

Relaciones genéticas entre poblaciones: análisis multivariado

María Virginia Albeza^{1,2,3}; Noemí Acreche^{1,2,3}; Cori Ramón⁴;
Antonia Picornell⁴ y José A. Castro Ocón⁴

1. CIUNSa

2. Facultad de Ciencias Naturales

3 Facultad de Humanidades, Universidad Nacional de Salta
Avenida Bolivia 5150 (4400) Salta, Argentina

4. Universitat de les Illes Balears (UIB). Islas Baleares, España
mvalbeza@unsa.edu.ar

RESUMEN

Las relaciones genéticas entre poblaciones humanas evaluadas a través de medidas de distancia genética permiten visualizar las diferencias observadas al considerar numerosos loci. A partir de microsatélites autosómicos analizados en cinco localidades de la Puna y el Valle Calchaquí de la provincia de Salta y de otras provincias argentinas, cuyos datos se obtuvieron de publicaciones, se evaluaron, en trabajos anteriores, coeficientes de distancia genética que fueron validados mediante diferentes técnicas. Los resultados obtenidos indican un gran número de asociaciones que no permiten optar por el “mejor dendrograma” por lo que se propone una metodología estadística alternativa como el análisis discriminante que refleja con más claridad la vinculación (o distancia) entre las poblaciones.

Palabras clave

Microsatélites, distancia genética, análisis multivariado, Puna, Valle Calchaquí

ABSTRACT

Genetic distances evaluated in human populations are a powerful tool to detected differences when a set of loci is included. In a previous work autosomal microsatellites from five populations of Puna and Calchaqui Valleys in Salta and several Argentine provinces (published by other authors) were considered. Different genetic distance coefficients were estimated and validated. Several dendrograms were constructed and from the results it is impossible to select the “best dendrogram”. In this work we propose an alternative methodology, multivariate analysis to classify and study group relations.

Keywords

Microsatellites, genetic distance, multivariate analysis, Puna, Calchaqui Valley



La variabilidad genética, base del proceso evolutivo, origen de gran parte de la diversidad biológica de los seres vivos y en particular la diversidad humana es un tema de interés para la Antropología Biológica. En este sentido, conocer y cuantificar la variabilidad intra e interpoblacional es aún un objetivo central independientemente del método de aproximación empleado para poblaciones humanas actuales.

Para ello es imprescindible conocer tanto la estructura genética como demográfica de las poblaciones a fin de indagar acerca de la sujeción de las diferentes poblaciones a los factores que pueden producir diferenciación y en consecuencia cambios microevolutivos. Se debe tener presente que, en ciertas ocasiones, variables como distribución diferencial en espacios geográficos, movimientos migratorios, aspectos sociales y culturales, selección de parejas, entre otras, modifican la estructura de una población lo que se refleja de manera indirecta en lo genético/evolutivo.

Sin lugar a dudas, el conocimiento y descripción de una población es el análisis básico de la diversidad genética y puede encararse tanto a partir de diferentes caracteres (o marcadores) como a diferente escala, esto es a nivel de poblaciones, regiones o entre ellas (Albeza *et al.* 2010).

Por otra parte, las relaciones entre poblaciones son evaluadas mediante distintas medidas de distancia genética, las que dan un peso diferencial a los factores evolutivos y se visualizan gráficamente a través de dendrogramas, los que pueden presentar algún grado de distorsión como consecuencia del proceso de agrupación de los conglomerados (ligamiento simple o vecino más cercano - NJ, ligamiento completo o vecino más lejano y ligamiento promedio o UPGM); sin embargo, e independientemente del método de agrupamiento utilizado, es imprescindible validar el ajuste de los mismos (Albeza *et al.* 2010).

Así como existen diferentes coeficientes de distancia genética (y/o similitud) y análisis de agrupamientos, hay cuatro técnicas que permiten validar los dendrogramas obtenidos: validación interna, externa, relativa y bootstrapping o método de remuestreo (Salzano & Callegari-Jacques 1988 e IPGRI & Cornell University 2004), aunque lamentablemente no siempre son puestas en práctica.

