



**ZTF-FCT**  
Zientzia eta Teknologia Fakultatea  
Facultad de Ciencia y Tecnología

**GRADO EN BIOLOGIA**

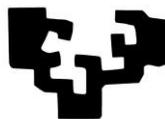
**TRABAJO DE FIN DE GRADO**

**DIVERSIDAD GENÉTICA DE  
POLIMORFISMOS AUTOSÓMICOS EN  
POBLACIONES AFRICANAS**

**JANIRE ALLENDE LÓPEZ**

**Leioa, Julio 2013**

eman ta zabal zazu



Universidad  
del País Vasco

Euskal Herriko  
Unibertsitatea

## ÍNDICE

Abstract.....	2
Introducción.....	2
Objetivos.....	4
Materiales y métodos.....	5
Resultados.....	8
Discusión.....	17
Referencias.....	19
Anexos.....	21

## **ABSTRACT**

African continent is the source of all the anatomically modern humans, who later spread all over the world. Africa has the main level of genetic variation, thus, it is an interesting region to study. In this research the aim is to analyze how is distributed the genetic variability in Africa regarding every single studied polymorphism, such as simple nucleotide polymorphism (SNP), blood types and microsatellites (STR); and to compare the information that each type of marker offers. It's observed a differentiation, above all Khoi-San and afroasiatics, but with differences depending of which polymorphism has been studied. In general, the natural selection seems to have affected similarly the SNPs of coding genes and blood types.

El continente africano es la fuente de todos los seres humanos anatómicamente modernos, que más tarde se dispersaron por todo el planeta. África posee el mayor nivel de variación genética, por lo que es una región de interés a estudiar. En este trabajo se ha analizado como se distribuye la variabilidad genética en África con cada uno de los polimorfismos estudiados, polimorfismos de nucleótido simple (SNP), grupos sanguíneos y microsatélites (STR) y se ha comparado la información que ofrece cada tipo de marcador. Se observa una diferenciación sobre todo de las familias Khoi-San y afroasiáticos pero con diferencias según que tipo de polimorfismo que se ha estudiado. En general la selección natural parece haber afectado de forma similar a SNPs de genes codificantes y a grupos sanguíneos.

## **INTRODUCCIÓN**

África ha sido identificado como la parte del mundo donde el nivel de la variación genética humana es mayor y los estudios genómicos han confirmado también que las poblaciones africanas tienen los menores niveles de desequilibrio de ligamiento (Reich *et al*, 2001).

África es una región importante para estudiar la diversidad genética humana debido a su compleja historia de la población y la dramática variación en el clima, la dieta y la exposición a las enfermedades infecciosas, que se traducen en altos niveles de variación genética y fenotípica de las poblaciones africanas. También posee una gran diversidad lingüística (Figura 1) y cultural, ya que hay más de 2.000 grupos

etnolingüísticos distintos, que hablan lenguas que constituyen casi un tercio de las lenguas del mundo.



**Figura 1.** Mapa de la distribución de idiomas africanos dentro de África. Las lenguas africanas se han clasificado en cuatro grandes familias lingüísticas: Niger-Congo (hablado principalmente por poblaciones agricultoras con una amplia distribución geográfica), Afro-Asiática (hablado por el norte y el este de África por pastores y agricultores), Nilo-Saharan (hablado en el este y por los pastores de África central) y Khoi-San (principalmente en el sur y este de África por poblaciones cazadoras-recolectoras y ganaderas).

Los africanos viven en climas que van desde el desierto más grande del mundo y la selva tropical, a la sabana, pantanos y tierras altas de montaña. Algunos de estos entornos han sufrido cambios dramáticos en el transcurso de la evolución humana moderna. Estas poblaciones practican una amplia gama de patrones de subsistencia, incluyendo varios modos de agricultura, pastoreo, caza y recolección en todo el continente, tal vez en respuesta a esta variabilidad del medio ambiente a través del tiempo y el espacio geográfico.

Varios estudios han indicado que las poblaciones ancestrales estuvieron geográficamente estructuradas antes de que los humanos modernos emigraran de África (Campbell y Tishkoff, 2008). Un reciente análisis de la variabilidad de la forma craneal en fósiles humanos anatómicamente modernos (que data de 200 a 60 kya) de África y el Medio Oriente, informó de un alto nivel de divergencia morfológica entre estos homínidos fósiles que se interpretó como evidencia de la estructura de la población ancestral en el Pleistoceno de África (Grunz *et al*, 2009). Por lo tanto, podría decirse que, una cantidad considerable de la diversidad genética y fenotípica puede haber estado presente en una etapa temprana de la evolución humana moderna. Posteriormente, la expansión geográfica de un pequeño número de humanos anatómicamente modernos

fuera de África dio lugar a un cuello de botella poblacional y la consiguiente pérdida de la diversidad genética. El tamaño de la población ancestral que dejó África se estima en alrededor de 1.000 varones y hembras fundadoras eficaces basados en loci autosómicos de microsatélites (Liu *et al*, 2006) o de alrededor de 1.500 varones y hembras fundadoras eficaces basados en un combinado de datos de ADNmt, el cromosoma Y-, y cromosoma X (Garrigan *et al*, 2007).

El patrón de variación genética en las poblaciones africanas modernas se ve influenciado por la historia demográfica (cambios en el tamaño de la población, y eventos de migración y mestizaje), así como las fuerzas de locus específicos, tales como la selección natural, la recombinación y mutación. Por ejemplo, la migración de los hablantes bantúes agrícolas de África Occidental por toda el África subsahariana en los últimos ~ 4000 años y la posterior mezcla con las poblaciones indígenas, ha tenido un gran impacto en los patrones de variación de las poblaciones africanas modernas (Pilkington *et al*, 2008).

## **OBJETIVOS**

Existen numerosos estudios que analizan la diversidad genética en poblaciones africanas. No obstante, no existe ninguna publicación que considere simultáneamente la información genética procedente de polimorfismos clásicos, polimorfismos de nucleótido simple y microsatélites. Además es importante resaltar que los distintos mecanismos evolutivos afectan de forma diferencial a estos polimorfismos. Así, además de la deriva genética y la mutación, en los polimorfismos que se transcriben a proteínas como en los grupos sanguíneos o los SNPs de genes, la selección natural podría también modificar las frecuencias alélicas. En cambio, los STR no se ven afectados por la selección natural.

Ante esta situación en el presente trabajo se plantean dos objetivos:

1. Analizar la distribución de la variabilidad genética en poblaciones africanas con los distintos tipos de polimorfismos.
2. Comparar la información proporcionada por los distintos tipos de marcadores para tratar de identificar que mecanismos evolutivos moldean las distribuciones de las frecuencias alélicas en poblaciones africanas.

Para alcanzar estos objetivos, se han creado 4 bases de datos con los distintos tipos de polimorfismos y se ha realizado el análisis estadístico de las poblaciones.

## **MATERIALES Y MÉTODOS**

### **1. Bases de datos**

Se han creado cuatro bases de datos para el análisis de diferentes polimorfismos.

La primera está formada por frecuencias de polimorfismos de nucleótido simple y se generó a partir de 96 SNPs (Anexo 1) y 15 poblaciones (San, Bantu, Pigmeos Biaka, Hausa, Ibo, Lisongo, Pigmeos Mbuti, Yoruba, Chagga, Jews, Masai, Sandawe, Zaramo, Mandenka y Mozabite) (Anexo 2). Los datos se han recogido en la base de datos ALFRED (<http://alfred.med.yale.edu/>) de donde se han seleccionado poblaciones que representen todo el continente africano y SNPs que hayan sido estudiados en todas estas poblaciones.

Una segunda base de datos con información sobre grupos sanguíneos con 17 alelos o haplotipos (ABO\*A, ABO\*B, ABO\*O, FY\*A, FY\*B\*0, KEL\*K, KEL\*k, MN\*M, MN\*N, RH\*CDE, RH\*Cde, RH\*CdE, RH\*Cde, RH\*cDE, RH\*cDe, RH\*cdE y RH\*cde) y 32 poblaciones (Anexo 2).

