visuetti | Departamento de Genética y Biología Molecular, Facultad de Ciencias Naturales Exactas y Tecnología, Universidad de Panamá, Ciudad de Panamá, Panamá.
- Fernando Díaz | Sección de Hematología, The Panama Clinic, Ciudad de Panamá, Panamá.
- Germán Espino | Departamento de Medicina, Facultad de Medicina, Universidad de Panamá, Ciudad de Panamá, Panamá. • Sección de Hematología, Hospital Santo Tomás, Ministerio de Salud, Ciudad de Panamá, Panamá. • Centro Hemato Oncológico Panamá, Ciudad de Panamá, Panamá.
- Gonzalo Carrasco | Departamento de Estadística, Facultad de Ciencias Naturales Exactas y Tecnología, Universidad de Panamá, Ciudad de Panamá, Panamá.
- Gumercindo Lorenzo | Instituto Nacional de Estadística y Censo, Contraloría General de la República, Ciudad de Panamá, Panamá. • Escuela de Estadística, Facultad de Ciencias Naturales, Exactas y Tecnología, Universidad de Panamá, Ciudad de Panamá, Panamá.
- Idalina Cubilla-Batista | Centro de Investigación Médica, Hospital Punta Pacífica (Pacífica Salud), Ciudad de Panamá • Hospital Rafael Estévez, Caja de Seguro Social, Aguadulce, Panamá. • Sistema Nacional de Investigación (SNI), SENACYT, Clayton, Ciudad de Panamá, Panamá.
- Isabel Blanco | Centro de Investigación Médica, Hospital Punta Pacífica (Pacífica Salud), Ciudad de Panamá.
- Jorge Lasso De La Vega | Instituto Oncológico Nacional, Ministerio de Salud, Ciudad de Panamá, Panamá.

- José Robinson-Duggon | Departamento de Bioquímica, Facultad de Ciencias Naturales Exactas y Tecnología, Universidad de Panamá, Ciudad de Panamá, Panamá.
- Julio C. Castillo | Centro de Biodiversidad y Descubrimiento de Drogas, INDICASAT-AIP, Clayton, Ciudad de Panamá, Panamá. • Departamento de Genética y Biología Molecular, Facultad de Ciencias Naturales Exactas y Tecnología, Universidad de Panamá, Ciudad de Panamá, Panamá. • Sistema Nacional de Investigación (SNI), SENACYT, Panamá.
- Kathia Canto Díaz | Departamento de Genética y Biología Molecular, Facultad de Ciencias Naturales, Exactas y Tecnología, Universidad de Panamá, Ciudad de Panamá, Panamá.
- Lucía Jorge-Nebert | Department of Environmental Health, and Center for Environmental Genetics (CEG), University of Cincinnati Medical Center, Cincinnati, OH, USA.
- Madián Poveda | Departamento de Genética y Biología Molecular, Facultad de Ciencias Naturales, Exactas y Tecnología, Universidad de Panamá, Ciudad de Panamá, Panamá.
- Marina Pacheco | Departamento de Patología, Complejo Hospitalario Metropolitano, Caja de Seguro Social, Ciudad de Panamá, Panamá. • Sistema Nacional de Investigación (SNI), SENACYT, Clayton, Ciudad de Panamá, Panamá. • Nathan Herrera | Departamento de Genética y Biología Molecular, Facultad de Ciencias Naturales, Exactas y Tecnología, Universidad de Panamá, Ciudad de Panamá, Panamá.
- Orlando Montenegro | Market Access Director for Acino Pharma LATAM Region, Ciudad de Panamá, Panamá.
- Ovidio Durán | Departamento de Genética y Biología Molecular, Facultad de Ciencias Naturales Exactas y Tecnología, Universidad de Panamá, Ciudad de Panamá, Panamá.
- Pablo Noel | Sección de Dermatología, Hospital Punta Pacífica (Pacífica Salud), Ciudad de Panamá, Panamá.
- Reynaldo Arosemena | Sección de Dermatología, Hospital Santo Tomás, Ministerio de Salud, Ciudad de Panamá, Panamá.

- Rolando Gittens | Centro de Biodiversidad y Descubrimiento de Drogas, INDICASAT-AIP, Clayton, Ciudad de Panamá, Panamá. • Sistema Nacional de Investigación (SNI), SENACYT, Clayton, Ciudad de Panamá, Panamá.
- Tomás A. Diez | Departamento de Bioquímica, Facultad de Medicina, Universidad de Panamá, Ciudad de Panamá, Panamá.
- Tomas Hrbek | Laboratório de Evolução e Genética Animal, Universidade Federal do Amazonas, Manaus, Brazil.
- Vicente Berovides-Alvarez | Laboratorio de Genética Ecológica, Facultad de Biología, Universidad de La Habana, La Habana, Cuba.
- Vijayasaradhi Setaluri | Department of Dermatology, University of Wisconsin School of Medicine and Public Health, Madison, WI, USA; William S. Middleton Memorial Veterans Hospital, Madison, WI, USA.
-

INTRODUCCIÓN

Este libro es el producto de más de 20 años de investigaciones sobre la estructura genética de la población panameña y las posibles implicaciones biomédicas de esta composición genética. La obra se centra en la utilización de marcadores moleculares polimórficos de tipo STRs, INDELS, SNPs para responder preguntas acerca de cuáles fueron los genes ancestrales que contribuyeron a la composición genética de la población panameña actual y la proporción de su contribución. Se demuestra que a pesar de que la población actual se ajusta a un modelo trihíbrido en el cual los genes de europeos principalmente españoles, amerindios y africanos se mezclaron, la proporción en la contribución de genes ancestrales es muy particular y difiere de otras poblaciones que se ajustan a este modelo. Incluso la proporción o contribución de genes ancestrales no es igual a nivel de provincias. En algunas provincias se evidencia el predominio de genes africanos mientras que en otras de españoles e incluso de amerindios. Esta mezcla de tres grupos étnicos y los diversos estadios de melexación en nuestro país fue lo que el insigne historiador panameño Roberto de la Guardia (1928-2012) denominó “melexa panamensis”.

En la obra se integran datos etnohistóricos que apoyan los resultados genéticos obtenidos con marcadores moleculares. Se aborda la interrogante respecto al origen de los Cuevas, un grupo amerindio extinto y se sugiere que estos son descendientes de los Chibchas. Con respecto a los Ngöbé y Emberá se revela la existencia de alelos únicos en el locus TPOX que podrían ser distintivos de cada grupo. El análisis de nueve loci STRs en estas dos poblaciones de amerindios evidencia bajos niveles de polimorfismo genético por lo que se recomienda la utilización de loci adicionales para pruebas forenses y de paternidad. Un aporte interesante de la obra es que pone en evidencia la relación entre las diferencias en frecuencias de genes ancestrales por provincia en la población mestiza y las diferencias en la proporción en que se observan en las provincias ciertas enfermedades como el cáncer de próstata y las enfermedades cerebrovasculares, lo cual tiene importantes implicaciones biomédicas específicamente desde la perspectiva de la medicina preventiva.

En este mismo contexto se evalúan varios marcadores polimórficos con la finalidad de revelar posibles asociaciones entre estos polimorfismos y la predisposición o susceptibilidad a desarrollar ciertas enfermedades. Un dato interesante que surge de análisis del polimorfismo en la región promotora del gen de paraoxonasa de humanos es la baja frecuencia del

genotipo -108CC en los Ngöbé, lo que podría estar asociado con una mayor susceptibilidad al envenenamiento con pesticidas organofosforados siendo este grupo uno de los más expuestos a este tipo de pesticida por ser la mano de obra por excelencia en la producción agrícola de tierras altas de la provincia de Chiriquí. En varios capítulos de la obra se analiza la posible asociación de varios polimorfismos presentes en la población panameña con enfermedades como artritis reumatoidea y diabetes mellitus tipo I y II. Específicamente en el capítulo 7 se evalúa la asociación del polimorfismo Pro12Ala en el exón B de receptor activado por proliferadores de peroxisomas gamma-2 encontrándose una mayor predisposición en mujeres con genotipo pro12ala a desarrollar diabetes tipo II. En los capítulos finales del libro utilizando secuenciación genómica se identifican en individuos Ngöbé dieciséis polimorfismos de un solo nucleótido con significancia clínica en los que se incluyen entre otros MTHFR y TP53 que podrían tener implicaciones biomédicas en esta población y presumiblemente en la población mestiza con un alto componente de estos genes ancestrales.

En general el libro es un aporte a la escasa o ninguna información sobre la estructura genética de nuestra población en particular la relacionada a polimorfismos asociados a enfermedades no transmisibles que en nuestro país son la principal causa de muerte.

C A P Í T U L O 1

Ascendencia genética de
la población panameña:
estructura polimórfica,
genes amerindios chibchas
y perspectivas biológicas
sobre las enfermedades

*Edgardo Castro-Pérez, Diomedes E. Trejos, Tomas Hrbek,
Vijayasarah Setaluri y Carlos Ramos Delgado.*

Resumen

La posición geográfica única de Panamá ha sometido a la población del país a un complejo proceso de mezcla desde antes de tiempos precolombinos. Los registros etnohistóricos indican que las oleadas migratorias procedentes de España y de grupos ancestrales de África occidental son parte del acervo genético panameño. Informes anteriores sugieren que varios grupos amerindios participaron en el proceso de mestizaje, pero su contribución específica al trasfondo genético no se comprende bien porque múltiples tribus e idiomas ocuparon Panamá. Además, los registros históricos no son claros y la mayoría de las tribus amerindias se extinguieron. Se piensa que cuando llegaron los europeos, los Chibchas y los Cuevas eran los grupos amerindios más numerosos que habitaban Panamá. Los Chibchas ocuparon la región occidental de Panamá y sus descendientes modernos son los Ngöbe. Los Cuevas habitaron la región oriental de Panamá, sin embargo, se extinguieron, pero están asociados etnolingüísticamente con los Chibchas o con los grupos amerindios que hablan Chocó. Para identificar los genes Amerindios en la población mestiza, recopilamos datos genéticos de Ngöbe y Amerindios Emberá, pertenecientes a familias amerindias Chibchas y Chocoes, respectivamente. Se analizaron quince marcadores STR autosómicos polimórficos para caracterizar en detalle la estructura genética y la mezcla de la población del país. Nosotros encontramos que la población de Panamá muestra niveles relativamente altos de contribución de las tres poblaciones ancestrales: 24% africanos, 25% europeos y 51% amerindios. El componente genético Amerindio es de origen chibcha y predomina en las regiones occidental y oriental del país. Proponemos que los genes Chibchas fueron transmitidos en su mayoría a los mestizos modernos desde ancestros Ngöbe y grupos de amerindios Cuevas extintos. En general, los mestizos panameños son altamente polimórficos con mezclas que varían dependiendo de la provincia, como lo demuestran los parámetros de diversidad y las distancias genéticas calculadas. Dados estos diversos patrones de mezcla entre provincias, supusimos que también podrían estar asociados con diferencias en la incidencia de enfermedades en provincias. Por lo tanto, calculamos parámetros de relevancia biomédica, incluyendo datos forenses y epidemiológicos, de las enfermedades más importantes de la población como el cáncer de próstata y los trastornos cerebro-cardiovasculares. El cáncer de próstata mostró una alta incidencia en las provincias con mayor porcentaje de genes africanos y los trastornos cardiovasculares cerebrales mostraron una presencia notable en provincias con un mayor aporte de

genes europeos y genes africanos moderadamente alto. En general, estos resultados sugieren que las proporciones de ascendencia genética determinadas por eventos etnohistóricos podrían tener un impacto significativo en la salud, los estilos de vida y las enfermedades de estas poblaciones.

Palabras claves: Chibchas, Chocoos, ascendencia genética, amerindios, enfermedades no transmisibles.

Abstract

Panama's unique geographic position has subjected the country's population to a complex admixture process since pre-Columbian times. Ethno-historical records indicate that migratory waves from Spain and Western African ancestral groups are part of the Panamanian gene pool. Previous reports suggest that several Amerindian groups participated in the miscegenation process, but their specific contribution to the genetic background is poorly understood because multiple tribes and languages occupied Panama. Moreover, historical records are unclear, and most Amerindian tribes suffered extinction. It is thought that by the arrival of Europeans, Chibchan and Cueva were the largest Amerindian groups inhabiting Panama. Chibchan occupied the Western region of Panama, and their modern descendants are the Ngöbe Amerindians. Cueva Amerindians inhabited Eastern Panama. However, they are extinct but ethno-linguistically associated with Chibchan or Chocoan-speaking Amerindian groups. To identify Amerindian genes in the mestizo population, we gathered genetic data from Ngöbe and Embera Amerindians, belonging to Chibchan and Chocoan Amerindian families, respectively. Fifteen polymorphic STR autosomal markers were analyzed to characterize the country's population's genetic structure and admixture. We found that the population of Panama shows relatively high levels of contribution from all three ancestral populations: 24% African, 25% European, and 51% Amerindian. The Amerindian genetic component is from Chibchan origins and is highly distributed within Western and Eastern regions. We propose that Chibchan genes were mostly passed to modern mestizos from Ngöbe ancestors and extinct Cueva Amerindian groups. Panamanian mestizos are highly polymorphic and differentially admixed among the country's provinces, as evidenced by diversity parameters and genetic distances calculated. Given these highly diverse admixture patterns among provinces, we surmised that they might also be associated with differences in disease incidence. Therefore, we computed parameters of biomedical relevance,

including forensic and epidemiological data of the major diseases in the population, and focused on prostate cancer and cerebrum-cardiovascular disorders. Prostate cancer showed higher incidence in provinces with the highest African admixture, and cerebrum-cardiovascular disorders showed higher incidence in provinces with the highest European admixture and moderately high African genes. These results suggest that genetic ancestry proportions shaped by ethnohistorical events might significantly impact these populations' health/cultural lifestyles and disease patterns.

Keywords: Chibchan, Chocoan, genetic ancestry, Amerindians, non-communicable diseases.

Introducción

La posición geográfica única de Panamá ha sometido a la población del país a diferentes procesos evolutivos desde la época precolombina. La población de Panamá ha sido objeto de uno de los más intrincados procesos de amalgamamiento de razas y culturas en América en un territorio y población relativamente pequeña. Sin embargo, la composición de la mezcla racial de la población actual no está claramente establecida. El primer intento de determinar la mezcla genética de la población panameña fue realizado por Arias et al. (2002), quien analizó una muestra de 4200 sujetos de todo el país utilizando un sistema de dos marcadores genéticos clásicos (ABO y Rh). Los resultados obtenidos pueden no ser representativos de la población panameña actual debido a que a inicios de los años 2000 se estimaba 3.05 millones de habitantes, mientras que los estimados más recientes indican una población de poco más de 4 millones de habitantes.

Estudios genéticos más recientes con ADN mitocondrial (Perego et al. 2012) y análisis del cromosoma Y (Grugni et. 2015) fueron realizados en la población panameña moderna. Estos estudios demostraron que la mayoría de los linajes de ADNmt (83%) son amerindios, 14% africanos y sólo 2% europeos (Perego et al. 2012). Además, estudios del cromosoma Y mostraron una contribución del 50% de amerindios y alrededor del 44,1% de Eurasia occidental y otros grupos menores no identificados (Grugni et al. 2015). Sin embargo, aunque estos hallazgos son interesantes la información proporcionada con marcadores uniparentales podría no ser representativa del acervo genético autosómico de panameños modernos. Por ejemplo, Arias et al. (2002) demostró que la mezcla de Panamá a prin-

cipios de la década de 2000 fue 38% de genes africanos, 36% amerindios y 25% europeos, sugiriendo así que los patrones de mezcla de los panameños modernos podrían ser más complejos.

Se sabe que durante la conquista europea y poco después del descubrimiento del Océano Pacífico por Vasco Núñez de Balboa en 1513, Panamá se convirtió en un mercado estratégico y puerto de intercambio de plata, oro, esclavos y mercancías transportadas desde América del Sur (ciudad de Panamá, en el Pacífico costa) por el Camino de Cruces a través del istmo hasta Portobelo en la costa caribeña y luego enviada a Europa y viceversa (Poveda-Ramos, et otros, 2004). Esta intensa actividad comercial impulsó otros eventos como patrones de olas migratorias masivas, junto con la mezcla cultural y racial. Al principio, los mestizos eran una minoría, pero finalmente dominaron a la mayoría de la población hasta días presentes. Varias minorías participaron en la mezcla de Panamá, pero los tres principales grupos ancestrales de este mestizaje de la población se agrupa etnohistóricamente como europeos, africanos y amerindios (Jaén-Suárez, 1978; 1998).

La contribución genética europea provino principalmente de muchos miles de españoles (principalmente hombres, contribuyendo con los genes del cromosoma Y) que llegaron durante y después de la conquista. Los primeros africanos llegaron a Panamá desde África Occidental para reemplazar a los amerindios esclavos varones, lo que ayudó a la supervivencia y recuperación de los varones amerindios después del cuello de botella, pero además aportaron con genes del cromosoma Y. Varias aproximaciones estiman que entre los siglos 16 y 17 llegaron más de 35.000 africanos a Panamá, sin embargo, dado que el contrabando era muy común, se cree que las cifras reales son mucho mayores (Jaén-Suárez, 1978).

La segunda ola de africanos (afro-caribeños) llegó a Panamá durante la fiebre del oro de California entre 1850 y 1855 para la construcción del Ferrocarril de Panamá. Más de 80.000 trabajadores afrocaribeños llegaron desde las islas del Caribe, pero la mayoría de ellos regresaron a sus países. La tercera ola de migración africana (África-Caribe) estuvo asociada con la construcción del Canal de Panamá 1881-1914 (Maloney, 1993). Se emplearon más 60.000 trabajadores, de los cuales el 44,1% eran de Barbados, el 24% eran de todas las demás Antillas, el 25% de Europa, de los cuales el 18% eran españoles, y el 6,9% restante eran de Centro y Latinoamérica (Maloney, 1993). Estas cifras eran muy significativas en aquellos tiempos, ya que la población total del país se estimaba en 316.054 habitantes (Maloney, 1993).

Los ancestros de los mestizos modernos de Panamá se remontan desde antes de la conquista de Panamá en 1510 por Rodrigo de Bastidas. Cuando los españoles llegaron al Istmo de Panamá, había varios grupos pequeños de amerindios que poblaban la zona. Datos etnohistóricos y demográficos indican que casi 500.000 amerindios poblaron tierras panameñas a principios del siglo XVI, pero los grupos de lengua chibcha y cueva fueron las dos poblaciones amerindias más grandes que habitaron el istmo (Jaén-Suárez, 1998). Los informes arqueológicos y etnohistóricos indican que, durante la época de la conquista española, muchos grupos chibchas habitaban la parte suroeste del istmo desde la porción occidental de la provincia de Coclé hasta la parte oriental del el Norte de Costa Rica. Con base en datos etnohistóricos y lingüísticos se sabe que los descendientes de los Chibchan son los actuales Ngöbe (también conocidos como Ngawbe-Guaymi), al igual que otras tribus más pequeñas (Jopling, 1994; Cooke, 1982, Barrantes et al., 1990; ver mapas en la Figura 1). La población Ngöbe es un grupo amerindio bien caracterizado perteneciente a una rama lingüística de los chibcha que aún habita el Occidente de Panamá (Aproximadamente 260.000 habitantes, censo de Panamá) y pequeñas regiones de Costa Rica (Barrantes et al., 1990, Jorge et otros, 2002). El lado oriental del istmo estaba habitado por otro gran grupo amerindio, el grupo de lengua cueva (o simplemente Cueva), del Golfo de Urabá y del Río Atrato (Colombia) hasta el “Río Indio” en el lado caribeño de Panamá y el río Mata Ahogado en el lado Pacífico de Panamá (Romoli, 1987, Jopling, 1994), (Figura 1). Se cree que el pueblo cueva se extinguió en 1550, principalmente debido a enfermedades, esclavitud y guerras contra los españoles (Romoli, 1987). El grupo lingüístico al que pertenecen no está claro, pero a menudo se hace referencia a los cueva como chocoanos/paezanos parlantes, pero otros creen que fueron Chibchas (Loewen, 1963, Constenla, 1991). No se cuenta con suficiente información respecto a estos dos grupos, ya que no se ha abordado con precisión en estudios de mezcla racial la contribución de estas poblaciones ancestrales al acervo genético actual. Con la finalidad de determinar la ascendencia de los mestizos panameños modernos, analizamos 15 marcadores autosómicos polimórficos de repeticiones cortas en tándem (STR) para caracterizar en detalle la estructura genética de la población y mezcla por provincia/regiones y la mezcla total del país. Además, abordamos la interrogante de la ascendencia del mestizo panameño moderno con respecto a las tribus amerindias (Chibcha o Chocoan). Determinamos que la población panameña está compuesta por 24% africanos, 25% genes europeos y 51% amerindios, y que el componente amerindio es de origen chibcha.

Los panameños son altamente polimórficos y diversos entre provincias. Consideramos que estas diferencias en la mezcla entre provincias también podrían estar asociadas con variaciones en la incidencia de ciertas enfermedades por provincia. Por lo tanto, calculamos parámetros de importancia biomédica, incluidos los forenses y datos epidemiológicos de las principales enfermedades que afectan a la población. Nos centramos en el cáncer de próstata y desórdenes cerebro-cardiovasculares, y encontramos que el cáncer de próstata mostró mayor incidencia en las provincias con mayor mezcla africana; mientras que las enfermedades cerebro-cardiovasculares mostraron mayor incidencia en las provincias con mezcla europea más alta y contribución de genes africanos moderadamente alta. En conjunto, estos hallazgos sugieren que la variabilidad genética ancestral moldeada por eventos etnohistóricos, podría desempeñar un papel importante en los patrones de enfermedades encontradas en estas poblaciones.

Materiales y métodos

Selección de Voluntarios Mestizos

En Panamá, la mayor parte de la población mestiza se concentra en las grandes zonas urbanas. Por lo tanto, recolectamos muestras de sangre de alrededor de 800 sujetos mestizos no emparentados de las principales ciudades de siete provincias de la República de Panamá: Panamá, Colón, Coclé, Herrera, Los Santos, Veraguas y Chiriquí. Los sujetos, la colecta y el manejo de las muestras se realizaron de acuerdo con los estándares internacionales sugeridos por la Comisión de ADN de la Sociedad Internacional de Genética Forense (Comisión de ADN 1994; Bar et al. 1997). Las personas seleccionadas para el estudio fueron entrevistadas para anotar confidencialmente la información biográfica, incluyendo nombres, lugares de nacimiento y ascendencia genética de sus padres y abuelos. Personas que no nacieron en Panamá o cuyos padres no eran panameños no fueron considerados para el estudio. Además, las personas pertenecientes a cualquier tribu amerindia o con un origen recientemente evidente de otras regiones como Asia no fueron incluidos para el análisis. Se obtuvo el consentimiento informado apropiado de los individuos seleccionados. Debido a la inaccesibilidad, dos provincias no pudieron ser incluidas en este estudio: Bocas del Toro y Darién. Sin embargo, estas provincias son relativamente pequeñas en número de habitantes, representando sólo el 3,2% (Bocas del Toro) y el 1,2% (Darién) de la población total de la Re-

pública de Panamá, según INEC. Además, los datos del censo indican que Bocas del Toro y Darién están compuestas principalmente por amerindios y afrodescendientes, que se encuentran relativamente aislados, mostrando así bajos niveles de mezcla con la mayoría de la población mestiza. Por tanto, la falta de datos de las provincias de Bocas del Toro y Darién no afectan significativamente los resultados generales del país, debido a la eficiencia en el muestreo las provincias representaron más del 95% de la población mestiza.

Extracción de sangre, amplificación por PCR y genotipado

Se recolectaron muestras de sangre periférica en un tubo con anticoagulante EDTA. Luego, utilizando pipetas de plástico, se colocaron unas gotas de sangre en el centro de cada uno de los cuatro círculos de Tarjetas FTA (Centro de pruebas de ADN Inc. Eules, Texas). El papel FTA se cortó en ~2 mm de diámetro y se procedió con un lavado según el procedimiento descrito por el fabricante. Las amplificaciones de PCR se realizaron utilizando la porción de la tarjeta FTA recortada y lavada. Las reacciones de PCR se realizaron en un Termociclador Applied Biosystems modelo 9700 en un volumen final de 25 μ L. Los STRs analizados fueron D8S1179, D21S11, D7S820, CSF1PO, D3S1358, TH01, D13S317, D16S539, vWA, TPOX, D18S51, D5S818, FGA, D19S433 y D2S1338 mediante reacciones de PCR multiplex, utilizando el kit comercial de amplificación AmpFISTR (*Applied Biosystem*, Foster City, CA, EE. UU.) según los protocolos del fabricante. Los productos amplificados y escaleras de referencia proporcionadas por el kit fueron visualizadas mediante electroforesis capilar usando el analizador genético ABI Prism™ 310.

Modelo de mezcla genética y datos de población ancestral

Varias poblaciones en el mundo han participado en los procesos de mezcla de la población panameña. Sin embargo, una amplia evidencia apoya que los principales componentes ancestrales del mestizaje de la población son los europeos provenientes de España, africanos occidentales y

Amerindios. Estudios previos, utilizando ABO y Sistemas sanguíneos Rh (Arias et al., 2002), ADNmt (Perego et al. 2012) y el cromosoma Y (Grugni et al. 2015), también apoyan este patrón ancestral trihíbrido. Por ello hemos diseñado un modelo trihíbrido de ascendencia en nuestros análisis. Como se mencionó anteriormente, los datos etnohistóricos y genéticos indican que los españoles trajeron el componente genético europeo a Panamá. Para probar el componente genético español, utilizamos datos de STR de población de España proporcionados por Camacho et al., (2007). También está claro que todos los genes africanos se originaron en África Occidental. Por lo tanto, seleccionamos y agrupamos datos de los STRs de dos estudios de la población de África occidental: Angola (Beleza et al., 2004) y Guiné-Bissau (Goncalves et otros, 2002). Agrupamos datos de dos poblaciones porque la información etnohistórica no es precisa sobre el origen de la migración de esclavos. Sin embargo, varios estudios genéticos indican que la mayoría de las poblaciones africanas occidentales (incluidas Angola y Guiné-Bissau) son muy similares entre ellas (Adeyemo et al., 2005). En nuestro análisis y en las frecuencias alélicas obtenidas en estudios previos confirmamos esta similitud, ya que no se observaron diferencias significativas entre las poblaciones de Angola y Guiné-Bissau (datos no mostrados).

El componente genético amerindio es más complejo y no está claro debido a lo incompleto de los registros etnohistóricos y a la temprana extinción de la mayoría las tribus de amerindios. Los grupos Ngöbe y Cueva eran las principales tribus amerindias presentes en el istmo de Panamá durante la conquista europea, y estos dos grupos podrían también ser los principales antepasados amerindios de los mestizos modernos. El aporte de cada uno de los grupos amerindios a la composición genética de Panamá se desconoce. Aunque los Ngöbe se caracterizan porque su lengua es la chibcha, el origen de los Cuevas no está claro, ya que están extintos. Algunos autores sugieren que los Cuevas eran Chocoos, otros los consideran como Chibchan (Loewen, 1963, Constenla, 1991). Para abordar esta interrogante, incluimos en los análisis microsátélites de secuencias cortas repetitivas en tándem de Ngöbe (Chibchan) y Embera (Chocoan) previamente reportadas por nosotros (Castro et al., 2007). Escogimos a los Emberá como modelo amerindio Chocó para identificar en principio los genes de Cueva porque los Emberá se caracterizan por ser chocoanos/paezanos parlantes y en la actualidad habitan parte del territorio ancestral Cueva (aproximadamente 31,284, según censo), disperso en áreas desde Colombia hasta Panamá (Figura 1). Los Emberá han sido genética y lingüísticamente asociados con el grupo amerindio Waunaan, que es otra tribu chocoana/paezana que también habita Panamá y Colombia, pero genética y lingüísticamente se-

parados de los Chibchan-Kunas y, aún más distantes lingüísticamente, de los Ngöbe (Kolman y Bermingham, 1997; Constenla, 1991). La proximidad geográfica en la era precolombina de los Cueva y los Emberá podría sugerir que entre ellos pudo ocurrir un intercambio cultural y genético. Los Emberá y Kunas experimentaron oleadas migratorias sucesivas al oeste trasladándose hacia el Darién, por lo que su población se concentra principalmente en la Comarca Emberá-Waunaan. Debido a su aislado estilo de vida y reciente migración, el contacto de los Emberá con los mestizos panameños es muy bajo. Los Ngöbe en conjunto con los Emberá, como cualquier otra tribu de amerindios, contribuyen con menos de 2% a la mezcla en mestizos (Barrantes et al., 1990; Kolman et al., 1995) (Figura 2).

Análisis de los datos

Después de descartar las muestras/datos de sujetos que no cumplían con los criterios de inclusión para la selección de mestizos, se lograron analizar 650 mestizos panameños no emparentados. Se utilizó el programa *STRUCTURE* (Pritchard et al., 2000; Falush et al., 2003) para estimar los patrones de mezcla y la estructura de la población total, dividida por provincias/regiones. El programa PopGene, versión 1.32 (disponible en <http://www.ualberta.ca/~fyeh/>) fue utilizado para calcular otros parámetros poblacionales. Las frecuencias genotípicas observadas y esperadas se determinaron utilizando dos algoritmos (Nei, 1973 y 1978) y la prueba de razón de verosimilitud (*G*-cuadrado), las cuales fueron utilizadas para estimar equilibrio de Hardy-Weinberg por locus, por provincia y en todo el país.

Las frecuencias alélicas calculadas y computadas, fueron utilizadas para estimar estadísticos como la diversidad de genes, el número efectivo de alelos (*N_e*) (Hartl y Clark, 1989), el número observado de alelos (*N_a*), y el índice de información de Shannon (*I*) (Shannon y Weaver, 1949) por locus. Las estadísticas *F* (*F_{IT}*, *F_{ST}* y *F_{IS}*) fueron calculadas utilizando dos enfoques (Hartl y Clark, 1989) y Vertedero (1990); para una jerarquía de muestreo de tres niveles: individuales, subpoblaciones (provincias) y en general población (el país). El flujo genético se estimó como número de migrantes (*N_m*) de *F_{ST}* (Slatkin y Barton, 1989) para provincias. Se calculó la identidad genética no sesgada (Nei, 1972) y distancia genética (Nei, 1978) entre provincias, y se construyó un dendrograma basado en la distancia de Nei utilizando el método UPGMA con el programa Neighbor de Phylip versión 3.5c de Joe Felsenstein. Se utilizó la plantilla de libro de MS Excel PowerStatsV12 (Promega Corp.) para calcular los parámetros foren-

ses: coincidencia probabilidad y poder de discriminación. Los parámetros de paternidad también se estimaron, incluyendo el poder de exclusión e índice típico de paternidad.

Resultados

Mezcla de genes ancestrales y estructura poblacional

Las estimaciones ancestrales de la mezcla genética se muestran en la tabla. 1. Los panameños modernos son un grupo trihíbrido con contribuciones de países africanos (24%), europeos (25%) y Amerindios Ngöbe/Chibchan (51%). Los genes chibchas están presentes en todo el país con niveles relativamente altos de contribución dentro de todas las provincias, pero en su mayoría concentrados en Coclé (70%), Chiriquí (64%), Veraguas (50%) y Panamá (50%), pero en menor cantidad en Los Santos (33%) y Colón (34%). La provincia de Colón mostró la mayor proporción de genes africanos (47%), seguidos por la provincia de Panamá (27%) y Los Santos (25%). La cantidad más baja de genes africanos se encontró en Coclé (10,2%) y Chiriquí (14,8%). Una proporción significativa de genes europeos fueron 42% en Los Santos, 40% Herrera y el más bajo fue en Colón (18,6%). En general, Panamá como país mostró una alta heterogeneidad genética (86%). Las provincias más heterogéneas fueron Panamá y Colón con el 85%, y Veraguas con 82%. Las provincias menos heterogéneas fueron Coclé (70%) y Chiriquí (72%). Las medidas de heterogeneidad para Herrera y Los Santos fueron 79% y 76%, respectivamente. Además, realizamos múltiples análisis comparativos de varianza, que mostraron patrones significativamente diversos de estructura genética y mezcla entre todas las provincias, excepto entre Herrera y Los Santos (Figura 2). El grupo Embera/Chocoan no contribuyó al perfil genético de la población panameña actual (sólo el 2,6% en promedio). Las estimaciones de mezcla se complementaron con los datos de distancias genéticas entre provincias y se evidenciaron en un dendrograma basado en información de loci STR (Figura 3). Se muestran dos conglomerados principales, uno que incluye Veraguas, Panamá, Chiriquí y Coclé, y el otro conglomerado formado por Herrera, Los Santos y Colón. Los patrones de mezcla entre provincias que se muestran en la Figura 2 son similares a la distribución de agrupamiento mostrada en el dendrograma, sin embargo, existen algunas diferencias sujetas a análisis de distancia entre provincias y fueron evidentes en un dendrograma basado en información de loci STR (Figura 3). Se muestran

dos grupos principales, uno que incluye Veraguas, Panamá, Chiriquí y Coclé, y el otro conglomerado formado por Herrera, Los Santos y Colón. Los patrones de mezcla entre provincias que se muestran en la Figura 2 son similares a la distribución de agrupamiento que se muestra en la dendrograma, sin embargo, existen algunas diferencias sujetas a consideración. Coclé, Chiriquí y Veraguas parecen ser muy similares en mezcla/estructura, pero Veraguas está un poco separada, aunque en un mismo conglomerado. Herrera y Los Santos no mostraron diferencias significativas en el patrón de estructura genética, pero Los Santos está unido a Colón, mientras que Herrera se agrupa un poco separada, aunque en el mismo grupo.

Variaciones de frecuencias alélicas

Las frecuencias alélicas totales de la población panameña de los quince loci STR de mestizos se muestran en la Tabla 2. Es importante destacar que faltan datos de algunos loci en algunas poblaciones ancestrales. Sin embargo, debido al alto porcentaje de Genes Ngöbe/Chibchan en provincias como Coclé, Chiriquí y Veraguas, y la falta de información de algunos loci ancestrales amerindios, es razonable suponer que muchos de los alelos ancestrales con mayor frecuencia en mestizos panameños provienen de los antepasados Chibchas. La Tabla 3 muestra los alelos con mayor frecuencia por provincia, en el país y en las poblaciones ancestrales. Por ejemplo, datos de africanos y europeos ancestrales en los loci D5S818 muestran sus frecuencias alélicas más altas en el alelo D5S818*12. En contraste, en la mayoría de las provincias el alelo más frecuente es el alelo D5S818*11, lo que probablemente implica que el D5S818*11 fue adquirido por los chibchas ancestrales. Un razonamiento similar podría en principio utilizarse para inferir a partir de alelos alelos D2S1338*23 y D18S51*14 los datos ancestrales de los Ngöbe que se perdieron. Adicionalmente, en la Tabla 3 se muestra que los alelos con mayor frecuencia en Embera/Chocoan para los loci TH01 y TPOX son *8 y *6, respectivamente, pero en Ngöbe son *6 y *11, respectivamente. Por otro lado, en las poblaciones ancestrales de África y España, los alelos con frecuencias más altas en los loci TH01 fueron *7 y *9.3, respectivamente. Sin embargo, todas las poblaciones mestizas de Panamá que se cruzaron con Ngöbe presentan el alelo TH01*6 como el de mayor frecuencia. De manera similar, TPOX*11 fue el alelo con mayor frecuencia en todas las muestras de mestizos cruzados con Ngöbe y africanos como alelos ancestrales más altos en mestizos, pero el *6 presente en embera/chocoanos. Es poco probable que los

Emberá/Choco contribuyeran con alelos que desaparecieron dado que son genéticamente diferentes a los Ngöbe/Chibchan, como se muestra en la Figura 2, lo cual también es apoyado por trabajos previos con ADN nuclear y mtDNA de esas poblaciones amerindias (Kolman y Bermingham, 1997). Esto lo abordamos posteriormente desde una perspectiva comparativa entre genes autosómicos de mestizos amerindios (en este trabajo) con el ADNmt de mestizos de un reporte previo (Perego et al. 2012).

HWE, diversidad genética y estadística F

Las estadísticas de heterocigosidad, variación genética y diversidad de loci se muestran en la Tabla 3. Los loci con el mayor número de alelos polimórficos (N_a) fueron D18S51 (20), D21S11 (19) y FGA (18), mientras que los loci con menos alelos polimórficos fueron TH01, D13S317 y TPOX, con 8 alelos polimórficos. Los loci con los mayores números efectivos de alelos (N_e) fueron D2S1338 (8.687) y D18S51 (8.433), y los menores fueron TPOX (3.405) y D3S1358 (3.503). Los loci más diversos según el índice de Shannon (I) fueron D18S51 y D2S1338 con valores de 2.303 y 2.275 respectivamente. Encontramos la diversidad de Shannon más baja en TPOX (1.442) y D3S1358 (1.468). Los loci D18S51 (87,40%) y D2S1338 (87,30%) mostraron la mayor heterocigosidad, pero TPOX (69,20%) y D3S1358 (71,90%) presentaron la más baja.

Los valores de HWE indican que la mayoría de los loci cumplen con las expectativas de equilibrio. La diversidad genética también se abordó mediante el uso de estadísticas F para evaluar la información sobre diversidad de subpoblaciones (provincias). Los valores estadísticos FIS evidencian un alto polimorfismo dentro de cada provincia, con la mayor heterocigosidad informativa en D21S11 (-0.049). Todos los loci presentaron valores de FST bajos entre las provincias, lo que indica un alto polimorfismo. Valores FST igualmente bajos han sido reportado en otras poblaciones mestizas en América Latina (Assis-Poiars et al., 2010; Rubi-Castellanos, et al., 2009). Sin embargo, nuestro dendrograma y las distancias genéticas indican diferencias importantes no detectables en el índice de diferenciación

FST. Consistente con los valores de FIS, el valor FIT sugiere también una alta heterocigosidad en la población general, siendo también el más informativo D21S11 con -0,037. Aunque los valores del flujo genético (N_m) están probablemente más relacionados con el flujo de genes ancestrales que con la migración entre provincias, se estimó con base en FST para provin-

cias, mostrando los niveles más altos de flujo de genes en vWA (45.680), D8S1179 (40.592) y D5S818 (38.000).

Parámetros Genéticos Forenses e Implicaciones Antropológicas

Las repeticiones cortas en tándem analizadas en este estudio son de importancia para los estudios biomédicos en casos forenses, antropología forense y pruebas de paternidad. La alta heterogeneidad genética, los polimorfismos y las diferencias en la mezcla entre provincias generó la interrogante de; son estos niveles de diversidad de utilidad en aplicaciones antropológicas. El alto polimorfismo genético y la variación en la diversidad encontrada podría requerir el cálculo por provincia de estos estadísticos. Por lo tanto, calculamos estadísticos de utilidad en pruebas forenses y de paternidad. Se calcularon estadísticos en cada locus, por provincia y a nivel todo el país (Tabla 3). Estos datos sugieren que los loci más informativos en aplicaciones forenses en la población panameña son D18S51, FGA y D21S11. Sin embargo, los otros loci también son muy informativos cuando se utilizan en combinación.

El análisis combinado de 15 loci STR arroja para la población panameña una probabilidad de coincidencia de 1 en 5.3×10^{17} y un poder de discriminación de 1.000, valores con aplicación forense de gran utilidad. Parámetros con aplicación en paternidad como el poder de exclusión (0,999998992) y el índice de paternidad combinado (924654.0396) también demostraron ser estadísticamente muy informativos para el país. Se observaron resultados similares en cada provincia, validando así estos loci como excelentes marcadores con aplicaciones biomédicas en genética forense (datos no mostrados).

Parámetros genéticos forenses e implicaciones antropológicas

Numerosos estudios epidemiológicos indican que la ascendencia genética es importante cuando se analizan grupos en lugar de individuos (Rosenberg et al. 2005; Tang et al., 2005; Kittles y Weiss 2003). Sin embargo, es importante señalar que la ascendencia genética por sí sola no es la causa de las enfermedades, ya que en el desarrollo de estas es el resultado de

varias interacciones entre genes, medio ambiente y estilos de vida (Nebert et al., 2003; Ober y Vercelli 2011; Bravo et al., 2011; Scatliff et al., 2011). Por lo tanto, la ascendencia genética es un marcador indirecto de utilidad para evaluar posibles factores de riesgo y disparidades en algunos fenotipos relacionados con la salud. Por ejemplo, se ha descrito que en Panamá durante los procesos de migración y mezcla, varias enfermedades afectaron la población de manera selectiva a ciertas razas (Jaén Suárez, 1998). De hecho, se ha reportado que muchas enfermedades diezmaron las poblaciones amerindias; mientras que los antillanos (africanos-caribeños) eran más resistentes a la mayoría de las infecciones (Poveda-Ramos, et al., 2004). Estudios farmacogenéticos indican que existen diferencias metabólicas de las tribus amerindias en relación con los europeos y africanos que explican diferencias en la biotransformación de fármacos debido a polimorfismos a nivel de enzimas (Jorge et al., 1999; Jorge-Nebert et al., 2002; Jorge y Arias, 1995). Los estudios epidemiológicos indican una alta incidencia de síndromes asociados a enfermedades metabólicas en amerindios como las del metabolismo de lípidos y diabetes como resultado de un cambio al estilo de vida de la sociedad moderna (Hanis et al., 1991; Chakraborty y Weiss 1986; Gower y otros, 2003; Torun y otros, 2002; Klimentidis et al., 2009). Estos fenotipos/enfermedades son altamente heredables y factores de alto riesgo para otras enfermedades complejas como las enfermedades cardiovasculares, accidentes cerebrovasculares y cánceres (Basu et al., 2009). La heterogeneidad observada y las diferencias en las proporciones de genes ancestrales en las provincias y la gran cantidad de datos genéticos epidemiológicos reportados en la literatura nos motivaron a investigar la posible relación entre las diferencias regionales de nuestra mezcla de genes en la incidencia de muertes por enfermedades. Las estadísticas de los últimos 20 años del Ministerio de Salud indican que las dos causas principales de las muertes por enfermedad en la República de Panamá son enfermedades cerebrales/cardiovasculares y cánceres. Con base en estas estadísticas nos concentramos en el cáncer de próstata y en enfermedades cerebro/cardiovasculares. Con respecto al cáncer de próstata, recopilamos datos desde 1985 hasta 1997 (después de 1998, la mayoría de los pacientes con cáncer son trasladados y tratados en la ciudad de Panamá); en el caso de las enfermedades cerebro/cardiovasculares recopilamos datos desde 1998 a 2010. La incidencia de estas enfermedades fue evaluada estadísticamente mediante análisis ANOVA unidireccional y comparado entre provincias. Determinamos las diferencias significativas en la incidencia de enfermedades entre provincias. Aún más, descubrimos que la incidencia de estas enfermedades parece correlacionarse con el componente genético ancestral de cada provincia. El cáncer

de próstata mostró una incidencia significativamente mayor en Colón y Panamá, provincias con la mayor proporción de genes africanos (Figura 4A). Además, los ataques cerebrales y las enfermedades cardiovasculares presentaron una incidencia significativamente mayor en provincias como Herrera y Los Santos (Figura 4B), las cuales son las provincias con mayor proporción de mezcla europea, pero también con una proporción moderadamente alta de genes africanos.

Discusión

Los mestizos panameños están heterogéneamente mezclados y son altamente polimórficos

El análisis de estructura genética de la población panameña, basados en 15 marcadores STR autosómicos, mostraron que nuestra población es altamente polimórfica como lo indican las diferencias en mezcla entre individuos y provincias (Figura 2). La distribución geográfica de los genes amerindios, africanos y europeos es consistente con los datos etnohistóricos (Jaén-Suárez, 1978; 1998). Patrones alélicos y parámetros de diversidad mostrados en las Tablas 1 y 3 también evidencian la diversidad genética entre provincias. Las distancias genéticas y el dendrograma (Figura 3) también respaldan estos hallazgos. La estadística, los parámetros forenses y de paternidad son consistentes con los parámetros de diversidad genética y heterogenidad calculados para cada loci y provincia. Parámetros forenses y de paternidad de 15 loci STR loci demostraron que la información proporcionada por estos marcadores es muy importante y útil en casos forenses y aplicaciones antropológicas en la población general y provincias. La población de Panamá presenta niveles relativamente altos de contribución de las tres poblaciones ancestrales: 24% africanos, 25% europeos y 51% amerindios. Estos resultados difieren de los reportados con anterioridad por Arias et al. (2002), quienes estimaron una contribución de 38% de genes africanos, un 36% de genes amerindios y un 25% de genes europeos. Nuestros resultados podrían ser explicados con base en el crecimiento demográfico y patrones migratorios en los últimos años, pero también podrían estar relacionados con el hecho de utilizar dos sistemas genéticos/clásicos *versus* 15 marcadores STR informativos. A diferencia de Panamá, la mayoría de los países latinoamericanos y las ciudades están dominadas por uno o dos orígenes genéticos (Tabla 1).

Costa Rica, Medellín (Colombia) y Ciudad de México están compuestos principalmente por genes amerindios y europeos. Uruguay, Buenos Aires (Argentina) y Santiago (Chile) están compuestos en su mayoría por europeos. Los genes africanos predominan en Barbados y Jamaica.

Los genes amerindios son de origen chibcha

El componente amerindio de los mestizos panameños es Ngöbe/Chibchan, ya que no hay una contribución significativa de ascendencia Emberá/Chocoana. Nuestro estudio arrojó que en promedio la contribución de genes Ngöbe/Chibchan es de 46% y sólo el 2,6% Embera/Chocoan. Considerando la distribución geográfica actual y precolombina de las tribus de habla Ngöbe/Chibchan, no es sorprendente encontrar genes Chibchas en Coclé, Chiriquí y Veraguas, pero también se encuentran genes chibchas con predominio en el este de las provincias de Panamá y Colón. Sin embargo, dada la incertidumbre de los datos sobre los Cuevas, que a veces son clasificados como Chocoanos que habitaban las provincias de Panamá y Colón, nuestros hallazgos podrían implicar que los Cueva eran Chibchas o simplemente no fueron capaces contribuir significativamente al proceso de mezcla debido a su temprana extinción. Sin embargo, Arias et al. (1992) informó que un grupo aislado de mestizos de ascendencia amerindia (Cholos de Coclé) de las montañas centrales de Panamá (Provincia de Coclé) presentaron marcadores tanto Ngöbe-Chibchan como Kuna-Chibchas. En particular, encontraron una frecuencia relativamente alta (0,52) de la variante alélica Pepa-Kun de la peptidasa A, que es indetectable en Ngöbe y Chibchans en el occidente de Panamá (Barrantes et al., 1990). La variante del alelo Pepa-Kun se encuentra con alta frecuencia (0,42) en amerindios Kuna-Chibchan del oriente y grupos de habla chibcha de Colombia (Arias et al., 1992; Barrantes et al., 1990). Es poco probable que los Kunas hayan pasado el alelo Pepa-Kun en tan alta frecuencia a los mestizos de Coclé, debido a que cuando los españoles llegaron a Panamá, los Kunas habitaban el Caribe al noroeste de Colombia, cerca de Chocoan Emberás (ver mapas). Esta opinión también está respaldada por informes anteriores del análisis del ADNmt de mestizos panameños, que muestran que el 51% del ADNmt amerindio correspondía a haplogrupo A2, un haplogrupo muy abundante en Ngöbe y Kuna, pero baja en Emberá/Chocoans (Perego et al. 2012). El mismo estudio también evidenció que los mestizos panameños mostraban baja frecuencia de haplotipos específicos de Emberá/

Chocoan. Actualmente, los Kunas se encuentran dispersos en el noreste de la costa caribeña de las provincias de Panamá y Darién. Los Kunas son un grupo altamente endogámico y relativamente aislado, y su contacto con mestizos es reciente, de allí que no ha contribuido significativamente al mestizaje. Por lo tanto, esto apoya la posibilidad de que los Cuevas fueran Chibchas y, por ende, podrían haber transmitido una fracción de sus genes chibchas a provincias como Panamá y Colón. Además, algunos autores han sugerido que los Cuevas tenían afinidad lingüística con los Chibchan Kuna (Loukotka 1968; Greenberg, 1987), apoyando por tanto la hipótesis de que los Cuevas son de origen Chibcha y su afinidad genética con los Kunas. Los análisis del ADN mitocondrial y nuclear también confirman la relación genética de los Chibcha con los Kunas y los Ngöbes (Kolman y Bermingham, 1997).

Es dudoso que los genes chibchas en las provincias de Panamá y Colón hayan sido el resultado de migraciones recientes de provincias centrales durante la construcción del ferrocarril y el Canal de Panamá, porque sólo el 1% de los trabajadores eran de origen local (Maloney, 1993). La mayoría de los pueblos y ciudades fueron fundadas por los españoles en poblados originalmente ocupados por amerindios (mapa, Figura 1), lo cual fue la semilla del mestizaje desarrollado principalmente *in situ*. Estas poblaciones se mantuvieron relativamente aisladas desde tiempos coloniales con poca migración a la ciudad, hasta 1962 y 1964 cuando se construyó el Puente de las Américas y la Carretera Panamericana, respectivamente. La construcción de estas infraestructuras fomentó la migración desde el interior a ciudades de Panamá Oeste, Ciudad de Panamá y San Miguelito en la segunda mitad del siglo XX, por lo que es posible que algunos genes Ngöbe/Chibchan hayan arribado a la provincia de Panamá y Colón en tiempos más recientes. No obstante, la mayoría los datos demográficos indican que la población moderna de las provincias de Panamá y Colón han crecido *in situ* desde la época colonial (Jaén-Suárez 1998).

La Ciudad de Panamá fue fundada en 1519, en el sitio de una aldea cueva, y posteriormente se incorporaron aldeas amerindias próximas, sitios de descanso y ciudades/pueblos satélite más pequeños. En consecuencia, en la Ciudad de Panamá se incorporaron habitantes de Santa María del Darién, Acla, Nombre de Dios, Chepo y otras ciudades/pueblos pequeños en territorio Cueva (Figura 1). Por ejemplo, Santa María del Darién y Acla fueron abandonadas en 1524 y 1527 respectivamente, y todas sus poblaciones, incluidos los amerindios y mestizos, se mudaron a la Ciudad de Panamá. Además, numerosos documentos que datan de los siglos XVI y

XVII, mencionan que las aldeas amerindias fueron rápidamente absorbidas debido al rápido mestizaje, dispersión y muertes (Jaén-Suárez, 1998). Ciudad de Panamá y Portobelo (en la provincia de Colón), las más grandes del istmo, absorbieron la mayoría de las ciudades más pequeñas y pueblos amerindios. En la Provincia de Colón, la ciudad de Colón, tal como la conocemos hoy, fue fundada en 1852, pero su población es el resultado de la fusión de Nombre de Dios, Portobelo, y otros pueblos y aldeas pequeñas. El pueblo cueva habitó las provincias de Panamá y Colón durante unos 30-40 años después de la llegada de europeos (Romoli 1987), probablemente tiempo suficiente para transmitir algunos genes. Múltiples fuentes apoyan que muchos los amerindios se integraron a la nueva sociedad europea colonial, contribuyendo así aún más a la mezcla (Torres de Arauz 1972; Castellero Calvo, 1995).

La distribución de genes africanos y europeos son consistentes con las olas migratorias

Aunque la contribución de genes africanos y europeos al mestizaje general fue de 24% y 25% respectivamente, la distribución de estos y el momento en que se incorporaron en cada provincia es diferente. Como se mencionó anteriormente, los datos etnohistóricos indican que los genes africanos llegaron por primera vez a Panamá en el siglo XVI directamente de África occidental, cuyos descendientes son hoy llamados afro-hispanos. En la segunda mitad del XVIII y principios del XIX, el segundo flujo de genes africanos provino de las Antillas, cuyos descendientes son llamados afro-caribeños. Nuestros datos genéticos también apoyan esta teoría. Las provincias de Colón y Panamá mostraron el componente más alto de los genes africanos. Estas dos provincias eran los principales lugares de distribución del tráfico de esclavos africanos durante la colonia, pero también el más importante foco migratorio de las Antillas (África-Caribe) durante la construcción del ferrocarril y el Canal de Panamá. En 1911, el censo de la ciudad de Colón evidenció que el 56% de los habitantes eran descendientes directos de africanos (Jaén-Suárez, 1998). Sin embargo, un importante porcentaje de genes africanos también se encontraron en todas las demás provincias, probablemente adquiridos durante la esclavitud en la época colonial. Un gran número de esclavos africanos escaparon hacia selvas del istmo (Arias et al. 2002). Esto también es consistente con que Los Santos y Herrera también mostraron ni-

veles relativamente altos de contribución africana, junto con los niveles más altos de genes europeos que sugieren patrón co-migratorio de los españoles con sus esclavos africanos en esas provincias. Esto se explica por las prácticas de poligamia de hombres españoles con esclavas, incrementando el bagaje genético europeo junto con el africano. En cambio, la provincia de Colón, al tener la mayor cantidad de genes africanos, mostró la menor contribución de europeos. El patrón de mezcla que se muestra en la Tabla 1 y en la Figura 2, también confirma poca contribución de genes europeos en las provincias, excepto Los Santos y Herrera, en las cuales son muy similares. Estas dos provincias no son montañosas y fueron rápidamente colonizadas, lo que implica que los amerindios pronto murieron o huyeron a zonas remotas de las montañas de Chiriquí, Bocas del Toro y Veraguas (Arias et al., 1992). Los esclavos africanos reemplazaron a los amerindios, lo que explica la baja contribución de amerindios en Herrera y Los Santos comparado con Chiriquí y Veraguas. Chiriquí mostró la menor cantidad de genes africanos, probablemente como resultado de su inaccesibilidad en términos de topografía montañosa y largas distancias en la época colonial (Arias et al., 2002).

La incidencia regional de enfermedades es comparable con la ascendencia genética

Los patrones altamente polimórficos y las diferencias ancestrales de la distribución entre provincias nos hicieron considerar que estas diferencias ancestrales pudieran estar correlacionadas con la incidencia por provincia de ciertas enfermedades. Nos centramos en algunas de las principales causas de muertes por enfermedad en diferentes provincias y regiones, por lo que seleccionamos el cáncer de próstata y las enfermedades cerebrales/cardiovasculares. Curiosamente, el cáncer de próstata mostró la mayor incidencia en las provincias con mayor ascendencia africana (Colón y Panamá) (Figura 4A). Una cantidad considerable de estudios reportan diferencias significativas en el riesgo de padecer de cáncer de próstata con una tasa de incidencia de 60% mayor en hombres en comparación con los hombres europeo-americanos (Bock, et al., 2009; Zeigler-Johnson et al., 2008; Robbins et al., 2007). Es importante destacar que, debido a su proximidad, la mayoría de los pacientes con cáncer de próstata de la provincia de Colón suelen ser tratados en hospitales de la Ciudad de Panamá, lo que puede explicar una incidencia ligeramente mayor de esta condición en Panamá que en Colón.

Las enfermedades de accidentes cerebrovasculares/cardiovasculares mostraron la mayor incidencia en las provincias con mayor proporción de genes europeos y una proporción moderadamente alta de genes africanos (Figura 4B). Los accidentes cerebrovasculares y los factores de riesgo de enfermedades cardíacas incluyen hipertensión, diabetes, obesidad, niveles altos de colesterol, y son condiciones de salud consistentemente asociadas con europeos-americanos y africanos-estadounidenses (Cheng et al., 2010; Hankey, 1999; Donnan et al., 2008). Además, las enfermedades cardíacas y los accidentes cerebrovasculares son las principales causas de muertes en afroamericanos y descendientes de europeos en Estados Unidos (Estadísticas Vitales Nacionales de EE.UU., 2007 y 2009, Estadísticas de enfermedades cardíacas y accidentes cerebrovasculares 2011; Centros para el Control y la Prevención de Enfermedades, Casper y otros, 2003; Roger et al., 2011). Por lo tanto, la mayor incidencia de enfermedades de accidentes cerebrovasculares/cardiovasculares en Los Santos y Herrera, con altas tasas europeas y ascendencia africana, podría implicar un factor de riesgo combinado más alto que sólo en Europa y África.

Los resultados generados en esta investigación proporcionan evidencia de cómo los acontecimientos etnohistóricos dieron forma a la distribución geográfica de poblaciones, mientras que los análisis genéticos realizados en este trabajo revelaron las diferencias en el mestizaje de nuestra población. Adicionalmente, el intercambio cultural modeló estilos de vida que conjuntamente con el acervo genético contribuyeron con las diferencias en la distribución de enfermedades que pueden ser explicadas por la ascendencia genética. En conjunto, estos resultados presentan un enfoque antropológico de los mestizos panameños como grupo modelo para estudiar fenotipos complejos relacionados con la salud. Además, se sugiere la existencia de interacciones entre genes y entre genes y el ambiente, particularmente en variantes genéticas que podrían ser poco comunes en las poblaciones ancestrales. Esta hipótesis se refuerza al observar mayores incidencias de enfermedades cerebrovasculares y cardiovasculares en provincias como Los Santos y Herrera, donde hay una mayor contribución de genes europeos y africanos.” Es importante destacar la necesidad de utilizar modelos celulares *in vitro* para el estudio de enfermedades con miras al desarrollo de la medicina genómica y mejora de las políticas sanitarias en poblaciones mestizas, siendo Panamá uno de los más impresionantes ejemplos de procesos complejos de mestizaje.

Tabla 1. Estimaciones de la mezcla de genes ancestrales por provincia y comparación con otras ciudades y países de América Latina.

*La provincia de Panamá incluye datos agrupados de las provincias de Panamá y Panamá Oeste.

	Africano	Europeo	Amerindio
Veraguas	0.11	0.39	0.5
Panamá (2provincias)*	0.27	0.23	0.5
Los Santos	0.25	0.42	0.33
Herrera	0.2	0.4	0.4
Colón	0.47	0.19	0.34
Coclé	0.1	0.2	0.7
Chiriquí	0.15	0.21	0.64
Panama (El país)	0.24	0.25	0.51
Costa Rica ^a	0.09	0.61	0.3
Medellín (Colombia) ^b	0.09	0.66	0.25
Ciudad de México ^c	0.05	0.26	0.69
Barbados ^d	0.89	0.1	0.02
Jamaica ^d	0.84	0.12	0.03
Sao Paulo (Brazil) ^e	0.14	0.79	0.07
Puerto Rico ^f	0.21	0.64	0.15
Uruguay ^g	0.05	0.84	0.1
Santiago (Chile) ^h	0	0.57	0.43
Argentina ⁱ	0.025	0.78	0.194

^aMorera et al., 2003, ^bWang et al., 2008, ^cJuarez-Cedillo et al., 2008,

^dBenn-Torres et al. 2008, ^eFerreira et al., 2006, ^fVia et al., 2011,

^gHidalgo et al 2005, ^hSans, 2000, ⁱSeldin et al., 2007.

Tabla 2. Frecuencias alélicas de 15 loci STR en la población mestiza de Panamá.

ALELO/ LOCUS	D8S1179	D21S11	D7S820	CSE1PO	D3S1358	TH01	D13S317	D16S539	D2S1338	D19S433	VWA	TPOX	D18S51	D5S818	FGA
3						0.001									
5						0.002									
6						0.379		0.001			0.016			0.001	
7			0.007	0.014		0.248					0.008			0.05	
8	0.004		0.123	0.023	0.098	0.076	0.016				0.419			0.022	
9	0.006		0.081	0.017	0.136	0.179	0.151		0.002		0.089			0.116	
9.3						0.131									
10	0.056		0.262	0.228	0.005	0.075	0.154	0.002		0.039	0.005	0.056			
11	0.078		0.277	0.277		0.184	0.311	0.021	0.001	0.307	0.004	0.315			
11.2									0.001						
12	0.124		0.199	0.359	0.003	0.311	0.23	0.07		0.12	0.081	0.284			
12.2								0.024							
13	0.292		0.045	0.075	0.005	0.118	0.121	0.008	0.3	0.006	0.002	0.072	0.144		
13.2								0.087					0.002		
14	0.275		0.005	0.006	0.115	0.052	0.016	0.237	0.059	0.156	0.012				
14.2								0.032					0.001		
15	0.133		0.001	0.001	0.442	0.005	0.001	0.121	0.11	0.116	0.001				
15.2						0.001		0.05							
16	0.026		0.24			0.035	0.033	0.38	0.143						
16.2								0.017							
16.6								0.001							
17	0.006		0.119			0.133	0.002	0.254	0.19						
17.2								0.001							
18			0.07			0.041	0.116	0.078	0.004						
18.2									0.006						
19			0.005			0.117	0.06	0.057	0.066						

Tabla 3. Parámetros de diversidad y variación genética, y comparaciones de alelos con las frecuencias más altas* entre provincias y poblaciones ancestrales en 15 loci STR.

LOCUS:	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D13S317	D16S539	D2S1338	D19S433	vWA	TPOX	D18S51	D5S818	FGA
3 Alelos más altos:															
Panamá (país)	13	30	11	12	15	6	12	11	23	13	16	8	17	11	24
	-0.29	-0.27	-0.28	-0.36	-0.44	-0.38	-0.31	-0.31	-0.16	-0.3	-0.38	-0.42	-0.19	-0.32	-0.22
Veraguas	13	30	11	12	15	6	12	11	23	13	16	8	14	11	25
	-0.28	-0.31	-0.31	-0.33	-0.51	-0.44	-0.34	-0.39	-0.21	-0.28	-0.38	-0.42	-0.18	-0.32	-0.26
Panamá (2provincias)*	14	30	10	12	15	6	12	11	22	13	16	8	17	11	24
	-0.3	-0.25	-0.26	-0.39	-0.44	-0.36	-0.33	-0.32	-0.15	-0.3	-0.36	-0.4	-0.21	-0.32	-0.21
Los Santos	13, 14 (0.27)	30	10	10, 11, 12 (0.29)	15	6	11	11	17	13	16	8	16	11	23
	-0.3	-0.31	-0.31	-0.32	-0.32	-0.32	-0.32	-0.45	-0.22	-0.37	-0.37	-0.52	-0.2	-0.32	-0.18
Herrera	14	30	11	11	15	6	12	12	20	14	16	8	17	11	23
	-0.3	-0.35	-0.29	-0.34	-0.38	-0.31	-0.28	-0.28	-0.15	-0.34	-0.33	-0.39	-0.22	-0.39	-0.2
Colón	13, 14 (0.31)	28, 29 (0.21)	10, 11 (0.25)	11, 12 (0.32)	15	6	12	11	17	13	16	8	17	12	24
	-0.39	-0.3	-0.28	-0.33	-0.15	-0.28	-0.33	-0.33	-0.15	-0.28	-0.38	-0.42	-0.19	-0.32	-0.2
Coclé	13	30	11	12	15	6	12	12	23	13	16	11	14	11	24
	-0.34	-0.25	-0.3	-0.4	-0.48	-0.39	-0.29	-0.33	-0.19	-0.35	-0.44	-0.39	-0.21	-0.3	-0.27
Chiriquí	13	30	11	12	15	6	12	11	23	13	16	8	17	11	24
	-0.36	-0.34	-0.31	-0.37	-0.46	-0.39	-0.34	-0.27	-0.19	-0.32	-0.39	-0.49	-0.2	-0.31	-0.24
Poblaciones Ancestrales:															
Ngöbe	-	-	11	12	-	6	9	10	-	-	16	11	-	-	-
	-0.43	-0.46	-0.86	-0.56	-0.37	-0.6	-0.48	-	-	-	-	-	-	-	-
Embera	-	-	11	12	-	8	9	10	-	-	16	6	-	-	-
	-0.45	-0.43	-0.55	-0.33	-0.37	-0.56	-0.54	-	-	-	-	-	-	-	-
Guine-bissau	14	30	10	12	16	7	12	11	-	-	16	11	17	12	22
	-0.34	-0.2	-0.38	-0.34	-0.31	-0.35	-0.48	-0.32	-0.26	-0.32	-0.22	-0.4	-0.22	-0.4	-0.23

Angola	14	28	10	10	15	7	12	11	19	13	15,16 (0.23)	8	15	12	24	
	-0.36	-0.22	-0.32	-0.32	-0.34	-0.4	-0.39	-0.31	-0.16	-0.25		-0.31	-0.21	-0.36	-0.19	
España	13	30	10	11	16	9.3	12	12	17	14	17	8	15	12	21	
	-0.28	-0.27	-0.3	-0.33	-0.26	-0.26	-0.3	-0.32	-0.28	-0.34	-0.27	-0.53	-0.17	-0.38	-0.21	
Variación Génica, HWE, Estadísticos F, y Forenses:																
N a	1294	1290	1290	1284	1294	2316	1294	1294	1284	1912	1294	1294	1290	1294	1294	
Na b	10	19	10	9	9	8	8	11	13	17	10	8	20	10	18	
Ne c	4.903	6.273	4.795	3.781	3.503	3.994	5.236	4.749	8.687	5.578	4.135	3.405	8.433	4.542	7.471	
I d	1.781	2.105	1.702	1.507	1.468	1.524	1.803	1.683	2.275	2.022	1.65	1.442	2.303	1.709	2.175	
Heterocigotos (%)	80.6	85.9	79.2	73.7	71.9	72.4	80.4	80.8	87.3	81.2	75.4	69.2	87.4	77.3	85.3	
P-HWE e	0.953	0.999	0.827	0.947	0.803	0.269	0.695	0.6	0.123	0.988	0.718	0.891	1	0.701	0.992	
FIS f	-0.018	-0.049	-0.035	-0.005	-0.01	0.021	-0.005	-0.03	0.022	0.008	0.023	0.002	-0.004	0.03	0.009	
FST g	0.006	0.011	0.011	0.009	0.012	0.008	0.014	0.014	0.009	0.012	0.005	0.01	0.012	0.007	0.01	
FIT h	-0.012	-0.037	-0.024	0.004	0.001	0.028	0.009	-0.015	0.032	0.02	0.028	0.012	0.009	0.036	0.019	
Nm i	40.592	22.934	22.625	29.177	20.484	32.77	17.438	17.134	26.406	20.887	45.68	24.21	20.048	38	25.69	
MP j	0.068	0.047	0.073	0.115	0.122	0.106	0.06	0.078	0.026	0.052	0.088	0.136	0.027	0.081	0.032	
PD k	0.932	0.953	0.927	0.885	0.878	0.894	0.94	0.922	0.974	0.948	0.912	0.864	0.973	0.919	0.968	
PE l	0.61	0.713	0.584	0.488	0.458	0.466	0.607	0.613	0.741	0.622	0.517	0.416	0.743	0.549	0.7	
TPI m	2.58	3.55	2.4	1.9	1.78	1.81	2.56	2.6	3.95	2.66	2.03	1.63	3.97	2.2	3.4	

*La provincia de Panamá incluye datos agrupados de las provincias de Panamá y Panamá Oeste. Las frecuencias más altas se muestran entre paréntesis debajo de sus alelos correspondientes. aN = número de cromosomas analizados, bNa = Número observado de alelos, cNe = Número efectivo de alelos, dI = Índice de información de Shannon (Lewontin, 1972), eP-HWE = Probabilidad de equilibrio de Hardy-Weinberg calculada con análisis de G cuadrado, fFIS = Coeficiente de consanguinidad individual en subpoblaciones (provincias), gFST = Coeficiente de consanguinidad de subpoblación (Índice de fijación), hFIT = Coeficiente de consanguinidad individual en toda la población, iNm = Flujo genético estimado a partir de $F_{st} = 0,25(1 - F_{st})/F_{st}$, jMP = Probabilidad de coincidencia, kPD = Poder de discriminación, lPE = Poder de exclusión, mTPI = Índice de paternidad típico.

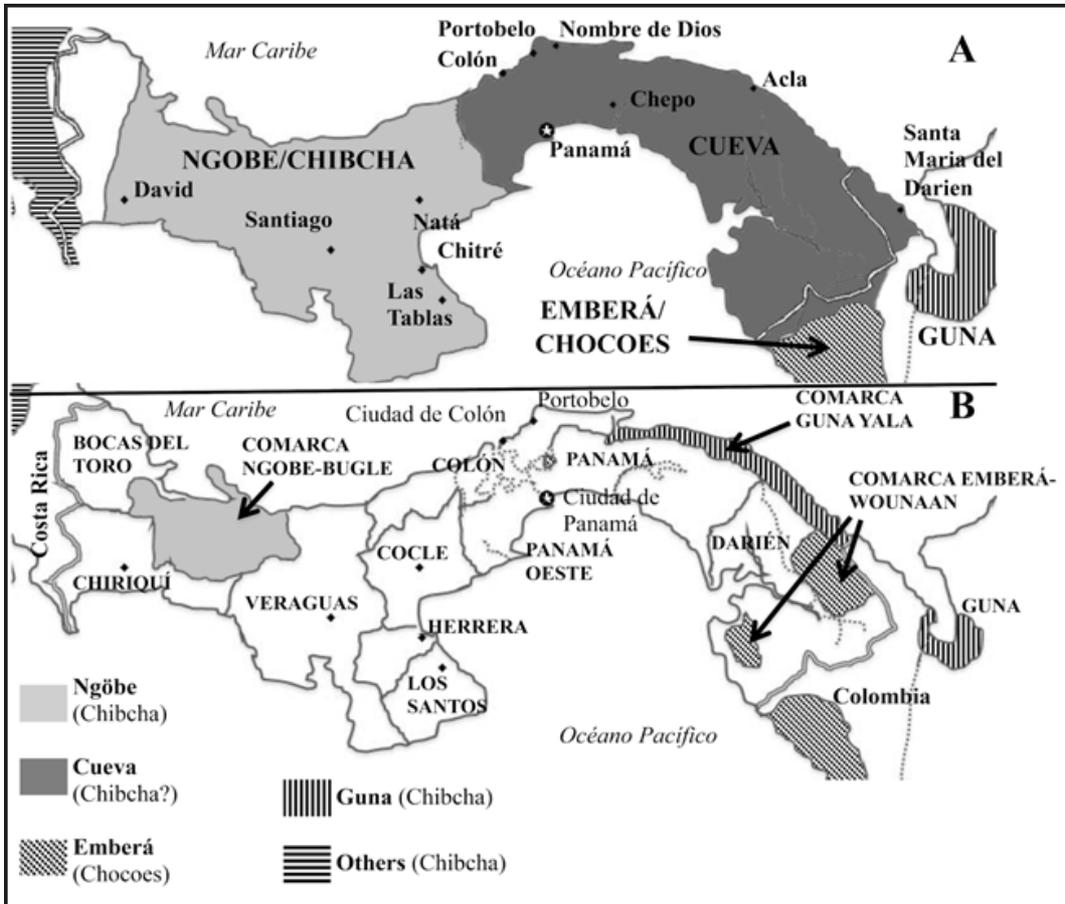


Figura 1. A. Distribución precolombina de amerindios chibchas (actuales Ngobes, Gunas y otros) y chocoes (actuales Emberá y Wounaan) y principales ciudades fundadas por los españoles durante la época colonial. La mayoría de estas ciudades fueron fundadas por los españoles en poblados originalmente ocupados por amerindios, lo cual fue la semilla del mestizaje. La mayoría de estas ciudades aún existen y se convirtieron en la ciudad capital de las actuales provincias panameñas. B. Distribución actual de las tribus amerindias sobrevivientes principalmente en comarcas; y división política de las provincias panameñas actuales. Los datos genéticos para las provincias de Panamá y Panamá Oeste se combinaron en un solo grupo. Las regiones urbanas y ciudades capitales de provincia son los sitios donde se colectaron las muestras de ADN de los participantes. Estas son regiones actualmente habitadas principalmente por mestizos. Este mapa fue adaptado con base en Barrantes et al., (1990), Arias et al., 1990, Romoli, (1987) y Jopling (1994).