A partir del análisis de STRs (marcadores moleculares autosómicos, microsátélites), en muestras de Cobres y San Antonio de los Cobres (localidades pertenecientes a la región de la Puna en Salta) y de Cachi, San José y El Barrial (Valle Calchaquí), Albeza (2008) analizó la estructura genética de estas poblaciones y sus relaciones mediante diferentes medidas de distancia genética (Nei '72, Nei '78, Prevosti, Reynolds & Slatkin) y el coeficiente de similitud BAND (específico para microsátélites).

Se extendió el análisis de distancias genéticas a dieciséis poblaciones incluyendo datos publicados para otras regiones del país y una muestra de la población arqueológica de Pampa Grande (Salta) (Albeza 2008 y Albeza *et al.* 2010).

De los resultados obtenidos se observó una gran variedad de asociaciones entre las poblaciones consideradas que, si bien en ocasiones presentaban una correlación cofenética alta indicando "buen" o "muy buen ajuste", no reflejaban las mismas relaciones en todos los casos. El coeficiente de similitud BAND, específico para los marcadores genéticos estudiados, tampoco mostró con claridad las relaciones entre las poblaciones incluidas (Albeza 2008 y Albeza *et al.* 2010).

A partir de mil repeticiones y con el objeto de realizar una prueba de bootstrap, tanto para las cinco poblaciones de la Provincia de Salta como para el conjunto de 16 poblaciones, el árbol de consenso obtenido, por regla de mayoría extendida, en líneas generales presenta los mismos inconvenientes.

Se realizó un test de Mantel para evaluar tanto la correlación con la matriz de distancia geográfica en cada caso y el aislamiento por distancia como entre diferentes marcadores genéticos (se incluyó en el análisis las poblaciones para las que se contaba con información de grupos sanguíneos) para los tres conjuntos poblacionales (País, Salta - 5 poblaciones y Salta - 7 poblaciones) y todos los coeficientes tanto de distancia como de similitud genéticos (BAND, Nei '72, Nei '78, Prevosti, Reynolds y Slatkin). Los resultados obtenidos indican bajísimas correlaciones ($-0,23 < r < 0,40$) con niveles de significación que varían entre 0,22 y 0,98 por lo que se concluyó que las distancias genéticas no tienen relación con la proximidad geográfica entre poblaciones ni entre marcadores genéticos ($r = 0,10163$; $t = 2,3918$, $p = 0,9916$) (Albeza 2008 y Albeza *et al.* 2010).

Esta reducida sensibilidad de los coeficientes de distancia genética y sus correspondientes dendrogramas ya fue señalada por Acreche (2006) habiendo detectado diversas y dispares asociaciones entre 32 poblaciones analizadas a partir de datos de marcadores genéticos clásicos (grupos sanguíneos) a nivel de zona, etnia, y población.

Si bien, las medidas de distancia genética son una herramienta extremadamente útil para establecer y evaluar relaciones entre poblaciones, no siempre pueden ser validadas por diferentes técnicas por lo que su aplicación no resulta adecuada para el análisis de diferentes marcadores. En este sentido, se propone la posibilidad de incorporar técnicas alternativas y complementarias en el estudio de poblaciones humanas como análisis discriminante.

Materiales y métodos

A partir de datos de STRs autosómicos analizados en muestras de las localidades de Cobres y San Antonio de los Cobres pertenecientes a la región de la Puna y de Cachi, San José y El Barrial del Valle Calchaquí, y los obtenidos de la bibliografía para poblaciones de diferentes regiones del país, se realizó un análisis multivariado a fin de ampliar y profundizar la interpretación de los resultados provenientes de distancias genéticas.

Los STRs (short tandem repeats) o microsátélites son marcadores moleculares de secuencias cortas (2 - 7 bp) repetidas en tándem que se encuentran ampliamente distribuidos en todo el genoma. Son marcadores multialélicos que presentan un elevado grado de polimorfismo (Tomàs *et al.* 1997). El polimorfismo de estos marcadores moleculares se basa en el número de repeticiones de una secuencia unitaria y cada alelo se identifica por el número de repeticiones de esta secuencia.

