

# MICROSATELITES, DISTANCIAS GENÉTICAS Y ESTRUCTURA DE POBLACIONES NATIVAS SUDAMERICANAS

Darío A. Demarchi\*

CONICET. Laboratorio de Bioantropología. Museo de Antropología. Facultad de Filosofía y Humanidades. Universidad Nacional de Córdoba. Argentina

**PALABRAS CLAVE** Amerindios; STRs; heterocigosis; historia biológica; estructura poblacional; modelo de Harpending y Ward

**RESUMEN** En este trabajo se investigaron las relaciones genéticas entre 17 poblaciones nativas sudamericanas en relación a 15 microsatélites (STRs) autosómicos, utilizando 3 distancias genéticas –  $D_{ST}$ ,  $D_A$  y  $(\delta u)^2$  – que se ajustan a diferentes postulados teóricos. A través de diferentes técnicas de análisis (escalamiento multidimensional, correlación y correlación parcial de matrices) se puso a prueba si las distancias genéticas reflejaban las relaciones interpopulacionales esperadas a partir de la distribución geográfica o de relaciones lingüísticas entre las poblaciones. Además, se estimó en qué grado las distintas medidas de distancias genéticas eran influenciadas por la diversidad ( $H_e$ ) de cada población. Los mapas genéticos muestran, principalmente para  $D_{ST}$  y  $D_A$ , que las poblaciones aisladas y con bajo tamaño efectivo ( $N_e$ ) aparecen como *outliers*,

mientras que las poblaciones con alto  $N_e$  y mayor flujo génico ocupan una posición central a bajos valores de distancia unas de otras y sin un patrón definido de agrupamiento. La falta de asociación entre distancias genéticas y lingüísticas o geográficas y por otra parte, la alta correlación negativa entre  $H_e$  y distancias genéticas promedio por población confirman ese patrón, demostrando que la mayor parte de la variación interpopulacional puede ser explicada en función del grado de diversidad intrapoblacional. Es decir, las distancias genéticas no reflejan relaciones filogenéticas, lingüísticas o geográficas, sino más bien eventos demográficos recientes tales como cuellos de botella genético, efecto fundador o migración externa masiva. Este hecho puede ser comprobado por medio de otra metodología analítica, el modelo de Harpending y Ward. *Rev Arg Antrop Biol* 11(1):73-88, 2009.

**KEY WORDS** Amerindians; STRs; heterozygosity; population structure; biological history; Harpending and Ward Model

**ABSTRACT** In this study the genetic relationships among 17 South American native populations are investigated using 15 autosomal STRs, and 3 genetic distances –  $D_{ST}$ ,  $D_A$ , and  $(\delta u)^2$  – that conform to different theoretical models. Several analytical techniques (multidimensional scaling, correlation and partial correlation matrices analyses) were employed to test whether or not genetic distances reflect the intergroup affinities expected from geographical or linguistic relationships. Additionally, it was investigated to what extent the different measures of genetic distances were influenced by the diversity ( $H_e$ ) of each population. The genetic maps show (principally for the case of  $D_{ST}$  and  $D_A$ ) that isolated populations with low effective size ( $N_e$ ) appear as outliers, whereas populations with high gene flow and greater  $N_e$  occupy a central position on the plot, at low values of distance from each other and without a clear pattern of clustering. The lack of association between genetic distances and

linguistic or geographical distances, and the high negative correlations between genetic distances and  $H_e$  confirm this pattern, showing that most of the observed variation can be explained by the levels of within group diversity. It can be concluded that the genetic distances do not reflect phylogenetic, geographical, or linguistic relationships, but only recent demographic events such as genetic bottleneck, founder effect or massive migrations. This fact can be verified using other analytical tools, such as, the method of Harpending and Ward. *Rev Arg Antrop Biol* 11(1):73-88, 2009.

\*Correspondencia a: Darío A. Demarchi. Laboratorio de Bioantropología. Museo de Antropología. Facultad de Filosofía y Humanidades. Universidad Nacional de Córdoba. Av. Hipólito Yrigoyen 174. 5000 Córdoba. Argentina. E-mail: demarchi@ffyh.unc.edu.ar

La intención de escribir este artículo surgió luego de asistir a una presentación realizada durante las “Octavas Jornadas Nacionales de Antropología Biológica” (Distancia y Genética-Salta, Albeza M.V. et al., *Rev Arg Antrop Biol* 9(1):43, 2007) y de la discusión que se generó después de la misma. En síntesis, en ese estudio se calcularon las distancias genéticas entre poblaciones de diferentes regiones de Argentina a partir de las frecuencias alélicas de las mismas en un número variable de microsatélites autosómicos. Se observó que las relaciones biológicas entre poblaciones no reflejaban los vínculos esperados a partir de la localización de las mismas, poniéndose en duda a partir de esos resultados la utilidad de estos marcadores para investigar la variación genética de las poblaciones humanas.

En este trabajo se investigaron las relaciones genéticas entre 17 poblaciones nativas sudamericanas en relación a 15 STRs autosómicos, utilizando 3 distancias genéticas  $-D_{ST}$ ,  $D_A$  y  $(\delta u)^2$ - que se ajustan a diferentes postulados teóricos. Son sus objetivos repasar algunas propiedades de estos marcadores, ciertos conceptos básicos acerca de los alcances y limitaciones del análisis de distancias genéticas y finalmente, proponer un método de análisis alternativo que tenga en cuenta la estructura de la población.

### Microsatélites

Los microsatélites o STRs (*short tandem repeats*) son secuencias cortas de ADN, generalmente de 1 a 4 pares de nucleótidos, que se repiten en bloque a lo largo de la molécula de ADN. En an-

tropología biológica, estos polimorfismos se han convertido en los marcadores de elección en un número creciente de estudios, tanto a escala regional (Sala et al., 1998; 2006; Albeza et al., 2002; Hutz et al., 2002; Kohlrausch et al., 2005; Crosetti et al., 2008) como continental (Salzano y Callegari-Jacques, 2006; Belle y Barbutani, 2007). Existen numerosas ventajas asociadas al uso de microsatélites: (1) Son encontrados en gran número y esparcidos uniformemente a través del genoma; (2) Dada su alta tasa de mutación (entre  $10^{-3}$  y  $10^{-4}$ ), la mayoría son polimórficos, aún en poblaciones donde existe baja variabilidad en otros marcadores (proteínas, ADN mitocondrial); (3) La mayoría son selectivamente neutros, con lo cual resultan compatibles con los postulados de la genética de poblaciones; (4) Son relativamente fáciles de tipificar y los distintos alelos pueden ser caracterizados sin ambigüedades; (5) El análisis por PCR (*Polymerase Chain Reaction*), de pequeños fragmentos de ADN permite la tipificación de muestras muy degradadas; (6) Por sobre las otras ventajas, la existencia de kits comerciales utilizados rutinariamente en genética forense hacen posible la tipificación rápida y eficiente de, por ahora, hasta 15 sistemas genéticos (y más de 100 alelos) a partir de una sola reacción de PCR y a un costo cada vez más accesible.