La tercera base de datos con 52 STR (Anexo 1) y 8 poblaciones (Pigmeos Biaka, Pigmeos Mbuti, Bantúes de Kenya, San, Yoruba, Mandenka, Bantúes de South Africa y Mozabites) (Anexo 2) creada a partir de la base de datos popSTR (STR population data comparisons) de la cual se han utilizado todos los STR disponibles (<http://spsmart.cesga.es/popstr.php>).

Dado que la anterior base de datos presenta pocas poblaciones, se ha creado otra base de datos con los STR más comunes en más poblaciones. Está formada por 12 STR y 25 poblaciones (Anexo 2), obtenida a partir de la base de datos Autosomal STR DNA Database (<http://www.strdna-db.org/>) de la cual se seleccionaron únicamente 12 de los STR (CSF1PO, D3S1358, D5S818, D7S820, D8S1179, D13S317, D18S51, D21S11, FGA, TH01, TPO, VWA).

## 2. Programas informáticos

Existen diferentes programas informáticos para realizar los tratamientos de datos que se realizan habitualmente a partir de frecuencias. En este trabajo se han utilizado los programas GeDis, PAST, Phylip y Treeview.

El programa GeDis ha sido diseñado para cubrir las necesidades básicas de tratamiento de datos relativos a frecuencias génicas en Antropogenética. Se ha diseñado un programa con una entrada de datos sencilla y relativamente flexible, se han implementado algunos de los métodos más comunes para el tratamiento de los datos y se ha facilitado la exportación de datos a otros programas. (Peña *et al*, 2009). Con GeDis se han obtenido las distancias genéticas  $F_{ST}$  de las cuatro bases de datos, para realizar matrices de distancias  $F_{ST}$  de cada par de poblaciones en cada una de las bases de datos. La distancia  $F_{ST}$  mide el grado de diferenciación génica entre las poblaciones que se ha obtenido a partir de las frecuencias alélicas.

PAST es un paquete de análisis de datos gratuito fácil de usar originalmente destinado a la paleontología, pero que ahora también es popular en muchos otros campos. Incluye estadística común, representación gráfica y modelado de funciones. Mediante PAST se ha realizado un análisis de escalamiento multidimensional (MDS) para cada una de las bases de datos. El MDS es una técnica multivariante de interdependencia que trata de representar en un espacio geométrico de pocas dimensiones las proximidades existentes entre un conjunto de objetos. Además PAST permite asignar colores a las poblaciones, por lo que se han dividido en las diferentes familias lingüísticas africanas (Afro-asiática, Nilo-Sahariana, Niger-Congo, Khoi-San y Pigmeos).

PAST también permite realizar el Test de Mantel, por lo que se introducen las matrices de las distancias genéticas  $F_{ST}$  por parejas, para obtener la correlación entre las tres bases de datos principales. Para ello se seleccionan las distancias  $F_{ST}$  de poblaciones comunes o de la misma familia lingüística entre las bases de datos (Tabla1).

SNP	Gr.Sanguíneos	SNP	STR	Gr.Sanguíneos	STR
Chagga	BantuNE	San	San	BantuNE	BantuNE
Lisongo	BantuNW	Bantu	BantuSA	BantuSE	BantuSA
Bantu	BantuSE	Biaka	Biaka	San	San
San	San	Mbuti	Mbuti	Bereberes	Mozabite
Sandawe	Sandawe	Yoruba	Yoruba	PBiaka	Biaka
Masai	Masai	Chagga	BantuNE	PMbuti	Mbuti
Mozabite	Bereberes	Mandenka	Mandenka	Akan	Yoruba
Hausa	Hausa	Mozabite	Mozabite		
Biaka	PBiaka				
Mbuti	PMbuti				
Yoruba	Akan				

**Tabla 1.** Listado de poblaciones seleccionadas para realizar el análisis de correspondencias y el Test de Mantel entre SNP, grupos sanguíneos y STR.

PHYLIP (*PHYL* o *geny I nference P ackage*) es un paquete de programas para entre otras cosas inferir filogenias (árboles evolutivos). Para empezar se ha realizado para cada base de datos un *Outfile* mediante el programa Gendist, donde se obtendrán las distancias  $F_{ST}$ . A partir de este archivo y con el programa Neighbor se crea un *Outtree*. Los árboles realizados han sido neighbor-joining (NJ), que se visualizan en un formato sin raíz.

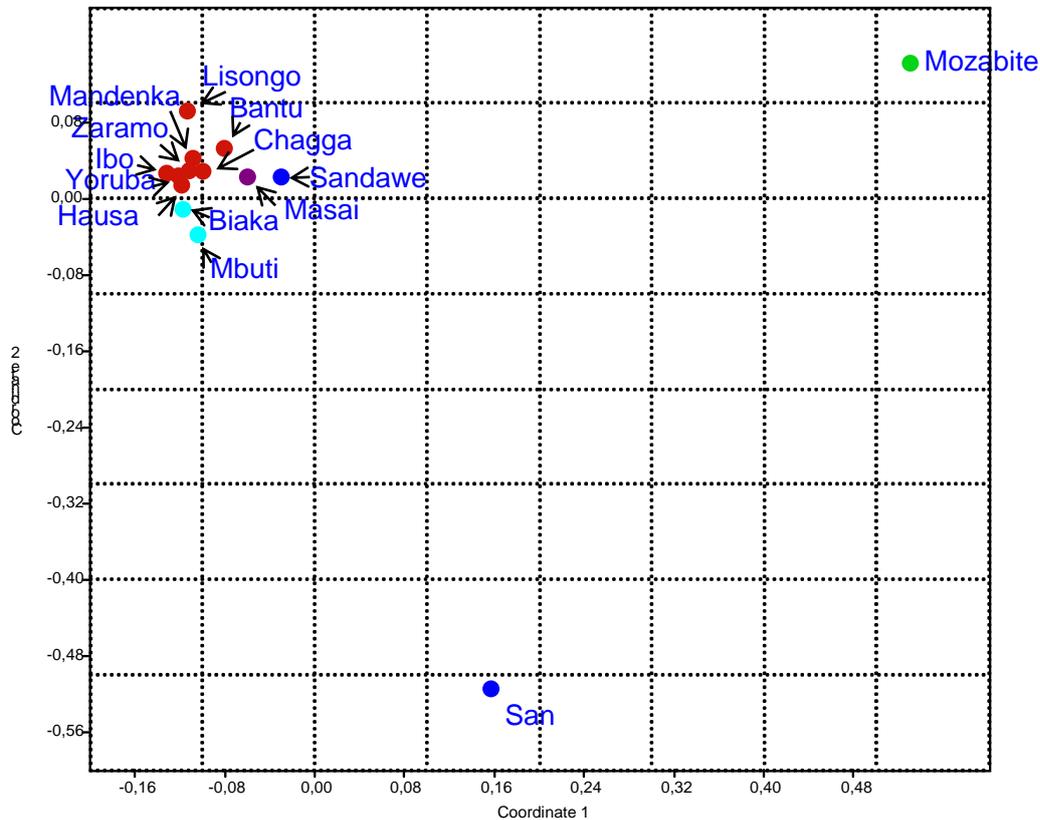
A continuación se ha realizado un bootstrap mediante el programa sepboot. Bootstrap es una técnica que permite estimar los niveles de confianza de una hipótesis filogenética, estima distribuciones muestrales a partir de repetidos *resampling* de los datos de una muestra original. Es decir, se han creado 1000 pseudomuestras del árbol filogenético original para comprobar la fiabilidad de la longitud de las ramas. Y tras utilizar de nuevo los programas Gendist y Neighbor se ha obtenido un nuevo árbol con los valores bootstrap, de los cuales han sido seleccionados los que son mayores del 50% y se han trasladados a las ramas del primer árbol filogenético que ha sido creado, ya que el árbol generado por bootstrap no da información sobre la longitud de las ramas.