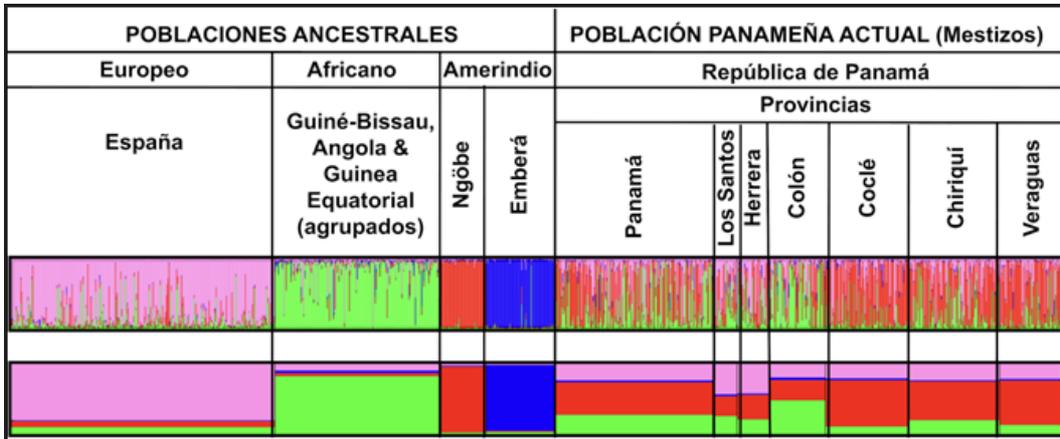


Figura 2. Estructura genética de la población panameña actual. Los análisis se realizaron utilizando STRUCTURE con datos genéticos de 15 marcadores STR de 650 individuos no relacionados de diferentes provincias de Panamá y de datos de estos marcadores publicados previamente de poblaciones ancestrales. Los bloques de códigos de colores corresponden a grupos de genes de poblaciones ancestrales. Los genes de origen europeo se observan en color rosado, los genes de origen africano en color verde, los genes de origen Chibcha (Ngobe) se observan en color rojo y los genes de origen Chocoe (Emberá) se observan en color azul. En la barra superior cada línea vertical representa las proporciones de genes de un solo sujeto (mezcla genética ancestral individual), mientras que la barra inferior representa la sumatoria promedio del aporte ancestral para cada provincia analizada o para cada grupo ancestral. Aunque todas las poblaciones ancestrales muestran algún nivel muy bajo de mezcla, el mismo no es significativo y se pueden distinguir como grupos genéticamente diferentes con patrones genéticos característicos de cada una. La población Ngobe es muy poco diversa con menos de 2% de mezcla, similarmente ocurre en la población Emberá. Se demuestra que la población del país es principalmente mestiza (trihíbrida), altamente polimórfica y que la misma está confirmada por individuos con altas y variables proporciones de mezclas de genes ancestrales de origen amerindio (Ngobe/Chibcha), africano (Oeste de África) y europeo (España). Se muestran los datos por provincia. Los datos genéticos para las provincias de Panamá y Panamá Oeste se combinaron en un solo grupo (Panamá). Para este estudio no se pudieron recolectar muestras en las provincias de Darién y Bocas del Toro, cuyas poblaciones actuales son proporcionalmente muy bajas, por lo que incluirlas en los análisis tendría un impacto muy reducido en los datos presentados para el país. Sin embargo, datos etnográficos, históricos y genéticos indican que Bocas del Toro y Darién son provincias con alta ascendencia africana e indígena (Arias et al., 2002; Ramos et al., 2018), por lo que podrían contribuir un poco más en estos grupos de genes ancestrales para los datos del país que presentamos a continuación. La población total del país mostró un 51% de genes de origen indígena, 25% europeo y 24% africano. Los datos demuestran que los genes indígenas del panameño son todos de origen Ngobe (Chibcha), tanto en las provincias del oeste (Coclé, Los Santos, Herrera, Veraguas, Chiriquí), como en las provincias del este (Panamá, Panamá Oeste y Colón) del país. Estos datos sugieren que independientemente de su lenguaje, los Cueva, que habitaron el este del país, pudieron ser genéticamente Chibcha. Alternativamente, se puede interpretar que independientemente de si los Cueva eran genéticamente chibcha (como los Ngobe) o chocoes (como los Emberá), se extinguieron sin tener una contribución significativa al mestizaje del país. No hay contribución significativa de los genes chocoes (Emberá, en color azul), los cuales se observan como ruido de fondo en una proporción menor del 2% en todas las provincias y

en las poblaciones ancestrales Ngobe, españoles y africanos.

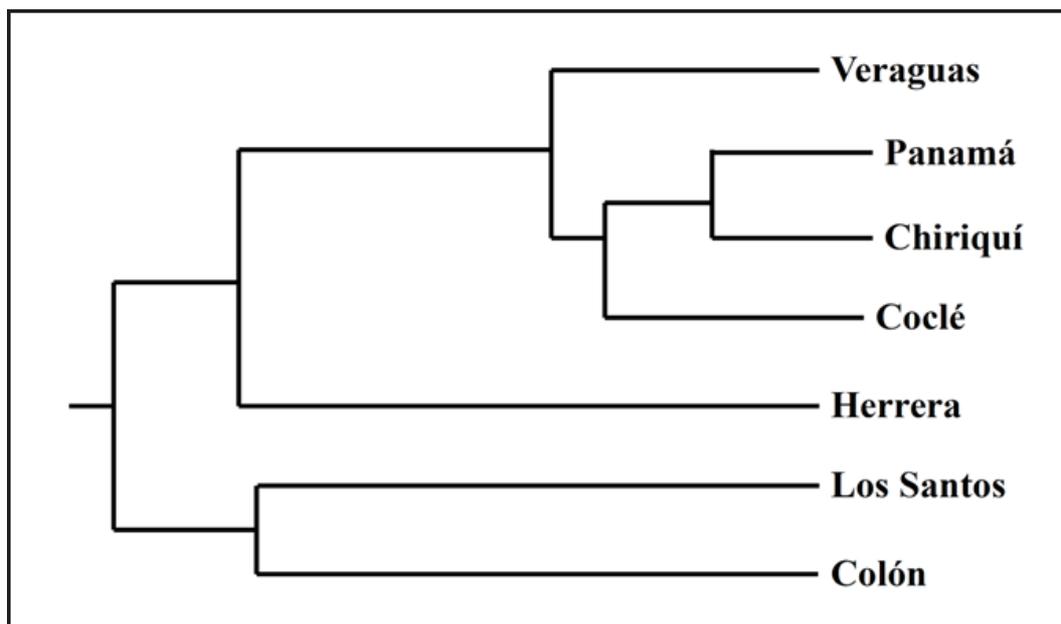


Figura 3. Dendrograma basado en distancia genética de Nei (1978), construido mediante el método del algoritmo UPGMA basado en la información de 15 loci STR de 650 individuos de 7 provincias de Panamá. Se observan dos ramificados principales que aluden al origen ancestral de los genes que las dominan. Las provincias de Veraguas, Panamá (incluye Panamá y Panamá Oeste), Chiriquí y Coclé con proporciones similares de genes ancestrales, pero dominadas principalmente por genes de origen indígena; y las provincias de Los Santos y Colón, con menos genes indígenas, pero alta contribución en genes de origen africano y europeo. La provincia de Herrera se encuentra en medio de ambos grupos, aunque su mestizaje es muy similar a Los Santos.

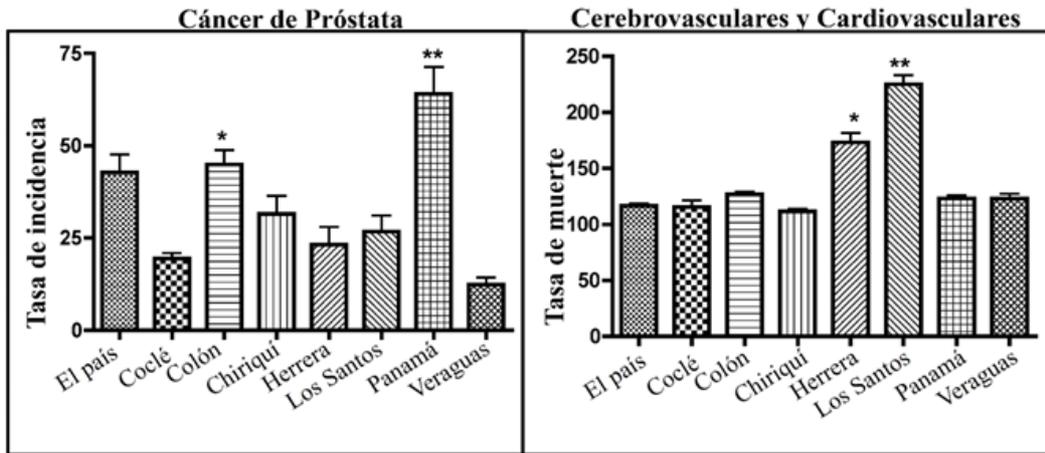


Figura 4. Incidencia y muertes en proporción per cápita por cada 100,000 habitantes de dos grupos de enfermedades que principalmente afectan al país por provincia. Los datos de la provincia de Panamá incluyen Panamá y Panamá Oeste. A. Incidencia de cáncer de próstata por provincia en habitantes varones 1985-1997. Izquierda: ANOVA unidireccional y la postcorrección estadística de Bartlett determinó que las provincias con mayor proporción de genes africanos (47 % Colón y 27% Panamá) mostraron las mayores incidencias de cáncer de próstata en comparación con todas las demás provincias del país con menor contribución de genes africanos ($P < 0,001$ para Panamá y $P < 0,01$ para Colón). Derecha: Muertes causadas por accidentes cerebrovasculares y cardiovasculares por provincia 1998-2010. ANOVA unidireccional y postcorrección estadística de Bartlett demostró que Los Santos y Herrera, ambas provincias con la mayor proporción de genes europeos (42 y 40%, respectivamente) y genes africanos moderadamente altos (25% y 20%, respectivamente) mostraron las tasas más altas de muertes por accidentes cerebrovasculares y cardiovasculares, en comparación con todas las demás provincias que muestran una menor proporción de mezcla europea y africana ($P < 0,001$ para ambas provincias).

Literatura citada

1. Adeyemo AA, Chen G, Chen Y, et al. 2005. Genetic structure in four West African population groups *BMC Genetics* 6:38.
2. Arias TD, Barrantes R, Jorge LF, et al. 1992. Los Cholos de Coclé: determinación de su mezcla racial y orígenes genéticos. *Rev Med Panama Sep*;17(3):180-7.
3. Arias TD, Castro E, Ruiz E, et al. 2002. La mezcla racial de la población panameña. *Rev. Med. Panama* 17:5-17.
4. Assis Poiaras L, Osorio PS, Spanhol FA, et al. 2010. Allele frequencies of 15 STRs in a representative sample of the Brazilian population. *Forensic Science International: Genetics* 4: e61-e63.
5. Bar WB, Brinkmann B, Budowle, et al. 1997. DNA recommendations: Further report of the DNA Commission of the ISFH regarding the use of short tandem repeat systems. *Int. J. Legal Med.* 110(4):175-176.
6. Barrantes R, Smouse PE, Mohrenweiser HW, et al. 1990. Microevolution in lower Central America: Genetic characterization of the Chibcha-speaking groups of Costa Rica and Panama, and a consensus taxonomy based on genetic and linguistic affinity. *Am. J. Hum. Genet.* 46(1):63-84.
7. Basu A, Tang H, Lewis CE, et al. 2009. Admixture mapping of quantitative trait loci for blood lipids in African-Americans. *Hum Mol Genet.* Jun 1;18(11):2091-8.
8. Belezá S, Alves C, Reis F, et al. 2004. 17 STR data (AmpF/STR Identifiler and Powerplex 16 System) from Cabinda (Angola). *Forensic Science International* 141,193-196.
9. Benn-Torres J, Bonilla C, Robbins CM, et al. 2008. Admixture and Population Stratification in African Caribbean Populations. *Annals of Human Genetics* 72,90-98.
10. Bock CH, Schwartz AG, Ruterbusch JJ, et al. 2009. Results from a prostate cancer admixture mapping study in African-American men. *Hum Genet.* Nov;126(5):637-42.
11. Boletín Estadístico, Dirección Nacional de Planificación: Departamento de Registros y Estadísticas de Salud. Ministerio de Salud, República de Panamá. 2010.
12. Bravo V, Rodríguez T, van Wendel de Joode B, et al. 2011. Monitoring pesticide use and associated health hazards in Central America. *Int J Occup Environ Health.* Jul-Sep;17(3):258-69.
13. Camacho MV, Benito C, Figueiras AM. 2007. Allelic frequencies of the 15 STR loci included in the AmpFISTR1 Identifiler™ PCR Amplification Kit in an autochthonous sample from Spain. *Forensic Science International* 173, 241-245.
14. Casper ML, Barnett E, Williams GI, et al. 2003. *Atlas of Stroke Mortality: Racial, Ethnic, and Geographic Disparities in the United States.* Atlanta, GA: Department of Health and Human Services, Centers for Disease Control and Prevention.

15. Castillero Calvo A. 1995. Conquista, evangelización y resistencia. Panamá: Editorial Mariano Arosemena, Instituto Nacional de Cultura.
16. Castro EA, Trejos DE, Berovides-alvarez V, et al. 2007. Genetic Polymorphism and Forensic Parameters of Nine Short Tandem Repeat Loci in Ngöbé and Emberá Amerindians of Panama. *Human Biology*, Oct. 79, no. 5, pp. 563–577.
17. Chakraborty R, Weiss KM. 1986. Frequencies of complex diseases in hybrid populations. *Am J Phys Anthropol*. Aug;70(4):489–503.
18. Cheng CY, Reich D, Coresh J, et al. 2010. Admixture Mapping of Obesity-related Traits in African Americans: The Atherosclerosis Risk in Communities (ARIC) Study. *Obesity (Silver Spring)*. Mar;18(3):563–72.
20. Constenla Umaña A. 1991. Las lenguas del área intermedia: introducción a su estado areal. Universidad de Costa Rica, San José, Costa Rica. 216p.
21. Cooke RG. 1982. Los guaymíes sí tienen historia. El pueblo guaymí y su futuro. Panamá: Comité Patrocinador del “Foro sobre el guaymí y su futuro” y Centro de Estudios y Acción Social.
22. DNA Commission, International Society for Forensic Hemogenetics. 1994. Report concerning further recommendations of the DNA Commission of the ISFH regarding PCR-based polymorphisms in STR (short tandem repeat) system. *Int. J. Legal Med.* 107(3):156–160.
23. Donnan GA, Fisher M, Macleod M, et al. 2008. Stroke. *Lancet* 371 (9624): 1612–23.
24. Falush D, Stephens M, Pritchard JK. 2003. Inference of Population Structure Using Multilocus Genotype Data: Linked Loci and Correlated Allele Frequencies. *Genetics* 164: 1567–1587.
25. Ferreira LB, Mendes-Junior CT, Wiezel CE, et al. 2006. Genomic ancestry of a sample population from the state of São Paulo, Brazil. *Am J Hum Biol*. Sep-Oct; 18(5):702–5.
26. Gonçalves R, Jesus J, Fernandes AT, et al. 2002. Genetic profile of a multi-ethnic population from Guiné-Bissau (west African coast) using the new PowerPlex16 System kit. *Forensic Science International* 129, 78–80.
27. Gower BA, Fernández JR, Beasley TM, et al. 2003. Using Genetic Admixture to Explain Racial Differences in Insulin-Related Phenotypes. *Diabetes* 52:1047–1051.
28. Greenberg JH. 1987. Language in the Americas. Stanford, CA: Stanford University Press.
29. Grugni V, Battaglia V, Perego UA, Raveane A, Lancioni H, Olivieri A, et al. (2015) Exploring the Y Chromosomal Ancestry of Modern Panamanians. *PLoS ONE* 10(12): e0144223. doi:10.1371/journal.pone.0144223
30. Hanis CL, Hewett-Emmett D, Bertin TK, et al. 1991. Origins of U.S. Hispanics. Implications for diabetes. *Diabetes Care* 14, 618–627.
31. Hankey GJ. 1999. Smoking and risk of stroke. *Journal of Cardiovascular Risk* 6 (4): 207–11.

32. Hartl DL, Clark AG. 1989. Principles of population genetics. 2nd ed. Sinauer Associates, Sunderland, MA.
33. Hidalgo PC, Bengochea M, Abilleira D, et al. 2005. Genetic Admixture Estimate in the Uruguayan Population Based on the Loci LDLR, GYPA, HBG, GC and D7S8. *Int J Hum Genet*, 5(3): 217-222.
34. Jaén-Suárez O. 1978. La población del istmo de Panamá del siglo XVI al siglo XX. Panamá: Impresora de La Nación.
35. Jaén-Suárez O. 1998. La Población del Istmo de Panamá. Estudio de Geohistoria, Madrid, Ediciones de Cultura Hispánica, Agencia Española de Cooperación Internacional.
36. Jopling CF. 1994. Indios y Negros en Panamá en los Siglos XVI y XVII. Centro de Investigaciones Regionales de Mesoamérica, Antigua, Guatemala, pp. 524-525.
37. Jorge LF, Arias TD. 1995. Human biodiversity and its effects on the pharmacological variability: CYP2D6 and NAT2 enzymes in Amerind populations of Panama, Colombia and Costa Rica. *Rev Med Panama*. Sep;20(3):98-107.
38. Jorge LF, Eichelbaum M, Griese EU, et al. 1999. Comparative evolutionary pharmacogenetics of CYP2D6 in Ngawbe and Emberá Amerindians of Panama and Colombia: role of selection versus drift in world populations. *Pharmacogenetics Apr*. 9(2):217-28.
39. Jorge-Nebert LF, Eichelbaum M, Griese EU, et al. 2002. Analysis of six SNPs of NAT2 in Ngawbe and Emberá Amerindians of Panama and determination of the Emberá acetylation phenotype using caffeine. *Pharmacogenetics Jan*12(1):39-48.
40. Juárez-Cedillo T, Zuñiga J, Acuña-Alonzo V, et al. 2008. Genetic admixture and diversity estimations in the Mexican Mestizo population from Mexico City using 15 STR polymorphic markers. *Forensic Science International: Genetics* 2: e37-e39.
41. Kittles RA, Weiss KM. 2003. Race, ancestry, and genes: implications for defining disease risk. *Annual Review of Genomics and Human Genetics* 4: 33-67.
42. Klimentidis YC, Miller GF, Shriver MD. 2009. The Relationship Between European Genetic Admixture and Body Composition Among Hispanics and Native Americans. *Am. J. Hum. Biol.* 21:377-382.
43. Kolman CJ, Bermingham E, Cook R, et al. 1995. Reduced mtDNA diversity in the Ngôbé Amerinds of Panamá. *Genetics* 140:275-283.
44. Kolman CJ, Bermingham E. 1997. Mitochondrial and nuclear DNA diversity in the Choco and Chibcha Amerinds of Panama. *Genetics*. Nov;147(3):1289-302.
45. Loewren JA. 1963. Chocó II: Phonological problems. *International Journal of American Linguistics*, 29 (4), 357-371.
46. Loukotka G. 1968. Classification of South American Indian Languages. Latin American Center, University of California, Los Angeles.
47. Maloney G. 1993. El Canal de Panamá y los trabajadores antillanos Panamá 1920: cronología de una lucha. Autoridad del Canal de Panamá, República de Panamá. Obra se-

- lecta. Panamá: Instituto Nacional de Cultura.
48. Morea B, Barrantes R, Marin-Rojas R. 2003. Gene admixture in the Costa Rican population. *Ann Hum Genet.* Jan;67(Pt 1):71-80.
 49. “National Institute of Statistics and Census” Censos Nacionales de Panamá. 2010. Contraloría General de la República. Dirección de Estadística y Censo, Panamá, República de Panamá.
 50. “National Institute of Statistics and Census” Proceso de Transición Demográfica en Panamá. 2016. Contraloría General de la República. Dirección de Estadística y Censo, Panamá, República de Panamá.
 51. Contraloría General de la República. Dirección de Estadística y Censo, Panamá, República de Panamá.
 52. National Vital Statistics Reports. 2007;56(5). Hyattsville, MD: National Center for Health Statistics.
 53. National Vital Statistics Reports. 2009;57(14). Hyattsville, MD: National Center for Health Statistics.
 54. Nebert DW, Jorge-Nebert L, Vesell ES. 2003. Pharmacogenomics and “individualized drug therapy”: high expectations and disappointing achievements. *Am J Pharmacogenomics.* 3(6):361-70.
 55. Nei M. 1972. Genetic distance between populations. *Am Nat* 106:283-292.
 56. Nei M. 1973. Analysis of gene diversity in subdivided populations. *Proc Natl Acad Sci USA* 70:3321-3323.
 57. Nei M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 89:583-590.
 58. Ober C, Vercelli D. 2011. Gene-environment interactions in human disease: nuisance or opportunity? *Trends in Genetics*, Vol. 27, No. 3 107 – 115.
 59. Perego UA, Lancioni H, Tribaldos M, Angerhofer N, Ekins JE, et al. (2012) Decrypting the Mitochondrial Gene Pool of Modern Panamanians. *PLoS ONE* 7(6): e38337. doi:10.1371/journal.pone.0038337
 60. Poveda-Ramos G. 2004. La construcción del ferrocarril de Panamá. *Dyna*, Revista de la facultad de minas, Universidad nacional de Colombia, Medellín. *Dyna*, Vol 71(143): 1-12.
 61. Pritchard JK, Stephens M, Donnelly P. 2000. Inference of Population Structure Using Multilocus Genotype Data. *Genetics* 155: 945-959.
 62. Registro Nacional del Cáncer, Principales Causas de Tumores Malignos. Departamento de Registros y Estadísticas de Salud. Dirección Nacional de Planificación: Ministerio de Salud, República de Panamá. 2007.
 63. Robbins C, Torres JB, Hooker S, et al. 2007. Confirmation study of prostate cancer risk variants at 8q24 in African Americans identifies a novel risk locus. *Genome Res.* 17: 1717-1722.

64. Roger VL, Go AS, Lloyd-Jones DM, et al. 2011. Heart Disease and Stroke Statistics-2011 Update: Heart Disease and Stroke Statistics. *Circulation.*;123:e18-e209.
65. Romoli K. 1987. Los de la lengua de Cueva: Los grupos indígenas del istmo oriental en la época de la conquista española. Bogotá: Instituto Colombiano de Antropología, Instituto Colombiano de Cultura (Ediciones Tercer Mundo).
66. Rosenberg NA, Mahajan S, Ramachandran S, et al. 2005. Clines, clusters, and the effect of study design on the inference of human population structure. *PLoS Genetics* 1(6): e70.
67. Rubi-Castellanos R, Martínez-Cortés G, Muñoz-Valle JF, et al. 2009. Pre-Hispanic Mesoamerican Demography Approximates the Present-Day Ancestry of Mestizos Throughout the Territory of Mexico. *Am J Phys Anthropol* 139:284-294.
68. Sans M. 2000. Admixture studies in Latin America: from the 20th to the 21st century. *Human Biology*, 72:155-177.
69. Scatliff CE, Koski KG, Scott ME. 2011. Diarrhea and novel dietary factors emerge as predictors of serum B12 in Panamanian children. *Food Nutr Bull.* Mar;32(1):54-9.
70. Seldin MF, Tian C, Shigeta R, et al. 2007. Argentine population genetic structure: large variance in Amerindian contribution. *Am J Phys Anthropol.* Mar;132(3):455-62.
71. Shannon CE, Weaver W. 1949. The mathematical theory of communication. Univ. of Illinois Press, Urbana.
72. Slatkin M, Barton NH. 1989. A comparison of three indirect methods for estimating average levels of gene flow. *Evolution* 43:1349-1368.
73. Tang H, Quertermous T, Rodriguez B, et al. 2005. Genetic structure, self-identified race/ethnicity, and confounding in case-control association studies. *American Journal of Human Genetics* 76 (2): 268-75.
74. Torres de Arauz R. 1972. Panorama actual de las culturas indígenas panameñas En: Panamá sus etnias y el canal. Biblioteca de la Nacionalidad. Autoridad del Canal de Panamá.
75. Torun B, Stein AD, Schroeder D, et al. 2002. Rural-to urban migration and cardiovascular disease risk factors in young Guatemalan adults. *Int J Epidemiol.* Feb;31(1):218-26. USA 70:3321-3323.
76. Via M, Gignoux CR, Roth LA, et al. 2011. History Shaped the Geographic Distribution of Genomic Admixture on the Island of Puerto Rico. *PLoS ONE* 6(1): e16513.
77. Wang J. 2003. Maximum-Likelihood Estimation of Admixture Proportions From Genetic Data. *Genetics*, Vol. 164, 747-765.
78. Wang S, Ray N, Rojas W, et al. 2008. Geographic Patterns of Genome Admixture in Latin American Mestizos. *PLoS Genet* 4(3): e1000037.
79. Weir BS. 1990. Genetic Data Analysis. Sinauer Associates, Sunderland, MA.
80. Zeigler-Johnson CM, Spangler E, Jalloh M, et al. 2008. Genetic susceptibility to prostate cancer in men of African descent: implications for global disparities in incidence and outcomes. *Can J Urol.* February; 15(1): 3872-3882.

C A P Í T U L O 2

Polimorfismo genético y
parámetros forenses de
nueve loci de repeticiones
cortas en tándem en
amerindios Ngöbe y
Emberá de Panamá

*Edgardo Castro-Pérez, Diomedes E. Trejos, Vicente
Berovides-Alvarez, Tomás D. Arias y Carlos Ramos Delgado*

Resumen

Se analizaron nueve loci STR (*CSF1PO*, *TPOX*, *TH01*, *F13A01*, *FESFPS*, *VWA*, *D16S539*, *D7S820* y *D13S317*) en individuos amerindios Ngobé y Emberá no relacionados de Panamá. La prueba de chi-cuadrado demostró diferencias estadísticamente significativas ($P < 0,001$) en las frecuencias alélicas para todos los marcadores excepto uno (*D16S539*; $P < 0,01$). En ambas poblaciones los alelos con las frecuencias más altas fueron los mismos en siete loci. Sin embargo, también hubo diferencias notables en el locus *TPOX*, que mostró sus frecuencias más altas en los alelos 11 (0,48) y 6 (0,54) para los Ngobé y Emberá, respectivamente. Curiosamente, estos alelos están presentes en una población y ausentes en la otra lo que sugiere que podrían ser distintivos de cada población. Estos resultados demuestran que, a pesar de que cada población pertenece a un grupo lingüístico diferente (Chibchan [Ngobé] y Chocoan [Emberá]), ambos mantienen algunas similitudes en la distribución de las frecuencias de alélicas. Tres loci (*TPOX*, *VWA*, y *F13A01*) en Ngobé y dos loci (*TH01* y *TPOX*) en Emberá no se encontraban en equilibrio de Hardy-Weinberg. El análisis de los marcadores STR demuestra que, a pesar de sus bajos niveles de polimorfismos genéticos, la mayoría podrían ser informativos para fines forenses, mostrando un poder combinado de discriminación de 0,9999 para ambas poblaciones amerindias. Sin embargo, el poder de exclusión en Ngobé fue bajo, particularmente en los loci *TH01* (0,04) *FESFPS* (0,08). Los poderes de exclusión combinados fueron 0,9338 y 0,9890 en Ngobé y Emberá, respectivamente. Además, el índice de paternidad combinado en los Ngobé fue considerablemente bajo (2,58), y en Emberá fue de 40.44, también bajo. Los bajos niveles de polimorfismo genético sugieren que el uso de loci adicionales que complementen los nueve loci es recomendable para pruebas forenses y de paternidad en ambas poblaciones, particularmente en los Ngobé.

Palabras claves: 15 Loci STR, diversidad genética, parámetros forenses, parámetros de paternidad, amerindios Ngobe y Embera.

Abstract

Nine STR loci (*CSF1PO*, *TPOX*, *TH01*, *F13A01*, *FESFPS*, *VWA*, *D16S539*, *D7S820*, and *D13S317*) were analyzed in unrelated Ngobe and Embera Amerindians of Panama. The chi-square test demonstrated statistically significant differences ($P < 0.001$) in the allele frequencies for all markers except one (*D16S539*; $P < 0.01$). Both populations shared their alleles with the highest

frequencies in seven loci. However, there were also noticeable differences at the *TPOX* locus, which showed its highest frequencies at alleles 11 (0.48) and 6 (0.54) for the Ngobe and Embera, respectively. Interestingly, these alleles are present in one population and absent in the other, suggesting they could be distinctive for each population. These results demonstrate that, even though each population belongs to a different linguistic stock [Chibchan (Ngobe) and Chocoan (Embera)], both retain some similarities in their allele-frequency distributions. Three loci (*TPOX*, *VWA*, and *F13A01*) in the Ngobe and two (*TH01* and *TPOX*) in the Embera departed from Hardy-Weinberg equilibrium. The analysis of the STR markers demonstrates that, despite their low levels of genetic polymorphisms, most of them could be informative for forensic purposes, showing a combined power of discrimination of 0.9999 for both Amerindian populations. However, powers of exclusion in the Ngobe were meager, particularly at the *TH01* (0.04) and *FESFPS* (0.08) loci. The combined powers of exclusion were 0.9338 and 0.9890 for the Ngobe and the Embera, respectively. Furthermore, the combined typical paternity index in the Ngobe was considerably low (2.58), and in the Embera, it was 40.44, which is also very low. The low genetic polymorphism levels suggest that additional loci supplementing the battery of the nine loci is recommended for paternity and forensic tests in both populations, particularly for the Ngobe.

Keywords: 15 STR loci, genetic diversity, forensic parameters, paternity parameters, Ngobe and Embera Amerindians.

Introducción

Se ha demostrado que las repeticiones cortas en tándem (STRs) son excelentes marcadores polimórficos para la caracterización genética tanto de individuos como de poblaciones.

Debido a su gran diversidad interétnica en las poblaciones a nivel mundial, los STR han contribuido a una comprensión de las relaciones genéticas, la evolución y la migración entre poblaciones (Rangel-Villalobos et al. 2000). Además, el rápido desarrollo de la tecnología del PCR ha hecho posible el uso generalizado de STRs en pruebas de paternidad y medicina forense (Mastana y Singh 2002). Recientemente, estos marcadores han sido utilizados en estudios de mezcla étnica (trihíbrida) en poblaciones mestizas (Cerdeña-Flores et al. 2002a, 2002b). La aplicación de estos marcadores al estudio de poblaciones de amerindios es un área excitante de la genética de poblaciones porque los amerindios representan la colonización humana más reciente de un continente previamente deshabitado (30.000-12.000 años B.R; Batista et al. 1995).

La evidencia arqueológica, etnohistórica, lingüística y genética indica que las tribus Ngobé (también conocidas como Ngawbe-Guaymi) y Emberá se desarrollaron *in situ* durante varios milenios, procedente de grupos migratorios de paleo-amerindios que llegaron a poblar las Américas desde el noreste de Asia (Kolman et al. 1995). Estos paleo-amerindios cruzaron el puente terrestre de Bering desde Mongolia/Manchuria/región sureste de Siberia (Neel et al. 1994). Aunque los Ngobé y los Emberá viven muy cerca y representan dos familias lingüísticas distintas: Chibchan y Choco/Paezan, respectivamente (Kolman y Bermingham 1997). Recientemente las autoridades de Costa Rica han informado que los Ngobé han estado emigrando de Panamá. El movimiento constante de los Ngobé ha sido bien documentado en las fronteras de ambos países.

La población Ngobé se originó, desarrolló y aún habita el oeste de Panamá, especialmente en la comarca Ngobé-Buglé, que es un territorio que abarca 6,673 km² (Figura 1). La población Ngobé está compuesta por aproximadamente 142,986 miembros (Dirección de Estadística y Censo 2000) y constituye la tribu amerindia pura más grande que existe hoy en día (Jorge et al. 1999). La tribu Emberá es considerada una población de origen amazónico que se asentó en el este de Panamá en los últimos 100 años a través de la constante migración hacia el oeste desde la provincia del Chocó en Colombia (Jorge-Nebert et al. 2002). Habitan principalmente en la Comarca Emberá-Wounaan, un territorio de 4,398 km² (Figura 1), con una población de aproximadamente 19,422 individuos (Dirección de Estadística y Censo 2000). Sin embargo, también se distribuyen en la República de Colombia, con una población de 40,000 individuos (Jorge-Nebert et al. 2002).

Aunque estos dos grupos amerindios representan menos del 8% del total de la población de Panamá (Dirección de Estadística y Censo 2000), es importante tener en cuenta que su contribución genética en la población mestiza de Panamá es aproximadamente del 36% (Arias et al. 2002). La población mestiza es una mezcla de caucásicos, negros y amerindios, y constituye el 85% de la población panameña.

Este alto nivel de mezcla amerindia es significativo debido a las implicaciones biomédicas de susceptibilidad a algunas enfermedades y trastornos metabólicos conocidos como el Síndrome del Nuevo Mundo (Salzano y Callegari-Jacques 1988). Estudios genéticos previos en estas dos poblaciones amerindias de Panamá realizados con marcadores clásicos (Barrantes et al. 1990), secuencias de ADN mitocondrial y nuclear (Kolman y Bermingham 1997) reportan una baja diversidad genética.



Figura 1. Distribución geográfica de las comarcas de amerindios Ngobé (gris) y Emberá (negro) en el oeste y este, respectivamente, de la República de Panamá, donde las muestras fueron recolectadas.

Sin embargo, a pesar de que se utilizaron marcadores STR, no se obtuvieron datos de frecuencia de alelos, lo que limitó el conocimiento y la aplicación de estos datos. Otros estudios genéticos, incluyendo investigaciones farmacogenéticas sobre polimorfismos de las enzimas N-acetiltransferasa-2 (NAT2) y citocromo P450 CYP2D6 (debrisoquina 4-hidroxisilasa), mostraron diferencias a nivel de fenotipos y genotipos como metabolizadores lentos y rápidos (Jorge et al. 1999; Jorge-Nebert et al. 2002). Por lo tanto, son necesarios más estudios sobre la naturaleza y el alcance de la diversidad genética de loci STRs en estas poblaciones de amerindios.

Las tribus amerindias de Panamá y las de otros países latinoamericanos están siendo constantemente absorbidas por comunidades no nativas, cambiando así la estructura genética de la población (Rangel-Villalobos et al. 2000). Esta tendencia hacia la aculturación y la mezcla étnica observadas entre los amerindios probablemente se acelere en un futuro próximo. La pregunta es si los grupos perderán su identidad biológica (genética) y cultural, o si resistirán este proceso de homogeneización (Salzano y Callegari-Jacques 1988). En consecuencia, se requiere de manera urgente información sobre los acervos genéticos de estas poblaciones nativas para contribuir a resolver y prevenir los problemas biomédicos de poblaciones de amerindios y mestizos. La información genética también podría ser útil para reconstruir en detalle los acontecimientos históricos que condujeron al asentamiento de América.

Los nueve loci STR descritos en este estudio han sido de gran utilidad en Panamá en la resolución de delitos y determinación de paternidad. A pesar de su uso, estos loci no han sido utilizados en estudios de genética de poblaciones en nuestro país. Aunque se han reportado una gran cantidad de estudios sobre STR en diferentes poblaciones a nivel mundial (Mastana y Singh 2002; Yunis et al. 2000), las bases de datos están restringidas en gran medida a poblaciones étnicamente definidas; poblaciones aisladas con tamaños efectivos pequeños y relativamente escasas (Sun et al. 2003). Debido a que los Ngobé y los Emberá también habitan en Costa Rica y Colombia, respectivamente, el impacto de estos resultados tiene también relevancia en estos países.

En este estudio se analizó la diversidad genética de las poblaciones amerindias Ngobé y Emberá, utilizando nueve marcadores STR. Las frecuencias alélicas obtenidas en cada población fueron comparadas. También se calcularon algunos parámetros poblacionales para evaluar la deficiencia alélica y de heterocigotos. Además, se incluyeron algunos parámetros forenses para realizar cálculos más precisos con fines de identificación genética en estas poblaciones. Estos datos contribuirán a una mejor comprensión de las estimaciones de mezclas de la población mestiza de Panamá, la cual representa la mayor parte del país.

Materiales y métodos

Selección de amerindios y colecta de muestras.

Los individuos seleccionados para el estudio fueron entrevistados para obtener la información biográfica que incluía nombres, lugares de nacimiento e idiomas hablados por sus padres y abuelos. El consentimiento informado fue obtenido de las autoridades comunitarias competentes y de los individuos seleccionados. Los individuos seleccionados para este estudio exhibían las características fenotípicas de amerindios, nacieron en su comarca, hablaban la lengua nativa, y tenía padres y abuelos pertenecientes al mismo grupo amerindio.

Las muestras de Ngobé se obtuvieron de individuos de cuatro asentamientos ubicados en regiones muy separadas a lo largo de la comarca Ngobe-Buglé, la cual es un vasto territorio en la provincia de Bocas del Toro, en el oeste de Panamá (ver Figura 1).

Las muestras Emberá fueron obtenidas de individuos en cinco asentamientos Emberá distribuidos en toda la comarca Emberá-Wounaan en la provincia del Darién en el este Panamá (ver Figura 1).

Todas las muestras de sangre en ambas poblaciones se obtuvieron de individuos no relacionados (hasta al menos el cuarto grado de consanguinidad), voluntarios sanos de ambos sexos. Se obtuvieron muestras de sangre venosa (10 ml) mediante venopunción y transferidas a tubos con anticoagulantes EDTA o ACD. Los tubos fueron rotados suavemente para garantizar que el EDTA y la sangre se mezclaran adecuadamente. Posteriormente, la sangre fue almacenada a 4°C, durante un máximo de 3-4 días y luego a -20°C si no se procesaba inmediatamente.

En el caso de las muestras de sangre de los Emberá, se extrajo de los tubos 125 µL de sangre utilizando pipetas plásticas, que fueron colocados en el centro de cada uno de los cuatro círculos de las tarjetas FTA (DNA Testing Center Inc., Eules, Texas). Se utilizó una perforadora de papel para cortar un círculo de 1 mm de diámetro. El círculo pequeño de papel fue lavado de acuerdo con el procedimiento descrito por el fabricante. La amplificación de los STRs se realizó directamente del ADN en el papel tratado.

El ADN de muestras de Ngobé se extrajo de leucocitos periféricos mediante una modificación del procedimiento descrito por Miller et al. (1988). La calidad del ADN fue verificada mediante electroforesis en gel de agarosa utilizando el marcador de ADN K562 como referencia (Promega Corp., Madison, Wisconsin).

Este estudio fue aprobado por el Comité de Ética de la Universidad de Panamá y cumplió con las leyes de la República de Panamá. Los individuos que no hablaban una lengua amerindia, presentaban alguna evidencia obvia de mezcla reciente con otras poblaciones amerindias o no amerindias (por ejemplo, pelo rizado) y no nacieron en regiones amerindias o personas cuyos padres no cumplían estos criterios fueron excluidos del estudio.

Amplificación STR y Genotipado

Las amplificaciones se realizaron con un Termociclador Perkin Elmer modelo 2400 (Perkin Elmer, Waltham, Massachusetts). Todas las reacciones en cadena de la polimerasa se llevaron a cabo en un volumen total de 25µL conteniendo aproximadamente 50 ng de ADN genómico (Ngobé) o un círculo de papel FTA de 1 mm diámetro (Emberá). Los nueve loci STR (CSF1PO, TPOX, TH01, F13A01, FESFPS, VWA, D16S539, D7S820 y D13S317) fueron tipificados utilizando el kit GenePrint STR Systems (Promega). El PCR multiplex, la electroforesis en poliacrilamida, la detección mediante tinción con nitrato de plata y la identificación de alelos se realizó de acuerdo con los protocolos recomendados por el fabricante. Además, se

cumplieron con todas las recomendaciones de la Comisión de ADN de la Sociedad Internacional de genética forense (Comisión de ADN 1994; Bar et al. 1997).

Análisis estadístico

Los análisis estadísticos se llevaron a cabo utilizando el software PopGene, versión 1.32 (disponible en <http://www.ualberta.ca/~fyeh/>), para calcular frecuencias alélicas. Se determinaron las frecuencias genotípicas observadas y esperadas utilizando el algoritmo de Levene (1949) mediante dos métodos: pruebas de chi-cuadrado y razón de verosimilitud (G_2) para determinar equilibrio de Hardy-Weinberg por locus por población. Se determinaron también medidas de diversidad genética, como el número efectivo de alelos (N_e) (Hartl y Clark 1989) por locus y combinados, el número observado de alelos (N_a) y el índice de información de Shannon (I) (Shannon y Weaver 1949) por locus y combinado. Se utilizó el software Statistica, versión 6.6, para comparar frecuencias alélicas por locus de ambas poblaciones por el método de Hedrick (1985). Los niveles de heterocigosidad observada (H_o) y el número de alelos por locus (N_a) se compararon mediante la prueba de Wilcoxon-Mann-Whitney (Siegel y Castellan 1998).

La plantilla de Microsoft Excel PowerStatsV12 (Promega Corp.) se utilizó para calcular parámetros forenses, que incluyeron la probabilidad de coincidencia y poder de discriminación. Además, se estimaron varios parámetros de paternidad: poder de exclusión, índice típico de paternidad y porcentaje de homocigotos y heterocigotos.

Resultados

Las frecuencias alélicas en cada locus STR, así como el porcentaje de homocigotos y heterocigotos, se presentan en la Tabla 1. Debido a que algunas muestras de ADN no se pudieron amplificar de manera óptima en ciertos loci, el número de cromosomas analizados difieren en cierta medida de un locus a otro.

Tabla 1. Frecuencias alélicas de nueve loci STR en amerindios Ngöbe (NG) y Emberá (EM) de Panamá.

Alelo	CSFIPO		TPOX		TH01		F13A01		FESFPS		VWA		D16S539		D7S820		D13S317	
	NG	EM	NG	EM	NG	EM	NG	EM	NG	EM	NG	EM	NG	EM	NG	EM	NG	EM
3-2							0.48	0.38										
3-3								0.01										
4							0.31	0.11										
5	0.00	0.00	0.01	0.00	0.00	0.00	0.13	0.24					0.00	0.00				
6	0.00	0.00	0.00	0.54	0.86	0.00	0.07	0.02					0.00	0.00	0.00	0.00		
7	0.00	0.00	0.00	0.39	0.13	0.01	0.01	0.23	0.00	0.00			0.00	0.00	0.00	0.00	0.00	0.00
8	0.00	0.05	0.40	0.02	0.00	0.55	0.00	0.01	0.00	0.00			0.01	0.00	0.00	0.05	0.00	0.01
9	0.06	0.04	0.01	0.00	0.00	0.01	0.00	0.00	0.00	0.00			0.15	0.18	0.01	0.01	0.56	0.33
9-3			0.05	0.01	0.00													
10	0.03	0.27	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.10			0.37	0.37	0.18	0.25	0.04	0.13
11	0.33	0.14	0.48	0.00	0.00	0.20	0.00	0.00	0.77	0.65	0.00	0.00	0.25	0.19	0.43	0.45	0.05	0.10
12	0.46	0.43	0.11	0.00	0.23	0.00	0.00	0.00	0.21	0.14	0.00	0.00	0.22	0.22	0.37	0.21	0.10	0.22
13	0.13	0.07	0.01	0.00	0.01	0.00	0.00	0.01	0.11	0.11	0.00	0.00	0.00	0.03	0.01	0.04	0.17	0.14
14	0.00	0.00				0.00	0.00	0.00	0.00	0.00	0.06	0.06	0.00	0.01	0.00	0.00	0.09	0.07
15	0.00	0.00				0.00	0.00	0.00			0.09	0.09	0.00	0.00				
16						0.00	0.00				0.60	0.56						
17											0.23	0.19						
18											0.01	0.11						
19											0.01	0.00						
20											0.00	0.00						
21											0.00	0.00						
Homocigosidad (%)	23.3	28.4	39.3	50.5	75.8	38.2	51	32.6	64.6	41.3	54	42.2	43.9	27.1	31.6	36.5	43.9	15.6
Heterocigosidad (%)	76.7	71.6	60.7	49.5	24.2	61.8	49	67.4	35.4	58.7	46	57.8	56.1	72.9	68.4	63.5	56.1	84.4
Cromosomas totales	120	176	122	186	124	178	98	92	96	92	100	90	114	192	114	192	114	192

Tabla 2. Patrones de distribución de frecuencias alélicas por locus en las poblaciones Ngöbe y Emberá.

Grupo	Locus	Patrón de Frecuencia Alélica
I	TPOX, TH01	Los alelos más frecuentes y otros alelos para cada población son diferentes, excepto uno.
II	FESFPS, D13S317, F13A01	Alelos predominantes similares en ambas poblaciones, pero difieren en sus frecuencias (>0.1); lo mismo para otros alelos.
III	D7S820, CSF1PO, VWA	Menos diferencias en las frecuencias de los alelos predominantes para ambas poblaciones (<0.1); otros alelos difieren, pero con bajas frecuencias
IV	D16S539	Todos los alelos muestran pocas diferencias entre las poblaciones

Para todos los loci, la prueba de chi-cuadrado mostró diferencias estadísticamente significativas ($P < 0,001$) entre las frecuencias alélicas de ambas poblaciones, excepto para el locus D16S539 ($P < 0,01$).

En siete loci STR los alelos con las frecuencias más altas fueron los mismos en ambas poblaciones de amerindios. Estos alelos fueron CSF1PO*12, F13A01 *3.2, FESFPS*11, VWA*16, D16S539*10, D7S820*11 y D13S317*9. Sin embargo, se observaron diferencias en locus TPOX, en el que las frecuencias más altas se encontraron en los alelos *11 (0,48) y *6 (0,54) para los Ngobé y Emberá, respectivamente. Del mismo modo en el locus TH01 los alelos con las frecuencias más altas fueron *6 (0,86) y *8 (0,55) para Ngobé y Emberá respectivamente. Con la excepción de D16S539, los otros ocho loci mostraron diferencias estadísticamente significativas en las frecuencias alélicas en ambas poblaciones. Estas diferencias generaron patrones específicos, dando lugar a cuatro grupos (Tabla 2). El primer grupo contiene el loci que mejor diferencia a ambas poblaciones (TPOX y TH01). Los grupos II y III permiten también diferenciar las poblaciones, pero en menor grado, ya que tienen los mismos alelos predominantes.

En los Ngöbe, la heterocigosidad observada más alta fue de 76,7% en el locus CSF1PO y la más baja en el locus TH01 (24,2 %). En cambio, en los Emberá las heterocigosidades más altas y más bajas fueron 84,4% y 49,5% en los loci D13S317 y TPOX, respectivamente. En todos los loci Emberá, los niveles de heterocigosidad observada y esperada fueron más altos que en los Ngöbe (para H_0 hubo diferen-

cias significativas, con base a la prueba de Wilcoxon-Mann-Whitney $P = 0,025$) (Tabla 3). Se observaron pocas diferencias en el valor promedio entre ambas estimaciones por población y por locus. Los Emberá también mostraron una proporción menor de déficit de heterocigotos por locus, por los valores positivos del índice F_{IS} . Esta tendencia hacia niveles más elevados de heterocigosidad en la población Emberá fue también evidenciada por su diversidad alélica estimada utilizando tres medidas (Tabla 4). Los Emberá mostraron valores medios de todos los índices más elevados que los Ngöbe; las diferencias de N_a fueron estadísticamente significativas (prueba de Wilcoxon-Mann-Whitney $P = 0,031$). Estas diferencias son causadas principalmente por los loci TPOX, F13A01 y D13S317, los cuales presentan siete alelos cada uno. Por otro lado, los Ngöbe sólo tienen dos loci (VWA y D13S317) con seis alelos cada uno.

Tabla 3. Valores promedio y desviaciones típicas de los niveles de heterocigosidad observados y esperados para las poblaciones Ngöbe y Emberá

Población	H_o^a	H_e^b	Diferencia Promedio	Diferencia Por Locus	Deficiencia Heterocigotos (%)
Ngöba	0.522 (0.158)	0.567 (0.160)	0.045	0.071	66.7
Emberá	0.656 (0.102)	0.665 (0.087)	0.007	0.043	55.5

^aHeterocigosidad observada (desviación típica entre paréntesis).

^bHeterocigosidad esperada (desviación típica entre paréntesis).

La distribución de las frecuencias genotípicas para la mayoría de los loci (seis en Ngöbe y ocho en los Emberá) no mostraron desviación del equilibrio de Hardy-Weinberg. En ambas pruebas, el locus F13A01 en Ngöbe y el TPOX en Emberá mostraron desequilibrio genético ($P < 0,05$). Otros dos loci en Ngöbe también mostraron desequilibrio (VWA y D16S539), pero con $P < 0,010$. Estos loci mostraron deficiencias significativas de heterocigotos, con diferencias de más de 0,10 entre los valores observados y esperados como resultado de los bajos niveles de diversidad, probablemente causado por un cuello de botella poblacional y un desequilibrio genético.

Los parámetros forenses y de paternidad se muestran en la Tabla 5. La probabilidad de coincidencia en Ngöbe osciló entre 0,124 y 0,602 en los loci D16S539 y TH01, respectivamente. De manera similar, en la población Emberá la probabilidad de coincidencia varió de 0,085 a 0,279 en los loci D13S317 y TPOX.

Table 4. Parámetros de diversidad alélica por locus en las poblaciones Ngöbe (NG) y Emberá (EM).

Locus	N _o ^a		N _e ^b		F ^c		FIS ^d	
	NG	EM	NG	EM	NG	EM	NG	EM
CSF1PO	5	6	2.9777	3.2103	1.2619	1.3828	-0.1543	-0.0688
TPOX	5	7	2.6278	2.5351	1.1019	1.1042	0.0208	-0.0205
TH01	3	5	1.3135	2.2638	0.4303	0.9582	-0.0136	0.114
F13A01	5	7	2.8858	3.7518	1.218	1.4698	0.2505	0.0812
FESFPS	4	4	1.5679	2.1428	0.6225	1.0239	0.0222	-0.1006
VWA	6	5	2.354	2.7931	1.1221	1.2612	0.2003	0.0654
D16S539	5	6	3.6805	3.9343	1.3721	1.4801	0.2188	0.0223
D7S820	5	6	2.8203	3.2022	1.1255	1.3417	-0.0601	0.0912
D13S317	6	7	2.7476	4.7665	1.3343	1.7003	0.1173	-0.0678
AV/TypDeve	4.89	5.89	2.55	3.1778	1.0654	1.3076	-	-
	(0.92)	(1.05)	(0.73)	(0.86)				

^a Número observado de alelos por locus.

^b Número efectivo de alelos por locus.

^c Índice de diversidad de Shannon.

^d Índice de Fijación.

^e Promedio y desviación típica (entre paréntesis) para cada parámetro por población

Tabla 5. Parámetros forenses y de paternidad para nueve loci STR en las poblaciones amerindias Ngöbe y Emberá.

Locus	Ngöbe				Emberá			
	MP ^a	PDb	PE ^c	TPI ^d	MP	PD	PE	TPI
CSF1PO	0.2	0.8	0.539	2.14	0.131	0.869	0.453	1.76
TPOX	0.255	0.745	0.299	1.27	0.279	0.721	0.183	0.99
TH01	0.602	0.398	0.042	0.66	0.23	0.77	0.313	1.31
D16S539	0.124	0.876	0.247	1.14	0.115	0.885	0.475	1.85
D7S820	0.202	0.798	0.404	1.58	0.154	0.846	0.336	1.37
D13S317	0.173	0.827	0.247	1.14	0.085	0.915	0.683	3.2
F13A01	0.184	0.816	0.179	0.98	0.133	0.867	0.389	1.53
FESFPS	0.465	0.535	0.088	0.77	0.254	0.746	0.276	1.21
VWA	0.245	0.755	0.155	0.93	0.172	0.828	0.265	1.18
Combinado	-	0.9999	0.9338	2.58	-	0.9999	0.989	40.44

^a Probabilidad de concordancia

^b Poder de discriminación

^c Poder de exclusión

^d Índice típico de paternidad.

Tabla 6. Comparaciones de alelos con las frecuencias más altas para nueve loci STR en Ngöbe, Emberá y otras poblaciones amerindias representativas de América del Sur y del Norte y en poblaciones asiáticas y en otras poblaciones de referencia de los Estados Unidos.

Alelo con la frecuencia más alta y su frecuencia									
Población	CSF1PO	TPOX	TH01	F13A01	FESFPS	VWA	D16S539	D7S820	D13S317
Amerindios de Centroamérica									
Ngöbe (Panamá)	12 (0.46)	11 (0.48)	6 (0.86)	3.2 (0.48)	11 (0.77)	16 (0.60)	10 (0.37)	11 (0.43)	9 (0.56)
Emberá (Panamá-Colombia)	12 (0.43)	6 (0.54)	8 (0.55)	3.2 (0.38)	11 (0.65)	16 (0.56)	10 (0.37)	11 (0.45)	9 (0.33)
Amerindios de Suramérica									
Kichwa (Ecuador)	10 and 12 (0.31)	8 (0.51)	6 (0.48)	–	–	16 (0.60)	10 (0.29)	11 (0.37)	9 (0.24)
Terena (Brasil)	11 (0.35)	8 (0.47)	7 (0.48)	–	–	16 (0.41)	11 (0.33)	11 (0.39)	12 (0.33)
Mapuche (Argentina)	12 (0.36)	–	6 (0.53)	7 (0.34)	11 (0.50)	16 (0.58)	–	–	–
Tehuelche (Argentina)	12 (0.41)	–	7 (0.43)	3.2 (0.35)	11 (0.71)	16 (0.42)	–	–	–
Wichi (Argentina)	12 (0.58)	–	7 (0.72)	5 (0.35)	11 (0.54)	16 (0.51)	–	–	–
Amerindios de Norteamérica									
Otomi (Sierra Madre, México)	13 (0.42)	8 (0.47)	7 (0.41)	–	–	16 (0.40)	12 (0.35)	11 (0.33)	9 (0.35)
Otomi (Valle Ixmiquilpan, México)	13 (0.52)	8 (0.47)	7 (0.43)	–	–	16 (0.36)	10 and 12 (0.29)	11 (0.42)	9 (0.30)
Huasteco (México)	13 (0.41)	8 (0.54)	7 (0.43)	–	–	16 (0.56)	11 (0.30)	11 (0.31)	9 (0.36)
Poblaciones Asiáticas									
Han (China)	12 (0.37)	8 (0.51)	9 (0.51)	–	–	14 (0.25)	9 (0.28)	11 (0.34)	8 (0.28)
Tu (China)	12 (0.45)	8 (0.50)	9 (0.47)	–	–	17 (0.28)	–	11 (0.30)	11 (0.24)
Sala (China)	12 (0.38)	8 (0.51)	9 (0.51)	–	–	16 (0.31)	–	11 (0.31)	8 and 11 (0.25)
Japón	12 (0.42)	8 (0.43)	9 (0.39)	–	–	17 (0.28)	9 (0.34)	11 (0.35)	8 (0.26)
Poblaciones de referencia de									
Estados Unidos									
Afro-americanos	10 (0.27)	8 (0.35)	7 (0.38)	5 (0.34)	11 (0.36)	16 (0.26)	11 (0.32)	10 (0.35)	12 (0.41)
Europeo-americanos	12 (0.33)	8 (0.53)	9.3 (0.33)	7 (0.33)	11 (0.44)	17 (0.26)	11 (0.32)	10 (0.30)	11 and 12 (0.31)
Hispano-americanos	12 (0.36)	8 (0.50)	7 (0.31)	7 and 5 (0.23)	11 (0.45)	16 (0.30)	11 (0.30)	10 (0.28)	12 (0.24)

El poder de la discriminación en Ngöbe osciló entre 0,398 y 0,876 para los loci TH01 y D16S539, respectivamente, mientras que en la población Emberá el poder de discriminación osciló entre 0,721 (TPOX) y 0,915 (D13S317). El poder combinado de discriminación fue de 0,9999 en ambas poblaciones amerindias. El poder de exclusión en la población Ngöbe osciló entre 0,042 y 0,539 para TH01 y CSF1PO, respectivamente, y el poder combinado de exclusión para nueve loci fue 0,9338. En los Emberá, el poder de exclusión varió de 0,183 a 0,683 en los loci TPOX y D13S317 respectivamente y el poder de exclusión combinado fue 0,9890. El índice

de paternidad típica en Ngöbe varió de 0,66 a 2,14 para TH01 y CSF1PO, respectivamente mientras que el valor combinado fue 2,58. En los Emberá, el índice de paternidad típico varió entre 0,99 y 3,2 en TPOX y D13S317 respectivamente, y el índice de paternidad combinada fué 40,44.

Discusión

Las comparaciones de alelos entre los Ngöbe y los Emberá revelaron algunas similitudes y discrepancias entre las frecuencias. La importancia de estos resultados se pueden apreciar mejor cuando se contrastan con los estudios de poblaciones representativas de amerindios de América del Norte y del Sur, poblaciones de Asia (noreste asiático) y poblaciones de referencia de Estados Unidos (Tabla 6). Para simplificar, en la Tabla 6 incluimos sólo las frecuencias más altas para estas poblaciones. Aunque no realizamos análisis estadísticos, encontramos algunas similitudes y discrepancias interesantes entre las poblaciones. En siete de los nueve loci STR, los alelos con las frecuencias más altas eran los mismos en ambas poblaciones (Ngöbe y Emberá). Estas similitudes también son frecuentes en otras poblaciones amerindias de México (Barrot et al. 2005), Colombia (Yunis et al. 2005), Ecuador (González-Andrade y Sánchez 2004), Brasil (da Silva et al. 2004), y Argentina (Sala et al. 1998) e incluso en algunas poblaciones del noreste asiático de China (Chen y Li 2004; Wang et al. 2003) y Japón (Hashiyada et al. 2003). Sin embargo, también se observaron diferencias en el locus TPOX, con las frecuencias más altas para los alelos *11 (0,48) y *6 (0,54) para Ngöbe y Emberá, respectivamente. Es interesante observar que estos alelos están presentes en una población y ausente en la otra, lo que sugiere que podrían ser distintivos para cada población. Además, el alelo TPOX más frecuente en todas las demás poblaciones fue *8 (Tabla 6). El alelo TPOX*11 mostró una frecuencia relativamente alta (aproximadamente 0,32) en afroamericanos, hispanoamericanos y caucásicos americanos (datos de Promega Corp.), Tu y Sala (dos poblaciones chinas; Chen y Li 2004), japoneses (Hashiyada et al. 2003) y en amerindios Terena de Brasil (da Silva et al. 2004). Además, el alelo TPOX*6 presentó frecuencias bajas en estas mismas poblaciones.

En el locus TH01, el alelo predominante en Ngöbe fue *6 (está también presente en dos poblaciones de amerindios del sur), mientras que en los Emberá fue *8; dicho sea de paso, los Emberá son la única población con mayor frecuencia de este alelo (ver Tabla 6). Es importante señalar que el alelo TH01 *6 en Ngöbe mostró una frecuencia notablemente mayor

(0,86) con respecto a las otras poblaciones. El alelo TH01*8 tuvo frecuencias bajas en los hispanoamericanos (0,01) y en las poblaciones Tu (0,09) y Sala (0,03), pero fue el alelo más frecuente en amerindios Terena (0,48).

También se observaron otros alelos exclusivos en Ngöbe y Emberá, aunque estos ocurrieron en bajas frecuencias (ver Tabla 1).

En todas las poblaciones amerindias y asiáticas el alelo *11 fue el alelo predominante en el locus D7S820, mientras que en las poblaciones de referencia el alelo predominante fué *10. Aunque faltan algunos datos, en el locus FESFPS todas las poblaciones el alelo predominante fue el mismo (*11). El alelo CSFIPO*12 fue el más común en los Ngöbe (0,46) y en los Emberá (0,43) y también en todas las poblaciones asiáticas, estadounidenses caucásicos, hispanoamericanos y la mayoría de los amerindios del sur, a diferencia de los amerindios del norte en donde el alelo con mayor frecuencia para este locus fue *13. En el locus F13A01 el alelo *3.2 fue el más común en los Ngobé (0,48), Emberá (0,38) y tehuelche (0,35), a diferencia de los mapuches y wichí, que mostraron frecuencias más altas para los alelos *7 (0,34) y *5 (0,35), respectivamente.

En el locus VWA, el alelo con mayor frecuencia fue *16 en todas las poblaciones amerindias como también en los sala y los afroamericanos e hispanoamericanos, aunque en proporciones menores. Por otro lado, en el locus D16S539, las frecuencias más altas fueron para el alelo *10 con el mismo valor (0,37) para Ngöbe y Emberá, y con una frecuencia de 0.29 para los Kichwa y Otomí del Valle de Ixmiquilpan. Sin embargo, en Terena la frecuencia más alta fue la del alelo *11 con 0,33; similar a la de los afroamericanos (0,32), los caucásicos (0,32) y los hispanos estadounidenses (0,30). Finalmente, en el locus D13S317 las frecuencias más altas correspondieron al alelo *9 en Ngöbe (0,56), Emberá (0,33), kichwa (0,24) y amerindios del norte (aproximadamente 0,33) mientras que el alelo *12 en este locus fue el de mayor frecuencia en terena (0,33), afroamericanos (0,41) y Americanos caucásicos (0,31). En la población Tu, el alelo más común en el El locus D13S317 fué *11 (0,24) y los alelos *8 y *11 fueron los de mayor frecuencia (0,25) en la población Sala.

El análisis de parámetros forenses y de paternidad para estos nueve loci STR demuestra que, a pesar de sus bajos niveles de polimorfismos genéticos, la mayoría de ellos podría ser informativos para fines forenses, mostrando un poder combinado de discriminación de 0,9999 para ambas poblaciones amerindias. Sin embargo, los poderes de exclusión en Ngöbe fue muy baja, particularmente en los loci TH01 y FESFPS, los cuales

presentaron valores de 0,04 y 0,08, respectivamente. Los poderes combinados de exclusión fueron 0,9338 y 0,9890 en los Ngöbe y los Emberá, respectivamente.

En contraste, los poderes combinados de exclusión en los afroamericanos, los caucásicos americanos y los hispanoamericanos fueron 0,9993, 0,9985 y 0,9986, respectivamente, lo que demuestra que el uso de loci adicionales que complementen la batería de nueve loci STR se recomienda para pruebas de paternidad y propósitos forenses, particularmente en la población Ngöbe. Recomendaciones similares con relación al uso de loci adicionales fueron sugeridas por Sun et al. (2003) en poblaciones aisladas. Además, un valor aceptable para el índice típico de paternidad combinada es de aproximadamente 90; sin embargo, para los Ngöbe fue considerablemente menor (2.58) y para los Emberá fue 40.44, que también es bastante bajo.

En resumen, a pesar de tener orígenes diferentes (los Ngöbe pertenecientes a la familia lingüística chibcha y los emberá a la chocona), ambas poblaciones amerindias conservan algunas similitudes en las distribuciones de las frecuencias alélicas; sin embargo, existen diferencias debido a la presencia de variaciones específicas, como también a frecuencias alélicas inusualmente altas en algunos loci polimórficos, como fue informado previamente por Barrantes et al. (1990) utilizando marcadores clásicos. Estas similitudes también son compartidas con otras poblaciones amerindias y con algunas poblaciones del noreste de Asia que pertenecen a grupos que son similares a los amerindios; estos grupos asiáticos están filogenéticamente relacionados y comparten una ascendencia común. De hecho, la gran cantidad de alelos compartidos, los bajos valores promedios del número de alelos (N_a) y el número efectivo de alelos (N_e) (ver Tablas 1, 2 y 4) y las deficiencias de heterocigotos son consistentes con los bajos niveles de diversidad genética en estas poblaciones amerindias, que vivieron durante muchos años en aislamiento del resto de la sociedad (Young 1971). Este fenómeno fue confirmado estimando por separado el índice de información de Shannon (I) en los Ngöbe y Emberá; valores combinados de 1,06 y 1,31, respectivamente fueron observados. Además, en la población Ngöbe tres loci (TPOX, VWA y F13A01) y en Emberá dos loci (77/07 y TPOX) que se desviaron del equilibrio de Hardy-Weinberg, lo que probablemente fue causado por la deficiencia de heterocigotos.

Los bajos niveles de diversidad también se muestran en la heterocigosidad promedio que fue del 65% en los Emberá y del 52,2% en los Ngöbe. Esto también está estrechamente relacionado a sus altos niveles de ho-

heterocigosidad, especialmente en los loci TH01 (75,8%) y FESFPS (64,6%) en los Ngöbe y el locus TPOX (50,5%) en los Emberá. El nivel promedio de heterocigosidad observado en los Emberá fue similar al de los Tarahumaras (66.8%), una población amerindia mexicana (Rangel-Villalobos et al. 2000), y los Wichi (63,5 %), población amerindia argentina (Sala et al. 1998). Kolman et al. (1995) también informaron niveles bajos similares de diversidad en los Ngöbe usando ADN mitocondrial, y propusieron que esta población había sufrido recientemente un cuello de botella demográfico. De hecho, la evidencia etnohistórica apoya la hipótesis de que las poblaciones de Ngöbe sufrieron una drástica reducción seguida de un aumento en el tamaño de la población con la consiguiente pérdida de diversidad genética o variabilidad (Jorge et al. 1993).

Estas características son típicas de la mayoría de las poblaciones amerindias, que son consideradas poblaciones aisladas pequeñas que han sufrido cuellos de botella recientes (Sun et al. 2003). Los registros históricos sugieren que las enfermedades y los exterminios masivos puede haber causado un colapso demográfico durante la conquista europea (Kolman et al. 1995; Jorge et al. 1993). Una interpretación alternativa postula que los cuellos de botella se originaron a partir de la etnogénesis de pequeñas poblaciones fundadoras del Paleo-amerindios y rápidas fragmentaciones in situ en grupos tribales separados con baja niveles de mezcla genética entre ellos (Kolman et al. 1995). Estos procesos promueven la rápida fijación o extinción de genes o alelos en presencia de pequeños tamaños efectivos de población, como es el caso de los amerindios (Barrantes 1990).

Las diferencias en la variabilidad genética entre las dos poblaciones podrían deberse a cualquiera de estos procesos microevolutivos con efectos negativos más evidentes entre los amerindios Ngöbe.

Literatura citada

1. Arias, T. D., E. A. Castro, E. Ruiz et al. 2002. La mezcla racial de la población panameña. *Rev. Med. Panama* 17:5—17.
2. Bar, W., B. Brinkmann, B. Budowle et al. 1997. DNA recommendations: Further report of the DNA Commission of the ISFH regarding the use of short tandem repeat systems. *Int. J. Legal Med.* 110(4): 175-176.
3. Barrantes, R. 1990. An evolutionary hypothesis on the absence of the Diego antigen (Di-a) in Chibcha Amerindians. *Rev. Biol. Trop.* 38(2A):277-282.
4. Barrantes, R., R E. Smouse, H. W. Mohrenweiser et al. 1990. Microevolution in lower Central America: Genetic characterization of the Chibcha-speaking groups of Costa Rica and Panama, and a consensus taxonomy based on genetic and linguistic affinity. *Am. J. Hum. Genet.* 46(1):63-84.
5. Barrot, C, C. Sanchez, M. Ortega et al. 2005. Characterization of three Amerindian populations from Hidalgo State (Mexico) by 15 STR-PCR polymorphisms. *Int. J. Legal Med.* 119(2):111-115.
6. Batista, O., C. J. Kolman, and E. Bermingham. 1995. Mitochondrial DNA diversity in the Kuna Amerinds of Panama. *Hum. Mol. Genet.* 4(5):921-929.
7. Cerda-Flores, R. M., B. Budowle, L. Jin et al. 2002a. Maximum-likelihood estimates of admixture in northeastern Mexico using 13 short tandem repeat loci. *Am. J. Hum. Biol.* 14:429-439.
8. Cerda-Flores, R. M., M. C. Villalobos-Torres, H. A. Barrera-Saldana et al. 2002b. Genetic admixture in three Mexican Mestizo populations based on *D1S80* and *HLA-DQA1* loci. *Am. J. Hum. Biol.* 14:257-263.
9. Chen, T., and S. Li. 2004. Genetic polymorphism at nine STR loci in two ethnic groups of China. *Forensic Sci. Int.* 146(1):53-55.
10. da Silva, D. A., A. C. S. Goes, and E. F. Carvalho. 2004. Eight short tandem repeats (STR) frequencies for descendants from Terena indigenous Brazilian group. *Forensic Sci.* 49(2):398-399.
11. Dirección de Estadística y Censo, Contraloría General de la República de Panamá. 2000. *Censos Nacionales de Población y Vivienda: Panamá en Cifras*. Available at <http://www.contraloria.gob.pa/DEC/cuadros.aspx?ID=1703>
12. DNA Commission, International Society for Forensic Hemogenetics. 1994. Report concerning further recommendations of the DNA Commission of the ISFH regarding PCR-based polymorphisms in STR (short tandem repeat) system. *Int. J. Legal Med.* 107(3): 156-160.
13. Gonzalez-Andrade, R, and D. Sanchez. 2004. Genetic profile of the Kichwas (Quichuas) from Ecuador by analysis of STR loci. *Hum. Biol.* 76(5):723-730.
14. Hartl, D. L., and A. G. Clark. 1989. *Principles of Population Genetics*, 2nd ed. Sunderland, MA: Sinauer Associates.

15. Hashiyada, M., Y. Itakura, T. Nagashima et al. 2003. Polymorphism of 17 STRs by multiplex analysis in Japanese population. *Forensic Sci. Int.* 133:250-253.
16. Hedrick, P. W. 1985. *Genetics of Populations*. Boston: Jones and Barttlet.
17. Jorge, L. F, T. D. Arias, U. Griese et al. 1993. Evolutionary pharmacogenetics of *CYP2D6* in Ngawbe Guaymi of Panama: Allele-specific PCR detection of the *CYP2D6B* allele and RFLP analysis. *Pharmacogenetics* (3):231-238.
18. Jorge, L. E, M. Eichelbaum, E. U. Griese et al. 1999. Comparative evolutionary pharmacogenetics of *CYP2D6* in Ngawbe and Embera Amerindians of Panama and Colombia: Role of selection versus drift in world populations. *Pharmacogenetics* 9(2):217-228.
19. Jorge-Nebert, L. R, M. Eichelbaum, E. U. Griese et al. 2002. Analysis of six SNPs of *NAT2* in Ngawbe and Embera Amerindians of Panama and determination of the Embera acetylation phenotype using caffeine. *Pharmacogenetics* 12(1):39-48.
20. Kolman, C. J., and E. Bermingham. 1997. Mitochondrial and nuclear DNA diversity in the Choco and Chibcha Amerinds of Panama. *Genetics* 147(3):1289-1302.
21. Kolman, C. J., E. Bermingham, R. Cooke et al. 1995. Reduced mtDNA diversity in the Ngobe Amerinds of Panama. *Genetics* 140(1):275-283.
22. Levene, H. 1949. On a matching problem in genetics. *Ann. Math. Stat.* 20:91-94.
23. Mastana, S., and P. P. Singh. 2002. Population genetic study of the STR loci (*HUMCS-F1PO*, *HUMTPOX*, *HUMTHoI*, *HUMLPL*, *HUMF13A01*, *HUMF13B*, *HSFESFPS*, and *HUMVWA*) in North Indians. *Ann. Hum. Biol.* 29(6):677-684.
24. Miller, S. A., o.o. Dykes, and H. F. Polesky. 1988. A simple salting-out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* 16:1215.
25. Neel, J. V., R. J. Biggar, and R. I. Sukernik. 1994. Virologic and genetic studies relate Amerind origins to the indigenous people of the Mongolia/Manchuria /southeastern Siberia region. *Proc. Natl. Acad. Sci. USA* 91(22):10,737-10,741.
26. Rangel-Villalobos, H., F. Rivas, L. Sandoval et al. 2000. Genetic variation among four Mexican populations (Huichol, Purepecha, Tarahumara, and Mestizo) revealed by two VNTRs and four STRs. *#wm. Aof.* 72(6):983-995.
27. Sala, A., G. Penacino, and o. Corach. 1998. Comparison of allele frequencies of eight STR loci from Argentinian Amerindian and European populations. *Hum. Biol.* 70(5):937-947.
28. Salzano, R M., and S. M. Callegari-Jacques. 1988. *South American Indians: A Case Study in Evolution*. Oxford, UK: Clarendon, 186-187.
29. Shannon, C. E., and W. Weaver. 1949. *The Mathematical Theory of Communication*. Urbana: University of Illinois Press.
30. Siegel, S., and N. J. Castellan. 1998. *Nonparametric Statistics for the Behavioral Sciences*. New York: McGraw-Hill.

31. Sun, G., S. T. McGarvey, R. Bayoumi et al. 2003. Global genetic variation at nine short tandem repeat loci and implications on forensic genetics. *Eur. J. Hum. Genet.* 11(1):39—49.
32. Wang, W., H. Jia, Q. Wang et al. 2003. STR polymorphisms of forensic loci in the northern Han Chinese population. *J. Hum. Genet.* 48:337-341.
33. Young, R D. 1911. *Ngawbe: The Tradition and Change Among the Western Guaymi of Panama*. Illinois Studies in Anthropology, no. 7. Urbana: University of Illinois Press.
34. Yunis, J. J., O. Garcia, I. Uriate et al. 2000. Population data on 6 short tandem repeat loci in a sample of Caucasian-Mestizos from Colombia. *Int. J. Legal Med.* 113(3):175—178.
35. Yunis, J. J., O. Garcia, and E. J. Yunis. 2005. Population frequencies for *CSF1PO*, *TPOX*, *TH01*, *F13A01*, *FES/FPS*, and *VWA* in seven Amerindian populations from Colombia. *J. Forensic Sci.* 50(3):682-684.