### Distancias génicas

Las distancias genéticas pueden ser divididas en dos grandes grupos. El primero incluye las llamadas distancias geométricas, que simplemente reflejan la similitud entre poblaciones con respecto a las

frecuencias alélicas que éstas presentan en determinados sistemas genéticos, sin asumir *a priori* supuesto evolutivo alguno. A este grupo corresponden la distancia *cuerda* (Cavalli-Sforza y Edwards, 1967), la distancia de Rogers (Rogers, 1972) y la  $D_A$  (Nei et al., 1983). Constituyen simplemente una visión geométrica de las distancias entre puntos en una hiperesfera cuyo número de dimensiones es igual al de alelos considerados en la comparación. A pesar de ser relativamente antigua, la distancia  $D_A$  es utilizada frecuentemente con microsatélites debido a que se ha demostrado que es la que mejor representa las relaciones poblacionales, tanto a partir de datos generados por computadora (Takezaki y Nei, 1996) como empíricos (Takezaki y Nei, 2008).

El segundo grupo de distancias comprende a aquellas que se basan en modelos evolutivos, donde la variación es atribuida principalmente a la acción de la mutación y la deriva, asumiendo que los marcadores utilizados no están sujetos a la selección natural. La gran mayoría de estas medidas de distancia, se ajustan al modelo de alelos infinitos (Infinite Alleles Model-IAM). Según este modelo cada mutación crea aleatoriamente un alelo nuevo y por tanto, cada estado alélico está relacionado por igual a cualquier otro. En el caso particular de los microsatélites, las distancias se calculan sin tomar en cuenta el número de repeticiones en que difieren dos estados alélicos. Esto implica que el tamaño no importa. La distancia basada en el modelo IAM más utilizada es la distancia estándar ( $D_a$ ) de Nei (1972), que también ha demostrado ser adecuada para el análisis de microsatélites (Takezaki y Nei, 2008).

La distancia de Nei ha sido formulada a partir de los siguientes supuestos: tasa de mutación constante e igual para todos los loci; poblaciones con tamaños efectivos iguales y constantes a través de las generaciones; las poblaciones están en equilibrio mutación-deriva (Nei, 1972). Bajo el modelo de alelos infinitos se supone que la  $D_a$  se incrementa de manera lineal con respecto al tiempo, en caso de que se mantenga un balance mutación-deriva a través del proceso evolutivo investigado (Takezaki y Nei, 1996).

Otra de las distancias basadas en el modelo IAM, muy utilizadas es la de co-ancestralidad de Reynolds (Reynolds et al., 1983). A diferencia de la distancia  $D_{ST}$ , la distancia de Reynolds parte del supuesto de ausencia de mutación. La deriva génica es la única fuerza actuando sobre las frecuencias alélicas de las poblaciones y el incremento en las distancias génicas depende en forma lineal del tamaño de la población, que no es constante y varía entre poblaciones. Así, si el tamaño efectivo de una población se duplica, la deriva génica tendrá lugar más lentamente y en consecuencia, las distancias se incrementarán en la mitad del tiempo (Reynolds et al., 1983). Esta distancia es eficiente para comparaciones entre poblaciones con pequeños tamaños efectivos y que han divergido hace un número relativamente bajo de generaciones (Weir, 1996). En el caso de microsatélites, la distancia de Reynolds parece inapropiada debido a que estos polimorfismos poseen una tasa de mutación demasiado alta como para desdeñar su efecto.

El modelo de mutación gradual (Stepwise Mutation Model-SMM), especí-

ficamente desarrollado para explicar la evolución de microsátélites, propone que cuando mutan los microsátélites ganan o pierden una sola repetición en tandem. Esto implica que dos alelos que difieren en una sola repetición están más relacionados (tienen un ancestro común más reciente) que alelos que difieren en varias repeticiones. En otras palabras, el tamaño del fragmento sí importa al realizar cualquier análisis estadístico. Existen varias distancias desarrolladas a partir de este modelo. Entre las más utilizadas se encuentran la  $(\delta u)^2$  (Goldstein et al., 1995) y la  $D_{sw}$  (Shriver et al., 1995).

### Poblaciones y marcadores

En la Tabla 1 se presentan las 17 poblaciones nativas sudamericanas incluidas en el estudio, su referencia bibliográfica, su localización geográfica aproximada y la filiación lingüística de acuerdo a Lokotka (1968). En el mejor de los casos, se cuenta con información sobre 15 STRs (todos ellos tetraméricos), correspondiente a los loci D2S1338, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D19S433, D21S11, FGA, vWA, CSF1PO, TPOX y TH01. En los trabajos de Hutz et al. (2002), Kohlrausch et al. (2005) y Crossetti et al. (2008), se utilizó el kit AmpFISTR Identifier kit (Applied Biosystems, Foster City, California). En el trabajo de Sala et al. (2006) se utilizó el kit PowerPlex 16 (Promega Corp., Madison), mientras que en el de Albeza et al. (2002) se utilizaron los kits AmpFISTR Profiler Plus PCR amplification kit (PE Applied Biosystems) y GenePrint STR Systems kit (Promega). Debido a diferen-

cias en los distintos kits utilizados, las distancias obtenidas entre los distintos pares de poblaciones no fueron establecidas a partir del mismo número de marcadores. Las distancias obtenidas entre las 9 poblaciones descritas por Kohlrausch et al. (2005) y Crossetti et al. (2008) se basan en la información de los 15 microsátélites. Las distancias entre esas poblaciones y las tipificadas en Hutz et al. (2002) y Sala et al. (2006) se basan en 12 STRs, mientras que las distancias de todas las poblaciones respecto a la de la puna salteña (Albeza et al., 2002) se calcularon a partir de sólo 10 sistemas genéticos.