Finalmente se ha utilizado el programa TreeView para visualizar los árboles filogenéticos. TreeView es un programa que proporciona una forma sencilla de ver los contenidos de un archivo con formato de árbol para la visualización de las filogenias.

## RESULTADOS

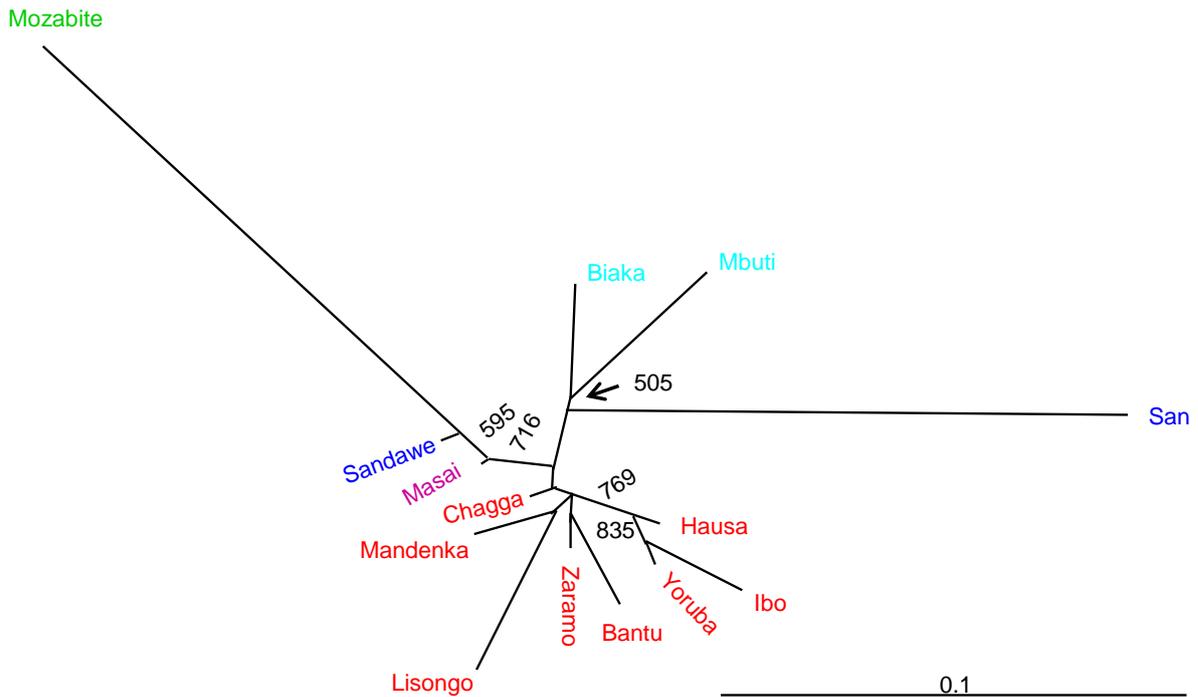
### 1. Análisis de las poblaciones africanas mediante SNPs.

En el análisis mediante 96 SNPs con 15 poblaciones africanas, se pueden observar patrones similares entre el MDS (Figura 2) y el árbol NJ (Figura 3).



**Figura 2.** MDS de las frecuencias de los SNPs. Verde: Afroasiáticos. Azul oscuro: Khoi-San. Azul claro: Pigmeos. Rojo: Niger-Congo. Morado: Nilo-Sahariano.

Se observa una diferenciación de las poblaciones Mozabite de lengua afroasiática y de la población San de lengua Khoi-San, con el resto de poblaciones. En cuanto a las demás poblaciones, se encuentran agrupadas, incluyendo pigmeos, Sandawe de lengua Khoi-San, Masai (Nilo-sahariano) y las poblaciones de la lengua Niger-Congo. No obstante, se aprecia dentro de este grupo una distribución de las poblaciones dependiendo de su familia lingüística, agrupándose por un lado las poblaciones de la familia Niger-Congo y por otro lado los pigmeos, donde Mbuti se encuentra más alejado del resto de poblaciones.

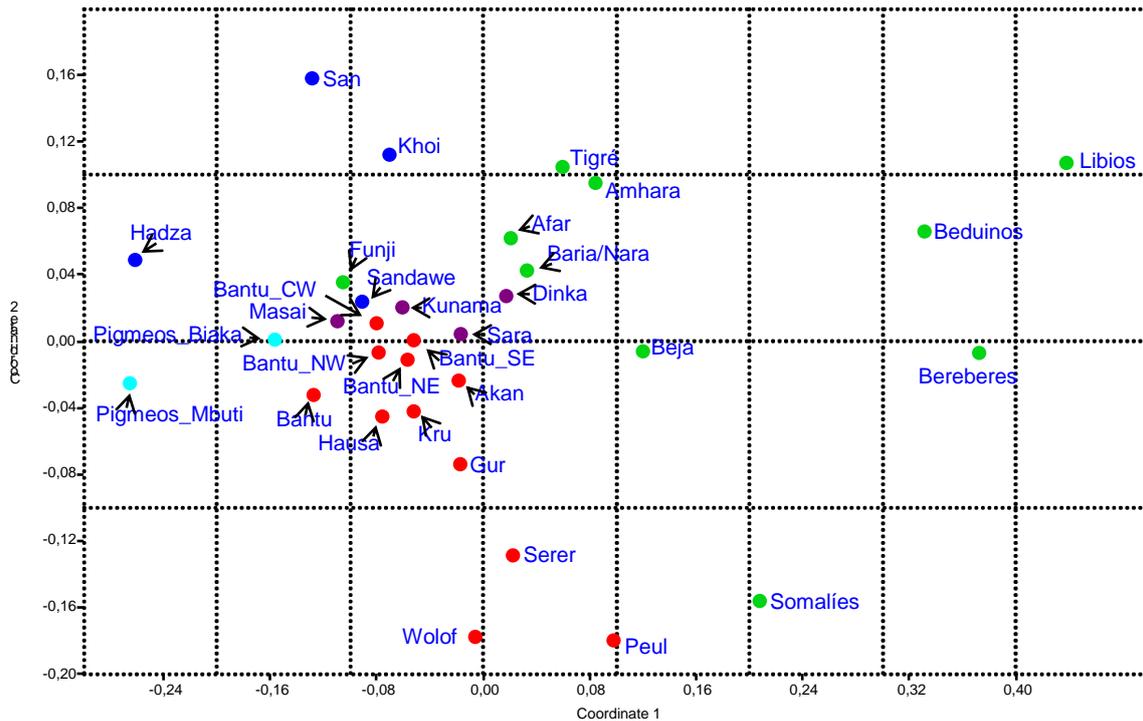


**Figura 3.** Árbol filogenético de los SNPs. Verde: Afroasiáticos. Azul oscuro: Khoi-San. Azul claro: Pigmeos. Rojo: Niger-Congo. Morado: Nilo-Sahariano.

Se observa de nuevo una agrupación de las poblaciones Niger-Congo que se diferencian del resto de poblaciones. Las poblaciones de Yoruba e Ibo se unen en la misma rama con un valor bootstrap de 83,5% y ambos se unen a los Hausa con un valor de 76,9%. Los pigmeos (Mbuti y Biaka) se unen en la misma rama con un valor bootstrap de 50,5%, y también están relacionados con los San. Por último tenemos los Masai, Sandawe y Mozabite unidos por una rama con un valor bootstrap de 71,6% y a los Sandawe y Mozabite unidos con un valor de 59,5%.

## 2. Análisis de las poblaciones africanas mediante grupos sanguíneos.

Una vez analizados los 17 alelos o haplotipos de 32 poblaciones de África, se ha observado que en el MDS (Figura 4) y en el árbol filogenético (Figura 5), las poblaciones estudiadas siguen un patrón similar de distribución.