Las características particulares que presentan los STRs, considerando que reúnen las que debe presentar un "buen marcador molecular" para maximizar su utilidad (tener una buena distribución a lo largo del genoma y un alto grado de polimorfismo) además de que la técnica para análisis debe ser rápida, práctica y repetible con alto grado de fiabilidad (Cal Teba, 2001), han posibilitado su uso en diferentes campos: genética de poblaciones, genética forense y genética clínica, entre otros.

Acreche (2006) señala que la aplicación de técnicas de análisis multivariado amplía el margen de la interpretación de la información obtenida ya que permiten tomar en cuenta todas las variables simultáneamente estudiando su contribución relativa a la discriminación o clasificación según el caso.

El análisis factorial, nombre genérico para un conjunto de técnicas multivariantes, reduce la dimensionalidad de un conjunto de variables con el fin de describir a los sujetos a partir de un menor número de ellas garantizando la menor pérdida de información.

No se estudia el posible efecto de un conjunto de variables independientes sobre una o más dependientes sino que se trata de extraer información significativa de una matriz formada a partir de un conjunto de variables independientes (Acreche, 2006).

En este trabajo, los casos (muestras, poblaciones) fueron evaluados según criterio de pertenencia a cada zona geográfica definida a priori y las frecuencias alélicas fueron consideradas como variables independientes. Se consideraron cuatro zonas:

- Salta: Cobres, San Antonio de los Cobres, San José, El Barrial, Cachi (Albeza 2008), Salta (Marino *et al.* 2006 a) y Pampa Grande (Carnese, comunicación personal)
- Patagonia: Río Negro y Chubut (Marino *et al.* 2006 b)
- Cuyo-Centro: Mendoza y Buenos Aires y Santa Fe (Marino *et al.* 2006 c)
- NE-Mesopotamia: Misiones, Chaco, Corrientes y Formosa (Marino *et al.* 2006 a) y se estimaron para cada grupo valores medios, desvíos y se realizó un test F de varianzas de muestras.

El test de lambda de Wilks se realizó para comprobar el poder discriminante de las variables consideradas en el análisis y la aproximación F (Rao 1952) se usó como estimador univariado de la cualidad de cada variable como discriminador.

El estadístico λ , basado en el principio de razón de verosimilitud generalizada, expresa la proporción de variabilidad total no debida a las diferencias entre grupos y permite contrastar la hipótesis nula de que las medias multivariantes (los centroides) son iguales.

La variable con mayor F fue introducida en el análisis y se estimó la Distancia Mínima de Mahalanobis entre los grupos. Esta distancia es una medida ajustada que remueve toda la correlación entre variables por medio de un tensor métrico. Esta distancia generalizada se estableció como criterio de selección de variables, las que pueden ser excluidas del análisis si la inclusión de una nueva variable prueba la ineffectividad de una previa (Acreche 2006).

El vector de coeficientes para las funciones discriminantes fue computado de una ecuación determinante y el poder discriminante de la función obtenida fue probado con un coeficiente de correlación canónica. Los centroides de grupos (scores medios) y dispersiones se obtuvieron para cada función y la identificación de pertenencia de cada sujeto al grupo se realizó en base a su score. La probabilidad de pertenencia a un grupo fue establecida inicialmente como proporcional al número de casos en cada grupo y esta probabilidad inicial con el score obtenido de la función de clasificación determina el grupo en el que tiene la mayor probabilidad de pertenencia (Acreche 2006).

La comparación entre el grupo real y el predicho, expresado como porcentaje de casos correctamente clasificados, se considera un estimador de la distancia entre grupos.

Si bien se analizaron catorce microsatélites en la Puna y once en el Valle Calchaquí (Albeza, 2008), al incluir datos de la bibliografía, el set de variables se reduce a nueve. Se consideraron 16 poblaciones y las frecuencias de 153 alelos correspondientes a 9 loci (D3S317, VWA, FGA, D8S1179, D21S11, D18S51, D5S818, D13S317 y D7S820).