### Procesamiento estadístico

Las relaciones intermuestrales se analizaron a través de 3 medidas de distancias, basadas cada una de ellas en supuestos teóricos diferentes, de acuerdo a lo comentado más arriba y cuyos algoritmos se presentan a continuación. Una descripción más detallada puede encontrarse en Takezaki y Nei (2008).

Siendo  $X_u$  = a la frecuencia alélica del alelo u en la población X y  $Y_u$  = a la frecuencia alélica del alelo u en la población Y

$$D_A = 1 - \sum_u \sqrt{X_u Y_u}$$

$$D_a = -1 \ln(J_{XY} / \sqrt{J_X J_Y})$$

$$(\delta\mu)^2 = (\mu_x - \mu_y)^2$$

La distancia geométrica  $D_A$  (Nei et al., 1983), las distancias  $D_a$  (Nei, 1972) y  $(\delta\mu)^2$  (Goldstein et al., 1995), basadas en los modelos IAM y SMM, respectivamen-

te, así también como la heterocigosis esperada ( $H_e$ ) se calcularon con el programa Poptree, de N. Takezaki ([www.bio.psu.edu/people/faculty/nei/lab/software.htm](http://www.bio.psu.edu/people/faculty/nei/lab/software.htm)).

A partir de las matrices de distancias se construyeron mapas genéticos mediante escalamiento multidimensional (non-metric multidimensional scaling, NMS, Kruskal, 1964a, b). Este es un método de ordenación que representa las relaciones entre  $n$  muestras en un espacio de  $n$ -dimensiones, de modo tal que las distancias entre las muestras correspondan tanto como sea posible a las distancias originales calculadas entre ellas. El NMS es un método iterativo de búsqueda para la ordenación y ubicación de  $x$  entidades en  $n$  dimensiones (ejes) que haga mínimo el “stress” en la configuración final. Stress es un concepto que se refiere a la magnitud de la desviación entre las distancias existentes entre los puntos en la matriz de distancias original y las distancias en el espacio reducido de ordenación. Esta técnica es recomendable cuando se trabaja con datos de distribución no normal o bien en escalas arbitrarias o discontinuas (McCune, 1991). En la práctica los resultados del NMS no difieren demasiado de los obtenidos a partir del análisis de coordenadas principales. Sin embargo, el NMS posee la propiedad de preservar con mayor fidelidad las distancias pequeñas. Esto es debido a que un método que maximiza las varianzas, como lo es el de coordenadas principales, da más peso a las distancias de mayor magnitud en detrimento de las de valores más bajos (Rohlf, 1993).

La matriz de distancias geográficas (geodésicas) entre poblaciones se obtuvieron a partir de coordenadas aproximadas

en latitud y longitud utilizando el programa GenAlEx 6 (Peakall y Smouse, 2006).

Siguiendo la clasificación de Loukotka (1968) se obtuvo una matriz de distancias lingüísticas ordinales utilizando un criterio simple: se asignó un valor de 1 a las distancias entre poblaciones que hablan la misma lengua; un valor de 2 entre poblaciones de la misma familia lingüística, pero que hablan idiomas distintos; un valor de 3 a las distancias entre poblaciones de familias lingüísticas diferentes pero de la misma división y de 4 a las distancias entre poblaciones clasificadas en distintos grupos geográficos culturales (Tabla 1).

Se calculó la correlación entre matrices de distancias genéticas, geográficas y lingüísticas y se obtuvo su significación estadística mediante el test de Mantel (1967). Se utilizó también el test de Smouse-Long-Sokal (Smouse et al., 1986) para realizar correlaciones parciales entre matrices de distancias genéticas y lingüísticas, dejando como covariable (es decir, suprimiendo su influencia) la matriz de distancias geográficas. El coeficiente de correlación  $r$  de Pearson se utilizó para valorar el grado de asociación entre la diversidad media por población ( $H_e$ ) y la distancia media de cada población a todas las restantes.

Los mapas génicos por NMS y los test de Mantel y Smouse-Long-Sokal se llevaron a cabo con el programa NTSyS 2.1 (Exeter Software).

### **Modelo de Harpending y Ward (1982)**

Este modelo permite evaluar visualmente los roles relativos cumplidos por las fuerzas evolutivas direccionales o siste-

TABLA 1. Poblaciones sudamericanas incluidas en el estudio, su filiación lingüística de acuerdo a Loukotka (1967) y localización geográfica aproximada

Población	Long	Lat	Clasificación Cultural-geográfica	División	Familia
1. Pilagá	59,6	24,4	Paleo-Americano	Chaco	Guaycurú
1. Toba Chaco	60,6	26,0	Paleo-Americano	Chaco	Guaycurú
1. Toba Formosa	58,3	26,2	Paleo-Americano	Chaco	Guaycurú
1. Wichí Chaco Formosa	61,5	24,6	Paleo-Americano	Chaco	Mataco
1. Wichí	62,2	24,0	Paleo-Americano	Chaco	Mataco
2. Ayoreo	58,0	16,0	Paleo-Americano	Chaco	Zamuco
2. Caingang	52,0	27,0	Paleo-Americano	Brasil	Central Kaingang
3. Xavante	51,4	13,2	Paleo-Americano	Brasil	Central Gê
2. Aché	56,2	23,0	Bosque	Tropical	Nor-Central Tupi
2. Guaraní	55,0	23,0	Bosque	Tropical	Nor-Central Tupi
3. Gaviao	61,8	10,1	Bosque	Tropical	Nor-Central Tupi
3. Suruí	61,1	10,5	Bosque	Tropical	Nor-Central Tupi
3. Zoró	60,2	10,2	Bosque	Tropical	Nor-Central Tupi
3. Wai-Wai	57,5	0,4	Bosque	Tropical	Noreste Karaib
4. Puna	66,2	24,1	Andino	Sur-Central	Aymara
5. Mapuche Chubut	65,1	43,2	Andino	Sur	Mapuche
5. Mapuche Río Negro	62,6	40,5	Andino	Sur	Mapuche

Referencias: 1 Crossetti et al., 2008; 2 Kohlrausch et al., 2005; 3 Hutz et al., 2002; 4 Albeza et al., 2002; 5 Sala et al., 2006.

máticas vs. aleatorias en la diferenciación genética de las poblaciones de una región determinada. Se calcula mediante la regresión de los valores de diversidad genética sobre las desviaciones de las frecuencias alélicas de cada población en relación al centroide del arreglo (la frecuencia promedio de la población total). De acuerdo a este modelo, bajo presión sistemática uniforme (flujo génico) desde el exterior de la región, existe una correlación lineal y negativa entre la diversidad genética observada y la distancia genética de cada población ( $r_{ij}$ ) al centroide (según el modelo, éste sería el antecesor común antes de la subdivisión). Las poblaciones que recibieron desde el exterior flujo génico mayor al promedio tendrán una diversidad mayor a la predicha y ocuparán una posición por encima de la recta de regresión. De manera contraria, las poblaciones que experimentaron aislamiento (dando lugar a la

acción de la deriva génica) mostrarán valores por debajo de la recta de regresión. Una descripción más detallada puede encontrarse en Reddy et al. (2001).

