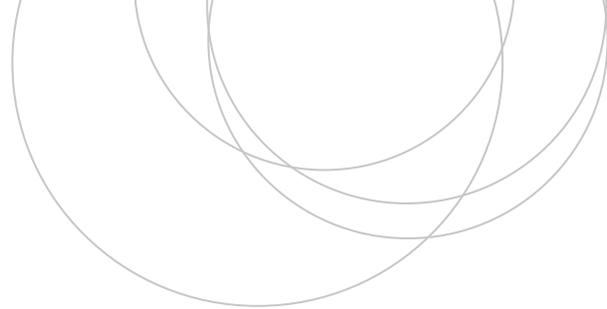




Universidad
del País Vasco

Euskal Herriko
Unibertsitatea

ZIENTZIA
ETA TEKNOLOGIA
FAKULTATEA
FACULTAD
DE CIENCIA
Y TECNOLOGÍA



Trabajo Fin de Grado
Grado en Biología

Variabilidad genética en poblaciones humanas: efecto de la selección natural en los genes del cromosoma 9

Autora:
Naiara Villarías González

Director:
Mikel Iriondo Orensanz

ÍNDICE

INTRODUCCIÓN.....	3
HIPÓTESIS Y OBJETIVOS.....	5
MATERIALES Y MÉTODOS	6
RESULTADOS.....	8
DISCUSIÓN	11
GEN GLIS3.....	12
GEN SMARCA2.....	13
CONCLUSIONES	17
BIBLIOGRAFÍA	17
ANEXO	22

RESUMEN

Mediante este estudio se ha querido analizar el efecto de la selección natural sobre los genes del cromosoma 9. Para ello, se han obtenido los datos de 129 SNPs partiendo de los genotipos de 2504 individuos pertenecientes a 26 poblaciones. Mediante el cálculo de la F_{ST} se puede distinguir entre la acción de la selección natural y la acción de otros procesos neutrales como deriva génica y flujo génico, los cuales afectan a todo el genoma por igual.

Tras seleccionar aquellos SNPs con una F_{ST} mayor que la media, se han escogido dos de ellos para explicar la distribución de las frecuencias en las diferentes poblaciones. Cambios en el gen GLIS3 se ha observado que pueden causar diabetes, por lo que es posible que la distribución de las frecuencias se deba a adaptaciones a la escasez de recursos. El gen SMARCA2, por su parte, interactúa con C/EBP- β , que participa en la acción de la respuesta inmune. Por ello, las frecuencias observadas pueden deberse a una adaptación a la resistencia a microorganismos.

PALABRAS CLAVE

Cromosoma 9, selección natural, SNP, GLIS3, diabetes, SMARCA2, inmunidad

ABSTRACT

Through this study, we wanted to analyze the effect of natural selection on the genes of chromosome 9. For this purpose, we have obtained the data of 129 SNPs using the genotypes of 2504 individuals belonging to 26 populations. By calculating the F_{ST} , it is possible to distinguish between the action of natural selection and the action of other neutral processes such as gene drift and gene flow, which affect the entire genome equally.

After selecting those SNPs with a F_{ST} greater than the average, two of them have been selected to explain the distribution of frequencies in the different populations. Changes in the GLIS3 gene have been observed to be a cause of diabetes, so it is possible that the frequency distribution is due to adaptations to the scarcity of resources. The SMARCA2 gene, for its part, interacts with C/EBP- β , which participates in the action of the immune response. Therefore, the observed frequencies may be due to an adaptation to resistance to microorganisms.

KEY WORDS

Chromosome 9, natural selection, SNP, GLIS3, diabetes, SMARCA2, immunity

INTRODUCCIÓN

El linaje humano moderno surgió en África hace 300 ka (De Castro y Martín-Torres, 2022), y desde entonces se ha expandido por todo el globo. Hace 170.000 años llegaron a Próximo Oriente (Hershkovitz *et al.*, 2018), y, más tarde, hace 80.000 años, se dirigieron hacia el sudeste de Asia siguiendo la ruta de la costa (Liu *et al.*, 2015). 15.000 años más tarde, llegaron a Australia y se expandieron por Nueva Guinea y las islas más próximas (Clarkson *et al.*, 2017). Partiendo del centro de Asia, hace 54.000 años llegaron a Europa (Metz *et al.*, 2023). En cuanto a la colonización de América, individuos procedentes de Siberia atravesaron el estrecho de Bering hace entre 15.000 y 20.000 años (Achilli *et al.*, 2013). Esta expansión por los diferentes continentes, dio lugar a una gran variabilidad. Dicha diversidad se ha debido en gran medida a los mecanismos microevolutivos: mutación, deriva génica, flujo génico y selección natural.

Si bien el ADN es una molécula altamente estable que se replica con una gran exactitud, pueden surgir cambios en su estructura y errores en la replicación. Una mutación se define como un cambio heredable (ya sean células u otros organismos) en la secuencia de la información genética del ADN. Las mutaciones son importantes porque la capacidad de los organismos para adaptarse a los cambios ambientales depende de la presencia de variación genética (producida precisamente por mutaciones). Desde el punto de vista evolutivo, se pueden clasificar en distintos tipos: deletéreas (siendo algunas de ellas letales), neutrales o ventajosas. El efecto de la mutación se da a lo largo de todo el genoma (Jiménez y Espino, 2012; Pierce, 2023).

Según a qué se deban las mutaciones, pueden dar lugar a distintos tipos de polimorfismos. Los polimorfismos de ADN generalmente se definen como variantes naturales para las cuales el alelo más común tiene una frecuencia de no más del 99%. La gran mayoría de los polimorfismos del ADN humano se pueden dividir en dos grupos: los que se basan en sustituciones de nucleótidos (comúnmente llamados "SNP", del inglés *single-nucleotide polymorphism*) y los que se basan en la inserción o eliminación de uno o más nucleótidos (indels). Los SNPs son numerosos y se encuentran por todo el genoma, en promedio uno cada 1 a 2kb. Cabe destacar que la mayoría de los SNP presentes en una población se originaron a partir de una única mutación ocurrida en un cromosoma en particular en un individuo y se esparcieron al resto de la población (Pierce, 2023; Weber *et al.*, 2002).

Otra fuerza evolutiva que afecta a las frecuencias génicas poblacionales es la deriva génica, que se define como el cambio en las frecuencias génicas debido al muestreo aleatorio de gametos que se produce en la reproducción en poblaciones de censo finito. Cuanto más pequeña es la muestra, mayor es la probabilidad de que su composición se desvíe de la del

acervo genético parental. Las causas pueden ser varias: mantener un tamaño reducido durante varias generaciones, que se produzca un efecto fundador o que se dé un cuello de botella. La deriva génica produce un cambio en las frecuencias alélicas dentro de una población, así como la reducción de la variación genética dentro de las poblaciones, dándose su efecto a lo largo de todo el genoma (Jiménez y Espino, 2012; Pierce, 2023).

Otro proceso que puede causar cambios en las frecuencias alélicas es la entrada de alelos provenientes de otras poblaciones, denominado en general migración o flujo génico. Esto se debe a que, generalmente, una población no constituye una única unidad panmíctica, sino que se encuentra subdividida en subpoblaciones. La migración tiene dos efectos importantes: aumenta la similitud entre los acervos genéticos de las diferentes poblaciones y suma variación genética a las poblaciones receptoras. Igual que en el caso anterior, su efecto se da a lo largo de todo el genoma (Fontdevila y Moya, 2000; Pierce, 2023).

Finalmente, el último mecanismo evolutivo, y en el que se va a profundizar en este estudio, es la selección natural. Mientras que los otros mecanismos microevolutivos afectan a todo el genoma, la selección actúa localmente y está restringida a regiones genómicas particulares (Quintana-Murci, 2016).