C A P Í T U L O 3

Análisis de 30 marcadores
INDELs polimórficos en
la población panameña:
estimaciones de mezcla
racial, estructura
poblacional y parámetros
forenses

*Carlos Ramos Delgado, Edgardo Castro-Pérez, Concepción
Molina-Jirón C. y Diomedes E. Trejos.*

Resumen

Los loci de repeticiones cortas en tándem corto (STR) se han empleado con éxito en estudios de genética forense, paternidad y análisis antropológicos en las poblaciones mestizas y amerindias de Panamá desde hace varios años. Sin embargo, los informes indican que el uso de STRs podría estar limitado en casos de muestras de ADN degradadas debido al tamaño del producto de PCR y en paternidad debido a las tasas de mutación relativamente altas. Por lo tanto, como complemento de los STR, se han desarrollado marcadores con mayor eficiencia de amplificación, como los de inserción/delección (INDEL). Sin embargo, no se tiene información sobre los polimorfismos genéticos y la distribución de los INDELS en la población panameña. Usando el kit Investigator® DIPplex (Qiagen), reportamos por primera vez el perfil genético de 30 marcadores INDEL en la población panameña. Las estimaciones de la mezcla genética y estructura poblacional demuestran que la población panameña esta mezclada diferente y es altamente polimórfica. Curiosamente, las estimaciones de mezclas fueron muy similares a nuestros estudios anteriores utilizando STRs en los que se reportan contribuciones genéticas amerindias, africanas y europeas. Los INDELS mostraron tres grupos de genes en estas proporciones: 0,46 (grupo 1), 0,24 (grupo 2) y 0,30 (grupo 3) versus 0,51 (amerindios), 0,24 (africanos) y 0,25 (europeos) para STRs. También encontramos que tanto los INDELS como los STRs indicaron que la distribución de genes ancestrales difiere entre provincias. Además, la variación de la frecuencia alélica demostró que todos los loci cumplieron con las expectativas de Hardy-Weinberg y muestran una alta diversidad como lo indica el promedio de heterocigosidad observada (0,4477), heterocigosidad esperada (0,4553) y alta similitud con poblaciones de referencia de Estados Unidos y del mundo. Además, los parámetros estadísticos forenses mostraron una gran importancia en el análisis combinado, poder de discriminación (0,9999999999987) y la probabilidad de coincidencia combinada ($1,27 \times 10^{-12}$), pero relativamente menor significancia en el poder combinado de exclusión (0,992655332). Los parámetros forenses calculados indican que los marcadores INDELS pueden ser efectivamente utilizados en la población panameña para usos forenses, pero deben ser complementados con marcadores STRs adicionales en pruebas de paternidad.

Palabras claves: inserción/delección; marcadores INDEL; repeticiones cortas en tándem; Loci STR; mezcla de genes ancestrales; parámetros forenses

Abstract

For several years, short tandem repeat (STR) loci have been successfully employed for forensic genetic, paternity, and anthropological analyses in the admixed and Amerindian populations of Panamanians. Nevertheless, reports indicate that the use of STRs might be limited in cases involving degraded DNA samples due to their PCR amplicon size and in paternity because of their relatively high mutation rates. Therefore, as a complement to STRs, markers with higher PCR efficiency, such as insertion/deletion (INDEL), have been developed. However, the genetic polymorphisms and distribution of INDELS in the Panamanian population were unknown. Using the Investigator® DIPplex kit (Qiagen), we report here for the first time the genetic profile of 30 INDEL markers in the Panamanian population. Gene admixture estimates and population structure indicated that the Panamanian population is differentially admixed and highly polymorphic. Interestingly, admixture estimates were similar to our previous STR report, indicating Amerindian, African, and European gene contributions. INDELS showed three gene clusters in these proportions: 0.46 (cluster 1), 0.24 (cluster 2), and 0.30 (cluster 3) versus 0.51 (Amerindian), 0.24 (African), and 0.25 (European) for STRs. We also found that both INDELS and STRs indicated that ancestral gene distribution is heterogeneous among the provinces across the country. Furthermore, allele frequency variation showed that all loci accomplished Hardy-Weinberg expectations and displayed high diversity as indicated by the average of observed heterozygosity (0.4477), expected heterozygosity (0.4553), and high similarity with reference U.S. and world populations. Additionally, forensic statistic parameters showed a strong significance in the combined power of discrimination (0.9999999999987) and the combined matching probability (1.27×10^{-12}) but relatively lower importance in the combined power of exclusion (0.992655332). The calculated forensic parameters indicate that the Panamanian population can effectively use INDEL markers for forensic purposes. Still, they should be complemented with additional markers, such as STRs, for paternity analyses.

Keywords: insertion/deletion; INDEL markers; short tandem repeats; Loci STR; ancestral gene admixture; forensic parameters

Introducción

El genotipaje con repeticiones cortas en tándem (STR) es uno de los marcadores moleculares más ampliamente utilizados en la identificación forense en humanos en todo el mundo (Shewale y Liu, 2013; Wang et al. 2014). Los STRs se han empleado con éxito para análisis forenses y de paternidad en la población panameña por más de 15 años. Además, hemos utilizado marcadores STRs para caracterizar a las poblaciones amerindias Ngöbe y Emberá de Panamá (también ubicados en Costa Rica y Colombia, respectivamente) (Castro et al. 2007). Recientemente, determinamos la ascendencia y la estructura genética trihíbrida con estimaciones de mestizaje de la población de Panamá utilizando loci STR [4]. Estos marcadores han mostrado un alto polimorfismo y una adecuada variación genética para análisis forenses y de paternidad en todas las provincias del país. Sin embargo, varios reportes indican que el uso de STRs podría en principio estar limitado por el tamaño del producto de PCR y por las tasas de mutación relativamente altas (Wang et al. 2014; Torres et al. 2014; TZ et al 2015).

En particular las muestras incluidas en parafina, ADN antiguo y ADN degradado, todas ellas comúnmente encontradas en casos forenses son sustratos difíciles para la amplificación de STRs (Martínez-Cortés, 2016; Pereira et al. 2012). Por lo tanto, como complemento a los STRs, se han desarrollado marcadores adicionales como los polimorfismos de un solo nucleótido (SNP) y los de inserción/ deleción (INDEL) con una mayor eficiencia para amplificar muestras de ADN degradadas (Weber et al. 2002; da costa et al. 2012). Desafortunadamente, el genotipaje basado en sistemas SNPs requiere de una instrumentación más sofisticada y una química de procesamiento de las muestras más laboriosa que lo hace incompatible con la tecnología de genotipado de STRs actualmente disponible en la mayoría de los laboratorios forenses [TZ et al. 2015; LaRue et al 2012]. En cambio, los INDELS combinan las mejores características de los SNPs y de los STRs, ya que generan productos de PCR mucho más pequeños que los STRs, tasas de mutación de loci binarios más bajas, pero con una plataforma de genotipado directa mediante electroforesis capilar como con los STRs (TZ et al. 2015; Martínez-Cortés; LaRue et al 2012). Por eso, se ha propuesto la implementación de los marcadores INDEL como complemento para análisis forense en la población panameña junto con los STRs. Sin embargo, los polimorfismos alélicos y la distribución de la variación genética de los INDELS en el país no se conoce.

El kit Investigator® DIPplex (Qiagen) con un panel 30 marcadores INDELS bialélicos más amelogenina ha demostrado tener valor en aplicación forense en varias poblaciones a nivel mundial (Fondevilla 2014; Investigator Diplex 2014). En este estudio se analizaron 332 individuos no relacionados de diferentes provincias del país con el Kit del panel Investigator® DIPplex. Es importante destacar que nuestro trabajo anterior utilizando STRs demostró que la población panameña muestra una estructura trihíbrida. La mezcla trihíbrida mestiza esta compuesta de genes amerindios, africanos y europeos (Castro-Pérez et al. 2016). Además, varios estudios han demostrado la utilidad de marcadores INDELS en estimaciones de mezcla de genes en múltiples poblaciones (Pereira et al. 2012; Santos et al. 2010; Resques et al. 2010; Saloum et al. 2013)

Por lo tanto, para ampliar aún más nuestros estudios de mezcla con STRs, determinamos la estructura genética de la población mestiza de Panamá utilizando este otro tipo de marcador como también la proporción de la mezcla genética. Calculamos las frecuencias alélicas y los parámetros de diversidad de cada locus y parámetros estadísticos de importancia forense y de paternidad. Los resultados mostraron que la estructura genética y las estimaciones de mezcla utilizando marcadores INDELS son consistentes con los reportes anteriores utilizando loci STRs. Además, los parámetros estadísticos forenses mostraron una alta significancia en el poder de discriminación combinado ($0,9999999999987$) y en la probabilidad de coincidencia combinado ($1,27 \times 10^{-12}$), pero menor en el poder de exclusión combinado ($0,992655332$). Los datos generados indican que los marcadores INDELS podrían ser aplicables a la población panameña en la mayoría de los casos forenses, pero deben complementarse con marcadores adicionales como STRs para pruebas de paternidad y parentesco.

Materiales y métodos

Selección de mestizos, recolección de sangre y extracción de ADN.

Seleccionamos 332 individuos (50% varones y 50% mujeres) no emparentados de origen mestizo, con padres y abuelos nacidos en Panamá. El número de individuos por provincia fue determinado proporcionalmente según el porcentaje de contribución de cada provincia a la población total del país según lo informado por las estadísticas nacionales del gobierno Figura 1. Se escogió este modelo porque tiene la ventaja de ser una buena representación de la población total del país. La desventaja de este modelo es que con base al cálculo regional de estadísticas de población, algunas de las provincias tendrían un número relativamente bajo de muestras. Sin

embargo, la recolección aleatoria de ADN de individuos no relacionados y de marcadores no sesgados podría ser suficiente para obtener una visión general de los genes en cada provincia. Las muestras se recolectaron principalmente en áreas urbanas donde la mayoría de las poblaciones mestizas (mezcladas) están concentradas. Por el contrario, áreas con bajas cantidades de población mestiza, como la amerindia, conocidas como Comarcas, no fueron consideradas en el muestreo.

Se obtuvieron muestras de sangre venosa (10 ml) mediante flebotomía y se recogieron en tubos que contenían tampón anticoagulante EDTA. Los tubos fueron suavemente mezclados para asegurar que el EDTA y la sangre estuvieran adecuadamente mezclados. Después de la extracción, las muestras de sangre se procesaron inmediatamente o fueron almacenados a 4°C, por un máximo de 1-4 días. Aproximadamente 125 µL de sangre fueron colocadas en el centro de cada uno de los cuatro círculos de Tarjetas FTA® preparadas comercialmente (tecnología de tarjetas Whatman® FTA®).

Se utilizó un abre hueco para cortar un pequeño trozo circular con sangre de la tarjeta FTA. El pequeño trozo fue procesado para la extracción de ADN de la muestra según el procedimiento de DNA IQ™ Kit para Maxwell® 16 (Promega® Corporation, Madison WI, EE.UU.). Las muestras de ADN fueron eluidas con 300 µL de buffer de elución. La concentración de ADN fue determinada y las muestras utilizadas para la amplificación por PCR.



Figura 1. Mapa de las provincias de Panamá con los sitios de recolección de las muestras. El número de muestras colectadas por provincia que representa el porcentaje de contribución a la población del país es Bocas del Toro n = 9 (2,8%) Coclé n = 23 (7%), Colón n = 25 (7,4%), Chiriquí n = 48 (14%), Darién n = 5 (1,4%), Herrera n = 14 (4%), Los Santos n = 12 (3,4%), Panamá (incluye Panamá y Panamá Oeste) n = 174 (53%), Veraguas n = 24 (7%).

Amplificación por PCR y genotipado de alelos de Inserción-Delección (INDELs)

La amplificación de 30 marcadores INDELs se realizó mediante PCR multiplex en un termociclador GeneAmp® PCR System 9700 (Applied Biosystems™) según el manual descrito en el kit Investigator® DIPplex (Qiagen). Los alelos presentes en cada genotipo fueron determinados mediante electroforesis capilar, utilizando un secuenciador ABI PRISM 3500 y de acuerdo al protocolo establecido por el kit Investigator® DIPplex (Qiagen). Los tamaños de los alelos se identificaron mediante comparación del patrón de migración del estándar de tamaño ADN BTO 550 (Qiagen®) y el Software de recopilación de datos versión 1.0 y software GeneMapper® ID-X versión 1.2 (Thermo Fisher Scientific Inc.) como indica el Fabricante del kit Investigator® DIPplex.

Estimaciones de mezcla racial y parámetros forenses poblacionales

Se utilizó el software STRUCTURE (Falush et al. 2000, Pritchard y Stephens 2000) versión 2.3.4 para estimar las proporciones de mezcla y la estructura de la población utilizando 5000 períodos de preparación y 50 000 repeticiones MCMC después de la preparación. El número de poblaciones asumidas fue de uno a ocho y el log de la probabilidad de cada K se calculó utilizando el software Structure Harvester (2012) que implementa el método de Evanno (Earl y von Holdt, 2012). Las frecuencias alélicas de inserciones (Dip+), de deleciones (Dip-) y los parámetros poblacionales se calcularon utilizando PopGene (Yeh et al. 1997), versión 1.32. La heterocigosidad observada y esperada se determinó utilizando el algoritmo de Nei (Nei 1978) y la prueba de razón de verosimilitud (G-cuadrado) se realizó para estimar las expectativas de Hardy-Weinberg por locus. Se utilizó la plantilla MS Excel de trabajo de PowerStatsV12 (Promega Corporation, Madison WI, EE. UU.) para calcular los parámetros forenses y de paternidad, probabilidad de emparejamiento (MP), poder de discriminación (PD), poder de exclusión (PE), índice de paternidad típico (IPT) y contenido de información polimórfica.

Resultados

Estimaciones de la mezcla genética

Después de múltiples permutaciones bayesianas de la estructura, los análisis mostraron que los K con la mayor probabilidad logarítmica $L(K)$ estimada fueron K_2 (11 ensayos), K_3 (12 ensayos) y K_4 (9 ensayos) correspondientes a probabilidades promedio de -1287.01 , -12944.04 y -13013.91 , respectivamente (Figura 2).

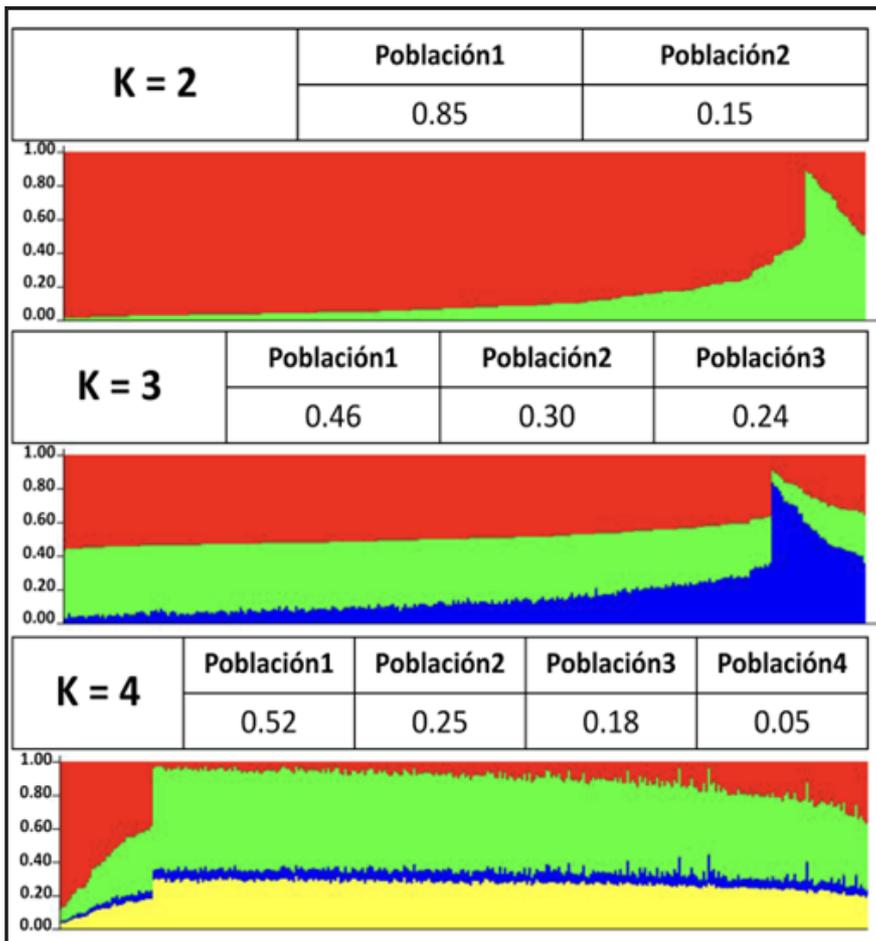


Figura 2. Estimaciones de mezcla ancestral para el país (total) y proporciones de cada conglomerado de genes en la población con colores predefinidos, según las permutaciones realizadas con STRUCTURE. El “STRUCTURE Harvester” calculó las probabilidades L_n más altas graficadas $L(K)$ con valores para K entre 2 y 4 ($K = 2 - 4$).

K2 presentó proporciones de mezcla de 0,85 y 0,15; las proporciones de mezcla de K3 fueron 0,46, 0,24 y 0,30; mientras que K4 presentó proporciones de mezcla de 0,52, 0,25, 0,18 y 0,05. Desafortunadamente, no existen datos disponibles de INDELs de poblaciones ancestrales; por lo tanto, para evaluar estos resultados comparamos las proporciones de mezcla generadas con INDELs con nuestro trabajo anterior con marcadores STRs (Castro-Pérez, 2016).

Tabla 1. Estimaciones de mezcla de genes por provincia, el país total y comparación de marcadores INDEL con loci STR.

PROVINCIA	INDEL ^a			STR ^b		
	Grupo 1	Grupo 2	Grupo 3	Amerindio	Africa no	Europeo
<i>El país (total)</i>	0.46	0.24	0.30	0.51	0.24	0.25
<i>Bocas del Toro</i>	0.18	0.14	0.69	n.d	n.d	n.d
<i>Coclé</i>	0.51	0.26	0.23	0.70	0.10	0.20
<i>Colón</i>	0.30	0.54	0.16	0.34	0.47	0.19
<i>Chiriquí</i>	0.61	0.19	0.20	0.64	0.15	0.21
<i>Darién</i>	0.41	0.35	0.24	n.d	n.d	n.d
<i>Herrera</i>	0.52	0.37	0.11	0.40	0.20	0.40
<i>Los Santos</i>	0.48	0.41	0.11	0.33	0.25	0.42
<i>Panamá (2provincias)^c</i>	0.61	0.24	0.14	0.50	0.27	0.23
<i>Veraguas</i>	0.59	0.16	0.25	0.50	0.11	0.39

^aPresent Study; ^bCastro-Perez et al. 2016; n.d = not determined; ^cIncluye las provincias de Panamá y Panamá Oeste.

Es importante destacar que nuestros estudios con STRs implicaron el uso de datos ancestrales de Amerindios (Ngobe-Chibchan), africanos occidentales (Angola y Guiné-Bissau) y europeos (España) utilizando un modelo trihíbrido (K3). Además, el modelo de mezcla trihíbrida es altamente consistente con datos etnohistóricos [Jaén-Suárez 1978; Jaén-Suárez, 1998] y de marcadores genéticos de ADN clásico (Arias et al. 2002), mitocondrial (Perego 20122015) y cromosoma Y (Grugni et al. 2015). Por lo tanto, comparamos las proporciones de mezcla K3 por provincia y en todo el país generadas con INDELs con los datos STRs de nuestro estudio anterior. Curiosamente, observamos que las estimaciones de mezcla con ambos marcadores; INDELs y STRs (Tabla 1) fueron muy similares para el país. Con INDELs 0,46 (grupo 1), 0,24 (grupo 2) y 0,30 (grupo 3) ver-

sus 0,51 (amerindios), 0,24 (africanos) y 0,25 (europeos) con STRs. Además, las proporciones de mezcla entre las provincias con INDELS también mostraron patrones generalmente similares, pero con algunas variaciones en comparación con los STRs (Tabla 1). Estos datos indican que en la población panameña tanto los INDELS como los STRs muestran proporciones de mezcla de genes comparables.

Diversidad genética y estructura poblacional

La proporción de mezcla entre las provincias utilizando como marcadores los INDELS mostraron que la población panameña está mezclada heterogéneamente y es altamente polimórfica (Figura 3). Similar a los resultados obtenidos con loci STRs, el análisis con INDELS generó una distribución diversa de genes en todas las provincias que se asemejan a los genes ancestrales amerindios, africanos y europeos (Tabla 1). El gráfico de triángulo (Figura 3) representa el conjunto de genes de la población panameña general que muestra tres grupos de genes que se comportan de manera similar a las poblaciones parentales de STRs: amerindios (grupo 1), africano (grupo 2) y europeo (grupo 3). Los puntos dentro del triángulo representan los individuos de cada provincia, quienes estaban agrupados y distribuidos espacialmente según su afinidad de mezcla individual con cada población ancestral. La estructura del gráfico de barras (Figura 3, abajo) muestra las proporciones diferenciales de genes ancestrales por individuos y demuestra la variación genética polimórfica entre provincias. Los INDELS demostraron que las provincias de Chiriquí y Panamá muestran la mayor proporción de genes de tipo amerindio (grupo 1), ambos con 0.61, seguido de Veraguas (0.59), Herrera (0,52) y Coclé (0,51). Los STRs también mostraron las mismas provincias con la mayor mezcla de genes amerindios, pero en diferentes proporciones (Tabla 1).

Además, la mezcla de tipo africano (grupo 2) determinada por los INDELS mostraron que, como se esperaba, la provincia de Colón presenta la mayor ascendencia genética africana (0,54), seguida de los Santos (0,41) y Herrera (0,37). Los loci STRs también mostraron que la mayor proporción de genes africanos se observó en Colón (0,47), pero con alguna variación en las provincias que a continua-

ción se enlistan. En los estudios con STRs, Colón es seguida por la provincia de Panamá con 0,27, Los Santos (0,25) y Herrera (0,20). Los INDELS también indicaron que los genes de tipo europeo (grupo 3) muestran una contribución relativamente menor al acervo genético en todas las provincias, que van desde 0,11 para Herrera y Los Santos, a 0.25 para Veraguas. Una excepción sobre la contribución de genes europeos fue Bocas, mostrando un inusual 0,69. La comparación con los STRs también indica niveles relativamente más bajos de contribución de genes europeos entre las provincias que van desde 0,19 para Colón hasta 0,42 para Los Santos. En conjunto, estos datos demuestran que tanto los INDELS como los STRs presentan patrones de mezcla y estructura genética similares.

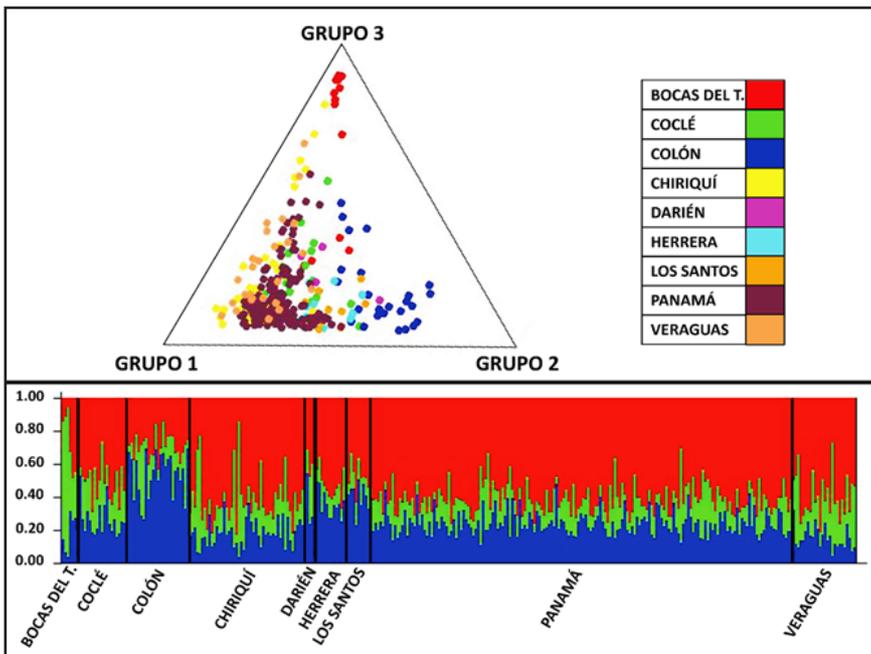


Figura 3. Estimaciones de mezcla genética ancestral calculadas utilizando STRUCTURE basado en 30 marcadores INDEL polimórficos. Arriba: Gráfico de triángulo para K=3 que muestra el análisis de agrupamiento con estimaciones de ascendencia. Cada punto dentro del triángulo representa a personas individuales con sus correspondientes códigos de colores de las provincias donde fue recolectada la muestra de ADN. Se observa una amplia variación en todas las provincias con una alta contribución de mezcla de población ancestral en el grupo 1. Basado en estudios genéticos previos (Arias et al., 2002; Castro-Pérez et al., 2016), esto sugiere que el grupo 1 correspondería a genes de origen indígena. Abajo: Gráfico de barras que muestra la estructura polimórfica en todo el país por provincia. Las barras verticales representan personas únicas y sus correspondientes mezclas genéticas ancestrales individuales.

Variación de las frecuencias alélicas

Las frecuencias alélicas de 30 INDELS; inserciones (Dip+) y deleciones (Dip-) de la población mestiza total de Panamá se muestran en Tabla 2. Las frecuencias de inserciones variaron entre 0,2024 (locus HLD58) y 0,8030 (locus HLD64) con un promedio de 0,48349; mientras que las frecuencias para las deleciones oscilaron entre 0,1970 (locus HLD64) y 0,7976 (locus HLD58) y 0,51651 como promedio. Los valores de heterocigosidad observados y esperados, y el equilibrio de Hardy-Weinberg también se muestran (Tabla 2). La heterocigosidad observada (H_o) osciló entre 0,2961 (locus HLD58) a 0,5287 (locus HLD136) con un promedio de 0,4477. La heterocigosidad esperada (H_e) mostró valores que oscilaban entre 0,3175 (locus HLD64) a 0,4997 (locus HLD131) con un promedio de 0,4553, que se acerca al promedio de H_o (0,4477). Sin embargo, cuando se compara la heterocigosidad observada con la esperada para cada marcador, encontramos que 5 loci (HLD 131, HLD 114, HLD 122, HLD 81 y HLD 128) muestran una H_o baja, en contraste con la H_e , que van desde 0,003 (locus HLD131) a 0,0487 (HLD128), lo que sugiere una deficiencia de heterocigosidad. Sin embargo, tras la corrección de Bonferroni ($p=0,0017$) todos los loci mostraron HWE (Carvalho 2015). Estos datos indican que los marcadores INDELS muestran en la población panameña una variación altamente polimórfica.

Tabla 2. Frecuencias alélicas y genotípicas de 30 marcadores INDELS en la población panameña.

LOCUS	ID	Alelo Dip-	Alelo Dip+	H_o	H_e	HWE
HLD77	rs1611048	0.6485	0.3515	0.4773	0.4571	0.3764
HLD45	rs2307959	0.3973	0.6027	0.4743	0.4789	0.8404
HLD131	rs1611001	0.4879	0.5121	0.4018	0.4997	0.003
HLD70	rs2307652	0.2613	0.7387	0.3776	0.3861	0.6707
HLD6	rs1610905	0.4517	0.5483	0.5106	0.4953	0.5944
HLD111	rs1305047	0.6843	0.3157	0.3958	0.4321	0.1195
HLD58	rs1610937	0.7976	0.2024	0.2961	0.3229	0.1233
HLD56	rs2308292	0.3833	0.6167	0.4592	0.4729	0.5407

Continúa en la próxima pag.

HLD118	rs16438	0.6888	0.3112	0.429	0.4287	0.9885
HLD92	rs17174476	0.6692	0.3308	0.4441	0.4428	0.9775
HLD93	rs2307570	0.4486	0.5514	0.4985	0.4947	0.9117
HLD99	rs2308163	0.4486	0.5514	0.4683	0.4947	0.3173
HLD88	rs8190570	0.4018	0.5982	0.5015	0.4807	0.4476
HLD101	rs2307433	0.4834	0.5166	0.5136	0.4994	0.6257
HLD67	rs1305056	0.4743	0.5257	0.5196	0.4987	0.4611
HLD83	rs2308072	0.6409	0.3591	0.4532	0.4605	0.7084
HLD114	rs2307581	0.5303	0.4697	0.4441	0.4982	0.0394
HLD48	rs28369942	0.2455	0.7545	0.3776	0.3712	0.8144
HLD124	rs6481	0.6576	0.3424	0.429	0.4497	0.4031
HLD122	rs8178524	0.6348	0.3652	0.4018	0.4647	0.0163
HLD125	rs16388	0.7379	0.2621	0.4169	0.3875	0.1923
HLD64	rs1610935	0.197	0.803	0.2991	0.3175	0.2534
HLD81	rs17879936	0.3167	0.6833	0.3807	0.4332	0.0217
HLD136	rs16363	0.4409	0.5591	0.5287	0.4927	0.1782
HLD133	rs2067235	0.6341	0.3659	0.4894	0.4639	0.3674
HLD97	rs17238892	0.5136	0.4864	0.4955	0.4996	0.8147
HLD40	rs2307956	0.5854	0.4146	0.4894	0.4852	0.9513
HLD128	rs2307924	0.596	0.404	0.429	0.4807	0.0487
HLD39	rs17878444	0.6037	0.3963	0.5045	0.4776	0.3093
HLD84	rs3081400	0.4343	0.5657	0.5257	0.4916	0.3067

Ho=Observed Heterozygosity, He=Expected Heterozygosity, HWE=Hardy-Weinberg Equilibrium.

Parámetros genéticos forenses de la población

Se muestran los valores calculados de los parámetros forenses y de paternidad con INDELS (todos los loci combinados) por provincia y país; y por loci para la población total (Tabla 3). Encontramos que el promedio de la probabilidad de coincidencia (MP) fue 0,4032, y oscila entre el marcador más informativo para este parámetro HLD64

con 0,519, y el menos informativo el locus HLD131 con 0,341. Sin embargo, la probabilidad de coincidencia combinada (CMP) para el país fue de $1,27 \times 10^{-12}$ y relativamente alta en la mayoría de las provincias. Entre las provincias, la CMP varió de 1.37219×10^{-07} (Darién) y 4.17645×10^{-09} (Bocas Del Toro) a valores relativamente altos de 9.17764×10^{-13} (Panamá), 8.34379×10^{-12} (Colón) y 4.56524×10^{-12} (Chiriquí) seguido por Los Santos (2.13145×10^{-11}), Herrera (2.10797×10^{-11}), Veraguas (1.76484×10^{-11}) y Coclé (1.69054×10^{-11}).

Además, el poder de discriminación (PD) mostró un promedio de 0,597, siendo el locus más informativo para este parámetro HLD131 con un valor de 0,659, y el menos informativo HLD64 con 0,481. Sin embargo, al igual que la CMP, el poder de discriminación combinado (CPD) también fue relativamente alto en el país y la mayoría de las provincias. En el país, el CPD fue 0,9999999999987, y en las provincias varió de 0,99999986 para Darién y 0,9999999999991 para la provincia de Panamá. Además, el poder de exclusión (PE) mostró un promedio de 0,15 y osciló entre el marcador más informativo para este parámetro (locus HLD136) con 0,215, a los loci menos informativos HLD58 y HLD64, ambos con 0,062.

En contraste con el CMP y el CPD, el poder de exclusión combinado (CPE) mostró una significancia relativamente menor a nivel de país (0,992655332) y provincias con un rango de 0,991455029 para la provincia de Panamá hasta 0,999564716 en la provincia de Herrera. Una excepción fue la provincia de Bocas del Toro que mostró un CPE de 1, lo que podría estar relacionado con el tamaño relativamente pequeño de la muestra, como se mencionó en los métodos. El índice típico de paternidad (TPI) por loci fue en promedio de 0,915 con un rango de 0,710 (locus HLD58) a 1,065 (locus HLD136). El contenido de información polimórfica (PIC) mostró un promedio de 0,3504 y varió de 0,266 (locus HLD64) a 0,375 (loci HLD131, HLD101 y HLD97). La mayoría de los valores PIC fueron cercanos a 0,375, excepto 5 loci (HLD64, HLD58, HLD48, HLD70 y HLD125) que mostraron valores inferiores al promedio. En general, las estadísticas combinadas de 30 INDEL para análisis forense y paternidad demuestran que la mayoría de los cálculos son altamente significativos para la aplicación forense de estos marcadores en la población panameña.

Tabla 3. Parámetros estadísticos combinados de genética forense y paternidad por provincia, y el total del país y por loci en el país.

PROVINCE	CMP	CPD	CPE		
Bocas del T.	4,18E-04	0.999999996	0.994436451		
Cocle	1,69E-06	0.9999999998	0.997620088		
Colon	8,34E-07	0.99999999992	0.996414523		
Chiriqui	4,57E-07	0.99999999995	0.995164563		
Darien	1,37E-02	0.99999986	1		
Herrera	2,11E-06	0.9999999998	0.999564716		
Los santos	2,13E-06	0.9999999997	0.998165382		
Panama	9,18E-08	0.999999999991	0.991455029		
Veraguas	1,76E-06	0.9999999998	0.998657597		
Total Country	1.27	×	0.3308	0.4441	0.4428
10-12	0.999999999987	0.992655332			0.4947
LOCUS	MP	PD	PE	TPI	PIC
HLD77	0.409	0.591	0.170	0.959	0.352
HLD45	0.384	0.616	0.166	0.951	0.364
HLD131	0.341	0.659	0.115	0.836	0.375
HLD70	0.450	0.550	0.101	0.803	0.312
HLD6	0.385	0.615	0.197	1.022	0.373
HLD111	0.407	0.593	0.111	0.828	0.339
HLD58	0.513	0.487	0.062	0.710	0.271
HLD56	0.384	0.616	0.153	0.922	0.361
HLD118	0.416	0.584	0.133	0.876	0.339
HLD92	0.409	0.591	0.143	0.899	0.345
HLD93	0.380	0.620	0.186	0.997	0.372
HLD99	0.367	0.633	0.162	0.943	0.372
HLD88	0.395	0.605	0.189	1.003	0.365
HLD101	0.383	0.617	0.200	1.028	0.375
HLD67	0.387	0.613	0.205	1.041	0.374
HLD83	0.394	0.606	0.148	0.912	0.354
HLD114	0.353	0.647	0.142	0.897	0.374
HLD48	0.466	0.534	0.100	0.801	0.302
HLD124	0.397	0.603	0.133	0.878	0.349
HLD122	0.377	0.623	0.116	0.838	0.356

Continúa en la próxima pag.

HLD125	0.457	0.543	0.123	0.855	0.312
HLD64	0.519	0.481	0.062	0.711	0.266
HLD81	0.404	0.596	0.102	0.805	0.339
HLD136	0.399	0.601	0.215	1.065	0.371
HLD133	0.405	0.595	0.177	0.976	0.356
HLD97	0.372	0.628	0.182	0.988	0.375
HLD40	0.384	0.616	0.177	0.976	0.368
HLD128	0.366	0.634	0.133	0.877	0.366
HLD39	0.400	0.600	0.193	1.012	0.364
HLD84	0.394	0.606	0.205	1.041	0.370

Discusión

La variación de la mezcla genética con INDELS es comparable a la de STRs

Los cálculos de la estructura poblacional mostraron que la probabilidad logarítmica más alta $L(K)$ de los modelos de probabilidad de mezcla fueron para K_2 , K_3 y K_4 (Figura 2). Sin embargo, varios estudios previos han demostrado que los modelos K_2 y K_4 están pobremente sustentados genética y etnohistóricamente. En contraste, el modelo trihíbrido (K_3) (Figura 2, centro) está mejor sustentado como lo demostramos con loci STR (Castro-Pérez 2016) y marcadores clásicos (Arias et al. 2002), al igual que otros grupos de investigación que utilizaron ADNmt (Perego et al. 2012) y análisis del cromosoma Y (Grugni 2016). Además, los datos etnohistóricos fuertemente sostienen que la población panameña es resultado del proceso de mezcla que involucra amerindios, ancestros africanos y europeos (Jaén-Suárez 1978; Jaén-Suárez 1998). Por lo tanto, todos los análisis adicionales de INDELS se realizaron con base en un modelo de mezcla trihíbrido (K_3) que muestra que la población general está compuesta por tres poblaciones ancestrales en las siguientes proporciones 0,46, 0,24 y 0,30 (Figura 2, centro). Curiosamente, las proporciones de mezcla y la distribución de genes son altamente consistentes con nuestro informe anterior utilizando loci STR (Tabla 1). Además, las proporciones de mezcla varían en provincia como se evidencia en la Tabla 1 y en el gráfico de barras de Estructura (Figura 3, abajo). Estas similitudes con los STRs sugieren que, a pesar de la ausencia de datos ancestrales de INDELS, las proporciones observadas corresponden a ancestros amerindios (grupo 1), africanos

(grupo 2) y europeos (grupo 3). De acuerdo con esto, el gráfico de triángulo (Figura 3, arriba) muestra tres grupos de genes que se asemejan a tres genes poblaciones parentales correspondientes a amerindios (grupo 1), africanos (grupo 2) y europeo (grupo 3). Es importante destacar que, aunque las proporciones ancestrales en INDELs y la distribución dentro de las provincias muestran algunas variaciones en comparación con los STRs, estas variaciones podrían estar relacionadas con la falta de datos ancestrales de INDELs, el tamaño relativamente pequeño de la muestra en algunas provincias o simplemente diferencias entre ambos métodos. Otra fuente de variación podría ser el hecho de que las muestras utilizadas para los INDELs fueron diferentes a las utilizadas en el análisis con STRs. Se utilizaron muestras de ADN (332 sujetos, la mitad hombres/mujeres) para los análisis de los INDELs que representan un muestreo totalmente nuevo, por lo que no se esperaban exactamente los mismos resultados que los observados con STRs (650 sujetos, mitad hombres/mujeres). Sin embargo, es interesante el hecho de que los análisis muestran una alta similitud con ambos métodos y diferentes muestreos. Esta reproducibilidad de métodos y muestreos es muy útil para múltiples aplicaciones de estos marcadores, incluyendo antropología y biomedicina, pero particularmente en paternidad y casos de genética forense.

Las frecuencias alélicas son muy similares a las de africanos e hispanicos estadounidenses

La distribución de la mezcla de genes y la estructura poblacional observada con marcadores INDELs fue altamente consistente con la observada con STRs en la población panameña. Sin embargo, cómo estos INDELs en la población panameña se comparan con otras poblaciones del mundo se desconocía. Consideramos por lo tanto importante comparar los datos de las frecuencias alélicas (Tabla 2) con poblaciones de referencia en EE.UU. como caucásicos-americanos, afroamericanos y asiáticos, en las cuales se utilizaron INDELs en casos de paternidad y análisis forense (Fondevilla 2014). Encontramos que la frecuencia alélica de 11 loci en la población panameña era muy similar a la de los afroamericanos estadounidenses (loci HLD77, HLD70, HLD111, HLD58, HLD118, HLD92, HLD93, HLD122, HLD125, HLD64 y HLD133). Además, 9 loci en panameños (loci HLD131, HLD99, HLD88, HLD101, HLD67, HLD114, HLD136, HLD97 y HLD39) fueron muy similares a la población hispanoamericana de EE. UU., mientras que el locus HLD48 fue similar tanto a la población afroamericana como a

la hispanoamericana. Dos loci mostraron similitudes combinadas en frecuencias alélicas con poblaciones Hispanoamericanas y caucásicos de los Estados Unidos, uno de los alelos con hispanoamericanos y en el otro con caucásicos (loci HLD56 y HLD40), mientras que el locus HLD45 mostró similitud en frecuencia en uno de los alelos con Hispanoamericanos y en el otro con asiáticos ambas de Estados Unidos. La frecuencia alélica para el locus HLD83 solo era similar a asiáticos, mientras que en el alelo del locus HLD84 la frecuencia fue similar a todas las poblaciones de referencia, excepto a los asiáticos. Curiosamente, cuatro loci (HLD6, HLD124, HLD81 y HLD128) no fueron comparables a las poblaciones de referencia de EE. UU., lo que muestra patrones de frecuencia únicos para la población panameña (Tabla 2). Es importante destacar que los valores de heterocigosidad (H_o y H_e) también mostraron variación entre loci. La heterocigosidad observada para estos loci varió entre 0,2961 (locus HLD58) y 0,5287 (locus HLD136), con un promedio de 0,4477. La heterocigosidad esperada varió entre 0.3175 (HLD64) a 0.4997 (HLD131) con un promedio de 0.4553, cerca al valor teórico esperado para los marcadores bialélicos (0,500). En general, el rango de variación alélica y heterocigosidad, junto con similitudes en el polimorfismo alélico con poblaciones de referencia, indica una alta diversidad genética de estos loci para uso forense en la población panameña.

Los marcadores polimórficos INDELS como complementos efectivos de STRs

Los parámetros forenses y de paternidad indican que cuando se combinan, los 30 marcadores INDELS son adecuados para la mayoría de las aplicaciones forenses en la población panameña (Cuadro 3). La probabilidad de coincidencia combinada (CMP) para los panameños fue $1,27 \times 10^{-12}$, lo cual es aceptable para uso forense, similar a otras poblaciones del mundo que utilizan actualmente INDEL para casos forenses. Por ejemplo, CMP para las poblaciones chinas oscilan entre $1,8-3,17 \times 10^{-11}$ (Wang et al. 2014), mientras que las poblaciones de referencia de EE. UU. son relativamente más altas en estadounidenses caucásicos con $3,65 \times 10^{-13}$. Sin embargo, los hispanoamericanos estadounidenses ($2,12 \times 10^{-12}$) y los afroamericanos estadounidenses ($1,43 \times 10^{-11}$) [11] muestran CMP muy similar a la población panameña. Además, Brasil ($3,4 \times 10^{-13}$) (Ferreira et al. 2015), que también tiene una población mestiza como la de Panamá, mostró una CMP mayor

que la obtenida en los panameños. Además, el poder combinado de discriminación para los panameños (CPD) fue 0.9999999999987, que es similar a los reportados en poblaciones mestizas de Brasil (0,9999999999997) (Ferreira et al. 2015) y México (0,99999999) (Martínez-Cortés et al. 2016). El poder combinado de exclusión (CPE) (Cuadro 3) para los panameños fue 0.992655332, que es relativamente menor que otras poblaciones, incluidas todas las poblaciones de referencia de EE. UU., con 0.999999999 (LaRue et al. 2012) y también más bajo que Brasil con 0,997250806 (Ferreira et al. 2015). Sin embargo, el CPE de los mestizos panameños es mayor que Somalia (0,9860) (Neuvonen et al. 2012) y China (0,988) (Wang et al. 2014). El índice típico de paternidad (TPI), arrojó valores entre 0,710 (locus HLD58) y 1.065 (locus HLD136), con un promedio de 0.9150; relativamente más bajo que el valor generado a partir de STRs. El contenido de información polimórfica (PIC) osciló entre 0,266 (locus HLD64) y 0,375. Estos valores en conjunto indican que, en la población panameña, los marcadores INDELS son adecuados para la mayoría de los análisis genéticos forenses, en especial para aquellos que involucran muestras de ADN degradado. Sin embargo, los INDELS deben complementarse con STRs u otros marcadores de paternidad y parentesco. Aunque la degradación de ADN suele ser menos común en las muestras utilizadas para pruebas de paternidad, éstas podrían ser más susceptibles a mutaciones comunes en los loci STRs que en marcadores INDELS.

Conclusión

Los resultados obtenidos en esta investigación proporcionan evidencia a favor del uso de marcadores INDELS en la población panameña con aplicaciones forenses, de paternidad y antropológicas. Los parámetros estadísticos forenses de INDELS son comparables a los reportados para poblaciones de referencia de EE. UU. y poblaciones mundiales. Además, las estimaciones de la mezcla de genes utilizando tanto INDELS como el STRs demuestran que ambos marcadores son suficientemente poderosos para detectar variaciones muy similares de ascendencia genética observados en todo el país.

En conjunto, estos datos sugieren que los marcadores INDELS pueden ser afectivamente aplicables a la población panameña para la mayoría de las aplicaciones forenses, pero deben ser complementados con marcadores adicionales para los análisis de paternidad.

Literatura citada

1. Arias TD, Castro E, Ruiz E, Barrantes R, Jorge-Nebert L, et al. (2002) Racial mix of the panamanian population. *Rev Med Panama* 27: 5-17.
2. Carvalho A, Pinheiro MF (2013) Population data of 30 insertion/ deletion polymorphisms from a sample taken in the north of Portugal. *Int J Legal Med* 127: 65-67.
3. Castro EA, Trejos DE, Berovides-Alvarez V, Arias TD, Ramos CW, et al. (2007) Genetic polymorphism and forensic parameters of nine short tandem repeat loci in Ngöbé and Emberá Amerindians of Panama. *Hum Biol* 79: 563-577.
4. Castro-Pérez E, Trejos DE, Hrbek T, Setaluri V, Ramos CW, et al. (2016) Genetic ancestry of the panamanian population: Polymorphic structure, chibchan amerindian genes; and biological perspectives on diseases. *Inter J Biol Anthropology* 9: 1-14.
5. da Costa Francez PA, Rodrigues EM, de Velasco AM, dos Santos SE (2012) Insertion-deletion polymorphisms-utilization on forensic analysis. *Int J Legal* 126: 491-496.
6. Earl DA, vonHoldt BM (2012) Structure harvester: A website and program for visualizing structure output and implementing the Evanno method. *Conservation Genetics Resources* 4: 359-361.
7. Falush D, Stephens M, Pritchard JK (2000) Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics* 164: 1567-1587.
8. Ferreira Palha TJB, Ribeiro Rodrigues EM, Cavalcante GC, Marrero A, de Souza IR, et al. (2015) Population genetic analysis of insertion-deletion polymorphisms in a Brazilian population using the Investigator DIPplex kit. *Forensic Sci Int Genet* 19: 10-14.
9. Fondevila M, Phillips C, Santos C, Pereira R, Gusmão L, et al. (2014) Forensic performance of two insertion-deletion marker assays. *Int J Legal Med* 126: 725-737.
10. Grugni V, Battaglia V, Perego UA, Raveane A, Lancioni H, et al. (2015) Exploring the Y chromosomal ancestry of modern panamanians. *PLoS One* 10: e0144223.
11. Investigator DIPplex Handbook (2014) QIAGEN.
12. Jaén-Suárez O (1998) The population of the isthmus of panama, study of geo-history, Agencia Española de Cooperación Internacional, Spain.
13. Jaén-Suárez O (1998) The population of the isthmus of panama, study of geo-history, Agencia Española de Cooperación Internacional, Spain
14. LaRue BL, Ge J, King JL, Budowle B (2012) A validation study of the Qiagen investigator DIPplex® kit; An INDEL-based assay for human identification. *Int J Legal Med* 126: 533-540.
15. Martínez-Cortés G, García-Aceves M, Favela-Mendoza AF, Muñoz-Valle JF, Velarde-Felix JS, et al. (2016) Forensic parameters of the Investigator DIPplex kit (Qiagen) in six Mexican populations. *Int J Legal Med* 130: 683-685.

16. Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 89: 583-590. Jaén-Suárez O (1978) The population of the isthmus of panama XVI century to the Xx: Impresora de La Nación, Spain.
17. Neuvonen AM, Palo JU, Hedman M, Sajantila A (2012) Discrimination power of Investigator DIPplex loci in Finnish and Somali populations. *Forensic Sci Int Genet* 6: e99-e102.
18. Perego UA, Lancioni H, Tribaldos M, Angerhofer N, Ekins JE, et al. (2012) Decrypting the mitochondrial gene pool of modern Panamanians. *PLoS One* 7: e38337.
19. Pereira R, Phillips C, Pinto N, Santos C, dos Santos SE, et al. (2012) Straightforward inference of ancestry and admixture proportions through ancestry-informative insertion-deletion multiplexing. *PLoS One* 7: e29684.
20. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155: 945-959.
21. Resque RL, Freitas NoS, Rodrigues EM, Guerreiro JF, Santos NP, et al. (2010) Estimates of interethnic admixture in the Brazilian population using a panel of 24 X-linked insertion/deletion markers. *Am J Hum Biol* 22: 849-852.
22. Saloum de Neves MF, Pereira R, Vianna R, Rodolfo Beuttenmüller de AA, Leite Góes Gitai D, et al. (2013) Revisiting the genetic ancestry of Brazilians using autosomal AIM-Indels. *PLoS One* 8: e75145.
23. Santos NP, Ribeiro-Rodrigues EM, Ribeiro-Dos-Santos AK, Pereira R, Gusmão L, et al. (2010) Assessing individual interethnic admixture and population substructure using a 48-insertion-deletion (INSEL) ancestry-informative marker (AIM) panel. *Hum Mutat* 31: 184-190.
24. Shewale JG, Liu RH (2013) *Forensic DNA analysis: Current practices and emerging technologies*: CRC press, Taylor and Francis Group, USA.
25. Torres SR, Uehara CJ, Sutter-Latorre AF, de Almeida BS, Sauerbier TS, et al. (2014) Population genetic data and forensic parameters of 30 autosomal InDel markers in Santa Catarina state population, Southern Brazil. *Mol Biol Rep* 41: 5429-5433.
26. TZ Gao, LB Yun, W He, Y Gu, YP Hou, et al. (2015) The application of multi-InDel as supplementary in paternity cases with STR mutation. *Forensic Sci Int Genet* 5: e218-e219.
27. Wang Z, Zhang S, Zhao S, Hu Z, Sun K, et al. (2014) Population genetics of 30 insertion-deletion polymorphisms in two Chinese populations using Qiagen Investigator® DIPplex kit. *Forensic Sci Int Genet* 11: e12-e14.
28. Weber JL, David D, Heil J, Fan Y, Zhao C, et al. (2002) Human diallelic insertion/deletion polymorphisms. *Am J Hum Genet* 71: 854-862.
29. Yeh FC, Yang RC, Boyle TBJ, Ye ZH, Mao JX, et al. (1997) PopGene, the user-friendly shareware for population genetic analysis, molecular biology and biotechnology center.

C A P Í T U L O 4

Asociación clínica
inmunológica y molecular
de los alelos DRB1, DQA1
y DQB1 con la diabetes
mellitus tipo I en la
población de Panamá

Brenda de Mayorga y Olga Chen

Resumen

La Diabetes Mellitus del tipo I es una enfermedad autoinmune que se caracteriza por la infiltración linfocítica de los islotes de Langerhans y la aparición de autoanticuerpos contra los antígenos de las células β .

Los genes Clase II del MHC están asociados a Diabetes Mellitus del tipo I. Múltiples estudios han demostrado que DRB1*, DQA1* y DQB1* están asociados a la enfermedad. En este estudio se encontró que DR3, DR4 y DQ están asociados con la enfermedad, incrementándose el riesgo en los heterocigotos DR3/DR4. Después de realizar análisis genéricos de los alelos DRB1* se determinaron que los subtipos del alelo DRB1* 04, observándose que el más frecuente en la población de pacientes estudiados es el DRB1*04011-0414 y el *0416-*0404. Además, se encontraron el DRB1*0407 y *0404 sólo en pacientes con Diabetes Mellitus Tipo I. En los análisis moleculares específicos de DRB1* se encontró que el DRB1*05031-05032, DRB1*0501 y DRB1*0502 y DRB1*0201 están mayormente asociados a la enfermedad con una alta predisposición o riesgo de padecer de Diabetes tipo I.

El análisis de los alelos DQA1* y DQB1* se realizó mediante frecuencias fenotípicas y genotípicas sin establecer la asociación con la Diabetes Mellitus Tipo I debido a la falta de grupo de control para estos alelos. Estos estudios pueden ser continuados y comparados con testigos para determinar el riesgo relativo y la frecuencia etiológica.

Palabras claves

HLA, Diabetes mellitus, PCR-SSP, MHC

Abstract

Type I Diabetes Mellitus is an autoimmune disease characterized by lymphocytic infiltration of Langerhans's islets and autoantibodies' appearance against cell antigens. MHC Class II genes are associated with type I Diabetes Mellitus. Multiple studies have shown that DRB1*, DQA1*, and DQB1* are associated with the disease. In this study, DR3, DR4, and DQ were associated with the disease, with increased risk in DR3/DR4 heterozygotes. After performing a genetic analysis of the DRB1* alleles, it was determined that the subtypes of the DRB1* 04 alleles, observing that the most frequent in the population of patients studied are DRB1*04011-0414 and *0416-*0404. In addition, DRB1*0407 and *0404 were found only in

patients with Type I Diabetes Mellitus. In the specific molecular analyses of DRB1*, DRB1*05031-05032, DRB1*0501 DRB1*0502, and DRB1*0201 were found to be primarily associated with disease with a high predisposition or risk for Type I Diabetes. The analysis of DQA1* and DQB1* alleles was performed by phenotypic and genotypic frequencies without establishing the association with Diabetes Mellitus Type I due to the lack of a control group for these alleles. These studies can be continued and compared with controls to determine relative risk and etiological frequency.

Keywords

HLA, Diabetes mellitus, PCR-SSP, MHC

Introducción

La Diabetes Mellitus Tipo 1, también conocida como Diabetes Mellitus Insulina Dependiente (IDDM) o diabetes juvenil, es una enfermedad autoinmune de tipo crónico cuya deficiencia en la producción de insulina se debe a la destrucción de las células Beta de páncreas por los linfocitos T que infiltran los islotes de Lagerhans y a la presencia de auto-anticuerpos contra los anticuerpos de las células β (Eisenbarth, 1986; Harrinson et al., 1990; Atkinson et al., 1994; Dorman, 1997).

Diversos estudios han revelado que la Diabetes Mellitus Tipo I, es una enfermedad poligénica en la que participan los genes del Complejo Mayor de Histocompatibilidad (MHC), cuya función general es la inducción y regulación de la respuesta inmune mediante la presentación de péptidos. Los productos de estos genes han sido bien caracterizados, determinándose que influyen drásticamente en el desarrollo de la enfermedad (Nepom & Kwok 1998; Rothbarb & Geftter, 1991; Sheehy et al, 1989; Davies et al., 1994; Hashimoto et al., 1994). El complejo mayor de histocompatibilidad está localizado en el brazo corto del cromosoma 6 y en humanos ha sido denominado complejo del antígeno leucocitario humano (HLA) y está compuesto de tres tipos de genes; I, II y III. Los genes tipo I comprenden los loci A, B, C; cuya función es reconocer antígenos extraños por linfocitos T citotóxicos. El tipo II representado por el locus HLA-D con subregiones DR, DP y DQ, responsable de reconocer antígenos extraños por linfocitos cooperadores. Los de tipo III codifican para algunos componentes de la cascada del complemento (Roitt et al., 1998). Los genes tipo I y II del sistema HLA están asociados con la predisposición o protección contra la enfermedad (Huang, 1996).

Diversos estudios sugieren la existencia de diferencias interraciales entre el complejo HLA y la diabetes tipo I (Todd et al. 1989). Alrededor de 1 90% de los blancos caucásicos que padecen de diabetes tipo I son homocigotos HLA-DR3 y HLA-DR4 o heterocigóticos DR3/DR4 (Sheehy et al. 1989; Ronnigen, 1993; Rossini et al. 2000; Tiwari & Terasaki, 1985; Uchigata, 1992; Nepom, 1990).

En japoneses la combinación DR4/DR9 y DR8/DR9 incrementa la susceptibilidad a la diabetes tipo I (Bertrams & Uchigata, 1990).

Los loci DQA* y DQB* están estrechamente implicados en la susceptibilidad y resistencia de la expresión de la diabetes tipo I. La asociación principal ha sido con los haplotipos DQB1*0302, DQA1*0301 y DQB1*0201, DQA*0501 (Gorodezki, 1995).

En este estudio se determinó mediante PCR-SSP la tipificación de los alelos del sistema HLA. Se determinaron las frecuencias alélicas y genotípicas en pacientes con diabetes tipo I, el Riesgo Relativo (RR) y la Fracción Etiológica (FE) para establecer el grado de asociación de estos alelos con la enfermedad. La información generada en este estudio es de suma importancia, ya que permitirá identificar de manera temprana sujetos con susceptibilidad a desarrollar la enfermedad como modelo de estrategia de medicina preventiva.

Materiales y métodos

Muestra

Se obtuvieron un total de 169 muestras de sangre de las que 119 correspondieron individuos sanos provenientes de la población general y 50 individuos pertenecientes a 12 familias en las que se encontraron individuos sanos (29) y enfermos (21) con diabetes mellitus tipo I referidos del servicio de endocrinología del complejo Hospitalario Metropolitano de la Caja de Seguro Social. La sangre obtenida fue utilizada para la extracción de ADN.

Se preparó un formulario para recopilar las características de la población control y en estudio. Entre las características estaban: sexo, edad, aparición de enfermedad, DR01 baja resolución, DR0104 alta resolución, DQ α , DQ β y DR.

El procesamiento de las muestras se realizó en el Laboratorio de Inmunología Molecular del Complejo de la Caja de Seguro Social. Se estudiaron doce familias, que incluían padres e hijos; de las cuales obtuvimos

21 muestras provenientes de individuos afectados por la enfermedad. Esto sugiere la existencia de un marcador de predisposición genética. Luego se procedió a la determinación de antígenos HLA entre los pacientes y los sanos.

Extracción de ADN

El ADN se extrajo a partir de sangre periférica utilizando el protocolo DTAB/CTAB de Dynal-SSP (Dynal, Norway). El ADN se almacenó a -70°C hasta su posterior utilización para la tipificación mediante una Reacción en Cadena de la Polimerasa de Cebadores de Secuencias Específicas (PCR-SSP).

Amplificación y tipificación de los alelos HLA

La tipificación de los alelos se realizó mediante el uso de la técnica de la reacción en cadena de la polimerasa con cebadores de secuencia específica (PCR-SSP) de acuerdo al procedimiento Dynal-SSP (Dynal Biotech, Hamburg, Alemania). Estos cebadores contienen secuencias específicas y amplifican el segundo exón polimórfico DRB, DRA y DRQ, los cuales varían en tamaño y secuencia.

Las condiciones de amplificación fueron 94°C 2 minutos, seguido de 10 ciclos de 94°C 10 segundos, 65°C 60 segundos y 20 ciclos de 94°C 10 segundos, 61°C 50 segundos y 72°C 30 segundos.

Igualmente, se procedió a la identificación de los subtipos de cada uno de los alelos utilizando cebadores genéricos (baja resolución) y cebadores específicos (alta resolución). Todos los productos de PCR fueron guardados a -20°C para proseguir con la caracterización mediante corrida en gel de agarosa al 2%.

Electroforesis

Cinco microlitros de los productos de PCR, incluyendo los controles, se corrieron en gel de agarosa al 2% por aproximadamente 30 minutos a 260 voltios para su identificación, como indica el protocolo Dynal-SSP. Luego

se examinaron en un transiluminador (*Fisher Scientific*) para determinar la presencia de bandas correspondientes para cada cebador y se documentaron con las fotografías.

Interpretación

La asignación de los alelos se realiza con base a los productos de amplificación generados los cuales se comparan con la tabla de Dynal-SSP para asignación de alelos. Cada tubo con la solución de cebador tiene además un control interno positivo que amplifica un segmento del gen de la hormona de crecimiento de 429 pares de bases.

Análisis estadístico

Los alelos presentes en los pacientes y testigos fueron comparados por chi cuadrado (X^2) con la corrección de Yates.

Las frecuencias génicas y antigénicas fueron calculadas con la siguiente fórmula:

$$FA = \frac{\text{portadores}}{N} \times 100$$

$$FG = I - \sqrt{1 - A}$$

N = total de individuos de la muestra; FA = frecuencia antigénica; FG = frecuencia génica (alélica)

El Riesgo Relativo es una estimación de la intensidad de la asociación que indica cuantas veces más riesgo de desarrollar la enfermedad tiene un individuo portador del antígeno en cuestión.

El riesgo relativo (RR) se calculó de acuerdo al método de Wolf utilizando cuatro casillas:

$$RR = \frac{ad}{bc}$$

RR = Riesgo Relativo; a = pacientes con el antígeno; b = pacientes sin el antígeno; c = testigos con el antígeno; d = testigos sin el antígeno

La fracción etiológica es una medida relativa de cuanto de la asociación se debe al marcador en cuestión. Mientras más se acerca el valor a uno, más se debe la enfermedad al marcador.

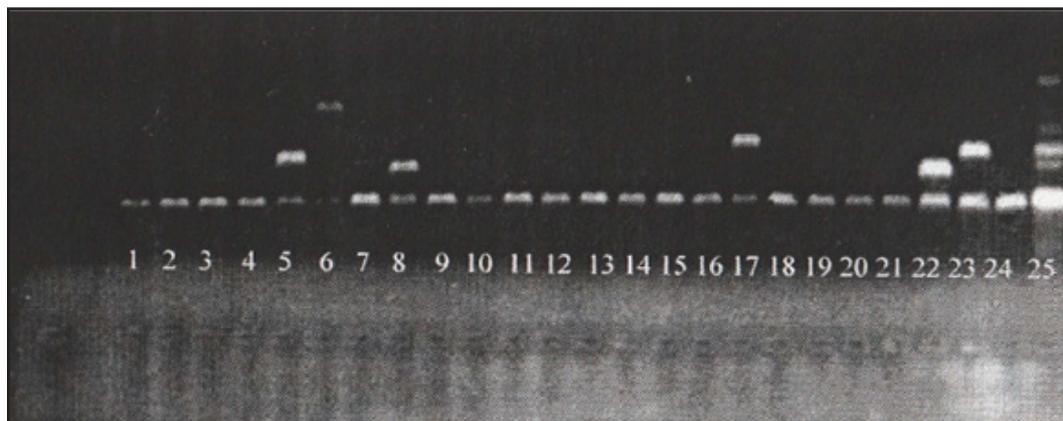
La fracción etiológica se calculó según la siguiente fórmula:

$$FE = \frac{RR - 1}{RR} \times \frac{a}{a+b}$$

Resultados y discusión

Tipificación de los alelos

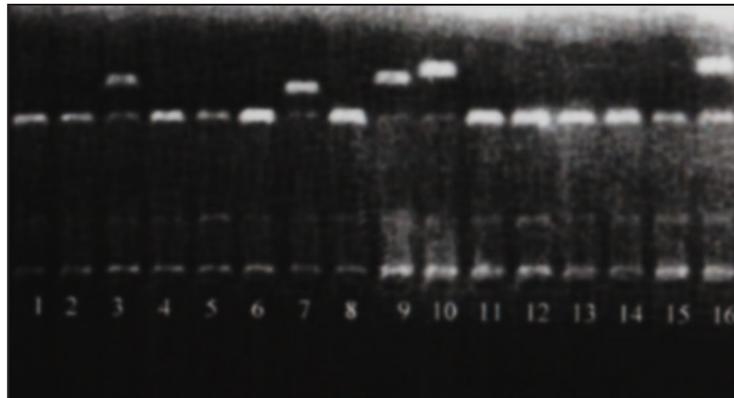
La tipificación de los alelos se llevó a cabo comparando los productos de amplificación con las tablas de Dynal-SSP. Los patrones de amplificación del alelo DRB1* 03 con los diferentes subtipos se muestra en la figura 1.



carril	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Banda	-	-	-	-	+	-	-	+	-	-	-	-	-	-	-	-	+	-	-	-	-	+	+	-	+
Tamaño	195	195	210	215	220	80	85	260	185	160 215	195	205	165 175 185	105	200	195 200 210	175	200 210	145 150	150 220	170	230	200 215	260	

Figura 1. Patrón de bandas de DRB1* y subtipo. El patrón de amplificación observado en los carriles 5, 8, 17, 22, 23 y 25 se corresponde en la tabla con el alelo DRB1*03. Los productos de amplificación presentan tamaños en los carriles antes mencionados de 220, 80, 260, 175, 230 y 200-215 respectivamente.

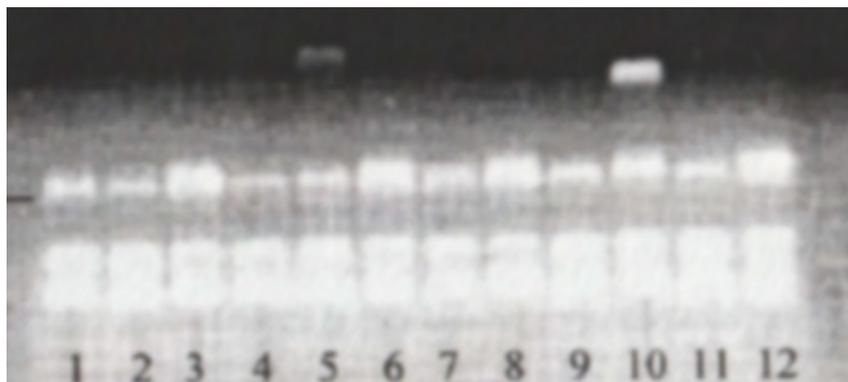
En la figura 2 se observa amplificación en los carriles 3, 7, 9, 10 y 16 característica del alelo DRB1* 04 subtipo DRB1*0407.



carril	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Banda	-	-	+	-	-	-	+	-	+	+	-	-	-	-	-	+
Tamaño	210	220	210	220	170	110	260	260	185	210	195	205	165 175 185	105	200	195 200 210

Figura 2. Patrón de amplificación del subtipo DRB1* 0407. Los productos de amplificación de 220, 260, 210, 260, 170, y 150 pb en los carriles 3, 7, 9, 10 y 16 respectivamente, corresponden al patrón de amplificación de DRB1* 0407 presentado en la tabla.

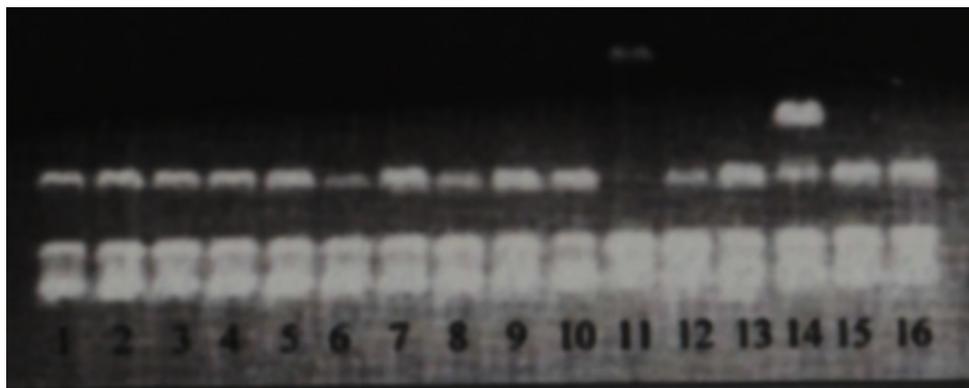
En la figura 3 se observa el patrón de amplificación de los alelos DQB1* 0201 y 0203. La presencia de un producto de amplificación de 105 bp en el carril 5 indica la presencia del alelo DQB1* 0201, mientras que la banda de 140 bp en el carril 10 representa la presencia del alelo DQB1* 0203



carril	1	2	3	4	5	6	7	8	9	10	11	12
Banda	-	-	-	-	+	-	-	-	-	+	-	-
Tamaño	120 130	165	205	165	105	255	205	175	145	140	120	65 130 195

Figura 3. Amplificación específica de los alelos DQB1* 0201 y DQB1* 0203. Se observa un producto de amplificación de 105 bp en el carril 5 correspondiente al alelo DQB1* 0201 y un producto de 140 bp correspondiente al alelo DQB1* 0203.

La figura 4 muestra el patrón de amplificación del alelo DQA1* con bandas de 190 y 220 en los carriles 11 y 12 respectivamente, como se presenta en la tabla de referencia para la tipificación de alelos.



carril	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Banda	-	-	-	-	-	-	-	-	-	-	+	-	-	+	-	-
Tamaño	145	175	145	175	220	95	175	185	215	220	190	90	90	220	200	115

Figura 4. Patrón de amplificación del alelo DQA1* 0401. Producto de amplificación de 190 y 220 bp en los carriles 11 y 14 de la tabla.

Los marcadores genéricos del HLA encontrados en este estudio se muestran en la tabla 1. Se observó que el alelo con mayor frecuencia antigénica (FAG) fue DR3 con 76.2 en los pacientes y 14.2 en los testigos. Con relación al riesgo relativo (RR), los pacientes presentan un RR =19.20. Por otro lado, los pacientes con el alelo DR4 presentan una frecuencia antigénica de 61.9 y los testigos 39.5 y un RR = 2.5. Resultados similares se reportaron en la población mexicana (Gorodesky et al. 1996). Los alelos DQ2 presentaron un RR de 78%, mientras que DQ8 fue de 20.5. Por otro lado, los alelos DQ5, DQ6 y DQ7 presentaron riesgos relativos bajos y se observan raramente en pacientes con IDDM, lo que sugiere podrían estar implicados en protección.

Tabla 1. Marcadores HLA en panameños con IDDM. FAg, frecuencia antigénica; FE, fracción etiológica; RR, riesgo relativo.

ANTIGENO	P (21)	T (119)		
HLA	FAg	FAg	RR	FE
DR3	76.2	14.2	19.2	0.72
DR4	61.9	39.5	2.5	0.37
DR11	8.0	6.0	1.4	0.037
DR3/DR4	28.6	1.7	23.4	0.27
DQ2	66.2	5.0	78.0	0.65
DQ5	66.7	32.5	4.5	0.51
DQ6	61.9	27.5	4.3	0.48
DQ7	61.9	35.0	3.0	0.41
DQ8	61.9	5.0	20.5	0.59

Los alelos más frecuentes del DRB1*04 en la población estudiada fueron DRB1*04011-04014 con una frecuencia de 0.32 (0.17) y el DRB1*0416-0424 con 0.28 (0.14) (Tabla 2). La subtipificación del alelo DRB1*04 se observó que 8 pacientes con diabetes mellitus tipo I eran portadores del alelo DRB1*0407 mientras que sólo dos pacientes del DRB1*0404. No se observaron estos alelos en la población sana. La presencia de estos alelos en pacientes podría sugerir su relación con la enfermedad.

El alelo DRB1*0404 que se presentó con la menor frecuencia, se reporta con una frecuencia relativamente alta en pacientes mestizo mexicanos con IDDM (Gorodezky et al. 1997).

Tabla 2. Frecuencia de los alelos DRB1*04 en la población estudiada. N, cantidad de individuos; FF, frecuencia fenotípica; FA, frecuencia alélica

DRB1* 04	N	FF	FA
0407	8	0.17	0.08
0404	2	0.04	0.02
04011-04014	16	0.34	0.17
0416-0424	14	0.29	0.14
04051-04052	7	0.15	0.07

Con relación a la frecuencia de los alelos DQA1* (Tabla 3) se observó que los alelos DQA1* 0601, *03011, *05011- *05013 y *0303 son los que mayor frecuencia fenotípica y alélica presentan.

Tabla 3. Frecuencia de los alelos DQA1* en la población estudiada. N, cantidad de individuos; FF, frecuencia fenotípica; FA, frecuencia alélica

DQA1*	N	FF	FA	FE
03011	27	0.19	0.01(0.14)	FE
05011-05013	25	0.18	0.09 (0.13)	0.72
0601	30	0.21	0.11(0.16)	0.37
0401	3	0.02	0.01 (0.016)	0.037
01021-01022	11	0.08	0.04 (0.059)	0.27
0104	1	0.01	0.01(0.005)	0.65
0303	16	0.11	0.06 (0.086)	0.51
0101	11	0.08	0.04 (0.059)	0.48
0105	7	0.05	0.02 (0.038)	0.41
0101	11	0.08	0.04 (0.059)	0.59

Con relación a la frecuencia de los alelos DQB1* los más frecuentes en orden descendientes son DQB1* 05031-05032, *0501, *0502 y *0201 (tabla 4).

Tabla 4. Frecuencias de los alelos DQB1* en la población estudiada.

DQB1*	N	FF	FA
0301	28	0.107	0.054
0501	31	0.119	0.060
0502	30	0.115	0.057
05031-05032	34	0.130	0.065
06011	22	0.085	0.042
0201	30	0.115	0.057
0302	15	0.058	0.029
0304	15	0.058	0.029
0202	2	0.008	0.004
0504	2	0.008	0.004
03032	23	0.088	0.011

Continúa en la próxima pag.

0305	2	0.008	0.004
0307	1	0.004	0.002
0401	7	0.027	0.0135
0402	7	0.027	0.0135
0504	3	0.011	0.006
0202	8	0.030	0.015

Conclusiones

Se demuestra en este estudio la asociación de DR3, DR4, DQ2 y DQ8 con la diabetes mellitus tipo I. Los alelos DRB1* 0301, *0102 y *0101 presentan un alto riesgo como se demostró por la alta frecuencia en individuos con IDDM en la población estudiada. Los subtipos del alelo DRB1*04 con mayor frecuencia fueron DRB1*04011-0414 y el DRB1*04016-0424. Además, sólo se encontraron en individuos con IDDM los alelos DRB1*0407, *04051-*04051 y *0404, mas no en individuos sanos, lo cual sugiere una posible contribución de estos con la enfermedad.

Con relación al alelo DQA1* los genotipos y fenotipos DQA1*0601, *03011, *05011 y *0303 fueron los mayormente expresados. En el caso de los alelos DQB1*, los fenotipos y genotipos más expresados fueron DQB1*05031-05032, *0501, *0502 y *0201.

Se observó un efecto sinérgico del haplotipo DR3/DR4 en la población enferma, no obstante, los heterocigotos DR3/DR4 se asocian a la enfermedad; sin embargo, los homocigotos DR3/DR3 presentan un riesgo similar de padecer de IDDM que los heterocigotos. Es posible que esto se deba a la presencia de otros genes que incidan sobre el desarrollo de la condición (McGrail et al. 2023).