## RESULTADOS

La Figura 1 es una representación bidimensional de la matriz de distancias  $D_A$ . El stress es de 0.149 lo cual, según la estimación de Kruskal (1964a), corresponde a un buen ajuste. Otra manera de medir la bondad de ajuste es calculando la correlación entre las matrices de distancias originales y las distancias euclidianas entre los puntos proyectados en los ejes de ordenación. En este caso el resultado fue de  $r = 0.932$ , un valor muy alto que confirma la eficiencia de esta técnica para representar matrices de distancias. Lo primero que se destaca en el gráfico es que la mayoría de las poblaciones se aglutinan en un área pe-

queña (resaltada en un rectángulo), a muy bajos valores de distancias unas de otras y sin una estructura de agrupamiento definida. Una mirada más detallada permite advertir allí algunas similitudes esperables, tales como las que presentan las dos muestras Tobas y la de Wichís de Formosa. Por otra parte, los Pilagá aparecen más cerca de los puneños y de los Wai Wai que de los otros grupos chaqueños y los Mapuches de Chubut de los Caingang que de los Mapuches de Río Negro. Fuera de esa pequeña área, el resto del espacio de variación genética es ocupado por los Aché, los Ayoreo, los Suruí y los Wichí del Chaco, a grandes valores de distancia entre sí y con respecto a las demás muestras, constituyendo de esta manera *outliers* dentro del conjunto. El mapa génico obtenido a partir de la  $D_{ST}$  es prácticamente idéntico al de la Figura 1, por lo que se ha omitido su presentación.

En la Figura 2 se presenta el mapa génico bidimensional construido a partir de la distancia  $(\delta\mu)^2$ . El stress es aquí de 0.212 y la correlación entre la matriz de distancias originales y su representación en el mapa génico es de  $r = 0.801$ . Estos valores indican un ajuste mucho menor que en el caso de la Figura 1, lo cual sugiere que la relación entre las muestras no puede ser reducida de manera conveniente a un espacio de sólo 2 dimensiones. No se observa aquí el apiñamiento que ocurría en la Figura 1, sino que las muestras se distribuyen más uniformemente a través del espacio de ordenación. Persiste sin embargo una escasa estructura en la distribución de las muestras. Si bien los Ayoreo, Suruí, Aché y Wichí del Chaco continúan ocupando posiciones próximas a los límites del gráfico, no aparecen a grandes valores de distancia de las otras muestras. Continúa aquí el agrupamiento entre las mues-

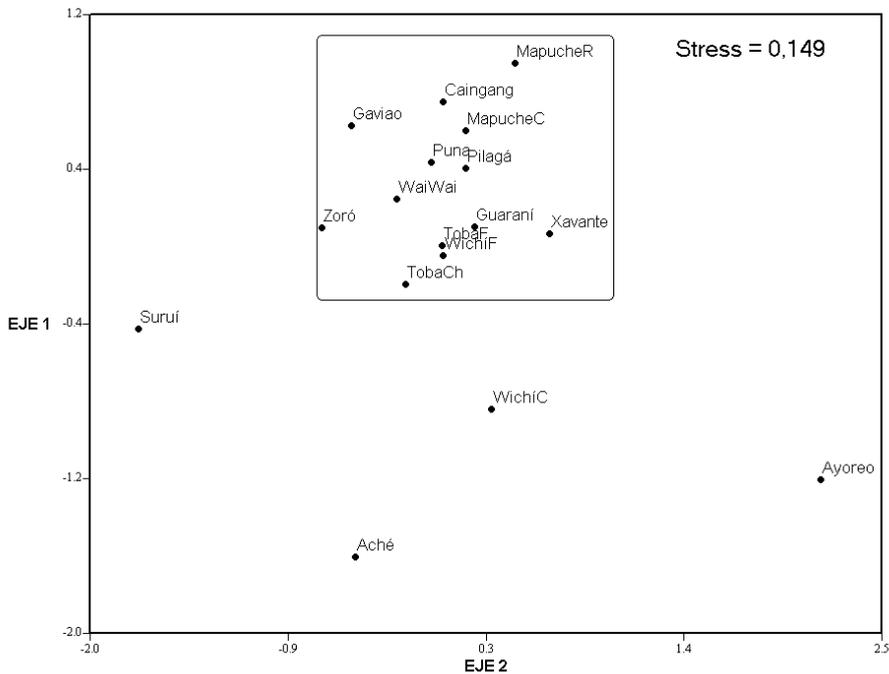


Fig. 1. Escalamiento bidimensional de la matriz de distancias  $D_A$ .

tras chaqueñas (ambas Tobas y los Wichís de Formosa), mientras que las muestras Mapuches aparecen ahora más próximas entre sí aunque, en rigor de verdad, la distancia entre los Mapuches de Chubut y los Zoró de Brasil es menor a la que presentan las dos muestras Mapuches entre sí.