La selección natural tiene lugar cuando los individuos con rasgos adaptativos producen un mayor número de descendientes que los que no son portadores de dichos rasgos. Cuando estos rasgos tienen una base genética, aparecen en mayor frecuencia en la siguiente generación, lo que hace que las poblaciones se adapten mejor a sus ambientes (Pierce, 2023).

Desde el punto de vista de una población, la selección natural se puede manifestar de diferentes formas. La *selección purificadora*, o *selección negativa*, se refiere al proceso por el cual las mutaciones deletéreas son eliminadas de la población. Esta selección puede ser direccional en contra del alelo recesivo o en contra del alelo dominante. La selección también puede ocurrir cuando una mutación nueva es favorable, denominada como *selección positiva*, que se cree que es una de las formas por las que se da la evolución adaptativa. Asimismo, puede darse la *selección balanceadora* (también llamada *sobredominancia*), en la que el heterocigoto tiene una aptitud más alta que la de los dos homocigotos, por lo que ambos alelos son favorecidos, y ninguno es eliminado de la población. Finalmente, la *selección diversificadora* o *infradominancia* es en la que el heterocigoto tiene menor aptitud que los dos homocigotos. La infradominancia induce un equilibrio inestable. En este caso, las frecuencias alélicas no cambian en tanto se mantenga el equilibrio, pero si alguna otra fuerza evolutiva altera el punto de equilibrio, se alejarán del equilibrio hasta que finalmente un alelo alcance la fijación (Pierce, 2023; Quintana-Murci, 2016).

Desde el punto de vista de varias poblaciones, la selección natural se puede diferenciar como selección diversificadora o selección balanceadora. La selección natural diversificadora se refiere al hecho de que en dos subpoblaciones de una misma especie que viven en diferentes ambientes se seleccionan diferentes alelos para un mismo locus. La selección balanceadora, sin embargo, ocurre cuando la selección natural actúa de igual forma en ambas subpoblaciones, dando lugar a frecuencias similares (Masimirembwa *et al.*, 2014).

Distinguir entre la selección natural y los procesos neutrales, como la deriva génica y el flujo de génico, es fundamental para este tipo de estudios. En los últimos años, para probar la acción de la selección natural sobre rasgos cuantitativos en poblaciones subdivididas, se han utilizado comparaciones de dos estadísticos: F_{ST} y Q_{ST} . F_{ST} o índice de fijación es una medida de la diferenciación de la población definida como la relación entre comunidad reproductiva y la varianza total del alelo (Miller *et al.*, 2008).

Uno de los objetivos centrales de la biología es comprender la relación entre genotipo y fenotipo. La secuencia del genoma humano de referencia (obtenida en 2004) proporciona una base para el estudio de la genética humana, pero la investigación sistemática de la variación humana requiere un conocimiento más profundo (The 1000 Genomes Project Consortium, 2010).

Los avances en la tecnología de secuenciación de ADN han permitido la secuenciación de genomas individuales, iluminando los vacíos en la primera generación de bases de datos (como el Proyecto Internacional HapMap) que contienen en su mayoría sitios de variantes comunes. De esta forma, el objetivo del Proyecto 1000 Genomas es descubrir, genotipar y proporcionar información precisa sobre haplotipos en todas las formas de polimorfismo del ADN humano en múltiples poblaciones humanas (The 1000 Genomes Project Consortium, 2010).

HIPÓTESIS Y OBJETIVOS

Se parte de la hipótesis de que, aunque las fuerzas evolutivas que actúan sobre todo el genoma (mutación, deriva génica y flujo génico) han diferenciado a las poblaciones humanas en cierta medida, la selección natural habría diferenciado algunos genes en mayor medida debido a diferencias en el efecto ambiental de distintas regiones.

El objetivo de este trabajo es detectar la selección natural sobre los genes del cromosoma 9 en base a datos de SNPs. Para ello se utilizará la base de datos 1000 genomas de la que se obtendrán los genotipos. Asimismo, mediante el programa BayeScan se detectará la acción de

la selección natural sobre los polimorfismos, y, en varios de ellos, se tratará de explicar cómo ha actuado.

MATERIALES Y MÉTODOS

En primer lugar, para la obtención de los datos, se ha utilizado el buscador del NCBI para acceder a la base de datos de 1000 genomas. En este buscador aparecen los resultados del proyecto 1000 genomas, que buscaba proporcionar una descripción de la variación genética humana mediante la aplicación de la secuenciación del genoma completo a un conjunto de individuos de múltiples poblaciones (The 1000 Genomes Project Consortium, 2015). Desde esta base de datos se puede ver la información acerca de los diferentes SNPs del genoma humano en las distintas poblaciones (Anexo).

Se han escogido aquellos pertenecientes al cromosoma 9 que cumplieran unos requisitos: 1) tener una MAF (frecuencia del alelo menos común, *Minor Allele Frequency* en inglés) mínima del 0,4 para obtener unos resultados robustos y 2) estar a una distancia de 1cM (en nuestra especie aproximadamente $7.5 \cdot 10^5$ pb; Lodish *et al.*, 2012) para que los SNPs sean independientes. Se ha descargado el genotipo de los SNPs seleccionados, y se han elaborado dos documentos con ellos. En uno de ellos (formato PED) se ha incluido la información de los individuos, así como los alelos obtenidos en cada SNP. En el otro documento (formato INFO) se ha incluido el nombre de cada polimorfismo y su ubicación en el cromosoma.

Posteriormente, para comprobar si los SNPs elegidos estaban ligados o no, se ha utilizado el programa Haploview, que proporciona un cálculo de estadísticas de desequilibrio de ligamiento a partir de datos de genotipos (al igual que otra información como patrones de haplotipos) (Barrett *et al.*, 2004). Los resultados se muestran en un gráfico triangular con cuadrados coloreados según la naturaleza del resultado (Tabla 1).

Tabla 1. Esquema de color estándar que utiliza el programa Haploview.

	$D' < 1$	$D' = 1$
$LOD < 2$	Blanco	Azul
$LOD \geq 2$	Rosa/Rojo claro	Rojo intenso

Después, para ver la agrupación y cercanía de los diferentes individuos y poblaciones considerando estos polimorfismos, se ha querido hacer un gráfico “Structure”. En primer lugar, se ha usado el programa PGDSpider, un programa Java que puede leer 27 formatos de archivos y exportar los datos a otros 29. Básicamente, permite convertir archivos entre diferentes programas de genética de poblaciones (Lischer y Excoffier, 2011). De esta forma, se ha podido

obtener un documento con el formato adecuado para el análisis. Con ello, se puede realizar un proyecto en el programa Structure, que utiliza los Métodos de Montecarlo basados en cadenas de Markov (MCMC) para clasificar los individuos en grupos a partir de datos de genotipo multilocus. Básicamente, asume que todo el material genético de los individuos proviene de uno o más K (número de agrupaciones) grupos no observados (Falush *et al.*, 2007). Se han utilizado los parámetros de 5000 *Length of Burnin Period* y 5000 *Number of MCMC Reps after Burnin*. Además, se han especificado las gráficas desde $K=1$ hasta $K=8$, con 10 iteraciones cada una.