Literatura citada

1. Atkinson, M. A., Bowman, M. A., Campbell, L., Darrow, B. L., Kaufman, D. L. and Macclaren, N. K. 1994. Cellular Immunity to a Determinant Common to Glutamate Decarboxylase and Coxsackie Virus in Insulin-dependent Diabetes. *J. Clin. Invest.* Volume 94, November 1994, 2125-2129
2. Bertrams J.& Uchigata K. 1990. An investigation of iapanese subject Susceptibility to type insulinodependent diabetes mellitus. *Human Immunology.* 33 24-28
3. Dorman, J. 1997. Molecular Epidemiology of Insulin dependent diabetes mellitus: WHO Dlamond Project. *Gaceta Medica Mexico.* Vol 133. Suppl 1; 151-154
4. Davies, J. L., Kawaguchi, Y., Bennett, S T., Copeman, J B., Cordell, H J., Pritchard, L E., Reed, P W., Gough, S C., Jenkins, S C., Palmer, S M et al. 1994. A genome-wide search for human type 1 diabetes susceptibility genes. *Nature.* 371(6493):130-6
5. Eisenbarth G.S. 1986. Type 1 diabetes mellitus. A chronic autoimmune disease. *New England Journal of Medicine.* 314 145-153.
6. Gorodezky, C., Olivares, A., Debazo, H., Rodríguez, L., Altamirano, N., Robles, C. 1995. Los mecanismos moleculares de susceptibilidad y de protección dependientes del MHC en la diabetes tipo I en mexicanos. *Gac. méd. Méx*; 131(4): 395-402,
7. Harrison L.C., Campbell I.L., et al. 1990. Type 1 diabetes: Immunopathology. *Advanced Endocrinology Metabolism.* 1:97-117.
8. Hashimoto L., Hafhta C., Beressi J., Cambon-Thomsefl, A., y col. 1994. Genetic mapping of a susceptibility locus for insulin dependent diabetes mellitus on cromosoma 11 q. *Nature* 371 86-94.
9. Huang W., Connor E., De La Rosa T., Muir A., Schatz D., Silverstein J., Crockett S., She X. J. 1996. Although DR3-DQB1* may be associated with multiple component diseases of the autoimmunity polyglandular syndromes, the human leukocytes antigen DR4-DQB110302 haplotype is implicated only in beta cell autoimmunity. *Journal of Clinical Endrocrinology and Metabolism.* 51: 1-5.
10. McGrail C, Chiou J, Elgamal R, Luckett M, Oram R, Benaglio P, Gaulton K. 2023. Genetic discovery and risk prediction for type 1 diabetes in individuals without high-risk HLA-DR3/DR4 haplotypes. <https://doi.org/10.1101/2023.11.11.23298405>
11. Nepom, G. T and Kwok, W. W. 1998. Molecular basis of HLA-DQ associations with IDDM. *Diabetes*, vol. 47: 1177-1184
12. Roitt Ivan, Brostoff Joanathan, Male David. 1998. *Inmunology.* Fifth Edition. Mosby International Ltd.
13. Roomingen KS., Sprukland A., Tart BD., Drummond, B. & et al. 1992. HLA Class II as-sociations in insulin dependent diabetes mellitus among Blacks, caucasians and Japo-nese. *Journal of clinical Endocrinology and Metabolism.* vol. 80, No.9:456-461

14. Rossini A., Yogish K-, Wilson B., Stromenges J. & Chella D. 2000. Co Expression of HLA DR3 and DQ8 results in the development of spontaneoses insulitis and loss of tolerance to GAD 65 in transgenic mice. *Diabetes*, Vol. 49. April 2000. 548-534.
15. Rothbard JB., Geftter ML. 1991. Interaction between immunogenetic peptides and MHC. *Annual Reviews of Immunology*. 9 (527-566)
16. Sheehy M. J. Scharf S. J., Rowe J. R. et al. 1989. A diabetes susceptible haplotype is best defined by a combination of HLA-DR and DQ alleles. *Journal of clinical investigation*. 83:830-83.
17. Tewari, J. L. & Terasaki, P. I. 1985. *HLA and Disease Associations*. Springer-Verlag. 472 pp.
18. Todd J A, Mijovic C., Fletcher J., Jenkins D., Broadwell A. R. & Barnett A. M. 1989. Identification of susceptibility loci for insulin dependent diabetes mellitus by trans-racial gene mapping. *Nature*. Vol 338:587-589
19. Uchigata Y., Kuwata S., Tokunaga K. et al. 1992. Strong association of insulin autoimmune syndrome with HLA-DR4. *Lancet* 339:393-394

C A P Í T U L O 5

Polimorfismo genético y actividad de la paraoxonase en la población panameña

Ana E. Tejada E., Tomás A. Diez, Gonzalo Carrasco y Carlos Ramos Delgado

Resumen

La Paraoxonasa (PON1) humana es una enzima asociada a las lipoproteínas de alta densidad (HDL) que protege contra el daño oxidativo tanto a las lipoproteínas de baja densidad (LDL) como a las HDL. La enzima está involucrada en la hidrólisis de metabolitos altamente tóxicos de pesticidas organofosforados. En el presente estudio, se determinó la distribución de los polimorfismos de PON1 en la posición -108 de la región promotora y en las posiciones 55 y 192 de la región codificadora en la población Panameña (n=187). El genotipo de cada individuo fue determinado mediante PCR-RFLP. Las frecuencias genotípicas encontradas en la posición 55 fueron: LL 0.631, LM = 0.331, y MM = 0.037; en la posición 192: QQ = 0.331, QR = 0.497, y RR = 0.171; y en la posición -108: CC = 0.428, CT = 0.428, y TT = 0.117. Las frecuencias alélicas más altas fueron: Q = 0.581, L = 0.796 y C = 0.655. Sin embargo, el grupo indígena panameño, Gnöbe Bugle, presentó una frecuencia baja para el genotipo -108CC (0.125), lo que podría estar asociado a un mayor riesgo al envenenamiento por pesticidas organofosforados. Las actividades enzimáticas de PON1 arilesterasa y diazoxonasa fueron determinadas en muestras de la población panameña de blancos y negros. El haplotipo más frecuente en la población total fue -108CC/55LL/192QR; éste haplotipo fue el segundo más frecuente en la población Negra Panameña y el más frecuente en la población Blanca Panameña. La actividad arilesterasa para este haplotipo fue de 116.16 U/mL y 146.30 U/mL en Negros y Blancos Panameños, respectivamente. La actividad diazoxonasa fue de 10.09 U/mL y 11.25 U/mL en los grupos antes mencionados. Relaciones entre la actividad arilesterasa y los polimorfismos de PON1 son discutidos en este artículo.

Palabras claves

Polimorfismo, paraoxonasa humana, población panameña, actividad enzimática.

Abstract

Human paraoxonase (PON1) is an enzyme associated with high-density lipoproteins (HDL) that protects against oxidative damage to both low-density lipoproteins (LDL) and HDL and is involved in the hydrolysis of highly toxic metabolites of organophosphate pesticides. The present study determined the distribution of PON1 polymorphisms at position

-108 of the promoter region and positions 55 and 192 of the coding region in the Panamanian population (n=187). PCR-RFLP determined the genotype of each individual. The genotypic frequencies found at position 55 were LL 0.631, LM = 0.331, and MM = 0.037; at position 192: QQ = 0.331, QR = 0.497, and RR = 0.171; and at position -108: CC = 0.428, CT = 0.428, and TT = 0.117. The highest allelic frequencies were: Q = 0.581, L = 0.731, L = 0.731, and MM = 0.037. 581, L = 0.796 and C = 0.655. However, the Panamanian indigenous group, Gnöbé Bugle, presented a low frequency for the -108CC genotype (0.125), which could be associated with a higher risk of poisoning by organophosphate pesticides. The enzymatic activities of PON1 arylesterase and diazoxonase were determined in samples from the Panamanian population of whites and blacks. The most frequent haplotype in the total population was -108CC/55LLL/192QR; this haplotype was the second most frequent in the Black Panamanian population and the most frequent in the White Panamanian population; the arylesterase activity for this haplotype was 116.16 U/mL and 146.30 U/mL in Blacks and White Panamanians, respectively. Diazoxonase activity was 10.09 U/mL and 11.25 U/mL in the groups above. The relationship between arylesterase activity and PON1 polymorphisms is discussed in this article.

Keywords

Polymorphism, human paraoxonase, Panamanian population, enzymatic activity

Introducción

La paraoxonasa humana (PON1) es una enzima asociada a partículas HDL que es capaz de reducir la oxidación de los componentes LDL y HDL, además de hidrolizar metabolitos altamente tóxicos derivados de pesticidas organofosforados (Brophy et al., 2000). Varios estudios han encontrado que PON1 actúa como agente protector contra el desarrollo de la placa aterogénica debido a su capacidad de hidrolizar los lípidos oxidados presentes en la pared arterial, evitando la oxidación de las partículas de LDL (Durrington et al., 2001). Es bien sabido que la acumulación de partículas de LDL oxidadas es uno de los factores implicados en las enfermedades cardíacas como consecuencia del desarrollo de aterosclerosis. La enzima es una proteína de 355 residuos de aminoácidos codificada por el gen PON1

que se encuentra en el cromosoma 7q21.3-22.1 (Furlong et al., 2006). En humanos, la actividad enzimática de PON1 puede variar de un individuo a otro. Este hecho se debe a la presencia de una cantidad considerable de polimorfismos en el gen PON1 tanto en la región codificante como en la región reguladora (Furlong et al., 2005). Al parecer, estos polimorfismos explican las diferencias observadas en la actividad sérica de PON1 y la concentración sérica de esta enzima. En la región codificante, los polimorfismos en las posiciones 55 y 192 son los más comunes en los individuos, y afectan la eficiencia de la hidrólisis de los metabolitos de los pesticidas. El polimorfismo en la posición 192 genera dos isoformas, Q y R, dependiendo de la presencia de un residuo de glutamina o arginina, respectivamente. El producto del alelo R es más eficaz que el Q para hidrolizar el paraoxón; sin embargo, ocurre lo contrario con la hidrólisis del somán y el sarín (Furlong et al., 2006). El polimorfismo en la posición 55 también genera dos isoformas, L y M (leucina o metionina en esa posición), que es en parte responsable de los niveles séricos diferenciales observados en PON1 humana. Los individuos que poseen el alelo M presentan niveles más bajos de esta enzima en sangre.

Los polimorfismos en la región reguladora en la posición -108 afectan el nivel de expresión del gen PON1. Dos alelos están asociados a esta posición, ya sea un nucleótido C o T. La presencia del alelo C produce un nivel dos veces mayor de PON1 en suero (Costa et al., 2003; Furlong et al., 2005). Aparentemente, el polimorfismo -108 se encuentra dentro de un sitio de unión para Sp1, un factor de transcripción ubicuo común en genes sin TATA como PON1.

Anteriormente se informaron variaciones en las frecuencias alélicas para cada polimorfismo de PON1 en diferentes poblaciones (Castaño et al., 2006; Rojas et al., 2005; Brophy et al., 2001; Holland et al., 2006). Sin embargo, existen muy pocos estudios sobre la distribución de estos polimorfismos en poblaciones de América del Sur y Centroamérica a pesar de que el uso de pesticidas organofosforados es una práctica común en la región.

Teniendo en cuenta los problemas de intoxicación por plaguicidas organofosforados, además de que las enfermedades cardíacas son una de las más frecuentes causas de muerte en Panamá, consideramos importante determinar la distribución de los polimorfismos de PON1 en las posiciones -108, 55 y 192 en una muestra de la población panameña.

Materiales y métodos

Sujetos

La muestra de población estuvo compuesta por 187 individuos no emparentados (hombres y mujeres adultos) de cuatro grupos raciales diferentes: 50 panameños caucásicos, 40 panameños negros, 49 mestizos y 48 amerindios ngöbe-buglé. Sólo se consideraron individuos que mostraran el fenotipo representativo de cada grupo racial. El criterio adicional fue que ambos padres tuvieran el mismo fenotipo, por lo que no había duda sobre la principal contribución genética ancestral. Se recogieron muestras de sangre mediante venopunción en tubos de 5 ml que contenían Na-EDTA. Cada muestra fue convenientemente identificada y guardada en una caja refrigerada con hielo hasta su traslado al laboratorio, en un lapso no mayor a una hora. Esta investigación fue aprobada por el Comité Nacional de Investigación en Bioética (Instituto Memorial Gorgas, Panamá).

Genotipo PON1

Las muestras de sangre se centrifugaron a 4000 rpm en una centrifuga refrigerada para separar los glóbulos rojos y blancos del plasma. El plasma se utilizó inmediatamente para determinar las actividades de PON1 en sujetos caucásicos y negros-panameños. El ADN se aisló de la fracción celular utilizando el kit de purificación de ADN Puregene (Gentra Systems Inc., Minneapolis, Minnesota, EE. UU.) de acuerdo con las instrucciones de este fabricante, o siguiendo el método descrito por Sambrook & Russel (2001) y almacenado en $-20\text{ }^{\circ}\text{C}$ hasta el análisis polimórfico. El genotipo PON1 en las posiciones -108, 55 y 192 se determinó mediante reacción en cadena de la polimerasa seguida de digestión de restricción específica del polimorfismo y electroforesis. La amplificación se realizó utilizando 100-200 ng de ADN, 1 U de Taq polimerasa (Promega) y los cebadores (Integrated DNA technologies), en un volumen final de 25 μL (Brophy et al., 2001; Campo et al., 2004).

Los cebadores GACCGCAAGCCACGCCTTTCTGTGCACC y TGAAAGACT-TAAACTGCCAGTC se utilizaron para amplificar la región polimórfica PON1 -108. La región PON1 55 se amplificó con los cebadores CCTGCAATAATATGAAAACAACCTG y TGAAAGACTT AACTGCCAGTC. Los cebadores TATTGTTGCTGTGGGACCTGAG y GACATACTTGCCATCGGGTGAA se utilizaron para amplificar la región polimórfica PON1 192. Se utilizó un ciclador térmico GeneAmp 2400 (Perkin-Elmer) para amplificar las re-

giones variantes bajo las siguientes condiciones de PCR: temperatura de desnaturalización inicial de 960 C durante 5 min, temperatura de hibridación de 630 C durante 1 min (excepto para la variante 192 Q/R, donde se utilizó una temperatura de 610 C), temperatura de extensión de 720 C por 1 min, y luego 940 C por 1 min. Para cada determinación se alcanzó una temperatura de extensión final de 720 C durante 10 min. Estas condiciones se repitieron durante 25, 30 y 30 ciclos para los polimorfismos PON1 -108, 55 y 192, respectivamente. Se digirieron alícuotas de 15 mL de los productos amplificados a 370 C durante 16 horas usando 1 U de las siguientes endonucleasas de restricción (New England Biolabs): BstUI para -108 C/T, NlaIII para 55 L/M, y AlwI para el polimorfismo PON1 192 Q/R.

La presencia del alelo -108C da como resultado bandas digeridas de 67 pb y 52 pb (en lugar de una banda no digerida de 119 pb, que indica la presencia de un alelo T). Las bandas digeridas (106 pb y 66 pb) indicaron la presencia del alelo M, mientras que la ausencia de digestión indicó la presencia del alelo L. El alelo R se determinó por la presencia de dos bandas; 172 pb y 66 pb y una banda no digerida de 238 pb indicaron la presencia del alelo Q (Figura 1).

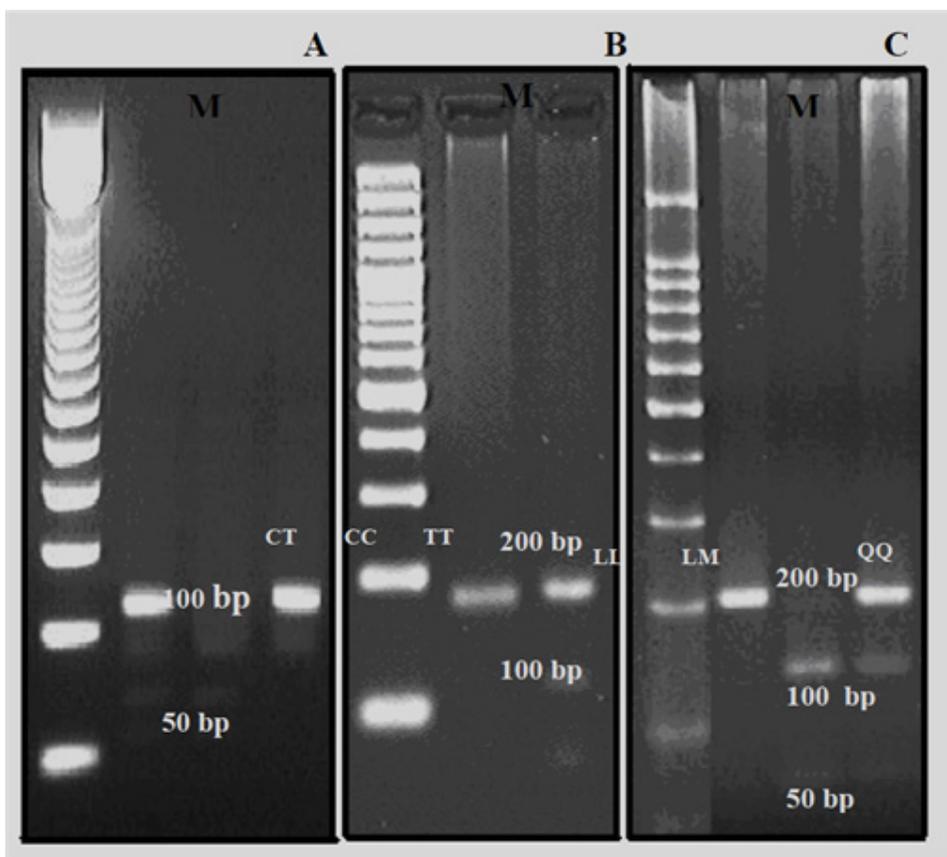


Figura 1. Geles de agarosa que muestran los polimorfismos de PON1 en las posiciones -108, 55 y 192. PON1 -108 genotipos (A): M, marcador molecular (Escalera de ADN de 100 pb, Fermentas); CT (bandas de 119, 67 y 52 pb); CC (bandas de 67 y 52 pb) y TT (banda de 119 pb). PON1 55 genotipos (B): M, marcador molecular (Ladder de 100 pb, Invitrogen); LL (banda de 172 pb); LM (bandas de 172, 106 y 66 pb); MM (datos no mostrados). PON1 192 genotipos (C): M, marcador molecular (Escalera de ADN de 100 pb, Fermentas); QQ (banda de 238 pb); RR (bandas de 172 y 66 pb); y QR (bandas de 238, 172 y 66 pb).

Actividad de PON1

Las actividades enzimáticas de PON1 solo se analizaron en suero de individuos caucásicos y panameños negros utilizando fenilacetato y diazoxón como sustratos (Rosenblat et al., 2003; Leary & Edwards, 2005). No hubo muestras de suero de los grupos Ngöbe y Mestizo. La actividad de arilésterasa se midió mediante la producción de fenol a partir de la hidrólisis de fenilacetato 10 mM usando Tris-HCl 20 mM, pH 8,0, CaCl₂ 0,9 mM, en un

volumen total de 3 ml. Cada determinación se realizó por duplicado en un espectrofotómetro Shimatzu colocando el aparato en modo cinético a 25° C. La reacción se inició añadiendo el sustrato.

La actividad de la diazoxonasa se midió mediante la producción de pirimidinol, a partir de la hidrólisis de diazoxón 1 mM, usando Tris-HCl 50 mM, pH 8,0, CaCl₂ 1 mM, en un volumen final de 3 ml. Cada medición también se realizó por duplicado en el aparato bajo el mismo modo cinético a 25° C.

Análisis estadístico

Los datos fueron analizados mediante el programa Epi Info™ 3.3.2. Las frecuencias alélicas y genotípicas de PON1 -108, 55 y 192 se obtuvieron por recuento directo. Se utilizó el test de chi-cuadrado para evaluar la concordancia de las frecuencias genotípicas con las expectativas de Hardy-Weinberg. Las diferencias de las actividades de PON1 entre genotipos se evaluaron utilizando las capacidades del programa EPIDAT mediante la prueba de Pearson con correcciones de Yates.

Resultados

Frecuencias alélicas y genotípicas de PON 1

Las frecuencias de genotipo y alelo para los polimorfismos de PON1 en las posiciones -108, 55 y 192, en cuatro de los principales grupos raciales panameños, se resumen en la Tabla 1. Se encontró una buena concordancia entre las frecuencias de genotipo observadas y esperadas según el equilibrio de Hardy-Weinberg para toda la población. Sin embargo, cuando la población se segregó en grupos raciales, se observó una desviación del equilibrio de Hardy-Weinberg (HWE) en individuos caucásicos para el genotipo PON1 192 ($\chi^2=4.6$, $p = 0,032$).

PON1 -108	Whole population (n=187)	Caucasian (n=50)	Mestizo (n=49)	Black (n=40)	Ngöbé (n=48)
C	0.655	0.730 (73)	0.724 (71)	0.787 (63)	0.396 (38)
T	0.344	0.270 (27)	0.275 (27)	0.213 (17)	0.604 (58)
CC	0.428	0.500 (25)	0.510 (25)	0.600 (24)	0.125 (6)
CT	0.428	0.460 (23)	0.428 (21)	0.375 (15)	0.542 (26)
TT	0.117	0.040 (2)	0.061 (3)	0.025 (1)	0.333 (16)
χ^2	0.0136	1.360	0.296	0.606	0.839
PON1 55					
L	0.796	0.750 (75)	0.796 (78)	0.775 (62)	0.864 (83)
M	0.203	0.250 (25)	0.204 (20)	0.225 (18)	0.135 (13)
LL	0.631	0.600 (30)	0.592 (29)	0.600 (24)	0.729 (35)
LM	0.331	0.300 (15)	0.408 (20)	0.350 (14)	0.270 (13)
MM	0.037	0.100 (5)	0 (0)	0.050 (2)	0 (0)
χ^2	0.293	2.005	3.275	0.002	1.290
PON1 192					
Q	0.581(217)	0.610 (61)	0.561 (55)	0.387 (31)	0.729 (70)
R	0.419(157)	0.390 (39)	0.438 (43)	0.613 (49)	0.270 (26)
QQ	0.331(62)	0.300 (15)	0.326 (16)	0.125 (5)	0.542 (26)
QR	0.497(93)	0.620 (31)	0.469 (23)	0.525 (21)	0.375 (18)
RR	0.171(32)	0.080 (4)	0.204 (10)	0.35 (14)	0.083 (4)
χ^2	0.212	4.6	0.137	0.482	0.182

Tabla 1. Genotipos y frecuencias alélicas de los polimorfismos de PON1 en las posiciones -108, 55 y 192 entre 187 sujetos panameños correspondientes a los cuatro grupos raciales principales. Los paréntesis representan el número de individuos o cantidades de alelos. HWE se calculó usando χ^2 .

Los alelos PON1 más frecuentes en las posiciones -108, 55 y 192 detectados en toda la población fueron C (0,655), L (0,796) y Q(0,581), respectivamente. Asimismo, los genotipos más frecuentes fueron CT (0,428), LL (0,631) y QR (0,497). El alelo más frecuente en la posición 55 en cada grupo racial fue L. En la posición 192, Q fue el más frecuente en caucásicos, mestizos y ngöbes, pero no en negros. En caucásicos, mestizos y negros, el alelo C en la posición -108 mostró una frecuencia mayor que el alelo T, en contraste con los ngöbes, donde T fue el alelo más frecuente.

Las frecuencias alélicas entre grupos raciales mostraron diferencias estadísticas para PON1 192 en caucásicos - negros ($p=0,0048$), mestizos - ne-

gros ($p=0.031$), Mestizos-Ngöbe ($p=0.0249$) y Negros-Ngöbe ($p<0.0001$). En la posición 55, no se observaron diferencias estadísticas entre razas. La frecuencia del alelo en la posición -108 en Ngöbes fue estadísticamente diferente de la de los otros grupos ($p < 0,0001$) (Tabla 2).

Tabla 2. Valores de p obtenidos de la comparación por pares de frecuencias alélicas entre diferentes grupos raciales. Los valores de p corresponden a chi cuadrado calculado mediante la prueba de Pearson con correcciones de Yates.

		Caucasian	Black	Mestizo	Ngöbé
PON1 -108	Caucasian	--			
	Black	0.4731	--		
	Mestizo	0.9422	0.4268	--	
	Ngöbé	<0.0001	<0.0001	<0.0001	--
PON1 55	Caucasian	--			
	Black	0.8298	--		
	Mestizo	0.4588	0.8769	--	
	Ngöbé	0.0698	0.1735	0.2941	--
PON1 192	Caucasian	--			
	Black	0.0048	--		
	Mestizo	0.5807	0.0310	--	
	Ngöbé	0.1164	<0.0001	0.0249	--

Las frecuencias alélicas para la población panameña obtenidas en este estudio se compararon con las frecuencias alélicas para otras poblaciones reportadas previamente (Tabla 3).

Tabla 3. Frecuencias alélicas de los polimorfismos PON1 -108, 55 y 192 en diferentes poblaciones del mundo en comparación con las encontradas en la población panameña.

Population	n	PON1 -108			PON1 55			PON1 192		
		C	T	P*	L	M	P*	Q	R	P*
AMERICA										
Panamanian ^a	187	0.66	0.34		0.80	0.20		0.58	0.42	
African-Americans ^b	117	0.85	0.15	< 0.0001				0.37	0.63	< 0.0001
Caribbean-Hispanics ^b	203	0.65	0.35	0.9608				0.54	0.46	0.2826
Costa Ricans	518							0.757	0.243	< 0.0001
Peruvians ^b	89	0.61	0.39	0.0501				0.539	0.461	0.4454
Mexicans ^c	214	0.45	0.55	0.0650	0.84	0.16	0.1580	0.510	0.490	0.0526
Mexican-Mestizos ^b	182							0.522	0.478	0.1294
Canadians ^b	865	0.48	0.52	0.1898				--	--	--
African-Brazilians ^b	70							0.471	0.529	0.0351
European-Brazilians ^b	101							0.693	0.307	0.0101
Caucasian-Americans ^b	82	0.38	0.62	0.0036				0.730	0.270	0.0012
Americans-USA ^b	376	0.50	0.50	0.5916				0.728	0.272	< 0.0001
Washington ^d	376				0.64	0.36	0.0001			
California ^e	260				0.82	0.18	0.5230			
Cayapa Indians ^b (EC)	83							0.211	0.789	< 0.0001
Chileans ^b NS	195							0.569	0.431	0.8152
Chileans ^b ES	129							0.663	0.337	0.0441
EUROPE										
Northern Irish ^b	170							0.712	0.288	0.0003
English ^b	282							0.710	0.290	< 0.0001
Finnish ^b	169							0.737	0.263	< 0.0001
Dutch ^b	250							0.716	0.284	< 0.0001
Germans ^b	2784							0.718	0.282	< 0.0001
Italians ^b (Sardinia)	161							0.752	0.248	< 0.0001
French ^b	125							0.765	0.235	< 0.0001
Spanish ^b	141	0.38	0.62	0.0005				0.700	0.300	0.0024
Turkish ^b	381							0.692	0.308	0.0003
ASIA										
Japanese ^b	132	0.48	0.52	0.3886				0.402	0.591	< 0.0001
Chinese ^b	475	--	--	--				0.352	0.648	< 0.0001
Thais ^b	202	0.25	0.75	< 0.0001				0.710	0.290	0.0114
Indians ^b	165							0.67	0.33	0.0173
Indians ^b ND	80							0.82	0.17	< 0.0001
AFRICA										
Ethiopians ^b	169							0.592	0.401	0.7954
Beninese ^b	98							0.388	0.612	< 0.0001

p* calculated using Pearson test with Yates corrections

^aThis study. ^bCastaño et al. 2006, ^cRojas et al. 2005, ^dBrophy et al. 2001, ^eHolland et al. 2006.

Haplotipos y actividad de arilesterasa y diazoxonasa

Las actividades de arilesterasa y diazoxonasa para los haplotipos PON1 en negros y caucásicos se presentan en la Tabla 4. Las actividades de PON1 no revelaron diferencias significativas ($p > 0,05$) entre grupos para los mismos haplotipos. Los individuos con el haplotipo -108CT/55LM/192QR tuvieron la actividad de arilesterasa más alta (153,25 U/mL), mientras que aquellos con el haplotipo 108TT/55LM/192QR exhibieron las actividades de arilesterasa (48,1 U/mL) y diazoxonasa (2,38 U/mL) más bajas. La mayor actividad de diazoxonasa se asoció con el haplotipo -108TT/55LM/192QQ (13,56 U/mL). El haplotipo más frecuente en toda la población fue -108CC/55LL/192QR. En los negros, este es el segundo haplotipo más frecuente. Las actividades de arilesterasa para este último haplotipo fueron 116,16 U/ml y 146,30 U/ml en negros y caucásicos, respectivamente. Las actividades de diazoxonasa para el mismo último haplotipo fueron 10,09 U/mL y 11,25 U/mL, en negros y caucásicos, respectivamente.

Tabla 4. Actividades de arilesterasa y diazoxonasa según diferentes genotipos de PON1.

Whole Population (n= 90) n	Genotype at Position			Black (n=40)			Caucasian (n=50)		
	-108	55	192	n	Arylesterase activity* mean (U/mL) ± SD	Diazoxonase activity* mean (U/mL) ± SD	n	Arylesterase activity* mean (U/mL) ± SD	Diazoxonase activity* mean (U/mL) ± SD
18	CC	LL	QR	6	116.16±33.75	10.09±3.97	12	146.39±39.09	11.25±2.88
12	CC	LM	QR	7	119.53±26.87	9.03±2.05	5	113.40±18.49	8.55±2.03
11	CT	LL	QR	4	137.11±34.41	10.75±2.67	7	124.05±27.13	10.52±3.34
7	CT	LL	RR	5	145.97±16.27	9.44±2.05	2	137.31±1.56	8.67±1.59
7	CC	LL	RR	5	134.38 ± 40.67	9.03±2.18	2	151.18±25.30	9.5±2.85
6	CT	LL	QQ	1	*95.97	*9.67	5	136.45± 20.88	13.4±2.72
6	CT	LM	QR	2	153.25±1.74	13.39±1.08	4	130.62±26.94	9.09±2.56
5	CC	LL	QQ	3	132.09±8.90	11.86±1.83	2	127.54± 17.74	12.53±1.45
3	CT	LM	RR	3	138.07±4.96	9.76±1.25	0	--	--
2	TT	LM	QR	1	*93.83	*6.81	1	*48.10	*2.38
2	CC	MM	QR	1	*126.64	*8.79	1	*129.66	*8.67
2	CC	MM	QQ	1	*118.56	*9.01	1	ND	--
2	CC	LM	QQ	0	--	--	2	127.00±28.34	12.08±3.63
2	CT	MM	QQ	0	--	--	2	106.80±20.95	10.11±2.85
2	CT	LM	QQ	0	--	--	2	118.08±12.36	11.79±1.66
1	CC	LM	RR	1	*149.65	*11.80	0	--	--
1	TT	LM	QQ	0	--	--	1	*147.22	*13.56
1	CT	MM	QR	0	--	--	1	*138.7	*10.71

$p > 0,005$ evaluado mediante prueba t de Student. * No se calcularon los valores medios ni la DE (desviación estándar).

Discusión

Hasta donde sabemos, este es el primer estudio que informa las frecuencias de polimorfismos de PON1 en las posiciones -108, 55 y 192 en la población panameña. La frecuencia del alelo PON1 -108C en toda la población panameña fue considerablemente mayor que la del alelo T. La frecuencia del alelo -108C para los panameños (0,66) es mayor que la de cualquier otra población, excepto para los afroamericanos (0,850), y muy similar a la de los hispanos caribeños (0,65) y los peruanos (0,61). Este resultado puede indicar que los panameños, así como los hispanos caribeños y los peruanos poseen un nivel plasmático relativamente más alto de PON1, lo que significa que estas poblaciones son menos susceptibles a la toxicidad de los pesticidas organofosforados. Las similitudes observadas en las frecuencias del alelo C entre panameños y otras poblaciones se explicarían como consecuencia de una contribución genética ancestral similar en el período de colonización. Las mayores diferencias de este polimorfismo se observaron cuando lo comparamos con las poblaciones tailandesa, caucásica americana y española, afroamericana e hispana caribeña ($p < 0,005$) (Tabla 3).

En cuanto al polimorfismo PON1 55, el alelo L tuvo la mayor frecuencia en la población panameña (0.80), similar a los mexicanos (0.84) y californianos (0.82). Por el contrario, la frecuencia de este alelo fue significativamente mayor que la observada en los residentes de Washington (Holland et al., 2006). Debido al hecho de que el alelo L se ha asociado con niveles más altos de PON1 en plasma, esperamos que la población panameña esté parcialmente protegida contra las enfermedades cardiovasculares. Curiosamente, cuando comparamos las frecuencias del alelo L y el porcentaje de mortalidad causada por cardiopatía isquémica, los países con menores frecuencias del alelo L exhiben un mayor porcentaje de mortalidad causada por esta enfermedad. Al respecto, Estados Unidos tiene una frecuencia del alelo L de 0.64, teniendo un 21 % de mortalidad; Panamá, con una frecuencia de 0.80 para el mismo alelo, presenta el 12 % de mortalidad, y México, cuya frecuencia del alelo L es de 0.84, la causa de mortalidad atribuida a la cardiopatía isquémica es del 11 % (Organización Mundial de la Salud. Mortality Country Fact Hoja, 2006).

Como se indica en la Tabla 3, la frecuencia observada para el alelo Q de PON1 192 en panameños fue ligeramente mayor (0,58) que en los hispanos caribeños (0,54) (Chen et al., 2003), peruanos (0,54) (Cataño et al., 2006), mexicanos (0,51) (Rojas-García et al., 2005), mestizos mexicanos

(0,52) (Gamboa et al., 2006) y chilenos NS (0,57) (Acuña et al., 2004). Los panameños tienen una frecuencia relativamente baja (0,419) del alelo R que se asocia a una alta actividad de arilesterasa. La actividad arilesterasa de PON1 solo proporciona una estimación de la sensibilidad a los pesticidas organofosforados debido al hecho de que la actividad arilesterasa se correlaciona con la hidrólisis del paraoxón, que es el catabolito del paratió (pesticida organofosforado). De acuerdo con estos resultados, se espera que la población panameña sea más susceptible a la intoxicación por compuestos organofosforados. Cuando comparamos las actividades de arilesterasa y diazoxonasa entre panameños negros y caucásicos, no se observaron diferencias significativas para las actividades de ambas enzimas. Este resultado podría explicarse en términos de la complejidad genética de la población panameña, que es resultado de una mezcla no sólo de estas dos razas, sino también del aporte nativo-panameño.

Como se mencionó anteriormente, la población panameña es el resultado de una mezcla de varios grupos raciales con importantes contribuciones de los nativos americanos, negros y españoles. De acuerdo con este hecho, esperábamos algunas similitudes entre los grupos raciales panameños y los grupos ancestrales correspondientes. Las frecuencias de los alelos PON1 192 y PON1 55 de los panameños negros se parecen más a las informadas para otras poblaciones negras, como los afroamericanos, los afrobrasileños y los afrobenineses, que las de las poblaciones no negras. La población negra panameña es descendiente de la población negra africana que llegó al istmo como esclavos en el período de la colonización o como trabajadores de las Antillas durante la construcción del Canal de Panamá.

El grupo caucásico panameño tenía una frecuencia del alelo PON1 192 más similar a la de los españoles que las poblaciones negras africanas y americanas, incluido el panameño negro. Sin embargo, la frecuencia de los alelos L y C es bastante diferente a la observada en otros grupos caucásicos (Cataño et al., 2006). Inesperadamente, las frecuencias del alelo PON1 192 para Ngöbe fueron muy diferentes de las reportadas para los indios Cayapa, aunque ambos grupos tienen el mismo origen lingüístico chibcha (Kolmand y Bermingham, 1997).

Encontramos que el grupo nativo panameño Ngöbe posee altas frecuencias de los alelos T y Q, lo que los hace más susceptibles a algunas intoxicaciones por organofosforados. Irónicamente, los Ngöbe trabajan principalmente como agrícolas en las tierras altas de Panamá, donde el uso de pesticidas organofosforados es bastante elevado. Es bien sabido que la exposición a altas dosis de estos compuestos tiene efectos profundos en el sistema nervioso central, y evidencia reciente sugiere que la exposición

crónica a bajos niveles puede afectar el desarrollo neurológico. Este último efecto es importante en mujeres embarazadas, recién nacidos e incluso en bebés, por lo que se deben tener consideraciones especiales con este grupo racial.

La mayor actividad arilesterasa de PON1 se observó, por separado, en individuos con los genotipos -108CC, 55LL y 192 RR. Resultados similares obtuvieron Rojas-García et al. (2005) en una población mexicana, en contraste con los resultados obtenidos por Brophy et al. (2001), en los que los individuos con genotipo PON1 192QQ se asociaron con la mayor actividad de arilesterasa de PON1. Otros investigadores no han encontrado relación entre estos genotipos y la actividad arilesterasa de PON1 en poblaciones caucásicas, afroamericanas y caribeñas (Chen et al., 2003).

Las actividades de arilesterasa y diazoxonasa para los haplotipos PON1 en panameños negros y caucásicos se presentan en la Tabla 4. La actividad arilesterasa de PON1, en función de diferentes haplotipos, reveló que la actividad de arilesterasa más alta (151,18 U/mL) se observó en panameños caucásicos con el - Haplotipo 108CC/55LL/192RR. Este haplotipo generalmente se asocia con una alta actividad de arilesterasa (Rojas-García et al., 2004). Sin embargo, estos valores de actividad son inferiores a los reportados para el mismo haplotipo en mexicanos (Rojas-García et al., 2004). Estas diferencias se explicarían como consecuencia de modificaciones en el protocolo utilizado para medir la actividad de la arilesterasa. En los panameños negros, el haplotipo asociado con la mayor actividad de arilesterasa (149,65 U/mL) fue -108CC/55LM/192RR. Aunque la actividad de arilesterasa para el haplotipo -108CC/55LL/192RR en panameños negros fue menor que la del haplotipo -108CT/55LM1/92QR, los valores obtenidos pueden no ser confiables porque solo dos individuos en nuestro estudio mostraron este haplotipo y también variaciones en las actividades han sido reportadas para individuos con el mismo haplotipo (Rojas-García et al., 2004).

En resumen, los resultados presentados en este estudio representan la única información disponible hasta el momento sobre las frecuencias de los polimorfismos asociados al gen PON1 en la población panameña. Las frecuencias alélicas obtenidas se asemejan a las reportadas para otras poblaciones latinas; sin embargo, los Nögbé Bugles, población amerindia panameña, mostraron una baja frecuencia para el genotipo -108CC, lo que significa que son más susceptibles al envenenamiento con pesticidas organofosforados. Además de la contribución de nuestro estudio, reconocemos la necesidad de realizar estudios más extensos sobre la relación entre la variabilidad de PON1 y las diferencias en la susceptibilidad al envenenamiento por pesticidas y el desarrollo de enfermedades vasculares.

Literatura citada

1. Acuña, M., L. Eaton & L. Cifuentes. 2004. Genetic variants of the paraoxonase (PON1 and PON2) in the Chilean population. *Hum Biol.* 76: 299-305.
2. Brophy, V., R.L. Jampsa, J.B. Clendenning, L.A. Mckinstry & C.E. Furlong. 2001. Effects of 5' regulatory-region polymorphisms on paraoxonase-gene (PON1) expression. *Am. J. Hum. Genet.* 68:1428-1436.
3. Brophy, V., M. Hastings, J. Clendenning, R. Richter, G. Jarvik & C. Furlong. 2001. Polymorphisms in the human paraoxonase (PON1) promoter. *Pharmacogenetics.* 11:77-84.
4. Cataño, H., J. Cueva, A. Cardenas, V. Izaguirre, A. Zavaleta, E. Carranza & A. Hernández. 2006. Distribution of paraoxonase-1 gene polymorphisms and enzyme activity in a Peruvian population. *Environmental and molecular mutagenesis.* 47: 699-706.
5. Chen, J., M. Kumar, W. Chan, G. Berkowitz & J. Wetmur. 2003. Increased influence of genetic variation on PON1 activity in neonates. *Environ Health Perspect.* 111: 1403-1409.
6. Costa, L., T. Cole, G. Jarvik & C. Furlong. 2003. Functional genomics of the paraoxonase (PON1): effects on pesticide sensitivity, Cardiovascular and Drug Metabolism. *An. Rev. Med.* 54: 371 - 392.
7. Durrington, P., B. Mackness & M. Mackness. 2001. Paraoxonase and atherosclerosis. *Arterioscler Thromb. Vasc. Biol.* 372: 473-480.
8. Furlong, C., T. Cole, G. Jarvik, C. Pettan, G. Geiss, R. Richter, D. Shih, A. Tward, A. Lulis & L. Costa. 2005. Role of Paraoxonase (PON1) status in pesticides sensitivity: genetic and temporal determinants. *Neuro Toxicology.* 26: 651-659.
9. Furlong, C., N. Holland, R. Richter, A. Bradman, A. Ho & B. Eskenazi. 2006. PON1 status of farmworker mothers and children as a predictor of organophosphate sensitivity. *Pharmacogenetic and Genomics.* 16: 183-190.
10. Gamboa, R., J. Zamora, J. Rodríguez-Perez, J. Frogoso, G. Cardoso, C. Posadas-Romero & G. Vargas-Alarcón. 2006. Distribution of paraoxonase PON1 gene polymorphisms in Mexican populations. Its role in the lipid profile. *Exp. Mol. Pathol.* 80: 85-90.
11. Holland, N., C. Furlong, M. Bastaki, R. Richter, A. Bradman, K. Huen, K. Beckman & B. Eskenazi. 2006. Paraoxonase polymorphisms, haplotype, and enzyme activity in latino mothers and newborns. *Environmental Health. Perspectives.* 114: 985-991.
12. C.J. Kolman & E. Bermingham. 1997. Mitochondrial and Nuclear DNA diversity in the Choco and Chibchan Amerinds of Panama. *Genetics.* 147: 1289-1302.
13. Leary, K. & R. Edward. 2005. Genetic and other sources of variation in the activity of serum paraoxonase/diazoxonase in human: consequences for risk from exposure to diazinon. *Pharmacogenetics and Genomics.* 15: 51-60.
14. Rojas-García, A. E., M.J. Solís-Heredia, B. Piña-Guzmán, L. Vega, L. López-Carrillo & L. Quintanilla-Vega. 2005. Genetic polymorphisms and activity of PON1 in a Mexican population. *Toxicology and Applied Pharmacology.* 205:282-289.

15. Rosemlat, M., O. Grunfeld, T. Hayek & M. Aviram. 2003. Serum paraoxonase activity and extent of lipid peroxidation are not affected by increased levels of human apolipoprotein A-I: studies in transgenic mice. *Clin. Chem. Lab. Med.* 40: 9-14.
16. Sambrook, J. & D. Russell. 2001. *Molecular cloning a laboratory manual*. Third edition, Volumen 1 CSHL Press, U.S.A.
17. World Health Organization. Mortality Country Fact Sheet, 2006. <http://www.who.int/countries/en/>

C A P Í T U L O 6

Determinación de alelos
HLA-DRB1 en pacientes
panameños con artritis
reumatoidea severa

*Diana Mora, Orlando Montenegro
y Carlos Ramos Delgado*

RESUMEN

Se determinó la presencia de alelos HLA-DRB1*0101 y HLA-DRB1*04 con epítotope reumatoideo en una muestra de la población panameña, formada por 50 individuos con diagnóstico positivo de artritis reumatoidea (AR) severa y 35 individuos no afectados utilizados como controles. La determinación de los alelos se efectuó mediante reacción en cadena de la polimerasa con cebadores de secuencia específica (PCR-SSP). La frecuencia de alelos con epítotope reumatoideo SE (+) en individuos con AR severa fue de 68 % y significativamente mayor ($p < 0.005$) que la observada en los individuos controles (34%). El riesgo relativo (R.R.) asociado a la presencia de estos alelos fue de 6.00. El alelo con SE (+) más frecuente en los individuos con AR severa fue HLA-DRB1*0101, con una frecuencia de 38% *versus* 17% en el grupo control. No se observó en pacientes panameños asociación de ninguno de los alelos HLA-DRB1*04 con la AR severa cuando se consideran de manera independiente. El único alelo que de manera independiente mostró asociación con la AR severa fue HLA-DRB1*0101, $p < 0.05$ con un R.R. de 2.96.

Palabras claves

Artritis rematoidea, HLA-DRB1 *0101, Riesgo relativo, población panameña

Abstract

The presence of HLA-DRB1*0101 and HLA-DRB1*04 alleles with rheumatoid epitope was determined in a sample of the Panamanian population, consisting of 50 individuals with a positive diagnosis of severe rheumatoid arthritis (RA) and 35 unaffected individuals used as controls. Allele determination was performed by polymerase chain reaction with sequence-specific primers (PCR-SSP). The frequency of alleles with rheumatoid epitope SE (+) in individuals with severe RA was 68 % and significantly higher ($p < 0.005$) than that observed in control individuals (34%). The relative risk (R.R.) associated with these alleles was 6.00. The most frequent allele with SE (+) in individuals with severe RA was HLA-DRB1*0101, with a frequency of 38% *versus* 17% in the control group. No association of any of the HLA-DRB1*04 alleles with severe RA was observed in Panamanian patients when considered independently. The only allele independently associated with severe RA was HLA-DRB1*0101, $p < 0.05$, with an R.R. of 2.96.

Keywords

Arthritis rheumatoid, HLA-DRB1 *0101, Relative risk, Panamanian population.

Introducción

La artritis reumatoidea (AR) es una enfermedad multifactorial compleja, en la que factores innatos de susceptibilidad desempeñan un papel crucial (Maini et al., 1995). La enfermedad afecta principalmente las articulaciones y los tejidos circundantes, pero también puede afectar otros sistemas de órganos. Las articulaciones de las muñecas, los pies y los dedos de las manos son generalmente las más afectadas. La deformación de las articulaciones como resultado de la destrucción progresiva del cartílago, erosiones óseas, y la inflamación y ruptura de los tendones es característica. Esta enfermedad afecta alrededor del 3% de la población adulta con predominio en mujeres en una proporción de 3:1 y una incidencia máxima entre los 25-50 años (van Schaaardenburg & Breedveld 1994).

A pesar que la causa de la AR se desconoce, se sugiere que la exposición a un agente ambiental, principalmente bacterias, puede contribuir al desarrollo de la enfermedad (Albani & Carson 1996; Krause et al., 1996). Existen además factores genéticos que contribuyen en 30-50% con el riesgo a desarrollar la enfermedad. Aproximadamente, un tercio de esos factores genéticos están en el complejo mayor de histocompatibilidad (MHC) y cuya localización es 6p21.3 (Kilding et al., 2004). Estudios previos demuestran asociación de los alelos HLA-DRB1 que codifican el “epítoto compartido” SE (+); ubicado en la posición 70-74, y cuya secuencia es (Q/R)(K/R)RRA y el riesgo a desarrollar AR (Ruíz-Morales et al., 2004; Bongi et al., 2004; Gorman & Criswell 2002). Recientemente se ha sugerido que factores adicionales localizados hacia la porción telomérica del MHC contribuyen también con el riesgo a desarrollar la enfermedad (Scholand et al., 2003; Zanelli et al., 2001).

La naturaleza multifactorial de la AR hace de ésta una enfermedad compleja en la que intervienen factores genéticos y no genéticos. Por lo tanto, es concebible que la interacción de otros genes contribuyan también con el riesgo o la severidad de la AR. Se han realizado estudios que demuestran diferencias en asociación a la enfermedad en diferentes grupos étnicos (Gorman & Criswell 2002; Listing et al., 2000; Hajeer et al., 2000; Wakitani et al., 1997). En adición, las manifestaciones clínicas como también la

frecuencia y el tipo de alelo asociado a AR varían de un grupo étnico a otro (Del Rincón et al., 2003; McDaniel et al., 1995). Estas diferencias justifican la necesidad de realizar estudios en diferentes poblaciones de manera que se pueda establecer una mejor relación entre la enfermedad y la heterogeneidad genética de cada población.

La población panameña es el resultado de una mezcla de diferentes grupos étnicos que han contribuido a la heterogeneidad genética actual característica de nuestra población, por lo que no es recomendable asumir que los alelos asociados a AR en otras poblaciones son los mismos en la población panameña. En nuestro país no se tiene información sobre la distribución de alelos HLA-DRB1 con SE (+) en pacientes con AR severa, lo cual implica que se desconoce qué alelos están asociados a esta condición y las frecuencias en que se presentan. En este trabajo se determinó que alelos HLA-DRB1 están presentes en pacientes panameños con AR severa, y si existe asociación entre la condición y la presencia de SE (+) y los alelos que exhiben asociación.

Materiales y métodos

Muestras

Se analizaron 85 muestras; 50 de pacientes del Servicio de Reumatología del Complejo Hospitalario Dr. Arnulfo Madrid con diagnóstico médico de artritis reumatoidea severa, y 35 donantes sanos del Banco de sangre de la misma institución; utilizados como controles. Se colectaron 4.5 mL de sangre venosa de cada individuo y se transfirieron a tubos con EDTA.

Extracción de ADN

Las muestras de sangre fueron centrifugadas a 3000 rpm por 15 minutos a 4 °C. El plasma resultante fue descartado y la capa rica en glóbulos blancos y plaquetas, transferida a un tubo y resuspendida en 4 mL de solución de lisis I (5 mM MgCl). La suspensión fue centrifugada a 3000 rpm por 15 minutos a 4 °C y el sobrenadante resultante descartado. Las células fueron una vez más resuspendidas en 4 mL de la misma solución y centrifugadas bajo las condiciones antes descritas. El sobrenadante fue descartado y el botón celular resuspendido en 4 mL de solución de lisis II (5 mM MgCl, 0.1 % Nonidet NP-40) y centrifugado a 3000 rpm por 15 minutos a 4 °C. El sobrenadante fue nuevamente descartado y el botón celular resuspendido

en 200 μ L de Tris-EDTA. Se añadieron 800 μ L de proteinasa K (0.3 mg/mL) y se incubó a 56 °C por 2 horas. Luego se procedió a realizar una extracción con igual volumen de fenol-cloroformo- alcohol isoamílico (25:24:1). La suspensión fue centrifugada a 14,000 rpm por 5 minutos, el sobrenadante transferido a un tubo Eppendorf y el ADN precipitado mediante la adición de 300 mM NaCl y dos volúmenes de etanol absoluto. El ADN precipitado fue recuperado mediante centrifugación a 14,000 rpm, lavado con etanol al 70% y resuspendido en TE 1X (10 mM Tris, 1 mM EDTA pH 8.0).

PCR-SSP

La determinación de los alelos HLA-DRB1 presentes en cada individuo se realizó mediante el procedimiento descrito por Lee et al., 1996. El procedimiento utilizado permite la tipificación los alelos HLA-DRB1*0101, HLA-DRB1*0402, HLA-DRB1*0404, HLA-DRB1*0406, HLA-DRB1*0407. El procedimiento no discrimina entre los alelos HLA-DRB1*0405 y 0409 (HLA-DRB1*0405/*0409); HLA-DRB1*0410 y *0411 (HLA-DRB1*0410/*0411); HLA-DRB1*0403 y *0406 (HLA-DRB1*0403/*0406) y HLA-DRB1*0401, *0408 y *0409 (HLA-DRB1*0401/*0408/*0409). Se prepararon tres mezclas de reacción diferentes (1, 2 y 3). Cada mezcla contenía 2 o más cebadores para la amplificación específica de diferentes alelos (Cuadro 1). Se realizaron tres reacciones de amplificación con cada muestra; cada una con una mezcla diferente de cebadores. Cada mezcla contenía 0.4 μ M de cada cebador de secuencia específica, 0.08 μ M de cada uno de los cebadores utilizados para la amplificación del control interno, 56 mM KCl, 1.7 mM MgCl₂, 11 mM Tris-HCl, pH 8.3, 0.0011% gelatina, 250 μ M de cada dNTPs. Se utilizaron en cada reacción 2 μ g de ADN y 0.5 U de Amplitaq (Perkin-Elmer). El volumen final de la reacción fue de 50 μ L. Las condiciones de amplificación fueron: 30 ciclos de 94 °C 20s, 59 °C 50s y 72 °C 20s. La tipificación de los alelos se realizó con base en el tamaño de los productos de amplificación observados con cada mezcla (Fig. 1).

Análisis estadístico

Se determinó la frecuencia de individuos con alelos HLA-DRB1 y se realizó la prueba de chi cuadrada para determinar la probable asociación utilizando una tabla de 2 x 2 (Steel & Torrie 1995) y la chi cuadrada corregida de Yates, 1934. Se determinó el riesgo relativo según Wolf (1955) para cada uno de los alelos con SE (+) de manera independiente y también combinada.

Cuadro 1. Alelos HLA-DRB1, productos esperados y cebadores en cada mezcla.			
	(pb)	Cebador 5'	Cebador 3'
Mezcla N° 1	210	GCGGGTGCGGTTGCTGGAA	TGCACTGTGAAGCTCTCAC
DRB1*0101			
Mezcla N° 2	111	GTTTCTTGGAGCAGGTAAACA	GCTGTCGAAGCGCACGG
DRB1*0406	171	GTTTCTTGGAGCAGGTAAACA	TGTTCCAGTACTCGGCGCT
DRB1*0405, 0409, 0410, 0411	222	GTTTCTTGGAGCAGGTAAACA	TCTGCAGTAGGTGTCCACCT
DRB1*0403, 0406, 0407	261	GTTTCTTGGAGCAGGTAAACA	CTGCACTGTGAAGCTCTCAC
DRB1*0401, 0405, 0407, 0408, 0409			
Mezcla N° 3	214	GTTTCTTGGAGCAGGTAAACA	GTCAACCGCGGCCCGCTC
DRB1*0402	261	GTTTCTTGGAGCAGGTAAAC	CTGCACTGTGAAGCTCTCCA
DRB1*0402, 0403, 0404, 0406			
Control Interno (C5, C3)	796	TGCCAAGTGGAGCACCCAA	GCATCTTGCTCTGTGCAGA

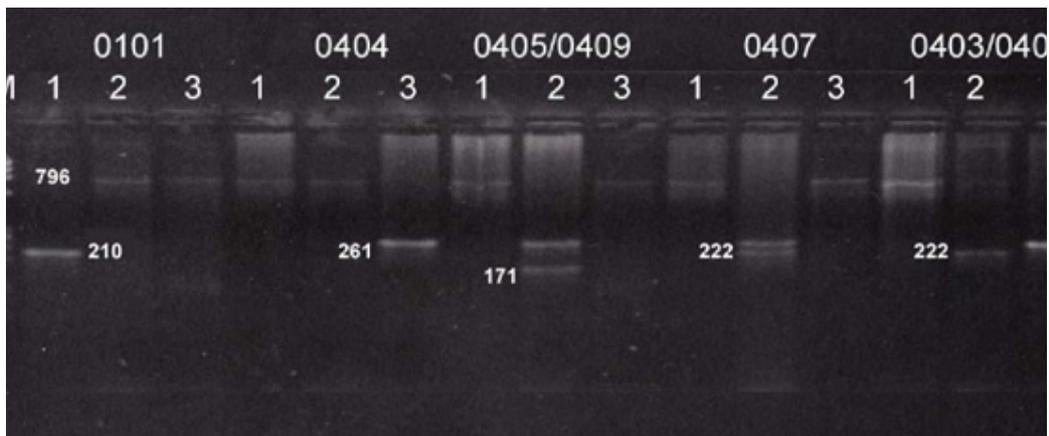


Fig. 1. Productos del PCR-SSP de algunos de los alelos HLA-DRB1 utilizados. M, marcador ϕ X174 Hae III. Se pueden apreciar los productos esperados de cada alelo con las diferentes mezclas de cebadores y una banda de 796 pb del control interno.

Resultados y discusión

La frecuencia de individuos con al menos un alelo HLA-DRB1 con SE (+) y el riesgo relativo para cada uno esos alelos con respecto al grupo control se muestra en el Cuadro 2. La frecuencia de individuos con alelos SE (+) (68% vs 34%, P no corregida < 0.005) fue significativamente mayor en pacientes con AR que en el grupo control. Frecuencias similares para la presencia de alelos con SE (+) fueron reportadas en pacientes japoneses (Toda et al., 1994; Higami et al., 1997; Yukioka et al., 1998) y en pacientes españoles con AR (González-Escribano et al., 1999). Sin embargo, el alelo más fuertemente asociado a AR en japoneses, griegos, polinesios y españoles es HLA-DRB1*0405. Por otro lado, HLA-DRB1*0101 es el alelo que se asocia con AR en italianos, israelíes, indostanes, vascos y algunos españoles (Yelamos et al., 1993; De Juan et al., 1994). En nuestro estudio, el alelo más frecuente en pacientes con AR severa fue HLA-DRB1*0101 y el único que de manera independiente se asocia con esta condición. La frecuencia de este alelo fue significativamente mayor (38% vs 17%, P no corregida < 0.05) en pacientes con AR severa que la observada en el grupo control. Por otro lado, a pesar de que las frecuencias de los alelos HLA-DRB1*0405/*0409 (20% vs 8.5%), HLA-DRB1*0410/0411 (4% vs 2.85%), HLA-DRB1*0401/*0408 (8% vs 5.7%) y HLA-DRB1*0404 (8% vs 5.7%) fueron mayores en pacientes con AR severa con respecto al grupo control, las diferencias observadas no fueron estadísticamente significativas. El riesgo relativo para la presencia de alelos HLA-DRB1* con SE (+) fue de 6.0, mientras que para la presencia de HLA-DRB1*0101 de 2.96. Este valor es considerablemente mayor que el reportado por Wakitani et al. (1997) de 1.8, pero muy similar (2.7) al reportado por Yukioka et al. (1998) en la población japonesa.

La no asociación observada entre AR y la presencia de subtipos HLA-DRB1*04 (HLA-DRB1*0404, HLA-DRB1*0405/*0409, HLA-DRB1*0410/*0411 y HLA-DRB1*0401/*0408/*0409) contrasta con lo reportado en otros estudios en los cuales los alelos HLA-DRB1*04 son los más frecuentes en pacientes con AR severa (Gorman & Criswell 2002). Es posible que en nuestra población el alelo más común en pacientes con AR sea HLA-DRB1*0101 en lugar de los alelos HLA-DRB1*04. Los resultados de este estudio deben ser validados mediante un estudio con un número mayor de pacientes. Este trabajo representa el primero, en el cual se proporciona información sobre la distribución y asociación de HLA-DRB1 y AR severa en pacientes panameños.

Cuadro 2. Frecuencia de individuos con alelos HLA-DRB1 que codifican el SE y riesgo relativo en pacientes panameños con AR severa.

HLA-DRB1* con SE(+)	Control (N=35)	Pacientes (N=50)
0101	17%	38% (2.96)*
0405/0409	8.5%	20% (2.34)
0410/0411	2.8%	4% (1.42)
0401/0408	5.7%	8% (3.21)
0404	5.7%	8% (3.31)
HLA-DRB1*SE(+)	34%	68% (6.00)**

Números en paréntesis corresponden al riesgo relativo

HLA-DRB1* SE (+) significa todos los alelos con SE (+)

* Valor de P no corregido es significativo comparado con el grupo control.

** Valor de P corregido es significativo comparado con el grupo control.

Conclusiones

La asociación entre AR y alelos HLA-DRB1* SE (+) ha sido previamente demostrada en la gran mayoría de las poblaciones estudiadas. En pacientes panameños con AR severa la frecuencia de alelos HLA-DRB1 con SE (+) es significativamente mayor que en los individuos que no padecen de la enfermedad demostrándose una vez más la asociación entre estos alelos y la AR severa. Con excepción de HLA-DRB1*0101 ninguno de los alelos HLA-DRB1* con SE (+) mostró por sí sólo asociación con la AR severa.

Literatura citada

1. Albani, S., & D. A. Carson. 1996. A multistep molecular mimicry hypothesis for the pathogenesis of rheumatoid arthritis. *Immunol. Today* 17:466-470.
2. Bongioanni S. M., B. Porfirio, G. Rombola, E. Palasciano, E. Beneforti & G. Bianucci. 2004. Shared-epitope HLA-DRB1 alleles and sex ratio in Italian patients with rheumatoid arthritis. *Joint Bone Spine*. 71(1):24-8.
3. De Juan, M. D., I. Belmonte, J. Barado, J. Martínez-Lasso, M. Figueroa, A. Arauz-Villena & E. Cuadrado. 1994. Differential associations of HLA-DR antigens with rheumatoid arthritis (RA) in Basques: high frequency of DR1 and DR10 and lack of associations with HLA-DR4 of any of its subtypes. *Tissue Antigens* 43(5):320-3.
4. Del Rincón I, D. F. Battafarano, R. A. Arroyo, F. T. Murphy, M. Fischbach & A. Escalante. 2003. Ethnic variation in the clinical manifestation of rheumatoid arthritis: role of HLA-DRB1 alleles. *Arthritis Rheum*. 49(2):200-8.
5. González-Escribano, M. F., R. Rodríguez, A. Valenzuela, A. García & A. Núñez-Roldan. 1999. Complex association between HLA-DRB1 genes and female rheumatoid arthritis: results from a prospective study. *Hum Immunol* 60:1259-1265.
6. Gorman, J. D., & L.A. Criswell. 2002. The shared epitope and severity of rheumatoid arthritis. *Rheumatic Diseases Clinics of North America*. 28(1):265-274. *Tecnociencia*, Vol. 6, N° 2 133
7. Hajeer, A. H., A. Dababneh, R. F. Makki, W. Thomson, K. Poulton M. A. Gonzalez-Gay, C
8. Garcia-Porrua, D. L. Matthey & W. E. Ollier. 2000. Different gene loci within the HLA-DR and TNF regions are independently associated with susceptibility and severity in Spanish rheumatoid arthritis patients. *Tissue Antigens*. 55:319-325.
9. Higami, K., M. Acoda, Y. Matsuda, H. Ueda & S. Kashiwazaki. 1997. Lack of association of HLA-DRB1 genotype with radiologic progression in Japanese patients with early rheumatoid arthritis. *Arthritis Rheum*. 40(12):2241-7.
10. Kilding, R., M.M. Iles, J. M. Timms, J. Worthington & A. G. Wilson. 2004. Additional genetic susceptibility for rheumatoid arthritis telomeric of the DRB1 locus. *Arthritis Rheum*. 50(3):763-9.
11. Krause, A., T. Kamradt & G. Gurmester 1996. Potential infectious agents in the induction of arthritides. *Curr. Opin. Rheumatol*. 8:203-209.
12. Krause, A., T. Kamradt & G. Gurmester 1996. Potential infectious agents in the induction of arthritides. *Curr. Opin. Rheumatol*. 8:203-209.
13. Lee K, Park T., Park Y., Oh M., & Kim Y. 1996. DNA polymorphism analysis of the HLA-DRB1 gene using polymerase chain reaction sequence specific primers (PCR-SSP) among Korean subjects. *J Biochem Mol Biol*. 29(1):45-51.
14. Listing, J., R. Rau, B. Muller, R. Alten, E. Gromnica-Ihle, D. Hagemann, & A. Zink. 2000. HLA-DRB1 genes, rheumatoid factor and elevated C-reactive protein :independent risk factors of radiographic progression in early rheumatoid arthritis. *J Rheumatol*. 27(9).2100-9.

15. Maini, R. N., C. Q. Chu & M. Feldmann. 1995. Aetiopathogenesis of rheumatoid arthritis. In B. Henderson, J. C. W. Edwards, & E. R. Pettipher (eds.), *Mechanisms and models in rheumatoid arthritis*. pp. 25-46. Academic Press Ltd., London, England.
16. McDaniel, D. O., Alarcon, G. S., Pratt, P. W., Reveille, J. D. 1995. Most African-American patients with rheumatoid arthritis do not have rheumatoid antigenic determinant (epitope) *Ann Intern. Med.* 123 (3): 232-3.
17. Ruíz – Morales, J. A., G. Vargas – Alarcon, P. O. Flores – Villanueva, C. Villarreal-Garza, G. Hernández-Pacheco, J. K. Yamamoto-Furusho, J. M. Rodríguez-Pérez, N. Pérez-Hernández, M. Rull, M. H. Cardiel, & J. Granados. 2004. HLA-DRB1 alleles encoding the “shared epitope” are associated with susceptibility to developing rheumatoid arthritis whereas HLA-DRB1 alleles encoding an aspartic acid at position 70 of the beta-chain are protective in Mexican mestizos. *Hum Immunol.* 65(3):262-9.
18. Schonland, S. O., C. Lopez, T. Widmann, J. Zimmer, E. Bryl, J. J. Goronzy, C. M. We-lyand. 2003. Premature telomeric loss in rheumatoid arthritis is genetically determined and involves both myeloid and lymphoid cell lineages. *Proc. Natl. Acad. Sci. USA.* 100(23):13471-6.
19. Steel, R. G. & J. H. Torrie. 1995. *Bioestadística: Principios y Procedimientos*. Editorial McGraw-Hill, USA, 584 pp.
20. Toda, Y., Y. Minamikawa, S. Akagi, H. Sugano, Y. Mori, H. Nishimura, S. Arita, Y. Sugin, R. Ogawa. 1994. Rheumatoid-susceptible alleles of HLA-DRB1 are genetically recessive to non-susceptible alleles in the progression of the bone destruction in the wrist and fingers of patients with RA. *Ann Rheum Dis.* 53:587-92.
21. van Schaardenburg, D. & F.C. Breedveld. 1994. Elderly-onset rheumatoid arthritis. *Sem Arthritis Rheum* 23:367-78.
22. Wakitani, S., N. Murata, Y. Toda, R. Ogawa, T. Kaneshige, Y. Nishimura & T. Ochi. 1997. The relationship between HLA-DRB1 alleles and diseases subsets of rheumatoid arthritis in Japanese. *British J of Rheumatol.* 36:630-6.
23. Woolf, B. 1955. On estimating the relation between blood group and disease. *Ann Hum Genet.* 19:251-3.
24. Yates, P. 1934. Contingency tables involving small numbers and the X2 test. *J Royal Stat Soc Suppl.* 1:217-235.
25. Yelamos, J., J. R. García-Lozano, I. Moreno, I. Aguilera, M. F. González, A. García & A. Nuñez-roldan, B. Sánchez. 1993 association of HLA-DR4-Dw15 (DRB1*0405) and DR10 with rheumatoid arthritis in a Spanish population. *Arthritis Rheum* 36 (6):811-4.
26. Yukioka, M., S. Wakitani, N. Murata, Y. Toda, R. Ogawa, T. Kaneshige & T. Ochi. 1998. Elderly-onset rheumatoid arthritis and its association with HLA-DRB1 alleles in Japanese. *Br J Rheumatol* 37(1):98-101.
27. Zanelli E., G. Jones, M. Pascual, P. Eerligh, A.R. van der Slik, A.H., Zwinderman, W. Verduyn, G.M. Schreuder, E. Roovers, F.C. Breedveld, R.R. de Vries, J. Martin & M. J. Giphart. 2001. The telomeric part of the HLA region predisposes to rheumatoid arthritis independently of class II loci. *Hum Immunol.* 62(1):75-84.

C A P Í T U L O 7

Asociación del
polimorfismo pro12ala
en el receptor activado
por proliferadores de
peroxisomas gamma-2
(PPAR γ 2) en una muestra
de la población panameña

*Evelyn Visuetti, Tomás A. Diez, Magali de Chial, Edgardo
Castro-Pérez y Carlos Ramos Delgado*

Resumen

La diabetes mellitus es un grupo de trastornos metabólicos comunes que comparten el fenotipo de hiperglucemia. El polimorfismo Pro12Ala en el exón B del gen que codifica para PPAR γ 2 ha sido asociado en algunas poblaciones con diversos factores de riesgo relacionados con la diabetes tipo 2.

El objetivo de este estudio fue determinar si existe asociación del polimorfismo Pro12Ala de PPAR γ 2 y la diabetes mellitus tipo 2 en una muestra de la población panameña.

En esta investigación se evaluó la posible asociación del polimorfismo Pro12Ala en una muestra de 195 individuos (99 controles y 96 experimentales). El polimorfismo Pro12Ala se detectó mediante el análisis de la reacción en cadena de la polimerasa y polimorfismo de longitud de fragmentos de restricción (PCR-RFLP).

Se observó una mayor frecuencia del genotipo Pro12Ala en el grupo experimental (DM2) con respecto al grupo control (0.114 vs 0.010). La frecuencia del alelo Ala12 (alelo de menor frecuencia) en nuestro estudio fue de 2.8%.

Este trabajo representa el primer estudio sobre asociación entre diabetes mellitus tipo 2 y el polimorfismo Pro12Ala (rs18012829) en una muestra de casos y controles en Panamá. La razón de momios (OR) evidenció asociación con un riesgo relativo de 1.9735 en aquellos con genotipo Pro12Ala. La frecuencia del genotipo Pro12Ala fue mayor en mujeres, lo cual sugiere que estas pueden estar más predispuestas a desarrollar esta condición.

Palabras claves

PPAR γ 2, polimorfismo, diabetes mellitus, población panameña

Abstract

Diabetes mellitus is a group of common metabolic disorders that share the phenotype of hyperglycemia. In some populations, the Pro12Ala polymorphism in exon B of the gene that codes for PPAR γ 2 has been associated with various risk factors related to type 2 diabetes. The objective of this study was to determine if there is an association between the Pro12Ala polymorphism of PPAR γ 2 and type 2 diabetes mellitus in a sample of

the Panamanian population. This research evaluated the possible association of the Pro12Ala polymorphism in a sample of 195 individuals (99 controls and 96 experimental). The Pro12Ala polymorphism was detected by polymerase chain reaction and restriction fragment length polymorphism (PCR-RFLP) analysis. A higher frequency of the Pro12Ala genotype was observed in the experimental group (DM2) compared to the control group (0.114 vs 0.010). Our study's frequency of the Ala12 allele (lowest frequency allele) was 2.8%. This work represents the first study on the association between type 2 diabetes mellitus and the Pro12Ala polymorphism (rs18012829) in a sample of cases and controls in Panama. The odds ratio (OR) was associated with a relative risk of 1.9735 in those with the Pro12Ala genotype. The frequency of the Pro12Ala genotype was higher in women, suggesting that they may be more predisposed to developing this condition.

Keywords

PPAR γ 2, polymorphism, diabetes mellitus, Panamanian population.

Introducción

La diabetes mellitus es un grupo de trastornos metabólicos comunes que comparten el fenotipo de hiperglucemia. Existen varios tipos de diabetes mellitus y son causados por una interacción compleja de factores genéticos y ambientales. Los factores que contribuyen a la hiperglucemia incluyen la reducción de la secreción de insulina, la disminución de la utilización de glucosa y el aumento o la producción de glucosa (Janani & Ranjitha, 2014). La ausencia o poca producción de insulina se correlaciona con la presencia de pocos transportadores de glucosa (GLUTs) en la membrana y, por ende, una reducción en el transporte de glucosa hacia el interior de la célula (AsocMexDiabetes, 2021).

Los dos tipos principales de diabetes mellitus son el tipo 1 y el tipo 2 (DM1, DM2). La diabetes mellitus tipo 1 se caracteriza principalmente por ser autoinmune, ya que el propio sistema inmunológico ataca y destruye las células beta que se encuentran en el páncreas, que son las encargadas de producir la insulina y, por lo tanto, se produce poca o ninguna insulina. Además, suele desarrollarse en la infancia o durante la adolescencia, pero puede presentarse a cualquier edad (Lucier & Weinstock, 2023).

La diabetes mellitus tipo 2 (DM2) es la más común y representa la ma-

yoría de los casos; generalmente, suele aparecer en personas mayores de 40 años, aunque podría presentarse en diferentes rangos de edades. Se caracteriza por niveles elevados de glucosa en sangre, provocado por la resistencia a la insulina. Las células no responden normalmente a la insulina y, por lo tanto, el cuerpo necesita cada vez más insulina para que la glucosa ingrese a las células y pueda ser utilizada como fuente de energía. Esto causa que la glucosa se acumule en la sangre, generando niveles de glucosa mayores de 125 mg/dl (Westman, 2021).

Los Receptores Activados por Proliferadores de Peroxisomas son factores de transcripción activados por ligandos. Existen tres isoformas en las que se incluye el Receptor Activado por Proliferadores de Peroxisomas Gamma-2 (PPAR γ 2) (Zárate et al., 2006).

Las variaciones o polimorfismos de un nucleótido (SNPs) son responsables de más del 80 % de la variación entre individuos DM2 (Sarhangi et al. 2020). Estas variaciones en uno o más genes, conjuntamente con variaciones en el ambiente y el estilo de vida, han sido asociadas en diversas poblaciones a la susceptibilidad a ciertas enfermedades (Kirk et al. 2002).

Existen polimorfismos de un nucleótido (SNPs) en el gen que codifica para PPAR γ 2 ha sido asociado con el riesgo de padecer de DM2 (Sarhangi et al. 2020) o con un efecto protector hacia el desarrollo de la enfermedad (Hara et al., 2000). El polimorfismo Pro12Ala en el exón B del gen que codifica para PPAR γ 2 ha sido asociado en algunas poblaciones con el riesgo o susceptibilidad de padecer de diabetes mellitus tipo 2, mientras que en otras ha sido asociado con un efecto protector o incluso no se ha evidenciado asociación con la enfermedad.

Según el Ministerio de Salud de Panamá, el 14% de la población panameña padece DM2, lo que representa una alta incidencia. Además, de acuerdo con la OMS, alrededor de 450 mil personas en Panamá se encuentran en prediabetes, por lo que es importante evaluar si el polimorfismo Pro12Ala en nuestra población está asociado a DM2 y si constituye un factor protector o de riesgo de la enfermedad.