### Prueba de correlación y correlación parcial de matrices

La representación bidimensional de las matrices de distancias constituye un método exploratorio útil para estimar, de manera visual y aproximada, las similitudes genéticas entre las muestras. Sin embargo, para poder cuantificar de manera inequívoca la asociación entre relaciones gené-

ticas, geográficas y lingüísticas y obtener una valoración estadística de la misma, se recurrió a la técnica de correlación de matrices. En la Tabla 2 se presentan los valores de correlación entre matrices de distancias biológicas, geográficas y lingüísticas, cuya significación estadística fue establecida a partir del test de Mantel (1967). En la misma tabla se presentan también los valores de  $r$  entre la  $H_e$  y las distancias medias (genéticas, geográficas y lingüísticas) por población. La correlación observada entre  $D_A$  y  $D_{ST}$  es extremadamente alta, es decir que, a pesar de partir de diferentes postulados teóricos representan casi exactamente las mismas relaciones intermuestrales, tal como ya se observara en los mapas genéticos. La correlación en-

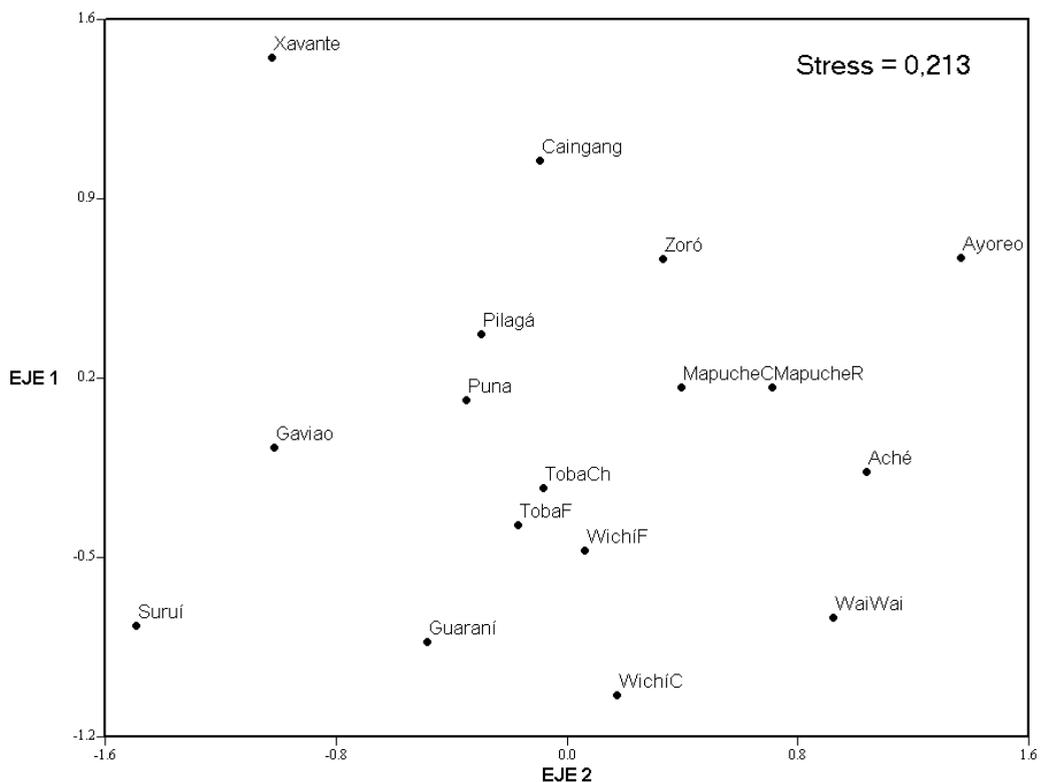


Fig. 2. Escalamiento bidimensional de la matriz de distancias  $(\delta\mu)^2$ .

MICROSATELITES Y DISTANCIAS GENETICAS

tre ambas y  $(\delta\mu)^2$  es mucho más baja, si bien significativa. Tal vez el resultado más relevante de la Tabla 2 es que, a pesar de algunas similitudes esperables observadas en los mapas génicos, ninguna de las tres matrices de distancias genéticas utilizadas muestra valores de  $r$  significativos ni con las distancias geográficas ni con las lingüísticas. Es decir, las relaciones genéticas (al menos a partir de estos marcadores) no reflejan el fenómeno de aislamiento por distancia o de posibles migraciones pasadas, ni de flujo génico favorecido por la existencia de vínculos lingüísticos o de un origen común reciente. Los resultados del análisis de correlación parcial indican que, después de eliminar el factor geográfico (sin duda importante en un área del tamaño de Sudamérica), la asociación entre vínculos lingüísticos y variación genética continúa siendo muy baja.

Por otra parte, la correlación negativa entre  $H_e$  y las distancias  $D_S$  y  $D_A$  es muy alta y pone en evidencia que estas distancias son extremadamente sensibles a la diversidad intramuestral. También se

observa una correlación negativa entre  $H_e$  y  $(\delta\mu)^2$  aunque mucho menor, hecho que demuestra que esta distancia no es tan dependiente de la diversidad.

La correlación entre las matrices de distancias geográficas y lingüísticas es alta y estadísticamente significativa. Este resultado es previsible ya que frecuentemente, poblaciones cercanas geográficamente hablan lenguas similares. En cambio, no es tan esperable que la  $H_e$  se asocie positivamente (si bien no llega a ser estadísticamente significativa) con las distancias geográficas promedio y en menor medida, con las distancias lingüísticas. La teoría prevería una correlación negativa, dado que poblaciones aisladas geográficamente tenderían a mostrar valores bajos de diversidad por reducción de flujo génico. Un análisis más detallado de los datos permite descubrir que este resultado se debe a que las dos poblaciones Mapuche, con alta heterocigosis (probablemente debido al mestizaje), son las que presentan los mayores valores de distancia geográfica promedio.

TABLA 2. Correlación entre matrices de distancias genéticas, lingüísticas (LEN) y geográficas (GEO) y entre distancias medias por población y diversidad intragrupal ( $H_e$ )

	$D_S$	$D_A$	$(\delta\mu)^2$	GEO	LEN
$D_A$	0,9101				
$(\delta\mu)^2$	0,5401	0,5091			
Geográfica	-0,131	-0,165	0,063		
Lenguaje	0,084	0,146	0,116	0,5111	
$H_e$	-0,8171	-0,8021	-0,5061	0,367	0,271
<sup>2</sup> Lenguaje GEO	0,176	0,187	0,107	-----	-----

1  $p < 0.01$

2 correlación parcial tomando las distancias geográficas como covariable.