**Figura 4.** MDS de grupos sanguíneos. Verde: Afroasiáticos. Azul oscuro: Khoi-San. Azul claro: Pigmeos. Rojo: Niger-Congo. Morado: Nilo-Sahariano.

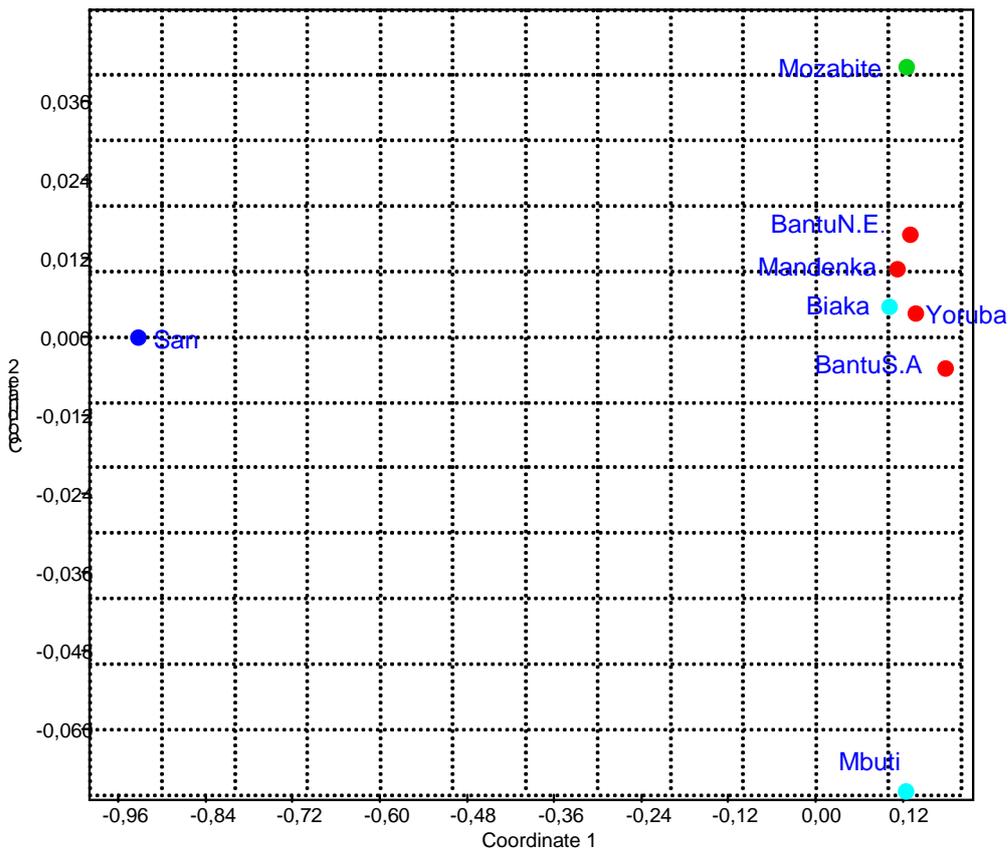
Los afroasiáticos se encuentran agrupados a un lado del gráfico, siendo las poblaciones del norte de África (Bereberes, Beduinos y Libios) las más alejadas. Las poblaciones Niger-Congo están próximas entre sí y siguen una distribución geográfica, siendo las poblaciones del oeste de África (Serer, Peul y Wolof) las que se encuentran más alejadas. También se observa una agrupación entre las poblaciones de Bantués de la familia Niger-Congo. De las cuatro poblaciones Khoi-Sanidas, Sandawe es la que más se aproxima a la familia Niger-Congo, mientras que las otras tres se encuentran algo alejadas, siendo la población San la más alejada del resto de poblaciones. Las poblaciones Nilo-Saharianas están muy próximas entre sí y se encuentran situadas entre la familia Niger-Congo y la familia afroasiática. Por último, se observa que los pigmeos están próximos entre sí y junto a la familia Niger-Congo, estando la población Mbuti ligeramente más alejada.



Las poblaciones de afroasiáticos, excepto los Funji, se encuentran agrupadas. Por un lado observamos las poblaciones de Libios y Bereberes unidos en una rama con un valor bootstrap de 51% y a los Libios, Bereberes y Beduinos con un valor bootstrap de 75,6%, todos ellos relacionados a los Somalies. Mientras que los Tigre, Amhara, Afar y Baria/Nara se encuentran en otra rama, las poblaciones Tigre y Amhara están unidas en una rama con un valor bootstrap de 50,8 %. Los pigmeos (Mbuti y Biaka) y los Hadza están relacionados con un valor bootstrap de 73%. En cuanto a la familia Khoi-Sanida, se observa que los Khoi y los San están en la misma rama. Por último las familias de Niger-Congo y Nilo-Saharanos se agrupan en el centro del árbol filogenético.

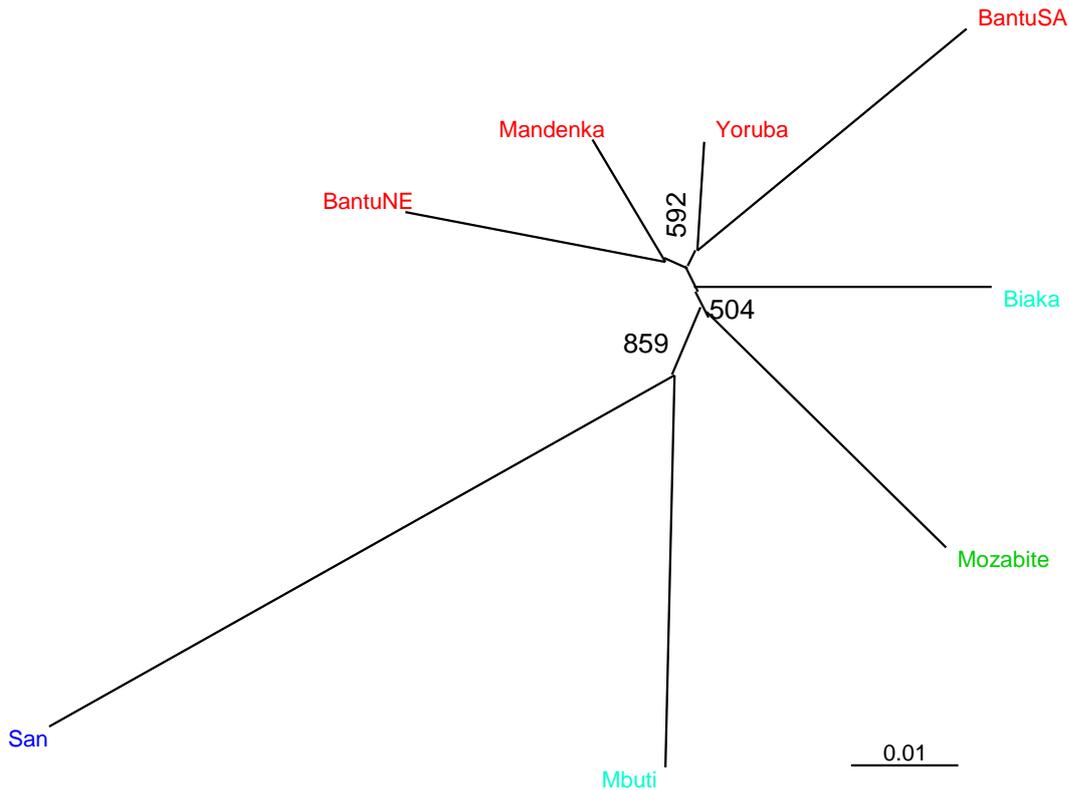
### 3. Análisis de las poblaciones africanas mediante STR.