A continuación, con el objetivo de observar cuál es la agrupación (valor de K en el gráfico Structure) más adecuada, se ha utilizado el programa Structure Harvester. Éste recopila los resultados generados por el programa Structure y evalúa los valores de probabilidad en múltiples valores de K y cientos de iteraciones para detectar la cantidad de grupos genéticos que se ajustan mejor a los datos (Earl y vonHoldt, 2012). Asimismo, hay que tener en cuenta que se habían realizado 10 iteraciones, con lo que se tienen 10 gráficos diferentes para cada valor de K debido a 2 fenómenos principales: “cambio de etiquetas” (los grupos que infiere el programa Structure se asignan arbitrariamente) y “multimodalidad” (los análisis de datos dan dos o más soluciones distintas). De esta forma, para realizar un gráfico promedio de los anteriores, se ha utilizado el programa Clumpak, que elimina el “cambio de etiqueta” e identifica conjuntos de ejecuciones similares, generando así una solución consenso (Kopelman *et al.*, 2015).

Por otra parte, para detectar el efecto de la selección natural, se comparan las frecuencias entre poblaciones mediante la F_{ST} (índice de fijación) de los diferentes SNPs. Para ello, se ha vuelto al documento que tenía el formato Structure y se ha convertido en formato BayeScan mediante el uso del programa PGDSpider. Este documento se ha introducido en el programa BayeScan, que utiliza un método bayesiano para estimar la probabilidad de que cada locus esté sujeto a selección (Foll y Gaggiotti, 2008). Utilizando el programa RStudio, se ha realizado un gráfico con estos resultados, para lo que se ha utilizado la función “plot_bayescan”. Además, de los 3 documentos que devuelve el programa Bayescan, se ha utilizado el que se llama “*Nombreakchivo_fst*” para obtener una tabla en Excel con la que se pueda trabajar. De esta tabla se han tomado los SNPs con una F_{ST} estadísticamente más alta o más baja que la media (q valor menor que 0,05).

Finalmente, se ha representado su presencia en las diferentes poblaciones mediante el buscador *Geography of Genetic Variants* (GGV), que muestra la distribución geográfica de una variante genética en un mapa (Marcus y Novembre, 2016).

RESULTADOS

Tras la búsqueda en la base de datos 1000 Genomas, se seleccionaron un total de 129 SNPs repartidos por todo el cromosoma, a excepción de la zona del centrómero, que es pobre en genes. Estos SNPs tienen un MAF mínimo de 0,4, y están separados por, al menos, 1cM. Por lo tanto, para realizar este trabajo, se han obtenido 323.016 genotipos.

A continuación, estos polimorfismos se sometieron a un análisis de desequilibrio de ligamiento, cuyo resultado se muestra en la Figura 1. Ya sea atendiendo al color de cada cuadrado (ver Tabla 1) o al número que aparece en su interior, se puede observar que la probabilidad de ligamiento es muy baja en casi todos los casos, salvo alguna excepción puntual.

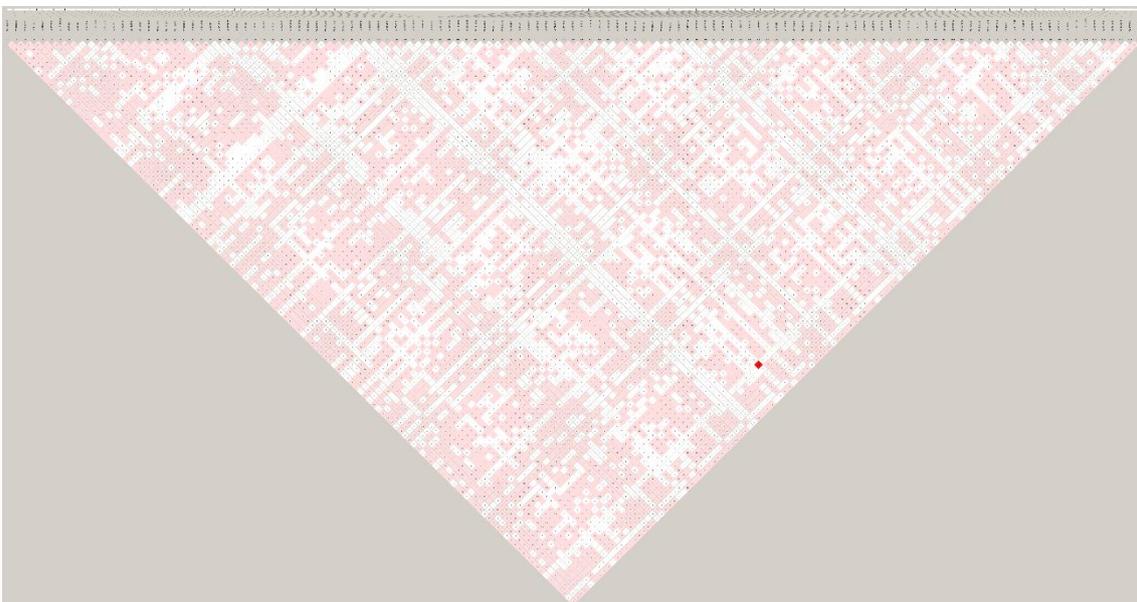


Figura 1. Análisis de desequilibrio de ligamiento realizado mediante el programa Haploview. En la parte superior se indican los nombres de los SNPs analizados, y debajo se observa el resultado mediante un esquema de colores.

Posteriormente, mediante el programa STRUCTURE se realizaron 80 análisis (desde 1 hasta 8 agrupaciones) en los que se mostraba la agrupación y cercanía de los diferentes individuos y poblaciones. Dado que no todos los valores de K son adecuados, se analizaron los gráficos para ver qué cantidad de grupos genéticos se ajustan mejor a los datos. Para ello, se usa el valor de deltaK, que se calcula como el logaritmo de la probabilidad de los datos para una K dada entre su varianza (Evanno *et al.*, 2005). En la Figura 2 se puede ver que el mayor valor de deltaK se obtiene para K=2, seguido de K=3, por lo que estos son los gráficos más adecuados.

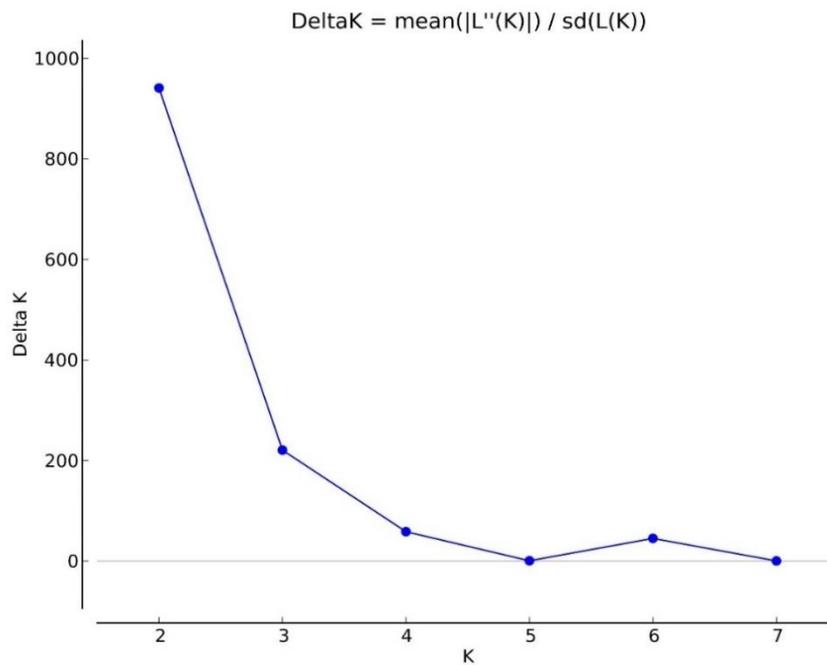


Figura 2. Representación gráfica de deltaK frente a K, donde se observa que el valor de K más adecuado es 2, seguido de K = 3.