### Modelo de Harpending y Ward

Otra manera de abordar el estudio de la variación genética en una región determinada, es a través del análisis de la estructura de la población, es decir, al balance entre la acción de las fuerzas evolutivas sistemáticas -o direccionales- y aleatorias. Para llevar a cabo este propósito se utilizó el modelo de Harpending y Ward (1982), cuyos supuestos básicos se mencionaron previamente. En la Figura 3 se observa a los Wichís y Tobas de Formosa cerca del origen (que correspondería a la población ancestral, según el modelo), levemente por debajo de la línea de regresión y próximos entre sí. Esta posición sugiere un tamaño efectivo relativamente grande (por lo cual se ha diferenciado poco de la población ancestral) y por otra parte, la existencia de flujo génico entre ambas poblaciones. Los Wai-Wai, Guaraníes, Tobas de Chaco y Xavantes se ubican próximos a la línea de regresión, de acuerdo a lo que predice el modelo (es decir en equilibrio entre la  $H_e$  observada y la distancia al origen). La ubicación por encima de la recta de las muestras Pilagá, Puna, ambas Mapuche y Caingang, sugiere que estas poblaciones han recibido flujo génico diferencial desde el exterior (es decir, de poblaciones no incluidas en el análisis), probablemente no amerindio. Reforzando esta suposición, se ha observado en los Pilagá una mayoría de linajes paternos no nativos (Demarchi y Mitchell, 2004) y algunos linajes maternos africanos (datos no publicados). Por otra parte, es conocida la existencia de un componente no amerindio en el pool génico de las poblaciones

Mapuche (Carnese et al., 1996) y de la Puna, si bien no se conoce exactamente el grado de mestizaje en esta región (Albeza et al., 2002). Lo mismo puede decirse de los Caingang, quienes han sufrido en las últimas décadas un alto grado de aculturación (Kohlrausch et al., 2005). El resultado más notorio de este análisis es la posición que ocupan en el gráfico los Suruí, Ayoreo y Aché, extremadamente distantes del origen y con una diversidad genética muy baja en relación a las demás poblaciones. Esta ubicación sugiere aislamiento prolongado y tamaños efectivos pequeños. Si bien las poblaciones Ayoreo y Aché no tienen tamaños poblacionales tan bajos, lo cual si se verifica entre los Suruí, sus tamaños efectivos se encuentran muy reducidos, tal cual lo evidencia el bajo número de linajes maternos existente en esas poblaciones (Dornelles et al., 2004; Schmitt et al., 2004). Esta pérdida de diversidad genética sería el reflejo de la acción de fuerzas aleatorias tales como deriva genética, cuello de botella o efecto fundador. Los Wichís del Chaco también se ubican por debajo de la línea de regresión y bastante lejos de los demás grupos del Chaco argentino, lo cual concuerda con lo encontrado en otros trabajos donde ya se había observado la pérdida de diversidad en esta población (Demarchi et al., 2001; Demarchi y Mitchell, 2004). El  $F_{ST}$  que se obtiene como un subproducto de este análisis a partir del promedio de las distancias ( $r_{ij}$ ) de cada muestra al centroide, es relativamente alto (0.055), pero son estas últimas 4 poblaciones las que más contribuyen a ese valor de diferenciación interpoblacional.

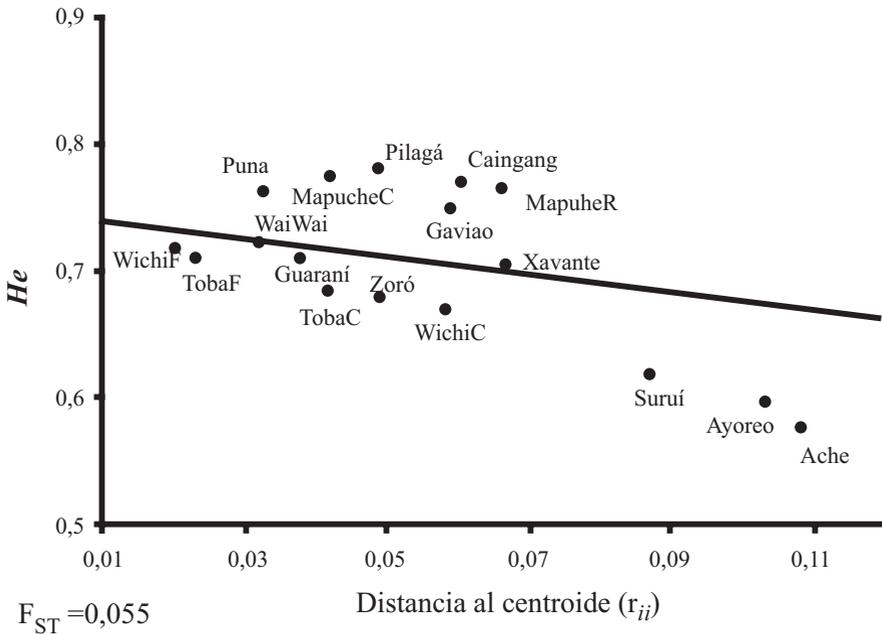


Fig. 3. Análisis de la estructura poblacional mediante el método de Harpending y Ward (1982).

## DISCUSION

La variación biológica entre poblaciones humanas puede ser investigada por medio de diferentes herramientas metodológicas. Sin embargo, en la bibliografía internacional se destaca nítidamente el uso de distancias biológicas. La representación gráfica de las matrices de distancias a través de dendrogramas (UPGMA o Neighbor Joining trees) o mapas génicos (componentes principales, NMS, coordenadas principales) son comúnmente usados para hacer inferencias con respecto a patrones geográficos, lingüísticos o etnohistóricos (Cavalli-Sforza et al., 1994). La interpretación más simple y a menudo la única considerada por los investigadores, es que distancias ge-

néticas menores corresponden a pares de poblaciones con un ancestro común más reciente (Relethford, 1996). Sin embargo, las distancias biológicas se ven afectadas fuertemente por la estructura de la población, es decir el equilibrio entre flujo y deriva genética, la distribución geográfica o el tamaño efectivo de la misma, si bien estos factores son raramente tenidos en cuenta (Harpending y Jenkins, 1973; Relethford, 1996). Cuando una población ancestral se divide en dos poblaciones descendientes, se supone que las frecuencias alélicas de éstas son idénticas a la de origen y que la división produjo inmediatamente aislamiento reproductivo. Sin embargo, cuando se produce un cuello de botella, los valores de distancia genética aumentan rápidamente. De este modo,

la tasa de evolución varía considerablemente aunque la tasa de mutación sigue siendo la misma. Es decir, la diversidad genética tiene un gran impacto en las distancias génicas. Así, la magnitud de los valores de distancia será exagerada para las poblaciones con menor diversidad. Por otra parte, todas las distancias genéticas tienden a cero al aumentar el tamaño poblacional: teóricamente, en poblaciones infinitamente grandes las distribuciones de alelos serían idénticas, conforme a la distribución de probabilidad de alelos (Takezaki y Nei, 1996). Obviamente, éste no es un hallazgo del autor de este artículo pero resulta conveniente recordarlo dado que es a menudo olvidado a la hora de explicar las posibles causas que determinan las distancias genéticas entre poblaciones.