En el análisis mediante 52 STR y 8 poblaciones, se ha observado tanto en el análisis MDS (Figura 6) como en el árbol filogenético (Figura 7) que las poblaciones San y Mbuti son las más alejadas del resto.



**Figura 6.** MDS de microsatélites. Verde: Afroasiáticos. Azul oscuro: Khoi-San. Azul claro: Pigmeos. Rojo: Niger-Congo.

Solo aparece un representante de la familia Khoi-San (los San), muy alejada del resto de las poblaciones. Y una única población afroasiática (Mozabite) que no se agrupa con las demás poblaciones. La familia Niger-Congo se encuentra agrupada y las dos poblaciones de pigmeos (Mbuti y Biaka) se encuentran muy alejadas entre sí, siendo Biaka la más próxima a la familia Niger-Congo.

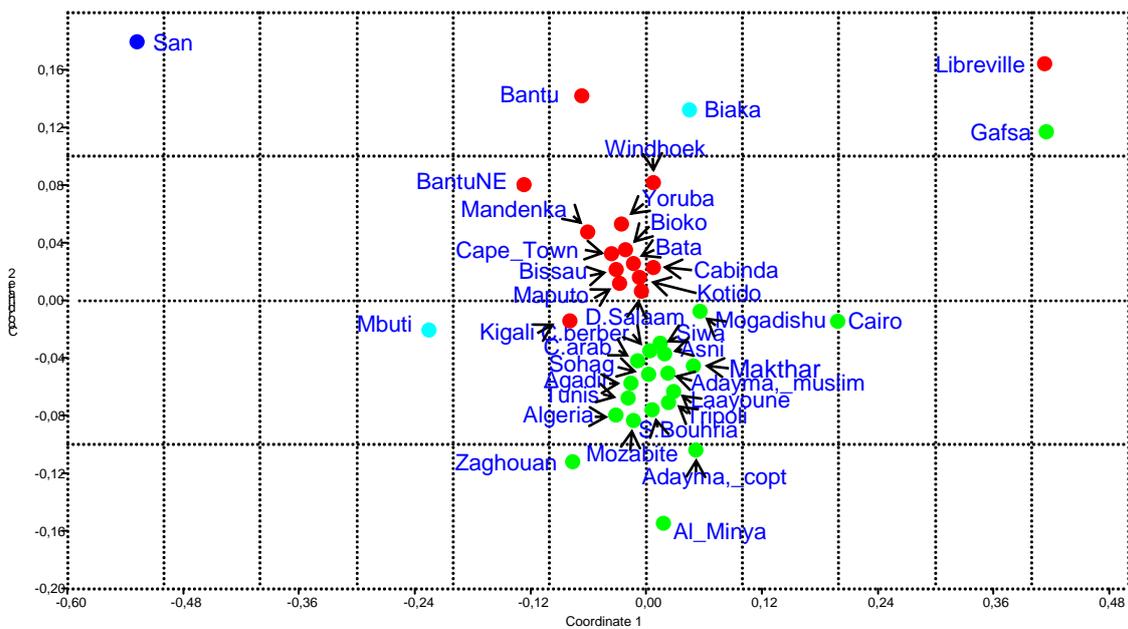


**Figura 7.** Árbol filogenético de microsatélites. Verde: Afroasiáticos. Azul oscuro: Khoi-San. Azul claro: Pigmeos. Rojo: Niger-Congo.

Las poblaciones Niger-Congo están relacionadas con los pigmeos Biaka, con un valor bootstrap de 50,4%. En cambio los pigmeos Mbuti, se encuentran agrupados a los San, con un valor bootstrap de 85,9%.

#### 4. Análisis de las poblaciones africanas mediante STR12.

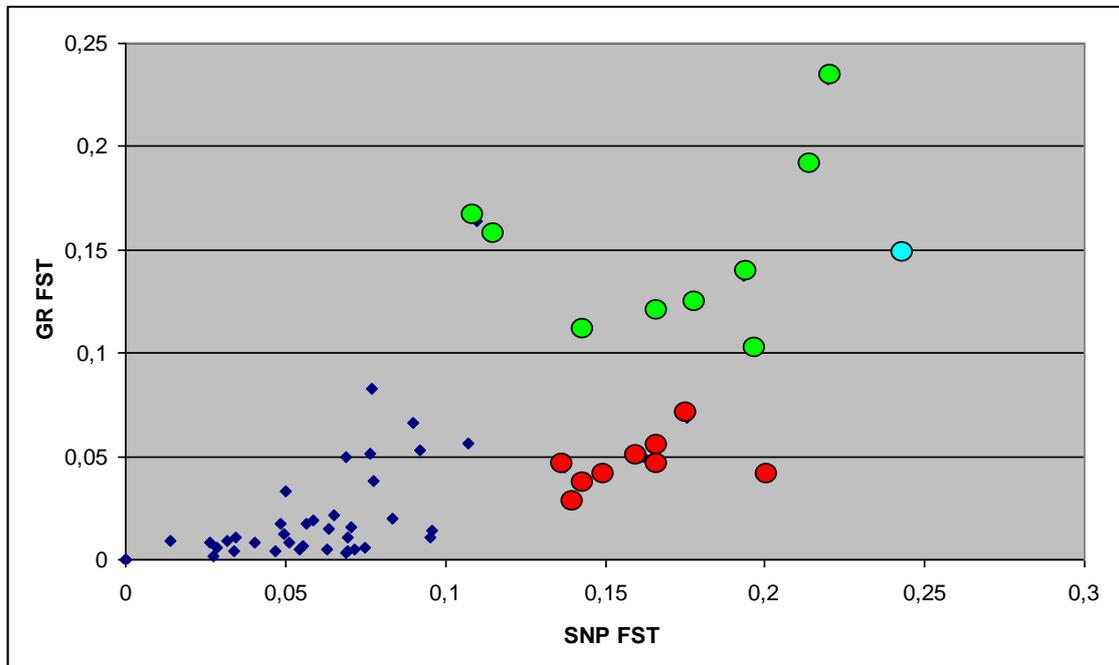
Tras el análisis de 12 STR y 25 poblaciones africanas, se ha obtenido un MDS (Figura 8) en donde se observa que la población San se encuentra muy alejada del resto de poblaciones. También aparecen dos poblaciones (Libreville y Gafsa) diferenciadas, aunque al no haber ningún dato que nos sugiera una razón para su diferenciación, posiblemente su localización sea debida a errores en el genotipado. Los pigmeos están próximos al resto de poblaciones, pero alejados entre sí. Se observa una clara diferenciación entre las familias afroasiáticas y los Niger-Congo, estando las poblaciones de cada familia, agrupadas entre sí. En la figura 8 la mayoría de las poblaciones son afroasiáticas y Niger-Congo, no se encuentran Nilo-Saharianos y únicamente dos poblaciones de pigmeos (Mbuti y Biaka) y una de Khoi-San (San).



**Figura 8.** MDS de microsatélites. Verde: Afroasiáticos. Azul oscuro: Khoi-San. Azul claro: Pigmeos. Rojo: Niger-Congo.