Con los 20 análisis seleccionados (10 de cada valor de K) se realizaron 2 gráficos promedio, uno para K=2 y otro para K=3, que se pueden ver en la Figura 3. En el gráfico de K=2 se puede observar que se separan claramente las poblaciones africanas (azul) de las demás (naranja). Esto sería indicativo del origen africano de las poblaciones (Gibbons, 2017). Las poblaciones de Barbados (ACB) y afroamericanos (ASW) son poblaciones americanas. Sin embargo, dado que se sabe que tienen un componente africano muy grande (Murray *et al.*, 2010), se han colocado en el gráfico como africanas. Esto es lo que se puede observar en la imagen. Dentro de estas se puede ver que los afroamericanos tienen cerca de un 35% de componente no africano, mientras que la población de Barbados sólo tiene un 15% de componente no africano. Asimismo, cabe destacar que hay una ligera mezcla con africanos en algunas de las poblaciones no africanas. Entre estas, en la muestra de Puerto Rico se observa una mezcla de alrededor del 15% con población africana, y, en la muestra de Colombia, alrededor del 5%.

Por otra parte, el gráfico K=3 clasifica las poblaciones no africanas. Se puede ver que aparecen separadas las poblaciones europeas (naranja) y las poblaciones del este asiático (China, Vietnam y Japón) (morado). En un lugar intermedio se pueden ver las poblaciones de Sri Lanka, Bangladesh, India y Pakistán, como también se ha observado en otros estudios (Li *et al.*, 2008). En cuanto a las poblaciones americanas, se puede observar un alto porcentaje de origen europeo, debido principalmente a los procesos de migración ocurridos desde 1492.

Asimismo, se aprecia una parte de origen asiático. Esto se explica porque el poblamiento de América se dio a partir de Asia, mediante una población ancestral que cruzó el Estrecho de Bering hace más de 15000 años. De esta forma, este componente representa el origen de los indígenas americanos (Reich *et al.*, 2012).

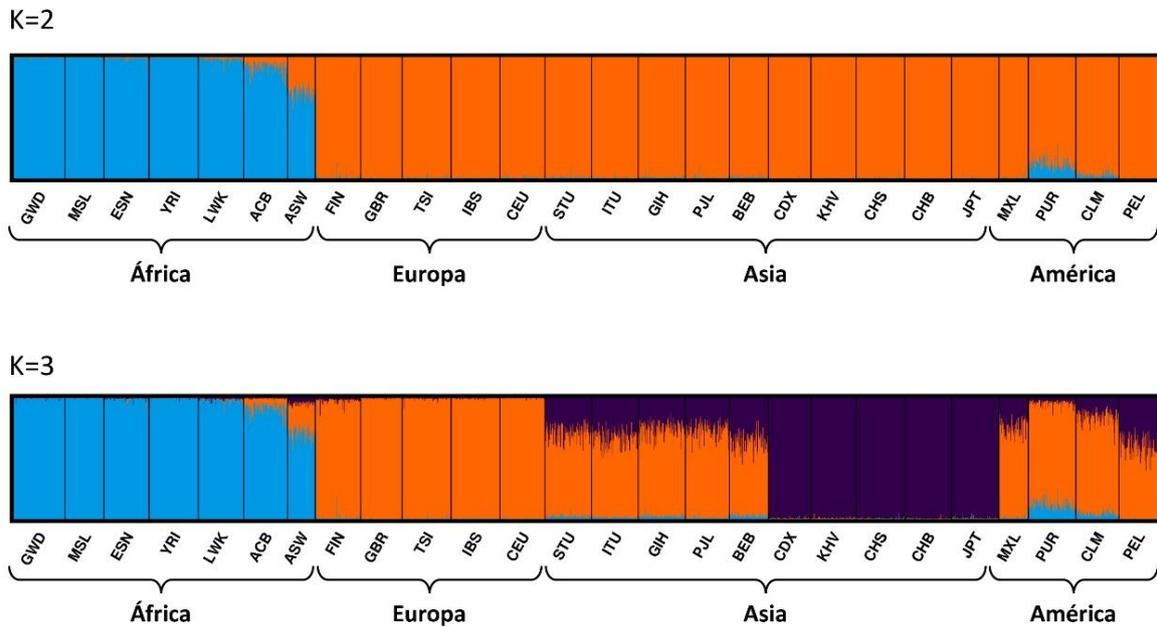


Figura 3. Gráfico Structure realizado para 2 y 3 componentes ($K = 2$ y $K = 3$). Se pueden ver las 26 poblaciones ordenadas según el continente, así como los códigos de las mismas.

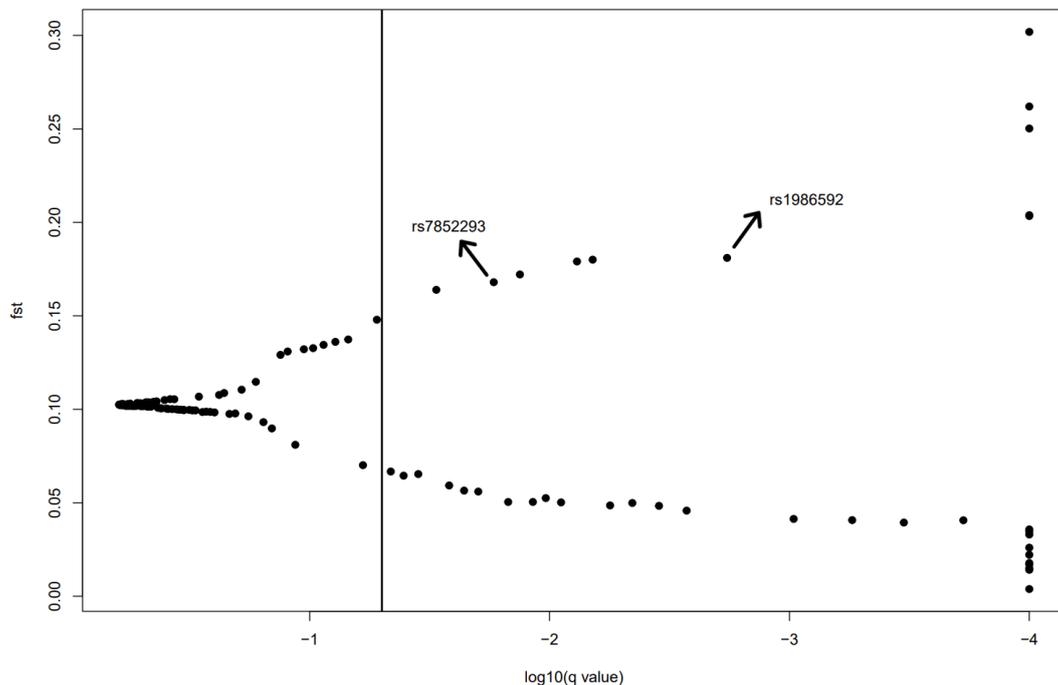


Figura 4. Representación de F_{ST} frente al logaritmo del q valor para los 129 SNPs utilizando los valores obtenidos mediante el programa BayeScan. La línea vertical representa el logaritmo decimal de $q = 0,05$, cuyo valor es $-1,3$.

Por otra parte, para ver los efectos de la selección natural se calculó la F_{ST} para cada polimorfismo, obteniéndose un valor promedio de 0,0985. En la Figura 4 se representan todos los SNPs seleccionados, y aparecen a la derecha de la línea vertical aquellos que tienen una F_{ST} significativamente diferente de los demás. Dentro de estos, en la parte superior están aquellos cuya frecuencia alélica se diferencia más de lo esperado (F_{ST} mayor que la media). En la parte inferior, sin embargo, se sitúan aquellos que están más conservados de lo esperado (F_{ST} menor que la media).

Como se puede observar, 11 de ellos tienen una F_{ST} significativamente mayor que la media (Tabla 2), y 29, una F_{ST} significativamente menor que la media.