La Figura 3 permite observar que las poblaciones nativas sudamericanas incluidas en el análisis, han tenido historias biológicas muy dispares. Algunas, como los Tobas de Chaco y Formosa y los Wichí de Formosa, han mantenido tamaños efectivos relativamente grandes y altas tasas de migración dentro de la región (Demarchi et al., 2001; Cabana et al., 2006). Otras, como los Pilagá, los puneños o los Mapuche parecen haber recibido un considerable flujo génico no nativo, lo cual ha incrementado su diversidad genética. Por último, algunas comunidades (como los Aché, Ayoreo, Suruí y los Wichí del Chaco) han sufrido en algún momento de su historia evolutiva una reducción muy drástica de sus tamaños efectivos y de su diversidad genética, hecho que ha acelerado su diferenciación con respecto a los demás pueblos de la región.

Diferencias genéticas muy marcadas entre poblaciones no necesariamente se corresponden con barreras físicas evidentes o grandes distancias geográficas. Particularmente en Sudamérica, se han observado barreras genéticas importantes entre poblaciones separadas por sólo unos pocos kilómetros (Belle y Barbujani, 2007). El pequeño tamaño efectivo de las poblaciones cazadoras-recolectoras y su tendencia a dividirse a lo largo de líneas familiares contribuyen a incrementar el efecto de la deriva génica (Crawford, 1998). En consecuencia, cuando la fisión es seguida por aislamiento, las divergencias genética y lingüística derivadas de esos eventos reforzarán la independencia evolutiva entre las poblaciones, incluso en ausencia de otras barreras reproductivas (Belle y Barbujani, 2007).

Es necesario además tener presente que la colonización europea y sus consecuencias han modificado en gran medida la composición genética de la población aborigen actual. Por un lado, hubo una reducción de linajes por guerras, hambrunas y epidemias (Crawford, 1998). Por el otro, durante 4 siglos se produjo el mestizaje con europeos y africanos, pero también entre etnias locales debido a movimientos poblacionales surgidos a partir de la colonización (encomiendas, misiones evangélicas, zafrá azucarera, etc.). Resulta entonces necesario tener precaución a la hora de emitir hipótesis sobre eventos evolutivos que ocurrieron hace cientos de generaciones, a partir del análisis de distancias genéticas entre poblaciones contemporáneas, siendo indispensable valerse también de toda la información disponible proveniente de la

arqueología, antropología física, lingüística y de cualquier otra fuente confiable de evidencias.

Finalmente, en relación a los marcadores genéticos utilizados, no puede dejar de mencionarse que las altas tasas de mutación que presentan los microsátelites, sumadas a limitaciones físicas o selectivas en el límite superior de los tamaños alélicos, lleva a que algunos estados alélicos ocurran independientemente en diferentes poblaciones (Mank y Avise, 2003). De esta forma, perfiles genéticos similares pueden ser consecuencia de homoplasia (convergencia evolutiva) en vez de flujo génico o ancestro común reciente. Esto puede conducir a una estimación errónea del grado de similitud entre poblaciones (Mank y Avise, 2003). Dado que la probabilidad de homoplasia aumenta con el tiempo de divergencia, la utilización de STRs en estudios a escala continental, entre poblaciones que han divergido hace cientos de generaciones, debería ser hecha con precaución. A pesar de esos problemas, debido a sus altos niveles de polimorfismos los STRs son de gran utilidad para investigar las relaciones genéticas entre poblaciones cercanas entre sí (Takezaki y Nei, 2008), abarcando áreas geográficas y rangos temporales más reducidos.

## CONCLUSIONES

Las diferencias en las historias evolutivas y formas de vida de las poblaciones conducen a valores de diversidad muy diferentes, lo que lleva a aumentar las distancias génicas en algunos casos mientras que las disminuyen en otros, no reflejan-

do las verdaderas relaciones evolutivas entre las mismas. En poblaciones que han sobrellevado períodos de cuellos de botella genéticos, los alelos sobrevivientes pueden servir de diagnóstico sobre su propia historia evolutiva, pero son poco informativos o podrían llevar a confusión sobre las relaciones evolutivas con otras poblaciones.

Por estos motivos, llevar a cabo análisis de distancias génicas incluyendo poblaciones aisladas con bajos tamaños efectivos junto a otras de mayor tamaño, el flujo génico (muchas veces con aporte multiétnico), no constituye un modo eficiente para estudiar los patrones de variación genética en una región geográfica dada, a no ser que estos factores históricos y demográficos sean conocidos y en consecuencia, valorados en su justa dimensión. Cuando esos factores son desconocidos, resulta provechosa la utilización de otros métodos de análisis, como el de Harpending y Ward, que permiten investigar la variación genética en una región a partir de hipótesis específicas sencillas que pueden ser puestas a prueba simplemente comparando la variación intrapoblacional observada y la esperada en un escenario de flujo génico uniforme.

## AGRADECIMIENTOS

Daniel Corach proporcionó gentilmente los datos sobre las frecuencias génicas de las muestras de Mapuches de Río Negro y Chubut. Agradezco a Graciela Bailliet, Mariana Fabra y Angelina García por la lectura crítica y las sugerencias realizadas sobre una versión anterior de este manuscrito.