En esta investigación se evaluó la posible asociación del polimorfismo Pro12Ala en una muestra de 195 individuos (99 controles y 96 experimentales) de la población panameña. No existen estudios previos de asociación de este polimorfismo con DM2 en nuestra población y, por lo tanto, este trabajo constituye el primer estudio, en el cual se establece asociación entre este polimorfismo y una de las condiciones de mayor morbimortalidad en nuestro país.

Materiales y métodos

Sitio de estudio

La investigación se realizó en el Laboratorio de Genética y Biología Molecular, ubicado dentro de la Universidad de Panamá, Campus Central Octavio Méndez Pereira (8°59'09° N, 79°31'57'' W).

Grupos de estudio

Noventa y seis (96) individuos con diabetes mellitus tipo 2, mayores de 40 años, previamente diagnosticados conformaron el grupo experimental, y noventa y nueve (99) individuos voluntarios sanos, también mayores de 40 años, conformaron el grupo control. Se registraron, cuando fue posible, datos básicos como nombre, apellido, sexo, edad, medidas de altura, peso e índice de masa corporal (IMC). En ningún momento estos datos fueron compartidos o comprometidos con personas que no estaban involucradas en la investigación. La participación en el estudio fue voluntaria con consentimiento informado por parte de los participantes. El estudio fue aprobado por el comité de bioética de la Universidad de Panamá.

Obtención de la muestra de sangre

Para la obtención de la muestra de sangre se seleccionó el dedo medio, al cual se le aplicó presión desde la falange media hacia la distal con el propósito de concentrar la sangre en esta región. Manteniendo la presión, se procedió a limpiar la región con alcohol al 70% y, posteriormente, se realizó una punción con una lanceta estéril. La sangre fue colectada en una tarjeta FTA e identificada con el código correspondiente (C control y E experimental-padece DM2) seguido del número que identifica la muestra. Una vez colectada la muestra, se presionó el dedo con algodón impregnado en alcohol 70% hasta detener el sangrado. Las muestras obtenidas fueron almacenadas a 4 °C para su posterior utilización en la extracción de ADN.

Extracción de ADN

Se recortaron de la tarjeta FTA pequeños pedazos de aproximadamente 0.5 mm, conteniendo sangre. Los pedazos fueron colocados en un tubo Eppendorf de 1.5 ml conteniendo 200 microlitros de buffer de extracción (TL Buffer de Omega BIO-TEK) e incubados durante 1 hora a 55 °C. Una vez lograda la lisis de los glóbulos blancos, la extracción del ADN se realizó de acuerdo con el procedimiento descrito en el manual E.Z.N.A. Tissue DNA Kit (Omega BIO-TEK). El ADN extraído fue almacenado a -20°C para luego ser utilizado en la reacción de amplificación.

Determinación de la concentración y calidad del ADN

La concentración de ADN fue determinada mediante espectrofotometría utilizando el Espectrofotómetro de micromuestra Nanodrop™ 2000 de ThermoScientific, el cual posee un rango de espectral (190-840nm) para medir diferentes tipos de muestra. La concentración de ADN se determinó mediante la lectura de la absorbancia a 260 nm y la calidad mediante el cociente de las lecturas a 260 y 280 nm.

Reacción en Cadena de la Polimerasa (PCR)

La reacción de amplificación se realizó en un volumen final de 30 µL conteniendo 2 µL (~6.0 ng) de ADN, 12.5 µL de Mastermix 2X (abm MegaFi™ Pro Fidelity), 1 µL de cada primer (10 µM) y 11 µL de agua libre de nucleasas. Las condiciones de amplificación fueron de 94 °C durante 8 minutos, seguido de 35 ciclos a 94 °C por 50 segundos, 59 °C por 50 segundos, 72 °C por 1 minuto y una extensión final de 5 min a 72 °C. Se utilizó un termociclador T100 Thermal Cycler de BIO-RAD.

Los cebadores utilizados en la PCR fueron 5'-GCCAATTCAAGCCCA-GTC-3' y 5'-GATATGTTTGCA GACAGTGTATCAGTGAAGGAATC-GCTTTCCG-3' (Priya et al. 2016). Estos cebadores amplifican una región de 270 bp del exón B del gen PPAR γ 2.

Electroforesis del producto de PCR

Con la finalidad de determinar la efectividad de la amplificación, se realizó una electroforesis en gel de agarosa al 1% en TAE IX. Para tal propósito, se cargaron 5 μ L de la reacción de amplificación y la electroforesis se realizó a 60 voltios durante 1 hora. En el primer carril se cargaron los marcadores de peso molecular 100 bp (Bioline) para verificar la generación de un producto de amplificación correspondiente al tamaño esperado (270 bp).

Digestión del producto de PCR

Una vez verificada la amplificación de la región esperada, se procedió a digerir 10 μ L (~30 ng) del producto de amplificación con la enzima de restricción BstU1 (New England Biolabs). La digestión se realizó a 60 °C durante 3 horas. La reacción de digestión se llevó a cabo en un volumen final de 25 μ L y 0.5 μ L de la enzima (5 U/ μ L).

Determinación del genotipo

El patrón de digestión generado a partir de cada muestra indica el genotipo presente en cada individuo. Debido a que el polimorfismo Pro12Ala resulta del cambio de un nucleótido que genera un sitio de restricción de la enzima de digestión BstU1, los individuos que poseen un alelo Ala (heterocigotos) generarán un patrón de restricción de dos bandas: una banda de 270 bp correspondiente al alelo Prolina que no posee un sitio de digestión para la enzima BstU1 y otra banda de 227 bp correspondiente al alelo Alanina, el cual es di-

gerido por la enzima, generando dos bandas: una de 227 bp y otra de 43 bp; esta última no se observa en el gel. De esta manera, los individuos heterocigotos (Pro/Ala) presentan una banda de 270 y otra de 227 bp, mientras que los individuos homocigotos Pro/Pro presentan un patrón de una sola banda de 270 bp. El genotipo homocigoto Ala12Ala estaría presentando una banda de 227 bp ya que, como se mencionó con anterioridad, la banda de 43 pb no se observa en la electroforesis.

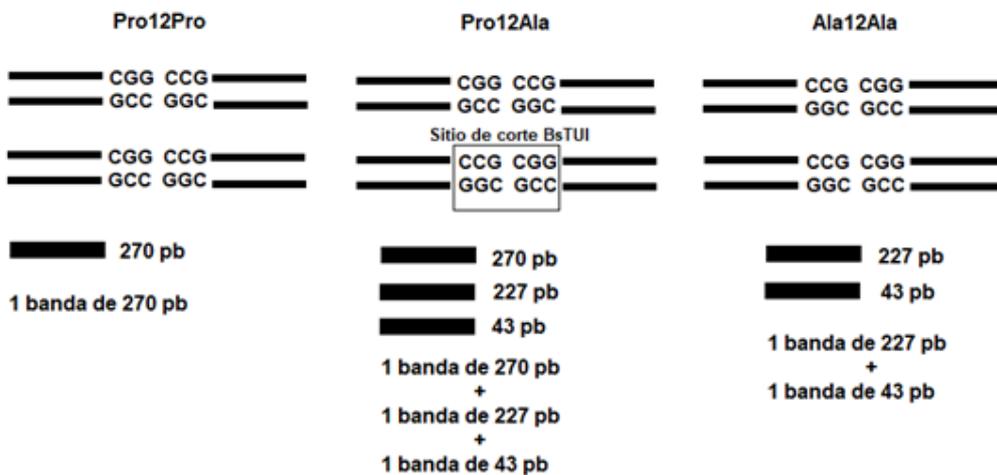


Figura 1. Productos de la digestión, sitios de restricción.

Electroforesis de los productos de digestión

El patrón de digestión fue evidenciado mediante electroforesis en gel de agarosa al 3% en TAE IX conteniendo 2 μ L de GelRed (Bio-tium). Diez microlitros (10 μ L) del producto de digestión de cada muestra fueron cargados en el gel. En cada gel se cargaron, además, 0.5 μ L de marcador de peso molecular 100 bp Ladder (Bioland). La electroforesis se realizó a 45 voltios durante 2 horas. El patrón de digestión fue visualizado en un fotodocumentador UV Gel Doc EZ BioRad.

Frecuencias Genotípicas y alélicas

Las frecuencias genotípicas y alélicas fueron calculadas tanto en la muestra total como en cada grupo (control y experimental). La frecuencia de cada genotipo fue calculada utilizando la siguiente relación:

$f(\text{ProPro}) = \text{cantidad de individuos con genotipo ProPro} / \text{Total de individuos}$.

$f(\text{Pro12Ala}) = \text{cantidad de individuos con genotipo Pro12Ala} / \text{Total de individuos}$.

$f(\text{AlaAla}) = \text{cantidad de individuos con genotipo AlaAla} / \text{Total de individuos}$.

La frecuencia de cada alelo fue calculada utilizando la siguiente relación:

$$f(\text{Pro}) = \frac{2(\text{cantidad individuos con genotipo ProPro} + \text{cantidad de individuos Pro12Ala})}{2(\text{total de individuos})}$$

$$f(\text{Ala}) = \frac{2(\text{cantidad individuos con genotipo AlaAla} + \text{cantidad de individuos Pro12Ala})}{2(\text{total de individuos})}$$

Equilibrio de Hardy-Weinberg (HWE)

Una vez determinadas las frecuencias alélicas y genotípicas totales, se procedió a determinar si las mismas se encontraban en equilibrio de Hardy-Weinberg, en donde las frecuencias genotípicas esperadas están dadas por la siguiente ecuación:

$$p^2 + 2pq + q^2 = 1$$

El HWE también fue determinado en cada grupo por separado (casos y experimentales).

Se utilizó el calculador en línea (wpcalc.com/en/equilibrium-hardy-weinberg/) para determinar si las frecuencias observadas se ajustaban o no al HWE.

Determinación de asociación entre el polimorfismo Pro12Ala y DM2

La asociación del genotipo Pro12Ala fue evaluada mediante una prueba de chi cuadrado (χ^2), en la cual las hipótesis a probar son:

H₀: Existe asociación entre el genotipo Pro/Ala y DM2

H₁: No existe asociación entre el genotipo Pro/Ala y la DM2

Para comprobar la asociación entre el genotipo Pro/Ala y DM2, se realizó una prueba de Razón de Momios (odds ratio) para determinar el grado de asociación entre estas variables.

Razón de Momios

Para determinar si existe asociación entre el polimorfismo Pro/Ala y la DM2, la razón de momios fue calculada utilizando el programa Epi Info 7.2.

Epi Info es un software de dominio público que utiliza la comunidad global de profesionales e investigadores de la salud. Esta herramienta proporciona a profesionales carentes de experiencia en tecnologías de la información, formularios simples de ingreso de datos y creación de bases de datos y análisis de datos, mapas y gráficos epidemiológicos. Se encuentra disponible tanto para Windows, Mobile y Web & Cloud (<https://www.cdc.gov/epiinfo/index.html>).

La razón de momios fue calculada utilizando la siguiente relación:
RM = ad/bc

CASOS (EXPERIMENTALES)

Enf E (DM2 + Pro/Ala) = a

Enf E (DM2 + Pro/Pro) = c

CONTROLES (CONTROLES)

Controles E (sanos + Pro/Ala) = b

Controles E (sanos + Pro/Pro) = d

Resultados

Se analizaron 195 muestras de las cuales 96 corresponden a pacientes con diagnóstico certero de DM2 (experimentales) y 99 individuos sanos (controles).

El genotipo presente en cada individuo se obtuvo mediante PCR-RFLP, resultando 183 individuos con genotipo Pro/Pro, 12 Pro12Ala y 0 individuos Ala/Ala.

El genotipo presente se determinó mediante el tamaño de las bandas generadas después del corte con la enzima de restricción BstUI. En la figura 7 se observan los genotipos identificados en las muestras 15C (control), 33E, 34E, 35E, 36E, 38E y 39E (experimentales). En los carriles 11 y 12 se observan los productos de PCR de la muestra 15C sin digerir y digerido, respectivamente, correspondientes a un individuo de género femenino. En el carril 11 se observa que el producto de amplificación de esta muestra migra ligeramente más que el marcador de peso molecular, que corresponde a 300 bp, lo cual indica que el producto de amplificación es de 270 bp, como se esperaba. En el carril 12, la misma muestra, luego de la digestión con BstUI, se observa que la banda de 270 persiste, lo cual indica que en el fragmento amplificado de este individuo no existe sitio de restricción para la enzima BstUI; es decir, no ha ocurrido el cambio de citosina (C) por guanina (G) en el sitio rs1801282 que genere el sitio de restricción de la enzima BstUI. Por lo tanto, el genotipo asignado a la muestra 15C es Pro/Pro; es decir, homocigoto para el alelo con el residuo de aminoácido prolina en la posición mencionada (rs1801282).

Los carriles 13 y 14 corresponden a productos de amplificación no digeridos (270 bp) de las muestras 33E y 34E, mientras que los carriles 15 y 16 corresponden a los productos de amplificación digeridos con BstUI de esas mismas muestras. Se puede observar en estos carriles la banda 270 bp correspondiente al producto de amplificación que no posee sitio de restricción de BstUI, y otra banda de 227 bp que resulta de la digestión con la mencionada enzima. El producto de PCR con sitio de restricción para BStUI genera dos bandas: 227 y 43 bp. Es importante señalar que la banda de 43 bp no se observa en la electroforesis. Este patrón corresponde a individuos heterocigotos Pro/Ala debido a que se generan dos productos de amplificación, uno que es digerido y otro que no es digerido, como resultado de la sustitución de C por G en uno de los dos alelos. Las muestras 35E, 36E y 38E corresponden a individuos homocigotos Pro/Pro mientras que la muestra 39E corresponde a un individuo heterocigoto (Pro/Ala).

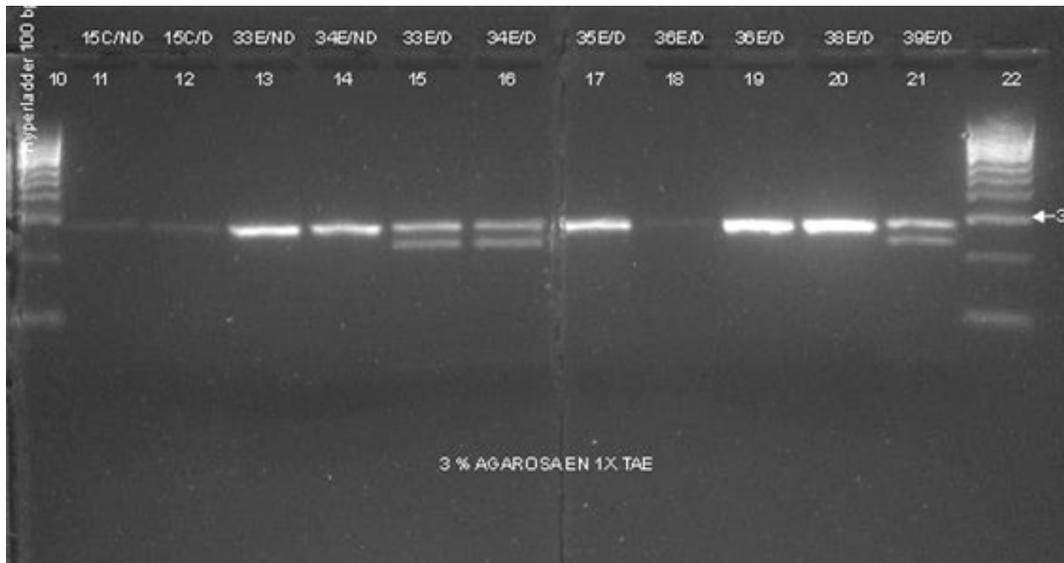


Figura 2. Detección del polimorfismo Pro12Ala mediante PCR. Carril 10 y 22 muestran los marcadores de peso molecular. Los carriles 11, 12, 13, 14, 17, 18, 19, 20 muestran el genotipo Pro12Pro (un solo fragmento de 270 pb). Los carriles 15, 16 y 21 muestran el genotipo Pro12Ala con los dos fragmentos de 270 pb y 227 pb que resultan del sitio de restricción adicional en estos individuos. El fragmento de 43 pb no se observa.

No se observó la presencia del genotipo homocigoto (Ala/Ala) en los individuos en estudio, tomando en consideración la ausencia de una sola banda de 227 bp en las muestras digeridas.

Las frecuencias alélicas y genotípicas de ambas poblaciones (experimental y control) se muestran en la Tabla 1. Se observa la ausencia de individuos con genotipo (Ala/Ala) dentro de nuestra muestra.

El alelo Pro se presentó con una frecuencia total de 0.969, mientras que la frecuencia total del alelo Ala es de 0.030. La frecuencia del alelo Pro en individuos que no padecen de DM2 (controles) fue de 0,995, mientras que en individuos con DM2 fue de 0,943. La frecuencia del alelo Ala fue mayor en individuos con DM2 que en aquellos que no padecen la condición. El genotipo con mayor frecuencia fue Pro/Pro en la muestra total y en ambos grupos (control y experimental). Por otro lado, los heterocigotos Pro/Ala se observaron con mayor frecuencia en individuos con DM2 que en individuos sanos.

Tabla 1. Frecuencias alélicas y genotípicas. N, cantidad de individuos; Cont., controles; Exp., experimentales; y f, frecuencia.

N (195)	f(Pro)	f(Ala)	f(ProPro)	f(Pro12Ala)	f(AlaAla)
Cont. 99	0.995	0.005	0.989	0.011	0
Exp. 96	0.943	0.057	0.885	0.114	0
Total. 195	0.969	0.030	0.938	0.061	0

Tabla 2. Distribución de los individuos por género según su genotipo. F (femenino), M (masculino) y SD (sin determinar).

	Controles			Experimentales		
	Pro/ Pro	Pro/Ala	Ala/Ala	Pro/ Pro	Pro/Ala	Ala/Ala
Sexo	0.943	0.057	0.885	0.114	0	
F	51	1	0	42	6	0
M	42	0	0	43	3	0
SD	5	0	0	0	2	0
Total	98	1	0	85	11	0

En ambos grupos (control y experimental) el genotipo Pro12Ala fue observado mayormente en mujeres. En las muestras experimentales, de los 11 individuos heterocigotos, 6 corresponden al género femenino, 3 al masculino y en los 2 heterocigotos restantes no se registró la información del género.

Razón de momios (odds ratio)

La RM fue de 12.6824 con un Riesgo Relativo (RR) estimado de 1.9735 (Tabla 3). En nuestro estudio, el Riesgo Relativo de desarrollar DM2 es mayor en individuos heterocigotos Pro/Ala comparado con respecto a aquellos que no poseen este genotipo. Un valor RM mayor de 1 indica que el genotipo Pro12Ala se comporta como un factor de riesgo.

$$RM = ad/bc$$

$$RM = (11)(98)/(1)(85)$$

$$RM = 1078/85$$

$$RM = 12.68$$

RM CON FRECUENCIAS

$$RM = (0.114)(0.989)/(0.010)(0.885)$$

$$RM = 0.112/0.009$$

$$RM = 12.44$$

ALTO

Tabla 3. Análisis mediante Epi info 7.2

Parámetros basados en la posibilidad			
	Estimado	Inferior	Superior
dds Ratio	12.6824	1.6041	100.2708
OMLE Odds Ratio	12.5599	2.0797	278.0071
Fisher Exact		1.7599	551.3208
Parámetros basados en el riesgo			
	Estimado	Inferior	Superior
Riesgo relativo	1.9735	1.5667	2.4861
Diferencia de riesgo	45.2186	27.9918	62.4454
Pruebas estadísticas			
	Chi cuadrado	2-tailed p	
No corregido	9.2129	0.00240438	

Discusión

La variación de un sólo nucleótido en el exón B del gen que codifica para PPAR γ 2 (Pro12Ala) ha sido estudiada en varias poblaciones como una variante genética asociada a DM2. El alelo Ala12 ha sido asociado a reducción del riesgo o protección de padecer diabetes mellitus tipo 2 (Altshuler et

al., 2000; Hara et al., 2000; Frederiksen et al., 2002; Poulsen et al., 2003; Memisoglu, et al., 2003; Andrulionyté et al., 2004; Doney et al., 2004; Pinterová et al., 2004; Moon et al., 2005; Jaziri, et al., 2006; Soriguer et al., 2006; Meshkani et al., 2007). Por el contrario, en otras poblaciones la presencia de este alelo ha sido asociado con predisposición o riesgo de desarrollar DM2 (Evans et al., 2001; Herder et al., 2008; Kilpelainen et al., 2008; Florez et al., 2007; Gouda et al., 2010; Ghoussani et al., 2005; Zeggini et al., 2005; Chistiakov et al., 2010; Vergotine et al., 2014; Majid et al., 2017). En otros estudios no se observó asociación entre este alelo y la diabetes mellitus tipo 2 (Mancini et al., 1999; Bouassida et al., 2005; Badii et al., 2008; Malecki et al., 2003; Moffett et al., 2005).

Independientemente de los resultados contradictorios en cuanto a asociación o no de Ala12 con DM2, el alelo de menor frecuencia (poco común) es el que codifica para alanina en la posición 12 (Ala12), en todas las poblaciones estudiadas. La frecuencia de este alelo varía entre 2% a 23% en diferentes grupos étnicos (Stunvoll & Haring, 2002). En individuos sanos, la frecuencia de este alelo es de 1.7 a 21.6%; en caucásicos varía entre 5.9% y 21.6%, y en individuos de ascendencia asiática entre 1.7% y 9.3% (Gouda et al., 2010).

En nuestro estudio, la frecuencia del alelo Ala12 fue de 2.8%, frecuencia que se ubica en el rango de los valores de las frecuencias reportadas para este alelo a nivel mundial. No obstante, la frecuencia fue mucho menor que la observada en niños mexicanos (Stryjecki et al., 2015), Amerindios (Cañizales-Quintero et al., 2007) y niños españoles (Vales-Villamarín et al., 2021). La frecuencia del alelo Ala12 observada en nuestro estudio se aproxima más al 2.1 % observada en negros americanos (Fornage et al., 2005) y al 2.2 % reportada en algunas poblaciones asiáticas (Fatemeh Namvaran et al., 2010).

Las frecuencias genotípicas en la muestra total, como también en ambos grupos control y el experimental, se encuentran en equilibrio de Hardy-Weinberg. Las frecuencias genotípicas en la muestra total fueron 93.8 %, 6.1% y 0 % para los genotipos Pro12Pro, Pro12Ala y Ala12Ala, respectivamente. Obviamente no se observaron individuos homocigotos Ala12Ala en nuestro trabajo. La ausencia de homocigotos Ala12Ala ha sido reportada en otros estudios en la población iraní y en los nativos de Java en Indonesia (Namvaran et al., 2010; Danawati et al., 2005). Frecuencias genotípicas de 97.4 % (Pro/Pro), 2.5% (Pro/Ala) y 0 % (Ala/Ala) un tanto similares a las nuestras fueron observadas en un estudio con 540 nativos de la isla de Java (Danawati et al., 2005).

Las frecuencias genotípicas por grupo fueron significativamente diferentes como demuestra el valor de chi cuadrado. Se observó una mayor frecuencia del genotipo Pro12Ala en el grupo experimental (DM2) con respecto al grupo control (0.114 vs 0.010). Interesantemente, la frecuencia de este genotipo en mujeres con DM2 fue mayor que en hombres con esta condición. Resultados similares se reportaron en nativos americanos de la etnia Oji-Cree (Hegele et al., 2000).

En nuestro estudio, el polimorfismo Pro12Ala muestra asociación con DM2 como un factor de riesgo, como ha sido reportado en varias investigaciones y contrario a otros en los que se asocia como un factor protector. En cuanto a la asociación o no como factor protector o de riesgo, las diferencias pueden ser atribuidas al hecho de que la DM2 es una enfermedad multifactorial en la cual intervienen tanto componentes genéticos y componentes ambientales (Chritiansen et al., 2023). La mezcla racial es un factor genético que juega un papel importante, ya que se ha demostrado que el riesgo de padecer diabetes mellitus tipo 2 varía entre grupos raciales o étnicos (Weiss et al., 1984).

Se ha demostrado que los Indios Americanos poseen una de las tasas más altas de diabetes en el mundo (West 1978; Bennett 1983).

Estudios de prevalencia de diabetes mellitus tipo 2 en poblaciones producto de la mezcla entre indios y caucásicos, como por ejemplo en poblaciones de hispanos del suroeste de Estados Unidos, muestran una prevalencia intermedia a la observada en indios y caucásicos (Gardner et al., 1984; Iyengar et al., 1991).

La obesidad, el aumento en el consumo de calorías y grasas, la reducción en el consumo de carbohidratos complejos y la poca actividad física son factores ambientales asociados a un incremento en la incidencia y prevalencia de diabetes mellitus tipo 2 (Hamman, 1992; Torun et al., 2002; Klimentidis et al., 2009). A pesar de que se ha cuestionado el rol de los factores ambientales, debido a la carencia de este tipo específico de estudio estrictamente ambientales, es importante mencionar que, en investigaciones en Hawaianos que migraron a Japón (Kawate et al., 1979;) y en Yemenies que migraron a Israel (Cohen et al., 1961), se observó un incremento de esta patología en los migrantes con relación a individuos del mismo acervo genético que no migraron, lo cual indica que el medio es un factor que se relaciona con el desarrollo de la condición. Esta interacción entre genes y ambiente podría explicar la inconsistencia en los resultados de estudios de asociación genética con la DM2. Estudios de análisis

de interacción logística de parejas de SNPs evidenciaron que la presencia de ciertas parejas de SNPs se asocian significativamente como factor de riesgo de padecer DM2 a pesar de que, individualmente, no mostraron asociación (Irgam et al., 2021). La interacción entre el estilo de vida (dieta, actividad física) y las variantes genéticas relacionadas con DM2 deben ser estudiadas en cada grupo étnico (Hur et al., 2022).

Un aspecto adicional a considerar son las modificaciones epigenéticas en la patogénesis de DM2 (Ahmed et al., 2020). Estudios realizados hace algunas décadas han demostrado la relación entre epigenética y diabetes mellitus tipo 2. Se han identificado alteraciones en el patrón de metilación del ADN en los islotes pancreáticos, tejido adiposo, músculo esquelético e hígado en individuos con DM2. De hecho, factores no genéticos asociados al riesgo de padecer de DM2, como la obesidad, la dieta, inactividad física, envejecimiento y entorno intrauterino, han sido asociados con modificaciones epigenéticas (Ling et al., 2022).

En nuestro país, la mezcla racial responde a un modelo trihíbrido complejo con contribuciones de africanos, europeos y amerindios (Castro et al., 2016). Esta compleja mezcla racial podría aportar componentes genéticos característicos que, aunados a los factores ambientales propios de nuestra sociedad y estilo de vida, respondan, de manera muy particular, al desarrollo o no de la DM2. Por lo tanto, algunos componentes genéticos presentes en nuestra población, no considerarse como factores de riesgo frente a ciertos factores ambientales o estilos de vida o bien ser considerados factores de riesgo, dependiendo del género.

En resumen, este es el primer estudio sobre el polimorfismo Pro12Ala en el gen PPAR γ 2 en la población panameña, encontrando una posible asociación de riesgo entre este polimorfismo y la DM2.

Conclusiones

Este trabajo representa el primer estudio sobre asociación entre diabetes mellitus tipo 2 y el polimorfismo Pro12Ala (rs18012829) a partir de una muestra de casos y controles de la población panameña. La razón de momios (OR) evidenció la asociación con un riesgo relativo de 1.9735 en aquellos con genotipo Pro12Ala. La frecuencia del genotipo Pro12Ala fue mayor en mujeres, lo cual sugiere que estas pueden estar más predispuestas a desarrollar esta condición. Los resultados observados en este trabajo no necesariamente son aplicables a otras poblaciones, ya que el acer-

vo genético de la población panameña, el estilo de vida y las condiciones medioambientales son muy particulares.

Se hace necesario realizar estudios de asociación de genoma completo (GWAS), en los cuales se consideren varios marcadores de manera simultánea y se correlacionen estos con patrones epigenéticos y estilos de vida, mediante un análisis de metadatos en una muestra mayor de la población panameña considerando un diseño de casos y controles.

Literatura citada

1. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM *et al* (2015) A global reference for human genetic variation. *Nature* 526(7571):68–74.
2. 9th edition | IDF Diabetes Atlas. (s. f.-b). Copyright © IDF Diabetes Atlas 2024. All Rights Reserved. <https://diabetesatlas.org/atlas/ninth-edition/>
3. Ahmed, S. A. H., Ansari, S. A., Mensah-Brown, E. P. K. & Emerald, B. S. 2020. Clinical Epigenéticas. 12:104 <https://doi.org/10.1186/s13148-020-00896-4>
4. Akhlaghipour, I., Bina, A.R., Mogharrabi, M.R. *et al*. Single-nucleotide polymorphisms as important risk factors of diabetes among Middle East population. *Hum Genomics* 16, 11 (2022). <https://doi.org/10.1186/s40246-022-00383-2>
5. Al-Shali KZ, House AA, Hanley AJ, Khan HM, Harris SB, Zinman B, Mamakeesick M, Fenster A, Spence JD, Hegele RA. Genetic variation in PPARG encoding peroxisome proliferator-activated receptor gamma associated with carotid atherosclerosis. *Stroke*. 2004 Sep;35(9):2036–40. doi: 10.1161/01.STR.0000138784.68159.a5. Epub 2004 Jul 29. PMID: 15284449.
6. Altshuler, D., Hirschhorn, J., Klannemark, M. *et al*. The common PPAR γ Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nat Genet* 26, 76–80 (2000). <https://doi.org/10.1038/79216>
7. Alwi ZB (2005) The use of SNPs in pharmacogenomics studies. *Malays J Med Sci MJMS* 12(2):4–12
8. AsocMexDiabetes. (2021, 30 junio). TRANSPORTADORES DE GLUCOSA (GLUTS). *Asociación Mexicana*. <https://www.amdiabetes.org/post/transportadores-de-glucosa-gluts>
9. Atención, M. 2., & del paciente con Diabetes mellitus tipo, C. y. T. (s/f). *Experto en el MANEJO DE*. Campuspanamericana.com. Recuperado el 26 de septiembre de 2023, de [http://aula.campuspanamericana.com/_Cursos/Curso01417/Temario/Experto_Citologia_Sangre/02%20M2T1%20\(DM\).pdf](http://aula.campuspanamericana.com/_Cursos/Curso01417/Temario/Experto_Citologia_Sangre/02%20M2T1%20(DM).pdf)
10. Atug O., Tahan V., Eren F., Tiftikci A., Imeryuz N., Hamzooglu H.O., Tozun N. Pro12Ala-polymorphism in the peroxisome proliferator-activated receptor-gamma (PPAR γ) gene in inflammatory bowel disease. *J. Gastrointest. Liver Dis.* 2008; 17(4): 433-437.
11. Badii, R., Bener, A., Zirie, M., Al-Rikabi, A. C., Şimşek, M., Al-Hamaq, A. O., Ghous-saini, M., Froguel, P., & Wareham, N. (2007). Lack of association between the Pro-12ALA polymorphism of the PPAR- γ 2 gene and Type 2 diabetes mellitus in the Qatari consanguineous population. *Acta Diabetologica*, 45(1), 15–21. <https://doi.org/10.1007/s00592-007-0013-8>
12. Barcia, M. A. L., Mero, L. H. A., Muentes, R. A. Z., & Tejena, J. E. S. (2022). Actualización diabetes tipo I. *Análisis del comportamiento de las líneas de crédito a través de la corporación financiera nacional y su aporte al desarrollo de las PYMES en Guayaquil 2011-2015*, 6(suppl 1), 267–274. [https://doi.org/10.26820/recimundo/6.\(suppl1\).junio.2022.267-274](https://doi.org/10.26820/recimundo/6.(suppl1).junio.2022.267-274)

13. Bennett PH: Diabetes in developing countries and unusual populations. In *Diabetes in Epidemiologic Perspective*, Mann JI, Pyorala K, and Teuscher A, Eds. Churchill-Livingstone, Edinburgh, 1983, pp 43-57.
14. Cañizales-Quinteros, Aguilar-Salinas, C., Ortiz-López, M G, Rodríguez-Cruz, M, Villarreal-Molina, M T, Coral-Vázquez, R, Huertas-Vázquez, A, Hernández-Caballero, A, López-Alarcón, M, Brito-Zurita, O R, Domínguez-Banda, A, Martínez-Sánchez, L R, Canto-de Cetina, T, Vilchis-Dorantes, G, Rosas-Vargas, H, Granados-Silvestre, M A, Medeiros-Domingo, A, Menjivar, M, Tusié-Luna M T. 2007. Association of PPAR γ Pro12Ala variant with larger body mass index in Mestizo and Amerindian populations of Mexico. *Hum Biol.* 79(1):111-9. doi: 10.1353/hub.2007.0022.
15. Carvajal K, Hernández-Esquivel ML, Moreno-Sánchez R. PPARs, síndrome metabólico y enfermedad cardíaca. *Arco Cardiol Mex.* 2007;77 (Suplemento: 4):66-76.
16. Chistiakov, D. A. et al. Te PPAR γ Pro12Ala variant is associated with insulin sensitivity in Russian normoglycaemic and type 2 diabetic subjects. *Diab. Vasc. Dis. Res.* 7, 56-62. <https://doi.org/10.1177/1479164109347689> (2010)
17. Cho YM, Kim DH, Lee KH, Jeong SW, Kwon OJ. The IRE1 α -XBP1s pathway promotes insulin-stimulated glucose uptake in adipocytes by increasing PPAR γ activity. *Exp Mol Med.* 2018 Aug 15;50(8):1-15. doi: 10.1038/s12276-018-0131-0. PMID: 30111834; PMCID: PMC6093883.
18. Christiansen, C. E., Arathimos, R., Pain, O., Molokhia, M., Bell J. T. & Lewis, C. M. 2023. Stratified genome-wide association analysis of type 2 diabetes reveals subgroups with genetic and environmental heterogeneity. *Human Molecular Genetics* 32 (16): 2638-2645.
19. Cohen AM, Bavly S, and Poznanski R: Change of diet of Yemenite Jews in relation to diabetes and ischemic heart disease. *Lancet* 21399-1401, 1961.
20. Contraloría General de la República, Panamá en Cifras, 2020.
21. Cruz, WTN (s/f). *En Panamá, la prevalencia de pacientes diabéticos es de doscientos mil personas*. Gob.pa. Recuperado el 8 de septiembre de 2023, de <https://prensa.css.gob.pa/2020/11/13/en-panama-la-prevalencia-de-pacientes-diabeticos-es-de-doscientos-mil-personas/>
22. Cubillo Espinoza A. Tamizaje de diabetes gestacional: técnica de un paso vrs. dos pasos. *Rev.méd.sinerg.* [Internet]. 1 de octubre de 2021 [citado 27 de septiembre de 2023];6(10):e724. Disponible en: <https://revistamedicasinergia.com/index.php/rms/article/view/724>
23. Danawati, C. W., Nagata, M., Moriyama, H., Hara, K., Yasuda, H., Nakayama, M., Kotani, R., Yamada, K., Sakata, M., Kurohara, M., Wiyono, P., Asdie, H. A. H., Sakaue, M., Taniguchi, H., & Yokono, K. (2005). A possible association of Pro12ALA polymorphism in peroxisome proliferator-activated receptor γ 2 gene with obesity in native Javanese in Indonesia. *Diabetes/Metabolism Research and Reviews*, 21(5), 465-469. <https://doi.org/10.1002/dmrr.543>

24. Danawati, C. W., Nagata, M., Moriyama, H., Hara, K., Yasuda, H., Nakayama, M., Kotani, R., Yamada, K., Sakata, M., Kurohara, M., Wiyono, P., Asdie, H. A. H., Sakaue, M., Taniguchi, H., & Yokono, K. (2005b). A possible association of Pro12ALA polymorphism in peroxisome proliferator-activated receptor γ 2 gene with obesity in native Javanese in Indonesia. *Diabetes/Metabolism Research and Reviews*, 21(5), 465-469. <https://doi.org/10.1002/dmrr.543>
25. De la Asociación Española de Pediatría de atención primaria, I. y. A. (s/f). *Actualización en diabetes*. Aepap.org. Recuperado el 25 de septiembre de 2023, de https://www.aepap.org/sites/default/files/pags._445-452_actualizacion_en_diabetes.pdf
26. Dean L, McEntyre J. The Genetic Landscape of Diabetes [Internet]. Bethesda (MD): National Center for Biotechnology Information (US); 2004. Chapter 3, Genetic Factors in Type 2 Diabetes. 2004 Jul 7. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK1665/>
27. Deeb SS, Fajas L, Nemoto M, Pihlajamaki J, Mykkanen L, Kuusisto J, Laakso M, Fujimoto W, Auwerx J: A Pro12Ala substitution in PPAR γ 2 associated with decreased receptor activity, lower body mass index and improved insulin sensitivity. *Nature genetics* 1998, 20(3):284-287.
28. Deeb, S., Fajas, L., Nemoto, M. *et al.* A Pro12Ala substitution in PPAR γ 2 associated with decreased receptor activity, lower body mass index and improved insulin sensitivity. *Nat Genet* 20, 284-287 (1998). <https://doi.org/10.1038/3099>
29. Definition, Diagnosis and Classification of Diabetes Mellitus and its Complications. Part 1: Diagnosis and Classification of Diabetes Mellitus (WHO/NCD/NCS/99.2). Ginebra, Organización Mundial de la Salud, 1999.
30. *Diabetes Tipo 1*. (2019, junio 21). Federación Española de Diabetes FEDE; Federación Española de Diabetes (FEDE). <https://fedesp.es/diabetes/tipos/diabetes-tipo-1/>
31. Diabetes. (2021, 10 noviembre). Organización Mundial del Salud. Recuperado 6 de septiembre de 2023, de <https://www.who.int/es/news-room/fact-sheets/detail/diabetes>
32. Diagnóstico y manejo de la diabetes de tipo 2 (HEARTS-D). Washington, D.C.: Organización Panamericana de la Salud; 2020. Licencia: CC BY-NC-SA 3.0 IGO
33. *Diagnóstico y manejo de la diabetes de tipo 2*. (s/f). Paho.org. Recuperado el 24 de junio de 2023, de https://iris.paho.org/bitstream/handle/10665.2/53007/OPSWN-MHNV200043_spa.pdf?sequence=1&isAllowed=y
34. Douglas JA, Erdos MR, Watanabe RM, Braun A, Johnston CL, Oeth P, Mohlke KL, Valle TT, Ehnholm C, Buchanan TA, Bergman RN, Collins FS, Boehnke M, Tuomilehto J: The peroxisome proliferator-activated receptor- γ 2 Pro12A1a variant: association with type 2 diabetes and trait differences. *Diabetes* 2001, 50(4):886-890.
35. E Castro-Pérez, D E Trejos, T Hrbek, V Setaluri, C W Ramos. Genetic Ancestry of the Panamanian Population:
36. Entendiendo la Diabetes. (2022). Federación Internacional de Diabetes. Recuperado 7 de septiembre de 2023, de <https://worlddiabetesday.org/es/sobre/entendiendo-la-diabetes/>

37. *Epi InfoTM* | CDC. (s. f.). <https://www.cdc.gov/epiinfo/index.html>
38. Especial, C. (2022, 21 marzo). *En nuestro país, aproximadamente 181,779 personas mayores de 18 años viven con diabetes*. En Segundos Panama. <https://ensegundos.com.pa/2022/03/21/en-nuestro-pais-aproximadamente-181779-personas-mayores-de-18-anos-viven-con-diabetes/>
39. Evans, D. *et al.* Association between the P12A and c1431t polymorphisms in the peroxisome proliferator activated receptor gamma (PPAR gamma) gene and type 2 diabetes. *Exp. Clin. Endocrinol. Diabetes* 109, 151–154, 10.1055/s-2001-14838 (2001).
40. Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. Report of the expert committee on the diagnosis and classification of diabetes mellitus. *Diabetes Care*. 2003 Jan;26 Suppl 1:S5–20. doi: 10.2337/diacare.26.2007.s5. PMID: 12502614.
41. Federación Internacional de Diabetes. Atlas de diabetes de la FID, 10ª ed. Bruselas, Bélgica: 2021. Disponible en: <https://www.diabetesatlas.org>
42. Fornage, M. T; Jacobs D. R. J; Steffes M. W.; Gross M. D.; Bray M. S., Schreiner, P. J. 2005. Inverse effects of the PPARc2 Pro12Ala polymorphism on measures of adiposity over 15 years in African Americans and whites The CARDIA study. *Metabolism Clinical and Experimental* 54 910 – 917
43. Frederiksen L. Brødbæk K. Fenger M. Jørgensen T. Borch-Johnsen K. Madsbad S. Urhammer SA. Comentario: estudios del polimorfismo Pro12Ala del gen PPAR-gamma en la cohorte danesa MONICA: la homocigosidad del alelo Ala confiere un menor riesgo de síndrome de resistencia a la insulina. *J Clin Endocrinol Metab.* 2002; 87 : 3989–3992.
44. Fujioka K. Pathophysiology of type 2 diabetes and the role of incretin hormones and beta-cell dysfunction. *JAAPA*. 2007 Dec;Suppl:3–8. doi: 10.1097/01720610-200712000-00001. PMID: 18217245.
45. García-García C. Diabetes mellitus gestacional. *Med Int Mex* 2008;24:148–156.
46. Gardner LI, Stern MP, Haffner SM, Gaskill SP, Hazuda HP, Relethford JH, and Eifler CW: Prevalence of diabetes in Mexican Americans: relationship to percent of gene pool derived from Native American genetic admixture. *Diabetes* 3336–92, 1984.
47. Giné, S. (2021, noviembre 14). ¿Qué es la glucemia y cuál es su valor normal? – Diabetes mellitus. Escuela de Postgrado de Medicina y Sanidad. <https://postgradomedicina.com/glucemia-valores-normales-diabetes/>
48. Gouda HN, Sagoo GS, Harding AH, Yates J, Sandhu MS, Higgins JP. The association between the peroxisome proliferator-activated receptor-gamma2 (PPARG2) Pro12Ala gene variant and type 2 diabetes mellitus: a HuGE review and meta-analysis. *Am J Epidemiol.* 2010 Mar 15;171(6):645–55. doi: 10.1093/aje/kwp450. Epub 2010 Feb 23. PMID: 20179158; PMCID: PMC2834889.
49. Gouda, H. N., Sagoo, G. S., Harding A-H., Yates, J., Sandhu, M. S. & Higgins J. P. 2009. The Association Between the Peroxisome Proliferator-Activated Receptor-g2 (PPARG2) Pro12Ala Gene Variant and Type 2 Diabetes Mellitus: A HuGE Review and Meta-Analysis. *Am. J. Epidemiol* 171:645–655

50. Gurnell M., Savage D., Chatterjee K., and O’Rahilly S. The Metabolic Syndrome: Peroxisome Proliferator-Activated Receptor and Its Therapeutic Modulation. *J. Clin. Endocrinol. Metab.* 2003. 88: 2412-2421.
51. Hara K, Okada T, Tobe K, Yasuda K, Mori Y, Kadowaki H, Hagura R, Akanuma Y, Kimura S, Ito C, Kadowaki T. The Pro12Ala polymorphism in PPAR gamma2 may confer resistance to type 2 diabetes. *Biochem Biophys Res Commun.* 2000 Apr 29;271(1):212-6. doi: 10.1006/bbrc.2000.2605. PMID: 10777704.
52. Hara K, Okada T, Tobe K, Yasuda K, Mori Y, Kadowaki H, Hagura R, Akanuma Y, Kimura S, Ito C, Kadowaki T: The Pro12Ala polymorphism in PPAR gamma2 may confer resistance to type 2 diabetes. *Biochemical and biophysical research communications* 2000, 271(1):212-216.
53. Hegele, R. A., Cao, H., Harris, S. B., Zinman, B., Hanley, A. J. G. & Anderson C. M. 2000. *The Journal of Clinical Endocrinology & Metabolism.* Vol. 85, No. 5. 2014-19
54. Hegele, R. A., Cao, H., Harris, S. B., Zinman, B., Hanley, A. J., & Anderson, C. M. (2000). Peroxisome Proliferator-Activated Receptor- γ 2 P12A and Type 2 diabetes in Canadian Oji-Cree*. *The Journal of Clinical Endocrinology & Metabolism*, 85(5), 2014-2019. <https://doi.org/10.1210/jcem.85.5.6610>
55. Hemminki, K., Li, X., Sundquist, K. & Sundquist, J. Familial risks for type 2 diabetes in Sweden. *Diabetes Care* 33, 293-297 2010.
56. Hur, H.J.; Yang, H.J.; Kim, M.J.; Lee, K.-H.; Kim, M.-S.; Park, S. 2022. Association of Polygenic Variants with Type 2 Diabetes Risk and Their Interaction with Lifestyles in Asians. *Nutrients*, 14, 3222. <https://doi.org/10.3390/nu14153222>
57. IDF SCHOOL OF DIABETES. (s/f). idfdiabeteschool.org. Recuperado el 8 de septiembre de 2023, de <https://idfdiabeteschool.org/type-2-diabetes>
58. Irgam Kumuda, Battini Sriteja Reddy, Sai Gayathri Hari, Swathi Banapuram and Battini Mohan Reddy. The genetic susceptibility profile of type 2 diabetes and reflection of its possible role related to reproductive dysfunctions in the southern Indian population of Hyderabad. 2021. *BMC Medical Genomics.* 14:272
59. Ismail, A., & Tanasova, M. (2022). Importance of GLUT transporters in disease diagnosis and treatment. *International Journal of Molecular Sciences*, 23(15), 8698. <https://doi.org/10.3390/ijms23158698>
60. Iyengar S, Hamman RF, Kamboh MI, Marshall JA, Baxter J, Majumder PP: Amerindian admixture among the Anglo and Hispanic ethnic groups in the San Luis Valley, Colorado. *Am J Phys Anthropol Suppl.* 12:97-98, 1991.
61. Janani C, Ranjitha Kumari BD. PPAR gamma gene – A review. *Diab Met Syndr: Clin Res Rev* (2014), <http://dx.doi.org/10.1016/j.dsx.2014.09.015>
62. Julián Ramírez-Bello, Gilberto Vargas-Alarcón, Carlos Tovilla-Zárate y José Manuel Fragoso. Polimorfismos de un solo nucleótido (SNP): implicaciones funcionales de los SNP reguladores (rsNP) y de los SNP-ARN estructurales (srNP) en enfermedades complejas. *Gaceta Médica de México.* 2013;149: 220-8

63. Kahn CR. Banting Lecture. Insulin action, diabetogenes, and the cause of type II diabetes. *Diabetes*. 1994 Aug;43(8):1066–84. doi: 10.2337/diab.43.8.1066. PMID: 8039601.
64. Katsarou A, Gudbjörnsdottir S, Rawshani A, Dabelea D, Bonifacio E, Anderson BJ, Jacobsen LM, Schatz DA, Lernmark Å. Type 1 diabetes mellitus. *Nat Rev Dis Primers*. 2017 Mar 30;3:17016. doi: 10.1038/nrdp.2017.16. PMID: 28358037.
65. Katsarou A, Gudbjörnsdottir S, Rawshani A, Dabelea D, Bonifacio E, Anderson BJ, Jacobsen LM, Schatz DA, Lernmark Å. Type 1 diabetes mellitus. *Nat Rev Dis Primers*. 2017 Mar 30;3:17016. doi: 10.1038/nrdp.2017.16. PMID: 28358037.
66. Kawate R, Yamakido M, Nishimoto Y, Bennett PH, Hamman RF, and Knowler WC: Diabetes mellitus and its vascular complications in Japanese migrants and on the island of Hawaii. *Diabetes Care* 2:161–170, 1979.
67. Kharroubi AT, Darwish HM. Diabetes mellitus: the epidemic of the century. *World J Diabetes*. 2015;6(6):850.
68. Kirk, B., Feinsod, M., Favis, R., Kliman, R. & Barany, F. 2002. Single Nucleotide Polymorphism seeking long term association with complex disease. *Nucleic Acid Research*. Vol 30 No 15 3295–3311
69. Klimentidis YC, Miller GF, Shriver MD. 2009. The Relationship Between European Genetic Admixture and Body Composition Among Hispanics and Native Americans. *Am. J. Hum. Biol.* 21:377–382. La diabetes, ¿Representa o no un obstáculo para lograr un embarazo mediante reproducción asistida? (2018, noviembre 14). *IVI PA*. <https://ivi.com.pa/blog/la-diabetes-representa-o-no-un-obstaculo-para-lograr-un-embarazo-mediante-reproduccion-asistida/>
70. Li S, He C, Nie H, Pang Q, Wang R, Zeng Z, Song Y. G Allele of the rs1801282 Polymorphism in PPAR γ Gene Confers an Increased Risk of Obesity and Hypercholesterolemia, While T Allele of the rs3856806 Polymorphism Displays a Protective Role Against Dyslipidemia: A Systematic Review and Meta-Analysis. *Front Endocrinol (Lausanne)*. 2022 Jun 29;13:919087. doi: 10.3389/fendo.2022.919087. PMID: 35846293; PMCID: PMC9276935.
71. Ling, C., Bacos, K. & Rönn, T. Epigenetics of type 2 diabetes mellitus and weight change — a tool for precision medicine?. *Nat Rev Endocrinol* 18, 433–448 (2022). <https://doi.org/10.1038/s41574-022-00671-w>
72. Majid, M., Masood, A., Kadla, S. A., Hameed, I. & Ganai, B. A. Association of Pro12Ala polymorphism of peroxisome proliferator-activated receptor gamma 2 (PPAR γ 2) gene with type 2 diabetes mellitus in ethnic Kashmiri population. *Biochem Genet.* 55, 10–21, <https://doi.org/10.1007/s10528-016-9765-6> (2017)
73. Medina-Pérez, EA, Sánchez-Reyes, A, Hernández-Peredo, AR, Martínez-López, MA, Jiménez-Flores, CN, Serrano-Ortiz, I, Maqueda-Pineda, AV, Islas-Cruz, DN, & Cruz-González, M. (2017). Diabetes gestacional. Diagnóstico y tratamiento en el primer nivel de atención. *Medicina interna de México*, 33(1), 91–98. Recuperado el 26 de septiembre de 2023, de http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0186-48662017000100091&lng=es&tlng=es.

74. Mori H, Ikegami H, Kawaguchi Y, Seino S, Yokoi N, Takeda J, Inoue I, Seino Y, Yasuda K, Hanafusa T, Yamagata K, Awata T, Kadowaki T, Hara K, Yamada N, Gotoda T, Iwasaki N, Iwamoto Y, Sanke T, Nanjo K, Oka Y, Matsutani A, Maeda E, Kasuga M: The Pro12 - >Ala substitution in PPAR-gamma is associated with resistance to development of diabetes in the general population: posible involvement in impairment of insulin secretion in individuals with type 2 diabetes. *Diabetes* 2001, 50(4):891-894.
75. Namvaran, F., Rahimi-Moghaddam, P., Azarpira, N. 2010. Genotyping of peroxisome proliferator-activated receptor gamma (PPAR- γ) polymorphism (Pro12Ala) in Iranian population. *JRMS* 2011; 16(3): 291-296
76. OPS. *Guías ALAD de Diagnóstico, Control y Tratamiento de la Diabetes mellitus tipo 2, 2009 (In Spanish)*. (s. f.). OPS/OMS | Organización Panamericana de la Salud. <https://www.paho.org/es/documentos/ops-guias-alad-diagnostico-control-tratamiento-diabetes-mellitus-tipo-2-2009-spanish>
77. Organización Mundial de la Salud, Informe Mundial sobre la Diabetes, noviembre de 2021.
78. Ortega-Domínguez B., Herrera-Ramírez M. y Tecalco-Cruz A. Receptores nucleares: del núcleo al citoplasma. *Revista Especializada en Ciencias Químico-Biológicas*. 2015; 18(2):131-143.
79. Palacios, Anselmo, Durán, Maritza, & Obregón, Oswaldo. (2012). Factores de riesgo para el desarrollo de diabetes tipo 2 y síndrome metabólico. *Revista Venezolana de Endocrinología y Metabolismo*, 10(Supl. 1), 34-40. Recuperado en 01 de septiembre de 2023, de http://ve.scielo.org/scielo.php?script=sci_arttext&pid=S1690-31102012000400006&lng=es&tlng=es.
80. Pérez, Omidres, Saba, Tony, Padrón, Miguel A, & Molina V, Rafael. (2012). Diabetes mellitus gestacional. *Revista Venezolana de Endocrinología y Metabolismo*, 10(Supl. 1), 22-33. Recuperado el 26 de septiembre de 2023, de http://ve.scielo.org/scielo.php?script=sci_arttext&pid=S1690-31102012000400005&lng=es&tlng=es.
81. Pínterová D., Cerná, M., Kolostova, K., Novota, P., Cimburová, M., Romzová, M., Kubena, A., Anđel, M. 2004. *Folia Biologica (Praga)* 50, 153-156
82. Polymorphic Structure, Chibchan Amerindian Genes; and Biological Perspectives on Diseases. *The Internet Journal of Biological Anthropology*. 2016 Volume 9 Number 1.
83. Poulsen P, Andersen G, Fenger M, Hansen T, Echwald T M., Vølund, A., Beck-Nielsen, H., Pedersen, O. & Allan Vaag. 2003. Impact of two common polymorphisms in the PPAR γ gene on glucose tolerance and plasma insulin profiles in monozygotic and dizygotic twins: thrifty genotype, thrifty phenotype, or both? *Diabetes*. 52(1):194-198.
84. Powell E, Kuhn P, Xu W. Nuclear Receptor Cofactors in PPARgamma-Mediated Adipogenesis and Adipocyte Energy Metabolism. *PPAR Res*. 2007;2007:53843. doi: 10.1155/2007/53843. PMID: 17389765; PMCID: PMC1783724.
85. Priya SS, Sankaran R, Ramalingam S, Sairam T, Somasundaram LS. Genotype Phenotype Correlation of Genetic Polymorphism of PPAR Gamma Gene and Therapeutic Response to Pioglitazone in Type 2 Diabetes Mellitus- A Pilot Study. *J Clin Diagn Res*. 2016 Feb;10(2):FC11-4. doi: 10.7860/JCDR/2016/16494.7331. Epub 2016 Feb 1. PMID: 27042481; PMCID: PMC4800546.

86. Pro12Ala polymorphism. *Diabetes* 51 (8), 2341–2347. <https://doi.org/>
87. Quintanilla Rodríguez BS, Mahdy H. *Diabetes Gestacional*. [Actualizado el 8 de agosto de 2023]. En: StatPearls [Internet]. Isla del Tesoro (FL): StatPearls Publishing; 2023 enero-. Disponible en: <https://www.ncbi.nlm.nih.gov/books/NBK545196/>
88. R. Verner Codoceo, *Prevención de diabetes mellitus 2*, *Revista Médica Clínica Las Condes*, Volume 21, Issue 5, 2010, Pages 741-748, ISSN 0716-8640, [https://doi.org/10.1016/S0716-8640\(10\)70595-2](https://doi.org/10.1016/S0716-8640(10)70595-2). (<https://www.sciencedirect.com/science/article/pii/S0716864010705952>)
89. Richard F. Hamman. Genetics and environmental determinants of Non-Insulin Dependent Diabetes Mellitus (NIDDM). *DiabetesIMetabolism Reviews*, Vol. 8, No. 4, 287-338 (1992)
90. Ryan EA, O' Sullivan MJ, Skyler JS. Insulin action during pregnancy: Studies with the euglycemic clamp technique. *Diabetes* 1985;34:380–389
91. Salazar Campos N, Sandí Ovares N, Mejía Arens C. Diabetes mellitus tipo I: retos para alcanzar un óptimo control glicémico. *Rev.méd.sinerg.* [Internet]. 1 de septiembre de 2020;5(9):e452. Disponible en: <https://revistamedicasinergia.com/index.php/rms/article/view/452>
92. Sanghera DK, Ortega L, Han S, Singh J, Ralhan SK, Wander GS, Mehra NK, Mulvihill JJ, Ferrell RE, Nath SK, Kamboh MI. Impact of nine common type 2 diabetes risk polymorphisms in Asian Indian Sikhs: PPAR2 (Pro12Ala), IGF2BP2, TCF7L2 and FTO variants confer a significant risk. *BMC Med Genet*. 2008 Jul 3;9:59. doi: 10.1186/1471-2350-9-59. PMID: 18598350; PMCID: PMC2481250.
93. Sanhueza J., Valenzuela A. Receptores nucleares y regulación de la expresión génica por ácidos grasos poliinsaturados: algo más que producción de energía y esencialidad *Revista Chilena de Nutrición*. 2006; 33(2): 1-18.
94. Sarhangi, N., Sharifi, F., Hashemian, L. *et al*. PPAR2 (Pro12Ala) genetic variant and risk of T2DM: a systematic review and meta-analysis. *Sci Rep* 10, 12764 (2020). <https://doi.org/10.1038/s41598-020-69363-7>
95. Síntomas de la diabetes tipo 2 . (s/f). diabetes.org. Recuperado el 24 de junio de 2023, de <https://diabetes.org/diabetes/type-2/symptoms>
96. Sladek, R. *et al*. A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* 445, 881–885 (2007) [Saxena, R. *et al*. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* 316, 1331–1336 (2007).
97. Souto, J. Estudio del polimorfismo Pro 12Ala en gen PPAR2 y de su relación con la insulino resistencia en una muestra de pacientes con Diabetes Mellitus [en línea] Tesis de grado. Universidad de la República (Uruguay). Facultad de Ciencias 2015
98. Staiger H, Machicao F, Fritsche A, Häring H-U. Pathomechanisms of type 2 diabetes genes. *Endocr Rev*. 2009;30(6):557–85

99. Stryjecki C, Peralta-Romero J, Alyass A, Karam-Araujo R, Suarez F, Gomez-Zamudio J, Burguete-Garcia A, Cruz M & Meyre D. 2015. Association between PPAR- γ 2 Pro12Ala genotype and insulin resistance is modified by circulating lipids in Mexican children. *Scientific Reports* | 6:24472 | DOI: 10.1038/srep24472
100. Tönjes A, Stumvoll M. The role of the Pro12Ala polymorphism in peroxisome proliferator-activated receptor gamma in diabetes risk. *Curr Opin Clin Nutr Metab Care*. 2007 Jul;10(4):410-4. doi: 10.1097/MCO.0b013e3281e389d9. PMID: 17563457.
101. Torun B, Stein AD, Schroeder D, et al. 2002. Rural-tourban migration and cardiovascular disease risk factors in young Guatemalan adults. *International Journal of Epidemiology*. 2002; 31: 218-226
102. Uauy R., Martínez J., Rojas C. Nutrición molecular, papel del sistema PPAR en el metabolismo lipídico y su importancia en obesidad y diabetes mellitus. *Rev. Méd. Chile*. 2000; 128(4):437-446.
103. Vales-Villamarín C, de Dios O, Pérez-Nadador I, Gavela-Pérez T, Soriano-Guillén L and Garcés C (2021) PPARc2 Pro12Ala Polymorphism is Associated in Children With Traits Related to Susceptibility to Type 2 Diabetes. *Front. Pharmacol*. 12:763853. doi: 10.3389/fphar.2021.763853
104. Varillas, C., Blanco, S., Couso, B., Gastelu-Iturri, J., & Reboredo, R. (2005). Diabetes gestacional: su complejidad y repercusión en la evolución del embarazo y salud del recién nacido. *Progresos de obstetricia y ginecología (Internet)*, 48(6), 289-296. [https://doi.org/10.1016/s0304-5013\(05\)72397-9](https://doi.org/10.1016/s0304-5013(05)72397-9)
105. Vegiopoulos, A., Rohm, M. & Herzig, S. Adipose tissue: between the extremes. *EMBO J*. 36, 1999-2017 (2017).
106. Vergotine, Z., Yako, Y. Y., Kengne, A. P., Erasmus, R. T. & Matsha, T. E. Proliferator-activated receptor gamma Pro12Ala interacts with the insulin receptor substrate 1 Gly972Arg and increase the risk of insulin resistance and diabetes in the mixed ancestry population from South Africa. *BMC Genet*. 15, 10, <https://doi.org/10.1186/1471-2156-15-10> (2014).
107. Weiss KM, Ferrell RE, and Hanis CL: A New World syndrome of metabolic disease with a genetic and evolutionary basis. *Yearbook Phys Anthropol* 27:153-178, 1984.
108. West KM: *Epidemiology of Diabetes and its Vascular Lesions*. Elsevier Biomedical Press, New York, 1978.
109. Zárata A, Saucedo R, Hernández M. El receptor de peroxisoma-proliferador-activado gamma (PPAR γ) es un factor de transcripción multigénico de versatilidad metabólica. *Acta Med*. 2005; 3(1):53-54.

C A P Í T U L O 8

Secuenciación genómica del componente amerindio de la población panameña revela factores de riesgo genético a enfermedades no transmisibles

Edgardo Castro-Pérez, Carlos W. Ramos-Delgado, Lucía Jorge-Nebert, Beatrice Di Biase, Kathia Canto-Díaz, Madián Poveda, Germán Espino, Vijayasaradhi Setaluri, Carmen Mela-Sánchez, Magaly de Chial, Olga Chen, Julio C. Castillo, Carlos M. Restrepo, Rolando Gittens, Diomedes E. Trejos, Jorge Lasso De La Vega, Brenda de Mayorga, Gumercindo Lorenzo, Ovidio Durán.

Resumen

Las enfermedades no transmisibles (ENT) como tumores malignos (cánceres), diabetes, enfermedades cardio y cerebrovasculares, entre otras, representan las principales causas de muerte por enfermedades en Panamá. Evidencias indican que el riesgo de padecer de ENT está asociado con disparidades ancestrales en la población panameña, sugiriendo que el componente genético ancestral es un factor determinante en la distribución, riesgo y muertes por ENT en nuestro país. A pesar del alto número de casos y muertes, no existen estudios científicos en Panamá que aborden la epidemiología genética de estas enfermedades principalmente porque no se conocen los polimorfismos genéticos asociados con el riesgo de padecer ENT. La identificación de estos polimorfismos patogénicos tendría importantes implicaciones como comprender la interacción de estos genes con factores ambientales y estilos de vida, y el riesgo de desarrollar ENT. Además, contribuiría al desarrollo de mejores estrategias de salud para identificar grupos o regiones de mayor riesgo a padecer determinadas enfermedades. Para la identificación de estos polimorfismos de riesgo en otras poblaciones del mundo, se han utilizado estrategias de análisis por secuenciación de genomas o exomas. Sin embargo, en Panamá existe diversidad de orígenes genéticos y mezclas ancestrales, principalmente indígena, europea y africana, por lo que estos datos genómicos reportados de otras poblaciones mundiales no son necesariamente aplicables en nuestro país. Adicionalmente, en Panamá no se han realizado estudios de secuenciación genómica, por lo tanto, estos estudios genómicos serían un paso fundamental para comprender estas enfermedades en nuestra población. Para abordar esto, en esta investigación secuenciamos el genoma completo de diez (10) amerindios Ngöbe, quienes son los principales ancestros de los panameños modernos y dominan el componente genético de la población con un 51%. Hipotetizamos que una proporción significativa de los polimorfismos genéticos asociados con ENT en la población panameña podrían tener su origen en esta población amerindia ancestral y estar distribuidos en la población Ngöbe actual. Para investigar esta hipótesis, nos enfocamos en el análisis bioinformático del exoma, lo que nos permitió identificar múltiples marcadores polimórficos posiblemente patogénicos asociados con ENT. Los análisis genómicos demostraron que los marcadores moleculares clínicamente más informativos asociados a patologías fueron los polimorfismos de nucleótidos simples (SNPs), entre los cuales se identificaron dieciséis (16) genes con polimorfismos posi-

blemente patogénicos. Algunas de estas variantes fueron seleccionadas para estudios más detallados como determinar su frecuencia alélica en la población Ngöbe (próximos dos capítulos). La alta frecuencia de estos polimorfismos no sólo tendría un impacto importante en la salud de los Ngöbes, sino también en la población mestiza panameña que representa el mayor grupo del país y exhibe una alta proporción de genes de origen Ngöbe. Los datos genómicos presentados aquí representan los primeros genomas humanos secuenciados en la población panameña y unos de los primeros reportados en la región.

Palabras claves: Secuenciación genómica, análisis de exoma, marcadores moleculares, polimorfismos patogénicos, factores de riesgo genético, genética ancestral Ngöbe.

Abstract

Non-communicable diseases (NCDs) such as malignant tumors (cancers), diabetes, and cardio- and cerebrovascular diseases, among others, represent the leading causes of death from diseases in Panama. Despite the high number of cases and deaths, there are no scientific studies in Panama that address the genetic epidemiology of these diseases, mainly because the genetic polymorphisms associated with a greater risk of suffering from them are unknown. Identifying these pathogenic polymorphisms would have significant implications, such as understanding the interaction between genes and these genes with environmental factors and lifestyles and their roles in the risk of developing NCDs. Knowledge of these polymorphisms would also contribute to creating better health strategies to identify groups or regions at higher risk of suffering from certain diseases. To identify these polymorphisms associated with a higher risk of NCDs in other populations around the world, mainly of European origin, genome or exome sequencing analysis strategies have been used. However, in Panama, there is high diversity in miscegenation and heterogeneity of genetic origins with different ancestral mixtures, mainly indigenous, European, and African, so the genomic data reported from other populations worldwide, primarily European, are not necessarily applicable to the country. Additionally, no genomic sequencing studies have been carried out in the Panamanian population, so it is necessary to analyze local genomes to identify these polymorphisms. Analyzing the genomes of patients and the country's general population would be a fundamental step toward understanding these diseases in the popu-

lation. However, the high heterogeneity and diversity of the population, mostly mestizos, and the highly polygenic and multifactorial nature (genes, environment, and lifestyles) of these diseases make it challenging to generate exhaustive data to interpret all the variables comprehensively. Moreover, evidence indicates that the risk of NCDs is also associated with ancestral disparities in the Panamanian population. This suggests that the ancestral genetic component is a highly determining factor in the country's prevalence, distribution, and deaths from NCDs. Therefore, studying the ancestral genes of the population could be a strategy that leads to identifying these polymorphisms. To address these problems and disparities in the Panamanian population, in this pilot research, we have sequenced the complete genome of ten (10) Ngöbe Amerindians, who are the principal ancestors of modern Panamanians and are the ones who dominate the genetic component of the population with a proportion of 51% as previously demonstrated. Therefore, we hypothesize that a significant magnitude of the genetic polymorphisms associated with NCDs in the Panamanian population could originate in this ancestral Amerindian population and be distributed in the country's general population. Furthermore, genetic reports indicate that the Ngöbes show low levels of diversity and miscegenation, which would facilitate the identification and interpretation of polymorphisms associated with that population. We focused on exome bioinformatic analysis to investigate this hypothesis, which allowed us to identify multiple possible pathogenic markers associated with increased NCD risk. Genomic analyses demonstrated that the most clinically informative molecular markers related to pathologies were single nucleotide polymorphisms (SNPs), among which sixteen (16) genes were identified with polymorphisms previously reported in other populations worldwide as pathogenic. Some of these variants were selected for more detailed studies, such as determining their allelic frequency in the Ngöbe population (next chapter). The data suggest that the high frequency of these polymorphisms, along with other genes and environmental factors such as lifestyle, could significantly impact the prevalence of these diseases not only in the Ngöbes but also in the general population of the country represented mainly by mestizos with a high proportion of genes of Ngöbe origin. To confirm whether or not there is an association between these polymorphisms and a higher risk of suffering from these diseases in the population, ongoing studies analyze the genetic patterns of these SNPs in groups of diagnosed patients and compare them with control or low-risk groups. The genomic data presented here represent the first human genomes sequenced in the Panamanian population and some of the first reported in the region.

Keywords: Genomic sequencing, exome analysis, molecular markers, pathogenic polymorphisms, genetic risk, Ngöbe ancestral genetics.

INTRODUCCIÓN

Las enfermedades no transmisibles (ENT), también conocidas como enfermedades crónicas, son afecciones de larga duración y de lenta progresión. Son enfermedades de etiología multifactorial, desarrollo poco predecible, de distintos factores de riesgo y, salvo raras excepciones, de un origen no infeccioso (OMS, 2005; Kelly et al., 2016; OMS, 2016). La OMS ha estimado que las ENT causan la muerte de 41 millones de personas cada año, un equivalente al 74% de todas las muertes a nivel global. Cuando no conducen a la muerte del enfermo, frecuentemente impactan negativamente la calidad de vida, resultando particularmente invalidantes en diferentes grados (World Health Organization, 2016). Consistente con estas tendencias mundiales, las cifras del Ministerio de Salud de Panamá (MINSA) y del Instituto Nacional de Estadísticas y Censo (INEC, Contraloría General de la República) muestran que la principal causa de muerte por enfermedades en el país son las ENT, siendo el cáncer y las enfermedades cardio-vasculares las más frecuentes, seguido de la diabetes y otras enfermedades metabólicas como obesidad (Ministerio de Salud, 2018; Reyes, 2022). Adicionalmente, es bien conocida la asociación entre condiciones como diabetes, obesidad e hipertensión con el riesgo de padecer de diferentes ENT incluyendo algunos cánceres y enfermedades vasculares, entre otras (Mariscal Davy, 2021). Se estima que el número de pacientes que sufren de ENT aumentará significativamente tanto a nivel mundial como en Panamá, resultado de una mayor expectativa de vida. En Panamá, la atención hacia estas enfermedades ha generado políticas preventivas y de cuidado, incluidas en el Plan Estratégico Nacional de Prevención y Control integral del cáncer, enfermedades no transmisibles y sus factores de riesgo (2014-2019 y 2019-2029, MINSA/CSS/OMS-OPS; Inst. Oncológico, 2019) y en el plan estratégico nacional para el desarrollo de la ciencia, la tecnología e innovación (PENCYT-SENACYT) 2019-2024. Estos planes contemplan, además, los objetivos de desarrollo sostenible (ODS) de la ONU-PNUD, en particular, el objetivo #3 sobre salud y bienestar.

La falta de disponibilidad de estadísticas respecto a las enfermedades no transmisibles que afectan al panameño por provincia, ha dificultado tener una idea acerca de las tendencias de estas enfermedades por región. Sin

embargo, para enfermedades como el cáncer, diabetes y enfermedades cardiovasculares, hay pocas estadísticas disponibles, pero muy generales, las cuales indican que existen considerables diferencias per cápita entre provincias en la prevalencia para este grupo de enfermedades. ¿A qué se deben estas diferencias? ¿Son debidas a factores dietéticos, climáticos o ambientales como la contaminación, estilos de vida diferentes, biológicos como factores genéticos-ancestrales (indígena, europea o africana) como sugieren algunos estudios previos? (Castro-Pérez et al., 2016; Ramos et al., 2018). Por ejemplo, múltiples estudios demuestran que la ancestralidad africana es uno de los factores de riesgo fuertemente asociado con el cáncer de próstata (Bock et al., 2009; Robbins et al., 2007; Zeigler-Johnson et al., 2008). Consistente con estos reportes, en Panamá las provincias de Panamá y Colón muestran la mayor prevalencia per cápita de cáncer de próstata y son también, las provincias con mayor proporción de genes de origen africano (Arias et al., 2002; Castro-Pérez et al., 2016; Ramos et al., 2018). Caso contrario lo representan las enfermedades cardiovasculares, las cuales son más frecuentes en pacientes de ascendencia europea (Cheng et al., 2010; Donnan et al., 2008; Hankey, 1999). Similarmente, en Panamá, las provincias de Herrera y Los Santos tienen la mayor incidencia de muertes cardiovasculares y son estas provincias en las que los estudios genéticos revelan la mayor proporción de genes ancestrales europeos (Arias et al., 2002; Castro-Pérez et al., 2016; Ramos et al., 2018). En cuanto a las enfermedades metabólicas, como diabetes, obesidad y el síndrome metabólico, las evidencias indican que tienen una mayor incidencia en poblaciones indígenas de acuerdo a datos del Centro de Control y Prevención de Enfermedades en los Estados Unidos (Mariscal Davy, 2021). Bocas del Toro, Chiriquí, Panamá y Colón son las provincias con mayor prevalencia de obesidad y en las que dominan genes de origen indígena (Ngöbe) (Sasson et al., 2014; Arias et al., 2002; Castro-Pérez et al., 2016; Ramos et al., 2018). Además, varios estudios indican que la obesidad, diabetes y el síndrome metabólico son condiciones que se asocian con mayor riesgo de algunos cánceres, hipertensión y otras condiciones crónicas (Mariscal Davy, 2021).

En Panamá estos estudios genéticos se han realizado utilizando marcadores generales como microsatélites STR (del inglés Short Tandem Repeats) (Castro-Pérez et al., 2016) e InDels (Inserciones/delecciones) (Ramos et al., 2018) y marcadores proteicos clásicos (Arias et al. 2002). Estos estudios han abordado el problema de la epidemiología genética de manera indirecta con marcadores moleculares generales, encontrando una fuerte asociación entre la distribución y prevalencia de muertes por ENT. A pe-

sar de dicha asociación, no hay hasta el momento estudios genéticos o genómicos que hayan identificado en la población panameña los genes y polimorfismos específicos, y su relación con factores de riesgo ambientales para estas enfermedades. Identificar estos polimorfismos es esencial para comprender la etiología y distribución de estas enfermedades, sus factores de riesgo genético y su relación con riesgos ambientales y estilos de vida, contaminación, etc. Además, esto contribuiría a mejorar su diagnóstico y el posible desarrollo de mejores tratamientos y de políticas de prevención más acordes con la genética autóctona, sentando las bases para una medicina genómica, de precisión y prevención en el país.