Tabla 2. Se muestran la localización, el gen al que pertenecen, el valor de F_{ST} y el q valor de los 11 SNPs con una F_{ST} mayor que la media.

SNP	Localización	Gen	F_{ST}	q valor
rs1986592	2.103.950	SMARCA2	0,1810	0,00182
rs7852293	4.299.453	GLIS3	0,1680	0,01709
rs2762462	12.699.776	TYRP1	0,1791	0,00768
rs6474916	15.424.291	SNAPC3	0,1639	0,02966
rs4977421	20.341.734	MLLT3	0,2620	< 0,00001
rs1888329	31.372.120	LINC01243	0,1800	0,00661
rs4745462	78.511.028	PCSK5	0,1721	0,01329
rs4877796	86.304.008	UBQLN1	0,2039	0,00003
rs954585	109.642.696	ZNF462	0,2034	0,00009
rs9792422	123.891.984	CNTRL	0,2502	< 0,00001
rs2153920	125.991.542	STRBP	0,3019	< 0,00001

De estos SNPs se han seleccionado 2 para profundizar en las causas por las que presentan diferencias entre poblaciones: rs1986592 (del gen SMARCA2) y rs7852293 (del gen GLIS3).

DISCUSIÓN

Este trabajo pretende detectar el efecto de la selección natural en poblaciones humanas. Con dicho fin, se han obtenido genotipos de 2504 individuos de 26 poblaciones para 129 SNPs presentes en el cromosoma 9. Tras confirmar que no estaban ligados mediante un análisis de desequilibrio de ligamiento, se han calculado los valores de F_{ST} . Este estadístico mide la diferenciación de la población, definida como la relación entre comunidad reproductiva y la varianza total. El valor medio de F_{ST} entre las poblaciones es 0,0985. Esta diferenciación habrá

ocurrido por efecto de los mecanismos microevolutivos (mutación, deriva génica y flujo génico). No obstante, se han detectado aquellos SNPs que se diferencian más que la media, lo cual no puede explicarse más que por efecto de la selección natural diversificadora.

De los 11 SNPs para los que se ha demostrado que la selección natural ha afectado favoreciendo un alelo en unas poblaciones y otro alelo en otras diferenciándolas, se han seleccionado rs1986592 del gen SMARCA2 y rs7852293 del gen GLIS3 para profundizar en su estudio.

GEN GLIS3

Mediante los análisis realizados, se ha calculado la F_{ST} entre las poblaciones y se ha obtenido un valor de 0,168 con un q valor de 0,01709 para el SNP rs7852293 localizado en el gen GLIS3. Esto demuestra que los grupos analizados son distintos, por lo que la selección natural ha actuado sobre este gen diferenciando las poblaciones.

La distribución de la frecuencia alélica de este polimorfismo se puede ver en la Figura 5. Llama la atención que el alelo C es dominante en la mayoría de las poblaciones. Sin embargo, en África, este alelo es el menos común. Asimismo, dentro de las poblaciones de América, sólo aquellas con un alto porcentaje de origen africano tienen una alta frecuencia del alelo T.

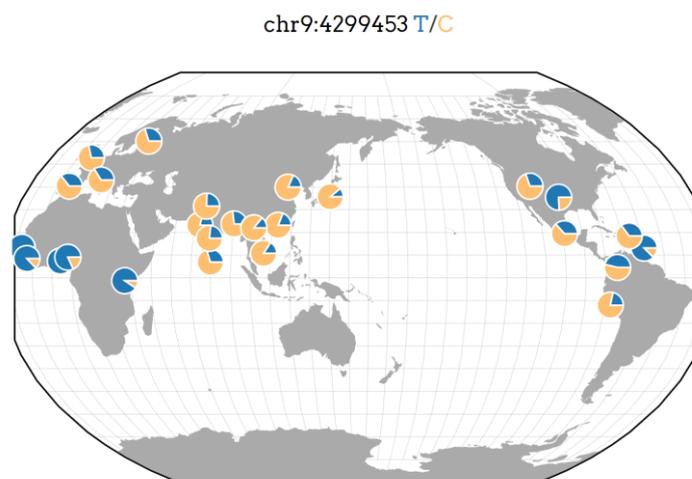


Figura 5. Representación de las frecuencias del SNP rs7852293 en las diferentes poblaciones, con el alelo T en azul y el C en amarillo.

El gen en el que se encuentra este polimorfismo es GLIS3, un gen es miembro de la familia de proteínas con dedos de zinc similares a Krüppel que se compone de tres proteínas: Glis1-3. Codifica una proteína nuclear que funciona como represor y activador de la transcripción y está específicamente involucrada en el desarrollo de las células beta del páncreas, la tiroides, los ojos, el hígado y los riñones (Alghamdi *et al.*, 2017; Weizmann Institute of Science, 2023).

Mutaciones en este gen se han relacionado con el desarrollo de diabetes. De hecho, la eliminación genética de GLIS3 en ratones y humanos induce diabetes neonatal, mientras que los SNPs se han asociado con diabetes tipo 1 y tipo 2. La deficiencia de GLIS3 reduce significativamente la generación de células endocrinas, en particular de células β , lo que provoca hiperglucemia (Scoville y Jetten, 2021).

Salvo en contadas excepciones (causas monogénicas), la diabetes se debe principalmente a una combinación de factores genéticos y ambientales (como la dieta). Asimismo, varios estudios en EEUU han demostrado que los afroamericanos tienen una mayor prevalencia e incidencia de diabetes que los blancos caucásicos (Scoville y Jetten, 2021; Dibato *et al.*, 2023).

Se han desarrollado varias hipótesis para explicar el beneficio evolutivo de la diabetes, entre las que destaca la hipótesis del genotipo ahorrador de Neel (1962). Según esta hipótesis, la diabetes tipo 2 es el resultado de un genotipo ahorrador que se ha vuelto perjudicial debido al progreso. La sobreproducción de insulina es el mecanismo conservador de la energía cuando la ingesta de alimentos es irregular. El genotipo fue seleccionado para mantener la homeostasis de la glucosa y proteger al organismo de la escasez de alimentos a través de la resistencia a la insulina. De esta forma, la rápida descarga de insulina en respuesta a la hiperglucemia fue ventajosa en el pasado, ya que permitía acumular depósitos de grasa en tiempos de abundancia para su uso eficiente en momentos de hambruna y escasez. Sin embargo, cuando los individuos adaptados a este tipo de ambientes migran a otros lugares con un mejor acceso a los alimentos, puede dar lugar a la aparición de diabetes (Neel, 1962).

En resumen, se ha demostrado que la selección natural ha afectado al gen GLIS3 a lo largo de la historia evolutiva del ser humano diferenciando las poblaciones de distintas regiones, y que podría estar relacionado con una adaptación a la dieta y a la escasez de recursos.

GEN SMARCA2

Con los análisis realizados se ha calculado la F_{ST} entre las poblaciones y se ha obtenido un valor de 0,181 con un q valor de 0,00182 para el SNP rs1986592 localizado en el gen SMARCA2. Esto demuestra que los grupos analizados son distintos, por lo que la selección natural ha actuado sobre este gen diferenciando las poblaciones.