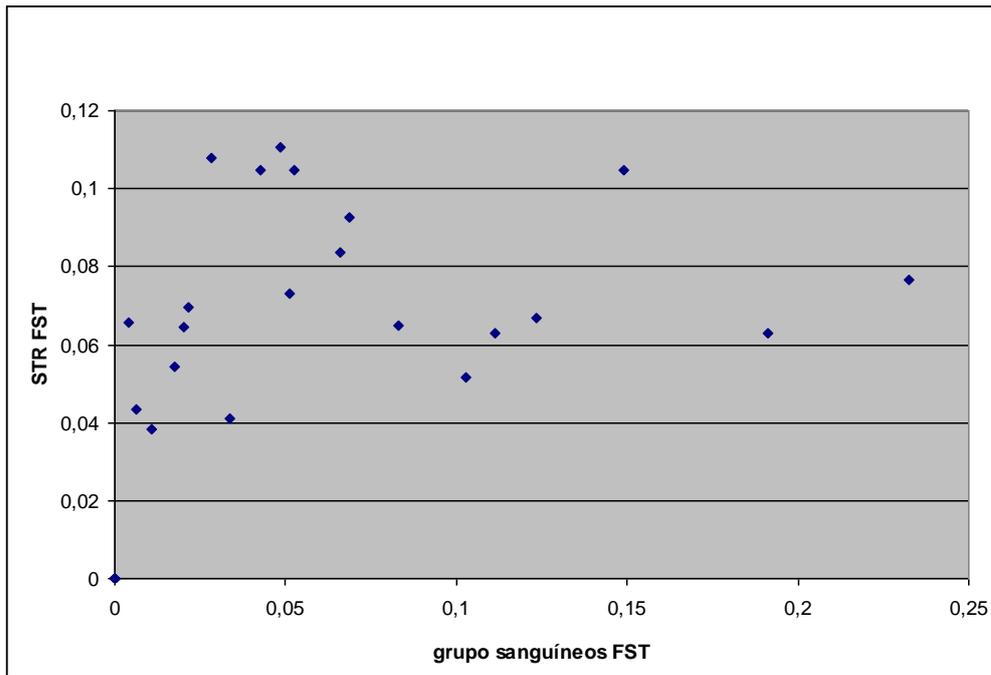
## 5. Comparación entre polimorfismos.

Se ha comparado mediante correlación y el Test de Mantel los diferentes polimorfismos entre sí.



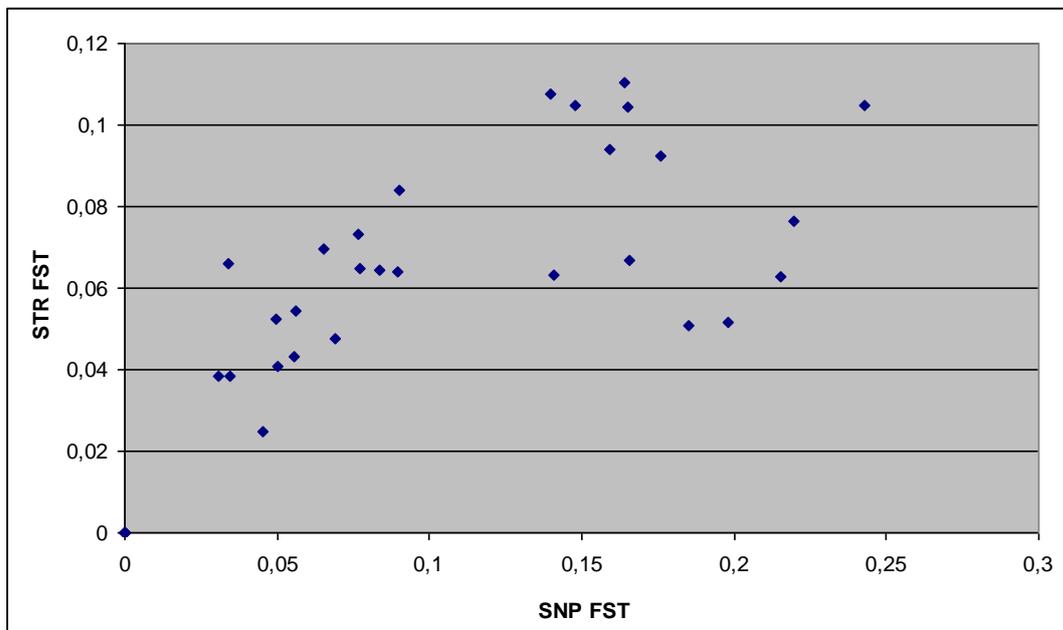
**Figura 9.** Gráfica que representa las distancias  $F_{ST}$  de SNP frente a las distancias  $F_{ST}$  de los grupos sanguíneos. En rojo están representados los puntos de la población San, en verde los puntos de la población Mozabite y en azul claro, el punto donde coinciden ambas poblaciones.

Los datos de los SNPs y grupos sanguíneos (Figura 9), tienen una alta correlación, con una R de 0.7561 y una significación de 0.003724. La población San (los puntos rojos) y la población Mozabite (puntos verdes), son las que menos correlación tienen. El punto azul, es el valor de la matriz de las distancias  $F_{ST}$  entre San y Mozabite.



**Figura 10.** Gráfica que representa las distancias  $F_{ST}$  de los grupos sanguíneos frente a las distancias  $F_{ST}$  de los microsatélites.

Las distancias entre los datos de STR comparados con los grupos sanguíneos (Figura 10), no están correlacionados ya que tienen una R muy baja ( $R=0.1301$ ) y no presentan significación ( $p= 0.2683$ ).



**Figura 11.** Gráfica que representa las distancias  $F_{ST}$  de SNP frente a las distancias  $F_{ST}$  de microsatélites.

Los datos de SNPs frente a los microsatélites, no presentan una alta correlación ( $R=0.5878$ ) ni tampoco significación ( $p=0.09664$ ).

## DISCUSIÓN

En este trabajo, con el objetivo de analizar la diversidad genética en África, se han analizado poblaciones africanas con tres tipos de marcadores. Además se ha comparado la información ofrecida por cada tipo de polimorfismo. Si analizamos en conjunto las imágenes obtenidas de todas las bases de datos, se observa en general que las poblaciones de cada familia lingüística se parecen genéticamente, como por ejemplo las familias afroasiáticas y las Niger-Congo.

La familia Khoi-San muestra una distribución parecida en todos los polimorfismos estudiados, siendo la población San la que se encuentra alejada del resto de poblaciones, excepto en los estudios realizados con grupos sanguíneos. La familia Khoe-San se separó del resto de poblaciones hace  $\geq 100,000$  años, pero la estructura de la población dentro de los Khoe-San remonta a hace unos 35.000 años (Schlebusch *et al*, 2012). Los Khoi, se encuentran en una posición intermedia entre los San y el resto de poblaciones, esto se debe a que son poblaciones ganaderas y se han mezclado más. Los Sandawe han sido estudiados tanto en SNPs como en grupos sanguíneos y presentan un gran parecido a las poblaciones de la familia Niger-Congo. Esto se debe a la proximidad geográfica de los Sandawe (Tanzania) con los Niger-Congo, que ha facilitado el intercambio de genes entre estas poblaciones. Los Hadza han sido estudiados únicamente con un polimorfismo (grupos sanguíneos) y se ha observado que presentan flujo génico con la población de pigmeos Mbuti.

La familia Niger-Congo es genéticamente más homogénea en comparación con las poblaciones africanas anteriores, probablemente por un menor aislamiento entre las distintas poblaciones de la familia Niger-Congo. También se observa en los análisis una distribución acorde con la geografía, siendo las poblaciones del oeste de África (Serer, Peul y Wolof), las que se encuentran ligeramente más alejadas. En general las muestras de los Bantúes aparecen agrupadas, probablemente por la reciente y rápida expansión de los Bantúes (de un origen común en Camerún y Nigeria) (Reed *et al*, 2006).

Los representantes de la familia Nilo-Sahariana muestran relación con las poblaciones de Niger-Congo debido a su proximidad geográfica, aunque han sido usadas pocas poblaciones de esta familia lingüística (para STR no aparecen poblaciones Nilo-saharianas, en los análisis de SNPs se encuentra una población y para grupos sanguíneos 4 poblaciones). En el análisis de grupos sanguíneos, las poblaciones Nilo-Saharianas se sitúan de manera intermedia entre algunas poblaciones de afroasiáticos (las del Este de África) y los Niger-Congo. Probablemente se han mezclado con ambas familias lingüísticas, ya que geográficamente se sitúan entre las dos.