Para identificar polimorfismos genéticos asociados a mayor riesgo de padecer ENT, se han realizado numerosos estudios de secuenciación genómica y/o de exoma en diferentes poblaciones del mundo (principalmente europeos y más recientemente asiáticos) generando cientos de reportes y miles de polimorfismos asociados a múltiples enfermedades (Ross et al., 2020). Sin embargo, estos polimorfismos identificados en otras poblaciones del mundo no son necesariamente aplicables a la población panameña con una mezcla ancestral única y muy compleja con genes de origen indígenas, europeos y africanos (Arias et al., 2002; Castro Pérez et al., 2016; Ramos et al., 2018). Además, un típico genoma humano tiene más de 3.5 millones de polimorfismos tan sólo de marcadores tipo SNP, los cuales varían de una población a otra, por lo que es necesario analizar el genoma de cada población para identificar los polimorfismos clínicamente relevantes para dicha población en particular (Ross et al., 2020). Los datos genómicos reportados hasta la fecha corresponden principalmente a genomas secuenciados de poblaciones de ascendencia europea (cerca del 80%) y más recientemente algunas africanas (mayormente afro-americanos, 2%) y asiáticas (10%) (Gurdasani et al., 2019; Sirugo et al., 2019). Este tipo de estudios, donde se secuencian genomas o exomas, no se han realizado en la población panameña y a nivel de poblaciones latinoamericanas e indígenas los datos genómicos son muy pocos (alrededor del 1%) (Gurdasani et al., 2019; Sirugo et al., 2019; Popejoy & Fullerton, 2016). La falta de genomas secuenciados de poblaciones latinoamericanas demuestra que hay muy poca diversidad de genomas disponibles en las bases de datos, por lo que los datos genómicos de otras poblaciones del mundo no son representativos de la genética del panameño con altos niveles de mestizaje. La alta diversidad y heterogeneidad de los genomas panameños, es un factor que dificulta la generación e interpretación de datos genómicos, ya que se requeriría secuenciar muchos genomas para obtener la mayor significancia posible. Además, el carácter altamente poligénico y multifac-

torial (genes, ambiente y estilos de vida) de estas enfermedades, dificulta aún más la interpretación integral de todas las variables y sus roles como factores de riesgo.

Para abordar estos problemas, en esta investigación piloto nos hemos embarcado en identificar por primera vez los polimorfismos genéticos de importancia biomédica asociados a las enfermedades en la población panameña mediante secuenciación del genoma completo de diez (10) individuos de la población Ngöbe. Los reportes previos señalan que, aunque los Ngöbe representan menos del 5% de la población actual del país, sus genes aportan la mayor contribución al trasfondo genético ancestral del panameño, permeando con aproximadamente 51% de genes Ngöbe en el acervo genético de la población (Castro et al., 2007; Castro-Pérez et al., 2016; Ramos et. al., 2018). Dada la alta contribución ancestral de la población Ngöbe en el trasfondo genético de la población actual del país, esto sugiere que gran parte de los genes asociados a mayor riesgo de ENT y problemas biomédicos podrían tener un origen en esta población ancestral (Arias, 1991). Por lo tanto, hipotetizamos que secuenciar el genoma de algunos individuos de la población ancestral Ngöbe sería un buen punto de partida para un estudio piloto que podría conducirnos a la identificación de polimorfismos genéticos asociados a ENT en el panameño. Además, los reportes genéticos indican que la población Ngöbe muestra muy bajos niveles de diversidad y mestizaje, por lo que es una población homogénea que facilitaría la identificación de polimorfismos asociados de manera específica a la misma y su interpretación.

Aunque secuenciamos el genoma completo de los diez individuos Ngöbe, los análisis bioinformáticos para identificar polimorfismos patogénicos de riesgo a ENT los enfocamos en el exoma (regiones codificantes de proteínas y RNAs), ya que las evidencias indican que aproximadamente, el 80-90% de las enfermedades conocidas son causadas a nivel de exones (Sant Joan de Déu Barcelona Hospital, 2014). Se analizaron cuatro marcadores moleculares polimórficos incluyendo variantes estructurales (SVs), variantes en número de copias (CNVs), inserciones/delecciones (InDels) y polimorfismos de nucleótidos simples (SNPs). Los análisis determinaron que los marcadores moleculares más informativos clínicamente y posiblemente asociados a ENT fueron los SNPs. Entre ellos, se identificaron dieciséis (16) genes con polimorfismos clínicamente más relevantes, algunas de las cuales fueron seleccionadas para estudios más detallados, parte de los cuales son presentados en el siguiente capítulo. La alta frecuencia de estos polimorfismos patogénicos en la población Ngöbe im-

plicaría que estos polimorfismos, en conjunto con otros genes y factores ambientales como estilos de vida, podrían contribuir en la prevalencia y distribución de estas enfermedades no sólo de esta población indígena, sino también en la población general del país mayormente mestiza (alrededor del 70%), la cual exhibe una alta proporción (51%) de genes ancestrales de origen Ngöbe. Los datos genómicos presentados aquí representan los primeros genomas humanos secuenciados en la población panameña y unos de los primeros reportados en la región.

Metodología

Obtención de muestras y análisis preliminar de ADN

Para el proyecto se utilizaron muestras de ADN amerindio (Ngöbe) de individuos no relacionados, en al menos el cuarto nivel de consanguinidad, obtenidas a partir de sangre periférica. Estas muestras son parte del banco de ADN del Departamento de Genética y Biología Molecular de la Universidad de Panamá colectados en proyectos desarrollados y publicados previamente (Arias et al., 1993; Jorge-Nebert et al., 2002; Petersen et al., 1991; Castro et al., 2007). Se verificó la concentración (mayor o igual a 100ng/ μ L como mínimo) y la razón de pureza por contaminación por proteínas (260/280nm) y por otros contaminantes (260/230nm) en las muestras por espectrofotométrica en NanoDrop (ThermoFisher). La calidad del ADN se evaluó por electroforesis en agarosa al 1% teñida con GelRed en amortiguador TAE1X, donde se escogieron muestras con bandas de alto peso molecular (banda de alrededor de 20Kb) como indicativo de buena calidad (poca degradación).

Sexado de las muestras de ADN mediante detección del Cromosoma Y por PCR

Con el objetivo de garantizar un balance en la proporción de genomas de varón y mujer secuenciados, se verificó el sexo de las muestras mediante detección del cromosoma Y por PCR. Para identificar el linaje masculino de las muestras se utilizaron un par de cebadores de oligonucleótidos: Y1 y Y2, que flanquean un fragmento de 170 pb de las repeticiones alfoides

del cromosoma Y humano (Y1: ATG ATA GAA CGG AAA TAT G; Y2: AGT AGA ATG CAA AGG GCT CC, (Witt & Erickson, 1989; Wolfe et al., 1985). Los cebadores fueron sintetizados por la compañía IDTDNA Technologies y purificados mediante precipitación estándar (*standard desalting*). La mezcla de reacción para amplificar esta región del cromosoma Y consistió en 15 µL de Mastermix (2X Blastaq™ PCR Taq MasterMix, abm Cat. #G895), 3 µL de cada primer (5 µM cada uno), 1 µL de ADN y 8 µL de agua libre de nucleasas en un volumen final de 30 µL. La PCR se realizó en un termociclador Applied Biosystems 2720 y las condiciones utilizadas para amplificar el cromosoma Y fueron 94 °C durante 8 min, 30 ciclos a 94 °C por 1 min, 55 °C por 1 min, 72 °C por 2 min, y al final 12 °C por un tiempo indefinido. Los productos de amplificación de 170 pb fueron resueltos en electroforesis de agarosa al 1% teñida con GelRed en tampón TAE 1X. Como control utilizamos un estándar de peso molecular (100bp ladder). Como control positivo de amplificación utilizamos el gen nuclear PPAR γ 2, del cual teníamos cebadores disponibles en el laboratorio: forward 5'- AAG GAA TCG CTT TCC G-3' y un cebador reverse 5'- GCC AAT TCA AGC CCA GTC -3' reportados previamente (Priya et al., 2016). La mezcla de reacción en la amplificación del gen control consistió en 15 µL de Mastermix (abm Cat.#G895), 2 µL de ambos primers (10µM cada uno), 1 µL de ADN y 10 µL de agua libre de nucleasas en un volumen final de 30 µL. Las condiciones empleadas en el termociclador (Applied Biosystems 2720) para la amplificación del gen control fueron 94 °C durante 8 min, 35 ciclos a 94 °C por 50s, 50 °C por 50s, 72 °C por 1 min, una extensión final a 72 °C por 5min y al final 12 °C por un tiempo indefinido. El producto de 270 pb fue resuelto en electroforesis de agarosa al 1% teñida con GelRed en amortiguador TAE 1X. Como control utilizamos un estándar de peso molecular (100 bp ladder). Todos los resultados de las electroforesis tanto muestras de ADN como amplificaciones por PCR fueron documentados con un Gel Doc™ EZ System (BioRad).

Secuenciación NGS en Illumina

Una vez se determinó el sexo de las muestras, se escogieron cinco (5) muestras identificadas como varones y cinco (5) identificadas como féminas. Las muestras fueron preparadas para envío a Estados Unidos para los servicios de secuenciación de nueva generación (NGS) del genoma completo y análisis bioinformático preliminar. El proceso de secuenciación de nueva generación (NGS, del inglés *Next Generation Sequencing*) fue

llevado a cabo por la compañía Novogene, utilizando los “kits” de la empresa Illumina, fabricante de secuenciadores de NGS (Folletos de Illumina, 2014/2015; Westenberger, 2020).

Brevemente, las muestras son procesadas para verificar con controles de calidad adicionales mediante análisis de concentración con el sistema “Qubit” y para la calidad se utiliza el sistema “TapeStation”. Las muestras fueron luego fragmentadas por endonucleasas a fin de generar extremos pegajosos compatibles con oligonucleótidos adaptadores que son unidos con ADN ligasa. Estos fragmentos con adaptadores fueron amplificados por PCR, purificados en columnas y se seleccionan por calidad y tamaño mediante el Qubit y el sistema TapeStation. Los fragmentos purificados constituyen la biblioteca genómica, los cuales son ligados a los chips de secuenciación de Illumina. La secuenciación NGS genera racimos (“clusters”) de miles de fragmentos de ADN que son sintetizadas y secuenciadas *in situ*. Los fragmentos secuenciados generan lecturas (*reads*) tienen una longitud de entre 200 y 400 pares de bases (Folletos de Illumina, 2014/2015; Westenberger, 2020; Rubio et al., 2020).

Análisis bioinformático NGS preliminar

Los análisis preliminares fueron realizados por Novogene con programas de análisis y plataformas de supercomputadoras. Los análisis bioinformáticos preliminares realizados por Novogene incluyeron control de calidad de las secuencias, alineamiento con el genoma humano de referencia, estadísticos de profundidad y de cobertura de secuenciación (Rubio et al., 2020). Además, incluyó la identificación preliminar (datos crudos sin procesar), anotación y estadísticas de miles de marcadores moleculares de cuatro tipos SV, CNV, InDels y SNPs (explicados más adelante), para cada genoma secuenciado.

Para el control de calidad se descartaron las lecturas de extremos emparejados que contenían contaminación del adaptador, y se determinó el total de lecturas que se secuenciaron (*reads*) y cuántas de ellas eran limpias o contenían secuencias no específicas (“N” o con alguna otra ambigüedad). Para garantizar los análisis posteriores, se requiere que la calidad de la mayoría de las bases sea superior a Q20, que indica una tasa de error de secuenciación máxima de 1/100, así como una tasa de secuenciación correcta del 99%. De acuerdo con la característica de secuenciación de la plataforma Illumina, para generar un buen análisis de los datos es reque-

rido que el porcentaje medio de las secuencias sea de Q30, es decir, una tasa de error de secuenciación máxima de 1/1000.

Alineamiento y mapeo con el genoma humano de referencia

Se mapearon las lecturas limpias de extremos emparejados y se alinearon con el genoma humano de referencia con *Burrows-Wheeler Aligner* (BWA v0.7.17) (Li y Durbin, 2009) y para exportar los archivos alineados en formato “BAM”. Los archivos BAM se ordenaron con SAMtools (Samtools v1.8) (Li et al., 2009; Danecek et al., 2021) y las lecturas duplicadas se marcaron con Picard (Picard v2.18.9) (<https://broadinstitute.github.io/picard/>). Luego, se obtuvieron los archivos BAM finales y después se calculó la cobertura y la profundidad de cobertura con base en BAM para cada genoma (Rubio et al., 2020).

Marcadores moleculares analizados

Se identificaron y anotaron miles de marcadores moleculares de cuatro tipos (SV, CNV, InDels y SNPs) para cada genoma secuenciado. A continuación, explicamos brevemente cada uno de estos marcadores moleculares. Las variaciones estructurales (SV, del inglés *Structural Variations*) son variantes genéticas de tamaño relativamente grande (>50 pb) e incluyen cinco tipos: translocaciones (desplazamiento de un segmento de ADN a un nuevo lugar en el mismo cromosoma u otros cromosomas en el genoma), deleciones (pérdida de un segmento de ADN), duplicaciones, llamadas también amplificaciones (aumento de una o más copias adicionales de un fragmento de ADN, y se repiten justo al lado del fragmento original), inserciones (adición de uno o más nucleótidos en un segmento de ADN) e inversiones (un segmento de ADN se desprende y se vuelve a unir dentro de su posición original, pero con la orientación opuesta). Los InDels (inserciones/deleciones) son otro tipo de marcadores moleculares, pero de tamaño menor de 50 pb. Las variaciones en número de copias (CNVs, del inglés *Copy Number Variations*) son marcadores genéticos que conducen a variaciones en el número de copias de fragmentos relativamente largos (más de 50 pb) entre individuos. Hay dos tipos de CNVs: ganancia y pérdida de copias. Los polimorfismos de nucleótidos simples (SNPs, del inglés *Single Nucleotide Polimorphisms*), también conocidos como variantes de nucleótidos simples (SNVs, del inglés *Single Nucleotide Variations*), constituyen la clase más amplia de variantes genéticas en el genoma y consisten en la sustitución (reemplazo) de un nucleótido por

otro y pueden ser de dos tipos: transiciones (sustitución de una base púrica por otra base púrica [A o G], o de una base pirimidínica por otra base pirimidínica [C o T]); o transversiones (sustitución de una purina [A o G] por una pirimidina [C o T] o viceversa) (Collins y Jukes, 1994; Klug et al., 2006; Griffiths et al., 2008).

Mapeo, identificación y anotación de marcadores polimórficos

Luego del mapeo y alineamiento con el genoma de referencia, se realizaron las anotaciones de variantes con la herramienta ANNOVAR v2015Dec14 (Wang et al., 2010) incluyendo cambios en la codificación de proteínas, regiones genómicas afectadas por las variantes, comparación en frecuencia alélica con poblaciones de referencia, predicción de patogenicidad, riesgo, asociación, etc.

Las variantes genéticas de interés biomédico y clínico de SNPs e InDels fueron identificadas usando Genome Analysis Tool Kit (GATK v4.0) Haplotype Caller (McKenna et al., 2010; Poplin et al., 2017). Los SVs fueron detectados utilizando DELLY v0.8.7 (Rausch et al., 2012); y los CNVs fueron detectados utilizando control-FREEC v11.4 (Boeva et al., 2011; Boeva et al., 2012). Las regiones de interés se enfatizaron creando intervalos utilizando Bedtools (Quinlan y Hall, 2010) y la profundidad de cobertura de secuenciación de los intervalos fue determinada utilizando GATK (profundidad de cobertura). Estos análisis generaron enormes tablas de MS Excel similares a la **Tabla 1** (ejemplo enfocado en SNPs) para cada marcador molecular (SV, CNV, InDels, SNP) y cada genoma que fueron sometidas a análisis más exhaustivos de miles de genes candidatos utilizando estrategias de *big data* con filtros de MS Excel que refinaron la identificación precisa de los genes candidatos posiblemente implicados en ENT.

Inicialmente se desconocía cuál de los cuatro marcadores resultaría en la identificación de variantes asociadas a ENT, por lo que decidimos estudiarlos todos. Estos análisis consistieron en una búsqueda jerárquica de varias selecciones de genes utilizando diferentes criterios recomendados por la literatura para el tipo de análisis en varias etapas. En particular, implementamos los parámetros y criterios recomendados por las Guías y Estándares de la Asociación Médica Americana de Genética y Genómica Médica para el análisis e interpretación de variantes moleculares (Li et al., 2017; Richards et al., 2015) y usando criterios de las bases de datos

clínicas que recomiendan estas guías. Por ejemplo, en la **Tabla 1** presenta datos de SNPs, donde en la columna Func (función), se filtró seleccionando *exonic* pues es en los exones donde se encuentran los cambios de aminoácidos, que están evidenciados en la columna “AAChange”. Sucesivamente, se hizo un filtro en la columna “CLNSIG” (significancia clínica) para que resultaran solo las variantes con las opciones de filtro: “*affects, Affects_association, association, association_risk factor, drug response, pathogenic, pathogenic protective y risk factor*”. Después de la detección de variantes genómicas, se realizó la anotación de variantes con la herramienta ANNOVAR (Wang et al., 2010). Esta información generada en las tablas fue, además, contrastada con otros criterios de las bases de datos clínicas en múltiples aspectos, incluidos cambios en la codificación de proteínas, regiones genómicas afectadas por las variantes, frecuencia de alelos, predicción de efectos nocivos, entre otros. Las principales bases de datos utilizadas fueron las siguientes:

1. Se utilizaron las bases de datos “RefSeq” y “Gencode” para identificar regiones genómicas afectadas por las variantes y posibles cambios en la proteína.
2. Se anotaron las características de las regiones genómicas afectadas por las variantes, como las citobandas (*cytoband*), ARNs pequeños, sitios diana regulatorios de microRNA mamífero conservados, regiones conservadas de vertebrados, sitios de unión de factor de transcripción, repeticiones, etc.
3. La predicción de la nocividad de mutaciones se realizó con los puntajes de SIFT (*Sorting Intolerant From Tolerant*) (Ng y Henikoff, 2003); PolyPhen2 (*Polymorphism Phenotyping v2*, Adzhubei et al., 2013), MutationAssessor (Reva et al., 2011), LRT (*likelihood ratio test*) y CADD (*Combined Annotation Dependent Depletion*, Kircher et al., 2014). Se usaron los puntajes de GERP++ (*Genomic Evolutionary Rate Profiling*, Davydov et al., 2010) para acceder a las conservaciones de mutaciones.
4. A fin de encontrar las frecuencias alélicas alternativas en poblaciones reportadas previamente, se usaron bases de datos establecidas como 1000 Human Genome, Exome Aggregation Consortium (ExAC), Genome Aggregation Database (gnomAD) y exome sequencing Project (ESP).
5. Se usaron las bases de datos dbSNP (SNP data base), COSMIC (*Catalogue Of Somatic Mutations In Cancer*), (Tate et al., 2019), OMIM (*Online*

Mendelian Inheritance In Man), GWAS (*Genome Wide Association Studies*) Catalog (Sollis et al., 2023) y HGMD (*Human Gene Mutation Database*), (Stenson et al., 2003) para encontrar información reportada sobre las variantes, como los principales (top) SNPs en GWAS y asociaciones a cáncer u otras enfermedades.

- Se proporcionó información adicional sobre la anotación funcional o de las vías aplicando las bases de datos, incluyendo *Gene Ontology*, KEGG (*Kyoto Encyclopedia of Genes and Genomes*, Kanehisa et al., 2008), Reactome (Croft et al., 2011), Biocarta y PID (*Pathway Interaction Database*, Schaefer et al., 2009).

Tabla 1. Algunos criterios de análisis para identificar polimorfismos patogénicos.

GeneN	Func	Gene	ExonicFunc	AAChange	CLNDN	CLNSIG
MTHFR	exonic	NM_001330358,NM_001330358	missense SNV	MTHFR:NM_001330358:exon5:c.788T>p.A263V,I	Neoplasm_of_stomach Gastrointestin	drug_response
SDC3	exonic	NM_014654	missense SNV	SDC3:NM_014654:exon4:c.C986T>p.T329I	Obesity_association_with	association
FAAH	exonic	NM_001441	missense SNV	FAAH:NM_001441:exon3:c.C385A>p.P129T	Polysubstance_abuse_susceptibility_t	risk_factor
IL6R	exonic	NM_000565	missense SNV	IL6R:NM_000565:exon9:c.A1073C>p.D358A	Interleukin_6_serum_level_of_quant	association
FCGR2A	exonic	NM_001136219,NM_021136219	missense SNV	FCGR2A:NM_001136219:exon4:c.A500G>p.H167R	Lupus_nephritis_susceptibility_to Pse	drug_response
EPHX1	exonic	NM_000120,NM_001136219	missense SNV	EPHX1:NM_000120:exon3:c.T337C>p.Y113H,EPHX	EPOXI DE_HYDROLASE_1_POLYMORPHI	drug_response
GCKR	exonic	NM_001486	missense SNV	GCKR:NM_001486:exon15:c.T1337C>p.L446P	Fasting_plasma_glucose_level_quantit	association
SLC11A1	exonic	NM_000578	missense SNV	SLC11A1:NM_000578:exon15:c.G1627A>p.D543N	Buruli_ulcer_susceptibility_to	risk_factor
GHRL	exonic	NM_001134941,NM_001134941	missense SNV	GHRL:NM_001134941:exon2:c.A116T>p.Q39L,GH	Obesity	risk_factor
XPC	exonic	NM_004628	missense SNV	XPC:NM_004628:exon16:c.C2815A>p.Q939K	Xeroderma_pigmentosum not_specifi	drug_response
ADD1	exonic	NM_001119,NM_001281	missense SNV	ADD1:NM_001119:exon10:c.G1378T>p.G460W,AI	Hypertension_salt-sensitive_essential	drug_response
TLR1	exonic	NM_003263	missense SNV	TLR1:NM_003263:exon4:c.A743G>p.N248S	Leprosy_5	risk_factor
UGT2B15	exonic	NM_001076	missense SNV	UGT2B15:NM_001076:exon1:c.T253G>p.Y85D	oxazepam_response_-_Other lorazep	drug_response
ADRB2	exonic	NM_000024	missense SNV	ADRB2:NM_000024:exon1:c.G46A>p.G16R	salbutamol_response_-_Efficacy salm	drug_response
PPARGC1B	exonic	NM_001172698,NM_001172698	missense SNV	PPARGC1B:NM_001172698:exon4:c.G490C>p.A16	Obesity_variation_in	association
OR2J3	exonic	NM_001005216	missense SNV	OR2J3:NM_001005216:exon1:c.A337G>p.T113A	C3HEX_ability_to_smell	Affects
OR2J3	exonic	NM_001005216	missense SNV	OR2J3:NM_001005216:exon1:c.G677A>p.R226Q	C3HEX_ability_to_smell	Affects
PLA2G7	exonic	NM_001168357,NM_001168357	missense SNV	PLA2G7:NM_001168357:exon11:c.T1136C>p.V379	Asthma_and_atopy_susceptibility_to	risk_factor

Selección de SNPs clínicamente patogénicos asociados a ENT

Los criterios y parámetros descritos arriba fueron analizados en el conjunto de los datos de las diferentes variantes/polimorfismos, identificados preliminarmente en las tablas de MS Excel para cada uno de los cuatro marcadores moleculares para cada genoma con el objetivo de generar una lista más reducida con los genes y variantes de mayor significancia clínica y probabilidad de tener un papel patogénico/riesgo a ENTs. Luego de realizar estos análisis exhaustivos basados en los criterios mencionados se determinó que los marcadores CNV, SV e InDel no eran muy informativos en términos de significancia biomédica relacionadas a ENT según las bases de datos clínicas (ver resultados). Sin embargo, los SNPs sí mostraron significancia clínica patogénica en riesgo, por lo que enfocamos el resto de los análisis basados en marcadores SNPs.

Para los análisis de SNPs, filtramos los datos de la columna CLNSIG, que indicaba la significancia clínica de cada variante. Centramos nuestra atención en todas las variantes que tenían los valores de *Affects*, *Affects_association*, *association*, *association_risk factor*, *drug response*, *pathogenic*, *pathogenic protective* y *risk factor*, que son términos y criterios que se usan en las guías para describir variantes identificadas en genes que causan trastornos mendelianos (Richards et al., 2015). Revisamos todos los genes que estaban asociados a condiciones patogénicas como cánceres, diabetes y enfermedades cardiovasculares, entre otras, en la columna CLNDN (ClinVar, *Clinical Variants*). Para cada gen y sus variantes polimórficas de esas categorías, nos enfocamos en los datos de las columnas como SIFT, Polyphen2_HDIV, Polyphen2_HVAR, LRT, Mutation Taster (Schwarz et al., 2014), MutationAssessor y FATHMM (*Functional Analysis through Hidden Markov Models*, Rogers et al., 2018) para identificar cuáles y cuántas variantes funcionales o deletéreas resultaban conectadas a ENT. También investigamos esas variantes polimórficas en distintas bases de datos, que comprendían dbSNP, ClinVar, OMIM, UniProt y VarSome para verificar si la información sobre su patogenicidad estaba actualizada. Un criterio adicional que utilizamos es que la variante identificada estuviese presente en al menos la mitad, es decir, cinco (5) de los diez genomas secuenciados. La presencia del polimorfismo en cinco o más de los genomas podría indicar una posible mayor probabilidad de que el mismo estuviese en alta frecuencia no sólo en la población (ancestral) Ngöbe, sino también en el resto de la población mestiza, que representa la mayor parte del país y exhibe un alto trasfondo genético de origen Ngöbe (Castro-Pérez et al., 2016; Ramos et al., 2018). Adicionalmente, se investigó información sobre estos genes/polimorfismos en la literatura para confirmar cuáles estaban más asociados a ENT. Los múltiples análisis realizados de manera exhaustiva, las comparaciones con varias bases de datos, así como parámetros y criterios explicados arriba, resultaron en un listado de 16 genes con variantes/polimorfismos SNPs patogénicos de riesgo que según estos análisis mostraron mayor probabilidad de ser asociados a diferentes ENT (ver resultados, **Tabla 5**).

Resultados

Concentración y Pureza del ADN

Se cuantificaron las muestras mediante NanoDrop, de las cuales se presentan algunas muestras representativas en la **Tabla 2**. Todas las muestras de ADN mostraron una concentración superior de 100 ng/μL. En cuanto a la razón 260/280nm y la razón de 260/230nm, la mayoría mostraron valores superiores a 1.8 para 260/280nm y entre 1.0 y 2.0 para 260/230nm. Todas estas mediciones generaron valores que estaban dentro de los parámetros mínimos de acuerdo con los estándares de calidad requeridos por la compañía que realizaría la secuenciación genómica Novogene (**Tabla 2**).

Tabla 2. Concentración y pureza de ADN determinadas en NanoDrop.

#	Muestra/Código	ng/μL	260/280nm	260/230nm
1	P11/2	420.5	1.81	1.91
2	732/2	199.7	1.82	1.27
3	P2/3	1008.7	1.80	2.06
4	P2	342.1	1.83	0.92
5	P6	242.8	1.83	0.74
6	736D	160.6	1.80	2.01
7	852	143.3	1.81	1.78

Calidad del ADN

Se seleccionaron las muestras con mejor concentración para analizar la calidad de las mismas mediante electroforesis en agarosa. Esta técnica nos permitió verificar si había degradación del ADN, es decir, bajo peso molecular o manchas de *smear*. Los análisis demostraron que la mayoría de las muestras tenían ADN de alto peso molecular con muy poca o ninguna señal de degradación. La **Figura 1** evidencia algunas muestras representativas de la buena calidad del ADN. La buena calidad de las muestras de ADN fue confirmada mediante análisis en TapeStation por Novogene (datos no mostrados).

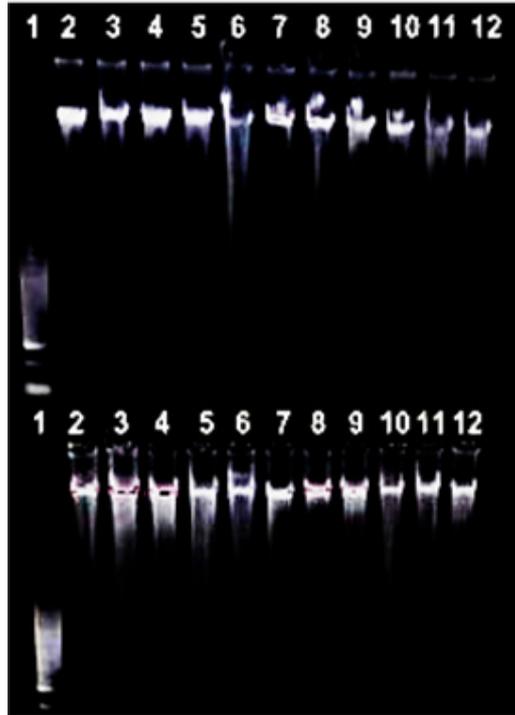


Figura 1. Electroforesis en agarosa de ADN genómico. **Arriba:** Carril 1 100bp PCR Ladder; carriles 2-12: muestras 709/2, 717, 727/2, 723/2, P14/2, 713/2, 722/2, 846, 712/2, 842, 718/2. **Abajo:** Carril 1 100 bp PCR Ladder; carriles 2-12: muestras 847, 714, 844, 713, 728/2, 836, 850, 843b, 706/3, 729, 711/2. Las muestras presentaron bandas de alto peso molecular.

Sexado de las muestras mediante detección del cromosoma Y por PCR

El sexado de las muestras se realizó mediante identificación molecular del cromosoma Y por PCR, amplificando un fragmento de 170 pb de las repeticiones alfoides de dicho cromosoma. El gen nuclear PPAR γ 2 fue utilizado como control positivo de amplificación. La **Figura 2** y la **Tabla 3** evidencian los resultados de algunas muestras representativas la amplificación exitosa de ambos genes y la determinación del sexo de cada muestra, respectivamente. Como se esperaba, todas las muestras amplificaron el gen

control PPAR γ 2, excepto el Master Mix control negativo (C-), mientras que en el cromosoma Y amplificaron las muestras P11/2, P2/3, P/2, 852, P13/2, P10/2 y P6/2 resultaron ser varones; mientras que las muestras 732/2, P6 y 736D resultaron ser femeninas.

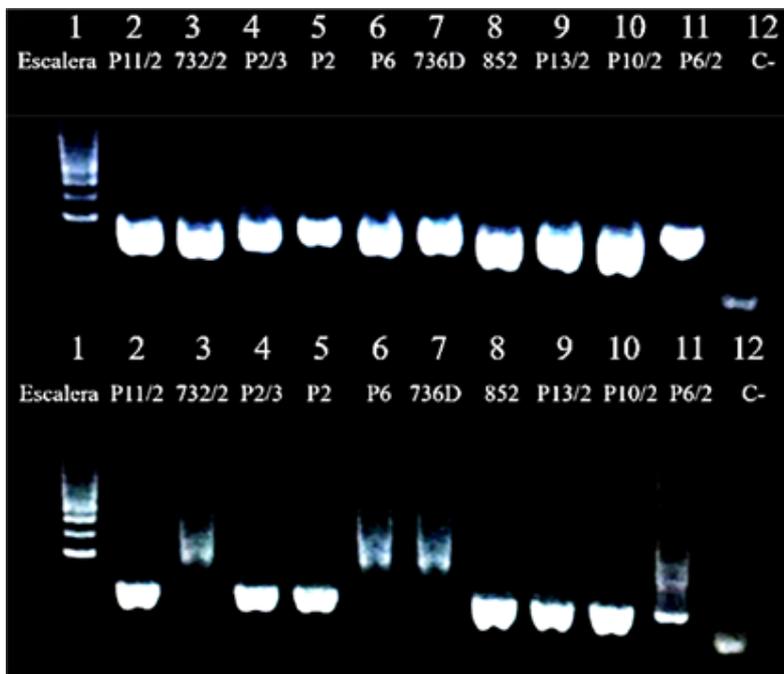


Figura 2. Electroforesis en gel de agarosa de PCR del gen nuclear control (arriba) y el fragmento del cromosoma Y (abajo). Arriba: gen control PPAR γ 2, de izquierda a derecha muestras P11/2, 732/2, P2/3, P2, P6, 736D, 852, P13/2, P10/2, P6/2 y Master Mix control negativo. Abajo: PCR de cromosoma Y, de izquierda a derecha muestras: P11/2, 732/2, P2/3, P2, P6, 736D, 852, P13/2, P10/2, P6/2 y Master Mix control negativo (C-).

Tabla 3. Resultados representativos de los análisis de PCR con el control de amplificación (gen nuclear PPAR γ 2) y PCR del cromosoma Y verificados en electroforesis en agarosa.

Carril	Código	PPAR γ 2	Cromosoma Y	Resultado
1	Ladder	N.A.	N.A.	N.A.
2	P11/2	+	+	Varón
3	732/2	+	-	Fémica
4	P2/3	+	+	Varón
5	P2	+	+	Varón
6	P6	+	-	Fémica
7	736D	+	-	Fémica
8	852	+	+	Varón
9	P13/2	+	+	Varón
10	P10/2	+	+	Varón
11	P6/2	+	+	Varón
12	M.Mix control -	-	-	N.A.

Análisis de secuencias genómicas de nueva generación:

Tasa de error, puntuación de calidad y composición de bases.

La **Figura 3** muestra gráficas representativas de la distribución de errores en cada posición de cada lectura/read (*error rate distribution along reads*), así como su puntuación de calidad (**Q**, *quality score distribution along reads*) y el contenido de bases G/C y A/T. Los análisis demuestran que la tasa de error de las secuencias estuvo en 0.03% en todos los genomas (**Figura 3A**, **Tabla 4**), mientras que la puntuación de calidad fue superior a Q30 en más del 90% de las secuencias (**Figura 3B** y **Tabla 4**). En cuanto a la composición de bases, las mismas estuvieron alrededor del 30% cada una para A/T (A + T = 60%) y alrededor del 20% cada una para G/C (G + C = 40%) con un promedio de 41.335% y bases no específicas "N" menos del 0.00% (**Figura 3C** y **Tabla 4**). Esto demuestra que no hay sesgo en el contenido del apareamiento de bases de los genomas secuenciados.

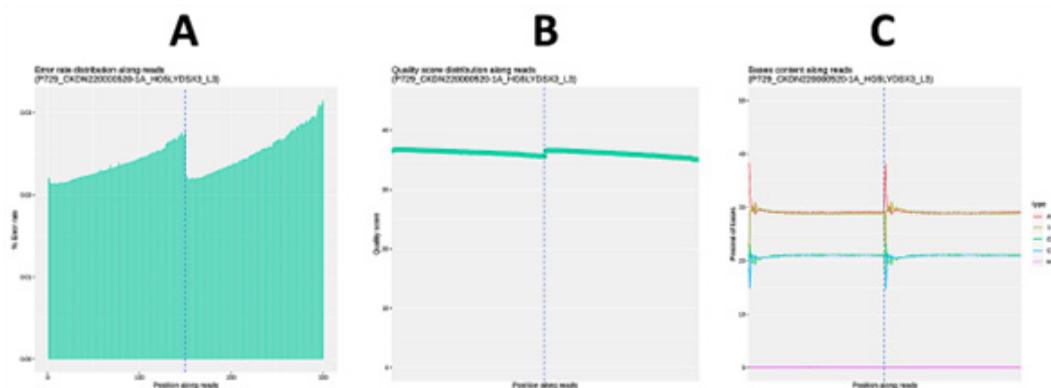


Figura 3. Gráficas representativas de la distribución en tasa de error, calidad y contenido de bases a lo largo de las lecturas (reads). **A.** Tasa de error a lo largo de las lecturas muestra que el error fue entre 0.02% y 0.035%. **B.** La puntuación de calidad (Q) fue superior a Q30 en más del 90% de las secuencias (Tabla 4). **C.** El contenido de bases muestra proporciones iguales de A/T (30% cada una) y G/C (20% cada una, promedio de 41.335%), como se esperaba, no hay discrepancias en proporción de bases.

Las lecturas fueron analizadas para determinar posibles errores y eliminar cualquier secuencia contaminante relacionada con los adaptadores. Los genomas secuenciados con códigos P729, P11/2, P723/2, P849 y P728/2 corresponden a varones (hombres, H) y los genomas con códigos P732/2, P722/2, P14/2, P713 y P715 corresponden a féminas (mujeres, M). Los análisis demostraron que todos los genomas presentaron un porcentaje superior al 99.83% de lecturas limpias, menos del 0.00% de las secuencias eran no específicas (“N”) y menos del 0.00% eran secuencias de baja calidad, mientras que tan solo menos del 0.17% fueron secuencias asociadas con los adaptadores (Figura 4). Estos datos demuestran que los genomas presentaron excelente calidad con secuencias obtenidas altamente confiables.

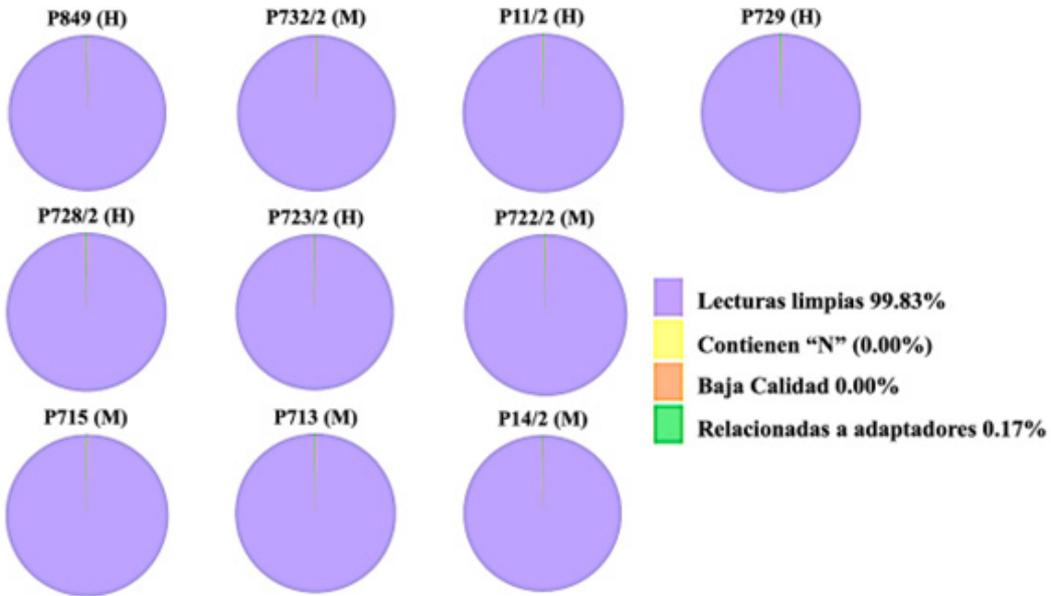


Figura 4. Gráficas de pastel mostrando la calidad en las lecturas (sin procesar) de las secuencias genómicas. En el código "H" indica genoma hombre (varón) y "M" indica genoma de mujer (fémica). Las gráficas ilustran que los datos de los diez genomas secuenciados son de alta calidad con aproximadamente 99.83% de lecturas limpias (violeta) y el porcentaje de error de 0.03% (Tabla 4). En todos los genomas el porcentaje de secuencias no específicas ("N", amarillo) o de baja calidad (anaranjado) es menor de 0.00%. El porcentaje de secuencias relacionadas con adaptadores es del 0.17% (verde).

Mapeo con el genoma de referencia, cobertura y profundidad

La **Figura 5** muestra la distribución de profundidad de todas las bases en cada genoma, la cual fue en promedio 32 entre todas las lecturas (**5A**, **Tabla 4**) y la distribución acumulativa de profundidades en todas las bases de cada genoma, demostrando que el 100% de las bases mapeadas tuvieron profundidad mayor de cero (**5B**).

La **Tabla 4** muestra las estadísticas de calidad, mapeo, cobertura y profundidad en cada genoma secuenciado. Para estos análisis se utilizaron solamente las lecturas limpias. La puntuación de cali-

dad de las secuencias fue de 97.666% para Q20 y 93.583% para Q30. En cuanto a la cobertura con el genoma humano de referencia, el promedio fue de 99.292% con una profundidad promedio de 32.231 veces entre todos los genomas. La cobertura con el genoma de referencia fue superior a 4 veces en al menos 98 – 99%; superior a 10 veces en al menos 97 – 98%; y superior a 20 veces en al menos 84 – 96% de todas las secuencias genómicas.

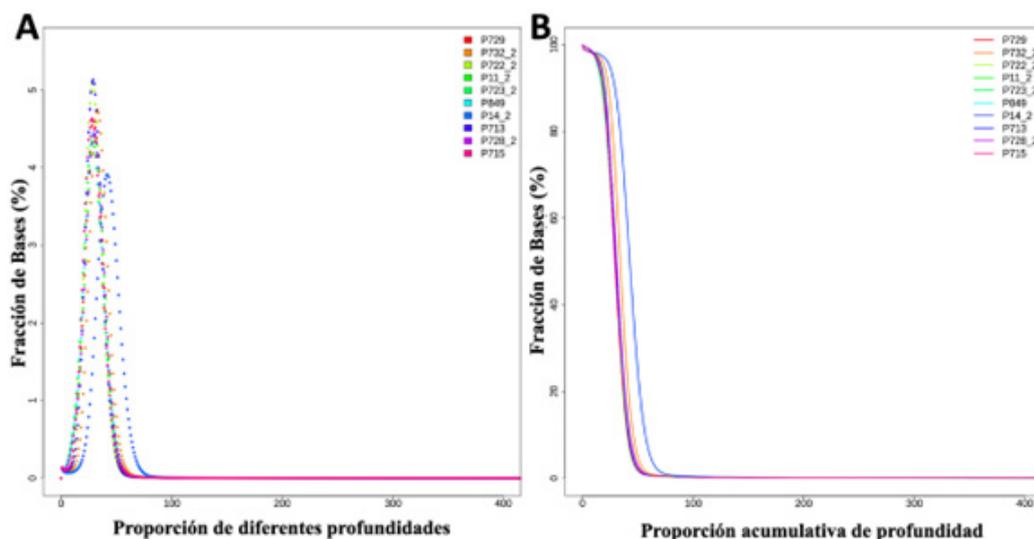


Figura 5. Proporción de la profundidad en los genomas secuenciados. A. Distribución de todas las bases secuenciadas a diferentes profundidades de cada genoma. El eje de x indica la profundidad de secuenciación y el eje de y indica las proporciones de bases secuenciadas a diferentes profundidades para cada genoma. La gráfica se distribuye alrededor del promedio de profundidad en las secuencias que fue de 32 veces en todos los genomas. B. Proporción acumulativa de profundidades de todas las bases de los genomas. El eje de x indica la profundidad de secuenciación, mientras que el eje de y indica la proporción de bases secuenciadas por encima de las diferentes profundidades. Se observa que el 100% de las bases muestran una profundidad de secuenciación mucho mayor de cero (0X).

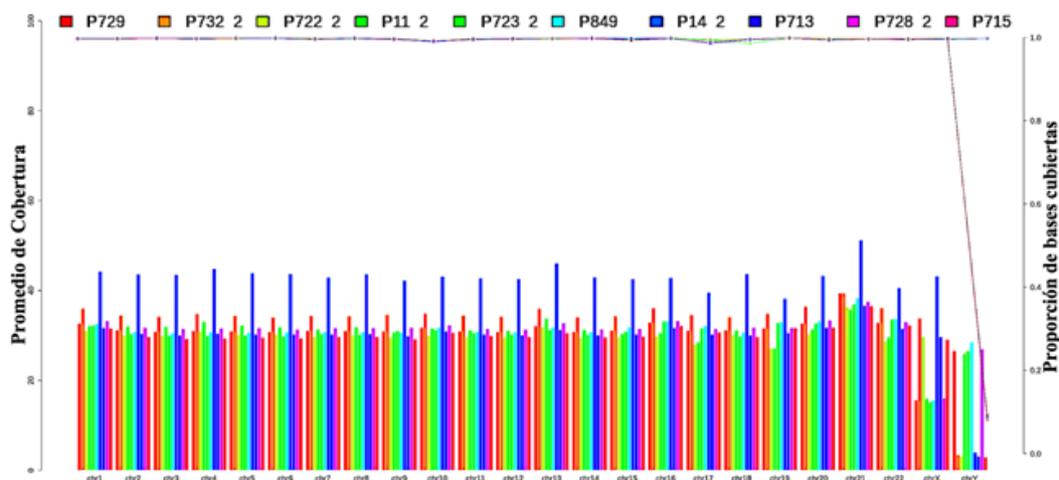
Tabla 4. Estadísticas de calidad, mapeo, cobertura y profundidad en cada genoma.

Muestra:	P729	P732/2	P722/2	P11/2	P723/2	P849	P14/2	P713	P728/2	P715
Efectividad (%)	99.82	99.83	99.85	99.82	99.83	99.81	99.85	99.82	99.81	99.83
Error (%)	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03
Q20 (%)	97.68	97.36	97.66	97.85	97.85	97.41	97.39	97.84	97.89	97.73
Q30 (%)	93.67	92.94	93.62	93.82	93.92	93.06	92.96	93.89	94.09	93.86
GC (%)	41.67	41.11	40.5	40.51	41.87	41.65	40.36	41.7	41.76	42.22
Lecturas duplicadas (%)	20.41	21.74	20.89	21.59	24.29	23.80	23.38	16.38	20.55	24.63
Lecturas mapeadas con genoma de referencia (%)	99.79	99.85	99.89	99.86	99.88	99.85	99.87	99.87	99.90	99.86
Lecturas mapeadas correctamente (%)	98.37	98.44	98.19	98.61	98.48	98.42	98.17	98.43	98.41	98.47
Promedio de Profundidad	30.75	34.68	30.03	30.88	30.16	30.71	43.3	30.51	31.25	30.04
Cobertura del genoma completo (%)	99.69	98.94	98.94	99.64	99.66	99.64	98.96	98.93	99.62	98.90
Cobertura al mayor o igual a 4X (%)	99.37	98.63	98.60	99.25	99.31	99.29	98.68	98.58	99.27	98.54
Cobertura al mayor o igual a 10X (%)	97.99	98.12	97.87	97.82	97.69	97.81	98.26	97.94	97.94	97.77
Cobertura al mayor o igual a 20X (%)	87.34	94.73	88.85	87.23	84.82	86.41	96.97	90.27	87.69	86.72

La **Figura 6** muestra las estimaciones del mapeo que incluye cobertura y profundidad de secuenciación promedio para cada cromosoma en cada genoma secuenciado. Los análisis demostraron que el promedio de cobertura en cromosomas autosómicos fue de 0.9965455 con un rango mínimo de 0.990 y un máximo de 0.999. Con respecto a los cromosomas sexuales, el cromosoma X el promedio de cobertura fue de 0.997 con un rango mínimo de 0.996 y un máximo de 0.998. Para el cromosoma Y (sólo hombres) la cobertura fue de 0.998 en todas las muestras. Por otro lado, las estadísticas de profundidad en los cromosomas autosómicos estuvieron en un rango máximo de profundidad entre 51.22 veces y un mínimo de 27.1 veces con un promedio de 32.8078182. Los cromosomas sexuales mostraron una profundidad promedio de 33.056 para el cromosoma X en mujeres con un rango mínimo de 29.02 y un máximo de 43.16. En hombres el cromosoma X mostró un promedio de profundidad de 15.618 con un rango mínimo de 15.21 y un máximo de 15.97.

El cromosoma Y (sólo hombres) mostró una cobertura promedio de 26.844 con un rango mínimo de 25.86 y un máximo de 28.41.

Figura 6. Promedio de profundidad (gráfico de barras) y cobertura (gráfico de líneas de puntos) en cada cromosoma para cada genoma secuenciado. El eje de x indica el número del cromosoma en cada genoma (ver código de colores). El eje de y de la izquierda indica la profundidad de cobertura promedio, mientras que el eje de y de la derecha indica la proporción de bases cubiertas a dicha profundidad. La profundidad de cobertura de cada cromosoma es calculada con el volumen de datos de secuenciación dividido sobre la longitud total del cromosoma en el genoma de referencia (volumen secuenciado/longitud del cromosoma de referencia). Al calcular la cobertura, la fórmula es: la longitud total cubierta dividido sobre la longitud total de cada cromosoma (longitud cubierta/la longitud total). La cobertura en todos los genomas fue superior al 98.9%.



Identificación de marcadores moleculares y polimorfismos genómicos:

Detección de marcadores y polimorfismos

En esta etapa, se abordó la principal pregunta de esta investigación: ¿cuáles son las variantes y polimorfismos genéticos posiblemente asociadas a ENT en el panameño?, específicamente en los genomas de la población amerindia Ngöbe, los ancestros de la población panameña y que contribuyen en un 51% al genoma del panameño actual. Los análisis preliminares resultaron en la identificación y anotación de enormes listas en MS Ex-

cel con miles de variantes/polimorfismos posiblemente asociados a ENT. Para cada marcador molecular (SV, CNV, InDel y SNP), se identificaron preliminarmente miles de variantes/polimorfismos genéticos potencialmente patogénicos que resultaron del análisis simultáneo con miles de datos clínicos, donde se han reportado estas variantes asociadas a ENT en diferentes poblaciones del mundo. Inicialmente no se sabía cuál de los cuatro marcadores moleculares resultaría en la identificación de variantes/polimorfismos patogénicos asociados a ENT, por lo que decidimos estudiarlos todos basándonos en varios criterios de patogenicidad (explicados en los métodos). Estos análisis nos permitieron identificar unos pocos marcadores candidatos y enfocarnos en algunos con mayor probabilidad de ser patogénicos y asociados a ENT.

Identificación preliminar de genes y polimorfismos

Los análisis preliminares de marcadores moleculares y anotación de polimorfismos fueron realizados por Novogene. Aunque secuenciamos el genoma completo de los diez individuos, los análisis bioinformáticos para identificar genes y polimorfismos patogénicos los enfocamos solamente en el exoma, ya que se estima que aproximadamente el 80-90% de las enfermedades conocidas son causadas a nivel de exones (regiones codificantes de RNA y/o proteínas). El genoma humano comprende alrededor de 3.2 billones de bases, de las cuales casi el 1.2% corresponde al exoma y representa aproximadamente 180,000 exones y unas 30 millones de bases (Maróti et al., 2018). Los análisis bioinformáticos preliminares incluyeron control de calidad de las secuencias, alineamiento con el genoma humano de referencia, estadísticos de profundidad y de cobertura de secuenciación explicados arriba (Rubio et al., 2020). Además, incluyó la identificación preliminar, anotación y estadísticas de miles de marcadores moleculares de tipo SV, CNV, InDel y SNPs, para cada genoma secuenciado.

Identificación de Variaciones Estructurales (SV)

Las Figuras 7 y 8 muestran, respectivamente, las proporciones de SV y tipos de cada SV (ganancias y pérdidas) identificados en cada genoma. Los análisis demostraron que la mayor parte de las variantes identificadas en los SV correspondían en deleciones (3,656 en promedio), seguido de duplicaciones (1,161.9 en promedio), translocaciones (1,059 en promedio), inversiones (627.4 en promedio) e inserciones (74.2 en promedio) (Figura 7). Adicionalmente, los datos demostraron que la mayor parte de las SV estaban relacionadas con ganancias (Figura 8). Los análisis y búsquedas en bases de datos clínicas demostraron que los SV identificados no eran significativos clínicamente como marcadores polimórficos para ENT, por lo que no se continuó su estudio.

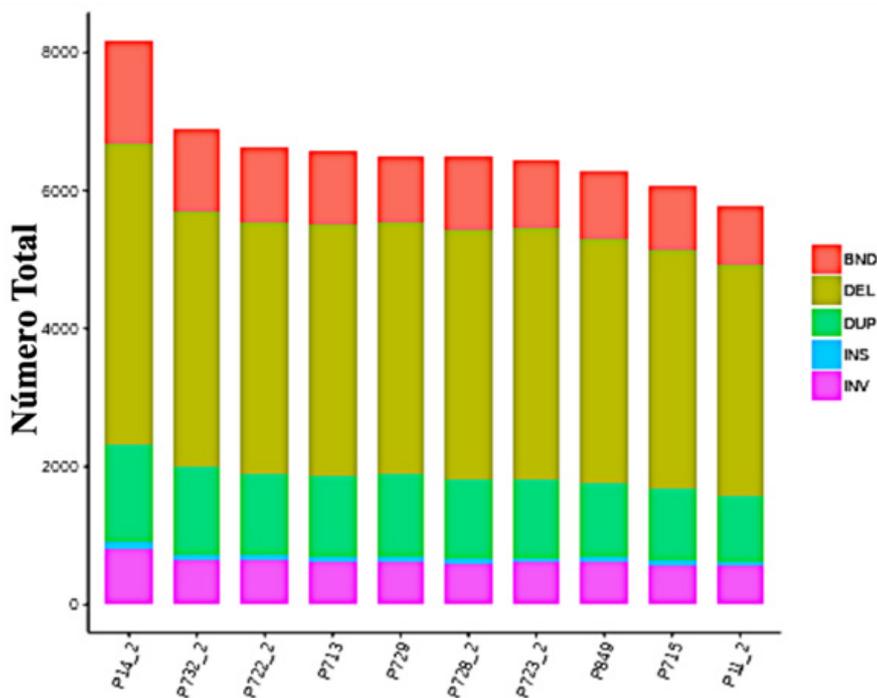


Figura 7. Número de diferentes tipos de Variaciones Estructurales (SV) en cada genoma. El eje x representa los diferentes genomas y el eje y representa la proporción de cada tipo de SV, según el código de colores mostrado. BND = translocaciones, DEL = deleciones, DUP = duplicaciones, INS = inserciones, INV = inversiones.

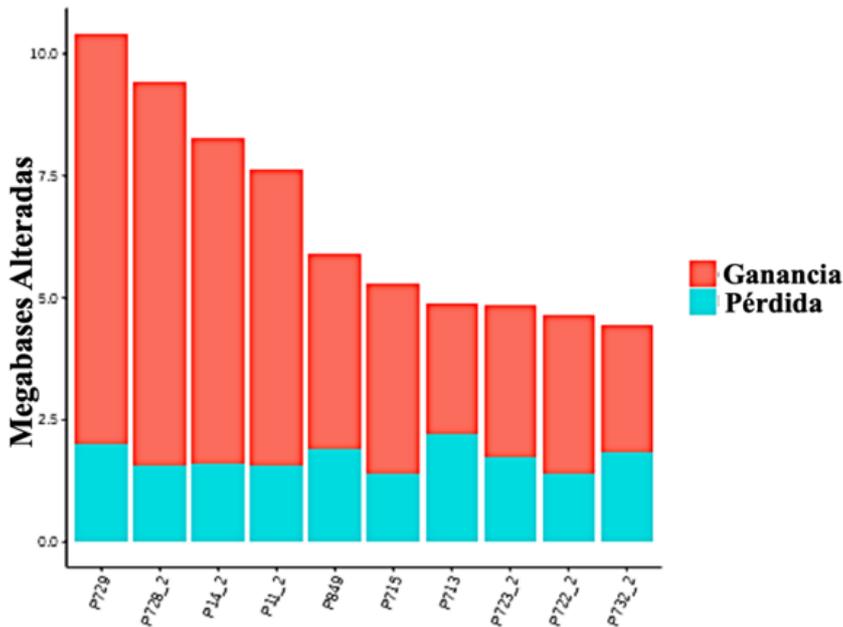


Figura 8. Variaciones estructurales en número de copias de acuerdo a ganancia o pérdida de tamaño en cada genoma. El eje x representa cada genoma y el eje y representa el tamaño total de las regiones genómicas afectadas por ganancias y pérdidas (Mb).

Identificación de Inserciones/ Deleciones (InDels)

Las **Figura 9** muestra las gráficas de pastel donde se clasifica la distribución de InDels en las regiones genómicas codificantes (CDS) y no codificantes (gráficas de la izquierda en cada muestra), y si los cambios representan cambios en aminoácidos o cambio en marco de lectura (gráficas de la derecha de cada muestra). En total, se identificaron 771,665.6 InDels en promedio en todos los genomas, de los cuales, 92,392 en promedio eran novedosos (reportados por primera vez) en todos los genomas. Además, se demostró que para las regiones codificantes (CDS) las deleciones con cambio en marco de lectura representan unas 142 y las inserciones con cambio en marco de lectura representan unas 108, respectivamente, en promedio para todos los genomas. Adicionalmente, se determinó que las ganancias en codones de terminación (stop gain) representan 10.5, mien-

tras que las pérdidas de codones de terminación (stop loss) representan 1.6 en promedio, respectivamente, para todos los genomas. Los análisis y búsquedas en bases de datos clínicos demostraron que los InDels eran muy poco informativos clínicamente como marcadores polimórficos para ENT, por lo que se discontinuó el estudio de los InDels.

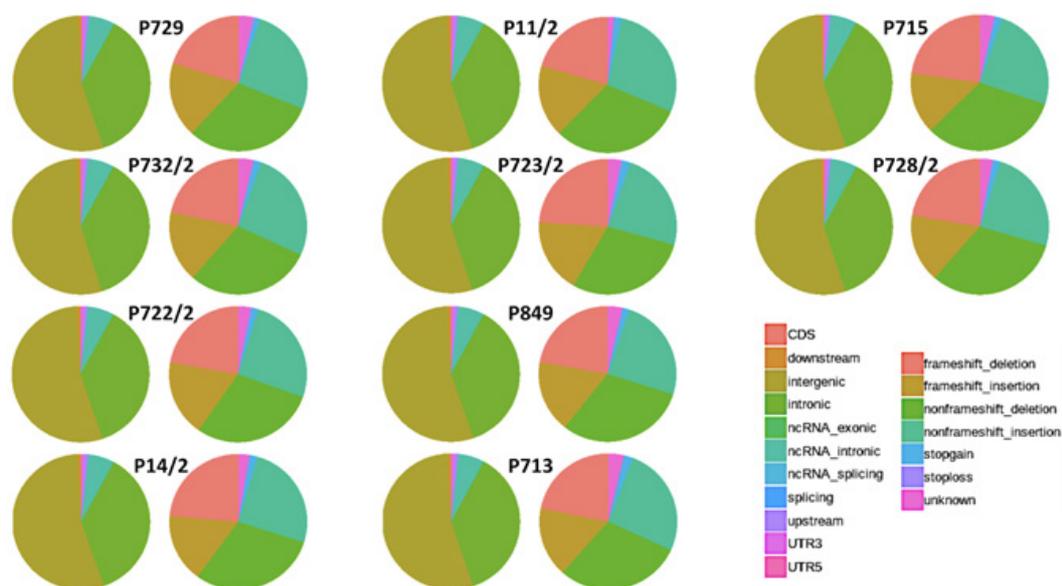


Figura 9. Gráficas de pastel demostrando la clasificación de InDels identificados en las regiones genómicas (izquierda) y número de diferentes tipos de InDels en las regiones codificantes (derecha).

Identificación de Variaciones en Número de Copias (CNV)

La **Figura 8** muestra las variaciones estructurales en cuanto a variaciones en número de copias (CNV). Los datos demostraron un promedio de 347 CNVs en todos los genomas, de los cuales la mayor parte correspondían a ganancias (209 en promedio), seguidas de 138 pérdidas en promedio, respectivamente para todos los genomas. Aunque muy interesantes, los CNVs demostraron ser muy poco informativos clínicamente para riesgo genético a ENT.

Identificación de Polimorfismos de Nucleótidos Simples (SNPs)

La **Figura 10** muestra las gráficas de pastel con los conteos totales de SNPs (gráficos de la izquierda) y sus variantes en regiones codificantes (CDS) (gráficos de la derecha) para cada genoma. En promedio se identificaron un total de 3,537,845.3 SNPs entre los genomas, de los cuales en promedio 98,247 eran novedosos (no reportados previamente). En las regiones codificantes los SNPs sinónimos (no generan cambios en aminoácidos) corresponden en promedio a 11,397.7; y los SNPs sin sentido corresponden a 10,796.5 (generan cambios en aminoácidos), en promedio, respectivamente para todos los genomas. Los codones de parada ganados (stop gain) y perdidos (stop loss) corresponden en promedio 97.6 y 11.2, respectivamente para todos los genomas. La relación entre transiciones/transversiones (ts/tv) fue de 2.01, mientras que la relación ts/tv entre los SNPs novedosos fue de 1.171, respectivamente en todos los genomas.

Se realizaron varios análisis para la búsqueda e identificación de SNPs en múltiples bases de datos clínicas (ver métodos) y se concluyó que los SNPs fueron mucho más informativos clínicamente para identificar asociaciones con ENT. Por lo tanto, en las siguientes investigaciones nos enfocamos en algunos SNPs para determinar su frecuencia en la población Ngöbe (ver capítulo siguiente). Los 16 genes con SNPs de mayor significancia clínica identificados se muestran en la **Tabla 5**. Entre estos SNPs nos enfocamos en los polimorfismos rs1801133 (gen MTHFR, asociado a cánceres, enfermedades vasculares y otras), rs1042522 (gen TP53, asociado a diferentes tipos de cánceres y cánceres hereditarios), rs13266634 (gen SLC30A8, asociado a diabetes tipo 2) y rs4684677 (gen GHRL, asociado a obesidad, diabetes y síndrome metabólico) (**Tabla 5**). Estos genes con SNPs patogénicos se identificaron en al menos 5 de los genomas secuenciados, por lo que fueron seleccionados para genotipar sus frecuencias en una muestra de 60 individuos de la población Ngöbe (ver capítulo siguiente).

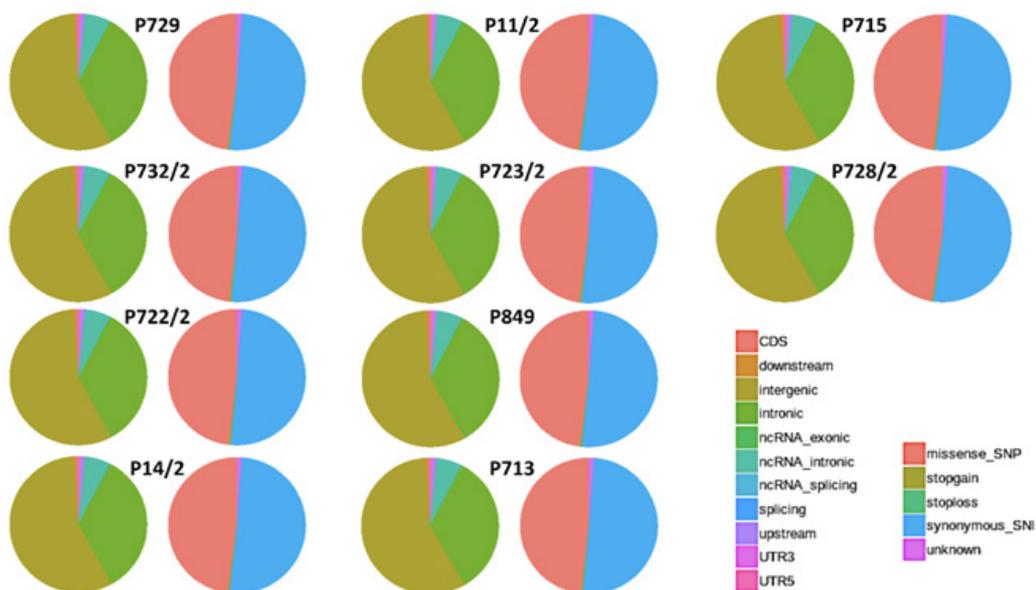


Figura 10. Gráficas de pastel mostrando los conteos en regiones genómicas (izquierda) y clasificación en regiones codificantes (derecha) de los SNPs encontrados en cada genoma.

Tabla 5. Polimorfismos SNPs de mayor significancia clínica identificados.

Cromosoma	Referencia	Mutado	Gen	Significancia Clínica
chr1	C	T	MTHFR*	Tumores malignos, deficiencia de MTHFR (homocistinuria), tromboembolismo, enfermedad cardíaca
chr1	G	A	SDC3	Obesidad
chr1	T	C	EPHX1	Deficiencia de epóxido hidrolasa soluble, problemas del metabolismo en absorción de lípidos (hipercolesterolemia), hipertensión, preeclampsia/eclampsia
chr3	G	T		Síndrome metabólico, obesidad, diabetes
	T	A		
chr5	G	A	FGFR4	Progresión y metástasis del cáncer
chr6	A	G	SOD2	Complicaciones vasculares en diabetes
<i>Continúa en la próxima pag.</i>				
chr7	C	A	PPP1R3A	Resistencia a la insulina (diabetes tipo 2)
chr8	C	T	SLC30A8	Diabetes tipo 2
chr12	G	A	OAS1	Diabetes tipo 1
chr13	C	T	IRS2	Diabetes tipo 2, obesidad
chr16	A	G	IL4R	Alergias (atopia), asma, progresión lenta del síndrome de inmunodeficiencia adquirida
chr16	G	A	NQO1	Cánceres de mama y pulmón, leucemias, toxicidad del benceno.
chr17	C	G	TP53*	Síndrome Li-Fraumeni, cánceres hereditarios
Chr10	T	C	STOX1	Preeclampsia/eclampsia, hipertensión
Chr17	T	C	AKAP10	Enfermedad cardíaca
Chr2	C	A		
	A	G		
	T	C		

*Polimorfismos patogénicos seleccionados para genotipado en 60 individuos Ngöbe (ver capítulo siguiente).

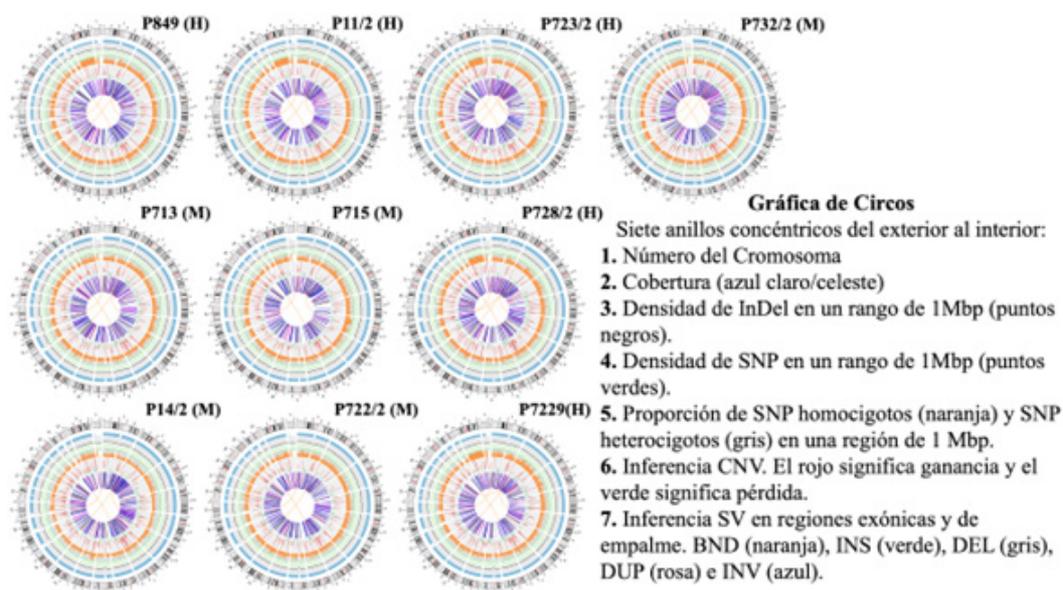


Figura 11. Gráfica de “Circos” representando la cobertura por cromosoma, InDels, SNPs, CNVs y SVs en cada genoma. La figura consta de siete círculos concéntricos (anillos) contados desde el exterior hacia el interior explicados en la figura. BND = translocaciones, INS= inserciones, DEL = deleciones, DUP = duplicaciones e INV = inversiones.

DISCUSIÓN

Los análisis de genomas Ngöbe condujeron a la identificación de polimorfismos de relevancia clínica en esta población

Esta es la primera vez que se secuencian genomas humanos y que se reportan los genes específicos y los polimorfismos de posible importancia clínica asociados a algunas patologías en la población panameña. Este tipo de estudios para identificar genes asociados a enfermedades ha sido reportado previamente en otros países donde también utilizan la ascendencia genética como un marcador estratégico para hacer inferencias sobre el papel potencial de la genética ancestral en las variantes genéticas asociadas a enfermedades (Wassel et al., 2009; Reiner et al 2007; Chande et al., 2021). Esto es importante, ya que estudiar la genética y genómica de enfermedades complejas y multifactoriales en poblaciones muy hete-

rogéneas como la panameña por el intenso mestizaje, representa un gran reto de investigación.

En Panamá, la población tiene tres orígenes genéticos ancestrales principales: indígenas (Ngöbe) con un 51%, seguido de europeos (España) con un 25% y africanos (del Oeste de África, lo cual incluye afro-coloniales y afro-caribeños) con un 24%, proporciones determinadas mediante marcadores moleculares generales como microsatélites (STRs) e InDels. La distribución de estos genes no es, sin embargo, homogénea, mostrando un mestizaje y polimorfismo diferencial entre las provincias del país (Castro Pérez et al., 2016; Ramos et al., 2018). En estos mismos estudios se encontró una asociación entre la ascendencia ancestral y la prevalencia de estas enfermedades. Es decir, dependiendo del origen ancestral que predomina en cada provincia, este predice (al menos en parte) la prevalencia de la enfermedad basado en las disparidades de dicha enfermedad en la población ancestral dominante (Castro-Pérez et al., 2016; Bock, et al., 2009; Zeigler-Johnson et al., 2008; Robbins et al., 2007; Cheng et al., 2010; Hankey, 1999; Donnan et al., 2008). Realizar estudios en poblaciones ancestrales homogéneas y de mayor predominio, como los Ngöbe, ofrece la ventaja de facilitar el análisis e interpretación de los datos (Castro et al., 2007). Adicionalmente, los datos reportados aquí permiten identificar el posible origen genético ancestral de mutaciones y polimorfismos asociadas a enfermedades en la población mestiza moderna que representa la mayor parte de la población del país con alrededor del 70% y poseen, además, alta contribución mayoritariamente de genes Ngöbe con un 51%. Esto sugiere que el componente ancestral amerindio/Ngöbe del acervo genético del país contribuiría en gran medida al riesgo a padecer ENT, como fue sugerido previamente en Panamá y otros países (Castro-Pérez et al., 2016; Ramos et al., 2018; Chande et al., 2021).

Los polimorfismos patogénicos identificados guardan relación con las principales enfermedades que afectan a la población panameña

Datos elaborados por la Contraloría General de la República, mediante el Instituto Nacional de Estadística y Censo (INEC) y del MINSA sobre las defunciones ocurridas en los últimos 20-30 años, indican como principales causas de muerte las enfermedades del sistema cardiovascular y cerebro-

vascular, y los tumores malignos (cáncer). Los cánceres han dominado en los últimos años, siendo los principales el de próstata, mama, colo-rectal, cuello uterino, estómago, así como de cabeza/cuello. Le siguen las enfermedades endocrinas y metabólicas, de las cuales la principal es la diabetes mellitus y aquellas asociadas a la obesidad.

Los análisis de datos de secuenciación genómica han demostrado que las variantes polimórficas de mayor significancia clínica identificadas en el exoma corresponden a SNPs, en particular en los que se sustituye un aminoácido por otro. Los análisis demostraron que en promedio 98,247 de los SNPs identificados eran novedosos (no reportados previamente), por lo que se requerirán análisis adicionales y para determinar la significancia de los mismos que serán publicados en futuros reportes. Los SNPs identificados y asociados a ENT más numerosos tienden a estar relacionados con el cáncer, enfermedades vasculares, diabetes, obesidad y síndrome metabólico, lo cual es consistente con los patrones de enfermedades y muertes en la población.

La lista de genes y sus polimorfismos patogénicos en la Tabla 5 corresponden a aquellos identificados en al menos cinco de los diez genomas secuenciados. Entre los polimorfismos asociados a enfermedades vasculares y cardíacas se identificaron en los genes MTHFR, EPHX1, TTN (3 polimorfismos), STOX1, AKAP10 y SOD2 (dbSNP, 2023; SNPedia, 2023). Entre polimorfismos asociadas a tumores malignos se identificaron en los genes NQO1, FGFR4, MTHFR, TP53 y EPHX1 (dbSNP, 2023; SNPedia, 2023). Para las variantes genéticas asociadas a diabetes, enfermedades nutricionales, endocrinas, obesidad y metabolismo se identificaron polimorfismos en los genes SDC3, EPHX1, SLC30A8, GHRL (2 polimorfismos), IRS2, OAS1, SOD2 y PPP1R3A (dbSNP, 2023; SNPedia, 2023). Es importante mencionar que algunos polimorfismos pueden interactuar entre sí y estar asociados con varias enfermedades, o actuar como agravantes de condiciones pre-existentes. Por ejemplo, MTHFR está asociado tanto con cáncer como con enfermedades cardiovasculares, EPHX1 y STOX1 se encuentran asociados con hipertensión arterial, la cual es un agravante de eventos cerebro y cardiovasculares. Otros polimorfismos contribuirían en mayor riesgo tanto en obesidad como en diabetes, como en los genes IRS2 y GHRL, mientras que SOD2 contribuiría a problemas vasculares y a severidad en diabetes. Por otro lado, entre los genes identificados, el ILR4 es el único cuya variante no está asociada al tipo de enfermedades que causan más muertes en el país, pues está asociada a alergias, asma y progresión lenta de la adquisición del Síndrome de Inmunodeficiencia Adquirida. Este polimorfismo podría

ser (teóricamente) un SNP que potencialmente ofrecería cierta ventaja en la severidad o rapidez de avance de esta enfermedad infecciosa, pero, a la vez, podría causar altos niveles de inflamación que complicarían otras condiciones asociadas a ENT (dbSNP, 2023; SNPedia, 2023).