La distribución de la frecuencia alélica de este polimorfismo se puede ver en la Figura 6. Destaca que el alelo C es el más frecuente en las poblaciones africanas, mientras que en las poblaciones europeas y del oeste asiático lo es el alelo T. En las poblaciones de Asia del este, sin embargo, se observa que la frecuencia de los dos alelos está equilibrada. Por otra parte, en

las poblaciones americanas se observan valores intermedios. Aquellas con un alto porcentaje africano se puede ver que son más parecidas a las poblaciones africanas. Sin embargo, el resto de poblaciones de América se parecen más a las asiáticas y europeas.

chr9:2103950 C/T

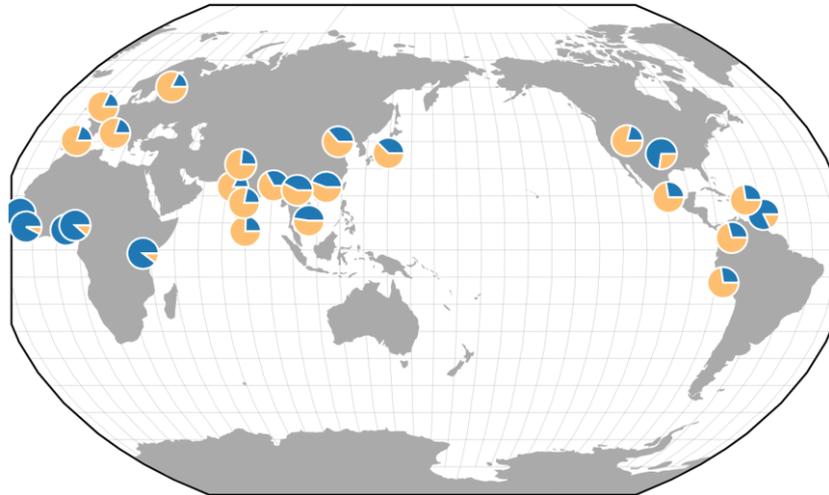


Figura 6. Representación de las frecuencias del SNP rs1986592 en las diferentes poblaciones, con el alelo C en azul y el T en amarillo.

El gen SMARCA2 es un regulador de cromatina dependiente de actina. La proteína sintetizada por este gen está involucrada en la activación transcripcional y la represión de ciertos genes mediante la remodelación de la cromatina. Forma parte del complejo de remodelación de cromatina SWI/SNF, que lleva a cabo actividades enzimáticas clave cambiando la estructura de la cromatina al alterar los contactos de las histonas de ADN dentro de los nucleosomas (UniProt Consortium, 2023). Se ha observado que interacciona con C/EBP- β (o CEBPB). El TAD (dominio de transactivación) del factor de transcripción específico de la secuencia de ADN C/EBP- β recluta el complejo SWI/SNF para iniciar la activación de genes silenciosos en la cromatina. Además, se ha descubierto que el TAD interactúa directamente con SMARCA2 (Kowenz-Leutz y Leutz, 1999).

C/EBP- β es un gen que codifica un factor de transcripción, cuya actividad es importante en la regulación de genes involucrados en varias funciones: reproducción femenina, función y regeneración del hígado, adipogénesis, hematopoyesis, así como en respuestas inmunes (innatas y adaptadas) e inflamatorias (Pless *et al.*, 2008). Entre estas funciones destacan las dos últimas.

En cuanto a la respuesta inflamatoria, durante la fase inicial los homodímeros de C/EBP activan diversos promotores de genes asociados con la inflamación. Durante la fase tardía, C/EBP- β se induce y reemplaza uno de los componentes del homodímero, para dar lugar a un activador más potente y para regular estos genes recién activados (Kinoshita *et al.*, 1992).

Por otra parte, en cuanto a la respuesta inmune, los macrófagos adquieren una mayor capacidad bactericida y tumoricida cuando se activan por completo tras su interacción con microorganismos o tras la estimulación por moléculas, entre las que destacan citocinas como el interferón- γ (IFN- γ). IFN- γ es una citocina producida por las células T y las células *natural killer*, que es fundamental para mediar las funciones antivirales y antitumorales, y que induce la expresión de C/EBP- β . Además, se han identificado varios factores de transcripción (entre los que se incluye C/EBP- β) que pueden desempeñar un papel en la regulación del programa de diferenciación y activación de los macrófagos. Se cree que C/EBP- β puede ser necesario para la expresión coordinada de un grupo de genes específicos de macrófagos, así como para la adquisición del fenotipo de macrófagos (Roy *et al.*, 2000; Tanaka *et al.*, 1995).

Cabe destacar que se ha demostrado la importancia de C/EBP- β (en el artículo se denomina NF-IL6) mediante el uso de ratones C/EBP- β (-/-). Las bacterias, tras ser fagocitadas mueren o se inhibe su crecimiento, para lo cual son críticos IFN- γ y TFN α . Sin embargo, los ratones C/EBP- β (-/-) tienen defectos para matar bacterias y tumores. Asimismo, se ha visto que es posible que C/EBP- β sea responsable de la expresión de componentes del mecanismo que mejora la actividad bactericida mediada por NO (Tanaka *et al.*, 1995).

La relación entre la función inmunitaria de este gen y la distribución de los alelos estaría en que a medida que los humanos salieron de África, las poblaciones se encontraron con distintos patógenos, y la selección natural aumentó la prevalencia de alelos ventajosos para los nuevos ecosistemas (Karlsson *et al.*, 2014). Si bien es cierto que también se encontraron con diversas condiciones climáticas y nutricionales, los estudios de genética de poblaciones han indicado que la inmunidad y las funciones de defensa del huésped se encuentran entre las funciones más frecuentemente sujetas a selección natural (Quach y Quintana-Murci, 2017). De hecho, los patógenos y las enfermedades infecciosas han sido y siguen siendo las principales amenazas para la salud y la supervivencia humanas (Barreiro y Quintana-Murci, 2009). A principios del siglo XX, durante el proceso de industrialización, las enfermedades daban como resultado una esperanza de vida promedio de alrededor de 40 años, llegando a ser tan sólo de 20 años durante algunas epidemias (Siddle y Quintana-Murci, 2014)). Hoy en día, existen algunas poblaciones como los Aeta que son cazadores-recolectores, y que tienen una esperanza de vida promedio al nacer de 16,5 años (Migliano *et al.*, 2007). Además en la actualidad, a pesar del progreso médico, las enfermedades infecciosas continúan siendo una de las principales causas de mortalidad, siendo responsables del 48% de las muertes de personas menores de 45 años en todo el mundo (Prugnolle *et al.*, 2005). Sin duda, a lo largo de la historia de la especie

humana, la selección natural ha modificado de forma importante la frecuencia alélica de algunos genes.

Al analizar los genes que afectan a la inmunidad, es más probable que se observen gradientes de selección direccional positiva en poblaciones con alta prevalencia y abundancia de patógenos, mientras que las poblaciones con menor riesgo de infección podrían estar sujetas a una selección estabilizadora, o incluso, a una selección direccional negativa (Seppälä, 2014).

En cuanto a las causas por las que se ha podido seleccionar este SNP, pueden ser diversas. Por una parte, se ha observado que las personas de ascendencia africana generalmente muestran una respuesta más fuerte a la infección bacteriana que las personas de ascendencia europea, en particular para los genes de respuesta inflamatoria (Quach y Quintana-Murci, 2017). En este caso, se podría decir que el alelo C de este SNP favorece una respuesta más fuerte a la infección bacteriana. Debido a que tras la salida de los humanos de África éstos se encontraron con otros microorganismos diferentes, se pudo favorecer una respuesta menos intensa.