Resultados

Los grupos se definieron en función de las zonas que habitan y las variables dependientes incluidas fueron las frecuencias de los alelos de los sistemas analizados. El conjunto de las frecuencias incorporadas no es homogéneo por lo que se realizaron sucesivos análisis.

De las 153 variables ingresadas, doce fueron incluidas: D3S1358-11, D3S1358-12, D3S1358-13, D3S1358-14, D3S1358-15, D3S1358-16, D3S1358-17, D3S1358-18, D3S1358-20, VWA-11, VWA-13 y VWA-14 (se indica el locus seguido del alelo). Se obtuvo un porcentaje de correcta clasificación del 100%.

Tres funciones canónicas, con muy altas correlaciones canónicas, dan cuenta del 100 % de la varianza (Tabla 1). λ de Wilks altamente significativos para las dos primeras (Tabla 2), permiten rechazar la hipótesis de que los grupos comparados tienen promedios iguales en las variables discriminantes, indicando gran diferencia entre grupos. Se puede entonces interpretar que la variabilidad entre zonas es sustancialmente mayor que la variabilidad al interior de las mismas.

De los coeficientes estandarizados (Tabla 3) se observa que los alelos D3S1358-13 y VWA-13 son las variables que más contribuyen a diferenciar los grupos en el eje definido por la función 1, en tanto que considerando la segunda función, los alelos de mayor contribución son D3S1358-15 y D3S1358-16.

En la Tabla 4 se observa que la función 1 separa con muy buena definición los centroides de las poblaciones pertenecientes a la zona Cuyo-Centro en un extremo y NE-Mesopotamia en el otro (valor positivo más alto y más bajo respectivamente). Salta queda separada del resto de las poblaciones considerando el eje definido por la función 2.

En la Figura 1 se ubican los centroides y las poblaciones clasificadas en el espacio definido por las dos primeras funciones

Tabla 1. Varianza explicada por tres funciones discriminantes

Función	Eigenvalue	% Varianza	% Acumulado	Correlación Canónica
1	226,547	82,8	82,8	0,998
2	44,494	16,3	99,1	0,989
3	2,424	0,9	100	0,841

Tabla 2. Lambda de Wilks (λ)

Funciones	λ	Chi-cuadrado	gl	Sig.
1	0,000	73,329	36	0,000
2	0,006	35,338	22	0,036
3	0,292	8,615	10	0,569

Tabla 3. Coeficientes estandarizados de las funciones discriminantes canónicas

Variables	Función		
	1	2	3
D3S1358-11	7,857	1,175	-0,719
D3S1358-12	-21,404	-1,405	-1,008
D3S1358-13	15,865	1,378	1,237
D3S1358-14	-10,295	2,672	0,141
D3S1358-15	-3,486	8,063	14,698
D3S1358-16	5,556	5,137	13,616
D3S1358-17	0,241	3,130	5,473
D3S1358-18	0,608	4,139	7,042
D3S1358-20	0,516	-0,026	0,980
VWA-11	-1,103	-0,991	-1,113
VWA-13	13,915	2,218	1,001
VWA-14	-7,828	1,386	-1,497

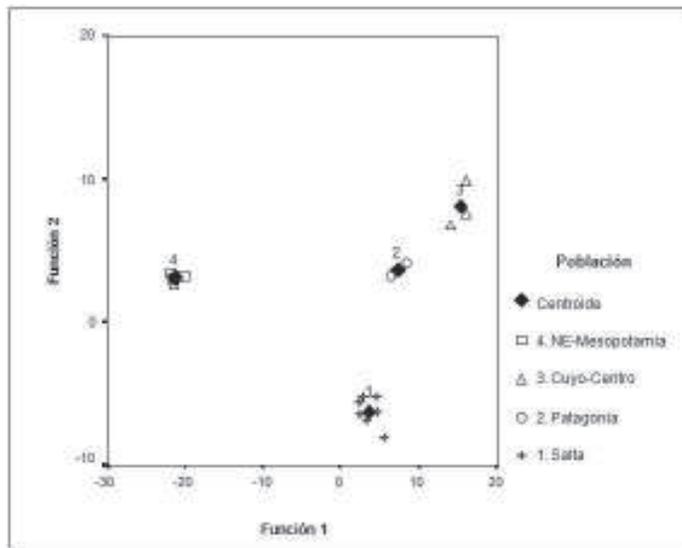


Figura 1.