La mayoría de las variantes identificadas en los genomas Ngöbe están asociadas a ENT, que son consistentes con las principales causas de muertes en las últimas décadas en el país, por lo que es posible que estos polimorfismos contribuyan a la aparición y/o severidad de las mismas. Sin embargo, más estudios genéticos y genómicos serán necesarios para confirmar estas asociaciones. Por ejemplo, estudios en curso analizan los patrones de estos SNPs en el ADN de pacientes diagnosticados con diferentes enfermedades y se comparan los genotipos de los SNPs con grupos control que no padezcan dichas enfermedades. Estas investigaciones tienen como objetivo confirmar o no estas posibles asociaciones genéticas y genómicas.

Interacciones de factores ambientales y estilos de vida con genes y ENT

Los análisis genómicos reportados aquí proporcionan las primeras evidencias de los genes y polimorfismos posiblemente asociados clínicamente a patologías en la población del país. Sin embargo, también es muy importante comprender que “asociación” no necesariamente significa causalidad, por lo que es imprescindible destacar que la forma en cómo estos genes interactúan con otros genes/polimorfismos, así como con factores ambientales, sociales, contaminación, estilos de vida, entre otros, tiene un impacto significativo en la etiología de las ENT. Por ejemplo, ha sido reportado que un solo polimorfismo o mutación no necesariamente es suficiente para causar la formación de cáncer, ya que los reportes indican que al menos dos mutaciones son requeridas para inducir tumores malignos (Knudson 2001; Hahn et al., 2003; Dankort et al., 2009). Por lo tanto, cabe destacar que la patogenicidad de algunos polimorfismos necesita ser interpretada en el contexto de su interacción con otros genes/polimorfismos o mutaciones, lo cual requiere análisis poblacionales mucho más exhaustivos llamados “estudios asociativos de todo el genoma” o en inglés *Genome Wide Association Studies* (GWAS). En los GWAS se analizan los genotipos de cientos o miles de SNPs en genomas de cientos o miles de pacientes enfermos y se comparan con genomas de grupos de pacientes control que no padezcan la enfermedad. Estos estudios son especialmen-

te complejos porque las enfermedades no transmisibles se consideran de origen multifactorial (genes, ambiente, estilos de vida, etc.) y de carácter poligénico (muchos genes involucrados). Además, la población panameña muestra muy altos niveles de mestizaje de tres poblaciones ancestrales (indígena, europea y africana), lo cual podría dificultar aún más la interpretación y análisis de datos con combinaciones heteroalélicas de polimorfismos patogénicos que podrían ser muy raros o inexistentes en las poblaciones ancestrales. Por todo lo anterior, el estudio piloto presentado aquí es una investigación pionera que permitiría un punto de partida para expandir al resto de la población e ir generando algunas luces preliminares que sirvan de referencia a los problemas de epidemiología genética de estas enfermedades en Panamá.

Por otro lado, se sabe que sólo un 5-10% de los tumores malignos son hereditarios o de tendencia familiar, mientras que la mayoría son esporádicos o adquiridos por mutaciones somáticas no hereditarias (NIH-National Cancer Institute, 2023). En este sentido, el cáncer es considerado una enfermedad genética molecular que afecta genes causantes de cáncer, principalmente oncogenes, los cuales sufren mutaciones activantes en genes de proliferación celular y mutaciones inactivantes en genes supresores de tumores, genes de reparación del ADN o del control del ciclo celular, genes de apoptosis, entre otros. Las afecciones de estos genes pueden ser heredadas o adquiridas en células somáticas tal como se describió anteriormente. Las principales mutaciones al genoma son introducidas a este por hábitos de los seres humanos o por acciones ambientales y no a nivel hereditario. Entre estos hábitos se han asociado fuertemente, por ejemplo, el tabaco, bebidas alcohólicas, la dieta con alimentos ultraprocesados y bajas en frutas y vegetales, obesidad y falta de actividad física, contaminación ambiental, la exposición a la luz solar ultravioleta y a algunos medicamentos. Además, infecciones virales como el VIH, virus del papiloma, virus de la hepatitis; e infecciones bacterianas como *H. Pylori* entre otras (NIH-National Cancer Institute, 2023). La exposición a estos factores ambientales causaría un efecto acumulativo de mutaciones a lo largo de los años en la vida de las personas, lo cual se refleja con la edad de diagnóstico de la mayoría de los cánceres que es en promedio superior a los 60 años. Por lo tanto, minimizar la exposición a estos factores de riesgo a lo largo de los años contribuiría a disminuir significativamente el riesgo de cáncer, así como retrasar procesos de envejecimiento y otras enfermedades. Por lo tanto, promover estilos de vida saludables como controlar el peso y tener dietas ricas en frutas y vegetales, las cuales proveen antioxidantes y previenen deficiencias vitamínicas, fortaleciendo así

el sistema inmune para combatir células cancerosas y los radicales libres causantes de mutaciones. Las estrategias de prevención incluirían, además, vacunarse contra infecciones virales asociadas a cánceres, mantener actividad física y chequeos médicos regulares, ya que, de ser detectado algún cáncer u otra condición de salud, el diagnóstico temprano aumenta las probabilidades de supervivencia (CDC, 2023; NIH-NCI, 2023).

Otro mecanismo molecular asociado con el desarrollo de cáncer (y con otras enfermedades) son los cambios epigenéticos y epigenómicos (epi = superficie; genética/genómica = herencia), los cuales hacen referencia a los cambios químicos introducidos en la superficie del ADN posterior a su replicación y que no afectan su secuencia de bases (no involucra mutaciones). Por ejemplo, agregarle grupos metilo ($-CH_3$) directamente al ADN o por medio de modificaciones (acetilaciones, metilaciones, entre otras) a proteínas que se unen al ADN llamadas histonas. Estos cambios epigenéticos controlan la actividad de los genes (expresión genética), es decir, regulan cuando un gen “se prende” (se expresa) produciendo RNA y proteínas que llevan a cabo las funciones celulares; o cuando un gen “se apaga” (se silencia, no produce RNA ni proteínas). Estos cambios epigenéticos tienen un efecto muy similar a las mutaciones, por ejemplo, el silenciamiento de genes supresores tumorales (que normalmente debería estar encendido) como TP53 por medio de la metilación del ADN, podría ser devastador para una célula y desembocar en un tumor maligno. Es decir, el gen TP53 frecuentemente encontrado en tumores malignos con mutaciones inactivantes, es inactivado (silenciado) por metilación del ADN, conduciría a la formación de un tumor maligno, similar a una mutación inactivante, a pesar de que su secuencia no haya tenido cambios (mutaciones). Múltiples evidencias reportadas indican que estos patrones y cambios epigenéticos y epigenómicos son dinámicos en las células, y que los mismos son modificables por factores ambientales como la dieta, medicamentos y estilos de vida. Es decir, algunos estilos de vida y ambientes no saludables podrían desregular estos patrones epigenéticos, llevando a la adquisición de cambios epigenéticos de metilación patogénicos que conduzcan a cáncer u otras enfermedades; mientras que estilos de vida saludables promoverían patrones epigenéticos protectores (Herceg, 2007; Sapienza e Issa, 2016; Baylin y Jones, 2016; Takeshima y Ushijima, 2019; Cheng et al., 2019; Lu et al., 2020; Desaulniers et al., 2021; Mancarella y Plass, 2021; Usui et al., 2023; Maleknia et al., 2023). Los procesos epigenéticos y mutacionales no son mutuamente excluyentes, ya que ambos pueden interactuar contribuyendo simultáneamente en procesos de carcinogénesis, por lo que en ambos casos es importante promover estilos

de vida saludables como los mencionados arriba. Aún no se han realizado estudios epigenéticos ni epigenómicos en la población panameña y su relación con el ambiente y estilos de vida, por lo que esta sería otra área de investigación que complementaría los estudios genéticos y genómicos como los presentados en este libro.

En cuanto a enfermedades cardio y cerebrovasculares, los reportes indican que los principales factores de riesgo incluyen presión arterial alta, altos niveles de colesterol y triglicéridos (dietas altas en grasas saturadas como carnes rojas y ultra procesadas), diabetes, fumar, obesidad o sobrepeso, falta de actividad física, así como dietas altas en sodio que incrementan la presión arterial (CDC, 2023; NIH-National Heart, Lung and Blood Institute, [NIH-NHLBI], 2023; American Heart Association). Por otro lado, los factores de riesgo asociados al desarrollo de diabetes incluyen tener prediabetes, haber tenido diabetes gestacional, obesidad o sobrepeso (asociado a dietas altas en grasas saturadas y altas en carbohidratos/azúcares), historia familiar, poca actividad física, y ascendencia ancestral africana, hispánica/latina, indígena y asiática (CDC, 2023; NIH-National Institute of Diabetes and Digestive and Kidney Diseases [NIH-NIDDK]).

Es conocido que algunos de estos estilos de vida que representan un factor de riesgo para ENT los aprendemos cultural y familiarmente, por lo que muchas veces, además de heredar los genes, también heredamos estas tradiciones y las continuamos por costumbre o simplemente por desconocimiento. Por lo tanto, es imprescindible que las Instituciones competentes del Estado, escuelas y las familias fomenten la educación en salud preventiva con buenos hábitos y estilos de vida, ya que independientemente de la edad y estado de salud, introducir estos cambios genera un efecto positivo. Además, es imprescindible que el Estado invierta en mejorar el acceso y la calidad de los servicios de salud, y que las entidades competentes del Estado sean muy celosas de que se cumplan con los niveles máximos permitidos de contaminación y toxicidad del ambiente tanto en el aire, suelo y las aguas, así como de los alimentos y medicamentos que consumimos. Esto es muy importante, ya que algunos contaminantes pueden ser tóxicos a largo plazo a concentraciones tan bajas como en partes por millón. Por ejemplo, algunos países de la región han implementado que en las etiquetas de los alimentos se mencione en letras grandes de manera explícita cuando algún alimento representa un riesgo para la salud por ENT. Un factor de riesgo ambiental crítico en los últimos años ha sido la contaminación de prácticamente todos los ambientes debido a los micro y nano plásticos, así como los llamados químicos eternos

(*Forever Chemicals*), los cuales son un tipo de plásticos. Además, se deben mantener la vigilancia de los niveles de metales pesados en el ambiente y productos de consumo, de los cuales sus efectos aún se están investigando, pero definitivamente no representan buenas noticias para la salud humana y para otras especies.

Lamentablemente, otro factor de riesgo reportado en estas y otras enfermedades incluye los bajos niveles socioeconómicos tales como la pobreza, el nivel de educación y la falta de acceso a servicios de salud eficientes. Esto es una tendencia en la mayoría de los países y también en Panamá (Quiel et al., 2016; CDC, 2023; NIH, 2023). En este sentido, no es un secreto que en Panamá los altos niveles de pobreza también están bastante asociados con la ascendencia africana e indígena. Por lo tanto, en algunos casos podría ser difícil separar si la causa de algunas enfermedades y/o muertes fue por una combinación de factores sociales como la pobreza asociada a poca educación en prevención de salud, poco acceso a una buena nutrición o falta de acceso a buenos servicios de salud, o los genes, el ambiente, etc. Para combatir esto, es responsabilidad del estado hacer las inversiones necesarias en salud pública para mejorar el acceso y la calidad de los servicios de salud en la población. Adicionalmente, es muy importante que cada individuo y familia se informe y promueva estilos de vida saludables y que el Estado fomente e invierta en educación y prevención de estas enfermedades, y desarrolle políticas para reducir la pobreza. También es fundamental que el Estado promueva la investigación celular y molecular de estas enfermedades asignando más fondos y reduciendo la burocracia para la ejecución de la investigación. Por ejemplo, el apoyo a más investigación básica sobre la identificación de los genes y polimorfismos, como los reportados aquí, podría contribuir al desarrollo de medicina molecular de precisión y mejores estrategias de diagnóstico y prevención en grupos de la población que posean estas variantes. Un programa modelo para este tipo de investigaciones es el Programa PARI (*Panama Aging Research Initiative*) que lleva a cabo INDICASAT, donde se investigan las disparidades en pacientes de Alzheimer en el país. En el Programa PARI se sigue longitudinalmente una cohorte de pacientes a los que se miden de manera integral múltiples parámetros de riesgo como biomarcadores en sangre, biomarcadores en líquido cefalorraquídeo, estudios genéticos y genómicos, estilos de vida, entre otros (Villarreal et al., 2015; Britton et al., 2019; Villarreal et al., 2019). Es muy importante que más modelos similares al Programa PARI se implementen en la población utilizando cohortes de pacientes longitudinales de las principales enfermedades que afectan al país. Estos estudios conducirían a una mejor comprensión de estas en-

fermedades, sus disparidades, factores de riesgo, así como sus factores protectores y preventivos a nivel genético, molecular, social y su relación con el ambiente y estilos de vida. Por otro lado, es importante que el Estado también promueva el desarrollo de modelos moleculares y celulares con aplicaciones biotecnológicas que conduzcan al posible desarrollo de patentes y nuevos tratamientos de estas enfermedades. Para lograr esto, no sólo es necesario asignar más fondos para la investigación y el desarrollo (I+D) que actualmente es apenas del 0.08% del PIB (esto sin contar al Instituto Smithsonian, con el cual llega tan solo al 0.13%), uno de los porcentajes más bajos en los países de la región y de economía similar en el mundo (p.ej. Algunas cifras aproximadas de los últimos años: Colombia 0.25%, Bolivia y Perú 0.3%, Costa Rica 0.34%, Chile 0.38%, México 0.5%, Uruguay 0.96, Brasil 1.15%, Singapur 2.16%, EE.UU. 3.4%, Taiwan 3.0%, Israel 4.1%). Para transformar esto, el Estado ha propuesto desde hace años la idea de llegar al 1% del PIB, lo cual no se ha cumplido. Además de asignar más fondos, se necesitan otras medidas que involucren la reducción de los procesos burocráticos para las convocatorias por fondos y la administración de los proyectos de investigación. Hoy en día, estos procesos de convocatorias y administración de fondos son abrumadoramente complejos y muy ineficientes. Por ejemplo, comprar un reactivo común puede tardar entre 6 – 12 meses en llegar, mientras que comprar un equipo de laboratorio puede tardar como mínimo unos dos años. Los procesos administrativos podrían simplificarse tanto en las convocatorias para fondos, como en los informes de las investigaciones, e igualmente hacer menos rígidos los procesos de compras. Hacer investigación científica es un proceso muy complejo, desafiante y competitivo, aún bajo las mejores condiciones como en Estados Unidos. Cuando a esto le sumamos estas limitantes en cantidad de fondos y procesos burocráticos, se obstaculiza mucho más la competitividad de las investigaciones del país.

CONCLUSIONES

Los análisis genómicos nos condujeron a la identificación por primera vez de variantes genéticas de importancia biomédica que podrían afectar a la población Ngöbe y, presumiblemente, a la población mestiza general del país derivada de esta población indígena ancestral. Presumimos que los alelos patogénicos en la población ancestral Ngöbe, podría haberse transferido a la población mestiza actual y posiblemente contribuir a la alta prevalencia y riesgo de padecer algunas ENT en la población general del

país. Sin embargo, dada la complejidad genética y multifactorial de estas enfermedades, la interpretación de los resultados presentados aquí aún está en desarrollo por tratarse de un estudio preliminar en el país. Investigaciones en curso confirmarían o no estas posibles asociaciones entre los polimorfismos identificados y el riesgo de padecer las respectivas enfermedades. Estos resultados representan los primeros genomas humanos secuenciados de la población panameña y uno de los primeros en la región, por lo que es un estudio piloto de referencia para la población.

Aunque algunos factores de riesgo genéticos como historia familiar y ancestralidad no son modificables, sin embargo, la mayoría de los factores de riesgo asociados ENT relacionados con el ambiente y estilos de vida son modificables o al menos podemos minimizar la exposición de los mismos. Adicionalmente, aunque los factores genéticos desempeñen un papel en la etiología de estas enfermedades, también es muy probable que las personas con antecedentes familiares de estas enfermedades compartan entornos comunes, hábitos y estilos de vida similares que pueden aumentar su riesgo. Las condiciones socioeconómicas, así como la contaminación y toxicidad ambiental, también son factores de riesgo modificable a nivel familiar y a nivel del Estado. Tanto los individuos, como las familias y el estado comparten niveles de responsabilidad en promover cambios en educación en estilos de vida saludables y hacer las inversiones necesarias en salud pública para mejorar el acceso y la calidad de los servicios de salud en la población y en la calidad del ambiente.

Literatura citada

1. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet*. 2013 Jan; Chapter 7:Unit7.20.
2. Arias, T., (1991). Aproximación antropológica a algunos problemas biomédicos de las poblaciones amerindias. *Hombre y cultura*. Universidad de Panamá. Centro de Investigaciones Antropológicas, volumen 1, número 1, páginas 99–114.
3. Arias, T., Castro, E., Ruiz, E., Barrantes, R., & Jorge-Nerbert, L. (2002). La mezcla racial de la población panameña. *Revista Médica de Panamá - ISSN 2412-642X*, 27, 5–17.
4. Arias, T. D., Jorge, L. F., Griese, E.-U., Inaba, T., & Eichelbaum, M. (1993). Polymorphic N-acetyltransferase (NAT2) in Amerindian populations of Panama and Colombia: high frequencies of point mutation 857A, as found in allele S3/M3. *Pharmacogenetics*, 2(6), 328–331.
5. Baylin SB, Jones PA. Epigenetic Determinants of Cancer. *Cold Spring Harb Perspect Biol*. 2016 Sep 1;8(9):a019505.
6. Bock, C. H., Schwartz, A. G., Ruterbusch, J. J., Levin, A. M., Neslund-Dudas, C. et al. (2009). Results from a prostate cancer admixture mapping study in African-American men. *Human Genetics*, 126(5), 637.
7. Boeva V, Popova T, Bleakley K, Chiche P, Cappo J, et al. Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. *Bioinformatics*. 2012 Feb 1;28(3):423–5.
8. Boeva V, Zinovyev A, Bleakley K, Vert JP, Janoueix-Lerosey I, et al. Control-free calling of copy number alterations in deep-sequencing data using GC-content normalization. *Bioinformatics*. 2011 Jan 15;27(2):268–9.
9. Britton GB, O'Bryant SE, Johnson LA, Hall JR, Villarreal AE, Oviedo DC, Pérez-Lao AR, Carreira MB. Inflammatory Biomarkers, Depressive Symptoms and Falls Among the elderly in Panama. *Curr Aging Sci*. 2019;11(4):236–241.
10. Castro, E., Trejos, D., Berovides-Alvarez, V., Arias, T., & Ramos, C. (2007). Genetic Polymorphism and Forensic Parameters of Nine Short Tandem Repeat Loci in Ngöbe and Emberá. *Human Biology*, 79(5), 563–577.
11. Castro-Pérez, E., Trejos, D. E., Hrbek, T., Setaluri, V., & Ramos, C. W. (2016). Genetic Ancestry of the Panamanian Population: Chibchan Amerindian Genes; and Biological Perspectives on Diseases. *The Internet Journal of Biological Anthropology*, 9(1).
12. Chande, A., Nagar, S., Rishishwar, L., Mariño, L., et al (2021). The Impact of Ethnicity and Genetic Ancestry on Disease Prevalence and Risk in Colombia. *Frontiers in genetics*, 12, 690366.
13. Cheng CY, Reich D, Coresh J, et al. 2010. Admixture Mapping of Obesity-related Traits in African Americans: The Atherosclerosis Risk in Communities (ARIC) Study. *Obesity (Silver Spring)*. Mar;18(3):563–72.

14. Cheng, Y., He, C., Wang, M. *et al.* Targeting epigenetic regulators for cancer therapy: mechanisms and advances in clinical trials. *Sig Transduct Target Ther* 4, 62 (2019).
15. Chiurillo, M., Morán, Y., Cañas, M., Valero G., (2010). Distribución del Polimorfismo del Codón 72 de TP53 en Venezuela: Implicaciones Étnicas y Geográficas. *Interciencia*. 35. 202-206.
16. Collins DW, Jukes TH. Rates of transition and transversion in coding sequences since the human-rodent divergence. *Genomics*. 1994 Apr;20(3):386-96.
17. Croft D, O'Kelly G, Wu G, Haw R, Gillespie M, et al., Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res*. 2011 Jan;39(Database issue):D691-7.
18. Dankort D, Curley DP, Cartlidge RA, Nelson B, Karnezis AN, et al. Braf(V600E) cooperates with Pten loss to induce metastatic melanoma. *Nat Genet*. 2009 May;41(5):544-52.
19. Davydov EV, Goode DL, Sirota M, Cooper GM, Sidow A, Batzoglou S. Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Comput Biol*. 2010 Dec 2;6(12):e1001025.
20. Danecek P, Bonfield JK, Liddle J, et al. Twelve years of SAMtools and BCFtools. *Gigascience*. 2021;10(2):giab008.
21. Desaulniers D, Vasseur P, Jacobs A, Aguila MC, Ertych N, Jacobs MN. Integration of Epigenetic Mechanisms into Non-Genotoxic Carcinogenicity Hazard Assessment: Focus on DNA Methylation and Histone Modifications. *Int J Mol Sci*. 2021 Oct 11;22(20):10969.
22. Donnan GA, Fisher M, Macleod M, et al. 2008. Stroke. *Lancet* 371 (9624): 1612-23.
23. Frosst, P., Blom, H. J., Milos, R., Goyette, P., Sheppard, C. A., et al. (1995). A candidate genetic risk factor for vascular disease: a common mutation in methylenetetrahydrofolate reductase. *Nature Genetics*, 10(1), 111-113.
24. Goyette, P., Sumner, J. S., Milos, R., Duncan, A. M. V., Rosenblatt, D. S., et al. (1994). Human methylenetetrahydrofolate reductase: isolation of cDNA, mapping and mutation identification. *Nature Genetics*, 7(2), 195-200.
25. Graydon, J. S., Claudio, K., Baker, S., Kocherla, M., Ferreira, M., et al. (2019). Ethnogeographic prevalence and implications of the 677C>T and 1298A>C MTHFR polymorphisms in US primary care populations. *Biomarkers in Medicine*, 13(8), 649-661.
26. Griffiths, A.J.F.; Wessler, S.R.; Lewontin, R.C. y Carroll, S.B. (2008) *Genética*. 9ª. edición. McGraw-Hill. Interamericana. Madrid.
27. Gurdasani, D., Barroso, I., Zeggini, E. et al. (2019). Genomics of disease risk in globally diverse populations. *Nat Rev Genet*, 20, 520-535.
28. Hahn WC, Weinberg RA. Rules for making human tumor cells [published correction appears in *N Engl J Med*. 2003 Feb 13;348(7):674.]. *N Engl J Med*. 2002;347(20):1593-1603.

29. Hankey, G. J. (1999). Smoking and risk of stroke. *Journal of Cardiovascular Risk*, 6, 207–211.
30. Herceg Z. Epigenetics and cancer: towards an evaluation of the impact of environmental and dietary factors. *Mutagenesis*. 2007 Mar;22(2):91–103.
31. Huang, L. (2014). Zinc and its transporters, pancreatic β -cells, and insulin metabolism. En *Vitamins and Hormones* (1.a ed., Vol. 95). Elsevier Inc.
32. Huang, Q., Yin, J. Y., Dai, X. P., Wu, J., Chen, X., et al. (2010). Association analysis of SL-C30A8 rs13266634 and rs16889462 polymorphisms with type 2 diabetes mellitus and repaglinide response in Chinese patients. *European Journal of Clinical Pharmacology*, 66(12), 1207–1215.
33. Human Genome Variation Sequence. (s. f.). Nomenclatura internacional para el informe de variaciones de secuencia del genoma humano. En *Grupo CAHT*. Grupo CAHT. Recuperado 8 de abril de 2023, de <https://www.grupocaht.com/wp-content/uploads/2019/04/Nomenclatura-Internacional.pdf>.
34. Illumina. An introduction to Next-Generation Sequencing Technology (2015). (https://www.illumina.com/content/dam/illumina-marketing/documents/products/illumina_sequencing_introduction.pdf).
35. Illumina. Nextera XT DNA Library Preparation Kit data sheet (2014). (www.illumina.com/documents/products/datasheets/datasheet_nextera_xt_dna_sample_prep.pdf).
36. Illumina Inc. (2017). Illumina sequencing introduction. *Illumina sequencing introduction, October, 1–8*.
37. Instituto Conmemorativo Gorgas de Estudios de la Salud (2015).
38. Instituto Conmemorativo Gorgas de Estudios de la Salud. (2011). *Sistema de Información de Enfermedades Cardiovasculares en Panamá. 2001–2009*.
39. Instituto Nacional de Estadística y Censo (INEC). (2015). *Principales enfermedades en Panamá*. CONTRALORÍA GENERAL DE LA REPÚBLICA Instituto Nacional de Estadística y Censo. <https://www.inec.gob.pa/archivos/P7731Comentarios.pdf>
40. Sapienza C, Issa JP. Diet, Nutrition, and Cancer Epigenetics. *Annu Rev Nutr*. 2016 Jul 17;36:665–81.
41. *Sistema de Información Geográfico de Enfermedades Cardiovasculares y Diabetes en la República de Panamá. Años 2001–2013*. Instituto Conmemorativo Gorgas.
42. Stenson PD, Ball EV, Mort M, et al. Human Gene Mutation Database (HGMD): 2003 update. *Hum Mutat*. 2003;21(6):577–581.
43. Instituto Oncológico Nacional (2019). *Instituto Oncológico de Panamá*. Instituto Oncológico Nacional.
44. Instituto Nacional de Estadística y Censo (2018). *DEFUNCIONES Y TASA DE MORTALIDAD DE LAS PRINCIPALES CAUSAS DE MUERTE EN LA REPÚBLICA, POR SEXO, SEGÚN CAUSA*.

45. <https://www.inec.gob.pa/archivos/P0579518620191205111309Cuadro%2011.pdf>
46. Jorge-Nebert, L. F., Eichelbaum, M., Griese, E.-U., Inaba, T., & Arias, T. D. (2002). Analysis of six SNPs of NAT2 in Ngawbe and Embera Amerindians of Panama and determination of the Embera acetylation phenotype using caffeine. En *Pharmacogenetics* (Vol. 12).
47. Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, et al. KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* 2008 Jan;36(Database issue):D480-4.
48. Kelly, P. M., Weinberg, A. D., Bernell, S., & Howard, S. W. (2016). Use Your Words Carefully: What Is a Chronic Disease? *Front. Public Health*, 4(159).
49. Khalil, R., Al-Awaida, W. J., Al-Ameer, H. J., Jarrar, Y., Imraish, A., et al. (2021). Investigation of ACE rs4646994, MTHFR rs1801133 and VDR rs2228570 Genotypes in Jordanian Patients with Fibromyalgia Syndrome. *Endocrine, metabolic & immune disorders drug targets*, 21(10), 1920–1928.
50. Kircher M, Witten DM, Jain P, O’Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet.* 2014 Mar;46(3):310-5.
51. Klug, W.S., Cummings, M.R. y Spencer, C.A. (2006) *Conceptos de Genética*. 8ª. edición. Pearson Prentice Hall. Madrid: 213–239.
52. Knudson AG. Two genetic hits (more or less) to cancer. *Nat Rev Cancer.* 2001;1(2):157-162.
53. Leclerc, D., Sibani, S., & Rozen, R. (2013). Molecular Biology of Methylenetetrahydrofolate Reductase (MTHFR) and Overview of Mutations/Polymorphisms. En *Madame Curie Bioscience Database [Internet]*. Landes Bioscience.
54. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009;25(14):1754-1760.
55. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 2009;25(16):2078-2079.
56. Li, M. M., Datto, M., Duncavage, E. J., Kulkarni, S., Lindeman, N. I. et al. (2017). Standards and Guidelines for the Interpretation and Reporting of Sequence Variants in Cancer: A Joint Consensus Recommendation of the Association for Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists. En *Journal of Molecular Diagnostics* (Vol. 19, Número 1, pp. 4–23). Elsevier B.V.
57. Lu, Y., Chan, YT., Tan, HY. *et al.* Epigenetic regulation in human cancer: the potential role of epi-drug in cancer therapy. *Mol Cancer* 19, 79 (2020).
58. Maleknia M, Ahmadirad N, Golab F, Katebi Y, Haj Mohamad Ebrahim Ketabforoush A. DNA Methylation in Cancer: Epigenetic View of Dietary and Lifestyle Factors. *Epigenet Insights.* 2023 Sep 15;16:25168657231199893.
59. Mancarella, D., Plass, C. Epigenetic signatures in cancer: proper controls, current challenges and the potential for clinical translation. *Genome Med* 13, 23 (2021).

60. Mariscal Davy, R. R. (2021). Características relacionadas al sobrepeso y obesidad en estudiantes de la facultad de medicina de la Universidad de Panamá en diciembre 2018. *Ciencia e Investigación Médico Estudiantil Latinoamericana*, 25(1).
61. Maróti, Z., Boldogkoi, Z., Tombácz, D., Snyder, M., & Kalmár, T. (2018). Evaluation of whole exome sequencing as an alternative to BeadChip and whole genome sequencing in human population genetic analysis. *BMC Genomics*, 19(778), 1–13.
62. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010;20(9):1297–1303.
63. Ministerio de Salud. (2018). *Análisis de Situación de Salud. Macro Visión Nacional de Salud.*
64. Ministerio de Salud (MINSA). (2019). *Boletín Estadístico del 2019.* Anuarios estadísticos. <https://www.minsa.gob.pa/contenido/anuario-estadistico-del-2019>.
65. Organización Panamericana de la salud. (2016). *LAS ENFERMEDADES NO TRANSMISIBLES Y SUS FACTORES DE RIESGO, ABORDAJE EN PANAMÁ.* [https://www3.paho.org/pan/dmdocuments/Las enfermedades no transmisibles y sus factores de riesgo, abordaje en Panamá.pdf](https://www3.paho.org/pan/dmdocuments/Las%20enfermedades%20no%20transmisibles%20y%20sus%20factores%20de%20riesgo,%20abordaje%20en%20Panam%C3%A1.pdf)
66. REPÚBLICA, C. G. D. LA, & Censo, I. N. de E. y. (2021). *Mortalidad Panamá 2021.* <https://www.inec.gob.pa/archivos/P053342420221229142845Comentario.pdf>
67. Naidoo, P., Naidoo, R., Ramkaran, P., Chuturgoon, A., (2020). Effect of maternal HIV infection, BMI, and NOx air pollution exposure on birth outcomes in South African pregnant women genotyped for the p53 Pro72Arg (rs1042522). *International journal of immunogenetics*, 47(5), 414–429.
68. Neves, W. A., & Hubbe, M. (2005). Cranial morphology of early Americans from Lagoa Santa, Brazil: Implications for the settlement of the New World. *Proceedings of the National Academy of Sciences of the United States of America*, 102(51), 18309–18314.
69. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* 2003 Jul 1;31(13):3812–4.
70. Olivier, M., Hollstein, M., & Hainaut, P., (2010). TP53 mutations in human cancers: origins, consequences, and clinical use. *Cold Spring Harbor perspectives in biology*, 2(1), a001008.
71. Organización Panamericana de la Salud (OPS). 2022 <https://www.paho.org/es/temas/cancer>
72. *PANAMÁ: PERFIL DE ENFERMEDADES CARDIOVASCULARES.* (2014). Organización Panameña de la Salud.
73. Petersen, D., Kong, A., Jorge, L., Nebert, D., & Arias, T. (1991). Debrisoquine polymorphism: novel CYP2D6 gene Bam HI restriction fragment length polymorphism in the Ngawbé Guaymí Indian of Panama. *Pharmacogenetics*, 1(3), 136–142.
74. Plan Estratégico Nacional para la Prevención y Control del Cáncer, 2019–2029, MINSA–Panamá, CSS–Panamá y OPS/OMS.

75. Plan estratégico nacional para el desarrollo de la ciencia, la tecnología e innovación (PENCIYT) 2019-2024, SENACYT.
76. Popejoy, A., Fullerton, S. (2016). Genomics is failing on diversity. *Nature*, 538, 161–164.
77. Poplin R, Ruano-Rubio V, DePristo MA, Fennell TJ, Carneiro MO, et al. (2017). Scaling accurate genetic variant discovery to tens of thousands of samples *bioRxiv*, 201178.
78. Priya, S. S., Sankaran, R., Ramalingam, S., Sairam, T., & Somasundaram, L. S. (2016). Genotype phenotype correlation of genetic polymorphism of PPAR gamma gene and therapeutic response to pioglitazone in type 2 diabetes mellitus- a pilot study. *Journal of Clinical and Diagnostic Research*, 10(2), 11–14.
79. Quiel L, Moreno Velásquez I, Gómez B, Motta J, Herrera-Ballesteros V. Social determinants and cardiovascular disease mortality in Panama, 2012-2016. *BMC Public Health*. 2019 Feb 15;19(1):199.
80. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010 Mar 15;26(6):841-2.
81. Ramos, C., Castro-Pérez, E., Molina-Jiron, C., & Trejos, D. (2018). Analysis of 30 INDEL Polymorphic Markers in the Panamanian Population: Gene Admixture Estimates, Population Structure and Forensic Parameters. *Journal of Forensic Research*, 09(01).
82. Rausch T, Zichner T, Schlattl A, Stuetz AM, Benes V, Korbel JO. DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics*. 2012 Sep 15;28(18):i333-i339.
83. Reiner, A., Carlson, C., Ziv, E., Iribarren, C., Jaquish, C., Nickerson, D., (2007) Genetic ancestry, population sub-structure, and cardiovascular disease-related traits among African American participants in the CARDIA Study. *Hum Genet*. 121: 565–575.
84. Reva B, Antipin Y, Sander C: Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic acids research* 2011, 39(17):e118.
85. Reyes, G. (2022, abril 8). Enfermedades no transmisibles representan el 52% del total de las causas de muertes. *La Prensa*. <https://www.prensa.com/sociedad/enfermedades-no-transmisibles-representan-el-52-del-total-de-las-causas-de-muertes/>
86. Reyes, L., Godfrey, D., Ming, L. J., MacLean, C., Gonzalez, F. J., & Madrigal, L. (2021). The distribution in native populations from Mexico and Central America of the C677T variant in the MTHFR gene. *American Journal of Human Biology*, 33(6).
87. Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., et al. (2015). Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine*, 17(5), 405–424.
88. Robbins, C., Torres, J. B., Hooker, S., Bonilla, C., Hernandez, W., et al. (2007). Confirmation study of prostate cancer risk variants at 8q24 in African Americans identifies a novel risk locus. *Genome Research*, 17(12), 1717.

89. Rogers MF, Shihab HA, Mort M, Cooper DN, Gaunt TR, Campbell C. FATHMM-XF: accurate prediction of pathogenic point mutations via extended features. *Bioinformatics*. 2018 Feb 1;34(3):511–513.
90. Rosenberg, N., Murata, M., Ikeda, Y., Opare-Sem, O., Zivelin, A., et al. (2002). The Frequent 5,10-Methylenetetrahydrofolate Reductase C677T Polymorphism Is Associated with a Common Haplotype in Whites, Japanese, and Africans. En *Am. J. Hum. Genet* (Vol. 70).
91. Roychoudhury, A. K., & Nei, Masatoshi. (1988). *Human Polymorphic Genes: World Distribution*. Oxford University Press.
92. Rubio, S., Pacheco-Orozco, R. A., Gómez, A. M., Perdomo, S., & García-Robles, R. (2020). Secuenciación de nueva generación (NGS) de ADN: presente y futuro en la práctica clínica. *Universitas Médica*, 61(2), 1–15.
93. Sant Joan de Déu Barcelona Hospital. (2014, noviembre 11). *El exoma y su papel en el diagnóstico de errores congénitos del metabolismo*.
94. Sasson, M., Lee, M., Jan, C., Fontes, F., & Motta, J. (2014). Prevalence and Associated Factors of Obesity among Panamanian Adults. 1982–2010. *PLOS ONE*, 9(3), e91689-.
95. Schaefer CF, Anthony K, Krupa S, Buchoff J, Day M, et al. PID: the Pathway Interaction Database. *Nucleic Acids Res*. 2009 Jan;37(Database issue):D674–9.
96. Schwarz JM, Cooper DN, Schuelke M, Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. *Nat Methods*. 2014 Apr;11(4):361–2.
97. Sirugo, G., Williams, S. M., & Tishkoff, S. A. (2019). The Missing Diversity in Human Genetic Studies. *Cell*, 177(1), 26–31.
98. Sollis E, Mosaku A, Abid A, Buniello A, Cerezo M, et al. The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. *Nucleic Acids Res*. 2023 Jan 6;51(D1):D977–D985.
99. Takeshima, H., Ushijima, T. Accumulation of genetic and epigenetic alterations in normal cells and cancer risk. *npj Precis. Onc*. 3, 7 (2019).
100. Tate JG, Bamford S, Jubb HC, et al. COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res*. 2019;47(D1):D941–D947.
101. Usui G, Matsusaka K, Huang KK, Zhu F, Shinozaki T, et al. (2023). Integrated environmental, lifestyle, and epigenetic risk prediction of primary gastric neoplasia using the longitudinally monitored cohorts. *eBioMedicine*. 2023; 98, 104844, December. The Lancet Discovery Science.
102. Villarreal AE, Grajales S, Lopez L, Britton GB; Panama Aging Research Initiative. Cognitive Impairment, Depression, and Cooccurrence of Both among the Elderly in Panama: Differential Associations with Multimorbidity and Functional Limitations. *Bio-med Res Int*. 2015;2015:718701.

103. Villarreal AE, Pérez-Lao AR, Oviedo DC, Grajales S, Carreira MB, Britton GB. The Panama Aging Research Initiative Longitudinal Study. *MEDICC Rev.* 2019 Apr-Jul;21(2-3):54-58.
104. Wang, K., Li, M., & Hakonarson, H. (2010). Annovar: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research*, 38(16), e164.
105. Wang, X., Fu, J., Li, Q., & Zeng, D. (2016). Geographical and ethnic distributions of the MTHFR C677T, A1298C and MTRR A66G gene polymorphisms in Chinese populations: A meta-analysis. *PLoS ONE*, 11(4).
106. Wassel, C., Pankow, J., Peralta, C., Choudhry, S., Seldin, M., Arnett, D., (2009) Genetic ancestry is associated with subclinical cardiovascular disease in blacks and Hispanics from the Multi-Ethnic Study of Atherosclerosis. *Circ Cardiovasc Genet.* 2: 629–636
107. Westenberger, S. (2020). Illumina Sequencing Overview: Library Prep to Data Analysis with Scott Westenberger, PhD - View Webinar | Ambry Genetics. En *Ambry Genetics*. Ambry Genetics. <https://www.ambrygen.com/providers/webinar/134/illumina-sequencing-overview-library-prep-to-data-analysis-with-scott-westenberger-phd>
108. Witt, M., & Erickson, R. P. (1989). A rapid method for detection of Y-chromosomal DNA from dried blood specimens by the polymerase chain reaction. En *Hum Genet* (Vol. 82).
109. Wolfe, J., Darling, S. M., Erickson, R. P., Craig, I. W., Buckle, V. J., (1985). Isolation and characterization of an alphoid centromeric repeat family from the human Y chromosome. *Journal of Molecular Biology*, 182(4), 477–485.
110. World Health Organization. (2005). *WHO steps surveillance manual: the WHO stepwise approach to chronic disease risk factor surveillance*. WHO.
111. World Health Organization. (2016). *Noncommunicable diseases*. <https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases>
112. Zeigler-Johnson, C. M., Spangler, E., Jalloh, M., Gueye, S. M., Rennert, H., & Rebbeck, T. R. (2008). Genetic susceptibility to prostate cancer in men of African descent: implications for global disparities in incidence and outcomes. *The Canadian journal of urology*, 15(1), 3872–3882.

CAPÍTULO 9

Identificación de dos
SNPs en los genes MTHFR
(rs1801133) y TP53
(rs1042522) como factores
de riesgo a cánceres y
otras enfermedades en
amerindios Ngöbe de
Panamá: significancia
biomédica en la población
mestiza panameña

Edgardo Castro-Pérez, Carlos Ramos-Delgado, Germán Espino, Beatrice Di Biase, Kathia Canto-Díaz, Madián Poveda, Caleb Herrera, Nathan Herrera, Vijayasradhi Setaluri, Reynaldo Arosemena, Carmen Mela-Sánchez, Marina Pacheco, Jorge Lasso De La Vega, Pablo Noel, Fernando Díaz, Idalina Cubilla-Batista, Isabel Blanco, José Robinson-Duggon.

Resumen

Las enfermedades no transmisibles (ENT) representan las principales causas de muerte por enfermedades en Panamá. Reportes indican que existen disparidades ancestrales en el riesgo de padecer ENT en la población del país, es decir, la prevalencia discrepa de acuerdo al trasfondo ancestral predominante como africano, europeo o amerindio. Estas evidencias sugieren fuertemente que el componente genético de la población es un factor de riesgo altamente determinante en la distribución, susceptibilidad y muertes por ENT en el país. Sin embargo, no existen estudios biomédicos sobre la población que aborden científicamente la epidemiología genética de estas enfermedades porque se desconocen los genes y los polimorfismos genéticos asociados con las mismas en Panamá. Identificar estos polimorfismos es fundamental para comprender la susceptibilidad genética y su relación con factores de riesgo ambiental y estilos de vida en la población. Además, contribuiría a mejorar las estrategias de salud en grupos susceptibles, su diagnóstico y posible desarrollo de tratamientos. Estudios previos han demostrado que el componente ancestral de la población panameña es dominado por genes de origen Ngöbe con un 51%. Hipotetizamos que una significativa parte de los polimorfismos genéticos asociados a enfermedades de la población panameña puede tener un origen en esta población ancestral amerindia. Por lo tanto, con el objetivo de identificar polimorfismos genéticos posiblemente asociadas a ENT en la población panameña, en el capítulo anterior se secuenció el genoma de diez individuos Ngöbe. Los análisis bioinformáticos condujeron a la identificación de dieciséis (16) genes con polimorfismos SNP posiblemente asociados a ENT como candidatos patogénicos en esta población. En el presente capítulo, ampliar los estudios enfocados en dos de esos polimorfismos SNP: uno en el gen MTHFR (rs1801133, 677C>T) y otro en el gen TP53 (rs1042522, 215C>G), los cuales fueron genotipados en una muestra de 60 individuos Ngöbe para determinar su frecuencia alélica en esta población. El polimorfismo rs1801133 ha sido asociado previamente en otras poblaciones de mayor riesgo de hiperhomocisteinemia y tiene asociación a cánceres y a enfermedades vasculares; mientras

que el polimorfismo rs1042522 ha sido asociado a diferentes tipos de cáncer. Los resultados demostraron que en la población Ngöbe, el polimorfismo rs1801133 (MTHFR) tiene una frecuencia alélica para los alelos “C” y “T” de 15% y 85%, respectivamente; así como una frecuencia genotípica de 0% para los genotipos homocigoto CC, 31% heterocigoto CT y 69% para el homocigoto TT. Por otro lado, el polimorfismo rs1042522 (TP53) mostró una frecuencia de 56% y 44%, para los alelos “C” y “G”, respectivamente; y una frecuencia genotípica de 24% para los genotipos homocigoto CC, 64% heterocigoto CG, y 12% para el homocigoto GG. Para confirmar si hay asociación o no de estos SNPs con estas enfermedades estudios en curso analizan los patrones genéticos de los mismos en grupos de pacientes diagnosticados o de alto riesgo y se comparan con grupos control o de bajo riesgo en la población mestiza del país.

Palabras clave: polimorfismo de nucleótido simple (SNP), genotipos patogénicos, riesgo genético, genes ancestrales Ngöbe.

Abstract

Non-communicable diseases (NCDs) represent the leading causes of death from diseases in Panama. Reports indicate ancestral disparities in the risk of NCDs in the country's population; their prevalence differs according to the predominant ancestral background, such as African, European, or Amerindian. This evidence strongly suggests that the genetic component of the population is a highly determining risk factor in the country's distribution, susceptibility, and deaths from NCDs. However, there are no biomedical studies on the population that scientifically address the genetic epidemiology of these diseases because the genes and genetic polymorphisms associated with them in Panama are unknown. Identifying these polymorphisms is essential to understanding genetic susceptibility and its relationship with the population's environmental risk factors and lifestyles. In addition, it would contribute to improving health strategies in susceptible groups, their diagnosis, and the possible development of treatments. Previous studies have shown that genes of Ngöbe origin dominate the ancestral component of the Panamanian population with 51%. We hypothesize that a significant part of the genetic polymorphisms associated with diseases of the Panamanian population may have an ori-

gin in this ancestral Amerindian population. Therefore, in the previous chapter, the genome of ten Ngöbe individuals was sequenced to identify genetic polymorphisms possibly related to NCDs in the Panamanian population. Bioinformatics analyses identified sixteen (16) genes with SNP polymorphisms perhaps associated with NCDs as pathogenic candidates in this population. In this chapter, we expand the studies focused on two of these SNP polymorphisms: one in the MTHFR gene (rs1801133, 677C>T) and another in the TP53 gene (rs1042522, 215C>G), which were genotyped in a sample of 60 Ngöbe individuals to determine their allelic frequency in this population. The rs1801133 polymorphism has previously been associated in other populations with a higher risk of hyperhomocysteinemia and is associated with cancers and vascular diseases. In contrast, the rs1042522 polymorphism has been associated with different types of cancer. The results showed that in the Ngöbe population, the rs1801133 (MTHFR) polymorphism has an allelic frequency for the “C” and “T” alleles of 15% and 85%, respectively, as well as a genotypic frequency of 0% for the homozygous CC, 31% heterozygous CT and 69% for the homozygous TT genotypes. On the other hand, the rs1042522 (TP53) polymorphism showed a frequency of 56% and 44% for the “C” and “G” alleles, respectively, and a genotypic frequency of 24% for the homozygous CC, 64% for heterozygous CG, and 12% for the homozygous GG genotypes. To confirm whether or not these SNPs are associated with these diseases, ongoing studies analyze their genetic patterns in groups of diagnosed or high-risk patients and compare them with control or low-risk groups in the country’s mestizo population.

Keywords: single nucleotide polymorphism (SNP), pathogenic genotypes, genetic risk, Ngöbe ancestral genes.

Introducción

Las estadísticas del MINSA y del Instituto Nacional de Estadística y Censo (INEC, Contraloría General de la República) indican que en los últimos 30 años las tasas de morbilidad y mortalidad más altas en la población panameña son debido a las enfermedades no transmisibles (ENT). Estas enfermedades incluyen cánceres, enfermedades cardiovasculares, cerebrovasculares y otras enfermedades relacionadas con problemas metabólicos como obesidad y diabetes (Estadísticas de Salud, MINSA Panamá, 2018). Estos reportes indican que la prevalencia de estas enfermedades ha ido en aumento en los últimos años y el cáncer ha liderado los incrementos más significativos tanto en prevalencia como en muertes en el país. Por ejem-

plo, cifras oficiales del Registro Nacional Poblacional del Cáncer indican que entre los años 2000 - 2009 la incidencia de cáncer por cada 100.000 habitantes fue en promedio de 43.7; mientras que estudios más recientes reportan una prevalencia de 146,8, lo cual equivale a más de 3,000 muertes anuales (OPS, 2022; Politis, et al., 2015; Inst. Gorgas 2015; Jorge y Haughton 2018; MINSA 2018; The Global Cancer Observatory, OMS 2020; Sung et al., 2021). Estos datos indican, además, que el cáncer de próstata es el cáncer más devastador en el país con unos 1,500 nuevos casos anuales que representan el 19% de todos los cánceres y es la principal causa de muerte por cáncer en el país. Le siguen el cáncer de mama con más de 1,000 nuevos casos anuales (14%), el más alto entre las mujeres. Luego, cáncer colo-rectal con aproximadamente 750 casos (9.5%), estómago 470 casos (6.1%), cérvico-uterino con 346 casos (4.5%) y el resto de múltiples cánceres con 3,570 nuevos casos de varios tipos que representan el 46.5% (Jorge y Haughton 2018; Motta y Politis, et al., 2015; MINSA 2018; OMS 2020; Sung et al., 2021). Dado que la mayoría de los cánceres se detectan en pacientes mayores de 60 años se estima que, con el aumento tanto en expectativa de vida como en el envejecimiento de la población, para el 2030 el número de diagnósticos de cáncer aumentará significativamente tanto en el mundo como en el país (The Global Cancer Observatory, OMS, 2020). Este incremento en casos ya se refleja en cifras recientes, donde se ha detectado un aumento significativo en el número de pacientes de cáncer menores de 50 años (Koh et al., 2023; Giannakis y Ng, 2023).

A pesar de que el cáncer es una de las enfermedades más prevalentes y está entre las principales causas de muerte en Panamá, actualmente no se estudia científicamente de una manera integral en la población panameña. En particular, algo que se desconoce es su epidemiología genética y la relación entre factores genéticos y ambientales en el riesgo de padecer esta enfermedad en el país. Estudiar la genética del cáncer en nuestra población es muy importante, ya que múltiples reportes señalan que existen diferencias en genética ancestral asociadas con disparidades significativas tanto en la prevalencia como en muertes; así como en la respuesta terapéutica/farmacogenética asociada a la sobrevivencia de algunos tratamientos (Shavers y Brown, 2002; O'Donnell et al., 2009; Guerrero et al., 2018; Walsh y Goh, 2019; Shubeck et al., 2023; Pathak et al., 2023; Roy et al., 2023; Stabellini et al., 2023; Zhao et al., 2023; Terman et al., 2023). Por ejemplo, múltiples estudios señalan que hombres cuyo trasfondo ancestral muestra alta contribución de genes africanos, presentan una prevalencia 60% mayor y una tasa de mortalidad entre 2 a 3 veces más alta al cáncer de próstata que hombres de ascendencia europea (Cotter et al.,

2002; Hoffman et al., 2001; Thompson et al., 2001; Robbins et al., 2007; Zeigler-Johnson et al., 2008; Bock et al., 2009; Farrell et al., 2013).

Consistente con estos reportes, estudios genéticos de la población panameña con marcadores microsatélites (STRs) e (InDels) indican que el cáncer de próstata muestra por muy lejos la mayor prevalencia y muertes en las provincias de Panamá y Colón, las cuales son las provincias con la mayor proporción de genes ancestrales de origen africano (Castro-Pérez et al., 2016; Ramos et al., 2018). Sin embargo, se desconocen los genes y los polimorfismos asociados a mayor riesgo de padecer este y otros cánceres. Identificar los polimorfismos involucrados en cánceres es fundamental para comprender el riesgo de cómo estos genes interactúan con factores ambientales y estilos de vida que aumentan la probabilidad de desarrollar tumores malignos. Adicionalmente, la identificación de estos polimorfismos podría contribuir a mejorar diagnósticos y tratamientos; así como al desarrollo de la medicina genómica de precisión y dirigir políticas de salud a grupos susceptibles.

Para abordar esto, en el capítulo anterior secuenciamos por primera vez el genoma completo de 10 amerindios Ngöbe, quienes son los ancestros de los panameños modernos y predominan en el componente genético de la población con un 51%. Por lo tanto, hipotetizamos que una proporción significativa de los polimorfismos genéticos posiblemente asociados con riesgo a enfermedades como el cáncer en la población panameña podría tener su origen en esta población amerindia ancestral. Los análisis genómicos demostraron que los marcadores moleculares más informativos asociados clínicamente a ENT fueron los SNPs, entre los cuales se identificaron dieciséis variantes posiblemente patogénicas de mayor significancia clínica. Entre estos polimorfismos seleccionamos dos SNPs posiblemente asociados a tumores malignos (cáncer) para estudios más detallados y determinar su frecuencia alélica en una muestra de 60 individuos Ngöbe: rs1801133 en el gen MTHFR y el polimorfismo rs1042522 en el gen TP53, lo cual presentamos en este capítulo.

Estos polimorfismos han sido previamente asociados a cánceres, trastornos vasculares y congénitos, entre otros, en diferentes poblaciones del mundo que describimos brevemente a continuación. El gen MTHFR codifica una enzima llamada metilentetrahidrofolato reductasa, la cual es un regulador clave en el metabolismo del folato y de la homocisteína. El gen MTHFR está localizado en el cromosoma 1 (1p36.22.) y el mayor producto del gen humano MTHFR es una proteína de 77 kDa con 697 aminoácidos

(Goyette et al., 1995; Gaughan et al., 2000; NCBI, 2023; NCBI, 2023; Mudd et al., 1972). Se han descrito múltiples variantes clínicamente relevantes en el gen MTHFR, entre las cuales el polimorfismo rs1801133 identificado en el genoma Ngöbe, es una mutación puntual en la posición 665 de la región codificante del exón 5 donde se reemplaza citosina o timina (C/T) (c.665C>T) y es también conocido comúnmente como C677T. Este polimorfismo codifica el aminoácido alanina (A/Ala) (C, en el ADN) o valina (V/Val) (T, en el ADN) en la posición 222 (p.Ala222Val) de la proteína (den Dunnen et al., 2016; Human Genome Variation Sequence, s. f.). El SNP rs1801133 ha sido identificado como un factor de riesgo para varios trastornos que incluyen enfermedades vasculares, neurológicas, varios tipos de cánceres, diabetes y pérdida del embarazo (Froese et al., 2018; Rosenberg et al., 2002; Goyette et al., 1994; Goyette et al., 1995; Rozen, 1997; Frosst et al., 1995; Leclerc et al., 2013).

Por otro lado, el gen TP53 se encuentra en el brazo corto del cromosoma 17 (17p13) y en humanos codifica una proteína de 393 aminoácidos llamada proteína tumoral 53 (“*tumor protein p53*”, o “p53”), la cual es un factor de transcripción y es un supresor de tumores, por lo que es conocida como el guardián del genoma (Ziener et al., 1982; Bourdon et al., 2005; Surget et al., 2013). La proteína p53 responde a diversos estreses e interacciones celulares que aún no se conocen en su totalidad estimulando la transcripción de diversos genes inhibidores de proliferación que detienen el ciclo celular en la transición de G₁-S, estimulan apoptosis, senescencia y la reparación del ADN o generando cambios en el metabolismo (Chumakov et al., 2000; Bargonetti et al., 2002; Rangel-López et al., 2006; Ingaramo et al., 2018; Kumar et al., 2020). TP53 es el gen más mutado en cánceres (base de datos IARC TP53, R13; Olivier et al., 2010; Surget et al., 2013; de Andrade et al., 2022). En particular, el polimorfismo rs1042522 identificado en el genoma Ngöbe corresponde a un SNP en la posición 215 del exón 4 donde se reemplaza citosina o guanina (C/G) (c.215C>G) codificando el aminoácido prolina (P/Pro) (C, en el ADN) o arginina (R/Arg) (G, en el ADN) en el codón 72 de la proteína (p.Pro72Arg) (Mohammed Basabaeen et al., 2019). Este polimorfismo ha sido asociado a varios cánceres, incluyendo cánceres hereditarios (Brady et al., 2010; Litviakov et al., 2010), uterino (Klug et al., 2009), páncreas (Naccarati et al., 2010; Sonoyama et al., 2011), mama (Johnson et al., 2007; Ahmed et al., 2023), hueso (Huang et al., 2019), colorectal (Polakova et al., 2009); próstata (Fan et al., 2017), pulmón (Han et al., 2008) y gástrico, entre otros (Akulevich et al., 2009; Huang et al., 2011; Rangel-López et al., 2006). Además, rs1042522 ha sido asociado a

enfermedad coronaria (Omrani-Nava et al., 2018; Khan et al., 2016), artritis reumatoide (Lee et al., 2012); obesidad (Kung et al., 2016) y diabetes (Bonfigli et al., 2013) entre otras condiciones de salud.

Con el objetivo de determinar la proporción de estos polimorfismos posiblemente patogénicos en la población Ngöbe, determinamos el genotipo de estos SNPs en una muestra de 60 individuos de esta población amerindia ancestral que representa la mayor parte del trasfondo genético del panameño con un 51% (Castro-Pérez et al., 2016). Se realizó PCR y secuenciación de ADN para el genotipado de ambos polimorfismos. Los resultados indican que en la población Ngöbe, el polimorfismo rs1801133 de MTHFR muestra una frecuencia alélica de 15% para el alelo C (alanina) y de 85% para el alelo T (valina); así como una frecuencia genotípica de 0% para los genotipos homocigoto CC, 31% heterocigoto CT y 69% para el homocigoto TT. Por otro lado, el polimorfismo rs1042522 en TP53 mostró una frecuencia de 62% y 38%, para el alelo C (prolina) y para el alelo G (arginina) respectivamente; y una frecuencia genotípica de 52% para los genotipos homocigoto CC, 20% heterocigoto CG y 28% para el homocigoto GG. La alta frecuencia de los alelos mutantes en la población Ngöbe, quienes son los ancestros indígenas de la población panameña actual, y con la mayor contribución de genes (51%), sugiere que estos alelos podrían estar presentes en la población del país. De confirmarse su presencia, estos alelos podrían contribuir junto con otros polimorfismos y factores ambientales a la prevalencia de enfermedades no transmisibles como tumores malignos y vasculares, entre otras. Para confirmar si hay asociación o no de estos SNPs con estas enfermedades, estudios en curso analizan los patrones genéticos de estos polimorfismos en grupos de pacientes diagnosticados y se comparan con grupos control de bajo riesgo en la población general del país.

Metodología

Obtención de muestras, calidad del ADN y sexado

Se utilizaron muestras de ADN Ngöbe del banco de ADN del Departamento de Genética y Biología Molecular de la Universidad de Panamá colectados en proyectos desarrollados, y publicados previamente (Arias et al., 1993; Jorge-Nebert et al., 2002; Petersen et al., 1991; Castro et al., 2007). Se verificó la concentración, pureza y calidad de las muestras, y fueron sexadas como describimos en la metodología del capítulo anterior.

Identificación de los polimorfismos rs1801133 y rs1042522 en genomas Ngöbe

Los análisis bioinformáticos del exoma de diez individuos Ngöbe descritos en el capítulo anterior nos condujo a la identificación de múltiples variantes génicas posiblemente asociadas a ENT. Seleccionamos algunos de estos polimorfismos de relevancia clínica posiblemente patogénicos asociados a cánceres y enfermedades vasculares para estudios más detallados, ya que representan las principales causas de muerte en la población panameña. En particular, nos enfocamos en los polimorfismos rs1801133 del gen MTHFR, el cual está asociado a mayor riesgo de cánceres, enfermedades vasculares y otras descritas en la introducción; y el SNP rs1042522 del gen TP53, asociado principalmente a mayor riesgo de cánceres. Ambos polimorfismos fueron identificados en al menos 5 de los 10 genomas secuenciados en el capítulo anterior. Realizamos el genotipado de ambos polimorfismos mediante PCR y secuenciación de ADN del producto de PCR, y determinamos la frecuencia alélica de ambos polimorfismos en una muestra de 60 individuos no relacionados de la población Ngöbe.

Amplificación por PCR y secuenciación del polimorfismo rs1801133 (MTHFR)

Para el genotipado del polimorfismo rs1801133 realizamos amplificación por PCR y secuenciación de ADN de la región del gen MTHFR que contiene el SNP con algunas adaptaciones de métodos descritos previamente (Khalil et al., 2021) y explicados más abajo. Las secuencias obtenidas fueron luego analizadas mediante varias estrategias que incluyeron análisis de restricción *in silico* (PCR-Secuenciación-RFLP-*in silico*), según fue reportado previamente (Khalil et al., 2021). Se amplificó una región del exón 5 del gen MTHFR que contiene el sitio del SNP con primers (cebadores) y condiciones reportadas previamente (Khalil et al., 2021). La mezcla de reacción fue realizada en un volumen final de 30 μ l que consistieron en 15 μ l de master mix [2X Blastaq™ PCR Taq MasterMix, Cat.# G895, Applied Biological Materials (abm) Inc.], 3 μ l de cada primer (5 μ M cada primer), 2 μ l de ADN y 7 μ l de agua. Como control negativo utilizamos una mezcla de reacción con todos los componentes, excepto que el volumen de ADN (2 μ l)

fue reemplazado con agua libre de nucleasa. Se usaron los primers con las siguientes secuencias (Khalil et al., 2021) forward (FW) y reverse (RV): FW (5'-TGA AGG AGA AGG TGT CTG CGG GA-3') y RV (5'-AGG ACG GTG CGG TGA GAG TG-3'). Las condiciones del termociclador fueron las siguientes: 95 °C por 10 min, seguido de 35 ciclos a 94 °C por 30 s, 62 °C por 30 s y 72 °C por 30 s; y una extensión final de 72 °C por 10 min. Se usaron 3 µL de los productos de PCR para resolverlos y visualizarlos mediante electroforesis en gel de agarosa al 1% en tampón TAE 1X y teñidos con GelRed. Se visualizó una banda del tamaño esperado de 198 pares de bases, la cual fue verificada utilizando un estándar de peso molecular (100 bp PCR ladder). Los resultados fueron foto-documentados digitalmente en un Gel Doc™ EZ System (BioRad). El restante volumen de los productos de PCR (60 muestras en total) fueron embalados en frío y enviados a la empresa Psomagen en la ciudad de Nueva York, Estados Unidos, para secuenciación de ADN de ambas hebras de los primers utilizados para su amplificación.

Amplificación por PCR y secuenciación del polimorfismo rs1042522 (TP53)

El genotipado del polimorfismo rs1042522 se realizó combinando métodos de análisis moleculares de amplificación por PCR y secuenciación de ADN del producto de PCR junto con métodos de análisis de restricción *in silico* (PCR-Secuenciación-RFLP-*in silico*). Esta estrategia fue adaptada basada en métodos descritos previamente (Naidoo et al., 2020). Se amplificó una región del exón 4 del gen TP53 que contiene el sitio del SNP. La reacción fue realizada en 30 µl que consistieron en 15 µl de master mix 2X Blastaq™ PCR Taq MasterMix, Cat.# G895, Applied Biological Materials (abm) Inc.), 3 µl de cada cebador (5µM cada cebador), 2 µl de ADN y 7 µl de agua. Como control negativo utilizamos una mezcla de reacción con todos los componentes, excepto que el DNA templado fue reemplazado con agua. Las secuencias de los primers utilizados fueron: FW-5'-TTG CCG TCC CAA GCA ATG GAT GA-3' y RV-5'- TCT GGG AAG GGA CAG AAG ATG AC-3'. Las condiciones de termociclador fueron las siguientes: 96 °C por 12 min, seguido de 35 ciclos a 94 °C por 30 s, 55 °C por 30 s y 72 °C por 30 s; seguido por una extensión final a 72 °C por 5 min. Los productos de PCR fueron separados mediante electroforesis en gel de agarosa al 1% y teñidos con GelRed. Los resultados fueron foto-documentados digitalmente en un Gel Doc™ EZ System (BioRad). Se visualizaron las bandas del

tamaño esperado de 131 pares de bases, la cual fue verificada utilizando un estándar de peso molecular (100 bp PCR ladder). El restante volumen de los productos de PCR (60 muestras en total) fueron embalados en frío y enviados a la empresa Psomagen en la ciudad de Nueva York, Estados Unidos, para secuenciación de ADN de ambas hebras de los primers utilizados para su amplificación.

Procesamiento y análisis de secuencias para el genotipado

Los datos de secuenciación recibidos consistieron en 60 secuencias *forward* (FW) y 60 secuencias *reverse* (RV), es decir, ambas hebras de cada muestra de ADN/PCR, representando un total de 120 secuencias para cada SNP genotipado. Estas secuencias confirmaron la identidad molecular y posición de los SNP identificados originalmente en los genomas Ngöbe (Capítulo anterior). También utilizamos como estándar las secuencias de cDNA y proteína de referencia del genoma humano del GenBank de los genes *MTHFR* (cDNA: [NM_001330358.2] y proteína [NP_001317287]); y *TP53* ([NM_000546.6] y proteína [NP_000537.3]) en la región correspondiente a la posición del respectivo SNP en estudio. Para determinar el genotipo, a estas secuencias se les realizaron varios análisis y controles con programas y herramientas en el orden como se muestran en la **Tabla 1**.

Tabla 1. Análisis y herramientas utilizadas en el análisis de secuencias para el genotipado.

Paso	Análisis	Programa
1	Ensamblaje de ambas hebras FW y RV	SEQUENCHER 4.1.4.
2	Limpieza y verificación de secuencias	SEQUENCHER 4.1.4.
3	Obtención de secuencias consenso	SEQUENCHER 4.1.4.
4	Verificación del tamaño del fragmento	SEQUENCHER 4.1.4.
5	Verificar la identidad molecular de la secuencia	NCBI-BLAST
6	Análisis de restricción in silico y verificar el patrón de restricción que identifica la posición del alelo/SNP	SEQUENCHER 4.1.4. y/o NEBCutter v3.0.17.
7	Alineamiento de secuencias de nucleótidos y verificar la posición del SNP	MEGA 11
8	Secuencias de proteínas con un predictor de traducción	ORFfinder (Translate tool) de ExpASy
9	Alineamiento de las secuencias de proteínas y verificar la posición del cambio de aminoácido que genera el SNP	MEGA 11
10	Determinar las frecuencias alélicas y genotípicas	PopGene v1.32

Ensamblaje, limpieza, secuencias consenso y verificación del tamaño

Los cuatro primeros pasos para el procesamiento y análisis de las secuencias se realizaron con el programa SEQUENCHER 4.1.4. (Gene Codes Corporation, Ann Arbor, MI USA). Las secuencias de cada hebra de los cebadores FW y RW se importaron a partir de archivos en formato con extensión “.ab1” y se ensamblaron usando los parámetros del programa. Una vez ensambladas, se pudo verificar el tamaño del fragmento en los cromatogramas. La limpieza consistió en revisar los cromatogramas y corregir las bases ambiguas de la secuencia basado en la hebra (FW o RV) de mejor calidad, se eliminaron los segmentos cortos de los extremos con errores o secuencias muy am-

biguas o no específicas y se corrigieron bases según el caso. Al mismo tiempo, se identificó la posición del correspondiente SNP mediante el análisis de restricción *in silico* en SEQUENCHER (explicado más adelante). Al final, se exportó la secuencia consenso para cada muestra en formato FASTA, la cual sería utilizada para los siguientes análisis. Para los homocigotos se generó una sola secuencia consenso porque mostraban un solo pico en el SEQUENCHER en la posición correspondiente a la base del SNP, es decir, era el mismo alelo/base de ADN del gen en ambos cromosomas. Para los heterocigotos, se generaron y exportaron dos secuencias consenso, una por cada alelo, lo cual se verificaba mediante la presencia de picos dobles (sobrepuestos con 2 colores diferentes) en los cromatogramas en la posición del correspondiente SNP (explicado en resultados).

Verificación de la identidad molecular del fragmento secuenciado

La identidad molecular de todas las secuencias de ADN fue verificada en la base de datos de las secuencias del genoma humano de referencia depositadas en GenBank (NCBI-NLB-NIH) utilizando la herramienta *Basic Local Alignment Search Tool* (BLAST) para búsqueda de nucleótidos (Blastn). La búsqueda confirmó las regiones genómicas y transcritos de mayor homología (similitud) con las secuencias de referencia del correspondiente gen en el genoma humano de referencia. Una vez confirmada la identidad molecular de las secuencias nucleotídicas (ADN) con el correspondiente gen (MTHFR o TP53), las mismas fueron analizadas con la herramienta ORFfinder (Translate tool) de “ExPASy” para obtener la predicción de su secuencia de aminoácidos de la proteína. La identidad de las secuencias de proteínas fue confirmada mediante la herramienta Protein BLAST (Blastp) en el genoma humano de referencia. Tanto las secuencias de nucleótidos como de proteínas obtenidas fueron confirmadas y todas coincidieron con su correspondiente gen (MTHFR o TP53). Por lo tanto, no hubo discrepancias en la identidad molecular de las secuencias obtenidas con su correspondiente gen del genoma humano de referencia.

Genotipado mediante análisis de restricción *in silico*

Realizamos análisis de restricción *in silico* de las secuencias nucleotídicas obtenidas para verificar la posición y cambio de base de los correspondientes SNPs, ya que en la posición de los mismos ocurre un sitio de reconocimiento de endonucleasas de restricción de manera específica reportado previamente para MTHFR (Khalil et al., 2021) y para TP53 (Naidoo et al., 2020) que explicamos a continuación. Para los análisis de restricción *in silico* utilizamos la herramienta de análisis de restricción incorporada en el programa SEQUENCHER 4.1.4. y/o con la plataforma NEBCutter v3.0.17.

Para la secuencia de MTHFR se han reportado métodos de genotipado del SNP rs1801133 mediante PCR-RFLP, donde el producto de PCR secuenciado de 198 pb es digerido con la endonucleasa de restricción HinfI generando un patrón de restricción de bandas que revela la identidad del genotipo en el polimorfismo rs1801133 (Khalil et al., 2021). Esta endonucleasa (secuencia de restricción 5'-G*ANTC-3'/3'-CTNA*G-5') corta los productos de PCR en la posición 665 del exón 5 donde está el SNP [citosina (C), o timina (T)], lo cual genera la sustitución de alanina/Ala o valina/Val en la posición 222 (p.Ala222Val) en la proteína. HinfI corta la secuencia del producto de PCR en los alelos homocigotos (T/T) generando dos fragmentos, uno de 175 pb y otro de 23 pb. Por el contrario, en los alelos homocigotos (C/C) este sitio de restricción se pierde y no es reconocido, por lo que el producto de PCR no es cortado dejando la banda original de 198 pb intacta. Por lo tanto, para los individuos con alelos heterocigotos (C/T, mostraban picos dobles en el cromatograma) la digestión con la endonucleasa genera 3 fragmentos uno de 198 pb correspondiente al cromosoma con el alelo "C", sin cortar; y dos fragmentos para el cromosoma con el alelo "T", uno de 175 pb y otro de 23 pb (Khalil et al., 2021).

Para la secuencia de TP53 también se han reportado métodos de genotipado del SNP rs1042522 mediante PCR-RFLP, donde el producto de PCR secuenciado de 131 pb es digerido con la endonucleasa de restricción BstUI (secuencia de restricción 5'-CG*CG-3'/5'-GC*GC-3') generando un patrón de restricción de bandas que revela la identidad y genotipo del polimorfismo rs1042522 (Naidoo et al., 2020). De acuerdo con este método, el alelo que corresponde a homocigoto C/C en el ADN en la posición 215 del exón 4 del gen da como resultado que no hay corte del producto de PCR

por lo que se observa la banda original intacta de 131 pb. Por otro lado, la variante homocigótica del alelo homocigoto G/G en el ADN, produce dos fragmentos, uno de 81 y otro de 50 pb. El genotipo heterocigoto que corresponde a C/G en el ADN (mostraban picos dobles en el cromatograma), produce tres bandas una de 131 pb correspondiente al cromosoma con el alelo “C” sin cortar; y dos fragmentos para el cromosoma con el alelo “G” uno de 81 pb y otro de 50 pb (Naidoo et al., 2020).

El programa Sequencher permite realizar análisis de restricción *in silico* donde se seleccionan las endonucleasas específicas, lo cual conduce a la identificación de la posición del SNP en las secuencias. Complementariamente, se utilizó la herramienta NEBcutter. Los análisis de restricción *in silico* permitieron confirmar la identidad (genotipo) y posición de los SNPs en las secuencias de ambos genes en sus correspondientes cromatogramas de todas las secuencias. El genotipo y posición de los SNPs fue luego confirmado en los alineamientos de secuencias de ADN y de proteínas.

Alineamiento de secuencias de nucleótidos y proteínas

Las secuencias nucleotídicas y de proteínas se alinearon con las correspondientes secuencias de referencia encontradas en la base de datos del NCBI-GenBank. Se generaron dos archivos en formato FASTA, uno para las secuencias de nucleótidos y otro para las secuencias de proteínas con las 60 secuencias analizadas para cada gen. Cada archivo FASTA contenía la misma longitud en cantidad de bases o cantidad de aminoácidos. Cada archivo en FASTA fue alineado en el programa *Molecular Evolutionary Genetics Analysis 11* (MEGA 11). Luego de alinearse, se verificó la posición del nucleótido o del aminoácido en la correspondiente posición del SNP de cada muestra analizada.

Determinación de frecuencias alélicas y frecuencias genotípicas

Una vez verificado el genotipo de las 60 muestras analizadas, se determinó la frecuencia alélica y la frecuencia genotípica de cada SNP con el programa PopGene v1.32 (Yeh et al., 1997). Se organizaron los datos en el formato requerido por el programa. Luego, se importó el archivo y se seleccionaron

las barras *Codominant Data* y *Diploid Data* para hacer el análisis, en el cual se calcularon las frecuencias alélicas y genotípicas de los correspondientes SNPs. Para observar su contexto general, diversidad y significancia, las frecuencias alélicas y genotípicas determinadas para ambos SNPs en la población Ngöbe fueron comparadas con las frecuencias alélicas reportadas de ambos polimorfismos en otras poblaciones del mundo.