Por otra parte, también es posible que la causa de su selección haya sido un microorganismo concreto. El modo por el cual el ser humano se adapta a los patógenos depende de una gran cantidad de factores, como el tipo de microorganismo, la diferente presencia temporal y espacial, la patogenicidad variable, la virulencia, las estrategias de infección y la velocidad a la que evolucionan. De esta forma, diferentes especies y tipos de patógenos pueden imponer diferentes presiones selectivas sobre la función inmunológica (Barreiro y Quintana-Murci, 2017; Sepälä, 2014). Sin embargo, como en muchos otros casos, a pesar de que los estudios de correlación respaldan firmemente la selección impulsada por patógenos, la identificación de aquellos que ejercen presión sobre genes concretos sigue siendo ardua (Siddle y Quintana-Murci, 2014). De esta forma, pueden ser muy variados los microorganismos que hayan provocado esta selección, especialmente patógenos que han matado y matan a una gran cantidad de personas. Entre estos (aunque pueden ser muchos otros) se encuentran *Plasmodium* que provoca malaria en África sub-Sahariana y en el sudeste asiático o *Yersinia pestis* que provoca la peste negra. De hecho, en el caso de la peste negra, cabe destacar que algunas bacterias sobreviven tras ser fagocitadas por los macrófagos y se propagan al ganglio linfático, donde se replican sin control (Barreiro y Quintana-Murci, 2009; Klunk *et al.*, 2022). Como ya se ha comentado anteriormente, este es uno de los resultados observados en el estudio realizado por Tanaka *et al.* (1995).

En resumen, se ha demostrado que la selección natural ha afectado al gen SMARCA2 a lo largo de la historia del ser humano diferenciando las poblaciones de distintas regiones, y que podría estar relacionado con una adaptación a la resistencia a microorganismos.

CONCLUSIONES

A partir de los resultados obtenidos, se pueden extraer varias conclusiones acerca de los dos genes que se han analizado más profundamente.

En primer lugar, en cuanto a las frecuencias del SNP del gen GLIS3, cabe destacar la diferencia de las poblaciones de origen africano respecto al resto de poblaciones. Este gen está involucrado en el desarrollo de células β del páncreas, y la presencia de ciertos polimorfismos se ha asociado con la aparición de diabetes. Estas observaciones pueden explicarse mediante la hipótesis del genotipo ahorrador, que afirma que la sobreproducción de insulina es el mecanismo conservador de energía cuando la ingesta de alimentos es irregular. Por lo tanto, podría ser una adaptación a la dieta y escasez de recursos del continente africano.

En segundo lugar, en cuanto a las frecuencias del SNP del gen SMARCA2, se ha observado que predomina uno de los alelos en las poblaciones de origen africano, mientras que en las poblaciones europeas y de Asia occidental este alelo es minoritario. En Asia oriental, sin embargo, se han dado valores intermedios. Este gen interactúa con C/EBP- β , que participa en la respuesta inmune. A lo largo de la historia del ser humano, los patógenos y las enfermedades infecciosas han sido las principales amenazas para la salud y supervivencia, y de hecho lo siguen siendo. Por ello, este gen podría ser una adaptación a la resistencia a microorganismos.

El análisis del genoma y su relación con la selección natural pueden aportar información fundamental acerca de la historia evolutiva del ser humano. Asimismo, en este trabajo se han detectado otros genes en los que ha actuado la selección natural, por lo que sería de interés profundizar en ellos para entender mejor la evolución del ser humano.

BIBLIOGRAFÍA

- Achilli, A., Perego, U. A., Lancioni, H., Olivieri, A., Gandini, F., Kashani, B. H., ... Torroni, A. (2013). Reconciling migration models to the Americas with the variation of North American native mitogenomes. *Proceedings of the National Academy of Sciences of the United States of America*, 110(35), 14308-14313. <https://doi.org/10.1073/pnas.1306290110>

- Alghamdi, K. A., Alsaedi, A. B., Aljasser, A., Altawil, A., & Kamal, N. M. (2017). Extended clinical features associated with novel GLIS3 mutation: a case report. *BMC Endocrine Disorders*, 17(1). <https://doi.org/10.1186/s12902-017-0160-z>
- Barreiro, L. B., & Quintana-Murci, L. (2009). From evolutionary genetics to human immunology: how selection shapes host defence genes. *Nature Reviews Genetics*, 11(1), 17-30. <https://doi.org/10.1038/nrg2698>
- Barrett, J. C., Fry, B., Maller, J., & Daly, M. (2004). HaploView: Analysis and visualization of LD and Haplotype maps. *Bioinformatics*, 21(2), 263-265. <https://doi.org/10.1093/bioinformatics/bth457>
- Clarkson, C., Jacobs, Z., Marwick, B., Fullagar, R., Wallis, L. A., Smith, M., ... Pardoe, C. (2017). Human occupation of Northern Australia by 65,000 years ago. *Nature*, 547(7663), 306-310. <https://doi.org/10.1038/nature22968>
- De Castro, J. N., & Martínón-Torres, M. (2022). The origin of the Homo sapiens lineage: When and where? *Quaternary International*, 634, 1-13. <https://doi.org/10.1016/j.quaint.2022.08.001>
- Dibato, J., Montvida, O. & Sanjoy, P. K. (2023). Racial disparity in the co-occurrence of depression and type 2 diabetes mellitus. An electronic medical record study involving African American and White Caucasian adults from the US. *Journal of Affective Disorders*, 330, 173-179. <https://doi.org/10.1016/j.jad.2023.02.097>
- Earl, D., & vonHoldt, B. M. (2011). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, 4(2), 359-361. <https://doi.org/10.1007/s12686-011-9548-7>
- Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology*, 14(8), 2611-2620. <https://doi.org/10.1111/j.1365-294x.2005.02553.x>
- Falush, D., Stephens, M., & Pritchard, J. K. (2007). Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular Ecology Notes*, 7(4), 574-578. <https://doi.org/10.1111/j.1471-8286.2007.01758.x>
- Foll, M., & Gaggiotti, O. E. (2008). A Genome-Scan method to identify selected Loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics*, 180(2), 977-993. <https://doi.org/10.1534/genetics.108.092221>
- Fontdevila, A., & Moya, A. (2000). *Introducción a la genética de poblaciones*. Editorial síntesis

- Gibbons, A. (2017). Oldest members of our species discovered in Morocco. *Science*, 356(6342), 993-994. <https://doi.org/10.1126/science.356.6342.993>
- Hershkovitz, I., Weber, G., Quam, R., Duval, M., Grün, R., Kinsley, L., ... Weinstein-Evron, M. (2018). The earliest modern humans outside Africa. *Science*, 359(6374), 456-459. <https://doi.org/10.1126/science.aap8369>
- Jiménez, C. B., & Espino, F. J. (2012). *Genética: conceptos esenciales*. Editorial medica panamericana
- Karlsson, E. K., Kwiatkowski, D. P., & Sabeti, P. C. (2014). Natural selection and infectious disease in human populations. *Nature Reviews Genetics*, 15(6), 379-393. <https://doi.org/10.1038/nrg3734>
- Kinoshita, S., Akira, S., & Kishimoto, T. (1992). A member of the C/EBP family, NF-IL6 beta, forms a heterodimer and transcriptionally synergizes with NF-IL6. *Proceedings of the National Academy of Sciences of the United States of America*, 89(4), 1473-1476. <https://doi.org/10.1073/pnas.89.4.1473>
- Klunk, J., Vilgalys, T. P., Demeure, C. E. *et al.* (2022). Evolution of immune genes is associated with the Black death. *Nature*, 611(7935), 312-319. <https://doi.org/10.1038/s41586-022-05349-x>
- Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak: A program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, 15(5), 1179-1191. <https://doi.org/10.1111/1755-0998.12387>
- Kowenz-Leutz, E., & Leutz, A. (1999). A C/EBP β Isoform Recruits the SWI/SNF Complex to Activate Myeloid Genes. *Molecular Cell*, 4(5), 735-743. [https://doi.org/10.1016/s1097-2765\(00\)80384-6](https://doi.org/10.1016/s1097-2765(00)80384-6)
- Li, J., Absher, D., Tang, H., Southwick, A., Casto, A. M., Ramachandran, S., ... Myers, R. M. (2008). Worldwide human relationships inferred from Genome-Wide patterns of variation. *Science*, 319(5866), 1100-1104. <https://doi.org/10.1126/science.1153717>
- Lischer, H. E. L., & Excoffier, L. (2011). PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, 28(2), 298-299. <https://doi.org/10.1093/bioinformatics/btr642>
- Liu, W., Martín-Torres, M., Cai, Y., Xing, S., Tong, H., Pei, S., ... Wu, X. (2015). The earliest unequivocally modern humans in southern China. *Nature*, 526(7575), 696-699. <https://doi.org/10.1038/nature15696>