En cuanto a los afroasiáticos, parecen estar divididos en dos grupos, las poblaciones del norte de África (Libios, Beduinos y Bereberes) se encuentran más alejadas del resto por el aislamiento geográfico del desierto del Sahara, mientras que el cruzamiento de las demás poblaciones afroasiáticas con otras familias lingüísticas ha sido más probable. En concreto se ha observado que en los análisis con grupos sanguíneos y SNPs los afroasiáticos se encuentran muy alejados genéticamente del resto, mientras que con STR no tanto. Este resultado de que con los grupos sanguíneos y los SNPs obtengamos aparentemente una información distinta de la de los STRs, tiene relación con lo observado en pigmeos (Mbuti), donde con STR se difiere más. Es decir, la situación contraria. En un estudio realizado con ADNmt, se concluye que la variabilidad genética de las poblaciones afroasiáticas en el Este de África, aunque esté fuertemente influenciada por los procesos migratorios, conserva vestigios de los estratos más antiguos (Boattini *et al*, 2013).

Los pigmeos presentan proximidad genética a la familia Niger-Congo. Las dos poblaciones estudiadas, Mbuti y Biaka, se encuentran relacionadas al estudiar las frecuencias de SNP y grupos sanguíneos, pero no con los microsatélites.

Para intentar comprender todo esto, se ha hecho un estudio de correlación y Test de Mantel, y se ha observado que los dos polimorfismos que son codificantes, es decir los SNP y los grupos sanguíneos, están correlacionados. La selección natural puede haber homogeneizado ambos polimorfismos, siendo neutral en los microsatélites. Mientras que los grupos sanguíneos con STR y SNPs con STR, no presentan correlación.

En el caso de los afroasiáticos, con marcadores neutrales no se diferencian mucho, pero con los polimorfismos codificantes sí. La selección natural puede ser la responsable de la diferenciación observada en estas poblaciones afroasiáticas, ya que el clima en el Norte de África es distinto a la de la mayor parte del resto de África. Es decir, la selección natural ha aumentado el efecto diferenciador en aquellos polimorfismos sobre los que actúa.

En el caso de los Mbuti ocurre lo contrario, es decir, se diferencia con marcadores neutrales (STR), lo que indica un origen antiguo y un aislamiento muy marcado. Pero se asemejan a las poblaciones circundantes con los marcadores codificantes (grupos sanguíneos y SNPs). En este caso, al presentar las mismas condiciones ambientales, la selección natural ha asemejado sus frecuencias alélicas.

Los San se diferencian del resto de poblaciones al analizar los SNPs y STR, pero no con los grupos sanguíneos. En este caso, la selección natural solo parece haber homogeneizado las frecuencias alélicas con las poblaciones circundantes de los genes responsables de los grupos sanguíneos.

## **REFERENCIAS**

Boattini A, Castrì L, Sarno S, Useli A, Cioffi M, Sazzini M, Garagnani P, De Fanti S, Pettener D, Luiselli D. mtDNA variation in East Africa unravels the History of Afro-Asiatic groups. *Amer Journal of Physic Anthro* 2013;150:375–385.

Campbell MC, Tishkoff SA. African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. *Annu Rev Genomics Hum Genet* 2008;9:403–433.

Garrigan D, Kingan SB, Pilkington MM, Wilder JA, Cox MP, Soodyall H, Strassmann B, Destro-Bisol G, de Knijff P, Novelletto A, Friedlaender J, Hammer MF. Inferring human population sizes, divergence times and rates of gene flow from mitochondrial, X and Y chromosome resequencing data. *Genetics* 2007;177:2195–2207.

Gunz P, Bookstein FL, Mitteroecker P, Stadlmayr A, Seidler H, Weber GW. Early modern human diversity suggests subdivided population structure and a complex out-of-Africa scenario. *Proc Natl Acad Sci U S A* 2009;106:6094–6098.

Liu H, Prugnolle F, Manica A, Balloux F. A geographically explicit genetic model of worldwide human-settlement history. *Am J Hum Genet* 2006;79:230–237.

Peña JA, Alfonso MA, Pérez AM, García S, Gómez L. GeDis: Un programa para análisis de datos en Antropogenética. *Antropo* 2009;20, 49-56.

Pilkington MM, Wilder JA, Mendez FL, Cox MP, Woerner A, Angui T, Kingan S, Mobasher Z, Batini C, Destro-Bisol G, Soodyall H, Strassmann BI, Hammer MF. Contrasting signatures of population growth for mitochondrial DNA and Y chromosomes among human populations in Africa. *Mol Biol Evol* 2008;25:517–525.

Reed FA, Tishkoff SA. African human Diversity, origins and migrations. *Curr Opin Genet Dev* 2006;16:597-605.

Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, Lavery T, Kouyoumjian R, Farhadian SF, Ward R, Lander ES. Linkage disequilibrium in the human genome. *Nature* 2001;411:199–204.

Schlebusch CM, Skoglund P, Sjödin P, Gattepaille LM, Hernandez D, Jay F, Li S, De Jongh M, Singleton A, Blum MJB, Soodyall H, Jakobsson M. Genomic variation in seven Khoe-San groups reveals adaptation and complex African History. *Science* 2012;338:374–379.

Tishkoff SA, Reed FA, Friedlaender FR, Ehret C, Ranciaro A, Froment A, Hirbo JB, Awomoyi AA, Bodo JM, Doumbo O, Ibrahim M, Juma AT, Kotze MJ, Lema G, Moore JH, Mortensen H, Nyambo TB, Omar SA, Powell K, Pretorius GS, Smith MW, Thera MA, Wambebe C, Weber JL, Williams SM. The genetic structure and history of Africans and African Americans. *Science* 2009;324:1035–1044.

## **Anexo 1.** Listado de SNPs y microsatélites.

### SNPs

Alcohol Dehydrogenasa 1B	exon 3 Arg 48 His C_2688471 C_2688455_1 Intron 3_Rse I Rs 1229982
Dopamina Receptor D2	C_11339234_20 HGV base SNP 000002593 Intron 1 T/C Rs 4581480
CD4 antigen (pSS)	rs 1055141 Rs7957426
Solute carrier family 6	exon 9 Dde I (A13436)
Proto-oncogene	C_3204345_10
Acyl_coenzyme Adehydronesace C_4	rs 1146581
Nerve growth factor receptor	rs 2119930 rs 2671666
LIM domain containg preferred translocatin portner in lipona	rs 1513181
CUB and sushi miltiple domain 1	rs 10108220
Solute carrier family 44 member 5	rs 6593584
Formin_like 1	rs 1552458
Tetratricopeptido repeat domain 21 A	rs 9809104
Keratin 26	rs 975078
Park 2_co_regulated	rs 4458655
Estadysplesim I receptor	C_1645496_10
Human inmunodeficieng virus tipe I	rs 1325502
SPHK1 interacting protein	rs 1343845
F_Box and lucine-rich repeat protein 17	rs 988436
Catenin $\alpha$ 2	rs 134000937

Neurobeadin	rs 7997709
Serina/threonine kinasa 32 B	rs 9291090
Between LOC100419685 & LOC100132132	rs 723021
Between OCM & LMTK2	rs 705308
Thioredoxin reductase 2	rs 1012157
Aldehyde dehydrogenase 2 family	C_340538_10
Homo box B1	rs 12510722
$\alpha$ -polypeptide 5 chipolypeptide	rs 1230155
$\alpha$ -polypeptide 6 chipolypeptide	rs 2051428
Hemoglobin $\beta$ pseudogen 1	C_9599122_16 rs 10837628
Gastric inhibitory polypeptide	rs 11671664 rs 4802262
Taste receptor type 2	C_9506827_10
VP510 domain receptor protein SORCC3	rs 1404784
Microtubule-associated protein tan	rs 16940742
Prohibitin	rs 1049620 rs 17637472
Synoptogyrin 2	rs 2854701
Echinoderm microtubule associated protein	rs 2286756
Radial spokehead-like 1	rs 1548029
Lymphocyte-activation gene 3	rs 870849
Hemoglobin gamma 6	rs 2855039
5rC family associated phosphoprotein	rs 11079811
Zinc finger protein, multi type 2	C_16041659_10
Tumor necrosis factor	C_609154_10
Sarcoglycon delta	C_11158875_10