Resultados

Amplificación de los polimorfismos mediante PCR y secuenciación de ADN

Los SNPs fueron genotipados mediante amplificación por PCR y secuenciación Sanger de la región del exón que contiene cada SNP respectivo. La **Figura 1** muestra la amplificación de la región del gen *MTHFR* que contiene al SNP rs1801133 con una banda de 198 pb. La **Figura 2** muestra la amplificación de la región del gen *TP53* que contiene al SNP rs1042522 con una banda de 131 pb. El tamaño de las bandas amplificadas fue consistente con los tamaños esperados, según el estándar de peso molecular “100bp PCR ladder” por lo que fueron enviadas para secuenciación Sanger a la empresa Psomagen (N.Y.C., EE.UU.) para continuar el proceso de genotipado. Una vez recibidas las secuencias, las mismas fueron procesadas en SEQUENCHER para verificar la calidad de las mismas, y se realizaron varios análisis (**Tabla 1**) para generar secuencias consenso y determinar el genotipo de cada muestra cuyos resultados presentamos a continuación para cada polimorfismo.

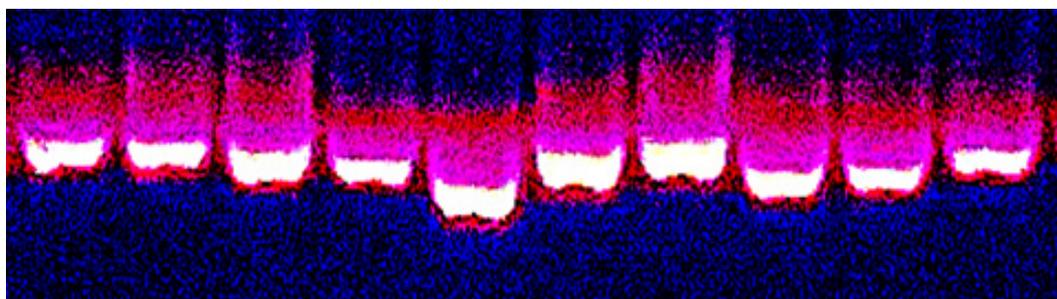


Figura 1. Electroforesis en gel de agarosa de productos de PCR representativos correspondientes a la región del gen *MTHFR* que contiene el polimorfismo rs1801133. Se observa una banda de 198 pb.

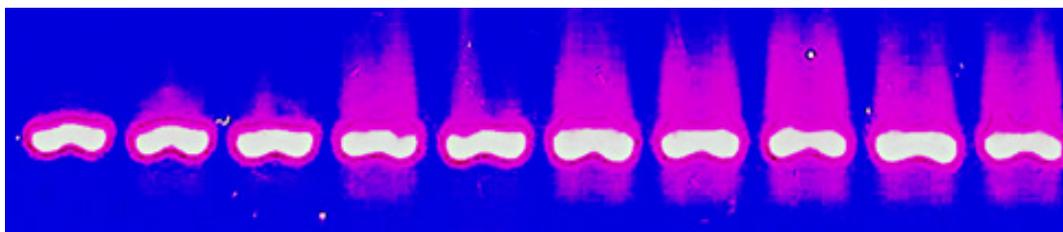


Figura 2. Electroforesis en gel de agarosa de productos de PCR representativos correspondientes a la región del gen TP53 que contiene el polimorfismo rs1042522. Se observa una banda de 131 pb.

Análisis de secuencias y genotipado poblacional de alelos de rs1801133 de MTHFR

La **Figura 3** muestra algunos genotipos representativos identificados en cromatogramas de SEQUENCHER para el SNP rs1801133 del gen MTHFR. Se observan homocigotos TT y heterocigotos CT. Las flechas indican la posición del SNP. No se identificaron homocigotos CC en las 60 muestras analizadas. El genotipo de cada individuo fue anotado en tablas para su posterior análisis de frecuencia alélica y genotípica usando PopGene. La identidad molecular de todas las secuencias fue verificada en BLAST, mostrando un porcentaje de similitud de 99% o superior y un “E-value” menor de $1e-50$. Una vez verificada la identidad en BLAST, las secuencias fueron traducidas a proteínas con la herramienta ORFfinder (Translate tool) de ExPASy. La **Figura 4** muestra los alineamientos de las secuencias de ADN (nucleótidos) mientras que la **Figura 5** muestra los alineamientos de las secuencias de proteínas (aminoácidos). Las flechas indican la posición del SNP. Estos análisis permitieron confirmar el genotipo de cada muestra para el SNP rs1801133. Por consiguiente, procedimos a realizar los análisis poblacionales para determinar las frecuencias alélicas y las frecuencias genotípicas en PopGene. Los análisis poblacionales determinaron que en la población Ngöbe, el polimorfismo rs1801133 de MTHFR muestra una frecuencia alélica de 15% para el alelo C (alanina) y de 85% para el alelo T (valina) (**Figura 6**, izquierda). Las frecuencias genotípicas fueron determinadas en 69% para homocigotos TT y 31% para heterocigotos CT. Como se mencionó previamente, no se identificaron homocigotos CC (0%) (**Figura 6** derecha).

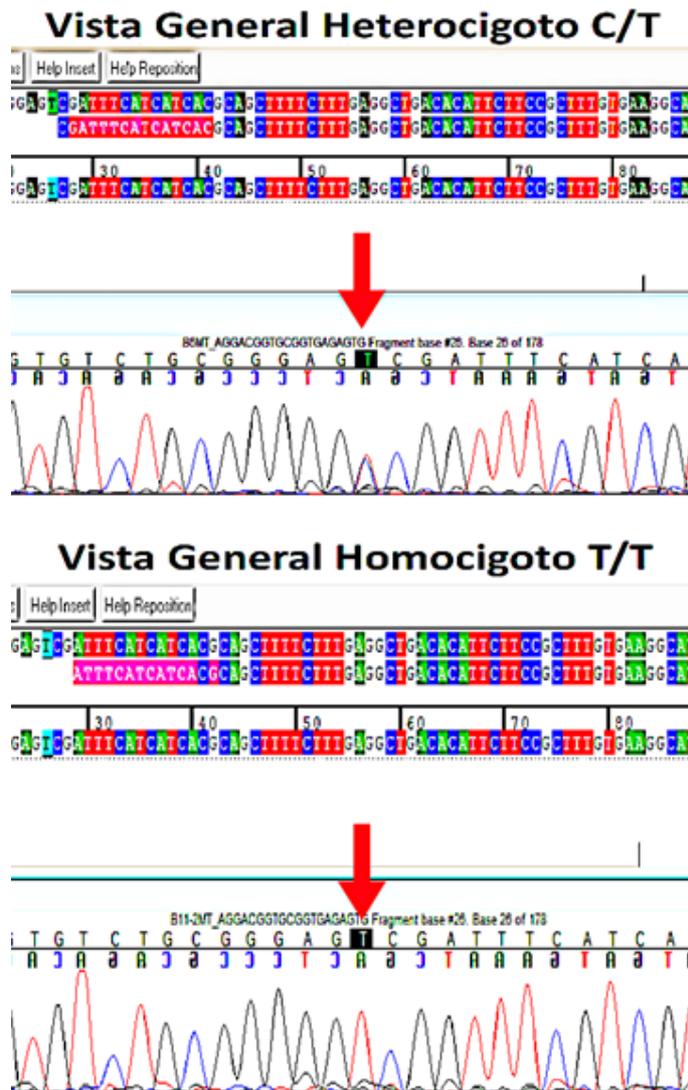


Figura 3. Cromatogramas representativos mostrando la posición del SNP rs1801133 del gen MTHFR en SEQUENCHER. **Arriba**, heterocigoto C/T, se observa picos dobles sobrepuestos (azul para citosina y rojo para timina) en la posición del SNP (flecha). **Abajo**, homocigoto T/T, se observa un solo pico (rojo para timina) en la posición del SNP (flecha). La posición del polimorfismo T es también el sitio de la secuencia de reconocimiento de corte de la endonucleasa HinfI 5'-GANTC-3'. No se identificaron homocigotos C/C en las muestras analizadas, por lo que no se muestran imágenes de ese genotipo.

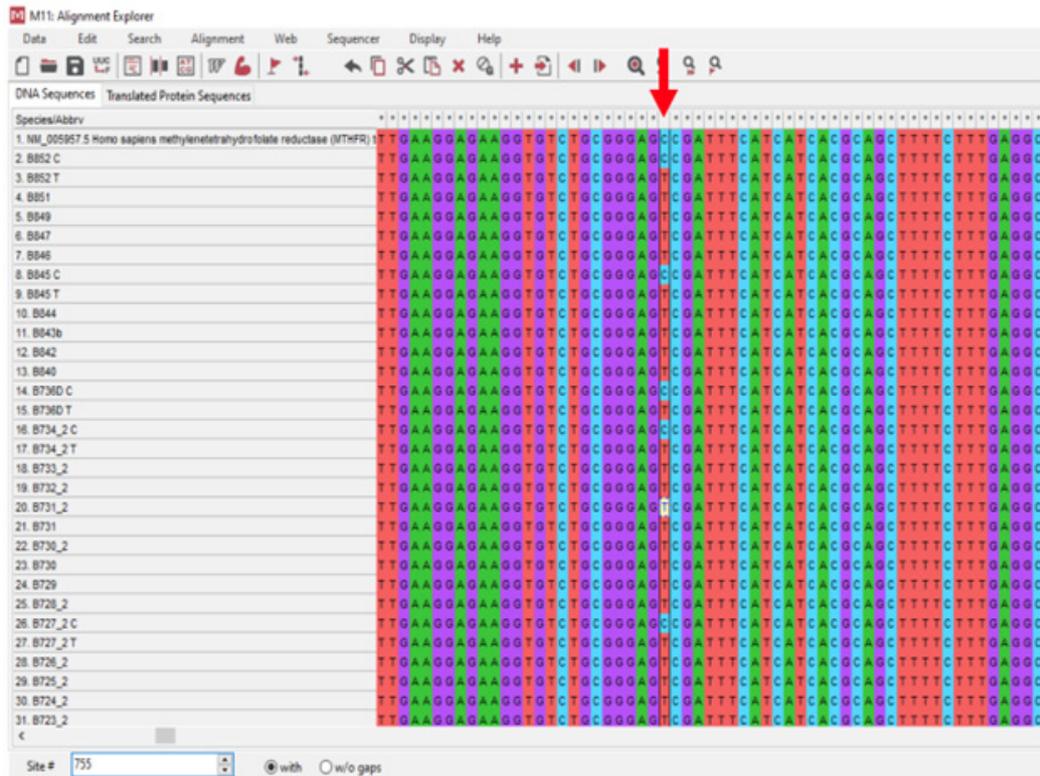


Figura 4. Alineamiento de las secuencias de nucleótidos (ADN). Las secuencias de ADN de las muestras fueron alineadas con la secuencia de referencia nucleotídica de MTHFR del Genbank (arriba, primera secuencia). La flecha indica la posición del SNP rs1801133 con las variantes polimórficas C o T.

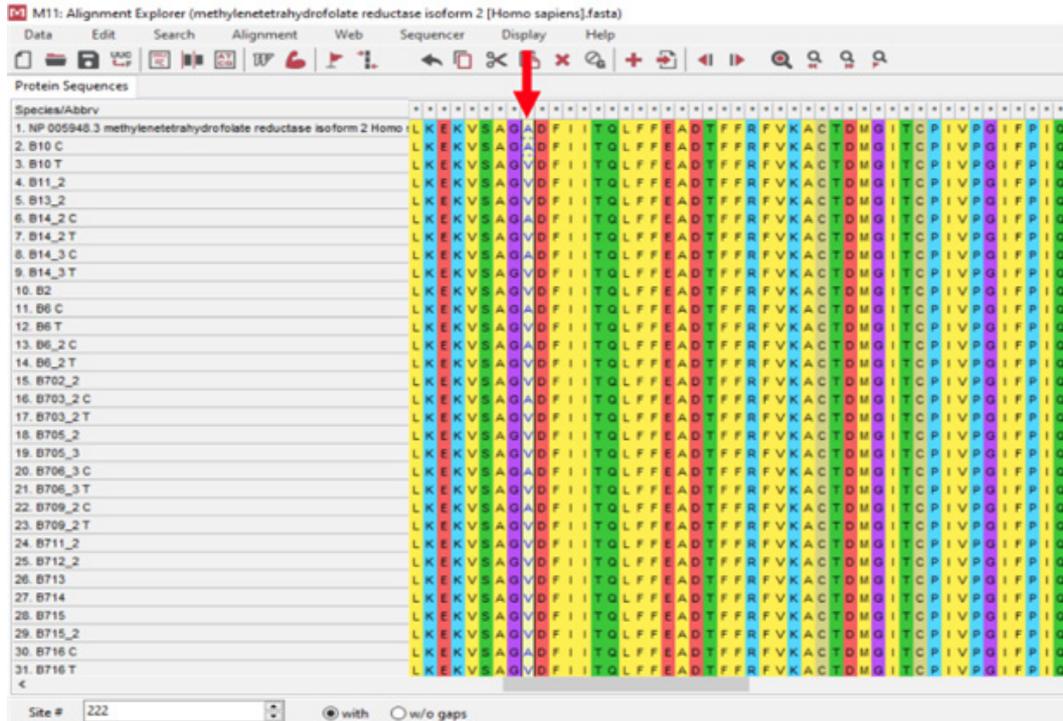


Figura 5. Alineamiento de las secuencias de aminoácidos (proteínas). Las secuencias de nucleótidos fueron traducidas a proteínas con la herramienta ORFfinder (Translate tool) y fueron luego alineadas en MEGA. Se muestra la secuencia de referencia proteínica MTHFR (arriba, primera secuencia). La flecha indica la posición del polimorfismo donde ocurre el cambio de aminoácido de alanina (A) o valina (V).

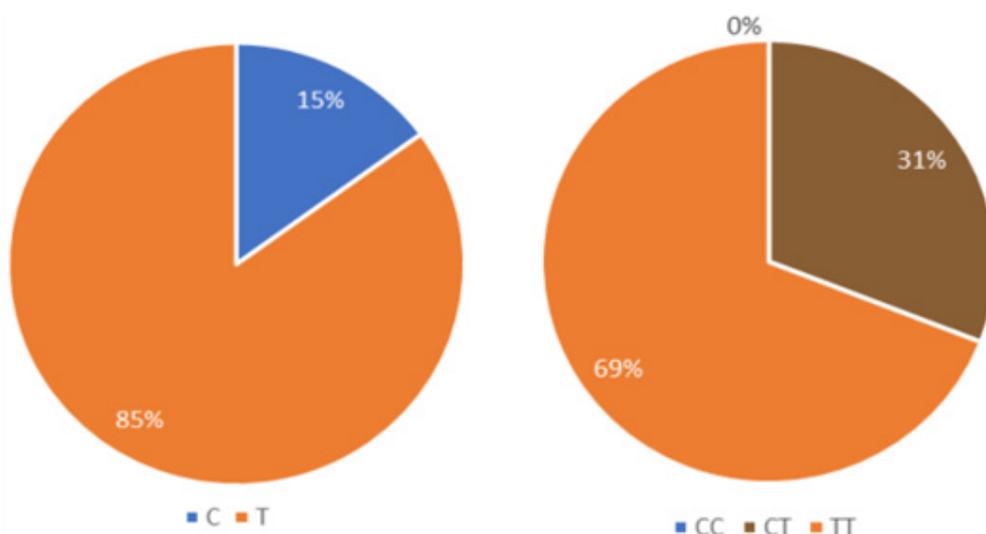


Figura 6. Frecuencias alélicas y genotípicas del SNP rs1801133 de MTHFR en la población Ngöbe. Izquierda: Frecuencia alélica del lelo C (azul, 15%); y T (anaranjado, 85%). Derecha: Frecuencia genotípica para los genotipos identificados CT (heterocigotos, marrón, 31%) y homocigotos TT (anaranjado, 69%). No se observaron genotipos homocigotos CC (0%) en las 60 muestras genotipadas.

Frecuencias alélicas y genotípicas de rs1801133 en otras poblaciones del mundo

Con el objetivo de observar su contexto general, diversidad y significancia, los patrones alélicos para el SNP rs1801133 del gen MTHFR observados en Ngöbes, fueron comparados con los patrones alélicos de otras poblaciones del mundo genéticamente diferentes. En primer lugar, se contrastaron los patrones alélicos de un estudio sobre poblaciones de referencia de Estados Unidos (EE.UU.) de grupos afroamericanos, caucásicos (europeos) e hispánicos (Graydon et al., 2019). En este estudio se reportó que el porcentaje del alelo 677T es más elevado en los hispánicos (42%), más bajo en los afroamericanos (16%) e intermedio en los caucásicos (32%), (Tabla 2). Además, el mismo estudio reportó una frecuencia genotípica para el homocigoto TT más alta para los grupos hispánicos que en caucásicos y afroamericanos. (Tabla 2).

Tabla 2. Frecuencias alélicas y genotípicas del SNP rs1801133 de MTHFR en poblaciones de referencia afroamericana, caucásica e hispánica de Estados Unidos*.

Población	Frecuencia Alélica (%)		Frecuencia Genotípica (%)		
	C	T	677CC	677CT	677TT
Ngöbe	15	85	0	31	69
Afroamericanos	84	16	46.4	23.0	8.1
Caucásicos	68	32	29.8	39.4	29.1
Hispánicos	58	42	23.8	37.6	62.8

*Tabla adaptada de Graydon et al., (2019).

En contraste, otros estudios de metaanálisis de diferentes regiones de la población China (Tabla 3), (Wang et al., 2016) y en otro estudio de meta análisis de poblaciones de México y de países de América Central (Tabla 4), (Reyes et al., 2021), se reportaron frecuencias alélicas muy similares a los reportados en el presente estudio en Ngöbe. Estos estudios reportan una frecuencia alélica más alta para el alelo T y muy alta frecuencia genotípica de homocigoto TT y heterocigotos, pero muy baja frecuencia de homocigotos CC.

Tabla 3. Frecuencias alélicas y genotípicas del SNP rs1801133 de MTHFR en distintos grupos étnicos de China*.

Región/ país	Grupo étnico	Frecuencias Genotípicas (%)			Frecuencias alélicas	
		CC	CT	TT	C	T
Panamá	Ngöbe	0	31	69	0.15	0.85
Henan	Han	18	45	37	0.41	0.59
Shandong	Han	15	46	39	0.38	0.62
Shanxi	Han	19	49	32	0.43	0.57
Hebei	Han	17	49	34	0.42	0.58
Guangxi	Han	14	48	37	0.39	0.61
Henan	Han	16	43	41	0.38	0.62

*Tabla adaptada de Wang et al., (2016).

Tabla 4. Frecuencias alélicas y genotípicas del SNP rs1801133 de MTHFR en poblaciones amerindias de México y Centroamérica*.

Población	País	Frecuencias genotípicas (%)			Frecuencias alélicas	
		CC	CT	TT	C	T
Ngöbe (n=60)	Panamá	0	31	69	0.15	0.85
Bribri (n=110)	Costa Rica	9	49	42	0.34	0.66
Cabecar (n=147)	Costa Rica	3	49	57	0.2297	0.7703
Chinanteco (n=81)	México	5	27	68	0.19	0.81
Trikís (n=89)	México	0	16	84	0.08	0.92
Chorotega (n=133)	Costa Rica	14	56	31	0.41	0.59
Huastecos (n=79)	México	11	40	48	0.32	0.68
Huetar (n=153)	Costa Rica	14	52	35	0.4	0.6
Huicholes (n=50)	México	16	56	28	0.44	0.56
Jakaltecos (n=40)	Guatemala	8	25	68	0.2	0.8
Kaqchibel (n=36)	Guatemala	0	30	69	0.15	0.85
Yucatán Maya (n=54)	México	13	52	35	0.39	0.61
Mazatecos (n=59)	México	7	36	58	0.25	0.75
Mixe (n=89)	México	12	48	39	0.38	0.62
Mixtecas (n=124)	México	10	40	50	0.3	0.7
Nahuas (n=135)	México	5	43	52	0.27	0.73
Otomí (n=220)	México	13	38	49	0.32	0.68

*Tabla adaptada de Reyes et al., (2021).

Análisis de secuencias y genotipado poblacional de alelos de rs1042522 de TP53

La **Figura 8** muestra cromatogramas representativos de SEQUENCHER que indican el genotipo de alelos homocigoto C/C (**izquierda**); heterocigoto G/C (**centro**) y homocigoto G/G (**derecha**) del polimorfismo rs1042522 en TP53. Se observa el sitio de restricción donde la endonucleasa BstUI corta el producto de PCR secuenciado y que también corresponde a la po-

sición del SNP. Se obtuvo una secuencia consenso y se verificó la identidad molecular de todas las secuencias en BLAST. Estas secuencias de ADN fueron luego traducidas a secuencias proteína con la herramienta ORFfinder (*Translate tool*) de ExPASy. La **Figura 9** muestra los alineamientos de las secuencias de ADN (nucleótidos, a la izquierda) y el alineamiento de las secuencias de proteínas (aminoácidos, a la derecha). Esto permitió confirmar el genotipo de los alelos de cada muestra, cuantificarlos y proceder con los análisis poblacionales de las frecuencias alélica y genotípica en PopGene.

Los análisis poblacionales determinaron que el polimorfismo rs1042522 en TP53 mostró una frecuencia de 56% y 44%, para el alelo C (prolina) y (G) (arginina), respectivamente (**Figura 10**, izquierda). Las frecuencias genotípicas demostraron un 64% para heterocigotos (CG), 12% homocigotos GG y 24% de homocigotos CC (**Figura 10**, derecha).

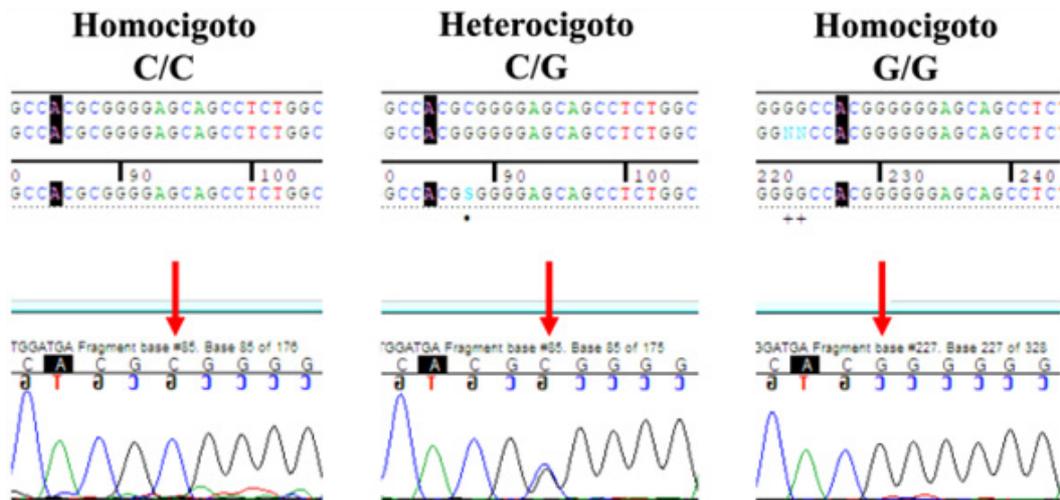


Figura 8. Cromatogramas representativos mostrando la posición del SNP rs1042522 del gen TP53 en SEQUENCHER. Se muestra genotipo homocigoto C/C (izquierda), genotipo heterocigoto C/G (centro); y homocigoto G/G (derecha). El programa SEQUENCHER permite identificar el sitio de corte de la endonucleasa BstUI que corresponde a la secuencia de restricción 5'-CG*CG-3'/ 5'-GC*GC-3' y coincide con la posición de la secuencia del SNP con genotipo C.

Frecuencias alélicas del SNP rs1042522 de TP53 en otras poblaciones del mundo

Con el objetivo de observar su contexto, diversidad y significancia, las frecuencias alélicas y genotípicas del polimorfismo rs1042522 de TP53 determinados en la población Ngöbe fueron comparadas con los patrones alélicos en otras poblacionales del mundo (Tabla 5), (Chiurillo et al., 2010). Se muestra que en poblaciones de Costa Rica, España, Argentina, Colombia y Finlandia el alelo predominante es el G (arginina). Mientras que, en general, el alelo C (prolina) es más frecuente en poblaciones como Japón (Asia) y Uganda (África).

Tabla 5. Frecuencia genotípica y alélica del SNP rs1042522 de TP53 en poblaciones del mundo*.

Población	Frecuencia Genotípica (%)			Muestra	Frecuencia Alélica (%)	
	Pro/Pro	Pro/Arg	Arg/Arg		Pro	Arg
Ngöbe (Panamá)	24	64	12	60	56	44
Costa Rica	10	35	55	175	-	-
Argentina	11.1	48.6	40.3	109	35.3	64.7
Colombia	10.2	45.7	44.1	186	33	67
España	10.2	43.5	46.3	90	32	68
Finlandia	5.3	38	56.7	171	24.3	75.7
Japón	18	46	36	110	59.5	40.5
Uganda	47	47.8	5.2	115	70.9	29.1
Turquía	8.7	45.2	46.1	115	68.7	31.3
Túnez	7	54	39	83	34.3	68.7

*Tabla adaptada de Chiurillo et al., (2010).

Discusión

Múltiples trabajos e información disponible en bases de datos genómicas, como dbSNP, COSMIC, OMIM, GWAS Catalog y HGMD, indican que los polimorfismos analizados han sido previamente asociados como factores de riesgo genético de cánceres, enfermedades vasculares y cierto número de

otras enfermedades. Sin embargo, la información de estas bases de datos es revisada y actualizada cada cierto tiempo por paneles de expertos a medida que se van reportando nuevas investigaciones sobre los polimorfismos. Estas actualizaciones generan nuevas evidencias que confirmarían o no las asociaciones de riesgo de las variantes génicas con las correspondientes enfermedades. Algunas veces los reportes generan asociaciones conflictivas o sin confirmar (por ejemplo, “uncertain significance”), las cuales asignan diferentes valores o puntajes (*scores*) a los polimorfismos, lo cual puede diferir entre diferentes poblaciones. Por lo tanto, es necesario contrastar los datos obtenidos en este reporte con las actualizaciones más recientes entre las diferentes bases de datos, las cuales pueden variar un poco en su información, dependiendo de la rapidez con la que se actualizan. Además, es importante revisar y discutir los hallazgos generados con las publicaciones y metaanálisis (combinación estadística de los resultados de múltiples estudios) más recientes que aún no aparecen actualizadas en las bases de datos.

Polimorfismo rs1801133 de MTHFR, un SNP asociado a cánceres, enfermedades vasculares y otras identificado en la población Ngöbe

El SNP rs1801133, también conocido como C677T, Ala222Val y A222V, recopila una variante del gen MTHFR, que codifica una enzima implicada en el metabolismo del folato. Según SNPedia, los individuos homocigotos rs1801133(T;T) tienen ~30% de la actividad enzimática MTHFR esperada, y los heterocigotos rs1801133(C;T) tienen ~65% de actividad, en comparación con el genotipo más común, rs1801133(C;C). Hasta la fecha, las bases de datos como CLNSIG, ClinVar, GWASCatalog, MutationAssessor y CADD, clasifican esta variante con significancia incierta, interpretaciones conflictivas, categoría media o probablemente benigna para algunas enfermedades. Sin embargo, la base de datos COSMIC reporta que este SNP está asociado a varios tipos de cáncer incluyendo mama, gástrico, leucemia mieloide aguda, pulmón, próstata y carcinoma urotelial de vejiga (Mutation Overview Page MTHFR_ENST00000376590 - p.A222V (Substitution-Missense)). Además, según la base de datos OMIM, rs1801133 está asociado a la homocistinuria debido a la deficiencia de MTHFR, a la susceptibilidad a los defectos del tubo neural, a la susceptibilidad del tromboembolismo y a la susceptibilidad a enfermedades vasculares. Se-

gún la base de datos HGMD, este SNP tiene asociación con enfermedades cardiovasculares (HGMD® home page, s. f.). Adicionalmente, SIFT clasifica a esta variante como deletérea (score=0.002), probablemente patogénica para Polyphen2_HDIV (0.998), probablemente patogénica para Polyphen2_HVAR (0.941), deletérea para LRT (0.000) y deletéreo para FATHMM (-4.03).

Reportes más recientes respaldan que rs1801133 está asociado con varios tipos de cáncer, incluyendo gástrico en poblaciones asiáticas (Han et al., 2021), Neuroblastoma (Olivera et al., 2020); leucemia linfoblástica (Chung-Filho et al., 2020) y mama (Omran et al., 2021), entre otros. El polimorfismo ha sido asociado al Parkinson en poblaciones mestizas e indígenas de México (Romero-Gutiérrez et al., 2021) y a enfermedades inmunes como Lupus (Zhou et al., 2020) y artritis (Nakano et al., 2021; Poornima et al., 2022; Arida et al., 2022). El SNP ha sido asociado a defectos congénicos incluyendo labio endido (Li et al., 2020), vasculares (Yuan et al., 2020) y del corazón (Sun et al., 2021). Además, rs1801133 ha sido relacionado con hipertensión (Fan et al., 2022; Mabhida et al., 2022ab; Kong et al., 2022; Wu et al., 2022; Fan et al., 2022; Chiu et al., 2023), con enfermedades cardiovasculares (Raina et al., 2020; Bouzidi et al., 2020; Shivkar et al., 2022), cerebrovasculares (Ames et al., 2022) y trombosis (Ames et al., 2023). Este polimorfismo ha sido involucrado con problemas metabólicos, incluyendo obesidad (Damavandi et al., 2022), diabetes, y altos niveles de colesterol y triglicéridos (Meng et al., 2019; Pathak et al., 2022; Asadi et al., 2024; Lapik et al., 2021).

Las frecuencias alélicas de rs1801133 de MTHFR en poblaciones del mundo sugieren que el polimorfismo podría constituir un factor de riesgo moderado para algunos trastornos. Desde el punto de vista de la población mundial, esto podría representar una carga considerable, ya que la variante es relativamente común (Leclerc et al., 2013). En Ngöbes no se detectaron homocigotos CC, pero sí un alto porcentaje de homocigotos TT y de heterocigotos. Además, la frecuencia del alelo C es de apenas 0.1545, mientras que la frecuencia del alelo T es remarcablemente alta con 0.8455. Esto contrasta con lo que indican las bases de datos en dbSNP, puesto que, a nivel mundial, el alelo C tiende a ser mayor que el alelo T. Por ejemplo, en Allele Frequency aggregator (ALFA, total/global): C=0.659842, T=0.340158; 1000Genomes: C=0.7546, T=0.2454; TopMed: C=0.708757, T=0.291243 (rs1801133 RefSNP Report - dbSNP - NCBI, 2022). Los datos demuestran que, en general, las poblaciones africanas poseen la menor frecuencia del alelo T: 0.090 (1000 Genome Projects Phase 3), 0.108 (gno-

mAD), 0.12 (ALFA). Los asiáticos exhiben una frecuencia mayor que los africanos en T 0.339 (ALFA). Los europeos muestran la segunda frecuencia más elevada en T 0.296 (1000 Genomes), 0.315 (gnomAD), 0.349 (ALFA). Los latinoamericanos muestran la frecuencia más elevada en T con 0.474 (1000 Genomes) y 0.503 ([gnomAD] [rs1801133 {SNP}] - Population genetics - *Homo sapiens* - Ensembl genome browser 109, s. f.]. Por ejemplo, el alelo T es mayor en frecuencia que el C en varias poblaciones latinas, lo cual se acerca a lo observado en la población Ngöbe (Graydon et al., 2019). Por otro lado, estudios en poblaciones de China demostraron que el alelo T supera al alelo C en frecuencia (Wang et al., 2016). Similarmente, se ha reportado que en grupos indígenas de México y de países de América Central, como Guatemala y Costa Rica, el alelo T es mucho más alto que el C (Reyes et al., 2021), lo cual es consistente con la población Ngöbe. La alta frecuencia del alelo T en Ngöbes, en otras poblaciones indígenas, en poblaciones mestizas latinoamericanas, así como en poblaciones asiáticas, sugiere que el alelo T podría tener mayor afinidad y posible origen en poblaciones indígenas, así como un origen ancestral aún más antiguo en poblaciones paleoamerindias asiáticas que poblaron el continente que derivaron posteriormente en los indígenas actuales y permearon a las poblaciones mestizas modernas. Adicionalmente, la altísima frecuencia del alelo T en las poblaciones indígenas de México, Centroamérica y en Ngöbes podría estar relacionado con los fenómenos de deriva génica, como los de cuello de botella reportados previamente en indígenas debido a la conquista europea y enfermedades que en esa época redujeron a la debacle poblacional que redujo significativamente su diversidad genética (Castro-Pérez et al., 2016; Castro et al., 2007). Dado que es la primera vez que este polimorfismo es identificado en el país, estudios en curso tienen el objetivo de confirmar la posible asociación patogénica de este SNP en la población panameña genotipando grupos de pacientes diagnosticados o grupos de alto riesgo, y compararlos con grupos control o de bajo riesgo en diferentes enfermedades.

Polimorfismo rs1042522 de TP53, un SNP asociado a cánceres identificado en la población Ngöbe

Este SNP se presenta con el alelo G menos común que codifica una arginina en el codón (posición) 72 de la proteína (donde normalmente hay una C que codifica prolina). El SNP se llama comúnmente variante Arg72, aun-

que P72R y Arg72Pro también son comunes en la literatura. El polimorfismo ha sido asociado principalmente con cánceres, pero también algunas otras enfermedades. La significancia clínica del polimorfismo rs1042522 de TP53 ha sido revisada en varias bases de datos y herramientas disponibles. Por ejemplo, en VarSome (Kopanov et al., 2019) se combinan varios predictores con información de diferentes bases de datos lo cual genera varios tipos de asociaciones del polimorfismo a cánceres con diferentes puntajes (scores). Por ejemplo, para MetaRNN el SNP es moderadamente benigno (Benign Moderate, 0.1661), pero para MetaLR (0.8875) y MetaSVM (0.8579) muestra respaldo como patogénico (Pathogenic Supporting), mientras que para FATHMM es moderadamente patogénico (pathogenic moderate, puntajes alrededor de -5.24, -5.75, -2.05). Sin embargo, para BayesDeladdAF (0.08668) y MutationTaster (0.9999) su estado es *Uncertain*. Adicionalmente, para ClinVar la significancia clínica del SNP es considerada incierta (*uncertain-significance*) o benigno (benigno) para síndrome de Li-Fraumeni (cánceres hereditarios de mama y de ovario), pero es patogénico para carcinoma oral, según las últimas evaluaciones de 2019 y 2021. Sin embargo, publicaciones más recientes sugieren que el polimorfismo está asociado a mayor riesgo en varios tipos de cáncer incluyendo cáncer cervical (Iordanishvili et al., 2023), cáncer de próstata (Duncan et al., 2023), así como algunas leucemias, incluyendo leucemia linfoblástica aguda de células B (B-ALL), leucemia linfoblástica aguda de células T (T-ALL) (Usenova et al., 2023; Ounalli et al., 2023) y leucemia linfocítica crónica (Basabaeen et al., 2019). Además, publicaciones recientes asocian este SNP con mayor riesgo para cáncer de mama (Cardoso et al., 2023), cáncer de pulmón de células no pequeñas (NSCLC) (Karaağaç et al., 2023) y melanoma (Ozola et al., 2019). Por otro lado, el polimorfismo fue asociado también con problemas congénicos como aneuploidías asociadas a la edad de la madre (Chan et al., 2023) y mayor riesgo de enfermedad arterial coronaria (Omrani-Nava et al., 2018; Khan, et al., 2016; Vidal et al., 2019; Caamaño et al., 2009), artritis reumatoide (Lee, et al., 2012), obesidad (Kung et al., 2016) y diabetes entre otras (Bonfigli et al., 2013). Con el objetivo de confirmar la posible asociación patogénica de este SNP en la población panameña estudios en curso analizan este y otros SNPs en grupos de pacientes diagnosticados o de alto riesgo de estas enfermedades y se comparan con grupos control o de bajo riesgo de padecer las mismas.

Las frecuencias alélicas obtenidas en este estudio nos permiten por primera vez demostrar que el alelo de mayor proporción en la población amerindia Ngöbe es el alelo mutante (G) que codifica arginina con 44%, la cual es una frecuencia bastante alta. En cuando a las frecuencias alé-

licas de rs1042522 en TP53 en otras poblaciones del mundo, la base de datos NIH-dbSNP muestra que a nivel global la frecuencia más alta es en el alelo silvestre con $C=0.714337$ (ALFA, total/global), pero varía según diferentes estudios en $C=0.5429$ (1000Genomes, global); $C=0.6174$ (TOPMED); $C=0.6265$ (GnomAD), lo cual podría estar más relacionado con sobrerrepresentación de grupos europeos (90388 europeos de un total de 105516), los cuales muestran una alta frecuencia con $C=0.73664$. Esta base de datos indica que, en grupos africanos, el alelo predominante en frecuencia es el G ($G=0.6854$). Por otro lado, el alelo de mayor frecuencia en grupos asiáticos es el silvestre (patogénico) con $C=0.6226$ (East Asian), Japón $C=0.6502$ y Korea $C=0.6554$. Esta misma base de datos indica que en grupos latinoamericanos, el alelo de mayor proporción es el silvestre con reportes de frecuencias alélicas de $C=0.579$ - $C=0.659$. Esto sugiere que la frecuencia de $C=0.6200$ en Ngöbes guarda más similitud con las frecuencias de poblaciones asiáticas como lo respaldan múltiples reportes genéticos sobre el origen asiático de los pueblos amerindios (Castro et al., 2007; Arnaiz-Villena et al., 2010).

Los resultados muestran que las frecuencias genotípicas de las variantes del rs1042522 se distribuyen en 20% población heterocigotos, 28% homocigotos mutantes y un 52% homocigotos de tipo silvestre. Estudios previos en Irán y Brasil han demostrado que el alelo mutante que codifica arginina prevalece más en personas de piel clara que viven en latitudes altas, mientras que el alelo silvestre que codifica prolina está más asociado a poblaciones de piel oscura que viven cerca al ecuador (Thurrow et al., 2011). Aunque se necesitan más estudios para confirmar si esto es así en Panamá, estos reportes podrían ser una guía para explicar los patrones de incidencia y distribución de algunos cánceres, así como de las muertes por esta enfermedad. Por ejemplo, en un reporte previo se identificó la distribución polimórfica en incidencia y muertes de algunas enfermedades como el cáncer entre las provincias del país (Castro-Pérez et al. 2016). Este reporte sugiere, además, que la distribución en incidencia y muertes por estas enfermedades podría estar asociada con la proporción de genes ancestrales que contribuyen con factores de riesgo genético. En particular, el cáncer de próstata, en el cual se observó que el mismo mostraba mayor incidencia y muertes en provincias con mayor proporción de genes africanos (Castro-Pérez et al., en 2016). Sin embargo, aunque algunas poblaciones del mundo muestran alta frecuencia del alelo mutante, el efecto de interacciones del mismo con el ambiente, estilos de vida, así como con otros genes determinarían el efecto real en riesgo a cáncer y otras enfermedades, como se discutió previamente en la población panameña

(Castro-Pérez et al., 2016). Más estudios confirmarán estas posibles asociaciones en la población panameña, donde se identificará la frecuencia de estos SNPs en pacientes o grupos de la población de alto riesgo versus grupos controles no enfermos.

Conclusiones

Los análisis genómicos nos condujeron por primera vez a la identificación de dos polimorfismos de importancia biomédica que afectan a la población Ngöbe rs1801133 en el gen MTHFR y rs1042522 en el gen TP53, y determinamos la frecuencia de ambos en una muestra de 60 individuos no relacionados de esta población. La frecuencia alélica de la variante silvestre (C) de rs1042522 resulta ser de 0.15 y la del alelo mutante (T) de 0.85, así como una frecuencia genotípica de 0% para los genotipos homocigoto CC, 31% heterocigoto CT y 69% para el homocigoto TT. Las frecuencias identificadas son consistentes con algunos reportes en otras poblaciones. Por otro lado, el polimorfismo rs1042522 (TP53) mostró una frecuencia de 0.56 y 0.44, para los alelos silvestre “C” y mutante “G”, respectivamente; y una frecuencia genotípica de 0.24 para los genotipos homocigoto CC, 0.64 heterocigoto CG, y 0.12 para el homocigoto GG.

Estos datos sugieren que, dada la alta frecuencia de los alelos mutantes, presumiblemente, la población mestiza del país derivada de esta población indígena ancestral podría presentar alta frecuencia de este alelo. De confirmarse esto, el polimorfismo podría contribuir, junto con otros factores genéticos y ambientales, a la incidencia y muertes por cánceres y otras enfermedades asociadas al mismo tanto en Ngöbes como en mestizos. Para confirmar si hay asociación o no de estos SNPs con estas enfermedades, estudios en curso analizan los patrones genéticos de los mismos en grupos de pacientes diagnosticados o de alto riesgo y se comparan con grupos control o de bajo riesgo en la población mestiza del país.

Literatura citada

1. Ahmed, S., Safwat, G., Moneer, M.M. et al. Prevalence of TP53 gene Pro72Arg (rs1042522) single nucleotide polymorphism among Egyptian breast cancer patients. *Egypt J Med Hum Genet* 24, 24 (2023).
2. Akulevich NM, Saenko VA, Rogounovitch TI, et al. Polymorphisms of DNA damage response genes in radiation-related and sporadic papillary thyroid carcinoma. *Endocr Relat Cancer*. 2009;16(2):491-503.
3. Ames PRJ, D'Andrea G, Arcaro A, Marottoli V, Iannaccone L, Maraglione M, Gentile F. Liver Cirrhosis Patients Homozygous for MTHFR C677T Develop Portal Vein Thrombosis 8 Years Earlier Than Wild Type. *Dig Dis Sci*. 2023 Apr;68(4):1332-1338.
4. Ames PRJ, D'Andrea G, Marottoli V, Arcaro A, Iannaccone L, Gentile F, Maraglione M. Juvenile patients with the homozygous MTHFR C677T genotype develop ischemic stroke 5 years earlier than wild type. *J Thromb Thrombolysis*. 2022 Aug;54(2):330-338.
5. Arias, T. D., Jorge, L. F., Griese, E.-U., Inaba, T., & Eichelbaum, M. (1993). Polymorphic N-acetyltransferase (NAT2) in Amerindian populations of Panama and Colombia: high frequencies of point mutation 857A, as found in allele S3/M3. *Pharmacogenetics*, 2(6), 328-331.
6. Arida A, Nezos A, Papadaki I, Sfikakis PP, Mavragani CP. Osteoprotegerin and MTHFR gene variations in rheumatoid arthritis: association with disease susceptibility and markers of subclinical atherosclerosis. *Sci Rep*. 2022 Jun 9;12(1):9534.
7. Arnaiz-Villena A, Parga-Lozano C, Moreno E, Areces C, Rey D, Gomez-Prieto P. The Origin of Amerindians and the Peopling of the Americas According to HLA Genes: Admixture with Asian and Pacific People. *Curr Genomics*. 2010 Apr;11(2):103-14.
8. Asadi S, Rahimi Z, Kohsari M, Babajani F, Amiri M, Jalilian N, Naseri R, Haghazari L. Assessing the possible association between MTHFR (rs1801133) and GPx-1 (rs1050450) polymorphisms with the risk of type 2 diabetes, diabetic neuropathy, and diabetic retinopathy. *Mol Biol Rep*. 2024 Apr 29;51(1):583.
9. Bargonetti, J., Manfredi, J., (2002). Multiple roles of the tumor suppressor p53. *Curr Opin Oncol*. 14 86-91.
10. Bock, C. H., Schwartz, A. G., Ruterbusch, J. J., Levin, A. M., Neslund-Dudas, C., Land, S. J., Wenzlaff, A. S., Reich, D., McKeigue, P., Chen, W., Heath, E. I., Powell, I. J., Kittles, R. A., & Rybicki, B. A. (2009). Results from a prostate cancer admixture mapping study in African-American men. *Human Genetics*, 126(5), 637.
11. Boletín de la Organización Panamericana de la Salud (OPS). 2022.
12. Boletín Estadístico, Anuario 2018, Ministerio de Salud, Dirección de Planificación, Departamento de Registros y Estadísticas de Salud. Panamá y Contraloría General de la República, Dirección de Estadísticas y Censo.

13. Bonfigli, A. R., Sirolla, C., Testa, R., Cucchi, M., Spazzafumo, L., Salvioli, S., Franceschi, C. (2013). The p53 codon 72 (Arg72Pro) polymorphism is associated with the degree of insulin resistance in type 2 diabetic subjects: A cross-sectional study. *Acta Diabetologica*, 50, 429–436.
14. Bourdon JC, Fernandes K, Murray-Zmijewski F, Liu G, Diot A, Xirodimas DP, Saville MK, Lane DP. p53 isoforms can regulate p53 transcriptional activity. *Genes Dev.* 2005 Sep 15;19(18):2122–37.
15. Bouzidi, N., Hassine, M., Fodha, H. et al. Association of the methylene-tetrahydrofolate reductase gene rs1801133 C677T variant with serum homocysteine levels, and the severity of coronary artery disease. *Sci Rep* 10, 10064 (2020).
16. Brady, C. A., & Attardi, L. D. (2010). p53 at a glance. *Journal of cell science*, 123(Pt 15), 2527–2532.
17. Caamaño, L., Saavedra, N., Jaramillo, P., Lanás, C., Lanás, F., (2009). Polimorfismo Pro72Arg del gen TP53 se asocia a enfermedad coronaria en individuos chilenos. *Revista chilena de cardiología*, 28(2), 151–157.
18. Cardoso YMN, Resque RL, Yoshio D, B Miranda AR, Secco M, Do N Rodrigues AS. Association between the Arg72Pro Genotypes in the TP53 Gene and Ile655Val in the HER2 Gene and the Risk of Developing Breast Cancer in the Population of Amapá, Northern Brazil. *Asian Pac J Cancer Prev.* 2023;24(1):157–162.
19. Castro-Pérez, E., Trejos, D. E., Hrbek, T., Setaluri, V., & Ramos, C. W. (2016). Genetic Ancestry of the Panamanian Population: Chibchan Amerindian Genes; and Biological Perspectives on Diseases. *The Internet Journal of Biological Anthropology*, 9(1).
20. Castro, E., Trejos, D., Berovides-Alvarez, V., Arias, T., & Ramos, C. (2007). Genetic Polymorphism and Forensic Parameters of Nine Short Tandem Repeat Loci in Ngöbe and Emberá. *Human Biology*, 79(5), 563–577.
21. Chan Y, Xu W, Feng Y, et al. Association of TP53 rs1042522 G>C, MDM2 rs2279744 T>G, and miR-34b/c rs4938723 T>C polymorphisms with aneuploidy pregnancy susceptibility. *BMC Pregnancy Childbirth.* 2023;23(1):624.
22. Chiu MH, Chang CH, Tantoh DM, Hsu TW, Hsiao CH, Zhong JH, Liaw YP. Susceptibility to hypertension based on MTHFR rs1801133 single nucleotide polymorphism and MTHFR promoter methylation. *Front Cardiovasc Med.* 2023 Oct 2;10:1159764.
23. Chiurillo, M., Morán, Y., Cañas, M., Valero G., (2010). Distribución del Polimorfismo del Codón 72 de TP53 en Venezuela: Implicaciones Étnicas y Geográficas. *Interciencia*. 35. 202 206.
24. Chumakov, P., (2000) Function of the p53 gene: choice between life and death *Biochim* 65: 28–40.
25. Chung-Filho AA, Brisson GD, Vieira TMF, Chagas-Neto P, Soares-Lima SC, Pombo-de Oliveira MS; on behalf the Brazilian Collaborative Study Group of Infant Acute Leukaemia (BCSGIAL). MTHFR rs1801133 polymorphism is associated with increased risk of B-cell precursor lymphoblastic leukaemia with recurrent genetic aberrations of fetal origin. *Cancer Epidemiol.* 2020 Apr;65:101693.

26. Cotter MP, Gern RW, Ho GYF, Chang RY, & Burk RD (2002) Role of family history and ethnicity on the mode and age of prostate cancer presentation. *The Prostate* 50:216-221.
27. Damavandi N, Soleymaniniya A, Bahrami Zadegan S, Samiee Aref MH, Zeinali S. Development of a genetic risk score for obesity predisposition evaluation. *Mol Genet Genomics*. 2022 Nov;297(6):1495-1503.
28. de Andrade KC, Lee EE, Tookmanian EM, Kesserwan CA, Manfredi JJ, Hatton JN, Loukissas JK, Zavadil J, Zhou L, Olivier M, Frone MN, Shahzada O, Longabaugh WJR, Kratz CP, Malkin D, Hainaut P, Savage SA. The TP53 Database: transition from the International Agency for Research on Cancer to the US National Cancer Institute. *Cell Death Differ*. 2022 May;29(5):1071-1073.
29. den Dunnen, J. T., Dalgleish, R., Maglott, D. R., Hart, R. K., Greenblatt, M. S., McGowan Jordan, J., Roux, A. F., Smith, T., Antonarakis, S. E., & Taschner, P. E. M. (2016). HGVS Recommendations for the Description of Sequence Variants: 2016 Update. *Human Mutation*, 37(6), 564-569.
30. Duncan A, Noursome D, Ricks R, et al. Association of TP53 Single Nucleotide Polymorphisms with Prostate Cancer in a Racially Diverse Cohort of Men. *Biomedicine*. 2023;11(5):1404.
31. Fan S, Hao ZY, Zhang M, Liang CZ. Association between the rs1042522 polymorphism in TP53 and prostate cancer risk: An updated meta-analysis. *Chronic Dis Transl Med*. 2017 May 25;3(2):95-104.
32. Fan Y, Wu L, Zhuang W. Methylenetetrahydrofolate Reductase Gene rs1801133 and rs1801131 Polymorphisms and Essential Hypertension Risk: A Comprehensive Analysis. *Cardiovasc Ther*. 2022 Feb 22;2022:2144443.
33. Farrell J, Petrovics G, McLeod DG, & Srivastava S (2013) Genetic and molecular differences in prostate carcinogenesis between African American and Caucasian American men. *Int J Mol Sci* 14:15510-15531.
34. Froese, D. S., Kopec, J., Rembeza, E., Bezerra, G. A., Oberholzer, A. E., Suormala, T., Lutz, S., Chalk, R., Borkowska, O., Baumgartner, M. R., & Yue, W. W. (2018). Structural basis for the regulation of human 5,10-methylenetetrahydrofolate reductase by phosphorylation and S-adenosylmethionine inhibition. *Nature Communications*, 9(1).
35. Frosst, P., Blom, H. J., Milos, R., Goyette, P., Sheppard, C. A., Matthews, R. G., Boers, G. J. H., den Heijer, M., Kluijtmans, L. A. J., van den Heuvel, L. P., & Rozen, R. (1995). A candidate genetic risk factor for vascular disease: a common mutation in methylenetetrahydrofolate reductase. *Nature Genetics*, 10(1), 111-113.
36. Gaughan, D. J., Barbaux, S., Kluijtmans, L. A. J., & Whitehead, A. S. (2000). The human and mouse methylenetetrahydrofolate reductase (MTHFR) genes: Genomic organization, mRNA structure and linkage to the CLCN6 gene. *Gene*, 257(2), 279-289.
37. Giannakis M, Ng K. A common cancer at an uncommon age. *Science*. 2023;379(6637):1088-1090.

38. Goyette P, Frosst P, Rosenblatt DS, Rozen R. Seven novel mutations in the methylenetetrahydrofolate reductase gene and genotype/phenotype correlations in severe methylenetetrahydrofolate reductase deficiency. *Am J Hum Genet.* 1995;56(5):1052-1059.
39. Goyette, P., Sumner, J. S., Milos, R., Duncan, A. M. V., Rosenblatt, D. S., Matthews, R. G., & Rozen, R. (1994). Human methylenetetrahydrofolate reductase: isolation of cDNA, mapping and mutation identification. *Nature Genetics*, 7(2), 195–200.
40. Graydon, J. S., Claudio, K., Baker, S., Kocherla, M., Ferreira, M., Roche-Lima, A., Rodríguez-Maldonado, J., Duconge, J., & Rúaño, G. (2019). Ethnogeographic prevalence and implications of the 677C>T and 1298A>C MTHFR polymorphisms in US primary care populations. *Biomarkers in Medicine*, 13(8), 649–661.
41. Guerrero, S., López-Cortés, A., Indacochea, A. et al. Analysis of Racial/Ethnic Representation in Select Basic and Applied Cancer Research Studies. *Sci Rep* 8, 13978 (2018).
42. Han Z, Sheng H, Gao Q, Fan Y, Xie X. Associations of the MTHFR rs1801133 polymorphism with gastric cancer risk in the Chinese Han population. *Biomed Rep.* 2021 Jan;14(1):14.
43. Hoffman RM, Gilliland FD, Eley JW, Harlan LC, Stephenson RA, Stanford JL, et al. (2001) Racial and ethnic differences in advanced-stage prostate cancer: the prostate cancer outcomes study. *Journal of the National Cancer Institute* 93:388-395.
44. Huang, C., Su, C., Chu, J., Huang, S., Pu, Y., et al. (2011). The polymorphisms of P53 codon 72 and MDM2 SNP309 and renal cell carcinoma risk in a low arsenic exposure area. *Toxicology and Applied Pharmacology.* 257 (3): 349–55.
45. Ingaramo, M., Sánchez, J., Dekanty, A., (2018). Regulation and function of p53: A perspective from Drosophila studies. *Mechanisms of development*, 154, 82–90.
46. Instituto Conmemorativo Gorgas de Estudios de la Salud (2015). Sistema de Información Geográfico de Enfermedades Cardiovasculares y Diabetes en la República de Panamá. Años 2001–2013. Instituto Conmemorativo Gorgas.
47. Instituto Nacional de Estadística y Censo (2018). Defunciones y tasa de mortalidad de las principales causas de muerte en la república, por sexo, según causa.
48. Iordanishvili S, Metreveli T, Lipartia E, et al. The HPV-TP53-MALAT1 Axis: Unraveling interactions in cervical cancer development. *PLoS One.* 2023;18(10):e0291725.
49. Johnson N, Fletcher O, Palles C, et al. Counting potentially functional variants in BRCA1, BRCA2 and ATM predicts breast cancer susceptibility. *Hum Mol Genet.* 2007;16(9):1051-1057.
50. Jorge-Nebert LF, Eichelbaum M, Griese EU, Inaba T, Arias TD. Analysis of six SNPs of NAT2 in Ngawbe and Embera Amerindians of Panama and determination of the Embera acetylation phenotype using caffeine. *Pharmacogenetics.* 2002 Jan;12(1):39-48.
51. Jorge, V., Haughton, M., (2018). Las enfermedades no transmisibles y sus factores de riesgo, abordaje en Panamá desde la perspectiva de salud universal. *Hacia la Salud Universal en Panamá: Documento de la OPS y OMS.*

52. Karaağaç M, Geredeli Ç, Yıldırım MS, et al. The XRCC1 and TP53 gene polymorphisms are associated with advanced-stage disease and early distant metastasis at diagnosis in non-small cell lung cancer. *J Cancer Res Ther.* 2023;19(5):1248-1254.
53. Khalil, R., Al-Awaida, W. J., Al-Ameer, H. J., Jarrar, Y., Imraish, A., al Bawareed, O., Qawadri, R., al Madhoun, F., & Obeidat, L. (2021). Investigation of ACE rs4646994, MTHFR rs1801133 and VDR rs2228570 Genotypes in Jordanian Patients with Fibromyalgia Syndrome. *Endocrine, metabolic & immune disorders drug targets*, 21(10), 1920–1928.
54. Khan, S., Phulukdaree, A., Ramkaran, P., Moodley, D., & Chuturgoon, A. A. (2016). The Arg72 variant of the p53 functional polymorphism (rs1042522) is associated with coronary artery disease in young South Africans of Indian ancestry. *Gene*, 593, 261–264.
55. Klug, S., Ressing, M., Koenig, J., Abba, M., Agorastos, T., et al. (2009). “TP53 codon 72 polymorphism and cervical cancer: a pooled analysis of individual data from 49 studies”. *The Lancet. Oncology.* 10 (8): 772–84.
56. Koh B, Tan DJH, Ng CH, Fu CE, Lim WH, Zeng RW, Yong JN, Koh JH, Syn N, Meng W, Wijarnpreecha K, Liu K, Chong CS, Muthiah M, Luu HN, Vogel A, Singh S, Yeoh KG, Loomba R, Huang DQ. Patterns in Cancer Incidence Among People Younger Than 50 Years in the US, 2010 to 2019. *JAMA Netw Open.* 2023 Aug 1;6(8):e2328171.
57. Kong Y, Zheng J, Li L, Lu L, Wang J. Association of MTHFR Polymorphisms with H-Type Hypertension: A Systemic Review and Network Meta-Analysis of Diagnostic Test Accuracy. *Int J Hypertens.* 2022 Mar 22;2022:2861444.
58. Kopanos C, Tsiolkas V, Kouris A, Chapple CE, Albarca Aguilera M, Meyer R, Massouiras A. VarSome: the human genomic variant search engine. *Bioinformatics.* 2019 Jun 1;35(11):1978-1980.
59. Kumar, et al. (2020). Molecular basis of Cancer: Role of Genetic and Epigenetic Alterations. En Saunders (Elsevier), ed. Robbins & Cotran Pathologic Basis of Disease (en inglés) (10th edición). Elsevier. ISBN 9780323531139.
60. Kung, C.-P., Leu, J.-J., Basu, S., Khaku, S., Anokye-Danso, F., Liu, Q., Murphy, M. E. (2016). The P72R polymorphism of p53 predisposes to obesity and metabolic dysfunction. *Cell Reports*, 14, 2413–2425.
61. Lapik IA, Ranjit R, Galchenko AV. Impact of KCNJ11 rs5219, UCP2 rs659366, and MTHFR rs1801133 Polymorphisms on Type 2 Diabetes: A Cross-Sectional Study. *Rev Diabet Stud.* 2021 Spring;17(1):21-29.
62. Leclerc, D., Sibani, S., & Rozen, R. (2013). Molecular Biology of Methylene-tetrahydro-folate Reductase (MTHFR) and Overview of Mutations/Polymorphisms. En Madame Curie Bioscience Database [Internet]. Landes Bioscience.
63. Lee, Y., Bae, S., Choi, S., Ji, J., & Song, G. (2012). Associations between the p53 codon 72 polymorphisms and susceptibility to systemic lupus erythematosus and rheumatoid arthritis: A meta-analysis. *Lupus*, 21, 430–437.

64. Li Q, Xu L, Jia X, Saleem K, Zaib T, Sun W, Fu S. SNPs in folate pathway are associated with the risk of nonsyndromic cleft lip with or without cleft palate, a meta-analysis. *Biosci Rep.* 2020 Mar 27;40(3):BSR20194261.
65. Litviakov NV, Denisov EV, Takhauov RM, et al. Association between TP53 gene AR-G72PRO polymorphism and chromosome aberrations in human cancers. *Mol Carcinog.* 2010;49(6):521-524.
66. Mabhida SE, Muhamed B, Sharma JR, Apalata T, Nomatshila S, Mabasa L, Benjeddou M, Masilela C, Ziqubu K, Shabalala S, Johnson R. Methylenetetrahydrofolate Reductase Polymorphism (rs1801133) and the Risk of Hypertension among African Populations: A Narrative Synthesis of Literature. *Genes (Basel).* 2022a Apr 1;13(4):631.
67. Mabhida SE, Sharma JR, Apalata T, Masilela C, Nomatshila S, Mabasa L, Fokkens H, Benjeddou M, Muhamed B, Shabalala S, Johnson R. The association of MTHFR (rs1801133) with hypertension in an indigenous south African population. *Front Genet.* 2022b Jul 22;13:937639.
68. Meng Y, Liu X, Ma K, Zhang L, Lu M, Zhao M, Guan MX, Qin G. Association of MTHFR C677T polymorphism and type 2 diabetes mellitus (T2DM) susceptibility. *Mol Genet Genomic Med.* 2019 Dec;7(12):e1020.
69. Ministerio de Salud, Panamá. (2018). Análisis de Situación de Salud. Macro Visión Nacional de Salud.
70. Mohammed Basabaen AA, Abdelgader EA, Babekir EA, et al. TP53 Gene 72 Arg/Pro (rs1042522) Single Nucleotide Polymorphism Contribute to Increase the Risk of B-Chronic Lymphocytic Leukemia in the Sudanese Population. *Asian Pac J Cancer Prev.* 2019;20(5):1579-1585.
71. Mudd SH, Uhlendorf BW, Freeman JM, Finkelstein JD, Shih VE. Homocystinuria associated with decreased methylenetetrahydrofolate reductase activity. *Biochem Biophys Res Commun.* 1972;46(2):905-912.
72. Naccarati A, Pardini B, Polakova V, et al. Genotype and haplotype analysis of TP53 gene and the risk of pancreatic cancer: an association study in the Czech Republic. *Carcinogenesis.* 2010;31(4):666-670.
73. Naidoo, P., Naidoo, R., Ramkaran, P., Chuturgoon, A., (2020). Effect of maternal HIV infection, BMI, and NOx air pollution exposure on birth outcomes in South African pregnant women genotyped for the p53 Pro72Arg (rs1042522). *International journal of immunogenetics*, 47(5), 414-429.
74. Nakano M, Yui H, Kikugawa S, Tokida R, Sakai N, Kondo N, Endo N, Haro H, Shimodaira H, Suzuki T, Kato H, Takahashi J, Nakamura Y. Associations of LRP5 and MTHFR Gene Variants with Osteoarthritis Prevalence in Elderly Women: A Japanese Cohort Survey Randomly Sampled from a Basic Resident Registry. *Ther Clin Risk Manag.* 2021 Sep 29;17:1065-1073.
75. O'Donnell PH, Dolan ME. Cancer pharmacoethnicity: ethnic differences in susceptibility to the effects of chemotherapy. *Clin Cancer Res.* 2009 Aug 1;15(15):4806-14.

76. Olivera GG, Yáñez Y, Gargallo P, Sendra L, Aliño SF, Segura V, Sanz MÁ, Cañete A, Castel V, Font De Mora J, Hervás D, Berlanga P, Herrero MJ. MTHFR and VDR Polymorphisms Improve the Prognostic Value of MYCN Status on Overall Survival in Neuroblastoma Patients. *Int J Mol Sci.* 2020 Apr 14;21(8):2714.
77. Olivier M, Hollstein M, Hainaut P. TP53 mutations in human cancers: origins, consequences, and clinical use. *Cold Spring Harb Perspect Biol.* 2010 Jan;2(1):a001008.
78. Omran MH, Fotouh BE, Shousha WG, Ismail A, Ibrahim NE, Ramadan SS. Strong Correlation of MTHFR Gene Polymorphisms with Breast Cancer and its Prognostic Clinical Factors among Egyptian Females. *Asian Pac J Cancer Prev.* 2021;22(2):617–626.
79. Omrani-Nava V, Hedayatizadeh-Omran A, Alizadeh-Navaei R, et al. TP53 single nucleotide polymorphism (rs1042522) in Iranian patients with coronary artery disease. *Biomed Rep.* 2018;9(3):259–265.
80. Ounalli A, Moumni I, Mechaal A, et al. TP53 Gene 72 Arg/Pro (rs1042522) single nucleotide polymorphism increases the risk and the severity of chronic lymphocytic leukemia. *Front Oncol.* 2023;13:1272876.
81. Ozola A, Ruklisa D, Pjanova D. The complementary effect of rs1042522 in TP53 and rs1805007 in MC1R is associated with an elevated risk of cutaneous melanoma in Latvian population. *Oncol Lett.* 2019;18(5):5225–5234.
82. Pathak D, Shrivastav D, Verma AK, Alsayegh AA, Yadav P, Khan NH, Al-Harbi AI, Khan MI, Bihade K, Singh DD, Beg MMA. Role of metabolizing MTHFR gene polymorphism (rs1801133) and its mRNA expression among Type 2 Diabetes. *J Diabetes Metab Disord.* 2022 Feb 19;21(1):511–516.
83. Pathak S, Zajac KK, Annaji M, et al. Clinical outcomes of chemotherapy in cancer patients with different ethnicities. *Cancer Rep (Hoboken).* 2023;6 Suppl 1(Suppl 1):e1830.
84. Petersen, D., Kong, A., Jorge, L., Nebert, D., & Arias, T. (1991). Debrisoquine polymorphism: novel CYP2D6 gene Bam HI restriction fragment length polymorphism in the Ngawbé Guaymí Indian of Panama. *Pharmacogenetics*, 1(3), 136–142.
85. Polakova V, Pardini B, Naccarati A, et al. Genotype and haplotype analysis of cell cycle genes in sporadic colorectal cancer in the Czech Republic. *Hum Mutat.* 2009;30(4):661–668.
86. Politis M, Higuera G, Chang LR, Gomez B, Bares J, Motta J. Trend Analysis of Cancer Mortality and Incidence in Panama, Using Joinpoint Regression Analysis. *Medicine (Baltimore).* 2015 Jun;94(24):e970.
87. Poornima S, Daram S, Devaki RK, Merugu R, Subramanyam K. Association of MTHFR gene polymorphism C677T (rs1801133) studies with early primary knee osteoarthritis in a South Indian population: a hospital-based study. *Afr Health Sci.* 2022 Mar;22(1):338–343.
88. Raina JK, Sharma M, Panjaliya RK, Dogra V, Bakaya A, Kumar P. Association of ESR1 (rs2234693 and rs9340799), CETP (rs708272), MTHFR (rs1801133 and rs2274976) and MS (rs185087) polymorphisms with Coronary Artery Disease (CAD). *BMC Cardio-vasc Disord.* 2020 Jul 18;20(1):340.

89. Ramos, C., Castro-Pérez, E., Molina-Jiron, C., & Trejos, D. (2018). Analysis of 30 INDEL Polymorphic Markers in the Panamanian Population: Gene Admixture Estimates, Population Structure and Forensic Parameters. *Journal of Forensic Research*, 09(01).
90. Rangel-López, A., Piña-Sánchez, P., Salcedo, M., (2006). Variaciones genéticas del gen supresor de tumores TP53: relevancia y estrategias de análisis. *Revista de investigación clínica*, 58(3), 254-264.
91. Reyes, L., Godfrey, D., Ming, L. J., MacLean, C., Gonzalez, F. J., & Madrigal, L. (2021). The distribution in native populations from Mexico and Central America of the C677T variant in the MTHFR gene. *American Journal of Human Biology*, 33(6).
92. Robbins, C., Torres, J. B., Hooker, S., Bonilla, C., Hernandez, W., Candreva, A., Aha-ghotu, C., Kittles, R., & Carpten, J. (2007). Confirmation study of prostate cancer risk variants at 8q24 in African Americans identifies a novel risk locus. *Genome Research*, 17(12), 1717.
93. Romero-Gutiérrez E, Vázquez-Cárdenas P, Moreno-Macías H, Salas-Pacheco J, Tusié-Luna T, Arias-Carrión O. Differences in MTHFR and LRRK2 variant's association with sporadic Parkinson's disease in Mexican Mestizos correlated to Native American ancestry. *NPJ Parkinsons Dis*. 2021 Feb 11;7(1):13.
94. Rosenberg, N., Murata, M., Ikeda, Y., Opare-Sem, O., Zivelin, A., Geffen, E., & Seligsohn, U. (2002). The Frequent 5,10-Methylenetetrahydrofolate Reductase C677T Polymorphism Is Associated with a Common Haplotype in Whites, Japanese, and Africans. *En Am. J. Hum. Genet (Vol. 70)*.
95. Roy AM, Patel A, Catalfamo K, et al. Racial and Ethnic Disparity in Preoperative Chemotherapy Sensitivity and Survival in Patients With Early-Stage Breast Cancer. *JAMA Netw Open*. 2023;6(11):e2344517. Published 2023 Nov 1.
96. Rozen, R. (1997). Genetic predisposition to hyperhomocysteinemia: deficiency of methylenetetrahydrofolate reductase (MTHFR). *Thrombosis and Haemostasis*, 78(1), 523-536.
97. Shavers VL, Brown ML. Racial and ethnic disparities in the receipt of cancer treatment. *J Natl Cancer Inst*. 2002;94(5):334-357.
98. Shivkar RR, Gawade GC, Padwal MK, Diwan AG, Mahajan SA, Kadam CY. Association of MTHFR C677T (rs1801133) and A1298C (rs1801131) Polymorphisms with Serum Homocysteine, Folate and Vitamin B12 in Patients with Young Coronary Artery Disease. *Indian J Clin Biochem*. 2022 Apr;37(2):224-231.
99. Shubeck S, Zhao F, Howard FM, Olopade OI, Huo D. Response to Treatment, Racial and Ethnic Disparity, and Survival in Patients With Breast Cancer Undergoing Neoadjuvant Chemotherapy in the US. *JAMA Netw Open*. 2023 Mar 1;6(3):e235834.
100. Sonoyama, T., Sakai, A., Mita, Y., Yasuda, Y., Kawamoto, H., et al. (2011). TP53 codon 72 polymorphism is associated with pancreatic cancer risk in males, smokers, and drinkers. *Molecular medicine reports*, 4(3), 489-495.
101. Stabellini, N., Cullen, J., Cao, L. et al. Racial disparities in breast cancer treatment patterns and treatment related adverse events. *Sci Rep* 13, 1233 (2023).

102. Sun M, Wang T, Huang P, Diao J, Zhang S, Li J, Luo L, Li Y, Chen L, Liu Y, Wei J, Song X, Sheng X, Qin J. Association analysis of maternal MTHFR gene polymorphisms and the occurrence of congenital heart disease in offspring. *BMC Cardiovasc Disord*. 2021 Jun 14;21(1):298.
103. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin*. 2021 May;71(3):209–249.
104. Surget S, Khoury MP, Bourdon JC. Uncovering the role of p53 splice variants in human malignancy: a clinical perspective. *Onco Targets Ther*. 2013;7:57–68. Published 2013 Dec 19.
105. Terman, E., Sheade, J., Zhao, F. et al. The impact of race and age on response to neoadjuvant therapy and long-term outcomes in Black and White women with early-stage breast cancer. *Breast Cancer Res Treat* 200, 75–83 (2023).
106. The Global Cancer Observatory, WHO, (2020).
107. Thompson IM, Tangen CM, Tolcher A, Crawford ED, Eisenberger M, & Moinpour CM (2001) Association of African-American ethnic background with survival in men with metastatic prostate cancer. *Journal of the National Cancer Institute* 93:219–225.
108. Thurow, H., Haack, R., Hartwig, F., Oliveira, I., Dellagostin, O., Gigante, D., Horta, B., Collares, T., Seixas, F., (2011). TP53 gene polymorphism: importance to cancer, ethnicity, and birth weight in a Brazilian cohort. *Journal of Biosciences*. 36 (5): 823–31.
109. Usenova A, Akhunbaev S, Tumanbaev A, Lim E. The Relationship between Gene Polymorphisms of the XRCC1 and TP53 with the Gender of Children with Acute Leukemia. *Asian Pac J Cancer Prev*. 2023;24(2):613–621.
110. Walsh R, Goh BC. Population diversity in oncology drug responses and implications to drug development. *Chin Clin Oncol*. 2019;8(3):24.
111. Wang, X., Fu, J., Li, Q., & Zeng, D. (2016). Geographical and ethnic distributions of the MTHFR C677T, A1298C and MTRR A66G gene polymorphisms in Chinese populations: A meta-analysis. *PLoS ONE*, 11(4).
112. Wu H, Huang Q, Yu Z, Zhong Z. Association of ALDH2 rs671 and MTHFR rs1801133 polymorphisms with hypertension among Hakka people in Southern China. *BMC Cardiovasc Disord*. 2022 Mar 27;22(1):128.
113. Yeh FC, Yang RC, Boyle TBJ, Ye ZH, Mao JX, et al. (1997) PopGene, the user-friendly shareware for population genetic analysis, molecular biology and biotechnology centre, University of Alberta, Canada.
114. Yuan H, Fu M, Yang X, Huang K, Ren X. Single nucleotide polymorphism of MTHFR rs1801133 associated with elevated Hcy levels affects susceptibility to cerebral small vessel disease. *PeerJ*. 2020 Feb 20;8:e8627.
115. Zhao F, Miyashita M, Hattori M, et al. Racial Disparities in Pathological Complete Response Among Patients Receiving Neoadjuvant Chemotherapy for Early-Stage Breast Cancer. *JAMA Netw Open*. 2023;6(3):e233329. Published 2023 Mar 1.

116. Zhou HY, Yuan M. MTHFR polymorphisms (rs1801133) and systemic lupus erythematosus risk: A meta-analysis. *Medicine (Baltimore)*. 2020 Oct 2;99(40):e22614.
117. Ziemer MA, Mason A, Carlson DM. Cell-free translations of proline-rich protein mRNAs. *J Biol Chem*. 1982;257(18):11176-11180.