Clasificación de 16 poblaciones (dos primeras funciones discriminantes canónicas)

discriminantes. Del análisis de los datos se observa que con las variables utilizadas se alcanza una muy buena discriminación de las poblaciones agrupadas según zona.

Discusión

Las medidas de distancia genética y su representación gráfica a través de dendrogramas cuyo objetivo es evaluar las relaciones entre poblaciones son las técnicas tradicionales y más empleadas en el campo de la genética de poblaciones.

Independientemente del coeficiente utilizado para medir estas vinculaciones, existen distorsiones en sus representaciones por lo que deben explicarse las relaciones entre poblaciones en función de otros factores, como por ejemplo convergencia aleatoria o alguna relación particular entre ellas como consecuencia de migraciones.

Si bien en cada coeficiente de distancia genética, como fuera

señalado, se da un peso diferencial al accionar de algunos factores evolutivos, la principal discrepancia radica en asumir si la diferencia es consecuencia de efectos de mutación y deriva génica o sólo del factor estocástico (Albeza 2008). Por otra parte, las medidas de distancia genética más adecuadas dependerán no sólo de los marcadores empleados en el análisis sino también de los diferentes modelos mutacionales y los tiempos evolutivos en cada caso (Takezaki & Nei 1996).

Pérez Lezaun *et al.* (1997) señalan, respecto a los microsatélites, que si bien tanto sus aplicaciones como su validación en el campo forense ha sido ampliamente aceptada, su utilidad o el grado de información en el campo de la genética de poblaciones es aún materia de debate, principalmente cuando se trata del análisis de las medidas de distancia genética.

Pérez-Lezaun *et al.* (1997) en función de los datos obtenidos

Tabla 4. Valor de los centroides de grupo en cada función

ZONA	Función		
	1	2	3
Salta	3,539	-6,267	0,252
Patagonia	7,377	3,644	-3,379
Cuyo-Centro	15,279	8,030	1,366
Patagonia	-21,341	3,123	0,223

para un gran número de poblaciones, sostienen que las diferencias genéticas observadas se deberían más a un fenómeno de deriva que a procesos mutacionales. Asimismo, señalan que la fidelidad con la que una medida de distancia genética refleja el rol de la mutación y la deriva en la evolución y la variación de los microsatélites es crucial para su adecuación en estudios de genética humana.

Estas dificultades teórico-metodológicas alertan acerca del empleo de diferentes coeficientes de distancia genética ampliamente difundidos en estudios de poblaciones a partir de marcadores genéticos clásicos y que probablemente no sean de aplicación ni adecuados con marcadores moleculares.

Salzano and Callegari-Jacques (1988) señalan que el uso de un creciente número de variables incrementa sin lugar a dudas, la veracidad de las comparaciones realizadas entre poblaciones y es en sí misma una herramienta para visualizar globalmente las relaciones obtenidas.

En este sentido, se debe interpretar que el análisis de variables, no sólo de las estrictamente genéticas, sino también de las provenientes de otros campos de estudio, permite conocer y comprender en profundidad la historia genética de una población. Así, la aplicación de otras herramientas estadísticas como las técnicas del análisis multivariado permiten combinar un conjunto mayor de variables.

Además de los resultados ya obtenidos por Acreche (2006) en los que se evidencia la confiabilidad y pertinencia de la utilización de estas técnicas, Acreche & Albeza (2010) analizaron datos genéticos a partir de análisis discriminante e indican que las poblaciones pueden agruparse en función de variables criterio, por lo que se puede interpretar que la altitud, el grupo lingüístico y, en menor medida, la nacionalidad son elementos que sostienen el aislamiento de las poblaciones locales y demuestran que la proximidad de los centroides Argentina y Bolivia incluidas en el trabajo es indicativa de la similitud genética entre sus poblaciones con respecto a los marcadores utilizados.