Ca <sup>2+</sup> dependn activator	rs 2501440
Phenylamina Hydroxylasa	rs 10860936 rs 4764916
Alcohol Dehydrogenase 7 class IV	C_119423061 exon 6 Hinf I rs 1348276
Solute carrier family 6	C_1841702_10
Dopamina Receptor D3	rs 167771
Angiotensinogen	rs699
Lactasa	C_15974400_10
Dopamina β-hydroxylasa	C_2253949_10
Neural cell adhesion molecule I	rs 2117912
Hypothetial LOC 388553	rs 4802274
Similar to Histone H2B291B	rs 10934250
Tetratricopeptide repeat domain 12	rs 17115388
KIM 1267	rs 1078997
Tubulin tyrosina ligasa member 6	rs 2032844
Intergenic between PPLAP17 & KIF16D	rs 12480506
Similar to MUF1 protein	C_8240638_10
D1P2 disco-interacting protein 2homologB	C_470465_1
Intergenic between RASGEF1A & FXYD4	rs 2243492
Queunine tRNA-riboxy 1 transferase	rs 6782373
Olfactory receptor family 52	rs 1378744
GRAM domain containing 1C	rs 4422272
Triportite motif_containing 5	rs 11038628
Dysbidin domain containing 1	rs 11648785
Intergenic between MAP3K14 & ARHGAP27	rs 4792814
Symplelin	rs 10500292

Chromodomain protein, Y-like	rs 1040045
Spleen tyrosine kinasa	rs 2306040
Astrotactin 2	rs 10513300
F-box protein 46	rs 11668847
Intergenic between LOC100652994 & LINC00161	C_2528441_20
Intergenic between LOC100422456 & TAS2R16	rs 10260099
Adenylasa kinasa domain containing 1	C_8705703_10
Roundabout exon guidance receptor	rs 6548616
Intergenic between LOC100130572 & SEMA3A	rs 10954737
Nucleosoma assembly protein 1-like pseudogene 3	rs 12629908
TBCI domain family, member 4	rs 9530435
Intergenic between GAPDHP36 & LOC131054	rs 2030763
ATP- binding cassette, subfamily A	rs 957718
Leucine rich repeat and Ig domain containing 2	rs 10511828

### STR

CSF1PO	D1S1656	D2S441	D2S1338	D3S1358	D5S818
D7S820	D8S1179	D10S1248	D12S391	D13S317	D16S539
D18S51	D19S433	D21S11	D22S1045	FGA	TH01
TPOX	VWA	SE33	Penta_D	Penta_E	D6S1043
D2S1360	D3S1744	D4S2366	D5S2500	D6S474	D7S1517
D8S1132	D10S2325	D21S2055	LPL	F13B	FESFPS
F13A01	Penta_B	Penta_C	D1S1679	D2S427	D3S2406
D3S4545	D5S1457	D7S2201	D9S1118	D11S1304	D12S297
D14S1426	D15S822	D21S102	D9S1120		

**Anexo 2.** Tabla de poblaciones estudiadas con cada polimorfismo con su país y familia lingüística correspondiente.

SPN			STR			STR12			Grupos sanguíneos		
Población	País	Lengua	Población	País	Lengua	Población	País	Lengua	Población	País	Lengua
San	Namibia	Khoisan	Pigmeos Biaka	Rep. Centroafricana	Pigmeo	Algeria	Argelia	Afroasiática	Bantu CE	Tanzania	Niger-Congo
Bantu	Congo	Niger-Congo	Pigmeos Mbuti	Congo	Pigmeo	Adayma, muslim	Egipto	Afroasiática	Bantu C'w	Congo	Niger-Congo
Biaka	Congo	Pigmeo	Bantu N.E.	Kenya	Niger-Congo	Adayma, copt	Egipto	Afroasiática	Bantu NE	Kenia	Niger-Congo
Hausa	Nigeria	Niger-Congo	San	Namibia	Khoisan	Al Minya	Egipto	Afroasiática	Bantu NW	Guinea	Niger-Congo
Ibo	Nigeria	Niger-Congo	Yoruba	Nigeria	Niger-Congo	Cairo	Egipto	Afroasiática	Bantu SE	Sudáfrica	Niger-Congo
Lisongo	Rep. Centroafricana	Niger-Congo	Mandenka	Senegal	Niger-Congo	Siwa	Egipto	Afroasiática	Beduinos	Argelia	Afroasiática
Mbuti	Congo	Pigmeo	Bantu	Sudáfrica	Niger-Congo	Sohag	Egipto	Afroasiática	Amhara	Etiopía	Afroasiática
Yoruba	Nigeria	Niger-Congo	Mozabite	Argelia	Afroasiática	Tripoli	Libia	Afroasiática	Beja	Sudán	Afroasiática
Chagga	Tanzania	Niger-Congo				Agadir	Marruecos	Afroasiática	Afar	Etiopía	Afroasiática
Masai	Tanzania	Nilo-Saharan				Asni	Marruecos	Afroasiática	Somalíes	Somalia	Afroasiática
Sandawe	Tanzania	Khoisan				Casablanca, arab	Marruecos	Afroasiática	Tigré	Eritrea	Afroasiática
Zaramo	Tanzania	Niger-Congo				Casablanca, berber	Marruecos	Afroasiática	Hadza	Tanzania	Khoisan
Mandenka	Senegal	Niger-Congo				Sidi Bouhria	Marruecos	Afroasiática	Khoi	Namibia	Khoisan
Mozabite	Argelia	Afroasiática				Gafsa	Tunez	Afroasiática	San	Botswana	Khoisan
						Makthar	Tunez	Afroasiática	Sandawe	Tanzania	Khoisan
						Tunis	Tunez	Afroasiática	Baia/Mara	Eritrea	Nilo-Saharan
						Zaghuan	Tunez	Afroasiática	Funji	Sudán	Nilo-Saharan
						Laayoune	Marruecos	Afroasiática	Kunama	Eritrea	Nilo-Saharan
						Cabinda	Congo	Niger-Congo	Masai	Kenia	Nilo-Saharan
						Bata	Guinea	Niger-Congo	Sara	Chas	Nilo-Saharan
						Bioko	Camerún	Niger-Congo	Dinka	Sudán	Nilo-Saharan
						Libreville	Gabón	Niger-Congo	Bereberes	Argelia	Afroasiática
						Bissau	Guinea-Bissau	Niger-Congo	Hausa	Nigeria	Afroasiática
						Maputo	Mozambique	Niger-Congo	Libios	Libia	Afroasiática
						Windhoek	Namibia	Niger-Congo	Pigmeos Biaka	Congo	Pigmeo
						Kigali	Ruanda	Niger-Congo	Pigmeos Mbuti	Congo	Pigmeo
						Mogadishu	Somalia	Afroasiática	Gur	Burkina Faso	Niger-Congo
						Cape Town	Sudáfrica	Niger-Congo	Kru	Liberia	Niger-Congo
						Dares Salaam	Tanzania	Niger-Congo	wolof	Senegal	Niger-Congo
						Kotido	Uganda	Niger-Congo	Peul	Guinea	Niger-Congo
						Biaka	Congo	Pigmeo	Serer	Senegal	Niger-Congo
						Mbuti	Congo	Pigmeo	Akan	Ghana	Niger-Congo
						Bantu NE	Kenya	Niger-Congo			
						San	Namibia	Khoisan			
						Yoruba	Nigeria	Niger-Congo			
						Mandenka	Senegal	Niger-Congo			
						Bantu	Sudáfrica	Niger-Congo			
						Mozabite	Argelia	Afroasiática			