- Lodish, H., Berk, A., Kaiser, C. A., Krieger, M., Bretscher, A., Ploegh, H., Amon, A., & Scott, M. P. (2013). *Molecular Cell Biology*. New York: W. H. Freeman.
- Marcus, J., & Novembre, J. (2016). Visualizing the geography of genetic variants. *Bioinformatics*, 33(4), 594-595. <https://doi.org/10.1093/bioinformatics/btw643>
- Masimirembwa, C., Dandara, C., & Hasler, J. A. (2014). Population Diversity and Pharmacogenomics in Africa. *Elsevier eBooks*, 971-998. <https://doi.org/10.1016/b978-0-12-386882-4.00043-8>
- Metz, L., Lewis, J., & Slimak, L. (2023). Bow-and-arrow, technology of the first modern humans in Europe 54,000 years ago at Mandrin, France. *Science Advances*, 9(8). <https://doi.org/10.1126/sciadv.add4675>
- Migliano, A. B., Vinicius, L., & Lahr, M. M. (2007). Life history trade-offs explain the evolution of human pygmies. *Proceedings of the National Academy of Sciences of the United States of America*, 104(51), 20216-20219. <https://doi.org/10.1073/pnas.0708024105>
- Miller, J., Wood, B., & Hamilton, M. J. (2008). F_{ST} and Q_{ST} under neutrality. *Genetics*, 180(2), 1023-1037. <https://doi.org/10.1534/genetics.108.092031>
- Murray, T., Beaty, T. H., Mathias, R. A., Rafaels, N., Grant, A. V., Faruque, M. U., ... Barnes, K. C. (2010). African and non-African admixture components in African Americans and an African Caribbean population. *Genetic Epidemiology*, 34(6), 561-568. <https://doi.org/10.1002/gepi.20512>
- Neel, J. V. (1962). Diabetes Mellitus: A “Thrifty” Genotype Rendered Detrimental by “Progress”? *American journal of human genetics*, 14(4), 353–362.
- Pierce, B. (2023). *Fundamentos de Genética*. Editorial medica panamericana
- Pless, O., Kowenz-Leutz, E., Knoblich, M., Lausen, J., Beyermann, M., Walsh, M. J., & Leutz, A. (2008). G9a-mediated Lysine Methylation Alters the Function of CCAAT/Enhancer-binding Protein-β. *Journal of Biological Chemistry*, 283(39), 26357-26363. <https://doi.org/10.1074/jbc.m802132200>
- Prugnolle, F., Manica, A., Charpentier, M. J. E., Guégan, J., Guernier, V., & Balloux, F. (2005). Pathogen-Driven selection and worldwide HLA Class I diversity. *Current Biology*, 15(11), 1022-1027. <https://doi.org/10.1016/j.cub.2005.04.050>
- Quach, H., & Quintana-Murci, L. (2017). Living in an adaptive world: Genomic dissection of the genus Homo and its immune response. *Journal of Experimental Medicine*, 214(4), 877-894. <https://doi.org/10.1084/jem.20161942>

- Quintana-Murci, L. (2016). Genetic and Epigenetic Variation of Human Populations: An Adaptive Tale. *Comptes Rendus Biologies*, 339(7-8), 278-283. <https://doi.org/10.1016/j.crv.2016.04.005>
- Reich, D., Price, A. L., Campbell, D., Tandon, A., Mazières, S., Ray, N., ... Ruiz-Linares, A. (2012). Reconstructing Native American population history. *Nature*, 488(7411), 370-374. <https://doi.org/10.1038/nature11258>
- Roy, S. K., Sj, W., Weihua, X., Hu, J., & Kalvakolanu, D. V. (2000). CCAAT/Enhancer-binding Protein- β Regulates Interferon-induced Transcription through a Novel Element. *Journal of Biological Chemistry*, 275(17), 12626-12632. <https://doi.org/10.1074/jbc.275.17.12626>
- Scoville, D. K., & Jetten, A. M. (2021). GLIS3: a critical transcription factor in Islet B-Cell generation. *Cells*, 10(12), 3471. <https://doi.org/10.3390/cells10123471>
- Seppälä, O. (2014). Natural selection on quantitative immune defence traits: a comparison between theory and data. *Journal of Evolutionary Biology*, 28(1), 1-9. <https://doi.org/10.1111/jeb.12528>
- Siddle, K. J., & Quintana-Murci, L. (2014). The Red Queen's Long Race: Human adaptation to Pathogen pressure. *Current Opinion in Genetics & Development*, 29, 31-38. <https://doi.org/10.1016/j.gde.2014.07.004>
- Tanaka, T., Akira, S., Yoshida, K., Umemoto, M., Yoneda, Y., Shirafuji, N., Fujiwara, H., Suematsu, S., Yoshida, N., & Kishimoto, T. (1995). Targeted disruption of the NF-IL6 gene discloses its essential role in bacteria killing and tumor cytotoxicity by macrophages. *Cell*, 80(2), 353-361. [https://doi.org/10.1016/0092-8674\(95\)90418-2](https://doi.org/10.1016/0092-8674(95)90418-2)
- The 1000 Genomes Project Consortium (2010). A map of human genome variation from population-scale sequencing. *Nature* 467, 1061–1073. <https://doi.org/10.1038/nature09534>
- The 1000 Genomes Project Consortium. (2015). A global reference for human Genetic Variation. *Nature*, 526(7571), 68-74. <https://doi.org/10.1038/nature15393>
- UniProt Consortium. (2023). P51531 · SMCA2_HUMAN. *UniProt*. <https://www.uniprot.org/uniprotkb/P51531/entry> (último acceso 04/07//2023).
- Weber, J. L., David, D. E., Heil, J., Fan, Y., Zhao, C., & Marth, G. T. (2002). Human Diallelic Insertion/Deletion Polymorphisms. *American Journal of Human Genetics*, 71(4), 854-862. <https://doi.org/10.1086/342727>
- Weizmann Institute of Science. (2023). GLIS3. GeneCards The Human Database. <https://www.genecards.org/cgi-bin/carddisp.pl?gene=GLIS3> (último acceso 15/07/2023).

ANEXO

Tabla con el código y el número de individuos genotipados de cada población.

Código	N	Población	Código	N	Población
ACB	96	Barbados	IBS	107	España
ASW	61	Afroamericanos	ITU	102	India
BEB	86	Bengali, Bangladesh	JPT	104	Japón
CDX	93	Dai, China	KHV	99	Kinh, Vietnam
CEU	99	Europeos de América	LWK	99	Luhya, Kenya
CHB	103	Beijing, China	MSL	85	Mende, Sierra Leona
CHS	105	Han, China	MXL	64	California, EEUU
CLM	94	Colombia	PEL	85	Lima, Perú
ESN	99	Esan, Nigeria	PJL	96	Punjabi, Pakistán
FIN	99	Finlandia	PUR	104	Puerto Rico
GBR	91	Reino Unido	STU	102	Sri Lanka
GIH	103	Gujarati, India	TSI	107	Toscani, Italia
GWD	113	Gambia	YRI	108	Yoruba, Nigeria