De acuerdo a los resultados obtenidos y del análisis de los mismos, en el presente trabajo se pone de manifiesto que las técnicas de análisis multivariado han sido más informativas que los coeficientes de distancia genética en el estudio de poblaciones de Salta y han reflejado su vinculación tanto a nivel zonal como con el resto del país, por lo que deben ensayarse e incorporarse nuevas estrategias que permitan evaluar las relaciones entre poblaciones humanas además de las tradicionales.

Conclusiones

De los diferentes métodos utilizados para evaluar las relaciones genéticas entre poblaciones a partir de frecuencias genéticas de microsatélites autosómicos, el análisis discriminante permite explicar con mayor consistencia y confiabilidad su asociación (y/o separación) y clasificar con un porcentaje de correcta clasificación del 100 %, su pertenencia a zonas definidas a priori diferenciando claramente las poblaciones de Salta de las del resto del país.

Referencias

- Acreche, N. 2006. Microevolución en poblaciones andinas. Universidad Nacional de Salta: Talleres Gráficos Continuos Salta SH. Salta, Argentina.
- Acreche, N & Albeza, MV. 2010. El siglo de las migraciones: entre el pluralismo y la fusión. Primer simposio internacional interdisciplinario Aduanas del Conocimiento La traducción y la constitución de las disciplinas entre el Centenario y el Bicentenario. Consultado en: http://www.expoesia.com/media/ponencia_acreche%20y%20Albeza.pdf
- Albeza, MV. 2008. Variabilidad Genética Poblacional en Salta: Análisis de STRs. Tesis Doctoral. Facultad de Ciencias Naturales. Universidad Nacional de Salta.
- Albeza, MV.; Acreche, N.; Ramón, MM.; Picornell, A & Castro Ocón, JA. 2010. Relaciones Genéticas en localidades de Salta, Argentina: ¿Qué reflejan las medidas de distancia? Revista Argentina de Antropología Biológica, 12: 37-46.
- Cal Teba, ML. 2001. Análisis de polimorfismos de ADN microsatélite del cromosoma Y. Estudio de la población de Galicia y aplicaciones forenses. Tesis Doctoral. Universidad de Santiago de Compostela, España
- IPGRI and Cornell University. 2004. Medidas de la diversidad genética. Consultado en: www.ipgri.cgiar.org. Febrero de 2014.
- Marino, M.; Sala, A & Corach, D. 2006 a. Genetic analysis of the populations from Northern and Mesopotamian provinces of Argentina by means of 15 autosomal STRs. Forensic Sci Int 160: 224-230.
- Marino, M.; Sala, A & Corach, D. 2006 b. Genetic attributes of 15 autosomal STRs in the population of two patagonian provinces of Argentina. Forensic Sci Int 160: 84-88.
- Marino, M.; Sala, A & Corach, D. 2006 c. Population genetic analysis of 15 autosomal STRs loci in the central region of Argentina. Forensic Sci Int 161: 72-77.
- Pérez-Lezaun, A; Callafell, F; Mateu, E; Comas, D; Bosch, E & Bertranpetit, J. 1997. Allele frequencies for 20 microsatellites in a worldwide population survey. Hum. Hered. 47: 189-196.
- Rao, CR.1952. Advanced Statistical Methods in Biological Research. J Wiley and Sons, Inc., New York.
- Salzano, F & Callegari-Jacques, SM. 1988. South American Indians. A Case Study in Evolution. Clarendon Press. Oxford.
- SPSS for windows, release 10.0.7 (1 jun 2000). Statistical Package for the Social Sciences.
- Takezaki, N & Nei, M. 1996. Genetic distance and reconstruction of phylogenetic trees from microsatellite DNA. Genetics 144: 389-399
- Tomas, C; Picornell, A; Castro, JA & Ramón, MM. 1997. STRs: marcadores polimorficos del genoma humà. Rev. Cie. (IEB), 19: 53-66.