## LITERATURA CITADA

- Albeza MV, Picornell A, Acreche N, Tomas C, Castro JA, Ramon MM. 2002. Genetic variability at 14 STR loci in the Puna population of north western Argentina. *Int J Legal Med* 116:126-132.
- Albeza MV, Acreche N, Ramón RM, Picornell A, Castro JA. 2007. Distancia y Genética – Salta. *Rev Arg Antrop Biol* 9 (1):43.
- Belle EM, Barbujani G. 2007. Worldwide analysis of multiple microsatellites: language diversity has a detectable influence on DNA diversity. *Am J Phys Anthropol* 133:1137-1146.
- Cabana G, Merriwether AD, Hunley KL, Demarchi DA. 2006. Is the genetic structure of Gran Chaco populations unique? Interregional perspectives on Native South American mitochondrial DNA variation. *Am J Phys Anthropol* 131:108-119.
- Carnese FR, Caratini A, Goicoechea A, Weimer TA, Franco M, Hutz MH, Salzano FM. 1996. Demography and blood genetics of Argentinian Mapuche Indians. *Int J Anthropol* 11(2-4 ):33-42.
- Cavalli-Sforza LL, Edwards AW. 1967. Phylogenetic analysis: models and estimation procedures. *Am J Hum Genet* 19:233-257.
- Cavalli-Sforza LL, Piazza A, Menozzi P. 1994. The history and geography of human genes. Princeton: Princeton University Press.
- Crawford MH. 1998. The origins of Native Americans. Evidence from anthropological genetics. New York: Cambridge University Press.
- Crossetti SG, Demarchi DA, Raimann PE, Salzano FM, Hutz MH, Callegari-Jacques SM. 2008. Autosomal STR genetic variability in the Chaco native population: homogeneity or heterogeneity? *Am J Human Biology* 20:704-711.
- Demarchi DA, Mitchell RJ. 2004. Genetic structure and gene flow in Gran Chaco populations of Argentina: evidence from Y-chromosome markers. *Hum Biol* 76:413-429.
- Demarchi DA, Panzetta-Dutari GM, Motran CC, Basualdo MA, Marcellino AJ. 2001. Mitochondrial DNA haplogroups in Amerindian populations from the Gran Chaco. *Am J Phys Anthropol* 115:199-203.
- Dornelles CL, Battilana J, Fagundes NJ, Freitas LB, Bonatto SL, Salzano FM. 2004. Mitochondrial DNA and Alu insertions in a genetically peculiar population: the Ayoreo Indians of Bolivia and Paraguay. *Am J Hum Biol* 16:479-488.
- Goldstein DB, Ruiz-Linares A, Cavalli-Sforza LL, Feldman MW. 1995. Genetic absolute dating based on microsatellites and the origin of modern humans. *Proc Natl Acad Sci USA* 92:6723-6727.
- Harpending HC, Jenkins T. 1973. Genetic distance among southern African populations. En: Crawford MH, Workman PL, editores. *Methods and theories of anthropological genetics*. Albuquerque: University of New Mexico Press. p 177-199.
- Harpending HC, Ward RH. 1982. Chemical systematics and human populations. En: Nitecki M, editor. *Biochemical aspects of evolutionary biology*.

- Chicago: University of Chicago Press. p 213-256.
- Hutz MH, Callegari-Jacques SM, Almeida SE, Armbrorst T, Salzano FM. 2002. Low levels of STRP variability are not universal in American Indians. *Hum Biol* 74:791-806.
- Kohlrausch FB, Callegari-Jacques SM, Tsuneto LT, Petzl-Erler ML, Hill K, Hurtado AM, Salzano FM, Hutz MH. 2005. Geography influences microsatellite polymorphism diversity in Amerindians. *Am J Phys Anthropol* 126:463-470.
- Kruskal J. 1964a. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* 29:1-27.
- Kruskal J. 1964b. Nonmetric multidimensional scaling: a numerical method. *Psychometrika* 29:115-129.
- Loukotka C. 1968. Classification of South American Indian languages. Los Angeles: Latin American Center, University of California.
- Mank JE, Avise JC. 2003. Microsatellite variation and differentiation in North Atlantic Eels. *J Hered* 94:310-314.
- Mantel N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Res* 27:209-220.
- McCune B. 1991. Multivariate analysis on the OC ORD system. Department of General Science, Oregon State University.
- Nei M. 1972. Genetic distance between populations. *Am Nat* 106:283-292.
- Nei M, Tajima F, Tateno Y. 1983. Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. *J Mol Evol* 19:153-170.
- Peakall R, Smouse PE. 2006. GENALEX 6: Genetic Analysis in Excel. Population genetic software for teaching and research. *Mol Ecol Notes* 6:288-295.
- Reddy BM, Demarchi DA, Malhotra KC. 2001. Patterns of biological variation among the 20 endogamous groups of Dhangar Caste-cluster of Maharashtra, India. *Coll Antropol* 25:425-442.
- Relethford JH. 1996. Genetic drift can obscure population history: Problem and solution. *Hum Biol* 68:29-44.
- Reynolds J, Weir BS, Cockerham CC. 1983. Estimation of the coancestry coefficient: basis for a short-term genetic distance. *Genetics* 105:767-779.
- Rogers JS. 1972. Measures of genetic similarity and genetic distances. *Studies in Genet VII* 7213:145-153.
- Rohlf FJ. 1993. Numerical taxonomy and multivariate analysis system. New York: Exeter Software.
- Sala A, Penacino G, Corach D. 1998. Comparison of allele frequencies of eight STR loci from Argentinian Amerindian and European populations. *Hum Biol* 70:937-947.
- Sala A, Marino ME, Bobillo MC, Corach D. 2006. Evaluación de la correlación entre atributos genéticos y apellidos: Características de STRs autosómicos en individuos con apellido Mapuche y Europeo. *Rev Arg Antrop Biol* 8:9-21.
- Salzano FM, Callegari-Jacques S. 2006. Amerindian and non Amerindian autosome molecular variability—a test analysis. *Genetica* 126:237-242.
- Schmitt R, Bonatto SL, Freitas LB, Mutschner VC, Hill K, Hurtado AM, Salzano FM. 2004. Extremely limited mitochondrial DNA variability among the

- Aché Natives of Paraguay. *Ann Hum Biol* 31:87-94.
- Smouse PE, Long JC, Sokal RR. 1986. Multiple regression and correlation extensions of the Mantel test matrix correspondence. *Syst Zool* 35:627-632.
- Shriver MD, Jin L, Boerwinkle E, Deka R, Ferrell RE, Chakraborty R. 1995. A novel measure of genetic distance for highly polymorphic tandem repeat loci. *Mol Biol Evol* 12:914-920.
- Takezaki N, Nei M. 1996. Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics* 144:389-399.
- Takezaki N, Nei M. 2008. Empirical tests of the reliability of phylogenetic trees constructed with microsatellite DNA. *Genetics* 178:385-392.
- Weir BS. 1996. *Genetic Data Analysis II: Methods for discrete population genetic data*. Sunderland: Sinauer Assoc.